Received: 13 February 2017 | Revised: 15 August 2017 | Accepted: 27 September 2017

DOI: 10.1111/mec.14376

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Effects of assortative mate choice on the genomic and morphological structure of a hybrid zone between two bird subspecies

Georgy A. Semenov[1,2] | Elizabeth S. C. Scordato[3] | David R. Khaydarov[2] | Chris C. R. Smith[3] | Nolan C. Kane[3] | Rebecca J. Safran[3]

[1]Department of Ecology & Evolutionary Biology, The University of Arizona, Tucson, AZ, USA

[2]Institute of Systematics and Ecology of Animals, Novosibirsk, Russia

[3]Department of Ecology and Evolutionary Biology, The University of Colorado, Boulder, CO, USA

Correspondence
Georgy A. Semenov, Department of Ecology & Evolutionary Biology, The University of Arizona, Tucson, AZ, USA.
Email: georgysemenov@email.arizona.edu

Funding information
Russian Foundations for Basic Research, Grant/Award Number: 14-04-01492, 12-04-31042 to GAS, 13-04-10063; the National Science Foundation, Grant/Award Number: DEB-CAREER 1149942; Simpson Postdoctoral Fellowship from the University of Arizona

Abstract

Phenotypic differentiation plays an important role in the formation and maintenance of reproductive barriers. In some cases, variation in a few key aspects of phenotype can promote and maintain divergence; hence, the identification of these traits and their associations with patterns of genomic divergence is crucial for understanding the patterns and processes of population differentiation. We studied hybridization between the *alba* and *personata* subspecies of the white wagtail (*Motacilla alba*), and quantified divergence and introgression of multiple morphological traits and 19,437 SNP loci on a 3,000 km transect. Our goal was to identify traits that may contribute to reproductive barriers and to assess how variation in these traits corresponds to patterns of genome-wide divergence. Variation in only one trait—head plumage patterning—was consistent with reproductive isolation. Transitions in head plumage were steep and occurred over otherwise morphologically and genetically homogeneous populations, whereas cline centres for other traits and genomic ancestry were displaced over 100 km from the head cline. Field observational data show that social pairs mated assortatively by head plumage, suggesting that these phenotypes are maintained by divergent mating preferences. In contrast, variation in all other traits and genetic markers could be explained by neutral diffusion, although weak ecological selection cannot be ruled out. Our results emphasize that assortative mating may maintain phenotypic differences independent of other processes shaping genome-wide variation, consistent with other recent findings that raise questions about the relative importance of mate choice, ecological selection and selectively neutral processes for divergent evolution.

KEYWORDS
assortative mating, hybridization, introgression, reproductive isolation, wagtail

1 | INTRODUCTION

Phenotypes mediate social and sexual interactions and are an important component in the evolution and maintenance of reproductive barriers (Coyne & Orr, 2004; Grant & Grant, 1997; Mayr, 1963; Price, 2008). However, not all aspects of phenotypic differentiation between populations will contribute to reductions in gene flow during speciation (Shaw & Mullen, 2011). When previously isolated populations become reconnected by hybridization and introgression, differences in some aspects of phenotype can persist over time while other traits become homogenized (Key, 1968). The maintenance of trait differences in secondary contact may be important in

preventing the complete merging of previously isolated, yet closely related lineages (Harrison & Larson, 2014; Wu, 2001). Mechanisms that maintain phenotype differences in the face of potential gene flow—including the genetic architecture underlying phenotype variation—are important for understanding the speciation process (Butlin et al., 2012; Seehausen et al., 2014).

Hybrid zones provide unique opportunities to study the role of phenotypic variation when populations are in secondary contact (Barton & Hewitt, 1989; Harrison, 1990). Hybrid zones can act as selective filters that "catch" phenotypic traits and corresponding genomic regions involved in reproductive barriers while allowing selectively neutral traits and associated loci to introgress freely unless they are physically linked to a locus under selection (Barton & Bengtsson, 1986; Barton & Gale, 1993; Bazykin, 1969; Key, 1968; Payseur, 2010). This generally results in a steeper spatial transition (i.e., narrower geographic clines) for loci and phenotypes involved in reproductive isolation. However, variation in cline width can be generated by different processes; for example, high dispersal and strong selection or low dispersal and weak selection can both generate narrow clines, although these scenarios will leave distinct genetic footprints (Barton & Hewitt, 1985; Brelsford & Irwin, 2009). In the former case, an excess of homozygous genotypes and strong associations between parental allele combinations (linkage disequilibrium) towards the hybrid zone centre will lead to a bimodal distribution (i.e., predominance of parental genotypes and few intermediates) and increased variance in corresponding quantitative traits. In the case of low dispersal and weak selection, many individuals will show evidence of hybridization and introgression, resulting in reduced linkage disequilibrium, low phenotypic variance, and a comparatively unimodal phenotypic and genotypic distribution due to the high frequency of intermediates (Barton & Gale, 1993; Gay, Crochet, Bell, & Lenormand, 2008; Jiggins & Mallet, 2000; Nurnberger, Barton, MacCallum, Gilchrist, & Appleby, 1995). Importantly, recent secondary contact alone may result in a narrow cline (Gompert & Buerkle, 2016) and nonunimodal distribution of phenotypes, increased trait variance and linkage disequilibrium in the absence of selection (Gay et al., 2007). In addition, high dispersal propensities of parental forms can potentially lead to a nonunimodal distribution of phenotypes and genotypes despite unrestricted hybridization (Nagylaki, 1976). However, if a cline is maintained by a balance between selection and dispersal, its width will be narrower than the expectation under a neutral diffusion model that accounts for the number of generations since secondary contact and dispersal capabilities of an organism (Barton & Gale, 1993; Endler, 1977). Especially in a case of traits maintained by weak selection, dominance relationships among alleles, epistasis, effects of genetic architecture and phenotypic plasticity might complicate the interpretation of phenotypic patterns. Despite these complexities, simultaneous analysis of geographic clines and patterns of phenotypic distributions and variances across the hybrid zone can help to understand the factors affecting zone structure and identify traits whose variation is unlikely to be explained by selectively neutral processes alone.

A comparison of introgression in phenotypic traits with variation in genomic markers can be further useful for inferring the genetic architecture of reproductive barriers (Payseur & Rieseberg, 2016). When multiple regions distributed across the genome are associated with reproductive isolation (e.g., multiple traits with polygenic control), introgression will be coincident between phenotypic traits and many unlinked genetic markers, leading to concordant geographic clines (e.g., (Mettler & Spellman, 2009; Ruegg, 2008; Scordato et al., 2017) and a genome-wide increase in linkage disequilibrium towards the hybrid zone centre (Barton & Gale, 1993; Brelsford & Irwin, 2009; Gay et al., 2008). By contrast, if few genes (e.g., few traits with relatively simple genetic control) underlie reproductive barriers, introgression will be limited to a small subset of loci, and appear otherwise unrestricted throughout the genome (Poelstra et al., 2014; Toews et al., 2016).

Despite numerous studies examining patterns of genomic and phenotypic variation across hybrid zones, it is often unclear what mechanisms underlie reproductive isolation. An advantage of empirical studies of hybrid zones is the opportunity to use field-based observations to assess these potential mechanisms (Harrison, 1986). For example, assortative mating can maintain divergence in traits associated with social or sexual signalling, and patterns of pair composition can be used to evaluate the likelihood of assortative mating. Thus, combining patterns of phenotypic and genomic variation in hybrid zones with observations of mating patterns can collectively provide insights about the processes underlying population divergence.

The white wagtail (*Motacilla alba*) is a small passerine bird comprised of nine named subspecies that exhibit remarkable variation in plumage and body size (Alström & Mild, 2003; Badyaev, Gibson, & Kessel, 1996). Subspecies *alba* and *personata* (Semenov & Yurlov, 2010; Semenov, Yurlov, & Khaydarov, 2010; Sushkin, 1938), *personata* and *baicalensis* (Red'kin, 2003), and *lugens* and *leucopsis* (Nazarenko, 1968; Stepanyan, 1983, 2003) form narrow zones of hybridization which are presumably maintained by selection (Stepanyan, 1983), although the possibility of recent secondary contact has not been evaluated. Plumage differences in these hybrid zones persist despite very little genetic differentiation (Alström & Ödeen, 2002; Li et al., 2016; Pavlova et al., 2005). The white wagtail complex is therefore a promising system for identifying the phenotypic and genetic basis of reproductive isolation, as high genetic homogeneity makes it more likely that characters showing restricted introgression are actually those related to reproductive isolation, rather then selectively neutral loci affected by linked selection. Assortative mating by phenotype has been observed between several subspecies pairs (Lobkov, 2011; Nazarenko, 1968), suggesting that phenotypic differences might be maintained by divergent mating preferences, although this assumption has not been examined. Here, we analysed patterns of hybridization, introgression and assortative mating between two subspecies, *personata* and *alba*. The subspecies differ in several morphological traits, including back colour, head and neck patterning, and wing coverts—traits that often function as inter- and intrasexual signals in birds (Seddon et al., 2013). Multiple aspects of

morphological differentiation may therefore contribute to reproductive isolation in wagtails. We used field-based data on breeding pair composition, a comprehensive set of morphological traits that distinguish the two subspecies, and 19,437 SNP loci sampled across 3,000 km transect to test what and whether traits, loci and patterns of assortative mating contribute to reproductive barriers.

2 | METHODS

2.1 | Field sampling

We collected 318 breeding adults (Table S1) along a transect spanning the hybrid zone in the western foothills of the Altai mountains as well as from remote allopatric populations of *alba* (West Siberia, samples A and B, Figure 1) and *personata* (Kazakhstan and Uzbekistan, O and P, Figure 1) during May–June of 2012 and 2014. We determined the appropriate sampling transect location and direction (Figure 1) using information from the literature (Berezovikov & Reznichenko, 2014; Berezovikov, Samusev, Khrokov, & Egorov, 2007; Cramp, 1988; Glutz Von Blotzheim & Bauer, 1985; Red'kin, 2003; Semenov & Yurlov, 2010; Semenov et al., 2010; Stepanyan, 2003; Sushkin, 1938; Zalesskii, 1927) and from data associated with 613 museum study skins from the Zoological Museum of Moscow State University (Moscow, Russia) and the Zoological Institute of the Russian Academy of Sciences (Saint Petersburg, Russia). In the field, the location of the hybrid zone was inferred from the coloration of the head plumage, the most noticeable diagnostic trait between subspecies. All field-collected specimens were prepared as study skins with one spread wing and deposited in the collection of the Institute of Systematics and Ecology of Animals (Novosibirsk, Russia; for ISEA specimen numbers and sampling details see Table S1). Sexing was performed by gonad examination. To ensure sampling of locally breeding individuals, we only collected birds exhibiting territorial behaviour and checked for presence of seminal glomus in males and enlarged ovaries in females as indicators of breeding status.

2.2 | Phenotypic measurements

The *alba* and *personata* subspecies differ in the proportion of black vs. white plumage colour on the head and sides of the neck (hereafter *head*), intensity of black vs. grey colour on the back (hereafter *back*), and proportion of white colour on the wing (hereafter *wing*, Figure 2). To quantify plumage variation among individuals, we took standardized digital photographs of all study skins. We measured the absolute number of white pixels on the *head*, the mean colour of the *back* and the ratio of white to black pixels on the *wing* using IMAGEJ (Abramoff, Magalhaes, & Ram, 2004). We measured each trait three times and averaged across replicates. Measurements were taken for eight morphological traits describing body size: weight, body length, wing length, tail length, tarsus length, bill length, bill depth and bill width. Variation in morphology was summarized using principal component analysis (PCA). We used PC1 scores as a proxy of body size variation (hereafter *body size*) in the geographic cline, variance and individual trait distribution analyses. Only males were used in analyses of morphological traits due to moderate but significant sexual dimorphism in plumage and body size (G. Semenov, unpublished data).

2.3 | Genotyping by sequencing

Samples of breast muscle were stored in 96% ethanol prior to laboratory analyses. Genomic DNA was extracted either using the DNeasy Blood and Tissue Kit (Qiagen Inc.) following manufacturers' recommendations, or a standard phenol–chloroform extraction. We used a restriction-fragment-based protocol to construct reduced genomic complexity libraries for each individual (Gompert et al., 2012; Parchman et al., 2012) following Safran et al. (2016). We digested genomic DNA with the restriction endonucleases EcoRI and MseI and ligated unique barcodes to each digested fragment. We then pooled samples and used standard Illumina PCR primers to amplify barcoded fragments. We separated the PCR product on a 2% agarose gel and excised the gel band corresponding to approximately 350–450 bp in length to ensure we sequenced standard-sized fragments. We purified gel fragments using the QIAquick Gel Extraction Kit (Qiagen Inc). Concentration and quality of the pooled library were evaluated on an Agilent BioAnalyzer. Genomic samples from locations A and O were processed in a separate Illumina lane and were later excluded from analyses due to strong lane effect on SNP calling.

We sequenced 294 individuals in multiplex on a single lane of Illumina Hiseq 2500 at the University of Texas Genomic Sequencing and Analysis Facility in Austin. We discarded reads that aligned to the Illumina oligo database, the *phiX* genome or the *E. coli* genome. Incorrectly sequenced barcodes with a single mismatch were identified and corrected, and barcode and adapter sequences were removed from reads. This yielded 45.7 million 84- to 86- base pair single-end reads. No reference genome exists for the white wagtail, so a subset of 25 million reads sampled randomly from across all individuals was used to construct a de novo pseudoreference assembly using SEQMAN NGEN 11.1.0 (DNASTAR). We required a 94% minimum match percentage, yielding 375,000 short contigs and N50 of 87 bp. One individual had fewer than 1,000 reads and was excluded from further analysis.

Reads from all individuals were then aligned to the de novo reference using BWA 0.7.12 (Li & Durbin, 2009) with the following settings: trim ends of reads with quality below 10, exclude trimmed reads smaller than 35 bp, maximum number of differences of 4, seed length of 20, maximum number of differences in the seed of 2, stop searching after 20 equally best hits. Variants were called using SAMTOOLS v. 1.2 and BCFTOOLS v. 1.2 (Li et al., 2009) with the following settings: 0.6 minimum proportion of samples covered, variant if $p < .01$, full prior and scaled substitution mutation rate of 0.001. We retained only common variants with a minor allele frequency of 5% or greater as these are most relevant for analyses of population structure (Gompert et al., 2012), and included only one randomly selected SNP per 100-bp locus to avoid tightly linked loci.

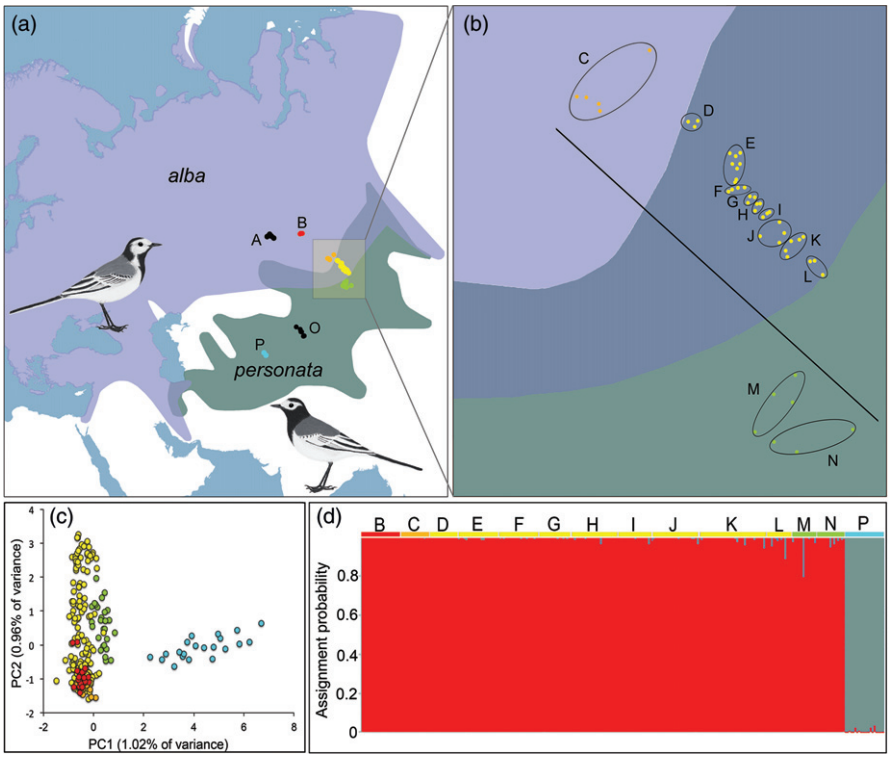

FIGURE 1 Breeding distribution of white wagtail taxa (1), hybrid zone and sampling points in central Siberia (2), principal component analysis of genome-wide covariance matrix with 19,437 loci (3) and results of Bayesian clustering analysis of STRUCTURE with 14 semidiagnostic loci (4). Violet indicates distribution of *alba*, dark green is *personata*, colour overlap corresponds to the hybrid zone. Coloured dots are sampling localities, colour scheme of PCA and top panels in STRUCTURE correspond to sampling localities. Ovals in (2) encircle sample groups used in analyses, capital letters are sample IDs. The black line on (2) shows the direction of the transect. Genomic samples A and O were processed in a separate Illumina lane and were later excluded from analyses due to strong lane effect on SNP calling

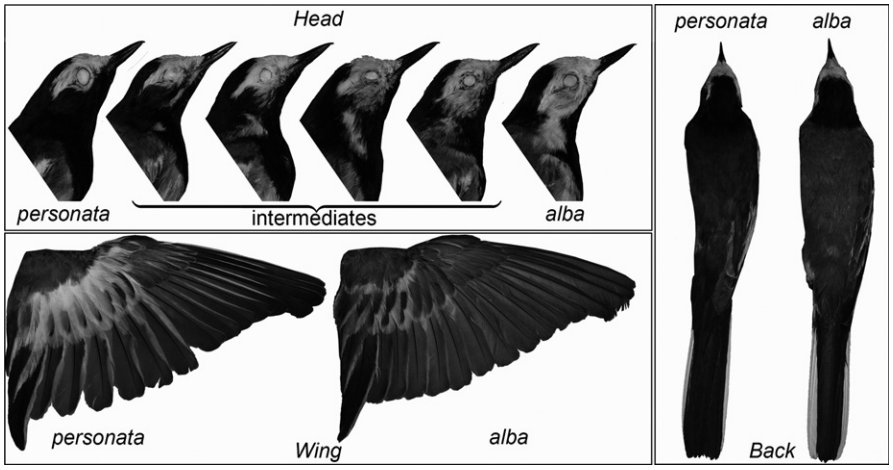

FIGURE 2 Variation in head and neck sides, wing coverts and back colour in *personata* and *alba* subspecies of the white wagtail. Intermediate head plumage patterning found in the hybrid zone (samples D–L) was presented by only few distinct categories. Wing and back plumage have continuous variation between parental extremes

To incorporate uncertainty due to variation in genotyping coverage (Gompert et al., 2012, 2014), we used a Bayesian model to estimate population allele frequencies for each SNP using the point estimates of genotypes from *bcftools* (described in (Gompert et al., 2012)). This model incorporates uncertainty due to variation in coverage across individuals and loci. Populations were defined a priori, genotypes and population allele frequencies were treated as unknown model parameters, and genotype probabilities and allele frequencies were simultaneously estimated for each population. Observed population allele frequencies were used as priors. We obtained posterior probabilities for parameters using Markov chain Monte Carlo (MCMC). Each analysis consisted of 1,000 steps of burn-in followed by a single MCMC chain iterated for 2,000 steps, with samples retained from every other step, resulting in 1,000 samples from the posterior distribution.

We obtained the mean genotype (scale of 0–2, with 0 and 2 as alternative homozygotes and 1 as the heterozygote) from the

posterior distribution for each SNP in each individual. In analyses that required SNP calls (rather than continuous probabilities), we considered genotype probabilities ranging from 0–0.1 to 1.9–2 as the two alternative homozygous forms, and 0.95–1.05 as heterozygous; all other values were treated as missing data. Individuals with >40% and loci with >30% missing data were excluded, resulting in a data set of 276 individuals and 19,437 SNP loci.

2.4 | Population structure

To assess evidence for population structure across the transect, we used the Bayesian assignment model in the program STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000). We ran STRUCTURE for *K* ranging from 1 to 4, with three replicates per *K*, using the admixture model with correlated allele frequencies and no prior sampling information. We ran STRUCTURE on the complete set of 19,437 SNPs and on a subset of loci with an allele frequency difference >0.4 between samples B and P. The runs were set for 500,000 generations of burn-in followed by 1,000,000 MCMC chains. The resulting Q-scores were summarized between runs using CLUMPP (Jakobsson & Rosenberg, 2007). As an alternative method to assess population structure without a predefined number of clusters, we conducted PCA based on eigenvalue decomposition of the covariance matrix for 19,437 SNPs using TASSEL 5.0 (Bradbury et al., 2007). The PC1 scores were used as a proxy of genome-wide variation in subsequent analyses. We hereafter refer to PC1 scores as "genomic ancestry" for convenience; this variable is representative of genome-wide covariance. To ensure that the geographic cline for genomic ancestry had decreasing values towards the end of the transect (a requirement for CFIT7 analyses, see below), we transformed the genomic ancestry by multiplying it by negative one.

2.5 | Variance analysis

To describe introgression in quantitative traits and molecular markers, we used an approach analogous to an analysis of linkage disequilibrium, that is based on a variance distribution (Gay et al., 2007). We estimated the parameter, V_{mix}, which is the expected variance under a model of complete parental admixture due to introgression (equation 6 in Gay et al. (2007)) for the four quantitative traits (*head*, *wing*, *back* and *body size*) and genomic ancestry. V_{mix} is defined as:

$$V_{mix} \approx pV_1 + (1-p)V_2,$$

where p is the admixture coefficient representing the proportion of individuals originating from population 1, and V_1 and V_2 are phenotypic variances in parental populations 1 and 2, respectively. We determined the admixture coefficient of each sample from mean trait values of parental samples:

$$p_{sample} = (\mu_{sample} - \mu_2)/(\mu_1 - \mu_2),$$

where μ_1 and μ_2 are means of parental samples 1 and 2.

We combined samples for parental *alba* from West Siberia (A-B) and parental *personata* from Kazakhstan and Uzbekistan (O-P) to estimate parental means and variances for morphological traits. We then used linear regression to calculate the 95% confidence intervals for V_{mix} along the hybrid zone and adjacent parental samples (samples C-N, Figure 1) and compared the observed variance and its 95% confidence interval to the expected range of variances as expressed by V_{mix}. A significant increase in observed variance compared to V_{mix} would suggest that both parental values are present in a population at higher than expected proportions, which may indicate selection against intermediates or assortative mate choice. Similar inferences can be made by analysing variance in genomic ancestry as a proxy of genome-wide linkage disequilibrium. Increased variance in genomic ancestry near the hybrid zone centre implies coexistence between divergent genotypes and suggests limited interbreeding, whereas low variance suggests high rates of genomic admixture or lack of genomic differentiation (Barton & Gale, 1993).

2.6 | Geographic cline analysis and individual value distribution modelling

The distance along the transect was set to zero at sample A and then estimated as the minimum distance from each sample to the lines spanning A-B-C and N-O-P (Figure 1, Table S1). For samples C-M, we estimated the shortest distance from the approximate centre of the ellipse encircling each sample to the transect line (Figure 1). To estimate the width and centre of geographic clines, we used the monotone cline function (Szymura & Barton, 1986) implemented in the program CFIT7 (Gay et al., 2008) to fit sigmoid geographic clines for the four morphological traits (*head*, *back*, *wing* and *body size*), genomic ancestry and individual SNPs with allele frequency differences >0.4 between samples B and P. The cline width was estimated as 4/maximum slope (Barton & Gale, 1993).

To test whether clines share similar positions in space (i.e., similar centres) and similar rates of introgression (i.e., similar slopes), we used the likelihood search method implemented in Cfit7. First, we estimated likelihoods for individual *head*, *wing*, *back*, *body size* and genomic ancestry clines and the maximum likelihood of the full model where all five clines are combined and have independent parameters for slope and centre. We then tested a set of models where cline parameters were constrained for either slope, centre, or both slope and centre. The CFIT7 input file enables the user to select parameters to be shared among clines, that is constrained among them. For instance, the full model for two clines with independent centres and slope parameters will be encoded as "cline A: centre = 1, slope = 2; cline B: centre = 3, slope = 4." For the full model with a constrained centre, the parameter set will be "cline A: centre = 1, slope = 2; cline B: centre = 1, slope = 3," etc. We compared likelihoods for individual clines using likelihood ratio tests (with significance threshold at $p = .05$). If the likelihood of an individual cline with constrained parameters was significantly lower than when unconstrained, we assigned this cline independent parameters and reran the full model. Every fit was checked for convergence by repeating the fit using different random seeds. The goal was to find

an optimal model where as many parameters as possible for individual clines can be constrained, without reducing the likelihoods for individual clines. We then compared maximum likelihoods of the unconstrained full model with models where (i) all centres were constrained, (ii) all slopes were constrained, (iii) both centres and slopes were constrained and (iv) the optimal full model with maximal possible number of constrained parameters using AIC (Akaike's information criterion) (Akaike, 1973) and AICc corrected for small sample sizes (Burnham & Anderson, 2002). We considered differences between full models to be significant if AIC and AICc were larger than 2 points (Raufaste et al., 2005).

We used CFIT7 (Gay et al., 2008) to test the correspondence of individual trait value distributions to five models of hybridization and introgression: (i) *bimodal-no-introgression*, (ii) *bimodal-with-introgression*, (iii) *trimodal-no-introgression*, (iv) *trimodal-with-introgression* and (v) *unimodal* (Figure 3, upper panel). These models represent evolutionary scenarios with decreasing values of dispersal and selection (Gay et al., 2008). If dispersal is high and hybridization is rare or absent, contact between two lineages results in an excess of homozygous genotypes and strong linkage disequilibrium between parental alleles, corresponding to the i) *bimodal-no-introgression* model. If hybridization happens more often there may be limited introgression due to episodes of hybridization and backcrossing, although recent generation hybrids are absent, which fits the (ii) *bimodal-with-introgression* model. With frequent hybridization, some individuals may form a well-defined group of F1 hybrids, which do not backcross due to strong selection, leading to the (iii) *trimodal-no-introgression* model, or hybrids may occasionally interbreed with parentals resulting in the (iv) *trimodal-with-introgression* model. Finally, in the case of low dispersal and weak or absent selection, the effects of linkage disequilibrium and homozygous excess will fade away over time due to recombination and introgression, resulting in a hybrid swarm and (v) *unimodal* distribution (Gay et al., 2008; Jiggins & Mallet, 2000). We fit the five models to *head*, *wing*, *back*, *body size* and genomic ancestry, and selected the best model using AIC and AICc with the significance thresholds described above.

To determine the likelihood that geographic clines are maintained by a balance between selection and dispersal, we estimated the cline width (w) expected under neutral diffusion model (Barton & Gale, 1993):

$$w = 2.51\sigma\sqrt{t},$$

where σ is a standard deviation (SD) of postnatal breeding dispersal distance and t is the number of generations since secondary contact. We used breeding dispersal distances for the subspecies *Motacilla alba yarrellii* (SD = 24.5 km) as these were the only available dispersal estimates for the white wagtail to our knowledge (Dougall, 1991). Hybridization between *alba* and *personata* in the region of our study has been known for at least one century (Sushkin, 1938), thus we considered that contact happened 100 generations ago, given the generation time of 1 year typical for many small passerines (Ehrlich, Dobkin, & Wheye, 1988).

2.7 | Test for assortative mate choice

In 2013, we collected observations for 47 social pairs from six sites near samples H and I, corresponding to the centre of the *head* geographic cline. Social mates were assigned based on observations of parental behaviour (feeding nestlings in a discrete nest location). We divided phenotypes into "parental" individuals corresponding to parental *head alba* and *personata* phenotypes, and "intermediates" falling outside the parental phenotypic variation limits (Figure 2). We used head patterning for delimitation between parental phenotypes and hybrids as it is the only morphological trait that reliably separates *alba* and *personata* from a distance (see Results). We used randomization tests to determine whether the observed composition of social pairs was more phenotypically similar than expected by chance in the population. For the randomization procedure, males and females were selected from the observed set of phenotypes to create sets of 47 pairs with random pairing using 10,000-bootstrap replicates. The randomized pairs were grouped into three categories: SAME included pairs either between two individuals of the same parental phenotype or two hybrids; NEAR included one parental and one hybrid, and FAR consisted of the two different parental phenotypes. We then calculated 95% confidence intervals for the randomized number of pairs in each category and compared them with the observed values. Calculations were performed in SAS 9.2 (SAS Institute Inc.).

3 | RESULTS

3.1 | Population structure

Genome-wide divergence between *alba* and *personata* was very low, with PC1 and PC2 explaining only 1.02% and 0.96% of total genome-wide covariance for 276 individuals using 19,437 SNPs (Figure 1). Only 14 loci (0.07%) had allele frequency differences >0.4 at the ends of the transect, further indicating that allopatric *alba* and *personata* have similar allelic variation at most markers throughout their genomes. Genomic variation across the transect was predominately driven by sample group P (parental *personata* from Uzbekistan), which was clearly differentiated from all other samples along PC1 (Figure 1). Genomic variation contrasted with patterns of plumage variation. Ninety-six per cent of individuals (n = 16) in samples M and N had head plumage patterns that were characteristic of *personata*, and were therefore classified as parental *personata* in the field, but were genomically very close to *alba* (Figure 1). Furthermore, there was no detectable genome-wide divergence between allopatric parental *alba* and all individuals sampled from within the hybrid zone, despite clear morphological differences (Figure 1).

Similarly, the STRUCTURE analysis revealed a very weak signal of population structure at all values of K for the set of 19,437 SNPs (Fig. S1). Putative parental populations at the ends of transect were virtually indistinguishable in their assignment probabilities at K = 2. There was weak differentiation between *alba* (samples B-D), and *personata* from Uzbekistan (sample P) at K = 3 and K = 4, but none of the samples formed differentiated clusters (Fig. S1). When we reran

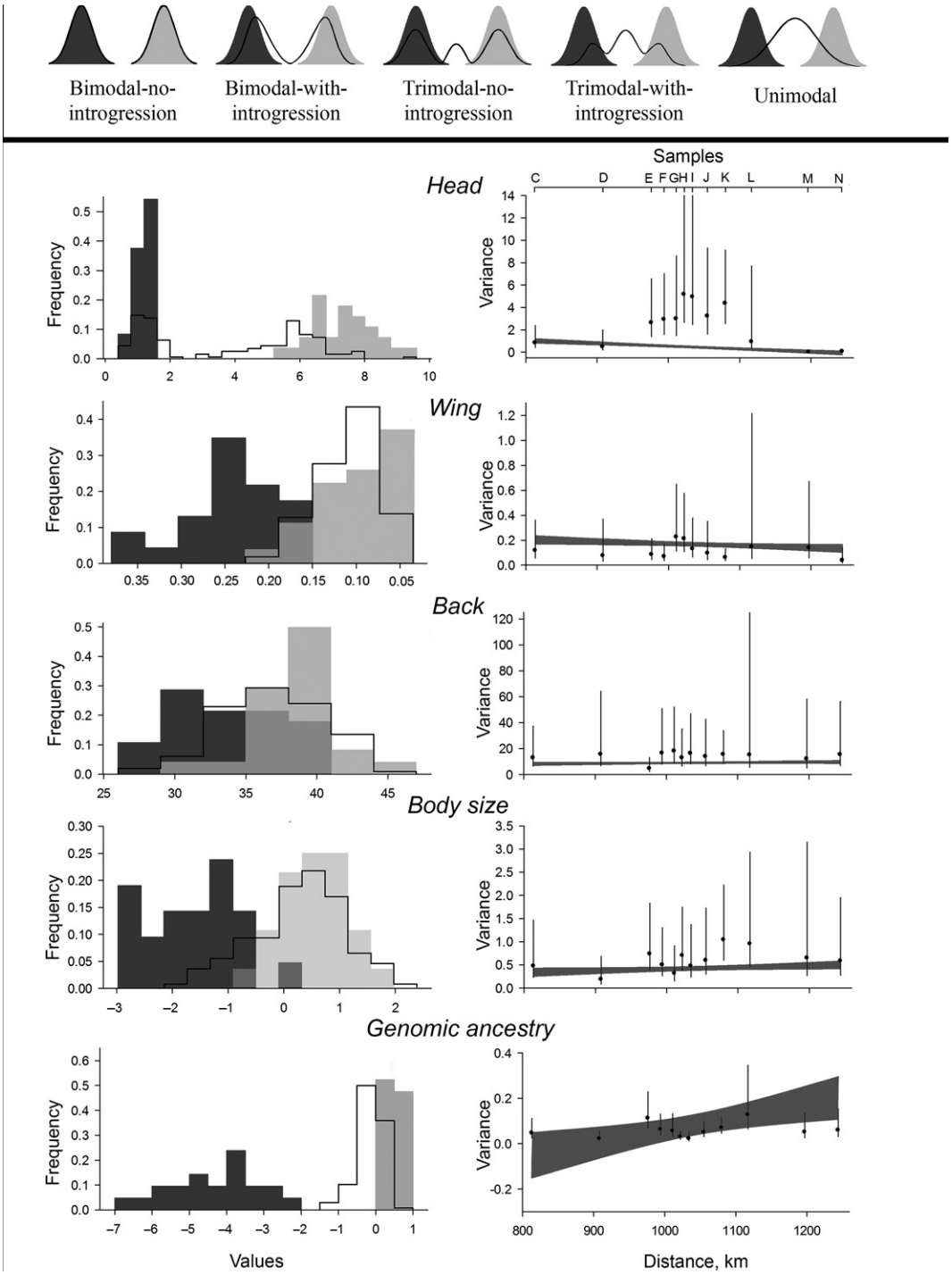

FIGURE 3 Distribution of individual values and results of variance analysis. Upper panel illustrates five scenarios of hybridization implemented in CFIT7 models. Dark grey and light grey indicate distributions in two parental populations, and black line shows distribution in the hybrid zone. Bottom left: Dark grey distribution corresponds to pure *personata*, light grey to pure *alba*, black outline to the hybrid zone (samples D–L). Bottom right: Grey area shows 95% confidence interval for variance V_{mix} expected under complete admixture of parental variances; dots and vertical lines correspond to observed variance and its 95% confidence interval. Distribution in the hybrid zone is nonunimodal for *head* but appears unimodal for other morphological traits and the genomic ancestry. There is an increase in variance for *head* in seven samples across the hybrid zone, but no noticeable increase in variance for other traits and genomic ancestry except sample K for *body size*. These results show that parental *head* phenotypes are present across the hybrid zone at proportions higher than expected under complete admixture. Note that in the hybrid zone all distributions except *back* are skewed towards *alba*, suggesting mismatch between variation patterns of *head* and other traits and genetic markers

STRUCTURE using only the subset of 14 semidiagnostic loci (those with allele frequency differences >0.4), we found clear separation between Uzbekistan *personata* (sample P) and all other samples at all values of *K* (Figure 1, Fig. S1). We thus used STRUCTURE assignment probabilities ($K = 2$ for 14 semidiagnostic loci) to classify individual genotypes as parental *alba* (assignment probabilities, AP > .995, lower limit in the sample B), parental *personata* (AP < .035, upper limit in the sample P) or intermediates (.036–.994, Figure 1, Table S2). Consistent with the results of the PCA, most individuals in samples M and N (70%, $n = 19$) had parental *alba* genotypes, with the remainder ($n = 8$) being intermediate, although very close to *alba* (AP = .794–.999), despite exhibiting plumage characteristic of *personata* (Figure 1, Table S2). Likewise, in the hybrid zone (samples D-L), 81% of individuals ($n = 155$) had ancestry corresponding to parental *alba* irrespective of their plumage phenotype, and the remaining ($n = 36$) were classified as intermediates, but were again much closer to *alba* (AP = .886–.999) (Figure 1, Table S2).

To further explore the loci differentiated across the plumage transition, we removed sample P and again searched for semidiagnostic loci. We looked for loci with an allele frequency difference >0.4 between *alba* (sample B) and *personata* from the Altai mountains adjacent to the hybrid zone (samples M and N combined). Only four loci met this search criterion, and variation in these loci had no direct correspondence with plumage patterns (Fig. S2).

3.2 | Phenotypic variation and variance patterns across the transect

Parental populations of *alba* and *personata* were significantly differentiated in all morphological and plumage traits (Figure 2), but only *head* had nonoverlapping parental distributions (Figures 3, Table S2). PC1 and PC2 explained 46.24% and 15.93%, respectively, of the variance in PCA on eight morphometric variables, with major differentiation between allopatric *alba* (samples A and B) and *personata* (samples O and P) captured by variation along PC1 (see Table S3 and Fig. S3 for details). In the phenotypic hybrid zone, 24% of individuals ($n = 114$, males) had intermediate *head* plumage categories, suggesting that these phenotypes are likely a product of hybridization and recombination (Figure 2). Variation in *head* plumage appeared discrete and was mostly confined to four intermediate categories, which may suggest a simple genetic architecture for this trait. Despite some evidence for phenotypic intermediates, the majority of individuals in the hybrid zone exhibited *head* plumage characteristic of either parental *alba* or *personata*, resulting in a nonunimodal distribution for this trait (Figure 3). Consistent with this distribution, observed variance in *head* was significantly greater than the variance V_{mix} expected under complete admixture in samples E-K (Figure 3) and variance increased towards the *head* cline centre, suggesting that ongoing admixture does not result in a hybrid swarm. In contrast, distributions of the other morphological traits and the genomic ancestry were unimodal (Figure 3). Variance V_{mix} was not greater than expected under a scenario of complete admixture in *wing* and *back*, and exhibited a slight but significant increase in the *body size* metric in sample K (Figure 3).

3.3 | Geographic clines and individual value distribution modelling

Geographic positions of cline centres were highly variable: the *head* cline centre was located at 1,032 km along the transect (between samples H and I, Table S1), whereas other cline centres were displaced 139–323 km south-eastwards, in the region populated by individuals with *personata*-type head plumage (Table 1; Figure 4). The *head* cline was also 3–6 times narrower than clines of the other quantitative traits (Table 1; Figure 4). The likelihood tests implemented in CFIT7 were used to analyse concordance between cline position and rates of introgression, and revealed that the models with constraints on either slope, centre, or both for all five clines should be rejected, suggesting differential introgression (Table 1). The model with the best fit had the slope constrained for *wing*, *back*, *body size* and genomic ancestry, but independent for *head*, and had independent centres for the *head* and genomic ancestry clines but constrained centres for *wing*, *back* and *body size*. This model was significantly better than the completely unconstrained model (Table 1). This result indicates that (i) *wing*, *back*, *body size* and genomic ancestry clines have similar widths despite differences in estimates of width for their individual clines and thus may have similar rates of introgression, and (ii) position of cline centres are similar for *wing*, *back* and *body size*, indicating that the temporal and spatial processes that affect the position of these clines may be similar. Centres of geographic clines for individual SNPs were located 602–1,923 km away from the *head* centre and corresponded to areas inhabited by populations with *personata*-type head plumage (Figure 4). Cline width of individual SNPs was highly variable (465–4,156 km) and exceeded cline width of *head* by ≈ 4–37 times (Figure 4).

Results from model-based testing of the five CFIT7 hybridization scenarios (Table 2) were concordant with the observed distributions of morphological traits and genomic ancestry (Figure 3). The *trimodal-with-introgression* model was significantly more likely for *head* (Table 2) compared to the other four models, consistent with presence of both parental types and intermediates across the hybrid zone and in the *head* cline centre. The *unimodal* model had the best fit for *wing*, *body size* and genomic ancestry. The *bimodal-no-introgression* and *unimodal* models, representing the opposite ends of the continuum of tested scenarios, were equally likely for *back*, suggesting that variation in this trait is a poor fit to the CFIT7 models.

Expected cline width under a neutral diffusion scenario due to unrestricted hybridization and introgression and assuming a conservative time since secondary contact of 100 years was 615 km. The *head* cline was over five times narrower (112 km, Table 1), suggesting that it cannot be maintained by dispersal or recent contact alone, and hence is likely maintained by selection. Cline width for *wing* did not differ from a scenario of neutral diffusion (722 km, Table 1), and the *back* and *body size* clines were only about 10%–20% narrower (559 and 492 km, Table 1). Although the genomic ancestry cline was about two times narrower (318 km, Table 1), this estimate should be considered with caution given the gap of sampling in the region of genomic transition (Figures 1 and 4), which also could have affected

TABLE 1 Likelihood search for common centre and slope in morphological and genomic ancestry clines. Width is cline width in kilometres. Centre is cline centre in kilometres from the sample A. Nb param is the number of parameters in model. The best model is boldfaced

Model	No constraint			Centre constrained		Slope constrained		Slopes and centres constrained		Four slopes and three centres constrained			
Cline	Width	Centre	Log L	Log L	*p*	Log L	*p*	Log L	*p*	Width	Centre	Log L	*p*
Head	112	1,032	−220.2	−253.4	<.0001	−229.2	.0012	−334.6	<.0001	112	1032	−220.2	1.0000
Wing	722	1,171	−438.3	−438.5	.9830	−440.1	.4449	−447.6	.0193	385	1209	−439.6	.7385
Back	559	1,237	−44.1	−48.7	.0584	−48.4	.0730	−64.0	<.0001	385	1209	−45.4	.7641
Body size	492	1,248	−192.1	−195.5	.1536	−193.3	.7284	−201.3	.0193	385	1209	−192.3	.9963
Genomic ancestry	318	1,355	−56.8	−111.0	<.0001	−58.1	.6286	−112.0	<.0001	385	1386	−57.6	.9005
Nb param			42		38		38		34				37
AIC			2,215.5		2,411.3		2,242.5		2,629.5				**2,212.2**
AICc			2,219.1		2,414.3		2,245.5		2,631.9				**2,215.0**

the width estimates for *wing*, *back* and *body size*. If we assumed a longer time since secondary contact or a higher postnatal dispersal distance, the expected cline width under a neutral diffusion model is wider. Thus, there may also be weak selection maintaining other morphological clines as well as genomic ancestry. Together, the results of geographic cline, variance and trait distribution analyses indicate that discrete differences in head plumage are likely maintained by selection in the hybrid zone. If selection drives divergence in other aspects of morphology (particularly *body size*), it appears to be comparatively much weaker.

3.4 | Test for assortative mate choice

Randomization tests revealed significant overrepresentation of pairs in the SAME category (homotypic pairs and pairs between two intermediates, N_{exp} = 9.0–21.6, $N_{ob\ s}$= 27) and underrepresentation in the FAR (heterotypic pairs N_{exp} = 8.6–21.2, N_{obs} = 6) categories compared to an expectation of random mating, indicating that the composition of social pairs was not random with respect to *head* plumage (Figure 5). These results suggest that variation in head plumage in the hybrid zone is maintained via assortative pairing between parental phenotypes, either because of strong premating isolation, reduced fitness of hybrids, or both.

4 | DISCUSSION

In this study, we examined patterns of genomic and phenotypic variation across the hybrid zone between the *alba* and *personata* white wagtail subspecies and assessed the role of assortative mating in maintaining reproductive isolation and genetic differentiation. Multiple lines of evidence suggest that one highly differentiated aspect of phenotype, head and neck plumage, is maintained by assortative mating. In contrast, several other divergent plumage and morphometric characters showed little or no evidence for being maintained by selection, suggesting that nonrandom mating is related to *head* plumage specifically, or perhaps other correlated traits that were not measured. One of the most surprising findings was the dramatic displacement of geographic clines for head plumage compared to all other traits and genomic ancestry. The transition in *head* plumage from *alba* to *personata* phenotypes occurred across populations otherwise genomically and morphologically closely corresponding to *alba*, whereas cline centres for other traits and a genomic ancestry occurred in a geographic region we had attributed to *personata* based on head plumage.

4.1 | Divergence, hybridization and introgression in wagtails

Hybrid zones are typically defined as narrow regions in which genetically distinct populations meet, mate and produce hybrids (Barton & Hewitt, 1985). In this context, interpretation of patterns of phenotypic and genomic variation across the wagtail transect is

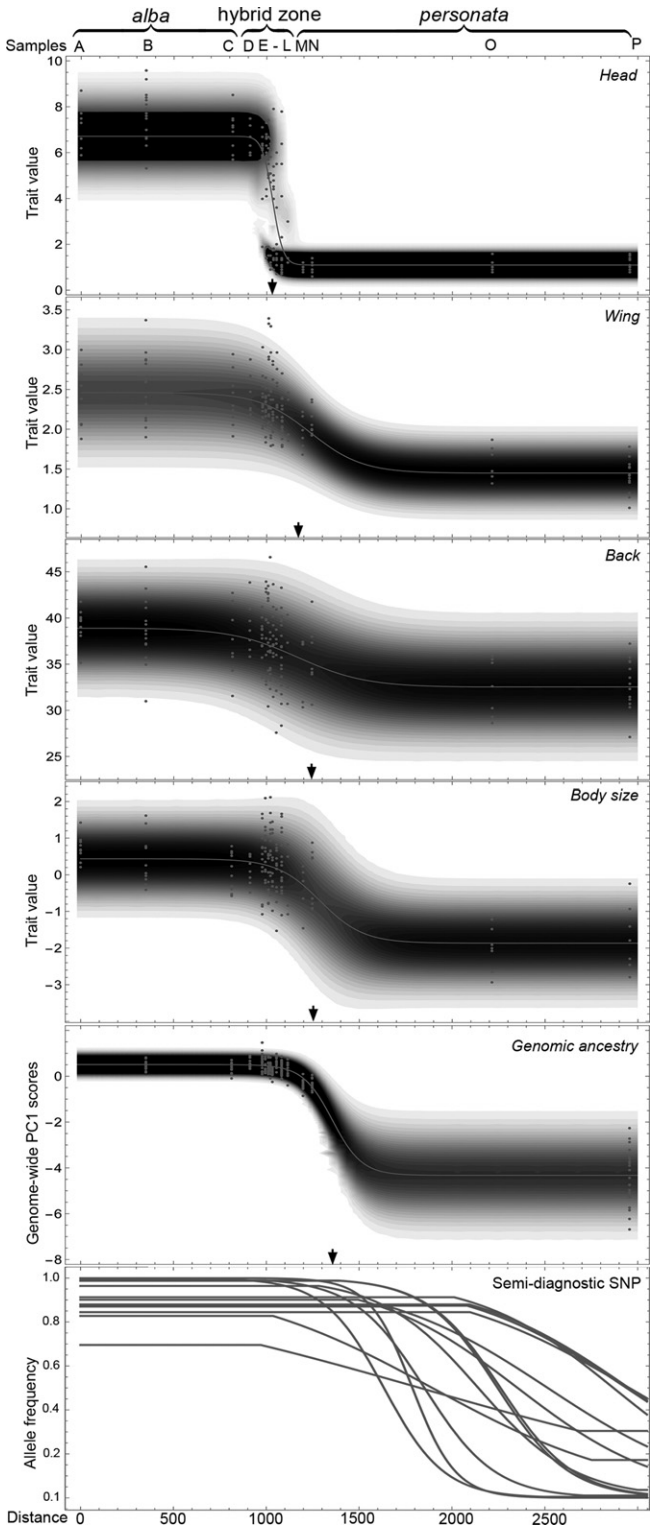

FIGURE 4 Geographic clines for *head*, *wing*, *back*, *body size*, genomic ancestry and 14 semi-diagnostic SNP loci. Shading shows probability density at a given distance. Arrows indicate positions of cline centers. *X*-axis reflects distance along the transect in kilometers (below) and position of sampling localities (above). Gray dots are individual values. Note abrupt transition in *head* plumage and displaced position of this cline compared to other morphological traits, the genomic ancestry and semi-diagnostic SNPs. Samples M and N appear to be parental *personata* in *head* plumage but are *alba*-like or intermediate in other morphological traits, genomic ancestry and allele frequencies of semi-diagnostic loci

not straightforward. The presence of several intermediate classes of *head* pattern, (Figure 2), a nonunimodal distribution of trait values, increased trait variance in the *head* cline centre, and a narrower zone width compared to a neutral diffusion scenario are all patterns consistent with a hybrid zone maintained by a balance between dispersal and selection, that is a tension zone (Barton & Hewitt, 1985; Endler, 1977; Gay et al., 2008; Key, 1968). However, analysis of SNP markers using STRUCTURE, PCA (Figure 1), variance analysis and geographic clines (Figure 3, Table 2) revealed that there is no corresponding genomic evidence for hybridization or introgression of molecular markers.

The vast majority of genotypes recovered by STRUCTURE using 14 semidiagnostic loci were classified as *alba* in the phenotypic hybrid zone and in adjacent samples of phenotypic *personata* (Figure 1). No genotypes corresponding to parental *personata* were detected in the hybrid zone. These results were consistent with geographic cline analysis, where the genomic ancestry cline was displaced >300 km south-eastwards from the phenotypic hybrid zone, in the region inhabited by birds with *personata* head patterns, but intermediate or *alba*-like body size measurements, wing and back plumage (Table 1, Figure 4, Table S2). The *head* cline was therefore embedded in populations mostly resembling parental *alba*, suggesting introgression of a restricted set of *personata* alleles underlying head plumage.

Variation in *head* plumage in wagtails seems limited to four discrete categories (Figure 2) rather than forming a continuum of intermediate phenotypes. This categorical variation suggests that head patterning may be underlain by few loci with large phenotypic effects, or by a few genomic regions, such as chromosomal inversions, where multiple genes are closely linked and protected from recombination. None of the semidiagnostic loci exhibited variation concordant with distinct *head* plumage patterns (Figure 4, Fig. S2), suggesting that our set of 19,437 loci did not capture any SNPs tightly linked to the genes underlying *head* pattern.

Divergent *head* plumage alleles could have arisen when *alba* and *personata* were in allopatry and then spread asymmetrically following secondary contact. Examination of historical records describing the distribution of *alba* and *personata* in the beginning of the 20th century in our study region (Sushkin, 1938; Zalesskii, 1927) suggests that the hybrid zone has moved approximately 60 km north-westward over the last 90–100 years. As head plumage is the diagnostic trait used to distinguish *alba* and *personata* in the field, it is likely that this comparison reflects historical movement of the *head* cline, with the *personata* head phenotype displacing the *alba* type. Plumage coloration is among the most prominent signal types involved in sexual selection in birds (Price, 2008; Scordato, Symes, Mendelson, & Safran, 2014; Seddon et al., 2013), and asymmetric

TABLE 2 Comparison of five CFIT7 models of hybridization scenarios for four morphological traits and genomic ancestry. The best models are boldfaced

	Head			Wing			Back			Body size			Genomic ancestry		
Model	Log L	AIC	AICc	Log L	AIC	AICc	Log L	AIC	AICc	Log L	AIC	AICc	Log L	AIC	AICc
Bimodal-no-introgression	−243.3	498.6	499.0	−56.2	124.3	124.8	**−439.4**	**890.7**	**891.2**	−179.0	370.0	370.5	−112.9	237.7	238.0
Bimodal-with-introgression	−229.1	474.2	475.0	−58.4	132.8	133.6	−438.1	892.3	893.1	−174.1	364.2	365.1	−59.4	134.8	135.3
Trimodal-no-introgression	−227.7	479.5	481.2	−44.8	113.7	115.5	−438.2	900.4	902.2	−173.6	371.2	373.1	−73.8	171.7	172.9
Trimodal-with-introgression	**−220.2**	**468.4**	**470.7**	−44.0	116.0	118.5	−449.6	927.1	929.6	−174.6	377.1	379.7	−59.4	146.8	148.4
Unimodal	−269.2	552.4	552.9	**−44.1**	**102.3**	**102.9**	**−438.3**	**890.5**	**891.2**	**−173.5**	**360.9**	**361.6**	**−56.8**	**127.7**	**128.1**

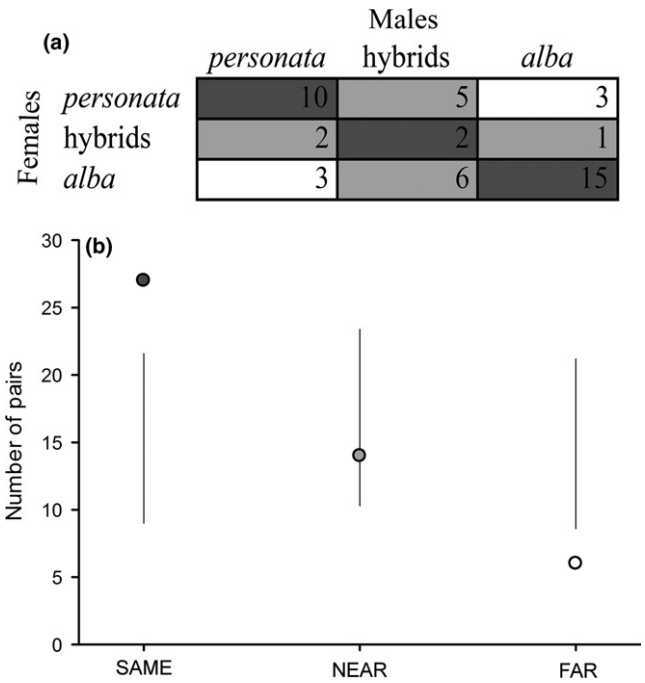

FIGURE 5 Composition of 47 social pairs observed in the centre of *head* geographic cline (a) and the results of randomization statistics (b). Dark grey—pairs from the SAME category, that is homotypic pairs and pairs between two hybrids; light grey—NEAR category that includes pairs between parental and hybrid; white—FAR category of heterotypic pairs. Vertical lines are 95% confidence intervals for number of pairs expected under random mating. Circles are observed values

introgression of traits with signalling functions has been reported in several vertebrate systems. In *Lepomis* sunfish (Garner & Neff, 2013), *Malurus* fairy wrens (Baldassarre & Webster, 2013; Baldassarre, White, Karubian, & Webster, 2014), *Manacus* manakins (Brumfield, Jernigan, McDonald, & Braun, 2001; Stein & Uy, 2006) and *Myzomela* honeyeaters (Sardell & Uy, 2016), asymmetric introgression is likely due to sexual or social selection favouring one signal type. Although we observed assortative mating by head plumage in the hybrid zone, assortative mating was not perfect; that is, we still observed some heterotypic pairings. Simulations suggest that a hybrid zone moves readily if alleles affecting mate choice have even a slight difference in selective advantage, even in the face of positive assortative mating and strong hybrid disadvantage (Brodin, Haas, & Hansson, 2013). This process can be easily overlooked in an empirical study, as documenting of such patterns requires large sample sizes for mating pairs. Moreover, there could be a strong assortative mating among social pairs, but asymmetric patterns of extrapair mating with respect to sexual signals (Baldassarre & Webster, 2013). Finally, nonadditive inheritance or epistasis may play a role in inheritance of head plumage. Therefore, the exact mechanism underlying asymmetric introgression of *head* phenotypes in wagtails remains to be determined.

The centres of clines for all traits except *head* were located in the region of northern Kazakhstan (Figure 1 and 4) that was not covered by sampling in our study. Despite substantial sampling efforts, we were unable to collect even a single individual in this region due to extremely low wagtail population density, suggesting that northern and southern *personata* may be mostly allopatric due to a distributional gap associated with inappropriate habitats. Hybrid zones tend to be trapped in areas of low-population density (Barton, 1979; Hewitt, 1975), and the consistent location of all clines except *head* may indicate that a genomic transition between wagtail populations is associated with a population density depression and consequent reductions in gene exchange between somewhat geographically isolated populations. Finally, the cline in *body size* followed the elevational gradient of the Altai mountains, representing a transition from steppe to taiga vegetation. In the vast majority of its distributional range, *personata* is associated with higher elevations than *alba* (Stepanyan, 1983; Sushkin, 1938). Weak ecological selection could therefore maintain the observed differences in *body size* and perhaps *wing* and *back*, as these traits share similar introgression rates (Table 1).

The majority of genomic differentiation we observed was driven by sample P from Uzbekistan, the southernmost data point of the transect, which also had comparatively high within-sample variance (Figures 1 and 4). Although this pattern of high variance may potentially be a sequencing artefact, for example due to lower read coverage or a higher proportion of missing data in the sample P, we did not observe any noticeable difference between sample P and other

samples for the number of filtered reads (Fig. S4) or proportion of missing data (sample P, 23% missing; average all other samples 20%, limits 8–28%). It is more likely that the genetic diversity of southern *personata* populations has a biological explanation. Analysis of mtDNA markers in the white wagtail previously revealed three haplotype groups with 0.4–1.3% sequence divergence that are associated with different parts of Eurasia and likely represent lineages that were isolated during the Pleistocene (Li et al., 2016; Pavlova et al., 2005). Previous studies revealed regions of admixture between two divergent mtDNA lineages in eastern and western parts of Eurasia (Li et al., 2016; Pavlova et al., 2005). However, our sample P from Uzbekistan included all three mtDNA haplotype groups in a single location and is the most genetically diverse white wagtail population sampled anywhere in the world (complete ND2 gene, G. Semenov unpublished data). These data suggest that southern *personata* in Central Asia could have experienced admixture between divergent lineages from different parts of Eurasia, and hence retains high genetic diversity, as revealed by mtDNA and genomic markers. This genetic diversity persists despite all individuals collected in sample P appearing to have typical *personata*-type plumage, emphasizing that plumage divergence in wagtails is highly discordant with variation in genetic markers, as was previously suggested (Pavlova et al., 2005).

The differences we observe in patterns of divergence and introgression between morphological traits and genetic markers may reflect a complex set of processes shaping different parts of the wagtail genome over space and time. Distinct head plumages appear to be maintained by divergent mating preferences, although any selection associated with these nominal subspecies differences likely targets a very small fraction of the wagtail genome, perhaps a few loci or genomic regions of major phenotypic effect. Broadly, the results of our study suggest that the mechanisms maintaining distinct *head* plumage types in the white wagtail may differ from the mechanisms that affect variation throughout most of the genome. Additional sampling and use of methods with higher genomic resolution are needed to resolve the mechanisms underlying complex introgression patterns in the two wagtail subspecies.

4.2 | Assortative mate choice, phenotypic signals and speciation

Previous studies have revealed similar patterns of discordance between phenotypic and genomic variation (Campagna et al., 2016; Mason & Taylor, 2015; Poelstra et al., 2014; Toews et al., 2016), wherein populations that are divergent in phenotype are genomically very similar. It is often speculated that these phenotypic differences could be maintained by assortative mate choice, but this is rarely tested in wild populations. Here, we show that divergence in a plumage signal may indeed maintained by divergent mating preferences.

Nonrandom mating has historically been considered a mechanism that can drive strong reproductive isolation (Fisher, 1958; Lande, 1981) and consequent reductions in gene exchange that lead to the completion of speciation (Jiggins & Mallet, 2000; Jiggins, Naisbit, Coe, & Mallet, 2001; Kirkpatrick & Ravigne, 2002; Mallet, 2005; Price, 2008). However, recent discoveries have revealed that genomic differentiation associated with conspicuous phenotypic differences can sometimes be maintained due to strong assortative mating in a very small fraction of otherwise homogenous genomes, particularly in insects, *Heliconius* butterflies (Kronforst et al., 2006; Van Belleghem et al., 2017), *Anopheles* mosquitoes (Turner, Hahn, & Nuzhdin, 2005) and birds, *Corvus* crows (Poelstra et al., 2014). The results of our study further suggest that the integrity of divergent phenotypic traits associated with assortative mating can be maintained within mostly undifferentiated genomes. These discoveries raise questions about the degree to which assortative mate choice alone contributes to the process of population divergence and speciation. For example, the patterns we found could be due to recent divergence between wagtail subspecies, wherein only plumage genes have evolved, and the rest of the genome remains comparatively undifferentiated. This would support an important role of assortative mating (and perhaps sexual selection) with respect to a few key genes as a driver of the earliest stages of speciation. Alternatively, the lack of genomic differentiation we observe may have little to do with the age of divergence, but exist because selection targeting few genomic regions is not effective enough to prevent admixture throughout most of the genome (Abbott et al., 2013). Additional studies that explore the mechanisms underlying associations between genomic and phenotypic variation and needed to more broadly examine the role of assortative mating in divergent evolution. In particular, comparing patterns of hybridization and introgression in hybrid zones at different stages of the speciation continuum in the same species can be a fruitful avenue for future research (Wagner & Mandeville, 2017).

The coloration of head plumage is the most pronounced phenotypic difference between the two wagtail subspecies studied in this study, and hence could be a speciation phenotype, that is a phenotypic trait whose divergence promotes speciation. Furthermore, a wide diversity of head plumages in other white wagtail subspecies (Alström & Mild, 2003) suggests that head and neck patterning might be a hot spot of morphological evolution in this species. Although the importance of particular traits to speciation might be limited to particular taxonomic groups, certain genes or gene pathways often play a prominent role across multiple divergence events (Hubbard, Uy, Hauber, Hoekstra, & Safran, 2010). For instance, mutations in the melanogenesis pathway genes can lead to discrete phenotypic variation involved in divergent evolution in mammals (Nachman, Hoekstra, & D'Agostino, 2003; Steiner, Weber, & Hoekstra, 2007), reptiles (Rosenblum, Hoekstra, & Nachman, 2004) and birds (Mundy, 2005; Uy, Moyle, Filardi, & Cheviron, 2009). Although the genetic architecture of phenotypes in the above examples is well characterized, the genomic basis of more continuous phenotypes is just beginning to be unravelled (Campagna et al., 2016; Poelstra, Vijay, Hoeppner, & Wolf, 2015; Toews et al., 2016; Van Belleghem et al., 2017) and requires the identification and analysis of such characters across a wide range of taxa.

ACKNOWLEDGEMENTS

We thank Maxim Mitropolskiy, Irina Malykh, Margarita Valukhina and Denis Nuridjanov for fieldwork assistance and Dmitry Taranenko for help processing field materials. We thank Nicolle Mode for help with randomization statistics and Laurene Gay for comments on CFIT7 analysis. We are grateful to Yaroslav Red'kin, Evgeniy Koblik, Andrei Lissovsky, Sergei Drovetski and members of the Safran, Duckworth and Badyaev laboratories for feedback. Comments from Robb Brumfield and an anonymous reviewer substantially improved the manuscript. Yaroslav Red'kin assisted with access to the collections of the Zoological Museum of Moscow State University. We thank Vladmir Loskot for granting access to collections at the Zoological Institute of the Russian Academy of Sciences. Brittany Jenkins did the genomic library preparations. We thank Alex Buerkle for help with genomic analyses. This work was supported by grants from the Russian Foundations for Basic Research 14-04-01492 and 12-04-31042 to GAS, 13-04-10063 to Zhanna Resnikova and GAS, the National Science Foundation DEB-CAREER 1149942 to RJS and by the G.G. Simpson Postdoctoral Fellowship from the University of Arizona to GAS. We are grateful to Yulia Sheina for the wagtail drawings used in Figure 1. Sampling was carried out according to the programme of fundamental scientific research of the Russian Academy of Sciences and Siberian Zoological Museums with necessary collecting permissions obtained in corresponding regions and countries.

CONFLICT OF INTEREST

No authors have any conflict of interest to declare.

DATA ACCESSIBILITY

Data sets from this study are available on Dryad, https://doi.org/10.5061/dryad.fg49r. RAD-seq reads are available on the NCBI Short Read Archive, BioProject PRJNA412922.

AUTHOR CONTRIBUTION

G.A.S. designed the study. G.A.S and D.R.K. collected data in the field. G.A.S. did museum skin preparation. G.A.S. analysed the data with input from E.S.C.S., N.C.K., C.C.R.S. and R.J.S. E.S.C.S., N.C.K., C.C.R.S. and R.J.S. designed the genomic work. E.S.C.S., N.C.K. and C.C.R.S. performed genomic assembly and S.N.P. calling. G.A.S. wrote the manuscript with input from E.S.C.S., R.J.S., N.C.K., C.C.R.S. and D.R.K. All authors signed off on the final manuscript.

ORCID

Georgy A. Semenov http://orcid.org/0000-0002-7218-7885
Elizabeth S. C. Scordato http://orcid.org/0000-0003-0224-8280
Chris C. R. Smith http://orcid.org/0000-0002-6470-3413
Rebecca J. Safran http://orcid.org/0000-0002-4762-2638

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246.

Abramoff, M. D., Magalhaes, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International*, *11*, 36–42.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *International symposium on information theory* (pp. 267–281). Budapest, Hungary : Akad'emiai Kiadi.

Alström, P., & Mild, K. (2003). *Pipits & wagtails of Europe, Asia and North America*. London, UK: Christopher Helm Publishers Ltd.

Alström, P., & Ödeen, A. (2002). Incongruence between mitochondrial DNA, nuclear DNA and non-molecular data in the avian genus Motacilla: Implications for estimates of species phylogenies. In P. Alström (Ed.), *Species limits and systematics in some passerine birds* (pp. 16–21). Uppsala: Acta Universitatis Upsaliensis.

Badyaev, A. V., Gibson, D. D., & Kessel, B. (1996). White Wagtail (Motacilla alba). In A. Poole & F. Gill (Eds.), *The birds of North America*. Philadelphia, PA: The Academy of Natural Sciences and The American Ornithologists' Union, Washington, D.C.

Baldassarre, D. T., & Webster, M. S. (2013). Experimental evidence that extra-pair mating drives asymmetrical introgression of a sexual trait. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 1–7.

Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution*, *68*, 2644–2657.

Barton, N. H. (1979). Gene flow past a cline. *Heredity*, *43*, 333–339.

Barton, N. H., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *57*, 357–376.

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), *Hybrid zones and the evolutionary process* (pp. 13–45). New York, NY: Oxford University Press.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology Evolution and Systematics*, *16*, 113–148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*, 497–503.

Bazykin, A. D. (1969). Hypothetical mechanism of speciation. *Evolution*, *23*, 685–687.

Berezovikov, N., & Reznichenko, S. (2014). Nesting of the masked wagtail Motacilla personata and cases of its hybridization with the white wagtail Motacilla alba in Bayanaul (North-East Kazakhstan). *Russian Ornithological Journal*, *23*(956), 73–78.

Berezovikov, N., Samusev, I., Khrokov, V., & Egorov, V. (2007). Passerine birds of Irtysh floodplains, & Altai foothills. *Russian Ornithological Journal*, *16*(371), 1031–1055.

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*, 2633–2635.

Brelsford, A., & Irwin, D. E. (2009). Incipient speciation despite little assortative mating: The yellow-rumped warbler hybrid zone. *Evolution*, *63*, 3050–3060.

Brodin, A., Haas, F., & Hansson, B. (2013). Gene-flow across the European crow hybrid zone – a spatial simulation. *Journal of Avian Biology*, *44*, 281–287.

Brumfield, R. T., Jernigan, R. W., McDonald, D. B., & Braun, M. J. (2001). Evolutionary implications of divergent clines in an avian (Manacus: Aves) hybrid zone. *Evolution*, *55*, 2070–2087.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*, 2 ed. New York, NY: Springer-Verlag.

Butlin, R., Debelle, A., Kerth, C., Snook, R. R., Beukeboom, L. W., & Castillo Cajas, R. F. … (2012). What do we need to know about speciation? *Trends in Ecology & Evolution*, *27*, 27–39.

Campagna, L., Repenning, M., Silveirad, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation driven by sexual selection. *Science Advances*, *3*(5), 1–11.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*.

Cramp, S. (1988). *Handbook of the birds of Europe, the Middle East and North Africa*. Oxford, UK: Oxford University Press.

Dougall, T. W. (1991). Winter distribution and associated movements of northern Pied Wagtails Motacilla alba yarrellii, as shown by ringing. *Ringing & Migration*, *12*, 1–15.

Ehrlich, P. R., Dobkin, D. S., & Wheye, D. (1988). *The birder's handbook Simon & Schuster*. New York, NY: London, Toronto, Sydney.

Endler, J. A. (1977). *Geographic variation, speciation and clines*. Princeton, NJ: Princeton University Press.

Fisher, R. A. (1958). *The genetical theory of natural selection*. New York, NY: Dover.

Garner, S. R., & Neff, B. D. (2013). Alternative male reproductive tactics drive asymmetrical hybridization between sunfishes (*Lepomis* spp.). *Biology Letters*, *9*, 1–4.

Gay, L., Crochet, P., Bell, D., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution*, *62*, 2789–2806.

Gay, L., Neubauer, G., Zagalska-Neubauer, M., Debain, C., Pons, J. M., David, P., & Crochet, P. A. (2007). Molecular and morphological patterns of introgression between two large white-headed gull species in a zone of recent secondary contact. *Molecular Ecology*, *16*, 3215–3227.

Glutz Von Blotzheim, U. N., & Bauer, K. M. (1985). *Handbuch der Vögel Mitteleuropas. Passeriformes*.

Gompert, Z., & Buerkle, C. A. (2016). What, if anything, are hybrids: Enduring truths and challenges associated with population structure and gene flow. *Evolutionary Applications*, *9*, 909–923.

Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, *23*, 4555–4573.

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, *66*, 2167–2181.

Grant, P. R., & Grant, B. R. (1997). Genetics and the origin of bird species. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 7768–7775.

Harrison, R. G. (1986). Pattern and process in a narrow hybrid zone. *Heredity*, *56*, 337–349.

Harrison, R. G. (1990). Hybrid zones: Windows on evolutionary processes. *Oxford Surveys in Evolutionary Biology*, *7*, 69–128.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*(Suppl. 1), 795–809.

Hewitt, G. M. (1975). Sex-chromosome hybrid zone in grasshopper Podisma pedestris (Orthoptera – Acrididae). *Heredity*, *35*, 375–387.

Hubbard, J. K., Uy, J. A., Hauber, M. E., Hoekstra, H. E., & Safran, R. J. (2010). Vertebrate pigmentation: From underlying genes to adaptive function. *Trends in Genetics*, *26*, 231–239.

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

Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, *15*, 250–255.

Jiggins, C. D., Naisbit, R. E., Coe, R. L., & Mallet, J. (2001). Reproductive isolation caused by colour pattern mimicry. *Nature*, *411*, 302–305.

Key, K. H. L. (1968). The concept of stasipatric speciation. *Systematic Zoology*, *17*, 14–22.

Kirkpatrick, M., & Ravigne, V. (2002). Speciation by natural and sexual selection: Models and experiments. *American Naturalist*, *159*, 22–35.

Kronforst, M. R., Young, L. G., Kapan, D. D., McNeely, C., O'Neill, R. J., & Gilbert, L. E. (2006). Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 6575–6580.

Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences of the United States of America*, *78*, 3721–3725.

Li, X., Dong, F., Lei, F., Alström, P., Zhang, R., Ödeen, A., … Yang, X. (2016). Shaped by uneven Pleistocene climate: Mitochondrial phylogeographic pattern and population history of white wagtail *Motacilla alba* (Aves: Passeriformes). *Journal of Avian Biology*, *47*, 263–274.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078–2079.

Lobkov, E. G. (2011). Kamchatka wagtail Motacilla (alba) lugens (Gloger, 1829): Variability, relationships with the spectacled white wagtail *Motacilla alba ocularis* (Swinhoe, 1860), and the taxonomic status. *Far Eastern Journal of Ornithology*, *2*, 27–55.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*, 229–237.

Mason, N. A., & Taylor, S. A. (2015). Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology*, *24*, 3009–3025.

Mayr, E. (1963). *Animal species and evolution*. Cambridge, UK: Belknap Press.

Mettler, R. D., & Spellman, G. M. (2009). A hybrid zone revisited: Molecular and morphological analysis of the maintenance, movement, and evolution of a Great Plains avian (Cardinalidae: Pheucticus) hybrid zone. *Molecular Ecology*, *18*, 3256–3267.

Mundy, N. I. (2005). A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proceedings of Biological Sciences*, *272*, 1633–1640.

Nachman, M. W., Hoekstra, H. E., & D'Agostino, S. L. (2003). The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 5268–5273.

Nagylaki, T. (1976). Clines with variable migration. *Genetics*, *83*, 867–886.

Nazarenko, A. A. (1968). On the character of interrelations of two forms of pied wagtails in south Ussuriland. *Problems of Evolution*, *1*, 195–201.

Nurnberger, B., Barton, N. H., MacCallum, C., Gilchrist, J., & Appleby, M. (1995). Natural selection on quantitative traits in the Bombina hybrid zone. *Evolution*, *49*, 1224–1238.

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21*, 2991–3005.

Pavlova, A., Zink, R. M., Rohwer, S., Koblik, E. A., Red'kin, Y. A., Fadeev, I. V., & Nesterov, E. V. (2005). Mitochondrial DNA and plumage evolution in the white wagtail Motacilla alba. *Journal of Avian Biology*, *36*, 322–336.

Payseur, B. A. (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, *10*, 806–820.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*, 2337–2360.

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., … Wolf, J. B. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, *344*, 1410–1414.

Poelstra, J. W., Vijay, N., Hoeppner, M. P., & Wolf, J. B. (2015). Transcriptomics of colour patterning and coloration shifts in crows. *Molecular Ecology*, *24*, 4617–4628.

Price, T. (2008). Speciation in birds.

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

Raufaste, N., Orth, A., Belkhir, K., Senet, D., Smadja, C., Baird, S. J. E., … Boursot, P. (2005). Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society*, *84*, 593–616.

Red'kin, Y. A. (2003). New data on spatial and reproductive relationships of some close forms of Passeriformes in Tyva, 201–206.

Rosenblum, E. B., Hoekstra, H. E., & Nachman, M. W. (2004). Adaptive reptile color variation and the evolution of the MC1R gene. *Evolution*, *58*, 1794–1808.

Ruegg, K. (2008). Genetic, morphological, and ecological characterization of a hybrid zone that spans a migratory divide. *Evolution*, *62*, 452–466.

Safran, R. J., Scordato, E. S. C., Wilkins, M. R., Hubbard, J. K., Jenkins, B. R., Albrecht, T., … Kane, N. C. (2016). Genome-wide differentiation in closely related populations: The roles of selection and geographic isolation. *Molecular Ecology*, *25*, 3865–3883.

Sardell, J. M., & Uy, J. A. (2016). Hybridization following recent secondary contact results in asymmetric genotypic and phenotypic introgression between island species of Myzomela honeyeaters. *Evolution*, *70*, 257–269.

Scordato, E. S., Symes, L. B., Mendelson, T. C., & Safran, R. J. (2014). The role of ecology in speciation by sexual selection: A systematic empirical review. *Journal of Heredity*, *105*, 782–794.

Scordato, E. S. C., Wilkins, M. R., Semenov, G., Rubtsov, A. S., Kane, N. C., & Safran, R. J. (2017). Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Molecular Ecology*, *511*, 229.

Seddon, N., Botero, C. A., Tobias, J. A., Dunn, P. O., Macgregor, H. E., Rubenstein, D. R., … Safran, R. J. (2013). Sexual selection accelerates signal evolution during speciation in birds. *Proceedings of Biological Sciences*, *280*, 20131065.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*, 176–192.

Semenov, G. A., & Yurlov, A. K. (2010). Relationships between masked and white wagtails on the south of Siberia. *Ornithologia*, *36*, 7–21.

Semenov, G., Yurlov, A., & Khaydarov, D. (2010). Hybridization of *Motacilla alba Linnaeus*, 1758, and M. (a.) personata Gould, 1861, in the south of Siberia. *Contemporary Problems of Ecology*, *3*(5), 579–586.

Shaw, K. L., & Mullen, S. P. (2011). Genes versus phenotypes in the study of speciation. *Genetica*, *139*, 649–661.

Stein, A. C., & Uy, J. A. C. (2006). Unidirectional introgression of a sexually selected trait across an avian hybrid zone: A role for female choice? *Evolution*, *60*, 1476–1485.

Steiner, C. C., Weber, J. N., & Hoekstra, H. E. (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biology*, *5*, e219.

Stepanyan, L. S. (1983). *Superspecies and twin-species in the avifauna of USSR*. Moscow, Russia: Nauka.

Stepanyan, L. S. (2003). *Conspectus of the ornithological fauna of Russia and adjacent territories (within the borders of the USSR as a historic region)*. Moscow, Russia: Academkniga.

Sushkin, P. (1938). *Birds of Soviet Altai and adjacent parts of North West Mongolia*. Moscow-Leningrad: Academy of Sciences of the USSR.

Szymura, J. M., & Barton, N. H. (1986). Genetic analysis of a hybrid zone between the fire-bellied toads, Bombina bombina and *B. variegata*, near Cracow in southern Poland. *Evolution*, *40*, 1141–1159.

Toews, D. P., 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 Biology*, *26*, 2313–2318.

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in Anopheles gambiae. *PLoS Biology*, *3*, e285.

Uy, J. A., Moyle, R. G., Filardi, C. E., & Cheviron, Z. A. (2009). Difference in plumage color used in species recognition between incipient species is linked to a single amino acid substitution in the melanocortin-1 receptor. *American Naturalist*, *174*, 244–254.

Van Belleghem, S. M., Rastas, P., Papanicolaou, A., Martin, S. H., Arias, C. F., Supple, M. A., … Papa, R. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. *Nature Ecology & Evolution*, *1*, 52.

Wagner, C. E., & Mandeville, E. G. (2017). Speciation, species persistence and the goals of studying genomic barriers to gene flow. *Journal of Evolutionary Biology*, *30*, 1512–1515.

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*, 851–865.

Zalesskii, I. (1927). Distribution borders of the Masked wagtail (Motacilla personata Gould.) in West Siberia. *Uragus*, *2*, 7–8.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Semenov GA, Scordato ESC, Khaydarov DR, Smith CCR, Kane NC, Safran RJ. Effects of assortative mate choice on the genomic and morphological structure of a hybrid zone between two bird subspecies. *Mol Ecol*. 2017;00:1–15. https://doi.org/10.1111/mec.14376