



Can extreme MHC class I diversity be a feature of a wide geographic range? The example of Seba's short-tailed bat (*Carollia perspicillata*)

Tamar Qurkhuli¹ · Nina Schwensow¹ · Stefan Dominik Brändel^{1,2} · Marco Tschapka^{1,2} · Simone Sommer¹

Received: 13 May 2019 / Accepted: 14 August 2019 / Published online: 13 September 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The major histocompatibility complex (MHC) is one of the most diverse genetic regions under pathogen-driven selection because of its central role in antigen binding and immunity. The highest MHC variability, both in terms of the number of individual alleles and gene copies, has so far been found in passerine birds; this is probably attributable to passerine adaptation to both a wide geographic range and a diverse array of habitats. If extraordinary high MHC variation and duplication rates are adaptive features under selection during the evolution of ecologically and taxonomically diverse species, then similarly diverse MHC architectures should be found in bats. Bats are an extremely species-rich mammalian group that is globally widely distributed. Many bat species roost in multitudinous groups and have high contact rates with pathogens, conspecifics, and allospecifics. We have characterized the MHC class I diversity in 116 Panamanian Seba's short-tailed bats (*Carollia perspicillata*), a widely distributed, generalist, neotropical species. We have detected a remarkable individual and population-level diversity of MHC class I genes, with between seven and 22 alleles and a unique genotype in each individual. This diversity is comparable with that reported in passerine birds and, in both taxonomic groups, further variability has evolved through length polymorphisms. Our findings support the hypothesis that, for species with a geographically broader range, high MHC class I variability is particularly adaptive. Investigation of the details of the underlying adaptive processes and the role of the high MHC diversity in pathogen resistance are important next steps for a better understanding of the role of bats in viral evolution and as carriers of several deadly zoonotic viruses.

Keywords MHC I exon 2 diversity · Copy number variation · Length polymorphism · Bat · *Carollia perspicillata* · Panama

Introduction

Genes involved in immune protection are evolutionarily important as they influence the survival of individuals. One of the most important vertebrate gene families for adaptive immunity against pathogens is the major histocompatibility complex (MHC) (reviewed by Apanius et al. 1997; Kaufman

2018). Within the MHC, the different genes have evolved specific functions, such that the MHC region of a vertebrate genome usually comprises groups of so-called non-classic and classic MHC genes, the latter of which code for molecules that bind and present antigens to the immune system for recognition and elimination (Klein 1986). Two main classes of classic MHC molecules are involved in adaptive immunity, both presenting antigens to T cells: class I molecules are specialized in presenting antigens from intracellular pathogens such as viruses, whereas class II molecules bind, among others, antigens of extracellular origin, e.g., from bacteria and helminths (Dengjel et al. 2005; Murphy and Weaver 2017; reviewed by Rock et al. 2016). However, several non-classic class I molecules are also known to present peptides to T cells, including HLA-E in humans and Qa1 and M3 in mice (Adams and Luoma 2013; D'Souza et al. 2019; Fischer Lindahl et al. 1997; Hansen et al. 2016).

In most outbred vertebrate species, antigen-presenting MHC genes have a high population-wide number of alleles (Robinson et al. 2000), populations show high rates of

Tamar Qurkhuli and Nina Schwensow shared first authorships

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00251-019-01128-7>) contains supplementary material, which is available to authorized users.

✉ Simone Sommer
Simone.Sommer@uni-ulm.de

¹ Institute for Evolutionary Ecology and Conservation Genomics, University of Ulm, Albert-Einstein Allee 11, 89081 Ulm, Germany

² Smithsonian Tropical Research Institute, Apartado, 0843-03092 Panamá, República de Panamá

heterozygosity (Aguilar et al. 2004; Sommer 2005; Huchard et al. 2010), and the sequence variation among alleles is usually high (Alcaide et al. 2008). The allelic diversity evolves by point mutations, recombination, and an increased non-synonymous over synonymous nucleotide substitution rate, especially within the antigen-binding sites (ABS) (reviewed by Bernatchez and Landry 2003; Sommer 2005). In addition, functional variability can increase through gene duplications. Some old copies have functions that are under strong balancing selection and even outlast host speciation events, a mechanism referred to as trans-species polymorphism (Eirín-López et al. 2012; Klein 1986; Nei et al. 1997). The number of MHC loci often differs between closely related species (reviewed by Kelley et al. 2005) and, even within species, individuals with different numbers of MHC genes are a common finding (Alcaide et al. 2014; Bonhomme et al. 2008; Siddle et al. 2010). Functional differences between the duplicated loci of the MHC can arise through ‘birth-and-death processes’ in which some duplicated genes mutate and gain new functions, whereas others might mutate and become pseudogenes (Nei et al. 1997). Overall, the high MHC diversity found in many species is assumed to reflect historic and current selective and evolutionary processes maintained by some sort of pathogen-driven selection and/or mate choice (reviewed by Sommer 2005).

The highest MHC variability, both in terms of number of individual alleles and gene copies, has so far been found in passerine birds. Sedge warblers (*Acrocephalus schoenobaenus*) possess up to 65 alleles corresponding to a minimum of 33 MHC class I loci per individual (Biedrzycka et al. 2017). Strong evidence has also been found for birth-and-death processes within the MHC of these birds as different groups of alleles showed different degrees of divergence and diversity, and evidence for positive (i.e. diversifying) selection (Biedrzycka et al. 2017). The extremely complex and highly diverse MHC gene region of several passerine bird species has been suggested to be the result of selection experienced by passerines during adaptive radiation into a wide variety of habitats with associated diverse pathogen pressures, combined with small body sizes, a short generation time and, consequently, a rapid rate of evolution (Westerdahl et al. 2005; Alcaide et al. 2013). An extraordinary high MHC diversity combined with a large geographic range (but not small body size and short generation time) has been observed in other bird species (e.g., greater flamingos *Phoenicopterus roseus*: Gillingham et al. 2017; Eurasian coot *Fulica atra*: Alcaide et al. 2014). If a high MHC allelic diversity and duplication rate is a general adaptive mechanism for species-rich animal groups that occupy a variety of different habitats and that experience elevated levels of pathogen contact, similarly diverse MHC architectures to those in passerines and other bird species should occur in bats.

Bats (Chiroptera) are ecologically, immunologically, and epidemiologically highly important species. They diverged from other Eutherian mammals early in evolution (Teeling et al. 2005) and represent the second largest order of mammals with more than 1300 species (Fenton and Simmons 2015). The diverse species of bats have adapted to a variety of habitats, including disturbed environments involving frequent contact with humans and livestock, and have an almost global distribution (Simmons 2005). Bats also represent natural reservoirs for a wide range of pathogens (Luis et al. 2013). A few studies have investigated the MHC constitution of bats. MHC class II DRB variation was studied in neotropical bat species (*Noctilio albiventris*, *N. leporinus* (Schad et al. 2011, a, b), *Molossus molossus*, *Desmodus rotundus* (Salmier et al. 2016); *Myotis velifer*, *M. vivesi* (Richman et al. 2010); *Artibeus jamaicensis* (Real-Monroy et al. 2014); *Carollia perspicillata* (Schad et al. 2012b); *Saccopteryx bilineata* (Mayer and Brunner 2007; Schad et al. 2012b). Surprisingly, despite its importance for viral antigen presentation, studies determining MHC class I gene variability in bat species are rare, but all have found evidence for gene duplications. Santos et al. (2016) have detected at least three expressed MHC class I genes in *S. bilineata*. The Australian black flying fox (*Pteropus alecto*) is reported to have between six (Ng et al. 2016) and nine (Papenfuss et al. 2012) MHC class I genes. To our knowledge, the highest number of MHC class I genes has been observed in the Egyptian rousette (*Rousettus aegyptiacus*) in which 12 genes and seven pseudogenes have been detected (Pavlovich et al. 2018). This is remarkable from an immunological viewpoint, as bats maintain and disseminate more than 80 zoonotic viruses (Olival et al. 2017). Indeed, bats represent natural reservoirs for a wide range of emerging and re-emerging viruses and host a greater number of zoonotic viruses per species than any other taxa (Luis et al. 2013). Examples of bat-borne zoonotic viruses are the rabies virus (RABV) and the related lyssaviruses, Nipah and Hendra viruses and the Ebola virus (Baker et al. 2013; Moratelli and Calisher 2015). Bat-borne coronavirus (CoV; family Coronaviridae) species are the progenitors of the viruses that have caused severe epidemics of acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) with high fatality rates in humans (reviewed by Banerjee et al. 2019). Bats also host astroviruses (AstV; family Astroviridae (Fischer et al. 2017), which are related to human strains (Chu et al. 2008) causing infectious diarrhoea in children, in the elderly and in immunocompromised people (Donato and Vijaykrishna 2017).

Here, we have aimed to characterize the MHC class I diversity in a widely distributed neotropical bat species with a geographical range spanning from Mexico to the southern part

of Brazil (Barquez et al. 2015, Fig. S1), namely Seba's short-tailed bat (*Carollia perspicillata*), in order to test the hypothesis that a bat species that occupies a large geographic range and that is presumably confronted with many pathogens harbors high MHC diversity. Like other bat species, *C. perspicillata* is a natural reservoir and a host of re-emerging pathogens, such as rabies virus (RABV) (Salmón-Mulanovich et al. 2009), Coronaviruses (CoV) (Anthony et al. 2013; Carrington et al. 2008; Corman et al. 2013), Polyomaviruses (PyVs) (Fagrouch et al. 2012), and Astroviruses (AstV) (Wasimuddin et al. 2018). In our study region in Panama, *C. perspicillata* has been found to be prevalent both in undisturbed large forest fragments and in disturbed habitats with potential contact to anthropogenic structures, livestock, and pet animals. We have further aimed to assess evidence for pathogen-driven selection and the evolutionary relationship between MHC class I alleles. Our results add further evidence to the hypothesis that pathogen richness and adaptation to a wide range of different habitats may be a general mechanism influencing rates of MHC gene duplication, length polymorphism, and variability, as recently suggested for passerine and other bird species.

Materials and methods

Study area and sampling

We sampled bats at 11 different locations in the Panama Canal area, Central Panama, between October 2013 and October 2015 (Fig. S1). Six sites were located in continuous tropical lowland rainforest forests surrounding the Gatún lake situated within the 5400 ha of the protected Barro Colorado Nature Monument (BCNM). Five sites were located in rainforest fragments (1.5–51 ha) embedded in an agricultural matrix (Fig. S1). Details of the overall project design are described in Schmid et al. (2018) and Brändel et al. (in review). The capture and sampling procedures followed Panamanian regulations (Autoridad Nacional del Ambiente, República de Panamá), SE/A-75-13 to SE/A-28-17, and were ethically approved by the Smithsonian Tropical Research Institute (IACUC protocol. 2014-0101-2016-2016-0627-2019). Briefly, we trapped bats continuously between October 2013 and October 2015 by using ground-level mist nets (6 m × 2.5 m, 70/2 denier, 16 mm mesh size; ECOTONE, Gdynia, Poland) that were installed at sunset and controlled every 20 min for six consecutive hours. We kept all captured bats temporarily in clean soft cloth bags until processing. After species identification, we took two small (ca. 3 mm²) tissue samples from the wing and stored them in 90% ethanol. Immediately after processing, we released all animals at their trapping site (Brändel et al. (in review)).

MHC class I gene amplification and sequencing

We isolated genomic DNA with an ammonium acetate precipitation method (Nicholls et al. 2000). To design primers, we downloaded and aligned 38 MHC class I exon 2 and 3 sequences from six bat species and from a corresponding human sequence (for species and accession numbers, see Table S1) from Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>) (Sayers et al. 2019) and used the primer design tool primer3 v2.3.7 (Rozen and Skaletsky 2000).

We tested four different primer pairs amplifying MHC I exon 2 and 3 (primer sequences available on request) and performed a test sequencing run with five randomly chosen individuals by using Nanokit Reagent Kit v2 chemistry for 500 cycles on the Illumina MiSeq platform (Illumina, Inc.) and two independent technical replicates per individual. As repeatability (97%) was higher for exon 2 than for exon 3, we accordingly selected the primer pair *Cape-exon-2-forward* (*Cape_f*): 5'-TGAGGTATTTCTACACCGCCG-3' and *Cape-exon-2-reverse* (*Cape_r*): 5'-TGGTTGTAGTAGCCGCNC-3'. These primers flank a 223-base-pair fragment that corresponds to the human MHC class I sequence, which includes functionally relevant antigen-binding sites (Fig. S2).

In order to generate individually barcoded libraries, we tagged amplicons following the approach of the Fluidigm System (Access Array™ System for Illumina Sequencing Systems, © Fluidigm Corporation), which required two consecutive rounds of polymerase chain reactions (PCR). In the first reaction, we amplified the 223-base-pair fragment by using the target-specific primers described above flanked by the common sequence tags CS1 and CS2 as required by the Fluidigm protocol. To optimize cluster identification during sequencing, four random base pairs were added between the CS1 primer and *Cape_f* forward primer (thus using CS1-NNNN-*Cape_f* and CS2-*Cape_r*). The second amplification reaction attached the unique sample barcodes and sequencing adapters to the product of the first PCR (additional information about the laboratory procedures can be found in Supplementary File 1).

We prepared one library (pool) with 48 samples (two technical replicates per individual for 23 animals and two 'no template' PCR controls that had been processed together with the 23 animals). The second library consisted of 192 samples (two technical replicates for 95 individuals and two 'no template' PCR controls) for Illumina-sequencing according to the manufacturer's protocol. To increase the nucleotide diversity of the amplicon libraries, we spiked them with a 6% PhiX Control v3 Library. We performed paired-end sequencing on an Illumina MiSeq machine by using Nanokit Reagent Kit v2 chemistry for 500 cycles for library 1 and Reagent Kit v2 chemistry for 500 cycles for library 2.

For allele calling and genotyping and to eliminate artefacts, we followed our previously published bioinformatics pipeline

(Santos et al. 2016; Sommer et al. 2013; Gillingham et al. 2019). Briefly, we merged the paired reads and cleaned the data from the PhiX sequences with the merging tool of FLASH (Fast Length Adjustment of SHort reads) software (Magoč and Salzberg 2011). The subsequent steps were performed to identify all artefacts. First, we removed low-quality reads (mean Phred quality score < 30) by using the software FASTX-Toolkit v0.0.13 (Gordon and Hannon 2010). Next, chimaera detection and removal were performed by using the default settings of UCHIME (Edgar et al. 2011). Further, we employed a local BLAST tool (Altschul et al. 1990) to compare all the remaining reads against a locally built Mammalian MHC I exon 2 database and to remove all detected non-MHC sequences, i.e., artefacts. Finally, individual fasta files were concatenated into a single file with all the reads aligned. Shannon entropy analysis and oligotyping were performed with the Oligotyping tool v2.2 (Eren et al. 2013). The positions (components) in the alignment with higher-than-background entropy suggesting polymorphic sites were used to cluster divergent variants and to distinguish true alleles from a background variation attributed to sequencing errors, i.e., artefacts. After artefact removal, we subsequently assigned alleles to individuals by using a customized Python script (Santos et al. 2016; Gillingham et al. 2019). Replicates were handled blindly. At the final stage of allele calling, we accepted a sequence as a true MHC allele if it was (1) present in both independent replicates of an individual and (2) if it was (under this condition) genotyped in at least two different individuals. We aligned all identified MHC I exon 2 alleles by using ClustalW 2.1 (Thompson et al. 1994) and named them according to their frequency in the final data set, following the standard nomenclature (Klein et al. 1990). By using our very conservative approach, we underestimated rather than overestimated allelic diversity. All alleles are deposited in GenBank (Accession Numbers MN213375–MN213615).

Data analysis

MHC allelic diversity estimates were calculated using MEGA5.2 with bootstrapping at 1000 (Tamura et al. 2011) under the option of ‘pairwise deletion’ if indels were present. Mean pairwise dissimilarity over all *C. perspicillata* amino acid pairs at a given alignment position was calculated in Geneious 11.1.5 (<https://www.geneious.com>) (Kearse et al. 2012) and presented as 1–(minus) ‘sequence identity’ (identity is 1 when all the amino acids are the same at that location of the alignment; it is 0 when all the amino acids are different).

We investigated the possible occurrence of recombination events between all identified sequences by using the exploratory recombination scan implemented in the program RDP v.4.95 (Martin et al. 2015) by applying the methods RDP (Martin and Rybicki 2000), GENECONV (Sawyer 1999), CHIMAERA (Posada and Crandall 2001), MAXCHI (Smith

1992), BOOTSCAN (Martin et al. 2005), SISCAN (Gibbs et al. 2000), 3SEQ (Boni et al. 2007), and LARD (Holmes et al. 1999) with a significance threshold of $p < 0.05$. After excluding any sequence with evidence of recombination, we tested for codons under positive selection (i.e., positions with increased amino acid variation) by using the HyPhy software (Kosakovsky Pond et al. 2005) available at the Datamonkey public webserver (Delport et al. 2010; Kosakovsky Pond and Frost 2005a). We applied the methods SLAC (single-likelihood ancestor counting), FEL (fixed effects likelihood) and REL (random effects likelihood) (Kosakovsky Pond and Frost 2005b), MEME (mixed effects model of evolution) (Murrell et al. 2012) and FUBAR (fast, unconstrained Bayesian approximation) (Murrell et al. 2013). For MEME, FEL, and SLAC selection tests, significance was set to $p = 0.1$ and, for FUBAR, we present the results where posterior probability was 0.9. We also investigated signs of positive selection by using CodeML integrated in the program PAML 4.9 (Yang 2007), which is based on maximum likelihood procedures of different models of nucleotide sequence evolution to identify species-specific positively selected codon sites ($\omega = dN/dS > 1$). We explored the models M1a (nearly neutral), M2a (positive selection), M7 (β : ω ratio varies among sites according to a β -distribution), and M8 ($\beta + \omega$: similar to M7 but with an additional site class that allows $\omega > 1$). The models M2a and M8 assume that codons are under the pressure of positive selection and are tested against the null models M1a and M7, which are based on neutrality theory (i.e., that the majority of mutations are synonymous and therefore are neither selected for nor selected against) (Kimura 1977, 1983). A posterior probability that each identified site is under positive selection was calculated by using the empirical Bayes procedure (Yang et al. 2005). We accepted sites being under positive selection when the Bayes Empirical Bayes (BEB) was supported at least at the 95% confidence interval level. Overall, we considered a site to be under positive selection when this was supported by at least four out of six selection tests.

To depict the evolutionary relationship between the MHC class I exon 2 alleles, we aligned three representative amino acid sequences of *C. perspicillata* with MHC class I exon 2 amino acid sequences of other mammals: human (*Homo sapiens*; Genbank accession number: NG_023187), black flying fox (*Pteropus alecto*; sequence 1 derived from KP862827.1; sequence 2: KP862825.1; sequence 3: KP862826.1), horse (*Equus caballus*; NC_009163.3), cow (*Bos taurus*; NC_037350.1), dog (*Canis lupus familiaris*; NC_006600.3), pig (*Sus scrofa*; AB646190.1), opossum (*Monodelphis domestica*; sequence 1: NM_001079820.1; sequence 2: NM_001171835.1) and platypus (*Ornithorhynchus anatinus*; AY112715.1). A neighbor-net phylogenetic network (Bryant and Moulton 2004) was built by using the program SplitsTree4 (Huson and Bryant 2006).

Results

Sequencing results

After excluding two individuals for which the replicate failed in the sequencing run, we successfully genotyped 116 individuals of *C. perspicillata*. In total, the two sequencing runs produced 11,383,399 reads (average depth per sample = 48,234; SD = 21,873). Sequencing depth did not affect the number of identified alleles per sample (data not shown); therefore, we concluded that a depth of approximately 20,000 reads was sufficient to identify even rare alleles in a single genotype. The allele calling repeatability (i.e., congruence of two individual replicates) was 92%. In 8% of the samples, the replicates did not fully overlap and single alleles with lower read coverage were not identified in both replicates. We subsequently took a conservative approach (i.e., preferring to underestimate diversity) by confirming only those alleles that were identified in both independent replicates of at least two individuals.

Diversity and phylogenetic relationships of MHC class I exon 2 alleles

We detected 241 unique nucleotide sequences (subsequently referred to as ‘alleles’) (Cape-A*001–Cape-A*241, Fig. S2). One allele, Cape-A*008, contained a stop codon and thus probably belonged to a non-expressed pseudogene. All the other 240 alleles could be translated into 208 unique amino acid sequences (Fig. S2).

The observed *C. perspicillata* alleles differed in their length and could be assigned to three different groups (Table 1). The majority of alleles had the expected length of 223 bp. The length was expected according to our designed primers and an alignment of related species (Table S1). Hereafter, we refer to these as ‘no deletion’ alleles. The second group of alleles had a length of 211 bp; these alleles exhibited a deletion of 12 bps (i.e., four codons, hereafter referred to as ‘4 codon deletion’ alleles). The third group of alleles was even shorter at only 208 bp in length. They showed a 15-bp deletion (i.e.,

five codons, hereafter referred to as ‘5 codon deletion’ alleles). Most of the identified nucleotide alleles ($N = 200$) belonged to the ‘no deletion’ alleles, $N = 19$ alleles were ‘5 codon deletion’ alleles and $N = 22$ were ‘4 codon deletion’ alleles (Table 1). The group of ‘4 codon deletion’ alleles was least polymorphic at both the nucleotide and amino acid levels ($d_{\text{nuc}} = 0.015$; $d_{\text{aa}} = 0.031$) compared with the ‘no deletion’ alleles ($d_{\text{nuc}} = 0.091$; $d_{\text{aa}} = 0.166$) and the ‘5 codon deletion’ alleles ($d_{\text{nuc}} = 0.096$; $d_{\text{aa}} = 0.184$).

A comparison of the mean pairwise dissimilarity over all *C. perspicillata* amino acid sequences at a given alignment position showed that the variability of ‘no deletion’ and ‘5 codon deletion’ alleles was concentrated on similar positions in these allele groups (Fig. 1).

An alignment of the representatives of each group of *C. perspicillata* MHC class I exon 2 amino acid alleles with the MHC class I sequences of other mammals demonstrated that ‘no deletion’ alleles were similar to the alleles of the black flying fox (*P. alecto*), ‘4 deletion’ alleles were different from all the other sequences compared. The ‘5 codon deletion’ alleles had a similar location of the deletions and thus a similar architecture across diverse groups of mammals (Fig. 2).

The network analysis revealed a well-supported cluster consisting of ‘4 codon deletion’ alleles only (Fig. 3). In contrast, the ‘5 codon deletion’ alleles were not monophyletic. They either clustered with other ‘5 codon deletion’ alleles or were more closely related to ‘no deletion’ alleles. No further clearly separated clusters were observed.

The MHC class I exon 2 alleles differed in their frequency in the overall data set. Moreover, within each group of alleles, we detected alleles that were clearly more frequent than all others (Fig. S3). Whereas 35% of the group of ‘no deletion’ alleles (i.e., 70 out of 200 alleles) occurred in more than five individuals, most alleles of the ‘5 deletion’ (12 out of 19, 63%) and ‘4 deletion’ group (19 out of 22, 86%) were rare, i.e., found in < 5 individuals (Fig. S3). The allele Cape-A*008, which contained a stop codon, was found in 27 individuals (23%).

The total number of nucleotide alleles per individual differed between individuals and ranged from 7 ($N = 3$ individuals, 3% of the samples) to 22 ($N = 2$; 2%, Fig. 4) suggesting

Table 1 Sequence polymorphism (measured as genetic distance) between *C. perspicillata* MHC class I exon 2 alleles. The overall mean distance was calculated among all alleles of each group (i.e., ‘no deletion’, ‘5 codon deletion’, and ‘4 codon deletion’ alleles) and among

	# Alleles _{nuc}	Length _{nuc}	d_{nuc}	# Alleles _{aa}	Length _{aa}	d_{aa}	PSS
‘No deletion’ alleles	200	223	0.091 ± 0.011	178	73	0.166 ± 0.029	12
‘5 codon deletion’ alleles	22	208	0.096 ± 0.012	17	68	0.184 ± 0.028	2
‘4 codon deletion’ alleles	19	211	0.015 ± 0.005	13	69	0.031 ± 0.013	0
All alleles	241	–	0.122 ± 0.011	208	–	0.224 ± 0.027	–

PSS positively selected sites

all MHC I exon2 alleles for both nucleotide (nuc) and amino acid (aa) sequences \pm standard error. The first and the last codons of the nucleotide sequences are incomplete and were excluded when translating nucleotide alleles into amino acid alleles

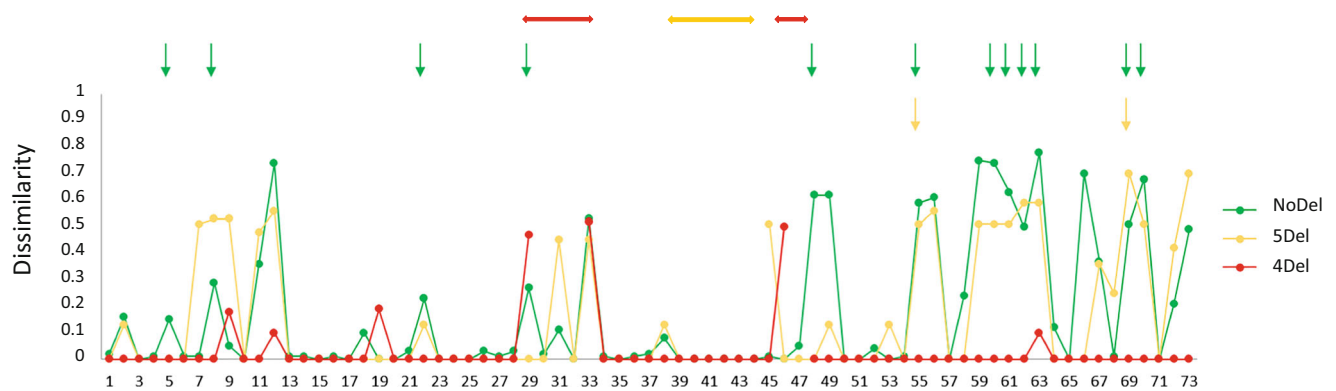


Fig. 1 Variability and positively selected sites of *C. perspicillata* MHC class I exon 2 amino acid alleles. Sequence dissimilarity per amino acid position is shown separately for the alignments of ‘no deletion’ alleles (green), ‘5 codon deletion’ alleles (yellow) and ‘4 codon deletion’ alleles (red). Green arrows indicate positively selected sites identified by ≥ 4 out

of 6 methods of selection testing in ‘no deletion’ alleles, whereas yellow arrows indicate those under positive selection in ‘5 codon deletion’ alleles. The location of the indel positions in the alignment (Fig. S1) is marked with double-headed arrows

gene copy number variation and the presence of between at least 4 and 11 loci per individual. The average number of alleles per individual was 13.13 (± 3.38 SD). All individuals carried between 5 and 18 ‘no deletion’ alleles (mean 11.25 ± 2.91 SD, Fig. 4). Whereas 13 (11.2%) individuals had only ‘no deletion’ alleles, all other individuals had additionally only ‘4 deletion’ alleles (29.3%), only ‘5 deletion’ alleles (18.1%) or both (41.4%). All individuals that carried ‘4 deletion alleles’ always carried either one (38%) or two (33%) alleles of this type (Fig. 4). Similarly, carriers of ‘5 deletion’ alleles had either one (38%) or two (21%) alleles of this type, with two exceptions (two individuals carried three and four alleles of ‘5 deletion’ alleles, respectively). Irrespective of the total number of alleles, each investigated individual had a unique genotype, i.e., a unique combination of MHC I exon 2 alleles. The individual genotypes are indicated in the Supplementary File 2.

Tests of recombination and selection

In the alignment of all alleles (i.e., all three length groups aligned together), we detected evidence for recombination in

two alleles (Table S2). Additionally, we aligned each group of alleles separately. In these three within-group alignments, we detected recombination signals in three of the ‘no deletion’ alleles and in four of the ‘5 codon deletion’ alleles with at least one of the eight applied methods. We did not find evidence for recombination in the ‘4 codon deletion’ alleles. In total, nine alleles were identified as recombinant (Table S2). All nine alleles for which we found evidence for recombination in each test (plus the allele with a stop codon) were removed from the subsequent analyses.

We detected no evidence for positive selection on ‘4 codon deletion’ alleles (Figs. 1 and S2). In contrast, we found strong evidence for positive selection on two codons of the ‘5 codon deletion’ alleles (amino acids 50 and 64) and on 12 codons of the ‘no deletion’ alleles (amino acids 5, 8, 22, 29, 48, 55, 60, 61, 62, 63, 69, 70, Figs. 1 and S2). This difference might be attributable to the overall higher number of ‘no deletion’ alleles, because the testing of ten subsets of 20 randomly chosen ‘no deletion’ alleles revealed a similar number of positively selected sites (mean $2.5 \pm$ SD 1.35) as in ‘5 codon deletion’ alleles. Two positively selected sites of the ‘5 codon deletion’ alleles fell on the same location as the positively selected sites

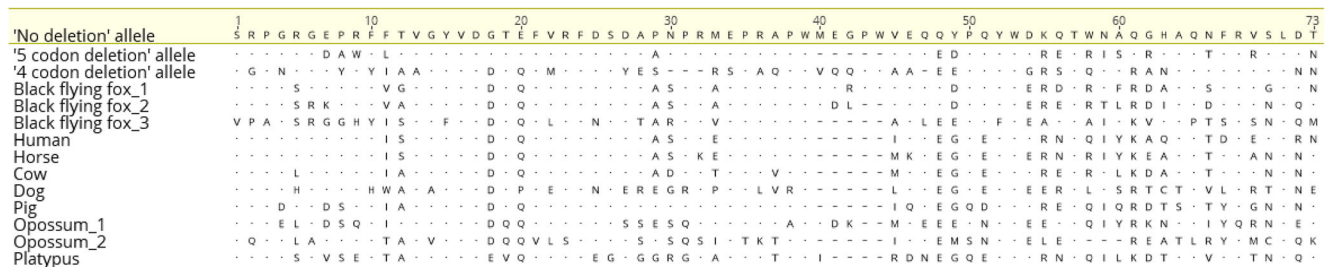


Fig. 2 Position of indels in the MHC class I gene of selected mammals. Three representatives of *C. perspicillata* MHC class I exon 2 amino acid alleles (one of each length group) are aligned with sequences from human (*Homo sapiens*), black flying fox (*Pteropus alecto*), horse (*Equus caballus*), cow (*Bos taurus*), dog (*Canis lupus familiaris*), pig (*Sus scrofa*), opossum (*Monodelphis domestica*), and platypus

(*Ornithorhynchus anatinus*) (Table S1). The alignment starts at amino acid position 13 of the human molecule. The indels of the ‘5 codon deletion’ alleles *C. perspicillata* align to the same location as the insertions of other species’ MHC alleles, indicating a common characteristic or evolutionary important trait in bats

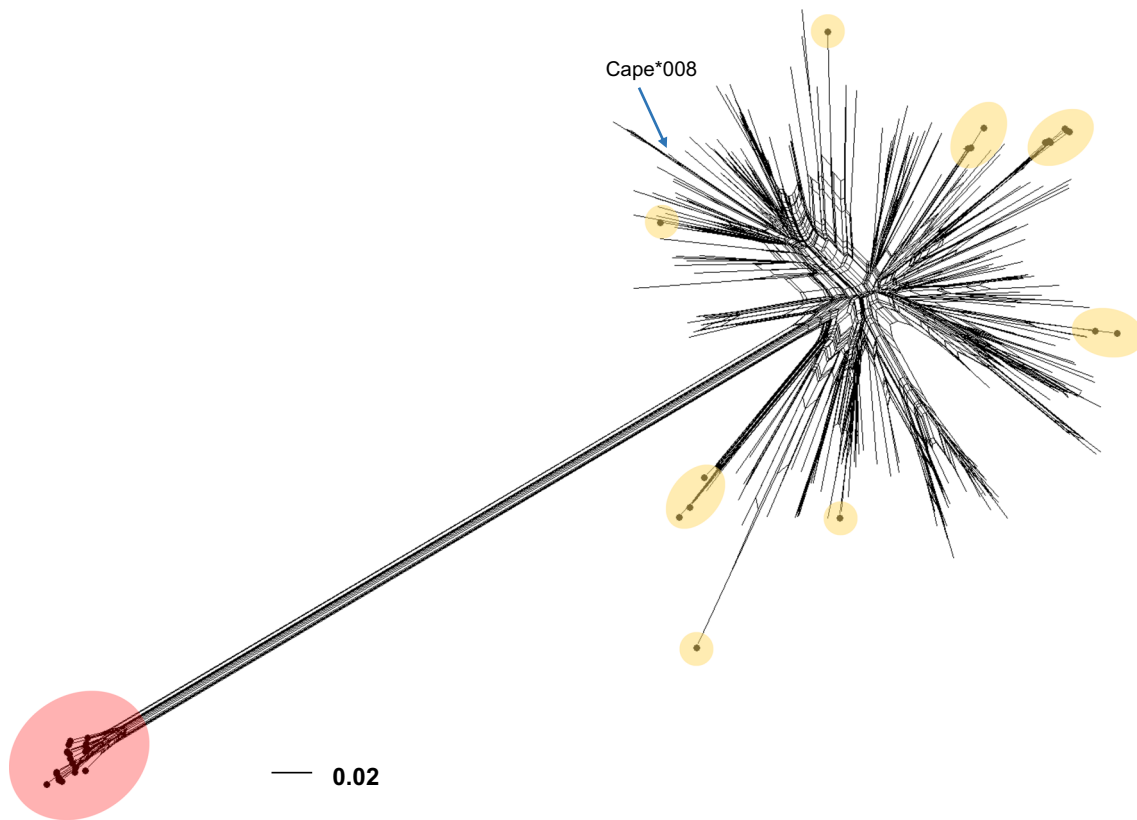


Fig. 3 A Neighbour-net network depicting the evolutionary relationships between the *C. perspicillata* MHC I exon 2 alleles. ‘4 codon deletion’ alleles cluster together separately from all the other alleles, whereas ‘5 codon deletion’ alleles are intermixed with ‘no deletion’ alleles and

neither of these two groups is monophyletic. ‘4 codon deletion’ alleles are marked with red circles. ‘5 codon deletion’ alleles are marked with yellow circles. The allele Cape-A*008 containing a stop codon is indicated

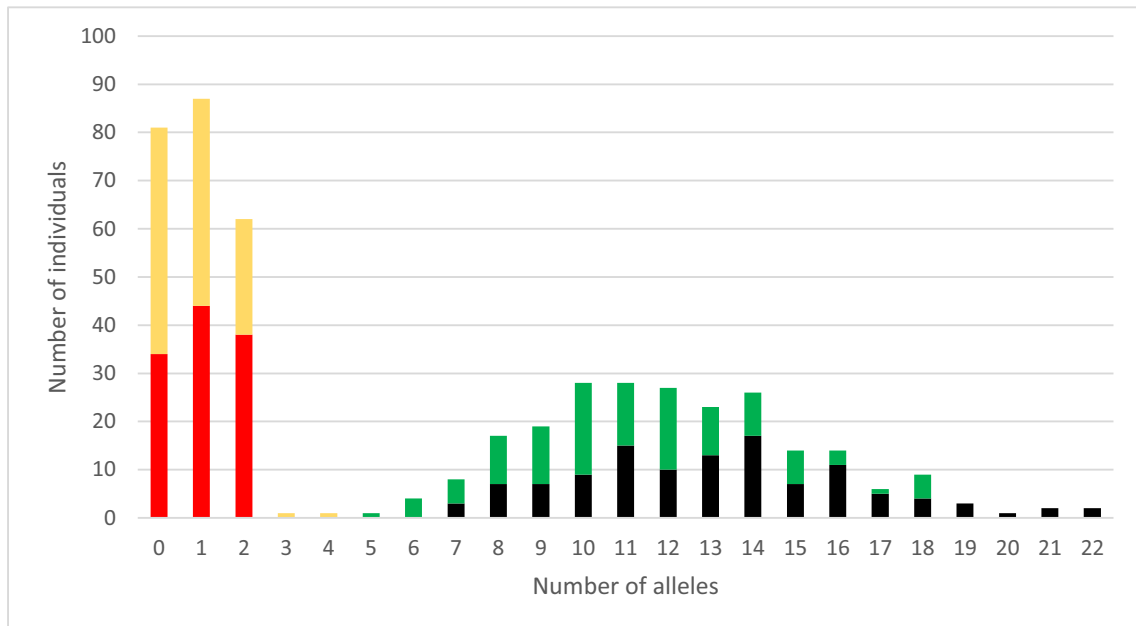


Fig. 4 Frequency distribution of the number of MHC class I exon 2 alleles observed in *Carollia perspicillata* individuals. Shown are the counts of all alleles together (black, 7–22 alleles per individual), only

‘no deletion’ alleles (green, 5–18 alleles per individual), only ‘4 codon deletion’ alleles (red, 0–2 alleles per individual) and only ‘5 codon deletion’ alleles (yellow, 0–5 alleles per individual)

in ‘no deletion’ alleles (amino acids 55 and 69 in the alignment of all 241 peptide alleles). The positions of the positively selected sites in ‘no deletion’ alleles (amino acids 55, 69, and 70) matched or was in close proximity to known ABS in the corresponding human MHC sequence (Bjorkman and Parham 1990). Further sites corresponding to human ABSs were indicated as positively selected in *C. perspicillata* only with one or two methods (amino acids 12, 33, 65, Fig. S2). Not all the variable sites of the sequences were under positive selection or the selection signal was too weak to be detected in ‘5 codon deletion’ and ‘no deletion’ alleles (Fig. 1).

Discussion

We have tested the hypothesis that species with a large geographic range and that have contact with many pathogens harbour a diverse immunogenic repertoire. We characterized the MHC class I exon 2 diversity of the neotropical Seba’s short-tailed bat, *C. perspicillata*, which is a well-known reservoir species for zoonotic virus infections, and which is widely distributed in various habitats all over Central America (Fig. S1). We found that the degree of MHC I exon 2 variability in terms of number of alleles, gene duplications and length polymorphism is indeed extremely high. Individuals carry between seven and 22 different alleles, with each sequenced bat having a unique genotype (despite a highly conservative bioinformatics pipeline). We also found nine possibly recombinant alleles in total; these were removed prior to selection analysis. We detected various signs of pathogen-mediated selection acting on the MHC class I loci, such as high nucleotide diversity among the observed 240 MHC class I alleles and high d_N/d_S rates in potentially antigen binding sites of sequences. Possible underlying selective mechanisms are balancing selection or directional positive (diversifying) selection, both of which are difficult to discriminate in empirical studies of wild species (Spurgin and Richardson 2010). Few alleles occurred at higher frequency, whereas the majority of alleles were present only in a few individuals. The diversity observed in *C. perspicillata* is comparable with the variability reported from the passerine bird species *Arcrocephalus schoenobaenus* (Biedrzycka et al. 2017) and seems to support the hypotheses that, particularly in species with a geographically broad range, a high MHC class I variability indicates high adaptability that might be selected for during range extension (Otting et al. 2005; Prugnolle et al. 2005). *Arcrocephalus schoenobaenus* has an even wider geographical range than *C. perspicillata*, spanning from its breeding regions in the Northern Europe through to Eastern Asian and Russia and its migration route and wintering grounds covering nearly the whole African continent (Fig. S1). The *C. perspicillata* MHC class I diversity also exceeds the high variability reported for the three previously studied bat species

with large geographic ranges (Santos et al. 2016; Ng et al. 2016; Papenfuss et al. 2012; Pavlovich et al. 2018). Overall, all these findings are in congruence with the hypothesis that adaptation during dispersal over large geographic ranges facilitates the evolution highly diverse MHC class I genes.

Moreover, we determined a complex organization of MHC loci as a result of length polymorphism caused by insertions or deletions (indels) and gene duplication events involving copy number variation (CNV), thus increasing the adaptive immunogenetic variation in MHC genes, both at an individual and at a species level. The 241 identified alleles could be grouped into ‘no deletion’ alleles (223 bp length), into ‘4 codon deletion’ alleles (211 bp), or into ‘5 codon deletion’ alleles (208 bp). The alleles had an uneven frequency distribution and all individuals carried ‘no deletion’ alleles. Whereas 11.2% had only ‘no deletion’ alleles, 29.3% carried additionally ‘4 codon deletion’ alleles, 18.1% possessed additionally ‘5 codon deletion’ alleles, and 41.4% possessed additionally both types of ‘deletion’ alleles. Interestingly, less than half of the genotyped animals (41%) carried alleles from all three groups. The length variation among MHC alleles did not cause a reading frame shift. Of note, the three observed groups of alleles (‘no deletion’, ‘4 codon deletion’, ‘5 codon deletion’) showed different signatures of selection suggesting a change in function. The group of ‘no deletion’ alleles was the most variable with 200 nucleotide alleles and 12 amino acid positions with strong evidence for positive selection, especially at the positions that are antigen binding sites in the corresponding human sequence or are in close proximity to them (Bjorkman and Parham 1990). Aligning the ‘no deletion’ and ‘5 codon deletion’ alleles showed that the variability concentrated on similar positions in these allele groups. Positively selected sites in ‘5 codon deletion’ alleles matched with the positively selected sites in ‘no deletion’ alleles. Although the 19 identified ‘5 deletion’ alleles also showed evidence of positive selection, fewer positions were clearly identified. This difference might be attributable to the overall much higher number of ‘no deletion’ alleles, because the testing of ten subsets of 20 randomly chosen ‘no deletion’ alleles revealed a similar number of positively selected sites as in ‘5 codon deletion’ alleles. Another explanation might be that the ‘5 codon deletion’ alleles are losing or changing their function and are no longer under strong selection pressure. In contrast to the two groups of ‘no deletion’ alleles and ‘5 codon deletion’ alleles, no evidence for positive selection on the ‘4 codon deletion’ alleles was observed. Together with the lower genetic distance among alleles and the formation of a distinct cluster in the network analysis, this suggests that the ‘4 codon deletion’ alleles are subject to a different selection pressure that might have evolved a different function.

Indels, together with gene duplication events and allele rearrangements, play an important role in MHC class I evolution (Hosomichi et al. 2008; Miska et al. 2002). Although

indels of various length and location are more frequent in introns than in exons (Hughes and Yeager 1997), they have been observed in MHC class I and II genes of both mammalian and non-mammalian species (Minias et al. 2018; Santos et al. 2018; Schaschl et al. 2008). Indels are considered as the main path to genomic divergence between closely related species such as humans and chimpanzees (Anzai et al. 2003). Indels in MHC genes increase the structural diversity of the molecule and, therefore, expand the antigen-binding repertoire. This enhances the adaptive potential of an individual immune system. Indels in classic MHC genes are common in birds, especially in passerines, and are thought to have an adaptive role (Biedrzycka et al. 2017; Bonneaud et al. 2004; Minias et al. 2018; Balasubramaniam et al. 2017). The same might be true for Chiroptera, as the pattern of indels in the MHC genes is similar between these two evolutionary divergent groups. For bats, length variation (3 and 5 amino acid indels) has been reported in the MHC class I genes of the black flying fox (*Pteropus alecto*), and in the Chinese horseshoe bat (*Rhinolophus sinicus*), the common vampire bat (*Desmodus rotundus*), the great roundleaf bat (*Hipposideros armiger*) and in the natal long-fingered bat (*Miniopterus natalensis*) (Ng et al. 2016; Abduriyim et al. 2019). The indels of the *C. perspicillata* alleles align to the same location as the indels in the other bat species. Compared to eutherian mammals other than bats, they apparently code for alleles with longer antigen binding domains (Abduriyim et al. 2019). This suggests a common characteristic or evolutionary important trait in bats: it may cause an increased binding spectrum, as a longer allele (3-codon indel allele Ptal-N01:01) present antigens with a broader length range than shorter MHC class I molecules from other mammals (Wynne et al. 2016).

The MHC gene families possibly evolve through birth-and-death processes (Eirín-López et al. 2012) and produce copy number variation (CNV) in MHC gene families as the results of gene duplication events. CNVs are common in various taxa, varying between species and within a species (Doxiadis et al. 2006; de Groot et al. 2016; Siddle et al. 2010; van der Wiel et al. 2018). CNVs might have different levels of gene expression (classified as major and minor expression) and might be an alternative to an allelic polymorphism at one gene (Otting et al. 2005). In *C. perspicillata*, gene duplications apparently happened particularly frequently with ‘no deletion’ alleles. Individuals carry between five and 18 of such alleles, suggesting that gene CNV occurs between individuals and that at least three to nine loci have been amplified. Similarly, MHC class I gene duplication (at least six loci) has also been found in another bat species (Ng et al. 2016). We think that, in *C. perspicillata*, the variation in the number of amplified loci is unlikely to be attributable to null alleles because we performed extensive primer testing experiments and because all the primer pairs tested gave the results indicating CNV.

In our network analysis, ‘no deletion’ and ‘5 codon deletion’ alleles clustered together, whereas the ‘4 codon deletion’ alleles formed a single distinct cluster. No distinct clusters indicating locus-specific phylogenetic relationships are apparent within the ‘no deletion’ alleles. This suggests that gene duplication and/or gene conversion happened only recently. As a consequence, some alleles might still be shared between loci (Reusch et al. 2004) and/or potential gene conversion and inter-locus recombination events could blur sequence differences between these loci in *C. perspicillata*. As mentioned above, ‘no deletion’ and ‘5 codon deletion’ alleles cluster together. One possible explanation for this close relationship is that the ‘5 codon deletion’ alleles evolved several times. In the network, ‘5 codon deletion’ alleles are usually placed apart and do not form a monophyletic group (Fig. 3). On the other hand, almost all individuals (with two exceptions) carried either one or two ‘5 codon deletion’ alleles. The simplest interpretation is that only a single locus for this group of alleles is present. Another possible explanation is that frequent recombination between the alleles lead to the observed results of the network analysis. The recombination tests that we applied did not find strong evidence for recombination. However, their power may be limited if the recombination is frequent and the recombination blocks are narrow (‘micro recombination’), as has been suggested for MHC genes (Biedrzycka et al. 2017; Klitz et al. 2012). In contrast to this, the ‘4 codon deletion’ alleles formed a single distinct cluster. All individuals that carried a ‘4 codon deletion’ allele (71% of the study subjects) harbored either one or two different ‘4 codon deletion’ alleles suggesting that a single locus is involved that has evolved earlier in time.

The common pattern shared between passerines and bats suggests that length variation and gene duplications cause important adaptive variation affecting the structural organization of the MHC molecules that bind and present antigens to the T cells. As a consequence of length variation, the primary structure of the MHC molecule changes; it will also have a different folding pattern and thus specificity to a different antigen molecule. This might increase the spectrum of pathogens an organism and a population as a whole can resist, supporting the idea that, in species with a geographically broad range, a high MHC class I variability is particularly adaptive. Further studies focusing on expression differences of MHC alleles and loci are required (Schwensow et al. 2019). Drews et al. (2017) reported expression level differences between several MHC class genes with high and low polymorphism in three different passerine species (*Passer domesticus*, *P. hispaniolensis*, *P. monachus*). Only 50% of the highly polymorphic alleles were found to be expressed (at a variable level), whereas alleles with low polymorphism had uniformly low expression, suggesting functional differences between classic and non-classic MHC class I genes (Drews et al. 2017). The extraordinary genetic diversity at the nucleotide level of MHC class I

exon 2 of *C. perspicillata* should also be translatable into expressed/functional (peptide) diversity. Future investigations aimed at understanding whether and how MHC genes and alleles are associated with differences in disease resistance should elucidate their adaptive role and the unsolved mystery of asymptomatic viral infections in bats.

Acknowledgments We thank the Smithsonian Tropical Research Institute in Panamá for providing the essential infrastructure and, especially, Oris Acevedo and Belkys Jimenez for their constant help during our fieldwork. We are grateful to Rachel Page for all her support in realizing this project. We extend our thanks to all field assistants, to Kerstin Wilhelm and Ulrike Stehle for excellent technical assistance, and to Pablo Santos and all team members of the Sommer lab for fruitful discussions. We are also grateful to Theresa Jones for language editing.

Funding information This research was funded by the German Science Foundation (DFG) and is part of the DFG Priority Program SPP 1596/2 Ecology and species barriers in emerging infectious diseases (SO 428/ 9-1, 9-2; TS 81/7-1, 7-2).

References

- Abduriyim S, Zou D-H, Zhao H (2019) Origin and evolution of the major histocompatibility complex class I region in eutherian mammals. *Ecol Evol* 9(13):7861–7874. <https://doi.org/10.1002/ece3.5373>
- Adams EJ, Luoma AM (2013) The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. *Annu Rev Immunol* 31(1):529–561. <https://doi.org/10.1146/annurev-immunol-032712-095912>
- Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc Natl Acad Sci* 101:3490–3494. <https://doi.org/10.1073/pnas.0306582101>
- Alcaide M, Edwards SV, Negro JJ et al (2008) Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (*Falco naumanni*). *Mol Ecol* 17:2652–2665. <https://doi.org/10.1111/j.1365-294X.2008.03791.x>
- Alcaide M, Liu M, Edwards SV (2013) Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ* 1:e86. <https://doi.org/10.7717/peerj.86>
- Alcaide M, Muñoz J, la Puente JM et al (2014) Extraordinary MHC class II B diversity in a non-passerine, wild bird: the Eurasian coot *Fulica atra* (Aves: Rallidae). *Ecol Evol* 4:688–698. <https://doi.org/10.1002/ece3.974>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Anthony SJ, Ojeda-Flores R, Rico-Chavez O, Navarrete-Macias I, Zambrana-Torrel CM, Rostal MK, Epstein JH, Tipps T, Liang E, Sanchez-Leon M, Sotomayor-Bonilla J, Aguirre AA, Avila-Flores R, Medellin RA, Goldstein T, Suzan G, Daszak P, Lipkin WI (2013) Coronaviruses in bats from Mexico. *J Gen Virol* 94:1028–1038. <https://doi.org/10.1099/vir.0.049759-0>
- Anzai T, Shiina T, Kimura N, Yanagiya K, Kohara S, Shigenari A, Yamagata T, Kulski JK, Naruse TK, Fujimori Y, Fukuzumi Y, Yamazaki M, Tashiro H, Iwamoto C, Umehara Y, Imanishi T, Meyer A, Ikeo K, Gojobori T, Bahram S, Inoko H (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proc Natl Acad Sci* 100:7708–7713. <https://doi.org/10.1073/pnas.1230533100>
- Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Crit Rev Immunol* 17:179–224. <https://doi.org/10.1615/CritRevImmunol.v17.i2.40>
- Baker ML, Schountz T, Wang L-F (2013) Antiviral immune responses of bats: a review. *Zoonoses Public Health* 60:104–116. <https://doi.org/10.1111/j.1863-2378.2012.01528.x>
- Balasubramaniam S, Mulder RA, Sunnucks P, Pavlova A, Melville J (2017) MHC class II β exon 2 variation in pardalotes (Pardalotidae) is shaped by selection, recombination and gene conversion. *Immunogenetics* 69:101–111. <https://doi.org/10.1007/s00251-016-0953-7>
- Banerjee A, Kulcsar K, Misra V, Frieman M, Mossman K (2019) Bats and coronaviruses. *Viruses* 11:41. <https://doi.org/10.3390/v11010041>
- Barquez R, Perez S, Miller B, Diaz M (2015) *Carollia perspicillata*. The IUCN red list of threatened species 2015: e.T3905A22133716. <https://doi.org/10.2305/IUCN.UK.2015-4.RLTS.T3905A22133716.en>. Downloaded on 21 June 2019
- Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol* 16:363–377. <https://doi.org/10.1046/j.1420-9101.2003.00531.x>
- Biedrzycka A, O'Connor E, Sebastian A, Migalska M, Radwan J, Zając T, Bielański W, Solarz W, Ćmiel A, Westerdahl H (2017) Extreme MHC class I diversity in the sedge warbler (*Acrocephalus schoenobaenus*); selection patterns and allelic divergence suggest that different genes have different functions. *BMC Evol Biol* 17: 159. <https://doi.org/10.1186/s12862-017-0997-9>
- Bjorkman PJ, Parham P (1990) Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu Rev Biochem* 59:253–288. <https://doi.org/10.1146/annurev.bi.59.070190.001345>
- Bonhomme M, Doxiadis GG, Heijmans CM et al (2008) Genomic plasticity of the immune-related Mhc class I B region in macaque species. *BMC Genomics* 9:514. <https://doi.org/10.1186/1471-2164-9-514>
- Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176: 1035–1047. <https://doi.org/10.1534/genetics.106.068874>
- Bonneaud C, Sorci G, Morin V et al (2004) Diversity of Mhc class I and IIB genes in house sparrows (*Passer domesticus*). *Immunogenetics* 55:855–865. <https://doi.org/10.1007/s00251-004-0648-3>
- Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21:255–265. <https://doi.org/10.1093/molbev/msh018>
- Carrington CVF, Foster JE, Zhu HC, Zhang JX, Smith GJD, Thompson N, Augustine AJ, Ramkissoon V, Adesiyun AA, Guan Y (2008) Detection and phylogenetic analysis of group 1 coronaviruses in south American bats. *Emerg Infect Dis* 14:1890–1893. <https://doi.org/10.3201/eid1412.080642>
- Chu DKW, Poon LLM, Guan Y, Peiris JSM (2008) Novel Astroviruses in insectivorous bats. *J Virol* 82:9107–9114. <https://doi.org/10.1128/JVI.00857-08>
- Corman VM, Rasche A, Diallo TD, Cottontail VM, Stocker A, Souza BFCF, Correa JI, Carneiro AJB, Franke CR, Nagy M, Metz M, Knornschild M, Kalko EKV, Ghanem SJ, Morales KDS, Salsamendi E, Spinola M, Herrler G, Voigt CC, Tschapka M, Drosten C, Drexler JF (2013) Highly diversified coronaviruses in neotropical bats. *J Gen Virol* 94:1984–1994. <https://doi.org/10.1099/vir.0.054841-0>
- D'Souza MP, Adams E, Altman JD, Birnbaum ME, Boggiano C, Casorati G, Chien YH, Conley A, Eckle SBG, Früh K, Gondré-Lewis T, Hassan N, Huang H, Jayashankar L, Kasmar AG, Kunwar N, Lavelle J, Lewinsohn DM, Moody B, Picker L, Ramachandra L,

- Shastri N, Parham P, McMichael AJ, Yewdell JW (2019) Casting a wider net: Immunosurveillance by nonclassical MHC molecules. *PLoS Pathog* 15(2):e1007567. <https://doi.org/10.1371/journal.ppat.1007567>
- de Groot NG, Heijmans CMC, van der Wiel MKH, Blokhuis JH, Mulder A, Guethlein LA, Doxiadis GGM, Claas FHJ, Parham P, Bontrop RE (2016) Complex MHC class I gene transcription profiles and their functional impact in orangutans. *J Immunol* 196:750–758. <https://doi.org/10.4049/jimmunol.1500820>
- Delpont W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457. <https://doi.org/10.1093/bioinformatics/btq429>
- Dengjel J, Schoor O, Fischer R, Reich M, Kraus M, Muller M, Kreymborg K, Altenberend F, Brandenburg J, Kalbacher H, Brock R, Driessen C, Rammensee HG, Stevanovic S (2005) Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. *Proc Natl Acad Sci* 102:7922–7927. <https://doi.org/10.1073/pnas.0501190102>
- Donato C, Vijaykrishna D (2017) The broad host range and genetic diversity of mammalian and avian Astroviruses. *Viruses* 9:102. <https://doi.org/10.3390/v9050102>
- Doxiadis GGM, Rouweler AJM, de Groot NG, Louwse A, Otting N, Verschoor EJ, Bontrop RE (2006) Extensive sharing of MHC class II alleles between rhesus and cynomolgus macaques. *Immunogenetics* 58:259–268. <https://doi.org/10.1007/s00251-006-0083-8>
- Drews A, Strandh M, Råberg L, Wester Dahl H (2017) Expression and phylogenetic analyses reveal paralogous lineages of putatively classical and non-classical MHC-I genes in three sparrow species (Passer). *BMC Evol Biol* 17(1):152. <https://doi.org/10.1186/s12862-017-0970-7>
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>
- Eirín-López JM, Rebordinos L, Rooney AP, Rozas J (2012) The birth-and-death evolution of multigene families revisited. *Repetitive DNA* 7:170–196. <https://doi.org/10.1159/000337119>
- Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML (2013) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* 4:1111–1119. <https://doi.org/10.1111/2041-210X.12114>
- Fagrouch Z, Sarwari R, Lavergne A, Delaval M, de Thoisy B, Lacoste V, Verschoor EJ (2012) Novel polyomaviruses in south American bats and their relationship to other members of the family Polyomaviridae. *J Gen Virol* 93:2652–2657. <https://doi.org/10.1099/vir.0.044149-0>
- Fenton MB, Simmons NB (2015) Bats: a world of science and mystery. University of Chicago Press
- Fischer Lindahl K, Byers DE, Dabhi VM, Hovik R, Jones EP, Smith GP, Wang CR, Xiao H, Yoshino M (1997) H2-M3, a full-service class Ib histocompatibility antigen. *Annu Rev Immunol* 15(1):851–879. <https://doi.org/10.1146/annurev.immunol.15.1.851>
- Fischer K, Pinho dos Reis V, Balkema-Buschmann A (2017) Bat Astroviruses: towards understanding the transmission dynamics of a neglected virus family. *Viruses* 9:34. <https://doi.org/10.3390/v9020034>
- Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16:573–582. <https://doi.org/10.1093/bioinformatics/16.7.573>
- Gillingham MAF, Bechet A, Courtiol A, Rendón-Martos M, Amat JA, Samraoui OO, Sommer S, Cezilly F (2017) Very high MHC class IIB diversity without spatial differentiation in the Mediterranean population of greater flamingos. *BMC Evol Biol* 17(1):56
- Gillingham MAF, Montero BK, Wilhelm K, Grudzus K, Sommer S, Santos PSC (2019) A novel workflow to improve multi-locus genotyping of wildlife species: an experimental set-up with a known model system. *bioRxiv*:638288. <https://doi.org/10.1101/638288>
- Gordon A, Hannon GJ (2010) Fastx-toolkit: FASTQ/a short-reads pre-processing tools
- Hansen SG, Wu HL, Burwitz BJ, Hughes CM, Hammond KB, Ventura AB, Reed JS, Gilbride RM, Ainslie E, Morrow DW, Ford JC, Selseth AN, Pathak R, Malouli D, Legasse AW, Axthelm MK, Nelson JA, Gillespie GM, Walters LC, Brackenridge S, Sharpe HR, Lopez CA, Fruh K, Korber BT, McMichael AJ, Gnanakaran S, Sacha JB, Picker LJ (2016) Broadly targeted CD8⁺ T cell responses restricted by major histocompatibility complex E. *Science* 351(6274):714–720. <https://doi.org/10.1126/science.aac9475>
- Holmes EC, Worobey M, Rambaut A (1999) Phylogenetic evidence for recombination in dengue virus. *Mol Biol Evol* 16:405–409. <https://doi.org/10.1093/oxfordjournals.molbev.a026121>
- Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, Nishibori M, Inoko H, Hanzawa K, Shiina T (2008) Contribution of mutation, recombination, and gene conversion to chicken Mhc-B haplotype diversity. *J Immunol* 181:3393–3399. <https://doi.org/10.4049/jimmunol.181.5.3393>
- Huchard E, Knapp LA, Wang J et al (2010) MHC, mate choice and heterozygote advantage in a wild social primate. *Mol Ecol* 19:2545–2561. <https://doi.org/10.1111/j.1365-294X.2010.04644.x>
- Hughes AL, Yeager M (1997) Comparative evolutionary rates of introns and exons in murine rodents. *J Mol Evol* 45:125–130. <https://doi.org/10.1007/PL00006211>
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267. <https://doi.org/10.1093/molbev/msj030>
- Kaufman J (2018) Unfinished business: evolution of the MHC and the adaptive immune system of jawed vertebrates. *Annu Rev Immunol* 36:383–409. <https://doi.org/10.1146/annurev-immunol-051116-052450>
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics* 56:683–695. <https://doi.org/10.1007/s00251-004-0717-7>
- Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267:275–276. <https://doi.org/10.1038/267275a0>
- Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press
- Klein J (1986) Natural history of the major histocompatibility complex. Wiley
- Klein J, Kasahara M, Gutknecht J, Figueroa F (1990) Origin and function of Mhc polymorphism. 1939-1989 Fifty Years Prog Allergy 49:35–50. <https://doi.org/10.1159/000418223>
- Klitz W, Hedrick P, Louis EJ (2012) New reservoirs of HLA alleles: pools of rare variants enhance immune defense. *Trends Genet* 28:480–486. <https://doi.org/10.1016/j.tig.2012.06.007>
- Kosakovsky Pond SL, Frost SDW (2005a) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533. <https://doi.org/10.1093/bioinformatics/bti320>
- Kosakovsky Pond SL, Frost SDW (2005b) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22:1208–1222. <https://doi.org/10.1093/molbev/msi105>

- Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679. <https://doi.org/10.1093/bioinformatics/bti079>
- Luis AD, Hayman DTS, O’Shea TJ et al (2013) A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc R Soc B* 280:20122753. <https://doi.org/10.1098/rspb.2012.2753>
- Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
- Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16:562–563. <https://doi.org/10.1093/bioinformatics/16.6.562>
- Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified Bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retrovir* 21:98–102. <https://doi.org/10.1089/aid.2005.21.98>
- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol* 1. <https://doi.org/10.1093/ve/vev003>
- Mayer F, Brunner A (2007) Non-neutral evolution of the major histocompatibility complex class II gene *DRB1* in the sac-winged bat *Saccopteryx bilineata*. *Heredity* 99:257–264. <https://doi.org/10.1038/sj.hdy.6800989>
- Minias P, Pikus E, Whittingham LA, Dunn PO (2018) A global analysis of selection at the avian MHC. *Evolution* 72:1278–1293. <https://doi.org/10.1111/evo.13490>
- Miska KB, Harrison GA, Hellman L, Miller RD (2002) The major histocompatibility complex in monotremes: an analysis of the evolution of Mhc class I genes across all three mammalian subclasses. *Immunogenetics* 54:381–393. <https://doi.org/10.1007/s00251-002-0484-2>
- Moratelli R, Calisher CH (2015) Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses? *Mem Inst Oswaldo Cruz* 110:1–22. <https://doi.org/10.1590/0074-02760150048>
- Murphy K, Weaver C (2017) *Janeway’s immunobiology*. Garland Science
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8:e1002764. <https://doi.org/10.1371/journal.pgen.1002764>
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K (2013) FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Mol Biol Evol* 30:1196–1205. <https://doi.org/10.1093/molbev/mst030>
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci* 94:7799–7806. <https://doi.org/10.1073/pnas.94.15.7799>
- Ng JHJ, Tachedjian M, Deakin J, Wynne JW, Cui J, Haring V, Broz I, Chen H, Belov K, Wang LF, Baker ML (2016) Evolution and comparative analysis of the bat MHC-I region. *Sci Rep* 6:21256. <https://doi.org/10.1038/srep21256>
- Nicholls JA, Double MC, Rowell DM, Magrath RD (2000) The evolution of cooperative and pair breeding in thornbills *Acanthiza* (Pardalotidae). *J Avian Biol* 31:165–176. <https://doi.org/10.1034/j.1600-048X.2000.310208.x>
- Olival KJ, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL, Daszak P (2017) Host and viral traits predict zoonotic spillover from mammals. *Nature* 546:646–650. <https://doi.org/10.1038/nature22975>
- Otting N, Heijmans CMC, Noort RC, de Groot NG, Doxiadis GGM, van Rood JJ, Watkins DI, Bontrop RE (2005) Unparalleled complexity of the MHC class I region in rhesus macaques. *Proc Natl Acad Sci* 102:1626–1631. <https://doi.org/10.1073/pnas.0409084102>
- Papenfuss AT, Baker ML, Feng Z-P, Tachedjian M, Cramer G, Cowled C, Ng J, Janardhana V, Field HE, Wang LF (2012) The immune gene repertoire of an important viral reservoir, the Australian black flying fox. *BMC Genomics* 13:261. <https://doi.org/10.1186/1471-2164-13-261>
- Pavlovich SS, Lovett SP, Koroleva G et al (2018) The Egyptian rousette genome reveals unexpected features of bat antiviral immunity. *Cell* 173:1098–1110.e18. <https://doi.org/10.1016/j.cell.2018.03.070>
- Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci* 98:13757–13762. <https://doi.org/10.1073/pnas.241370698>
- Prugnolle F, Manica A, Charpentier M, Guégan JF, Guemier V, Balloux F (2005) Pathogen-driven selection and worldwide HLA class I diversity. *Curr Biol* 15:1022–1027. <https://doi.org/10.1016/j.cub.2005.04.050>
- Real-Monroy MD, Martínez-Méndez N, Ortega J (2014) MHC-DRB exon 2 diversity of the Jamaican fruit-eating bat (*Artibeus jamaicensis*) from Mexico. *Acta Chiropterol* 16:301–314. <https://doi.org/10.3161/150811014X687260>
- Reusch TB, Schaschl H, Wegner KM (2004) Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback. *Immunogenetics* 56:427–437
- Richman AD, Herrera MLG, Ortega-García S et al (2010) Class II DRB polymorphism and sequence diversity in two vesper bats in the genus *Myotis*. *Int J Immunogenet* 37:401–405. <https://doi.org/10.1111/j.1744-313X.2010.00941.x>
- Robinson J, Malik A, Parham P, Bodmer JG, Marsh SGE (2000) IMGT/HLA database—a sequence database for the human major histocompatibility complex. *Tissue Antigens* 55:280–287
- Rock KL, Reits E, Neefjes J (2016) Present yourself! By MHC class I and MHC class II molecules. *Trends Immunol* 37:724–737. <https://doi.org/10.1016/j.it.2016.08.010>
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics methods and protocols*. Springer, pp 365–386
- Salmier A, de Thoisy B, Crouau-Roy B, Lacoste V, Lavergne A (2016) Spatial pattern of genetic diversity and selection in the MHC class II DRB of three Neotropical bat species. *BMC Evol Biol* 16:229. <https://doi.org/10.1186/s12862-016-0802-1>
- Salmón-Mulanovich G, Vásquez A, Albújar C, Guevara C V, Laguna-Torres A, Salazar M, Zamalloa H, Cáceres M, Gómez-Benavides J, Pacheco V, Contreras C, Kochel T, Niezgodá M, Jackson FR, Velasco-Villa A, Rupprecht C, Montgomery JM (2009) Human rabies and rabies in vampire and nonvampire bat species, southeastern Peru, 2007. *Emerg Infect Dis* 15:1308–1311. <https://doi.org/10.3201/eid1508.081522>
- Santos PSC, Courtiol A, Heidel AJ, Höner OP, Heckmann I, Nagy M, Mayer F, Platzer M, Voigt CC, Sommer S (2016) MHC-dependent mate choice is linked to a trace-amine-associated receptor gene in a mammal. *Sci Rep* 6:38490. <https://doi.org/10.1038/srep38490>
- Santos PSC, Mezger M, Kolar M, Michler FU, Sommer S (2018) The best smellers make the best choosers: mate choice is affected by female chemosensory receptor gene diversity in a mammal. *Proc R Soc B Biol Sci* 285:20182426. <https://doi.org/10.1098/rspb.2018.2426>
- Sawyer SA (1999) GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis. St Louis
- Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I (2019) GenBank. *Nucleic Acids Res* 47:D94–D99. <https://doi.org/10.1093/nar/gky989>
- Schad J, Dechmann DK, Voigt CC, Sommer S (2011) MHC class II DRB diversity, selection pattern and population structure in a neotropical bat species, *Noctilio albiventris*. *Heredity* 107:115–126
- Schad J, Dechmann DKN, Voigt CC, Sommer S (2012a) Evidence for the ‘good genes’ model: association of MHC class II DRB alleles with

- ectoparasitism and reproductive state in the neotropical lesser bulldog bat, *Noctilio albiventris*. PLoS One 7:e37101. <https://doi.org/10.1371/journal.pone.0037101>
- Schad J, Voigt CC, Greiner S, Dechmann DKN, Sommer S (2012b) Independent evolution of functional MHC class II DRB genes in New World bat species. Immunogenetics 64:535–547. <https://doi.org/10.1007/s00251-012-0609-1>
- Schaschl H, Tobler M, Plath M et al (2008) Polymorphic MHC loci in an asexual fish, the amazon molly (*Poecilia formosa*; Poeciliidae). Mol Ecol 17:5220–5230. <https://doi.org/10.1111/j.1365-294X.2008.03997.x>
- Schmid J, Rasche A, Eibner G, Jeworowski L, Page RA, Corman VM, Drost C, Sommer S (2018) Ecological drivers of *Hepacivirus* infection in a Neotropical rodent inhabiting landscapes with various degrees of human environmental change. Oecologia 188(1):289–302
- Schwensow N, Castro-Prieto A, Wachter B, Sommer S (2019) Immunological MHC supertypes and allelic expression: how low is the functional MHC diversity in free-ranging Namibian cheetahs? Conserv Genet 20:65–80. <https://doi.org/10.1007/s10592-019-01143-x>
- Siddle HV, Marzec J, Cheng Y, Jones M, Belov K (2010) MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. Proc R Soc Lond B Biol Sci 277: 2001–2006. <https://doi.org/10.1098/rspb.2009.2362>
- Simmons NB (2005) Order Chiroptera pp. In: Wilson DE, Reeder DM (eds) Mammal species of the world: a taxonomic and geographic reference. Johns Hopkins University Press, Baltimore, pp 312–529
- Smith JM (1992) Analyzing the mosaic structure of genes. J Mol Evol 34: 126–129. <https://doi.org/10.1007/BF00182389>
- Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16. <https://doi.org/10.1186/1742-9994-2-16>
- Sommer S, Courtiol A, Mazzoni CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. BMC Genomics 14: 542. <https://doi.org/10.1186/1471-2164-14-542>
- Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc Lond B 277:979–988. <https://doi.org/10.1098/rspb.2009.2084>
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. <https://doi.org/10.1093/molbev/msr121>
- Teeling EC, Springer MS, Madsen O, Bates P, O'Brien S, Murphy WJ (2005) A molecular phylogeny for bats illuminates biogeography and the fossil record. Science 307:580–584. <https://doi.org/10.1126/science.1105113>
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680. <https://doi.org/10.1093/nar/22.22.4673>
- van der Wiel MKH, Doxiadis GGM, de Groot N, Otting N, de Groot NG, Poirier N, Blancho G, Bontrop RE (2018) MHC class I diversity of olive baboons (*Papio anubis*) unravelled by next-generation sequencing. Immunogenetics 70:439–448. <https://doi.org/10.1007/s00251-018-1053-7>
- Wasimuddin, Brändel SD, Tschapka M et al (2018) Astrovirus infections induce age-dependent dysbiosis in gut microbiomes of bats. ISME J 12:2883–2893. <https://doi.org/10.1038/s41396-018-0239-1>
- Westerdahl H, Waldenström J, Hansson B, Hasselquist D, von Schantz T, Bensch S (2005) Associations between malaria and MHC genes in a migratory songbird. Royal Society of London. Proceedings B. Biol Sci 272(1571):1511–1518. <https://doi.org/10.1098/rspb.2005.3113>
- Wynne JW, Woon AP, Dudek NL, Croft NP, Ng JHJ, Baker ML, Wang L-F, Purcell AW (2016) Characterization of the antigen processing machinery and endogenous peptide presentation of a bat MHC class I molecule. J Immunol 196(11):4468–4476. <https://doi.org/10.4049/jimmunol.1502062>
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118. <https://doi.org/10.1093/molbev/msi097>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.