Parasite-Mediated Evolution of the Functional Part of the MHC in Primates

Parasite-mediated evolution of non-synonymous substitution rate at the functional part of the MHC in primates

László Z. Garamszegi[1]* and Charles L. Nunn[2]

[1]*Department of Evolutionary Ecology, Estacion Biologica de Donana-CSIC, Seville, Spain*

[2]*Department of Human Evolutionary Biology, Peabody Museum, Harvard University, Cambridge, USA;*

Word count: 8084

Running head: Evolution of MHC polymorphism in primates

Correspondence to LZG

Address: Department of Evolutionary Ecology, Estacion Biologica de Donana-CSIC, c/ Americo Vespucio, s/n, 41092, Seville, Spain

Tel: (+34) 954 232 340

Fax: (+34) 954 621 125

E-mail: laszlo.garamszegi@ebd.csic.es

ABSTRACT

The major histocompatibility complex (MHC) is a key model of genetic polymorphism, but the mechanisms underlying its extreme variability are debated. Most hypotheses for MHC diversity focuses on pathogen-driven selection and predict that MHC polymorphism evolves under the pressure of a diverse parasite fauna. Several studies reported that certain alleles offer protection against certain parasites, yet it remains unclear whether variation in parasite pressure more generally covaries with allelic diversity and rates of molecular evolution of MHC across species. We tested this prediction in a comparative study of 41 primate species. We characterized polymorphism of the exon 2 of DRB region of the MHC class II. Our phylogenetic analyses controlled for potential effects of population size, geographic origin and body mass and revealed that nematode species richness associates positively with non-synonymous nucleotide substitution rate at the functional part of the molecule. We failed to find evidence for allelic diversity being strongly related to parasite species richness. Continental distribution was a strong predictor of both allelic diversity and substitution rate, with higher values in Malagasy and Neotropical primates. These results indicate that parasite pressure can influence different estimates of MHC polymorphism, while geography plays an independent role in the natural history of MHC.

Keywords: balancing selection, helminths, host-parasite coevolution, immune defence, immunogenetics, parasitism, phylogenetic comparative methods.

INTRODUCTION

Host-parasite dynamics involve selection processes at the genetic level, which favour virulence genes in parasites on the one hand, and antigen recognition genes in hosts on the other hand. The major histocompatibility complex (MHC) serves as a molecular basis for immune recognition and reaction in most vertebrates (Klein & Ohuigin, 1994; Hedrick, 2002). MHC products are expressed as glycoproteins and function to bind and present antigens that trigger the appropriate immune response from T-lymphocytes (Tizard, 2002). There are two main classes of MHC that are responsive to different types of parasites (Klein, 1986). MHC class I molecules present peptides from intracellular parasites (e.g. viruses), while MHC class II molecules react to extracellular parasites (e.g. nematodes). Given this functional link between parasites and immune response mediated by MHC, this gene complex is thought to be under strong selection from parasites (e.g. Apanius *et al.*, 1997; Bernatchez & Landry, 2003; Piertney & Oliver, 2006).

MHC contains the most polymorphic set of genes among all nuclear-encoding genes, which is expressed by the number of alleles maintained and also by the differences in nucleotide sequence among alleles (Hughes & Yeager, 1998). The second exon of the *DRB* region of the MHC class II genes is a key model of this variability because it codes the functional part of the molecule that accomplishes peptide presentation (the antigen-binding sites [ABS]) (Ohta, 1998; Tizard, 2002). Contact residues of the ABS consistently exhibit a higher rate of non-synonymous than synonymous substitutions, which implies selection for changes in the amino acid sequence. Accordingly, balancing selection may preserve polymorphism at the MHC, and particularly within the ABS, which allows populations or species to maintain alleles to function against a diverse array of antigens (e.g. Apanius *et al.*, 1997; Bernatchez & Landry, 2003; Piertney & Oliver, 2006).

Three main mechanisms have been proposed to explain how balancing selection can operate on MHC polymorphism through antagonistic host-parasite relationships. The “heterozygote advantage” hypothesis posits that heterozygous individuals enjoy selective advantage over homozygous individuals because, by having two different alleles, they can combat a broader

spectrum of parasites (Doherty & Zinkernagel, 1975). The "negative frequency dependent" hypothesis presumes that rare MHC alleles incur benefits against pathogen strains that can evade common MHC alleles (Takahata & Nei, 1990). In this process, the co-evolutionary arms race between parasite antigenicity and host recognition selects for cyclic changes in the susceptible/resistant MHC alleles and thus maintains MHC alleles in flux. Finally, according to the "fluctuating selection" hypothesis, temporal and/or geographical variation in the type and prevalence of pathogens may result in fluctuations in parasite-mediated selection that can drive MHC diversity by selecting different sets of MHC alleles at different time and/or space (Hill, 1991).

These hypotheses are not mutually exclusive and have received mixed support in studies of both laboratory and wild populations (Spurgin & Richardson, 2010). Most of these studies focused on the expression of a particular allele or allele combination within a host population in relation to the prevalence of one or a few parasite species. If an allele or allele combination is found to be associated with the absence of a parasite, it is often treated as evidence for parasite-mediated balancing selection. However, the time scale associated with the inter-individual context does not capture the long-term dynamics between allele frequencies and parasite pressure acting at the population level (Apanius *et al.*, 1997; Westerdahl *et al.*, 2004; Charbonnel & Pemberton, 2005). Moreover, the heterozygote advantage hypotheses can be more powerfully tested when multiple parasites are considered because the advantage of heterozygotes is manifested in a multi-parasite context (McClelland *et al.*, 2003).

An alternative approach to investigate questions about the evolutionary role of parasites in mediating MHC polymorphism is to compare populations or species that differ in levels of disease risk. Such comparisons can be used to make inferences about the preservation of polymorphic genes at organizational levels above the individual, and to identify factors that select for the maintenance of MHC polymorphism over longer phylogenetic time scales. The relationship between MHC polymorphism and parasitism at a between-population or between-species level is highly relevant

for hypotheses about parasite-mediated balancing selection. Under strong parasite pressure, populations or species can be expected to maintain more MHC alleles because this increases the chances of individuals having rare alleles and/or heterogeneous allele-combinations, or enhances spatio-temporal variations across subpopulations. Alternatively, species may have a high rate of non-synonymous substitution rate without necessarily accumulating MHC alleles. This is because allelic diversity and substitution rate reflect different phenomenon. The former refers to the number of alleles preserved in a population irrespective to the sequential difference between these alleles (which can cover several amino acids or a just a single substitution), while the latter deals with sequence variability regardless of the number of functioning alleles on which this variability is preserved. Therefore, it might be that several alleles are maintained but with small differences, which only offer protection against a narrow spectrum of antigens. On the other hand, few alleles accumulating multiple amino acid substitutions can involve resistance to many parasite species. Consequently, pathogen-driven selection forces can favour species that have either more MHC alleles or a higher non-synonymous substitution rate or both.

Some recent studies have investigated why some populations maintain more alleles or a higher substitution rate than others (e.g. Kim *et al.*, 199; Boyce *et al.*, 1997; Landry & Bernatchez, 2001; Miller *et al.*, 2001; Schad *et al.*, 2005), with two of them focusing on the role of parasites in mediating this diversity at the across-population level (Wegner *et al.*, 2003; Prugnolle *et al.*, 2005). Less attention has been paid to interspecific patterns of MHC variation (Lehman *et al.*, 2004; Schaschl *et al.*, 2006). Two studies have assessed how species-specific selective parasite pressures shaped variation in MHC diversity across species. In a phylogenetic analysis of 14 species of cyprinid fish, Šimková and her co-workers identified an association between nucleotide diversity (sequence variability) of the exon 2 of *DAB* genes belonging to MHC class IIB and ectoparasite richness (Šimková *et al.*, 2006). Similarly, de Bellocq et al. (2008) revealed that rodent species that face a rich helminth fauna also maintain increased allelic polymorphism at the MHC class II.

To our knowledge, however, no study has investigated the evolutionary consequences of having species rich parasite fauna for allelic diversity and sequence variability simultaneously. Such distinction would be important, because as note above, different estimates of MHC polymorphism may represent qualitatively different outcomes that are relevant to different processes. A previous study in primates showed that both substitution rate and allelic diversity in the exon 2 of the primate *Mhc-DRB* gene (Figure 1) are species-specific traits, and these species-specific variations are prevalent across different lineages of primates (Garamszegi *et al.*, 2009b). This indicates that selection forces may operate on MHC traits at the species level, which prompted us to test hypotheses about MHC polymorphism using interspecific comparisons of both allelic diversity and substitution rate.

Specifically, in this paper, we investigate whether allelic variation and non-synonymous nucleotide substitution rate at the functional part of the MHC (Figure 1) covary with measures of parasitism across species. We assessed parasitism as species richness of the entire parasite fauna, but we specifically focused on the links between MHC polymorphism and nematode richness. We adopted this focus because molecules of the MHC class II generally are used to recognize extracellular parasites such as nematodes, and most within-population studies demonstrated a link between the presence of MHC alleles and nematode prevalence (e.g. Paterson *et al.*, 1998; Buitkamp *et al.*, 1999; Ditchkoff *et al.*, 2005; Meyer-Lucht & Sommer, 2005; Schad *et al.*, 2005; Tollenaere *et al.*, 2008). Accordingly, if parasites drive polymorphism at MHC genes, we predicted that nematode species richness covaries positively with allelic diversity and nucleotide substitution rate. Due to evolutionary time constraints, recent and virulent parasites are expected to involve selection forces for MHC diversity of weak magnitude (Klein & Ohuigin, 1994). Nematodes have a long-lasting a co-evolutionary history with their hosts (Sorci *et al.*, 2003; Nieberding *et al.*, 2005). Thus, we specifically focused on the effects of nematode parasite species richness rather than on the virulence of individual pathogens.

To test these predictions, we developed a novel dataset that integrates genetic and parasitological data. We controlled for the potentially confounding effects of host phylogenetic history, genetic drift, population demography, geography and life history, as these factors mediating both parasitism and MHC polymorphism can, in theory drive spurious correlations between the focal traits. In addition, we used an effect size framework for evaluating the predictions, as this enabled us to investigate the strength of the effects and the precision by which these effects can be estimated from the given sample (2-3 key references here).

MATERIAL AND METHODS

Our MHC data relies on 51 studies representing 2500 animals and 1174 sequences. The parasite data came from *The Global Mammal Parasite Database* (Nunn & Altizer, 2005), which at the time of the analyses included data from 447 studies of approximately 68,000 wild animals representing 116 primate species (total number of individuals is approximate because not all studies reported sample sizes, and sometimes multiple studies analyzed samples from the same individuals). These studies identified 629 different parasites to the species level. The actual data used were from a subset of this dataset, representing 41 primate species for which matching data on MHC were also available.

MHC Data

Details on the collection methods are given in Garamszegi et al. (2009b). These authors focused on describing interspecific variation in MHC diversity, whereas we investigated whether parasitological and other factors account for MHC diversity in primates. Briefly, we extracted information on within-lineage polymorphism of the exon 2 of *Mhc-DRB* (Figure 1) for all primates from the literature in an attempt to recover all published data through careful searches in the IPD/MHC database (http://www.ebi.ac.uk/ipd/mhc, Robinson *et al.*, 2003), Web of Science and GenBank. Sequences that are derived from common ancestry in different species, that have known gene products and peptide-binding grooves that are highly similar, and that could therefore select

the same peptide for T-cell activation, can be considered to belong to the same lineage (Geluk *et al.*, 1993). In this framework, we relied on the standard nomenclature and organization, in which *DRB#* (e.g. *DRB1*, *DRB5* or *DRBW*) labels loci that are composed of lineages in the form of *DRB#*##* (e.g. *DRB1*03, DRB3*05*, *DRB*W28* or *DRB*Wb*), and these lineages host alleles that are denoted *DRB#*####* (e.g. *DRB1*0301, DRB3*0504, DRB*W706*, or *DRB*Wb01*). Therefore, we treated human (HLA) orthologues and non-orthologues (those with a 'W' workshop number) in the same way, for which justification is given in a previous analysis (Garamszegi *et al.*, 2009b). As our study focused on polymorphism at the within-lineage level, we gathered information on the number of alleles detected in each lineage in each species. The number of animals sampled also was recorded. We then imported and aligned the exon 2 nucleotide sequences in the program MEGA (Kumar *et al.*, 2008), and estimated non-synonymous (d_N) substitution rate at the contact residues of the ABS for each lineage after excluding pseudogenes. Sequences with codon or nucleotide insertions or deletions, with premature stop codons together with alleles from the DRB6 locus were considered as pseudogenes, as these may code non-functional proteins (Hughes, 1995). We treated the following 16 ABS contact residues to be relevant: 9, 11, 13, 28, 30, 37, 38, 57, 61, 67, 70, 71, 74, 78, 82, 86 (Brown *et al.*, 1988; 1993; see also Figure 1). The aligned sequences can be found in Garamszegi et al. (2009b), while the updated IPD/MHC database lists the corresponding GenBank accession numbers.

For the estimation of substitution rates, we used the Nei and Gojobori (1986) method with the Jukes and Cantor (1969) correction for multiple hits, which computes the number of non-synonymous differences (d_N) between each pair of sequences after normalizing for the potential number of non-synonymous sites and by correcting for multiple substitutions. For each group of lineages within species, an arithmetic average of d_N was computed for all possible pair-wise comparisons of sequences. We repeated this process for synonymous mutation rate (d_S). The corrected Nei and Gojobori approach is the most common way to estimate substitution rate (e.g. Hedrick *et al.*, 2001; Harf & Sommer, 2005; Schad *et al.*, 2005; Abbott *et al.*, 2006; Huchard *et al.*,

2006; Schwensow *et al.*, 2007). When we used other methods, such as the Li-Wu-Luo method (Li *et al.*, 1985) or Kumar's method (Nei & Kumar, 2000, page 64), we obtained values that correlated very highly with our estimates of substitution rate ($r > 0.997$, $P < 0.001$, $N = 19$ species, when estimating substitution rate for the *DRB1*03* lineage using different methods).

We refer to non-synonymous substitution rate at the ABS simply as d_N-ABS. d_N-ABS strongly correlates with different substitution rates at different sites of exon 2 (Garamszegi *et al.*, 2009b), and this estimate generally serves as a basis for tests of positive selection on the MHC, which assess the frequency of non-synonymous substitutions relative to the frequency of synonymous substitutions (d_S-ABS). We avoided calculating the d_N:d_S ratio for our phylogenetic analyses. Correlations with ratios may be difficult to interpret because a given pattern may arise from the effect of the numerator, the denominator or the combination of the two (Sokal & Rohlf, 1995). However, at the statistical level, we also tested for biological effects acting on d_N-ABS, while the neutral mutation rate (d_S-ABS) was held constant by including it in our statistical models (see below).

We created a dataset at the level of lineage, which we further processed for the comparative tests that were based on species-specific values. For each lineage, we entered the number of alleles detected and substitution rate as calculated above. Then, we built a General Linear Mixed Model (GLMM) to deal with species and lineage effects simultaneously, as both substitution rate and allele number vary across species and lineages. For our comparative tests of interspecific associations, we were interested in expressing MHC polymorphism as a species-specific trait while lineage-specific effects were held constant. To achieve this, we used d_N-ABS as the dependent variable, while entering species as a main factor and lineage as a random factor. We applied square-root-arcsine transformation on d_N-ABS in order to achieve normally distributed rate data (Sokal & Rohlf, 1995). After estimating species effects in this model (Garamszegi *et al.*, 2009b), we calculated species-specific values of substitution rate in the form of Least Square (LS) means that are thus independent of lineage-specific effects. Note that LS means are associated with quantifiable error, and these

errors can be accounted for in the level of analyses as measurement error (see below).

To obtain species-specific estimates of allelic variation, we adopted a similar modelling philosophy. Accordingly, we constructed a model with number of alleles ($\log_{10}$-transformed for normality) as a dependent variable and species and lineage as factors (main and random, respectively). As above, we obtained LS means (and their errors) of allele counts for each species, which were thus independent of lineages. These species-specific allele counts were then corrected for sample size, as more alleles are discovered when more individuals are screened. This correction was achieved in a linear regression on LS means of allele counts with the $\log_{10}$-transformed number of individuals as the independent variable, from which we derived the residuals to reflect relative allele number, i.e. allelic variation. We used species-specific estimates of substitution rate and allelic variation in the subsequent phylogenetic analyses to test for the determinants of MHC polymorphism at the interspecific level.

Originally, we adopted the most recent primate taxonomy for the species assignment (Wilson & Reeder, 1993), but for correspondence with the parasite data, we followed the species concept that is followed in the *Global Mammal Parasite Database* (Nunn & Altizer, 2005). Thus, we considered the *Papio* group as different species (*P. cynocephalus, P. hamadryas,* and *P. ursinus*); *Aotus nancymaae* and *A. nigriceps* as *A. azarai*; *A. trivirgatus* and *A. vociferans* as *A. trivirgatus*; and *Microcebus myoxinus* as *M. rufus*. In these cases, we categorized alleles according to this species scenario (by removing identical alleles if necessary), and calculated allele counts and substitution rates accordingly. MHC traits for *Lepilemur leucopus* were treated at the genus-level and thus matched with parasite richness calculated for *Lepilemur.*

Parasite species richness

The diversity of parasite communities, measured here as parasite richness, may provide a reliable estimate of the evolutionary impact of parasites on host species (Poulin, 1995). While individual parasites select for qualitative defence, parasite species richness is more likely to favour

quantitative defence, as hosts that are exposed to several parasites should have a sophisticated self/non-self recognition system, which is the main function of MHC. *The Global Mammal Parasite Database* is the most comprehensive collection of published records of parasitic organisms from free-living primates (Nunn & Altizer, 2005). From this resource, we counted the total number of parasite species found in each host species. Parasite richness data is informative only if research effort is controlled because, as noted for MHC allelic diversity, the number of parasites discovered is a positive function of the number of hosts studied (Walther *et al.*, 1995). As the primary measure of research effort, we followed previous studies by using the number of citations for each host species (and common taxonomic variants; Nunn et al. 2003, 2004), in our case from an online database, PrimateLit (http:// primatelit.library.wisc.edu/). We then calculated residuals from the $\log_{10}$-transformed species richness/research effort regression line to derive relative parasite species richness. We performed this procedure by focusing separately on nematode parasites, which as noted above are one of the most relevant parasites for the MHC class II (e.g. Paterson *et al.*, 1998; Buitkamp *et al.*, 1999; Ditchkoff *et al.*, 2005; Meyer-Lucht & Sommer, 2005; Schad *et al.*, 2005; Tollenaere *et al.*, 2008). Other measures of sampling effort are available, including quantifying the number of individuals sampled. We preferred using citation counts rather than animals sampled because many of the original studies failed to provide sample sizes or gave the number of samples collected (rather than the number of individuals). In addition, some studies of primate parasites focused on intensive sampling for singular zoonotic parasites and pathogens, resulting in huge sample sizes that fail to capture the number or types of parasite species that were screened by each study (see also Nunn et al. 2003).

Confounding variables

We controlled for several potentially confounding variables. First, given that the strength of positive selection on a background purifying selection can be assessed by the non-synonymous substitution rate relative to synonymous substitution rate (Bernatchez & Landry, 2003; Piertney &

Oliver, 2006), the comparison of non-synonymous substitutions might be more informative when silent mutation rates are included as a covariate in the statistical model. To achieve this normalization, we included d_S-ABS in the statistical analyses, in which d_N-ABS was used as the dependent variable. Although such correction is warranted theoretically, in practice we expect it will have minor effects on the results because d_N:d_S ratios were larger than 1 for most of the DRB lineages in primates, and thus show unambiguous evidence for selection (Suarez *et al.*, 2006; Garamszegi *et al.*, 2009b).

Second, we statistically controlled for population size as a surrogate of effective population size because genetic drift is one of the key determinants of the total number of alleles segregating in a population; thus, allelic richness should be a function of effective population size (Hedrick, 1985). Because all populations are finite, genetic variability will be eroded with time, resulting in larger populations maintaining higher levels of genetic variation than smaller populations. Similarly, host population size is an important epidemiological determinant of parasite population growth via density dependent constraints, and thus may also affect host-parasite interactions (Anderson & May, 1978). Moreover, larger host populations may represent larger "islands" for pathogens and thus should support a greater number of susceptible individuals (Nunn *et al.*, 2003; Hughes & Page, 2007). Unfortunately, effective population size data based on genetic data are unavailable for the majority of species in our dataset. We therefore estimated observed population size as density (individuals/km^2) * distribution area (km^2) (see Møller *et al.*, 2008 for relevance) using data from Nunn et al. (2003) and Harcourt et al. (2005). However, from the literature (Yu *et al.*, 2004; Won & Hey, 2005; Stevison & Kohn, 2009; Wlasiuk & Nachman, 2010), we could locate effective population size data for six species in our dataset that showed a suggestive positive correlation with our estimate ($r = 0.795$, 95% CI = -0.047 to 0.977, $N = 6$, $P = 0.059$). This indicated that our surrogate measure is reliable.

Third, we controlled for geography in terms of the geographic location of the different primate species. As compared to other primates, Malagasy primates (lemurs) have higher densities,

smaller distribution ranges and smaller body sizes (Harcourt *et al.*, 2005), all of which might influence MHC population genetics. Moreover, some primate radiations, including those in the Neotropics and Madagascar, originated from small bottlenecked founding populations harbouring few ancestral DRB genes (Go *et al.*, 2002). In addition, the evolution of MHC traits likely followed specific directions in different continents after colonization, as the size of the founding populations and subsequent selection patterns were different (Trtkova *et al.*, 1995; Antunes *et al.*, 1998; Suarez *et al.*, 2006). To control for these effects, we discriminated species from four realms (Madagascar, Africa, Asia, America) and included these codes as a covariate in the analyses.

Finally, body mass reflects a suite of fundamental life history and demographic parameters that can affect both parasite species richness and MHC polymorphism (see e.g. O'Brien & Evermann, 1988; Finch & Rose, 1995; Lochmiller, 1996; Clayton & Walther, 2001; Poulin & Morand, 2004; Vitone *et al.*, 2004). Hence, we obtained body mass data from Smith and Jungers (1997). Log_{10}-transformed body mass was strongly associated with geographic origin ($F_{3, 41} = 5.935$, $P = 0.002$) and with log_{10}-transformed population size ($F_{1, 36} = 10.448$, $P = 0.003$). Consequently, including them together as independent variables in the same regression model would induce collinearity problems. To avoid this problem, we calculated residual body mass from a model with geographic origin and log_{10}-transformed population size as predictors. These residuals were used in the phylogenetic models with multiple predictors to reflect life history that is independent of the geographic and demographic effects we covered in this study.

The full comparative dataset is provided in the electronic supplementary material (ESM).

Comparative analyses

Closely related species may share ecological, molecular and life history traits relevant to the predictions of interest here. Therefore, approaches are needed that examine phylogenetic signal and control for similarity in MHC characteristics among species due to common descent. To do this, we applied phylogenetic generalized least squares (PGLS) models (Martins & Hansen, 1997; Pagel,

1999), which incorporate a matrix of the expected covariances among species based on likelihood ratio statistics. This method enabled us to estimate the importance of phylogenetic corrections by calculating the phylogenetic scaling parameter lambda (λ), which varies between 0 (phylogenetic independence) and 1 (trait evolution corresponds to a Brownian motion model under the given branch lengths) (Freckleton *et al.*, 2002). We conducted analyses using the maximum likelihood estimate of λ; thus, we corrected for phylogenetic effects as much as the data required. We also investigated more complex models that incorporated a weighting factor for the number of individuals studied and considered the errors associated with the species-specific estimates of MHC polymorphism (Garamszegi & Møller, 2007). However, the incorporation of these error terms due to heterogeneity in sample size did not improve the maximum likelihood of the models, indicating that such adjustments are not needed. Therefore, we present results based on unweighted models.

The PGLS analyses were performed in the R statistical environment (R Development Core Team, 2007), with additional unpublished phylogenetic functions provided by R. Freckleton (University of Sheffield). The phylogeny used to calculate V originated from the consensus tree of Version 1 of the 10kTrees Project (http://10ktrees.fas.harvard.edu/), which provides a Bayesian inference of primate phylogeny (Arnold et al. 2010). While it may be interesting to also incorporate parasite phylogeny into the analysis, sufficient data on parasite phylogeny is currently unavailable, as are methods for incorporating both host and parasite phylogenies in a single comparative analysis. We did, however, examine variation in parasites based on parasite taxonomy. Most species in our nematode parasite counts come from different genera, resulting in a strong association between species and genus diversity ($r=0.98$, $N=35$).

Although we attempted to process all potentially available data, sample size remains modest in the interspecific context, which has statistical consequences in terms of statistical power and the precision of estimates. When applying a null hypothesis testing framework, insufficient power can increase the risk of type II errors (i.e., failing to reject the null hypothesis when it is false). This problem becomes robust when significance levels are adjusted in order to balance type I errors (i.e.,

rejecting the null hypothesis when it is true) due to multiple testing. To avoid such errors, we followed recent statistical recommendations that shift the focus from significance levels to effect sizes, as null-hypothesis testing at low statistical power would likely dismiss moderate effects with evolutionary importance (Nakagawa & Cuthill, 2007; Garamszegi *et al.*, 2009a). Accordingly, we present effect sizes (such as correlation effect size "r" sensu Cohen, 1988) and the associated 95% confidence intervals (95 % CI). Our interpretations are based on the strength of biological effects and the precision with which they can be measured from the available data. We provide P values for illustrative purposes but avoid emphasizing the statistical significance of the results.

RESULTS

First, we tested for the relationship between overall parasite species richness and MHC-DRB polymorphism in terms of allelic diversity and non-synonymous substitution rate at the antigen-binding sites of the molecule (d_N-ABS). The phylogenetic models generally failed to detect strong relationships between species richness and these MHC traits (allelic diversity: $\lambda = 0.586$, $r = -0.012$, 95% CI = -0.318 to 0.297, N = 41, P = 0.936; d_N-ABS: $\lambda = 0.891$, $r = 0.143$, 95% CI = -0.190 to 0.446, N = 37, P = 0.403). We repeated these analyses for nematode species richness and found that the estimated range of the effect size for the relationship between d_N-ABS and nematode diversity mostly covers a positive association (allelic diversity: $\lambda = 0.810$, $r = 0.060$, 95% CI = -0.279 to 0.386, N = 35, P = 0.732; d_N-ABS: $\lambda = 0.852$, $r = 0.294$, 95% CI = -0.067 to 0.587, N = 31, P = 0.109, Figure 2A).

We then developed a multi-predictor phylogenetic model that controlled for d_S-ABS, population size, geographic range and body mass (d_S-ABS was considered in the analyses of substitution rates only). Accordingly, we entered these confounding variables as covariates in addition to the focal variables into the phylogenetic model, and assessed if these factors had any effect on the strength of the relationship between parasite burden and estimates of MHC polymorphism. In this multi-predictor approach, we found that the positive association between nematode species richness and d_N-ABS remained, with the effect size covering a 95% confidence

range that excluded zero (Table 1, Figure 2B). The models also revealed that both allelic variation and d_N-ABS varied systematically among the four major geographic areas in which primates are found (Table 1). Specifically, we found that primates from Madagascar and from the New World had a higher degree of MHC polymorphism than primates from Asia and Africa (Figure 3).

DISCUSSION

Given the limited sample size, the 95% CIs of the estimated effects show that there is a considerable uncertainty around our effect size estimates; thus, it is premature to make inferences about the strength of the effects. Yet, our study stands on the largest sample size available so far that tested for the interspecific relationship between parasite burden and MHC polymorphism. Summarizing our results within the effect size statistical framework that we adopted, the relationship between d_N-ABS and nematode burden can be weak or strong, but at least the currently available data show that when confounds are held constant it is highly likely to be positive. On the other hand, we can be certain that the relationship for allelic diversity is unlikely to be strong, but based on the current data we should retain the possibility that it can go weakly in the positive or negative directions or even be of zero magnitude. Concerning the effect of geography, most of the 95% CI ranges exceed $r = 0.3$, which would suggest a strong effect for the geographic variation in MHC polymorphism. We interpret our results in the light of these effect sizes (Nakagawa & Cuthill, 2007).

Our results support the hypothesis that higher diversity of nematodes favours higher nucleotide substitution rates in order to maintain different alleles at varying frequency (Takahata & Nei, 1990) or in heterozygote combination (Doherty & Zinkernagel, 1975), or to allow temporal/spatial variations in relation to fluctuating pathogen regime (Hill, 1991). Therefore, our results corroborate evolutionary theories of MHC polymorphism based on host-parasite dynamics, but our correlative findings do not allow us to discriminate among these three possible mechanisms. Exposure to many different parasites may select for increased antigen recognition, which is achieved by higher rates of non-synonymous nucleotide substitution at the functional part of the

molecule (Ohta, 1991; Yeager & Hughes, 1999). If the evolutionary arms race between hosts and parasites affects substitution rate at the MHC, our results further suggest that this effect has minor, if any, consequences for the number of alleles maintained in a species. This would suggest that nematode-driven host-parasite dynamics promote substitutions, but even if it leads to the emergence of new alleles, these are not necessarily preserved over evolutionary time scales. Accordingly, directional selection in an evolutionary arms race would entail selection of weak magnitude for allele and antigen diversity but would favour rapid evolution at non-synonymous sites. This might be effective, for example, if the coevolutionary dynamics select for novel antigens in the nematode species. Accordingly, an allele that once provided resistance against a certain antigen might not be worth conserving because it soon becomes non-protective against a more rapidly evolving pathogen. Host immunogenetics therefore plays an important role in the co-evolutionary process between hosts and parasites.

The relationship between parasites and MHC polymorphism does not generally apply to all MHC and parasite traits, as it specifically applies to non-synonymous substitution rate and nematode parasite burden. Our analyses failed to detect a comparable relationship between MHC allelic variation and overall parasite species richness. Therefore, it remains an open question as to whether other parasite traits are relevant for the accumulation of MHC alleles, or whether allelic variation itself is simply unresponsive to parasite-mediated selective pressures. It remains possible, for example, that it is not the number of parasites that primarily favours greater numbers of MHC alleles, but rather the presence of specific pathogens and the harm they cause. On the other hand, allelic variation may be mediated by factors other than parasites, such as mate choice for particular genotypes or gene combinations (Penn & Potts, 1999) or maternal-foetal interactions required for proper implantation (Apanius *et al.*, 1997). It is also plausible that a complex association exists between effective population size, allelic diversity and parasitism, as unobservable bottlenecks and range expansions might have occurred under parasite pressure having unpredictable consequences for the preservation of parasite resistance genes (Bonhomme *et al.*, 2007). In general, multiple

selective forces may operate on MHC polymorphism, which may mask any effect of parasite species richness on allelic diversity (Spurgin & Richardson, 2010). These alternative hypotheses require further investigation, and our study provides a blueprint for how such analyses could be conducted in primates and other groups of organisms.

We considered some factors that are likely to shape MHC polymorphism due to demographic, geographic and life-history effects. Our multi-predictor phylogenetic modelling (Table 1) revealed that both allelic diversity and d_N-ABS vary across four major geographic regions independently of parasite-related and other traits. Such large-scale geographic variation in MHC characteristics is well known in humans, which can be explained by differences in the evolutionary history of human populations (Blanco-Gelaz *et al.*, 2001; Gibert & Sanchez-Mazas, 2003; Solberg *et al.*, 2008). The primate data at the interspecific level show that species from Madagascar (lemurs) and South America (monkeys) harbour the most variable set of MHC lineages (Figure 2). Lemurs and New World monkeys likely originated from small founder populations (Trtkova *et al.*, 1995; Go *et al.*, 2002). As a result, the present-day diversity arose from severe bottlenecks, and now includes extensive allelic diversification relative to lineage diversification. This would cause high sequence variation within the few remaining lineages and is consistent with higher rates of diversification after the bottleneck (Go et al. 2002). In contrast, the radiation of Old World monkeys occurred from large populations, and subsequent evolution of the MHC can be typified by a conserved polymorphism at the within-lineage level (Satta, 2001). The relationship between continental distribution and MHC polymorphism is in accordance with the "fluctuating selection" hypothesis at the global scale, as it predicts spatial variation in MHC diversity due to different subsets of alleles being selected in different localities as mediated by the local selection forces caused by parasites (Hill, 1991). However, the degree by which continent-specific parasite pressure mediates MHC polymorphism to vary across realms awaits further investigations.

In summary, our analyses offer new insights to the evolutionary origins of MHC diversity. While most of the previous studies have demonstrated links between specific parasites and

particular MHC alleles, here we integrated the most comprehensive data currently available on MHC genetics and primate parasite diversity to investigate evolutionary hypotheses at a broader, interspecific scale. These analyses revealed two new findings, namely that nematode diversity covaries with substitution rates across primates, and that MHC polymorphism varies among the four major biogeographic areas inhabited by primates. Importantly, the analyses further demonstrate that comparative studies of MHC diversity can productively test hypotheses about MHC ecology and evolution. Hence, these analyses open the doors for future comparative investigations of MHC diversity and evolution, including tests of hypotheses involving mate choice, demography and environmental predictors of disease risk.

ACKNOWLEDGEMENT

L. J. Matthews and three anonymous referees provided constructive criticism. We are grateful to L. Wolff for her help in the literature search and C. Brezine for assistance with queries related to the parasite database. During this study, LZG was supported by a grant from the Research Foundation, Flanders (FWO, Belgium) and subsequently, from the Spanish National Research Council (CSIC, Spain). CLN was supported by the Max Planck Society and Harvard University.

REFERENCES

Abbott, K.M., Wickings, E.J. & Knapp, L.A. 2006. High levels of diversity characterize mandrill (Mandrillus sphinx) Mhc-DRB sequences. *Immunogenetics* **58:** 628-640.

Anderson, R.M. & May, R.M. 1978. Regulation and stability of host-parasite population interactions. 1. Regulatory processes. *Journal of Animal Ecology* **47:** 219-247.

Antunes, S.G., de Groot, N.G., Brok, H., Doxiadis, G., Menezes, A.A.L., Otting, N. & Bontrop, R.E. 1998. The common marmoset: A new world primate species with limited Mhc class II variability. *Proceedings of the National Academy of Sciences USA* **95:** 11745-11750.

Apanius, V., Penn, D., Slev, P.R., Ruff, L.R. & Potts, W.K. 1997. The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology* **17:** 179-224.

Bernatchez, L. & Landry, C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *Journal of Evolutionary Biology* **16:** 363-377.

Blanco-Gelaz, M.A., Lopez-Vazquez, A., Garcia-Fernandez, S., Martinez-Borra, J., Gonzalez, S. & Lopez-Larrea, C. 2001. Genetic variability, molecular evolution, and geographic diversity of HLA-B27. *Human Immunology* **62:** 1042-1050.

Bonhomme, M., Blancher, A., Jalil, M.F. & Crouau-Roy, B. 2007. Factors shaping genetic variation in the MHC of natural non-human primate populations. *Tissue Antigens* **70:** 398-411.

Boyce, W.M., Hedrick, P.W., Muggli-Cockett, N.E., Kalinowski, S., Penedo, M.C.T. & Ramey-II, R.R. 1997. Genetic variation of Major Histocompatibility Complex and microsatellite loci: a comparison in bighorn sheep. *Genetics* **145:** 421-433.

Brown, J.H., Jardetzky, T., Saper, M.A., Samraoui, B., Bjorkman, P.J. & Wiley, D.C. 1988. A hypothetical model of the foreign antigen-binding site of class-II histocompatibility molecules. *Nature* **332:** 845-850.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. & Wiley, D.C. 1993. 3-Dimensional structure of the human class-II histocompatibility antigen HLA-DR1. *Nature* **364:** 33-39.

Buitkamp, J., Feichtlbauer-Huber, P. & Stear, M.J. 1999. Association between MHC class II alleles and resistance to the parasitic nematode Ostertagia circumcincta. *Archiv Fur Tierzucht-Archives of Animal Breeding* **42:** 11-21.

Charbonnel, N. & Pemberton, J. 2005. A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. *Heredity* **95:** 377-388.

Clayton, D.H. & Walther, B.A. 2001. Influence of host ecology and morphology on the diversity of Neotropical bird lice. *Oikos* **94:** 455-467.

Cohen, J. 1988. *Statistical power analysis for the behavioural sciences*, 2 edn. Lawrence Erlbaum Associates, Hillsdale.

de Bellocq, J.G., Charbonnel, N. & Morand, S. 2008. Coevolutionary relationship between helminth diversity and MHC class II polymorphism in rodents. *Journal of Evolutionary Biology* **21:** 1144-1150.

Ditchkoff, S.S., Hoofer, S.R., Lochmiller, R.L., Masters, R.E. & Van Den Bussche, R.A. 2005. MHC-DRB evolution provides insight into parasite resistance in white-tailed deer. *Southwestern Naturalist* **50:** 57-64.

Doherty, P.C. & Zinkernagel, R.M. 1975. Enhanced immunological surveillance in mice heterozygous at h-2 gene complex. *Nature* **256:** 50-52.

Finch, C.E. & Rose, M.R. 1995. Hormones and the physiological architecture of life history evolution. *Quarterly Review of Biology* **70:** 1-52.

Freckleton, R.P., Harvey, P.H. & Pagel, M. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist* **160:** 712-726.

Garamszegi, L.Z. & Møller, A.P. 2007. Prevalence of avian influenza and host ecology. *Proceedings of the Royal Society of London, Series B* **274:** 2003-2012.

Garamszegi, L.Z., Calhim, S., Dochtermann, N., Hegyi, G., Hurd, P.L., Jørgensen, C., Kutsukake, N., Lajeunesse, M.J., Pollard, K.A., Schielzeth, H., Symonds, M.R.E. & Nakagawa, S. 2009a. Changing philosophies and tools for statistical inferences in behavioral ecology. *Behavioral Ecology* **20:** 1363-1375.

Garamszegi, L.Z., de Groot, N.G. & Bontrop, R.E. 2009b. Correlated evolution of nucleotide substitution rates and allelic variation in *Mhc-DRB* lineages of primates. *BMC Evolutionary Biology* **9:** 73.

Geluk, A., Elferink, D.G., Slierendregt, B.L., van Meijgaarden, K.E., de Vries, R.R., Ottenhoff, T.H. & Bontrop, R.E. 1993. Evolutionary conservation of major histocompatibility complex- DR/peptide/T cell interactions in primates. *Journal of Experimental Medicine* **177:** 979-987.

Gibert, M. & Sanchez-Mazas, A. 2003. Geographic patterns of functional categories of HLA-DRB1 alleles: a new approach to analyse associations between HLA-DRB1 and disease. *European Journal of Immunogenetics* **30:** 361-374.

Go, Y., Satta, Y., Kawamoto, Y., Rakotoarisoa, G., Randrianjafy, A., Koyama, N. & Hirai, H. 2002. Mhc-DRB genes evolution in lemurs. *Immunogenetics* **54:** 403-417.

543 Harcourt, A.H., Coppeto, S.A. & Parks, S.A. 2005. The distribution-abundance (density)
544 relationship: its form and causes in a tropical mammal order, Primates. *Journal of*
545 *Biogeography* **32:** 565-579.
546 Harf, R. & Sommer, S. 2005. Association between major histocompatibility complex class II DRB
547 alleles and parasite load in the hairy-footed gerbil, Gerbillurus paeba, in the southern
548 Kalahari. *Molecular Ecology* **14:** 85-91.
549 Hedrick, P.W. 1985. *Genetics of populations*. Jones and Bartlett Publishers, Boston.
550 Hedrick, P.W., Parker, K.M. & Lee, R.N. 2001. Using microsatellite and MHC variation to identify
551 species, ESUs, and MUs in the endangered Sonoran topminnow. *Molecular Ecology* **10:**
552 1399-1412.
553 Hedrick, P.W. 2002. Pathogen resistance and genetic variation at MHC loci. *Evolution* **56:** 1902-
554 1908.
555 Hill, A.V.S. 1991. HLA associations with malaria in Africa: some implications for MHC evolution.
556 In: *Molecular evolution of the major histocompatibility complex* (J. Klein & D. Klein, eds),
557 pp. 403-419. Springer, Berlin.
558 Huchard, E., Cowlishaw, G., Raymond, M., Weill, M. & Knapp, L.A. 2006. Molecular study of
559 MHC-DRB in wild chacma baboons reveals high variability and evidence for trans-species
560 inheritance. *Immunogenetics* **58:** 805-816.
561 Hughes, A.L. 1995. Origin and evolution of HLA class-I pseudogenes. *Molecular Biology and*
562 *Evolution* **12:** 247-258.
563 Hughes, A.L. & Yeager, M. 1998. Natural selection at major histocompatibility complex loci of
564 vertebrates. *Annual Review of Genetics* **32:** 415-435.
565 Hughes, J. & Page, R.D.M. 2007. Comparative tests of ectoparasite species richness in seabirds.
566 *BMC Evolutionary Biology* **7:** 227.
567 Jukes, T.H. & Cantor, C.R. 1969. Eolution of protein molecules in Mammalian protein metabolism.
568 In: *Mammalian protein metabolsism* (H. N. Munro, ed., pp. 21-132. Academic Press, New
569 York.
570 Kim, T.J., Parker, K.M. & Hedrick, P.W. 199. Major histocompatibility complex differentiation in
571 Sacramento river chinook salmon. *Genetics* **151:** 1115-1122.
572 Klein, J. 1986. *Natural history of the major histocompatibility complex*. Wiley & Sons, New York.
573 Klein, J. & Ohuigin, C. 1994. MHC polymorphism and parasites. *Philosophical Transactions of the*
574 *Royal Society of London, Series B* **346:** 351-357.
575 Kumar, S., Dudley, J.P., Nei, M. & Tamura, K. 2008. MEGA: A biologist-centric software for
576 evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* **9:** 299-
577 306.
578 Landry, C. & Bernatchez, L. 2001. patibility complex and microsatellite loci in Atlantic salmon
579 (*Salmo salar*). *Molecular Ecology* **10:** 2525-2539.
580 Lehman, N., Decker, D.J. & Stewart, B.S. 2004. Divergent patterns of variation in major
581 histocompatibility complex class II alleles among antarctic phocid pinnipeds. *Journal of*
582 *Mammalogy* **85:** 1215-1224.
583 Li, W.-H., Wu, C.-I. & Luo, C.-C. 1985. A new method for estimating synonymous and
584 nonsynonymous rates of nucleotide substitution considering the relative likelihood of
585 nucleotide and codon changes. *Molecular Biology and Evolution* **2:** 150-174.
586 Lochmiller, R.L. 1996. Immunocompetence and animal population regulation. *Oikos* **76:** 594-602.
587 Martins, E.P. & Hansen, T.F. 1997. Phylogenies and the comparative method: a general approach to
588 incorporating phylogenetic information into the analysis of interspecific data. *The American*
589 *Naturalist* **149:** 646-667.
590 McClelland, E.E., Penn, D.J. & Potts, W.K. 2003. Major histocompatibility complex heterozygote
591 superiority during coinfection. *Infection and Immunity* **71:** 2079-2086.
592 Meyer-Lucht, Y. & Sommer, S. 2005. MHC diversity and the association to nematode parasitism in
593 the yellow-necked mouse (*Apodemus flavicollis*). *Molecular Ecology* **14:** 2233-2243.

Miller, K.M., Kaukinen, K.H., Beacham, T.D. & Withler, R.E. 2001. Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. *Genetica* **111:** 237-257.

Møller, A.P., Garamszegi, L.Z. & Spottiswoode, C. 2008. Genetic similarity, breeding distribution range and sexual selection. *Journal of Evolutionary Biology* **21:** 213-225.

Nakagawa, S. & Cuthill, I.C. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* **82:** 591-605.

Nei, M. & Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3:** 418-426.

Nei, M. & Kumar, S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York.

Nieberding, C., Libois, R., Douady, C.J., Morand, S. & Michaux, J.R. 2005. Phylogeography of a nematode (Heligmosomoides polygyrus) in the western Palearctic region: persistence of northern cryptic populations during ice ages? . *Molecular Ecology* **14:** 765-779.

Nunn, C.L., Altizer, S., Jones, K.E. & Sechrest, W. 2003. Comparative test of parasite species richness in primates. *The American Naturalist* **162:** 597-614.

Nunn, C.L. & Altizer, S. 2005. The global mammal parasite database: An online resource for infectious disease records in wild primates. *Evolutionary Anthropology* **14:** 1-2.

O'Brien, S.J. & Evermann, J.F. 1988. Interactive influence of infectious disease and genetic diversity in natural populations. *Trends in Ecology & Evolution* **3:** 254-259.

Ohta, T. 1991. Role of diversifying selection and gene conversion in evolution of Major Histocompatibility Complex loci. *Proceedings of the National Academy of Sciences of the United States of America* **15:** 6716-6720.

Ohta, T. 1998. On the pattern of polymorphisms at major histocompatibility complex loci. *Journal of Molecular Evolution* **46:** 633-638.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401:** 877-884.

Paterson, S., Wilson, K. & Pemberton, J.M. 1998. Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (Ovis aries L.). *Proceedings of the National Academy of Sciences of the United States of America* **95:** 3714-3719.

Penn, D.J. & Potts, W.K. 1999. The evolution of mating preferences and major histocompatibility complex. *The American Naturalist* **153:** 145-164.

Piertney, S.B. & Oliver, M.K. 2006. The evolutionary ecology of the major histocompatibility complex. *Heredity* **96:** 7-21.

Poulin, R. 1995. Phylogeny, ecology, and the richness of parasite communities in vertebrates. *Ecological Monographs* **65:** 283-302.

Poulin, R. & Morand, S. 2004. *Parasite biodiversity*. Smithsonian Institution Press, Washington DC.

Prugnolle, F., Manica, A., Charpentier, M., Guegan, J.F., Guernier, V. & Balloux, F. 2005. Pathogen-driven selection and worldwide HLA class I diversity. *Current Biology* **15:** 1022-1027.

R Development Core Team 2007. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robinson, J., Waller, M.J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L.J., Stoehr, P. & Marsh, S.G.E. 2003. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Research* **31:** 311-314.

Satta, Y. 2001. Comparison of DNA and protein polymorphisms between humans and chimpanzees. *Genes and Genetic Systems* **76:** 159-168.

Schad, J., Ganzhorn, J.U. & Sommer, S. 2005. Parasite burden and constitution of major histocompatibility complex in the malagasy mouse lemur, *Microcebus murinus*. *Evolution* **59:** 439-450.

Schaschl, H., Wandeler, P., Suchentrunk, F., Obexer-Ruff, G. & Goodman, S.J. 2006. Selection and recombination drive the evolution of MHC class II DRB diversity in ungulates. *Heredity* **97:** 427-437.

Schwensow, N., Fietz, J., Dausmann, K.H. & Sommer, S. 2007. Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity* **99:** 265-277.

Šimková, A., Ottova, E. & Morand, S. 2006. MHC variability, life-traits and parasite diversity of European cyprinid fish. *Evolutionary Ecology* **20:** 465-477.

Smith, R.J. & Jungers, W.L. 1997. Body mass in comparative primatology. *Journal of Human Evolution* **32:** 523-559.

Sokal, R.R. & Rohlf, F.J. 1995. *Biometry*, 3rd edn. W. H. Freeman & Co., New York.

Solberg, O.D., Mack, S.J., Lancastera, A.K., Single, R.M., Tsai, Y., Sanchez-Mazas, A. & Thomson, G. 2008. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies. *Human Immunology* **69:** 443-464.

Sorci, G., Skarstein, F., Morand, S. & Hugot, J.-P. 2003. Correlated evolution between host immunity and parasite life histories in primates and oxyurid parasites. *Procedings of the Royal Society of London, Series B* **270.:** 2481–2484.

Spurgin, L.G. & Richardson, D.S. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society of London Series B-Biological Sciences* **277:** 979-988.

Stevison, L.S. & Kohn, M.H. 2009. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Molecular Ecology* **18:** 2457-2475.

Suarez, C.F., Patarroyo, M.E., Trujillo, E., Estupinan, M., Baquero, J.E., Parra, C. & Rodriguez, R. 2006. Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages. *Immunogenetics* **58:** 542-558.

Takahata, N. & Nei, M. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of Major Histocompatibility Complex loci. *Genetics* **124:** 967-978.

Tizard, I.R. 2002. *Veterinary Immunology: An introduction*. W. B. Saunders, Philadelphia.

Tollenaere, C., Bryja, J., Galan, M., Cadet, P., Deter, J., Chaval, Y., Berthier, K., Salvador, A.R., Voutilainen, L., Laakkonen, J., Henttonen, H., Cosson, J.F. & Charbonnel, N. 2008. Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. *Journal of Evolutionary Biology* **21:** 1307-1320.

Trtkova, K., Mayer, W.E., O'hUigin, C. & Klein, J. 1995. Mhc-DRB genes and the origin of new world monkeys. *Molecular Phylogenetics and Evolution* **4:** 408-419.

Vitone, N.D., Altizer, S. & Nunn, C.L. 2004. Body size, diet and sociality influence the species richness of parasitic worms in anthropoid primates. *Evolutionary Ecology Research* **6:** 183–199.

Walther, B.A., Cotgreave, P., Price, R.D., Gregory, R.D. & Clayton, D.H. 1995. Sampling effort and parasite species richness. *Parasitology Today* **11:** 306-310.

Wegner, K.M., Reusch, T.B. & Kalbe, M. 2003. Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology* **16:** 224-232.

Westerdahl, H., Hansson, B., Bensch, S. & Hasselquist, D. 2004. Between-year variation of MHC allele frequencies in great reed warblers: selection or drift? *Journal of Evolutionary Biology* **17:** 485-492.

Wilson, D.E. & Reeder, D.M., eds. 1993. *Mammal species of the world: a taxonomic and geographic reference*. Smithsonian Institution Press, Washington, USA.

Wlasiuk, G. & Nachman, M.W. 2010. Promiscuity and the rate of molecular evolution at primate immunity genes. *Evolution* **in press**.

Won, Y.-J. & Hey, J. 2005. Divergence population genetics of chimpanzees. *Molecular Biology and Evolution* **22:** 297-307.

Yeager, M. & Hughes, A.L. 1999. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunological Reviews* **167:** 45-58.

Yu, N., Jensen-Seaman, M.I., Chemnick, L., Ryder, O. & Li, W.-H. 2004. Nucleotide diversity in gorillas. *Genetics* **166:** 1375-1383.

TABLES

Table 1. Multi-predictor phylogenetic models of MHC-DRB polymorphism that investigated the effect of parasite species richness, population size, geographic range and body size. Note that the models did not require phylogenetic adjustments for the data (λ was estimated to be zero), probably because most of the phylogenetic variation was captured by geographic origin.

Model	Effect statistics
Allelic diversity	
Full model	$\lambda = 0.000$, $F_{6,28} = 3.713$, N = 35 (P = 0.008)
Species richness of all parasites	r = -0.061, 95% CI = -0.412 to 0.306 (P = 0.750)
Population size	r = 0.019, 95% CI = -0.343 to 0.377 (P = 0.920)
Geographic origin	r = 0.626, 95% CI = 0.343 to 0.805 (P = 0.003)
Body mass (residual)	r = 0.099, 95% CI = -0.271 to 0.443 (P = 0.604)
Full model	$\lambda = 0.000$, $F_{6,23} = 3.144$, N = 30 (P = 0.021)
Nematode species richness	r = 0.064, 95% CI = -0.340 to 0.447 (P = 0.763)
Population size	r = -0.018, 95% CI = -0.411 to 0.379 (P = 0.930)
Geographic origin	r = 0.662, 95% CI = 0.361 to 0.838 (P = 0.004)
Body mass (residual)	r = 0.038, 95% CI = -0.363 to 0.427 (P = 0.858)
d_N-ABS	
Full model	$\lambda = 0.000$, $F_{7,24} = 9.814$, N = 32 (P < 0.001)
Species richness of all parasites	r = 0.358, 95% CI = -0.035 to 0.654 (P = 0.073)
d_S-ABS	r = 0.697, 95% CI = 0.424 to 0.854 (P < 0.001)
Population size	r = 0.199, 95% CI = -0.204 to 0.544 (P = 0.330)
Geographic origin	r = 0.582, 95% CI = 0.251 to 0.791 (P = 0.017)
Body mass (residual)	r = -0.135, 95% CI = -0.496 to 0.267 (P = 0.512)
Full model	$\lambda = 0.000$, $F_{7,19} = 10.31$, N = 27 (P < 0.001)
Nematode species richness	r = 0.507, 95% CI = 0.097 to 0.770 (P = 0.019)
d_S-ABS	r = 0.701, 95% CI = 0.387 to 0.870 (P < 0.001)
Population size	r = 0.200, 95% CI = -0.253 to 0.582 (P = 0.384)
Geographic origin	r = 0.677, 95% CI = 0.347 to 0.858 (P = 0.008)
Body mass (residual)	r = -0.153, 95% CI = -0.549 to 0.298 (P = 0.507)

712 **FIGURE LEGENDS**

713 Figure 1. The organization of the MHC class II region within the chromosome (CEN: centromere
714 TEL: telomere) coding molecules that are expressed by antigen presenting cells. These proteins are
715 denoted as *DP-DR* proteins and composed of an invariable α-chain (coded by the *A* genes) and a
716 variable β-chain (coded by the B genes). Depending on the individual, different *Mhc-DRB* genes
717 each occupying different loci (such as *DRB1*, *DRB2* or *DRB*W*) may be present in the
718 chromosome. Such individual-specific compositions are distinguished as *DR* haplotypes (the map
719 shows the example of an individual that harbours three *DRB* loci). The enlargement in the middle
720 shows the structure of the *Mhc-DRB1* gene as assembled by exons and introns (marked with ||). The
721 numbers reflect the length of the corresponding nucleotide sequences in base pair. The enlargement
722 in the bottom provides information on the position of the 16 contact residues that codes the amino
723 acids of the antigen binding sites. Based on the similarity of sequences in terms of nucleotide
724 composition, *DRB* alleles within each locus identified in a species can be arranged into different
725 allelic lineages (such as *DRB1*01*, *DRB1*03* or *DRB*W2*) that appear as groups of highly related
726 alleles in a phylogenetic analysis of sequences. Given its specific organization and the large number
727 of potential allele combination, the DRB region is the most polymorphic part of the MHC class II
728 gene complex.

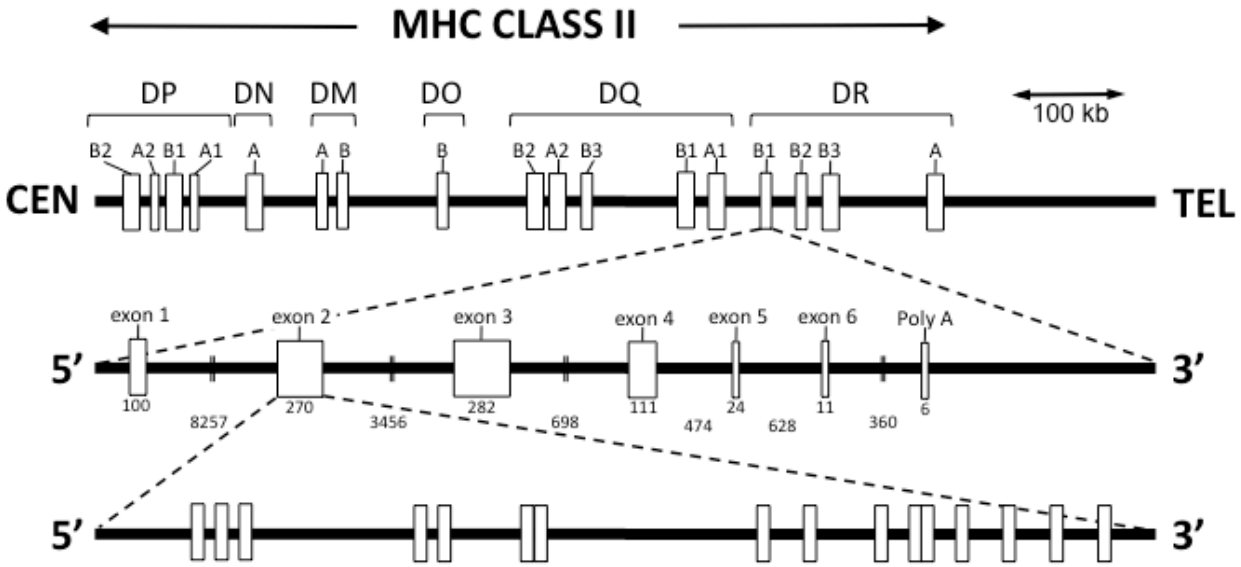
729

730 Figure 2. The relationship between MHC-DRB polymorphism in terms of non-synonymous
731 substitution rate at the ABS (d_N-ABS) and nematode species burden in primates without (A) and
732 with (B) considering confounding variables. A) The bivariate relationship between the two traits.
733 Points correspond to the appropriately transformed species-specific data (N = 31). B) The
734 relationship between traits controlling for synonymous substitution rate (d_S-ABS), population size,
735 geography and body size. Points are residuals taken from the relevant phylogenetic model (Table 1,
736 N = 27). Lines are regression lines.

737

738 Figure 3. The pair-wise relationship between A) geographic range and allelic variation of the *Mhc-*
739 *DRB*, and B) between geographic range and non-synonymous nucleotide substitution rate at the
740 ABS. Columns represent mean values, error bars give SE, while the numbers within columns show
741 the number of species in the corresponding realm.
742
743

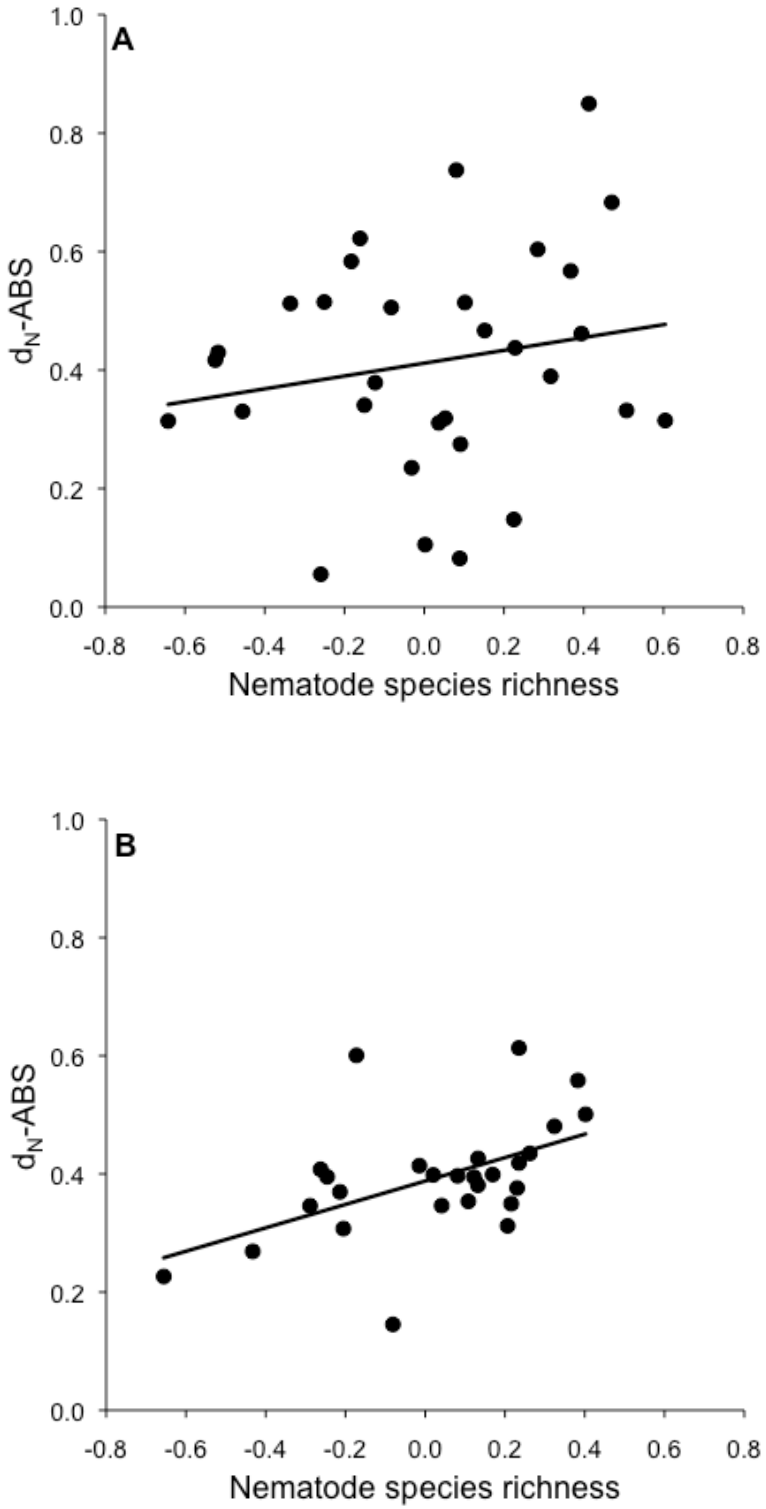
744 **FIGURE 1**



745

746

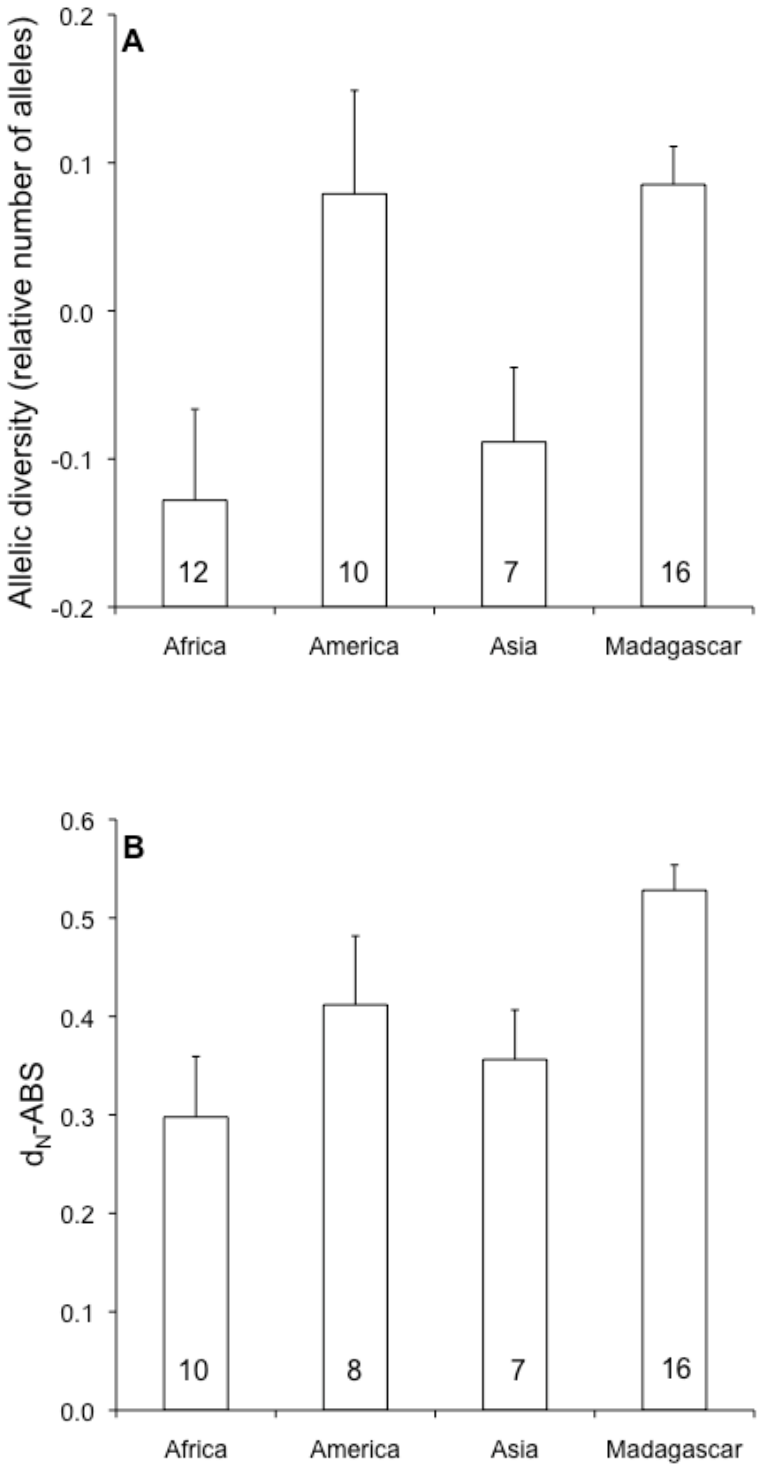
747 **FIGURE 2**



748

749

750 **FIGURE 3**



751