Patterns of Evolution of MHC Class II Genes of Crows (*Corvus*) Suggest Trans-species Polymorphism

John A Eimes, Andrea K Townsend, Irem Sepil, Isao Nishiumi, Yoko Satta

A distinguishing characteristic of genes that code for the major histocompatibility complex (MHC) is that alleles often share more similarity between, rather than within species. There are two likely mechanisms that can explain this pattern: convergent evolution and trans-species polymorphism (TSP), in which ancient allelic lineages are maintained by balancing selection and retained by descendant species. Distinguishing between these two mechanisms has major implications in how we view adaptation of immune genes. In this study we analyzed exon 2 of the MHC class IIB in three passerine bird species in the genus *Corvus*: jungle crows (*Corvus macrorhynchos japonensis*) American crows (*C. brachyrhynchos*) and carrion crows (*C. corone orientalis*). Carrion crows and American crows are recently diverged, but allopatric, sister species, whereas carrion crows and jungle crows are more distantly related but sympatric species, and possibly share pathogens linked to MHC IIB polymorphisms. These patterns of evolutionary divergence and current geographic ranges enabled us to test for trans-species polymorphism and convergent evolution of the MHC IIB in crows. Phylogenetic reconstructions of MHC IIB sequences revealed several well supported interspecific clusters containing all three species, and there was no biased clustering of variants among the sympatric carrion crows and jungle crows. The topologies of phylogenetic trees constructed from putatively selected sites were remarkably different than those constructed from putatively neutral sites. In addition, trees constructed non-synonymous substitutions from a continuous fragment of exon 2 had more, and generally more inclusive, supported interspecific MHC IIB variant clusters than those constructed from the same fragment using synonymous substitutions. These phylogenetic patterns suggest that recombination, especially gene conversion, has partially erased the signal of allelic ancestry in these species. While clustering of positively selected amino acids by supertyping revealed a single supertype shared by only jungle and carrion crows, a pattern consistent with convergence, the overall phylogenetic patterns we observed suggest that TSP, rather than convergence, explains the interspecific allelic similarity of MHC IIB genes in these species of crows.

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

1 Introduction

2   The major histocompatibility complex (MHC) is an unusual example of a functional gene

3 complex that exhibits high levels of polymorphism. The MHC is a multigene cluster that encodes

4 molecules that bind and present peptides to T-cells in vertebrates, initiating a cascade of

5 immunological responses to pathogens (Janeway et al. 2005). The MHC is the most polymorphic

6 coding gene family in vertebrate genomes (Klein 1986) and the maintenance of this

7 polymorphism is usually attributed to pathogen-mediated balancing selection (Sommer 2005),

8 although sexual selection, as well as molecular mechanisms such as recombination and gene

9 conversion, may play important roles (Andersson and Mikko 1995, Martinsohn et al. 1999,

10 Zelano and Edwards 2002).

11   One distinguishing characteristic of MHC genes is that alleles often share more similarity

12 between, rather than among species. Two mechanisms have been proposed to explain this

13 "trans-specific" similarity of MHC alleles. The first explanation, called "trans-species

14 polymorphism" (TSP), proposes that orthologous MHC allelic lineages are maintained by

15 balancing selection, often over macro-evolutionary time-scales, and persist through speciation

16 events (Klein et al. 1998). Strong evidence for TSP in MHC IIB genes has been found in a wide

2

17 range of taxa, particularly those of mammals and fish (Klein et al. 1993, Graser et al. 1996,

18 Garrigan and Hedrick 2003, Lenz et al. 2013).

19   An alternative explanation for trans-specific MHC allele clustering is convergent or parallel

20 evolution.   The term molecular convergent evolution has been used to describe similar, but

21 distinct phenomena (e.g. convergence between non-orthologous genes or within orthologues

22 only) (Zhang and Kumar 1997, Yeager and Hughes 1999, Kriener et al. 2000); and debate exists

23 for distinguishing molecular parallel evolution from molecular convergent evolution as well

24 (Arendt and Reznick 2008, Pearce 2011). For the purpose of this study, we define convergence

25 as any interspecific allele similarity that arises from adaptation to similar selective pressures

26 (Gustafsson and Andersson 1994, Hughes 1999).

27   A common method used for identifying TSP, and in some cases for distinguishing TSP from

28 convergence, is the comparison of codons thought to be under selection, such as those found in

29 the MHC peptide binding region (PBR), with synonymous substitutions in surrounding codons

30 or flanking introns. Alleles that are similar by decent should retain a signature of ancestry at

31 selectively neutral sites. On the other hand, convergence is indicated if alleles are more similar

32 at putatively selected sites than at neutral sites (Klein et al. 1998). Specifically, convergence and

33 TSP can be tested by constructing different phylogenetic trees and comparing clustering patterns

34 of trees constructed from non-synonymous mutations at codons likely to be under selection with

35 trees constructed from synonymous substitutions at putatively neutral sites (Graser et al. 1996,

36 Kupfermann et al. 1999, Kriener et al. 2000).

37   While many studies have provided evidence for TSP in MHC IIB allelic lineages, only a few

38 have provided evidence for convergence. One study of MHC DRB sequences in New World

3

39   monkeys found that PBR amino acid motifs were shared among species but some sequences of

40   the flanking introns sorted intra-specifically (Kriener et al. 2000). Another study that favored

41   convergence over TSP in MHC class IIB based their conclusions on different codon usage in

42   shared amino acids among three Musteloid species, two that were sympatric and one that was

43   allopatric (Srithayakumar et al. 2012).

44   While evidence for TSP in MHC genes has been found in birds (Vincek et al. 1997, Sato et al.

45   2001, Richardson and Westerdahl 2003, Alcaide et al. 2007, Burri et al. 2008, Alcaide et al.

46   2013), few studies to date have provided convincing evidence of TSP in MHC IIB genes in

47   passerines. The strongest evidence for TSP in passerine MHC IIB genes was shown in a study of

48   recently diverged Darwin's finches, where several well supported interspecific allelic lineages

49   were described (Sato et al. 2011). Well supported interspecific clustering of variants was also

50   reported in closely related species within the Muscicapidae family (Zagalska-Neubauer et al.

51   2010).

52   The unusually complex nature of the passerine MHC may partially explain why TSP is

53   difficult to detect in this order. The MHC of passerines is among the most complex in terrestrial

54   vertebrates, exhibiting highly duplicated classical MHC class I and II genes and more than ten

55   functional class II B loci have been reported in some species. This contrasts sharply with other

56   bird taxa such as those within the Galliformes, where one to three MHC IIB loci have been

57   reported (Balakrishnan et al. 2010, Bollmer et al. 2010, Eimes et al. 2012, Sepil et al. 2012). It is

58   likely that the unusually high levels of polymorphism observed in the MHC of songbirds is the

59   result of a combination of several factors, including gene duplication, subsequent recombination

60   (including gene conversion), and balancing selection (Hess and Edwards 2002). Because of the

61   potential for recombination between non-homologous loci, the role of recombination, especially

4

62   gene conversion, may be exaggerated in passerines relative to other vertebrate taxa in shaping

63   observed patterns of interspecific MHC variation between, as well as within species. The best

64   evidence for this can be seen in phylogenetic trees of songbirds, where multi-locus MHC IIB

65   sequences tend to cluster by species, rather than by locus (e.g. Sutton et al. 2013). This lack of

66   trans-specific clustering of orthologous loci in passerines is thought to be due to concerted

67   evolution as a result of gene conversion (Hess and Edwards 2002).

68   In this study, we characterized MHC IIB loci in three closely related species of crows: jungle

69   crow (*Corvus macrorhynchos japonensis*), carrion crow (*C. corone orientalis*), and American

70   crow (*C. brachyrhynchos*).   We also tested for TSP and convergence by taking advantage of the

71   unusual arrangement of phylogeny and geographical ranges of the three species.   The

72   contemporary range of eastern carrion crows and jungle crows overlaps in East Asia

73   (southeastern China, the Korean peninsula and the Japanese archipelago) while American crows

74   are limited to North America (Madge and Burn 1993, Haring et al. 2012).   Mitochondrial DNA

75   (mtDNA) phylogenies, however, place carrion crows and American crows in a monophyletic

76   group, with jungle crows paraphyletic to these two species (Haring et al. 2012).   The genus

77   *Corvus* likely has a Palearctic origin, and mtDNA phylogenies suggest that carrion crows and

78   American crows diverged in the Nearctic relatively recently; while jungle crows represent a

79   single, older, lineage originating in the south-east Asian tropics (Haring et al. 2012).

80   Crows are suitable species for assessing TSP and convergence because their habitat, diet and

81   social behavior are likely to render them susceptible to a diverse array of pathogens. Crows are

82   omnivores that forage, in part, on carrion and human refuse (McGowan 2001); and scavengers,

83   in general, could be under selection for more robust immune responses because of contact with

84   pathogen-rich carcasses (Blount 2003).   Furthermore, all three of the *Corvus* species

5

85 investigated in this study are found in densely human-populated areas (Ali and Ripley 1972,

86 Richner 1989, McGowan 2001, Kurosawa et al. 2003), and some data suggest that species with

87 the ability to exploit urban areas have disproportionately strong immune responses (Moller

88 2009). Finally, most crows are highly gregarious, often foraging and roosting in large communal

89 flocks in the non-breeding season (Madge and Burn 1993, Verbeek and Caffrey 2002), which

90 could promote pathogen transmission among conspecifics (Bull 1994, Frank 1996, Møller et al.

91 2001).

92   For this study, we assumed that the two sympatric species (carrion crows and jungle crows),

93 which have similar habitats and co-occur in east Asia, likely share a more similar suite of

94 pathogens with each other than with American crows. Diverse pathogen burdens are well-

95 documented in all three species, and a comparative study of jungle crows and carrion crows

96 revealed that ten specific helminth species were shared between the two species (Mizuno 1984,

97 Miller et al. 2010, Wheeler et al. 2014). There have been several surveys of intestinal parasites

98 in American crows; however no report to date has identified these Asian helminth species in

99 American crows, except for one anecdotal case (Jones 1968, Hendricks et al. 1969, Cawthorn et

100 al. 1980, Naderman and Pence 1980, Mizuno 1984, Miller et al. 2010). A strong relationship

101 between MHC class IIB and resistance to helminth parasites has been established (Goüy de

102 Bellocq et al. 2008, Zhang and He 2013). For example, Froeschke and Sommer (2012) reported

103 both a positive and negative relationship between helminth parasites and specific MHC alleles in

104 rodents. Thus, similar selective pressures may have acted on the MHC IIB of carrion crows and

105 jungle crows, and a signal of convergent evolution may be detectable.

106   We tested for TSP among all three species of crows and convergence among carrion crows

107 and jungle crows by analyzing a 246 bp fragment of MHC IIB exon 2 generated by 454

6

108   pyrosequencing. For the analyses, we used five different nucleotide partitions as well as

109   supertyping, which clusters functionally equivalent amino acid sequences into units of selection

110   (Sette and Sidney 1999, Lund et al. 2004, Sidney et al. 2008). We also estimated species

111   divergence times for all three species using published, as well as newly generated, mtDNA

112   Control Region (CR) sequences.

113   **Materials and Methods**

114   *Sample collection*

115   We analyzed the data from 18 individuals of each species. American crows were sampled from

116   nestlings (one bird per nest) in Yolo County, California (38°32′ N, 121°45′ W) on the campus of

117   the University of California, Davis, in May and June 2012. Collection methods were approved

118   by the Institutional Animal Care and Use Committee of the University of California, Davis

119   (Permit Number: 16897). Approximately 50 μL of whole blood was preserved in Queens's lysis

120   buffer, and gDNA was extracted using a standard phenol/chloroform protocol followed by

121   ethanol precipitation. Previously extracted gDNA from the blood of jungle crows and carrion

122   crows was obtained from the Japanese National Museum of Nature and Science (Tsukuba,

123   Japan) and the Yamashina Institute of Ornithology (Chiba, Japan). These samples were

124   collected between 1994-2010 from the entire range of the Japanese archipelago, although most

125   jungle crows were from the Tokyo area (16 from Tokyo, 1 each from Hokkaido and Okinawa),

126   whereas carrion crow collection was more evenly distributed across the Japanese archipelago (6

127   from Kagoshima, 1 from Tsushima, 4 from Nara, 1 from Osaka and 6 from Tokyo; Fig. S1).

128   One *m*L of whole blood for RNA extraction was collected from three juvenile jungle crows

129   trapped at Ueno Zoo in Tokyo, Japan in 2012. The blood was flash frozen on site and stored at -

7

130   80 °C. Jungle crow RNA was extracted from whole blood using the RNeasy Protect Animal

131   Blood System (Qiagen).

132   *Amplification of MHC class IIB exon 2*

133   In order to isolate MHC IIB exon 2, which codes for the PBR, jungle crow cDNA from two

134   individuals (U2 and U3) was synthesized with the LongRange 2Step RT-PCR System (Qiagen)

135   using the manufacturer's recommended amplification protocol. MHC IIB was then targeted in

136   the cDNA using the primers MHC05 (Miller and Lambert 2004), which anneals within exon 1 in

137   several passerines (positions variable), and *Coma*IIbex3R, which was designed by aligning the

138   conserved regions of IIB exon 3 from several passerines (positions 19-38). These primers

139   amplified a 359 bp cDNA fragment. We verified that the amplicon was MHC IIB by cloning the

140   fragment from both individuals (TOPO XL One Shot-Invitrogen).   Twenty-four clones from

141   each individual were selected and whole colony PCR was performed using the TOPO vector

142   primers M13 F and M13 R. The 20 μL PCR contained a pluck of colony cells as template,

143   0.625 μM of each primer, 0.5 mM of each dNTP, 1.5 mM MgCl2, 1X PCR buffer and 0.5 U of

144   ExTaq (Takara) polymerase. The cycling conditions included an initial denaturation at 94 °C

145   for 2 min followed by 30 cycles of 94, 55 and 72 °C each for 1 min and a final extension of 72

146   °C for 10 min.   We sequenced each amplicon in both directions using M13 primers (Applied

147   Biosystems 3130xl). We performed a BLAST search and identified eight sequences that aligned

148   with known passerine MHC IIB.

149   We targeted intron 1 by aligning the cDNA-derived amplicons using MUSCLE in Geneious

150   6.1.4 (Biomatters) and a reverse primer, *Coma*IIbex2R, was designed in a conserved region of

151   exon 2 and paired with MHC05.   Using gDNA from two jungle crows (*Coma*5061 and

8

152   *Coma*7509), a 516 bp fragment extending from exon 1 to exon 2 was amplified using 50-100 ng

153   of gDNA, 1.0 μM of each primer, 0.5 mM of each dNTP, 1.5 mM MgCl2, 1X PCR buffer and

154   0.5 U of ExTaq (Takara) polymerase (25 μL total volume). The cycling conditions included an

155   initial denaturation at 94 °C for 30s followed by 30 cycles of 94 (20s), 60 (20s) and 72 (45s) °C

156   and a final extension of 72 °C for 10 min. Each of these two amplicons was cloned and 10

157   colonies were sequenced in each direction using the cloning and whole colony PCR protocol

158   described above. Sequences were aligned using MUSCLE and a forward primer flanking exon

159   2, *Coma*iF2, was designed for the (100%) conserved region at position 282-304 of intron 1.

160   Next, we isolated intron 2 by pairing *Coma*iF2 with *Coma*IIBex3R using the same individuals

161   and PCR, cloning, and whole colony PCR conditions described above. When amplifying from

162   exon 2 to exon 3 in jungle crows, we discovered indels that resulted in large gaps at different

163   locations in intron 2. 454 pyrosequencing is most efficient when sequences are of equal length,

164   and for this reason, we designed a single degenerate reverse primer, *ComaEx2RA*, between

165   positions 249-268 of exon 2. Finally, we confirmed that the primers *Coma*iF2 and *ComaEx2RA*

166   amplified MHC IIB loci in all three crow species. Primer sequences and annealing temperatures

167   are listed in Table S1.

168   *454 pyrosequencing*

169   Fusion primers were synthesized (Fasmac-Japan) by ligating the standard Roche multiplex

170   identifiers (MIDs) to both the forward and reverse MHC IIB primers and the 454 adaptor primers

171   (the Roche "Basic Amplicon" design). In order to minimize the formation of PCR artifacts, we

172   reduced the cycle number and primer concentration and eliminated the final extension step

173   during amplicon generation (Lenz and Becker 2008). PCRs contained 25-50 $n$gs of template,

174   0.5 μM of each fusion primer, 0.5 mM of each dNTP, 1.5 mM MgCl2, 1X PCR buffer, 5%

9

175 DMSO and 0.5 U of ExTaq (Takara) polymerase (25 μL total volume). The cycling conditions

176 included an initial denaturation at 94 °C for 30s followed by 28 cycles of 94 (20s), 60 (20s) and

177 72 (45s) °C.

178   PCR amplicons were purified using the Agencourt AMPure XP System (Beckman Coulter)

179 and recovered DNA was checked for quality and primer dimers on a 1.5% agarose gel. Samples

180 were then quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) and pooled in

181 equimolar amounts according to the Roche GS Junior Titanium Amplicon Library Preparation

182 Method Manual. The quality and concentration of the pooled samples was checked on a 2100

183 Bioanalyzer (Agilent Technologies) and then sequenced using the Roche GS Junior Titanium

184 Sequencing System at the Japanese National Museum of Nature and Science in Tsukuba, Japan.

185   Amplicons were bi-directionally sequenced, and reads that passed the initial 454 Roche Junior

186 quality filter were de-multiplexed using jMHC (Stuglik et al. 2011). jMHC is especially efficient

187 for de-multiplexing bi-directionally sequenced double-tagged amplicons because it bins only

188 those sequences that contain both forward and reverse primers and their associated MIDs with no

189 ambiguous characters ("N"s) in the primers, MIDs or target sequence. Sorted FASTA files

190 (excluding primers and MIDs) were then imported into Geneious and aligned with published

191 MHC class IIB sequences.   Sequences containing indels were removed from the data set at this

192 time.

193 *Bioinformatics variant validation and genotyping*

194 Variants were accepted as "true" if they occurred in at least three reads ($r > 2$) in both of two

195 independent PCRs of the same individual. The criteria of ($r > 2$) is derived from the estimate of

196 a read containing at least one sequence error for a 171 bp fragment of (HLA) DRB exon 2 of

10

197 ~0.11 (Galan et al. 2010). Thus, the probability that an identical error occurs three times in the

198 same sample is ~10^{-8} (Galan et al. 2010). At this stage, the most likely types of artefacts

199 remaining in the data set are either single base substitutions or fragment chimeras generated

200 during PCR. Although both of these artifacts are randomly generated, if these errors occur early

201 in the PCR, they could be replicated and represented at relatively high frequencies in individual

202 amplicons. To further reduce the probability of erroneously including single base substitutions,

203 we classified sequences as true variants only when they differed by at least two nucleotide

204 substitutions from more common variants. Galan et al. (2010) reported an average chimera

205 frequency of 0.06 and some chimeras occurred at frequencies > 0.1; thus, elimination of

206 chimeras using frequency thresholds may erroneously inflate variant estimates. It is highly

207 unlikely that two independent PCRs would each generate identical chimeras (where as many as

208 10 different variants could recombine) or single base substitution errors three times ($r > 2$). By

209 validating variants with a second PCR for each individual (rather than a second PCR within or

210 across individuals) chimeras are eliminated, yet variants that are rare in the population are not

211 erroneously eliminated from the data set. The total number of reads that were included at this

212 stage was then designated as "net reads" for the purposes of genotype confidence level

213 calculations.

214   We used the program 'Negative Multinomial' to calculate the number of sequences that are

215 necessary for amplifying all the variants at least three times with a confidence level of 99%

216 (Galan et al. 2010). While the program is limited to eight possible variants, the relationship

217 between n, net reads, r, the number of times each variant must be sampled, and m, the maximum

218 number of variants per amplicon, is linear. Our cloned sequences indicated a maximum of 20

219 variants per individual which corresponds to a minimum of 175 reads for 99.9% confidence;

11

220 however this assumes equal sampling of variants. Because amplification bias is likely when co-

221 amplifying 10 loci, we expected this bias to be reflected in over- and under-representation of

222 some variants (Burri et al. 2014). Therefore, we increased the minimum required net reads for

223 genotyping to 275, which represents a frequency of 1.1% (amplification bias of ~ one order of

224 magnitude) for a variant that has exactly three copies in an amplicon represented by 275 net

225 reads.

226 *Tests for selection and MHC IIB phylogenetics*

227 For each species, we generated an alignment of validated variants, and identified the codons

228 that have been shown to comprise the PBR of HLA in humans (Brown et al. 1993). While most

229 MHC studies rely on the HLA PBR codons identified in humans to estimate selection, it has not

230 been shown that these same codon positions apply universally or that selection is limited to

231 peptide binding codons. Therefore, in order to identify specific codons that are likely under

232 selection in crows, we performed a Wu-Kabat analysis on the amino acid alignment of verified

233 alleles in all three species. A Wu-Kabat plot predicts amino acids that are likely subject to

234 selection by identifying positions of high variability. Variability is calculated by dividing the

235 number of amino acids at a position by the frequency of the most common amino acid (Wu and

236 Kabat 1970). For each species, we identified codons/amino acids as likely under selection if the

237 Wu-Kabat variability metric was ≥ twice the mean of Wu-Kabat values for all sites (Wu and

238 Kabat 1970, Bos and Waldman 2006). These sites are denoted WuK-PS (PS = polymorphic site).

239 We tested for selection on the HLA PBR sites by calculating the ratios of non-synonymous (dN)

240 and synonymous mutations (dS) using a Z test (modified Nei-Gojobori- Jukes Cantor correction)

241 in MEGA v. 6.0 (Tamura et al. 2011). For each species, individual codons were tested for

242 selection using the HYPHY package in MEGA. This method, known as the counting method,

12

243   estimates the number of nonsynonymous and synonymous mutations that have occurred at each

244   codon by using a maximum likelihood reconstruction of the ancestral state of each sequence.

245   The test statistic $dN - dS$ is used for detecting codons that have undergone selection and for

246   positive values the probability of rejecting the null hypothesis of neutral evolution (p-value) is

247   calculated (Suzuki and Gojobori 1999, Pond and Frost 2005). Significant p values are denoted as

248   positively selected sites. This method is more computationally compatible with large data sets

249   than empirical or hierarchal Bayesian approaches (Yang et al. 2000), yet has been shown to

250   provide nearly identical results with a sufficient number of sequences (Pond and Frost 2005).

251   We constructed nucleotide phylogenies from five different partitions of exon 2: variable sites

252   identified by the Wu-Kabot plot (WuK-PS, 33bp), non-WuK-PS sites (213 bp), HLA PBR sites

253   (66 bp), non-HLA PBR sites (180 bp) and the contiguous 246 bp fragment of exon 2. For each

254   analysis, 24 different nucleotide substitution models were tested for best fit (ML) in MEGA and

255   the Jukes-Cantor (JC) Model + G (Gamma) + I (invariant sites) was chosen based on the

256   Bayesian Information Criterion (BIC) score.

257

258   Supertyping

259   Functionally equivalent MHC IIB alleles were clustered using the supertyping method described

260   in Sepil et al. (2012), a study that also examined highly duplicated MHC loci in a passerine

261   species. Briefly, amino acid sites identified as positively selected sites were aligned and

262   characterized by five physicochemical descriptor variables: $z1$ (hydrophobicity),$z2$ (steric bulk),

263    $z3$ (polarity), $z4$ and $z5$ (electronic effects). These descriptor variables were placed into a matrix

264   and subjected to K-means clustering algorithm (Sandberg et al. 1998, Doytchinova and Flower

265   2005). Discriminant analysis of principle components (DAPC) was used for describing clusters

13

266 in the 'adegenet' package in R (Jombart et al. 2010). We performed three different supertype

267 analyses, using three different partitions: (i) the nine positively selected sites identified in the

268 HYPHY analysis that were shared among all three species; (ii) WuK-PS sites that were shared

269 among all three species; and (iii) HLA PBR codons.

270 *DNA phylogenetics and species divergence*

271 To ensure that the mtDNA phylogenies used in Haring et al. (2012) were consistent with our

272 populations, we tested a subset of carrion crows (N = 4), jungle crows (N = 6) and American

273 crows (N = 4) at the same mtDNA region using the primers CR-Cor+ and Phe-Cor- (Table S1)

274 and the same PCR conditions as Haring et al. (2012), followed by Sanger sequencing as

275 described above. We aligned these sequences with four mtDNA sequences from each species

276 that were used in the original phylogenetic construction by Haring et al. (2012). The Hasegawa-

277 Kishino-Yano (HKY) model (Lenz et al. 2013) had the lowest BIC scores in the best-fit

278 substitution model (ML). HKY ML trees with 500 bootstrap replicates were constructed in

279 MEGA v. 6.0.

280 Using the mtDNA sequences, we estimated divergence times for the three crow species: the

281 divergence of carrion crow/American crow clade from jungle crow, and the divergence between

282 carrion crows and American crows. Divergence times to the most recent common ancestor were

283 estimated in BEAST v. 2.0 (Bouckaert et al. 2014). We used the HKY site model and a Yule

284 process speciation prior for the branching rates. We tested the assumption of a molecular clock

285 using the maximum likelihood method which compares the ML value for the given topology

286 with and without the molecular clock constraints under the HKY model (MEGA v. 6.0). We ran

287 two separate analyses using two different mtDNA substitution rates (per site, per My) previously

14

288 estimated for crows (*C. macrorynchos* = 0.0567555 and *C. coronoides* = 0.0603431) (Nabholz et

289 al. 2009). The Markov chain Monte Carlo (MCMC) analyses were run for 10^7 generations

290 (10,000 iteration burn in); the mean and 95% highest posterior density interval (HPD) for

291 divergence times were calculated in Tracer v. 1.6 (Bouckaert et al. 2014) .

292

Results

294 Pyrosequencing

295 The goal of the first 454 run was to estimate the number of IIB loci and the minimum read

296 number required for accurate genotyping (99.9% confidence). Six individuals from each species

297 were sequenced and a total of 125,839 reads passed the initial 454 filter. After excluding reads of

298 less than 270 bp, there were 76,891 reads with an average length of 340 nucleotides. Amplicons

299 had between 8,384 and 483 initial reads (mean = 1,213) that were sorted by jMHC. These

300 samples had an average of 376 net reads (range 171-3,211). The maximum number of variants

301 (differing by at least 2 nucleotides) found in a single individual was 20, which suggests that

302 crows have up to 10 IIB loci per haplotype. The required minimum read number was calculated

303 as 211; however, we increased the cutoff to 275 net reads to account for amplification bias (see

304 methods).

305 For the second 454 run, a total of 107,279 reads passed the initial 454 filter. After excluding

306 reads of less than 270 bp, there were 90,203 reads with an average length of 340 nucleotides. A

307 total of 18 individuals (with replicates) from each species met the criteria for variant validation

308 of at least 275 net reads and at least three identical reads per variant in both of two independent

15

309   PCRs of the same individual. All DNA sequences were deposited in NCBI Genbank: Accession

310   numbers: MHC IIB: TBA; crow mtDNA: KM246294-308; crow nuclear intronic: KM246309-23

311   *MHCIIB variation and selection*

312   The 248 bp fragment was conserved at the final two base pair positions in all three species and

313   was trimmed to 246 bp. The number of validated MHC IIB nucleotide variants for each species

314   was: jungle crows = 89, carrion crows = 81 and American crows = 67. Individuals had a range

315   of 7-20 variants, indicating that the MHC IIB in crows exhibits copy number variation (Table 1),

316   although it is possible that the range of variant number could be explained by other phenomena,

317   including null alleles, allele sharing across loci, homozygous vs. heterozygous loci or the variant

318   validation methodology (where sequences with indels were excluded from the data set). Few

319   variants were shared among species and there was no discernable pattern of allele sharing

320   relating to phylogeny or sympatry: Four unique variants were shared among American crows and

321   jungle crows, three among American crows and carrion crows, three among carrion crows and

322   jungle crows and four variants were shared by all three species.

323   The codon-based Z test for selection (non-synonymous/synonymous mutations) averaged

324   across the PBR yielded a significant *p* value for American crows and carrion crows but not for

325   jungle crows (Table 1). The Wu-Kabot plot identified 11 amino acid sites that were highly

326   variable in all three species: 6, 23, 25, 32, 33, 52, 61, 62, 65, 73 and 81. Eight of these sites (73

327   %) corresponded to known peptide binding sites in HLA (Fig. 1). Among the three species,

328   jungle crows exhibited the most amino acid variation, with five hypervariable (>20 Wu-Kabat

329   index) amino acid sites (Fig.1). The HYPHY test for selection identified nine positively selected

16

330   sites that were shared among the three species (2, 4, 33, 52, 55, 66, 68, 72 and 81), but only 5 of

331   these (56%) corresponded to the HLA PBR (Fig. 1).

332   In total, we recovered 172 unique amino acid sequences among the three species (70 in jungle

333   crows, 58 in carrion crows and 44 in American crows). We detected no species-specific motifs

334   or motifs that were shared exclusively between two species. Within-species amino acid diversity

335   was relatively high (0.245 ± 1.5), similar to that which we calculated for scrub jays (*Aphelocoma*

336   *coerulescens*), another corvid species (23.3 %; N = 11; NCBI (U23958-65; 72,73,75) (Edwards

337   et al. 1995). An alignment (82 sites) of common amino acid fragments from the three crow

338   species with four other passerine species, and a falconiform species for comparison, highlights

339   the high level of amino acid diversity found among and between species of passerines (Fig. S2).

340   *MHC IIB phylogenies*

341   We observed a similar pattern among the five different partitions of exon 2: MHC IIB variants

342   formed several supported (bootstrap values ≥50), interspecific clusters in trees comparing

343   putatively functional sites. Interspecific clusters were also observed when comparing putatively

344   neutral sites, but there were fewer of these supported clusters, and in general, they were less

345   inclusive. This pattern was most striking in trees constructed using the contiguous 246 bp exon

346   2 fragment (Fig. 3). For instance, a tree that was constructed by comparing non-synonymous

347   substitutions (Fig. 3A) shows six well supported clusters, each containing all three crow species,

348   whereas a tree constructed by comparing synonymous substitutions of the same fragment (Fig.

349   3B), shows only three such clusters. Supported clustering at putatively functional sites was also

350   more inclusive in these exon 2 trees, with 44% of all variants (105/237) forming supported

351   interspecific clades compared to just 18% (32/237) in the tree using synonymous sites. The trees

352   constructed from WuK-PS and non-Wuk-PS sites (Fig. S3A and S3B) as well as those

17

353 constructed from the HLA PBR and non-PBR sites (Fig. S4A and S4B) displayed similar, though

354 less distinct patterns. Among all partitions, the most common (supported) clustering was among

355 all three species. We did not observe biased clustering of variants among the sympatric carrion

356 crows and jungle crows; or among the more recently diverged, allopatric, American crows and

357 carrion crows.

358 Supertyping

359 Among the three species, there were nine shared positively selected sites in exon 2. Five of these

360 nine sites corresponded to HLA PBR sites (Fig. 1). From the 140 unique positively selected sites

361 amino acid variants a total of eight supertypes were identified (Fig. 4). Only one supertype

362 (Supertype 4) was limited to carrion crows and jungle crows, whereas the rest of the supertypes

363 were shared among the three species.   The supertypes generated by the WuK-PS and the HLA

364 PBR codons were not well supported and thus excluded from our analysis.

365 mt*DNA phylogeny and divergence estimate*

366 The CR-Cor+ and Phe-Cor- primers amplified an 836-920 bp fragment of the mtDNA CR region

367 in the three species of crows. The ML tree placed carrion crows and American crows in a

368 separate clade from jungle crows with 100% bootstrap support (Fig. 2), thus confirming that the

369 jungle crows and carrion crow populations from this study conform to the phylogeny of Haring

370 et al. (2012).   The ML-based test of the molecular clock failed to reject the null hypothesis of

371 equal evolutionary rate throughout the tree ($p = 0.85$); thus we used a strict clock setting for the

372 MCMC iterations in the BEAST program.   Using the mtDNA substitution rate of

373 0.0567555/site/My the estimated mean divergence time to MRCA of all species was 0.8992 Ma

374 with a 95% HPD interval of 0.7055, 1.1147 Ma (effective sample size (ESS) = 8581.7, median =

375   0.8942, standard deviation = 0.107).   The estimated divergence to the MRCA of carrion crows

376   and American crows was 0.4372 Ma with a 95% HPD interval of 0.3036, 0.5801 Ma (ESS =

377   7682.7, median = 0.4319, standard deviation = 0.0723) (Fig. 2). Divergence estimates using the

378   higher substitution rate (*C. coronoides* = 0.0603431/site/My) yielded similar values: the MRCA

379   of all species = 0.8509 Ma, HPD interval = 0.6578, 1.0486 Ma; MRCA of carrion crows and

380   American crows = 0.4143 Ma, HPD interval = 0.2881, 0.5458 Ma.

381   **Discussion**

382   We observed strong interspecific clustering of MHC IIB variants in three closely related species

383   of crows (*Corvus* sp.), and most of the supported clusters contained all three species; however,

384   the clustering was much more inclusive, and better supported, when comparing non-synonymous

385   substitutions at putatively selected sites than synonymous substitutions at putatively non-selected

386   sites (Fig. 3; Fig. S3; Fig. S4). This pattern was observed in all of the phylogenies constructed

387   from the different partitions of exon 2, but was most obvious in the trees constructed using the

388   entire 246 fragment of exon 2 (Fig. 3). This pattern of strong clustering among functional sites is

389   suggestive of convergent evolution (Klein et al. 2007). In fact, if this study been limited to the

390   sympatric carrion crows and jungle crows, we may have erroneously concluded that convergent

391   evolution explains these patterns; however, inclusion of sequences from the allopatric American

392   crow, and the resulting interspecific clustering of all three species at functional sites, suggests

393   that the pattern observed in the phylogenies is not convergence, but rather trans-species

394   polymorphism.

395   Our results indicate that the MHC IIB in crows exhibits the same characteristics of the MHC

396   IIB observed in other passerine species (Bollmer et al. 2010, Zagalska-Neubauer et al. 2010),

19

397 namely, high levels of polymorphism, highly duplicated loci, probable copy number variation of

398 loci and a strong signature of selection. Consistent with similar studies, we also observed highly

399 polymorphic amino acid positions and detected selection at sites other than the HLA PBR (Fig.

400 1), which indicates that either the PBR in passerines comprises different codon positions than the

401 human MHC or that regions other than the PBR are under selection (Anmarkrud et al. 2010,

402 Bollmer et al. 2010, Zagalska-Neubauer et al. 2010). The Wuk-Kabot plot was a better fit to

403 the HLA PBR than the HYPHY probability based test for selection (Fig. 1), which may indicate

404 that such probability tests are not reliable when identifying selected sites across many duplicated

405 loci if the relationships between these loci are unknown (e.g. orthologous versus paralogous

406 relationships). For example, it is unlikely that the hyper-polymorphic amino acid sites at

407 positions 6, 23, 32, 62, and 65 are not subject to selection. This highlights the importance of

408 using a combination of different exon 2 partitions when evaluating selection or establishing

409 putative neutral and selected sites for phylogenetic analyses in complicated systems such as the

410 passerine MHC.

411 Several factors may explain the relatively high MHC IIB neutral genetic variation between

412 allelic lineages we observed in these crow species. First, if these allelic clusters represent

413 ancient lineages, they are likely millions of years older than the MRCA of all three species,

414 which allows sufficient time for random mutations to be incorporated, and the likelihood that

415 these neutral mutations persist is increased by their close association with selected sites (Klein et

416 al. 1998, O'huigin et al. 2000). Thus, the majority of the synonymous polymorphisms observed

417 in neutral sites were likely inherited, rather than generated since species divergence. Other

418 factors that may contribute to the relatively high synonymous nucleotide variation between

419 species observed at non-selected sites relative to selected sites are the combined effects of

20

420   recombination through genetic exchange and gene conversion.   In a study of MHC IIB in

421   sticklebacks (Gasterosteidae), Lenz et al. (2013) observed a similar pattern of clustering of

422   selected sites (in this case, the HLA PBR) but a lack of clustering in the remaining sites of exon

423   2, a pattern similar to the one reported here. The authors reasoned that selected sites may evolve

424   under site-specific functional constraints that maintain similar amino acid motifs, and that any

425   new variants generated by recombination are more likely to be retained if they share this

426   functional similarity. Intraspecific recombination and gene conversion could then erode the

427   signal of common allelic ancestry over time.

428   Gene conversion is thought to explain why passerine MHC sequences often sort by species,

429   rather than by locus, on phylogenetic trees (Hess and Edwards 2002). Elevated rates of gene

430   conversion have been documented in birds, and one study reported a gene conversion rate over a

431   magnitude higher than the neutral mutation rate (Spurgin et al. 2011).   It is possible that the

432   reason why the crow MHC IIB sequences do not sort intra-specifically is that insufficient time

433   has passed since species divergence for gene conversion to homogenize the neutral sites within

434   exon 2. The patterns we observed were similar to those described for MHC IIB in Darwin's

435   finches in that the sequences sorted in a manner that would be expected from a single,

436   undifferentiated species (Sato et al. 2011). The similar results between our study and that of Sato

437   et al. (2011) are likely due to the relatively recent divergence times of the focal taxa in both

438   studies. In the Galapagos, Darwin finches' radiated from a founding population approximately

439   2.5 Ma, while the MRCA to the three crow species was estimated at approximately 900,000

440   years ago.

441   While we did not detect evidence of convergence between the sympatric carrion crow and

442   jungle crow in the MHC IIB phylogenies, the supertype analysis revealed a pattern consistent

21

443 with convergent evolution. While seven of the eight supertypes were shared among the three

444 species, one supertype, ST-4, was exclusive to carrion crows and jungle crows, which might

445 reflect selective pressure from a specific, shared parasite. None of the other seven supertypes

446 were exclusive to either a single species or group of two species.   This supertyping method

447 classifies MHC II B loci into clusters based on the physio-chemical properties of the amino acids

448 in the positively selected sites of the PBR. Several studies of primate MHC have shown that

449 different MHC alleles bind very similar peptide motifs, suggesting significant overlap in peptide

450 binding repertoires (Sette and Sidney 1999, Lund et al. 2004, Sidney et al. 2008). Thus, different

451 amino acid sequences can be functionally equivalent, and grouped together as a supertype,

452 considered a unit of selection. Recently, (Sepil et al. 2013) showed that different MHC class I

453 supertypes confer resistance to malaria in great tits (*Parus major*). In our study, supertypes ST-3

454 and ST-4 grouped together, away from the other six supertypes, which may suggest that these

455 groups are functionally important for a specific type of parasite, whereas the other supertypes

456 may confer resistance to more common, generalist parasites (Fig 4).

457

458 **Conclusion**

459 In the present study we describe strong evidence for trans-specific polymorphism at the MHC

460 IIB in three species of crows. Our results are consistent with two other studies that examined

461 MHC IIB in closely related passerine taxa and indicate that TSP in passerines may be common,

462 but due to intra-specific gene conversion, most easily detectable among recently diverged

463 species.

464   Given that the allopatric American crow and carrion crow diverged approximately 430,000

465 years ago, we expected to find some signal of convergence of MHC IIB genes between the

22

466   sympatric jungle crow and carrion crow in the MHC IIB phylogenies. While we did find

467   patterns consistent with convergence in the phylogenetic analyses, any convergence would have

468   had to occur among all three species, before the MRCA split approximately 900,000 years ago.

469   While this scenario cannot be ruled out, the results presented here are better explained by TSP.

470

475   **References**

476   Alcaide, M., S. V. Edwards, and J. J. Negro. 2007. Characterization, polymorphism, and
477   evolution of MHC class II B genes in birds of prey. *Journal of Molecular Evolution*
478   65:541-554.
479   Alcaide, M., M. Lui, and S. V. Edwards. 2013. Major histocompatibility complex class I
480   evolution in songbirds: universal primers, rapid evolution and base compositional shifts
481   in exon 3. *PeerJ 1:e86 http://dx.doi.org/10.7717/peerj.86*
482   Ali, S. and S. Ripley. 1972. Handbook of the Birds of India and Pakistan. Oxford University
483   Press, Bombay.
484   Andersson, L. and S. Mikko. 1995. Generation of MHC Class II Diversity by Intra- and
485   Intergenic Recombination. *Immunological Reviews* 143:5-12.
486   Anmarkrud, J. A., A. Johnson, L. Bachman, and J. T. Lifjeld. 2010. Ancestral polymorphism in
487   exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes. *Journal of Evolutionary*
488   *Biology* 23:1206-1217.
489   Arendt, J. and D. Reznick. 2008. Convergence and parallelism reconsidered: what have we
490   learned about the genetics of adaptation? *Trends in Ecology Evolution* 23:26-32.
491   Balakrishnan, C., R. Ekblom, M. Volker, H. Westerdahl, R. Godinez, H. Kotkiewicz, D. Burt, T.
492   Graves, D. Griffin, W. Warren, and S. Edwards. 2010. Gene duplication and
493   fragmentation in the zebra finch major histocompatibility complex. *BMC Biology* 8:29.
494   Blount, J. D. 2003. Do individual branches of immune defence correlate? A comparative case
495   study of scavenging and non-scavenging birds. *Oikos* 102:340-350.
496   Bollmer, J. L., P. O. Dunn, L. A. Whittingham, and C. Wimpee. 2010. Extensive MHC Class II
497   B Gene Duplication in a Passerine, the Common Yellowthroat (Geothlypis trichas).
498   *Journal of Heredity* 101:448-460.

23

Bos, D. H. and B. Waldman. 2006. Evolution by Recombination and Transspecies Polymorphism in the MHC Class I Gene of Xenopus laevis. *Molecular Biology and Evolution* 23:137-143.

Bouckaert, R., J. Heled, D. Kuhnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10:e1003537.

Brown, J. H., T. S. Jardetzky, J. C. Gorga, L. J. Stern, R. G. Urban, J. L. Strominger, and D. C. Wiley. 1993. 3-dimensional structure of the human class-II histocompatibility antigen Hla-Dr1. *Nature* 364:33-39.

Bull, J. J. 1994. Perspective - Virulence. *Evolution* 48:1423-1437.

Burri, R., H. N. Hirzel, N. Salamin, A. Roulin, and L. Fumagalli. 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., M. Promerova, J. Goebel, and L. Fumagalli. 2014. PCR-based isolation of multigene families: lessons from the avian MHC class IIB. *Molecular Ecology Resources* 14:778-788.

Cawthorn, R., R. Anderson, and I. Barker. 1980. Lesions caused by Diplotriaena tricuspis (Nematoda: Diplotriaenoidea) in the American crow, Corvus brachyrhynchos Brehm. *Canadian Journal of Zoology* 58:1892-1898.

Doytchinova, I. A. and D. R. Flower. 2005. In Silico Identification of Supertypes for Class II MHCs. *Journal of Immunology* 174:7085-7095.

Edwards, S. V., M. Grahn, and W. K. Potts. 1995. Dynamics of Mhc evolution in birds and crocodilians: amplification of class II genes with degenerate primers. *Molecular Ecology* 4:719-729.

Edwards, S. V., E. K. Wakeland and W. K. Potts. 1995. Contrasting histories of avian and mammalian Mhc genes revealed by class II B sequences from songbirds. Proceedings of the National Academy of Sciences, USA. 92: 12200-4.

Eimes, J. A., K. M. Reed, K. M. Mendoza, J. L. Bollmer, L. A. Whittingham, Z. W. Bateson, and P. O. Dunn. 2012. Greater prairie-chickens have a compact MHC-B with a single class 1A locus. *Immunogenetics* 87:435-435.

Frank, S. A. 1996. Models of parasite virulence. *Quarterly Review of Biology* 71:37-78.

Galan, M., E. Guivier, G. Caraux, N. Charbonnel, and J. F. Cosson. 2010. A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. *BMC Genomics* 11:296.

Garrigan, D. and P. W. Hedrick. 2003. Detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* 57:1707-1722.

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

Graser, R., C. OhUigin, V. Vincek, A. Meyer, and J. Klein. 1996. Trans-species polymorphism of class II Mhc loci in danio fishes. *Immunogenetics* 44:36-48.

Gustafsson, K. and L. Andersson. 1994. Structure and polymorphism of horse MHC class II DRB genes: convergent evolution in the antigen binding site. *Immunogenetics* 39:355-358.

Haring, E., B. Daubl, W. Pinsker, A. Kryukov, and A. Gamauf. 2012. Genetic divergences and intraspecific variation in corvids of the genus Corvus (Aves: Passeriformes: Corvidae) - a

24

545   first survey based on museum specimens. *Journal of Zoological Systematics and*
546   *Evolutionary Research* 50:230-246.

547 Hendricks, L., R. Harkema, and G. Miller. 1969. Helminths of the Crow, Corvus brachyrhynchos
548   Brehm, 1822, in North Carolina1. *Proceedings of the Helminthological Society of*
549   *Washington* 36:150-152.

550 Hess, C. and S. Edwards. 2002. The evolution of the major histocompatibility complex in birds.
551   *Bioscience* 52:423-431.

552 Hughes, A. L. 1999. Adaptive evolution of genes and genomes. Oxford University Press, New
553   York, New York.

554 Janeway, C., P. Travers, M. Walport, and M. Shlomchick. 2005. Immunobiology: the immune
555   system in health and disease. Garland Press, New York.

556 Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a
557   new method for the analysis of genetically structured populations. *Bmc Genetics* 11.

558 Jones, J. 1968. Some Parasites of the Common Crow, Corvus Brachyrhynchos Brehm, from
559   Ohio. *The Ohio Journal of Science* 68:25-31.

560 Klein, D., H. Ono, C. O'HUigin, V. Vincek, T. Goldschmidt, and J. Klein. 1993. Extensive MHC
561   variability in cichlid fishes of Lake Malawi. *Nature* 364:330-334.

562 Klein, J. 1986. The natural history of the major histocompatibility complex. John Wiley, New
563   York.

564 Klein, J., A. Sato, S. Nagl, and C. O'hUigín. 1998. Molecular trans-species polymorphism.
565   *Annual Review of Ecology and Systematics* 29:1-21.

566 Kriener, K., C. O'HUigin, H. Tichy, and J. Klein. 2000. Convergent evolution of major
567   histocompatibility complex molecules in humans and New World monkeys.
568   *Immunogenetics* 51:169-178.

569 Kupfermann, H., Y. Satta, N. Takahata, H. Tichy, and J. Klein. 1999. Evolution of Mhc-DRB
570   introns: Implications for the origin of primates. *Journal of Molecular Evolution* 48:663-
571   674.

572 Kurosawa, R., R. Kono, T. Kondo, and Y. Kanai. 2003. Diet of jungle crows in an urban
573   landscape. *Global Environmental Research* 7:193-198.

574 Lenz, T. L. and S. Becker. 2008. Simple approach to reduce PCR artefact formation leads to
575   reliable genotyping of MHC and other highly polymorphic loci - Implications for
576   evolutionary analysis. *Gene* 427:117-123.

577 Lenz, T. L., C. Eizaguirre, M. Kalbe, and M. Milinski. 2013. Evaluating patterns of convergent
578   evolution and trans-species polymorphism at MHC immunogenes in two sympatric
579   stickleback species. *Evolution* 67:2400-2412.

580 Lund, O., M. Nielsen, C. Kesmir, A. G. Petersen, C. Lundegaard, P. Worning, C. Sylvester-Hvid,
581   K. Lamberth, G. Roder, S. Justesen, S. Buus, and S. Brunak. 2004. Definition of
582   supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics*
583   55:797-810.

584 Madge, S. and H. Burn. 1993. Crows and jays. Helm, London, UK.

585 Martinsohn, J. T., A. B. Sousa, L. A. Guethlein, and J. C. Howard. 1999. The gene conversion
586   hypothesis of MHC evolution: a review. *Immunogenetics* 50:168-200.

587 McGowan, K. 2001. Demographic and behavioral comparisons of suburban and rural American
588   Crows. Kluwer Academic Press, Norwell, MA.

589 Miller, A. D., A. K. Townsend, K. J. McGowan, A. Clark, A. L. Glaser, L. A. Patrican, E.
590   Dobson, and E. L. Buckels. 2010. Non-west nile virus-associated mortality in a poulation

25

591   of Amercian crows (*Corvus brachyrhynchos*): a gross and histopathologic study. *Journal*
592   *of Veterinary Diagnostic Investigation* 22:289-295.

593 Miller, H. C. and D. M. Lambert. 2004. Gene duplication and gene conversion in class II MHC
594   genes of New Zealand robins (Petroicidae). *Immunogenetics* 56:178-191.

595 Mizuno, F. 1984. Studies on the parasite fauna of the eastern carrion crow, Corvus corone
596   orientalis Eversman, and the Japanese jungle crow, Corvus macrorhynchos japonensis
597   *Japanese Journal of Veterinary Research* 32:105.

598 Moller, A. P. 2009. Successful city dwellers: a comparative study of the ecological
599   characteristics of urban birds in the Western Palearctic. *Oecologia* 159:849-858.

600 Møller, A. P., S. Merino, C. R. Brown, and R. J. Robertson. 2001. Immune Defense and Host
601   Sociality: A Comparative Study of Swallows and Martins. *American Naturalist* 158:136-
602   145.

603 Nabholz, B., S. Glemin, and N. Galtier. 2009. The erratic mitochondrial clock: variations of
604   mutation rate, not population size, affect mtDNA diversity across birds and mammals.
605   *Bmc Evolutionary Biology* 9.

606 Naderman, J. and D. Pence. 1980. Helminths of the Common Crow, Corvus brachyrhynchos
607   Brehm, from West Texas. *Proceedings of the Helminthological Society of Washington*
608   47:100-105.

609 O'huigin, C., Y. Satta, A. Hausmann, R. L. Dawkins, and J. Klein. 2000. The implications of
610   intergenic polymorphism for major histocompatibility complex evolution. *Genetics*
611   156:867-877.

612 Pearce, T. 2011. Convergence and Parallelism in Evolution: A Neo-Gouldian Account. *The*
613   *British Journal for the Philosophy of Science* 63:429-448.

614 Pond, S. L. K. and S. D. W. Frost. 2005. Not so different after all: A comparison of methods for
615   detecting amino acid sites under selection. *Molecular Biology and Evolution* 22:1208-
616   1222.

617 Richardson, D. S. and H. Westerdahl. 2003. MHC diversity in two Acrocephalus species: the
618   outbred Great reed warbler and the inbred Seychelles warbler. *Molecular Ecology*
619   12:3523-3529.

620 Richner, H. 1989. Habitat-Specific Growth and Fitness in Carrion Crows (Corvus-Corone-
621   Corone). *Journal of Animal Ecology* 58:427-440.

622 Sandberg, M., L. Eriksson, J. Jonsson, M. Sjostrom, and S. Wold. 1998. New chemical
623   descriptors relevant for the design of biologically active peptides. A multivariate
624   characterization of 87 amino acids. *Journal of Medicinal Chemistry* 41:2481-2491.

625 Sato, A., W. E. Mayer, H. Tichy, P. R. Grant, B. R. Grant, and J. Klein. 2001. Evolution of Mhc
626   class II B genes in Darwin's finches and their closest relatives: birth of a new gene.
627   *Immunogenetics* 53:792-801.

628 Sato, A., H. Tichy, P. R. Grant, B. R. Grant, T. Sato, and C. O,ÄôhUigin. 2011. Spectrum of
629   MHC Class II Variability in Darwin's Finches and Their Close Relatives. *Molecular*
630   *Biology and Evolution* 28:1943-1956.

631 Sepil, I., S. Lachish, A. E. Hinks, and B. C. Sheldon. 2013. Mhc supertypes confer both
632   qualitative and quantitative resistance to avian malaria infections in a wild bird
633   population. *Proceedings of the Royal Society B-Biological Sciences* 280: 20130134.

634 Sepil, I., H. K. Moghadam, E. Huchard, and B. C. Sheldon. 2012. Characterization and 454
635   pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal
636   complexity in a passerine system. *Bmc Evolutionary Biology* 12.

Sette, A. and J. Sidney. 1999. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 50:201-212.

Sidney, J., B. Peters, N. Frahm, C. Brander, and A. Sette. 2008. HLA class I supertypes: a revised and updated classification. *BMC Immunology* 9:1.

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* 2:16.

Spurgin, L. G., C. van Oosterhout, J. C. Illera, S. Bridgett, K. Gharbi, B. C. Emerson, and D. S. Richardson. 2011. Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* 20:5213-5225.

Srithayakumar, V., S. Castillo, J. Mainguy, and C. J. Kyle. 2012. Evidence for evolutionary convergence at MHC in two broadly distributed mesocarnivores. *Immunogenetics* 64:289-301.

Stuglik, M. T., J. Radwan, and W. Babik. 2011. jMHC: software assistant for multilocus genotyping of gene families using next-generation amplicon sequencing. *Molecular Ecology Resources* 11:739-742.

Sutton, J. T., B. C. Robertson, C. E. Grueber, J. A. L. Stanton, and I. G. Jamieson. 2013. Characterization of MHC class II B polymorphism in bottlenecked New Zealand saddlebacks reveals low levels of genetic diversity. *Immunogenetics* 65:619-633.

Suzuki, Y. and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. *Molecular Biology and Evolution* 16:1315-1328.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: Molecular evolutionary genetics analysis using likelihood, distance and parsimony methods. *Molecular Biology and Evolution* 28:2731-2739.

Verbeek, N. and C. Caffrey. 2002. The American Crow (*Corvus brachyrhynchos*).*in* A. Poole and F. Gill, editors. Birds of North America. Birds of North America, Inc., Philadelphia.

Vincek, V., C. O'Huigin, and J. Klein. 1997. How large was the founding population of Darwin's finches? *Proceedings of the Royal Society B-Biological Sciences.* 264:111-118.

Wheeler, S. S., L. W. Woods, W. M. Boyce, C. D. Eckstrand, S. A. Langevin, W. K. Reisen, and A. K. Townsend. 2014. West Nile virus and non-West Nile virus mortality and coinfection of American crows (Corvus brachyrhynchos) in California. *Avian Diseases* 58:255-261.

Wu, T. T. and E. A. Kabat. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *Journal of Experimental Medicine* 132:211-250.

Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431-449.

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

Zagalska-Neubauer, M., W. Babik, M. Stuglik, L. Gustafsson, M. Cichon, and J. Radwan. 2010. 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *Bmc Evolutionary Biology* 10:395.

Zelano, B. and S. V. Edwards. 2002. An Mhc component to kin recognition and mate choice in birds: predictions, progress, and prospects. *American Naturalist* 160:S 225- 237.

Zhang, J. and S. Kumar. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. *Molecular Biology and Evolution* 14:527-536.

27

682 Zhang, M. and H. He. 2013. Parasite-mediated selection of major histocompatibility complex
683   variability in wild brandt's voles (Lasiopodomys brandtii) from Inner Mongolia, China.
684   *Bmc Evolutionary Biology* 13:149.

685

Figures and Tables

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

691

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

692

693

694 **Figure 1.** A Wu-Kabat plot of amino acid sequences from a 246 bp fragment of MHC IIB Exon 2 for
695 three species of crows. Blue circles represent HLA PBR codons identified by Brown et al. 1993. Red
696 ovals identify positively selected sites (PSS) identified by HYPHY. Variability is represented on the Y
697 axis and site position is on the X axis. Histogram key: Red: jungle crow; yellow: carrion crow; blue:
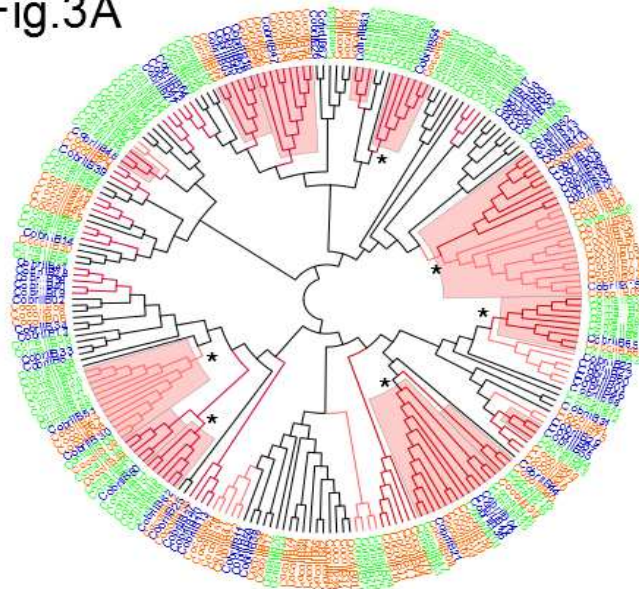698 American crow.

699

700

701

28
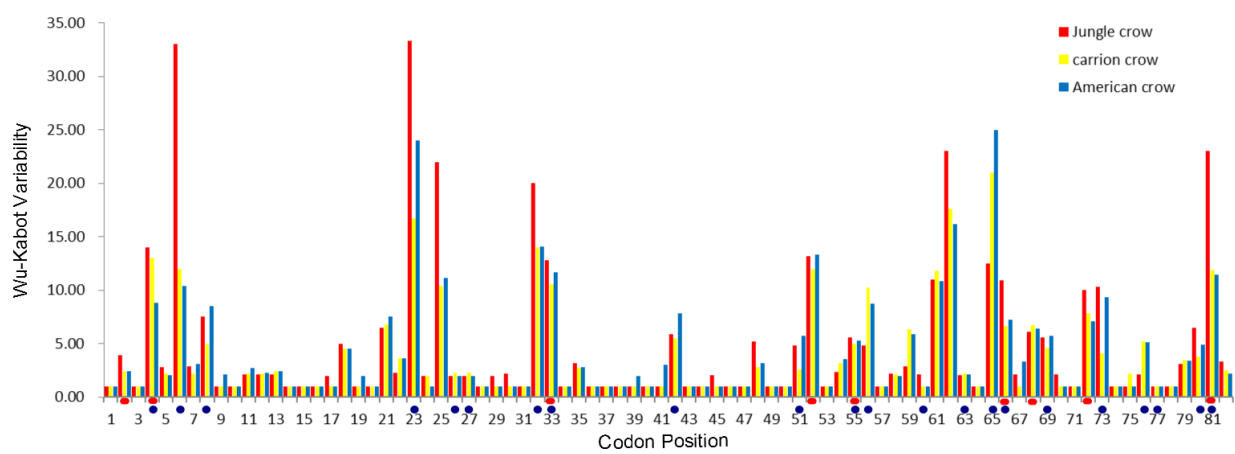
702 **Figure 2.** A ML tree (Hasegawa-Kishino-Yano (HKY)) using 836-920 bp fragments of the mtDNA
703 control region of jungle, carrion and American crows. Relative branch lengths are shown. At each node is
704 the estimated species divergence time in millions of years (above line) and 95% highest posterior density
705 interval (HPD) (below line) from MCMC simulations performed in BEAST v. 2.0. The mtDNA
706 mutation rate used in the divergence time estimate was 0.0567555 (Nabholz et al. 2009). Accession
707 numbers with an asterisk are from this study, all others are from Haring et al. 2012. Coco = carrion crow;
708 Cobr = American crow; Coma = jungle crow.
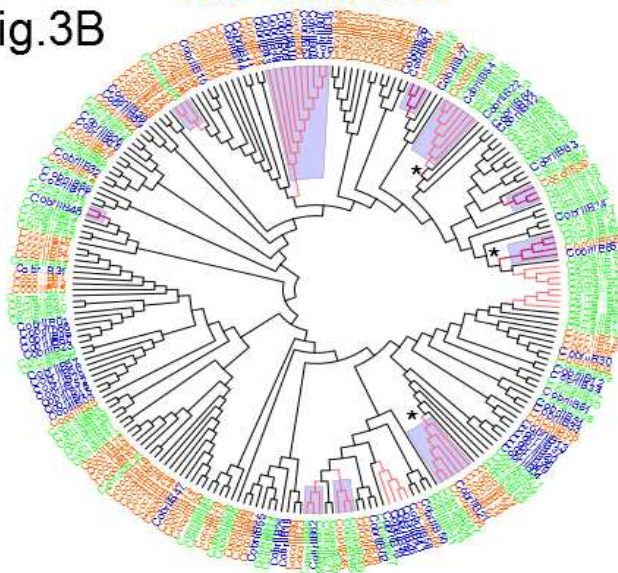
709

710

711

29
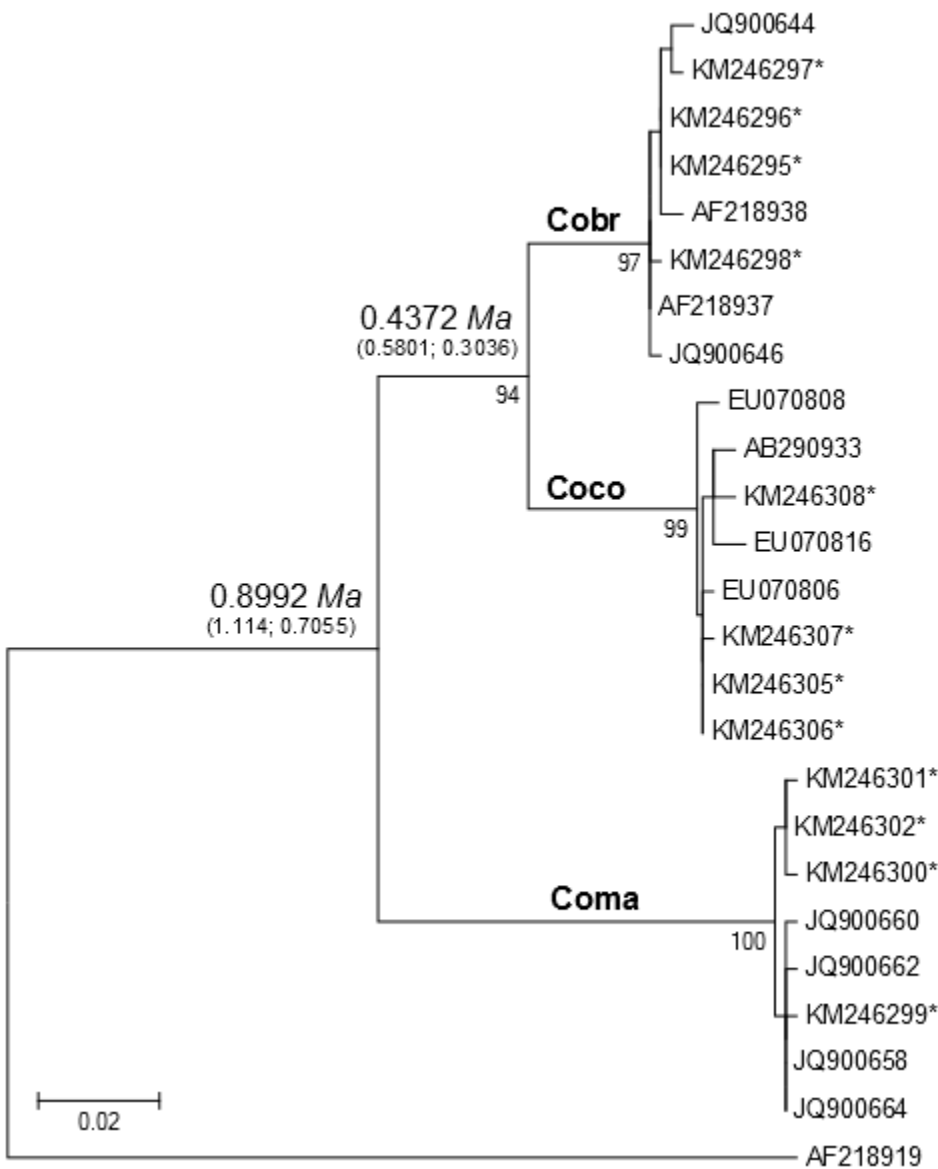
712 **Figure 3.** NJ trees (Jukes Cantor, 500 bootstrap replicates) of all MHC IIB variants in this study using a
713 246 bp fragment of exon 2. Branches with bootstrap support are indicated: light red = >50%; dark red =
714 >70. Supported nodes represented by all three species are numbered. Coco (orange) = carrion crow;
715 Cobr (blue) = American crow; Coma (green) = jungle crow. Fig. 3A: Tree constructed by comparing non-
716 synonymous substitutions/ nonsynonymous site. Red shading indicates supported interspecific clades.
717 Fig. 3B: Tree constructed by comparing synonymous substitutions/synonymous site. Blue shading
718 indicates supported interspecific clades.

719

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

725

726

727