Received: 16 April 2018 | Revised: 5 March 2019 | Accepted: 29 March 2019

DOI: 10.1111/mec.15091

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America

Philip Lavretsky[1,2,3] | Jeffrey M. DaCosta[4,5] | Michael D. Sorenson[4] | Kevin G. McCracken[3,6,7,8] | Jeffrey L. Peters[2]

[1]Department of Biological Sciences, University of Texas at El Paso, El Paso, Texas

[2]Department of Biological Sciences, Wright State University, Dayton, Ohio

[3]Department of Biology, University of Miami, Miami, Florida

[4]Biology Department, Boston College, Chestnut Hill, Massachusetts

[5]Biology Department, Boston College, Boston, Massachusetts

[6]Department of Marine Biology and Ecology, Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, Miami, Florida

[7]Human Genetics and Genomics, Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida

[8]Institute of Arctic Biology and University of Alaska Museum, University of Alaska Fairbanks, Fairbanks, Alaska

Correspondence
Philip Lavretsky, Department of Biological Sciences, University of Texas at El Paso, El Paso, TX.
Email: plavretsky@utep.edu

Funding information
James A. Kushlan endowment in Waterbird Biology and Conservation at the University of Miami; Black Duck Joint Venture; Safari Club International El Paso Chapter

Abstract

Recently evolved species typically share genetic variation across their genomes due to incomplete lineage sorting and/or ongoing gene flow. Given only subtle allele frequency differences at most loci and the expectation that divergent selection may affect only a tiny fraction of the genome, distinguishing closely related species based on multi-locus data requires substantial genomic coverage. In this study, we used ddRAD-seq to sample the genomes of five recently diverged, New World "mallards" (*Anas* spp.), a group of dabbling duck species characterized by diagnosable phenotypic differences but minimal genetic differentiation. With increased genomic sampling, we aimed to characterize population structure within this group and identify genomic regions that may have experienced divergent selection during speciation. We analyzed 3,017 autosomal ddRAD-seq loci and 177 loci from the Z-chromosome. In contrast to previous studies, the ddRAD-seq data were sufficient to assign individuals to their respective species or subspecies and to generate estimates of gene flow in a phylogenetic framework. We find limited evidence of contemporary gene flow between the dichromatic mallard and several monochromatic taxa, but find evidence for historical gene flow between some monochromatic species pairs. We conclude that the overall genetic similarity of these taxa likely reflects retained ancestral polymorphism rather than recent and extensive gene flow. Thus, despite recurring cases of hybridization in this group, our results challenge the current dogma predicting the genetic extinction of the New World monochromatic dabbling ducks via introgressive hybridization with mallards. Moreover, ddRAD-seq data were sufficient to identify previously unknown outlier regions across the Z-chromosome and several autosomal chromosomes that may have been involved in the diversification of species in this recent radiation.

KEYWORDS
ddRADseq, hybridization, introgression, population genomics, sex chromosome, speciation

1 | INTRODUCTION

Recently diverged species share genetic polymorphisms due to their common ancestry (incomplete lineage sorting [ILS]) and are often characterized by "porous" genomes that are open to gene flow during ongoing or secondary contact (Malinsky, Svardal, et al., 2018a; Mallet, Besansky, & Hahn, 2016; Rheindt & Edwards, 2011; Seehausen, 2004). This shared variation can make it difficult to detect subtle patterns of population structure. Methodological advances over the past decade, however, now permit researchers to efficiently and economically sample hundreds to thousands of loci scattered across the genome (Funk, McKay, Hohenlohe, & Allendorf, 2012; Oyler-McCance, Oh, Langin, & Aldridge, 2016; Rice, Rudh, Ellegren, & Qvarnström, 2011). In addition to providing sufficient power for multi-locus "diagnosis" of closely related species and populations (Ellegren, 2008; Stapley et al., 2010; Toews et al., 2015), these methods may provide sufficient coverage of the genome to detect genetic regions involved in phenotypic divergence and speciation (Abbott et al., 2013; Nosil & Schluter, 2011; Rice et al., 2011; Seehausen, 2004; Wolf, Lindell, & Backström, 2010; Wu & Ting, 2004). Of the various high-throughput genomic methods, restriction-site-associated DNA sequencing (RAD-seq; Miller, Dunham, Amores, Cresko, & Johnson, 2007), and related methods (e.g., ddRAD, GBS, etc; Andrews, Good, Miller, Luikart, & Hohenlohe, 2016), have been particularly transformative for non-model organisms (Andrews et al., 2016; Davey & Blaxter, 2010; Ellegren, 2014).

The "mallard complex" exemplifies the challenges of identifying diagnostic genetic markers for recently diverged taxa. Five members of the mallard group occur in North America, the sexually dichromatic mallard (*Anas platyrhynchos*) and four monochromatic taxa: American black duck (*A. rubripes*; "black duck"), Mexican duck (*A [p.] diazi*), Florida mottled duck (*A. fulvigula fulvigula*), and West Gulf Coast mottled duck (*A. f. maculosa*). Although each of these species/subspecies is phenotypically distinguishable, they have not achieved reciprocal monophyly in mitochondrial DNA (mtDNA) and are only weakly differentiated in nuclear allele frequencies (Avise, Ankney, & Nelson, 1990; Johnson & Sorenson, 1999; Lavretsky et al., 2015; Lavretsky, McCracken, & Peters, 2014b; McCracken, Johnson, & Sheldon, 2001; Peters et al., 2014). Along with recent divergence and ILS, hybridization with mallards upon secondary contact is thought to contribute to the genetic similarity of these taxa (Ankney, Dennis, Wishard, & Seeb, 1986; Avise et al., 1990; Johnson & Sorenson, 1999; Lavretsky et al., 2015; Lavretsky, McCracken, et al., 2014b; McCracken et al., 2001).

Among the monochromatic species, black ducks have the highest rates of hybridization and introgression with mallards (Lavretsky, Janzen, & McCracken, 2019), and previous studies have suggested either that the two forms are conspecific (Ankney et al., 1986), or that black ducks have suffered a complete breakdown of their genetic distinctiveness (Mank, Carlson, & Brittingham, 2004). Although the frequency of mixed pairs and hybrid individuals remains uncertain (Heusmann, 1988; Johnsgard, 1967; Kirby, Reed, Dupuis, Obrecht, & Quist, 2000), black ducks and mallards are genetically indistinguishable based on allozymes (Ankney et al., 1986), microsatellites (Mank et al., 2004), and sequence data from a limited number of mitochondrial and nuclear genes (Johnson & Sorenson, 1999; Lavretsky, Hernández Baños, & Peters, 2014a; Lavretsky, McCracken, et al., 2014b; McCracken et al., 2001). ddRAD-seq methods, however, have provided sufficient genomic coverage and allele frequency differences to identify population structure, hybrid individuals, founder events, and genomic regions putatively under divergent selection between mallards and Mexican ducks (Lavretsky et al., 2015), between the two mottled duck subspecies (Peters et al., 2016), and more recently between mallards and black ducks (Lavretsky et al., 2019). Regardless, assessing the relative roles of gene flow and ILS in explaining the genetic similarity across New World (NW) mallard taxa remains difficult.

Given diagnosable phenotypes, but a lack of diagnostic neutral genetic diversity (Φ_{ST} range for 17 introns = 0.011–0.043; Lavretsky, Hernández Baños, et al., 2014a), the primary objective of this study was to test for genome-wide genetic differentiation among all five NW mallard taxa, and especially between the mallard and black duck. Specifically, we address the following questions: (a) are allele frequency differences from thousands of loci sufficient for distinguishing among these taxa, (b) is there evidence of genetic regions showing elevated divergence suggestive of divergent selection, and (c) is the Z sex-chromosome more divergent than the autosomes? In addition, we use a phylogenetic framework, as implemented in the program TreeMix (Pickrell & Pritchard, 2012), to test the null hypothesis that recent divergence and ILS is sufficient to explain the genetic similarity among taxa.

2 | MATERIALS AND METHODS

2.1 | Sampling, DNA extraction, library preparation, and data processing

A total of 178 samples (28–43 per taxon) representing the five closely related NW dabbling ducks (Figure 1; Table S1) were included in our analyses. We used published ddRAD-seq data for FL and WGC mottled ducks (BioProject PRJNA343361, Peters et al., 2016), and Mexican ducks (BioProject SRP064125, Lavretsky et al., 2015) (BioProject PRJNA516035, Lavretsky et al., 2019). Only Mexican ducks from interior Mexico were used to limit the influence of potentially high rates of recent introgression with mallards in the northern portion of their range (U.S. populations; Aldrich & Baer, 1970; Hubbard, 1977) and potential biases resulting from a likely recent founder event in coastal habitats in Sonora (i.e., Sonora; Williams, 1980; Perez-Arteaga, Gaston, & Kershaw, 2002; Lavretsky et al., 2015). In addition to previously published data for 17 mallards (SRP # SRP064125 [Sample IDs 4095849–4095865]; Lavretsky et al., 2015), 11 additional NW mallards were sampled for this analysis. We also analyzed new ddRAD-seq data for 28 black ducks sampled across their range (Figure 1; BioProject Sample IDs see Table S1).

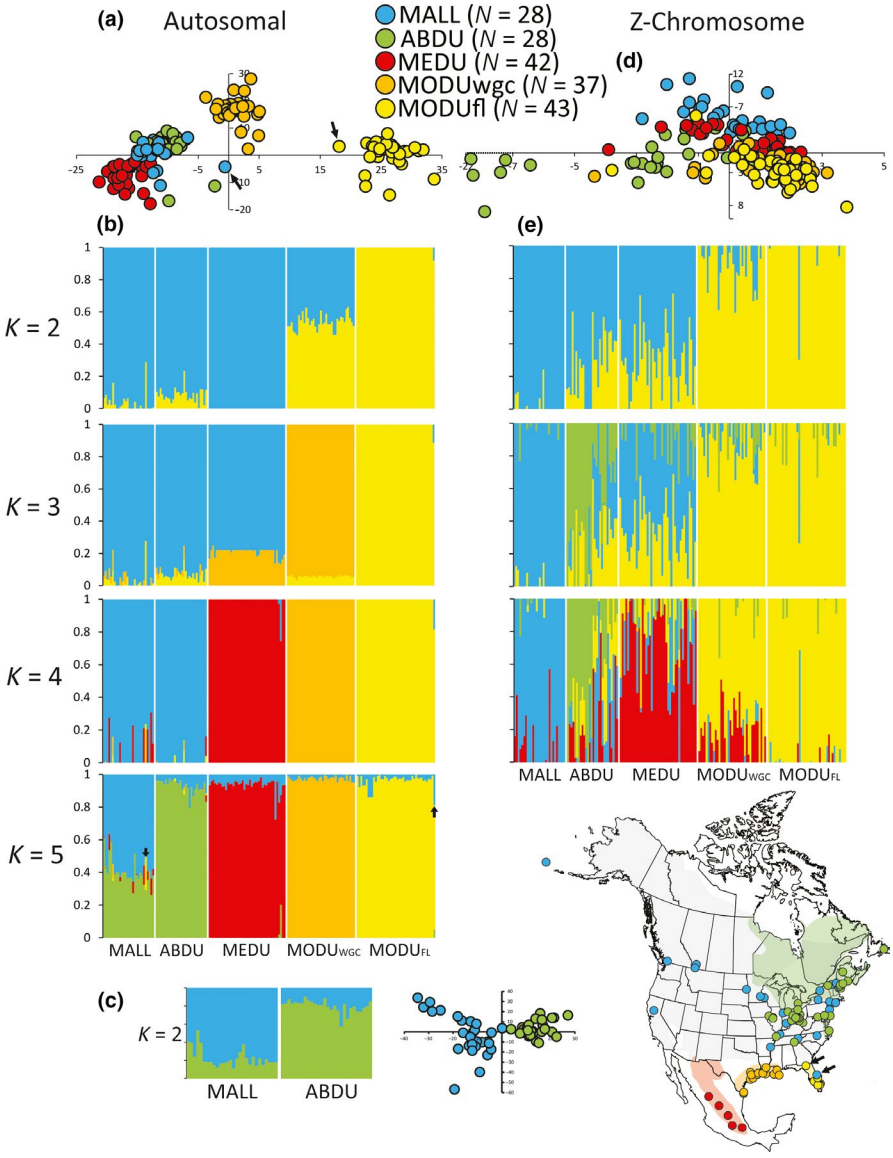

FIGURE 1 Lower right: breeding distributions (adapted from Baldassarre, 2014) and sampling localities are colour coded for mallards (blue), American black ducks (green), Mexican ducks (red), Florida mottled ducks (yellow), and West-Gulf Coast mottled ducks (orange) (Table S1). Principal component analyses (PCA, PC1 on the x-axis, PC2 on the y-axis; top, left) for (a) autosomal and (d) Z-linked loci (*N* = number of samples). ADMIXTURE results for (b) 3,017 autosomal loci and (e) 177 Z-linked loci. (c) PCA and ADMIXTURE analyses for mallards and black ducks only. Arrows highlight two samples consistently identified as admixed across PCA, ADMIXTURE, and fineRADstructure (see Figure 3) analyses

For the 11 new mallards and 28 black ducks, genomic DNA was extracted using a DNeasy Blood & Tissue kit following the manufacturer's protocol (Qiagen). Extractions were quantified using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific Inc.) to ensure a minimum concentration of 0.02 μg/μl. Preparation of multiplexed fragment libraries followed steps outlined in DaCosta and Sorenson (2014) (also see Lavretsky et al., 2015). The samples were pooled in equimolar concentrations, and 151 base pair (bp), single-end sequencing was completed on an Illumina HiSeq 2,500 at the Tufts University Core Genomics Facility. Illumina reads have been deposited in NCBI's Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra; SRA data PRJNA516035 & PRJNA530757).

Raw Illumina reads were demultiplexed and processed using the computational pipeline described by DaCosta and Sorenson (2014) (Python scripts available at http://github.com/BU-RAD-seq/ddRAD-seq-Pipeline) and following steps outlined in Lavretsky et al. (2015). Prior to demultiplexing, 151 bp fragments were truncated to the first 100 bp to coincide with previously collected Mexican duck, mottled duck, and mallard data that were based on 100 bp single-end sequences. The software pipeline clusters filtered reads into putative loci based on sequence similarity and genomic position as determined by BLAST to the reference mallard sequence (Kraus et al., 2011; Huang et al., 2013; chromosomal assembly provided by T. Farault, unpublished data), aligns reads within each putative locus, and infers haplotypes for individual samples at each locus. Genotypes were called as homozygous if ≥93% of reads were identical at variable positions, or heterozygous if a second allele was represented by at least 29% of reads. Between these limits, the genotype was flagged as ambiguous (for more information see DaCosta & Sorenson, 2014). Additionally, to further limit the effect of sequencing error, we required a minimum sequencing depth of five reads to score an allele, such that a minimum of 10 reads was required to score a locus as homozygous or heterozygous; alleles with <5x coverage were scored as missing data. Loci with <20% missing genotypes were retained for downstream analyses, and final output files (e.g., fasta, NEXUS, ADMIXTURE) were generated with custom

python scripts (Lavretsky et al., 2016). Input files for fineRADstructure analyses were created using a custom python script from Stryjewski and Sorenson (2017).

A representative sequence from each ddRAD locus was aligned to the assembled mallard genome (Kraus et al., 2011; Huang et al., 2013; chromosomal assembly provided by T. Farault, unpublished data). This permitted separation of autosomal and Z-linked loci in downstream analyses and the chromosomal positions of ddRAD-seq loci that showed elevated differentiation. Moreover, identifying Z-linked loci permitted us to properly code these loci as having two alleles in males (homogametic sex) and one allele in females (heterogametic sex) when formatting the data for different analyses.

2.2 | Population structure

Composite pairwise estimates of relative divergence (Φ_{ST}) for autosomal and Z-linked ddRAD-seq loci were calculated in the R package PopGenome (Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014) using a concatenated data set for each category; indels were treated as missing data. We used a simple Mantel test as implemented in the ZT program (Bonnet & Van de Peer, 2002) to test whether relative divergence estimates from ddRAD-seq data were significantly correlated with those from previous work based on a much smaller number of loci (Lavretsky, Hernández Baños, et al., 2014a).

Next, we assessed population structure using biallelic single nucleotide polymorphisms (SNPs), with singletons (i.e., rare allele observed in only one individual) excluded, and without a priori information on population or species identity. Population structure was first visualized using a principal component analysis (PCA) in R (i.e., "prcomp"), with scoring of biallelic SNPs as described by Novembre and Stephens (2008). To accommodate Z-linked loci, male genotypes were coded as 0, 0.5, or 1 (heterozygote = 0.5), whereas females were coded as 0 or 1 (also see Lavretsky et al., 2015). Second, maximum likelihood estimates of population assignments for each individual were obtained using ADMIXTURE v.1.3 (Alexander & Lange, 2011; Alexander, Novembre, & Lange, 2009, 2012 Admixture 1.22 Software manual). Autosomal and Z-linked SNPs were formatted for the ADMIXTURE analyses using PLINK (Purcell et al., 2007), following steps outlined in Alexander, Novembre, and Lange (2013). Importantly, the current version of ADMIXTURE permits analysis of sex-linked markers (Shringarpure, Bustamante, Lange, & Alexander, 2016). Analyzing autosomal and Z-linked markers separately, each ADMIXTURE v.1.3 analysis was run with a 10-fold cross validation (CV), and with a quasi-Newton algorithm employed to accelerate convergence (Zhou, Alexander, & Lange, 2011). To limit any possible stochastic effects from single analyses, we ran 100 iterations at each value of *K* (= number of populations; *K* = 1–10). Each analysis used a block relaxation algorithm for point estimation and terminated once the change in the log-likelihood of the point estimations increased by <0.0001. The optimum *K* was based on the average of CV-errors across the 100 analyses per *K* value; however, additional values of *K* were examined to test for further structural resolution. We then used the program CLUMPP v.1.1 (Jakobsson & Rosenberg, 2007) to determine the robustness of the assignments of individuals to populations at each *K* value. The PopHelper (Francis, 2016) R package was used to convert ADMIXTURE outputs into CLUMPP input files at each *K* value. In CLUMPP, we employed the Large Greedy algorithm and 1,000 random permutations. Final admixture proportions for each *K* value and per sample assignment probabilities (*Q* estimates; the log likelihood of group assignment) were based on CLUMPP analyses of all 100 replicates per *K* value.

Recognizing the potential pitfalls of interpreting mixed ancestry based on maximum likelihood assignments in ADMIXTURE alone (Lawson, Van Dorp, & Falush, 2018), we further assessed recent patterns of coancestry using fineRADstructure (Malinsky, Trucchi, Lawson, & Falush, 2018b), which includes RADpainter v 0.1 and finestructure (Lawson, Hellenthal, Myers, & Falush, 2012). In short, fineRADstructure derives a matrix of coancestry coefficients based on the distribution of identical or nearest neighbour haplotypes among samples. Each individual's coancestry at each locus is equally divided among all other individuals with identical haplotypes, or in the case of a unique allele, all other individuals with the "nearest neighbour" haplotype. Thus, rare haplotypes defined by rare SNPs, which are on average of more recent origin (Kimura & Ohta, 1973), contribute the most to the coancestry index, providing a measure that emphasizes recent coancestry. This analysis is also completed without a priori information on population or species identity. A burn-in of 100,000 iterations, followed by 100,000 Markov chain Monte Carlo iterations were completed, followed by tree building using default parameters. To visualize the results, we used the R scripts fineradstructureplot.r and finestructurelibrary.r (available at http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html).

2.3 | Outliers and tests of selection

We used the program LDhat v 2.1 (McVean & Auton, 2007) to estimate per taxon Tajima's D (Tajima, 1989) and Fu and Li (1993) *D* statistics for the concatenated autosomal and Z-linked loci, respectively, to assess possible differences in signatures of selection or demographic history among taxa and between the autosomes and sex chromosome.

Next, to visualize patterns of differentiation across the genome, pairwise per locus Φ_{ST} values were calculated in the R package PopGenome (Pfeifer et al., 2014), and plotted in excel by chromosomal position (i.e., Manhattan plots). BayeScan v. 2.1 (Foll & Gaggiotti, 2008), which has relatively low rates of false positives (<1%) for populations with low overall differentiation (Pérez-Figueroa, García-Pereira, Saura, Rolán-Alvarez, & Caballero, 2010), as is observed in the NW "mallard" radiation (Φ_{ST} estimates range from 0.011 to 0.043; Lavretsky, Hernández Baños, et al., 2014a), was used to test for outlier loci. BayeScan employs a reversible-jump MCMC method by calculating a posteriori probability models with and without selection across loci. The program also distinguishes between positive/diversifying selection ($\alpha > 0$) and balancing/purifying selection ($\alpha < 0$). Analyses included 20 pilot runs of 5,000 steps each,

followed by 100,000 burn-in steps and 10,000 sampling steps with a thinning interval of 10 for a total of 200,000 iterations. The prior odds parameter for the neutral model was set at $\log_{10}(10)$ (Posterior Odds > 1.0). We allowed a probability of false discovery (qval) of 0.05. Three separate analyses were run, one with autosomal and Z-linked loci analyzed together, and two more analyzing autosomal and Z-linked loci separately. Finally, to further assess relationships among samples at outlier loci, haplotype networks were constructed for four loci representing the most significant Φ_{ST} outliers, including one locus on each of four different chromosomes.

Finally, per locus absolute divergence (i.e., d_{XY}; Nei & Li, 1979), nucleotide diversity, and Tajima's *D* were calculated in the R package PopGenome (Pfeifer et al., 2014). We plotted both Tajima's *D* and d_{XY} values against nucleotide diversity. Under a strict scenario of divergence with ongoing gene flow, we expect absolute divergence to be significantly higher for loci that are resistant to introgression, assuming a sufficient amount of time has passed since initial divergence (Cruickshank & Hahn, 2014). Alternatively, under a scenario of post-speciation selection leading to elevated Φ_{ST} at outlier loci, we expect a strong correlation between d_{XY} and nucleotide diversity at neutral loci, whereas outliers should have moderate/low values of diversity and negatively skewed Tajima's *D* compared to genome-wide values (Cruickshank & Hahn, 2014). We calculated Kolmogorov–Smirnov *D* statistics and associated *p*-values (Friedman & Rafsky, 1979) using the "KStest" function in the R program GSAR (Rahmatallah, Zybailov, Emmert-Streib, & Glazko, 2017), with 1,000 permutations, to determine whether the distributions of calculated summary statistics from putative outlier loci are statistically different from putative nonoutlier loci. A significant difference ($p < 0.01$) in distributions would further support the inference that outlier loci have been subject to different evolutionary processes (i.e., selection, genetic drift, gene flow) than the rest of the genome.

2.4 | Testing for gene flow in a phylogenetic context

The program TreeMix version 1.12 (Pickrell & Pritchard, 2012) was used to test for gene flow in a phylogenetic context. Specifically, TreeMix simultaneously estimates a maximum likelihood (ML) species tree and the direction and weight (*w*) of gene flow among taxa based on allele frequencies. An ML species tree without migration is built first, and then migration events are sequentially added until the ln(Likelihood) is maximized. To test between tree models with and without gene flow we applied a likelihood ratio test.

Next, we tested whether shared genetic variation among taxa was best explained by recent ancestry or gene flow by calculating the *f*4-statistic (Keinan, Mullikin, Patterson, & Reich, 2007) as implemented in the fourpop software within TreeMix. Given inferred relationships among four taxa (e.g., A,B; C,D), a significant *f*4-statistic (i.e., *Z* score > |3|; $p < 0.0001$) rejects the assumed relationships: either those relationships are incorrect or there has been admixture in the history of the populations. We acknowledge that the lack of an outgroup that we can safely assume has been unaffected by gene flow complicates inferences regarding the proportional contribution of parental taxa into a putative hybrid taxon based on the *f*4-statisic; however, it remains a useful statistic to test whether the data are consistent with simple genetic drift along a phylogeny (Reich, Thangaraj, Patterson, Price, & Singh, 2009).

Finally, we tested whether any of the focal populations resulted from the admixture of two other populations using *f*3-statistics (Reich et al., 2009) as implemented in the threepop software within TreeMix. Specifically, for a given triplet of populations (e.g., A; B,C), a negative *f*3-statistic with a significant *Z*-score (*Z* score < −3; $p < 0.0001$) suggests that population A is the product of admixture between B and C (Reich et al., 2009). Given the lack of a rooted phylogeny, we tested all possible triplets. We also used TreeMix analyses to test whether putatively non-neutral loci (as defined by BayeScan) showed reduced gene flow and different evolutionary histories as compared to putatively neutral loci. In total, six data sets were analyzed in TreeMix: (a) all autosomal SNPs, (b) SNPs from putatively neutral autosomal loci, (c) SNPs from putatively non-neutral autosomal loci, (d) all Z-linked SNPs, (e) SNPs from putatively neutral Z-linked loci, and (f) SNPs from putatively non-neutral Z-linked loci. Finally, 1,000 bootstrap replicates were used to assess the robustness of the inferred phylogenetic relationships using the "treemix.bootstrap.sh" bash script, and plotted with the "treemix.bootstrap.R" script, both of which are parts of the R package BITE (Milanesi et al., 2017).

3 | RESULTS

We recovered 3,194 ddRAD-seq loci that met our coverage and missing data criteria; 3,017 loci (280,240 aligned base pairs; 44,412 SNPs) were assigned to autosomes and 177 loci (16,171 aligned base pairs; 1,708 SNPs) to the Z-chromosome (Figure S1). These loci were broadly distributed across all chromosomes except chromosome-17, with the number of loci per chromosome proportional to chromosome size (Figure S1). Final data sets comprised loci with an average median sequencing depth of 133 reads per locus per individual (median range = 30–760 reads/locus/individual), and on average, both alleles were scored for 98% of individuals per locus.

3.1 | Population structure

Genetic differentiation among all five NW taxa was generally higher for the Z-chromosome (overall $\Phi_{ST} = 0.12$) as compared to autosomes (overall $\Phi_{ST} = 0.032$), resulting in an overall Z:Autosomal (Z:A) Φ_{ST} ratio of 3.75. Pairwise comparisons yielded similar trends, with Φ_{ST} values for comparisons between mallards and the monochromatic taxa ranging from 0.010 to 0.048 for all autosomal loci combined and from 0.088 to 0.21 for all Z-linked loci; Z:A Φ_{ST} ratios ranged from 4.4 to 13.7 (Figure 2). In pairwise comparisons between monochromatic taxa, the Z:A Φ_{ST} ratio ranged from 0.82 to 4.64; only in the comparison of the two mottled duck subspecies (Z:A = 0.82) and perhaps between Mexican ducks and either mottled duck subspecies (Z:A ~ 2.0) was this ratio close to neutral expectations (i.e., Φ_{ST} Z:Autosomal ≤ 1.33; Lavretsky et al., 2015,

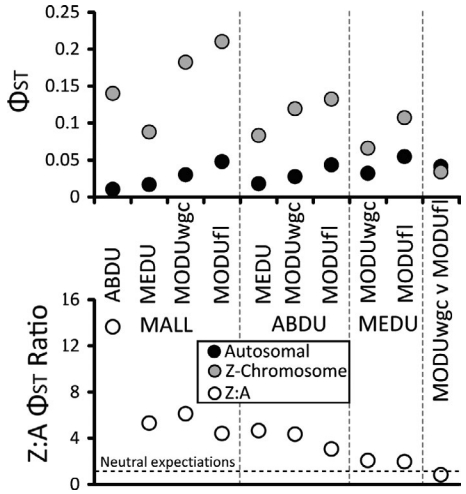

FIGURE 2 Top panel: composite pairwise Φ_{ST} estimates for 3,017 autosomal and 177 Z-linked loci for mallards (MALL), American black ducks (ABDU), Mexican ducks (MEDU), West-Gulf Coast mottled ducks ($MODU_{WGC}$), and Florida mottled ducks ($MODU_{FL}$). Bottom panel: pairwise Z:Autosomal Φ_{ST} ratios with the dotted line denoting the neutral expectation under assumptions of constant population sizes and equal variance in reproductive success in males and females, respectively

2016; Figure 2). Pairwise Φ_{ST} values based on autosomal ddRAD loci were strongly and significantly correlated (based on a simple Mantel test) with those obtained from both mtDNA and 17 nuclear introns (Figure S2; Lavretsky, Hernández Baños, et al., 2014a). In contrast, Φ_{ST} values for Z-linked ddRAD loci were not correlated with values obtained from any of the other data sets (Figure S2).

Prinipal components analysis and ADMIXTURE analyses of autosomal loci were based on 15,687 biallelic SNPs (out of 44,412 total SNPs), after excluding singletons. The first two principal component axes clearly separate the two mottled duck populations from each other and from the other taxa, whereas Mexican ducks, black ducks, and mallards clustered adjacent to one another (Figure 1a). ADMIXTURE analysis of autosomal loci identified an optimal value of K = 2 (Figure S3), at which Florida mottled ducks are distinguished from mallards, black ducks, and Mexican ducks, whereas WGC mottled ducks are mixed, with 46%–63% assignment to the Florida population (Figure 1b). However, there was a linear increase in cross-validation (CV) values, indicating that higher values of K might reveal biologically meaningful information (Janes et al., 2017). Increasing values of K up to five provided additional, interpretable resolution of population structure in which individuals of each monochromatic taxon are assigned to a different cluster (Figure 1b). At K = 4, four individual samples are assigned to more than one group (i.e., ≥10% assignment to a second group), including two black ducks, one Mexican ducks, and one Florida mottled duck. At K = 5, eleven individual samples are assigned to more than one group, including four black ducks, three Mexican ducks, and four Florida mottled ducks. Finally, analyzing black ducks and mallards separately results in better discrimination of these two species (Figure 1c).

For Z-linked loci, analyses were based on 359 biallelic SNPs, after excluding singletons. With this smaller data set, the four species show some differentiation along the first two PC axes (Figure 1d), whereas the two mottled duck subspecies are largely overlapping. At K = 3 (the optimal value; Figure S3), individual mottled ducks and mallards are generally assigned to different groups, whereas many black ducks and all Mexican ducks are assigned to two or three groups (Figure 1e). Additional resolution is apparent at K = 4, at which Mexican ducks tend to have the highest values for assignment to a fourth population. Values of K = 4 and 5 were nearly equally probable based on cross-validation values (Figure S3), but no additional interpretable resolution was achieved at $K \geq 5$ for the Z-linked data set.

Analysis of autosomal loci in fineRADstructure, which emphasizes recent coancestry, produced results similar to the ADMIXTURE analysis with K = 5, with five well-supported groups comprising the five a priori taxa in our analysis (Figure 3). Although recent coancestry is greatest between mallards and black ducks, the two species form distinct groups in the analysis, and no individual black duck has relatively higher coancestry with mallards as compared to other black duck samples. Thus, we conclude that higher coancestry between black ducks and mallard is likely the result of more recent common ancestry, but might also reflect episodes of historical admixture. Also consistent with the ADMIXTURE analysis, Florida mottled ducks have the lowest levels of recent coancestry with other taxa, and particularly with Mexican ducks, indicating a longer period of time since isolation, lower rates of ongoing gene flow, and/or smaller population sizes resulting in greater genetic drift. Whereas ADMIXTURE results indicated up to eleven samples as potentially being admixed, only two samples show evidence of recent mixed ancestry in the fineRADstructure results: a Florida mottled duck with elevated mallard coancestry and a mallard, also collected in Florida, with elevated mottled duck coancestry. Both of these samples show similar evidence of admixture in autosomal ADMIXTURE analysis (arrows in Figures 1a,c and 3). Thus, ADMIXTURE appears to have overestimated the number of recent generation hybrids (Lawson et al., 2018) in comparison to the fineRADstructure results, in which none of the black duck, Mexican duck, and WGC mottled duck samples appear to be admixed. Finally, we note that six mallards sampled in different eastern states (i.e., New York, North Carolina, New Jersey, Florida, and Tennessee) show a much higher than average level of coancestry with each other as compared to the average level for all mallards (Figure 3).

3.2 | Outlier loci

BayeScan identified the same set of autosomal loci as outliers whether autosomal loci were analyzed alone or together with Z-linked loci. However, no outliers were found when analyzing Z-linked loci only, likely the result of the elevated average level of differentiation observed for the Z-chromosome (Figure 2), a context in which BayeScan is less effective (Pérez-Figueroa et al., 2010). While we acknowledge that some caution is required in making inferences about Z-linked loci when analyzed together with autosomal markers (Lavretsky et al., 2015), results from such analyses may

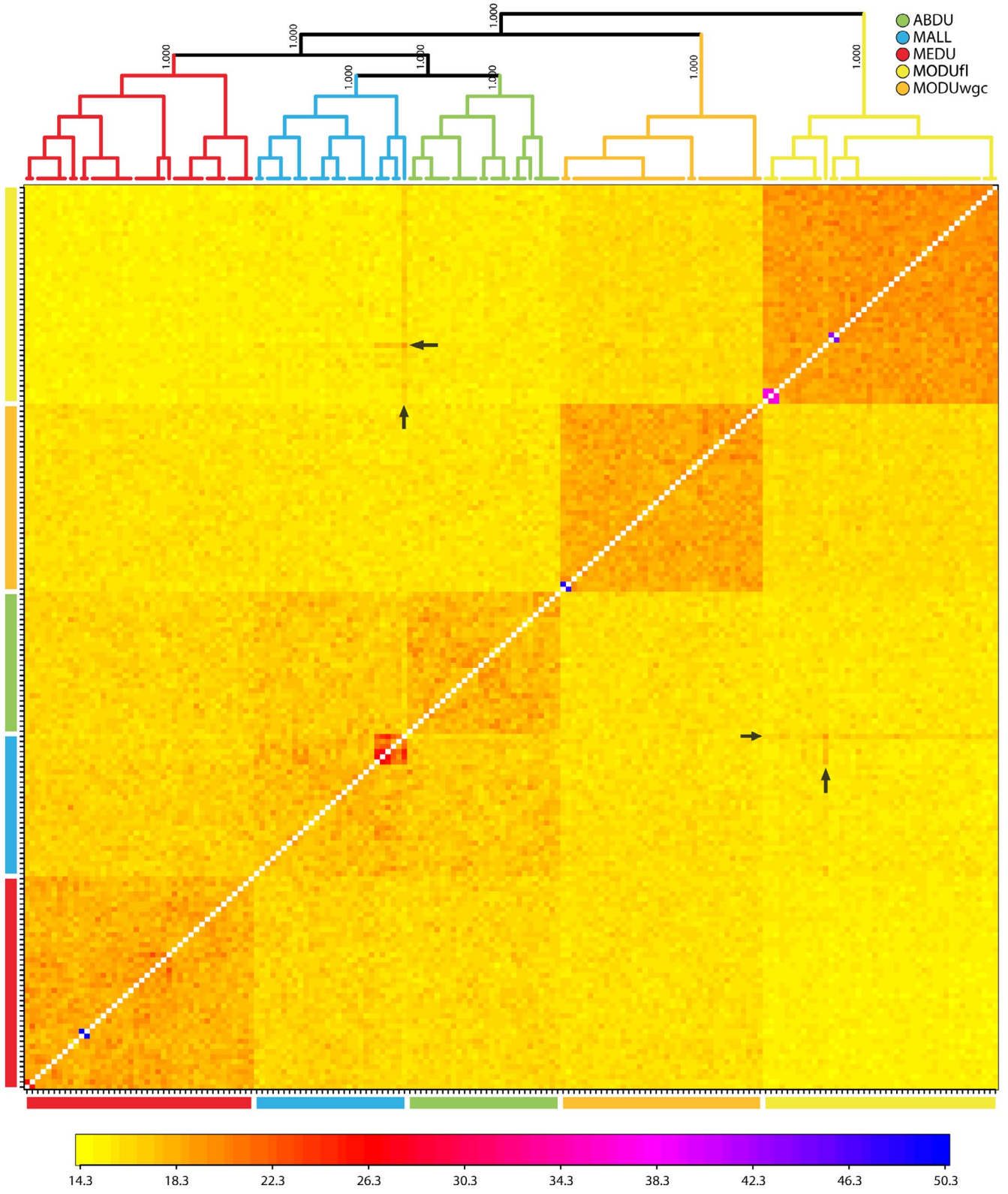

FIGURE 3 Coancestry matrix from fineRADstructure based on autosomal ddRAD-seq loci. Pairwise coefficients of coancestry are colour coded from low (yellow) to high (blue). The dendrogram depicts a clustering of individual samples based on the pairwise matrix of coancestry coefficients. Arrows highlight two samples identified as admixed across PCA, ADMIXTURE (Figure 1), and fineRADstructure analyses

be informative when the same outliers are consistently recovered across multiple pairwise comparisons of different species pairs. In the combined analysis, a set of eight Z-linked loci were identified as significant outliers in two or more of the four comparisons between mallards and each of the monochromatic taxa; three of these loci were outliers in all four comparisons (Table S2). All of the Z-linked

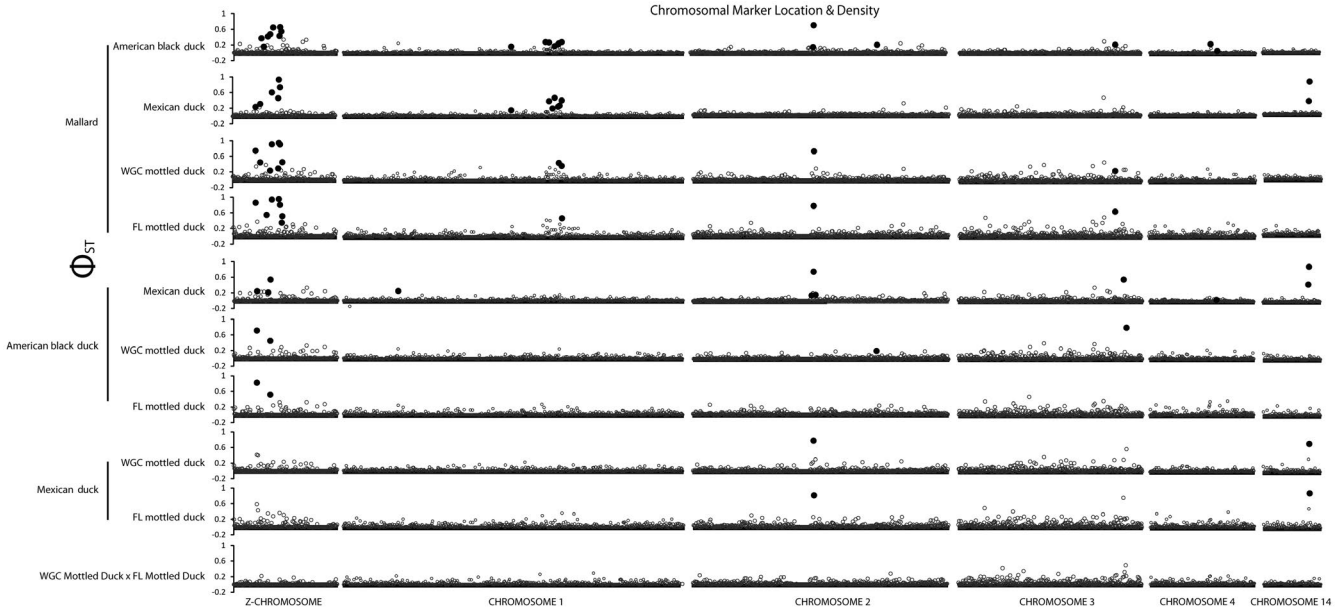

FIGURE 4 Distribution of Φ_{ST} values for chromosomes with significant outliers (chromosomes Z, 1, 2, 3, 4, and 14) for pairwise comparisons between mallards (MALL), American black ducks (ABDU), Mexican ducks (MEDU), West-Gulf Coast mottled ducks ($MODU_{WGC}$), and Florida mottled ducks ($MODU_{FL}$). Black dots denote markers identified by BayeScan as putatively under positive selection (Table S2)

outliers fell within an ~21 Mbp region (positions 1.7E7–3.8E7 bp; Figure 4 and Figure S4). In pairwise comparisons among the monochromatic taxa, BayeScan identified two or three Z-linked outliers in comparisons of black ducks with each of the other three taxa (Figure 4 and Figure S4, Table S2). Mallards and black ducks were the only species with negative values of Tajima's *D* and Fu & Li's *D* for all Z-chromosome loci combined (Table 1), which is consistent with either positive selection or population expansion.

While combining autosomal and Z-linked loci in the same analysis has the potential to produce false positives for Z-linked loci, it should decrease the false positive rate for autosomal loci. BayeScan identified a smaller percentage of autosomal loci (0.007%–0.05%) as compared to Z-linked loci (3.4%–4.5%) as putatively under diversifying selection. Several autosomal regions included outlier loci in multiple pairwise comparisons. In particular, one to six outlier loci depending on the pairwise comparison, were identified within an ~11 Mbp region (1.0E8–1.2E8 bp) on chromosome 1 when comparing mallards to each of the monochromatic taxa. An outlier locus on chromosome 14 (position ~1.6E7; also see Lavretsky et al., 2015) was detected in all four comparisons involving Mexican ducks, suggesting directional selection at this or a linked locus in Mexican ducks only. Another locus on chromosome 2 (starting position ~ 6.6E8) was identified as an outlier when comparing mallards or Mexican ducks to black ducks or mottled ducks. Additional outliers were detected in pairwise comparisons involving black ducks (e.g., Z-chromosome, chromosome 3; Figure 4 and Figure S4). Haplotype networks for the four most extreme outlier loci on chromosomes 1 (position 111,050,764), 2 (position 65,815,089), 14 (position 15,899,148), and the Z-chromosome (position 35,470,645), respectively, show allele frequency differences but no fixed differences or species diagnostic SNPs (Figure 5). Finally, the pairwise comparison of mottled duck subspecies was the only comparison that did not yield outlier loci; these taxa were also the only ones with positive values of Tajima's *D* and Fu & Li's *D* (Table 1).

We observed strong positive correlations between values of d_{XY} and nucleotide diversity for both autosomal and Z-linked nonoutlier loci (Figure 6). With the exception of the locus from chromosome 14 (position ~ 1.6E7), at which Mexican ducks are nearly fixed for an allele that is rare or absent in the other four taxa (Figures 4 and 5), autosomal outliers were largely characterized by relatively low d_{XY} values and by low nucleotide diversity. Likewise, Z-linked loci had generally lower values of both d_{XY} and nucleotide diversity as compared to autosomal loci. However, several Z-chromosome loci representing outliers between mallards and each of the monochromatic taxa were characterized by relatively high values of d_{XY} relative to nucleotide diversity (Figure 6). Finally, although distributions of nucleotide diversity and Tajima's *D* were statistically similar between outlier and nonoutlier loci (all KS-test *p*-values ≥ 0.50) for both autosomal and Z-linked loci (Figure S5), one apparent exception was a set of loci within the outlier region on chromosome 1. This region was characterized by low nucleotide diversity and negative values of Tajima's *D* in mallards only, suggesting the possibility of directional selection affecting this chromosomal region in the mallard lineage.

3.3 | Testing for gene flow in a phylogenetic context

TreeMix analyses were used to estimate the direction and magnitude of gene flow among taxa. First, unrooted phylogenies were generally robust, with bootstrap support ≥90% across all nodes (Figure 7). Next, treeMix analyses of all SNPs or nonoutlier SNPs included up to three or four connections indicating gene flow between

TABLE 1 Nucleotide and haplotype diversity for 3,017 autosomal and 177 Z-chromosome loci for five New World Mallard Complex taxa. Estimated census size and *Ne* (= Watterson's $\theta_{autosomal}/4\ \mu$) using all ddRAD autosomal markers and the previously estimated nuclear mutation rate of 1.2×10^{-9} substitutions/site/year for autosomal markers in ducks (Peters, Zhuravlev, Fefelov, Humphries, & Omland, 2008). (BELOW) Estimates of Watterson's θ, Tajima's *D*, and Fu and Li *D* statistic for all five taxa across autosomal or Z-linked loci

	Nucleotide diversity		Haplotype diversity			
	Autosomal	Z-chromosome	Autosomal	Z-chromosome	Census size	Autosomal *Ne*
Mallard	0.0067	0.0029	0.32	0.17	11,640,000[a]	3,239,965
American black duck	0.0068	0.0032	0.33	0.14	540,600[a]	3,403,731
Mexican duck	0.0066	0.0028	0.32	0.15	55,500[b]	2,196,592
WGC Mottled duck	0.0065	0.0025	0.31	0.12	135,000[c]	1,921,988
FL Mottled duck	0.0062	0.0026	0.30	0.12	35,000[c]	1,225,744

	Autosomes			Z-chromosome		
	Watterson's θ	Tajima's *D*	Fu and Li *D* statistic	Watterson's θ	Tajima's *D*	Fu and Li *D* statistic
Mallard	1,555	−1.19[d]	−2.07[d]	46	−1.19[d]	−1.96[d]
American black duck	1,634	−1.22[d]	−1.82[d]	29	−1.25[d]	−2.26[d]
Mexican duck	1,054	−1.14[d]	−1.23[d]	32	−0.56	−0.31
WGC Mottled duck	923	−0.71	−0.41	23	−0.43	0.15
FL Mottled duck	588	−0.26	0.33	16	0.037	0.28

[a]U.S. Fish and Wildlife Service 2015;
[b]Perez-Arteaga et al., 2002;
[c]Delany & Scott, 2006;
[d]Denotes significant values of Tajima's *D* or Fu and Li *D* statistic.

pairs of taxa (Figure 7); in contrast, adding gene flow to the model did not significantly improve the likelihood for trees based on either autosomal or Z-linked SNPs from outlier loci (Table S3). Likewise, *f*4-statistics provided no evidence of gene flow for autosomal and Z-linked outliers (Table S4). While these results are consistent with resistance to gene flow at outlier loci, it is also possible that they are simply a consequence of the substantially smaller number of SNPs in these data partitions (Table S3). For all SNPs and nonoutlier SNPs, gene flow was inferred for several pairs of taxa, including gene flow from Mexican ducks into WGC mottled ducks, from black ducks into mottled ducks, and from mallards into either black ducks (SNPs from autosomal nonoutlier loci), or Mexican ducks (all Z-linked SNPs) (Figure 7; Table S3). In addition, gene flow from WGC mottled ducks into mallards was inferred when analyzing SNPs from Z-linked nonoutlier loci only.

Results of the "four population tests" (*f*4-statistics) for autosomal loci were generally consistent with the above inferences about gene flow. For each set of four populations ($n = 5$), there are three unique unrooted trees resulting in a total of 15 unique tests. Of these, the six tests with the strongest departures from null expectation were based on trees in which the two mottled duck populations were separated; in these cases, the significant statistical result is presumably an artifact of assuming an incorrect tree, making "gene flow" necessary to counteract the erroneous assumption and explain the similarity of the two mottled duck populations. Of the remaining nine tests, five were statistically significant, including three tests consistent with gene flow between Mexican ducks and WGC mottled ducks, and three tests consistent with gene flow between black ducks and Florida mottled ducks (one test was consistent with both of these connections; Table S4). A significant result in any individual test may be the consequence of assuming an incorrect tree, but the recurrence of gene flow inferences for the two population pairs noted above and the geographic proximity of populations within each pair, respectively, suggest a legitimate signal of historical gene flow.

For the nine trees that did not separate the mottled duck populations, none of the tests based on Z-linked loci or autosomal outlier loci produced a significant result, which may be a function of limited data, rather than strong evidence against gene flow. Finally, none of the "three population" tests (*f*3-statistics) yielded statistically significant evidence of admixture (Table S4).

4 | DISCUSSION

4.1 | Population structure and gene flow within a recent radiation

Recently diverged species share ancestral polymorphisms throughout much of their genomes and are unlikely to have many fixed differences. Therefore, detecting portions of the genome with diagnostic allele frequency differences or that may be indicative of diversifying selection in independent evolutionary lineages

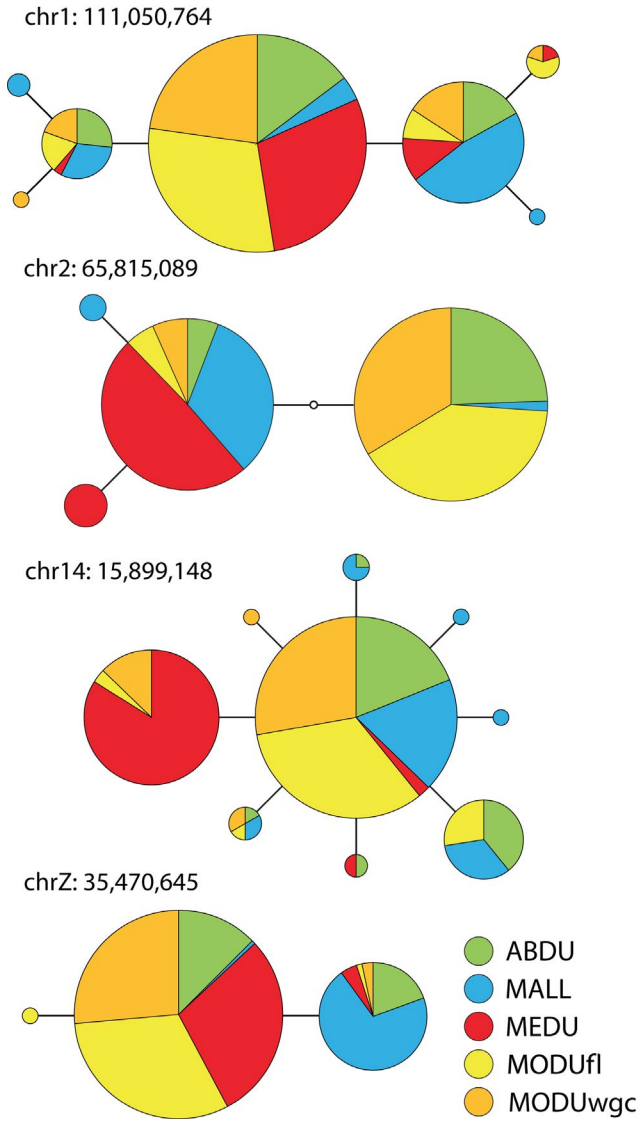

FIGURE 5 Haplotype networks depicting allelic variation for the four most extreme outlier loci on chromosomes 1 (position 111,050,764), 2 (position 65,815,089), 14 (position 15,899,148), and Z (position 35,470,645) (Figure 4). Each haplotype network includes two alleles per individual for each autosomal locus and one (females) or two (males) alleles for the Z-linked locus

requires the sampling of a large number of loci (Funk et al., 2012; Oyler-McCance et al., 2016; Rice et al., 2011). In this study, the thousands of loci recovered using ddRAD-seq did not include any perfectly diagnostic SNPs, but did include a sufficient number of loci with differences in allele frequencies to allow multi-locus assignment of individuals to populations. Importantly, complementing recent studies showing multi-locus genetic discrimination of mallards and Mexican ducks (Lavretsky et al., 2015) and the two mottled duck subspecies (Peters et al., 2016), respectively, we demonstrate that mallards and black ducks are also distinguishable using ddRAD-seq data (Figures 1 and 3).

Although ADMIXTURE and fineRADstructure were largely consistent in their assignment of individuals to populations, the programs suggest different inferences about the prevalence of individuals with recent mixed ancestry. ADMIXTURE results suggest that several individuals among the monochromatic taxa show some evidence of mixed ancestry, but only one mottled duck from Florida showed evidence of elevated coancestry with mallards in the fineRADstructure analysis (Figures 1 and 3). Given the greater sensitivity of fineRADstructure to recent ancestry (Malinsky, Trucchi, et al., 2018b), we conclude that putative signals of admixture in the ADMIXTURE analysis are likely due to shared ancestral variation (i.e., ILS), and that the occurrence of hybrid individuals in our data set is low, consistent with other recent analyses of the NW "mallards" (Ford, Selman, & Taylor, 2017; Lavretsky et al., 2015; Peters et al., 2016). We note, however, that we intentionally avoided sampling in geographic regions where mallards and Mexican ducks come into contact, and Peters et al. (2016) only examined mottled ducks that were phenotypically "pure," though this sampling protocol would not necessarily exclude subsequent generation backcrosses. Finally, these results are consistent with recent studies demonstrating that evidence of hybrid ancestry from population assignment programs (e.g., ADMIXTURE, STRUCTURE) should be interpreted cautiously and confirmed using multiple methods (Lawson et al., 2018).

In general, our results are at odds with expectations for a group of birds known for high rates of hybridization (Baldassarre, 2014; Ottenburghs, Ydenberg, Van Hooft, Van Wieren, & Prins, 2015). Secondary contact between various monochromatic taxa and the dichromatic mallard has long been assumed to result in high rates of hybridization (Champagnon et al., 2013; Guay & Tracey, 2009; Lavretsky, Hernández Baños, et al., 2014a; US Fish & Wildlife Service, 2013), and in some cases, concern about the possibility of genetic extinction (Rhymer, 2006; Rhymer & Simberloff, 1996). Moreover, high rates of gene flow have been invoked to explain similar levels of molecular variation despite substantial differences in known census sizes (Table 1; Avise et al., 1990; Lavretsky et al., 2015; Lavretsky, McCracken, et al., 2014b; McCracken et al., 2001; Peters et al., 2014). Our results, however, suggest that none of the sampled groups are extensively admixed based on *f*3-statistics (Table S4), let alone being at risk for merging into a hybrid swarm. Whereas TreeMix identified gene flow from mallards into either Mexican ducks or black ducks in two different data partitions (Figure 7; Table S3), *f*4-statistics were equivocal with respect to rejecting a null hypothesis of no gene flow involving mallards (Table S4). In addition, other recent studies have detected a relatively low frequency of hybrids and/or recent backcrosses—for example, between mallards and either mottled ducks (~5%; Peters et al., 2016; Ford et al., 2017) or Mexican ducks (~2%; Lavretsky et al., 2015). Thus, although hybridization is known to occur between mallards and each of the monochromatic species, our results suggest that contemporary gene flow and introgression may be lower than assumed.

In contrast to results for mallards, both TreeMix analyses and *f*4-statistics were suggestive of greater gene flow between geographically proximate pairs of monochromatic taxa: Mexican ducks and WGC mottled ducks, and black ducks and FL mottled ducks, respectively (Figure S5; Tables S3 and S4).

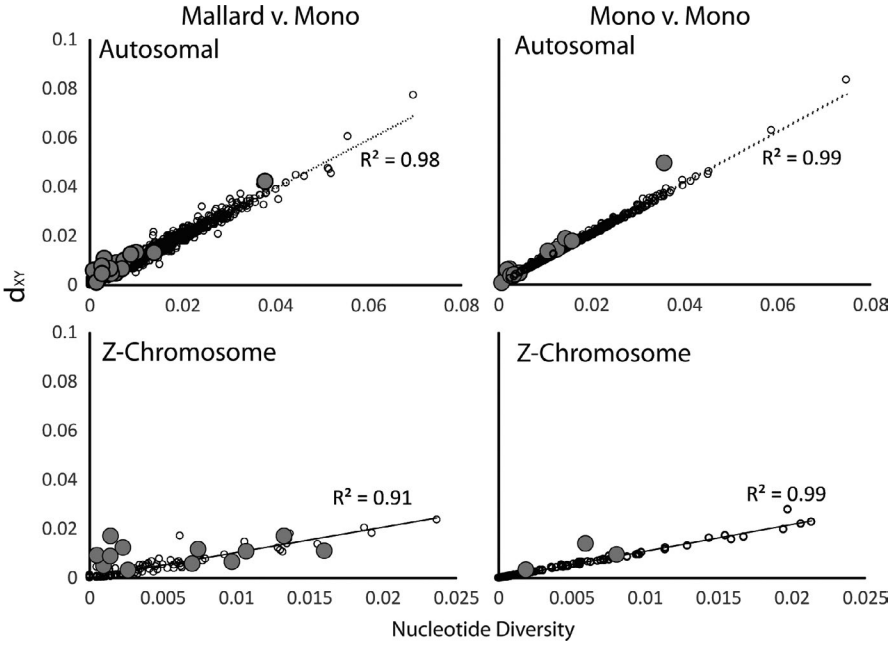

FIGURE 6 Relationship between absolute divergence (d_{XY}) and average nucleotide diversity for comparison of mallards versus monochromatic species (left panels) and comparisons among monochromatic species (right panels). Average per locus values across taxa (nucleotide diversity) and comparisons (d_{XY}) are plotted for 3,017 autosomal and 177 Z-linked loci. Grey dots denote loci identified by BayeScan as putatively under positive selection between mallards and any of the monochromatic species or between pairs of monochromatic species (Table S2)

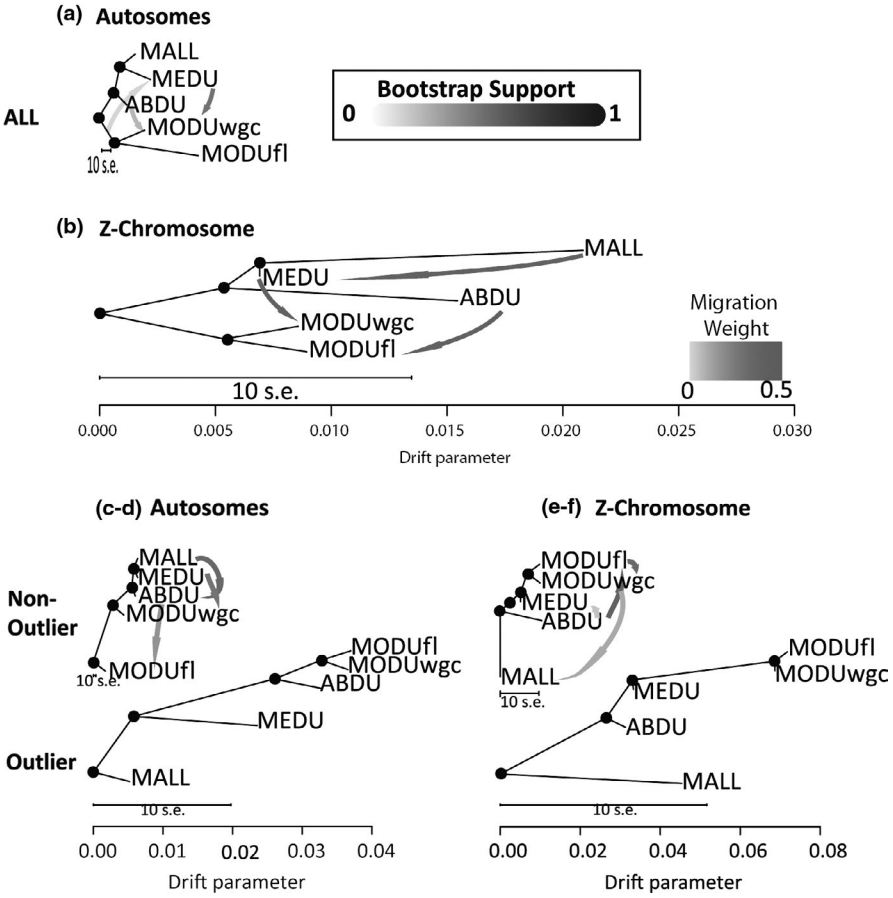

FIGURE 7 TreeMix maximum likelihood species trees based on biallelic SNPs for six data sets: (a) all autosomal SNPs, (b) all Z-linked SNPs, (c) putatively neutral autosomal SNPs, (d) putatively non-neutral autosomal SNPs, (e) putatively neutral Z-linked SNPs, and (f) putatively non-neutral Z-linked SNPs. Arrows depict possible gene flow events that increase the likelihood of each tree, color coded by migration weight (Table S3). Loci identified as putatively under diversifying selection in BayeScan analyses (Table S2) were used for autosomal outlier analyses, whereas partitioned Z-chromosome phylogenies were based on loci within and outside the ~21 Mbp outlier region (1.7E7–3.8E7 bp; Figures 4 and 5), respectively

4.2 | Genomics of mallards and black ducks

Among the pairwise comparisons within the NW mallard group, black ducks and mallards are the most similar genome-wide (Figures 1–3; Figure S4). Given mostly small allele frequency differences across their genomes, discriminating between black ducks and mallards based on genetic data has been an elusive goal. Their genetic similarity has been attributed to both recent ancestry (~180,000 years before present; Lavretsky, Hernández Baños, et al., 2014a) and high rates of ongoing hybridization (Ankney et al., 1986; Avise et al., 1990; Johnson

& Sorenson, 1999; Lavretsky et al., 2015; Lavretsky, McCracken, et al., 2014b; McCracken et al., 2001). Although our sample sizes were not large enough to robustly test for current hybridization, the lack of obvious hybrids or backcrossed individuals in the data set, as well as nonsignificant *f*3-statistics across mallard and black duck comparisons (Table S4), suggest little direct evidence of contemporary gene flow. Conversely, one TreeMix analysis (i.e., autosomal nonoutlier SNPs only; Figure 7c) assigned a migration weight of 0.40 from mallards into black ducks (Table S3), suggesting that the two may have experienced gene flow in the recent past. Thus, the high level of coancestry between mallards and black ducks (Figure 3) may be the combined result of retained ancestral variation and bouts of gene flow throughout the divergence process. Rather than being each other's closest relatives from a phylogenetic perspective, the genome-wide similarity of mallards and black ducks relative to the other taxa may be due to substantial introgression following secondary contact, perhaps followed by a subsequent reduction in gene flow over time (e.g., Stryjewski & Sorenson, 2017). Future work would benefit from a landscape approach to further test for evidence of hybridization and introgression by comparing geographic regions with varying proportions of mallards and black ducks.

4.3 | Outlier distribution and genomic diversity in the NW "mallard" group

In addition to providing the power to genetically distinguish these closely related dabbling duck species, the genomic coverage offered by ddRAD-seq was sufficient to identify regions of elevated divergence that may have been involved in speciation and/or the phenotypic diversification of the group (Figure 4; Figure S4). By comparing multiple species, we were able to identify genomic regions likely affected by divergent selection in a specific lineage or taxon.

Regions on the Z-chromosome and chromosome 1, respectively, included multiple, tightly clustered outliers that were detected in comparisons between the dichromatic mallard and each of the monochromatic species (Figures 3 and 5). Moreover, despite low values of d_{XY}, the chromosome 1 outliers were characterized by low nucleotide diversity and negative Tajima's *D* in mallards; in contrast, there was high variance in both d_{XY} and nucleotide diversity for the same loci in the monochromatic taxa (Figure S5). Given that these outliers were less prominent or absent in comparisons between monochromatic taxa, it seems likely that these genomic regions were influenced by directional selection within the lineage leading to contemporary mallards.

Another case of apparent selection was found within the Mexican duck lineage on chromosome 14. One or two loci within a small region of chromosome 14 were identified as outliers in each pairwise comparison involving Mexican ducks. Furthermore, this locus was nearly fixed in Mexican ducks for an allele that was rare or absent in the other taxa (Figure 5). Thus, we conclude that selection acting within Mexican ducks has likely influenced this region on chromosome 14. A similar pattern was found in a chromosome 2 outlier, at which an allele nearly fixed in both Mexican ducks and mallards was rare in the remaining taxa (Figure 5).

The Z-chromosome had generally higher levels of overall differentiation (Figure 2), and exhibited patterns of pairwise differentiation that were uncorrelated with pairwise estimates for autosomal loci, introns, and mitochondrial DNA (Figure S2). Moreover, the ratio of Z-chromosome to autosomal Φ_{ST} was highest and deviated most strongly from neutral expectations in pairwise comparisons involving mallards (Figure 2; e.g., ≥1.33; Caballero, 1995; Whitlock & McCauley, 1999; Dean, Harrison, Wright, Zimmer, & Mank, 2015). Although demographic processes can result in skewed ratios (Van Belleghem et al., 2018), modelling of strictly neutral divergence suggests that $\Phi_{ST\ Z:Autosomal}$ ratios > 5 could be generated only if the effective population size of Z-linked loci is 10%–20% the effective size for autosomal loci (as compared to the standard expectation of 75%), which is unlikely in ducks (Lavretsky et al., 2015). Although sex chromosomes are often found to harbour outliers thought to be important in speciation (Dhami, Joseph, Roshier, & Peters, 2016; Ellegren et al., 2012; Lavretsky et al., 2015; Martin et al., 2013; Minvielle, Ito, Inoue-Murayama, Mizutani, & Wakasugi, 2000; Phadnis & Orr, 2009; Pryke, 2010; Ruegg, Anderson, Boone, Pouls, & Smith, 2014; Saether et al., 2007; Sutter, Beysard, & Heckel, 2013), caution is warranted in making conclusions about the relative importance of loci with different modes of inheritance based on data representing a relatively small proportion of the genome. In particular, it is important to consider that higher linkage disequilibrium in sex chromosomes relative to autosomes (Bergero & Charlesworth, 2009) increases the probability of capturing outlier loci when using reduced representation methods (e.g., ddRAD-seq; Samuk et al., 2017).

Overall, we find intriguing evidence that selection has contributed to enhanced divergence among mallards and closely related species in a small number of genomic regions. In some cases, patterns of nucleotide diversity and Tajima's *D* are consistent with the effects of positive or directional selection, especially in the mallard (chromosomes 1 & Z) and Mexican duck (chromosome 14). However, the recent diversification of these species (NW Mallard taxa diverged ~300,000 years before present; Lavretsky, Hernández Baños, et al., 2014a) and consequently strong correlation between nucleotide diversity and d_{XY} (Figure 6; (Martin, Davey, & Jiggins, 2015) limits our ability to make inferences regarding divergence with gene flow versus post-speciation selection in generating these outliers (Cruickshank & Hahn, 2014).

5 | CONCLUSIONS

Estimates of Φ_{ST} from a large ddRAD-seq data set were nearly identical to estimates obtained from 17 nuclear introns for the same species (Lavretsky, Hernández Baños, et al., 2014a), suggesting that previous studies with fewer markers provided accurate estimates of overall genomic differentiation. In contrast to previous research, however, the ~150-fold increase in number of loci provided sufficient power to assign individuals to their respective taxonomic groups and identify putative hybrid individuals (Figures 1 and 3). In general, our results suggest a lack of widespread contemporary gene flow, challenging long-standing concerns about the possible genetic extinction of the

NW monochromatic dabbling ducks via introgressive hybridization with mallards (Rhymer, 2006; Rhymer & Simberloff, 1996). Finally, despite the limited sampling of the genome achieved in a ddRAD-seq data set and the expectation of sampling mostly neutral variation (Hoban et al., 2016; Lowry et al., 2016), we identified several small genomic regions as putative outliers likely affected by selection either before or after speciation (Figures 4, 6, 7; Figure S4 and Table S2). It is likely that other interesting genomic regions went undetected due to the low genomic coverage (>0.03%) and spacing between (~350 kbp on average) the ddRAD-seq loci in our data set (Catchen et al., 2017; Lowry et al., 2016). We also acknowledge that outlier detection methods are biased towards loci of large effect, and likely miss those with small additive effects (Harrisson, Pavlova, Telonis-Scott, & Sunnucks, 2014). Thus, in addition to strong diversifying selection, as well as the contributions of simple genetic drift (Lavretsky et al., 2015; Peters et al., 2016), differences between the mallard and its close relatives may include quantitative traits influenced by many loci (Rockman, 2012; Stölting et al., 2013; Yeaman & Whitlock, 2011). Future studies will require whole genome sequence data for population samples to more definitively determine the relative contributions of these mechanisms to the diversification of the "mallard" clade.

ACKNOWLEDGEMENTS

We are grateful to Ken Richkus (U.S. Fish and Wildlife Service) and the North American Flyway "wing bees" for providing samples for this study. We also thank Evangeline Shank, Michaela Woods, and Brittany Bowers for their contributions in sample preparation. This research was funded by the American Black Duck Joint Venture, Safari Club International El Paso Chapter, the James A. Kushlan endowment in Waterbird Biology and Conservation at the University of Miami, the College of Sciences and Mathematics at Wright State University, and the University of Texas at El Paso.

AUTHOR CONTRIBUTION

P.L. & J.L.P. conceptualized and collected data. P.L., J.L.P., M.D.S., J.M.D. & K.G.M. analyzed and supported data acquisition. P.L., J.L.P., M.D.S., J.M.D., & K.G.M. equally contributed to the writing of this manuscript.

DATA ARCHIVING

Illumina ddRAD-Seq Reads: NCBI's Sequence Read Archive data PRJNA516035 & PRJNA530757 Other data files (e.g. FASTA files; ADMIXTURE, and fineRADstructure input files): Dryad accession doi:10.5061/dryad.5gk37t8.

ORCID

Philip Lavretsky https://orcid.org/0000-0002-5904-8821
Jeffrey M. DaCosta https://orcid.org/0000-0002-8590-2083

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Aldrich, J. W., & Baer, K. P. (1970). Status and Speciation in the Mexican Duck (*Anas diazi*). *The Wilson Bulletin*, *82*, 63–73.

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*, 246. https://doi.org/10.1186/1471-2105-12-246

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

Alexander, D. H., Novembre, J., & Lange, K. (2012). Admixture 1.22 Software manual.

Andrew, R. L., & Rieseberg, L. H. (2013). Divergence is focused on few genomic regions early in speciation: Incipient speciation of sunflower ecotypes. *Evolution*, *67*, 2468–2482. https://doi.org/10.1111/evo.12106

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*, 81–92. https://doi.org/10.1038/nrg.2015.28

Ankney, C. D., Dennis, D. G., Wishard, L. N., & Seeb, J. E. (1986). Low genic variation between black ducks and mallards. *The Auk*, *103*, 701–709.

Avise, J. C., Ankney, D. C., & Nelson, W. S. (1990). Mitochondrial gene trees and the evolutionary relationship of mallard and black ducks. *Evolution*, *44*, 1109–1119. https://doi.org/10.1111/j.1558-5646.1990.tb03829.x

Baldassarre, G. (2014). *Ducks, geese, and swans of north*. Baltimore, MD: America Johns Hopkins University Press.

Bergero, R., & Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends in Ecology & Evolution*, *24*, 94–102. https://doi.org/10.1016/j.tree.2008.09.010

Bonnet, E., & Van de Peer, Y. (2002). zt: A software tool for simple and partial Mantel tests. *Journal of Statistical Software*, *7*, 1–12.

Caballero, A. (1995). On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics*, *139*, 1007–1011.

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, *17*(3), 362–365. https://doi.org/10.1111/1755-0998.12669

Champagnon, J., Crochet, P.-A., Kreisinger, J., Čížková, D., Gauthier-Clerc, M., Massez, G., … Guillemain, M. (2013). Assessing the genetic impact of massive restocking on wild mallard. *Animal Conservation*, *16*, 295–305. https://doi.org/10.1111/j.1469-1795.2012.00600.x

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157. https://doi.org/10.1111/mec.12796

DaCosta, J. M., & Sorenson, M. D. (2014). Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS ONE*, *9*, e106713. https://doi.org/10.1371/journal.pone.0106713

Davey, J. W., & Blaxter, M. L. (2010). RADSeq: Next-generation population genetics. *Briefings in Functional Genomics*, *9*, 416–423. https://doi.org/10.1093/bfgp/elq031

Dean, R., Harrison, P. W., Wright, A. E., Zimmer, F., & Mank, J. E. (2015). Positive selection underlies Faster-Z evolution of gene expression in birds. *Molecular Biology and Evolution*, *32*, 2646–2656. https://doi.org/10.1093/molbev/msv138

Delany, S., & Scott, D. (2006). *Waterbird population estimates* (4th ed.). Wetlands International Global Series, the Netherlands.

Dhami, K. K., Joseph, L., Roshier, D. A., & Peters, J. L. (2016). Recent speciation and elevated Z-chromosome differentiation between sexually monochromatic and dichromatic species of Australian teals. *Journal of Avian Biology*, *47*, 92–102. https://doi.org/10.1111/jav.00693

Ellegren, H. (2008). Sequencing goes 454 and takes large-scale genomics into the wild. *Molecular Ecology*, *17*, 1629–1631. https://doi.org/10.1111/j.1365-294X.2008.03699.x

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*, 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., … Wolf, J. B. W. (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, *491*, 756–760. https://doi.org/10.1038/nature11584

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*, 977–993. https://doi.org/10.1534/genetics.108.092221

Ford, R. J., Selman, W., & Taylor, S. S. (2017). Hybridization between Mottled Ducks (*Anas fulvigula maculosa*) and Mallards (*A. platyrhynchos*) in the western Gulf Coast region. *The Condor*, *119*, 683–696.

Francis, R. M. (2016). Pophelper: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32.

Friedman, J. H., & Rafsky, L. C. (1979). Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics*, *7*(4), 697–717. https://doi.org/10.1214/aos/1176344722

Fu, Y.-X., & Li, W.-H. (1993). Statistical tests of neutrality of mutations. *Genetics*, *133*, 693–709.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*, 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Guay, P. J., & Tracey, J. P. (2009). Feral mallards: A risk for hybridisation with wild pacific black ducks in Australia? *Victorian Naturalist*, *126*, 87–91.

Harrisson, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, *7*, 1008–1025. https://doi.org/10.1111/eva.12149

Heusmann, H. W. (1988). Influence of wintering mallards on hybridization in American black ducks. *Journal of Field Ornithology*, *59*, 258–261.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*, 379–397. https://doi.org/10.1086/688018

Huang, Y., Li, Y., Burt, D. W., Chen, H., Zhang, Y., Qian, W., … Li, N. (2013). The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nature Genetics*, *45*, 776–783. https://doi.org/10.1038/ng.2657

Hubbard, J. P. (1977). The biological and taxonomic status of the Mexican Duck (No. 16). Santa Fe, New Mexico: New Mexico Department of Game and Fish Bulletin.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*, 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K=2 conundrum. *Molecular Ecology*, *26*, 3594–3602.

Johnsgard, P. A. (1967). Sympatry changes and hybridization incidence in mallards and black ducks. *American Midland Naturalist*, *77*, 51–63. https://doi.org/10.2307/2423425

Johnson, K. P., & Sorenson, M. D. (1999). Phylogeny and biogeography of dabbling ducks (genus: *Anas*): A comparison of molecular and morphological evidence. *The Auk*, *116*, 792–805. https://doi.org/10.2307/4089339

Keinan, A., Mullikin, J. C., Patterson, N., & Reich, D. (2007). Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genetics*, *39*, 1251–1255. https://doi.org/10.1038/ng2116

Kimura, M., & Ohta, T. (1973). The age of a neutral mutant persisting in a finite population. *Genetics*, *75*, 199–212.

Kirby, R. E., Reed, A., Dupuis, P., Obrecht, H. H. III, & Quist, W. J. (2000). *Description and identification of American black duck, mallard, and hybrid wing plumage. Biological Science Report* (p. 26). Springfield, VA: U.S. Geological Survey, Biological Resources Division.

Kraus, R. H. S., Kerstens, H. H. D., Van Hooft, P., Crooijmans, R. P. M. A., Van Der Poel, J. J., Elmberg, J., … Groenen, M. A. M. (2011). Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). *BMC Genomics*, *12*, 150. https://doi.org/10.1186/1471-2164-12-150

Lavretsky, P., Dacosta, J. M., Hernández-Baños, B. E., Engilis, A. Jr, Sorenson, M. D., & Peters, J. L. (2015). Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. *Molecular Ecology*, *24*, 5364–5378. https://doi.org/10.1111/mec.13402

Lavretsky, P., Hernández Baños, B. E., & Peters, J. L. (2014a). Rapid radiation and hybridization contribute to weak differentiation and hinder phylogenetic inferences in the new world mallard complex (*Anas* spp.). *The Auk*, *131*, 524–538.

Lavretsky, P., Janzen, T., & McCracken, K. G. (2019). Identifying hybrids & the genomics of hybridization: Mallards & American black ducks of eastern North America. *Ecology & Evolution*, *9*, 3470–3490.

Lavretsky, P., McCracken, K. G., & Peters, J. L. (2014b). Phylogenetics of a recent radiation in the mallards and allies (Aves: *Anas*): Inferences from a genomic transect and the multispecies coalescent. *Molecular Phylogenetics and Evolution*, *70*, 402–411.

Lavretsky, P., Peters, J. L., Winker, K., Bahn, V., Kulikova, I., Zhuravlev, Y. N., … McCracken, K. G. (2016). Becoming pure: Identifying generational classes of admixed individuals within lesser and greater scaup populations. *Molecular Ecology*, *25*, 661–674. https://doi.org/10.1111/mec.13487

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, *8*, e1002453. https://doi.org/10.1371/journal.pgen.1002453

Lawson, D. J., Van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*, 3258. https://doi.org/10.1038/s41467-018-05257-7

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2016). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152. https://doi.org/10.1111/1755-0998.12635

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018a). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, *2*, 1940.

Malinsky, M., Trucchi, E., Lawson, D., & Falush, D. (2018b). RADpainter and fineRADstructure: population inference from RADseq data. *bioRxiv*, 057711.

Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays*, *38*, 140–149.

Mank, J. E., Carlson, J. E., & Brittingham, M. C. (2004). A century of hybridization: Decreasing genetic distance between American black ducks and mallards. *Conservation Genetics*, *5*, 395–403. https://doi.org/10.1023/B:COGE.0000031139.55389.b1

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., … Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research*, *23*, 1817–1828.

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, *32*, 244–257.

McCracken, K. G., Johnson, W. P., & Sheldon, F. H. (2001). Molecular population genetics, phylogeography, and conservation biology of the mottled duck (*Anas fulvigula*). *Conservation Genetics*, *2*, 87–102.

McVean, G., & Auton, A. (2007). *LDhat 2.1: A package for the population genetic analysis of recombination*. Oxford, UK: Department of Statistics.

Milanesi, M., Capomaccio, S., Vajana, E., Bomba, L., Fernando Garcia, J., & Colli, L. (2017). BITE: An R package for biodiversity analyses. *bioRxiv*, 181610.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, *17*, 240–248. https://doi.org/10.1101/gr.5681207

Minvielle, F., Ito, S., Inoue-Murayama, M., Mizutani, M., & Wakasugi, N. (2000). Brief communication. Genetic analyses of plumage color mutations on the Z chromosome of Japanese quail. *Journal of Heredity*, *91*, 499–501. https://doi.org/10.1093/jhered/91.6.499

Nei, M., & Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*, 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, *26*, 160–167. https://doi.org/10.1016/j.tree.2011.01.001

Novembre, J., & Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, *40*, 646–649. https://doi.org/10.1038/ng.139

Ottenburghs, J., Ydenberg, R. C., Van Hooft, P., Van Wieren, S. E., & Prins, H. H. (2015). The Avian Hybrids Project: Gathering the scientific literature on avian hybridization. *Ibis*, *157*, 892–894. https://doi.org/10.1111/ibi.12285

Oyler-McCance, S. J., Oh, K. P., Langin, K. M., & Aldridge, C. L. (2016). A field ornithologist's guide to genomics: Practical considerations for ecology and conservation. *The Auk*, *133*, 626–648. https://doi.org/10.1642/AUK-16-49.1

Perez-Arteaga, A., Gaston, K. J., & Kershaw, M. (2002). Population trends and priority conservation sites for Mexican Duck *Anas diazi*. *Bird Conservation International*, *12*, 35–52. https://doi.org/10.1017/S0959270902002034

Pérez-Figueroa, A., García-Pereira, M., Saura, M., Rolán-Alvarez, E., & Caballero, A. (2010). Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, *23*, 2267–2276. https://doi.org/10.1111/j.1420-9101.2010.02093.x

Peters, J. L., Lavretsky, P., DaCosta, J. M., Bielefeld, R. R., Feddersen, J. C., & Sorenson, M. D. (2016). Population genomic data delineate conservation units in mottled ducks (*Anas fulvigula*). *Biological Conservation*, *203*, 272–281. https://doi.org/10.1016/j.biocon.2016.10.003

Peters, J. L., Sonsthagen, S. A., Lavretsky, P., Rezutek, M., Johnson, W. P., & McCracken, K. G. (2014). Interspecific hybridization contributes to high genetic diversity and apparent effective population size in an endemic population of mottled ducks (*Anas fulvigula maculosa*). *Conservation Genetics*, *15*, 509–520. https://doi.org/10.1007/s10592-013-0557-9

Peters, J. L., Zhuravlev, Y., Fefelov, I., Humphries, E. M., & Omland, K. E. (2008). Multilocus phylogeography of a Holarctic duck: Colonization of North America from Eurasia by gadwalls (*Anas strepera*). *Evolution*, *62*, 1469–1483.

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*, 1929–1936. https://doi.org/10.1093/molbev/msu136

Phadnis, N., & Orr, H. A. (2009). A single gene causes both male sterility and segregation distortion in Drosophila hybrids. *Science*, *323*, 376–379. https://doi.org/10.1126/science.1163934

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*, e1002967. https://doi.org/10.1371/journal.pgen.1002967

Pryke, S. R. (2010). Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. *Evolution*, *64*, 1301–1310.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

Rahmatallah, Y., Zybailov, B., Emmert-Streib, F., Glazko, G. (2017). GSAR: Bioconductor package for Gene Set analysis in R. *BMC Bioinformatics*, *18*, 61. https://doi.org/10.1186/s12859-017-1482-6

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*, 489–494. https://doi.org/10.1038/nature08365

Rheindt, F. E., & Edwards, S. V. (2011). Genetic introgression: An integral but neglected component of speciation in birds. *The Auk*, *128*, 620–632. https://doi.org/10.1525/auk.2011.128.4.620

Rhymer, J. M. (2006). Extinction by hybridization and introgression in anatine ducks. *Acta Zoologica Sinica*, *52*(Supplement), 583–586.

Rhymer, J. M., & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, *27*, 83–109.

Rice, A. M., Rudh, A., Ellegren, H., & Qvarnström, A. (2011). A guide to the genomics of ecological speciation in natural animal populations. *Ecology Letters*, *14*, 9–18. https://doi.org/10.1111/j.1461-0248.2010.01546.x

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution*, *66*, 1–17. https://doi.org/10.1111/j.1558-5646.2011.01486.x

Ruegg, K., Anderson, E. C., Boone, J., Pouls, J., & Smith, T. B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, *23*, 4757–4769. https://doi.org/10.1111/mec.12842

Saether, S. A., Saetre, G.-P., Borge, T., Wiley, C., Svedin, N., Andersson, G., … Qvarnstrom, A. (2007). Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science*, *318*, 95–97. https://doi.org/10.1126/science.1141506

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology*, *26*(17), 4378–4390. https://doi.org/10.1111/mec.14226

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology and Evolution*, *19*, 198–207. https://doi.org/10.1016/j.tree.2004.01.003

Shringarpure, S. S., Bustamante, C. D., Lange, K., & Alexander, D. H. (2016). Efficient analysis of large datasets and sex bias with ADMIXTURE. *BMC Bioinformatics*, *17*, 218. https://doi.org/10.1186/s12859-016-1082-x

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., … Slate, J. (2010). Adaptation genomics: The next generation. *Trends in Ecology & Evolution*, *25*, 705–712. https://doi.org/10.1016/j.tree.2010.09.002

Stölting, K. N., Nipper, R., Lindtke, D., Caseys, C., Waeber, S., Castiglione, S., & Lexer, C. (2013). Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology*, *22*, 842–855. https://doi.org/10.1111/mec.12011

Stryjewski, K. F., & Sorenson, M. D. (2017). Mosaic genome evolution in a recent and rapid avian radiation. *Nature Ecology & Evolution*, *1*, 1912. https://doi.org/10.1038/s41559-017-0364-7

Sutter, A., Beysard, M., & Heckel, G. (2013). Sex-specific clines support incipient speciation in a common European mammal. *Heredity*, *110*, 398–404. https://doi.org/10.1038/hdy.2012.124

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595.

Toews, D. P. L., Campagna, L., Taylor, S. A., Balakrishnan, C. N., Baldassarre, D. T., Deane-Coe, P. E., … Winger, B. M. (2015). Genomic approaches to understanding population divergence and speciation in birds. *The Auk*, *133*, 13–30. https://doi.org/10.1642/AUK-15-51.1

U.S. Fish and Wildlife Service (2015). *Waterfowl population status*, *2015*. Washington, DC: U.S. Department of the Interior.

US Fish and Wildlife Service (2013). Review of captive-reared mallard regulations on shooting preserves—Final report. US Fish and Wildlife Service, Washington: Division of Migratory Bird Management.

Van Belleghem, S. M., Baquero, M., Papa, R., Salazar, C., McMillan, W. O., Counterman, B. A., … Martin, S. H. (2018). Patterns of Z chromosome divergence among Heliconius species highlight the importance of historical demography. *Molecular Ecology*, *27*(19), 3852–3872.

Whitlock, M. C., & McCauley, D. E. (1999). Indirect measures of gene flow and migration: FST[ne]1/(4Nm+1). *Heredity*, *82*, 117–125.

Williams, S. O. I. (1980). *The Mexican Duck in Mexico: Natural history, distribution, and population status*. Colorado State University.

Wolf, J. B. W., Lindell, J., & Backström, N. (2010). Speciation genetics: Current status and evolving approaches. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 1717–1733.

Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. *Nature Reviews Genetics*, *5*, 114–122. https://doi.org/10.1038/nrg1269

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration–selection balance. *Evolution*, *65*, 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

Zhou, H., Alexander, D., & Lange, K. (2011). A quasi-Newton acceleration for high-dimensional optimization algorithms. *Statistics and Computing*, *21*, 261–273. https://doi.org/10.1007/s11222-009-9166-3

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Lavretsky P, DaCosta JM, Sorenson MD, McCracken KG, Peters JL. ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. *Mol Ecol*. 2019;00:1–16. https://doi.org/10.1111/mec.15091