

University of Groningen

Avian diversification across space

van Els, Paul

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

van Els, P. (2018). *Avian diversification across space*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter II: High latitudes and high genetic diversity: Phylogeography of a widespread boreal bird, the gray jay (*Perisoreus canadensis*)

Paul van Els ^{a,b}, Carla Cicero ^c, John Klicka ^a

^aUniversity of Nevada, Las Vegas, Marjorie Barrick Museum of Natural History, 4505 S. Maryland Parkway, Las Vegas, NV 89154-4012, United States

^b School of Life Sciences, University of Nevada Las Vegas, 4505 S. Maryland Parkway, Box 454004, Las Vegas, NV 89154, United States ^c Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720-3160, United States

Published in Molecular Phylogenetics and Evolution (2012)

Abstract

We describe range-wide phylogeographic variation in gray jays (*Perisoreus canadensis*), a boreal Nearctic corvid that occurs today primarily in recently glaciated regions. Phylogenetic analysis of mitochondrial DNA (1041 base pairs ND2 gene; N = 205, 50 localities) revealed four reciprocally monophyletic groups. One widespread clade occurs across the North American boreal zone, from Newfoundland to Alaska and southwest into Utah. Three other clades occur at lower latitudes in the montane West in Colorado, the northern Rocky Mountains, and the Pacific Northwest respectively. The geographic distribution of clades in gray jays corresponds with a general pattern that is emerging for boreal taxa, having one widespread northern clade and one or more geographically restricted southwestern clades. Population genetic analyses indicate that the larger boreal clade is genetically structured and harbors significantly more genetic diversity than those clades occurring at lower latitudes. Species distribution modeling (SDM) revealed multiple putative Pleistocene refugia including several occurring at higher latitudes. We suggest that multiple post-glacial colonization routes, some of which originate from these northern refugia, are responsible for the relatively high genetic diversity at high latitudes. Conversely, lower latitude clades show little variation, probably as a result of historical restriction to smaller geographical areas with smaller long-term population sizes. This 'upside-down' pattern of genetic diversity contrasts with the conventional view that populations of north-temperate species occupying previously glaciated habitats should possess lower levels of diversity than their southern counterparts.

1. Introduction

Pleistocene glacial cycles have long been recognized as important in shaping the diversity and distributions of North American birds (e.g. Rand, 1948; Mengel, 1964). For most birds breeding in the northern temperate zone, ranges were presumably shifted southward in response to glacial

advances and formerly widespread species were often fragmented into multiple isolated populations (hereafter, the geographic locations of these populations are referred to in the broad sense as refugia) where among population genetic divergence then occurred. These populations often retain a unique 'genetic footprint' after postglacial expansion (Hewitt, 1996) that allows their evolutionary history to be inferred.

For example, demographic parameters such as ancestral and contemporary population sizes, degree of connectedness with other populations, and time since isolation began, can all be estimated using population genetic methods. Important information is also encoded within each population by the relative amount of genetic diversity present. The conventional view suggests that in regions affected by Pleistocene climatic oscillations, populations nearest refugial areas should harbor the highest amounts of genetic diversity (Provan and Bennett, 2008). The expected pattern of 'southern richness and northern purity' (Hewitt, 1996, 1999) is generally attributed to prolonged population stability nearer southern refugia, coupled with the loss of variation during post-glacial expansion. Populations (or groups of populations) that have been isolated by one or more glacial cycles may be identified via phylogeographic analyses as possessing significant geographic genetic structuring and in some cases by forming reciprocally monophyletic groups (clades). Molecular data for birds have shown that the effects of the most recent glacial advances (Late Pleistocene) are more often seen within species, in the form of well-marked subspecies and genetically structured populations (Avise and Walker, 1998). For example, phylogeographic surveys of several widely distributed temperate zone birds, such as blue grouse (*Dendragapus obscurus*; Barrowclough et al., 2004); white-breasted nuthatch (*Sitta carolinensis*; Spellman and Klicka, 2007); brown creeper (*Certhia americana*; Manthey et al., 2011), and Swainson's thrush (*Catharus ustulatus*; Ruegg and Smith, 2002) identify population differentiation that can be attributed to climate change during the recent glacial periods. Current distributions of these species suggest southward displacement and range-size reductions during the late glacial maximum (LGM). A special case of LGM range displacement that is less well explored involves those species restricted to contemporary boreal habitats, as their distribution was replaced entirely (or nearly so) by ice during glacial advances. Although the fossil record has long suggested the presence of southern glacial refugia (macrorefugia, sensu Rull, 2009, 2010) for such high latitude taxa during glacial maxima, only recently have phylogeographic studies given support to the existence of previously unknown, 'cryptic' refugia (Provan and Bennett, 2008; microrefugia of Rull, 2009, but see Stewart et al., 2010). In North America, hypothesized refugial areas now also include the northern Pacific coast (Arbogast, 1999; Cook et al., 2006), the Queen Charlotte Islands (Haida Gwaii; Byun et al., 1997; Burg et al., 2005; Topp and Winker, 2008), Beringia (Fleming and Cook, 2002; Anderson et al., 2006), and the North Atlantic Coast and the Grand Banks off of Newfoundland (Gill et al., 1993; Zink and Dittmann, 1993). While a few phylogeographic studies on boreal birds have been done (e.g. pine grosbeak [*Pinicola enucleator*], Drovetski et al., 2010; fox sparrow [*Passerella i. iliaca*], Zink, 1994; American three-toed woodpecker [*Picoides dorsalis*], Zink et al., 2002), none have explicitly addressed the question of precisely where these species persisted in North America during the LGM (see Waltari et al., 2007) and what routes may have been used for re-colonization. Inferences from the genetic

structure of the broadly distributed gray jay (*Perisoreus canadensis*) may provide insights into the complex postglacial history of North America's boreal taxa.

The gray jay is a sedentary corvid with a distribution entirely restricted to boreal habitats. Currently, the species ranges from Newfoundland through much of Canada to western Alaska, and south to northern California. Isolated populations also occur in the Black Hills of South Dakota, Utah, the southern Rockies, and the White Mountains of Arizona. Geographic-genetic structuring of gray jays is suggested by morphological variation. While as many as 14 subspecies have been recognized (Miller, 1933; Phillips, 1986; Swarth, 1918), the AOU (American Ornithologists' Union, 1998) currently recognizes two distinct morphological assemblages (Pacific Northwest, Taiga and Rocky Mountains; whereas others (e.g. Sibley, 2000) identify three morphologically discrete groups (Taiga, Pacific, Rocky Mountains). Throughout most of its current distribution it is closely associated with the occurrence of black and white spruce (*Picea mariana*, *P. glauca*), the dominant conifer trees in the North American boreal forest (Payette et al., 2001). Regardless of where gray jays weathered the LGM, a nearly complete range displacement over large geographic distances is likely to have affected the genetic characteristics of populations of this species. Examining genetic structure at the population level will allow important inferences to be made regarding the recent evolutionary history of the gray jays and will provide a framework for a discussion of the recent history of other obligately boreal taxa. Herein, we use phylogeographic and demographic analysis of mitochondrial DNA and climate-based species distribution models (SDMs) to investigate the geographic distribution of genetic variation within the species and to infer the locations of glacial refugia, colonization routes, apparent boundaries to gene flow, and potential areas of secondary contact between formerly isolated groups.

2. Materials and methods

2.1. Sample and data collection

We obtained tissue samples for 205 individuals from 50 localities (Fig. 1), representing all six subspecies recognized in the most recent taxonomic treatment of this species (Strickland and Ouellet, 1993). We focused our sampling on the western part of the species' range, where morphological variation is most pronounced and where populations of the species are restricted to suitable montane or coastal coniferous habitat. Since sister taxon (Siberian Jay, *Perisoreus infaustus*) samples were not available, we used another corvid taxon, the American crow (*Corvus brachyrhynchos*), to root all analyses. We extracted DNA from all samples using a DNeasy extraction tissue extraction kit (QIAGEN, Valencia, California) following the manufacturer's protocol. Because of its high variability and rapid coalescence time, mtDNA is the molecule of choice for studies such as ours, where the main objective is to determine the geographic limits of recently evolved groups (see Zink and Barrowclough, 2008). Thus, we amplified the mitochondrial DNA (mtDNA) gene NADH dehydrogenase subunit 2 (ND2) via polymerase

chain reaction (PCR) using the primers L5215 (Hackett, 1996) and HTrpC (Smithsonian Tropical Research Institute, Balboa, Panamá) in 12.5 μ l reactions. We used the program Sequencher v. 4.9 (Gene Codes Corporation, Ann Arbor, Michigan) to align complementary DNA strands, detect stop codons, and translate genetic information into amino acids. We found no gaps, insertions, or deletions in our data.

2.2. Phylogenetic analyses

We constructed phylogenetic trees using both maximum likelihood and Bayesian methods. We used JMODELTEST v. 0.1.1 (Posada, 2008) and Akaike Information Criterion to identify the evolutionary model that most accurately described our data. We performed a molecular clock analysis in MEGA 5.0 using the maximum likelihood method (Tamura et al., 2011). We used GARLI v 1.0 (Zwickl, 2006) to perform maximum likelihood analyses with 1000 boot-strap replicates, setting our evolutionary model according to the JMODELTEST output. Bayesian analyses were executed in BEAST v. 1.6.1 (Drummond and Rambaut, 2007). For our Bayesian analyses we used a GTR+I+C model (AIC = 7100.4081) and ran our analyses for 15 million generations, while sampling every 1000 generations. We used a constant growth model for the coalescent prior and constrained our analyses to the monophyletic groups present within gray jays. Lognormal distributions for substitution model parameters were specified. Posterior output was analyzed in TRACER v. 1.5 and the burn-in was determined to be 10%. Because ND2 appears to evolve in birds at a slightly faster rate than cytochrome-b (Smith and Klicka, 2010), a substitution rate prior of 0.0125 (2.5% divergence per million years, my) was used in our strict clock analysis instead of the cytochrome-b rate of 0.01 (2%/my). Divergence estimates were inferred for a reduced dataset, which included three individuals from each geographically structured group, because of problems associated with model parameter variance across heterogeneous datasets (e.g., Guirer and Burbrink, 2008). Values for Time to Most Recent Common Ancestor (TMRCA) are based on the relaxed-clock tree. To visualize results with detailed topology, we used a strict-clock tree including all individuals in combination with the timescale from our chronogram based on a relaxed clock. Finally, to show structure within clades, we constructed median-joining haplotype networks using the program NETWORK v.4.5.1.6 (Bandelt et al., 1999).

2.3. Analyses of genetic diversity

We designated 15 populations using individuals from single localities or from pooling geographically proximate localities that are not separated by putative barriers to gene flow (range of $n = 6-22$, 180 individuals total). We calculated Fu's F_s , Tajima's D , nucleotide and haplotype diversity, as well as the proportion of private haplotypes for each clade and for each population using DNASP v. 5.10.01 (Rozas et al., 2009). We examined the amount of genetic structure

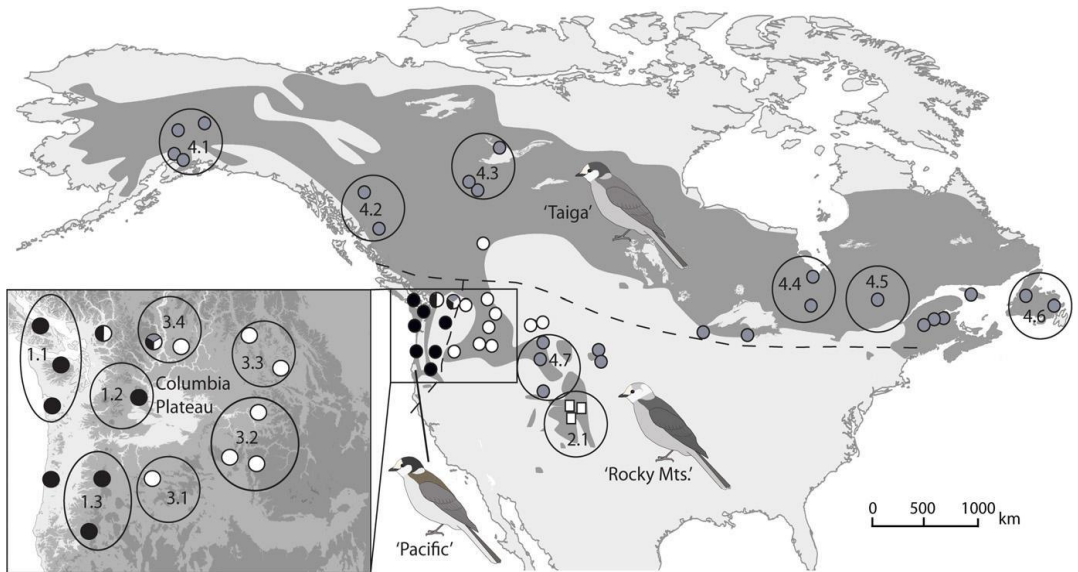


Fig. 1. Distribution of clades within the gray jay (*Perisoreus canadensis*). Black circles = Pacific, white squares = Colorado, white circles = Transcascades, gray circles = Boreal. Current distribution is indicated by gray shading. Populations designations are as follows: 1.1 Western Washington, 1.2 Central Washington, 1.3 Western Oregon, 2.1 Colorado, 3.1 Central Oregon, 3.2 Northeast Oregon, 3.3 Northern Idaho, 3.4 Northeast Washington, 4.1 Alaska, 4.2 British Columbia, 4.3 North-central Canada, 4.4 Ontario, 4.5 Québec, 4.6 Newfoundland, 4.7 Central Rocky Mountains. Three distinctive morphotypes (Sibley, 2000) are shown, with their distributions indicated by dashed lines.

Table 1. Gray jay sampling localities (this study, n = 50) with clade composition (B = Boreal, P = Pacific, C = Colorado, T = Transcascades, locality sample sizes, population assignment, and number of haplotypes per locality indicated.

Clade	State/province	County/location	n	Subpopulation
B	Alaska	Anchorage	4	Alaska
B	Alaska	Denali	1	Alaska
B	Alaska	Fairbanks North Star	4	Alaska
B	Alaska	Matanuska-Susitna	1	Alaska
B	British Columbia	Skeena	7	British Columbia
B	British Columbia	New Hazelton	3	British Columbia
B	Alberta	Mackenzie	5	NC Canada
B	Northern Territories	Yellowknife	7	NC Canada
B	Ontario	Kenora	2	Ontario
B	Ontario	Cochrane	6	Ontario
B	Québec	Côte-Nord	6	Québec
B	Québec	Saguenay	8	Québec
B	Newfoundland	Division 4	3	Newfoundland
B	Newfoundland	Division 5	4	Newfoundland
B	New Brunswick	Victoria	1	n/a
B	New Brunswick	Northumberland	1	n/a
B	Maine	Aroostook	3	n/a

B	Michigan	Dickinson	2	n/a
B	Minnesota	Pine	2	n/a
B	South Dakota	Custer	1	n/a
B	South Dakota	Lawrence	2	n/a
B	Utah	Uintah	3	n/a
B	Wyoming	Teton	4	n/a
B	Montana	Park	2	n/a
B	Washington	Okanogan	1	NE Washington
C	Colorado	Gunnison	3	n/a
C	Colorado	Clear Creek	1	n/a
C	Colorado	Summit	2	n/a
P	British Columbia	Vancouver Island	2	W Washington
P	Washington	Clallam	3	W Washington
P	Washington	Whatcom	9	W Washington
P	Washington	Okanogan	2	NE Washington
P	Washington	Jefferson	1	W Washington
P	Washington	Pacific	6	W Oregon
P	Washington	Kittitas	10	C Washington
P	Oregon	Linn	5	W Oregon
P	Oregon	Lincoln	6	W Oregon
P	Oregon	Jackson	5	W Oregon
T	Washington	Okanogan	6	NE Washington
T	Washington	Whatcom	1	W Washington
T	Washington	Ferry	3	NE Washington
T	Oregon	Wallowa	11	NE Oregon
T	Oregon	Wheeler	10	C Oregon
T	Oregon	Crook	7	C Oregon
T	Idaho	Boundary	7	W Montana
T	Idaho	Clearwater	8	C Idaho
T	Idaho	Valley	2	C Idaho
T	Idaho	Adams	2	C Idaho
T	Montana	Meagher	2	W Montana
T	Montana	Sanders	1	W Montana
T	Montana	Judith Basin	1	W Montana
T	Alberta	Pincher	1	W Montana

within and among populations of the three largest clades by performing an AMOVA (Analysis of Molecular Variance) in ARLEQUIN v. 3.11 (Excoffier and Schneider, 2005). In addition to calculating pairwise genetic differences (F_{st}) among populations, we performed a Mantel test to examine isolation by distance effects in intra-clade populations using ARLEQUIN. We tested for recent population expansion in ARLEQUIN and DnaSP using mismatch distributions with Harpending's raggedness index, Ramos-Onzins and Rozas' R_2 , and significant sum of squared deviations. Mismatch analyses also generate estimates of the time since expansion (s) and relative population sizes before (H_0) and after (H_1) population expansion occurs. We used the statistic s to calculate the time (t , in years) at which individual populations began expansion with $s = 2ut$, where $u = 2lk$ and l is the mutation rate (set at $1.25 \cdot 10^{-8}$) and k is the length of sequence (Rogers

and Harpending, 1992). Finally, we used MEGA 5.0 to calculate uncorrected sequence divergence between clades.

2.4. Species distribution modeling

We used MAXENT v. 3.3.3e (Phillips et al., 2006) to develop SDMs from recent sight records (Avian Knowledge Network, 2009) and 19 bioclimatic layers related mostly to temperature and precipitation (Worldclim Database v. 1.5, Hijmans et al., 2005). Duplicate data and records with poorly defined geographical references were removed and sight records were spatially homogenized to avoid discrepancies between areas of high and low observer visitation. We assumed that climatologic variables rather than biotic parameters defined gray jay niches, and that niches are conserved over time (Wiens and Graham, 2005). We modeled the current distribution of gray jays and then reconstructed the paleo-distribution of gray jays during the Last Glacial Maximum (LGM, 21,000 years ago). Paleo-SDMs were made based on two climatic models, the Community Climate System Model (CCSM, Collins et al., 2006) and the Model for Interdisciplinary Research on Climate (MIROC, Hasumi and Emori, 2004). We used a 25% random test percentage. In ArcGIS v. 9.2 (ESRI Inc., Redlands, CA) we classified probability of niche occurrence data into 10 equivalent classes of 0.1. Current-range models were then adjusted to overlap as much as possible with the known range of gray jays, by removing the five lowest probability classes. Paleo-models were constructed with the same number of probability classes.

3. Results

3.1. Phylogenetic analyses

Bayesian and ML analyses revealed four clades within *P. canadensis*. A widely distributed boreal clade (hereafter referred to as 'Boreal'), ranges from Newfoundland and northern New England through the northern United States and most of boreal Canada to western Alaska, and south to Wyoming and Utah and the Black Hills of South Dakota (Fig. 1). Three additional clades occupy distinct regions of western North America. These include a clade restricted to the southern Rocky Mountains ('Colorado'); another occurring from southwestern coastal Oregon north to the Olympic Peninsula and adjacent British Columbia ('Pacific'); and a clade distributed east of the Cascades in Oregon and Washington, ranging east to central Montana, and north at least to the southern Alberta Rocky Mountains ('Transcascades'). Topologies differed little between Bayesian and maximum likelihood analyses, with the exception of a few terminal nodes. Nodal support was 1.0 for interior nodes, whereas bootstrap values ranged more widely (63–100%). Individuals from the Pacific, Transcascades, and Boreal clades were sympatric in north-central Washington in Okanogan County and individuals from the Pacific and Transcascades clades co-occurred in

nearby Whatcom County. Haplotype network analyses revealed considerable phylogeographic structure in the Boreal clade (Fig. 2), but relatively little in the Pacific and Transcascades clades (see star-shaped networks, Fig. 2).

3.2. Population genetic analyses

Within our total sample of 205 individuals (Table 1), we detected 127 polymorphic sites, with an average of 24 nucleotide differences per individual. We found a total of 89 haplotypes (Table 2, Fig. 2), the most common of which was present in 49 individuals from the Transcascades clade. Haplotype diversity was variable, with high values for most Boreal clade populations (except for the peripheral Newfoundland and Central Rockies populations, 0.974 overall) and slightly lower values for populations in the Pacific clade (0.834). The Transcascades clade was relatively depauperate with respect to haplotype diversity (0.471) whereas our data based on a small sample size ($n = 6$) suggest that the Colorado clade may also have high haplotype diversity with five out of six individuals possessing unique haplotypes. Nucleotide diversity was consistently low across populations from all clades but lowest by an order of magnitude in most Transcascades clade populations. Summing nucleotide diversity across clades (Table 2) revealed that the Boreal clade harbored more than twice as much genetic diversity as the Pacific clade and an order of magnitude more than the Transcascades clade. Demography statistics provided conflicting results (Table 2). The mismatch statistics SSD and Harpending's raggedness index (r) were able to reject a hypothesis of exponential population growth for only three of our 14 populations. In contrast, the more conservative Fu's F_s and Tajima's D statistics (not shown) taken together identified a single population (NC Canada) with a genetic signature of recent population growth; and the R_2 statistic indicated a lack of population expansion throughout. Limited expansion was indicated by the recency of our estimates of time since expansion (t , Table 2). These range from only 4–19 thousand years ago (kya) and average slightly higher in the Boreal clade populations. Pairwise F_{st} comparisons (Table 3) indicated that significant population genetic structuring occurs within the widespread Boreal ($P = 0.0000$) clade but none was evident within either the Pacific ($P = 0.1476$) or Transcascades ($P = 0.0997$) clades.

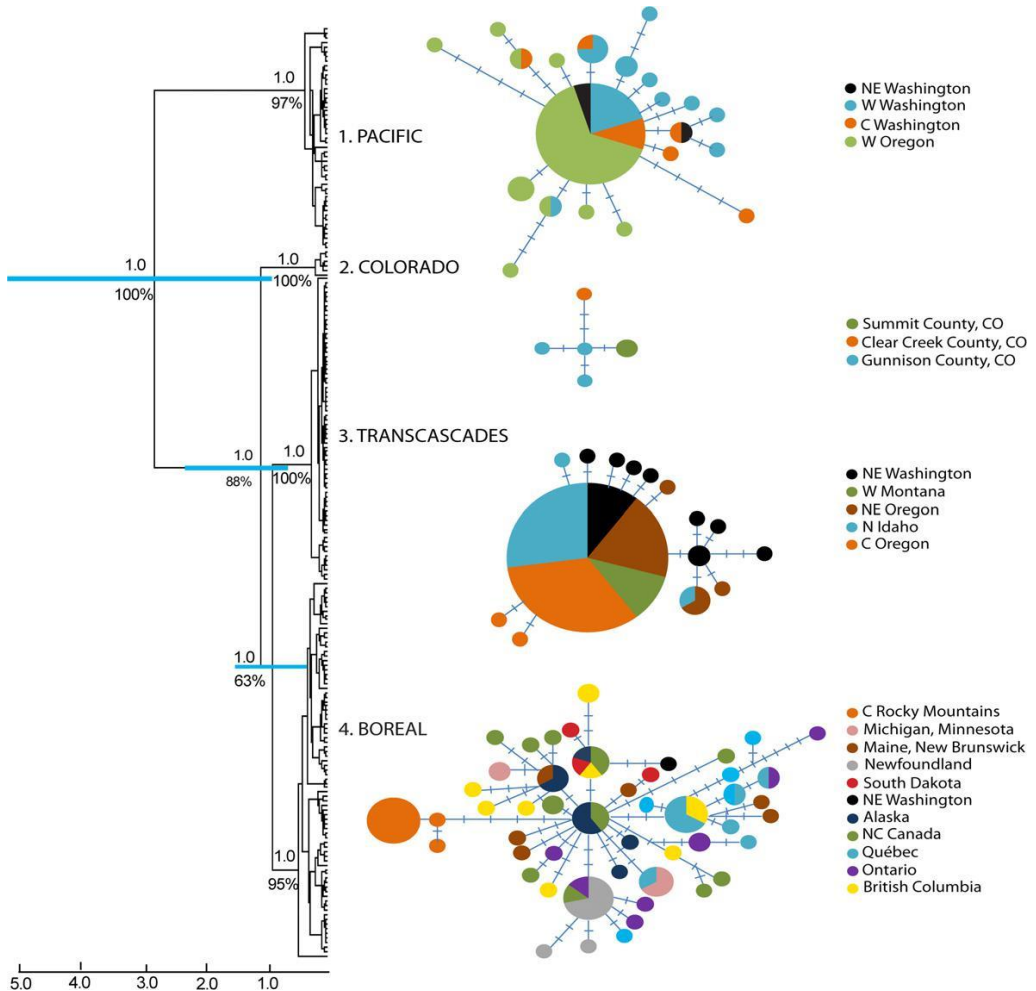


Fig. 2. Median-joining haplotype networks and Bayesian chronogram indicating structure within gray jays (*Perisoreus canadensis*). Values on top of nodes are posterior probabilities and lower values are maximum likelihood bootstrap percentages. The scale at the bottom indicates evolutionary time in millions of years. The blue bars in the tree represent 95% HPD intervals. For the Bayesian analysis, the TrN+I+C model of evolution (AIC = 7097.0960) and a relaxed clock were the most suitable fit to our data (ln L = 3078.24 vs. ln L = 2950.75, df = 136, $p < 0.01$).

The Newfoundland and Central Rockies populations differed significantly from all other sampled Boreal populations suggesting isolation during and since the LGM. Similarly, the Alaska sample differs from the eastern populations (Newfoundland, Québec, Ontario) but not from those populations more proximate geographically (British Colombia, north central Canada). Isolation by distance effects were detected in the Boreal clade as shown by a Mantel test ($R^2 = 0.5689$, $P = 0.006$), an unsurprising result given the apparent genetic structure and the geographic breadth of its distribution. Such effects were not detected in the more geographically restricted Pacific ($R^2 = 0.0231$, $P = 0.1032$) and Transcascades clades ($R^2 = 0.1024$, $P = 0.2257$). Uncorrected genetic

Table 2. Genetic diversity indices and mismatch statistics for clades and 14 populations within *Perisoreus canadensis*. Shown are: sample size (N), number of haplotypes (#H), number of private haplotypes (H_{priv}), haplotype diversity (H_{div}), nucleotide diversity (p), Fu's F_s (F_s), Ramos-Onsins and Rozas R_2 statistic (R_2), mismatch sum of squared deviation (SSD), Harpending's raggedness index (r), and time since population expansion (t). Significance is indicated by asterisks (= <0.02, = <0.01 for Fu's F_s and Ramos-Onsins and Rozas R_2 , = <0.05, = <0.01 for SSD, and Harpending's raggedness index). Numeric population designations refer to localities identified in Fig. 1.

Sample	N	#H	H_{priv}	H_{div}	p	F_s	R_2	SSD	r	t
Pacific	50	21		0.834	0.0018			0.0059	0.0790	10,374
1.1 W Washington	12	7	3	0.872	0.0017	2.784	0.1201	0.0061	0.0420	13,832
1.2 C Washington	10	8	5	0.956	0.0028	3.607	0.1158	0.3055	0.1812	4226
1.3 W Oregon	11	4	2	0.378	0.0010	0.477	0.2408	0.0245	0.1467	
Colorado	6	5		0.933	0.0018	2.263	0.1583			
Transcascades	66	15		0.471	0.0007					
3.1 C Oregon	17	2	0	0.118	0.0001	0.748	0.2353	0.3708	0.0370	11,527
3.2 NE Oregon	15	6	3	0.648	0.0011	2.451	0.0855	0.0126	0.0667	5864
3.3 N Idaho	15	3	2	0.257	0.0004	0.831	0.1805	0.0100	0.3964	7674
3.4 NE Washington	13	8	8	0.859	0.0015	4.567	0.0740	0.0092	0.1037	6532
Boreal	83	48		0.973	0.0040					
4.1 Alaska	10	5	2	0.867	0.0014	1.320	0.1606	0.0155	0.1333	6532
4.2 British Columbia	10	8	6	0.956	0.0038	2.629	0.0989	0.0231	0.0642	17,291
4.3 NC Canada	12	11	8	0.985	0.0035	7.040	0.0603	0.0031	0.0347	14,271
4.4 Ontario	8	7	5	0.964	0.0043	2.182	0.1589	0.0558	0.1327	17,291
4.5 Québec	14	10	7	0.923	0.0035	3.743	0.0796	0.0103	0.0264	18,828
4.6 Newfoundland	7	3	2	0.524	0.0008	0.237	0.2428	0.2500	0.1798	7300
4.7 C Rockies	9	3	3	0.417	0.0006	0.532	0.1848	0.2608	0.1690	6121

Table 3. Pairwise F_{st} values for all *Perisoreus* population samples, by clade, and AMOVA results. Asterisks indicate statistical significance after a Bonferroni correction. For AMOVA results we report the percentage of variation within populations for each clade (%Var_w) and F_{st} values.

Population no.:	1	2	3	4	5	6	%Var _w	F_{st}
Pacific							98.16	0.0184
1.1 W Wash.								
1.2 C Wash.	0.0455							
1.3 W Oregon	0.0187	0.0169						
Transcascades							97.33	0.0267
3.1 C Oregon								
3.2 NE Oregon	0.0306							
3.3 N Idaho	0.0276	0.0084						
3.4 NE Washington	0.8872	0.0046	0.0616					
Boreal							69.98	0.3002
4.1 Alaska								
4.2 British Columbia	0.0061							
4.3 NC Canada	0.0007	0.0030						
4.4 Ontario	0.1314	0.0601	0.0579					
4.5 Québec	0.2065	0.1285	0.1653	0.0498				
4.6 Newfoundland	0.6541	0.4582	0.4706	0.3503	0.4823			
4.7 C Rockies	0.7740	0.5816	0.6021	0.6245	0.6586	0.8863		

distances (p) between the Boreal, Transcascades and Colorado clades were similar (1.5–1.9%), whereas the distance from these to the Pacific clade was much greater (4.6–5.1%). TMRCA mean values indicate that the basal divergence between the Pacific clade and all other gray jays occurred around 3 million years ago (2.97 mya; 95% HPD: 0.89–5.52) and the more recently derived Colorado and Transcascades clades diverged around 1.10 mya (95% HPD: 0.34–2.16) and 0.72 mya (95% HPD: 0.19–1.38) respectively.

3.3. Species distribution modeling

Area Under Curve (AUC) values indicate the predictive capacity of a model, with values typically ranging from 0.5 (no capacity) to 1.0 (high capacity). For our data, AUC training values pertaining to the Receiving Operator Characteristic (ROC) curve produced by MAXENT were 0.970 and testing values were 0.967. Given the present distribution of gray jays (Fig. 3a, Strickland and Ouellet, 1993), the current-range model (Fig. 3b) described the northern boundaries of the distribution of the species relatively well. The paleo-SDM (Fig. 3c) identified areas of overlap with the current-range model that include southern Alaska, Pacific Canada, Newfoundland, and parts of the montane West. The most conspicuous departure from the current-range model was the predicted widespread occurrence in southeastern North America. CCSM-based models resulted in distributions nested entirely within those based on MIROC so we only show the latter.

4. Discussion

4.1. Evolutionary history of gray jays: broad sense

The gray jay is the sole New World representative of an other-wise Old World assemblage (see Bonaccorso and Peterson, 2007). Divergence from its putative sister taxon Siberian jay (*P. infautus*, Ericson et al., 2005) likely occurred when the ancestral gray jay clade crossed Beringia and became isolated in North America. Over time, it spread southward filling all accessible boreal habitats. While such habitats are widespread and homogeneous in distribution throughout higher latitudes, they become more restricted in size and less regularly spaced in lower latitude montane systems. These properties of low-latitude boreal habitats, in combination with cyclic climatic changes throughout the Pleistocene, have led to isolation and differentiation within the gray jay. The end result of this process is the genetic and geographic structure reflected in the four monophyletic clades that we identified. Precisely how these divergences occurred is not certain, but a temporal sequence is indicated by our data. We suggest that the phylogeographic pattern observed today is the result of a widespread and highly mobile northern ancestral form that has sequentially budded off daughter clades at the southern extent of its range. An initial Late Pliocene divergence resulted in an independently evolving Pacific clade. The geological and climatic events driving this event remain unclear. The timing of this divergence is consistent with vegetation records that suggest a recent uplift of the Cascade/Sierra Nevada chain 5–2 mya (Graham, 1999)

but this timing is also at odds with a more recent interpretation of the geological record that suggests a mid-Miocene origin for the northern Cascades (Mitchell and Montgomery, 2006). The subsequent origin of first the Colorado (1.1 mya) and then the Transcascades (0.7 mya) clades are presumably the result of distributional shifts and population fragmentation in response to successively later Pleistocene climatic oscillations. Evidence that this budding process is ongoing can be seen in our Central Rockies population (Figs. 1 and 3). Although embedded within the Boreal clade, this now isolated and southern-most Boreal population is reciprocally monophyletic and separated from all other clade members by a minimum of four mutational steps.

4.2. Gray jay phylogeography

Our LGM species distribution model (Fig. 3c) indicates a highly fragmented gray jay distribution at 21 kya. In the US Southwest, the Pacific clade evidently remained restricted to the Coast, Cascade, and Sierra Nevada ranges, presumably separated from the Transcascades clade by the Columbia Plateau (Fig. 3d). The SDM for the Transcascades clade suggests a southward retraction into the southern Great Basin; the model further suggests that the Colorado (southern Rockies) clade remained near where it occurs to-day, although likely at lower altitudes. That each of these clades remains monophyletic through one or more glacial cycles suggests that down-slope refugia (and population sizes) were small enough to avoid mixing during glacial periods. Our genetic data also provide an additional perspective regarding where these montane gray jays might have occurred during the LGM. In both the Pacific and Transcascades clades, the highest amounts of nucleotide diversity are observed in the northernmost sampled populations (Table 2), with the lowest nucleotide diversity in the south. This result is not consistent with a southern refugial population that expanded northward as the glacier retreated (as suggested by Fig. 3c and d), but rather it suggests that within these isolated montane systems, jays persisted in higher numbers in the north, perhaps in spruce forests along the southern margins of the ice sheet. A more complex LGM history is inferred for elements of the Boreal clade. Our LGM model suggests that a minimum of four widely separated LGM refugia were used (Fig. 3d), a history that is consistent with the geographic genetic structuring that is observed in the haplotype network (Fig. 2) and also reflected in genetic diversity indices (i.e. 'genetic footprints').

Genetic structuring (Fig. 2) that is centered on the Central Rockies, Newfoundland, and Québec population samples indicates a history of isolation in these areas. The central Rockies population at present is geographically centered between the Transcascades clade to the northwest and the Colorado clade to the southwest, and otherwise isolated from most of the Boreal clade by several hundred kilometers of unsuitable (mostly grassland) habitat. Our data suggest that this is a small population that may have persisted in this region over several glacial cycles.

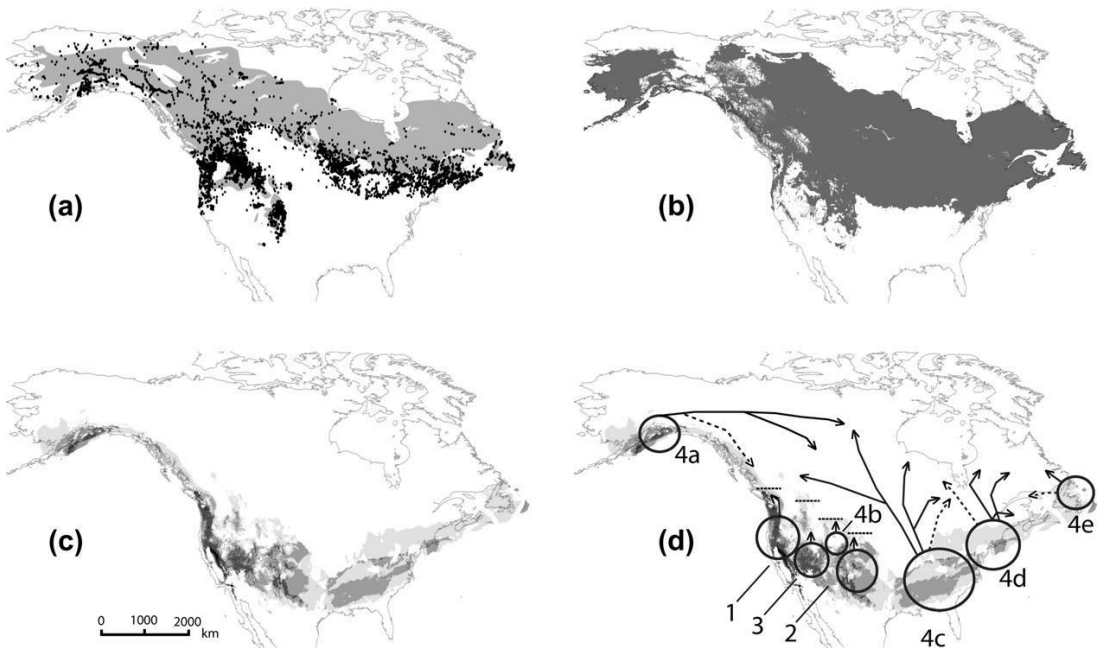


Fig. 3. Current and Last Glacial Maximum (LGM) species distribution model (SDM) of gray jay (*Perisoreus canadensis*). Map showing: (a) current-range indicated by gray shading and localities used (Avian Knowledge Network, 2009) in SDM reconstruction (b) current-range SDM, (c) putative LGM distribution of suitable niches according to the MIROC model, darker shades representing higher probabilities of suitable niche occurrence, and (d) a post-LGM re-colonization scenario that is consistent with our data. Solid circles represent refugia of 1. Pacific, 2. Colorado, 3. Transcascades, and 4. Boreal clades. Letters indicate gray jay refugia that overlap with known black (*Picea mariana*; Jaramillo-Correa et al., 2004) and white spruce (*Picea glauca*; Ritchie, 1987) refugia. Arrows represent putative re-colonization routes, dashed arrows indicate less certain routes. Horizontal dashed lines indicate the approximate northern limits of current distribution for western montane clades.

A similarly isolated population in the Black Hills of South Dakota was apparently colonized independently, and instead shares genetic affinities with birds from more northerly regions. A Newfoundland (area) refugium is also indicated by our data. Newfoundland birds form a distinct genetic cluster along with samples taken from British Columbia, north central Canada, and Ontario. That all seven Newfoundland samples are part of this clade indicates that the presence of additional widely scattered haplotypes is likely a consequence of incomplete clade sorting rather than ongoing, unidirectional gene flow. A third structured subclade comprised mostly of haplotypes from Québec (Fig. 2) is consistent with the presence of a third refugium in either the northeastern or southeastern United States. The LGM history of birds from the remaining boreal populations is less clear. Alaska and north central Canada haplotypes are located at the center of the Boreal clade network, suggesting the existence of a Beringia refugium. Additional evidence is provided by F_{st} values from Alaska and British Columbia samples that differ significantly from all sampled eastern populations, although this could also reflect isolation by distance. We note that the Alaska and British Columbia samples appear to have had very different evolutionary histories. Alaska has relatively lower nucleotide and haplotype diversity with only two private alleles. In

contrast, British Columbia has extremely high genetic diversity, six private alleles, and lacks the genetic structure observed in the Alaska population. In terms of genetic diversity indices, the British Columbia population is similar to the other Canadian interior populations sampled (NC Canada, Ontario, Québec) in having a high number of private alleles, high nucleotide diversity, and high haplotype diversity. We suggest that the high levels of genetic diversity seen in jays at high latitude localities are a consequence of the admixture of divergent clades colonizing these regions from multiple isolated refugia (Fry and Zink, 1998; Petit et al., 2003) that probably include Beringia, Maritime Canada (Grand Banks, Newfoundland), and larger and genetically diverse southern populations both east and west of the Appalachians. The existence of the latter refugia are suggested by our paleo-SDM and fossil evidence indicating that gray jays occurred at least as far south as Tennessee during the LGM (Parmalee and Klipper, 1982). Our data suggest a caveat to other phylogeographers when interpreting the significance of private haplotypes. High numbers of private haplotypes are usually considered signatures of stable and isolated populations, and such populations are often thought to be source populations or occupants of former refugial areas (Slatkin, 1985; Provan and Bennett, 2008). Our four interior high latitude populations (Fig. 1, Nos. 4.2–4.5) all possess high numbers of private haplotypes and high genetic diversity, yet all four localities were certainly under ice sheets during the LGM. The extant genetic diversity observed necessarily came from elsewhere. This anomaly may be explained by insufficient geographic sampling. Although we consider our geographic coverage sufficient and population level sampling thorough, additional sampling may reveal that many of these putative private haplotypes are indeed present at low frequencies throughout the interior Boreal clade distribution.

4.3. Southern richness, northern purity

The ‘southern richness, northern purity’ (Hewitt, 1996, 1999; Taberlet et al., 1998) paradigm has often been invoked for organisms that show a distinct pattern of high genetic diversity in the southern part of their distributions and genetically depauperate populations in the north. This pattern is expected under the pioneer model of post-glacial colonization (Nichols and Hewitt, 1994) in which high diversity in southern samples is attributed to greater population stability nearer refugial areas and lower northern diversity is due to a rapid re-colonization of formerly glaciated areas by small numbers of long-distance leading edge dispersers. This model does not appear to hold for gray jays. In this system, mtDNA variation is ‘upside down’, with variation highest in the widespread northern Boreal clade and distinctly lower in most (but not all, Table 2) southern clades. High genetic variation in the Boreal clade is better explained by the phalanx model (Nichols and Hewitt, 1994) of colonization whereby ancestral population variability is retained due to a slow northward spread by a large southern refugial population in a broad front. The already high genetic variation in the northward moving front was then enhanced by the addition of unique haplotypes from additional cryptic northern refugia. To explain the slowed pace of colonization in the phalanx model, a barrier (preventing more rapid colonization) is often invoked (Runck and Cook, 2005). The ice sheets themselves likely slowed colonization in this

way. Spruce (*Picea* spp.) trees are a common denominator of most forests where gray jays presently occur (Del Hoyo et al., 2009) and late glacial pollen records suggest that both black and white spruce (*P. mariana*, *P. glauca*) were present just south of the ice when glacial retreat began (Ritchie, 1987). For both species, re-colonization of the current range was rapid (McLeod and MacDonald, 1997). Spruce pollen was present in north central Saskatchewan 13,100 (*mariana*) and 12,650 (*glauca*) ya, reaching the northern limits of the Northwest Territories around 10,000 – 9300 ya (*mariana* and *glauca*, respectively; McLeod and MacDonald, 1997). It is probable that gray jays in southern and eastern refugia tracked these habitats northward at this time. Rapid east to west colonization of greater boreal North America was facilitated for both spruce and gray jays by a habitat corridor located between the Cordilleran and Laurentide ice sheets (Pielou, 1991; Dyke et al., 2002). This ice-free zone was first occupied by spruce beginning around 10,000 ya (MacDonald and McLeod, 1996; Runck and Cook, 2005). Our estimates of time since population expansions began (t, Table 2) are consistent with this interpretation, averaging 17,000 years since expansion for the four interior boreal populations. In contrast, the different evolutionary histories of the gray jay populations in the southern montane clades are reflected in shorter time estimates, averaging 8600 years.

4.4. Comparative phylogeography of boreal taxa

The close relationship between gray jays and coniferous habitats dominated by spruce is revealed by comparison of putative refugial areas. Molecular studies of black spruce recovered four putative LGM refugia that align remarkably well with those proposed for Boreal clade jays including two large southern (US) populations located east and west of the Appalachian range and a probable northeastern (Maritime Canadian provinces) refuge (Jaramillo-Correa et al., 2004). A northeastern coastal refugium was also proposed for another boreal conifer, the Jack pine (*Pinus banksiana*; Godbout et al., 2010). An additional proposed black spruce refugium in the Pacific Northwest is consistent with that proposed for the Pacific clade of gray jays. White spruce also occupied distinct glacial refugia east and west of the Appalachians (De Lafontaine et al., 2010) leading to genetically distinct populations east and west of the Great Lakes, a pattern also suggested by our data. A third white spruce clade was inferred to have a Beringian refugium, a region also suggested as an LGM refugium for jays. Overall, the genetic patterns recovered for gray jays are identical to those that would be expected if they simply tracked spruce habitats throughout the last glacial cycle. Gray jays appear to have shared evolutionary histories with a number of additional co-distributed boreal forest species, as a number of common biogeographic patterns are beginning to emerge. Among the most prominent of these is a widespread trans-continental boreal clade, distinct from one or multiple lower latitude montane western clades. The pattern occurs in other temperate forest birds (e.g. pine grosbeak [Drovetski et al., 2010], fox sparrow [Zink, 1994]), and also in mammals (e.g. northern flying squirrel [*Glaucomys sabrinus*, Arbogast, 1999], black bear [*Ursus americanus*, Byun et al., 1997], and amphibians (e.g. wood frog [*Rana sylvatica*, Lee-Yaw et al., 2008])). In these studies, lower latitude western clades occur

in the Rocky Mountain, Cascade, and (or) Sierra Nevada ranges. In the case of multiple western clades, a geographic break is most often found between the Rocky Mountains and the Cascades/Sierra Nevada. Estimated divergence times between boreal and western montane clades in these studies fall within the mid-to-late Pleistocene (from 0.13 mya to 1.2 mya), a temporal window that also captures the estimated divergence times for the Colorado and Transcascade clades of gray jays. The population source for postglacial expansion of boreal forest taxa into Central Canada and through to Alaska is typically refugia in southeastern North America, whereas descendants of western refugial clades generally remain restricted west of the Rocky Mountains with northern expansions (if they occur at all) limited to the Pacific coast west of the Coast Ranges (Runck and Cook, 2005, and references therein). Commonly associated with this pattern is the presence of a suture zone in southern British Columbia (Remington, 1968, but see Swenson and Howard, 2004) where many plant and animal taxa that have north–south discontinuities (including gray jays) come into secondary contact (Soltis et al., 1997; Maroja et al., 2007). Our insufficient sampling in this region does not allow us to accurately infer current levels of gene exchange between gray jay clades across this zone. However, the current distribution of jays, the lack of apparent physical barriers, and clinal morphological variation all suggest that a zone of introgression is likely to be found in south central British Columbia and southwestern Alberta. Another pattern emerging from multiple phylogeographic studies is the Canadian Maritime Provinces as a source of biodiversity. Along with the conifers previously mentioned, this region harbored refugial populations of wood frog (Lee-Yaw et al., 2008), flying squirrel (*Glaucomys volans*, Petersen and Stewart, 2006) and Maritime ringlet butterfly (*Coenonympha nipisiquit*, Sei and Porter, 2007). A LGM refugium near Newfoundland has also been suggested for several bird species, including song sparrow (*Melospiza melodia*, Fry and Zink, 1998), and boreal chickadee (*Poecile hudsonicus*, Gill et al., 1993).

4.5. Gray jays, refugia, and the LGM

The glacial refugium hypothesis has dominated studies of ice age biogeography for some time (Stewart et al., 2010). An abundance of paleontological and palynological evidence supports the traditional view that many temperate taxa persisted through glacial maxima in lower latitude refugia where climatic conditions were less extreme (Provan and Bennett, 2008). Indeed, our analyses suggest that gray jays have followed this pattern. Recent advances in phylogeographic and population genetic methods have led to fresh insights and much new discussion regarding LGM refugia. The new evidence indicates that in addition to the traditional southern refugium for temperate zone species, a multitude of other refugial areas were used (Hewitt, 2004; Provan and Bennett, 2008, see review by Stewart et al., 2010), including northern polar and cryptic southern refugia for cold-adapted species, and for temperate species, several cryptic northern refugia (Stewart et al., 2010; ‘microrefugia’ of Rull, 2009) in ice-free north-temperate coastal areas. An additional layer of complexity is added when considering sedentary taxa with montane distributions where distinctions are made between nunatak, peripheral, and lowland refugia

(Holderegger and Thiel-Egenter, 2009). The gray jay response to the recent glacial maximum makes an interesting test case for many of these newly proposed ideas. Our data indicate differential regional responses to glacial advances and these responses have led to genetic diversification and differences in levels of genetic variation. Since reciprocal monophyly has been retained through one or more glacial cycles for each montane gray jay clade, glacial populations probably never became large and likely were restricted to nearby down-slope regions ('peripheral refugia' of Holderegger and Thiel-Egenter, 2009) rather than spreading out in larger and more widespread 'lowland refugia' (Holderegger and Thiel-Egenter, 2009). In contrast, most gray jays from the lower elevation trans-Canadian portion of their distribution appeared to follow the general pattern established for many temperate taxa (Hewitt, 2004) in moving south to avoid harsh glacial conditions ('southern refugia' of Stewart et al., 2010) followed by re-colonization of northern regions. Some gray jays also survived in high latitude coastal refugium ('cryptic northern refugia' of Stewart et al., 2010). Our data provide a signature of such a refuge in Newfoundland and also suggest that one was also located on the west coast, likely in Beringia. This range of regional responses to glacial advances has led to the distribution of genetic diversity observed today with the counter-intuitive net result of higher genetic diversity at higher latitudes and lower genetic diversity in the south. Southern populations, constrained to isolated mountainous regions, necessarily have a smaller region to occupy and consequently reduced population sizes accompanied by reduced genetic variability. The much higher levels seen across most of the high latitude boreal distribution are presumably the result of genetic structuring occurring in multiple refugial populations during glacial periods (including a presumably large southern refuge) interconnected by high levels gene flow during interglacials.

Acknowledgments

We thank Brian T. Smith, Robert M. Zink, and Robert W. Bryson Jr. for critically reviewing our manuscript, as well as for analytical assistance and thoughtful discussions. Kevin Winker and an anonymous reviewer also provided many helpful comments. We thank Jaime A. Chaves for assistance with gray jay illustrations. Garth M. Spellman at Black Hills State University and Sharon M. Birks and Sievert Rohwer of the Burke Museum of Natural History at the University of Washington were most helpful in providing specimens. This work was funded in part by NSF DEB 0815057 (to J.K.) and the Barrick Museum Foundation.