

# MHC class I diversity in chimpanzees and bonobos

Vincent Maibach<sup>1</sup> · Jörg B. Hans<sup>1</sup> · Christina Hvilsom<sup>2</sup> · Tomas Marques-Bonet<sup>3,4,5</sup> · Linda Vigilant<sup>1</sup>

Received: 15 February 2017 / Accepted: 20 April 2017 / Published online: 16 June 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Major histocompatibility complex (MHC) class I genes are critically involved in the defense against intracellular pathogens. MHC diversity comparisons among samples of closely related taxa may reveal traces of past or ongoing selective processes. The bonobo and chimpanzee are the closest living evolutionary relatives of humans and last shared a common ancestor some 1 mya. However, little is known concerning MHC class I diversity in bonobos or in central chimpanzees, the most numerous and genetically diverse chimpanzee subspecies. Here, we used a long-read sequencing technology (PacBio) to sequence the classical MHC class I genes *A*, *B*, *C*, and *A-like* in 20 and 30 wild-born bonobos and chimpanzees, respectively, with a main focus on central chimpanzees to assess and compare diversity in those two species. We describe in total 21 and 42 novel coding region sequences for the two species, respectively. In addition, we found evidence for a reduced MHC class I diversity in bonobos as compared to central chimpanzees as well as to

western chimpanzees and humans. The reduced bonobo MHC class I diversity may be the result of a selective process in their evolutionary past since their split from chimpanzees.

**Keywords** *Pan troglodytes* · *Pan paniscus* · PacBio · Great apes · Next-generation sequencing

## Introduction

The major histocompatibility complex (MHC), one of the most gene dense regions within higher vertebrates, plays a critical role in the mammalian immune response (Codner et al. 2012; Geraghty et al. 2002; Kelley et al. 2005). Genes within this complex code for cell surface proteins, which can be grouped into different classes. MHC class I molecules defend against intracellular pathogens by binding intracellular peptides and presenting them to cytotoxic T cells (Sommer 2005). The need to bind a wide range of intracellular peptides selects for a large number of different alleles, making the MHC one of the most polymorphic gene families within mammalian genomes (Geraghty et al. 2002; Kelley et al. 2005; Klein 1987; Kulski et al. 2002; Mayer et al. 1988; Rock et al. 2016; Trowsdale and Parham 2004). In addition, some MHC class I allotypes are involved in the interaction with killer cell immunoglobulin-like receptors (KIR receptors) expressed predominantly on natural killer cells (NK cells) which activate and regulate NK cell responses (Parham 2005; Trowsdale 2001).

Orthologues of the human class I genes *HLA-A*, *-B*, and *-C*, have been found in different great ape species, including chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) which shared a common ancestor with humans about 7 to 8 million years ago (Adams and Parham 2001; Bontrop 2006; Langergraber et al. 2012; Lawlor et al. 1988; Mayer et al. 1988; Prado-Martinez et al. 2013). Those loci are designated

**Electronic supplementary material** The online version of this article (doi:10.1007/s00251-017-0990-x) contains supplementary material, which is available to authorized users.

✉ Vincent Maibach  
vincent\_maibach@eva.mpg.de

- <sup>1</sup> Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
- <sup>2</sup> Copenhagen Zoo, Roskildevej 38, 2000 Frederiksberg, Denmark
- <sup>3</sup> Institute of Evolutionary Biology (UPF-CSIC), PRBB, Dr. Aiguader 88, 08003 Barcelona, Catalonia, Spain
- <sup>4</sup> Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010 Barcelona, Spain
- <sup>5</sup> CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain

in chimpanzees and bonobos as *Patr-A*, *Patr-B*, *Patr-C*, *Papa-A*, *Papa-B*, and *Papa-C*, respectively (Klein et al. 1990). Because polymorphism in the common ancestor may be retained or lost in speciating lineages, MHC alleles may exhibit trans-species polymorphism, in which some alleles from a certain MHC locus are more closely related to alleles of the orthologous locus in a different species than to alleles from the same locus in the same species (Gyllensten and Erlich 1989; Klein 1987; Klein et al. 1998; Lawlor et al. 1988; Mayer et al. 1988). Furthermore, not all MHC loci are present in all great apes and humans. For example, half of chimpanzees but no bonobos surveyed possess an *A-like* locus, which is nonpolymorphic and orthologous to the orangutan class I *A* and the 5' part of the *HLA-Y* pseudogene (Adams et al. 2001; Gleimer et al. 2011).

The presence or absence, as well as the relative diversity of alleles at MHC class I loci may give insights into the selective pressures shaping the evolutionary histories of these loci in great apes. For example, researchers proposed that exposure to SIVcpz or another highly pathogenic virus 2–3 million years ago may be responsible for the low diversity found in a sample of chimpanzee MHC class I loci *A*, *B* and *C* intron 2 sequences, as well as the low variation observed in a survey using microsatellite loci across the chimpanzee MHC class I and II regions (de Groot and Bontrop 2013; de Groot et al. 2008, 2002). More direct evidence for the influence of SIVcpz was inferred from a study of MHC class I *B* allelic variation in a wild population of 125 eastern chimpanzees (*P.t. schweinfurthii*) including three communities varying in SIVcpz prevalence (Wroblewski et al. 2015). The community with the highest prevalence of SIVcpz had significantly higher frequencies of particular *B* alleles, including an allele with high similarity to the *HLA-B\*57* allele associated in humans with less viral load and delayed progression of HIV infection (Wroblewski et al. 2015).

It is, however, important to note that sampling at nearly 90 field sites revealed SIV infection in only two of the four chimpanzee subspecies, namely the central (*P.t. troglodytes*) and eastern (*P.t. schweinfurthii*) subspecies, but not the western (*P.t. verus*) or Nigeria-Cameroonian (*P.t. ellioti*) subspecies (Heuverswyn et al. 2007; Keele et al. 2006; Rudicell et al. 2010; Santiago et al. 2003; Santiago et al. 2002; Sharp and Hahn 2011; Worobey et al. 2004). This could mean that SIVcpz, or other pathogens, may still be in the process of differentially shaping the MHC diversity of chimpanzee subspecies today. Comparisons are difficult because MHC class I diversity has largely been characterized from representatives of just the western or eastern subspecies. After over 20 years of MHC class I studies in chimpanzees, the number of individuals characterized from the different subspecies greatly varies, with data available from approximately 85 western chimpanzees and 130 eastern chimpanzees but only some 15 central chimpanzees, indicating that central chimpanzees are clearly underrepresented in the description of MHC class I diversity. Furthermore, it is interesting to note that individuals from the western and eastern chimpanzee

subspecies apparently have no common alleles at the *B* locus, suggesting that there may indeed be pronounced differences in MHC variation between the subspecies (Wroblewski et al. 2015). Chimpanzee subspecies also differ in demographic histories. Comparisons revealed highest genetic diversity and effective population size in the central chimpanzee (Becquet et al. 2007; Fischer et al. 2006, 2004; Hey 2010; Prado-Martinez et al. 2013; Wegmann and Excoffier 2010). Central chimpanzees also showed signs of population growth, whereas western chimpanzees and eastern chimpanzees showed signs of a decrease in effective population size (Caswell et al. 2008; de Manuel et al. 2016; Fischer et al. 2011, 2004). The different demographic histories of the subspecies which may have influenced also the shaping of the MHC diversity in the past, the different pathogenic pressures of the subspecies and the underrepresentation of central chimpanzees in the description of MHC class I diversity all present a compelling argument to analyze MHC diversity specifically in central chimpanzees. This would contribute to the understanding of the whole diversity in this species and could eventually reveal signs of selection in these important loci in central chimpanzees in particular.

The closest relative of the chimpanzee is the bonobo, and these species last shared a common ancestor between 1 and 2.6 mya (Langergraber et al. 2012; Prufer et al. 2012). As with the different chimpanzee subspecies, chimpanzees and bonobos may have experienced different selective pressures influencing the MHC diversity in those two species, and it is notable that SIVcpz has not been detected in bonobos (Li et al. 2012; Sharp and Hahn 2011). Although MHC variation has been characterized in far more than 100 chimpanzees, assessment of bonobo MHC class I diversity has been limited to a few individuals in various studies adding up to four, four, and nine individuals for the *A*, *B*, and *C* class I loci, respectively (Cooper et al. 1998; Lawlor et al. 1995; Martinez-Laso et al. 2006; McAdam et al. 1994; McAdam et al. 1995). Characterization of MHC variation in additional representatives of the bonobo as well as the chimpanzee will allow assessment of diversity within as well as across these species and may contribute to understanding the evolutionary history of the human MHC.

Although studies of great ape genome diversity are increasingly common (Cagan et al. 2016; de Manuel et al. 2016; Prado-Martinez et al. 2013; Scally et al. 2012) MHC variation is more difficult to assess due to the highly repetitive nature of the region and the short read lengths of most next-generation sequencing platforms, which both make the subsequent assembly and unambiguous reconstruction of the alleles challenging. However, the PacBio RS II system features sequencing of longer molecules and has been used to sequence complete MHC alleles, making this platform a practical solution for MHC typing of humans and non-human primates (Mayor et al. 2015; Westbrook et al. 2015). Here, we used the PacBio RS II system to sequence the MHC class I *A*, *B*, *C*, and *A-like* loci in a sample of wild-born bonobos and chimpanzees,

including three different chimpanzee subspecies. We use these data and published information from humans in order to describe the MHC class I diversity in bonobos and elucidate comparative levels of diversity between those species.

## Materials and methods

### Samples

We analyzed 21 bonobo (*P. paniscus*) and 30 chimpanzee (*P. troglodytes*) individuals, including representatives of three of the four chimpanzee subspecies (20 *P.t. troglodytes*, 4 *P.t. verus*, and 6 *P.t. schweinfurthii*). Blood samples of 20 bonobos (*P. paniscus*) and 20 central chimpanzees (*P.t. troglodytes*) were collected in African sanctuaries (Lola ya Bonobo sanctuary Kinshasa, Democratic Republic of Congo, and Tchimpounga sanctuary, Jane Goodall Institute, Republic of Congo, respectively) in 2007 and 2008 by Michel Halbwax and Anne Fischer (Fischer et al. 2011). Individuals were considered unrelated as described in Fischer et al. (2011). We also included one female bonobo (Ulindi, blood sample collected in the Leipzig Zoo in 2011), who was the source for the bonobo genome project (Prufer et al. 2012). Details regarding the samples are listed (Table S1). Genomic DNAs were extracted using the Gentra Puregene Blood Kit (QIAGEN) or QIAamp DNA Blood Mini Kit (QIAGEN). All individuals in this study were either born in the wild or have wild-born parents (Ulindi).

### DNA amplification

We amplified full-length PCR products for each of the four loci (*A*, *B*, *C*, and *A-like*) separately using a primer set for each locus designed to span all exons and introns (see Table S2). Primers for the *B* and *C* loci were taken from Hans et al. (2017). Products for *A*, *B*, *C*, and *A-like* targets were 3412–3440 bp, 3019–3049 bp, 3067–3080 bp long and 3156 bp, respectively. PCRs were set up in a 50  $\mu$ l volume containing 5 $\times$  Crimson LongAmp Taq Reaction Buffer (New England BioLabs, Ipswich, Massachusetts, USA), 2.5 mM of each dNTP, 10  $\mu$ M of each forward and reverse primer, 1 U of Crimson LongAmp Taq DNA Polymerase (New England BioLabs, Ipswich, Massachusetts, USA) and 50 ng DNA. Amplifications started with an initial denaturation step of 2 min at 94  $^{\circ}$ C, followed by 40 cycles of 20 s at 94  $^{\circ}$ C, 60 s at 68  $^{\circ}$ C, 8 min at 68  $^{\circ}$ C and finished by a final extension step of 15 min at 68  $^{\circ}$ C. We conducted for all individuals two independent PCR replicates for each locus and verified successful amplification with gel electrophoresis. Amplicons were purified with AGENCOURT AMPure XP PCR purification (Agencourt Bioscience Corporation, Beverly, Massachusetts, USA) following suppliers instructions, except that we changed

the ratio between PCR product and AMPure XP Magnetic Particle Solution to 1:1. DNA concentrations of purified amplicons were measured using NanoDrop 1000 (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Not all chimpanzees have the *A-like* locus, and it has not been detected in bonobos (Adams et al. 2001). To test whether our primers for the *A-like* locus (Table S2) were able to identify all individuals having the *A-like*<sup>+</sup> haplotype and confirm the lack of this locus in bonobos, for all samples, we performed an additional PCR based on a *Patr-A-like* typing system described in Adams et al. (2001). PCR was done using AmpliTaq Gold with GeneAmp Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and contained the following reagents in a total of 20  $\mu$ l: 10 $\times$  GeneAmp PCR Buffer II, 25 mM MgCl<sub>2</sub>, 2.5 mM dNTPs, 10  $\mu$ M primer AL4S, 10  $\mu$ M primer AL5R, 0.5 units AmpliTaq Gold DNA Polymerase, and 50 ng DNA. After 9 min at 95  $^{\circ}$ C, we used 30 cycles with 30 s at 94  $^{\circ}$ C, 30 s at 58  $^{\circ}$ C, 1 min at 72  $^{\circ}$ C and followed by a final elongation of 10 min at 72  $^{\circ}$ C to amplify the 423-bp product. Successful amplification was checked with gel electrophoresis.

### Pacific Bioscience library preparation

We prepared SMRTbell libraries for PacBio RS II sequencing following the manufacturer's protocol for Multiplex SMRT Sequencing with PacBio Barcoded Adapters (Pacific Biosciences, Menlo Park, CA, USA). Prior to library preparation, amplicons of the three (*A*, *B*, *C*) or four (*A*, *B*, *C*, *A-like*) different loci, respectively, were pooled in equimolar amounts for each replicate of each individual. Each library contained ten such pooled samples with a DNA concentration of 100 ng per sample. Library preparations were performed according to supplier's instructions except that after the damage repair step, we again added 1  $\mu$ l of DNA Damage Repair Mix (Pacific Biosciences, Menlo Park, CA, USA) to each library and incubated the mixture for 30 min at 37  $^{\circ}$ C. This second damage repair step improved the sequencing by preventing early terminations in the sequencing process due to damages in the DNA template. Sequencing was performed on a PacBio RS II system (Pacific Biosciences, Menlo Park, CA, USA) using P6-C4 chemistry and 240 min movies.

### Sequence data processing

We used the Long Amplicon Analysis tool based on the Quiver framework within the Pacific Bioscience SMRT Analysis System version 2.3.0.140936 (Pacific Biosciences, Menlo Park, CA, USA) to generate high-quality consensus sequences for each amplicon. The amplicons were demultiplexed and assigned to distinct loci according to their primer sequences. For each individual we compared results from the two replicates of each amplicon of each locus. In all but three

**Table 1** Chimpanzee MHC class I alleles obtained from 30 individuals. Novel alleles are written in bold. Novel alleles include changes either in the introns or in the coding region sequence or in the protein sequence (allotype). Novel protein sequences (allotypes) are indicated with an asterisk. The three different chimpanzee subspecies *Pt. schweinfurthii* (P.t.s.), *Pt. verus* (P.t.v.), and *Pt. troglodytes* (P.t.t.) are indicated

|  | Patr-A                | Patr-B                | Patr-C                | Patr-AL               | Subspecies |
|--|-----------------------|-----------------------|-----------------------|-----------------------|------------|
| Bihati                                   | A*15:01:01:01         | A*08:01:01:01         | <b>B*07:04*</b>       | <b>AL*01:01:01:04</b> | P.t.s.     |
| Cleo                                     | A*22:01               | A*08:01:01:01         | <b>B*38:02*</b>       | <b>AL*01:01:01:06</b> | P.t.s.     |
| Diana                                    | <b>A*27:01*</b>       | A*15:01:01:01         | B*22:04               | <b>AL*01:01:01:08</b> | P.t.s.     |
| Maya                                     | <b>A*15:01:01:02</b>  | <b>A*13:01:01:02</b>  | <b>B*23:05</b>        | C*09:02:01:01         | P.t.s.     |
| Tongo                                    | A*15:01:01:01         | A*08:01:01:01         | <b>B*30:02*</b>       | <b>AL*01:01:01:04</b> | P.t.s.     |
| Trixie                                   | A*15:02               | A*06:03               | <b>B*07:05*</b>       | AL*01:02              | P.t.s.     |
| Alice                                    | A*09:01               | <b>B*36:01</b>        | B*01:01               | AL*01:01:01:01        | P.t.v.     |
| Berta                                    | A*03:02               | <b>B*20:01:02</b>     | B*01:01               | AL*01:01:01:01        | P.t.v.     |
| Cindy                                    | A*05:01               | B*13:01               | C*12:01               | AL*01:01:01:01        | P.t.v.     |
| Linda                                    | A*06:02               | B*04:02               | C*16:01               | —                     | P.t.v.     |
| Agnagui                                  | <b>A*25:01:01:01*</b> | A*17:01:01:01         | <b>B*12:03*</b>       | —                     | P.t.t.     |
| Bailele                                  | <b>A*26:01*</b>       | A*17:04*              | B*18:01               | <b>AL*01:01:01:02</b> | P.t.t.     |
| Bayokele                                 | A*10:01               | A*18:01               | <b>B*11:03*</b>       | <b>AL*01:01:01:03</b> | P.t.t.     |
| Bimangou                                 | A*13:01:01:01         | <b>A*17:04*</b>       | B*07:02               | <b>AL*01:01:01:05</b> | P.t.t.     |
| Botsomi                                  | A*18:01               | B*07:03               | <b>B*35:02*</b>       | —                     | P.t.t.     |
| Casimir                                  | <b>A*27:01*</b>       | A*08:01:01:01         | <b>B*30:02*</b>       | <b>AL*01:01:01:06</b> | P.t.t.     |
| Castro                                   | A*13:01:01:01         | <b>A*25:01:01:02*</b> | B*22:04               | <b>AL*01:01:01:05</b> | P.t.t.     |
| Chinoc                                   | A*15:01:01:01         | <b>B*22:06*</b>       | B*06:03               | <b>AL*01:01:01:07</b> | P.t.t.     |
| Clara_T                                  | A*16:01               | <b>B*40:01:01:01*</b> | <b>B*22:06*</b>       | <b>AL*01:01:01:06</b> | P.t.t.     |
| Dzeke                                    | <b>A*25:01:01:01*</b> | A*12:01               | <b>B*11:03*</b>       | —                     | P.t.t.     |
| Elikia                                   | A*17:01:01:01         | <b>A*17:04*</b>       | <b>B*11:03*</b>       | —                     | P.t.t.     |
| Fan Tuek                                 | <b>A*03:06*</b>       | A*13:01:01:01         | <b>B*22:06*</b>       | <b>AL*01:04*</b>      | P.t.t.     |
| Gao                                      | <b>A*03:05*</b>       | A*15:01:01:01         | <b>B*19:03*</b>       | C*13:03               | P.t.t.     |
| Golfi                                    | <b>A*26:01*</b>       | A*18:01               | <b>B*40:01:01:01*</b> | C*03:02:02:01         | P.t.t.     |
| Grand Maitre                             | <b>A*03:07*</b>       | <b>A*15:03*</b>       | <b>B*23:07*</b>       | <b>AL*01:01:01:09</b> | P.t.t.     |
| Imphondo                                 | A*16:01               | <b>A*03:07</b>        | B*35:02               | AL*01:02              | P.t.t.     |
| Loufombou                                | A*17:01:01:01         | A*08:02               | <b>B*19:04*</b>       | <b>AL*01:01:01:05</b> | P.t.t.     |
| Lufino                                   | <b>A*08:01:01:02</b>  | A*18:01               | B*23:02               | <b>AL*01:01:01:06</b> | P.t.t.     |
| Marcelle                                 | A*18:01               | <b>B*23:02</b>        | B*06:02               | —                     | P.t.t.     |
| Moka                                     | A*18:01               | <b>B*33:01:01:01</b>  | <b>B*11:03*</b>       | —                     | P.t.t.     |
| No. of different alleles                 | 30                    | 41                    | 29                    | 11                    |            |
| No. of different coding region sequences | 25                    | 39                    | 26                    | 3                     |            |
| No. of different allotypes               | 25                    | 39                    | 25                    | 3                     |            |
| No. of novel alleles                     | 13                    | 22                    | 17                    | 9                     |            |
| No. of novel coding region sequences     | 8                     | 20                    | 13                    | 1                     |            |
| No. of novel allotypes                   | 8                     | 18                    | 11                    | 1                     |            |

**Table 2** Bonobo MHC class I alleles obtained from 21 individuals. Novel alleles are written in bold. Novel alleles include changes either in the introns or in the coding region sequence or in the protein sequence (allotype). Novel protein sequences (allotypes) are indicated with an asterisk

|  | <i>Papa-A</i>         |                       | <i>Papa-B</i>     |                   | <i>Papa-C</i>         |                       |
|--|-----------------------|-----------------------|-------------------|-------------------|-----------------------|-----------------------|
| Api                                      | <i>A*03:06*</i>       | <i>A*03:03:01:02*</i> | <i>B*01:03*</i>   | <i>B*07:01:02</i> | <i>C*04:01</i>        | <i>C*03:02</i>        |
| Bandundu                                 | <i>A*05:03*</i>       |                       | <i>B*07:01:02</i> |                   | <i>C*03:02</i>        |                       |
| Bili_L                                   | <i>A*06:03:01:01*</i> | <i>A*08:03*</i>       | <i>B*02:02</i>    | <i>B*15:01</i>    | <i>C*03:01:01:03</i>  | <i>C*06:01:01:02*</i> |
| Boende                                   | <i>A*03:06*</i>       |                       | <i>B*15:01</i>    |                   | <i>C*06:01:01:02*</i> |                       |
| Bolobo                                   | <i>A*05:03*</i>       |                       | <i>B*07:01:02</i> |                   | <i>C*03:02</i>        |                       |
| Fizi                                     | <i>A*03:02*</i>       | <i>A*08:03*</i>       | <i>B*09:01</i>    | <i>B*19:02</i>    | <i>C*04:02*</i>       | <i>C*03:02</i>        |
| Isiro                                    | <i>A*05:03*</i>       | <i>A*06:03:01:01*</i> | <i>B*02:02</i>    | <i>B*07:01:02</i> | <i>C*03:01:01:03</i>  | <i>C*03:02</i>        |
| Keza                                     | <i>A*09:02*</i>       | <i>A*08:02*</i>       | <i>B*01:03*</i>   | <i>B*17:01</i>    | <i>C*04:01</i>        | <i>C*05:02*</i>       |
| Kikwit                                   | <i>A*06:03:01:02*</i> | <i>A*05:03*</i>       | <i>B*09:01</i>    | <i>B*07:01:02</i> | <i>C*04:02*</i>       | <i>C*03:02</i>        |
| Kisantu                                  | <i>A*05:03*</i>       | <i>A*08:03*</i>       | <i>B*07:01:02</i> |                   | <i>C*03:02</i>        |                       |
| Kubulu                                   | <i>A*03:06*</i>       | <i>A*06:03:01:03*</i> | <i>B*15:01</i>    | <i>B*19:01</i>    | <i>C*06:01:01:02*</i> | <i>C*05:01*</i>       |
| Likasi                                   | <i>A*06:03:01:01*</i> |                       | <i>B*02:02</i>    |                   | <i>C*03:01:01:03</i>  |                       |
| Lipopo                                   | <i>A*03:05*</i>       |                       | <i>B*19:02</i>    | <i>B*16:01</i>    | <i>C*03:03:01:01*</i> | <i>C*05:03*</i>       |
| Lodja                                    | <i>A*08:03*</i>       | <i>A*03:02*</i>       | <i>B*07:01:02</i> | <i>B*19:02</i>    | <i>C*03:02</i>        | <i>C*03:03:01:01*</i> |
| Lomami                                   | <i>A*08:03*</i>       | <i>A*06:03:01:03*</i> | <i>B*19:02</i>    | <i>B*14:01</i>    | <i>C*03:03:01:01*</i> | <i>C*04:03*</i>       |
| Malou_L                                  | <i>A*08:03*</i>       | <i>A*04:02*</i>       | <i>B*19:02</i>    | <i>B*07:01:02</i> | <i>C*03:02</i>        | <i>C*03:03:01:01*</i> |
| Matadi                                   | <i>A*08:03*</i>       | <i>A*06:03:01:02*</i> | <i>B*15:02*</i>   | <i>B*07:01:02</i> | <i>C*03:02</i>        | <i>C*06:01:01:02*</i> |
| Max                                      | <i>A*05:03*</i>       | <i>A*06:03:01:01*</i> | <i>B*02:02</i>    | <i>B*15:01</i>    | <i>C*03:01:01:03</i>  | <i>C*06:01:01:02*</i> |
| Semwendwa                                | <i>A*06:02</i>        | <i>A*04:02*</i>       | <i>B*01:02*</i>   | <i>B*08:01:01</i> | <i>C*04:02*</i>       | <i>C*03:03:01:02</i>  |
| Tshilomba                                | <i>A*03:02*</i>       | <i>A*03:06*</i>       | <i>B*14:01</i>    | <i>B*07:01:02</i> | <i>C*03:02</i>        | <i>C*04:03*</i>       |
| Ulindi                                   | <i>A*05:03*</i>       | <i>A*03:02*</i>       | <i>B*07:01:02</i> |                   | <i>C*03:02</i>        |                       |
| No. of different alleles                 | 13                    |                       | 13                |                   | 11                    |                       |
| No. of different coding region sequences | 11                    |                       | 13                |                   | 10                    |                       |
| No. of different allotypes               | 11                    |                       | 13                |                   | 10                    |                       |
| No. of novel alleles                     | 13                    |                       | 3                 |                   | 8                     |                       |
| No. of novel coding region sequences     | 11                    |                       | 3                 |                   | 7                     |                       |
| No. of novel allotypes                   | 11                    |                       | 3                 |                   | 7                     |                       |

individuals (Malou\_L, Lodja, and Cindy), the replicate sequences were identical. For these three individuals, we found at one locus each (*Papa-C*, *Papa-C*, and *Patr-A*, respectively), two alleles in one replicate in contrast to only one allele in the other replicate. Furthermore, the solitary allele showed a deletion within the coding region at the position which distinguished the other two alleles, suggesting that the algorithm sometimes had difficulties in distinguishing two very similar alleles. For identification of the correct alleles for these three individuals, we used the minor variant protocol of the SMRT Analysis (Pacific Biosciences, Menlo Park, CA, USA), which aligns sequences to a reference and calls variants within that alignment. For each individual, we used in both replicates all three alleles as a reference separately, resulting in six analysis runs per individual. Furthermore, each alignment generated by the minor variant protocol was checked by eye, using samtools version 0.1.19 (Li et al. 2009) for visualizing the data. The minor variant analysis (data not shown) and

the visualization of the alignment confirmed that the two alleles found within one replicate are the true alleles of each individual. The fact that those alleles were already described in the literature or we found them in other individuals within our data set and alleles with the deletion had premature stop codons also supported this finding. The spurious alleles with the deletion were discarded. We sent all full-length MHC sequences to the IPD-MHC database for official designation of novel identified alleles (Groot et al. 2012).

### Checking for artifacts

To ensure that we had enough template for the sequencing process, we used 40 PCR cycles (see the “DNA amplification” section). Although within the range of the manufacturer’s recommendation, this is a relatively large number of cycles and may promote the formation of chimeric sequences (Kanagawa 2003; Lenz and Becker 2008). We



**Table 3** Comparison of the number of full-length alleles, coding region sequences, and protein sequences among bonobos ( $n = 21$ ), chimpanzees including all three subspecies ( $n = 30$ ), and chimpanzees only from the central subspecies ( $n = 20$ )

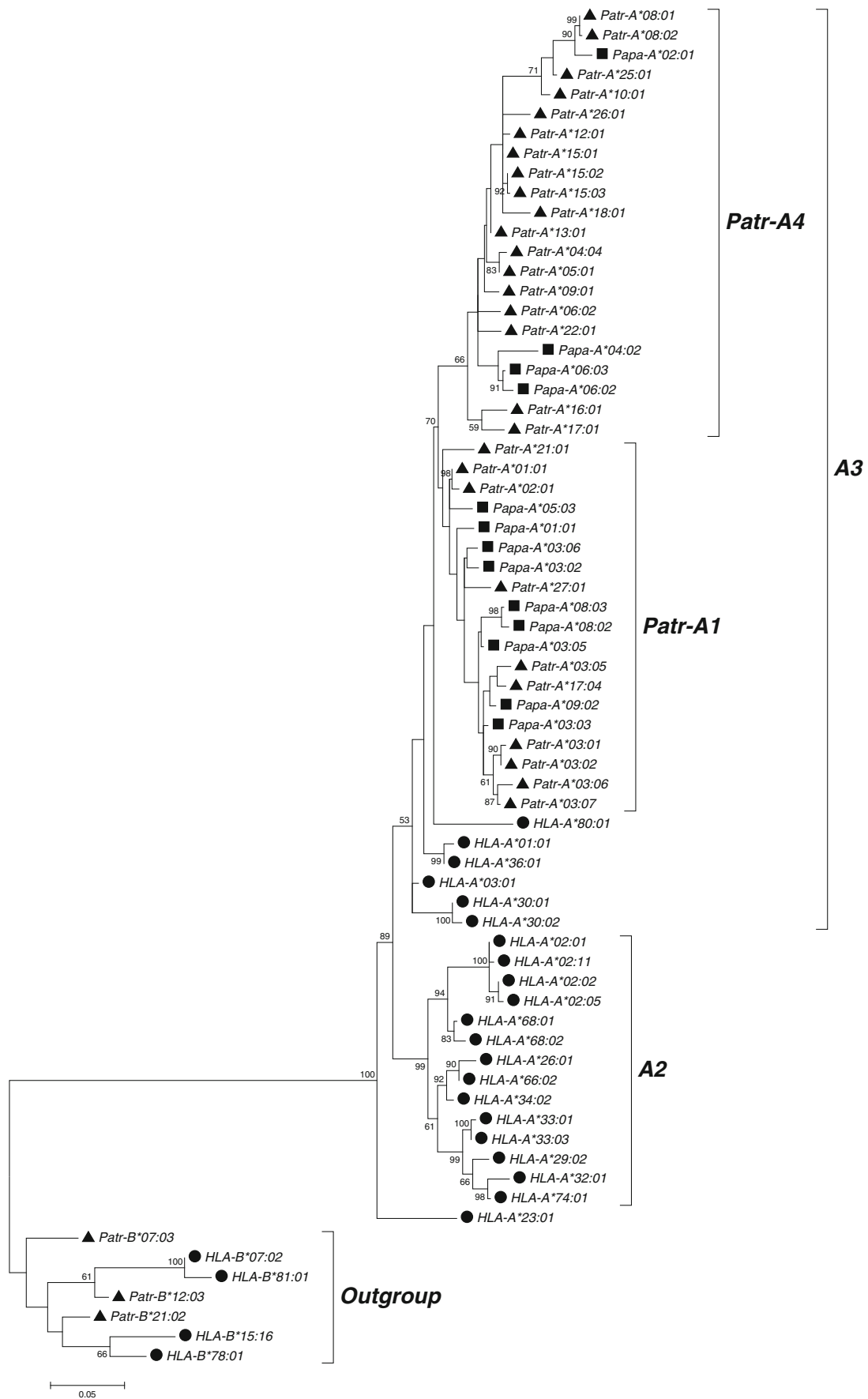
| Species                     | A      |                        |                         | B      |                        |                         | C      |                        |                         | A-like |                        |                         |
|-----------------------------|--------|------------------------|-------------------------|--------|------------------------|-------------------------|--------|------------------------|-------------------------|--------|------------------------|-------------------------|
|                             | Bonobo | Chimpanzee<br>(3 ssp.) | Chimpanzee<br>(central) | Bonobo | Chimpanzee<br>(3 ssp.) | Chimpanzee<br>(central) | Bonobo | Chimpanzee<br>(3 ssp.) | Chimpanzee<br>(central) | Bonobo | Chimpanzee<br>(3 ssp.) | Chimpanzee<br>(central) |
| No. full-length alleles     | 13     | 30                     | 20                      | 13     | 41                     | 26                      | 11     | 29                     | 20                      | 11     | 8                      |                         |
| No. coding region sequences | 11     | 25                     | 18                      | 13     | 39                     | 25                      | 10     | 26                     | 17                      | 3      | 3                      |                         |
| No. protein sequences       | 11     | 25                     | 18                      | 13     | 39                     | 25                      | 10     | 25                     | 16                      | 3      | 3                      |                         |

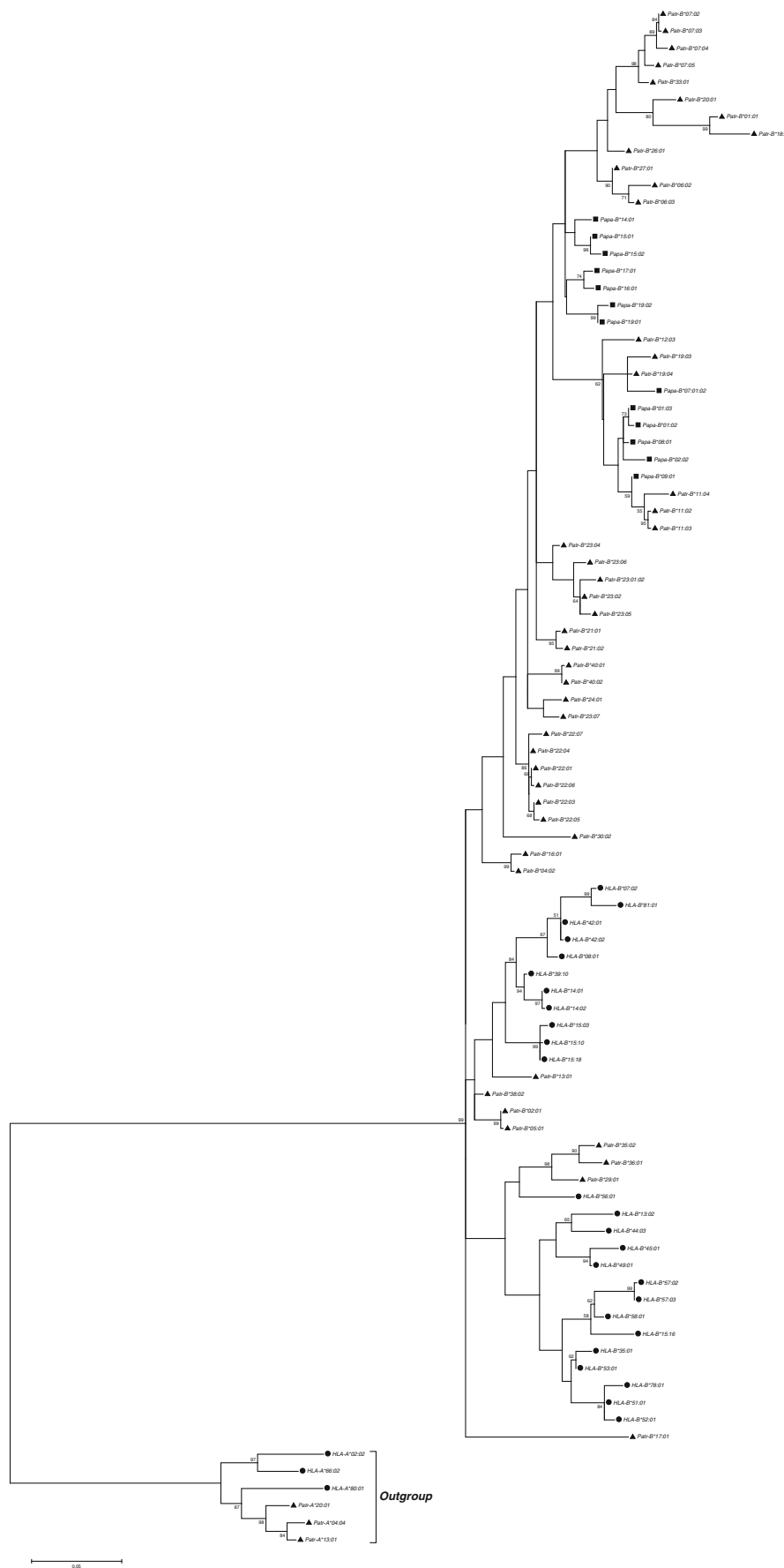
**Fig. 1** Molecular phylogenetic tree of MHC class I A sequences reconstructed using the maximum likelihood method based on the Tamura 3-parameter model. The model used a discrete gamma distribution allowing evolutionary rate differences among sites. The tree with the highest log likelihood ( $-5799.1552$ ) is shown. Confidence in the branching patterns is indicated by bootstrap values and only relevant values ( $\geq 50$ ) are shown. The scale indicates the number of substitutions per site. The tree was rooted with an outgroup consisting of the indicated *MHC-B* sequences. The three species humans, chimpanzees, and bonobos are represented by circles, triangles, and squares, respectively, in front of the sequence names. Different phylogenetic groups are defined on the right side of the tree, showing the two *MHC-A* lineages A2 and A3 and the two chimpanzee- and bonobo-specific groups within the A3 lineage *Patr-A1* and *Patr-A4*

undertook three measures to reduce the possibility of and check for potential chimeric sequences. First, because chimeric sequences arise mainly during late PCR cycles from incomplete PCR products of earlier cycles, we used an excessively long elongation time of 8 min (see the “DNA amplification” section) to ensure complete elongation of the products and reduce incomplete PCR products to a minimum (Kanagawa 2003; Lenz and Becker 2008). Second, we did every PCR of every locus and of every individual in duplicate, as it is unlikely that identical chimeric sequences would occur in both PCR replicates. Furthermore, any PCR artifacts should have a lower copy number than the target amplicon, which should result in a lower number of reads after the sequencing process. We did not recognize any discrepancies between the two replicates in terms of different sequences or number of reads, except those detailed above (see the “Sequence data processing” section). Lastly, the Long Amplicon Analysis tool of the Pacific Bioscience SMRT Analysis System used for generating the consensus sequences incorporates a chimera detection step using the UCHIME algorithm, which marks all potential chimera sequences and removes them from the further analysis (Edgar et al. 2011; Pacific Biosciences 2014). In summary, using those steps above, we are confident that the sequences reported here do not include chimera sequences.

### Data analysis

We used MEGA 6.06 to construct maximum likelihood trees including all different bonobo and chimpanzee coding region sequences from our data set (Tamura et al. 2013). We added all bonobo and chimpanzee sequences from the literature which were not represented by our own sequences. For human references we included sequences for the A, B, and C loci from Yoruba individuals from the HapMap project (The International HapMap C 2005). In total, the HapMap project sampled 30 triplets of Yoruba individuals (parents and one adult offspring). We took representative alleles from all parents (60 individuals) for the construction of phylogenetic trees. For the construction, the best substitution model was







◀ **Fig. 2** Molecular phylogenetic tree of MHC class I *B* sequences reconstructed using the maximum likelihood method based on the Tamura 3-parameter model. The model used a discrete gamma distribution allowing evolutionary rate differences among sites. The tree with the highest log likelihood (−7151.1817) is shown. Confidence in the branching patterns is indicated by bootstrap values and only relevant values ( $\geq 50$ ) are shown. The scale indicates the number of substitutions per site. The tree was rooted with an outgroup consisting of the indicated *MHC-A* sequences. The three species humans, chimpanzees, and bonobos are represented by circles, triangles, and squares, respectively, in front of the sequence names

selected using MEGA 6.06. We used a Tamura 3-parameter model with discrete gamma distribution with invariable sites (G + I) for all trees.

We calculated nucleotide diversity  $\Pi$  at each locus for 20 bonobos, after excluding Ulindi to have matched sample size to central chimpanzees and 20 central chimpanzees, including both alleles for each individual, with R (version 3.1.3) and the R package “pegas” (version 0.9). We estimated  $\Pi$  for introns and exons separately. Additionally, we included data from 20 published western chimpanzees (Adams et al. 2000) and estimated the nucleotide diversity. For comparison with humans, we reduced the Yoruba HapMap dataset to 30 individuals (every first individual of triplet) and chose randomly 20 individuals from this reduced dataset.

We conducted a permutation test (Adams and Anthony 1996) in R (version 3.1.3) for pairwise comparisons of diversity. We pooled all individuals of the two taxa being compared, resampled individuals (to account for non-independence of alleles in the same individual) independently of their taxa origin to compose two groups of the same size as the original two taxa, and calculated the nucleotide diversity of each new group and the absolute difference in nucleotide diversity between the two groups. The absolute difference in nucleotide diversity between the two groups of each permutation was compared to the absolute difference in nucleotide diversity of the original two taxon groups. The *p* value was calculated as the proportion of absolute differences of every permutation greater than or equal to the difference of the two original species (based on 10,000 permutations into which we included the original data as one permutation). We accounted for multiple testing and used a Bonferroni correction. After correction, *p* values below or equal to 0.002 were defined as being significant.

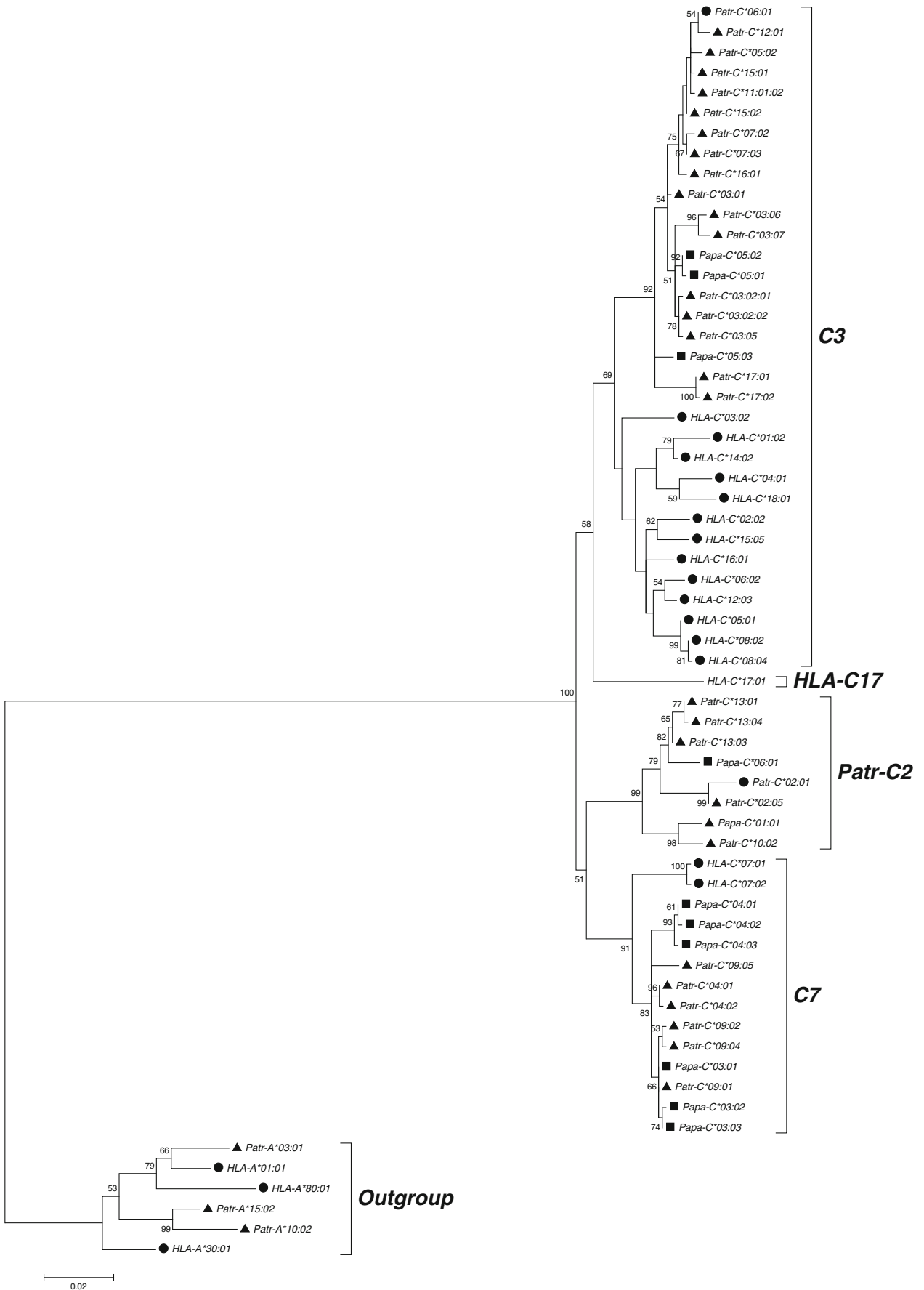
## Results

In our sample of 30 chimpanzees and 21 bonobos, we found 148 different full-length MHC class I alleles (Genbank accession numbers: KY613033–KY613180), including 42 and 21 novel coding region sequences in chimpanzees and bonobos,

respectively. From our chimpanzee samples, we obtained 30, 41, 29, and 11 different full-length alleles for the *Patr-A*, *-B*, *-C*, and *-A-like* loci, respectively (Table 1). We found 25, 39, 26, and 3 different coding region sequences for the four loci, respectively, including 8 *A*, 20 *B*, 13 *C*, and 1 *A-like* sequences, which have not been described previously (Table 1). Of these novel coding region sequences *Patr-B\*20:01:02*, *Patr-B\*23:01:02*, *Patr-C\*03:02:02*, and *Patr-C\*11:01:02* each had one synonymous substitution to known chimpanzee MHC coding region sequences. The translation of all coding region sequences from our chimpanzee sample resulted in 25, 39, 25, and 3 different and functional protein sequences for the *A*, *B*, *C*, and *A-like* loci. The results of the *A-like* screening were in concordance with the successful amplification and sequencing of a full-length *A-like* sequence for individuals carrying the *A-like*<sup>+</sup> haplotype.

Our bonobo sample revealed 13, 13, and 11 different full-length alleles for the *Papa-A*, *-B*, and *-C* loci, respectively (Table 2). We found no *A-like* allele or had a positive *A-like* screening result within our bonobo samples, which is consistent with the presence of the *A-like* locus only in chimpanzees. After removing introns from the full-length alleles, we obtained 11, 13, and 10 different coding region sequences for the three loci *A*, *B*, and *C*, respectively, and of these 11, 3, and 7 were not previously described. Each coding region sequence from our bonobo sample predicted a different and functional protein sequence. The number of different full-length alleles, coding region sequences, and protein sequences for each of three loci (*A*, *B*, and *C*) was higher in the chimpanzee sample compared to the bonobo sample, even if considering only the 20 central chimpanzees (Table 3).

To examine the evolutionary relationships of the alleles from the two *Pan* species, we reconstructed phylogenetic trees from the coding region sequences for the *A*, *B*, and *C* loci. We included chimpanzee and bonobo coding region sequences from the literature to depict all available species diversity and all known phylogenetic groups. We found that for the *A* locus, human alleles clustered into two lineages *A2* and *A3*, as previously described (Lawlor et al. 1990) (Fig. 1). As expected, chimpanzee and bonobo sequences clustered only within the human *A3* lineage of the *A* locus. Furthermore, chimpanzee alleles formed two groups of moderate bootstrap support previously termed *Patr-A1* and *Patr-A4* (Adams et al. 2000), with bonobo alleles found in both chimpanzee groups. Phylogenetic analysis of the alleles from the *B* locus produced a tree with notably low bootstrap support (Fig. 2). Nonetheless, it is notable that in contrast to bonobos, chimpanzee alleles were found in more phylogenetic clades, hinting at a greater diversity in this species at this locus. The phylogenetic tree for the *C* locus (Fig. 3) featured four different lineages named *C3*, *C7*, *HLA-C17*, and *Patr-C2* with intermediate bootstrap support



**Fig. 3** Molecular phylogenetic tree of MHC class I *C* sequences reconstructed using the maximum likelihood method based on the Tamura 3-parameter model. The model used a discrete gamma distribution allowing evolutionary rate differences among sites. The tree with the highest log likelihood (−4873.9977) is shown. Confidence in the branching patterns is indicated by bootstrap values and only relevant values ( $\geq 50$ ) are shown. The scale indicates the number of substitutions per site. The tree was rooted with an outgroup consisting of the indicated *MHC-A* sequences. The three species humans, chimpanzees, and bonobos are represented by circles, triangles, and squares, respectively, in front of the sequence names. Different phylogenetic groups are defined on the right side of the tree, showing the two lineages *C3* and *C7* and as well the species specific lineages *HLA-C17* and *Patr-C2*

(Adams et al. 2000). As was previously described, both chimpanzee and human alleles are found in the lineages *C3* and *C7*, whereas each species also has an additional lineage where sequences from the other species are not found (*Patr-C2* and *HLA-C17*) (Adams et al. 2000). Bonobo alleles clustered together with chimpanzee alleles within the *C3*, *C7*, and *Patr-C2* lineages. However, chimpanzees seem to have much more variety within the *C3* lineage, suggesting again relatively lower diversity in bonobos.

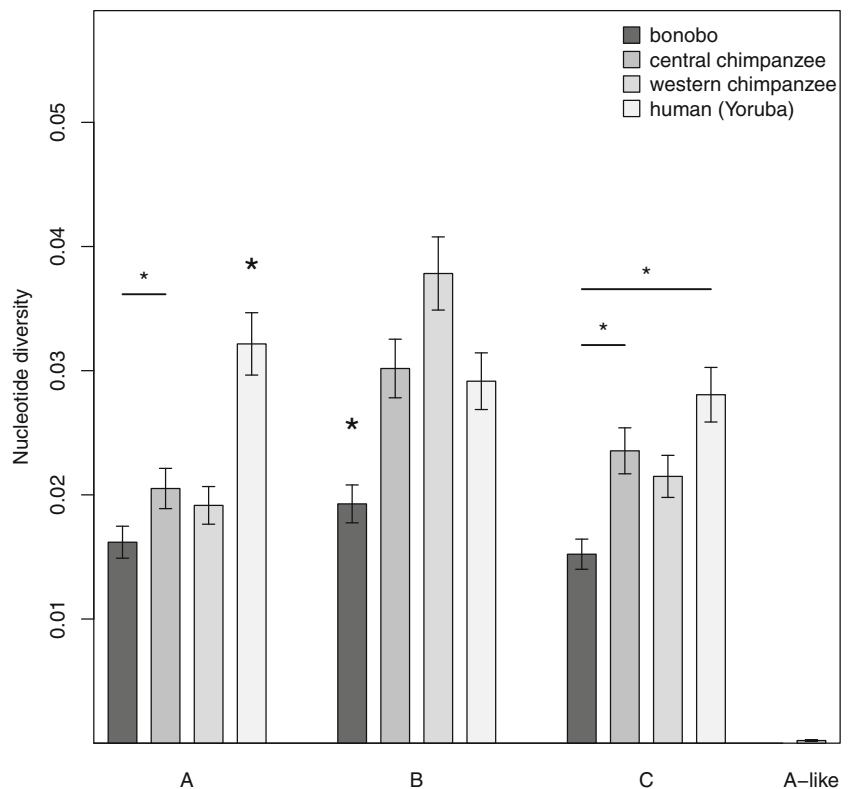
After seeing that bonobos have relatively fewer alleles than chimpanzees and that the bonobo alleles appear to be less widespread in the phylogenetic trees, we next estimated nucleotide diversities of all exons to quantitatively assess whether bonobos have a less diverse set of *A*, *B*, and *C* alleles than chimpanzees. Because chimpanzees as a species

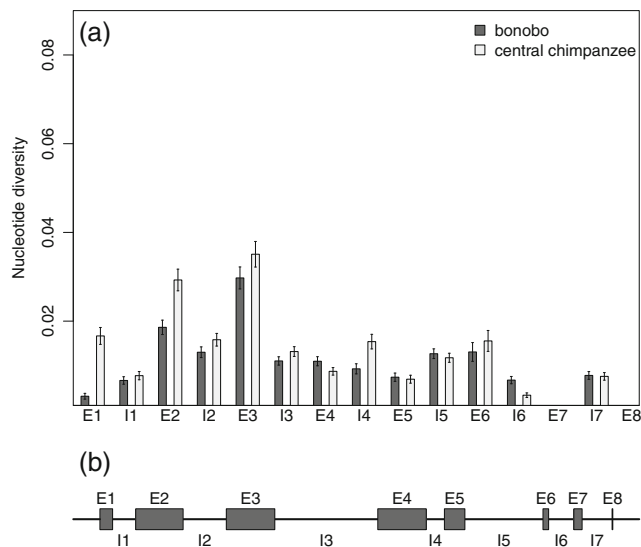
exhibit more genetic structure than bonobos (Becquet et al. 2007; Kawamoto et al. 2013; Prado-Martinez et al. 2013), we compared nucleotide diversity estimates from our bonobo samples with those from our central chimpanzee samples as well as with published western chimpanzees (Adams et al. 2000) and humans (Yoruba) (The International HapMap C 2005).

Our dataset of 20 bonobos exhibited significantly lower nucleotide diversity than our dataset of 20 central chimpanzees and humans at all three loci (Fig. 4 and supplement tables S3 and S4 for the exact values), while bonobos and western chimpanzees had no significant differences in diversity estimates at the *A* and *C* loci. At the *A* locus, our sample of humans had significantly higher nucleotide diversity than any of the *Pan* taxa. At the *B* locus, however, our sample of 20 bonobos had less nucleotide diversity than any of the other taxa. Compared to the other three loci *A*, *B*, and *C*, central chimpanzees had extremely low nucleotide diversity at the *A-like* locus.

To understand the patterns of relative nucleotide diversity across the MHC class I *A*, *B*, and *C* genes, we used our sequences of the entire genes from bonobos and central chimpanzees to estimate nucleotide diversity for each individual exon and intron (Figs. 5, 6, and 7). We found that there was no particular exon or intron within the genes which could explain the lower nucleotide diversity of bonobos. It was rather a general reduced diversity, e.g., lower diversity across all exons and introns. The comparison

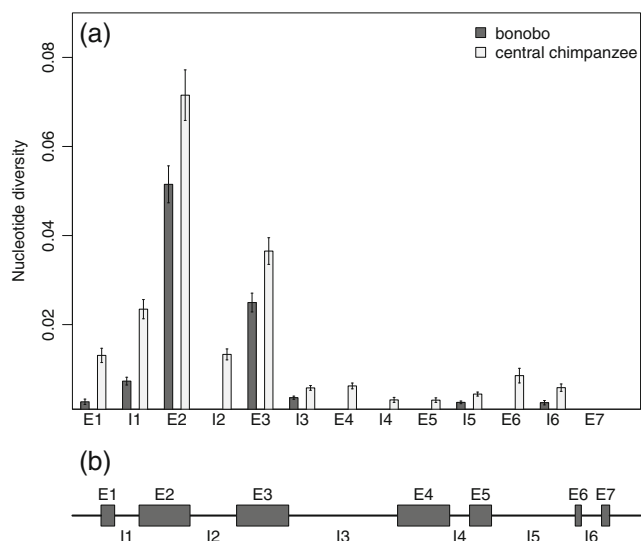
**Fig. 4** Estimated nucleotide diversity ( $\Pi$ ) of all exons for the different MHC loci *A*, *B*, *C*, and *A-like* DNA sequences from 20 unrelated bonobos and central chimpanzees reported here and 20 western chimpanzees and 20 humans (Yoruba) from the literature (Adams et al. 2000; The International HapMap C 2005). Error bars represent the standard error of the mean. Significant differences are indicated by stars above the individual bars, whereby *small stars* indicate significant differences between two bars linked by a line and *big stars* indicate significant differences to every other bar at this particular locus (permutation test, 10,000 resamplings, Bonferroni correction,  $p \leq 0.002$ )



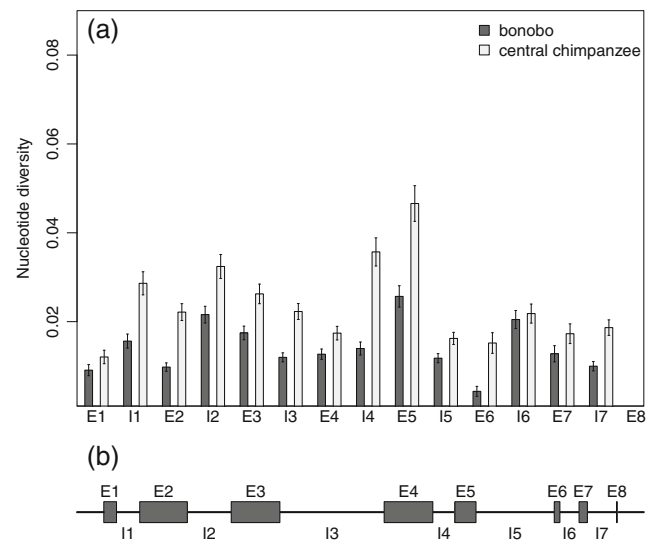


**Fig. 5** **a** Estimated nucleotide diversity ( $\pi$ ) at the class IA locus for each exon (E1–E8) and intron (I1–I7) for 20 bonobos and 20 central chimpanzees. *Error bars* represent standard error of the mean. **b** Organization of exons and introns of the *MHC-A* gene. The length of the boxes indicates the relative length of the particular exons and introns, where I3 is 583 bp long and E8 is 5 bp long

between the three different genes reveals that the distribution of nucleotide diversity of the different exons and introns was not similar between the three loci. For both species, the nucleotide diversity for the A locus (Fig. 5) was highest in exon 2 and exon 3, followed by lower nucleotide diversity for other regions of the gene. The nucleotide diversity of the B locus (Fig. 6) follows the same pattern, but with greater differences between the highly diverse exons 2



**Fig. 6** **a** Estimated nucleotide diversity ( $\pi$ ) of the class IB locus for each exon (E1–E7) and intron (I1–I6) for 20 bonobos and 20 central chimpanzees. *Error bars* represent standard error of the mean. **b** Organization of exons and introns of the *MHC-B* gene. The length of the boxes indicates the relative length of the particular exons and introns, where I3 is 574 bp long and E7 is 44 bp long



**Fig. 7** **a** Estimated nucleotide diversity ( $\pi$ ) of the class IC locus for each exon (E1–E8) and intron (I1–I7) for 20 bonobos and 20 central chimpanzees. *Error bars* represent standard error of the mean. **b** Organization of exons and introns of the *MHC-C* gene. The length of the boxes indicates the relative length of the particular exons and introns, where I3 is 588 bp long and E8 is 5 bp long

and 3 and the other regions of the gene. Nucleotide diversity of those other regions in bonobos was occasionally zero (intron 2, exon 4, intron 4, exon 6, exon 6, exon 7), which means that all bonobo B alleles within our sample set share the same sequence in those parts of the gene. The pattern of nucleotide diversity for the C locus differed from the other two loci (Fig. 7). At the C locus, the highest nucleotide diversity for both species was within exon 5. The nucleotide diversity of exon 2 and exon 3, which had the highest diversity values in the A and B loci, had relatively low nucleotide diversity compared to other parts of the gene.

## Discussion

This study of MHC class I diversity in the chimpanzee and bonobos identified 43, 54, 40, and 11 full-length alleles at the A, B, C, and A-like loci. Of these, 19, 23, 20, and 1 defined coding region sequences had not been previously described, thereby emphasizing the extraordinary polymorphism at this gene complex and the utility of investigating additional individuals to fully describe the whole range of MHC sequences in the genus *Pan*. In particular, the new sequences from our study doubles the number of known bonobo MHC class I alleles for each of the three loci A, B, and C.

The lower nucleotide diversity at the A locus of chimpanzees and bonobos compared to humans is not surprising due to the absence of one of the two lineages at the A locus in those two species (Adams et al. 2000; Lawlor et al. 1995; McAdam et al. 1995). Human A alleles cluster in the two ancient

lineages A2 and A3 where chimpanzee and bonobo A alleles are only related to alleles of the A3 lineage (Lawlor et al. 1990; McAdam et al. 1995). It has been suggested by de Groot et al. (2002) that the loss of the A2 lineage in chimpanzees was due to a selective sweep, perhaps caused by a widespread simian immunodeficiency virus (SIV) infection. Under this assumption, alleles from the A2 lineage offered insufficient protection against this infection and were therefore lost during evolution (Gleimer et al. 2011). The fact that bonobos also contain sequences only related to the A3 lineage indicates that if the selective sweep happened, it occurred before the separation of the two species. However, despite the loss of this specific A lineage, the *Patr-A-like* gene in chimpanzees has an overlapping peptide binding repertoire to human alleles from the A2 lineage, giving chimpanzees with this locus the ability to bind peptides specific to the A02 supertype (Gleimer et al. 2011). This is not possible in bonobos as they lack the *A-like* locus, and whether the immune response of bonobos is affected by this or if bonobos compensate this lineage loss in a different way requires further investigation.

In comparison to central chimpanzees, our results indicate that bonobos have a decreased MHC class I diversity shown by the lower number of different alleles, the phylogenetic trees and the decreased nucleotide diversity at all three loci. However, this lower diversity was larger at the B and C loci compared to A locus. One possible explanation for this reduced diversity in bonobos is that it simply reflects an overall reduced genetic diversity in bonobos. Sequencing of different non-coding autosomal regions distributed over different parts of the genome as well as sequencing of the whole genome for several individuals revealed reduced nucleotide diversity and a lower effective population size in bonobos compared to central chimpanzees (Fischer et al. 2006; Fischer et al. 2011; Prado-Martinez et al. 2013). However, western chimpanzees and humans also exhibit low diversity and effective population size (Li and Sadler 1991; Nam et al. 2017; Takahata et al. 1995; Zhao et al. 2000). If the MHC diversity reflects only the overall genetic diversity of the species, we would expect also a reduced MHC diversity in western chimpanzees and humans as compared to central chimpanzees, but our results show that published western chimpanzee and human sequences have equal or higher MHC nucleotide diversity than central chimpanzee sequences. This suggests that the low MHC diversity in bonobos is unlikely to reflect only the overall low genetic diversity of this species.

The critical role of the MHC in relation to immune functions rather suggests that the reduced bonobo MHC diversity may be explained by a selective process caused by pathogen exposure in the past history of bonobos since their split from chimpanzees. In such a scenario, exposure to certain pathogens would change the frequencies of certain MHC alleles within the population and may lead to a loss of certain MHC alleles and lineages. This loss could occur due to following

scenario: in a selective sweep, certain MHC alleles give protection against certain parasites or pathogens and are therefore selected within the population leading to an increase in frequency while disadvantageous MHC alleles at this locus are lost via purifying selection. Because we found the greatest difference in diversity at the B locus, this process of selection could have affected mainly this locus and due to genetic hitchhiking also the closely located C locus (Anzai et al. 2003; Shiina et al. 2006). Evidence for the role of MHC class I alleles in increased protection or susceptibility against diseases have been observed in humans. Human alleles *HLA-B\*57*, *B\*27*, or *B\*53* are associated with an increased protection against AIDS or malaria, respectively, whereas the allele *HLA-B\*35* was linked to rapid progression to AIDS after HIV infection (Altfeld et al. 2003, 2006; Carrington et al. 1999; Feeney et al. 2004; Gao et al. 2001; Hill et al. 1991; Hill et al. 1992; Kaslow et al. 1996; Migueles et al. 2000). It remains to be investigated which pathogens could be responsible for the loss in MHC diversity in bonobos. Potential candidates could be pathogens which also have a critical impact on health conditions in the two closely related species chimpanzees and humans like malaria or SIV/HIV. Although identical methods have been used to screen bonobo and chimpanzee samples from the wild, SIV has not been detected in bonobos thus far (Li et al. 2012; Sharp and Hahn 2011). The apparent absence of SIV in bonobos could be explained by the assumption that this species never experienced SIV or exhibited a superior defense against this type of virus similar as proposed for western chimpanzees (de Groot et al. 2002). Other studies showed that bonobos are confronted with malaria, ebola, monkeypox, and trypanosomiasis in their natural habitat and that pathogens have an impact on their distribution (Inogwabini and Leader-Williams 2012; Krief et al. 2010). One of those pathogens could also have been important in shaping the bonobo MHC diversity in their evolutionary past but this remains speculative at this point.

In concordance with our results, a recently published study by Wroblewski et al. (2017) investigated sequences of exons 2 and 3 of MHC class I B in representatives of several wild bonobo populations and found low diversity. Interestingly, the allele with the highest frequency (*Papa-B\*07:01*) in their study was also the most common bonobo B allele found in our study (Table 2).

Our study demonstrates the utility of typing MHC alleles using PCR and the PacBio RS II sequencing technology in set of wild-born individuals. The comparison between chimpanzees and bonobos revealed a lower MHC diversity in bonobos, representing perhaps the result of a selective process after the split from chimpanzees. Investigating also the effect of this low MHC diversity on the parasite and pathogen resistance in this species could contribute to our understanding of the relationship between MHC diversity and immunity.



**Acknowledgements** Open access funding provided by Max Planck Society. We thank Christoph König from Pacific Biosciences, Aida Andrés, Janet Kelso, and Montgomery Slatkin for the helpful discussion. We are thankful for assistance by the MPI-EVA sequencing group and thank Birgit Nickel and Kathrin Köhler for the assistance with the samples and Roger Mundry for the statistical advice. We kindly thank Centre de Conservation pour Chimanzés, Chimfunshi Wildlife Orphanage Trust, Jeunes Animaux Confisqués Au Katanga, Stichting AAP, Furuviksparken, Lola ya Bonobo sanctuary and Tchimpounga sanctuary for generously providing samples.

#### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** The sample collection and the sample handling were performed in accordance with all relevant regulations and guidelines and where necessary CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) permissions for sample shipment were obtained.

**Funding** This study received financial support by the Max Planck Society.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Adams DC, Anthony CD (1996) Using randomization techniques to analyse behavioural data. *Anim Behav* 51:733–738. doi:10.1006/anbe.1996.0077
- Adams EJ, Cooper S, Parham P (2001) A novel, nonclassical MHC class I molecule specific to the common chimpanzee. *J Immunol* 167:3858–3869
- Adams EJ, Cooper S, Thomson G, Parham P (2000) Common chimpanzees have greater diversity than humans at two of the three highly polymorphic MHC class I genes. *Immunogenetics* 51:410–424. doi:10.1007/s002510050639
- Adams EJ, Parham P (2001) Species-specific evolution of MHC class I genes in the higher primates. *Immunol Rev* 183:41–64
- Altfeld M et al (2003) Influence of HLA-B\*57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS* 17:2581–2591. doi:10.1097/01.aids.0000096870.36052.b6
- Altfeld M et al (2006) HLA alleles associated with delayed progression to AIDS contribute strongly to the initial CD8(+) T cell response against HIV-1. *PLoS Med* 3:e403. doi:10.1371/journal.pmed.0030403
- Anzai T et al (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proc Natl Acad Sci* 100:7708–7713. doi:10.1073/pnas.1230533100
- Becquet C, Patterson N, Stone AC, Przeworski M, Reich D (2007) Genetic structure of chimpanzee populations. *PLoS Genet* 3:e66. doi:10.1371/journal.pgen.0030066
- Bontrop RE (2006) Comparative genetics of MHC polymorphisms in different primate species: duplications and deletions. *Hum Immunol* 67:388–397. doi:10.1016/j.humimm.2006.03.007
- Cagan A et al (2016) Natural selection in the great apes. *Mol Biol Evol* 33(12):3268–3283. doi:10.1093/molbev/msw215
- Carrington M et al (1999) HLA and HIV-1: heterozygote advantage and B\*35-Cw\*04 disadvantage. *Science* 283:1748–1752. doi:10.1126/science.283.5408.1748
- Caswell JL, Mallick S, Richter DJ, Neubauer J, Schirmer C, Gnerre S, Reich D (2008) Analysis of chimpanzee history based on genome sequence alignments. *PLoS Genet* 4:e1000057. doi:10.1371/journal.pgen.1000057
- Codner GF, Stear MJ, Reeve R, Matthews L, Ellis SA (2012) Selective forces shaping diversity in the class I region of the major histocompatibility complex in dairy cattle. *Anim Genet* 43:239–249. doi:10.1111/j.1365-2052.2011.02239.x
- Cooper S, Adams EJ, Wells RS, Walker CM, Parham P (1998) A major histocompatibility complex class I allele shared by two species of chimpanzee. *Immunogenetics* 47:212–217
- de Groot NG, Bontrop RE (2013) The HIV-1 pandemic: does the selective sweep in chimpanzees mirror humankind's future? *Retrovirology* 10:1–15. doi:10.1186/1742-4690-10-53
- de Groot NG et al (2008) Pinpointing a selective sweep to the chimpanzee MHC class I region by comparative genomics. *Mol Ecol* 17:2074–2088. doi:10.1111/j.1365-294X.2008.03716.x
- de Groot NG et al (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. *Proc Natl Acad Sci U S A* 99:11748–11753. doi:10.1073/pnas.182420799
- de Manuel M et al (2016) Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* 354:477–481. doi:10.1126/science.aag2602
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200. doi:10.1093/bioinformatics/btr381
- Feeney ME et al (2004) Immune escape precedes breakthrough human immunodeficiency virus type 1 viremia and broadening of the cytotoxic T-lymphocyte response in an HLA-B27-positive long-term nonprogressing child. *J Virol* 78:8927–8930. doi:10.1128/JVI.78.16.8927-8930.2004
- Fischer A, Pollack J, Thalmann O, Nickel B, Paabo S (2006) Demographic history and genetic differentiation in apes. *Current biology: CB* 16:1133–1138. doi:10.1016/j.cub.2006.04.033
- Fischer A et al (2011) Bonobos fall within the genomic variation of chimpanzees. *PLoS One* 6:e21605. doi:10.1371/journal.pone.0021605
- Fischer A, Wiebe V, Paabo S, Przeworski M (2004) Evidence for a complex demographic history of chimpanzees. *Mol Biol Evol* 21:799–808. doi:10.1093/molbev/msh083
- Gao X et al (2001) Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N Engl J Med* 344:1668–1675. doi:10.1056/nejm20010513442203
- Geraghty DE, Daza R, Williams LM, Vu Q, Ishitani A (2002) Genetics of the immune response: identifying immune variation within the MHC and throughout the genome. *Immunol Rev* 190:69–85
- Gleimer M et al (2011) Although divergent in residues of the peptide binding site, conserved chimpanzee Patr-AL and polymorphic human HLA-A\*02 have overlapping peptide-binding repertoires. *J Immunol* 186:1575–1588. doi:10.4049/jimmunol.1002990
- Groot NG 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. doi:10.1007/s00251-012-0617-1



- Gyllenstein UB, Erlich HA (1989) Ancient roots for polymorphism at the HLA-DQ alpha locus in primates. *Proc Natl Acad Sci U S A* 86: 9986–9990
- Hans JB, Bergl RA, Vigilant L (2017) Gorilla MHC class I gene and sequence variation in a comparative context. *Immunogenetics*:1–21. doi:10.1007/s00251-017-0974-x
- Heuvels FV et al (2007) Genetic diversity and phylogeographic clustering of SIVcpzPtt in wild chimpanzees in Cameroon. *Virology* 368:155–171
- Hey J (2010) The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol Biol Evol* 27:921–933. doi:10.1093/molbev/msp298
- Hill AVS et al (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature* 352:595–600
- Hill AVS et al (1992) Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature* 360:434–439
- Inogwabini B-I, Leader-Williams N (2012) Effects of epidemic diseases on the distribution of bonobos. *PLoS One* 7:e51112. doi:10.1371/journal.pone.0051112
- Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng* 96:317–323. doi:10.1016/S1389-1723(03)90130-7
- Kaslow RA et al (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med* 2:405–411
- Kawamoto Y et al (2013) Genetic structure of wild bonobo populations: diversity of mitochondrial DNA and geographical distribution. *PLoS One* 8:e59660. doi:10.1371/journal.pone.0059660
- Keele BF et al (2006) Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* 313:523–526. doi:10.1126/science.1126531
- Kelley J, Walter L, Trowsdale J (2005) Comparative genomics of major histocompatibility complexes. *Immunogenetics* 56:683–695. doi:10.1007/s00251-004-0717-7
- Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum Immunol* 19:155–162. doi:10.1016/0198-8859(87)90066-8
- Klein J et al (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* 31:217–219. doi:10.1007/BF00204890
- Klein J, Sato A, Nagl S, O'HUig xed n C (1998) Molecular trans-species polymorphism. *Annu Rev Ecol Syst* 29 1-C1
- Krief S et al (2010) On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from bonobos. *PLoS Pathog* 6:e1000765. doi:10.1371/journal.ppat.1000765
- Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol Rev* 190:95–122. doi:10.1034/j.1600-065X.2002.19008.x
- Langergraber KE et al (2012) Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc Natl Acad Sci* 109:15716–15721. doi:10.1073/pnas.1211740109
- Lawlor DA, Edelson BT, Parham P (1995) Mhc-A locus molecules in pygmy chimpanzees: conservation of peptide pockets. *Immunogenetics* 42:291–295
- Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* 335:268–271
- Lawlor DA, Warren E, Ward FE, Parham P (1990) Comparison of class I MHC alleles in humans and apes. *Immunol Rev* 113:147–185. doi:10.1111/j.1600-065X.1990.tb00040.x
- 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. doi:10.1016/j.gene.2008.09.013
- Li H et al (2009) The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25(25):2078–2079
- Li WH, Sadler LA (1991) Low nucleotide diversity in man. *Genetics* 129: 513–523
- Li Y et al (2012) Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir. *J Virol* 86:10776–10791. doi:10.1128/jvi.01498-12
- Martinez-Laso J, Gomez-Casado E, Armaiz-Villena A (2006) Description of seven new non-human primate MHC-B alleles. *Tissue Antigens* 67:85–88. doi:10.1111/j.1399-0039.2005.00496.x
- Mayer WE, Jonker M, Klein D, Ivanyi P, van Seventer G, Klein J (1988) Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO J* 7:2765–2774
- Mayor NP et al (2015) HLA typing for the next generation. *PLoS One* 10:e0127153. doi:10.1371/journal.pone.0127153
- McAdam SN, Boyson JE, Liu X, Garber TL, Hughes AL, Bontrop RE, Watkins DI (1994) A uniquely high level of recombination at the HLA-B locus. *Proc Natl Acad Sci U S A* 91:5893–5897
- McAdam SN, Boyson JE, Liu X, Garber TL, Hughes AL, Bontrop RE, Watkins DI (1995) Chimpanzee MHC class I A locus alleles are related to only one of the six families of human A locus alleles. *J Immunol* 154:6421–6429
- Miguel SA et al (2000) HLA B 5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A* 97:2709–2714. doi:10.1073/pnas.050567397
- Nam K et al (2017) Evidence that the rate of strong selective sweeps increases with population size in the great apes. *Proc Natl Acad Sci* 114:1613–1618. doi:10.1073/pnas.1605660114
- Pacific Biosciences (2014) ConsensusTools v2.3.0 Documentation. <https://github.com/PacificBiosciences/SMRT-Analysis/wiki/ConsensusTools-v2.3.0-Documentation>. Accessed 28.03.2017
- Parham P (2005) Influence of KIR diversity on human immunity. In: Gupta S, Paul WE, Steinman R (eds) *Mechanisms of lymphocyte activation and immune regulation X: innate immunity*. Springer US, Boston, MA, pp 47–50. doi:10.1007/0-387-24180-9\_6
- Prado-Martinez J et al (2013) Great ape genetic diversity and population history. *Nature* 499:471–475. doi:10.1038/nature12228
- Prüfer K et al (2012) The bonobo genome compared with the chimpanzee and human genomes *Nature* 486:527–531 doi:<http://www.nature.com/nature/journal/v486/n7404/abs/nature11128.html#supplementary-information>
- Rock KL, Reits E, Neefjes J (2016) Present yourself! By MHC class I and MHC class II. *Molecules Trends in Immunology* 37:724–737. doi:10.1016/j.it.2016.08.010
- Rudicell RS et al (2010) Impact of simian immunodeficiency virus infection on chimpanzee population dynamics. *PLoS Pathog* 6:e1001116. doi:10.1371/journal.ppat.1001116
- Santiago ML et al (2003) Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). *J Virol* 77:7545–7562
- Santiago ML et al (2002) SIVcpz in wild chimpanzees. *Science* 295:465. doi:10.1126/science.295.5554.465
- Scally A et al (2012) Insights into hominid evolution from the gorilla genome sequence. *Nature* 483:169–175. doi:10.1038/nature10842
- Sharp PM, Hahn BH (2011) Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine* 1:a006841. doi:10.1101/cshperspect.a006841
- Shiina T et al (2006) Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in

- humans via hitchhiking diversity. *Genetics* 173:1555–1570. doi:10.1534/genetics.106.057034
- Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool* 2:16
- Takahata N, Satta Y, Klein J (1995) Divergence time and population size in the lineage leading to modern humans. *Theor Popul Biol* 48:198–221. doi:10.1006/tpbi.1995.1026
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. doi:10.1093/molbev/mst197
- The International HapMap C (2005) A haplotype map of the human genome *Nature* 437:1299–1320 doi:[http://www.nature.com/nature/journal/v437/n7063/supinfo/nature04226\\_S1.html](http://www.nature.com/nature/journal/v437/n7063/supinfo/nature04226_S1.html)
- Trowsdale J (2001) Genetic and functional relationships between MHC and NK receptor. *Genes Immunity* 15:363–374. doi:10.1016/S1074-7613(01)00197-2
- Trowsdale J, Parham P (2004) Mini-review: defense strategies and immunity-related genes. *Eur J Immunol* 34:7–17. doi:10.1002/eji.200324693
- Wegmann D, Excoffier L (2010) Bayesian inference of the demographic history of chimpanzees. *Mol Biol Evol* 27:1425–1435. doi:10.1093/molbev/msq028
- Westbrook CJ et al (2015) No assembly required: full-length MHC class I allele discovery by PacBio circular consensus sequencing. *Hum Immunol* 76:891–896. doi:10.1016/j.humimm.2015.03.022
- Worobey M et al (2004) Origin of AIDS: contaminated polio vaccine theory refuted. *Nature* 428:820. doi:10.1038/428820a
- Wroblewski EE et al (2015) Signature patterns of MHC diversity in three Gombe communities of wild chimpanzees reflect fitness in reproduction and immune defense against SIVcpz. *PLoS Biol* 13:e1002144. doi:10.1371/journal.pbio.1002144
- Wroblewski EE et al (2017) Bonobos maintain immune system diversity with three functional types of MHC-B. *J Immunol* 198(9):3480–3493. doi:10.4049/jimmunol.1601955
- Zhao Z et al (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. *Proc Natl Acad Sci U S A* 97:11354–11358