The Evolution of Innate Immune Genes: Purifying and Balancing Selection on β-Defensins in Waterfowl

Joanne R. Chapman,*[,1] Olof Hellgren,[2] Anu S. Helin,[1] Robert H. S. Kraus,[3,4] Ruth L. Cromie,[5] and Jonas Waldenström*[,1]

[1]Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, Kalmar, Sweden

[2]Molecular Ecology & Evolution Lab (MEMEG), Department of Biology, Lund University, Lund, Sweden

[3]Department of Biology, University of Konstanz, Konstanz, Germany

[4]Department of Migration and Immuno-Ecology, Max Planck Institute for Ornithology, Radolfzell, Germany

[5]Wildfowl & Wetlands Trust, Slimbridge, Gloucestershire, United Kingdom

***Corresponding authors:** E-mails: joanne.chapman@lnu.se; jonas.waldenstrom@lnu.se.

Associate editor: Dr. Ryan Hernandez

Abstract

In disease dynamics, high immune gene diversity can confer a selective advantage to hosts in the face of a rapidly evolving and diverse pathogen fauna. This is supported empirically for genes involved in pathogen recognition and signalling. In contrast, effector genes involved in pathogen clearance may be more constrained. β-Defensins are innate immune effector genes; their main mode of action is via disruption of microbial membranes. Here, five β-defensin genes were characterized in mallards (*Anas platyrhynchos*) and other waterfowl; key reservoir species for many zoonotic diseases. All five genes showed remarkably low diversity at the individual-, population-, and species-level. Furthermore, there was widespread sharing of identical alleles across species divides. Thus, specific β-defensin alleles were maintained not only spatially but also over long temporal scales, with many amino acid residues being fixed across all species investigated. Purifying selection to maintain individual, highly efficacious alleles was the primary evolutionary driver of these genes in waterfowl. However, we also found evidence for balancing selection acting on the most recently duplicated β-defensin gene (*AvBD3b*). For this gene, we found that amino acid replacements were more likely to be radical changes, suggesting that duplication of β-defensin genes allows exploration of wider functional space. Structural conservation to maintain function appears to be crucial for avian β-defensin effector molecules, resulting in low tolerance for new allelic variants. This contrasts with other types of innate immune genes, such as receptor and signalling molecules, where balancing selection to maintain allelic diversity has been shown to be a strong evolutionary force.

Key words: **Antimicrobial peptides, host defense peptides, ecoimmunology, avian immune system, host-pathogen dynamics.**

Introduction

Antagonistic co-evolution between hosts and their pathogens is one of the major driving forces of molecular evolution of species (Paterson et al. 2010). As a consequence, host species are thought to maintain high standing allelic variation at immunity and disease resistance loci, to counter a rapidly evolving and diverse pathogen fauna (Haldane 1949; Sommer 2005). Indeed, genome-wide selection scans in vertebrates have demonstrated that one of the main classes of genes with evidence for positive (i.e., directional, balancing, or diversifying) selection are those with immune-related functions (e.g., Andrés et al. 2009; Pickrell et al. 2009; Barreiro and Quintana-Murci 2010). What is less well understood is the extent and role of allelic variation in individual immune genes, within and between species, and the role this plays in disease susceptibility (Magor and Magor 2001). In this regard, information pertaining to genes of the innate immune system are particularly lacking (Gura 2001; Vinkler and Albrecht 2009), especially with respect to non-model and non-human species (Magor and Magor 2001; Lazarus et al. 2002). This is despite the pivotal roles the innate immune system plays in pathogen recognition, presentation of pathogens to the adaptive immune system, and direct antimicrobial activities that can mitigate disease (Hoffmann et al. 1999).

There are two opposing views as to the main evolutionary forces acting on innate immune genes (Mukherjee et al. 2014). The first states that because innate immunity is ancient and crucial, selection has had sufficient time to select the most efficacious alleles and there is low tolerance for new variants; as such purifying selection is the main evolutionary driver of innate immune genes (e.g., Mukherjee et al. 2009; Majumder 2010). The other view states that given the rapid rate of pathogen evolution, selection for high allelic diversity enhances the flexibility and broad spectrum antimicrobial activity that typifies the innate immune system and, therefore, diversifying and balancing selection are the main evolutionary drivers (e.g., Hughes et al. 2005;

Ferrer-Admetlla et al. 2008). The reality is likely not so clear cut, with different genes undergoing different selective forces, and demographic features such as host population density, mating system, and pathogen load having important effects. In particular, it is likely that selective pressures differ between genes with different functional properties in the innate immune cascade, specifically whether genes are involved in sensing (the afferent arm) or eliminating (the efferent arm) infection (Beutler 2004). Afferent molecules, involved in pathogen recognition and signalling, need to recognize a diverse array of pathogens and have flexibility to adapt to emerging/evolving microbes, thus continual or episodic balancing selection to maximize diversity may be particularly important. In contrast, efferent molecules, with essential roles in pathogen elimination, may be more constrained and display low tolerance for new genetic variants, and, therefore, be maintained by purifying selection. Additionally, patterns of selection may be shaped by which major pathogen group(s) a given immune gene has coevolved with. Perhaps the best example of opposing selective forces acting on different functional components of the innate immune repertoire is the family of Toll-like receptors (TLRs). TLRs are receptor and signalling molecules that comprise two distinct domains. The extracellular leucine-rich repeat (LRR) domain is responsible for recognizing and binding pathogen ligands and is subject to balancing selection, whereas the intracellular Toll interleukin-receptor (TIR) domain is involved in signalling to other components of the innate immune cascade and is subject to purifying selection (Barreiro et al. 2009; Werling et al. 2009; Alcaide and Edwards 2011; Tschirren et al. 2011; Mikami et al. 2012; Grueber et al. 2014). These results demonstrate that selective forces can vary, even across small genomic scales, when the functional properties of innate immune genes differ.

Signatures of selection on other components of the innate immune system have been less well characterized. In particular, a better understanding of the selective pressures acting on microbiocidal molecules is of interest because they can rapidly terminate or mitigate infections before onset of disease. The family of host defense peptides (HDPs, also known as antimicrobial peptides, AMPs) are particularly interesting in this respect because they comprise one of the most ancient forms of antimicrobial defense (Wiesner and Vilcinskas 2010), are present in all three eukaryotic kingdoms (Zhu 2008), and have direct antimicrobial activity, for example, by interfering with microbial membranes and lipid coats (Brogden 2005; Klotman and Chang 2006). Furthermore, HDPs are well characterized with respect to structure, function and activity (reviewed in Zasloff 2002; Yeaman and Yount 2003; Brogden 2005; Melo et al. 2009; Nguyen et al. 2011; Li et al. 2012). HDPs are increasingly recognized as potential therapeutic agents, in some cases offering a potential alternative to conventional antibiotics to which resistance is a global problem (Hancock and Sahl 2006; Zhang and Sunkara 2014). An increased understanding of the evolutionary and molecular properties of these peptides is, therefore, crucial. Despite this, evolutionary analyses of HDPs have been rare, but provide the opportunity to elucidate the genetic basis of Darwinian evolutionary processes by linking immunogenetic diversity with functional properties of individual peptides (Tennessen 2005; Hellgren 2015). Predicting the likely selective forces acting on HDPs is not straight forward. Although they have an important efferent role in the direct killing and clearance of pathogens, which might promote purifying selection, they also have multifaceted roles in signalling to and recruiting other molecules of the immune system, and as such have complex immunomodulatory effects (reviewed in Hancock et al. 2016), which might promote balancing selection.

One of the main types of HDPs in vertebrates are defensins, small cationic amphipathic peptides of less than 100 amino acid residues that are characterized by a highly conserved motif of six cysteine (C) residues in the mature (functional) peptide, which pair to form three intramolecular disulphide bonds (Lai and Gallo 2009) that stabilize the molecule and help to protect from proteolysis (Campopiano et al. 2004). Defensins are further subdivided into α-, β-, and θ-defensins based on secondary and tertiary structure and the linking pattern of C residues to form disulphide bonds; all are activated by post-translational cleaving of the mature peptide from a propiece and signal peptide (Lai and Gallo 2009; Hellgren and Ekblom 2010). Their primary mode of action is the killing of pathogens via disruption of membranes (Brogden 2005; Lai and Gallo 2009); however, they are recognized to have multifaceted roles in immune defense (Gura 2001; Sugiarto and Yu 2004; Funderburg et al. 2007; Lai and Gallo 2009).

Here, we characterize patterns of immunogenetic diversity and selection on five β-defensin genes in wild mallards and other members of the waterfowl family Anatidae (Ducks, Geese, and Swans). β-defensins are the only class of defensin in birds and are the most ancient of the three classes (van Dijk et al. 2008); α-defensins are found exclusively in mammals and θ-defensins in some primates (Ganz 2003; Zhang and Sunkara 2014). The fundamental importance of β-defensins in impeding pathogens has been demonstrated in humans (e.g., Quiñones-Mateu et al. 2003; Wehkamp et al. 2005; Hazrati et al. 2006; Funderburg et al. 2007; Jarczak et al. 2013; Segat et al. 2014), and to a lesser extent in birds (Soman et al. 2009; Hellgren et al. 2010; Ma et al. 2011). However, the extent and role of allelic variation in wild bird populations has to date only been studied at a very limited population scale (Hellgren 2015) and an understanding of HDP diversity within and between wild avian populations and species, and across large spatial scales, is, therefore, lacking. Improving knowledge of the avian immune system is valuable, particularly with respect to waterfowl, given the interest in this taxon in relation to zoonotic pathogens such as avian influenza viruses, *Salmonella* spp. and *Campylobacter* spp. (Reed et al. 2003; Kruse et al. 2004; Causey and Edwards 2008). We address these current gaps in knowledge by surveying natural allelic variation in five avian β-defensin (AvBD) genes at two levels: (1) intra-specific variation among mallard ducks (*Anas platyrhynchos*) from local and global populations; and (2) inter-specific variation among diverse members of the waterfowl; and discuss patterns of immunogenetic diversity in the light of evolutionary forces acting on these genes in waterfowl.

Results and Discussion

Intra-Specific Variation: Limited Genetic Diversity of Mallard AvBD Genes

We characterized intra- and inter-population allelic diversity of five AvBD genes (*AvBD3b*, *AvBD4*, *AvBD5*, *AvBD10*, and *AvBD13*, mature peptide only) in mallards in order to compare and contrast patterns of diversity across their natural range. First, to characterize genetic diversity at a local, intra-population scale, we genotyped 274 wild mallards caught at Ottenby Bird Observatory, south-eastern Sweden. With the exception of *AvBD10*, this sampling was sufficient to detect the majority of alleles likely present in the sampled population (supplementary fig. S1, Supplementary Material online). Some individuals failed to amplify at certain loci, final sample sizes per locus were the following: *AvBD3b* = 256, *AvBD4* = 212, *AvBD5* = 238, *AvBD10* = 243, and *AvBD13* = 245. The mean ± SD number of loci amplified per individual was 4.32 ± 0.95. Patterns of allelic diversity at the amino acid level were similar for all five loci, whereby we detected 2–7 alleles per locus (fig. 1 and supplementary table S1, Supplementary Material online). In contrast, at the nucleotide level, many more alleles were detected for *AvBD10* (44 alleles) than the other four loci (7–12 alleles, fig. 1 and supplementary table S1, Supplementary Material online). At the nucleotide level, individual heterozygosity was generally low, with the exception of *AvBD10* for which nearly all (86%) of the individuals were heterozygous (fig. 2*A*). The percentage of individuals possessing at least one copy of the main (most frequent) nucleotide allele was much lower for *AvBD10* (51%) and *AvBD3b* (67%) than the other three loci (96–99%, fig. 2*B*). At the amino acid level, for the genes *AvBD4*, *AvBD5*, *AvBD10*, and *AvBD13*, a single allele was extremely common (fig. 1), whereby 99–100% of individuals possessed at least one copy of the most common amino acid allele and heterozygosity was low (0.5–5% of individuals, fig. 2). In contrast, for *AvBD3b* the most common amino acid allele was present in only 69% of individuals (with 46% of individuals being homozygous for this allele), and heterozygosity was notably higher (25%, fig. 2). In addition, a second *AvBD3b* allele was common, whereby 42% of individuals possessed at least one copy and 25% of individuals were homozygous for this allele (figs. 1 and 2).

Second, to characterize genetic diversity at a global, inter-population scale, 190 individuals from 16 global populations, representing the natural range of mallards, were genotyped (supplementary table S2, Supplementary Material online). Due to the non-amplification of certain individuals for some loci, final sample sizes were as follows: *AvBD3b* = 177, *AvBD4* = 170, *AvBD5* = 185, *AvBD10* = 172, and *AvBD13* = 168. The mean ± SD number of loci amplified per individual was 4.62 ± 0.77.

In comparison with the more extensive but geographically restricted sampling of mallards in Sweden, additional global amino acid alleles were detected for *AvBD3b* (five new alleles) and *AvBD13* (four new), but not for *AvBD10, AvBD4*, or *AvBD5*: the latter two loci also had the fewest amino acid alleles detected (two and four, respectively), suggesting little global genetic diversity for these two loci (supplementary table S1, Supplementary Material online). Global patterns of allelic diversity were very similar to those found on a local scale: the single common amino acid allele detected on a local scale in Sweden for *AvBD4*, *AvBD5*, *AvBD10*, and *AvBD13* was in fact common globally (fig. 3*A* and supplementary figs. S2 and S3, Supplementary Material online), whereas the two common *AvBD3b* alleles in Sweden were also common globally, although the relative frequency of these two alleles differed somewhat between populations (fig. 3*B*).

The limited global allelic diversity suggests that single AvBD alleles have been selected and maintained as the most efficacious in populations that likely differ in regional pathogen prevalence and composition. This fits with the primary non-specific mode of action for AvBDs, whereby their main target (microbial membranes) are likely much more highly conserved than other microbial features. Thus, increased AvBD diversity may not result in activity against a wider range of pathogens, rather selection may favor maintenance of the most efficacious alleles with the highest broad spectrum activity against microbial membranes. However, it is increasingly recognized that defensins and other HDPs have multifaceted roles in immune defense (Arnett and Seveau 2011; Hilchie et al. 2013) such as preventing cell wall synthesis (Yount and Yeaman 2013), disrupting secretion of bacterial toxins (Vega and Caparon 2012), and an array of immunomodulatory activities (reviewed in Choi et al. 2012; Hilchie et al. 2013). Given that defensins have complex and varied roles in the immune response to different groups of pathogens, and that prevalence and pathogenicity of these groups (e.g., fungi versus bacteria) will vary geographically (e.g., with climate), it is thus somewhat surprising that regional differences in AvBD alleles are not observed, even at continental scales.

Inter-Specific Variation: Allele Sharing across Species Divides

In order to assess the variability of AvBD alleles across species of waterfowl, we genotyped the same five β-defensin genes in two individuals from 43 species of waterfowl (Family Anatidae, supplementary table S3, Supplementary Material online). Primers designed to amplify mallard AvBD genes (supplementary table S4, Supplementary Material online) were successfully applied in other waterfowl species, with at least two loci successfully amplified in every species tested (supplementary table S5, Supplementary Material online). Between 4 (*AvBD4* and *AvBD13*) and 30 (*AvBD3b*) new amino acid alleles were detected in non-mallard waterfowl (supplementary table S1, Supplementary Material online). Nucleotide and haplotype diversity were much higher for *AvBD10* than the other four loci (supplementary table S1, Supplementary Material online), reflecting the high number of nucleotide alleles for this locus. For all five genes, identical amino acid alleles were shared across multiple species (fig. 4 and supplementary figs. S4–S7, Supplementary Material online). With the exceptions of *AvBD3b* and *AvBD13*, the most common mallard allele was also the most widely shared across waterfowl species (fig. 4 and supplementary figs. S5

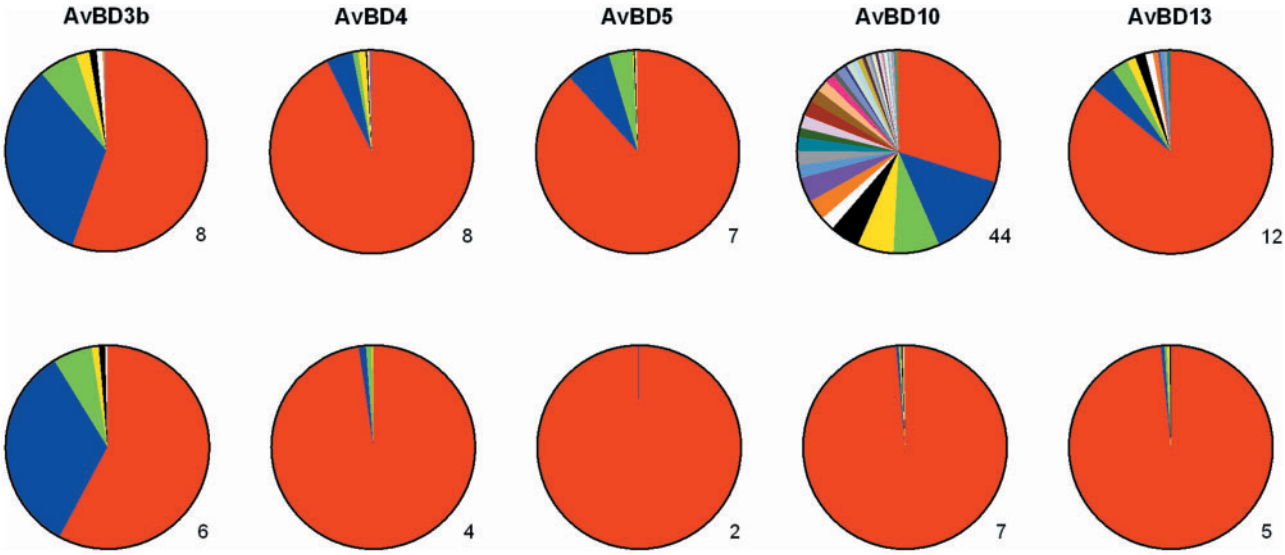

FIG. 1. Number and relative frequency of nucleotide (top panel) and amino acid (bottom panel) alleles in the locally sampled (Swedish) population. For each gene, the most common allele is shown in red, followed by blue, green, yellow etcetera. Numbers to the bottom right of each pie show the total number of alleles for that locus.

and S6, Supplementary Material online). For *AvBD13* (supplementary fig. S7, Supplementary Material online), the allele shared most widely among waterfowl was not present in mallards, whereas the most common global mallard allele was only detected in one other closely related species (*Anas crecca*). Less allele sharing was observed overall for *AvBD3b* than the other four genes, with the most widely shared allele being found in seven members of the Aythyini (supplementary fig. S4, Supplementary Material online). The most common mallard *AvBD3b* allele was found in two other dabbling duck species. At the nucleotide level, identical alleles were also observed across species for all genes, although to a lesser extent (supplementary fig. S8, Supplementary Material online).

Although waterfowl are the most readily hybridizing family of birds (Grant and Grant 1992; Ottenburghs et al. 2015), hybridization is unlikely to fully account for the observed widespread sharing of AvBD alleles. First, the occurrence of hybrid duck species decreases with phylogenetic distance (Tubaro and Lijtmaer 2002; Kraus et al. 2012), yet shared alleles were found among species, genera, and families separated by millions of years of evolution (Gonzalez et al. 2009) (fig. 4, supplementary figs. S4–S7, Supplementary Material online), and between species living in allopatry (supplementary table S3, Supplementary Material online). For example, the most common *AvBD4* and *AvBD10* alleles were shared among various families of ducks, as well as members of the *Dendrocygninae* (fig. 4 and supplementary fig. S5, Supplementary Material online) which is estimated to have diverged from the rest of the Anatidae approximately 49 million years before present (Gonzalez et al. 2009). Second, ducks are clearly not genetically identical at other genes, given that mitochondrial (Gonzalez et al. 2009) and nuclear (Lavretsky et al. 2014) genes can delineate Anatidae species boundaries. An alternative explanation for the widespread sharing of AvBD alleles may be convergence, driven by pathogen mediated selection, given the shared pathogen fauna of many waterfowl. Convergence has been inferred for HDPs in anuran amphibians (König and Bininda-Emonds 2011). Although we cannot exclude convergent evolution in our study, the widespread allele sharing detected at both the amino acid (fig. 4 and supplementary figs. S4–S7, Supplementary Material online) and nucleotide (supplementary fig. S8, Supplementary Material online) levels for all loci, and the observed phylogenetic signal, whereby alleles were more commonly shared by more closely related species, suggests that retention of ancient, highly efficacious alleles is a more parsimonious explanation for our data.

Signatures of Selection on β-Defensin Genes

We found evidence for purifying selection acting on individual amino acid residues for all genes, both across species and within mallards (figs. 5 and 6 and supplementary tables S6–S20, Supplementary Material online). Tajima's *D* was significantly negative for *AvBD3b* and *AvBD4* (supplementary table S1, Supplementary Material online). Furthermore, for every gene, between 32% (*AvBD3b*) and 84% (*AvBD4*) of amino acid residues were fixed across every species examined. Indeed, many residues were also fixed at the codon level (i.e., the same triplet codon was utilized by all species), with the extent of codon fixation varying between genes: 19% in *AvBD3b*, 62% in *AvBD4*, 49% in *AvBD5*, 37% in *AvBD10*, and 61% in *AvBD13* (fig. 5, residues shown in black). Given that the species represented in our inter-specific dataset represent upwards of 50 million years of evolution, this high level of codon fixation (i.e., lack of synonymous substitutions) was unexpected and suggests that codon usage bias may be an important factor in AvBD evolution. Unsurprisingly, the six conserved cysteines in each mature peptide were generally either fixed at the codon level or under strong purifying selection. In only one case did we find a non-synonymous

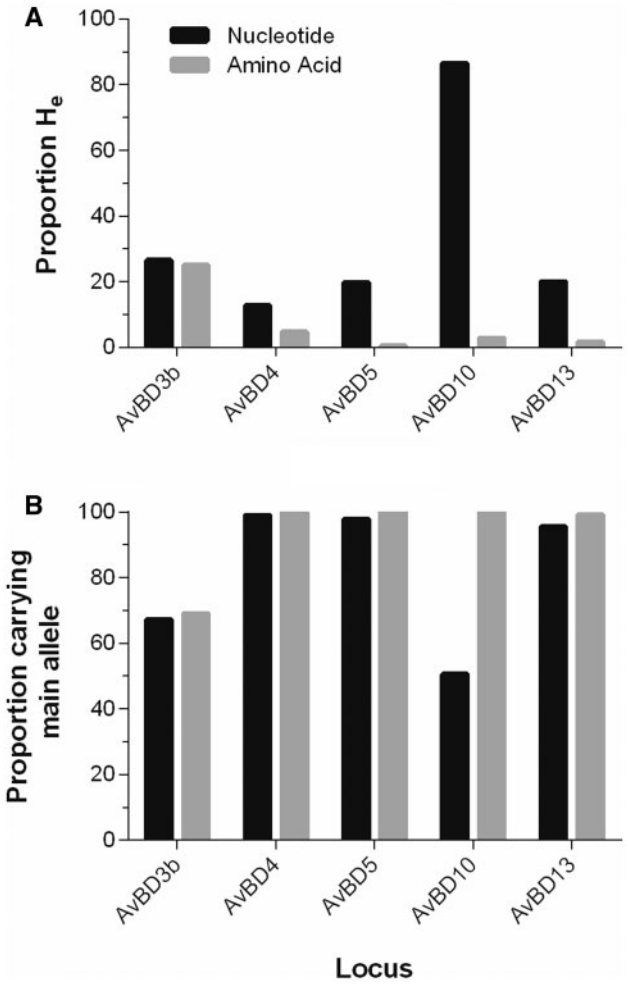

FIG. 2. Distribution of alleles in mallards sampled locally in Sweden: (*A*) proportion of heterozygous (H_e) individuals per locus; and (*B*) proportion of individuals carrying at least one copy of the most frequent (main) allele. In both cases, nucleotide alleles are shown as black bars and amino acid alleles as gray bars.

substitution at any of the cysteine residues, whereby a tyrosine replaced the second cysteine residue in *Branta ruficollis AvBD5*; both individuals were heterozygous for this substitution. Additionally, the final residue in every gene was always fixed at the codon level. Fixed residues, and those under purifying selection, were interspersed across the entire length of the mature peptide, for every gene (fig. 5). Additionally, we found evidence for diversifying selection acting on several residues in *AvBD3b* (figs. 5, 6 and supplementary tables S6, S11, and S16, Supplementary Material online). These positively selected residues occurred throughout the *AvBD3b* gene, with a notable cluster at the 3′-end, whereby 50% of the final eight residues were under some degree of positive selection. It has previously been suggested that residues under positive selection in AvBDs generally occur within two residues of the conserved cysteines (Cheng et al. 2015), but this was not fully supported for *AvBD3b* (fig. 5). Two *AvBD3b* amino acid alleles, found locally in Swedish mallards, were maintained in mallard populations globally (fig. 3*B*). This gene also displayed the highest amino acid allelic diversity and lowest level of allele sharing among waterfowl (supplementary fig. S4, Supplementary Material online).

These contrasting signatures of selection likely reflect the evolutionary origin of the β-defensin genes: *AvBD4*, *AvBD5*, *AvBD10*, and *AvBD13* have one-to-one orthologous genes across the avian phylogeny and are inferred to have evolved at least 100 million years ago (Cheng et al. 2015). In contrast, *AvBD3b* is a member of the AvBD3 cluster which shows lineage-specific gene duplications (Hellgren and Ekblom 2010; Lan et al. 2014; Cheng et al. 2015), implicating a more recent evolutionary origin for *AvBD3b*. Overall, these results demonstrate that both purifying and balancing selection can act on β-defensin genes in waterfowl, which are inferred to be tightly clustered on chromosome three in mallard (Huang et al. 2013; Cheng et al. 2015). Although purifying selection is the more pervasive evolutionary force in maintaining functionally important β-defensin alleles across populations and species, the evolutionary age of genes appears to play an important role (see also Cheng et al. 2015). Gene duplication tends to trigger a period of relaxed selection, whereby the original function of the gene is maintained by one of the duplicates, whereas the other(s) are freed to explore new functional space; if a new function is obtained then the duplicate(s) are subsequently maintained by purifying selection (Lynch and Conery 2000; Hurles 2004). Purifying selection on the β-defensin mature peptide has been inferred to drive low allelic diversity in some species (e.g., Simard et al. 2007; Tennessen and Blouin 2007; Lazzaro 2008); whereas positive selection for enhanced diversity has been inferred in others (e.g., Semple et al. 2003; Radhakrishnan et al. 2005; Hollox and Armour 2008; Viljakainen and Pamilo 2008), as well as for α-defensins in mammals (Lynn et al. 2004; Patil et al. 2004). It has been suggested that in general, moderate positive selection is the main evolutionary driver of host defense peptides (Tennessen 2005) and that HDPs are some of the most rapidly evolving genes in the genome of mammals (Peschel and Sahl 2006; Semple et al. 2006). However, a recent comparative analysis of a single individual from 53 avian species showed that purifying selection is pervasive in the AvBD gene family (Cheng et al. 2015), which agrees with our in-depth analysis of allelic variation of five loci in waterfowl.

It is notable that *AvBD10* has vastly more synonymous substitutions than the other four genes, but low functional diversity (60 nucleotide vs. 7 amino acid alleles in globally sampled mallards, supplementary fig. S3 and table S1, Supplementary Material online). What drives this difference remains an open question. One possibility is a tolerance for higher mutation rates at this locus. However, this fails to explain why increased variation is only observed at the nucleotide level. Thus, a higher mutation rate in *AvBD10* would need to be coupled with strong selection to purge most newly emerging amino acid alleles. Moreover, given that the main *AvBD10* amino acid allele is common globally in mallards, as well as in 27 other species of waterfowl, the selective pressure would need to be present through the entire mallard range, and across species divides. Additionally, this selective pressure would need to be strongly underdominant and linked to

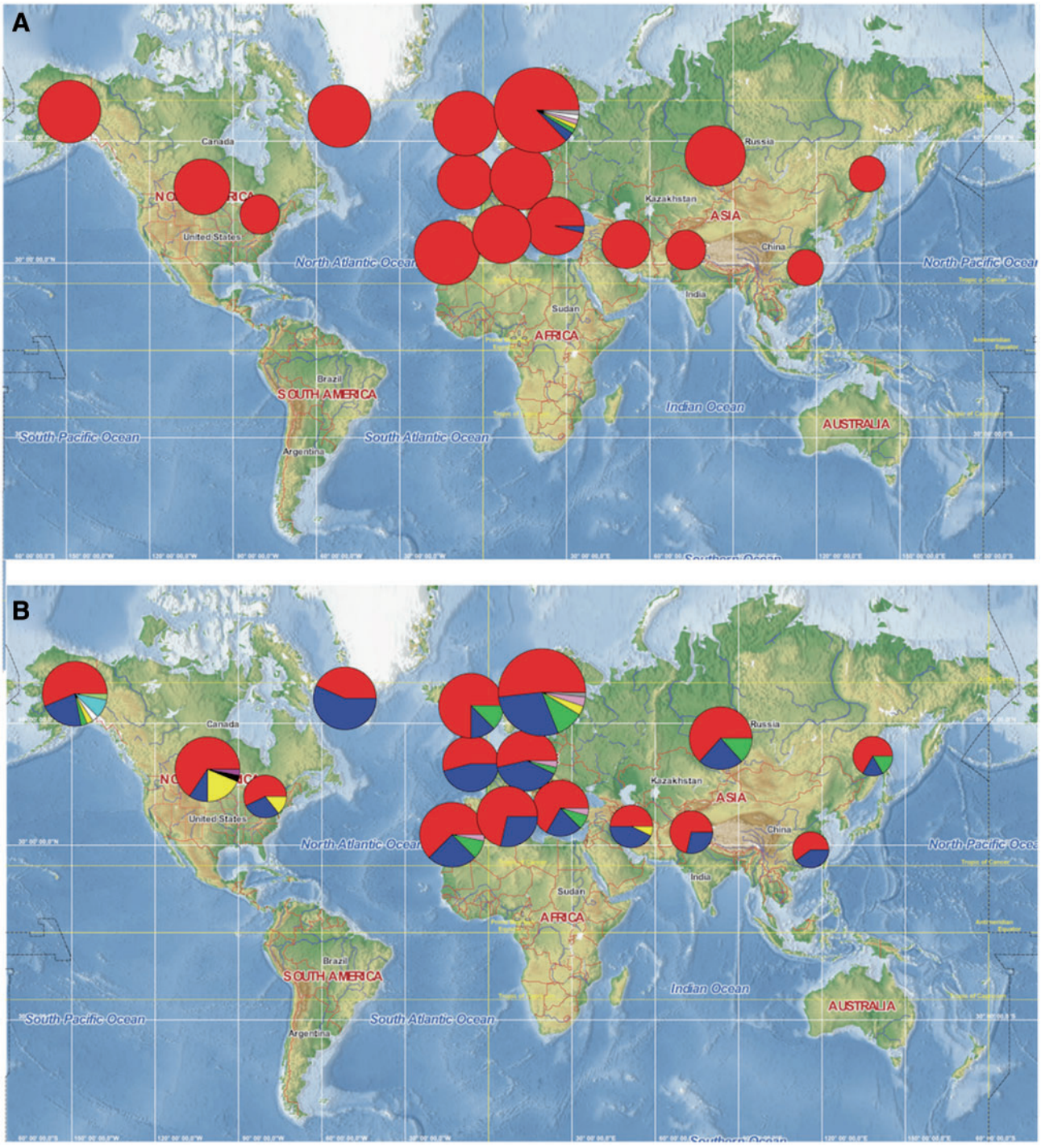

Fig. 3. Amino acid allele frequencies in mallard populations across their global distribution. The most common allele in Sweden and globally is shown in red (allele A), the next most common is in blue (allele B) etcetera (see supplementary fig S2, Supplementary Material online, for further details). Circle sizes represent number of alleles characterised per population, with the exception of Sweden, which has been scaled down by a factor of ten for ease of interpretation. (*A*) *AvBD10*, showing global dominance of allele A; (*B*) *AvBD3b*, showing global maintenance of two alleles. Both common alleles (i.e. red and blue pie segments) are present in every population studied; however, frequencies vary, for example, in the Faroe Islands allele B is rare, whereas in Greenland it is the most common allele.

survival, whereby individuals not homozygous for this allele are largely purged: across all mallards sampled at this locus ($n = 420$), we found only 10 (2.4%) heterozygous individuals (all with one copy of the common amino acid allele) and only a single individual with no copies of the most common allele.

Another possibility is that *AvBD10* is the most ancient of the waterfowl AvBD genes examined here, and has thus had a longer time to accumulate mutations, although this would still not explain the fact that there is low diversity at the amino acid level.

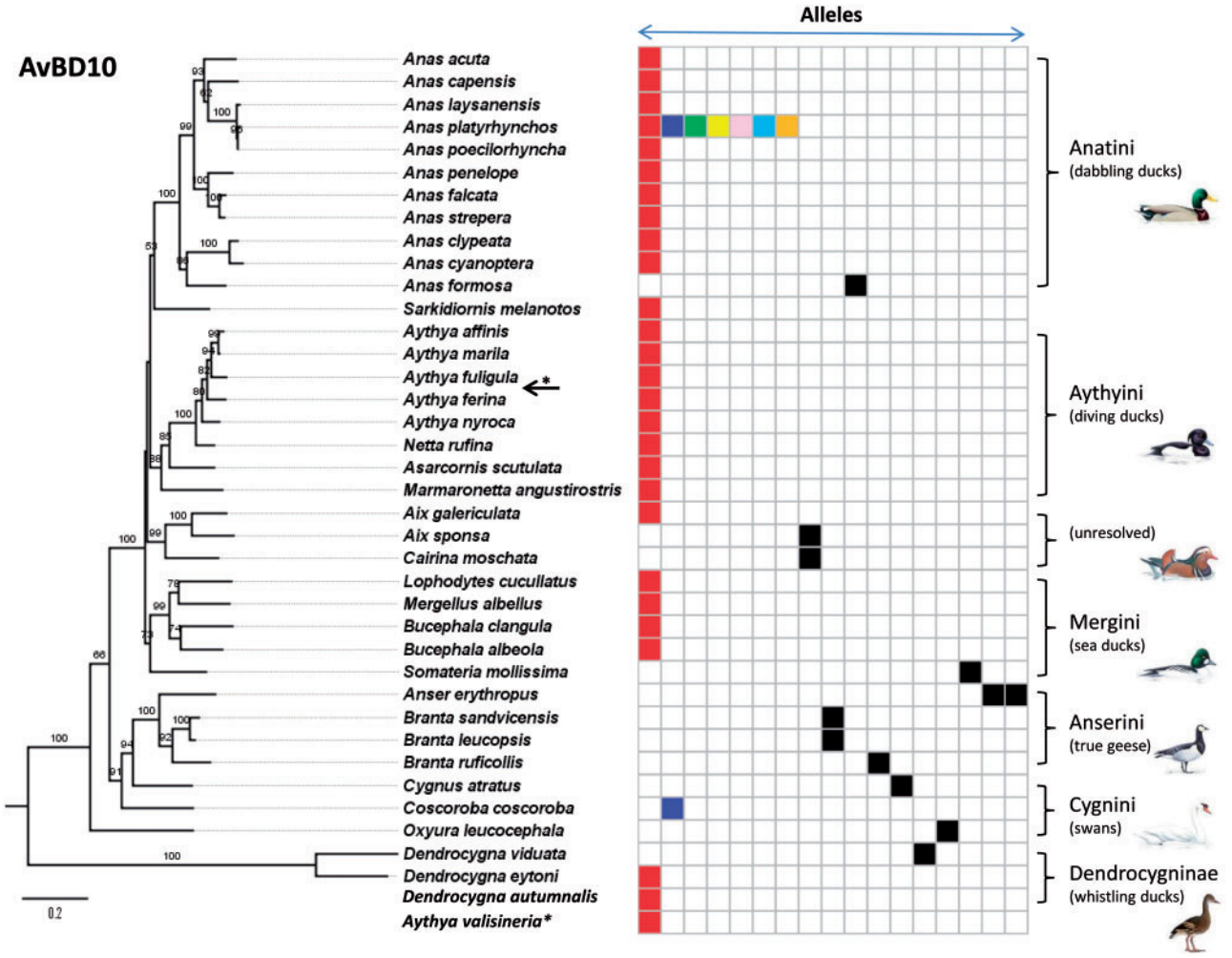

Fig. 4. Sharing of amino acid alleles amongst waterfowl for *AvBD10*. Left panel: PhyML phylogeny based on 1879 base pairs of *cytb* and *nd2* mtDNA genes concatenated, based on Gonzalez et al. (2009), but restricted to those species for which sequences were amplified for *AvBD10*. Bootstrap values are provided for branches with over 50% support. Two species without mtDNA sequence data are included at the bottom: *Dendrocygna autumnalis* is in approximately the correct phylogenetic position; *Aythya valisineria* belongs to the Aythyini tribe, its inferred position in the phylogeny is indicated with an arrow. Right panel: division of tribes and subfamilies, based on Gonzalez et al. (2009). Bird illustrations by Mike Langman (rspb-images.com, last accessed 15 August 2016). Middle panel: distribution of alleles across species, whereby coloured shading denotes mallard alleles (with red for the most common mallard allele, blue for the next most common etcetera) and black shading denotes alleles found only in non-mallard waterfowl. Species with shading in the same vertical column share the allele denoted by that column. The number of shaded boxes in a row horizontally from each branch tip denotes the number of alleles found in that species. Results for the other four loci, and for nucleotide alleles, are presented in supplementary figs. S4–S8. Supplementary Material online.

Host–Pathogen Dynamics

The co-evolutionary arms race between hosts and their pathogens can drive the maintenance of immunogenetic allelic diversity via balancing selection (Ferrer-Admetlla et al. 2008) driven by negative-frequency-dependant selection (Slade and Mccallum 1992) or heterozygote advantage (Slade and Mccallum 1992; Sommer 2005). This is clearly not the case for AvBD genes in waterfowl: we show low levels of immunogenetic diversity linked to purifying selection for four loci (*AvBD4*, *AvBD5*, *AvBD10*, and *AvBD13*), with somewhat enhanced diversity linked to balancing selection for the other locus (*AvBD3b*). A likely explanation for differences in allelic diversity between different types of immune gene is that the selective forces acting on immune genes vary with respect to function. This has recently been confirmed in a wild avian system, whereby recognition molecules (i.e., the MHC) showed higher population differentiation and diversity than signalling molecules (i.e., cytokines) in the greater prairie chicken (Bateson et al. 2015). Thus, one can postulate that afferent molecules, involved in recognition and signalling, need to recognize a wide array of pathogens, and are, therefore, under balancing selection to maintain diversity to counter a rapidly evolving pathogen fauna. In contrast, efferent molecules, involved in processing and eliminating pathogens, are more constrained and thus under purifying selection to maintain activity. If so, one might expect strong selection on pathogens to evolve resistance to effector molecules, such as β-defensins, given their comparatively low rate of

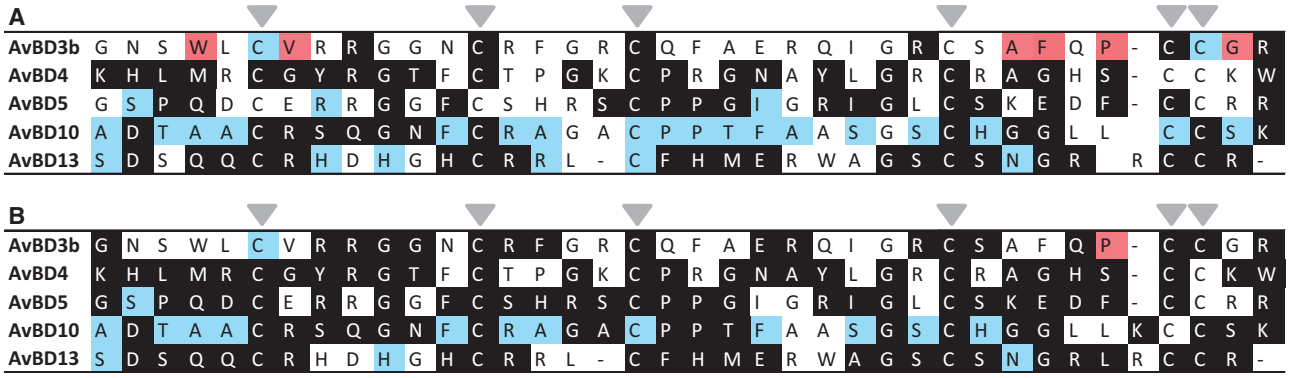

Fig. 5. Strength of support for selection acting on individual amino acid residues, (*A*) across the whole inter-specific dataset and (*B*) within mallards sampled globally. Positive selection on residues was tested via three methods, fixed effects likelihood (FEL), mixed effects model of evolution (MEME) and fast unconstrained Bayesian approximation (FUBAR). Where two or more tests showed significant evidence for positive selection acting on the residue, it is shaded in red. Negative selection on residues was tested via FEL and FUBAR. Where both tests showed significant evidence for negative selection acting on the residue, it is shaded in blue. The significance levels used in each test are presented in the Materials and Methods section. Residues shaded black are fixed at the codon level, whereby every individual included in the dataset shares the exact same triplet codon. Full results of selection tests for each locus are included as supplementary information (panel A, supplementary tables S6–S10, Supplementary Material online; panel B, supplementary tables S11–S15, Supplementary Material online). The allele shown for each locus is the most common in the dataset. The six conserved cysteine (C) residues that typify β-defensin genes are marked with gray triangles.

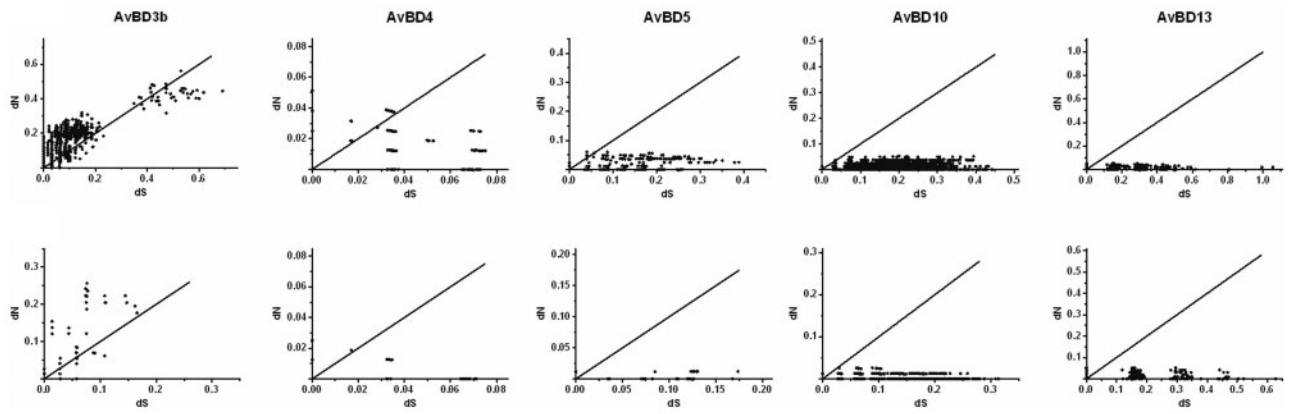

Fig. 6. Pairwise comparisons of the ratio (ω) of d_N (non-synonymous substitution rate) and d_S (synonymous substitution rate) between all unique alleles found in Anatidae (top panel) and mallards sampled globally (bottom panel) for *AvBD3b*; *AvBD4*; *AvBD5*; *AvBD10*; and *AvBD13*. The diagonal line represents $\omega = 1$ (neutral selection), points above the line represent $\omega > 1$ (positive selection) and those below the line $\omega < 1$ (negative selection).

diversification. That AvBDs are ancient molecules (Sugiarto and Yu 2004) that continue to display high antimicrobial activity (Soman et al. 2009; Hellgren et al. 2010; Ma et al. 2011) despite the low genetic diversity we show here, suggests that pathogens have been unable to evolve effective resistance to these peptides. One possible explanation is that β-defensins exploit a fundamental, common component of most pathogens, a negatively charged cellular membrane with a hydrophobic core (Zasloff 2002). Thus, the broad functional space occupied by β-defensins would require a drastic alteration in the pathogen to allow escape. Such large transitions are likely to be extremely rare events. When such escape is achieved by pathogens, this may subsequently promote gene duplication of host immune genes to allow new avenues of resistance to be explored. This could provide one explanation for why some AvBD genes are duplicated in birds and others are not. Interestingly, we found that for the more recently duplicated gene, *AvBD3b*, amino acid substitutions within each allele were significantly more likely to be classified as radical replacements than conservative replacements (Wilcox $W = -258.0$, $P = 0.02$, $n = 40$, supplementary table S21, Supplementary Material online). In contrast, there was no significant difference in the type of replacement within alleles for the other four genes ($P > 0.05$ in all cases). Radical replacements are more likely to have profound effects on the structure and function of the folded mature peptide than conservative substitutions (Smith 2003). Thus, the more recently duplicated *AvBD3b* may be subject to relaxed selection, allowing the gene to acquire new functions, for example evolution of alleles for resistance to emerging pathogens. Nevertheless, we found that hydrophobicity and net charge, whereas differing somewhat between genes, were broadly

similar for the different amino acid alleles within genes (supplementary table S21 and fig. S9, Supplementary Material online), suggesting functional constraints whereby charge and hydrophobicity are actively maintained due to their fundamental roles in β-defensin activity.

Conclusion

Just one to two β-defensin alleles per locus are maintained at high frequency in global populations of mallards, and have been conserved across species divides, indicating that these alleles likely represent the optimal available evolutionary solution for those species possessing them. Small changes in AvBD amino acid sequence can have large effects on peptide activity (Hellgren et al. 2010), which likely explains why we find such strong purifying selection on most AvBD genes in waterfowl. β-Defensins are a current target of pharmaceutical research as alternatives to traditional antibiotics (Brogden and Brogden 2011). We suggest that such research should focus on specific alleles that are widespread, within and across species, as these likely represent peptides favored over evolutionary timespans for their efficacy at dealing with pathogens of biological relevance to their hosts.

That we find differing patterns of selection on tightly clustered β-defensin genes with similar functions in the immune system provides a valuable insight into the rate and strength of selection acting on innate immune genes. Furthermore, the fact that individual alleles can be maintained in diverse species separated by millions of years of evolution, suggests that the evolutionary benefit to hosts in possessing specific AvBD alleles, or allelic combinations, must be much higher than previously recognized. Overall, our results show that although patterns of selection can differ between tightly clustered β-defensin genes, when considering each gene individually patterns of selection are remarkably similar within and between waterfowl species regardless of their geographic and evolutionary separation.

Materials and Methods

Sample Collection

For Swedish mallards, wild birds ($n = 265$) were captured in 2011 and 2012, in a duck trap at Ottenby Bird Observatory, Sweden (56°12′ N 16°24′ E). Further details of the duck trap and duck handling protocol can be found in Wallensten et al. (2007). Each bird was ringed, sexed, and aged and a small blood sample was collected. Less than 50 μl of blood was collected from the tarsus (2011) or brachial (2012) vein and stored in 500 μl SET buffer (0.15 M NaCl, 0.05 M Tris, 0.001 M EDTA) at −20 °C until required for extraction. Among sampled individuals, 11 (4%) were captured in both years, genotyped and scored independently to ensure repeatability of allele assignments. For global mallards, samples were collected as described in Kraus et al. (2011; 2013). Location and number of individuals used per population are provided in supplementary table S2 (Supplementary Material online). Anatidae samples were obtained from two sources (supplementary table S3, Supplementary Material online).

Laboratory Methods

DNA was extracted from wild-caught mallards and both captive and wild-caught Anatidae species via ammonium acetate precipitation, or as described in Kraus et al. (2011). The mallard genome (Huang et al. 2013) was searched to locate β-defensin genes, based on the conserved six cysteine motif C X_{4-8} C X_{3-5} C X_{9-13} C X_{4-7} CC, where X can be any amino acid. Primers (supplementary table S4, Supplementary Material online) were designed to span exon three, which codes for the entire mature peptide. PCR reactions were conducted in 25 μl volumes containing 1 × PCR buffer (containing Tris-Cl, KCl, $(NH_4)_2SO_4$), 2.5 mM $MgCl_2$, 0.2 μM of each primer (forward and reverse), 0.4 mM of each dNTP (Qiagen), and 1U *Taq* DNA polymerase (Qiagen). The thermocycling procedure consisted of an initial denaturing step of 94 °C for 5 min, followed by 40 cycles of 94 °C for 30 s, T_m °C for 30 s and 72 °C for 30 s, and a final extension step of 72 °C for 5 min. Optimal annealing temperatures, T_m, are listed in supplementary table S4 (Supplementary Material online). All PCR reactions were run on Applied Biosystems 9700 or 2720 Gene Amp Thermal cyclers. In most cases, amplicons were sequenced in both the forward and reverse directions. PCR products were sequenced (Eurofins, Germany), then analyzed in Geneious v. 8.0.4 (Kearse et al. 2012) and aligned using the MUSCLE algorithm (Edgar 2004). Alleles were defined as sequences differing by one or more single nucleotide polymorphisms (SNPs). Alleles found in a single individual (if not cloned, see below) were confirmed by a second independent PCR and sequencing analysis. Highly heterozygous individuals were cloned via the pGEM-T Vector System II system (Promega) and re-sequenced. For *AvBD10*, 340 individuals (82%) were heterozygous for more than one SNP. Of these, 239 were cloned. For the remaining 101 individuals, all of which possessed SNP combinations in common with cloned individuals, alleles were determined using PHASE v. 2.1 (Stephens et al. 2001) as implemented in DnaSP v. 5.10.01 (Librado and Rozas 2009). Allele phasing was conducted separately for the cross-species dataset so as not to bias results towards possession of mallard alleles. Sequences were trimmed to encompass just the mature β-defensin peptide, identification of the boundary between propiece and mature peptide was performed with reference to chicken and zebra finch sequences (fig. 4 in Hellgren and Ekblom 2010).

Evolutionary Analyses

Nucleotide sequences were translated into amino acid sequences and aligned in Geneious. Given that defensins are subject to gene duplication followed by rapid molecular evolution (van Dijk et al. 2008; Hellgren and Ekblom 2010) and that our β-defensin sequences were short (108–114 bp), traditional phylogenetic analyses that assume bifurcating processes may not be appropriate to represent evolutionary relationships (Moulton and Huber 2009). As such, phylogenetic relationships between nucleotide alleles were analyzed using median-joining haplotype networks (Bandelt et al. 1999), implemented in PopArt v. 1.7.2 (Leigh and Bryant

2015), which was additionally used to create allelic distribution maps. In these analyses, each individual was represented by two sequences (one per allele), whereby homozygotes had two identical sequences and heterozygotes had one sequence per allele. Sizes of circles in the resultant phylogenetic networks (supplementary fig. S3, Supplementary Material online) and distribution maps (fig. 3 and supplementary fig. S2, Supplementary Material online), therefore, represent the frequency of each allele in the population. For the phylogenies presented in fig. 4 and supplementary figs. S4–S7 (Supplementary Material online), mtDNA sequences derived from Gonzalez et al. (2009) were obtained for the relevant species (those for which we had amplified AvBD sequences for the relevant locus). Interested readers should refer to Gonzalez et al. (2009) for a more complete reconstruction of waterfowl phylogeny. Model testing was performed in MEGA v. 6.0.6 (Tamura et al. 2013) to determine the correct evolutionary model. Thereafter, PhyML trees (Guindon and Gascuel 2003) were constructed in Seaview v. 4.6 (Gouy et al. 2010), with 1000 bootsrap iterations to determine branch support, and edited in FigTree v. 1.4 (http://tree.bio.ed.ac.uk, last accessed 12 August 2016).

The ratio (ω) of non-synonymous (d_N) to synonymous (d_S) substitutions was determined in PAMLX (Xu and Yang 2013) using the YN00 method, whereby for each gene and dataset (local mallard, global mallard, Anatidae), ω was calculated for pairwise comparisons of unique alleles and plotted in GraphPad Prism v. 6.0 (www.graphpad.com, last accessed 12 August 2016). Results for local mallards are not presented due to paucity of data for some loci, but were qualitatively similar to those obtained for global mallards. Nucleotide and haplotype diversity and Tajima's Test of Neutrality (Tajima 1989) were performed in DnaSP v. 5.10.01 (Librado and Rozas 2009). Analysis of selection on individual amino acid residues was performed using HyPhy (Pond et al. 2005), implemented via the Datamonkey (Pond and Frost 2005a) web interface (http://www.datamonkey.org, last accessed 12 August 2016). We used three complementary methods to detect positive and negative selection. First, fixed effects likelihood (FEL), which has been shown to outperform counting (e.g., SLAC) and random effects (e.g., REL) based methods (Pond and Frost 2005b), including with small datasets (i.e., fewer than 50 sequences), which applies to many of our intra-specific analyses. Second, a mixed effects model of evolution (MEME), to detect footprints of positive selection (both pervasive and episodic) at individual sites; MEME has been suggested to provide more sensitive detection of positive selection than FEL but does not test for purifying selection (Murrell et al. 2012). Third, a fast unconstrained Bayesian approximation (FUBAR) which can detect evidence for both pervasive diversifying and purifying selection at individual sites (Murrell et al. 2013) and has also been suggested to outperform more traditional methods such as FEL and REL. Within HyPhy, model testing was initially performed in order to correctly specify the nucleotide substitution model, and this model was used for all subsequent tests. The following significance levels were used for selection analyses: $P \leq 0.1$ for FEL and MEME, posterior probability ≥ 90 for FUBAR. All HyPhy analyses were performed with user-specified trees, being PhyML trees inferred in SeaView v. 4.6 after model testing in MEGA v. 6.0 to determine the appropriate nucleotide substitution model. Analyses of selection were conducted at three levels: (1) entire inter-specific dataset, comprising Swedish mallards, global mallards, and other species of Anatidae (supplementary tables S6–S10, Supplementary Material online); (2) intra-specific dataset, comprising locally and globally sampled mallards (supplementary tables S11–S15, Supplementary Material online); and (3) intra-population dataset, comprising locally sampled (Swedish) mallards only (supplementary tables S16–S20, Supplementary Material online). To assess whether tree topology influenced the inference of selection on codons, HyPhy analyses were repeated on a subset of four representative datasets using randomly subsampled Bayesian trees. Briefly, Bayesian phylogenetic analyses were run in MrBayes v. 3.2.6 (Ronquist et al. 2012) with 1.5 million generations, after discarding 500,000 generations as burn-in, on four Markov chains. Subsequently, we used a modified version of an R script written by Heath Blackman (available at http://coleoguy.github.io/rseminar/wksheet5.pdf, last accessed 12 August 2016) to randomly draw five trees from the retained set of Bayesian trees, using R v. 3.0.2 (R Core Team 2013). The selection analyses were then re-run using each of these five subsampled Bayesian trees as the user-defined input tree in DataMonkey. Further details and a summary of results are presented in supplementary table S22 (Supplementary Material online), full results and methods can be obtained from the authors upon request.

Amino acid substitutions were categorized as polar-neutral (S, T, Y, C, N, and Q), polar-acidic (D and E), polar-basic (K, R, and H), and non-polar (G, A, V, L, I, F, P, M, and W) (Hanada et al. 2007). Replacements (with reference to the most common amino acid allele) within a category were considered conservative changes, whereas replacements between categories were considered radical (Hanada et al. 2007). To determine whether radical or conservative changes were more likely within alleles at each locus, non-parametric Wilcoxon matched-pairs tests were conducted in GraphPad Prism. Allele discovery (rarefaction) curves (supplementary fig. S1, Supplementary Material online) were estimated via a coverage-based rarefaction and extrapolation analysis in iNEXT (https://chao.shinyapps.io/iNEXTOnline, last accessed 12 August 2016) (Chao and Jost 2012; Hsieh et al. 2013) with an endpoint of twice the number of haplotypes as actually sampled (i.e., the number of individuals times four), 50 knots and 500 bootstrap replicates to estimate 95% confidence intervals.

Note that the gene named here as *AvBD3b* has also been called *AvBD16* (Huang et al. 2013) and *AvBD3.6* (Cheng et al. 2015). When comparing our *AvBD3b* sequences to those presented in Cheng et al. (2015) we found that although most (95%) of our sequences formed a clade with *AvBD3.6*, the remaining 5% formed into two distinct clades, one of which grouped with *AvBD3.5* (supplementary fig. S10, Supplementary Material online). We, therefore, cannot discount the possibility that our primers designed for *AvBD3b* occasionally amplified a different member of the *AvBD3*

duplicated family, possibly due to the presence of rare null alleles. As such, we repeated selection analyses with these sequences removed. Results are largely similar, whereby we show that after removal of these sequences the inference of balancing selection on *AvBD3b* was retained, and indeed somewhat strengthened, compared with the full dataset presented here (supplementary table S23 and fig. S11, Supplementary Material online). Because sequences of less than 200bp cannot be deposited in GenBank, the sequences for all nucleotide alleles are provided in FASTA format in supplementary file 4 (Supplementary Material online).

Supplementary Material

Supplementary figs. S1–S11, tables S1–S23, and supplementary material are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We particularly thank Michelle Wille for invaluable assistance with sample collection, figures, phylogenetic analyses and numerous helpful discussions. We additionally thank Staffan Bensch for comments on a previous version of this manuscript and Sara Larsson for laboratory assistance. This is publication number 296 from Ottenby Bird Observatory.

References

Alcaide M, Edwards SV. 2011. Molecular evolution of the Toll-like receptor multigene family in birds. *Mol Biol Evol*. 28:1703–1715.

Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, Gutenkunst RN, White TJ, Green ED, Bustamante CD, et al. 2009. Targets of balancing selection in the human genome. *Mol Biol Evol*. 26:2755–2764.

Arnett E, Seveau S. 2011. The multifaceted activities of mammalian defensins. *Curr Pharm Des*. 17:4254–4269.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 16:37–48.

Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O, Gicquel B, et al. 2009. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet*. 5:e1000562.

Barreiro LB, Quintana-Murci L. 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet*. 11:17–30.

Bateson ZW, Whittingham LA, Johnson JA, Dunn PO. 2015. Contrasting patterns of selection and drift between two categories of immune genes in prairie-chickens. *Mol Ecol*. 24:6095–6106.

Beutler B. 2004. Innate immunity: an overview. *Mol Immunol*. 40:845–859.

Brogden KA. 2005. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol*. 3:238–250.

Brogden NK, Brogden KA. 2011. Will new generations of modified antimicrobial peptides improve their potential as pharmaceuticals? *Int J Antimicrob Ag* 38:217–225.

Campopiano DJ, Clarke DJ, Polfer NC, Barran PE, Langley RJ, Govan JRW, Maxwell A, Dorin JR. 2004. Structure-activity relationships in defensin dimers: a novel link between β-defensin tertiary structure and antimicrobial activity. *J Biol Chem*. 279:48671–48679.

Causey D, Edwards SV. 2008. Ecology of avian influenza virus in birds. *J Infect Dis*. 197:S29–S33.

Chao A, Jost L. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93:2533–2547.

Cheng Y, Prickett MD, Gutowska W, Kuo R, Belov K, Burt DW. 2015. Evolution of the avian β-defensin and cathelicidin genes. *BMC Evol Biol*. 15:188.

Choi KY, Chow LNY, Mookherjee N. 2012. Cationic host defence peptides: multifaceted role in immune modulation and inflammation. *J Innate Immun*. 4:361–370.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res*. 32:1792–1797.

Ferrer-Admetlla A, Bosch E, Sikora M, Marquès-Bonet T, Ramírez-Soriano A, Muntasell A, Navarro A, Lazarus R, Calafell F, Bertranpetit J, et al. 2008. Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol*. 181:1315–1322.

Funderburg N, Lederman MM, Feng Z, Drage MG, Jacllowsky J, Harding CV, Weinberg A, Sieg SF. 2007. Human β-*defensin-3* activates professional antigen-presenting cells via Toll-like receptors 1 and 2. *Proc Natl Acad Sci U S A*. 104:18631–18635.

Ganz T. 2003. Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol*. 3:710–720.

Gonzalez J, Düttmann H, Wink M. 2009. Phylogenetic relationships based on two mitochondrial genes and hybridization patterns in Anatidae. *J Zool*. 279:310–318.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 27:221–224.

Grant PR, Grant BR. 1992. Hybridization of bird species. *Science* 256:193–197.

Grueber CE, Wallis GP, Jamieson IG. 2014. Episodic positive selection in the evolution of avian Toll-like receptor innate immunity genes. *PLoS One* 9:e89632.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.

Gura T. 2001. Innate immunity: ancient system gets new respect. *Science* 291:2068–2071.

Haldane JBS. 1949. Disease and evolution. *Ric Sci Suppl A*. 19:68–76.

Hanada K, Shiu SH, Li WH. 2007. The nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. *Mol Biol Evol*. 24:2235–2241.

Hancock REW, Haney EF, Gill EE. 2016. The immunology of host defence peptides: beyond antimicrobial activity. *Nat Rev Immunol*. 16:321–334.

Hancock REW, Sahl HG. 2006. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat Biotechnol*. 24:1551–1557.

Hazrati E, Galen B, Lu WY, Wang W, Yan OY, Keller MJ, Lehrer RI, Herold BC. 2006. Human α- and β-defensins block multiple steps in herpes simplex virus infection. *J Immunol*. 177:8658–8666.

Hellgren O. 2015. Allelic variation at innate immune genes (avian β-defensins), within a natural population of great tits. *J Avian Biol*. 46:113–118.

Hellgren O, Ekblom R. 2010. Evolution of a cluster of innate immune genes (β-defensins) along the ancestral lines of chicken and zebra finch. *Immunome Res*. 6:3.

Hellgren O, Sheldon BC, Buckling A. 2010. *In vitro* tests of natural allelic variation of innate immune genes (avian β-defensins) reveal functional differences in microbial inhibition. *J Evol Biol*. 23:2726–2730.

Hilchie AL, Wuerth K, Hancock REW. 2013. Immune modulation by multifaceted cationic host defense (antimicrobial) peptides. *Nat Chem Biol*. 9:761–768.

Hoffmann JA, Kafatos FC, Janeway CA, Ezekowitz RAB. 1999. Phylogenetic perspectives in innate immunity. *Science* 284:1313–1318.

Hollox EJ, Armour JAL. 2008. Directional and balancing selection in human beta-defensins. *BMC Evol Biol*. 8:113.

Hsieh TC, Ma KH, Chao A. 2013. iNEXT online: interpolation and extrapolation. Available from: https://chao.shinyapps.io/iNEXT/. Last accessed 12 August 2016

Huang YH, Li YR, Burt DW, Chen HL, Zhang Y, Qian WB, Kim H, Gan SQ, Zhao YQ, Li JW, et al. 2013. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat Genet*. 45:776–783.

Hughes AL, Packer B, Welch R, Chanock SJ, Yeager M. 2005. High level of functional polymorphism indicates a unique role of natural selection at human immune system loci. *Immunogenetics* 57:821–827.

Hurles M. 2004. Gene duplication: the genomic trade in spare parts. *Plos Biol*. 2:e206.

Jarczak J, Kościuczuk EM, Lisowski P, Strzalkowska N, Jóźwik A, Horbańczuk J, Krzyżewski J, Zwierzchowski L, Bagnicka E. 2013. Defensins: natural component of human innate immunity. *Hum Immunol*. 74:1069–1079.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.

Klotman ME, Chang TL. 2006. Defensins in innate antiviral immunity. *Nat Rev Immunol*. 6:447–456.

Kraus RHS, Kerstens HHD, van Hooft P, Megens HJ, Elmberg J, Tsvey A, Sartakov D, Soloviev SA, Crooijmans RPMA, Groenen MAM, et al. 2012. Widespread horizontal genomic exchange does not erode species barriers among sympatric ducks. *BMC Evol Biol*. 12:45.

Kraus RHS, Van Hooft P, Megens HJ, Tsvey A, Fokin SY, Ydenberg RC, Prins HHT. 2013. Global lack of flyway structure in a cosmopolitan bird revealed by a genome wide survey of single nucleotide polymorphisms. *Mol Ecol*. 22:41–55.

Kraus RHS, Zeddeman A, van Hooft P, Sartakov D, Soloviev SA, Ydenberg RC, Prins HHT. 2011. Evolution and connectivity in the world-wide migration system of the mallard: inferences from mitochondrial DNA. *BMC Genet*. 12:99.

Kruse H, Kirkemo AM, Handeland K. 2004. Wildlife as source of zoonotic infections. *Emerg Infect Dis*. 10:2067–2072.

König E, Bininda-Emonds ORP. 2011. Evidence for convergent evolution in the antimicrobial peptide system in anuran amphibians. *Peptides* 32:20–25.

Lai Y, Gallo RL. 2009. AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends Immunol*. 30:131–141.

Lan H, Chen H, Chen LC, Wang BB, Sun L, Ma MY, Fang SG, Wan QH. 2014. The first report of a Pelecaniformes defensin cluster: characterization of β-defensin genes in the crested ibis based on BAC libraries. *Sci Rep*. 4:e6923.

Lavretsky P, McCracken KG, Peters JL. 2014. Phylogenetics of a recent radiation in the mallards and allies (Aves: Anas): inferences from a genomic transect and the multispecies coalescent. *Mol Phylogenet Evol*. 70:402–411.

Lazarus R, Vercelli D, Palmer LJ, Klimecki WJ, Silverman EK, Richter B, Riva A, Ramoni M, Martinez FD, Weiss ST, et al. 2002. Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunol Rev*. 190:9–25.

Lazzaro BP. 2008. Natural selection on the *Drosophila* antimicrobial immune system. *Curr Opin Microbiol*. 11:284–289.

Leigh JW, Bryant D. 2015. POPART: full-feature software for haplotype network construction. *Methods Ecol Evol*. 6:1110–1116.

Li YM, Xiang Q, Zhang QH, Huang YD, Su ZJ. 2012. Overview on the recent study of antimicrobial peptides: origins, functions, relative mechanisms and application. *Peptides* 37:207–215.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.

Lynn DJ, Lloyd AT, Fares MA, O'Farrelly C. 2004. Evidence of positively selected sites in mammalian α-defensins. *Mol Biol Evol*. 21:819–827.

Ma D, Lin L, Zhang K, Han Z, Shao Y, Liu X, Liu S. 2011. Three novel *Anas platyrhynchos* avian β-defensins, upregulated by duck hepatitis virus, with antibacterial and antiviral activities. *Mol Immunol*. 49:84–96.

Magor BG, Magor KE. 2001. Evolution of effectors and receptors of innate immunity. *Dev Comp Immunol*. 25:651–682.

Majumder PP. 2010. Pathogen pressure and molecular evolutionary genetics of innate immunity genes in humans. In: Sharma VP, editor. Nature at work: ongoing saga of evolution. New Delhi: Springer India. p. 249–265.

Melo MN, Ferre R, Castanho MARB. 2009. Antimicrobial peptides: linking partition, activity and high membrane-bound concentrations. *Nat Rev Microbiol*. 7:245–250.

Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N. 2012. Molecular evolution of vertebrate Toll-like receptors: evolutionary rate difference between their leucine-rich repeats and their TIR domains. *Gene* 503:235–243.

Moulton V, Huber KT. 2009. Split networks: a tool for exploring complex evolutionary relationships in molecular data. In: Salemi M, Vandamme AM, Lemey P, editors. The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. Cambridge: Cambridge University Press. p. 631–653.

Mukherjee S, Ganguli D, Majumder PP. 2014. Global footprints of purifying selection on Toll-like receptor genes primarily associated with response to bacterial infections in humans. *Genome Biol Evol*. 6:551–558.

Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP. 2009. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proc Natl Acad Sci U S A*. 106:7073–7078.

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLK, Scheffler K. 2013. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Mol Biol Evol*. 30:1196–1205.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 8:e1002764.

Nguyen LT, Haney EF, Vogel HJ. 2011. The expanding scope of antimicrobial peptide structures and their modes of action. *Trends Biotechnol*. 29:464–472.

Ottenburghs J, Ydenberg RC, Van Hooft P, Van Wieren SE, Prins HHT. 2015. The Avian Hybrids Project: gathering the scientific literature on avian hybridization. *IBIS* 157:892–894.

Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, Quail M, Smith F, Walker D, Libberton B, et al. 2010. Antagonistic coevolution accelerates molecular evolution. *Nature*. 464:275–278.

Patil A, Hughes AL, Zhang GL. 2004. Rapid evolution and diversification of mammalian α-defensins as revealed by comparative analysis of rodent and primate genes. *Physiol Genomics*. 20:1–11.

Peschel A, Sahl HG. 2006. The co-evolution of host cationic antimicrobial peptides and microbial resistance. *Nat Rev Microbiol*. 4:529–536.

Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, et al. 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res*. 19:826–837.

Pond SLK, Frost SDW. 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.

Pond SLK, Frost SDW. 2005b. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*. 22:1208–1222.

Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 21:676–679.

Quiñones-Mateu ME, Lederman MM, Feng ZM, Chakraborty B, Weber J, Rangel HR, Marotta ML, Mirza M, Jiang B, Kiser P, et al. 2003. Human epithelial β-defensins 2 and 3 inhibit HIV-1 replication. *Aids*. 17:F39–F48.

R Core Team. 2013. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Radhakrishnan Y, Hamil KG, Yenugu S, Young SL, French FS, Hall SH. 2005. Identification, characterization, and evolution of a primate β-defensin gene cluster. *Genes Immun*. 6:203–210.

Reed KD, Meece JK, Henkel JS, Shukla SK. 2003. Birds, migration and emerging zoonoses: west nile virus, lyme disease, influenza A and enteropathogens. *Clin Med Res*. 1:5–12.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61:539–542.

Segat L, Zupin L, Moura RR, Coelho AVC, Chagas BS, de Freitas AC, Crovella S. 2014. *DEFB1* polymorphisms are involved in susceptibility to human papillomavirus infection in Brazilian gynaecological patients. *Mem I Oswaldo Cruz* 109:918–922.

Semple CA, Gautier P, Taylor K, Dorin JR. 2006. The changing of the guard: molecular diversity and rapid evolution of β-defensins. *Mol Divers*. 10:575–584.

Semple CAM, Rolfe M, Dorin JR. 2003. Duplication and selection in the evolution of primate β-defensin genes. *Genome Biol*. 4:R31.

Simard F, Licht M, Besansky NJ, Lehmann T. 2007. Polymorphism at the *defensin* gene in the *Anopheles gambiae* complex: testing different selection hypotheses. *Infect Genet Evol*. 7:285–292.

Slade RW, Mccallum HI. 1992. Overdominant *vs*. frequency-dependent selection at MHC loci. *Genetics* 132:861–862.

Smith NGC. 2003. Are radical and conservative substitution rates useful statistics in molecular evolution? *J Mol Evol*. 57:467–478.

Soman SS, Arathy DS, Sreekumar E. 2009. Discovery of *Anas platyrhynchos* avian beta-defensin 2 (*Apl_AvBD2*) with antibacterial and chemotactic functions. *Mol Immunol*. 46:2029–2038.

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

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 68:978–989.

Sugiarto H, Yu PL. 2004. Avian antimicrobial peptides: the defense role of β-defensins. *Biochem Biophys Res Commun*. 323:721–727.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30:2725–2729.

Tennessen JA. 2005. Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. *J Evol Biol*. 18:1387–1394.

Tennessen JA, Blouin MS. 2007. Selection for antimicrobial peptide diversity in frogs leads to gene duplication and low allelic variation. *J Mol Evol*. 65:605–615.

Tschirren B, Råberg L, Westerdahl H. 2011. Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (*TLR2*) during the evolutionary history of rodents. *J Evol Biol*. 24:1232–1240.

Tubaro PL, Lijtmaer DA. 2002. Hybridization patterns and the evolution of reproductive isolation in ducks. *Biol J Linn Soc*. 77:193–200.

Wallensten A, Munster VJ, Latorre-Margalef N, Brytting M, Elmberg J, Fouchier RAM, Fransson T, Haemig PD, Karlsson M, Lundkvist A, et al. 2007. Surveillance of influenza A virus in migratory waterfowl in northern Europe. *Emerg Infect Dis*. 13:404–411.

van Dijk A, Veldhuizen EJA, Haagsman HP. 2008. Avian defensins. *Vet Immunol Immunop*. 124:1–18.

Vega LA, Caparon MG. 2012. Cationic antimicrobial peptides disrupt the *Streptococcus pyogenes* ExPortal. *Mol Microbiol*. 85:1119–1132.

Wehkamp J, Fellermann M, Herrlinger KR, Bevins CL, Stange EF. 2005. Mechanisms of disease: defensins in gastrointestinal diseases. *Nat Rev Gastroenterol Hepatol*. 2:406–415.

Werling D, Jann OC, Offord V, Glass EJ, Coffey TJ. 2009. Variation matters: TLR structure and species-specific pathogen recognition. *Trends Immunol*. 30:124–130.

Wiesner J, Vilcinskas A. 2010. Antimicrobial peptides: the ancient arm of the human immune system. *Virulence* 1:440–464.

Viljakainen L, Pamilo P. 2008. Selection on an antimicrobial peptide defensin in ants. *J Mol Evol*. 67:643–652.

Vinkler M, Albrecht T. 2009. The question waiting to be asked: innate immunity receptors in the perspective of zoological research. *Folia Zool*. 58:S15–S28.

Xu B, Yang ZH. 2013. PAMLX: a graphical user interface for PAML. *Mol Biol Evol*. 30:2723–2724.

Yeaman MR, Yount NY. 2003. Mechanisms of antimicrobial peptide action and resistance. *Pharmacol Rev*. 55:27–55.

Yount NY, Yeaman MR. 2013. Peptide antimicrobials: cell wall as a bacterial target. *Ann NY Acad Sci*. 1277:127–138.

Zasloff M. 2002. Antimicrobial peptides of multicellular organisms. *Nature* 415:389–395.

Zhang G, Sunkara LT. 2014. Avian antimicrobial host defense peptides: from biology to therapeutic applications. *Pharmaceuticals*. 7:220–247.

Zhu SY. 2008. Discovery of six families of fungal defensin-like peptides provides insights into origin and evolution of the CS$\alpha\beta$ defensins. *Mol Immunol*. 45:828–838.