**Abstract**

Studying species interactions at hybrid zones allows biologists to understand the forces that promote speciation. Hybridization among *Sphyrapicus nuchalis*, *S. varius*, and *S. ruber* has long been acknowledged, and hybrid zones between *S. nuchalis/S. ruber* and *S. varius*/*S. ruber* have been characterized with both genetic and genomic data.Using a combination of next-generation genotyping by sequencing (GBS) and traditional genetic methods, we examined patterns of introgression in the poorly characterized *S. nuchalis/S. varius* contact zone. We found high introgression rates, with several early and many advanced generation hybrids along a 275 km stretch of Rocky Mountain foothill, pointing to a well-established hybrid zone with many hybrid individuals backcrossing with individuals from the parental species and each other. In the hybrid zone plumage coloration was a relatively poor indicator of overall genomic background, which could be attributed to the possible involvement of few large effect genes.

Keywords: Hybrid zone; introgression; Genotyping-by-Sequencing; red-naped sapsucker; yellow-bellied sapsucker

**Introduction**

Describing the process by which new species arise (speciation) is central to comprehending avian biodiversity, and speciation can be best studied by examining populations at intermediate stages of species divergence (Grant & Grant 1992). Populations are often geographically isolated, but individuals may interbreed, or hybridize, where they co-occur. Studying hybridization among such forms often reveals the extent of reproductive isolation between these populations and levels of divergence. Hybridization is richest and most informative in hybrid zones where the breeding ranges of two adjacent species overlap (Hewitt 1988, Short 1972), which are often called “natural laboratories” of speciation because they present many different combinations of parental alleles, much like biologists might create in a laboratory (Hewitt 1988). The populations may collapse into a hybrid swarm, they may remain distinct despite introgression, or they may be ephemeral or stable, recurring or novel, distributed broadly or restricted to a narrow zone (Grant & Grant 1992, Moore 1977, Short 1972). Describing hybrid zone dynamics helps us understand the evolutionary histories of species and predict their evolutionary trajectories (Grant & Grant 1992).

*Sphyrapicus* *varius*, *S. nuchalis*, and *S. ruber* are three species of North American woodpeckers that hybridize in sympatry. Hybridization within this species complex has long been of interest to biologists, who have studied the forces maintaining several hybrid zones and changes within them (Billerman 2016, Billerman et al. 2016, 2019, Howell 1952, Johnson & Johnson 1985, Natola & Burg 2018, Scott et al. 1976, Seneviratne et al. 2012, 2016, Trombino 1998). *Sphyrapicus ruber*/*S. nuchalis* and *S. ruber/S. varius* hybrid zoneshave been studied both behaviourally and genetically (Billerman et al. 2016, 2019, Cicero & Johnson 1995, Grossen et al. 2016, Johnson & Johnson 1985, Johnson & Zink 1983, Natola & Burg 2018, Seneviratne et al. 2016, 2012). Though these species hybridize, studies of mate choice in two hybrid zones show the forms mate assortatively, supporting their designations as separate species (Johnson & Johnson 1985, Seneviratne et al. 2012, 2016). Possible putative reproductive barriers between these species include different migratory strategies, breeding phenology, habitat preferences, and assortative mating (Billerman et al. 2016, Johnson & Johnson 1985, Natola & Burg 2018, Trombino 1998).

A *S. nuchalis/S. varius* hybrid zone exists in west-central Alberta (CAB), but it has not been well studied using genetic methods. Recent behavioural work (Jocelyn Hudon, unpublished data) and some genetic data (Natola & Burg 2018) document hybridization within this hybrid zone, but we know little about the extent of recent or past hybridization or genetic introgression between the two species. Since the *S. varius/S. ruber* and *S. nuchalis/S. ruber* hybrid zones are well characterized, describing the *S. varius/S. nuchalis* hybrid zone stands to greatly improve our understanding of reproductive interactions among these three species and the isolating barriers that assist their differentiation. Here we characterize genetically an area where the ranges of *S. nuchalis* and *S. varius* overlap and the two forms hybridize in order to delineate the geographical extent of the hybrid zone, quantify introgression rates and characteristics of the contact zone, and compare early vs late generation hybrids within the zone. Using traditional and next generation sequencing methods allows us to examine introgression patterns both across the genome and across the geographic range, which provides a broader context to investigate patterns of hybridization between *S. nuchalis* and *S. varius*.

**Methods**

*Sample acquisition*

We collected DNA samples from specimens in museum collections and birds caught with mist nets during the breeding season. We collected wild-caught samples from May to July to reduce the number of migrants caught. Birds were called in with playbacks and caught using 12 m mist nets. We took a small (<50 µL) sample of blood from the brachial vein, banded the birds, and took morphometric measurements and photographs. All birds were released on site and blood samples were stored in 99% ethanol. Museum specimens were selected from birds collected within the last 20 years during the breeding season to ensure data reflected contemporary patterns (Appendix 1). These specimens included material collected at the conclusion of a 15-year study of reproductive interactions between *S.* *varius* and *S. nuchalis* along the foothills of western Alberta between Strachan (53.4005˚ N, -117.8685˚ W) and the Porcupine Hills (49.0512˚ N, -113.9115˚ W), AB (Jocelyn Hudon, unpublished observations).

We examined the sapsuckers sampled for the extent of red on the nape, the degree of invasion of the malar stripe and cheek patch by red in males, the amount of red on the throat of females, and the amount and distribution of white markings on the back (male plumage shown in Shunk 2005). We classified individuals with any evidence of red on the nape, malar and throat, and limited amount of white on the back, that was not within the range expected for “pure” red-naped sapsuckers as hybrids.

*GBS methods*

GBS DNA extraction, processing

We extracted DNA using a standard phenol-chloroform extraction procedure and sent 63 samples (33 *S. nuchalis*, 18 *S. varius*, and 12 hybrids) to Cornell University’s Institute for Genomic Diversity (IGD) for GBS following Elshire et al. (2011) with the restriction enzyme *PstI*. We sent an additional 80 samples (39 *S. nuchalis*, 24 *S. varius*, and 17 hybrids) to the Genomic Sequencing and Analysis Facility (GSAF) at the University of Texas for double digest RADseq using the restriction enzyme pair *NlaIII* and *MluCI* following Peterson et al. (2012). In total, we sequenced 12 *S. varius* individuals from Saskatchewan (SK), five from northwest British Columbia (BC), seven from North Dakota (ND), two from North Carolina (NC), six individuals from Michigan (MI) and 11 from Alberta (AB). We also sequenced nine *S. nuchalis* individuals from Colorado (CO), 11 from northeast Oregon (OR), four from Montana (MT), six from Washington (WA), five from New Mexico (NM), six from South Dakota (SD), nine from Utah (UT), one from Wyoming (WY), 21 from AB. Additionally, we sequenced 28 hybrids from AB. Nine *S. nuchalis* individuals (two AB, three UT, two OR, and two CO) and one *S.* *varius* (AB) were sent to both facilities for a total of 143 samples, 60 collected within AB and 83 from outside AB.

SNP calling

Due to the lack of a reference genome for *Sphyrapicus*, we performed *de novo* SNP calling. For the Cornell dataset, we used the GBS UNEAK analysis pipeline version 3.0, which is an extension of the JAVA program TASSEL (Bradbury et al. 2007), to filter reads and call SNPs. Quality filtering removed any reads with incorrect, missing, or multiple restriction cut sites or barcodes. We then truncated reads to 64 bases and aligned them into identical sequence tags. We set a threshold of a minimum of five reads per tag for inclusion in the SNP calling process with the error tolerance rate set to 0.03 and a minimum minor allele frequency (MAF) of 0.05 for pairwise alignment identification of SNPs. We kept a maximum of one SNP per read to minimize linkage disequilibrium in downstream analyses.

We filtered, demultiplexed, and cleaned the UTexas data using the STACKS v 1.09 process\_radtags pipeline (J. Catchen et al., 2013; J. M. Catchen et al., 2011). We used the denovo\_map STACKS pipeline to identify SNPs *de novo* with both the number of reads required to create a stack and the number of mismatches allowed between loci set to four. We filtered UTexas SNPs for a MAF of 0.05 (as per the Cornell SNPs) using PLINK v 1.07 (Purcell et al. 2007). We kept only one SNP per read and SNPs with less than 20% missing genotype data across individuals for downstream genomic analyses.

Genomic analyses of hybrid zone individuals

Pairwise *F*ST was calculated using the software ARLEQUIN (Excoffier & Lischer 2010) for the Cornell dataset only, as SNP calling produced more SNPs in the Cornell dataset when compared to the UTexas dataset.

To determine the proportion of ancestry from each species in individuals from the hybrid zone, we used the program ADMIXTURE v 1.2.3 (Alexander et al., 2009). We ran ADMIXTURE on the two GBS datasets separately as they both contain different loci and different individuals. After the analyses, we combined the results to view the complete dataset together. We ran each ADMIXTURE analysis for K = 1-4, using a quasi-Newton algorithm for accelerated convergence (Zhou et al., 2011) and a 5-fold cross-validation. We used the default block relaxation algorithm to perform point estimation stopping the analyses when the change in the log-likelihood of point estimations between iterations increased by <0.0001. We determined the number of clusters that best fit the data by the K value with the lowest cross-validation error. We used the resulting ancestry coefficient values to define individuals as either *S. nuchalis* (Q ≤ 0.1), *S. varius* (Q ≥ 0.9), or admixed (0.1 < Q < 0.9) in accordance with Billerman et al. (2019).

Genomic structure of the hybrid zone

To determine the proportion of F1 and advanced generation hybrids within AB, we compared the hybrid index (HINDEX) score – determined by calculating the allele frequencies of the individual and comparing that to the allele frequencies of *a priori* defined parent populations – to the interspecific heterozygosity score for each individual within AB using the R package INTROGRESS v 1.22 (Gompert & Buerkle, 2010; Gompert & Buerkle, 2009; R Development Core Team, 2013). We defined parental individuals within AB by a hybrid index of either zero (*S. nuchalis*) or one (*S. varius*), with interspecific heterozygosities close to zero.

We plotted the interspecific heterozygosity against the hybrid index for each individual to distinguish pure individuals from F1 hybrids or advanced generation backcrosses. We used *S. varius* and *S. nuchalis* individuals from allopatric populations with high assignment of ancestry to their respective species in ADMIXTURE as *a priori* parental populations in the INTROGRESS analysis. To determine the genetic composition of individuals within the hybrid zone, we used all individuals located within AB (n = 60), regardless of their phenotypic identification. We also used the prepare.data function of INTROGRESS to calculate the allele frequency differential (δ) for each locus to identify loci with fixed differences – or loci with a low proportion of similar alleles – between the two species. We compared the results from INTROGRESS using only loci with δ = 1 to those using all loci.

*Traditional genetic marker methods*

DNA extraction, amplification, and sequencing

We selected a total of 207 samples from populations outside, near, and within the hybrid zone from each species, in addition to 43 individuals designated as phenotypic hybrids. We sampled fifteen *S. nuchalis* individuals from New Mexico (NM), along with 10 from Idaho (ID), and 41 from Alberta (AB). We sampled 57 *S. varius* individuals from AB, 14 from Saskatchewan (SK), 15 from Illinois (IL), and 13 from Nova Scotia/New Brunswick (NSNB). We extracted total genomic DNA from blood samples using a modified Chelex extraction (Walsh et al., 1991, Burg and Croxall 2001). Following extraction, we stored all samples at -20˚C.

We amplified a 370 bp segment of the α-enolase (Enol) nuclear gene in 17 individuals total from *S. nuchalis* and *S. varius* using the Enol8L731 and Enol9H912 primers (Table 1). The thermal cycling profile was one cycle of 120 s at 94 ˚C, 45 s at 54 ˚C, 60 s at 72 ˚C; 37 cycles of 30 s at 94 ˚C, 45 s at 54 ˚C, 60 s at 72 ˚C; one cycle of 300 s at 72 ˚C and 20 s at 4 ˚C. The 25 µL PCR reaction contained 5x Green GoTaq® Flexi buffer (Promega), 0.2 mM dNTP, 1 mM MgCl2, 0.4 µM primers Enol8L731 and Enol9H912, 0.5 U GoTaq® Flexi polymerase, and genomic DNA.

We amplified and sequenced a 450 bp region of the glyceraldehyde gene (GAPD) in a total of 11 individuals of both species using primers GAPD11L890 and GAPD12H950 (Table 1). The thermal cycling profile was similar to that used to amplify Enol, but with a 60 ˚C annealing temperature, and 0.8 mM MgCl2.

We sequenced 760 bp of an anonymous region in 21 individuals from both species using primers TP1F4 and TP1R5 (Table 1). The thermal cycling profile was similar to the profile used in Enol, but with a 48 ˚C annealing temperature, a 105 s extension time, 0.8 mM dNTP, and 2.5 mM MgCl2.

We sent successfully amplified samples to NanuQ sequencing service at McGill University, Montreal, Quebec for sequencing. We aligned sequences using MEGA v. 6 (Tamura et al. 2011).

SNP screening

We used the aligned Enol, GAPD, and anonymous nuclear marker sequences to identify SNPs. We detected a C/T SNP 213 bp from the 3’ end of the Enol8L731 primer. The GAPD sequences contained a 4 bp insertion/deletion that was associated with a CNC/ANG SNP 118 bp from the 3’ end of the GAPD11L890 primer. Sequences from the anonymous nuclear marker revealed a C/T SNP 84 bp from the 3’ end of the TP1F4 primer. We designed new primers for Enol and the anonymous nuclear marker to create cut sites associated with the SNPs to allow screening with restriction enzymes. We added M13 tags to the 5’ end either to allow screening on an acrylamide gel (Enol) or to increase size differences of digested products on an agarose gel (GAPD, anonymous nuclear marker).

We used a standard PCR protocol with a 10 µL reaction containing 0.1 mM dNTP, 0.4 µM primers, 0.25 U GoTaq® Flexi polymerase, genomic DNA, and varying amounts of MgCl2, 5x GoTaq® Flexi buffer (Promega), and the addition or omission of 0.04 µM fluorescent M13 tag (Table 2). Thermal cycling protocols were the same as GAPD, with different annealing temperatures (Table 2). We digested PCR products with restriction enzymes (Table 2) and a 1x buffer for a minimum of 3 hours at 37 ˚C.

We ran digested Enol products on a 6% acrylamide gel on the LI-COR 4300 DNA Analyzer and scored individuals as homozygous for T (187 bp) or C (41 bp and 146 bp), or heterozygous (41 bp, 146 bp, and 187 bp) for 187 individuals. We ran the GAPD and the anonymous nuclear marker restriction digests on a 3% agarose gel. We scored 174 individuals for the GAPD locus as CTC/deletion (469 bp), ATG/insertion (149 bp and 320 bp), or heterozygous (149 bp, 320 bp, and 469 bp). We scored the products at the anonymous nuclear marker as homozygous C (41 bp and 103 bp) or T (144 bp), or heterozygous (41 bp, 103 bp, and 144 bp) for 206 individuals.

To determine statistical significance of SNP variation between *S. nuchalis* and *S. varius*, between each species and hybrids, and between zones of allopatry and sympatry within each species, we used Fisher’s exact tests.

We ran genotype data at all three loci through STRUCTURE v. 2.3.4 (Pritchard et al. 2000) for ancestry assignment. We used a burn in of 10,000, MCMC length of 150,000, the loc priors setting and ran ten iterations for each K = 2-4. We used the lowest log-likelihood values to select optimal K for the data, and averaged Q values. We used the resulting ancestry coefficient values to define individuals as either *S. nuchalis* (Q ≤ 0.1), *S. varius* (Q ≥ 0.9), or admixed (0.1 < Q < 0.9) in accordance with Billerman et al. (2019).

**Results**

*GBS*

After filtering, the Cornell dataset contained 11,311 SNPs shared between *S. varius* and *S. nuchalis* while the UTexas dataset contained 1,638 such SNPs. The pairwise FST for the Cornell dataset was 0.26 (*p* < 0.05). ADMIXTURE identified K = 2 as having the lowest cross-validation error, and therefore the optimal number of clusters for both the UTexas and Cornell datasets.

The ADMIXTURE plot differentiated the individuals outside of AB into two distinct clusters – one for *S. varius* and one for *S. nuchalis* (Figure 1a). Individuals identified phenotypically as *S. nuchalis* within AB were more admixed than both *S. nuchalis* individuals outside the zone and *S. varius* within AB. *Sphyrapicus varius* within AB showed little or no admixture, with the exception of two individuals with 48% and 32% *S. nuchalis* ancestry (Figure 1a).

The hybrid individuals within AB had a mix of ancestry from both species. The amount of admixture varied throughout AB, with individuals located further north having a higher proportion of *S. varius* ancestry and those to the south, a greater proportion of *S. nuchalis* ancestry (Figure 2). Individuals with ancestry coefficient (Q) values ≥ 0.1 begin occurring at about 49.8˚ N and admixture is evident until about 52.3˚ N where no more individuals with Q ≤ 0.9 are found. This corresponds to a hybrid zone approximately 275 km long. Within AB, 22% of birds were classified as genomic *S. nuchalis* (Q ≤ 0.1), 30% as genomic *S. varius* (Q ≥ 0.9), and 48% as genomic admixed (0.1 < Q < 0.9).

Individuals within AB exhibited low overall interspecific heterozygosity (0.03 – 0.31; Figure 3). None of the AB individuals phenotypically identified as *S. varius* (n = 12) or *S. nuchalis* (n = 22) had interspecific heterozygosity of zero, or had hybrid indices of pure *S. varius* (HINDEX = 1.0) or *S. nuchalis* (HINDEX = 0.0), indicating that some alleles of these individuals were shared between species (Figure 3). While we observed a cline, the ancestry of hybrids within AB was not skewed towards one species or the other, but instead varied across the hybrid index between the two species, similar to the results with ADMIXTURE (Figures 1, 2 and 4). We found no F1 generation hybrids (HINDEX = 0.5, interspecific heterozygosity = 1.0) or early generation hybrids using all loci (Figure 3a), but we detected several using diagnostic loci (Figure 3b). The prevalence of samples with low interspecific heterozygosity is indicative of many advanced generation hybrids within the hybrid zone.

To determine if the low interspecific heterozygosity of hybrid individuals and the higher than expected interspecific heterozygosity of pure individuals is the result of widespread recombination or incomplete lineage sorting of ancestral traits, we identified a subset of 68 species-specific loci with an allele frequency differential (δ) of one (hereafter referred to as diagnostic loci). Although the interspecific heterozygosity of individuals differed between the diagnostic loci and all loci, the overall pattern was similar (Figure 3). The diagnostic loci identified three *S. varius* from AB as pure (HINDEX = 1.0, interspecific heterozygosity = 0); also, four *S. nuchalis* and two phenotypic hybrids were identified as pure *S. nuchalis* (HINDEX = 0, interspecific heterozygosity = 0) (Figure 3b). No individuals we phenotypically identified as belonging to one species was categorized as belonging to the other species using genomic data, but with diagnostic loci many phenotypic hybrids had genotypes similar to pure *S. nuchalis* or *S. varius (8 hybrids ∂ = 1 HINDEX ≤ 0.1, 6 hybrids ≥ 0.9)*, while 12 phenotypically pure individuals were genotypically backcrossed (∂ =1 HINDEX ≥ 0.1, ≤ 0.9, interspecific heterozygosity ≤ 0.85) and one was a clear genomic F1 (∂ = 1 HINDEX ≥ 0.1, ≤ 0.9, interspecific heterozygosity ≥ 0.85)(Figure 3b).

*Traditional genetic markers*

Enol showed nearly diagnostic SNPs, with a T allele being characteristic of *S. nuchalis* (110 of 130 alleles) and a C allele being more common in *S. varius* (123 of 164 alleles; p < 0.0001; Table 3, Figure 4). *Sphyrapicus nuchalis* had a significantly higher proportion of GAPD insertions (49 of 122 alleles) than *S. varius* (72 of 138 alleles; p < 0.0001; Table 3a,b), and *S. nuchalis* tended to have proportionally fewer T alleles (97 of 140) at the anonymous nuclear marker than *S. varius* (169 of 194 alleles; p < 0.0001) (Table 3a,b). The two species showed significant differences at all three loci when considering only parental populations outside central Alberta (Enol, GAPD, anonymous nuclear marker, p < 0.0001; Table 3b).

Birds identified as *S. nuchalis* within AB had significantly different allele frequencies from both *S. varius* in AB and phenotypic hybrids at all three loci (p ≤ 0.012) (Table 3b). However, *S. varius* had no loci with allele frequencies that differed significantly from hybrids (p ≥ 0.236).

Log likelihood values identified K = 2 as optimal for STRUCTURE. The clusters generally described the two species well, with most individuals in central AB expressing shared ancestry (Figure 1b). In AB, the STRUCTURE plot shows admixture extending through the northern extent of our transect, revealing more extensive introgression in phenotypically *S. varius* individuals than in phenotypically *S. nuchalis* individuals. Individuals with ancestry coefficient values ≥ 0.1 begin occurring at about 49.8˚ N and remain throughout the rest of the transect (up to 60.0˚ N), representing a transect of more than 1130 km. Within AB, 21% of birds were classified as genetic *S. nuchalis* (Q ≤ 0.1), 0% as genetic *S. varius* (Q ≥ 0.9), and 89% as genetically admixed (0.1 < Q < 0.9).

Of 45 individuals examined using both GBS and traditional methods, Q values assigned clusters similarly in 29 samples and differently in 16.

**Discussion**

*Traditional markers vs GBS*

We provide the results for traditional genetic markers for a methodological comparison with GBS because it is generally financially and practically more accessible to researchers. The traditional genetic data differentiated well between admixed and non-admixed populations (Figure 1a), and corroborated the GBS findings of strong admixture between the two species along a north-south gradient of western AB. However, GBS data documented less admixture of northern *S. varius* compared to the traditional methods. The discrepancy between the two methods is likely due to the lower resolution inherent in using three unfixed loci compared to thousands of loci with GBS, including many which were diagnostic. Because of this general limitation of the traditional loci approach, we refer primarily to the GBS results throughout the discussion.

*Rates of hybridization and introgression*

Our data record high levels of introgression and a continuum from early to advanced generation hybrids. Most individuals within west-central AB were admixed (48% GBS, 89% traditional loci with intermediate Q scores), compared to only a few individuals with intermediate Q scores in a *S. ruber/S. nuchalis* hybrid zone in northern California and Oregon (Billerman et al. 2019), and a low incidence of hybridization in the *S. ruber/S. varius* hybrid zone in central British Columbia (Seneviratne et al. 2016). Of the 60 samples analyzed using diagnostic GBS loci, four individuals, a phenotypic *S. nuchalis* and three hybrids, had interspecific heterozygosity scores expected of F1 hybrids (Figure 3b). Several other individuals had intermediate HINDEX scores and interspecific heterozygosity values, which is consistent with them being non-F1 early generation hybrids and backcrosses. The majority of admixed individuals, however, appear to be advanced generation hybrids (intermediate HINDEX scores, low interspecific heterozygosity). The presence of many advanced generation hybrids with low interspecific heterozygosity and a range of HINDEX scores indicates a well-established hybrid zone with many hybrid individuals backcrossing with individuals from the parental species and each other.

*Geographic, genomic, and genetic patterns of introgression*

Our data document genetic differentiation among allopatric populations of *S. nuchalis* and *S. varius*, that disappears within the hybrid zone. This is supported by the strong genetic differentiation of populations outside AB from individuals within AB, and the low differentiation between individuals phenotypically resembling the parent species in AB and hybrids. Admixture of individuals within and near the hybrid zone occurs largely along a north-south gradient, in part due to no suitable habitat being available to the west and east of the area of contact. *Sphyrapicus nuchalis* largely spreads from the south through low elevation passes across the Rockies, whereas *S. varius* reaches the contact zone through a narrow strip of forest between the Rocky Mountains and the Canadian prairies connected to the boreal forest to the north. Our sampling throughout approximately 1200 km of AB shows genomically admixed individuals present over approximately 275 km along the foothills east of the Rocky Mountains.

*Sphyrapicus nuchalis* and *S. varius* both favor secondary growth deciduous and mixed forests and specialize on similar plant species including aspen (*Populus spp.*) and birch (*Betula spp.*) trees (Walters et al., 2002a, 2002b). Shared habitat preferences increase the likelihood for the parental species to occupy the same area, and sympatry of the forms will increase the likelihood of admixture. Furthermore, the hybrid zone is located along a transitional environment between montane habitat to the south and the boreal forest to the north, where habitats used by the two species blend smoothly (Downing and Pettapiece 2006). In this blended environment intermediate phenotypes could be at an advantage over the parental phenotypes (the bounded hybrid superiority hypothesis of Moore 1977). Additionally, hybrid zones made up of mostly late generation hybrids could permit selection to create novel allelic combinations adapted to the intermediate environment (Hamilton & Aitken, 2013; Milne & Abbott, 2008; Pinheiro et al., 2016). However, we note that the hybrid zone appears to be moving, *S. varius* and its characteristics, shifting south slightly over the 15 years of study of reproductive interactions between the two forms in western Alberta (Jocelyn Hudon, pers obs.).

In this region of Alberta where the two species of sapsuckers meet, there is limited suitable habitat for both species, creating a density trough where the genetic make-up of individuals in the hybrid zone may be influenced by immigration from larger populations to the north and/or to the south. Limited opportunities to mate with conspecifics in such “density traps” may further promote hybridization (Hewitt 1988, McCracken et al. 2013).

*Plumage insights*

All pre-genomic species classifications in this study were made based on plumage. *Sphyrapicus nuchalis* and *S. varius* are the more difficult sapsuckers to differentiate by plumage, but these classifications were carefully made using an eight-point system developed by J. Hudon (Supplemental Materials). Despite meticulous classification, our phenotypic and genotypic classifications were not always concordant. Eight phenotypic hybrids were classified as genotypic parental types, and 13 phenotypic parental birds were genetically admixed (Figure 3b). This contrasts with the finding that plumage is generally a reliable predictor of ancestry in *S. nuchalis* and *S. ruber* (Billerman et al. 2019). This could be due to the plumage differences between *S. nuchalis* and *S. varius* being controlled by a few genes of large effect. Howell (1952) noted the different sapsucker species differ primarily in the presence or absence of red coloration at the tip of black and white feathers on the head, neck, and upper breast. There are other examples where color differences between closely related species are explained by few genes or small genomic regions (e.g., Toews et al. 2016, Wang et al. 2020, Aguillon et al. 2020). This means that species designations based on plumage should be taken with care, and analyses should be based upon genetic classifications and additional data wherever possible.

Plumage is often identified as a potentially important isolating barrier preventing the production of a hybrid swarm and genetic homogenization in sapsuckers (Billerman et al. 2019, Johnson & Zink 1983, Seneviratne et al. 2016). Though assortative mating has been observed or inferred in both *S. nuchalis/S. ruber* (Billerman et al. 2019, Johnson & Johnson 1985) and *S. varius/S. ruber* (Seneviratne et al. 2012, 2016) hybrid zones, no data has yet been gathered on whether *S. nuchalis* and *S. varius* mate assortatively based on plumage phenotypes. We note that they are the two forms that look most alike, though they are not each other’s closest relatives (Johnson and Zink 1983, Cicero and Johnson 1995). Indeed, hybridization between the two species does not appear to be prevented by their strong genetic differentiation (FST between *S. nuchalis* and *S. varius* (0.26 this study, 0.23 Grossen et al. 2016) is large compared to either species and *S. ruber* (0.18 *ruber/varius*, 0.06 *ruber/nuchalis*, Grossen et al. 2016). This raises an intriguing question: how can assortative mating limit introgression if mate choice is based on a trait that is not closely associated with ancestry, a phenomenon recently described by Semenov et al. (2017) in white wagtails (*Motacilla alba*). If genotype and phenotype are not highly correlated, assortative mating based on plumage may not lead to assortative mating at most loci, and what would appear to be an isolating barrier may actually permit widespread introgression.

*S. nuchalis/S. varius in the sapsucker hybridization complex*

We have shown that hybridization between *S. nuchalis*/*S. varius* occurs in a relatively stable hybrid zone 275 km in length in western Alberta, thus relatively wide compared to the 65 km for a S. *ruber/S. nuchalis* hybrid zone in southern British Columbia and 84 km for S. *ruber/S. varius* hybrid zone in northern British Columbia (775 and 400 km transects surveyed respectively; Seneviratne et al. 2016). Hybridization rates were quite high, and a variety of hybrid classes existed suggesting backcrossing occurs frequently. These are somewhat surprising results given that FST between *S. nuchalis* and *S. varius* (0.26 this study, 0.23 Grossen et al. 2016) is higher than that between either species and *S. ruber* (0.18 *ruber/varius*, 0.06 *ruber/nuchalis*, Grossen et al. 2016). It could be that the greater phenotypic similarities in S. nuchalis/varius facilitate more introgression than would be expected from their high genomic differentiation and more distant evolutionary history.

Plumage is not a reliable indicator of genotype among *S. nuchalis* and *S. varius* individuals, despite high plumage/genotype correlations in another sapsucker pair (Billerman et al. 2019). The contrasts amongst this and other well documented sapsucker hybrid zones are intriguing for such closely related species and they elicit a unique opportunity to try to decouple isolating barriers and hybridization rates in a future study integrating all three hybrid zone combinations.

**Acknowledgements**

We would like to acknowledge the assistance of R. Adams, B. Brinkman, K. Dohms, G. M. Erickson, B. Graham, C. Kaluthota, L. Lait, C. Macfarlane, P. Pulgarin, C. Scobie, and C. Welke in the field. We also thank the following museums for supplying samples: Royal Alberta Museum, Canadian Museum of Nature, Field Museum, Museum of Southwest Biology, and Royal Saskatchewan Museum. Funding has been provided by Natural Sciences and Engineering Research Council of Canada Discovery Grant (TMB), Alberta Innovates (TMB), and the North American Bluebird Society (LN).

**Data Accessibility**

Data used in this manuscript will be made available on Dryad upon acceptance.

**Author Contributions**

JH and TB initiated the study. JH collected and scored the phenotype of sapsuckers from the hybrid zone. AC processed and analyzed the GBS data, while LN processed and analyzed the traditional genetic markers data. LN synthesized and organized the results. All authors contributed to the project development, analyses, and writing.

**References**

Aguillon, S.M., Walsh, J., & Lovette, I.J. (2020). Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. bioRxiv 2020.07.10.197715;

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19(9), 1655–1664.

Billerman, S. M. (2016). An Integrative Approach to Understanding Hybrid Zone Movement and Introgression in Sphyrapicus Woodpeckers. University of Wyoming.

Billerman, S. M., Cicero, C., Bowie, R. C. K., & Carling, M. D. (2019). Phenotypic and genetic introgression across a moving woodpecker hybrid zone. Molecular Ecology, 28, 1692–1708.

Billerman, S. M., Murphy, M. A., & Carling, M. D. (2016). Changing climate mediates sapsucker (Aves : *Sphyrapicus*) hybrid zone movement. Ecology and Evolution, 6(22), 7976–7990.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19), 2633–2635.

Burg, T.M., & Croxall, J.P. (2001). Global relationships amongst black-browed and grey-headed albatrosses: analysis of population structure using mitochondrial DNA and microsatellites. Molecular Ecology, 10, 2647–2660.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. Molecular Ecology, 22(11), 3124–3140.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., & De Koning, D.-J. (2011). Stacks: Building and genotyping loci *de novo* from short-read sequences. G3, 1(3), 171–182.

Cicero, C., & Johnson, N. K. (1995). Speciation in sapsuckers (*Sphyrapicus*): III. Mitochondrial-DNA sequence divergence at the cytochrome-b locus. Auk, 112(3), 547–553.

Downing, D.J., & Pettapiece, W.W. (2006). Natural Regions Committee 2006. Natural Regions and Subregions of Alberta. Government of Alberta. Pub. No. T/852.

Friesen, V. L., Congdon, B. C., Walsh, H. E., & Birt, T. P. (1997). Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. Molecular Ecology, 6(11), 1047–1058.

Gompert, Z., & Buerkle, C. A. (2009). A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Molecular Ecology, 18(6), 1207–1224.

Gompert, Z., & Buerkle, C. A. (2010). INTROGRESS: A software package for mapping components of isolation in hybrids. Molecular Ecology Resources, 10(2), 378–384.

Grant, P. R., & Grant, B. R. (1992). Hybridization of bird species. Science, 256(5054), 193–197.

Grossen, C., Seneviratne, S. S., Croll, D., & Irwin, D. E. (2016). Strong reproductive isolation and narrow genomic tracts of differentiation among three woodpecker species in secondary contact. Molecular Ecology, 25, 4247–4266.

Hamilton, J. A., & Aitken, S. N. (2013). Genetic and morphological structure of a spruce hybrid (*Picea sitchensis x P. glauca*) zone along a climatic gradient. American Journal of Botany, 100(8), 1651–1662.

Hewitt, G. M. (1988). Hybrid zones - natural laboratories for evolutionary studies. Trends in Ecology & Evolution, 3(7), 158–167.

Howell, T. R. (1952). Natural history and differentiation in the yellow-bellied sapsucker. Condor, 54(5), 237–282.

Johnson, N. K., & Johnson, C. B. (1985). Speciation in sapsuckers (*Sphyrapicus*): II. Sympatry, hybridization, and mate preference in *S. ruber daggetti* and *S. nuchalis*. Auk, 102(1), 1–15.

Johnson, N. K., & Zink, R. M. (1983). Speciation in sapsuckers (*Sphyrapicus*): I. Genetic differentiation. Auk, 100(4), 871–884.

McCracken, K. G., Wilson, R. E., & Martin, A. R. (2013). Gene flow and hybridization between numerically imbalanced populations of two duck species on the subantarctic island of South Georgia. PLoS ONE, 8(12), 1–8.

Milne, R. I., & Abbott, R. J. (2008). Reproductive isolation among two interfertile *Rhododendron* species: Low frequency of post-F1 hybrid genotypes in alpine hybrid zones. Molecular Ecology, 17(4), 1108–1121.

Moore, W. S. (1977). An evaluation of narrow hybrid zones in vertebrates. The Quarterly Review of Biology, 52(3), 263–277.

Nadeau, N. J., Mundy, N. I., Gourichon, D., & Minvielle, F. (2007). Association of a single-nucleotide substitution in TYRP1 with roux in Japanese quail (*Coturnix japonica*). Animal Genetics, 38(6), 609–613.

Natola, L., & Burg, T. M. (2018). Population genetics and speciation of yellow-bellied, red-naped, and red-breasted sapsuckers (*Sphyrapicus varius*, *S. nuchalis*, and *S. ruber*). Journal of Heredity, 109(6), 663–674.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE, 7(5), e37135.

Pinheiro, F., Gouveia, T. M. Z. de M. E, Cozzolino, S., & Cafasso, D. (2016). Strong but permeable barriers to gene exchange between sister species of *Epidendrum*. American Journal of Botany, 103(8), 1472–1482.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3), 559–575.

Scott, D. M., Ankney, C. D., & Jarosch, C. H. (1976). Sapsucker hybridization in British Columbia: changes in 25 years. Condor, 78(2), 253–257.

Semenov, G. A., Smith, C. C. R., Scordato, E. S. C., Khaydarov, D. R., Kane, N. C., & Safran, R. J. (2017). Effects of assortative mate choice on the genomic and morphological structure of a hybrid zone between two bird subspecies. Molecular Ecology, 26, 6430–6444.

Seneviratne, S. S., Davidson, P., Martin, K., & Irwin, D. E. (2016). Low levels of hybridization across two contact zones among three species of woodpeckers (*Sphyrapicus* sapsuckers). Journal of Avian Biology, 47(6), 887–898.

Seneviratne, S. S., Toews, D. P. L., Brelsford, A., & Irwin, D. E. (2012). Concordance of genetic and phenotypic characters across a sapsucker hybrid zone. Journal of Avian Biology, 43(2), 119–130.

Short, L. L. (1972). Hybridization, taxonomy and avian evolution. Missouri Botanical Garden Press, 59(3), 447–453.

Shunk, S. A. (2005). Sphyrapicus anxiety. Identifying hybrid sapsuckers. Birding 37, 289–298.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28(10), 2731–2739.

Team, R. D. C. (2011). R: a language and environment for statistical computing. R Foundation for Statistical Computing.

Toews, D.P.L., Taylor, S.A., Vallender, R., Brelsford, A., Butcher, B.G., Messer, P.W., Lovette, I.J. 2016. Plumage genes and little else distinguish the genomes of hybridizing Warblers. Current Biol 26, 1-6.

Trombino, C. L. (1998). Species interactions in the hybrid zone between red-breasted (*Sphyrapicus ruber*) and red-naped (*Sphyrapicus nuchalis*) sapsuckers: Fitness consequences, reproductive character displacement, and nest site selection. Northern Illinois University.

Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex-100 as a medium for simple extraction of DNA for PCR-Based typing from forensic material. BioTechniques, 10(4), 506–513.

Walters, E. L., Miller, E. H., & Lowther, P. E. (2002a). Red-breasted sapsucker (*Sphyrapicus ruber*) and red-naped sapsucker (*Sphyrapicus nuchalis*). Birds of North America, *663*, 1–32.

Walters, E. L., Miller, E. H., & Lowther, P. E. (2002b). Yellow-bellied sapsucker. Birds of North America, 662, 2–24.

Wang, S., Rohwer, S., de Zwaan, D. E., Toews, D. P. L., Lovette, I. J., Mackenzie, J., & Irwin, D. (2020). Selection on a small genomic region underpins differentiation in multiple color traits between two warbler species. Evol Letters doi.org/10.1002/evl3.198

Zhou, H., Alexander, D., & Lange, K. (2011). A quasi-Newton acceleration for high-dimensional optimization algorithms. Statistics and Computing, 21(2), 261–273.

Tables.

Table 1 Primer sets used for the α-enolase, glyceraldehyde, and anonymous nuclear marker, primer sequences, and source.

|  |  |  |  |
| --- | --- | --- | --- |
| Gene | Primer Name | Primer sequence (5’-3’) | Source |
| Enol | Enol8 L731 | TGGACTTCAAATCCCCCGATGATCCCAGC | (Friesen et al., 1997)  (Friesen et al., 1997) |
| Enol9 H912 | CCAGGCACCCCAGTCTACCTGGTCAAA |
| Enol SapLM13 | GTCCTGTGAATGTTCTTTGAGGCGG | This study |
| GAPD | GAPD11 L890 | ACCTTTAATGCGGGTGCTGGCATTGC | (Friesen et al. 1997) |
| GAPD12 H950 | CATCAAGTCCACAACACGGTTGCTGTA | (Friesen et al. 1997) |
| anonymous nuclear marker | TP1F4 | CAGCTCTGCTGAACCTGTTG | (Nadeau et al., 2007)  (Nadeau et al., 2007) |
| TP1R5 | ATTGGTTTTAGTCACAAGCAAAAA |
| TP1 SapRM13 | GCTGTTGAGTTTTGGCTTACC | This study |

Table 2 Standard screening PCR and restriction digest protocols. Bolded, underlined nucleotides in restriction enzyme sequences denote SNP sites used for screening.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | PCR | | | | Restriction Digest | | | | |
|  | 5x buffer | MgCl2 | M13 tag | Ta | Enzyme | reaction  volume | enzyme  (U) | PCR  product | Gel |
| Enol | clear | 2.0 mM | added | 60˚ | *Fnu4HI*  *(5’-GCNG****C****-3’)* | 5 µL | 0.5 | 1 µL | Acrylamide |
| GAPD | green | 2.0 mM | omitted | 60˚ | *NlaIII*  *(5’-C****A****T****G****-3’)* | 10 µL | 1 | 6 µL | Agarose |
| anonymous nuclear marker | green | 2.5 mM | omitted | 52˚ | *MspI*  (5’-C**C**GG-3’) | 10 µL | 2 | 6 µL | Agarose |

Table 3 SNP assignments by locus and population (a). Samples sizes recorded as number of alleles (n). Comparisons among species, hybrids, and populations within Alberta (b). P-values from Fisher’s exact tests on the left, bold indicates comparisons are significant (p < 0.05), values to the right of the p-values indicate number of alleles (n).

a.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Enol | | | GAPD | | | anonymous nuclear marker | | |
|  | T | C | n | insertion | deletion | n | T | C | n |
| *S. nuchalis* NM | 28 | 2 | 30 | 24 | 6 | 30 | 23 | 7 | 30 |
| *S. nuchalis* ID | 20 | 0 | 20 | 19 | 1 | 20 | 13 | 7 | 20 |
| *S. nuchalis* AB | 62 | 18 | 80 | 66 | 6 | 72 | 53 | 29 | 82 |
| Hybrid | 29 | 51 | 80 | 52 | 34 | 86 | 76 | 10 | 86 |
| *S. varius* AB | 34 | 52 | 86 | 45 | 31 | 76 | 91 | 21 | 112 |
| *S. varius* SK | 3 | 25 | 28 | 18 | 10 | 28 | 27 | 1 | 28 |
| *S. varius* IL | 1 | 25 | 26 | 21 | 9 | 30 | 29 | 1 | 30 |
| *S. varius* NSNB | 3 | 21 | 24 | 2 | 2 | 4 | 22 | 2 | 24 |
| Total | 180 | 194 | 374 | 247 | 99 | 346 | 334 | 78 | 412 |

b.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Enol | | GAPD | | anonymous nuclear marker | |
| *S. nuchalis x* *S. varius* | **p < 0.0001** | 294 | **p < 0.0001** | 262 | **p < 0.0001** | 326 |
| AB *S. nuchalis x* AB *S. varius* | **p < 0.0001** | 166 | **p < 0.0001** | 150 | **p = 0.004** | 194 |
| AB *S. nuchalis x* hybrid | **p < 0.0001** | 158 | **p < 0.0001** | 156 | **p < 0.001** | 166 |
| AB *S. varius x* hybrid | 0.7536 | 168 | 0.4247 | 162 | 0.4318 | 200 |

A picture containing chart

Description automatically generated

Figure 1 ADMIXTURE plot of the sapsuckers using GBS data (a) and STRUCTURE plot using traditional genetic markers (b) showing *S. nuchalis* ancestry (black) and *S. varius* ancestry (white). The individuals within AB are organized south to north based on latitude. Violet outlines birds trapped within the 275 km hybrid zone identified using GBS q values. See Figure 2 for additional details on locations.

Website, map

Description automatically generated

Figure 2 Sampling locations of individuals phenotypically identified as *S. nuchalis* (black triangles), *S. varius* (white circles), and hybrids (light grey diamonds). Violet indicates location of hybrid zone identified using GBS loci. The STRUCTURE plot on the left indicates the proportion of the genotype that is *S. varius* ancestry (white) and *S. nuchalis* ancestry (black) using traditional genetic methods, ancestry calculated using GBS is represented in ADMIXTURE plot on the right. Individuals in the plots are arranged by ascending sampling latitude, with the most southern individual at the bottom of the plot and the most northern at the top. Violet boxes indicate birds sampled within the hybrid zone. PID columns show phenotypic species identification for each individual in the STRUCTURE and ADMIXTURE plots, with black representing *S. nuchalis*, white for *S. varius*, and grey for hybrids.

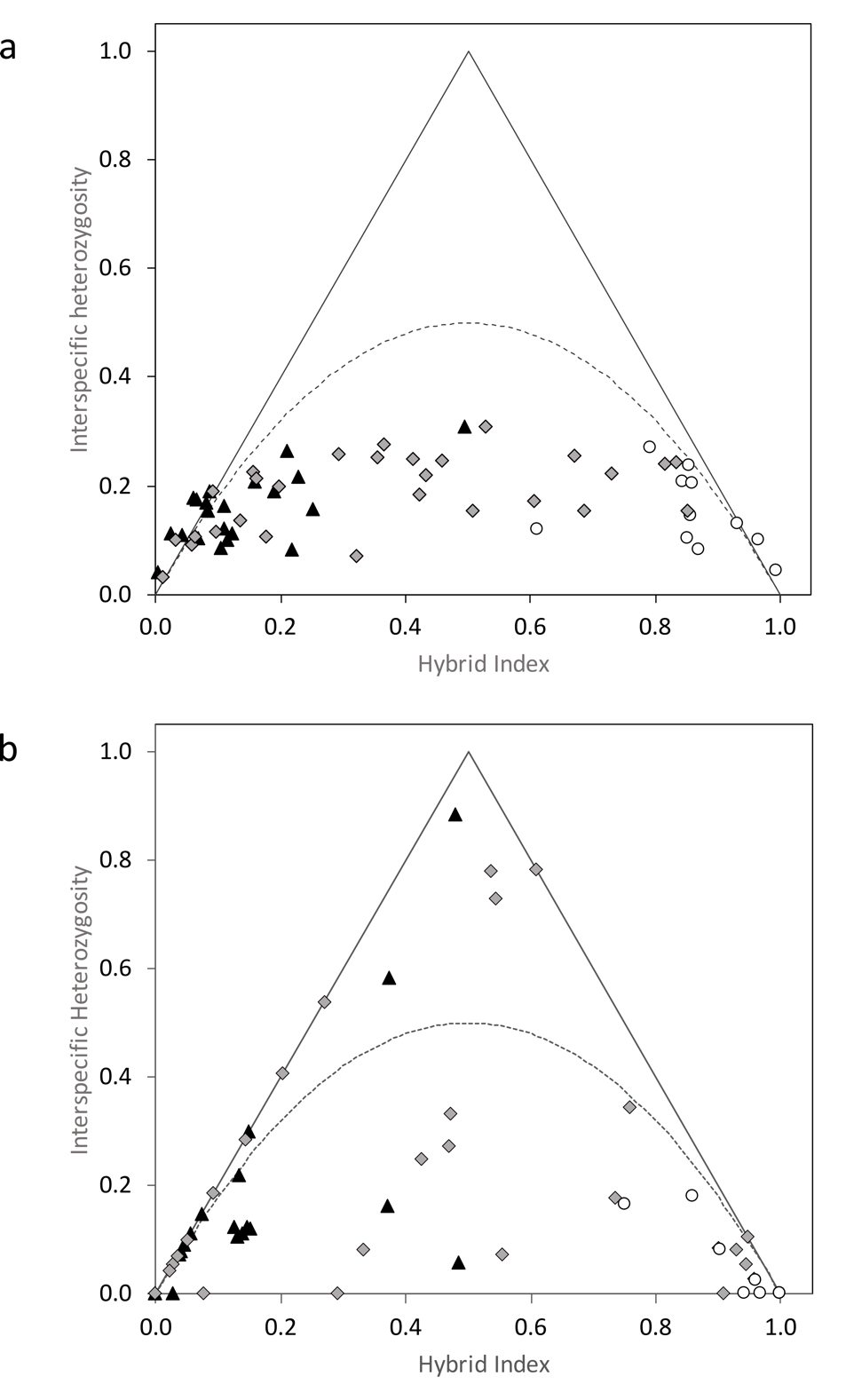


Figure 3 Interspecific heterozygosity and hybrid index (HINDEX) scores of each individual in AB using (a) all loci and (b) diagnostic loci. Symbols identify the forms sampled based on phenotype of the birds: pure *S. nuchalis* (black triangle), pure *S. varius* (white circles), and hybrids (grey diamonds). Individuals located on the solid line are back-crossed individuals. The dotted line indicates the expected distribution for individuals if mating was random throughout the zone.

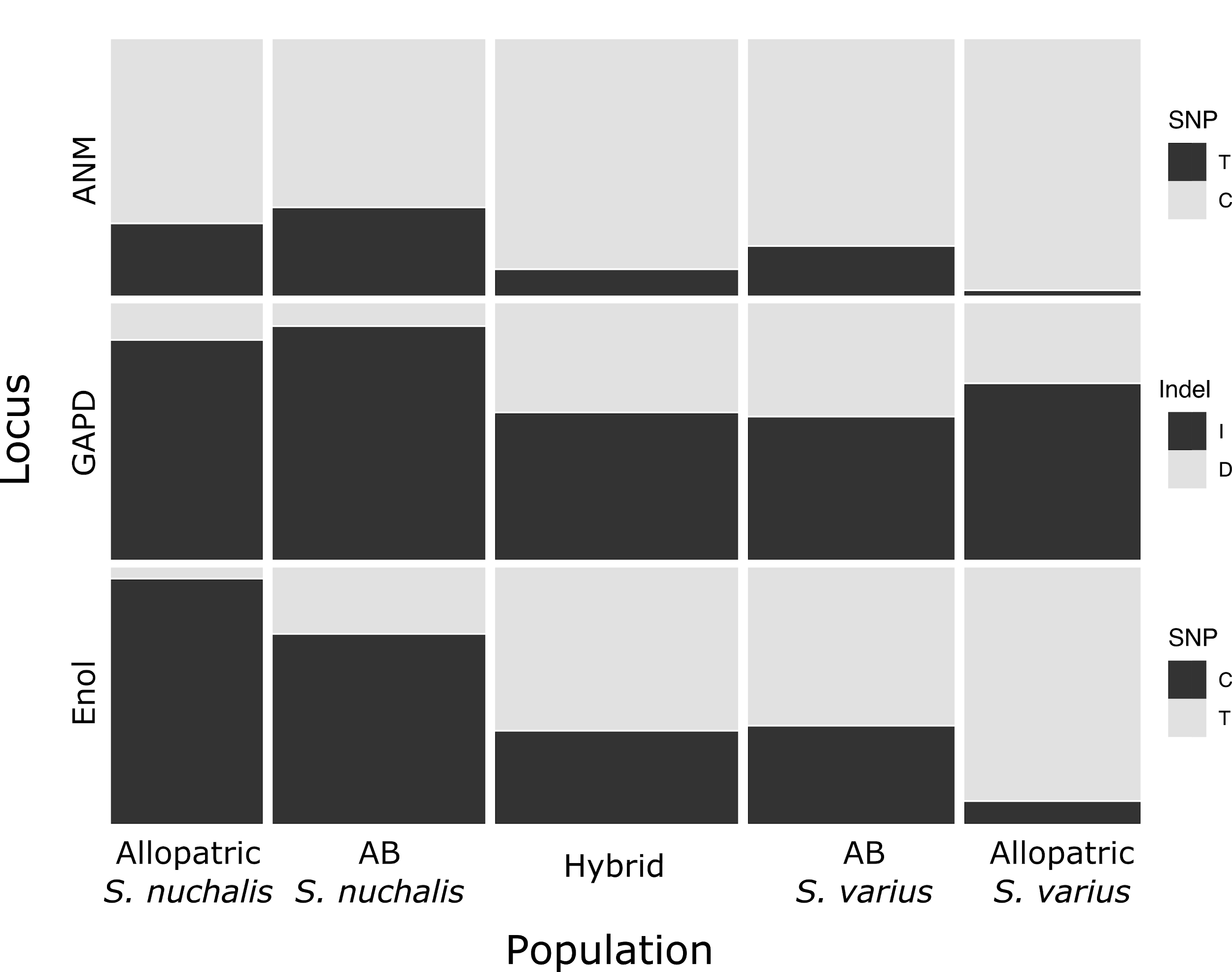


Figure 4 SNP assignments for Enol, GAPD, and anonymous nuclear marker for each population sampled. Sample sizes in Table 3.