

**Recent climate change and historical population structure predict spatial patterns of admixture between two host-specialized pine sawfly species**

Running title: Hybridization dynamics in pine sawflies

Ashleigh N. Glover<sup>1</sup> and Catherine R. Linnen<sup>1</sup>

<sup>1</sup>Department of Biology, University of Kentucky, Lexington, KY 40506, United States

Corresponding author: Ashleigh N. Glover, [angl226@uky.edu](mailto:angl226@uky.edu)

## **ABSTRACT**

Human disturbance can have profound effects on biodiversity, including increasing hybridization between reproductively isolated species. One approach for understanding how human activity affects hybridization dynamics is to evaluate correlations between disturbance (e.g., urbanization, temperature change) and hybridization. Because variation in hybridization can also arise from historical factors unrelated to recent human disturbance, it is essential to account for population structure to avoid spurious correlations. Here, we combine environmental and high-coverage whole-genome resequencing data to investigate how human disturbance and population structure affect hybridization dynamics between a pair of pine sawflies adapted to different pines, *Neodiprion lecontei* and *Neodiprion pinetum*. We find that *N. lecontei* and *N. pinetum* exhibit strikingly different patterns of population structure, which we hypothesize stem from differences in host use. We also find that recent admixture is both asymmetric and geographically variable. Linear regression analyses reveal that admixture proportion is predicted by indirect human disturbance (i.e., climate change) and not direct human disturbance (e.g., urbanization) in both *N. lecontei* and *N. pinetum*. Lastly, in *N. pinetum*, we find evidence of a spurious association between admixture and direct human disturbance that disappears when regression models account for population structure via inclusion of genetic principal component scores as covariates. Together, our data suggest that indirect human disturbance and population structure both contribute to geographic variation in admixture between *N. lecontei* and *N. pinetum*. Our study also highlights the importance of adequately controlling for population structure when attempting to identify environmental predictors (human disturbance-related or not) of hybridization.

**Keywords:** hybridization, human disturbance, *Neodiprion*, admixture, population structure, climate change

## **INTRODUCTION**

Humans are rapidly modifying the planet, both directly (e.g., construction, cropland expansion, atmospheric light pollution; Wilson et al. 2021; Hernandez and Suni 2022; Kreiner et al. 2022; Grabenstein and Taylor 2018) and indirectly (i.e., climate change; Garroway et al. 2010; IPCC, 2007). Both direct and indirect human disturbance can have profound effects on genetic and phenotypic variation within and between species. For example, novel selection pressures via urbanization have resulted in phenotypic and genomic changes across taxa (Benmazouz et al. 2023; Winchell et al. 2023; Johnson and Munshi-South 2017; Kreiner et al. 2022) while climate change has been linked to higher mortality in sheep and fishes (Hallett et al. 2004; Finney et al. 2000) and variation in reproductive success in marine birds (Croxall et al. 2002; Horreo et al. 2011). Human-mediated landscape changes and climate change can also alter species' abundances and distributions and facilitate hybridization between reproductively isolated species (Garroway et al. 2010; De Meester et al. 2018; Grabenstein and Taylor 2018; Grabenstein et al. 2023). For species that naturally co-occur, human disturbance can promote gene exchange via three non-mutually exclusive mechanisms: (1) breaking down physical or ecological barriers to gene flow by altering habitat structure and/or phenology (Grabenstein and Taylor 2018; Grabenstein et al. 2023); (2) disrupting assortative mating by interfering with visual, chemical, and acoustic mate-recognition cues (Fisher et al. 2006; Grabenstein and Taylor 2018); and (3) creating novel environments with relaxed selection against hybrids, increasing hybrid survival (Seifert et al. 2010; Hasselman et al. 2014; Grabenstein et al. 2023). Overall, human disturbance can promote gene exchange by reducing the strength of prezygotic and/or postzygotic isolation, although disturbances that reduce prezygotic reproductive isolating barriers are most likely to promote hybridization (Grabenstein and Taylor 2018). Importantly, any localized changes in selection pressures and/or expression of traits facilitating prezygotic isolation (e.g., via these mechanisms) can cause variation in hybridization across geographic space and time (Hasselman et al. 2014; Sianta et al. 2024).

Understanding hybridization dynamics between taxa is critical because hybridization, which occurs in an estimated 10% of animal species and 25% of plant species, can have important evolutionary consequences (Walsh et al. 2023; Aguillon et al. 2022). When introgression occurs via hybridization and backcrossing, the evolutionary trajectories of the hybridizing taxa can be altered drastically. For example, extensive admixture can result in the formation of a new hybrid species, thereby increasing biodiversity, or in the extinction of one of the parental species via genetic swamping, thereby decreasing biodiversity (DeVos et al. 2023; Aguillon et al. 2022). Introgression can also increase genetic diversity within the recipient species, potentially increasing its adaptive potential (Hasselman et al. 2014; Norris et al. 2015; Oziolor et al. 2019).

Despite the increased recognition that the environment can play a critical role in shaping geographical patterns of hybridization, the mechanisms responsible for generating variation in hybridization dynamics remain unknown in most hybridizing taxa. Additionally, our understanding of the consequences of such hybridization is largely incomplete (Mandeville et al. 2022; Grabenstein et al. 2023; Sianta et al. 2024). Characterizing patterns and establishing correlations between human disturbance and hybridization is a necessary first step for generating hypotheses to guide future work that explores the mechanisms that promote or prevent hybridization (Grabenstein et al. 2023). Relationships between indirect human disturbance (climate change) and hybridization have been recovered in a variety of taxa (e.g., Garroway et al. 2010; Muhlfeld et al. 2014, 2017; Sianta et al. 2024), but relatively few studies have been explicitly designed to explore relationships between direct human disturbance and hybridization (Grabenstein et al. 2023). Even fewer studies test for relationships between hybridization and both direct and indirect human disturbance simultaneously (but see Ortego et al. 2017; Muhlfeld et al. 2017).

When describing geographical patterns of hybridization and introgression, it is critical to account for population structure. In addition to the environment, population structure within species can produce geographic variation in hybridization and introgression. For example, local adaptation and/or geographic variation in historical overlap/isolation can cause geographic variation in the strength of reproductive isolating

barriers (Borge et al. 2005; Poikela et al. 2019; Glover et al. 2023), producing variable patterns of hybridization and introgression (Walsh et al. 2016; Sianta et al. 2024; Faske et al. 2024). Any recent changes in hybridization due to human disturbance are overlaid on these historical processes. Therefore, not accounting for population structure can produce spurious associations between measures of hybridization and human disturbance. The confounding effect of population structure is well recognized in genome-wide association studies (GWAS; Young 2024) and genome-environment association (GEA) studies (Forester et al. 2018). A common approach to account for population structure in these studies is to quantify population structure using principal component analysis (PCA) and include genetic PC scores as covariates in subsequent models testing for associations between genotype and phenotype/environment (Price et al. 2006; Forester et al. 2018; Privé et al. 2020). Another approach used to account for population structure is to include a proxy for population structure (e.g., sampling site; kinship matrix) as a random effect in models (Liu et al. 2011; Reutimann et al. 2023; Fachrul et al. 2023).

These approaches can be extended to studies testing for relationships between hybridization and human disturbance, although few of these existing studies account for population structure within species (but see Ortego et al. 2017; Muhlfeld et al. 2017; Grabenstein et al. 2023; DeVos et al. 2023). For example, to investigate the impact of human disturbance on hybridization between Californian *Quercus* oak species, Ortego et al. (2017) included latitude and longitude as model covariates while Grabenstein et al. (2023) included sampling site as a random effect in their model to test for a correlation between hybridization and human disturbance in *Poecile* chickadees. Notably, the efficacy of different models (i.e., covariates versus random effects) at preventing spurious associations between hybridization and human disturbance remains unexplored (but see Ortego et al. 2017). Overall, investigating population structure provides important context for interpreting patterns of hybridization and introgression (Sianta et al. 2024). Disentangling the relative roles of recent human disturbance and historical population structure in shaping patterns of hybridization and introgression will enhance our understanding of how taxa have and will continue to respond to human disturbance and

shed light on how humans impact biodiversity. To fill this knowledge gap, we investigated the influence of human disturbance and population structure on hybridization dynamics between a sister-species pair of pine sawflies, *Neodiprion lecontei* and *Neodiprion pinetum*.

Like all *Neodiprion* (Order: Hymenoptera; Family: Diprionidae), *N. lecontei* and *N. pinetum* feed on conifers and are dependent upon their host plant for all stages of their life cycle (Coppel and Benjamin 1965; Knerer and Atwood 1973; Davis et al. 2023). These sister species have large, mostly overlapping ranges in eastern North America but are adapted to *Pinus* species with different needle morphologies (Linnen and Farrell 2008, 2010; Glover et al. 2023). *Neodiprion pinetum* specializes on the thin-needled white pine (*Pinus strobus*) and is the only species in the genus to prefer this thin-needled host while *N. lecontei* avoids white pine and uses other *Pinus* species that have thicker needles (Wilson et al. 1992; Linnen and Farrell 2010; Bendall et al. 2017). In addition to this divergent host preference, *N. lecontei* and *N. pinetum* also differ in adult body size, female ovipositor morphology and other egg-laying traits (Bendall et al. 2017; Glover et al. 2023). These divergent host-use traits contribute to both prezygotic and postzygotic reproductive isolation (Bendall et al. 2017, 2023; Glover et al. 2023). Notably, no-choice mating assays suggest that the strength of prezygotic isolation varies across geographic space (Glover et al. 2023). Despite strong reproductive isolation between *N. lecontei* and *N. pinetum*, these species occasionally hybridize in nature and viable, fertile hybrids for both sexes and cross directions can be produced in the lab (Bendall et al. 2017, 2022, 2023). Finally, historical hybridization has probably occurred between *N. lecontei* and *N. pinetum*: a recent demographic analysis suggests that they diverged in sympatry with continuous and asymmetric gene flow (Bendall et al. 2022).

Here, we combine environmental and high-coverage whole-genome resequencing data to describe patterns of hybridization between *N. lecontei* and *N. pinetum* and explore relationships between human disturbance (direct and indirect) and hybridization while investigating the confounding effect of population structure. If disturbance alters aspects of a species' environment that are important for maintaining reproductive isolating barriers,

then individuals in more disturbed areas will have a higher proportion of their genome introgressed from the species with which they hybridize (e.g., Ortego et al. 2017; Coppi et al. 2020). For example, direct human disturbance (e.g., urbanization) creates patchy host plant distributions in areas where the hosts are native, potentially making it more difficult for insects with limited dispersal capabilities to find a conspecific mate (Lopez et al. 2018). Urbanization also often involves ornamental planting of non-native host plants, potentially bringing insect species specialized on these host plants together in an area where they would not naturally co-occur (Lopez et al. 2018). Notably, white pine (*N. pinetum*'s host) is one of the most widely planted pines in the eastern United States, often ornamentally planted around homes and businesses (Wendel and Smith 1990; Dickerson 2002). Both habitat fragmentation and ornamental planting could increase hybridization rates. Therefore, we predict that *N. lecontei* and *N. pinetum* individuals in areas that have more urban and agricultural land cover will have a higher proportion of their genome introgressed. Additionally, indirect human disturbance (i.e., climate change) includes temperature changes. As temperature is important for insect development, faster development time in one or both species can decrease temporal isolation, allowing more opportunities to hybridize (Larson et al. 2019). Therefore, we predict that *N. lecontei* and *N. pinetum* individuals in areas that have experienced a higher increase in temperature will have a higher proportion of their genome introgressed.

Finally, to investigate the confounding effect of population structure, we first describe population structure in *N. lecontei* and *N. pinetum*. Previous work with ddRAD data in *N. lecontei* revealed distinct genetic clusters that were dubbed “North”, “Central”, and “South” based on corresponding geographic regions in eastern North America (Bagley et al. 2017). Demographic modeling suggests these clusters were produced by historical isolation in three different pine refugia, with the North cluster particularly distinct (i.e., lowest heterozygosity; Bagley et al. 2017). Subsequent work in *N. lecontei* detected an additional East/West genetic break in the Central cluster (Lindstedt et al. 2022). Here, we report the first analysis of population structure in *N. pinetum* and the first analysis of population structure in *N. lecontei* with whole-genome resequencing (WGS) data.

Importantly, to compare population structure between *N. lecontei* and *N. pinetum*, having similar data (i.e., WGS) is ideal. Using results from these analyses, we fit models that differ in the approach used to account for population structure and assess how these approaches affect inferred relationships between human disturbance and hybridization.

## **MATERIALS AND METHODS**

### **Sampling, DNA extraction and sequencing**

We sampled *Neodiprion lecontei* and *N. pinetum* mid- to late-instar larval colonies from regions where the species co-occur across eastern North America (Figure 1; Table S1). Because we required an allopatric population of *N. lecontei* for an analysis of historical introgression (see below), we also sampled a Florida population of *N. lecontei*. Larvae were either immediately preserved in 100% ethanol (stored at -20°C) or reared to adults in the lab using standard lab protocols (Bendall et al. 2017; Harper et al. 2016) and subsequently preserved in 100% ethanol (stored at -20°C). To avoid sampling closely related individuals, we selected one individual from each colony, as each colony typically represents a group of siblings. *Neodiprion* pine sawflies are haplodiploid: males develop from unfertilized eggs and are haploid across their entire genome while females develop from fertilized eggs and are diploid across their entire genome (Bendall et al. 2022; Glover et al. 2024). Thus, to ensure that our samples were all diploid, we selected one adult female from each colony when available. From colonies where no adult females were available, we selected the largest larva, as relatively larger larvae are usually diploid (Bagley et al. 2017). Due to sample availability and to ensure we sampled a broad range of intraspecific variation in the more variable species (*N. lecontei*), we sampled 314 *N. lecontei* individuals (78 adult females and 236 larvae) and 94 *N. pinetum* individuals (56 adult females and 38 larvae). Notably, although sample size affects power in regression analyses, *N. lecontei* and *N. pinetum* exhibited similar regression patterns (see Results below), suggesting that sample size differences between the species did not affect our main conclusions.

We extracted DNA from either head and thorax tissue (adult females) or head and thoracic leg tissue (larvae) with a Qiagen DNeasy Blood & Tissue Kit following the standard manufacturer protocol, including an optional RNase A step. We measured DNA concentration with a Quant-iT dsDNA High-Sensitivity fluorescence assay (Invitrogen). Extracted DNA was sent to Admera Health (Plainfield, NJ) for library preparation and whole-genome resequencing. A separate library was prepared for each individual with a KAPA Hyper Prep Min PCR kit and whole-genome resequencing was performed with 150 bp paired-end sequencing technology in an Illumina NovaSeq X plus sequencer. To reach approximately 20x coverage per individual, all libraries were sequenced in two batches.

### **Quality filtering and variant calling**

Demultiplexed reads were cleaned using trimmomatic v0.39 (Bolger et al. 2014) to remove Illumina adapters, low quality bases (quality score < 3) from the beginning and end of each read, any window (width = 4 bp) that had an average quality score below 15, and any individual read with a total length of less than 36 bp. For each batch separately, we mapped the cleaned reads to the chromosome-level *N. lecontei* reference genome (iyNeoLeco1.1, GCA\_021901455.1; Herrig et al. 2024) using the BWA-MEM algorithm in bwa v0.7.17 (Li and Durbin 2009). We also included previously generated whole-genome resequencing reads from an individual of an outgroup species (*Neodiprion hetricki*; Herrig et al. 2024) so that we could quantify historical introgression (see below). We then used samtools v1.13 (Danecek et al. 2021; Li et al. 2009) to mark and remove PCR duplicates and remove ambiguously mapped reads/secondary alignments. Finally, we used samtools to merge the filtered reads from the two sequencing batches (for *N. lecontei* and *N. pinetum*) and index the resulting bam files.

We called all sites (variant and invariant) using bcftools mpileup v1.13 (Danecek et al. 2021). We chose bcftools mpileup because it has been shown to outperform GATK for non-human data (Lefouili and Nam 2022). We then used a custom script from Mark Ravinet ([https://github.com/markravinet/genotyping\\_pipeline/blob/main/3\\_filter\\_variants.nf](https://github.com/markravinet/genotyping_pipeline/blob/main/3_filter_variants.nf)) to remove indels. We performed additional filtering with vcftools v1.16 (Danecek et al. 2011)

to create two high-quality datasets for downstream analyses. The first dataset (“SNPs only”) was restricted to *N. lecontei* and *N. pinetum* only and included only biallelic SNPs present in at least 90% of individuals. We also required retained SNPs to have a minimum site quality score of 20, mean site depth between 10-40X, genotype depth between 10-60X, and a minor allele frequency (MAF) above 0.05. The second dataset (“all sites”) included *N. lecontei*, *N. pinetum*, and *N. hetricki*. This dataset was filtered with the same missingness, site quality, and depth criteria as above but included all sites (i.e., invariant and variant) with no MAF filtering.

To check for irregularities indicative of sample contamination (as part of our standard data quality control pipeline) or putative haploid males (for larvae), we used a custom script from David Marques and Joana Meier (<https://github.com/speciationgenomics/scripts/blob/master/checkHetsIndvVCF.sh>) to calculate and visualize allelic imbalance for all heterozygote calls (i.e., allele skew) within the “SNPs only” dataset. In a diploid individual with minimal contamination, heterozygous positions contain roughly equal numbers of reads for both alleles (Figure S1a). However, diploid individuals with high levels of contamination and putative haploid males will exhibit strong allelic imbalance (Figures S1b-S1v)—males should exhibit zero heterozygosity as they are haploid across their entire genome. We excluded any sample for which more than 49% of heterozygote calls displayed allele balance ratios less than 0.3. This cutoff was determined based on data visualization and a threshold that struck a balance between stringency and retaining adequate data. This resulted in exclusion of 12 larvae that were putatively haploid males and 9 individuals with high levels of contamination (Figures S1b-S1v; Table S1). For both datasets (“SNPs only” and “all sites”), we filtered out any low levels of contamination with a custom script by Joana Meier (<https://github.com/joanam/scripts/blob/master/allelicBalance.py>) by converting heterozygote genotypes with allele balance ratios less than 0.2 to homozygous for the more common allele. Notably, our results were not impacted by converting imbalanced heterozygote genotypes to homozygous for the more common allele rather than to missing (Figure S2). Also, our measure of hybridization (admixture proportions, see below) did not

correlate with allele skew or any other genotype quality metrics (Figure S3; Table S2). We then used vcftools v1.16 (Danecek et al. 2011) to refilter the resulting vcf files with the same criteria above while simultaneously excluding the 21 putative haploid male/highly contaminated individuals. This resulted in a final dataset of 387 individuals (299 *N. lecontei* and 88 *N. pinetum*; Table S1).

### **Population structure and genomic introgression**

Prior to population structure analyses, we performed linkage pruning on the “SNPs only” dataset using vcftools v1.16 (Danecek et al. 2011). We required SNPs to be at least 1,000 bp apart because previous analyses indicate linkage disequilibrium decays to approximately zero at this distance (Glover et al. 2024). We then used PLINK v1.9 (Chang et al. 2015) to perform three principal component analyses (PCAs): both species combined, within *N. lecontei* only, and within *N. pinetum*. These scripts and scripts for all subsequent analyses are deposited on DRYAD (Glover and Linnen 2025). Next, we used ADMIXTURE v1.3 (Alexander et al. 2009) with both species combined to further investigate population structure using a model-based approach and to estimate individual admixture proportions. We set the number of assumed populations (K) between 1 and 10 and ran 100 iterations for each value of K. The best supported number of populations (K) was based on the lowest averaged five-fold cross-validation (CV) error across all 100 iterations per K. Because we were interested in species-level admixture for downstream analyses, we estimated admixture proportions using assignment probabilities for K=2 even though this was not the best supported K (see below). Final assignment probabilities (i.e., admixture proportion) for each individual at K=2 were estimated across all 100 iterations using the LargeKGreedy algorithm for 2,000 repetitions as implemented in CLUMPAK (Kopelman et al. 2015).

We consider estimated admixture proportions to be a measure of recent introgression because the ADMIXTURE model assumes modern individuals were produced by recent admixture between K ancestral populations (Lawson et al. 2018). However, a critical consideration when interpreting admixture proportions is that some demographic scenarios—such as isolation by distance (IBD)—can give rise to spurious evidence of

discrete population structure with “admixed” individuals in the center of the range (Meirmans 2012; Wiens and Colella 2025). For this reason, it is important to rule out IBD before interpreting admixture proportions as evidence of hybridization (Wiens and Colella 2025). Fortunately, several observations from prior studies reject a model of IBD for *N. lecontei* and *N. pinetum*: (1) these two species have largely overlapping ranges and can often be found in the same locations (vs. living at opposite ends of a continuous geographic range as expected under IBD) (Linnen and Farrell 2010); (2) they form discrete and highly differentiated clusters even in tight sympatry (e.g.,  $F_{ST} = 0.61$  in Lexington, KY; Glover et al. 2024); and (3) demographic modeling suggests these species diverged with continuous gene exchange, with an estimated divergence time of ~1.5 million generations ago (Bendall et al. 2022). Additionally, preliminary analyses with hybrid indices (Wiens et al. 2025), an alternate metric for hybridization, indicated they are highly sensitive to the choice of reference populations for *N. lecontei* and *N. pinetum* and to genotyping error. In contrast, admixture proportions did not require specification of reference populations and were not correlated with any genotyping error metrics, including allele skew, read count, coverage depth, and missing data (Figure S3; Table S2). For these reasons, we used assignment probabilities (i.e., admixture proportion) as our estimate of species-level admixture for downstream analyses.

To further investigate regional patterns of introgression, we used Dsuite v0.5 (Malinsky et al. 2021) to calculate Patterson’s D (ABBA-BABA) statistic (Green et al. 2010; Durand et al. 2011). This analysis considers ancestral (“A”) and derived (“B”) alleles at each site for three populations and an outgroup defined as  $((P_1, P_2), P_3), O$  and tests for an excess of “ABBA” or “BABA” patterns (Durand et al. 2011; Martin et al. 2014). A D-statistic that is positive and significantly different from zero (i.e., excess of “ABBA”) indicates introgression between  $P_2$  and  $P_3$ . Conversely, a D-statistic that is negative and significantly different from zero (i.e., excess of “BABA”) indicates introgression between  $P_1$  and  $P_3$ . Notably, the D-statistic is well documented to detect ancient introgression (McFarlane and Pemberton 2019; Dagilis et al. 2022). However, because the D-statistic also captures recent introgression (Dagilis et al. 2022), additional analyses are required to date the timing

of introgression (e.g., Hibbins and Hahn 2019). This analysis was performed with the “all sites” dataset, which included allopatric *N. lecontei* ( $P_1$ ) and *N. hetricki* (the outgroup). Because our PCA and ADMIXTURE analyses revealed a distinct genetic break within *N. lecontei* that corresponded to “North” and “Central” geographic regions (see below; Bagley et al. 2017), we calculated and assessed the significance of the D-statistic for each region separately with the following relationships: (((allopatric *N. lecontei* (Florida), North *N. lecontei*), North *N. pinetum*), *N. hetricki*) and (((allopatric *N. lecontei* (Florida), Central *N. lecontei*), Central *N. pinetum*), *N. hetricki*).

### **Quantifying human disturbance**

To quantify direct human disturbance, we first downloaded 2021 land cover GIS layers for the eastern United States produced by the National Land Cover Database (Dewitz 2021; <https://www.mrlc.gov/viewer/>) and a 2020 land cover GIS layer for Canada produced by the Canada Centre for Remote Sensing (Commission for Environmental Cooperation (CEC) 2023; <http://www.cec.org/north-american-environmental-atlas/land-cover-30m-2020/>). These layers provide land cover data at a 30m x 30m spatial resolution. Next, for each sampling site, we used the *sf* v1.0-16 package (Pebesma 2018; Pebesma and Bivand 2023) in R v4.3.3 (R Core Team 2024) to create a 5km-radius buffer around the GPS coordinates. We chose this distance because it encompasses the estimated maximum dispersal distance of individual *Neodiprion* males (Östrand et al. 2001). Then, we used the *terra* v1.7-71 package (Hijmans 2024) to calculate the proportion of each land cover class within each 5km-radius buffer. Finally, to reduce all land cover classes—which are correlated to varying degrees—to a composite measure of disturbance at each sampling site, we used the *princomp*() function in the *stats* v4.3.3 package (R Core Team 2024) to perform a PCA of the land cover proportion data for each sampling site. Prior to performing the PCA, for the United States sampling sites, some land cover classes were grouped together (proportions added together) to be comparable with the Canadian land cover classes: all “Developed” land cover classes (i.e., “Developed Open”, “Developed Low Intensity”, “Developed Medium Intensity”, and “Developed High Intensity”) were grouped

as “Urban”, “Pasture and Hay” and “Cultured Crop” were grouped as “Cropland”, and “Woody Wetland” and “Emergent Herbaceous Wetland” were grouped as “Wetland”. PC1 separated disturbed sites (i.e., higher urban and agricultural land cover; negative PC values) from “natural” sites (i.e., more forest, shrubland and grassland land cover; positive PC values) while PC2 separated low vegetation sites (i.e., more barren, water and wetland land cover; positive PC values) from high vegetation (of various types) sites (i.e., negative PC values; Figure 2).

To quantify indirect human disturbance (i.e., climate change), we first downloaded monthly average daily maximum temperature (°C) GIS layers for North America produced by NASA (Thornton et al. 2022; [https://daac.ornl.gov/cgi-bin/dsvviewer.pl?ds\\_id=2131](https://daac.ornl.gov/cgi-bin/dsvviewer.pl?ds_id=2131)) for the years 1980-1989 and 2010-2019. These layers provide temperature data at a 1km x 1km spatial resolution. Next, for each sampling site, we used the *terra* v1.7-71 package (Hijmans 2024) to extract the average of daily maximum temperature for each of the 12 months within each year. Then, we calculated the average daily maximum temperature across all months for the years 1980-1989 and for the years 2010-2019 separately to estimate the average daily maximum temperature for each decade (Tmax). Finally, to quantify the change in average daily maximum temperature between the two decades at each sampling site, we subtracted the Tmax estimate for 1980-1989 from the Tmax estimate for 2010-2019 (change in Tmax (°C)). Thus, positive values indicated an increase in the average daily maximum temperature from the 1980’s to the 2010’s while negative values indicated a decrease in the average daily maximum temperature.

### **Statistical analyses**

To explore relationships between admixture proportion and human disturbance (direct and indirect) and investigate how the approach used to account for population structure within species affects these relationships, we performed multiple linear regression with the *glmmTMB* v1.1.9 package (Brooks et al. 2017) in R v4.3.3 (R Core Team 2024). Due to asymmetric gene flow between *N. lecontei* and *N. pinetum* (Bendall et al. 2022), we modeled each species separately. For *N. pinetum*, we excluded one individual

that was originally classified as *N. pinetum* at the time of collection (because it was collected on white pine) but was later identified by PCA and ADMIXTURE analyses as an F1 hybrid (see below). For all models, we used the beta family with a logit link and a zero-inflation formula to model the distribution of a response variable that was proportion data with zeros present in the dataset. Prior to running each model, all predictor variables were normal-quantile transformed to ensure that they were on the same scale.

Because we were interested in testing for relationships between admixture proportion and human disturbance while simultaneously exploring the potentially confounding effect of population structure, we fit three different models for each species separately. For all models, admixture proportion was the response variable and land cover PC1, land cover PC2 and change in average daily maximum temperature (Tmax) were included as predictor variables. However, each model differed in the approach used to account for population structure. In the first model, we included quantitative measures of population structure—specifically PC1 and PC2 scores from within-species genetic PCAs (see above)—as model covariates to control for the potentially confounding effect of population structure. For each species, the model took the form:

(M1) Admixture proportion  $\sim$  land cover PC1 + land cover PC2 + change in Tmax + genetic PC1 + genetic PC2

As a second approach to account for population structure, we included “grouped site number” as a random effect, where all sampling sites that were within 5km of each other were considered a “grouped site” and were assigned an arbitrary grouped site number. With this approach, population structure was not directly quantified. Instead, grouped sites were assumed to capture the genetic variation across geographic space (i.e., correspond to population structure) as individuals sampled in close geographic proximity tend to be more genetically similar (Bradburd et al. 2016). Thus, including grouped site number as a random effect allowed us to test for relationships between admixture proportion and measures of human disturbance while accounting for variability across

sites (i.e., population structure). Specifically, for each species, the model took the following form:

(M2) Admixture proportion  $\sim$  land cover PC1 + land cover PC2 + change in Tmax + (1|grouped site number)

Finally, we tested for relationships between admixture proportion and measures of human disturbance without including a control for population structure (i.e., no covariates or random effects were included). For each species, this third model took the form:

(M3) Admixture proportion  $\sim$  land cover PC1 + land cover PC2 + change in Tmax

## **RESULTS**

After mapping trimmed reads to the *N. lecontei* reference genome and filtering, we retained 46,957,406 variant and invariant sites (“all sites” dataset) and 207,049 linkage-pruned SNPs (from the “SNPs only” dataset). Average depth coverage across all retained individuals (299 *N. lecontei*, 87 *N. pinetum*, 1 F1 hybrid, and 1 *N. hetricki*) was 20.8x (range: 12.5x – 29.2x).

### **Patterns of population structure differ between *Neodiprion lecontei* and *N. pinetum***

When both species were included, the PCA clearly separated *N. lecontei* and *N. pinetum* and revealed different patterns of population structure within *N. lecontei* and *N. pinetum* (Figure 3a). PC1 separated *N. lecontei* and *N. pinetum* and identified one F1 hybrid (PC1 = 61.4% variance explained). While *N. pinetum* individuals were more tightly clustered together, PC2 separated *N. lecontei* into distinct clusters (PC2 = 17.4% variance explained). The PCA of only *N. lecontei* individuals recapitulated the distinct genetic breaks within *N. lecontei* that generally corresponded to geography (Figure 3b). PC1 separated *N. lecontei* from northern US states and Canada (“North”; positive side of PC1 axis) and *N. lecontei* from central US states (“Central”; negative side of PC1 axis) (PC1 = 59.4% variance

explained). PC2 further separated Central *N. lecontei* into two groups: individuals collected in northeastern US states and states east of the Appalachian Mountains (positive side of PC2 axis) and individuals collected in states west of the Appalachian Mountains (negative side of PC2 axis), although this split was much less pronounced than the North/Central split (PC2 = 9.76% variance explained). In contrast to *N. lecontei*, the PCA of only *N. pinetum* individuals revealed less discrete genetic clustering (Figure 3c). Rather, individuals were distributed more continuously across both PC1 (56.4% variance explained) and PC2 (12.2% variance explained).

The ADMIXTURE analysis revealed similar patterns of population structure as the PCAs. The best supported number of populations was K=3, although K=4 was also a reasonable solution (Figure S4). Regardless of K, all *N. pinetum* individuals formed one genetic group (Figures 3d-e, S5). At K=3, the remaining two genetic groups represented North and Central *N. lecontei* (Figure 3e). At K=4, the Central *N. lecontei* group was further split into eastern and western groups with extensive admixture between these groups (Figure S5). This analysis also detected the single F1 hybrid that was originally classified as *N. pinetum* at the time of collection (Figures 3d-e, S5).

### **Patterns of admixture differ between geographic regions**

Admixture proportions were low overall in both species (Figure 3d). Admixture proportions ranged from 0.0 to 0.019 in *N. lecontei* and from 0.0 to 0.05 in *N. pinetum*. For both species, admixture proportions varied among geographic regions (Figure 4). Admixture proportions were lower in the North (*N. lecontei*: mean  $0.0 \pm 0.0$  SE; *N. pinetum*: mean  $0.0 \pm 0.0$  SE) compared to the Central region (*N. lecontei*: mean  $2.92 \times 10^{-3} \pm 3.45 \times 10^{-4}$  SE; *N. pinetum*: mean  $6.37 \times 10^{-3} \pm 1.50 \times 10^{-3}$  SE). In contrast, estimates of Patterson's D statistic revealed evidence of introgression between *N. lecontei* and *N. pinetum* in both the North and Central regions (Figure 4). Patterson's D statistic (D) was significantly different from zero and positive for both regions, being lower in the Central ( $D = 0.019$ ,  $p = 2.35 \times 10^{-11}$ ) compared to the North ( $D = 0.046$ ,  $p = 7.61 \times 10^{-4}$ ).

## Hybridization consistently correlates with indirect human disturbance

We used multiple linear regression to explore whether admixture proportion correlated with direct and indirect human disturbance and whether these correlations were affected by the approach used to account for population structure. When genetic PC scores were included as model covariates (M1), admixture proportion significantly correlated with at least one genetic PC axis in each species (Table 1). Additionally, for both species, change in average daily maximum temperature (indirect human disturbance) was the only measure of human disturbance to significantly correlate with admixture proportion; the relationship was positive in both species (Figures 5a, b; Table 1). When we included grouped site as a random effect to account for population structure (M2) or did not include a control for population structure (M3), only change in average daily maximum temperature significantly and positively correlated with admixture proportion in *N. lecontei* (Figures 5c-f; Table 1). In *N. pinetum*, both change in average daily maximum temperature and land cover PC1 (direct human disturbance) significantly correlated with admixture proportion, with all relationships being positive (Figures 5c-f; Table 1).

## **DISCUSSION**

In this study, we used environmental data and high-coverage whole-genome resequencing data to describe patterns of hybridization between *N. lecontei* and *N. pinetum* and explore relationships between human disturbance and admixture while investigating how the approach used to account for population structure affected these relationships. We find that *N. lecontei* and *N. pinetum* exhibit strikingly different patterns of population structure (Figure 3) and that patterns of admixture vary among geographic regions and, potentially, between recent and historical time points (Figure 4). Additionally, we find that admixture proportion is consistently predicted by indirect human disturbance (i.e., change in average daily maximum temperature) and not direct human disturbance in both *N. lecontei* and *N. pinetum* (Figure 5; Table 1). Lastly, in *N. pinetum*, we find evidence of a spurious association between admixture proportion and direct human disturbance (i.e., land cover PC1). Here, we first discuss patterns of population structure in *N. lecontei*

and *N. pinetum*. Then, we discuss two non-mutually exclusive mechanisms that could explain how increasing temperatures have affected hybridization dynamics between *N. lecontei* and *N. pinetum*. Finally, we highlight how knowledge of population structure provides important context for interpreting patterns of admixture, offering a more complete picture of hybridization dynamics between taxa, and discuss the implications of model structure when testing for relationships between admixture and human disturbance.

### **Patterns of population structure differ between *N. lecontei* and *N. pinetum***

*Neodiprion lecontei* and *N. pinetum* exhibit strikingly different patterns of population structure (Figure 3). Our whole-genome resequencing analysis of population structure within *N. lecontei* recapitulates patterns previously described using ddRAD data (Bagley et al. 2017; Lindstedt et al. 2022). Demographic modeling suggests that the three distinct genetic clusters within *N. lecontei* (North, Central, and South) likely formed when populations were isolated in three glacial *Pinus* refugia within the US: one southwestern refugium, one mid-Atlantic refugium, and one northeastern refugium (Bagley et al. 2017). Additionally, each of the three primary *N. lecontei* lineages is currently associated with a different assemblage of *Pinus* species that differ in needle morphology (Bagley et al. 2017; Glover et al. 2023). Notably, adaptation to different *Pinus* species has been shown to produce rapid genetic and phenotypic differentiation in *N. lecontei*, even at a single site with no geographic isolation (Bagley et al. 2023). Thus, genetic divergence within *N. lecontei* was probably the result of genetic drift due to historical isolation (e.g., Qvarnström et al. 2016; Lozier et al. 2023) and adaptation to different local hosts and climates (e.g., Nelson et al. 2022; MacDonald et al. 2022).

In contrast to *N. lecontei*, *N. pinetum* does not show any evidence of discrete genetic breaks that might be attributable to historical isolation (Figures 3c-e, S5). This finding is consistent with analyses of genetic data for *Pinus strobus* (the only host of *N. pinetum*) that suggest that this pine species occurred in a single mid-Atlantic pine refugium (overlapping with *N. lecontei*) during the last glacial maximum (Zinck and Rajora 2016). Overall, we hypothesize that observed differences in patterns of population structure

between *N. lecontei* and *N. pinetum* are likely attributable to (1) differences in available *Pinus* refugia during the Pleistocene, and (2) different opportunities for divergent selection and host-associated differentiation for generalist (*N. lecontei*) vs. specialist (*N. pinetum*) *Pinus* feeders.

### **Potential mechanisms for correlations between increased temperature and hybridization**

Differences in population structure aside, we found that admixture proportions for both species were positively correlated with temperature change, and this finding was robust to the approach used to account for population structure (Table 1). One potential mechanism for increased hybridization due to increasing temperatures is via changes in adult sawfly body size. In general, as ectotherms, temperature has a strong effect on insect growth rate and on the rate of biochemical reactions (Davidowitz and Nijhout 2004; Chown and Gaston 2010). Higher temperatures often reduce total development time such that, even within a single species, individuals in warmer areas are often smaller than individuals in colder areas (reviewed in Chown and Gaston 2010). Within *Neodiprion*, females tend to be larger than males. Additionally, previous work demonstrated that *N. lecontei* females and males tend to be larger than *N. pinetum* females and males and that both species exhibit size-assortative mating: pairs that are more similar in size are more likely to mate (Glover et al. 2023). Thus, it is possible that in areas that have experienced a larger increase in temperature, *N. lecontei* and/or *N. pinetum* adult body sizes have decreased, increasing their willingness to hybridize. Due to widespread geographic variation in body size in nature (Ashton 2002; Chown and Gaston 2010; Stillwell 2010; Terribile et al. 2009; Auteri 2022) and widespread size-assortative mating across taxa (e.g., Bearhop et al. 2005; Jones et al. 2003; Greenway et al. 2016; Rougemont et al. 2015), experimental studies investigating the role of increasing temperatures in body size changes and subsequent hybridization dynamics across taxa would be a fruitful avenue for future research. These studies could also shed light on the role of plasticity compared to genetic adaptation to climate change,

an understudied topic in global change research (Chown and Gaston 2010; De Meester et al. 2018).

A second potential mechanism for increased hybridization due to increasing temperatures is via changes in phenology. Multiple ecological (e.g., synchronization with host) and physiological (e.g., development time) processes can lead to phenological divergence between naturally co-occurring species, causing temporal isolation between the species (Nelson et al. 2022). However, increasing temperatures can change the context of species interactions by potentially eroding temporal isolation: within insects, higher temperatures can accelerate growth rate (Chown and Gaston 2010) and/or alter seasonal timing of adult emergence (e.g., Buckley et al. 2015), thereby increasing overlap of adults from each species and opportunities for hybridization (Larson et al. 2019). Within *N. lecontei*, larval hatching and adult emergence from cocoons occur earlier when temperatures are higher (Benjamin 1955). Thus, it is possible that overlap in *N. lecontei* and *N. pinetum* adult emergence has increased in areas that have experienced a higher increase in temperature, resulting in more hybridization and introgression. Indeed, the breakdown of temporal isolation due to climate change has been observed in a variety of taxa, including plants (e.g., Rossetto et al. 2011; Theobald et al. 2017), insects (Taylor and Friesen 2017), and birds (e.g., Bom et al. 2023).

Importantly, increased hybridization due to temperature-related changes in adult body size and changes in phenology are not mutually exclusive processes. For example, increasing temperatures, and subsequently growth rates, can cause insects to increase the number of generations they have per year (Tobin et al. 2008; Altermatt 2010; Larson et al. 2019). This can impact both adult body size (e.g., result in smaller body size; Larson et al. 2019) and phenology (e.g., reduce temporal isolation between species; Dopman et al. 2010). Thus, it is possible that climate change has resulted in the breakdown of multiple prezygotic barriers to gene flow between *N. lecontei* and *N. pinetum*. However, ultimately, our inferences are based on correlative approaches and the specific mechanisms responsible for generating patterns of hybridization between *N. lecontei* and *N. pinetum* are unknown. Therefore, future manipulative experiments that investigate the temperature

dependence of reproductive traits in *Neodiprion* and other taxa are necessary to enhance our understanding of the long-term maintenance of species boundaries in the face of rapid global change.

Our finding that indirect human disturbance, but not direct human disturbance, consistently predicted admixture proportion in *N. lecontei* and *N. pinetum* contrasts with other studies testing for relationships between human disturbance (direct and indirect) and hybridization. For example, Ortego et al. (2017) found that wildfire frequency but not climatic variables (indirect disturbance) nor urbanization/land clearing for agriculture (direct human disturbance) predicted admixture in Californian *Quercus* oaks. Conversely, Muhlfeld et al. (2017) found that both mean summer stream temperature (indirect human disturbance) and road density (direct human disturbance) predicted admixture in *Oncorhynchus* trout. Overall, this highlights the importance of evaluating human-mediated hybridization in diverse taxa.

### **The importance of controlling for population structure when evaluating correlates of hybridization**

Although change in average daily maximum temperature from the 1980's to the 2010's consistently predicted admixture proportion in *N. lecontei* and *N. pinetum* (Figure 5; Table 1), increasing temperature is not the only factor affecting hybridization dynamics between these two species. For example, although some sampling sites in the North region (e.g., Canada) have experienced relatively high temperature increases (Figure 1), we found no evidence of recent admixture in either species (i.e., admixture proportions of zero for all individuals of both species; Figure 4).

Even before climate change, population structure within species and variation in historical contact between species could give rise to geographic variation in reproductive isolating barriers (e.g., Borge et al. 2005; Hoskin et al. 2005; Martin and Willis 2010; Matute et al. 2014; Poikela et al. 2019), and any recent changes in hybridization due to direct and indirect human effects are overlaid on these historical processes. For example, both species exhibit geographic clines in body size, but in opposite directions: whereas *N.*

*lecontei* males and females tend to increase in size with latitude, *N. pinetum* adult body size is negatively correlated with latitude (Glover et al. 2023). In *N. lecontei*, latitudinal clines in body size mirror latitudinal clines in host needle width, whereas in *N. pinetum*, body size clines are more likely driven by abiotic factors (Glover et al. 2023). Regardless of their selective causes, these opposing body size clines result in northern populations having greater body size disparities between *N. lecontei* and *N. pinetum* compared to populations at lower latitudes (hereafter “central” populations), possibly accounting for pronounced differences in admixture proportions between the two regions (Figure 4). In further support of this hypothesis, crosses between northern *N. pinetum* and central *N. lecontei* populations (bigger body size differences) yielded stronger prezygotic isolation than crosses between central populations of each species (smaller body size differences) (Glover et al. 2023).

Geographic variation in reproductive isolation could also stem from spatial and temporal variation in range overlap between the two species. For example, temporary isolation of *N. lecontei* populations from *N. pinetum* in some glacial refugia (i.e., in the southwestern and northeastern refugia), but not others (e.g., the mid-Atlantic refugium), may have provided spatially variable opportunities to fix alleles contributing to postzygotic isolation (Cutter 2012). Likewise, opportunities for hybridization—whether in primary or secondary contact—could have strengthened prezygotic isolation via reinforcement (Servedio and Noor 2003). Previous work lends support to this hypothesis: using no-choice mating assays, Glover et al. (2023) found that prezygotic isolation was stronger in two crosses that included *N. lecontei* and *N. pinetum* from areas where they co-occur compared to a cross involving allopatric populations of *N. lecontei* and *N. pinetum*, a pattern consistent with reinforcement (Servedio and Noor 2003; Butlin and Smadja 2018). However, more experimental work is needed to fully characterize geographic variation in pre- and postzygotic isolation, and more modelling work is needed to quantify spatial and temporal variation in gene flow between *N. lecontei* and *N. pinetum*.

Whatever the cause, there is evidence of geographic variation in the strength of reproductive isolation (Glover et al. 2023) and the amount of admixture (Figure 4) between

*N. lecontei* and *N. pinetum* that correlates to some degree with observed patterns of population structure (e.g., highest reproductive isolation and lowest admixture proportions correspond to the North *N. lecontei* cluster). Our finding that population structure likely correlates with admixture is consistent with several other studies (e.g., Hamlin and Arnold 2014; Mandeville et al. 2015; Lewanski et al. 2022), suggesting this may be a widespread phenomenon. Because population structure should also correlate with environmental variables, including our metrics of human disturbance, failure to account for this structure can give rise to spurious associations between human disturbance and admixture that have nothing to do with recent human activity.

To investigate the confounding effect of population structure, we compared the results of three models that differed in the approach used to control for population structure. Although our models consistently revealed a significant and positive relationship between indirect human disturbance (change in average daily maximum temperature) and admixture in both species (Figures 5a, c, e; Table 1), our data hint at a spurious association between direct human disturbance (land cover PC1) and admixture in *N. pinetum*. When we included genetic PC scores as covariates in the model (M1), we did not recover a significant correlation between land cover PC1 and admixture proportion (Figure 5b; Table 1). However, when we included grouped site number as a random effect (M2) or did not include a control for population structure in the model (M3), we recovered a significant and positive correlation (Figures 5d, f; Table 1). Further inspection of the data revealed that the significant relationships between land cover PC1 and admixture proportion in these models are probably driven by *N. pinetum* sampled from Virginia. These Virginia samples represent the most extreme PC2 values in the within-species PCA (Figure S6a). Additionally, these individuals have the highest admixture proportions, and all Virginia sampling sites load on the highly positive land cover PC1 axis (Figures S6b-d). Therefore, it is likely that the significant relationships between land cover PC1 and admixture proportion for models M2 and M3 in *N. pinetum* are confounded with population structure. These results highlight that methods for accounting for population structure can affect conclusions about relationships between hybridization and human disturbance.

Interestingly, it appears that including grouped site number as a random effect (M2) is insufficient to control for the confounding effect of population structure in *N. pinetum* compared to including genetic PCs as model covariates (M1). Although grouped sites may capture possible relatedness among samples, they likely do not capture broader patterns of population structure such as discrete structure (e.g., via historical isolation) or isolation by distance. While the effectiveness of different modeling approaches to correct for population structure has been investigated in GWAS (e.g., Price et al. 2010; Liu et al. 2011) and GEA studies (e.g., Forester et al. 2018), this topic remains unexplored in studies testing for relationships between hybridization and human disturbance (but see Ortego et al. 2017). Therefore, we suggest that studies investigating how correcting for population structure affects relationships between hybridization and human disturbance under different scenarios of population structure (e.g., with simulations) are a high priority. Overall, the genetics of reproductive isolation, in addition to environmental context, can cause variation in hybridization across geographic space (Sianta et al. 2024). Thus, accounting for variation within species is crucial to our understanding of human impacts on biodiversity.

## **CONCLUSIONS**

Our study documents geographic variation in recent admixture between two pine sawfly species that co-occur throughout much of eastern North America. Our study also adds to a growing body of literature implicating climate change in altering hybridization dynamics between species (e.g., Garroway et al. 2010; Muhlfeld et al. 2014, 2017; Sianta et al. 2024). Moving forward, high priorities for future work in this tractable insect system are to (1) uncover the causal mechanisms that produce observed correlations between temperature change and admixture, and (2) evaluate the evolutionary consequences (e.g., adaptive introgression) of hybridization. Our study also highlights the importance of considering population structure within species for a more complete picture of the factors affecting hybridization dynamics between species. Overall, accurately characterizing the patterns, causes, and consequences of human disturbance-mediated hybridization is

critical for our ability to better predict how biodiversity will be affected as our planet rapidly changes.

## **ACKNOWLEDGEMENTS**

We thank members of the Linnen lab for assistance with pine sawfly collection and rearing. We also thank three anonymous reviewers for helpful feedback. This work was supported by the USDA National Institute of Food and Agriculture Predoctoral Fellowship (2022-67011-36550) to ANG and the National Science Foundation DEB-CAREER-1750946 and DEB-2348574 to CRL. We thank the University of Kentucky Center for Computational Sciences and Information Technology Services Research Computing for their support and use of the Morgan Compute Cluster.

## **REFERENCES**

- Aguillon, S. M., Dodge, T. O., Preising, G. A., & Schumer, M. (2022). Introgression. *Current Biology*, 32(16), R865-R868.
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655-1664.  
<https://doi.org/10.1101/gr.094052.109>
- Altermatt, F. (2010). Climatic warming increases voltinism in European butterflies and moths. *Proceedings of the Royal Society B: Biological Sciences*, 277(1685), 1281-1287. <https://doi.org/10.1098/rspb.2009.1910>
- Ashton, K. G. (2002). Patterns of within-species body size variation of birds: Strong evidence for Bergmann's rule. *Global Ecology and Biogeography*, 11(6), 505–523.  
<https://doi.org/10.1046/j.1466-822x.2002.00313.x>
- Auteri, G. G. (2022). A conceptual framework to integrate cold-survival strategies: Torpor, resistance and seasonal migration. *Biology Letters*, 18(5), 20220050.  
<https://doi.org/10.1098/sbl.2022.0050>
- Bagley, R. K., Hurst, M. N., Frederick, J., Wolfe, J., Terbot II, J. W., Frost, C. J., & Linnen, C. R. (2023). Multiple mechanisms contribute to isolation by environment in the

- redheaded pine sawfly, *Neodiprion lecontei*. *Evolution*, 77(10), 2257-2276.  
<https://doi.org/10.1093/evolut/qp137>
- Bagley, R. K., Sousa, V. C., Niemiller, M. L., & Linnen, C. R. (2017). History, geography and host use shape genomewide patterns of genetic variation in the redheaded pine sawfly (*Neodiprion lecontei*). *Molecular Ecology*, 26(4), 1022-1044.  
<https://doi.org/10.1111/mec.13972>
- Bearhop, S., Fiedler, W., Furness, R. W., Votier, S. C., Waldron, S., Newton, J., Bowen, G. J., Berthold, P., & Farnsworth, K. (2005). Assortative mating as a mechanism for rapid evolution of a migratory divide. *Science*, 310(5747), 502–504.  
<https://doi.org/10.1126/science.1115661>
- Bendall, E. E., Bagley, R. K., Sousa, V. C., & Linnen, C. R. (2022). Faster-haplodiploid evolution under divergence-with-gene-flow: Simulations and empirical data from pine-feeding hymenopterans. *Molecular Ecology*, 31, 2348–2366.  
<https://doi.org/10.1111/mec.16410>
- Bendall, E. E., Mattingly, K. M., Moehring, A. J., & Linnen, C. R. (2023). A test of Haldane's rule in *Neodiprion* sawflies and implications for the evolution of postzygotic isolation in haplodiploids. *The American Naturalist*, 202(1), 40–54.  
<https://doi.org/10.1086/724820>
- Bendall, E. E., Vertacnik, K. L., & Linnen, C. R. (2017). Oviposition traits generate extrinsic postzygotic isolation between two pine sawfly species. *BMC Evolutionary Biology*, 17(1), 1–15. <https://doi.org/10.1186/s12862-017-0872-8>
- Benjamin, D. M. (1955). The biology and ecology of the red-headed pine sawfly. *USDA Technical Bulletins*, 118, 55.
- Benmazouz, I., Jokimäki, J., Juhász, L., Kaisanlahti-Jokimäki, M. L., Paládi, P., Kardos, G., ... & Kövér, L. (2023). Morphological changes in Hooded Crows (*Corvus cornix*) related to urbanization. *Frontiers in Ecology and Evolution*, 11, 1196075.  
<https://doi.org/10.3389/fevo.2023.1196075>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina

- sequence data. *Bioinformatics*, 30(15), 2114–2120.  
<https://doi.org/10.1093/bioinformatics/btu170>
- Bom, R. A., Piersma, T., Alves, J. A., & Rakhimberdiev, E. (2023). Global temperature homogenization can obliterate temporal isolation in migratory animals with potential loss of population structure. *Global Change Biology*, 30(1), e17069.  
<https://doi.org/10.1111/gcb.17069>
- Borge, T., Lindroos, K., Nádvorník, P., Syvänen, A. C., & Sætre, G. P. (2005). Amount of introgression in flycatcher hybrid zones reflects regional differences in pre and post-zygotic barriers to gene exchange. *Journal of Evolutionary Biology*, 18(6), 1416-1424.  
<https://doi.org/10.1111/j.1420-9101.2005.00964.x>
- Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2016). A spatial framework for understanding population structure and admixture. *PLoS Genetics*, 12(1), e1005703.  
<https://doi.org/10.1371/journal.pgen.1005703>
- Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), 378-400.  
<https://doi.org/10.32614/RJ-2017-066>
- Buckley, L. B., Nufio, C. R., Kirk, E. M., & Kingsolver, J. G. (2015). Elevational differences in developmental plasticity determine phenological responses of grasshoppers to recent climate warming. *Proceedings of the Royal Society B: Biological Sciences*, 282(1809), 20150441. <https://doi.org/10.1098/rspb.2015.0441>
- Butlin, R. K., & Smadja, C. M. (2018). Coupling, reinforcement, and speciation. *The American Naturalist*, 191(2), 155-172. <https://doi.org/10.1086/695136>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), s13742-015. <https://doi.org/10.1186/s13742-015-0047-8>
- Chown, S. L., & Gaston, K. J. (2010). Body size variation in insects: a macroecological perspective. *Biological Reviews*, 85(1), 139-169.  
<https://doi.org/10.1111/j.1469-185X.2009.00097.x>

- Commission for Environmental Cooperation (CEC). (2023). *2020 Land Cover of North America at 30 Meters*. North American Land Change Monitoring System. Canada Centre for Remote Sensing (CCRS), U.S. Geological Survey (USGS), Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Comisión Nacional Forestal (CONAFOR), Instituto Nacional de Estadística y Geografía (INEGI). Ed. 1.0, Raster digital data [30-m]. Available at <http://www.cec.org/north-american-environmental-atlas/land-cover-30m-2020/>
- Coppel, H., & Benjamin, D. (1965). Bionomics of the Nearctic pine-feeding diprionids. *Annual Review of Entomology*, 10, 69–96.  
<https://doi.org/10.1146/annurev.en.10.010165.000441>
- Coppi, A., Baker, A. J., Bettarini, I., Colzi, I., Echevarria, G., Pazzagli, L., ... & Selvi, F. (2020). Population genetics of *Odontarrhena* (Brassicaceae) from Albania: the effects of anthropic habitat disturbance, soil, and altitude on a Ni-hyperaccumulator plant group from a major serpentine hotspot. *Plants*, 9(12), 1686.  
<https://doi.org/10.3390/plants9121686>
- Croxall, J. P., Trathan, P. N., & Murphy, E. (2002). Environmental change and Antarctic seabird populations. *Science*, 297(5586), 1510-1514.  
<https://doi.org/10.1126/science.1071987>
- Cutter, A. D. (2012). The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. *Trends in Ecology & Evolution*, 27(4), 209-218.  
<https://doi.org/10.1016/j.tree.2011.11.004>
- Dagilis, A. J., Peede, D., Coughlan, J. M., Jofre, G. I., D'Agostino, E. R., Mavengere, H., ... & Matute, D. R. (2022). A need for standardized reporting of introgression: Insights from studies across eukaryotes. *Evolution Letters*, 6(5), 344-357.  
<https://doi.org/10.1002/evl3.294>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A.,

- Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008.  
<https://doi.org/10.1093/gigascience/giab008>
- Davidowitz, G., & Nijhout, H. F. (2004). The physiological basis of reaction norms: the interaction among growth rate, the duration of growth and body size. *Integrative and Comparative Biology*, 44(6), 443-449. <https://doi.org/10.1093/icb/44.6.443>
- De Meester, L., Stoks, R., & Brans, K. I. (2018). Genetic adaptation as a biological buffer against climate change: Potential and limitations. *Integrative Zoology*, 13(4), 372-391. <https://doi.org/10.1111/1749-4877.12298>
- DeVos, T. B., Bock, D. G., & Kolbe, J. J. (2023). Rapid introgression of non-native alleles following hybridization between a native *Anolis* lizard species and a cryptic invader across an urban landscape. *Molecular Ecology*, 32(11), 2930-2944.  
<https://doi.org/10.1111/mec.16897>
- Dickerson, J. (2002). Plant Fact Sheet: Eastern White Pine. *USDA Natural Resources Conservation Service*.
- Dewitz, J. (2021). National land cover database (NLCD) 2019 products. *US Geological Survey*, 10. <https://doi.org/10.5066/P9KZCM54>.
- Dopman, E. B., Robbins, P. S., & Seaman, A. (2010). Components of reproductive isolation between North American pheromone strains of the European corn borer. *Evolution*, 64(4), 881-902. <https://doi.org/10.1111/j.1558-5646.2009.00883.x>
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28(8), 2239-2252. <https://doi.org/10.1093/molbev/msr048>
- Fachrul, M., Karkey, A., Shakya, M., Judd, L. M., Harshegyi, T., Sim, K. S., ... & Inouye, M. (2023). Direct inference and control of genetic population structure from RNA sequencing data. *Communications Biology*, 6(1), 804.  
<https://doi.org/10.1038/s42003-023-05171-9>
- Faske, T. M., Agneray, A. C., Jahner, J. P., Osuna-Mascaró, C., Sheta, L. M., Richardson, B. A., ... & Parchman, T. L. (2024). Environment predicts the maintenance of

- reproductive isolation in a mosaic hybrid zone of rubber rabbitbrush. *Evolution*, 78(2), 300-314. <https://doi.org/10.1093/evolut/qpad207>
- Finney, B. P., Gregory-Eaves, I., Sweetman, J., Douglas, M. S., & Smol, J. P. (2000). Impacts of climatic change and fishing on Pacific salmon abundance over the past 300 years. *Science*, 290(5492), 795-799. <https://doi.org/10.1126/science.290.5492.795>
- Fisher, H. S., Wong, B. B., & Rosenthal, G. G. (2006). Alteration of the chemical environment disrupts communication in a freshwater fish. *Proceedings of the Royal Society B: Biological Sciences*, 273(1591), 1187-1193. <https://doi.org/10.1098/rspb.2005.3406>
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology*, 27(9), 2215-2233. <https://doi.org/10.1111/mec.14584>
- Garroway, C. J., Bowman, J., Cascaden, T. J., Holloway, G. L., Mahan, C. G., Malcolm, J. R., ... & Wilson, P. J. (2010). Climate change induced hybridization in flying squirrels. *Global Change Biology*, 16(1), 113-121. <https://doi.org/10.1111/j.1365-2486.2009.01948.x>
- Glover, A. N., & Linnen, C. R. (2025). Data from: Recent climate change and historical population structure predict spatial patterns of admixture between two host-specialized pine sawfly species. *Dryad*. doi: 10.5061/dryad.kh18932jw
- Glover, A. N., Bendall, E. E., Terbot, J. W., II, Payne, N., Webb, A., Filbeck, A., Norman, G., & Linnen, C. R. (2023). Body size as a magic trait in two plant-feeding insect species. *Evolution*, 77(2), 437-453. <https://doi.org/10.1093/evolut/qpac053>
- Glover, A. N., Sousa, V. C., Ridenbaugh, R. D., Sim, S. B., Geib, S. M., & Linnen, C. R. (2024). Recurrent selection shapes the genomic landscape of differentiation between a pair of host-specialized haplodiploids that diverged with gene flow. *Molecular Ecology*, 00, e17509. <https://doi.org/10.1111/mec.17509>
- Grabenstein, K. C., & Taylor, S. A. (2018). Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. *Trends in Ecology & Evolution*, 33(3), 198-212. <https://doi.org/10.1016/j.tree.2017.12.008>

- Grabenstein, K. C., Otter, K. A., Burg, T. M., & Taylor, S. A. (2023). Hybridization between closely related songbirds is related to human habitat disturbance. *Global Change Biology*, 29(4), 955-968. <https://doi.org/10.1111/gcb.16476>
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... & Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, 328(5979), 710-722. <https://doi.org/10.1126/science.1188021>
- Greenway, R., Drexler, S., Arias-Rodriguez, L., & Tobler, M. (2016). Adaptive, but not condition-dependent, body shape differences contribute to assortative mating preferences during ecological speciation. *Evolution*, 70(12), 2809–2822. <https://doi.org/10.1111/evo.13087>
- Hallett, T. B., Coulson, T., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Grenfell, B. T. (2004). Why large-scale climate indices seem to predict ecological processes better than local weather. *Nature*, 430(6995), 71-75. <https://doi.org/10.1038/nature02708>
- Hamlin, J. A., & Arnold, M. L. (2014). Determining population structure and hybridization for two iris species. *Ecology and Evolution*, 4(6), 743-755. <https://doi.org/10.1002/ece3.964>
- Harper, K. E., Bagley, R. K., Thompson, K. L., & Linnen, C. R. (2016). Complementary sex determination, inbreeding depression and inbreeding avoidance in a gregarious sawfly. *Heredity*, 117(5), 326–335. <https://doi.org/10.1038/hdy.2016.46>
- Hasselman, D. J., Argo, E. E., McBride, M. C., Bentzen, P., Schultz, T. F., Perez-Umphrey, A. A., & Palkovacs, E. P. (2014). Human disturbance causes the formation of a hybrid swarm between two naturally sympatric fish species. *Molecular Ecology*, 23(5), 1137-1152. <https://doi.org/10.1111/mec.12674>
- Hernandez, M., & Suni, S. S. (2022). Intersecting effects of landscape and body size on dispersal in bee populations. *Authorea Preprints*. <https://doi.org/10.22541/au.166373877.75451168/v1>
- Herrig, D. K., Ridenbaugh, R. D., Vertacnik, K. L., Everson, K. M., Sim, S. B., Geib, S. M.,

- Weisrock, D. W., & Linnen, C. R. (2024). Whole genomes reveal evolutionary relationships and mechanisms underlying gene-tree discordance in *Neodiprion* sawflies. *Systematic Biology*, syae036. <https://doi.org/10.1093/sysbio/syae036>
- Hibbins, M. S., & Hahn, M. W. (2019). The timing and direction of introgression under the multispecies network coalescent. *Genetics*, 211(3), 1059-1073. <https://doi.org/10.1534/genetics.118.301831>
- Hijmans, R. (2024). terra: Spatial Data Analysis, R package version 1.7-71. <https://CRAN.R-project.org/package=terra>. Accessed August 2024.
- Horreo, J. L., Machado-Schiaffino, G., Ayllon, F., Griffiths, A. M., Bright, D., Stevens, J. R., & Garcia-Vazquez, E. (2011). Impact of climate change and human-mediated introgression on southern European Atlantic salmon populations. *Global Change Biology*, 17(5), 1778-1787. <https://doi.org/10.1111/j.1365-2486.2010.02350.x>
- Hoskin, C. J., Higgie, M., McDonald, K. R., & Moritz, C. (2005). Reinforcement drives rapid allopatric speciation. *Nature*, 437(7063), 1353-1356. <https://doi.org/10.1038/nature04004>
- IPCC. (2007). *Climate Change 2007: Synthesis Report*. International Panel of Climate Change, Cambridge University Press, Cambridge.
- Johnson, M. T., & Munshi-South, J. (2017). Evolution of life in urban environments. *Science*, 358(6363), eaam8327. <https://doi.org/10.1126/science.aam8327>
- Jones, A., Moore, G., Kvarnemo, C., Walker, D., & Avise, J. (2003). Sympatric speciation as a consequence of male pregnancy in sea-horses. *Proceedings of the National Academy of Sciences*, 100(11), 6598–6603. <https://doi.org/10.1073/pnas.1131969100>
- Knerer, G., & Atwood, C. E. (1973). Diprionid sawflies: Polymorphism and speciation. *Science*, 179(4078), 1090–1099. <https://doi.org/10.1126/science.179.4078.1090>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, 15(5), 1179-1191. <https://doi.org/10.1111/1755-0998.12387>

- Kreiner, J. M., Latorre, S. M., Burbano, H. A., Stinchcombe, J. R., Otto, S. P., Weigel, D., & Wright, S. I. (2022). Rapid weed adaptation and range expansion in response to agriculture over the past two centuries. *Science*, 378(6624), 1079-1085.  
<https://doi.org/10.1126/science.abo7293>
- Larson, E. L., Tinghitella, R. M., & Taylor, S. A. (2019). Insect hybridization and climate change. *Frontiers in Ecology and Evolution*, 7, 348.  
<https://doi.org/10.3389/fevo.2019.00348>
- Lawson, D. J., Van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9(1), 3258.  
<https://doi.org/10.1038/s41467-018-05257-7>
- Lefouili, M., & Nam, K. (2022). The evaluation of Bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. *Scientific Reports*, 12(1), 11331. <https://doi.org/10.1038/s41598-022-15563-2>
- Lewanski, A. L., Golcher-Benavides, J., Rick, J. A., & Wagner, C. E. (2022). Variable hybridization between two Lake Tanganyikan cichlid species in recent secondary contact. *Molecular Ecology*, 31(19), 5041-5059. <https://doi.org/10.1111/mec.16636>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.  
<https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.  
<https://doi.org/10.1093/bioinformatics/btp352>
- Lindstedt, C., Bagley, R. K., Calhim, S., Jones, M., & Linnen, C. R. (2022). The impact of life stage and pigment source on the evolution of novel warning signal traits. *Evolution*, 76(3), 554-572. <https://doi.org/10.1111/evo.14443>
- Linnen, C. R., & Farrell, B. D. (2008). Comparison of methods for species-tree inference in the sawfly genus *Neodiprion* (Hymenoptera: Diprionidae). *Systematic Biology*, 57(6), 876–890. <https://doi.org/10.1080/10635150802580949>

- Linnen, C. R., & Farrell, B. D. (2010). A test of the sympatric host race formation hypothesis in *Neodiprion* (Hymenoptera: Diprionidae). *Proceedings of the Royal Society B: Biological Sciences*, 277(1697), 3131–3138. <https://doi.org/10.1098/rspb.2010.0577>
- Liu, N., Zhao, H., Patki, A., Limdi, N. A., & Allison, D. B. (2011). Controlling population structure in human genetic association studies with samples of unrelated individuals. *Statistics and its Interface*, 4(3), 317-326.
- Lopez, B. E., Urban, D., & White, P. S. (2018). Testing the effects of four urbanization filters on forest plant taxonomic, functional, and phylogenetic diversity. *Ecological Applications*, 28(8), 2197-2205. <https://doi.org/10.1002/eap.1812>
- Lozier, J. D., Strange, J. P., & Heraghty, S. D. (2023). Whole genome demographic models indicate divergent effective population size histories shape contemporary genetic diversity gradients in a montane bumble bee. *Ecology and Evolution*, 13(2), e9778. <https://doi.org/10.1002/ece3.9778>
- MacDonald, Z. G., Snape, K. L., Roe, A. D., & Sperling, F. A. (2022). Host association, environment, and geography underlie genomic differentiation in a major forest pest. *Evolutionary Applications*, 15(11), 1749-1765. <https://doi.org/10.1111/eva.13466>
- Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite – fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21, 584-595. <https://doi.org/10.1111/1755-0998.13265>
- Mandeville, E. G., Hall Jr, R. O., & Buerkle, C. A. (2022). Ecological outcomes of hybridization vary extensively in *Catostomus* fishes. *Evolution*, 76(11), 2697-2711. <https://doi.org/10.1111/evo.14624>
- Mandeville, E. G., Parchman, T. L., McDonald, D. B., & Buerkle, C. A. (2015). Highly variable reproductive isolation among pairs of *Catostomus* species. *Molecular Ecology*, 24(8), 1856-1872. <https://doi.org/10.1111/mec.13118>
- Martin, N. H., & Willis, J. H. (2010). Geographical variation in postzygotic isolation and its genetic basis within and between two *Mimulus* species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1552), 2469-2478.

- <https://doi.org/10.1098/rstb.2010.0030>
- Martin, S. H., Davey, J. W., & Jiggins, C. D. (2014). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, 32(1), 244-257. <https://doi.org/10.1093/molbev/msu269>
- Matute, D. R., Gavin-Smyth, J., & Liu, G. (2014). Variable post-zygotic isolation in *Drosophila melanogaster*/*D. simulans* hybrids. *Journal of Evolutionary Biology*, 27(8), 1691-1705. <https://doi.org/10.1111/jeb.12422>
- McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the true extent of introgression during anthropogenic hybridization. *Trends in Ecology & Evolution*, 34(4), 315-326. <https://doi.org/10.1016/j.tree.2018.12.013>
- Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21(12), 2839-2846. <https://doi.org/10.1111/j.1365-294X.2012.05578.x>
- Muhlfeld, C. C., Kovach, R. P., Al-Chokhachy, R., Amish, S. J., Kershner, J. L., Leary, R. F., ... & Allendorf, F. W. (2017). Legacy introductions and climatic variation explain spatiotemporal patterns of invasive hybridization in a native trout. *Global Change Biology*, 23(11), 4663-4674. <https://doi.org/10.1111/gcb.13681>
- Muhlfeld, C. C., Kovach, R. P., Jones, L. A., Al-Chokhachy, R., Boyer, M. C., Leary, R. F., ... & Allendorf, F. W. (2014). Invasive hybridization in a threatened species is accelerated by climate change. *Nature Climate Change*, 4(7), 620-624. <https://doi.org/10.1038/nclimate2252>
- Nelson, T. D., MacDonald, Z. G., & Sperling, F. A. (2022). Moths passing in the night: Phenological and genomic divergences within a forest pest complex. *Evolutionary Applications*, 15(1), 166-180. <https://doi.org/10.1111/eva.13338>
- Norris, L. C., Main, B. J., Lee, Y., Collier, T. C., Fofana, A., Cornel, A. J., & Lanzaro, G. C. (2015). Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proceedings of the National Academy of Sciences*, 112(3), 815-820. <https://doi.org/10.1073/pnas.1418892112>
- Ortego, J., Gugger, P. F., & Sork, V. L. (2017). Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian

- oak species. *New Phytologist*, 213(2), 942-955. <https://doi.org/10.1111/nph.14182>
- Östrand, F., Anderbrant, O., Jönsson, P., & Lytikäinen-Saarenmaa, P. (2001). Capture rates of the European pine sawfly, *Neodiprion sertifer*, in pheromone traps, with special regard to effects of wind speed. *Journal of Chemical Ecology*, 27, 1561-1574. <https://doi.org/10.1023/A:1010450006518>
- Oziolor, E. M., Reid, N. M., Yair, S., Lee, K. M., Guberman VerPloeg, S., Bruns, P. C., ... & Matson, C. W. (2019). Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science*, 364(6439), 455-457. <https://doi.org/10.1126/science.aav4155>
- Pebesma, E. J. (2018). Simple features for R: standardized support for spatial vector data. *The R Journal*, 10(1), 439-446. <https://doi.org/10.32614/RJ-2018-009>
- Pebesma, E., & Bivand, R. (2023). *Spatial data science: With applications in R*. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429459016>
- Poikela, N., Kinnunen, J., Wurdack, M., Kauranen, H., Schmitt, T., Kankare, M., ... & Hoikkala, A. (2019). Strength of sexual and postmating prezygotic barriers varies between sympatric populations with different histories and species abundances. *Evolution*, 73(6), 1182-1199. <https://doi.org/10.1111/evo.13732>
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904-909. <https://doi.org/10.1038/ng1847>
- Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Review Genetics*, 11(7), 459-463. <https://doi.org/10.1038/nrg2813>
- Privé, F., Luu, K., Blum, M. G., McGrath, J. J., & Vilhjálmsson, B. J. (2020). Efficient toolkit implementing best practices for principal component analysis of population genetic data. *Bioinformatics*, 36(16), 4449-4457. <https://doi.org/10.1093/bioinformatics/btaa520>
- Qvarnström, A., Ålund, M., McFarlane, S. E., & Sirkiä, P. M. (2016). Climate adaptation and

- speciation: particular focus on reproductive barriers in *Ficedula* flycatchers. *Evolutionary Applications*, 9(1), 119-134. <https://doi.org/10.1111/eva.12276>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Reutimann, O., Dauphin, B., Baltensweiler, A., Gugerli, F., Kremer, A., & Rellstab, C. (2023). Abiotic factors predict taxonomic composition and genetic admixture in populations of hybridizing white oak species (*Quercus* sect. *Quercus*) on regional scale. *Tree Genetics & Genomes*, 19(3), 22. <https://doi.org/10.1007/s11295-023-01598-7>
- Rossetto, M., Thurlby, K. A., Offord, C. A., Allen, C. B., & Weston, P. H. (2011). The impact of distance and shifting temperature gradient on genetic connectivity across a heterogeneous landscape. *BMC Evolutionary Biology*, 11, 1-11. <https://doi.org/10.1186/1471-2148-11-126>
- Rougemont, Q., Gaigher, A., Lasne, E., Côte J., Coke, M., Besnard, A.-L., Launey, S., & Evanno, G. (2015). Low reproductive isolation and highly variable levels of gene flow reveal limited progress towards speciation between European river and brook lampreys. *Journal of Evolutionary Biology*, 28(12), 2248–2263. <https://doi.org/10.1111/jeb.12750>
- Seifert, B., Kulmuni, J., & Pamilo, P. (2010). Independent hybrid populations of *Formica polyctena* X *rufa* wood ants (Hymenoptera: Formicidae) abound under conditions of forest fragmentation. *Evolutionary Ecology*, 24, 1219-1237. <https://doi.org/10.1007/s10682-010-9371-8>
- Servedio, M. R., & Noor, M. A. (2003). The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, 34(1), 339-364. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132412>
- Sianta, S. A., Moeller, D. A., & Brandvain, Y. (2024). The extent of introgression between incipient *Clarkia* species is determined by temporal environmental variation and mating system. *Proceedings of the National Academy of Sciences*, 121(12), e2316008121. <https://doi.org/10.1073/pnas.2316008121>

- Stillwell, R. C. (2010). Are latitudinal clines in body size adaptive?. *Oikos*, 119(9), 1387–1390. <https://doi.org/10.1111/j.1600-0706.2010.18670.x>
- Taylor, R. S., & Friesen, V. L. (2017). The role of allochry in speciation. *Molecular Ecology*, 26(13), 3330–3342. <https://doi.org/10.1111/mec.14126>
- Terribile, L. C., Olalla-Tárraga, M., Diniz-Filho, J. A. F., & Rodríguez, M. A. (2009). Ecological and evolutionary components of body size: Geographic variation of venomous snakes at the global scale. *Biological Journal of the Linnean Society*, 98(1), 94–109. <https://doi.org/10.1111/j.1095-8312.2009.01262.x>
- Theobald, E. J., Breckheimer, I., & HilleRisLambers, J. (2017). Climate drives phenological reassembly of a mountain wildflower meadow community. *Ecology*, 98(11), 2799–2812. <https://doi.org/10.1002/ecy.1996>
- Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., Kao, S., & Wilson, B. E. (2022). Daymet: Monthly Climate Summaries on a 1-km Grid for North America, Version 4 R1. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2131>
- Tobin, P. C., Nagarkatti, S., Loeb, G., & Saunders, M. C. (2008). Historical and projected interactions between climate change and insect voltinism in a multivoltine species. *Global Change Biology*, 14(5), 951–957. <https://doi.org/10.1111/j.1365-2486.2008.01561.x>
- Walsh, J., Rowe, R. J., Olsen, B. J., Shriver, W. G., & Kovach, A. I. (2016). Genotype-environment associations support a mosaic hybrid zone between two tidal marsh birds. *Ecology and Evolution*, 6(1), 279–294. <https://doi.org/10.1002/ece3.1864>
- Walsh, J., Billerman, S. M., Butcher, B. G., Rohwer, V. G., Toews, D. P., Vila-Coury, V., & Lovette, I. J. (2023). A complex genomic architecture underlies reproductive isolation in a North American oriole hybrid zone. *Communications Biology*, 6(1), 154. <https://doi.org/10.1038/s42003-023-04532-8>
- Wendel, G. W., & Smith, H. C. (1990). Eastern white pine (*Pinus strobus* L.). *Silvics of North America*, 1, 476–488.
- Wiens, B. J., & Colella, J. P. (2025). That’s Not a Hybrid: How to Distinguish Patterns of

- Admixture and Isolation By Distance. *Molecular Ecology Resources*, 25(3), e14039.  
<https://doi.org/10.1111/1755-0998.14039>
- Wiens, B. J., DeCicco, L. H., & Colella, J. P. (2025). triangulaR: An R package for identifying  
AIMs and building triangle plots using SNP data from hybrid zones. *Heredity*, 1-12.  
<https://doi.org/10.1038/s41437-025-00760-2>
- Wilson, A. A., Ditmer, M. A., Barber, J. R., Carter, N. H., Miller, E. T., Tyrrell, L. P., & Francis,  
C. D. (2021). Artificial night light and anthropogenic noise interact to influence bird  
abundance over a continental scale. *Global Change Biology*, 27(17), 3987-4004.  
<https://doi.org/10.1111/gcb.15663>
- Wilson, L. F., Wilkinson, R. C., & Averill, R. D. (1992). *Redheaded pine sawfly: Its ecology  
and management*. U.S. Dept. of Agriculture, Forest Service.
- Winchell, K. M., Campbell-Staton, S. C., Losos, J. B., Revell, L. J., Verrelli, B. C., & Geneva,  
A. J. (2023). Genome-wide parallelism underlies contemporary adaptation in urban  
lizards. *Proceedings of the National Academy of Sciences*, 120(3), e2216789120.  
<https://doi.org/10.1073/pnas.2216789120>
- Young, A. S. (2024). Genome-wide association studies have problems due to confounding:  
Are family-based designs the answer?. *PLoS Biology*, 22(4), e3002568.  
<https://doi.org/10.1371/journal.pbio.3002568>
- Zinck, J. W., & Rajora, O. P. (2016). Post-glacial phylogeography and evolution of a wide-  
ranging highly-exploited keystone forest tree, eastern white pine (*Pinus strobus*) in  
North America: single refugium, multiple routes. *BMC Evolutionary Biology*, 16, 1-17.  
<https://doi.org/10.1186/s12862-016-0624-1>

#### **DATA ACCESSIBILITY AND BENEFIT SHARING STATEMENT**

Raw *Neodiprion lecontei* and *Neodiprion pinetum* whole-genome resequencing  
reads are published in the NCBI SRA database (accession number PRJNA1221225). All  
input files and scripts required for reproducing the analyses within the manuscript are  
available on Dryad (<https://doi.org/10.5061/dryad.kh18932jw>). The environmental data  
were derived from resources available in the public domain: land cover in the United States

(<https://www.mrlc.gov/viewer/>) and Canada (<http://www.cec.org/north-american-environmental-atlas/land-cover-30m-2020/>); monthly average daily maximum temperature in the United States and Canada ([https://daac.ornl.gov/cgi-bin/dsvviewer.pl?ds\\_id=2131](https://daac.ornl.gov/cgi-bin/dsvviewer.pl?ds_id=2131)).

Benefits Generated: Benefits from this study accrue from the sharing of our data and results on public databases as described above.

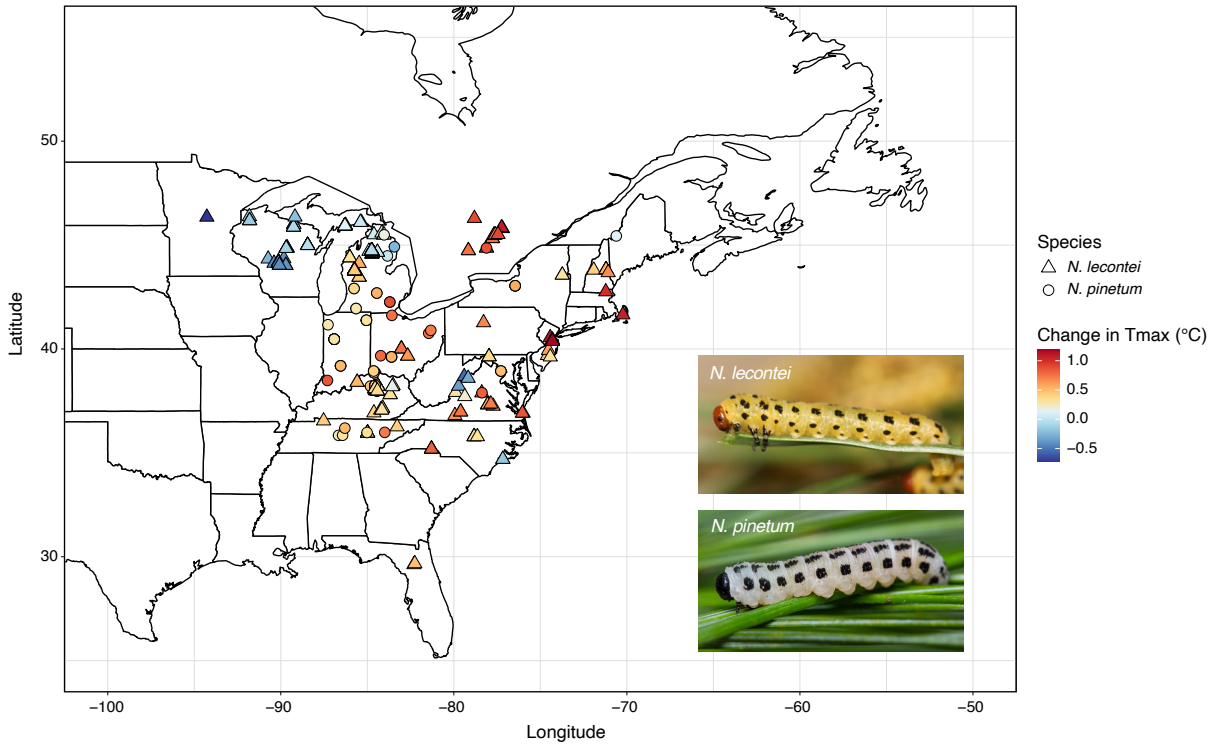
### **AUTHOR CONTRIBUTIONS**

ANG prepared DNA extractions for the resequencing dataset. ANG and CRL conceptualized and designed the study, performed data analysis, and wrote the manuscript.

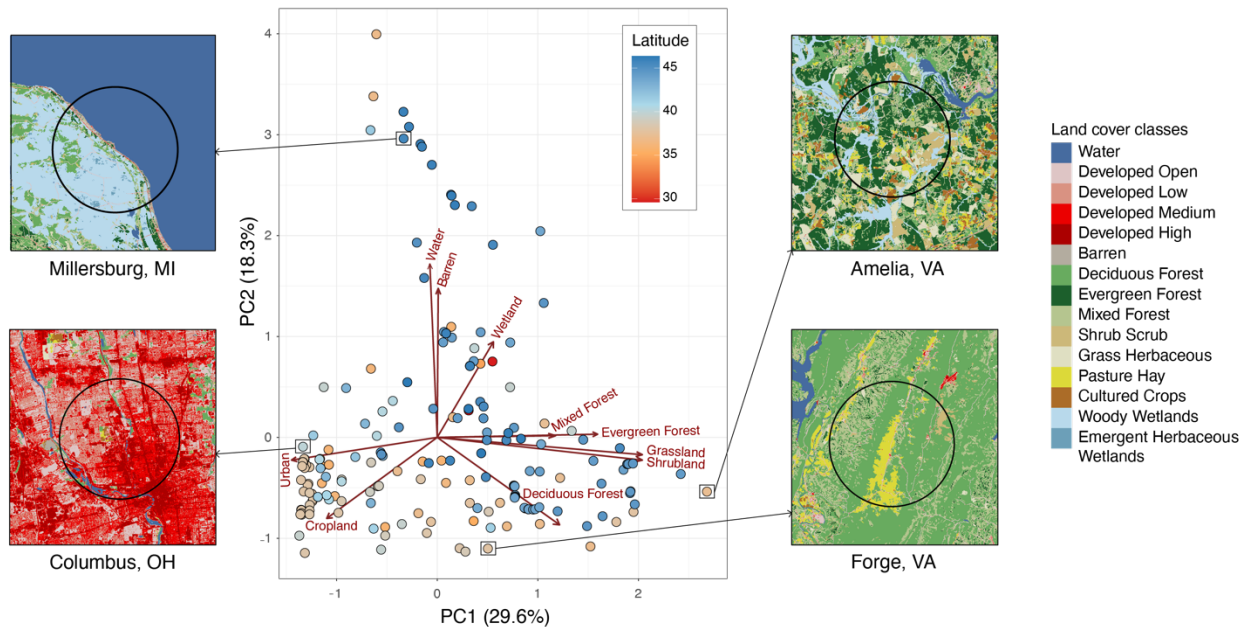
**Table 1. Results from multiple linear regression testing for relationships between admixture proportion and human disturbance in *Neodiprion lecontei* and *N. pinetum*.** All models used the beta family with a logit link and a zero-inflation formula. All predictor variables were normal-quantile transformed prior to running each model to ensure that they were on the same scale. Each model differed in approach used to account for population structure within each species: genetic PC scores included as model covariates (M1), grouped site number included as a random effect (M2), or no control for population structure (M3). Abbreviations: lcPC1, land cover PC1; lcPC2, land cover PC2; CTmax, change in average daily maximum temperature; genPC1, genetic PC1; genPC2, genetic PC2. Significant *p*-values (*p* < 0.05) are indicated in bold.

Model	Species	Predictor	Estimate	SE	z value	<i>p</i> -value
<b>M1: Genetic PC scores included as covariates</b>						
	<i>N. lecontei</i>	<b>(Intercept)</b>	-4.60	0.22	-20.97	<b>&lt; 2 x 10<sup>-16</sup></b>
		lcPC1	-0.028	0.082	-0.34	0.73
		lcPC2	-0.097	0.052	-1.86	0.062
		CTmax	0.19	0.063	3.06	<b>2.21 x 10<sup>-3</sup></b>
		genPC1	1.24	0.34	3.59	<b>3.26 x 10<sup>-4</sup></b>
		genPC2	-0.14	0.11	-1.35	0.18
	<i>N. pinetum</i>	<b>(Intercept)</b>	-4.34	0.15	-28.75	<b>&lt; 2 x 10<sup>-16</sup></b>
		lcPC1	0.011	0.10	0.11	0.91
		lcPC2	-0.13	0.10	-1.24	0.22
		CTmax	0.38	0.11	3.36	<b>7.67 x 10<sup>-4</sup></b>
		genPC1	-0.25	0.18	-1.38	0.17
		genPC2	0.52	0.089	5.81	<b>6.26 x 10<sup>-9</sup></b>
<b>M2: Site number included as random effect</b>						
	<i>N. lecontei</i>	<b>(Intercept)</b>	-5.22	0.093	-56.31	<b>&lt; 2 x 10<sup>-16</sup></b>
		lcPC1	0.079	0.078	1.02	0.31
		lcPC2	-0.044	0.064	-0.69	0.49
		CTmax	0.27	0.072	3.70	<b>2.16 x 10<sup>-4</sup></b>
	<i>N. pinetum</i>	<b>(Intercept)</b>	-4.35	0.15	-28.39	<b>&lt; 2 x 10<sup>-16</sup></b>
		lcPC1	0.27	0.12	2.24	<b>0.025</b>
		lcPC2	-0.20	0.15	-1.34	0.18
		CTmax	0.36	0.16	2.25	<b>0.024</b>
<b>M3: No accounting for population structure</b>						
	<i>N. lecontei</i>	<b>(Intercept)</b>	-5.21	0.089	-58.73	<b>&lt; 2 x 10<sup>-16</sup></b>
		lcPC1	0.060	0.071	0.84	0.40
		lcPC2	-0.076	0.053	-1.43	0.15
		CTmax	0.29	0.065	4.46	<b>8.17 x 10<sup>-6</sup></b>
	<i>N. pinetum</i>	<b>(Intercept)</b>	-4.21	0.14	-30.77	<b>&lt; 2 x 10<sup>-16</sup></b>
		lcPC1	0.30	0.095	3.11	<b>1.86 x 10<sup>-3</sup></b>
		lcPC2	-0.21	0.12	-1.78	0.076
		CTmax	0.37	0.13	2.80	<b>5.12 x 10<sup>-3</sup></b>

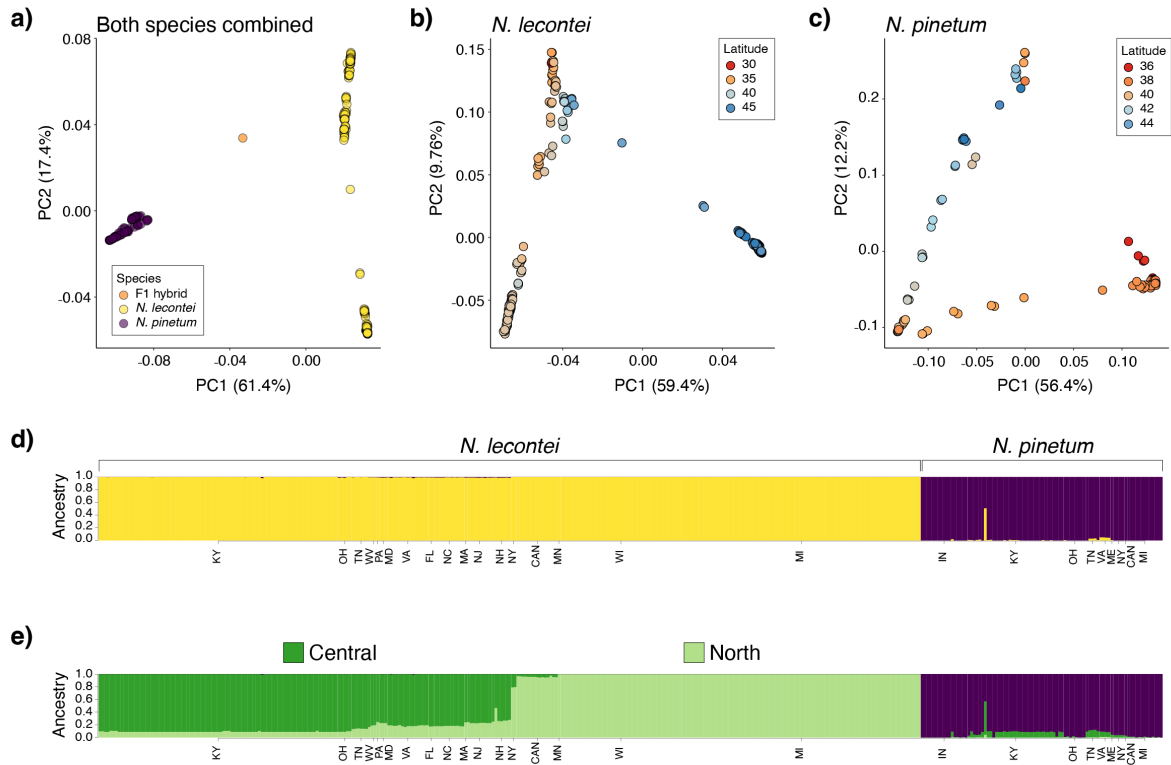
## FIGURES



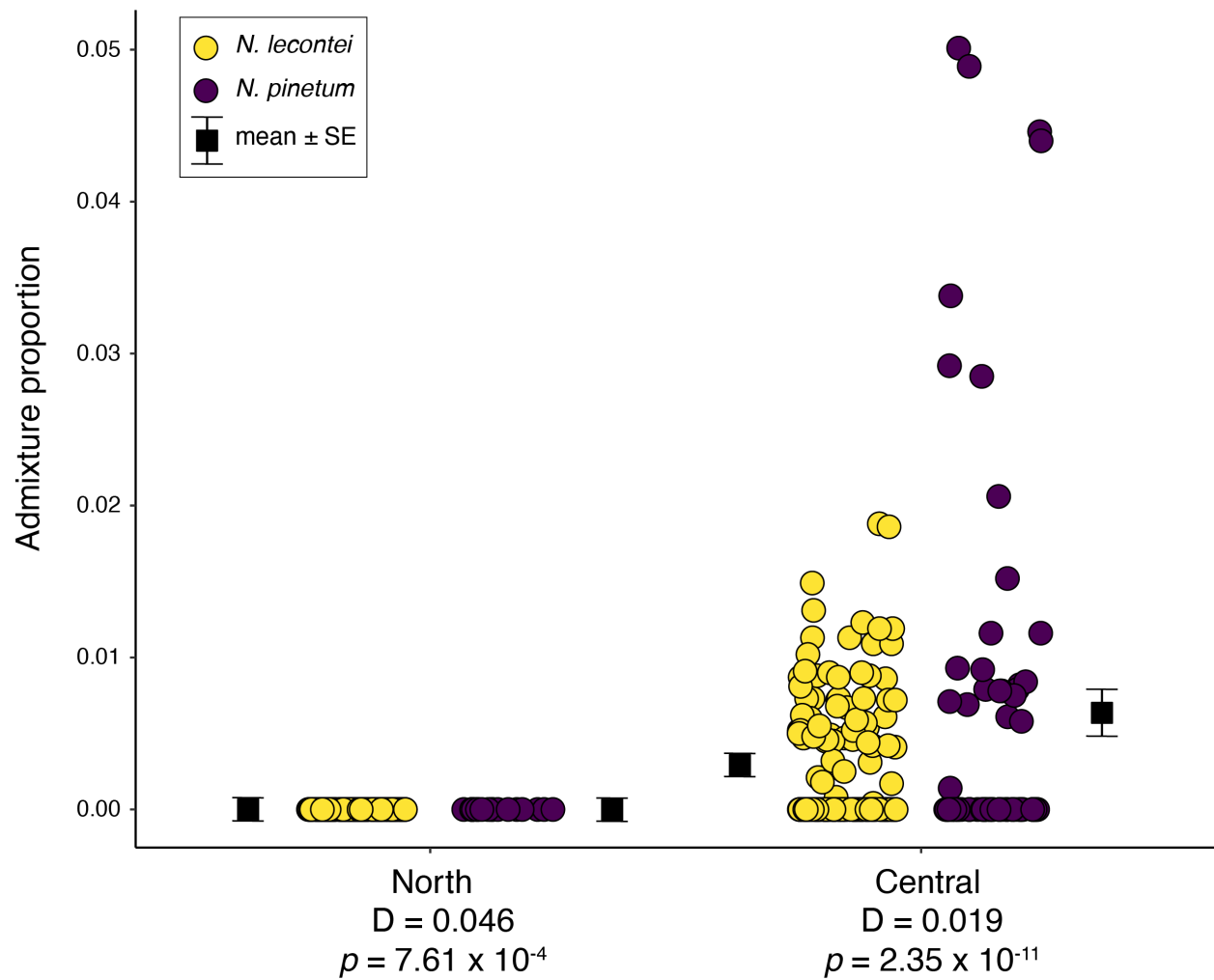
**Figure 1. Sampling map of *Neodiprion lecontei* and *N. pinetum*.** Map of sampling locations in eastern North America for *N. lecontei* and *N. pinetum* individuals used in this study. Each point represents a unique sampling site and is colored by the change in average daily maximum temperature from the 1980's to the 2010's (Change in Tmax) at that site (see Materials and Methods). Photographs of *N. lecontei* and *N. pinetum* larvae by Robin Bagley and Ryan Ridenbaugh.



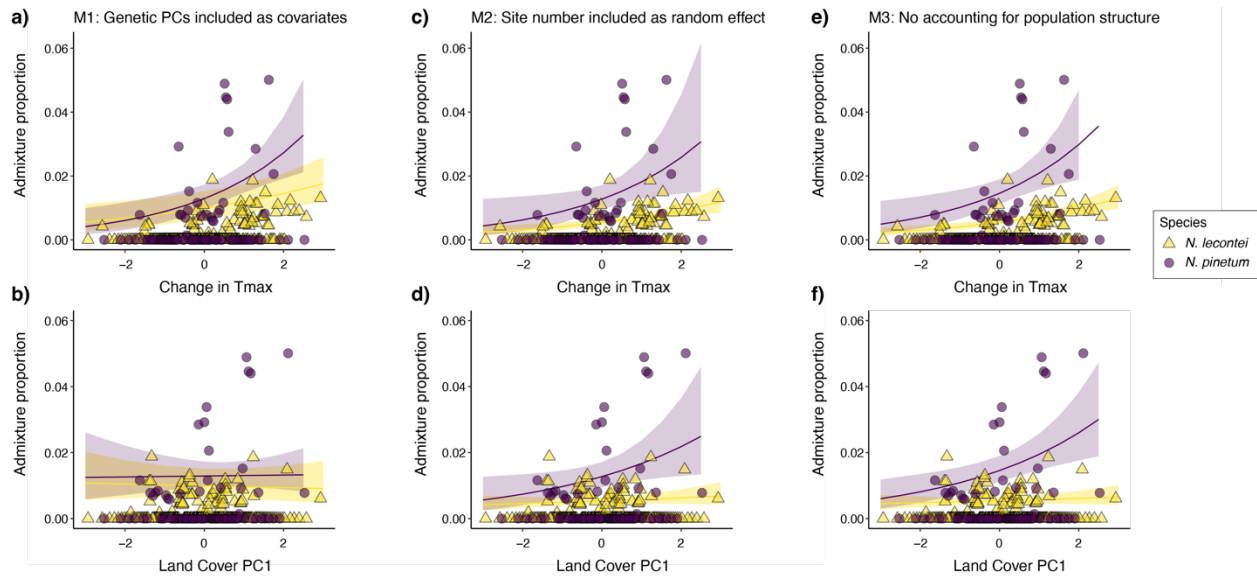
**Figure 2. Principal component analysis (PCA) of land cover proportions to quantify direct human disturbance at each sampling site.** Each point represents a sampling site and is colored by the latitude of the sampling site. For sampling sites within the United States, several land cover classes were grouped together (proportions added together) to be comparable with the Canadian land cover classes prior to performing the PCA: all four “Developed” land cover classes (see legend) were grouped as “Urban”; “Pasture Hay” and “Cultured Crops” were grouped as “Cropland”; and “Woody Wetlands” and “Emergent Herbaceous Wetlands” were grouped as “Wetland”. PC1 separates disturbed sites from “natural” sites while PC2 separates low vegetation sites from high vegetation sites. Land cover GIS layers for four sampling sites are provided to highlight extremes on each PC axis: highly disturbed (Columbus, OH) vs. “natural” (Amelia, VA) sites on PC1 and low vegetation (Millersburg, MI) vs. high vegetation (Forge, VA) sites on PC2. Within each land cover GIS layer, the black circle represents the 5km-radius buffer around the sampling site GPS coordinates, within which land cover proportions were calculated.



**Figure 3. Patterns of population structure in *Neodiprion lecontei* and *N. pinetum*.** Genetic principal component analysis (PCA) results for all individuals combined (a), *N. lecontei* individuals only (b), and *N. pinetum* individuals only (c). Each point represents an individual sawfly, and points are colored by either species (a) or the latitude the individual was sampled at (b, c). Ancestry coefficients from ADMIXTURE at (d) K=2 (representing species-level admixture) and at (e) K=3 (best supported K). Individuals within each species are grouped by sampling location (either US state or Canada (“CAN”)).



**Figure 4. Admixture proportions for *Neodiprion lecontei* and *N. pinetum* within each geographic region.** Each colored point represents an individual sawfly. The black squares represent the mean admixture proportion for each species within each geographic region with standard error bars. For both species, admixture proportions are lower in the North region (all zero) compared to the Central region. In contrast, both regions show evidence of introgression from the ABBA-BABA analysis (positive Patterson’s D statistic (D) that is significantly different from zero).



**Figure 5. Predictors of admixture proportion within *Neodiprion lecontei* and *N. pinetum* across models that vary in approach to account for population structure.** When genetic PC scores were included as model covariates (a, b), only change in average daily maximum temperature (Tmax) significantly predicted admixture proportion in both species (see Table 1). When site number was included as a random effect (c, d) and when no control for population structure was included (e, f), change in Tmax significantly predicted admixture proportion in both species and land cover PC1 significantly predicted admixture proportion in *N. pinetum* (see Table 1). Points represent individual sawflies. Trendlines with shaded 95% confidence intervals show predicted admixture proportions from multiple linear regression. Note: Although model results for both species are plotted together to aid with visualization, each of the three models was fit within each species separately.