Major Histocompatibility Complex (MHC) Evolution in Platyrrhine Monkeys: Duplications and Hybridization as Sources of Adaptive Genetic Variation

by

Adrian Melo Carrillo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2022

Doctoral Committee:

Professor Timothy James, Chair
Associate Professor Regina Baucom
Associate Professor Jeffrey Kidd
Professor Priscilla Tucker

Adrian Melo Carrillo

meloco@umich.edu

ORCID iD: 0000-0002-2557-9843

Dedication

A todos aquellos que pueden subir a la nube voladora

Acknowledgements

I would like to thank my main advisor Liliana Cortés-Ortiz. Her support, advice and care were fundamental in me successfully completing not only my PhD program but my master's as well. We have been working together since 2013 and I have experienced an academic and personal growth that would not have been possible without her guidance. It is hard to say goodbye to a lab that felt like home, but it is comforting knowing that we are all moving in a positive direction where great opportunities await… and this is only a see you later.

I would also like to thank Marcella Baiz and Susanna Campbell, both members of the Cortés-Ortiz Lab. Marcella helped me tremendously at the beginning of grad school, and then through at all of it. She was not only very kind but was always willing to help and contribute in any way she could. Susanna was someone I could bounce ideas off, she would have great advice and deep insights. She was someone that would always be there for me, and I will miss our daily conversations. I'm very happy I can call both my friends.

To my committee members Priscilla, Tim, Jeff, and Gina. I could not be happier to have chosen such an amazing group of scientists and kind people that supported me throughout my PhD training. My committee meetings and one-on-one interactions were always full of positive interactions and constructive ideas that led to the successful completion of my program.

Finally, I would like to thank the people that were involved in my teaching experiences. Laura Eidietis, Cindee Giffen and Jo Kurdziel, thank you for your kindness, advice, and fun moments. You were an important part of this journey and have contributed to my growth both

academically and personally. Alison Davis Rabosky and Stephen Smith, thank you for your support inside and outside our teaching commitments. I appreciated your kindness and your mentorship. Thank you all of you for being great educators and caring so much about the students.

Table of Contents

List of Tables

List of Figures

Abstract

Genetic variation is the foundation for a population to adapt to an environment that is constantly changing. Investigating the mechanisms through which genetic variation is acquired and maintained can inform how populations evolve in response to environmental pressures. One of the most variable genomic regions is the Major Histocompatibility Complex (MHC), the hallmark of adaptive immunity. In my dissertation, I explore mechanisms affecting MHC genetic variation with a focus on hybridization as a source of adaptive genetic variation.

In Chapter 1, I characterized three MHC Class II loci (DQA1 exon 2, DQB1 exon 2 and DQB2 exon 2) in four species of howler monkeys (genus *Alouatta*): *A. palliata*, *A. pigra*, *A. caraya* and *A. guariba*. These species had not been characterized for any of these MHC loci. Overall, platyrrhines, monkeys of the Americas, are an understudied primate group for MHC genes. The genotypes derived from this study of these four *Alouatta* species suggested duplications at the MHC-DQA1 and MHC-DQB1 genes. Different lines of evidence (genotypes, genome search, gene prediction and phylogenetic analysis) supported at least one duplication event for DQA1 and two duplication events for DQB1. Additionally, sequences from individuals of two more divergent platyrrhine lineages (*Aotus* and *Callithrix*) indicate these duplications may be present across platyrrhines and constitute trans-specific polymorphisms, in which polymorphic gene lineages persist through speciation events. These analyses support positive and balancing selection maintaining the diversity of MHC-DQ in platyrrhines.

In Chapter 2, I evaluated MHC genetic diversity across the distribution of *A. palliata*. I analyzed populations from Mexico, Costa Rica and Peru at seven MHC loci. I found that populations in Mexico were monomorphic at all seven loci while the other populations exhibited greater genetic diversity. These three populations also had signatures of population structure. My results suggested positive selection acting on DRB and DQB1-L3. The remaining loci appear to be evolving neutrally, mirroring patterns of genetic diversity at neutral markers. The Mexican population showed extremely low levels of genetic diversity at each MHC locus. This reduced functional diversity may have serious implications for the conservation of the Mexican population.

In Chapter 3, I assessed introgression of MHC alleles in the *Alouatta* hybrid zone (*A. palliata* x *A. pigra*). The level of genomic admixture (based on a hybrid index) for each individual analyzed in the hybrid zone has been previously estimated. I used these values to evaluate the level and direction of introgression at each locus. I found asymmetric introgression from *A. palliata* into *A. pigra* at all loci. Moreover, five out of six MHC loci analyzed showed increased introgression compared to neutral expectations, suggesting alleles at MHC loci from one species can confer advantages in a second species following hybridization. This is the first study to provide empirical evidence, in a primate hybrid zone, that supports previous claims of adaptive introgression of MHC genes in humans.

This dissertation investigated MHC genetic variation in platyrrhines and the mechanisms and sources responsible for this variation. I provided evidence for gene duplication and hybridization as important sources of adaptive variation maintained through selection. Also, I implemented MHC loci as markers for evaluating genetic diversity, which are more informative

of adaptive genetic variation than traditionally used neutral markers. This work provides new insight into the evolutionary history of MHC in platyrrhines and primates more generally.

Chapter 1: MHC-DQ Duplications in Platyrrhine Primates Maintained by Selection

Abstract

The Major Histocompatibility Complex (MHC) is a highly polymorphic multigene family with a key role in adaptive immunity. We characterized for the first time three MHC Class II loci (DQA1 exon 2, DQB1, exon 2 and DQB2 exon 2) in four species of howler monkeys (genus *Alouatta*). Using high-throughput sequencing, we genotyped 99 wild howler monkeys, including 57 *A. palliata*, 38 *A. pigra*, 2 *A. caraya* and 2 *A. guariba*. Genotyping results, showing more than two alleles per locus per individuals, suggested one duplication event at DQA1 and two duplication events at DQB1. We implemented different approaches to confirm these duplications. Genome search and gene prediction analyses, in *A. palliata*, supported duplications at DQA1 and DQB1. We also performed genome search analyses on *Callithrix jacchus* and *Aotus nancymaae*, using publicly available MHC-DQ sequences and their respective genomes to determine the presence of these putative duplications in platyrrhine primates. The results also supported these duplications. Additionally, maximum likelihood phylogenetic analyses including all four *Alouatta* species, *C. jacchus* and *A. nancymaae* were consistent with duplications at DQA1 and DQB1. Based on the divergence of the species we investigated, these duplications events occurred at least 20 MYA. Moreover, we found positive selection occurring mostly at antigen binding sites (ABS) in DQA1 and DQB1 based on dN/dS ratios. We propose that positive and balancing selection have maintained these duplications in platyrrhine primates.

Introduction

The Major Histocompatibility Complex (MHC) is a multigene family with a central role in adaptive immunity. This genomic region is characterized by its high gene density and extreme polymorphism across many genes (Edwards and Hedrick 1998). Such level of polymorphism is hypothesized to be a consequence of pathogen-driven balancing selection, in which multiple gene variants are maintained in populations facing a broad spectrum of pathogens that must be recognized to trigger specific immune responses (Doherty and Zinkernagel 1975; Jeffery and Bangham 2000; Bernatchez and Landry 2003; Prugnolle et al. 2005). The MHC is found across multiple species of jawed vertebrates but the genomic organization varies across taxa. In general, the current organization of MHC genes in human and non-human primates is similar, when compared to other vertebrates (Kelley et al. 2005).

MHC genes are traditionally divided by function into three different classes: MHC class I, MHC class II and MHC class III. In particular, MHC class II genes code for cell surface proteins that bind to peptides of pathogenic origin and present them to the appropriate T-cells (Garrigan and Hendrick 2003; Janeway et al. 2004). MHC class II includes the well characterized genes DP, DQ and DR. These genes are further subclassified into alpha (DPA, DQA and DRA) and beta (DPB, DQB and DRB) depending on the part of the heterodimer molecule they code for (Bernatchez and Landry 2003; Reche and Reinherz 2003). MHC class II gene organization and composition appears to be fairly conserved in various groups of primates with some considerations. Catarrhine monkeys, those from Asia and Africa, have lost their DQA2/DQB2 genes along their evolutionary history. They also show extreme variability for DR genes that is not shared with apes or platyrrhines (the monkeys of the Americas) (Bontrop et al. 1999, Doxiadis et al. 2007, Doxiadis et al. 2012). Humans, great apes and platyrrhines are

believed to share the same MHC class II organization, including four DP genes (two alpha and two beta), four DQ genes (two alpha and two beta) and three DR genes (one alpha and two beta) (Heijmans et al. 2020). However, platyrrhines remain a largely understudied taxonomic group for MHC genes in general, with only 24 out of 176 currently recognized species (IUCN 2021) studied for any MHC locus (IPD-MHC data base). Particularly, DQ genes have only been characterized in the families Aotidae (genus *Aotus*) (Gaur et al. 1992, Diaz et al. 2000, Otting et al. 2020), Callithrichidae (genera *Callithrix* and *Saguinus*) (Bidwell et al. 1994, Antunes et al. 1998, Otting et al. 2020) and Cebidae (genera *Cebus* and *Sapajus*) (Buckner et al. 2021). For the Atelidae family no DQ gene has ever been characterized, although one species of howler monkey (*Alouatta pigra*) has been studied at the DRB exon 2 MHC locus (Arguello-Sánchez et al. 2018).

Howler monkeys (*Alouatta*) have the largest distribution of any platyrrhine, ranging from southeast Mexico to northern Argentina (Cortés-Ortiz et al. 2003, Doyle et al. 2021). There are 14 recognized species (Cortés-Ortiz et al. 2015), which can be found in a variety of forest habitats, from pristine forests to highly disturbed landscapes (Milton 1998). Such broad distribution and differences in habitat may influence their exposure to different pathogens, consequently impacting the diversity of MHC genes.

Previously uncharacterized MHC loci of primate species may provide new insights into the evolutionary history of MHC in primates, more generally. In this chapter, we aimed to characterize for the first time DQA1 exon 2, DQB1 exon 2, and DQB2 exon 2 in four *Alouatta* species to the determine the level of intra and interspecific variation in the genus and compare it to that in other platyrrhines to evaluate the evolutionary history of these genes in the group. DQA1 and DQB1 are functionally relevant genes that encode for the antigen binding groove

within the MHC class II molecule, to which proteins from pathogenic origin attach with high specificity and trigger an appropriate immune response. Importantly, exon 2 from DQA1 and DQB1 contains antigen biding sites (ABS) that specifically code for the antigen biding groove (Bartl & Weissman 1994). Our results suggest gene duplications at DQA1 and DQB1 loci for all platyrrhine species. Although gene duplications at DQ loci have been well documented in other mammal species (Xu et al 1994, Yang et al 2005, Bryja et al 2006, Niranjan et al 2010, Ballingal et al. 2017, Pagán et al. 2018) they had never been confirmed in primates (but see Diaz et al. 2000 and Buckner et al. 2021). Additionally, our results suggest that these duplications have been maintained through selective processes.

Methods

Sampling, DNA extraction and amplification

We analyzed samples from 99 wild *Alouatta* individuals from four species, including 57 samples from *A. palliata* from Mexico (37 blood samples), Costa Rica (10 blood samples) and Peru (10 fecal samples), 38 blood samples from *A. pigra* from Mexico (N=32) and Guatemala (N=6), 2 tissue samples from *A. caraya* from Argentina, and 2 tissue samples from *A. guariba* from Argentina (Figure 1-1). Samples from Mexico were collected between 1998 and 2010. Samples from Costa Rica were kindly shared by Dr. K. Glader to the Cortés-Ortiz Lab. Samples from Peru were collected by Sergio Redondo and Fanny Cornejo. Fecal samples were obtained fresh from observable individuals immediately after defecation, collected in 15 ml tubes with 99% ethanol, refrigerated in the field and frozen before being shipped to the US and stored at –80°C

in our lab facilities. Details for the collection of samples from Mexico and Costa Rica can be found in Cortés-Ortiz et al. (2003, 2019) and Kelaita et al. (2011).

We extracted DNA from blood and fecal samples using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA) and the QIAamp DNA Stool Mini Kit (Qiagen, Valencia, CA), respectively. We followed manufacturer's protocol for both DNA extraction kits. We quantified DNA concentration in a Qubit 4 Fluorometer, using the Qubit dsDNA high-sensitivity assay (Invitrogen, Carlsbad, CA). All samples were standardized at a DNA concentration of 5 ng/µl.

Next Generation Sequencing Library Preparation

We followed the 16S protocol from Illumina (Illumina 2013) for next generation sequencing library preparation with modifications. Briefly, we performed a PCR to amplify the product of interest, and a second PCR to attach labeling indexes to the DNA fragments generated in the first PCR. The final products were pooled in equimolar DNA concentrations for sequencing. We performed PCRs using previously published primers targeting partial sequences of three exons: DQA1 exon 2 and DQB1 exon 2 (Gyllensten & Elrich 1988 and Yasunaga et al. 1996). As previously reported (Diaz et al 2000) the set of primers used for DQB1 also co-amplified DQB2 exon 2. We modified the first stage PCR to a 3-phase Touch-Down with an initial denaturing temperature of 95 °C for 5 minutes, followed by two cycles of 94 °C for 30 sec, annealing temperature (55 °C for DQA1 and 60 °C for DQB1 and DQB2) for 30 sec and 72 °C for 30 seconds (phase 1); two cycles of 94 °C for 30 sec, annealing temperature (50 °C for DQA1 and 55 °C for DQB1 and DQB2) for 30 sec and 72 °C for 30 seconds (phase 2); 30 cycles of 94 °C for 30 sec, annealing temperature (48 °C for DQA1 and 52 °C for DQB1 and DQB2)for 30 sec and 72 °C for 30 seconds (phase 3) and a final extension at 72 °C for 10 minutes. PCR total

reaction volume was 20 µl and included 5.5 µl of ultrapure water, 1 µl of forward primer at 10 µM, 1 µl of reverse primer at 10 µM, 10 µl of Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Ipswich, MA) and 2.5 µl of DNA at 5ng/µ. For the second (Index) PCR, our total reaction volume was 25 µl: 4 µl of ultrapure water, 3 µl of Nextera XT Index Primer 1 (Illumina, San Diego, CA), 3 µl of Nextera XT Index Primer 2 (Illumina, San Diego, CA), 12.5 µl of Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Ipswich, MA) and 2.5 µl of product from the first PCR. Index PCR thermocycling conditions were not modified from Illumina protocol. Samples were labeled with unique index combinations for individual identification.

Products of the indexed PCRs were quantified using fluorometric DNA quantification using the Qubit dsDNA high-sensitivity assay, diluted to a concentration of 4nM, and pooled together. We amplified and sequenced all samples in duplicate. Duplicates were amplified in independent PCRs and sequenced on different sequencing runs. We submitted our libraries for 250 bp pair-end MiSeq Nano 500 sequencing (Illumina, San Diego, CA) at The University of Michigan Advanced Genomics Core.

MHC genotyping

We filtered out raw sequences that had a phred score lower than 20 using Trimmomatic (Bolger et al. 2014). We merged pair-end sequences using FLASH (Magoc & Salzberg 2011) keeping default parameters and filtered sequences based on size (retaining fragments>130 bp) using PRINSEQ (Schmieder & Edwards 2011) to eliminate sequences that were not within the range of expected size.

We created an R function to determine preliminary genotypes for each individual at each locus. This function ranks all the unique reads of a particular locus by frequency in each

individual, eliminates all singletons, trims primers, and calculates the relative frequency of each unique read to the total number of reads in the individual. Once we determined individual putative alleles, we identified and eliminated chimeric sequences manually by aligning all sequences of each locus and evaluating for the distribution of the variable sites in the aligned sequences. Chimeric sequences were defined as sequences that combined two parental sequences that were found in high frequency within an individual. We confirmed that all putative alleles that remained after removing chimeras were MHC-DQ sequences comparing sequence identity to other published MHC-DQ sequences of closely related taxa using BLAST.

Following the protocol of Sommer et al. 2013 we determined the final genotypes. Briefly, to call true alleles we i) identified high frequency putative alleles in and individual, ii) corroborated allele presence in the duplicate, and iii) confirm that the frequency of each potential true allele sequence was higher than any known artifact (e.g., singletons, chimeras).

Multi-species genome search

The only genome assembly publicly available for *Alouatta* is a non-annotated assembly for *A. palliata* (GenBank accession number: GCA_004027835.1) We used this public assembly and the *A. palliata* sequences that we recovered from the three DQ loci to a) determine if the DQ sequences we obtained could be found in the genome assembly and if so, b) identify the number of genomic regions (scaffolds) containing MHC-DQ loci. We also searched the *Callithrix jacchus* (GenBank assembly accession: GCA_009663435.2) and *Aotus nancymaae* (GenBank assembly accession: GCA_000952055.2) genome assemblies for matches to MHC-DQ sequences and to identify the number of scaffolds containing MHC-DQ sequences, using published sequences reported for those species (IPD-MHC data base). We performed genome

searches using the blastn function, with default parameters, to located regions with high sequence similarity. Supplementary material Table 1-S6 includes all GenBank accession numbers for each MHC-DQ sequence used in our analyses.

Alouatta palliata gene prediction

As the *A. palliata* genome is not annotated we wanted to identify genes within the scaffolds to which our MHC-DQ sequences were matched. We used the software AUGUSTUS to predict genes on contigs that we identified as containing MHC-DQ sequences from the *A. palliata* genome. This program is a probabilistic Hidden Markov Model that identifies gene structure by detecting protein coding regions in a nucleotide sequence (Stanke & Morgensten 2005). Within a given contig AGUSTUS predicts different genes and outline the gene sequence, including coding and non-coding sequence fragments.

Phylogenetic analysis

To determine the relatedness of MHC sequences across platyrrhine primates we built maximum likelihood phylogenetic trees for each MHC-DQ locus (DQA1, DQB1 and DQB2) using the sequences obtained from the four *Alouatta* species, using all previously published DQA1 and DQB1 sequences from *C. jacchus* and *A. nancymaae* and one human consensus sequence for comparison. Sequences were aligned using the MUSCLE algorithm (Edgar 2004) and alignments were visually confirmed and edited manually when needed. Using jModelTest (Darriba et al. 2012) we determined the nucleotide substitution model to use in the tree building. The likelihood score derived from jModelTest supported Jukes-Cantor as the best nucleotide substitution model for all three loci. We assessed statistical significance using bootstrap analysis (1000 iterations).

Amino acid sequence alignment and conservation

To determine the level of conservation of alleles in each locus across howler monkeys, we used MEGAX (Stecher et al. 2020) to deduce the amino acid sequences for all alleles across all loci and reported all variable sites within each locus. We also aligned these sequences to their human homologue sequence using MUSCLE (Edgar 2004) to infer potential antigen binding sites (ABS) in our sequences. ABS in Human Leukocyte Antigen (HLA) genes, the human version of the MHC genes, have been characterized in previous studies for DQA1 and DQB1 (Otting et al. 1992, Kwok et al. 1996, Reche & Reinherz 2003, Siebold et al. 2003, Cocco et al. 2012, Sarri et al. 2018). We did not assign ABS to DQB2 as this locus does not code for any portion of the antigen binding groove (Garrigan and Hendrick 2003).

We used Jalview (Waterhouse et al. 2009) to graphically represent the amino acid alignments, indicating the degree of conservation at each site based on their physico-chemical properties (Livingstone and Barton 1993) (Supplementary Materials Figure 1-S4).

Selection analysis

Differences in the levels of genetic variation at a particular locus among species can be attributed to neutral or selective forces.We used PAMLX (Xu & Yang 2013) to evaluate if selection is responsible for the genetic variation present in DQA1, DQB1 and DQB2 in platyrrhines (the four *Alouatta* species, *A. nancymaae* and *C. jacchus*). Testing for selection PAMLX requires a phylogenetic tree of the sequences to perform the analysis. Phylogenies were built with PAUP (Swofford 2003) using a likelihood method with a neighbor joining clustering algorithm. We implemented the Jukes-Cantor model of nucleotide substitution as this model was also determined by jModelTest (Darriba et al. 2012) as the best model for this dataset. Within

PAMLX we used the program CodeML including site model M0 (one-ratio) and M2 (positive selection) to infer the dN/dS ratio across the gene phylogeny and at each codon, respectively. We determined sites undergoing positive selection using the Bayes Empirical Bayes (BEB) approach (Yang et al. 2005), which calculates the probability of each site experiencing positive selection while accounting for sampling errors.

Results

Genotyping and allele diversity

We successfully amplified and sequenced the three MHC-DQ loci (DQA1 exon 2, DQB1 exon 2 and DQB2 exon 2) in all species of howler monkeys. Our paired-end sequencing raw data yielded a total of 1,535,165 reads across all individuals (including duplicates). After filtering our data (based on sequence quality and fragment size), we recovered 550,411 sequences that were used for allele identification and genotyping. Table 1-1 shows the maximum and minimum number of reads per individual and for all individuals during each step of the filtering process.

Across the four *Alouatta* species, DQB1 was the locus that contained the greatest number of alleles while DQB2 contained the lowest number of alleles. We identified a total of 14, 26, and 8 unique alleles across the four *Alouatta* species for DQA1, DQB1 and DQB2, respectively (Table 1-2A). We were not able to confidently recover alleles for all loci for each individual of *A. palliata* from Peru, given the degraded nature of the DNA (extracted from fecal samples).

All DQA1 alleles had a length of 186 bp and were in the correct reading frame with no indication of pseudogenization. DQB1 alleles had a length of 247bp and 250bp, consistent with the human consensus allele, and also had no signatures of pseudogenization. Five DQB1 alleles

had a total length of 247 bp, presenting a one codon deletion at codon 55 when compared to the 83 codons of the human allele sequence, but they still were in the correct reading frame. All four *Alouatta* species had at least one allele of this size. The remaining 21 DQB1 alleles had a length of 250bp. For DQB2, all alleles had a length of 247 bp, but the two alleles found in *A. caraya*, (alleles 5 and 6), had an early stop codon in the sequence, suggesting pseudogenization. All other DQB2 alleles from the remaining howler monkey species had no indication of being pseudogenized.

We observed that, on average, all four *Alouatta* species have more than two alleles per locus for DQA1 and DQB1, but not for DQB2 (Table 1-3 and Table 1-S7). We use the following nomenclature for the different putative loci. First, we used the gene name, followed by a dash and "L#", where the # indicates the locus number, then another dash and the allele number. For example, allele number one from the putative first locus of DQA1 is named DQA1-L1-1. For DQB2 the "-L#" portion is omitted as there was no indication of multiple loci. This nomenclature is used in all analyses throughout this chapter.

We found shared alleles among *Alouatta* species for all loci. For DQA1, DQA1-L2-2 was shared between *A. palliata* and *A. caraya*, DQA1-L1-3 and DQA1-L1-5 were shared between *A. palliata* and *A. pigra*, and allele DQA1-L1-9 was shared between *A. caraya* and *A. guariba.* For DQB1, *A. palliata* and *A. pigra* share alleles DQB1-L1-1 and DQB1-L2-4, *A. pigra* and *A. caraya* share allele DQB1-L3-8 and *A. pigra* and *A. guariba* share allele DQB1-L1-9. For DQB2, *A. palliata* from Costa Rica and Peru and *A. pigra* share DQB2-2, but this allele was not found in *A. palliata* individuals from Mexico (Table 1-2B).

Multiple genome locations for DQA1 and DQB1 in the Alouatta palliata genome assembly

Based on the sequences we generated, we were able to identify five different scaffolds on the *A. palliata* assembly containing MHC-DQ sequences (Figure 1-1). For DQA1, we located two scaffolds (AloPal_scaffold_8772 and AloPal_scaffold_50203) to which our allele sequences matched. DQA1-L1-1 and DQA1-L2-2 were perfect matches for sequences within scaffolds 8772 and 50203, respectively. DQA1-L1-3 and DQA1-L1-5 also matched sequences in the 8772 scaffold and DQA1-L2-4 matched sequences in scaffold 50203, although these were not identical sequences. The multiple alleles that matched to scaffolds 8772 and 50203 aligned to the same position within their respective scaffold, indicating that they are likely alleles on two distinct loci.

DQB1 sequences corresponded with sequences in three scaffolds (AloPal_scaffold_89817, AloPal_scaffold_61825 and AloPal_scaffold_64423). DQB1-L1-1, DQB1-L2-2 and DQB1-L3-3 were identical to sequences within scaffolds 89817, 64423 and 61825, respectively. DQB1-L2-4 also matched scaffold 64423 and the remaining three alleles matched to scaffold 61825, although these sequences were not identical to those in the genome assembly. All alleles found in scaffolds 61825 and 64423 were found in the same position within each scaffold.

DQB2 sequences matched sequences within theAloPal_scaffold_8772. DQB2-1 was a perfect match to the sequence within this scaffold, while allele DQB2-2 also aligned to the same position but had some nucleotide differences compared to the published sequence. Information on the exact location of each locus on the *A. palliata* reference genome assembly and the characteristics of our blast analyses are summarized in Supplementary Materials (Table 1-S8).

Multiple genome locations also found for DQA1 and DQB1 in Callithrix jacchus and Aotus nancymaae

Both *C. jacchus* and *A. nancymaae* are platyrrhine species that have been investigated at different MHC genes and have publicly available genomes. *C. jacchus* has a significantly better-quality genome assembly (N_{50}= 98,198,953) when compared to *A. nancymaae* (N_{50}= 126,456) and *A. palliata* (N_{50}= 51,304). This results in longer continuous DNA sequence scaffolds. We identified a single scaffold (length 174,041,770 bp) containing MHC-DQ genes based on the published *C. jacchus* DQA1, DQB1 and DQB2 allele sequences (Figure 1-1). A total of eight previously published DQA1 allele sequences matched to two locations in the scaffold. For DQB1, eight sequences were also found in two locations of the scaffold. Caja-DQB1*23:03 was found as a perfect match. For DQB2 the two alleles aligned to the same location in the scaffold. We identified a total of five different locations in the *C. jacchus* genome assembly as containing MHC-DQ sequences (Supplementary Material, Table 1-S9).

In the *A. nancymaae* genome assembly we found three scaffolds where published MHC-DQ allele sequences aligned (Figure 1-1). A total of five DQA1 alleles matched to two scaffolds. For DQB1, 14 alleles matched to three locations of the genome. For DQB2, sequences of the two publicly available alleles aligned to one genome location. We identified a total of six different locations in the *A. nancymaae* genome assembly as containing MHC-DQ sequences (Supplementary Material, Table 1-S10).

MHC-DQ genes predicted within the A. palliata genome

We ran AUGUSTUS to predict genes using all the genome scaffolds identified as containing MHC-DQ sequences. Within each scaffold one to three genes were predicted and characterized

in terms of sequence, size, position within scaffold and coding regions (Table 1-4). For the *Alouatta palliata* genome scaffold 8772, AUGUSTUS predicted three genes. The gene coding sequences of the first gene contained the sequence of allele DQA1-L1-1. The gene coding sequence of the second gene contained allele DQB2-1. The third predicted gene did not contain any of our recovered sequences, but it appears to be a DQA gene based on sequence similarity.

For scaffold 50203 of the *Alouatta palliata* genome assembly, AUGUSTUS predicted two genes. The gene coding sequence of the first gene contained the sequence of allele DQA1-L2-2, while the second gene did not contain any of our previously identified sequences and the gene sequence identity did not match any MHC-DQ sequence publicly available on GenBank. For scaffold 89817, only one gene was predicted and this gene contained the sequence of allele DQB1-L1-1 in the gene coding sequence. For scaffold 61825 only one gene was predicted and the gene coding sequence contained the sequence of allele DQB1-L3-3. Lastly, one gene was predicted for scaffold 64423 and the gene coding sequence contained the sequence of allele DQB1-L2-2.

Differences in phylogenetic relationships across DQA1, DQB1 and DQB2 alleles

We built a phylogenetic gene tree for each locus using all platyrrhine sequences available (Figure 1-2). The topology of the DQA1 tree (Figure 1-2A) did not match the expected phylogenetic relationships among the taxa, but instead alleles clustered in a manner consistent with the putative loci identified with genotype and genome location analyses (see above). The DQA1 gene phylogeny showed four well-supported clades. One clade (Figure 1-2A, Clade A) containing sequences of all *Alouatta* species except for *A. palliata*. A second clade (Figure 1-2A, Clade B) that was formed by sequences from all six species. This group of sequences is further

subdivided in three groups containing *C. jacchus* sequences, *A. nancymaae* sequences and all four *Alouatta* species sequences, respectively. However, there is only strong support for the *C. jacchus* subgroup. The other two well supported clades are the ones containing the more divergent Caja-DQA1*28 alleles.

The DQB1 phylogeny contains two well-supported clades (Figure 1-2B). Clade A includes sequences from all six species. Clade B also contains sequences from all six species but it also has subgroups with significant statistical support. Subgroup C is comprised exclusively of *A. pigra* sequences. Subgroup D contains sequences of all four *Alouatta* species that presented the codon deletion at position 55. Lastly, subgroup E is a monophyletic group of *C. jacchus* sequences.

In the DQB2 gene tree (Figure 1-2C) all sequences follow the expected phylogenetic relationships of the three genera, with *Alouatta, Aotus* and *C. jacchus* alleles forming reciprocally monophyletic clades (clades A, B, and C, respectively).

Overall signatures of positive selection in DQB1 and in specific antigen binding sites in DQA1 and DQB1

Our PAML analyses detected signatures of positive selection in DQB1 across the entire sequence (dN/dS ratio ω= 1.26). The dN/dS ratios for DQA1 (ω=0.93) and DQB2 (ω=0.63) are indicative of neutral and purifying selection, respectively. We also tested for positive selection at individual amino acid sites in the three loci. In DQA1 six out of 62 aminoacid sites showed signatures of positive selection, from which five are antigen binding sites (ABS). In DQB1, 12 out of 80 sites were under positive selection, and nine of these are also ABS. Lastly from DQB2 six sites out of 79 showed positive selection (Table 1-5).

Discussion

We genotyped for the first time MHC-DQ loci in four *Alouatta* species. Individual genotypes of these species suggested duplications at MHC-DQA1 and MHC-DQB1 genes. Different lines of evidence (genotypes, genome search, gene prediction and phylogenetic analyses) supported at least one duplication event for DQA1 and two duplication events for DQB1 (Figure 1-3). Additionally, sequences from individuals of two more divergent platyrrhine lineages (*Aotus* and *Callithrix*) indicate these duplications may be present across platyrrhines and constitute a case of trans-specific polymorphisms, in which polymorphic gene lineages persist through speciation events. Our analyses support different forms of selection maintaining the diversity of MHC-DQ in platyrrhines.

Multiple lines of evidence support DQA1 and DQB1 duplications in Alouatta palliata

Multiple lines of evidence (genotypes, genome search, gene prediction and phylogenetic analysis) supported at least one duplication event for DQA1 resulting in two different loci and at least two duplication events in DQB1 resulting in three loci, while no evidence suggested a duplication at DQB2. The first two strong indicators of gene duplication came from our genotyping data. First, the average number of alleles per individual at both loci, in all four *Alouatta* species, was greater than two. Second, all *A. palliata* individuals from Mexico (N=38) had the same two alleles at DQA1, suggesting that, in the absence of a duplication event, all 38 individuals would be heterozygous at this locus. This is extremely unlikely, because based on inheritance laws of diploid individuals we would expect some proportion of these individuals to be homozygous for either allele. The genome search results supported the duplications at both loci since one allele from each putative locus aligned with a perfect match to the *A. palliata*

genome assembly. Moreover, these putative loci were also supported by phylogenetic analyses although we presented these results in a multi-species context.

Intrinsic characteristics of our sampling and the genome assembly provided additional support to our findings. The *A. palliata* genome assembly is based on an individual from a Mexican population (Zoonomia Consortium 2020) and *A. palliata* populations in Mexico have extremely low genetic diversity compared to other populations that occur at lower latitudes (Melo-Carrillo et al. 2020). The genome assembly has one of the highest levels of homozygosity among mammalian species studied to date (Zoonomia Consortium 2020). In fact, all MHC loci studied here are monomorphic in all *A. palliata* individuals from Mexico, but no in individual from Costa Rica of Peru. The low levels of genetic diversity increased the probability that the MHC-DQ alleles found in the Mexican individuals would also be found in the genome assembly, as it was the case. It is worth mentioning that *A. pigra* shared DQA1 and DQB2 alleles with *A. palliata*, but these alleles were only present *A. palliata* individuals from Costa Rica and Peru. This indicates that some ancestral variation has been maintained in both species but has been lost in the Mexican populations of *A. palliata*, likely due to recent bottleneck events (Melo-Carrillo et al. 2020).

Furthermore, the seven predicted genes were located in different parts of the genome assembly (i.e., different scaffolds) and consequently were different in terms of their sequence identity. From these seven predicted genes, six contained perfect matches of the six *A. palliata* alleles that corresponded to different loci. Importantly, these alleles were found in the coding region of the predicted genes as these were segments of the exons 2 of MHC-DQ genes. Gene prediction findings are concordant with the genome search results in terms of number of duplicated loci in DQA1 and DQB1. We were able two predict two DQA1 genes that contained

one allele from each DQA1 allele group. We also predicted three distinct DQB1 genes containing one allele from each DQB1 allele group.

The new loci designation (two DQA1 loci and three DQB1 loci) supported by the different lines of evidence was also concordant with the initial individual genotypes. Based on this new-loci designation individuals no longer rendered more than two alleles per locus and individuals from Mexico (N=38) were homozygous at all five loci. The high level of homozygosity observed across the genome in this taxon (Melo-Carrillo et al. 2019; Zoonomia Consortium 2020) is consistent with this lack of variation in MHC.

MHC-DQ duplications extend to other platyrrhines

Although genome assemblies are not available for other *Alouatta* species besides *A. palliata*, there are publicly available genomes for the more distantly related *A. nancymaae* and *C. jacchus*. Results from phylogenetic analyses of all publicly available MHC-DQ alleles in these two species and genome searches of these alleles in their respective published genomes, also supported duplications at DQA1 and DQB1.

The DQA1 phylogenetic tree provides support for two different loci in *C. jacchus* and in the *Alouatta* species contained in clade A. The Caja-DQA1*28 alleles are well-supported in two distinct clades, a strongly supported subgroup within clade B that is monophyletic to *C. jacchus* containing all the Caja-DQA1*29 alleles. Also, Caja-DQA1*27:02 is nested within clade B. Additionally, genome search results in *C. jacchus* also support at least two loci for DQA1. The three *Alouatta* species found in clade A (*A. pigra*, *A. caraya* and *A. guariba*) have also sequences in the well differentiated clade B, which could be indicative of a duplication. These two clades indicating two loci are concordant with the genotypes in all three species as no individuals

present more than two alleles per locus (Supplementary Material Table, 1-S11). The support for a DQA1 duplication is not as strong for *A. palliata* based on the phylogenetic tree. The subgroups containing DQA-L1-alleles and DQA1-L2 alleles show a moderate statistical strength. We do not observe any indication of duplication based on the tree topology for *A. nancymaae*. Yet, DQA1 sequences aligned to two locations within the genome.

Loci designation in the DQB1 based on *A. palliata* analyses was concordant with the clustering of the different alleles in the phylogenetic analyses. We observed three clades corresponding to putative loci for the *Alouatta* species: clade A, clade B and subgroup D. Unlike DQA1, we did find phylogenetic evidence of gene duplication from *A. nancymaae*. Besides the 22*01 alleles in clade A, other *A. nancymaae* sequences are found in clade B. Genome search results for this species also revealed three distinct genome locations for DQB1 genes. *C. jacchus* sequences in the phylogeny are found in two clades (A and B). However, genome search analyses found three distinct locations for DQB1 sequences. There is evidence for a fourth locus in *A. pigra* based on subgroup C and the overall allele diversity of DQB1 in *A. pigra*. We were unable to fully reconcile the genotyping with precise number of loci for DQB1 in *A. pigra*, particularly for DQB1-L3. We believe there is another duplication in this species but we need more evidence to confidently support this claim.

The *A. palliata* results served as the foundation to explore the duplication events in DQA1 and DQB1 in the context of the *Alouatta* genus and more distantly related platyrrhines. We found support for two loci in DQA1 and three loci in DQBI across all six species. Some results suggested even more duplication events at DQB1 but we do not have the same depth of evidence as for *A. palliata*, in which these extra duplications were not detected, to make such

claim. However, phylogenetic analyses and genome searches from the different species agree with the loci designation that derived from *A. palliata* findings.

Trans-species polymorphism and balancing selection in MHC-DQ

We wanted to evaluate the role of selection in maintaining the observed diversity in MHC-DQ genes. We were particularly interested in DQA1 and DQB1 not only because of the duplication events but also because they encode the alpha-1 and beta-1 MHC molecule subunits involved in antigen recognition.

Phylogenetic trees, in both DQA1 and DQB1, presented allele grouping that did not correspond to the species trees. These different groups of sequences that are more closely related to other sequences found in different species rather than sequences found in the same species are indicative of ancestral polymorphic gene lineages that were passed from their common ancestors to modern species, known as trans-species polymorphisms (Klein 1987), and remained in each species likely due to selective pressures. Constant and strong selection is necessary for genetic polymorphisms to persist throughout speciation events (Nei & Rooney 2005, Charlesworth 2006). It is not common to find an overall ω (dN/dS) > 1, indicating positive selection, across an entire sequence as not all sites experience selection equally (Hughes & Nei 1988). In the case of DQB1 we did find an overall ω>1 and our analyses showed that 12 amino acid sites underwent positive selection in this locus. In DQA1 we did not find an overall ω>1, but six amino acid sites were under positive selection. Importantly, a great proportion of these amino acid sites under positive selection (83% for DQA1 and 75% for DQB1) were in the ABS. The maintenance of polymorphic gene lineages through selection is how balancing selection is defined (Charlesworth 2006, Key et al 2014). MHC genes have been well documented to be under balancing selection

given their extreme levels of polymorphism that persist through speciation events and their role in adaptive immunity (Azevedo et al. 2015, Edjsmon & Radwan 2015, Teixeira et al 2015). Our evidence supports the notion that balancing selection is maintaining the observed diversity in MHC-DQ genes in platyrrhines as polymorphic gene lineages remain present in long-diverged species. Other studies in mammals (Bryja et al. 2006, Ballingall et al. 2018, Pagan et al. 2018) have reported similar findings in which balancing selection appears to be maintaining duplicated loci beyond speciation events. We believe similar and numerous pathogens interacting with distantly related species in the neotropics are likely to be imposing the selective pressures behind this genetic variation.

Concluding remarks

We present for the first-time robust evidence for MHC-DQ duplications in platyrrhine primates that have likely been maintained through balancing selection. Based on the taxa that were included in this study MHC-DQ duplications arose at least 20 MYA (Perelman et al. 2011). More distantly related platyrrhine species need to be characterized at MHC-DQ loci to have a better time estimate. Future studies could benefit from characterizing longer MHC sequences (e.g., complete genes) as this may provide greater inference regarding functionality, a key aspect commonly missing in MHC studies in non-model organisms. In closing, this study provides novel insight into the evolutionary history of MHC in platyrrhine primates.

References

Antunes SG, De Groot NG, Brok H *et al.* (1998) The common marmoset: A new world primate species with limited Mhc class II variability. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 11745–11750.

Arguello-Sánchez LE, Arguello JR, García-Feria LM *et al.* (2018) MHC class II DRB variability in wild black howler monkeys (*Alouatta pigra*), an endangered new world primate. *Animal Biodiversity and Conservation*, **41**, 389–404.

Azevedo L, Serrano C, Amorim A, Cooper DN (2015) Trans-species polymorphism in humans and the great apes is generally maintained by balancing selection that modulates the host immune response. *Human Genomics*, 4–9.

Ballingall KT, Lantier I, Todd H, Lantier F, Rocchi M (2017) Structural and functional diversity arising from intra- and inter-haplotype combinations of duplicated DQA and B loci within the ovine MHC. *Immunogenetics 2017 70:4*, **70**, 257–269.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363–377.

Bidwell JL, Lu P, Wang Y *et al.* (1994) DRB, DQA, DQB AND DPB NUCLEOTIDE SEQUENCES OF SANGUINUS OEDIPUS B95-8. *International Journal of Immunogenetics*, **21**, 67–77.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bontrop RE, Otting N, de Groot NG, Doxiadis GGM (1999) Major histocompatibility complex class II polymorphisms in primates. *Immunological Reviews*, **167**, 339–350.

Bryja J, Galan M, Charbonnel N, Cosson JF (2006) Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). *Immunogenetics 2006 58:2*, **58**, 191–202.

Buckner JC, Jack KM, Melin AD *et al.* (2021) Major histocompatibility complex class II DR and DQ evolution and variation in wild capuchin monkey species (Cebinae). *PLOS ONE*, **16**, e0254604.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, **2**, 379–384.

Cocco E, Sardu C, Pieroni E *et al.* (2012) HLA-DRB1-DQB1 haplotypes confer susceptibility and resistance to multiple sclerosis in sardinia. *PLoS ONE*, **7**.

Cortés-Ortiz L, Bermingham E, Rico C *et al.* (2003) Molecular systematics and biogeography of the genus, *Alouatta*. *Molecular Phylogenetics and Evolution*, **26**, 64–81.

Cortés-Ortiz L, Roos C, Zinner D (2019) Introduction to Special Issue on Primate Hybridization and Hybrid Zones. *International Journal of Primatology 2019 40:1*, **40**, 1–8.

Cortés-Ortiz L, Rylands AB, Mittermeier RA (2015) The Taxonomy of Howler Monkeys: Integrating Old and New Knowledge from Morphological and Genetic Studies. In: *Howler Monkeys: Adaptive Radiation, Systematics, and Morphology* (eds Kowalewski MM, Garber PA, Cortés-Ortiz L, Urbani B, Youlatos D), pp. 55–84. Springer US, New York.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods 2012 9:8*, **9**, 772–772.

Diaz D, Naegeli M, Rodriguez R *et al.* (2000) Sequence and diversity of MHC DQA and DQB genes of the owl monkey *Aotus nancymaae*. *Immunogenetics*, **51**, 528–537.

Doherty P, Zinkernagel R (1975) Enhanced immunological surveillance in mice heterozygous at

the H-2 gene complex. *Nature*.
Doxiadis GGM, De Groot N, Claas FHJ *et al.* (2007) A highly divergent microsatellite facilitating fast and accurate DRB haplotyping in humans and rhesus macaques.
Doxiadis GGM, Hoof I, De Groot N, Bontrop RE (2012) Evolution of HLA-DRB genes. *Molecular Biology and Evolution*, **29**, 3843–3853.
Doyle ED, Prates I, Sampaio I *et al.* (2020) Molecular phylogenetic inference of the howler monkey radiation (Primates: *Alouatta*). *Primates 2020 62:1*, **62**, 177–188.
Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
Edwards S V, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *TREE*, **13**, 305–311.
Ejsmond MJ, Radwan J (2015) Red Queen Processes Drive Positive Selection on Major Histocompatibility Complex (MHC) Genes. *PLoS Computational Biology*, **11**, 1–14.
Gaur LK, Heise ER, Thurtle PS, Nepom GT (1992) Conservation of the HLA-DQB2 locus in nonhuman primates. *The Journal of Immunology*, **148**, 943–948.
Genereux DP, Serres A, Armstrong J *et al.* (2020) A comparative genomics multitool for scientific discovery and conservation. *Nature*, **587**, 240–245.
Garrigan, D. & Hedrick, P.W. 2003. Perspective: detecting adaptive molecular polymorphism, lessons from the MHC. Evolution 57: 1707–1722.
Gyllensten UB, Erlich H a (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 7652–6.
Heijmans CMC, de Groot NG, Bontrop RE (2020) Comparative genetics of the major histocompatibility complex in humans and nonhuman primates. *International Journal of Immunogenetics*, **47**, 243–260.
Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335**, 167–170.
Illumina (2013) *16S Metagenomic Sequencing Library Preparation*. https://web.uri.edu/gsc/files/16s-metagenomic-library-prep-guide-15044223-b.pdf
IPD-MHC Database. https://www.ebi.ac.uk/ipd/mhc/
IUCN/SSC Primate Specialist Group http://www.primate-sg.org/
Janeway, C.A., Travers, P., Walport, D. & Shlomchik, M.J. 2004. Immunobiology: The Immune System in Health and Disease. Garland Publishing, New York, NY
Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? *Microbes and Infection*, **2**, 1335–1341.
Kelaita M, Dias PAD, Aguilar-Cucurachi MS, Canales-Espinosa D, Cortés-Ortiz L (2011) Impact of Intrasexual Selection on Sexual Dimorphism and Testes Size in the Mexican Howler Monkeys *Alouatta palliata* and A . *pigra*. *American Journal of Physical Anthropology*, **187**, 179–187.
Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics*, **56**, 683–695.
Key FM, Teixeira JC, de Filippo C, Andrés AM (2014) Advantageous diversity maintained by balancing selection in humans. *Current Opinion in Genetics and Development*, **29**, 45–51.
Klein J (1987) Origin of major histocompatibility complex polymorphism: The trans-species

hypothesis. *Human Immunology*, **19**, 155–162.
Klein J, Bontrop RE, Dawkins RL *et al.* (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics*, **31**, 217–219.
Kwok BW, Domeier ME, Johnson ML, Nepom GT, Koelle DM (1996) *HLA-DQB1 Codon 57 Is Critical for Peptide Binding and Recognition*.
Livingstone CD, Barton GJ (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Bioinformatics*, **9**, 745–756.
Magoc T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.
Melo-Carrillo A, Dunn JC, Cortés-Ortiz L (2020) Low genetic diversity and limited genetic structure across the range of the critically endangered Mexican howler monkey (*Alouatta palliata* mexicana). *American Journal of Primatology*, **82**, 1–1.
Milton K (1998) Physiological Ecology of Howlers (*Alouatta*): Energetic and Digestive Considerations and Comparison with the Colobinae. *International Journal of Primatology 1998 19:3*, **19**, 513–548.
Nei M, Rooney A (2005) NIH Public Access. *Annual review of genetics*, **39**, 121–152.
Niranjan SK, Deb SM, Kumar S *et al.* (2010) Allelic diversity at MHC class II DQ loci in buffalo (Bubalus bubalis): Evidence for duplication. *Veterinary Immunology and Immunopathology*, **138**, 206–212.
Otting N, de Groot NG, Bontrop RE (2020) Full-length MHC class II alleles in three New World monkey species. *Hla*, **95**, 163–165.
Otting BN, Kenter M, Van Weeren P, Jonker M (1992) Mhc-DQB repertoire variation in hominoid and Old World primate species. *The Journal of Immunology*, **149**, 461–470.
Pagán HJT, Ferrer T, O'Corry-Crowe G (2018) Positive selection in coding regions and motif duplication in regulatory regions of bottlenose dolphin MHC class II genes. *PLOS ONE*, **13**, e0203450.
Perelman P, Johnson WE, Roos C *et al.* (2011) A molecular phylogeny of living primates. *PLoS Genetics*, **7**, 1–17.
Reche PA, Reinherz EL (2003) Sequence Variability Analysis of Human Class I and Class II MHC Molecules: Functional and Structural Correlates of Amino Acid Polymorphisms. *Journal of Molecular Biology*, **331**, 623–641.
Sarri CA, Papadopoulos GE, Papa A *et al.* (2018) Amino acid signatures in the HLA class II peptide-binding region associated with protection/susceptibility to the severe West Nile Virus disease. *PLoS ONE*, **13**, 1–20.
Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**, 863–864.
Siebold C, Hansen BE, Wyer JR *et al.* (2003) *Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy*.
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.
Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, **33**, W465–W467.
Stecher G, Tamura K, Kumar S (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS (C Russo, Ed,). *Molecular Biology and Evolution*, **37**, 1237–1239.
Swofford, D.L. (2003) PAUP*. Phylogenetic analysis using parsimony (* and other methods).

Version 4. Sinauer Associates, Sunderland.
Teixeira JC, De Filippo C, Weihmann A *et al.* (2015) Long-term balancing selection in LAD1 maintains a missense trans-species polymorphism in humans, chimpanzees, and bonobos. *Molecular Biology and Evolution*, **32**, 1186–1196.
Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
Xu A, Park C, Lewin HA (1994) Both DQB genes are expressed in BoLA haplotypes carrying a duplicated DQ region. *Immunogenetics 1994 39:5*, **39**, 316–321.
Xu B, Yang Z (2013) pamlX: A Graphical User Interface for PAML. *Molecular Biology and Evolution*, **30**, 2723–2724.
Yang G, Yan J, Zhou K, Wei F (2005) Sequence Variation and Gene Duplication at MHC DQB Loci of Baiji (Lipotes vexillifer), a Chinese River Dolphin. *Journal of Heredity*, **96**, 310–317.
Yasunaga S, Kimura A, Hamaguchi K, Rønningen KS, Sasazuki T (1996) Different contribution of HLA-DR and -DQ genes in susceptibility and resistance to insulin-dependent diabetes mellitus (IDDM). *Tissue Antigens*, **47**, 37–48.

Tables

Table 1-1. Number of reads after each filtering step for all Alouatta samples, including the minimum (Min.) and maximum (Max.) number of reads per sample, the mean number of reads across all samples, and the total number of reads in all samples. The total number of samples is 198 as we had duplicates for each individual (N=99 individuals).

Filtering step	Min	Max	Mean	Total
Initial number of reads	3,745	46,737	7,486	1,535,165
Quality filter (phred >20)	2,287	45,378	5,602	1,148,810
Merging sequences	2,190	43,610	5,346	1,096,312
Fragment size (>130 bp)	768	29,800	2,684	550,411

Table 1-2. Allele diversity and shared alleles. A) Number of individuals (*N*) per species and number of alleles found in each species per locus. The bottom row shows the total number of unique alleles across all species per locus. B) Pair-wise species comparison showing the number of alleles that are shared in the following order DQA1\DQB1\DQB2.

A)

Species	*N*	DQA1	DQB1	DQB2
A. palliata	57	5	7	2
A. pigra	38	5	13	3
A. caraya	2	4	6	2
A. guariba	2	4	4	2
	Total unique	14	26	8

B)

	A. palliata	***A. pigra***	***A. caraya***	***A. guariba***
A. palliata				
A. pigra	2\2\1			
A. caraya	1\0\0	0\1\0		
A. guariba	0\0\0	0\1\0	1\0\0	

Table 1-3. Average number of alleles per individual per locus in four Alouatta species.

		Average allele number per locus		
	N	**DQA1**	**DQB1**	**DQB2**
A. palliata	57	2.1	3.1	1.2
A. pigra	38	2.8	4.7	1.6
A. caraya	2	3	4	2
A. guariba	2	2.5	3.5	1

Table 1-4. Numbers of genes predicted for the four contigs that were identified as containing MHC-DQ genes in the *A. palliata* genome assembly. We include *length of the scaffold* in base pairs. *Gene* refers to a particular gene that was predicted and *start* and *end positions* indicated the spanning of each predicted gene within the scaffold. The last column contains the allele that was found in each of the predicted genes. If none of the identified MHC-DQ alleles in *A. palliata* were matched to any gene, we indicated it as NA.

Scaffold ID	Scaffold length	Gene	Start position	End position	Allele found
AloPal_scaffold_8772	128650	1	10015	16204	DQA1-L1-1
AloPal_scaffold_8772	128650	2	25346	60414	DQB2-1
AloPal_scaffold_8772	128650	3	72533	76878	NA
AloPal_scaffold_50203	30189	1	3657	9043	DQA1-L2-2
AloPal_scaffold_50203	30189	2	22713	25046	NA
AloPal_scaffold_89817	9103	1	1022	7250	DQB1-L1-1
AloPal_scaffold_61825	21639	1	9239	15903	DQB1-L3-3
AloPal_scaffold_64423	20065	1	9846	12740	DQB1-L2-2

Table 1-5. Selection in DQA1, DQB1 and DQB2. Tests of selection in the entire gene sequences of each MHC-DQ locus (DQA1, DQB1 and DQB2**)** analyzed in this study (ω = dN/dS), as well as particular amino acid sites under positive selection within each locus and their posterior probability, based on analyses BEB approach on PAML Antigen binding sites (ABS) are indicated with an asterisks and the proportion of sites under positive selection that are ABS is given at the bottom.

	DQA1		**DQB1**		**DQB2**	
	ω = 0.93		ω = 1.26		ω = 0.63	
	Site	Probability	Site	Probability	Site	Probability
	22*	0.715	13*	0.982	32	0.985
	38	0.998	26*	1	56	0.946
	52*	1	29	0.999	59	0.973
	53*	0.717	37*	0.997	62	0.864
	65*	0.998	56*	0.724	70	0.584
	66*	1	57*	1	83	0.55
			61*	0.979		
			70	1		
			71*	0.8		
			75	0.999		
			85*	1		
			90*	0.529		
ABS proportion	**0.83**		**0.75**		**N/A**	

Figures

Figure 1-1. Putative loci genome position in A) C. jacchus, B) A. nancymaae and C) A. palliata. Loci are positioned in the same order they are found in the genome. All loci are found in different genome locations. If the loci are joined by a black line this indicates they are found on the same scaffold, scaffold identification number is below the locus name. Below each locus is a list of the different alleles that aligned to that genome position, if an allele is marked with an asterisk that allele, is a perfect match.

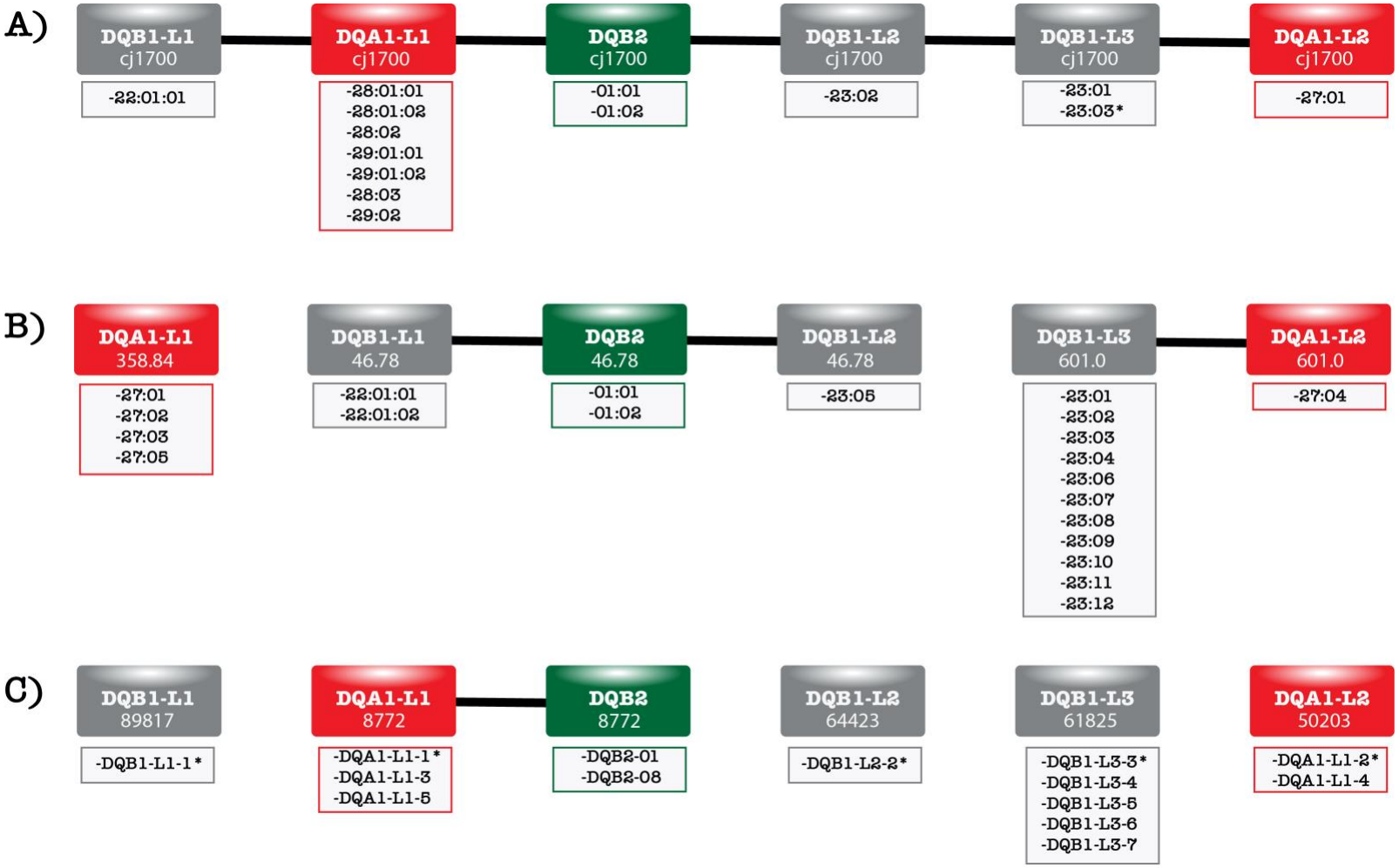
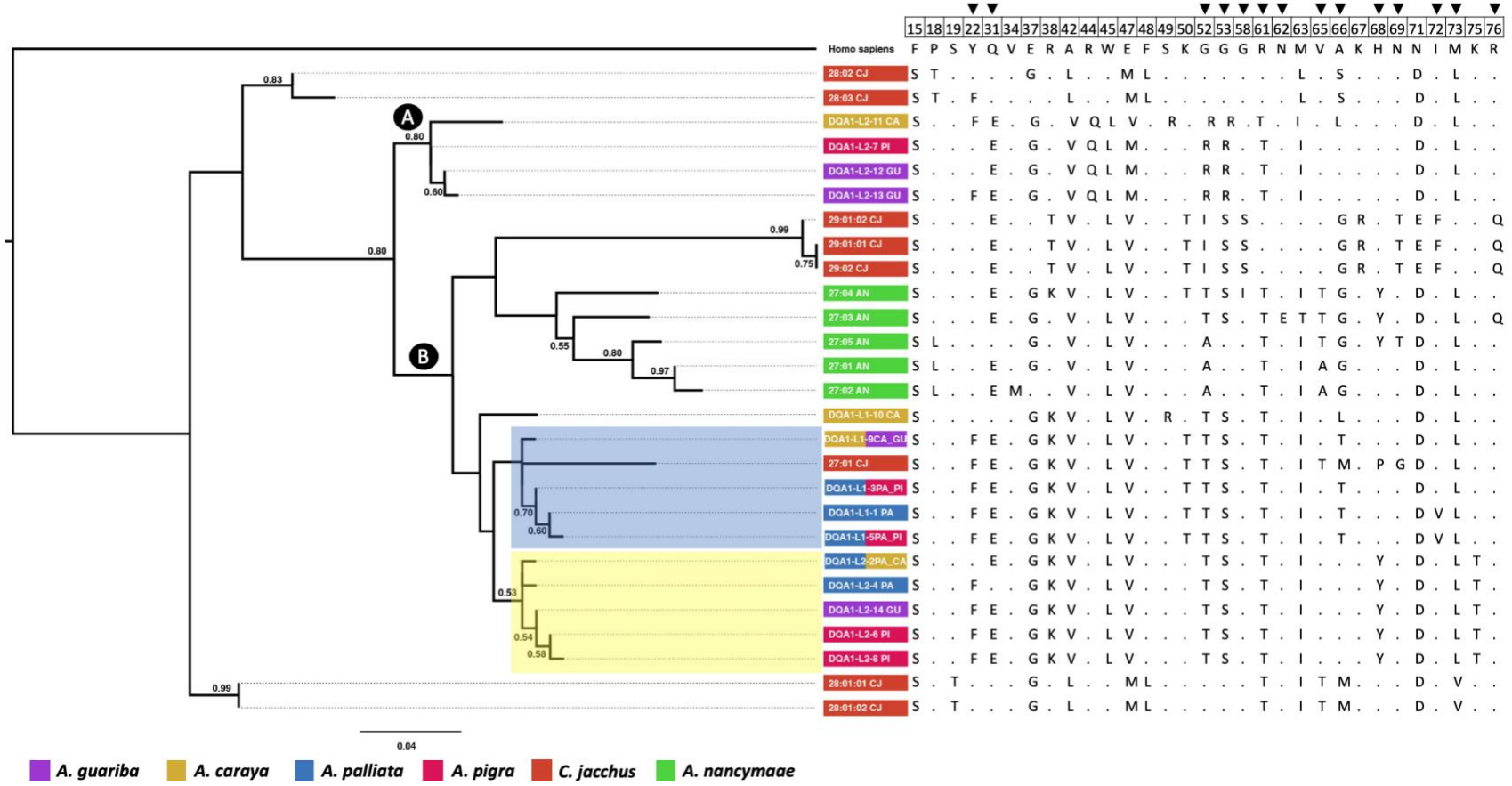
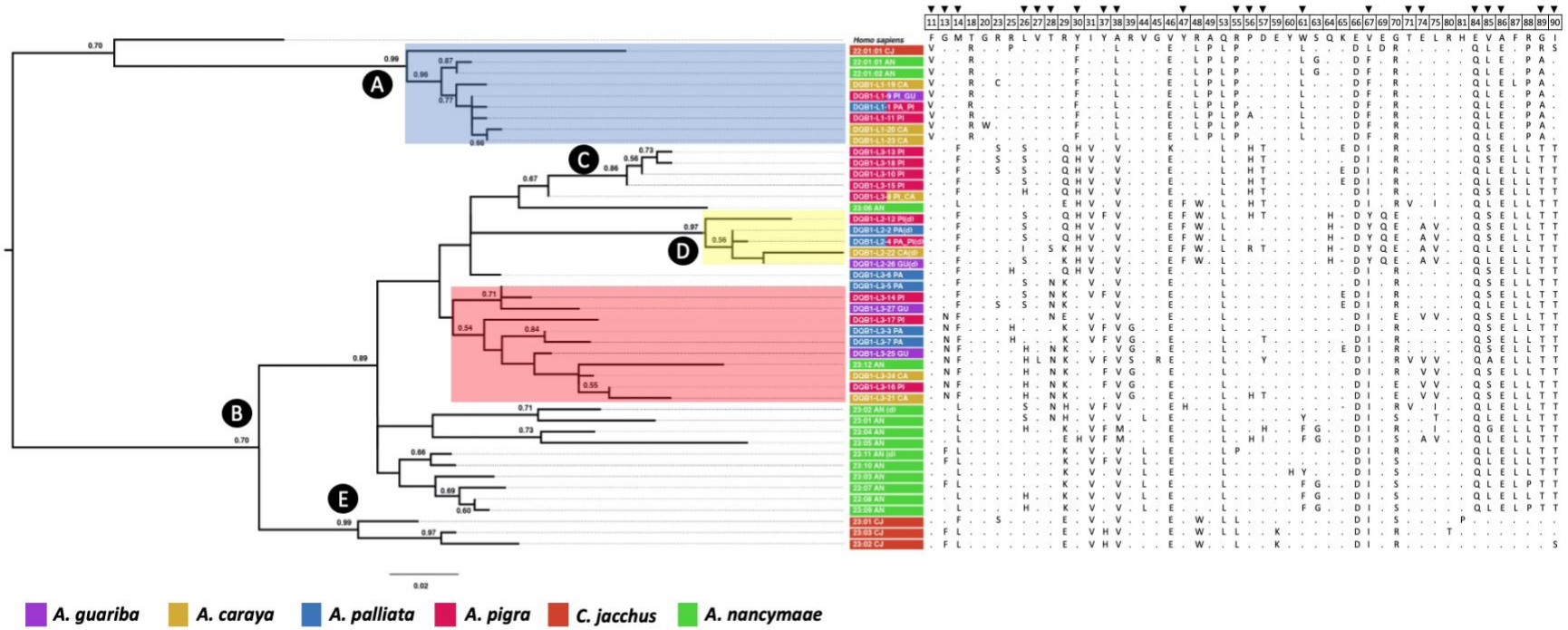


Figure 1-2. Phylogenetic relationships of MHC-DQ alleles for A) DQA1, B) DQB1, and C) DQB2. Variable aligned amino acid sites across species are shown to the right of each allele sequence in the tree. Phylogenetic trees include bootstrap values at each branch, branches with no bootstrap value had a value lower than 0.50. Informational clusters were designated at the base of the node with a black circle and a white letter. Sequence names are embedded in colored rectangles representing each species, double-colored rectangles indicate shared alleles. *C. jacchus* = maroon, *A. nancymaae* = green, *A. palliata* = blue, *A. pigra* = red, *A. caraya* = gold and *A. guariba* = purple, also explained in the legend below the image. Amino acid alignments show only the variable sites where dots indicate conservation of the site compared to the reference human sequence. Top row shows the amino acid position and black triangles indicate antigen binding sites. For A) DQA1, alleles belonging to the DQA1-L1 locus are highlighted within the blue box and alleles belonging to the DQA1-L2 locus are highlighted within the yellow box. For B) DQB1, alleles belonging to the DQB1-L1 locus are highlighted within the blue box, alleles belonging to the DQA1-L2 locus are highlighted within the yellow box and alleles belonging to the DQB1-L3 locus are highlighted within the red box.

A)



B)


A
B
C
D
E
A. guariba
A. caraya
A. palliata
A. pigra
C. jacchus
A. nancymaae

C)

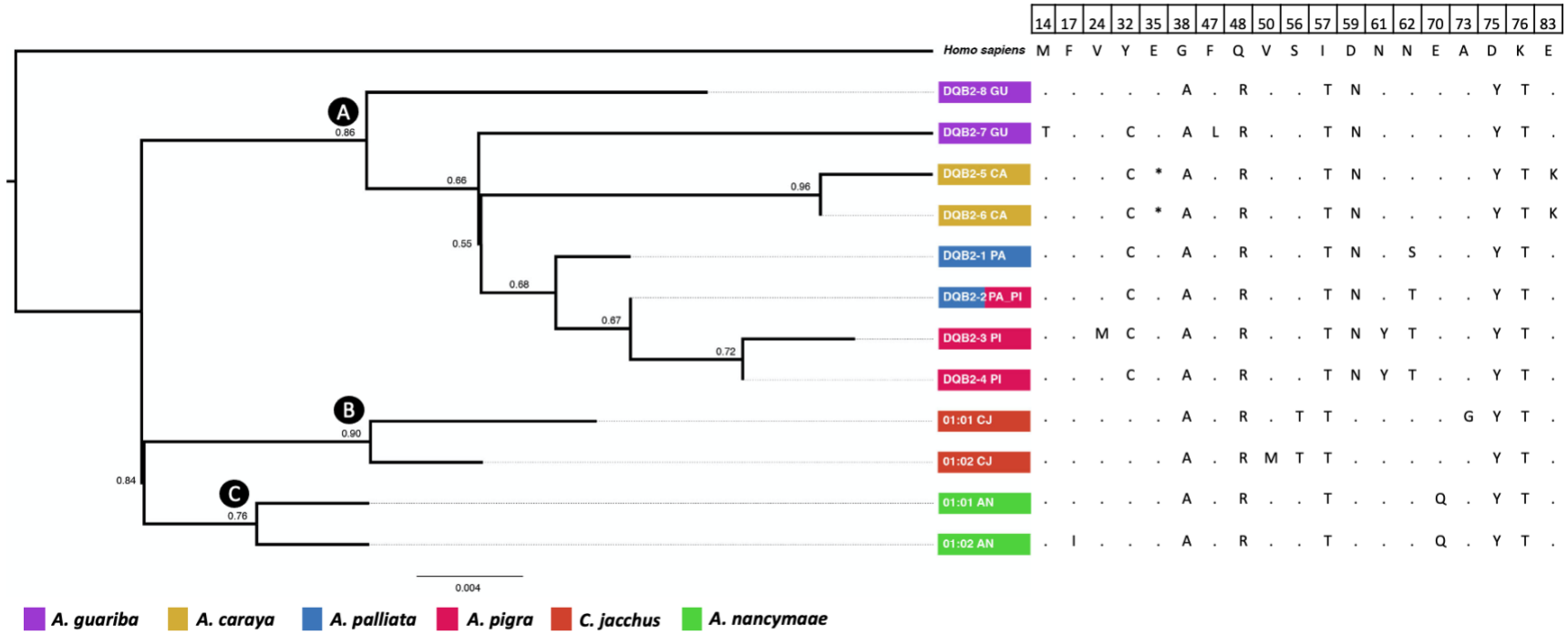
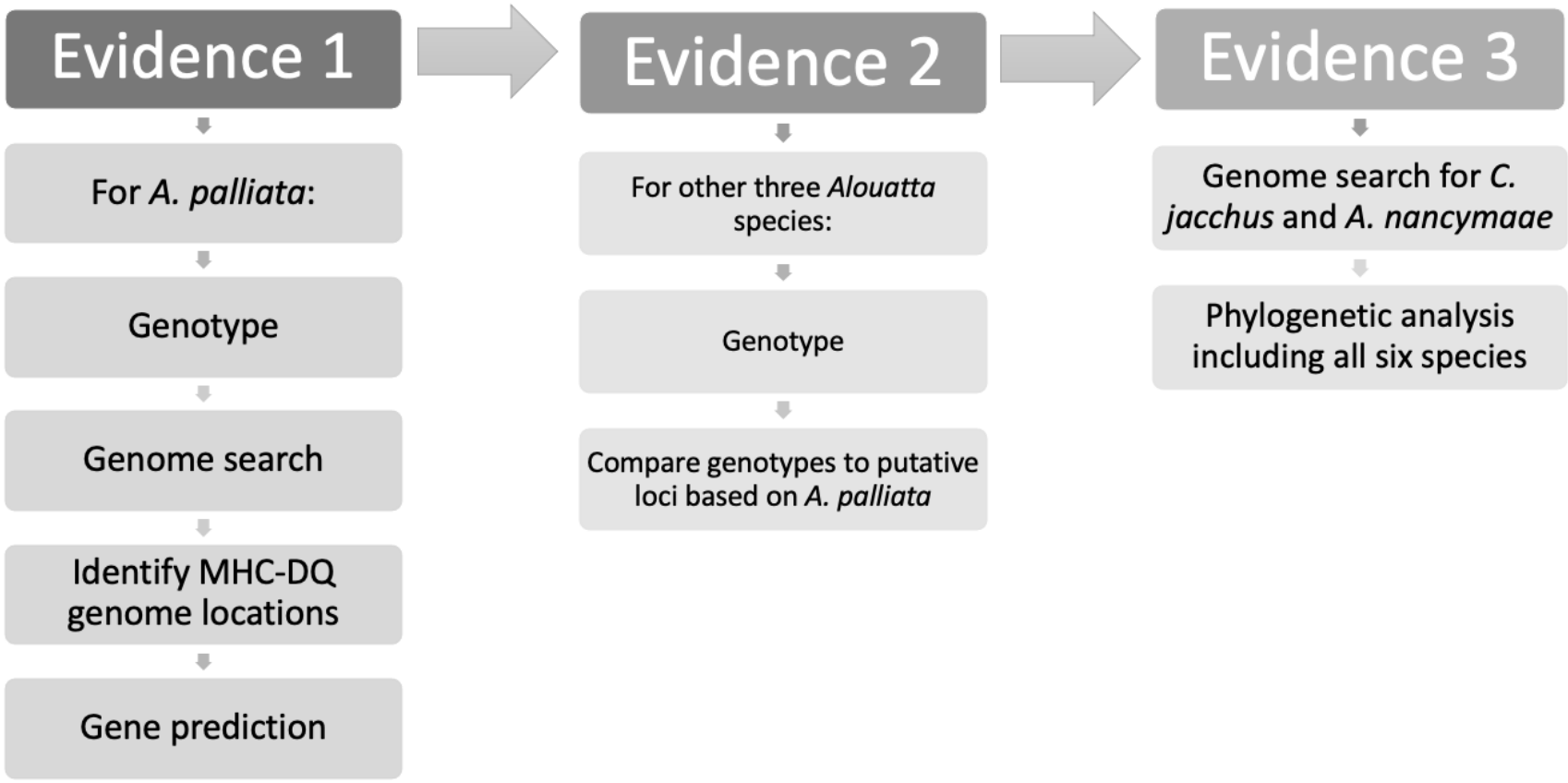

14 17 24 32 35 38 47 48 50 56 57 59 61 62 70 73 75 76 83
Homo sapiens M F V Y E G F Q V S I D N N E A D K E
DQB2-8 GU . . . . . A . R . . T N . . . . Y T .
DQB2-7 GU T . . C . A L R . . T N . . . . Y T .
DQB2-5 CA . . . C * A . R . . T N . . . . Y T K
DQB2-6 CA . . . C * A . R . . T N . . . . Y T K
DQB2-1 PA . . . C . A . R . . T N . S . . Y T .
DQB2-2 PA_PI . . . C . A . R . . T N . T . . Y T .
DQB2-3 PI . . M C . A . R . . T N Y T . . Y T .
DQB2-4 PI . . . C . A . R . . T N Y T . . Y T .
01:01 CJ . . . . . A . R . T T . . . . G Y T .
01:02 CJ . . . . . A . R M T T . . . . . Y T .
01:01 AN . . . . . A . R . . T . . . Q . Y T .
01:02 AN . I . . . A . R . . T . . . Q . Y T .
A
0.86
0.66
0.96
0.55
0.68
0.67
0.72
B
0.90
0.84
C
0.76
0.004
A. guariba
A. caraya
A. palliata
A. pigra
C. jacchus
A. nancymaae

Figure 1-3. Schematic representation of different lines of evidence supporting gene duplication at MHC-DQA1 and MHC-DQB1.



Supplementary Materials

Table 1-S6. Aotus nancymaae and Callithrix jacchus MHC-DQ sequences. This table contains a list of all the MHC-DQ sequences publicly available from these two species that were used in our study. We provide the accession number and the allele name based on the IPD-MHC nomenclature.

Allele	Gene	Species	Accession number
Aona-DQA1*27:01	DQA1	*Aotus nancymaae*	NHP00001
Aona-DQA1*27:02	DQA1	*Aotus nancymaae*	NHP00002
Aona-DQA1*27:03	DQA1	*Aotus nancymaae*	NHP00003
Aona-DQA1*27:04	DQA1	*Aotus nancymaae*	NHP00004
Aona-DQA1*27:05	DQA1	*Aotus nancymaae*	NHP00005
Caja-DQA1*01:01	DQA1	*Callithrix jacchus*	NHP00085
Caja-DQA1*27:01	DQA1	*Callithrix jacchus*	NHP00086
Caja-DQA1*28:01:01	DQA1	*Callithrix jacchus*	NHP10384
Caja-DQA1*28:01:02	DQA1	*Callithrix jacchus*	NHP10385
Caja-DQA1*28:02	DQA1	*Callithrix jacchus*	NHP10386
Caja-DQA1*29:01:01	DQA1	*Callithrix jacchus*	NHP10387
Caja-DQA1*29:01:02	DQA1	*Callithrix jacchus*	NHP10388
Caja-DQA1*28:03	DQA1	*Callithrix jacchus*	NHP10389
Caja-DQA1*29:02	DQA1	*Callithrix jacchus*	NHP10390
Aona-DQB1*22:01:01	DQB1	*Aotus nancymaae*	NHP00006
Aona-DQB1*22:01:02	DQB1	*Aotus nancymaae*	NHP00007
Aona-DQB1*23:01	DQB1	*Aotus nancymaae*	NHP00008
Aona-DQB1*23:02	DQB1	*Aotus nancymaae*	NHP00009
Aona-DQB1*23:03	DQB1	*Aotus nancymaae*	NHP00010
Aona-DQB1*23:04	DQB1	*Aotus nancymaae*	NHP00011
Aona-DQB1*23:05	DQB1	*Aotus nancymaae*	NHP00012
Aona-DQB1*23:06	DQB1	*Aotus nancymaae*	NHP00013
Aona-DQB1*23:07	DQB1	*Aotus nancymaae*	NHP00014
Aona-DQB1*23:08	DQB1	*Aotus nancymaae*	NHP00015

Aona-DQB1*23:09	DQB1	*Aotus nancymaae*	NHP00016
Aona-DQB1*23:10	DQB1	*Aotus nancymaae*	NHP00017
Aona-DQB1*23:11	DQB1	*Aotus nancymaae*	NHP00018
Aona-DQB1*23:12	DQB1	*Aotus nancymaae*	NHP00019
Caja-DQB1*22:01:01	DQB1	*Callithrix jacchus*	NHP00087
Caja-DQB1*22:01:02	DQB1	*Callithrix jacchus*	NHP00088
Caja-DQB1*23:01	DQB1	*Callithrix jacchus*	NHP00089
Caja-DQB1*23:02	DQB1	*Callithrix jacchus*	NHP00090
Caja-DQB1*23:03	DQB1	*Callithrix jacchus*	NHP10391
Caja-DQB1*23:04	DQB1	*Callithrix jacchus*	NHP10392
Aona-DQB2*01:01	DQB2	*Aotus nancymaae*	NHP00020
Aona-DQB2*01:02	DQB2	*Aotus nancymaae*	NHP00021
Caja-DQB2*01:01	DQB2	*Callithrix jacchus*	NHP00091
Caja-DQB2*01:02	DQB2	*Callithrix jacchus*	NHP00092

Table 1-S7. Individual Genotypes. List of alleles identified per individual across four *Alouatta* species at each locus. A) DQA1, B) DQB1 and C) DQB2. N/A = not amplified.

A) DQA1

Species	Individual ID	Alleles	
A. palliata	S-001	DQA1-L1-1	DQA1-L2-2
A. palliata	S-002	DQA1-L1-1	DQA1-L2-2
A. palliata	S-003	DQA1-L1-1	DQA1-L2-2
A. palliata	S-004	DQA1-L1-1	DQA1-L2-2
A. palliata	S-005	DQA1-L1-1	DQA1-L2-2
A. palliata	S-087	DQA1-L1-1	DQA1-L2-2
A. palliata	S-088	DQA1-L1-1	DQA1-L2-2
A. palliata	S-089	DQA1-L1-1	DQA1-L2-2
A. palliata	S-090	DQA1-L1-1	DQA1-L2-2
A. palliata	S-091	DQA1-L1-1	DQA1-L2-2
A. palliata	S-092	DQA1-L1-1	DQA1-L2-2
A. palliata	S-143	DQA1-L1-1	DQA1-L2-2
A. palliata	S-144	DQA1-L1-1	DQA1-L2-2
A. palliata	S-145	DQA1-L1-1	DQA1-L2-2
A. palliata	S-146	DQA1-L1-1	DQA1-L2-2
A. palliata	S-147	DQA1-L1-1	DQA1-L2-2
A. palliata	S-172	DQA1-L1-1	DQA1-L2-2
A. palliata	S-173	DQA1-L1-1	DQA1-L2-2
A. palliata	S-174	DQA1-L1-1	DQA1-L2-2
A. palliata	S-175	DQA1-L1-1	DQA1-L2-2
A. palliata	S-176	DQA1-L1-1	DQA1-L2-2
A. palliata	S-204	DQA1-L1-1	DQA1-L2-2
A. palliata	S-205	DQA1-L1-1	DQA1-L2-2
A. palliata	S-206	DQA1-L1-1	DQA1-L2-2
A. palliata	S-207	DQA1-L1-1	DQA1-L2-2
A. palliata	S-208	DQA1-L1-1	DQA1-L2-2
A. palliata	S-209	DQA1-L1-1	DQA1-L2-2
A. palliata	S-608	DQA1-L1-1	DQA1-L2-2
A. palliata	S-609	DQA1-L1-1	DQA1-L2-2
A. palliata	S-610	DQA1-L1-1	DQA1-L2-2
A. palliata	S-611	DQA1-L1-1	DQA1-L2-2
A. palliata	S-612	DQA1-L1-1	DQA1-L2-2
A. palliata	S-613	DQA1-L1-1	DQA1-L2-2

A. palliata	S-614	DQA1-L1-1	DQA1-L2-2		
A. palliata	S-615	DQA1-L1-1	DQA1-L2-2		
A. palliata	S-617	DQA1-L1-1	DQA1-L2-2		
A. palliata	S-618	DQA1-L1-1	DQA1-L2-2		
A. palliata	S-030	DQA1-L1-1	DQA1-L2-2		
A. palliata	S-031	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	
A. palliata	S-032	DQA1-L1-1	DQA1-L2-4	DQA1-L1-5	
A. palliata	S-033	DQA1-L1-1	DQA1-L2-2		
A. palliata	S-034	DQA1-L1-1	DQA1-L2-4	DQA1-L1-5	
A. palliata	S-035	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	
A. palliata	S-036	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	
A. palliata	S-037	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	
A. palliata	S-038	DQA1-L1-1	DQA1-L2-2	DQA1-L1-5	
A. palliata	S-039	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	
A. palliata	Peru-A2	DQA1-L1-1	DQA1-L2-2	DQA1-L1-3	
A. palliata	Peru-A4	DQA1-L1-1	DQA1-L1-3		
A. palliata	Peru-A5	DQA1-L1-1			
A. palliata	Peru-A10	DQA1-L1-3			
A. palliata	Peru-A16	DQA1-L1-1	DQA1-L1-3		
A. palliata	Peru-G1	NA			
A. palliata	Peru-G5	DQA1-L1-3			
A. palliata	Peru-G6	DQA1-L1-3			
A. palliata	Peru-G11	DQA1-L2-2	DQA1-L1-3		
A. palliata	Peru-G24	DQA1-L1-1	DQA1-L2-2	DQA1-L1-3	
A. pigra	S-149	DQA1-L1-3	DQA1-L1-5	DQA1-L2-7	DQA1-L2-8
A. pigra	S-150	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-151	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-153	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-185	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8	
A. pigra	S-186	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-187	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8	
A. pigra	S-188	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-189	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-190	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-191	DQA1-L1-3	DQA1-L1-5	DQA1-L2-7	
A. pigra	S-193	DQA1-L1-3	DQA1-L1-5	DQA1-L2-7	
A. pigra	S-194	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-195	DQA1-L1-3	DQA1-L2-6		

A. pigra	S-196	DQA1-L1-3	DQA1-L1-5	DQA1-L2-7	
A. pigra	S-219	DQA1-L1-3	DQA1-L2-8		
A. pigra	S-222	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-224	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-225	DQA1-L1-3	DQA1-L1-5	DQA1-L2-7	DQA1-L2-8
A. pigra	S-227	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-228	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-229	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8	
A. pigra	S-234	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8	
A. pigra	S-242	DQA1-L1-3	DQA1-L2-8		
A. pigra	S-248	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-279	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8	
A. pigra	S-289	DQA1-L1-3	DQA1-L2-8		
A. pigra	S-490	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-491	DQA1-L1-3	DQA1-L2-8		
A. pigra	S-494	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-504	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-505	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-507	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-508	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-509	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-510	DQA1-L1-3	DQA1-L2-6		
A. pigra	S-511	DQA1-L1-3	DQA1-L2-6	DQA1-L2-7	
A. pigra	S-526	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8	
A. caraya	H1	DQA1-L1-9	DQA1-L1-10	DQA1-L2-2	
A. caraya	21C	DQA1-L1-9	DQA1-L1-10	DQA1-L2-11	
A. guariba	109	DQA1-L1-9	DQA1-L2-12		
A. guariba	89	DQA1-L1-9	DQA1-L2-14	DQA1-L2-13	

B) DQB1

Species	Individual ID	Alleles		
A. palliata	S-001	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-002	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-003	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-004	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-005	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-087	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-088	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-089	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-090	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-091	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-092	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-143	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-144	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-145	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-146	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-147	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-172	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-173	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-174	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-175	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-176	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-204	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-205	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-206	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-207	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-208	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-209	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-608	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-609	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-610	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-611	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-612	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-613	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-614	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-615	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3
A. palliata	S-617	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3

A. palliata	S-618	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3		
A. palliata	S-030	DQB1-L1-1	DQB1-L2-2	DQB1-L3-5		
A. palliata	S-031	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3	DQB1-L2-4	DQB1-L3-7
A. palliata	S-032	DQB1-L1-1	DQB1-L3-3	DQB1-L2-4	DQB1-L3-7	
A. palliata	S-033	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3		
A. palliata	S-034	DQB1-L1-1	DQB1-L3-3	DQB1-L2-4	DQB1-L3-7	
A. palliata	S-035	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-7	
A. palliata	S-036	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-7	
A. palliata	S-037	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3	DQB1-L2-4	
A. palliata	S-038	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3		
A. palliata	S-039	DQB1-L1-1	DQB1-L2-2	DQB1-L3-3	DQB1-L2-4	DQB1-L3-7
A. palliata	Peru-A2	DQB1-L1-1	DQB1-L2-2	DQB1-L3-6		
A. palliata	Peru-A4	DQB1-L1-1	DQB1-L2-2			
A. palliata	Peru-A5	DQB1-L1-1	DQB1-L2-2	DQB1-L3-6		
A. palliata	Peru-A10	DQB1-L1-1	DQB1-L2-2	DQB1-L3-6		
A. palliata	Peru-A16	DQB1-L3-6				
A. palliata	Peru-G1	DQB1-L1-1	DQB1-L2-2	DQB1-L3-6		
A. palliata	Peru-G5	DQB1-L3-6				
A. palliata	Peru-G6	DQB1-L1-1	DQB1-L2-2	DQB1-L3-6		
A. palliata	Peru-G11	DQB1-L1-1	DQB1-L2-2	DQB1-L3-6		
A. palliata	Peru-G24	DQB1-L1-1	DQB1-L2-2	DQB1-L3-6		
A. pigra	S-149	DQB1-L1-1	DQB1-L2-4	DQB1-L1-9	DQB1-L2-12	DQB1-L3-18
A. pigra	S-150	DQB1-L3-8	DQB1-L3-10	DQB1-L1-11		
A. pigra	S-151	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11	
A. pigra	S-153	DQB1-L1-1	DQB1-L2-4	DQB1-L1-9	DQB1-L3-8	DQB1-L3-10
A. pigra	S-185	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L2-12	DQB1-L3-18
A. pigra	S-186	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11	
A. pigra	S-187	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L2-12	DQB1-L3-18
A. pigra	S-188	DQB1-L1-1	DQB1-L2-4	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10
A. pigra	S-189	DQB1-L1-1	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	DQB1-L1-11
A. pigra	S-190	DQB1-L1-1	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	DQB1-L1-11
A. pigra	S-191	DQB1-L1-1	DQB1-L2-4	DQB1-L1-11	DQB1-L3-13	DQB1-L3-15
A. pigra	S-193	DQB1-L1-1	DQB1-L2-4	DQB1-L1-9	DQB1-L3-17	
A. pigra	S-194	DQB1-L1-1	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	DQB1-L1-11
A. pigra	S-195	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11	DQB1-L3-17
A. pigra	S-196	DQB1-L2-4	DQB1-L1-9	DQB1-L3-17		
A. pigra	S-219	DQB1-L1-9	DQB1-L2-12	DQB1-L3-13	DQB1-L3-16	DQB1-L3-18
A. pigra	S-222	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11	DQB1-L3-13	DQB1-L3-15

A. pigra	S-224	DQB1-L2-4	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L3-18	
A. pigra	S-225	DQB1-L1-1	DQB1-L1-9	DQB1-L2-12	DQB1-L3-18		
A. pigra	S-227	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11		
A. pigra	S-228	DQB1-L1-1	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	DQB1-L1-11	
A. pigra	S-229	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11	DQB1-L2-12	DQB1-L3-18
A. pigra	S-234	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11	DQB1-L2-12	DQB1-L3-18
A. pigra	S-242	DQB1-L1-9	DQB1-L2-12	DQB1-L3-13	DQB1-L3-15	DQB1-L3-18	
A. pigra	S-248	DQB1-L1-1	DQB1-L2-4	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	
A. pigra	S-279	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L2-12	DQB1-L3-18	
A. pigra	S-289	DQB1-L1-11	DQB1-L2-12	DQB1-L3-13	DQB1-L3-15	DQB1-L3-18	
A. pigra	S-490	DQB1-L1-1	DQB1-L2-4	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	
A. pigra	S-491	DQB1-L1-9	DQB1-L1-11	DQB1-L2-12	DQB1-L3-18		
A. pigra	S-494	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11		
A. pigra	S-504	DQB1-L3-8	DQB1-L1-9	DQB1-L1-11	DQB1-L3-13	DQB1-L3-15	DQB1-L3-16
A. pigra	S-505	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L3-16		
A. pigra	S-507	DQB1-L1-11	DQB1-L3-13	DQB1-L3-15			
A. pigra	S-508	DQB1-L3-8	DQB1-L3-10	DQB1-L1-11	DQB1-L3-13	DQB1-L3-15	
A. pigra	S-509	DQB1-L3-8	DQB1-L3-10	DQB1-L1-11	DQB1-L3-18		
A. pigra	S-510	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11	DQB1-L3-13	DQB1-L3-15
A. pigra	S-511	DQB1-L2-4	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11	
A. pigra	S-526	DQB1-L3-8	DQB1-L1-9	DQB1-L3-10	DQB1-L1-11	DQB1-L2-12	DQB1-L3-18
A. caraya	H1	DQB1-L1-19	DQB1-L1-20	DQB1-L3-21	DQB1-L2-22		
A. caraya	21C	DQB1-L3-24	DQB1-L1-19	DQB1-L1-23	DQB1-L3-8		
A. guariba	109	DQB1-L1-9	DQB1-L2-26	DQB1-L3-25			
A. guariba	89	DQB1-L1-9	DQB1-L3-27	DQB1-L2-26	DQB1-L3-25		

C) DQB2

Species	Individual ID	Alleles
A. palliata	S-001	DQB2-1
A. palliata	S-002	DQB2-1
A. palliata	S-003	DQB2-1
A. palliata	S-004	DQB2-1
A. palliata	S-005	DQB2-1
A. palliata	S-087	DQB2-1
A. palliata	S-088	DQB2-1
A. palliata	S-089	DQB2-1
A. palliata	S-090	DQB2-1
A. palliata	S-091	DQB2-1
A. palliata	S-092	DQB2-1
A. palliata	S-143	DQB2-1
A. palliata	S-144	DQB2-1
A. palliata	S-145	DQB2-1
A. palliata	S-146	DQB2-1
A. palliata	S-147	DQB2-1
A. palliata	S-172	DQB2-1
A. palliata	S-173	DQB2-1
A. palliata	S-174	DQB2-1
A. palliata	S-175	DQB2-1
A. palliata	S-176	DQB2-1
A. palliata	S-204	DQB2-1
A. palliata	S-205	DQB2-1
A. palliata	S-206	DQB2-1
A. palliata	S-207	DQB2-1
A. palliata	S-208	DQB2-1
A. palliata	S-209	DQB2-1
A. palliata	S-608	DQB2-1
A. palliata	S-609	DQB2-1
A. palliata	S-610	DQB2-1
A. palliata	S-611	DQB2-1
A. palliata	S-612	DQB2-1
A. palliata	S-613	DQB2-1
A. palliata	S-614	DQB2-1
A. palliata	S-615	DQB2-1
A. palliata	S-617	DQB2-1

A. palliata	S-618	DQB2-1	
A. palliata	S-030	DQB2-2	
A. palliata	S-031	DQB2-1	
A. palliata	S-032	DQB2-1	DQB2-2
A. palliata	S-033	DQB2-1	
A. palliata	S-034	DQB2-1	DQB2-2
A. palliata	S-035	DQB2-1	
A. palliata	S-036	DQB2-1	
A. palliata	S-037	DQB2-1	
A. palliata	S-038	DQB2-1	
A. palliata	S-039	DQB2-1	
A. palliata	Peru-A2	DQB2-2	
A. palliata	Peru-A4	DQB2-2	
A. palliata	Peru-A5	DQB2-1	DQB2-2
A. palliata	Peru-A10	DQB2-2	
A. palliata	Peru-A16		
A. palliata	Peru-G1	DQB2-1	DQB2-2
A. palliata	Peru-G5	DQB2-2	
A. palliata	Peru-G6	DQB2-1	DQB2-2
A. palliata	Peru-G11	DQB2-1	DQB2-2
A. palliata	Peru-G24	DQB2-2	
A. pigra	S-149	DQB2-2	DQB2-3
A. pigra	S-150	DQB2-3	
A. pigra	S-151	DQB2-3	
A. pigra	S-153	DQB2-2	DQB2-3
A. pigra	S-185	DQB2-3	
A. pigra	S-186	DQB2-3	
A. pigra	S-187	DQB2-3	
A. pigra	S-188	DQB2-2	DQB2-3
A. pigra	S-189	DQB2-2	DQB2-3
A. pigra	S-190	DQB2-2	DQB2-3
A. pigra	S-191	DQB2-2	DQB2-4
A. pigra	S-193	DQB2-2	DQB2-3
A. pigra	S-194	DQB2-2	DQB2-3
A. pigra	S-195	DQB2-3	
A. pigra	S-196	DQB2-2	DQB2-3
A. pigra	S-219	DQB2-3	
A. pigra	S-222	DQB2-3	

A. pigra	S-224	DQB2-3	
A. pigra	S-225	DQB2-2	DQB2-3
A. pigra	S-227	DQB2-3	
A. pigra	S-228	DQB2-2	DQB2-3
A. pigra	S-229	DQB2-3	
A. pigra	S-234	DQB2-3	
A. pigra	S-242	DQB2-3	
A. pigra	S-248	DQB2-2	DQB2-3
A. pigra	S-279	DQB2-3	
A. pigra	S-289	DQB2-3	DQB2-4
A. pigra	S-490	DQB2-2	DQB2-3
A. pigra	S-491	DQB2-3	
A. pigra	S-494	DQB2-3	
A. pigra	S-504	DQB2-3	DQB2-4
A. pigra	S-505	DQB2-3	
A. pigra	S-507	DQB2-4	
A. pigra	S-508	DQB2-4	
A. pigra	S-509	DQB2-4	
A. pigra	S-510	DQB2-3	DQB2-4
A. pigra	S-511	DQB2-3	
A. pigra	S-526	DQB2-3	
A. caraya	H1	DQB2-5	DQB2-6
A. caraya	21C	DQB2-5	DQB2-6
A. guariba	109	DQB2-7	
A. guariba	89	DQB2-8	

Table 1-S8. Alouatta palliata genome search. Blastn search results are shown for each allele that matched the *A. palliata* in the genome assembly. We include the name and length of the scaffold that each allele matched to, as well as the start and end position of the sequence within that scaffold. We also report the percent of the allele sequence that aligned to the genome (query cover) and the similarity (i.e., how many characters in each sequence are identical) of the allele sequence to the sequence in the assembly (percent identity). We also report E-values to provide a quantitative estimate of the quality of the match, the lower the value the better the match. Rows that are colored in light grey represent perfect matches.

Locus	Allele	Scaffold	Contig length (bp)	Start position	End position	Query cover	Percent identity	E-value	Accession
DQA1	DQA1-L1-1	AloPal_scaffold_8772	128650	11269	11454	100%	100%	9.00E-93	PVKV010004392.1
DQA1	DQA1-L1-3	AloPal_scaffold_8772	128650	11269	11454	100%	99.46%	4.00E-91	PVKV010004392.1
DQA1	DQA1-L1-5	AloPal_scaffold_8772	128650	11269	11454	100%	99.46%	4.00E-91	PVKV010004392.1
DQA1	DQA1-L2-2	AloPal_scaffold_50203	30189	4904	5089	100%	100%	9.00E-93	PVKV010025127.1
DQA1	DQA1-L2-4	AloPal_scaffold_50203	30189	4904	5089	100%	98.92%	2.00E-89	PVKV010025127.1
DQB1	DQB1-L1-1	AloPal_scaffold_89817	9103	5475	5724	100%	100%	3.00E-128	PVKV010044961.1
DQB1	DQB1-L3-3	AloPal_scaffold_61825	21639	14092	14341	100%	100%	3.00E-128	PVKV010030942.1
DQB1	DQB1-L3-7	AloPal_scaffold_61825	21639	14092	14341	100%	99.20%	7.00E-125	PVKV010030942.1
DQB1	DQB1-L3-6	AloPal_scaffold_61825	21639	14092	14341	100%	97.20%	2.00E-116	PVKV010030942.1
DQB1	DQB1-L3-5	AloPal_scaffold_61825	21639	14092	14341	100%	96.40%	3.00E-113	PVKV010030942.1
DQB1	DQB1-L2-2	AloPal_scaffold_64423	20065	10916	11162	100%	100%	1.00E-126	PVKV010032242.1
DQB1	DQB1-L2-4	AloPal_scaffold_64423	20065	10916	11162	100%	99.60%	7.00E-125	PVKV010032242.1
DQB2	DQB2-2	AloPal_scaffold_8772	128650	57006	57252	100%	100%	1.00E-126	PVKV010004392.1
DQB2	DQB2-8	AloPal_scaffold_8772	128650	57006	57252	100%	99.60%	7.00E-125	PVKV010004392.1

Table 1-S9. Callithrix jacchus genome search. Blastn search results are shown for each publicly available allele for *C. jacchus* in the genome assembly. We include the name and length of the scaffold that each allele was matched to, and the start and end position of the sequence within that scaffold. We also report the percent of the allele sequence that aligned to the genome (query cover) and the similarity (i.e., how many characters in each sequence are identical) of the allele sequence to the sequence in the assembly (percent identity) We also reported E-values to provide a quantitative estimate of the quality of the match, the lower the value the better the match. Rows that are colored in light grey represent perfect matches.

Locus	Allele	Contig	Contig Length	Start position	End position	Query Length	Percent identity	E value	Accession
DQA1	DQA1*27:01	cj1700	174,041,770	37733179	37733477	100	97.232	2.23E-144	NC_048386.1
DQA1	DQA1*28:01:01	cj1700	174,041,770	36889854	36890142	100	99.308	4.72E-146	NC_048386.1
DQA1	DQA1*28:01:02	cj1700	174,041,770	36889854	36890142	100	99.308	4.72E-146	NC_048386.1
DQA1	DQA1*28:02	cj1700	174,041,770	36889854	36890142	100	99.308	4.72E-146	NC_048386.1
DQA1	DQA1*29:01:01	cj1700	174,041,770	36889854	36890142	100	98.616	1.04E-142	NC_048386.1
DQA1	DQA1*29:01:02	cj1700	174,041,770	36889854	36890142	100	99.308	4.80E-146	NC_048386.1
DQA1	DQA1*28:03	cj1700	174,041,770	36889854	36890142	100	97.578	1.03E-137	NC_048386.1
DQA1	DQA1*29:02	cj1700	174,041,770	36889854	36890142	100	98.616	1.04E-142	NC_048386.1
DQA1	DQA1*29:02	cj1700	174,041,770	36889854	36890142	100	96.54	1.05E-132	NC_048386.1
DQB1	DQB1*22:01:01	cj1700	174,041,770	37750930	37750684	100	98.785	5.20E-121	NC_048386.1
DQB1	DQB1*23:01	cj1700	174,041,770	36905041	36904757	100	98.947	3.70E-142	NC_048386.1
DQB1	DQB1*23:02	cj1700	174,041,770	36907816	36907570	100	99.19	1.12E-122	NC_048386.1
DQB1	DQB1*23:03	cj1700	174,041,770	36905041	36904757	100	100	3.67E-147	NC_048386.1
DQB2	DQB2*01:01	cj1700	174,041,770	37691149	37690907	100	98.354	1.11E-117	NC_048386.1
DQB2	DQB2*01:02	cj1700	174,041,770	37691149	37690907	100	98.765	2.38E-119	NC_048386.1

Table 1-S10. Aotus nancymaae genome search. Blastn search results are shown for each publicly available allele for *A. nancymaae* in the genome assembly. We include the contig that each allele was matched to and the length of this contig. The start and end position of the sequence within that contig. What percent of the allele sequence was aligned to the genome (query cover) and the percent identity between the allele and the sequence it was matched to in the genome. We also reported E-values to provide a quantitative estimate of the quality of the match, the lower the value the better the match.

Locus	Allele	Contig	Contig length	Start position	End position	Query cover	Percent identity %	E value	Accession
DQA1	DQA1*27:01	601.0	446445	26979	26754	100	93.81	1.00E-91	NW_018508875.1
DQA1	DQA1*27:02	601.0	446445	26979	26754	100	92.92	3.00E-88	NW_018508875.1
DQA1	DQA1*27:03	601.0	446445	26979	26754	98	100	1.00E-112	NW_018508875.1
DQA1	DQA1*27:04	358.84	14324	2204	1979	100	94.27	3.00E-93	NW_018508875.1
DQA1	DQA1*27:05	601.0	446445	26979	26754	100	94.69	6.00E-95	NW_018508875.1
DQB1	DQB1*22:01:01	46.78	581944	187240	186999	100	99.174	1.81E-120	NW_018506613.1
DQB1	DQB1*22:01:02	46.78	581944	187240	186999	100	99.587	3.89E-122	NW_018506613.1
DQB1	DQB1*23:01	601.0	446445	7174	7407	96	92.308	2.44E-89	NW_018508875.1
DQB1	DQB1*23:02	601.0	446445	7174	7407	96	94.872	3.07E-98	NW_018508875.1
DQB1	DQB1*23:03	601.0	446445	7174	7407	96	94.017	5.22E-96	NW_018508875.1
DQB1	DQB1*23:04	601.0	446445	7174	7407	96	94.017	5.22E-96	NW_018508875.1
DQB1	DQB1*23:05	46.78	581944	27314	27073	100	90.574	1.9E-85	NW_018506613.1
DQB1	DQB1*23:06	601.0	446445	7174	7407	96	93.617	2.43E-94	NW_018508875.1
DQB1	DQB1*23:07	601.0	446445	7174	7413	99	90.265	3.21E-78	NW_018506613.1
DQB1	DQB1*23:08	601.0	446445	7174	7407	96	95.299	5.18E-101	NW_018508875.1
DQB1	DQB1*23:09	601.0	446445	7174	7413	99	95	1.11E-102	NW_018508875.1
DQB1	DQB1*23:10	601.0	446445	7174	7407	96	95.726	1.11E-102	NW_018508875.1
DQB1	DQB1*23:11	601.0	446445	7174	7407	96	94.017	6.65E-95	NW_018508875.1
DQB1	DQB1*23:12	601.0	446445	7174	7407	96	93.162	1.13E-92	NW_018508875.1
DQB2	DQB2*01:01	46.78	581944	120020	119782	100	98.745	3.87E-117	NW_018506613.1
DQB2	DQB2*01:02	46.78	581944	120020	119782	100	98.745	3.87E-117	NW_018506613.1

Table 1-S11. Individual Genotypes according to new loci designation. List of alleles identified per individual across four *Alouatta* species at each locus. A) DQA1, B) DQB1 and C) DQB2. N/A = not amplified.

A) DQA1

Species	Individual ID	DQA1-L1		DQA1-L2	
A. palliata	S-001	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-002	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-003	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-004	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-005	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-087	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-088	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-089	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-090	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-091	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-092	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-143	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-144	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-145	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-146	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-147	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-172	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-173	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-174	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-175	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-176	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-204	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-205	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-206	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-207	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-208	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-209	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-608	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-609	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-610	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-611	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-612	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-613	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2

A. palliata	S-614	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-615	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-617	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-618	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-030	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-031	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4
A. palliata	S-032	DQA1-L1-1	DQA1-L1-5	DQA1-L2-4	DQA1-L2-4
A. palliata	S-033	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
A. palliata	S-034	DQA1-L1-1	DQA1-L1-5	DQA1-L2-4	DQA1-L2-4
A. palliata	S-035	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4
A. palliata	S-036	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4
A. palliata	S-037	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4
A. palliata	S-038	DQA1-L1-1	DQA1-L1-5	DQA1-L2-2	DQA1-L2-2
A. palliata	S-039	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4
A. palliata	Peru-A2	DQA1-L1-1	DQA1-L1-3	DQA1-L2-2	DQA1-L2-2
A. palliata	Peru-A4	DQA1-L1-1	DQA1-L1-3	NA	NA
A. palliata	Peru-A5	DQA1-L1-1	DQA1-L1-1	NA	NA
A. palliata	Peru-A10	DQA1-L1-3	DQA1-L1-3	NA	NA
A. palliata	Peru-A16	DQA1-L1-1	DQA1-L1-3	NA	NA
A. palliata	Peru-G1	NA	NA	NA	NA
A. palliata	Peru-G5	DQA1-L1-3	DQA1-L1-3	NA	NA
A. palliata	Peru-G6	DQA1-L1-3	DQA1-L1-3	NA	NA
A. palliata	Peru-G11	DQA1-L1-3	DQA1-L1-3	DQA1-L2-2	DQA1-L2-2
A. palliata	Peru-G24	DQA1-L1-1	DQA1-L1-3	DQA1-L2-2	DQA1-L2-2
A. pigra	S-149	DQA1-L1-3	DQA1-L1-5	DQA1-L2-8	DQA1-L2-7
A. pigra	S-150	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-151	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-153	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-185	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8
A. pigra	S-186	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-187	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8
A. pigra	S-188	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-189	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-190	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-191	DQA1-L1-3	DQA1-L1-5	DQA1-L2-7	DQA1-L2-7
A. pigra	S-193	DQA1-L1-3	DQA1-L1-5	DQA1-L2-7	DQA1-L2-7
A. pigra	S-194	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-195	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6

A. pigra	S-196	DQA1-L1-3	DQA1-L1-5	DQA1-L2-7	DQA1-L2-7
A. pigra	S-219	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-8
A. pigra	S-222	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-224	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-225	DQA1-L1-3	DQA1-L1-5	DQA1-L2-8	DQA1-L2-7
A. pigra	S-227	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-228	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-229	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8
A. pigra	S-234	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8
A. pigra	S-242	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-8
A. pigra	S-248	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-279	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8
A. pigra	S-289	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-8
A. pigra	S-490	DQA1-L1-3	DQA1-L1-5	DQA1-L2-6	DQA1-L2-7
A. pigra	S-491	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-8
A. pigra	S-494	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-504	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-505	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-507	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-508	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-509	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-510	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
A. pigra	S-511	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-7
A. pigra	S-526	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-8
A. caraya	H1	DQA1-L1-9	DQA1-L1-10	DQA1-L2-2	DQA1-L2-2
A. caraya	21C	DQA1-L1-9	DQA1-L1-10	DQA1-L2-11	DQA1-L2-11
A. guariba	109	DQA1-L1-9	DQA1-L1-9	DQA1-L2-12	DQA1-L2-12
A. guariba	89	DQA1-L1-9	DQA1-L1-9	DQA1-L2-13	DQA1-L2-14

B) DQB1

Species	Individual ID	DQB1-L1		DQB1-L2		DQB1-L3	
A. palliata	S-001	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-002	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-003	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-004	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-005	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-087	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-088	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-089	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-090	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-091	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-092	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-143	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-144	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-145	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-146	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-147	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-172	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-173	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-174	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-175	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-176	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-204	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-205	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-206	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-207	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-208	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-209	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-608	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-609	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-610	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-611	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-612	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-613	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-614	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-615	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3
A. palliata	S-617	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3

A. palliata	S-618	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
A. palliata	S-030	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-5	DQB1-L3-5	
A. palliata	S-031	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-3	DQB1-L3-7	
A. palliata	S-032	DQB1-L1-1	DQB1-L1-1	DQB1-L2-4	DQB1-L2-4	DQB1-L3-3	DQB1-L3-7	
A. palliata	S-033	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
A. palliata	S-034	DQB1-L1-1	DQB1-L1-1	DQB1-L2-4	DQB1-L2-4	DQB1-L3-3	DQB1-L3-7	
A. palliata	S-035	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-7	DQB1-L3-7	
A. palliata	S-036	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-7	DQB1-L3-7	
A. palliata	S-037	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-3	DQB1-L3-3	
A. palliata	S-038	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
A. palliata	S-039	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-3	DQB1-L3-7	
A. palliata	Peru-A2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	
A. palliata	Peru-A4	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	NA	NA	
A. palliata	Peru-A5	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	
A. palliata	Peru-A10	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	
A. palliata	Peru-A16	NA	NA	NA	NA	DQB1-L3-6	DQB1-L3-6	
A. palliata	Peru-G1	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	
A. palliata	Peru-G5	NA	NA	NA	NA	DQB1-L3-6	DQB1-L3-6	
A. palliata	Peru-G6	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	
A. palliata	Peru-G11	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	
A. palliata	Peru-G24	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	
A. pigra	S-149	DQB1-L1-1	DQB1-L1-9	DQB1-L2-4	DQB1-L2-12	DQB1-L3-18		
A. pigra	S-150	DQB1-L1-11	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	
A. pigra	S-151	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	
A. pigra	S-153	DQB1-L1-1	DQB1-L1-9	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	
A. pigra	S-185	DQB1-L1-9	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-8	DQB1-L3-10	DQB1-L3-18
A. pigra	S-186	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	
A. pigra	S-187	DQB1-L1-9	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-8	DQB1-L3-10	DQB1-L3-18
A. pigra	S-188	DQB1-L1-1	DQB1-L1-9	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	
A. pigra	S-189	DQB1-L1-1	DQB1-L1-11	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	
A. pigra	S-190	DQB1-L1-1	DQB1-L1-11	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	
A. pigra	S-191	DQB1-L1-1	DQB1-L1-11	DQB1-L2-4	DQB1-L2-4	DQB1-L3-13	DQB1-L3-15	
A. pigra	S-193	DQB1-L1-1	DQB1-L1-9	DQB1-L2-4	DQB1-L2-4	DQB1-L3-17		
A. pigra	S-194	DQB1-L1-1	DQB1-L1-11	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	
A. pigra	S-195	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	DQB1-L3-17
A. pigra	S-196	DQB1-L1-9	DQB1-L1-9	DQB1-L2-4	DQB1-L2-4	DQB1-L3-17		
A. pigra	S-219	DQB1-L1-9	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-13	DQB1-L3-14	DQB1-L3-18
A. pigra	S-222	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-10	DQB1-L3-13	DQB1-L3-15

A. pigra	S-224	DQB1-L1-9	DQB1-L1-9	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	DQB1-L3-18	
A. pigra	S-225	DQB1-L1-1	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-18			
A. pigra	S-227	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10		
A. pigra	S-228	DQB1-L1-1	DQB1-L1-11	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10		
A. pigra	S-229	DQB1-L1-9	DQB1-L1-11	DQB1-L2-12	DQB1-L2-12	DQB1-L3-8	DQB1-L3-10	DQB1-L3-18	
A. pigra	S-234	DQB1-L1-9	DQB1-L1-11	DQB1-L2-12	DQB1-L2-12	DQB1-L3-8	DQB1-L3-10	DQB1-L3-18	
A. pigra	S-242	DQB1-L1-9	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-13	DQB1-L3-14	DQB1-L3-18	
A. pigra	S-248	DQB1-L1-1	DQB1-L1-9	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10		
A. pigra	S-279	DQB1-L1-9	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-8	DQB1-L3-10	DQB1-L3-18	
A. pigra	S-289	DQB1-L1-11	DQB1-L1-11	DQB1-L2-12	DQB1-L2-12	DQB1-L3-13	DQB1-L3-14	DQB1-L3-18	
A. pigra	S-490	DQB1-L1-1	DQB1-L1-9	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10		
A. pigra	S-491	DQB1-L1-9	DQB1-L1-11	DQB1-L2-12	DQB1-L2-12	DQB1-L3-18			
A. pigra	S-494	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10		
A. pigra	S-504	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-13	DQB1-L3-15	DQB1-L3-16
A. pigra	S-505	DQB1-L1-9	DQB1-L1-9	NA	NA	DQB1-L3-8	DQB1-L3-10	DQB1-L3-16	
A. pigra	S-507	DQB1-L1-11	DQB1-L1-11	NA	NA	DQB1-L3-13	DQB1-L3-15		
A. pigra	S-508	DQB1-L1-11	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	DQB1-L3-13	DQB1-L3-15
A. pigra	S-509	DQB1-L1-11	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	DQB1-L3-18	
A. pigra	S-510	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	DQB1-L3-13	DQB1-L3-15
A. pigra	S-511	DQB1-L1-9	DQB1-L1-11	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10		
A. pigra	S-526	DQB1-L1-9	DQB1-L1-11	DQB1-L2-12	DQB1-L2-12	DQB1-L3-8	DQB1-L3-10	DQB1-L3-17	
A. caraya	H1	DQB1-L1-19	DQB1-L1-20	DQB1-L2-22	DQB1-L2-22	DQB1-L3-21	DQB1-L3-21		
A. caraya	21C	DQB1-L1-19	DQB1-L1-23	NA	NA	DQB1-L3-8	DQB1-L3-24		
A. guariba	109	DQB1-L1-9	DQB1-L1-9	DQB1-L2-26	DQB1-L2-26	DQB1-L3-25	DQB1-L3-25		
A. guariba	89	DQB1-L1-9	DQB1-L1-9	DQB1-L2-26	DQB1-L2-26	DQB1-L3-25	DQB1-L3-27		

C) DQB2

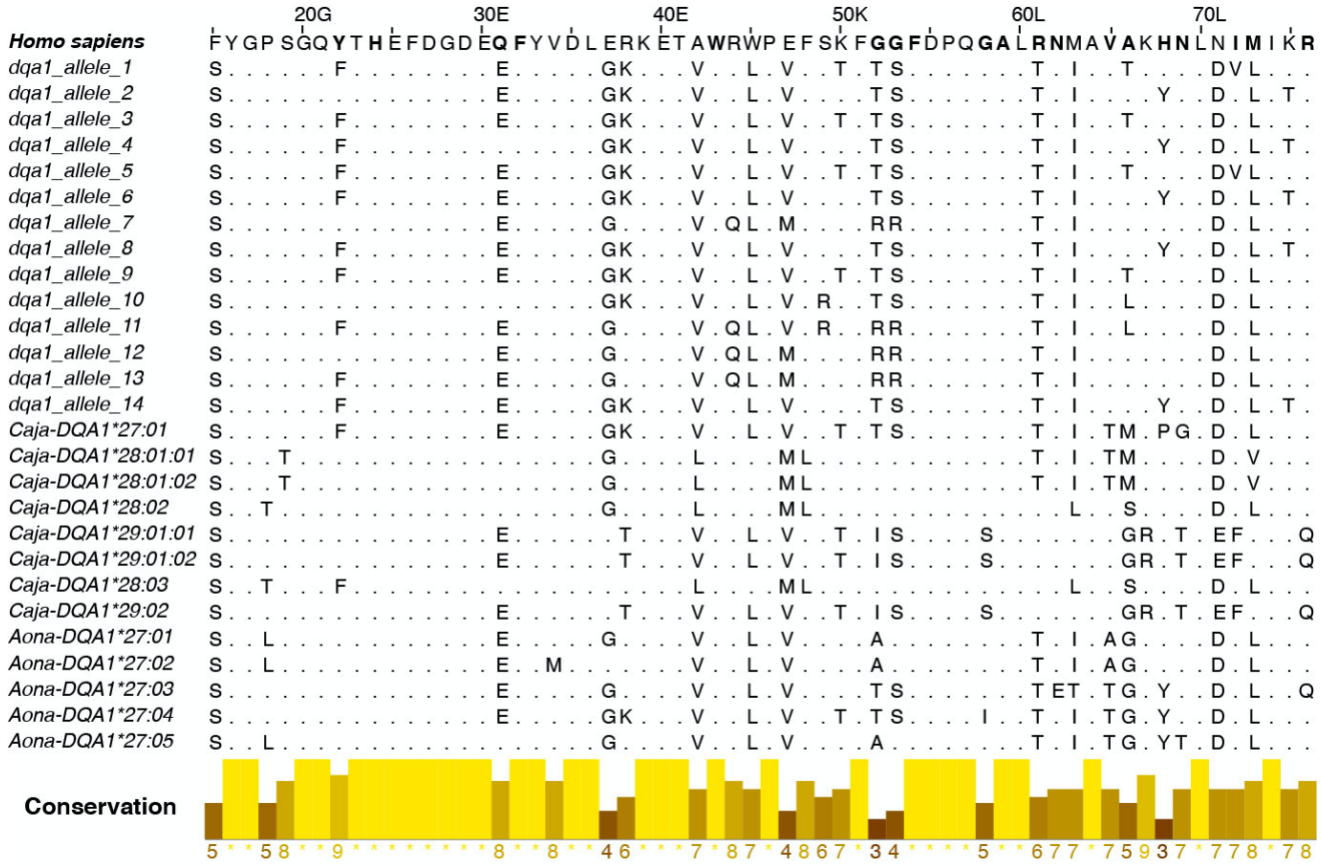
Species	Individual ID	DQB2	
A. palliata	S-001	DQB2-1	DQB2-1
A. palliata	S-002	DQB2-1	DQB2-1
A. palliata	S-003	DQB2-1	DQB2-1
A. palliata	S-004	DQB2-1	DQB2-1
A. palliata	S-005	DQB2-1	DQB2-1
A. palliata	S-087	DQB2-1	DQB2-1
A. palliata	S-088	DQB2-1	DQB2-1
A. palliata	S-089	DQB2-1	DQB2-1
A. palliata	S-090	DQB2-1	DQB2-1
A. palliata	S-091	DQB2-1	DQB2-1
A. palliata	S-092	DQB2-1	DQB2-1
A. palliata	S-143	DQB2-1	DQB2-1
A. palliata	S-144	DQB2-1	DQB2-1
A. palliata	S-145	DQB2-1	DQB2-1
A. palliata	S-146	DQB2-1	DQB2-1
A. palliata	S-147	DQB2-1	DQB2-1
A. palliata	S-172	DQB2-1	DQB2-1
A. palliata	S-173	DQB2-1	DQB2-1
A. palliata	S-174	DQB2-1	DQB2-1
A. palliata	S-175	DQB2-1	DQB2-1
A. palliata	S-176	DQB2-1	DQB2-1
A. palliata	S-204	DQB2-1	DQB2-1
A. palliata	S-205	DQB2-1	DQB2-1
A. palliata	S-206	DQB2-1	DQB2-1
A. palliata	S-207	DQB2-1	DQB2-1
A. palliata	S-208	DQB2-1	DQB2-1
A. palliata	S-209	DQB2-1	DQB2-1
A. palliata	S-608	DQB2-1	DQB2-1
A. palliata	S-609	DQB2-1	DQB2-1
A. palliata	S-610	DQB2-1	DQB2-1
A. palliata	S-611	DQB2-1	DQB2-1
A. palliata	S-612	DQB2-1	DQB2-1
A. palliata	S-613	DQB2-1	DQB2-1
A. palliata	S-614	DQB2-1	DQB2-1

A. palliata	S-615	DQB2-1	DQB2-1
A. palliata	S-617	DQB2-1	DQB2-1
A. palliata	S-618	DQB2-1	DQB2-1
A. palliata	S-030	DQB2-2	DQB2-2
A. palliata	S-031	DQB2-1	DQB2-1
A. palliata	S-032	DQB2-1	DQB2-2
A. palliata	S-033	DQB2-1	DQB2-1
A. palliata	S-034	DQB2-1	DQB2-2
A. palliata	S-035	DQB2-1	DQB2-1
A. palliata	S-036	DQB2-1	DQB2-1
A. palliata	S-037	DQB2-1	DQB2-1
A. palliata	S-038	DQB2-1	DQB2-1
A. palliata	S-039	DQB2-1	DQB2-1
A. palliata	Peru-A2	DQB2-2	DQB2-2
A. palliata	Peru-A4	DQB2-2	DQB2-2
A. palliata	Peru-A5	DQB2-1	DQB2-2
A. palliata	Peru-A10	DQB2-2	DQB2-2
A. palliata	Peru-A16	NA	NA
A. palliata	Peru-G1	DQB2-1	DQB2-2
A. palliata	Peru-G5	DQB2-2	DQB2-2
A. palliata	Peru-G6	DQB2-1	DQB2-2
A. palliata	Peru-G11	DQB2-1	DQB2-2
A. palliata	Peru-G24	DQB2-2	DQB2-2
A. pigra	S-149	DQB2-2	DQB2-3
A. pigra	S-150	DQB2-3	DQB2-3
A. pigra	S-151	DQB2-3	DQB2-3
A. pigra	S-153	DQB2-2	DQB2-3
A. pigra	S-185	DQB2-3	DQB2-3
A. pigra	S-186	DQB2-3	DQB2-3
A. pigra	S-187	DQB2-3	DQB2-3
A. pigra	S-188	DQB2-2	DQB2-3
A. pigra	S-189	DQB2-2	DQB2-3
A. pigra	S-190	DQB2-2	DQB2-3
A. pigra	S-191	DQB2-2	DQB2-4
A. pigra	S-193	DQB2-2	DQB2-3
A. pigra	S-194	DQB2-2	DQB2-3
A. pigra	S-195	DQB2-3	DQB2-3
A. pigra	S-196	DQB2-2	DQB2-3

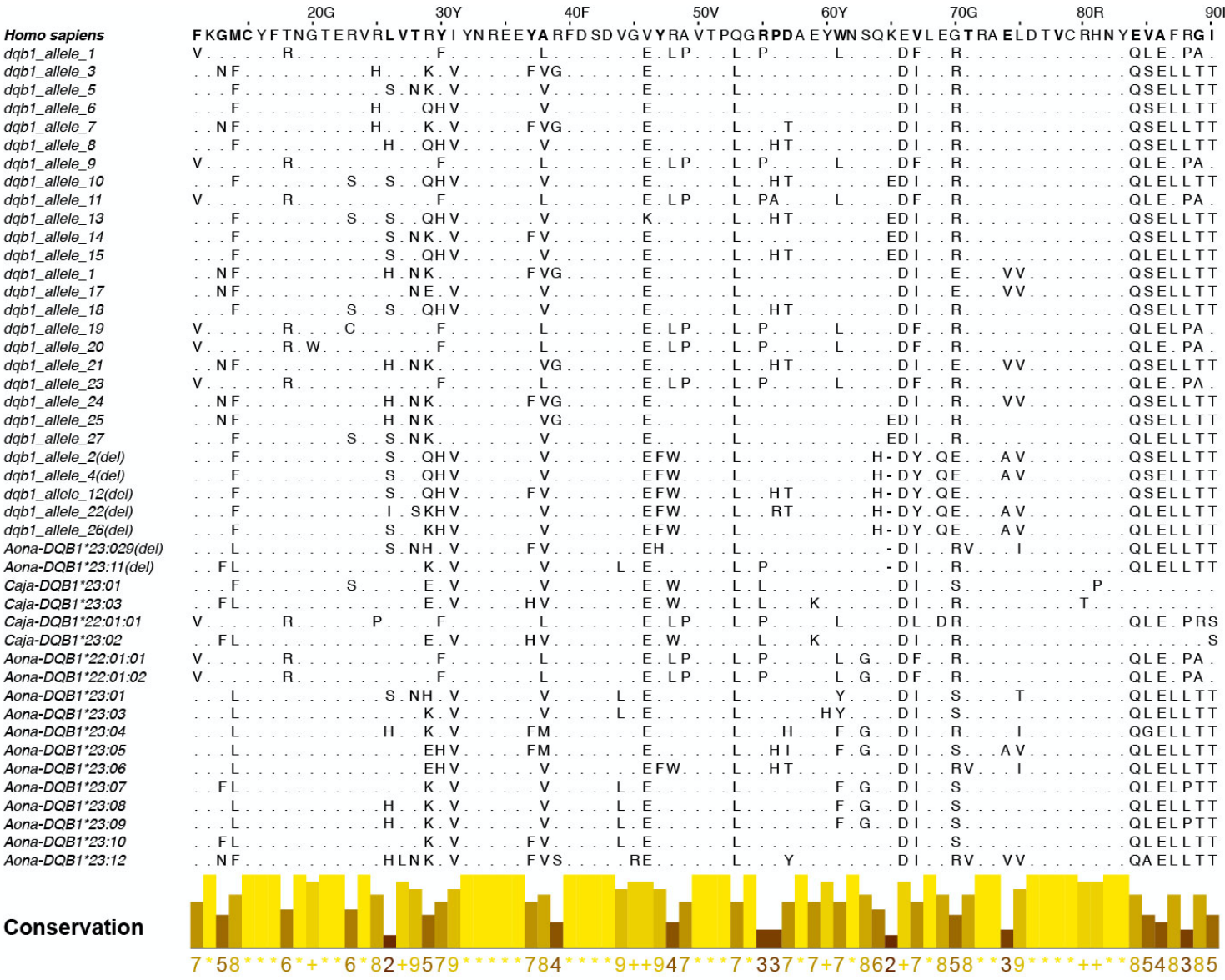
A. pigra	S-219	DQB2-3	DQB2-3
A. pigra	S-222	DQB2-3	DQB2-3
A. pigra	S-224	DQB2-3	DQB2-3
A. pigra	S-225	DQB2-2	DQB2-3
A. pigra	S-227	DQB2-3	DQB2-3
A. pigra	S-228	DQB2-2	DQB2-3
A. pigra	S-229	DQB2-3	DQB2-3
A. pigra	S-234	DQB2-3	DQB2-3
A. pigra	S-242	DQB2-3	DQB2-3
A. pigra	S-248	DQB2-2	DQB2-3
A. pigra	S-279	DQB2-3	DQB2-3
A. pigra	S-289	DQB2-3	DQB2-4
A. pigra	S-490	DQB2-2	DQB2-3
A. pigra	S-491	DQB2-3	DQB2-3
A. pigra	S-494	DQB2-3	DQB2-3
A. pigra	S-504	DQB2-3	DQB2-4
A. pigra	S-505	DQB2-3	DQB2-3
A. pigra	S-507	DQB2-4	DQB2-4
A. pigra	S-508	DQB2-4	DQB2-4
A. pigra	S-509	DQB2-4	DQB2-4
A. pigra	S-510	DQB2-3	DQB2-4
A. pigra	S-511	DQB2-3	DQB2-3
A. pigra	S-526	DQB2-3	DQB2-3
A. caraya	H1	DQB2-5	DQB2-6
A. caraya	21C	DQB2-5	DQB2-6
A. guariba	109	DQB2-7	DQB2-7
A. guariba	89	DQB2-8	DQB2-8

Figure 1-S4. Amino acid sequence alignments. Alignments include sequences from all four *Alouatta* species, *C. jacchus* and *A. nancymaae* for all three loci A) DQA1, B) DQB1 and C) DQB2. The reference alignment sequence is a human consensus sequence for each locus. Site numbers are indicated on top of the reference sequence based on the human sequence. ABS are indicated in bold in the reference sequence. An amino acid conservation bar plot is displayed below the alignment. Each bar corresponds to the amino acid above. The higher the values the more conserved the amino acid is across all sequences. Completely conserved sites are indicated with "*" and sites with mutations where all properties are conserved are indicated with "+".

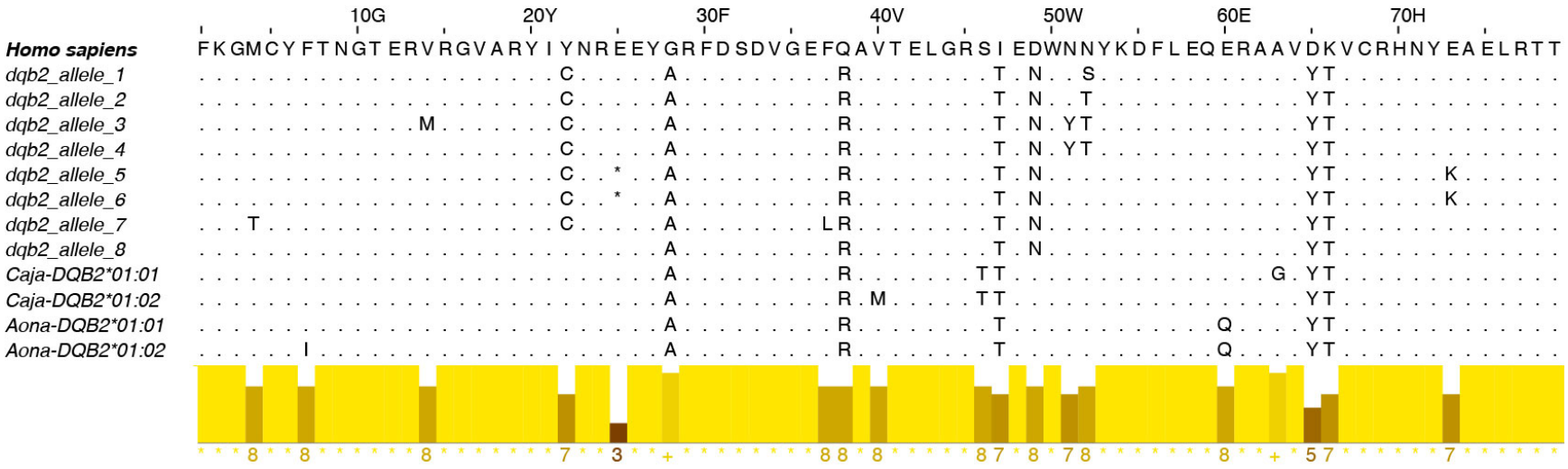
A)



B)


20G 30Y 40F 50V 60Y 70G 80R 90I
Homo sapiens FKGMCYFTNGTERVRLVTRYIYNREEYARFDSDVGVYRAVTPQGRPDAEYWNSQKEVLEGTRAELDTVCRHNYEVAFRGI
dqb1_allele_1
dqb1_allele_3
dqb1_allele_5
dqb1_allele_6
dqb1_allele_7
dqb1_allele_8
dqb1_allele_9
dqb1_allele_10
dqb1_allele_11
dqb1_allele_13
dqb1_allele_14
dqb1_allele_15
dqb1_allele_1
dqb1_allele_17
dqb1_allele_18
dqb1_allele_19
dqb1_allele_20
dqb1_allele_21
dqb1_allele_23
dqb1_allele_24
dqb1_allele_25
dqb1_allele_27
dqb1_allele_2(del)
dqb1_allele_4(del)
dqb1_allele_12(del)
dqb1_allele_22(del)
dqb1_allele_26(del)
Aona-DQB1*23:029(del)
Aona-DQB1*23:11(del)
Caja-DQB1*23:01
Caja-DQB1*23:03
Caja-DQB1*22:01:01
Caja-DQB1*23:02
Aona-DQB1*22:01:01
Aona-DQB1*22:01:02
Aona-DQB1*23:01
Aona-DQB1*23:03
Aona-DQB1*23:04
Aona-DQB1*23:05
Aona-DQB1*23:06
Aona-DQB1*23:07
Aona-DQB1*23:08
Aona-DQB1*23:09
Aona-DQB1*23:10
Aona-DQB1*23:12
Conservation
7*58***6*+**6*82+9579*****784****9++947***7*337*7+7*862+7*858**39****++**8548385

C)


10G
20Y
30F
40V
50W
60E
70H
Homo sapiens FKGMCYFTNGTERVRGVARYIYNREEYGRFDSDVGEFQAVTELGRSIEDWNNYKDFLEQERAAVDKVCRHNYEAELRTT
dqb2_allele_1
dqb2_allele_2
dqb2_allele_3
dqb2_allele_4
dqb2_allele_5
dqb2_allele_6
dqb2_allele_7
dqb2_allele_8
Caja-DQB2*01:01
Caja-DQB2*01:02
Aona-DQB2*01:01
Aona-DQB2*01:02

Chapter 2: Major Histocompatibility Complex (MHC) Variation Across the Distribution of the Mantled Howler Monkey (*Alouatta palliata*): Population Structure and Low Genetic Diversity in Mexican Populations

Abstract

Genetic variation is fundamental for populations to adapt and consequently persist in an ever-changing environment. Traditionally, studies that evaluate genetic variation in wild populations rely on neutral genetic markers, however, these markers do not convey any functional information. In this study we analyzed seven loci from the Major Histocompatibility Complex (MHC) (DQA1-L1, DQA1-L2, DQB1-L1, DQB1-L2, DQB1-L3, DQB2 and DRB) to evaluate the genetic diversity of a broadly distributed primate, the mantled howler monkey (*A. palliata*). MHC is a genomic region with a key role in adaptive immunity. Using high-throughput sequencing, we analyzed a total of 57 wild *A. palliata* individuals from Mexico (N=37), Costa Rica (N=10) and Peru (N=10). Individuals from Mexico show the lowest genetic diversity as they were all homozygotes at all seven MHC loci. Both, Costa Rica and Peru individuals were more genetically diverse overall regardless the substantially smaller sample size compared to Mexico. We also found populations structure based on F_{ST} and AMOVA analyses. F_{ST} values were high in all our comparisons: Mexico-Costa Rica (0.36), Mexico-Peru (0.85) and Costa Rica-Peru (0.51) but Mexico-Peru comparison was not significant. AMOVA results were also significant showing that most of the genetic variation (68%) is found among populations. Lastly, we found evidence for positive selection at DRB and DQB1-L3, based on

dN/dS ratios, while the rest of the loci appear to be evolving neutrally. MHC genetic variation across the geographic distribution of *A. palliata* appeared to be predominantly shaped by neutral processes, mirroring studies of genetic diversity at neutral markers. Moreover, the extremely low MHC genetic diversity found in *A. palliata*'s Mexican population is of concern, in terms of conservation, given the crucial function of these loci.

Introduction

The ability of a population to adapt to an ever-changing environment is fundamentally determined by its genetic diversity. In Eukaryotes, genetic diversity originates from mutations, gene flow and recombination, but it is the interaction of selective and neutral forces what defines genetic diversity over evolutionary time (Frankham et al. 2002). Genetic drift changes allele frequencies at random from generation to generation decreasing genetic diversity. Selective pressures may maintain, reduce, or increase genetic diversity depending on the type of selection operating (Allendorf. 1986). Estimates of genetic diversity are traditionally derived from neutral markers which inherently do not provide information on adaptive variation (Avise 2009). In contrast, highly variable functional genomic regions are ideal candidates to investigate the adaptive potential of populations. The Major Histocompatibility Complex (MHC) is a multi-gene family characterized by its elevated polymorphism and its crucial role in adaptive immunity encoding molecules that present antigens for pathogen recognition (Edwards and Hedrick 1998).

It has long been hypothesized that such a level of polymorphism is a consequence of pathogen-driven selection in which more genetic variants being present in a population enables the recognition and appropriate immune response to a broader spectrum of pathogens (Doherty and Zinkernagel 1975; Jeffery and Bangham 2000; Bernatchez and Landry 2003). However,

these host and pathogen interactions are not geographically homogeneous. When considering species or populations with broad distribution ranges, MHC genetic diversity and allelic composition may not be uniformly distributed across space. Recently, MHC and other functional genes have been proposed as more suitable genetic marker for conservation studies as they may be more informative when evaluating the potential adaptability of a population (Manlik et al. 2019, Teixeira & Huber 2020). The mantled howler monkey (*Alouatta palliata*) is currently listed as vulnerable according to the IUCN Red List (Cortés-Ortiz et al. 2020), including the subspecies *A. palliata mexicana*, which is restricted to Mexico and listed as endangered (Cuarón et al. 2020)

The mantled howler monkey's geographic range extends from the Tumbes region in northern Peru through the Pacific coast of northwestern South America and into Central America to the Honduras-Guatemala border, with a disjunct population in southeast Mexico (Cortés-Ortiz et al. 2015; Rylands et al. 2006). Genetic diversity studies of neutral markers on different populations of *A. palliata* indicate a reduction in genetic diversity from Panama to Mexico (Winkler et al. 2004, Ellsworth & Hoelzer 2006, Ruíz-García et al. 2007, Milton et al. 2009, Jasso-del-Toro et al. 2016, Melo-Carrillo et al. 2020).

Given the extensive distribution of the mantled howler monkey, we predict genetic differentiation among populations when examining MHC genetic regions. Studies have of MHC genetic diversity across the geographic distribution of a species suggest that the strength of selection, at a local level, results in genetic differentiation among different populations within a species distribution (Hendrickson et al. 2000, Miller et al. 2001, Ekblom et al 2007, Alcaide et al. 2008, Cammen et al. 2010, Herdegen et al. 2014, Li et al. 2016, Lei et al. 2016, Talarico et al. 2019). In fewer cases, neutral processes appear to be the main drivers of the geographical

distribution of MHC genetic variation (Cutrera et al. 2017, Gillingham et al. 2017, Andre et al. 2017). Studies on primates (Schad et al. 2004, Bonhomme et al. 2007, Zhang et al. 2018) have provided evidence of MHC genetic differentiation at the population level because of local host-pathogen interactions.

Most of the studies that evaluate MHC genetic diversity over space rely on a single MHC genetic marker. DRB exon 2 has been the standard MHC marker for evaluating genetic diversity due to its observed high variability but this is not the case in every system (Mainguy et al. 2007, Saka et al. 2018). In humans and non-human primates, DRB does appear to be more variable than the more traditionally conserved DQA and DQB (Bontrop et al. 2006). To evaluate the extent of variation at MHC class II loci, we analyzed DRB exon 2 (DRB) and six MHC-DQ loci: DQA1 exon 2 Locus 1 (DQA1-L1), DQA1 exon 2 Locus 2 (DQA1-L2), DQB1 exon 2 Locus 1 (DQB1-L1), DQB1 exon 2 Locus 2 (DQB1-L2), DQB1 exon 2 Locus 3 (DQB1-L3) and DQB2 exon 2 (DQB2e2). Our aim was to assess the diversity and geographical variation of MHC loci in *A. palliata* populations. We i) genotyped 58 wild individuals sampled from Mexico, Costa Rica, and Peru at the seven MHC loci, ii) estimated MHC genetic diversity in each population, iii) evaluated population structure, iv) examined the amino acid variation at antigen binding sites (ABS) and v) assessed selection comparing ratios of synonymous and non-synonymous substitutions.

Methods

Sampling, DNA extraction and amplification

We analyzed blood (N = 47) and fecal (N = 10) samples from a total of 57 wild *A. palliata* individuals from three main geographic locations: Mexico (blood samples from nine groups,

N=37), Costa Rica (blood samples from an undetermined number of groups in a single location, N=10 individuals) and in Peru (fecal samples from an undetermined number of groups in a single location, N=10 individuals) (Figure 2-1). Samples from Mexico were collected between 1998 and 2010. Samples from Costa Rica were kindly shared by Dr. K. Glader to the Cortés-Ortiz Lab. Samples from Peru were collected by Sergio Redondo and Fanny Cornejo. Fecal samples were obtained fresh from observable individuals immediately after defecation then collected in 15 ml tubes with 99% ethanol, refrigerated in the field at -20°C before being stored at -80°C in our lab facilities. Details on the collection of samples from Mexico and Costa Rica can be found in Cortés-Ortiz et al. (2003, 2019) and Kelaita et al. (2011).

We extracted DNA from blood samples using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA). For fecal samples, we used the QIAamp DNA Stool Mini Kit (Qiagen, Valencia, CA). We followed manufacturer's protocol for both DNA extraction kits. We quantified DNA concentration in a Qubit 4 Fluorometer and using Qubit dsDNA high-sensitivity assay (Invitrogen, Carlsbad, CA). All samples were standardized at a DNA concentration of 5 ng/µl. We performed PCRs using previously published primers targeting partial sequences of three exons: DRB exon 2 (DRB), DQA1 exon 2 (DQA1) and DQB1 (DQB1) exon 2 (Tiercy et al. 1990, Gyllensten & Elrich 1988 and Yasunaga et al. 1996). As previously reported (Diaz et al. 2000) the set of primers used for DQB1 also co-amplified DQB2. Moreover, duplications were detected in both DQA1 and DQB1, resulting in two and three distinct loci respectively (as explained in Chapter I). Here, we sequenced DRBe2 for the first time in *A. palliata* with no evidence of duplication. In total we used seven loci to assess MHC Class II diversity: DQA1-2L1, DQA1-L2, DQB1-L1, DQB1-L2, DQB1-L3, DQB2 and DRB.

Next Generation Sequencing Library Preparation

We followed the 16S protocol from Illumina (Illumina 2013) for next generation sequencing library preparation with modifications. Briefly, the protocol consists in a PCR to amplify the product of interest, and another PCR to attach labeling indexes to the sequences generated in the first PCR. The final products are pooled in equimolar DNA concentrations for sequencing. We modified the first stage PCR to a 3-phase Touch-Down with an initial denaturing temperature of 95 °C for 5 minutes, followed by two cycles of 94 °C for 30 sec, annealing temperature (55 °C for DQA1, 60 °C for DQB1 and DQB2, 68 °C for DRB) for 30 sec and 72 °C for 30 seconds (phase 1); two cycles of 94 °C for 30 sec, annealing temperature (50 °C for DQA1, 55 °C for DQB1 and DQB2, 63 °C for DRB) for 30 sec and 72 °C for 30 seconds (phase 2); 30 cycles of 94 °C for 30 sec, annealing temperature (48 °C for DQA1, 52 °C for DQB1 and DQB2, 58 °C for DRB) for 30 sec and 72 °C for 30 seconds (phase 3) and a final extension at 72 °C for 10 minutes. PCR total volume was 20 µl and included 5.5 µl of ultrapure water, 1 µl of forward primer at 10 µM, 1 µl of reverse primer at 10 µM, 10 µl of Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Ipswich, MA) and 2.5 µl of DNA at 5ng/µ. For the second (Index) PCR, our total reaction volume was 25 µl: 4 µl of ultrapure water, 3 µl of Nextera XT Index Primer 1 (Illumina, San Diego, CA), 3 µl of Nextera XT Index Primer 2 (Illumina, San Diego, CA), 12.5 µl of Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Ipswich, MA) and 2.5 µl of product from the first PCR. Samples were labeled with unique index combinations for individual identification.

Products of the indexed PCRs were quantified using fluorometric DNA quantification using the Qubit dsDNA high-sensitivity assay, diluted to a concentration of 4nM, and then pooled together. We amplified and sequenced all samples in duplicate. Duplicates were amplified in independent PCRs and sequenced on different sequencing runs. We submitted our

libraries for 250 bp pair-end MiSeq Nano 500 sequencing (Illumina, San Diego, CA) at The University of Michigan Advanced Genomics Core.

MHC genotyping

We filtered out raw sequences that had a phred score lower than 20 using Trimmomatic (Bolger et al. 2014). We merged pair-end sequences using FLASH (Magoc & Salzberg 2011) keeping default parameters and filtered sequences based on size (retaining fragments>130 bp) using PRINSEQ (Schmieder & Edwards 2011) to eliminate sequences that were not within the range of expected size.

We created an R function to determine preliminary genotypes for each individual at each locus. This function ranks all the unique reads of a particular locus by frequency in each individual, eliminates all singletons, trims primers, and calculates the relative frequency of each unique read to the total number of reads in the individual. Once we determined individual putative alleles, we identified and eliminated chimeric sequences manually by aligning all sequences of each locus and evaluating for the distribution of the variable sites in the aligned sequences. Chimeric sequences were defined as sequences that combined two parental sequences that were found in high frequency within an individual. All putative alleles that remained after removing chimeras were confirmed as MHC-DQ sequences comparing sequence identity to other published MHC-DQ sequences of closely related taxa. Following the protocol of Sommer et al. 2013 we determined the final genotypes. Briefly, to call true alleles we i) identified high frequency putative alleles in and individual, ii) corroborated allele presence in the duplicate, and iii) confirm that the frequency of each potential true allele sequence was higher than any known artifact (e.g., singletons, chimeras).

MHC genetic diversity and population structure

We genotyped each individual at each locus and calculated allele frequencies, genotype frequencies, private alleles and observed and expected heterozygosity for each locus in each population.

To analyze population structure, we grouped samples by geographical location (Mexico, Costa Rica and Peru) and included all loci except DRBe2 because we were unable to recover any sequences for Peruvian individuals for this locus. We used the R package Hierfstat (Goudet 2005) to calculate pairwise F_{ST} and a 95% bootstrap confidence interval to determine statistical significance. We used the program GenoDive (Meirmans 2020) to perform an analysis of molecular variance (AMOVA) and used 10,000 permutations to assess statistical significance.

Functional analyses

We deduced the amino acid sequences for all alleles across all loci and reported all variable sites within each locus using MEGAX (Stecher et al. 2020). We also aligned these sequences, using MUSCLE (Edgar 2004), to their human homologue sequence to infer potential antigen binding sites (ABS) in our sequences. ABS in Human Leukocyte Antigen (HLA) genes, the human version of the MHC genes, have been characterized in previous studies (Otting et al. 1992, Kwok et al. 1996, Reche & Reinherz 2003, Siebold et al. 2003, Cocco et al. 2012, Sarri et al. 2018). We did not assign ABS to DQB2e2 as HLA-DQB2 does not code for any portion of the antigen binding groove (Garrigan and Hendrick 2003). Nevertheless, we still assessed the nucleotide and amino acid variation at this locus within *A. palliata*.

We evaluated amino acid replacement at ABS in alleles at each locus. Amino acid changes can be radical or conservative depending on the properties of each amino acid.

Conservative changes tend to not alter protein product function while radical changes are likely to do so (Sneath 1966, Grantham 1974). We implemented the index of amino acid dissimilarities by Xia and Xie (2002) to determine conservative vs radical amino acid replacement. This index builds on widely used indexes (Grantham 1974 and Miyata et al. 1979) by incorporating neighbor preference into the model. The index amino acid dissimilarity scale spans from values 0.06 to 5.23, the lower the value the more conservative the amino acid replacement is. The median value for all possible 190 amino acid replacement is 2.365, we used this threshold to determine if an amino acid change is conservative or radical.

Selection analysis

Populations are subject to distinct selective pressures that may genetically differentiate them. To determine signatures of selection we used the overall average and pairwise Z-test of Selection per locus, as implemented in MEGAX (Stecher et al. 2020). The Z-test of selection estimates the number of synonymous substitutions per synonymous sites (d_S) and the number of non-synonymous substitutions per non-synonymous sites (d_N) and their respective variances between pairs of sequences and among all sequences. Based on these estimates it tests the null hypothesis of neutral evolution ($d_N=d_S$) versus an alternative hypothesis (positive selection -$d_N>d_S$- or purifying selection -$d_N<d_S$-). We used 1000 bootstrap replications to assess statistical significance. We included only DRB, DQA1-L1 and DQB1-L2 for this analysis. These loci had more than two different alleles across all 58 individuals, allowing for statistical significance.

Results

MHC sequencing and genotyping

We amplified and obtained sequences of exon 2 for seven MHC Class II genes: DQA1-L1, DQA1-L2, DQB1-L1, DQB1-L2, DQB1-L3, DQB2 and DRB. However, we were unable to obtain DRB sequences from the fecal samples from individuals from Peru. Individuals A-16 and G-5 from Peru presented missing data at three loci and were excluded from the analyses.

Comparing our sequences to human homologue sequences, and assuming identical sizes between human and *A. palliata* sequences, we estimated a coverage of the total exon 2 sequence of 91% for DRB, 75% for DQA1, 93% for DQB1 and 93% for DQB2. Our paired-end sequencing raw data yielded a total of 868,376 reads across all individuals. After filtering our data (based on quality and size), we recovered a total of 311,344 sequences that were used for the genotyping process (please refer to Supplementary Materials Table 2-S5 for a more detailed description of the filtering process and the number of reads).

Low MHC genetic diversity

For DRB (length = 245 bp), we found a total of four alleles. Individuals from Mexico were all homozygous for allele DRB-1. Individuals from Costa Rica presented all four alleles (DRB-1, DRB-2, DRB-3, DRB-4) combined in 5 genotypes. Alleles DRB-2, DRB-3 and DRB-4 are private alleles for Costa Rica.

For DQA1-L1 (length = 188 bp), we found a total of three alleles. All individuals from Mexico were homozygous for DQA1-L1-1. Individuals from Costa Rica showed two alleles (DQA1-L1-1, DQA1-L1-5) yielding two genotypes. Individuals from Peru harbor three alleles

(DQA1-L1-1, DQA1-L1-3) and two genotypes. DQA1-L1-5 and DQA1-L1-3 are private alleles for Costa Rica and for Peru respectively. For DQA1-L2 (length = 188 bp), we found a total of three alleles. All individuals from Mexico were homozygous for allele DQA1-L2-2. Individuals from Costa Rica presented two alleles (DQA1-L2-2, DQA1-L2-4) and three genotypes. Individuals from Peru were all homozygous for allele DQA1-L2-2. Allele DQA1-L2-4 is private in Costa Rica.

For DQB1-L1 (length = 250 bp), we found only one allele across all individuals from Mexico, Costa Rica and Peru. For DQB1-L2 (length = 247 bp), we found a total of two alleles. All individuals from Mexico were homozygous for DQB1-L2 -2. Individuals from Costa Rica presented two alleles (DQB1-L2 -2, DQB1-L2 -4) yielding three different genotypes. All individuals from Peru were homozygous for DQB1-L2 -2. DQB1-L2 -4 was a private allele in Costa Rica. For DQB1-L3 (length = 250 bp), we found a total of four different alleles. Individuals from Mexico were all homozygous for DQB1-L3-3. Individuals from Costa Rica presented three alleles (DQB1-L3-3, DQB1-L3-5 and DQB1-L3-7) and four different genotypes. Individuals from Peru were all homozygous for DQB1-L3-6. Both DQB1-L3-5 and DQB1-L3-7 are private alleles in Costa Rica and DQB1-L3-6 is a private allele in Peru.

For DQB2 (length = 247 bp), we found a total of two alleles. All individuals from Mexico were homozygous for DQB2-1. Individuals from Costa Rica showed two alleles (DQB2-1, DQB2-2) yielding three different genotypes. All Individuals from Peru harbor two alleles (DQB2-1, DQB2-2) for a total of two different genotypes. All individual genotypes, allele frequencies and genotype frequencies for the different populations are reported in Supplementary Materials Table 2-S6, 2-S7 and 2-S8, respectively.

In terms of heterozygosity, individuals from Mexico are homozygotes at all seven alleles. Individuals from Peru are only heterozygotes at DQA1-L1 and DQB2. For Costa Rica there was at least one heterozygote individual at each locus, except for DQB1-L1 which appears to be very conserved across all geographic regions. Costa Rica presented private alleles at all loci except for DQB1-L1 and DQB2. Peru presented private alleles at locus DQA1-L1 and DQB1-L3. Mexico did not have private alleles at any locus. Genetic diversity results are summarized in Table 2-1.

Signatures of population structure

Both pairwise F_{ST} analysis and AMOVA detected signatures of population structure when comparing the three geographic regions. F_{ST} values were high in all our comparisons: Mexico-Costa Rica (0.36), Mexico-Peru (.85) and Costa Rica-Peru (0.51) but they were not statistically significant for the Mexico-Peru comparison. AMOVA results showed that most of the genetic variation (68%) is found among populations, while 28.8% of the genetic variation is found within populations (Table 2-2).

Amino acid sequence diversity

Our assessment of the variability in DRB nucleotide and amino acid sequences across all individuals revealed 34 variable sites in the nucleotide sequence (34bp/245bp) that translated to 23 variable sites in the amino acid sequence (23aa/81aa) in DRBe2. Both DQA1-L1 and DQA1-L2 presented two variable sites in the nucleotide sequence (2bp/186bp) but DQA1-L1 had only one variable site in the amino acid sequence (1aa/62aa) while DQA1-L2 had two variable sites (2aa/62aa). We only found one allele for DQB1-L1 across all individuals. DQB1-L2 presented one variable site in the nucleotide sequence (1bp/247bp) and no variable sites in the amino acid

sequence (0aa/82) (Table 2-3). DQB1-L3 had 15 variable sites in the nucleotide sequence (15bp/250bp) and nine variable sites in the amino acid sequence (9aa/83aa). DQB2 presented one variable site in the nucleotide sequence (1bp/247bp) which produced an amino acid change at one site (1aa/82) (Table 2-3).

We aligned our amino acid sequences to their respective human homologue with known ABSs and identified 24 ABS for the *A. palliata* DRB fragment sequenced in this study. 15 ABS were variable in *A. palliata*, one was conserved in *A. palliata* but different from the human sequence, and eight were conserved in both humans and howler monkeys. In both, DQA1-L1 and DQA1-L2 sequences we identified 19 ABS. For DQA1-L1 one ABS was variable in *A. palliata*, five were conserved in *A. palliata* but distinct from the human sequence, and 13 were conserved in both humans and howler monkeys. For DQA1-L2 two ABS were variable in *A. palliata*, three were conserved in *A. palliata* but distinct from the human sequence, and 14 were conserved in both humans and howler monkeys. In both DQB1-L1 and DQB1-L3 we located 25 ABS. For DQB1-L1 eight ABS differ between *A. palliata* and the human homologue. There was no ABS variability within *A. palliata* because only one allele was present in all individuals for this locus. For DQB1-L3 six ABS were variable in *A. palliata*, seven were conserved in *A. palliata* but distinct from the human sequence, and 12 were conserved in both humans and howler monkeys (Table 2-3). We did not compare homologue human sequences to DQB2 as it does not code for any portion of the antigen binding groove.

We evaluated whether amino acid changes at ABS across *A. palliata* alleles at each locus were radical or conservative. In DRB, we found three radical amino acid changes at positions 11, 56 and 61. No radical amino acid replacements were found for DQA1-L1 and DQA1-L2. For DQB1-L3 we found one radical change at position 56. For reference, we also evaluated the

amino acid changes between *A. palliata* and humans. In DRB, we found five radical amino acid changes at positions 9, 11, 13, 56 and 61. In both DQA1-L1 and DQA-L2, we found one radical amino acid replacement at position 52. In DQB1-L1, we found two radical changes at position 55 and 61. In DQB1-L3, we found three radical changes at position 57, 89 and 90. Amino acid sequence alignment and indexes values can be found in Figure 2-2.

Positive selection in DRB and DQB1-L3

We found evidence of positive selection in DRB and DQB1-L3. When implementing the Z-test of Selection, we were unable to reject the neutral evolution hypothesis under the scenarios of positive selection or purifying selection for DQA1-L1 in the overall average and in one pairwise comparison. For DQB1-L3 we found evidence of positive selection in the pairwise comparisons for DQB1-L3-5 and DQB1-L3-3 ($p<0.05$) and for DQB1-L3-5 and DQB1-L3-7 ($p<0.05$). Neutrality could not be rejected on the overall average test, but positive selection was close to reach significance ($p=0.07$). In DRB, we found that allele DRB-3 was under positive selection when compared to the rest of the alleles, and the neutral hypothesis was rejected under the scenario of positive selection ($p<0.05$) in the overall average. (Table 2-4).

Discussion

The diversity of MHC Class II loci in *A. palliata* varies across populations throughout its distribution range. Measures of genetic diversity that are based on functional loci can provide a better understanding of a population's extinction risk than using neutral loci (Manlik et al. 2019, Teixeira & Huber 2020). We analyzed samples of 57 individuals from three geographical locations: Mexico, Costa Rica and Peru. High levels of genetic diversity are one of the hallmarks

of MHC, however, genetic diversity results were remarkably low for the species in general but more so for individuals from Mexico.

Why is MHC genetic diversity so low in A. palliata?

The population in Mexico had the lowest level of genetic diversity despite being the most thoroughly sampled geographic region. All Mexican individuals were homozygotes at all seven loci. Costa Rica and Peru were more genetically diverse. Overall Costa Rica was the most genetically diverse geographic region, although Peru had the highest observed heterozygosity for DQA1-L1 and DQB2 (Table 2-1). DRB is considered the most polymorphic locus from MHC Class II genes and studies of Neotropical primates with similar number of individuals to those analyzed in this study have found significantly more alleles than those found in this study. A study on *Alouatta pigra*, a sister taxon to *A. palliata*, found 21 different alleles in 49 individuals sampled in Mexico (Arguello-Sanchéz, 2018). That level of diversity was already considered low compared to other platyrrhine species. Studies on two other platyrrhine primates, Nancy ma's night monkey (*Aotus nancymaae)* and the common marmoset (*Callithrix jacchus)* reported 34 and 21 alleles, respectively (Antunes et al. 1998, Gyllensten et al. 1994). Again, both studies analyzed samples from a smaller number of individuals than this study.

We believe that *A. palliata* populations that occur at lower latitudes than Mexico may have comparable levels of genetic diversity at DRB locus to other platyrrhines. In this study we were limited by the number of individuals for the DRB locus that were not from Mexico (N=13). Analysis of genetic diversity at neutral loci of non-Mexican *A. palliata* population (Ellsworth and Hoelzer 2006, Ruiz-García et al 2007, Milton et al. 2009) reveal similar levels of neutral genetic diversity as other Neotropical primate species such as *C. jacchus* (Nievergelt et al 2000,

Malukiewicz et al 2015), *Aotus azarai* (Babb et al. 2011), *Alouatta belzebul* (Bastos et al. 2010) and *Alouatta caraya* (Oklander et al. 2017). Populations of *A. palliata* that occur in Mexico have extremely low levels of genetic diversity at neutral loci (Melo-Carrilo et al. 2020) that are also seen in this study for MHC genes.

A recent comparative genomic study across 240 eutherian mammal species (Zoonomia Consortium 2020) included an *A. palliata* individual from Mexico. This study also found extremely high levels of homozygosity across the entire genome of this individual when compared genomes of individuals of other species. It was ranked 7th for the SoH (segments of homozygosity) metric from the 240 species. The evolutionary history of *A. palliata* infers an origin in South America and subsequent migration waves that resulted in the colonization of Central America and Mexico (Cortés-Ortiz et al., 2003; Ford, 2006). Mexican populations stand at the forefront of this demographic expansion that likely resulted in a founder event that decreased the genetic variability in the northernmost distribution of the species.

Only a few studies have analyzed genes DQA1, DQB1 and DQB2 genes in platyrrhines. One study for *C. jacchus* found two DQA1 alleles, four DQB1 alleles and two DQB2 alleles in 25 individuals (Antunes et al. 1998). A more recent study by Otting et al. 2020 found eight alleles for DQA1 and three alleles for DQB1 in 24 individuals in *C. jacchus*. This study also analyzed *Aotus lemurinus* and *Saguinus oedipus*. In *A. lemurinus* they found three DQA1 alleles and one DQB1 allele across nine individuals. In *S. oedipus* they found five DQA1 alleles in 13 individuals. Diaz et al. 2000 found five alleles for DQA1, 14 alleles for DQB1 and two alleles for DQB2 in 19 *A. nancymaae* individuals. Our results are similar to the findings in other studies when comparing diversity of DQA1, DQB1 and DQB2 genes. These genes are traditionally considered less diverse than DRB genes. Moreover, our results from amino acid sequence

demonstrate that these genes are more conserved than DRB sequences, even when compared with human homologue sequences.

Selection operates differently at each locus

We found evidence for selection at DRB1 and DQB1-L3. The rest of the loci (DQA1-L1, DQA-L2, DQB1-L1 and DQB2) were fairly conserved across all individuals and no signatures of selection were detected. DRB showed signatures of selection when looking at amino acid replacements and based on the Z-test of Selection. DRB showed the most variation in amino acid sequences in comparison to the rest of the other loci. DRB also presented the most radical changes within *A. palliata*, specifically allele DRB-3 presented radical amino acid replacements at three ABS. DRB-3 also showed signatures of positive selection when compared to the rest of the alleles in the Z-test. DRB-3 allele was private to Costa Rica and was also the most abundant allele in that population.

DQB1-L3 indicated signatures of positive selection in DQB1-L3-5, although neutrality could not be rejected in an overall average. DQB1-L3 also presented a radical amino acid change at a ABS, specifically DQB1-L3-7. This allele was in high frequency in individuals from Costa Rica and was also a private allele for that geographical region. These results may be indicative of local adaptation at these two loci. However, we cannot confirm the adaptive value of these results as we do not possess information on the specific function of each allele. Studies that evaluate the immune function of each allele and the prevalence of local pathogens could determine the adaptability that each allele may confer.

Population structure is mostly driven by neutral processes

Both F_{ST} and AMOVA results support genetic differentiation between the different geographic regions. Other studies on primate wild populations have found population structure at MHC loci when evaluating their distribution (Schad et al. 2004, Bonhomme et al. 2007, Zhang et al. 2018). These studies have found that selection at a local scale is the main factor for genetically differentiating populations. However, we consider that drift maybe outweighing selection in genetically differentiating the populations when considering all MHC class II loci.

From all the loci (N=6) that were included for these analyses only one (DQB1-L3) presented signatures of positive selection. The rest of the loci appear to be evolving neutrally as neutrality could not be statistically rejected as a explaining hypothesis. Most of the amino acid replacements were not radical and should not significantly change the final protein product. As mentioned earlier, all Mexican individuals were homozygous at all loci. Previous studies have shown low levels of genetic diversity when looking at neutral genetic markers in this geographic region concluding that founder events through its evolutionary history have shaped the current genetic diversity (Ellsworth & Hoelzer 2006, Dunn et al. 2013, Jasso-del Toro et al. 2016, Melo Carrillo et al. 2020). Taken together we believe that neutral processes are likely to be outweighing selective forces in genetically differentiating *A. palliata* populations.

Conservation implications of low MHC genetic diversity

It is difficult to draw conclusions for the species based on our small sampling outside of Mexico. However, it is important to acknowledge the severity of the Mexican populations' situation. *A. palliata* individuals in Mexico are listed as an endangered subspecies *Alouatta palliata mexicana*. Different studies have already determined very low levels of genetic diversity in

Mexican populations of *A. palliata*, based on genetic studies of neutral markers (Ellsworth & Hoelzer 2006, Dunn et al. 2013, Jasso-del Toro et al. 2016, Melo Carrillo et al. 2020). Here, we are also reporting very low levels of genetic diversity at a functional and highly polymorphic genomic region. Low levels of functional genetic diversity indicate a restricted capacity to adapt to environmental changes. Novel pathogens are particularly a big threat for the species. This is not new for the genus as brown howler monkey (*Alouatta guariba*) and black and gold howler monkey (*Alouatta caraya*) experienced significant population size reduction due to yellow fever from 2007 to 2009 (Holzman et al 2010). If a similar outbreak occurred the populations of *A. palliata* could be severely affected by it, particularly in Mexico. Low MHC genetic diversity compromises a proper immune response as pathogens can go undetected and successfully infect and individual.

Concluding remarks

We assessed the distribution and composition MHC Class II genetic diversity in *A. palliata*. We analyzed seven distinct MHC Class II loci and examined 58 individuals pertaining to three distinct geographic locations (Mexico, Costa Rica and Peru). We found overall low levels of genetic diversity across the species with all Mexican individuals (N= 38) being homozygous at all seven loci. The three populations are genetically differentiated based on F_{ST} and AMOVA analyses. We found evidence of selection acting at DRB and DQB1-L3. Our small sampling outside of Mexico may be preventing us from capturing more genetic variation. However, the evolutionary history of this species undergoing founder events as its range expanded into Mexico have importantly reduce the genetic diversity in this region. Low levels of functional genetic diversity signify a reduced ability to adapt to new selective pressures ultimately threatening the

long-term persistence of a population. Mexican populations are at a greater risk of extinction from all the populations considered in this study.

References

Alcaide M, Edwards S V., Negro JJ, Serrano D, Tella JL (2008) Extensive polymorphism and geographical variation at a positively selected MHC class II B gene of the lesser kestrel (Falco naumanni). *Molecular Ecology*, **17**, 2652–2665.

Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity. Zoo Biology, 5, 181–190.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

André A, Millien V, Galan M, Ribas A, Michaux JR (2017) Effects of parasite and historic driven selection on the diversity and structure of a MHC-II gene in a small mammal species (Peromyscus leucopus) undergoing range expansion. *Evolutionary Ecology*, **31**, 785–801.

Antunes SG, De Groot NG, Brok H *et al.* (1998) The common marmoset: A new world primate species with limited Mhc class II variability. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 11745–11750.

Arguello-Sánchez LE, Arguello JR, García-Feria LM *et al.* (2018) MHC class II DRB variability in wild black howler monkeys (*Alouatta pigra*), an endangered new world primate. *Animal Biodiversity and Conservation*, **41**, 389–404.

Avise JC (2009) Phylogeography: Retrospect and prospect. *Journal of Biogeography*, **36**, 3–15.

Babb PL, Mcintosh AM, Fernandez-Duque E, Fiore A Di, Schurr TG (2011) An Optimized Microsatellite Genotyping Strategy for Assessing Genetic Identity and Kinship in Azara's Owl Monkeys (Aotus azarai). *Folia Primatol*, **82**, 107–117.

Bastos HB, Gonçalves EC, Ferrari SF, Silva A, Schneider MPC (2010) Genetic structure of red-handed howler monkey populations in the fragmented landscape of eastern Brazilian Amazonia. Genetics and Molecular Biology, 33, 774–780.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363–377.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bonhomme M, Blancher A, Jalil MF, Crouau-Roy B (2007) Factors shaping genetic variation in the MHC of natural non-human primate populations. *Tissue Antigens*, **70**, 398–411.

Bontrop RE (2006) Comparative Genetics of MHC Polymorphisms in Different Primate Species: Duplications and Deletions. *Human Immunology*, **67**, 388–397.

Bontrop RE, Otting N, de Groot NG, Doxiadis GGM (1999) Major histocompatibility complex class II polymorphisms in primates. *Immunological Reviews*, **167**, 339–350.

Cammen K, Hoffman JI, Knapp LA, Harwood J, Amos W (2011) Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (Halichoerus grypus). *Molecular Ecology*, **20**, 740–752.

Cocco E, Sardu C, Pieroni E *et al.* (2012) HLA-DRB1-DQB1 haplotypes confer susceptibility and resistance to multiple sclerosis in sardinia. *PLoS ONE*, **7**.

Cortés-Ortiz L, Bermingham E, Rico C, Rodriguez-Luna E, Sampaio I, Ruiz-García M (2003) Molecular systematics and biogeography of the genus, *Alouatta*. *Molecular Phylogenetics and Evolution*, **26**, 64–81.

Cortes-Ortíz, L., Rosales-Meda, M., Williams-Guillén, K., Solano-Rojas, D., Méndez-Carvajal, P.G., de la Torre, S., Moscoso, P., Rodríguez, V., Palacios, E., Canales-Espinosa, D., Link, A., Guzman-Caro, D. & Cornejo, F.M. 2021. *Alouatta palliata* (amended version of 2020

assessment). *The IUCN Red List of Threatened Species* 2021: e.T39960A190425583.
Cortés-Ortiz, L., Rylands, A. B., & Mittermeier, R. A. (2015). The taxonomy of howler monkeys: Integrating old and new knowledge from morphological and genetic studies. In M. Kowalewski, P. A. Garber, L. Cortés-Ortiz, B. Urbani & D. Youlatos (Eds.), Howler monkeys: Adaptive radiation, systematics, and morphology (pp. 55–84). New York, NY: Springer.
Cutrera AP, Mora MS (2017) Selection on MHC in a context of historical demographic change in 2 closely distributed species of tuco-Tucos (ctenomys australis and C. Talarum). *Journal of Heredity*, **108**, 628–639.
Cuarón, A.D., Shedden, A., Rodriguez-Luna, E., de Grammont, P.C. & Link, A. 2020. *Alouatta palliata* ssp. mexicana. The IUCN Red List of Threatened Species 2020: e.T925A17978896.
Diaz D, Naegeli M, Rodriguez R *et al.* (2000) Sequence and diversity of MHC DQA and DQB genes of the owl monkey *Aotus nancymaae*. *Immunogenetics*, **51**, 528–537.
Doherty P, Zinkernagel R (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*.
Dunn JC, Shedden-González A, Cristóbal-Azkarate J *et al.* (2014) Limited genetic diversity in the critically endangered Mexican howler monkey (*Alouatta palliata* mexicana) in the Selva Zoque, Mexico. *Primates*, **55**, 155–160.
Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
Edwards S V, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics to sexual selection. *TREE*, **13**, 305–311.
Ekblom R, Sæther SA, Jacobsson P *et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (Gallinago media). *Molecular Ecology*, **16**, 1439–1451.
Ellsworth, J. A., & Hoelzer, G. A. (2006). Genetic evidence on the historical biogeography of Central American Howler monkeys. In S. M. Lehman & J. G. Fleagle (Eds.), Primate biogeography: Progress and prospects (pp. 81–103).
Ford, S. M. (2006). The biogeographic history of Mesoamerican primates. In A. Estrada, P. A. Garber, M. Pavelka & L. Luecke (Eds.), New perspectives in the study of Mesoamerican primates (pp. 81–114). New York, NY: Springer.
Frankham, R., Briscoe, D. A., & Ballou, J. D. (2002). Introduction to conservation genetics, Cambridge: Cambridge University Press.
Genereux DP, Serres A, Armstrong J *et al.* (2020) A comparative genomics multitool for scientific discovery and conservation. *Nature*, **587**, 240–245.
Gillingham MAF, Béchet A, Courtiol A *et al.* (2017) Very high MHC Class IIB diversity without spatial differentiation in the mediterranean population of greater Flamingos. *BMC Evolutionary Biology*, **17**, 1–11.
Goudet, J. (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. Molecular Ecology Notes. **5**, 184-186.
Gyllensten U, Bergström T, Josefsson A *et al.* (1994) The cotton-top tamarin revisited: Mhc class I polymorphism of wild tamarins, and polymorphism and allelic diversity of the class II DQA1, DQB1, and DRB loci. *Immunogenetics*, **40**, 167–176.
Gyllensten UB, Erlich H a (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 7652–6.

Hendrickson SL, Mayer GC, Wallen EP, Quigley K (2000) Genetic variability and geographic structure of three subspecies of tigers (Panthera tigris) based on MHC class I variation. *Animal Conservation*, **3**, 135–143.

Herdegen M, Babik W, Radwan J (2014) Selective pressures on MHC class II genes in the guppy (Poecilia reticulata) as inferred by hierarchical analysis of population structure. *Journal of Evolutionary Biology*, **27**, 2347–2359.

Illumina (2013) *16S Metagenomic Sequencing Library Preparation*. https://web.uri.edu/gsc/files/16s-metagenomic-library-prep-guide-15044223-b.pdf

IPD-MHC Database. https://www.ebi.ac.uk/ipd/mhc/

IUCN/SSC Primate Specialist Group http://www.primate-sg.org/

Jasso-del Toro C, Márquez-Valdelamar L, Mondragón-Ceballos R (2016) Diversidad genética en grupos de monos aulladores de manto (*Alouatta palliata* mexicana) en la Reserva de la Biosfera Los Tuxtlas (Veracruz, México). *Revista Mexicana de Biodiversidad*, **87**, 1069–1079.

Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? *Microbes and Infection*, **2**, 1335–1341.

Kelaita M, Dias PAD, Aguilar-Cucurachi MS, Canales-Espinosa D, Cortés-Ortiz L (2011) Impact of Intrasexual Selection on Sexual Dimorphism and Testes Size in the Mexican Howler Monkeys *Alouatta palliata* and A . *pigra*. *American Journal of Physical Anthropology*, **187**, 179–187.

Klein J, Bontrop RE, Dawkins RL *et al.* (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics*, **31**, 217–219.

Kwok BW, Domeier ME, Johnson ML, Nepom GT, Koelle DM (1996) *HLA-DQB1 Codon 57 Is Critical for Peptide Binding and Recognition*.

Lenormand C, Hanau D, De La Salle H *et al.* (2012) Molecule Cells and Encode a New HLA Class II Langerhans Specifically Expressed in Human Genes Are HLA-DQB2 and HLA-DQA2. *The Journal of Animal Ecology*, **188**, 3903–3911.

Lei W, Fang W, Zhou X, Lin Q, Chen X (2016) Population genetic diversity and geographical differentiation of MHC class II DAB genes in the vulnerable Chinese egret (Egretta eulophotes). *Conservation Genetics*, **17**, 1459–1468.

Li J, Shen H, Wang H *et al.* (2016) Diversifying selection is the dominant factor affecting the geographical variation of MHC class II genes in the Omei tree frog. *Journal of Zoology*, **300**, 197–204.

Magoc T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.

Mainguy J, Worley K, Coté SD, Coltman DW (2007) Low MHC DRB class II diversity in the mountain goat: past bottlenecks and possible role of pathogens and parasites. *Conservation Genetics*, **8**, 885–891.

Manlik O, Krützen M, Kopps AM *et al.* (2019) Is MHC diversity a better marker for conservation than neutral genetic diversity? A case study of two contrasting dolphin populations. *Ecology and Evolution*, **9**, 6986–6998.

Meirmans, P.G. (2020), GENODIVE version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids, Molecular Ecology Resources DOI: 10.1111/1755-0998.13145.

Malukiewicz J, Boere V, Fuzessy LF *et al.* (2015) Natural and Anthropogenic Hybridization in Two Species of Eastern Brazilian Marmosets (*Callithrix jacchus* and C. penicillata) (P Prentis, Ed,). *PLOS ONE*, **10**, e0127268.

Melo-Carrillo A, Dunn JC, Cortés-Ortiz L (2020) Low genetic diversity and limited genetic structure across the range of the critically endangered Mexican howler monkey (*Alouatta palliata* mexicana). *American Journal of Primatology*, **82**, 1–10.

Miller KM, Kaukinen KH, Beacham TD, Withler RE (2001) *Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon*.

Milton, K., Lozier, J. D., & Lacey, E. A. (2009). Genetic structure of an isolated population of mantled howler monkeys (*Alouatta palliata*)on Barro Colorado Island, Panama. Conservation Genetics, 10(2), 347–358.

Nievergelt CM, Digby LJ, Ramakrishnan U, Woodruff DS (2000) *Genetic Analysis of Group Composition and Breeding System in a Wild Common Marmoset (Callithrix jacchus) Population*.

Oklander LI, Kowalewski MM, Corach D (2010) Genetic Consequences of Habitat Fragmentation in Black-and-Gold Howler (*Alouatta* caraya) Populations from Northern Argentina. *International Journal of Primatology*, **31**, 813–832.

Otting BN, Kenter M, Van Weeren P, Jonker M (1992) Mhc-DQB repertoire variation in hominoid and Old World primate species. *The Journal of Immunology*, **149**, 461–470.

Otting N, de Groot NG, Bontrop RE (2020) Full-length MHC class II alleles in three New World monkey species. *Hla*, **95**, 163–165.

Reche PA, Reinherz EL (2003) Sequence Variability Analysis of Human Class I and Class II MHC Molecules: Functional and Structural Correlates of Amino Acid Polymorphisms. *Journal of Molecular Biology*, **331**, 623–641.

Ruiz-García, M., Escobar-Armel, P., Alvarez, D., Mudry, M., Ascunce, M., Gutiérrez-Espeleta, G., & Shostell, J. M. (2007). Genetic variability in four *Alouatta* species measured by means of nine DNA microsatellite markers: Genetic structure and recent bottlenecks Folia Primatologica, 78(2), 73–87.

Rylands, A. B., Groves, C. P., Mittermeier, R. A., Cortés-Ortiz, L., & Hines, H. J. J. (2006). Taxonomy and distributions of Mesoamerican primates. In A. Estrada, P. Garber, M. Pavelka & L. Luecke (Eds.), New perspectives in the study of Mesoamerican primates: distribution, ecology, behavior and conservation (pp. 29–79). New York, NY: Springer.

Saka T, Nishita Y, Masuda R (2018) Low genetic variation in the MHC class II DRB gene and MHC-linked microsatellites in endangered island populations of the leopard cat (Prionailurus bengalensis) in Japan. *Immunogenetics*, **70**, 115–124.

Sarri CA, Papadopoulos GE, Papa A *et al.* (2018) Amino acid signatures in the HLA class II peptide-binding region associated with protection/susceptibility to the severe West Nile Virus disease. *PLoS ONE*, **13**, 1–20.

Schad J, Sommer S, Ganzhorn JU (2004) MHC variability of a small lemur in the littoral forest fragments of southeastern Madagascar. *Conservation Genetics*, **5**, 299–309.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**, 863–864.

Siebold C, Hansen BE, Wyer JR *et al.* (2003) *Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy*.

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.
Stecher G, Tamura K, Kumar S (2020) Molecular Evolutionary Genetics Analysis (MEGA) for macOS (C Russo, Ed,). *Molecular Biology and Evolution*, **37**, 1237–1239.
Talarico L, Babik W, Marta S, Pietrocini V, Mattoccia M (2019) MHC structuring and divergent allele advantage in a urodele amphibian: a hierarchical multi-scale approach. *Heredity*, **123**, 593–607.
Teixeira JC, Huber CD (2021) The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences*, **118**, 1–10.
Winkler, L. A., Zhang, X., Ferrell, R., Wagner, R., Dahl, J., Peter, G., & Sohn, R. (2004). Geographic microsatellite variability in Central American howling monkeys. International Journal of Primatology, 25(1), 197–210.
Yasunaga S, Kimura A, Hamaguchi K, Rønningen KS, Sasazuki T (1996) Different contribution of HLA-DR and -DQ genes in susceptibility and resistance to insulin-dependent diabetes mellitus (IDDM). *Tissue Antigens*, **47**, 37-48.
Yu, L. P., and M. J. Sheehy. 1991. The cryptic HLA-DQA2 ("DX alpha") gene is expressed in human B cell lines. J. Immunol. **147:** 4393–4397.
Zhang P, Huang K, Zhang B *et al.* (2018) High polymorphism in MHC-DRB genes in golden snub-nosed monkeys reveals balancing selection in small, isolated populations. *BMC Evolutionary Biology*, **18**.

Tables

Table 2-1. Genetic diversity parameters per geographic region. Number of individuals (N), number of alleles (NA), number of genotypes (NG), private alleles (PA), observed heterozygosity (HO) and expected heterozygosity (HE). These values are presented per locus and for each geographic region. NA, NG, HO and HE are also presented across all individuals in the Total column.

	Mexico						Costa Rica						Peru						Total				
	N	NA	NG	PA	HO	HE	N	NA	NG	PA	HO	HE	N	NA	NG	PA	HO	HE	N	NA	NG	HO	HE
DRBe2	38	1	1	0	0	0	10	4	5	3	0.200	0.767	0	NA	NA	NA	NA	NA	48	4	6	0.100	0.370
DQA1e2_L1	38	1	1	0	0	0	10	2	2	1	0.300	0.267	9	2	2	1	0.667	0.500	57	3	5	0.322	0.246
DQA1e2_L2	38	1	1	0	0	0	10	2	3	1	0.500	0.522	3	1	1	0	0	0	51	3	2	0.167	0.180
DQB1e2_L1	38	1	1	0	0	0	10	1	1	0	0	0	8	1	1	0	0	0	56	1	1	0	0
DQB1e2_L2	38	1	1	0	0	0	10	3	4	2	0.400	0.622	8	1	1	1	0	0	56	4	5	0.133	0.206
DQB2e2_L3	38	1	1	0	0	0	10	2	3	1	0.500	0.522	9	1	1	0	0	0	57	2	3	0.167	0.172
DQB2e2	38	1	1	0	0	0	10	2	3	0	0.200	0.344	9	2	2	0	0.500	0.392	57	2	3	0.233	0.242

Table 2-2. AMOVA and pairwise FST results for the three different geographic locations: Mexico, Costa Rica (C.R.) and Peru. Asterisks on F_{ST} values indicate statistical significance based on a 95% confidence interval bootstrap and *ns* indicates no statistical significance.

	Percentage variation			
AMOVA	**Among populations**	**Within populations**	***p* value**	**Pairwise F_{ST}**
Mexico-C.R.-Peru	68.00%	28.80%	0.0001	Mexico-C.R. = 0.3640* Mexico-Peru = 0.8458^{ns} C.R.-Peru = 0.5087*

Table 2-3. Conservation and general information across all MHC loci. Nucleotide sequence length (SL (bp)), number of variable sites within nucleotide sequence (VS (bp)), amino acid sequence length (SL (aa)), number of variable sites within amino acid sequence (VS (aa)), antigen binding sites within each exon (ABS), number of antigen binding sites that are variable among *A. palliata* individuals (ABS VS), number of antigen binding sites that are conserved in *A. palliata* (ABS CSA) and number of antigen binding sites that are conserved in both *A. palliata* and *H. sapiens* (ABS CSS).

Locus	SL (bp)	VS (bp)	SL (aa)	VS (aa)	ABS	ABS VS	ABS CSS	ABS CSA
DQA1-L1	186	2	62	1	19	5	2	12
DQA1-L2	186	2	62	2	19	5	2	12
DQB1-L1	250	NA	83	NA	25	8	NA	17
DQB1-L2	247	1	82	0	NA	NA	NA	NA
DQB1-L3	250	15	83	9	25	6	7	12
DQB2	247	1	82	1	NA	NA	NA	NA
DRB	245	34	81	23	24	15	1	8

Table 2-4. Pairwise and overall average codon-based Z-test of selection. We present here *P* values for the rejection of the neutrality hypothesis under positive selection ($d_N > d_S$) to the left of the slash and purifying selection ($d_N < d_S$) to the right A) DQA1-L1, B) DQB1-L3 and C) DRB. Significant values are bolded.

A)

DQA1-L1	DQA1-L1-1	DQA1-L3	DQA1-L5
DQA1-L1-1			
DQA1-L1-3	0.15 / 1.00		
DQA1-L1-5	1.00 / 0.15	1.00/ 0.25	
Overall average	1.00/0.25		

B)

DQB1_L3	DQB1_L3-3	DQB1_L3-5	DQB1_L3-6	DQB1_L3-7
DQB1_L3-3				
DQB1_L3-5	**0.02** / 1.00			
DQB1_L3-6	0.23 / 1.00	0.34 / 1.00		
DQB1_L3-7	0.36 / 1.00	**0.03** / 1.00	0.21 / 1.00	
Overall average	0.07/1.00			

C)

DRB	DRB-1	DRB-2	DRB-3	DRB-4
DRB-1				
DRB-2	0.07 / 1.00			
DRB-3	**0.00** / 1.00	**0.00** / 1.00		
DRB-4	1.00 / 0.23	1.00 / 0.37	**0.02** / 1.00	
Overall average	**0.03/**1.00			

Figures

Figure 2-1. The three main sampling locations in Mexico, Costa Rica and Peru are indicated by red circles. The green shading represents the disjunct distribution of *A. palliata*. The green rectangle shows a close-up of the location of the nine different groups that were sampled in Mexico.

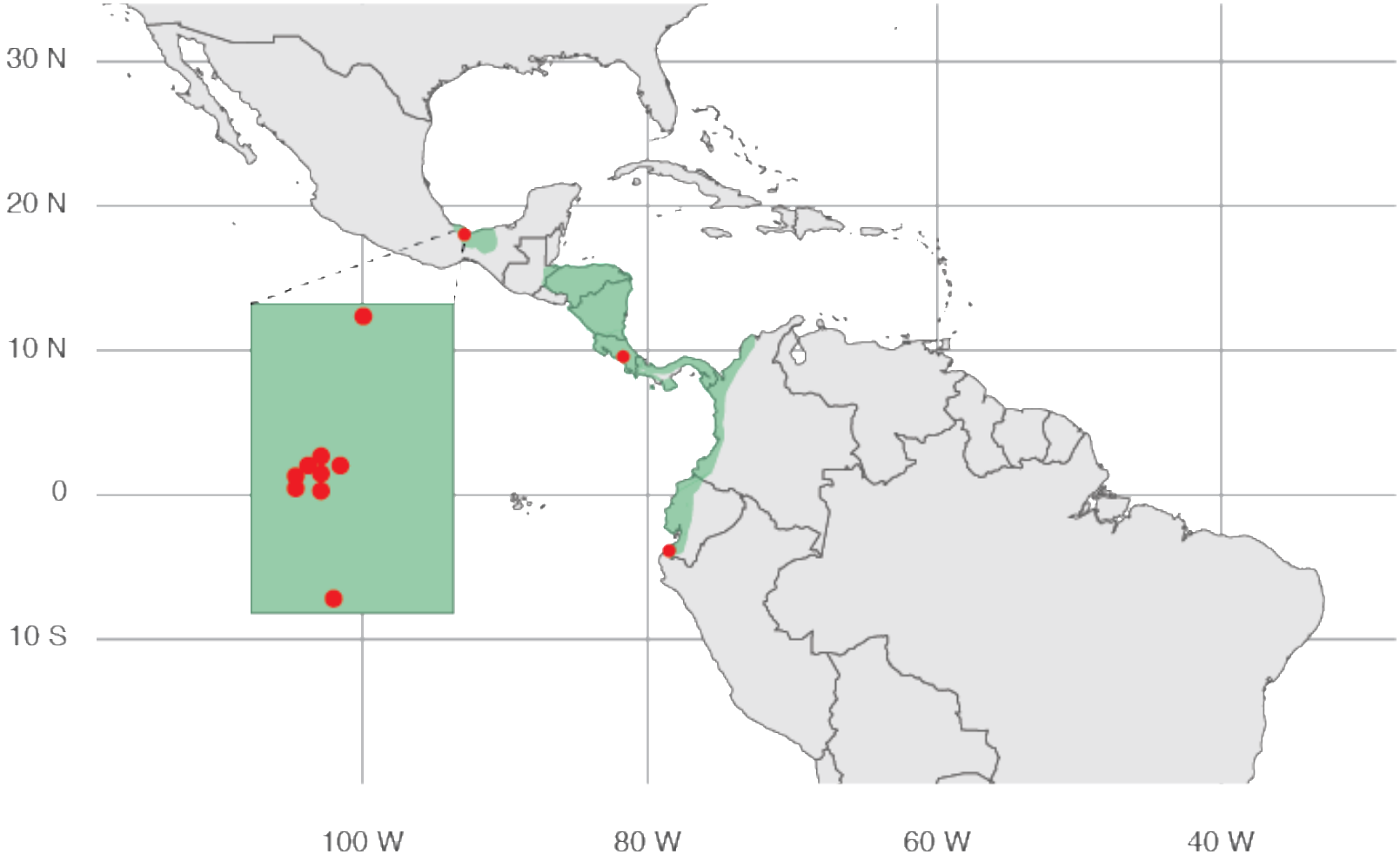
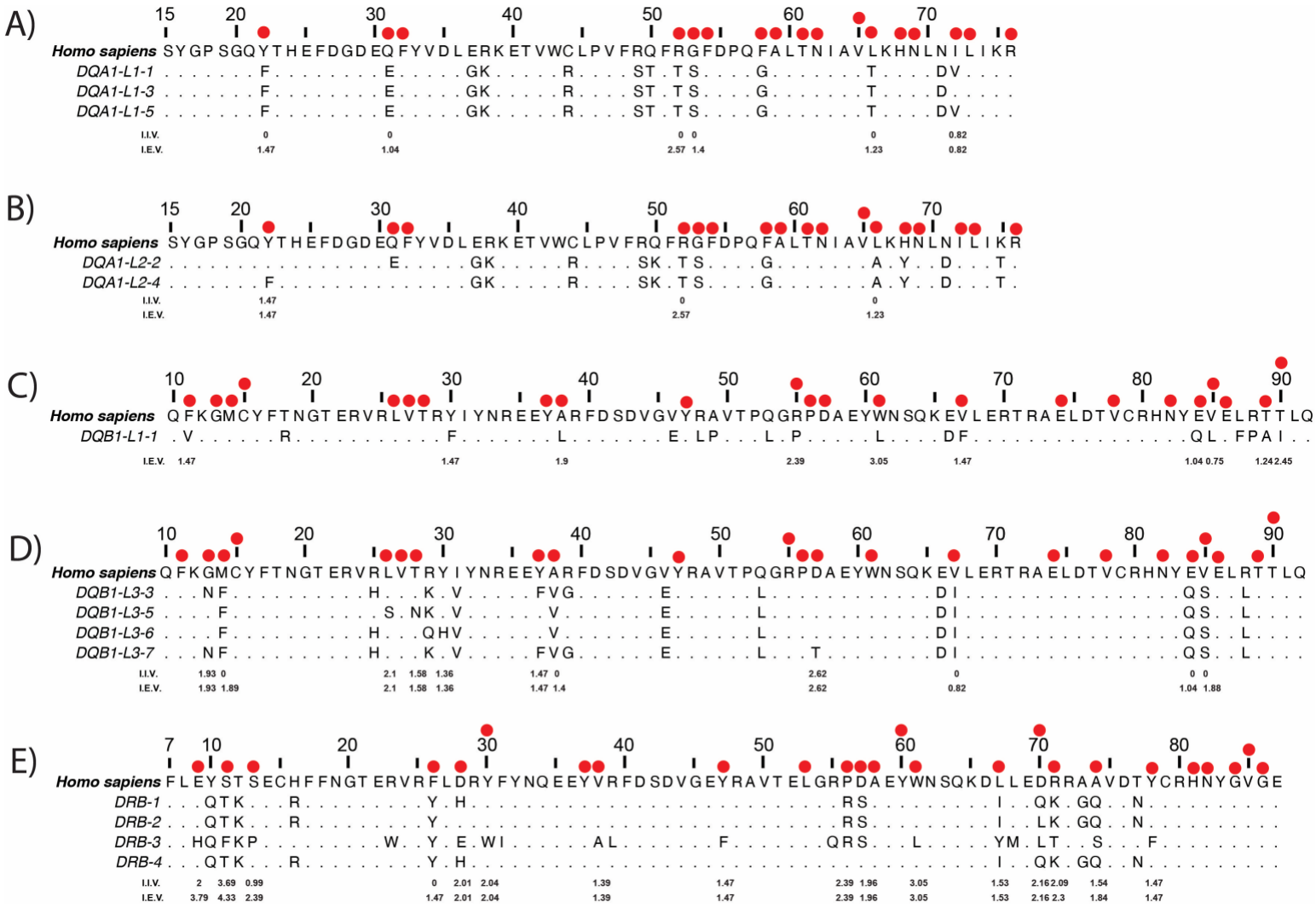


Figure 2-2. Amino acid sequence alignment and antigen binding sites. For A) DQA1-L1, B) DQA1-L2, C) DQB1-L1, D) DQB1-L3 and E) DRB. *A. palliata* sequences are aligned to a human consensus sequence (on top). Site numbering is based on the human sequence. Antigen binding sites are indicated by red circles. I.I.V. (Intraspecific Index Value) indicates the highest amino acid dissimilarity index value among *A. palliata* alleles. E. I. V (intErspecific Index Value) indicates the highest amino acid dissimilarity index value comparing *A. palliata* alleles to the human reference sequence.



Supplementary Materials

Table 2-S5. Read numbers after filtering steps. This table contains the minimum (Min.) and maximum (Max.) number of reads per sample, the average number of reads across all samples and the total number of reads in 116 samples. The total number of samples is 116 as we had duplicates for each individual (58 total individuals).

Filtering step	Min.	Max.	Average	Total
Initial number of reads	3,745	46,737	7,486	868,376
Quality filter (phred >20)	2,287	45,378	5,602	649,832
Merging sequences	2,190	43,610	5,346	620,136
Fragment size (>130 bp)	768	29,800	2,684	311,344

Table 2-S6. Individual genotypes. In the first and second column we present the individual identification code and geographic location respectively. The rest of the columns contain the alleles found in each locus. The different loci are indicated on the top row. Missing data is indicated as NA.

Species	Individual ID	DQA1-L1		DQA1-L2		DQB1-L1		DQB1-L2		DQB1-L3		DQB2		DRB	
A. palliata	S-001	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-002	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-003	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-004	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-005	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-087	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-088	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-089	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-090	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-091	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-092	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-143	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-144	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-145	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-146	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-147	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-172	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-173	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-174	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-175	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-176	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-204	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-205	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-206	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-207	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-208	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-209	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-608	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-609	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-610	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-611	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-612	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-613	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-614	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-615	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-617	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-618	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-1
A. palliata	S-030	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-5	DQB1-L3-5	DQB2-2	DQB2-2	DRB-2	DRB-2
A. palliata	S-031	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-3	DQB1-L3-7	DQB2-1	DQB2-1	DRB-1	DRB-3
A. palliata	S-032	DQA1-L1-1	DQA1-L1-5	DQA1-L2-4	DQA1-L2-4	DQB1-L1-1	DQB1-L1-1	DQB1-L2-4	DQB1-L2-4	DQB1-L3-3	DQB1-L3-7	DQB2-1	DQB2-2	DRB-1	DRB-3
A. palliata	S-033	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-4
A. palliata	S-034	DQA1-L1-1	DQA1-L1-5	DQA1-L2-4	DQA1-L2-4	DQB1-L1-1	DQB1-L1-1	DQB1-L2-4	DQB1-L2-4	DQB1-L3-3	DQB1-L3-7	DQB2-1	DQB2-2	DRB-2	DRB-3
A. palliata	S-035	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-7	DQB1-L3-7	DQB2-1	DQB2-1	DRB-1	DRB-3
A. palliata	S-036	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-7	DQB1-L3-7	DQB2-1	DQB2-1	DRB-1	DRB-3
A. palliata	S-037	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-4	DRB-3
A. palliata	S-038	DQA1-L1-1	DQA1-L1-5	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	DQB2-1	DQB2-1	DRB-1	DRB-4
A. palliata	S-039	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-4	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-4	DQB1-L3-3	DQB1-L3-7	DQB2-1	DQB2-1	DRB-2	DRB-3
A. palliata	Peru-A2	DQA1-L1-1	DQA1-L1-3	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	DQB2-2	DQB2-2	DRB-1	DRB-3
A. palliata	Peru-A4	DQA1-L1-1	DQA1-L1-3	NA	NA	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	NA	NA	DQB2-2	DQB2-2	NA	NA
A. palliata	Peru-A5	DQA1-L1-1	DQA1-L1-1	NA	NA	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	DQB2-1	DQB2-2	NA	NA
A. palliata	Peru-A10	DQA1-L1-3	DQA1-L1-3	NA	NA	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	DQB2-2	DQB2-2	NA	NA
A. palliata	Peru-A16	DQA1-L1-1	DQA1-L1-3	NA	NA	NA	NA	NA	NA	DQB1-L3-6	DQB1-L3-6	NA	NA	NA	NA

A. palliata	Peru-G1	NA	NA	NA	NA	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	DQB2-1	DQB2-2	NA	NA
A. palliata	Peru-G5	DQA1-L1-3	DQA1-L1-3	NA	NA	NA	NA	NA	NA	DQB1-L3-6	DQB1-L3-6	DQB2-2	DQB2-2	NA	NA
A. palliata	Peru-G6	DQA1-L1-3	DQA1-L1-3	NA	NA	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	DQB2-1	DQB2-2	NA	NA
A. palliata	Peru-G11	DQA1-L1-3	DQA1-L1-3	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	DQB2-1	DQB2-2	NA	NA
A. palliata	Peru-G24	DQA1-L1-1	DQA1-L1-3	DQA1-L2-2	DQA1-L2-2	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-6	DQB1-L3-6	DQB2-2	DQB2-2	NA	NA

Table 2-S7. Allele frequencies per geographic regions.

Allele	Mexico	Costa Rica	Peru
DRB-1	1	0.3	NA
DRB-2	0	0.2	NA
DRB-3	0	0.35	NA
DRB-4	0	0.15	NA
DQA1-L1-1	1	0.85	0.33
DQA1-L1-3	0	0	0.67
DQA1-L1-5	0	0.15	0
DQA1-L2-2	1	0.55	1
DQA1-L2-4	0	0.45	0
DQB1-L1-1	1	1	1
DQB1-L2-2	1	0.55	1
DQB1-L2-4	0	0.45	0
DQB1-L3-3	1	0.5	0
DQB1-L3-5	0	0.1	0
DQB1-L3-6	0	0	1
DQB1-L3-7	0	0.4	0
DQB2-1	1	0.8	0
DQB2-2	0	0.2	1

Table 1-S8. Genotype frequencies per geographic regions.

Genotype	Mexico	Costa Rica	Peru
DRB-1/DRB-1	1	0	NA
DRB-1/DRB-3	0	0.4	NA
DRB-1/DRB-4	0	0.2	NA
DRB-2/DRB-2	0	0.1	NA
DRB-2/DRB-3	0	0.2	NA
DRB-3/DRB-4	0	0.1	NA
DQA1-L1-1/DQA1-L1-1	1	0.7	0
DQA1-L1-1/DQA1-L1-5	0	0.3	0
DQA1-L1-1/DQA1-L1-3	0	0	0.67
DQA1-L1-3/DQA1-L1-3	0	0	0.33
DQA1-L2-2/DQA1-L2-2	1	0.3	1
DQA1-L2-2/DQA1-L2-4	0	0.5	0
DQA1-L2-4/DQA1-L2-4	0	0.2	0
DQB1-L1-1/DQB1-L1-1	1	1	1
DQB1-L2-2/DQB1-L2-2	1	0.3	1
DQB1-L2-2/DQB1-L2-4	0	0.5	0
DQB1-L2-4/DQB1-L2-4	0	0.2	0
DQB1-L3-3/DQB1-L3-3	1	0.3	0
DQB1-L3-3/DQB1-L3-7	0	0.4	0
DQB1-L3-5/DQB1-L3-5	0	0.1	0
DQB1-L3-6/DQB1-L3-6	0	0	1
DQB1-L3-7/DQB1-L3-7	0	0.2	0
DQB2-1/DQB2-1	1	0.7	0
DQB2-1/DQB2-2	0	0.2	0
DQB2-2/DQB2-2	0	0.1	1

Chapter 3: Asymmetric and Increased Introgression at MHC Loci in a Primate Hybrid Zone

Abstract

When distributions of two genetically divergent lineages overlap and they breed and produce fertile offspring a hybrid zone is formed. Hybridization may result in introgression, which is the transfer of genetic material into a heterospecific genomic background. Studies on human evolution have found evidence of ancestral hybridization as a source of adaptive genetic variation, specifically for immune-related genes. However, relevant empirical data is missing from studies of observable hybridization. We use a non-human primate hybrid zone, the *Alouatta* hybrid zone (*A. palliata* x *A. pigra*), to evaluate introgression at the Major Histocompatibility Complex (MHC), a genomic region fundamental to adaptive immunity. We analyzed a total of 124 wild individuals (*A. palliata* (N=37), *A. pigra* (N=38) and individuals within the hybrid zone (N=49)) at seven MHC loci (DQA1-L1, DQA1-L2, DQB1-L1, DQB1-L2, DQB1-L3, DQB2 and DRB). We compared introgression at MHC loci to neutral expectations derived from a panel of 24 neutral markers (microsatellites). We found asymmetric introgression from *A. palliata* into *A. pigra* at all seven loci but not in the other direction. Moreover, we found that alleles at five loci were introgressing significantly differently from neutral expectations. DQA1-L1 (p=0.01), DQA1-L2 (p=0.00), DQB1-L1 (p=0.04) and DQB2 (p=0.02) presented introgression concordant with over-dominance while DQB1-L2 (p=0.01) pattern of introgression was concordant with under-dominance. We believe selection is maintaining *A. palliata* alleles in the genomic

background of A. pigra. Taken together our results provide evidence of the potential of hybridization as a source of adaptive variation.

Introduction

The Major Histocompatibility Complex (MHC) is a multigene family that plays a key role in adaptive immunity (Gorer et al. 1948; Hughes and Yeager 1998; Dawkins et al. 1999; Beck and Trowsdale 2000; Klein et al. 2007). The MHC represents the most polymorphic and gene dense region known in mammals (Kelley et al. 2005). Because of its role in pathogen recognition, MHC extreme polymorphism has been hypothesized to be a result of balancing selection imposed by pathogens (Doherty and Zinkernagel 1975; Jeffery and Bangham 2000; Bernatchez and Landry 2003). As genetic variability is maintained in MHC, natural selection favors novel genetic variants (Ejsmond & Radwan 2015). We hypothesize that introgression, through species hybridization, may be an important source of genetic variation that is potentially adaptive for MHC genes. What distinguishes introgression from other sources of genetic variation (standing variation, mutation and intraspecific geneflow) is the higher probability of incorporating a potentially adaptive allele in a genomic region where genetic variability is favored. The first reason is the ready access to a larger collection of genetic variants in contrast with the variability that mutation alone can create and the time it will require to do so. The other reason is the higher probability of incorporating a functional novel genetic variant when compared to intraspecific geneflow or standing variation. Not only this variation would be novel but also functionally divergent. Both, intraspecific geneflow and standing variation, by definition, are less likely to confer a novel genetic variation than interspecific geneflow.

Different studies have shown that introgression can in fact serve as a source of adaptive genetic variation (Rieseberg 2011; Hedrick 2013; Suarez-Gozalez et al. 2018). Most of these studies showcase particular cases in which a specific selective force (herbivore resistant (Whitney et al. 2015; Le Corre et al. 2020), pesticide resistance (Song et al. 2012), Mullerian mimicry (Pardo-Diaz et al. 2012) maintain heterospecific variants in the recipient species. However, introgression in genes in which variability is favored (e.g., MHC, *cds* gene, *S* -self-incompatibility- gene) should not be case or taxa dependent but more of a generalized pattern. Studies on different non-model organisms have provided evidence that supports introgression at MHC genes (Grossen et al. 2014; Sagonas et al. 2018; Dudek et al. 2019; Pohjoismäki et al. 2021; Šimková et al. 2021; Talarico et al. 2021). Also, introgression has been reported on human HLA (Human Leukocyte Antigen) genes, the human version of MHC, as a result of archaic hybridization (Abi-Rached et al. 2011; Racimo et al. 2015; Zhang et al 2020; but see Yasukochi & Ohashi 2017). Human MHC introgression studies are based on ancient admixture inferences as hybridization with archaic hominins is no longer observable. However, current primate hybrid zones present an opportunity to empirically assess the effect of hybridization on the genetic variation of MHC in the context of human evolution given the close phylogenetic relationship.

Alouatta palliata and *Alouatta pigra* are sister species that hybridize in southern Mexico. These species display notable differences in their ecology, morphology (Kelaita and Cortés -Ortiz 2013), behavior (Kappeler and Schaik 2002), and general genetic composition (Cortés-Ortiz 2003; Steinberg et al. 2014). They diverged about 3 MYA (Cortés -Ortiz et al. 2003) and share a complex evolutionary history involving a series of invasion waves from South America into Central America that defined their current distribution (Ford 2006). *A. palliata* ranges from northwestern Peru through the Pacific coast of northern South America up to southeast Mexico,

while *A. pigra* has a more restricted distribution, constrained to the Yucatan peninsula (Mexico, Belize and Northern Guatemala) (Barrueta-Rath et al 2003; Estrada et al 2002). Despite their mostly parapatric distribution, these species have a small area of overlap in the state of Tabasco, Mexico. Hybridization has been documented with viable and fertile offspring resulting in a narrow hybrid zone (Cortés -Ortiz et al. 2007). Hybrid individuals show a spectrum of phenotypes that correlate with their genomic background (Kelaita et al. 2011; Kelaita and Cortés -Ortiz 2013; Ho et al. 2014; Kitchen et al. 2019). Moreover, F1 hybrids have not been found in the hybrid zone but backcrossing has been substantial (Cortés -Ortiz et al. 2007; Kelaita et al. 2011; Kelaita and Cortés -Ortiz 2013; Cortés –Ortiz et al. 2015). Genome wide analyses in the *Alouatta* hybrid zone have revealed reduced introgression for loci residing in the X-chromosome, hinting at a role in reproductive isolation (Cortés –Ortiz et al. 2018; Baiz et al. 2020). These same studies have also shown differential introgression across the genome, indicating that some alleles have decreased or increased introgression on different loci when compared to neutral expectations. We predict that MHC alleles will show increased introgression when compared to neutrality.

Our aim was to evaluate the effect of hybridization on the genetic diversity of MHC genes in the *Alouatta* hybrid zone and assess introgression at MHC loci. We genotyped a total of 124 individuals (37 *A. palliata* individuals, 38 *A. pigra* individuals and 49 admixed individuals) at seven different MHC Class II loci (DQA1 exon 2 locus 1, DQA1 exon 2 locus 2, DQB1 exon 2 locus 1, DQB1 exon 2 locus 2, DQB1 exon 2 locus 3, DQB2 exon 2 and DRB exon 2) using high-throughput sequencing. We compared genetic diversity levels of both parental populations to admixed individuals. Admixture levels, proportion of each parental population ancestry in admixed individuals, were previously determined using a collection of 24 polymorphic neutral

markers (microsatellites) (Cortés –Ortiz et al. 2018). Using these same microsatellites, we constructed neutral introgression expectation and compared the introgression patterns of MHC loci to determine differential introgression. The results from this study highlight the role of hybridization as a source of genetic variation that can potentially be adaptive based on empirical evidence from a current primate hybrid zone.

Methods

Sampling and DNA extraction

We analyzed blood samples from a total of 124 wild *Alouatta* individuals located in geographically distinct groups. We included in our analysis 37 samples from *A. palliata* (nine groups), 38 samples from *A. pigra* (16 groups) and 49 samples from individuals within the hybrid zone (29 groups) (Figure 3-1). *A. palliata* and *A. pigra* individuals are referred as to parental populations. Individuals from both parental populations are considered non-admixed based on the significant geographical distance from the hybrid zone and previous admixture analyses (Cortés-Ortiz et al. 2019). Samples were collected between 1998 and 2010. Details for the collection of samples can be found in Cortés-Ortiz et al. (2003, 2019) and Kelaita et al. (2011).

We extracted DNA from blood samples using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA). We followed manufacturer's protocol for DNA extraction kits. We quantified DNA concentration in a Qubit 4 Fluorometer and using Qubit dsDNA high-sensitivity assay (Invitrogen, Carlsbad, CA). All samples were standardized at a DNA concentration of 5 ng/µl.

Microsatellite amplification and genotyping

We used a published data set on the same individuals (Nidiffer and Cortés-Ortiz 2015) that implemented a panel of 24 polymorphic microsatellites (Supplementary Information Table 3-S4). Samples were sequenced at the University of Michigan Sequencing Core and alleles were scored using GeneMarker V 1.5. Details of the genotyping process and amplification conditions can be found in Nidiffer and Cortés-Ortiz 2015 and Cortés-Ortiz et al. 2019.

MHC amplification and Next Generation Sequencing Library Preparation

We performed PCRs using previously published primers targeting partial sequences of three different exons: DRB exon 2 (DRB), DQA1 exon 2 (DQA1) and DQB1 (DQB1) exon 2 (Tiercy et al. 1990, Gyllensten & Elrich 1988 and Yasunaga et al. 1996). As previously reported (Diaz et al 2000) the set of primers used for DQB1 also co-amplified DQB2. Moreover, duplications were detected in both DQA1 and DQB1, resulting in two and three distinct loci respectively (as explained in Chapter I). In total we amplified seven MHC loci: DRB, DQA1-2L1, DQA1-L2, DQB1-L1, DQB1-L2, DQB1-L3 and DQB2.

We followed the 16S protocol from Illumina (Illumina 2013) for next generation sequencing library preparation with modifications. Briefly, the protocol consists in a PCR to amplify the product of interest, and another PCR to attach labeling indexes to the sequences generated in the first PCR. The final products are pooled in equimolar DNA concentrations for sequencing. We modified the first stage PCR to a 3-phase Touch-Down with an initial denaturing temperature of 95 °C for 5 minutes, followed by two cycles of 94 °C for 30 sec, annealing temperature (55 °C for DQA1, 60 °C for DQB1 and DQB2, 68 °C for DRB) for 30 sec and 72 °C for 30 seconds (phase 1); two cycles of 94 °C for 30 sec, annealing temperature (50 °C

for DQA1, 55 °C for DQB1 and DQB2, 63 °C for DRB) for 30 sec and 72 °C for 30 seconds (phase 2); 30 cycles of 94 °C for 30 sec, annealing temperature (48 °C for DQA1, 52 °C for DQB1 and DQB2, 58 °C for DRB) for 30 sec and 72 °C for 30 seconds (phase 3) and a final extension at 72 °C for 10 minutes. PCR total volume was 20 µl and included 5.5 µl of ultrapure water, 1 µl of forward primer at 10 µM, 1 µl of reverse primer at 10 µM, 10 µl of Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Ipswich, MA) and 2.5 µl of DNA at 5ng/µ. For the second (Index) PCR, our total reaction volume was 25 µl: 4 µl of ultrapure water, 3 µl of Nextera XT Index Primer 1 (Illumina, San Diego, CA), 3 µl of Nextera XT Index Primer 2 (Illumina, San Diego, CA), 12.5 µl of Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Ipswich, MA) and 2.5 µl of product from the first PCR. Samples were labeled with unique index combinations for individual identification.

Products of the indexed PCRs were quantified using fluorometric DNA quantification using the Qubit dsDNA high-sensitivity assay, diluted to a concentration of 4nM, and then pooled together. We amplified and sequenced all samples in duplicate. Duplicates were amplified in independent PCRs and sequenced on different sequencing runs. We submitted our libraries for 250 bp pair-end MiSeq Nano 500 sequencing (Illumina, San Diego, CA) at The University of Michigan Advanced Genomics Core.

MHC genotyping

We filtered out raw sequences that had a phred score lower than 20 using Trimmomatic (Bolger et al. 2014). We merged pair-end sequences using FLASH (Magoc & Salzberg 2011) keeping default parameters and filtered sequences based on size (retaining fragments>130 bp) using

PRINSEQ (Schmieder & Edwards 2011) to eliminate sequences that were not within the range of expected size.

We created an R function to determine preliminary genotypes for each individual at each locus. This function ranks all the unique reads of a particular locus by frequency in each individual, eliminates all singletons, trims primers, and calculates the relative frequency of each unique read to the total number of reads in the individual. Once we determined individual putative alleles, we identified and eliminated chimeric sequences manually by aligning all sequences of each locus and evaluating for the distribution of the variable sites in the aligned sequences. Chimeric sequences were defined as sequences that combined two parental sequences that were found in high frequency within an individual. All putative alleles that remained after removing chimeras were confirmed as MHC-DQ and MHC-DRB sequences comparing sequence identity to other published MHC-DQQ and MHC-DRB sequences of closely related taxa using BLAST. Following the protocol of Sommer et al. 2013 we determined the final genotypes. Briefly, to call true alleles we i) identified high frequency putative alleles in and individual, ii) corroborated allele presence in the duplicate, and iii) confirm that the frequency of each potential true allele sequence was higher than any known artifact (e.g., singletons, chimeras).

MHC genetic diversity

We calculated number of alleles for each population (the two parental populations and the admixed population) and determined shared alleles. We also calculated expected and observed heterozygosity using GenoDive (Meirmans 2020). Expected and observed heterozygosity were estimated for all seven MHC loci excluding DQB1-L3 given the uncertainty of gene copy numbers at this locus for *A. pigra* individuals.

Admixture analyses

Individuals within the hybrid zone were assigned a hybrid index value indicating the proportion of genomic ancestry from either parental population. The hybrid index is based on the allele frequencies at each microsatellite locus for each parental species, where a value of 1 represents an *A. pigra* individual and a value of 0 represents an *A. palliata* individual (as detailed in Cortés-Ortiz et al. 2019). Previous studies on the *Alouatta* hybrid zone (Cortés-Ortiz et al. 2019, Baiz et al. 2020) have revealed a bimodal distribution in terms of ancestry for the admixed individuals (Supplementary Material Figure 3-S4). This means that most of the individuals within the hybrid zone have either a very similar genomic background to *A. pigra* (a hybrid index value close to 1) or to *A. palliata* (a hybrid index value close to 0). Therefore, we were able to select a representative sample of admixed individuals in terms of their hybrid index value. We included 24 individuals with a hybrid index close to 1 (hereafter referred as *pigra*-like), 23 individuals with a hybrid index value close to 0 (hereafter referred as *palliata*-like) and 2 individuals with a hybrid index value close to 0.5. Admixed individuals and their hybrid index value are listed in Supplementary Materials Table 3-S5.

Introgression analyses

We implemented introgress (Gompert and Buerkle 2009; Gompert and Buerkle 2010) to analyze if introgression at any MHC loci deviates from neutral expectations based on genomic clines. Loci that exhibit signatures of non-neutral introgression are determined comparing the likelihoods of both a regression and a neutral model based on the empirical data. Neutral introgression is simulated with a permutation process. We computed 10,000 permutations and estimated statistical significance. Neutral introgression is defined by the expected number of

homozygotes (for either *A. palliata* or *A. pigra*) and heterozygotes given the observed data. If homozygotes are more or less common than neutral expectations this would be indicative of positive or negative selection. If the number of heterozygotes does not correspond to neutral expectations this would be indicative of under-dominance and over-dominance (Nolte et al. 2009, Teeter et al. 2010). Introgress analysis relies on co-dominant genetic markers. Since some individuals had more than two alleles at DQB1_L3 in *A. pigra* it was not possible for us to determine homozygosity vs heterozygosity and therefore we did not include DQB1-L3 for this analysis.

Results

MHC Genotyping and genetic diversity

We successfully amplified and sequenced seven MHC loci: DQA1 exon 2 locus 1 (DQA1-L1), DQA1 exon 2 locus 2 (DQA1-L2), DQB1 exon 2 locus 1 (DQB1-L1), DQB1 exon 2 locus 2 (DQB1-L2), DQB1 exon 2 locus 3 (DQB1-L3), DQB2 exon 2 (DQB2) and DRB exon 2 (DRB) in *A. palliata*, *A. pigra* and admixed individuals. Our paired-end sequencing raw data yielded a total of 1,922,833 reads across all individuals (including duplicates). After filtering our data (based on sequence quality and fragment size), we recovered 689,404 sequences that were used for allele identification and genotyping. Supplementary Materials Table 3-S6 shows the maximum and minimum number of reads per individual and total during each step of the filtering process.

Both DQA1-L1 and DQA1-L2 alleles had a length of 186 bp and were in the correct reading frame with no indication of pseudogenization. DQB1-L1 and DQB1-L3 had a length of

250bp and no signatures of pseudogenization. DQB1-L2 had a total length of 247 bp, presenting a one codon deletion at codon 55 when compared to the 83 codons of the human allele sequence, but they still were in the correct reading frame. For DQB2, all alleles had a length of 247 bp with no indication of pseudogenization. All DRB alleles had a length of 245bp with no indication of pseudogenization.

In terms of genetic diversity, all *A. palliata* individuals were homozygous at all seven loci. This was also the case for all *palliata*-like admixed individuals. DQB1-L3 was the locus with the highest number of alleles and DQB1-L1 was the locus with the greater heterozygosity (HO) in both *A. pigra* and *pigra*-like admixed individuals. The rest of the loci varied in number of alleles and heterozygosity in *A. pigra* and *pigra*-like admixed individuals (Table 3-1 and Table 3-2). The two individuals with intermediate hybrid index values were heterozygous at all loci with one allele coming from one parental population the other allele coming from the other parental population. We did not detect any *A. palliata* alleles for DQB1-L3, also for this locus we cannot determine heterozygosity vs homozygosity as we were not able to determine the total number of loci. Only allele DQB1-L1-1 was shared between *A. palliata* and *A. pigra*, the rest of the alleles were unique to each species. Individual genotypes for all loci in admixed individuals are provided in Supplementary Material Table 3-S7.

Asymmetric differential introgression at MHC loci

All *palliata*-like individuals presented no introgressed alleles from *A. pigra*. *Pigra*-like individuals presented introgressed alleles from *A. palliata* at all loci (Figure 3-2). DQB1-L1 is the only locus in which both parental populations share one allele. In this case, the ancestry identity of admixed individuals for this locus is based on probability given that all *A. palliata*

individuals are homozygous for this locus and *A. pigra* has three alleles. DRB locus had the highest number of *pigra*-like individuals with introgressed alleles (42% of total individuals) while DQB1-L3 had the lowest number of individuals with introgressed alleles (12.5% of total individuals) (Table 3-3).

Genomic clines showed significant differential introgression at five loci. DQA1-L1 (p=0.01), DQA1-L2 (p=0.00), DQB1-L1 (p=0.04) and DQB2 (p=0.02) present a similar pattern of introgression based on cline shape consistent with over-dominance. DQB1-L2 genomic cline (p=0.01) shows signatures of under-dominance. While DRB genomic cline is not significantly different from neutral expectations (p=0.20), 42% of *pigra*-like individuals present *A. palliata* alleles in homozygote (N=3) and heterozygote (N=7) form (Figure 3--3).

Discussion

All seven MHC Class two genes presented asymmetric introgression from *A. palliata* into *A. pigra* genomic background with no introgression in the other direction. Introgression patterns were different at each locus and five out of six loci showed statistically significant differential introgression. The patterns of differential introgression are concordant with adaptive introgression further supported by selection analyses in previous studies (Chapter I and Chapter II). We believe selection is the most likely mechanism maintaining these heterospecific MHC genetic variants in the genomic background of *A. pigra* suggesting that hybridization is a source of adaptive genetic variation for MHC loci as ancient human admixture studies have proposed.

A. palliata ancestry decreases MHC genetic diversity in admixed individuals

Two characteristics of our data enabled us to confidently determine the ancestry source of MHC alleles in admixed individuals. First, all *A. palliata* individuals from the parental populations are homozygotes at all MHC loci. Second, all MHC loci presented species-specific alleles. From our genotyping, we were able to determine that *palliata*-like individuals did not have any introgressed alleles at any MHC locus but *pigra*-like individuals presented *A. palliata* alleles at all loci. We also compared levels of genetic diversity (HO) of admixed individuals to both parental species. For our analyses of genetic diversity, we separated admixed individuals into *palliata*-like and *pigra*-like. This allowed us to evaluate at a finer scale how the genomic ancestry affected genetic diversity in admixed individuals. We observed that overall HO, and HO at each locus, was lower in admixed individuals when compared to the *A. pigra* population. When we excluded the *palliata*-like individuals from the analyses the level of heterozygosity increased as all *palliata*-like admixed individuals were homozygous at all loci. Still, *A. pigra* individuals presented an overall higher genetic diversity than *pigra*-like admixed individuals. From previous studies (Melo-Carrillo et al. 2020, Zoonomia Consortium 2020, Chapter II), we have evidence that *A. palliata* populations in Mexico have extremely reduced genetic diversity genome-wide and at MHC loci. This is further reflected in the admixed individuals as *A. palliata* ancestry decreases MHC genetic diversity in admixed individuals.

Asymmetric and differential introgression due to selection

All MHC loci presented asymmetric introgression from *A. palliata* into *A. pigra*. Moreover, five out of the six MHC loci showed significant non-neutral introgression. DQA1-L1, DQA1-L2, DQB1-L1 and DQB2 genomic clines are concordant with over-dominance, in which alleles from

both parental populations are maintained in a heterozygote form more frequently than what is expected under neutral expectations. Overdominance, by heterozygote advantage, has long been proposed as a mechanism for maintaining variation in MHC (Doherty & Zinkernagel 1975).

Heterozygote individuals may be more resistant to infectious diseases as they can present a more diverse set of antigens to initiate an immune response (Hedrick 2012). DQB1-L2, genomic cline showed a pattern of underdominance. In this case admixed individuals that are *pigra*-like present *A. palliata* alleles in a homozygote form with a frequency that exceeds neutral introgression expectations. It is probable that individual homozygotes for DQB1-L2-2 allele may have a fitness advantage over alternative genotypes. We have shown (Chapter I) that selection acts differently at each locus for MHC. Theoretical and empirical studies also provide evidence that not all MHC loci undergo the same selective pressures. However, *A. palliata* MHC alleles are introgressing into *A. pigra* genomic background while *A. pigra* alleles do not introgress.

We had initially hypothesized introgression to be asymmetric from *A. pigra* to *A. palliata* based on the low MHC genetic diversity in *A. palliata* and selection favoring variability in MHC loci. However, we did not observe any *A. pigra* allele in *palliata*-like individuals. Other studies that have looked at MHC introgression, in different systems, have found symmetrical introgression (Grossen et al. 2014, Dudek et al. 2019, Pohjoismäki et al. 2021) or asymmetric introgression from the more genetically diverse population (Šimková et al. 2021) contrasting with our results. We believe that selection might be maintaining these alleles in the heterospecific genomic background. The *Alouatta* hybrid zone, where all admixed individuals occur, is very narrow (approximately 20 km. wide) (Cortés-Ortiz et al. 2007, 2015) and the surrounding geographic area has very similar environmental conditions, therefore it is reasonable to assume that all admixed individuals are experiencing similar selective pressures. This could

indicate that MHC alleles found in *A. palliata* are a beneficial in the environment where admixed individuals occur. Moreover, we have evidence of positive selection at these loci (Chapter I and Chapter II). Our data does not allow us to directly link MHC genetic variation to functional variation. However, MHC-DQ and MHC-DR genes have been associated with infectious disease resistance and susceptibility to autoimmune diseases (Matzaraki et al. 2017). It is probable that *A. palliata* alleles are conferring an advantage in this respect suggesting that hybridization could be a source of adaptive genetic variation.

Other processes can also result in asymmetric introgression such as movement of hybrid zones over time. The movement of a hybrid zone can create asymmetric introgression as the hybrid zone shifts from the range of one population into the range of the other population. It is unlikely that hybrid zone movement is generating the pattern we are observing for MHC alleles. Baiz et al. (2020) study on the *Alouatta* hybrid zone showed that introgression is asymmetric but in the other direction for most alleles across the genome. Therefore, if the movement of the hybrid zone would be causing asymmetric introgression at MHC loci we would observe the same pattern that is more common across the genome. Alternatively, assortative mating could also result in asymmetric introgression. We discarded this possibility as *palliata*-like and *pigra*-like admixed individuals are equally abundant (Kelaita and Cortés-Ortiz 2013; Cortés-Ortiz et al. 2019) and our previous genetic analyses of uni-parentally and bi-parentally inherited markers (mtDNA and Y chromosome, and autosomal SNPs, respectively) show that crosses occur in both directions (Cortes-Ortiz et al. 2019). This indicates that there is no preference in mating or difference in abundance of individuals with one parental population ancestry.

Concluding remarks

To our knowledge, this is the first study to provide empirical evidence of hybridization as a source of potential adaptive genetic variation in primates for MHC. We observed asymmetric introgression at all MHC genes from *A. palliata* to *A. pigra*. Moreover, most of these loci showed differential introgression concordant with adaptive introgression. We propose that *A. palliata* alleles are conferring a selective advantage to admixed individuals based on the results of this study and positive selection occurring at these genes according to our previous studies. We acknowledge that adaptive introgression cannot be stated without evidence on functionality. However, alternative explanations appear to be less likely. Moreover, hybridization has been proposed as an important source of adaptive genetic variation in humans. Different genes have been identified as potential genomic regions that acquired beneficial genetic variation through introgression. Importantly many of these genes have an immune function including MHC genes (or HLA). The results from this study provide additional evidence to support hybridization as a source of genetic variation in MHC genes in humans. The phylogenetic proximity of platyrrhines to humans, as opposed to other studies that have evaluated MHC genes variability in the context of hybridization, highlights the relevance of this study in the context of human evolution.

References

Abi-Rached L, Jobin M, Luo M, Dalva K (2011) The Shaping of Modern Human Admixture with Archaic Humans. *Science*, **334**, 89–94.

Baiz MD, Tucker PK, Mueller JL, Cortés-Ortiz L (2020) X-linked signature of reproductive isolation in humans is mirrored in a howler monkey hybrid zone. *Journal of Heredity*, **111**, 419–428.

Beck S, Trowsdale J (2000) T HE H UMAN M AJOR H ISTOCOMPATIBILITY C OMPLEX : Lessons from the DNA Sequence. *Annu. Rev. Ecol. Evol. Syst.*, **1**, 117–1137.

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, **16**, 363–377.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Le Corre V, Siol M, Vigouroux Y, Tenaillon MI, Délye C (2020) Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 25618–25627.

Cortés-Ortiz L, Agostini I, Aguiar ML *et al.* (2015) Hybridization in Howler Monkeys: Current Understanding and Future Directions. In: *Howler Monkeys: Adaptive Radiation, Systematics, and Morphology* (eds Kowalewski MM, Garber PA, Cortés-Ortiz L, Urbani B, Youlatos D), pp. 107–131. Springer US, New York.

Cortés-Ortiz L, Bermingham E, Rico C *et al.* (2003) Molecular systematics and biogeography of the genus, *Alouatta*. *Molecular Phylogenetics and Evolution*, **26**, 64–81.

Cortés-Ortiz L, Duda TF, Canales-Espinosa D *et al.* (2007) Hybridization in large-bodied new world primates. *Genetics*, **176**, 2421–2425.

Cortés-Ortiz L, Nidiffer MD, Hermida-Lagunes J *et al.* (2018) Reduced Introgression of Sex Chromosome Markers in the Mexican Howler Monkey (*Alouatta palliata* × A. *pigra*) Hybrid Zone. *International Journal of Primatology 2018 40:1*, **40**, 114–131.

Cortés-Ortiz L, Roos C, Zinner D (2019) Introduction to Special Issue on Primate Hybridization and Hybrid Zones. *International Journal of Primatology 2019 40:1*, **40**, 1–8.

Dawkins R, Leelayuwat C, Gaudieri S *et al.* (1999) Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunological reviews*, **167**, 275–304.

Diaz D, Naegeli M, Rodriguez R *et al.* (2000) Sequence and diversity of MHC DQA and DQB genes of the owl monkey *Aotus nancymaae*. *Immunogenetics*, **51**, 528–537.

Doherty P, Zinkernagel R (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*.

Dudek K, Gaczorek TS, Zieliński P, Babik W (2019) Massive introgression of major histocompatibility complex (MHC) genes in newt hybrid zones. *Molecular Ecology*, **28**, 4798–4810.

Ejsmond MJ, Radwan J (2015) Red Queen Processes Drive Positive Selection on Major Histocompatibility Complex (MHC) Genes. *PLoS Computational Biology*, **11**, 1–14.

Estrada A, Mendoza A, Castellanos L *et al.* (2002) Population of the black howler monkey (*Alouatta pigra*) in a fragmented landscape in Palenque, Chiapas, Mexico. *American Journal of Primatology*, **58**, 45–55.

Ford, S. M. (2006). The biogeographic history of Mesoamerican primates. In a. Estrada, P. A.

Garber, M. S. M. Pavelka, & L. Luecke (Eds.), New perspectives in the study of Mesoamerican primates: Distribution, ecology, behavior and conservation (pp. 81–114). Developments in primatology: Progress and prospects. New York: Springer science+business media.

Genereux DP, Serres A, Armstrong J *et al.* (2020) A comparative genomics multitool for scientific discovery and conservation. *Nature*, **587**, 240–245.

Gompert Z, Alex Buerkle C (2010) introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.

Gorer PA, Lyman S, Snell GD (1948) Studies on the Genetic and Antigenic Basis of Tumour Transplantation. Linkage between a Histocompatibility Gene and "Fused" in Mice. *Proceedings of the Royal Society B: Biological Sciences*, **135**, 499–505.

Grossen C, Keller L, Biebach I, International T, Genome G (2014) Introgression from Domestic Goat Generated Variation at the Major Histocompatibility Complex of Alpine Ibex. *PLoS Genetics*, **10**.

Gyllensten UB, Erlich H a (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 7652–6.

Hedrick PW (2012) What is the evidence for heterozygote advantage selection? *Trends in Ecology and Evolution*, **27**, 698–704.

Hedrick PW (2013) Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**, 4606–4618.

Ho L, Cortés-Ortiz L, Dias PAD *et al.* (2014) Effect of ancestry on behavioral variation in two species of howler monkeys (*Alouatta pigra* and A. *palliata*) and their hybrids. *American Journal of Primatology*, **76**, 855–867.

Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates Reproduced with permission of the copyright owner . Further reproduction prohibited without permission . *Annual review of genetics*, **32**, 415–435.

Jeffery KJM, Bangham CRM (2000) Do infectious diseases drive MHC diversity? *Microbes and Infection*, **2**, 1335–1341.

Kappeler PM, Van Schaik CP (2002) Evolution of Primate Social Systems. *International Journal of Primatology 2002 23:4*, **23**, 707–740.

Kelaita MA, Cortés-Ortiz L (2013) Morphological variation of genetically confirmed *Alouatta Pigra* × A. *palliata* hybrids from a natural hybrid zone in Tabasco, Mexico. *American Journal of Physical Anthropology*, **150**, 223–234.

Kelaita M, Dias PAD, Aguilar-Cucurachi MS, Canales-Espinosa D, Cortés-Ortiz L (2011) Impact of Intrasexual Selection on Sexual Dimorphism and Testes Size in the Mexican Howler Monkeys *Alouatta palliata* and A . *pigra*. *American Journal of Physical Anthropology*, **187**, 179–187.

Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics*, **56**, 683–695.

Kitchen DM, Bergman TJ, Dias PAD *et al.* (2019) Temporal but Not Acoustic Plasticity in Hybrid Howler Monkey (*Alouatta palliata* × A. *pigra*) Loud Calls. *International Journal of Primatology*, **40**, 132–152.

Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annual review of genetics*, **41**, 281–304.
Magoc T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957–2963.
Matzaraki V, Kumar V, Wijmenga C, Zhernakova A (2017) The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biology 2017 18:1*, **18**, 1–21.
Meirmans PG, Liu S, Van Tienderen PH (2018) The Analysis of Polyploid Genetic Data. *Journal of Heredity*, 283–296.
Melo-Carrillo A, Dunn JC, Cortés-Ortiz L (2020) Low genetic diversity and limited genetic structure across the range of the critically endangered Mexican howler monkey (*Alouatta palliata* mexicana). *American Journal of Primatology*, **82**, 1–1.
Nidiffer MD, Cortés-Ortiz L (2015) Intragroup genetic relatedness in two howler monkey species (*Alouatta pigra* and A. *palliata*): Implications for understanding social systems and dispersal. *American Journal of Primatology*, **77**, 1333–1345.
Pardo-Diaz C, Salazar C, Baxter SW *et al.* (2012) Adaptive Introgression across Species Boundaries in Heliconius Butterflies. *PLOS Genetics*, **8**, e1002752.
Pohjoismäki JLO, Michell C, Levänen R, Smith S (2021) Hybridization with mountain hares increases the functional allelic repertoire in brown hares. *Scientific Reports 2021 11:1*, **11**, 1–14.
Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E (2015) Evidence for archaic adaptive introgression in humans. *Nature reviews. Genetics*, **16**, 359–371.
Rieseberg L (2011) Adaptive Introgression: The Seeds of Resistance. *Current Biology*, **21**, R581–R583.
Sagonas K, Runemark A, Antoniou A *et al.* (2019) Selection, drift, and introgression shape MHC polymorphism in lizards. *Heredity*, **122**, 468–484.
Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**, 863–864.
Šimková A, Gettová L, Civáňová K *et al.* (2021) Diversity of MHC IIB genes and parasitism in hybrids of evolutionarily divergent cyprinoid species indicate heterosis advantage. *Scientific Reports 2021 11:1*, **11**, 1–18.
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.
Song S, Dey DK, Holsinger KE (2012) NIH Public Access. *Theoretical Population Biology*, **100**, 130–134.
Steinberg ER, Nieves M, Mudry MD (2014) Cytogenetics Multiple sex chromosome systems in howler monkeys. *Comparative Cytogenetics*, **8**, 43–69.
Suarez-Gonzalez A, Lexer C, Cronk QCB (2018) Adaptive introgression: a plant perspective. *Biology Letters*, **14**.
Talarico L, Marta S, Rossi AR *et al.* (2021) Balancing selection, genetic drift, and human-mediated introgression interplay to shape MHC (functional) diversity in Mediterranean brown trout. *Ecology and Evolution*, **11**, 10026–10041.
Teeter KC, Thibodeau LM, Gompert Z *et al.* (2010) THE VARIABLE GENOMIC ARCHITECTURE OF ISOLATION BETWEEN HYBRIDIZING SPECIES OF HOUSE MICE. *Evolution*, **64**, 472–485.
Tiercy J -M, Jeannet M, Mach B (1990) A new HLA-DRB1 allele within the DRw52 supertypic

specificity (DRw13-DwHAG): Sequencing and direct identification by oligonucleotide typing. *European Journal of Immunology*, **20**, 237–241.
Whitney KD, Randell RA, Rieseberg LH (2015) Adaptive Introgression of Herbivore Resistance Traits in the Weedy Sunflower Helianthus annuus. *https://doi.org/10.1086/504606*, **167**, 794–807.
Yasukochi Y, Ohashi J Elucidating the origin of HLA-B*73 allelic lineage: Did modern humans benefit by archaic introgression? *Immunogenetics*.
Yasunaga S, Kimura A, Hamaguchi K, Rønningen KS, Sasazuki T (1996) Different contribution of HLA-DR and -DQ genes in susceptibility and resistance to insulin-dependent diabetes mellitus (IDDM). *Tissue Antigens*, **47**, 37–48.
Zhang X, Kim B, Lohmueller KE, Huerta-Sánchez E (2020) The Impact of Recessive Deleterious Variation on Signals of Adaptive Introgression in Human Populations. *Genetics*, **215**, 799–812.

Tables

Table 3-1. Number of alleles at each locus in each population. Overall observed (HO) and expected (HE) heterozygosity for each population.

Individuals	DQA1-L1	DQA1-L2	DQB1-L1	DQB1-L2	DQB1-L3	DQB2	DRB	HO	HE
A. palliata	1	1	1	1	1	1	1	0	0
A. pigra	2	2	3	2	7	3	3	0.624	0.617
Admixed	2	3	3	3	6	2	4	0.315	0.604
Admixed *palliata*-like	1	1	1	1	1	1	1	0	0
Admixed *pigra*-like	2	3	3	3	6	2	4	0.56	0.646

Table 3-2. Observed (HO) and expected (HE) heterozygosity at each locus for each population. HO is located on the left side of the slash, and HE is located on the right side.

Populations	DQA1-L1	DQA1-L2	DQB1-L1	DQB1-L2	DQB1-L3	DQB2	DRB
A. palliata	0	0	0	0	NA	0	0
A. pigra	0.342/0.287	0.171/0.391	0.658/0.623	0.04/0.517	NA	0.421/0.474	0.432/0.516
Admixed	0.188/0.503	0.25/0.578	0.333/0.589	0.103/0.439	NA	0.188/0.503	0.25/0.53
Admixed-*palliata*-like	0	0	0	0	0	0	0
Admixed-*pigra*-like	0.273/0.24	0.409/0.583	0.545/0.592	0.231/0.609	NA	0.227/0.206	0.409/0.606

Table 3-3. Total number and proportion (in percentage) of individuals with introgressed alleles at each locus for each population. Information is presented for *palliata*-like and *pigra*-like populations. The last column (Total unique) indicates the total number of individuals that presented an introgressed allele in at least one locus. Introgressed alleles cannot be determined for DQB1-L1 as it is the only locus with shared alleles.

Population	*N*	DQA1-L1	DQA1-L2	DQB1-L1	DQB1-L2	DQB1-L3	DQB2	DRB	Total unique
***palliata*-like**	23	0	0	0	0	0	0	0	
***pigra*-like**	24	10 (42%)	10 (42%)	NA	6 (25%)	3 (12%)	6 (25%)	10 (42%)	12 (50%)

Figures

Figure 3-1. Sampling locations for A. palliata, A. pigra and admixed individuals. Top map indicates the general location of the sampling region. Bottom map shows the distribution of *A. palliata* (light grey), *A. pigra* (dark grey) and the hybrid zone (where both distributions overlap) along with group locations and sampling size. inset map shows a zoomed view of the group locations where admixed individuals where sampled.

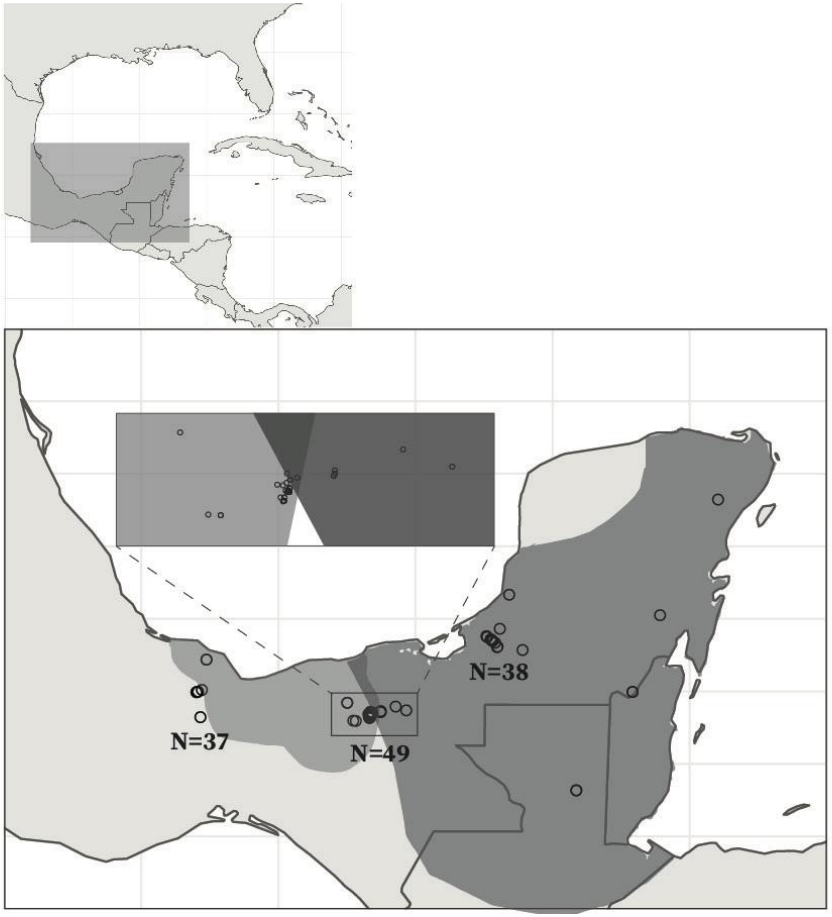


Figure 3-2. Color-coded genotype of admixed individuals. Each row represents an individual and each column represents a locus respectively labeled. Gold blocks indicate homozygotes for *A. palliata*, blue blocks indicate homozygotes for *A. pigra*, green blocks indicate heterozygotes and white blocks indicate missing data. Rows containing *pigra*-like individuals are delineated on the right side of the graph by a dashed line (small gaps), intermediate individuals by a solid line and *palliata* like individuals by a dashed (large gaps).

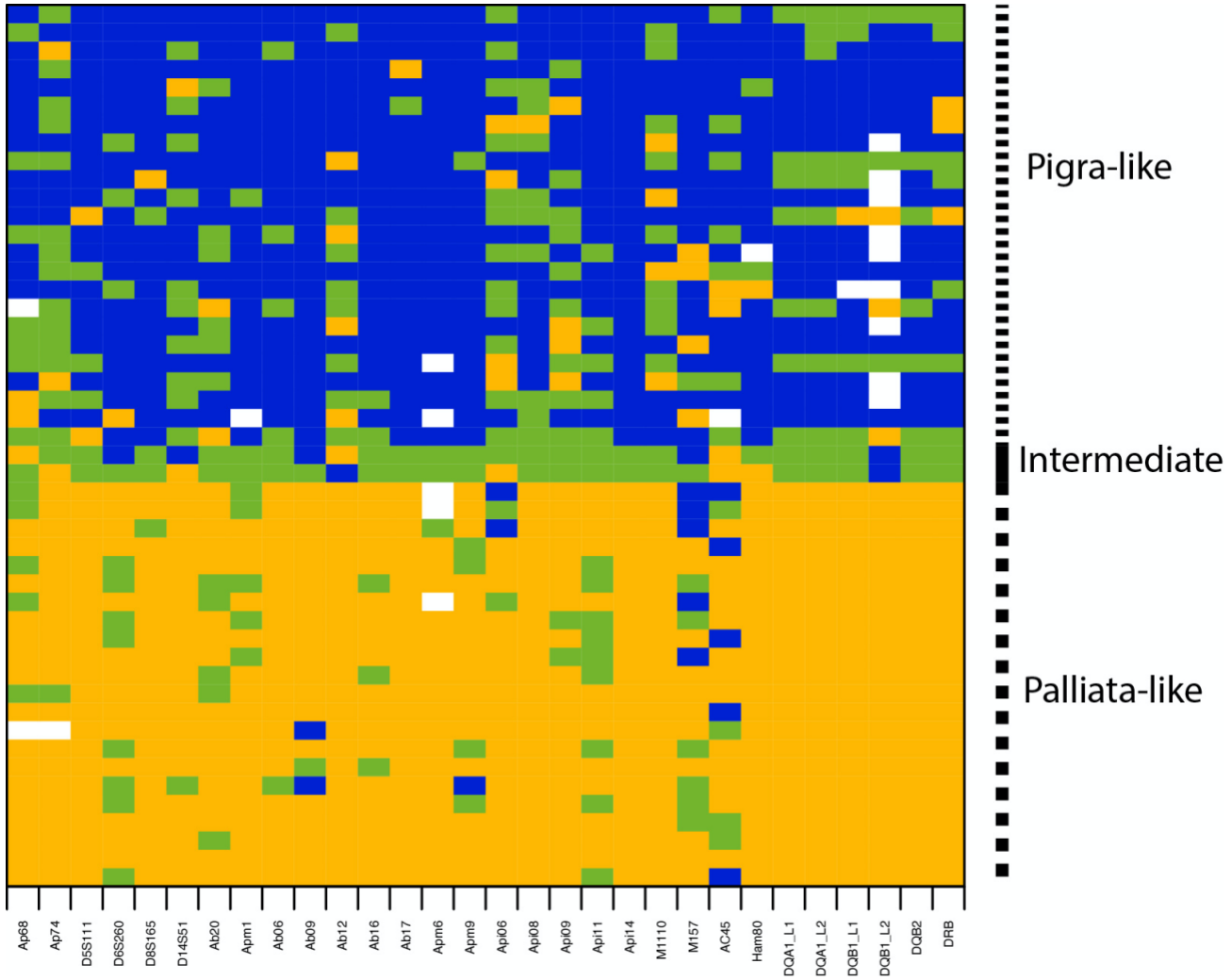
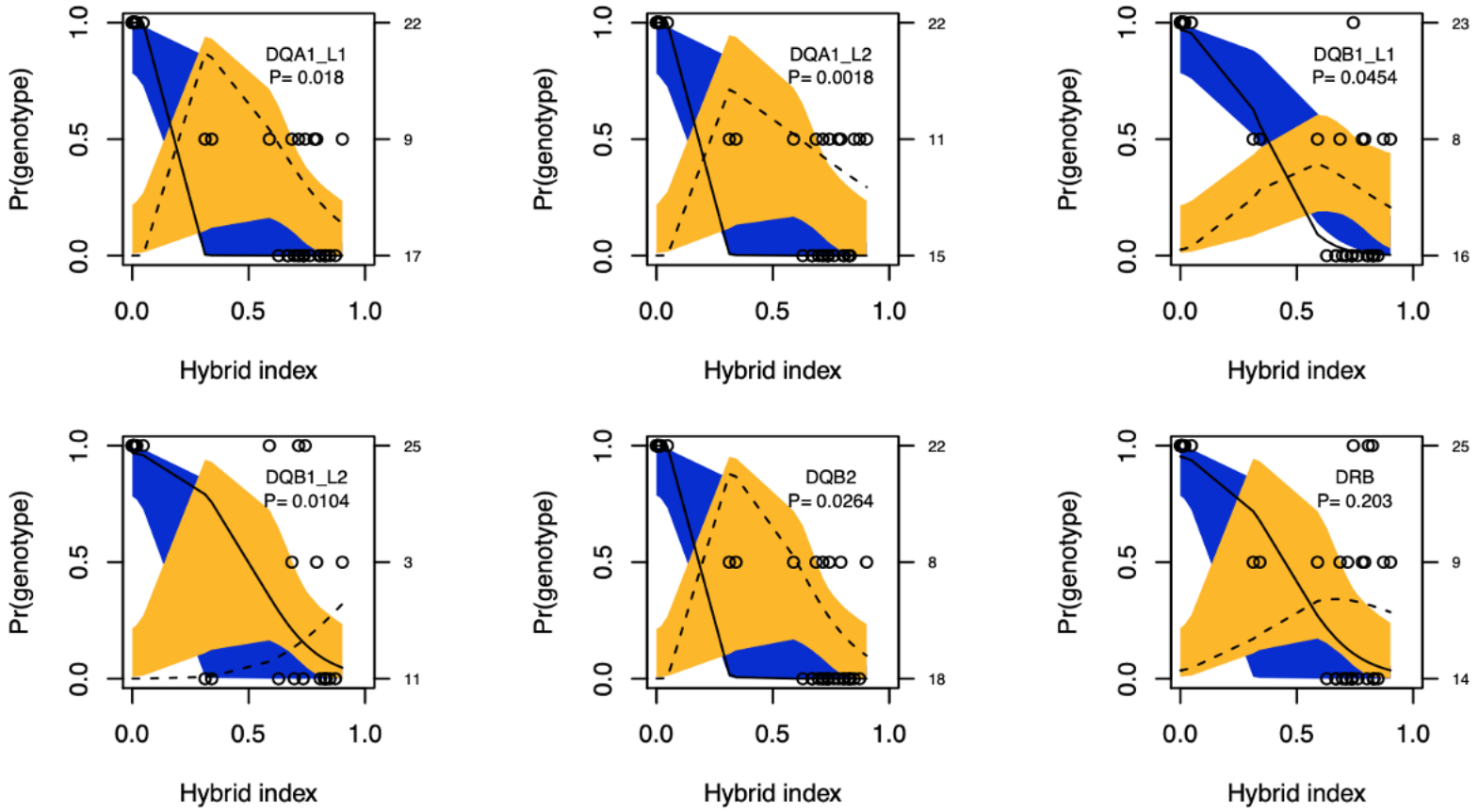


Figure 3-3. Genomic clines for MHC loci. Hybrid index (HI) value indicates the proportion of alleles derived from *A. palliata* (zero indicating a non-admixed *A. palliata* individual and one a non-admixed *A. pigra* individual). Y-axis indicates the probability of genotypes on the left and the number of individuals homozygous for *A. palliata* (top), heterozygous for both populations (middle) and homozygous for *A. pigra* (bottom). Each circle represents an individual. Colored clines denote 95% confidence intervals according to neutral introgression for homozygous *A. palliata* (blue) and heterozygotes (gold). Solid line and dashed line are estimated clines based on the observed genotypes for *A. palliata* homozygotes and heterozygotes, respective. Locus name and *p-value*, indicating departure from neutrality, are given on the top right corner. DQA1-L1, DQA1-L2, DQB1-L1 and DQB2 show significant differential introgression with a cline pattern corresponding to over-dominance. DQB1-L2 and DRB present a cline pattern corresponding to underdominance, only DQB1-L2 present statistically significant differential introgression.



Supplementary Materials

Table 3-S4. Microsatellites used in this study and sources.

Locus	Source
Ap68	Ellsworth and Hoelzer (1998)
Ap74	Ellsworth and Hoelzer (1998)
D5S111	Cortés-Ortiz et al. 2010)
D6S260	Cortés-Ortiz et al. 2010)
D8S165	Cortés-Ortiz et al. 2010)
D14S51	Cortés-Ortiz et al. 2010)
Ab06	Goncalves et al. (2004)
Ab07	Goncalves et al. (2004)
Ab09	Goncalves et al. (2004)
Ab12	Goncalves et al. (2004)
Ab16	Goncalves et al. (2004)
Ab17	Goncalves et al. (2004)
Ab20	Goncalves et al. (2004)
Apm1	Goncalves et al. (2004)
Apm6	Cortés-Ortiz et al. (2010)
Apm9	Cortés-Ortiz et al. (2010)
Api06	Cortés-Ortiz et al. (2010)
Api08	Cortés-Ortiz et al. (2010)
Api09	Cortés-Ortiz et al. (2010)
Api11	Cortés-Ortiz et al. (2010)
Api14	Cortés-Ortiz et al. (2010)
1110	Di Fiore and Fleischer (2004)
157	Di Fiore and Fleischer (2004)
AC45	Oklander et al. (2007)
Ham80	Katoh et al. (2009)

Table 3-S5. Admixed individuals' hybrid index values. First column contains the population designation of each individual. Middle column lists the identification code. Last column indicates the hybrid index value (HI). HI of zero indicates a full *A. palliata* individual and HI of one indicates a full *A. pigra* individual.

Designation	Individuals	Hybrid Index value
palliata-like	S-010	0.05
palliata-like	S-093	0.05
palliata-like	S-094	0.05
palliata-like	S-097	0.05
palliata-like	S-099	0.05
palliata-like	S-101	0.00
palliata-like	S-158	0.02
palliata-like	S-159	0.04
palliata-like	S-160	0.00
palliata-like	S-162	0.03
palliata-like	S-293	0.02
palliata-like	S-296	0.04
palliata-like	S-334	0.13
palliata-like	S-335	0.08
palliata-like	S-337	0.05
palliata-like	S-339	0.07
palliata-like	S-434	0.03
palliata-like	S-561	0.04
palliata-like	S-589	0.07
palliata-like	S-625	0.07
palliata-like	S-629	0.05
palliata-like	S-630	0.03
pigra-like	S-015	0.94
pigra-like	S-155	0.74
pigra-like	S-163	0.89
pigra-like	S-165	0.92
pigra-like	S-297	0.96
pigra-like	S-301	0.85
pigra-like	S-305	0.81
pigra-like	S-307	0.84
pigra-like	S-308	0.92
pigra-like	S-313	0.83
pigra-like	S-317	0.97
pigra-like	S-322	0.87

pigra-like	S-323	0.90
pigra-like	S-329	0.94
pigra-like	S-408	0.93
pigra-like	S-410	0.93
pigra-like	S-429	0.81
pigra-like	S-430	0.87
pigra-like	S-529	0.98
pigra-like	S-540	0.87
pigra-like	S-543	0.79
pigra-like	S-546	0.87
pigra-like	S-597	0.85
pigra-like	S-601	0.86
Intermediate	S-098	0.46
Intermediate	S-157	0.50

Table 3-S6. Number of reads after each filtering step for all individuals (*A. palliata*, *A. pigra* and admixed), including the minimum (Min.) and maximum (Max.) number of reads per sample, the mean number of reads across all samples, and the total number of reads in all samples. The total number of samples is 248 as we had duplicates for each individual (N=124 individuals).

Filtering step	Min	Max	Average	Total
Initial number of reads	3,745	46,737	7,486	1,922,833
Quality filter (phred >20)	2,287	45,378	5,602	1,438,913
Merging sequences	2,190	43,610	5,346	1,373,158
Fragment size (>130 bp)	768	29,800	2,684	689,404

Table 3-S7. Individual Genotypes for admixed individuals. List of individual genotypes for A) DQA1-L1 and DQA1-L2, B) DQB1-L1, DQB1-L2 and DQB1-L3, C) DQB2 and D) DRB. Only alleles present (but not genotypes) are provided for A. *pigra*, *pigra*-like and intermediates individuals for DQB1-L3 as number of loci could not be determined. N/A = not amplified.

A)

Designation	Individuals	DQA1-L1		DQA1-L2	
palliata-like	S-010	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-093	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-094	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-097	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-099	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-101	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-158	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-159	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	*S-160*	*DQA1-L1-1*	*DQA1-L1-1*	*DQA1-L2-2*	*DQA1-L2-2*
palliata-like	S-162	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-293	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-296	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-334	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-335	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-337	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-339	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-434	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-561	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-589	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-625	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-629	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
palliata-like	S-630	DQA1-L1-1	DQA1-L1-1	DQA1-L2-2	DQA1-L2-2
pigra-like	S-015	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-8
pigra-like	S-155	DQA1-L1-3	DQA1-L1-1	DQA1-L2-6	DQA1-L2-2
pigra-like	S-163	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
pigra-like	S-165	DQA1-L1-1	DQA1-L1-3	DQA1-L2-8	DQA1-L2-2
pigra-like	S-297	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-2
pigra-like	S-301	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
pigra-like	S-305	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
pigra-like	S-307	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
pigra-like	S-308	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6

pigra-like	S-313	DQA1-L1-1	DQA1-L1-3	DQA1-L2-6	DQA1-L2-2
pigra-like	S-317	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-2
pigra-like	S-322	DQA1-L1-3	DQA1-L1-1	DQA1-L2-8	DQA1-L2-2
pigra-like	S-323	DQA1-L1-3	DQA1-L1-1	DQA1-L2-6	DQA1-L2-2
pigra-like	S-329	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-8
pigra-like	S-408	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
pigra-like	S-410	DQA1-L1-3	DQA1-L1-3	DQA1-L2-7	DQA1-L2-7
pigra-like	S-429	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
pigra-like	S-430	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
pigra-like	S-529	DQA1-L1-3	DQA1-L1-1	DQA1-L2-8	DQA1-L2-2
pigra-like	S-540	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
pigra-like	S-543	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-8
pigra-like	S-546	DQA1-L1-3	DQA1-L1-3	DQA1-L2-8	DQA1-L2-6
pigra-like	S-597	DQA1-L1-3	DQA1-L1-3	DQA1-L2-6	DQA1-L2-6
pigra-like	S-601	DQA1-L1-1	DQA1-L1-3	DQA1-L2-2	DQA1-L2-6
Intermediate	S-098	DQA1-L1-1	DQA1-L1-3	DQA1-L2-2	DQA1-L2-6
Intermediate	S-157	DQA1-L1-1	DQA1-L1-3	DQA1-L2-2	DQA1-L2-8

B)

Designation	Individuals	DQB1-L1		DQB1-L2		DQB1-L3		
palliata-like	S-010	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-093	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-094	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-097	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-099	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-101	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-158	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-159	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-160	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-162	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-293	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-296	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-334	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-335	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-337	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-339	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-434	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-561	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-589	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-625	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-629	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
palliata-like	S-630	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-3	
pigra-like	S-015	DQB1-L1-9	DQB1-L1-11	DQB1-L2-12	DQB1-L2-12	DQB1-L3-13	DQB1-L3-14	DQB1-L3-17
pigra-like	S-155	DQB1-L1-1	DQB1-L1-11	DQB1-L2-2	DQB1-L2-2	DQB1-L3-8	DQB1-L3-10	
pigra-like	S-163	DQB1-L1-11	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	
pigra-like	S-165	DQB1-L1-1	DQB1-L1-9	DQB1-L2-12	DQB1-L2-2	DQB1-L3-17	DQB1-L3-3	
pigra-like	S-297	DQB1-L1-9	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-17		
pigra-like	S-301	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	
pigra-like	S-305	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	
pigra-like	S-307	DQB1-L1-9	DQB1-L1-11	DQB1-L2-12	DQB1-L2-12	DQB1-L3-8	DQB1-L3-10	DQB1-L3-17
pigra-like	S-308	DQB1-L1-11	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	
pigra-like	S-313	DQB1-L1-1	DQB1-L1-9	DQB1-L2-2	DQB1-L2-12	DQB1-L3-3	DQB1-L3-17	
pigra-like	S-317	DQB1-L1-1	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-17		
pigra-like	S-322	DQB1-L1-1	DQB1-L1-1	DQB1-L2-2	DQB1-L2-2	DQB1-L3-3	DQB1-L3-16	
pigra-like	S-323	DQB1-L1-1	DQB1-L1-11	NA	NA	DQB1-L3-13	DQB1-L3-14	DQB1-L3-16
pigra-like	S-329	DQB1-L1-9	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-17		
pigra-like	S-408	DQB1-L1-9	DQB1-L1-11	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8	DQB1-L3-10	
pigra-like	S-410	DQB1-L1-9	DQB1-L1-9	DQB1-L2-4	DQB1-L2-4	DQB1-L3-8		
pigra-like	S-429	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	

pigra-like	S-430	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	
pigra-like	S-529	DQB1-L1-1	DQB1-L1-9	DQB1-L2-2	DQB1-L2-12	DQB1-L3-17		
pigra-like	S-540	DQB1-L1-9	DQB1-L1-9	NA	NA	DQB1-L3-8	DQB1-L3-10	
pigra-like	S-543	DQB1-L1-9	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-17		
pigra-like	S-546	DQB1-L1-9	DQB1-L1-9	DQB1-L2-12	DQB1-L2-12	DQB1-L3-8	DQB1-L3-10	DQB1-L3-17
pigra-like	S-597	DQB1-L1-9	DQB1-L1-11	NA	NA	DQB1-L3-8	DQB1-L3-10	
pigra-like	S-601	DQB1-L1-9	DQB1-L1-9	DQB1-L2-2	DQB1-L2-2	DQB1-L3-8	DQB1-L3-10	
Intermediate	S-098	DQB1-L1-1	DQB1-L1-11	DQB1-L2-2	DQB1-L2-2	DQB1-L3-8	DQB1-L3-10	
Intermediate	S-157	DQB1-L1-1	DQB1-L1-9	DQB1-L2-2	DQB1-L2-12	DQB1-L3-17	DQB1-L3-17	

C)

Designation	Individuals	DQB2	
palliata-like	S-010	DQB2-1	DQB2-1
palliata-like	S-093	DQB2-1	DQB2-1
palliata-like	S-094	DQB2-1	DQB2-1
palliata-like	S-097	DQB2-1	DQB2-1
palliata-like	S-099	DQB2-1	DQB2-1
palliata-like	S-101	DQB2-1	DQB2-1
palliata-like	S-158	DQB2-1	DQB2-1
palliata-like	S-159	DQB2-1	DQB2-1
palliata-like	S-160	DQB2-1	DQB2-1
palliata-like	S-162	DQB2-1	DQB2-1
palliata-like	S-293	DQB2-1	DQB2-1
palliata-like	S-296	DQB2-1	DQB2-1
palliata-like	S-334	DQB2-1	DQB2-1
palliata-like	S-335	DQB2-1	DQB2-1
palliata-like	S-337	DQB2-1	DQB2-1
palliata-like	S-339	DQB2-1	DQB2-1
palliata-like	S-434	DQB2-1	DQB2-1
palliata-like	S-561	DQB2-1	DQB2-1
palliata-like	S-589	DQB2-1	DQB2-1
palliata-like	S-625	DQB2-1	DQB2-1
palliata-like	S-629	DQB2-1	DQB2-1
palliata-like	S-630	DQB2-1	DQB2-1
pigra-like	S-015	DQB2-3	DQB2-1
pigra-like	S-155	DQB2-3	DQB2-1
pigra-like	S-163	DQB2-3	DQB2-3
pigra-like	S-165	DQB2-1	DQB2-3
pigra-like	S-297	DQB2-3	DQB2-3
pigra-like	S-301	DQB2-3	DQB2-3
pigra-like	S-305	DQB2-3	DQB2-3
pigra-like	S-307	DQB2-3	DQB2-3
pigra-like	S-308	DQB2-3	DQB2-3
pigra-like	S-313	DQB2-3	DQB2-1
pigra-like	S-317	DQB2-3	DQB2-3
pigra-like	S-322	DQB2-1	DQB2-3
pigra-like	S-323	DQB2-3	DQB2-3
pigra-like	S-329	DQB2-3	DQB2-3

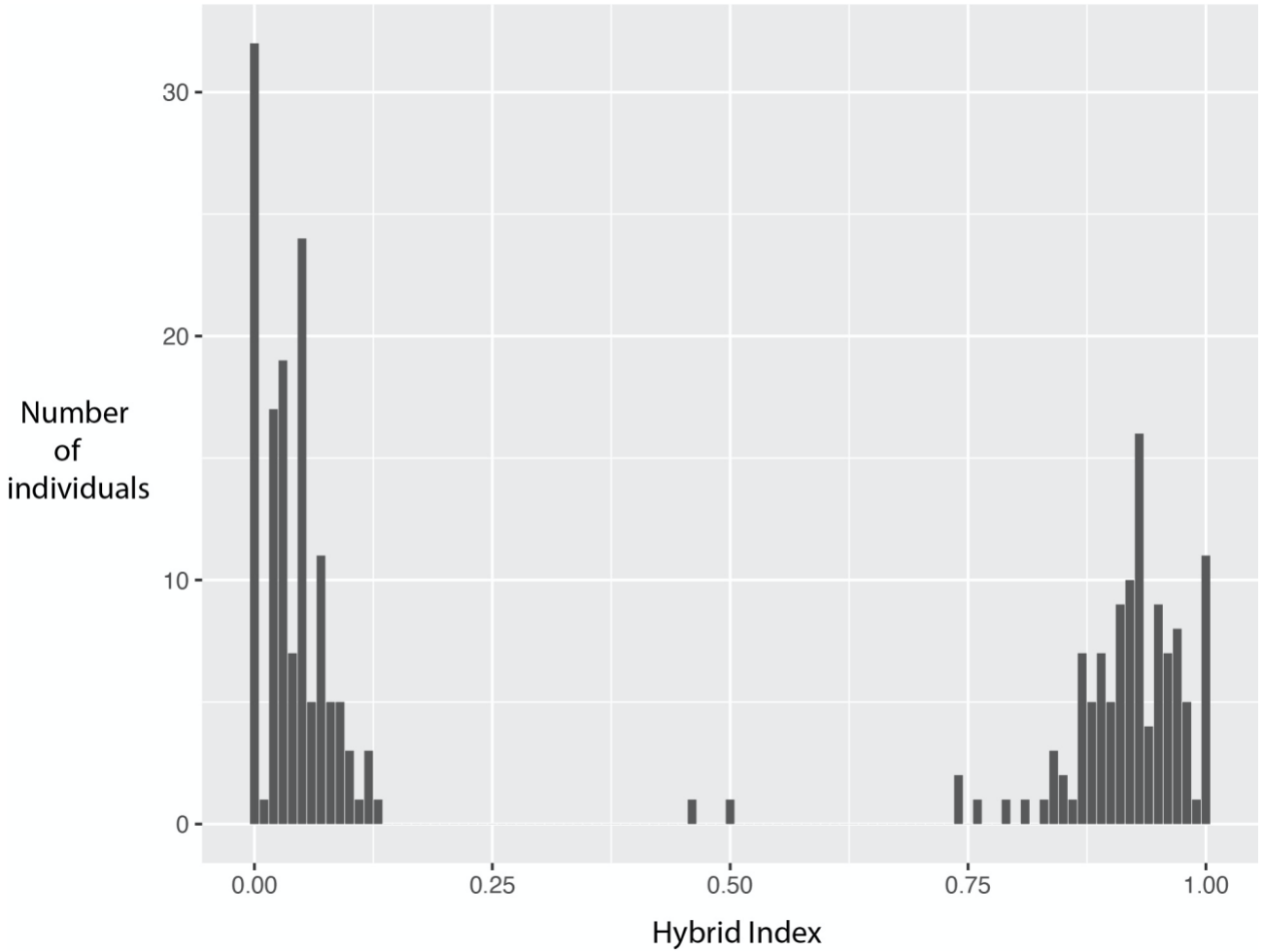
pigra-like	S-408	DQB2-3	DQB2-3
pigra-like	S-410	DQB2-3	DQB2-3
pigra-like	S-429	DQB2-3	DQB2-3
pigra-like	S-430	DQB2-3	DQB2-3
pigra-like	S-529	DQB2-1	DQB2-3
pigra-like	S-540	DQB2-3	DQB2-3
pigra-like	S-543	DQB2-3	DQB2-3
pigra-like	S-546	DQB2-3	DQB2-3
pigra-like	S-597	DQB2-3	DQB2-3
pigra-like	S-601	DQB2-1	DQB2-3
Intermediate	S-098	DQB2-1	DQB2-3
Intermediate	S-157	DQB2-1	DQB2-3

D)

Designation	Individuals	DRB	
palliata-like	S-010	DRB-1	DRB-1
palliata-like	S-093	DRB-1	DRB-1
palliata-like	S-094	DRB-1	DRB-1
palliata-like	S-097	DRB-1	DRB-1
palliata-like	S-099	DRB-1	DRB-1
palliata-like	S-101	DRB-1	DRB-1
palliata-like	S-158	DRB-1	DRB-1
palliata-like	S-159	DRB-1	DRB-1
palliata-like	S-160	DRB-1	DRB-1
palliata-like	S-162	DRB-1	DRB-1
palliata-like	S-293	DRB-1	DRB-1
palliata-like	S-296	DRB-1	DRB-1
palliata-like	S-334	DRB-1	DRB-1
palliata-like	S-335	DRB-1	DRB-1
palliata-like	S-337	DRB-1	DRB-1
palliata-like	S-339	DRB-1	DRB-1
palliata-like	S-434	DRB-1	DRB-1
palliata-like	S-561	DRB-1	DRB-1
palliata-like	S-589	DRB-1	DRB-1
palliata-like	S-625	DRB-1	DRB-1
palliata-like	S-629	DRB-1	DRB-1
palliata-like	S-630	DRB-1	DRB-1
pigra-like	S-015	DRB-5	DRB-5
pigra-like	S-155	DRB-5	DRB-1
pigra-like	S-163	DRB-5	DRB-5
pigra-like	S-165	DRB-1	DRB-5
pigra-like	S-297	DRB-5	DRB-5
pigra-like	S-301	DRB-5	DRB-1
pigra-like	S-305	DRB-7	DRB-7
pigra-like	S-307	DRB-7	DRB-7
pigra-like	S-308	DRB-5	DRB-5
pigra-like	S-313	DRB-1	DRB-5
pigra-like	S-317	DRB-1	DRB-5
pigra-like	S-322	DRB-1	DRB-1
pigra-like	S-323	DRB-5	DRB-1
pigra-like	S-329	DRB-5	DRB-5

pigra-like	S-408	DRB-1	DRB-1
pigra-like	S-410	DRB-1	DRB-1
pigra-like	S-429	DRB-5	DRB-5
pigra-like	S-430	DRB-5	DRB-5
pigra-like	S-529	DRB-1	DRB-5
pigra-like	S-540	DRB-5	DRB-5
pigra-like	S-543	DRB-5	DRB-5
pigra-like	S-546	DRB-5	DRB-7
pigra-like	S-597	DRB-8	DRB-5
pigra-like	S-601	DRB-8	DRB-5
Intermediate	S-098	DRB-1	DRB-5
Intermediate	S-157	DRB-1	DRB-5

Figure 3-S4. Bimodal distribution of admixed individuals. Hybrid index values (HI) for a total of 254 individuals within the *Alouatta* hybrid zone. HI of zero indicates a full *A. palliata* individual and HI of one indicates a full *A. pigra* individual. This is a modified figure from the results found in Cortés-Ortiz et al. 2019.



Chapter 4: Conclusion

In my dissertation, I evaluated MHC genetic variation in platyrrhines and the mechanisms and sources responsible for this variation. I focused on hybridization as a source of adaptive genetic variation. This topic has been understudied in animals, but it likely represents an important source of adaptive genetic variation, particularly for immune genes as previous studies have suggested. I used the *Alouatta* hybrid zone in Mexico (*A. palliata* x *A. pigra*) as a system to address my questions. To explore if my findings were more broadly present in platyrrhines I incorporated data that I generated from two additional howler species (*A. caraya* and *A. guariba*) and publicly available data from two more distantly related species of platyrrhines (*Callithrix jacchus* and *Aotus nancymaae*). The overall findings of this work provide new insight into the evolutionary history of MHC genes in platyrrhines, and primates more generally. Moreover, my work contributes empirical evidence of hybridization as a source of genetic variation for immune genes, supporting claims in the field of human evolution.

In Chapter 1, I presented substantial evidence (genotypes, genome search, gene prediction and phylogenetic analyses) for MHC-DQ gene duplication events in platyrrhine primates based on the novel characterization of DQA1 and DQB1 genes for four *Alouatta* species (*A. palliata*, *A. pigra*, *A. caraya* and *A. guariba*). Also, my work revealed that these duplication events in the genus *Alouatta* also occurred in more distantly related species *C. jacchus* and *A. nancymaae,* setting the time of these events at least 20 MYA. Moreover, MHC-DQA1 and MHC-DQB1 loci represent a case of trans-species polymorphisms, in which

polymorphic gene lineages persist beyond speciation events. Selection analyses indicate that this genetic variation has been maintained by selective pressures, also that positive selection is mostly occurring at antigen biding sites. Based on this evidence, I propose that pathogens are the main selective force behind the genetic variation patterns we observed in MHC-DQ loci in platyrrhines.

In Chapter 2, I found differences in MHC genetic diversity and population structure across the geographic range of *A. palliata*. The Mexican population showed extremely low levels of genetic diversity as all individuals were homozygous at all seven MHC loci. Both Costa Rica and Peru populations presented higher levels of genetic diversity despite the fact that fewer individuals were sampled in those geographic areas. *A. palliata* likely underwent multiple founder events as its range expanded into Mexico, which significantly reduced the genetic diversity in this region. These results have conservation implications for the Mexican population given the important function of MHC genes in adaptive immunity and the extremely low levels of genetic diversity. A comparative study on pathogen abundance and prevalence, across *A. palliata's* distribution could complement this study. The results could be paired with MHC genetic variation to have a more accurate representation of the vulnerability of each population.

In Chapter 3, I found evidence for hybridization as a source of adaptive variation for MHC genes in a primate hybrid zone. I had originally hypothesized asymmetric introgression from *A. pigra* into *A. palliata* genomic background given the low MHC genetic variation in *A. palliata*. However, I observed asymmetric introgression at all seven MHC loci from *A. palliata* into *A. pigra*. Moreover, five of these MHC loci showed statistically significant increased introgression when compared to neutral expectations. I proposed that the *A. palliata* alleles are advantageous in the environment that the admixed individuals occur. This is the first study to

provide empirical evidence for hybridization as a source of adaptive variation for MHC genes in primates. My work supports claims of ancestral hybridization as a source of adaptive genetic variation in immune genes for humans. The recent divergence of platyrrhines and humans, compared to other studies that have evaluated MHC in the context of hybridization, emphasizes the significance of this study in the context of human evolution. One of the limitations of this study is the comparison of MHC introgression to neutral introgression derived from microsatellite markers. Future work could derive rates of introgression at genome-wide level. The results from these analyses would inform us on how introgression occurring at MHC loci compares to introgression occurring across the entire genome.

The overall findings of my work indicate that gene duplications and hybridization have an important role in the evolution of MHC in platyrrhines and that MHC loci can be implemented as informative markers in conservation genetic studies. MHC gene duplications in platyrrhines have been maintained by selection through speciation events implying an essential function in the adaptive immunity of these primates. Introgression, through hybridization, has transferred MHC alleles from one species into another. These alleles have been maintained in the heterospecific background as they are likely favorable in the environment where admixed individuals live. I propose that introgression at MHC loci could be a generalized phenomenon in vertebrate hybrid zones, given the characteristics that are favored in the MHC genomic regions: high genetic variation and novel genetic variation. My work provides empirical evidence for introgression at MHC genes. The close phylogenetic relationship between platyrrhines and humans highlights the relevance of this work to the field of human evolution, in which empirical evidence is currently lacking. MHC loci can be used to estimate genetic variation in conservation

studies. The high variability and the intrinsic adaptive value of MHC genes, contrast the commonly used neutral markers that cannot convey any information on adaptive variation.

In general, two aspects should be considered to improve future work on MHC genetic variation in non-model organism. First, sequencing longer MHC reads. With sequencing technologies improving constantly it will become easier and cheaper to sequence longer reads. Investigating complete MHC genes brings a completely new realm for understanding MHC variation at a DNA level but more so at a protein level. Second, assessing MHC functional variation. This is a more challenging task, particularly for non-model organisms. But if possible, evaluating MHC functional variation should be a priority objective. It provides direct evidence of the adaptive value of different genetic variants that otherwise is just inferred, and it is currently quite scarce.