Parasite-mediated evolution of the functional part of the MHC in primates

L. Z. GARAMSZEGI* & C. L. NUNN†
*Department of Evolutionary Ecology, Estacion Biologica de Doñana-CSIC, Seville, Spain
†Department of Human Evolutionary Biology, Peabody Museum, Harvard University, Cambridge, MA, USA

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), whereas 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 manifested both as a large number of alleles and by the differences in nucleotide sequence among alleles (Hughes & Yeager, 1998). The second exon of the DRB region of the MHC class II. Our phylogenetic analyses controlled for the potential effects of neutral mutation rate, population size, geographic origin and body mass and revealed that nematode species richness associates positively with nonsynonymous 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 the different estimates of MHC polymorphism, whereas geography plays an independent role in the natural history of MHC.

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 focus 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 the potential effects of neutral mutation rate, population size, geographic origin and body mass and revealed that nematode species richness associates positively with nonsynonymous 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 the different estimates of MHC polymorphism, whereas 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.

Correspondence: László Z. Garamszegi, Department of Evolutionary Ecology, Estacion Biologica de Doñana-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

JOURNAL OF EVOLUTIONARY BIOLOGY © 2010 EUROPEAN SOCIETY FOR EVOLUTIONARY BIOLOGY

JOURNAL OF EVOLUTIONARY BIOLOGY 24 (2011) 184–195

DOI: 10.1111/j.1420-9101.2010.02156.x
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 times and/or locations (Hill, 1991).

These hypotheses are not mutually exclusive and have received mixed support in the studies of both laboratory and wild populations (Spurgin & Richardson, 2010). Most of these studies focused on the presence 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 necessarily 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 the 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 nonsynonymous substitution rate without necessarily accumulating MHC alleles. This is because allelic diversity and substitution

---

**Fig. 1** The organization of the MHC class II region within the chromosome (CEN: centromere TEL: telomere)-coding molecules that are expressed by antigen-presenting cells. These proteins are denoted as DP-DR proteins and composed of an invariable α-chain (coded by the A genes) and a variable β-chain (coded by the B genes). Depending on the individual, different Mhc-DRB genes each occupying different loci (such as DRB1, DRB2 or DRB*W) may be present in the chromosome. Such individual-specific compositions are distinguished as DR haplotypes (the map shows the example of an individual that harbours three DRB loci). The enlargement in the middle shows the structure of the Mhc-DRB1 gene as assembled by exons and introns (marked with II). The numbers reflect the length of the corresponding nucleotide sequences in base pairs. The enlargement in the bottom provides information on the position of the 16 contact residues that codes the amino acids of the antigen-binding sites. Based on the similarity of sequences in terms of nucleotide composition, DRB alleles within each locus identified in a species can be arranged into different allelic lineages (such as DRB1*01, DRB1*03 or DRB*W2) that appear as groups of highly related alleles in a phylogenetic analysis of sequences. Given its specific organization and the large number of potential allele combinations, the DRB region is the most polymorphic part of the MHC class II gene complex. MHC, major histocompatibility complex.
rate reflect different phenomenon. The former refers to the number of alleles preserved in a population irrespective of the sequence divergence between these alleles, whereas the latter refers to sequence variability regardless of the number of functioning alleles on which this variability is preserved. Therefore, several alleles with low sequence divergence could be maintained, resulting in protection against a narrow spectrum of antigens; or few very divergent alleles could be maintained resulting in resistance to many parasite species. Consequently, pathogen-driven selection forces can favour species that have either more MHC alleles or a higher nonsynonymous substitution rate or both.

Some recent studies have investigated why some populations maintain more alleles or a higher substitution rate than others (e.g. Boyce et al., 1997; Kim et al., 1999; 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 II 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 a distinction would be important, because as noted 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 (Fig. 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 nonsynonymous nucleotide substitution rate at the functional part of the MHC (Fig. 1) covary with the 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. Because of 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 co-evolutionary history with their hosts (Sorci et al., 2003; Nieberding et al., 2005). Hence, 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 data set that integrates genetic and parasitological data. We controlled for the potentially confounding effects of host phylogenetic history, neutral mutation rate, 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. Given that data on MHC polymorphism are currently available only for a modest number of species we adopted an effect size framework for evaluating the predictions. By shifting the focus from statistical significance, this approach enabled us to investigate the strength of the effects and the precision by which these effects can be estimated from the currently available sample (Nakagawa & Cuthill, 2007).

Materials and methods

Our MHC data rely 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 analysed 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 data set, 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). This earlier study focused on the covariation of different estimates of MHC polymorphism that is prevalent across allelic lineages, whereas here we investigated whether parasitological and other factors...
account for MHC diversity in an across-species context. For this purpose, we relied on information on the species-specific estimates of the within-lineage polymorphism of the exon 2 of Mhc-DRB in primates (Fig. 1; see also definitions and calculations elsewhere), for which data were extracted 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 nonorthologues (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 nonsynonymous (dN) 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 nonfunctional 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 Fig. 1). The aligned sequences can be found in Garamszegi et al. (2009b), and the updated IPD/MHC database lists the corresponding GenBank accession numbers.

For the estimation of substitution rates, we used the Nei & Gojobori (1986) method with the Jukes & Cantor (1969) correction for multiple hits, which computes the number of nonsynonymous differences (dN) between each pair of sequences after normalizing for the potential number of nonsynonymous sites and by correcting for multiple substitutions. For each group of lineages within species, an arithmetic average of dN was computed for all possible pairwise comparisons of sequences. We repeated this process for synonymous mutation rate (dS). 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 nonsynonymous substitution rate at the ABS simply as dN-ABS; dN-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 the tests of positive selection on the MHC, which assess the frequency of nonsynonymous substitutions relative to the frequency of synonymous substitutions ($dN/dS$). We avoided calculating the $dN/dS$ 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 dN-ABS, and the neutral mutation rate ($dS$-ABS) was held constant by including it in our statistical models.

We created a data set 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 earlier. Then, we built a General Linear Mixed Model 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 whereas lineage-specific effects were held constant. To achieve this, we used $dS/\bar{d}_K$-ABS as the dependent variable, and entered species as a main factor and lineage as a random factor. We applied square-root-arcsine transformation on $dS/\bar{d}_K$-ABS 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. We also derived species-specific estimates of $dS/\bar{d}_K$-ABS in the same way. 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 aforementioned, 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
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). Although individual parasites select for qualitative defence, parasite species richness is more likely to favour quantitative defence, as hosts that are exposed to several defence, parasite species richness is more likely to favour evolutionary impact of parasites on host species (Poulin, 2002). Thus, we considered the Papio group as different species (Papio cynocephalus, Papio hamadryas, and Papio ursinus); Aotus nancymaeae and Aotus nigriceps as Aotus azarae; Aotus trivirgatus and Aotus vociferans as A. trivirgatus and Microcebus myoxinus as Microcebus 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.

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 nonsynonymous substitution rate relative to synonymous substitution rate (Bernatchez & Landry, 2003; Pietrny & Oliver, 2006), the comparison of nonsynonymous substitutions might be more informative when silent mutation rates are included as a covariate in the statistical model. To achieve this normalization, we included dS-ABS in the statistical analyses, in which dS-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 dS : dN 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 data set. We therefore estimated observed population size as density (individuals per km²) x distribution area (km²) (see Møller, 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; Wasiuk & Bachman, 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 the 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 would have been detected with more complete screening for parasites (see also Nunn et al., 2003).
2010), we could locate effective population size data for six species in our data set 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. When 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 and 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 & 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 independent variables in the same regression model that would induce collinearity problems. To avoid this problem, 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.

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 because of 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 ($\lambda$), 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 $\lambda$; 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 did not improve the 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 in these calculations 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).

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 to balance type I errors (i.e. rejecting the null hypothesis when it is true) owing 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).

Results

First, we tested for the relationship between overall parasite species richness and MHC-DRB polymorphism in terms of allelic diversity and nonsynonymous substitution rate at the ABS of the molecule (d$_{NS-ABS}$). The phylogenetic models generally failed to detect strong relationships between species richness and these MHC traits (allelic diversity: $\lambda = 0.586$, $r = -0.012$, 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. When 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 and 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 & 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 data set is provided in the Supporting information.
95% CI = −0.318 to 0.297, \( n = 41, P = 0.936; d_N\text{-ABS}: \hat{\lambda} = 0.891, \hat{r} = 0.143, 95\% \text{ CI} = −0.190 \text{ to } 0.446, \ n = 37, P = 0.403 \). We repeated these analyses for nematode species richness and found that the estimated range of the effect size for the relationship between \( d_N\text{-ABS} \) and nematode diversity mostly covers a positive association (allelic diversity: \( \hat{\lambda} = 0.810, \hat{r} = 0.060, 95\% \text{ CI} = −0.279 \text{ to } 0.386, \ n = 35, P = 0.732; d_N\text{-ABS}: \hat{\lambda} = 0.852, \hat{r} = 0.294, 95\% \text{ CI} = −0.067 \text{ to } 0.587, \ n = 31, P = 0.109, \ Fig. 2a).

We then developed a multi-predictor phylogenetic model that controlled for \( d_S\text{-ABS} \), population size, geographic range and body mass (\( d_S\text{-ABS} \) 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 whether 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 \( d_N\text{-ABS} \) remained, with the effect size covering a 95% confidence range that excluded zero (Table 1, Fig. 2b). The models also revealed that both allelic variation and \( d_N\text{-ABS} \) 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 (Fig. 3).

**Discussion**

Given the limited sample size, the 95% CI of the estimated effects show that there is 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 \( d_N\text{-ABS} \) and nematode burden can be concluded as 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 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 nonsynonymous 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 nonsynonymous 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 nonprotective 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 nonsynonymous 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 conse-

### Table 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Effect statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allelic diversity</td>
<td></td>
</tr>
<tr>
<td>Full model</td>
<td>( \lambda = 0.000, F_{2,25} = 3.713, n = 35 )</td>
</tr>
<tr>
<td>Species richness of all parasites</td>
<td>( r = -0.061, 95% )</td>
</tr>
<tr>
<td>Population size</td>
<td>( r = 0.019, 95% )</td>
</tr>
<tr>
<td>Geographic origin</td>
<td>( r = 0.626, 95% )</td>
</tr>
<tr>
<td>Body mass (residual)</td>
<td>( r = 0.099, 95% )</td>
</tr>
<tr>
<td>Full model</td>
<td>( \lambda = 0.000, F_{2,23} = 3.144, n = 30 )</td>
</tr>
<tr>
<td>Nematode species richness</td>
<td>( r = 0.064, 95% )</td>
</tr>
<tr>
<td>Population size</td>
<td>( r = -0.018, 95% )</td>
</tr>
<tr>
<td>Geographic origin</td>
<td>( r = 0.662, 95% )</td>
</tr>
<tr>
<td>Body mass (residual)</td>
<td>( r = 0.038, 95% )</td>
</tr>
<tr>
<td>( d_{\text{ABS}} )</td>
<td>[ \lambda = 0.000, F_{2,24} = 9.814, n = 32 ]</td>
</tr>
<tr>
<td>Species richness of all parasites</td>
<td>( r = 0.358, 95% )</td>
</tr>
<tr>
<td>( d_{\text{ABS}} )</td>
<td>( r = 0.697, 95% )</td>
</tr>
<tr>
<td>Population size</td>
<td>( r = 0.199, 95% )</td>
</tr>
<tr>
<td>Geographic origin</td>
<td>( r = 0.582, 95% )</td>
</tr>
<tr>
<td>Body mass (residual)</td>
<td>( r = -0.135, 95% )</td>
</tr>
<tr>
<td>Full model</td>
<td>( \lambda = 0.000, F_{1,19} = 10.31, n = 27 )</td>
</tr>
<tr>
<td>Nematode species richness</td>
<td>( r = 0.507, 95% )</td>
</tr>
<tr>
<td>( d_{\text{ABS}} )</td>
<td>( r = 0.701, 95% )</td>
</tr>
<tr>
<td>Population size</td>
<td>( r = 0.200, 95% )</td>
</tr>
<tr>
<td>Geographic origin</td>
<td>( r = 0.677, 95% )</td>
</tr>
<tr>
<td>Body mass (residual)</td>
<td>( r = -0.153, 95% )</td>
</tr>
</tbody>
</table>

MHC, major histocompatibility complex; ABS, antigen-binding sites.
quences 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 because of 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 (Fig. 2). Lemurs and New World monkeys likely originated from small founder populations (Trikova 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).

In summary, our analyses offer new insights to the evolutionary origins of MHC diversity. A previous study showed that different estimates of MHC polymorphism, such as allelic diversity and sequence variation, evolve in a correlated fashion (Garamszegi et al., 2009b). Here, we extended this work by demonstrating that parasitism and geographic distribution can account for MHC polymorphism, but these factors can incur independent selection forces for different MHC traits. Although most of the previous studies have demonstrated links between specific parasites and particular MHC alleles, 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 into MHC diversity and evolution, including tests of hypotheses involving mate choice, demography and environmental predictors of disease risk.

Acknowledgments

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


