nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-018-0509-3

The evolution of immunity in relation to colonization and migration

Emily A. O'Connor*, Charlie K. Cornwallis, Dennis Hasselquist, Jan-Åke Nilsson and Helena Westerdahl

Colonization and migration have a crucial effect on patterns of biodiversity, with disease predicted to play an important role in these processes. However, evidence of the effect of pathogens on broad patterns of colonization and migration is limited. Here, using phylogenetic analyses of 1,311 species of Afro-Palaearctic songbirds, we show that colonization events from regions of high (sub-Saharan Africa) to low (the Palaearctic) pathogen diversity were up to 20 times more frequent than the reverse, and that migration has evolved 3 times more frequently from African- as opposed to Palaearctic-resident species. We also found that resident species that colonized the Palaearctic from Africa, as well as African species that evolved long-distance migration to breed in the Palaearctic, have reduced diversity of key immune genes associated with pathogen recognition (major histocompatibility complex class I). These results suggest that changes in the pathogen community that occur during colonization and migration shape the evolution of the immune system, potentially by adjusting the trade-off between the benefits of extensive pathogen recognition and the costs of immunopathology that result from high major histocompatibility complex class I diversity.

The global distribution of biodiversity is continually being shaped by colonization events, species range expansions and migration between different geographical regions[1–3]. Perhaps the most enigmatic of these processes is seasonal migration, which is evident each year as billions of animals, from blue whales to songbirds, make biannual journeys between their breeding and wintering grounds (Fig. 1). An important factor predicted to determine whether animals move into new areas is the pathogen fauna of their native and novel ranges[4,5]. We use the term 'pathogen' in a broad sense to cover all infectious agents. When disease risk is high in native ranges, selection can favour dispersal to new areas where pathogen pressure is lower ('pathogen escape')[6–8]. Alternatively, entering new environments can result in the opposite pattern if new pathogens, to which species are not adapted, increase mortality; ('pathogen exposure')[9]. However, evidence of the effect of pathogens on patterns of animal movement is mixed and often limited to single species over ecological timescales[5,10]. Consequently, the general effect of pathogens on the evolution of colonization and migration is unknown, despite the potential implications for the spread of zoonotic diseases and climate-driven host range-shifts[2,8].

The main challenge in testing how pathogen escape versus pathogen exposure influences colonization and migration has been identifying multiple independent evolutionary events where species experience a change in pathogen pressure (for example, movement between areas of low and high pathogen diversity) while simultaneously quantifying pathogen-mediated selection. Here, we address this issue using phylogenetic data on 1,311 species of passerine birds (songbirds) to reconstruct colonization events across the Palaearctic and sub-Saharan African regions (hereafter, 'Africa'). These colonization events represent switches between areas with low (the Palaearctic) and high (Africa) pathogen diversity, due to an inverse relationship between pathogen richness and latitude[11–20]. The number of pathogens that species are exposed to has repeatedly been shown to increase closer to the Equator across a broad range of hosts[13–17], including birds[18–20]. First, we examine the evolutionary patterns of colonization between Africa and the Palaearctic among resident species, and test whether this has altered selection on immune genes (major histocompatibility complex class I (MHC-I)). Second, we reconstruct the evolutionary history of migratory species and then examine how breeding in the Palaearctic and wintering in Africa has influenced selection on immune genes (MHC-I) in long-distance migratory birds relative to resident species.

Results

Evolutionary patterns of colonization and their influence on immune genes. Consistent with the idea that colonization events are associated with parasite escape, we found—using stochastic character mapping (SCM)—that the rate of evolutionary transitions from high (Africa) to low (Palaearctic) pathogen areas was 16 times higher than the reverse (Fig. 2; Africa to the Palaearctic = 0.050, 95% quantile interval (QI) = 0.040 to 0.060; Palaearctic to Africa = 0.003, 95% QI = 0.001 to 0.003; percentage of stochastic maps where estimates were different (p) = 0.005; Supplementary Table 1). Similarly, reverse-jump Markov chain Monte Carlo (RJ-MCMC) transition rate models estimated that the transition rate from Africa to the Palaearctic was 20 times higher than from the Palaearctic to Africa (Fig. 2; Africa to the Palaearctic = 0.040, 95% credible interval (CI) = 0.030 to 0.050, Palaearctic to Africa = 0.002, CI = 0.001 to 0.003, percentage of transition rates that were higher from Africa to the Paleaearctic than the reverse = 0.001; Supplementary Table 2). These results suggest that the predominant pattern of colonization in Afro-Palaearctic passerine birds has been movement 'out of Africa' northwards to inhabit the Palaearctic.

Moving out of Africa to low-pathogen areas in the north is expected to reduce pathogen pressure. One way of testing this prediction is to examine selection on immune genes in species that have colonized low-pathogen areas relative to closely related species that have remained in high-pathogen areas. A central process to activating the adaptive immune response is the presentation of antigens

Molecular Ecology and Evolution Laboratory, Lund University, Lund, Sweden. *e-mail: emily.o_connor@biol.lu.se

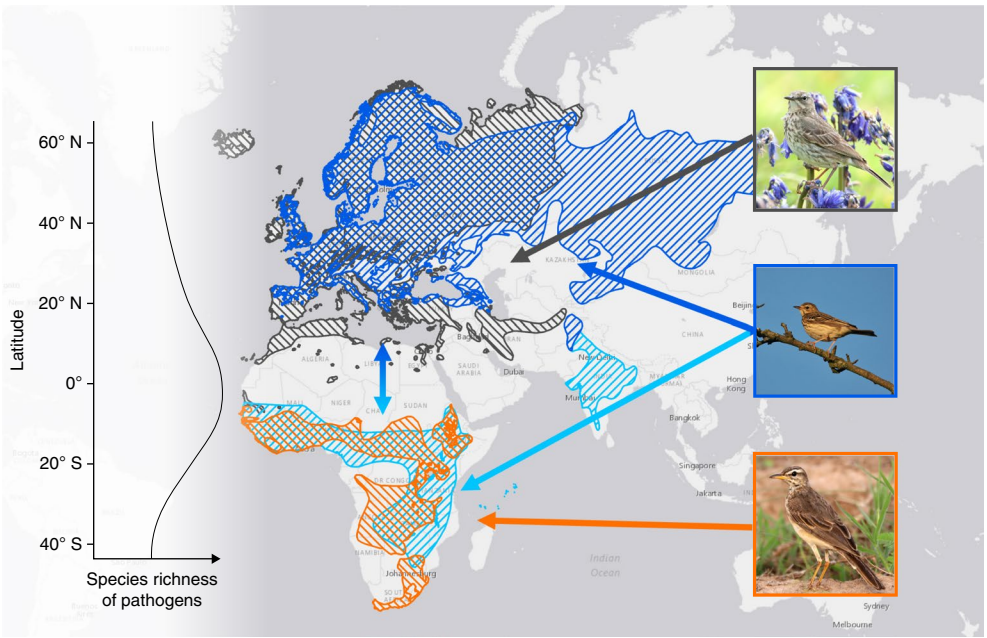

Fig. 1 | Colonization and migration change the pathogens that species face. Colonizations and migrations between Africa and the Palaearctic lead to changes in pathogen pressure due to a latitudinal gradient in pathogen richness[11–20]. Left: schematic representation of the increase in pathogen species richness at lower latitudes (adapted from ref. [11], University of Oxford). For example, Meadow pipits (*Anthus pratensis*, grey) live year-round in the Palaearctic and experience fewer pathogens than plain-backed pipits (*Anthus leucophrys*, orange), which remain year-round in Africa. Tree pipits (*Anthus trivialis*, turquoise (wintering range) and blue (breeding range)) experience both pathogen faunas, as they migrate annually between the Palaearctic and Africa. Map created in ArcMap 10.2.2 using distribution data provided by BirdLife International (ref. [58]). Photographs courtesy of Ray Seagrove (top), Roger Wasley (middle) and Christopher Berry (bottom).

to T cells by MHC proteins. Having more MHC alleles (number of different alleles per individual), as well as greater genetic divergence between alleles (P-distance)—hereafter collectively termed 'MHC diversity'—increases the range of antigens that can be presented to T cells by MHC proteins[21,22]. We used high-throughput sequencing to examine MHC-I of 25 resident species. These species were carefully selected to span the entire passerine clade, while representing closely related species that occur in Africa and the Palaearctic, to ensure comparisons across multiple independent evolutionary colonization events between Africa and the Palaearctic (Fig. 2).

Consistent with a latitudinal gradient in pathogen richness driving the differences in MHC-I diversity between African and Palaearctic species, we found that MHC-I diversity declined with distance from the Equator (Fig. 3). The two MHC-I diversity measures (the number of alleles and allele divergence) showed qualitatively similar patterns of decline, but differed quantitatively. The reduction in allele number with latitude was most pronounced close to the Equator; that is, in Africa (Fig. 3; African: posterior mode (PM) = −0.068, CI = −0.109 to −0.020, pMCMC (the number of iterations greater or less than zero for regressions, or the number of iterations when one level is greater or less than the other level when comparing groups, divided by the total number of iterations) < 0.01; Palaearctic: PM = 0.029, CI = −0.031 to 0.084, pMCMC = 0.15; Supplementary Table 3). In contrast, the decline in allele divergence was mostly at higher latitudes; that is, in the Palaearctic (Fig. 3; African: PM = −0.007, CI = −0.029 to 0.013, pMCMC = 0.21; Palaearctic: PM = −0.049, CI = −0.072 to −0.011, pMCMC < 0.01; Supplementary Table 4). In addition to analysing the relationship between the centroid latitude of species ranges and MHC-I diversity, we tested whether accounting for latitudinal span (the size of the range from north to south of African and Palaearctic species) influenced our results. We found that latitudinal span did not change the estimates of the relationship between latitude and MHC diversity (Supplementary Tables 3–8). The latitudinal gradient in MHC-I diversity led to Palaearctic species generally having lower MHC-I diversity compared with African species (Fig. 3; not controlling for latitude: P-distance PM = −0.071, CI = −0.338 to 0.139, pMCMC = 0.18; number of alleles PM = −0.237, CI = −0.804 to 0.194, pMCMC = 0.12; number of alleles excluding data from *Erithacus rubecula* (identified as an outlier in a box-and-whiskers plot) PM = −0.548, CI = −0.937 to −0.036, pMCMC = 0.01; controlling for latitude: P-distance PM = 0.564, CI = −0.076 to 1.313, pMCMC = 0.04; number of alleles PM = 0.705, CI = −0.794 to 2.423, pMCMC = 0.14; Supplementary Tables 3 and 4).

An alternative explanation for the differences between African and Palaearctic species in MHC-I diversity is that it reflects differences in neutral genetic processes as a result of demographic effects; for example, founder events during colonization or post-glacial bottlenecks in populations at high latitudes[23]. To investigate this possibility, we tested whether the inclusion of two measures of neutral genetic processes—synonymous substitutions across MHC-I alleles and 'haplotype redundancy' (whereby more than one allele on the nucleotide level encodes the same amino acid sequence)—in our statistical models altered the results. We used estimates from MHC genes, as opposed to genome-wide markers, as genes under balancing selection, such as MHC, are differentially affected by demographic events compared with neutral genetic markers[24–26]. Furthermore, the MHC region is known for having distinct patterns of evolution compared with the rest of the genome, such as a high evolutionary rate and frequent gene duplications[27–30].

We found that including synonymous substitutions across MHC-I alleles or haplotype redundancy did not explain the differences in MHC-I diversity between resident African and Palaearctic species (Supplementary Tables 3 and 4). There were also two African species with Palaearctic ancestry in our dataset (*Parus leucomelas* and *Turdus pelios;* Supplementary Tables 1 and 9) and they had among the highest MHC-I diversity of any of the species tested (Supplementary Fig. 1 and Supplementary Table 10).

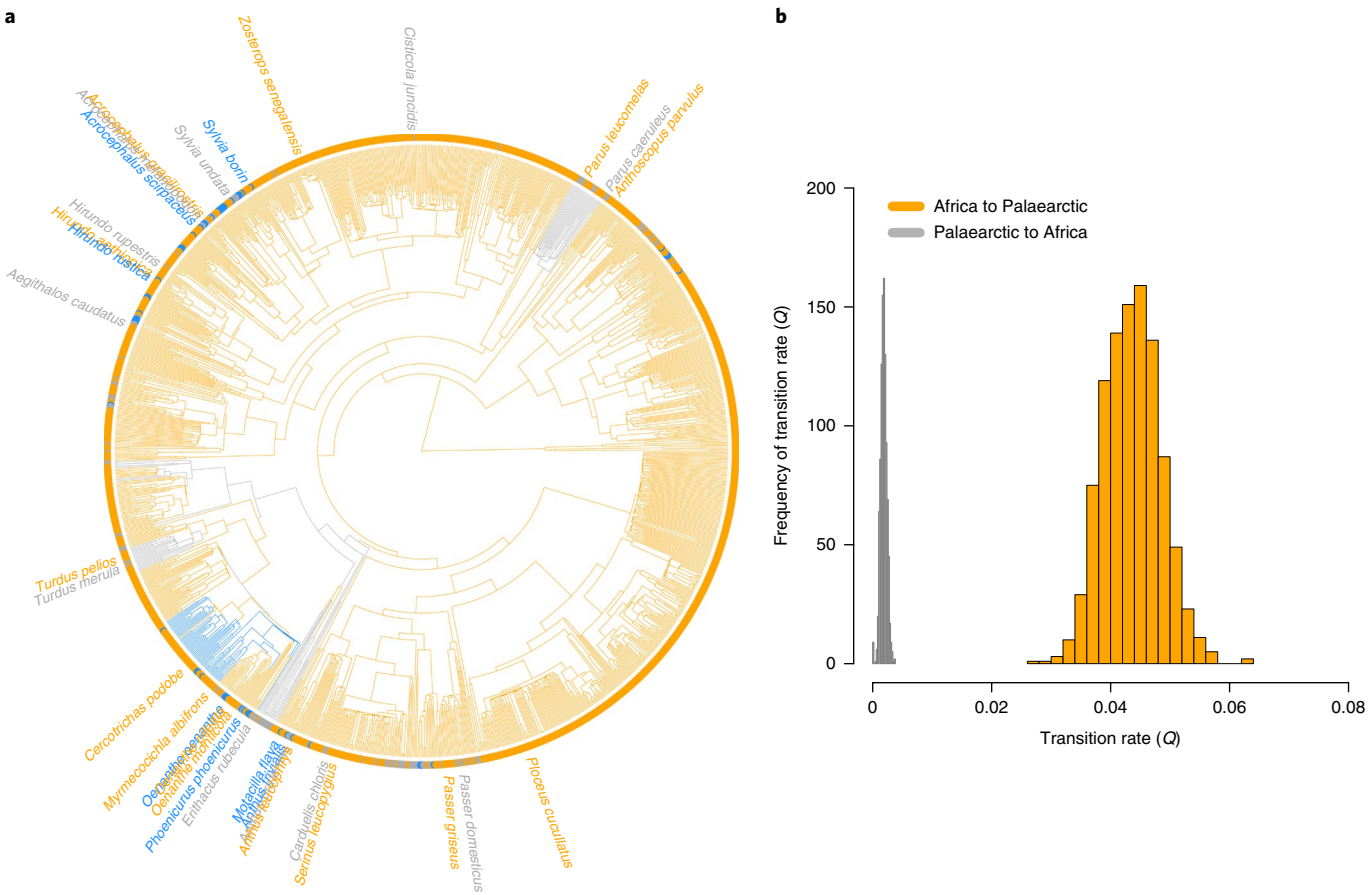

Fig. 2 | Evolutionary history of colonization events. a, Phylogenetic distribution of evolutionary transitions between inhabiting Africa (orange), inhabiting the Palaearctic (grey) and being migratory between the two (blue), depicted on a maximum clade credibility tree of the 1,000 trees used in the analysis. n = 1,311 species. Branches are coloured according to their predicted ancestral states estimated using a BPMM. Dots at the tips indicate species' current distributions and labelled tips indicate the 32 species that were MHC-I genotyped. **b**, Posterior distribution of transition rate estimates from the RJ-MCMC analysis showing that colonizations from Africa to the Palaearctic (orange) were significantly higher than transitions from the Palaearctic to Africa (grey).

This is the opposite pattern to the one expected if MHC-I diversity in Palaearctic species were constrained by demographic effects. Although this observation is based on just two species, it demonstrates that colonizations of Africa, and the associated elevated pathogen exposure, have been coupled with an increase in MHC-I diversity. In fact, the difference between *T. pelios* (the African resident) and *Turdus merula* (the Palaearctic resident) represented the greatest rate of change in MHC-I in our dataset: *T. pelios* had, on average, 28 more MHC-I alleles per individual than *T. merula*, which were estimated to have diverged around 10 million years ago (a mean change of 2.82 alleles Myr^{-1}).

We also examined evidence of positive selection on MHC-I through the proportion of positively selected sites (PSSs) and the strength of positive selection on the PSSs (omega values (ω)). These measures provide an indication of past selection on protein-coding changes[31]. Although we detected PSSs in nearly all the species tested, the proportion of PSSs and strength of positive selection did not differ either between African and Palaearctic species or across latitudes (Supplementary Fig. 2; proportion of PSSs: African versus Palaearctic species PM = −0.042, CI = −0.762 to 0.454, pMCMC = 0.28; latitude PM < 0.001, CI = −0.018 to 0.012, pMCMC = 0.42; ω: African versus Palaearctic species PM = 0.146, CI = −0.376 to 0.837, pMCMC = 0.17; latitude PM = 0.0072, CI = −0.010 to 0.021, pMCMC = 0.25; Supplementary Table 11). This suggests that while lower pathogen richness in the Palaearctic may have led to a reduction in MHC diversity, the strength of positive selection exerted on the antigen binding sites of the remaining alleles does not appear to have changed significantly. It is known that certain MHC alleles may specialize on particular pathogens[32–34] and that many avian pathogens that are present in Africa are missing in the Palaearctic, whereas others are found in both regions[35]. Together, this may have resulted in species that colonized the Palaearctic from Africa losing the alleles that no longer interacted with pathogens, while other alleles, which continue to interact with pathogens, remained and exhibit a similar signal of positive selection.

Evolution of migration and its influence on immune genes. Next, we examined the MHC-I diversity and evolutionary origins of long-distance migratory species that move biannually between Africa and the Palaearctic. These species provide a unique opportunity to tease apart whether pathogen escape or pathogen exposure within species influences immune-gene evolution. For instance, migration may enable an escape from pathogens during certain critical times of the year[6,18,36]. Alternatively, migrants may experience higher pathogen pressure due to being exposed to two separate pathogen communities[37].

We tested these two alternative hypotheses and found that MHC-I diversity of long-distance migrants was significantly lower than that of African residents, but similar to Palaearctic residents (Fig. 4; number of alleles: migrants versus African

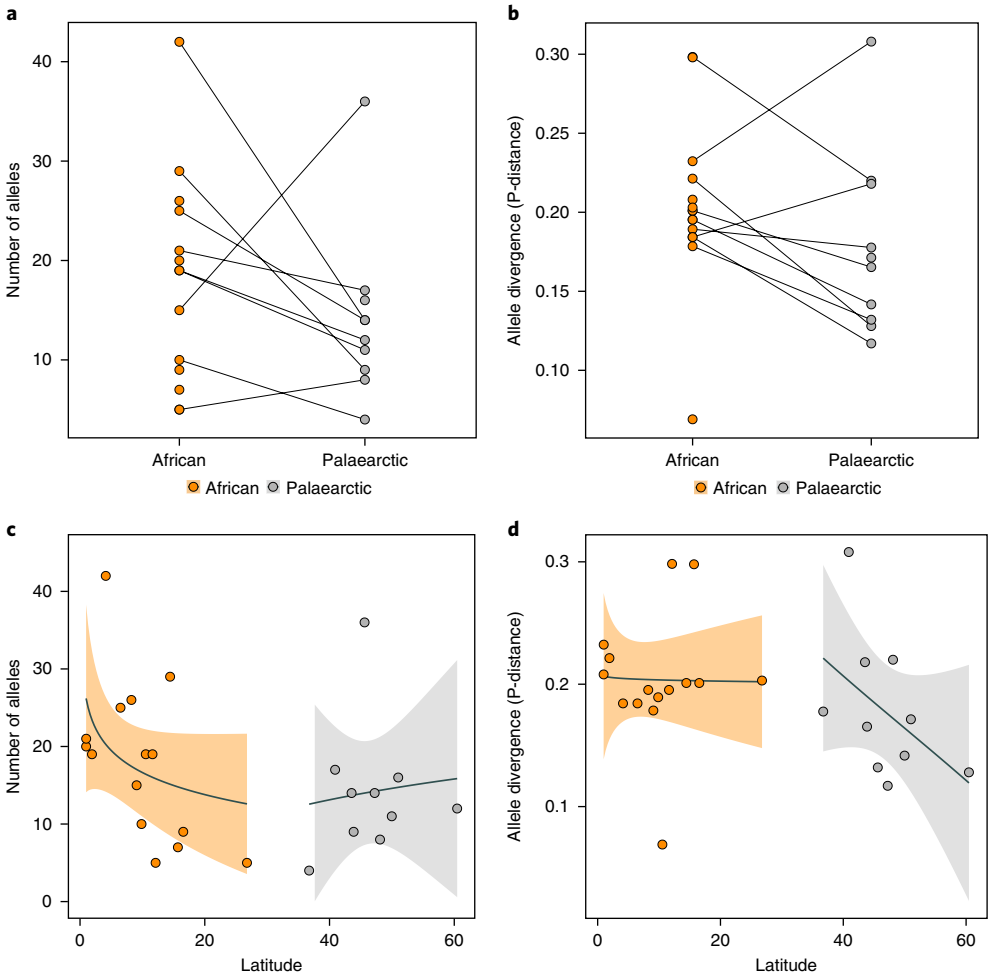

Fig. 3 | Changes in MHC-I diversity during the colonization process. a,**b**, MHC-I diversity (mean number of MHC-I alleles (**a**) and mean allelic divergence (**b**)) in African (orange, $n=15$ species) and Palaearctic (grey, $n=10$ species) residents with lines indicating taxonomically paired species (species from the same family or genus). **c**,**d**, Relationship between the absolute centroid latitude of species' ranges and MHC-I diversity (mean number of MHC-I alleles (**c**) and mean allelic divergence (**d**)) in African and Palaearctic residents. Latitude is plotted as absolute decimal degrees, regardless of orientation, to reflect the latitidinal gadient in parasite diversity seen both north and south of the Equator. Plotted lines are from linear regressions (log mean number of alleles in **c** and logit mean P-distance in **d**), with shaded areas representing 95% confidence intervals. No species were sampled from the Sahara Desert due to the paucity of bird species inhabiting this region.

residents PM = −0.471, CI = −0.902 to −0.119, pMCMC = 0.006; migrants versus Palaearctic residents PM = −0.100, CI = −0.707 to 0.701, pMCMC = 0.47; P-distance: migrants versus African residents PM = −0.199, CI = −0.390 to 0.004, pMCMC = 0.03; migrants versus Palaearctic residents PM = −0.048, CI = −0.491 to 0.296, pMCMC = 0.28; Supplementary Tables 5–8). These results arise despite the ancestral-state reconstructions indicating that all sequenced migrant species originated from African residents and thus, based on phylogenetic history, are expected to have high MHC-I diversity (Fig. 2 and Supplementary Tables 1 and 9). In line with our results on resident species, we found no evidence for a difference in positive selection on MHC-I between migrants and either Palaearctic or African residents (Supplementary Fig. 2 and Supplementary Tables 12 and 13). These results suggest that the evolution of migration to breed in more northerly regions is associated with pathogen escape, similar to resident species that have colonized the Palaearctic, resulting in a reduction in MHC-I diversity. Migrants may further reduce their pathogen exposure by breeding at particularly high latitudes, as we found that the species in our dataset bred significantly further north than the Palaearctic residents (PM = −7.553, CI = −13.941 to −1.935, pMCMC < 0.01; Supplementary Table 14).

Discussion

Moving away from areas with high pathogen richness—either by colonizing more northerly regions or migrating to breed at higher latitudes—appears to have resulted in an evolutionary loss of genetic diversity in the immune system of passerine birds. This implies that high MHC diversity is costly to maintain, potentially due to an increased risk of immunopathology through self-reactive T cells[38,39]. However, evidence of the costs of high MHC diversity has been difficult to demonstrate. Our data suggest that the relative benefits of recognizing and eliminating a diverse range of pathogens are increased in African resident birds compared with the possible costs of immunopathology. For Palaearctic species, the reverse seems to be true, leading to different levels of optimal MHC-I diversity between these regions. The rate at which such immunological adaptations evolve remains to be accurately dated. However, examination of the divergence times between the Palaearctic and migratory species that we genotyped for MHC-I and their most recent African ancestors across the dated phylogeny[40] indicates that this has occurred within the past 10.6 Myr (range: 4.5 to 27.2 Myr).

The reduction in MHC-I diversity seen in migratory species indicates that selection from pathogens during the breeding phase is particularly important for shaping the evolution of immunity. This may

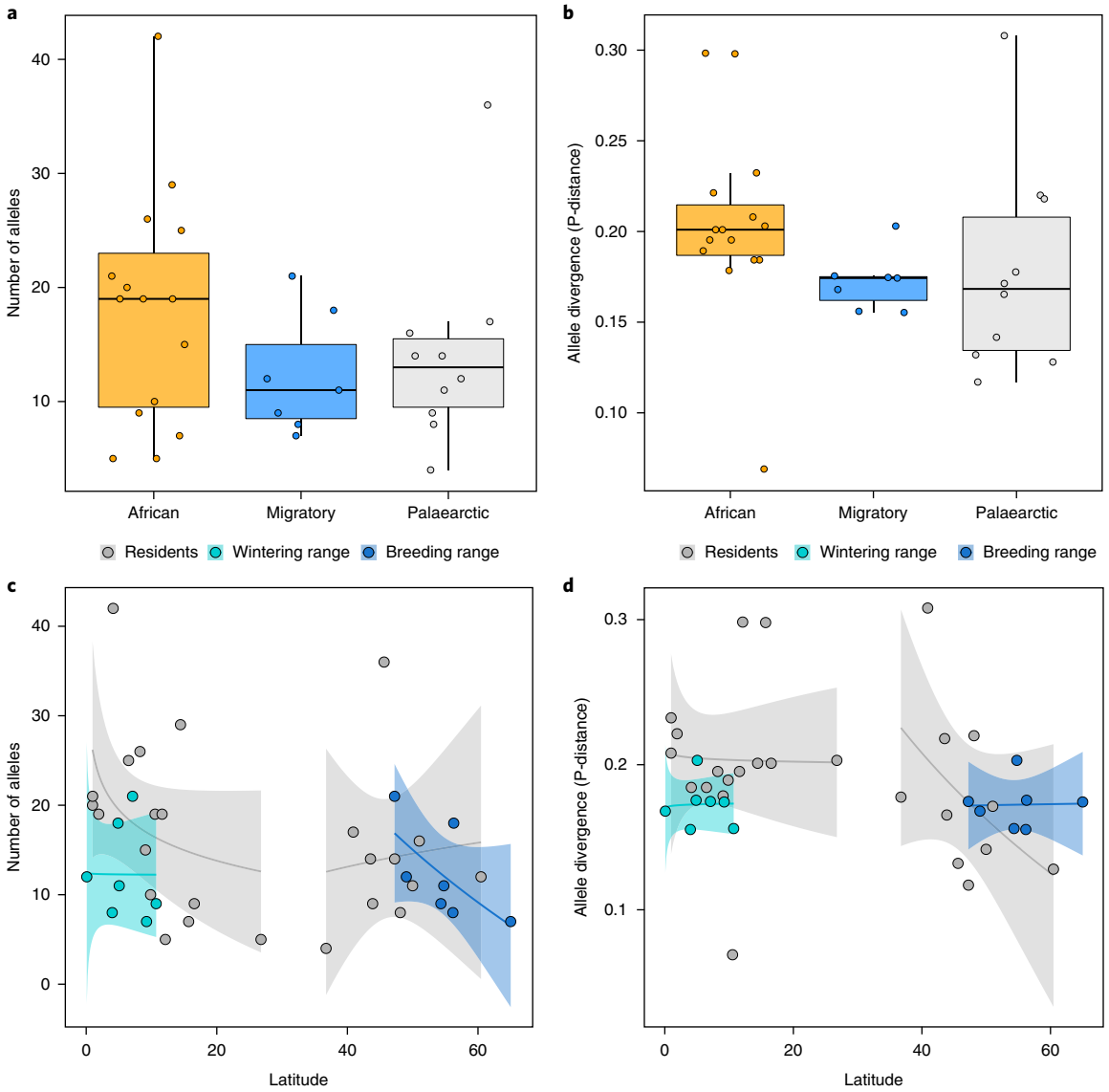

Fig. 4 | Changes in MHC diversity during the evolution of migration. a,**b**, MHC-I diversity (number of MHC-I alleles (**a**) and allelic divergence (**b**)) in African residents (orange, $n=15$ species), Palaearctic residents (grey, $n=10$ species) and migratory species (blue, $n=7$ species) shown in box plots (the central line depicts the median, the lower and upper hinges correspond to the first and third quartiles, and the whiskers extend to the highest value within 1.5 × the interquartile range) overlaid with the individual data points for each species. **c**,**d**, Relationship between the absolute centroid latitude of species' ranges and MHC-I diversity (mean number of MHC-I alleles (**c**) and mean allelic divergence (**d**)) in migratory species in their breeding (blue) and wintering (turquoise) ranges superimposed over that of resident species (grey). Latitude is plotted as absolute decimal degrees, regardless of orientation, to reflect the latitidinal gadient in parasite diversity seen both north and south of the Equator. Plotted lines are from linear regressions (log mean number of alleles in **c** and logit mean P-distance in **d**), with shaded areas representing 95% confidence intervals.

be due to a number of factors. First, breeding imposes high physiological demands on adults, so the trade-off between investment in reproduction and immunity is expected to be strongest during this time[41,42]. Second, exposure to pathogens during the juvenile phase, when the immune system is naïve, may be particularly important for the development of immunity. These two factors combined may mean that selection from pathogens on immune-gene diversity in migratory species is more similar to that of Palaearctic resident species, as opposed to African resident species. Nevertheless, processes occurring during the wintering period in Africa may also play a role in the reduction of MHC-I diversity. For example, migrants could avoid high pathogen pressure by occupying regions with low local pathogen richness[18,43].

The evolutionary origins of migration in the Afro-Palaearctic radiation from Africa northwards is different from that found in the New World migratory passerines. In the New World clade, migration appears to have evolved in North America and facilitated colonization of the tropics[44]. As such, the New World radiation of passerines includes Neotropical migrant and resident species that have evolved from temperate ancestors in North America. Although the immune genes of these species are yet to be characterized, the New World radiation provides a contrasting situation to test how colonization of high-pathogen areas is related to evolution of the immune system. We would expect Neotropical migrants to have similar levels of immune-gene diversity to the temperate residents in North America, as our results from the Afro-Palaearctic radiation suggest that the pathogen richness of the breeding range imposes the greatest selection on the immune system. However, Neotropical residents, in contrast with African residents, have faced an increase in pathogen richness as a result of colonizing from the

north, leading to the prediction that their immune-gene diversity should have increased over evolutionary time.

Climate-change models and empirical data show that pathogen ranges are expanding northwards[45,46]. Our results predict that host range-shifts to higher latitudes may help buffer species from this increased pathogen exposure, potentially negating the need for changes in immunity in response to climate change. However, when species hit their northern range limits, it remains to be seen whether they will have the capacity to evolve greater pathogen recognition, given the reduced diversity in their immune systems, or if these species will be vulnerable to reductions in population size and potentially extinctions.

Methods

Reconstructing colonization events between regions of high (Africa) and low (Palaearctic) pathogen richness. *Species selection.* The International Ornithological Congress (IOC) World Bird List version 6.2 (ref. [47]) and *Handbook of the Birds of the World Alive* (HBW) database[48] were used to identify passerine species that are either resident in the Palaearctic, resident in sub-Saharan Africa (hereafter 'Africa') or migrate between these two regions (hereafter 'migrants'). First, we extracted all species classed in the order Passeriformes from the IOC list. From this list, we extracted all species that had a breeding or non-breeding range in the Palaearctic or Africa. Species were classified as Palaearctic residents if they had a breeding range within the Palaearctic and did not have a non-breeding range elsewhere. Similarly, species were classified as African residents if they had a breeding range in Africa and no recorded non-breeding range elsewhere. Species were classified as migrants if they had a recorded breeding range in the Palaearctic and non-breeding range in Africa. In cases where the classification of ranges was broad within the IOC list, we cross-referenced these against the distribution maps in the HBW database for more information. We excluded any sedentary species with distribution ranges that spanned both the Palaearctic and Africa. In the final dataset, there were 1,311 species classified as either Palaearctic residents, African residents or migrants. The full list of species and classifications can be found in Supplementary Table 15.

Phylogenetic trees. Ref. [40] generated phylogenetic trees for nearly all bird species using Bayesian phylogenetic methods. We downloaded 1,500 trees from the posterior distribution for the 1,311 species in our dataset using the subset tool on the Bird Tree website (http://birdtree.org/), specifying the Hackett 'all-species backbone tree'[49]. We used this distribution of trees in all our analyses to account for the uncertainty in estimated phylogenies.

Evolutionary transition rate analyses. We examined the evolutionary transition rates between African-resident, Palaearctic-resident and migratory species across the 1,311 passerine species using two different methods: SCM and RJ-MCMC estimation. We used these two methods to check the robustness of our results to differing model assumptions (for details and estimation techniques, see refs [50,51]). Both modelling techniques gave very similar results (Supplementary Tables 1, 2 and 9).

We used SCM to determine the number and direction of transitions between Palaearctic-resident, African-resident and migratory species. In brief, this approach calculates the conditional likelihood that each node is a Palaearctic resident, African resident or migratory species, which is based on the estimated transition rate matrix (Q) between resident or migratory states and the length of the branch associated with that node. Based on these conditional likelihoods, ancestral states at each node are stochastically simulated and used in combination with observations at the tips to reconstruct a character history along each branch. Each character history is simulated using a continuous-time Markov chain where changes between states and the time spent in each state are modelled as a Poisson process (see ref. [51] for details).

We built 10 stochastic character maps for 1,000 trees (trees 501 to 1,500 from the downloaded sample—the same number of trees used in other analyses after burn-in periods) using an all-rates-different Q matrix with empirical Bayes estimation in the 'phytools' R package (R version 3.2.4)[52]. We used these 10,000 stochastic character maps to estimate the transition rates between Palaearctic-resident, African-resident or migratory states. We extracted the most likely state (the state with highest posterior probability) for each node for each tree in the sample of 1,000 trees. The mean number and s.d. of each type of transition were calculated across the 1,000 trees (Supplementary Table 1).

To estimate transition rates between African, Palaearctic and migratory states, we used the multistate module in BayesTraits version 2 with RJ-MCMC estimation[50]. We fitted an all-rates-different Q matrix to estimate the transition rates between all three states and ran models resampling from 1,000 trees (trees 501 to 1500 from the downloaded sample). We examined the likelihood of transitions occurring by examining the proportion of models visited by the RJ-MCMC algorithm where the rates were assigned to zero (Supplementary Table 2). We tested whether transition rates were significantly different from each other by calculating the mean difference in transition rates and 95% CI (significantly different when it did not span 0), together with the percentage of transition rates where one transition rate was higher than the other. We used hyper priors where values were drawn from a uniform distribution with a range 0–100 to seed the mean and variance of an exponential prior to reduce uncertainty over prior selection[50]. The prior settings were chosen according to the estimated range of transition rates obtained using analyses with maximum-likelihood estimation. We also examined the sensitivity of our models to prior selection by running models with gamma priors seeded using hyper priors and recovered similar results. We only present the results from the models using exponential priors as the mixing properties of the MCMC from these models were better than other priors. We ran each model three times for a total of 6,000,000 iterations and a burn-in of 1,000,000 iterations, and sampled every 5,000 iterations.

We examined the convergence of models by repeating each analysis three times and examining the correspondence between chains in R using the 'coda' package version 0.16-1 (ref. [53]) by: (1) visually inspecting the traces of the MCMC posterior estimates and their overlap; (2) calculating the autocorrelation and effective sample size of the posterior distribution of each chain; and (3) using Gelman and Rubin's convergence diagnostic test, which compares within- and between-chain variance using a potential scale reduction factor[54]. Values substantially higher than 1.1 indicate chains with poor convergence properties. The potential scale reduction factor was less than 1.1 for all the parameter estimates presented.

Reconstructing the ancestral states of the MHC-I-typed species. We estimated the probability that the ancestors of the 1,311 passerine species described in the above section 'Species selection' evolved from African-resident, Palaearctic-resident or migratory species, and extracted this information for the 32 species we MHC-I typed (see below for details of the MHC-I species). Again, we used two different analytical approaches to ensure the robustness of our estimates: SCM and Bayesian phylogenetic mixed models (BPMMs) (Supplementary Tables 1 and 9).

From the SCM models described in the section 'Evolutionary transition rate analyses', we estimated the ancestral state for each of the species in our MHC-I dataset by: (1) taking the most frequently estimated ancestral state for each species across the ten simulations per tree to reduce uncertainty in estimating states for individual trees; and (2) calculating the proportion of trees where that node was assigned to each state for each species across the 1,000 phylogenetic trees (Supplementary Table 1). This was to account for uncertainty in estimating states across phylogenetic trees.

We used a BPMM with a multinomial response variable (1 = Palaearctic residents, 2 = migrants, 3 = African residents) with a logit link function to estimate the probability that each node came from Africa, came from the Palaearctic or was migratory in the R package MCMCglmm[55]. For random effects, we examined three different priors: inverse-Wishart priors (V = 1, nu = 0.002) and two parameter expanded priors (Fisher prior: V = 1, nu = 1, alpha.mu = 0, alpha.V = 1,000) and (χ2 prior: V = 1, nu = 1,000, alpha.mu = 0, alpha.V = 1)[56]. In multinomial models, the residual variance cannot be identified and this was therefore was set to 1. For the intercept, we specified a prior of mu = 0, V = σ2units + π2/3, which is approximately flat on the probability scale when a logit link function is used. This improved the mixing properties of the chain. Models were iterated across a sample of 1,500 trees. For each tree, we set a burn-in of 49,999 iterations and saved iteration 50,000. Estimates from the last iteration from tree i were used to as starting parameter values for tree $i+1$. We discarded samples from the first 500 trees, saving one estimate from each of 1,000 trees. The convergence of models was examined in the same way as described in the section 'Evolutionary transition rate analysis'.

Characterizing MHC-I genes across species. *Species selection for MHC-I characterization.* It is possible to download MHC sequence data of various passerine species from GenBank to study selection on MHC genes (for example, ref. [57]). However, we did not use this approach as differences in the use and performance of primer pairs between studies can result in non-comparable sets of alleles being available for different species. Furthermore, the MHC data available on GenBank represent a highly mixed phylogenetic distribution of species and it is sometimes not possible to ascertain where samples were collected from. This makes it extremely difficult to infer the evolutionary processes influencing MHC genes and guard against biases in sampling. Instead, we selected 32 species across the parvorder Passerida[58] for the collection of blood samples for MHC-I genotyping. These species were chosen to include sets of species across multiple taxonomic families that included Palaearctic residents, sub-Saharan African (African) residents and birds that breed in the Palaearctic but perform long-distance migrations to winter in Africa (migrants). Where possible, sets of Palaearctic-resident, African-resident and migrant species were selected from the same genus or taxonomic family. We used the HBW resource[48] for information on the distribution and movement of different bird species. We selected 10 Palaearctic-resident species, 15 African-resident species and 7 migrant species. These species spanned 14 families and 21 genera and were selected to provide a broad phylogenetic spread across Passerida (Supplementary Fig. 3). The full list of species and their classifications are provided in Supplementary Table 16.

Latitude of species' ranges. We estimated the centroid latitude of each species distribution as well as the number of degrees latitude spanned by each species' distribution range ('latitudinal span'). In the case of the migrants, there were two ranges (their breeding and wintering ranges). Details of each species' distribution were obtained from ESRI data shapefiles provided by BirdLife International[58] using the geospatial programming software ArcMap version 10.2.2 (http://desktop.arcgis.com/en/arcmap/). Full details of how data were extracted from the shapefiles can be found in the Supplementary Methods.

For the purposes of data analyses, the absolute latitude was used, whereby only the decimal-degrees distance from the Equator was considered and not the orientation (north or south). This approach was to account for the latitudinal gradient in parasite diversity seen both north and south of the Equator[11,12].

Sample collection and preparation. Blood samples (20–40 µl) were collected from the brachial veins of three individuals from each species before they were released back into the wild (for details, including in species where fewer than three individuals were sampled, see Supplementary Table 16). Three individuals should be sufficient to gain MHC-I estimates that are comparable across species (see Supplementary Methods). The total sample size was 81 individuals. Blood was stored in 500 µl of buffer (150 mM saline (NaCl), 1 mM EDTA, 50 mM Tris, pH 8.0) at −20 °C until required for DNA extraction. Genomic DNA was extracted using a standard salt-based protocol and stored at −20 °C.

Polymerase chain reaction (PCR) and 454 sequencing. Exon 3 amplicons of MHC-I alleles were amplified using three different combinations of three forward and three reverse primers (Supplementary Fig. 4). These primers were designed based on known MHC-I sequences for 12 Passerida species[59], which are spread across the parvorder Passerida[60]. Duplicate PCRs were performed on each individual for each of the three primer pairs. Amplicons were sequenced using bi-directional pyrosequencing on the 454 GS FLX system by 454/Roche at the Lund University Sequencing Facility (Faculty of Science). Full details of the primers used, primer performance, PCR conditions and 454 sequencing can be found in the Supplementary Methods.

The raw 454 data were de-multiplexed, clustered and filtered using the programme AmpliSAS[61] with the aid of the technical replicates. Full details of these steps can be found in the Supplementary Methods.

Genotyping and classification of alleles. The final set of verified alleles for each individual for each of the three primer pairs was compiled to merge any identical alleles. Alleles were considered 'identical' and merged if they had 100% sequence identity for the full length of their sequence overlap. We used the De Novo Assemble option within Geneious R 9.0.2 with customized sensitivity settings to identify and merge identical alleles and create a FASTA file of the final genotype for each individual. Merged alleles could be up to 80 amino acids in length (amino acid positions 2–81) if fragments were amplified by all three primer pairs with identical sequences in their overlapping sections (Supplementary Fig. 5). For details on the agreement between the three primer pairs, see Supplementary Table 10. All verified alleles were uploaded to GenBank (accession codes: MF477947–MF478976).

Estimating MHC-I diversity and selection on alleles. All alleles were manually aligned in the BioEdit version 7.2.5 Sequence Alignment Editor. This was first performed on the alleles for each individual separately and then on the alleles found in each species. Putatively functional alleles were identified by the presence of two highly conserved cysteine residues in this region (positions 11 and 74; Supplementary Fig. 5). These residues are fundamental to disulphide-bridge formation and thus to peptide binding[62]. Alleles were considered non-functional if they were missing either of the aforementioned cysteines, contained stop codons (as this region is within an exon), or contained deletions or insertions that shifted the reading frame. Only functional alleles were included in the analyses of this study. The positions of codons believed to represent the peptide-binding region (PBR) in this fragment of exon 3 for MHC-I were inferred from the PBR of human leukocyte antigen A[63] (Supplementary Fig. 5).

MHC-I diversity was calculated on an individual level using two main approaches. First, the number of different MHC-I alleles on a nucleotide level was counted. Second, the amino acid sequence divergence (P-distance) between the alleles possessed by each individual was calculated using Molecular Evolutionary Genetics Analysis version 6 (ref. [64]). P-distance was calculated on the 33 amino acids that represented the area of overlap between the 3 primer pairs (positions 35–67; Supplementary Fig. 5) as these amino acid positions were represented in every individual for every species. P-distance was calculated separately for the PBR sites in these 33 amino acids (6 amino acid positions) and the non-PBR sites (27 amino acid positions). We checked that the P-distance estimates calculated from the 33-amino-acids section were representative of estimates calculated from the full length of the sequence (81 amino acids) by also calculating P-distance using the full length of the sequence for the alleles where this was possible. These two P-distance estimates were highly correlated (PBR sites: $n=32$, Spearman's rank correlation coefficient (r_s) = 0.88, $P<0.001$; non-PBR sites: $n=32$, $r_s=0.94$, $P<0.001$).

Estimating selection on MHC-I. To estimate selection on MHC-I, we calculated metrics of positive selection across the sequence overlap between the three primer pairs (99 nucleotides; 33 amino acids) for each species. These metrics were the proportion of sites under positive selection and the strength of positive selection (ω) for these sites. For this, we used the species-level alignment of alleles. M0 + gamma and M5 gene trees were generated in CodonPhyML[65] for the set of alleles belonging to each species. Both trees were used to estimate PSSs. However, as the M5 trees yielded the highest log likelihood, only the results using the M5 tree are presented. The proportion of PSSs and ω of PSSs estimated using the two trees were highly correlated (proportion of PSSs: $n=32$, $r_s=0.98$, $P<0.001$; ω: $n=32$, $r_s=0.98$, $P<0.001$). The CODEML programme within Phylogenetic Analysis by Maximum Likelihood (PAML)[31] was used to test the following models of codon evolution: M0, M3, M7, M8 and M8a. The best-fitting model was selected based on log likelihood ratio tests between the nested models (M0 versus M3, M7 versus M8, and M8 versus M8a). In all but two species (*Oenanthe monticola* and *Serinus leucopygius*), there was evidence of sites under significant positive selection. The Bayes empirical Bayes method in PAML was used to estimate which sites were under positive selection using a 0.80 threshold for significance[66].

As recombination can influence the outcome of selection tests, we also estimated the number of sites under positive selection and recombination rates simultaneously using the programme omegaMap[67]. The numbers of PSSs identified in species using the Bayes empirical Bayes approach from the M8 model in PAML and omegaMap were correlated ($n=32$, $r_s=0.38$, $P=0.03$). We identified PSSs (posterior probability of selection above 0.90) in every species tested using omegaMap. Thus, PSSs were identified even in the presence of recombination, suggesting that the PSS results from the M8 models are robust to the effects of recombination. Full details of the PSS results can be found in Supplementary Table 17.

Measure of neutral genetic processes. Two measures of neutral genetic processes within the MHC region were estimated: 'haplotype redundancy' and the synonymous substitution rate (dS) across the non-PBR sites of MHC-I. Haplotype redundancy refers to more than one allele, at the level of the nucleotide sequence, coding for the same allele at the level of the amino acid sequence. Lower haplotype redundancy is expected in populations that have undergone demographic processes that reduce population sizes and increase the impact of genetic drift; for example, genetic bottlenecks and founder events[68]. Haplotype redundancy was calculated by dividing the number of unique alleles at the nucleotide-sequence level by the number of alleles at the amino acid sequence level for each individual. The second measure of neutral genetic processes was dS of the non-PBR sites. Under the expectations of neutral theory, dS reflects the mutation rate and thereby neutral genetic variation, which is expected to be higher in populations with a more stable demographic history[68,69]. We assume a comparable mutation rate between the species in our dataset, as dS has been shown to be similar between Old World warblers and Old World flycatchers in multiple genes[70] and these families span the phylogenetic range of the species in our dataset (Supplementary Fig 3). dS across the non-PBR sites of MHC-I for each species was calculated using the Nei-Gojobori model in Molecular Evolutionary Genetics Analysis (MEGA) version 6 (ref. [64]). The area of sequence overlap between the three primer pairs was used for these analyses to make the results comparable across species. Haplotype redundancy and dS were not statistically linked in our dataset (PM = −0.374, CI = −0.139 to 1.80, pMCMC = 0.41) as, although both measures are determined by synonymous mutations, they are calculated on different levels: dS reflects the rate of synonymous mutations per synonymous site, whereas haplotype redundancy reflects the occurrence of one or more synonymous mutations across the entire allele. Haplotype redundancy and dS were included in the main statistical models on the resident species to test whether incorporating the effect of neutral genetic processes influenced the relationship between region and latitude on MHC diversity (Supplementary Tables 3 and 4).

Phylogenetic trees. For the purpose of statistical analyses, 1,500 phylogenies for the 32 species in which we genotyped MHC-I were obtained using the subset tool on the Bird Tree website (http://birdtree.org/)[40], specifying the Hackett all-species backbone tree[49].

Statistical analyses. The effect of being an African resident, Palaearctic resident or migrant, as well as latitude, on estimates of MHC-I diversity and positive selection were analysed using three subsets of data. First, we tested whether MHC-I diversity and the strength of positive selection on these genes was related to latitude and was different between African and Palaearctic resident species. The dataset used for these tests contained only the resident species (Palaearctic residents and African residents). Second, we tested whether there are differences in MHC-I diversity and positive selection on these genes between migratory species and: (1) Palaearctic-resident species controlling for differences in latitude to the migrants' breeding ranges; and (2) African-resident species controlling for differences in latitude to the migrants' wintering ranges. The approach of separating these datasets enabled us to assess the role of the breeding and wintering ranges of migrants independently and to make comparisons with the species inhabiting the corresponding ranges year-round (the breeding ranges of migrants versus the Palaearctic species ranges, and the wintering ranges of migrants versus the African species ranges).

All analyses were conducted using BPMMs using the R package 'MCMCglmm'[55]. The error distributions used to model the different response variables were: (1) a Poisson error distribution for the number of alleles; (2) a binomial error distribution for P-distance, the proportion of PSSs and ω; and (3) a Gaussian error distribution for the centroid latitude of the distributions of Palaearctic species and the breeding ranges of migrants.

We tested the effect of resident geographical range (African resident or Palaearctic resident) and migration, latitude and their interaction on the MHC-I measures by fitting them as fixed effects. For the number of alleles and P-distance, we had data on multiple individuals per species. To account for the non-independence of these data and estimate the amount of variation within species in our response variables, we fitted species as a random effect. We accounted for the effect of phylogenetic relationships on the response variables by incorporating a variance–(co)variance matrix of evolutionary distances calculated from the species phylogeny fitted as a random effect.

To account for uncertainty in phylogenetic relationships, we ran our models on 1,500 trees. For each tree, models were run for 10,000 iterations with a burn-in of 9,999 and the final iteration was saved. As each set of 10,000 iterations was conducted on each of 1,500 trees, this generated 1,500 posterior samples from which we discarded the first 500 samples as burn-in. From the final 1,000 samples, we estimated the effect of each parameter on the response variable using the PM and 95% CIs. Terms were considered statistically significant when 95% CIs did not span 0 and pMCMC values (the number of iterations greater or less than zero for regressions, or the number of iterations when one level is greater or less than the other level when comparing groups, divided by the total number of iterations) were less than 0.05 (ref. [55]).

We specified inverse-Wishart priors ($V=1$, $nu=0.002$) for random effects and the default priors in MCMCglmm for fixed effects (independent normal priors with zero mean and large variance (10^{10})). These priors led to all models converging (see the section 'Evolutionary transition rate analyses' for assessment of model convergence).

Full details of all statistical tests conducted in this study and their parameter estimates can be found in the Supplementary Information (Supplementary Tables 1–9 and 11–14), along with the R code used to run all statistical models.

By choosing multiple independent evolutionary transitions, where closely related species have switched between being African residents, Palaearctic residents and migrants, variation in multiple population-level and life-history characteristics is implicitly accounted for as they are known to be highly phylogenetically conserved[71,72]. However, as body size is correlated with multiple life-history variables and ecological constraints that could interact with pathogen exposure[72–74], we re-ran our main analyses with the mean body mass of species[75] included as a covariate to account for potentially confounding factors. We found no qualitative difference in the results (Supplementary Tables 3–8).

Ethics. All birds used in this study were released back into the wild after blood samples had been taken. The project was approved by the Ethical Committee of Lund/Malmö, and blood samples were collected under a licence issued by the relevant authority in Sweden (Jordbruksverket) with the licence code M 160-11.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability. R script containing the code for all analyses run are provided in the Supplementary Information.

Data availability. Sequence data that support the findings of this study have been deposited in GenBank with the accession codes MF477947–MF478976.

Received: 4 September 2017; Accepted: 19 February 2018;
Published online: 9 April 2018

References

1. Bauer, S. & Hoye, B. J. Migratory animals couple biodiversity and ecosystem functioning worldwide. *Science* **344**, 1242552 (2014).
2. Pecl, G. T. et al. Biodiversity redistribution under climate change: impacts on ecosystems and human well-being. *Science* **355**, eaai9214 (2017).
3. Helmus, M. R., Mahler, D. L. & Losos, J. B. Island biogeography of the Anthropocene. *Nature* **513**, 543–546 (2014).
4. Sax, D. F. et al. Ecological and evolutionary insights from species invasions. *Trends Ecol. Evol.* **22**, 465–471 (2007).
5. Torchin, M. E., Lafferty, K. D., Dobson, A. P., McKenzie, V. J. & Kuris, A. M. Introduced species and their missing parasites. *Nature* **421**, 628–630 (2003).
6. Altizer, S., Bartel, R. & Han, B. A. Animal migration and infectious disease risk. *Science* **331**, 296–302 (2011).
7. Mitchell, C. E. & Power, A. G. Release of invasive plants from fungal and viral pathogens. *Nature* **421**, 625–627 (2003).
8. Westerdahl, H. et al. in *Animal Movement Across Scales* (eds Hansson, L. A. & Åkesson, S.) Ch. 8 (Oxford Univ. Press, Oxford, 2014).
9. Keane, R. M. & Crawley, M. J. Exotic plant invasions and the enemy release hypothesis. *Trends Ecol. Evol.* **17**, 164–170 (2002).
10. Almberg, E. S., Cross, P. C., Dobson, A. P., Smith, D. W. & Hudson, P. J. Parasite invasion following host reintroduction: a case study of Yellowstone's wolves. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2840–2851 (2012).
11. Guégan, J.-F., Prugnolle, F. & Thomas, F. in *Evolution in Health and Disease* (eds Stearns, S. C. & Koella, J. C.) Ch. 2 (Oxford Univ. Press, Oxford, 2008).
12. Guernier, V., Hochberg, M. E. & Guégan, J.-F. Ecology drives the worldwide distribution of human diseases. *PLoS Biol.* **2**, e141 (2004).
13. Bordes, B., Guégan, J. F. & Morand, S. Microparasite species richness in rodents is higher at lower latitudes and is associated with reduced litter size. *Oikos* **120**, 1889–1896 (2011).
14. Nunn, C. L., Altizer, S. M., Sechrest, W. & Cunningham, A. A. Latitudinal gradients of parasite species richness in primates. *Divers. Distrib.* **11**, 249–256 (2005).
15. Yang, X. B. & Feng, F. Ranges and diversity of soybean fungal diseases in North America. *Phytopathology* **91**, 769–775 (2001).
16. Rohde, K. Ecology and biogeography of marine parasites. *Adv. Mar. Biol.* **43**, 1–83 (2002).
17. Wellman, F. L. More diseases on crops in the tropics than in the temperate zone. *Ceiba* **14**, 17–28 (1968).
18. Clark, N. J., Clegg, S. M. & Klaassen, M. Migration strategy and pathogen risk: non-breeding distribution drives malaria prevalence in migratory waders. *Oikos* **125**, 1358–1368 (2016).
19. Mendes, L., Piersma, T., Lecoq, M., Spaans, B. & Ricklefs, R. E. Disease-limited distributions? Contrasts in the prevalence of avian malaria in shorebird species using marine and freshwater habitats. *Oikos* **109**, 396–404 (2005).
20. Merino, S. et al. Haematozoa in forest birds from southern Chile: latitudinal gradients in prevalence and parasite lineage richness. *Austral Ecol.* **33**, 329–340 (2008).
21. Lenz, T. L. Computational prediction of MHC II–antigen binding supports divergent allele advantage and explains trans-species polymorphism. *Evolution* **65**, 2380–2390 (2011).
22. Murphy, K., Janeway C. A. Jr., Travers, P., Walport, M. & Ehrenstein, M. *Janeway's Immunobiology* 7th edn (Garland Science, New York, 2008).
23. Hughes, A. L. & Hughes, M. A. K. Coding sequence polymorphism in avian mitochondrial genomes reflects population histories. *Mol. Ecol.* **16**, 1369–1376 (2007).
24. Aguilar, A. et al. High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc. Natl Acad. Sci. USA* **101**, 3490–3494 (2004).
25. Schierup, M. H., Vekemans, X. & Charlesworth, D. The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genet. Res.* **76**, 51–62 (2000).
26. Van Oosterhout, C. et al. Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution* **60**, 2562–2574 (2006).
27. Sorci, G. Immunity, resistance and tolerance in bird–parasite interactions. *Parasite Immunol.* **35**, 350–361 (2013).
28. Eizaguirre, C., Lenz, T. L., Kalbe, M. & Milinski, M. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. *Nat. Comm.* **3**, 621 (2012).
29. Piertney, S. B. & Oliver, M. K. The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7–21 (2006).
30. Edwards, S. V. & Hedrick, P. W. Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol. Evol.* **13**, 305–311 (1998).
31. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
32. Kiepiela, P. et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* **432**, 769–775 (2004).
33. McKiernan, S. M. et al. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology* **40**, 108–114 (2004).
34. Wallny, H. J. et al. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proc. Natl Acad. Sci. USA* **103**, 1434–1439 (2006).
35. Hellgren, O. et al. Detecting shifts of transmission areas in avian blood parasites—a phylogenetic approach. *Mol. Ecol.* **16**, 1281–1290 (2007).
36. Piersma, T. Do global patterns of habitat use and migration strategies co-evolve with relative investments in immunocompetence due to spatial variation in parasite pressure? *Oikos* **80**, 623–631 (1997).
37. Westerdahl, H., Wittzell, H. & von Schantz, T. Mhc diversity in two passerine birds: no evidence for a minimal essential Mhc. *Immunogenetics* **52**, 92–100 (2000).
38. Abbas, A. K., Lichtman, A. H. & Pillai, S. *Basic Immunology: Functions and Disorders of the Immune System* 4th edn (Elsevier Saunders, Philadelphia, 2014).
39. Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B. H. & Milinski, M. Parasite selection for immunogenetic optimality. *Science* **301**, 1343 (2003).

40. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
41. Buehler, D. & Piersma, T. Travelling on a budget: predictions and ecological evidence for bottlenecks in the annual cycle of long-distance migrants. *Philos. Trans. R. Soc. B* **363**, 247–266 (2008).
42. Hasselquist, D. Comparative immunoecology in birds: hypotheses and tests. *J. Ornithol.* **148**, 571–582 (2007).
43. Yohannes, E. et al. Isotope signatures in winter moulted feathers predict malaria prevalence in a breeding avian host. *Oecologia* **158**, 299–306 (2008).
44. Winger, B. M., Barker, F. K. & Ree, R. H. Temperate origins of long-distance seasonal migration in New World songbirds. *Proc. Natl Acad. Sci. USA* **111**, 12115–12120 (2014).
45. Harvell, C. D. et al. Climate warming and disease risks for terrestrial and marine biota. *Science* **296**, 2158–2162 (2002).
46. Altizer, S., Ostfeld, R. S., Johnson, P. T., Kutz, S. & Harvell, C. D. Climate change and infectious diseases: from evidence to a predictive framework. *Science* **341**, 514–519 (2013).
47. Gill, F. & Donsker, D. *IOC World Bird List* Version 6.2 (International Ornithological Congress, 2016); https://doi.org/10.14344/IOC.ML.6.2
48. Del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & de Juana E. *Handbook of the Birds of the World Alive* (Lynx Edicions, Barcelona, 2016).
49. Hackett, S. J. et al. A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768 (2008).
50. Pagel, M. A. & Meade, A. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* **167**, 808–825 (2006).
51. Bollback, J. P. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinforma.* **7**, 88 (2006).
52. Revell, L. J. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
53. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R. News* **6**, 7–11 (2006).
54. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–511 (1992).
55. Hadfield, J. MCMC methods for multi-response generalised linear mixed models: the MCMCglmm R package. *J. Stat. Softw.* **33**, 1–22 (2010).
56. de Villemereuil, V., Gimenez, O. & Doligez, B. Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits. *Methods Ecol. Evol.* **4**, 260–275 (2013).
57. Minias, P., Whittingham, L. A. & Dunn, P. O. Coloniality and migration are related to selection on MHC genes in birds. *Evolution* **71**, 432–441 (2017).
58. *Bird Species Distribution Maps of the World* Version 5.0. (BirdLife International & NatureServe, 2015).
59. Barker, F. K., Cibois, A., Schikler, P., Feinstein, J. & Cracraft, J. Phylogeny and diversification of the largest avian radiation. *Proc. Natl Acad. Sci. USA* **101**, 11040–11045 (2004).
60. O'Connor, E. A., Strandh, M., Hasselquist, D., Nilsson, J. Å. & Westerdahl, H. The evolution of highly variable immunity genes across a passerine bird radiation. *Mol. Ecol.* **25**, 977–989 (2016).
61. Sebastian, A., Herdegen, M., Migalska, M. & Radwan, J. AmpliSAS: a web server for multilocus genotyping using next-generation amplicon sequencing data. *Mol. Ecol. Res.* **16**, 498–510 (2016).
62. Peaper, D. R. & Cresswell, P. Regulation of MHC class I assembly and peptide binding. *Annu. Rev. Cell Dev. Biol.* **24**, 343–368 (2008).
63. Bjorkman, P. et al. The foreign antigen binding site and T cell recognition regions. *Nature* **329**, 512–518 (1987).
64. Tamura, T., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
65. Gil, M., Zanetti, M. S., Zoller, S. & Anisimova, M. CodonPhyML: fast maximum likelihood phylogeny estimation under codon substitution models. *Mol. Biol. Evol.* **30**, 1270–1280 (2013).
66. Yang, Z., Wong, W. S. & Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118 (2005).
67. Wilson, D. J. & McVean, G. Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* **172**, 1411–1425 (2006).
68. Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer Associates, Sunderland, 1998).
69. Carnaval, A. C., Hickerson, M. J., Haddad, C. F., Rodrigues, M. T. & Moritz, C. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science* **323**, 785–789 (2009).
70. Lanfear, R., Ho, S. Y., Love, D. & Bromham, L. Mutation rate is linked to diversification in birds. *Proc. Natl Acad. Sci. USA* **107**, 20423–20428 (2010).
71. Garamszegi, L. Z. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology* (Springer, Berlin, 2014).
72. Brown, J. H. in *Macroecology* Ch. 5 (Univ. Chicago Press, Chicago, 1995).
73. Blueweiss, L., Fox, H., Kudzma, V., Nakashima, D., Peters, R. & Sams, S. Relationships between body size and some life history parameters. *Oecologia* **37**, 257–272 (1978).
74. Nunn, C. L., Altizer, S., Jones, K. E. & Sechrest, W. Comparative tests of parasite species richness in primates. *Am. Nat.* **162**, 597–614 (2003).
75. Dunning, J. B. Jr. *CRC Handbook of Avian Body Masses* (CRC Press, Boca Raton, 2013).

Acknowledgements

This report received support from the Centre for Animal Movement Research, financed by a Linnaeus grant (349–2007–8690) from the Swedish Research Council and Lund University, the Swedish Research Council (621–2011–3674 and 2015–05149 to H.W., 621–2013–4386 to J.-Å.N., 621–2013–4357 and 2016–04391 to D.H., and 2010–5641 to C.K.C.), the Crafoord Foundation (20110600 to H.W.), the Royal Physiographic Society (Schyberg Foundation; 2011-04-13 to H.W.) and a Wallenberg Academy Fellowship to C.K.C. We are grateful to O. Hellgren, L. Råberg, B. Hansson, S. Bensch, J. Neto, M. Melo, U. Ottosson and A. Marzal for assistance with sampling. We thank M. Anisimova and A. Busin (Institute of Applied Simulation, Zurich University of Applied Science) who designed and implemented the pipeline for the positive selection and recombination analysis. We also thank I. Ekström for help with the graphics.

Author contributions

All authors contributed to the study design. All data collection and laboratory work was performed by E.A.O. Data analyses were conducted by E.A.O. and C.K.C. All authors contributed to interpreting the data and writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0509-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to E.A.O.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Emily O'Connor

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

In the case of the birds for MHC genotyping: blood samples from upto 3 individuals per species were used. This sample size was based upon the inter-individual variation we anticipated in MHC diversity from previous studies of MHC class I gene in other passerine birds. 32 species were sampled to test the effect of migratory strategy (African resident, Palearctic resident and migratory) on MHC diversity. This sample size was determined by sampling species representing each of the three migratory strategies within genus from taxonomic families that represented the full phylogenetic span of the Passerida parvorder.

2. Data exclusions

Describe any data exclusions.

No data was excluded from the analyses.

3. Replication

Describe whether the experimental findings were reliably reproduced.

For the MHC genotyping three individuals from each species were sampled to represent independent replicates within species (biological replicates). Each individual was sequenced twice (technical replicates) for each of the three primer pairs used (thus each individual was sequenced a total of six times).

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

This study does not describe an experiment where randomization would have been applicable. However, species from each migratory strategy (i.e. African resident, Palearctic resident and migratory) were evenly distributed across each of the three 454 sequencing runs in order to ensure any differences between the runs were balanced across the groups of interest.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Not applicable

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	R was used for all statistical analyses. The customized R scripts for all analyses have been provided as supplementary material.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	The blood and DNA samples for the 32 bird species used for MHC genotyping are held at Lund University and can be made available to other researcher upon reasonable request made to the corresponding author. The sequences for all MHC-I alleles identified in this study have been uploaded to GenBank (Accession codes: MF477947 - MF478976).

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Not applicable

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Not applicable
b. Describe the method of cell line authentication used.	Not applicable
c. Report whether the cell lines were tested for mycoplasma contamination.	Not applicable
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	Not applicable

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Non-lethal blood samples were collected from upto 3 individuals from the following bird species (individuals were caught in the wild and released directly after sampling):
Acrocephalus melanopogon
Anthus pratensis
Carduelis chloris
Parus caeruleus
Passer domesticus
Turdus merula
Aegithalos caudatus
Erithacus rubecula
Hirundo rupestris
Sylvia undata
Acrocephalus scirpaceus
Anthus trivialis
Motacilla flava
Oenanthe oenanthe
Phoenicurus phoenicurus
Sylvia borin
Hirundo rustica
Acrocephalus gracilirostris
Anthoscopus parvulus
Anthus leucophrys
Cercotrichas podobe
Myrmecocichla albifrons
Oenanthe monticola
Oenanthe pileata
Parus leucomelas
Passer griseus
Ploceus cucullatus
Serinus leucopygius
Turdus pelios
Zosterops senegalensis
Hirundo aethiopica
Cisticola juncidis

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Not applicable