

# Rapid identification of major histocompatibility complex class I haplotypes using deep sequencing in an endangered Old World monkey

Noah D. Simons<sup>1</sup> · Maria Jose Ruiz-Lopez<sup>2</sup> · Colin A. Chapman<sup>3</sup> · Tony L. Goldberg<sup>4,6</sup> · Julie A. Karl<sup>6</sup> · Roger W. Wiseman<sup>6</sup> · Patrick S. Bohn<sup>6</sup> · David H. O'Connor<sup>5,6</sup> · Nelson Ting<sup>1,2</sup>

Received: 25 June 2015 / Accepted: 7 January 2016 / Published online: 22 January 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** Immunogenetic data from wild primate populations have been difficult to obtain, due to logistic and methodological constraints. We applied a well-characterized deep sequencing method for MHC I typing, developed for macaques, to a population of wild red colobus to assess the feasibility of identifying MHC I-A/B haplotypes. Ten individuals produced sufficient data from blood and tissue samples to assign haplotypes. Eighty-two sequences were classified as red colobus MHC I alleles distributed across six MHC I loci. Individuals averaged ~13 k reads across six MHC I loci, with 83 % of all alleles representing either MHC I-A or MHC I-B loci. This study not only represents an important advance in the identification and genotyping of MHC in the endangered red colobus but also shows the potential for using this approach in other endangered wild primates.

**Keywords** Major histocompatibility complex · Next generation sequencing · Primates · Immunogenetics · Red colobus

Red colobus monkeys (genus *Procolobus*; Grubb et al. 2003; Oates and Davies 1994) are among the most threatened African primates due to hunting and habitat modification/loss (Struhsaker 2005; Ting 2008). While habitat modification can drive disease susceptibility (Brearley et al. 2013), there remains a lack of immunogenetic data for red colobus, which precludes the advancement of disease association studies. Few wild, endangered primates have been characterized for immune genes, including major histocompatibility complex (MHC). MHC genotyping of wild primates remains challenging principally because of logistic constraints and because wild primates such as red colobus are non-model organisms (Sommer et al. 2013).

We tested a well-characterized MHC deep sequencing pipeline, developed for macaques (Karl et al. 2014), in wild red colobus. We genotyped a highly polymorphic 156 base pair region of MHC class I exon 2. Nine blood samples from Ugandan red colobus (*Procolobus rufomitratus tephrosceles*) from Kibale National Park were collected in EDTA in 2006 and 2010. One tissue sample from Temminck's red colobus (*Procolobus badius temminckii*) was collected from Abuko Nature Reserve in the Gambia (Fig. 1). Genomic DNA was extracted using QIAamp DNA Mini Kits (Qiagen) following the manufacturers protocol and normalized to 10 ng/μl. Primers used were (SBT195-F: GCCTCGCTCTGGTTG-TAG-TAG, SBT195-R: GGGCTACGTGGACGACAC; Karl et al. 2014). PCR amplification in 25 μl reactions contained 10 ng of DNA, 12.5 nM of each primer, 100 nM of

**Electronic supplementary material** The online version of this article (doi:10.1007/s12686-016-0516-0) contains supplementary material, which is available to authorized users.

✉ Nelson Ting  
nting@uoregon.edu

<sup>1</sup> Department of Anthropology, University of Oregon, Eugene, OR 97405, USA

<sup>2</sup> Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97405, USA

<sup>3</sup> Department of Anthropology, McGill School of Environment, Montreal, QC H3A 2T7, Canada

<sup>4</sup> Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI 53725, USA

<sup>5</sup> Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI 53715, USA

<sup>6</sup> Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI 53715, USA

**Fig. 1** Haplotypes for MHC I-A/B, demographic information, sample type, and number of reads for Ugandan and Temminick's red colobus

ID	Species	Haplotypes		Biomaterial	Location	Sex	# Reads Identified
		MHC I-A	MHC I-B				
1	<i>P. r. tephrosceles</i>	A001a	B001	Whole Blood	Uganda	M	9192
		A001a	B002				
#	<i>P. r. tephrosceles</i>	A001a	B001	Whole Blood	Uganda	M	7566
		A001a	B004				
3	<i>P. r. tephrosceles</i>	A001a	B003	Whole Blood	Uganda	M	9201
		A001a	B003				
4	<i>P. r. tephrosceles</i>	A001a	B001	Whole Blood	Uganda	M	34611
		A001a	B003				
5	<i>P. r. tephrosceles</i>	A001a	B001	Whole Blood	Uganda	M	9241
		A001a	B003				
6	<i>P. r. tephrosceles</i>	A001a	B004	Whole Blood	Uganda	F	10359
		A001a	B003				
7	<i>P. r. tephrosceles</i>	A001a	B001	Whole Blood	Uganda	M	7332
		A002	B005				
8	<i>P. r. tephrosceles</i>	A002	B002	Whole Blood	Uganda	M	10280
		A002	B004				
9	<i>P. r. tephrosceles</i>	A001b	B002	Whole Blood	Uganda	M	10534
		A003	B006				
10	<i>P. badius</i>	A004	B007	Liver	Gambia	M	13251
		A005	B007				

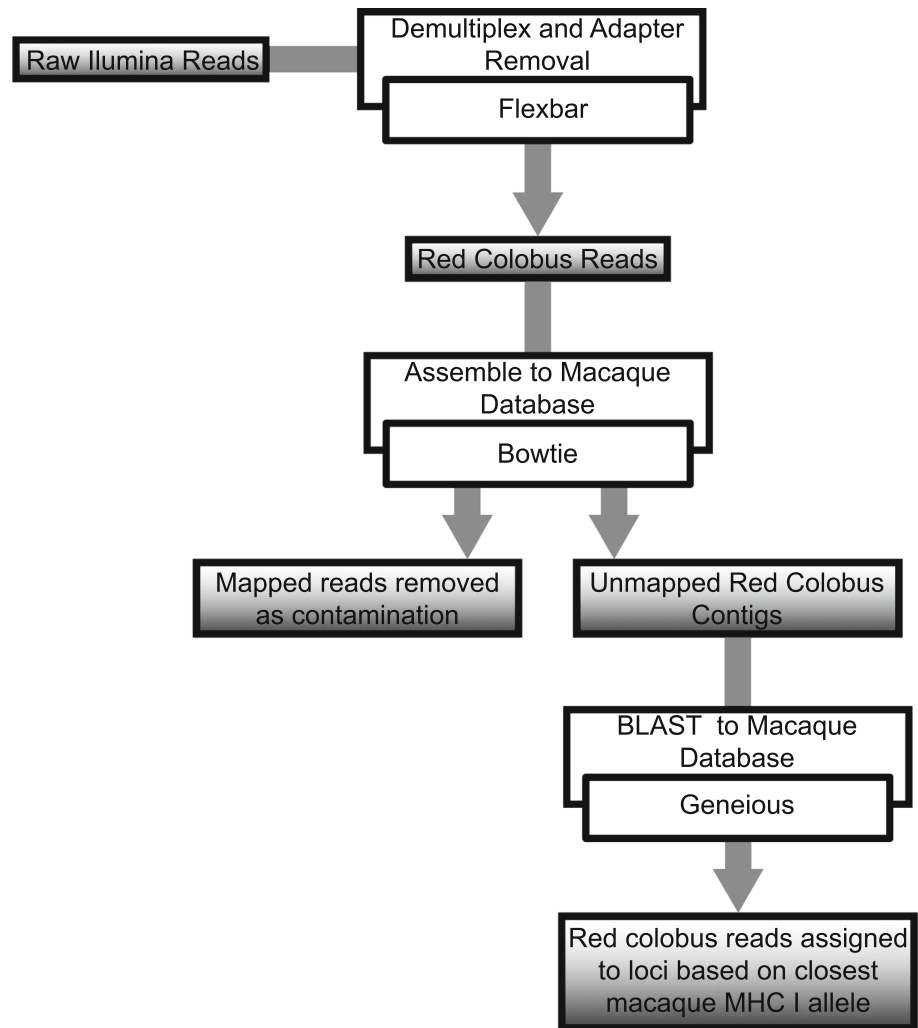
Fluidigm barcoded Illumina adapter primers, 12.5  $\mu$ l 2 $\times$  Phusion High Fidelity PCR Master Mix (New England BioLabs), and 7.5  $\mu$ l of water. PCR conditions included an initial denaturation for 3 min (98  $^{\circ}$ C), followed by 25 cycles of 5-s denaturation (98  $^{\circ}$ C), 10-s annealing (60  $^{\circ}$ C) and 20-s extension (72  $^{\circ}$ C), followed by a final 5-min extension (72  $^{\circ}$ C). PCR products were purified using an Agilent AMPure XP (Beckman Coulter) bead cleanup. Purified products were quantified using PicoGreen chemistry (Thermo Fisher Scientific). We pooled together equimolar concentrations of each sample and carried out two additional AMPure XP cleaning steps to remove remaining dimers. The library was quantified using the Qubit<sup>TM</sup> High-Sensitivity dsDNA Assay (Invitrogen) and sequenced as part of a single, paired-end run on an Illumina MiSeq using the 500-cycle MiSeq Reagent Kit v2 (Illumina, San Diego, CA, USA).

Raw Illumina reads were demultiplexed and primer and adapter sequences removed using Flexbar (Dodt et al. 2012). Primers were trimmed in left trim mode with a minimum overlap equal to the length of the primer with no unknown bases allowed. Sequences were then trimmed from the right to a maximum length of 156 base pairs. Paired-end reads were merged using FLASH under default settings (Magoc and Salzberg 2011). Merged reads were

mapped against a curated database of known macaque MHC class I alleles from the Wisconsin National Primate Research Center (WNPRC) using Bowtie, allowing zero mismatches (Langmead et al. 2009). Mapped reads were removed as contamination (<0.0023 %), and unmapped reads were retained as red colobus sequences. Sequences were searched against the WNPRC macaque database using BLAST in Geneious Pro v. 6.1.2 (Biomatters) allowing a 1 % mismatch (Fig. 2). Sequences were assigned to a locus based on closest match in the macaque database. A minimum threshold of 100 $\times$  coverage was used, and we retained only sequences that matched the expected sequence size (156)  $\pm$  3 base pairs in order to remove sequences in which a reading frame shift had occurred. We translated the sequences and searched for stop codons using the second base position as the start of the reading frame, and sequences containing a stop codon were designated as pseudogenes. Alleles with a Mean Per Amplicon Frequency (MPAF) <0.01 were visually checked for chimeras and sequencing errors (no chimeras identified, one allele removed as sequencing error) following Zagalska-Neubauer et al. (2010).

In total, 82 alleles (including two pseudogenes) were identified that mapped closely to six macaque (*Macaca mulatta*) MHC I loci (doi:10.5061/dryad.6r40g).

**Fig. 2** Bioinformatic pipeline for novel MHC I haplotype identification



Haplotypes were determined based on patterns of allele sharing across individuals (see Supplemental Material S1). The majority of haplotypes represented MHC I-A (6 haplotypes, 16 alleles) and MHC I-B (7 haplotypes, 41 alleles), suggesting that, like macaques, MHC I has undergone extensive duplication events in red colobus. Number of reads per individual ranged from 7332 to 34,611 (average = 12,156). Six Ugandan red colobus were homozygous for the A001a, one was homozygous for A002, and three were heterozygotes. A005 and A006 were restricted to Temminick's red colobus. One Ugandan red colobus was homozygous for B003, and eight were heterozygous. B007 was restricted to Temminick's red colobus.

We have demonstrated that a protocol for genotyping MHC I in macaques can be applied successfully to a non-model primate, the red colobus. This finding shows that rapid and accurate MHC genotyping in endangered primates is currently feasible, given the availability of suitable biomaterials. Because our primers were designed for a distantly related Old World monkey, we predict that this

approach will be successful across the family Cercopithecidae. Applying our method more broadly should facilitate studies of the genetic determinants of disease susceptibility as well as assessments of immunogenetic diversity in isolated primate populations.

## References

- Brearley G, Rhodes J, Bradley A, Baxter G, Seabrook L, Lunney D, Liu Y, McAlpine C (2013) Wildlife disease prevalence in human-modified landscapes. *Biol Rev* 88:427–442
- Dodt M, Roehr JT, Ahmed R, Dieterich C (2012) FLEXBAR-flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* 1(3):895–905
- Grubb P, Butynski TM, Oates JF, Bearder SK, Disotell TR, Groves CP, Struhsaker TT (2003) Assessment of the diversity of African primates. *Int J Primatol* 24:1301–1357
- Karl JA, Heimbruch KE, Vriezen CE, Mironczuk CJ, Dudley DM, Wiseman RW, O'Connor DH (2014) Survey of major histocompatibility complex class II diversity in pig-tailed macaques. *Immunogenetics* 66(11):613–623

- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25
- Magoc T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21):2957–2963
- Oates J, Davies G (1994) *Colobine monkeys: their ecology, behaviour and evolution*. Cambridge University Press, Cambridge, p 415
- 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 Genom* 14(1):542
- Struhsaker TT (2005) Conservation of red colobus and their habitats. *Int J Primatol* 26:525–538
- Ting N (2008) *Molecular systematics of red colobus monkeys (Procolobus [Piliocolobus]): understanding the evolution of an endangered primate*. Ph.D. Dissertation. City University of New York Graduate Center
- Zagalska-Neubauer M, Babik W, Stuglik M, Gustafsson L, Cichoń M, Radwan J (2010) 454 sequencing reveals extreme complexity of the class II major histocompatibility complex in the collared flycatcher. *BMC Evol Biol* 10(1):395