What mandrills leave behind: using fecal samples to characterize the major histocompatibility complex in a threatened primate

Anna Weber[1†], Jackie Lighten[2], Cock van Oosterhout[3], Amour Guibinga Mickala[4], Stephan Ntie[4,5], Patrick Mickala[4], David Lehmann[5,6], Katharine Abernethy[6,7], Nicola Anthony[1]

1. *Department of Biological Sciences, University of New Orleans, 2000 Lakeshore Drive, New Orleans, Louisiana, 70148 USA*
2. *Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter, England, UK*
3. *School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, England, UK*
4. *Laboratoire de Biologie Moléculaire et Cellulaire (LABMC), Département de Biologie, Université des Sciences et Techniques de Masuku, BP 941, Franceville, Gabon*
5. *Agence Nationale des Parcs Nationaux, Rue Bana Ba Kengue, Quartier Haut de Gué Gué B.P. : 20379 Libreville, Gabon*
6. *Division of Biological and Environmental Sciences, University of Stirling, FK9 4LA, Scotland, UK*
7. *Institut de Recherche en Ecologie Tropicale, CENAREST, Gros Bouquet, Libreville, Gabon*

Corresponding author, Nicola Anthony, nanthony@uno.edu, ORCID 0000-0003-4878-517X

[†]ORCID ID of first author: 0000-0002-8370-3187

Abstract

The major histocompatibility complex (MHC) can be useful in guiding conservation planning because of its influence on immunity, fitness, and reproductive ecology in vertebrates. The mandrill (*Mandrillus sphinx*) is a threatened primate endemic to central Africa. Considerable research in this species has shown that the MHC is important for disease resistance, mate choice, and reproductive success. However, all previous MHC research in mandrills has focused on an inbred semi-captive population, so their genetic diversity may have been underestimated. Here we expand our current knowledge of mandrill MHC variation by performing next-generation sequencing of non-invasively collected fecal samples from a large wild horde in central Gabon. We observe MHC lineages and alleles shared with other primates, and we uncover forty-five putative new class II MHC DRB alleles, including representatives of the DRB9 pseudogene, which has not previously been identified in mandrills. We also document methodological challenges associated with fecal samples in NGS-based MHC research. Even with high read depth, the replicability of alleles from fecal samples was lower than that of tissue samples, and allele assignments are inconsistent between sample types. Further, the common assumption that variants with very high read depth should represent true alleles does not appear to be reliable for fecal samples. Nevertheless, the use of degraded DNA in the present study still enabled significant progress in quantifying immunogenetic diversity and its evolution in wild primates.

Keywords: Noninvasive samples, major histocompatibility complex, Illumina sequencing, replicability, *Mandrillus sphinx,* primates

Introduction

The genes of the major histocompatibility complex (MHC) have drawn considerable attention from conservationists over past decades due to their role in adaptive immunity (Manlik et al., 2019; Sommer, 2005). MHC loci encode cell-surface glycoproteins that bind antigenic peptides from pathogens and present them to T cells to initiate an immune response (Kaufman et al., 1984; Unanue, 1984). The amino acids in the antigen binding groove determine its binding properties, so variability in these amino acids enables immune responses against multiple pathogens (Matsumura et al., 1992; Ou et al., 1998). Class II MHC genes, which bind extracellular peptides, are under strong positive selection and include some of the most polymorphic loci in the vertebrate genome (Hughes & Nei, 1989; Klein et al., 1993; Radwan et al., 2020). These crucial functional genes have been used to study host-parasite co-evolution (Biedrzycka et al., 2018; Hedrick, 2002), delineate conservation units (Vásquez-Carrillo et al., 2014; Zhu et al., 2013), and infer migratory connectivity (Rodríguez et al., 2011).

MHC diversity has been well characterized in humans and some non-human primates, especially those used in medical research (Bontrop et al., 1999; Knapp et al., 1997; Otting et al., 2002). In humans, the most diverse MHC gene family, the class II DRB, includes nine lineages denoted DRB1-9 (Robinson et al., 2020). Of these, DRB1, 3, 4, and 5 are functional and highly variable, each containing hundreds or thousands of alleles (Klein et al., 2007). The lineages of the class II DRB predate primate speciation and are conserved across species (Geluk et al., 1993; Kelley et al., 2005; Slierendregt et al., 1992). This phenomenon, known as trans-species polymorphism (TSP), has been documented in multiple MHC markers, such as the DQB and DPB genes (Doxiadis et al., 2006; Otting et al., 2002; Song et al., 2016). Shared lineages have also been identified in primate species as distantly related as humans and owl monkeys (*Aotus* spp.) (Suárez et al., 2006), although it remains unclear whether this similarity is due to TSP or convergent evolution.

In addition to its role in immunity, the MHC also influences primate reproductive ecology. Heterozygote advantage has been documented in baboons (*Papio ursinus*) (Huchard et al., 2010) and rhesus macaques (*Macaca mulatta*) (Sauermann et al., 2001). Several primates use the MHC as a criterion for mate choice, preferring heterozygous or dissimilar mates, which can serve as an inbreeding avoidance mechanism (Dandine-Roulland et al., 2019; Huchard et al., 2013; Setchell et al., 2010). The MHC's influence on immunity and reproduction also makes it an important factor in conservation planning. An understanding of a species' MHC diversity provides information on local adaptation, vulnerability to parasites, and can guide decision-making for translocations or captive breeding programs (Sommer, 2005).

One primate in which the importance of the MHC has been clearly demonstrated is the mandrill (*Mandrillus sphinx*). Mandrill class II MHC DRB variation has previously been characterized in a single captive population located at the Centre International de Recherches Médicales (CIRMF) in Gabon (Abbott et al., 2006; Setchell et al., 2009). Male reproductive success increases with their MHC diversity, and females preferentially mate with dissimilar males (Setchell et al., 2010), resulting in higher immunogenetic variability for offspring (Setchell et al., 2013). MHC alleles are also related to the male's odor profile and dramatic red facial coloration (Setchell et al., 2009, 2011).

Evidence for the MHC's role in mandrill reproductive ecology is strong, but our understanding of their level of immunogenetic variability in the wild remains limited by the difficulties of sampling from

natural populations. Although MHC data have been obtained from many CIRMF individuals (n=155), the colony is inbred, having originated in 1983 from fifteen individuals that were confiscated from poachers and are of unknown origin (Charpentier et al., 2005; Wickings, 1995). Genetic diversity in the captive colony can be assumed to be less than a wild horde, which can number hundreds of individuals (Abernethy et al., 2002; Guibinga Mickala et al., 2022). Therefore, information that relies solely on the captive horde may provide an incomplete characterization of MHC diversity and thus hinder our ability to quantify adaptive diversity in wild populations and detect evolutionary patterns such as TSP.

Since mandrills are declining in numbers (Abernethy & Maisels, 2019) and are highly elusive in their dense forest habitat, collecting high-quality genetic samples from wild populations is problematic. Non-invasive sampling, such as collection of feces, is a possible alternative, since many samples can be collected with minimal contact with the target species. A prime opportunity for non-invasive sampling of wild mandrills exists in the forest-savannah mosaic in northern Lopé National Park (LNP), Gabon, which is home to a horde of nearly one thousand individuals (Abernethy et al., 2002; Guibinga Mickala et al., 2022). The mosaic habitat allows access to the resident horde, and since the 1980s, these mandrills have been the focus of research by staff at the Station d'Etudes des Gorilles et Chimpanzees (SEGC) (Abernethy et al., 2002; Guibinga Mickala et al., 2022; Harrison, 1988; Lahm, 1986; Rogers et al., 1996; Telfer et al., 2003; White et al., 2010). Some SEGC mandrills have been previously fitted with radio collars to aid in locating the group, simplifying non-invasive sampling of feces.

Despite their advantages, fecal samples generally contain degraded DNA and are prone to allelic dropout and cross-contamination (Morin et al., 2001; Taberlet et al., 1999). Fecal samples have been used successfully in past MHC studies, primarily using older sequencing methods such as Sanger sequencing (Arguello-Sánchez et al., 2018; Yu et al., 2018; Zhang et al., 2018), single-stranded conformation polymorphism (Maruya et al., 1996), or denatured gradient gel electrophoresis (Huchard et al., 2006; Knapp et al., 1997; Setchell et al., 2009). To date, Hans et al. (2015) remains the only case where MHC variation was characterized from wild primate fecal samples using targeted next-generation sequencing (NGS), allowing the most comprehensive description of gorilla MHC that was available at the time.

Although Hans et al. (2015) employed rigorous quality controls in their study, the reliability of MHC alleles identified from non-invasive samples using NGS has not been thoroughly evaluated. Several methods of assigning MHC alleles from thousands of NGS reads have been shown to have good replicability between sequencing runs (83.6-98% depending on the method) (Biedrzycka et al., 2017; Lighten et al., 2014; Million & Lively, 2022), but these measures have only been estimated using high quality samples such as tissue. Furthermore, some of the assumptions underlying these methods, namely that the sequences with the highest read depths within their amplicons should usually represent true alleles (Cummings et al., 2010; Lighten et al., 2014; Sommer et al., 2013), have never been tested using degraded DNA.

Here, we examine the reliability of MHC allele assignment from fecal samples using NGS by pursuing three objectives. First, we test whether allele assignments from fecal and tissue samples are equally replicable across repeated NGS runs. Second, we compare allele assignments from paired fecal and tissue samples from four individuals in a single sequencing run. Third, we explore the assumption

that very high depth sequences in an amplicon are more likely to represent genuine alleles than are low-depth sequences. If this assumption holds true in degraded samples, then variants with high read depth within their amplicons should be more repeatable than low-depth variants. Next, we expand the current characterization of the mandrill MHC to a wild population for the first time. We compare MHC variation in the wild to that of the previously-studied CIRMF colony, and we look for evidence of lineage-sharing between mandrills and other primates. Lastly, we quantify adaptive diversity by clustering alleles into functional supertypes.

Methods

Sample Collection

Fresh samples of feces (<6 hours) were collected daily from the SEGC horde during July or August of 2016 and 2017, when adult males and females are both present (Abernethy et al., 2002). Researchers located the horde using radio telemetry, then followed the horde on foot, collecting ~1 cm^3 of dung per sample. A total of 638 fecal samples were collected and stored in Falcon tubes with 25mL of silica, as recommended by previous studies (Soto-Calderón et al. 2009). Samples were also collected between 2016 and 2018 from 22 radio-collared mandrills. From these collared individuals, we collected fourteen blood samples, nine samples of plucked hair with attached bulb tissue, and nine fecal samples. DNA from fecal samples was extracted using the QIAamp DNA Stool Mini Kit (Qiagen, CA), while the DNeasy Blood & Tissue Kit (Qiagen, CA) was used for blood and hair samples.

PCR Amplification and Sequencing

Previous studies of mandrill MHC have used a single pair of primers that are expected to amplify a 252-base fragment from all MHC loci (Abbott et al., 2006; Setchell et al., 2005). Shorter fragments amplify more reliably from degraded DNA (Butler et al., 2003; Wiegand & Kleiber, 2001), so we designed a new pair of primers interior to the previously-used binding sites based on available mandrill MHC sequences. Our primers (forward 5'-TTCTTCAAYGGGACGGAGC-3', reverse 5'-GTGTCTGCAGTAGGTGTCC-3') amplify a 157-nucleotide fragment of the second exon of the DRB gene, encompassing close to 60% of the peptide binding region (PBR), an important target of selection (Brown et al., 1993). Illumina linker sequences were also added to the 5' end of each primer to facilitate library preparation and sequencing. PCRs for all samples were performed in a total volume of 25µl, with 12.5 µL 2X GoTaq HotStart Polymerase Master Mix (Promega), 0.5 mM $MgCl_2$, 400 nM of each primer, 8.25 µL of water, and 2 µL of template DNA. Cycling conditions were as follows: 5 minutes initial denaturation at 98°C, followed by 38 cycles of 95°C for 30 seconds, 65°C for 30 seconds, and 72°C for 60 seconds, and a ten-minute final extension at 72°C. Successful amplification of each sample was confirmed by agarose gel electrophoresis.

To quantify allele replicability and establish allele assignments for each sample, three Illumina runs were performed, all using 150 cycles of paired-end sequencing (Figure 1). Library preparation, which included addition of barcodes, sample pooling, and quantification, was performed at the Georgia Genomics and Bioinformatics Core. We aimed to achieve high read depth (≥5,000 reads) for most samples in each run, although targeting a specific depth is challenging in fecal samples due to their

variation in target DNA concentration. The first sequencing run, performed in 2018, was an Illumina MiSeq run of 192 pooled PCR products, each generated from a unique non-invasive sample. In 2019, a MiSeq Nano run was conducted using the samples of blood (n=14), hair (n=6), and feces (n=4) that successfully amplified from the radio-collared mandrills. The four fecal samples originated from four individuals from which blood was also collected, while the remaining ten blood samples and all six hair samples were all from different individuals. These 24 samples were pooled with replicate PCRs products for twenty-three randomly-selected non-invasive samples from the 2018 run. Finally, 192 replicate PCR products (comprising nine of the blood and hair samples and 183 previously sequenced noninvasive samples) were pooled into another standard MiSeq run in 2021. For repeatability calculations and development of consensus allele assignments, these three runs were considered in pairs as shown in Figure 1. A schematic of the overall workflow is shown in Figure 2.

Replicability of MHC allele assignments from feces and tissue

To quantify allele replicability between sequencing runs, data from each Illumina run was processed independently (Figure 2a). Trimmomatic and cutPrimers were used to trim primer and adapter sequences (Bolger et al., 2014; Kechin et al., 2017). Paired end reads were merged and cleaned using ampliMERGE and ampliCLEAN in the ampliSAT pipeline (Sebastian et al., 2016). Merged reads were discarded if their average Phred score was below 30 or if their length was outside the range of 152-160 bases. In ampliSAS (Sebastian et al., 2016), highly similar sequences were then clustered together using a 1% substitution error threshold, a 0.001% indel error threshold, and a 30% minimum dominant frequency threshold (MDF). To differentiate clusters representing true alleles from sequence artifacts, we then applied the degree of change (DOC) method (Lighten et al., 2014) as implemented in ampliSAS on amplicons containing at least 100 reads after filtering. In the DOC method, the read depths of each variant cluster in an amplicon are added cumulatively in descending order, and the first and second derivative of the resulting cumulative depth curve are used to identify a breakpoint in read depth that separates true alleles from sequence artifacts.

After applying the DOC method to all three Illumina datasets, we calculated a replicability score for each sample (R_A), defined as the proportion of variants in the sample that replicated between sequencing runs. Evidence for statistically significant differences in the average R_A values calculated from replicated blood/hair and fecal samples was assessed using Mann-Whitney U tests. Specifically, comparison were made between datasets B and C, A and C, and C and D, as defined in Figure 1. Because datasets B and D have very different sample sizes (n=9 and n=183 R_A scores respectively), their average R_A scores were compared by randomly selecting 1000 sets of nine values from D to generate a distribution of scores. Statistical significance was assessed based on whether the average value of B was greater than the 95th percentile of the permuted distribution of scores from non-invasive samples. The proportions of blood/hair and fecal samples with "perfect" R_A scores (R_A=1) were also compared using Fisher's Exact Test. Finally, within-run R_A scores were also calculated between the paired blood and fecal samples that were included in the MiSeq Nano run.

Since the DOC method has been shown to have lower replicability in amplicons with less than 5,000 reads (Biedrzycka et al., 2017), the R_A scores for dataset D were also calculated excluding

amplicons below that read depth. As an alternative to the DOC method, we also assigned alleles to the samples in the 2018 and 2021 runs using the default assignment method implemented in AmpliSAS (Sebastian et al., 2016). Replicability scores in dataset D were then recalculated from these allele assignments.

Replicability of high-depth sequence variants

Next, we explored whether a variant's read depth is associated with its status as an allele or an artifact in degraded samples, following the assumption that true alleles should be more repeatable than artifacts across runs. We focused on non-invasive samples for this assessment and therefore used only the 2018 and 2021 MiSeq runs, which contain the bulk of the fecal samples. For each amplicon in each run, alleles were ranked according to their depth, and each allele's relative frequency was calculated (allele depth divided by total amplicon depth). We then combined these data from all alleles in both runs and split the data into a training set and a testing set (80% and 20% of the alleles, respectively). Logistic regressions were performed on the training sets to determine whether variant rank or frequency are significant predictors of whether a variant replicates across runs. Because variant rank and frequency are closely related, variables were tested in separate models. The strength of each variable's influence was assessed by calculating odds ratios, and the predictive power of the model was quantified by calculating the proportion of variants in the test set for which the model correctly predicted replication status.

Individual Allele Assignment

Due to the poor replicability observed when Illumina runs are processed independently (see Results), we devised a method to generate consensus allele assignments from paired runs (Figure 2b). Our strategy relies on several assumptions. First, we assumed that given high amplicon sequencing depth, all alleles should be detected even if at very low read depth. Past studies show that some alleles amplify less efficiently than others, resulting in consistently lower read depths (Sommer et al., 2013). True alleles may therefore appear at depths lower than putative sequence artifacts. This problem may be exacerbated in fecal samples because they tend to be highly degraded and subject to allelic dropout (Morin et al., 2001), leading to poor replicability. However, given high read depths for each amplicon, we assume detection of all alleles, even if depth for a particular allele is very low (<1% of amplicon depth). Second, artificial variants resulting from random single-base substitution or indels are unlikely to replicate at an appreciable depth between runs. Some bases are more error-prone than others (Gilles et al., 2011), resulting in a higher probability of replicable errors at these positions. However, because artifacts are clustered with their parent sequences in the ampliSAT pipeline (Sebastian et al., 2016), repeatable errors are still unlikely to reoccur in two independent runs. Third, chimeras and indels in homopolymeric regions may be more likely to replicate between runs, but they can be easily identified and removed after alignment to the parent sequence(s). Fourth, alleles resulting from cross-contamination between samples during PCR are unlikely to be replicated, as PCR products for each Illumina run were prepared independently.

Considering these four assumptions, our strategy to assign individual MHC alleles was as follows (Figure 2b): each Illumina dataset was reanalyzed using the ampliSAS tool (Sebastian et al., 2016), using

less stringent clustering and filtering parameters. Data presented here (see Results) shows that low-depth variants are sometimes repeatable, suggesting that they may be true alleles, so we aimed to initially retain these variants by using more relaxed parameters. The chosen settings are still expected to cluster most artificial variants resulting from substitution or indel errors with their parent alleles. Substitution, indel, and minimum dominant frequency thresholds were therefore set to 1%, 0.01%, and 10% respectively. Minimal filtering parameters were applied, retaining variants with relative frequency greater than 0.05%. For each sample, the ampliSAS-generated fasta files of variants from two runs were then entered into a custom Python program that extracted all replicated variants. For each non-invasive sample, replicated variants were extracted from the 2018 and the 2021 runs. For the blood and hair samples, variants from the 2019 and 2021 runs were used. Resulting variants were then aligned in MEGA-X (Kumar et al., 2018), where replicated artifacts resulting from indels in homopolymer regions were easily identified by their introduction of alignment gaps relative to otherwise-identical parent sequences. Replicated chimeras were also detected visually and by generating neighbor-joining trees for each sample, since chimeras result from recombination of parent sequences and contain no unique mutations. After removing replicated sequence artifacts, all remaining variants were considered putative true alleles.

Our method is similar to that of Sommer et al., (2013) in that we rely on paired sequencing runs. A key difference is that the previous approach assumes that the variant with the greatest read depth in each amplicon is a true allele, and all other alleles in the sample are identified based on their replicability, relative frequency, and their similarity to other variants. Our data shows that even the highest-depth variant in an amplicon may not be replicable (see Results), so our method makes no assumptions about relative read depth indicating allele veracity.

An unrooted neighbor-joining tree was then generated in MEGA-X (Kumar et al., 2018), using alleles identified in the SEGC horde as well as previously-published mandrill MHC DRB sequences downloaded from the Immuno-Polymorphism Database (IPD) (Maccari et al., 2020). When possible, lineages of novel variants were identified based on monophyly with previously described DRB loci (Abbott et al., 2006; Setchell et al., 2009). Because our sequences do not include the full exon, novel alleles were named arbitrarily by number within their lineages instead of following the nomenclature described by de Groot et al. (2019). Here, the putative lineage for novel alleles is given a prefix “p,” so, for instance, alleles that form a monophyletic group with the DRB3 lineage would be named pDRB3-1, pDRB3-2, and so forth. Alleles that do not form a monophyletic group with a known lineage are designated pDRB, and sequences that are identical to a previously described allele are given the same name, with the addition of the “p” prefix.

Assessing lineages shared with other primates

To look for patterns of lineage-sharing between mandrills and other primates, and to further validate the identity of the alleles, each sequence was subjected to BLAST searches in IPD (Maccari et al., 2020) and NCBI’s Genbank (Clark et al., 2016). Evolutionary relationships between lineages were visualized by generating an unrooted neighbor-joining tree using MHC-DRB sequences from mandrills and three well-studied species: olive baboons (*Papio anubis*), crab-eating macaques (*Macaca fascicularis*), and chimpanzees (*Pan troglodytes*).

Sequencing short amplicons can cause an underestimate of true diversity (Llaurens et al., 2012), for instance, if the fragment length is insufficient to capture polymorphisms differentiating alleles. To quantify the amount of diversity lost due to our use of a short amplicon, all MHC-DRB alleles were downloaded from the IPD for 15 species of cercopithecoids, platyrrhines, and apes (Supplementary Table 1). The alleles were aligned with those from the present study and trimmed to equal length. We counted the number of unique alleles before and after trimming to estimate the loss of allelic diversity.

Assessing functional diversity

Alleles were classified into supertypes following Doytchinova and Flower (2005). A supertype represents a group of alleles with similar physio-chemical properties in the amino acids comprising the antigen binding sites (ABS). Because ABS diversity is functionally important, amino acid sites involved in peptide binding are likely under positive selection. To identify these positively-selected sites (PSS), per-codon signatures of positive selection were evaluated by comparing rates of synonymous (dS) and nonsynonymous (dN) mutations. The dN/dS comparison was performed using MEME (Murrell et al., 2012) and FUBAR (Murrell et al., 2013) within the DataMonkey server (Weaver et al., 2018). MEME compares the dN and dS rates at each site, accounting for the possibility that strength of selection may vary across phylogenetic branches, while FUBAR assumes pervasive selection. The positions of ABS codons are likely to be similar within a species, so to increase statistical power, this analysis was applied to a dataset including both novel and previously-published alleles.

Codons under positive selection were then extracted from the sequences. We quantified each allele's binding properties at the PSS using five physiochemical measurements (z-scores): z1 (hydrophobicity), z2 (steric bulk), z3 (polarity) and z4 and z5 (electronic effects) (Sandberg et al., 1998). Z-scores were then transformed by principal components analysis and alleles classified into supertypes using a Discriminant Analysis of Principal Components (DAPC) (Jombart et al., 2010) in the R package adegenet (Jombart & Ahmed, 2011). This classification method maximizes the between-group z-score variance while minimizing within-group variance, generating allele groups with functionally distinct PSS.

Results

Replicability of MHC allele assignments from feces and tissue

After filtering, 2018 and 2021 MiSeq runs had comparable read depths, although nearly 60% of the reads from the 2021 run were lost during filtering (Table 1). As an expected consequence of the sequencing technology used, the 2019 MiSeq Nano run had much lower read depth. According to a Fisher Exact test comparing allele counts in each run, the 2021 run detected fewer alleles per sample than the 2018 run ($p=0.005$), although there was no significant difference in the number of alleles detected between the 2018 MiSeq and the 2019 MiSeq Nano runs ($p=0.86$) or the 2019 and 2021 runs ($p=0.07$).

Replicate sequence data were obtained from 181 of the 183 non-invasive samples sequenced in the 2018 and 2021 MiSeq runs. Twenty-two non-invasive samples and nine blood/hair samples sequenced successfully in both the 2019 and 2021 runs. The average R_A score from the nine blood/hair samples ($R_A=0.76$, dataset B in Figure 1 and Table 2) is higher than the entire distribution of average

scores calculated from 1000 randomly-drawn sets of nine non-invasive samples from dataset D (maximum R_A from permuted data=0.69), indicating a significant difference in replicability. The blood/hair samples also had a significantly higher percentage of perfect R_A scores (dataset B, 55.6%) than the replicated non-invasive samples (dataset C, 13.6%) (Fisher Exact Test, p=0.016) (Table 2). The 22 non-invasive samples sequenced in both the 2019 and 2021 runs also had significantly lower R_A scores (dataset C, R_A=0.31) than the blood/hair samples (dataset B, R_A=0.76) (p=0.003) (Table 2). There was no significant difference in R_A scores between any of the sets of non-invasive samples (datasets A and C, and C and D, in Figure 1 and Table 2).

For the four mandrills with blood and fecal samples sequenced in the MiSeq Nano run, allele assignments varied by sample type. Within-run R_A scores between paired blood and fecal samples were 0, 0.33, 0.33, and 1. Blood samples also obtained between 2.9 and 4 times more read depth than did the fecal samples (read depth range=3368-4489 and 962-1535 for blood and fecal samples respectively).

When R_A scores from dataset D were calculated using only amplicons with >5,000 reads (n=154 from dataset D), replicability levels did not differ (mean R_A = 0.32). Average scores were also very similar when alleles were assigned using AmpliSAS's built-in assignment method (mean R_A = 0.35).

Replicability of high-depth variants

In logistic regressions, both variant rank and frequency were significant predictors of variant replication (p<0.001 for each model). For each step downward in variant rank, for instance, from the highest depth to the second highest depth, variants were 25.9% (95% CI: 20.8-30.8%) less likely to replicate across runs. Unsurprisingly, similar results were apparent from the model using variant frequency as a predictor. For every 1% increase in a variant's within-amplicon frequency, that variant is on average 3.1% (95% CI: 2.4%-3.8%) more likely to replicate in a subsequent run. In the test dataset, the models successfully predicted whether or not a variant would replicate in 65.1% and 65.5% of cases when using rank and frequency respectively.

Despite these associations, the variants with the highest depth in each amplicon in the 2018 run failed to replicate in 36.8% of amplicons in the 2021 run, and the top two variants failed to re-occur 45.6% of the time. In histograms of variant frequency and replication status (Figure 3), it is evident that, although many low-frequency variants fail to replicate, an appreciable number do re-occur in subsequent runs.

MHC characterization from consensus genotypes

Of the 181 non-invasive fecal samples that successfully replicated in the 2018 and 2021 runs, consensus allele assignments could be generated for 170 samples, representing at least 162 individuals based on microsatellite data in an unrelated study (Guibinga Mickala et al., 2022). For the remaining 11 samples, no variants replicated between runs, despite the relaxed clustering and filtering parameters used for this step. All nine replicated blood and hair samples had replicated variants and could be assigned alleles.

Sixty-two alleles were detected in the SEGC horde, 17 of which match those reported in the CIRMF semi-captive mandrills (n=155) (Abbott et al., 2006; Setchell et al., 2009) (Supplementary Fig. 1). We found a maximum of 11 alleles per sample, with an average of 3.62 (± 1.88), suggesting up to six MHC-DRB loci. The most common allele (Masp-pDRB9-1) appeared in 111 samples (58.4%). Eighteen alleles were only identified in a single animal, but in accordance with our assignment procedure, they were considered true alleles because they replicated across runs. One of these matched the previously published mandrill allele Masp-DRB1*04:02 (Abbott et al., 2006), lending support to the validity of these singletons (Supplementary Fig. 1).

When these alleles were combined in a neighbor-joining tree with 23 previously-published alleles (Abbott et al., 2006; Setchell et al., 2009), two monophyletic groups were apparent (Figure 4). The first ("clade 1") contains 63 alleles with unique amino acid sequences, representing DRB lineages 1, 3, 5, and 6. All previously-described mandrill alleles belong to this clade. The second monophyletic group ("clade 2") includes 22 alleles that are highly divergent from clade 1, with 20 unique amino acid sequences. In BLAST searches of the IPD (Maccari et al., 2020), clade 2 alleles closely match the primate DRB9 lineage. Masp-pDRB9-1, the most common allele in the dataset (Supplementary Fig. 1), shares 97% identity with the crab-eating macaque DRB9 (*Macaca fascicularis)* (GenBank accession number MW679616.1). Individuals possessing Masp-pDRB9-1 have on average 0.96 more alleles than those that do not, a significant difference according to a Mann-Whitney U test (Z=0.03, p=0.002). All clade 2 alleles contain a single nucleotide deletion between exon sites 82 and 83 and an insertion at position 170. Two alleles, Masp-pDRB9-12 and Masp-pDRB9-21, contain a stop codon caused by a G to T substitution at amino acid sites 35 and 53 respectively, and an additional 2-base deletion. Seven clade 2 sequences have deletions of 1-3 bases. By contrast, clade 1 contains no deletions, and the only stop codon present is in a sequence belonging to the DRB6 pseudogene (Bontrop et al., 1999).

Lineage-sharing between primate species

Lineage-sharing between mandrills and olive baboons, chimpanzees, and crab-eating macaques can be observed in all DRB loci (Figure 5). Some alleles cluster more closely by DRB lineage than by species, although considerable paraphyly exists.

Of the 62 sequences, four are identical to alleles in other primates. Thirty mandrill samples possessed the sequence Masp-pDRB-17, which matches Paan-DRB*W001:02 (IPD accession NHP04483), previously identified in olive baboons. Two individuals carry a DRB6 allele fragment, designated pDRB6-2, from the same species (Paan-DRB6*01:05, IPD accession NHP04517). A sequence(Masp-pDRB5-3) matching crab-eating macaque's Mafa-DRB5*03:02 (IPD accession NHP00322) (Blancher et al., 2006; Leuchte et al., 2004), appears in two mandrill samples. Finally, one sample includes Masp-pDRB5-2, which matches Patr-DRB5*03:11 (IPD accession NHP00891), previously found in chimpanzees (Fan et al., 1989; Kenter et al., 1992).

After downloading MHC-DRB sequences from fifteen nonhuman primates and counting the number of unique sequences before and after trimming, we determined that an average of 13.46% (±12.18) of each species' allelic diversity was lost by using the shorter amplicon (Supplementary Table 1). When the shortened sequences (n=856 alleles from all species combined) were compared across

species, 684 unique alleles were found. However, alleles identical across the shortened length were almost always between species in the same genus. The only exception was between gorillas (*Gorilla gorilla*) and chimpanzees, for which alleles Gogo-DRB6*01:03N and Patr-DRB6*01:05N were identical after trimming. While some allelic diversity within mandrills was likely lost due to our use of a short amplicon, the shared sequences between mandrills and baboons, macaques, and chimpanzees are unusual and may be indicative of biological phenomena.

Functional diversity

Clade 1 and clade 2 alleles may be subjected to different selective pressures and so were analyzed separately in MEME. The pseudogenic DRB6 sequences were also excluded. There was no evidence for selection on the clade 2 alleles, which is characteristic of loss of functionality. In the clade I alleles, nine codons had significantly higher rates of nonsynonymous than synonymous mutations ($p<0.1$), indicating positive selection. The nine sites (amino acids 5, 9, 15, 33, 34, 41, 47, 48, and 51) correspond to DRB exon sites 28, 32, 38, 56, 57, 64, 70, 71, and 74. All of these, with the exception of exon site 28, were also identified by the program FUBAR (Murrell et al., 2013). Seven of these PSS match the ABS identified in humans using x-ray crystallography (Figure 6) (Brown et al., 1993). According to the DAPC, these sequences cluster into six functional supertypes containing 5-14 alleles each (Figure 4). Wild mandrills possess an average of 1.9 (±1.03) supertypes each, with a maximum of five per individual.

Discussion

Replicability of MHC sequencing

We found that between-run repeatability of MHC alleles is poor in fecal samples compared to blood and plucked hair. Furthermore, paired blood and fecal samples from the same individuals did not yield equivalent assignments, although only four pairs were compared, and read depth was low for these samples. Studies have shown that replication is critical to producing reliable data from degraded samples (Morin et al., 2001; Taberlet et al., 1996), and it is unsurprising that the same would apply to NGS. More research is needed to determine a suitable number of replicates for reliable allele assignments from non-invasive samples, and an appropriate method to differentiate alleles from artifacts.

One caveat to this study is that the technical variation between the runs used for replicability tests may have reduced replicability scores overall. The 2021 run appeared to be of lower quality, given the high proportion of reads lost during filtering and the identification of fewer alleles per sample after DOC processing. The 2019 run used MiSeq Nano, which produces lower read depth than a MiSeq run. However, the 2019 and 2021 runs included both feces and blood/hair, so any negative bias in R_A scores would have affected all sample types. Indeed, the scores for blood/hair samples are lower than previously documented, although this could also be a result of stochasticity due to the small sample size (Biedrzycka et al., 2017; Lighten et al., 2014). However, because all samples were affected, the reduced R_A scores of fecal samples relative to blood/hair is still meaningful.

We also found that a variant's replication across runs is related to its within-amplicon rank and frequency. However, low-frequency sequences were sometimes repeatable, while the highest-ranking

variants frequently failed to reappear in a second run. If replicability reflects a variant's status as an allele or an artifact, then neither frequency nor rank appear to be particularly useful in classifying sequences from non-invasive samples. However, all MHC genotyping methods thus far have relied on these factors (Babik et al., 2009; Grogan et al., 2016; Lighten, van Oosterhout, Paterson, et al., 2014; Sommer et al., 2013; Stutz & Bolnick, 2014), generally assuming that the top-ranking variant in an amplicon is a true allele. While this assumption may be valid for high-quality samples, it appears unsubstantiated for degraded DNA. If these methods were applied to a single run of non-invasive samples, the error rate would likely be very high. Our consensus allele-calling method circumvents this issue, since it considers all replicated variants regardless of depth as potential alleles before error screening.

The high read depth in this study allowed the identification of poorly-amplifying alleles, but depth also leads to a caveat. Biedrzycka et al. (2017) showed that consensus assignment methods relying on sequence replication (Sommer et al., 2013) are prone to a high false discovery rate when read depths are high, presumably because higher depth increases the probability of recovering replicated errors. We observed such artifacts, but they were removed from the consensus during error-screening. Furthermore, if false allele discovery occurred commonly in high-depth amplicons, then the number of alleles included in a sample's consensus should increase with sample depth. We found no evidence of such a trend (data not shown).

MHC diversity in wild mandrills

The described allele assignment method requires validation using individuals with known allele assignments, so our individual assignments must be interpreted cautiously. However, we have confidence in the alleles identified, and our results yield interesting conclusions.

Twice as many alleles were identified in the wild population compared to the captive horde, although only six supertype clusters were formed compared to eleven in a previous study (Setchell et al., 2009, 2010). This difference could be due to our novel primers and next-generation sequencing. The lower allelic diversity in the captive colony may also be due to their documented inbreeding depression (Charpentier et al., 2005). We identified up to eleven alleles per animal, four more than in the captive horde (Setchell et al., 2009). This finding is partially explained by the "clade 2 alleles," which account for up to three alleles per animal and are characterized by mutations that may disrupt transcribed amino acids. The compensatory indels observed at base 82 and 170 have also been identified in human DRB9 alleles (Haas et al., 1987), supporting our classification of this clade as DRB9. This lineage may contribute to the copy number variation commonly observed in the MHC (Bontrop, 2006; Doxiadis et al., 2012), since the average allele count of individuals possessing the Masp-pDRB9-1 allele is higher by almost one. This finding suggests that this allele may be fixed on certain haplotypes, which could explain its frequent occurrence.

The DRB9 is an ancient MHC fragment and has been amplified in gorillas, orangutans, chimpanzees, and crab-eating macaques (Gongora, Figueroa, and Klein 1996), although publicly-available sequences remain limited. Here, the lineage may have been revealed by our use of novel primers. Previous studies of mandrill MHC have used the same primers that were thought to amplify all

DRB loci across primates (Abbott et al., 2006; Setchell et al., 2009). However, in the DRB9 sequence from the crab-eating macaque, the binding site for the reverse primer used by Setchell et al. (2009) contains mutations that would likely preclude primer binding at the 3' end. By contrast, our primer pair is complementary to alleles in clade 1 (DRB1-6) and clade 2 (DRB9) in macaques and mandrills. This suggests that using only one pair of primers may underestimate MHC diversity (see also Llaurens et al., 2012).

We have also noted four sequences that are shared between mandrills and olive baboons (Masp-pDRB-17 and Masp-pDRB6-2), crab-eating macaques (Masp-pDRB5-3), and chimpanzees (Masp-pDRB5-2). Mutations may have occurred outside the fragment sequenced here, but the shared sequences are more conserved than is typically found across genera, according to our analyses of publicly available MHC data. These similarities could be explained by TSP, which is known to occur in the primate MHC (Geluk et al., 1993; Suárez et al., 2006; Yasukochi & Satta, 2014).

However, one of the shared sequences, Masp-pDRB6-2, is likely a pseudogene, having lost its function a very long time ago (Figueroa et al., 1991). It is unclear why this allele would have been preserved through positive selection, but pseudogenes may instead be maintained through balancing selection. The MHC accumulates recessive deleterious mutations as a "sheltered load" (van Oosterhout, 2009). These mutations are rarely expressed in the homozygous condition and can accumulate without being removed through purifying selection. When a locus evolves into a pseudogene and positive selection relaxes, purifying selection may continue to act on the peri-MHC region. This could retain the same haplotypes over evolutionary time and explain TSP of non-functional pseudogenes (van Oosterhout, 2009).

Another possible explanation for the sequences shared with olive baboons (pDRB-17 and pDRB6-2) is historical introgression. Introgression of baboons (*Papio hamadryas*) has been documented in geladas (*Theropithecus gelada*) (Dunbar & Dunbar, 1974) and kipunji (*Rungwecebus kipunji*) (Roberts et al. 2010; Zinner, Arnold, and Roos 2009). As central African forest cover has shifted over the millennia (Maley, 1996; Maley et al., 2018), mandrill and baboon habitats may have overlapped. The two species are group-living and primarily terrestrial (Abernethy et al., 2002; Higham et al., 2009), and they may be infected by similar pathogens. Since MHC alleles are thought to be under pathogen-mediated selection (Spurgin & Richardson, 2010), if a beneficial allele in baboons was introduced to the mandrill population, positive selection may increase that allele's frequency. Although this is a highly speculative hypothesis, it could explain the incidence of the pDRB-17 allele in mandrills.

Future directions

This study highlights the advantages and challenges of non-invasive genetic research. Fecal samples enabled MHC characterization from many wild individuals, but technical obstacles remain that inhibit effective use of targeted NGS with degraded samples. Optimization of sampling, laboratory, and sequencing procedures may be required, and the importance of replication in generating allele assignments merits further investigation. Up to eight replicates have been encouraged for microsatellite data from fecal samples (Taberlet et al., 1996), but this number would be cost-prohibitive for many NGS studies. It remains to be seen whether two replicates are sufficient, and the inclusion of within-run

replicates would be beneficial in future studies to rule out potential batch effects during sequencing. Our procedure should also be validated using an individual with known MHC alleles, and such an individual was unavailable in this study.

Alternative sequencing approaches may also be explored. Genotyping-in-thousands by sequencing (GT-seq) (Campbell et al., 2015) and targeted capture enrichment (Fontsere et al., 2021) show promise for non-invasive samples (Hayward et al., 2022; Natesh et al., 2019; Schmidt et al., 2020), but these may be less useful for *de novo* sequencing of specific MHC loci. Portable sequencers such as the Oxford Nanopore's MinION offer the potential to sequence samples in the field, which may improve data quality since fecal DNA degrades with time until extraction (Soto-Calderón et al. 2009). The MinION's PCR-free technology also eliminates PCR-generated sequence errors (Payne et al., 2021; Wanner et al., 2021). However, the MinION sequencer has a high error rate otherwise (Loit et al., 2019), and further optimization for fecal samples would be required.

Despite problems with replicability, lineage-sharing was still detected. The degree to which the nonfunctional DRB9 is preserved across primates warrants investigation, as it is likely widespread despite having been reported in only a few species. Identical allele fragments shared between mandrills and other primates should also be confirmed through sequencing of the entire DRB region, as these alleles may have implications for the evolution of disease resistance in primates.

Our study focuses on the southern mandrill subpopulation, but collecting MHC data from the genetically-distinct mandrill population north of the Ogooué River may reveal insights into their evolution and historical introgression with olive baboons (Telfer et al., 2003). Northern mandrills are in closer proximity to the baboon range and thus may be more likely to have been involved in introgression events. If introgression has occurred, baboon genetic material may also be detectable in northern mandrill populations. MHC alleles could have been passed to the southern population through historical immigration, but nuclear gene flow between these populations has never been assessed.

Although the exact role of MHC diversity in species viability is not well understood (Radwan et al., 2010), our findings of diverse DRB genes bode well for the species' ability to adapt to shifting pathogen pressures over time. Our results also reveal interesting evolutionary relationships between mandrills and other primates, and studies of other populations will be integral to uncovering the roles of TSP and potential introgression in the evolution of primate MHC.

Acknowledgements

We deeply appreciate the aid we received from the staff of the Station d'Etudes des Gorilles et Chimpanzés (SEGC), Gabon, who acted as field guides and assisted in sample collection. We thank the many undergraduate assistants who helped with lab work for this project: Ibraheem Hachem, Justine Davis, Shyla Irthum, Kaleb Hill, Gina Kissee, Claire Melancon, Patrick Hall and Gabrielle Sehon. We also thank the Agence National des Parcs Nationaux (ANPN) and the Centre National de la Recherche Scientifique et Technologique (CENAREST) for research permits No. AR0023/17/MESRSFC/CENAREST/CG/CST/CSAR; AR0036/16/MESRSFC/CENAREST/CG/CST/CSAR and

LNP entry permits No. AE17 O 1 6/PR/ANPN/SE/CS/AFKP; AE16025/PR/ANPN/SE/CS/AFKP. We also thank the Audubon Nature Institute for their support.

References

Abbott, K. M., Wickings, E. J., & Knapp, L. A. (2006). High levels of diversity characterize mandrill (Mandrillus sphinx) Mhc-DRB sequences. *Immunogenetics*, *58*(8), 628–640. https://doi.org/10.1007/s00251-006-0132-3

Abernethy, K., & Maisels, F. (2019). *Mandrillus sphinx*. The IUCN Red List of Threatened Species. https://www.iucnredlist.org/species/12754/17952325

Abernethy, K., White, L. J. T., & Wickings, E. J. (2002). Hordes of mandrills (Mandrillus sphinx): Extreme group size and seasonal male presence. *Journal of Zoology, Proceedings of the Zoological Society of London*, *258*, 131–137.

Arguello-Sánchez, L. E., Arguello, J. R., García-Feria, L. M., García-Sepúlveda, C. A., Santiago-Alarcon, D., & Espinosa de los Monteros, A. (2018). MHC class II DRB variability in wild black howler monkeys (Alouatta pigra), an endangered new world primate. *Animal Biodiversity and Conservation*, *41*(2), 389–404. https://doi.org/10.32800/abc.2018.41.0389

Babik, W., Taberlet, P., Ejsmond, M. J., & Radwan, J. (2009). New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources*, *9*(3), 713–719. https://doi.org/10.1111/j.1755-0998.2009.02622.x

Biedrzycka, A., Bielański, W., Ćmiel, A., Solarz, W., Zając, T., Migalska, M., Sebastian, A., Westerdahl, H., & Radwan, J. (2018). Blood parasites shape extreme major histocompatibility complex diversity in a migratory passerine. *Molecular Ecology*, *27*(11). https://doi.org/10.1111/mec.14592

Biedrzycka, A., Sebastian, A., Migalska, M., Westerdahl, H., & Radwan, J. (2017). Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. *Molecular Ecology Resources*, *17*(4), 642–655. https://doi.org/10.1111/1755-0998.12612

Blancher, A., Tisseyre, P., Dutaur, M., Apoil, P. A., Maurer, C., Quesniaux, V., Raulf, F., Bigaud, M., & Abbal, M. (2006). Study of Cynomolgus monkey (Macaca fascicularis) MhcDRB (Mafa-DRB) polymorphism in two populations. *Immunogenetics*, *58*(4), 269–282. https://doi.org/10.1007/s00251-006-0102-9

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bontrop, R. E. (2006). Comparative Genetics of MHC Polymorphisms in Different Primate Species: Duplications and Deletions. *Human Immunology*, *67*(6), 388–397. https://doi.org/10.1016/j.humimm.2006.03.007

Bontrop, R. E., Otting, N., De Groot, N. G., & Doxiadis, G. G. M. (1999). Major Histocompatibility Complex Class II polymorphisms in primates. *Immunological Reviews*, *167*, 339–350. https://doi.org/10.1016/B978-0-12-374994-9.10001-4

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L., & Wiley, D. C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*, *364*(6432), 33–39. https://doi.org/10.1038/364033a0

581 Butler, J. M., Shen, Y., & Mccord, B. R. (2003). The development of reduced size STR amplicons as tools
582 for analysis of degraded DNA. *Journal of Forensic Science*, *48*(5), 1054–1064.

583 Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-Thousands by sequencing (GT-seq):
584 A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology*
585 *Resources*, *15*(4), 855–867. https://doi.org/10.1111/1755-0998.12357

586 Charpentier, M., Setchell, J. M., Prugnolle, F., Knapp, L. A., Wickings, E. J., Peignot, P., & Hossaert-McKey,
587 M. (2005). Genetic diversity and reproductive success in mandrills (Mandrillus sphinx). *Proceedings*
588 *of the National Academy of Sciences of the United States of America*, *102*(46), 16723–16728.
589 https://doi.org/10.1073/pnas.0507205102

590 Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2016). GenBank. *Nucleic Acids*
591 *Research*, *44*(D1), D67–D72. https://doi.org/10.1093/nar/gkv1276

592 Cummings, S. M., McMullan, M., Joyce, D. A., & van Oosterhout, C. (2010). Solutions for PCR, cloning
593 and sequencing errors in population genetic analysis. *Conservation Genetics*, *11*(3), 1095–1097.
594 https://doi.org/10.1007/s10592-009-9864-6

595 Dandine-Roulland, C., Laurent, R., Dall'Ara, I., Toupance, B., & Chaix, R. (2019). Genomic evidence for
596 MHC disassortative mating in humans. *Proceedings. Biological Sciences*, *286*(1899), 20182664.
597 https://doi.org/10.1098/rspb.2018.2664

598 de Groot, N. G., Otting, N., Maccari, G., Robinson, J., Hammond, J. A., Blancher, A., Lafont, B. A. P.,
599 Guethlein, L. A., Wroblewski, E. E., Marsh, S. G. E., Shiina, T., Walter, L., Vigilant, L., Parham, P.,
600 O'Connor, D. H., & Bontrop, R. E. (2019). Nomenclature report 2019: major histocompatibility
601 complex genes and alleles of Great and Small Ape and Old and New World monkey species.
602 *Immunogenetics*. https://doi.org/10.1007/s00251-019-01132-x

603 Doxiadis, G. G. M., De Vos-Rouweler, A. J. M., De Groot, N., Otting, N., & Bontrop, R. E. (2012). DR
604 haplotype diversity of the cynomolgus macaque as defined by its transcriptome. *Immunogenetics*,
605 *64*(1), 31–37. https://doi.org/10.1007/s00251-011-0561-5

606 Doxiadis, G. G. M., Rouweler, A. J. M., De Groot, N. G., Louwerse, A., Otting, N., Verschoor, E. J., &
607 Bontrop, R. E. (2006). Extensive sharing of MHC class II alleles between rhesus and cynomolgus
608 macaques. *Immunogenetics*, *58*(4), 259–268. https://doi.org/10.1007/s00251-006-0083-8

609 Doytchinova, I. A., & Flower, D. R. (2005). In Silico Identification of Supertypes for Class II MHCs. *The*
610 *Journal of Immunology*, *174*(11), 7085–7095. https://doi.org/10.4049/jimmunol.174.11.7085

611 Dunbar, R. I. M., & Dunbar, P. (1974). On hybridization between Theropithecus gelada and Papio anubis
612 in the wild. *Journal of Human Evolution*, *3*(3), 187–192. https://doi.org/10.1016/0047-
613 2484(74)90176-6

614 Fan, W., Kasahara, M., Gutknecht, J., Klein, D., Mayer, W. E., Jonker, M., & Klein, J. (1989). Shared class II
615 MHC polymorphisms between humans and chimpanzees. *Human Immunology*, *26*(2), 107–121.
616 https://doi.org/10.1016/0198-8859(89)90096-7

617 Figueroa, F., Colm, O., Inoki, H., & Klein, J. (1991). Primate DRB6 pseudogenes: clue to evolutionary
618 origin of the HLA-DR2 haplotype. *Immunogenetics*, *34*, 324–337.

619 Fontsere, C., Frandsen, P., Hernandez-Rodriguez, J., Niemann, J., Scharff-Olsen, C. H., Vallet, D., Le
620 Gouar, P., Ménard, N., Navarro, A., Siegismund, H. R., Hvilsom, C., Gilbert, M. T. P., Kuhlwilm, M.,

621 Hughes, D., & Marques-Bonet, T. (2021). The genetic impact of an Ebola outbreak on a wild gorilla
622 population. *BMC Genomics*, *22*(1), 1–12. https://doi.org/10.1186/s12864-021-08025-y

623 Geluk, A., Elferink, D. G., Slierendregt, B. L., Van Meijgaarden, K. E., Vries, R. R. P., Ottenhoff, T. H. M., &
624 Bontrop, R. E. (1993). Evolutionary conservation of major histocompatibility complex-DR/peptide/T
625 cell interactions in primates. *Journal of Experimental Medicine*, *177*(4), 979–987.
626 https://doi.org/10.1084/jem.177.4.979

627 Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T., & Martin, J. F. (2011). Accuracy and quality
628 assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics*, *12*.
629 https://doi.org/10.1186/1471-2164-12-245

630 Gongora, R., Figueroa, F., & Klein, J. (1996). The HLA-DRB9 gene and the origin of HLA-DR haplotypes.
631 *Human Immunology*, *51*(1), 23–31. https://doi.org/10.1016/S0198-8859(96)00189-9

632 Grogan, K. E., McGinnis, G. J., Sauther, M. L., Cuozzo, F. P., & Drea, C. M. (2016). Next-generation
633 genotyping of hypervariable loci in many individuals of a non-model species: Technical and
634 theoretical implications. *BMC Genomics*, *17*(1), 1–16. https://doi.org/10.1186/s12864-016-2503-y

635 Guibinga Mickala, A., Weber, A., Ntie, S., Gahlot, P., Lehmann, D., Mickala, P., Abernethy, K., & Anthony,
636 N. (2022). Estimation of the census (Nc) and effective (Ne) population size of a wild mandrill
637 (Mandrillus sphinx) horde in the Lope National Park, Gabon using a non-invasive genetic approach.
638 *Conservation Genetics*.

639 Haas, D. A., Boss, J. M., Strominger, J. L., & Spies, T. (1987). A highly diverged B1 Exon in the DR region of
640 the human MHC: Sequence and evolutionary implications. *Immunogenetics*, *25*, 15–20.

641 Hans, J. B., Haubner, A., Arandjelovic, M., Bergl, R. A., Fünfstück, T., Gray, M., Morgan, D. B., Robbins, M.
642 M., Sanz, C., & Vigilant, L. (2015). Characterization of MHC class II B polymorphism in multiple
643 populations of wild gorillas using non-invasive samples and next-generation sequencing. *American*
644 *Journal of Primatology*, *77*(11), 1193–1206. https://doi.org/10.1002/ajp.22458

645 Harrison, M. J. S. (1988). The mandrill in Gabon's rain forest--ecology, distribution, and status. *Oryx*,
646 *22*(4), 218–228. https://doi.org/10.1017/S0030605300022365

647 Hayward, K. M., Clemente-Carvalho, R. B. G., Jensen, E. L., de Groot, P. V. C., Branigan, M., Dyck, M.,
648 Tschritter, C., Sun, Z., & Lougheed, S. C. (2022). Genotyping-in-thousands by sequencing (GT-seq) of
649 noninvasive faecal and degraded samples: A new panel to enable ongoing monitoring of Canadian
650 polar bear populations. *Molecular Ecology Resources*, *June 2021*, 1–13.
651 https://doi.org/10.1111/1755-0998.13583

652 Hedrick, P. W. (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution*, *56*(10), 1902–
653 1908. https://doi.org/10.1111/j.0014-3820.2002.tb00116.x

654 Higham, J. P., Warren, Y., Adanu, J., Umaru, B. N., Maclarnon, A. M., Sommer, V., & Ross, C. (2009).
655 Living on the edge: Life-history of olive baboons at Gashaka-Gumti National Park, Nigeria.
656 *American Journal of Primatology*, *71*(4), 293–304. https://doi.org/10.1002/ajp.20651

657 Huchard, E., Baniel, A., Schliehe-Diecks, S., & Kappeler, P. M. (2013). MHC-disassortative mate choice
658 and inbreeding avoidance in a solitary primate. *Molecular Ecology*, *22*(15), 4071–4086.
659 https://doi.org/10.1111/mec.12349

660 Huchard, E., Cowlishaw, G., Raymond, M., Weill, M., & Knapp, L. A. (2006). Molecular study of Mhc-DRB

in wild chacma baboons reveals high variability and evidence for trans-species inheritance. *Immunogenetics*, *58*(10), 805–816. https://doi.org/10.1007/s00251-006-0156-8

Huchard, E., Knapp, L. A., Wang, J., Raymond, M., & Cowlishaw, G. (2010). MHC, mate choice and heterozygote advantage in a wild social primate. *Molecular Ecology*, *19*(12), 2545–2561. https://doi.org/10.1111/j.1365-294X.2010.04644.x

Hughes, A. L., & Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(3), 958–962. https://doi.org/10.1073/pnas.86.3.958

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(94). https://doi.org/10.1371/journal.pcbi.1000455

Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A., & Strominger, J. (1984). The class II molecules of the human and murine major histocompatibility complex. *Cell*, *36*(1), 1–13. https://doi.org/10.1016/0092-8674(84)90068-0

Kechin, A., Boyarskikh, U., Kel, A., & Filipenko, M. (2017). cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *Journal of Computational Biology*, *24*(11), 1138–1143. https://doi.org/10.1089/cmb.2017.0096

Kelley, J., Walter, L., & Trowsdale, J. (2005). Comparative genomics of major histocompatibility complexes. *Immunogenetics*, *56*(10), 683–695. https://doi.org/10.1007/s00251-004-0717-7

Kenter, M., Otting, N., Anholts, J., Jonker, M., Schipper, R., & Bontrop, R. E. (1992). Mhc-DRB diversity of the chimpanzee (Pan troglodytes). *Immunogenetics*, *37*(1), 1–11. https://doi.org/10.1007/BF00223539

Klein, J., Sato, A., & Nikolaidis, N. (2007). MHC, TSP, and the Origin of Species: From Immunogenetics to Evolutionary Genetics. *Annual Review of Genetics*, *41*(1), 281–304. https://doi.org/10.1146/annurev.genet.41.110306.130137

Klein, J., Satta, Y., & O'hUigin, C. (1993). The molecular descent of the major histocompatibility complex. *Annual Review of Immunology*, *11*, 269–295.

Knapp, L. A., Cadavid, L. F., Eberle, M. E., Knechtle, S. J., Bontrop, R. E., & Watkins, D. I. (1997). Identification of new Mamu-DRB alleles using DGGE and direct sequencing. *Immunogenetics*, *45*, 171–179.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lahm, S. A. (1986). Diet and Habitat Preference of Mandrillus sphinx in Gabon: Implications of foraging strategy. *American Journal of Primatology*, *26*, 9–26.

Leuchte, N., Berry, N., Köhler, B., Almond, N., LeGrand, R., Thorstensson, R., Titti, F., & Sauermann, U. (2004). MhcDRB-sequences from cynomolgus macaques (Macaca fascicularis) of different origin.

Tissue Antigens, *63*(6), 529–537. https://doi.org/10.1111/j.0001-2815.2004.0222.x

Lighten, J., van Oosterhout, C., Paterson, I. G., Mcmullan, M., & Bentzen, P. (2014). Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (Poecilia reticulata). *Molecular Ecology Resources*, *14*(4), 753–767. https://doi.org/10.1111/1755-0998.12225

Llaurens, V., McMullan, M., & Van Oosterhout, C. (2012). Cryptic MHC polymorphism revealed but not explained by selection on the class IIB peptide-binding region. *Molecular Biology and Evolution*, *29*(6), 1631–1644. https://doi.org/10.1093/molbev/mss012

Loit, K., Adamson, K., Bahram, M., Puusepp, R., Anslan, S., Kiiker, R., Drenkhan, R., & Tedersood, L. (2019). Relative performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) thirdgeneration sequencing instruments in identification of agricultural and forest fungal pathogens. *Applied and Environmental Microbiology*, *85*(21). https://doi.org/10.1128/AEM.01368-19

Maccari, G., Robinson, J., Hammond, J. A., & Marsh, S. G. E. (2020). The IPD Project: a centralised resource for the study of polymorphism in genes of the immune system. *Immunogenetics*, *72*(1–2), 49–55. https://doi.org/10.1007/s00251-019-01133-w

Maley, J. (1996). The African rain forest - Main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. In *Proceedings of the Royal Society of Edinburgh Section B: Biological Sciences* (Vol. 104, Issue 1996). https://doi.org/10.1017/S0269727000006114

Maley, J., Doumenge, C., Giresse, P., Mahé, G., Philippon, N., Hubau, W., Lokonda, M. O., Tshibamba, J. M., & Chepstow-Lusty, A. (2018). Late Holocene forest contraction and fragmentation in central Africa. *Quaternary Research (United States)*, *89*(1), 43–59. https://doi.org/10.1017/qua.2017.97

Manlik, O., Krützen, M., Kopps, A. M., Mann, J., Bejder, L., Allen, S. J., Frère, C., Connor, R. C., & Sherwin, W. B. (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*(12), 6986–6998. https://doi.org/10.1002/ece3.5265

Maruya, E., Saji, H., & Yokoyama, S. (1996). PCR-LIS-SSCP (low ionic strength single-stranded conformation polymorphism) - A simple method for high-resolution allele typing of HLA-DRB1, -DQB1, and -DPB1. *Genome Research*, *6*(1), 51–57. https://doi.org/10.1101/gr.6.1.51

Matsumura, M., Fremont, D. H., Peterson, P. A., & Wilson, I. A. (1992). Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science*, *257*(5072), 927–934. https://doi.org/10.1126/science.1323878

Million, K. M., & Lively, C. M. (2022). Trans-specific polymorphism and the convergent evolution of supertypes in major histocompatibility complex class II genes in darters (Etheostoma). *Ecology and Evolution*, *12*(1), 1–10. https://doi.org/10.1002/ece3.8485

Morin, P. A., Chambers, K. E., Boesch, C., & Vigilant, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (Pan troglodytesverus). *Molecular Ecology*, *10*(7), 1835–1844.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A fast, unconstrained bayesian AppRoximation for inferring selection. *Molecular Biology and Evolution*, *30*(5), 1196–1205. https://doi.org/10.1093/molbev/mst030

741 Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012).
742 Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*(7).
743 https://doi.org/10.1371/journal.pgen.1002764

744 Natesh, M., Taylor, R. W., Truelove, N. K., Hadly, E. A., Palumbi, S. R., Petrov, D. A., & Ramakrishnan, U.
745 (2019). Empowering conservation practice with efficient and economical genotyping from poor
746 quality samples. *Methods in Ecology and Evolution*, *10*(6), 853–859. https://doi.org/10.1111/2041-
747 210X.13173

748 Otting, N., De Groot, N. G., Doxiadis, G. G. M., & Bontrop, R. E. (2002). Extensive Mhc-DQB variation in
749 humans and non-human primate species. *Immunogenetics*, *54*(4), 230–239.
750 https://doi.org/10.1007/s00251-002-0461-9

751 Ou, D., Mitchell, L. A., & Tingle, A. J. (1998). A new categorization of HLA DR alleles on a functional basis.
752 *Human Immunology*, *59*(10), 665–676. https://doi.org/10.1016/S0198-8859(98)00067-6

753 Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B. J., & Loose, M. (2021). Readfish enables targeted
754 nanopore sequencing of gigabase-sized genomes. *Nature Biotechnology*, *39*(4), 442–450.
755 https://doi.org/10.1038/s41587-020-00746-x

756 Radwan, J., Babik, W., Kaufman, J., Lenz, T. L., & Winternitz, J. (2020). Advances in the Evolutionary
757 Understanding of MHC Polymorphism. *Trends in Genetics*, *36*(4), 298–311.
758 https://doi.org/10.1016/j.tig.2020.01.008

759 Radwan, J., Biedrzycka, A., & Babik, W. (2010). Does reduced MHC diversity decrease viability of
760 vertebrate populations? *Biological Conservation*, *143*(3), 537–544.
761 https://doi.org/10.1016/j.biocon.2009.07.026

762 Roberts, T. E., Davenport, T. R. B., Hildebrandt, K. B. P., Jones, T., Stanley, W. T., Sargis, E. J., & Olson, L.
763 E. (2010). The biogeography of introgression in the critically endangered African monkey
764 Rungwecebus kipunji. *Biology Letters*, *6*(2), 233–237. https://doi.org/10.1098/rsbl.2009.0741

765 Robinson, J., Barker, D., Georgiou, X., Cooper, M., Flicek, P., & Marsh, S. (2020). The IPD-IMGT/HLA
766 Database. *Nucleic Acids Research*, *43*, D948–D955.

767 Rodríguez, A., Alcaide, M., Negro, J. J., & Pilard, P. (2011). Using major histocompatibility complex
768 markers to assign the geographic origin of migratory birds: Examples from the threatened lesser
769 kestrel. *Animal Conservation*, *14*(3), 306–313. https://doi.org/10.1111/j.1469-1795.2010.00431.x

770 Rogers, M., Abernethy, K., & Fontaine, B. (1996). Ten days in the life of a mandrill horde in the Lope
771 Reserve Gabon. *American Journal of Primatology*, *40*(4), 297–313.

772 Sandberg, M., Eriksson, L., Jonsson, J., Sjöström, M., & Wold, S. (1998). New chemical descriptors
773 relevant for the design of biologically active peptides. A multivariate characterization of 87 amino
774 acids. *Journal of Medicinal Chemistry*, *41*(14), 2481–2491. https://doi.org/10.1021/jm9700575

775 Sauermann, U., Nurnberg, P., Bercovitch, F. B., Berard, J. D., Trefilov, A., Widdig, A., Kessler, M.,
776 Schmidtke, J., & Krawczak, M. (2001). Increased reproductive success of MHC class II heterozygous
777 males among free-ranging rhesus macaques. *Human Genetics*, *108*(3), 249–254.
778 https://doi.org/10.1007/s004390100485

779 Schmidt, D. A., Campbell, N. R., Govindarajulu, P., Larsen, K. W., & Russello, M. A. (2020). Genotyping-in-
780 Thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA

781 samples to support studies in molecular ecology. *Molecular Ecology Resources*, *20*(1), 114–124.
782 https://doi.org/10.1111/1755-0998.13090

783 Sebastian, A., Herdegen, M., Migalska, M., & Radwan, J. (2016). Amplisas: A web server for multilocus
784 genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources*, *16*(2),
785 498–510. https://doi.org/10.1111/1755-0998.12453

786 Setchell, J. M., Abbott, K. M., Gonzalez, J. P., & Knapp, L. A. (2013). Testing for post-copulatory selection
787 for major histocompatibility complex genotype in a semi-free-ranging primate population.
788 *American Journal of Primatology*, *75*(10), 1021–1031. https://doi.org/10.1002/ajp.22166

789 Setchell, J. M., Charpentier, M., Abbott, K. M., Wickings, E. J., & Knapp, L. A. (2009). Is brightest best?
790 Testing the Hamilton-Zuk hypothesis in Mandrills. *International Journal of Primatology*, *30*(6), 825–
791 844. https://doi.org/10.1007/s10764-009-9371-0

792 Setchell, J. M., Charpentier, M., Abbott, K. M., Wickings, E. J., & Knapp, L. A. (2010). Opposites attract:
793 MHC-associated mate choice in a polygynous primate. *Journal of Evolutionary Biology*, *23*(1), 136–
794 148. https://doi.org/10.1111/j.1420-9101.2009.01880.x

795 Setchell, J. M., Charpentier, M., & Wickings, E. J. (2005). Sexual selection and reproductive careers in
796 mandrills (Mandrillus sphinx). *Behavioral Ecology and Sociobiology*, *58*(5), 474–485.
797 https://doi.org/10.1007/s00265-005-0946-2

798 Setchell, J. M., Vaglio, S., Abbott, K. M., Moggi-Cecchi, J., Boscaro, F., Pieraccini, G., & Knapp, L. A.
799 (2011). Odour signals major histocompatibility complex genotype in an Old World monkey.
800 *Proceedings of the Royal Society B: Biological Sciences*, *278*(1703), 274–280.
801 https://doi.org/10.1098/rspb.2010.0571

802 Slierendregt, B. L., van Noort, J. T., Bakas, R. M., Otting, N., Jonker, M., & Bontrop, R. E. (1992).
803 Evolutionary stability of transspecies major histocompatibility complex class II DRB lineages in
804 humans and rhesus monkeys. *Human Immunology*, *35*(1), 29–39. https://doi.org/10.1016/0198-
805 8859(92)90092-2

806 Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and
807 conservation. *Frontiers in Zoology*, *2*, 1–18. https://doi.org/10.1186/1742-9994-2-16

808 Sommer, S., Courtiol, A., & Mazzoni, C. J. (2013). MHC genotyping of non-model organisms using next-
809 generation sequencing: A new methodology to deal with artefacts and allelic dropout. *BMC*
810 *Genomics*, *14*(1). https://doi.org/10.1186/1471-2164-14-542

811 Song, X., Zhang, P., Huang, K., Chen, D., Guo, S., Qi, X., He, G., Pan, R., & Li, B. (2016). The influence of
812 positive selection and trans-species evolution on DPB diversity in the golden snub-nosed monkeys
813 (Rhinopithecus roxellana). *Primates*, *57*(4), 489–499. https://doi.org/10.1007/s10329-016-0544-0

814 Soto-Calderón, I. D., Ntie, S., Mickala, P., Maisels, F., Wickings, E. J., & Anthony, N. M. (2009). Effects of
815 storage type and time on DNA amplification success in tropical ungulate faeces. *Molecular Ecology*
816 *Resources*, *9*(2), 471–479. https://doi.org/10.1111/j.1755-0998.2008.02462.x

817 Spurgin, L. G., & Richardson, D. S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and
818 misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1684), 979–988.
819 https://doi.org/10.1098/rspb.2009.2084

820 Stutz, W. E., & Bolnick, D. I. (2014). Stepwise threshold clustering: A new method for genotyping MHC

loci using next-generation sequencing technology. *PLoS ONE*, *9*(7), 25–27. https://doi.org/10.1371/journal.pone.0100587

Suárez, C. F., Patarroyo, M. E., Trujillo, E., Estupiñán, M., Baquero, J. E., Parra, C., & Rodriguez, R. (2006). Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages. *Immunogenetics*, *58*(7), 542–558. https://doi.org/10.1007/s00251-006-0127-0

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P., & Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, *24*(16), 3189–3194.

Taberlet, P., Waits, L. P., & Luikart, G. (1999). Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, *14*(8), 323–327.

Telfer, P. T., Souquière, S., Clifford, S. L., Abernethy, K., Bruford, M. W., Disotell, T. R., Sterner, K. N., Roques, P., Marx, P. A., & Wickings, E. J. (2003). Molecular evidence for deep phylogenetic divergence in Mandrillus sphinx. *Molecular Ecology*, *12*(7), 2019–2024. https://doi.org/10.1046/j.1365-294X.2003.01877.x

Unanue, E. R. (1984). Antigen-presenting function of the macrophage. *Annual Review of Immunology*, *2*, 395–428.

van Oosterhout, C. (2009). A new theory of MHC evolution: Beyond selection on the immune genes. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1657), 657–665. https://doi.org/10.1098/rspb.2008.1299

Vásquez-Carrillo, C., Friesen, V., Hall, L., & Peery, M. Z. (2014). Variation in MHC class II B genes in marbled murrelets: Implications for delineating conservation units. *Animal Conservation*, *17*(3), 244–255. https://doi.org/10.1111/acv.12089

Wanner, N., Larsen, P. A., McLain, A., & Faulk, C. (2021). The mitochondrial genome and Epigenome of the Golden lion Tamarin from fecal DNA using Nanopore adaptive sequencing. *BMC Genomics*, *22*(1), 1–11. https://doi.org/10.1186/s12864-021-08046-7

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution*, *35*(3), 773–777. https://doi.org/10.1093/molbev/msx335

White, E. C., Dikangadissi, J. T., Dimoto, E., Karesh, W. B., Kock, M. D., Abiaga, N. O., Starkey, R., Ukizintambara, T., White, L. J. T., & Abernethy, K. (2010). Home-range use by a large horde of wild Mandrillus sphinx. *International Journal of Primatology*, *31*(4), 627–645. https://doi.org/10.1007/s10764-010-9417-3

Wickings, E. J. (1995). Genetic self-management in a captive colony of mandrills (Mandrillus sphinx) as revealed by DNA minisatellite fingerprints. *Electrophoresis*, *16*, 1678–1683.

Wiegand, P., & Kleiber, M. (2001). Less is more - Length reduction of STR amplicons using redesigned primers. *International Journal of Legal Medicine*, *114*(4–5), 285–287. https://doi.org/10.1007/s004140000162

Yasukochi, Y., & Satta, Y. (2014). A human-specific allelic group of the MHC DRB1 gene in primates. *Journal of Physiological Anthropology*, *33*(1), 1–9. https://doi.org/10.1186/1880-6805-33-14

Yu, L., Nie, Y., Yan, L., Hu, Y., & Wei, F. (2018). No evidence for MHC-based mate choice in wild giant pandas. *Ecology and Evolution*, *8*(17), 8642–8651. https://doi.org/10.1002/ece3.4419

Zhang, P., Huang, K., Zhang, B., Dunn, D. W., Chen, D., Li, F., Qi, X., Guo, S., & Li, B. (2018). High polymorphism in MHC-DRB genes in golden snub-nosed monkeys reveals balancing selection in small, isolated populations. *BMC Evolutionary Biology*, *18*(1), 1–13. https://doi.org/10.1186/s12862-018-1148-7

Zhu, Y., Wan, Q. H., Yu, B., Ge, Y. F., & Fang, S. G. (2013). Patterns of genetic differentiation at MHC class i genes and microsatellites identify conservation units in the giant panda. *BMC Evolutionary Biology*, *13*(1). https://doi.org/10.1186/1471-2148-13-227

Zinner, D., Arnold, M. L., & Roos, C. (2009). Is the new primate genus Rungwecebus a baboon? *PLoS ONE*, *4*(3). https://doi.org/10.1371/journal.pone.0004859

Statements and Declarations

Funding

We are grateful for the financial support provided by the Freeport McMoran Endowed Chair awarded by the Audubon Nature Institute to Nicola Anthony (University of New Orleans, US), and for support from the University of New Orleans Office of Research (ORSP) (Award #CON000000002361).

Competing Interests

The authors have no financial or non-financial interests to disclose.

Author Contributions

All authors collaboratively conceived of the study topic and design. AW, AGM, and DL collected samples. AW and AGM performed laboratory work, and AW, JL, CvO, and NA planned and performed data analysis. The manuscript was written by AW, with input from all authors.

Data Accessibility Statement

Replicability data, MHC allele sequences, consensus MHC allele assignments, and Python scripts will be stored at DataDryad.org. MHC nucleotide and amino acid sequences have also been provided in the Supplementary Material.

888 **Tables**

889

Table 1. Summary information from each of the three sequencing runs

	Raw reads	Filtered reads	Mean depth per amplicon	Mean allele count
2018 Miseq	7,429,011	4,215,792	21,957 (SD=19,001)	3.91 (SD=2.02)
2019 Miseq Nano	289,775	120,608	2,566 (SD=1,444)	3.48 (SD=1.70)
2021 Miseq	13,952,325	5,077,399	28,052 (SD=18,032)	3.83 (SD=2.76)

890

891

Table 2. Replicability statistics for all Illumina runs

Runs Compared	R_A Score Dataset	Sample Type	Mean R_A	SD R_A	%R_A=1	%R_A=0
2018 Miseq & 2019 Miseq Nano	A	Non-invasive (n=22)	0.44	0.30	13.6	9.1
2019 Miseq Nano	B	Blood/Hair (n=9)	0.76	0.29	55.6	0
& 2021 Miseq	C	Non-invasive (n=22)	0.31	0.33	13.6	27.3
2018 Miseq &	D	Non-invasive (n=181)	0.32	0.30	9.9	21.5
2021 Miseq	D	1000 random draws of n=9	0.05-0.69	0.07-0.48	0-44.4	0-77.8

R_A= proportion of variants that replicated between runs. Letters A, B, C, and D correspond to sequencing run comparisons shown in Figure 1. Significant R_A score comparisons: B&C (p=0.003), D&B (R_A score from B > the 95th percentile of score distribution from D).

892

893

894

Figures

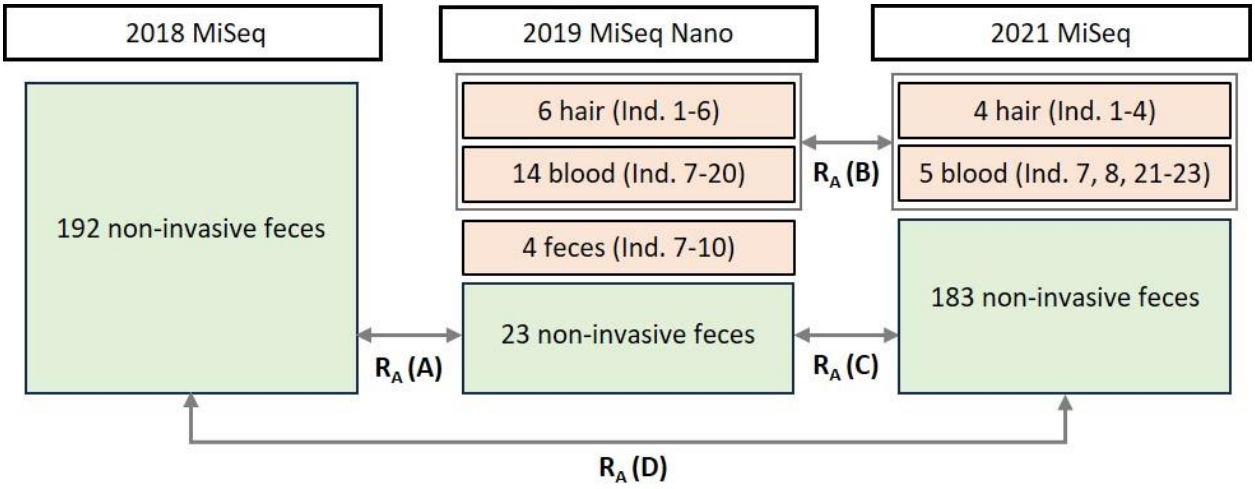


Fig. 1. Between-run replicability scores calculated between different sample types in three independent Illumina runs. Gray arrows show the pairs of Illumina runs used to calculate R_A scores, and different datasets of R_A scores are denoted by letters A (n=23 replicated fecal samples), B (n=9 replicated tissue samples), C (n=23 replicated fecal samples), and D (n=183 replicated fecal samples). Pink boxes represent samples from radio-collared mandrills, and green boxes represent non-invasively collected fecal samples.

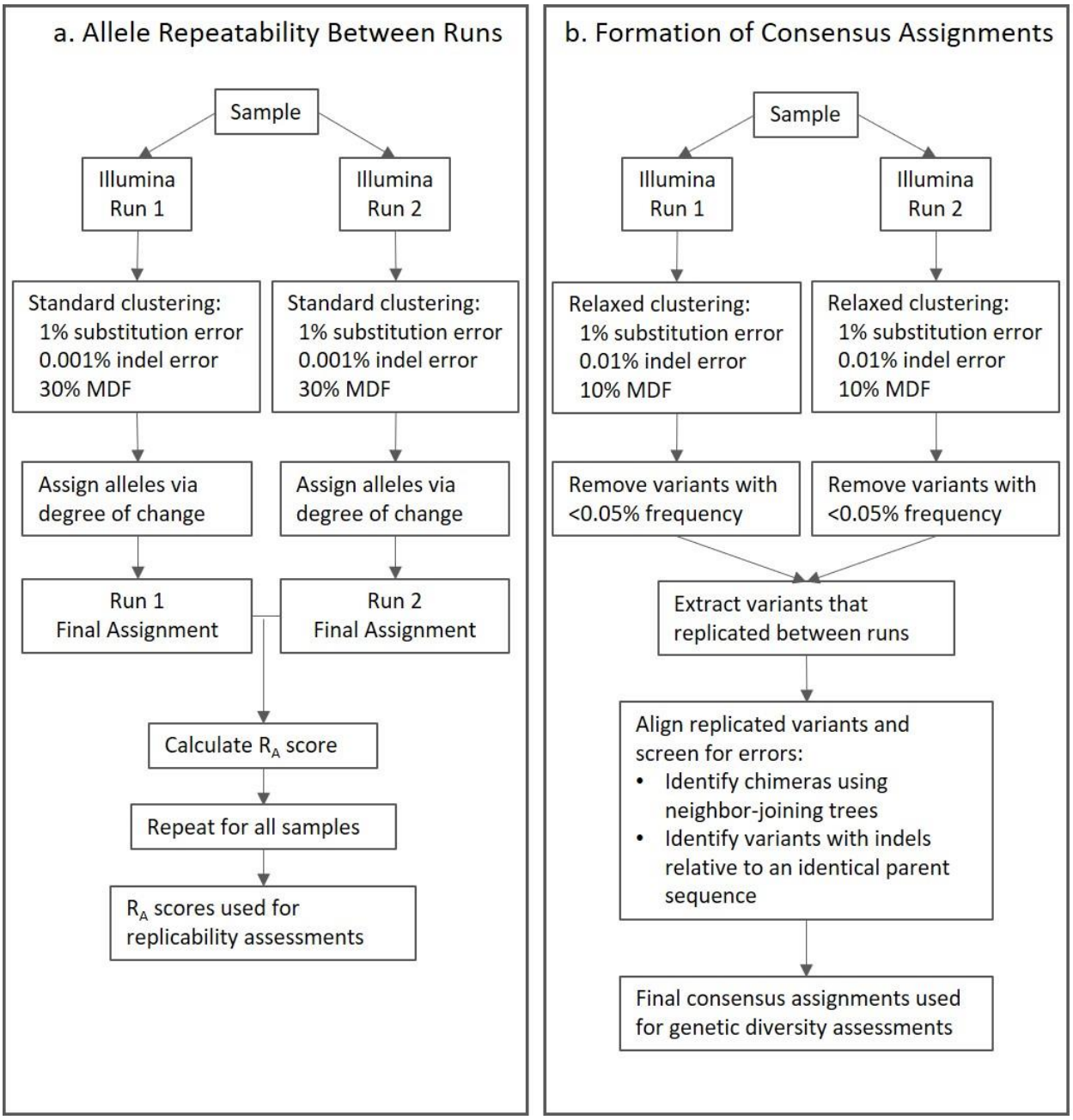
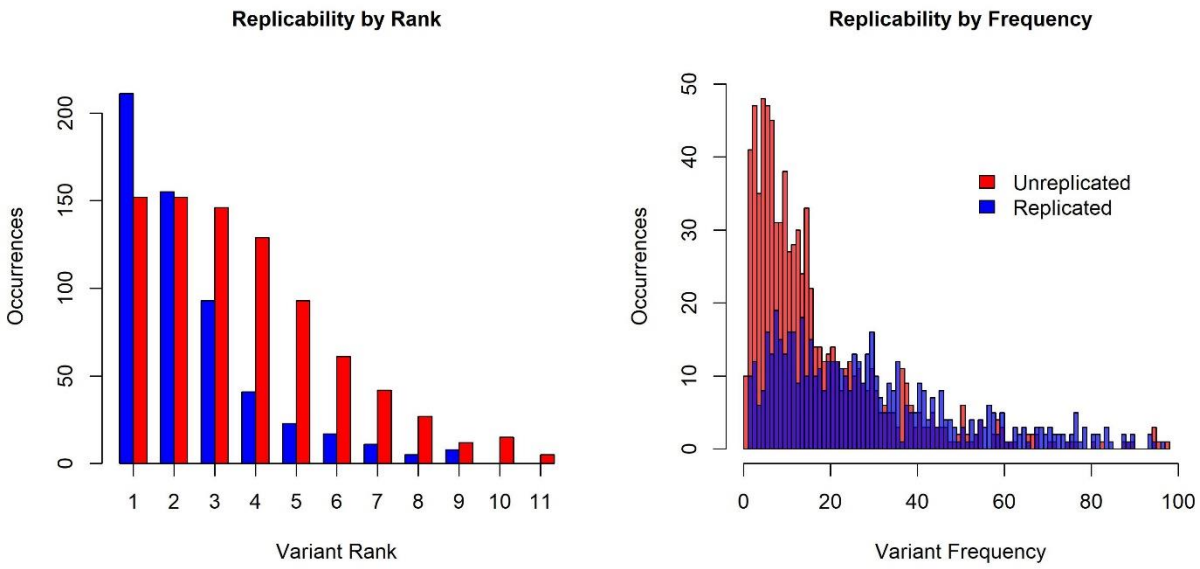


Fig. 2. The above figure shows the processes undertaken for each sample to determine its repeatability score between runs (*a*) and its consensus allele assignment (*b*). Pairs of sequencing runs used for each sample are shown in Figure 1. MDF = minimum dominant frequency threshold



911

Fig. 3. Graphs showing relative frequencies (right) and per amplicon ranks (left) of variants that replicated (blue) and those that did not (red), between the 2018 and 2021 Illumina Miseq runs.

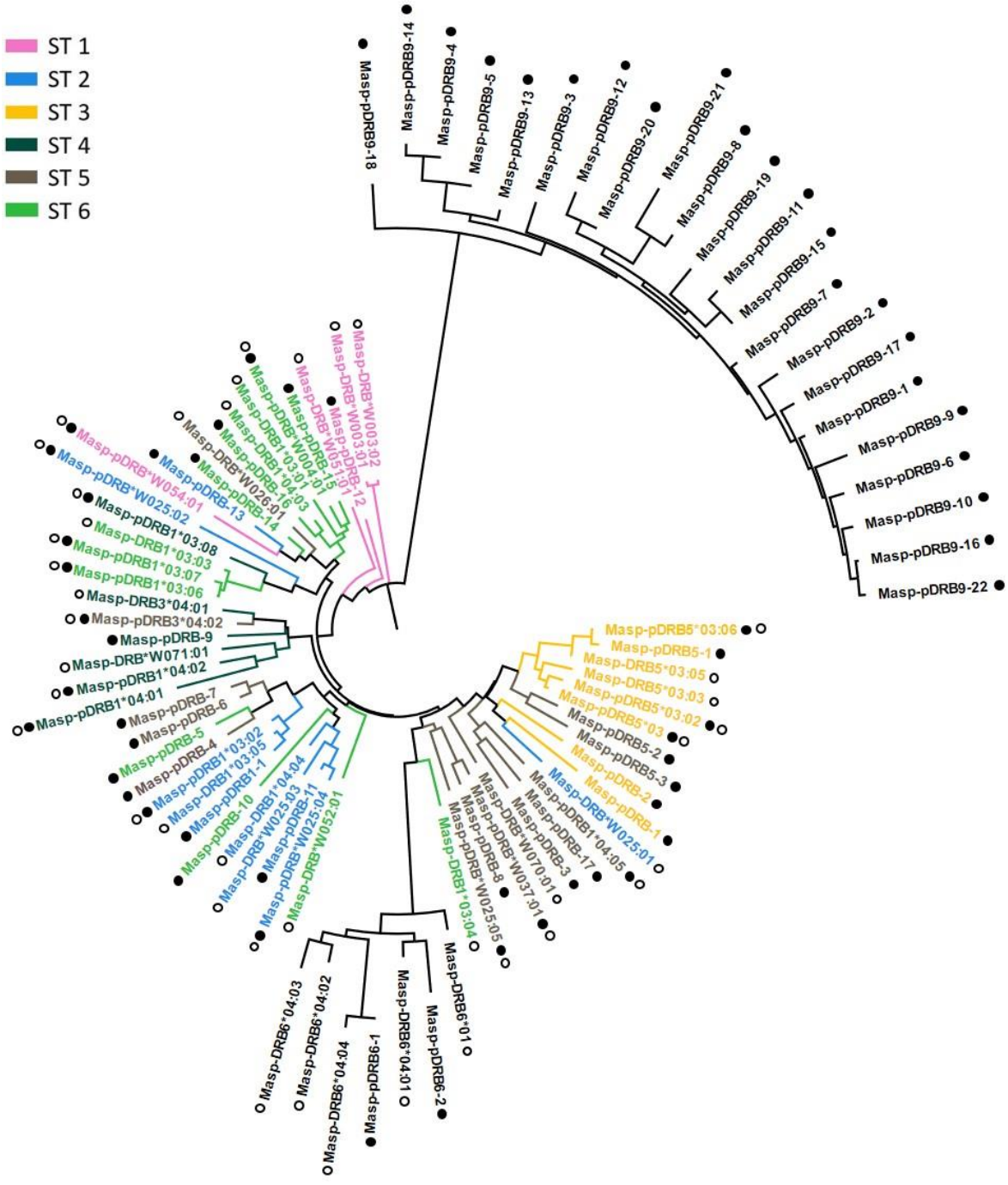


Fig. 4. Neighbor-joining tree of all known mandrill MHC-DRB alleles. Alleles are colored by supertype (ST), and nonfunctional loci (DRB6 and putative DRB9) are represented in black. Alleles found in the captive and wild populations are designated with open and filled circles, respectively.

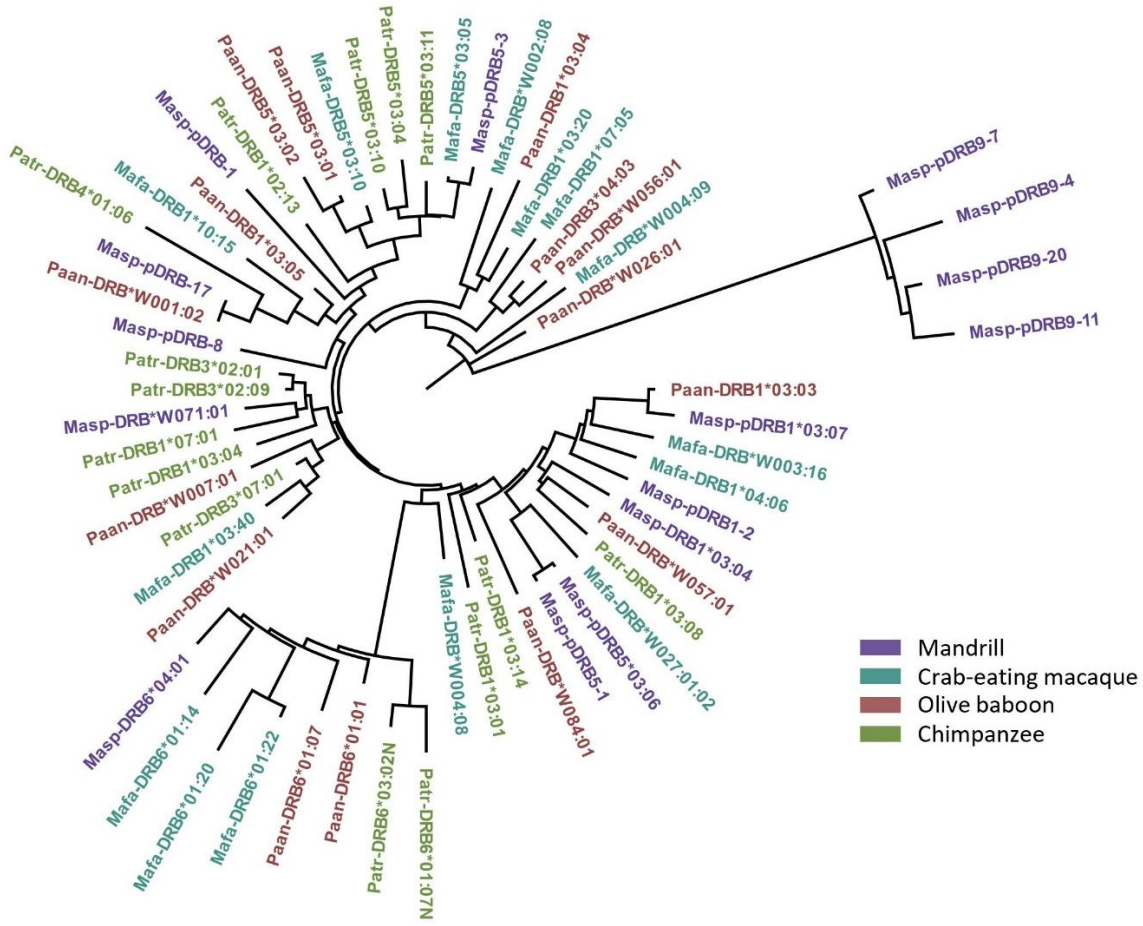
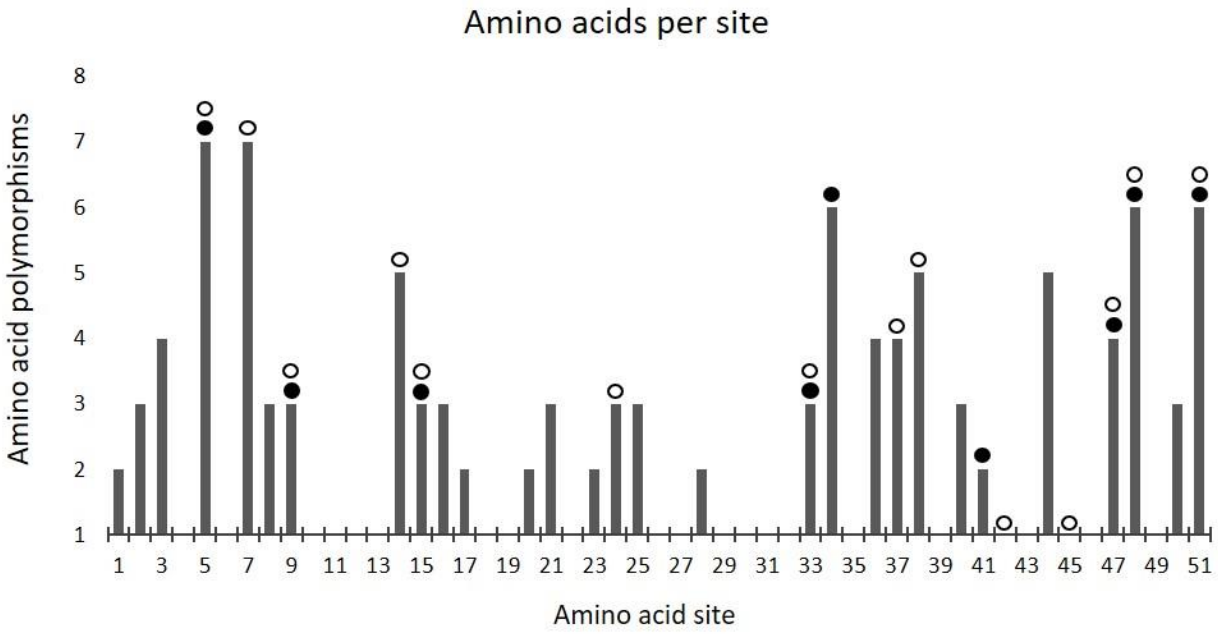


Fig. 5 An unrooted neighbor-joining tree showing DRB lineages from mandrills and other primates. For simplicity, only twenty sequences per species are shown.



923

Fig. 6. A plot of amino acid polymorphisms at each codon. Black circles indicate sites from the current study with significant evidence for positive selection, and open circles indicate sites identified in Brown et al. (1993) as part of the HLA peptide binding region.