

EVOLUTIONARY MECHANISMS  
SHAPING MHC VARIATION  
IN SYMPATRIC LEMURS

Dissertation  
for the award of the degree  
“Doctor rerum naturalium” (Dr. rer. nat.)  
of the Georg-August-Universität Göttingen

within the doctoral program Biology  
of the Georg-August-University School of Science (GAUSS)

Submitted by  
Eva Kaesler, born Pechoušková

From Praha, Czech Republic

Göttingen 2015



## **THESIS COMMITTEE:**

Prof. Dr. Peter M. Kappeler, Department of Sociobiology and Anthropology, Georg-August-Universität Göttingen, Behavioral Ecology and Sociobiology Unit, German Primate Center

Prof. Dr. Eckhard W. Heymann, Department of Sociobiology and Anthropology, Georg-August-Universität Göttingen, Behavioral Ecology and Sociobiology Unit, German Primate Center

## **MEMBERS OF THE EXAMINATION BOARD:**

Reviewer: Prof. Dr. Peter M. Kappeler

Second Reviewer: Prof. Dr. Eckhard W. Heymann

## **FURTHER MEMBERS OF THE EXAMINATION BOARD:**

Dr. Elise Huchard, Evolutionary Ecology Department, Centre d'Ecologie Fonctionnelle et Evolutive Montpellier, France

PD Dr. Christian Roos, Primate Genetics Laboratory, German Primate Center

Prof. Dr. Julia Ostner, Primate Social Evolution Group, Courant Research Centre Evolution of Social Behavior, Georg-August-Universität Göttingen

Dr. Oliver Schülke, Primate Social Evolution Group, Courant Research Centre Evolution of Social Behavior, Georg-August-Universität Göttingen

DATE OF THE ORAL EXAMINATION: 19 NOVEMBER 2015



*Dedicated to*



*my beloved Fifinka*

# CONTENTS

SUMMARY .....	1
ZUSAMMENFASSUNG.....	3
GENERAL INTRODUCTION.....	6
CHAPTER 1 .....	18
The Major Histocompatibility Complex and Primate Behavioral Ecology: New Tools and Future Questions	
CHAPTER 2.....	44
MHC class II variation in a rare and ecological specialist mouse lemur reveals lower allelic richness and contrasting selection patterns compared to a generalist and widespread sympatric congener	
CHAPTER 3.....	72
Evolutionary origins of shared MHC polymorphism in sympatric lemurs	
APPENDICES .....	105
GENERAL DISCUSSION .....	164
REFERENCES .....	181
ACKNOWLEDGEMENTS.....	215
CURRICULUM VITAE .....	217
DECLARATION .....	219

# SUMMARY

The major histocompatibility complex (MHC) is a multigene cluster characterized by remarkable polymorphism and a complex evolutionary history. The MHC genes assume a central role in the adaptive immune response of vertebrates and their polymorphism is thought to affect the functional plasticity of immune responses against heterogeneous pathogenic pressures. Beside of their immune function, MHC genes are thought to be involved in a variety of non-immune functions, such as mate choice, individual or kin recognition and other reproductive functions. This makes them a particularly interesting molecular marker to address some fundamental evolutionary questions, such as evolution of molecular adaptations or proximate mechanisms maintaining species diversification at the molecular level.

The challenges imposed by genotyping highly polymorphic MHC genes have long hampered progress in conducting large-scale studies that would enable addressing such questions, especially in non-model organisms. Introducing Next-generation sequencing (NGS) therefore represent a major break-through in this aspect. I first review advantages as well as challenges linked to the use of NGS for MHC genotyping before envisioning the revolutionary implications of integration of NGS into MHC-based research in evolutionary ecology and primatology.

Moreover, the remarkable MHC polymorphism often crosses species boundaries, with similar alleles or allelic motifs shared across species. Understanding the origin and mechanisms underlying the maintenance of this polymorphism across different species may improve our understanding of the evolution of host-immune response. Here, I therefore aimed to broaden our understanding of the evolution and the maintenance of MHC class II polymorphism at the community level. First, I explored MHC variation at the two highly polymorphic loci, DRB and DQB, and patterns of molecular selection acting upon them in four sympatric con-familiar lemur species - *Microcebus murinus*, *M. berthae*, *Cheirogaleus medius* and *Mirza coquereli* (*Cheirogaleidae*).

I observed contrasting patterns of MHC variation and molecular selection, but also considerable functional and structural overlap among MHC alleles of these species. These lemurs present contrasting aspects of their ecology that has been suggested to affect

## SUMMARY

---

MHC variation and the level of pathogen exposure. I found out that demographic factors may exert a stronger influence than pathogen-driven selection on current levels of standing allelic richness, especially in species with more pronounced ecological vulnerability.

I then attempted to elucidate the origin of MHC allelic similarity. Shared MHC polymorphism at coding regions has been suggested to reflect either the operation of convergent selection in the presence of parallel selective pressures (e.g. shared parasites), or via allelic co-ancestry and the long-term maintenance of MHC sequence motifs over multiple speciation events. I integrated MHC data with parasitological screening of gut helminth communities to investigate whether similar parasite pressure may select functionally similar MHC alleles in different host species. I detected link between shared parasite pressures and the distribution of functionally similar MHC alleles across *Cheirogaleidae*. Moreover, I found out significant associations between functionally similar MHC alleles and particular helminth infestation in closely related *Microcebus* sp. that suggested potential functional convergence between these species. This finding indicates that shared descent rather than convergent evolution might be responsible for this functional convergence. Finally, I explored MHC data presented by this thesis and those previously obtained for other *Cheirogaleidae* to examine patterns of structural integrity of MHC alleles as a sign of allelic co-ancestry. I provide quantitative evidence that co-ancestry is the primary mechanism responsible for the retention of MHC sequence motifs between species that diverged up to 30 million years ago.

My findings highlight the importance of moving from simple host-parasites models to community level approach that may provide more realistic picture of host-pathogen co-evolution. Moreover, I show that common selective pressures may contribute to rather slow-down MHC divergence across species and populations and that the exceptional polymorphism is in fact concentrated within species rather than among species. Last, functionally similar alleles, or allelic motifs, have been maintained over multiple speciation events, presumably accumulated through interactions between hosts and parasites over long evolutionary time scales. This stresses out the importance to protect standing genetic variability as it might be slow to regenerate when eroded



# ZUSAMMENFASSUNG

Der Haupthistokompatibilitätskomplex (major histocompatibility complex, MHC) ist ein Multigen-Cluster, welcher durch seinen bemerkenswert hohen Polymorphismus sowie seine komplexe Evolutionsgeschichte geprägt ist. Die MHC-Gene spielen eine zentrale Rolle in der adaptiven Immunantwort von Wirbeltieren, wobei der Polymorphismus mutmaßlich einen Einfluss auf die funktionelle Plastizität der Immunantwort auf heterogene pathogene Selektionsdrücke hat.

Des Weiteren wird vermutet, dass MHC-Gene in weitere, nicht-immunspezifische Funktionen involviert sind, wie z.B. Partnerwahl, Verwandten- und Individuenerkennung sowie weitere Fortpflanzungsfunktionen. Es handelt sich daher bei den MHC-Genen um einen besonders interessanten molekularen Marker für die Beantwortung fundamentaler evolutionärer Fragen, wie z.B. die Evolution der molekularen Adaption oder der molekularen Mechanismen der Speziesdiversifikation.

Die Genotypisierung solch hochpolymorpher Gene, insbesondere außerhalb von Modellorganismen, wurde erst durch das aufkommen des Next Generation Sequencing (NGS) ermöglicht. Zu Beginn werde ich Vorteile und Herausforderungen der Anwendung von NGS für die Genotypisierung der MHC-Gene erläutern, bevor ich auf die revolutionären Implikationen von NGS für die MHC-basierte Forschung im Bereich evolutionärer Ökologie und Primatologie eingehe.

MHC Polymorphismus überschreitet häufig Artgrenzen, das heißt es werden ähnliche Allele bzw. Allel motive über Artgrenzen hinweg gefunden. Wissen über den Ursprung sowie die Mechanismen, welche diesen Polymorphismus über Artgrenzen hinweg erhalten, könnten unser Verständnis der Evolution der Wirtsimmunantwort erhöhen. Diese Arbeit zielte daher auf ein breiteres Verständnis der Evolution und des Erhalts des Klasse II MHC Polymorphismus auf der Gemeinschaftsebene ab.

Als erstes untersuchte ich in den vier sympatrischen con-familiären Lemurenarten *Microcebus murinus*, *M. berthae*, *Cheirogaleus medius* and *Mirza coquereli* (*Cheirogaleidae*) die MHC-Variationen der zwei hochpolymorphen Loci DRB und DQB sowie die molekulare Selektion, der sie ausgesetzt sind. Hierbei beobachtete ich gegensätzliche Muster der MHC-Variationen und der molekularen Selektion, aber auch

eine funktionelle und strukturelle Überlappung der MHC-Gene dieser Spezies. Diese Lemuren zeigen gegensätzliche ökologische Aspekte, welche möglicherweise einen Einfluss auf die MHC-Variationen und Pathogen Exposition haben. Meine Ergebnisse zeigen, dass die Demographie einen stärkeren Einfluss auf die momentane Allelreichheit ausübt, als pathogen-basierte Selektion, insbesondere in Arten mit erhöhter ökologischer Gefährdung.

Des Weiteren untersuchte ich mögliche Ursachen der MHC Allelähnlichkeit. Es wurde vorgeschlagen, dass das Vorliegen eines gemeinsamen MHC Polymorphismus in den kodierenden Regionen auf konvergenter Selektion aufgrund von parallelen Selektionsdrücken (z.B. gemeinsame Parasiten), oder aber auf einer allelischen co-Abstammung und der langfristigen Erhaltung der MHC-Sequenzmotive über mehrere Speziationen hinaus beruht. Um zu untersuchen, ob ähnliche parasitäre Selektionsdrücke auf funktionell ähnliche MHC-Allele selektieren, habe ich die Ergebnisse eines parasitologischen Screenings des Verdauungstraktes auf Würmer mit MHC-Daten kombiniert. Innerhalb der Cheirogaleidae konnte ich eine Korrelation zwischen gemeinsamen parasitären Selektionsdrücken und der Verteilung funktionell ähnlicher MHC-Allele finden. Außerdem fand ich eine signifikante Assoziation von funktionell ähnlichen MHC-Allelen und bestimmten Wurminfektionen in eng verwandten *Microcebus* sp., was eine potentielle funktionelle Konvergenz zwischen diesen eng verwandten Arten indiziert. Schlußendlich untersuchte ich MHC-Daten dieser Dissertation sowie zuvor gewonnene Daten auf Muster struktureller Integrität innerhalb der MHC-Allele als Zeichen allelischer co-Abstammung. Ich zeige quantitative Beweise, dass die co-Abstammung der primäre Mechanismus des Erhalts von MHC-Sequenzmotiven zwischen Spezies ist, deren Artbildung bis zu 30 Millionen Jahre zurück liegt.

Diese Funde heben hervor, wie wichtig es für ein realistischeres Bild der Wirt-Parasit Koevolution ist, den Fokus von einfachen Wirt-Parasit-Modellen auf Gemeinschaftsebenenmodelle zu legen. Ich zeige außerdem, dass gemeinsame Selektionsdrücke zu einer Verlangsamung der art- und populationsübergreifenden MHC-Divergenz führen und dass der außergewöhnliche Polymorphismus eher innerhalb einer Art als zwischen den Arten zu finden ist. Funktionell ähnliche Allele oder Allelmotive blieben über mehrere Artbildungen hinweg erhalten und wurden möglicherweise durch Wirt-Parasit-Interaktionen über lange evolutionäre Zeiträume hinweg akkumuliert. Dies

hebt die Wichtigkeit des Schutzes der bestehenden genetischen Variabilität hervor, da sie nach Verlust nur langsam wieder regeneriert werden kann.

# GENERAL INTRODUCTION

## MHC AS A PLEIOTROPIC MARKER IN EVOLUTIONARY BIOLOGY

The major histocompatibility complex (MHC) is a multigene cluster consisting of highly polymorphic immune genes with an elaborate architecture (Hedrick 2002; Garrigan and Hedrick 2003; Piertney and Oliver 2006; Spurgin and Richardson 2010). MHC genes encode MHC molecules, cell-surface glycoproteins, that are actively involved in initial phases of immune-response against intracellular (MHC class I) or extracellular pathogens (MHC class II) by peptide (antigen) presentation at peptide-binding region (PBR) to T-lymphocytes (Bernatchez and Landry 2003). Each MHC molecule can bind only a limited array of antigens derived from pathogens (Rammensee et al. 1995) and thereby initiate an antigen-specific immune response (Trowsdale 1993). The majority of MHC polymorphism is concentrated at peptide-binding region (PBR) (Hughes and Nei 1988, 1989; Bernatchez and Landry 2003; Hughes and Yeager 1998). The immuno-repertoire of an individual, i.e. the array of recognized antigens, can be increased by a number of mechanisms, such as gene duplication, increasing individual level of allelic diversity, increasing polymorphism at PBR, or alternatively, outside PBR (ca 4 Å; Nielsen et al. 2007) that may alter the 3D positioning of PBR contact residues (Bjorkman and Burmeister 1994). Consequently, it also determines the binding specificity of each encoded MHC molecule to a limited range of antigens (Rammensee et al. 1995). MHC genes have also been found to be involved in a variety of non-immune functions in all major vertebrate taxa, including mate choice, individual or kin recognition thought to be mediated by olfactory signalling and a number of other reproductive functions (reviewed in Apanius et al. 1997; Penn and Potts 1999; Younger et al. 2001; Penn 2002; Milinski 2006; Yamazaki and Beauchamp 2007; Havlicek and Roberts 2009; Spurgin and Richardson 2010; Ruff et al. 2012).

Due to their immune function, MHC genes are thought to be under strong selection pressure driven largely by balancing selection mediated by host-pathogen

interactions which are thought to promote its extreme polymorphism (reviewed in Apanius et al. 1997; Bernatchez and Landry 2003; Piertney and Oliver 2006; Spurgin and Richardson 2010). Compelling evidence for this hypothesis stems from reported associations between higher MHC diversity in populations or individuals exposed to a higher diversity of parasites (Wegner et al. 2003; Šimková et al. 2006; Dionne et al. 2007; de Bellocq et al. 2008; Prugnolle et al. 2005; Garamszegi and Nunn 2011; Tobler et al. 2014), or associations between possession of certain MHC alleles and single or multiple pathogen prevalence at the population or at the community level (e.g. Hill et al. 1991; Wegner et al. 2003; Froeschke and Sommer 2005; Meyer-Lucht and Sommer 2005; Schad et al. 2005; Schwensow et al. 2007, 2010a,b; Oliver et al. 2009; Froeschke and Sommer 2012; Sepil et al. 2013; Sommer et al. 2014; Pilosof et al. 2014).

Based on the assumption of co-dependency between strength of selective pressure exerted by pathogens and its link to the evolution of MHC polymorphism, several potentially non-exclusive evolutionary mechanisms have been proposed to explain it. Up to date, hypotheses of negative frequency-dependent selection (Snell 1968; Bodmer 1972; Borghans et al. 2004), based on the principles of dynamic MHC allelic frequency fluctuation in response to co-evolving parasite communities, and of heterozygote advantage (Doherty and Zinkernagel 1975) extended to divergent allele advantage hypothesis, that would increase the host flexibility towards heterogenous pathogenic pressure (Wakeland et al. 1990; Lenz 2011) have received largest empirical support (reviewed in Apanius et al. 1997; Spurgin and Richardson 2010). In addition, variation of parasite-mediated selection patterns in space and time can induce selection for a specific set of MHC alleles (i.e. fluctuating selection; Hedrick 1999; Spurgin and Richardson 2010; Eizaguirre et al. 2009a, 2012).

Patterns and strength of selection exerted on MHC genes are likely to be affected by the diversity and frequencies of MHC alleles present in a host population, thus its demography (Hedrick 1972; Borghans et al. 2004), and by the composition and pathogenicity of the local parasite communities (Eizaguirre et al. 2009a; 2012). The capability of populations to keep up with co-evolving parasites depends on the contemporary genetic diversity (O'Brien and Evermann 1988; Hedrick 2001; Frankham et al. 2002). Thus, loss of genetic diversity, and thereby adaptive flexibility, may have critical consequences on fitness as well as spread of parasites due to increasing

frequencies of common genotypes (Frankham et al. 2002; Meagher 1999; Spielman et al. 2004). Therefore, the capacity to effectively respond to changing pathogenic pressure might be severely disrupted in small fragmented populations where limited gene flow might result in loss of standing genetic variation through genetic drift and inbreeding (Wright 1969; Keller and Waller 2002; Frankham et al. 2002; O'Brien and Evermann 1988). Moreover, introduction of novel pathogenic pressures might challenge the capability of such population to keep up with co-evolving parasites on a long-term.

The resistance of an individual against pathogens may affect its health, long-term fitness and survival (reviewed e.g. in Albon et al. 2002; Degen 2006; Pedersen and Greives 2008), and consequently, may influence the value of given genotypes for a potential mate. In that context, the type of selection affecting balanced MHC polymorphism linked to disease resistance may be also sexual selection (Milinski 2006; Penn and Potts 1999). In that sense, particular alleles that confer resistance against particular or multiple infestations, or allelic combinations that would expand the antigen binding repertoire of an individual facing heterogenous pathogenic pressure, might be targeted by mate choice (reviewed in Apanius et al. 1997; Milinski 2006; Spurgin and Richardson 2010) that appears to be context-dependent (Setchell and Huchard 2010; Kamiya et al. 2014). These targeted alleles might be species- or locally specific, reflecting the composition and selective impact of the present parasite community (Blais et al. 2007; Eizaguirre et al. 2009a, 2011, 2012; Eizaguirre and Lenz 2010; Matthews et al. 2010).

The crucial role of MHC genes in initial phases of immune response and its link to fitness and long-term survival of populations and mate choice, as well as complex routes leading to maintenance of their extreme polymorphism, predestinate MHC genes to be among the most interesting marker to study proximate mechanisms of molecular adaptation (Sommer 2005) and processes leading to population diversification (Eizaguirre et al. 2009a). Understanding which selective mechanisms or ecological factors have a crucial impact on the observed patterns of standing MHC variation (in terms of number of specific alleles 'allelic richness' as well as in terms of nucleotide or amino acid diversity 'allelic divergence') in species facing environmental and demographic challenges will facilitate future assessments of their ability to cope with rapid environmental changes, demographic fluctuations and novel selective pressures.

## EVOLUTIONARY MECHANISMS DRIVING MHC POLYMORPHISM

The origin and maintenance of remarkably complex and variable patterns of MHC polymorphism across species or populations and the presence of shared orthologous (homologous sequence descending from the same ancestral sequence) allelic lineages or identical allelic motifs across multiple speciation events, have received lots of attention (Klein et al. 2007). Inconsistencies between MHC genealogy manifested by inter-specific clustering of MHC alleles and species phylogeny has been reported for a variety of vertebrate species, including mammals (e.g. Lundberg and McDevitt 1992; Edwards et al. 1997; Otting et al. 2002; Bryja et al. 2006; Huchard et al. 2006; Cutrera and Lacey 2007; Xu et al. 2009; Kamath and Getz 2011; Kuduk et al. 2012; Yao et al. 2014), birds (e.g. Burri et al. 2008; Li et al. 2011; Alcaide et al. 2007, 2013; Eimes et al. 2015), fish (e.g. Graser et al. 1996; Figueroa et al. 2000; Ottová et al. 2005; Lenz et al. 2013), reptiles (e.g. Jaratlerdsiri et al. 2014; Stiebens et al. 2013) or amphibians (e.g. Kiemnec-Tyburczy et al. 2010; Tracy et al. 2015); reviewed in Azevedo et al. 2015; Tesicky and Vinkler 2015. Two major evolutionary mechanisms have been proposed to explain this intriguing presence of identical alleles in MHC of distant species that is often limited to coding regions of MHC alleles.

First, sharing of possibly advantageous orthologous allelic lineages, or allelic motifs, across taxa and over macro-evolutionary time-scales (trans-species polymorphism, ‘TSP’) has been attributed to similarity by descent, which is thought to be maintained by balancing selection (Klein 1987, Klein et al. 1998, 2007; Figueroa et al. 1988; McConell et al. 1988). The empirical evidence that enabled the foundation of this concept is based on serological isolation of mouse MHC H2 antigen showing identical peptide map between mouse and rat by Klein and Arden (1982), the presence of identical structural modifications (Figueroa et al. 1988) or retroposon-tagged mapping of transmission patterns (McConnell et al. 1988). Alternatively, MHC allelic similarity limited to peptide binding region (PBR) of the coding regions can be induced by the operation of convergent evolution. Convergent evolution assumes a presence of

overlapping selective pressures from parasites (e.g. in sympatry, or ecologically overlapping niches) leading to developing convergent adaptations in different species independently and regardless of their taxonomy (Gustafsson and Andersson 1994; Yeager and Hughes 1999; Christin et al. 2010). Specifically, alleles that are similar by descent at peptide-binding regions (PBR) are expected to retain a signature of co-ancestry also at sites that are selectively neutral, or under purifying selection, such as PBR- neighbouring codons ('non-PBR'), or non-coding introns. In contrast, independent convergence would be indicated if MHC alleles are more similar at PBR, but not at the non-PBR, or introns (Klein et al. 1998; Kriener et al. 2000). Moreover, other intriguing characteristics, such as shared polymorphism located within a CpG dinucleotide (the occurrence of C base next to G within linear DNA sequence) has been envisaged to provide information whether polymorphism in question is shared by descent, or is identical by state (i.e. occurs independently in different lineages). Cytosin in CpG can be methylated, which can alter the DNA expression (see e.g. Azevedo et al 2015 for a detailed review)

Both qualitative or quantitative indications supportive of either scenario in shaping MHC similarity across species have been reported (e.g. Convergent evolution: Andersson et al. 1991; Trtkova et al. 1995; Yeager and Hughes 1999; Figueroa et al. 2000; Kriener et al. 2000a, 2001; Srithayakumar et al. 2012; Pilosof et al. 2014; Co-ancestry: Lundberg and McDevitt 1992; Klein et al. 1993b; Graser et al. 1996; Garrigan and Hedrick 2003; Xu et al. 2009; Lenz et al. 2013; Tobler et al. 2014; Eimes et al. 2015), but only few studies have attempted to disentangle between these hypotheses directly. It remains challenging to quantitatively distinguish between these two evolutionary scenarios under natural settings, especially due to difficulties to identify and quantify major selective pressure and due to the complexity of MHC genotyping that would require genotyping large PBR coding and non-coding regions (e.g. introns) (see e.g. Zhang and Kumar 1997; Kriener et al. 2000a).

The complexity of MHC genotyping has for a long time hampered the large-scale population or community level studies that would enable addressing these and other major evolutionary questions, especially when dealing with non-model species. The main challenges are associated with the use of low quality non-invasive DNA samples in rare, threatened or otherwise protected species that may impair genotyping reliability, the extreme allelic polymorphism and frequent loci duplication, primer design that would



enable capturing the full allelic variability, or the correct identification of non-functional pseudogenes that might require obtaining invasive samples for gene-expression studies (reviewed in Janeway et al. 2005; Piertney and Oliver 2006; Babik 2010; Lighten et al. 2014a).

Majority of studies have thus focused on sequencing short fragments containing the antigen-binding sites and studies that would investigate also non-coding introns are relatively rare (see e.g. Zhang and Kumar 1997; Kriener et al. 2000; Lenz et al. 2013). Due to all these challenges, most MHC studies reporting TSP have suggested that the maintenance of ancient alleles by balancing selection is the most likely mechanism to explain lack of species-specific polymorphism at the coding sites.

Moreover, other molecular mechanisms, such as intra-locus recombination leading to gene conversion, incomplete lineage sorting by chance (i.e. allele sharing in closely related species that gradually disappears; Klein et al. 1998; Tracy et al. 2015), or allelic introgression between hybridizing species have been indicated to play a role in shaping allelic similarity (Andersson and Mikko 1995; Martinsohn et al. 1999; Wegner and Eizaguirre 2012; Nadachowska-Brzyska 2012; Hedrick 2013), but the empirical evidence remains limited.

Interactions between host MHC and parasite communities have been suggested to trigger and speed up fundamental evolutionary processes such as early population diversification preceding speciation events. MHC divergence could emerge in a context of local adaptation to differentiated parasite communities and later speed-up reproductive isolation via assortative mating (for example, choice for particular MHC-alleles) during episodes of adaptive radiation (Haldane 1949; Eizaguirre et al. 2009a; Maan and Seehausen 2011). Despite some evidence that divergent parasite pressures may promote selection for divergent MHC alleles in different populations (Blais et al. 2007; Ekblom et al. 2007; Eizaguirre et al. 2009a, 2011, 2012; Eizaguirre and Lenz 2010; Matthews et al. 2010), empirical evidence for a role of parasite-mediated selection in generating reproductive isolation between closely related species is scarce (Schluter 2001; Buckling and Rainey 2002; Eizaguirre et al. 2009a; Abbott et al. 2013). In contrast, balancing selection has been shown to homogenize, rather than diversify, MHC alleles in sympatric species after speciation, despite differences in parasite communities and erosion of background genomic diversity (Fraser and Neff 2010; Tobler et al. 2014). Balancing

selection could in theory over-ride the power of divergent parasite-mediated pressures and thereby hinder species diversification and speciation. Moreover, the presence of similar alleles, or allelic motifs, across different species exposed to different parasite pressures may suggest structural constraints on MHC evolution favoring certain allelic motifs (Gustafsson and Anderson 1994; Andersson et al. 1991).

Thus, understanding the origin of MHC allelic similarity will assist to comprehend which evolutionary processes might have majorly contributed to shape MHC variation and maintenance of similar allelic motifs, hence homogenizing the genetic diversity at coding regions in closely related but also more distantly related species, over evolutionary time-scales.

## THESIS AIMS

The introduction of Next-generation sequencing (NGS) technologies has brought a breakthrough in the feasibility of large-scale studies of species with unknown MHC constitution by overcoming numerous technical difficulties associated with traditional methods used for MHC genotyping. In the first chapter of my thesis, I review the progress in MHC genotyping techniques prior and after introducing NGS before focusing on new challenges emerging during processing complex high-throughput datasets generated by cost-effective, but also error-prone NGS technologies. I then highlight the potential of NGS technologies for future large-scale studies in evolutionary biology and primatology.

In the second chapter, I aim at understanding the processes responsible for shaping MHC variation among closely related, yet ecologically differentiated sympatric species. Here I attempted to understand the effects of demographic and ecological variation would influence patterns of MHC variation and molecular selection in two sympatric congeneric species. To do this, I compare patterns of MHC variation and tracks of molecular selection acting on the two highly polymorphic loci – DRB and DQB - in two sympatric lemurs species (*Microcebus* sp.), the rare ecological specialist Madame Berthae's (*Microcebus berthae*) mouse lemur and the common generalist grey (*Microcebus murinus*) mouse lemur. Using NGS (454 pyrosequencing), I describe MHC class II variation in *M. berthae* for the first time and compare my findings to the MHC data previously obtained for its sympatric congener *M. murinus* over comparable spatial and temporal scale in Kirindy forest, western Madagascar. These species present both a contrasting ecology and demography, two factors that have been shown to influence patterns of selection at MHC.

*M. berthae* is currently endangered (B1ab.i-iii; Andriaholinirina et al. 2014) and highly endemic to the dry forest of the Menabe region in western Madagascar (Schmid and Kappeler 1994; Ganzhorn et al. 2001; Schäfler and Kappeler 2014). It exhibits lower population densities than its widely distributed and abundant sympatric congener (Schäfler and Kappeler 2014). These two species contrast in the level of ecological flexibility and social cohesion, with *M. berthae* relying on specialist diet that is thought to promote intense scramble competition leading to spatial over-dispersion and lower rates

of social interactions among conspecifics (Dammhahn and Kappeler 2009). In contrast, a generalist diet may facilitate spatial proximity and social interactions among conspecifics of *M. murinus* (Dammhahn and Kappeler 2008a, b, 2010), and thus parasite transmission between them as well as it might promote exposure to more diverse array of pathogens. Moreover, larger population size and generalist diet is thought to increase risk of exposure to heterogenous parasites (e.g. Chen et al. 2008; Vitone et al. 2004; Nunn et al. 2003; Hughes and Pages 2007). These ecological aspects are thought to promote higher MHC allelic richness (number of distinct alleles) and allelic divergence (nucleotide or amino acid differentiation) as well as higher level of parasitism (Hedrick 1985; Anderson and May 1978; Vitone et al. 2004; Nunn et al. 2003; Hughes and Page 2007; reviewed in Morand 2015).

Based on these assumptions, I predicted allelic richness to be higher and the strength of molecular selection to be more prominent in *M. murinus*, given its potentially greater exposure to parasites and its larger population size. Alternatively, if ecological differentiation of *M. berthae* would promote relaxed pathogen-mediated pressure, I would expect to detect weaker tracks of pathogen-driven selection on MHC alleles compared to *M. murinus*.

These two species represent a good example of species with close life-histories facing similar environmental challenges, such as ongoing loss of habitat or highly seasonal environment, but showing different degrees of ecological vulnerability. Indications of demographic or genetic bottlenecks have been reported across various lemur species (Louis et al. 2005; Olivieri et al. 2008; Markolf et al. 2008; Radespiel et al. 2008; Craul et al. 2009; Razakamaharavo et al. 2010; Holmes et al. 2013; Baden et al. 2014). However, little is known about potential fitness consequences and the capability of these populations to maintain effective level of functional (adaptive) genetic variation facing heterogenous and increasing environmental and anthropogenic pressures. Consequently, the evaluation of standing MHC variation over longer time scales in species with contrasting ecological vulnerability may assist further assessments of future viability of lemur species facing similar environmental settings.

In the third chapter, I examine the patterns of functional and neutral MHC similarity across sympatric and allopatric lemur species of the family *Cheirogaleidae* in an attempt

to determine which evolutionary mechanism, common ancestry or convergent evolution, are likely responsible for shaping MHC variation beyond genus level.

Using NGS, I first describe MHC class II variation of the two highly polymorphic loci (DRB and DQB) in four sympatric mouse and dwarf lemurs (*M. murinus*, *M. berthae*, *Cheirogaleus medius* and *Mirza coquereli*) and assessed the divergence in pathogen-driven selective pressures among these species by evaluating the overlap in their gastrointestinal parasite communities. Gastrointestinal macroparasites (e.g. helminths) are well suited to study the effects of pathogen-driven selection. Helminths represent the most prevalent macroparasite group of endoparasites (Weil et al. 2006) and especially gastrointestinal nematodes can have a pronounced impact on human and animal health (Stear et al. 1997; Mas-Coma et al. 2008). Although their effects are usually non-lethal, they may impose significant energetic demands, and consequently have detrimental effects on the overall host fitness, fecundity or even mortality rates in domestic as well as wild host species (e.g. Coyne and Smith 1994; Stien et al. 2002; Moran et al. 2006). Moreover, in wild populations, where non-invasive sampling is often required, screening of gut parasites has become a common approach to assess the level of parasitism of a population or community.

I then employ a bioinformatic approach to characterise MHC alleles according to the binding specificities of amino acid residues lining their peptide-binding groove (PBR) and classify functionally equivalent groups into so called ‘MHC supertypes’ (Sette and Sidney 1999; Southwood et al. 1998; Trachtenberg et al. 2003). This approach allows me to define functional overlap (similarities) between encoded MHC molecules that aids to clarify functional proximity between MHC alleles of different hosts. Using this classification, I then explore whether parasite-mediated selection is likely to contribute to the distribution of MHC supertypes across species by evaluating whether shared parasite pressures may select for shared MHC supertypes in different host species.

At last, I integrate MHC data of four sympatric *Cheirogaleidae* with those previously obtained for other allopatric populations of *M. murinus* and two other *Microcebus* sp. to investigate patterns of codon usage at functionally relevant (PBR) versus neutral sites of MHC alleles. Here I attempt to determine whether MHC allelic similarity at PBR (in case of supertype overlap) stems from independent convergence occurring after speciation event, or rather from common descent. I predict that if common

descent is responsible for allelic similarity at PBR across different host species, allelic structural (codon) similarity would also show higher similarity than expected by chance at adjacent neutral sites (non-PBR) that are not directly involved in antigen binding. In contrast, the operation of convergent evolution would be expected to promote only amino acid, but not codon similarity, at PBR, whereas phylogenetic signal would be retained at non-PBR (Klein et al. 2007; Lenz et al. 2013).

A community of four sympatric lemurs provides an ideal setting to investigate these questions. First, they are exposed to broadly similar environmental conditions, with slight differentiation in their diet (Dammhahn and Kappeler 2014), the level of social interactions and demography that vary to a certain degree among the four species regardless of their phylogeny (Dammhahn and Kappeler 2014; Dammhahn and Kappeler 2005, 2008a; Kappeler et al. 2002; Kappeler 2003; Eberle and Kappeler 2002, 2004, 2006; Fietz et al. 1999 a,b; Mittermeier et al. 2008; Markolf et al. 2008; Schäffler and Kappeler 2014). All these factors have been previously suggested to affect the level of parasitism and MHC variation at the population level to a certain level (e.g. Nunn et al. 2003; Vitone et al. 2004; Hughes and Page 2007; Chen et al. 2008; reviewed in Godfrey 2013; Morand 2015) and may provide an additional opportunity to evaluate the impact of ecological overlap on patterns of MHC allelic similarity. Second, previous studies have suggested parallel parasite-mediated selection affecting MHC supertype overlap in two of these species – *M. murinus* and *C. medius* (Schwensow et al. 2010b) as well as an effect of DRB genes on individual fitness and mate choice (Schad et al. 2005; Schwensow et al. 2007, 2008a,b, 2010a,b; Huchard et al. 2013; Sommer et al. 2014), supporting the assumption of adaptive value of the DRB in *Cheirogaleidae*. Third, variation in phylogenetic relatedness allows me to test the relative contribution of the two evolutionary scenarios in species pairs presenting a gradient of phylogenetic relatedness and may even allow me to estimate MHC allelic longevity. Fourth, the absence of ongoing reproduction minimizes the risk of confounding factors, such as allelic introgression (Klein et al. 2007; Lenz et al. 2013). Overall, identifying which evolutionary mechanism is more likely to contribute to maintain MHC allelic similarity across multiple speciation events will provide valuable information of the origin of MHC polymorphism as well as on the longevity of MHC alleles (or shared allelic motifs) across larger evolutionary time-scale.



© Eva Pechouskova / *Microcebus murinus*



© Eva Pechouskova / *Cheirogaleus medius*



© Eva Pechouskova / *Mirza coquereli*



© Eva Pechouskova / *Microcebus berthae*

# CHAPTER 1

## THE MAJOR HISTOCOMPATIBILITY COMPLEX AND PRIMATE BEHAVIORAL ECOLOGY: NEW TOOLS AND FUTURE QUESTIONS

Elise Huchard and Eva Pechouskova

This is Golden Open Access version

Originally published in:

*International Journal of Primatology* (2013) **35**:11–31. doi:10.1007/s10764-013-9700-1

Publisher: Springer

Copyright: © Springer Science+Business Media New York 2013

18





## **Abstract**

Since the serendipitous discovery of the effect of the Major Histocompatibility Complex (MHC) on mate choice in laboratory mice nearly 40 years ago, there has been sustained interest in the role that MHC genes may play in vertebrate sexual behaviour. However, the challenges posed by MHC genotyping have long hampered progress in this area. We briefly introduce the documented links between MHC and behaviour, before presenting an overview of the genotyping methods that were available before the introduction of new sequencing technologies. We then clarify why next-generation sequencing represents a major break-through in MHC genotyping by reviewing the recent successes - and pitfalls - of pioneer studies applying these techniques, before envisioning their revolutionary implications for future MHC studies in evolutionary ecology and primatology. We hope that our practical guidance to the design of MHC-based projects will promote and facilitate the integration of a MHC component into the research agendas of primatologists.

**Keywords:** Major Histocompatibility Complex, sexual behaviour, mate choice, kin discrimination, primates, next-generation sequencing, MHC genotyping

---

## Introduction

The important immune function of genes of the Major Histocompatibility Complex (MHC) was known well before the serendipitous discovery of their potential influence on behaviour in 1976 (Yamazaki et al. 1976). MHC molecules are cell surface glycoproteins responsible for recognizing foreign peptides (‘antigens’), and presenting them to specialized immune cells (such as lymphocytes) to initiate the appropriate immune response. MHC is a large cluster of highly polymorphic genes, and this variability is thought to be selected by the great variety of pathogenic agents (Doherty and Zinkernagel 1975; Apanius et al. 1997; Hughes and Yeager 1998). Accumulating evidence suggests that MHC genes have several non-immune functions, including a role in reproduction and materno-foetal interactions as well as a role in behaviour and mate choice thought to be mediated by olfactory communication (Alberts and Ober 1993; Potts and Wakeland 1993; Penn and Potts 1999). It is, however, still unclear whether sexual selection and reproduction play a significant role in maintaining MHC polymorphism. Due to their influence on a variety of fitness-related traits, such as disease resistance, reproduction and mate choice, and the complexity of the selective mechanisms that generate and maintain their polymorphism, MHC genes have emerged as a popular model for evolutionary biologists and behavioural ecologists (Apanius et al. 1997; Sommer 2005; Milinski 2006; Piertney and Oliver 2006). Here we will first outline the potential that studying MHC variability in wild populations offers for understanding animal behaviour, before explaining the technical challenges that have thus far restricted rapid progress in this field, and highlight novel approaches that will broaden the use and scope of MHC studies in behavioural ecology and primatology. We provide a glossary to guide the non-specialist reader through molecular biology terminology.

## **MHC and Behavior**

### **MHC and Mate Choice**

Evidence for MHC-dependent mate choice comes from all major vertebrate taxa. Various mating strategies have been described, including choice for partners possessing (i) compatible, (ii) diverse or (iii) particular MHC genotypes (Penn and Potts 1999; Milinski 2006; Yamazaki and Beauchamp 2007; Ruff et al. 2012).

In the first case, choosing a compatible partner to favour a good combination of genes in the offspring (Trivers 1972) may take the form of choice for maximally or optimally dissimilar partners, or for similar partners. Choice for MHC dissimilar partners may help to limit the deleterious effects of inbreeding or, alternatively, increase offspring diversity at specific MHC genes. A high MHC diversity may help fighting against a greater variety of pathogens ('heterozygote advantage': Doherty and Zinkernagel 1975) but may also increase the risk of reacting against self-derived peptides (and thus of autoimmune disease), so that optimal diversity may sometimes be more advantageous (Woelfing et al. 2009). Finally, choosing a similar mate may in particular contexts limit potential outbreeding costs.

In the second case, it is less clear why females may choose diverse (or heterozygous) partners, because paternal heterozygosity is not transmitted to offspring. Such a strategy may nevertheless be observed when females choose to mate with healthy partners if heterozygous individuals are more vigorous than homozygotes, or if it favours the transmission of rare MHC alleles to offspring because the possession of a rare allele is generally associated with heterozygosity at the considered locus (Apanius et al. 1997). This latter form of mate choice does not differ from MHC-disassortative mate choice.

In the third case, female choice may also target particular MHC genotypes conferring protection against dominant pathogenic pressures in a given environment. As possessing specific MHC genotypes may often be more important than possessing a high MHC diversity for parasite and disease resistance (Apanius et al. 1997), this form of MHC-dependent mate choice may be common. However, it is difficult to detect in nature, as testing for the effects of particular MHC alleles requires large sample sizes, and few

studies – none in primates - have so far been able to generate these large-scale tests (i.e., Eizaguirre et al. 2009b).

Primate studies have detected choice for partners possessing dissimilar (Schwensow et al. 2008a; Setchell et al. 2010), diverse genotypes (Sauermaun et al. 2001; Schwensow et al. 2008a; Schwensow et al. 2008b) or no mate choice at all (Huchard et al. 2010) suggesting that mate choice may be flexible and context-dependent (Setchell and Huchard 2010). For example, mate choice for dissimilarity may be favoured in inbred populations, while mate choice for particular genotypes may be favoured under high pathogenic pressure. Consequently, much remains to be done to understand the evolutionary drivers of MHC-dependent mate choice by measuring its fitness benefits across a range of environmental conditions and in populations exhibiting contrasted genetic structures.

### **MHC and Social Signaling: A Wider Behavioral Function of MHC?**

An obvious challenge in understanding MHC-biased mate choice is identifying its proximate basis, through modes of signalling and perception of genetic information among individuals. MHC contributes to the discrimination of self and non-self at the molecular level and may also contribute to individual and kin discrimination at the organismic level, with behavioural consequences that may extend further than mate choice, by affecting cooperation among kin (Brown and Eklund 1994; Ruff et al. 2012), parent-offspring recognition (Yamazaki et al. 2000) or even species diversification (Eizaguirre et al. 2009a). The role of MHC genes as determinants of sensed genetic individuality through olfactory cues has been highlighted in various vertebrate taxa (Penn 2002; Yamazaki and Beauchamp 2007). A MHC signature in odours has been identified by several studies in rodents (Willse et al. 2006; Kwak et al. 2008; Kwak et al. 2011) as well as in a primate species, the mandrill (*Mandrillus sphinx*, Setchell et al. 2011). In rodents, MHC similarity can be perceived beyond overall genomic similarity (Younger et al. 2001a; Yamazaki and Beauchamp 2007) and some studies suggest this may be the case in humans, too (Havlicek and Roberts 2009a). Although physiological pathways linking MHC genes and odour production are far from being understood at the organismic level (Kwak et al. 2011), the tight genomic linkage between MHC and a large cluster of olfactory receptor genes observed in humans and rodents raises the possibility of a

functional connection between both, where the possession of particular MHC genotypes might be associated with particular olfactory abilities or preferences (Ehlers et al. 2000; Younger et al. 2001a). This functional association is further supported by the activation of vomeronasal receptors by MHC derived peptides in rodents (Leinders-Zufall et al. 2004). More importantly, a recent study has demonstrated that particular HLA alleles directly influence the production of specific volatile organic compounds at the cellular level, leading to a cell-specific odour “fingerprint” (Aksenov et al. 2012).

It is, however, important to realize that MHC genes are not the only candidates for the genetic basis of chemical communication. For example, Major Urinary Proteins (MUPs) have been linked to the individual odour-based regulation of species-specific behaviour in mice (Hurst et al. 2001; Hurst et al. 2005; Cheetham et al. 2007) and genes coding for MUPs represent an equivalent polymorphic complex but with no implications for immune function. Comparing their respective effect on social and sexual behaviour would thus be very interesting (Thom et al. 2008). Recent findings of functional MUP genes in non-human primates such as mouse lemurs suggest that the role of MUPs in olfactory signalling is not limited to rodents, though functional MUP genes have not been detected in humans (Logan et al. 2008).

Overall, increasing evidence suggests that MHC genes directly influence socio-sexual behaviour. However, much of this evidence remains correlative, except for experiments performed on humans and rodents, and the ecological and evolutionary significance of MHC-dependent behaviour, as well as its universality, remains to be established.

## **MHC Genotyping: Historical Challenges and New Alternatives**

The single main obstacle to elucidating the role of MHC in behaviour, and more generally in population ecology, lies in the challenges of MHC typing. These challenges currently limit the scale of studies, include difficulties associated with low-quality DNA from non-invasive (fecal) samples and occasionally result in unreliable genotyping. The next section reviews the approaches adopted by molecular ecologists to overcome these problems, before introducing the contribution of new sequencing technologies.

---

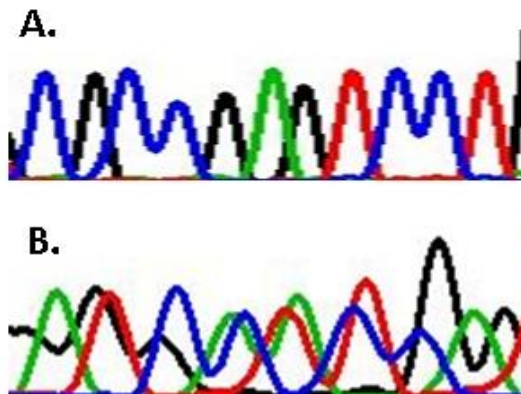
## Why Is MHC Notoriously Difficult to Genotype?

Three features of MHC genes pose major challenges to MHC typing (Bernatchez and Landry 2003; Piertney and Oliver 2006; Babik 2010): (i) its extreme allelic polymorphism, (ii) the presence of pseudogenes (non-functional genes) and (iii) the frequent gene duplications that have resulted in variation in the number of loci within and across species. The parts of MHC proteins involved in binding foreign antigenic peptides determine the functional differences between MHC alleles and frequently represent the regions of interest for behavioural ecologists, thus setting the targets for genotyping. They represent the most variable regions of MHC, which can count over 8500 alleles in humans (Robinson et al. 2011). This polymorphism precludes the application of molecular tools routinely available to type genetic polymorphisms, such as sequence-specific priming, except when extensive background information is available regarding MHC variability in the species of interest, as is the case in humans and some organisms used in biomedical research like laboratory rodents or rhesus macaques (*Macaca mulatta*). A widespread alternative consists in sequencing short fragments containing the antigen-binding sites but a critical difficulty there consists in designing primers which may amplify the full range of allelic variation (Babik 2010). Failing this step may cause investigators to miss an unknown fraction of the MHC variation in the genotyping process, which will cast doubts on the results of downstream analyses. Designing efficient primers requires characterization of the primer binding region and may take place as an incremental process consisting in cloning and amplifying the region of interest repeatedly until full characterization (e.g. Babik 2010; Sepil et al. 2012).

Second, identifying whether pseudogenes are present is required if the project goal is to characterize functional variation. It is sometimes possible to identify a pseudogene simply by examining its nucleotide sequence, for example if it contains a stop codon. However, not all pseudogenes show such features and their identification therefore requires ensuring that all alleles are expressed by comparing genotypes obtained by RNA (cDNA) and genomic DNA amplification. RNA should be extracted from tissues expressing the MHC molecules targeted, because certain MHC genes are only present in blood and lymphoid organs (Janeway et al. 2005). Although this step should represent a

standard for MHC studies (Knapp 2007), it is not always possible to acquire invasive samples from subjects of field studies.

Finally, MHC loci are frequently duplicated, resulting in multiple co-occurring copies in the genome. This means that DNA amplification for a single individual often yields more than the two sequences expected if this individual is heterozygous. The co-amplification of multiple alleles has traditionally represented the single main challenge to MHC genotyping. A PCR product containing an allelic mix cannot be sequenced using traditional sequencing methods (Sanger sequencing), as a mix of sequences renders the sequencing chromatogram unreadable (see Fig. 1). Sanger sequencing therefore requires a preliminary separation of the different alleles present in the PCR product of a given individual, and a variety of methods have been proposed to overcome this challenge. These are briefly detailed below.



**Figure 1.** A concrete example of the difficulties of MHC sequencing: two example chromatograms obtained from Sanger (traditional) sequencing of two MHC PCR products. Each coloured peak represents one of the four nucleotides (A, T, G or C). Panel A: chromatogram from a pure PCR product containing only one allele. Panel B: chromatogram from a PCR product containing an allelic mix. This is a problem because Sanger sequencing methods rely on reading a single nucleotide at each time interval.

## **What Are the Methods Commonly Used to Genotype MHC in Evolutionary Ecology?**

We aim to provide a brief, non-exhaustive, overview of the most common methods used to genotype MHC in organisms for which extensive background knowledge on MHC variability is unavailable, as is frequently the case for the species studied by field



---

primatologists (summarized in Table 1). More information can be found elsewhere (Babik 2010). The “gold standard” method of separating the different alleles present in the initial PCR product is cloning the PCR product. Molecular cloning involves inserting a particular fragment of DNA into the purified DNA genome of a self-replicating genetic element—generally a virus or a plasmid - which is then introduced into a bacterial cell and cultivated to generate a large population of bacteria containing identical DNA molecules (e.g. Alberts 2002). Consequently, sequencing multiple bacterial colonies for one given PCR product allows the identification of multiple alleles present in the PCR product because each colony “selects” one (and only one) allele at random. In concrete terms, cloning 10 to 20 PCR products (corresponding to 10 to 20 study subjects) may typically require a week of work and the sequencing of 30 to 50 bacterial colonies for each PCR product to ensure that all alleles present have been identified, which is financially costly.

As a result, alternative methods relying on sensitive electrophoresis have been developed which are substantially quicker and cheaper than cloning. These methods typically occur in two main steps. The first step consists of migrating the different alleles present in the initial PCR product on a gel, in order to separate them according to their differential nucleotide composition using Denaturing Gradient Gel Electrophoresis or DGGE (Fig 2) (Myers et al. 1987; Knapp et al. 1997; Knapp 2005b; Huchard et al. 2006) or according to the differential 3D conformation of a single DNA strand using Single Strand Conformational Polymorphism or SSCP (Orita et al. 1989; Sommer et al. 2002). The second step consists of excising each separate band (which corresponds to one allele) from the gel and re-amplifying it to sequence it. A gel may accommodate 10 to 20 PCR products (typically 10-20 study subjects) and preparing and running it may take a working day. Although more efficient than cloning, this method is still impractical for processing more than one hundred individuals. Electrophoretic methods are also difficult to optimize in order to obtain a resolution able to separate alleles differing by only a point mutation (Knapp 2005b), which are common in MHC genes. As a result, they often require extra separation steps, including cloning (Huchard et al. 2008).

Finally, a last approach consists of identifying microsatellite polymorphisms tightly linked to a variable MHC region of interest, for example located in the intron of the MHC gene targeted (Doxiadis et al. 2007; de Groot et al. 2008). This method can easily be

carried out on a large number of samples but requires considerable optimization to identify appropriate microsatellite polymorphisms and link them to the corresponding MHC allele.

**Table 1.** Summary table of the most common Major Histocompatibility Complex (MHC) genotyping methods used in non-model organisms. As a very rough indication of the financial costs, each € symbol represents 5-15€, and one sample represents one individual for one gene (so one individual for two genes would cost twice as much). A low optimization effort may represent few days to few weeks (less than a month) of work for one person, while a high optimization effort would likely represent more than that.

<b>Genotyping method</b>	<b>Cost per sample</b>	<b>Initial optimization effort</b>	<b>Throughput<sup>1</sup></b>	<b>High quality DNA required?</b>	<b>Example study in primates</b>
Cloning	€€€	Low	+	No	(Lukas et al. 2004)
Electrophoresis-based methods	€-€€ <sup>2</sup>	High	++	Yes	DGGE: (Knapp et al. 1997) SSCP: (Schad et al. 2004)
Microsatellite typing	€	High	+++	No	(Doxiadis et al. 2007)
Next-generation sequencing	€-€€ <sup>3</sup>	Low	++++ <sup>4</sup>	Possibly not	(Huchard et al. 2012)

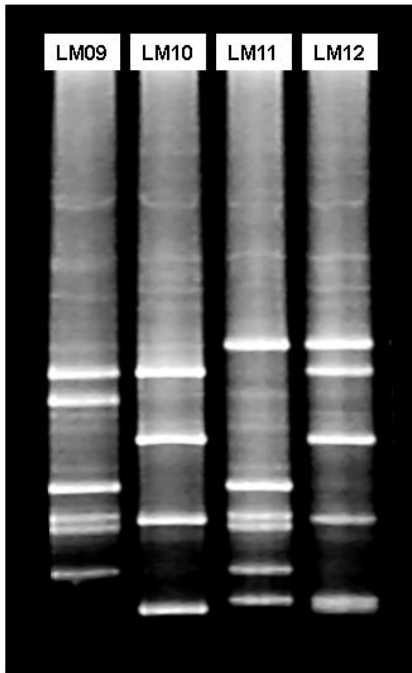
<sup>1</sup> Throughput indicates the potential of the method to handle large samples.

<sup>2</sup> These methods may turn out to be more onerous than initially thought when extra separation steps are required.

<sup>3</sup> Costs proportionally decrease as the number of samples increases, but are also conditional on the coverage required, which largely depends on the number of co-amplified loci.

<sup>4</sup> While time spent in the lab is usually much shorter than with other methods, time spent sorting alleles using bioinformatics and statistic tools is substantial, and may typically require a minimum of 3 months per project.

**Figure 2.** Example of an electrophoresis gel (DGGE) separating the different alleles of a mixed PCR product. PCR products of a variable fragment of MHC class II DRB are shown for four chacma baboons (*Papio ursinus*) identified by different codes with different MHC genotypes. The thickness of the lowest band of the individual LM12 reveals the limits of the resolution of this gel in separating alleles: two alleles co-migrated in this location, and additional separation steps were thus required to separate these sequences.



### What Is the Alternative Offered by New Sequencing Technologies?

The recent introduction of Next Generation Sequencing (NGS) to MHC genotyping shows a great deal of promise to circumvent some of the challenges inherent to the complexity of MHC structure. It greatly improves the scope of MHC studies by considerably extending the scale of the projects that can be undertaken. 454-technology has now been successfully applied to MHC genotyping in a number of studies (Babik et al. 2009; Wegner 2009; Galan et al. 2010; Zagalska-Neubauer et al. 2010; Sepil et al. 2012) including one in primates (Huchard et al. 2012). Similar to cloning in principle, but without the need for living recombinant organisms, NGS provides individual reads of each single DNA molecule present in a sample, thus avoiding the delicate step of separating the multiple alleles present in a PCR product (Shendure and Ji 2008; Metzker 2010). The fragment length that can be analysed using 454-pyrosequencing ranges from 30

approximately 100 to 500 base pairs, which represents the typical length of the most variable MHC exons that are routinely sequenced by MHC studies. We briefly describe below the basic principle of 454-sequencing that is useful for MHC genotyping and recommend some well-illustrated reviews for additional information regarding the principle and chemistry of next-generation sequencing (Shendure and Ji 2008; Metzker 2010). In a first step, the DNA of each individual is amplified with a set of tagged primers, that is primers that contain a unique label (in the form of 4 to 10 nucleotides) specific to each individual involved in the experiment. In a second step, PCR products from all individuals can be mixed (as they are identified by their individual tag) and the clonal amplification of each DNA molecule is performed by an “emulsion PCR”. During this amplification step, each single DNA strand from the DNA-mix is captured by a different bead and the bead-DNA complex is encapsulated into a water droplet in an oil solution. A PCR amplifies each DNA molecule enclosed into the “emulsion bubble” leading to thousand copies of the same DNA template in each bead. In a final step, the beads are deposited in a PicoTiterPlate for pyrosequencing. Each bead binds to one micro-well of the PicoTiter Plate (which contains around  $10^5$  to  $10^6$  wells depending on the sequencer model) and the PCR products enclosed in each bead are sequenced in parallel using a bioluminescence method. A sequencing run takes 4-6 hours and provides an output consisting of  $10^5$  to  $10^6$  sequences (commonly referred to as “sequencing reads”), each of which can be attributed to one individual thanks to its primer tag, allowing a list of the reads possessed by each individual to be established. The average number of reads per individual is referred to as the ‘coverage’ and is adjusted at the optimization stage: if a maximum of two alleles per locus is expected, a coverage of 20-40 reads / individual may be sufficient to ensure that both alleles are captured (Galan et al. 2010). If the number of gene copies targeted is greater, a higher coverage is required. Given that the total number of sequences per sequencing run is fixed, the coverage determines the number of individuals that can be sequenced in any given run. With a targeted coverage of 100 reads per individual,  $10^4$  individuals can be genotyped in one run. It is important to keep a comfortable safety margin when establishing a targeted coverage as there is extensive variation in the number of reads across individuals (and possibly alleles). One sequencing run, from the initial amplification stage to the final sequencing stage, typically takes one full week of work.

The utility of NGS technology comes at the cost of frequent genotyping errors. Pioneering studies have suggested that this difficulty may be overcome by stringent quality control, allowing true and false alleles to be sorted (Babik et al. 2009; Wegner 2009; Galan et al. 2010; Zagalska-Neubauer et al. 2010; Huchard et al. 2012; Sepil et al. 2012). The high frequency of false alleles is, in fact, not specific to next-generation sequencing and is common in any MHC genotyping method that relies on DNA amplification; it is very frequent in cloning (Huchard et al. 2012) and present in electrophoretic methods (Knapp 2005b). However, the high throughput of next-generation sequencing precludes manual allele sorting, and requires bioinformatics to automate this process. Although several authors have provided valuable guidance on how to proceed (Babik et al. 2009; Wegner 2009; Galan et al. 2010; Zagalska-Neubauer et al. 2010; Sepil et al. 2012), this task should not be underestimated and may take three months of work (or more) for a given project. It represents the main drawback of 454-sequencing applied to MHC genotyping, which otherwise has the key advantages of relatively low costs per sample, high throughput and rapid optimization (once primers are identified). It may also allow relatively smooth genotyping of the most complex multilocus systems with extensive allelic polymorphism (Zagalska-Neubauer et al. 2010; Sepil et al. 2012).

### **Designing an MHC Project: A Practical Guide**

The next section provides some practical advice to the design of a MHC project in behavioural ecology. It is broken into three sections focussing on identifying the most suitable study system, the best genotyping approach, and the most appropriate scale of projects in the time frame and funding conditions of a typical research project.

#### **(1) What Is the Ideal Study System?**

It is a species for which: (i) prior information on MHC variability and suitable primers are available for several MHC loci; (ii) high quality DNA can be obtained from a large number of recognizable individuals in the wild or in free-ranging populations; (iii) high resolution behavioural and life-history data can be obtained from the same individuals as well as, if possible, physiological and parasitological data (although the need for the latter obviously depends on the research question to

be addressed); (iv) multigenerational data are available as well as a molecular-based pedigree, which will allow validation of MHC genotyping and the investigation of multiple questions relating to MHC's influence on mate choice and patterns of parentage; and (v) a captive population of the same species is accessible so that some questions can be addressed using an experimental approach, allowing further mechanistic exploration of the processes under scrutiny. Although not many primate models score all these points, some of them do. The phylogenetic proximity of anthropoid primates to humans may facilitate the use of molecular tools developed on HLA, the human equivalent of the MHC, in many of these species (de Groot et al. 2002; Doxiadis et al. 2007). In addition, primatology counts a number of long-term studies documenting the behaviour of many individuals throughout their life (Kappeler and Watts 2012). There is certainly no need to have all these advantages to start a comprehensive MHC study, but combining a couple of them may considerably increase the scope of the questions that can be addressed.

## **(2) Which Genotyping Approach Should Be Adopted?**

The constraints and the opportunities of the study species should guide the approach used (see Table 1). In brief, and regardless of the method employed, difficulties in optimizing the genotyping are to be expected if only low quality DNA (from non-invasive samples) is available and if appropriate primers have to be designed. Low quality DNA will also require that most or all samples are genotyped in duplicate or triplicate to ensure reliability (Lukas et al. 2004; Knapp 2005a). The standards of the field are progressively moving away from studies that investigate variation in a single short DNA fragment, so future projects should plan to analyse variation in more than one MHC gene, especially since sequencing technologies now provide this possibility. Sequencing a higher number of variable loci will increase the resolution of measures of MHC similarity among partners, or of MHC diversity within individuals, and may help to identify the most important MHC regions for mate choice or social signals (Huchard et al. 2013). In studies where DNA is available from a large number of individuals, there may be a trade-off between the number of individuals and of loci to be sequenced, the optimal value of which will depend on the research question, the ease with which many loci can be sequenced,

and the extent of phenotypic data available for some or all individuals. Previous knowledge regarding the relative variability of different MHC loci in the target species will be helpful to identify the best genes to be studied. The topic of the project will also guide this decision. The MHC is divided into several main clusters, including the class I region, which is typically involved in resistance against intracellular pathogens such as viruses, and the class II region, which is typically involved in resistance against extracellular parasites such as macro-parasites and bacteria (Janeway et al. 2005). Whereas many primate studies have focussed on a few variable genes in MHC class II (typically DRB and DQB) which are known to be important in parasite resistance (Setchell and Huchard 2010), less is known regarding the ecological and evolutionary consequences of MHC class I variation, which is known to influence a number of reproductive functions including materno-foetal interactions (Alberts and Ober 1993), but notoriously more difficult to genotype in Old World Monkeys due to the complexity of its organisation (e.g., Otting et al. 2005).

Any large-scale project – e.g. more than 100 individuals – should turn to NGS. The number of genotyping platforms increases rapidly, and access to sequencing facilities should never be a limiting factor given that a full MHC project may only take 6 hours of sequencing time. Access to help for the bioinformatics following NGS may be more limiting. Regardless of the genotyping method, it is essential to evaluate the genotyping quality by running duplicates for a large subset of samples, or by comparing genotypes among parent-offspring dyads to estimate error rates. Baseline levels of MHC variability observed in the study population should be published along with a demonstration of the genotyping reliability, even if the primary interest of the principal investigator is in behavioural ecology (Huchard et al. 2008; Huchard et al. 2012; Sepil et al. 2012). This is facilitated by the fact that standards in the field of immunogenetics encourage investigators to describe new MHC alleles (for any given species) in publications detailing their characteristics, rather than simply depositing them in public databases. Unreliable MHC genotyping may cast doubts on downstream analyses, and the robustness of this field of research critically depends on the use of rigorous and transparent procedures.



**(3) What Scale Can Be Envisaged in the Realm of a Typical Research Project?**

MHC genotyping, as already emphasized, can be time-consuming and care should be taken to avoid underestimating the optimization stage – typically the most time-consuming part of the whole process – especially in conditions where difficulties are to be expected. Projects should use reasonably large sample sizes, and starting a MHC project with less than fifty individuals is risky as most analyses may fall short of power, except perhaps through an experimental approach. Genotyping 500-1000 samples in the context of a typical three-year project is realistic using NGS, although handling a high number of samples may require extra time at other stages of the project, such as genotyping these for neutral markers to establish pedigrees or to infer demographic influences on genetic variation. It is good practice to take the time necessary to validate and publish genotyping quality before moving on to the next stages, which should also factor into planned time schedules. Practically, although a large-scale MHC project can be planned in the context of a three-year project, it may be incompatible with extensive fieldwork, and it may thus be safer to use behavioural data that are already available. Enough time should be reserved for data analyses, which may typically require tools from adjacent fields including population genetics and molecular evolution, as well as sophisticated statistics. Acquiring a basic understanding of these disciplines is necessary and may be time-consuming.

**What is the Future of MHC Studies in Primatology?**

New sequencing technologies dramatically increase the scale and scope of studies aimed at elucidating the mechanistic bases and ultimate causes of MHC-biased behaviour, and promises rapid progress in the near future. The broad socio-ecological diversity found in the primate order (Kappeler and van Schaik 2002; Mitani et al. 2012) and the richness of primate socio-sexual behaviour and signalling (Dixson 1998; Kappeler 2002) offer opportunities to address important questions regarding the role of MHC genes in sexual and kin selection. In addition, resources available to primatologists, including some key molecular resources designed for human research and transferable to closely related

species (Tung et al. 2010), a wealth of comparative data across primate species and populations, as well as a number of multigenerational individually-based field studies (Kappeler and Watts 2012) place primates at the forefront of future research. Here, we highlight some specific questions and topics to which the study of primates may provide valuable contributions.

### **Test the Links Between MHC and Behavior at the Genomic Scale**

The sequencing of the full genome of several primate species has been achieved (Tung et al. 2010) and is ongoing for several others. This information will be very helpful for the design of densely spaced markers throughout the MHC region and the whole genome. As a result, primate studies may stand at the forefront of the number of MHC genes and markers examined. Marker design will offer insights into whether specific MHC loci exert an influence on behaviour, and may, in turn, help to identify the pathway linking genotype and phenotype, as specific MHC genes may have specific functions, for example in reproductive physiology. In addition, these influential MHC loci may be linked to other functional genes causing the observed patterns given the high level of linkage disequilibrium observed in the MHC region (Horton et al. 2004), and genomic scans may help to detect effects of this kind. For example, the tight genomic linkage observed between MHC and olfactory receptor genes in humans and rodents (Younger et al. 2001b) may suggest that genetic diversity or dissimilarity at olfactory receptor genes may generate observed behavioural biases attributed to MHC genes.

In addition, it is important to contrast MHC variation measured over an extensive region with wider genomic diversity using markers of the same nature (e.g., Single Nucleotide Polymorphisms) to determine whether observed behavioural effects are MHC specific or not (see e.g. Chaix et al. 2008; Laurent and Chaix 2012 for an example of this design in humans). This remains an outstanding question in MHC evolutionary ecology, as most studies have examined the links between behaviour and a small MHC fragment (Huchard et al. 2010) without necessarily controlling for wider genomic diversity. When they have controlled for it, they have often used crude estimates in the form of multilocus heterozygosity measured over a handful of microsatellite loci (Balloux et al. 2004). As a result, it is often impossible to exclude the possibility that MHC-disassortative mate

choice is a simple by-product of inbreeding avoidance based on MHC-independent cues. Genome-wide studies are costly and may only be applicable to a subsample of individuals, so documenting the effect size of the MHC-dependent behavioural bias of interest using a traditional genotyping approach on the mastersample may provide a good starting point to calculate the minimum number of individuals to be included in the genome-wide study.

### **Understand How Social and Mating Systems Drive MHC-Associated Mate Choice**

The social diversity of primates, including solitary, pair-living and group-living taxa, and the corresponding diversity in mating systems, with monogamous, polyandrous, polygynous and polygynandrous mating patterns (Kappeler and van Schaik 2002) offers an opportunity to examine the universality and context-dependence of mate choice rules (Setchell and Kappeler 2003). Individual mating strategies appear to be constrained by group size and social structure. For example, alpha males may largely monopolize reproduction in mixed sex groups of polygynous species (Cowlshaw and Dunbar 1991; Port and Kappeler 2010) and compromise the expression of mating preferences of subordinate males and females. In addition to the social system, demographic effects on population genetic structure may influence the expression of individual mating strategies. For example, research in humans suggests that MHC-disassortative mate choice only operates in close, inbred populations (Chaix et al. 2008) and further efforts are needed to understand the links between population genetic structure and the evolution and plasticity of mate choice. Primatologists can play a key role in addressing these questions, because primate societies have been comparatively well-studied and provide a wealth of detailed data on behaviour and ecology (Setchell and Kappeler 2003).

### **Investigate Variations in MHC-Dependent Mate Choice Within and Across Individuals, and Their Consequences**

Detailed individually-based longitudinal datasets developed by primatologists over decades can be used to measure within-individual consistency in MHC-dependent mate

choice, and to identify factors driving variation, such as fluctuations in social context (Gowaty 2004) (e.g., the pool of available partners) or reproductive state (Havlicek and Roberts 2009b). Some large-scale studies can be used to examine the extent and causes of inter-individual variation in mating decisions, as well as the fitness consequences of MHC-dependent mate choice, which are far from being established, and even further from being measured (Setchell and Huchard 2010). Primatologists can also pinpoint the mechanisms mediating MHC-biased mate choice. Whereas most published studies of MHC-dependent mate choice rely on indirect measures of mate choice based on biased paternities, it is now important to understand whether these biases occur prior or after copulation and detailed records of mating behaviour available in many primate studies can shed light on this question. In addition, the primate order counts some of the brightest and most colourful sexual signals found in mammals, and recent work has highlighted potential connections between the possession of particular MHC genotypes, individual condition and the intensity of these signals in primates (Setchell et al. 2009; Huchard et al. 2010b). Future studies may attempt to confirm these intriguing findings using larger sample sizes.

### **Test the Role of MHC in Kin Discrimination**

Whereas kin recognition has long been thought to be mediated by familiarity through stable bonds created during early development (Waldman 1988), recent studies suggest that individuals may recognize and preferentially associate with unfamiliar relatives such as paternal kin in promiscuous societies where paternity is uncertain (reviewed by Widdig 2007). These observations suggest alternative mechanisms of kin discrimination, potentially through self-referent phenotype matching, the comparison between own and other's phenotypes (Widdig 2007). Many cues reflect relatedness in non-human primates, including visual appearance (Alvergne et al. 2009; Parr et al. 2010; Kazem and Widdig 2013), vocalisations (Rendall et al. 1996; Kessler et al. 2012) and odours (Charpentier et al. 2008; Célerier et al. 2010). Despite a long held belief that monkeys and apes are microsmatic and that their main communicatory channels rely on visual rather than olfactory cues (Heymann 2006), recent developments in the study of primate olfactory communication have revealed that scents reflect genome-wide diversity and genetic

relatedness in ring-tailed lemurs *Lemur catta* (Charpentier et al. 2008; Charpentier et al. 2010) and MHC diversity and dissimilarity in mandrills (Setchell et al. 2011). Testing whether the MHC plays a role in kin discrimination is thus an important goal for future research, in primates and other animals. The best evidence so far comes from a recent study in *Xenopus laevis* showing that within sibships, tadpoles associate preferentially with MHC-similar full-sibs, with a positive linear relationship between MHC similarity and the intensity of association among partners (Villinger and Waldman 2012). In addition, new findings indicate that humans can discriminate the odour of self from non-self MHC peptide ligands (Milinski et al. 2013). Establishing the role of MHC in kin discrimination in non-human primates will nonetheless represent a challenging task, because it requires showing that affiliation or association among unfamiliar individuals depends on their MHC similarity, independently of their degree of kinship (which may be reflected in other phenotypic cues). To reach that goal, studies will have to compare the social interactions of unfamiliar individuals who exhibit a similar level of kinship but vary in MHC similarity. Experimental designs in captive colonies or comparisons of interactions among paternal kin in long-term field studies of polygynandrous primates – where familiarity is an unreliable paternity cue - may be able to tackle this challenge.

### **Elucidate a Potential Role of MHC in Species Diversification**

One of the major riddles in our understanding of the evolution of socio-ecological diversity represents the identification of traits responsible for the regulation of gene flow across populations. Sexual selection, along with selection against hybridization or ecological selection, has been considered to be of major importance in this process (reviewed in Hoskin and Higgie 2010). Adaptive traits linking mate choice to environmental changes, sometimes referred to as “magic traits” (Gavrilets 2004), may play a crucial role in the speciation process if they contribute to the translation of adaptation to a new environment into a shift in mating preference. MHC genes have been proposed to represent “magic genes” (Eizaguirre et al. 2009a). Several studies have suggested that pools of MHC genes are shaped by local pathogen communities that may vary across populations (Wegner et al. 2003; Blais et al. 2007; Eizaguirre et al. 2009a) and that mate choice favouring locally adapted MHC genotypes might consequently

accelerate reproductive separation among populations (Eizaguirre et al. 2009b; Eizaguirre et al. 2012). This framework represents a new angle linking MHC to biodiversity at an evolutionary scale. Several primate radiations, featuring several cryptic species, incomplete lineage sorting, or hybridization *in situ* (Weisrock et al. 2010) provide an excellent opportunity to unravel the tangled influence of genetic and ecological factors in leading to reproductive isolation among closely related species.

### **Conclusions**

MHC figures among the first set of functional genes that have been identified as influencing behaviour in laboratory rodents. Sustained research efforts invested in understanding the mechanisms and evolutionary function of this intriguing connection have long been hampered by the challenges posed by MHC genotyping. The first studies to apply NGS to MHC genotyping successfully have thus opened a major technological lock. NGS dramatically expands the potential scope of studies, which can now be undertaken at the genomic and population scale. This opens exciting opportunities to resolve old outstanding questions, such as understanding whether, and to what extent, MHC causes behavioural variation, and whether, and to what extent, sexual selection contributes to the generation and maintenance of MHC polymorphism. NGS will also help to exploit the full potential of longitudinal individually-based primate field studies that have run over decades, and represent the historical strength of primatology. Primatologists thus have key assets at hand to pioneer the next generation of studies addressing the links between MHC and behaviour.

**Box 1 – Glossary.**

**454- pyrosequencing** (Roche diagnostics): A method of DNA sequencing by synthesis, using a large-scale parallel pyrosequencing system capable of sequencing roughly 400-600 megabases of DNA per 10-hour.

**ABS (Antigen-Binding Site):** The physical location of MHC molecules where antigens bind and are presented to T cells.

**Antigen:** Any molecule recognized by an antibody or T-cell receptor.

**cDNA (complementary DNA):** Single-stranded DNA produced from an RNA template, lacking introns (introns are the non-coding parts of genes).

**DGGE (Denaturing Gradient Gel Electrophoresis):** a technique used for separation of DNA fragments according to their mobility under increasingly denaturing conditions.

**Electrophoresis:** A technique used to separate DNA, RNA, or protein molecules based on their size and electrical charge.

**Exon:** The sections of a gene that are translated into a protein.

**Haplotype:** Set of alleles, which are found in adjacent locations (loci) on a single chromosome and inherited as a physically linked set from each parent.

**HLA (Human Leucocyte Antigen):** The human major histocompatibility complex.

**Heterozygosity:** Two different alleles at a locus in diploid organisms, resulting from inheritance of different alleles from each parent.

**MHC (Major Histocompatibility Complex):** A complex of tightly linked genes coding for molecules involved in the detection of non-self antigens and their presentation to specialized immune cells such as lymphocytes.

**Molecular cloning** involves inserting a particular fragment of DNA into the purified DNA genome of a self-replicating genetic element, generally a virus or a plasmid, which is then introduced in a bacterial cell and cultivated to generate a large population of bacteria containing identical DNA molecules.

**MUPs (Major Urinary Proteins):** A subfamily of a larger protein family called lipocalins. MUPs are found in abundance in the urine and other secretions of many animals. They are encoded by a cluster of genes, located adjacent to each other on a single stretch of DNA, varying in number between species: from about 21 functional genes in mice to none in humans.

**NGS (Next-Generation Sequencing):** DNA sequencing method, extending basic principles of Sanger sequencing across millions of reactions in a massively parallel fashion. Rather than being limited to a single or a few DNA fragments as in Sanger sequencing, NGS generates hundreds of gigabases of data in a single sequencing run.

**PCR (Polymerase chain reaction):** Technique used to amplify DNA sequences using specific primers.

**Polymorphism (in genetics):** presence of different alleles at a given locus.

**Primer:** Short DNA sequences used to target the part of the genome to be amplified in PCR.

**Primer binding region:** A region of DNA where a single-stranded primer binds to start replication; or a duplication of a complementary DNA sequence during PCR.

**Pseudogenes:** Gene or allele that is not coding for a functional protein because it contains mutations that disrupt the sense of the sequence.

**RNA (Ribonucleic Acid):** A single-stranded molecule similar to DNA, often resulting from DNA transcription and performing multiple vital roles in the coding, decoding, regulation, and expression of genes.

**Sanger sequencing:** A method of DNA sequencing, based on a selective incorporation of chain-terminating radioactively or fluorescently labelled dideoxynucleotides by DNA polymerase during *in vitro* DNA replication. The resulting DNA fragments are heat-denatured and separated by size using gel electrophoresis.

**SSP (Sequence-Specific Priming):** Also known as allele-specific PCR, a molecular typing method consisting in designing PCR primers that amplify one (and only one) specific allele of a polymorphic gene. Genotyping relies on using all specific-primers designed for each known allele to identify which allele(s) is/are possessed by the individual.

**SNP (Single Nucleotide Polymorphism):** A type of polymorphism involving variation of a single base pair.

**SSCP (Single-Strand Chain Polymorphism):** A conformational difference of single-stranded nucleotide sequences of identical length, a characteristic which allows distinguishing the sequences by means of gel electrophoresis (separating different conformations).

**Stop codons:** A trinucleotide sequence within a messenger RNA (mRNA) molecule signalling a halt to protein synthesis.



**Acknowledgements**

We thank Lauren Brent and Amanda Melin for inviting us to contribute to this timely Special Issue as well as Dieter Lukas, Lauren Brent and two anonymous reviewers for providing detailed and valuable comments on an earlier draft. We are also very grateful to Christian Roos, Christina Albrecht and Lutz Walter for introducing us to 454-technology, Christina Oberdieck for her help in the lab while undergoing large-scale genotyping, Markus Brameier for his expertise in bioinformatics, Peter Kappeler for hosting our research and contributing the large-scale dataset that formed the basis of our recent MHC research and Susanne Schliehe-Diecks for genotyping over 800 mouse lemurs for a set of microsatellites within a few months (and without the help of NGS). EH was funded by a DFG research grant (number HU 1820/1-1) when applying NGS to MHC, and EP is funded by a SAW project, the Leibniz Graduate School (SAW-2011-DPZ-4).

# CHAPTER 2

## MHC CLASS II VARIATION IN A RARE AND ECOLOGICAL SPECIALIST MOUSE LEMUR REVEALS LOWER ALLELIC RICHNESS AND CONTRASTING SELECTION PATTERNS COMPARED TO A GENERALIST AND WIDESPREAD SYMPATRIC CONGENER

Eva Pechouskova, Melanie Dammhahn, Markus Brameier,  
Claudia Fichtel, Peter M. Kappeler and Elise Huchard

Published in:  
*Immunogenetics* (2015) 67:229-245. DOI: 10.1007/s00251-015-0827-4

Publisher: Springer Berlin Heidelberg  
Copyright: The Author(s) 2015. This article is published with Open Access at  
Springerlink.com



## Abstract

The polymorphism of immunogenes of the major histocompatibility complex (MHC) is thought to influence the functional plasticity of immune responses and, consequently, the fitness of populations facing heterogeneous pathogenic pressures. Here, we evaluated MHC variation (allelic richness and divergence) and patterns of selection acting on the two highly polymorphic MHC class II loci (DRB and DQB) in the endangered primate Madame Berthe's mouse lemur (*Microcebus berthae*). Using 454 pyrosequencing, we examined MHC variation in a total of 100 individuals sampled over 9 years in Kirindy Forest, Western Madagascar, and compared our findings with data obtained previously for its sympatric congener, the grey mouse lemur (*Microcebus murinus*). These species exhibit a contrasting ecology and demography that were expected to affect MHC variation and molecular signatures of selection. We found a lower allelic richness concordant with its low population density, but a similar level of allelic divergence and signals of historical selection in the rare feeding specialist *M. berthae* compared to the widespread generalist *M. murinus*. These findings suggest that demographic factors may exert a stronger influence than pathogen-driven selection on current levels of allelic richness in *M. berthae*. Despite a high sequence similarity between the two congeners, contrasting selection patterns detected at DQB suggest its potential functional divergence. This study represents a first step towards unravelling factors influencing the adaptive divergence of MHC genes between closely related but ecologically differentiated sympatric lemurs and opens new questions regarding potential functional discrepancy that would explain contrasting selection patterns detected at DQB.

**Keywords:** Primates, *Cheirogaleidae*, *Microcebus berthae*, 454 pyrosequencing

## Introduction

The island of Madagascar, one of the world's biodiversity hotspots, has faced rapid deforestation over the last century, resulting in population fragmentation of many endemic primates (Lemuriformes) (Ganzhorn et al. 2001; Mittermeier et al. 1992; Schwitzer et al. 2013a,b), 94% of which are currently classified as threatened (Schwitzer et al. 2013a). Many of these species are highly arboreal, inhabit restricted biogeographic ranges and exhibit fast life histories and higher population turnover compared to most anthropoid primates; however, the average life span is thought to be compromised by high extrinsic mortality pressure in the wild populations (Kraus et al. 2008; Fichtel 2012; Kappeler 2012). These characteristics make lemurs particularly vulnerable to an ongoing habitat degradation that may disrupt gene flow and cause local demographic fluctuations. This may in turn result in an irreversible loss of genetic diversity in lemurs, especially those species with restricted spatial distribution. Given the ongoing rate of deforestation, rapid surveys of remaining genetic diversity and its potential consequences for the future viability of these populations are essential to determine conservation priorities (Kremen et al. 2008). Although difficult to detect in small populations (Chikhi et al. 2010), signals of demographic or genetic bottlenecks or decreasing genetic diversity at neutral markers seem present across lemur taxa (Louis et al. 2005; Olivieri et al. 2008; Markolf et al. 2008; Radespiel et al. 2008; Craul et al. 2009; Razakamaharavo et al. 2010; Holmes et al. 2013; Baden et al. 2014). However, little is known regarding its potential fitness consequences and the capacity of such populations to maintain an effective level of functional (adaptive) genetic variation (allelic richness and divergence), sufficient to ensure their health and survival.

Genes of the major histocompatibility complex (MHC) are well suited to study the adaptive maintenance of genetic variation given their immune function (Klein 1986) and extreme polymorphism (e.g. Hedrick 2002; Garrigan and Hedrick 2003; Sommer 2005; Piertney and Oliver 2006; Spurgin and Richardson 2010). MHC genes are coding for MHC molecules (cell surface glycoproteins) that trigger the immune response by presenting antigens at antigen-binding sites (ABS) to T-lymphocytes, which then activate further components of the immune system. Each MHC molecule can bind only a limited array of antigens, and a number of mechanisms may increase the spectrum of antigens

recognized, including gene duplications, extensive allelic diversity, polymorphism at ABS (Hughes and Nei 1988), or polymorphism outside ABS that may alter the 3D positioning of ABS contact residues (Bjorkman and Burmeister 1994).

Balancing selection, primarily exerted by pathogen pressure, is thought to represent one of the major forces driving MHC polymorphism (reviewed in Bernatchez and Landry 2003; Sommer 2005; Piertney and Oliver 2006; Milinski 2006; Spurgin and Richardson 2010). Three non-mutually exclusive evolutionary mechanisms have been proposed to explain this polymorphism. First, due to the codominant expression of MHC alleles and the function of MHC molecules in the immune response, MHC heterozygous individuals may be at an advantage in a population facing heterogeneous pathogen pressures (i.e. heterozygote advantage; Doherty and Zinkernagel 1975). Second, MHC alleles that are advantageous in protection against dominant pathogens in a given environment (Apanius et al. 1997) may temporarily rise in frequency, until pathogens evolve resistance to the most common host alleles, which are then progressively replaced by rarer alleles (i.e. frequency-dependent selection; Snell 1968; Bodmer 1972; Borghans et al. 2004). Finally, parasite communities typically vary in space and time and may thereby select distinct sets of MHC alleles in host populations (i.e. fluctuating selection; Hedrick 1999; Spurgin and Richardson 2010; Eizaguirre et al. 2009a, 2012). As a result, patterns of selection on MHC genes are likely to be durably affected by the diversity and frequency of MHC alleles present in the host population and, hence, by its demography (Hedrick 1972; Borghans et al. 2004). The interruption of gene flow across fragmented populations and the reduction of effective population sizes can therefore result in a loss of genetic diversity through genetic drift and inbreeding (Wright 1969; Keller and Waller 2002; Frankham et al. 2002) and may disrupt balancing selection (Hughes and Yeager 1998) that might consequently compromise population capacity to respond to changing pathogenic pressures (O'Brien and Evermann 1988).

In support of this assumption, a possible link between highly divergent or specific MHC genotypes and individual fitness (Schad et al. 2005; Schwensow et al. 2007, 2010a, b; Sommer et al. 2014) and mate choice (Schwensow et al. 2008a, b; Huchard et al. 2013) has been suggested for two widely distributed lemurs – *Microcebus murinus* and *Cheirogaleus medius* (*Cheirogaleidae*, Primates). Consequently, the evaluation of MHC variation retained in populations of endangered confamilial species can provide valuable

---

insights to assess their viability. The complexity of MHC genotyping has long impaired detailed genetic studies of free-ranging species with unknown genomic organisation. Next-generation sequencing (NGS) technologies promises progress in this area by overcoming some of the technical difficulties associated with the complexity of MHC genotyping and by allowing cost-effective processing of large datasets (reviewed in Babik 2010; Huchard and Pechouskova 2014; Koboldt et al. 2013; Lighten et al. 2014a).

Here, we investigated MHC variation (allelic richness and divergence) and patterns of selection of two highly polymorphic MHC class II genes, DRB and DQB, in the endangered Madame Berthe's mouse lemur, *Microcebus berthae* (B1ab.i-iii, Andriaholinirina et al. 2014), by genotyping a total of 100 individuals sampled over 9 years in three study areas. This world's smallest primate (ca. 30g) is endemic to the dry forests of the Menabe region in western Madagascar (Schmid and Kappeler 1994; Ganzhorn et al. 2001; Schäffler and Kappeler 2014), which has recently been identified as a priority site for conservation (Schwitzer et al. 2013a). The distribution of *M. berthae* is restricted to an area of less than 810km<sup>2</sup> within two forest fragments and a narrow corridor connecting them (e.g. Rasoloarison et al. 2000; Schäffler and Kappeler 2014). Its population density varies across its geographic range (30-100 individuals/km<sup>2</sup>) and seems to be affected by habitat heterogeneity and anthropogenic disturbances (Schwab and Ganzhorn 2004; Schäffler and Kappeler 2014).

Next, we compared our findings with data obtained across a similar temporal scale from a population of *M. murinus*, a sympatric congener (e.g. Weisrock et al. 2010), which in comparison to *M. berthae* presents several key ecological and demographic differences. *M. berthae* is a feeding specialist relying mostly on dispersed fast depleting resources, such as homopteran secretions or arthropods. This feeding strategy is thought to promote an intense scramble competition leading to spatial avoidance, overdispersion and lower rate of social interactions among conspecifics (Dammhahn and Kappeler 2009). In contrast, the more opportunistic feeding niche of *M. murinus*, including diverse plant and animal matter, seems to reduce competition and facilitate spatial proximity and social interactions among conspecifics (Dammhahn and Kappeler, 2008a, b, 2010). The ecological flexibility of *M. murinus* is also reflected by its wide distribution across southern and western Madagascar. Its population density is higher and population size larger in Kirindy forest than those of *M. berthae* (Eberle and Kappeler 2004; Dammhahn

and Kappeler 2005, 2008b). We expect these inter-specific contrasts to influence MHC variation. First, given that allelic richness appears to be a function of effective population size (Hedrick 1985), we expect MHC allelic richness of *M. berthae* to be lower than in *M. murinus*. Second, the larger population of *M. murinus* might harbour a more diverse array of pathogens (Anderson and May 1978; Nunn et al. 2003; Hughes and Page 2007; see also Altizer et al. 2007), and this effect might be enhanced by a broader feeding niche and more frequent encounters with conspecifics offering a greater chance of pathogen encounter and transmission. In contrast, low population densities and low rates of social interactions in combination with a narrow feeding niche could result in relaxed pathogen-mediated pressure in *M. berthae*. As such, we predict to detect weaker tracks of pathogen-driven selection on *M. berthae* MHC alleles compared to *M. murinus*.

## Methods

### Sample collection and DNA extraction

DNA samples were collected from *M. berthae* from three sub-populations captured between 2005 and 2013 using baited Sherman life traps set within 25-ha study areas (N5, CS7 and Savannah) located within a 12.500-ha concession of Kirindy Forest of the Centre National de Formation, d'Etude et de Recherche en Environnement et Foresterie (CNFEREF) de Morondava (Madagascar: 44° 39' E, 20° 03' S, Kappeler and Fichtel 2012). The centres of the study areas N5-CS7 and Savannah-CS7 are situated ca 2-2.5 km and N5-Savannah ca 4-4.5 km away from each other, respectively. In total, we collected samples from 100 individuals, with sample sizes reflecting contrasting population densities and sampling effort across the years at each study area (Electronic supplementary material, see Appendix 1 pp. 102-108: ESM 1; N5:  $n=80$ ; 42♂/37♀/1 n.a.; CS7:  $n=14$ , 1♂/13 n.a.; Savannah:  $n=6$ , 3♀/3♂). At first capture, each individual was briefly immobilised with 10 µl Ketanest 100 (s.c.) (Rensing 1999), individually marked with a sub-dermal microtransponder (Trovan, Usling, Germany) for other studies (e.g. Dammhahn and Kappeler 2005; 2008a,b; 2010), and a small ear biopsy of 2-3 mm<sup>2</sup> was taken and preserved in 70 % ethanol. Genomic DNA was extracted from ear biopsies following standard protocol (Qiagen QIAmp DNA Mini-Kit, Qiagen Germany). We have



adhered to the Guidelines for the Treatment of Animals in Behavioral Research and Teaching (Animal Behavior 2006, 71: 245-253) and the legal requirements of the country (Madagascar) in which the fieldwork was carried out.

### **454 library preparation**

PCR amplification targeting the two loci of the most variable parts of the MHC class II region, DRB and DQB, was performed using primers that flank the functionally important ABS and captured the full variability in the congener *M. murinus* (Schad et al. 2004; Averdam et al. 2011). PCR reaction mix and amplification conditions are summarised in ESM 2. Each individual PCR product (further referred to as amplicon) was electrophoresed on 1 % agarose gel to verify successful amplification. Primer design and the preparation of locus-specific amplicon libraries were described elsewhere (Huchard et al. 2012). Sequencing was conducted according to standard protocols for GS Junior sequencing (Roche, 454 pyrosequencing). All sequencing reads retrieved from a total of six sequencing runs were processed according to a post-sequencing quality control procedure following Huchard et al. (2012).

### **454 library processing**

#### *Allelic discrimination and evaluation of the number of loci*

Artefactual alleles introduced by PCR or sequencing errors and assessing the sequencing depth necessary for reliable genotyping are well-known technical challenges associated with NGS that might compromise the reliability of assessments of MHC polymorphism (reviewed in Babik 2010; Lighten et al. 2014a).

Here, we adjusted some of the filtering steps proposed by previous authors (e.g. Babik et al. 2009; Galan et al. 2010; Zagalska-Neubauer et al. 2010; Huchard et al. 2012) to discriminate true versus artefactual alleles, relying on two central assumptions: (1) Artefactual alleles should show high similarity to one of the two parental sequences they originated from within amplicons, either by single point mutation or indels causing a shift in the reading frame, or by recombination of the two parental sequences (chimeras), and (2) artefactual alleles should be relatively rare, compared to true alleles, both across and

within amplicons. In contrast to studies mentioned above, we did not identify artefactual alleles based on global allelic frequency thresholds established across amplicons but evaluated allele status both across and within each amplicon.

Both manual alignment (Multalin; Corpet 1988) and two numeric indices of allelic frequency were used to critically evaluate a potential number of loci and to discriminate true alleles from artefacts: (1) the mean per amplicon frequency (MPAF) of any given allele as the proportion of reads from an amplicon assigned to this allele, averaged across all amplicons possessing this allele, and (2) the relative per amplicon frequency (RPAF) of each allele as the proportion of reads retrieved for each given allele within a given amplicon. We predicted that artefactual alleles should be relatively rare, compared to true alleles across amplicons (reflected by low MPAF) and within amplicons (reflected by low RPAF). While MPAF can help to identify artefacts returned at low frequencies across sequencing runs, RPAF can help to identify heterogeneities in the distribution of within-amplicon allelic frequency. These can be either artefactual alleles that occur non-randomly across sequencing runs or at high frequencies only in few amplicons, skewing their MPAF (i.e. run-specific sequencing errors or homopolymers) (see also Lenz and Becker 2008; Harismendy et al. 2009; Gilles et al. 2011; Sommer et al. 2013; Lighten et al. 2014a, b), or cross-amplicon contaminations of true alleles (with high MPAF and low RPAF). The occurrence of such DNA carryover contaminants has recently been acknowledged as an underrated source of genotyping errors associated with NGS (Huchard et al. 2012; Li and Stoneking 2012; Sommer et al. 2013; Lighten et al. 2014a, b). Here, we did not systematically eliminate these suspected contaminants by using a fixed threshold for a minimum frequency per amplicon (here referred to as RPAF) under which alleles are filtered out within amplicons (e.g. Sepil et al. 2012; Huchard et al. 2012), to avoid eliminating true alleles and thereby generating potential allelic dropout (van Oosterhout et al. 2006; Sommer et al. 2013). Rather, the status of amplicons suspected of contaminations by true alleles was clarified by replicating affected amplicons.

In the first step of the allele sorting procedure, we attempted to evaluate whether target genes may be duplicated in order to assess the sequencing depth required to ensure reliable genotyping. Although there was no indication of a loci duplication in *M. murinus* (Averdam et al. 2011; Huchard et al. 2012), we could not assume the same in *M. berthae*

due to extensive variation in the genomic organisation of MHC within and across species (e.g. Kelley et al. 2005; Winternitz and Wares 2013; Lighten et al. 2014b). Therefore, we investigated the MPAF and RPAF distribution of the most common to the least common alleles across all amplicons, assuming that in the case of non-duplication, we would detect a notable drop between the two most common and remaining alleles (see also Babik et al. 2009; Huchard et al. 2012). Next, we evaluated our findings by manual alignment of all alleles within each amplicon and attempted to discriminate artefactual from true alleles based on our assumptions - similarity to parental allele and low MPAF and RPAF. When the status of alleles remained ambiguous, affected amplicons were replicated. Finally, we replicated those few amplicons that passed allele sorting with more true alleles than expected given the estimated number of loci to check whether this came from the genuine locus duplication or from an artefact.

#### *Assessment of minimum sequencing depth and genotyping reliability*

Based on the estimated number of loci, we proceeded with a final screening step to evaluate the sequencing depth necessary for accurate genotyping using the program 'Negative Multinomial' developed by Galan et al. (2010). Amplicons that did not return a sufficient amount of reads based on this estimate were re-genotyped.

The efficiency of allele sorting was then evaluated through two steps. First, assuming that artefacts originated during a single PCR reaction or introduced by a genotyping mistake would not occur independently in many amplicons, we investigated the relationship between the MPAF of each allele and the number of amplicons possessing this allele. We then correlated both values before and after allele sorting. Here, we expected a positive relationship before allele sorting - due to the fact that artefacts should display both low MPAF and be retrieved in few individuals only - that would disappear after allele sorting (Babik et al. 2009; Huchard et al. 2012). Second, a set of amplicons were additionally replicated for each loci within independent sequencing runs (DRB  $n=23$ ; DQB  $n=27$ ) to assess the reliability of our genotyping for each locus within.

## Sequence analysis and phylogeny reconstruction

All true alleles were retained for downstream analysis, including those that were possessed by one individual only to prevent the elimination of rare alleles leading to an underestimation of allelic richness and to avoid generating false homozygotes. However, these alleles were not submitted in public repositories.

Allelic divergence was evaluated by computing average pairwise distances (number of differences) among all pairs of nucleotide and amino acid sequences in each locus in MEGA 6 (Tamura et al. 2013). Evolutionary relationships between amino acid sequences of both loci found by this study in *M. berthae* and previously in *M. murinus* were constructed in MEGA, using a neighbour-joining algorithm with Poisson correction (Saitou and Nei 1987; Zuckerkandl and Pauling 1965). *Mimu* sequences originated from 664 individuals captured within the study area CS7 between the years 2000 and 2010. These *Mimu* alleles were retrieved using 454 technology and comparable allele validation steps (see Huchard et al. 2012). The repeatability of sequence alignment was determined by a bootstrap analysis with 1000 replications.

Allelic richness of both loci detected within the study population of *M. berthae* was compared to those previously described for *M. murinus* (Huchard et al. 2012). To evaluate the number of alleles detected for a given sampling effort, we conducted a permutation test in R ([www.r-project.org](http://www.r-project.org)). Here, we randomly selected 10 individuals and counted the number of distinct alleles detected. This procedure was repeated 100 times to calculate a mean and SD for each sampling effort, adding 10 individuals at each step until 100. This procedure was conducted for both *M. berthae* and *M. murinus* and plotted for both loci to illustrate the number of distinct alleles detected per given sampling effort.

## Population genetic analysis

Linkage disequilibrium between DRB and DQB loci was tested using a likelihood ratio test, where the null hypothesis of no association between loci (linkage equilibrium) is compared to the hypothesis of a possible association (Slatkin and Excoffier 1996). The significance of the procedure is found by computing the null distribution of this ratio under the hypothesis of linkage equilibrium using 10,000 permutations implemented in

---

Arlequin 3.5.1.3. (Excoffier and Lischer 2010). Using GENEPOP v 1.2 (Rousset and Raymond 1995), a null allelic frequency was estimated by maximum likelihood estimation (EM algorithm; Dempster et al. 1977), and deviations from Hardy-Weinberg equilibrium (HWE) were calculated for each locus separately using the exact U-score test (Rousset and Raymond 1995), with the alternative hypothesis predicting heterozygote excess. The extent of genetic differentiation among sub-populations (regardless of sex and year cohort) was examined by pairwise  $F_{ST}$  at both loci using Arlequin (10,000 permutations, Wright 1965). Comparisons between adult males and females and across year cohorts were not included due to uneven and small sample sizes (see above and ESM 1).

### **Test of positive selection**

The presence of positively selected sites (PSS) was investigated in both genes separately. PSS are characterised by  $\omega > 1$  with  $\omega = d_N/d_S$  and  $d_N$  and  $d_S$  being the relative amounts of substitutions at non-silent ( $d_N$ ) and silent ( $d_S$ ) codon sites. First, we investigated the strength of positive selection using the likelihood ratio modelling approach. We compared two models of the codon evolution: the null model, where  $\omega < 1$  and varies according to the beta distribution (model M7), and a model allowing an additional class of sites, where  $\omega > 1$ , to account for a possible occurrence of PSS (model M8) using a likelihood ratio test (LRT) (in Yang et al. 2000). If model M8 fits the data better than M7, PSS were identified through Bayes Empirical Bayes (BEB) procedure and retained for evaluation by the next step if statistically significant (Yang et al. 2005), using the package CodeML implemented in PAML 4.7 (Yang 2007).

As a second approach, we estimated values of  $d_N$  and  $d_S$  and their standard errors by calculating the pairwise number of silent and non-silent substitutions (Nei and Gojobori 1986) applying Jukes-Cantor correction for multiple hits implemented in MEGA. This rather conservative approach considers all possible evolutionary pathways (excluding termination codons) leading from one codon to another as equally probable and is thereby expected to provide conservative (minimum) estimates of numbers of substitutions compared to the positive selection hypothesis  $\omega > 1$  (Nei and Kumar 2000). A codon based Z-test of selection was performed to test whether both PSS ( $\omega > 1$ ) and non-

PSS ( $\omega < 1$ ), identified by the previous approach using BEB procedure, were under positive selection. Furthermore, we compared  $\omega$  in ABS and non-ABS. Their location was derived from referential human sequences (HLA-DQB, HLA-DRB; Bondinas et al. 2007) and compared with previously detected PSS and non-PSS. Finally, overall values of  $d_N$  and  $d_S$  were calculated.

## Results

### Allelic discrimination and evaluation of the number of loci

A total of 321 unique sequences were retrieved for DRB from 148 amplicons (including 49 replicates, see below) and 105 sequences for DQB from 130 amplicons (including 32 replicates), in the range of 162-170bp (excluding primers). For more details and statistics of sequencing outcome, see ESM 3. The distribution of MPAF and RPAF from the first to sixth most common alleles averaged across all amplicons revealed a notable drop of frequency between the two most common and remaining alleles, suggesting no duplication of either locus (ESM 4).

Within 321 DRB sequences, 286 (89%) displayed  $MPAF < 0.05$  and were identified by manual alignment as artefacts (95 %) or contaminants (5 %), following our criteria. From the remaining 35 (11 %) alleles with  $MPAF \geq 0.05$ , 13 alleles were identified as artefacts. Additionally, five alleles occurring within a single amplicon ( $MPAF$  0.05-0.08) were identified as contaminants and discarded (see below). The remaining 17 alleles ( $MPAF$  0.08-0.69) were retained as true alleles.

In DQB, 74 (70 %) of all 105 sequences displayed  $MPAF < 0.05$ , and all of those were eliminated as artefacts (95 %) or contaminants (5 %). Among 31 remaining sequences with  $MPAF \geq 0.05$ , four more sequences were eliminated as artefacts and five as contaminants ( $MPAF$  0.06-0.24). The remaining 22 sequences were retained as true alleles. In both loci, artefacts with  $MPAF \geq 0.05$  mostly represented homopolymer indels - occurring either at inconsistent frequencies across sequencing runs, within a single sequencing run or amplicon - and their elimination was further supported by running replicates of affected amplicons. Nine out of 14 contaminants were eliminated by running

replicates and three in DRB and two in DQB occurred within one amplicon that was due to extensive inflation of artefacts, and thereby inconclusive genotype, excluded in each locus. Overall, 304 (95 %) sequences in DRB and 83 (79 %) in DQB were eliminated. After allele sorting, only three (out of 96 individuals) and six (out of 98 individuals) amplicons had more than two true alleles in DRB and DQB, respectively. In all of them, the existence of a third and in one case fourth allele could be excluded by re-sequencing affected amplicons, which further confirmed our assumption that both loci were non-duplicated.

### **Assessment of minimum sequencing depth and genotyping reliability**

Based on our conclusions of no loci duplication, we used the probabilistic model (Galan et al. 2010) and estimated that a minimum of 18 reads were required for reliable genotyping of each given locus with a confidence level of 0.95. All amplicons with <18 reads (DRB 26; DQB 5) were re-genotyped and replaced in the dataset, except of three amplicons in DRB that did not return >18 reads in the second genotyping attempt. Additionally, 23 amplicons for DRB and 27 for DQB were genotyped in replicates to estimate genotyping reliability and all of them showed a perfect reproducibility of assigned genotypes.

The correlation between the MPAF of each allele and the number of amplicons possessing this allele was significant before allele sorting in both loci (Pearson's correlation, DRB  $n=321$  alleles,  $r=0.78$ ,  $P<10^{-15}$ ; DQB  $n=105$  alleles,  $r=0.69$ ,  $P<10^{-15}$ ), largely driven by the presence of alleles that had both low MPAF and low frequency (ESM 5). This correlation disappeared in DQB after discarding artefacts (Pearson's correlation  $n=22$  alleles,  $r=0.22$ ,  $P=0.33$ ) and was weakened in DRB, though still significant (Pearson's correlation, DRB  $n=17$  alleles,  $r=0.64$ ,  $P<0.006$ ).

### **Sequence analysis and phylogeny reconstruction**

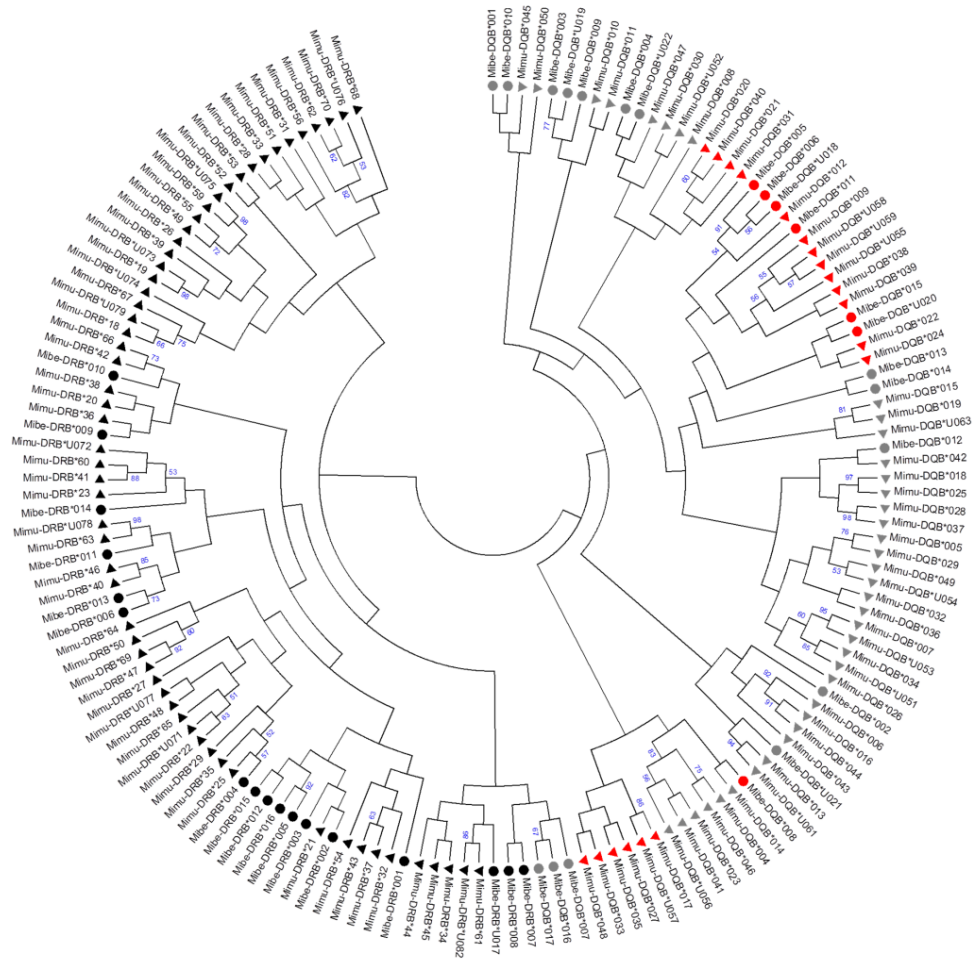
From the total of 17 *Mibe*-DRB and 22 *Mibe*-DQB sequences found in 96 and 98 individuals of *M. berthae*, respectively, none have been described previously. Accession numbers of these alleles as well as the full nucleotide sequence of alleles occurring in

only one individual are given in Electronic supplementary dataset, see Appendix 1 pp. 109-111.

Average nucleotide divergence (number of differences) between sequences was comparably high in both loci (mean $\pm$ SD; DRB 17.07 $\pm$ 2.44, DQB 14.44 $\pm$ 2.16). Among *Mibe*-DRB sequences, we found 47 (29 %) variable nucleotide sites and 35 (21 %) among *Mibe*-DQB sequences. Each nucleotide sequence of both loci translated into a unique amino acid sequence and the absence of stop codons suggests that all sequences can encode functional proteins. Amino acid sequences revealed 23 (43 %; DRB) and 23 (41 %; DQB) variable sites out of 54 and 56 sites (see below), and comparable amino acid divergence was found in both loci (mean $\pm$ SD; DRB 11.73 $\pm$ 2.0, DQB 10.76 $\pm$ 1.82).

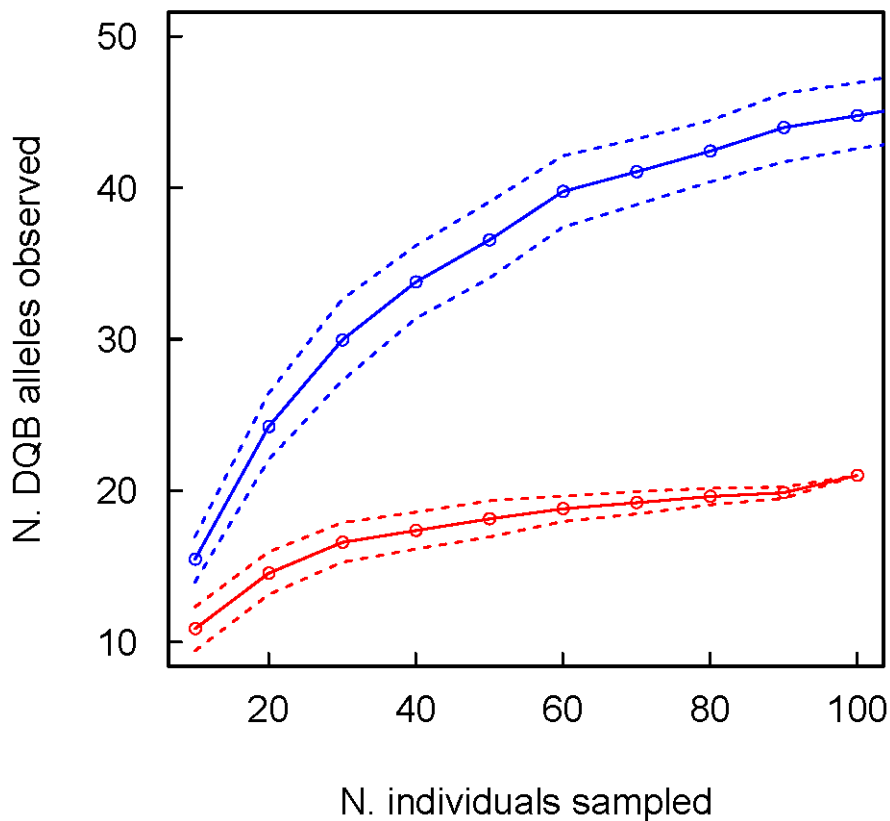
The reconstruction of evolutionary relationships between amino acid sequences of both species revealed two distinct loci-specific clusters, with the exception of three *Mibe*-DQB sequences (*Mibe*-DQB\*017, \*016 and \*007), that clustered separately among DRB sequences of both *M. berthae* and *M. murinus* (Fig. 1). These three *Mibe* sequences were retrieved independently at least in two and up to 11 individual amplicons across different sequencing runs, and their presence and loci identity was confirmed by replicating from one to three amplicons possessing affected sequences. Moreover, these three sequences clustered with other DQB sequences when including only *Mibe* sequences in the analysis (data not shown). There was no clear separation between sequences of the two species in either locus (Fig. 1), and the level of amino acid divergence was comparable within (see above and Huchard et al. 2012) and between allelic pools of the two species (*Mimu* vs. *Mibe*-DRB 11.6 $\pm$ 2.1; DQB 11.3 $\pm$ 1.9). Additionally, an insertion (two codons) causing fragment length polymorphism in DQB, homologous to the one previously described in 19 *Mimu*-DQB sequences (Huchard et al. 2012), was detected in 7 *Mibe*-DQB sequences. This insertion did not result in a shift of the reading frame, and there was no evidence for stop codons that would indicate a loss of function. These sequences created large distinct clusters, except for one *Mibe* sequence (*Mibe*-DQB\*008, Fig. 1).





**Fig. 1** Evolutionary relationships between amino acid sequences for 17 *Mibe*-DRB (black circles) and 22 *Mibe*-DQB sequences (grey circles for sequences without 6-bp insertion and red circles for sequences with the insertion) described in this study, including 59 *Mimu*-DRB (black triangles) and 58 *Mimu*-DQB sequences (grey triangles for sequences without two-codon insertion and red triangles for sequences with it) described in Huchard et al. (2012). The tree configuration was derived using neighbour-joining algorithm (Bootstrap 1000; Poisson correction) in MEGA 6. Only bootstrap values exceeding 50 % are shown. Accession numbers and nucleotide sequences of *M. berthae* are presented in Appendix (Table 3)

Allelic richness (number of distinct alleles) estimated for a given sampling effort (number of sampled individuals) through re-sampling procedure in *M. berthae* and *M. murinus* is shown in Fig. 2. The estimated mean of distinct alleles detected per sampling effort is lower in *M. berthae* in both loci, indicating lower overall allelic richness for a given sampling effort in this study population.



**Fig. 2** Estimation of allelic richness for a given sampling effort through re-sampling procedure, showing the number of distinct alleles detected when randomly drawing an increasing number of individuals from our sample in *M. berthae* (red) and *M. murinus* (blue). Given the similarity of the observed pattern between DRB and DQB, only the plot for DQB loci is shown. The *dotted lines* indicate the standard deviation around the estimated mean (*solid line*)

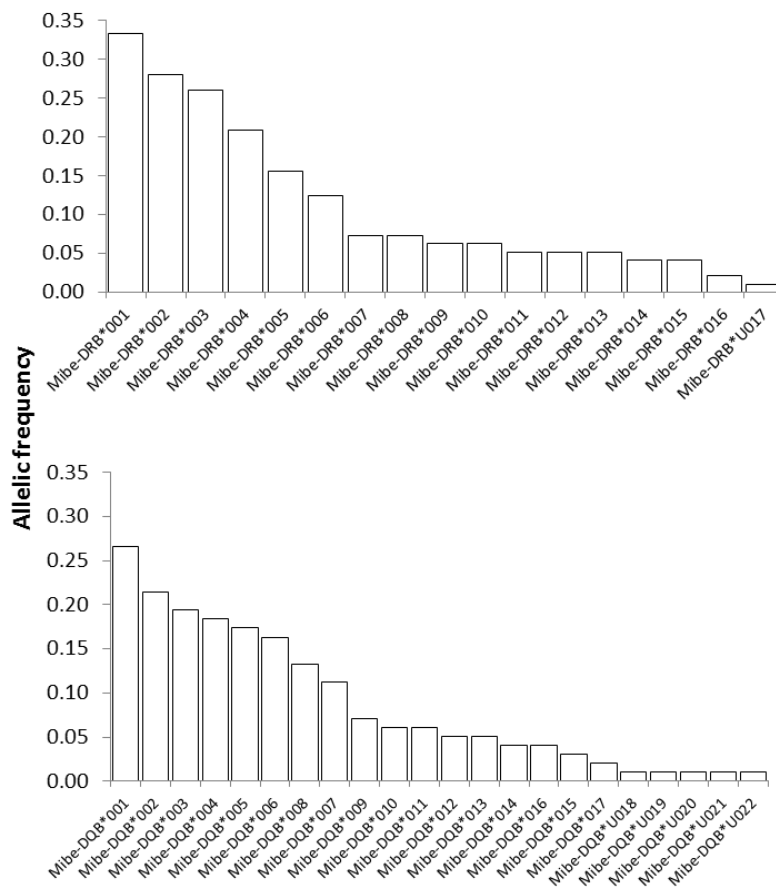
---

### Population genetic analysis

The null hypothesis of linkage equilibrium between loci could be rejected ( $\chi^2=810.33$ ,  $df=396$ ,  $P<10^{-06}$ ). Allelic distribution patterns were relatively similar in both genes across all sub-populations with allelic frequencies varying widely within each locus from 1 to 33 % in DRB and 1 to 27 % in DQB (Fig. 3).

The estimated frequency of null alleles was low in both loci ( $<0.01$ ). The comparable observed and expected level of heterozygosity could be found at both loci across all sub-populations (DRB,  $H_O=0.91$ ,  $H_E=0.90$ ; DQB,  $H_O$  and  $H_E=0.92$ ), and heterozygote excess was not detected in either locus (DRB,  $Fis^{W\&C}=-0.007$ ,  $P=0.58$ ; DQB,  $Fis^{W\&C}=0.005$ ,  $P=0.80$ ).

Pairwise comparisons did not reveal any genetic differentiation in either loci among the three sub-populations (for all pairs,  $F_{ST}<0.001$ ;  $P=0.49-0.84$ ), suggesting intact/ongoing gene flow among them. The allelic frequencies within each sub-population and across year cohorts of the largest sub-population (N5) are shown in ESM 6.



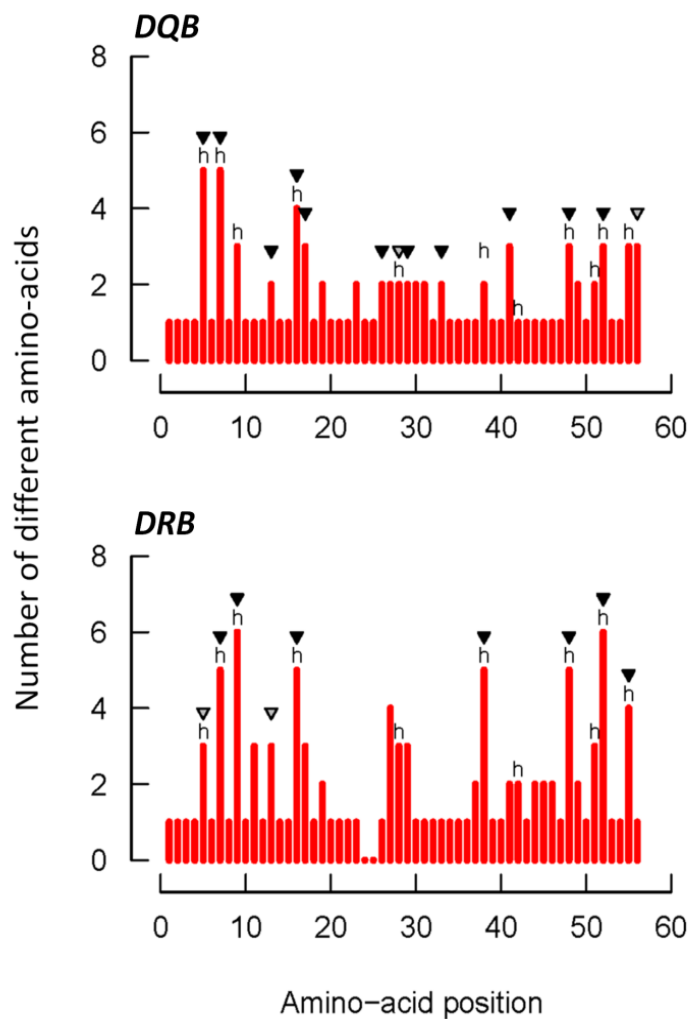
**Fig. 3** The distribution of allelic frequencies (i.e. rate of occurrence) of 17 MHC-DRB and 22 MHC-DQB alleles within the study population of *M. berthae* (DRB,  $n_{ind}=96$ ; DQB,  $n_{ind}=98$ )

### Test of positive selection

The significant deviation of the LRT statistics from a  $\chi^2$  distribution allowed rejection of the null model assuming neutral evolution (M7) in favor of a model allowing for a class of sites being subjected to diversifying selection (M8) for both loci (Table 1). In DRB, nine PSS were identified (CI 99 %,  $n=7$ ; CI 95 %,  $n=2$ ). Eight of those occurred at homologous positions with HLA-ABS and another one was located within a three amino acid distance (Fig. 4). In DQB, 13 PSS were detected (CI 99 %,  $n=11$ ; CI 95 %,  $n=2$ ). Six out of 13 PSS were homologous to HLA-ABS, seven other located within one to four amino acid distance (Fig. 4). In comparison, two out of three DQB-PSS described in *M.*

*murinus* (Huchard et al. 2012) were homologous to those identified in *M. berthae* (positions 5 and 16). Additionally, six PSS were homologous between the two species in DRB.

The estimated values of  $d_N$  and  $d_S$  ( $\pm$ SE) through evolutionary pathways method (Table 2) confirmed the results of the first analysis by revealing an elevated  $d_N$  relative to  $d_S$  in all ABS versus non-ABS, and in all PSS versus non-PSS. Codon-based Z-tests of selection indicated that all PSS and ABS have been affected by positive selection in both loci.



**Fig. 4** Amino acid variation plots for *Mibe*-DQB and *Mibe*-DRB alleles. Human antigen-binding sites (ABS) are indicated with the letter *h* (Bondinas et al. 2007), and positively selected sites (PSS) are indicated by *black* ( $P > 99\%$ ) and *grey triangles* ( $P > 95\%$ ). The insertion of two codons at positions 24-25 in seven DQB alleles causes a gap in sequences of DRB loci

**Table 1** Evaluation of the goodness of fit for different models of codon evolution and estimated parameter values

Model	LnL <sup>a</sup>	Kappa (ts/tv)	AIC	$\Delta$ AIC <sup>b</sup>	Parameters
<b><i>MHC-DRB</i></b>					
M0-one $\omega$	-922.27	0.79	1846.12	184.38	$\omega^c=0.61$
M7-nearly neutral with $\beta^g$	-839.49	0.86	1680.7	18.96	
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ ) <sup>h</sup>	-830.14	0.73	1661.74	Best	$p_0^d=0.78, p_1^e=0.22, \omega^f=3.01$
<b><i>MHC-DQB</i></b>					
M0-one $\omega$	-873.89	1.34	1750.45	150.84	$\omega=1.29$
M7-nearly neutral with $\beta$	-810.82	1.66	1624.97	25.36	
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ )	-798.3	1.5	1599.61	Best	$p_0=0.73, p_1=0.27, \omega=4.24$

AIC Akaike information criterion, Kappa (ts/tv) transition/transversion rate

<sup>a</sup> Log likelihood of a model

<sup>b</sup> Difference between the value of the AIC of a given model and the best model

<sup>c</sup>  $d_N/d_S$

<sup>d</sup> Proportion of sites with  $\omega \leq 1$

<sup>e</sup> Proportion of positively selected sites ( $\omega > 1$ )

<sup>f</sup> Estimated value of  $\omega$  for sites under positive selection

<sup>g</sup> For all sites,  $\omega \leq 1$  and the  $\beta$  distribution approximates  $\omega$  variation

<sup>h</sup> Proportion of sites evolves with  $\omega > 1$

**Table 2** Results of evolutionary pathways method (MEGA 6) to estimate values of  $d_N$ ,  $d_S$  ( $\pm$ SE) for ABS and non-ABS defined by homology with HLA, and for PSS and non-PSS identified by Bayes Empirical Bayes (BEB) analysis (PAML).

Positions	Number of codons in each category	$d_N$	$d_S$	Z	P
<b>MHC-DRB</b>					
ABS	11	0.59 $\pm$ 0.09	0.05 $\pm$ 0.04	5.038	0.000
Non-ABS	43	0.05 $\pm$ 0.02	0.06 $\pm$ 0.03	-0.470	0.639
PSS	9	0.75 $\pm$ 0.08	0.17 $\pm$ 0.14	3.226	0.002
Non-PSS	45	0.06 $\pm$ 0.02	0.04 $\pm$ 0.02	0.484	0.629
All	54	0.13 $\pm$ 0.03	0.06 $\pm$ 0.03	1.921	0.057
<b>MHC-DQB</b>					
ABS	11	0.35 $\pm$ 0.11	0.01 $\pm$ 0.01	3.123	0.002
Non-ABS	43	0.07 $\pm$ 0.02	0.04 $\pm$ 0.02	1.493	0.138
PSS	13	0.50 $\pm$ 0.09	0.04 $\pm$ 0.04	4.912	0.000
Non-PSS	41	0.04 $\pm$ 0.01	0.03 $\pm$ 0.02	0.783	0.435
All	54	0.12 $\pm$ 0.03	0.03 $\pm$ 0.02	3.174	0.002

From Nei and Gojobori 1986, Bondinas et al. 2007, Yang et al. 2005.

P probability of  $d_N=d_S$  using Z-test of selection. All positions containing gaps and missing data were eliminated

## Discussion

This is a first description of the polymorphism of two MHC class II genes in the endangered Madame Berthe's mouse lemur. We detected a total of 17 *Mibe*-DRB and 22 *Mibe*-DQB unique sequences which showed high divergence and tracks of past positive selection. Below, we compare patterns of variation and selection in *M. berthae* with those previously described for *M. murinus*, a closely related sympatric congener, that differs in several key aspects of its demography and ecology and discuss potential implications of these results on population viability.

### MHC variation and selection patterns

A total of 17 *Mibe*-DRB and 22 *Mibe*-DQB unique sequences were detected within members of the three sub-populations of *M. berthae* in Kirindy forest, Western Madagascar. Comparison with 59 DRB and 58 DQB sequences of *M. murinus* obtained previously from the same study site (Huchard et al. 2012) revealed no clear separation of amino acid sequences between the two species in either locus (Fig. 1). Nucleotide and amino acid sequence similarity between alleles of the two species, as well as the presence of a two-codon insertion located at the same position in 19 *Mimu*- and 7 *Mibe*-DQB sequences, could indicate the retention of MHC motifs in both loci during periods of time exceeding the evolutionary split between species (trans-species polymorphism; Klein, 1987). MHC sequence similarity limited to exons encoding peptide binding regions have been detected many times between species that were sometimes distantly related (summarised in Klein et al. 2007; Lenz et al. 2013) and may represent examples of trans-species polymorphism or result from convergent evolution (independent evolution of similar traits in response to similar ecological pressures) (e.g. Klein et al. 2007). In addition, the reconstruction of evolutionary relationships between amino acid sequences of the two loci revealed a partial paraphyly, with three *Mibe*-DQB sequences clustering within DRB sequences. A BLAST search of the affected *Mibe*-DQB nucleotide sequences revealed *Mimu*-DQB sequences as a close match. Moreover, when analysed separately, *Mibe* sequences generated two distinct loci-specific clusters (data not shown) supporting their correct assignment to DQB loci. Additionally, these sequences occurred within



---

multiple individuals and their presence was confirmed by replication, which excluded the possibility of sequencing run-specific amplification mismatch. Such paralogy has been also reported previously in *M. murinus* by Huchard et al. (2012) and may result from a combined effect of tight physical linkage, shared origin and high functional similarity between the two loci.

In contrast to 17 *Mibe*-DRB and 22 *Mibe*-DQB sequences obtained by this study from the total of ca. 100 individuals sampled over 9 years in three study areas, 59 *Mimu*-DRB and 58 *Mimu*-DQB originated from 664 individuals of *M. murinus* sampled over a comparable period of time within a single study area (Huchard et al. 2012). Allelic richness (number of distinct alleles) for a given sampling effort (number of sampled individuals) that was estimated through re-sampling procedure in dataset of both species revealed the average number of distinct alleles detected for a given sampling effort to be ca. two-fold lower in *M. berthae* compared to *M. murinus* (Fig. 2). This finding indicated lower allelic richness in *M. berthae*, where sampling of ca. 60 individuals, compared to ca. 200 individuals in *M. murinus*, would allow to capture most alleles present in the study population, deduced from an inflection in the graph illustrating the relationship between allelic richness and sampling effort (see Fig. 2 and Huchard et al. 2012). Moreover, the allelic distribution across year cohorts within the largest sub-population of *M. berthae* (N5) suggests that reported allelic richness within this study site may be overestimated, since some alleles were detected exclusively within earlier cohorts and seem to have disappeared within current generations (e.g. after 2008-2009) (ESM 6). However, sample size is too small to interpret apparent fluctuations in allelic frequencies that might be further enhanced by a progressive displacement of the *M. berthae* population, located at the periphery of the study area (Dammhahn and Kappeler 2008b), or by individual migrations. The lower allelic richness found in *M. berthae* matched our predictions based on its overall lower population densities and population size relative to *M. murinus* (Dammhahn and Kappeler 2005, 2008b). Even though the reasons for lower population density of *M. berthae* are unknown, factors such as a narrow feeding niche promoting intense intra-specific scramble competition, larger home ranges and less cohesive social networks in *M. berthae* (Dammhahn and Kappeler 2005; 2008a; 2009; 2010), when compared to spatially more clumped generalist *M. murinus* (e.g. Eberle and Kappeler 2004), are thought to contribute to the naturally lower population densities in

this species (Dammhahn and Kappeler 2008b). Additionally, we observed a notable decrease in the number of newly captured individuals across the years in the most densely populated study area (N5) despite of a comparable capture effort across years (ESM 1). This pattern could either be the result of a decreasing population size or, alternatively, a spatial exclusion from the study area by its superior competitor (Dammhahn and Kappeler 2008b), whose sub-population is shifting in recent years into the areas previously exclusively occupied by *M. berthae* (data not shown).

Finally, the small population size, specialist diet and lower rate of social interactions among conspecifics of *M. berthae* could to some extent promote a limited array of pathogens and its transmission across conspecifics (reviewed in Edwards and Potts 1996; Nunn et al. 2003; Vitone et al. 2004; Rifkin et al. 2012). This could in turn result in relaxed selection, possibly manifested not only by lower allelic richness and/or divergence but also by less tracks of selection on MHC sequences. In *M. berthae*, allelic divergence in both loci, as well as strong evidence of past historical balancing selection on MHC sequences (Table 1 and 2; Fig. 4), are comparable to patterns described in *M. murinus* (Huchard et al. 2012) and do not support the idea of a weaker pathogen-driven selection in *M. berthae* compared to *M. murinus*. Thus, population size rather than weak selection seems to constrain allelic richness in this population.

In addition, nine PSS were detected across 17 *Mibe*-DRB sequences, and 13 across 22 *Mibe*-DQB sequences, suggesting that DQB may be of equal or higher functional importance than DRB in this species. This contrasts with previous findings in *M. murinus*, where DRB was suggested to be under stronger diversifying selection than DQB based on their relative number of PSS (11 vs. 3) (Huchard et al. 2012). Contrasting selection patterns could reflect divergent functions of this locus in *M. berthae* versus *M. murinus*. Under this scenario, we may also expect different levels of allelic variation (richness and divergence) between the two loci. This is the case in neither *M. murinus* nor *M. berthae*. This may be due to the fact that (i) selection pressures acting on both loci are not independent given their tight linkage, (ii) allelic variation reflects variation in demography and not simply selection, or (iii) allelic variation and signatures of past positive selection reflect the strength of selection over different time scales. An elevated rate of non-synonymous mutations requires a long time to accumulate (Bryja et al. 2007)

---

as well as to vanish after the disappearance of selection (Garrigan and Hedrick 2003), whereas fluctuations in allelic variation may be more dynamic.

### **Implications for population resistance**

To assess the adaptive significance and fitness consequences of MHC variation, it is essential to distinguish the relative importance of different measures of MHC polymorphism (Garamszegi and Nunn 2011). The high level of amino acid divergence among *Mibe* alleles or their effective combination within individual genotypes may buffer potential detrimental effects of lower allelic richness for pathogen resistance. Thus, persistence of certain *Mibe* alleles across several generations and study areas (ESM 6) could be facilitated by their divergence (i.e. divergent allele advantage hypothesis) (Richman et al. 2001; Schwensow et al. 2010a; Lenz et al. 2009; Lenz 2011; Froeschke and Sommer 2012; Sepil et al. 2013) or by the effect of MHC-dependent mate choice favoring specific alleles conferring resistance against dominant pathogens (e.g. Hill et al. 1991; Schad et al. 2005; Schwensow et al. 2007, 2010a; Axtner and Sommer 2012; Kloch et al. 2013). However, high allelic divergence may not be sufficient to maintain effective flexibility of the immune response in the long-term when allelic richness is low. In small populations with limited gene flow, genetic drift may weaken the capability of balancing selection to maintain high levels of MHC polymorphism through disappearance of rare allelic variants (Hartl and Clark 1997; Ejsmond and Radwan 2011). This might in turn compromise the capability of the host's immune system to keep up with the evasive mechanisms of the current, or newly introduced pathogens. However, whether and how MHC variation found in *M. berthae* translates into population viability remains to be tested by the integration of genetic data with further health and survival assessment.

Overall, the empirical evidence supporting a link between MHC variation and fitness remains equivocal across taxa (reviewed in Acevedo-Whitehouse and Cunningham 2006; Radwan et al. 2010; Winternitz et al. 2013). Some populations that have undergone a demographic bottleneck seem to cope with critically low MHC variation (e.g. Ellegren et al. 1993; Mikko et al. 1995; Babik et al. 2005; Gangoso et al. 2012), or low MHC allelic richness compensated by high allelic divergence (e.g. Radwan et al. 2007; Castro-Prieto et al. 2011), while others retained high levels of MHC variation despite facing a

bottleneck that simultaneously lowered neutral genetic diversity (Aguilar et al. 2004; Hedrick and Hurt 2012; Oliver and Piertney 2012). Although the consequences of decreasing MHC variation might be undetectable over long periods of time, it might eventually compromise the ability of small or isolated populations to resist to ever-changing pathogen pressures in the future, over time scales that may be difficult to measure in most empirical studies (reviewed in Radwan et al. 2010; Spurgin and Richardson 2010). Therefore, the continuous long-term demographic monitoring of populations for which estimates of MHC variation have been established at one or several points in time may, in the future, help us to refine our understanding of the time-scale over which such processes are acting, especially in relatively short-lived species.

**Compliance with ethical standards**

**Research involving animals:** We have adhered to the Guidelines for the Treatment of Animals in Behavioral Research and Teaching (Animal Behavior 2006, 71: 245-253) and the legal requirements of the country (Madagascar) in which the fieldwork was carried out.

**Funding:** This study was funded by a SAW project of Leibniz Graduate School for the Foundations of Primate Social Behavior issued by Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e.V. (SAW-2011-DPZ-4) and field work was supported by the DPZ and DFG (Ka 1082/10-1&2).

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

**Acknowledgements**

We would like to thank all people who have contributed to the collection and processing of DNA samples, with a special thanks to the Kirindy field assistants, notably Bruno Tsiverimana, Tiana Andrianjanahary, Jean-Claude de Beroboka and Mamy Razafindrasamba, as well as to Christina Glaschke. We also thank the Département de Biologie Animale, Université d'Antananarivo, the CAFF of the Direction des Eaux et Forêts and the CNFEREF Morondava for granting a research permit and to Prof. Lutz Walter and Dr. Christian Roos for use of the GS Junior sequencer and Nico Westphal for providing technical help. This research was funded by a SAW project of Leibniz Graduate School for the Foundations of Primate Social Behavior issued by Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e.V. (SAW-2011-DPZ-4) and field work was supported by the DPZ and DFG (Ka 1082/10-1&2).

# CHAPTER 3

## EVOLUTIONARY ORIGINS OF SHARED MHC POLYMORPHISM IN SYMPATRIC LEMURS

Eva Kaesler (born Pechouskova), Peter M. Kappeler, Markus  
Brameier, Janina Demeler, Cornelia Kraus, Josué H. Rakotoniaina,  
Anni. M. Hämäläinen and Elise Huchard

Submitted to *Molecular Ecology*



## Abstract

Genes of the Major Histocompatibility Complex (MHC) assume a central role in the adaptive immune response of vertebrates. They exhibit a remarkable polymorphism, which often crosses species boundaries, with similar alleles or allelic motifs shared across species, a phenomenon known as ‘trans-species polymorphism’. Trans-species polymorphism may reflect the operation of parallel selective pressures, either via convergence and the emergence of functionally similar alleles, or via co-ancestry and the long-term maintenance of MHC sequence motifs over multiple speciation events. Here we investigate the origins of MHC similarity across several confamilial lemur species (*Cheirogaleidae*). We examined MHC class II variation in two highly polymorphic loci (DRB and DQB) and evaluated the overlap of gut-parasite communities in four sympatric lemurs. We tested for parasite-MHC associations across species to determine whether similar parasite pressure may select similar MHC alleles in different host species. Next, we integrated our MHC data with those previously obtained from other *Cheirogaleidae* to investigate the relative contributions of convergence versus co-ancestry to trans-species polymorphism by contrasting patterns of codon usage at functional and neutral sites. Our results indicate that parasites shared across host species may select for functionally similar MHC alleles and some alleles may affect risk of multiple infections in several host species implying that the dynamics of MHC-parasite co-evolution should be envisaged at the community level. We further show that co-ancestry is the primary mechanism responsible for the retention of MHC sequence motifs between species that diverged up to 30 million years ago, suggesting that MHC polymorphism is capitalized over multiple speciation events and may be slow to regenerate when eroded.



---

## Introduction

The Major Histocompatibility Complex (MHC) is a multigene cluster characterized by remarkable polymorphism and a complex evolutionary history. The MHC genes encode MHC molecules, cell surface glycoproteins, that trigger the initial phase of an immune response against intracellular (MHC class I) or extracellular pathogens (MHC class II), by binding antigenic peptides at peptide-binding regions (PBR). Antigens are presented to T-lymphocytes that initiate an antigen-specific immune response (Trowsdale 1993). Amino acid polymorphism lining the peptide-binding groove (Hughes and Nei 1988) or its close proximity (ca 4 Å, Nielsen et al. 2007), may alter the 3D-positioning of amino acid contact residues (Bjorkman and Burmeister 1994) that determines the binding specificity of each encoded MHC molecule and consequently spectrum of antigens recognized (Rammensee et al. 1995). The individual immune repertoire, that is, which pathogens can or cannot be recognized by its adaptive immune response, is then determined by MHC genotype.

MHC polymorphism is thought to be largely maintained by balancing selection mediated by co-evolving parasites through selective mechanisms such as divergent allele advantage (Wakeland et al. 1990; Richman and Kohn 1999; Lenz 2011) or overdominance and negative-frequency dependent selection (Klein et al. 1993a; Spurgin and Richardson 2010). Support for a pathogen-mediated selection on MHC stems from associations between certain MHC alleles and the prevalence of some pathogens at the population or community level (e.g. Hill et al. 1991; Schwensow et al. 2007, 2010a,b; Wegner et al. 2003; Froeschke and Sommer 2012; Sepil et al. 2013; Sommer et al. 2014; Pilosof et al. 2014), or from positive correlations between MHC polymorphism and parasite richness across species (de Bellocq et al. 2008; Garamszegi and Nunn 2011). Host-parasite co-evolution can be seen as a dynamic interplay of direct and indirect effects between multiple hosts and parasites shaping host MHC variation (Pilosof et al. 2014). For instance, according to the ‘rare-allele advantage hypothesis’ (e.g. Snell 1968, Bodmer 1972; Spurgin and Richardson 2010), rare alleles may confer selective advantage over more frequent alleles given that co-evolving pathogens might not have enough time to evolve evasive mechanisms. Alterations in parasite resistance in one host species, e.g. via changes in allelic frequencies, can influence parasite prevalence in a host population,

which can have cascading effect on individuals of other host species by altering selection imposed by the parasite on their MHC diversity. Thus, it is important to consider the whole community rather than single species to comprehend the effects of parasite-mediated selection on the evolution of MHC polymorphism (Pilosof et al. 2014).

The extensive MHC polymorphism often crosses species boundaries, with similar alleles or sequence motifs shared among species confounding MHC-based phylogenetic reconstructions, a phenomenon referred to as ‘trans-species polymorphism’ (‘TSP’, Klein 1987, 2007). Two fundamental evolutionary scenarios have been proposed to explain it. First, the MHC allelic similarity across taxa may originate from common ancestry (CA) provided distinct orthologous allelic lineages are passed from ancestral to descendant species across speciation events (Klein 1987; Figueroa et al. 1988; McConell et al. 1988; reviewed in Klein et al. 2007; Azevedo et al. 2015; Tesicky and Vinkler 2015). Alternatively, similar parasite pressures may select for similar allelic motifs in different species that are limited to the PBR through the operation of convergent evolution (CE), independently of host taxonomy (Gustafsson and Andersson 1994; Yeager and Hughes 1999; Christin et al. 2010). If co-ancestry is responsible for allelic similarity at coding regions, adjacent regions that are not directly involved in antigen recognition (e.g. non-PBR, or flanking introns) would also show higher similarity than expected by chance. In contrast, convergent evolution would promote amino acid (but not necessarily codon) similarity limited to coding regions, but should not affect the structural (nucleotide) composition of non-coding regions that are expected to segregate according to the species phylogeny (Klein et al. 2007; Lenz et al. 2013).

So far, qualitative or quantitative indications supportive of either scenario were reported (e.g. CE: Andersson et al. 1991; Trtkova et al. 1995; Yeager and Hughes 1999; Kriener et al. 2000, 2001; Srithayakumar et al. 2012; Pilosof et al. 2014, CA: Lundberg and McDevitt 1992; Klein et al. 1993b; Graser et al. 1996; Garrigan and Hedrick 2003; Lenz et al. 2013; Eimes et al. 2015; Tobler et al. 2014), but the majority of instances where inconsistent MHC genealogy was observed were attributed to common ancestry. Simultaneous examination of MHC constitution, including the evolution of non-coding regions (‘introns’, Zhang and Kumar 1997; Kriener et al. 2000), and MHC-parasite interactions would be an ideal setting to address the relative contribution of the two

---

evolutionary mechanisms to MHC allelic similarity. However, this approach remains challenging in natural populations and non-model organisms (Lenz et al. 2013).

Additional insights regarding the interactions between pathogens and host immune response can be achieved by reaching beyond nucleotide or amino acid similarity to examine (1) the functional similarity of encoded MHC proteins, as well as (2) patterns of codon usage, as an indication of co-ancestry. Each encoded MHC molecule can bind a limited array of antigens, and the antigen specificity is determined by physio-chemical properties of amino acids positioned at PBR. Consequently, MHC alleles encoding MHC molecules with similar antigen-binding affinities can be classified into functionally equivalent groups, referred to as ‘MHC supertypes’ (Sette and Sidney 1999; Southwood et al. 1998; Trachtenberg et al. 2003; Lund et al. 2004). Supertype classification has proven valuable to investigate associations between the binding specificity of MHC alleles possessed by a given host, and host susceptibility to pathogens within and across species (e.g. Trachtenberg et al. 2003; Schwensow et al. 2007, 2010b; Sepil et al. 2013; Pilosof et al. 2014). Functional similarity across species (i.e. alleles from different species belonging to a same supertype) would suggest the operation of convergent evolution if allelic structural similarity is absent from non-PBR regions. Alternatively, if shared polymorphism at PBR is a consequence of common descent, we expect the similarity of amino acid sequences to reflect the similarity of nucleotide sequences, and so, similar patterns of codon usage (Lundberg and McDevitt 1992; Lenz et al. 2013).

Here we aim to address the relative contribution of convergent evolution and common ancestry to MHC allelic similarity by simultaneous examination of MHC constitution and parallel selective pressures in a community of four sympatric lemur species – *Microcebus berthae*, *M. murinus*, *Mirza coquereli* and *Cheirogaleus medius* (Primates, *Cheirogaleidae*). This lemur community presents three key advantages for the present study. First, they vary in phylogenetic relatedness, allowing us to test the relative contribution of the two evolutionary scenarios across pairs of species that have diverged more or less recently. We may expect their contribution to vary with the degree of species relatedness, which may provide insights about the potential longevity of MHC alleles. Second, these species are reproductively isolated and the absence of ongoing hybridization minimizes the risk of confounding factors, such as allelic introgression (see also Lenz et al. 2013; Klein et al. 2007). Third, all these species are small-bodied

(<500g), nocturnal omnivores facing broadly similar environmental conditions. This community includes rare solitary and spatially more dispersed *M. berthae* and *M. coquereli*, relatively abundant and pair-living *C. medius* and abundant and socially cohesive *M. murinus* (Dammhahn and Kappeler 2005, 2008b; Kappeler et al. 2002; Markolf et al. 2008; Fietz et al. 1999a,b; Mittermeier et al. 2008; Eberle and Kappeler 2002, 2006). They exhibit different levels of ecological overlap in their diet and in their demography in dyadic fashion that is incongruent with their phylogeny (Dammhahn and Kappeler 2014; Thiele et al. 2013). This may provide an opportunity to assess the importance of ecological overlap explaining patterns of MHC variation in these species, given that diet, host population density and spatial proximity have been indicated to affect the level of parasitism (e.g. Vitone et al. 2004; Hughes and Page 2007; Chen et al. 2008; Godfrey 2013; Morand 2015). A considerable overlap in parasite communities, dietary ranges and functional proximity of MHC alleles reported between *M. murinus* and *C. medius* (Schwensow et al 2010b; Dammhahn and Kappeler 2014) suggest a certain degree of ecological convergence present between at least these two species.

Here we proceed in three main steps. First, we assess the extent of overlap in gastro-intestinal gut parasites across four sympatric lemurs and compare patterns of their MHC class II allelic variation in the two highly polymorphic loci (DQB and DRB) derived by this and a previous study (*M. berthae*, Pechouskova et al. 2015).

Second, to evaluate the impact of pathogen-mediated selection on shaping MHC allelic variation across species, we evaluate the functional overlap between MHC alleles using supertype classification. Next, we integrate these data with helminth prevalence across species, to test whether species with similar parasite communities present similar patterns of MHC supertype frequencies. We further examine MHC-parasite interactions at a community and species level to test whether shared parasite pressures may select for functionally equivalent MHC alleles across host species, according to two distinct scenarios of parasite-driven selection. First, if each MHC supertype protects against one or few parasite species within and across species, we expect to detect a correlation between the frequency of particular MHC superotypes and the prevalence of particular parasites within and across species. The effect of particular helminths on host fitness remains unknown in *Cheirogaleidae* and the most common genera presumably show rather mild pathogenicity (Raharivololona and Ganzhorn 2010; Irwin and Raharison

2009). In contrast, multiple infestations can impose additional energetic demands causing detrimental effect on host fitness and survival and thereby reduced reproductive output (reviewed e.g. in Morand 2015). Thus alternatively, if some MHC supertypes have a broader and more generalistic effect than others and may protect against a diverse array of pathogens within and across species, we expect to detect a correlation between the frequency of particular MHC supertypes and parasite richness within and across species.

Third, to evaluate whether MHC functional similarity stems from common descent or from independent convergence, we evaluate the potential co-ancestry of (1) orthologous alleles that are functionally similar and belong to a same supertype and (2) orthologous peptide motifs at the PBR in species with various degrees of phylogenetic proximity by integrating our MHC data with those previously acquired for other *Cheirogaleidae*. Under the co-ancestry hypothesis, we predict that (1) MHC alleles belonging to the same supertype show higher structural similarity at non-PBR than alleles belonging to different supertypes, and (2) amino acids at PBR show higher codon similarity across species than expected by chance. In contrast, under the operation of convergent evolution, selection for similar amino acids, but not codons, would be expected across species. Moreover, allelic similarity should be predominantly limited to PBR (Lenz et al. 2013), especially in more distantly related species. Lastly, we expect the relative contribution of co-ancestry and convergence to vary with the degree of phylogenetic relatedness of species, with co-ancestry being the primary mechanism responsible for allelic similarity among the most closely related species.

## **Materials and Methods**

### **Study site, DNA collection and MHC genotyping**

The four lemur species live in sympatry within 12,500-ha Kirindy Forest/ CNFEREF (Centre National de Formation, d'Etude et de Recherche en Environnement et Foresterie) in western Madagascar (Kappeler and Fichtel 2012). Members of each species have been regularly captured between the years 1993 and 2013 at four study sites, using Sherman

and Tomahawk live traps. Further details on sampling design and sample size distribution of each species are presented in Supporting Information 1; see Appendix 2, pp. 113-145. The capture and DNA sampling and extraction protocols were identical to those previously described elsewhere (e.g. Pechouskova et al. 2015; Supporting Information 1). PCR amplification targeting the two highly polymorphic loci of the MHC class II region, DRB and DQB followed by MHC genotyping (454 pyrosequencing; Roche, France) were conducted as described in Pechouskova et al. (2015); see Supporting Information 1 for a brief summary.

All animal handling and sample treatments were in compliance with animal care regulations and applicable national laws of Germany and Madagascar (CITES: 429C-EA10/MG07, 430C-EA10/MG07), approved by the appropriate Animal Use and Care committees of Germany (Bundesministerium für Naturschutz, BfN) and the legal requirements of Madagascar (Ministère de l'Environnement et des Eaux et Forêts, MINEEF).

### **Parasite communities screening at species level**

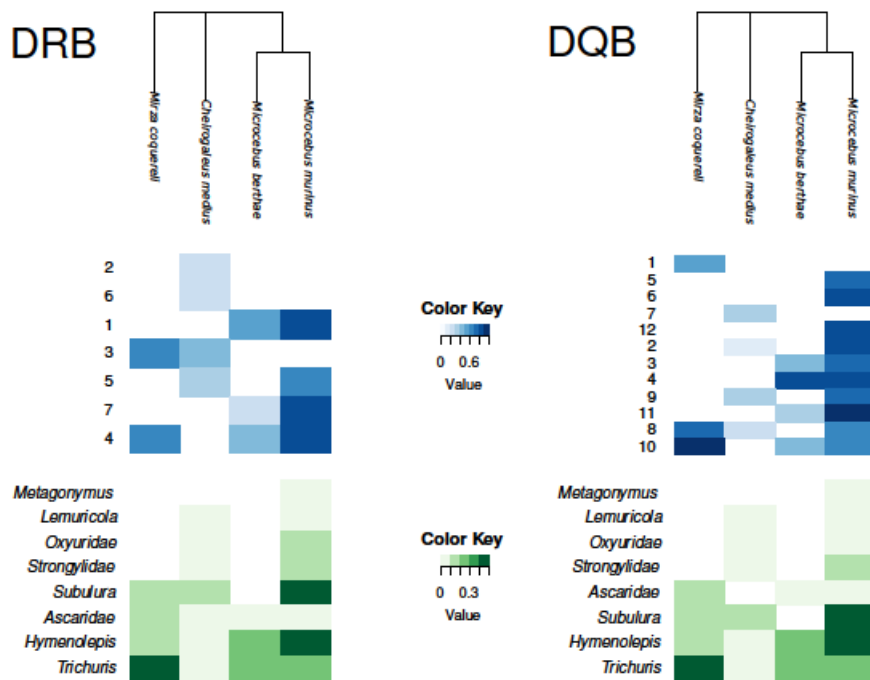
Coprosopic samples used in this study were collected during animal handling and from clean traps between September and December 2012. Samples were immediately homogenized in 10% formalin and stored for further examination. Helminth eggs were extracted using two different techniques: (a) standard FLOTAC protocol (Cringoli et al. 2010) and (b) Ritchie's formol - ether concentration method (Ritchie 1948) and their presence was determined microscopically following criteria for egg-morphotype classification up to the genus level (Raharivololona et al. 2006, 2009; Irwin and Raharison 2009). Only morphotypes that could be identified reliably (8 genera, Supporting Information 1) were retained for the subsequent analyses. The amount of samples collected for each host species is presented in Supporting Information 1.

At first, a simple quality control procedure was performed to evaluate the repeatability of helminth egg detection. The implementation and the outcome of this analysis are presented in Supporting Information 1.

Next, to evaluate the extent of overlap among helminth communities in the four host species, we tested whether two random samples collected from the same species are

more similar than two random samples from different species. A higher proportion of individuals of *M. murinus* (42 %) and *C. medius* (41 %) were sampled only once, compared to *M. coquereli* (13 %) and *M. berthae* (7 %). Thus, we homogenized sampling effort across species by including only those individuals that were sampled at least twice throughout two different sampling months. To control for seasonal variability, we performed the same test including only a subset of individuals that were sampled on the same trapping session only (within 1-3 days apart). The number of individuals included in each test is provided in Supporting Information 1. For all tests, Jaccard dissimilarity index was calculated for each pair of samples, using the function ‘vegdist’ implemented in the package ‘vegan’ (Oksanen et al. 2015), and averaged across groups (e.g. ‘same’ versus ‘different’ species). The resulting average was compared using Mann-Whitney *U* tests.

Finally, to assess the extent of parasite sharing among the four host species, we created a host-parasite matrix based on helminth prevalence (the proportion of individuals within each species infected by a given parasite) within each host species, following Pilosof et al. (2014). The host-parasite matrix (see Fig. 1) was transformed into a dissimilarity matrix using the *Růžička index* (RI), a quantitative version of the Jaccard index that allows direct comparisons between hosts (Speed et al. 2010, Tamás et al. 2001), implemented in the R package ‘vegan’. A RI value of 1 indicates maximum dissimilarity, where no parasites are shared between a pair of hosts and a RI value of 0 indicates that parasites infect hosts with a similar prevalence. All analyses were conducted in R v.3.2.1. (R Core Team 2014).



**Fig. 1** Host-parasite matrix (on top, in *green*) based on helminth prevalence defined as the fraction of individuals within host species (columns) infected by a given parasite (rows), sorted from generalist to specialist genera ; and host-MHC supertype matrix (bottom, in *blue*) based on supertype frequencies defined as the number of infected individuals of a given host species (columns) possessing a given supertype (rows), divided by the total number of individuals possessing this supertype within the same host species. Phylogenetic relationships between the four host species are visualised using an ultrametric phylogenetic tree based on 1140bp of cytochrome b mitochondrial gene sequences

### MHC variation and patterns of molecular selection across species

Allelic richness for each locus was compared across the four species, while controlling for sampling effort (number of individuals trapped in a given species), using a simple permutation test set to evaluate the number of alleles present in the population based on the number of alleles retrieved for a given sampling effort (Huchard et al. 2012). Physical association between the two loci was examined using tests of linkage disequilibrium in each host species implemented in Arlequin v.3.5.1.3. (using 10,000 permutations; Excoffier and Lischer 2010).



---

The presence of positive selection operating on nucleotide sequences was tested following the method developed by Yang et al. (2000, 2005), using the program CodeML implemented in the PAML v.4.7 software (Yang 2007). The implementation of the method is summarized in Supporting Information 1.

### **Characterization of MHC supertypes**

Classification of MHC alleles into ‘supertypes’ consists in characterizing MHC alleles by the binding specificity of amino acid residues lining the peptide-binding groove (here represented by positively selected sites ‘PSS’ identified in the previous step of the analysis) to assess their functional proximity within and across species.

At first, we generated a sequence alignment for each locus that was restricted to PSS. Given that the PSS distribution was not homologous across species for either locus, we created a conservative consensus alignment that would enable designing of supertypes across species. We included all PSS detected in any species (20 DRB and 12 DQB) to ensure that any functionally important site was retained in the analysis. Each amino acid from the DRB (or DQB) PSS alignment was characterized by a vector of five physiochemical descriptor variables that are thought to play a key role in determining antigen-binding specificities (Sandberg et al. 1998): z1 (hydrophobicity), z2 (steric bulk), z3 (polarity), z4 and z5 (electronic effects); as described by Doytchinova and Flower (2005).

The resulting matrix was then subjected to a K-means clustering algorithm with increasing number of groups using the function ‘find.clusters’ implemented in ‘adeget’ R package (Jombart et al. 2010) following Sepil et al. (2012). Here, different clustering solutions are compared using the Bayesian Information Criterion (BIC) to identify the optimal number of clusters. A minimal number of clusters is manually selected when BIC statistics decreases by a negligible amount, indicated by a minimum or an elbow in the curve of BIC values as a function of cluster number.

The supertype classification was validated in two steps as proposed by Pilosof et al. (2014). First, we repeated the procedure by running 500 automated classifications using the automated clustering criterion ‘smoothNgoesup’ implemented in the package ‘adeget’, that reduces sensitivity to small fluctuations in the decrease of the BIC statistics and therefore the risk of identifying a local minimum. Second, we compared

results of manual classification with the automated one using a measure of normalized mutual information (MI) based on information theory (Danon et al. 2005). Here an index of certainty is calculated with its maximum value (1) reached when clusters created by both approaches contain only identical alleles, and its minimum value (0) when only different alleles are present.

### **Selection of similar supertypes across species: parasite-host-supertype associations**

To examine whether parasite-driven selection is likely to contribute to the distribution of MHC supertypes across species, we examined associations between helminth infestations and supertype frequency using two different approaches. Here a total of 187 individuals (*M. berthae*  $n=14$ ; *M. murinus*  $n=128$ ; *M. coquereli*  $n=8$ ; *C. medius*  $n=37$ ) for which both MHC and data on helminth prevalence were available were included in following analyses.

#### ***Supertype-parasite associations at the community level***

First, we investigated the extent of association between the frequency of MHC supertypes and helminth prevalence across all host species, to test whether species with similar helminth communities may also show similar MHC supertype distributions. Following the procedure described by Pilosof et al. (2014), we created a matrix describing the frequency of MHC supertypes in different host species and then transformed it into a distance matrix using the *Růžička index* (RI) in the same manner as described above for the host-parasite matrix (see Fig. 1). To control for a potential phylogenetic signal among the four host species, we constructed a phylogenetic tree based on previously published cytochrome b mitochondrial gene sequences (1140bp; NCBI Genebank, *M. coquereli* EU835932; *C. medius* EU825326 in Groeneveld 2008, 2009; *M. berthae* GU327166, *M. murinus* GU327178 in Weisrock et al. 2010). The tree was constructed using a ML algorithm (GTR substitution model, 1000 bootstrap replications; Tavaré 1986) implemented in PhyML 3.0 (Guindon et al. 2010); and is displayed in Fig 1. We then inferred phylogenetic distances using the function ‘cophenetic.phylo’ implemented in the R package ‘ape’ (Paradis et al. 2004). Finally, we performed a partial Mantel test to

---

evaluate the significance of the correlation between the host-supertype and host-parasite matrices, while controlling for host phylogeny. To examine whether uneven sample sizes available for each species could influence the supertype-parasite association patterns we repeated the partial Mantel test controlling for host abundance (see also Pilosof et al. 2014; Supplementary methods).

***Supertype-parasite associations at the species level***

If each MHC supertype inhibits or favors infection by a limited array of parasites, we expect to detect a correlation between the frequency of some MHC superotypes and the prevalence of some parasites within and across species. In a first approach, we therefore examined the strength of supertype-helminth associations across samples and individuals from different species, using multivariate generalized linear mixed models (GLMMs) with helminth presence/absence in a given sample as a response variable (binomial error distribution, logit-link function). Alternatively, if some MHC superotypes may have a broader, more generalistic effect than others and inhibit or favor infection by a diverse array of parasites, we expect to detect a correlation between the frequency of some MHC superotypes and parasite richness within and across species. In a second approach, we therefore examined whether some superotypes may influence parasite richness measured by the number of helminth genera present in a given sample, using multivariate GLMMs with parasite richness in a given sample as a response variable (Poisson-error distribution, log-link function).

For building our models, we initially considered the eight most common helminth genera (Supporting Information 1). Four of those (*Hymenolepis* sp., *Subulura* sp., *Trichuris* sp. and *Ascaris* sp.) accounted for 87% of the total helminth infections and were found in 3-4 lemur species. These were considered for models examining the effect of superotypes on one helminth genus. Host species for which infestation by a particular parasite was present in less than 2 individuals or 2 observations were not considered in given models, in order to ensure reliable calculation of model estimates. For models on parasite richness, all helminth genera and all host species were included (2<sup>nd</sup> approach).

For both approaches, MHC superotypes that occurred in one host species only (Supporting Information 1) were excluded from further analysis, as we aimed at testing whether the possession of a given supertype may be linked with the presence of a given

parasite in more than one host species. We accounted for monthly variation in parasite prevalence and repeated observations of a same individual, by including month of data collection and host individual identity as random effects. Host species and MHC supertype were fitted as fixed effects. Given that pairwise correlations among DRB (labelled ‘SR’) and among DQB (‘SQ’) superotypes showed relatively low level of collinearity (DRB  $r < 0.23$ ; DQB  $r < 0.26$ ), all relevant SR and SQ superotypes were considered in a single model. For multi-host models, only superotypes present in at least two host species were included in a given model. The significance of fixed effects was evaluated by comparing a model with and a model without the supertype of interest using a likelihood ratio test and Bonferroni’s correction for multiple testing. Significant effects detected across species (‘multi-host models’) were tested in each species separately (‘single-host models’) to ensure that such effects were not driven by the most abundant host (*M. murinus*) only. Statistical significance of variables was tested using the full model to avoid problems associated with stepwise model selection (Whittingham et al. 2006; Mundry and Nunn 2009). All models were conducted using function ‘glmer’ implemented in the R package ‘lme4’ (Bates et al. 2015).

### **The role of common ancestry vs. convergent evolution in maintaining MHC functional similarity**

To investigate whether MHC functional similarity stems largely from the operation of convergent evolution (CE) or from common ancestry (CA), we investigated the origin of sequence similarity at sites that are either neutral or under purifying selection (non-PSS) and at sites that are under positive selection (PSS).

We first tested whether the functional similarity of alleles belonging to the same supertype is likely to arise from common descent, or independently by convergence. To do this, we compared nucleotide similarity at non-PSS between alleles belonging to a same versus different superotypes, by computing average pairwise nucleotide distances of non-PSS (number of differences) within and between superotypes for each locus separately in MEGA 6 (Tamura et al. 2013). Under the CA scenario, we expect alleles belonging to the same supertype to be more similar at both PSS and non-PSS compared to alleles belonging to different superotypes. Under the CE scenario, we expect alleles belonging to

---

the same supertype to be more similar at PSS, but not at non-PSS, compared to alleles belonging to different superotypes.

Second, we examined whether the phylogeny of MHC alleles based on non-PSS sequence partitions reflects species phylogenies, in contrast to the phylogeny of PSS sequence partition. This would be expected if inter-specific overlap at PSS arose independently from convergence (Yeager and Hughes 1999; Lenz et al. 2013). For the tree reconstruction we used DRB and DQB sequences available for the four *Cheirogaleidae* from Kirindy and DRB sequences described at distant study sites within *Microcebus* sp. - *M. murinus*, *M. griseorufus* and *M. rufus* (Sommer et al. 2014; see Supporting Information 2; in Appendix 2; pp. 146-158). For these additional sequences, the positioning of PSS was extrapolated. The non-PSS sequence alignment was derived from the overall sequence alignment (169bp) by randomly selecting equal amount of non-PSS as of PSS in each locus to obtain sequence alignments of equal length. All trees were reconstructed in PhyML using the most likely model of sequence evolution estimated by jModelTest for each tree separately (500 bootstrap replicates; Guindon et al. 2010; Posada 2008).

Third, we examined codon usage patterns at PSS, following a procedure described by Lenz et al. (2013). First, a Monte Carlo simulation (further referred to as MC sampling) was performed to generate a distribution of the proportion of identical codons expected to occur between species if CE is the prevailing mechanism generating allelic similarity at PSS. To do so, we calculated the number of identical amino acids at PSS in a pairwise fashion between all pairs of species, and recorded for each pair of identical amino acid, whether the corresponding codons were identical or not. Then, we created a drawing pool used for the MC sampling that consists of codon frequencies derived for each species separately from the actual codon frequencies present across the whole exon alignment. For MC sampling, the amino acid specific number of potential codons was randomly drawn 1000 times and cases of identical codons between species were recorded. As a result, we obtained a simulated probability distribution of the proportion of identical codons expected under the CE scenario ( $CE_d$ ).

Next, the observed codon similarity between species was compared with the codon similarity expected within species, approximating a scenario of recent common descent. Here, we performed pairwise comparisons of all alleles within each species and

recorded how often the same amino acid was coded by the same or by a different codon. This amino acid-specific identity count served as a second drawing pool, from which the number of identical amino acids observed between species at PSS (the same number as above) was randomly drawn 1000 times to obtain a probability distribution of the expected proportion of identical codons within species ( $CA_d$ ). Both procedures were repeated for each species pair within each locus separately. For inter-specific comparison, we used the same set of sequences as for the phylogenetic reconstruction (see above). All above simulations were performed using customized Perl scripts available upon request from Lenz et al. (2013).

Finally, the proportion of identical codons observed at PSS between species (and populations) was compared against the simulated distributions ( $CA_d$ , and  $CE_d$ ). To compute the relative contribution of CE versus CA in each species, we located each species on an axis with two ends, where each end corresponds to the value of the median of the simulated distribution of each evolutionary scenario ( $CE_d$ ,  $CA_d$ ). This provided a score indicative of the relative contribution of co-ancestry (further referred to as ‘score indicative of co-ancestry’) for a species pair as:  $(\text{number of identical codons between species} - \text{median } CE_d) / (\text{median } CA_d - \text{median } CE_d)$ .

Next, we tested whether the scores indicative of co-ancestry obtained for all pairwise comparisons would correlate with phylogenetic distances among the species. We would expect the relative contribution of co-ancestry to correlate with phylogeny, if co-ancestry is majorly responsible for allelic similarity at PSS; and this effect to be greater in more closely related species. Phylogenetic distances among the species were inferred from a phylogenetic tree based on cytochrome b mitochondrial gene sequences for the Kirindy community (see above) and for allopatric populations of *M. rufus* and *M. griseorufus* (NCBI Genebank, *M. rufus* GU327245; *M. griseorufus* GU327345 in Weisrock et al. 2010).

Finally, we tested whether the proportion of identical amino acids and the scores indicative of co-ancestry obtained for all pairwise comparisons would correlate with the distance in helminth prevalences (obtained in the previous steps of the analysis), as CE requires shared parasite pressures. We would expect the relative contribution of CE to be greater in species with more similar helminth prevalences independently of the phylogenetic signal, if CE is the main contributing factor to the functional overlap among

alleles. Given that the proportion of shared codons between species is not independent from phylogenetic distance, we performed a partial Mantel test controlling for it.

## Results

### Parasite communities screening at species level

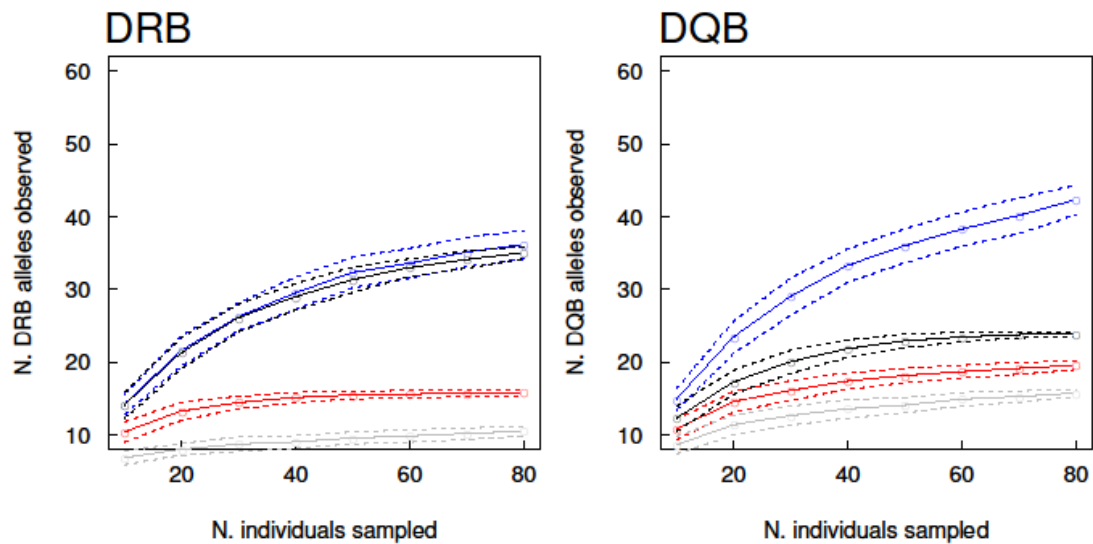
The overview of helminth infestations across four *Cheirogaleidae* are presented in Supporting Information 1. Four out of eight helminth genera (*Ascaris* sp., *Trichuris* sp., *Hymenolepis* sp. and *Subulura* sp.) were present in all four or at least three different host species. The proportions of infected individuals ranged from 32 % in *C. medius* to 82 % in *M. murinus*, which also showed the highest parasite richness (Supporting Information 1). Comparison of parasite communities across species revealed that samples collected from the same host species were more similar ( $d_J$ , mean $\pm$ SD 0.68 $\pm$ 0.29) than samples from different species ( $d_J$ , mean $\pm$ SD 0.91 $\pm$ 0.19) when including samples collected from a given individual over at least two different months (Mann-Whitney *U* test,  $W=3914187$ ,  $P<0.001$ ), as well as during the same trapping session (same host species:  $d_J$ , mean $\pm$ SD 0.82 $\pm$ 0.28 vs. different host species:  $d_J$ , mean $\pm$ SD 0.94 $\pm$ 0.19) (Mann-Whitney *U* test,  $W=1582817$ ,  $P<0.001$ ). The host similarity matrix suggested that *M. coquereli* and *M. berthae* are most similar concerning helminth prevalences (Supporting Information 1). In addition, seven out of eight genera were shared among *C. medius* and *M. murinus*.

### MHC variation and patterns of molecular selection across species

The overview of successfully genotyped individuals per each species is presented in Supporting Information 1. We present a total of 22 DRB and 41 DQB exon MHC class II sequences described here for the first time. Details on all sequences detected or newly described by this study are summarized in Supporting Information 2. In either species, we did not find any support for loci duplication (see Supporting Information 1) and both loci were in tight linkage disequilibrium in all three species (*C. medius*  $\chi^2=900.56$ ,  $P<0.01$ , *M. murinus*  $\chi^2=2289.84$ ,  $P<0.001$ ; *M. coquereli*  $\chi^2=539.01$ ,  $P<0.001$ ).

The estimated allelic richness (number of distinct alleles) was highest for *M. murinus* in DQB and comparably high for *M. murinus* and *C. medius* in DRB when controlling for sampling effort. *M. berthae* followed by *M. coquereli* displayed the lowest allelic richness in both loci (Fig. 2).

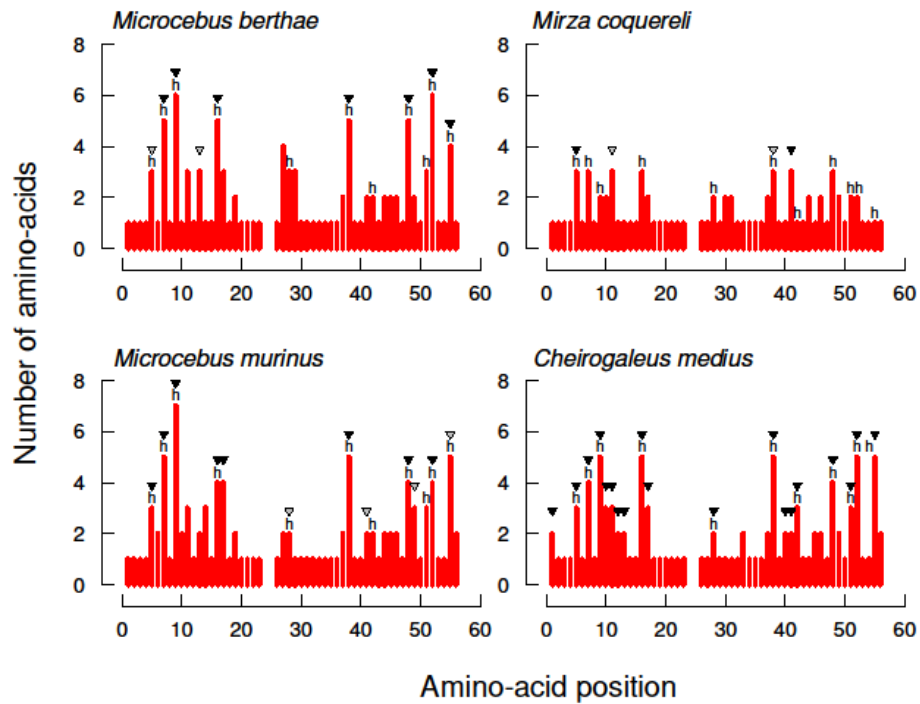
The presence of positively selected sites ‘PSS’ was supported in all species and loci, except for the DRB locus in *M. coquereli* (Supporting Information 1). The distribution of PSS showed different patterns across species (Fig. 3) especially at DQB between *M. berthae* ( $n_{PSS}=11$ ) and all three other *Cheirogaleidae*; *C. medius* ( $n_{PSS}=1$ ), *M. murinus* ( $n_{PSS}=2$ ) and *M. coquereli* ( $n_{PSS}=2$ ).



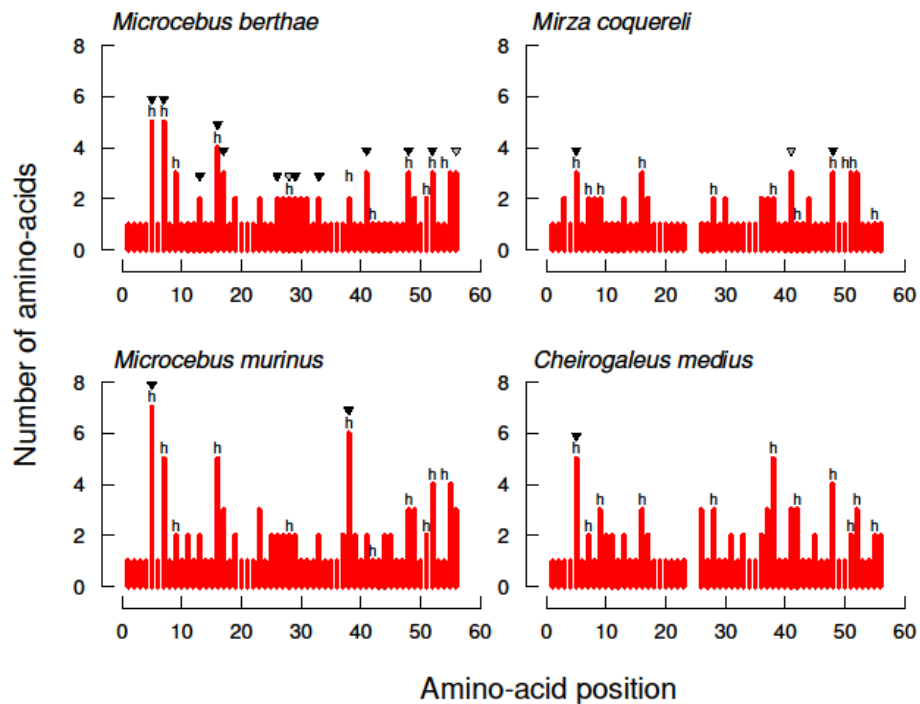
**Fig. 2** Allelic richness for a given sampling effort estimated through a resampling procedure, showing the number of distinct alleles detected at DRB and DQB when randomly drawing an increasing number of individuals for *M. murinus* (blue), *C. medius* (black), *M. coquereli* (grey) and *M. berthae* (red). The dotted lines indicate the standard deviation around the estimated mean (solid line)



## DRB



## DQB



**Fig. 3** Amino acid variation plots for MHC-DRB and DQB in four sympatric lemurs. Human Antigen-Binding Sites (ABS) are indicated by the letter 'h' and positively selected sites are indicated by black ( $P > 99\%$ ) and grey triangles ( $P > 95\%$ )

## Characterisation of supertypes

The optimal clustering solutions that displayed lowest BIC suggested 7 clusters in DRB and 12 in DQB (Supporting Information 1), which was in concordance with the number of clusters obtained by averaging 500 automated classifications ( $\text{mean}_{\text{CLUSTERS}} \pm \text{SD}$ , DRB  $7.86 \pm 0.57$ , 95 % CI (7,9); DQB  $12.57 \pm 1.78$ , 95 % CI (9,16)). The average normalized mutual information (MI) across the 500 automatic classifications was  $0.84 \pm 0.07$  in DRB and  $0.79 \pm 0.05$  in DQB, indicating high certainty and a great overlap in the classification of alleles into particular supertypes.

## Selection of similar supertypes across species: parasite-host-supertype associations

### *Supertype-parasite associations at the community level*

The community level analysis relating cross-species similarities in helminth prevalence and in the distribution of MHC supertypes lacked significance in both loci (Partial Mantel test controlling for phylogeny, DRB  $r=0.76$ ,  $P=0.13$ ; DQB,  $r=0.77$ ,  $P=0.17$ ; Supporting Information 1). Controlling for host abundance did not markedly alter the outcome (DRB  $r=0.53$ ;  $P=0.13$ ; DQB,  $r=0.67$ ,  $P=0.17$ ).

### *Supertype-parasite associations at the species level (GLMMs)*

First, we investigated whether the possession of a given supertype affects individual infestation by some common helminth genera that are shared across host species (*Hymenolepis* sp., *Subulura* sp., *Trichuris* sp. and *Ascaris* sp.). In DRB, a significant link between the possession of SR4 and the presence of *Hymenolepis* sp. was detected in the multi-host model including the two *Microcebus* sp. (Table 1, SR4  $\chi^2_{(1)}=12.46$ ,  $P>0.001$ ) and the direction of the effect was consistent in both hosts when tested in single-host models, with SR4 favoring infestations by *Hymenolepis*.

In the second step, we investigated, across species, the congruence of supertype effects on the risk of multiple infestations, i.e. parasite richness. A link between the possession of SQ4, shared by the two *Microcebus* sp. and parasite richness was present in

the multi-host model (SQ4  $\chi^2=8.09$ ,  $P<0.005$ ), and remained significant when tested in single-host models for both *Microcebus* sp (Table 1).

Other associations that were detected using multi-host models, but which proved undetectable when tested in single-host models were not considered further.

**Table 1.** Associations between MHC supertypes and (a) prevalence of the shared helminth *Hymenolepis* sp. and (b) infection by multiple parasites ('parasite richness'); all models were analysed using GLMMs with response variables: (a) 'helminth prevalence' in a given sample (yes/no) and (b) 'parasite richness' (number of parasite species present in a given sample). For each approach, we present multi-host models including all relevant species and supertypes that fulfilled our criteria for testing and single host models testing significant predictor derived from multi-host models. Host species and the possession of several MHC supertypes (yes/no) were fitted as fixed effects, and individual identity and month of collection as random effects. Only models were significant associations were found both in multi- and single-host models are shown.

Model	Fixed effect	Estimate	SE	z value	Pr(> z )	CI 95 %	
						2.5%	97.5%
(a) Effect of DRB supertypes on infection by <i>Hymenolepis</i> sp.							
<b>Multi-host model</b>							
<i>including M. berthae &amp; M. murinus</i> (ref category: <i>M. berthae</i> )							
	(Intercept)	-3.38	1.03	-3.28	0.001	-5.401	-1.360
	<i>M. murinus</i>	1.13	0.72	1.56	0.119	-0.288	2.540
	SR1	0.47	0.44	1.05	0.293	-0.404	1.339
	SR4	1.62	0.49	3.31	0.001	0.662	2.583
	SR7	0.22	0.40	0.54	0.591	-0.573	1.007
<b>Single host models</b>							
<i>M. berthae</i>							
	(Intercept)	-3.53	1.13	-3.13	0.002	-5.747	-1.322
	SR4	3.13	1.33	2.35	0.019	0.514	5.744
<i>M. murinus</i>							
	(Intercept)	-1.71	0.64	-2.69	0.007	-2.958	-0.465
	SR4	1.22	0.44	2.76	0.006	0.354	2.081
(b) Effect of DQB supertypes on parasite richness							
<b>Multi-host model</b>							
Including all four <i>Cheirogaleidae</i> species (ref category: <i>C. medius</i> )							
	(Intercept)	-1.22	0.28	-4.29	<0.001	-1.777	-0.663
	<i>M. berthae</i>	-0.74	0.44	-1.66	0.096	-1.612	0.132
	<i>M. coquereli</i>	1.91	0.52	3.65	0.000	0.885	2.943
	<i>M. murinus</i>	1.41	0.28	4.96	0.000	0.855	1.971
	SQ1	-2.24	0.62	-3.64	0.000	-3.445	-1.035
	SQ2	-0.23	0.16	-1.41	0.159	-0.541	0.088
	SQ3	-0.39	0.16	-2.47	0.014	-0.702	-0.081
	SQ4	0.49	0.17	2.91	0.004	0.161	0.821
	SQ5	-0.46	0.26	-1.74	0.082	-0.979	0.059
	SQ8	-0.11	0.23	-0.47	0.640	-0.550	0.338
	SQ9	-0.36	0.23	-1.61	0.107	-0.808	0.079
	SQ10	-0.22	0.19	-1.20	0.230	-0.591	0.142
	SQ11	0.24	0.16	1.49	0.137	-0.075	0.546
	SQ12	-0.15	0.15	-1.06	0.291	-0.438	0.131
<b>Single host models</b>							
<i>M. murinus</i>							
	(Intercept)	-0.08	0.00	-33.16	<0.001	0.083	-0.074
	SQ4	0.49	0.00	206.04	<0.001	0.481	0.490
<i>M. berthae</i>							
	(Intercept)	-3.45	1.07	-3.23	0.001	-5.545	-1.360
	SQ4	2.21	1.12	1.96	0.0495	0.005	4.409

---

## The role of a common ancestry vs. convergent evolution in maintaining MHC functional similarity

The relative contribution of common ancestry (CA) vs. convergent evolution (CE) in shaping functional similarity of MHC alleles in *Cheirogaleidae* was evaluated in three steps.

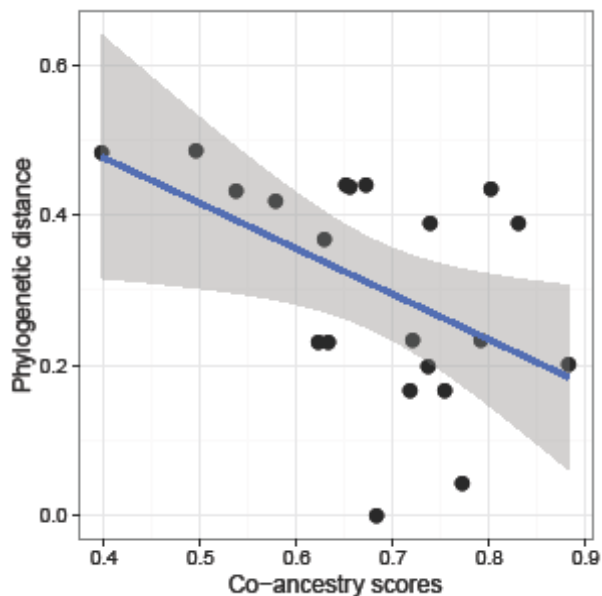
First, the nucleotide similarity of non-PSS was marginally or significantly higher among alleles belonging to the same vs. different supertypes in both loci (mean $\pm$ SD for same vs. different supertype, Student's t-test, DRB 2.21 $\pm$ 0.79 vs. 2.87 $\pm$ 0.97,  $P=0.08$ ; DQB 5.00 $\pm$ 1.35 vs. 8.05 $\pm$ 1.85,  $P<0.001$ ), suggesting that at least in DQB, the functional similarity of alleles belonging to the same supertype is more likely to be explained by co-ancestry.

Second, 128 DQB and 235 DRB nucleotide sequences were used to reconstruct phylogenies based on (1) 20 DRB and 12 DQB PSS sequence partitions and (2) a complementary alignment of equal length (20 DRB and 12 DQB codons), using a subset of randomly selected non-PSS. There was a lack of species-specific clustering within the tree based on PSS in both loci that persisted also within the tree based on non-PSS, but with less supported and less inclusive clusters. Although higher and better supported at PSS, allelic similarity is not limited to these sites, as expected if allelic similarity at PSS originated from similarity by descent.

Third, we analyzed codon usage patterns at PSS by calculating the proportion of shared codons at PSS in a pairwise fashion between all pairs of species and evaluated its relative distance from the expected codon similarity distribution (obtained by MC sampling) under the two evolutionary scenarios ( $CE_d$ ,  $CA_d$ ). The proportions of shared codons at PSS, scores indicative of co-ancestry, phylogenetic distances and median of the  $CE_d$ ,  $CA_d$  obtained for all pair-wise comparisons are summarized in Supporting Information 1. For all pairwise comparisons in both loci, the proportions of identical codons was significantly higher than expected under the scenario of  $CE_d$ , and notably closer but still significantly lower than expected codon similarity within species ( $CA_d$ ) (for all comparisons  $P<0.001$ , one-tailed t-test). In DRB, we found a significant correlation between phylogenetic distances and the scores indicative of co-ancestry ( $r=-0.48$ ,  $P<0.03$ ; Fig. 4). In DQB, the direction of the correlation was the same, but

weakened probably due to lack of statistical power compared to DRB (6 vs. 21 comparisons) ( $r=-0.58$ ,  $P=0.23$ ). This indicates that functional similarity of MHC alleles in closely related species stems largely from their co-ancestry, rather than convergence.

Finally, if allelic similarity at PSS emerged independently from convergence, we expected to find greater amino acid similarity at PSS, and lower scores indicative of co-ancestry, in species with more similar parasites. Here, no relationship could be detected in either locus (Partial Mantel test, proportion of shared amino acids at PSS vs. parasite distances, DRB  $r=-0.49$ ;  $P=0.79$ ; DQB  $r=0.26$ ;  $P=0.38$ ; scores indicative of co-ancestry vs. parasite distance, DRB  $r=-0.15$ ,  $P=0.63$ ; DQB  $r=-0.63$ ,  $P=0.92$ ), confirming that co-ancestry may play a greater role than shared parasite pressures in maintaining functional similarity among MHC alleles belonging to different species.



**Fig. 4** The relative contribution of convergent evolution and co-ancestry to MHC allelic functional similarity in six lemur species (*Cheirogaleidae*) (including one distant population) that exhibit various levels of phylogenetic relatedness. Scores indicative of co-ancestry were derived from proportion of identical codons at PSS in relation to the simulated distribution of codon usage expected under the scenario of convergent evolution and co-ancestry. The proportion of identical codons at PSS was calculated in a pairwise fashion among sequences of all pairs of species (or populations).

---

## Discussion

Our study addressed the relative contribution of convergent evolution and co-ancestry in shaping functional MHC class II variation within a lemur community (*Cheirogaleidae*). To examine the potential of parallel parasite-mediated selection across species, expected under convergent evolution, we integrated MHC data with an assessment of helminth prevalence to test whether particular MHC supertypes are associated with infection by particular parasites, or with a risk of multiple infestations in four lemur species. We further investigated the relative contribution of the two evolutionary mechanisms to MHC allelic similarity across species by inferring allelic co-ancestry from nucleotide similarity at sites directly involved in interactions with pathogens versus other sites, using MHC sequences from several lemur species with variable phylogenetic relatedness. Below, we first discuss patterns of parasite prevalence and MHC variation in the four lemur species before evaluating the links between multi-host parasite communities and the distribution of functionally similar MHC alleles across host species. Finally, we discuss the potential contribution of co-ancestry versus convergence in shaping MHC allelic similarity.

### **Parasite community overlap, MHC variation and selection patterns**

First, we evaluated the potential of parallel selection pressures across hosts, expected to be substantial under the convergent evolution hypothesis, by assessing the extent of overlap in helminth communities. Although the composition of helminth communities differed across species, four out of eight helminth genera were present in more than two host species. *M. berthae* and *M. coquereli* appeared as the most similar host species based on helminth prevalence. They share three out of four helminth genera with both direct and indirect transmission routes, which may reflect the substantial overlap in their diets, i.e. the amount of fruit/animal matter (Dammhahn and Kappeler 2014). Similarly, overlap in helminth prevalence and dietary ranges have been described between *C. medius* and *M. murinus* (Schwensow et al. 2010b; Dammhahn and Kappeler 2014). These infection patterns were confirmed in our dataset, with seven out of eight helminth genera shared between these two species.

*M. murinus*, the most abundant host, displayed the highest allelic richness in both loci as well as the highest levels of parasitism (proportion of infected individuals and parasite richness), which may reflect intense pathogen-driven selection on MHC. Several other selective pressures may contribute to explain higher allelic richness, such as large population size, that can simultaneously provide larger reservoir for pathogens and support greater number of susceptible individuals (e.g. Nunn et al. 2003; Hughes and Page 2007; Morand 2015). Additionally, intensity of selection on MHC has been previously shown to increase in social vs. solitary species (Hambuch and Lacey 2002; de Bellocq et al. 2009; Srithayakumar et al. 2012). Thus, spatial proximity and the richer social life of *M. murinus*, compared to other cheirogaleids, may contribute to explain increased parasite exposure (see Pechouskova et al. 2015 for a more detailed discussion).

Divergent patterns of past positive selection were observed across the four *Cheirogaleidae*. In *C. medius*, a striking difference between the two loci, with 19 positively selected sites (PSS) detected in DRB and only one in DQB, was also reflected by a slightly lower allelic richness in DQB (24) vs. DRB (36). In other sympatric lemurs, allelic richness was higher in DQB than in DRB, but only in *M. berthae* it was also accompanied by a high PSS density. Divergent patterns of past selection at different loci may indicate a potential differentiation of the functional role of these loci in different hosts, as was previously suggested in *M. murinus*, where DRB was under stronger diversifying selection and more influential on MHC-based mate choice than DQB (Huchard et al. 2012; 2013). Past and current demographic fluctuations most likely contribute substantially to the overall variation in MHC polymorphism across species (not between loci of the same species), and allelic richness was, accordingly, higher in the two most common species (*M. murinus* and *C. medius*).

### **Convergent evolution or co-ancestry maintains MHC functional similarity?**

The operation of convergent evolution implies that shared pathogen-driven pressures may select for similar MHC alleles or allelic motifs across species (Gustafsson and Andersson 1994; Christin et al. 2010), but the empirical evidence supporting this hypothesis is rare. To disentangle between convergence and similarity by descent requires simultaneous examination of the evolution of coding and non-coding regions as well as identification of



---

dominant parallel selective pressures, which remains a challenging task (summarized in Klein et al 2007; Lenz et al. 2013).

Here, we attempted to integrate both approaches to test the potential of parallel selective pressures in shaping allelic similarity at coding sites across different species. To do this, we first evaluated the extent of functional overlap among MHC alleles by classifying them according to their binding specificity into 7 DRB and 12 DQB supertypes across the four host species. Except for two DRB and two DQB supertypes, all were shared by two to four species. We subsequently detected an association between species similarity in the distribution of MHC supertypes and helminth prevalence (DRB  $r=0.76$ , DQB:  $r=0.77$ ), which was comparable in strength to the association previously found in a multi-host-parasite ecological network of 11 rodent species and 26 helminth taxa (DRB  $r=0.62$ ; Pilosof et al. 2014), suggesting that MHC-parasite associations may often occur beyond the host species level. However, with only four lemur species (vs. 11 rodent species for Pilosof et al. 2014), we lacked statistical power to reach significance.

Next, we examined whether the possession of particular supertypes would influence particular helminth infestations across individuals from different species. We found that a possession of a particular supertype (SR4) favors infestation by a common helminth - *Hymenolepis* sp. (*Cestoda*) - in the two *Microcebus* sp. MHC-parasite associations shared by two very closely related species, *M. murinus* and *M. berthae*, which diverged less than ca 9 Mya (Thiele et al. 2013), might however reflect the long-term maintenance of particular allelic motifs with a similar evolutionary origin, rather than convergent evolution in response to shared parasites. Association of particular allelic motifs with infestation by particular helminths has already been described in more distantly related *M. murinus* and *C. medius*, where *Mimu*-DRB\*28 was associated with increased risk of *Ascaris* sp. infection and showed a functional overlap with a particular *Chme*-DRB sequence (Schwensow et al. 2010b), but the underlying mechanism of this functional convergence was not identified. In our study, this allele was grouped into SR4, which was associated with increased risk of infestation by *Hymenolepis* sp. in the two *Microcebus* sp. and also with a higher risk of multiple infestations in *M. murinus* (data not shown). Moreover, the possession of a particular DQB supertype promoted the risk of multiple infestations in the two *Microcebus* sp. indicating that particular alleles or allelic motifs may be associated with multiple parasites. Disadvantageous character of some MHC

alleles in connection to parasite prevalence (see e.g. Harf and Sommer 2005; Meyer-Lucht and Sommer 2005; Froeschke and Sommer 2005, 2012; Bonneaud et al. 2006; Loiseau et al. 2008) has been interpreted as a support for ‘rare allele advantage hypothesis’, further enhanced by antagonistic effects of some alleles that can explain their persistence and frequency shifts in the population (Froeschke and Sommer 2012).

Despite a relatively high amount of protein encoding genes estimated to be possessed by intestinal helminths (Pearce and Tarlton 2002), immune responses of mammalian hosts inducing a typical CD4+ T helper cell type 2 (Th2) cytokine response (Finkelman et al. 2004; Perrigoue et al. 2008; Allen and Maizels 2011) seem to vary very little. Host immune system may have thereby only limited ability to distinguish among different helminth species (Finkelman et al. 2004). Some alleles (or supertypes) may have a generalist profile by interacting directly with “ubiquitous” antigenic peptides, which may be part of molecules underlying the basic architecture or physiology of many parasites (see also Wegner et al. 2003; Tollenaere et al. 2008; Oliver et al. 2009; Froeschke and Sommer 2012; Pilosof et al. 2014, Tobler et al. 2014), in a similar way as large-spectrum antibiotic or anti-parasitic agents may inhibit or kill organisms belonging to broad taxonomic groups. Such “broad-spectrum” supertypes may then confer protection against several parasites from any parasite community, and would consequently likely be under purifying selection. They may rarely undergo the drastic drops in frequency expected under frequency-dependent selection and may easily be passed-on from one host species to the next during speciation events.

More generally, envisioning MHC-parasite interactions at the community level may have profound implications for our conceptual understanding on MHC-parasite dynamics. For example, the lack of host-specificity of many parasites, together with the sharing of MHC supertypes among several species, may largely disrupt or inhibit the co-evolutionary cycles (Pilosof et al. 2014; Tobler et al. 2014) that have long been assumed to occur under negative frequency-dependent selection (Snell 1968; Bodmer 1972; Spurgin and Richardson 2010). Indeed, theory proposes that parasites may often evolve resistance against MHC alleles (or supertypes) that rise in frequency, and these alleles may subsequently drop in frequency after losing their advantage. As envisaged by Pilosof et al. (2014), a multi-host perspective may cast doubts on the plausibility of such dynamics, because both MHC alleles and parasites that are common in one host may be rare in

---

others, so changes in MHC-parasite interactions may commonly occur at the community, rather than at the single population scale. As a result, it may be urgent to move from the population level framework prevailing in the evolutionary literature on MHC towards more integrative perspectives envisaging MHC-host-parasite dynamics at the community level.

Finally, we explored whether the inter-specific functional similarity among MHC alleles in species with gradient phylogenetic relatedness stems from their long maintenance over time (similarity by descent), or arose independently as a result of convergence. First, higher similarity at non-PSS of MHC alleles belonging to the same vs. different supertype indicated that their similarity arose from co-ancestry. Second, while phylogenetic reconstructions based on PSS alignment lacked species-specific clustering, as described previously in other taxa (e.g. Kriener et al. 2000; Lenz et al. 2013; Eimes et al. 2015), the inter-specific clustering was less supported but still present in the tree based on neighbouring non-PSS codons, similarly to patterns recently described in crows or in ardeid birds (Li et al. 2011; Eimes et al. 2015). Inter-specific clustering at non-PSS rather supports the co-ancestry of similar alleles belonging to different species and the maintenance of similar allelic motifs at non-PSS sites over long periods of time might be facilitated by their close proximity to PSS (Klein et al. 1998; O'hUigín et al. 2000). It is possible that co-ancestry at PSS has been to date largely overlooked to date due to lack of site-specific analyses (Lenz et al. 2013).

Further support for the co-ancestry of shared alleles across species comes from the codon usage analysis, which showed that when different species exhibited similar amino acids in the anticipated antigen-binding groove of their MHC molecules, these amino acids were often encoded by similar codons (or, strictly speaking, these codons were more similar than expected by chance). This suggests that sequences of orthologous alleles show signs of co-ancestry even at PSS, which strongly supports the co-ancestry hypothesis by indicating that functional similarity is inherited and arises from the long-term retention of alleles or allelic motifs across multiple speciation events, rather than resulting from independent evolutionary pathways. This pattern was similar to those previously observed among two stickleback species (Lenz et al. 2013). Moreover, the scores indicative of co-ancestry for DRB were significantly correlated with phylogenetic distances among species, suggesting that functionally similar alleles are progressively lost as species

diverge. In contrast, the relative contribution of convergence versus co-ancestry to allelic functional similarity did not increase among species that harbored more similar parasite communities. This may indicate that convergent evolution contributes very little to trans-species allelic similarity, even though similar parasite pressures should also favor the long-term retention of related alleles, so that such relationship may also be expected under the co-ancestry hypothesis. In addition to our small sample size, it is possible that our non-invasive gut parasite screening did not accurately capture the similarity of parasite communities in different host species, which encompasses a variety of other extracellular parasites and pathogens.

It has been argued that alternative explanations to co-ancestry and convergence may generate the observed patterns. For example, allelic introgression through hybridization, or gene conversion and recombination, may produce structurally and functionally similar alleles across species (Wegner and Eizaguirre 2012). However, allelic introgression is rather unlikely to affect patterns observed in *Cheirogaleidae* since there is no recent hybridization, except for the distant populations of *M. murinus* and *M. griseorufus* (Gligor et al. 2009; Sommer et al. 2014). The operation of convergent evolution has rarely been shown and may predominantly explain allelic similarity limited to coding regions among species that split at least 30 Mya (e.g. human and bovine: Andersson et al. 1991; primates and rodent: Yeager et al. 1997; humans and New world monkeys: Kriener et al. 2001; skunks and racoons: Srithayakumar et al. 2012). The estimated divergence times are smaller in our sample of *Cheirogaleidae* (*Microcebus* sp. ca. 7–11 Ma, *Microcebus* sp. vs. *Mirza* sp. 12-18 Ma), except perhaps for the divergence time separating *Cheirogaleus medius* from other genera (*Cheirogaleidae* 19-26 Ma; Thiele et al. 2013) which could, in theory, be long enough to detect potential effect of convergent evolution at the sequence level. In contrast, well-supported species segregation at non-PBR partition compared to PBR partitions of similar length as in our DRB (21 codons) that would suggest potential convergent evolution, was detected between the two species of stickleback that diverged relatively recently (7 Ma, Bell and Foster 1994). However, this has been attributed to a possible intra-specific recombination and gene conversion (Lenz et al. 2013). Additionally, balancing selection could in theory over-ride the power of divergent parasite-mediated pressures promoting species diversification as it has been noted to homogenize, rather than diversify, MHC alleles in sympatric species after speciation,

despite of differences in parasite communities and erosion of background genomic diversity (e.g. Fraser and Neff 2010; Tobler et al. 2014). Our study adds to an emerging empirical body by showing that trans-species polymorphism primarily reflects the long-term maintenance of MHC alleles, even among species that diverged about 30 Mya ago, suggesting that some polymorphisms may survive for million years and multiple speciation events.

In conclusion, our results show that parasite communities are partially shared by sympatric lemurs. Functionally similar MHC alleles (supertypes), or allelic motifs, may affect susceptibility to similar parasites and some supertypes may encompass higher risk of multiple infections across several host species in a similar manner. Moreover, we show that allelic similarity at coding as well as non-coding sites resulted from the long term maintenance of ancestral sequence motifs or allelic lineages across speciation events, sometimes for as long as 20-30 million years. These results have at least two important implications. First, they contribute to a growing body of evidence suggesting that the evolutionary ecology of MHC should move on from the population level framework towards a more integrative, multi-host and multi-infection perspective. This may lead us to revisit some of the fundamental mechanisms envisioned to explain the exceptional MHC polymorphism. Second, they imply that MHC polymorphism is a cumulative capital, gathered through the interactions between multiple hosts and diverse parasite communities over evolutionary times and passed on during speciation events, a richness which may be slowly recovered when eroded, and should become a focus for conservation.

### **Acknowledgements**

We would like to thank all people who have contributed to the collection and processing of DNA samples, with a special thanks to the Kirindy field assistants, Melanie Dammhahn and Christina Glaschke. We also thank the Département de Biologie Animale, Université d'Antananarivo, the CAFF of the Direction des Eaux et Forêts and the CNFEREF Morondava for granting a research permit. Special thanks go to Prof. Lutz Walter and Dr. Christian Roos for use of the sequencing platform, and Nico Westphal for technical help. This research was funded by a SAW project of Leibniz Graduate School for the Foundations of Primate Social Behaviour issued by Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e.V. (SAW-2011-DPZ-4) and field work was supported by the DPZ and DFG (Ka1082/10-1&2).

# APPENDICES

# APPENDIX 1

## CHAPTER 2

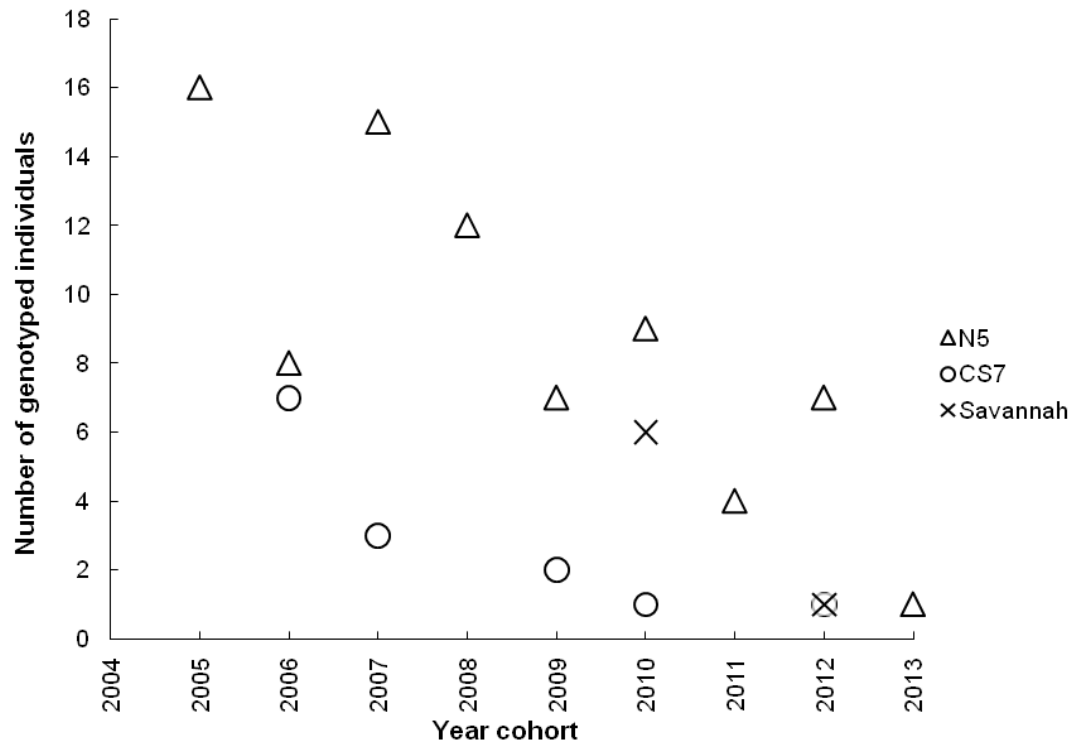
MHC class II variation in a rare and ecological specialist mouse lemur reveals lower allelic richness and contrasting selection patterns compared to a generalist and widespread sympatric congener

(1) ELECTRONIC SUPPLEMENTARY MATERIAL

(2) ELECTRONIC SUPPLEMENTARY DATASET



## ELECTRONIC SUPPLEMENTARY MATERIAL



**ESM 1** Sample size distribution of newly captured individuals within three study areas (N5, Savannah and CS7) between the years 2005-2013 (2010 for Savannah). Only individuals included in the downstream analyses are shown. Trapping sessions for N5 were conducted 6-10 times a year and less regularly in Savannah and within the area of the known species distribution in CS7

ESM 2 PCR reaction mix and conditions.

***Roche High Fidelity PCR system***

---

<b><i>PCR reaction mix</i></b>	2.5µl Buffer; 1.8mM MgCl <sub>2</sub> ; 0.5 dNTPs;		
	0.25 High Fidelity Taq Polymerase;		
	2µl of genomic DNA template.		

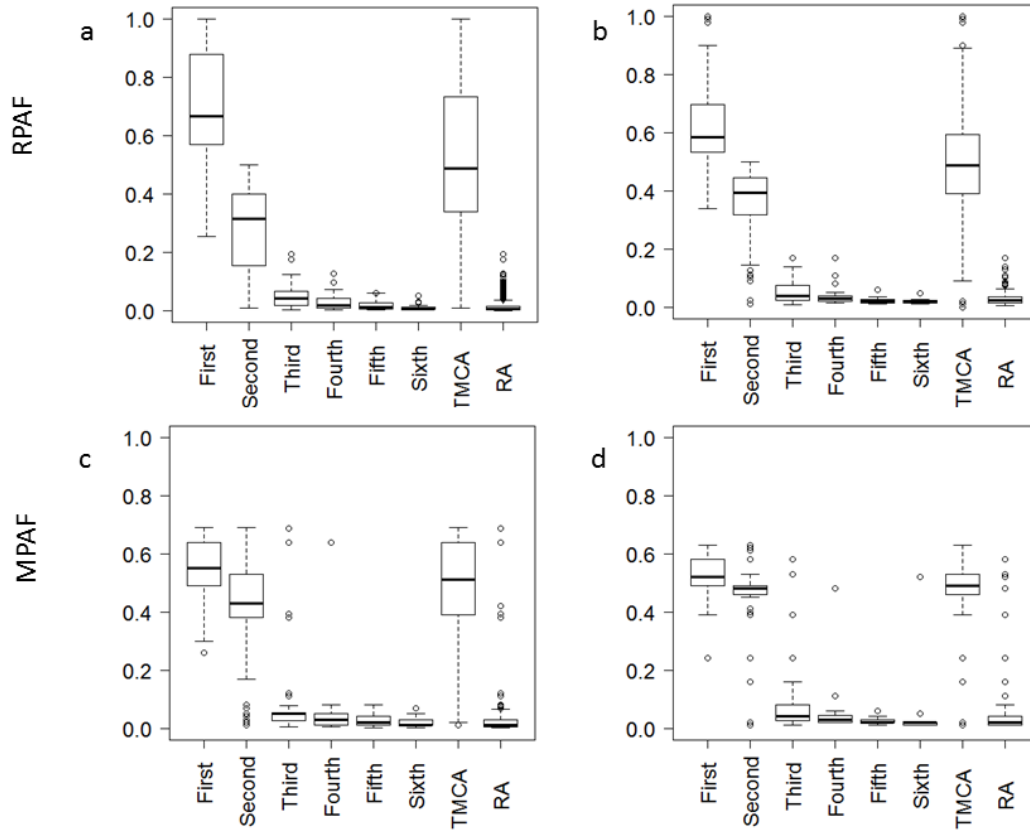
		Temperature	Time
<b><i>PCR conditions</i></b>	1x	95°C	2 min
	35x	60 <sup>1</sup> -62 <sup>2</sup> °C	30 s
		72°C	45 s
	1x	72°C	7 min

---

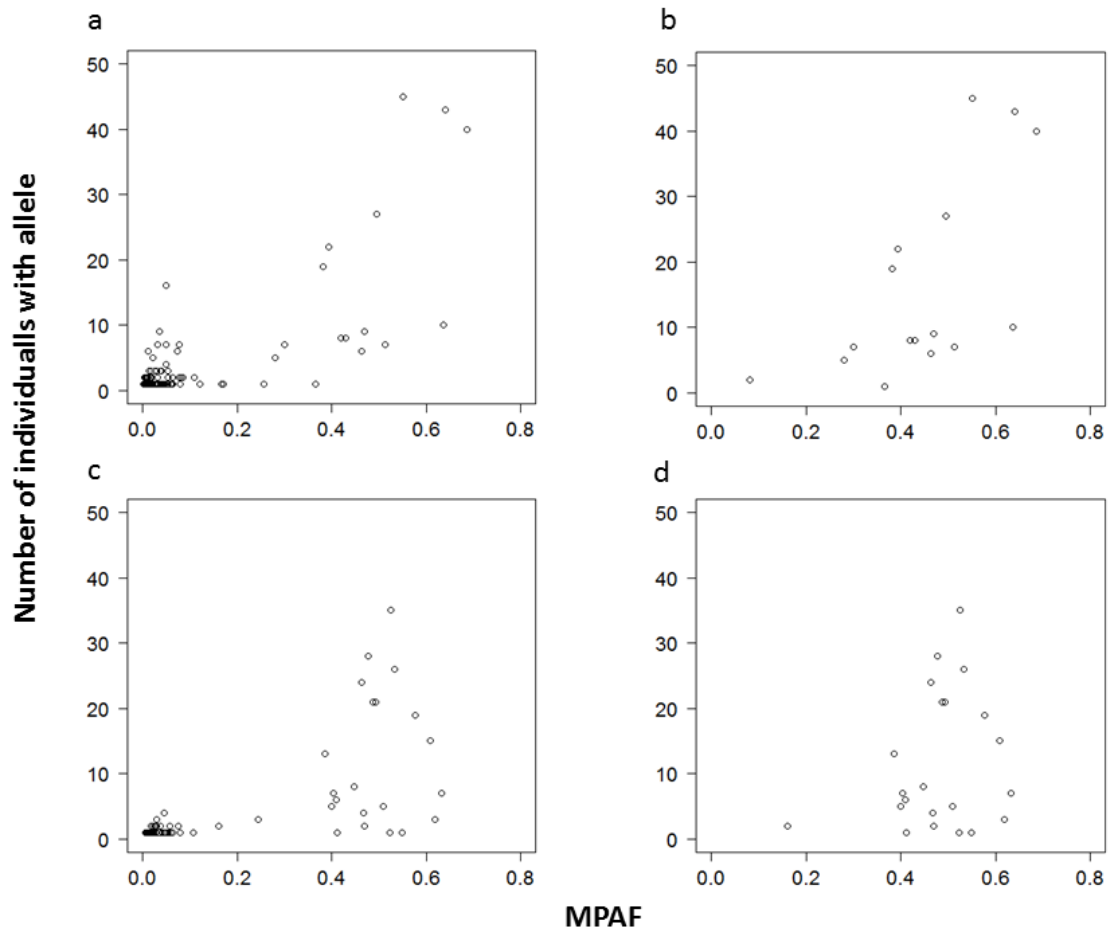
<sup>1</sup>DRB; <sup>2</sup>DQB

**ESM 3** Summary statistics of sequencing outcome for all reads passing the initial quality control procedure that required: i) a minimum of five reads for each sequence within individual amplicon, ii) a minimum two different sequences per amplicon, iii) minimum sequence frequency of 5% among all sequences of an amplicon (Huchard et al. 2012).

	<i><b>DRB</b></i>	<i><b>DQB</b></i>
<i><b>Nr_reads per amplicon</b></i>		
Mean±SD	302.50±34.64	184.20±16.68
Median	169.5	101.5
Min-max	5-2255	5-832
<i><b>Nr_sequences per amplicon</b></i>		
Mean ±SD	44.51±0.98	2.72±0.18
Median	2	2
Min-max	1-130	1-14
<i><b>Total amplicons</b></i>	148	130
<i><b>Total individuals</b></i>	96	98
<i><b>Total reads</b></i>	44 770	23 948



**ESM 4** Distribution of relative per amplicon frequency (RPAF) for DRB (a) and DQB (b) and mean per-amplicon frequency (MPAF) for DRB (c) and DQB (d) of the first to sixth most common variants across amplicons and TMCA (two most common variants) and RA (remaining variants)



**ESM 5** Mean per amplicon frequency (MPAF) in relation to allelic frequency (number of individuals carrying a given allele) for DRB alleles before (a) and after (b) allele sorting and for DQB alleles before (c) and after (d) allele sorting

**ESM 6** The overview of the allelic frequency distribution within each sub-population N5, Savannah and CS7 (DQB: a, DRB: b) and across year cohorts 2005-2013 (DQB: c, DRB: d) within the largest sub-population (N5). Sequences *Mibe*-DQB\*U018, \*U019, \*U020, \*U021, \*U022 and *Mibe*-DRB U017 were found in one individual only.

(a)	Allele_DQB	N5	Savannah	CS7
	<i>Mibe</i> -DQB*001	0.14	0.10	0.11
	<i>Mibe</i> -DQB*002	0.13	0.20	0.04
	<i>Mibe</i> -DQB*003	0.12	-	0.07
	<i>Mibe</i> -DQB*004	0.08	0.20	0.18
	<i>Mibe</i> -DQB*005	0.09	0.10	0.11
	<i>Mibe</i> -DQB*006	0.10	-	0.04
	<i>Mibe</i> -DQB*007	0.06	-	0.07
	<i>Mibe</i> -DQB*008	0.06	-	0.11
	<i>Mibe</i> -DQB*009	0.04	0.10	-
	<i>Mibe</i> -DQB*010	0.02	0.10	0.07
	<i>Mibe</i> -DQB*011	0.03	-	0.04
	<i>Mibe</i> -DQB*012	0.03	-	-
	<i>Mibe</i> -DQB*013	0.03	-	0.04
	<i>Mibe</i> -DQB*014	0.02	0.10	-
	<i>Mibe</i> -DQB*015	0.01	-	0.07
	<i>Mibe</i> -DQB*016	0.03	-	-
	<i>Mibe</i> -DQB*017	0.01	0.10	-
	<i>Mibe</i> -DQB*U018	-	-	0.04
	<i>Mibe</i> -DQB*U019	-	-	0.04
	<i>Mibe</i> -DQB*U020	0.01	-	-
	<i>Mibe</i> -DQB*U021	0.01	-	-
	<i>Mibe</i> -DQB*U022	0.01	-	-
(b)	Allele_DRB	N5	Savannah	CS7
	<i>Mibe</i> -DRB*001	0.19	-	0.19
	<i>Mibe</i> -DRB*002	0.16	0.10	0.12
	<i>Mibe</i> -DRB*003	0.14	0.10	0.08
	<i>Mibe</i> -DRB*004	0.12	0.20	0.04
	<i>Mibe</i> -DRB*005	0.07	0.10	0.15
	<i>Mibe</i> -DRB*006	0.06	-	0.12
	<i>Mibe</i> -DRB*007	0.03	0.10	0.04
	<i>Mibe</i> -DRB*008	0.03	0.10	0.08
	<i>Mibe</i> -DRB*009	0.03	0.10	0.04
	<i>Mibe</i> -DRB*010	0.03	-	0.04
	<i>Mibe</i> -DRB*011	0.03	-	0.04
	<i>Mibe</i> -DRB*012	0.03	-	-
	<i>Mibe</i> -DRB*013	0.03	0.10	-
	<i>Mibe</i> -DRB*014	0.01	-	0.08
	<i>Mibe</i> -DRB*015	0.02	0.10	-
	<i>Mibe</i> -DRB*016	0.01	-	-
	<i>Mibe</i> -DRB*U017	0.01	-	-

(c)	Allele_DQB	2005	2006	2007	2008	2009	2010	2011	2012	2013
	<i>Mibe-DQB*001</i>	0.06	0.13	0.20	0.21	0.07	0.22	0.25	-	-
	<i>Mibe-DQB*002</i>	0.09	0.13	0.10	0.21	0.07	0.11	0.13	0.21	-
	<i>Mibe-DQB*003</i>	0.13	0.13	0.20	0.04	0.07	0.11	0.13	0.14	-
	<i>Mibe-DQB*004</i>	0.09	0.13	0.10	0.04	0.14	-	-	-	0.50
	<i>Mibe-DQB*005</i>	0.13	-	0.03	0.13	0.21	0.06	-	0.14	-
	<i>Mibe-DQB*006</i>	0.09	0.13	-	0.13	-	0.11	0.13	0.29	0.50
	<i>Mibe-DQB*007</i>	0.09	0.06	0.03	0.04	0.14	0.06	-	-	-
	<i>Mibe-DQB*008</i>	0.03	0.06	0.07	0.04	0.07	0.06	-	0.21	-
	<i>Mibe-DQB*009</i>	0.06	-	-	0.04	-	-	0.38	-	-
	<i>Mibe-DQB*010</i>	0.03	0.06	0.03	-	-	-	-	-	-
	<i>Mibe-DQB*011</i>	0.09	-	0.03	-	-	0.06	-	-	-
	<i>Mibe-DQB*012</i>	0.03	0.06	0.03	0.04	0.07	-	-	-	-
	<i>Mibe-DQB*013</i>	-	-	0.07	-	-	0.17	-	-	-
	<i>Mibe-DQB*014</i>	0.03	-	-	0.04	-	0.06	-	-	-
	<i>Mibe-DQB*015</i>	-	-	-	-	0.07	-	-	-	-
	<i>Mibe-DQB*016</i>	-	0.06	0.07	-	0.07	-	-	-	-
	<i>Mibe-DQB*017</i>	0.03	-	-	-	-	-	-	-	-
	<i>Mibe-DQB*U018</i>	-	-	-	-	-	-	-	-	-
	<i>Mibe-DQB*U019</i>	-	-	-	-	-	-	-	-	-
	<i>Mibe-DQB*U020</i>	-	0.06	-	-	-	-	-	-	-
	<i>Mibe-DQB*U021</i>	-	-	0.03	-	-	-	-	-	-
	<i>Mibe-DQB*U022</i>	-	-	-	0.04	-	-	-	-	-
(d)	Allele_DRB	2005	2006	2007	2008	2009	2010	2011	2012	2013
	<i>Mibe-DRB*001</i>	0.22	0.13	0.03	0.25	0.3	0.17	0.13	0.43	0.25
	<i>Mibe-DRB*002</i>	0.19	0.13	0.2	0.08	0.1	0.11	0.5	0.14	-
	<i>Mibe-DRB*003</i>	0.06	0.13	0.2	0.21	-	0.22	0.25	-	0.25
	<i>Mibe-DRB*004</i>	0.09	0.13	0.1	0.21	-	0.11	0.13	0.21	-
	<i>Mibe-DRB*005</i>	0.09	0.13	0.1	0.08	0.1	-	-	-	-
	<i>Mibe-DRB*006</i>	0.03	0.06	0.07	0.04	-	0.06	-	0.21	-
	<i>Mibe-DRB*007</i>	0.09	-	-	0.04	0.1	-	-	-	-
	<i>Mibe-DRB*008</i>	0.03	0.06	0.03	-	-	-	-	-	0.25
	<i>Mibe-DRB*009</i>	-	0.06	0.03	-	0.1	0.06	-	-	-
	<i>Mibe-DRB*010</i>	0.09	-	0.03	-	-	0.06	-	-	-
	<i>Mibe-DRB*011</i>	-	-	0.07	-	-	0.17	-	-	-
	<i>Mibe-DRB*012</i>	0.03	0.06	0.03	0.04	0.1	-	-	-	-
	<i>Mibe-DRB*013</i>	0.03	0.06	0.07	-	-	-	-	-	-
	<i>Mibe-DRB*014</i>	-	0.06	-	-	0.1	-	-	-	-
	<i>Mibe-DRB*015</i>	0.03	-	-	0.04	-	0.06	-	-	-
	<i>Mibe-DRB*016</i>	-	-	-	-	0.1	-	-	-	0.25
	<i>Mibe-DRB*U017</i>	-	-	0.03	-	-	-	-	-	-

## ELECTRONIC SUPPLEMENTARY DATASET



**Table 3** A List and accession numbers of *Mibe*-DRB and -DQB sequences described by this study. The number of individuals (N) from which sequence was retrieved and length (bp) is given. Sequences retrieved from one individual only were not submitted to a public repository to ensure the storage of high quality sequences only

<b>Allele name</b>	<b>N</b>	<b>Length</b>	<b>Genbank accession number/nucleotide sequence (5`-3`)</b>
<i>Mibe</i> -DRB*001	33	163	LN610539
<i>Mibe</i> -DRB*002	27	163	LN610540
<i>Mibe</i> -DRB*003	25	163	LN610541
<i>Mibe</i> -DRB*004	21	163	LN610542
<i>Mibe</i> -DRB*005	17	163	LN610543
<i>Mibe</i> -DRB*006	11	163	LN610544
<i>Mibe</i> -DRB*007	7	163	LN610545
<i>Mibe</i> -DRB*008	6	163	LN610546
<i>Mibe</i> -DRB*009	6	163	LN610547
<i>Mibe</i> -DRB*010	6	163	LN610548
<i>Mibe</i> -DRB*011	5	163	LN610549
<i>Mibe</i> -DRB*012	5	163	LN610550
<i>Mibe</i> -DRB*013	5	163	LN610551
<i>Mibe</i> -DRB*014	4	163	LN610552
<i>Mibe</i> -DRB*015	4	163	LN610553
<i>Mibe</i> -DRB*016	2	163	LN610554
<i>Mibe</i> -DRB*U017	1	163	CAGCGGGTGCGGCTCCTGGTGAGAGGCATCTACAACCGCGAGGAGTTCCTGCGCTACGACAGCGACGTGG GCAAGTACCGGGCGGTGACGGAGCTGGGCCGCGGACGCCGAGTCCTTGAACCGCCAGCAGGACCACCT GGAGCAGAGGCGGGCCGCGGTGG
<i>Mibe</i> -DQB*001	25	163	LN610555
<i>Mibe</i> -DQB*002	20	163	LN610556
<i>Mibe</i> -DQB*003	19	163	LN610557
<i>Mibe</i> -DQB*004	18	163	LN610558
<i>Mibe</i> -DQB*005	17	169	LN610559
<i>Mibe</i> -DQB*006	16	169	LN610560
<i>Mibe</i> -DQB*007	13	163	LN610561
<i>Mibe</i> -DQB*008	11	169	LN610562
<i>Mibe</i> -DQB*009	7	163	LN610563
<i>Mibe</i> -DQB*010	6	163	LN610564
<i>Mibe</i> -DQB*011	6	169	LN610565

Allele name	N	Length (bp)	Genbank accession number/nucleotide sequence (5`-3`)
<i>Mibe-DQB*012</i>	5	163	LN610566
<i>Mibe-DQB*013</i>	5	163	LN610567
<i>Mibe-DQB*014</i>	5	163	LN610568
<i>Mibe-DQB*015</i>	3	169	LN610569
<i>Mibe-DQB*016</i>	3	163	LN610570
<i>Mibe-DQB*017</i>	2	163	LN610571
<i>Mibe-DQB*U018</i>	1	169	CAGCGGGTGCGGAGTGTGAACAGATACATCTACAACCAGGAGGAGTTCGTGCGCTTCGACAGCGACATCG GCTTGGGCGAGTACCTGGCGGTGACGGAGCTGGGCCGGCCGGAGGCCGAGCACTGGAACCGCCAGCAGGA CCTCCTGGAGCAGAAGCGGGCCGCGGTGG
<i>Mibe-DQB*U019</i>	1	163	CAGCGGGTGCGGTATGTGACCAGATACATCTACAACCGCGAGGAGACCGTGCCTTCGACAGCGACGTGG GCGAGTACCTGGCCATGACGCCGCTGGGCCGGCCGGACGCCGAGTACTGGAACCGCCAGCAGGACATCCT GGAGCAGACGCGGGCCGAGCTGG
<i>Mibe-DQB*U020</i>	1	169	CAGCGGGTGCGGCTTGTGACCAGATACATCTACAACCAGGAGGAGTTCGTGCGCTTCGACAGCGACATCG GCTTGGGCGAGTACCGGGCCGTGACGGAGCTGGGCCGGCCGGACGCCGAGTCCTGGAACCGCCAGCAGGA CTTCATGGAGCAGAGGCCGGCCGAGGTGG
<i>Mibe-DQB*U021</i>	1	163	CAGCGGGTGCGGCATGTGGTCAGACACATCTACAACCGGGAGGAGTACGTGCGCTTCGACAGCGACGTGG ACGAGTACCGGCCGGTGACGGAGCTGGGCCGGCCGGACGCCGAGTACTGGAACCGCCAGCAGGACATCAT GGAGCGGAAGCGGGCCGAGCTGG
<i>Mibe-DQB*U022</i>	1	163	CAGCGGGTGCGGCTTGTGACCAGATACATCTACAACCGCGAGGAGTACGTGCGCTTCGACAGCGACGTGG GCGAGTACCGGGCCGTGACGCCGCTGGGCCGGCCGGACGCCGAGTACTGGAACCGCCAGCAGGACTTCCT GGAGCAGACGCGGGCCGAGCTGG

# APPENDIX 2

## CHAPTER 3

Evolutionary origins of shared MHC polymorphism in sympatric lemurs

(1) SUPPORTING INFORMATION 1

(2) SUPPORTING INFORMATION 2

## SUPPORTING INFORMATION 1

## TABLE OF CONTENTS

## SUPPLEMENTARY METHODS

1. Study site, DNA collection and MHC genotyping .....	1	<b>Chyba! Záložka není definována.</b>	8
1.1. Capture and DNA sampling protocol.....	1	<b>Chyba! Záložka není definována.</b>	8
1.2. MHC genotyping.....	1	<b>Chyba! Záložka není definována.</b>	8
1.3. Sample size and population genetics.....		<b>Chyba! Záložka není definována.</b>	19
2. Processing 454 sequencing data.....	1	<b>Chyba! Záložka není definována.</b>	2
2.1. Allele sorting and evaluation of loci duplication .....			122
2.2. Evaluation of allele sorting efficiency and genotyping reliability .....			124
3. Parasite communities screening .....			124
3.1. Sampling effort across host species .....			124
3.2. Quality control tests .....			124
3.3. Sample sizes for inter-specific comparison of helminth communities.....			125
4. MHC variation, patterns of molecular selection and supertype classification.....			127
4.1. Test of positive selection.....			127

## SUPPLEMENTARY RESULTS

5. Processing 454 data.....			128
5.1. Allele sorting and evaluation of loci duplication .....			128
5.2. Evaluation of allele sorting efficiency and genotyping reliability.....		<b>Chyba! Záložka není definována.</b>	
6. Parasite communities screening.....			137
6.1. The overview of helminth infestations across hosts.....			137
6.2. Quality control test.....			137
6.3. Interspecific comparison of helminth communities .....			137
7. MHC variation, patterns of molecular selection and supertype classification.....		<b>Chyba! Záložka není definována.</b>	
7.1. Test of positive selection.....		<b>Chyba! Záložka není definována.</b>	
8. Characterisation of supertypes.....			142
9. Selection of similar supertypes across species: parasite-host-supertype associations .....			144
10. Co-ancestry vs. convergence maintaining MHC functional similarity.....		<b>Chyba! Záložka není definována.</b>	
10.1. Phylogenetic reconstruction of PSS and non-PSS sequence alignment.....			146
10.2. Codon usage analysis .....			150

## SUPPLEMENTARY REFERENCES

## Supplementary methods

### 1. Study site, DNA collection and MHC genotyping

#### 1.1. Capture and DNA sampling protocol

Regular trapping sessions targeting the four lemur species have been conducted within 12,500-ha Kirindy Forest concession in western Madagascar (44°39' E, 20°03' S, Kappeler and Fichtel 2012) at four different study sites study areas (locally known as N5, CS7, Savannah and Kirindy village); 6-10 times a year since 1993 in N5 and CS7 and less regularly in Savannah (since 2010) and Kirindy village (since 2012) using Sherman and Tomahawk live traps baited with pieces of banana. Traps were set on trail intersections every 25 or 50 m within each study site (e.g. Dammhahn and Kappeler 2008). The centres of the study areas N5 & CS7 and Savannah & CS7 are situated ca. 2-2.5 km and N5 & Savannah ca. 4-4.5 km away from each other. The study area Kirindy village is roughly 7-9 km distant from the two closest areas (N5, CS7) and even more distant from Savannah (ca 11km). All study areas, except for Kirindy village, are inter-connected by continuous forest growth.

At first capture, each individual was briefly anaesthetized (0.01-0.04 mL Ketanest 100, s.c.) (Rensing 1999), individually marked by injecting a sub-dermal micro-transponder (Trovan, Usling, Germany), and a small ear biopsy of 2-3 mm<sup>2</sup> was taken and preserved in 70 % ethanol.

#### 1.2. MHC genotyping

Genomic DNA was extracted from small ear biopsies following standard protocol (Qiagen QIAmp DNA Mini-Kit, Qiagen Germany). PCR amplification targeting the two highly polymorphic loci of the MHC class II region, DRB and DQB, was performed using primers flanking the functionally important exon II that captured the full allelic variation in *M. murinus* and *M. berthae* for both loci (Schad et al. 2004; Averdam et al. 2011; Huchard et al. 2012; Pechouskova et al. 2015) and for DRB in *C. medius* (Schwensow et al. 2007). PCR reaction-mix and amplification conditions for both loci were identical to those described previously for *M. berthae* (Pechouskova et al. 2015). Sequencing primer design and the preparation of locus-specific sequencing libraries were

conducted following Huchard et al. (2012). MHC genotyping was performed using high throughput sequencing (454 pyrosequencing, Roche, France) following standard bi-directional amplicon sequencing guidelines for GS Junior sequencing (Roche, France). The sequencing data of *M. berthae* have been described in detail in Pechouskova et al. (2015).

### **1.3. Sample size and population genetics**

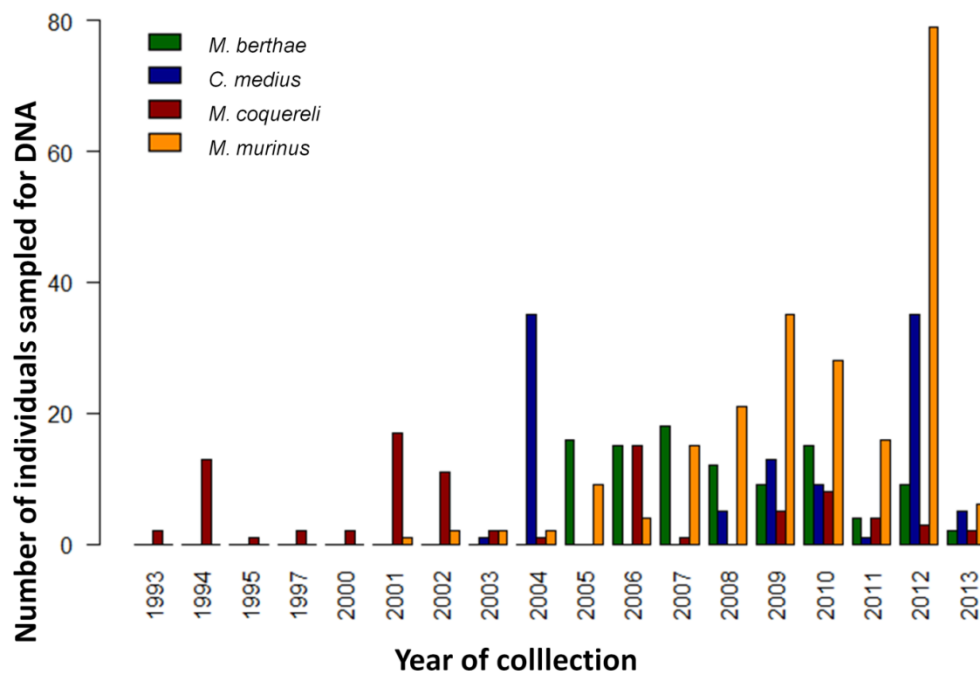
The amount of individuals captured within four study areas (N5, CS7, Savannah and Kirindy village; in Kirindy forest) during the years 1993 to 2013, which were included in this study, are presented in Table S1. Their sample size distribution across year cohorts is displayed in Fig. S1.

In order to test whether individuals from a same species coming from different areas are part of a same population, we investigated whether sub-populations captured at different study areas may represent reproductively independent units ( $F_{ST}$ , 10,000 permutations, Arlequin 3.5.1.3., Excoffier and Lischer 2010). Due to limited sample size in *M. coquereli*, we only conducted this test in *C. medius* (*Cm*) and *M. murinus* (*Mm*). Despite spatial distance between sampling areas, we found no indication of population differentiation in either host species (*Cm*  $F_{ST} < 0.05$ , *Mm*  $F_{ST} < 0.03$ ). Thus, all samples originating from the four study areas were pooled in species-specific manner for the downstream analysis.

**Table S1.** The total number of individuals per each host species captured across four study areas during the years 1993-2013. The sample sizes available for *M. berthae* are shown for comparison (in Pechouskova et al. 2015).

	CS7*			Kirindy village			N5			Savannah			Subtotal
	♀	♂	N.A.	♀	♂	N.A.	♀	♂	N.A.	♀	♂	N.A.	
<i>M. berthae</i>	-	1	13	-	-	-	37	42	1	3	3	-	100
<i>M. murinus</i>	40	23	-	2	10	-	47	41	-	35	22	-	220
<i>C. medius</i>	10	1	-	5	4	-	30	36	2	13	3	-	104
<i>M. coquereli</i>	17	12	2	-	-	-	24	9	10	12	3	-	89

\*Additionally, 13 previously captured and genotyped *M. murinus* individuals from CS7 study area presented previously in Huchard et al. (2012), for which parasite data were available, were included in downstream analyses.



**Figure S1.** The number of individuals captured for the first time per each host species during our sampling period (1993-2013)

## 2. Processing 454 sequencing data

Raw sequencing data derived from a total of six independent sequencing runs were processed and evaluated following previously described protocol of initial post-sequencing quality control procedure (in Huchard et al. 2012). After the initial quality control, a large proportion of artefactual alleles might still be present in the dataset. Therefore, various protocols aimed at discriminating true vs. artefactual alleles have been proposed (reviewed in Babik 2010; Lighten et al. 2014).

Here we first describe steps implemented in allele sorting procedure for each species (and locus) separately (A) before quality control steps employed to assess allele sorting efficiency and reliability of our genotyping (B).

Sequences that were identified as true alleles after our allele sorting procedure, but were retrieved in one individual only were not submitted to public repositories to ensure storage of high quality sequences only. However, all sequences were included in downstream analyses to prevent elimination of rare alleles that could bias further population genetic analyses by generating null alleles.

### 2.1. Allele sorting and evaluation of loci duplication

Here we employed a multi-step approach to identify artefactual alleles generated by random PCR or sequencing errors, but also the non-random occurrence of run-specific errors (i.e. homopolymer under- or over calls), or cross-amplicon contaminations. Allele screening steps employed by this study are identical to those previously described for *M. berthae* (Pechouskova et al. 2015) and for *M. murinus* (Huchard et al. 2012).

Briefly, we first established two numeric indices characterising the frequency of each allele both within and across amplicons (each individual PCR product) based on: (i) the amount of sequencing reads retrieved for one particular allele averaged across all amplicons possessing it (mean per amplicon frequency ‘MPAF’) and (ii) the relative proportion of reads retrieved for one particular allele out of the total amount of reads returned within one given amplicon (relative per amplicon frequency ‘RPAF’). Three main types of artefactual alleles occurred within our dataset and were identified: (1) Artefactual alleles (further referred to as artefacts) originating from PCR or sequencing errors showing high similarity to parental true alleles within affected amplicons, either by



single point mutation or reading frame disrupting indels, or by recombination of the two parental sequences (chimeras); artefacts occur at low frequencies across and within amplicons (Babik et al. 2009); (2) Primer mismatches originating from occasional amplification of non-targeted locus (i.e., DRB locus is amplified by DQB primers or the reverse), as a consequence of using degenerate primers and of the high sequence similarity between DQB and DRB (see also Huchard et al. 2012). They never occurred among the two most common alleles (TMCA), they were identified in the final genotype of the individual amplicon, and they were not found in replicates. Additionally, these sequences mostly showed low MPAF and RPAF; (3) Cross-amplicon contaminants of true alleles that displayed high MPAF and were present in numerous amplicons, but showed low RPAF within affected amplicons and were absent from replicates.

For *M. murinus*, a previous study using 454 sequencing with a similar sequencing design has suggested in DRB and DQB as a reliable cut off value to discard artefactual alleles (Huchard et al. 2012). Here, we used this threshold for the initial screening steps and eliminated alleles with  $MPAF < 0.05$  in this species. However, we did not use this threshold to systematically eliminate artefactual alleles in the two other species (*C. medius* and *M. coquereli*) and followed the protocol previously implemented for *M. berthae* by Pechouskova et al. (2015).

Briefly, we first visualised the distribution of MPAF and RPAF from the 1<sup>st</sup> most common to the 6<sup>th</sup> most common allele within each amplicon of each species and locus, expecting an abrupt drop-off between the two most common alleles (TMCA) and the remaining alleles (RA) of each amplicon in the absence of loci duplication (see also Babik et al. 2009; Huchard et al. 2012). This step can provide an initial estimation of whether loci are duplicated or not, which is necessary to determine the sequencing depth (the total amount of reads per amplicon) necessary for reliable genotyping.

Subsequently, we re-evaluated the estimated number of loci by manual alignment of all alleles within each amplicon and identified artefactual alleles. Alleles remaining after allele sorting steps were retained as true alleles. Amplicons were systematically replicated when the presence of a given allele remained equivocal, or when the number of alleles retrieved was higher than expected given the estimated number of loci.

Last, using the acquired information of the number of loci determined in previous steps we used the program ‘Negative Multinomial’ developed by Galan et al. (2010) to

compute the amount of reads necessary for reliable genotyping. Throughout the sequencing process, all amplicons that did not return a sufficient amount of reads based on this estimate were re-genotyped.

## **2.2. Evaluation of allele sorting efficiency and genotyping reliability**

The efficiency of allele-sorting was evaluated through several steps. First, assuming that most artefacts would not occur independently in many amplicons, we would expect a positive correlation between the MPAF of each allele and the number of amplicons possessing this allele before allele sorting that disappears after allele sorting (Babik et al. 2009; Huchard et al. 2012). Second, given the expected variability of coding loci in MHC genes, the allelic divergence (i.e. rate of nucleotide substitutions) between true alleles of the same amplicon should be relatively high (e.g. Kelley et al. 2005; Babik et al. 2009). Here we calculated the average substitution rate between true alleles within each amplicon in MEGA 6 (Tamura et al. 2013) and compared it to the average substitution rate acquired across all true alleles detected for a given locus and species. Last, for each species and for each locus, a number of independent replicates were included across different sequencing runs to assess the reliability of retrieved genotypes.

## **3. Parasite communities screening at species level**

### **3.1. Sampling effort across host species**

The overview of coproscopic samples collected per each sampling month (September–December; 2012) for each host species is presented in Fig. S2.

### **3.2. Quality control tests**

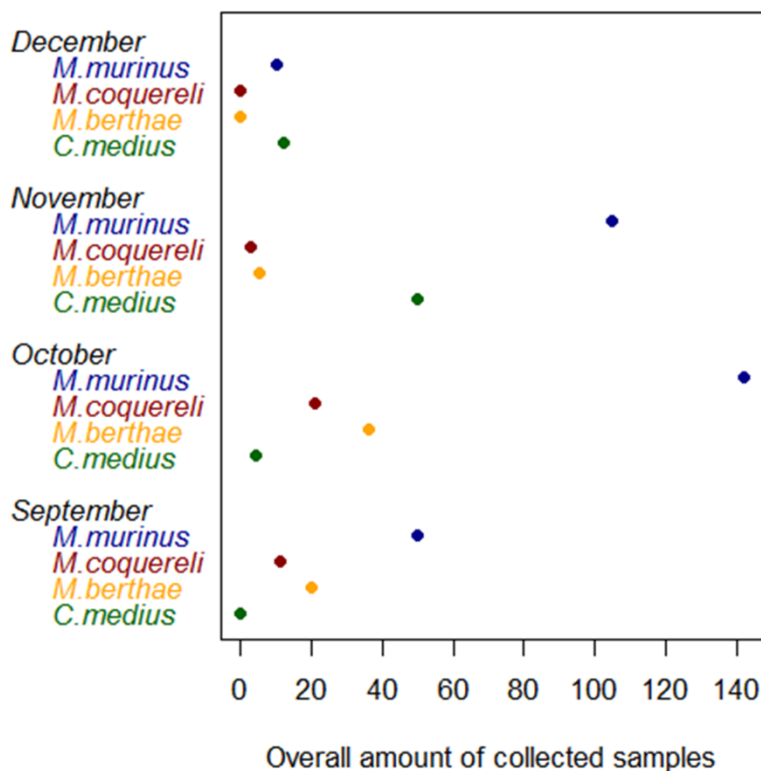
The repeatability of helminth egg detection was evaluated by performing a simple quality control procedure in several steps. First, we examined how different methods used to extract egg morphotypes (standard FLOTAC protocol ‘flotation’ and Ritchie’s formol-ether concentration method ‘sedimentation’) may affect the composition of helminth communities. To do this, we tested whether two samples from a same individual collected on the same trapping day (further referred to as ‘sample sets’) are (1) more similar if analyzed using the same ( $n_{sample\ sets}=55$ ) than different methods ( $n_{sample\ sets}=46$ ). Next, we

evaluated the repeatability of helminth egg detection within an individual by testing whether samples from a same individual collected on the same trapping day are (2) more similar to each other than any random sample within the same species regardless of the method used ( $n_{\text{sample sets}}$  per each host species, *M. murinus*=52, *M. berthae*=11, *C. medius*=26, *M. coquereli*=12) and also when comparing only samples analyzed by the same method ( $n_{\text{sample sets}}$  per each host species for flotation, *M. murinus*=20, *M. berthae*=11, *C. medius*=16, *M. coquereli*=12; and sedimentation, *M. murinus*=52, *M. berthae*=7, *C. medius*=26, *M. coquereli*=3). For all tests, we calculated Jaccard dissimilarity index for each pair of samples, using the function ‘vegdist’ implemented in the package ‘vegan’ (Oksanen et al. 2015), and averaged across groups (e.g., ‘same’ versus ‘different’). The resulting average was then compared using Mann-Whitney *U* tests.

### **3.3. Sample sizes for inter-specific comparison of helminth communities**

Tests of the inter-specific differentiation in helminth communities pooled several analyzed samples from a same individual that were collected within the same trapping session (max. 1-3 days apart), or in at least two different months within the sampling period (September-December 2012).

The amount of individuals included per each host species for each inter-specific comparisons are summarized in Table S2.



**Figure S2.** Overview of sampling effort (the amount of collected samples) per each host species within sampling month (September–December 2012). A fraction of these counts was collected from the same individuals (repeated records)

**Table S2.** The overview of individuals per each host species used for inter-specific comparisons of helminth communities.

	Inter-specific comparison	
	collected during same trapping session*	collected within two different months**
<i>M. berthae</i>	10	13
<i>C. medius</i>	25	27
<i>M. coquereli</i>	8	8
<i>M. murinus</i>	46	55
Total individuals	89	103

\*including 2-4 repeats per individual across all host species

\*\* including 4-11 repeats per individual across all host species

## 4. MHC variation and patterns of molecular selection across species

### 4.1 Test of positive selection

The presence of positive selection operating on nucleotide sequences is indicated by an increased ratio of non-silent ( $d_N$ ) to silent ( $d_S$ ) substitutions at codon sites. These positively selected sites (PSS) are characterized by  $\omega > 1$  ( $\omega = d_N/d_S$ ). The strength of positive selection was globally estimated using the likelihood ratio modeling approach where two models of codon evolution are compared: (i) the null model that assumes  $\omega < 1$  and varies according to the beta distribution (model M7), and (ii) an alternative model allowing an additional class of sites to account for a possible occurrence of PSS where  $\omega > 1$  (model M8) (Yang et al. 2000). If model M8 fits the data better than M7, PSS are identified through Bayes Empirical Bayes (BEB) procedure and retained for the next steps if statistically significant (CI 99 %) (Yang et al. 2005). The analysis was computed using the package CodeML implemented in PAML 4.7 (Yang 2007) for each locus and each species separately.

## Supplementary results

### 5. Processing 454 data

The summary of sequencing outcome is presented in Table S3. Below we describe details on allele sorting and loci duplication (A) before we briefly discuss allele sorting efficiency and reliability of our genotyping (B).

#### 5.1. Allele sorting and evaluation of loci duplication

The distribution of RPAF from the 1<sup>st</sup> to the 6<sup>th</sup> sixth most common allele averaged across all amplicons revealed a notable drop between the TMCA and RA in all three species, suggesting no indication of loci duplication (Fig. S3). Below, we describe the allele sorting procedure in more detail for each locus and species separately.

##### *M. murinus*

In DRB, 43 out of 204 *Mimu*-DRB sequences retrieved from 219 amplicons (including 24 replicates) were retained as true alleles (MPAF 0.09-0.83), including 2 new alleles. In DQB, 53 out of 156 *Mimu*-DQB sequences were retrieved out of 225 amplicons and retained as true alleles (MPAF 0.17-0.74); 13 of them were described for the first time.

The total of 18 DRB and 9 DQB amplicons possessed more than two true alleles. Given that loci duplication seems unlikely due to RPAF distribution in our dataset and based on previous findings (Schad et al. 2004, Averdarm et al. 2011; Huchard et al. 2012, Sommer et al. 2014), surplus alleles in 15 DRB and one DQB amplicons were removed given their low RPAF compared to TMCA and excluded by replication in three DQB amplicons. The remaining four DRB and five DQB amplicons were removed from further analysis due to inconclusive genotypes. The minimum amount of sequencing reads necessary for reliable genotyping was estimated on 18 reads. Conclusively, 4 DRB (and none for DQB) amplicons for which <18 reads were retrieved were successfully re-genotyped.

##### *C. medius*

A total of 137 sequences were retrieved for DRB locus from 118 amplicons (including 11 replicates). Importantly, we noticed that some of the identified primer-mismatches DQB  
128

sequences were described as DRB sequences in public repositories by previous studies using the same DRB primer pair (*Chme*-DRB\*3, \*16, \*30, \*19, \*37, \*45; Supporting Information 2), suggesting that the problem of primer-mismatch may occur regularly with these primers. However, all these sequences and three other sequences not detected within this study (*Chme*-DRB\*29, \*32, \*34) clustered with DQB sequences (Fig. S4) and their closest match using BLAST search were DQB sequences of other species (e.g. *M. murinus*); confirming our classification into the DQB locus.

Overall, 36 *Chme*-DRB sequences (MPAF 0.20-0.82), including 9 newly described, were retained as true alleles.

In DQB, genotyping 112 amplicons (including 10 replicates) yielded 120 sequences; 24 DQB sequences (MPAF 0.27-1.00), were retained as true alleles; 12 of which were detected for the first time.

After allele sorting, two DRB and one DQB amplicons possessed more than two true alleles. In both loci, the 3<sup>rd</sup> allele showed low RPAF compared to RPAF of TMCA and were therefore removed. Additionally, in one DRB amplicon, the existence of the 3<sup>rd</sup> allele (RPAF 0.08) was excluded by replication. These results supported our predictions of no duplication of either locus. Conclusively, 5 out of 9 DRB amplicons and 3 out of 4 DQB amplicons that returned <18 reads were re-genotyped and replaced in the dataset. Amplicons that could not be replicated were discarded from further analysis.

### ***M. coquerelli***

In DRB, 128 sequences were retrieved from 142 amplicons (incl. 53 replicates). Here the majority of artefactual alleles were represented by single point homopolymer indels or mutations (,HP<sup>+</sup>) (MPAF>0.26) that were highly run-specific and showed relatively high frequency of occurrence (1-12 amplicons). The unusually high frequency of HP in *M. coquerelli* was likely due to a high frequency of single-base repeats characterising the nucleotide sequence of both DRB and DQB exon 2. The artefactual status of each HP was verified by replication of at least one, but mostly numerous amplicons. In total, 11 newly described *Mico*-DRB alleles (MPAF 0.15-0.87) were retained as true alleles.

In DQB, we retrieved 76 DQB sequences from 124 amplicons (incl. 36 replicates). Similarly, as for DRB all artefactual alleles with MPAF>0.10 represented single point indels (HP) and occurred in one to 35 amplicons. Overall, both RPAF and MPAF of

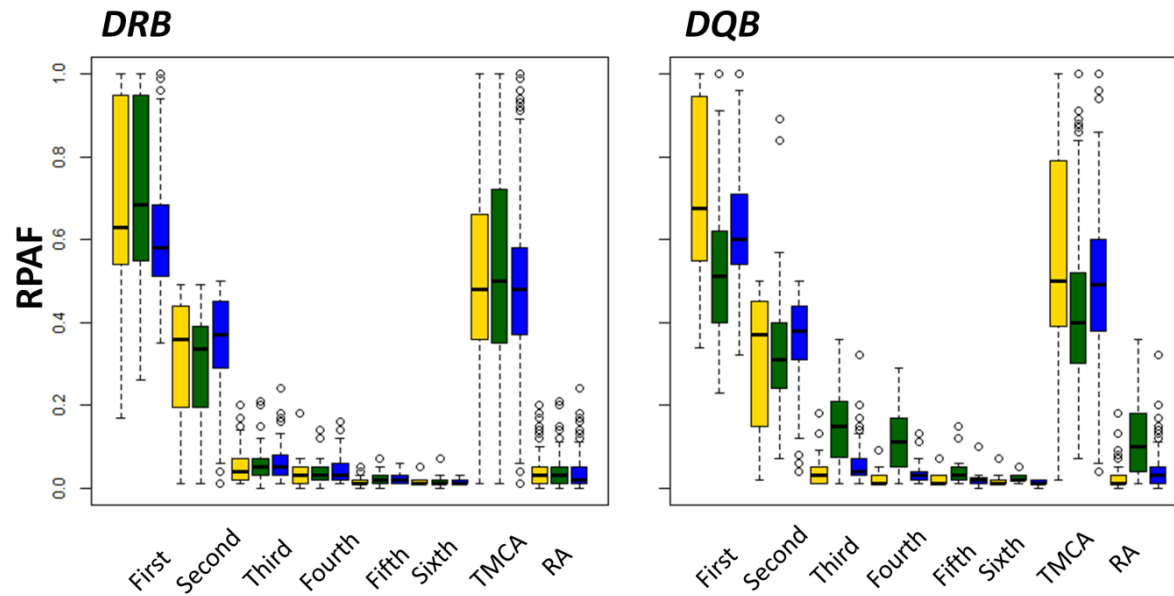
particular HP were very heterogeneous across sequencing runs (some runs were unaffected by some HP). Conclusively, 17 HP-affected amplicons were replicated, or triplicated and confirmed the artefactual status of single or multiple HP sequences. Overall, 16 unique *Mico*-DQB sequences were retained as true alleles (MPAF 0.12-0.75).

After allele sorting, 3 DRB and 1 DQB amplicons possessed more than two alleles. In DRB, the existence of a 3<sup>rd</sup> allele was rejected due to their low RPAF (DRB 0.02, 0.04; DQB 0.01) versus RPAF of TMCA (DRB 0.45-0.83, DQB 0.10-0.84), or via replication of affected amplicons. One amplicon, where the existence of a 3<sup>rd</sup> allele could not be evaluated through replication was excluded from further analysis. In this species, examination of the RPAF distribution from 1<sup>st</sup> to 6<sup>th</sup> most common allele correctly predicted no loci duplication in DRB, with a notable drop present between the TMCA and RA. In DQB, the RPAF distribution of TMCA was distorted by the elevated frequency of HP. Thus, our allele sorting procedure, identifying two true alleles in all individuals but one, was the main basis to infer an absence of locus duplication. Consequently, 29 DRB amplicons and 4 DQB amplicons with <18 reads were successfully re-genotyped and replaced in the dataset.

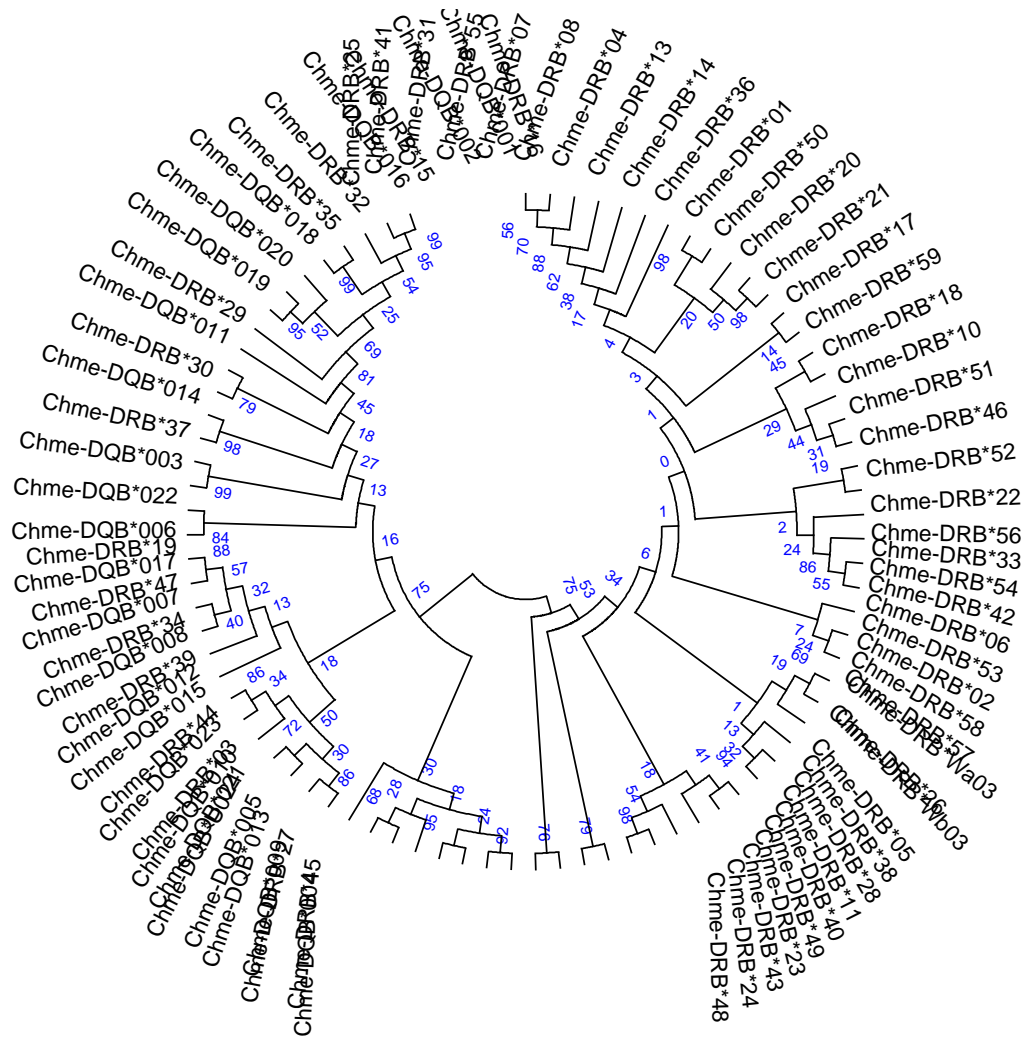


**Table S3.** Summary of sequencing outcome in DRB and DQB for the three *Cheirogaleidae* described in this study. Sequencing data for both loci from ca 100 individuals of *M. berthae* is described in Pechouskova et al. 2015.

	<i>Mirza coquereli</i>		<i>Cheirogaleus medius</i>		<i>Microcebus murinus</i>	
	DRB	DQB	DRB	DQB	DRB	DQB
<b>Nr. reads / amplicon</b>						
Mean±SD	239.61±25.35	128.82±10.32	243.58±22.86	204.25±24.05	236.39±15.56	144.84±11.46
Median	156	84	175	102	154	101
Min	5	18	5	5	12	15
Max	1557	565	1456	1215	1592	1629
Total	34025	15974	28742	22876	51770	32590
<b>Nr. sequences / amplicon</b>						
Mean±SD	3.04±0.18	3.50±0.12	2.87±0.21	2.81±0.22	3.08±0.14	2.52±0.09
Median	2	3	2	2	2	2
Min	1	1	1	1	1	1
Max	14	8	19	13	13	14
Total	142	124	118	112	219	225
Replicates of amplicons with <18 reads	29	4	5	3	4	-
Independent replicates	24	32	6	7	20	19
Total of succesfully genotyped individuals	88	88	100	97	192	201
N. sequences retrieved	128	76	137	118	204	156
N. true sequences retrieved	11	16	36	24	43	53



**Figure S3.** The distribution of the relative per amplicon frequency (RPAF) for DRB and DQB from the 1<sup>st</sup> to the 6<sup>th</sup> most common alleles across amplicons of showing all three species: *C. medius* (gold), *M. coquereli* (green) and *M. murinus* (blue) and for the TMCA (two most common alleles) and RA (remaining alleles) accordingly



**Figure S4.** The phylogenetic reconstruction of DRB and DQB nucleotide sequences of *C. medius* described in this study, or deposited in Genbank up to date. For accession numbers see the overview in Supporting Information 2. The gene tree was constructed using Neighbour-joining algorithm with Jukes-Cantor correction (1000 bootstrap) in MEGA 6 (Saitou and Nei 1987; Felsenstein 1985; Jukes and Cantor 1969; Tamura et al. 2013)

## 5.2. Evaluation of allele sorting efficiency and genotyping reliability

The overall efficiency of our allele-sorting procedure was evaluated through several steps. According to our assumption that artefactual alleles should not occur independently in many amplicons, we found a positive correlation between the MPAF of each allele and the number of amplicons possessing it. This correlation disappeared after allele sorting in all cases, except for DRB for *C. medius* (Fig. S5).

Next, for each species and locus, average pairwise substitution rates between alleles retrieved within same amplicon were comparable to the average substitution across the entire allelic pool of a given locus and species, and indicated that true alleles present in a given amplicon represent a random selection of all true alleles (t-tests,  $P > 0.05$  in all cases).

Last, 100 % (20 DRB, 19 DQB) of replicated amplicons included for *M. murinus* within independent sequencing runs demonstrated full repeatability of retrieved genotypes. Out of 6 DRB and 7 DQB replicates included for *C. medius*, all DRB and 5 DQB displayed perfect repeatability of acquired genotypes. The remaining 2 DQB replicates were affected by allelic drop out originating most likely from amplification bias between the TMCA, resulting in falsely homozygous genotypes. Both 2<sup>nd</sup> most common alleles were retrieved among sequences returning with  $< 5$  copies, the threshold used to filter low-quality sequences during initial trimming steps (for details see Huchard et al. 2012). From 24 DRB and 32 DQB replicates included for *M. coquereli*, all DRB and 22 DQB displayed a perfect repeatability of retrieved genotypes. The remaining 10 DQB replicates were affected by an extensive inflation of run-specific homopolymers (HP) with its parental allele (the allele these HP originated from) returning in  $< 5$  copies in each affected amplicon. Moreover, 6 of these amplicons that originated from one sequencing run were all characterised by low overall amount of reads (mean $\pm$ SE,  $49.71 \pm 5.77$ ); none of these amplicons was retrieved with non-matching true alleles.

To investigate the potential of a similar allelic dropout occurring within the rest of our dataset, we randomly selected 30 homozygotes and manually investigated all sequences retrieved for these amplicons that returned in  $< 5$  copies. We found no further indication of allelic dropout. Conclusively it seems a good precaution to increase the threshold for genotyping reliability when extensive HP inflation occurs, as well as to

perform random control of the sequences retrieving in <5 copies, especially in homozygous individuals, to prevent including false homozygous genotypes.

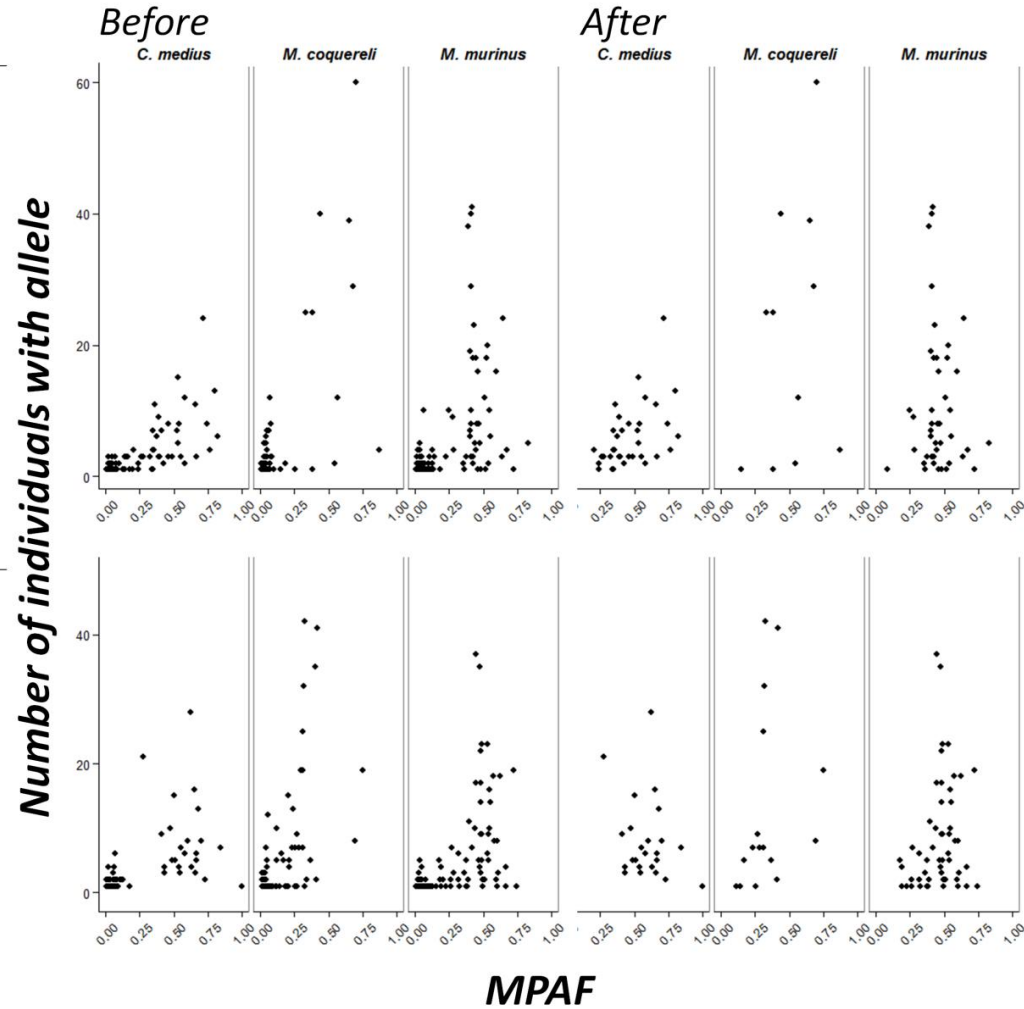
Eighty-nine percent of 108 replicates were repeatable. For the remaining 11 % we could identify the cause of non-repeatability, namely dropout of the 2<sup>nd</sup> most common allele. For *M. berthae*, we reported 100 % repeatability of genotypes for DRB and DQB in 23 and 27 amplicons (Pechouskova et al. 2015).

## DRB

	Before	After
<i>C. medius</i>	$r = 0.72$ $P < 0.001$ $n = 137$	$r = 0.46$ $P < 0.05$ $n = 36$
<i>M. coquereli</i>	$r = 0.68$ $P < 0.001$ $n = 148$	$r = 0.27$ $P < 0.42$ $n = 11$
<i>M. murinus</i>	$r = 0.57$ $P < 0.001$ $n = 213$	$r = -0.05$ $P < 0.75$ $n = 43$

## DQB

	Before	After
<i>C. medius</i>	$r = 0.60$ $P < 0.001$ $n = 118$	$r = -0.27$ $P < 0.20$ $n = 24$
<i>M. coquereli</i>	$r = 0.55$ $P < 0.001$ $n = 81$	$r = 0.25$ $P < 0.34$ $n = 16$
<i>M. murinus</i>	$r = 0.58$ $P < 0.001$ $n = 159$	$r = 0.22$ $P < 0.12$ $n = 53$



**Figure S5.** Mean per amplicon frequency (MPAF) in relation to allelic frequency (number of individuals carrying a given allele) for DRB and DQB alleles before and after allele sorting. The table on the left shows the strength of the correlation between MPAF of each allele and the number of amplicons possessing it **before** and **after** allele sorting for DRB and DQB, with n representing number of alleles included in each calculation

## **6. Parasite communities screening at species level**

### **6.1. Overview of helminth infestations across hosts**

The overview of helminth genera that could be reliably identified by microscopic evaluation and the amounts of individuals per each host species infected by particular helminth is presented in Table S4.

### **6.2. Quality control test**

Before characterizing the overlap in helminth communities across host species, we first performed a set of quality control tests to evaluate the repeatability of helminth egg detection using two different extraction methods (FLOTAC ‘flotation’; Ritchie’s formol-ether concentration method ‘sedimentation’). We detected that two samples collected from the same individual on the same trapping day were more similar when using the same (Jaccard dissimilarity index ( $d_j$ ), mean $\pm$ SD 0.56 $\pm$ 0.44) versus different methods ( $d_j$ ; mean  $\pm$  SD, 0.85 $\pm$ 0.29) (Mann-Whitney,  $W=1250$ ,  $P=0.008$ ). However, samples collected from the same individual on the same day were more similar to each other regardless of the method used ( $d_j$ ; mean  $\pm$  SD, 0.69 $\pm$ 0.41) than to any random sample from the same species ( $d_j$ ; mean  $\pm$  SD, 0.88 $\pm$ 0.26) ( $W=341571$ ,  $P < 0.001$ ). Similar results were obtained when including only samples analyzed by the same method (mean $\pm$ SD, same day vs. random sample, flotation 0.21 $\pm$ 0.35 vs. 0.96 $\pm$ 0.20,  $W=2402.5$ ,  $P < 0.001$ ; sedimentation, 0.67 $\pm$ 0.40 vs. 0.86 $\pm$ 0.27; Mann-Whitney  $W= 16640$ ,  $P < 0.001$ ). These findings show that repeatability of helminth egg detection is higher within than between the two extraction techniques. Nevertheless, given that samples collected from the same individual were significantly more similar to each other than to any random sample within the host species, regardless of the method used samples analyzed by both methods were integrated and used for further analysis.

### **6.3. Inter-specific comparison of helminth communities**

The dissimilarity matrix indicating the extent of overlap in helminth prevalence for each pair of host species are presented in Table S5.

**Table S4.** The number of individuals infected by a given helminth genus that were included in the analysis testing the correlation between cross-species similarity in MHC supertype frequencies and in helminth communities

Phylum	Family	Genus	Life cycle	Infection status	<i>Cheitrogateus medius</i>	<i>Microcebus berthae</i>	<i>Microcebus murinus</i>	<i>Mirza coquereli</i>	Number of individuals with infection
Nematoda	<i>Ascaridida</i>	<i>Ascaris</i> sp.	<i>Direct</i>	-	35	13	121	7	11
				+	2	1	7	1	
	Oxyuridae	<i>Lemuricola</i> sp.	<i>Direct</i>	-	35	14	124	8	6
				+	2	0	4	0	
				-	35	14	115	8	
				+	2	0	13	0	
	Strongylida	<i>Strongylida</i> sp.	<i>Direct</i>	-	36	14	113	8	16
				+	1	0	15	0	
	Subuluridae	<i>Subulura</i> sp.	<i>Indirect</i> <sup>1</sup>	-	31	14	43	7	92
				+	6	0	85	1	
Trichuridae	<i>Trichuris</i> sp.	<i>Direct</i>	-	35	11	92	4	45	
			+	2	3	36	4		
Cestoda	Hymenolepididae	<i>Hymenolepis</i> sp.	<i>Indirect</i> <sup>1</sup>	-	36	11	74	7	59
				+	1	3	54	1	
Trematoda	Heterophyidae	<i>Metagonimus</i> sp.	<i>Indirect</i> <sup>2</sup>	-	37	14	124	8	4
				+	0	0	4	0	

<sup>1</sup> Arthropod intermediate host, <sup>2</sup> one or more intermediate hosts (e.g. snail and arthropod)



**Table S5.** A dissimilarity matrix showing the extent of overlap in helminth prevalence within each pair of host species. Here an RI value of 1 indicates maximum dissimilarity, with no parasites shared between a pair of hosts and RI value of 0 indicates that parasites infect hosts with a similar prevalence.

	<i>C. medius</i>	<i>M. berthae</i>	<i>M. murinus</i>
<i>M. berthae</i>	0.831		
<i>M. murinus</i>	0.763	0.719	
<i>M. coquereli</i>	0.752	0.574	0.706

## 7. MHC variation and patterns of molecular selection

### 7.1. Test of positive selection

The outcome of the likelihood modelling approach testing the presence of positive selection operating on nucleotide sequences by comparing two models of codon evolution is presented in Table S6.

**Table S6.** Evaluation of the goodness of fit for different models of codon evolution and estimated parameter values.

Model	LnL	Kappa (ts/tv)	AIC	$\Delta$ AIC	Parameters	$\chi^2$	d.f.	P	np
<b><i>Mirza coquereli</i></b>									
<b><i>DRB</i></b>									
M0-one $\omega$	-474.85	2.32	954.34	39.97	$\omega=0.66$				
M7-nearly neutral with $\beta$	-456.32	2.70	918.03	3.66		3.09	2	n.s.	23
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ )	-454.77	2.41	914.37	<b>Best</b>	$p_0=0.62, p_1=0.38, \omega=2.27$				25
<b><i>DQB</i></b>									
M0-one $\omega$	-557.97	2.42	1120.78	85.20	$\omega=0.62$				
M7-nearly neutral with $\beta$	-527.18	2.88	1060.12	24.55		24.29	2	***	33
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ )	-515.03	2.76	1035.58	<b>Best</b>	$p_0=0.94, p_1=0.06, \omega=9.07$				35
<b><i>Microcebus berthae</i></b>									
<b><i>DRB</i></b>									
M0-one $\omega$	-922.27	0.79	1846.12	184.38	$\omega=0.61$				
M7-nearly neutral with $\beta$	-839.49	0.86	1680.70	18.96		18.69	2	***	35
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ )	-830.14	0.73	1661.74	<b>Best</b>	$p_0=0.78, p_1=0.22, \omega=3.01$				37
<b><i>DQB</i></b>									
M0-one $\omega$	-881.92	1.33	1766.50	150.52	$\omega=1.32$				
M7-nearly neutral with $\beta$	-819.36	1.65	1642.03	26.05		25.74	2	***	45
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ )	-806.49	1.50	1615.98	<b>Best</b>	$p_0=0.65, p_1=0.35, \omega=4.84$				47

Model	LnL	Kappa (ts/tv)	AIC	$\Delta$ AIC	Parameters	$\chi^2$	d.f.	P	np
<b><i>Cheirogaleus medius</i></b>									
<b><i>DRB</i></b>									
M0-one $\omega$	-1286.02	1.27	2574.58	263.63	$\omega=5.19$				
M7-nearly neutral with $\beta$	-1184.93	1.55	2372.95	62.01		61.46	2	***	73
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ )	-1154.20	1.27	2310.94	<b>Best</b>	$p_0=0.63, p_1=0.37, \omega=11.56$				75
<b><i>DQB</i></b>									
M0-one $\omega$	-791.05	1.99	1586.08	171.10	$\omega=1.04$				
M7-nearly neutral with $\beta$	-743.27	2.50	1491.54	76.55		77.15	2	***	49
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ )	-704.70	2.80	1414.98	<b>Best</b>	$p_0=0.98, p_1=0.02, \omega=76.25$				51
<b><i>Microcebus murinus</i></b>									
<b><i>DRB</i></b>									
M0-one $\omega$	-1359.95	1.05	2722.01	277.69	$\omega=1.03$				
M7-nearly neutral with $\beta$	-1239.76	1.30	2482.11	37.79		37.42	2	***	87
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ )	-1221.05	1.11	2444.32	<b>Best</b>	$p_0=0.75, p_1=0.25, \omega=3.56$				89
<b><i>DQB</i></b>									
M0-one $\omega$	-1728.08	1.51	3459.18	456.03	$\omega=0.66$				
M7-nearly neutral with $\beta$	-1555.91	1.75	3115.31	112.17		81.42	2	***	105
M8-positive selection with $\beta$ ( $\omega_0 \leq 1, \omega_1 > 1$ )	-1515.20	1.64	3033.68	<b>Best</b>	$p_0=0.95, p_1=0.05, \omega=6.15$				107

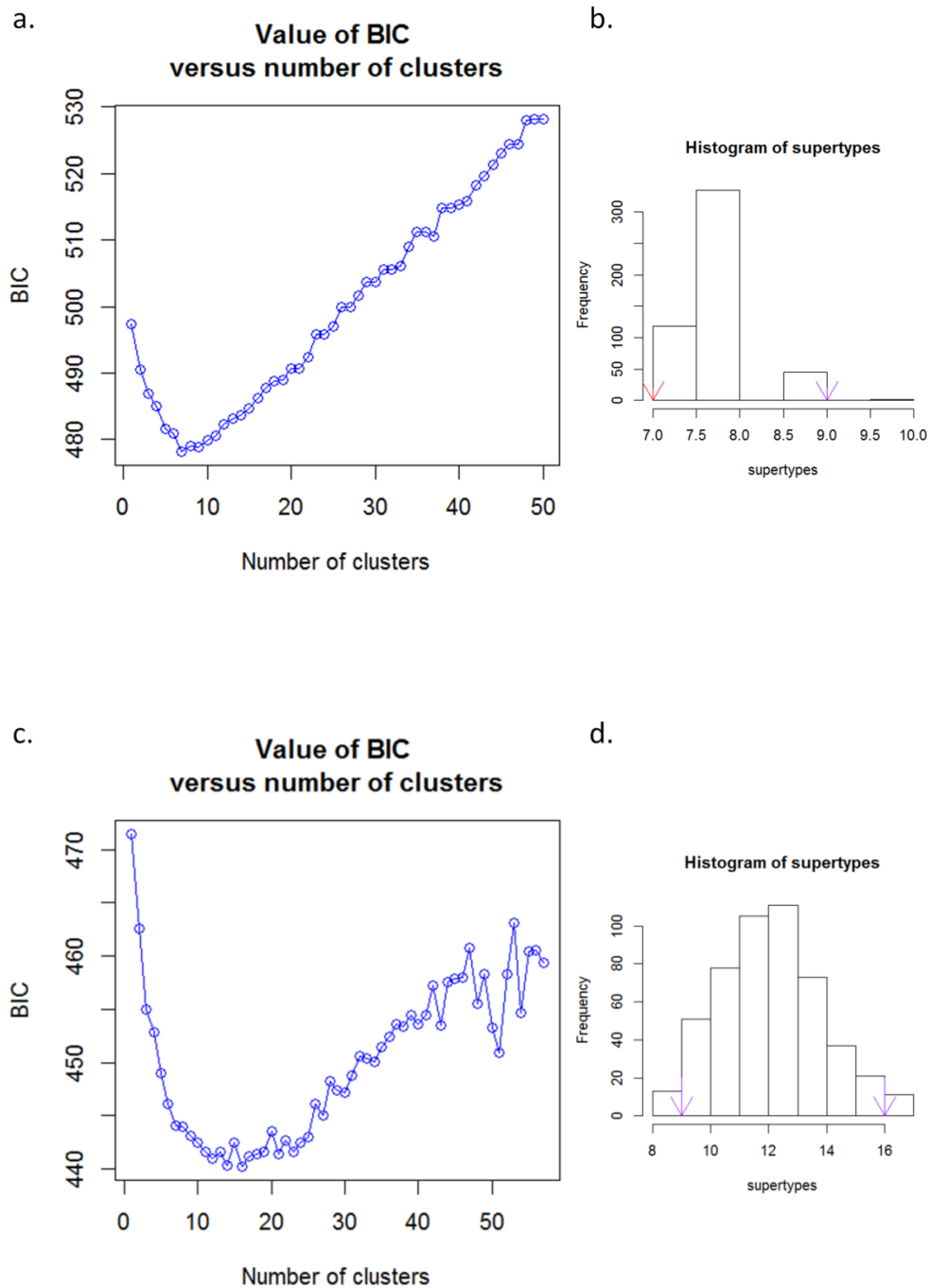
$\omega$ -dN/dS; nearly neutral with beta-for all sites  $\omega \leq 1$  and the beta distribution approximates  $\omega$  variation;

positive selection-a proportion of sites evolves with  $\omega > 1$ ;  $p_0$ -proportion of sites with  $\omega \leq 1$ ,  $p_1$ -proportion of positively selected sites ( $\omega > 1$ ),  $\omega_1$ -estimated value of  $\omega$  for sites under positive selection;

AIC-Akaike information criterion,  $\Delta$ AIC-difference between the value of the AIC of a given model and the best model

## **8. Characterisation of supertypes**

The overview of supertype clustering solutions for DRB and DQB are presented in Figure S6. An optimal number of clusters is chosen when BIC statistic decreases by a negligible amount, indicated by a minimum or an elbow in the curve of BIC values as a function of cluster number.



**Figure S6.** Distribution of BIC values for an increasing number of supertype clusters. The optimal number of supertypes was identified as 7 in DRB (Panel a) and 12 in DQB (Panel c) as indicated by a minimum in the curve. The distribution of the number of clusters obtained through 500 automated classifications is shown on the right for DRB (Panel b) and for DQB (Panel d)

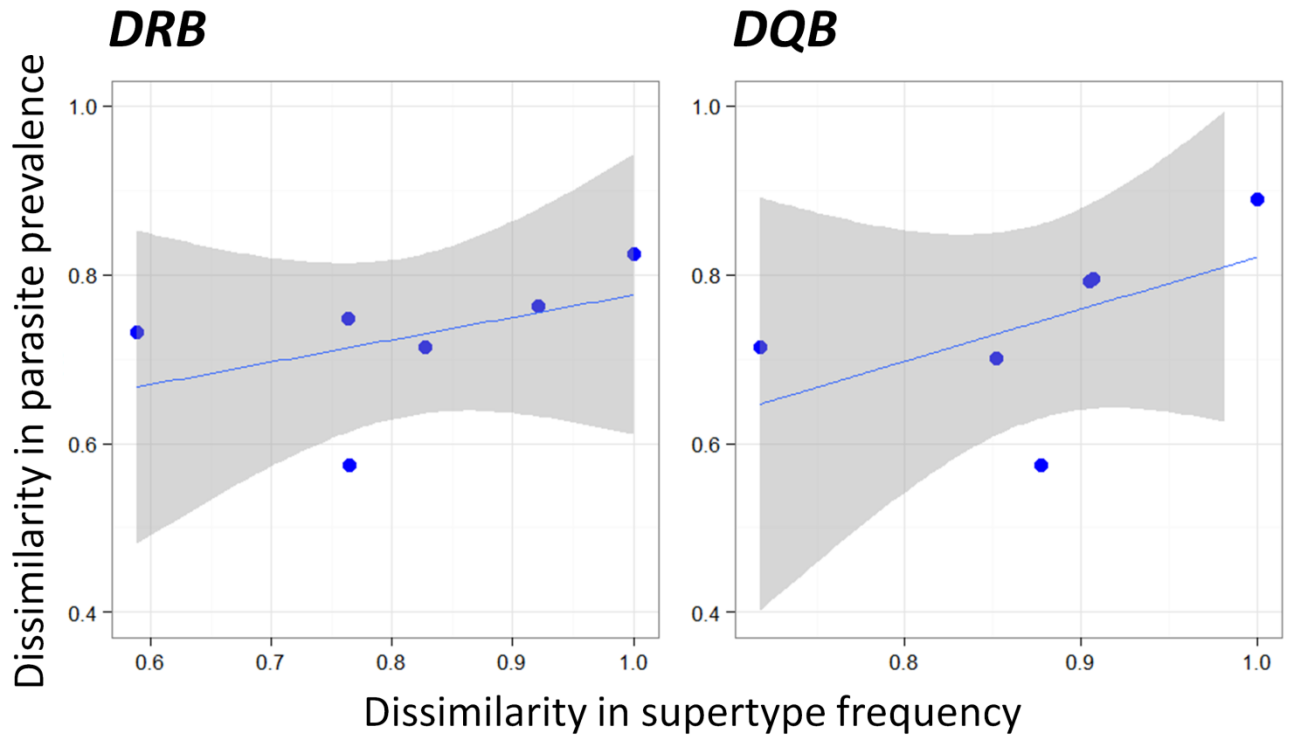
## 9. Selection of similar supertypes across species: parasite-host-supertype associations

The overview of supertype frequencies (the amount of individuals possessing a given supertype) within each species for DRB and DQB are presented in Table S7. Correlation between dissimilarities in infestation by particular helminths to the dissimilarities in possession of particular supertypes across the four host species (Partial mantel test controlling for phylogenetic signal) is displayed in Fig. S7.

**Table S7.** Supertype frequencies for DRB (labelled „SR1 - SR7“) and DQB (labelled „SQ1 – SQ12“) across four host species.

<b>DRB</b>	SR1	SR2	SR3	SR4	SR5	SR6	SR7	Supertype homozygots	<i>n<sub>ind</sub></i>
<i>M. berthae</i>	12	-	-	2	-	-	8	6	14
<i>M. murinus</i>	59	-	-	77	3	-	51	48	119
<i>C. medius</i>	-	8	20	1	12	17	-	12	35
<i>M. coquereli</i>	-	-	8	3	-	-	-	5	8
<b>Subtotal</b>	<b>71</b>	<b>8</b>	<b>28</b>	<b>83</b>	<b>15</b>	<b>17</b>	<b>59</b>	<b>71</b>	<b>176</b>

<b>DQB</b>	SQ1	SQ2	SQ3	SQ4	SQ5	SQ6	SQ7	SQ8	SQ9	SQ10	SQ11	SQ12	Supertype homozygots	<i>n<sub>ind</sub></i>
<i>M. berthae</i>	-	-	2	7	-	-	-	-	-	6	10	1	2	14
<i>M. murinus</i>	1	28	30	19	13	28	-	9	13	21	23	42	19	124
<i>C. medius</i>	-	6	-	-	3	-	12	17	12	1	-	2	15	34
<i>M. coquereli</i>	7	-	-	-	-	-	-	4	-	2	-	-	3	8
<b>Subtotal</b>	<b>8</b>	<b>34</b>	<b>32</b>	<b>26</b>	<b>16</b>	<b>28</b>	<b>12</b>	<b>30</b>	<b>25</b>	<b>30</b>	<b>33</b>	<b>45</b>	<b>39</b>	<b>180</b>



**Figure S7.** The relationship between pairwise dissimilarities in helminth prevalence and supertype frequencies among the four host species for DRB (a) and DQB (b); each data point represents a value obtained for each pair of species (6 comparisons)

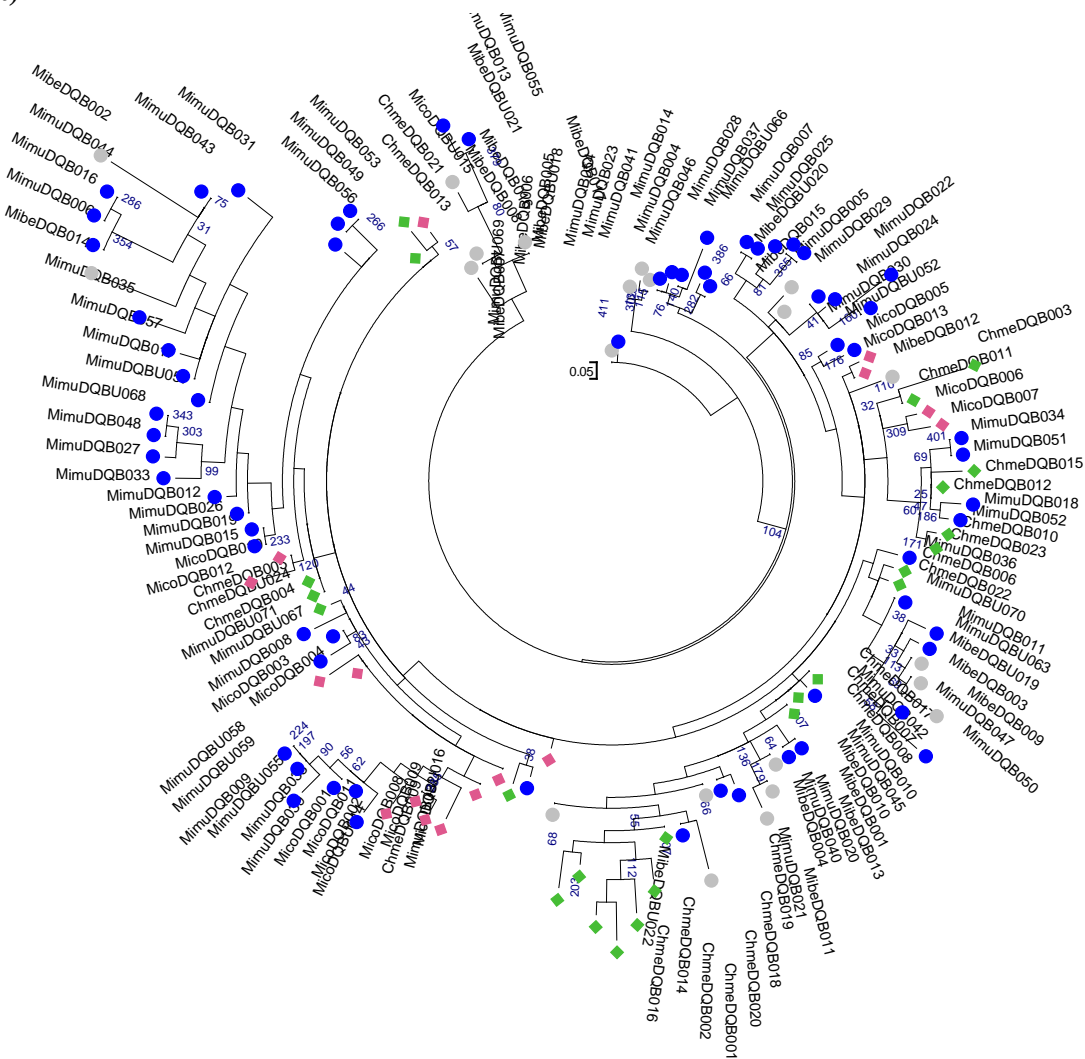
## 10. The role of a common ancestry vs. convergent evolution in maintaining MHC functional similarity

### 10.1. Phylogenetic reconstruction of PSS and non-PSS sequence alignment

**Figure S8.** The phylogenetic reconstruction of PSS and non-PSS sequence alignment for 128 DQB (Panel A) and 235 DRB (Panel B) sequences (Supplementary Information 2); *full circle blue*=*M. murinus*; *full circle grey*=*M. berthae*; *full circle red*=*M. rufus*; *full circle orange*=*M. griseorufus*; *green diamond*=*C. medius*; *pink diamond*=*M. coquereli*

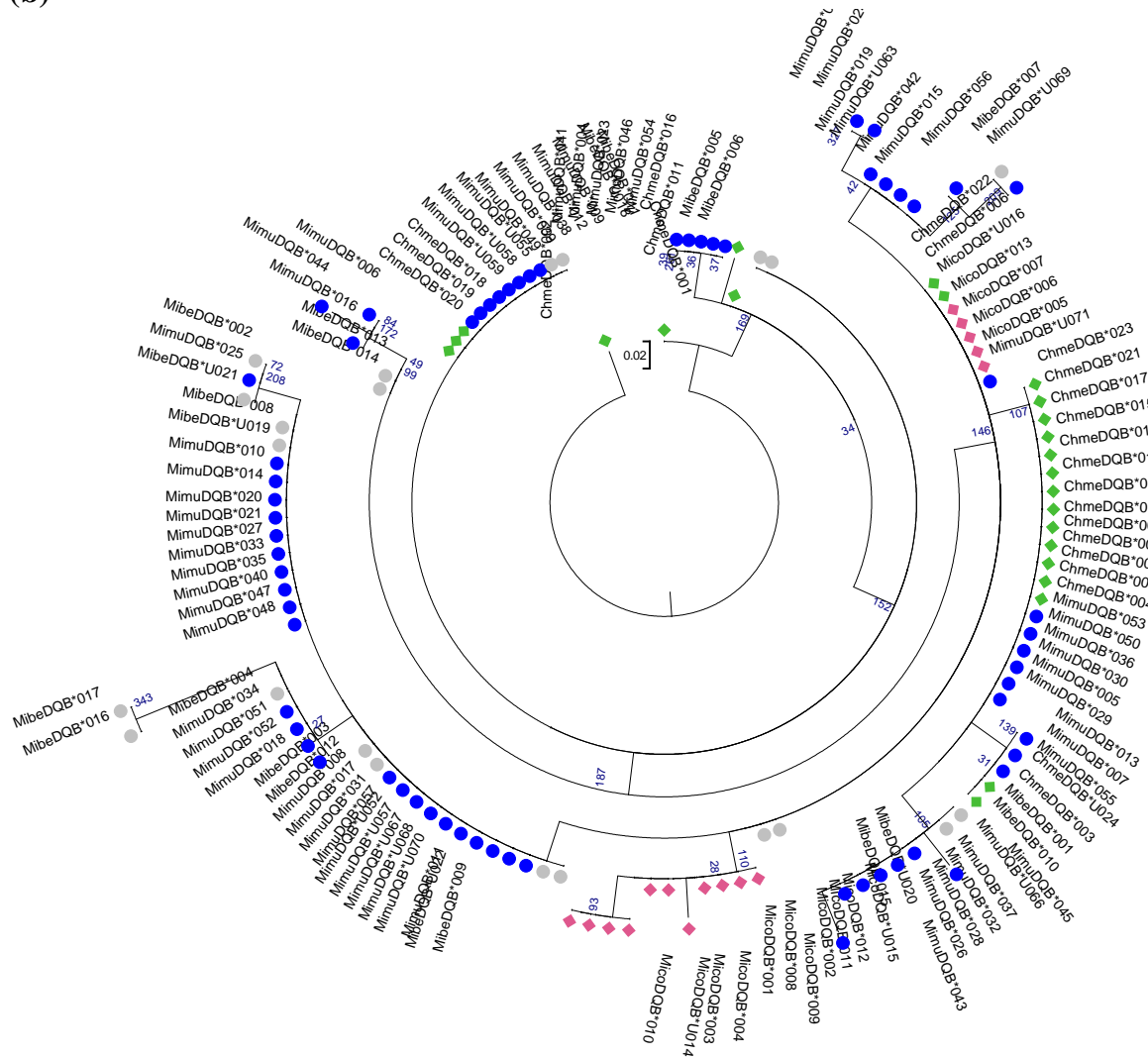
**Panel A.** Gene trees based on 12 DQB PSS (a) and a sequence partition of complimentary length based on 12 randomly selected non-PSS DQB codons (b). The tree configuration was derived by the ML algorithm implemented in PhyML (a: TIM2+I+G substitution model, p-inv=0.13,  $\gamma$  shape=1.01; b: TPM1uf+G,  $\gamma$  shape=0.27); jModelTest) and its reliability tested by 500 bootstrap replicates. Only bootstrap values exceeding 25 % are shown

(a)

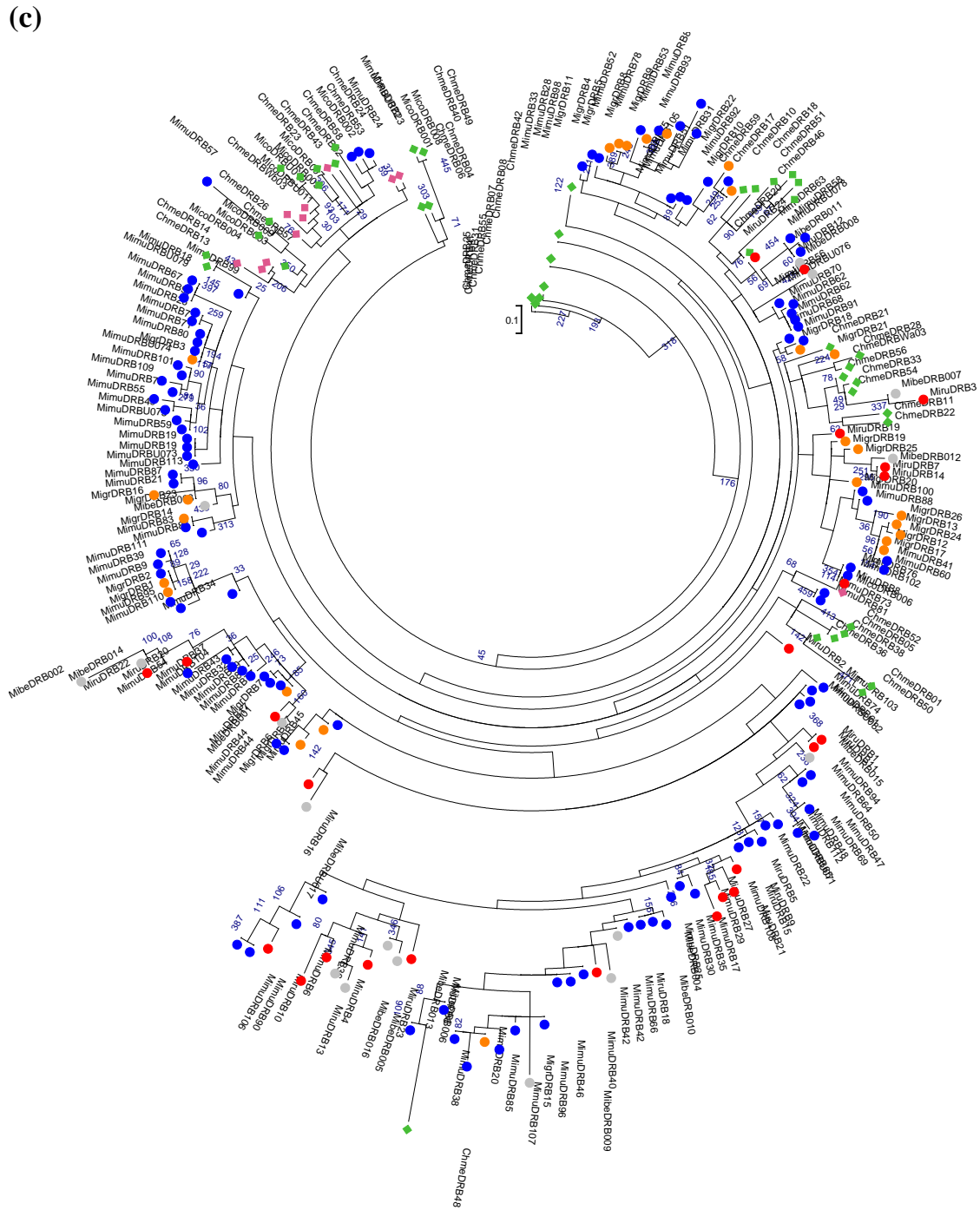




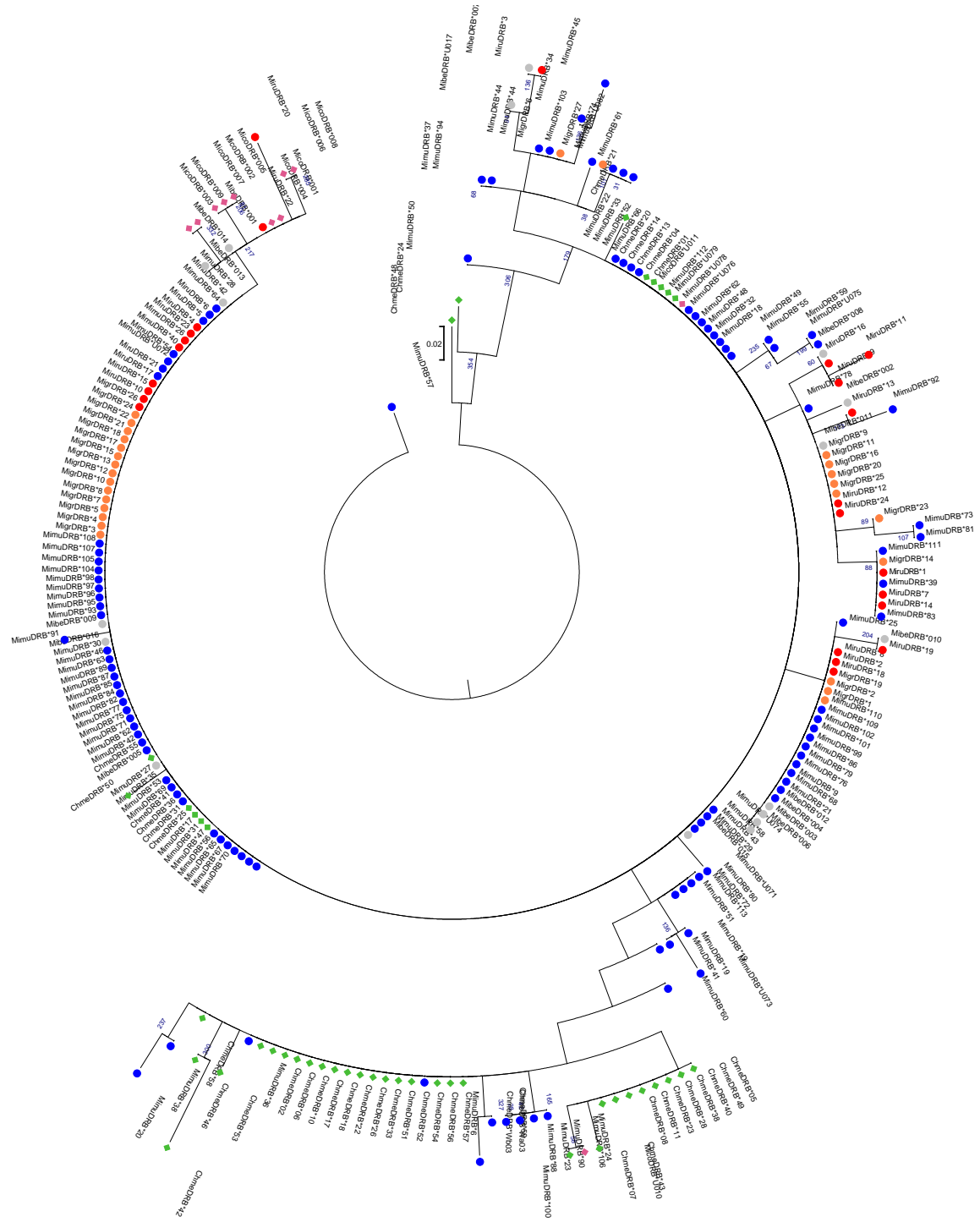
(b)



**Panel B.** Gene trees based on 20 DRB PSS (c) and 20 randomly selected non-PSS codons (d). The tree configuration was derived by the ML algorithm implemented in PhyML (c: GTR+G,  $\gamma$  shape=0.61; d: GTR+G,  $\gamma$  shape=0.37); jModelTest) and its reliability tested by 500 bootstrap replicates. Only bootstrap values exceeding 25 % are shown



(d)



## 10.2. Codon usage analysis

The outcome of codon usage analysis and simulations showing proportions of shared codons at PSS, scores indicative of co-ancestry, phylogenetic distances and median of the  $CE_d$ ,  $CA_d$  obtained for all pair-wise comparisons is summarized in Table S8.

**Table S8.** The overview of medians of the expected proportions of shared codons expected under the scenarios of convergent evolution ( $CE_d$ ) and common ancestry ( $CA_d$ ) obtained from a simulated distributions, proportions of identical codons at PSS, scores indicative of coancestry and phylogenetic distance between all pairs of species in DRB and DQB.

Species pair		Proportion of identical codons	median CE	median CA	Scores indicative of coancestry	Phylogenetic distance
<b>DRB</b>						
<i>M. murinus (distant)</i>	<i>M. murinus (Kirindy)</i>	0.964	0.919	0.986	0.683	<0.001
<i>M. rufus</i>	<i>M. berthae</i>	0.975	0.929	0.989	0.772	0.043
<i>M. griseorufus</i>	<i>M. murinus (Kirindy)</i>	0.969	0.918	0.986	0.754	0.167
<i>M. murinus (distant)</i>	<i>M. griseorufus</i>	0.967	0.906	0.991	0.717	0.167
<i>M. griseorufus</i>	<i>M. berthae</i>	0.975	0.922	0.994	0.737	0.199
<i>M. rufus</i>	<i>M. griseorufus</i>	0.983	0.918	0.992	0.882	0.202
<i>M. murinus (Kirindy)</i>	<i>M. berthae</i>	0.966	0.932	0.985	0.633	0.232
<i>M. murinus (distant)</i>	<i>M. berthae</i>	0.968	0.928	0.992	0.623	0.232
<i>M. rufus</i>	<i>M. murinus (Kirindy)</i>	0.972	0.925	0.984	0.792	0.235
<i>M. murinus (distant)</i>	<i>M. rufus</i>	0.97	0.922	0.989	0.721	0.235
<i>M. griseorufus</i>	<i>M. coquereli</i>	0.975	0.934	0.998	0.629	0.367
<i>M. murinus (Kirindy)</i>	<i>M. coquereli</i>	0.979	0.939	0.987	0.83	0.388
<i>M. murinus (distant)</i>	<i>M. coquereli</i>	0.978	0.936	0.993	0.738	0.388
<i>M. griseorufus</i>	<i>C. medius</i>	0.954	0.902	0.992	0.578	0.419
<i>M. berthae</i>	<i>M. coquereli</i>	0.966	0.936	0.991	0.537	0.432
<i>M. rufus</i>	<i>M. coquereli</i>	0.982	0.939	0.992	0.801	0.435
<i>C. medius</i>	<i>M. coquereli</i>	0.959	0.904	0.988	0.655	0.436
<i>M. murinus (Kirindy)</i>	<i>C. medius</i>	0.96	0.907	0.985	0.671	0.44
<i>M. murinus (distant)</i>	<i>C. medius</i>	0.959	0.903	0.99	0.651	0.44
<i>M. berthae</i>	<i>C. medius</i>	0.94	0.908	0.989	0.397	0.483
<i>M. rufus</i>	<i>C. medius</i>	0.947	0.906	0.989	0.495	0.486
<b>DQB</b>						
<i>M. murinus (Kirindy)</i>	<i>M. berthae</i>	0.962	0.777	0.98	0.909	0.232
<i>M. murinus (Kirindy)</i>	<i>M. coquereli</i>	0.943	0.664	0.972	0.907	0.388
<i>M. murinus (Kirindy)</i>	<i>C. medius</i>	0.917	0.762	0.977	0.719	0.44
<i>M. berthae</i>	<i>M. coquereli</i>	0.939	0.661	0.978	0.878	0.432
<i>M. berthae</i>	<i>C. medius</i>	0.907	0.761	0.982	0.662	0.483
<i>C. medius</i>	<i>M. coquereli</i>	0.942	0.681	0.969	0.906	0.436

## SUPPORTING INFORMATION 2

The overview of all sequences described or included in this study. The length of previously described sequences was optimized to match the fragment length investigated by the present study (169bp) and resulting nucleotide and /or amino acid identity was noted. Some of the newly described *Mimu* sequences submitted by this study were previously detected by Huchard et al. (2012), but not submitted due to their rare occurrence. All sequences that were newly described by this study translated into unique amino acids and contained no stop codons.

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
<b>DRB</b>					
<b><i>M. murinus</i></b>					
AJ431270.1	<i>Mimu-DRB*6</i>	Schad et al. 2004; Sommer et al. 2014			
AJ555838.1	<i>Mimu-DRB*9</i>	Schad et al. 2004; Sommer et al. 2014		<i>Mimu-DRB*003</i> <i>Migr-DRB*1</i>	<i>Mimu-DRB*003</i> <i>Migr-DRB*1</i>
EU137061	<i>Mimu-DRB*17</i>	Schwensow et al. 2008a			
EU137062	<i>Mimu-DRB*18</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137063	<i>Mimu-DRB*19</i>	Schwensow et al. 2008a; Huchard et al. 2012; Sommer et al. 2014	yes		
EU137064	<i>Mimu-DRB*20</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137065	<i>Mimu-DRB*21</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137066	<i>Mimu-DRB*22</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137067	<i>Mimu-DRB*23</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137068	<i>Mimu-DRB*24</i>	Schwensow et al. 2008a			
EU137069	<i>Mimu-DRB*25</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137070	<i>Mimu-DRB*26</i>	Schwensow et al. 2008a; Huchard et al. 2012		<i>Mimu-DRB*71</i> <i>Mimu-DRB*97</i> <i>Migr-DRB*3</i>	<i>Mimu-DRB*71</i> <i>Mimu-DRB*97</i> <i>Migr-DRB*3</i>
EU137071	<i>Mimu-DRB*27</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137072	<i>Mimu-DRB*28</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137073	<i>Mimu-DRB*29</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137074	<i>Mimu-DRB*30</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes	<i>Mimu-DRB*35</i>	<i>Mimu-DRB*35</i>
EU137075	<i>Mimu-DRB*31</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137076	<i>Mimu-DRB*32</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137077	<i>Mimu-DRB*33</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
EU137078	<i>Mimu-DRB*34</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137079	<i>Mimu-DRB*35</i>	Schwensow et al. 2008a	yes	<i>Mimu-DRB*30</i>	<i>Mimu-DRB*30</i>
EU137080	<i>Mimu-DRB*36</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137081	<i>Mimu-DRB*37</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137082	<i>Mimu-DRB*38</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137083	<i>Mimu-DRB*39</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137084	<i>Mimu-DRB*40</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137085	<i>Mimu-DRB*41</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137086	<i>Mimu-DRB*42</i>	Schwensow et al. 2008a; Huchard et al. 2012; Sommer et al. 2014	yes		
EU137087	<i>Mimu-DRB*43</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137088	<i>Mimu-DRB*44</i>	Schwensow et al. 2008a; Huchard et al. 2012; Sommer et al. 2014	yes		
EU137089	<i>Mimu-DRB*45</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137090	<i>Mimu-DRB*46</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137091	<i>Mimu-DRB*47</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137092	<i>Mimu-DRB*48</i>	Schwensow et al. 2008a; Huchard et al. 2012			
EU137093	<i>Mimu-DRB*49</i>	Schwensow et al. 2008a; Huchard et al. 2012			
EU137094	<i>Mimu-DRB*50</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137095	<i>Mimu-DRB*51</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137096	<i>Mimu-DRB*52</i>	Schwensow et al. 2008a; Huchard et al. 2012			
EU137097	<i>Mimu-DRB*53</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137098	<i>Mimu-DRB*54</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137099	<i>Mimu-DRB*55</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137100	<i>Mimu-DRB*56</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		
EU137101	<i>Mimu-DRB*57</i>	Schwensow et al. 2008a			
EU137102	<i>Mimu-DRB*58</i>	Schwensow et al. 2008a			
EU137103	<i>Mimu-DRB*59</i>	Schwensow et al. 2008a; Huchard et al. 2012	yes		<i>Mimu-DRB*U075</i>
HE801954	<i>Mimu-DRB*60</i>	Huchard et al. 2012			
HE801955	<i>Mimu-DRB*61</i>	Huchard et al. 2012	yes		
HE801956	<i>Mimu-DRB*62</i>	Huchard et al. 2012; Sommer et al. 2014	yes		
HE801957	<i>Mimu-DRB*63</i>	Huchard et al. 2012			
HE801958	<i>Mimu-DRB*64</i>	Huchard et al. 2012	yes		

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
HE801959	<i>Mimu-DRB*65</i>	Huchard et al. 2012	yes		
HE801960	<i>Mimu-DRB*66</i>	Huchard et al. 2012	yes		
HE801961	<i>Mimu-DRB*67</i>	Huchard et al. 2012	yes		
HE801962	<i>Mimu-DRB*68</i>	Huchard et al. 2012			
HE801963	<i>Mimu-DRB*69</i>	Huchard et al. 2012			
HE801964	<i>Mimu-DRB*70</i>	Huchard et al. 2012	yes		
	<i>Mimu-DRB*U071</i>	Huchard et al. 2012			
	<i>Mimu-DRB*U072</i>	Huchard et al. 2012			
	<i>Mimu-DRB*U073</i>	Huchard et al. 2012			
	<i>Mimu-DRB*U074</i>	Huchard et al. 2012			
	<i>Mimu-DRB*U075</i>	Huchard et al. 2012			<i>Mimu-DRB*59</i>
	<i>Mimu-DRB*U076</i>	Huchard et al. 2012			
	<i>Mimu-DRB*U077</i>	Huchard et al. 2012; submitted by this study <i>Mimu-DRB*112</i>	yes		
	<i>Mimu-DRB*U078</i>	Huchard et al. 2012			
	<i>Mimu-DRB*U079</i>	Huchard et al. 2012			
	<i>Mimu-DRB*U082</i>	Huchard et al. 2012		<i>Mimu-DRB*74</i>	<i>Mimu-DRB*74</i>
	<i>Mimu-DRB*U083</i>	Pechouskova et al.; submitted by this study as <i>Mimu-DRB*113</i>	yes		
<b><i>M. murinus</i></b>					
KF183544	<i>Mimu-DRB*71</i>	Sommer et al. 2014		<i>Migr-DRB*3</i> <i>Mimu-DRB*26</i> <i>Mimu-DRB*97</i>	<i>Migr-DRB*3</i> <i>Mimu-DRB*26</i> <i>Mimu-DRB*97</i>
KF183545	<i>Mimu-DRB*72</i>	Sommer et al. 2014		<i>Mimu-DRB*80</i>	<i>Mimu-DRB*80</i>
KF183546	<i>Mimu-DRB*73</i>	Sommer et al. 2014		<i>Mimu-DRB*81</i> <i>Mimu-DRB*2</i>	<i>Mimu-DRB*81</i> <i>Mimu-DRB*2</i>
KF183547	<i>Mimu-DRB*74</i>	Sommer et al. 2014			<i>Mimu-DRB*U082</i>
KF183548	<i>Mimu-DRB*75</i>	Sommer et al. 2014			
KF183549	<i>Mimu-DRB*76</i>	Sommer et al. 2014		<i>Mimu-DRB*102</i> <i>Miru-DRB*8</i>	<i>Mimu-DRB*102</i> <i>Miru-DRB*8</i>
KF183550	<i>Mimu-DRB*77</i>	Sommer et al. 2014		<i>Migr-DRB*7</i>	<i>Migr-DRB*7</i>
KF183551	<i>Mimu-DRB*78</i>	Sommer et al. 2014		<i>Migr-DRB*9</i>	<i>Migr-DRB*9</i>
KF183552	<i>Mimu-DRB*79</i>	Sommer et al. 2014			



Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
KF183553	<i>Mimu-DRB*80</i>	Sommer et al. 2014		<i>Mimu-DRB*72</i>	<i>Mimu-DRB*72</i>
KF183554	<i>Mimu-DRB*81</i>	Sommer et al. 2014		<i>Mimu-DRB*73</i>	<i>Mimu-DRB*73</i>
				<i>Mimu-DRB*2</i>	<i>Mimu-DRB*2</i>
KF183555	<i>Mimu-DRB*82</i>	Sommer et al. 2014			
KF183556	<i>Mimu-DRB*83</i>	Sommer et al. 2014		<i>Migr-DRB*14</i>	<i>Migr-DRB*14</i>
KF183557	<i>Mimu-DRB*84</i>	Sommer et al. 2014			
KF183558	<i>Mimu-DRB*85</i>	Sommer et al. 2014		<i>Migr-DRB*15</i>	<i>Migr-DRB*15</i>
KF183559	<i>Mimu-DRB*86</i>	Sommer et al. 2014			
KF183560	<i>Mimu-DRB*87</i>	Sommer et al. 2014			
KF183561	<i>Mimu-DRB*88</i>	Sommer et al. 2014			
KF183562	<i>Mimu-DRB*89</i>	Sommer et al. 2014			
KF183563	<i>Mimu-DRB*90</i>	Sommer et al. 2014			
KF183564	<i>Mimu-DRB*91</i>	Sommer et al. 2014			
KF183565	<i>Mimu-DRB*92</i>	Sommer et al. 2014			
KF183566	<i>Mimu-DRB*93</i>	Sommer et al. 2014			
KF183567	<i>Mimu-DRB*94</i>	Sommer et al. 2014			
KF183568	<i>Mimu-DRB*95</i>	Sommer et al. 2014			
KF183569	<i>Mimu-DRB*96</i>	Sommer et al. 2014			
KF183570	<i>Mimu-DRB*97</i>	Sommer et al. 2014		<i>Migr-DRB*3</i>	<i>Migr-DRB*3</i>
				<i>Mimu-DRB*26</i>	<i>Mimu-DRB*26</i>
				<i>Mimu-DRB*71</i>	<i>Mimu-DRB*71</i>
KF183571	<i>Mimu-DRB*98</i>	Sommer et al. 2014			
KF183572	<i>Mimu-DRB*99</i>	Sommer et al. 2014			
KF183573	<i>Mimu-DRB*100</i>	Sommer et al. 2014			
KF183574	<i>Mimu-DRB*101</i>	Sommer et al. 2014			
KF183575	<i>Mimu-DRB*102</i>	Sommer et al. 2014		<i>Mimu-DRB*76</i>	<i>Mimu-DRB*76</i>
				<i>Miru-DRB*8</i>	<i>Miru-DRB*8</i>
KF183576	<i>Mimu-DRB*103</i>	Sommer et al. 2014			
KF183577	<i>Mimu-DRB*104</i>	Sommer et al. 2014			
KF183578	<i>Mimu-DRB*105</i>	Sommer et al. 2014			
KF183579	<i>Mimu-DRB*106</i>	Sommer et al. 2014			

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
KF183580	<i>Mimu-DRB*107</i>	Sommer et al. 2014			
KF183581	<i>Mimu-DRB*108</i>	Sommer et al. 2014			
KF183582	<i>Mimu-DRB*109</i>	Sommer et al. 2014			
KF183583	<i>Mimu-DRB*110</i>	Sommer et al. 2014			
KF183584	<i>Mimu-DRB*111</i>	Sommer et al. 2014			
<b><i>M. rufus</i></b>					
KF183585	<i>Miru-DRB*1</i>	Sommer et al. 2014			
KF183586	<i>Miru-DRB*2</i>	Sommer et al. 2014			
KF183587	<i>Miru-DRB*3</i>	Sommer et al. 2014			
KF183588	<i>Miru-DRB*4</i>	Sommer et al. 2014			
KF183589	<i>Miru-DRB*5</i>	Sommer et al. 2014			
KF183590	<i>Miru-DRB*6</i>	Sommer et al. 2014			
KF183591	<i>Miru-DRB*7</i>	Sommer et al. 2014		<i>Miru-DRB*14</i>	<i>Miru-DRB*14</i>
KF183592	<i>Miru-DRB*8</i>	Sommer et al. 2014		<i>Mimu-DRB*76</i>	<i>Mimu-DRB*76</i>
				<i>Miru-DRB*102</i>	<i>Miru-DRB*102</i>
KF183593	<i>Miru-DRB*9</i>	Sommer et al. 2014			
KF183594	<i>Miru-DRB*10</i>	Sommer et al. 2014			
KF183595	<i>Miru-DRB*11</i>	Sommer et al. 2014			
KF183596	<i>Miru-DRB*12</i>	Sommer et al. 2014		<i>Mibe-DRB*011</i>	<i>Mibe-DRB*011</i>
KF183597	<i>Miru-DRB*13</i>	Sommer et al. 2014			
KF183598	<i>Miru-DRB*14</i>	Sommer et al. 2014		<i>Miru-DRB*7</i>	<i>Miru-DRB*7</i>
KF183599	<i>Miru-DRB*15</i>	Sommer et al. 2014			
KF183600	<i>Miru-DRB*16</i>	Sommer et al. 2014			
KF183601	<i>Miru-DRB*17</i>	Sommer et al. 2014			
KF183602	<i>Miru-DRB*18</i>	Sommer et al. 2014			
KF183603	<i>Miru-DRB*19</i>	Sommer et al. 2014			
KF183604	<i>Miru-DRB*20</i>	Sommer et al. 2014			
KF183605	<i>Miru-DRB*21</i>	Sommer et al. 2014			
KF183606	<i>Miru-DRB*22</i>	Sommer et al. 2014			
KF183607	<i>Miru-DRB*23</i>	Sommer et al. 2014			
KF183608	<i>Miru-DRB*24</i>	Sommer et al. 2014			

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
<b><i>M. griseorufus</i></b>					
KF183517	<i>Migr-DRB*1</i>	Sommer et al. 2014		<i>Mimu-DRB*9</i>	<i>Mimu-DRB*9</i>
KF183518	<i>Migr-DRB*2</i>	Sommer et al. 2014			
KF183519	<i>Migr-DRB*3</i>	Sommer et al. 2014		<i>Mimu-DRB*71</i> <i>Mimu-DRB*97</i> <i>Mimu-DRB*26</i>	<i>Mimu-DRB*71</i> <i>Mimu-DRB*97</i> <i>Mimu-DRB*26</i>
KF183520	<i>Migr-DRB*4</i>	Sommer et al. 2014		<i>Migr-DRB*5</i>	<i>Migr-DRB*5</i>
KF183521	<i>Migr-DRB*5</i>	Sommer et al. 2014		<i>Migr-DRB*4</i>	<i>Migr-DRB*4</i>
KF183522	<i>Migr-DRB*6</i>	Sommer et al. 2014			
KF183523	<i>Migr-DRB*7</i>	Sommer et al. 2014		<i>Mimu-DRB*77</i>	<i>Mimu-DRB*77</i>
KF183524	<i>Migr-DRB*8</i>	Sommer et al. 2014			
KF183525	<i>Migr-DRB*9</i>	Sommer et al. 2014		<i>Mimu-DRB*78</i>	<i>Mimu-DRB*78</i>
KF183526	<i>Migr-DRB*10</i>	Sommer et al. 2014			
KF183527	<i>Migr-DRB*11</i>	Sommer et al. 2014			
KF183528	<i>Migr-DRB*12</i>	Sommer et al. 2014		<i>Migr-DRB*17</i>	<i>Migr-DRB*17</i>
KF183529	<i>Migr-DRB*13</i>	Sommer et al. 2014			
KF183530	<i>Migr-DRB*14</i>	Sommer et al. 2014		<i>Mimu-DRB*83</i>	<i>Mimu-DRB*83</i>
KF183531	<i>Migr-DRB*15</i>	Sommer et al. 2014		<i>Mimu-DRB*85</i>	<i>Mimu-DRB*85</i>
KF183532	<i>Migr-DRB*16</i>	Sommer et al. 2014			
KF183533	<i>Migr-DRB*17</i>	Sommer et al. 2014		<i>Migr-DRB*12</i>	<i>Migr-DRB*12</i>
KF183534	<i>Migr-DRB*18</i>	Sommer et al. 2014			
KF183535	<i>Migr-DRB*19</i>	Sommer et al. 2014			
KF183536	<i>Migr-DRB*20</i>	Sommer et al. 2014			
KF183537	<i>Migr-DRB*21</i>	Sommer et al. 2014			
KF183538	<i>Migr-DRB*22</i>	Sommer et al. 2014			
KF183539	<i>Migr-DRB*23</i>	Sommer et al. 2014			
KF183540	<i>Migr-DRB*24</i>	Sommer et al. 2014			
KF183541	<i>Migr-DRB*25</i>	Sommer et al. 2014			
KF183542	<i>Migr-DRB*26</i>	Sommer et al. 2014			
KF183543	<i>Migr-DRB*27</i>	Sommer et al. 2014			

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
<b><i>C. medius</i></b>					
EF194225	<i>Chme-DRB*01</i>	Schwensow et al. 2007	yes		
EF194226	<i>Chme-DRB*02</i>	Schwensow et al. 2007	yes		
EF194227	<i>Chme-DRB*03</i>	Schwensow et al. 2007		<i>Chme-DQB*010</i>	<i>Chme-DQB*010</i>
EF194228	<i>Chme-DRB*04</i>	Schwensow et al. 2007	yes		
EF194229	<i>Chme-DRB*05</i>	Schwensow et al. 2007	yes		
EF194230	<i>Chme-DRB*06</i>	Schwensow et al. 2007	yes		
EF194231	<i>Chme-DRB*07</i>	Schwensow et al. 2007	yes		
EF194232	<i>Chme-DRB*08</i>	Schwensow et al. 2007			
EF194233	<i>Chme-DRB*10</i>	Schwensow et al. 2007	yes		
EF194234	<i>Chme-DRB*11</i>	Schwensow et al. 2007	yes		
EF194235	<i>Chme-DRB*13</i>	Schwensow et al. 2007	yes		
EF194236	<i>Chme-DRB*14</i>	Schwensow et al. 2007	yes		
EF194237	<i>Chme-DRB*15</i>	Schwensow et al. 2007		<i>Chme-DQB*016</i>	<i>Chme-DQB*016</i>
EF194238	<i>Chme-DRB*16</i>	Schwensow et al. 2007		<i>Chme-DQB*001</i>	<i>Chme-DQB*001</i>
EF194239	<i>Chme-DRB*17</i>	Schwensow et al. 2007	yes		
EF194240	<i>Chme-DRB*18</i>	Schwensow et al. 2007	yes		
EF194241	<i>Chme-DRB*19</i>	Schwensow et al. 2007		<i>Chme-DQB*017</i>	<i>Chme-DQB*017</i>
EF194242	<i>Chme-DRB*20</i>	Schwensow et al. 2007	yes		
EF194243	<i>Chme-DRB*21</i>	Schwensow et al. 2007			
EF194244	<i>Chme-DRB*22</i>	Schwensow et al. 2007	yes		
EF194245	<i>Chme-DRB*23</i>	Schwensow et al. 2007	yes		
EF194246	<i>Chme-DRB*24</i>	Schwensow et al. 2007		<i>Mimu-DRB*13</i>	<i>Mimu-DRB*13</i>
EF194247	<i>Chme-DRB*25</i>	Schwensow et al. 2007			
EF194248	<i>Chme-DRB*26</i>	Schwensow et al. 2007	yes		
EF194249	<i>Chme-DRB*27</i>	Schwensow et al. 2007		<i>Chme-DQB*009</i>	<i>Chme-DQB*009</i>
EF194250	<i>Chme-DRB*28</i>	Schwensow et al. 2007	yes		
EF194251	<i>Chme-DRB*29</i>	Schwensow et al. 2007			
EF194252	<i>Chme-DRB*30</i>	Schwensow et al. 2007		<i>Chme-DQB*014</i>	<i>Chme-DQB*014</i>
EF194253	<i>Chme-DRB*31</i>	Schwensow et al. 2007			
EF194254	<i>Chme-DRB*32</i>	Schwensow et al. 2007			

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
EF194255	<i>Chme-DRB*33</i>	Schwensow et al. 2007	yes		
EF194256	<i>Chme-DRB*34</i>	Schwensow et al. 2007			
EF194257	<i>Chme-DRB*35</i>	Schwensow et al. 2007		<i>Chme-DQB*018</i>	<i>Chme-DQB*018</i>
EF194258	<i>Chme-DRB*36</i>	Schwensow et al. 2007	yes		
EF194259	<i>Chme-DRB*37</i>	Schwensow et al. 2007		<i>Chme-DQB*003</i>	<i>Chme-DQB*003</i>
EF194260	<i>Chme-DRB*38</i>	Schwensow et al. 2007	yes		
EF194261	<i>Chme-DRB*39</i>	Schwensow et al. 2007		<i>Chme-DQB*012</i>	<i>Chme-DQB*012</i>
EF194262	<i>Chme-DRB*40</i>	Schwensow et al. 2007	yes		
EF194263	<i>Chme-DRB*41</i>	Schwensow et al. 2007	yes		
EF194264	<i>Chme-DRB*42</i>	Schwensow et al. 2007			
EF194265	<i>Chme-DRB*43</i>	Schwensow et al. 2007	yes		
EF194266	<i>Chme-DRB*44</i>	Schwensow et al. 2007		<i>Chme-DQB*023</i>	<i>Chme-DQB*023</i>
EF194267	<i>Chme-DRB*45</i>	Schwensow et al. 2007		<i>Chme-DQB*004</i>	<i>Chme-DQB*004</i>
EF194268	<i>Chme-DRB*46</i>	Schwensow et al. 2007			
EF194269	<i>Chme-DRB*47</i>	Schwensow et al. 2007		<i>Chme-DQB*007</i>	<i>Chme-DQB*007</i>
EF194270	<i>Chme-DRB*48</i>	Schwensow et al. 2007			
EF194271	<i>Chme-DRB*49</i>	Schwensow et al. 2007	yes		
EF194272	<i>Chme-DRB*50</i>	Schwensow et al. 2007	yes		
AB078306	<i>Chme-DRB*Wa03</i>	Go et al. 2002; Schwensow et al. 2007	yes		
AB078304	<i>Chme-DRB*Wb03</i>	Go et al. 2002; Schwensow et al. 2007	yes		
	<i>Chme-DRB*51</i>	this study	yes		
	<i>Chme-DRB*52</i>	this study	yes		
	<i>Chme-DRB*53</i>	this study	yes		
	<i>Chme-DRB*54</i>	this study	yes		
	<i>Chme-DRB*55</i>	this study	yes		
	<i>Chme-DRB*56</i>	this study	yes		
	<i>Chme-DRB*57</i>	this study	yes		
	<i>Chme-DRB*58</i>	this study	yes		
	<i>Chme-DRB*59</i>	this study	yes		

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
<b><i>M. coquereli</i></b>					
	<i>Mico-DRB*001</i>	this study	yes		
	<i>Mico-DRB*002</i>	this study	yes		
	<i>Mico-DRB*003</i>	this study	yes		
	<i>Mico-DRB*004</i>	this study	yes		
	<i>Mico-DRB*005</i>	this study	yes		
	<i>Mico-DRB*006</i>	this study	yes		
	<i>Mico-DRB*007</i>	this study	yes		
	<i>Mico-DRB*008</i>	this study	yes		
	<i>Mico-DRB*009</i>	this study	yes		
	<i>Mico-DRB*U010</i>	this study	yes		
	<i>Mico-DRB*U011</i>	this study	yes		
<b><i>DQB</i></b>					
<b><i>M. murinus</i></b>					
HQ222948.1	<i>Mimu-DQB*004</i>	Huchard et al. 2012	yes		
HQ222949.1	<i>Mimu-DQB*005</i>	Huchard et al. 2012	yes		
HQ222950.1	<i>Mimu-DQB*006</i>	Huchard et al. 2012	yes		
HQ222951.1	<i>Mimu-DQB*007</i>	Huchard et al. 2012	yes		
HQ222952.1	<i>Mimu-DQB*008</i>	Huchard et al. 2012	yes		
HQ222953.1	<i>Mimu-DQB*009</i>	Huchard et al. 2012	yes		
HQ222954.1	<i>Mimu-DQB*010</i>	Huchard et al. 2012	-		
HE801914	<i>Mimu-DQB*011</i>	Huchard et al. 2012	yes		
HE801915	<i>Mimu-DQB*012</i>	Huchard et al. 2012	yes		
HE801916	<i>Mimu-DQB*013</i>	Huchard et al. 2012	yes		
HE801917	<i>Mimu-DQB*014</i>	Huchard et al. 2012	yes		
HE801918	<i>Mimu-DQB*015</i>	Huchard et al. 2012	yes		
HE801919	<i>Mimu-DQB*016</i>	Huchard et al. 2012	yes		
HE801920	<i>Mimu-DQB*017</i>	Huchard et al. 2012	yes		
HE801921	<i>Mimu-DQB*018</i>	Huchard et al. 2012	yes		

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
HE801922	<i>Mimu-DQB*019</i>	Huchard et al. 2012	yes		
HE801923	<i>Mimu-DQB*020</i>	Huchard et al. 2012	yes		
HE801924	<i>Mimu-DQB*021</i>	Huchard et al. 2012	yes		
HE801925	<i>Mimu-DQB*022</i>	Huchard et al. 2012	yes		
HE801926	<i>Mimu-DQB*023</i>	Huchard et al. 2012	yes		
HE801927	<i>Mimu-DQB*024</i>	Huchard et al. 2012	yes		
HE801928	<i>Mimu-DQB*025</i>	Huchard et al. 2012	yes		
HE801929	<i>Mimu-DQB*026</i>	Huchard et al. 2012	yes		
HE801930	<i>Mimu-DQB*027</i>	Huchard et al. 2012	yes		
HE801931	<i>Mimu-DQB*028</i>	Huchard et al. 2012	yes		<i>Mimu-DQB*037</i>
HE801932	<i>Mimu-DQB*029</i>	Huchard et al. 2012	-		
HE801933	<i>Mimu-DQB*030</i>	Huchard et al. 2012	yes		
HE801934	<i>Mimu-DQB*031</i>	Huchard et al. 2012	yes		
HE801935	<i>Mimu-DQB*032</i>	Huchard et al. 2012	yes		
HE801936	<i>Mimu-DQB*033</i>	Huchard et al. 2012	yes		
HE801937	<i>Mimu-DQB*034</i>	Huchard et al. 2012	yes		
HE801938	<i>Mimu-DQB*035</i>	Huchard et al. 2012	yes		
HE801939	<i>Mimu-DQB*036</i>	Huchard et al. 2012	-		
HE801940	<i>Mimu-DQB*037</i>	Huchard et al. 2012	yes		<i>Mimu-DQB*028</i>
HE801941	<i>Mimu-DQB*038</i>	Huchard et al. 2012	yes		
HE801942	<i>Mimu-DQB*039</i>	Huchard et al. 2012	yes		
HE801943	<i>Mimu-DQB*040</i>	Huchard et al. 2012	yes		
HE801944	<i>Mimu-DQB*041</i>	Huchard et al. 2012	-		
HE801945	<i>Mimu-DQB*042</i>	Huchard et al. 2012	yes		
HE801946	<i>Mimu-DQB*043</i>	Huchard et al. 2012	yes		
HE801947	<i>Mimu-DQB*044</i>	Huchard et al. 2012	yes		
HE801948	<i>Mimu-DQB*045</i>	Huchard et al. 2012	yes		
HE801949	<i>Mimu-DQB*046</i>	Huchard et al. 2012	yes		
HE801950	<i>Mimu-DQB*047</i>	Huchard et al. 2012	-		
HE801951	<i>Mimu-DQB*048</i>	Huchard et al. 2012	-		
HE801952	<i>Mimu-DQB*049</i>	Huchard et al. 2012	yes		

Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
HE801953	<i>Mimu-DQB*050</i>	Huchard et al. 2012	-		
	<i>Mimu-DQB*U051</i>	Huchard et al. 2012; submitted by this study as <i>Mimu-DQB*051</i>	yes		
	<i>Mimu-DQB*U052</i>	Huchard et al. 2012	-		
	<i>Mimu-DQB*U053</i>	Huchard et al. 2012; submitted by this study as <i>Mimu-DQB*052</i>	yes		
	<i>Mimu-DQB*U054</i>	Huchard et al. 2012; submitted by this study as <i>Mimu-DQB*053</i>	yes		
	<i>Mimu-DQB*U055</i>	Huchard et al. 2012	-		
	<i>Mimu-DQB*U056</i>	Huchard et al. 2012; submitted by this study as <i>Mimu-DQB*054</i>	yes		
	<i>Mimu-DQB*U057</i>	Huchard et al. 2012	-		
	<i>Mimu-DQB*U058</i>	Huchard et al. 2012	-		<i>Mimu-DQB*U059</i>
	<i>Mimu-DQB*U059</i>	Huchard et al. 2012	-		<i>Mimu-DQB*U058</i>
	<i>Mimu-DQB*U061</i>	Huchard et al. 2012; submitted by this study as <i>Mimu-DQB*055</i>	yes		
	<i>Mimu-DQB*U063</i>	Huchard et al. 2012	-		
	<i>Mimu-DQB*056</i>	this study	yes		
	<i>Mimu-DQB*057</i>	this study	yes		
	<i>Mimu-DQB*U066</i>	this study	yes		
	<i>Mimu-DQB*U067</i>	this study	yes		
	<i>Mimu-DQB*U068</i>	this study	yes		
	<i>Mimu-DQB*U069</i>	this study	yes		
	<i>Mimu-DQB*U070</i>	this study	yes		
	<i>Mimu-DQB*U071</i>	this study	yes		
<b><i>C. medius</i></b>	<i>Chme-DQB*001</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*16</i>	<i>Chme-DRB*16</i>
	<i>Chme-DQB*002</i>	this study	yes		
	<i>Chme-DQB*003</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*37</i>	<i>Chme-DRB*37</i>
	<i>Chme-DQB*004</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*45</i>	<i>Chme-DRB*45</i>
	<i>Chme-DQB*005</i>	this study	yes		
	<i>Chme-DQB*006</i>	this study	yes		
	<i>Chme-DQB*007</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*47</i>	<i>Chme-DRB*47</i>
	<i>Chme-DQB*008</i>	this study	yes		
	<i>Chme-DQB*009</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*27</i>	<i>Chme-DRB*27</i>
	<i>Chme-DQB*010</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*03</i>	<i>Chme-DRB*03</i>



Accession nr.	DRB allele	Described by or detected by	This study	Co-identity nucleotides (169 bp)	Co-identity amino acid (54 positions)
	<i>Chme-DQB*011</i>	this study	yes		
	<i>Chme-DQB*012</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*39</i>	<i>Chme-DRB*39</i>
	<i>Chme-DQB*013</i>	this study	yes		
	<i>Chme-DQB*014</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*30</i>	<i>Chme-DRB*30</i>
	<i>Chme-DQB*015</i>	this study	yes		
	<i>Chme-DQB*016</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*15</i>	<i>Chme-DRB*15</i>
	<i>Chme-DQB*017</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*19</i>	<i>Chme-DRB*19</i>
	<i>Chme-DQB*018</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*35</i>	<i>Chme-DRB*35</i>
	<i>Chme-DQB*019</i>	this study	yes		
	<i>Chme-DQB*020</i>	this study	yes		
	<i>Chme-DQB*021</i>	this study	yes		
	<i>Chme-DQB*022</i>	this study	yes		
	<i>Chme-DQB*023</i>	this study; Schwensow et al. 2007	yes	<i>Chme-DRB*44</i>	<i>Chme-DRB*44</i>
	<i>Chme-DQB*U024</i>	this study	yes		
<b><i>M. coquereli</i></b>					
	<i>Mico-DQB*001</i>	this study	yes		
	<i>Mico-DQB*002</i>	this study	yes		
	<i>Mico-DQB*003</i>	this study	yes		
	<i>Mico-DQB*004</i>	this study	yes		
	<i>Mico-DQB*005</i>	this study	yes		
	<i>Mico-DQB*006</i>	this study	yes		
	<i>Mico-DQB*007</i>	this study	yes		
	<i>Mico-DQB*008</i>	this study	yes		
	<i>Mico-DQB*009</i>	this study	yes		
	<i>Mico-DQB*010</i>	this study	yes		
	<i>Mico-DQB*011</i>	this study	yes		
	<i>Mico-DQB*012</i>	this study	yes		
	<i>Mico-DQB*013</i>	this study	yes		
	<i>Mico-DQB*U014</i>	this study	yes		
	<i>Mico-DQB*U015</i>	this study	yes		
	<i>Mico-DQB*U016</i>	this study	yes		

## GENERAL DISCUSSION

In the first chapter of my thesis, I have reviewed technical challenges associated with the complexity of MHC genotyping that has, for a long time, impaired progress in MHC-based studies of free-ranging species. I focused on highlighting the potential of employing Next-Generation sequencing ('NGS') to address complex questions in evolutionary biology at the genomic and at the population scale.

In the two following chapters, I aimed to broaden our understanding of the evolution of MHC polymorphism at the community level.

First, I examined whether contrasting aspects of demography and ecology would affect patterns of selection at MHC. To do this, I examined the variation of two MHC class II genes (DRB, DQB) in two ecologically differentiated congeneric mouse lemur species (*M. murinus* and *M. berthae*, *Cheirogaleidae*) that were sampled over comparable spatial and temporal scales. These species show contrasting aspects of their ecology - such as dietary flexibility, social cohesion and population size and density - that has been previously suggested to affect the strength and diversity of selective pressures exerted by parasites on MHC.

Second, I attempted to elucidate which evolutionary mechanism – common ancestry or convergent evolution – is most likely to explain the origins of MHC similarity in a community of several confamiliar lemur species (*Cheirogaleidae*). First, I compared patterns of MHC variation across four sympatric mouse and dwarf lemurs (*M. berthae*, *M. murinus*, *Mirza coquereli* and *Cheirogaleus medius*) and evaluated the potential of overlapping selective parasite pressures among them, which would be required in the presence of convergent evolution, by gut parasite screening. I then investigated the links between functional proximity of MHC alleles and shared parasite pressures across different species to determine whether parallel selection from parasites could contribute to the distribution of functionally similar MHC alleles 'MHC supertypes', independently of host phylogeny.

Last, I integrated my new MHC data with those previously obtained for other *Cheirogaleidae* in order to investigate whether inter-specific MHC allelic similarity at

codon sites that are presumably affected by the operation of positive selection is likely to arise from the operation of convergent selection as a result of shared selective pressures, or from co-ancestry across different evolutionary time-scales.

Below, I first briefly summarize practical implications and future perspectives associated with the introduction of NGS technologies into MHC research before comparing patterns of MHC variation and molecular selection across four sympatric *Cheirogaleidae* and discussing potential effects of contrasting demography and parasite-mediated pressure on the levels of MHC variation of each host species. Finally, I evaluate whether shared selective pressures from gut-parasites could have independently affected the distribution of MHC supertypes across different host species, as assumed under the hypothesis of convergent evolution, or whether MHC similarity is more likely to arise from the maintenance of ancestral allelic lineages across species boundaries.

## NEXT-GENERATION SEQUENCING:

### NEW CHALLENGES AND PERSPECTIVES FOR MHC RESEARCH

Introducing NGS into MHC research presents a major potential to expand the scope of the longitudinal field-based studies aimed at addressing various evolutionary questions related to MHC in different systems and offers opportunities to conduct such studies at the genomic scale and at the population and community level (reviewed e.g. in Koboldt et al. 2013; Lighten et al. 2014a). Up to date, a large amount of MHC studies relied on sequencing a single locus. However, the simultaneous sequencing of a higher number of variable MHC loci of a possibly adaptive value is now increasingly feasible when using high-throughput sequencing and may increase the resolution of addressing some major evolutionary questions (e.g. Babik 2010; Siddle et al. 2010; Huchard et al. 2013).

NGS has numerous advantages compared to traditional MHC genotyping approaches, such as a relatively low cost per sample, low optimization effort and high throughput of data. However, they also are particularly prone to sequencing errors, particularly 454 amplicon sequencing used by this study. These sequencing errors may compromise the clarity of the final dataset and are manifested by relatively high frequencies of artefactual (vs. true) alleles, or also cross-amplicon contaminations. The

artefactual alleles may originate during PCR or sequencing procedures and may include homopolymer over- or under calls ('homopolymer: sequence region including repeats of the same base', Moore et al. 2006; Brockman et al. 2008), increased frequency of single point indels (insertion/deletion) or single-point mutations (Babik et al. 2009), or frequent occurrence of sequence chimeras (i.e. recombination of the two parental sequences, Babik 2010). Nonetheless, the consequences of such effects can be diminished by combining precautionary PCR preparation (reconditioning steps), including independent replicate reactions (using differently labeled sequencing primers), accurate primer design and sufficient depth of sequencing coverage (e.g. Lenz and Becker 2008; Babik 2010; Galan et al. 2010; Llaurens et.al 2012; Burri et. al. 2014). Although I applied multiple quality control steps that enabled dealing with a high rate of sequencing errors, manual allele sorting and evaluation of potential loci duplication still represents major time investment when dealing with species of previously unknown background MHC variability.

In contrast, using large-scale bioinformatic tools (see e.g. Lighten et al. 2014a for an in depth review) promises a certain level of automatization in that aspect. Nevertheless, automatization of sequencing data processing may also induce under-estimation of standing MHC variability, e.g. due to unequal amplification efficiency of alleles within an individual resulting in allelic dropout (van Oosterhout et al. 2006; Sommer et al. 2013). Or, alternatively, over-estimation by incorrect identification of sequencing artefacts that may return with high amount of sequencing reads and / or at high frequencies (Li and Stoneking 2012; Sommer et al. 2013; Lighten et al. 2014b), such as in the case of homopolymer errors in *M. coquereli* reported in this study. In each case, manual investigation of sequencing outcome, comparing genotypes among parent-offspring dyads to estimate error rates (e.g. in Huchard et al. 2012), as well as including independent replicates remain costly, but valuable tools to ensure high quality control and assessment of genotyping reliability. Moreover, ongoing development of bioinformatics tools aiming at screening sequencing artefact (e.g. Ferrandiz-Rovira et al. 2015; Fijarczyk and Babik 2015; Sebastian et al. 2015) and future developments in third-generation single molecule sequencing that relies on nanopore technology (reviewed e.g. in Lighten et al. 2014a) will hopefully also help to overcome problems associated with de novo multigene amplicon genotyping when using current second-generation sequencing (or 'NGS') technologies.

---

## MHC VARIATION IN SYMPATRIC LEMURS

### GENE ARCHITECTURE

The MHC configuration is not stable across, or sometimes also within, species on an evolutionary time scale. Besides being part of a complex multigene family, most MHC class I and II genes have undergone multiple independent gene duplication events (Dawkins et al. 1999; Kulski et al. 2002). Specifically, the number of loci may substantially vary across species, populations and individuals according to the birth-and-death scenario of gene duplications and deletions (Nei and Hughes 1992; Nei et al. 1997). This leads to so called ‘copy number variation’, when different number of alleles are present due to differential gene duplication (e.g. Eimes et al. 2011; Cheng et al. 2012; Winternitz et al. 2013; Lighten et al. 2014b). In concordance with some earlier findings, I did not find any support for a duplication of MHC DRB and DQB locus in *Cheirogaleidae* (*M. murinus*, *M. berthae* DRB & DQB: Schad et al. 2004; Averdam et al. 2011; Huchard et al. 2012; Sommer et al. 2014), which is in contrast to some earlier notifications of potential duplication in DRB (Schwensow et al. 2007, 2008a; Go et al. 2002). The lack of loci duplication in *Cheirogaleidae* is in contrast to MHC configuration described in some other primates, e.g. New- or Old- World monkeys, and apes (e.g. Trtkova et al. 1993, 1995; Slierendregt et al. 1994; Doxiadis et al. 2000, 2010; Otting et al. 2002; Bontrop 2006; Klein et al. 2007; de Groot et al. 2008).

In comparison, one of the most complex MHC constitutions with highly duplicated MHC class I and II genes and more than ten functional loci have been described in passerines. This extreme level of polymorphism has been attributed to arise by a combined effect of several factors, i.e. inter- or intra-locus recombination leading to gene conversion, which is frequently observed between and within species of birds or fish (e.g. Shum et al. 2001; Hess and Edwards 2002; Edwards et al. 1995a; Reusch and Langefors 2005; Spurgin et al. 2011; Lenz et al. 2013; Promerová et al. 2013; Sutton et al. 2013). Such processes may lead to the rapid formation of novel MHC alleles (Spurgin et al. 2011; Zhao et al. 2013), and may enhance copy number variation across genomic loci (Völker et al. 2010; van Oosterhout 2013). Specifically, gene conversion is thought to be largely responsible for species-specific clustering between non-homologous (paralogous)

loci observed frequently in birds compared to placental mammals where orthologous relationships among genes have often been retained (e.g. Edwards et al. 1995b).

Although little is known about the influence of these mechanisms on MHC variation in *Cheirogaleidae*, there is no indication of frequent recombination between non-homologous loci given that for all species there is clear locus-specific clustering (here DRB and DQB) (see also Trowsdale et al. 1995; Gu and Nei 1999). Both loci are located in close proximity on the same chromosome in *M. murinus* (Averdam et al. 2011) and in all species the two loci are in tight linkage disequilibrium suggesting physical linkage with no indication of recombination hotspot between them. Overall, a tight physical association between DRB and DQB is well conserved across mammals including primates (Bontrop et al. 1999; Kelley et al. 2005) and probably reflects their functional similarity.

The lack of loci duplication and clear locus separation further envisage a great potential to study the evolution of MHC or other fundamental MHC-related evolutionary questions in *Cheirogaleidae*, given that targeting individual loci, which often requires the use of degenerate (general) PCR primers for multilocus genotyping (Babik 2010), the complexity of correct loci assignment and the simultaneous amplification of multiple loci present major technical difficulties in MHC genotyping. These technical challenges often impair progress in large-scale studies in non-model organisms. The use of NGS technologies diminishes the need for many steps aimed at separating multiple co-amplified alleles from duplicated loci (e.g. Metzker et al. 2010). However, when multiple alleles per individual are sequenced, it remains necessary to increase sequencing depth within individual sample that lowers the final amount of individuals that can be sequenced within one sequencing run, as well as it imposes additional costs and workload.

### MHC CLASS II VARIATION AS A FUNCTION OF DEMOGRAPHY OR CONTRASTING SELECTION FROM PARASITES ?

The ability of balancing selection to maintain sufficient levels of MHC polymorphism may depend on the population size of host species (Hedrick 1972). In small fragmented populations, the overall loss of genetic diversity can be enhanced by genetic drift and inbreeding that may counter-act balancing selection and, consequently, compromise the

---

capability of such population to keep up with co-evolving parasite communities due to the progressive disappearance of rare allelic variants (Wright 1969; O'Brien and Evermann 1988; Hartl and Clark 1997; Hughes and Yeager 1998; Keller and Waller 2002; Frankham et al. 2002; Ejsmond and Radwan 2011).

To examine the potential effects of contrasting demography on patterns of MHC variation in *Cheirogaleidae*, I first compared MHC allelic richness (number of distinct alleles in a population) in the two highly polymorphic MHC loci in two congeneric sympatric species with contrasting population size and population densities – *M. murinus* and *M. berthae* (*Cheirogaleidae*); sampled over comparable spatial and temporal scale. While controlling for sampling effort (the overall number of sampled individuals), I found allelic richness to be ca. twofold lower in rare *M. berthae* than that of the more abundant and spatially clumped *M. murinus*. This finding matched our predictions of lower allelic richness in species that exhibits both lower population size and density. Moreover, the allelic frequency distribution over a period of nine years furthermore indicated the progressive disappearance of alleles in *M. berthae*, suggesting that reported allelic richness may be in fact over-estimated in this species. Similar patterns were obtained by including two other species (*C. medius* and *M. coquereli*) that complemented the gradient of contrasting population size and densities, ranging from the least abundant and rather solitary *M. berthae* and *M. coquereli* to relatively abundant, but spatially more dispersed pair-living *C. medius* and the most abundant and spatially clumped *M. murinus*. This further confirms that the level of allelic richness might be primarily affected by constraints imposed by demography and this effect appears most prominent in species facing demographic fluctuations, such as *M. coquereli* (Markolf et al. 2008) and *M. berthae* (see discussion in chapter 2).

Alternatively, small overall population size, lower rate of social cohesion and reduced spatial proximity, as well as a specialist diet in *M. berthae* could in theory promote relaxed parasite-mediated pressure due to exposure to a limited array of parasites and its transmission among conspecifics (e.g. Nunn et al. 2003; Altizer et al. 2007; Chen et al. 2008; Morand 2015), when compared to abundant and spatially clumped generalist *M. murinus*. In that case, I expected reduced parasite pressure to be reflected by diverging patterns of MHC allelic richness, but also of allelic divergence (measured as a nucleotide or amino acid distances between alleles of the same species). However, I found

that allelic divergence and the strength of selection detected at the molecular level were comparable in *M. berthae* and *M. murinus*, which did not support the assumption of relaxed pathogen-mediated pressure as the primary cause of contemporary lower allelic richness in this species.

In line with assumptions that high population density, spatial cohesion and generalist diet may increase the encountered spectrum of pathogens as well as their transmission probability (e.g. Nunn et al. 2003; Vitone et al. 2004; Chen et al. 2008; Hughes and Page 2007; reviewed in Morand 2015), I found the highest levels of parasitism in the most abundant and spatially cohesive generalist *M. murinus*. In contrast, the proportion of infected individuals and the overall diversity of parasites were lowest in the two rare lemurs – *M. berthae* and *M. coquereli*. However, sample sizes available for these two species were too low to deduce a potential impact of lower parasitism on patterns of observed MHC variation. Interestingly, these two species were most similar concerning the prevalence of shared helminth genera that could be affected by a considerable overlap in their diets and extensive home ranges (Dammhahn and Kappeler 2014; Schäfler 2011). A considerable overlap of shared parasites was also previously found between *C. medius* and *M. murinus* (Schwensow et al. 2010b) and confirmed by this study. Given that these species also show close proximity in their dietary ranges according to a recent stable isotope analysis (Dammhahn and Kappeler 2014), it seems plausible that dietary overlap may influence the array of shared gastrointestinal parasites in *Cheirogaleidae*. The possible effect of diet flexibility on the structure and diversity of parasite communities has been previously suggested, e.g. Vitone et al. 2004; Chen et al. 2008; Watts and Alexander 2012, but remains to be addressed in *Cheirogaleidae*.

Apart from contrasting patterns of allelic richness, I also observed divergent patterns of past positive selection at the sequence level across the four *Cheirogaleidae*. The density of sites assumed to be under positive selection (here positively selected sites, ‘PSS’), thought to represent one of the indications of selection acting on MHC alleles, was also contrasted between the two loci within each lemur species. Given the overall structural conservation of MHC class II molecules within mammals (Jones et al. 2006), the functional homology was assumed between empirically determined antigen-binding sites (‘ABS’) in humans (Human Leukocyte Antigen ‘HLA-ABS’; Brown et al. 1993) and ABS in the rest of mammalian and other orders. In lemurs, PSS were detected in close proximity to but not always at identical positions, than empirically determined human



ABS (Bondinas et al. 2007) by this and previous studies (e.g. Huchard et al. 2012; Schwensow et al. 2007, 2008a; Sommer et al. 2014). This incomplete overlap might occur due to the fact that adjacent amino acid residues, not directly involved in antigen-binding but possibly affecting the 3D-structural conformation of the peptide-binding groove, might be equally crucial in determining the range of bound peptides (Yang and Swanson 2002; Furlong and Yang 2008; Lenz 2011).

I observed higher density of PSS at DRB, except for *M. berthae* where DQB appeared to be under stronger selection compared to the three other species. Allelic richness was generally higher in DQB, except for *C. medius* where I observed lower allelic richness in DQB accompanied by only 1 PSS. In spite of the high level of allelic similarity and physical linkage between the two investigated MHC loci, suggesting that selection might not affect each locus entirely independently, these contrasting patterns could indicate a potential differentiation in the function of these loci, as has already been suggested for *M. murinus*. Here DRB was suggested to be under stronger diversifying selection and of higher importance than DQB in MHC-based mate choice (Huchard et al. 2012; 2013). While differences in allelic richness are most likely affected by demographic fluctuations, it seems plausible that contrasting patterns of selection at the two loci might come from functional differences. Nevertheless, the rates of non-synonymous substitutions primarily reflect past selection, given the long time needed for these substitutions to accumulate, but also to disappear (Bryja et al. 2007; Garrigan and Hedrick 2003). This means that rates of non-synonymous mutations may remain elevated long after selection has ceased. Thus, further studies directly addressing potential functional differences of different MHC loci by investigating contemporary selective pressures that may be affecting contrasting PSS distribution would be desirable.

Overall, it is likely that demography and progressive disappearance of rare alleles over time are largely responsible for divergent patterns of allelic richness in these lemur species. Although the high allelic divergence might buffer potential detrimental effects of decreasing allelic richness in *M. berthae*, it might not provide enough flexibility on longer time scale to face co-evolving or novel parasites. Sampling ca 60 (out of 100) individuals of *M. berthae* allowed capturing the entire allelic pool available across 9 years and four study areas, whereas it was necessary to sample more than 200 individuals (out of > 600) to catch the full allelic pool in *M. murinus*. The progressive disappearance of rare alleles

over time raises concerns regarding the viability of this small and isolated population of highly endemic and endangered species which may not be able to face ever-changing pathogen pressure on a long-term. The increasing risk of inbreeding and susceptibility to a demographic or genetic bottleneck has been detected in other lemur species, including *M. coquereli*, that are facing increasing environmental pressure all over Madagascar (Louis et al. 2005; Olivieri et al. 2008; Markolf et al. 2008; Radespiel et al. 2008; Craul et al. 2009; Razakamaharavo et al. 2010; Holmes et al. 2013; Baden et al. 2014).

## MHC EVOLUTIONARY PATHWAYS

### SHARED SELECTIVE PRESSURE OR CO\_ANCESTRY SHAPING MHC VARIATION ACROSS LEMUR SPECIES ?

In the final steps, I examined patterns of functional and neutral MHC similarity across sympatric and allopatric lemur species of the family *Cheirogaleidae* in an attempt to determine which evolutionary mechanism, common ancestry or convergent evolution, was most likely to shape MHC variation beyond genus level.

I first evaluated the extent of functional overlap among MHC alleles in different species and classified them based on their antigen-binding specificities into functionally similar super-groups, so called ‘supertypes’. I found extensive supertype overlap across the four sympatric *Cheirogaleidae* with all but two supertypes in each locus found in at least two and up to all four species. Using supertype classification enabled me to further investigate the interactions between parasite infestations and host immune response by examining (1) the functional similarity of encoded MHC proteins and (2) patterns of codon usage at the functionally important sites (here PSS) and allelic similarity at sites that are not predicted to be involved in antigen-binding (here non-PSS).

In the first step, I tested whether shared selective pressures could play a role in shaping allelic similarity, i.e. functional convergence, at coding sites across different species. Despite lacking significance with only four species included in the analysis, I found an indication of a strong association between the frequency distribution of MHC supertypes and helminth prevalence across species. This means that host species that are infected by the same parasite species with similar prevalence tend to harbor same MHC supertypes with similar frequencies. My results were corresponding to those previously

obtained by another study focusing on a community of 11 rodent species, which reported that more than half of the variability in MHC diversity present within this community might be shaped by host–parasite interactions, with reciprocal effects of the changes in MHC diversity on host–parasite interactions (Pilosof et al. 2014).

Next, I examined whether there is a link between the possession of particular supertypes and particular helminth infestations that would be detectable across individuals in different species (and not just within one species). Associations between particular DRB alleles and helminth infestations (*Ascaris* sp. and *Hymenolepis* sp.) have also been previously indicated by other studies in *Microcebus* sp. and in *C. medius* (Schad et al. 2005; Schwensow et al. 2007, 2010a, b; Sommer et al. 2014), as well as in other bird or mammal species (e.g. Westerdahl et al. 2005, 2013; Bonneaud et al. 2006; Loiseau et al. 2008; de Belloq et al. 2008; Froeschke and Sommer 2012; Zhang and He 2013; Pilosof et al. 2014). I found significant associations between the possession of particular supertypes and infestation by a common helminth, such as *Hymenolepis* sp. (*Cestoda*) and *Subulura* sp. (*Nematoda*), but these were mostly detectable within single species. However, the association between the possession of one particular DRB supertype (labelled SR4) and *Hymenolepis* sp. and another DQB supertype (labelled SQ4) and overall parasite richness was detected in both species of *Microcebus*. The functional overlap of alleles associated with *Ascaris* sp. infestation, was already reported between more distantly related *M. murinus* and *C. medius* (Schwensow et al. 2010b). However, it remained unclear whether this functional similarity arose as a result of convergent evolution due to shared parasite pressure, or from a common descent.

Overall, my results supported some of the previous associations found between the possession of particular alleles (or supertypes) and particular helminth infestations or higher parasite burden obtained by a previous study conducted in Kirindy forest (Schwensow et al. 2010a, b). These findings indicate that these associations do not occur randomly and that also other associations arising through adaptations to similar selective pressures in different host species might be present within our dataset, but limited sample sizes precludes more detailed analyses. Interestingly, *Ascaris* sp. infestations that are thought to impose a significant burden on host fitness (Crompton and Nesheim 2002) and were previously linked to functional overlap in the two more distantly related *Cheirogaleidae*, were almost absent at our study site (Hämäläinen et al. 2015). Given that

the occurrence of this helminth with direct transmission route seems to be related to the level of precipitation (Sommer et al. 2014), its absence in our dataset could be affected by timing of our sampling (dry season), since a previous study in Kirindy conducted during the rainy season indicated *Ascaris* sp. as the most prevalent helminth in both *C. medius* and *M. murinus* (Schwensow et al. 2010b). However, the seasonal effect on *Ascaris* sp. prevalence has not been confirmed recently (Hämäläinen et al. 2015).

More specifically, associations of particular MHC alleles (or supertypes) with multiple parasites detected here and also in previous studies (e.g. Wegner et al. 2003; Tollenaere et al. 2008; Oliver et al. 2009; Froeschke and Sommer 2012; Pilosof et al. 2014; Tobler et al. 2014) may indicate the existence of generalist alleles (or supertypes) with broad-spectrum antigen recognition. These alleles might be consequently under purifying selection, and may not be affected by the drastic drops in allelic frequency anticipated under the scenario of negative-frequency dependent selection, which assumes that the frequency of alleles dynamically interacts with co-evolving parasite pressure. This may in turn facilitate the persistence of such alleles in host populations over long evolutionary time scales. Moreover, rather limited capabilities of antigen recognition assumed to occur in the mammalian hosts in contrast to the high diversity of protein-encoding genes assumed to be possessed by intestinal helminths, suggest that mammalian hosts might dispose of limited capabilities to distinguish between particular helminth species (e.g. Pearce and Tarlton 2002; Finkelman et al. 2004). The helminth genera detected in our study species also do not seem to be specialized on certain hosts, but exhibit low host specificity, as indicated by the fact that the majority of infections identified across all four species were represented by 3-4 common genera in this and other studies (*M. murinus*: Hämäläinen et al. 2015; Sommer et al. 2014). It has been pointed out that lack of host specificity, or also mild pathogenicity, and functional overlap of MHC alleles across different host species may disrupt the parasite-host co-evolutionary cycles that are assumed to occur under frequency-dependent selection (e.g. Snell 1968; Spurgin and Richardson 2010). Rare alleles may provide better resistance due to insufficient time for co-evolving parasites to evolve evasive mechanisms, that would be required for co-evolution of particular MHC alleles in response to particular pathogens (Penn 2001; Tobler et al. 2014; Pilosof et al. 2014). Thus, lack of host specificity may in fact promote selection for more generalistic immune responses (Tobler et al. 2014). This

mechanism may be more complicated when considering a community of hosts instead of single host populations.

Moreover, little is known about interactions between multiple helminth species within a host, about the pathogenicity of given helminths, as well as about potential antagonistic (opposing) interactions of particular MHC alleles in response to infestations. These allelic effects can be both in the sense of contrasting susceptibility of individuals possessing particular allele to the infection of one particular parasite species, and contrasting susceptibility of individual(s) possessing particular allele to the infection imposed by different parasites. Such contrasting effects, i.e. susceptibility to one, but protection against another (perhaps more crucial) parasite species, of particular MHC allele, may remain often unobserved and may contribute to their maintenance in a population and explain frequency shifts within the population. For instance, such antagonistic effects of alleles associated with blood-borne diseases or helminth infestations were detected in passerines and rodents (Apanius et al. 1997; Loiseau et al. 2008, 2011; Froeschke and Sommer 2012). The maintenance of disadvantageous alleles associated with increased risk of infection which is commonly observed across taxa and also in our study (e.g. Hendel et al. 1999; Segal and Hill 2003; Harf and Sommer 2005; Meyer-Lucht and Sommer 2005; Froeschke and Sommer 2005; Schad et al. 2005; Bonneaud et al. 2006; Loiseau et al. 2008; Schwensow et al. 2010b; Froeschke and Sommer 2012; Westerdahl et al. 2013; Kamath et al. 2014) could therefore be explained by pleiotropic effects of given allele(s) or MHC-affected competition between two parasites species (or lineages) (see e.g. Zhang et al. 2002; Loiseau et al. 2008).

Finally, I investigated patterns of codon usage at the functionally important sites thought to be under positive selection (PSS) that are assumed to be involved in antigen binding; and allelic similarity at the sites that are not (non-PSS). Here I aim to disentangle, whether allelic similarity at PSS, manifested by the supertype overlap across species, stems largely from the operation of convergent selection or from the maintenance of similar alleles or allelic motifs over time (allelic co-ancestry). The overlap of binding affinities at PBR (here represented by PSS) among species resulting in alleles from different species belonging to the same supertype, or a similar supertype frequency distribution across species that are exposed to similar pathogenic pressures, would be indicated to arise from independent convergence provided that allelic similarity is limited,

or absent from non-PBR (here non-PSS) regions. Alternatively, if shared polymorphism at PBR is a consequence of common descent, the composition of codons coding for identical amino acids at PBR should show high structural identity (Lundberg and McDevitt 1992; Lenz et al. 2013) and the signal of co-ancestry may be present also at sites not involved in antigen binding.

I found higher structural (nucleotide) similarity of alleles from different species that belonged to the same supertype than of alleles that belonged to different superotypes. I also found, inter-specific clustering at PSS and the lack of species-specific clustering at non-PSS. All these findings indicate that similarity of alleles at PSS stems largely from their common descent, rather than convergence. Moreover, the presence of identical indels, such as 6bp insertion among some DQB sequences of *M. murinus* and *M. berthae*, are highly improbable to occur independently after the speciation event (Klein et al. 2007). The high structural similarity of these alleles that resulted in the formation of a monophyletic allelic cluster suggests their similarity by co-ancestry, rather than independent convergence.

I gained more support for this conclusion by investigating patterns of codon usage at PSS which showed that the inter-specific sharing of amino acids was accompanied by a higher proportion of codon identity than expected by chance. In addition, the proportion of codon identity shared across species was a function of the phylogenetic proximity of species, suggesting that allelic similarity is progressively lost as species diverge. In contrast, I detected no relationship between shared parasite pressures and proportion of shared amino acids at PSS. Additionally, associations detected between the possession of particular superotypes and helminth infestations between the two most closely related species, *M. murinus* and *M. berthae*, which diverged less than 9 Mya (Thiele et al. 2013), would rather suggest that this functional overlap arose from the long term maintenance of allelic motifs over time.

Up to date, only small amount of studies have conducted such site-specific analyses that might have obscured the fact that similarity at PSS- (or PBR-) encoding regions across species might arise more often from co-ancestry (Lenz et al. 2013) also in more distantly related species, than has been acknowledged and supported by quantitative data in the literature.

In addition, it would be desirable to investigate also non-coding regions of the MHC loci (e.g. introns), since that would provide a better resolution for phylogenetic

inference across species, first because these regions are often longer than exons, and second because they may be under lower selective pressure (Zhang and Kumar 1997; Kriener et al. 2000; Lenz et al. 2013).

So far, MHC allelic similarity has been majorly attributed to independent convergence among taxa that split more than 30 Mya (e.g. Andersson et al. 1991; Trtkova et al. 1995; Yeager et al. 1997; Kriener et al. 2001; Srithayakumar et al. 2012). Even though the split between *Cheirogaleus medius* and other *Cheirogaleidae* species investigated by this study is larger and close to 30 Mya (19-26 Mya), in contrast to divergence times estimated within *Microcebus* sp. (7-11 Mya) and between *M. coquereli* and *Microcebus* sp (12-18 Mya) (Thiele et al. 2013), I did not find support for the fact that allelic similarity between *Cheirogaleus* and other sympatric *Cheirogaleidae* would stem from functional convergence in response to shared parasite pressures, as previously suggested.

In addition, a previous study on Malagasy lemurs has reported the sharing of identical DRB alleles of partial exon 2 among lemur species within genus, as well as across families that diverged up to 40 Mya (Go et al. 2002) and attributed this finding to trans-species polymorphism ('sharing of identical alleles by common descent'). In contrast, neither I nor other studies focusing on lemur MHC have confirmed the occurrence of identical alleles at DRB loci among closely related or more distantly related lemurs, except for instances with ongoing hybridization (Sommer et al. 2014). The majority of the samples used in the study of Go et al. (2002) originated from a captive zoo colony where the effect of hybridization between closely related species cannot be fully excluded. In addition, the potential contribution of other mechanisms, such as allele introgression or convergence, that could contribute to this allele sharing were not investigated.

Overall, given the extreme polymorphism of MHC genes within and across species and given their importance in immune response, most studies have focused on the occurrence of trans-species polymorphism (often referred to as 'sharing of identical alleles or allelic motifs across species boundaries') of MHC class I and class II genes (reviewed in Tesicky and Vinkler 2015). However, MHC genes are not the only marker where shared polymorphism across species has been described (reviewed in Azevedo et al. 2015). For instance, putative shared polymorphism between great apes and humans

have been identified also for histo-blood ABO group, two antiviral genes (ZC3HAV1 and TRIM5), an autoimmunity-related gene (LAD1) and several non-coding genomic segments with a putative regulatory role. These instances involve genes with key roles in the immune system and that in a broader sense appear to confer genetic resistance to pathogens. Although the selective mechanisms are still largely unknown, trans-species polymorphism appears to be a more often re-curring event than anticipated in the evolution of the primate immune system. The identification of such polymorphisms and of their underlying mechanism ('identity-by descent' or 'identity by state') shall therefore improve our understanding of the evolution of host-immune response, with implications for preventive medicine. The identification of functionally important variants present across species boundaries and presumably maintained by balancing selection in human and animal genomes may represent a promising new approach in evolutionary medicine to recognize naturally occurring disease resistance or susceptibility alleles with potential applications in human medicine, animal breeding or conservation. For instance, trans-specifically shared alleles may be linked to deleterious (or beneficial) mutations accumulated in the neighbourhood of the sites maintained by the balancing selection (van Oosterhout 2009) and its identification may improve personalized approach in future medicine (reviewed in Azevedo et al. 2015; Tesicky and Vinkler 2015).



## CONCLUSION

In my thesis, I investigated patterns of MHC class II variation of two highly polymorphic loci in four sympatric lemurs (*Cheirogaleidae*). I found that demographic factors may exert a stronger influence than pathogen-driven selection at current patterns of MHC allelic diversity at the population level and this effect is most apparent in species facing contemporary density fluctuations. I detected considerable structural as well as functional allelic similarity across the four *Cheirogaleidae*. I detected partial overlap in helminth communities across the four species that was more prominent in species with similar ecological traits, e.g. dietary overlap. This overlapping parasite pressures seem to affect to a certain extent patterns of susceptibility to helminth infestations in at least two different host species based on the detected associations between host MHC and helminth infestation, as well as it seems to affect the frequency distribution of functionally similar MHC alleles across species. By investigating patterns of allelic similarity at sites thought to be involved in antigen binding and at sites that are not, I showed that allelic similarity largely resulted from the long term maintenance of ancestral sequence motifs, or allelic lineages, across multiple speciation events, sometimes for as long as 20 - 30 million years ago.

My results contribute to current voices aiming at expanding view on MHC evolution from the one-to-one host-parasite framework to multiple hosts vs. multiple parasites complex of community level interactions. Interactions between multiple infestations as well as between parasites and multiple hosts and their immune response present a relatively novel and challenging niche for future research that would allow us to obtain a more comprehensive picture that can lead us closer to understanding proximate processes underlying the dynamics of disease susceptibility as well as the diversification of populations in response to changing pathogenic pressures. This new perspective may also result in the need to revisit some of the fundamental mechanisms that have been proposed to explain the exceptional levels of MHC polymorphism.

Furthermore, understanding the mechanisms driving adaptations to ever-changing selective pressures and those maintaining standing genetic variability in critically important immune genes is crucial for further assessments of conservation impact on species with increased ecological vulnerability. The loss of genetic variability may show

## CONCLUSION

---

no immediate consequences, perhaps because it can be compensated to a certain level by other mechanisms, such as persistence of highly divergent alleles that may enable recognition of higher diversity of antigens within an individual host, or specific alleles that confer resistance against crucial pathogen (e.g. Wakeland 1990; Apanius et al. 1997; Lenz 2011; Spurgin and Richardson 2010). However, it is likely to have deleterious consequences on the long term, and may limit population capacity to face moving parasite pressures. Thus, the monitoring of the MHC diversity of small and fragmented populations which now becomes more feasible with the introduction of high-throughput sequencing technologies shall become a standard procedure accompanying conservation plans. Finally, the long-term maintenance of functionally similar alleles, or allelic motifs, shows that MHC diversity has accumulated slowly, through interactions between hosts and parasites over long evolutionary time scales, and stresses the importance of protecting current genetic variability which might take another million years to recover, if not lost irreversibly.

## REFERENCES

- Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.
- Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology and Evolution*, **21**, 433–438.
- Aguilar A, Roemer G, Debenham S *et al.* (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 3490–3494.
- Aksenov AA, Gojova A, Zhao W *et al.* (2012) Characterization of volatile organic compounds in human leukocyte antigen heterologous expression systems: A cell’s “chemical odor fingerprint.” *ChemBioChem*, **13**, 1053–1059.
- Alberts B (2002) Isolating, cloning and sequencing DNA. In: Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P (Eds.), *Molecular Biology of the cell* (4<sup>th</sup> ed.). New York: Garland Science.
- Alberts SC, Ober C (1993) Genetic variability in the major histocompatibility complex: a review of non-pathogen-mediated selective mechanisms. *Yearbook of Physical Anthropology*, **36**, 71–89.
- Albon SD, Stien a., Irvine RJ *et al.* (2002) The role of parasites in the dynamics of a reindeer population. *Proceedings of the Royal Society of London B: Biological Sciences*, **269**, 1625–1632.
- Alcaide M, Edwards S V., Negro JJ (2007) Characterization, Polymorphism, and Evolution of MHC Class II B Genes in Birds of Prey. *Journal of Molecular Evolution*, **65**, 541–554.
- Alcaide M, Liu M, Edwards S V. (2013) Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ*, **1**, e86.
- Allen JE, Maizels RM (2011) Diversity and dialogue in immunity to helminths. *Nature Reviews Immunology*, **11**, 375–388.
- Altizer S, Nunn CL, Lindenfors P (2007) Do threatened hosts have fewer parasites? A comparative study in primates. *The Journal of Animal Ecology*, **76**, 304–314.

## REFERENCES

---

- Alvergne A, Huchard E, Caillaud D *et al.* (2009) Human ability to recognize kin visually within primates. *International Journal of Primatology*, **30**, 199–210.
- Anderson RM, May RM (1978) Regulation and stability of host-parasite population interactions. *Journal of Animal Ecology*, **47**, 219–247.
- Andersson L, Sigurdardottir S, Borsch C, Gustafsson K (1991) Evolution of MHC polymorphism - extensive sharing of polymorphic sequence motifs between human and bovine DRB alleles. *Immunogenetics*, **33**, 188–193.
- Andersson L, Mikko S (1995) Generation of MHC class II diversity by intra- and intergenic recombination. *Immunological Reviews*, **143**, 5–12.
- Andriaholinirina N, Baden A, Blanco M *et al.* (2014) *Microcebus berthae*. The IUCN Red List of Threatened Species. Version 2014.3. <[www.iucnredlist.org](http://www.iucnredlist.org)>.
- Apanius V, Penn D, Slev P, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Critical Review of Immunology*, **17**, 179–224.
- Arden B, Klein J (1982) Biochemical comparison of major histocompatibility complex molecules from different subspecies of *Mus musculus*: evidence for trans-specific evolution of alleles. *Proceedings of the National Academy of Sciences of the United States of America*, **79**, 2342–2346.
- Averdam A, Kuschal C, Otto N *et al.* (2011) Sequence analysis of the grey mouse lemur (*Microcebus murinus*) MHC class II DQ and DR region. *Immunogenetics*, **63**, 85–93.
- Axtner J, Sommer S (2012) The functional importance of sequence versus expression variability of MHC alleles in parasite resistance. *Genetica*, **140**, 407–420.
- Azevedo L, Serrano C, Amorim A, Cooper DN (2015) Trans-species polymorphism in humans and the great apes is generally maintained by balancing selection that modulates the host immune response. *Human Genomics*, **9**, 21.
- Babik W (2010) Methods for MHC genotyping in non-model vertebrates. *Molecular Ecology Resources*, **10**, 237–251.
- Babik W, Durka W, Radwan J (2005) Sequence diversity of the MHC DRB gene in the Eurasian beaver (*Castor fiber*). *Molecular ecology*, **14**, 4249–4257.
- Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular ecology resources*, **9**, 713–719.

- Baden AL, Holmes SM, Johnson SE *et al.* (2014) Species-level view of population structure and gene flow for a critically endangered primate (*Varecia variegata*). *Ecology and Evolution*, **4**, 2675–2692.
- Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology*, **13**, 3021–3031.
- Bates D, Maechler M, Bolker B, Walker S (2015) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9, <https://CRAN.R-project.org/package=lme4>.
- Bell MA, Foster SA (1994) The evolutionary biology of the threespine stickleback. Oxford Univ. Press, Oxford UK.
- 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.
- Bjorkman PJ, Burmeister WP (1994) Structures of two classes of MHC molecules elucidated: crucial differences and similarities. *Current Opinion in Structural Biology*, **4**, 852–856.
- Blais J, Rico C, van Oosterhout C *et al.* (2007) MHC adaptive divergence between closely related and sympatric African cichlids. *PloS one*, **8**, 1-12.
- Bodmer WF (1972) Evolutionary Significance of the HL-A System. *Nature*, **237**, 139–145.
- Bondinas GP, Moustakas AK, Papadopoulos GK (2007) The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. *Immunogenetics*, **59**, 539–553.
- Bonneaud C, Pérez-Tris J, Federici P, Chastel O, Sorci G (2006) Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution*, **60**, 383–389.
- Bontrop RE, Otting N, de Groot NG, Doxiadis GGM (1999) Major histocompatibility complex class II polymorphisms in primates. *Immunological Reviews*, **167**, 339–350.
- Borghans JA, Beltman JB, De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics*, **55**, 732–739.
- Brown JH, Jardetzky TS, Gorga JC *et al.* (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33–39.

## REFERENCES

---

- Brown JL and Eklund A (1994) Kin recognition and the major histocompatibility complex – an integrative review. *American Naturalist*, **143**, 435–461.
- Bryja J, Galan M, Charbonnel N, Cosson JF (2006) Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (*Arvicolinae*). *Immunogenetics*, **58**, 191–202.
- Bryja J, Charbonnel N, Berthier K, Galan M, Cosson JF (2007) Density-related changes in selection pattern for major histocompatibility complex genes in fluctuating populations of voles. *Molecular ecology*, **16**, 5084–5097.
- Buckling A, Rainey PB (2002) The role of parasites in sympatric and allopatric host diversification. *Nature*, **420**, 496–499.
- Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L (2008) Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Molecular Biology and Evolution*, **25**, 1180–1191.
- Burri R, Promerová M, Goebel J, Fumagalli L (2014) PCR-based isolation of multigene families: lessons from the avian MHC class IIB. *Molecular Ecology Resources*, **14**, 778–788.
- Castro-Prieto A, Wachter B, Sommer S (2011) Cheetah paradigm revisited: MHC diversity in the world's largest free-ranging population. *Molecular biology and evolution*, **28**, 1455–1468.
- Célerier A, Huchard E, Alvergne A *et al.* (2010) Detective mice assess relatedness in baboons using olfactory cues. *Journal of Experimental Biology*, **213**, 1399–1405.
- Chaix R, Cao C, Donnelly P (2008) Is Mate Choice in Humans MHC-Dependent? *PLoS Genetics*, **4**, 1–5.
- Charpentier MJE, Boulet M, Drea CM (2008) Smelling right: the scent of male lemurs advertises genetic quality and relatedness. *Molecular Ecology*, **17**, 3225–3233.
- Charpentier MJE, Crawford JC, Boulet M, Drea CM (2010) Message ‘scent’: lemurs detect the genetic relatedness and quality of conspecifics via olfactory cues. *Animal Behaviour*, **80**, 101–108.
- Cheetham SA, Thom MD, Jury F *et al.* (2007) The genetic basis of individual-recognition signals in the mouse. *Current Biology*, **17**, 1771–1777.
- Chen HW, Liu WC, Davis AJ *et al.* (2008) Network position of hosts in food webs and their parasite diversity. *Oikos*, **117**, 1847–1855.

- 
- Cheng Y, Stuart A, Morris K *et al.* (2012) Antigen-presenting genes and genomic copy number variations in the Tasmanian devil MHC. *BMC Genomics*, **13**, 87.
- Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA (2010) The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, **186**, 983–995.
- Christin PA, Weinreich DM, Besnard G (2010) Causes and evolutionary significance of genetic convergence. *Trends in Genetics*, **26**, 400–405.
- Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic acids research*, **16**, 10881–10890.
- Cowlishaw G, Dunbar RIM (1991) Dominance rank and mating success in male primates. *Animal Behaviour*, **41**, 1045–1056.
- Coyne MJ, Smith G (1994) Trichostrongylid parasites of domestic ruminants. In: *Parasitic And Infectious Diseases: Epidemiology and Ecology* (ed. Smith MESG), pp. 235–247. Academic Press, San Diego, CA.
- Craul M, Chikhi L, Sousa V *et al.* (2009) Influence of forest fragmentation on an endangered large-bodied lemur in northwestern Madagascar. *Biological Conservation*, **142**, 2862–2871.
- Cringoli G, Rinaldi L, Maurelli MP, Utzinger J (2010) FLOTAC: new multivalent techniques for qualitative and quantitative copromicroscopic diagnosis of parasites in animals and humans. *Nature protocols*, **5**, 503–515.
- Crompton DWT, Nesheim MC (2002) Nutritional impact of intestinal helminthiasis during the human life cycle. *Annual Review of Nutrition*, **22**, 35–59.
- Cutrera AP, Lacey E a. (2007) Trans-species polymorphism and evidence of selection on class II MHC loci in tuco-tucos (*Rodentia: Ctenomyidae*). *Immunogenetics*, **59**, 937–948.
- Dammhahn M, Kappeler PM (2005) Social System of *Microcebus berthae*, the World's Smallest Primate. *International Journal of Primatology*, **26**, 407–435.
- Dammhahn M, Kappeler PM (2008a) Comparative Feeding Ecology of Sympatric *Microcebus berthae* and *Microcebus murinus*. *International Journal of Primatology*, **29**, 1567–1589.

## REFERENCES

---

- Dammhahn M, Kappeler PM (2008b) Small-scale coexistence of two mouse lemur species (*Microcebus berthae* and *Microcebus murinus*) within a homogeneous competitive environment. *Oecologia*, **157**, 473–483.
- Dammhahn M, Kappeler PM (2009) Females go where the food is: Does the socio-ecological model explain variation in social organisation of solitary foragers? *Behavioral Ecology and Sociobiology*, **63**, 939–952.
- Dammhahn M, Kappeler PM (2010) Scramble or contest competition over food in solitarily foraging mouse lemurs (*Microcebus* spp.): New insights from stable isotopes. *American Journal of Physical Anthropology*, **141**, 181–189.
- Dammhahn M, Kappeler PM (2014) Stable isotope analyses reveal dense trophic species packing and clear niche differentiation in a Malagasy primate community. *American Journal of Physical Anthropology*, **153**, 249–259.
- Danon L, Diaz-Guilera A, Duch J, Arenas A (2005) Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, **9**, P09008.
- Dawkins R, Leelayuwat C, Gaudieri S *et al.* (1999) Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunological Reviews*, **167**, 275–304.
- Degen, AA (2006) Effect of macroparasites on the energy budget of small mammals. In: Morand, S., Krasnov, B.R., Poulin, R. (Eds.), *Micromammals and Macroparasites*. Springer, Tokyo.
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM Algorithm, *Journal of the Royal Statistical Society*, **39**, 1–38.
- Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L (2007) Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in atlantic salmon. *Evolution*, **61**, 2154–2164.
- Dixson AF (1998) *Primate sexuality: Comparative studies of the prosimians, monkeys, apes, and human beings*. Oxford: Oxford University Press.
- Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, **256**, 50–52.
- Doxiadis GGM, Otting N, de Groot NG, Noort R, Bontrop RE (2000) Unprecedented polymorphism of Mhc-DRB region configurations in rhesus macaques. *The Journal of Immunology*, **164**, 3193–3199.



- 
- Doxiadis GG, de Groot N, Claas FHJ *et al.* (2007) A highly divergent microsatellite facilitating fast and accurate DRB haplotyping in humans and rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 8907–8912.
- Doxiadis GGM, de Groot N, de Groot NG *et al.* (2010) Extensive DRB region diversity in cynomolgus macaques: recombination as a driving force. *Immunogenetics*, **62**, 137–147.
- Doytchinova IA, Flower DR (2005) In silico identification of supertypes for class II MHCs. *Journal of Immunology (Baltimore, Md. : 1950)*, **174**, 7085–7095.
- Eberle M, Kappeler P (2002) Mouse lemurs in space and time: a test of the socioecological model. *Behavioral Ecology and Sociobiology*, **51**, 131–139.
- Eberle M, Kappeler PM (2004) Sex in the dark: determinants and consequences of mixed male mating tactics in *Microcebus murinus*, a small solitary nocturnal primate. *Behavioral Ecology and Sociobiology*, **57**, 77–90.
- Eberle M, Kappeler PM (2006) Family insurance: Kin selection and cooperative breeding in a solitary primate (*Microcebus murinus*). *Behavioral Ecology and Sociobiology*, **60**, 582–588.
- Edwards SV, Grahn M, Potts WK (1995a) Dynamics of MHC evolution in birds and crocodylians: amplification of class II genes with degenerate primers. *Molecular Ecology*, **4**, 719–729.
- Edwards SV, Wakeland EK, Potts WK (1995b) Contrasting histories of avian and mammalian MHC genes revealed by class II B sequences from songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 12200–12204.
- Edwards SV, Potts WK (1996) Polymorphism of MHC genes: implications for conservation genetics of vertebrates. In: Smith TB, Wayne RK (Eds.) *Molecular Genetic Approaches to Conservation*, Oxford University Press, Oxford, pp 214–237.
- Edwards SV, Chesnut K, Satta Y, Wakeland EK (1997) Ancestral polymorphism of MHC class II genes in mice: implications for balancing selection and the mammalian molecular clock. *Genetics*, **146**, 655–68.

## REFERENCES

---

- Ehlers A, Beck S, Forbes S a *et al.* (2000) MHC-linked olfactory receptor loci exhibit polymorphism and contribute to extended HLA / OR-Haplotypes. *Genome Research*, **10**, 1968–1978.
- Eimes JA., Bollmer JL, Whittingham LA. *et al.* (2011) Rapid loss of MHC class II variation in a bottlenecked population is explained by drift and loss of copy number variation. *Journal of Evolutionary Biology*, **24**, 1847–1856.
- Eimes JA, Townsend AK, Sepil I, Nishiumi I, Satta Y (2015) Patterns of evolution of MHC class II genes of crows (*Corvus*) suggest trans-species polymorphism. *PeerJ*, **3**, e853.
- Eizaguirre C, Lenz TL, Traulsen A, Milinski M (2009a) Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. *Ecology Letters*, **12**, 5–12.
- Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M (2009b) MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Molecular Ecology*, **18**, 3316–3329.
- Eizaguirre C, Lenz TL (2010) Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *Journal of Fish Biology*, **77**, 2023–2047.
- Eizaguirre C, Lenz TL, Kalbe M, Milinski M (2012) Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecology letters*, **15**, 723–731.
- Eizaguirre C, Lenz TL, Sommerfeld RD *et al.* (2011) Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. *Evolutionary Ecology*, **25**, 605–622.
- Ejsmond MJ, Radwan J (2011) MHC diversity in bottlenecked populations: a simulation model. *Conservation Genetics*, **12**, 129–137.
- Eklom R, Saether SA, Jacobsson P *et al.* (2007) Spatial pattern of MHC class II variation in the great snipe (*Gallinago media*). *Molecular Ecology*, **16**, 1439–1451.
- Ellegren H, Hartmant G, Johansson M, Andersson L (1993) Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 8150–8153.

- 
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39**, 783–791.
- Ferrandiz-Rovira M, Bigot T, Allainé D, Callait-Cardinal M-P, Cohas a (2015) Large-scale genotyping of highly polymorphic loci by next-generation sequencing: how to overcome the challenges to reliably genotype individuals? *Heredity*, **114**, 485–93.
- Fichtel C (2012) Predation. In: Mitani JC, Call J, Kappeler P, Palombit R, Silk J (Eds.) The evolution of primate societies. University of Chicago Press, Chicago: 169–194.
- Fietz J (1999a) Monogamy as a rule rather than exception in nocturnal lemurs: the case of the fat-tailed dwarf lemur, *Cheirogaleus medius*. *Ethology*, **105**, 255–272.
- Fietz J (1999b) Demography and floating males in a population of *Cheirogaleus medius*. In: Rakotosamimanana B, Rasaminanana H, Ganzhorn JU (Eds.) New Directions in Lemur Studies, pp 159–172. New York, Kluwer Academic, Plenum Publishers.
- Figuroa F, Günther E, Klein J (1988) MHC polymorphism pre-dating speciation. *Nature*, **335**, 265–267.
- Figuroa F, Mayer WE, Sülthmann H *et al.* (2000) MHC class II B gene evolution in east african cichlid fishes. *Immunogenetics*, **51**, 556–575.
- Fijarczyk A, Babik W (2015) Detecting balancing selection in genomes: limits and prospects. *Molecular Ecology*, **24**, 3529–3545.
- Finkelman FD, Shea-Donohue T, Mooris SC *et al.* (2004) Interleukin-4- and interleukin-13- mediated host protection against intestinal nematode parasites. *Immunological Reviews*, **201**, 139–155.
- Frankham R, Ballou JD, Briscoe DA (2002) Introduction to conservation genetics. Cambridge University Press, Cambridge.
- Fraser BA, Neff BD (2010) Parasite mediated homogenizing selection at the MHC in guppies. *Genetica*, **138**, 273–278.

## REFERENCES

---

- Froeschke G, Sommer S (2005) MHC class II DRB variability and parasite load in the striped mouse (*Rhabdomys pumilio*) in the Southern Kalahari. *Molecular Biology and Evolution*, **22**, 1254–1259.
- Froeschke G, Sommer S (2012) Insights into the complex associations between MHC class II DRB polymorphism and multiple gastrointestinal parasite infestations in the striped mouse. *PloS One*, **7**, 1-11.
- Furlong RF, Yang Z (2008) Diversifying and purifying selection in the peptide binding region of DRB in mammals. *Journal of Molecular Evolution*, **66**, 384–394.
- Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*, **11**, 296.
- Gangoso L, Alcaide M, Grande JM *et al.* (2012) Colonizing the world in spite of reduced MHC variation. *Journal of Evolutionary Biology*, **25**, 1438–1447.
- Ganzhorn JU, Lowri PP II, Schatz GE, Sommer S (2001) The biodiversity of Madagascar: one of the world's hottest hotspots on its way out. *Oryx*, **35**, 1-3.
- Garamszegi LZ, Nunn CL (2011) Parasite-mediated evolution of the functional part of the MHC in primates. *Journal of Evolutionary Biology*, **24**, 184–195.
- Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution*, **57**, 1707–1722.
- Gavrilets S (2004) *Fitness landscapes and the origin of species*. Princeton, NJ: Princeton University Press.
- Gilles A, Megléc E, Pech N *et al.* (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics*, **12**, 245.
- Gligor M, Ganzhorn JU, Rakotondravony D *et al.* (2009) Hybridization between mouse lemurs in an ecological transition zone in southern Madagascar. *Molecular Ecology*, **18**, 520–533.
- Godfrey SS (2013) Networks and the ecology of parasite transmission: a framework for wildlife parasitology. *International Journal for Parasitology: Parasites and Wildlife*, **2**, 235–245.
- Goüy de Bellocq J, Charbonnel N, Morand S (2008) Coevolutionary relationship between helminth diversity and MHC class II polymorphism in rodents. *Journal of Evolutionary Biology*, **21**, 1144–1150.

- Goüy De Bellocq J, Suchentrunk F, Baird SJE, Schaschl H (2009) Evolutionary history of an MHC gene in two leporid species: characterisation of Mhc-DQA in the European brown hare and comparison with the European rabbit. *Immunogenetics*, **61**, 131–144.
- Gowaty PA (2004) Sex roles, contests for the control of reproduction, and sexual selection. In: Kappeler PM and Van Schaik CP (Eds.) *Sexual Selection in Primates: new and comparative perspectives*, pp. 37-54. Cambridge, UK: Cambridge University Press.
- Graser R, Oh'Uigin C, Vincek V, Meyer A, Klein J (1996) Trans-species polymorphism of class II Mhc loci in danio fishes. *Immunogenetics*, **44**, 36–48.
- Groeneveld LF, Rasoloarison RM, Weisrock D, Kappeler PM (2008) Dwarf lemurs in space and time: phylogeography of the genus *Cheirogaleus*. In: Groeneveld L (2008) *Species delimitation, phylogeography and population genetics of the endemic malagasy dwarf lemurs (genus Cheirogaleus)*.
- Groeneveld LF, Weisrock DW, Rasoloarison RM, Yoder AD, Kappeler PM (2009) Species delimitation in lemurs: multiple genetic loci reveal low levels of species diversity in the genus *Cheirogaleus*. *BMC Evolutionary Biology*, **9**, 30.
- De Groot NG, Otting N, Doxiadis GGM *et al.* (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11748–11753.
- De Groot N, Doxiadis GGM, de Vos-Rouweler AJM *et al.* (2008) Comparative genetics of a highly divergent DRB microsatellite in different macaque species. *Immunogenetics*, **60**, 737–748.
- Gu X, Nei M (1999) Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Molecular Biology*, **16**, 147–156.
- Guindon S, Dufayard JF, Lefort V *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.
- Gustafsson K, Andersson L (1994) Structure and Polymorphism of horse MHC Class-II DRB genes: convergent evolution in the antigen-binding site. *Immunogenetics*, **39**, 355–358.
- Haldane JBS (1949) Disease and evolution. *La Ricerca Scientifica*, **19**, 68–76.

## REFERENCES

---

- Hämäläinen A, Raharivololona B, Ravoniarimbinina P, Kraus C (2015) Host sex and age influence endoparasite burdens in the gray mouse lemur. *Frontiers in Zoology*, **12**, 25.
- Hambuch TM, Lacey EA (2002) Enhanced selection for MHC diversity in social tuco-tucos. *Evolution*, **56**, 841–845.
- Harf R, Sommer S (2005) Association between major histocompatibility complex class II DRB alleles and parasite load in the hairy-footed gerbil, *Gerbillurus paeba*, in the southern Kalahari. *Molecular Ecology*, **14**, 85–91.
- Harismendy O, Ng PC, Strausberg RL *et al.* (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology*, **10**, R32.
- Hartl DL, Clark AG (1997) Principles of population genetics. Sinauer associates, Sunderland, MA.
- Havlicek J, Roberts SC (2009) MHC-correlated mate choice in humans: a review. *Psychoneuroendocrinology*, **34**, 497–512.
- Hedrick PW (1972) Maintenance of genetic variation with a frequency-dependent selection model as compared to the overdominant model. *Genetics*, **72**, 771–775.
- Hedrick PW (1985) Genetics of populations. Jones and Bartlett Publishers, Boston, MA
- Hedrick PW (1999) Balancing selection and MHC. *Genetica*, **104**, 207–214.
- Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology and Evolution*, **16**, 629–636.
- Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution*, **56**, 1902–1908.
- Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**, 4606–4618.
- Hedrick PW, Hurt CR (2012) Conservation genetics and evolution in an endangered species: research in Sonoran topminnows. *Evolutionary Applications*, **5**, 806–819.
- Hendel H, Caillat-Zucman S, Lebuane H *et al.* (1999) New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. *Journal of Immunology (Baltimore, Md. : 1950)*, **162**, 6942–6946.
- Hess CM, Edwards SV (2002) The Evolution of the major histocompatibility complex in birds. *BioScience*, **52**, 423.

- 
- Heymann EW (2006) The neglected sense: olfaction in primate behaviour, ecology and evolution. *American Journal of Primatology*, **68**, 519–524.
- Hill AVS, Allsopp CEM, Kwiatkowski D *et al.* (1991) Common west african HLA antigens are associated with protection from severe malaria. *Nature*, **352**, 595–600.
- Holmes SM, Baden AL, Brenneman RA. *et al.* (2013) Patch size and isolation influence genetic patterns in black-and-white ruffed lemur (*Varecia variegata*) populations. *Conservation Genetics*, **14**, 615–624.
- Horton R, Wilming L, Rand V *et al.* (2004) Gene map of the extended human MHC. *Nature Reviews Genetics*, **5**, 889–899.
- Hoskin CJ, Higgie M (2010) Speciation via species interactions: the divergence of mating traits within species. *Ecology Letters*, **13**, 409–420.
- Huchard E, Albrecht C, Schliehe-Diecks S *et al.* (2012) Large-scale MHC class II genotyping of a wild lemur population by next generation sequencing. *Immunogenetics*, **64**, 895–913.
- Huchard E, Baniel A, Schliehe-Diecks S, Kappeler PM (2013) MHC-disassortative mate choice and inbreeding avoidance in a solitary primate. *Molecular Ecology*, **22**, 4071–4086.
- Huchard E, Cowlshaw G, Raymond M, Weill M, Knapp LA (2006) Molecular study of MHC-DRB in wild chacma baboons reveals high variability and evidence for trans-species inheritance. *Immunogenetics*, **58**, 805–816.
- Huchard E, Knapp LA, Wang J, Raymond M, Cowlshaw G (2010a) MHC, mate choice and heterozygote advantage in a wild social primate. *Molecular Ecology*, **19**, 2545–2561.
- Huchard E, Pechouskova E (2013) The major histocompatibility complex and primate behavioral ecology: new tools and future questions. *International Journal of Primatology*, **35**, 11–31.
- Huchard E, Raymond M, Benavides J *et al.* (2010b) A female signal reflects MHC genotype in a social primate. *BMC Evolutionary Biology*, **10**, 96.
- Huchard E, Weill M, Cowlshaw G, Raymond M, Knapp LA (2008) Polymorphism, haplotype composition, and selection in the Mhc-DRB of wild baboons. *Immunogenetics*, **60**, 585–598.

## REFERENCES

---

- Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335**, 167–170.
- Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 958–62.
- Hughes J, Page RDM (2007) Comparative tests of ectoparasite species richness in seabirds. *BMC Evolutionary Biology*, **7**, 227.
- Hughes AL, Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, **32**, 415–435.
- Hurst JL, Payne CE, Nevison CM *et al.* (2001) Individual recognition in mice mediated by major urinary proteins. *Nature*, **414**, 631–634.
- Hurst JL, Thom MD, Nevison CM, Humphries RE, Beynon RJ (2005) MHC odours are not required or sufficient for recognition of individual scent owners. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 715–724.
- Irwin MT, Raharison JL (2009) A review of the endoparasites of the lemurs of Madagascar. *Malagasy Nature*, **2**, 66–93.
- Janeway C, Travers P, Walport M, Shlomchik M (2005) Immunobiology: the immune system in health and disease (6<sup>th</sup> ed.). New York: Garland.
- Jaratlerdsiri W, Isberg SR, Higgins DP, Miles LG, Gongora J (2014) Selection and trans-species polymorphism of major histocompatibility complex class II genes in the order *Crocodylia*. *PloS One*, **9**, e87534.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.
- Jones EY, Fugger L, Strominger JL, Siebold C (2006) MHC class II proteins and disease: a structural perspective. *Nature Reviews Immunology*, **6**, 271–282.
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN, (Ed.) Mammalian Protein Metabolism, pp. 21-132, Academic Press, New York.
- Kamath PL, Getz WM (2011) Adaptive molecular evolution of the major histocompatibility complex genes, DRA and DQA, in the genus *Equus*. *BMC Evolutionary Biology*, **11**, 128.



- Kamiya T, O'Dwyer K, Westerdahl H, Senior a., Nakagawa S (2014) A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity. *Molecular Ecology*, **23**, 5151–5163.
- Kappeler PM (2002) Sexual selection in primates: new and comparative perspectives. *Evolutionary Anthropology*, **11**, 173-175.
- Kappeler PM, van Schaik CP (2002) Evolution of primate social systems. *International Journal of Primatology*, **23**, 707-740.
- Kappeler PM, Wimmer B, Zinner D, Tautz D (2002) The hidden matrilineal structure of a solitary lemur: implications for primate social evolution. *Proceedings of the Royal Society of London B: Biological Sciences*, **269**, 1755–1763.
- Kappeler PM (2003) *Mirza coquereli*, Coquerel's dwarf lemur. In: Goodman SM, Benstead JP (Eds.) *The natural history of Madagascar*, pp. 1316-1318 Chicago: University of Chicago Press.
- Kappeler (2012) The behavioral ecology of strepsirrhines and tarsiers In: Mitani JC, Call J, Kappeler PM, Palombit RA, Silk JB (Eds.) *The evolution of primate societies*. University of Chicago Press, Chicago: 17-42.
- Kappeler PM, Fichtel C (2012) A 15-year perspective on the social organization and life history of sifaka in Kirindy Forest. In: PM Kappeler, Watts DP (Eds.) *Long term field studies of primates*. Springer-Verlag Berlin Heidelberg.
- Kappeler PM, & Watts DP (Eds.) (2012) *Long-term Field Studies of Primates*. Heidelberg: Springer.
- Kazem AJN, Widdig A (2013) Visual phenotype matching: cues to paternity are present in Rhesus Macaque Faces. *PLoS One*, **8**, e55846.
- Keller LF, Waller DM (2002) Inbreeding effects in wild populations. *Trends in Ecology and Evolution*, **17**, 230-241.
- Kelley J, Walter L, Trowsdale J (2005a) Comparative genomics of major histocompatibility complexes. *Immunogenetics*, **56**, 683–695.
- Kessler SE, Scheumann M, Nash LT, Zimmermann E (2012) Paternal kin recognition in the high frequency / ultrasonic range in a solitary foraging mammal. *BMC Ecology*, **12**, 26.

## REFERENCES

---

- Kiemnec-Tyburczy KM, Richmond JQ, Savage a E, Lips KR, Zamudio KR (2012) Genetic diversity of MHC class I loci in six non-model frogs is shaped by positive selection and gene duplication. *Heredity*, **109**, 146–155.
- Klein J (1986) Natural History of the major histocompatibility complex. John Wiley and Sons, New York.
- Klein J (1987) Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Human Immunology*, **19**, 155–162.
- Klein J, Satta Y, O’huigin C, Takahata N (1993a) The molecular descent of the major histocompatibility complex. *Annual Review of Immunology*, **11**, 269–295.
- Klein D, Ono H, O’huigin C *et al.* (1993b) Extensive MHC variability in cichlid fishes of Lake Malawi. *Nature*, **364**, 330–334.
- Klein J, Sato A, Nagl S, O’huigin C (1998) Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics*, **29**; 1–21.
- Klein J, Sato A, Nikolaidis N (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annual Review of Genetics*, **41**, 281–304.
- Kloch A, Baran K, Buczek M, Konarzewski M, Radwan J (2013) MHC influences infection with parasites and winter survival in the root vole *Microtus oeconomus*. *Evolutionary Ecology*, **27**, 635–653.
- Knapp LA (2005a) The ABCs of MHC. *Evolutionary Anthropology*, **14**, 28–37.
- Knapp LA (2005b) Denaturing gradient gel electrophoresis and its use in the detection of major histocompatibility complex polymorphism. *Tissue Antigens*, **65**, 211–219.
- Knapp LA (2007) Selection on MHC: A matter of form over function. *Heredity*, **99**, 241–242.
- Knapp LA, Cadavid LF, Eberle ME *et al.* (1997) Identification of new *Mamu*-DRB alleles using DGGE and direct sequencing. *Immunogenetics*, **45**, 171–179.
- Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The Next-generation sequencing revolution and its impact on genomics. *Cell*, **155**, 27–38.
- Kraus C, Eberle M, Kappeler PM (2008) The costs of risky male behaviour: sex differences in seasonal survival in a small sexually monomorphic primate. *Proceedings of the Royal Society of London B: Biological Sciences*, **275**, 1635–1644.

- Kremen C, Cameron A, Moilanen A *et al.* (2008) Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, **320**, 222–226.
- Kriener K, O’hUigin C, Tichy H, Klein J (2000) Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics*, **51**, 169–178.
- Kriener K, O’hUigin C, Klein J (2001) Independent origin of functional MHC class II genes in humans and New World monkeys. *Human Immunology*, **62**, 1–14.
- Kuduk K, Babik W, Bojarska K *et al.* (2012) Evolution of major histocompatibility complex class I and class II genes in the brown bear. *BMC Evolutionary Biology*, **12**, 197.
- Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunological Reviews*, **190**, 95–122.
- Kwak J, Willse A, Matsumura K *et al.* (2008) Genetically-based olfactory signatures persist despite dietary variation. *PLoS One*, **3**, 10.
- Kwak J, Willse A, Preti G, Yamazaki K, Beauchamp GK (2010) In search of the chemical basis for MHC odourtypes. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 2417–2425.
- Laurent R, Chaix R (2012) MHC-dependent mate choice in humans: why genomic patterns from the HapMap European American dataset support the hypothesis. *BioEssays*, **34**, 267–271.
- Leinders-Zufall T, Brennan P, Widmayer P *et al.* (2004) MHC class I peptides as chemosensory signals in the vomeronasal organ. *Science*, **306**, 1033–1037.
- Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci: implications for evolutionary analysis. *Gene*, **427**, 117–123.
- Lenz TL, Wells K, Pfeiffer M, Sommer S (2009) Diverse MHC IIB allele repertoire increases parasite resistance and body condition in the Long-tailed giant rat (*Leopoldamys sabanus*). *BMC Evolutionary Biology*, **9**, 269.
- Lenz TL (2011) Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution*, **65**, 2380–2390.

## REFERENCES

---

- Lenz TL, Eizaguirre C, Kalbe M, Milinski M (2013) Evaluating patterns of convergent evolution and trans-species polymorphism at mhc immunogenes in two sympatric stickleback species. *Evolution*, **67**, 2400–2412.
- Li M, Stoneking M (2012) A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biology*, **13**, R34.
- Li L, Zhou X, Chen X (2011) Characterization and Evolution of MHC class II B genes in Ardeid birds. *Journal of Molecular Evolution*, **72**, 474–483.
- Lighten J, van Oosterhout C, Bentzen P (2014a) Critical review of NGS analyses for de novo genotyping multigene families. *Molecular Ecology*, **23**, 3957–3972.
- Lighten J, van Oosterhout C, Paterson IG, McMullan M, Bentzen P (2014b) Ultra-deep Illumina sequencing accurately identifies MHC class IIb alleles and provides evidence for copy number variation in the guppy (*Poecilia reticulata*). *Molecular Ecology Resources*, **14**, 753–767.
- Llaurens V, McMullan M, van Oosterhout C (2012) Cryptic MHC polymorphism revealed but not explained by selection on the class IIB peptide-binding Region. *Molecular Biology and Evolution*, **29**, 1631–1644.
- Logan DW, Marton TF, Stowers L (2008) Species specificity in major urinary proteins by parallel evolution. *PloS One*, **3**, e3280.
- Loiseau C, Zoorob R, Garnier S *et al.* (2008) Antagonistic effects of a MHC class I allele on malaria-infected house sparrows. *Ecology Letters*, **11**, 258–265.
- Loiseau C, Zoorob R, Robert A *et al.* (2011) Plasmodium relictum infection and MHC diversity in the house sparrow (*Passer domesticus*). *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 1264–1272.
- Louis, ER Jr., Ratsimbazafy JH, Razakamaharavo VR *et al.* (2005) Conservation genetics of black and white ruffed lemurs, *Varecia variegata*, from southeastern Madagascar. *Animal Conservation*, **8**, 105–111.
- Lukas D, Bradley BJ, Nsubuga AM *et al.* (2004) Major histocompatibility complex and microsatellite variation in two populations of wild gorillas. *Molecular Ecology*, **13**, 3389–3402.
- Lund O, Nielsen M, Kesmir C *et al.* (2004) Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics*, **55**, 797–810.

- Lundberg AS, McDevitt HO (1992) Evolution of major histocompatibility complex class II allelic diversity: direct descent in mice and humans. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 6545–6549.
- Maan ME, Seehausen O (2011) Ecology, sexual selection and speciation. *Ecology Letters*, **14**, 591–602.
- Markolf M, Roos C, Kappeler P (2008) Genetic and demographic consequences of a rapid reduction in population size in a solitary lemur (*Mirza coquereli*). *The Open Conservation Biology Journal*, **2**, 21–29.
- Martinsohn JT, Sousa a B, Guethlein L a, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics*, **50**, 168–200.
- Mas-Coma S, Valero MA, Bargues MD (2008) Effects of climate change on animal and zoonotic helminthiasis. *Revue Scientifique et Technique*, **27**, 443–452.
- Matthews B, Harmon LJ, M’Gonigle L, Marchinko KB, Schaschl H (2010) Sympatric and allopatric divergence of MHC genes in threespine stickleback. *PloS One*, **5**, e10948.
- McConnell TJ, Talbot WS, McIndoe RA, Wakeland EK (1988) The origin of MHC class II gene polymorphism within the genus *Mus*. *Nature*, **332**, 651–654.
- Meagher S (1999) Genetic diversity and *Capillaria hepatica* (Nematoda) prevalence in Michigan deer mouse populations. *Evolution*, **53**, 1318.
- Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics*, **11**, 31–46.
- Meyer-Lucht Y, Sommer S (2005) MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). *Molecular Ecology*, **14**, 2233–2243.
- Mikko S, Andersson L (1995) Low major histocompatibility complex class II diversity in European and North American moose. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 4259–4263.
- Milinski M (2006) The major histocompatibility complex, sexual selection, and mate choice. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 159–186.
- Milinski M, Croy I, Hummel T, Boehm T (2013) Major histocompatibility complex peptide ligands as olfactory cues in human body odour assessment. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**, 20122889.

## REFERENCES

---

- Mitani JC, Call J, Kappeler PM, Palombit RA and Silk JB (Eds.) (2012) *The Evolution of primate societies*. Chicago: The University of Chicago Press.
- Mittermeier RA, Konstant WR, Nicoll ME, Langrand O (1992) *Lemurs of Madagascar: an action plan for their conservation, 1993–1999*. IUCN, Gland, Switzerland.
- Mittermeier R, Louis E, Hawkins F, Langrand O, Ganzhorn J, Konstant W, Rasoloarison R, Rajaobelina S, Richardson M (2008) *Lemurs of Madagascar, 3<sup>rd</sup> edition*. Conservation International.
- Moore MJ, Dhingra A, Soltis PS *et al.* (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC plant biology*, **6**, 17.
- Morand S, Bouamer S, Hugot JP (2006) Nematodes. In: Morand S, Krasnov BR, Poulin R (Eds.) *Micromammals and Macroparasites: From Evolutionary Ecology to Management*, pp. 63-79. Springer-Verlag, Tokyo.
- Morand S (2015) (macro-) Evolutionary ecology of parasite diversity: from determinants of parasite species richness to host diversification. *International Journal for Parasitology: Parasites and Wildlife*, **4**, 80–87.
- Mundry R, Nunn CL (2009) Stepwise model fitting and statistical inference: turning noise into signal pollution. *American Naturalist*, **173**, 119-123.
- Myers RM, Fischer SG, Lerman LS, Maniatis T (1985) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Research*, **13**, 3131–3145.
- Nadachowska-Brzyska K, Zieliński P, Radwan J, Babik W (2012) Interspecific hybridization increases MHC class II diversity in two sister species of newts. *Molecular Ecology*, **21**, 887–906.
- Nei M, Gojoborit T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418–426.
- Nei M, Hughes AL (1992) Balanced polymorphism and evolution by the birth- and-death process in the MHC loci. In: Tsuji K, Aizawa M, Sasazuki T (Eds.) *Eleventh histocompatibility workshop and conference*, pp 27– 38. Oxford University Press, Oxford, England,
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 7799–7806.

- 
- Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York.
- Nielsen M, Lundegaard C, Blicher T *et al.* (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One*, **2**:e796
- Nunn CL, Altizer S, Jones KE, Sechrest W (2003) Comparative test of parasite species richness in primates. *The American Naturalist* **162**:597–614.
- O'Brien SJ, Evermann JF (1988) Interactive influence of infectious disease and genetic diversity in natural populations. *Trends in Ecology and Evolution*, **3**, 254–259.
- O'hUigin C, Satta Y, Hausmann A, Dawkins RL, Klein J (2000) The implications of intergenic polymorphism for major histocompatibility complex evolution. *Genetics*, **156**, 867–877.
- Oksanen J, Blanchet FG, Kindt R *et al.* (2015) Vegan: Community Ecology Package. R package version 2.3-0. <http://CRAN.R-project.org/package=vegan>
- Oliver MK, Telfer S, Piertney SB (2009) Major histocompatibility complex (MHC) heterozygote superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proceedings of the Royal Society of London B: Biological sciences*, **276**, 1119–1128.
- Oliver MK, Piertney SB (2012) Selection maintains MHC diversity through a natural population bottleneck. *Molecular Biology and Evolution*, **29**, 1713–1720.
- Olivieri GL, Sousa V, Chikhi L, Radespiel U (2008) From genetic diversity and structure to conservation: genetic signature of recent population declines in three mouse lemur species (*Microcebus* spp.). *Biological Conservation*, **141**, 1257–1271.
- van Oosterhout C, Joyce DA, Cummings SM *et al.* (2006) Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution*, **60**, 2562–2574.
- van Oosterhout C (2009) A new theory of MHC evolution: beyond selection on the immune genes. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 657–665.
- Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand

## REFERENCES

---

- conformation polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 2766–2770.
- Otting N, de Groot N, Doxiadis G, Bontrop R (2002) Extensive Mhc-DQB variation in humans and non-human primate species. *Immunogenetics*, **54**, 230–239.
- Otting N, Heijmans CMC, Noort RC *et al.* (2005) Unparalleled complexity of the MHC class I region in rhesus macaques. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 1626–1631.
- Ottová E, Šimková A, Martin J-F *et al.* (2005) Evolution and trans-species polymorphism of MHC class II $\beta$  genes in cyprinid fish. *Fish and Shellfish Immunology*, **18**, 199–222.
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Parr LA, Heintz M, Lonsdorf E, Wroblewski E (2011) Visual kin recognition in nonhuman primates: (*Pan troglodytes* and *Macaca mulatta*): Inbreeding avoidance or male distinctiveness? *Journal of Comparative Psychology*, **124**, 343–350.
- Pearce EJ, Tarlton RL (2002) Overview of the parasitic pathogens. In: Kaufmann SHE, Sher A, Ahmed R (Eds.) *Immunology of infectious diseases* ASM Press, Washington. pp 39–52.
- Pechouskova E, Dammhahn M, Brameier M *et al.* (2015) MHC class II variation in a rare and ecological specialist mouse lemur reveals lower allelic richness and contrasting selection patterns compared to a generalist and widespread sympatric congener. *Immunogenetics*, **67**, 229–245.
- Pedersen AB, Greives TJ (2008) The interaction of parasites and resources cause crashes in a wild mouse population. *Journal of Animal Ecology*, **77**, 370–377.
- Penn, DJ (2001) Coevolution: host-Parasite. In: eLS. John Wiley & Sons Ltd, Chichester.
- Penn DJ (2002) The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology*, **108**, 1–21.
- Penn DJ, Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. *The American Naturalist*, **153**, 145–164.
- Perrigoue JG, Marshall FA, Artis D (2008) On the hunt for helminths: innate immune cells in the recognition and response to helminth parasites. *Cellular Microbiology*, **10**, 1757–1764.



- Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.
- Pilosof S, Fortuna MA, Cosson J-F *et al.* (2014) Host-parasite network structure is associated with community-level immunogenetic diversity. *Nature Communications*, **5**, 5172.
- Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.
- Port M, Kappeler PM (2010) The utility of reproductive skew models in the study of male primates, a critical evaluation. *Evolutionary Anthropology*, **19**, 46–56.
- Potts WK, Wakeland EK (1993) Evolution of MHC genetic diversity: a tale of incest, pestilence and sexual preference. *Trends in Genetics*, **9**, 408–412.
- Promerová M, Králová T, Bryjová A, Albrecht T, Bryja J (2013) MHC class IIB exon 2 polymorphism in the Grey partridge (*Perdix perdix*) is shaped by selection, recombination and gene conversion. *PloS One*, **8**, e69135.
- Prugnolle F, Manica A, Charpentier M *et al.* (2005) Pathogen-driven selection and worldwide HLA class I diversity. *Current Biology*, **15**, 1022–7.
- Radespiel U, Rakotondravony R, Chikhi L (2008) Natural and anthropogenic determinants of genetic structure in the largest remaining population of the endangered golden-brown mouse lemur, *Microcebus ravelobensis*. *American Journal of Primatology*, **70**, 860–870.
- Radwan J, Kawałko A, Wójcik JM, Babik W (2006) MHC-DRB3 variation in a free-living population of the European bison, *Bison bonasus*. *Molecular Ecology*, **16**, 531–540.
- Radwan J, Biedrzycka A, Babik W (2010) Does reduced MHC diversity decrease viability of vertebrate populations? *Biological Conservation*, **143**, 537–544.
- Raharivololona B (2006) Gastrointestinal parasites of *Cheirogaleus* sp. and *Microcebus murinus* in the littoral forest of Mandena, Madagascar. *Lemur News*, **11**, 31–35.
- Raharivololona B (2009) Parasites gastro-intestinaux de *Microcebus murinus* de la forêt littorale de Mandena, Madagascar. *Madagascar Conservation and Development*, **4**.

## REFERENCES

---

- Raharivololona B, Ganzhorn J (2010) Seasonal variations in gastrointestinal parasites excreted by the gray mouse lemur *Microcebus murinus* in Madagascar. *Endangered Species Research*, **11**, 113–122.
- Rammensee HG, Friede T, Stevanović S.(1995) MHC ligands and peptide motifs: first listing. *Immunogenetics*, **41**, 178–228.
- Rasoloarison RM, Goodman SM, Ganzhorn U (2000) Taxonomic revision of mouse lemurs (*Microcebus*) in the western portions of Madagascar. *International Journal of Primatology*, **21**, 963-1019.
- Razakamaharavo VR, McGuire SM, Vasey N, Louis EE, Brenneman RA (2010) Genetic architecture of two red ruffed lemur (*Varecia rubra*) populations of Masoala National Park. *Primates*, **51**, 53–61.
- R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rendall D, Rodman PS, Emond RE (1996) Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour*, **51**, 1007–1015.
- Rensing S (1999) Immobilization and anesthesia of nonhuman primates. *Primate report*, **55**, 33–38.
- Richman AD, Herrera LG, Nash D (2001) MHC class II beta sequence diversity in the deer mouse (*Peromyscus maniculatus*): implications for models of balancing selection. *Molecular Ecology*, **10**, 2765–2773.
- Richman a D, Kohn JR (1999) Self-incompatibility alleles from *Physalis*: implications for historical inference from balanced genetic polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 168–172.
- Ritchie LS (1948) An ether sedimentation technique for routine stool examinations. *Bulletin of the 639 US Army Medical Department United States Army Medical Department*, **8**, 326.
- Reusch TBH, Langefors Å (2005) Inter- and intralocus recombination drive MHC class IIB gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. *Journal of Molecular Evolution*, **61**, 531–541.
- Rifkin JL, Nunn CL, Garamszegi LZ (2012) Do animals living in larger groups experience greater parasitism? A meta-analysis. *The American Naturalist*, **180**, 70–82.

- 
- Robinson J, Halliwell JA, McWilliam H *et al.* (2013) The IMGT/HLA database. *Nucleic Acids Research*, **41**, D1222-D1227.
- Rousset F, Raymond M (1995) Testing heterozygote excess and deficiency. *Genetics*, **140**, 1413–1419.
- Ruff JS, Nelson AC, Kubinak JL, Potts WK (2012) MHC signalling during social communication. In: Lopez-Larrea C (Ed.) *Self and non-self*, pp. 290-313. New York: Landes Bioscience and Springer Science+Business Media.
- Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.
- Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *Journal of Medicinal Chemistry*, **41**, 2481–2491.
- Sauermann U, Nürnberg P, Bercovitch FB *et al.* (2001) Increased reproductive success of MHC class II heterozygous males among free-ranging rhesus macaques. *Human Genetics*, **108**, 249–254.
- Schad J, Sommer S, Ganzhorn JU (2004) MHC Variability of a Small Lemur in the Littoral Forest Fragments of Southeastern Madagascar. *Conservation Genetics*, **5**, 299–309.
- Schad AJ, Ganzhorn JU, Sommer S (2005) Parasite burden and constitution of major histocompatibility complex in the malagasy mouse lemur, *Microcebus murinus*. *Evolution*, **59**, 439–450.
- Schad J, Dechmann DKN, Voigt CC, Sommer S (2012) Evidence for the “good genes” model: association of MHC class II DRB alleles with ectoparasitism and reproductive state in the neotropical lesser bulldog bat, *Noctilio albiventris*. *PLoS One*, **7**, e37101.
- Schäffler L, Kappeler PM (2014) Distribution and abundance of the world’s smallest primate, *Microcebus berthae*, in central western Madagascar. *International Journal of Primatology*, **35**, 557–572.
- Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372–380.

## REFERENCES

---

- Schmid J, Kappeler PM (1994) Sympatric mouse lemurs (*Microcebus* spp.) in western Madagascar. *Folia Primatologica*, **63**, 162–170.
- Schwab D, Ganzhorn JU (2004) Distribution, population structure and habitat use of *Microcebus berthae* compared to those of other sympatric cheirogalids. *International Journal of Primatology*, **25**, 307–330.
- Schwensow N, Fietz J, Dausmann KH, 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.
- Schwensow N, Eberle M, Sommer S (2008a) Compatibility counts: MHC-associated mate choice in a wild promiscuous primate. *Proceedings of the Royal Society of London B: Biological sciences*, **275**, 555–564.
- Schwensow N, Fietz J, Dausmann K, Sommer S (2007b) MHC-associated mating strategies and the importance of overall genetic diversity in an obligate pair-living primate. *Evolutionary Ecology*, **22**, 617–636.
- Schwensow N, Eberle M, Sommer S (2010a) Are there ubiquitous parasite-driven major histocompatibility complex selection mechanisms in gray mouse lemurs? *International Journal of Primatology*, **31**, 519–537.
- Schwensow N, Dausmann K, Eberle M, Fietz J, Sommer S (2010b) Functional associations of similar MHC alleles and shared parasite species in two sympatric lemurs. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, **10**, 662–8.
- Schwitzer C, Mittermeier RA, Johnson SE *et al.* (2013) Averting lemur extinctions amid Madagascar's political crisis. *Science*, **343**, 842–843.
- Schwitzer C, Mittermeier RA, Davies N, Johnson S, Ratsimbazafy J, Razafindramanana J, Louis Jr EE, Rajaobelina S (Eds.) (2013a) Lemurs of Madagascar a strategy for their conservation 2013–2016. IUCN SSC Primate Specialist Group, Bristol.
- Sebastian A, Herdegen M, Migalska M, Radwan J (2015) AMPLISAS: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Molecular Ecology Resources*, e-published.
- Segal S, Hill AVS (2003) Genetic susceptibility to infectious disease. *Trends in Microbiology*, **11**, 445–448.

- Sepil I, Moghadam HK, Huchard E, Sheldon BC (2012) Characterization and 454 pyrosequencing of major histocompatibility complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evolutionary Biology*, **12**, 68.
- Sepil I, Lachish S, Hinks AE, Sheldon BC (2013) MHC supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population MHC supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**, 20130134.
- Setchell JM, Kappeler PM (2003) Selection in relation to sex in primates. *Advances in the Study of Behaviour*, **33**, 87–176.
- Setchell JM, Charpentier MJE, Abbott KM, Wickings EJ, Knapp LA (2009) Is brightest best? Testing the Hamilton-Zuk hypothesis in mandrills. *International Journal of Primatology*, **30**, 825–844.
- Setchell JM, Charpentier MJE, Abbott KM, Wickings EJ, Knapp LA (2010) Opposites attract: MHC-associated mate choice in a polygynous primate. *Journal of Evolutionary Biology*, **23**, 136–48.
- Setchell JM, Huchard E (2010) The hidden benefits of sex: evidence for MHC-associated mate choice in primate societies. *BioEssays*, **32**, 940–948.
- Setchell JM, Vaglio S, Abbott KM *et al.* (2011) Odour signals major histocompatibility complex genotype in an Old World monkey. *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 274–280.
- Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics*, **50**, 201–212.
- Shendure J, Ji HL (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1135–1145.
- Shum BP, Guethlein L, Flodin LR *et al.* (2001) Modes of salmonid MHC Class I and II evolution differ from the primate paradigm. *The Journal of Immunology*, **166**, 3297–3308.
- Siddle HV, Marzec J, Cheng Y, Jones M, Belov K (2010) MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 2001–2006.

## REFERENCES

---

- Slatkin M, Excoffier L (1996) Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity*, **76**, 377–383.
- Slierendregt BL, Otting N, van Besouw N, Jonker M, Bontrop RE (1994) Expansion and contraction of rhesus macaque DRB regions by duplication and deletion. *Journal of Immunology*, **152**, 2298-2307.
- Snell GD (1968) The H-2 locus of the mouse: observations and speculations concerning its comparative genetics and its polymorphism. *Folia Biologica (Praha)*, **14**, 335–358.
- Sommer S, Schwab D, Ganzhorn J (2002) MHC diversity of endemic malagasy rodents in relation to geographic range and social system. *Behavioral Ecology and Sociobiology*, **51**, 214–221.
- Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, **2**, 16.
- Sommer S, Courtiol A, Mazzoni CJ (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics*, **14**, 542.
- Sommer S, Rakotondranary SJ, Ganzhorn JU (2014) Maintaining microendemic primate species along an environmental gradient - parasites as drivers for species differentiation. *Ecology and Evolution*, **4**, 4751-4765.
- Southwood S, Sidney J, Kondo A *et al.* (1998) Several common HLA-DR types share largely overlapping peptide binding repertoires. *Journal of Immunology (Baltimore, Md. : 1950)*, **160**, 3363–3373.
- Speed JDM, Cooper EJ, Jónsdóttir IS, Van Der Wal R, Woodin SJ (2010) Plant community properties predict vegetation resilience to herbivore disturbance in the Arctic. *The Journal of Ecology*, **98**, 1002–1013.
- Spielman D, Brook BW, Briscoe D a., Frankham R (2004) Does inbreeding and loss of genetic diversity decrease disease resistance? *Conservation Genetics*, **5**, 439–448.
- Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society of London B: Biological sciences*, **277**, 979–988.
- Spurgin LG, Van Oosterhout C, Illera JC *et al.* (2011) Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology*, **20**, 5213–5225.

- Srithayakumar V, Castillo S, Mainguy J, Kyle CJ (2012) Evidence for evolutionary convergence at MHC in two broadly distributed mesocarnivores. *Immunogenetics*, **64**, 289–301.
- Stear MJ, Bairden K, Duncan JL, Holmes PH, McKellar QA, et al. (1997) How hosts control worms. *Nature*, **389**, 27.
- Stiebens VA, Merino SE, Chain FJJ, Eizaguirre C (2013) Evolution of MHC class I genes in the endangered loggerhead sea turtle (*Caretta caretta*) revealed by 454 amplicon sequencing. *BMC Evolutionary Biology*, **13**, 95.
- Stien A, Irvine RJ, Ropstad E *et al.* (2002) The impact of gastrointestinal nematodes on wild reindeer: experimental and cross-sectional studies. *Journal of Animal Ecology*, **71**, 937–945.
- Šimková A, Ottová E, Morand S (2006) MHC variability, life-traits and parasite diversity of European cyprinid fish. *Evolutionary Ecology*, **20**, 465–477.
- Sutton JT, Robertson BC, Grueber CE, Stanton JAL, Jamieson IG (2013) Characterization of MHC class II B polymorphism in bottlenecked New Zealand saddlebacks reveals low levels of genetic diversity. *Immunogenetics*, **65**, 619–633.
- Tamás J, Podani J, Csontos P (2001) An extension of presence/absence coefficients to abundance data: a new look at absence. *Journal of Vegetation Science*, **12**, 401–410.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, **30**, 2725–9.
- Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA Sequences. *Lectures on Mathematics in the Life Sciences (American Mathematical Society)*, **17**, 57–86.
- Těšický M, Vinkler M (2015) Trans-species polymorphism in immune genes: general Pattern or MHC-restricted phenomenon? *Journal of Immunology Research*, **2015**, 1–10.
- Thiele D, Razafimahatratra E, Hapke A (2013) Discrepant partitioning of genetic diversity in mouse lemurs and dwarf lemurs--biological reality or taxonomic bias? *Molecular Phylogenetics and Evolution*, **69**, 593–609.
- Thom MD, Stockley P, Beynon RJ and Hurst JL (2008) Scent, mate choice and genetic

## REFERENCES

---

- heterozygosity. In: Hurst JL, Beynon RJ, Roberts SC, Wyatt TD (Eds.), *Chemical signals in vertebrates 11* (pp. 291-301). New York: Springer.
- Tobler M, Plath M, Riesch R *et al.* (2014) Selection from parasites favours immunogenetic diversity but not divergence among locally adapted host populations. *Journal of Evolutionary Biology*, **27**, 960–974.
- Tollenaere C, Bryja J, Galan M *et al.* (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.
- Trachtenberg E, Korber B, Sollars C *et al.* (2003) Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, **9**, 928–935.
- Tracy KE, Kiemiec-Tyburczy KM, DeWoody JA, Parra-Olea G, Zamudio KR (2015) Positive selection drives the evolution of a major histocompatibility complex gene in an endangered Mexican salamander species complex. *Immunogenetics*, **67**, 323–335.
- Trivers RL (1972) Parental investment and sexual selection. In: Campbell P (Ed.), *Sexual selection and the descent of man* (pp. 136-179). London: Heinemann.
- Trowsdale J (1993) Genomic structure and function in the MHC. *Trends in Genetics*, **9**, 117-122.
- Trowsdale J (1995) ‘Both bird and man and beast’: comparative organization of MHC genes. *Immunogenetics*, **41**, 1–17.
- Trtkova K, Kupfermann H, Grahovac B *et al.* (1993) Mhc-DRB genes of platyrrhine primates. *Immunogenetics*, **38**, 210–222.
- Trtkova K, Mayer WE, O’Hugin C, Klein J (1995) MHC-DRB genes and the origin of New World monkeys. *Molecular Phylogenetics and Evolution*, **4**, 408-419.
- Tung J, Alberts SC, Wray GA (2010) Evolutionary genetics in wild primates: combining genetic approaches with field studies of natural populations. *Trends in Genetics*, **26**, 353–362.
- Villinger J, Waldman B (2012) Social discrimination by quantitative assessment of immunogenetic similarity. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 4368–4374.
- Vitone ND, Altizer S, Nunn CL (2004) Body size, diet and sociality influence the species richness of parasitic worms in anthropoid primates. *Evolutionary Ecology Research*, **6**, 183–199.



- Völker M, Backström N, Skinner BM *et al.* (2010) Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Research*, **20**, 503–511.
- Wakeland EK, Boehme S, She JX *et al.* (1990) Ancestral polymorphisms of MHC class II genes: Divergent allele advantage. *Immunologic Research*, **9**, 115–122.
- Watts AG, Alexander SM (2012) Community variation of gastrointestinal parasites found in urban and rural coyotes (*Canis latrans*) of Calgary, Alberta. *Cities and Environment (CATE)*, **4**, Article 11.
- Waldman B (1988) The ecology of kin recognition. *Annual Review of Ecology and Systematics*, **19**, 543–571.
- Wegner KM, Reusch TBH, Kalbe M (2003) Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *Journal of Evolutionary Biology*, **16**, 224–232.
- Wegner KM (2009) Massive parallel MHC genotyping: titanium that shines. *Molecular Ecology*, **18**, 1818–1820.
- Wegner KM, Eizaguirre C (2012) New(t)s and views from hybridizing MHC genes: Introgression rather than trans-species polymorphism may shape allelic repertoires. *Molecular Ecology*, **21**, 779–781.
- Weil ZM, Martin II LB, Nelson RJ (2006) Interactions among immune, endocrine, and behavioural response to infection. In: Morand S, Krasnov BR, Poulin R (Eds.) *Micromammals and Macroparasites*, pp 443–473. Tokyo: Springer-Verlag.
- Weisrock DW, Rasoloarison RM, Fiorentino I *et al.* (2010) Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. *PloS One*, **5**, e9883.
- Westerdahl H, Waldenstrom J, Hansson B *et al.* (2005) Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 1511–1518.
- Westerdahl H, Stjernman M, Råberg L, Lannefors M, Nilsson J-Å (2013) MHC-I affects infection intensity but not infection status with a frequent avian malaria parasite in blue tits. *PLoS One*, **8**, e72647.
- Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP (2006) Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, **75**, 1182–1189.

## REFERENCES

---

- Widdig A (2007) Paternal kin discrimination: the evidence and likely mechanisms. *Biological Reviews*, **82**, 319–334.
- Willse A, Kwak J, Yamazaki K *et al.* (2006) Individual odortypes: interaction of MHC and background genes. *Immunogenetics*, **58**, 967–982.
- Winternitz JC, Minchey SG, Garamszegi LZ *et al.* (2013) Sexual selection explains more functional variation in the mammalian major histocompatibility complex than parasitism. *Proceedings of the Royal Society of London B: Biological Sciences*, **280**, 20131605.
- Winternitz JC, Wares JP (2013) Duplication and population dynamics shape historic patterns of selection and genetic variation at the major histocompatibility complex in rodents. *Ecology and Evolution*, **3**, 1552–68.
- Woelfling B, Traulsen A, Milinski M, Boehm T (2009) Does intra-individual major histocompatibility complex diversity keep a golden mean? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **364**, 117–128.
- Wright S (1965) The interpretation of population structure by F–statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.
- Wright S (1969) *Evolution and the genetics of populations, Vol 2: The theory of gene frequencies*. University of Chicago Press, Chicago.
- Xu SX, Ren WH, Li SZ *et al.* (2009) Sequence polymorphism and evolution of three Cetacean MHC genes. *Journal of Molecular Evolution*, **69**, 260–275.
- Yamazaki K, Beauchamp GK (2007) Genetic basis for MHC-dependent mate choice. *Advances in Genetics*, **59**, 129–145.
- Yamazaki K, Beauchamp GK, Curran M, Bard J, Boyse EA (2000) Parent-progeny recognition as a function of MHC odortype identity. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 10500–10502.
- Yamazaki K, Boyse EA, Miké V *et al.* (1976) Control of mating preferences in mice by genes in the major histocompatibility complex. *The Journal of Experimental Medicine*, **144**, 1324–1335.
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, **155**, 431–449.

- Yang Z, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular Biology and Evolution*, **19**, 49–57.
- Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107–1118.
- Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.
- Yao Y-F, Dai Q-X, Li J *et al.* (2014) Genetic diversity and differentiation of the rhesus macaque (*Macaca mulatta*) population in western Sichuan, China, based on the second exon of the major histocompatibility complex class II DQB (MhcMamu-DQB1) alleles. *BMC Evolutionary Biology*, **14**, 130.
- Yeager M, Hughes AL (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunological Reviews*, **167**, 45–58.
- Yeager M, Kumar S, Hughes AL (1997) Sequence convergence in the peptide-binding region of primate and rodent MHC class Ib molecules. *Molecular Biology and Evolution*, **14**, 1035–1041.
- Younger RM, Amadou C, Bethel G *et al.* (2001) Characterization of clustered MHC-linked olfactory receptor genes in human and mouse. *Genome Research*, **11**, 519–530.
- Zagalska-Neubauer M, Babik W, Stuglik M *et al.* (2010) 454 sequencing reveals extreme complexity of the class II major histocompatibility complex in the collared flycatcher. *BMC Evolutionary Biology*, **10**, 395.
- Zhang J, Kumar S (1997) Detection of convergent and parallel evolution at the amino acid sequence level. *Molecular Biology and Evolution*, **14**, 527–536.
- Zhang M, He H (2013) Parasite-mediated selection of major histocompatibility complex variability in wild brandt's voles (*Lasiopodomys brandtii*) from Inner Mongolia, China. *BMC Evolutionary Biology*, **13**, 149.
- Zhang XS, Wang J, Hill WG (2002) Pleiotropic model of maintenance of quantitative genetic variation at mutation-selection balance. *Genetics*, **161**, 419–433.
- Zhao M, Wang Y, Shen H *et al.* (2013) Evolution by selection, recombination, and gene duplication in MHC class I genes of two *Rhacophoridae* species. *BMC Evolutionary Biology*, **13**, 113.

## REFERENCES

---

Zuckerkindl E, Pauling L (1965) Evolutionary divergence and convergence in proteins.  
In: Bryson V, Vogel HJ (Eds.) *Evolving genes and proteins* (Eds.), pp. 97-166.  
Academic Press, New York.

# ACKNOWLEDGEMENTS

At first, I would like to thank my supervisors Peter Kappeler and Elise Huchard, without who I would not write this thesis today. Peter Kappeler for kindly enabling me to join his research group and work in Madagascar and for his support of this particular project. To Elise Huchard for guiding my project and for teaching me everything I didn't know about science, lab and field work and for being extremely engaged, kind and motivated supervisor. I would also like to thank Eckhard Heymann for his engagement in progress of this project and the rest of the committee, Christian Roos, Julia Ostner and Oliver Schülke, for evaluating my thesis. I would like to thank Claudia Fichtel for her overall help with my project.

I am especially grateful to Christina Glaschke for her massive help in the lab and with processing the samples, as well as for keeping the good spirits. I also thank to Nico Westphal for his help with 454-sequencing and to the whole Primate genetics lab for kindly welcoming me in their lab and for enabling me to work with their high-throughput sequencing platform. I wish to express special thanks to Markus Brameier for his highly efficient help with sequencing data processing.

I would like to thank my collaborators – Cornelia Kraus, Anni Hämäläinen and Josue H. Rakotoniaina for their help in the field, processing the samples and overall friendly conversations. Josue for his help with the parasitological analysis and Anni for her overall help with my project and for being great support in the field. Special thanks to Melanie Dammhahn for her massive help with everything, especially with the world's smallest primate, and for her extremely good spirit and pleasant company in the field and in Göttingen. I also thank Janina Demeler for her expertise in parasitological analysis.

I wish to thank my family for their endless support throughout my studies and especially to Jan Kaesler for being here the whole time for and with me. The biggest thanks to my cat and dogs, namely Ró, Fifinka, Köty, Aninka, Veny and Zrzunda, who always helped to keep my spirits up throughout the whole time.

I thank all people who have contributed to the collection and processing of DNA samples throughout the time and to Susanne Schliehe-Diecks for being Susanne. A special thanks goes to all Kirindy field assistants, notably Bruno Tsiverimana, Tiana Andrianjanahary, Jean-Claude de Beroboka and Mamy Razafindrasamba for their help

with captures, to Leon Razafimanantsoa and Rodin Rasoloarison for their logistic and administrative support.

Last, but not least, I wish to thank all my colleagues at the department for creating pleasant social and working environment. Foremost, I wish to thank my office mate Omer Nevo for his empathy and for providing great insights to my project. In addition, I would like to thank Henning Lahmann, Ulrike Walbaum, Anja Cziommer and Anja Engelke for providing administrative and technical support.

# CURRICULUM VITAE

## PERSONAL INFORMATION

Name: EVA KAESLER, born PECHOUŠKOVÁ  
Date of birth: 05. December, 1983  
Place of birth: Praha, Czech Republic  
Citizenship: Czech

## EDUCATION

2011 – **PhD candidate in Biology,**  
Leibniz-institute of Primate Research,  
Deutsches Primatenzentrum (DPZ) in Göttingen, Germany  
Thesis: Evolutionary mechanisms shaping MHC variation  
in sympatric lemurs.  
With Prof. Dr. Peter Kappeler, Dr. Elise Huchard  
Funded by: SAW project of Leibniz Graduate School,  
Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz  
e.V.

2006 – 2009 **Master of Science in Zoology, Ecology & Ethology,**  
Charles University of Prague, Czech Republic  
Thesis: Field study of microhabitat preferences and  
ecology of the hazel dormouse using nesting tubes.  
With Dr. Vladimír Vohralík

2003 – 2006 **Bachelor of Science in Biology,**  
Charles University of Prague  
Thesis: Field methods of research and ecology of hazel  
Dormouse.  
With Dr. Vladimír Vohralík

1997 – 2003 Gymnasium of Jan Neruda of music, specialisation: Oboe

## LIST OF PUBLICATIONS

**Pechouskova E**, Dammhahn M, Brameier M, Fichtel C, Kappeler P, Huchard E (2015) MHC class II variation in a rare and ecological specialist mouse lemur reveals lower allelic richness and contrasting selection patterns compared to a generalist and widespread sympatric congener. *Immunogenetics*: 67: 229-245.

Huchard E, **Pechouskova E** (2014) The major histocompatibility complex and primate behavioral ecology: New tools and future questions. *International Journal of Primatology*: 35 (1), 11-31.

Balounova Z, **Pechouskova E**, Rajchard J, Joza V, Sinko J (2013) World-wide distribution of the Bryozoan *Pectinatella magnifica* (Leidy 1851). *European Journal of Environmental Sciences*: 3(2), 96-100.



# DECLARATION

I hereby declare that I have written this thesis entitled “Evolutionary mechanisms shaping MHC variation in sympatric lemurs” independently and with no other aids or sources than quoted.

Eva Kaesler, born Pechoušková

In Göttingen, 19.10.2015