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 the complex composition of immune repertoire and the difficulty of analysis. To explore our understanding of bats' t-cell immune system, we established a reference library of TCR-β germline genes based on the gene annotation of *Rhinolophus ferrumequinum* (Rhinolophidae), *Phyllostomus discolor* (Phyllostomidae) and *Pipistrellus pipistrellus* (Vespertilionidae). Further, we undertook the analysis of the Rhinolophus affinis TCR-β repertoire. The bats' TCR-β repertoire were formed by the rearrangement of the V-D-J-C genes with a D-J/V-D deletion and insertion, showing a highly 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 be applied to the study of adaptive immune system in bats.

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(*11-13*), among others. However, specific viral response studies in bats have been less frequently reported(*14-16*). The main reasons are the lack of relevant reagents in cell biology studies of bats and disease models in bats, the difficulty of isolating virus strains in bats itself, and the huge diversity of bat species (2 orders, 21 families, more than 1350 species) (*17*). The difficulty of conducting bat-specific immune responses, coupled with the very limited information of the bat genome, has resulted in a near gap in the study of bat characteristics from TCR and BCR.

With the development of Next Generation Sequencing (NGS), High-Throughput Sequencing(HTS) technologies have been widely used for the analysis of the immune system, especially immune repertoire (*18*). 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, and deep sequencing of the repertoire has been widely used in the resolution of the adaptive immune system, assessment of individual health status, antibody development, and targeted detection and therapy (*19*)。IMGT/HighV-QUEST, TRUST, MiXCR, etc., various tools have been used to analyze HTS TCR/BCR repertoire (*20-24*)。In recent years, the bat1k project has published some of the high-quality bat genomic data, which has greatly helped our understanding of the bat immune system (*25, 26*), 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 (*27, 28*). 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 with mammals such as humans and mice.

In our previous work, we completed the annotation of T-cell receptor germline genes in the *Rhinolophus ferrumequinum* (Greater Horseshoe Bat), 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 (*28*). On this basis, we systematically annotated the TRB loci of the *Phyllostomus discolor*(Pale spear-nosed bat) and *Pipistrellus pipistrellus*(Common pipistrelle), and compared the TRB loci of the three bats at the genomic level. Using the conserved TRBC gene, we compared the TRBC Exon1 sequences of various bats, and designed primers to amplify the TCR β-chain CDR3 repertoire of *Rhinolophus affinis*(Intermediate horseshoe bat) by 5' rapid amplification of cDNA ends(5'race) method. This 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 number of germline genes obtained by different methods was counted while exporting Genebank files (Table 1, SupData1). The three bat species did not differ greatly in the structural distribution of TRB loci, and all maintained the classical structure of mammals: multiple TRBV genes involved in rearrangements upstream, a 1-3 D-J-C clusters downstream, and a reverse TRBV gene after the last D-J-C cluster (Figure 1B, Supplementary Figure 1). Consistent with mammals such as human and mouse (*29*). 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.

After naming the TRBV genes according to the ImMunoGeneTics(IMGT) guidelines, the numbers of *Phyllostomus discolor*, *Rhinolophus ferrumequinum*, and *Pipistrellus pipistrellus* TRBV genes were 100 (including 2 pseudogenes whose families could not be identified), 30, and 45, respectively; the nucleotide sequences were found by constructing a phylogenetic tree, and although the V gene family deletions differed among the three bat species, all genes could be found homologous to each other or in pigs and humans, and we did not find V gene families that could cluster separately in bats (Figure 2A).

Internal duplication of gene families is the underlying cause of differences in the number of germline genes, and this duplication process appears to have some degree of commonality. Both in terms of consistency of family duplication among species and overall numbers (Fig. 2B), Chiroptera are more similar to Artiodactyla in terms of the evolutionary direction of TRBV genes, random and large-scale, whereas carnivores are more similar to primates. With the exception of 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 special family, TRBV12 of *Phyllostomus discolor* containing 37 members, which is close to almost all TRBV genes of *Pipistrellus pipistrellus* (45) and even more than the total TRBV genes of *Rhinolophus ferrumequinum*(30), and even in the entire known mammalian range, no TRBV multigene family is found with more than 30 members (the only one 30 members is TRBV5 in goats)(*30*). In the amino acid multiple sequence alignment of TRBV genes, we found that not only the conserved sites 23Cys, 41Trp, 89Leu, 104Cys existed, but also Gln at position 6, Tyr at position 42 was extremely conserved, and Gln at position 44 in the *Pipistrellus pipistrellus* 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 *Phyllostomus discolor* implies an increased probability of gene mutation; therefore, the proportion of TRBV pseudogenes in *Phyllostomus discolor*, 17%, is the highest among the three bat species (7% in *Rhinolophus ferrumequinum* and 11% in *Pipistrellus pipistrellus*), and even so, the number of 97 functional genes (V/D/J genes) in *Phyllostomus discolor* far exceeds that of *Rhinolophus ferrumequinum* and *Pipistrellus pipistrellus* (44, 53) (Table 3).

The D, J, and C genes are named according to the location of the Cluster, and like primates and carnivorous mammals, the TRB locus of 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 genes in the second cluster of *Rhinolophus ferrumequinum*. All TRBJ genes in bats contain conserved FGNG motifs except for that TRBJ2-3(Figure 3B, Supplementary Figure 2B).

We compared the nucleotide sequences of the V and J genes of the three bat species and found that the homology of the same V and J gene families was very high, with most sequences having more than 70% similarity, except for TRBV17 (Figure 2C). Therefore, when constructing the reference database for analyzing the bat TCR repertoire, for the V gene, we grouped each TRBV gene of the three bat species into the large family: TRBV1-1, TRBV1-2, TRBV1-3; as the naming of J gene was determined according to the cluster, the naming method for grouping followed the naming of alleles in IMGT: TRBJ1-1*01, TRBJ1-1*02, TRBJ1-1*03 (Supplementary Table 4), and further a reference library for performing MiXCR tool analysis of bats was produced based on the nucleotide sequences and results of region delineation by AA sequences MSA.

2. TCRβ chain Repertoire Building、Sequencing、Analyzing

The Cytb(Cytochrome b) sequencing results of the 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), and were

determined to be *Rhinolophus affinis* after BLAST (Supplementary Table5). We compared the TRBC Exon1 nucleotide sequences of 15 bat species and designed primers in the conserved regions to construct TCRβ chain repertoire by 5'race, and the nucleotide sequence similarity of 16 TRBC Exon1 ranged from 77.66% to 98.9%, with the presence of a large number of fully conserved regions, suitable for repertoire building and sequencing using conserved regions for 5'race (Figure 4A).
The sequencing process and identification of the TCRβ chain repertoire of *Rhinolophus affinis* were shown in Figure 4B and Supplementary Figure 3, and there were significant peaks at the 600bp-700bp target bands, suggesting successful library construction. The sequenced raw sequences of the nine samples were analyzed by MiXCR. On average, 4.27 million TCR reads and 24904 clonetype were obtained for each bat sample according to the 5'race method; 15 million TCR reads and 1.23 million clonetype were obtained for each mouse sample; 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 sequencing depth of bat samples, all of them met the CDR3 sequence depth analysis requirements for HTS sequencing (Table 2, Table 6).
For the sequencing results of *Rhinolophus affinis*, we first exported 25 V gene family and 14 J gene family sequences contained in the Bat1 sample (Figure 5A), in order to suggest the completeness of the repertoire construction results. After categorizing the sequencing results of the three *Rhinolophus affinis*, we found that most of the sequences belonged to the annotated genes of *Rhinolophus affinis* output sequences belonged to *Rhinolophus ferrumequinum*. To verify the degree of bat species similarity, the Cytb sequencing results of the three experimental bats showed 87.9% and 74.6% similarity with the *Rhinolophus ferrumequinum* and *Pipistrellus pipistrellus*, respectively (Figure 5C), and the sequenced TRBC sequences of the three experimental bats showed 77.1%, 93.8% and 75.0% similarity with *Phyllostomus discolor*, *Rhinolophus ferrumequinum* and *Pipistrellus pipistrellus*, respectively (Figure 5D), the sequence comparison results further indicated that the sequences of *Rhinolophus affinis* were more similar to those of *Rhinolophus ferrumequinum*.
Based on the homology analysis of the phylogenetic tree, we compared the TRBV gene usagees in bats, humans, and mice (Figure 6A), and the results of V gene usage in all three species were neat, and the homologous genes expressed simultaneously were: TRBV3, TRBV13, TRB15, TRBV20, TRBV28, and TRBV30, among which TRBV15 and TRBV30 in the three species except for TRBV15 and TRBV30, which had similar expression ratios in the three species, all other homologous genes had some differences in access. In addition, we analyzed the usage of all TRBV genes in bats and humans (Figure 6B). The top 3 usage in bats were: TRBV12, TRBV18, and TRBV20, and in humans were: TRBV13, TRBV28, and TRBV29. The TRBV2 and TRBV10 gene families were not found in any of the three annotated bat genomes. TRBV9 and TRBV19 were not expressed in either bat1, while there was very low expression in bat2 and bat3 samples, the annotated sequences were only found in *Phyllostomus discolor*, and all four TRBV9 families were pseudogenes, and TRBV17 was very lowly expressed in all three bats; there are several possible explanations for these lowly expressed families, firstly the expression in *Rhinolophus affinis* 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 *Rhinolophus affinis* and therefore cannot be detected, and these mutations may also lead to functional genes becoming pseudogenes, a situation that is very common across species. For example, the TRBV1 gene in bats is a functional gene, while TRBV1 in humans is a pseudogene; however, 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. The three species did not show functional J gene non-expression, and all J gene families were expressed among the three species except TRBJ2-8 (pseudogene) in bats and TRBJ2-6 (pseudogene) in mice (Figure 6C). 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 6D), 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), it was found that among the top 10 high-frequency take motifs, 4 were shared by bat, human, and mouse. (Figure 6E, Figure 6F), we observed that most of the shared motifs among the 3 species were CASSN motifs, and since the CASS motifs in the V gene region belong to the antigen docking function directly related to the CDR3 region, this suggests that the cell populations that play a major role among repertoire of the 3 species may be similar. In other 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 7A). 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, and bat and mouse had almost the same insertion at the V 3' end and the J gene 5' end (Figure 7B,. 7C); 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 7D); the similarity of the bat, human, and mouse CDR3 group libraries was assessed by the four indices of overlap coefficient, Jaccard, Morisita, and Tversky (Figure 7E), 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.

Materials and Methods

1.TRB locus annotation

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 (*31*): To find germline genes in the sequence by homology with TR germline genes of 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), and the final results were summarized for all compared to germline gene fragments. 2. 12/23 RSS (Recombination Signal Sequence) scanning: similar method to previous studies(*28*). Briefly: we screened all RSS motif sequences 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 and Supplementary Table 1. Following the IMGT guidelines (*32*), We labelled 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) (Figure 1C, Supplementary Figure 1).

2. Germline gene naming and TRBV families' cluster

Family naming was performed for each germline gene annotated according to the IMGT guidelines: for V genes, nucleotide similarity greater than 75% was considered the same family; a biological evolutionary tree was constructed with other species 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 2), and performed nucleotide multiple sequence alignment of TRBV genes by MEGA7(Version 7.0.26) to construct a phylogenetic tree by Neighbor-joining method (Figure 2A), 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 evolutionary direction of the V gene, included: Artiodactyla: cattle, sheep, and pigs; Carnivores: dogs, cats, and ferrets; and primates: humans, crab-eating monkeys, and rhesus monkeys (Figure 2C) (https://www.imgt.org/IMGTrepertoire/). For D, J, and C genes, they 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(ClustalW program), with priority alignment of three conserved sites: 23CYS, 41TRP, and 104CYS, based on the known structure and specific positions of mammalian TRBV(*32*). 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) (Figure 3, Supplementary Figure 2). 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 Table4. 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 (Supplementary Table4).

4.Sample collection

Bat samples were collected in XIshui County, Zunyi City, Guizhou Province with the help from Zhou's group. Muscle tissues were used for genotype identification and then Cytb gene were amplificated with designed primers (Supplementary Table6), the PCR products 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 (Table 5), finally, 3 *Rhinolophus affinis* 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

15 known bats' TRBC genes were aligned by their nucleotide sequences then primers for TCR-β library construction were designed in the conserved region, repertoires of *Rhinolophus affinis* 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) (Supplementary Figure 3). The 5'race TCR-β repertoires construction and sequencing work of mice were

done by SEQHEALTH Company (Wuhan, China). The TCR-β repertoires of human were prepared ourselves with multiplex PCR library construction (Supplementary Table6), 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 human and mouse were also routinely pipeline analyzed by MiXCR (Version 3.0.13), the MiXCR output data of the 9 samples are shown in Table 2 and Supplementary Table 7. To check the quality of reads, we exemplify the sequences of bat1 sample in Figure 5, containing 25 TRBV genes as well as 14 TRBJ genes. The evolutionary relationship between the sequenced samples and the 3 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 3 species. We performed the following analyses: 1. TRBV gene and TRBJ gene usage, V-J gene pairing, and conserved motif tracking (Figure 6, Figure 7, and Supplementary Figure 5); 2. CDR3 length and AA usage, nucleotide deletion and insertion; 3. Repertoire overlap index statistics among the 3 species (Figure 8).

7. 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.

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 (*33*) ; Julia et al. found that CD4+ T cells in fox bats dominate the blood lymph, while CD8+ T cells dominate the spleen (*34*). These indicate the T/B cell immune in bats play 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 (*27, 28*) , 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. To establish bat germline gene database, the first work is to complete the annotation of TR/IG loci, unlike the usual genome annotation, which requires high genome quality due to the very large span of TR/IG loci on the genome, and the annotation of TR loci at the scaffold and contig level brings the result that most germline genetic information lost. In this study, we finally identified three bat species with high genomic quality from the 45 bat species in the existing database: *Rhinolophus ferrumequinum* (Rhinolophidae) (*28*), *Phyllostomus discolor* (Phyllostomidae), and *Pipistrellus pipistrellus* (Vespertilionidae). Second, the methodological approach to gene annotation, the central issue in the creation of germline gene databases is how to find all potential genes as possible based on specific structure features. Due to the short sequence D (10-15 bp) and J (30-50 bp) genes, in order to avoid losing sequences from homology search, we used RSS sequence search method at the same time, although there may be non-classical 23 RSS (23+/-1 nt) to influence the final result, but this kind of genes has very limited examples in human and mouse, moreover, pseudogenes that lack of RSS cannot involve in the rearrangement process. This means that these pseudogenes have nothing impact on productive repertoire formation and also pathogen response(*35*).

Although several studies have reported the presence of varying degrees of duplications in Chiroptera on immune-related genes. For the TCR, a large number of familial duplications events were only seen in Artiodactyla such as cattle (134 TRBV, 371 TRAV/DV)(*36, 37*)and goat（94 TRBV）(*30*), 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 pale spear-nosed bats and the large variation in TRBV gene numbers (30 and 100) among Chiroptera have never been reported in mammals. Because previous studies of primates: humans, rhesus monkeys, and crab-eating monkeys(*29, 38, 39*) , carnivores: cats, dogs, ferrets (*40-42*) and three species of camels(*43*), species of the same order differed very little in TRB loci both in terms of loci structure and number of germline genes. Phylogenetic analysis classified Chiroptera as a sister order to Carnivora, and Artiodactyla (*44*), and we found that the distribution of TRBV gene families in the 3 bat species was more similar to that of Artiodactyla, with random and huge duplications. Based on the results of the 3 bat species annotated 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. Notably, it has been reported that *Rhinolophus Hildebrandti* and *Pipistrellus pipistrellus* do not contain gut-associated lymphoid tissue (GALT) (*45, 46*) , and the chicken immunoglobulin

system is a typical GALT species that relies primarily on Gene Conversion to generate receptor diversity, but bats have not been seen to possess similar Gene Conversion-related in birds and rabbits (*47, 48*), 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 3 bats annotated.

The anchor residues in TRBV gene are associated with MHC peptide binding, although the MHC genes in bats more diversity than other mammalians, some researches have been reported and showed bats' MHC genes have highly similarity in sequences with human, mice, dog, cow et.al(*13, 49, 50*), the anchor residues in annotated TRBV and TRBJ genes are highly conserved, this finding supporting 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 *Rhinolophus affinis* bats, and comparative analysis with the human and mouse TCR β chain CDR3 repertoire. The 3 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, the impact of primers, we found that 25 TRBV genes as well as 14 TRBJ gene families were present in each experimental bat after 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 *Rhinolophus affinis* were firstly found to be very similar to the annotated genes of *Rhinolophus ferrumequinum*, with 99% of the C genes belonging to *Rhinolophus ferrumequinum*, and TRBV and TRBJ were also basically identical to the annotated V and J genes of *Rhinolophus ferrumequinum*. There is a high genetic conservation in the composition of specific immune response genes among different species of bats of the same family (*Rhinolophus affinis* and *Rhinolophus ferrumequinum*). 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, and 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 (*51*). 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 (*52-54*) , 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(*55-59*), 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 are essentially the same and have a high consistency in high-frequency motifs, whereas they show greater 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 common 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 *Rhinolophus ferrumequinum* (Rhinolophidae), *Phyllostomus discolor* (Phyllostomidae), and *Pipistrellus pipistrellus* (Vespertilionidae), as well as the first construction the repertoire, sequencing, and MiXCR analysis tools 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 basic data for comparative analysis in the study of the mechanism of virus tolerance in bats.

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 The 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.

Fig1 Results of bat locus annotation and TRB loci correlation analysis of three bat species. (A) Overview of this study design. (B) Overview plot of TRB loci of *Phyllostomus discolor*, the numbers represent the position in the 10th chromosome, and the colored boxes represent each gene annotated, namely 100 TRBV genes, 2 TRBD genes, 13 TRBJ genes, and 2 TRBC genes. (C) Homology comparison map of TRB loci for the 3 annotated bats.

Fig2 Analysis of TRBV gene homology in bats with different species. (A) NJ-phylogenetic tree constructed from functional TRBV gene nucleotide sequences of 3 species of bats, human, mouse, and pig, with each color representing 1 family and bootstrap is 1000. (B) Comparison of TRBV gene family distribution in 3 mammalian species: Carnivora, Artiodactyla, and Carnivora with 3 species of Chiroptera (https://www.imgt.org/ IMGTrepertoire/). (C) Nucleotide sequence similarity comparison of TRBV and TRBJ genes in 3 annotated bat species.

Fig3 *Pipistrellus pipistrellus* annotated germline gene 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 present in all TRBJ genes are marked in red.

Fig4 TCRβ chain repertoire library construction and sequencing process for the 3 species used in this study. (A) TCRβ chain repertoire library construction and sequencing of spleen tissues from 3 *Rhinolophus affinis* and 3 mouse was using the 5'race method and the PE150 sequencing platform. 3 individuals PBMC TCR β chain repertoire were constructed by our group by multiplex PCR method, PE150 sequencing platform. (B) The TRBC Exon1 nucleotide sequence alignment results of a total of 15 species of bats, including the annotated 3 species, were used to design amplification primers for the 5'race method of the Middle Chrysanthemum-headed bat.

Fig5 Statistical analysis of the sequencing results of *Rhinolophus affinis*. (A) The 25 TRBV genes and 14 TRBJ genes output in bat1 samples. (B) Attribution statistics of the sequencing results of the three *Rhinolophus affinis* bats in the annotated 3 bat species. (C) Nucleotide sequence comparison of Cytb sequencing results of the 3 bat samples with *Phyllostomus discolor*, *Rhinolophus ferrumequinum*, and *Pipistrellus pipistrellus*. (D) The partial sequences of C gene from bat1 samples were compared with those of the 3 annotated bats.

Fig6 Comparative analysis of TCR β chain group library genes and motif usage of bats, human and mouse. (A) Homologous TRBV gene usage analysis of the 3 species. (B) TRBV gene usage analysis of bats and humans. (C) TRBJ gene usage analysis. (D) J gene usage in cluster. (E) Top 30 5 AA length motif usage analysis. (F) Example of top 10 motif usage with shared motifs labeled in the same color.

Fig7 Comparative analysis of the CDR3 region features of the TCRβ chain repertoire in bat, human and mouse. (A) CDR3 region length analysis. (B) Comparison of deletion at the 3' end of V gene and 5' end of J gene. (C) Composition of the CDR3 region. (D) AA usage of the CDR3 region. (E) CDR3 sequence overlap index of the 3 species.

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

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

Supplementary Fig 1. Sequence maps of *Phyllostomus discolor* and *Pipistrellus pipistrellus* 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 *Phyllostomus discolor* germline genes. (A) Amino acid sequence comparison of 100 TRBV genes of *Phyllostomus discolor*. (B) Amino acid sequence comparison of 14 TRBJ genes of *Phyllostomus discolor*

Supplementary Fig 3. TCRβ chain repertoire building peaks of the three *Rhinolophus affinis*.

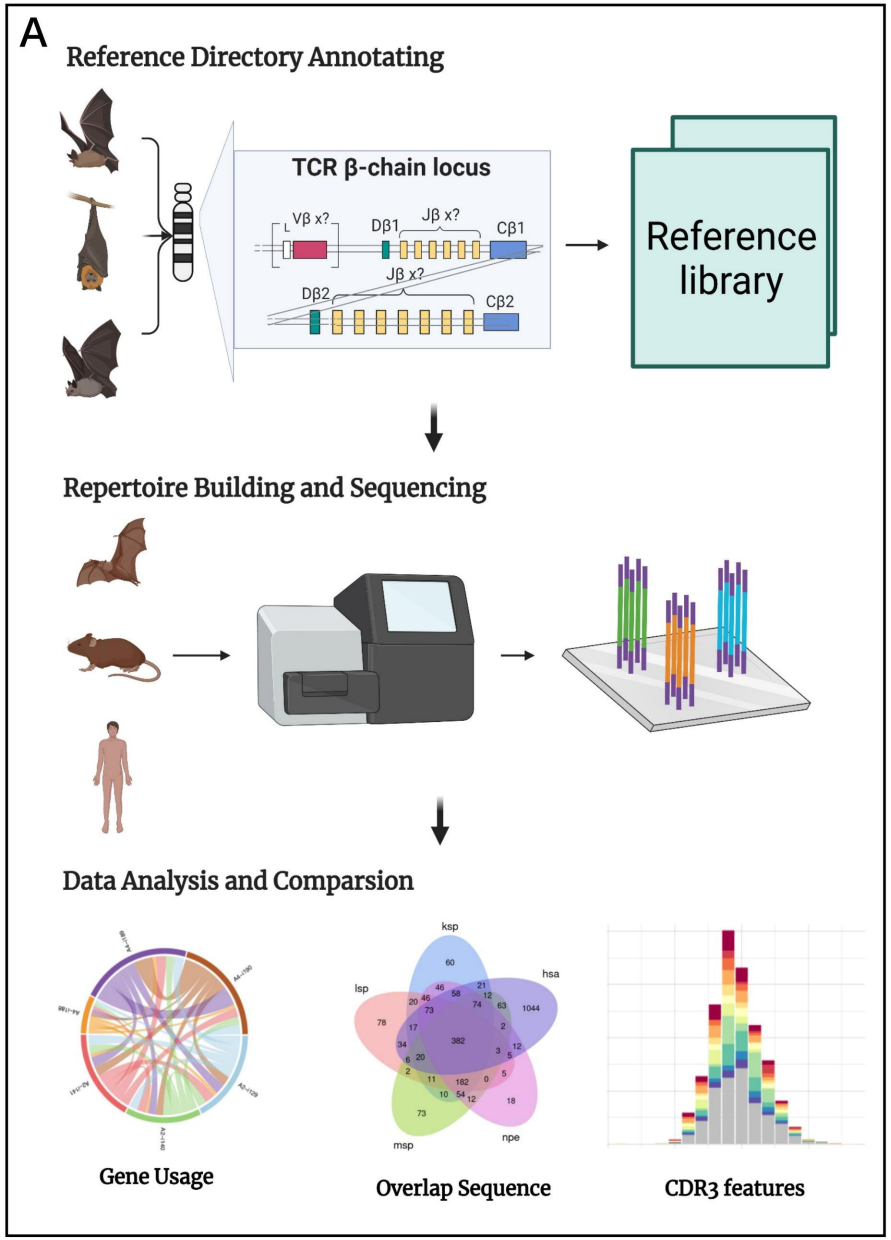
Supplementary Fig 4. Analysis of V and J pairing (the order is Bat2, Human2 and Mouse2, Bat3, Human3 and Mouse3, respectively).

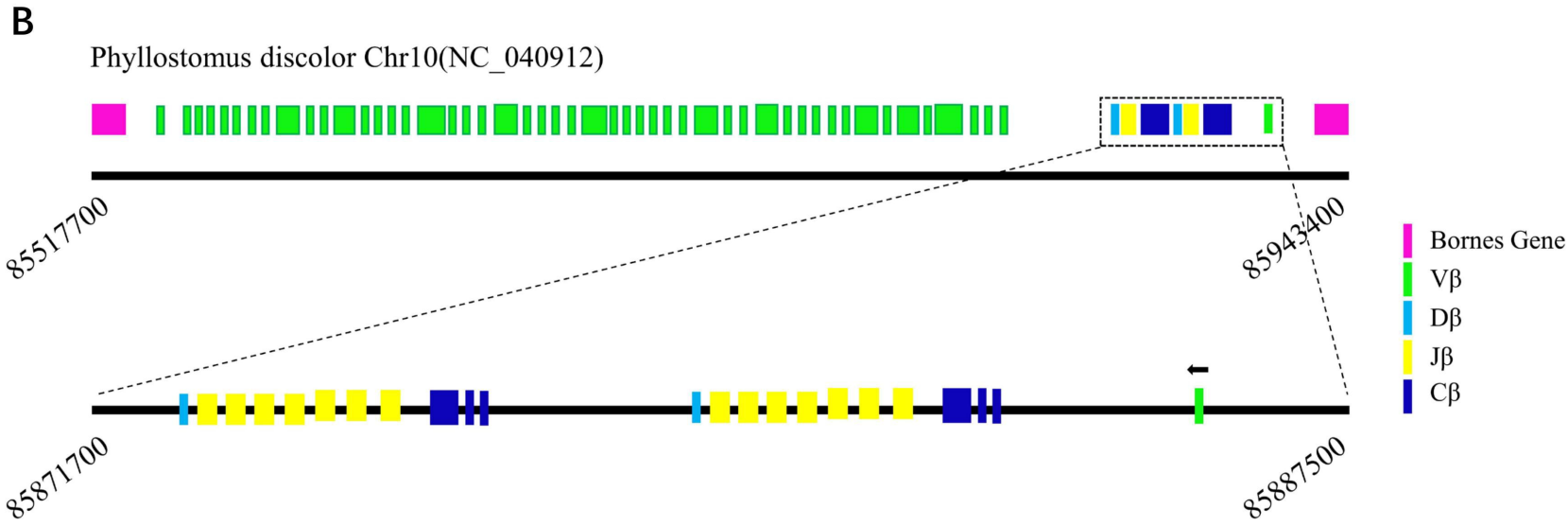
Reference

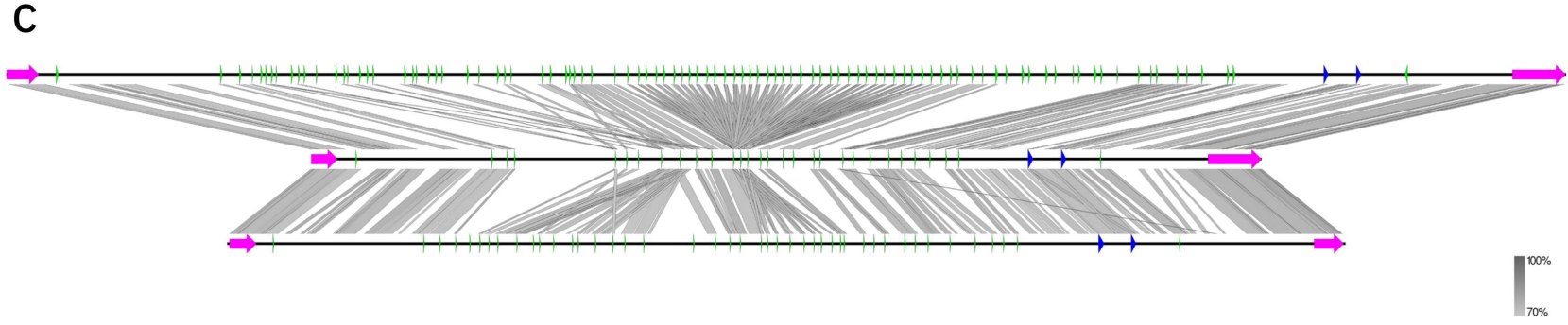
1. C. J. Burgin, J. P. Colella, P. L. Kahn, N. S. Upham, How many species of mammals are there? *Journal of Mammalogy* **99**, 1-14 (2018).
2. M. Letko, S. N. Seifert, K. J. Olival, R. K. Plowright, V. J. Munster, Bat-borne virus diversity, spillover and emergence. *Nat Rev Microbiol* **18**, 461-471 (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).
8. A. T. Irving, M. Ahn, G. Goh, D. E. Anderson, L. F. Wang, Lessons from the host defences of bats, a unique viral reservoir. *Nature* **589**, 363-370 (2021).
9. M. Ahn, J. Cui, A. T. Irving, L. F. Wang, Unique Loss of the PYHIN Gene Family in Bats Amongst Mammals: Implications for Inflammasome Sensing. *Sci Rep* **6**, 21722 (2016).
10. J. Xie *et al.*, Dampened STING-Dependent Interferon Activation in Bats. *Cell Host Microbe* **23**, 297-301 e294 (2018).
11. S. S. Pavlovich *et al.*, The Egyptian Rousette Genome Reveals Unexpected Features of Bat Antiviral Immunity. *Cell* **173**, 1098-1110 e1018 (2018).
12. H. E. Randolph, L. B. Barreiro, Holy Immune Tolerance, Batman! *Immunity* **48**, 1074-1076 (2018).
13. J. H. Ng *et al.*, Evolution and comparative analysis of the bat MHC-I region. *Sci Rep* **6**, 21256 (2016).
14. S. J. Anthony *et al.*, Further Evidence for Bats as the Evolutionary Source of Middle East Respiratory Syndrome Coronavirus. *mBio* **8**, (2017).
15. V. D. Menachery *et al.*, A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* **21**, 1508-1513 (2015).
16. X. Y. Ge *et al.*, Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535-538 (2013).
17. C. G. Sotero-Caio, R. J. Baker, M. Volleth, Chromosomal Evolution in Chiroptera. *Genes (Basel)* **8**, (2017).
18. A. Mori *et al.*, Next generation sequencing: new tools in immunology and hematology. *Blood Res* **48**, 242-249 (2013).
19. A. Six *et al.*, The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. *Front Immunol* **4**, 413 (2013).
20. S. Teraguchi *et al.*, Methods for sequence and structural analysis of B and T cell receptor repertoires. *Comput Struct Biotechnol J* **18**, 2000-2011 (2020).
21. V. G. Eltaf Alamyar , Shuo Li , Patrice Duroux , Marie-Paule Lefranc, IMGT/HIGHV-QUEST: THE IMGT® WEB PORTAL FOR IMMUNOGLOBULIN (IG) OR ANTIBODY AND T CELL RECEPTOR (TR) ANALYSIS FROM NGS HIGH THROUGHPUT AND DEEP SEQUENCING. *Immunome Research* **8**, 1-15 (2012).
22. D. A. Bolotin *et al.*, MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* **12**, 380-381 (2015).

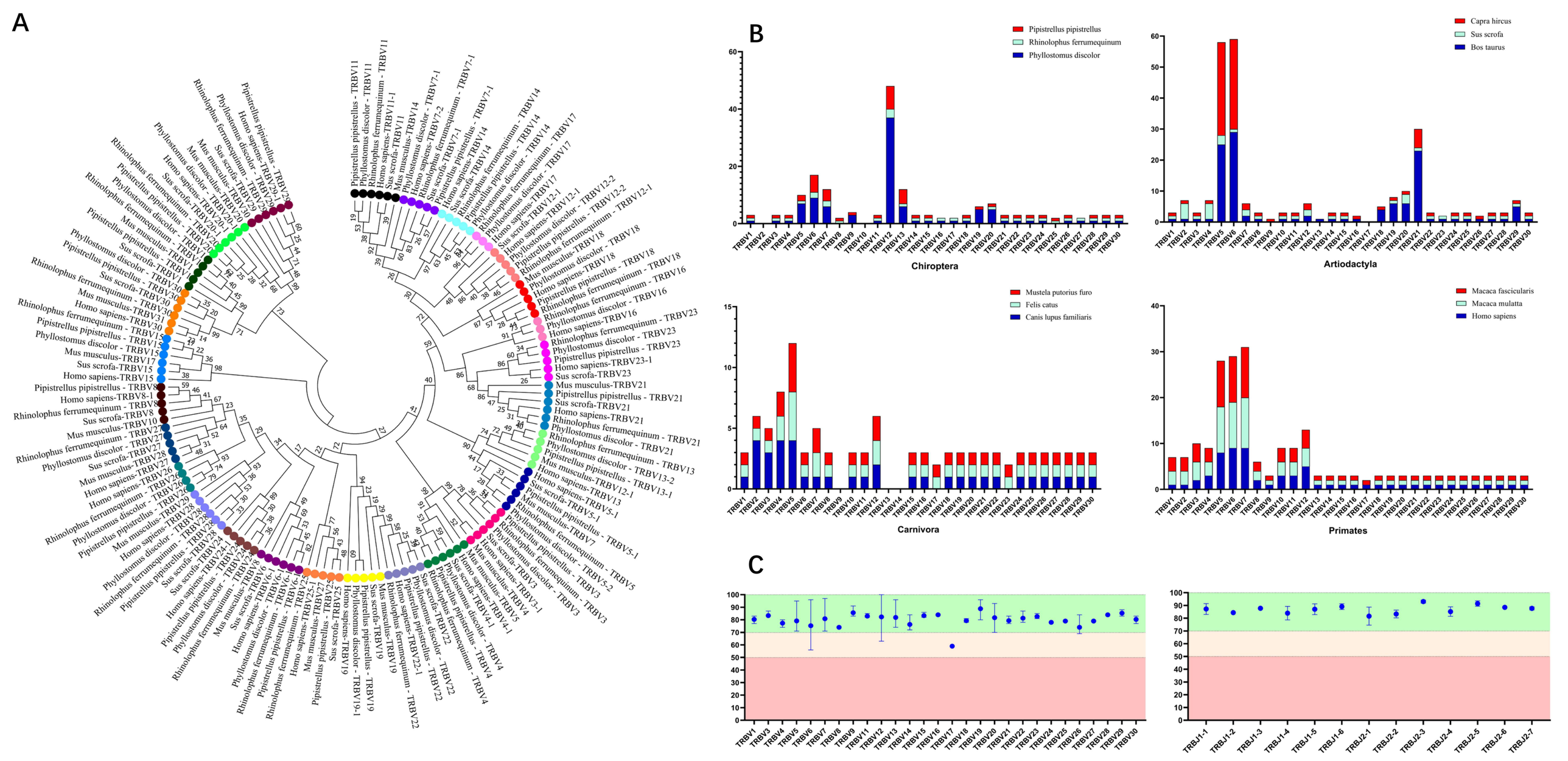
23. L. Song *et al.*, TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat Methods* **18**, 627-630 (2021).
24. S. Li *et al.*, IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nat Commun* **4**, 2333 (2013).
25. D. Jebb *et al.*, Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578-584 (2020).
26. A. Rhie *et al.*, Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737-746 (2021).
27. P. A. Larson *et al.*, Genomic features of humoral immunity support tolerance model in Egyptian rousette bats. *Cell Rep* **35**, 109140 (2021).
28. H. Zhou, L. Ma, L. Liu, X. Yao, TR Locus Annotation and Characteristics of Rhinolophus ferrumequinum. *Front Immunol* **12**, 741408 (2021).
29. P. Pegorier *et al.*, IMGT(R) Biocuration and Comparative Study of the T Cell Receptor Beta Locus of Veterinary Species Based on Homo sapiens TRB. *Front Immunol* **11**, 821 (2020).
30. F. Giannico *et al.*, The expansion of the TRB and TRG genes in domestic goats (Capra hircus) is characteristic of the ruminant species. *BMC Genomics* **21**, 623 (2020).
31. P. D. a. M.-P. L. Jérôme Lane, From IMGT-ONTOLOGY to IMGT/LIGMotif: the IMGT® standardized approach for immunoglobulin and T cell receptor gene identification and description in large genomic sequences. *BMC Bioinformatics* **11**, 1-16 (2010).
32. V. Giudicelli, M. P. Lefranc, Imgt-Ontology 2012. *Front Genet* **3**, 79 (2012).
33. C. A. Sarkar SK, Analysis of immunocompetent cells in the bat, Pteropus giganteus: isolation and scanning electron microscopic characterization. *Dev Comp Immunol* **15**, 423-430 (1991).
34. J. M. Martinez Gomez *et al.*, Phenotypic and functional characterization of the major lymphocyte populations in the fruit-eating bat Pteropus alecto. *Sci Rep* **6**, 37796 (2016).
35. K. B. a. G. E. W. Dale A.Ramsden, Conservation of sequence in recombination signal sequence spacers. *Nucleic Acids Research* **22**, 1785-1796 (1994).
36. T. Connelley, J. Aerts, A. Law, W. I. Morrison, Genomic analysis reveals extensive gene duplication within the bovine TRB locus. *BMC Genomics* **10**, 192 (2009).
37. K. D. Timothy K Connelley, Cassandra W Longhi and W Ivan Morrison, Genomic analysis offers insights into the evolution of the bovine TRA/TRD locus. *BMC Genomics* **15**, 1-17 (2014).
38. S. Jaiswal *et al.*, Identification and characterization of the T cell receptor (TCR) repertoire of the Cynomolgus macaque (Macaca Fascicularis). *bioRxiv*, 1-29 (2022).
39. H. Y. Greenaway *et al.*, Extraction and characterization of the rhesus macaque T-cell receptor beta-chain genes. *Immunol Cell Biol* **87**, 546-553 (2009).
40. M. Mineccia *et al.*, New insight into the genomic structure of dog T cell receptor beta (TRB) locus inferred from expression analysis. *Dev Comp Immunol* **37**, 279-293 (2012).
41. A. Radtanakatikanon *et al.*, Topology and expressed repertoire of the Felis catus T cell receptor loci. *BMC Genomics* **21**, 20 (2020).
42. B. Gerritsen *et al.*, Characterization of the ferret TRB locus guided by V, D, J, and C gene expression analysis. *Immunogenetics* **72**, 101-108 (2020).
43. R. Antonacci, M. Bellini, G. Linguiti, S. Ciccarese, S. Massari, Comparative Analysis of the TRB Locus in the Camelus Genus. *Front Genet* **10**, 482 (2019).
44. G. Tsagkogeorga, J. Parker, E. Stupka, J. A. Cotton, S. J. Rossiter, Phylogenomic analyses elucidate the evolutionary relationships of bats. *Curr Biol* **23**, 2262-2267 (2013).
45. S. Strobel, J. A. Encarnacao, N. I. Becker, T. E. Trenczek, Histological and histochemical analysis of the gastrointestinal tract of the common pipistrelle bat (Pipistrellus pipistrellus). *Eur J Histochem* **59**, 2477 (2015).
46. A. N. M. a. J. N. MAINA, The morphology of the intestine of the insectivorous horseshoe bat (Rhinolophus hildebvandti, Peters): a scanning electron and light microscopic study. *Afr. J. Ecol* **32**, 158-168 (1994).
47. V. A. Claude-Agnès Reynaud, Hélène Grimal, Jean-Claude Weill,, A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell* **48**, 379-388 (1987).
48. M. RG., Diversification of rabbit VH genes by gene-conversion-like and hypermutation mechanisms. *Immunol Rev* **162**, 49-54 (1998).
49. Z. Qu *et al.*, Structure and Peptidome of the Bat MHC Class I Molecule Reveal a Novel Mechanism Leading to High-Affinity Peptide Binding. *J Immunol* **202**, 3493-3506 (2019).
50. T. Qurkhuli, N. Schwensow, S. D. Brandel, M. Tschapka, S. Sommer, Can extreme MHC class I diversity be a feature of a wide geographic range? The example of Seba's short-tailed bat (Carollia perspicillata).

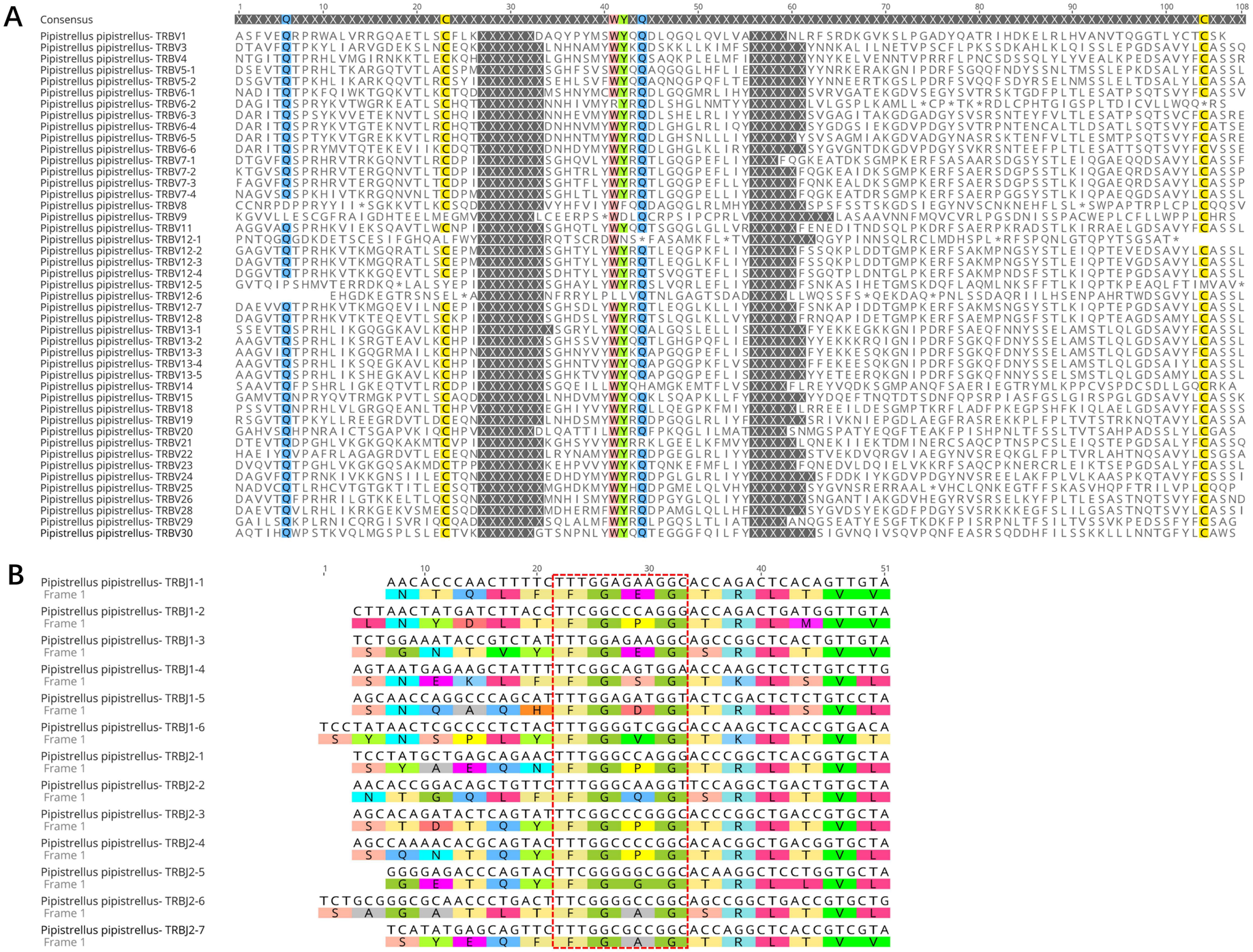
Immunogenetics **71**, 575-587 (2019).
51. S. J. Turner, P. C. Doherty, J. McCluskey, J. Rossjohn, Structural determinants of T-cell receptor bias in immunity. *Nat Rev Immunol* **6**, 883-894 (2006).
52. T. O. Hansen, A. B. Lange, T. Barington, Sterile DJH rearrangements reveal that distance between gene segments on the human Ig H chain locus influences their ability to rearrange. *J Immunol* **194**, 973-982 (2015).
53. L. Fu *et al.*, A comprehensive profiling of T- and B-lymphocyte receptor repertoires from a Chinese-origin rhesus macaque by high-throughput sequencing. *PLoS One* **12**, e0182733 (2017).
54. J. E. Park *et al.*, A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, (2020).
55. R. Alli, Z. M. Zhang, P. Nguyen, J. J. Zheng, T. L. Geiger, Rational design of T cell receptors with enhanced sensitivity for antigen. *PLoS One* **6**, e18027 (2011).
56. C. J. Cibotti R, Pannetier C, Delarbre C, Vergnon I, Kanellopoulos JM, Kourilsky P, Public and 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).
57. 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).
58. 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).
59. H. F. Koay *et al.*, Diverse MR1-restricted T cells in mice and humans. *Nat Commun* **10**, 2243 (2019).

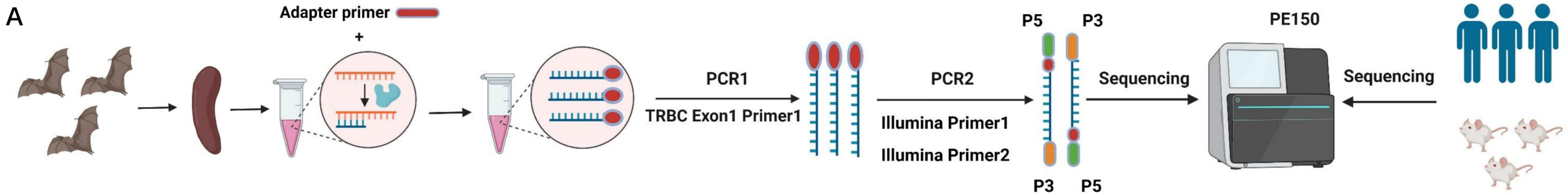

A
Reference Directory Annotating
TCR β-chain locus
Vβ x?
Dβ1
Jβ x?
Cβ1
Jβ x?
Dβ2
Cβ2
Reference library
Repertoire Building and Sequencing
Data Analysis and Comparsion
Gene Usage
Overlap Sequence
CDR3 features

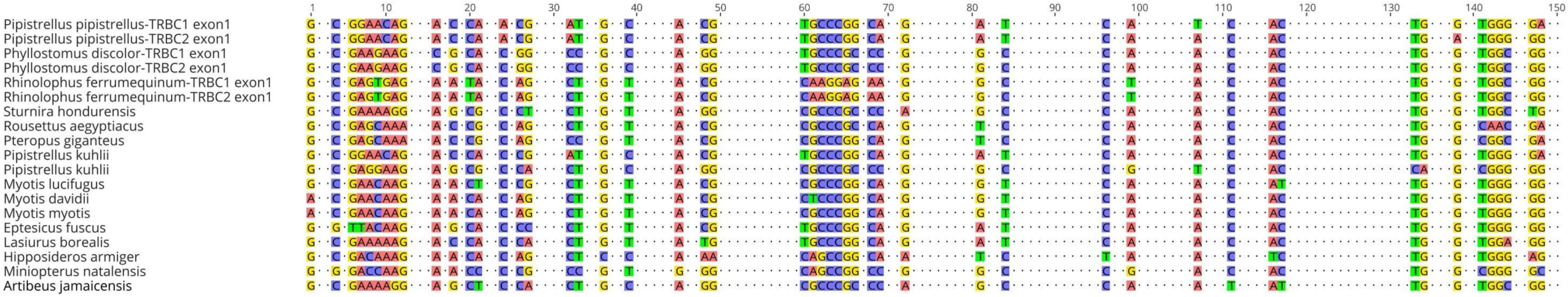

B
Phyllostomus discolor Chr10(NC_040912)
85517700
85943400
85871700
85887500
Bornes Gene
Vβ
Dβ
Jβ
Cβ


C
100%
70%


A
B
Pipistrellus pipistrellus
Rhinolophus ferrumequinum
Phyllostomus discolor
Chiroptera
Capra hircus
Sus scrofa
Bos taurus
Artiodactyla
Mustela putorius furo
Felis catus
Canis lupus familiaris
Carnivora
Macaca fascicularis
Macaca mulatta
Homo sapiens
Primates
C
TRBV1
TRBV2
TRBV3
TRBV4
TRBV5
TRBV6
TRBV7
TRBV8
TRBV9
TRBV10
TRBV11
TRBV12
TRBV13
TRBV14
TRBV15
TRBV16
TRBV17
TRBV18
TRBV19
TRBV20
TRBV21
TRBV22
TRBV23
TRBV24
TRBV25
TRBV26
TRBV27
TRBV28
TRBV29
TRBV30
TRBJ1-1
TRBJ1-2
TRBJ1-3
TRBJ1-4
TRBJ1-5
TRBJ1-6
TRBJ2-1
TRBJ2-2
TRBJ2-3
TRBJ2-4
TRBJ2-5
TRBJ2-6
TRBJ2-7


A
Consensus
Pipistrellus pipistrellus- TRBV1
Pipistrellus pipistrellus- TRBV3
Pipistrellus pipistrellus- TRBV4
Pipistrellus pipistrellus- TRBV5-1
Pipistrellus pipistrellus- TRBV5-2
Pipistrellus pipistrellus- TRBV6-1
Pipistrellus pipistrellus- TRBV6-2
Pipistrellus pipistrellus- TRBV6-3
Pipistrellus pipistrellus- TRBV6-4
Pipistrellus pipistrellus- TRBV6-5
Pipistrellus pipistrellus- TRBV6-6
Pipistrellus pipistrellus- TRBV7-1
Pipistrellus pipistrellus- TRBV7-2
Pipistrellus pipistrellus- TRBV7-3
Pipistrellus pipistrellus- TRBV7-4
Pipistrellus pipistrellus- TRBV8
Pipistrellus pipistrellus- TRBV9
Pipistrellus pipistrellus- TRBV11
Pipistrellus pipistrellus- TRBV12-1
Pipistrellus pipistrellus- TRBV12-2
Pipistrellus pipistrellus- TRBV12-3
Pipistrellus pipistrellus- TRBV12-4
Pipistrellus pipistrellus- TRBV12-5
Pipistrellus pipistrellus- TRBV12-6
Pipistrellus pipistrellus- TRBV12-7
Pipistrellus pipistrellus- TRBV12-8
Pipistrellus pipistrellus- TRBV13-1
Pipistrellus pipistrellus- TRBV13-2
Pipistrellus pipistrellus- TRBV13-3
Pipistrellus pipistrellus- TRBV13-4
Pipistrellus pipistrellus- TRBV13-5
Pipistrellus pipistrellus- TRBV14
Pipistrellus pipistrellus- TRBV15
Pipistrellus pipistrellus- TRBV18
Pipistrellus pipistrellus- TRBV19
Pipistrellus pipistrellus- TRBV20
Pipistrellus pipistrellus- TRBV21
Pipistrellus pipistrellus- TRBV22
Pipistrellus pipistrellus- TRBV23
Pipistrellus pipistrellus- TRBV24
Pipistrellus pipistrellus- TRBV25
Pipistrellus pipistrellus- TRBV26
Pipistrellus pipistrellus- TRBV28
Pipistrellus pipistrellus- TRBV29
Pipistrellus pipistrellus- TRBV30
B
Pipistrellus pipistrellus- TRBJ1-1
Frame 1
Pipistrellus pipistrellus- TRBJ1-2
Frame 1
Pipistrellus pipistrellus- TRBJ1-3
Frame 1
Pipistrellus pipistrellus- TRBJ1-4
Frame 1
Pipistrellus pipistrellus- TRBJ1-5
Frame 1
Pipistrellus pipistrellus- TRBJ1-6
Frame 1
Pipistrellus pipistrellus- TRBJ2-1
Frame 1
Pipistrellus pipistrellus- TRBJ2-2
Frame 1
Pipistrellus pipistrellus- TRBJ2-3
Frame 1
Pipistrellus pipistrellus- TRBJ2-4
Frame 1
Pipistrellus pipistrellus- TRBJ2-5
Frame 1
Pipistrellus pipistrellus- TRBJ2-6
Frame 1
Pipistrellus pipistrellus- TRBJ2-7
Frame 1


A
Adapter primer
+
PCR1
TRBC Exon1 Primer1
PCR2
Illumina Primer1
Illumina Primer2
P5
P3
P3
P5
Sequencing
PE150
Sequencing


B
Pipistrellus pipistrellus-TRBC1 exon1
Pipistrellus pipistrellus-TRBC2 exon1
Phyllostomus discolor-TRBC1 exon1
Phyllostomus discolor-TRBC2 exon1
Rhinolophus ferrumequinum-TRBC1 exon1
Rhinolophus ferrumequinum-TRBC2 exon1
Sturnira hondurensis
Rousettus aegyptiacus
Pteropus giganteus
Pipistrellus kuhlii
Pipistrellus kuhlii
Myotis lucifugus
Myotis davidii
Myotis myotis
Eptesicus fuscus
Lasiurus borealis
Hipposideros armiger
Miniopterus natalensis
Artibeus jamaicensis

A

CDR3 region | C gene

V1-2 J2-3
E P E D S A L Y L C A S S Q E G G S T G Q L Y F G T G S K L T V L E D L S E V K L P K V T V F E P
GAGCCAGAAGACTCTGCCTTGTATCTCTGTGCCAGCAGCCAAGAAGGGGGGTCCACCGGACAACTGTACTTTGGGACCGGTTCCAAGCTGACTGTGCTGGAGGACCTGAGTGAGGTGAAACTACCCAAGGTCACTGTGTTCGAACCA

V3-2 J1-2
P E D S A L Y L C A S S Q E R G A A A A A L T F G A G S R L T V L E D L S E V K L P K V T C S N H
CCAGAAGACTCTGCCTTGTATCTCTGTGCCAGCAGCCAAGAAAGGGGGGCTGCAGCCGCCGCCCTGACTTTCGGGGCCGGCAGCCGGCTGACCGTGCTCGAGGACCTGAGTGAGGTGAAACTACCCAAGGTCACGTGTTCGAACCAT

V4-2 J2-2
E P E D S A L Y L C A S S Q E G G S T G Q L Y F G T G S K L T V L E D L S E V K L P K V T V F E P
GAGCCAGAAGACTCTGCCTTGTATCTCTGTGCCAGCAGCCAAGAAGGGGGGTCCACCGGACAACTGTACTTTGGGACCGGTTCCAAGCTGACTGTGCTGGAGGACCTGAGTGAGGTGAAACTACCCAAGGTCACTGTGTTCGAACCA

V5-8 J2-6
E L R D A A L Y L C A S S P K S S A A A A L T F G A G S R L T V H E D L S E V K L P K V T V F E P
GAGCTGAGAGACGCAGCCCTGTATCTGTGTGCCAGCAGCCCAAAGAGCTCTGCAGCCGCCGCCCTGACTTTCGGGGCCGGCAGCCGGCTGACCGTGCACGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V6-11 J1-6
K S A S P S Q T S V Y F C A S R D G R S P L H F G T G T R L T V T E D L S E V K L P K V T V F E P
AAGTCGGCTTCCCCCTCCCAGACCTCTGTGTACTTCTGTGCCAGCAGGGACGGGCGTTCGCCCCTCCACTTTGGGACAGGCACCAGACTCACCGTGACAGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTCGAACCA

V7-7 J2-7
E Q R D S A V Y L C A S R I S G D R Y Y E Q F F G P G T K L T V L E D L S E V K L P K V T V F E P
GAGCAGAGGGACTCGGCCGTGTATCTCTGTGCCAGCAGGATTAGTGGGGACAGATACTATGAGCAGTTCTTCGGCCCCGGCACCAAGCTCACGGTCTTAGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTCGAACCA

V8-1 J2-4
S A S T N Q T S L Y L C A S S N N F Q D T Q Y F G A G T R L T V L E D L S E V K L P K V T V F E P
TCCGCCAGCACCAACCAGACCTCTCTGTACCTCTGTGCCAGCAGTAACAACTTCCAAGACACTCAGTATTTCGGCGCGGGCACCCGGCTGACCGTGCTGGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V11-2 J2-2
M G D S A V Y F C A S S P L S G A D T G Q L Y F G T G S K L T V L E D L S E V K L P K V T V F E P
CTGGGGGACTCGGCCGTGTATTTCTGTGCCAGCAGCCCCCTCTCCGGGGCGGACACCGGACAACTGTACTTTGGGACCGGTTCCAAGCTGACTGTGCTGGAGGACCTGAGTGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V12-26 J2-1
A G D S A L Y L C A S S L P W D D N Y A E Q F F G P G T R L T V L E D L R N V T P P K V T V F E P
GCAGGAGACTCAGCACTGTACCTCTGTGCCAGCAGTCTGCCCTGGGACGATAACTATGCTGAGCAGTTCTTCGGACCAGGGACACGACTCACCGTCCTAGAGGATCTGAGAAATGTGACTCCACCCAAGGTCACTGTGTTTGAACCA

V13-7 J1-2
S L Q L E D S A V Y L C A S S L D S E G D Y T F G P G T R L T V V E D L S E V K L P K V T V F E P
AGCTTGCAACTGGAAGACTCGGCCGTGTATCTCTGTGCCAGCAGCTTAGACAGTGAAGGGGATTATACCTTCGGCCCAGGGACCAGGCTGACAGTTGTAGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V18-2 J2-1
E P G D S A V Y F C A S S L S T G G S N E Q H F G P G T R L T V L E D L S E V K L P K V T V F E P
GAGCCGGGAGACTCCGCAGTGTATTTCTGTGCCAGTTCCTTATCCACAGGGGGTAGTAATGAGCAACACTTCGGGCCAGGAACGCGGCTCACGGTATTAGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V19-6 J2-4
A Q K N E M A V F L C A S S I L G S Q N T L Y F G A G T R L S V L E D L R N V T P P K V T V F E P
GCCCAGAAGAACGAGATGGCCGTTTTTCTCTGTGCCAGCAGTATATTGGGTAGTCAAAACACCTTGTACTTTGGTGCGGGCACCCGACTATCGGTGCTAGAGGATCTGAGAAATGTGACTCCACCCAAGGTCACTGTGTTTGAACCA

V20-6 J2-3
A N V H P E D S S L Y F C G A G H R T E T Q Y F G A G T R L T V L E D L S E V K L P K V T V F E P
GCGAATGTGCATCCTGAAGACAGCAGCCTCTACTTCTGTGGTGCTGGACACCGCACAGAAACTCAGTATTTCGGCGCAGGCACCCGGCTGACCGTGCTGGAGGACCTGAGTGAGGTGAAACTACCCAAGGTCACTGTGTTCGAACCA

V21-2 J2-5
T D S G D S A L Y F C A S S Q F G Q V D E Q F F G P G T K L T V L E D L S E V K L P K V T V F E P
ACTGATTCAGGCGACTCGGCACTGTATTTCTGTGCCAGCAGCCAATTCGGGCAGGTAGATGAGCAGTTCTTCGGCCCCGGCACCAAGCTCACGGTCTTAGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTCGAACCA

V22-2 J2-7
H T N Q T S L Y L C S G S A G A G S S Y E Q F V G P G T K L T V L E D L S E V K L P K V T V F E P
CACACCAACCAAACGTCCCTGTACCTATGCTCTGGCAGCGCAGGGGCTGGGAGCTCCTATGAGCAGTTCGTCGGCCCCGGCACCAAGCTCACGGTCTTAGAGGACCTGAGTGAGGTGAAACTACCCAAGGTCACTGTGTTCGAACCA

V23-2 J2-2
G D S A L Y F C A S S Q S L W G G P E G Q L Y F G T G S K L T V L E D L S E V K L P K V T V F E P
GGAGACTCAGCCCTGTATTTCTGTGCCAGCAGCCAATCGCTCTGGGGGGGTCCCGAGGGACAACTGTACTTTGGGACCGGTTCCAAGCTGACTGTGCTGGAGGACCTGAGTGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V24-2 J2-1
Q T A S P N Q T A L Y F C A S S P T G N E Q H F G P G T R L T V L E D L S E V K L P K V T V F E P
CAGACTGCCAGCCCCAACCAGACAGCTCTTTACTTCTGTGCCAGCAGTCCAACAGGGAATGAGCAACACTTCGGGCCAGGAACGCGGCTCACGGTATTAGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V25-1 J1-1
N S T S P S Q T S W Y F C A S S E Q G T E S Y F G E G T R L T V V E D L S E V K L P K V T V F E P
AATTCTACCAGCCCCTCACAGACCTCTTGGTACTTCTGTGCCAGCAGTGAACAGGGCACTGAAAGTTACTTTGGAGAAGGCACCAGACTCACGGTTGTAGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V26-2 J1-4
A S T N Q T S V Y F C A S S G Q G V H D K L F F G S G T K L S V L E D L S E V K L P K V T V F E P
GCCAGTACCAACCAGACATCTGTGTACTTCTGTGCCAGCAGTGGACAGGGGGTCACCGACAAACTATTCTTCGGCAGTGGAACCAAGCTCTCTGTTCTGGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTCGAACCA

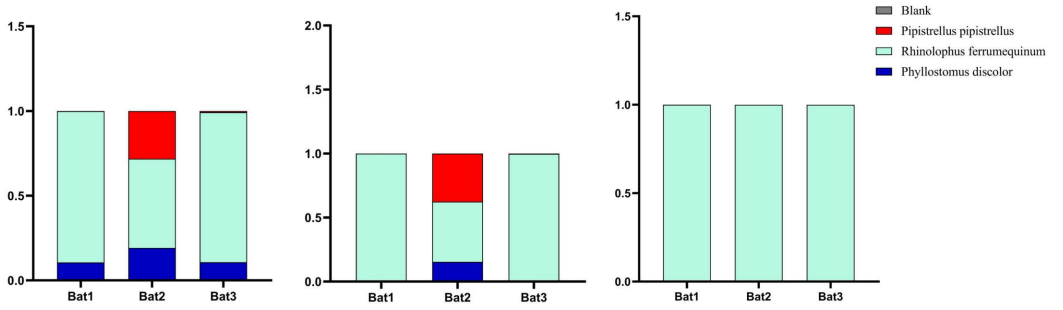
V27-2 J1-4
S T N Q T S V Y F C A S S A L R Q G N D K L F F G S G T K L S V L E D L S E V K L P K V T V F E P
AGCACCAACCAGACCTCTGTGTACTTCTGTGCCAGCAGTGCCTTAAGACAGGGTAACGACAAACTATTCTTCGGCAGTGGAACCAAGCTCTCTGTTCTGGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V28-2 J2-2
A S T N Q T S V Y F C A S R H G G R T G Q L Y F G T G S K L T V L E D L S E V K L P K V T V F E P
GCCAGTACCAACCAGACATCTGTGTACTTCTGTGCCAGCAGGCACGGGGGCGCACCGGACAACTGTACTTTGGGACCGGTTCCAAGCTGACTGTGCTGGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTCGAACCA

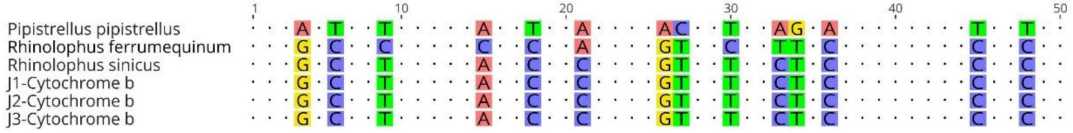
V29-2 J2-4
V R P E D S S L Y F C S V G R G S Y Q D T Q Y F G A G T R L T V L E D L S E V K L P K V T V F E P
GTGCGCCCCGAAGACAGCAGCCTTTACTTCTGCAGCGTTGGTCGGGGGTCGTACCAAGACACTCAGTATTTCGGCGCGGGCACCCGGCTGACCGTGCTGGAGGACCTGAGCGAGGTGAAACTACCCAAGGTCACTGTGTTTGAACCA

V30-2 J1-2
K L L L S H S G F Y L C A W S H T G A S D Y T F G S G T R L L V I E D L R N V T P P K V T V F E P
AAGCTGCTTCTCAGCCACTCTGGCTTCTACCTCTGTGCCTGGAGTCACACAGGAGCCTCCGACTACACCTTCGGCTCAGGGACCAGGCTTTTGGTAATAGAGGATCTGAGAAATGTGACTCCACCCAAGGTCACTGTGTTCGAACCA

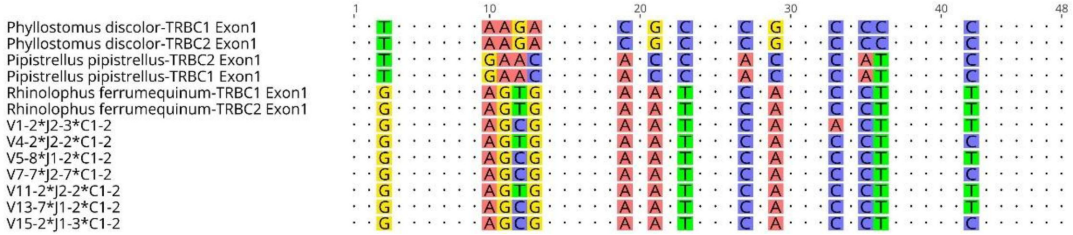
B

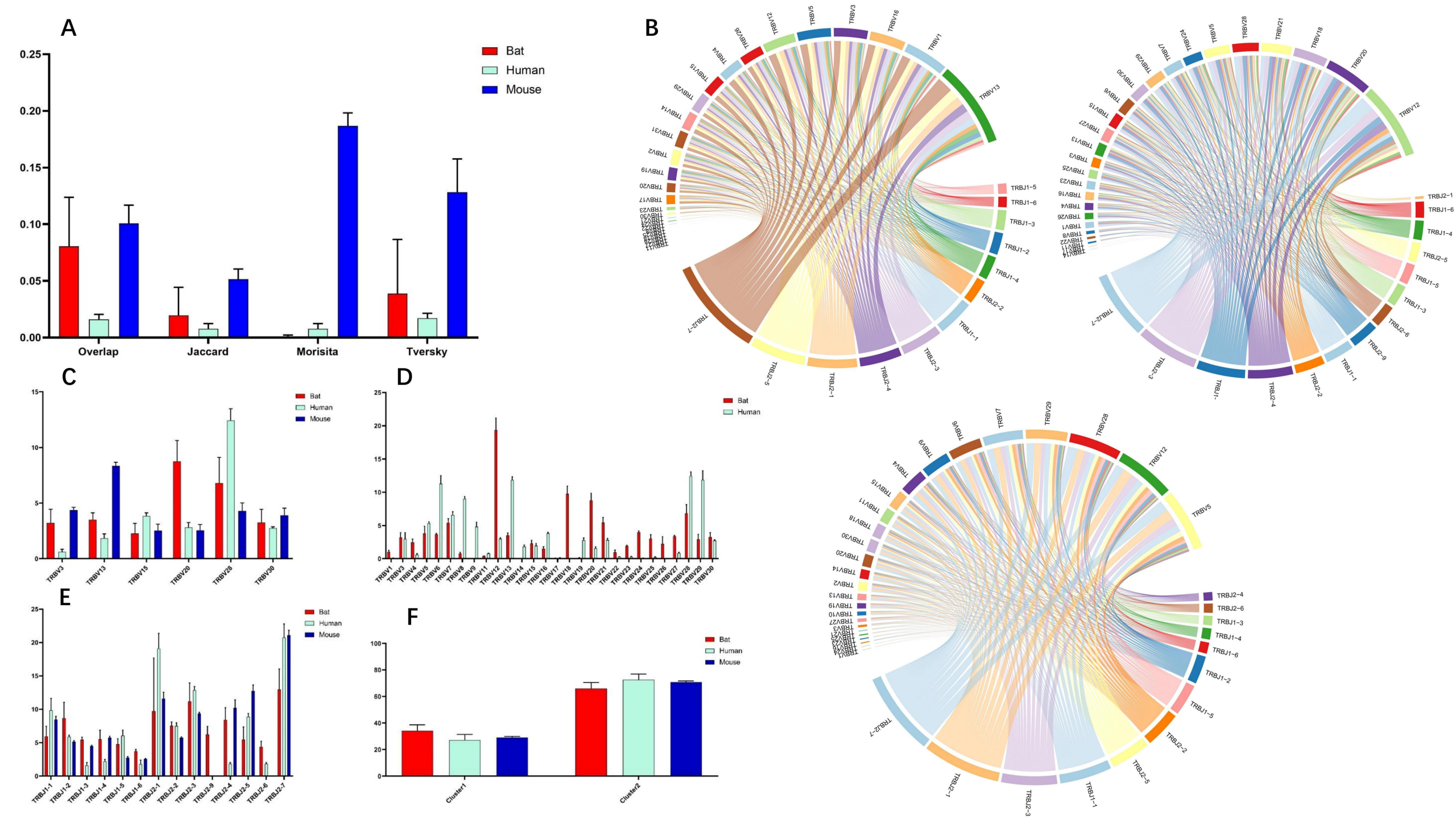


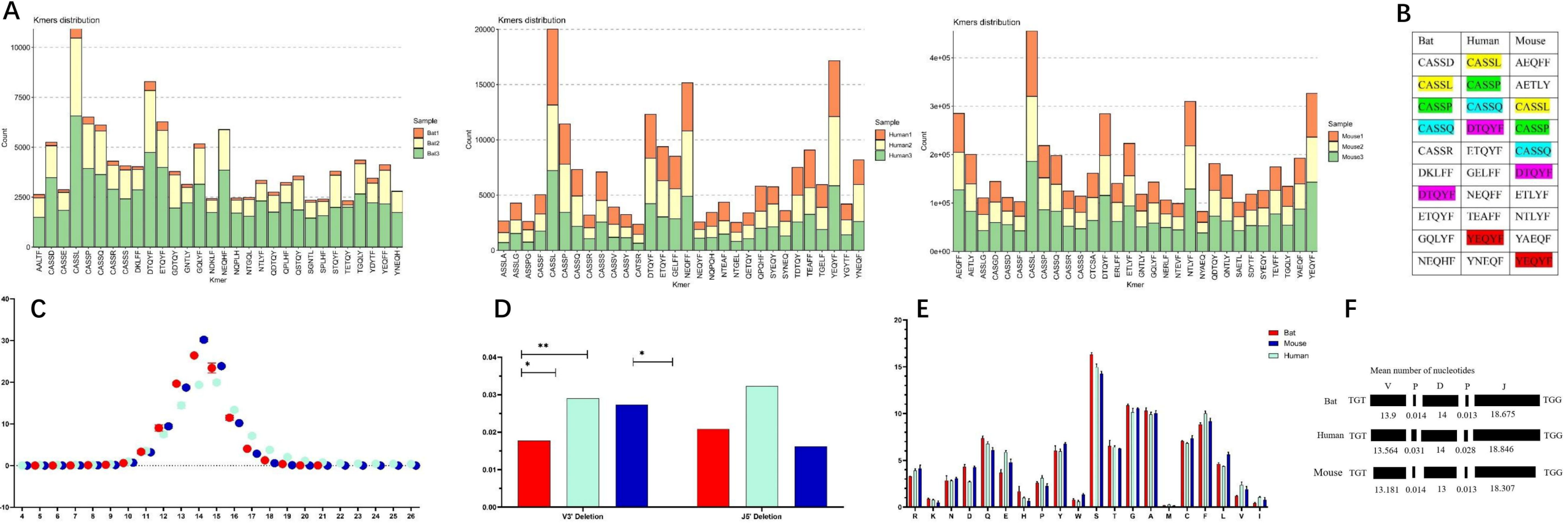
C



D




A
Bat
Human
Mouse
0.25
0.20
0.15
0.10
0.05
0.00
Overlap
Jaccard
Morisita
Tversky
B
Bat
Human
C
Bat
Human
Mouse
TRBV3
TRBV13
TRBV15
TRBV20
TRBV28
TRBV30
D
TRBV1
TRBV3
TRBV4
TRBV5
TRBV6
TRBV7
TRBV8
TRBV9
TRBV11
TRBV12
TRBV13
TRBV14
TRBV15
TRBV16
TRBV17
TRBV18
TRBV19
TRBV20
TRBV21
TRBV22
TRBV23
TRBV24
TRBV25
TRBV26
TRBV27
TRBV28
TRBV29
TRBV30
E
Bat
Human
Mouse
TRBJ1-1
TRBJ1-2
TRBJ1-3
TRBJ1-4
TRBJ1-5
TRBJ1-6
TRBJ2-1
TRBJ2-2
TRBJ2-3
TRBJ2-9
TRBJ2-4
TRBJ2-5
TRBJ2-6
TRBJ2-7
F
Bat
Human
Mouse
Cluster1
Cluster2


A
Kmers distribution
Count
Kmer
Sample
Bat1
Bat2
Bat3
Human1
Human2
Human3
Mouse1
Mouse2
Mouse3
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
C
D
V3' Deletion
J5' Deletion
E
Bat
Mouse
Human
F
Mean number of nucleotides
V
P
D
P
J
Bat
TGT
TGG
13.9
0.014
14
0.013
18.675
Human
TGT
TGG
13.564
0.031
14
0.028
18.846
Mouse
TGT
TGG
13.181
0.014
13
0.013
18.307