Received: 27 July 2016 | Revised: 10 June 2017 | Accepted: 13 June 2017

DOI: 10.1111/mec.14221

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Cis-regulatory evolution in a wild primate: Infection-associated genetic variation drives differential expression of *MHC-DQA1* in vitro

Noah D. Simons[1] | Geeta N. Eick[1] | Maria J. Ruiz-Lopez[2] | Patrick A. Omeja[3] | Colin A. Chapman[3,4] | Tony L. Goldberg[5,6] | Nelson Ting[1,2] | Kirstin N. Sterner[1]

[1]Department of Anthropology, University of Oregon, Eugene, OR, USA

[2]Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA

[3]Makerere University Biological Field Station, Fort Portal, Uganda

[4]Department of Anthropology and McGill School of Environment, McGill University, Montreal, QC, Canada

[5]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI, USA

[6]Global Health Institute, University of Wisconsin-Madison, Madison, WI, USA

Correspondence
Kirstin N. Sterner, Department of Anthropology, University of Oregon, Eugene, OR, USA.
Email: ksterner@uoregon.edu

Funding information
NIH, Grant/Award Number: TW009237; NSF, Grant/Award Number: BCS-1540459; National Geographic Society; NSERC; University of Oregon

Abstract

Few studies have combined genetic association analyses with functional characterization of infection-associated SNPs in natural populations of nonhuman primates. Here, we investigate the relationship between host genetic variation, parasitism and natural selection in a population of red colobus (*Procolobus rufomitratus tephrosceles*) in Kibale National Park, Uganda. We collected parasitological, cellular and genomic data to test the following hypotheses: (i) *MHC-DQA1* regulatory genetic variation is associated with control of whipworm (*Trichuris*) infection in a natural population of red colobus; (ii) infection-associated SNPs are functional in driving differential gene expression in vitro; and (iii) balancing selection has shaped patterns of variation in the *MHC-DQA1* promoter. We identified two SNPs in the *MHC-DQA1* promoter, both in transcription factor binding sites, and both of which are associated with decreased control of *Trichuris* infection. We characterized the function of both SNPs by testing differences in gene expression between the two alleles of each SNP in two mammalian cell lines. Alleles of one of the SNPs drove differential gene expression in both cell lines, while the other SNP drove differences in expression in one of the cell lines. Additionally, we found evidence of balancing selection acting on the *MHC-DQA1* promoter, including extensive trans-species polymorphisms between red colobus and other primates, and an excess of intermediate-frequency alleles relative to genome-wide, coding and noncoding RADseq data. Our data suggest that balancing selection provides adaptive regulatory flexibility that outweighs the consequences of increased parasite infection intensity in heterozygotes.

KEYWORDS
ecoimmunology, immunogenetics, major histocompatibility complex, red colobus, *Trichuris*

1 | INTRODUCTION

Understanding how regulatory genetic variation influences complex disease phenotypes is an important aim of biology (Albert & Kruglyak, 2015). *Cis*-regulatory genetic variation (i.e., variation in noncoding DNA that may influence the transcription of proximal genes), particularly that found in transcription factor binding sites, plays an important evolutionary role in controlling patterns of gene expression (Chen, van Nimwegen, Rajewsky, & Siegal, 2010; Hahn, 2007; Wittkopp & Kalay, 2012; Wray, 2007). Although growing evidence indicates that *cis*-regulatory variants influence immune function and may explain variation in the control of disease (Guo et al., 2015; Miura et al., 2008), a better understanding of how regulatory regions contribute to an individual's ability to control infections and how

variation in these regions shapes variation observed in infection intensity between individuals (e.g., "elite controllers" or long-term nonprogressors of HIV infection; Saag & Deeks, 2010) is needed.

Here, we investigate the relationship between host genetics and control of parasite infection in a natural population of nonhuman primates. Specifically, we identified *cis*-regulatory variation in the 5′ core promoter of *MHC-DQA1* in the Ugandan red colobus (*Procolobus rufomitratus tephrosceles*) that is associated with control of *Trichuris* infection and we test whether these infection-associated SNPs influence gene expression in vitro. We also examined natural selection acting on the *MHC-DQA1* promoter region, as a previous study (Loisel, Rockman, Wray, Altmann, & Alberts, 2006) found this region to be under long-term balancing selection across closely related primate taxa.

1.1 | Control of infection in natural populations: red colobus

Well-studied natural populations of nonhuman primates are valuable for understanding how host biology and environment influence disease dynamics, and they provide a powerful model for investigating human diseases (Nunn, 2012; Tung et al., 2009; Wolfe et al., 1998). Patterns of infectious disease represent complex and often prolonged interactions between a host and pathogen (Borghans, Beltman, & De Boer, 2004; Van Valen, 1973), with the host response being shaped by many factors, including ecology, demography, behaviour and genetics (Hart, 1990; Minchella, 1985; Zuk & Stoehr, 2002).

The Ugandan red colobus is an Old World monkey (family Cercopithecidae) that has been the subject of ecological studies at Kibale National Park, hereafter Kibale, in Uganda for over 46 years (Chapman, Struhsaker, & Lambert, 2005; Struhsaker, 2010). Despite being highly threatened (Oates & Ting, 2015; Struhsaker, 2005; Ting, 2008a), it is one of very few wild nonhuman primates where data are available on host behavioural ecology, host genetics and infectious disease. The Ugandan red colobus has consequently emerged as an important model for understanding disease transmission and prevalence (Bailey et al., 2014; Chapman et al., 2011; Ghai, Chapman, Omeja, Davies, & Goldberg, 2014; Ghai, Fugere, Chapman, Goldberg, & Davies, 2015; Goldberg et al., 2008; Lauck et al., 2011; Sibley et al., 2014; Thurber et al., 2013).

Over the past 11 years, detailed and ongoing studies have been conducted on the prevalence of gastrointestinal parasites in the Ugandan red colobus (e.g., Chapman, Speirs, Gillespie, Holland, & Austad, 2006; Chapman, Wasserman, et al., 2006; Gillespie, Chapman, & Greiner, 2005; Gillespie, Greiner, & Chapman, 2005). In general, gastrointestinal parasitism is known to influence reproductive fitness by inducing sickness behaviour (Ghai et al., 2015; Hart, 1988, 1990), modulating the immune system (MacIntosh, 2014), and inducing an inflammatory immune response (MacIntosh, 2014; Marzal, De Lope, Navarro, & Møller, 2005). For example, disease caused by *Trichuris* infection is considered a neglected tropical disease that affects an estimated ~800 million people worldwide and is responsible for a variety of clinical symptoms including developmental disabilities, growth stunting and anaemia (Stephenson, Holland, & Cooper, 2000). Although quantifying disease states and health in natural animal populations is difficult, *Trichuris* infection in the Ugandan red colobus was recently shown to be associated with sickness behaviours, including decreased grooming, moving and copulation (Ghai et al., 2015). Additionally, *Trichuris* has been identified as a multihost, potentially zoonotic, parasite in the Kibale primate community (Ghai, Simons, et al., 2014). Typical ecological and demographic factors have not been able to consistently predict patterns of *Trichuris* infection in red colobus living in Kibale. Factors including season (Gillespie, Greiner, et al., 2005), age (Gillespie, Greiner, et al., 2005), group size (Chapman, Rothman, & Hodder, 2009), forest type within park boundaries (Chapman, Speirs, et al., 2006; Gillespie, Chapman, et al., 2005) and sex (Gillespie, Greiner, et al., 2005) have shown no association with patterns of *Trichuris* infection, while Snaith, Chapman, Rothman, and Wasserman (2008) and Ghai et al. (2015) found group size and sex, respectively, to be associated with patterns of infection. The lack of consistent results in such studies suggests that factors other than ecology and demography, such as host genetics, may contribute to patterns of control of infection.

1.2 | Regulation of MHC class II gene expression and its role in disease

To examine whether host genetics influence the control of *Trichuris* infection, we chose to target genetic variation in the promoter region of *MHC-DQA1*, a member of the major histocompatibility complex class II gene family associated with a variety of both autoimmune and infectious disease phenotypes, including extracellular pathogens (e.g., helminths; Ghodke, Joshi, Chopra, & Patwardhan, 2005).

Although laboratory studies of model organisms can provide valuable information about the mechanisms of antigen processing and presentation, studies of natural populations have greater potential to provide insights into the role MHC expression plays in shaping disease dynamics because they better reflect the complexity of interactions between an organism's biology and environment (Ellegren & Sheldon, 2008; Wolfe et al., 1998). Studies of natural populations are able to incorporate variation often held constant in laboratory studies, such as environmental parameters and the potential effects of coinfection on infection duration, severity and transmission (Vaumourin, Vourc'h, Gasqui, & Vayssier-Taussat, 2015). This ecoimmunological approach has already provided interesting information on the relationship between MHC class II gene expression and disease susceptibility and infection intensity (Axtner & Sommer, 2011; Wegner, Kalbe, & Reusch, 2007). Although these studies demonstrate an association between MHC II gene expression and disease, they do not identify the specific genetic variants underlying these associations.

There is growing evidence that *cis*-regulatory variation plays an important role in driving patterns of gene expression (Backström, Shipilina, Blom, & Edwards, 2013; Khaitovich, Pääbo, & Weiss, 2005;

Otto et al., 2009), including variation in the promoter regions of MHC class II genes (Handunnetthi, Ramagopalan, Ebers, & Knight, 2010; Lee, Jong Kim, & Park, 2000). For *MHC-DQA1* in particular, studies have connected variation in the promoter to differential patterns of *MHC-DQA1* expression across various cell lines (Fernandez, Wassmuth, Knerr, Frank, & Haas, 2003), the binding affinity of transcription factors (Indovina et al., 1998) and the expression of reporter genes in cell-based assays (Loisel et al., 2006; Morzycka-Wroblewska, Munshi, Ostermayer, Harwood, & Kagnoff, 1997). Regulatory variation in *MHC-DQA1* has also been implicated in both infectious and autoimmune diseases (Britten, Mijovic, Barnett, & Kelly, 2009; Haas et al., 1995; Paz Bettinotti et al., 1993; Yao et al., 1993). However, characterization of *cis*-regulatory variation and the relationship between differential gene expression and control of infection remain far less understood compared to the relationship between variation in the protein-coding regions of MHC class II genes and disease phenotypes (as reviewed in Mach, Steimle, Martinez-Soria, & Reith, 1996; Klein, 2001; Neefjes, Jongsma, Paul, & Bakke, 2011).

1.3 | Functional *cis*-regulatory variation in nonhuman primates

To date, relatively few studies in nonhuman primates have identified and functionally characterized *cis*-regulatory variation. Previous research has largely focused on either characterizing functional variants or demonstrating associations between variants and disease (e.g., Clough, Kappeler, & Walter, 2011; Loisel et al., 2006). To our knowledge, only a single study has both identified and functionally characterized infection-associated *cis*-regulatory variation in a wild primate (Tung et al., 2009). Tung et al. (2009) identified a *cis*-regulatory variant in the *FY* (*DARC*) gene, an erythrocyte surface chemokine receptor, that was associated with infection by a malaria-like parasite (*Hepatocystis)* and influenced *FY* expression in vivo in a population of wild baboons. The functional consequences of this *cis*-regulatory variant on gene expression were also demonstrated in vitro using reporter assays.

We investigated the relationship between host genetics and control of *Trichuris* infection in red colobus in Kibale National Park, Uganda, to address three objectives: (i) identify SNPs in the promoter region of *MHC-DQA1* associated with control of *Trichuris* infection; (ii) functionally characterize these infection-associated SNPs; and (iii) test for evidence of balancing selection on this regulatory region.

2 | METHODS AND MATERIALS

2.1 | Ethics statement

All animal use followed the guidelines of the Weatherall Report (Weatherall, 2006) on the use of nonhuman primates in research and was approved by the Uganda Wildlife Authority, the Uganda National Council for Science and Technology, and the University of Oregon and University of Wisconsin Animal Care and Use Committees prior to initiation of the study (University of Oregon IACUC #15-06A; University of Wisconsin IACUC #A3368-01). Biological materials were shipped internationally under CITES permit #002290.

2.2 | Study system and sample collection

This research is part of the Kibale EcoHealth Project, which is a long-term study of the factors affecting the prevalence of infectious disease across the human and wildlife community in and surrounding Kibale National Park, Uganda (0°13′–0°41′N, 30°19′–30°32′E). One of the 13 primate species found in Kibale is the Ugandan red colobus, which is an Old World monkey in the subfamily Colobinae. In general, red colobus are arboreal and sensitive to habitat disturbance and prefer old growth forest (Nowak & Lee, 2011; Ruiz-Lopez et al., 2016; Struhsaker, 2005). The small camp (SC) group is a well-habituated social group that has been a focus of the Kibale EcoHealth Project since 2005 (Ghai et al., 2015; Goldberg et al., 2009; Simons et al., 2016). This group was composed of approximately 104 adult individuals in 2011 (Miyamoto, Allen, Gogarten, & Chapman, 2013). We chose to focus on 40 individuals (38 adults, two subadults; $M = 19$, $F = 21$) from the SC group for which we could reliably obtain both parasitological and genetic data. Parasitological data were generated from faecal samples (see Section 2.3), and genetic data were generated using either blood or faecal samples (see Section 2.4).

2.3 | Parasite data collection

We first sought to characterize *Trichuris* infection in all 40 individuals. Parasite infection was indexed as the number of *Trichuris* eggs per gram of faecal material. Faecal egg count should be recognized as an index that only approximately corresponds to adult nematode number due to variation in egg production over time (Hodder & Chapman, 2012). However, it is a commonly used, albeit imperfect, proxy for adult nematode infection intensity (Cabaret, Gasnier, & Jacquiet, 1998). For these 40 individuals, we collected a total of 658 faecal samples (see Table S1) between 2007 and 2012 as part of the Kibale EcoHealth Project's long-term study of parasitism and used these samples to generate faecal egg count data. The mean number of samples per individual was 16.5 ± 0.35 (range: 2–38; $M = 366$, $F = 292$).

Approximately 1 g of faeces was stored in 2 ml of formalin. Samples were processed with a modified ethyl acetate concentration method following Greiner and McIntosh (2009). Following sedimentation, multiple thin smears totalling one full gram of faecal sediment were used to identify and count *Trichuris* eggs at 10× objective magnification on a Leica DM2500 (Wetzlar, Germany) light microscope. Faecal egg counts are reported as eggs per gram of sedimented faeces.

2.4 | Sample collection, DNA extraction, and sequencing

DNA was extracted from blood ($n = 22$) or faecal ($n = 18$) samples to generate genetic data from all 40 individuals. Whole-blood

samples were collected from collared individuals between 2010 and 2013 and stored on Whatman FTA Classic Cards (GE Healthcare Life Sciences, Marlborough, MA, USA) using methods described in Lauck et al., 2011. Faecal samples were noninvasively collected during daily follows between June and August 2012. Approximately 2 g of faeces were placed in 5 ml of RNAlater® (Carlsbad, CA, USA), homogenized and stored at −20°C. Genomic DNA was extracted from dried blood spots using QIAamp DNA Mini Kits (Qiagen, Valencia, CA, USA) and from faecal samples using QIAamp DNA Stool Mini Kits (Qiagen).

The *MHC-DQA1* proximal core promoter region (280-bp region ~13 bp upstream of start site) for all 40 individuals was PCR-amplified on an Eppendorf Mastercycler (Hamburg, Germany) using primers DQApFor (5′-CAGACATGCACACACCAGAGAA-3′) and DQApRev (5′-GGATCATCYTCTTCCCAAGG-3′) from Loisel et al. (2006). The thermocycler protocol included the following: (i) an initial denaturation of 95°C for 3 min; (ii) 40 cycles of 95°C denaturing for 30 s, 56°C annealing for 30 s and 72°C extension for 90 s; and (iii) a final extension of 72°C for 7 min. Reactions were carried out in a 25 μl volume containing 1X GoTaq Green Mastermix (Promega, Madison, WI, USA), 0.16 μM of each primer and 2 μl (5–20 ng) of DNA. PCR amplicon clean-up followed the manufacturer's protocol for ExoSAP-IT (Affymetrix, Santa Clara, CA, USA), and sequencing was done on an ABI 3730 capillary sequencer (Applied Biosystems, Foster City, CA, USA) at the Center for Genome Research and Biocomputing of Oregon State University. Each PCR was conducted in duplicate, and sequence traces were cleaned, trimmed and aligned in GENEIOUS v.7 (Kearse et al., 2012).

2.5 | Genetic associations

To reduce the number of statistical tests, we restricted our analysis to SNPs in transcription factor binding sites (Figure 1a) with a minor allele frequency >0.05. Faecal egg counts were square-root-transformed; samples with an egg count of zero were transformed to 0.05 before square-root transformation (McDonald, 2009). Differences in egg count based on sex were tested using a *t* test. The

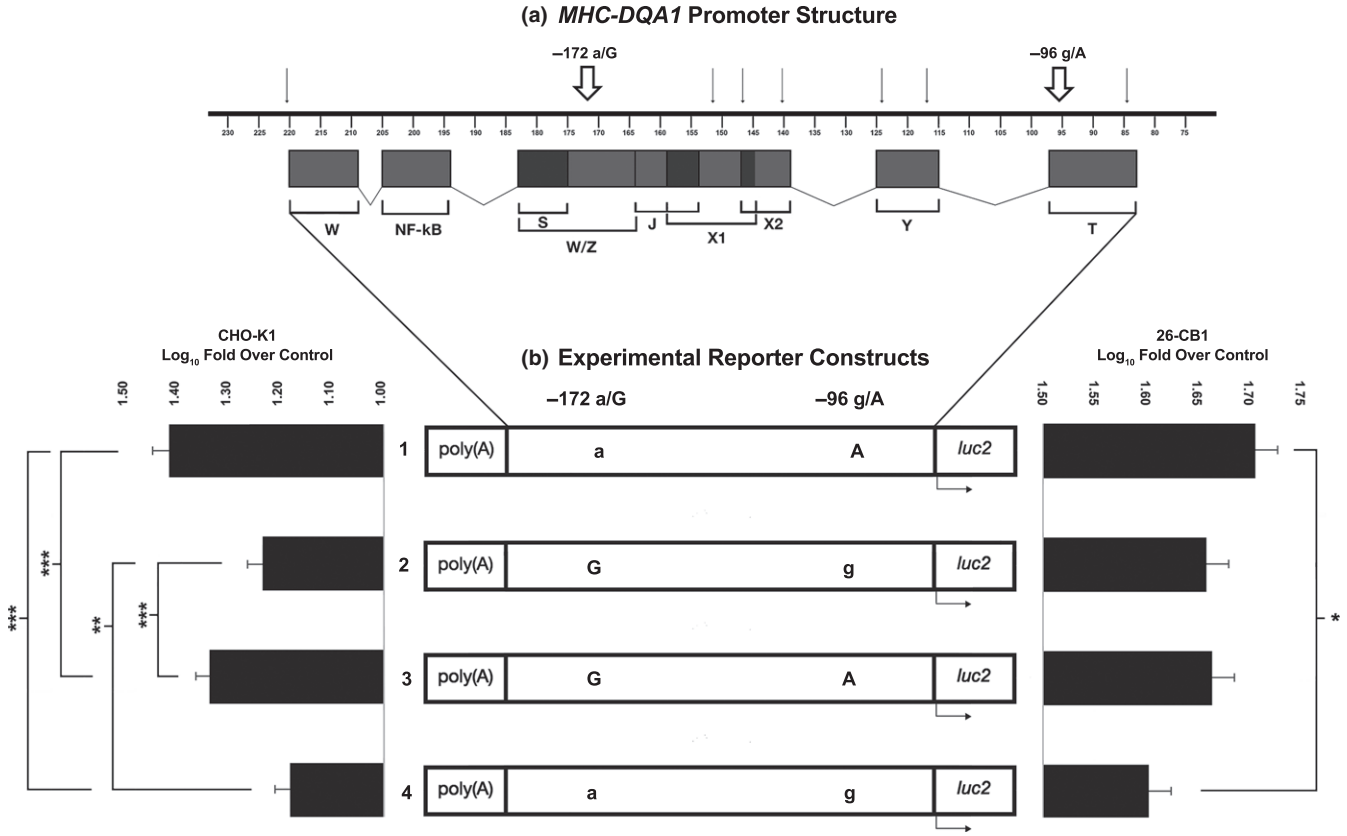

FIGURE 1 (a) *MHC-DQA1* promoter structure and SNPs. Transcription factor binding sites (based on Morzycka-Wroblewska & Harwood, 1993; Loisel et al., 2006) located in the proximal core promoter (number denotes position upstream of transcription start site) are shown as grey boxes. Areas where binding sites overlap are shaded in dark grey. All SNPs found within transcription factor binding sites are shown (black arrows), and the two SNPs associated with decreased control of *Trichuris* infection are indicated with open arrows. (b) Structure of experimental promoter constructs for both SNPs. Constructs 1–4 (labelled to the left of each construct) represent each of four possible allele combinations given two SNPs with two alleles each. For each of the four constructs, the ancestral (lower-case) and derived (upper-case) alleles are noted at both variable sites. Adjacent to each construct is the normalized luciferase activity (luciferase normalized to *Renilla* coreporter, fold over control calculated, then Log_{10}-transformed) of CHO-K1 cells (left) and baboon lymphoblast cells (26CB-1; right). Significant differential expression between constructs is noted (*$p < .05$, **$p < .005$, ***$p < .0005$; see also Table 1)

TABLE 1 Results of tests for differential expression between ancestral and derived alleles of SNP −172 a/G and SNP −96 g/A. For both SNPs, differences between the ancestral vs. derived were tested with each of the ancestral and derived alleles at the other SNP. See Figure 1b for combinations of alleles in each construct. Tests with significant differences between alleles are indicated in bold

SNP	Constructs	State at other SNP	Cell line	*p* Value
−96 g/A	**1 vs. 4**	**Ancestral**	**CHO-K1**	**1.2×10^{-11}**
−96 g/A	**2 vs. 3**	**Derived**	**CHO-K1**	**2.1×10^{-7}**
−172 a/G	**1 vs. 3**	**Ancestral**	**CHO-K1**	**.001**
−172 a/G	**2 vs. 4**	**Derived**	**CHO-K1**	**.005**
−96 g/A	**1 vs. 4**	**Ancestral**	**26CB-1**	**.03**
−96 g/A	2 vs. 3	Derived	26CB-1	.46
−172 a/G	1 vs. 3	Ancestral	26CB-1	.26
−172 a/G	2 vs. 4	Derived	26CB-1	.14

effect of genotype at each individual SNP on egg count was fitted using a generalized linear mixed model (GLMM). Our model included genotype and sex as fixed effects and included collection time (AM/PM categories) and individual ID as random effects. We also tested for interaction effects between infection-associated SNPs by fitting a GLMM using the full model including infection-associated SNPs. Additionally, we calculated two estimates of linkage disequilibrium between infection-associated SNPs, D′ and r^2. We performed a chi-square test to determine the significance of r^2. All tests were conducted in R (R Core Team 2013).

2.6 | Reporter assays

To test the functional consequences of the identified SNPs on gene expression in vitro, we designed four constructs representing all possible combinations of haplotypes for two SNPs with two states (Figure 1b). Reporter constructs were synthesized and cloned into the pGL4.10[*luc2*] vector by GenScript, and we confirmed construct identities by sequencing. We assessed the ability of the constructs to drive gene expression in two mammalian cell lines: Chinese hamster ovary cells (CHO-K1; ATCC), which are widely used for reporter assays (Xu et al., 2011), and baboon lymphoblasts (26CB-1, ATCC). CHO-K1 cells were grown in 96-well plates in a volume of 100 μl per well and transfected with 100 ng of one of the four reporter plasmid constructs (pGL4.10[*luc2*]) or control reporter plasmid (no cloned promoter) and 0.1 ng of pGL4.73 [*hRluc*/SV40] *Renilla* constitutive luciferase reporter plasmid using Lipofectamine and Plus Reagent in OPTIMEM (Invitrogen). After 4 hr, transfection medium was replaced with phenol-red-free α-MEM supplemented with 10% dextran–charcoal-stripped FBS (HyClone). Cells were then incubated at 37°C with 5% CO_2 for 24 hr and assayed using the Dual-Glo luciferase system (Promega). In addition to CHO-K1 cells, we replicated our experiments in baboon lymphoblast cells (26CB-1), which are primate-derived naïve immune cells. Baboon lymphoblast cells were grown to a cell density of 1×10^6/ml and 100 μl of cells were transfected with 5,000 ng of cloned reporter plasmid (pGL4.10[*luc2*]) and 100 ng of pGL4.73[*hRluc*/SV40] *Renilla* constitutive luciferase reporter plasmid using electroporation. Electroporation was used for baboon lymphoblast cells as the baboon cells were resistant to lipofection-mediated transfection. Cells were electroporated with a 25 ms square wave pulse at 120 V in a 0.2-cm cuvette with a Bio-Rad GenePulser XCell. Electroporated cells were immediately transferred to 500 μl of media in 48-well plates, then incubated at 37°C with 5% CO_2 for 24 hr and assayed using the Dual-Glo luciferase system (Promega). Levels of transcriptional activity were measured via fluorescence on a PerkinElmer Victor X5 Multilabel plate reader. Light units for each well were normalized to the *Renilla* coreporter units to control for differences in transfection efficiency and cell concentrations across wells. Each plate contained an equal number of control replicates with a promoterless construct, and *Renilla*-normalized luciferase values were then used to calculate the fold activation over the average control for that plate. For CHO-K1 cells, three separate assays were conducted, each with 12 technical replicates. For baboon lymphoblast cells, three separate assays were conducted, each with eight technical replicates.

All measured fluorescent replicates (CHO-K1 $n = 36$; baboon lymphoblast $n = 24$) were log-transformed and tested for normality using a Shapiro–Wilk test. Log-transformed data for both cell types met the assumption of normality, and differences between the transcriptional activity of the ancestral and derived haplotypes (determined based on conserved nucleotides across the alignment of primate taxa in Loisel et al., 2006) were tested using a *t* test. For each SNP, tests were conducted between the constructs with the ancestral and derived alleles at the site being tested, with both the ancestral and derived allele at the second site. All tests were conducted in R (R Core Team 2013).

2.7 | Balancing selection—trans-species polymorphism

To examine whether balancing selection may be acting on the *MHC-DQA1* promoter region, we investigated the presence of trans-species polymorphism between red colobus and other primates at this locus by reconstructing the *MHC-DQA1* promoter phylogeny and conducting two tree topology tests. This method allowed us to assess the presence of trans-species polymorphism and to test whether the *MHC-DQA1* promoter tree topology or the species tree topology fit the data significantly better than the other.

The *MHC-DQA1* promoter phylogeny was constructed with the following species: red colobus ($N = 40$, MF133402–MF133441), baboon ($N = 12$, DQ924473–DQ924484), macaque ($N = 4$, DQ924485–DQ924488), chimpanzee ($N = 4$, DQ924453–DQ924456), bonobo ($N = 3$, DQ924457–DQ924459), human ($N = 3$; AF099905.1, M97463, M97459), gorilla ($N = 2$, DQ924460, DQ92446), orangutan ($N = 7$, DQ924462–DQ924468) and gibbon ($N = 4$, DQ924469–DQ924472). Sequences were aligned using

ClustalW in MEGA7 (Kumar, Stecher, & Tamura, 2016) and trimmed to 268 base pairs. The best fit model of nucleotide evolution was determined using the maximum-likelihood method implemented in the model selection analysis in MEGA7. The best fit model (Kimura 2-parameter [K2 + G]) was used for phylogenetic reconstruction of the relationships between *MHC-DQA1* promoter sequences produced in this study and those from Loisel et al. (2006). Phylogenetic relationships were inferred using MRBAYES v3.2.5 (Ronquist & Huelsenbeck, 2003) using the K2 + G model, 5×10^6 Markov chain Monte Carlo replicates, with sampling every 100 generations after a 20% burn-in. Chains were determined to be converged at the end of the run based on the averaged standard deviation of split frequencies (0.003).

Phylogenetic incongruence between the species tree and *MHC-DQA1* promoter tree was quantified by two topology tests: the Shimodaira–Hasegawa (SH) test, and the approximately unbiased (AU) test as implemented in IQ-TREE (Nguyen, Schmidt, von Haeseler, & Minh, 2015). Both tests allow the rejection of alternative tree topologies based on likelihood, while the AU test provides a less biased bootstrap correction relative to the SH test. Both tests were run with 1,000 RELL replicates. The species phylogeny used for topology tests was the maximum clade credibility consensus tree from the 10K Trees project (Arnold, Matthews, & Nunn, 2010) and was consistent with previous phylogenies for included taxa (Perelman et al., 2011).

2.8 | Balancing selection—site-frequency spectra

We also investigated the evidence for balancing selection by evaluating the site-frequency spectrum, as a skew towards intermediate-frequency alleles relative to the genomewide spectrum is indicative of balancing selection (Andrés et al., 2009).

The site-frequency spectrum for a set of loci is shaped by demography and natural selection, and deviations from expectations under neutral evolution are used to infer patterns of selection. *MHC-DQA1* sequence data were phased to haplotypes using PHASE 2.1 (Stephens & Scheet, 2005) and SEQPHASE (Flot, 2010), and deviations from neutral evolution were evaluated by the site-frequency spectrum using the PEGAS package (Paradis, 2010) in R (R Core Team 2013) where the *x*-axis represents the distribution of allele frequencies and the *y*-axis represents the population frequency of those allele frequencies. Results indicative of balancing selection, such as an excess of intermediate-frequency alleles, can also be indicators of demographic history, such as a recent population contraction (Richman, 2000). To further account for demography in our evolutionary analyses, we analysed a preexisting genomewide restriction-site-associated DNA (RAD) data set consisting of 24 Ugandan red colobus individuals (Blanchet et al., 2015). These individuals all belong to the Kibale population, with 11 individuals from the SC group (eight of which overlap with the *MHC-DQA1* data set) and 13 individuals from a different social group (see Blanchet et al., 2015 for details on RAD data collection). We identified RAD loci, called SNPs for each individual, and selected polymorphic loci using the program STACKS v. 1.1 (available at http://creskolab.uoregon.edu/stacks/; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013).

We found 35,755 polymorphic RAD loci that were present in all 24 individuals. The consensus sequences for all the polymorphic loci found were compared against the NCBI NR protein database and UNIPROT database using the BLASTx function in BLAST v.2.2.27 (Altschul et al., 1997). In both cases, BLASTx parameters included an e-value of 1×10^{-6}, number of Blast hits = 20, and HSP length cut-off = 33. Of the 35,755 loci, 4,176 had Blast hits in either the NCBI NR or the UNIPROT database and were considered to be protein-coding loci. 31,579 did not have any hits and were considered noncoding. For each of these two groups, we used the module populations.pl in STACKS and calculated the minor allele frequency at each SNP. These analyses were carried out on the University of Oregon Applied Computational Instrument for Scientific Synthesis (ACISS) server. From the RAD data, we generated genomewide (protein-coding + noncoding), protein-coding, and noncoding folded site-frequency spectra in R (R Core Team 2013). While not all protein-coding loci show evidence of selection, and not all noncoding loci are neutral, we generated all three site-frequency spectra to serve as neutral expectations against which to interpret the results from the *MHC-DQA1* promoter. We performed a Welch two-sample *t* test to determine if the *MHC-DQA1* site-frequency spectrum differed significantly from the genomewide noncoding site-frequency spectrum.

3 | RESULTS

3.1 | Two SNPs in the core promoter of *MHC-DQA1* are associated with decreased control of *Trichuris* infection

The mean normalized egg count across the individuals was 0.88 ± 0.08 (M = 0.93 ± 0.13, F = 0.73 ± 0.12), and the sex difference was nonsignificant (p = .27). Prevalence of *Trichuris* infection across the sample (N = 40) was 97.5%, and no sex difference was observed (M = 90%; F = 89%; χ^2 (1) = 0.01, p = .92).

We identified nine SNPs located in transcription factor binding sites (Figure 1a). Of those nine SNPs, two had genotypes that were significantly associated with higher egg count. One SNP (−172 a/G; ancestral and derived alleles will be denoted with lower case and upper case, respectively, when discussing reporter constructs) is located 172 base pairs upstream of the transcription start site (TSS) in a conserved *cis*-regulatory element called the W/S/Z box, part of the SXY module. The second SNP (−96 g/A) is located 96 base pairs upstream of the TSS in a putative T-box element. Genotypes frequencies for −172 a/G were G/G: 0.15, A/A: 0.57, G/A: 0.27, and for −96 g/A were G/G: 0.62, A/A: 0.05; G/A: 0.32. We found no evidence for linkage disequilibrium between the two SNPs according to either D′ (0.07), or r^2 (0.004; χ^2 (1) = 0.14, p = .71). For both SNPs, the heterozygote genotype (Figure 2) was significantly associated with higher egg counts (−172 a/G, p = .02; −96 g/A, p = .02;

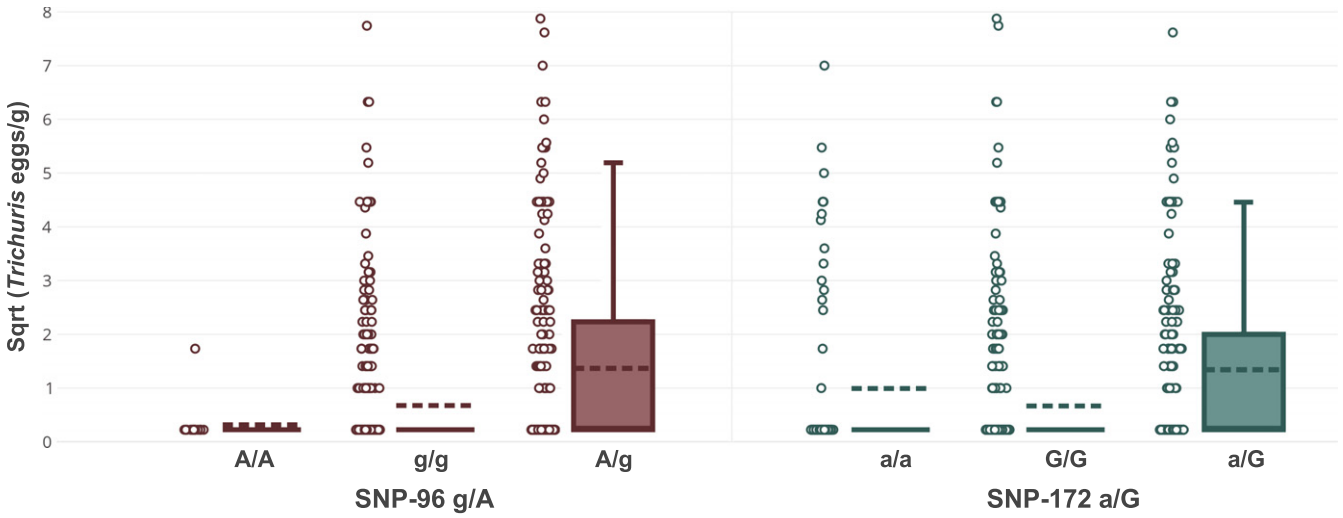

FIGURE 2 Infection intensity (square-root-transformed *Trichuris* egg counts) for each of the three possible genotypes for both of the two significant SNPs. For each genotype, a stripchart (L) and boxplot (R) are shown

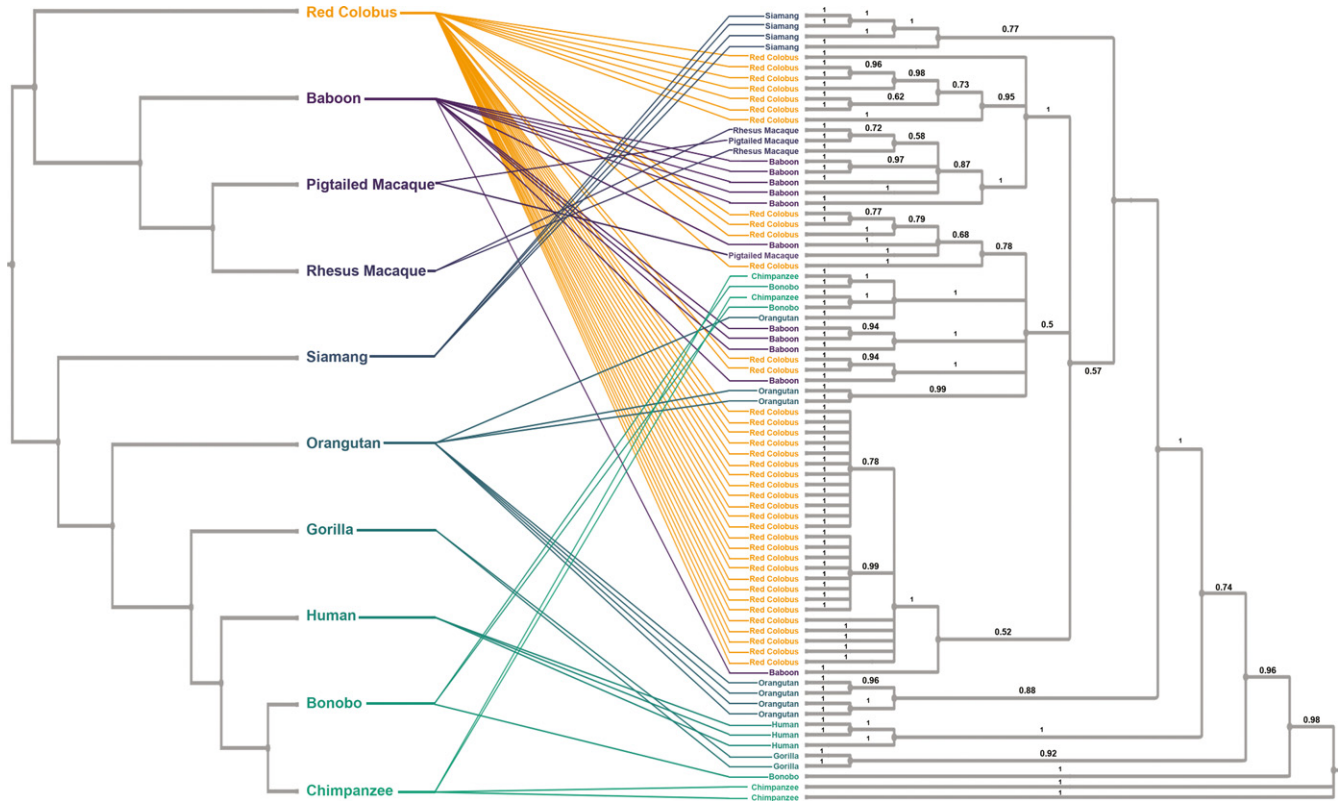

FIGURE 3 Trans-species polymorphism and phylogenetic incongruence between species tree (left) and *MHC-DQA1* promoter gene tree (right). Branch lengths are arbitrary for both trees, with posterior probability of branch support above branches on *MHC-DQA1* tree. Results of both topology tests indicated significant incongruence between the species tree and gene tree with the species tree topology being rejected in favour of the *MHC-DQA1* promoter tree topology: Shimodaira–Hasegawa (SH) test ($p \leq .0000$) and the approximately unbiased (AU) test ($p = .0002$)

See Table S2). Genotype was not significantly associated with egg counts for the remaining seven SNPs (−83 A/G, $p = .67$; −116 C/T, $p = .78$; −122 A/G, $p = .69$; −139 C/T, $p = .36$; −146 G/A, $p = .57$; −150 C/T, $p = .58$; −217 C/T, $p = .48$). No significant interaction effects were observed between the two significant SNPs ($z = 0.42$, $p = .66$).

3.2 | Infection-associated SNPs drive differential expression in vitro

MHC-DQA1 promoter sequences containing all combinations of the ancestral and derived states at each of the two SNP sites (−172 a/G and −96 g/A) drove differential expression in vitro (Table 1). Frequencies in our sample of the four allelic combinations from the four synthesized constructs (see Figure 1 for constructs) were construct 1: 0.0, construct 2: 0.05, construct 3: 0.22, and construct 4: 0.72. For SNP-172 a/G, the ancestral allele (a) drove differential luciferase expression compared to the derived allele (G) in CHO-K1 cells regardless if −96 g/A was in the ancestral (g) or derived (A) state ($p = .005$, $p = .001$, respectively). In baboon lymphoblast cells, the difference between the ancestral allele (−172a) and the derived allele (-172G) was nonsignificant (ancestral background, $p = .26$; derived background, $p = .14$). At SNP −96 g/A, the derived allele (A) drove significantly higher expression than the ancestral (g) allele in both CHO-K1 and baboon lymphoblast cells with the ancestral allele (a) at position −172 ($p = 1.2 \times 10^{-11}$, and $p = .03$, respectively). The derived allele at −96 (A) also drove significantly higher expression with the derived state at −172 (G) in CHO-K1 cells ($p = 2.1 \times 10^{-7}$), but not baboon lymphoblast cells ($p = .46$; Figure 1b).

3.3 | Balancing selection has shaped *MHC-DQA1* promoter variation in red colobus

The *MHC-DQA1* promoter phylogeny exhibited extensive trans-species polymorphism (Figure 3), and there was significant phylogenetic incongruence between the topologies with the species tree topology being rejected in favour of the *MHC-DQA1* promoter tree topology (SH test: $p < .0000$, AU test: $p = .0002$). This result deviates from expectations under neutrality and is considered robust evidence for balancing selection (Fijarczyk & Babik, 2015). It is also consistent with evidence of balancing selection on the *MHC-DQA1* promoter found across primates in other studies (Lindo et al., 2016; Loisel et al., 2006).

In addition, the folded site-frequency spectrum for *MHC-DQA1* deviated from the L-shaped spectrum expected under neutral evolution (Figure 4a) and showed an excess of intermediate-frequency alleles relative to genomewide ($p = .0003$), coding ($p = .0009$), and noncoding RADseq data ($p = .0003$), which is consistent with expectations for balancing selection acting on the *MHC-DQA1* promoter. The site-frequency spectra based on genomewide (35,755), protein-coding (4,176), and noncoding (31,579) loci conformed to the neutral expectation; all had the expected L-shaped curve (Figure 4b–d). While past changes in population size can mimic signatures of selection, our results from RAD data suggest that historic demographic changes are not responsible for the shape observed in the *MHC-DQA1* site-frequency spectrum (Allen et al., 2012; Gattepaille, Jakobsson, & Blum, 2013). Additionally, estimates of nucleotide diversity were higher for the *MHC-DQA1* promoter ($\pi = 5.94$) than for genomewide ($\pi = 0.27$, $\sigma^2 = 0.02$, $SE = 0.0008$), coding ($\pi = 0.26$, $\sigma^2 = 0.02$, $SE = 0.002$) and noncoding RADseq data ($\pi = 0.27$, $\sigma^2 = 0.02$, $SE = 0.0007$).

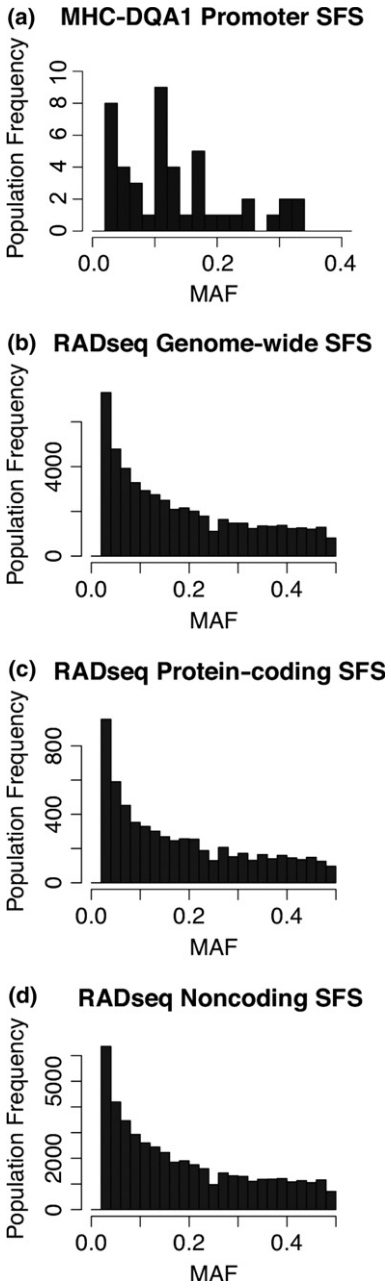

FIGURE 4 Folded site-frequency spectra (SFS; MAF: minor allele frequency). The *y*-axis represents the population frequencies of observed minor allele frequencies (*x*-axis). Panel (a) represents the SFS from the *MHC-DQA1* promoter; the distribution deviates from the L-shaped distribution expected under neutrality and is evidence of balancing selection maintaining an excess of intermediate-frequency alleles. Panels (b–d) represent SFS from 35,755 genomewide (protein-coding + noncoding), 4,176 protein-coding and 31,579 noncoding restriction-site-associated DNA (RAD) loci, respectively. The *MHC-DQA1* SFS deviates significantly from genomewide, coding and noncoding SFS ($p = .0003$, $p = .0009$, $p = .0003$, respectively)

4 | DISCUSSION

We identified and functionally characterized two infection-associated *cis*-regulatory SNPs in the promoter region of *MHC-DQA1* in a wild population of red colobus. For both SNPs, heterozygotes had decreased control of *Trichuris* infection when compared to homozygotes. We demonstrated experimentally that both SNPs are functional and capable of differentially driving gene expression in vitro. For both SNPs, the ancestral and derived alleles drove significant differential expression in CHO-K1 cells, regardless of the background SNP. We saw similar patterns of differential expression in baboon lymphoblasts, although for −172 a/G, the differences were nonsignificant. One possible explanation for the observed pattern of nonsignificance for this SNP in baboon lymphoblasts is the relatively large level of variation in electroporation efficiency compared to lipofection in our experiments. The relatively higher variation in our electroporation results could be due to electrical field distortion, local pH variation, excess heat or a multitude of other factors, all of which can affect cell viability and electroporation efficiency (Geng & Lu, 2013). Alternatively, these SNPs may not drive differential expression in baboon lymphoblasts; however, the consistency in expression patterns between these distinct mammalian cell lines and the larger relative variation in transfection efficiency among technical replicates when using electroporation vs. lipofection to transfect the two different cell lines suggest that the differences in expression are biologically relevant. In addition, we found evidence that this *cis*-regulatory region is under balancing selection. Taken together, the patterns observed here suggest a trade-off between *MHC-DQA1 cis*-regulatory variation maintained by balancing selection and decreased control of *Trichuris* infection in heterozygotes. This interpretation is predicated on two ideas: (i) that balancing selection in regulatory regions provides transcriptional flexibility and an evolutionary advantage and (ii) this advantage outweighs any fitness consequences that heterozygotes with decreased control of parasite infection would incur.

4.1 | Balancing selection and regulatory flexibility

Extensive evidence exists to suggest that balancing selection occurs in *cis*-regulatory regions of immune-related genes (Bamshad et al., 2002; Cagliani et al., 2008; Wilson et al., 2006), including MHC (Lindo et al., 2016; Liu et al., 2006; Loisel et al., 2006; Tan, Shon, & Ober, 2005). Similar to Loisel et al. (2006), we found extensive trans-species polymorphism between red colobus and deeply divergent taxa, including other Old World monkeys and apes. As the age of divergent lineages increases, gene tree/species tree incongruence due to incomplete lineage sorting becomes less likely, and presence of balanced polymorphisms across divergent taxa (and subsequent patterns of species/gene tree incongruence) is maintained exclusively by balancing selection (Fijarczyk & Babik, 2015). These findings are supported by the significant excess of intermediate-frequency alleles in the *MHC-DQA1* site-frequency spectrum analysis and demonstration of function in our in vitro reporter assays.

Although it is possible that the signatures of balancing selection observed are due to linkage disequilibrium with the *MHC-DQA1* exon 2 (Loisel et al., 2006; also see Lenz, Spirin, Jordan, & Sunyaev, 2016), the strength of hitchhiking decays rapidly over chromosomal distance (Hudson & Kaplan, 1988; Kelly & Wade, 2000) and empirical studies demonstrate that high levels of diversity in exon 2 decrease considerably in flanking intronic sequences within a few hundred base pairs of class II loci in humans (Fu et al., 2003). Additionally, the probability of hitchhiking at sites >1 kb from the selected site is low (Wiuf, Zhao, Innan, & Nordborg, 2004), particularly when considering that the *cis*-regulatory region investigated here is ~4 kb from the highly variable exon 2, and recombination, both historical and recent, has been shown to break down linkage between these sites in humans (Alaez, Vazquez-Garcia, & Gorodezky, 2001; Petronzelli, Kimura, Ferrante, & Mazzilli, 1995; Takahata & Satta, 1998). Our results also indicate the two infection-associated SNPs in the current study are not in significant linkage with each other. Alternatively, the alleles identified here may be too recent to have been removed from the population by negative selection. Although this is possible given that the derived alleles at both SNPs are fixed in red colobus relative to other primate taxa (Loisel et al., 2006), the alleles could be as old as the divergence between Cercopithecinae and Colobinae (by ~15 MYA, Sterner, Raaum, Zhang, Stewart, & Disotell, 2006; Ting, 2008b).

The variation produced and maintained by balancing selection in MHC class II *cis*-regulatory regions is thought to play several important roles. MHC class II expression is tightly regulated and largely restricted to professional antigen-presenting cells (APCs) including thymus epithelium, dendritic cells, B cells, activated T cells and macrophages. Therefore, high levels of diversity observed in MHC class II promoters are likely to provide adaptive potential for cell-specific expression (Guardiola, Maffei, & Lauster, 1996). Regulation of MHC class II is a critical process because it must provide an appropriately robust immune defence to pathogens while minimizing damage to the host (Ting & Trowsdale, 2002). Part of this critical balance is the ability to control expression according to cell-type specificity through both time (developmentally) and space (site of infection). Because natural selection cannot act on specific cell types in isolation, selection for differential expression at the organismal level allows a balanced immune response. Similar hypotheses regarding the adaptive value of flexibility in immunoregulation have also been made for regulatory variation observed in several cytokines (Daser, Mitchison, Mitchison, & Müller, 1996). The cell type-specific expression profiles resulting from regulatory variation described above have been previously demonstrated for *MHC-DQB1*, in which different promoter haplotypes drove expression differentially based on cell type (Beaty, Sukiennicki, & Nepom, 1999).

4.2 | Fitness trade-offs with *Trichuris* infection

Although it is possible that *Trichuris* infection does not have any fitness consequences, current views of sickness behaviour as an adaptive response to infection suggest that infection commonly has

fitness consequences (Poulin, 1995). Long-term maintenance of regulatory variation by balancing selection, enabling a flexible transcriptional response at the expense of decreased control of parasite infection as proposed here, should only occur if the fitness consequences of the parasites are relatively low. In long-lived host species, such as red colobus and other primates, it is difficult to quantify the fitness consequences of infection by specific parasites, particularly if those fitness consequences are low and can be masked by the behaviour of the host. Ghai et al. (2015) investigated sickness behaviours in response to *Trichuris* infection in the red colobus in Kibale and documented that sickness behaviours included increased resting, decreased energetic behaviours (including mating effort) and self-medicating through ingestion of specific anti-helminthic bark species. This suggests *Trichuris* infection likely has fitness consequences in the red colobus. However, our data suggest the fitness benefits conferred by balancing selection outweigh any potential fitness costs incurred by *Trichuris* infection, thus allowing balancing selection to maintain infection-associated alleles.

4.3 | Future directions

Studies of wild primates are often subject to small sample size because of difficulties of habituation and individual identification. The number of red colobus sampled is comparable to that used in other studies that addressed both *cis*-regulatory variation and function (Loisel et al., 2006), and *cis*-regulatory variant associations with parasite infection (Clough et al., 2011; Tung et al., 2009). The demonstration of a functional role of the two infection-associated SNPs also lends support to the biological relevance of the associations. In addition, to test the assumption that the fitness consequences of *Trichuris* infection are relatively low, it will be important to collect long-term reproductive and demographic data and incorporate these data into future models.

To further characterize how the *MHC-DQA1* core promoter influences gene expression in red colobus, it will be necessary to characterize other regulatory elements known to play a role in MHC class II expression and combine that information with measures of *MHC-DQA1* expression in vivo using a technique such as RNAseq. This will allow for an assessment of the role of individual regulatory elements in driving differences in gene expression, and ultimately control of parasite infection. In addition, it is currently unclear why heterozygotes at these two SNPs have decreased control of *Trichuris* infection. Although several examples exist in the literature that show disease associations with a heterozygote regulatory variant (Backström et al., 2013; Knapp et al., 2003; Morahan et al., 2002; Park, Martin, Zhang, Jegga, & Benson, 2012), no study has provided, either experimentally or theoretically, a viable mechanism for the phenomenon of why regulatory heterozygotes may be disadvantaged. While the experimental work necessary to untangle the complexity of the underling mechanism is outside the scope of this research, future research in this area would be highly valuable.

ACKNOWLEDGEMENTS

This research was funded by NIH grant TW009237 as part of the joint NIH–NSF Ecology of Infectious Disease programme and the UK Economic and Social Research Council, NSF BCS-1540459, National Geographic Society, NSERC and the University of Oregon. We thank the Uganda Wildlife Authority and Uganda National Council of Science and Technology for permission to conduct this research. We are grateful to Robert Basaija, Peter Tuhairwe, Clovice Kaganzi and Dr. Dennis Twinomugisha for assistance with logistics and fieldwork. We thank members of the Molecular Anthropology Group at University of Oregon for helpful feedback throughout the research process and valuable comments during manuscript preparation. We thank Dr. Ria Ghai for assistance with parasitological laboratory training, Dr. Ryan Raaum for valuable discussion on the manuscript, Dr. Jamie Bridgham for technical assistance with laboratory work, and comments from three anonymous reviewers that helped strengthen the manuscript.

DATA ACCESSIBILITY

Promoter sequences for all 40 individuals have been submitted to GenBank under Accession nos. MF133402–MF133441.

AUTHOR CONTRIBUTIONS

N.D.S., N.T., and K.N.S. designed the research; N.D.S., G.N.E., M.J.R.-L., P.A.O., C.A.C., T.L.G. and N.T. performed the research; N.D.S. analysed data; N.D.S. and K.N.S. wrote the manuscript with contributions from G.N.E., M.J.R.-L., C.A.C., T.L.G., and N.T.

REFERENCES

Alaez, C., Vazquez-Garcia, M. N., & Gorodezky, C. (2001). *DQA1* and *DQB1* promoter diversity and linkage disequilibrium with class II haplotypes in Mexican Mestizo population. *Genes and Immunity*, *2*(4), 216–221.

Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, *16*(4), 197–212.

Allen, J. M., Miyamoto, M. M., Wu, C. H., Carter, T. E., Ungvari-Martin, J., Magrini, K., & Chapman, C. A. (2012). Primate DNA suggests long-term stability of an African rainforest. *Ecology and Evolution*, *2*(11), 2829–2842.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.

Andrés, A. M., Hubisz, M. J., Indap, A., Torgerson, D. G., Degenhardt, J. D., Boyko, A. R., … Clark, A. G. (2009). Targets of balancing selection in the human genome. *Molecular biology and evolution*, *26*(12), 2755–2764.

Arnold, C., Matthews, L. J., & Nunn, C. L. (2010). The 10kTrees website: A new online resource for primate phylogeny. *Evolutionary Anthropology*, *19*, 114–118.

Axtner, J., & Sommer, S. (2011). *Heligmosomoides polygyrus* infection is associated with lower MHC class II gene expression in *Apodemus flavicollis*: Indication for immune suppression? *Infection, Genetics and Evolution*, *11*(8), 2063–2071.

Backström, N., Shipilina, D., Blom, M. P., & Edwards, S. V. (2013). *Cis*-regulatory sequence variation and association with Mycoplasma load in

natural populations of the house finch (*Carpodacus mexicanus*). *Ecology and Evolution*, *3*(3), 655–666.

Bailey, A. L., Lauck, M., Weiler, A., Sibley, S. D., Dinis, J. M., Bergman, Z., … O'Connor, D. H. (2014). High genetic diversity and adaptive potential of two simian hemorrhagic fever viruses in a wild primate population. *PLoS One*, *9*(3), e90714.

Bamshad, M. J., Mummidi, S., Gonzalez, E., Ahuja, S. S., Dunn, D. M., Watkins, W. S., … Ahuja, S. K. (2002). A strong signature of balancing selection in the 5′ *cis*-regulatory region of *CCR5*. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(16), 10539–10544.

Beaty, J. S., Sukiennicki, T. L., & Nepom, G. T. (1999). Allelic variation in transcription modulates MHC class II expression and function. *Microbes and Infection*, *1*(11), 919–927.

Blanchet, S., Bouchez, O., Chapman, C. A., Etter, P. D., Goldberg, T. L., Johnson, E. A., … Ting, N. (2015). Identification of SNP markers for the endangered Ugandan red colobus (*Procolobus rufomitratus tephrosceles*) using RAD sequencing. *Molecular Ecology Resources*, *15*, 684.

Borghans, J. A., Beltman, J. B., & De Boer, R. J. (2004). MHC polymorphism under host-pathogen coevolution. *Immunogenetics*, *55*(11), 732–739.

Britten, A. C., Mijovic, C. H., Barnett, A. H., & Kelly, M. A. (2009). Differential expression of HLA-DQ alleles in peripheral blood mononuclear cells: Alleles associated with susceptibility to and protection from autoimmune type 1 diabetes. *International Journal of Immunogenetics*, *36*(1), 47–57.

Cabaret, J. J., Gasnier, N., & Jacquiet, P. (1998). Faecal egg counts are representative of digestive-tract strongyle worm burdens in sheep and goats. *Parasite*, *5*, 137–142.

Cagliani, R., Fumagalli, M., Riva, S., Pozzoli, U., Comi, G. P., Menozzi, G., … Sironi, M. (2008). The signature of long-standing balancing selection at the human defensin beta-1 promoter. *Genome Biology*, *9*(9), R143.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140.

Chapman, C. A., Bowman, D. D., Ghai, R. R., Gogarten, J. F., Goldberg, T. L., Rothman, J. M., … Walsh, C. (2011). Protozoan parasites in group-living primates: Testing the biological island hypothesis. *American Journal of Primatology*, *73*, 1–8.

Chapman, C. A., Rothman, J. M., & Hodder, S. A. M. (2009). Can parasite infections be a selective force influencing primate group size? A test with red colobus. In M. A. Huffman & C. A. Chapman (Eds.), *Primate parasite ecology* (pp. 423–440). Cambridge: Cambridge University Press.

Chapman, C. A., Speirs, M. L., Gillespie, T. R., Holland, T., & Austad, K. M. (2006). Life on the edge: Gastrointestinal parasites from the forest edge and interior primate groups. *American Journal of Primatology*, *68*(4), 397–409.

Chapman, C. A., Struhsaker, T. T., & Lambert, L. E. (2005). Thirty years of research in Kibale National Park, Uganda, reveals a complex picture for conservation. *International Journal of Primatology*, *26*, 539–555.

Chapman, C. A., Wasserman, M. D., Gillespie, T. R., Speirs, M. L., Lawes, M. J., Saj, T. L., & Ziegler, T. E. (2006). Do food availability, parasitism, and stress have synergistic effects on red colobus populations living in forest fragments? *American Journal of Physical Anthropology*, *131*(4), 525–534.

Chen, K., van Nimwegen, E., Rajewsky, N., & Siegal, M. L. (2010). Correlating gene expression variation with *cis*-regulatory polymorphism in *Saccharomyces cerevisiae*. *Genome Biology and Evolution*, *2*, 697–707.

Clough, D., Kappeler, P. M., & Walter, L. (2011). Genetic regulation of parasite infection: Empirical evidence of the functional significance of an *IL4* gene SNP on nematode infections in wild primates. *Frontiers in Zoology*, *8*(1), 9.

Daser, A., Mitchison, H., Mitchison, A., & Müller, B. (1996). Non-classical-MHC genetics of immunological disease in man and mouse. The key role of pro-inflammatory cytokine genes. *Cytokine*, *8*(8), 593–597.

Ellegren, H., & Sheldon, B. C. (2008). Genetic basis of fitness differences in natural populations. *Nature*, *452*(7184), 169–175.

Fernandez, S., Wassmuth, R., Knerr, I., Frank, C., & Haas, J. P. (2003). Relative quantification of HLA-DRA1 and -DQA1 expression by real-time reverse transcriptase–polymerase chain reaction (RT–PCR). *European Journal of Immunogenetics*, *30*(2), 141–148.

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: Limits and prospects. *Molecular ecology*, *24*(14), 3529–3545.

Flot, J. F. (2010). SeqPHASE: A web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources*, *10*(1), 162–166.

Fu, Y., Liu, Z., Lin, J., Chen, W., Jia, Z., Pan, D., & Xu, A. (2003). Extensive polymorphism and different evolutionary patterns of intron 2 were identified in the HLA-DQB1 gene. *Immunogenetics*, *54*(11), 761–766.

Gattepaille, L. M., Jakobsson, M., & Blum, M. G. (2013). Inferring population size changes with sequence and SNP data: Lessons from human bottlenecks. *Heredity*, *110*(5), 409–419.

Geng, T., & Lu, C. (2013). Microfluidic electroporation for cellular analysis and delivery. *Lab on a Chip*, *13*(19), 3803–3821.

Ghai, R. R., Chapman, C. A., Omeja, P. A., Davies, T. J., & Goldberg, T. L. (2014). Nodule worm infection in humans and wild primates in Uganda: Cryptic species in a newly identified region of human transmission. *PLoS Neglected Tropical Diseases*, *8*(1), e2641.

Ghai, R. R., Fugere, V., Chapman, C. A., Goldberg, T. L., & Davies, T. J. (2015). Sickness behaviour associated with non-lethal infections in wild primates. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *282*, 20151436.

Ghai, R. R., Simons, N. D., Chapman, C. A., Omeja, P. A., Davies, T. J., Ting, N., & Goldberg, T. L. (2014). Hidden population structure and cross-species transmission of whipworms (*Trichuris* sp.) in humans and nonhuman primates in Uganda. *PLoS Neglected Tropical Diseases*, *8*, e3256.

Ghodke, Y., Joshi, K., Chopra, A., & Patwardhan, B. (2005). HLA and disease. *European Journal of Epidemiology*, *20*(6), 475–488.

Gillespie, T. R., Chapman, C. A., & Greiner, E. C. (2005). Effects of logging on gastrointestinal parasite infections and infection risk in African primates. *Journal of Applied Ecology*, *42*(4), 699–707.

Gillespie, T. R., Greiner, E. C., & Chapman, C. A. (2005). Gastrointestinal parasites of the colobus monkeys of Uganda. *Journal of Parasitology*, *91*(3), 569–573.

Goldberg, T. L., Chapman, C. A., Cameron, K., Saj, T., Karesh, W. B., Wolfe, N. D., … Slifka, M. K. (2008). Serologic evidence for novel poxvirus in endangered red Colobus monkeys, Western Uganda. *Emerging Infectious Diseases*, *14*(5), 801–803.

Goldberg, T. L., Sintasath, D. M., Chapman, C. A., Cameron, K. M., Karesh, W. B., Tang, S., … Switzer, W. M. (2009). Coinfection of Ugandan red colobus (*Procolobus* [*Piliocolobus*] *rufomitratus tephrosceles*) with novel, divergent delta-, lenti-, and spumaretroviruses. *Journal of Virology*, *83*, 11318–11329.

Greiner, E. C., & McIntosh, A. (2009). Collection methods and diagnostic procedures for primate parasitology. In M. A. Huffman & C. A. Chapman (Eds.), *Primate parasite ecology* (pp. 3–29). Cambridge: Cambridge University Press.

Guardiola, J., Maffei, A., & Lauster, R. (1996). Functional significance of polymorphism among MHC class II gene promoter. *Tissue Antigens*, *48*, 615–625.

Guo, H., Fortune, M. D., Burren, O. S., Schofield, E., Todd, J. A., & Wallace, C. (2015). Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate

causal genes in immune-mediated diseases. *Human Molecular Genetics*, *24*(12), 3305–3313.

Haas, J. P., Kimura, A., Truckenbrodt, H., Suschke, J., Sasazuki, T., Volgger, A., & Albert, E. D. (1995). Early-onset pauciarticular juvenile chronic arthritis is associated with a mutation in the Y-box of the HLA-DQA1 promoter. *Tissue Antigens*, *45*(5), 317–321.

Hahn, M. W. (2007). Detecting natural selection on *cis*-regulatory DNA. *Genetica*, *129*, 7–18.

Handunnetthi, L., Ramagopalan, S. V., Ebers, G. C., & Knight, J. C. (2010). Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. *Genes and Immunity*, *11*(2), 99–112.

Hart, B. L. (1988). Biological basis of the behavior of sick animals. *Neuroscience and Biobehavioral Reviews*, *12*(2), 123–137.

Hart, B. L. (1990). Behavioral adaptations to pathogens and parasites: Five strategies. *Neuroscience and Biobehavioral Reviews*, *14*(3), 273–294.

Hodder, S. A. M., & Chapman, C. A. (2012). Do nematode infections of red colobus (*Procolobus rufomitratus*) and black-and-white colobus (*Colobus guereza*) on humanized forest edges differ from those on nonhumanized forest edges? *International Journal of Primatology*, *33*, 845–859.

Hudson, R. R., & Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics*, *120*(3), 831–840.

Indovina, P., Megiorni, F., Ferrante, P., Apollonio, I., Petronzelli, F., & Mazzilli, M. C. (1998). Different binding of NF-Y transcriptional factor to DQA1 promoter variants. *Human Immunology*, *59*(12), 758–767.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Thierer, T. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

Kelly, J. K., & Wade, M. J. (2000). Molecular evolution near a two-locus balanced polymorphism. *Journal of Theoretical Biology*, *204*(1), 83–101.

Khaitovich, P., Pääbo, S., & Weiss, G. (2005). Toward a neutral evolutionary model of gene expression. *Genetics*, *170*, 929–939.

Klein, J. (2001). George Snell's first foray into the unexplored territory of the major histocompatibility complex. *Genetics*, *159*(2), 435–439.

Knapp, S., Hennig, B. J., Frodsham, A. J., Zhang, L., Hellier, S., Wright, M., … Thursz, M. R. (2003). Interleukin-10 promoter polymorphisms and the outcome of hepatitis C virus infection. *Immunogenetics*, *55*(6), 362–369.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, *33*(7), 1870–1874.

Lauck, M., Hyeroba, D., Tumukunde, A., Weny, G., Lank, S. M., Chapman, C. A., … Goldberg, T. L. (2011). Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan red colobus monkey discovered using direct pyrosequencing. *PLoS One*, *6*(4), e19056.

Lee, T. J., Jong Kim, S., & Park, J. H. (2000). Influence of the sequence variations of the HLA-DR promoters derived from human melanoma cell lines on nuclear protein binding and promoter activity. *Yonsei Medical Journal*, *41*(5), 593–599.

Lenz, T. L., Spirin, V., Jordan, D. M., & Sunyaev, S. R. (2016). Excess of deleterious mutations around HLA genes reveals evolutionary cost of balancing selection. *Molecular Biology and Evolution*, *33*(10), 2555–2564.

Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J., … Malhi, R. S. (2016). A time transect of exomes from a Native American population before and after European contact. *Nature Communications*, *7*, 13175.

Liu, X., Fu, Y., Liu, Z., Lin, B., Xie, Y., Liu, Y., … Zeng, K. (2006). An ancient balanced polymorphism in a regulatory region of human major histocompatibility complex is retained in Chinese minorities but lost worldwide. *American Journal of Human Genetics*, *78*(3), 393–400.

Loisel, D. A., Rockman, M. V., Wray, G. A., Altmann, J., & Alberts, S. C. (2006). Ancient polymorphism and functional variation in the primate MHC-DQA1 5′ *cis*-regulatory region. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(44), 16331–16336.

Mach, B., Steimle, V., Martinez-Soria, E., & Reith, W. (1996). Regulation of MHC class II genes: Lessons from a disease. *Annual Review of Immunology*, *14*, 301–331.

MacIntosh, A. J. J. (2014). Ecology and epidemiology of nematode infection in Japanese macaques. *Primate Research*, *30*(1), 23–51.

Marzal, A., De Lope, F., Navarro, C., & Møller, A. P. (2005). Malarial parasites decrease reproductive success: An experimental study in a passerine bird. *Oecologia*, *142*(4), 541–545.

McDonald, J. H. (2009). *Handbook of biological statistics*, 2nd ed. Baltimore, Maryland: Sparky House Publishing.

Minchella, D. J. (1985). Host life-history variation in response to parasitism. *Parasitology*, *90*(01), 205–216.

Miura, T., Brockman, M. A., Brumme, C. J., Brumme, Z. L., Carlson, J. M., Pereyra, F., … Baker, B. M. (2008). Genetic characterization of human immunodeficiency virus type 1 in elite controllers: Lack of gross genetic defects or common amino acid changes. *Journal of Virology*, *82*(17), 8422–8430.

Miyamoto, M. M., Allen, J. M., Gogarten, J. F., & Chapman, C. A. (2013). Microsatellite DNA suggests that group size affects sex-biased dispersal patterns in red colobus monkeys. *American Journal of Primatology*, *75*(5), 478–490.

Morahan, G., Huang, D., Wu, M., Holt, B. J., White, G. P., Kendall, G. E., & Holt, P. G. (2002). Association of IL12B promoter polymorphism with severity of atopic and non-atopic asthma in children. *Lancet*, *360*(9331), 455–459.

Morzycka-Wroblewska, E., & Harwood, J. (1993). Structure and evolution of the promoter regions of the DQA genes. *Immunogenetics*, *37*(5), 364–372.

Morzycka-Wroblewska, E., Munshi, A., Ostermayer, M., Harwood, J. I., & Kagnoff, M. F. (1997). Differential expression of *HLA-DQA1* alleles associated with promoter polymorphism. *Immunogenetics*, *45*(3), 163–170.

Neefjes, J., Jongsma, M. L., Paul, P., & Bakke, O. (2011). Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, *11*(12), 823–836.

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). Iq-tree: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, *32*(1), 268–274.

Nowak, K., & Lee, P. C. (2011). Demographic structure of Zanzibar red colobus populations in unprotected coral rag and mangrove forests. *International Journal of Primatology*, *32*(1), 24–45.

Nunn, C. L. (2012). Primate disease ecology in comparative and theoretical perspective. *American Journal of Primatology*, *74*(6), 497–509.

Oates, J., & Ting, N. (2015). Conservation consequences of unstable taxonomies: The case of the red colobus monkeys. In A. M. Behie & M. F. Oxenham (Eds.), *Taxonomic tapestries: The threads of evolutionary, behavioural and conservation research* (pp. 321–343). Acton: ANU Press.

Otto, W., Stadler, P. F., Lopez-Giraldez, F., Townsend, J. P., Lynch, V. J., & Wagner, G. P. (2009). Measuring transcription factor-binding site turnover: A maximum likelihood approach using phylogenies. *Genome Biology and Evolution*, *1*, 85–98.

Paradis, E. (2010). pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*, 419–420.

Park, J. K., Martin, L. J., Zhang, X., Jegga, A. G., & Benson, D. W. (2012). Genetic variants in *SCN5A* promoter are associated with arrhythmia phenotype severity in patients with heterozygous loss-of-function mutation. *Heart Rhythm: the Official Journal of the Heart Rhythm Society*, *9*(7), 1090–1096.

Paz Bettinotti, M., Kolek, A., Brünnler, G., Haas, P., Paul, C., Hochberger, M., ... Albert, E. D. (1993). Polymorphism of the 5′flanking region of the *HLA-DQA1* gene in coeliac disease. *International Journal of Immunogenetics*, *20*(5), 399–407.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A. M., ... Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, *7*(3), e1001342.

Petronzelli, F., Kimura, A., Ferrante, P., & Mazzilli, M. C. (1995). Polymorphism in the upstream regulatory region of DQA1 gene in the Italian population. *Tissue Antigens*, *45*(4), 258–263.

Poulin, R. (1995). 'Adaptive' changes in the behaviour of parasitized animals: A critical review. *International Journal of Parasitology*, *25*, 1371–1383.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Richman, A. (2000). Evolution of balanced genetic polymorphism. *Molecular Ecology*, *9*(12), 1953–1963.

Ronquist, F., & Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*, 1572–1574.

Ruiz-Lopez, M. J., Barelli, C., Rovero, F., Hodges, K., Roos, C., Peterman, W. E., & Ting, N. (2016). A novel landscape genetic approach demonstrates the effects of human disturbance on the Udzungwa red colobus monkey (*Procolobus gordonorum*). *Heredity*, *116*(2), 167–176.

Saag, M., & Deeks, S. G. (2010). How do HIV elite controllers do what they do? *Clinical Infectious Diseases*, *51*(2), 239–241.

Sibley, S. D., Lauck, M., Bailey, A. L., Hyeroba, D., Tumukunde, A., Weny, G., ... Friedrich, T. C. (2014). Discovery and characterization of distinct simian pegiviruses in three wild African old world monkey species. *PLoS One*, *9*, e98569.

Simons, N. D., Ruiz-Lopez, M. J., Chapman, C. A., Goldberg, T. L., Karl, J. A., Wiseman, R. W., ... Ting, N. (2016). Rapid identification of major histocompatibility complex class I haplotypes using deep sequencing in an endangered old world monkey. *Conservation Genetics Resources*, *8*, 23–26.

Snaith, T. V., Chapman, C. A., Rothman, J. M., & Wasserman, M. D. (2008). Bigger groups have fewer parasites and similar cortisol levels: A multi-group analysis in red colobus monkeys. *American Journal of Primatology*, *70*(11), 1072–1080.

Stephens, M., & Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics*, *76*, 449–462.

Stephenson, L. S., Holland, C. V., & Cooper, E. S. (2000). The public health significance of *Trichuris trichiura*. *Parasitology*, *121*(S1), S73–S95.

Sterner, K. N., Raaum, R. L., Zhang, Y. P., Stewart, C. B., & Disotell, T. R. (2006). Mitochondrial data support an odd-nosed colobine clade. *Molecular Phylogenetics and Evolution*, *40*(1), 1–7.

Struhsaker, T. T. (2005). Conservation of red colobus and their habitats. *International Journal of Primatology*, *26*, 525–538.

Struhsaker, T. T. (2010). *The red colobus monkeys: Variation in demography, behavior, and ecology of endangered species*. New York, NY: Oxford University Press.

Takahata, N., & Satta, Y. (1998). Footprints of intragenic recombination at HLA loci. *Immunogenetics*, *47*(6), 430–441.

Tan, Z., Shon, A. M., & Ober, C. (2005). Evidence of balancing selection at the *HLA-G* promoter region. *Human Molecular Genetics*, *14*(23), 3619–3628.

Thurber, M. I., Ghai, R. R., Hyeroba, D., Weny, G., Tumukunde, A., Chapman, C. A., ... Goldberg, T. L. (2013). Co-infection and cross-species transmission of divergent *Hepatocystis* lineages in a wild African primate community. *International Journal for Parasitology*, *43*, 613–619.

Ting, N. (2008a). *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.

Ting, N. (2008b). Mitochondrial relationships and divergence dates of the African colobines: Evidence of Miocene origins for the living colobus monkeys. *Journal of Human Evolution*, *55*, 312–325.

Ting, J. P. Y., & Trowsdale, J. (2002). Genetic control of MHC class II expression. *Cell*, *109*, S21–S33.

Tung, J., Primus, A., Bouley, A. J., Severson, T. F., Alberts, S. C., & Wray, G. A. (2009). Evolution of a malaria resistance gene in wild primates. *Nature*, *460*(7253), 388–391.

Van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory*, *1*, 1–30.

Vaumourin, E., Vourc'h, G., Gasqui, P., & Vayssier-Taussat, M. (2015). The importance of multiparasitism: Examining the consequences of co-infections for human and animal health. *Parasites & Vectors*, *8*(1), 1.

Weatherall, D. W. (2006). *The use of non-human primates in research*. London: Academy of Medical Sciences.

Wegner, K. M., Kalbe, M., & Reusch, T. B. (2007). Innate versus adaptive immunity in sticklebacks: Evidence for trade-offs from a selection experiment. *Evolutionary Ecology*, *21*(4), 473–483.

Wilson, J. N., Rockett, K., Keating, B., Jallow, M., Pinder, M., Sisay-Joof, F., ... Kwiatkowski, D. (2006). A hallmark of balancing selection is present at the promoter region of interleukin 10. *Genes and immunity*, *7*(8), 680–683.

Wittkopp, P. J., & Kalay, G. (2012). *Cis*-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, *13*, 59–69.

Wiuf, C., Zhao, K., Innan, H., & Nordborg, M. (2004). The probability and chromosomal extent of trans-specific polymorphism. *Genetics*, *168*(4), 2363–2372.

Wolfe, N. D., Escalante, A. A., Karesh, W. B., Kilbourn, A., Spielman, A., & Lal, A. A. (1998). Wild primate populations in emerging infectious disease research: The missing link? *Emerging Infectious Diseases*, *4*(2), 149.

Wray, G. A. (2007). The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics*, *8*, 206–216.

Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., ... Andersen, M. R. (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology*, *29*(8), 735–741.

Yao, Z., Kimura, A., Hartung, K., Haas, P. J., Volgger, A., Brünnler, G., ... Albert, E. D. (1993). Polymorphism of the *DQA1* promoter region (*QAP*) and *DRB1*, *QAP*, *DQA1*, *DQB1* haplotypes in systemic lupus erythematosus. *Immunogenetics*, *38*(6), 421–429.

Zuk, M., & Stoehr, A. M. (2002). Immune defense and host life history. *The American Naturalist*, *160*(4), S9–S22.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Simons ND, Eick GN, Ruiz-Lopez MJ, et al. *Cis*-regulatory evolution in a wild primate: Infection-associated genetic variation drives differential expression of *MHC-DQA1* in vitro. *Mol Ecol*. 2017;00:1–13. https://doi.org/10.1111/mec.14221