

*Genetics and population analysis***AdmixLD: Fast genome-scale inference of ancestry disequilibrium in hybrid zones**Yannick Z. Francioli^{1*}, Richard H. Adams², Kaas Ballard¹, Zachariah Gompert³, and Todd A. Castoe^{1*}.¹Department of Biology, University of Texas at Arlington, Arlington, TX²Department of Entomology and Plant Pathology, University of Arkansas, Fayetteville, AR³Department of Biology, Utah State University, Logan, UT USA

*To whom correspondence should be addressed.

Associate Editor: Russell Schwartz

Received on 15 May 2026; revised on 10 July 2026; accepted on 18 August 2026

Abstract

Summary: Hybrid zones represent powerful natural systems for studying reproductive isolation and speciation. One key genomic signature of genetic incompatibilities and epistatic interactions is linkage disequilibrium (LD) — non-random associations between loci from different lineage backgrounds, generated by selection against maladaptive allele combinations. However, admixture alone induces strong genome-wide LD in hybrid populations, obscuring selection-driven signals. Here, we present AdmixLD, a fast, scalable C++ tool for genome-wide LD scanning in hybrid zones that estimates LD using partial correlation to control for individual hybrid index. By removing admixture-driven covariance, AdmixLD enhances detection of locus-specific associations and enables genome-scale identification of candidate barrier loci and interacting genomic regions.

Availability: The software and its code source are available at <https://github.com/yzfranci/AdmixLD>, and scripts for the data analysis are available at <https://github.com/yzfranci/AdmixLDAnalysis>.

Contact: francioli.yannick@uta.edu (Y.F.), todd.castoe@uta.edu (T.C.)

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Natural hybrid zones between divergent populations provide valuable systems to study mechanisms of reproductive isolation and speciation. Hybrid zones, particularly those formed via secondary contact between previously isolated populations, have long been described as natural laboratories in which the consequences of gene flow, selection, and genetic incompatibilities can be observed directly (Barton and Hewitt 1985, Hewitt 1988, Harrison 1993, Abbott et al. 2013). When divergent genomes come into contact, hybridization disrupts coadapted gene complexes, separating alleles that evolved in different genetic backgrounds and generating incompatibilities that reduce hybrid fitness. Multi-locus interactions are therefore expected to play a central role in the evolution and maintenance of reproductive isolation between diverging lineages. A key genomic signature of such incompatibilities is the generation of non-random associations among loci — linkage disequilibrium (LD) — as selection acts against maladaptive allele

combinations in hybrid individuals (Szymura and Barton 1991, Nürnberger et al. 1995, Payseur 2010, Pool 2015). Such incompatibilities are recognized as pervasive features of natural hybrid zones, and their polygenic nature is consistent with reproductive isolation being driven by interactions across many loci (Arnegard et al. 2014, Schumer et al. 2014, Christie and Strauss 2018, Nikolakis et al. 2022).

When ongoing gene flow from parental populations generates spatially structured ancestry across a transect, mixture LD (Falush, Stephens, and Pritchard 2003) arises purely from covariance in shared ancestry proportions across individuals, independent of any selection against incompatible combinations (Barton and Hewitt 1985, Pfaff et al. 2001). This represents implications for sampling itself: because mixture LD scales with variation in admixture proportions across sampled individuals, overrepresentation of parental-type individuals from zone margins will further inflate inter-locus correlations. Admixture LD arises from a related but distinct process: in recently admixed genomes, recombination has not yet had time to break down long ancestry tracts inherited from parental

populations, generating correlations among physically distant loci that reflect shared ancestry tracts rather than selection (Falush, Stephens, and Pritchard 2003). Together, mixture LD and admixture LD form a pervasive background of ancestry-driven LD in hybrid populations that is unrelated to selection. LD in excess of this background is exceptionally informative for identifying barrier loci and genomic regions that impede gene flow, and for dissecting the genetic architecture underlying reproductive isolation (Gavrilets 2004, Nosil, Funk, and Ortiz-Barrientos 2009, Flaxman et al. 2014, Lindtke and Buerkle 2015, Ravinet et al. 2017). Accounting for individual genome-wide admixture and focusing on the correlation between residuals – excess ancestry LD – is therefore crucial to distinguish non-random associations among loci driven by selection from those generated by mixture LD alone, particularly in heavily structured systems such as hybrid zone transects. The statistical foundations for estimating LD in structured populations have been established by several key methodological contributions (Mackay and Powell 2007, Lin, Xing, and Xing 2012, Mangin et al. 2012), and Schumer and Brandvain (Schumer and Brandvain 2016) extended this framework explicitly to hybrid zones, providing a theoretical basis for separating mixture-driven from selection-driven associations in admixed genomes. Despite these advances, a practical gap remains: while fast, genome-scale tools for estimating LD in structured populations exist (Bercovich et al. 2025), they are designed for LD pruning rather than the detection of outlier loci, and are not tailored specifically for hybrid zone analyses. To address this, we present AdmixLD: a dedicated, scalable C++ software for genome-wide intra- and interchromosomal excess ancestry LD scans in hybrid zones that explicitly accounts for individual admixture proportions, facilitating the identification of selection-driven LD.

2 Methods

AdmixLD is implemented in C++ (v. 17). The tool accepts local ancestry tracts or SNPs as ancestry markers encoded in standard VCF/BCF files through HTSlib (Bonfield et al., 2021), as well as standard most likely sub-population (MSP) files from RFMIX (Maples et al., 2013). AdmixLD first computes a per-sample hybrid index (HI) – the genome-wide proportion of ancestry. When local ancestry tracts are provided, HI is estimated using a tract-length-weighted estimator: $HI_i = \sum_j d_{ij} \cdot L_j / (2 \cdot \sum_j L_j)$, where d_{ij} is the ancestry dosage (0 or 1 for phased haplotypes, 0, 1, or 2 for diploids) for individual i at tract j and L_j is the inferred tract length. When a VCF encodes SNPs rather than ancestry tracts, two options are available. If the markers are assumed ancestry-diagnostic (parental allele frequencies $p_1 \approx 1$, $p_2 \approx 0$), a non-weighted HI can be computed directly from dosages. Alternatively, if the variants are not fixed between the two parental populations, an allele frequencies file can be supplied. In this case, each marker is first polarized so that the ancestry-1 allele is the higher-frequency allele in parental population 1 ($p_1 \geq p_2$); for any marker where $p_2 > p_1$ the dosage is flipped, $d'_{ij} = d_{ij} - 1$ (with p_1 and p_2 swapped accordingly), where $d_{max} = 2$ for diploid and 1 for phased haplotype data. HI is then estimated using a δ^2 -weighted method-of-moments estimator (Buerkle 2005): $HI_i = [\sum_j (x_{ij}/2 - p_{2j}) \cdot \delta_j] / \sum_j \delta_j^2$ where $\delta_j = p_{1j} - p_{2j}$ is the allele frequency difference between parental populations at marker j . A leave-focus-chromosomes-out HI computation mode is also available to avoid confounding the focal chromosome's signal with its own contribution to the global HI estimate (Schumer and Brandvain, 2016). Alternatively, a file specifying hybrid indices can be provided directly by the user.

Each marker's dosage vector is residualized against the genome-wide ancestry and z-scored, isolating locus-specific ancestry variation from the background admixture signal common to all loci. For the j th marker with

dosage vector d_j across n individuals, residuals are computed using the expected dosage under Hardy–Weinberg equilibrium in an admixed population. When parental allele frequencies are not provided (markers assumed ancestry-diagnostic, $p_1 \approx 1$, $p_2 \approx 0$), residuals are computed as:

$$\epsilon_{ij} = d_{ij} - 2HI_i \text{ (diploid)} \quad \epsilon_{ij} = d_{ij} - HI_i \text{ (phased)}$$

When parental allele frequencies are provided, residuals incorporate the marker-specific allele-frequency contrast $\delta_j = p_{1j} - p_{2j} (\geq 0 \text{ after polarization})$:

$$\epsilon_{ij} = d_{ij} - 2(p_{2j} + \delta_j \cdot HI_i) \text{ (diploid)} \quad \epsilon_{ij} = d_{ij} - (p_{2j} + \delta_j \cdot HI_i) \text{ (phased)}$$

In both cases the residuals are mean-centered and standardized across samples, $Z_{ij} = (\epsilon_{ij} - \text{mean}(\epsilon_j)) / \text{SD}(\epsilon_j)$ – where $\text{mean}(\epsilon_j)$ and $\text{SD}(\epsilon_j)$ are computed across samples i for each marker j – and markers with zero residual standard deviation are excluded as uninformative. Excess ancestry LD is then quantified as the Pearson correlation between all pairwise standardized residual vectors.

AdmixLD supports both interchromosomal pairs (the default) and intrachromosomal pairs, the latter optionally filtered for maximum genomic distance threshold. To avoid computing and storing correlations across all marker pairs — which can produce extremely large output files at genome scale — AdmixLD estimates the empirical distribution of partial correlation values using reservoir sampling (default capacity: 200,000 values), which provides a memory-efficient approximation of the full distribution. This empirical distribution can then be used to inform the choice of an outlier detection threshold.

Additionally, we implemented a false discovery rate (FDR) framework based on an empirical null distribution (Efron 2004, 2012), which assumes that most tested pairs are neutral and only a minority reflect true signal. We calibrate a separate null distribution for each chromosome-pair block (interchromosomal) or distance bin (intrachromosomal). Within a block, all pairwise correlations are Fisher z -transformed ($z = \text{atanh}(r)$) and modeled as Gaussian, $N(\mu_0, \sigma_0^2)$, with $\mu_0 = \text{median}(z)$ and $\sigma_0 = 1.4826 \times \text{median}(|z - \text{median}(z)|)$, a robust estimator of the null distribution's spread that limits the influence of true signal in the tails. Each z -score is then recalibrated against its block-specific null distribution ($z^* = (z - \mu_0)/\sigma_0$) and converted to a one-sided p -value per tail, since the empirical null distribution is not guaranteed to be symmetric about zero. We then apply Storey's q -value procedure (Storey 2002, Storey and Tibshirani 2003) per tail to control the FDR at a user-specified threshold. To bound memory use, null distribution calibration draws on a reservoir sample of z -scores per block (200,000 by default) rather than the full pairwise set. For intrachromosomal scans, a minimal distance can be set for marker pairs and the null distribution calibration is done separately within log-spaced distance bins, as physical linkage inflates LD independent of admixture. The same framework was applied on traditional LD output using a custom Python script.

We demonstrated the application of AdmixLD using simulated hybrid zone data, which were generated using SLiM version 5.2 (Haller and Messer 2017). Each simulation modelled a 30-deme stepping-stone hybrid zone evolving under a Wright-Fisher process, with deme sizes of 100 individuals and an inter-deme migration rate of 0.1 per generation. Individual genomes consisted of three one-Morgan chromosomes, with 1,000 fixed ancestry markers distributed at equal intervals, and simulations were run for 500 generations. For each replicate, two VCF files of ancestry markers were output to evaluate the effect of sampling scheme: one containing 200 randomly sampled hybrids (balanced sampling), and one with unbalanced sampling comprising 50 individuals with hybrid index (HI) < 0.05 , 100 with $0.05 \leq HI \leq 0.95$, and 50 with $HI > 0.95$.

We simulated several incompatibility architectures. First, a neutral simulation without any selected loci. We then modelled classical two-

AdmixLD

locus Dobzhansky–Muller incompatibility scenarios (DMI) involving different numbers of loci pairs (1, 2, 3 and 6) with individual fitness defined as $w = 1 - s \times |d_1 - d_2| / 2$, where d_1 and d_2 are ancestry dosages at each paired locus (0, 1, or 2 alleles derived from population 2; loci are fully differentiated between parental populations) and $s = 0.5$. We also simulated a single-DMI scenario under weaker selection ($s = 0.05$). We modeled coadaptation incompatibility involving six fully differentiated loci (two per chromosome), where fitness was penalized by variance in dosage across loci: $w = 1 - s \times \text{Var}(d)$, where d is the vector of per-locus dosages and $s = 0.5$. Finally, to test the effect of non-fixed allele-frequency variance between populations, we repeated the six-locus coadaptation scenario, but with 1,000 markers spanning a range of allele frequencies. Each simulation type was run for 100 replicates.

Excess Ancestry LD was computed with AdmixLD on VCF files from SLiM, and benchmarked against traditional LD estimated with `vcftools --interchrom-geno-r2` (Danecek et al. 2011). Although `vcftools` is not designed to account for admixture structure, it provides a baseline for illustrating the magnitude of bias introduced by uncorrected ancestry covariance. For each simulation, the coefficient of determination (R^2) distribution was approximated by randomly sampling 200,000 marker pairs. To estimate overall R^2 distributions in each simulation replicate, we used 200,000 randomly sampled marker pair R^2 values. The resulting background distributions were then used to compute z-scores for selected loci R^2 for every simulation replicate, quantifying their deviation from the genomic background. Distributions for visualization purposes (violin plots) were obtained by pooling 200,000 randomly sampled marker pair R^2 values across five replicates per simulation type.

To assess signal enrichment at selected loci, we computed the median of the pooled true selected locus pair R^2 distribution across simulation replicates, along with their background z-scores (Supp. Table S2), and the 5th percentile to provide a conservative lower-bound estimate. Sensitivity and specificity of the FDR-based outlier detection were assessed using the true positive rate as the number of true hits among selected pairs divided by the total number of selected pairs across all replicates, and the false positive rate as the number of false hits divided by the total number of non-selected loci pairs, excluding pairs where both markers fall within 2 cM of a selected locus to avoid confounding effects from indirect selection. Additionally, we computed a region hit rate: the proportion of marker pairs with both markers within 2 cM of a selected locus that were called significant, used to quantify enrichment of hits near true loci relative to the background rate.

We also demonstrated an empirical application of AdmixLD and compared its performance against traditional LD estimation using published ancestry track data from a rattlesnake hybrid zone (Francioli et al. 2025). Both AdmixLD and `vcftools --interchrom-geno-r2/--geno-r2` were applied to compute Interchromosomal excess ancestry LD between local ancestry tracts of chromosomes 6 and 7, while intrachromosomal excess ancestry LD was computed for chromosome 3.

3 Results

Across all simulated scenarios, AdmixLD's background R^2 remained low and stable, ranging from 0.0126 (neutral) to 0.0275 (six-pair multi-DMI; Fig. 1A, Supp. Table S1). By contrast, traditional LD's background R^2 scaled with the extent of genome-wide selection, from 0.399 (neutral) to 0.836 (six-pair multi-DMI). These results highlight how the signal of admixture-driven and selection-driven associations are entangled when using traditional LD. In the single-DMI scenario with strong selection ($s = 0.5$), selected loci pairs showed extreme R^2 enrichment relative to

background, expressed as a z-score (deviation from the background R^2 distribution) of 26.53, far exceeding traditional LD ($z = 3.55$). As expected, reducing selection strength lowered enrichment: at $s = 0.05$, AdmixLD's median z-score dropped to 7.40, compared to 2.91 for traditional LD (Fig. 1A, Supp. Table S1).

Across all simulations, AdmixLD's z-scores for selected loci pairs remained well above traditional LD's, with one exception: the six-pair multi-DMI scenario under extreme genome-wide selection (Fig. 1A, Supp. Table S1). Enrichment declined as more independent DMI pairs were placed under selection: median z-scores fell from 26.53 (single strong DMI) to 15.11 (two pairs) to 8.81 (three pairs). This decline coincided with a rise in AdmixLD's own background R^2 (0.0149 to 0.0275), consistent with genome-wide selection increasingly affecting the background against which true selected loci pairs are compared. In the six-pair multi-DMI scenario, AdmixLD's median z-score dropped to 0.09 (5th percentile $z = -0.275$; vs. 1.04 and 0.931 for traditional LD), indicating that at this level of genome-wide selection, background and true-pair R^2 distributions were no longer distinguishable. Even here, however, a subset of selected pairs still showed extreme R^2 values under AdmixLD – highlighting a distinction that was obscured under traditional LD due to its far more inflated background R^2 (0.83; Fig. 1A). Notably, while detection power decreases as genome-wide selection increases (i.e., in the three-pair multi-DMI scenario above), the number of loci alone does not erode the signal: the pathway scenario, which involves six selected loci but penalizes fitness through variance in dosage (maximum $s = 0.5$), retained stronger enrichment (median $z = 13.4$, 5th percentile $z = 8.85$). The scenario with variable allele frequency markers showed an increase in the median z-score for both AdmixLD (13.4 to 16.2) and traditional LD (2.80 to 3.28).

Under unbalanced sampling, traditional LD's background R^2 increased substantially across all scenarios (e.g., 0.399 to 0.567 in the neutral simulation; Fig. 1A, Supp. Fig. S1, Supp. Table S1), while AdmixLD's background R^2 rose only marginally (0.0126 to 0.0162). This asymmetry across methods carried through to enrichment: AdmixLD's z-scores at selected loci pairs dropped only slightly relative to balanced sampling (e.g., 26.53 to 23.65 in single strong DMI) and remained far above that of traditional LD, which declined further under imbalance (e.g., 3.55 to 2.83).

At $\text{FDR} < 0.01$, the FDR-based outlier detection showed low false positive rates across all scenarios, with background hit rates ranging from 0.00008 (neutral) to 0.0057 (six-pair multi-DMI), broadly mirroring the slight increase in background R^2 with architecture complexity (Supp. Table S2). True positive rates (TPR) closely tracked the enrichment patterns described above: TPR was 1.0 for single-pair strong-DMI and two-pair multi-DMI, declined to 0.93 at three pairs, and collapsed to 0.12 at six pairs. In multi-DMI scenarios, cross pair hit rates – false positives among pairs combining loci from different true DMI pairs – were high and close to TPR, likely reflecting a coupling effect between DMI, and highlighting the challenge of separating distinct barrier loci. Detection power was sensitive to selection strength, as TPR dropped to 0.3 under weaker selection ($s = 0.05$). In contrast, both pathway architectures retained high TPR (0.89 with fixed markers, 0.998 with variable allele frequencies) despite also involving six loci, indicating that detection power depends on how selection is distributed among loci rather than based on selected locus count alone. AdmixLD outlier density concentrates specifically around selected loci (within 0.2cM: 10–1059× enriched over background, lowest at six-pair multi-DMI and highest for single-strong-DMI; Fig 1B, Supp.

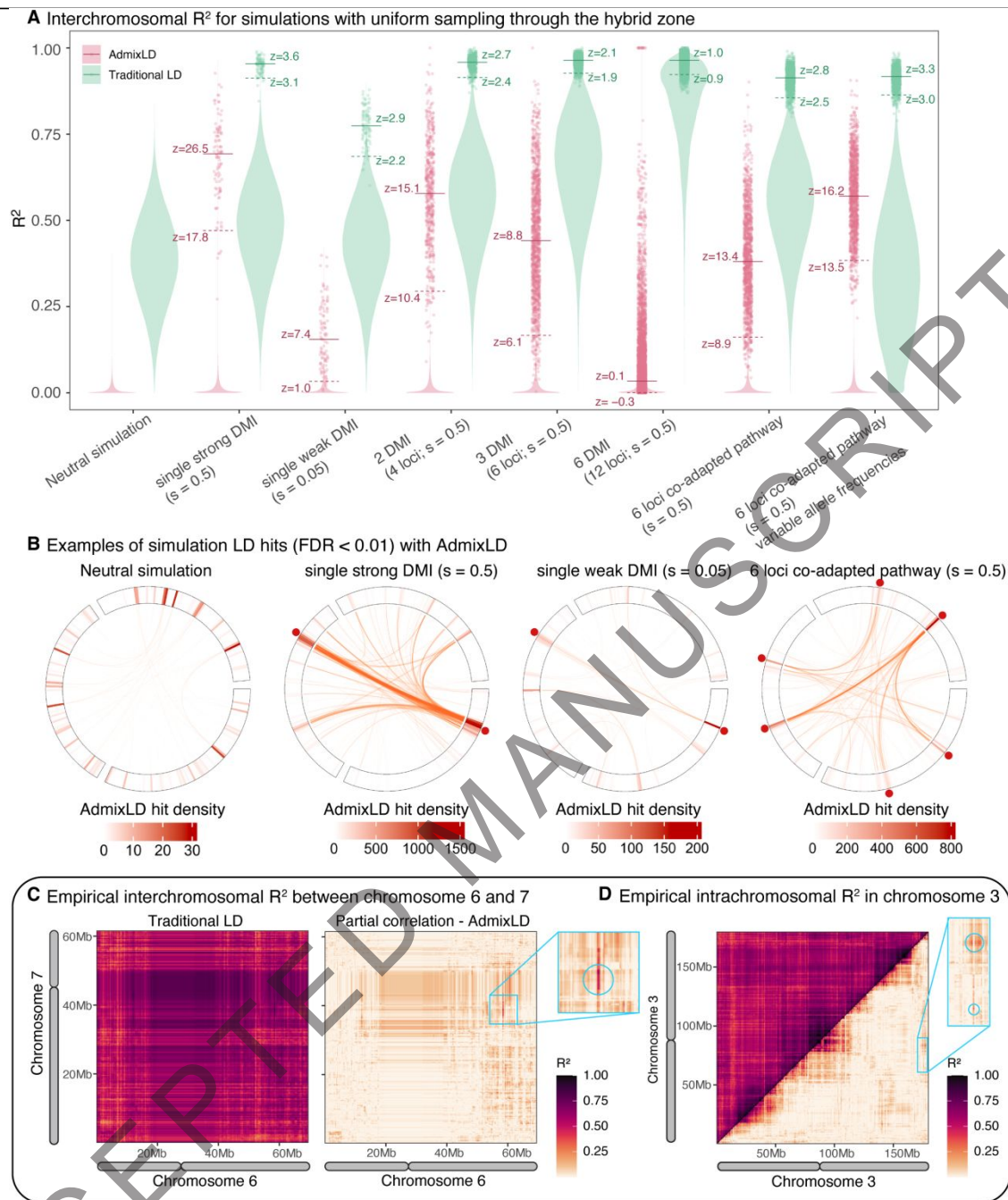


Figure 1. Comparison of excess ancestry LD (partial correlation) and traditional LD across different simulation scenarios, and on empirical data. (A) Overall R^2 distributions compared to selected loci pairs. Shaded distributions show background R^2 values across all locus pairs; colored points indicate R^2 values for incompatible locus pairs. Pink: excess ancestry LD (AdmixLD); green: traditional LD. Dashed and solid horizontal lines indicate the 5th percentile and median z-scores of incompatible locus pairs, respectively. (B) Structure of identified outlier pairs (using AdmixLD implemented FDR method) across simulated genomes, shown for single simulation replicates. Orange lines represent LD between loci pairs identified as outliers (FDR < 0.01), red dots show the location of loci under selection, and the outer track show the density of outlier hit (See Supplementary Figure 2 for all simulation scenarios). (C) Empirical interchromosomal and intrachromosomal LD, between chromosomes 6 and 7 in a rattlesnake hybrid zone, estimated using excess ancestry LD (right) and traditional LD (left). (D) Empirical intrachromosomal for chromosome 3 of the same hybrid zone, displayed as a triangular heatmap with excess ancestry LD in the upper-left triangle and traditional LD in the lower-right triangle. Blue rectangles zoom in regions with identified outliers, and blue circles show the location of specific outlier region pairs (FDR < 0.01 for panel C and FDR < 0.05 for panel D).

Fig. S2, Supp. Table S2), scaling with selection strength and structure, in contrast with the low-density background under neutrality. Notably, the enrichment dropped in the pathway case with variable allele frequency, likely due to lower correlation between the selected loci and surrounding loci, as well as a higher false positive rate. Applying the same FDR

framework to traditional LD yielded no outliers in any scenario, a direct consequence of its inflated background R^2 , which prevents selected loci pairs from being statistically distinguished from the background distribution regardless of true enrichment.

AdmixLD

AdmixLD took 3 minutes and 25 seconds single-threaded (Intel Core i7-3770 3.40GHz) with the default tile size (1,024 columns per tile; see Methods) to compute partial correlations between loci of the empirical dataset, including 580 million ancestry tracts pairs, demonstrating efficient genome-scale computation. In our empirical application, interchromosomal excess ancestry LD computed with AdmixLD revealed a prominent region of elevated R^2 between chromosomes 6 and 7, against a near-zero background (Figure 1C), highlighting strong localized signals. Using the implemented FDR framework, this region pair (~58 Mb on chromosome 6 and ~39Mb on chromosome 7) was the only significant outlier at $FDR < 0.01$ for this chromosome pair. Traditional LD estimated with vcftools, by contrast, showed broadly elevated R^2 across nearly the entire chromosome pair, obscuring discrete signals (Fig. 1C). Intrachromosomal excess ancestry LD (showed on chromosome 3; Fig. 1D) exhibited a clear distance-decay pattern, with R^2 highest among physically proximate loci and declining sharply with genomic distance, consistent with recombination breaking down ancestry associations. This pattern was largely absent in traditional LD estimates due to a uniformly elevated background. While no outliers were found at $FDR < 0.01$ on chromosome 3, two region pairs were significant at $FDR < 0.05$: ~175 Mb with ~84 Mb, and ~175 Mb with ~66 Mb (Fig. 1D).

4 Discussion

AdmixLD provides a computationally efficient approach to detect excess ancestry LD in hybrid zones, where admixture alone generates strong LD signals that challenge traditional methods (Barton and Hewitt 1985, Pfaff et al. 2001). By explicitly accounting for genome-wide admixture and using partial correlation (Lin, Xing, and Xing 2012, Schumer and Brandvain 2016), AdmixLD isolates LD in excess of the mixture LD background, enriching the signal for selection-driven LD regions and facilitating reliable identification of outlier locus pairs and genomic regions. AdmixLD also implements an outlier detection method based on an empirical null distribution (Efron 2004, 2012) and Storey's q-value FDR procedure (Storey 2002, Storey and Tibshirani 2003), which our simulations show performs well across a range of incompatibility architectures with low false positive rates. This built-in framework is intended as a convenient default rather than the only valid approach, and AdmixLD's outputs are provided in standard formats, allowing users to apply alternative FDR or outlier-detection methods. AdmixLD combined with this outlier detection method performed well across simulation scenarios (strong and weak DMIs, multi-locus coadapted pathways), achieving high true positive rates while keeping background false positive rates very low. It should be noted, however, that given the large number of background comparisons genome-wide, the absolute number of false positives can still be substantial even at a low per-comparison rate. Structural variation, alleles under local adaptation and assortative mating can also generate confounding LD signals unrelated to epistasis. Accordingly, AdmixLD's outlier detection would ideally be combined with complementary evidence. Enrichment of outlier density in the vicinity of candidate incompatibility loci, consistency of LD outliers across independent hybrid zone transects of the same species pair, and alternative statistics such as the decay of heterozygosity variance (Li, Schumer, and Bank 2022) could be used to corroborate inferences. Finally, strong genome-wide selection can itself elevate background LD (Feder et al. 2014), as selection acting across many loci simultaneously increasingly perturbs the ancestry structure against which any single locus

pair is compared, reducing the method's power to distinguish true incompatibilities from background. Controlling for mixture LD confers robustness under imbalanced sampling, which is crucial since hybrid zones are rarely sampled in a uniform way with equal representation across hybrid indices. By using tiled matrix multiplication, combined with OpenMP parallelization across chromosome pairs – or chromosomes for intrachromosomal inferences – AdmixLD provides efficient computation, enabling genome-scale analyses of large marker sets. In conclusion, AdmixLD fills an important methodological gap in hybridization and reproductive isolation research, which increasingly relies on whole genome datasets (Payseur and Rieseberg 2016, Moran et al. 2021). By explicitly correcting for genome-wide admixture, AdmixLD increases power to detect selection-driven excess ancestry LD compared to traditional LD measures, enabling more accurate identification of the genetic architecture of interacting loci and genomic regions under selection in hybrid zones.

Author contributions

Yannick Z. Francioli (Conceptualization, Software, Formal analysis, Visualization, Writing—original draft, Writing—review & editing), Rich H. Adams (Writing—review & editing), Kaas Ballard (Writing—review & editing), Zachariah Gompert (Conceptualization, Writing—review & editing), Todd A. Castoe (Conceptualization, Supervision, Writing—review & editing).

Conflict of interest

None declared.

Acknowledgements

This work was supported by NSF Grants DEB-1655571 (to TAC) and IOS-2307044 (to TAC and RHA).

Data availability

AdmixLD is available for download at <https://github.com/yzfranci/AdmixLD>, and it can be installed through bioconda (<https://anaconda.org/bioconda/admixld>). AdmixLD is also archived on Zenodo (<https://doi.org/10.5281/zenodo.21298533>). The scripts for the figures and analysis are available at <https://github.com/yzfranci/AdmixLDanalysis>. The empirical data used is available on Dryad at <https://doi.org/10.5061/dryad.kh18932k9>.

References

Abbott R, Albach D, Ansell S et al. Hybridization and speciation. *Journal of Evolutionary Biology* 2013;26(2):229–46.
Arnegard ME, McGee MD, Matthews B et al. Genetics of ecological divergence during speciation. *Nature* 2014;511(7509):307–11.
Barton NH, Hewitt GM. Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics* 1985;16:113–48.
Bercovich U, Rasmussen MS, Li Z et al. Measuring linkage disequilibrium and improvement of pruning and clumping in structured populations. *Genetics* 2025;229(3):iyaf009.
Buerkle CA. Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes* 2005;5(3):684–7.
Christie K, Strauss SY. Along the speciation continuum: Quantifying intrinsic and extrinsic isolating barriers across five million years of evolutionary divergence in California jewelflowers. *Evol* 2018;72(5):1063–79.
Danecek P, Auton A, Abecasis G et al. The variant call format and VCFtools. *Bioinformatics* 2011;27(15):2156–8.
Efron B. Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *Journal of the American Statistical Association* 2004;99(465):96–104.

- Efron B. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction, vol. 1. n.p.: Cambridge University Press, 2012.
- Falush D, Stephens M, Pritchard JK. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* 2003;164(4):1567–87.
- Feder JL, Nosil P, Wacholder AC et al. Genome-Wide Congealing and Rapid Transitions across the Speciation Continuum during Speciation with Gene Flow. *Journal of Heredity* 2014;105(S1):810–20.
- Flaxman SM, Wacholder AC, Feder JL et al. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology* 2014;23(16):art. 16. 24724861.
- Francioli YZ, Bernstein JM, Adams RH et al. Estimation of genome-wide coupling in rattlesnake hybrids provides insight into the process of speciation and its progress. *Nat Commun* 2025;16(1):10242.
- Gavrilets S. *Fitness Landscapes and the Origin of Species*. n.p.: Princeton University Press, 2004.
- Haller BC, Messer PW. SLiM 2: flexible, interactive forward genetic simulations. *Molecular Biology and Evolution* 2017;34(1):230–40.
- Harrison RG. Hybrids and hybrid zones, a historical perspective. In: Harrison RG (ed.), *Hybrid Zones and the Evolutionary Process*. Oxford: Oxford University Press, 1993, 3–12.
- Hewitt GM. Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology & Evolution* 1988;3(7):art. 7.
- Li J, Schumer M, Bank C. Imbalanced segregation of recombinant haplotypes in hybrid populations reveals inter- and intrachromosomal Dobzhansky-Muller incompatibilities. Preprint, bioRxiv, 14 Feb. 2022, 2021.08.04.454891.
- Lin CY, Xing G, Xing C. Measuring linkage disequilibrium by the partial correlation coefficient. *Heredity* 2012;109(6):401–2.
- Lindtke D, Buerkle CA. The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution* 2015;69(8):1987–2004.
- Mackay I, Powell W. Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science* 2007;12(2):57–63.
- Mangin B, Siberchicot A, Nicolas S et al. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 2012;108(3):285–91.
- Moran BM, Payne C, Langdon Q et al. The genomic consequences of hybridization. *eLife* 2021;10:e69016.
- Nikolakis ZL, Schield DR, Westfall AK et al. Evidence that genomic incompatibilities and other multilocus processes impact hybrid fitness in a rattlesnake hybrid zone. *Evol* 2022;76(11):2513–30.
- Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* 2009;18(3):art. 3.
- Nürnberg B, Barton N, MacCallum C et al. Natural Selection on Quantitative Traits in the *Bombina* Hybrid Zone. *Evolution* 1995;49(6):1224–38.
- Payseur BA. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources* 2010;10(5):art. 5. 21565092.
- Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. *Molecular Ecology* 2016;25(11):art. 11. 26836441.
- Pfaff CL, Parra EJ, Bonilla C et al. Population Structure in Admixed Populations: Effect of Admixture Dynamics on the Pattern of Linkage Disequilibrium. *The American Journal of Human Genetics* 2001;68(1):198–207.
- Pool JE. The Mosaic Ancestry of the *Drosophila* Genetic Reference Panel and the *D. melanogaster* Reference Genome Reveals a Network of Epistatic Fitness Interactions. *Mol Biol Evol* 2015;32(12):3236–51.
- Ravinet M, Faria R, Butlin RK et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol* 2017;30(8):1450–77.
- Schumer M, Brandvain Y. Determining epistatic selection in admixed populations. *Mol Ecol* 2016;25(11):2577–91.
- Schumer M, Cui R, Powell DL et al. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *eLife* 2014;3:e02535.
- Storey JD. A Direct Approach to False Discovery Rates. *J R Stat Soc Ser B Stat Methodol* 2002;64(3):479–98.
- Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 2003;100(16):9440–5.
- Szymura JM, Barton NH. The Genetic Structure of the Hybrid Zone Between the Fire-Bellied Toads *Bombina Bombina* and *B. Variegata*: Comparisons Between Transects and Between Loci. *Evolution* 1991;45(2):237–61.

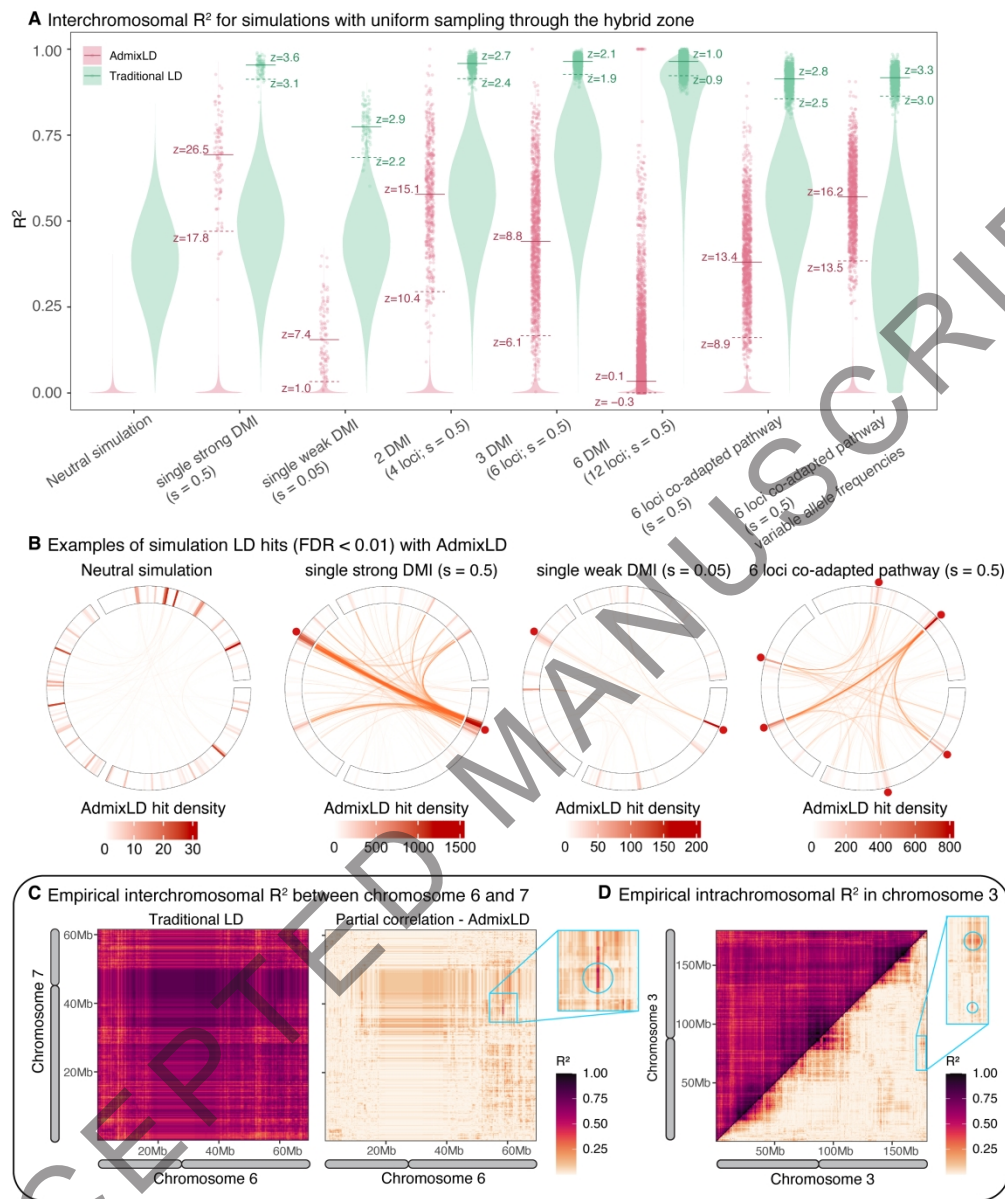


Figure 1. Comparison of excess ancestry LD (partial correlation) and traditional LD across different simulation scenarios, and on empirical data. (A) Overall R^2 distributions compared to selected loci pairs. Shaded distributions show background R^2 values across all locus pairs; colored points indicate R^2 values for incompatible locus pairs. Pink: excess ancestry LD (AdmixLD); green: traditional LD. Dashed and solid horizontal lines indicate the 5th percentile and median z-scores of incompatible locus pairs, respectively. (B) Structure of identified outlier pairs (using AdmixLD implemented FDR method) across simulated genomes, shown for single simulation replicates. Orange lines represent LD between loci pairs identified as outliers (FDR < 0.01), red dots show the location of loci under selection, and the outer track show the density of outlier hit (See Supplementary Figure 2 for all simulation scenarios). (C) Empirical interchromosomal and intrachromosomal LD. between chromosomes 6 and 7 in a rattlesnake hybrid zone, estimated using excess ancestry LD (right) and traditional LD (left). (D) Empirical intrachromosomal R^2 in chromosome 3 of the same hybrid zone, displayed as a triangular heatmap with excess ancestry LD in the upper-left triangle and traditional LD in the lower-right triangle. Blue rectangles zoom in on regions with identified outliers, and blue circles show the location of specific outlier region pairs (FDR < 0.01 for panel C and FDR < 0.05 for panel D).

D).

300x360mm (300 x 300 DPI)

ACCEPTED MANUSCRIPT