nature ecology & evolution

Article

https://doi.org/10.1038/s41559-025-02699-3

Long-distance seasonal migration to the tropics promotes genetic diversity but not gene flow in boreal birds

Received: 23 October 2024

Accepted: 26 March 2025

Published online: 20 May 2025

Check for updates

T. M. Pegan[1,9], A. A. Kimmitt[1,10], B. W. Benz[1], B. C. Weeks[2], Y. Aubry[3], T. M. Burg[4], J. Hudon[5], A. W. Jones[6,11], J. J. Kirchman[7], K. C. Ruegg[8] & B. M. Winger[1]

Differences in life history can cause co-distributed species to evolve contrasting population genetic patterns, even as they occupy the same landscape. In high-latitude animals, evolutionary processes may be especially influenced by long-distance seasonal migration, a widespread adaptation to seasonality. Although migratory movements are intuitively linked to dispersal and therefore promotion of gene flow, their evolutionary genetic consequences remain poorly understood. Using ~1,700 genomes from 35 co-distributed boreal-breeding bird species that differ in non-breeding latitude and thus migration distance, we find that most long-distance migrants unexpectedly exhibit spatial genetic structure, despite their strong movement propensity. This result suggests evolutionary effects of philopatry—the tendency of many migrants to return to the same breeding site year after year, resulting in restricted dispersal. We further demonstrate that migration distance and genetic diversity are strongly positively correlated in our study species. This striking relationship suggests that the adaptive seasonal shifts in biogeography inherent to long-distance migration may enhance population stability, preserving genetic diversity in long-distance migrants relative to shorter-distance migrants that winter in harsher conditions at higher latitudes. Our results suggest that the major impact of long-distance seasonal migration on population genetic evolution occurs through promotion of demographic stability, rather than facilitation of dispersal.

Co-occurring species provide layers of information about how evolution has shaped the distribution and abundance of genetic diversity across their shared landscape. Spatial genetic patterns can be concordant among co-occurring species, reflecting the extrinsic influence of the shared landscape on evolutionary processes[1]. However, intrinsic biological differences between species can also modulate population demographics in ways that influence evolution[2]. For example, variation in life history can affect the amount of genetic diversity in populations[3], while differences in dispersal influence the ecological and genetic connectivity of metapopulations[2,4], the dynamics of local genetic adaptation[5] and the formation of new species[6]. Comparison of spatial genetic patterns among species on a common landscape therefore provides a rich source of information about how ecological and evolutionary processes are influenced by species' intrinsic traits[7–9]. The spatial evolutionary consequences of species' intrinsic traits may be especially apparent within continuously distributed populations without strong external barriers to gene flow[4]. Yet, owing to the often subtle nature of genomic differentiation across continuous space, comparative studies

A full list of affiliations appears at the end of the paper. e-mail: wingerb@umich.edu

https://doi.org/10.1038/s41559-025-02699-3

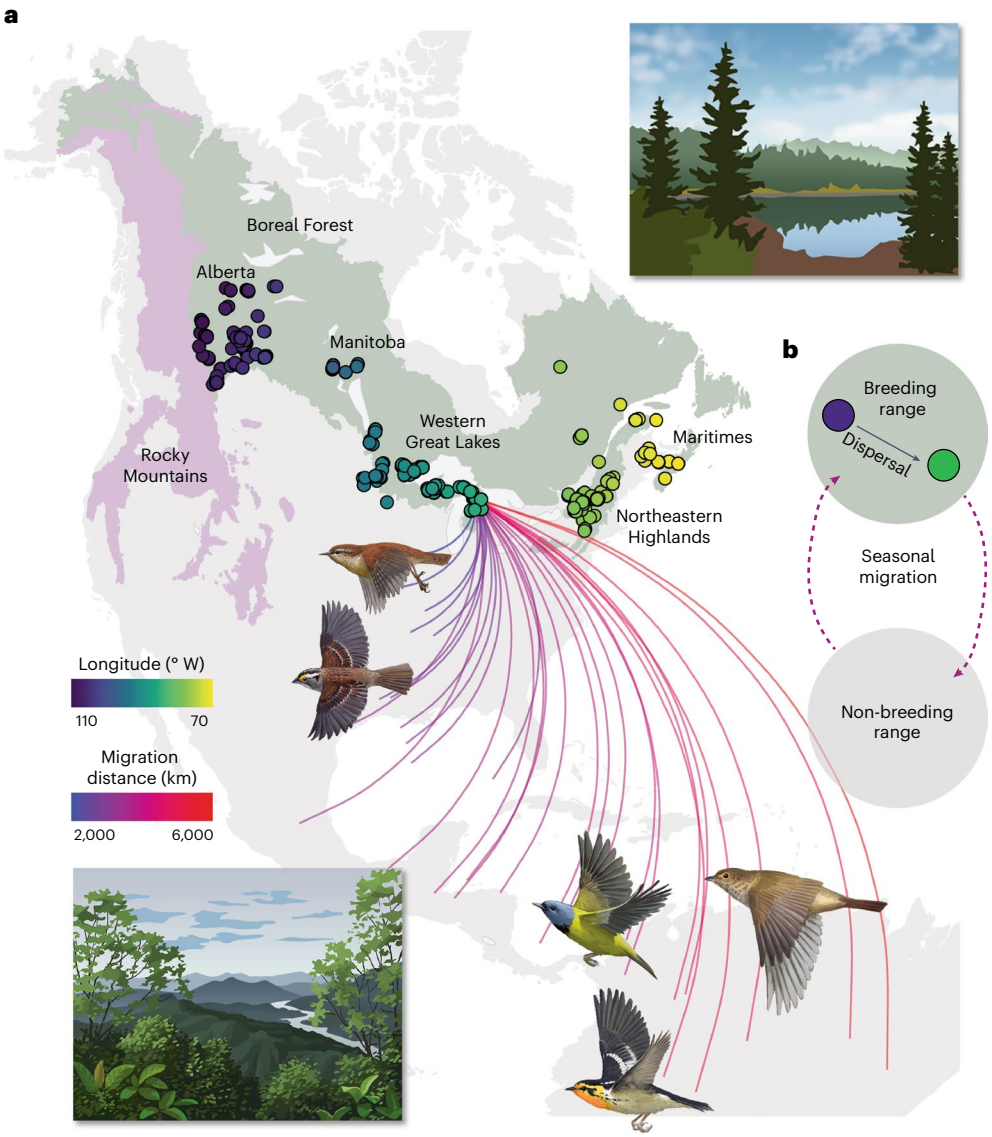


Fig. 1 | Ecogeographic context of the study. a, Study design. The North American boreal ecoregion is shown in green. Points, coloured by longitude, show sampling localities along the longitudinal axis of this ecoregion from Alberta, Canada east of the Rocky Mountains to the mainland Maritime provinces of Canada. The 35 species in this study are broadly co-distributed across this region and were sampled evenly across this study area. The nearby Rocky Mountains ecoregion (in pale purple) contains ecologically similar but often phylogeographically diverged populations. We did not sample the Rockies or the extremes of the boreal region in Alaska and Newfoundland, where phylogeographic breaks are known to occur, because of our focus on continuous spatial genetic variation and because not all species are found that far east or west. Curved lines depict species-level migration distances of the migratory species in the study, connecting an arbitrary point within their boreal breeding ranges to the centroid of each species' non-breeding distribution. **b**, Conceptual diagram of movement behaviours. Dispersal is a movement between breeding sites that can influence gene flow and is distinct from seasonal migration, a round-trip movement between breeding and non-breeding sites[41]. Artwork in **a** by John Megahan. Ecoregion shapefiles adapted with permission from ref. 70, Springer Nature Ltd.

with adequate genomic resolution to identify continuous spatial patterns from many co-distributed species have been scarce[10].

To test hypotheses about intrinsic drivers of spatial genetic evolution at a more powerful comparative scale than previously possible, we generated a dataset of 1,673 low-coverage whole genomes sampled from across the geographic ranges of 35 species of birds that co-occur throughout the North American boreal forest (15–67 genomes per species; Extended Data Table 1), sampling that required several decades of field work (Fig. 1). We then used a comparative phylogenetic framework to evaluate intrinsic factors that mediate differences in genetic diversity and patterns of spatial genetic differentiation across the continuous landscape shared by these species.

The ecogeographic setting of our study—the North American boreal forest (Fig. 1)—is particularly well-suited to test hypotheses about how species' attributes influence their spatial genetic patterns because it represents a vast region that lacks major geographic or environmental barriers, yet is large enough (>3,500 km in longitudinal breadth) for spatial genetic patterns to accrue, even in dispersive species[11]. Species tend not to show discrete population structure within the boreal ecoregion, whereas strong genetic differentiation between boreal populations and populations in neighbouring ecogeographic regions, such as the Rocky Mountains (Fig. 1), is more common (for example, refs. 12–15). Within our boreal forest study system, the expected spatial genetic consequence of dispersal limitation is a continuous pattern of isolation by distance, as opposed to deep differentiation driven by extrinsic barriers to gene flow. This expectation allows us to test whether traits that influence dispersal and population demographic processes affect the presence or strength of isolation by distance patterns across this large biome[6,16].

https://doi.org/10.1038/s41559-025-02699-3

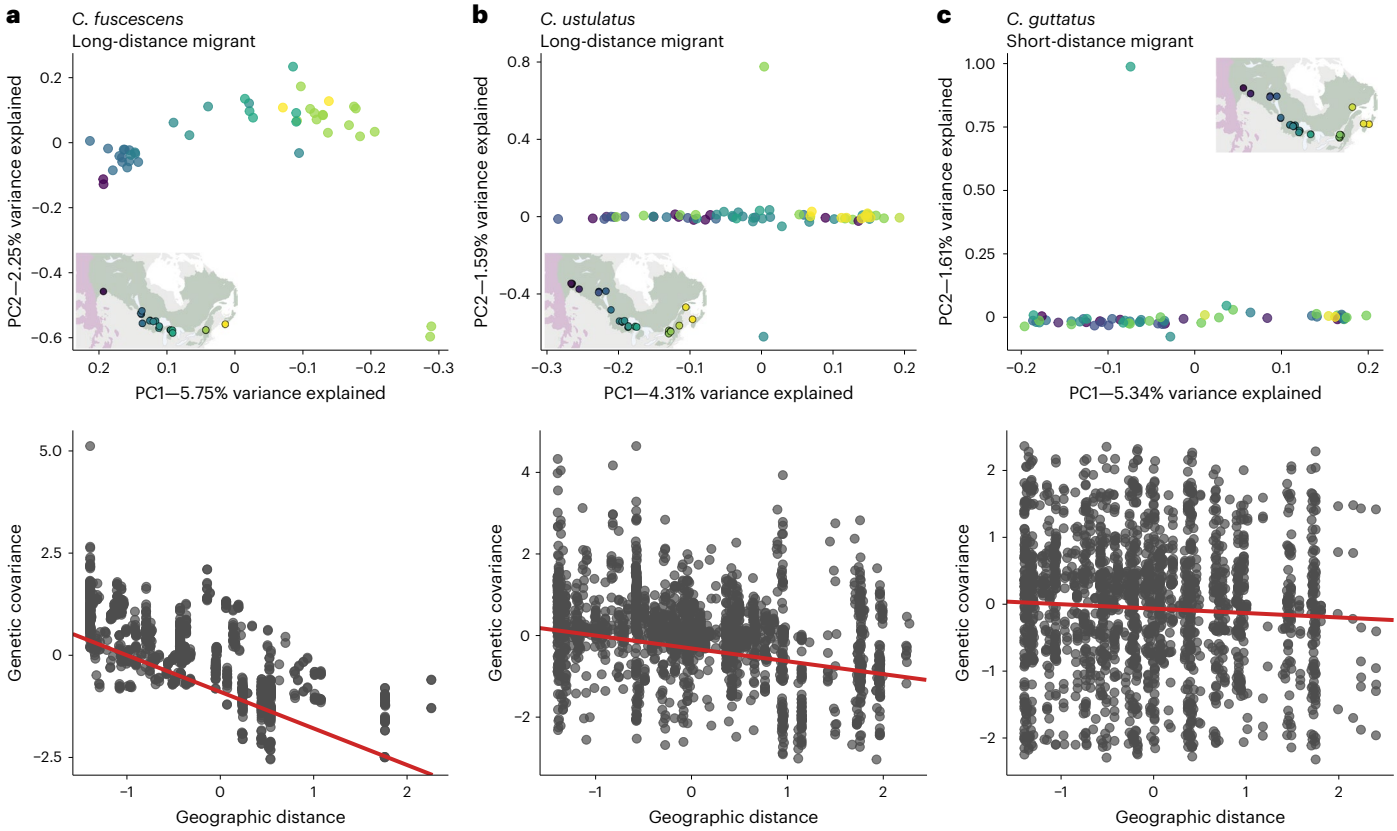


Fig. 2 | Interspecific variation in spatial genetic patterns. a–**c**, PCA and β_{IBD} plots from three example species in the genus *Catharus* demonstrate the variety of spatial genetic patterns present in boreal populations of the 35 study species (Supplementary Figs. 1–35). β_{IBD} is estimated as the slope of geographic distance and genetic covariance. The axis values in the scatterplots have been centred and standardized. The long-distance migrant *C. fuscescens* (**a**) shows a clear spatial genetic pattern of isolation by distance, whereas *C. ustulatus* (**b**) shows subtler isolation by distance and the shorter-distance migrant *C. guttatus* (**c**) shows no spatial structure. Point colours in PCA plots correspond to longitudinal locations of samples, as shown in the inset of each panel. The PC1 axis in **a** is reversed so the spatial variation it captures is shown from west to east, as in the sampling plot. Each PCA is accompanied by a scatterplot showing the relationship between genetic covariance and geographic distance, with a line indicating the slope and intercept of β_{IBD} for that species (Methods and Fig. 3).

An important life history trait with potential to shape population genetic evolution in the boreal avian assemblage is seasonal migration, because most bird species breeding at temperate latitudes migrate to lower latitudes after the breeding season to avoid harsh winter conditions[17] (Fig. 1). A species' migration distance covaries with aspects of biogeography and life history that may influence standing genetic diversity[3,17,18]. Furthermore, migratory behaviour may have uniquely strong influences on spatial evolution through its connection to organismal vagility and its putative influences on dispersal. It is intuitive to assume that long-distance migratory species have high dispersal potential, and there is evidence that migration promotes dispersal in migratory species compared with non-migratory species[19,20]. However, seasonal migration may also promote population divergence and even speciation by increasing behavioural or chronological isolation among populations (reviewed in refs. 21,22). Moreover, some studies have noted surprisingly low rates or distances of dispersal in migrants despite their high mobility[20,23], and migration appears to restrict geographic range expansion in some contexts[24,25]. Such paradoxical restrictions on dispersal in migrants are thought to be a result of high philopatry, or the tendency of migratory birds to return annually to the same breeding location[21,26,27] instead of dispersing away. Taken together, these observations suggest that migratory birds are characterized by traits that may both promote and suppress gene flow, but it is unclear how these conflicting effects of seasonal migration play out to influence spatial genetic differentiation over evolutionary time.

To shed new light on the evolutionary consequences of seasonal migration, we leverage the considerable variation among boreal-breeding bird species in their non-breeding ranges and thus their seasonal migration distances (Fig. 1). We first test the hypothesis that migratory boreal-breeding species have lower spatial genetic structure than the non-migratory species that are resident year-round in the boreal region, due to their higher mobility, which could lead to higher dispersal and gene flow. Furthermore, using only migratory species, we test the hypothesis that spatial structure scales negatively with migration distance such that the longest-migrating species experience the highest gene flow within their breeding range and therefore demonstrate the lowest spatial structure.

In addition to analysing the effects of seasonal migration on spatial genetic patterns, we evaluated two other species attributes that may influence spatial genetic differentiation: (1) body size, which is positively correlated with dispersal distance in birds[19]; and (2) species' associations with early successional boreal habitat (as opposed to mature boreal forest), because habitat ephemerality is associated with lower levels of population genetic structure[28]. We also included an estimate of population-wide genetic diversity (θ_π) as a third covariate to test whether genetic diversity predicts spatial genetic structure across species. A positive correlation may occur because low rates of dispersal between metapopulations can promote both genetic diversity and spatial genetic differentiation[29]. Moreover, in post-glacial landscapes, population expansion and invasion dynamics can affect both genetic diversity and the formation of spatial genetic structure in complex and interacting ways[30]. Such effects may be especially relevant in the boreal region, where breeding ranges were repeatedly shifted during glacial–interglacial cycles of the Pleistocene and current populations

have inhabited their contemporary spatial context for no more than 20 kyr, following the end of the Last Glacial Maximum[31].

Finally, the variation in migration distance among our study species coupled with the absence of strong geographical barriers provides a unique opportunity to test how migration, a key life history adaptation to seasonally fluctuating environments, influences standing genetic diversity in populations occupying the highly seasonal boreal latitudes. Genetic diversity is an important parameter that reflects the size and demographic dynamics of populations over their evolutionary history[32] and carries implications for future population persistence[33]. Prior work suggests that life history traits, including migratory behaviour, influence the evolutionary processes that shape genetic diversity[3,34,35], but the mechanisms by which life history parameters influence genetic diversity remain poorly understood[36,37]. The long-distance migrants in our study have been shown to have a slower life history than their co-distributed short-distance migrants[17], so our study design provides a detailed test of the relationship between life history and genetic diversity among co-distributed species sampled in a consistent manner from a common landscape.

Results and discussion

Migration and spatial genetic structure

Our analyses reveal that boreal bird species vary considerably in spatial population genetic patterns and genetic diversity despite their common landscape and broadly similar histories in the region[38] (Fig. 2, Supplementary Figs. 1–35 and Extended Data Table 1). To test predictors of spatial genetic structure, we used a multilevel modelling approach (Hierarchical Modelling of Species Communities; HMSC[39]) that estimates the relationship between predictor variables and the dependent variable within species while jointly estimating the effects of species' attributes on interspecific variation and accounting for phylogenetic relatedness among species. Specifically, we used HMSC to model the relationship between pairwise genetic covariance and pairwise geographic distance in all conspecific sample pairs, generating a slope of isolation by distance (β_{IBD}) for each species (Fig. 3), while simultaneously modelling the effects of all four species attributes—migration, mass, habitat association and genetic diversity—on the slope of isolation by distance (Fig. 3, γ coefficients). Under a model of isolation by distance, genetic covariance values are expected to be positive in geographically close pairs of samples and negative in geographically distant pairs that are more distantly related[16,40]. As such, β_{IBD} coefficients, which describe the slope of the IBD curve, are negative in species demonstrating isolation by distance, and species attributes with negative γ parameters are those associated with spatial partitioning of genetic structure.

We tested our two hypotheses about the effects of seasonal migration on spatial genetic structure using two models. The first model contained all 35 species in the study and included a binary γ variable indicating migratory status (migratory versus non-migratory). We used the second model to test whether effects of migration scale with migration distance (as a continuous variable); this model contained only migratory species ($N = 30$). We included the other three attributes (mean mass, genetic diversity and a binary variable indicating association with early successional habitat; Extended Data Fig. 1) as γ variables in both models.

We found that non-migratory species tend to display clearer spatial structure within the boreal region than migratory species, suggesting that migratory species experience higher rates of gene flow. Four out of the five non-migratory species in our study show clear spatial patterns in principal component analysis (PCA) plots, whereas many migratory species lack an evident geographic pattern (for example, Fig. 2c, and Supplementary Figs. 4–6, 25–30, 33 and 34). HMSC models also provide modest support for an effect of migratory status on spatial structure, where migrants are predicted to show weaker (less negative) β_{IBD} (mean posterior $\gamma = 0.29$, posterior support = 0.82; Fig. 3). Our conclusions about non-migratory behaviour are necessarily based on data from few non-migratory species, because most boreal bird species migrate. However, these conclusions are consistent with other studies that used mark–recapture data to demonstrate that non-migratory birds typically undergo reduced dispersal compared with migratory birds[19,20].

Although our results suggest that migratory birds experience more extensive gene flow than non-migratory birds, we found a contrasting effect of migration distance among the migratory species. In the model including only migratory species ($N = 30$), long-distance migration was not associated with increased gene flow. Rather, the γ coefficient for an effect of migration distance on β_{IBD} was negative ($\gamma = -0.081$, posterior support = 0.8; Fig. 3), suggesting somewhat higher spatial genetic differentiation in long-distance migrants relative to short-distance migrants. Although posterior support for this modelled relationship was modest, our results make it clear that some extremely long-distance migratory species in our study exhibit notable spatial structure or isolation by distance (for example, Fig. 2a, and Supplementary Figs. 23, 24, 32 and 35). These patterns may reflect limited gene flow across the boreal ecoregion, a remarkable pattern to discover in species that annually migrate distances roughly 1.5–2 times the entire breadth of the boreal breeding ranges where gene flow could occur.

Spatial genetic structure in long-distance migratory species may reflect a counterintuitive relationship between seasonal migration distance and philopatry[27,41]. While long-distance movement tends to be the most obvious and remarked-upon characteristic of seasonal migration, philopatry is equally fundamental—it is what distinguishes migration from other movement behaviours such as nomadism and dispersal[26]. When time in the breeding range is most limited, there is pressure to establish a breeding attempt quickly[42] and there are clear advantages of returning to a familiar area, rather than exploring elsewhere[43,44]. This effect may be strongest in long-distance migrants, which spend the least time in their breeding regions[17]. Spatial structure in long-distance migratory species may therefore reflect an evolutionary consequence of the remarkable navigational precision that long-distance migrants employ to travel around the globe only to return to the exact same breeding sites year after year (for example, refs. 45,46).

Other predictors of spatial structure

Aside from migration, the evidence for effects of other species attributes on population genetic patterns in the boreal avian assemblage was less clear. In our HMSC models, posterior support values for effects of genetic diversity on β_{IBD} were equal to or less than 0.7 in all cases, which means that our models do not support a correlation between genetic diversity and spatial genetic structure. We found similarly low support for hypothesized effects of body mass on spatial structure. We predicted that species associated with habitat in early stages of succession would show reduced spatial genetic structure. Early successional habitats are spatially ephemeral compared with mature habitats and studies in tropical bird species suggest that habitat ephemerality promotes gene flow and reduces signatures of spatial genetic structure[28]. However, our HMSC model using only migratory species demonstrated the opposite pattern, wherein early successional habitat was associated with stronger β_{IBD} (that is, stronger spatial structure; $\gamma = -0.11$, posterior support = 0.81; Fig. 3). Posterior support for this effect was substantially lower in the model with all species ($\gamma = -0.033$, posterior support = 0.61; Fig. 3). Our results demonstrate that the positive effects of habitat ephemerality on gene flow observed at tropical latitudes[28] are not present in the boreal avifauna. Effects of habitat on spatial structure may vary across latitudes and ecological contexts because these effects are probably driven by the dispersal behaviours and movement capacities of the birds in these habitats as much as by the characteristics of the habitats themselves[47,48], and even the non-migratory species in our study probably have higher dispersal ability than many tropical resident birds[49]. Additionally, the subtle differences between habitat types in our study system—in which many study species inhabit a variety of

https://doi.org/10.1038/s41559-025-02699-3

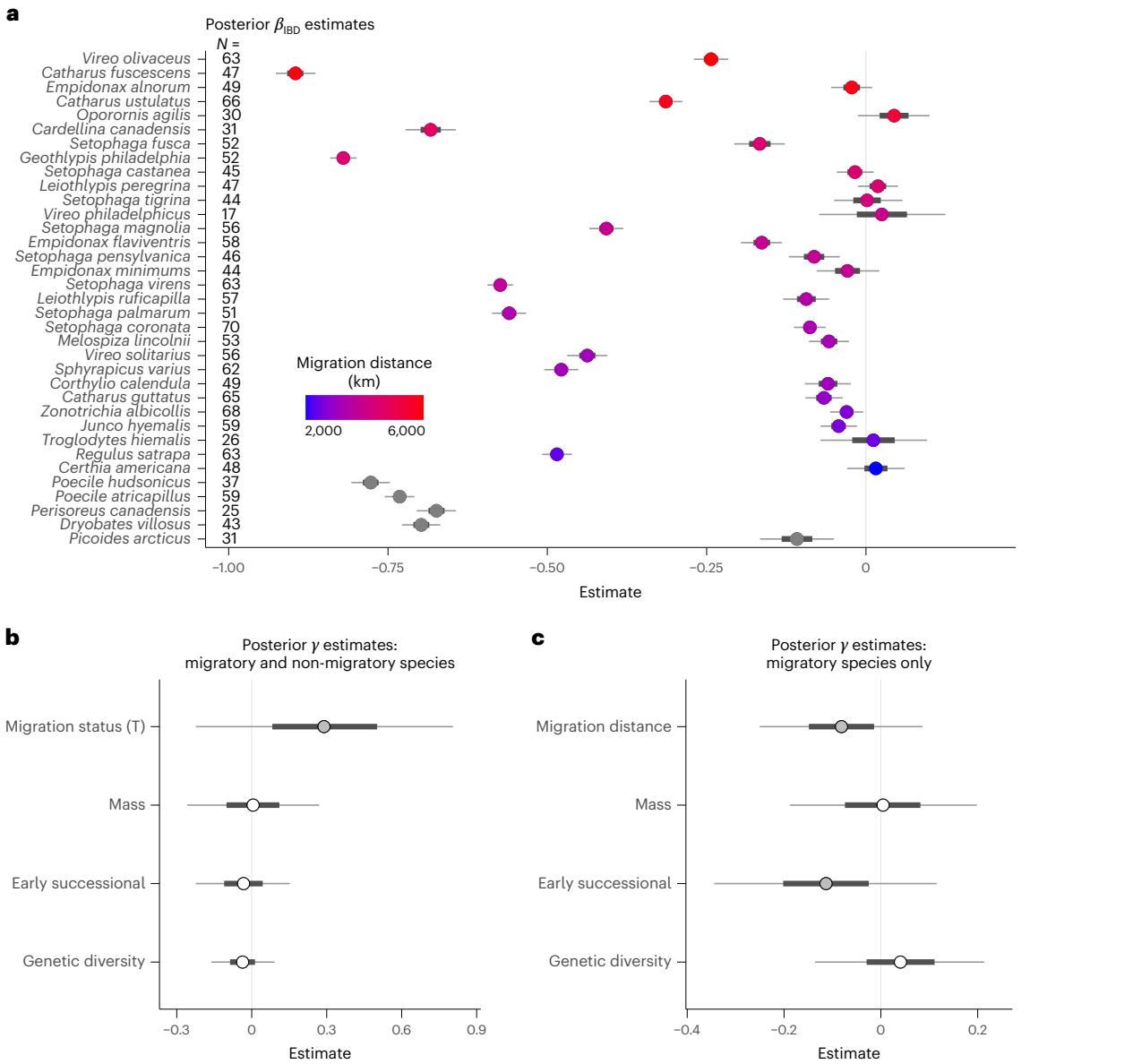


Fig. 3 | β_{IBD} and effects of species attributes on β_{IBD}. a–c, Posterior estimates of β_{IBD} (**a**) and γ (**b**,**c**) and their credible intervals from HMSC models. Points represent the mean posterior estimates, thick lines represent 50% credible intervals and thin lines represent 90% credible intervals. Species-specific slopes of β_{IBD} (**a**) are generated from the HMSC hierarchical mixed models. The sample size from each species used to create these estimates is shown next to the *y* axis. We created one model where migration was coded as a binary trait, and another including only the 30 migratory species with migration distance as a continuous trait. β_{IBD} estimates were highly similar across models, so we show results only from the full 35-species model, with points coloured by migratory status and migration distance (non-migratory = grey). Species are arranged on the *y* axis in order of migration distance. Most species show negative β_{IBD} coefficients, reflecting the negative relationship between genetic covariance and geographic distance expected under a model of isolation by distance. Other species show β_{IBD} coefficients close to 0, reflecting a weak or absent relationship between the spatial distribution of samples and their genetic covariance. Posterior γ estimates from HMSC models (**b**,**c**) reflect the influence each species characteristic (migration status or distance, mass, early successional habitat association and genetic diversity) on β_{IBD} across the 35 species shown in **a**. In **b** where migration status is treated as a binary variable, we show the effect of being migratory (status = True (T)) compared with being non-migratory. In each case, a negative value of γ indicates that a trait promotes spatial structure (that is, results in a more negative β_{IBD}), while a positive value of γ indicates that a trait reduces spatial structure. The central point is coloured by the position of 0 relative to the estimate's credible intervals, where 0 is within the 50% interval for white points and outside it for grey points.

boreal forest types—may be of less importance than in tropical systems wherein species have finer habitat specificity.

Overall, our HMSC models explained only a small proportion of interspecific variation in β_{IBD} (proportion of variance explained by effects of γ coefficients was 0.010 in the model with all species, and 0.03 in the model with migratory species only), but the models also identified strong phylogenetic signal in residual (unexplained) variation. The mean (± s.e.m.) posterior ρ coefficients were 0.86 ± 0.09 and 0.86 ± 0.10 for the 35-species and 30-species models, respectively (where ρ can range from 0, indicating no phylogenetic signal, to 1, indicating very strong phylogenetic signal[50]). In applications of HMSC to species distribution data, phylogenetic signal in residuals is interpreted to reflect unidentified traits that contribute to phylogenetic niche conservation[50], but we are not aware of previous examples where residual phylogenetic signal has been identified in spatial genetic patterns like isolation by distance. Our results thus raise new questions about how such signal could arise. Whereas inherited phenotypic traits are passed directly from ancestors to descendants, creating a

https://doi.org/10.1038/s41559-025-02699-3

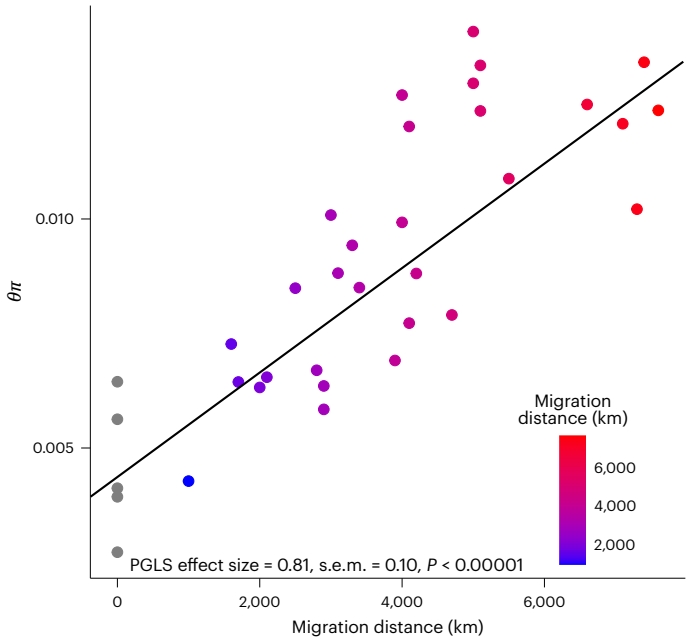


Fig. 4 | θ_π correlates with migration distance. Each point represents one species and points are coloured by migration distance, as in Figs. 1 and 3, where non-migrants are shown in grey. The intercept and slope (effect size) of the line were estimated with a PGLS model (two-sided, no multiple comparison). Exact *P* value = 1.1×10^{-9}.

strong expectation of phylogenetic signal in many phenotypes[51], spatial genetic patterns arise emergently through a fundamentally different process of interaction between individuals and the landscape. That boreal species demonstrate phylogenetic signal in their residual spatial genetic variation suggests that unidentified, phylogenetically conserved traits may have a strong influence on how these patterns form[50].

A broad conclusion of our analyses is therefore that much of the interspecific variation in spatial structure among boreal bird species may be driven by species' intrinsic attributes, rather than extrinsic processes that are random with respect to phylogeny. Although coarse-scale attributes such as general locomotory capacity predict spatial genetic patterns at broad phylogenetic scales[52], our results show that their explanatory power is lost within assemblages of closely related species, highlighting the difficulty of identifying attributes that play the most important roles in spatial evolution without more refined ecological and natural history information. Specifically, among our study species, we posit that much of the unexplained variation in spatial structure may be related to other ecological and behavioural factors affecting dispersal behaviour that are difficult to measure and thus could not be confidently included in the study as covariates. For example, species may vary in strength of territoriality and tendency for sex-biased dispersal, with consequences for the population dispersal kernel and therefore for spatial structure[53–55]. Boreal bird species may also vary in the extent to which they follow resource variation across their breeding grounds when choosing nesting sites instead of exhibiting philopatry. Indeed, it is noteworthy that the three species of warblers in our study whose population ecology is closely tied to spruce budworm outbreaks (*Leiothlypis peregrina*, *Setophaga castanea* and *Setophaga tigrina*[56]) show no discernible geographic structure in PCA (Supplementary Figs. 25, 28 and 34). Similarly, the non-migratory species with the weakest evidence of geographic structure is *Picoides arcticus* (Supplementary Fig. 3), a woodpecker species known to follow outbreaks of bark beetles after forest fires[57]. These results raise the possibility that species that follow prey with boom–bust population cycles may undergo longer-distance dispersal in response to food availability, resulting in high gene flow and weak spatial genetic structure.

Migration and genetic diversity

Our results also revealed an unexpectedly strong positive correlation between genetic diversity and migration distance (Fig. 4 and Extended Data Fig. 2). We found that non-migratory species tend to show the least standing genetic diversity in the boreal avian assemblage, implying that they have the smallest effective population sizes (N_e; ref. 36), and that genetic diversity (and N_e) increases with migration distance (Fig. 4; phylogenetic generalized least squares (PGLS) effect size = 0.81, s.e.m. = 0.10, $P < 0.00001$, $\lambda = 0.18$). Similarly, the mean heterozygosity of sampled individuals from each species correlates positively with migration distance (PGLS effect size = 0.77, s.e.m. = 0.10, $P < 0.00001$, $\lambda = 0.19$). The adjusted R^2 values for these PGLS models, which reflect variance explained given the phylogenetic tree and model of evolution applied by the PGLS model[58], are 0.67 and 0.62, respectively.

To ensure that these results were not driven by biases associated with low-coverage inference or reference mapping[59–61], we generated higher-coverage genomes from one representative from each of 31 species in our study (mean mapped coverage 14.4×, range 6.5–22.5×; Extended Data Table 2). We estimated heterozygosity in these genomes using mapped reads, as we did with our population genomic data, and with an additional method based on *k*-mer counting of unmapped reads[62], which allowed us to estimate heterozygosity without the potentially biasing influence of the reference genome. Remarkably, migration distance is a strong predictor of heterozygosity even of single individuals sampled from boreal populations and sequenced at higher coverage (Extended Data Fig. 2; from mapped reads, PGLS heterozygosity effect size = 0.78, s.e.m. = 0.15, $P = 0.00002$, $\lambda = 0.16$; from *k*-mer counting, PGLS heterozygosity effect size = 0.75, s.e.m. = 0.18, $P = 0.0003$, $\lambda = 0.28$). Heterozygosity from single individuals can sometimes deviate from population-wide levels of genetic diversity[63], but our single-individual high-coverage heterozygosity estimates are strongly positively correlated with population-wide genetic diversity (Pearson correlation coefficients 0.92–0.94, $P < 0.0001$). Individually, each estimate of heterozygosity and genetic diversity from low-coverage population data and from single high-coverage individuals is subject to important caveats (Extended Data Table 3). However, the strong congruence we find across all four of these estimates suggests that they reflect true interspecific variation in genetic diversity within individuals and populations, which is positively correlated with migration distance. These strongly congruent results from the nuclear genome contrast with patterns in the separate mitochondrial genome, where the relationship between migration distance and genetic diversity is much less clear[35].

Why do long-distance boreal migrants have higher nuclear genetic diversity? Long-distance migrants in our system have slower life history (higher annual survival, lower annual fecundity) than short-distance migrants[17]. Thus, the genetic diversity pattern we document (Fig. 4) contrasts with results from comparisons across animal taxa at much broader phylogenetic scales, which show that species with slow life history tend to exhibit low genetic diversity[3]. However, the mechanisms underlying a relationship between life history and genetic diversity are complex and poorly understood[36,37]. Genetic diversity is also thought to be influenced by effects of the environment on demography, including population bottlenecks[32] and unstable metapopulation dynamics[64]. Given the overlap in the breeding ranges of our study species, we propose that interspecific variation in genetic diversity in boreal birds is strongly influenced by demographic consequences of the non-breeding environment. Although recent work suggests that many boreal bird species did not experience recent severe bottlenecks at the Last Glacial Maximum[38], genetic diversity can also be eroded by frequent extinction–recolonization events in metapopulations, which create successive local bottlenecks and reduce genetic variation[65]. Boreal non-migratory species and short-distance migrants are exposed to challenging temperate-zone winter conditions during t he non-breeding season, which are known to limit population sizes[66,67].

By contrast, long-distance migrants evade temperate winter by migrating to warm and resource-rich tropical regions (Fig. 1), where they experience higher apparent annual adult survival than short-distance boreal-breeding migrants that winter at high latitudes[17]. Therefore, greater population stability on tropical compared with temperate non-breeding grounds may buffer long-distance migrants from the frequent loss of genetic diversity. The relationship between non-breeding latitude and genetic diversity among sympatrically breeding boreal bird species parallels a global relationship between latitude, genetic diversity, life history and population demographic stability that is present in many taxa[18], highlighting the importance of non-breeding ranges for migratory bird population demographics across ecological and evolutionary timescales. Moreover, our results suggest that in assemblages of migratory birds, interspecific variation in genetic diversity is strongly influenced by the demographic effects of seasonal biogeography.

Conclusions

Our results carry important implications for how interspecific variation in seasonal migration—a widespread adaptation to seasonal environments—interacts with evolutionary processes. We show that even long-distance migrants can exhibit sufficiently high philopatry to result in spatial segregation of breeding populations across a continuous landscape. Moreover, our comparative framework reveals a previously unknown positive correlation between migration distance and genetic diversity, suggesting that migration to tropical forests each year may have allowed long-distance migrants to experience greater long-term population stability than species whose annual cycles take place entirely within the temperate zone. However, many long-distance migrating species face contemporary instability in habitat availability and population size, with many species in steep decline[68]. The relatively high genetic diversity displayed by long-distance migratory species could potentially help their populations adapt to new conditions[33,69], but they will also have to persist through population declines on a scale that may have little precedent in their recent evolutionary histories.

Methods

Species and sampling

Our study system includes 35 co-distributed boreal forest bird species (Extended Data Table 1 and Extended Data Fig. 1). These species represent a core subset of the small-bodied species (<75 g) that breed widely across the boreal forest and the temperate–boreal transition (hemiboreal) habitat of our sampling area (corresponding to the 'Northern Forests' level I ecoregion[70]). Although all 35 species we sampled are found sympatrically during the breeding season throughout much of the boreal ecoregion, they exhibit variation in microhabitat preference and the extent of their geographic range beyond the boreal forest ecoregion[12,70–72]. Three species are woodpeckers (Piciformes), and the remaining are from 17 genera and 10 families of songbirds (Passeriformes). We focus on obligate migratory species and non-migratory species, meaning we did not include species primarily characterized by nomadic or irruptive movements (for example, Fringillidae, Sittidae and Bombycillidae) in our analyses. Our study species vary greatly in migratory strategy and distance but share many life history attributes (for example, mating system and age at first breeding season[17]).

We sampled species broadly and evenly along the longitudinal axis of this ecoregion (Fig. 1), from Alberta, Canada east of the Rocky Mountains in the west to Québec and the mainland Maritime provinces of Canada (New Brunswick and Nova Scotia) in the east, as well as intervening locations in central Canada (Manitoba) and the northern United States (northern Michigan, Minnesota and New York). All samples were collected during the breeding season. We sequenced DNA from 1,673 samples (mean ± s.d. = 47 ± 13 samples per species, range = 15–67 samples; Extended Data Table 1 and Supplementary Dataset 1). To facilitate even sampling, we strove to sample 15 individuals from each of the broad regions mentioned above, though sampling was lower for some populations (Supplementary Figs. 1–35). Most (88.3%) of the samples were ethanol-preserved or flash-frozen specimen-vouchered tissues provided by contributing natural history collections, while the remaining 11.7% came from unvouchered blood samples (Supplementary Dataset 1). Field sampling was approved by the University of Michigan Institutional Animal Care and Use Committee and all relevant permitting authorities (see Acknowledgements), and performed in accordance with the Ornithological Council's Guidelines for Use of Wild Birds in Research[73].

Our sampling was intentionally designed to capture continuous spatial variation across the eastern boreal region shared broadly by the species in our study. We did not sample populations in regions at the extreme edges of the boreal region (Alaska, Newfoundland) or in other ecoregions (such as the Rocky Mountains) because not all species' ranges extend into these regions, and many species distributed across these ecoregions have experienced phylogeographic breaks between them, even as the core boreal region typically comprises a single population[13–15,74–79]. Based on this past work, we expect spatial differentiation within our core boreal sampling region to be continuous rather than discrete. We therefore focus on quantifying continuous isolation by distance (IBD) rather than attempting to identify discrete subpopulations, especially because inference of discrete structure can be confounded by the presence of IBD in continuously distributed populations with geographic gaps in sampling[4,80].

Preparation of DNA sequence data

We extracted DNA using DNeasy Blood and Tissue Kits (Qiagen Sciences) or phenol–chloroform extraction. Libraries were prepared using a modified Illumina Nextera library preparation protocol[81,82] and then sequenced on an Illumina HiSeq, NovaSeq 6000 or NovaSeq X platform using paired-end sequencing of 150 bp reads. Library prep involves primers with sequences AATGATACGGCGACCACCGA and CAAGCAGAAGACGGCATACGA. Libraries used in analyses produced an average (± s.d.) of 32 million ± 10 million reads per sample (range 7 million–156 million).

Using a bioinformatic pipeline we also describe in ref. 27, we trimmed remaining adaptors and low-quality bases from demultiplexed data with AdapterRemoval v.2.3.1 using the --trimns and --trimqualities options[83]. To mitigate potential batch effects associated with differences between the NovaSeq and HiSeq platforms, we used fastp v.0.23.2[84] with the --cut_right option to remove low-quality read ends[85]. Following this filter, the mean (± s.d.) number of bases read per individual at a base quality of at least 30 was 4.3 billion ± 1.3 billion bases (range: 0.94 billion–21.2 billion).

We aligned samples to a reference genome from a related species[86–93] (Supplementary Dataset 2) using bwa mem v.0.7.15[94] and then sorted them using SAMtools v.1.13[95]. We selected reference genomes for each species representing the closest possible relative of that species with a chromosome-assembled genome on GenBank at time of analysis. We used usingclipOverlap in bamUtil v.1.0.14[96] to remove overlapping reads, and we used picard-tools v.2.8.1 (http://broadinstitute.github.io/picard/) to mark duplicate reads with the MarkDuplicates tool and to assign all reads to a new read-group with the AddOrReplaceReadGroups tool. We indexed bam alignment files with SAMtools. Samples had an average duplication rate of 7% (s.d. 4%, range 0.3–33%). Excluding duplicate reads, the mean mapping rate was 88% (s.d. 7%, range 38–98%). Finally, we re-aligned samples around indels using GATK v.3.7[97]. We applied the GATK RealignerTargetCreator tool to the entire dataset analysed for each species and applied the GATK IndelRealigner tool to each sample. We calculated the average genome-wide coverage for each alignment by dividing the total coverage of the bam file (calculated with the SAMtools depth function, filtered to exclude reads mapped with a quality lower than 30) by the total length of the reference genome. Bases with a coverage of 0 (including segments

of the reference genome where no reads were mapped) are included in these calculations, making them conservative. Across all aligned samples, average genome-wide coverage of high-quality mapped reads was 2.8× (s.d. 1×).

Data quality control using mitochondrial sequences

We confirmed species identity by examining at least one mitochondrial gene per successful assembly via BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) in Geneious v.2021.2.2. Mitochondrial genomes were assembled using NOVOplasty v.4.3.1[98] and analysed as described in refs. 35,38. We also visually inspected alignments of mitochondrial genes to check for evidence of cross-contamination and hybridization. Based on these assessments, we excluded 14 samples with an unexpected mitochondrial species identity due to specimen misidentification or hybridization, as well as 5 samples with evidence of mitochondrial chimerism, which is probably due to sample cross-contamination. None of these excluded samples are counted in sample totals reported elsewhere or included in downstream analyses.

Data filtering

Continuing the bioinformatic pipeline also described in ref. 27, we used ANGSD v.0.941[99] to calculate genotype likelihoods for all sites inferred by ANGSD to be single nucleotide polymorphisms (SNPs) at alpha level 0.05. We used minimum base and map quality thresholds of 30. We applied ngsParalog v.1.3.2[100] to filter out SNPs with a high likelihood of occurring within a mis-mapped or paralogous region. Next, we used PCAngsd v.1.10[101] to create a PCA separately for each chromosome. We observed many chromosomes with evidence of PCA clustering consistent with the presence of inversion polymorphisms[102,103], so we analysed each chromosome further using lostruct v.0.0.4[104] as implemented using PCAngsd with scripts available from https://github.com/alxsimon/local_pcangsd (v.0.0.1). Inversion polymorphisms and polymorphisms in large linked haplotypes associated with other forms of suppressed recombination can obscure spatial genetic patterns[40,105], so we removed all such regions from each species' dataset (Supplementary Dataset 3). We flagged regions as polymorphisms when they met two criteria: (1) lostruct identified the region as showing distinct local structure; and (2) PCA with data from the region showed distinct clustering into two or three groups. None of the regions we flagged with these criteria showed strong spatial structure. We removed entire microchromosomes (<35 Mb in length[106]) with flagged regions on them. To remove flagged regions from macrochromosomes without discarding the entire chromosome, we used lostruct plots to identify the affected regions and we removed the region with a buffer of at least 1 Mb on either side. When macrochromosomes showed evidence of more than one flagged region and/or if the affected region was not clearly identifiable using lostruct, we discarded the entire chromosome (Supplementary Dataset 3). Finally, we removed all reads mapped to sex chromosomes or unassembled scaffolds.

We also used chromosome-level PCA plots to further filter individuals. We removed individuals if they presented as a strong PCA outlier on multiple macrochromosomes. In all three of the sparrow species and four warbler species (*Leiothlypis peregrina*, *Leiothlypis ruficapilla*, *Setophaga castanea*, *Setophaga magnolia*), we detected sex-based clustering on autosomes as well as sex chromosomes. We filtered females out of these species because we have many more samples from males than females. Individuals filtered out at this stage are not included in the sample numbers we report here.

Estimation of spatial genetic structure within each species

After examining the data and applying the quality filters described above, we then used PCAngsd v.1.10[101] with the –admix option to create full-genome PCAs and admixture plots for each species. We plotted the results of these analyses in R v.4.2.0. PCAngsd also produces a matrix of covariances between each pair of individuals for each species, which we used in downstream modelling of isolation by distance as in ref. 27. We estimated isolation by distance using individual-based genetic covariance instead of more typical F_{ST} estimates because the latter require a priori definition of populations, which is not feasible within our continuous sampling area. For each pair of samples, we calculated geographic distance using the function distGeo from the R package geosphere v.1.5-18[107] on latitude/longitude coordinates.

Estimation of species attributes hypothesized to affect spatial structure

We created a species-level estimate for each of the attributes we hypothesized to influence spatial structure (seasonal migration, mass, habitat association and genetic diversity; Extended Data Table 1). We represented variation in migratory behaviour both as a binary variable (migratory versus non-migratory) and as a continuous estimate of migration distance, as represented by the distance between the centroids of species' breeding and wintering ranges[17,108]. Mean mass estimates for each species are from refs. 109,110. We used a binary variable to indicate whether species show association with early successional habitat based on our field experience with these species and information from ref. 110.

Estimation of genetic diversity using subsampled loci

To estimate genetic diversity, we created a subsampled set of genome loci for each species using scripts modified from https://github.com/markravinet/genome_sampler. Each subset was created by sampling random 2 kb loci at least 10 kb apart, which produces loci comprising roughly 10% of the genome (Supplementary Dataset 2). Regions removed by the inversion filters described above were excluded prior to random locus selection. After locus selection, we removed SNPs from the selected loci that were flagged by ngsParalog as described above. Individuals removed from PCA-based filters described above were excluded. The subsampled loci were stored in a Browser Extensible Data file and supplied as a site filter to ANGSD in the next steps. We estimated genetic diversity in two ways: first by estimating θ_π (specifically, $\hat{\theta}_\pi$ of ref. 99) on the entire dataset for each species, and second by estimating heterozygosity in each individual and taking the species-wide mean. To estimate θ_π, we created a species-wide Site Allele Frequency (SAF) file in ANGSD (with minimum base and map quality thresholds of 30) and used the program winsfs v.0.7.0[111] to generate and fold a species-wide one-dimensional Site Frequency Spectrum (SFS). We then used ANGSD's saf2theta and thetaStat functions on the SAF and SFS to generate estimates of θ_π. To calculate genome-wide θ_π, we divided the sum of θ_π across all loci by the sum of the number of sites analysed. To estimate individual heterozygosity, we used ANGSD to generate a separate SAF file per individual, then used winsfs to generate a separate one-dimensional SFS per individual. Individual heterozygosity is the second element in this SFS divided by the sum of the entire SFS[112]. We found that θ_π and mean heterozygosity were strongly correlated with each other and with additional estimates of heterozygosity that we generated from representative high-coverage genomes (see below; Pearson correlation coefficients 0.92–0.99, $P < 0.0001$ for all comparisons). Therefore, we used the θ_π estimate to represent genetic diversity in downstream hierarchical models of genetic structure described below.

Hierarchical modelling of the effects of traits on spatial structure

We used the R package HMSC v.3.0-13[39] to jointly estimate spatial structure for each species and the effects of species' traits, or attributes, on these patterns at the community level. This approach has advantages over analogous two-step approaches (those that model slopes generated from one model as the outcome variable in a second model) because hierarchical modelling propagates error across the levels of the model, making it ideal for use in this context (for example, ref. 113).

We model spatial structure as a linear relationship between pairwise genetic covariance (calculated in PCAngsd) and pairwise geographic distance (of sampling locations) for all conspecific pairs of samples in our dataset. These relationships are represented as β coefficients (which we refer to as β_{IBD}) for each species in the HMSC model. The effect of each attribute on interspecific variation in β_{IBD} is represented by the attribute's γ coefficient in the HMSC model. HMSC can use a phylogenetic covariance matrix as a random effect in models to account for phylogenetic relatedness[50], so we provided the HMSC model with a phylogenetic tree of our study species prepared using data from https://birdtree.org (ref. 114) following ref. 115 as described in refs. 35,108 (Extended Data Fig. 1). We centred and scaled all traits prior to modelling to have a mean of 0 and a standard deviation of 1. We verified that all HMSC models converged with an effective sample size > 300 (ref. 50).

To test whether migratory status (migratory versus non-migratory) affects spatial structure differently than variation in migration distance, we created two HMSC models. The first includes all 35 study species and includes migratory status as a binary trait in addition to the 5 other traits we investigated. The second includes only the 30 migratory species in our study and represents migration behaviour using our continuous estimate of migration distance.

We ran HMSC models with the options 'YScale = T' and 'distr = "normal"' and all other options at the default. We used the HMSC's default priors, which are designed to be generally applicable and are described in detail in ref. 50. We ran the models in four replicate Markov chain Monte Carlo (MCMC) chains. After a 10,000 iteration burn-in, we let the chains run for 75,000 iterations and thinned the output by 50, producing 1,500 samples per chain from the posterior distribution for each parameter[113]. We examined trace plots and effective sample sizes to ensure model convergence and adequate independence of samples.

We examined and visualized the posterior estimates and posterior support values for each model parameter of interest using functions from the R packages HMSC, MCMCvis v.0.16.3[116] and bayesplot v.1.10.0[117], which create summaries of model output across all four chains. We estimated trait R^2 values with the HMSC function computeVariancePartitioning(). We performed variance partitioning on 250 posterior samples from our models, as opposed to all 1,500 used in other analyses, for computational feasibility.

Assessing the relationship between genetic diversity and migration distance

We used the function pgls() from the R package caper v.1.0.3[58] to assess the relationship between our estimates of genetic diversity and migration distance in a phylogenetic framework. We modelled the effect of migration distance on θ_π and mean heterozygosity in separate models. We scaled the variables to have a mean of 0 and a standard deviation of 1 prior to modelling and we supplied the function with the same phylogenetic tree we used in the HMSC models, described above. We used the option 'lambda = "ML"', which causes the model to estimate a maximum likelihood estimate of λ (phylogenetic signal in model residuals).

Additional analyses of heterozygosity with higher-coverage genomes

To evaluate whether interspecific variation in heterozygosity was accurately captured by our low-coverage analyses, we re-sequenced one individual per species from all migratory species and one non-migrant (N = 31; Extended Data Table 2) at high coverage using leftover sequencing libraries after low-coverage data were generated. We selected one individual per species from the central portion of the boreal region (Michigan or Minnesota, USA; or Manitoba, Canada when no samples were available from the central USA). Within this region, we chose the individual male sample for each species with the highest weight of sequence-ready DNA library leftover after low-coverage sequencing. We re-pooled the libraries and sequenced them on a NovaSeq X machine as described for low-coverage samples above, except that we adjusted the number of samples included in the pool to obtain higher depth per sample.

We estimated the heterozygosity of these higher-coverage genomes in two ways. First, we processed the sequences as described above (Preparation of DNA sequence data), using the previously generated population-wide 'realigner target' file to realign around indels with GATK. Across all aligned high-coverage samples, average (± s.d.) genome-wide coverage of high-quality mapped reads was 14.4× (±3.6×). We then estimated the individual heterozygosity with ANGSD for each high-coverage sample, as described above (Estimation of genetic diversity using subsampled loci). Second, we estimated the heterozygosity of each genome directly from unmapped read files using GenomeScope v.1.0[62] to analyse k-mer distributions we created with Jellyfish v.2.2.3[118], with k-mer length set to 21. Jellyfish identifies all unique 21-base-long sequences (that is, k-mers where k = 21) in the genome and calculates the distribution of their frequencies, while GenomeScope fits models to this distribution to evaluate which sets of model parameters (including heterozygosity) result in the best model fit. To generate the k-mer distributions, we used fastq files that were trimmed and processed with fastp as described above. We uploaded each k-mer distribution to GenomeScope's web interface (http://genomescope.org/) with read length set to 150, k-mer length set to 21 and maximum k-mer coverage set to 1,000. GenomeScope models failed to converge for four genomes that had among the lowest sequencing depth in our higher-coverage dataset (Extended Data Table 2), so analyses with GenomeScope heterozygosity estimates use 27 species instead of 31.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequence data generated during the study have been uploaded to the NCBI Sequence Read Archive under BioProjects PRJNA1043688 and PRJNA1130443. Individual accession numbers for each sample are provided in Supplementary Dataset 1. All other data are available within the manuscript, the repositories described in the Code availability statement and the Supplementary Information files. Source data are provided with this paper.

Code availability

Bioinformatic code used to process sequence data are available via figshare at https://doi.org/10.6084/m9.figshare.27284553 (ref. 119). Code used to conduct comparative analyses are available via a Code Ocean capsule at https://doi.org/10.24433/CO.5578409.v1 (ref. 120).

References

1. Avise, J. C. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* **18**, 489–522 (1987).
2. Papadopoulou, A. & Knowles, L. L. Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proc. Natl. Acad. Sci. USA* **113**, 8018–8024 (2016).
3. Romiguier, J. et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261–263 (2014).
4. Bradburd, G. S. & Ralph, P. L. Spatial population genetics: it's about time. *Annu. Rev. Ecol. Evol. Syst.* **50**, 427–449 (2019).
5. Bachmann, J. C., Jansen van Rensburg, A., Cortazar-Chinarro, M., Laurila, A. & van Buskirk, J. Gene flow limits adaptation along steep environmental gradients. *Am. Nat.* **195**, E67–E86 (2020).
6. Singhal, S. et al. Does population structure predict rate of speciation? A comparative test across Australia's most diverse vertebrate radiation. *Am. Nat.* **192**, 432–447 (2018).

7. Edwards, S. V., Robin, V. V., Ferrand, N. & Moritz, C. The evolution of comparative phylogeography: putting the geography (and more) into comparative population genomics. *Genome Biol. Evol.* **14**, evab176 (2022).
8. Gagnaire, P. A. Comparative genomics approach to evolutionary process connectivity. *Evol. Appl.* **13**, 1320–1334 (2020).
9. Zbinden, Z. D., Douglas, M. R., Chafin, T. K. & Douglas, M. E. Riverscape community genomics: a comparative analytical approach to identify common drivers of spatial structure. *Mol. Ecol.* **32**, 6743–6765 (2022).
10. Riginos, C. & Jahnke, M. Comparative landscape genomics has arrived with a splash. *Mol. Ecol.* **32**, 6725–6728 (2023).
11. Ralston, J. & Kirchman, J. J. Continent-scale genetic structure in a boreal forest migrant, the blackpoll warbler (*Setophaga striata*). *Auk* **129**, 467–478 (2012).
12. Weir, J. T. & Schluter, D. Ice sheets promote speciation in boreal birds. *Proc. R. Soc. B* **271**, 1881–1887 (2004).
13. Ruegg, K. C. & Smith, T. B. Not as the crow flies: a historical explanation for circuitous migration in Swainson's thrush (*Catharus ustulatus*). *Proc. R. Soc. B* **269**, 1375–1381 (2002).
14. Graham, B. A. & Burg, T. M. Molecular markers provide insights into contemporary and historic gene flow for a non-migratory species. *J. Avian Biol.* **43**, 198–214 (2012).
15. Ralston, J. et al. Comparative phylogeographic analysis suggests a shared history among eastern North American boreal forest birds. *Ornithology* **138**, ukab018 (2021).
16. Slatkin, M. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**, 264–279 (1993).
17. Winger, B. M. & Pegan, T. M. Migration distance is a fundamental axis of the slow-fast continuum of life history in boreal birds. *Ornithology* **138**, ukab043 (2021).
18. Fonseca, E. M., Pelletier, T. A., Decker, S. K., Parsons, D. J. & Carstens, B. C. Pleistocene glaciations caused the latitudinal gradient of within-species genetic diversity. *Evol. Lett.* **7**, 331–338 (2023).
19. Paradis, E., Baillie, S. R., Sutherland, W. J. & Gregory, R. D. Patterns of natal and breeding dispersal in birds. *J. Anim. Ecol.* **67**, 518–536 (1998).
20. Weeks, B. C. et al. Morphological adaptations linked to flight efficiency and aerial lifestyle determine natal dispersal distance in birds. *Funct. Ecol.* **36**, 1681–1689 (2022).
21. Turbek, S. P., Scordato, E. S. C. & Safran, R. J. The role of seasonal migration in population divergence and reproductive isolation. *Trends Ecol. Evol.* **33**, 164–175 (2018).
22. Winker, K. On the origin of species through heteropatric differentiation: a review and a model of speciation in migratory animals. *Ornithol. Monogr.* **69**, 1–30 (2010).
23. Chu, J. J. & Claramunt, S. Determinants of natal dispersal distances in North American birds. *Ecol. Evol.* **13**, e9789 (2023).
24. Toews, D. P. L. Habitat suitability and the constraints of migration in New World warblers. *J. Avian Biol.* **48**, 1614–1623 (2017).
25. Bensch, S. Is the range size of migratory birds constrained by their migratory program? *J. Biogeogr.* **26**, 1225–1235 (1999).
26. Greenberg, R. & Marra, P. P. (eds) *Birds of Two Worlds: The Ecology and Evolution of Migration* (Johns Hopkins Univ. Press, 2005).
27. Kimmitt, A. A., Pegan, T. M., Jones, A. W., Winker, K. & Winger, B. M. How veeries vary: whole genome sequencing resolves genetic structure in a long-distance migratory bird. *Ornithology* **141**, ukad061 (2023).
28. Johnson, O. et al. Amazonian birds in more dynamic habitats have less population genetic structure and higher gene flow. *Mol. Ecol.* **32**, 2186–2205 (2023).
29. Wilkins, J. F. A separation-of-timescales approach to the coalescent in a continuous population. *Genetics* **168**, 2227–2244 (2004).
30. Hewitt, G. M. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* **58**, 247–276 (1996).
31. Dyke, A. S. & Prest, V. K. Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Géogr. Phys. Quatern.* **41**, 237–263 (1987).
32. Nei, M., Maruyama, T. & Chakraborty, R. The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1–10 (1975).
33. Frankham, R. Genetics and extinction. *Biol. Conserv.* **126**, 131–140 (2005).
34. García-Berro, A. et al. Migratory behaviour is positively associated with genetic diversity in butterflies. *Mol. Ecol.* **32**, 560–574 (2023).
35. Pegan, T. M., Berv, J. S., Gulson-Castillo, E. R., Kimmitt, A. A. & Winger, B. M. The pace of mitochondrial molecular evolution varies with seasonal migration distance. *Evolution* **78**, 160–173 (2024).
36. Ellegren, H. & Galtier, N. Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433 (2016).
37. Mackintosh, A. et al. The determinants of genetic diversity in butterflies. *Nat. Commun.* **10**, 3466 (2019).
38. Kimmitt, A. A. et al. Genetic evidence for widespread population size expansion in North American boreal birds prior to the Last Glacial Maximum. *Proc. R. Soc. B* **290**, 20221334 (2023).
39. Ovaskainen, O. et al. How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* **20**, 561–576 (2017).
40. Novembre, J. et al. Genes mirror geography within Europe. *Nature* **456**, 98–102 (2008).
41. Winger, B. M., Auteri, G. G., Pegan, T. M. & Weeks, B. C. A long winter for the Red Queen: rethinking the evolution of seasonal migration. *Biol. Rev.* **94**, 737–752 (2019).
42. Kokko, H. Competition for early arrival in migratory birds. *J. Anim. Ecol.* **68**, 940–950 (1999).
43. Stamps, J. A. & Davis, J. M. Adaptive effects of natal experience on habitat selection by dispersers. *Anim. Behav.* **72**, 1279–1289 (2006).
44. McNamara, J. M. & Dall, S. R. X. The evolution of unconditional strategies via the 'multiplier effect'. *Ecol. Lett.* **14**, 237–243 (2011).
45. DeLuca, W. V. et al. A boreal songbird's 20,000 km migration across North America and the Atlantic Ocean. *Ecology* **100**, e02651 (2019).
46. Conklin, J. R. et al. High dispersal ability versus migratory traditions: fine-scale population structure and post-glacial colonisation in bar-tailed godwits. *Mol. Ecol.* **33**, e17452 (2024).
47. Luna, L. W. et al. Late Pleistocene landscape changes and habitat specialization as promoters of population genomic divergence in Amazonian floodplain birds. *Mol. Ecol.* **32**, 214–228 (2023).
48. Talavera, A. & Tellería, J. L. Does microhabitat use affect population differentiation? A test with southwestern Palaearctic forest birds. *J. Ornithol.* **163**, 923–929 (2022).
49. Salisbury, C. L., Seddon, N., Cooney, C. R. & Tobias, J. A. The latitudinal gradient in dispersal constraints: ecological specialisation drives diversification in tropical birds. *Ecol. Lett.* **15**, 847–855 (2012).
50. Ovaskainen, O. & Abrego, N. *Joint Species Distribution Modeling With Applications in R* (Cambridge Univ. Press, 2020).
51. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
52. Medina, I., Cooke, G. M. & Ord, T. J. Walk, swim or fly? Locomotor mode predicts genetic differentiation in vertebrates. *Ecol. Lett.* **21**, 638–645 (2018).
53. D'Urban Jackson, J. et al. Polygamy slows down population divergence in shorebirds. *Evolution* **71**, 1313–1326 (2017).
54. Kwon, E. et al. Breeding site fidelity is lower in polygamous shorebirds and male-biased in monogamous species. *Behav. Ecol.* **33**, 592–605 (2022).

55. Li, M. H. & Merilä, J. Genetic evidence for male-biased dispersal in the Siberian jay (*Perisoreus infaustus*) based on autosomal and Z-chromosomal markers. *Mol. Ecol.* **19**, 5281–5295 (2010).
56. Drever, M. C., Smith, A. C., Venier, L. A., Sleep, D. J. H. & MacLean, D. A. Cross-scale effects of spruce budworm outbreaks on boreal warblers in eastern Canada. *Ecol. Evol.* **8**, 7334–7345 (2018).
57. Pierson, J. C., Allendorf, F. W., Drapeau, P. & Schwartz, M. K. Breed locally, disperse globally: fine-scale genetic structure despite landscape-scale panmixia in a fire-specialist. *PLoS ONE* **8**, e67248 (2013).
58. Orme, D. et al. Caper: comparative analyses of phylogenetics and evolution in R https://cran.r-project.org/web/packages/caper/index.html (2012).
59. Korunes, K. L. & Samuk, K. pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).
60. Prasad, A., Lorenzen, E. D. & Westbury, M. V. Evaluating the role of reference-genome phylogenetic distance on evolutionary inference. *Mol. Ecol. Resour.* **22**, 45–55 (2022).
61. Thorburn, D. M. J. et al. Origin matters: using a local reference genome improves measures in population genomics. *Mol. Ecol. Resour.* **23**, 1706–1723 (2023).
62. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
63. Excoffier, L. in *Handbook of Statistical Genetics* 3rd edn (eds Balding, D. J. et al.) 980–1020 (Wiley, 2007).
64. Pannell, J. R. & Charlesworth, B. Effects of metapopulation processes on measures of genetic diversity. *Phil. Trans. R. Soc. B* **355**, 1851–1864 (2000).
65. Banks, S. C. et al. How does ecological disturbance influence genetic diversity? *Trends Ecol. Evol.* **28**, 670–679 (2013).
66. Calvert, A. M., Walde, S. J. & Taylor, P. D. Nonbreeding-season drivers of population dynamics in seasonal migrants: conservation parallels across taxa. *Avian Conserv. Ecol.* **4**, 5 (2009).
67. Forsman, J. T. & Mönkkönen, M. The role of climate in limiting European resident bird populations. *J. Biogeogr.* **30**, 55–70 (2003).
68. Both, C. et al. Avian population consequences of climate change are most severe for long-distance migrants in seasonal habitats. *Proc. R. Soc. B* **277**, 1259–1266 (2010).
69. De Meester, L., Stoks, R. & Brans, K. I. Genetic adaptation as a biological buffer against climate change: potential and limitations. *Integr. Zool.* **13**, 372–391 (2018).
70. Omernik, J. M. & Griffith, G. E. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environ. Manag.* **54**, 1249–1266 (2014).
71. Cumming, S. G. et al. Climate and vegetation hierarchically structure patterns of songbird distribution in the Canadian boreal region. *Ecography* **37**, 137–151 (2014).
72. Stralberg, D. et al. Biogeography of boreal passerine range dynamics in western North America: past, present, and future. *Ecography* **40**, 1050–1066 (2017).
73. Fair, J., Paul, E., Jones, J. & Bies, L. (eds) *Guidelines to the Use of Wild Birds in Research* (Ornithological Council, 2023).
74. Milá, B., Smith, T. B. & Wayne, R. K. Speciation and rapid phenotypic differentiation in the yellow-rumped warbler *Dendroica coronata* complex. *Mol. Ecol.* **16**, 159–173 (2007).
75. Milá, B., McCormack, J. E., Castañeda, G., Wayne, R. K. & Smith, T. B. Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus *Junco*. *Proc. R. Soc. B* **274**, 2653–2660 (2007).
76. Colbeck, G. J., Gibbs, H. L., Marra, P. P., Hobson, K. & Webster, M. S. Phylogeography of a widespread North American migratory songbird (*Setophaga ruticilla*). *J. Hered.* **99**, 453–463 (2008).
77. van Els, P., Spellman, G. M., Smith, B. T. & Klicka, J. Extensive gene flow characterizes the phylogeography of a North American migrant bird: black-headed grosbeak (*Pheucticus melanocephalus*). *Mol. Phylogenet. Evol.* **78**, 148–159 (2014).
78. Hindley, J. A., Graham, B. A., Pulgarin-R, P. C. & Burg, T. M. The influence of latitude, geographic distance, and habitat discontinuities on genetic variation in a high latitude montane species. *Sci. Rep.* **8**, 11846 (2018).
79. Miller, C. V. et al. Genomics-informed conservation units reveal spatial variation in climate vulnerability in a migratory bird. *Mol. Ecol.* **33**, e17199 (2023).
80. Perez, M. F. et al. Assessing population structure in the face of isolation by distance: are we neglecting the problem? *Divers. Distrib.* **24**, 1883–1889 (2018).
81. Schweizer, T. M. & DeSaix, M. G. Cost-effective library preparation for whole genome sequencing with feather DNA. *Conserv. Genet. Resour.* **15**, 21–28 (2023).
82. Therkildsen, N. O. & Palumbi, S. R. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Mol. Ecol. Resour.* **17**, 194–208 (2017).
83. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
84. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
85. Lou, R. N. & Therkildsen, N. O. Batch effects in population genomic studies with low-coverage whole genome sequencing data: causes, detection and mitigation. *Mol. Ecol. Resour.* **22**, 1678–1692 (2021).
86. Zhang, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1321 (2014).
87. Ruegg, K. et al. Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecol. Lett.* **21**, 1085–1096 (2018).
88. Toews, D. P. L. et al. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr. Biol.* **26**, 2313–2318 (2016).
89. Feng, S. et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
90. Laine, V. N. et al. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* **7**, 10474 (2016).
91. Manthey, J. D., Klicka, J. & Spellman, G. M. The genomic signature of allopatric speciation in a songbird is shaped by genome architecture (Aves: *Certhia americana*). *Genome Biol. Evol.* **13**, evab120 (2021).
92. Friis, G., Vizueta, J., Ketterson, E. D. & Milá, B. A high-quality genome assembly and annotation of the dark-eyed junco *Junco hyemalis*, a recently diversified songbird. *G3* **12**, jkac083 (2022).
93. Sly, N. D. et al. Molecular parallelism in signaling function across different sexually selected ornaments in a warbler. *Proc. Natl Acad. Sci. USA* **119**, e2120482119 (2022).
94. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
95. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
96. Jun, G., Wing, M. K., Abecasis, G. & Kang, H. M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* https://doi.org/10.1101/gr.176552.114 (2015).
97. Van der Auwera, G. A. et al. From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* https://doi.org/10.1002/0471250953.bi1110s43 (2013).

98. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **45**, e18 (2016).
99. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinform.* **15**, 356 (2014).
100. Linderoth, T. *Identifying Population Histories, Adaptive Genes, and Genetic Duplication from Population-Scale Next Generation Sequencing*. PhD dissertation, Univ. California, Berkeley (2018).
101. Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
102. Harringmeyer, O. S. & Hoekstra, H. E. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nat. Ecol. Evol.* **6**, 1965–1979 (2022).
103. Ishigohoka, J. et al. Distinct patterns of genetic variation at low-recombining genomic regions represent haplotype structure. *Evolution* **78**, 1916–1935 (2024).
104. Li, H. & Ralph, P. Local PCA shows how the effect of population structure differs along the genome. *Genetics* **211**, 289–304 (2019).
105. Privé, F., Luu, K., Blum, M. G. B., McGrath, J. J. & Vilhjálmsson, B. J. Efficient toolkit implementing best practices for principal component analysis of population genetic data. *Bioinformatics* **36**, 4449–4457 (2020).
106. Waters, P. D. et al. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc. Natl Acad. Sci. USA* **118**, e2112494118 (2021).
107. Hijmans, R. J. Package 'geosphere' https://cran.r-project.org/web/packages/geosphere/index.html (2022).
108. Pegan, T. M. & Winger, B. M. The influence of seasonal migration on range size in temperate North American passerines. *Ecography* **43**, 1191–1202 (2020).
109. Dunning, J. B. J. *CRC Handbook of Avian Body Masses* (CRC Press, 2008).
110. *Birds of the World* (Cornell Lab of Ornithology, 2022).
111. Rasmussen, M. S., Garcia-Erill, G., Korneliussen, T. S., Wiuf, C. & Albrechtsen, A. Estimation of site frequency spectra from low-coverage sequencing data using stochastic EM reduces overfitting, runtime, and memory usage. *Genetics* **222**, iyac148 (2022).
112. Korneliussen, T. S. Heterozygosity https://www.popgen.dk/angsd/index.php/Heterozygosity (2017).
113. Zimova, M. et al. Body size predicts the rate of contemporary morphological change in birds. *Proc. Natl Acad. Sci. USA* **120**, e2206971120 (2023).
114. 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).
115. Rubolini, D., Liker, A., Garamszegi, L. Z., Moller, A. P. & Saino, N. Using the BirdTree.org website to obtain robust phylogenies for avian comparative studies: a primer. *Curr. Zool.* **61**, 959–965 (2015).
116. Youngflesh, C. MCMCvis: tools to visualize, manipulate, and summarize MCMC output. *J. Open Source Softw.* **3**, 640 (2018).
117. Gabry, J., Simpson, D., Vehtari, A., Betancourt, M. & Gelman, A. Visualization in Bayesian workflow. *J. R. Stat. Soc. A* **182**, 389–402 (2019).
118. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
119. Pegan, T. M. et al. Bioinformatic code for manuscript "Seasonal migration to the tropics promotes genetic diversity but not gene flow in boreal birds". *figshare* https://doi.org/10.6084/m9.figshare.27284553 (2025).
120. Pegan, T. M. et al. Seasonal migration to the tropics promotes genetic diversity but not gene flow in boreal birds. *Code Ocean* https://doi.org/10.24433/CO.5578409.v1 (2025).

Acknowledgements

For comments and advice, we thank G. Bradburd, R. Burner, K. Wacker, E. Gulson-Castillo, V. Gómez-Bahamón, J. Berv, Z. Hancock, M. Hack, A. Marshall, L. Knowles, D. Rabosky, M. Witynksi and A. Benavides. For providing samples, we thank the curators, collections staff and field collectors from the following institutions: American Museum of Natural History, Cleveland Museum of Natural History, Cornell University Museum of Vertebrates, New York State Museum, Royal Alberta Museum, Royal Ontario Museum, University of Alaska Museum of the North, University of Minnesota Museum of Natural History, University of California, Berkeley Museum of Vertebrate Zoology, University of Michigan Museum of Zoology and University of Washington Burke Museum. For additional samples, we thank J. Tremblay (Environment and Climate Change Canada). For field permits, we thank the United States Fish and Wildlife Service, the United States Forest Service, the Minnesota Department of Natural Resources, the Michigan Department of Natural Resources, the Canadian Wildlife Service of Environment and Climate Change Canada, Alberta Fish and Wildlife, and Manitoba Fish and Wildlife. Field sampling was approved by the University of Michigan Animal Care and Use Committee (no. PRO00010608). For assistance in the field, we thank C. Brennan, S. Campbell, S. DuBay, G. M. Erickson, M. M. Ferraro, A. FitzGerald, L. Gooch, E. Gulson-Castillo, J. Ralston, C. Scobie, H. Skeen, V. Ting, K. Wacker and members of the Burg lab. For assistance in the lab, we thank T. Schweizer, C. Rayne, R. Herman, J. Yan, C. Kaczmarek, M. Florkowski, M. Guza, C. Pajka, M. Hack and C. Jordan. Next-generation sequencing for this project was partially carried out in the Advanced Genomics Core at the University of Michigan. This research was also supported in part through computational resources and services provided by Advanced Research Computing (ARC), a division of Information and Technology Services (ITS) at the University of Michigan, Ann Arbor. Funding includes National Science Foundation grant DEB 2146950 (B.M.W.), National Science Foundation Graduate Research Fellowship DGE 1256260, Fellow ID 2018240490 (T.M.P.), Jean Wright Cohn Endowment Fund at the University of Michigan Museum of Zoology, Robert W. Storer Endowment Fund at the University of Michigan Museum of Zoology, Mary Rhoda Swales Museum of Zoology Research Fund at the University of Michigan Museum of Zoology, William G. Fargo Fund at the University of Michigan Museum of Zoology, William A. and Nancy R. Klamm Endowment funds at the Cleveland Museum of Natural History, University of Michigan Rackham Graduate Student Research Grant (T.M.P.), British Ornithologists Union Small Research Grant (T.M.P.), and American Museum of Natural History Chapman Research Grant (T.M.P.).

Author contributions

B.M.W. and T.M.P conceived the study. Genomic samples were contributed by all authors and were prepared for sequencing by T.M.P. and A.A.K. with methodological support from K.C.R. Data were analysed and visualized by T.M.P. and B.M.W. T.M.P. and B.M.W. wrote the paper with input and revisions from all authors. All authors approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-025-02699-3.

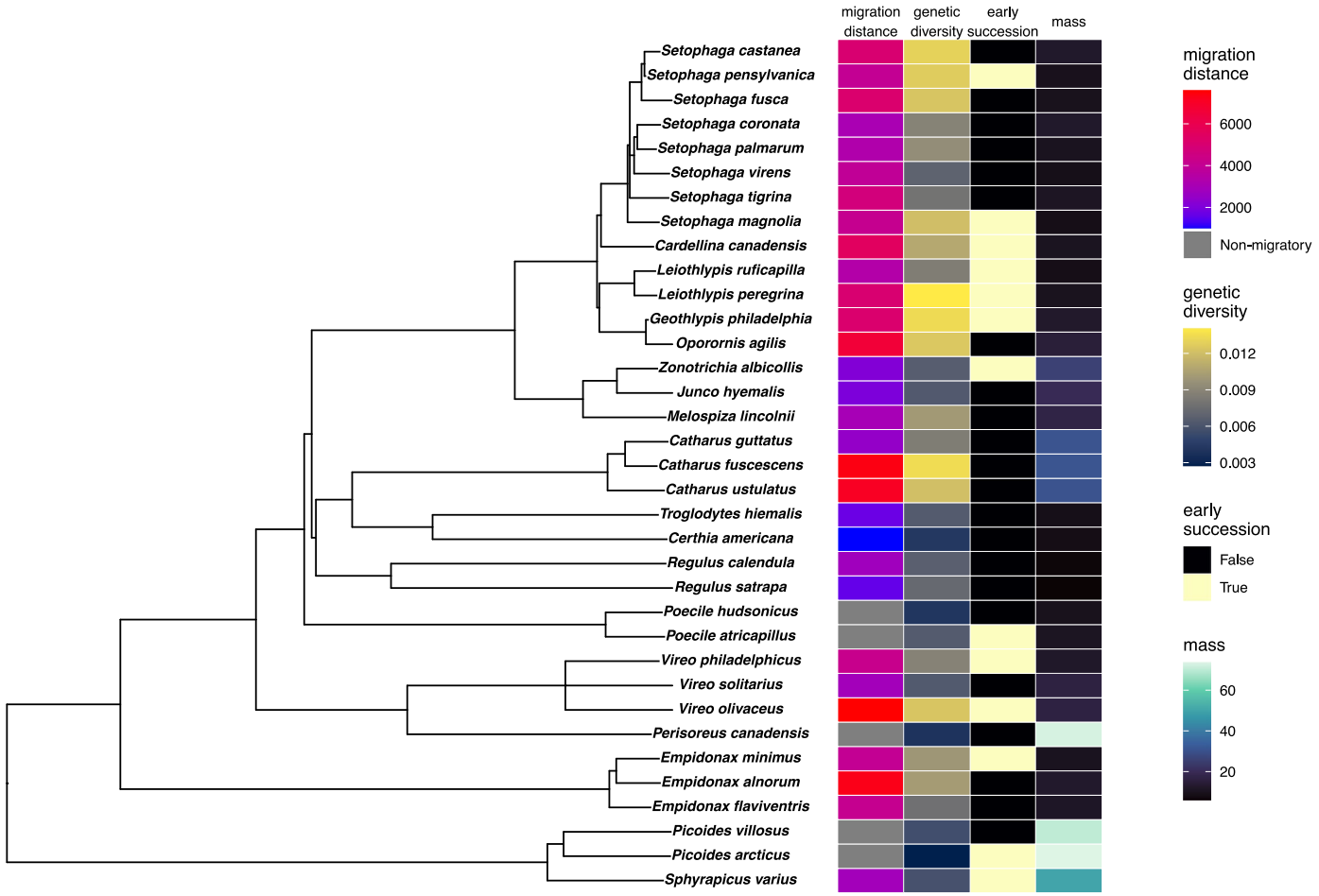
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-025-02699-3.

Correspondence and requests for materials should be addressed to B. M. Winger.

Peer review information *Nature Ecology & Evolution* thanks Leilton Luna and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

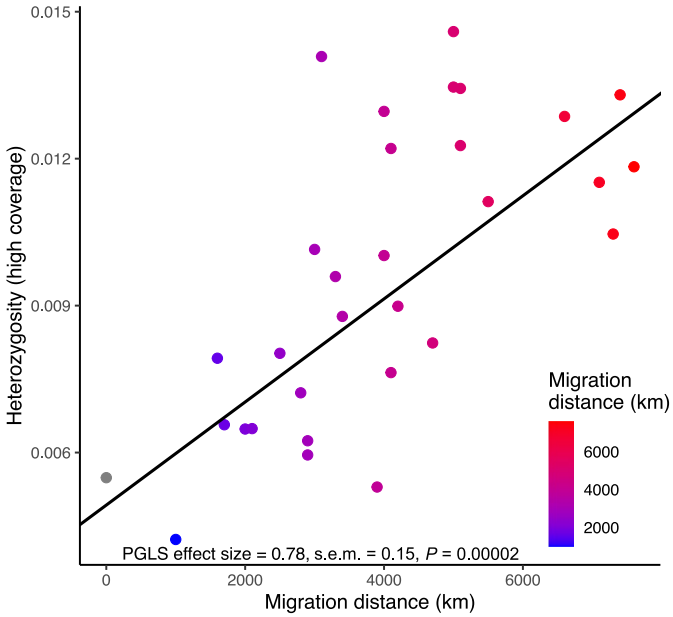
[1]Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [2]School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA. [3]Environment and Climate Change Canada, Québec City, Québec, Canada. [4]Department of Biological Sciences, University of Lethbridge, Lethbridge, Alberta, Canada. [5]Royal Alberta Museum, Edmonton, Alberta, Canada. [6]Cleveland Museum of Natural History, Cleveland, OH, USA. [7]New York State Museum, Albany, NY, USA. [8]Department of Biology, Colorado State University, Fort Collins, CO, USA. [9]Present address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [10]Present address: Department of Biology, Hofstra University, Hempstead, NY, USA. [11]Present address: Spring Island Trust, Okatie, SC, USA. ✉e-mail: wingerb@umich.edu

https://doi.org/10.1038/s41559-025-02699-3



Extended Data Fig. 1 | The phylogenetic relationship among the species in the study and species attributes included in HMSC models. The phylogenetic relationship is shown on the tree from birdtree.org used in our analyses. The tree is plotted alongside a heatmap representing interspecific variation in the four species attributes we investigated in HMSC models (migration distance, mass, association with early successional habitat, and genetic diversity). Genetic diversity estimates were calculated in this study (Supplementary Table 1) and values for the other three attributes come from published sources as described in the Methods.

https://doi.org/10.1038/s41559-025-02699-3



Extended Data Fig. 2 | Migration distance correlates with an estimate of heterozygosity from higher-coverage samples. Each point represents one species ($N = 27$) and points are colored by migration distance, as in main text figures, where non-migrants are shown in gray. This plot shows heterozygosity estimated by ANGSD from reference-mapped genomes. The intercept and slope of the line were estimated with a phylogenetic generalized least squares (PGLS) model.

https://doi.org/10.1038/s41559-025-02699-3

Extended Data Table 1 | Species used in analyses

Species	Family	*N*	Mig. dist. (km)	Mean mass (g)	Early successional habitat association	Genetic diversity (θ_π)	Mean hetero-zygosity	β_{IBD}	β_{IBD} posterior support
Picoides arcticus	Picidae	31	0	74	TRUE	0.0027	0.0020	-0.1083	0.9985
Dryobates villosus	Picidae	43	0	70.36	FALSE	0.0056	0.0051	-0.6978	1.0000
Sphyrapicus varius	Picidae	62	2900	50.3	TRUE	0.0058	0.0046	-0.4782	1.0000
Empidonax alnorum	Tyrannidae	49	7300	12.79	FALSE	0.0102	0.0098	-0.0223	0.8723
Empidonax flaviventris	Tyrannidae	58	4100	11.6	FALSE	0.0077	0.0071	-0.1635	1.0000
Empidonax minimus	Tyrannidae	44	4000	10.3	TRUE	0.0099	0.0095	-0.0291	0.8387
Perisoreus canadensis	Corvidae	25	0	73	FALSE	0.0039	0.0041	-0.6740	1.0000
Vireo olivaceus	Vireonidae	63	7600	16.7	TRUE	0.0124	0.0119	-0.2431	1.0000
Vireo philadelphicus	Vireonidae	17	4200	12.2	TRUE	0.0088	0.0087	0.0248	0.3338
Vireo solitarius	Vireonidae	56	2900	16.6	FALSE	0.0064	0.0061	-0.4373	1.0000
Poecile atricapillus	Paridae	59	0	10.8	TRUE	0.0064	0.0064	-0.7316	1.0000
Poecile hudsonicus	Paridae	37	0	9.8	FALSE	0.0041	0.0040	-0.7770	1.0000
Corthylio calendula	Regulidae	49	2800	6.68	FALSE	0.0067	0.0063	-0.0597	0.9960
Regulus satrapa	Regulidae	63	1600	6.23	FALSE	0.0073	0.0066	-0.4849	1.0000
Certhia americana	Certhiidae	48	1000	8.4	FALSE	0.0043	0.0043	0.0154	0.2773
Troglodytes hiemalis	Troglodytidae	26	1700	8.9	FALSE	0.0064	0.0065	0.0116	0.4107
Catharus fuscescens	Turdidae	47	7400	31.2	FALSE	0.0134	0.0127	-0.8952	1.0000
Catharus guttatus	Turdidae	65	2500	31	FALSE	0.0085	0.0080	-0.0660	1.0000
Catharus ustulatus	Turdidae	66	7100	30.8	FALSE	0.0121	0.0118	-0.3142	1.0000
Junco hyemalis	Passerellidae	59	2000	19.89	FALSE	0.0063	0.0063	-0.0429	0.9947
Melospiza lincolnii	Passerellidae	53	3000	17.4	FALSE	0.0101	0.0098	-0.0581	0.9988
Zonotrichia albicollis	Passerellidae	68	2100	25.9	TRUE	0.0065	0.0062	-0.0303	0.9768
Cardellina canadensis	Parulidae	31	5500	10.42	TRUE	0.0109	0.0108	-0.6831	1.0000
Geothlypis philadelphia	Parulidae	52	5100	12.61	TRUE	0.0134	0.0130	-0.8201	1.0000
Oporornis agilis	Parulidae	30	6600	15.2	FALSE	0.0125	0.0121	0.0442	0.0995
Leiothlypis peregrina	Parulidae	47	5000	10.02	TRUE	0.0141	0.0139	0.0188	0.1660
Leiothlypis ruficapilla	Parulidae	57	3400	8.73	TRUE	0.0085	0.0083	-0.0937	1.0000
Setophaga castanea	Parulidae	45	5000	12.59	FALSE	0.0130	0.0128	-0.0172	0.8332
Setophaga coronata	Parulidae	70	3100	12.51	FALSE	0.0088	0.0090	-0.0880	1.0000
Setophaga fusca	Parulidae	52	5100	9.7	FALSE	0.0124	0.0127	-0.1667	1.0000
Setophaga magnolia	Parulidae	56	4100	8.72	TRUE	0.0120	0.0118	-0.4074	1.0000
Setophaga palmarum	Parulidae	51	3300	10.3	FALSE	0.0094	0.0092	-0.5598	1.0000
Setophaga pensylvanica	Parulidae	46	4000	9.64	TRUE	0.0127	0.0127	-0.0813	0.9997
Setophaga tigrina	Parulidae	44	4700	11	FALSE	0.0079	0.0080	0.0017	0.4892
Setophaga virens	Parulidae	63	3900	8.8	FALSE	0.0069	0.0052	-0.5739	1.0000

Species used in analyses along with their taxonomic family and the number of samples used, values for species attributes used in HMSC models to infer γ coefficients (migration distance, mass, early successional habitat association, and genetic diversity), and posterior estimates of β_{IBD} and its mean posterior probability from the version of our HMSC model that included all species. β_{IBD} estimates from the other version of the HMSC model (without non-migratory species) were highly similar. We also include estimates of mean individual heterozygosity, which were not used in HMSC models.

https://doi.org/10.1038/s41559-025-02699-3

Extended Data Table 2 | Samples used to generate high-coverage genomes

Specimen number	Voucher museum	Species	BAM mean cov. depth	Het. (ANGSD)	Het. (Genome Scope)
UMMZ:Bird:248029	UMMZ	*Dryobates villosus*	16.81	0.0055	0.0052
75074	CMNH	*Sphyrapicus varius*	12.54	0.0060	0.0063
75110	CMNH	*Empidonax alnorum*	15.14	0.0105	0.0116
75067	CMNH	*Empidonax flaviventris*	15.47	0.0076	0.0088
UMMZ:Bird:248037	UMMZ	*Empidonax minimus*	16.46	0.0100	0.0118
UMMZ:Bird:248147	UMMZ	*Vireo olivaceus*	6.56	0.0118	Failed to converge
UMMZ:Bird:247356	UMMZ	*Vireo philadelphicus*	10.35	0.0090	0.0115
UMMZ:Bird:248151	UMMZ	*Vireo solitarius*	15.81	0.0062	0.0098
UMMZ:Bird:245292	UMMZ	*Corthylio calendula*	10.46	0.0072	0.0073
UMMZ:Bird:245950	UMMZ	*Regulus satrapa*	9.56	0.0079	0.0129
UMMZ:Bird:245242	UMMZ	*Certhia americana*	13.57	0.0042	0.0051
UMMZ:Bird:245234	UMMZ	*Troglodytes hiemalis*	11.32	0.0066	0.0060
UMMZ:Bird:247380	UMMZ	*Catharus fuscescens*	16.28	0.0133	0.0153
72443	CMNH	*Catharus guttatus*	7.80	0.0080	Failed to converge
72718	CMNH	*Catharus ustulatus*	9.93	0.0115	Failed to converge
72451	CMNH	*Junco hyemalis*	20.02	0.0065	0.0084
UMMZ:Bird:248059	UMMZ	*Melospiza lincolnii*	12.17	0.0101	Failed to converge
UMMZ:Bird:245355	UMMZ	*Zonotrichia albicollis*	18.91	0.0065	0.0098
UMMZ:Bird:248005	UMMZ	*Cardellina canadensis*	14.62	0.0111	0.0138
UMMZ:Bird:248040	UMMZ	*Geothlypis philadelphia*	14.11	0.0134	0.0149
UMMZ:Bird:245337	UMMZ	*Leiothlypis peregrina*	14.34	0.0146	0.0163
75054	CMNH	*Leiothlypis ruficapilla*	14.13	0.0088	0.0114
UMMZ:Bird:247273	UMMZ	*Oporornis agilis*	22.48	0.0129	0.0147
UMMZ:Bird:245367	UMMZ	*Setophaga castanea*	17.53	0.0135	0.0158
72396	CMNH	*Setophaga coronata*	10.81	0.0141	0.0144
75189	CMNH	*Setophaga fusca*	17.51	0.0123	0.0151
UMMZ:Bird:245321	UMMZ	*Setophaga magnolia*	16.32	0.0122	0.0153
UMMZ:Bird:245277	UMMZ	*Setophaga palmarum*	13.87	0.0096	0.0114
UMMZ:Bird:245312	UMMZ	*Setophaga pensylvanica*	17.03	0.0130	0.0153
76969	CMNH	*Setophaga tigrina*	16.34	0.0082	0.0097
UMMZ:Bird:245245	UMMZ	*Setophaga virens*	18.67	0.0053	0.0074

Samples used to generate high-coverage genomes are listed with the museum where each sample is vouchered, the species, the average genome-wide coverage depth of each alignment, and the average genome-wide heterozygosity we calculated with ANGSD. Additional metadata for each sample can be found in Dataset S1 and information about the reference genome used for each species in Dataset S2. UMMZ, University of Michigan Museum of Zoology; CMNH, Cleveland Museum of Natural History.

https://doi.org/10.1038/s41559-025-02699-3

Extended Data Table 3 | Comparison of strengths and caveats associated with each of four methods we used to estimate genetic diversity and heterozygosity

Method number	Method	Strengths	Caveats
1	θ_π from low-coverage population genomic data	The most appropriate metric to represent genetic diversity of the entire sampled population	Derived from low-coverage reference-mapped data: biases possible from missing data, ANGSD's specific algorithms, reference bias
2	Mean heterozygosity from low-coverage population genomic data	Directly comparable to high-coverage heterozygosity estimates listed below	All caveats above apply; in addition, individual heterozygosity values can sometimes deviate from population-wide diversity metrics
3	Heterozygosity from 1 high coverage individual per species – estimated with ANGSD	Mitigates biases arising from low coverage/missing data	Individual heterozygosity values can sometimes deviate from population-wide diversity metrics; reference bias or unidentified ANGSD biases still possible
4	Heterozygosity from 1 high coverage individual per species – estimated with GenomeScope	Mitigates all caveats listed for method 1: uses high coverage genomes, derived solely from *k*-mer graphs produced without reference mapping or ANGSD	Individual heterozygosity values can sometimes deviate from population-wide diversity metrics

Methods 1 and 2 rely on low-coverage population genomic data, while methods 3 and 4 use higher coverage genomes from single representatives per species. All four resulting parameters display highly consistent interspecific variation, such that all four parameters also correlate strongly and significantly with migration distance across species (see Results and discussion). The congruence of these results suggests that they are not being driven by caveats particular to each estimation method.

nature portfolio

Corresponding author(s): Benjamin Winger

Last updated by author(s): 24 February 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	All code and software, including software versions, are described in detail in the code associated with this manuscript, which is deposited in the associated Code Ocean capsule (comparative analyses and software https://doi.org/10.24433/CO.5578409.v1) and figshare repository (all details of bioinformatic analyses and software https://doi.org/10.6084/m9.figshare.27284553). The major pieces of software used are bwa (v0.7.15), ANGSD (v0.941), PCAngsd (v1.11), and the HMSC R package (v3.0-13).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The sequence data generated during the study have been uploaded to the NCBI Sequence Read Archive under BioProjects PRJNA1043688 and PRJNA1130443.

Individual accession numbers for each sample are provided in Dataset S1 in the Supplementary Information. All other data are available in within the manuscript, the repositories described in the Code Availability statement, and the Supplementary Information files.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No human data were used.
Reporting on race, ethnicity, or other socially relevant groupings	No human data were used.
Population characteristics	No human data were used.
Recruitment	No human data were used.
Ethics oversight	No human data were used.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	In this study we describe the amount of genetic diversity in co-distributed populations of 35 species of birds and we describe how genetic covariance between individuals varies in geographic space. We then evaluate the relationship between intrinsic species attributes and the population genetic patterns using hierarchical Bayesian modeling and phylogenetic generalized least squares modeling.
Research sample	We use genomic samples gathered from the co-distributed boreal breeding ranges of 35 species of boreal birds. We chose these 35 species because they represent a core subset of the assemblage of small-bodied migratory and resident birds in the North American boreal forest. Species excluded from analyses which otherwise meet similar criteria fall into two categories: species for which insufficient sampling exists from the study area, or species known to engage in nomadic or irruptive movements. A list of the 35 taxa used in analyses is provided in Table S1. We used samples that were collected from adult individuals during the breeding season. Both males and females were sampled.
Sampling strategy	We aimed to sample each species as broadly and evenly as possible throughout our study area, to the extent possible based on sample availability and species ranges. To aid in balanced sampling, we did not sample more than 15 individuals per species from any geographic subregion (Alberta; Manitoba; the Great Lakes region comprising northern Michigan and Minnesota; the Adirondack and Green Mountain region comprising New York and Vermont; Quebec; and the Canadian Maritimes). We did not include any species in analyses for which we had fewer than 15 samples total after filtering.
Data collection	For each sample, the key pieces of data are the locality and the species of the sample. Locality data were recorded with GPS points and provided by the museum institution loaning the sample and species identity was confirmed by using the NCBI BLAST tool on mitochondrial gene sequences.
Timing and spatial scale	Genomic sampling was based on samples collected within the eastern boreal belt ecoregion between Alberta and the Canadian Maritime provinces, including parts of central and eastern Canada and the United States. All samples were collected during the boreal avian breeding season (mostly the month of June with some samples from late May or July). Samples were collected between the years 1981 and 2022.
Data exclusions	We excluded samples from analyses only when they failed data quality filters.
Reproducibility	This study does not involve experimental manipulation. All analyses used to process genomic data, describe patterns, and test correlations are fully reproducible.
Randomization	Individuals were not randomized in this study. Sampling was determined by sample availability as described above
Blinding	Blinding was not possible or necessary in this study.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Environmental conditions were not relevant to data collection in this study.
Location	Fieldwork associated with the samples we used occurred at various localities (shown in Fig 1 and Dataset S1) throughout the boreal ecoregion in the United States and Canada.
Access & import/export	Most samples were loaned from existing natural history tissue collections at various institutions. Additional samples collected by the authors in the US and Canada were done with all necessary permits, including US Fish and Wildlife Service MB17024C, Canadian Wildlife Service 16-AB-SC002 and 19-MB/SK-SC001 and all necessary state, provincial and local authorizations, and UM IACUC PRO00010608. Importation of samples to the US from Canada were done in accordance with procedures described in "A Guide to the Permits and Procedures for Importing Bird Products into the United States for Scientific Research and Display, 2nd Edition" published by the Ornithological Council, July 2020 (https://birdnet.org/wp-content/uploads/2023/11/ImportGuide_November-2023.pdf), and with all necessary permits (USDA APHIS 130628, USFWS MB745448).
Disturbance	This study did not cause disturbance.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	This study did not involve laboratory animals.
Wild animals	This study involves 35 species of wild birds listed in Table S1.
Reporting on sex	No findings from this study are sex-specific.
Field-collected samples	No animals were housed for this research. Samples were collected in accordance with the Ornithological Council guidelines for use of wild birds in research (Fair et al 2023). Fair, J., E. Paul, J. Jones and L. Bies, Eds. 2023. Guidelines to the Use of Wild Birds in Research. Washington, D.C. Ornithological Council. https://birdnet.org/wp-content/uploads/2023/09/Guidelines-September-2023.pdf
Ethics oversight	All work was approved by the University of Michigan Institutional Animal Care & Use Committee (IACUC) protocol number PRO00010608

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	No plants were used in this study.
Novel plant genotypes	No plants were used in this study.
Authentication	No plants were used in this study.