American Journal of Primatology

RESEARCH ARTICLE

Characterization of MHC Class II B Polymorphism in Multiple Populations of Wild Gorillas Using Non-Invasive Samples and Next-Generation Sequencing

JÖRG B. HANS[1]*, ANNE HAUBNER[1], MIMI ARANDJELOVIC[1], RICHARD A. BERGL[2], TILLMANN FÜNFSTÜCK[1], MARYKE GRAY[3], DAVID B. MORGAN[4], MARTHA M. ROBBINS[1], CRICKETTE SANZ[5], AND LINDA VIGILANT[1]
[1]*Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*
[2]*North Carolina Zoological Park, Asheboro, North Carolina, USA*
[3]*International Gorilla Conservation Program, Kigali, Rwanda*
[4]*Lincoln Park Zoo, Chicago, Illinois, USA*
[5]*Washington University, Saint Louis, Missouri, USA*

Genes encoded by the major histocompatibility complex (MHC) are crucial for the recognition and presentation of antigens to the immune system. In contrast to their closest relatives, chimpanzees and humans, much less is known about variation in gorillas at these loci. This study explored the exon 2 variation of -DPB1, -DQB1, and -DRB genes in 46 gorillas from four populations while simultaneously evaluating the feasibility of using fecal samples for high-throughput MHC genotyping. By applying strict similarity- and frequency-based analysis, we found, despite our modest sample size, a total of 18 alleles that have not been described previously, thereby illustrating the potential for efficient and highly accurate MHC genotyping from non-invasive DNA samples. We emphasize the importance of controlling for multiple potential sources of error when applying this massively parallel short-read sequencing technology to PCR products generated from low concentration DNA extracts. We observed pronounced differences in MHC variation between species, subspecies and populations that are consistent with both the ancient and recent demographic histories experienced by gorillas. Am. J. Primatol. © 2015 Wiley Periodicals, Inc.

Key words: MHC genotyping; high-throughput sequencing; Illumina; fecal samples; non-model organisms

INTRODUCTION

The genes of the major histocompatibility complex (MHC) play a critical role in the immune response by coding for cell-surface glycoproteins which present antigens to immunocompetent cells. Within the MHC, two major gene classes can be distinguished: MHC class I molecules bind and present intracellular peptides to CD8+ T cells whereas the function of the heterodimeric MHC class II molecules is the recognition and presentation of peptides of extracellular origin to CD4+ T cells [Parham & Ohta, 1996; Watts, 1997]. MHC genes are exceptional for their polygenicity and high polymorphism, that is, there are several different MHC genes with multiple alleles constituting the repertoire of antigen-binding molecules with variation at both the intra- and interindividual level.

The gorilla MHC class II region is similar to those from the chimpanzee and human and includes, among others, three gene families designated Gogo-DP, -DQ, and -DR. Typically, polymorphism of MHC class II genes is mostly confined to exon 2 of both the A and B genes encoding the alpha-1 and beta-1 domains of the heterodimeric class II molecule, respectively. Within these domains, the most variable amino acid positions are the peptide and T cell contact residues. It appears that both the Gogo-DPA1 and -DPB1 genes show relatively low levels of polymorphism and each possess only one allelic lineage [Gyllensten et al., 1996; Otting & Bontrop, 1995]. The Gogo-DQA1 and -DQB1, however, display a high degree of exon 2 variation, resulting in numerous allelic lineages and alleles [Bontrop, 1994, and references therein; Gyllensten et al., 1990]. Whereas the apparently invariant DRA gene

Contract grant sponsor: Max Planck Society

Conflicts of interest: The authors declare no conflicts of interest.

*Correspondence to: Jörg B. Hans, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. E-mail: joerg_hans@eva.mpg.de

Received 23 March 2015; revised 8 July 2015; revision accepted 3 August 2015

DOI: 10.1002/ajp.22458
Published online XX Month Year in Wiley Online Library (wileyonlinelibrary.com).

© 2015 Wiley Periodicals, Inc.

is highly conserved among primates [Doxiadis et al., 2008; Fan et al., 1989], the primate DRB region has been subject to recurrent duplication events resulting in multiple gene copies [Bontrop, 1997]. As observed for chimpanzees, in gorillas the -DRB1, -DRB3, and -DRB5 genes are the most polymorphic [O'hUigin et al., 1993]. In gorillas, five different DRB region configurations, also referred to as DRB linkage groups, have been described with two to three functional DRB loci each [Bontrop, 1997; Kasahara et al., 1992; Kenter et al., 1993].

However, our knowledge of functional MHC variation in gorillas is limited because most studies have been confined to western gorillas, the predominant species in captivity [Lukas et al., 2004; Nsubuga et al., 2009]. Two species of gorillas are currently recognized: the western gorilla (*Gorilla gorilla*) and the eastern gorilla (*Gorilla beringei*), each of which have been further classified into two subspecies [Groves, 2001]. The initial split between the ancestral western and eastern gorilla populations has been dated to approximately 0.9–1.75 million years ago with subsequent genetic exchange persisting until as recently as 80,000–200,000 years ago [McManus et al., 2015; Scally et al., 2012; Thalmann et al., 2007]. Despite significant differences in genetic diversity between the two gorilla species, probably as a result of variable demographic histories, much less is known about their variation at functionally important loci [Fünfstück & Vigilant, 2015; Prado-Martinez et al., 2013; Scally et al., 2013]. Assessment of MHC variation can provide important insights into the susceptibility and resistance to infectious diseases which are of particular relevance for endangered species living in small and isolated populations [reviewed in Sommer, 2005]. However, genetic studies of wild great apes are particularly challenging because it is usually necessary to rely upon non-invasive samples such as feces, which are typically of low DNA quantity and quality [Taberlet et al., 1999]. Application of next-generation sequencing (NGS) has the potential to increase the efficiency and reliability of MHC genotyping studies even when dealing with difficult DNA templates [Babik, 2010; Perry et al., 2010]. However, multilocus systems such as the MHC are particularly prone to PCR-generated artefacts which can, unless recognized as such, lead to an overestimation of MHC diversity [Babik, 2010; Lenz & Becker, 2008; Sommer et al., 2013].

In the present pilot study, NGS was utilized to characterize the polymorphic exon 2 of the -DPB1, -DQB1, and -DRB genes from fecal samples of four gorilla populations representing both western gorilla subspecies and the two eastern mountain gorilla populations. On the basis of a rigorous data analysis workflow, we were able to compare MHC class II sequence variation of different gorilla subspecies while considering the different demographic histories experienced by gorillas. Finally, we highlight major issues associated with the use of non-invasive samples for MHC genotyping studies.

METHODS

Samples

We selected 46 DNA extracts from wild gorilla fecal samples analyzed in previous genotyping studies and thus known to represent different individuals. These include 22 western gorillas (*Gorilla gorilla*) consisting of 13 western lowland gorillas (*G. g. gorilla*) collected close or in the Goualougo Triangle in the Nouabale-Ndoki National Park, Republic of Congo [Fünfstück et al., 2014] and 9 Cross River gorillas (*G. g. diehli*) collected in the Kagwene Gorilla Sanctuary and the Mone River Forest Reserve, Cameroon [Arandjelovic et al., 2015]. Eastern gorillas (*Gorilla beringei*) were represented by 11 mountain gorillas (G.b. beringei) collected in the Virunga Massif [Gray et al., 2013] and 13 mountain gorillas from the Bwindi Impenetrable National Park, Uganda [Roy et al., 2014a]. Details of sample collection, DNA extraction and microsatellite genotyping to establish identity are given in the above mentioned references, respectively. In brief, DNA extracts were genotyped at 12 to 16 highly variable microsatellites with high ability to discriminate between closely related individuals [Gray et al., 2013]. Across all samples within each population, CERVUS version 3.0 [Kalinowski et al., 2007] was used to compare genotypes using the "identity check" function and evaluate the probability (P_{ID}) that a pair of samples with completely or partially matching genotypes were derived from the same individual (or a pair of siblings), as well as to evaluate whether individuals could be close relatives.

Amplification of MHC Loci and NGS Library Construction

PCR reactions were performed in a final volume of 20 μL consisting of 1× PCR Buffer II, 2 mM MgCl2, 16 μg bovine serum albumin, 200 μM of each dNTP, 200 μM of each primer, 0.5 units of AmpliTaq Gold DNA Polymerase (Life Technologies, Carlsbad, CA) and 2 μL DNA extract with concentrations ranging from approximately 20–300 pg/μL (Arandjelovic et al., 2015; Fünfstück et al., 2014). The complete DPB1 exon 2 (expected length including primers 338 bp) and partial DQB1 exon 2 (230 bp) were amplified using the primer sets UG19/UG21 [Gyllensten et al., 1996] and GH28/GH29 [Gyllensten et al., 1990], respectively. However, it has to be noted that the latter primer pair also co-amplifies the second exon of the highly homologous DQB2 pseudogene which is characterized by a 3-bp difference compared to DQB1 [Gyllensten et al., 1990; Otting et al., 1992]. For

the amplification of partial DRB exon 2 loci (296 bp), the forward primer GH46 was used as it has been shown to be complementary across a wide range of primates [Gyllensten et al., 1991]. As reverse primer, we used AB60 [Bentley et al., 2009] which amplifies all functional DRB loci in humans. However, multilocus genotyping is particularly prone to primer-template mismatches which can cause PCR bias in terms of unequal amplification efficiency of alleles [Wegner, 2009; Sommer et al., 2013]. Indeed, Primer-BLAST analysis revealed that for some gorilla DPB1 and DRB6 alleles there are single mismatches of primer and target sequences, albeit not at or close to the 3′ ends. However, as a precaution, a touchdown PCR protocol was applied in order to overcome potential primer-template mismatches [Don et al., 1991; Hecker & Roux, 1996; Roux, 1994]. Cycling conditions for the amplification of DPB1 followed a touchdown protocol with initial denaturation at 94°C for 9min followed by 30 sec denaturation at 94°C, 30 sec annealing at 70°C and 30 sec elongation at 72°C, two cycles of 30 sec at 94°C, 30 sec at 65°C and 30 sec at 72°C, 40 cycles of 30 sec at 94°C, 30 sec at 60°C, and 30 sec at 72°C, finishing with 1 hr final elongation at 72°C. For the amplification of DQB1, the initial denaturation was followed by 40 cycles of 30 sec at 94°C, 30 sec at 58°C, 30 sec at 72°C and a final elongation step of 30 min at 72°C. Cycling conditions for DRB were similar to DPB1 but annealing temperature started at 70°C for one cycle followed by two cycles at 65°C, two cycles at 60°C and 40 cycles at 55°C. For each sample, two independent PCR replicates were carried out and after confirmation of successful amplification by gel electrophoresis, amplicons of each locus were individually SPRI purified as previously described [Rohland & Reich, 2012]. The quantity and quality of each purified amplicon was checked using an Agilent DNA 1000 Kit and the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). The first replicates of each of the three loci were pooled by individual in equimolar ratios and prepared for Illumina high-throughput sequencing following the protocol of Meyer & Kircher [2010] with modifications described in Kircher et al. [2012]. A second library with different index combinations was prepared from the second replicate of the three amplicons per individual following the same procedure. After quantification with a DNA 1000 chip, the two libraries were combined in equimolar ratios and spiked into a human whole-genome shotgun library in order to increase the complexity because sequence homogeneity is expected to negatively influence both phasing and read quality [MiSeq Reporter User Guide, September 2014]. In the final sequencing library our MHC amplicons represented around 8% of the final pool. Paired-end sequencing was done using the Illumina MiSeq system and 2×250 bp reads.

Data Analysis

Data analysis consisted of (i) sequence preprocessing; (ii) discrimination of artefacts from true alleles through sequence clustering in combination with coverage depth information; and (iii) genotype estimation and validation through comparison with existing MHC information in gorillas. These steps are described in detail below.

Sequence Preprocessing

After base calling performed with Illumina's Bustard, paired-end reads were merged and demultiplexed by their unique index combination followed by trimming of adaptors using leeHom [Renaud et al., 2014]. Using prinseq-lite 0.20.4 [Schmieder & Edwards, 2011], we discarded merged reads showing an average Phred score below 30 or length differences greater or less than 5% to the expected product in order to account for indels. We used the software jMHC version 1.0.471 [Stuglik et al., 2011] for each locus separately to remove the priming sites, estimate the coverage per sequence and conduct hierarchical clustering of variants based on their frequency.

Discrimination of Artefacts From True Alleles

In order to separate true alleles from artefacts arising out of nucleotide misincorporations during PCR, we first retained only variant sequences that were observed at least twice [Sommer et al., 2013]. We then combined the remaining sequences from both replicates, constructed a multiple alignment using ClustalW and built a neighbor-joining tree based on the number of differences as implemented in the software package MEGA version 6.06 [Tamura et al., 2013]. Only sequences confirmed in both replicates were considered in the next step of the analysis. However, due to the possibility of allelic dropout, that is, the stochastic or preferential amplification of one allele despite the state of the locus being heterozygous, we also included sequences present in only one of the replicates if these were found to separately cluster in the tree. After re-evaluating the tree, variants differing in 1–2 bp (including length differences) to the most frequently observed sequence within each cluster were considered to have arisen through misincorporations and were excluded from further analysis as per previous recommendations [Lighten et al., 2014; Sommer et al., 2013; Stutz & Bolnick, 2014]. These steps are based on the assumption that true alleles are present at higher sequencing depths than artefacts. Importantly, artefact sequences are also expected to arise if

incompletely elongated template molecules act as a primer in subsequent PCR cycles [Kanagawa, 2003; Lenz & Becker, 2008; Meyerhans et al., 1990] or by so-called "jumping PCR", that is, the occurrence of template-switching during PCR amplification [Kanagawa, 2003; Odelberg et al., 1995; Pääbo et al., 1990]. The resulting chimeric sequence should therefore co-occur with its respective parent sequences, that is, the true alleles from which the chimera originated [Huber et al., 2004; Zagalska-Neubauer et al., 2010]. Moreover, since chimeras contain sequence motifs from potentially phylogenetically distinct alleles we expect that a dendrogram should reflect the hybrid origin of these sequences. We estimated the genetic distances between sequences using the Jukes and Cantor method [1969] and built a neighbor-joining tree using MEGA. The most frequent variant within each cluster was assumed to represent a putative allele. Lower-frequency variants were checked using visual inspection and if it was found that they could be a combination of the high-frequency sequences then the query was classified as chimera and discarded [Galan et al., 2010; Sommer et al., 2013; Stutz & Bolnick, 2014]. Finally, the most frequent sequences were compared to each other and where potential recombination of other high-frequency variants could explain the sequence we also considered them chimeras and removed them from the preliminary dataset.

Genotype Estimation and Validation

Our newly identified sequences were given provisional allelic designations by means of phylogenetic analyses utilizing published gorilla alleles either downloaded from the IPD-MHC database [de Groot et al., 2012; Robinson et al., 2013] or, if necessary, extracted from relevant publications. We reconstructed phylogenies using the neighbor-joining method with the Kimura 2-parameter distance model [Kimura, 1980] and 1000 bootstrap replicates in MEGA. Alleles were named according to the nomenclature of Klein et al. [1990].

Allelic dropout frequency was assessed by comparing the number of identified alleles between the two PCR replicates. For each locus separately, the allelic dropout rate was calculated as the number of times only one of the two alleles was present at an apparently heterozygous locus, divided by the total number of PCRs regardless of whether one or two alleles were identified.

For each individual, we expect a maximum of two alleles at each of the DPB1 and DQB1 loci. However, three individuals showed more than two DPB1 alleles in one of the replicates, probably due to cross-contamination. In two other individuals, we observed alleles in one of the replicates identical to human, indicating contamination. These five individuals were removed from subsequent DPB1 analysis. For DQB1, one individual was excluded from further analyses due to the observation of more than two alleles, probably as a result of cross-contamination. In contrast, the number of functional DRB genes present per chromosome can vary from two to three in gorillas. After designation of sequences, we were able to unambiguously allocate alleles to different DRB genes. Consequently, we did not accept more than two alleles belonging to the same gene per individual. Five individuals were removed from further DRB analyses because in one of the replicates more than two DRB alleles of the same gene were detected, probably due to cross-contamination. For one additional individual, we identified a human allele in one of the replicates, indicating contamination. This individual was removed from subsequent DRB analyses. Due to the occurrence of homozygous individuals within each population, we were able to infer DRB linkage groups, that is, the combinations of DRB loci occurring on a single chromosome.

Because our study exclusively relied on non-invasively collected samples and therefore did not involve animal handling or testing, we did not violate any regulations of the Deutsches Tierschutzgesetz and the US Public Health Service Policy on Humane Care and Use of Laboratory Animal. Our study complied entirely with ASP principles for the ethical treatment of non-human primates and all research was conducted in accordance with the German law.

RESULTS

A total of 1,107,555 contigs were accepted based on their high average accuracy ($\geq$Q30), expected length and correctly called priming sites. However, we observed a bias against longer amplicons, especially for DPB1, probably as a result of filtering out merged reads of low average quality. Indeed, a common phenomenon in sequencing-by-synthesis is the accumulation of low-quality calls towards the end of reads. Nevertheless, high coverage depths were obtained for all amplicons of DPB1, DQB1, and DRB, with totals of 202,823 (mean per replicate and individual: 2204 ± 479), 457,611 (4974 ± 817), and 447,121 (4860 ± 876), respectively. A total of 248,731 reads (22.46%) were observed once and were therefore discarded. More specifically, single sequences totaled up to 58,709 (mean per replicate and individual: 638 ± 139), 53,987 (587 ± 85), and 136,035 (1479 ± 492) for DPB1, DQB1, and DRB, respectively. Chimeric sequences were detected in 6 individuals (8.69%), in 44 individuals (95.65%), and in 46 individuals (100%) for DPB1, DQB1, and DRB, respectively. A total of 609,279 (55.01%) reads were identified as real alleles. For the DPB1, DQB1, and DRB loci, we obtained a total of 108,539, 330,320 and 170,420 reads, respectively. Total read number of both replicates for DPB1 amplicons ranged from 1682 to 3455 per individual, for DQB1 from 3706 to

9613 and for DRB from 1575 to 5047. Information at a single locus was discarded for eleven of the 46 individuals due to suspected cross-contamination in one of the two replicates, leading to an overall estimate of 3.98%. For DPB1, DQB1, and DRB, we observed an allelic dropout rate of 9.8%, 3.3%, and 13.0%, respectively. In addition, for two individuals we identified simultaneous allelic dropout at three DRB loci. We were conservative and excluded these individuals from further DRB analyses.

Gorilla MHC Genotypes and Linkage Groups

Among the 41 individuals, we observed a total of eight different DPB1 exon 2 DNA sequences, two of which had been previously described [Gyllensten et al., 1996]. These eight alleles translate into seven unique amino acid sequences. We found six different DPB1 alleles in western lowland gorillas and only two each in the Cross River gorillas and in the two mountain gorilla populations (Table I). Interestingly, the two alleles detected in Cross River gorillas represented a subset of the DPB1 variation found in western lowland gorillas. In addition, identical alleles were found across the two mountain gorilla populations (Table II).

We reconstructed a phylogeny using the six novel sequences and five published gorilla DPB1 exon 2 sequences found in six western lowland gorillas [Gyllensten et al., 1996] (Fig. 1). These sequences formed two main clusters in the neighbor-joining tree. Included in the first cluster are all previously described alleles and one novel allele all of which are only found in the western gorilla species, that is, western lowland and Cross River gorillas. The second cluster contains three novel alleles found in western lowland gorillas and the two novel alleles from mountain gorillas, thus representing a newly described allelic lineage.

Although, as expected, co-amplification of the second exon of the DQB2 pseudogene occurred, reliable DQB1 genotypes were obtained for 45 of the 46 individuals analyzed. Discrimination between DQB1 and DQB2 sequences was possible due to a 3-bp length difference consistent with earlier findings [Gyllensten et al., 1990; Otting et al., 1992]. All individuals were homozygous for the DQB2*01:03 allele [Gaur et al., 1992]. A total of seven different alleles were identified of which one allele has been previously described [Otting et al., 1992]. The remaining six alleles were novel at the DNA level. At the amino acid level, however, four alleles were identical to published sequences. All seven alleles were found in western lowland gorillas. In both Cross River and mountain gorillas, only two alleles (DQB1*N1 and DQB1*N5) were identified, respectively, representing only a subset of the DQB1 variation found in western lowland gorillas (Table I and II). In the tree analysis (Fig. 2), the 16 previously described alleles found among nine western lowland gorillas [Gaur et al., 1992; Gyllensten et al., 1990; Otting et al., 1992] and the seven novel sequences grouped into four clusters (DQB1*02, DQB1*05, DQB1*19, and DQB1*06). Despite the limited variation found in Cross River and mountain gorillas, no subspecies-specific clustering was observed.

As anticipated due to the duplicated nature of this locus, assessment of DRB results was more complex as compared to DPB1 and DQB1. Specific DRB6 alleles were characterized by relatively low but consistent sequencing depths which suggest that despite potential primer-template mismatches our PCR protocol amplifies all DRB variants in gorillas. However, one DRB6*02 allele could not unambiguously designated due to its similarity to two previously described alleles and our lack of sequence data outside of the targeted region. Nevertheless, this had no impact on functional analyses because the DRB6 locus is a pseudogene in gorillas [Corell et al., 1992; Figueroa et al., 1991; Kasahara et al., 1992]. We found a total of 24 DRB exon 2 sequences, six of which had not been previously reported (Table I). These 24 DRB alleles translate into 23 unique amino acid sequences with the ambiguous DRB6*02 allele being the only exception. We observed pronounced differences between the different gorilla subspecies regarding the total number of functional DRB alleles (Table II), with western lowland gorillas being the most diverse (11 different alleles) followed by the two populations of mountain gorillas (nine and six different alleles for Bwindi and the Virungas, respectively) and the Cross River gorillas (five different alleles). In addition, in mountain gorillas we observed only DRB1 alleles belonging to the DRB1*03 lineage whereas alleles of the DRB1*03 and DRB1*10 lineages were found in the two western gorilla subspecies.

In the phylogenetic tree (Fig. 3), our new sequences as well as 34 previously published alleles from a total of 52 western lowland gorillas and 19 mountain gorillas fell within six clusters (DRB1*03, DRB1*10, DRB3, DRB5, DRB*W, and DRB6) consistent with earlier findings [Lukas et al., 2004]. One previously published allele, Gogo-DRB5*02:01 [Kenter et al., 1993], was excluded due to its unusual placement in the tree, a finding consistent with the suggestion that this allele is actually a chimeric sequence [Lukas et al., 2004].

Inference of complete DRB linkage groups was possible for 34 out of 38 individuals (Fig. 4). Four individuals were excluded due to the absence of an allele, probably due to allelic dropout in both replicates. Although previously described in mountain gorillas [Lukas et al., 2004], we detected the linkage group I (DRB1/6/W) only in the western gorilla species whereas the linkage group II (DRB1/3/6/5) was found in all gorilla subspecies. The linkage group IV (DRB1/3) was only detected in mountain

TABLE I. Gorilla MHC Class II Alleles Identified in the Present Study and Their Respective Genebank Accession Numbers: WLG, Western Lowland Gorilla; CRG, Cross River Gorilla; MG-B, Mountain Gorillas from Bwindi; MG-V, Mountain Gorillas From Virunga; MG-BV, Mountain Gorillas From Bwindi and Virunga

Allele	Subspecies/population	Number of individuals	Accession number
*DPB1*01*	WLG	6	KP872242
	CRG	2	KP872248
*DPB1*04*	WLG	2	KP872243
	CRG	9	KP872249
DPB1*N1	WLG	1	**KP872244**
DPB1*N2	WLG	1	**KP872245**
DPB1*N3	WLG	2	**KP872246**
DPB1*N4	MG-BV	20	**KP872250**
DPB1*N5	MG-BV	8	**KP872251**
DPB1*N6	WLG	2	**KP872247**
DQB1*N1	WLG	4	**KP872252**
	CRG	9	**KP872259**
	MG-BV	11	**KP872261**
*DQB1*05:02*	WLG	2	KP872253
DQB1*N2	WLG	1	**KP872254**
DQB1*N3	WLG	1	**KP872255**
DQB1*N4	WLG	7	**KP872256**
DQB1*N5	WLG	2	**KP872257**
	CRG	3	**KP872260**
	MG-BV	19	**KP872262**
DQB1*N6	WLG	6	**KP872258**
*DRB1*03:07:01*	WLG	5	KP872263
	CRG	1	KP872267
*DRB1*03:08*	WLG	2	KP872264
*DRB1*03:09*	WLG	2	KP872265
	MG-B	2	KP872269
*DRB1*03:10*	MG-B	6	KP872270
*DRB1*03:11*	MG-B	7	KP872271
DRB1*N1	MG-V	8	**KP872272**
*DRB1*10:02*	WLG	1	KP872266
	CRG	9	KP872268
*DRB3*01:01*	WLG	10	KP872273
	CRG	1	KP872275
	MG-B	6	KP872276
*DRB3*01:07*	MG-BV	7	KP872277
*DRB3*01:08*	MG-BV	9	KP872278
*DRB3*04:01*	WLG	4	KP872274
	MG-BV	7	KP872279
*DRB5*01:01*	WLG	4	KP872280
DRB5*N1	WLG	5	**KP872281**
DRB5*N2	MG-V	2	**KP872287**
*DRB5*05:01*	WLG	1	KP872282
	CRG	1	KP872284
DRB5*N3	MG-V	8	**KP872288**
DRB5*N4	WLG	3	**KP872283**
*DRB5*05:06*	MG-B	6	KP872286
*DRB5*05:04:01*	MG-B	7	KP872285
*DRB*W8:02*	WLG	1	KP872289
	CRG	9	KP872290
*DRB6*01:02*	WLG	3	KP872292
	CRG	9	KP872294
*DRB6*02:04*	MG-BV	13	KP872296
DRB6*N1	MG-B	2	**KP872297**
DRB6?*	WLG	7	KP872291
	CRG	1	KP872293
	MG-B	7	KP872295

Novel alleles are highlighted in bold.

TABLE II. Variation at the Exon 2 of the Functional MHC Class II B Loci in Gorillas From the Present Study

	Number of functional MHC class II B alleles						
	DPB1	*DQB1*	*DRB1*	*DRB3*	*DRB5*	*DRB*W*	Total
Western gorillas							
Lowland	6	7	4	2	4	1	24
Cross River	2	2	2	1	1	1	9
Shared between western gorillas	2	2	2	1	1	1	9
Mountain gorillas							
Bwindi	2	2	3	4	2	–	13
Virunga	2	2	1	3	2	–	10
Shared between eastern gorillas	2	2	0	3	0	–	7
Shared between gorillas species	0	2	1	2	0	0	5

gorillas from the Virungas. Furthermore, we were able to confirm the gorilla linkage group DR8 (DRB3/5/6) described in Kasahara et al. [1992] in both western lowland gorillas and mountain gorillas from the Virungas. However, we did not detect the linkage group III (DRB1/6/5) in any of the individuals [Bontrop, 1997; Lukas et al., 2004]. Whereas multiple allelic combinations were also reported for the linkage group I [Lukas et al., 2004], we observed lineage polymorphism only for linkage groups II and DR8 (Fig. 4). The total number of linkage groups observed differed markedly between the gorilla subspecies. Western lowland gorillas exhibited six different combinations of linkage groups whereas in Cross River gorillas only two were identified. In mountain gorillas from Bwindi and the Virungas, we found four and three different combinations, respectively.

DISCUSSION

In the present study, we used massively parallel short-read sequencing for the characterization of MHC class II B genes using fecal samples derived from four wild gorilla populations. By applying strict similarity- and frequency-based analysis steps, we demonstrate that MHC genotyping from fecal samples is feasible but that stringent control of errors and artefacts is essential. In our analysis, we accepted only alleles that passed our validation tests, and so may have underestimated the variation in our sample. Nonetheless, and despite our moderate sample size, we here present the most comprehensive characterization of MHC variation in gorillas thus far.

MHC Class II Diversity in Gorillas

Among the gorillas analyzed, we identified a total of eight DPB1 alleles including six alleles which have not been previously described. Given that previous knowledge of DPB1 variation in gorillas was solely based on the characterization of six western lowland gorillas, our study contributes significantly to the understanding of DPB1 variation in gorillas [Gyllensten et al., 1996]. Phylogenetic tree analysis using our and previously described Gogo-DPB1 exon 2 sequences indicates the existence of two

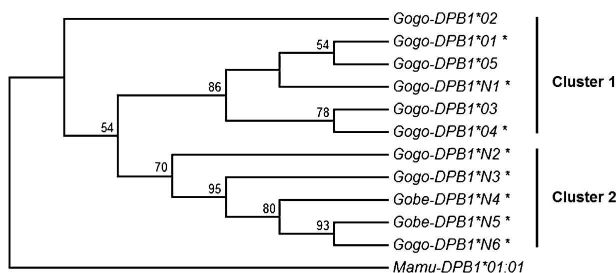

Fig. 1. Phylogenetic tree of gorilla DPB1 exon 2 sequences with the rhesus macaque allele Mamu-DPB1*01:01 taken as the outgroup [Otting et al., 1998]. Relevant bootstrap values over 50% are shown. Asterisks indicate alleles reported in this study.

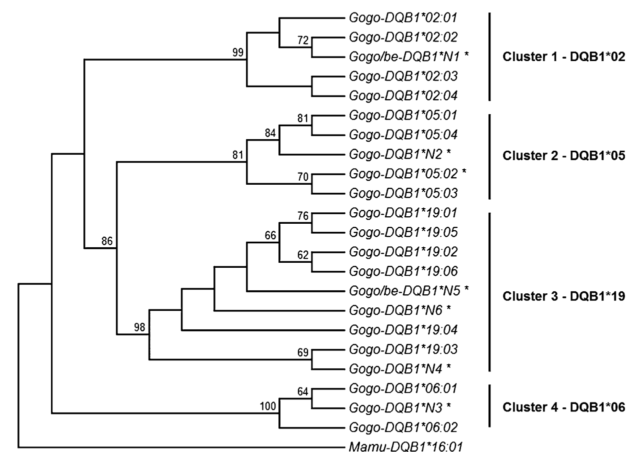

Fig. 2. Phylogenetic tree of gorilla DQB1 exon 2 sequences with the rhesus macaque allele Mamu-DQB1*16:01 taken as the outgroup [Otting et al., 1992]. Relevant bootstrap values over 50% are shown. Asterisks indicate alleles reported in this study. Gogo/be-DQB1 alleles were found in both western and eastern gorillas.

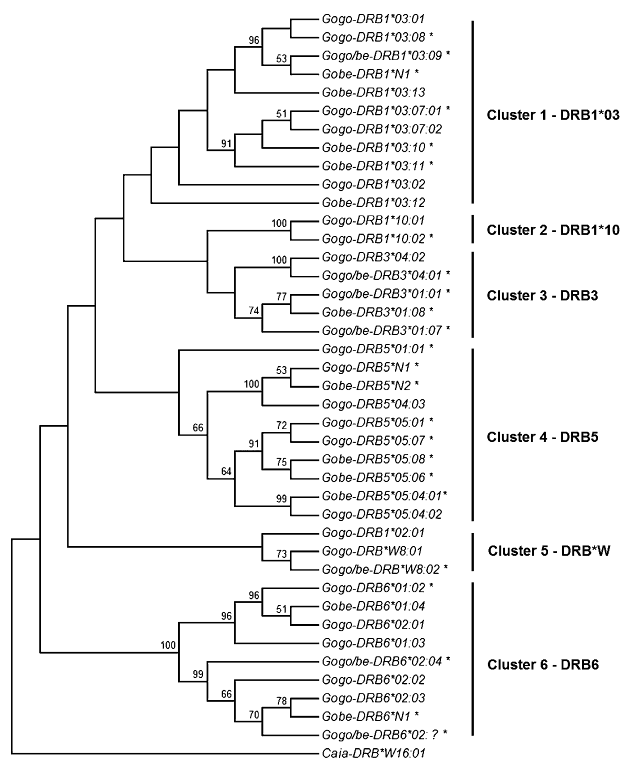

Fig. 3. Phylogenetic tree of gorilla DRB exon 2 sequences with the common marmoset allele Caja-DRB*W16:01 taken as the outgroup [Doxiadis et al., 2006]. Relevant bootstrap values over 50% are shown. Asterisks indicate alleles reported in this study. Gogo/be-DRB alleles were found in both western and eastern gorillas. Due to single synonymous substitutions and therefore identical amino acid sequences, the allele Gogo-DRB1*03:14 was renamed to Gogo-DRB1*03:07:02 whereas the allele Gogo-DRB5*05:04:02 was previously named Gogo-DRB5*05:05.

clusters including a newly described allelic lineage. Although it has been shown that the DPB1 locus displays extensive polymorphism in other hominids both in terms of allelic lineages and with regard to the number of alleles, gorillas seem to display comparatively low levels of variation with alleles mostly generated through point mutations rather than recombination-like events [Gyllensten et al., 1996; Otting et al., 1998]. However, given our moderate sample size and possible non-amplification of alleles, the characterization of additional gorillas or use of more extensive replication might increase the number of variants detected at the DPB1 locus.

For the DQB1 locus, we identified seven alleles including six alleles which have not been previously described. However, the majority of novel DQB1 exon 2 sequences differed only by a single non-synonymous substitution to previously described alleles. Consistent with the pattern of trans-species polymorphism observed for the DQB1 locus, that is, the occurrence of similar alleles or allelic lineages across species borders, identical alleles were identified between both gorilla species analyzed [Klein et al., 1998]. Although, as with the DPB1 locus, allelic polymorphism of the DQB1 locus is mainly generated through point mutations, the Gogo-DQB1*19 lineage in particular appears to evolve at a higher rate reflected by the accumulation of nucleotide substitutions [Otting et al., 1992]. In conclusion, gorillas seem to have maintained relatively high DQB1 variation comparable to chimpanzees and humans [Gaur et al., 1992; Gyllensten et al., 1990].

Not surprisingly, the majority of identified gorilla DRB alleles have been previously described since most studies focused on this particular locus [Acevedo-Whitehouse & Cunningham, 2006; Lukas et al., 2004]. However, unlike in chimpanzees and humans, the Gogo-DRB5 locus appears to be as polymorphic as the DRB1 locus whereas the DRB3 locus in gorillas displays an intermediate level of variability [Kenter et al., 1992]. The DRB region in human and chimpanzees is known to display both copy number and allelic variation [Brändle et al., 1992; Böhme et al., 1985; Slierendregt & Bontrop, 1994]. In chimpanzees, nine different region configurations with one to three functional DRB loci have been described [Bontrop, 1997; de Groot et al., 2009]. In contrast, only five DRB linkage groups are known in gorillas which suggest that compared to chimpanzees, gorillas are less variable in the DRB region [Bontrop, 1997; Lukas et al., 2004]. However, given the plasticity of the DRB region in primates and the limited number of gorillas analyzed thus far, we cannot completely rule out that we might have underestimated the allelic variation in gorillas at these loci.

In summary, gorillas appear to have comparatively low levels of MHC class II B polymorphism. Indeed, genome-wide data have shown that in western gorillas and mountain gorillas there is a marked reduction of genetic diversity close to the MHC indicative of a recent selective sweep [Scally et al., 2013; Xue et al., 2015]. It is noteworthy that a selective sweep in the MHC class I region of chimpanzees has been reported which also affected other regions of the MHC including the DRB region, possibly due to a retroviral infection [de Groot et al., 2002, 2008, 2010]. However, given our modest sample size and the possibility of sampling genetically homogeneous gorillas from substructured populations as well as potential underestimation of diversity due to non-amplification of alleles, at this point only tentative conclusions should be drawn from this study. We therefore strongly encourage more extensive studies that will enhance our knowledge of MHC variation in gorillas.

Interspecies Comparisons of MHC Class II Diversity in Gorillas

Comparison between the two gorilla species revealed that there are pronounced differences in

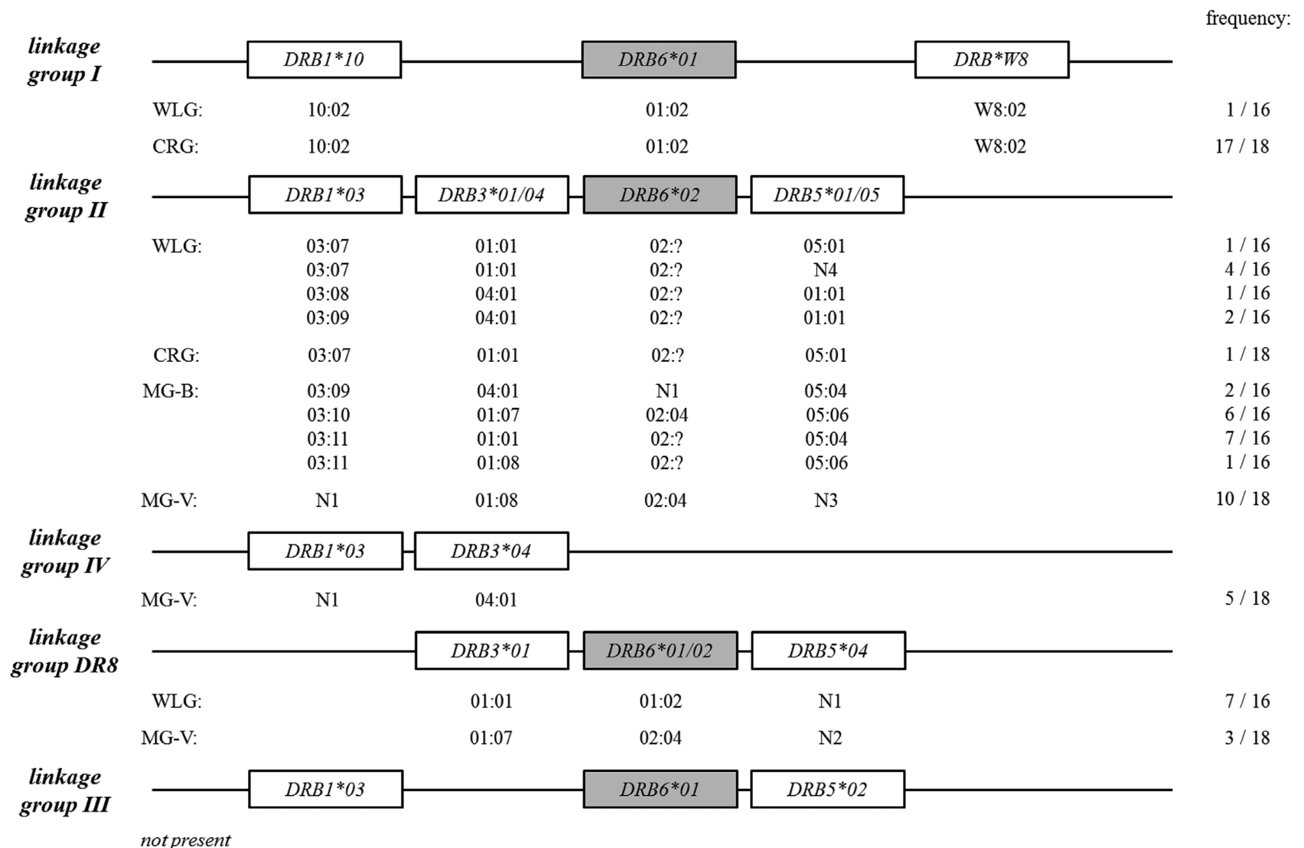

Fig. 4. DRB linkage groups identified in gorillas from the present study. White boxes represent functional loci whereas gray boxes indicate pseudogenes. Alleles present per linkage group as well as linkage group frequencies are listed for each gorilla subspecies: WLG, western lowland gorillas; CRG, Cross River gorillas; MG-B, mountain gorillas from Bwindi; MG-V, mountain gorillas from Virunga. DRB6*02:? Could not unambiguously be assigned to an allele (see text for details). Linkage group III was not detected in the present study.

the MHC class II B diversity, with western gorillas being the more diverse (Table II). This is in concordance with previous findings focusing on neutral or large-scale genomic autosomal variation showing that western gorillas have about twofold higher genetic diversity than eastern gorillas. This has been attributed to the higher effective population size (N_e) in the western species, both historically and today [Fünfstück & Vigilant, 2015; Scally et al., 2012, 2013; Thalmann et al., 2007]. Western gorillas are estimated today to number perhaps 100,000 individuals whereas eastern gorillas may number in the tens of thousands [IUCN, 2013]. While western gorillas are more or less continuously distributed throughout Central Africa, except for the Cross River gorilla which is restricted to the border region between Nigeria and Cameroon, eastern gorillas are patchily distributed primarily in the eastern Democratic Republic of Congo including mountain gorillas which are found only in two remnant and isolated populations in the Virunga Massif and the Bwindi Impenetrable Forest. Besides the higher effective population size, ongoing gene flow between western lowland gorilla populations might explain the maintenance of comparably high levels of genetic diversity despite the observation of moderate substructure within the western lowland gorilla population [Fünfstück et al., 2014; Scally et al., 2013; Thalmann et al., 2007].

In contrast, lower genetic diversity of eastern lowland gorillas is assumed to have resulted from lower ancestral N_e consistent with the smaller present-day population size. In addition, elevated levels of homozygosity indicate that eastern gorillas have been subjected to recurrent inbreeding both in the past and present indicating small and fragmented populations [Prado-Martinez et al., 2013; Scally et al., 2012]. Indeed, strong regional mtDNA differentiation among eastern gorilla populations suggests that eastern gorillas experienced a postglacial expansion from formerly isolated refugia [Anthony et al., 2007]. However, if the low MHC diversity observed in mountain gorillas also reflects patterns within eastern lowland gorillas remains unclear at this point.

Nevertheless, similar climatic fluctuations causing environmental changes during the Pleistocene have been proposed for the expansion and contraction

of the western gorilla population, ultimately leading to the isolation of the Cross River gorilla [Thalmann et al., 2011]. Initial divergence time between the ancestral western and Cross River gorilla populations has been dated to approximately 20,000–70,000 years ago with ongoing gene flow as long as 400 years ago [McManus et al., 2015; Thalmann et al., 2011]. In accordance, identical MHC alleles were found in Cross River gorillas and western lowland gorillas, respectively. However, MHC variation found in Cross River gorillas constitutes only a fraction of that in western lowland gorillas suggesting either samples were derived from subpopulations or the occurrence of a recent bottleneck leading to the low MHC diversity in this subspecies. Indeed, there is strong evidence that Cross River gorillas experienced a severe reduction in population size during the last 100–300 years, probably due to increased anthropogenic pressure [Bergl et al., 2008; Thalmann et al., 2011].

Today, major threats responsible for the drastic decline and isolation of gorilla populations include hunting for bushmeat, habitat loss, and fragmentation as well as spread of zoonotic diseases, particularly Ebola outbreaks [Bermejo et al., 2006; Junker et al., 2012; Le Gouar et al., 2009; Walsh et al., 2003]. Exceptionally, mountain gorilla populations are increasing as revealed by recent censuses [Gray et al., 2013; Robbins et al., 2011; Roy et al., 2014a]. Surrounded by high-density human populations, approximately 480 mountain gorillas of the Virunga Massif population are separated by only 30 km from the approximately 400 mountain gorillas living in the Bwindi Impenetrable Forest [Gray et al., 2013]. Recent estimates suggest an initial split time between the ancestral eastern lowland and mountain gorilla populations of about 10,000 years ago, followed by the split of the two mountain gorilla populations 5,000 years ago. Subsequent to their split from eastern lowland gorillas, mountain gorillas decreased substantially in effective population size [Bergl et al., 2008; Roy et al., 2014b]. As inferred by autosomal markers, genetic diversity measures of the two mountain gorilla populations are relatively similar. However, slightly higher values observed in Bwindi mountain gorillas suggest a higher ancestral N_e in this population [Bergl et al., 2008; Roy et al., 2014b] and in accordance, we found slightly higher MHC diversity in mountain gorilla from Bwindi than from the Virungas. Interestingly, none of the DRB linkage groups and only a few of the DRB alleles are shared between the mountain gorilla populations which suggest that the split between the two mountain gorillas was relatively abrupt without persisting gene flow between the two populations, or that variants were prone to stochastic loss in these small populations. The earlier study by Lukas et al. [2004] found comparable levels of DRB variation between mountain gorillas from Bwindi and western lowland gorillas. We also do not find pronounced differences in the total number of functional DRB alleles and linkage groups between mountain gorillas from Bwindi and western lowland gorillas. However, with regard to the other MHC class II loci analyzed the reduction in the total number of alleles observed in Bwindi mountain gorillas as compared to the western lowland gorillas is striking, which highlights the need to characterize more than one MHC locus in future studies aiming towards a more comprehensive understanding of the selective forces driving MHC diversity.

In summary, our results show that the gorilla populations studied display pronounced differences in variation at MHC class II loci which can be best explained by their different demographic histories.

Applicability of Non-Invasive Samples for MHC Sequencing

While cloning and Sanger sequencing approaches remain the “gold standard” for MHC characterization, NGS has the potential to render such traditional typing methods almost obsolete [Babik, 2010; Huchard & Pechouskova, 2014; Sommer et al., 2013; Wegner, 2009]. Although the possibility to simultaneously sequence millions of molecules appears appealing, the issues associated with amplicon-based MHC genotyping using NGS are still the same as for traditional typing methods [Sommer et al., 2013]. In the following, we provide an overview of the most common errors we encountered during our analysis. These are not exclusively but particularly related to the use of non-invasive samples and may therefore act as a guideline for beginners.

Even though analysis of multiple replicates is a standard approach in traditional MHC genotyping methods the majority of studies employing NGS omit or include only a limited number of replicates [e.g., Babik et al., 2009; Kloch et al., 2010; Promerová et al., 2012; Stutz & Bolnick, 2014]. However, recently it has been shown that independent replicates can detect the majority of PCR-generated artefacts, therefore highlighting their importance [Galan et al., 2010; Lighten et al., 2014; Sommer et al., 2013].

In the present study, DNA sequences were generated from two independent PCR replicates and after assignment of loci, we observed allelic dropout rates of 3–13% consistent with estimates reported in genotyping studies using fecal samples [Morin et al., 2001]. Allelic dropout promoted by low copy number templates has been identified as a major cause of error in other contexts which led to the application of methodological standards including multiple PCR replicates [Arandjelovic et al., 2009; Morin et al., 2001; Perry et al., 2010; Taberlet et al., 1999]. However, allelic dropout is also promoted by primer-template mismatches which can lead to

differences in amplification efficiencies of alleles [e.g., Babik, 2010; Kwok et al., 1990; Sommer et al., 2013]. Although methodological approaches, for example, touchdown PCR, can overcome non-complementary pairing to a certain extent, special emphasis should be placed on either the design or the choice of PCR primers in order to amplify the complete MHC variation. In non-model organisms like gorillas, however, there is typically little information available in terms of reference sequences. Therefore, amplification of MHC loci is often accomplished through the use of primers established in more- or less-related species, so called cross-species amplification, which can lead to an underestimation of MHC diversity. We thus cannot completely rule out that we might have missed certain alleles due to the non-amplification of certain alleles, particularly for the DRB region. Our levels of allelic dropout were high but consistent with those reported in microsatellite genotyping studies, and suggest that, if using the approach described here, more than two replicates should be analyzed to ensure detection of all variants and adequately describe MHC variation in gorillas [Arandjelovic et al., 2009].

The incorporation of independent PCR replicates can also detect artefacts arising through cross-contamination between samples, a potentially major source of error, especially when simultaneous handling of many samples is required [Li & Stoneking, 2012; Lighten et al., 2014; Prado-Martinez et al., 2013]. Based on our rigorous data analysis which fully utilizes the advantage of NGS by treating all DNA sequences equally rather than applying arbitrarily chosen thresholds and takes advantage of existing MHC information in gorillas, we observed a cross-contamination rate of 4%. Although the number of sequences of alleles potentially introduced through cross-contamination never exceeded the frequency of real alleles present in both replicates, we cannot preclude the possibility of other samples being affected. Additional independent replicates are therefore highly recommended. However, we are aware of the fact that these validation steps can only be accomplished with knowledge about the MHC structure of the species of interest which is often not available in non-model organisms. Interestingly, in this context a comparative microsatellite genotyping study using blood and fecal samples found a cross-contamination rate of around 1% which suggests that this kind of artefact might be more frequent than assumed, even without obvious signs for it [Fernando et al., 2002; see also Navidi et al., 1992].

In addition, we identified alleles identical to human in three individuals, indicating contamination probably introduced sporadically during sampling, DNA extraction, PCR set-up or in early stages of the library preparation. Due to the fact that low copy number templates are particular prone to the introduction of spurious DNA, it is highly recommended to adhere to rigorous guidelines and coherent standards to obtain reliable results, including use of extensive replicates [Lukas & Vigilant, 2005].

Moreover, comparison of replicates enables the identification of artifactual sequences arising through polymerase misincorporations. If introduced during early cycles of the PCR, these can potentially exceed the frequency of real alleles. The choice of the DNA polymerase and specifically its error rate has a profound impact on the quality of sequencing data [Brandariz-Fontes et al., 2015]. Different enzymes vary in their processivity and specificity, especially in the presence of templates of inferior quality and/or with low-copy number [Arezi et al., 2003; Moretti et al., 1998]. Modern DNA polymerases which combine proofreading activity with enhanced processivity, so-called high-fidelity DNA polymerases, dramatically reduce the incorporation of incorrect nucleotides. However, a growing number of studies suggest that the increased processivity also favors the formation of chimeric sequences [e.g., Ahn et al., 2012; Dabney & Meyer, 2012; Qiu et al., 2001].

Nevertheless, chimeric sequences are expected to accumulate during late PCR cycles, especially when an excessive number of PCR cycles are used [e.g., Lenz & Becker, 2008; Smyth et al., 2010]. Indeed, these issues are particularly important when dealing with low copy number templates such as fecal DNA, as has been shown in the present study. Therefore, several technical modifications have been suggested to reduce chimera formation including the use of minimum numbers of PCR cycles [Lenz & Becker, 2008; Liu et al., 2014; Smyth et al., 2010; Thompson et al., 2002]. However, most of these modifications cannot be adopted for MHC studies using fecal DNA, therefore emphasizing the need for strict data analysis. In this context, it is worth noting that chimeric sequences falsely identified as rare alleles may be more common than expected as has recently been shown in humans [Holcomb et al., 2014]. This might be of particular concern for MHC genotyping in non-model organisms since the high-throughput of NGS may erroneously inflate diversity.

Although many studies have recently encountered these issues, until now there is no clear consensus despite some suggestions on how to analyze MHC sequence data generated by NGS [Babik et al., 2009; Ferrandiz-Rovira et al., 2015; Lighten et al., 2014; Sommer et al., 2013; Stutz & Bolnick, 2014]. We therefore emphasize the call for rigorous and explicit data analysis standards in order to compare newly generated MHC datasets.

In sum, we have shown that non-invasive samples may be suitable for surveys of MHC variation in wildlife and can reveal the geographic distribution of variation more effectively than samples from captive individuals, but that extreme care is necessary in order to obtain valid results.

ACKNOWLEDGMENTS

We would like to thank Roland Schröder, Matthias Meyer and Janet Kelso for helpful discussions. This work was funded by the Max Planck Society.

REFERENCES

Acevedo-Whitehouse K, Cunningham AA. 2006. Is MHC enough for understanding wildlife immunogenetics? Trends in Ecology and Evolution 21:433–438.

Ahn JH, Kim BY, Song J, Weon HY. 2012. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. Journal of Microbiology 50:1071–1074.

Anthony NM, Johnson-Bawe M, Jeffery K, et al. 2007. The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. Proceedings of the National Academy of Sciences 104:20432.

Arandjelovic M, Guschanski K, Schubert G, et al. 2009. Two-step multiplex polymerase chain reaction improves the speed and accuracy of genotyping using DNA from noninvasive and museum samples. Molecular Ecology Resources 9:28–36.

Arandjelovic M, Bergl RA, Ikfuingei R, et al. 2015. Detection dog efficacy for collecting faecal samples from the critically endangered Cross River gorilla (*Gorilla gorilla diehli*) for genetic censusing. Royal Society Open Science 2:140423.

Arezi B, Xing W, Sorge JA, Hogrefe HH. 2003. Amplification efficiency of thermostable DNA polymerases. Analytical Biochemistry 321:226–235.

Babik W. 2010. Methods for MHC genotyping in non-model vertebrates. Molecular Ecology Recourses 10:237–251.

Babik W, Taberlet P, Ejsmond MJ, Radwan J. 2009. New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. Molecular Ecology Resources 9:713–719.

Bentley G, Higuchi R, Hoglund B, et al. 2009. High-resolution, high-throughput HLA genotyping by next-generation sequencing. Tissue Antigens 74:393–403.

Bergl RA, Bradley BJ, Nsubuga A, Vigilant L. 2008. Effects of habitat fragmentation, population size and demographic history on genetic diversity: the Cross River gorilla in a comparative context. American Journal of Primatology 70:848–859.

Bermejo M, Rodríguez-Teijeiro JD, Illera G, et al. 2006. Ebola outbreak killed 5000 gorillas. Science 314:1564.

Böhme J, Andersson M, Andersson G, et al. 1985. HLA-DR beta genes vary in number between different DR specificities, whereas the number of DQ beta genes is constant. Journal of Immunology 135:2149–2155.

Bontrop RE. 1994. Nonhuman primate MHC-DQA and -DQB second exon nucleotide sequences: a compilation. Immunogenetics 39:81–92.

Bontrop RE. 1997. The DR loci. In: Blancher A, Klein J, Socha WW, editors. Molecular biology and evolution of blood groups and MHC antigens in primates. New York: Springer. p 358–371.

Brandariz-Fontes C, Camacho-Sanchez M, Vilà C, et al. 2015. Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results. Scientific Reports 5:8056.

Brändle U, Ono H, Vincek V, et al. 1992. Trans-species evolution of MHC-DRB haplotype polymorphism in primates: organization of DRB genes in the chimpanzee. Immunogenetics 36:39–48.

Corell A1, Morales P, Varela P, et al. 1992. Allelic diversity at the primate major histocompatibility complex DRB6 locus. Immunogenetics 36:33–38.

Dabney J, Meyer M. 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. Biotechniques 52:87–94.

de Groot NG, Otting N, Doxiadis GG, et al. 2002. Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proceedings of the National Academy of Sciences 99:11748–11753.

de Groot NG, Heijmans CM, de Groot N, et al. 2008. Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics. Molecular Ecology 17:2074–2088.

de Groot NG, Heijmans CM, de Groot N, et al. 2009. The chimpanzee MHC-DRB region revisited: gene content, polymorphism, pseudogenes, and transcripts. Molecular Immunology 47:381–389.

de Groot NG, Heijmans CM, Zoet YM, et al. 2010. AIDS-protective HLA-B*27/B*57 and chimpanzee MHC class I molecules target analogous conserved areas of HIV-1/SIVcpz. Proceedings of the National Academy of Sciences 107:15175–15180.

de Groot NG, Otting N, Robinson J, et al. 2012. Nomenclature report on the major histocompatibility complex genes and alleles of Great Ape, Old and New World monkey species. Immunogenetics 64:615–631.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research 19:4008.

Doxiadis GG, van der Wiel MK, Brok HP, et al. 2006. Reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a New World primate species. Proceedings of the National Academy of Sciences 103:5864–5868.

Doxiadis GG, de Groot N, de Groot NG, Doxiadis II, Bontrop RE. 2008. Reshuffling of ancient peptide binding motifs between HLA-DRB multigene family members: old wine served in new skins. Molecular Immunology 45:2743–2751.

Fan WM, Kasahara M, Gutknecht J, et al. 1989. Shared class II MHC polymorphisms between humans and chimpanzees. Human Immunology 26:107–121.

Fernando P, Vidya TN, Rajapakse C, Dangolla A, Melnick DJ. 2003. Reliable noninvasive genotyping: fantasy or reality. Journal of Heredity 94:115–123.

Ferrandiz-Rovira M, Bigot T, Allainé D, Callait-Cardinal MP, Cohas A. 2015. Large-scale genotyping of highly polymorphic loci by next-generation sequencing: how to overcome the challenges to reliably genotype individuals. Heredity 114:485–493.

Figueroa F, O'hUigin C, Inoki H, Klein J. 1991. Primate DRB6 pseudogenes: clue to the evolutionary origin of the HLA-DR2 haplotype. Immunogenetics 34:324–337.

Fünfstück T, Vigilant L. 2015. The geographic distribution of genetic diversity within gorillas. American Journal of Primatology. doi: 10.1002/ajp.22427.

Fünfstück T, Arandjelovic M, Morgan DB, et al. 2014. The genetic population structure of wild western lowland gorillas (*Gorilla gorilla gorilla*) living in continuous rain forest. American Journal of Primatology 76:868–878.

Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF. 2010. A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. BMC Genomics 11:296.

Gaur LK, Heise ER, Thurtle PS, Nepom GT. 1992. Conservation of the HLA-DQB2 locus in nonhuman primates. Journal of Immunology 149:2530.

Gray M, Roy J, Vigilant L, et al. 2013. Genetic census reveals increased but uneven growth of a critically endangered mountain gorilla population. Biological Conservation 158:230–238.

Groves C. 2001. Primate taxonomy. Washington, DC: Smithsonian Institution Press.

Gyllensten UB, Lashkari D, Erlich HA. 1990. Allelic diversification at the class II DQB locus of the mammalian major histocompatibility complex. Proceedings of the National Academy of Sciences 87:1835–1839.
Gyllensten U, Sundvall M, Ezcurra I, Erlich HA. 1991. Genetic diversity at class II DRB loci of the primate MHC. Journal of Immunology 146:4368–4376.
Gyllensten U, Bergström T, Josefsson A, Sundvall M, Erlich HA. 1996. Rapid allelic diversification and intensified selection at antigen recognition sites of the MHC class II DPB1 locus during hominoid evolution. Tissue Antigens 47:212–221.
Hecker KH, Roux KH. 1996. High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. Biotechniques 20:478–485.
Holcomb CL, Rastrou M, Williams TC, et al. 2014. Next-generation sequencing can reveal in vitro-generated PCR crossover products: some artifactual sequences correspond to HLA alleles in the IMGT/HLA database. Tissue Antigens 83:32–40.
Huber T, Faulkner G, Hugenholtz P. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20:2317–2319.
Huchard E, Pechouskova E. 2014. The major histocompatibility complex and primate behavioral ecology: new tools and future questions. International Journal of Primatology 35:11–31.
IUCN. 2014. IUCN Red List of Threatened Species. v2014.3. http://www.iucnredlist.org. Downloaded on 19 March 2014.
Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. p 21–132.
Junker J, Blake S, Boesch C, et al. 2012. Recent decline in suitable environmental conditions for African great apes. Diversity and Distributions 18:1077–1091.
Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099–1106.
Kanagawa T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). Journal of Bioscience and Bioengineering 96:317–323.
Kasahara M, Klein D, Vincek V, Sarapata DE, Klein J. 1992. Comparative anatomy of the primate major histocompatibility complex DR subregion: evidence for combinations of DRB genes conserved across species. Genomics 14:340–349.
Kenter M, Otting N, Anholts J, et al. 1992. MHC-DRB diversity of the chimpanzee (*Pan troglodytes*). Immunogenetics 37:1–11.
Kenter M, Otting N, de Weers M, et al. 1993. MHC-DRB and -DQA1 nucleotide sequences of three lowland gorillas: implications for the evolution of primate MHC class II haplotypes. Human Immunology 36:205–218.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111–120.
Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research 40:e3.
Klein J, Bontrop RE, Dawkins RL, et al. 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217–219.
Klein J, Sato A, Nagl S, O'hUigin C. 1998. Molecular trans-species polymorphism. Annual Review of Ecology and Systematics 29:1–21.
Kloch A, Babik W, Bajer A, Siński E, Radwan J. 2010. Effects of an MHC-DRB genotype and allele number on the load of gut parasites in the bank vole *Myodes glareolus*. Molecular Ecology 19:255–265.
Kwok S, Kellogg DE, McKinney N, et al. 1990. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucleic Acids Research 18:999–1005.
Le Gouar PJ, Vallet D, David L, Bermejo M, et al. 2009. How ebola impacts genetics of western lowland gorilla populations. PloS ONE 4:e8375.
Lenz TL, Becker S. 2008. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci: implications for evolutionary analysis. Gene 427:117–123.
Li M, Stoneking M. 2012. A new approach for detecting low-level mutations in next-generation sequence data. Genome Biology 13:R34.
Lighten J, van Oosterhout C, Paterson IG, 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 14:753–767.
Liu J, Song H, Liu D, et al. 2014. Extensive recombination due to heteroduplexes generates large amounts of artificial gene fragments during PCR. PloS ONE 9:e106658.
Lukas D, Vigilant L. 2005. Reply: facts, faeces and setting standards for the study of MHC genes using noninvasive samples. Molecular Ecology 14:1601–1602.
Lukas D, Bradley BJ, Nsubuga AM, et al. 2004. Major histocompatibility complex and microsatellite variation in two populations of wild gorillas. Molecular Ecology 13:3389–3402.
McManus KF, Kelley JL, Song S, et al. 2015. Inference of gorilla demographic and selective history from whole-genome sequence data. Molecular Biology and Evolution 32:600–612.
Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010:pdb.prot5448.
Meyerhans A, Vartanian JP, Wain-Hobson S. 1990. DNA recombination during PCR. Nucleic Acids Research 18:1687–1691.
MiSeq Reporter User Guide. 2014. http://www.support.illumina.com. Downloaded on 19 March 2014.
Moretti T, Koons B, Budowle B. 1998. Enhancement of PCR amplification yield and specificity using AmpliTaq Gold DNA polymerase. Biotechniques 25:716–722.
Morin PA, Chambers KE, Boesch C, Vigilant L. 2001. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). Molecular Ecology 10:1835–1844.
Navidi W, Arnheim N, Waterman MS. 1992. A multiple-tubes approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. American Journal of Human Genetics 50:347–359.
Nsubuga AM, Holzman J, Chemnick LG, Ryder OA. 2009. The cryptic genetic structure of the North American captive gorilla population. Conservation Genetics 11:161–172.
Odelberg SJ, Weiss RB, Hata A, White R. 1995. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. Nucleic Acids Research 23:2049–2057.
O'hUigin C, Bontrop R, Klein J. 1993. Nonhuman primate MHC-DRB sequences: a compilation. Immunogenetics 38:165–183.
Otting N, Bontrop BE. 1995. Evolution of the major histocompatibility complex DPA1 locus in primates. Human Immunology 42:184–187.
Otting N, Kenter M, van Weeren P, Jonker M, Bontrop RE. 1992. MHC-DQB repertoire variation in hominoid and Old World primate species. Journal of Immunology 149:461–470.

Otting N1, Doxiadis GG, Versluis L, et al. 1998. Characterization and distribution of MHC-DPB1 alleles in chimpanzee and rhesus macaque populations. Human Immunology 59:656–664.
Pääbo S, Irwin DM, Wilson AC. 1990. DNA damage promotes jumping between templates during enzymatic amplification. The Journal of Biological Chemistry 265:4718–4721.
Parham P, Ohta T. 1996. Population biology of antigen presentation by MHC class I molecules. Science 272:67–74.
Perry GH, Marioni JC, Melsted P, Gilad Y. 2010. Genomic-scale capture and sequencing of endogenous DNA from feces. Molecular Ecology 19:5332–5344.
Prado-Martinez J, Sudmant PH, Kidd JM, et al. 2013. Great ape genetic diversity and population history. Nature 499:471–475.
Promerová M, Babik W, Bryja J, et al. 2012. Evaluation of two approaches to genotyping major histocompatibility complex class I in a passerine-CE-SSCP and 454 pyrosequencing. Molecular Ecology Resources 12:285–292.
Qiu X, Wu L, Huang H, et al. 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Applied and Environmental Microbiology 67:880–887.
Renaud G, Stenzel U, Kelso J. 2014. leeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Research 42:141.
Robbins MM, Gray M, Fawcett KA, et al. 2011. Extreme conservation leads to recovery of the Virunga mountain gorillas. PLoS ONE 6:e19788.
Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SG. 2013. IPD: the Immuno Polymorphism Database. Nucleic Acids Research 41:D1234–D1240.
Rohland N, Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research 2:939–946.
Roux KH. 1994. Using mismatched primer-template pairs in touchdown PCR. Biotechniques 16:812–814.
Roy J, Vigilant L, Gray M, et al. 2014a. Challenges in the use of genetic mark-recapture to estimate the population size of Bwindi mountain gorillas (*Gorilla beringei beringei*). Biological Conservation 180:249–261.
Roy J, Arandjelovic M, Bradley BJ, et al. 2014b. Recent divergences and size decreases of eastern gorilla populations. Biology Letters 10:20140811.
Scally A, Dutheil JY, Hillier LW, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. Nature 483:169–175.
Scally A, Yngvadottir B, Xue Y, et al. 2013. A genome-wide survey of genetic variation in gorillas using reduced representation sequencing. PLoS ONE 8:e65066.
Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863–864.
Slierendregt BL, Bontrop RE. 1994. Current knowledge on the major histocompatibility complex class II region in non-human primates. European Journal of Immunogenetics 21:391–402.
Smyth RP, Schlub TE, Grimm A, et al. 2010. Reducing chimera formation during PCR amplification to ensure accurate genotyping. Gene 469:45–51.
Sommer S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Frontiers in Zoology 2:16.
Sommer S, Courtiol A, Mazzoni CJ. 2013. MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. BMC Genomics 14:542.
Stuglik MT, Radwan J, Babik W. 2011. jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. Molecular Ecology Resources 11:739–742.
Stutz WE, Bolnick DI. 2014. Stepwise threshold clustering: a new method for genotyping MHC loci using next-generation sequencing technology. PLoS ONE 9:e100587.
Taberlet P, Waits LP, Luikart G. 1999. Noninvasive genetic sampling: look before you leap. Trend in Ecology and Evolution 14:323–327.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30:2725–2729.
Thalmann O, Fischer A, Lankester F, Pääbo S, Vigilant L. 2007. The complex evolutionary history of gorillas: insights from genomic data. Molecular Biology and Evolution 24:146–158.
Thalmann O, Wegmann D, Spitzner M, et al. 2011. Historical sampling reveals dramatic demographic changes in western gorilla populations. BMC Evolutionary Biology 11:85.
Thompson JR, Marcelino LA, Polz MF. 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucleic Acids Research 30:2083–2088.
Walsh P, Abernathy K, Bermejo M, et al. 2003. Catastrophic ape decline in western equatorial Africa. Nature 422: 611–614.
Watts C. 1997. Capture and processing of exogenous antigens for presentation on MHC molecules. Annual Review of Immunology 15:821–850.
Wegner KM. 2009. Massive parallel MHC genotyping: titanium that shines. Molecular Ecology 18:1818–1820.
Xue Y, Prado-Martinez J, Sudmant PH, et al. 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science 348:242–245.
Zagalska-Neubauer M, Babik W, Stuglik M, et al. 2010. 454 sequencing reveals extreme complexity of the class II major histocompatibility complex in the collared flycatcher. BMC Evolutionary Biology 10:395.