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

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

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

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

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

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 sexual selection, as well as molecular mechanisms such as recombination and gene conversion, may play important roles (Andersson and Mikko 1995, Martinsohn et al. 1999, Zelano and Edwards 2002).

One distinguishing characteristic of MHC genes is that alleles often share more similarity between, rather than among species. Two mechanisms have been proposed to explain this "trans-specific" similarity of MHC alleles. The first explanation, called "trans-species polymorphism" (TSP), proposes that orthologous MHC allelic lineages are maintained by balancing selection, often over macro-evolutionary time-scales, and persist through speciation events (Klein et al. 1998). Strong evidence for TSP in MHC IIB genes has been found in a wide

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

range of taxa, particularly those of mammals and fish (Klein et al. 1993, Graser et al. 1996, Garrigan and Hedrick 2003, Lenz et al. 2013).

An alternative explanation for trans-specific MHC allele clustering is convergent or parallel evolution. The term molecular convergent evolution has been used to describe similar, but distinct phenomena (e.g. convergence between non-orthologous genes or within orthologues only) (Zhang and Kumar 1997, Yeager and Hughes 1999, Kriener et al. 2000); and debate exists for distinguishing molecular parallel evolution from molecular convergent evolution as well (Arendt and Reznick 2008, Pearce 2011). For the purpose of this study, we define convergence as any interspecific allele similarity that arises from adaptation to similar selective pressures (Gustafsson and Andersson 1994, Hughes 1999).

A common method used for identifying TSP, and in some cases for distinguishing TSP from convergence, is the comparison of codons thought to be under selection, such as those found in the MHC peptide binding region (PBR), with synonymous substitutions in surrounding codons or flanking introns. Alleles that are similar by decent should retain a signature of ancestry at selectively neutral sites. On the other hand, convergence is indicated if alleles are more similar at putatively selected sites than at neutral sites (Klein et al. 1998). Specifically, convergence and TSP can be tested by constructing different phylogenetic trees and comparing clustering patterns of trees constructed from non-synonymous mutations at codons likely to be under selection with trees constructed from synonymous substitutions at putatively neutral sites (Graser et al. 1996, Kupfermann et al. 1999, Kriener et al. 2000).

While many studies have provided evidence for TSP in MHC IIB allelic lineages, only a few have provided evidence for convergence. One study of MHC DRB sequences in New World

monkeys found that PBR amino acid motifs were shared among species but some sequences of the flanking introns sorted intra-specifically (Kriener et al. 2000). Another study that favored convergence over TSP in MHC class IIB based their conclusions on different codon usage in shared amino acids among three Musteloid species, two that were sympatric and one that was allopatric (Srithayakumar et al. 2012).

While evidence for TSP in MHC genes has been found in birds (Vincek et al. 1997, Sato et al. 2001, Richardson and Westerdahl 2003, Alcaide et al. 2007, Burri et al. 2008, Alcaide et al. 2013), few studies to date have provided convincing evidence of TSP in MHC IIB genes in passerines. The strongest evidence for TSP in passerine MHC IIB genes was shown in a study of recently diverged Darwin's finches, where several well supported interspecific allelic lineages were described (Sato et al. 2011). Well supported interspecific clustering of variants was also reported in closely related species within the Muscicapidae family (Zagalska-Neubauer et al. 2010).

The unusually complex nature of the passerine MHC may partially explain why TSP is difficult to detect in this order. The MHC of passerines is among the most complex in terrestrial vertebrates, exhibiting highly duplicated classical MHC class I and II genes and more than ten functional class II B loci have been reported in some species. This contrasts sharply with other bird taxa such as those within the Galliformes, where one to three MHC IIB loci have been reported (Balakrishnan et al. 2010, Bollmer et al. 2010, Eimes et al. 2012, Sepil et al. 2012). It is likely that the unusually high levels of polymorphism observed in the MHC of songbirds is the result of a combination of several factors, including gene duplication, subsequent recombination (including gene conversion), and balancing selection (Hess and Edwards 2002). Because of the potential for recombination between non-homologous loci, the role of recombination, especially

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

gene conversion, may be exaggerated in passerines relative to other vertebrate taxa in shaping observed patterns of interspecific MHC variation between, as well as within species. The best evidence for this can be seen in phylogenetic trees of songbirds, where multi-locus MHC IIB sequences tend to cluster by species, rather than by locus (e.g. Sutton et al. 2013). This lack of trans-specific clustering of orthologous loci in passerines is thought to be due to concerted evolution as a result of gene conversion (Hess and Edwards 2002).

In this study, we characterized MHC IIB loci in three closely related species of crows: jungle crow (*Corvus macrorhynchos japonensis*), carrion crow (*C. corone orientalis*), and American crow (*C. brachyrhynchos*). We also tested for TSP and convergence by taking advantage of the unusual arrangement of phylogeny and geographical ranges of the three species. The contemporary range of eastern carrion crows and jungle crows overlaps in East Asia (southeastern China, the Korean peninsula and the Japanese archipelago) while American crows are limited to North America (Madge and Burn 1993, Haring et al. 2012). Mitochondrial DNA (mtDNA) phylogenies, however, place carrion crows and American crows in a monophyletic group, with jungle crows paraphyletic to these two species (Haring et al. 2012). The genus *Corvus* likely has a Palearctic origin, and mtDNA phylogenies suggest that carrion crows and American crows diverged in the Nearctic relatively recently; while jungle crows represent a single, older, lineage originating in the south-east Asian tropics (Haring et al. 2012).

Crows are suitable species for assessing TSP and convergence because their habitat, diet and social behavior are likely to render them susceptible to a diverse array of pathogens. Crows are omnivores that forage, in part, on carrion and human refuse (McGowan 2001); and scavengers, in general, could be under selection for more robust immune responses because of contact with pathogen-rich carcasses (Blount 2003). Furthermore, all three of the *Corvus* species

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

pyrosequencing. For the analyses, we used five different nucleotide partitions as well as supertyping, which clusters functionally equivalent amino acid sequences into units of selection (Sette and Sidney 1999, Lund et al. 2004, Sidney et al. 2008). We also estimated species divergence times for all three species using published, as well as newly generated, mtDNA Control Region (CR) sequences.

Materials and Methods

Sample collection

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

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

80 °C. Jungle crow RNA was extracted from whole blood using the RNeasy Protect Animal Blood System (Qiagen).

Amplification of MHC class IIB exon 2

In order to isolate MHC IIB exon 2, which codes for the PBR, jungle crow cDNA from two individuals (U2 and U3) was synthesized with the LongRange 2Step RT-PCR System (Qiagen) using the manufacturer's recommended amplification protocol. MHC IIB was then targeted in the cDNA using the primers MHC05 (Miller and Lambert 2004), which anneals within exon 1 in several passerines (positions variable), and *Coma*IIbex3R, which was designed by aligning the conserved regions of IIB exon 3 from several passerines (positions 19-38). These primers amplified a 359 bp cDNA fragment. We verified that the amplicon was MHC IIB by cloning the fragment from both individuals (TOPO XL One Shot-Invitrogen). Twenty-four clones from each individual were selected and whole colony PCR was performed using the TOPO vector primers M13 F and M13 R. The 20 μL PCR contained a pluck of colony cells as template, 0.625 μM of each primer, 0.5 *m*M of each dNTP, 1.5 *m*M MgCl2, 1X PCR buffer and 0.5 U of ExTaq (Takara) polymerase. The cycling conditions included an initial denaturation at 94 °C for 2 min followed by 30 cycles of 94, 55 and 72 °C each for 1 min and a final extension of 72 °C for 10 min. We sequenced each amplicon in both directions using M13 primers (Applied Biosystems 3130xl). We performed a BLAST search and identified eight sequences that aligned with known passerine MHC IIB.

We targeted intron 1 by aligning the cDNA-derived amplicons using MUSCLE in Geneious 6.1.4 (Biomatters) and a reverse primer, *Coma*IIbex2R, was designed in a conserved region of exon 2 and paired with MHC05. Using gDNA from two jungle crows (*Coma*5061 and

*Coma*7509), a 516 bp fragment extending from exon 1 to exon 2 was amplified using 50-100 ng of gDNA, 1.0 μM of each primer, 0.5 *m*M of each dNTP, 1.5 *m*M MgCl2, 1X PCR buffer and 0.5 U of ExTaq (Takara) polymerase (25 μL total volume). The cycling conditions included an initial denaturation at 94 °C for 30s followed by 30 cycles of 94 (20s), 60 (20s) and 72 (45s) °C and a final extension of 72 °C for 10 min. Each of these two amplicons was cloned and 10 colonies were sequenced in each direction using the cloning and whole colony PCR protocol described above. Sequences were aligned using MUSCLE and a forward primer flanking exon 2, *Coma*iF2, was designed for the (100%) conserved region at position 282-304 of intron 1. Next, we isolated intron 2 by pairing *Coma*iF2 with *Coma*IIBex3R using the same individuals and PCR, cloning, and whole colony PCR conditions described above. When amplifying from exon 2 to exon 3 in jungle crows, we discovered indels that resulted in large gaps at different locations in intron 2. 454 pyrosequencing is most efficient when sequences are of equal length, and for this reason, we designed a single degenerate reverse primer, *ComaEx2RA*, between positions 249-268 of exon 2. Finally, we confirmed that the primers *Coma*iF2 and *ComaEx2RA* amplified MHC IIB loci in all three crow species. Primer sequences and annealing temperatures are listed in Table S1.

454 pyrosequencing

Fusion primers were synthesized (Fasmac-Japan) by ligating the standard Roche multiplex identifiers (MIDs) to both the forward and reverse MHC IIB primers and the 454 adaptor primers (the Roche "Basic Amplicon" design). In order to minimize the formation of PCR artifacts, we reduced the cycle number and primer concentration and eliminated the final extension step during amplicon generation (Lenz and Becker 2008). PCRs contained 25-50 *n*gs of template, 0.5 μM of each fusion primer, 0.5 *m*M of each dNTP, 1.5 *m*M MgCl2, 1X PCR buffer, 5%

DMSO and 0.5 U of ExTaq (Takara) polymerase (25 μL total volume). The cycling conditions included an initial denaturation at 94 °C for 30s followed by 28 cycles of 94 (20s), 60 (20s) and 72 (45s) °C.

PCR amplicons were purified using the Agencourt AMPure XP System (Beckman Coulter) and recovered DNA was checked for quality and primer dimers on a 1.5% agarose gel. Samples were then quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) and pooled in equimolar amounts according to the Roche GS Junior Titanium Amplicon Library Preparation Method Manual. The quality and concentration of the pooled samples was checked on a 2100 Bioanalyzer (Agilent Technologies) and then sequenced using the Roche GS Junior Titanium Sequencing System at the Japanese National Museum of Nature and Science in Tsukuba, Japan.

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

Bioinformatics variant validation and genotyping

Variants were accepted as "true" if they occurred in at least three reads ($r > 2$) in both of two independent PCRs of the same individual. The criteria of ($r > 2$) is derived from the estimate of a read containing at least one sequence error for a 171 bp fragment of (HLA) DRB exon 2 of

~0.11 (Galan et al. 2010). Thus, the probability that an identical error occurs three times in the same sample is ~10^{-8} (Galan et al. 2010). At this stage, the most likely types of artefacts remaining in the data set are either single base substitutions or fragment chimeras generated during PCR. Although both of these artifacts are randomly generated, if these errors occur early in the PCR, they could be replicated and represented at relatively high frequencies in individual amplicons. To further reduce the probability of erroneously including single base substitutions, we classified sequences as true variants only when they differed by at least two nucleotide substitutions from more common variants. Galan et al. (2010) reported an average chimera frequency of 0.06 and some chimeras occurred at frequencies > 0.1; thus, elimination of chimeras using frequency thresholds may erroneously inflate variant estimates. It is highly unlikely that two independent PCRs would each generate identical chimeras (where as many as 10 different variants could recombine) or single base substitution errors three times ($r > 2$). By validating variants with a second PCR for each individual (rather than a second PCR within or across individuals) chimeras are eliminated, yet variants that are rare in the population are not erroneously eliminated from the data set. The total number of reads that were included at this stage was then designated as "net reads" for the purposes of genotype confidence level calculations.

We used the program 'Negative Multinomial' to calculate the number of sequences that are necessary for amplifying all the variants at least three times with a confidence level of 99% (Galan et al. 2010). While the program is limited to eight possible variants, the relationship between n, net reads, r, the number of times each variant must be sampled, and m, the maximum number of variants per amplicon, is linear. Our cloned sequences indicated a maximum of 20 variants per individual which corresponds to a minimum of 175 reads for 99.9% confidence;

however this assumes equal sampling of variants. Because amplification bias is likely when co-amplifying 10 loci, we expected this bias to be reflected in over- and under-representation of some variants (Burri et al. 2014). Therefore, we increased the minimum required net reads for genotyping to 275, which represents a frequency of 1.1% (amplification bias of ~ one order of magnitude) for a variant that has exactly three copies in an amplicon represented by 275 net reads.

Tests for selection and MHC IIB phylogenetics

For each species, we generated an alignment of validated variants, and identified the codons that have been shown to comprise the PBR of HLA in humans (Brown et al. 1993). While most MHC studies rely on the HLA PBR codons identified in humans to estimate selection, it has not been shown that these same codon positions apply universally or that selection is limited to peptide binding codons. Therefore, in order to identify specific codons that are likely under selection in crows, we performed a Wu-Kabat analysis on the amino acid alignment of verified alleles in all three species. A Wu-Kabat plot predicts amino acids that are likely subject to selection by identifying positions of high variability. Variability is calculated by dividing the number of amino acids at a position by the frequency of the most common amino acid (Wu and Kabat 1970). For each species, we identified codons/amino acids as likely under selection if the Wu-Kabat variability metric was ≥ twice the mean of Wu-Kabat values for all sites (Wu and Kabat 1970, Bos and Waldman 2006). These sites are denoted WuK-PS (PS = polymorphic site). We tested for selection on the HLA PBR sites by calculating the ratios of non-synonymous (*d*N) and synonymous mutations (*d*S) using a Z test (modified Nei-Gojobori- Jukes Cantor correction) in MEGA v. 6.0 (Tamura et al. 2011). For each species, individual codons were tested for selection using the HYPHY package in MEGA. This method, known as the counting method,

estimates the number of nonsynonymous and synonymous mutations that have occurred at each codon by using a maximum likelihood reconstruction of the ancestral state of each sequence. The test statistic *d*N - *d*S is used for detecting codons that have undergone selection and for positive values the probability of rejecting the null hypothesis of neutral evolution (*p*-value) is calculated (Suzuki and Gojobori 1999, Pond and Frost 2005). Significant *p* values are denoted as positively selected sites. This method is more computationally compatible with large data sets than empirical or hierarchal Bayesian approaches (Yang et al. 2000), yet has been shown to provide nearly identical results with a sufficient number of sequences (Pond and Frost 2005).

We constructed nucleotide phylogenies from five different partitions of exon 2: variable sites identified by the Wu-Kabot plot (WuK-PS, 33bp), non-WuK-PS sites (213 bp), HLA PBR sites (66 bp), non-HLA PBR sites (180 bp) and the contiguous 246 bp fragment of exon 2. For each analysis, 24 different nucleotide substitution models were tested for best fit (ML) in MEGA and the Jukes-Cantor (JC) Model + G (Gamma) + I (invariant sites) was chosen based on the Bayesian Information Criterion (BIC) score.

Supertyping

Functionally equivalent MHC IIB alleles were clustered using the supertyping method described in Sepil et al. (2012), a study that also examined highly duplicated MHC loci in a passerine species. Briefly, amino acid sites identified as positively selected sites were aligned and characterized by five physicochemical descriptor variables: z1 (hydrophobicity),z2 (steric bulk), z3 (polarity), z4 and z5 (electronic effects). These descriptor variables were placed into a matrix and subjected to K-means clustering algorithm (Sandberg et al. 1998, Doytchinova and Flower 2005). Discriminant analysis of principle components (DAPC) was used for describing clusters

in the 'adegenet' package in R (Jombart et al. 2010). We performed three different supertype analyses, using three different partitions: (i) the nine positively selected sites identified in the HYPHY analysis that were shared among all three species; (ii) WuK-PS sites that were shared among all three species; and (iii) HLA PBR codons.

DNA phylogenetics and species divergence

To ensure that the mtDNA phylogenies used in Haring et al. (2012) were consistent with our populations, we tested a subset of carrion crows (N = 4), jungle crows (N = 6) and American crows (N = 4) at the same mtDNA region using the primers CR-Cor+ and Phe-Cor- (Table S1) and the same PCR conditions as Haring et al. (2012), followed by Sanger sequencing as described above. We aligned these sequences with four mtDNA sequences from each species that were used in the original phylogenetic construction by Haring et al. (2012). The Hasegawa-Kishino-Yano (HKY) model (Lenz et al. 2013) had the lowest BIC scores in the best-fit substitution model (ML). HKY ML trees with 500 bootstrap replicates were constructed in MEGA v. 6.0.

Using the mtDNA sequences, we estimated divergence times for the three crow species: the divergence of carrion crow/American crow clade from jungle crow, and the divergence between carrion crows and American crows. Divergence times to the most recent common ancestor were estimated in BEAST v. 2.0 (Bouckaert et al. 2014). We used the HKY site model and a Yule process speciation prior for the branching rates. We tested the assumption of a molecular clock using the maximum likelihood method which compares the ML value for the given topology with and without the molecular clock constraints under the HKY model (MEGA v. 6.0). We ran two separate analyses using two different mtDNA substitution rates (per site, per My) previously

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

293 **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

PCRs of the same individual. All DNA sequences were deposited in NCBI Genbank: Accession numbers: MHC IIB: TBA; crow mtDNA: KM246294-308; crow nuclear intronic: KM246309-23

MHCIIB variation and selection

The 248 bp fragment was conserved at the final two base pair positions in all three species and was trimmed to 246 bp. The number of validated MHC IIB nucleotide variants for each species was: jungle crows = 89, carrion crows = 81 and American crows = 67. Individuals had a range of 7-20 variants, indicating that the MHC IIB in crows exhibits copy number variation (Table 1), although it is possible that the range of variant number could be explained by other phenomena, including null alleles, allele sharing across loci, homozygous vs. heterozygous loci or the variant validation methodology (where sequences with indels were excluded from the data set). Few variants were shared among species and there was no discernable pattern of allele sharing relating to phylogeny or sympatry: Four unique variants were shared among American crows and jungle crows, three among American crows and carrion crows, three among carrion crows and jungle crows and four variants were shared by all three species.

The codon-based Z test for selection (non-synonymous/synonymous mutations) averaged across the PBR yielded a significant *p* value for American crows and carrion crows but not for jungle crows (Table 1). The Wu-Kabot plot identified 11 amino acid sites that were highly variable in all three species: 6, 23, 25, 32, 33, 52, 61, 62, 65, 73 and 81. Eight of these sites (73 %) corresponded to known peptide binding sites in HLA (Fig. 1). Among the three species, jungle crows exhibited the most amino acid variation, with five hypervariable (>20 Wu-Kabat index) amino acid sites (Fig.1). The HYPHY test for selection identified nine positively selected

sites that were shared among the three species (2, 4, 33, 52, 55, 66, 68, 72 and 81), but only 5 of these (56%) corresponded to the HLA PBR (Fig. 1).

In total, we recovered 172 unique amino acid sequences among the three species (70 in jungle crows, 58 in carrion crows and 44 in American crows). We detected no species-specific motifs or motifs that were shared exclusively between two species. Within-species amino acid diversity was relatively high (0.245 ± 1.5), similar to that which we calculated for scrub jays (*Aphelocoma coerulescens*), another corvid species (23.3 %; N = 11; NCBI (U23958-65; 72,73,75) (Edwards et al. 1995). An alignment (82 sites) of common amino acid fragments from the three crow species with four other passerine species, and a falconiform species for comparison, highlights the high level of amino acid diversity found among and between species of passerines (Fig. S2).

MHC IIB phylogenies

We observed a similar pattern among the five different partitions of exon 2: MHC IIB variants formed several supported (bootstrap values ≥50), interspecific clusters in trees comparing putatively functional sites. Interspecific clusters were also observed when comparing putatively neutral sites, but there were fewer of these supported clusters, and in general, they were less inclusive. This pattern was most striking in trees constructed using the contiguous 246 bp exon 2 fragment (Fig. 3). For instance, a tree that was constructed by comparing non-synonymous substitutions (Fig. 3A) shows six well supported clusters, each containing all three crow species, whereas a tree constructed by comparing synonymous substitutions of the same fragment (Fig. 3B), shows only three such clusters. Supported clustering at putatively functional sites was also more inclusive in these exon 2 trees, with 44% of all variants (105/237) forming supported interspecific clades compared to just 18% (32/237) in the tree using synonymous sites. The trees constructed from WuK-PS and non-Wuk-PS sites (Fig. S3A and S3B) as well as those

constructed from the HLA PBR and non-PBR sites (Fig. S4A and S4B) displayed similar, though less distinct patterns. Among all partitions, the most common (supported) clustering was among all three species. We did not observe biased clustering of variants among the sympatric carrion crows and jungle crows; or among the more recently diverged, allopatric, American crows and carrion crows.

Supertyping

Among the three species, there were nine shared positively selected sites in exon 2. Five of these nine sites corresponded to HLA PBR sites (Fig. 1). From the 140 unique positively selected sites amino acid variants a total of eight supertypes were identified (Fig. 4). Only one supertype (Supertype 4) was limited to carrion crows and jungle crows, whereas the rest of the supertypes were shared among the three species. The supertypes generated by the WuK-PS and the HLA PBR codons were not well supported and thus excluded from our analysis.

mt*DNA phylogeny and divergence estimate*

The CR-Cor+ and Phe-Cor- primers amplified an 836-920 bp fragment of the mtDNA CR region in the three species of crows. The ML tree placed carrion crows and American crows in a separate clade from jungle crows with 100% bootstrap support (Fig. 2), thus confirming that the jungle crows and carrion crow populations from this study conform to the phylogeny of Haring et al. (2012). The ML-based test of the molecular clock failed to reject the null hypothesis of equal evolutionary rate throughout the tree ($p = 0.85$); thus we used a strict clock setting for the MCMC iterations in the BEAST program. Using the mtDNA substitution rate of 0.0567555/site/My the estimated mean divergence time to MRCA of all species was 0.8992 Ma with a 95% HPD interval of 0.7055, 1.1147 Ma (effective sample size (ESS) = 8581.7, median =

0.8942, standard deviation = 0.107). The estimated divergence to the MRCA of carrion crows and American crows was 0.4372 Ma with a 95% HPD interval of 0.3036, 0.5801 Ma (ESS = 7682.7, median = 0.4319, standard deviation = 0.0723) (Fig. 2). Divergence estimates using the higher substitution rate (*C. coronoides* = 0.0603431/site/My) yielded similar values: the MRCA of all species = 0.8509 Ma, HPD interval = 0.6578, 1.0486 Ma; MRCA of carrion crows and American crows = 0.4143 Ma, HPD interval = 0.2881, 0.5458 Ma.

Discussion

We observed strong interspecific clustering of MHC IIB variants in three closely related species of crows (*Corvus* sp.), and most of the supported clusters contained all three species; however, the clustering was much more inclusive, and better supported, when comparing non-synonymous substitutions at putatively selected sites than synonymous substitutions at putatively non-selected sites (Fig. 3; Fig. S3; Fig. S4). This pattern was observed in all of the phylogenies constructed from the different partitions of exon 2, but was most obvious in the trees constructed using the entire 246 fragment of exon 2 (Fig. 3). This pattern of strong clustering among functional sites is suggestive of convergent evolution (Klein et al. 2007). In fact, if this study been limited to the sympatric carrion crows and jungle crows, we may have erroneously concluded that convergent evolution explains these patterns; however, inclusion of sequences from the allopatric American crow, and the resulting interspecific clustering of all three species at functional sites, suggests that the pattern observed in the phylogenies is not convergence, but rather trans-species polymorphism.

Our results indicate that the MHC IIB in crows exhibits the same characteristics of the MHC IIB observed in other passerine species (Bollmer et al. 2010, Zagalska-Neubauer et al. 2010),

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

namely, high levels of polymorphism, highly duplicated loci, probable copy number variation of loci and a strong signature of selection. Consistent with similar studies, we also observed highly polymorphic amino acid positions and detected selection at sites other than the HLA PBR (Fig. 1), which indicates that either the PBR in passerines comprises different codon positions than the human MHC or that regions other than the PBR are under selection (Anmarkrud et al. 2010, Bollmer et al. 2010, Zagalska-Neubauer et al. 2010). The Wuk-Kabot plot was a better fit to the HLA PBR than the HYPHY probability based test for selection (Fig. 1), which may indicate that such probability tests are not reliable when identifying selected sites across many duplicated loci if the relationships between these loci are unknown (e.g. orthologous versus paralogous relationships). For example, it is unlikely that the hyper-polymorphic amino acid sites at positions 6, 23, 32, 62, and 65 are not subject to selection. This highlights the importance of using a combination of different exon 2 partitions when evaluating selection or establishing putative neutral and selected sites for phylogenetic analyses in complicated systems such as the passerine MHC.

Several factors may explain the relatively high MHC IIB neutral genetic variation between allelic lineages we observed in these crow species. First, if these allelic clusters represent ancient lineages, they are likely millions of years older than the MRCA of all three species, which allows sufficient time for random mutations to be incorporated, and the likelihood that these neutral mutations persist is increased by their close association with selected sites (Klein et al. 1998, O'huigin et al. 2000). Thus, the majority of the synonymous polymorphisms observed in neutral sites were likely inherited, rather than generated since species divergence. Other factors that may contribute to the relatively high synonymous nucleotide variation between species observed at non-selected sites relative to selected sites are the combined effects of

recombination through genetic exchange and gene conversion. In a study of MHC IIB in sticklebacks (Gasterosteidae), Lenz et al. (2013) observed a similar pattern of clustering of selected sites (in this case, the HLA PBR) but a lack of clustering in the remaining sites of exon 2, a pattern similar to the one reported here. The authors reasoned that selected sites may evolve under site-specific functional constraints that maintain similar amino acid motifs, and that any new variants generated by recombination are more likely to be retained if they share this functional similarity. Intraspecific recombination and gene conversion could then erode the signal of common allelic ancestry over time.

Gene conversion is thought to explain why passerine MHC sequences often sort by species, rather than by locus, on phylogenetic trees (Hess and Edwards 2002). Elevated rates of gene conversion have been documented in birds, and one study reported a gene conversion rate over a magnitude higher than the neutral mutation rate (Spurgin et al. 2011). It is possible that the reason why the crow MHC IIB sequences do not sort intra-specifically is that insufficient time has passed since species divergence for gene conversion to homogenize the neutral sites within exon 2. The patterns we observed were similar to those described for MHC IIB in Darwin's finches in that the sequences sorted in a manner that would be expected from a single, undifferentiated species (Sato et al. 2011). The similar results between our study and that of Sato et al. (2011) are likely due to the relatively recent divergence times of the focal taxa in both studies. In the Galapagos, Darwin finches' radiated from a founding population approximately 2.5 Ma, while the MRCA to the three crow species was estimated at approximately 900,000 years ago.

While we did not detect evidence of convergence between the sympatric carrion crow and jungle crow in the MHC IIB phylogenies, the supertype analysis revealed a pattern consistent

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

with convergent evolution. While seven of the eight supertypes were shared among the three species, one supertype, ST-4, was exclusive to carrion crows and jungle crows, which might reflect selective pressure from a specific, shared parasite. None of the other seven supertypes were exclusive to either a single species or group of two species. This supertyping method classifies MHC II B loci into clusters based on the physio-chemical properties of the amino acids in the positively selected sites of the PBR. Several studies of primate MHC have shown that different MHC alleles bind very similar peptide motifs, suggesting significant overlap in peptide binding repertoires (Sette and Sidney 1999, Lund et al. 2004, Sidney et al. 2008). Thus, different amino acid sequences can be functionally equivalent, and grouped together as a supertype, considered a unit of selection. Recently, (Sepil et al. 2013) showed that different MHC class I supertypes confer resistance to malaria in great tits (*Parus major*). In our study, supertypes ST-3 and ST-4 grouped together, away from the other six supertypes, which may suggest that these groups are functionally important for a specific type of parasite, whereas the other supertypes may confer resistance to more common, generalist parasites (Fig 4).

Conclusion

In the present study we describe strong evidence for trans-specific polymorphism at the MHC IIB in three species of crows. Our results are consistent with two other studies that examined MHC IIB in closely related passerine taxa and indicate that TSP in passerines may be common, but due to intra-specific gene conversion, most easily detectable among recently diverged species.

Given that the allopatric American crow and carrion crow diverged approximately 430,000 years ago, we expected to find some signal of convergence of MHC IIB genes between the

sympatric jungle crow and carrion crow in the MHC IIB phylogenies. While we did find patterns consistent with convergence in the phylogenetic analyses, any convergence would have had to occur among all three species, before the MRCA split approximately 900,000 years ago. While this scenario cannot be ruled out, the results presented here are better explained by TSP.

Acknowledgements

We wish to Naoyuki Takahata and Colm O'hUigín for advice during the experimental phase of the project and Peter Dunn for advice on the analysis and interpretations of results. We also thank Jon Fong for assistance of the phylogenetic analyses.

References

Alcaide, M., S. V. Edwards, and J. J. Negro. 2007. Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. *Journal of Molecular Evolution* 65:541-554.

Alcaide, M., M. Lui, and S. V. Edwards. 2013. Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ 1:e86 http://dx.doi.org/10.7717/peerj.86*

Ali, S. and S. Ripley. 1972. Handbook of the Birds of India and Pakistan. Oxford University Press, Bombay.

Andersson, L. and S. Mikko. 1995. Generation of MHC Class II Diversity by Intra- and Intergenic Recombination. *Immunological Reviews* 143:5-12.

Anmarkrud, J. A., A. Johnson, L. Bachman, and J. T. Lifjeld. 2010. Ancestral polymorphism in exon 2 of bluethroat (*Luscinia svecica*) MHC class II B genes. *Journal of Evolutionary Biology* 23:1206-1217.

Arendt, J. and D. Reznick. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology Evolution* 23:26-32.

Balakrishnan, C., R. Ekblom, M. Volker, H. Westerdahl, R. Godinez, H. Kotkiewicz, D. Burt, T. Graves, D. Griffin, W. Warren, and S. Edwards. 2010. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biology* 8:29.

Blount, J. D. 2003. Do individual branches of immune defence correlate? A comparative case study of scavenging and non-scavenging birds. *Oikos* 102:340-350.

Bollmer, J. L., P. O. Dunn, L. A. Whittingham, and C. Wimpee. 2010. Extensive MHC Class II B Gene Duplication in a Passerine, the Common Yellowthroat (Geothlypis trichas). *Journal of Heredity* 101:448-460.

499 Bos, D. H. and B. Waldman. 2006. Evolution by Recombination and Transspecies
500 Polymorphism in the MHC Class I Gene of Xenopus laevis. *Molecular Biology and*
501 *Evolution* 23:137-143.
502 Bouckaert, R., J. Heled, D. Kuhnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut,
503 and A. J. Drummond. 2014. BEAST 2: a software platform for Bayesian evolutionary
504 analysis. *PLoS Computional Biology* 10:e1003537.
505 Brown, J. H., T. S. Jardetzky, J. C. Gorga, L. J. Stern, R. G. Urban, J. L. Strominger, and D. C.
506 Wiley. 1993. 3-dimensional structure of the human class-II histocompatibility antigen
507 Hla-Dr1. *Nature* 364:33-39.
508 Bull, J. J. 1994. Perspective - Virulence. *Evolution* 48:1423-1437.
509 Burri, R., H. N. Hirzel, N. Salamin, A. Roulin, and L. Fumagalli. 2008. Evolutionary Patterns of
510 MHC Class II B in Owls and Their Implications for the Understanding of Avian MHC
511 Evolution. *Molecular Biology and Evolution* 25:1180-1191.
512 Burri, R., M. Promerova, J. Goebel, and L. Fumagalli. 2014. PCR-based isolation of multigene
513 families: lessons from the avian MHC class IIB. *Molecular Ecology Resources* 14:778-
514 788.
515 Cawthorn, R., R. Anderson, and I. Barker. 1980. Lesions caused by Diplotriaena tricuspis
516 (Nematoda: Diplotriaenoidea) in the American crow, Corvus brachyrhynchos Brehm.
517 *Canadian Journal of Zoology* 58:1892-1898.
518 Doytchinova, I. A. and D. R. Flower. 2005. In Silico Identification of Supertypes for Class II
519 MHCs. *Journal of Immunology* 174:7085-7095.
520 Edwards, S. V., M. Grahn, and W. K. Potts. 1995. Dynamics of Mhc evolution in birds and
521 crocodilians: amplification of class II genes with degenerate primers. *Molecular Ecology*
522 4:719-729.
523 Edwards, S. V., E. K. Wakeland and W. K. Potts. 1995. Contrasting histories of avian and
524 mammalian Mhc genes revealed by class II B sequences from songbirds. Proceedings of
525 the National Academy of Sciences, USA. 92: 12200-4.
526 Eimes, J. A., K. M. Reed, K. M. Mendoza, J. L. Bollmer, L. A. Whittingham, Z. W. Bateson, and
527 P. O. Dunn. 2012. Greater prairie-chickens have a compact MHC-B with a single class
528 1A locus. *Immunogenetics* 87:435-435.
529 Frank, S. A. 1996. Models of parasite virulence. *Quarterly Review of Biology* 71:37-78.
530 Galan, M., E. Guivier, G. Caraux, N. Charbonnel, and J. F. Cosson. 2010. A 454 multiplex
531 sequencing method for rapid and reliable genotyping of highly polymorphic genes in
532 large-scale studies. *BMC Genomics* 11:296.
533 Garrigan, D. and P. W. Hedrick. 2003. Detecting adaptive molecular polymorphism: lessons
534 from the MHC. *Evolution* 57:1707-1722.
535 Goüy de Bellocq, J., N. Charbonnel, and S. Morand. 2008. Coevolutionary relationship between
536 helminth diversity and MHC class II polymorphism in rodents. *Journal of Evolutionary*
537 *Biology* 21:1144-1150.
538 Graser, R., C. OhUigin, V. Vincek, A. Meyer, and J. Klein. 1996. Trans-species polymorphism
539 of class II Mhc loci in danio fishes. *Immunogenetics* 44:36-48.
540 Gustafsson, K. and L. Andersson. 1994. Structure and polymorphism of horse MHC class II
541 DRB genes: convergent evolution in the antigen binding site. *Immunogenetics* 39:355-
542 358.
543 Haring, E., B. Daubl, W. Pinsker, A. Kryukov, and A. Gamauf. 2012. Genetic divergences and
544 intraspecific variation in corvids of the genus Corvus (Aves: Passeriformes: Corvidae) - a

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

first survey based on museum specimens. *Journal of Zoological Systematics and Evolutionary Research* 50:230-246.
Hendricks, L., R. Harkema, and G. Miller. 1969. Helminths of the Crow, Corvus brachyrhynchos Brehm, 1822, in North Carolina1. *Proceedings of the Helminthological Society of Washington* 36:150-152.
Hess, C. and S. Edwards. 2002. The evolution of the major histocompatibility complex in birds. *Bioscience* 52:423-431.
Hughes, A. L. 1999. Adaptive evolution of genes and genomes. Oxford University Press, New York, New York.
Janeway, C., P. Travers, M. Walport, and M. Shlomchick. 2005. Immunobiology: the immune system in health and disease. Garland Press, New York.
Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *Bmc Genetics* 11.
Jones, J. 1968. Some Parasites of the Common Crow, Corvus Brachyrhynchos Brehm, from Ohio. *The Ohio Journal of Science* 68:25-31.
Klein, D., H. Ono, C. O'HUigin, V. Vincek, T. Goldschmidt, and J. Klein. 1993. Extensive MHC variability in cichlid fishes of Lake Malawi. *Nature* 364:330-334.
Klein, J. 1986. The natural history of the major histocompatibility complex. John Wiley, New York.
Klein, J., A. Sato, S. Nagl, and C. O'hUigín. 1998. Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics* 29:1-21.
Kriener, K., C. O'HUigin, H. Tichy, and J. Klein. 2000. Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics* 51:169-178.
Kupfermann, H., Y. Satta, N. Takahata, H. Tichy, and J. Klein. 1999. Evolution of Mhc-DRB introns: Implications for the origin of primates. *Journal of Molecular Evolution* 48:663-674.
Kurosawa, R., R. Kono, T. Kondo, and Y. Kanai. 2003. Diet of jungle crows in an urban landscape. *Global Environmental Research* 7:193-198.
Lenz, T. L. and S. Becker. 2008. Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - Implications for evolutionary analysis. *Gene* 427:117-123.
Lenz, T. L., C. Eizaguirre, M. Kalbe, and M. Milinski. 2013. Evaluating patterns of convergent evolution and trans-species polymorphism at MHC immunogenes in two sympatric stickleback species. *Evolution* 67:2400-2412.
Lund, O., M. Nielsen, C. Kesmir, A. G. Petersen, C. Lundegaard, P. Worning, C. Sylvester-Hvid, K. Lamberth, G. Roder, S. Justesen, S. Buus, and S. Brunak. 2004. Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* 55:797-810.
Madge, S. and H. Burn. 1993. Crows and jays. Helm, London, UK.
Martinsohn, J. T., A. B. Sousa, L. A. Guethlein, and J. C. Howard. 1999. The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* 50:168-200.
McGowan, K. 2001. Demographic and behavioral comparisons of suburban and rural American Crows. Kluwer Academic Press, Norwell, MA.
Miller, A. D., A. K. Townsend, K. J. McGowan, A. Clark, A. L. Glaser, L. A. Patrican, E. Dobson, and E. L. Buckels. 2010. Non-west nile virus-associated mortality in a poulation

of Amercian crows (*Corvus brachyrhynchos*): a gross and histopathologic study. *Journal of Veterinary Diagnostic Investigation* 22:289-295.
Miller, H. C. and D. M. Lambert. 2004. Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics* 56:178-191.
Mizuno, F. 1984. Studies on the parasite fauna of the eastern carrion crow, Corvus corone orientalis Eversman, and the Japanese jungle crow, Corvus macrorhynchos japonensis *Japanese Journal of Veterinary Research* 32:105.
Moller, A. P. 2009. Successful city dwellers: a comparative study of the ecological characteristics of urban birds in the Western Palearctic. *Oecologia* 159:849-858.
Møller, A. P., S. Merino, C. R. Brown, and R. J. Robertson. 2001. Immune Defense and Host Sociality: A Comparative Study of Swallows and Martins. *American Naturalist* 158:136-145.
Nabholz, B., S. Glemin, and N. Galtier. 2009. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *Bmc Evolutionary Biology* 9.
Naderman, J. and D. Pence. 1980. Helminths of the Common Crow, Corvus brachyrhynchos Brehm, from West Texas. *Proceedings of the Helminthological Society of Washington* 47:100-105.
O'huigin, C., Y. Satta, A. Hausmann, R. L. Dawkins, and J. Klein. 2000. The implications of intergenic polymorphism for major histocompatibility complex evolution. *Genetics* 156:867-877.
Pearce, T. 2011. Convergence and Parallelism in Evolution: A Neo-Gouldian Account. *The British Journal for the Philosophy of Science* 63:429-448.
Pond, S. L. K. and S. D. W. Frost. 2005. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* 22:1208-1222.
Richardson, D. S. and H. Westerdahl. 2003. MHC diversity in two Acrocephalus species: the outbred Great reed warbler and the inbred Seychelles warbler. *Molecular Ecology* 12:3523-3529.
Richner, H. 1989. Habitat-Specific Growth and Fitness in Carrion Crows (Corvus-Corone-Corone). *Journal of Animal Ecology* 58:427-440.
Sandberg, M., L. Eriksson, J. Jonsson, M. Sjostrom, and S. Wold. 1998. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *Journal of Medicinal Chemistry* 41:2481-2491.
Sato, A., W. E. Mayer, H. Tichy, P. R. Grant, B. R. Grant, and J. Klein. 2001. Evolution of Mhc class II B genes in Darwin's finches and their closest relatives: birth of a new gene. *Immunogenetics* 53:792-801.
Sato, A., H. Tichy, P. R. Grant, B. R. Grant, T. Sato, and C. O,ÄôhUigin. 2011. Spectrum of MHC Class II Variability in Darwin's Finches and Their Close Relatives. *Molecular Biology and Evolution* 28:1943-1956.
Sepil, I., S. Lachish, A. E. Hinks, and B. C. Sheldon. 2013. Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proceedings of the Royal Society B-Biological Sciences* 280: 20130134.
Sepil, I., H. K. Moghadam, E. Huchard, and B. C. Sheldon. 2012. Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *Bmc Evolutionary Biology* 12.

637 Sette, A. and J. Sidney. 1999. Nine major HLA class I supertypes account for the vast
638 preponderance of HLA-A and -B polymorphism. *Immunogenetics* 50:201-212.
639 Sidney, J., B. Peters, N. Frahm, C. Brander, and A. Sette. 2008. HLA class I supertypes: a
640 revised and updated classification. *BMC Immunology* 9:1.
641 Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology
642 and conservation. *Frontiers in Zoology* 2:16.
643 Spurgin, L. G., C. van Oosterhout, J. C. Illera, S. Bridgett, K. Gharbi, B. C. Emerson, and D. S.
644 Richardson. 2011. Gene conversion rapidly generates major histocompatibility complex
645 diversity in recently founded bird populations. *Molecular Ecology* 20:5213-5225.
646 Srithayakumar, V., S. Castillo, J. Mainguy, and C. J. Kyle. 2012. Evidence for evolutionary
647 convergence at MHC in two broadly distributed mesocarnivores. *Immunogenetics*
648 64:289-301.
649 Stuglik, M. T., J. Radwan, and W. Babik. 2011. jMHC: software assistant for multilocus
650 genotyping of gene families using next-generation amplicon sequencing. *Molecular*
651 *Ecology Resources* 11:739-742.
652 Sutton, J. T., B. C. Robertson, C. E. Grueber, J. A. L. Stanton, and I. G. Jamieson. 2013.
653 Characterization of MHC class II B polymorphism in bottlenecked New Zealand
654 saddlebacks reveals low levels of genetic diversity. *Immunogenetics* 65:619-633.
655 Suzuki, Y. and T. Gojobori. 1999. A method for detecting positive selection at single amino acid
656 sites. *Molecular Biology and Evolution* 16:1315-1328.
657 Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5:
658 Molecular evolutionary genetics analysis using likelihood, distance and parsimony
659 methods. *Molecular Biology and Evolution* 28:2731-2739.
660 Verbeek, N. and C. Caffrey. 2002. The American Crow (*Corvus brachyrhynchos*).*in* A. Poole
661 and F. Gill, editors. Birds of North America. Birds of North America, Inc., Philadelphia.
662 Vincek, V., C. O'Huigin, and J. Klein. 1997. How large was the founding population of Darwin's
663 finches? *Proceedings of the Royal Society B-Biological Sciences.* 264:111-118.
664 Wheeler, S. S., L. W. Woods, W. M. Boyce, C. D. Eckstrand, S. A. Langevin, W. K. Reisen, and
665 A. K. Townsend. 2014. West Nile virus and non-West Nile virus mortality and
666 coinfection of American crows (Corvus brachyrhynchos) in California. *Avian Diseases*
667 58:255-261.
668 Wu, T. T. and E. A. Kabat. 1970. An analysis of the sequences of the variable regions of Bence
669 Jones proteins and myeloma light chains and their implications for antibody
670 complementarity. *Journal of Experimental Medicine* 132:211-250.
671 Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen. 2000. Codon-substitution models for
672 heterogeneous selection pressure at amino acid sites. *Genetics* 155:431-449.
673 Yeager, M. and A. L. Hughes. 1999. Evolution of the mammalian MHC: natural selection,
674 recombination, and convergent evolution. *Immunological Reviews* 167:45-58.
675 Zagalska-Neubauer, M., W. Babik, M. Stuglik, L. Gustafsson, M. Cichon, and J. Radwan. 2010.
676 454 sequencing reveals extreme complexity of the class II Major Histocompatibility
677 Complex in the collared flycatcher. *Bmc Evolutionary Biology* 10:395.
678 Zelano, B. and S. V. Edwards. 2002. An Mhc component to kin recognition and mate choice in
679 birds: predictions, progress, and prospects. *American Naturalist* 160:S 225- 237.
680 Zhang, J. and S. Kumar. 1997. Detection of convergent and parallel evolution at the amino acid
681 sequence level. *Molecular Biology and Evolution* 14:527-536.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

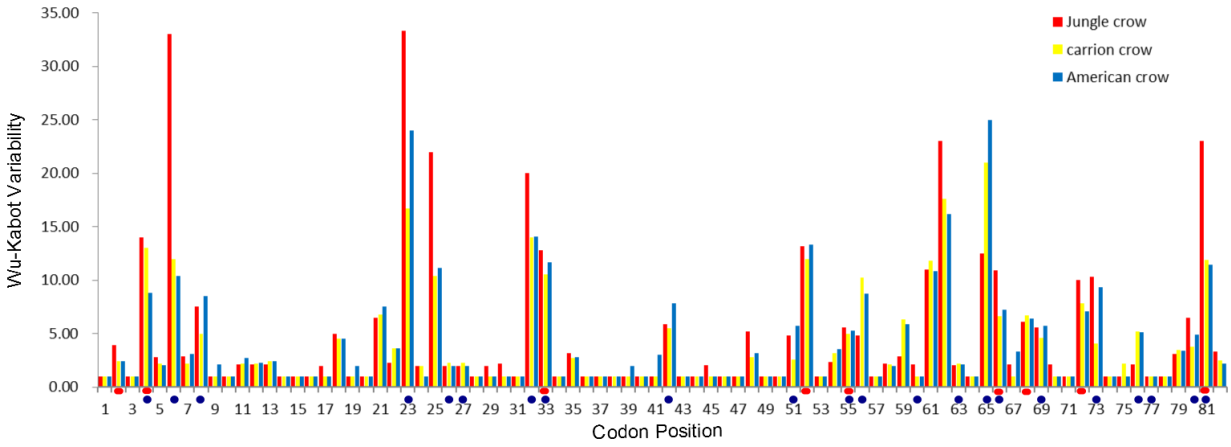
Zhang, M. and H. He. 2013. Parasite-mediated selection of major histocompatibility complex variability in wild brandt's voles (Lasiopodomys brandtii) from Inner Mongolia, China. *Bmc Evolutionary Biology* 13:149.

Figures and Tables

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.

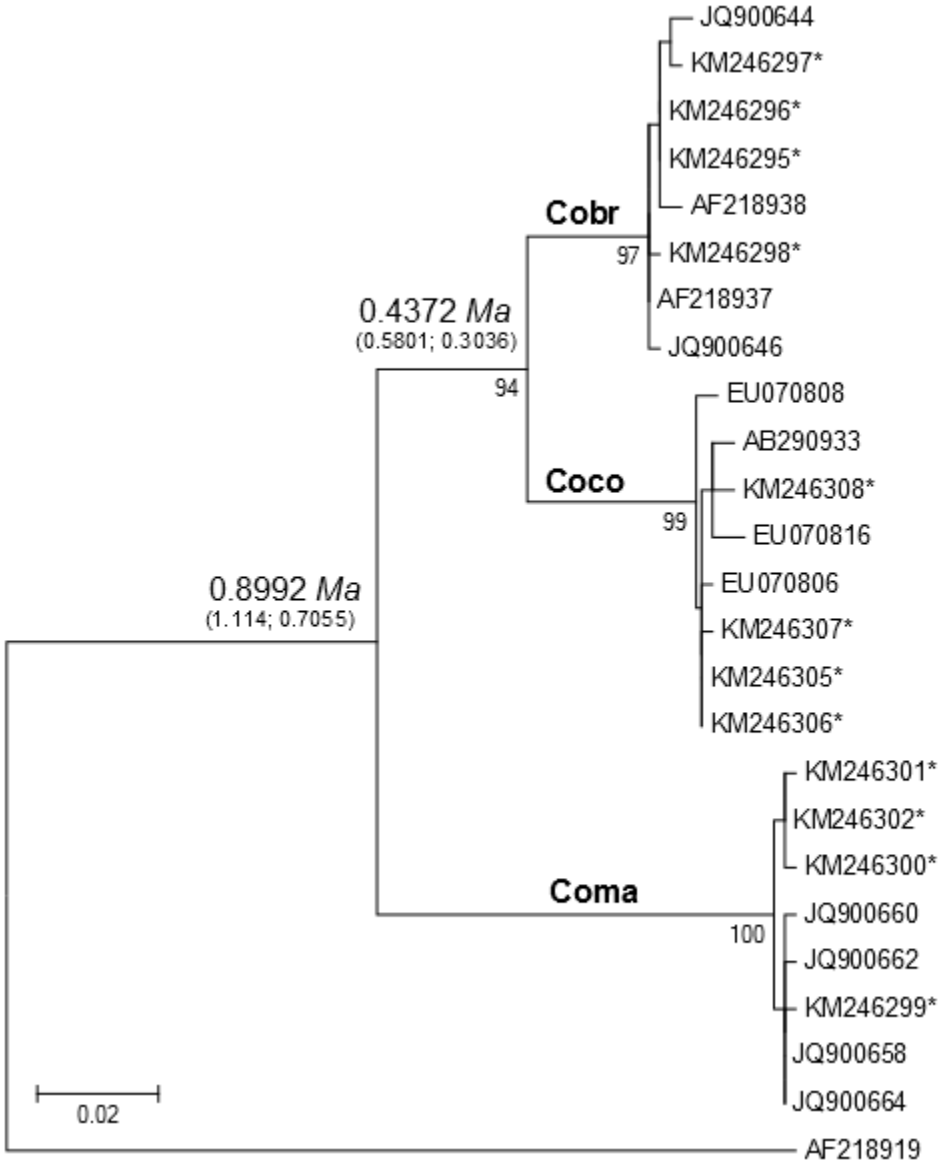
Species	Alleles (total)	Alleles (Indiv)	π	π_SPBR	π_NPBR	dN/dS PBR	*p* value Z-test
Jungle	89	7-18	0.16	0.29 ± 0.115	0.47 ± 0.09	1.6	0.09
Carrion	81	11-20	0.15	0.24 ± 0.02	0.45 ± 0.1	1.9	0.028
American	67	11-18	0.14	0.25 ± 0.02	0.44 ± 0.09	1.6	0.047

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

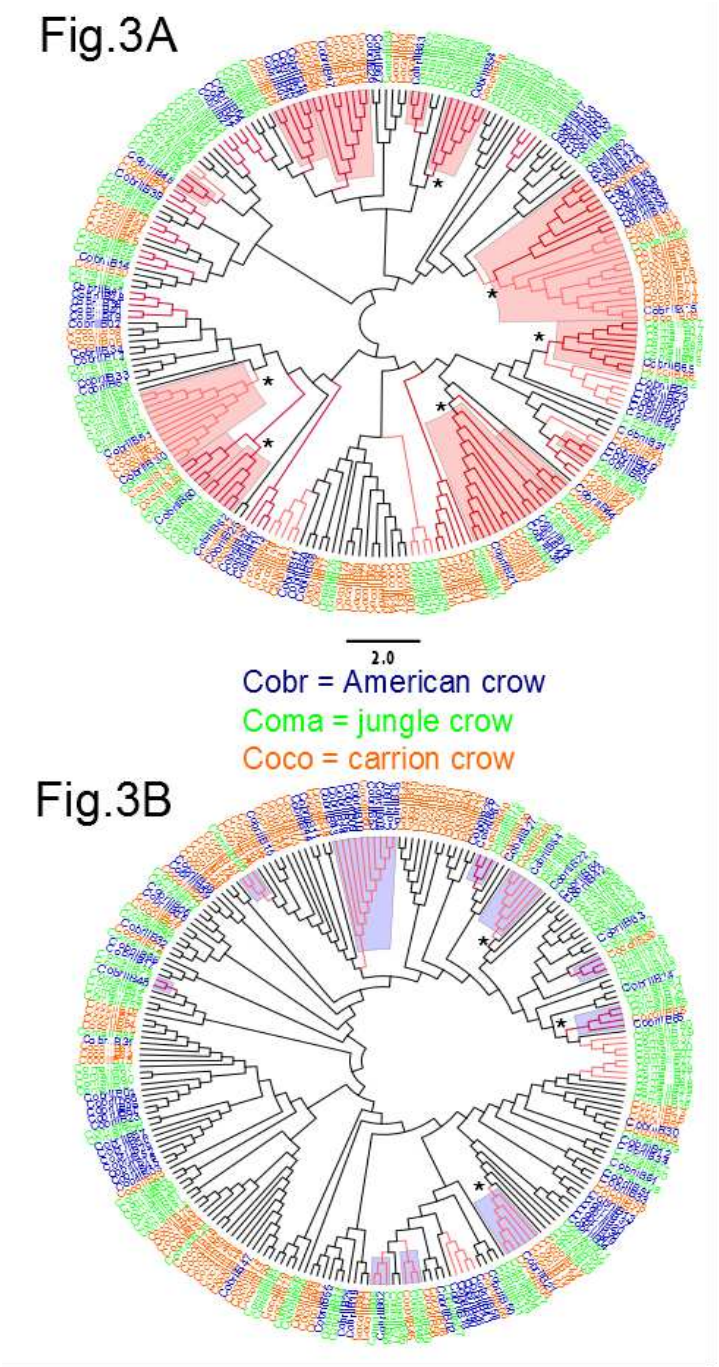


PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

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



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

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.621v2 | CC-BY 4.0 Open Access | rec: 8 Mar 2015, publ: 8 Mar 2015

Figure 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.