Covariation of genetic and epigenetic divergence in barn swallows

Sarah A. Mueller, Uthara Srinivasan, Zachary M. Laubach, et al.

Genome Res. published online June 16, 2026
Access the most recent version at doi:10.1101/gr.281470.125

Covariation of genetic and epigenetic divergence in barn swallows

Sarah A. Mueller,[1] Uthara Srinivasan,[1] Zachary M. Laubach,[2] Drew R. Schield,[2,3] Jan A.C. von Rönn,[4] Rebecca J. Safran,[2] and Jochen B.W. Wolf[1]

[1]*Department of Evolutionary Biology, Faculty of Biology, Ludwig Maximilian University of Munich, Planegg-Martinsried 82152, Germany;* [2]*Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA;* [3]*Department of Biology, University of Virginia, Charlottesville, Virginia 22904, USA;* [4]*Swiss Ornithological Institute, Sempach 6203, Switzerland*

As populations diverge, genetic differences accumulate across the genome. Although genetic differentiation is well documented, evolutionary change in epigenetic modifications is less well understood. Here, we integrate genomic and methylomic data from 168 barn swallows (*Hirundo rustica*) to examine the relationship between genetic and epigenetic population divergence at the resolution of CpG islands and single CpG sites. Sampling across all six extant subspecies and two hybrid zones, we assess the roles of genetic ancestry and environment in shaping DNA methylation patterns. Genome-wide methylation differentiation (P_{ST}) closely mirrors genetic differentiation (F_{ST}), with secondary contributions from geographic and environmental variation. DNA methylation patterns track the genomic landscape of sequence divergence, especially in CpG islands of nonregulatory regions. Across hybrid zones, DNA methylation levels are associated with nearby *cis*-acting genetic variation and global genomic ancestry. Moreover, CpG sites exhibit linkage disequilibrium with nearby SNPs, extending over several thousand base pairs within a putative inversion. These results suggest that although environmental effects on genome-wide DNA methylation are present, epigenomic patterns are costructured by the evolutionary processes shaping genetic variation.

[Supplemental material is available for this article.]

Evolutionary divergence arises from the accumulation of genetic, phenotypic, and ecological differences among populations (Dobzhansky 1937; Mayr 1942; Lewontin 1974; Bolnick et al. 2023). Its extent depends on the time since population separation and the interplay among mutation, genetic drift, migration, selection, and recombination. Well-established theories and empirical evidence indicate that the partitioning of genetic variation between populations, measured in relative (F_{ST}) or absolute (D_{xy}) terms (Nachman and Payseur 2012; Cruickshank and Hahn 2014), is promoted by genetic drift and is counterbalanced by migration, which typically declines with geographic distance (isolation by distance) (Slatkin 1985). Phenotypic divergence, as assessed by analogs of F_{ST} such as Q_{ST} or P_{ST} (Leinonen et al. 2013), may be further affected by divergent selection. If selection against nonadapted migrants is sufficiently strong among demes, this can enhance genome-wide divergence (isolation by environment) (Hendry 2004; Shafer and Wolf 2013).

Importantly, genetic differentiation does not accumulate uniformly across the genome (Wu 2001). Variation in mutation rate, recombination rate, and selection generates heterogeneous divergence landscapes (Cruickshank and Hahn 2014; Burri 2017), which, in turn, contain information about the underlying evolutionary processes (Ravinet et al. 2017; Wolf and Ellegren 2017; Edelman and Mallet 2021). Genome scans assessing the distribution of summary statistics of genetic variation along the genome are thus routinely used to infer process from pattern (Ellegren et al. 2012; Vijay et al. 2016; Tavares et al. 2018; Shang et al. 2023). Yet, the underlying processes can be challenging to pinpoint. Regions of elevated genetic differentiation, for example, may reflect divergent selection, which reduces effective migration locally (Gwee et al. 2025); indicate linked selection even in the absence of gene flow (Burri 2017); or flag a local genomic history of adaptive introgression (Rossi et al. 2024). Regions subject to divergent selection may further come into linkage disequilibrium (LD) if selection outweighs the rate of recombination, promoting genome-wide divergence in other parts of the genome (Barton and De Cara 2009; Schield et al. 2024). Yet, weaker, transient associations may also arise in single populations by genetic drift alone owing to cosampling of identical-by-descent haplotypes.

Although divergence of genetic variation has long been in focus, there is growing interest in whether epigenetic mechanisms might influence population divergence (Richards et al. 2017; Vogt 2017; Richards and Pigliucci 2021; Vernaz et al. 2022). Such interest necessitates an understanding of how epigenetic variation aligns with population divergence. Understanding epigenetic divergence relies on (1) accurately pinpointing the sources of epigenetic variation and (2) utilizing methods that relate epigenetic variation to known processes, like selection, linkage, or drift. Achieving this will require quantitative description with informative summary statistics that characterize epigenetic change along the genome, akin to genome scan approaches (de Carvalho et al. 2025). Adapting population genetic methods, which have a long history of identifying processes underlying genetic variation, to epigenetic data provides the possibility to elucidate whether

Corresponding authors: s.mueller@bio.lmu.de, j.wolf@biologie.uni-muenchen.de
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.281470.125.

epigenetic divergence landscapes mirror genomic divergence and to reveal which evolutionary processes (i.e., drift, divergence) could be influencing epigenetic variation.

Currently, the field of evolutionary epigenetics focuses on identifying differentially methylated CpG sites or regions and their environmental determinants. In rare cases, such differences can be linked to genetic variation, for example, via QTL approaches (Mueller et al. 2025). Such studies have shown varying degrees of correlated genetic–epigenetic population divergence (Carja et al. 2017; Johnson and Kelly 2020; Silliman et al. 2023; Merondun and Wolf 2025), and in plants, in which epigenetic variation appears to have a higher degree of autonomy (Quadrana and Colot 2016), there is evidence for a genetic basis of locally adaptive epigenetic variation (Dubin et al. 2015). At a regional genomic scale, previous studies provide further evidence for both *cis*- and *trans*-acting genetic influences on epigenetic profiles (Villicaña and Bell 2021; Merondun and Wolf 2025). However, studies fully characterizing epigenetic divergence landscapes in natural populations remain rare, particularly those considering genomic variation in the same individuals (Hagmann et al. 2015; van Moorsel et al. 2019; de Carvalho et al. 2025). Yet, such studies are essential to move from mere characterization of epigenetic differences to understanding how processes, like divergent selection, linkage, and drift, influence epigenetic divergence (Mueller et al. 2025). Whether epigenetic divergence promotes or hinders genetic divergence will ultimately depend on the source, heritability, and fitness consequences of DNA methylation (Greenspoon et al. 2022; Mueller et al. 2025).

Here, we integrate whole-genome resequencing (WGS) with reduced representation bisulfite sequencing (RRBS) for 168 barn swallows (*Hirundo rustica*), including individuals of all six known subspecies sampled from across the global species distribution (Fig. 1A). Our sampling further includes two hybrid zones between three subspecies *Hirundo rustica rustica/Hirundo rustica tytleri* and *H. r. rustica/Hirundo rustica gutturalis*. The parental subspecies show phenotypic, migratory, and genomic differences (Scordato and Safran 2014; Safran et al. 2016; Turbek et al. 2022) and evidence of backcrossing and reproductive isolation (Scordato et al. 2017, 2020; Schield et al. 2024). The system lends itself to decompose genetic, environmental, and methylation variation along several axes of increasing genomic resolution. First, pairwise population comparisons along a continuum of divergence within and among six subspecies allow us to assess genome-wide covariation in genetic (F_{ST}) and epigenetic differentiation (P_{ST}) in the native spatial and ecological context. Second, known heterogeneity in genetic differentiation, in part owing to sexual selection (Schield et al. 2024), permits comparison of local methylome divergence with the underlying genomic divergence landscape. Third, admixed populations from hybrid zones allow us to quantify the association between CpG methylation and genetic variants, disentangling *cis*- and *trans*-genetic associations, while largely controlling for environment and geography (i.e., a natural "common garden") (Mueller et al. 2025). Finally, examination of LD between CpG sites and genetic variants enables us to assess the range of *cis*-regulatory association in recombining and nonrecombining regions of the genome.

Results

RRBS of blood from adult barn swallows produced an average of 14 million reads per sample (range 9.5–336 million). After quality-filtering and excluding 1866 sex- and age-informative CpG sites distributed evenly across the genome (see Methods) (Supplemental Fig. S1), we retained two complementary data sets: (1) a site-based data set containing 376,106 single CpG sites (CpG sites hereafter) and (2) a region-based data set averaging DNA methylation across CpG islands and their flanks to obtain more stable estimates of CpG variation (Supplemental Fig. S2). In the region-based data set, DNA methylation values of 632,510 CpG sites were averaged across 11,456 CpG islands (CpG islands hereafter), and 171,127 CpG sites were averaged across 13,756 flanking regions (CpG island flanks hereafter, defined as 2 kb upstream of or downstream from CpG islands). We further annotated both data sets per their genomic context, classifying CpG sites and regions as promoters, exons, introns, transcription termination sites (TTSs), or intergenic regions. Although the relative representation of these categories varied, all functional categories were represented in each data set (Fig. 1G; Supplemental Table S1). WGS data generated for the same 168 individuals with an average coverage of 16.3 (range 1.8–63.1) yielded 4,985,699 segregating variants after filtering.

Covariation of genetic and DNA methylation variation: genome-wide

We first quantified genome-wide population differentiation for the SNP and CpG site–based data sets using principal component analysis (PCA). For the SNP data set, the first two principal component (PC) axes accounted for 14.34% of the variance and separated subspecies according to known population histories (Fig. 1B). The sedentary subspecies *Hirundo rustica savignii* (SA; $n = 8$) and *Hirundo rustica transitiva* (TR; $n = 8$) clustered together on PC1 axis, which further separated all other subspecies, except for a joint cluster of East Asian *tytleri* (TY; $n = 11$) and North American *Hirundo rustica erythrogaster* (ER; $n = 7$), reflecting a recolonization event from the Americas into Asia (Zink et al. 2006). F_{ST} values of pairwise population comparisons revealed complementary patterns. The highest F_{ST} values mirrored maximal distance on the PC1 axis between the sedentary subspecies and *gutturalis* (GU; $n = 12$). Populations of *rustica* (RUeu, $n = 27$; RUru, $n = 12$) showed very low levels of differentiation despite a wide distribution across Europe and Asia (Fig. 1C). Consistent with previous work, differentiation between parental populations of the *rustica/gutturalis* hybrid zone was elevated compared with the *rustica/tytleri* hybrid zone (Schield et al. 2024).

Population structure was partially recovered by PCA on PC2 axis for the CpG site–based data but explained less variation and was likely influenced by unaccounted variables (Supplemental Fig. S3). We therefore also conducted partial, distance-based redundancy analysis (dbRDA) of the CpG site–based data set controlling for a number of variables, including the impact of genetic variation. Overall, RDA accounted for 10.8% of the variation and largely mirrored population structure in the genetic data (Fig. 1D; Supplemental Table S2).

To quantify pairwise methylation differentiation between populations, we treated DNA methylation as a phenotypic trait and calculated P_{ST}, a surrogate measure of Q_{ST}, which quantifies the partitioning of phenotypic variance among populations (Fig. 1E). Because the additive genetic variance components for within- and between-population phenotypic variation were unknown, we chose the scalar parameter of the P_{ST} equation to ensure its distribution was comparable in scale to the observed F_{ST} (see Methods) (Fig. 1C,E). We then quantified covariation between P_{ST} and F_{ST}, which was robust to the choice of the unknown components of additive genetic variance (Supplemental Fig. S4).

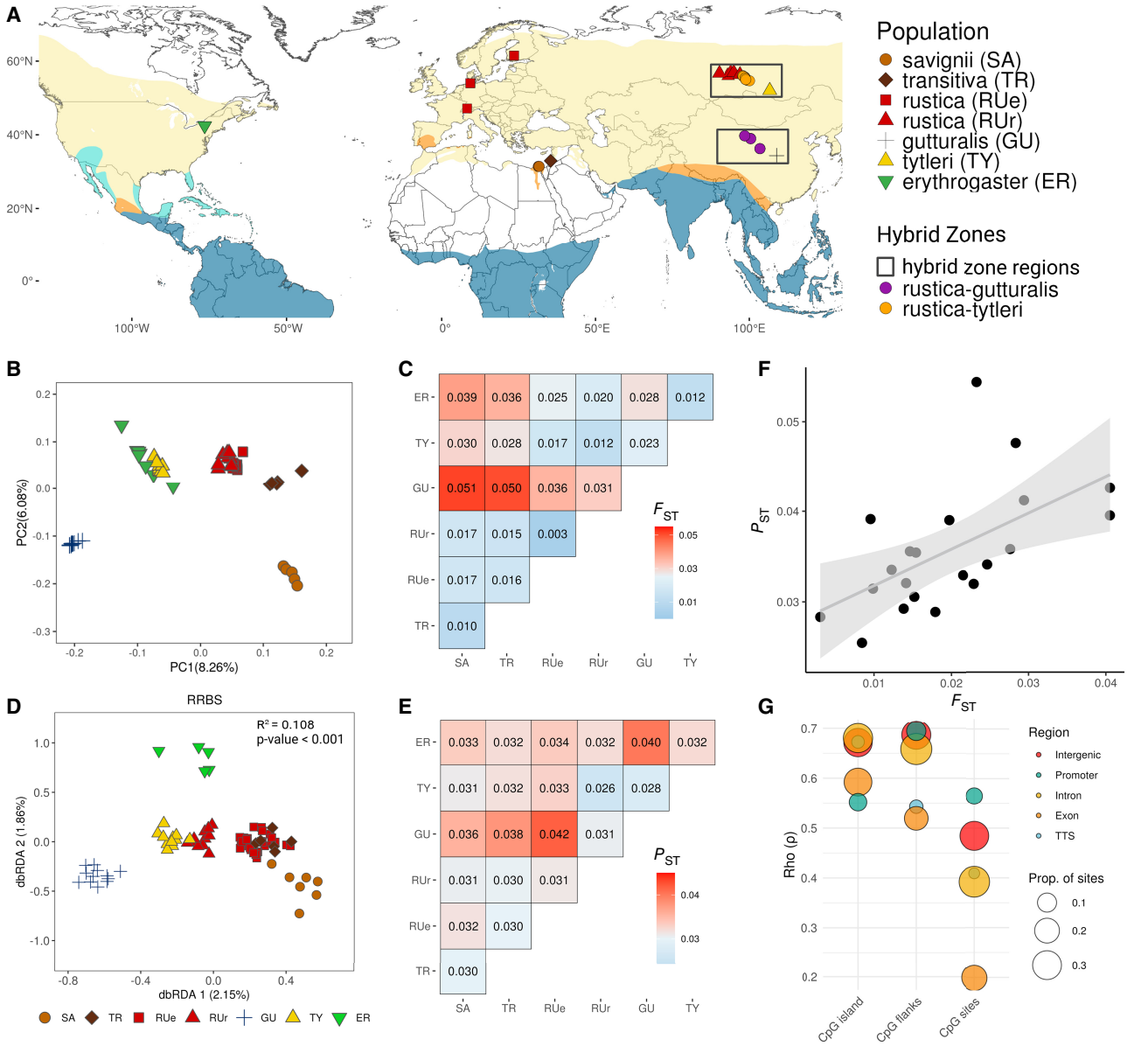

Figure 1. Genome-wide covariation of genetic and methylation differentiation in the barn swallow (*Hirundo rustica*). (*A*) Barn swallow individuals (n = 168) sampled across the global breeding range representing six different subspecies: *H. r. savignii* (SA; n = 8), *H. r. transitiva* (TR; n = 8), *H. r. rustica* (RU; n = 39: separated into European RUe, n = 27, and Russian RUr, n = 12), *H. r. gutturalis* (GU; n = 12), *H. r. tytleri* (TY; n = 11), and *H. r. erythrogaster* (ER; n = 7). Additionally, samples from two hybrid zones, one between *rustica* and *gutturalis* (n = 47) and another between *rustica* and *tytleri* (n = 36), were sampled. Distribution of the barn swallow habitat is shown, with distinctions for breeding (yellow) and nonbreeding or passage (light/dark blue) habitats for the migratory populations, with resident habitat (orange) for *savignii* and *transitiva*. (*B*) Principal component analysis of whole-genome resequencing data showing genetic population structure for the subset of 85 unadmixed individuals. (*C*) Heatmap of genome-wide pairwise F_{ST} estimates between subspecies, as well as between European and Asian populations of *rustica*. (*D*) RDA analysis of CpG sites on the same 85 individuals showing the best fit model including eight genetic, spatial, and environmental factors. (*E*) Heatmap of genome-wide pairwise P_{ST} values showing partial concordance with the genomic data set but with higher differentiation between GU and ER and GU and RUe. (*F*) Correlation of genome-wide F_{ST} and P_{ST} estimates for all CpG sites, for which one point represents one pairwise comparison between populations for each metric (ρ = 0.62, P-value = 0.003). (*G*) Comparison of the correlation strength separated by all CpG sites, CpG islands, and CpG flanks for each of the functional categories. The proportion of CpG sites/regions used for calculation is shown by size.

Population-wide F_{ST} and P_{ST} differentiation were significantly correlated across CpG islands (ρ = 0.69, P-value = 0.0007), across CpG island flanks (ρ = 0.66, P-value = 0.001), and for all CpG sites (ρ = 0.62, P-value = 0.003) (Fig. 1F; Supplemental Fig. S5). Elevated correlations observed for CpG islands and flanks likely reflects a combination of biological signal (Lövkvist et al. 2016) and the averaging of methylation values across multiple CpG sites (Supplemental Fig. S2).

The strength of this relationship also varied by choice of genomic context: It was consistently strong when estimating F_{ST} and P_{ST} for intergenic regions and introns and was lowest in exons (Fig. 1G). In addition to using F_{ST} as a metric for relative genetic

divergence, which has a clear theoretical relationship to P_{ST}, we also considered the absolute genetic divergence, D_{XY}, to test whether the accumulation of differences between populations may explain divergence in methylation levels better than the allelic differential captured by F_{ST}. In contrast to F_{ST}, P_{ST} did not show a significant correlation to D_{XY} for either the CpG region or sites for all annotation categories (Supplemental Figs. S5, S6).

Importantly, we did not find evidence that the correlations between P_{ST} and F_{ST} were inflated by environmental components covarying with genetic population structure. Neither P_{ST} nor F_{ST} showed covariation with environmental contrasts (Supplemental Table S3). Moreover, a notable, but not exclusive, contribution of genetic variation on CpG site differentiation was also detected in the partial dbRDA. All genetic, environmental, and spatial variables provided explained a significant, but small proportion of CpG site methylation variance with the genetic components explaining the largest component of variation (0.01 genetic, 0.008 spatial, 0.005 environmental) (Supplemental Table S2). Overall, these genome-wide results suggest that the divergence of epigenetic variation, as inferred from P_{ST}, closely aligns with differentiation of genetic variation with limited contribution of the environmental context.

Covariation of genetic and DNA methylation variation: along the genome

Next, we examined heterogeneity in genetic differentiation along the genome to assess the degree of covariation with differentiation in DNA methylation (for an example population comparison, see Fig. 2A). The 1% most highly differentiated CpG windows (10 kb) from P_{ST} estimates were concentrated around the most highly differentiated genetic regions (1% F_{ST} in 10 kb windows; hereafter F_{ST} outliers) (Fig. 2A). Observed distances between the top 1% of P_{ST} site estimates (hereafter P_{ST} outliers) and F_{ST} outliers were significantly lower than expected by chance in 72% of all possible pairwise population comparisons (Bonferroni-corrected type I error threshold $P<0.042$) (Fig. 2B; Supplemental Fig. S7). P_{ST} and F_{ST} outliers covaried across the genome in 10 kb windows regardless of whether we considered CpG islands (median ρ across all pairwise comparisons = 0.033), CpG island flanks (0.027), or single CpG sites (0.043) (Fig. 2C). Functional categories impacted the correlation. Intergenic regions had consistently higher correlations, whereas correlations were less consistent and weaker for TTSs and exonic regions (Fig. 2C). These results provide evidence that methylation divergence weakly covaries with the genetic differentiation landscape, in particular in putatively neutral, intergenic regions.

Covariation of genetic and DNA methylation differentiation: by ancestry in hybrid zones

Leveraging two hybrid zones in Asia (Fig. 1A), we further explored DNA methylation patterns in the admixed genomic background of hybrids, which can be represented as mosaics of ancestry blocks. Although the genetic variance in hybrids exceeds variation found within populations of a single subspecies, geographic and environmental variation is expected to be reduced. Hybrid zones thus enable the evaluation of the relative contributions of local *cis*-genetic ancestry and *trans*-acting genetic factors to differences in DNA

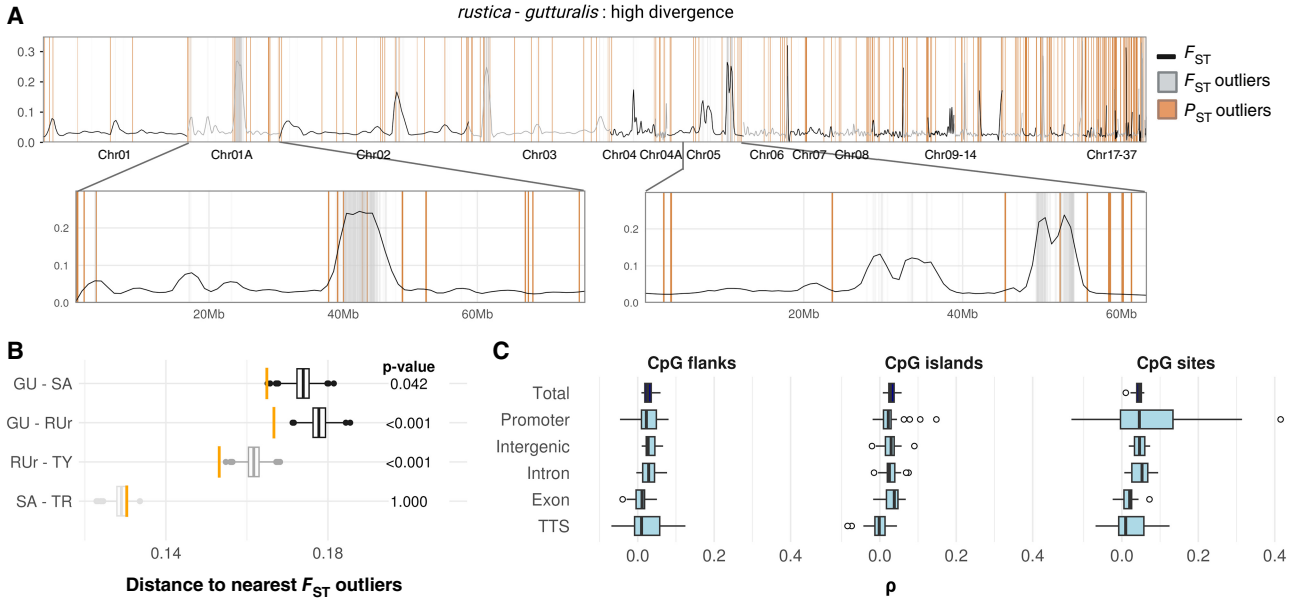

Figure 2. Covariation of genetic and methylation differentiation along the genome. (*A*) Pairwise comparison between *rustica* and *gutturalis* showing background F_{ST} in 10 kb windows (black line), along with the top 1% of F_{ST} windows (F_{ST} outliers, gray bars). CpG windows (10 kb) with elevated P_{ST} values (P_{ST} outliers, orange bars) are concentrated around chromosome ends and regions of elevated F_{ST}. (*B*) Empirical distances of outlier P_{ST} sites to the nearest F_{ST} outlier region, represented as proportion of chromosome length (orange) were significantly smaller than the expected distances obtained by random permutation of CpG sites (gray box plot spanning the first to third quantile with the median marked; whiskers extend to a maximum 1.5 times the interquartile range; outliers beyond are shown as points). Shown are four pairwise population comparisons representing high (*gutturalis/savignii* [GU-SA] and *gutturalis/rustica* Russia [GU-RUr]), medium (*rustica/tytleri* [RUr-TY]), and low (*savignii/transitiva* [SA-TR]) genome-wide F_{ST} values, shown in dark (high divergence) to light gray (low divergence) (for all 21 comparisons, see Supplemental Fig. S4). (*C*) Total correlation between F_{ST} and P_{ST} in 10 kb windows along the genome for CpG islands, CpG flanks, and all CpG sites, further broken down based on genomic region (promoter, intergenic, intron, exon, TTS). Methylation in intergenic regions and introns shows the highest consistent correlation given the F_{ST} background.

methylation while, in part, controlling for environmental variation between the parental populations.

We first characterized DNA methylation differences between the parental populations, yielding 1258 P_{ST} outliers in CpG sites (923, 1003), islands (152, 124) and flanks (183, 131; upper 1% of outliers), in *rustica*/*gutturalis* and *rustica*/*tytleri*, respectively. Pairwise comparison between *rustica* and *gutturalis* (genome-wide $F_{ST} = 0.03$) yielded higher estimates in P_{ST} outliers (mean = 0.24) compared with the comparison between the less genetically differentiated ($F_{ST} = 0.02$) *rustica* and *tytleri* (mean = 0.20, $P < 0.001$) (Fig. 3A,B). To examine the effect of ancestry, we then compared methylation levels of these outlier sites between parental populations and F_1 hybrids. Of the 1258 outliers, 281 islands/flanks and 724 sites showed significant differences between parental and/or hybrids in *rustica/gutturalis*, whereas 185 islands/flanks and 744 sites were significantly different in *rustica*/*tytleri*. Of these significant CpG regions/sites, 146 and 177 regions/sites were consistent with a dominant genetic architecture of *rustica* ancestry: Methylation was indistinguishable between *rustica* and the F_1 hybrids but differed significantly from the other parental group (*gutturalis* or *tytleri*). Dominance of *gutturalis* and *tytleri* was observed in 239 and 138 regions/sites, respectively. A codominant architecture, meaning that F_1 hybrid values were intermediate and differed significantly from both parentals, was only observed in one *rustica*/

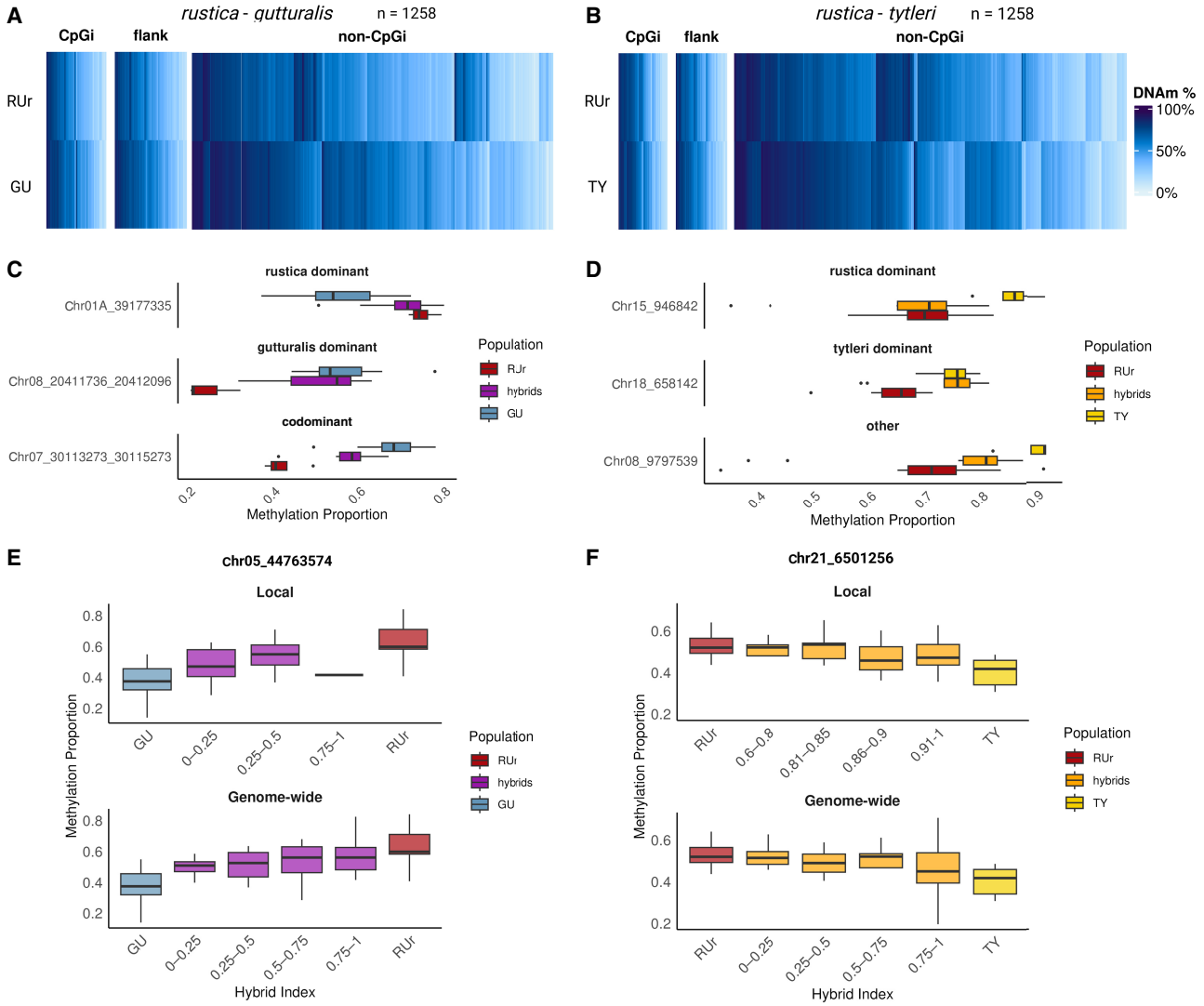

Figure 3. P_{ST} outliers between parental and hybrid individuals across the *rustica–gutturalis* and *rustica–tytleri* hybrid zones. (*A*,*B*) Mean methylation proportions of CpG islands, CpG flanks, and CpG sites identified as P_{ST} outliers (upper 1%) between parental populations. (GU) *gutturalis*, (RUr) *rustica*, (TY) *tytleri*. The *rustica/gutturalis* comparison has significantly higher P_{ST} estimates (mean = 0.24) compared with *rustica/tytleri* (mean = 0.20, $P < 0.001$). (*C*,*D*) Of these P_{ST} outliers, 79% and 73% were significantly differentially methylated between individuals with *rustica/gutturalis* and *rustica/tytleri* ancestry, respectively. The genetic architecture of these CpG regions/sites was consistent with dominance when hybrids shared values with one of the parents but differed from the other parent (dominant). Intermediate values of hybrids differing from both parents are consistent with codominance (codominant). Transgression, in which hybrid values exceed the variance of both parents, was not observed. One example from each category is represented here (no codominant loci were observed in the *rustica–tytleri* comparison), but further loci can be found in Supplemental Figures S8 and S9. (*E*,*F*) Methylation level as a function of *cis*-genetic (*above*) and a proxy of *trans*-genetic (*below*) ancestry measured as local and global hybrid index. (*E*) Example of a *rustica* dominant locus with evidence for an effect of the *cis*-genetic ancestry (*above*), but no effect of the *trans*-genetic ancestry (*below*). (*F*) Example of a nondominant CpG locus with evidence for an effect of both *cis*- (*above*) and *trans*-genetic ancestry (*below*).

gutturalis test. There was no evidence for transgression, in which hybrid values would significantly exceed any of the parental values. In the majority CpG islands/flanks/sites (*rustica/gutturalis* 619; *rustica/tytleri* 614), a specific genetic architecture could not be assigned and was classified as "other." CpG islands/flanks were more likely to be significant compared with CpG sites in the *rustica/gutturalis* comparison, likely because of lower measurement error when averaging across several CpG sites (Fisher's exact test *P*-value 0.02 and 0.63) (Supplemental Table S4). Example loci from each category are displayed in Figure 3, C and D, and a summary of tested loci can be found in Supplemental Figures S8 and S9.

Both a dominant and codominant architecture of the DNA methylation level could result from *cis*- or *trans*-genetic control (Wittkopp et al. 2004). To differentiate between the two, we fit a generalized linear model (GLM) with quasibinomial distribution and CpG methylation as the response using all parental and hybrid individuals. Explanatory variables consisted of the nearest SNP to the CpG region/site (*cis*-control), a local ancestry component ("extended" *cis*-control inferred from a 10 kb around the CpG region/site) and genome-wide hybrid indices that we used as a proxy for *trans*-control. Although genome-wide ancestry does not flag individual loci affecting methylation level, it provides a summary of cumulative, parental *trans*-effects present in the respective hybrids. The nearest SNP distances ranged from 1 bp–47,798 bp from the CpG region/site in *rustica*/*gutturalis* comparisons (median = 253 bp) and 1 bp–102,022 bp in *rustica*/*tytleri* (median = 276 bp). Local ancestry estimates were only available for 932 (*rustica*/*gutturalis*) and 853 (*rustica*/*tytleri*) tested CpG regions/sites. In cases in which these were missing, a model using only global ancestry and the nearest SNP was used.

Of the tested sites, the nearest SNP was significantly associated to 33 and 12 CpG islands/flanks and 85 and 40 CpG sites for *rustica*/*gutturalis* and *rustica*/*tytleri*, respectively (Fig. 3E,F). SNPs associated with significant changes in DNA methylation were within 2000 bp of the CpG region/site in *rustica*/*gutturalis* and *rustica*/*tytleri*, suggesting short-range *cis*-association, with 12 exceptions in the *rustica*/*gutturalis* comparison and two exceptions in *rustica/tytleri*. Global ancestry was significantly associated with CpG methylation in 71 and 79 cases, whereas local ancestry was significant in 62 and 32 cases in *rustica*/*gutturalis* and *rustica*/*tytleri*, respectively (Supplemental Table S5). Overall, hybrid DNA methylation appears to act as a quantitative trait associated with *cis*- and *trans*-genetic ancestry, with a strong impact of short-range *cis*-genetic effects of single allelic variants.

Covariation of genetic and DNA methylation variation: at base pair resolution

LD is a key statistic of genetic correlation that is influenced by recombination rate variation along the genome but also by population history and selection (Baird 2015). In the context of this study, LD may further provide insight into close-range association of *cis*-genetic variation and patterns of DNA methylation. To test for such an association, we encoded methylation values of all CpG sites as a categorical variable (see Methods) and incorporated it into linkage analysis. This allowed us to examine pairwise LD not only between SNPs but also between CpG sites and between pairs of SNP loci–CpG sites. For this purpose, we used the full CpG site data set without averaging across CpG islands or their flanks. The following analyses were conducted on the largest population, *rustica*.

LD between SNPs and between CpG sites showed a similar, rapid decay on average decreasing below R^2 values of 0.05 beyond physical distances of 1960 bp and 660 bp, respectively. LD between SNP–CpG pairs was mostly restricted to very close neighbors with R^2 values falling below 0.05 after 40 bp (Fig. 4A). For SNP–CpG pairs <40 bp apart, LD ranged from 0–0.93 with the majority of high LD pairs ($R^2 > 0.5$) occurring between SNPs and CpG sites with a median distance of 5 bp, similar to SNP–SNP LD with a median of 8 bp. LD decay was weakest in CpG islands for CpG–CpG pairs, particularly in islands of intergenic regions (Fig. 4A, bottom panel), a pattern not visible for LD between SNP–CpG (Supplemental Fig. S10). For CpG–CpG pairs, LD was also generally elevated in CG-rich microchromosomes, which was not seen for LD between SNP–CpG pairs but does appear between SNPs (Fig. 4B).

Elevated levels of both SNP–SNP and SNP–CpG linkage were observed on Chr 11, which we attribute to a 3.2 Mbp large inversion (Fig. 4C). Genetic variation of this region decomposes into three different clusters typical for inversions corresponding to the two alternative, homozygous diplotypes and the heterokaryotype (Supplemental Fig. S11). Average methylation across 21 highly linked CpG sites ($R^2 > 0.5$) mirrored genetic diplotype structure: DNA methylation was either hypo- or hypermethylated in the homozygote haplotypes, whereas the heterozygote showed intermediate values (haplotype 2) (Fig. 4D). Additionally, within the inversion, CpG sites showed LD (of $R^2 > 0.5$) to more than 100 SNPs with distances ranging from 2 bp to the entire inversion (Fig. 4E). These results suggest that genome-wide methylation is associated with the state of nearby genetic variation. It appears that given a lack of recombination, SNP–CpG site associations can extend across long haplotypes.

Discussion

We combined WGS and RRBS on the same 168 individuals across six subspecies of the barn swallow (*H. rustica*) in a single analytical framework. Treating DNA methylation of single CpG sites, islands and, flanks as a quantitative phenotype, we uncover covariance between genetic and DNA methylation divergence across multiple scales of genomic resolution.

DNA methylation and genetic divergence covary across the genome

At the genome-wide scale, DNA methylation divergence was best explained by genetic divergence, followed by geography and environment (Fig. 1). Evidence for isolation by ecology was weak, which may be partly owing to the limited number of accessible environmental parameters but may also reflect the life history of this migratory species integrating across a large number of habitats during the course of a year (Scordato et al. 2020; Turbek et al. 2022). P_{ST} covaried with F_{ST} rather than associating with environmental variables, particularly in intergenic regions, suggesting that a portion of DNA methylation divergence may share the dynamics of genetic drift. Considering the large measurement error associated with methylation levels of single CpG sites, this covariation is notable. Results for CpG islands (which average across several sites) support this notion, as correlation with genetic variation generally increased. Despite such technical limitations, our results concur with previous work showing that variation in DNA methylation is associated with genetic variation to different extents and depends on the genomic context (e.g., intergenic, intronic, exonic)

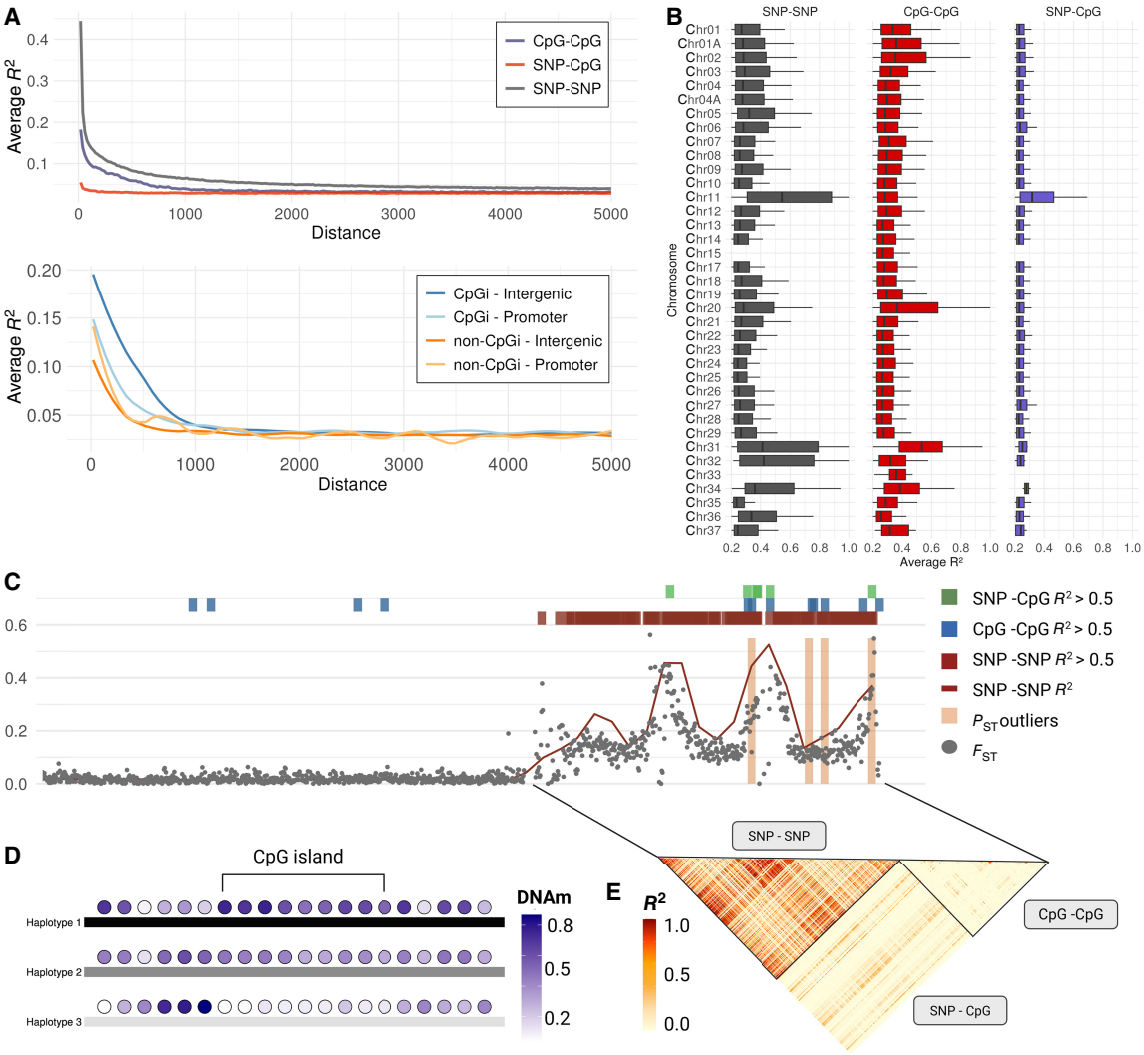

Figure 4. Linkage disequilibrium (LD) between and among SNP and CpG loci. (*A, top*) Decay of LD between pairs of SNP (SNP–SNP), between CpG sites (CpG–CpG), and between SNPs and CpG sites (SNP–CpG). LD between pairs of CpG–CpG is generally higher in CpG islands (CpGi) than CpG flanks or outside CpG islands (non-CpGi). It is also elevated in intergenic regions relative to promoters (*bottom*). (*B*) Chromosome average of pairwise LD between SNP–SNP, CpG–CpG, and SNP–CpG for loci with elevated LD ($R^2 > 0.2$). CpG pairs have elevated LD in several microchromosomes. LD between SNP–SNP and SNP–CpG pairs was elevated in Chr 11. (*C*) A candidate inversion on Chromosome 11 shows strong LD among SNPs (red), among CpG sites (blue), and between SNPs and CpG sites (green). It also harbors several P_{ST} outliers (orange bars) and regions of elevated F_{ST} (gray dots). The red lines show a running average of the LD for pairwise SNP–SNP comparisons. (*D*) Schematic showing average methylation at each of the 21 highly linked CpG sites ($R^2 > 0.6$) by diplotype of the inversion on Chr 11: diplotyes 1 and 2 represent the putatively recombining diplotypes; diplotype 3, the heterokaryotype. This region contains one CpG island. (*E*) Heatmap of LD along the inversion broken down by SNP–SNP, CpG–CpG, and SNP–CpG pairs. SNP–CpG pairs in this region show elevated levels of linkage to several SNPs, mirroring high LD between SNPs in this region.

(Dubin et al. 2015; Platt et al. 2015; Carja et al. 2017; Heckwolf et al. 2020; Hu et al. 2021; de Carvalho et al. 2025).

To better understand the source of covariation across the genome, we studied patterns of methylation and genetic divergence across the genome at a finer resolution. In addition to the global, genome-wide estimates, this analysis shed light on covariation in local genomic regions subject to linked selection, which include divergent sexual selection as well as background selection in the system (Schield et al. 2024). P_{ST} outliers clustered near F_{ST} peaks, and heterogeneity in P_{ST} covaried with the landscape of genetic differentiation (Fig. 2A,B), which is consistent with other studies attributing methylation divergence to variation in genetic divergence, in part contributed by selection (Venney et al. 2021; Ord et al. 2023; de Carvalho et al. 2025; Merondun and Wolf 2025). In contrast, in an evolution experiment exposing populations to contrasting environments, regions of elevated methylation divergence clustered in regions of reduced F_{ST} (Brennan et al. 2025). This urges for further investigation into the processes that may cause disparate patterns of P_{ST} and F_{ST}, including environmental control of CpG sites, selection, and time of divergence.

An important consideration is that all methylation data were generated from whole blood samples. Although blood methylation is expected to broadly reflect the patterns in other tissues (McKay et al. 2011; Derks et al. 2016), signals may differ substantially in tissues associated with phenotypic traits under selection, such as feather follicles in the swallow system (Schield et al.

2024). More generally, most variation in methylation is expected to occur among tissues (Merondun and Wolf 2025). Yet, depending on the source of methylation variation, different tissues might still have a similar degree of genome-wide covariation between genetic and epigenetic variation. Consistency in covariation is contingent on the proportion of spontaneous, *trans*-generationally stable, developmentally, and environmentally inert CpG sites (Mueller et al. 2025). Tissue specificity of epigenetic and genetic covariation clearly deserves more attention.

Local covariation persists in admixed hybrid genomes

Hybrid zones have been suggested to provide an underused, natural experimental setup to investigate the source of epigenetic variation and help to understand the relationship to genetic divergence (Mueller et al. 2025). The presence of hybrid zones in Asia, where there is known divergent selection on sexual signaling traits between *rustica/gutturalis* and *rustica/tytleri* (Schield et al. 2024), provided a complementary framework to study methylation patterns in the admixed genetic background of hybrids while reducing potentially confounding environmental variation. If methylation was associated with genetic variation, we expected higher CpG divergence in the genetically more diverged *rustica/gutturalis* comparison (Fig. 1). We indeed observed elevated P_{ST} in the *rustica/gutturalis* comparison than in *rustica/tytleri*. In 19% and 10% of 1005 and 929 tested P_{ST} outlier CpG regions/sites, the DNA methylation status of hybrids could be statistically predicted by the nearest SNP or local ancestry components. This result provides evidence that nearby *cis*-acting genetic variation affects, or is at least associated with, CpG methylation. Particularly in the *rustica/gutturalis* hybrid zone, this is more prevalent, which could be owing to elevated levels of divergence between these subspecies (Fig. 1). However, association in both hybrid zones with the global hybrid index also suggests a cumulative impact of genetic factors differing between ancestries. Given that global hybrid indices and geography covaried to some degree in this case, larger sample sizes would be needed to better disentangle the causes of the DNA methylation difference in hybrids.

Few studies have examined DNA methylation variation in the context of admixed hybrid genomes. Boman et al. (2024) and Höglund et al. (2020) provided evidence that both *cis*- and *trans*-acting genetic variation can influence DNA methylation. In line with Boman et al. (2024), who studied F_1 hybrids of collared and pied flycatchers, we found that genetic variants in close proximity to CpG sites were predictors of population-level differences in barn swallow hybrids. Notably, the correlation between genetic variation and DNA methylation within CpG islands was reduced in promoter regions, suggesting that functional constraints may limit the impact of *cis*-genetic variation in regulatory regions compared with intergenic regions. Such constraints could have important consequences for the regulation of gene expression. Boman et al. (2024) also reported misexpression in F_1 hybrids. Although we did not test the effect of DNA methylation on gene expression, we found no evidence for transgressive methylation levels, arguing against a simple linear contribution of DNA methylation variation to hybrid misexpression.

Association between SNP alleles and CpG methylation

In the hybrid zone analysis, the nearest SNP explained methylation proportions better than local ancestry, indicating that allele specific differences at single sites are associated with differences in nearby CpG methylation. In hybrid zones, *cis*-genetic associations between SNPs can arise because of admixture of differentiated parental population and can be enhanced by divergent selection in regions of the genome shielding parental haplotypes against interspecific recombination. Both processes act in the swallow hybrid zone (Schield et al. 2024). As we observe a stronger impact of the nearest SNP in the *rustica/gutturalis* hybrid zone, this could indicate that higher levels of genetic divergence impact nearby CpG methylation, either by induction or by linkage, as this hybrid zone shows strong signals of divergent selection.

To reduce any association arising from population divergence (F_{ST} vs. P_{ST}) followed by admixture, we quantified the degree of nonrandom associations between SNPs and CpG sites in a single population. Also in this context, we found a strong association of *cis*-acting SNP loci with CpG methylation in short distances (<100 bp). In the context of a single population, weak LD can also arise as a corollary of genetic drift, generating random, transient associations by cosampling CpG and SNPs in identical-by-descent haplotypes. Weak, short-range association can thus not be separated from a process of causal interactions between genetic and epigenetic variation.

We also found LD between genetic variants and CpG methylation to extend over substantially longer distances within an inversion on Chr 11, advocating a close association of CpG methylation with genetic variation in extended haplotypes in regions of reduced recombination. Studies in humans have similarly found that inversions, more so than point mutations, especially in the vicinity of breakpoints, are associated with nearby methylation at CpG sites (Zhang et al. 2019; Billingsley et al. 2024).

From pattern to process

The joint analysis of population-level variation in DNA methylation and genomic variation revealed covariation in patterns across multiple genomic scales. Several mechanisms could explain this association. One possibility is that CpG methylation is causally influenced by genetic variation that lies in binding sites or disrupts other regulatory proteins or activity (Martin-Trujillo et al. 2020; Villicaña and Bell 2021). Our finding that CpG methylation was predicted by adjacent SNPs, which are typically short distances away, is consistent with this idea and provides a plausible mechanism for generating population-specific methylation profiles during ontogenetic reprogramming (Reik et al. 2001). Second, CpG methylation can induce genetic mutations through deamination, such that differences in methylation between populations may translate into genetic differences (Hazarika et al. 2022). Presence/absence of TEs in the parental populations is a key candidate for differential methylated regions. Because we excluded repetitive sequences from this study, this mechanism is unlikely to explain *cis*- and *trans*-genetic associations observed here. Third, covariation may also arise if spontaneous epi-mutations (Johannes and Schmitz 2019) are subject to the same processes as genetic variation, including genetic drift and linked selection. This would likewise cause genetic and DNA methylation divergence to accumulate in similar regions of the genome, without a causal link between the two.

As outlined above, these contrasting mechanisms would also generate similar patterns of LD between CpG and SNP loci (Mueller et al. 2025). Under a scenario of linked selection with elevated LD between SNPs (Burri 2017), a CpG site that is a causally influenced by a genetic variant would naturally show LD to the other surrounding SNPs. However, the same pattern may arise if spontaneous or even environmentally induced, yet *trans*-

generationally stable CpG mutations are subject to selection and become associated with neighboring, neutral SNPs (Greenspoon et al. 2022; Planidin et al. 2025). Selection affecting levels of methylation is at least plausible for the subset influencing gene expression levels, which has been shown in several cases (Anastasiadi et al. 2018).

Disentangling the drivers of LD between SNP–CpG pairs, as well as the broader correlation between genetic and epigenetic variation, remains a substantial challenge. Ultimately, progress will require both theoretical predictions (Greenspoon et al. 2022; Planidin et al. 2025) and empirical work designed to isolate the sources of epigenetic variation (Hawe et al. 2022; Boman et al. 2024). With due caution, we suggest that our results are consistent with genetic variation causally influencing methylation for several reasons. First, the environmental contribution in our data appears secondary to the effects of genetic background. This makes environmentally induced epigenetic changes along ecological gradients less likely to be the primary driver of epigenetic and genetic covariance, especially given the importance of sexual selection in shaping population divergence in the system (Schield et al. 2024). Second, patterns of LD between CpG sites and SNPs were detected within a single *rustica* population, reducing the likelihood that LD caused by admixture or divergent selection alone explains the observed correlations. Finally, although *trans*-generationally stable spontaneous epimutations have been reported mainly in plants (Johannes and Schmitz 2019), we observed covariation spanning all major functional categories of the genome, although to different degrees. Taken together with the notion that epigenetic reprogramming in vertebrates is thought to constrain the flexibility of 5mC methylation (Reik et al. 2001), these considerations point to a plausible causal role of genetic variation in shaping methylation divergence in this system. Yet, ultimate characterization of the, likely variable, source of epigenetic variation across the genome will require experimental manipulation (Taudt et al. 2016).

In summary, our findings show that DNA methylation divergence in barn swallows is closely associated with the underlying genetic variation, particularly through *cis*-acting effects and in regions under divergent selection. Although our data cannot definitively resolve causality, they highlight genomic contexts in which covariation is strongest, namely, genomic regions subject to linked selection and extended haplotypes within a putative inversion. Such regions provide promising candidates for functional validation in tractable systems. Experimental approaches, such as CRISPR-mediated allele replacements or controlled crosses, could directly test whether altering specific genetic variants induces predictable methylation changes. Such studies are essential to disentangle the causal architecture of DNA methylation divergence in natural populations. Overall, our results are consistent with the view that DNA methylation acts in part as a genetically encoded phenotype, with variation shaped by local genomic context and evolutionary history.

Methods

Sampling design

A total of 180 blood samples from barn swallow (*H. rustica*) individuals from six different subspecies and their hybrids were sampled to investigate both epigenetic and genetic variation (Supplemental Table S6). Hybrids were represented by 36 samples from the *rustica–tytleri* hybrid zone in Russia and 47 samples from the *rustica–gutturalis* hybrid zone in China (Fig. 1; Schield et al. 2024). For a subset of 168 adult individuals, we generated (n = 27) or obtained publicly available WGS data (n = 141). These same samples were also subjected to RRBS, in addition to 12 age-heterogeneous individuals who were used to remove age-influenced DNA methylation sites from the final data set (for more details, see Supplemental Methods). In total, we generated 225 libraries including all individuals of adult life stage (>1 year, n = 168), the data set used for the estimation of age effects (n = 17), and two additional technical replicates for 10 individuals each of *rustica* and *gutturalis* populations (n = 40) flanking the hybrid zones to test for data reliability. For an overview on sample breakdown by population and sequencing method, see Supplemental Table S7.

DNA extraction and sequencing

DNA was extracted using Qiagen DNeasy kits with the addition of RNase A. WGS for the 27 *rustica* samples from Europe was conducted by Novogene using Illumina NovaSeq 6000 for 150 bp paired-end reads. RRBS was also performed by Novogene using Illumina NovaSeq 6000 for 150 bp paired-end reads with a lambda phage spike in to assess conversion efficiency (see Supplemental Methods). Raw sequencing data can be found at the NCBI BioProject database (see Data access) (for per-sample accession numbers, see Supplemental Table S6).

WGS data

Genotyping

Genotypes of publicly available data were directly obtained as VCF files from the authors (Schield et al. 2024) from a sequencing effort yielding an average sequence coverage of 16.3 (range 1.8–63.1). Genomic variants from all newly generated sequences were called using the same methods as previously described (Schield et al. 2024) to ensure consistency across data set, largely following the GATK v4.0.8.1 best-practices workflow (McKenna et al. 2010; Van der Auwera et al. 2013). The VCF files from both data sets were merged and filtered into three genomic data sets tailored to specific analyses: (1) all 4,985,699 SNP loci (with ≤10% missing data and 0.05 minor allele frequency) that survived quality filtering, (2) 2,468,648 unlinked SNPs for inference of population structure (PLINK v1.9, 50 10 0.1), and (3) a VCF file with invariant loci for window-based genetic diversity estimations in pixy (Korunes and Samuk 2021).

Statistical analysis

F_{ST}

Genome-wide, pairwise population F_{ST} values were calculated using VCFtools (v0.1.16) (Danecek et al. 2011). F_{ST} is a measure of genetic differentiation among populations, quantifying the proportion of the total genetic variation attributed to population level variation. Here, we separated Russian *rustica* samples from all European *rustica* samples, which we grouped into one population owing to the apparent lack of genetic structure, with analogous grouping for the methylation data set (Supplemental Fig. S12).

Principle component analysis

We performed a PCA using PLINK (v1.9) (Chang et al. 2015), allowing an unbiased estimation of the general trends in allele frequencies. This analysis was conducted on 85 barn swallow individuals from the six known subspecies on the unlinked, filtered SNP data set.

RRBS data

Data processing

Raw RRBS reads were trimmed with cutadapt v3.4 (Martin 2011), then mapped and methylation-called with Bismark v0.23.1 (Krueger and Andrews 2011) against the barn swallow reference genome (GCA_015227805.2) (Secomandi et al. 2023). We kept sites that mapped to major chromosomes and filtered out sites that did not map to predicted cut-sites, fell in repeat regions, overlapped with any SNP from WGS data, or had coverage of fewer than 20 or greater than 200 reads. We then generated site-level and region-level (CpG islands and CpG flanks) data sets for downstream analyses. In short, the site-level data set contained counts for all CpG sites, and the region-level data set averaged counts across annotated CpG islands and flanks (±2 kb from an island). In both site- and region-level data sets, consistent low (<10%) or high (>90%) methylation in all individuals, outlier individuals per site, and age-/sex-associated CpGs were removed. The final filtered data sets comprised 376,106 CpG sites and aggregated methylation across 11,456 CpG islands and 13,756 flanking regions. Full details including software versions, commands, and parameters for filtering are provided in the Supplemental Methods.

Statistical analysis: P_{ST} calculation

We calculated P_{ST}, an analog of Q_{ST}, which provides a statistic to quantify population-level differentiation for a phenotypic trait, here DNA methylation. Analogous to F_{ST}, P_{ST} is constructed as a ratio relating within-population to between-population variance components (Brommer 2011; Leinonen et al. 2013). As such, P_{ST} enables direct comparison of epigenetic and genetic divergence in natural populations (Johnson and Kelly 2020; Silliman et al. 2023). P_{ST} was calculated using the Pstat package (Da Silva and Silva 2018) in R (R Core Team 2025) using the proportion of methylation as the "phenotypic" variable of interest. We chose Pstat because the program uses a sum of squares approach, which should be robust to sample size differences between populations. Yet, simulating data sets, we found it to be sensitive to outliers, which prompted us to restrict the analyses to the data set with outlier removal described in the Supplemental Methods. P_{ST} values were calculated both for the filtered CpG site data set and the aggregate data set for CpG islands and flanks. The Pstat package requires a numeric value for the c/h^2 (csh in the package) parameter, where h^2 is the heritability of the trait within populations, and c denotes the proportion of the variance in the trait that is caused by effects of additive genetic variation between populations (Brommer 2011). After parameter testing, we chose a csh value of 0.05, which restricts the distribution of P_{ST} values to the range of F_{ST} values for illustrative purposes (Fig. 1C,F). As we are interested in the covariation across populations, we believe this choice does not impact the relationship between the two statistics (Supplemental Fig. S4; Supplemental Methods).

Joint analysis of WGS and RRBS data

Variance partitioning

We performed a partial dbRDA to quantify the influence of genetic, geographic and environmental variables on DNA methylation variation. Methylation proportion of CpG sites containing no missing data formed the response variable (122,842 CpG sites). Explanatory variables included eight factors, including the first PC axis from the PCA on the genetic variation, two environmental factors (temperature and precipitation), the subspecies and population assignment, and longitude and latitude as spatial variables. We used sex as the covariate to be held constant for the analysis. To minimize the impact of outliers, the DNA methylation proportion data were transformed using the Hellinger transformation (vegan package) (Dixon 2003). We used the forward.sel function with 1000 permutations in R to assess the statistical significance of explanatory variables during forward selection. At each step, adjusted R^2 (R_{adj}^2) was recalculated to retain only variables that significantly improved the model, resulting in a parsimonious RDA with the highest R_{adj}^2 (Dray et al. 2006).

P_{ST} versus F_{ST}

To investigate the relationships among genetic, epigenetic, geographic, and ecological factors, we constructed several distance matrices. We calculated the Haversine distance between the average sampling locations for each population to form a geographic distance matrix. We used the Bray–Curtis distance to measure the dissimilarity between populations based on ecological variables, including minimum and maximum temperature, precipitation, and pressure. We completed each analysis on Manhattan distances of methylation proportions between populations. We performed all analyses on all CpG sites as well as CpG islands, CpG island flanks, and non-CpG island sites. We used a genetic distance matrix, calculated using the adegenet package (Jombart 2008) in R, employing Manhattan distance on genetic markers. We performed a series of Mantel tests and partial Mantel tests to evaluate the correlations between these distance matrices using the vegan package (https://cran.r-project.org/web/packages/vegan/index.html). The Mantel test is a permutation-based procedure used to test the correlation between two matrices, and the partial Mantel test extends this by examining the relationship between two matrices while controlling for the effect of a third matrix. The DNA methylation distances always formed the response matrix, and we tested each of the three explanatory matrices (genetic, geographic, ecological) separately. We then extended to a partial Mantel test, testing each explanatory matrix while removing variation from the others.

Next, we investigated the relationship between phenotypic divergence (P_{ST}), in this case DNA methylation, to genetic variation along the genome using a genome-scan approach. In doing so, we wanted to evaluate for pairwise comparisons on whether methylation divergence (P_{ST} peaks) is more likely to occur in regions of high genomic differentiation (F_{ST} peaks). CpG sites show different rates of methylation and demethylation depending on the distance between CpGs (Lövkvist et al. 2016), with CpG islands generally being correlated to higher CpG density (Han et al. 2008). Therefore, we calculated P_{ST} separately on the CpG site and the CpG island/flanks data set. We then averaged methylation across 10 kb windows, a size chosen for comparison to the genomic data. Here, we only kept windows with a minimum of three CpG sites or CpG islands for the calculation. Each windowed methylation value was subsequently used for P_{ST} calculations. We looked at the correlation between average P_{ST} and F_{ST} per window across the genome compared with a randomized data set to determine how DNA methylation and genomic variation covary across the genome.

The second goal was to examine spatial co-occurrence of regions of elevated DNA methylation divergence with regions of elevated genomic differentiation. We calculated the minimum distance of the P_{ST} outliers (top 1% of P_{ST} values) to the nearest F_{ST} outliers (top 1% of F_{ST} values in 10 kb windows). We then compared this distribution to that of 1000 randomly sampled P_{ST} values compared with F_{ST} peaks to determine if the P_{ST} peaks are closer to F_{ST} peaks than would be expected by chance. Across all pairwise combinations of populations, we examined the

association between P_{ST} and the distance to highly diverged genomic regions.

Hybrid zone analysis

Two hybrid zones between three subspecies in Russia and China (*rustica/gutturalis* and *rustica/tytleri*) with known patterns of selection and genomic regions linked to traits underlying phenotypic differences were characterized by Schield et al. (2024). We first identified ancestry informative SNPs ($F_{ST} > 0.8$) between the parental populations for both hybrid zones and calculated individual hybrid indices for both parental and hybrid individuals using the introgress v 1.2.3 package in R (for hybrid triangle plots, see Supplemental Fig. S13; Gompert and Buerkle 2010). We also used the ancestry informative SNPs (*rustica/gutturalis* 73, *rustica/tytleri* 79) to assign each individual to a hybrid or parental class using NewHybrids software (Anderson and Thompson 2002) via parallelnewhybrid in R (Wringe et al. 2017). To evaluate the performance of NewHybrids and the ability of these SNPs to correctly identify hybrids, we used the hybriddetector package in R (Wringe et al. 2017). We discarded the first 100,000 generations as burn-in and estimated parameters from the following 200,000 MCMC algorithm iterations. This analysis resulted in a probability for each parental and hybrid of originating from one of six classes (pure parental, F_1, F_2, or backcross to either parental). Simulation-based validation showed high and stable performance for pure and F_1 classes, whereas later-generation hybrids showed variable assignment probabilities; therefore, downstream analyses focus on high-confidence assignments only (assignment probability > 0.9). For each hybrid zone, the identified pure parental samples and the F_1 hybrids were then used to characterize the genetic architecture of methylation at CpG loci at P_{ST} outliers (upper 1% of tested sites, $n = 1258$). To test differences in mean methylation among groups at each site, we applied one-way ANOVA with population as a factor in CpG sites at P_{ST} outliers. In cases in which overall ANOVA tests were significant ($P < 0.05$; $n = 1005$ *rustica-gutturalis*, $n = 929$ *rustica-tytleri*), we performed Tukey's HSD post hoc tests to identify pairwise group differences.

To evaluate the impact of *cis*- (local genetic ancestry) and *trans*-effects (genome-wide ancestry), we used a GLM to test the association of the methylation value at significant candidate CpG regions/sites with both genetic determinants and geography. We estimated global hybrid indices for each individual using introgress in R (Gompert and Buerkle 2010) based on the ancestry informative SNPs. For *cis*-effects, we included two explanatory variables. First, we included the closest SNP to the CpG region/site encoded as 0, 1, 2, where 1 represents the heterozygote. SNPs were chosen that had a minor allele frequency >0.1 to ensure no rare variants or possible genotyping errors were included. Second, local ancestry blocks were inferred as one 10 kb region around the focal CpG region/site for each hybrid individual using introgress; 10 kb was chosen as a compromise between maximizing the available data for ancestry estimation (average 50 SNPs) and minimizing recombination breaking up local ancestry. For introgress, we identified SNPs from the 10 kb flanking region that show differences in allele frequencies between parental populations of at least 0.1 and used these SNPs to estimate hybrid indices. This allowed us to represent hybrid genomes as mosaics of ancestry blocks, enabling the differentiation of *cis*- and *trans*-effects. Next, we fit a GLM with a quasibinomial distribution for methylation proportions at significant highly diverged P_{ST} regions/sites. The model included local ancestry and global hybrid index as explanatory variables. Model selection was performed using AIC to identify the best-fitting model for each locus. The significance of associations was determined using Wald tests with a significance threshold of $P < 0.05$.

Linkage analysis

To enable the calculation of LD between SNPs and CpG sites, the methylation data set was converted into epi-genotypes. For each CpG site, methylation proportions were categorized into three states: (1) hypo-methylated, <30% methylation was encoded as "0"; (2) intermediate methylation, 30%–70% methylation was encoded as "1"; and (3) hypermethylated, >70% methylation was encoded as "2." This categorization allowed us to treat methylation states analogously to diploid "epigenotypes" (e.g., homozygous and heterozygous states). Assuming that these "epi-genotypes" represent methylation states of the two chromosomes in single cells across the cell population under investigation, LD between CpG sites and SNPs may reflect an association of genetic and epigenetic variation on a haplotype. Alternatively, heterozygous epi-genotypes may refer to the signal of a heterogeneous cell population with half of the cells being methylated and half of the cells being unmethylated. Under this assumption, we would not expect to observe any LD.

For SNP and CpG data, epi-genotypes were further filtered for a minor allele frequency of 0.2 to minimize spurious effects of rare or erroneous SNP or CpG sites. LD was calculated using the R^2 metric, which quantifies the correlation between alleles at two loci using PLINK (v. 1.9) (Purcell et al. 2007) on the combined SNP and epi-genotyped CpG data set. We used `--ld-window 100` and `--ld-window-kb 70` as the window sizes for analysis, meaning LD was calculated for up to 100 pairs of loci within a maximum physical distance of 70 kb. After running PLINK, the loci were split post hoc into three categories of locus pairs: SNP–SNP, CpG–CpG, and SNP–CpG.

Data access

The raw RRBS data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA797941. Bioinformatic analysis pipelines are available at GitHub (https://github.com/EvoBioWolf/SWALLOW_methylation) and as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank M. Grüebler, M. Liedvogel, H. Lokki, S. Martens, and M. Piha for providing administrative and logistical support to conduct fieldwork in Switzerland, Germany, and Finland. Additionally, we thank M. Grüebler, S. Martens, H. Lokki, M. Piha, M. Briedis, M. Pilati, G. Rüppel, S. Schirmer, and I. Zwahlen for (help with) fieldwork in Switzerland, Germany, and Finland. S.A.M. was funded by the German Research Foundation (DFG; grant MU5038/2-1). This work was further supported by National Science Foundation (NSF) postdoctoral research fellowship grant DBI-1906188 to D.R.S., NSF CAREER grant DEB-1149942 to R.J.S., NSF grant IOS-DEB-1856266 to R.J.S., and LMU Munich (to J.B.W.W.). Computing was performed on the BioHPC hosted at Leibniz Rechenzentrum Munich funded by the German Research Foundation (grant INST 86/2050-1 FUGG to J.B.W.W.). Figures were created in BioRender (https://BioRender.com/27h6b0i).

Author contributions: S.A.M. and J.B.W.W. conceived of the project with input from D.R.S., J.A.C.v.R., and R.J.S. to project planning. R.J.S. and J.A.C.v.R. contributed samples for analysis. S.A.M. analyzed the data, with help from U.S. and D.R.S. and feedback from Z.M.L. S.A.M. and J.B.W.W. prepared the manuscript. All authors participated in discussions on results and edited and gave feedback on the manuscript.

References

Anastasiadi D, Esteve-Codina A, Piferrer F. 2018. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* **11:** 37. doi:10.1186/s13072-018-0205-1

Anderson EC, Thompson EA. 2002. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160:** 1217–1229. doi:10.1093/genetics/160.3.1217

Baird SJE. 2015. Exploring linkage disequilibrium. *Mol Ecol Resour* **15:** 1017–1019. doi:10.1111/1755-0998.12424

Barton NH, de Cara MAR. 2009. The evolution of strong reproductive isolation. *Evolution* **63:** 1171–1190. doi:10.1111/j.1558-5646.2009.00622.x

Billingsley KJ, Meredith M, Daida K, Jerez PA, Negi S, Malik L, Genner RM, Moller A, Zheng X, Gibson SB, et al. 2024. Long-read sequencing of hundreds of diverse brains provides insight into the impact of structural variation on gene expression and DNA methylation. bioRxiv doi:10.1101/2024.12.16.628723

Bolnick DI, Hund AK, Nosil P, Peng F, Ravinet M, Stankowski S, Subramanian S, Wolf JBW, Yukilevich R. 2023. A multivariate view of the speciation continuum. *Evolution* **77:** 318–328. doi:10.1093/evolut/qpac004

Boman J, Qvarnström A, Mugal CF. 2024. Regulatory and evolutionary impact of DNA methylation in two songbird species and their naturally occurring F_1 hybrids. *BMC Biol* **22:** 124. doi:10.1186/s12915-024-01920-2

Brennan RS, deMayo JA, Finiguerra M, Baumann H, Dam HG, Pespeni MH. 2025. Complementary genetic and epigenetic changes facilitate rapid adaptation to multiple global change stressors. *Proc Natl Acad Sci* **122:** e2422782122. doi:10.1073/pnas.2422782122

Brommer JE. 2011. Whither P_{ST}? The approximation of Q_{ST} by P_{ST} in evolutionary and conservation biology. *J Evol Biol* **24:** 1160–1168. doi:10.1111/j.1420-9101.2011.02268.x

Burri R. 2017. Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. *Mol Ecol* **26:** 3853–3856. doi:10.1111/mec.14167

Carja O, MacIsaac JL, Mah SM, Henn BM, Kobor MS, Feldman MW, Fraser HB. 2017. Worldwide patterns of human epigenetic variation. *Nat Ecol Evol* **1:** 1577–1583. doi:10.1038/s41559-017-0299-z

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4:** 7. doi:10.1186/s13742-015-0047-8

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* **23:** 3133–3157. doi:10.1111/mec.12796

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27:** 2156–2158. doi:10.1093/bioinformatics/btr330

Da Silva S, Silva AD. 2018. Pstat: an R package to assess population differentiation in phenotypic traits. *R J* **10:** 447. doi:10.32614/RJ-2018-010

de Carvalho CF, Planidin NP, Slate J, Riesch R, Gompert Z, Nosil P. 2025. Genomic regions exhibiting divergent methylation patterns covary with loci associated with mate choice traits in a stick insect. bioRxiv doi:10.1101/2025.07.30.667717v1

Derks MFL, Schachtschneider KM, Madsen O, Schijlen E, Verhoeven KJF, van Oers K. 2016. Gene and transposable element methylation in great tit (*Parus major*) brain and blood. *BMC Genomics* **17:** 332. doi:10.1186/s12864-016-2653-y

Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci* **14:** 927–930. doi:10.1111/j.1654-1103.2003.tb02228.x

Dobzhansky T. 1937. *Genetics and the origin of species*. Reprint. Columbia University Press, New York.

Dray S, Legendre P, Peres-Neto PR. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Modell* **196:** 483–493. doi:10.1016/j.ecolmodel.2006.02.015

Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, Paolo Casale F, Drewe P, Kahles A, Jean G, Vilhjálmsson B, et al. 2015. DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *eLife* **4:** e05255. doi:10.7554/eLife.05255

Edelman NB, Mallet J. 2021. Prevalence and adaptive impact of introgression. *Annu Rev Genet* **55:** 265–283. doi:10.1146/annurev-genet-021821-020805

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491:** 756–760. doi:10.1038/nature11584

Gompert Z, Buerkle CA. 2010. introgress: a software package for mapping components of isolation in hybrids. *Mol Ecol Resour* **10:** 378–384. doi:10.1111/j.1755-0998.2009.02733.x

Greenspoon PB, Spencer HG, M'Gonigle LK. 2022. Epigenetic induction may speed up or slow down speciation with gene flow. *Evolution* **76:** 1170–1182. doi:10.1111/evo.14494

Gwee CY, Metzler D, Fuchs J, Wolf JBW. 2025. Reconciling gene tree discordance and biogeography in European crows. *Mol Ecol* **34:** e17764. doi:10.1111/mec.17764

Hagmann J, Becker C, Müller J, Stegle O, Meyer RC, Wang G, Schneeberger K, Fitz J, Altmann T, Bergelson J, et al. 2015. Century-scale methylome stability in a recently diverged *Arabidopsis thaliana* lineage. *PLoS Genet* **11:** e1004920. doi:10.1371/journal.pgen.1004920

Han L, Su B, Li W-H, Zhao Z. 2008. CpG island density and its correlations with genomic features in mammalian genomes. *Genome Biol* **9:** R79. doi:10.1186/gb-2008-9-5-r79

Hawe JS, Wilson R, Schmid KT, Zhou L, Lakshmanan LN, Lehne BC, Kühnel B, Scott WR, Wielscher M, Yew YW, et al. 2022. Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function. *Nat Genet* **54:** 18–29. doi:10.1038/s41588-021-00969-x

Hazarika RR, Serra M, Zhang Z, Zhang Y, Schmitz RJ, Johannes F. 2022. Molecular properties of epimutation hotspots. *Nat Plants* **8:** 146–156. doi:10.1038/s41477-021-01086-7

Heckwolf MJ, Meyer BS, Häsler R, Höppner MP, Eizaguirre C, Reusch TBH. 2020. Two different epigenetic information channels in wild three-spined sticklebacks are involved in salinity adaptation. *Sci Adv* **6:** eaaz1138. doi:10.1126/sciadv.aaz1138

Hendry AP. 2004. Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evol Ecol Res* **6:** 1219–1236.

Höglund A, Henriksen R, Fogelholm J, Churcher AM, Guerrero-Bosagna CM, Martinez-Barrio A, Johnsson M, Jensen P, Wright D. 2020. The methylation landscape and its role in domestication and gene regulation in the chicken. *Nat Ecol Evol* **4:** 1713–1724. doi:10.1038/s41559-020-01310-1

Hu J, Wuitchik SJS, Barry TN, Jamniczky HA, Rogers SM, Barrett RDH. 2021. Heritability of DNA methylation in threespine stickleback (*Gasterosteus aculeatus*). *Genetics* **217:** 1–15. doi:10.1093/genetics/iyab001

Johannes F, Schmitz RJ. 2019. Spontaneous epimutations in plants. *New Phytol* **221:** 1253–1259. doi:10.1111/nph.15434

Johnson KM, Kelly MW. 2020. Population epigenetic divergence exceeds genetic divergence in the Eastern oyster *Crassostrea virginica* in the Northern Gulf of Mexico. *Evol Appl* **13:** 945–959. doi:10.1111/eva.12912

Jombart T. 2008. *Adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403–1405. doi:10.1093/bioinformatics/btn129

Korunes KL, Samuk K. 2021. Pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour* **21:** 1359–1368. doi:10.1111/1755-0998.13326

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27:** 1571–1572. doi:10.1093/bioinformatics/btr167

Leinonen T, McCairns RJS, O'Hara RB, Merilä J. 2013. Q_{ST}–F_{ST} comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nat Rev Genet* **14:** 179–190. doi:10.1038/nrg3395

Lewontin RC. 1974. *The genetic basis of evolutionary change*. Columbia University Press, New York.

Lövkvist C, Dodd IB, Sneppen K, Haerter JO. 2016. DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res* **44:** 5123–5132. doi:10.1093/nar/gkw124

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.j* **17:** 10–12. doi:10.14806/ej.17.1.200

Martin-Trujillo A, Patel N, Richter F, Jadhav B, Garg P, Morton SU, McKean DM, DePalma SR, Goldmuntz E, Gruber D, et al. 2020. Rare genetic variation at transcription factor binding sites modulates local DNA methylation profiles. *PLoS Genet* **16:** e1009189. doi:10.1371/journal.pgen.1009189

Mayr E. 1942. *Systematics and the origin of species*. Columbia University Press, New York.

McKay JA, Xie L, Harris S, Wong YK, Ford D, Mathers JC. 2011. Blood as a surrogate marker for tissue-specific DNA methylation and changes due

to folate depletion in post-partum female mice. *Mol Nutr Food Res* **55:** 1026–1035. doi:10.1002/mnfr.201100008
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20:** 1297–1303. doi:10.1101/gr.107524.110
Merondun J, Wolf JBW. 2025. DNA methylation reflects cis-genetic differentiation across the European crow hybrid zone. *Mol Ecol* **34:** e70026. doi:10.1111/mec.70026
Mueller SA, Merondun J, Lečić S, Wolf JBW. 2025. Epigenetic variation in light of population genetic practice. *Nat Commun* **16:** 1028. doi:10.1038/s41467-025-55989-6
Nachman MW, Payseur BA. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos Trans R Soc Lond B Biol Sci* **367:** 409–421. doi:10.1098/rstb.2011.0249
Ord J, Gossmann TI, Adrian-Kalchhauser I. 2023. High nucleotide diversity accompanies differential DNA methylation in naturally diverging populations. *Mol Biol Evol* **40:** msad068. doi:10.1093/molbev/msad068
Planidin NP, de Carvalho CF, Feder J, Gompert Z, House JD, Nosil P. 2025. Adaptive epigenetic divergence can facilitate ecological speciation. *Proc Biol Sci* **292:** 20251217. doi:10.1098/rspb.2025.1217
Platt A, Gugger PF, Pellegrini M, Sork VL. 2015. Genome-wide signature of local adaptation linked to variable CpG methylation in oak populations. *Mol Ecol* **24:** 3823–3830. doi:10.1111/mec.13230
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559–575. doi:10.1086/519795
Quadrana L, Colot V. 2016. Plant transgenerational epigenetics. *Annu Rev Genet* **50:** 467–491. doi:10.1146/annurev-genet-120215-035254
Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, Noor MAF, Mehlig B, Westram AM. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol* **30:** 1450–1477. doi:10.1111/jeb.13047
R Core Team. 2025. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.
Reik W, Dean W, Walter J. 2001. Epigenetic reprogramming in mammalian development. *Science* **293:** 1089–1093. doi:10.1126/science.1063443
Richards C, Pigliucci M. 2021. Epigenetic inheritance: a decade into the extended evolutionary synthesis. *PARADIGMI* **38:** 463–494. doi:10.30460/99624
Richards CL, Alonso C, Becker C, Bossdorf O, Bucher E, Colomé-Tatché M, Durka W, Engelhardt J, Gaspar B, Gogol-Döring A, et al. 2017. Ecological plant epigenetics: evidence from model and non-model species, and the way forward. *Ecol Lett* **20:** 1576–1590. doi:10.1111/ele.12858
Rossi M, Hausmann AE, Alcami P, Moest M, Roussou R, Van Belleghem SM, Wright DS, Kuo C-Y, Lozano-Urrego D, Maulana A, et al. 2024. Adaptive introgression of a visual preference gene. *Science* **383:** 1368–1373. doi:10.1126/science.adj9201
Safran RJ, Scordato ESC, Wilkins MR, Hubbard JK, Jenkins BR, Albrecht T, Flaxman SM, Karaardıç H, Vortman Y, Lotem A, et al. 2016. Genome-wide differentiation in closely related populations: the roles of selection and geographic isolation. *Mol Ecol* **25:** 3865–3883. doi:10.1111/mec.13740
Schield DR, Carter JK, Scordato ESC, Levin II, Wilkins MR, Mueller SA, Gompert Z, Nosil P, Wolf JBW, Safran RJ. 2024. Sexual selection promotes reproductive isolation in barn swallows. *Science* **386:** eadj8766. doi:10.1126/science.adj8766
Scordato ESC, Safran RJ. 2014. Geographic variation in sexual selection and implications for speciation in the Barn Swallow. *Avian Res* **5:** 8. doi:10.1186/s40657-014-0008-4
Scordato ESC, Wilkins MR, Semenov G, Rubtsov AS, Kane NC, Safran RJ. 2017. Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Mol Ecol* **26:** 5676–5691. doi:10.1111/mec.14276
Scordato ESC, Smith CCR, Semenov GA, Liu Y, Wilkins MR, Liang W, Rubtsov A, Sundev G, Koyama K, Turbek SP, et al. 2020. Migratory divides coincide with reproductive barriers across replicated avian hybrid zones above the Tibetan plateau. *Ecol Lett* **23:** 231–241. doi:10.1111/ele.13420
Secomandi S, Gallo GR, Sozzoni M, Iannucci A, Galati E, Abueg L, Balacco J, Caprioli M, Chow W, Ciofi C, et al. 2023. A chromosome-level reference genome and pangenome for barn swallow population genomics. *Cell Rep* **42:** 111992. doi:10.1016/j.celrep.2023.111992
Shafer ABA, Wolf JBW. 2013. Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecol Lett* **16:** 940–950. doi:10.1111/ele.12120
Shang H, Field DL, Paun O, Rendón-Anaya M, Hess J, Vogl C, Liu J, Ingvarsson PK, Lexer C, Leroy T. 2023. Drivers of genomic landscapes of differentiation across a *Populus* divergence gradient. *Mol Ecol* **32:** 4348–4361. doi:10.1111/mec.17034
Silliman K, Spencer LH, White SJ, Roberts SB. 2023. Epigenetic and genetic population structure is coupled in a marine invertebrate. *Genome Biol Evol* **15:** evad013. doi:10.1093/gbe/evad013
Slatkin M. 1985. Gene flow in natural populations. *Annu Rev Ecol Syst* **16:** 393–430. doi:10.1146/annurev.es.16.110185.002141
Taudt A, Colomé-Tatché M, Johannes F. 2016. Genetic sources of population epigenomic variation. *Nat Rev Genet* **17:** 319–332. doi:10.1038/nrg.2016.45
Tavares H, Whibley A, Field DL, Bradley D, Couchman M, Copsey L, Elleouet J, Burrus M, Andalo C, Li M, et al. 2018. Selection and gene flow shape genomic islands that control floral guides. *Proc Natl Acad Sci* **115:** 11006–11011. doi:10.1073/pnas.1801832115
Turbek SP, Schield DR, Scordato ESC, Contina A, Da X-W, Liu Y, Liu Y, Pagani-Núñez E, Ren Q-M, Smith CCR, et al. 2022. A migratory divide spanning two continents is associated with genomic and ecological divergence. *Evolution* **76:** 722–736. doi:10.1111/evo.14448
Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43:** 11.10.1–11.10.33. doi:10.1002/0471250953.bi1110s43
van Moorsel SJ, Schmid MW, Wagemaker NCAM, van Gurp T, Schmid B, Vergeer P. 2019. Evidence for rapid evolution in a grassland biodiversity experiment. *Mol Ecol* **28:** 4097–4117. doi:10.1111/mec.15191
Venney CJ, Wellband KW, Heath DD. 2021. Rearing environment affects the genetic architecture and plasticity of DNA methylation in Chinook salmon. *Heredity (Edinb)* **126:** 38–49. doi:10.1038/s41437-020-0346-4
Vernaz G, Hudson AG, Santos ME, Fischer B, Carruthers M, Shechonge AH, Gabagambi NP, Tyers AM, Ngatunga BP, Malinsky M, et al. 2022. Epigenetic divergence during early stages of speciation in an African crater lake cichlid fish. *Nat Ecol Evol* **6:** 1940–1951. doi:10.1038/s41559-022-01894-w
Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, Wolf JBW. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat Commun* **7:** 13195. doi:10.1038/ncomms13195
Villicaña S, Bell JT. 2021. Genetic impacts on DNA methylation: research findings and future perspectives. *Genome Biol* **22:** 127. doi:10.1186/s13059-021-02347-6
Vogt G. 2017. Facilitation of environmental adaptation and evolution by epigenetic phenotype variation: insights from clonal, invasive, polyploid, and domesticated animals. *Environ Epigenet* **3:** dvx002. doi:10.1093/eep/dvx002
Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. *Nature* **430:** 85–88. doi:10.1038/nature02698
Wolf JBW, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet* **18:** 87–100. doi:10.1038/nrg.2016.133
Wringe BF, Stanley RRE, Jeffery NW, Anderson EC, Bradbury IR. 2017. *Parallelnewhybrid*: an R package for the parallelization of hybrid detection using newhybrids. *Mol Ecol Resour* **17:** 91–95. doi:10.1111/1755-0998.12597
Wu C-I. 2001. The genic view of the process of speciation. *J Evol Biol* **14:** 851–865. doi:10.1046/j.1420-9101.2001.00335.x
Zhang Y, Yang L, Kucherlapati M, Hadjipanayis A, Pantazi A, Bristow CA, Lee EA, Mahadeshwar HS, Tang J, Zhang J, et al. 2019. Global impact of somatic structural variation on the DNA methylome of human cancers. *Genome Biol* **20:** 209. doi:10.1186/s13059-019-1818-9
Zink RM, Pavlova A, Rohwer S, Drovetski SV. 2006. Barn swallows before barns: population histories and intercontinental colonization. *Proc Biol Sci* **273:** 1245–1251. doi:10.1098/rspb.2005.3414

Received September 25, 2025; accepted in revised form June 10, 2026.