Cryptic genetic diversity and cytonuclear discordance characterize contact among Canada Jay (*Perisoreus canadensis*) morphotypes in western North America

Brendan. A. Graham[1], Carla Cicero[2], Dan Strickland[3], John G. Woods[4], Howard Coneybeare[5], Kimberly M. Dohms[6], Ildiko Szabo[7], and Theresa M. Burg[1]

1 University of Lethbridge, Department of Biological Sciences, 4401 University Dr. W, Lethbridge AB, T1K 3M4

2 Museum of Vertebrate Zoology 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720-3160

3 1063 Oxtongue Lake Rd, Dwight, ON, P0A 1H0

4 4210 Frederick Road, Armstrong, BC, V0E 1B4

5 1221 23rd Ave SW, Salmon Arm, BC, V1E 0A9

6 Canadian Wildlife Services, Environment and Climate Change Canada, 5421 Robertson Rd, Delta, BC, V4K 3N2

7 Beaty Biodiversity Museum, University of British Columbia, 2212 Main Mall, Vancouver, BC, V6T 1Z4

Cryptic genetic diversity and cytonuclear discordance characterize contact among Canada Jay (*Perisoreus canadensis*) morphotypes in western North America

Abstract

Three distinct Canada Jay (*Perisoreus canadensis*) morphotypes with easily recognizable plumage traits come into contact in western North America. Recent work demonstrated high genetic structure across the species' range; however, patterns of genetic variation in these contact zones remain unknown. We categorized 605 individuals into one of three morphotypes (Pacific, Rocky Mountain, and Boreal) based on plumage, and genotyped individuals at the mtDNA control region and 12 microsatellite loci to assess the extent of hybridization between morphotypes. Our data showed cryptic genetic diversity and high cytonuclear discordance among morphotypes within contact zones, which is likely the result of recent and historical admixture. The distributions of the Boreal and Pacific morphotypes each showed a strong association with a single, distinct genetic group, whereas the Rocky Mountain morphotype exhibited higher genetic diversity and was associated with multiple genotypes. Our analyses show the importance of considering both plumage and genetic traits when examining contact zones between closely related taxa. Finally the data presented in this study reaffirm that the Pacific morphotype is distinct from the Boreal and Rocky Mountain morphotypes based on genetic, phenotypic, and ecological data, indicating that the Pacific morphotype should be re-elevated to a full species.

Introduction

Many morphologically similar species occupy the same geographic area and have similar ecological requirements, thereby making it difficult to distinguish one species from another (Brown *et al.*, 2007; Elmer *et al.*, 2007). For these reasons biodiversity may be cryptic and underestimated. Technological advancements have improved our ability to quantify and monitor biodiversity (e.g. DNA barcoding, Hebert *et al.*, 2003), which is now thought to be much greater than originally estimated (Mora *et al.*, 2011; Pimm *et al.*, 2014). These technological advancements not only improve our ability to study, monitor, and quantify biodiversity.

Although our ability to monitor and quantify biodiversity has improved; determining whether two similar taxa qualify as "species" (i.e. are judged to be "sufficiently" isolated from each other reproductively) is complicated because many taxon pairs are capable, to varying extents, of interbreeding and producing viable hybrids when they come into contact in zones of sympatry (Bell, 1996; Bradbury *et al.*, 2014; Toews *et al.*, 2016). Hybridization is currently viewed as an important component in evolution, promoting genetic diversification (Smith *et al.*, 2003; Capblancq *et al.*, 2015; Kagawa & Takimoto, 2018). Many traits, including behaviour, ecology, life history, and morphology influence the potential for hybridization; therefore, examining areas where two or more morphologically similar taxa come into contact with each other will provide greater insight into factors driving evolution and speciation.

Multiple geographic regions exist where large numbers of closely related species come into contact (Swenson & Howard, 2005). In North America, many of these contact

zones are found in the west where several major mountain ranges, habitat breaks, and other geographic barriers reduce gene flow (Swenson & Howard, 2005). In addition to present-day barriers, many historical events have helped shape biodiversity and ecological patterns. In particular, Pleistocene glaciations, with alternating cycles of glacial and interglacial periods that began approximately 2.5 million years ago, had a profound effect on biodiversity. Isolation during these periods as well as post-glacial colonization patterns have shaped many of the observed phylogeographic patterns, and these patterns are consistent across taxa (Weir & Schluter, 2004). Thus, many sister species exhibit morphological or plumage differences that may be indicative of reproductive isolation (Schluter, 2009; Grossen *et al.*, 2016).

The Canada Jay (*Perisoreus canadensis*) is a widely distributed resident corvid found in the boreal and subalpine forests of North America (Strickland & Ouellet, 2020). In addition to the continent-spanning (Alaska to Newfoundland) Boreal morphotype, two other distinct morphotypes occur in the western U.S. and Canada (Box 1, Figure 1). The Pacific form occurs from northern California through the Coastal and Cascades Ranges of Oregon and Washington to south-central British Columbia. Compared to the Boreal form, Pacific birds are pale-breasted, have conspicuous white-shafted feathers in their dorsal plumage, and have a more extensive dark crown patch. The Rocky Mountain morphotype, while more similar to the Boreal form than to the Pacific morphotype, is usually paler and with the dark nuchal patch failing to reach as far forward as the eyes, leaving individuals in this form with a notably "white-headed" appearance. Rocky Mountain morphotypes range from southeastern and central British Columbia to

increasingly isolated and high-elevation Rocky Mountain locations as far south as New Mexico and Arizona. The three morphotypes show putative evidence of hybridization where they meet and sometimes overlap in southern and central British Columbia (i.e. birds with diagnostic plumage traits of two or more morphotypes).

The two western morphotypes were initially described as "varieties" of the Canada Jay (Ridgeway, 1873), but the Pacific morphotype was later proposed as a full species (Sharpe, 1877) and this was accepted by Ridgway (Ridgway, 1880) as *P. obscurus*, the "Oregon Jay" (Strickland, 2017; Strickland & Ouellet, 2020). This status was retained until 1944, when the putatively separate species was lumped back into *P. canadensis* (American Ornithologist Union, 1944). Two recent studies (van Els *et al.*, 2012; Dohms *et al.*, 2017) using analyses of mtDNA and microsatellite genetic variation, demonstrated population genetic structure with high diversity, four monophyletic groups based on mtDNA, and multiple contact zones. Because sampling within the contact zones was limited, questions remain about the genetic patterns within these areas.

We used genotypic and phenotypic data in conjunction with spatial distribution models to provide a comparative framework for examining morphotype variation within putative contact zones of Canada Jays. Specifically, we sought to examine the extent of admixture among morphotypes, whether morphotypes produce hybrids, and the geographic distribution of introgression. Given that little is known about the relationship between genetic and phenotypic patterns within this species, we also explored this question in greater detail.

Methods

Sampling

We collected blood, feather, and tissues samples from 605 individuals (Table S1) between 1990 and 2018 at 35 sampling sites (hereafter referred to as populations; Figure 1), including 334 samples used in prior studies (224 samples from Dohms *et al.*, 2017; 110 samples from van Els *et al.*, 2012). The remaining 271 samples were collected between 2012 and 2018 to increase the number of samples from western North America, focusing on areas near or within the putative contact zones. Of the 605 samples, 300 were collected from live birds and 305 from tissues vouchered by museum specimens (Table S1). A small blood sample (~50 µL) or tail feather was collected from live birds, and birds were banded with a unique aluminum numbered band to avoid re-sampling. We extracted DNA from all blood, feather, and tissue samples using a modified chelex reaction (Walsh *et al.*, 1991) as outlined by (Burg & Croxall, 2001).

Phenotype assignment

Analysis of plumage data for 199 museum skins confirmed that the three morphotypes are distinct from each other, and that several key plumage traits can be used to categorize individuals accurately (see supplemental methods); 93.4% of individuals were identified to the morphotype designated by the observers using discriminant function analysis following a cross-validated approach (see supplemental results). We then examined the phenotype for each of the 605 individuals and assigned them to one of the three morphotypes: Boreal, Pacific, or Rocky Mountain (details in

Box 1). In a small number of cases (n=15), we could not assign individuals to a specific morphotype because they possessed diagnostic plumage traits from at least two morphotypes; these birds were categorized as intergrades (examples of putative intergrades are shown in Figure S1). For live birds, we made notes on the appearance in the field and took photos of the crown, dorsal, and ventral colouration.

Spatial Distribution Modeling

Using photographs available from citizen science data platforms (eBird; www.ebird.org), we delineated the distribution of the three morphotypes in western North America to identify putative contact zones. We constructed separate spatial distribution models for each morphotype using georeferenced photos from eBird to estimate areas of suitable habitat. We reviewed over 7,000 photos taken in western North America that were available on October 28th 2019, and assigned birds to one of the three morphotypes; in many cases, multiple photos were available from the same area, and we reviewed all of these photos to ensure that multiple morphotypes were not present. Our final dataset contained 1,234 georeferenced points from eBird and 605 georeferenced points from our samples (Figure 1). Apparent intergrades (n=59) were identified and excluded from the ecological niche models (ENM) due to their intermediate morphology. Prior to modeling, we removed all duplicates and outliers for each morphotype dataset. We used the thinning function with a distance of 5 km in the R package Wallace (version 1.06; Kass *et al.*, 2018) to reduce any sampling bias. Our final

dataset included 156, 152, and 193 georeferenced points for the Boreal, Pacific, and Rocky Mountain morphotypes, respectively.

The settings in our model followed those outlined by Dohms *et al.* (2017). All three models were constructed with 10 BIOCLIM variables (annual mean temperature, mean diurnal range, isothermality, temperature seasonality, mean temperature of wettest quarter, annual precipitation, precipitation of driest month, precipitation seasonality, precipitation of warmest quarter, precipitation of coldest quarter), and 25% of our points were used to train the models while the remaining points were used to test the model.

Microsatellite and mtDNA genotyping

We genotyped individuals at 12 microsatellite loci developed for other avian species. Seven of the 12 primer sets (ApCo 30, ApCo37, ApCo 40, ApCo 41, ApCo 91, (Stenzler & Fitzpatrick, 2002); Ck.2 A5 A, (Tarr & Fleischer, 1998); MJG1, (Li *et al.*, 1997)) were used in a previous study of Canada Jays (Dohms *et al.*, 2017), while five additional primer sets (AIAAAAG13, (Delaney & Wayne, 2005); LTML8, (McDonald & Potts, 1994); Pdo5, (Otter *et al.*, 1998); PJGATA2, (Busch *et al.*, 2009); Lox 1, (Bensch *et al.*, 1997)) were added for this study. PCR techniques and thermocycler conditions followed those outlined in Dohms *et al.* (2017); for the five new primer sets, T_{A1} of 50°C and T_{A2} of 52°C for the two-step annealing process.

MtDNA

We genotyped the mtDNA control region for all 605 individuals. For the purpose of this study, we were interested in which individuals were assigned to one of the four-mtDNA clades (Pacific, Intermountain West, Boreal, and southwest Rockies) identified by van Els *et al.* (2012) and Dohms *et al.* (2017). Using the sequences from Dohms *et al.* (2017), we identified diagnostic SNPs and designed separate primers for each clade (Table S2) to amplify individuals based on the diagnostic sites. PCRs were conducted in 10 μL reactions with 5× GoTaq Goflexi Buffer, 2 mM MgCl$_2$, 0.2 μM dNTP, 1 μM of each forward and reverse primer, and 0.5 U of taq. Thermocycler conditions were the same as Dohms *et al.* (2017), with the exception of a 60°C annealing temperature. We ran PCR products on a 0.8% agarose gel to confirm amplification. To confirm that the newly designed primer sets identified the correct mtDNA lineage, we sequenced a subset of individuals in addition to screening previously sequenced samples from Dohms *et al.* (2017). Sequencing was completed at NanuQ (Genome Quebec).

Population structure analyses

We tested the 12 microsatellite loci for deviations from Hardy-Weinberg equilibrium and linkage disequilibrium using GENEPOP 4.7 (Raymond & Rousset, 1995). We corrected for multiple tests using the sequential Bonferroni method (Rice, 1989). To quantify and compare genetic diversity among populations, we calculated allelic richness using the heirfstat package in R, and both observed and unbiased expected heterozygosity in GenAlEx 6.5 (Peakall & Smouse, 2012). We compared the three

genetic diversity measurements among populations using a Kruskal-Wallis test in Past 3.0 (Hammer *et al.*, 2001).

We used STRUCTURE v2.3.4 (Pritchard *et al.*, 2000) to determine the number of clusters (K) across western North America based on the microsatellite data. We ran 10 iterations for each K (K = 1-6) with a burn-in of 25,000 chains, followed by 50,000 replicates, and did not include loc priors. To determine how many genetic clusters were present across our study area, we estimated K by calculating ΔK (Evanno *et al.*, 2005) and examining posterior probabilities as recommended by Pritchard *et al.* (2000). Following our initial analyses, we examined hierarchical genetic structure within genetic clusters using the same settings for K = 1 to 5.

We combined the mtDNA and microsatellite genotyping data to examine admixture between morphotypes. We considered an individual to be admixed if there was a mismatch between its mtDNA and nuclear genotype (e.g. an individual assigned to the 'Boreal' mtDNA lineage, but clustered with the 'Pacific' microsatellite group). Lastly, we combined both genetic datasets with the plumage classification data to examine patterns of cytonuclear discordance within and among morphotypes (examples shown in Figure S2).

To examine the geographic distribution of plumage and genetic variation across western North America, we created a phenoscape using plumage data and three genoscapes (mtDNA, microsatellite, and mtDNA and microsatellite combined data). Both the phenoscape and genoscapes were created in R (version 3.6.3) using the packages and scripts outlined by Eric Anderson and the Bird Genoscape Project

(https://eriqande.github.io/make-a-BGP-map/Make-a-BGP-map-Notebook.nb.html). For the range map included in these analyses, we used the Canada Jay range map in - Strickland & Ouellet (2020), for which digital maps are provided by Birds of the World (2016). We used the kriging method to interpolate Q-values generated from our STRUCTURE analyses for both plumage and genetics. This analysis overlays group membership onto the distribution map with each colour representing one of the identified plumage or genetic clusters (Ruegg *et al.*, 2014). For our analyses of plumage, mtDNA, and microsatellites combined, we included hybrids as a separate cluster to identify the extent of putative contact zones.

Results

Phenotypic patterns

Of the 605 birds examined, 231 were Pacific morphotypes, 272 were Rocky Mountain morphotypes, and 87 were Boreal morphotypes (Table 1). In addition we identified 15 putative hybrids, all located within contact zones west of the Rockies and east of the Cascades (Figure 2).

Spatial distribution modeling

The mean AUC was high (≥0.89) for each of our three spatial distribution models (Figure S2). Combined, the models predicted suitable habitat for all three morphotypes throughout the study area, with extensive overlaps in the Pacific Northwest, southern British Columbia, and northern Washington. All three morphotypes had a high

probability (>0.54) of occurring within the areas of overlap, although Pacific morphotypes had the highest probability (0.77). Overall, the models predicted greater overlap of suitable habitat between Boreal and Rocky Mountain morphotypes, especially in interior British Columbia and throughout the Rocky Mountains in southern Alberta and Montana.

Genetic analyses

The twelve microsatellite loci used in this study were all polymorphic, with several loci exhibiting high levels of polymorphism (mean: 16.58 ± 4.08 alleles; range: 5-52 alleles; Table S3). Overall, 22 of 420 (5%) loci × population comparisons exhibited departures from Hardy-Weinberg equilibrium ($p<0.003$) following corrections for multiple tests. Five of the twenty-two (23%) loci × population comparisons involved the Wyoming population; however, we left this population in our analyses because the high frequency of departures from Hardy-Weinberg equilibrium is consistent with patterns expected by the Wahlund effect due to the presence of multiple mtDNA lineages and microsatellite genetic clusters within this population. Only three of 2,310 (<1%) loci × population combinations showed evidence of departures from linkage disequilibrium ($p<0.0005$).

Genetic diversity measures (allelic richness, observed heterozygosity, and unbiased expected heterozygosity; $\chi^2=15.89\text{-}16.41$, $p>0.99$) were comparable across all 35 populations. Genetic diversity was lowest on Vancouver Island and in several of the southeast Rocky Mountain populations.

Of the 605 individuals genotyped for the mtDNA CR, 66.3% had Intermountain West or Pacific haplotypes, 22.5% had Boreal haplotypes, and 11.2% had southeast Rockies haplotypes. The geographic distribution of CR haplotypes matched those reported by Dohms *et al*. (2017), although we observed higher rates of mtDNA introgression (Figure 2) than previously reported by both van Els *et al.* (2012) and Dohms *et al.* (2017). At the population level, 11 of the 35 populations were composed of individuals from two different mtDNA lineages, while two populations (Joffre Lakes British Columbia, and Okanogan County Washington) were composed of individuals from three mtDNA lineages (Figure 2).

Microsatellite genetic patterns exhibited a similar geographic distribution to mtDNA patterns, although we observed some differences (Figure 2). The Evanno method revealed K=2 (ΔK=173.37; not shown) and K=3 (ΔK=87.91; Figure 2A) as the top models. At K=2, individuals from west of the Cascades and Vancouver Island formed one genetic cluster, while the remaining individuals formed the second genetic cluster. At K=3, these same western individuals again formed one genetic cluster, while individuals in the southeast Rockies (southwest Montana, Colorado, Utah, New Mexico and parts of Wyoming) separated from all other individuals in Alberta, British Columbia, Idaho, Washington, and Oregon. Given that this third genetic cluster encompassed such a broad geographic area and was composed of birds with both Boreal and Rocky Mountain phenotypes, we examined hierarchical structure within this group. STRUCTURE recognized two distinct clusters (ΔK=97.56; Figure 2B) in this second analysis; individuals from interior British Columbia and the northern Canadian Rocky

Mountains clustered together, while individuals located in southern British Columbia and the northwestern United States formed the second genetic cluster, although some populations had individuals assigned to the other genetic cluster. Across both STRUCTURE analyses, a portion of individuals (26 of 605; 4.3%) exhibited similar Q-values for two or more clusters and therefore could not be assigned to a single genetic cluster; we considered these birds as admixed if the assignment to the highest group was <50%.

Overall STRUCTURE (K=3) showed high accuracy and power with respect to assigning birds to the correct genetic cluster. In areas of allopatry, where only one morphotype is found, STRUCTURE accurately assigned individuals to the correct genetic cluster; only 22 individuals (4.8%) were assigned to the incorrect genetic cluster (Figure S3). In areas of sympatry, where two or morphotypes are present, assignments with STRUCTURE were less accurate as 39 individuals (27%) were assigned to the incorrect cluster.

Phenotypic and genetic patterns

Both Boreal and Pacific morphotypes were strongly associated with the Boreal and Pacific mtDNA and microsatellite genetic clusters, respectively (Figure 3). The Rocky Mountain morphotype was associated with three mtDNA clades (Intermountain West, Boreal, and southeast Rocky Mountain) and two microsatellite genetic clusters (Intermountain West and southeast Rocky Mountain); overall there was strong geographic structuring of mtDNA and microsatellite patterns within the Rocky Mountain

morphotype (Figure 3). Among the 15 individuals identified as putative hybrids, the majority (87%) of birds had Intermountain West mtDNA haplotypes, yet clustered with the Boreal (40%) or Pacific (47%) genetic clusters based on microsatellites. Individuals with either Pacific or Boreal morphotypes exhibiting mismatches between phenotype and genotype were located within contact zones. Among the Rocky Mountain morphotype, 39% of individuals had Intermountain West mtDNA haplotypes and clustered with the Intermountain West microsatellite group, 24% had southeast Rocky Mountain mtDNA haplotypes and clustered with the southeast Rocky Mountain microsatellite group, while 36% exhibited cytonuclear discordance (Figures 2 and 4). Finally, all 15 individuals with hybrid or intergrade morphotypes exhibited cytonuclear discordance.

Plumage, mtDNA, and microsatellite data all showed similar patterns across geographic space (Figure 4), although there were some differences. Intergrade or hybrid morphotypes were identified in southern British Columbia and northern Washington (examples shown in Figure S4). Genetic contact zones for both mtDNA and microsatellite markers were broader, and we identified several areas in addition to southern British Columbia and northern Washington where genetic groups combine; these include northeastern Oregon, southern Alberta and northern Montana, as well as Wyoming. When we combined both mtDNA and microsatellite markers, we identified three areas where cytonuclear discordance is more prominent: southern British Columbia and northern Washington, Wyoming, and Utah. Overall, 222 individuals

showed mismatches between microsatellite and mtDNA patterns demonstrating that cytonuclear discordance is prominent in this species.

Discussion

The overall genetic diversity and variation matches the general patterns reported by van Els *et al*. (2012) and Dohms *et al*. (2017); the current study, however, reveals high instances of cytonuclear discordance in putative contact zones due to the more intensive sampling undertaken in this study. The geographic distribution of morphotypes, genetic lineages, and cytonuclear discordance highlights the cryptic genetic diversity present within this species, especially in the genetically diverse Rocky Mountain morphotype. Further, genetic patterns for Canada Jays are indicative of long-term isolation, followed by secondary contact, and subsequent hybridization among the three morphotypes. Given the observed genetic patterns, it appears that range extensions following Pleistocene glaciation facilitated interbreeding and hybridization among morphotypes, as has been observed for other species (Chenuil *et al.*, 2019).

The three morphotypes come into contact across a broad geographic area in southern British Columbia and northern Washington. This area of overlap is narrower than the area of overlap predicted by spatial distribution modeling. The observed introgression and size of this contact zone matches patterns found in other plant and animal species (Toews & Irwin, 2008; Gugger *et al.*, 2010; Chavez *et al.*, 2011; Natola & Burg, 2018). Within the Pacific genetic lineage and morphotype, we observed extremely low levels of cytonuclear discordance west of the Cascades (3.5%) compared to birds

north and east of the Cascades (36%). Cytonuclear discordance was most prominent in previously glaciated areas suggesting secondary contact and subsequent hybridization was facilitated by post-glacial expansion, a pattern that has been shown for other species (Chavez *et al.*, 2011; Hinojosa *et al.*, 2019).

In addition, we detected secondary contact and hybridization in the Rocky Mountains where genetic patterns are characteristic of both recent and historical hybridization between Boreal and Rocky Mountain morphotypes. Spatial distribution modeling predicted extensive habitat overlap between Boreal and Rocky Mountain morphotypes in Alberta and Montana, an area of ongoing hybridization for several other species (Shafer *et al.*, 2011; Natola & Burg, 2018). Genetic diversity is known to be underestimated in the Rocky Mountains (Milá *et al.*, 2011), and this is also the case for Canada Jays despite the clear transition between Boreal and Rocky Mountain morphotypes (Strickland & Ouellet, 2020). Although Canada Jays from southeastern British Columbia south to New Mexico and Arizona are exclusively Rocky Mountain morphotypes, three of the four mtDNA lineages and two microsatellite groups are present in this area. Birds to the north and south of the Wyoming Basin are quite distinct from each other. Birds south of the Wyoming Basin form a divergent lineage consistent with phylogeographic breaks for other species in the area (Galbreath *et al.*, 2010; Graham & Burg, 2012), whereas birds north of the Wyoming Basin have either Boreal or Intermountain West haplotypes. The presence and frequency of Boreal haplotypes this far south, as well as the presence of a population of the Boreal morphotype in a disjunct area of white spruce in South Dakota (van Els *et al.*, 2012),

suggests that the ranges of white spruce and the associated boreal morphotype of the Canada Jay formerly extended much farther south than at present. The high frequency of cytonuclear discordance in this area appear to have been shaped by historical hybridization and subsequent isolation of genetic groups, given the frequency and geographic distribution of genotypes in this area (e.g. individuals with southeast Rocky Mountain microsatellite genotypes and Boreal mtDNA haplotypes). The genetic patterns are further complicated by what appears to be secondary contact with birds that have Intermountain West genotypes, given that we observed admixture between Intermountain West genotypes and those found both north and south of the Wyoming Basin.

All three morphotypes are associated with coniferous forests, particularly of spruce (*Picea* spp.). The Boreal morphotype is not found apparently outside the ranges of black spruce (*P. mariana*) or white spruce (*P. glauca*), and the Rocky Mountain morphotype is similarly restricted to the range of Engelmann spruce (*P. engelmannii*). The Pacific morphotype, however, shows no consistent relationship with any single conifer. It is closely associated with shoreline Sitka spruce (*P. sitchensis*) in California, for example, but absent from similar stands in British Columbia and Alaska. It is common in mainland areas of subalpine habitat where Engelmann spruce is often also present, but it is equally common in subalpine areas on Vancouver Island where spruce is absent and forests are dominated by mountain hemlock (*Tsuga mertensia*), amabilis fir (*Abies amabilis*), and yellow cedar (*Callitropsis nootkensis*; Strickland & Ouellet, 2020). Patterns of secondary contact and hybridization among Canada Jay morphotypes and genetic

lineages appear to be most prevalent in previously glaciated areas characterized by transitional forest habitat. Pacific morphotypes come into contact and hybridize with the other two morphotypes in the northern Cascades where coastal and interior forests meet. This pattern also is seen in two parapatric species of tree squirrel (*Tamiasciurus hudsonicus* and *T. douglasii*) that hybridize in the same region (Arboghast *et al.*, 2001). Similarly, patterns of secondary contact and hybridization between Boreal and Rocky Mountain morphotypes are congruent with patterns of secondary contact and hybridization between Engelmann spruce and white spruce; these species hybridize throughout Montana, southwest Alberta, and interior British Columbia (Daubenmire, 1974). The close association between habitat and hybridization likely reflects prolonged isolation of Canada jays in multiple glacial refugia (van Els *et al*., 2012; Dohms *et al*., 2017), and post-glacial expansion following the recolonization by obligate forest species (Williams, 2003).

Incomplete lineage sorting is often used to explain patterns of cytonuclear discordance (Toews & Brelsford, 2012). Given the strong geographic population structure among Canada Jays based on microsatellite markers, it seems unlikely that the discordance between nuclear and mtDNA genetic patterns is due to incomplete lineage sorting. We sought to determine whether the three morphotypes are associated with particular genotypes. As has been found in other studies, the relationship between plumage and genetic variation is often complex. Plumage patterns are often independent of mtDNA or microsatellite genetic patterns (Ribot *et al.*, 2009; Lehtonen *et al.*, 2009) because plumage appears to be linked with a relatively small number of

nuclear genes (Toews *et al.*, 2016; Uy *et al.*, 2016; Funk & Taylor, 2019). Although

microsatellite and mtDNA markers may be poor predictors of avian plumage variation,

both Boreal and Pacific morphotypes were strongly associated with Boreal and Pacific

mtDNA haplotypes and microsatellite genetic clusters, respectively. These patterns

likely stem from prolonged historical isolation and subsequent genetic differentiation.

By comparison, Rocky Mountain morphotypes were associated with several genotypes,

probably reflecting a complex history of hybridization and isolation. Our analysis of

photos from eBird and the identification of several individuals with hybrid plumage

suggests that further analysis is warranted. In particular, analysis with next generation

sequencing techniques may help to uncover the genes associated with plumage

variation in Canada Jays (Funk & Taylor, 2019).

Our study emphasizes the complex phylogeographic history of Canada Jays in

western North America (van Els *et al*., 2012; Dohms *et al*., 2017). By sampling contact

zones more intensively and analyzing phenotypic and genetic data together, we were

better able to delineate the extent of hybridization among morphotypes and gain

further insight into the dynamics of these contact zones. Our work provides a

foundation for future work on the basis of morphotype variation and the genes that are

associated with plumage variation within this species. The observed gene flow and

hybridization among morphotypes in areas of contact is comparable to the patterns

observed in other species complexes (Bell, 1996; Smith *et al.*, 2003; Grossen *et al*., 2016;

Chavez *et al*., 2011; Toews & Irwin, 2008). Further, the use of multiple approaches to

examine Canada Jay contact zones incorporates the criteria outlined in Sites & Marshall (2004).

Moving forward we suggest a review of the Pacific morphotype's taxonomic status is warranted. A growing body of research on the genetics, morphology, and behavioural ecology of Canada Jay morphotypes (van Els *et al.*, 2012, Dohms *et al.*, 2017, Strickland & Ouellet, 2020, Strickland & Doucet, *in press*) suggests that re-elevation of the Pacific morphotype to full species status may be justified.

Ethics Statement

All methods and procedures were approved by the University of Lethbridge animal welfare committee (#1901).

Data Accessibility

All genotyping and plumage data will be stored on Dryad and mtDNA sequences will be uploaded to GenBank.

Competing Interests

The authors declare they have no conflicts of interest.

Authors' contribution statement

All authors assisted with field and museum data collection. BAG, CC, DS, JW, and TB designed the study. BAG and KMD performed the analyses, and all authors assisted with the writing and editing of the manuscript.

References

Arboghast BS, Browne RA, Weigl PD. 2001. Evolutionary genetics and Pleistocene biogeography of North American tree squirrels (*Tamiasciurus*). *Journal of Mammology* 82: 302–319.

Bell DA. 1996. Genetic differentiation, geographic variation and hybridization in gulls of the *Larus glaucescens-occidentalis* complex. *The Condor* 98: 527–546.

Bensch S, Price T, Kohn J. 1997. Isolation and characterization of microsatellite loci in a Phyloscopus warbler. *Molecular Ecology* 6: 91–92.

Birds of the World. 2016. Bird species distribution maps of the world. Version 6.0.

Bradbury IR, Bowman S, Borza T, Snelgrove PVR, Hutchings JA, Berg PR, Rodríguez-Ezpeleta N, Lighten J, Ruzzante DE, Taggart C, Bentzen P. 2014. Long distance linkage disequilibrium and limited hybridization suggest cryptic speciation in Atlantic cod. *PLoS ONE* 9:e106380.

Brown DM, Brenneman RA, Koepfli KP, Pollinger JP, Milá B, Georgiadis NJ, Louis EE, Grether GF, Jacobs DK, Wayne RK. 2007. Extensive population genetic structure in the giraffe. *BMC Biology* 5: 1–13.

Burg TM, Croxall JP. 2001. Global relationships amongst black-browed and grey-headed albatrosses: analysis of population structure using mitochondrial DNA and microsatellites. *Molecular Ecology* 10: 2647–2660.

Busch JD, Benford R, Pearson T, Palmer E, Balda RP, Keim P. 2009. Development of polymorphic tetranucleotide microsatellites for pinyon jays (*Gymnorhinus cyanocephalus*). *Conservation Genetics* 10: 689–691.

Capblancq T, Després L, Rioux D, Mavárez J. 2015. Hybridization promotes speciation in *Coenonympha* butterflies. *Molecular Ecology* 24: 6209–6222.

Chavez AS, Saltzberg CJ, Kenagy GJ. 2011. Genetic and phenotypic variation across a hybrid zone between ecologically divergent tree squirrels (*Tamiasciurus*). *Molecular Ecology* 20: 3350–3366.

Chenuil A, Cahill AE, Délémontey N, Du Salliant du Luc E, Fanton H. 2019. Problems and questions posed by cryptic species. A framework to guide future studies. In: Casetta E, Marques da Silva J, Vecchi D, eds *From assessing to conserving biodiversity. History, philosophy and theory of the life ciences*, vol 24. Springer, Cham.

Daubenmire R. 1974. Taxonomic and ecologie relationships between *Picea glauca* and *Picea engelmannii* . *Canadian Journal of Botany* 52: 1545–1560.

Delaney KS, Wayne RK. 2005. Adaptive units for conservation: population distinction and historic extinctions in the island scrub-jay. *Conservation Biology* 19: 523–533.

Dohms KM, Graham BA, Burg TM. 2017. Multilocus genetic analyses and spatial modeling reveal complex population structure and history in a widespread resident North American passerine (*Perisoreus canadensis*). *Ecology and Evolution* 7: 9869–9889.

Elmer KR, Dávila JA, Lougheed SC. 2007. Cryptic diversity and deep divergence in an upper Amazonian leaflitter frog, *Eleutherodactylus ockendeni*. *BMC Evolutionary Biology* 7: 1–14.

van Els P, Cicero C, Klicka J. 2012. High latitudes and high genetic diversity: Phylogeography of a widespread boreal bird, the gray jay (*Perisoreus canadensis*). *Molecular Phylogenetics and Evolution* 63: 456–465.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.

Funk ER, Taylor SA. 2019. High-throughput sequencing is revealing genetic associations with avian plumage color. *The Auk* 136: 1–7.

Galbreath KE, Hafner DJ, Zamudio KR, Agnew K. 2010. Isolation and introgression in the Intermountain West: Contrasting gene genealogies reveal the complex biogeographic history of the American pika (*Ochotona princeps*). *Journal of Biogeography* 37: 344–362.

Graham BA, Burg TM. 2012. Molecular markers provide insights into contemporary and historic gene flow for a non-migratory species. *Journal of Avian Biology* 43: 198–214.

Grossen C, Seneviratne SS, Croll D, Irwin DE. 2016. Strong reproductive isolation and narrow genomic tracts of differentiation among three woodpecker species in secondary contact. *Molecular Ecology* 25: 4247–4266.

Gugger PF, Sugita S, Cavender-Bares J. 2010. Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: testing hypotheses from the fossil record. *Molecular Ecology* 19: 1877–1897.

Hammer Ø, Harper DA, Ryan PD. 2001. Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4: 9.

Hebert PDN, Ratnasingham S, de Waard J. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B* 270: S96–S99.

Hinojosa JC, Koubínová D, Szenteczki MA, Pitteloud C, Dincă V, Alvarez N, Vila R. 2019. A mirage of cryptic species: Genomics uncover striking mitonuclear discordance in the

butterfly *Thymelicus sylvestris*. *Molecular Ecology* 28: 3857–3868.

Kagawa K, Takimoto G. 2018. Hybridization can promote adaptive radiation by means of transgressive segregation. *Ecology Letters* 21: 264–274.

Kass JM, Muscarella R, Vilela B, Aiello-lammens ME. 2018. Wallace : A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods in Ecology and Evolution* 9: 1151–1156.

Lehtonen PK, Laaksonen T, Artemyev A V, Belskii E, Both C, Bures S, Bushuev A V, Krams I, Moreno J, Mägi M, Nord A, Potti J, Ravussin PA, Sirkiä PM, Saetre GP, Primmer CR. 2009. Geographic patterns of genetic differentiation and plumage colour variation are different in the pied flycatcher (*Ficedula hypoleuca*). *Molecular Ecology* 18: 4463–4476.

Li SH, Huang Y, Brown J. 1997. Isolation of tetranucleotide microsatellites from the Mexican jay *Aphelcoma ultramarina*. *Molecular Ecology* 6: 499–501.

McDonald DB, Potts WK. 1994. Cooperative display and relatedness among males in a lek-mating bird. *Science (New York, N.Y.)* 266: 1030–1032.

Milá B, Toews DPL, Smith TB, Wayne RK. 2011. A cryptic contact zone between divergent mitochondrial DNA lineages in southwestern North America supports past introgressive hybridization in the yellow-rumped warbler complex (Aves: *Dendroica coronata*). *Biological Journal of the Linnean Society* 103: 696–706.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on earth and in the ocean? *PLoS Biology* 9: 1–8.

Natola L, Burg TM. 2018. Population genetics and speciation of yellow-bellied, red-naped, and red-breasted sapsuckers (*Sphyrapicus varius, S. nuchalis,* and *S. rube*r).

Journal of Heredity 109: 663–674.

Otter KA, Ratcliffe LM, Michaud D, Boag PT. 1998. Do female black-capped chickadeesprefer high ranking males as extra-pair partners. *Behavioral Ecology and Sociobiology* 43: 25–36.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics* 28: 2537.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 344: 1246752.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

Ribot RFH, Berg ML, Buchanan KL, Komdeur J, Joseph L, Bennett ATD. 2009. Does the ring species concept predict vocal variation in the crimson rosella, *Platycercus elegans*, complex? *Animal Behaviour* 77: 581–593.

Rice W. 1989. Analyzing tables of statistical tests. *Evolution* 43: 223–225.

Ridgway R. 1873. The birds of Colorado. *Bulletin of the Essex Institute* 5: 194–199.

Ridgway R. 1880. Revisions of nomenclature of certain North American birds. *Proceedings of the U. S. National Museum* 3: 4–5.

Ruegg KC, Anderson EC, Paxton KL, Apkenas V, Lao S, Siegel RB, Desante DF, Moore F, Smith TB. 2014. Mapping migration in a songbird using high-resolution genetic markers.

Molecular Ecology 23: 5726–5739.

Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* 323: 737–741.

Shafer ABA, Côté SD, Coltman DW. 2011. Hot spots of genetic diversity descended from multiple Pleistocene refugia in an alpine ungulate. *Evolution* 65: 125–138.

Sharpe RB. 1877. *Catalogue of the Birds in the British Museum* 3: 104–106.

Smith PF, Konings A, Kornfield I. 2003. Hybrid origin of a cichlid population in Lake Malawi: Implications for genetic variation and species diversity. *Molecular Ecology* 12: 2497–2504.

Stenzler LM, Fitzpatrick JW. 2002. Isolation of microsatellite loci in the Florida scrub-jay *Aphelcoma coerulescens*. *Molecular Ecology Notes* 2: 547–550.

Strickland D. 2017. How the Canada jay lost its name and why it matters. *Ontario Birds* 35: 2-16.

Strickland DS, Doucet SM. In Press. The bird that changes colour without molting: How the Wisakedjak tricked the taxonimist. *Canadian Journal of Zoology*.

Strickland DS, Ouellet H. 2020. Canada jay (*Perisoreus canadensis*), version 1.0. In: PG Rodewald, eds. *Birds of the World*. *Cornell Lab of Ornithology, Ithaca, NY, USA.*

Swenson NG, Howard DJ. 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist* 166: 581–591.

Tarr CL, Fleischer RC. 1998. Primers for polymorphic GT microsatellites isolated from the Mariana crow, *Corvus kabaryi*. *Molecular Ecology* 7: 253–255.

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear

discordance in animals. *Molecular Ecology* 21:3907-3930.

Toews DPL, Irwin DE. 2008. Cryptic speciation in a Holarctic passerine revealed by genetic and bioacoustic analyses. *Molecular Ecology* 17: 2691–2705.

Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, Lovette IJ. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology* 26: 2313–2318.

Merican Ornithologist Union. 1944. Nineteenth Supplement to the American Ornithologists' Union Check-list of North American Birds. *Auk* 61: 308–312.

Uy JAC, Cooper EA, Cutie S, Concannon R, Poelstra JW, Moyle RG, Filardi CE. 2016. Mutations in different pigmentation genes are associated with parallel melanism in island flycatchers. *Proceedings of the Royal Society B* 283: 20160731.

Walsh P, Metzger D, Higuchi R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10: 506–513.

Weir JT, Schluter D. 2004. Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271: 1881–1885.

Williams JW. 2003. Variations in tree cover in North America since the last glacial maximum. *Global and Planetary Change* 35: 1–23.

Location	Location ID	allopatric/ sympatric	n	Pacific	Rocky Mountain	Boreal	Hybrid
Pacific Coast							
1) Vancouver Island, BC	VI-BC	allopatric	18	18	-	-	-
2) Olympic Peninsula, WA	OP-WA	allopatric	24	24	-	-	-
3) coastal Oregon	CoOR	allopatric	8	8	-	-	-
4) southern Oregon	SOR	allopatric	19	19	-	-	-
Cascade Mountains		allopatric					
5) northwest Cascade Mountains, WA	NWC-WA	allopatric	19	19	-	-	-
6) eastern Cascade Mountains, WA	EC-WA	allopatric	14	14	-	-	-
7) Mt. Rainier, WA	MtR-WA	allopatric	36	36	-	-	-
8) eastern Cascade Mountains, OR	EC-OR	allopatric	15	15	-	-	-
9) southeast Cascade Mountains, OR	SEC-OR	allopatric	18	18	-	-	-
west of Rocky Mountains BC		allopatric					
10) Joffre Lakes, BC	JL-BC	sympatric	16	16	-	-	-
11) Princeton, BC	P-BC	sympatric	13	1	-	2	10
12) Nickel Plate Provincial Park, BC	NPBC	sympatric	7	-	4	-	3
13) Manning Park, BC	MP-BC	sympatric	11	10	-	-	1
west of Rocky Mountains WA							
14) northeast Cascades, WA	NE-Casc	sympatric	13	13	-	-	-
west slope of Rocky Mountains							
15) Okanogan County, WA	OK-WA	sympatric	29	20	1	7	1
16) east of Okanagan Valley, BC	EOK-BC	sympatric	12	-	12	-	-
17) east of Okanogan, WA	EOK-WA	sympatric	19	-	19	-	-
18) Pend Oreille, WA	PO-WA	sympatric	4	-	4	-	-
19) Wallowa, OR	Wa-OR	allopatric	13	-	13	-	-
20) Umatilla National Forest, OR	UNF-OR	allopatric	10	-	10	-	-
21) Ochoco National Forest, OR	ONF-OR	allopatric	10	-	10	-	-
22) northern Idaho	NID	allopatric	15	-	15	-	-
23) central Idaho	CID	allopatric	25	-	25	-	-
north and interior BC							
24) northwestern British Columbia	NWBC	allopatric	25	-	-	25	-
25) central British Columbia	CBC	allopatric	13	-	-	13	-
26) Clinton, BC	CL-BC	sympatric	12	-	-	12	-
Alberta and Montana							
27) central Alberta	CAB	allopatric	28	-	-	28	-
28) southern Alberta	SAB	sympatric	13	-	13	-	-
29) northwestern Montana	NWMT	allopatric	12	-	12	-	-
30) southwestern Montana	SWMT	allopatric	7	-	7	-	-
southeast Rocky Mountains							
31) Wyoming	WY	allopatric	34	-	34	-	-
32) Colorado	CO	allopatric	44	-	44	-	-

33) southeast Colorado	SWCO	allopatric	12	-	12	-	-
34) Utah	UT	allopatric	24	-	24	-	-
35) New Mexico	NM	allopatric	13	-	13	-	-

Table 1: Location and location codes (for data presented in Figure 2) of the 35 populations examined in this study; also included is the total sample size from each population (n) and the number of individuals assigned to each of the three morphotypes (Pacific, Rocky Mountain, Boreal) or identified as putative hybrids. Numbers for each population correspond to the sample sites shown in Figure S2.

Box 1: Plumage features that separate Canada Jays (*Perisoreus canadensis*) into three morphotypes: Pacific (a+d), Rocky Mountain (b+e), and Boreal (c+f). The Pacific morphotype (a+d) is characterized by a much more extensive dark area on the head extending to, and sometimes past, the eye, a correspondingly reduced white forehead, white feather shaft streaking on the back, a pale breast and belly with little to no contrast with the pale throat and neck collar, and limited (if any) white edging on the tips of the secondaries and rectrices. The defining characteristic of the Rocky Mountain morphotype (b+e) is a much reduced nuchal patch that does not reach as far forward as the eye thereby contributing to a "white-headed" appearance. Otherwise, it shares the following characteristics with the Boreal morphotype: a gray, unstreaked back, a variably gray breast and belly contrasting with the white throat and neck collar, and white tips (not mere edgings) on the secondaries and rectrices. The Boreal morphotype (c+f) is similar to the Rocky Mountain form, but has a dark nuchal patch that always reaches to the eye. Photographs in a) and c) provided by Dan Strickland, b) was provided by ebird. Drawings of morphotypes by Howard Coneybeare.

Figure 1: a) Georeferenced points of 1,839 ebird photos showing locations of Pacific (green triangles), Boreal (red circles), and Rocky Mountain (yellow squares) morphotypes, along with hybrids or intergrades (purple diamonds). A subset of these points (n=501) were used to conduct species distribution models in Figure S2. Dark shading represents forest habitat, while white shading represents non-forest habitat. b)

Map showing the sample sites examined in this study. Numbers correspond with those listed in Table 1.

 Figure 2: Histogram showing the phenotypic, mtDNA, and microsatellite genetic cluster assignments. A: Morphotype assignment (top row) to Pacific (dark green), Boreal (red), Rocky Mountain (yellow), or intergrade morphotypes (purple). Four mtDNA clades (middle row) were present: Pacific (green), Intermountain West (blue), Boreal (red), and southeast Rocky Mountain (light blue). Microsatellite genetic cluster assignments (bottom row) show assignment probabilities at K=3 as determined using STRUCTURE: Pacific (green), southeast Rocky Mountain (light blue), and Boreal/Intermountain West (pink). B) Microsatellite assignment probabilities at K=2 in our hierarchical genetic structure analysis of the Boreal/Intermountain West cluster: Intermountain West (blue) and Boreal (red). Geographic regions are listed along the top and population codes (Table 1) are listed on the bottom.

Figure 3: Distribution of mtDNA, microsatellite genetic cluster, and combined mtDNA and microsatellite genotypes among the three morphotypes and intergrade morphotypes. For the combined dataset, the first four categories (Pacific, Intermountain West-IMW, Boreal, and Southeast) represent individuals that showed congruent patterns between mtDNA and microsatellites. The fifth category (Cyto-ND) represents individuals that exhibited cyto-nuclear discordance between mtDNA and microsatellite markers. Percent is listed on the y-axis and genotype is listed along the x-axis. For each of the main morphotypes, the expected genetic group is marked with an asterisk.

Figure 4: Geographic distribution of plumage (a) and genetic (b-d) variation for Canada Jays (*Perisoreus canadensis*) in western North America. 4a shows the distribution of the three morphotypes: Pacific (green), Boreal (red), and Rocky Mountain (yellow); inset in 4a shows the distribution of hybrids / intergrades (purple) sampled in this study. Colours for each group are consistent between figures 4-d: Pacific (green), Boreal (red), Intermountain West (blue), and southeast - Rocky Mountains (light blue) genotypes. In d) putative hybrids represent individuals that showed cytonuclear discordance between mtDNA and microsatellite patterns (orange).