A peer-reviewed version of this preprint was published in PeerJ on 19 March 2015.

View the peer-reviewed version (peerj.com/articles/853), which is the preferred citable publication unless you specifically need to cite this preprint.

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 https://doi.org/10.7717/peerj.853

Species divergence, selection and polymorphism in the MHC of crows

The relatively high level of polymorphism usually found in the major histocompatibility complex (MHC) is usually attributed to pathogen mediated selection; however, the timescale of selection is often unclear. Here we describe the MHC class II (IIB) in three passerine bird species in the genus Corvus: American, carrion and jungle crows. Carrion and American crows are recently diverged, but allopatric, sister species, whereas carrion and jungle crows are more distantly related but sympatric, likely sharing pathogens linked to MHC IIB polymorphisms. These patterns of evolutionary divergence and current geographic ranges enabled us to examine evidence for ancient versus recent selection, trans-species polymorphism, and convergent evolution of the MHC in closely related species. Among the three species, the MHC IIB genes were highly duplicated (7-20 variants per individual) and polymorphic (an average of 79 variants per species; N = 18/species). Phylogenetic reconstructions of MHC IIB revealed patterns that were inconsistent with the evolutionary histories of the species. Several well supported interspecific clusters were observed, indicating trans-species polymorphism within this genus. Clustering of positively selected amino acids by supertyping revealed a single supertype shared by only jungle and carrion crows, whereas all other supertypes were shared among the three species, a pattern consistent with convergent evolution.

PeerJ PrePrints

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

John A. Eimes, Department of Evolutionary Studies of Biosystems, Graduate University for Advanced Studies (SOKENDAI), Hayama, Japan

Andrea K. Townsend, Department of Biology, Hamilton College, Clinton, NY, USA

Irem Sepil, Department of Zoology, University of Oxford, Oxford, UK

Isao Nishiumi , Department of Zoology, National Museum of Nature and Science, Tsukuba, Japan

Yoko Satta, Department of Evolutionary Studies of Biosystems, Graduate University for Advanced Studies (SOKENDAI), Hayama, Japan

Corresponding author:
John A. Eimes
Department of Evolutionary Studies of Biosystems
Graduate University for Advanced Studies
Shonan Village, Hayama, 240-0115, Japan
Fax: 81-46-858-1542
Email: johneimes12@gmail.com

PeerJ PrePrints

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

Introduction

The major histocompatibility complex (MHC) is an unusual example of a functional gene complex that exhibits high levels of polymorphism. The MHC is a multigene cluster that encodes molecules that bind and present peptides to T-cells in vertebrates, initiating a cascade of immunological responses to pathogens (Janeway *et al.* 2005). The MHC is the most polymorphic coding gene family in vertebrate genomes (Klein 1986) and the maintenance of this polymorphism is usually attributed to pathogen-mediated balancing selection (Sommer 2005), although other factors such as gene conversion, recombination and sexual selection may play an important role (Andersson & Mikko 1995; Martinsohn *et al.* 1999; Zelano & Edwards 2002).

Many studies in a wide variety of taxa have shown a strong signature of selection at the peptide binding region of the MHC; however, positive selection is assumed to be ancient (Bernatchez & Landry 2003). Indeed, MHC allelic lineages are often maintained over macro-evolutionary time scales, and this phenomenon, known as trans-species polymorphism, has been well documented, particularly in primates (Klein *et al.* 1993) and fishes (Garrigan & Hedrick 2003). Recently, a great deal of effort has been directed toward describing how balancing selection acts to shape MHC variation on more recent evolutionary time scales by either comparing closely related species or isolated conspecific populations that differ in pathogen regimes (Spurgin & Richardson 2010). The aims of these studies are often to determine the specific kinds of balancing selection that are acting on the MHC of interest (i.e. rare allele vs. heterozygote advantage) (Penn *et al.* 2002; Seddon & Ellegren 2004) or to establish specific pathogen-host MHC genotype interactions (Bonneaud *et al.* 2006; Westerdahl *et al.* 2005). Despite a few studies that detected recent selection on the MHC, notably pollution-driven selection on MHC IIB in estuarine fishes (Cohen 2002) and specific genotype resistance to

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

24 chitrid fungus in frogs (Savage & Zamudio 2011), there is little direct evidence for contemporary
25 pathogen-mediated selection in wild populations. Furthermore, as MHC repertoires change over
26 time, pathogens are predicted to co-evolve, making associations difficult to establish. In
27 addition, different MHC alleles may confer resistance to the same or similar pathogen, as
28 demonstrated by MHC supertyping in humans and other species (Sidney *et al.* 2008), further
29 complicating MHC genotype-pathogen associations. Thus, the general timescale within which
30 selection on MHC molecules can be detected remains unclear.

31 The MHC of the order Passeriformes (perching songbirds) is among the most complex among
32 terrestrial vertebrates, exhibiting highly duplicated classical MHC class I and II genes. This
33 large-scale gene duplication has resulted in more than 10 functional class II B loci reported in
34 some species, which contrasts sharply with the rather condensed "minimal MHC" observed in
35 other bird taxa such as Galliformes (Balakrishnan *et al.* 2010; Bollmer *et al.* 2010; Eimes *et al.*
36 2012; Sepil *et al.* 2012). It is likely that the extant genetic patterns of the MHC in songbirds are
37 the result of a combination of gene duplication, subsequent recombination (including gene
38 conversion), and balancing selection (Hess & Edwards 2002). Because of the large number of
39 potentially recombining loci, the role of recombination, especially gene conversion, may be
40 exaggerated in passerines relative to other vertebrates in shaping the MHC genetic repertoire.
41 The best evidence for this is presented in phylogenetic trees of songbirds among which multi-
42 locus MHC sequences cluster by species, reflecting the species rather than gene phylogeny (e.g.
43 Sutton *et al.* 2013). Models that detect recombination are most effective in evaluating
44 homologous loci (Bruen *et al.* 2006), and estimates of recombination rates of highly duplicated
45 loci, especially in a system where the genetic architecture is not well described and the
46 phylogenetic relationships of genes are unknown (such as songbirds), are suspect. In general,

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

47 however, the extent to which the MHC repertoire in a given passerine species is the result of
48 neutral molecular events, such as gene conversion and recombination, or pathogen-mediated
49 selection, is unclear.

50 In this study, we assessed how ancestry, divergence time and selection shape MHC
51 polymorphisms in three species of crows: jungle crow (*C. macrorhynchos japonensis*), eastern
52 carrion crow (*C. corone orientalis*), and American crow (*C. brachyrhynchos*). Crows are a
53 suitable group of closely related species for assessing recent balancing selection because their
54 habitat, diet and social behavior could promote pathogen transmission, and, therefore, pathogen-
55 mediated selection. Crows are omnivores that forage, in part, on carrion and human refuse
56 (McGowan 2001), and scavengers, in general, could be under selection for more robust immune
57 responses because of contact with pathogen-rich carcasses (Blount 2003). Furthermore, all three
58 of the *Corvus* species investigated in this study can be found in densely human-populated areas
59 (Ali & Ripley 1972; Kurosawa *et al.* 2003; McGowan 2001; Richner 1989), and some data
60 suggest that species with the ability to exploit urban areas have disproportionately strong
61 immune responses (Moller 2009). Finally, most crows are highly gregarious, often foraging and
62 roosting in large communal flocks in the non-breeding season (Madge & Burn 1993; Verbeek &
63 Caffrey 2002), which could promote horizontal transfer among conspecifics (Bull 1994; Frank
64 1996; Møller *et al.* 2001).

65 Jungle crows, eastern carrion crows, and American crows have contrasting phylogenetic
66 relationships and geographic ranges. The contemporary range of eastern carrion and jungle
67 crows overlaps in East Asia (southeastern China, the Korean peninsula and the Japanese
68 archipelago) (Haring *et al.* 2012); American crows are limited to North America (Madge &
69 Burn 1993). Mitochondrial DNA phylogenies, however, place carrion and American crows in a

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

70 monophyletic group, with jungle crows paraphyletic to these two species (Haring *et al.* 2012).
71 The genus *Corvus* likely has a Palearctic origin, and *mt*DNA phylogenies suggest that carrion
72 and American crows diverged in the Nearctic relatively recently while jungle crows represent a
73 single, older, lineage originating in the south-east Asian tropics (Haring *et al.* 2012).

74 In this analysis, we make the assumption that the two sympatric species (the carrion and jungle
75 crows), which have similar habitats and co-occur throughout east Asia, share a more similar suite
76 of contemporary pathogens with each other than with the geographically distant American crow.
77 Diverse pathogen burdens are well-documented in all three species, but a comparative study has
78 only been done for jungle and carrion crows, where 10 specific helminth species were shared by
79 jungle and carrion crows (Miller *et al.* 2010; Mizuno 1984; Wheeler *et al.* 2014). There have
80 been several surveys of internal parasites in American crows, and no cases of any of these
81 specific species of parasites were reported (Hendricks *et al.* 1969; Jones 1968; Miller *et al.* 2010;
82 Naderman & Pence 1980). We could find only a single case of any of these specific helminths in
83 North America (Cawthorn *et al.* 1980). A strong relationship between different characteristics
84 of MHC class II and resistance to helminth parasites is well documented (Goüy de Bellocq *et al.*
85 2008; Zhang & He 2013) In particular, allele specificity and both a positive and negative
86 relationship to helminth parasite load has been shown in rodents (Froeschke & Sommer 2005;
87 2012). Thus, similar selective pressures may have acted on the MHC IIB of carrion and jungle
88 crows.

89 The primary aim of this study was to differentiate between ancient selection on the MHC IIB,
90 occurring before the split of the jungle crows and the carrion/American crow clades, and
91 relatively recent selection that occurred since both that divergence as well as the more recent
92 divergence of American and carrion crows. We define ancient selection as that occurring before

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

93 speciation and recent selection as that occurring after speciation. In order to assess the timescale
94 of any detectable selection, we constructed two independent phylogenies from neutral markers
95 and estimated species divergence time. Next, we analyzed data for all three species from 454
96 pyrosequencing of the MHC IIB exon 2, using both nucleotide and amino acid alignments as
97 well as supertyping, which clusters functionally equivalent amino acid sequences into units of
98 selection (Lund *et al.* 2004; Sette & Sidney 1999; Sidney *et al.* 2008). Finally we attempted to
99 detect recent pathogen-mediated selection by comparing sequences between more recently
100 diverged allopatric species (carrion and American crows) and more distantly related sympatric
101 species (carrion and jungle crows).

102 The phylogenetic and geographic relationships of the three species allowed us to test two
103 different hypotheses regarding the timescale and intensity required to detect selection: The first
104 hypothesis is that any detectable selection is ancient and that insufficient time has passed since
105 the divergence of crows for any new patterns within the MHC sequences to have emerged
106 between or among the three species. In this scenario, the MHC phylogeny should roughly mirror
107 phylogenies constructed from neutral markers. The second hypothesis is that strong pathogen-
108 mediated selection is detectable, and that convergent evolution acted to drive MHC similarity
109 among the sympatric jungle and carrion crows, which are known to share more similar pathogen
110 relative to American crows. In this scenario we predicted that the jungle and carrion crows
111 MHC IIB variants would cluster in phylogenies more often that the more closely related
112 American and carrion crow clusters or that jungle and carrion crows would share more
113 supertypes than American and carrion crows.

114

115 **Materials and Methods**

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

116 *Sample collection*

117 We analyzed the data from 18 individuals of each species. American crows were sampled from
118 nestlings (one bird per nest) in Yolo County, California (38°32′ N, 121°45′ W) on the campus of
119 the University of California, Davis, in May and June 2012. Collection methods were approved
120 by the Institutional Animal Care and Use Committee of the University of California, Davis
121 (Permit Number: 16897). Approximately 50 μL of whole blood was preserved in Queens's lysis
122 buffer, and gDNA was extracted using a standard phenol/chloroform protocol followed by
123 ethanol precipitation. Previously extracted gDNA from the blood of jungle and carrion crows
124 was obtained from the Japanese National Museum of Nature and Science (Tsukuba, Japan) and
125 the Yamashina Institute of Ornithology (Chiba, Japan). These samples were collected from
126 1994-2010 from the entire range of the Japanese archipelago, although most jungle crows were
127 from the Tokyo area (16 from Tokyo, 1 each from Hokkaido and Okinawa), whereas carrion
128 crow collection was more evenly distributed across the Japanese archipelago (6 from
129 Kagoshima, 1 from Tsushima, 4 from Nara, 1 from Osaka and 6 from Tokyo; Fig S1). One *m*L
130 of whole blood for RNA extraction was collected from three juvenile jungle crows trapped at
131 Ueno Zoo in Tokyo, Japan in 2012. The blood was flash frozen on site and stored at -80 °C.
132 Jungle crow RNA was extracted from whole blood using the RNeasy Protect Animal Blood
133 System (Qiagen).

134 *Amplification of MHC class IIB exon 2*

135 In order to isolate MHC IIB exon 2, which codes for the PBR, jungle crow cDNA from two
136 individuals (U2 and U3) was synthesized with the LongRange 2Step RT-PCR System (Qiagen)
137 using the manufacturer's recommended amplification protocol. MHC IIB was then targeted in
138 the cDNA using the primers MHC05 (Miller & Lambert 2004), which anneals within exon 1 in

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

139 several passerines (positions variable), and *Coma*IIbex3R, which was designed by aligning the
140 conserved regions of IIB exon 3 from several passerines. These primers amplified a 359 bp
141 cDNA fragment. We verified that the amplicon was MHCB IIB by cloning the fragment from
142 both individuals (TOPO XL One Shot-Invitrogen). Twenty-four clones from each individual
143 were selected and whole colony PCR was performed using the TOPO vector primers M13 F and
144 M13 R. The 20 μL PCR contained a pluck of colony cells as template, 0.625 μM of each
145 primer, 0.5 *m*M of each dNTP, 1.5 *m*M MgCl2, 1X PCR buffer and 0.5 U of ExTaq (Takara)
146 polymerase. The cycling conditions included an initial denaturation at 94 °C for 2 min followed
147 by 30 cycles of 94, 55 and 72 °C each for 1 min and a final extension of 72 °C for 10 min. We
148 sequenced each amplicon in both directions using M13 primers (Applied Biosystems 31310xl).
149 Using BLAST, we identified eight sequences that aligned with known passerine MHC IIB.

150 We targeted intron 1 by aligning the cDNA-derived amplicons using MUSCLE in Geneious
151 6.1.4 (Biomatters) and a reverse primer, *Coma*IIbex2R, was designed in a conserved region of
152 exon 2 and paired with MHC05. Using gDNA from two jungle crows (*Coma*5061 and
153 *Coma*7509), a 516 bp fragment extending from exon 2 to exon 1 was amplified using 50-100 ng
154 of gDNA, 1.0 μM of each primer, 0.5 *m*M of each dNTP, 1.5 *m*M MgCl2, 1X PCR buffer and
155 0.5 U of ExTaq (Takara) polymerase (25 μL total volume). The cycling conditions included an
156 initial denaturation at 94 °C for 30s followed by 30 cycles of 94 (20s), 60 (20s) and 72 (45s) °C
157 and a final extension of 72 °C for 10 min. Each of these two amplicons was cloned and 10
158 colonies were sequenced in each direction using the cloning and whole colony PCR protocol
159 described above. Sequences were aligned using MUSCLE and a forward primer flanking exon
160 2, *Coma*iF2, was designed for the (100%) conserved region at position 282-304 of intron 1.
161 Next, we isolated intron 2 by pairing *Coma*iF2 with *Coma*IIBex3R using the same individuals

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

162 and PCR, cloning, and whole colony PCR conditions described above. When amplifying from
163 exon 2 to exon 3 in jungle crows, we discovered indels that resulted in large gaps in different
164 locations in intron 2. 454 pyrosequencing is most efficient when sequences are of equal length,
165 and for this reason, we designed a single degenerate reverse primer, *ComaEx2RA*, from positions
166 249-268 of exon 2 by amplifying and aligning sequences from exon 2 to exon 3 in two individual
167 of each species. Finally, we confirmed that the primers *Coma*iF2 and *ComaEx2RA* amplified
168 MHC IIB loci in all three crow species. Primer sequences and annealing temperatures are listed
169 in Table S1.

170 *454 pyrosequencing*

171 Fusion primers were synthesized (Fasmac-Japan) by ligating the standard Roche multiplex
172 identifiers (MIDs) to both the forward and reverse gene specific primers and the 454 adaptor
173 primers (the Roche "Basic Amplicon" design). In order to minimize the formation of PCR
174 artifacts, we reduced the cycle number and primer concentration and eliminated the final
175 extension step during amplicon generation (Lenz & Becker 2008). PCRs contained 25-50 *n*gs of
176 template, 0.5 μM of each fusion primer, 0.5 *m*M of each dNTP, 1.5 *m*M MgCl2, 1X PCR buffer,
177 5% DMSO and 0.5 U of ExTaq (Takara) polymerase (25 μL total volume). The cycling
178 conditions included an initial denaturation at 94 °C for 30s followed by 28 cycles of 94 (20s), 60
179 (20s) and 72 (45s) °C.

180 PCR amplicons were purified using the Agencourt AMPure XP System (Beckman Coulter)
181 and recovered DNA was checked for quality and primer dimers on a 1.5% agarose gel. Samples
182 were then quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) and pooled in
183 equimolar amounts according to the Roche GS Junior Titanium Amplicon Library Preparation
184 Method Manual. The quality and concentration of the pooled samples was checked on a 2100

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

185 Bioanalyzer (Agilent Technologies) and then sequenced at the Japanese National Museum of
186 Nature and Science in Tsukuba, Japan.

187 Amplicons were bi-directionally sequenced, and reads that passed the initial 454 Roche Junior
188 quality filter were de-multiplexed using jMHC (Stuglik *et al.* 2011). jMHC is especially
189 efficient for de-multiplexing bi-directionally sequenced double-tagged amplicons because it bins
190 only those sequences that contain both forward and reverse primers and their associated MIDs
191 with no ambiguous characters ("N"s) in the primers, MIDs or target sequence. Sorted FASTA
192 files (excluding primers and MIDs) were then imported into Geneious and aligned with
193 published MHC class IIB sequences.

194 *Bioinformatics variant validation and genotyping*

195 Variants were accepted as "true" if they occurred in at least three reads ($r > 2$) in both of two
196 independent PCRs of the same individual. The criteria of ($r > 2$) is derived from the (Galan *et*
197 *al.* 2010) estimate of a read containing at least one sequence error for a 171 bp fragment of
198 (HLA) DRB exon 2 of ~0.11. Thus, the probability that an identical error occurs three times in
199 the same sample is $\sim 10^{-8}$ (Galan *et al.* 2010). The most likely types of artefacts remaining in the
200 data set at this stage are either single base substitutions or fragment chimeras generated during
201 PCR. Although both of these artifacts are randomly generated, if these errors occur early in the
202 PCR, they can be replicated and represented at relatively high frequencies in individual
203 amplicons. To further reduce the probability of errantly including single base substitution
204 errors, we only classified sequences as true variants if they differed by at least two nucleotide
205 substitutions. Galan et al. 2010 reported an average chimera frequency of 0.06 and some
206 chimeras occurred at frequencies > 0.1; thus, elimination of chimeras by using frequency
207 thresholds may errantly inflate variant estimates. It is highly unlikely that two independent

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

208 PCRs each generated identical chimeras (where as many as 10 different variants can recombine)
209 or single errant substitutions three times ($r > 2$). By validating variants with a second PCR for
210 each individual (rather than a second PCR within or across individuals) chimeras are eliminated,
211 yet variants that are rare in the population are not errantly eliminated from the data set. The
212 total number of reads that were included at this stage was then designated as "net reads" for the
213 purposes of genotype confidence level calculations.

214 We calculated the 99.9% confidence level for the genotyped individuals that each variant was
215 sampled three times using the program "negative multinomial" (Galan et al. 2010). While the
216 program is limited to eight possible variants, the relationship between n, net reads, r, the number
217 of times each variant must be sampled, and m, the maximum number of variants per amplicon, is
218 linear. Our cloned sequences indicated a maximum of 20 variants per individual which
219 corresponds to a minimum of 175 reads for 99.9% confidence; however, this assumes equal
220 sampling of variants. Because amplification bias is likely when co-amplifying 10 loci, we
221 expected this bias to be reflected in over- and under-representation of some variants. Therefore,
222 we increased the minimum required net reads for genotyping to 275, which represents a
223 frequency of 1.1% (amplification bias of ~ one order of magnitude) for a variant that has exactly
224 three copies in an amplicon represented by 275 net reads.

225 *Phylogenetics and tests for selection*

226 For each species, we generated an alignment of validated variants and identified and aligned
227 codons that have been shown to comprise the PBR of HLA in humans (Brown *et al.* 1993).
228 While most MHC studies rely on the HLA PBR codons identified in humans to estimate
229 selection, it has not been shown that these same codons positions apply universally or that
230 selection is limited to peptide binding codons. Therefore, in order to identify specific codons

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

231 that are likely under selection in crows, we performed a Wu-Kabat analysis on the amino acid
232 alignment of verified alleles in all three species. A Wu-Kabat plot predicts amino acids that are
233 likely subject to selection by identifying positions of high variability. Variability is calculated by
234 dividing the number of amino acids at a position by the frequency of the most common amino
235 acid (Wu & Kabat 1970). For each species, we identified codons/amino acids as likely under
236 selection if they fit two criteria: 1, the Wu-Kabat variability metric was ≥ twice the mean of all
237 Wu-Kabat values and, 2, if the site had ≥ 0.5 amino acid diversity. These sites are denoted WuK-
238 PS (PS = polymorphic site). We tested for historical, positive selection on the HLA PBR sites by
239 calculating the ratios of non-synonymous (*d*N) and synonymous mutations (*d*S) using a Z test
240 (modified Nei-Gojobori- Jukes Cantor correction) in MEGA v. 5.2 (Tamura *et al.* 2011). In
241 addition, individual codons were tested for selection using the HYPHY package in MEGA. This
242 method, known as a counting method, estimates the number of nonsynonymous and synonymous
243 mutations that have occurred at each codon by using a maximum likelihood reconstruction of the
244 ancestral state of each sequence. The test statistic *d*N - *d*S is used for detecting codons that have
245 undergone positive selection and for positive values the probability of rejecting the null
246 hypothesis of neutral evolution (*p*-value) is calculated (Pond & Frost 2005; Suzuki & Gojobori
247 1999). Significant *p* values are denoted as positively selected sites (PSS). This method is more
248 computationally compatible with large data sets than empirical or hierarchal Bayesian
249 approaches, yet has been shown to provide nearly identical results with a sufficient number of
250 sequences (Pond & Frost 2005).

251 We constructed nucleotide phylogenies using either synonymous or non-synonymous
252 substitutions, of the 248 bp exon 2 fragments, the PBR (HLA) and the WuK-PS codons. NJ
253 trees were constructed in MEGA. For each analysis, 24 different nucleotide substitution models

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

254 were tested in MEGA and the Jukes-Cantor (JC) Model + G (Gamma) + I (invariant sites) was
255 chosen based on the Bayesian Information Criterion (BIC) score.

256 The test of our hypothesis relies on the *mt*DNA phylogenetic relationships described in Haring
257 *et al*. (2012); however, only two of the 49 jungle and (eastern) carrion crows used in that study
258 were from Japan. To insure that the *mt*DNA phylogenies used in Haring *et al*. (2012) were
259 consistent with our populations, we tested a subset of carrion (N = 4), jungle (N = 6) and
260 American (N = 4) crows at the same *mt*DNA region using the primers CR-Cor+ and Phe-Cor-
261 (Table S1) to amplify the *mt*DNA CR region using the same PCR conditions as Haring et al.
262 (2012) followed by cloning and Sanger sequencing as described above. We aligned these
263 sequences with a subset of each species from those *mt*DNA sequences used in the original
264 phylogenetic construction by Haring *et al*. (2012).

265 If differences in MHC repertoire can be detected between the closely related American and
266 carrion crows, a calculation of the minimum time since species divergence may shed light on the
267 time scale necessary to shape MHC variation in recently diverged taxa. While generating a
268 highly accurate estimate of species divergence time in crows was not the aim of this paper, we
269 reasoned that at least two independent "rough" estimates that yield similar results (within 1 MY)
270 may be informative. We applied three methods to estimate species divergence time. First, we
271 aligned 2831 bp of concatenated sequences from five nuclear introns. Primers for four of these
272 fragments were designed by aligning at least six sequences of each intron from *Corvis cornix*
273 sequences downloaded from NCBI (Table S2). These introns were selected based on size (500-
274 650 bp) and suitability of primer design in conserved regions. The SPINZ intron was amplified
275 using primers designed to amplify this fragment in *Gallus sp.* (Sawai *et al.* 2010). Two
276 randomly chosen individuals from each species were included in this analysis. These fragments

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

277 were amplified using the standard PCR conditions described above and the primers listed in
278 Table S1. The mutation rate for nuclear genes was previous calculated for Galliformes as $1.17 \times$
279 10^{-9} per site per year (0.17%/MY) (Sawai et al. 2010), which is very similar to that of primates
280 (Takahata & Satta 1997). We used this rate combined with the calculated distance, *d*, from the
281 intronic sequences to estimate species divergence time of the three species of crows using a NJ
282 tree and the Jukes Cantor method for evolutionary distance in MEGA.

283 Gene divergences often predate species divergence and thus overestimate the true divergence
284 between closely related species (Klein *et al.* 1998; Nei & LI 1979). To correct for this, ancestral
285 polymorphisms were removed from pairwise comparisons by subtracting the average of the
286 intraspecific pairwise differences from the observed interspecific value to calculate the net
287 nucleotide divergence, $d_N = dXY - (dX + dY)/2$ (Arbogast *et al.* 2002; Nei & LI 1979). We
288 estimated net evolutionary divergence between each of the three species using the JC model
289 (with 500 bootstrap replicates) and equated this value, d_N, to $2\mu t$, where μ is the mutation rate
290 and t is time of species divergence (Arbogast et al. 2002).

291 The third estimate of species divergence time was based on the same mitochondrial sequences
292 used to generate the phylogeny described above. Nabholz *et al.* (2009) estimated the *mt*DNA
293 mutation rate in three species of *Corvus*: *C. coronoides, C. frugilegis and C. macrorhynchos.*
294 We used the mean of these estimates (0.058 (± 0.002) per site per million years) to calculate
295 divergence time. We used a NJ tree and the Jukes Cantor method for evolutionary distance in
296 MEGA.

297 Supertyping

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

298 Functionally equivalent MHC IIB alleles were clustered using the same methods of Sepil et al.
299 (2012), a study that also examined highly duplicated MHC loci in a passerine species. Briefly,
300 amino acid sites identified as positively selected (PSS) were aligned and characterized by five
301 physicochemical descriptor variables: z1 (hydrophobicity),z2 (steric bulk), z3 (polarity), z4 and
302 z5 (electronic effects). These descriptor variables were placed into a matrix and subjected to K-
303 means clustering algorithm (Doytchinova & Flower 2005; Sandberg *et al.* 1998). Discriminant
304 analysis of principle components (DAPC) was used for describing clusters in the 'adegenet'
305 package in R (Jombart *et al.* 2010).

306 **Results**

307 Pyrosequencing

308 The goal of the first 454 run was to estimate the number of IIB loci and the minimum read
309 number required for accurate genotyping (99.9% confidence). Six individuals from each species
310 were sequenced and a total of 125,839 reads passed the initial 454 filter. After excluding reads of
311 less than 270 bp there were 76,891 reads with an average length of 340 nucleotides. Amplicons
312 had between 8,384 and 483 initial reads (mean = 1,213) that were sorted by jMHC. These
313 samples had an average of 376 net reads (range 171-3,211). The maximum number of variants
314 (differing by at least 2 nucleotides) found in a single individual was 20, which suggests there are
315 10 IIB loci in crows. This value was used to calculate the minimum net read number necessary
316 for 99.9% genotyping confidence in the program "Negative Multinomial". The calculated
317 minimum read number was 211; however, we increased the cutoff to 275 net reads to at least
318 partially account for amplification bias (see methods).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

319 For the second 454 run, a total of 107,279 reads passed the initial 454 filter. After excluding
320 reads of less than 270 bp, there were 90203 reads with an average length of 340 nucleotides. A
321 total of 18 individuals (with replicates) from each species met the criteria of at least 275 net
322 reads. All DNA sequences were deposited in NCBI Genbank: Accession numbers: MHC IIB:
323 TBA; crow *mt*DNA: KM246294-308; crow nuclear intronic: KM246309-23

324 *MHCIIB variation and selection*

325 The number of validated MHC IIB nucleotide variants for each species was: jungle crows = 89,
326 carrion crows = 81 and American crows = 67. Individuals had a range of 7-20 variants,
327 indicating that the MHC IIB in crows exhibits copy number variation (Table 1). Few variants
328 were shared among species and there was no discernable pattern of allele sharing relating to
329 phylogeny or sympatry: American and jungle crows shared four alleles, American and carrion
330 crows shared three alleles, carrion and jungle crows shared three alleles. Four alleles were
331 shared among all species.

332 A codon Z test for positive selection (non-synonymous/synonymous mutations) averaged
333 across PBR yielded a significant *p* value for American and carrion crows but not for jungle
334 crows (Table 1). The Wu-Kabat plot of amino acid variability showed that several highly
335 variable amino acid positions in exon 2 are not known peptide binding sites in HLA; for all three
336 species, 15 of 22 (68%) HLA PBR sites overlapped with WuK-PS sites. Among the three
337 species, jungle crows exhibited the most amino acid variation, with six hypervariable (>20 Wu-
338 Kabat index) amino acid sites in the Wu-Kabat plot (Fig.1). The HYPHY test for selection
339 identified nine positively selected sites that were shared among the three species, but only 6 of
340 these corresponded to HLA PBR (Fig. 1). This result suggests that the PBR of crows is not

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

341 identical to that identified for HLA or that sites other than PBR are under strong selection in
342 crows.

343 In total, we recovered 172 unique amino acid sequences among the three species (70 jungle,
344 58 carrion and 44 American). We detected no species-specific motifs or motifs that were shared
345 exclusively between two species. Within-species amino acid diversity was relatively high (0.245
346 ± 1.5), similar to that which we calculated for scrub jays (*Aphelocoma coerulescens*) (23.3 %; N
347 = 11; NCBI (U23958-65; 72,73,75). An alignment (82 sites) of common amino acid fragments
348 from the three crow species with four other passerines (including scrub jays) and a falconiform
349 species highlights the high amino acid diversity found among and between species of passerines
350 (Fig. S2).

351 *Species divergence*

352 *mtDNA phylogeny and divergence estimate*

353 The CR-Cor+ and Phe-Cor- primers amplified an 836-920 bp fragment of the *mt*DNA CR region
354 in the three species of crows. These sequences were aligned with known crow CR sequences. A
355 NJ tree placed carrion and American crows in a separate clade from jungle crows with 100%
356 bootstrap support (Fig. 2), thus confirming that the jungle and carrion crow populations from this
357 study conform to the phylogeny of Haring *et al*. (2012). The *mt*DNA mutation rate of *Corvus*
358 was estimated at 0.058 ± 0.002 per site per million years, a rate at the higher end of the range
359 within passerines where a wide range of rates were reported , and which is not calibrated with
360 passerine fossil evidence (Nabholz *et al.* 2009). Nevertheless, if this value is used in the *mt*DNA
361 phylogeny of the crows, a linearized NJ tree estimates a divergence time of approximately 0.9
362 MY between jungle crows and the common ancestor to carrion and American crows and 0.45
363 MY for the split between American and carrion crows (Fig. 2).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

364 *Nuclear intron phylogeny and divergence estimate*

365 The phylogeny and estimated divergence of the concatenated 2831 bp fragment from five
366 nuclear introns is shown in Fig. S3. The range of pairwise distances, *d*, between species for the
367 five intron sequences individually was 0.002-0.006 and *d* for the entire concatenated fragment
368 was 0.04 (±0.001) (*d* within species was 0). Using the neutral nuclear mutation rate of 1.17 X
369 10^{-9} per site per year (estimated for *Gallus sp.,* Sawai *et al*. 2010), we calculated an approximate
370 divergence time between jungle crows and the branch point of the American and carrion crow
371 clade of 1.3 MY and a divergence of 0.52 MY for the spilt between American and carrion crows.
372 The estimates from the net nucleotide divergence method were similar: 0.427 ± 0.03 MY for the
373 split between carrion and American crows and 1.28 ± 0.30 MY for the divergence between
374 jungle crows and most recent common ancestor of the other two species. Both of these estimates
375 were consistent with the divergence time estimates provided by the *mt*DNA sequences.
376 However, unlike the *mt*DNA substitution rate, the nuclear gene mutation rate is not expected to
377 be highly variable between taxa, and is therefore probably more reliable. Using the divergence
378 time calculated with nuclear introns, we estimated a mitochondrial mutation rate in *Corvus* at
379 ~3.0 % per million years.

380 *MHC IIB phylogeny*

381 We observed the same general pattern in all of the phylogenies constructions (PBR, Wuk-PS,
382 exon 2): While there was a general trend for crow MHC IIB sequences to cluster by species, this
383 pattern was not as distinct as that seen in other passerine phylogenies (e.g. Sutton et al. 2013) as
384 there were many supported interspecific clusters (Fig. 3; Fig. S4-S6). While most supported
385 clusters were species specific, the trees revealed an unexpected pattern: the MHC IIB sequences
386 of the allopatric and more distantly related jungle and American crows were just as likely to

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

387 cluster together as any other species dyad in most of the phylogeny constructions and only
388 species specific clusters were more common among all clusters. This pattern was most obvious
389 in the tree constructed using non-synonymous substitutions in exon 2 (222-248 bp) (Fig. 3).

390 Supertyping

391 Among the three species, there were nine shared positively selected sites (PSS) in exon 2. Six of
392 these nine sites corresponded to HLA PBR sites and eight corresponded to Wuk-PS sites (Fig. 1).
393 From these, a total of eight supertypes, constructed from 140 unique PSS amino acid variants
394 were identified (Fig. 4). Only one supertype (Supertype 4) was limited to carrion and jungle
395 crows, whereas the rest of the supertypes were shared among the three species.

396 Discussion

397 We predicted that the MHC phylogeny would either mirror phylogenies constructed from neutral
398 markers, reflecting differentiation at the time of species divergence, or correspond, to some
399 degree, with current range overlap, reflecting convergent evolution between the sympatric jungle
400 and carrion crows. Neither of these predicted phylogenetic patterns were observed. If either
401 selection or divergence time were insufficient to shape MHC polymorphism in either a divergent
402 or convergent manner, then we would expect the sequences among the species to reflect a neutral
403 model-in which American and carrion crow MHC sequences should cluster together more often
404 than any other dyad. The mtDNA and nuclear intronic phylogenies show that sufficient time has
405 passed since species divergence for neutral mutations to establish well supported clades, yet
406 American and carrion crow MHC IIB sequences were actually less likely to cluster together than
407 those of the more distantly related, and allopatric, American and jungle crows. While MHC
408 sequences mostly clustered by species overall, in the phylogenies constructed from non-

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

409 synonymous substitutions (Fig. 3; Figs. S5-S6), inter- specific clustering was common. These
410 interspecific clusters may be the result of ancient selection, with these trans-species
411 polymorphisms remaining under constant selection into the present (Klein et al. 1998).

412 Analysis of the supertypes generated from the data did reveal a pattern consistent with
413 convergence evolution in carrion and jungle crows. While seven of the eight supertypes were
414 shared among the three species, one, supertype ST-4, was exclusive to carrion and jungle crows,
415 and could reflect selection pressure to a specific, shared parasite. None of the other seven
416 supertypes were exclusive to either a single species or group of two species. This supertyping
417 method classifies MHC II B loci into clusters based on the physio-chemical properties of the
418 amino acids in the positively selected sites of the PBR. Several studies of primate MHC have
419 shown that different MHC alleles bind very similar peptide motifs, suggesting significant overlap
420 in peptide binding repertories (Lund *et al.* 2004; Sette & Sidney 1999; Sidney *et al.* 2008). Thus,
421 different amino acid sequences can be functionally equivalent, and grouped together as a
422 supertype, can be considered a unit of selection. Recently, (Sepil *et al.* 2013) showed that
423 different MHC class I supertypes conferred resistance to malaria in great tits (*Parus major*).
424 Interestingly, in our study, supertypes ST-3 and ST-4 grouped together, away from the other six
425 supertypes, which may suggest that these groups are functionally important for a specific type of
426 parasite, whereas the others are combatting common, generalist parasites (Fig 4).

427 Our results, at both MHC and neutral loci, indicate that the *Corvus* genus is recently
428 diverged. Thus, while pathogen-mediated selection may be acting on MHC loci in crows,
429 phylogenetic methods of identifying it may be ineffective, as there may not be sufficient
430 evolutionary time to reshape ancient MHC lineages. Our divergence estimates suggest that there
431 has been at least 290,000 years since carrion and American crows diverged, but it is unclear how

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

432 long the species have been geographically isolated, allowing for selection from different
433 pathogens to differentiate the MHC. It is also unclear how long carrion and jungle crows have
434 shared a common range, and the extent to which they share common pathogens in East Asia is
435 unknown. Nevertheless, if pathogen-mediated selection is a powerful force in shaping MHC
436 variation, and this selection can be detected in recent time scales as some studies have found
437 (Spurgin & Richardson 2010), we expected to find some evidence of divergence between
438 American and carrion crows and possibly convergence in jungle and carrion crows at the MHC
439 IIB loci. While we did observe a single supertype exclusive to jungle and carrion crows, we
440 found no nucleotide or amino acids motifs that were either more common in the two sympatric
441 species or between the more closely related allopatric species that suggested either phylogenetic
442 ancestry or recent selection.

443 Perhaps our most interesting findings were the supported clusters that defied species ancestry
444 or included all three species. These clusters may a combination of incomplete lineage sorting
445 between these three closely related species (Pamilo & Nei 1988) and trans-species
446 polymorphisms that are maintained by balancing selection (Klein et al. 1998). Trans-species
447 polymorphism has been described in a host of vertebrate taxa, but many of these cases describe
448 identical, or near identical, alleles that are shared between species. Of the 239 unique sequences
449 of the MHC IIB among the three species of crow, only three identical variants were shared
450 between any two species. However, there were at 13 well supported interspecies clusters (> 70%
451 bootstrap support) of similar alleles (differing by up to 50% nucleotide identity) in the phylogeny
452 constructed using non-synonymous substitutions in exon 2 (Fig. 3). Thus, our example of
453 potential trans-species polymorphism describes similar, rather than identical, alleles shared
454 between species, which corresponds to the original description of trans-species polymorphisms

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

455 by Klein et al. (1998). Only a few studies have described trans-species polymorphism in birds,
456 and two distinct patterns can be seen. (Sato *et al.* 2011) described identical or nearly identical
457 alleles between recently diverged Darwin's Finches, while Bollmer et al. (2010), which analyzed
458 common yellowthroats (*Geothylypis trichas*), described a pattern more similar to the one shown
459 here, where well supported interspecific allele clusters had up to 0.05 nucleotide diversity.
460 Supported clusters in this study had up to 0.08 nucleotide diversity.

461 **Conclusion**

462 Given the relatively large number of MHC IIB loci that are commonly found in passerines, the
463 MHC-pathogen recognition system of this order may, counterintuitively, be less specific than
464 taxa that express few MHC loci. Thus, in systems with highly duplicated MHC, recent pathogen
465 mediates selection may be undetectable using traditional comparative analysis, such as fragment
466 alignment and phylogeny construction. The finding of a single supertype that supported
467 convergent evolution in jungle and carrion suggests that supertyping may be a more promising
468 method of detecting selection in systems with highly duplicated MHC regions. Our results
469 indicate trans-species polymorphisms in the MHC IIB are maintained by balancing selection in
470 recently diverged species of crows.

471 **Acknowledgements**

472 We wish to Naoyuki Takahata and Colm O'hUigín for advice during the experimental phase of the
473 project and Peter Dunn for advice on manuscript preparation.

474 **References**

475 Ali S, Ripley S (1972) *Handbook of the Birds of India and Pakistan* Oxford University Press, Bombay.
476 Andersson L, Mikko S (1995) Generation of MHC Class II Diversity by Intra- and Intergenic
477 Recombination. *Immunol Rev* **143**, 5-12.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

478 Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from
479 molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology*
480 *and Systematics* **33**, 707-740.
481 Balakrishnan C, Ekblom R, Volker M*, et al.* (2010) Gene duplication and fragmentation in the zebra finch
482 major histocompatibility complex. *BMC Biology* **8**, 29.
483 Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about
484 natural selection in 15 years? *J Evol Biol* **16**, 363-377.
485 Blount JD (2003) Do individual branches of immune defence correlate? A comparative case study of
486 scavenging and non-scavenging birds. *Oikos* **102**, 340-350.
487 Bollmer JL, Dunn PO, Whittingham LA, Wimpee C (2010) Extensive MHC Class II B Gene Duplication in a
488 Passerine, the Common Yellowthroat (Geothlypis trichas). *J Heredity* **101**, 448-460.
489 Bonneaud C, Chastel O, Federici P, Westerdahl H, Sorci G (2006) Complex Mhc-based mate choice in a
490 wild passerine. *Proceedings of the Royal Society B: Biological Sciences* **273**, 1111-1116.
491 Brown JH, Jardetzky TS, Gorga JC*, et al.* (1993) 3-dimensional structure of the human class-II
492 histocompatibility antigen Hla-Dr1. *Nature* **364**, 33-39.
493 Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of
494 recombination. *Genetics* **172**, 2665-2681.
495 Bull JJ (1994) Perspective - Virulence. *Evolution* **48**, 1423-1437.
496 Cawthorn R, Anderson R, Barker I (1980) Lesions caused by Diplotriaena tricuspis (Nematoda:
497 Diplotriaenoidea) in the American crow, Corvus brachyrhynchos Brehm. *Canadian Journal of*
498 *Zoology* **58**, 1892-1898.
499 Cohen S (2002) Strong Positive Selection and Habitat-Specific Amino Acid Substitution Patterns in Mhc
500 from an Estuarine Fish Under Intense Pollution Stress. *Molecular Biology and Evolution* **19**, 1870-
501 1880.
502 Doytchinova IA, Flower DR (2005) In Silico Identification of Supertypes for Class II MHCs. *Journal of*
503 *Immunology* **174**, 7085-7095.
504 Eimes JA, Reed KM, Mendoza KM*, et al.* (2012) Greater prairie-chickens have a compact MHC-B with a
505 single class 1A locus. *Genes & Genetic Systems* **87**, 435-435.
506 Frank SA (1996) Models of parasite virulence. *Quarterly Review of Biology* **71**, 37-78.
507 Froeschke G, Sommer S (2005) MHC Class II DRB Variability and Parasite Load in the Striped Mouse
508 (Rhabdomys pumilio) in the Southern Kalahari. *Mol Biol Evol* **22**, 1254-1259.
509 Froeschke G, Sommer S (2012) Insights into the complex associations between MHC class II DRB
510 polymorphism and multiple gastrointestinal parasite infestations in the striped mouse. *PLoS One*
511 **7**, e31820.
512 Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF (2010) A 454 multiplex sequencing method for
513 rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics*
514 **11**, 296.
515 Garrigan D, Hedrick PW (2003) Detecting adaptive molecular polymorphism: lessons from the MHC.
516 *Evolution* **57**, 1707-1722.
517 Goüy de Bellocq J, Charbonnel N, Morand S (2008) Coevolutionary relationship between helminth
518 diversity and MHC class II polymorphism in rodents. *J Evol Biol* **21**, 1144-1150.
519 Haring E, Daubl B, Pinsker W, Kryukov A, Gamauf A (2012) Genetic divergences and intraspecific
520 variation in corvids of the genus Corvus (Aves: Passeriformes: Corvidae) - a first survey based on
521 museum specimens. *Journal of Zoological Systematics and Evolutionary Research* **50**, 230-246.
522 Hendricks L, Harkema R, Miller G (1969) Helminths of the Crow, Corvus brachyrhynchos Brehm, 1822, in
523 North Carolina1. *Proceedings of the Helminthological Society of Washington* **36**, 150-152.
524 Hess C, Edwards S (2002) The evolution of the major histocompatibility complex in birds. *Bioscience* **52**,
525 423-431.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

526 Janeway C, Travers P, Walport M, Shlomchick M (2005) *Immunobiology: the immune system in health*
527 *and disease* Garland Press, New York.
528 Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for
529 the analysis of genetically structured populations. *Bmc Genetics* **11**.
530 Jones J (1968) Some Parasites of the Common Crow, Corvus Brachyrhynchos Brehm, from Ohio. *The*
531 *Ohio Journal of Science* **68**, 25-31.
532 Klein D, Ono H, O'HUigin C*, et al.* (1993) Extensive MHC variability in cichlid fishes of Lake Malawi.
533 *Nature* **364**, 330-334.
534 Klein J (1986) *The natural history of the major histocompatibility complex* John Wiley, New York.
535 Klein J, Sato A, Nagl S, O'hUigín C (1998) Molecular trans-species polymorphism. *Annual Review of*
536 *Ecology and Systematics* **29**, 1-21.
537 Kurosawa R, Kono R, Kondo T, Kanai Y (2003) Diet of jungle crows in an urban landscape. *Global*
538 *Environmental Research* **7**, 193-198.
539 Lenz TL, Becker S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping
540 of MHC and other highly polymorphic loci - Implications for evolutionary analysis. *Gene* **427**,
541 117-123.
542 Lund O, Nielsen M, Kesmir C*, et al.* (2004) Definition of supertypes for HLA molecules using clustering of
543 specificity matrices. *Immunogenetics* **55**, 797-810.
544 Madge S, Burn H (1993) *Crows and jays* Helm, London, UK.
545 Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC
546 evolution: a review. *Immunogenetics* **50**, 168-200.
547 McGowan K (2001) *Demographic and behavioral comparisons ofsuburban and rural American Crows*
548 Kluwer Academic Press, Norwell, MA.
549 Miller AD, Townsend AK, McGowan KJ*, et al.* (2010) Non-west nile virus-associated mortality in a
550 poulation of Amercian crows (*Corvus brachyrhynchos*): a gross and histopathologic study.
551 *Journal of Veterinary Diagnostic Investigation* **22**, 289-295.
552 Miller HC, Lambert DM (2004) Gene duplication and gene conversion in class II MHC genes of New
553 Zealand robins (Petroicidae). *Immunogenetics* **56**, 178-191.
554 Mizuno F (1984) Studies on the parasite fauna of the eastern carrion crow, Corvus corone orientalis
555 Eversman, and the Japanese jungle crow, Corvus macrorhynchos japonensis *Japanese Journal of*
556 *Veterinary Research* **32**, 105.
557 Moller AP (2009) Successful city dwellers: a comparative study of the ecological characteristics of urban
558 birds in the Western Palearctic. *Oecologia* **159**, 849-858.
559 Møller AP, Merino S, Brown CR, Robertson RJ (2001) Immune Defense and Host Sociality: A Comparative
560 Study of Swallows and Martins. *American Naturalist* **158**, 136-145.
561 Nabholz B, Glemin S, Galtier N (2009) The erratic mitochondrial clock: variations of mutation rate, not
562 population size, affect mtDNA diversity across birds and mammals. *Bmc Evolutionary Biology* **9**.
563 Naderman J, Pence D (1980) Helminths of the Common Crow, Corvus brachyrhynchos
564 Brehm, from West Texas. *Proceedings of the Helminthological Society of Washington* **47**, 100-105.
565 Nei M, LI W (1979) Mathematical model for studying genetic variation in terms of restriction
566 endonucleases. *Proceedings of the National Academy of Sciences USA* **76**, 5269-5273.
567 Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Mol Biol Evol* **5**, 568-583.
568 Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against
569 multiple-strain infections. *Proceedings of the National Academy of Sciences, USA* **99**, 11260-
570 11264.
571 Pond SLK, Frost SDW (2005) Not so different after all: A comparison of methods for detecting amino acid
572 sites under selection. *Molecular Biology and Evolution* **22**, 1208-1222.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

573 Richner H (1989) Habitat-Specific Growth and Fitness in Carrion Crows (Corvus-Corone-Corone). *Journal*
574 *of Animal Ecology* **58**, 427-440.
575 Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S (1998) New chemical descriptors relevant for the
576 design of biologically active peptides. A multivariate characterization of 87 amino acids. *Journal*
577 *of Medicinal Chemistry* **41**, 2481-2491.
578 Sato A, Tichy H, Grant PR*, et al.* (2011) Spectrum of MHC Class II Variability in Darwin's Finches and Their
579 Close Relatives. *Mol Biol Evol* **28**, 1943-1956.
580 Savage AE, Zamudio KR (2011) MHC genotypes associate with resistance to a frog-killing fungus.
581 *Proceedings of the National Academy of Sciences, U S A*.
582 Sawai H, Kim HL, Kuno K*, et al.* (2010) The Origin and Genetic Variation of Domestic Chickens with
583 Special Reference to Junglefowls Gallus g. gallus and G. varius. *PLoS One* **5**, -.
584 Seddon JM, Ellegren H (2004) A temporal analysis shows major histocompatibility complex loci in the
585 Scandinavian wolf population are consistent with neutral evolution. *Proceedings of the Royal*
586 *Society of London, B* **271**, 2283-2291.
587 Sepil I, Lachish S, Hinks AE, Sheldon BC (2013) Mhc supertypes confer both qualitative and quantitative
588 resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society*
589 *B-Biological Sciences* **280**.
590 Sepil I, Moghadam HK, Huchard E, Sheldon BC (2012) Characterization and 454 pyrosequencing of Major
591 Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system.
592 *Bmc Evolutionary Biology* **12**.
593 Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A
594 and -B polymorphism. *Immunogenetics* **50**, 201-212.
595 Sidney J, Peters B, Frahm N, Brander C, Sette A (2008) HLA class I supertypes: a revised and updated
596 classification. *BMC Immunol* **9**, 1.
597 Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and
598 conservation. *Frontiers in Zoology* **2**, 16.
599 Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and
600 misunderstandings. *Proceedings of the Royal Society B: Biological Sciences* **277**, 979-988.
601 Stuglik MT, Radwan J, Babik W (2011) jMHC: software assistant for multilocus genotyping of gene
602 families using next-generation amplicon sequencing. *Molecular Ecology Resources* **11**, 739-742.
603 Sutton JT, Robertson BC, Grueber CE, Stanton JAL, Jamieson IG (2013) Characterization of MHC class II B
604 polymorphism in bottlenecked New Zealand saddlebacks reveals low levels of genetic diversity.
605 *Immunogenetics* **65**, 619-633.
606 Suzuki Y, Gojobori T (1999) A method for detecting positive selection at single amino acid sites. *Mol Biol*
607 *Evol* **16**, 1315-1328.
608 Takahata N, Satta Y (1997) Evolution of the primate lineage leading to modern humans: Phylogenetic
609 and demographic inferences from DNA sequences. *Proceedings of the National Academy of*
610 *Sciences, U S A* **94**, 4811-4815.
611 Tamura K, Peterson D, Peterson N*, et al.* (2011) MEGA5: Molecular evolutionary genetics analysis using
612 likelihood, distance and parsimony methods. *Mol Biol Evol* **28**, 2731-2739.
613 Verbeek N, Caffrey C (2002) The American Crow (*Corvus brachyrhynchos*). In: *Birds of North America*
614 (eds. Poole A, Gill F). Birds of North America, Inc., Philadelphia.
615 Westerdahl H, Waldenström J, Hansson B*, et al.* (2005) Associations between malaria and MHC genes in
616 a migratory songbird. *Proceedings of the Royal Society B: Biological Sciences* **272**, 1511-1518.
617 Wheeler SS, Woods LW, Boyce WM*, et al.* (2014) West Nile virus and non-West Nile virus mortality and
618 coinfection of American crows (Corvus brachyrhynchos) in California. *Avian Dis* **58**, 255-261.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

619 Wu TT, Kabat EA (1970) An analysis of the sequences of the variable regions of Bence Jones proteins and
620 myeloma light chains and their implications for antibody complementarity. *Journal of*
621 *Experimental Medicine* **132**, 211-250.
622 Zelano B, Edwards SV (2002) An Mhc component to kin recognition and mate choice in birds:
623 predictions, progress, and prospects. *American Naturalist* **160**, S 225- 237.
624 Zhang M, He H (2013) Parasite-mediated selection of major histocompatibility complex variability in wild
625 brandt's voles (Lasiopodomys brandtii) from Inner Mongolia, China. *Bmc Evolutionary Biology*
626 **13**, 149.

627

PeerJ PrePrints

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

Tables and Figures

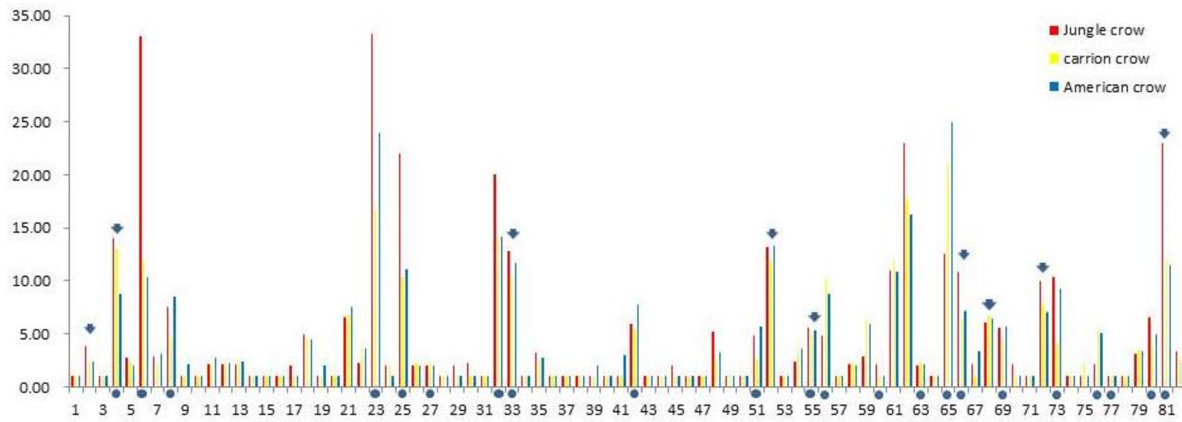
Table 1. MHC IIB variation and historical selection in three species of crows. N = 18 for each species. π = nucleotide diversity across 246 bp of exon 2, π_SPBR = synonymous mutation rate and π_NPBR = non-synonymous mutation rate at the HLA PBR identified by Brown et al. 1993, Z-test = significance (*p*) value using only HLA PBR codons.

Species	Alleles (total)	Alleles (Indiv)	π	π_SPBR	π_NPBR	dN/dS PBR	*p* value Z-test
Jungle	89	7-18	0.16	0.29 ± 0.115	0.47 ± 0.09	1.6	0.09
Carrion	81	11-20	0.15	0.24 ± 0.02	0.45 ± 0.1	1.9	0.028
American	67	11-18	0.14	0.25 ± 0.02	0.44 ± 0.09	1.6	0.047

PeerJ PrePrints

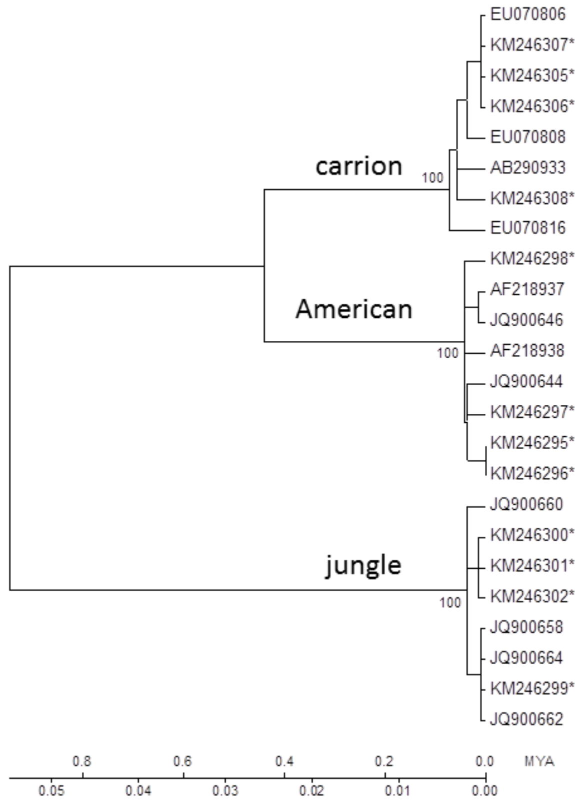
PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

Fig 1. Wu-Kabat plot of all amino acid sequences for three species of crows. Red: jungle crow; yellow: carrion crow; blue: American crow. Blue dots represent HLA PBR codons identified by Brown et al. 1993. Arrows identified positively selected sites (PSS) identified by HYFY. Variability is represented on the X axis and site position is on the y axis.



PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

Fig 2. A NJ tree (Jukes Cantor, 500 bootstrap replicates) using 836-920 bp fragments of the mtDNA control region of jungle, carrion and American crows. Accession numbers with an asterisk are from this study, all others are from Haring et al. 2012. The mtDNA mutation rate used to generate divergence time was 5.8 %/MY (Nabholz *et al*. 20009)



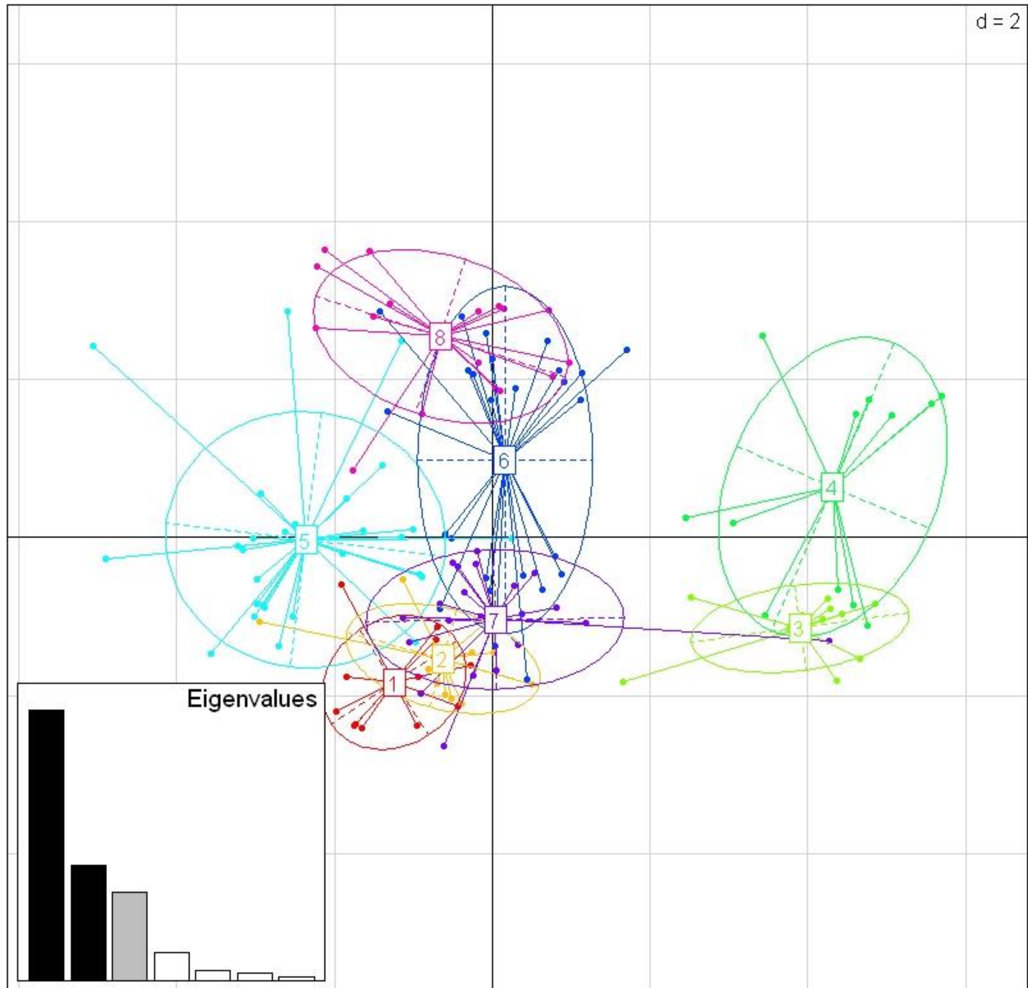
PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

PeerJ PrePrints

Fig 3. A NJ tree (500 bootstrap replicates, Jukes Cantor) of selected MHC IIB exon 2 variants (226-248 bp) using non-synonymous substitutions. Supported branches (Bootstrap value ≥ 0.5) of interspecies clusters for crows (Coma: *Corvus macrorhynchos*, Coco: *C. corone*, Cobr: *C. brachyrhynchos*) are denoted by horizontal red bars. Included are three other species of Passeriformes (Apco: *Aphelocoma coerulescens*, Getr: *Geothlypis trichas*, Tumi: *Turdus migratorius*) and one Falconiformes for an out group (Fati: *Falco tinnunculus*). Accession numbers for crows are [TBA]. Accession numbers for species are listed after label.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014

Fig. 4. DAPC scatterplot of the 8 Mhc supertypes. 12 PCs and three discriminant functions (dimensions) were retained during analyses, to describe the relationship between the clusters. The scatterplot show only the first two PCs (d = 2) of the DAPC of Mhc supertypes. The bottom graph illustrates the variation explained by the 12 PCs. Each allele is represented as a dot and the supertypes as ellipses.



PeerJ PrePrints

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v1 | CC-BY 4.0 Open Access | rec: 19 Nov 2014, publ: 19 Nov 2014