Faster-haplodiploid evolution under divergence-with-gene-flow: simulations and empirical data from pine-feeding hymenopterans

Short Running Title: Adaptive divergence under haplodiploidy

Authors: Emily E. Bendall[1,2], Robin K. Bagley[1,3], Vitor C. Sousa[4]*, and Catherine R. Linnen*[1]

Affiliations:

[1] Department of Biology, University of Kentucky, Lexington, Kentucky 40506, USA

[2] Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan 48109, USA (present address)

[3] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University at Lima, Lima, OH 45804, USA (present address)

[4] CE3C – Centre for Ecology, Evolution and Environmental Changes, Department of Animal Biology, Faculdade de Ciências da Universidade de Lisboa, University of Lisbon, Campo Grande 1749-016 Lisboa, Portugal

*These authors contributed equally. Correspondence: vmsousa@fc.ul.pt; catherine.linnen@uky.edu

Keywords: faster-X, local adaptation, gene flow, speciation, haplodiploidy, genomic differentiation

Abstract

Although haplodiploidy is widespread in nature, the evolutionary consequences of this sex determination mechanism are not well characterized. Here, we examine how genome-wide hemizygosity and a lack of recombination in haploid males affects genomic differentiation in populations that diverge via natural selection while experiencing gene flow. First, we simulated diploid and haplodiploid "genomes" (500-kb loci) evolving under an isolation-with-migration model with mutation, drift, selection, migration, and recombination; and examined differentiation at neutral sites both tightly and loosely linked to a divergently selected site. So long as there is divergent selection and migration, sex-limited hemizygosity and recombination cause elevated differentiation (i.e., produce a "faster-haplodiploid effect") in haplodiploid populations relative to otherwise equivalent diploid populations, for both recessive and codominant mutations. Second, we used genome-wide SNP data to model divergence history and describe patterns of genomic differentiation between sympatric populations of *Neodiprion lecontei* and *N. pinetum*, a pair of pine sawfly species (order: Hymenoptera; family: Diprionidae) that are specialized on different pine hosts. These analyses support a history of continuous gene exchange throughout divergence and reveal a pattern of heterogeneous genomic differentiation that is consistent with divergent selection on many unlinked loci. Third, using simulations of haplodiploid and diploid populations evolving according to the estimated divergence history of *N. lecontei* and *N. pinetum*, we found that divergent selection would lead to higher differentiation in haplodiploids. Based on these results, we hypothesize that haplodiploids undergo divergence-with-gene-flow and sympatric speciation more readily than diploids.

Introduction

In terms of both species richness and biomass, haplodiploid organisms account for a substantial proportion of terrestrial biodiversity (Forbes et al., 2018; Hölldobler & Wilson, 1990). Haplodiploidy (arrhenotoky), a sex determination mechanism in which females develop from fertilized eggs and are diploid, while males develop from unfertilized eggs and are haploid, has evolved repeatedly in diverse arthropod lineages and is the mode of sex determination for an estimated 12% of extant animal species (Blackmon et al., 2017; de la Filia et al., 2015; Hedrick & Parker, 1997; Normark, 2003). From a theoretical perspective, most work on haplodiploidy has focused on the evolution of eusociality (Hamilton, 1964a, 1964b, 1972; Rautiala et al., 2019, but see de la Filia et al., 2015). However, haplodiploid transmission genetics can have many other important evolutionary consequences. For example, when haplodiploid populations hybridize, only female hybrids are produced in the first generation and hybrid males are produced in the subsequent generation. This asymmetry may lead to higher rates of mitochondrial introgression compared to nuclear introgression (Linnen & Farrell, 2007; Patten et al., 2015) and may have consequences for the evolution of postzygotic isolation (Bendall et al., 2020; see also Ghenu et al., 2018; Nouhaud et al., 2020). Haplodiploidy is also expected to impact the evolution of sexually selected traits (Kirkpatrick & Hall, 2004; Reeve & Pfennig, 2003), mating systems (Boulton et al., 2015; Werren, 1993), parental care (Davies & Gardner, 2014; Gardner, 2012), sex ratios (Hamilton, 1967), and the outcomes of intra- and inter-locus conflicts (Hitchcock, 2021; Kraaijeveld, 2009). However, formal theory and empirical tests for the evolutionary consequences of haplodiploidy remain rare (de la Filia et al., 2015).

Here, we focus here on how haplodiploidy affects genomic differentiation in diverging populations and species. Similarities in transmission genetics between haplodiploid genomes and

67 X (or Z) chromosomes make it possible to draw on faster-X theory to generate predictions for
68 haplodiploids (Avery, 1984; Hedrick & Parker, 1997; Kraaijeveld, 2009). Within populations,
69 hemizygosity in XY and haplodiploid males will expose recessive or partially recessive
70 mutations to selection, thereby hastening the removal of deleterious recessive alleles and the
71 fixation of beneficial recessive alleles (Avery, 1984; Charlesworth, et al., 1987; Hedrick &
72 Parker, 1997). More efficient selection on novel hemizygous alleles will also impact linked
73 variation via hitchhiking (Betancourt et al., 2004) and background selection (Charlesworth,
74 2012), and these effects will be exacerbated by a lack of recombination in XY and haplodiploid
75 males (Betancourt et al., 2004; Lester & Selander, 1979; Owen, 1986). So long as adaptation is
76 driven primarily by new mutations that are at least partially recessive, faster-X theory predicts
77 higher adaptive substitution rates and greater genetic divergence at linked sites on sex
78 chromosomes and haplodiploid genomes relative to diploid autosomes when populations or
79 species diverge in isolation (Presgraves 2018, but see Wright et al., 2015).

80 Conversely, faster-X theory under models of divergence-with-gene-flow suggests that
81 hemizygous selection can facilitate adaptive differentiation at a selected site regardless of
82 dominance and starting allele frequency (Lasne et al., 2017). Additionally, when formally
83 isolated populations come into secondary contact, sex-limited hemizygosity and recombination
84 can reduce effective migration rates at neutral loci linked to loci involved in local adaptation
85 and/or hybrid incompatibilities (Fraïsse & Sachdeva, 2021; Fusco & Uyenoyama, 2011;
86 Muirhead & Presgraves, 2016). So long as gene flow accompanies divergence, faster-X theory
87 predicts that sex chromosomes and haplodiploid genomes will exhibit greater differentiation at
88 selected and linked sites compared to autosomal chromosomes.

89 Consistent with faster-X theory, comparative and population genomic data from diverse
90 taxa suggest that faster-X effects (i.e., elevated differentiation, divergence, and substitution rates
91 on sex chromosomes) are widespread in nature (Irwin, 2018; Meisel & Connallon, 2013;
92 Presgraves, 2018). However, these patterns are not necessarily caused by sex-limited
93 recombination and hemizygosity. Indeed, there are many other differences between sex
94 chromosomes and autosomes that can also produce differences in genetic differentiation,
95 including: differences in effective population size (N_e), mutation rate, recombination rate, gene
96 content, and susceptibility to meiotic drive (Frank, 1991; Hurst & Pomiankowski, 1991;
97 Meiklejohn et al., 2018; Patten, 2018). Because they lack sex chromosomes, haplodiploids are
98 potentially powerful models for investigating the impact of sex-limited hemizygosity and
99 recombination on genomic differentiation independent of sex-chromosome specific factors.
100 However, because they also lack anything analogous to diploid autosomes, haplodiploids do not
101 have a built-in benchmark for evaluating whether they exhibit "faster-haplodiploid" effects,
102 which we define as greater differentiation or divergence in haplodiploids relative to comparable
103 diploids. Fortunately, increasingly sophisticated tools for simulating genomic datasets evolving
104 under complex demographic and ecological scenarios (Hoban et al., 2012; Haller & Messer
105 2019; Hart et al., 2021) offer a potential strategy for detecting faster-haplodiploid effects:
106 simulate a benchmark diploid dataset with equivalent demographic history, recombination, and
107 mutation under neutral and adaptive scenarios.

108 To better understand the impact of haplodiploidy on genomic differentiation, we combine
109 simulations of haplodiploid and diploid genomes evolving under divergence-with-gene-flow
110 with an empirical case study of a haplodiploid species pair for which we have extensive
111 knowledge regarding the drivers of divergent selection and reproductive isolation, as well as

basic life history knowledge to parameterize simulations. *Neodiprion pinetum* (white pine sawfly) and *N. lecontei* (redheaded pine sawfly) are sister species with overlapping distributions in eastern North American (Linnen & Farrell, 2008, 2010). Because both species are pests of economically important pines, their basic ecology and life history is well described (Benjamin, 1955; Rauf & Benjamin, 1980; Coppel & Benjamin, 1965; Knerer & Atwood, 1973; Wilson et al., 1992). Reproductive adults emerge in spring after overwintering as cocoons. Females fly to their preferred host and attract haploid males via a sex pheromone. Mating takes place on the host plant, and females use their saw-like ovipositor to embed their full complement of eggs within the needles of a single pine branch. Larvae emerge and feed on pine needles before dispersing to the soil to spin a cocoon.

While *N. pinetum* and *N. lecontei* share many similarities, *N. pinetum* feeds exclusively on white pine (*Pinus strobus*) and *N. lecontei* tends to avoid this host. Differences between their hosts likely generate divergent selection on many different larval and adult traits (Coppel & Benjamin, 1965; Codella & Raffa, 2002; Lindstedt et al., 2020; Bendall et al., 2017). For example, differences in needle chemistry and thickness between the preferred hosts of *N. lecontei* and *N. pinetum* are associated with differences in egg size, female ovipositor morphology, and female egg-laying behaviors. These traits, which together determine the reproductive success of adult females, act as an ecological barrier to gene exchange in sympatric populations (Bendall et al., 2017). This previous work suggests that many regions of the genome are likely to be under divergent selection between these species. Moreover, a coalescent-based analysis revealed evidence of historical mitochondrial introgression, suggesting that this species pair has diverged with gene flow (Linnen & Farrell, 2007).

134 We hypothesize that adaptation to different pines and speciation-with-gene-flow in *N.*
135 *lecontei* and *N. pinetum* was facilitated by sex-limited hemizygosity and recombination. To test
136 this hypothesis, we: (1) simulate diploid and haplodiploid "genomes" (500-kb loci) evolving
137 under mutation, drift, divergent selection, migration, and recombination; (2) model the
138 divergence history and characterize patterns of genomic differentiation in sympatric populations
139 of *N. lecontei* and *N. pinetum*; and (3) use our estimated divergence history and other system-
140 specific details to parameterize simulations of haplodiploid and diploid genomes evolving under
141 varying levels of selection. Our data support a faster-haplodiploid effect in *Neodiprion* sawflies,
142 and based on our results, we suggest that such effects have promoted adaptation and speciation in
143 haplodiploid taxa.

144

145 **Methods**

146 *Simulation of haplodiploid and diploid chromosomes under divergence with gene flow*

147 To evaluate the effects of hemizygous selection and sex-limited recombination on
148 genomic differentiation patterns, we simulated populations of diploid autosomes and
149 haplodiploid chromosomes (Figure 1). We used SLiM v3 (Haller & Messer, 2019) to simulate
150 500-kilobase (kb) chromosomes evolving via mutation, drift, migration, and selection. We
151 considered an isolation-with-migration model with two populations that diverged at some time
152 (t_{div}) from an ancestral population, with symmetric gene flow (Figure 1b). Simulations consisted
153 of two phases. First, to enable the ancestral population to reach mutation-drift equilibrium, we
154 simulated neutral evolution of an ancestral population with an effective size of 1500 ($2N_e$), a
155 mutation rate of 2.5×10^{-7}/bp/generation, and a recombination rate of 2.5×10^{-7}/bp/generation for

156 10,000 generations (>4N_e generations). Second, to simulate divergence-with-gene-flow, the
157 ancestral population splits into two equally sized populations (2N_e) that exchange migrants at a
158 constant and symmetrical migration rate ($m=m_{12}=m_{21}$). The timing of this split coincides with
159 the onset of divergent natural selection on a polymorphic site (initial frequency of derived allele
160 *a* denoted as q_0) located at the middle of the chromosome (250-kb). We modeled selection under
161 a "parallel dominance" fitness model in which the derived allele *a* is favored in population 1 and
162 allele *A* (ancestral allele) is favored in population 2, its dominance is the same irrespective of the
163 population (Figure 1a, as in Moran (1959); Lasne et al., (2017)). Our model assumes there are
164 separate sexes, with equal numbers of males and females, without sex-specific migration rates,
165 and similar distributions of offspring numbers in males and females. Our model also assumes the
166 fitness of hemizygous males (*A* or *a*) is equal to the fitness of corresponding homozygous
167 females (*AA* or *aa*). Following the onset of selection, populations evolve for an additional 2000
168 generations.

169 The parameter values above are identical to a scaled mutation rate ($\theta=4N_e\mu L$) and
170 recombination rate ($\rho=4N_erL$) of a 500-kb chromosome in a population with an effective size of
171 100,000 and a mutation rate of $2.5\text{x}10^{-9}$/ bp/generation. We used a smaller effective population
172 size of 1500 and scaled the other parameters accordingly to reduce the computational burden of
173 forward simulations, as is usually done when using SLIM (e.g., Phung et al., 2016). Our chosen
174 divergence time correspond to a divergence with a scaled time $T_{div}/(4N_e)=2/3$, which is within the
175 range of values estimated for pairs of closely related populations and species across many taxa
176 (Hey & Pinho, 2012; Pinho & Hey, 2010), but lower than the threshold of $T_{div}/(4N_e)>1$ (and
177 $2N_em<1$) proposed by Hey and Pinho (2012) as a diagnostic for fully independent species.

178 We simulated diverging populations under all possible combinations of seven selection
179 coefficients (scaled $2N_es$ ~ 0, 10, 20, 40, 80, 100, 200), four migration rates (scaled $2N_em$ ~ 0.0,
180 0.5, 2.5, 5.1), two dominance coefficients (recessive h=0.01 and codominant h=0.50), and four
181 different starting allele frequencies (q_0 = 1/($2N_e$), 0.01, 0.10, and 0.50; Supplementary Table S1).
182 These parameters were chosen to capture a range of selection coefficients and migration rates,
183 including the neutral case (s=0) and the no-migration case (m=0). The starting allele frequencies
184 ranged from new (q_0=1/($2N_e$)) or rare (q_0=0.01) mutations to common variants (q_0=0.10 and
185 0.50). To investigate the impact of recombination rate on linked variation we repeated a subset of
186 these conditions (with q_0=0.10 and 0.50) under a lower recombination rate (r=0.1μ= $2.5\times10^{-}$
187 ^{8}/bp/generation). For each unique combination of parameters, we performed 1000 simulations.

188 To control for factors other than hemizygous selection and sex-limited recombination that
189 might also cause differences in genomic differentiation between diploids and haplodiploids, our
190 simulations were scaled so that haplodiploid and diploid chromosomes experience equivalent
191 effective levels of drift, migration, and recombination (Supplementary Table S1). Thus, we
192 adjusted the N_e to ensure that both the diploid and haplodiploid chromosomes have $2N_e$=1500,
193 which is the N_e of a hemizygous locus with N=1000 individuals (500 females with two copies
194 and 500 males with one copy). This corresponds in the diploid case to N_D=750 individuals,
195 obtained as N_D=$x\,N$ individuals, where x is a scaling factor that is 3/4 for 0.50 sex-ratio
196 (Supplementary Methods). Because a haplodiploid chromosome spends 2/3 of the time in the sex
197 in which it recombines, we scaled the diploid recombination rate as 2/3 of that of haplodiploids
198 to ensure identical average recombination rates in diploids and haplodiploids (Kong et al., 2002;
199 Wilfert et al., 2007). To confirm our scaling, we verified that the values of several summary
200 statistics measuring diversity, differentiation and linkage disequilibrium were identical for

201 neutral simulations for haplodiploid and diploid chromosomes, and that they converged to the
202 expected values under neutrality (Supplementary Figure S1).

203 For each replicate, we followed the trajectories of allele frequencies at the selected site in
204 both populations, which were used to compute the number of simulations that retained the
205 derived allele *a* in population 1. To investigate patterns of variation in samples rather than at the
206 population level, in the last generation we sampled 20 chromosomes of 500 kb from each
207 population. For each parameter combination, we computed average nucleotide diversity, D_{xy} and
208 weighted F_{ST} across simulations using the Hudson estimator (Bhatia et al., 2013), averaged
209 across all SNPs at three scales: (1) a 20-kb window centered on the selected site ("20-kb"); (2)
210 across the 500-kb chromosome ("500-kb"); and (3) scan of contiguous non-overlapping 20-kb
211 windows ("genome scan"). To measure patterns of linkage disequilibrium, we computed average
212 r^2 between all pairs of SNPs within 20-kb windows. To evaluate the effect of loss of the derived
213 allele *a*, we used two approaches: in the "all simulations" approach, the mean of summary
214 statistics (e.g., F_{ST}) was computed across all simulations, whereas in the "conditional" approach,
215 the mean of summary statistics is computed only across simulations where the derived allele was
216 retained in population 1 (Figure 1c). Thus, the "conditional" approach removes the effect of
217 allele loss. Parameter combinations for which fewer than 10 simulations were retained were
218 treated as missing data.

219 *Empirical data: estimating divergence history in a haplodiploid species pair*

220 *Population sampling*

221 We sampled 23 *N. pinetum* larvae and 44 *N. lecontei* larvae from Kentucky
222 (Supplementary Table S2). Larvae tend to be found in gregarious colonies of siblings in both

223 species. To ensure we were not sampling close relatives, each individual was collected from a
224 different colony. To maximize our chances of sequencing diploid female larvae, which tend to be
225 larger than haploid male larvae, we extracted DNA from large larvae and verified sex with
226 heterozygosity estimates. We also sampled 3 suspected field-caught hybrids (based on
227 intermediate larval pigmentation) from Kentucky and one lab-reared female F_1 as a control
228 (Supplementary Table S2). An additional 18 *N. lecontei* samples from an allopatric population in
229 Michigan and 1 *N. virginiana* from Blackstone, VA (37°06'47.2"N, 78°01'37.4"W) were
230 sequenced for use in demographic analyses.

231 *DNA sequencing*

232 We extracted DNA using a CTAB/phenol-chloroform-isoamyl alcohol method (Chen,
233 Rangasamy, Tan, Wang, & Siegfried, 2010). We visualized the DNA on a 0.8% agarose gel to
234 confirm quality. To quantify the DNA, we used a Quant-iT High-Sensitivity DNA Assay Kit
235 (Invitrogen – Molecular Probes, Eugene, OR, USA). For *N. pinetum*, *N. lecontei*, and hybrids,
236 we used a modified ddRAD sequencing protocol from Bagley et al. (2017) and Peterson et al.
237 (2012). We fragmented the DNA using NlaIII and EcoRI. We assigned each individual along
238 with additional samples from other projects to one of eight libraries. During adapter ligation, we
239 assigned each sample one 48 unique in-line barcodes (Supplementary Table S2). We used the 5-
240 10 bp variable length barcodes used in Burford Reiskind et al. (2016). We then pooled each
241 group of samples and size selected for a 379-bp fragment (+/- 76bp) on a PippinPrep (Sage
242 Science, Beverly, MA). We did 12 rounds of high-fidelity PCR amplification (Phusion High-
243 Fidelity DNA Polymerase, NEB, Ipswich, MA) using PCR primers that included one of 12
244 unique Illumina multiplex read indices (Supplementary Table S2). To allow for the detection of
245 PCR duplicates, we included a string of 4 degenerate bases next to the Illumina read index

246 (Schweyen et al., 2014). We used a Bioanalyzer 2100 (Agilent, Santa Clara, CA) to check
247 library quality. The libraries were sequenced at the University of Illinois' Roy J. Carver
248 Biotechnology Center, using two lanes of Illumina HiSeq 4000 and150-bp single-end reads.

249 For *N. virginiana*, which we used as an outgroup, we used 150-PE reads generated on an
250 Illumina Nextseq at the Univeristy of Georgia Genomics Facility (Vertacnik, 2020). Library
251 preparation and whole-genome shotgun sequencing were both completed at the sequencing
252 facility. We removed the adapters using cutadapt 1.16 and contaminants using the standard and
253 pine databases in Kraken (Martin, 2011; Wood & Salzberg, 2014).

254 *DNA processing and variant calling*

255 We aligned demultiplexed ddRAD reads to the *N. lecontei* reference genome (Nlec1.1
256 GenBank assembly accession number- GCA_001263575.2; Linnen et al., 2018; Vertacnik &
257 Linnen, 2015) using the very sensitive setting in bowtie2 (Langmead & Salzberg, 2012). We
258 only retained reads that aligned to one locus in the reference genome and had a Phred score
259 greater than 30. For the ddRAD dataset, we removed PCR duplicates using a custom script. We
260 called SNPs in samtools (Li et al., 2009). Male and female larvae are morphologically
261 indistinguishable. To identify putative haploid males, which are expected to have unusually low
262 heterozygosity, we computed per-individual heterozygosity (as in Bagley et al., 2017). No
263 individuals were excluded based on heterozygosity. We required all sites to have a minimum of
264 7x coverage and 50% missing data or less. We also removed SNPs with significantly more
265 heterozygotes than expected under Hardy-Weinberg equilibrium (an indicator of
266 genotyping/mapping error). We removed any individual that was missing more than 70% of the
267 data. We performed all filtering in VCFtools v0.1.13 (Danecek et al., 2011).

268 We created several datasets with subsets of individuals and additional filtering for each of
269 the population genetic analyses. We generated three data sets with minor allele filtering (MAF,
270 SNPs <0.01 removed): 1) sympatric *N. pinetum* and *N. lecontei* for genome-wide patterns of
271 divergence (36,935 SNPs), 2) sympatric *N. pinetum*, *N. lecontei,* and hybrids for admixture
272 analysis (35,649 SNPs), and 3) sympatric *N. pinetum*, *N. lecontei,* allopatric *N. lecontei*, and
273 outgroup *N. virginiana* for ABBA- BABA tests (12,905 SNPs). We also generated a down-
274 sampled dataset (described below) without a MAF filter for estimating site-frequency spectra
275 (SFS) that included sympatric *N. pinetum*, *N. lecontei*, and *N. virginiana* for demographic
276 analyses.

277 *Population structure, demographic analysis, and genomic differentiation*

278 To confirm that our morphological hybrids were genetically admixed, we used
279 Admixture v1.3.0 (Alexander et al., 2009) to estimate the proportion of ancestry for each
280 individual collected in Kentucky (*N. lecontei, N. pinetum,* lab-reared hybrids, and suspected
281 field-caught hybrids) from *K* populations for *K* =1- 5. We ran 100 replicates per *K* and chose the
282 *K* with the lowest cross-validation (CV) score (Supplementary Table S3). Additionally, to test
283 for introgression between sympatric *N. lecontei* (P1) and *N. pinetum* (P3), we performed an
284 ABBA-BABA test (Patterson et al., 2012) with Kentucky *N. lecontei* (P1), Michigan *N. lecontei*
285 (allopatric population, P2), Kentucky *N. pinetum* (P3) and *N. virginiana* (outgroup, P4). We used
286 custom R script to compute the ABBA-BABA assuming that the outgroup is not fixed for the
287 ancestral allele (Patterson et al., 2012), assessing significance with block-jackknife resampling
288 diving data into 645 blocks of ~20 SNPs.

289 To evaluate the timing and magnitude of gene flow between *N. lecontei* and *N. pinetum*,
290 we performed demographic modeling based on the site frequency spectrum (SFS) using the

291 composite likelihood method implemented in fastsimcoal2 v2.6 (Excoffier et al., 2013). For this
292 analysis, we used ddRAD data from sympatric populations of *N. lecontei* and *N. pinetum* filtered
293 as described above, with additional filters applied to satisfy analysis assumptions. First, to
294 minimize the impact of selection on demographic history estimates, we used the NCBI
295 *Neodiprion lecontei* Annotation Release 100 (updated to GCA_001263575.2) to exclude SNPs
296 that were in or within 1 kb of the start or end of a gene, thereby generating a set of putatively
297 neutral markers. Furthermore, to reduce bias in the SFS, we applied more stringent depth-of-
298 coverage filters, requiring a minimum depth of 10x and a maximum depth less than 2x the
299 median depth of coverage per individual. To build the 2D-SFS without missing data, each
300 scaffold was divided into non-overlapping 50-kb blocks, and we kept only blocks where the
301 median distance between SNPs was larger than 2bp. SNPs without missing data were obtained
302 for each block by downsampling four and six females from *N. pinetum* and *N. lecontei*,
303 respectively. This resulted in a downsampled dataset with 9,994 SNPs. To polarize the
304 ancestral/derived state of alleles and obtain the unfolded 2D-SFS we used data from *N.*
305 *virginiana*. To obtain the number of invariant sites in the 2D-SFS we assumed that the proportion
306 of SNPs removed because of extra filters was the same for invariant sites. Given a proportion of
307 number of SNPs to number of invariant sites before extra filters of ~0.046, the number of
308 invariant sites in the 2D-SFS after filters was set to 215,283.

309 We tested five alternative demographic scenarios: (1) divergence without gene flow, (2)
310 divergence with continuous bidirectional migration, (3) divergence in isolation followed by a
311 single bout of secondary contact (unidirectional gene flow), (4) divergence with bidirectional
312 migration that stops before divergence is complete, and (5) divergence in isolation followed by
313 continuous secondary contact (bidirectional). All models except the model of continuous gene

314 flow had an equal number of parameters, so we compared their likelihoods directly. We ran each
315 model 100 times starting from different parameter combinations, each run with 50 optimization
316 cycles (-l50) and approximating the expected SFS with 100,000 coalescent simulations (-
317 n100000). We selected the run with the highest likelihood to estimate parameter values.

318 To examine genome-wide patterns of genetic divergence, we computed F_{ST} and π in 100-
319 kb non-overlapping windows for *N. lecontei* and *N. pinetum* in VCFtools on the non-down
320 sampled dataset. To identify regions of the genome that were more or less differentiated than
321 expected under neutrality, we simulated 10,000 datasets under the inferred demographic history
322 for sawflies using coalescent simulations implemented in the R package *scrm* (Staab et al.,
323 2015). For each simulation, we computed F_{ST} as for the *Neodiprion* dataset (see above). Outlier
324 windows were defined as those above or below the 95% CI for F_{ST} obtained from the 10,000
325 simulations. Simulations were done assuming no recombination, 50% missing data (i.e., female
326 samples sizes of 0.5*23 for *N. pinetum* and 0.5*44 for *N. lecontei*), and scaling theta ($4N\mu$) such
327 that the average number of SNPs per window across simulations were similar to the observed in
328 *Neodiprion* dataset.

329 *Comparison of empirical haplodiploid data to simulated diploids and haplodiploids*

330 To evaluate the potential influence of haplodiploidy on genomic differentiation between
331 *N. lecontei* and *N. pinetum*, we used SLiM v3 to simulate haplodiploid and diploid "genomes"
332 evolving under divergent selection and the demographic history estimated for our focal species
333 pair. Because *N. pinetum* is on the derived host plant (Linnen & Farrell, 2010), we modeled *N.*
334 *pinetum* as population 1 (where derived allele *a* is favored), and *N. lecontei* as population 2. For
335 these simulations, we also assumed a sex ratio of 70 females:30 males based on previously

336 published sex ratios for *N. lecontei* and *N. pinetum* (Harper et al., 2016). As in our first set of
337 simulations, we scaled our simulations to ensure equivalent levels of drift, migration, mutation
338 rates and recombination between diploids and haplodiploids.

339 To reduce the computational burden of forward SLiM simulations we re-scaled
340 parameters such that under neutrality, the SFS obtained with SLiM was identical to the expected
341 SFS obtained under the demographic history inferred with fastsimcoal2. This was achieved by
342 ensuring that the scaled mutation rate $4N_A\mu L$ for L sites was identical in both cases, where N_A is
343 the ancestral effective size and μ is the mutation rate per site per generation. By considering a
344 mutation rate two orders of magnitude higher (μ=3.50x10^{-7} rather than the 3.50x10^{-9}
345 /bp/generation used for SFS-based inference) and that L=5x10^{5} sites in SLiM corresponds to
346 L=5x10^{4} sites in the 2D-SFS used for fastsimcoal2 (including SNPs and invariant sites), the
347 haploid effective population sizes were three orders of magnitude lower (328 for *N. pinetum*,
348 1093 for *N. lecontei* and 1982 for the ancestral population) and migration rates three orders of
349 magnitude higher (3.64x10^{-4} into *N. pinetum*, 1.71x10^{-5} into *N. lecontei*). The times of split were
350 scaled accordingly, resulting in 1548 generations (rather than 1.54x10^{6}). To obtain the number of
351 individuals N in SLiM that correspond to the above haploid effective sizes, we had to account for
352 the sex-ratio of 70 females:30 males (Supplementary Methods). Given the average of 19.02
353 SNPs in *Neodiprion* 100-kb windows (with gaps due to sparse ddRAD loci), the average number
354 of SNPs in SLiM simulations under neutrality of 2,257 would correspond to ~10Mb of a similar
355 ddRAD dataset. We considered two recombination rates r: (i) equal to the mutation rate
356 (r=μ=3.50x10^{-7}), and (ii) three times higher than mutation rate (r=1.05x10^{-6}) using an estimate of
357 3.43 cM/Mb based on a linkage map for *N. lecontei* constructed from an inter-population cross
358 (Linnen et al., 2018). In both cases, recombination rate for diploids was scaled by (2/3) as done

359 for the simulation study (see above). Because the per-locus selection estimate is unknown, we
360 simulated differentiation and under a wide range of selection coefficients s from 0.0 to 0.3. We
361 also simulated all combinations of two dominance coefficients (h = 0.01 and 0.50) and one
362 starting allele frequency (q_0 = 0.10) (Supplementary Table S4). We computed mean F_{ST} across
363 all 1000 simulations for each starting allele frequency, dominance, and selection coefficient
364 combination. These combined simulations can be thought of as a divergence history in which, on
365 average, there is a divergently selected site every 10Mb.

366

367 **Results**

368 *Faster-haplodiploid effects on genomic differentiation with migration*

369 Across all parameter combinations and for both window sizes (20-kb and 500-kb), we
370 found that F_{ST} between haplodiploid populations was always equal to or greater than F_{ST} between
371 diploid populations (Figure 2). Migration was required for haplodiploid F_{ST} to exceed diploid
372 F_{ST}. For both window sizes, we found that faster-haplodiploid effects (i.e., ratio of haplodiploid
373 F_{ST}: diploid $F_{ST} > 1$) were more pronounced in the recessive case (Figure 2a-d) than in the
374 codominant case (Figure 2e-h). For each dominance coefficient, the regions of parameter space
375 that *maximized* faster-haplodiploid effects depended on the window size used to calculate F_{ST}.
376 For sites tightly linked to the selected site (20-kb window), faster-haplodiploid effects were
377 maximized when migration was high ($2N_em \geq 2.5$) and selection was moderate ($10 \leq 2N_es \leq 40$)
378 (Figures 2b and 2f). By contrast, when we considered much larger windows of 500-kb, relative
379 differences between haplodiploid and diploid differentiation levels were maximized at higher
380 selection coefficients ($2N_es \geq 40$) but at similar migration rates ($2N_em \geq 2.5$). The same trend—

381 faster-haplodiploid effects maximized at higher selective coefficients for the 500-kb windows
382 than for the 20-kb windows—was found for all initial frequencies of the derived allele *a* for the
383 codominant case and for q_0=0.5 for the recessive case (Supplementary Figure S2). However, for
384 new (q_0=1/(2*N*)) or rare (q_0=0.01) recessive alleles, faster-haplodiploid effects were always
385 maximized at the highest selection coefficients, regardless of window size (Supplementary
386 Figure S2).

387 One mechanism leading to greater differentiation in haplodiploids was differential allele
388 loss. Simulations of haplodiploid populations had a higher probability of retaining the derived
389 allele across a wide range of parameter combinations (Figure 3, Supplementary Figure S3). For
390 recessive alleles, differences in allele retention between haplodiploids and diploids were
391 dependent on both starting allele frequency and selection strength, but relatively insensitive to
392 migration rate (Figure 3a; Supplementary Figure S3). By contrast, outside of a slightly elevated
393 probability of retaining the derived allele at lower selection coefficients ($10 \leq 2N_e s \leq 20$, inset
394 panel in Figure 3b), differences in allele retention between haplodiploid and diploid populations
395 were minimal for the codominant case (Figure 3b). For both recessive and codominant cases,
396 increasing migration had a limited but consistent effect, leading to a lower probability of allele
397 retention.

398 To investigate the impact of haplodiploidy on genomic differentiation via mechanisms
399 other than differential allele loss, we computed average F_{ST} for each parameter combination
400 conditional on retaining the derived allele at the selected locus. For the recessive case,
401 controlling for the impact of differential allele loss decreased the magnitude of faster-
402 haplodiploid effects across all parameter combinations (Fig. 4a vs Fig. 2b and Fig. 4e vs. Fig.
403 2d), indicating that the increased retention of the derived allele in haplodiploids contributed to

404 faster-haplodiploid evolution. Compared to the recessive case, conditioning on allele retention
405 had much less of an impact on the magnitude of faster-haplodiploid effects for the codominant
406 case. The heatmaps in Figure 5a and 5e are nearly identical to those in Figure 2f and 2h,
407 respectively. This is unsurprising since differences in allele loss/retention were minimal in the
408 codominant case (Figure 3). Finally, once we conditioned on retaining the derived allele, starting
409 allele frequency had little impact on patterns of faster-haplodiploid evolution (Supplementary
410 Figure S4; for comparison see Supplementary Figure S2).

411 The observation that haplodiploid F_{ST} tends to exceed diploid F_{ST} even after conditioning
412 on retaining the derived allele (Fig. 4a,b,e,f; Fig. 5a,b,e,f) indicates that mechanisms other than
413 differential allele retention must contribute to elevated differentiation in haplodiploids. To gain
414 intuition about these mechanisms note that with migration, there is a constant influx of
415 maladapted alleles into each population. For diploids, hybridization between immigrants and
416 residents is likely to generate hybrids that are heterozygous for the locally adaptive allele (*Aa*).
417 The heterozygotes (*Aa*) are important because the fate of maladapted alleles is determined by
418 differences in fitness between heterozygotes (w_{Aa}) and resident homozygotes (w_{aa} in population
419 1, w_{AA} in population 2)(Yeaman & Otto, 2011). Additionally, the fate of linked neutral sites
420 ultimately depends on recombination in *Aa* heterozygotes. For haplodiploids, there are no
421 heterozygote males (Figure 1a), and males do not recombine, which likely affects (i) how
422 efficiently selection removes maladaptive alleles at the selected site; and (ii) how recombination
423 breaks immigrant haplotypes allowing for gene flow at neutral sites that escape linkage to
424 maladapted alleles. The impact of the lack of heterozygous males in haplodiploids can be
425 determined by comparing haplodiploid and diploid allele frequency trajectories at the selected
426 site (which reveals differences in the efficacy of selection against maladapted alleles) and F_{ST}

427 scans along the 500-kb chromosome (which reveals differences in effective migration rate at
428 linked sites).

429 We examined allele trajectories and chromosome-wide F_{ST} patterns under high migration
430 ($2Nm$=5.1) and two selection intensities (moderate: $2Ns$=40 and strong: 2Ns=200). For sites
431 tightly linked to recessive alleles (20-kb windows), elevated haplodiploid differentiation was
432 observed for both selection coefficients (Figure 4b). Simulated allele trajectories at the site under
433 divergent selection reveal two possible sources of elevated differentiation. First, the derived
434 allele increased in frequency in population 1 in the haplodiploids more rapidly than in the
435 diploids (Figure 4c,d). Second, once an equilibrium between migration and selection was
436 reached, selection was more efficient at removing maladapted immigrant recessive *a* alleles in
437 population 2 in the haplodiploids. In population 2, haplodiploid *A* males had a higher fitness than
438 immigrant *a* males ($w_A > w_a$); whereas the fitness of diploid *AA* males was similar to *Aa* males
439 ($w_{AA} \approx w_{Aa}$, Figure 1a). As a result, the frequency of *a*-bearing haplotypes was lower in
440 population 2—and F_{ST} between populations was higher—in haplodiploids. We attribute these
441 differences in allele trajectories to differences in male fitness, since haplodiploid and diploid
442 females have the same fitness (Figure 1a) and we assumed a 50:50 sex-ratio with no sex-specific
443 migration rates.

444 For codominant and tightly linked sites (20-kb window), a faster-haplodiploid effect was
445 more pronounced for moderate selection than for strong selection (Figure 5a,b). This can be
446 explained using as an approximation for the efficiency of selection, the fitness difference
447 between heterozygotes and resident males for diploids ($\Delta w_d = w_{AA} - w_{Aa} = s/2$), and between
448 immigrant and resident males for haplodiploids ($\Delta w_h = w_A - w_a = s$, Figure 1a). Our results are
449 consistent with this approximation, showing that even for the codominant case, selection is more

450 efficient in haplodiploids ($s > s/2$). Under strong selection, by contrast, locally adaptive alleles
451 were nearly fixed in both populations (Figure 5d). Because there were few heterozygotes in
452 either population, differences in the efficacy of selection in haplodiploid and diploid males had a
453 negligible impact on differentiation at the selected site.

454 For 500-kb windows, a faster-haplodiploid effect was more pronounced for strong
455 selection than for moderate selection for both recessive (Figure 4e-f) and codominant (Figure 5e-
456 f) cases. At moderate selection coefficients, differentiation across the 500-kb region was elevated
457 at the selected site and nearby sites but dropped off quickly to the expected neutral F_{ST} (Figure
458 4g, 5g). This is consistent with moderately strong divergent selection reducing the effective
459 migration rate only at neutral sites tightly linked to the selected site. By contrast, under strong
460 selection, chromosome-wide differentiation was elevated above neutral expectations for both
461 haplodiploids and diploids, even at distant sites (Figure 4h, 5h). We attribute the elevated
462 differentiation across the whole 500-kb region in haplodiploids to reduced opportunities for
463 recombination between divergently selected haplotypes. Such recombination events ultimately
464 depend on the frequency of *Aa* heterozygotes, and for haplodiploids, a heterozygote *Aa* can only
465 occur and recombine in females.

466 Our results reveal that haplodiploidy reduced opportunities for recombination in
467 heterozygotes at two distinct phases. First, during the initial stages of divergence, haplodiploids
468 reached migration-selection equilibrium faster than diploids: approximately ~3x faster for the
469 recessive case (Figure 4c-d) and ~1.2x faster for the codominant case (Figure 5c-d). The faster
470 time to equilibrium in the codominant case resulted in a chromosome-wide faster-haplodiploid
471 effect even though there was no difference in F_{ST} at the selected site (Figure 5a,e). Second, once
472 migration-selection equilibrium was reached, the frequency of the maladapted allele in each

473 population tended to be lower in haplodiploids than diploids (Figures 4c-d, 5c-d), resulting in a
474 lower frequency of *Aa* heterozygotes and a smaller effective recombination rate. This was
475 especially pronounced in the recessive case. As expected, simulations with a lower
476 recombination rate caused faster-haplodiploid differentiation of 500-kb to more closely resemble
477 the 20-kb window results (Supplemental Figure S5).

478 Overall, our simulations suggest so long as there is migration, haplodiploidy will lead to
479 elevated differentiation at selected sites and linked neutral sites when populations diverge via
480 divergent selection. This "faster-haplodiploid effect" is produced under a wide range of selection
481 coefficients; regardless of whether selection acts on new mutations, rare standing genetic
482 variation, or common standing genetic variation; and regardless of whether selection acts on
483 recessive or codominant alleles (Figure 2, Supplementary Figure S2). Finally, while the effects
484 of hemizygous selection and sex-limited recombination tend to be most pronounced at selected
485 sites and tightly linked neutral sites (20-kb windows), faster-haplodiploid effects can extend far
486 beyond the selected site (500-kb windows, which translates to ~4.4 cM in our simulations).

487

488 *Demography and genomic differentiation in pine sawflies*

489 *Neodiprion lecontei* and *N. pinetum* differ in many host-related traits (Figure 6A). Our
490 admixture analysis of *N. lecontei*, *N. pinetum*, a lab-reared F_1 hybrid, and three suspected wild-
491 caught hybrids supported two distinct genetic clusters (K=2; Supplementary Table S3). Putative
492 wild hybrids were indistinguishable from the lab-reared hybrid, and all four individuals were
493 genetically admixed with approximately equal contributions from *N. pinetum* and *N. lecontei*
494 (Figure 6B). The only other admixed individual detected was morphologically indistinguishable

495 from *N. pinetum*, but an estimated ~13% of its genome came from *N. lecontei*. In addition to
496 finding evidence of recent admixture, an ABBA-BABA test revealed evidence of historical
497 introgression between sympatric *N. pinetum* and *N. lecontei* populations (D=0.18; P = 2.12 x 10^{-}
498 ^{15}). Finally, demographic models that had no migration or only a single burst of admixture were
499 far less likely than models that included continuous migration (Table 1). Together, these results
500 support a divergence-with-gene-flow scenario for *N. lecontei* and *N. pinetum* (Figure 6C).

501 Comparing three different models that included migration (staring after divergence,
502 stopping before the present day, or continuous migration), our SFS data were most likely under
503 the model that had the fewest parameters: a continuous migration model (Table 1). Maximum
504 likelihood parameter estimates under this model suggest that *N. pinetum* and *N. lecontei* diverged
505 ~1.5×10^6 generations ago. Assuming 1-3 generations per year for KY populations of these
506 species (Benjamin 1955; Rauf and Benjamin 1980; CL, personal observation), this estimate
507 suggests that *N. pinetum* and *N. lecontei* likely diverged between 0.5 and 1.5 million years ago.
508 Our parameter estimates also suggest that *N. pinetum* has a smaller N_e than *N. lecontei*, and that
509 migration rates have been asymmetric, with more migration from *N. lecontei* to *N. pinetum* than
510 the reverse (Figure 6C, Table1). Importantly, this model provides a good fit to the observed SFS
511 and other summary statistics (Supplementary Figure S6).

512 Despite continuous migration throughout divergence, genome-wide average F_{ST} was high
513 (F_{ST} = 0.63). However, differentiation levels varied widely across the genome, with localized
514 regions of both very high and very low F_{ST} (Figure 7). Using simulations according to the
515 inferred demographic history to generate 95% confidence intervals for F_{ST} under neutrality
516 revealed evidence of both high-F_{ST} and low-F_{ST} outliers in our empirical dataset (Figure 7).
517 These regions are candidates for divergent selection and adaptive introgression, respectively.

518 Nucleotide diversity (π) for both *N. pinetum* and *N. lecontei* also varied across the genome, but
519 *N. lecontei* had a higher average π (2.71×10^{-5}) than *N. pinetum* (1.95×10^{-5}) which is consistent
520 with the differences in effective population size between the two species (Table 1). Overall,
521 demographic modelling and genomic differentiation patterns are consistent with the hypothesis
522 that this species pair diverged with substantial gene flow, while experiencing divergent selection
523 at many unlinked locations throughout the genome.

524 *Expectations for faster-haplodiploid effects under inferred demographic model*

525 There were several differences between our simulations and our empirical system,
526 including lower migration rates and asymmetries in both effective population size and migration
527 rate (Table 1) as well as female-biased sex ratios (Harper et al., 2016). To capture some of these
528 system-specific characteristics, we simulated haplodiploid and diploid populations evolving
529 under the demographic model we estimated from our sawfly data. A comparison between our
530 observed summary statistics (SFS, F_{ST}, Π, D_{xy}, and r^2) for the putatively neutral intergenic
531 regions and summary statistics obtained from neutral simulations ($2Ns = 0$) shows that the
532 demographic model implemented in SLiM is working as expected, and that our simulated diploid
533 and haplodiploid chromosomes do not differ under neutrality (Supplementary Figure S6).

534 When we included divergent natural selection in our sawfly-parameterized simulations of
535 diploid and haplodiploid genomes, we once again observed faster-haplodiploid effects on the
536 "genome-wide" mean F_{ST} for both dominant and recessive alleles, rare and common genetic
537 variants, and under a range of selection coefficients (Figure 8). As observed for simulations
538 under the simpler isolation-with-migration model, the magnitude of this effect was highest for
539 rare, recessive alleles and moderate-to-strong selection. These simulations also demonstrate

540 faster-haplodiploid effects can be observed with female-biased sex ratios and with asymmetric
541 migration.

542

543 **Discussion**

544 Haplodiploid taxa are numerous and ecologically diverse (Forbes et al., 2018; Hölldobler
545 & Wilson, 1990). While haplodiploid diversity could be due low transition rates between
546 haplodiploidy and diploidy, it is also possible that haplodiploidy increases speciation rates
547 (Blackmon et al., 2017; Koevoets & Beukeboom, 2008; Lohse & Ross, 2015; Patten et al.,
548 2015). Here we explore one avenue through which haplodiploidy may facilitate speciation: by
549 increasing genomic differentiation and linkage disequilibrium between populations that diverge
550 with gene flow. Specifically, our simulations reveal that so long as there is both selection and
551 migration, haplodiploid populations will maintain higher levels of differentiation than
552 comparable diploid populations. This is true not only at the selected site, but also up to ~20 cM
553 away. With the sawfly empirical data, we identify a potential case of sympatric divergence via
554 adaptation to different hosts. Here, we discuss implications of these results for faster-X theory,
555 evolution in haplodiploids, and models of sympatric speciation. We also discuss some limitations
556 of our models and data and highlight priorities for future work.

557

558 *Faster-haplodiploid effects under divergence-with-gene-flow and relevance to faster-X theory*

559 Overall, our simulations demonstrate multiple mechanisms through which genomic
560 differentiation in haplodiploids is increased relative to diploids when populations diverge with
561 gene flow. Given similarities between the transmission genetics of haplodiploid genomes and X

562 chromosomes, these findings are also relevant to faster-X theory. As described below, our
563 simulations recapitulate several key results from previous work on faster-X theory, albeit in
564 some additional corners of parameter space (e.g., divergence-with-gene-flow via new or rare
565 mutations; *c.f.* Lasne et al., 2017). Based on our simulations, we can group faster-haplodiploid
566 effects and mechanisms into three distinct phases. In the first phase, the dynamics of a
567 divergently selected, low-frequency allele is mostly determined by the risk of loss due to drift
568 (Figure 3, Supplementary Figure S3). The increased efficacy of selection and the reduced
569 probability of allele loss under haplodiploidy and a divergence-with-gene-flow scenario is
570 analogous to classical faster-X divergence among isolated populations in which dominance has a
571 large impact on the outcomes (e.g., Charlesworth et al., 1987; Vicoso and Charlesworth 2009;
572 Meisel and Connallon 2013; Charlesworth et al., 2018). However, assuming sufficient
573 recombination, differential allele loss during phase 1 has a minimal impact on differentiation at
574 linked neutral sites (e.g., compare Figures 2d and 2h to Figures 4e and 5e, respectively).

575 Once populations escape phase 1 without losing the divergently selected allele, there is a
576 second transitional phase during which the increased efficacy of selection against locally
577 maladaptive alleles reduces effective migration rates and causes haplodiploid loci to differentiate
578 more rapidly than comparable diploid loci (Figures 4 and 5). Again, this is in line with classical
579 faster-X theory demonstrating shorter sojourn times for beneficial X-linked alleles in isolated
580 populations (Avery, 1984; Betancourt et al., 2004). Reduced sojourn times in phase 2 also reduce
581 opportunities for recombination between locally adaptive and maladaptive haplotypes, thereby
582 affecting linked variation (Figures 4 and 5), analogous to predictions for X-linked variation in
583 isolated populations (Betancourt et al., 2004; Owen, 1988).

584 Once diverging populations approach equilibrium between selection, migration, and drift,
585 they enter phase 3. In this phase, haplodiploidy increases the efficacy of selection against locally
586 maladapted immigrant alleles, resulting in higher allele frequency differences at hemizygous loci
587 compared to diploid loci (Figures 4 and 5). Consistent with deterministic results obtained under
588 similar demographic models (Lasne et al., 2017), we find that when populations remain
589 connected by gene flow, faster-haplodiploid effects occur irrespective of dominance. These
590 findings contrast with classical faster-X theory that has been developed for divergence in
591 isolation, which predicts increased substitution rate only when beneficial mutations are recessive
592 (e.g., Charlesworth et al., 1987; Vicoso & Charlesworth 2006; Meisel & Connallon 2013;
593 Charlesworth et al., 2018) or when codominance is accompanied by deviations from 50:50 sex
594 ratio (Vicoso & Charlesworth, 2009). Additionally, efficient selection against maladapted
595 migrant alleles in Phase 3 causes reduced opportunities for recombination in haplodiploids. This
596 mechanism produces faster-haplodiploid effects at neutral sites linked to both recessive and
597 codominant alleles (Figures 4 and 5). These results are also consistent with predictions from
598 deterministic continent-island models of secondary-contact (Fraïsse & Sachdeva, 2021; Fusco &
599 Uyenoyama, 2011; Muirhead & Presgraves, 2016). Despite several important differences
600 between our model and these secondary contact models, including divergence scenario,
601 migration direction, and the presence of drift, we reach qualitatively similar conclusions. These
602 similarities suggest that in the long term, after migration-selection-drift equilibrium is reached,
603 the impact of the initial phases is negligible.

604 Where our work departs most from previous faster-X theory is that—to better connect
605 theory to data—we have explicitly modelled the effects of sex-limited hemizygosity and
606 recombination on population genomic datasets. Here, a couple of surprises have emerged. First,

607 under some parameter combinations, faster-haplodiploid effects can be observed in loosely
608 linked neutral sites without corresponding effects at the selected site and tightly linked sites
609 (Figure 5h). These patterns, which are dependent on recombination rate, emerge when
610 haplodiploid populations diverge more rapidly than diploid populations, but ultimately reach the
611 same equilibrium allele frequency. In essence, more rapid differentiation and reduced
612 opportunities for recombination between divergently selected haplotypes in haplodiploids locks
613 linked variation into place, and moderate to strong selection prevents erosion of linkage even
614 with migration. In other words, haplodiploidy leads to larger genomic regions around the
615 selected site with reduced effective migration rate. One important implication of this finding is
616 that haplodiploidy can facilitate the establishment of beneficial mutations that appear in such
617 genomic regions (Yeaman et al., 2016, see below).

618 Second, while previous work demonstrates that hemizygous selection will give rise to
619 faster-X effects at selected and linked sites, there has been some uncertainty as to how much of
620 the genome is likely to be impacted when there is recurrent migration (Presgraves, 2018). Here,
621 we show that with strong selection ($2Ns$>100) and high migration ($2Nm$~5), regions of elevated
622 differentiation in haplodiploid chromosomes will be higher and wider than in corresponding
623 diploid chromosomes. Moreover, divergent selection at a single haplodiploid locus can reduce
624 gene flow relative to the diploid case even at neutral sites more than 250-kb away, corresponding
625 to >4.40 cM in simulations under the symmetric isolation-with-migration model (Figures 4 and
626 5), or >21.5 cM in simulations with sawfly-specific parameters (Figure 8). Additionally, the
627 continent-island models of Fusco & Uyenoyama (2011) and Muirhead & Presgraves (2016)
628 predict that selection at X-linked (and, analogously, haplodiploid) sites can also impact unlinked
629 neutral markers. Assuming that at equilibrium, adaptive divergence-with-gene-flow dynamics

630 can be reasonably well approximated by the continent-island model, we speculate that localized
631 reductions in gene flow surrounding hemizygous loci could extend to the chromosome-wide
632 level.

633 Although our work better connects theory to empirical data, our models make several
634 simplifying assumptions that could impact patterns of faster-haplodiploid differentiation.
635 Relaxing these assumptions and creating more complex, but realistic, models are therefore
636 potentially fruitful avenues for future research. First, the genetic architecture of adaptation to
637 novel niches is likely much more complex than the simple single-locus model considered here
638 (i.e., adaptation is likely due to many loci with variable effect sizes, dominance coefficients, and
639 non-additive interactions). Although there has been some work on how divergent selection on
640 multiple, possibly interacting, loci impact genomic differentiation (e.g., Yeaman et al., 2016;
641 Aeschbacher et al., 2017), this has not been investigated in the context of hemizygosity (but see
642 Fusco & Uyenoyama 2011; Fraïsse & Sachdeva 2021). Second, the divergence model we
643 considered was also relatively simple, ignoring sex-specific effects such as sex-biased migration,
644 sex-specific selection, and the absence of dosage compensation. For instance, Lasne et al. (2017)
645 predicted that sex-specific migration has a large impact on faster X effects in models with strong
646 migration ($m>>s$). Third, we have considered a parallel dominance fitness landscape and the
647 magnitude of faster-X differentiation might differ for other models (but see Lasne et al., 2017).
648 Fourth, we assumed that the fitness of haploid males was equivalent to that of diploid
649 homozygotes. Whether or not this is the case depends on mechanisms of dosage compensation
650 and allelic effects in haploid males, which are not well understood (Gardner 2012; Hitchcock
651 2021; but see Aron et al., 2005; Dearden et al., 2006; Glastad et al., 2014). Finally, we have
652 ignored the effects that removing deleterious mutations with similar effects across populations

653 may have on patterns of differentiation in haplodiploids and diploids (Charlesworth et al., 1993;
654 Charlesworth et al., 1997). Making precise predictions about chromosome-wide levels in
655 population genomic datasets will require modeling these more complex scenarios, as well as
656 considering local variation in mutation and recombination rate.

657

658 *Implications for speciation in haplodiploids*

659 One of the longest running debates in evolutionary biology is over the plausibility and
660 prevalence of sympatric speciation, the evolution of reproductive isolation in absence of
661 geographical isolation (Berlocher & Feder, 2002; Bolnick & Fitzpatrick, 2007; Foote, 2018; Via,
662 2001). Because of their pronounced host specialization and lifelong association with their host
663 plants, *Neodiprion* sawflies have been hypothesized to undergo sympatric speciation (Bush,
664 1975a, 1975b; Knerer & Atwood, 1973; Linnen & Farrell, 2010). Although gene flow has been
665 ubiquitous throughout *Neodiprion* divergence (Linnen & Farrell, 2007) and *N. pinetum*'s range is
666 nested within *N. lecontei*'s range, species ranges have changed too much to reconstruct the
667 geographic context of speciation from present day range overlap (Linnen & Farrell, 2010). Here,
668 demographic modeling revealed that the model that best explains patterns of genomic variation
669 in *N. lecontei* and *N. pinetum* does not include a period of isolation (Table 1; Figure 6C).
670 However, distinguishing between models of sympatric divergence and secondary contact is
671 difficult (Sousa & Hey, 2013). This difficulty appears to be true for *N. lecontei* and *N. pinetum*
672 as well: models that included continuous migration either starting after a brief period of isolation
673 (~64000 generations) or ending right before the present day (~2000 generations) explained the
674 data nearly as well as a continuous migration model (Table 1). Despite this uncertainty, our top
675 three models and ML parameter estimates all point to a scenario in which gene flow was present

676 throughout all or most of the divergence history of these two species.

677 Previous work also demonstrates that differences in the pines that *N. lecontei* and *N.*
678 *pinetum* use are likely to generate divergent selection on many different types of traits, including
679 female oviposition traits, correlated male traits, and larval physiology (Bendall et al., 2020, 2017;
680 Benjamin, 1955; Coppel & Benjamin, 1965; Rauf & Benjamin, 1980; Wilson et al., 1992).
681 Consistent with a "multifarous" or "multidimensional" model of divergent selection (Feder &
682 Nosil, 2010; Rice & Hostert, 1993; White & Butlin, 2021), multiple unlinked loci exceeded
683 expected levels of differentiation under neutrality (Figure 7). We also observed multiple regions
684 of unusually low differentiation, which could be explained by adaptive introgression. We
685 acknowledge, however, that other mechanisms besides divergent selection and adaptive
686 introgression can cause dips and valleys in genome scans (Cruickshank & Hahn, 2014; Ravinet
687 et al., 2017). Thus, interpretation of these genome scans would be improved by characterizing
688 the genomic landscape of recombination and gene density, as well as mapping loci underlying
689 divergently selected traits.

690 Together with previous work characterizing reproductive barriers in this species pair
691 (Bendall et al., 2017), our demographic modeling and genome scan results support a scenario in
692 which adaptation to different pine trees drove the evolution of reproductive isolation in the
693 presence of substantial gene flow. There is little debate that, given sufficiently strong selection,
694 genetic and phenotypic differences, such as divergently selected host-use traits, can be
695 maintained in the face of gene flow. Instead, the primary objection to sympatric speciation has
696 been that gene flow and recombination will tend to break up associations among favorable
697 combinations of alleles and between divergently selected loci and loci that confer reproductive
698 isolation (Felsenstein, 1981). Multiple mechanisms can help overcome this "selection-

recombination" antagonism, thereby aiding the evolution of reproductive isolation when there is gene flow, including pleiotropy ("magic traits" wherein loci underlying local adaptation also confer reproductive isolation (Servedio et al., 2011)) and genomic features that reduce recombination (e.g., chromosomes inversions (Kirkpatrick & Barton, 2006; Ravinet et al., 2017)). Because *Neodiprion* mate on the host plant, it is possible that alleles underlying divergent host preferences also produce habitat isolation (Linnen & Farrell, 2010). However, there is little evidence of chromosomal rearrangements that would reduce recombination in *N. lecontei*-*N. pinetum* hybrids (Geib and Sim, personal communication).

Even in the absence of "magic traits" and inversions, divergent selection can also facilitate the evolution of reproductive isolation through effects on linked variation (divergence hitchhiking; Via, 2009; Via & West, 2008) and, when there are multiple divergently selected loci, via a genome-wide reduction in the effective migration rate (genome hitchhiking; Barton & Bengtsson, 1986; Feder et al., 2012a,b; Flaxman et al., 2012, 2013). By simulating genomic differentiation under divergent selection, our estimated demographic model, and other system-specific details (sex ratio, recombination rate), we show that haplodiploid inheritance in *N. lecontei* and *N. pinetum* likely increased differentiation at selected and linked loci relative to a comparable diploid scenario (Figure 8). The impact of haplodiploidy was most pronounced for recessive mutations and at intermediate selection coefficients, with effects extending over sizable regions of the genome (comparable to ~10Mb). By increasing differentiation at linked sites, haplodiploidy could facilitate both divergence hitchhiking and genome hitchhiking, thereby promoting speciation-with-gene-flow. This hypothesis could be tested more directly via simulations that examine the impact of haplodiploidy on non-neutral linked variation and

721 interactions between multiple loci (Feder et al., 2012b; Feder & Nosil, 2010; Flaxman et al.,
722 2012; Nosil & Feder, 2012; Via, 2012; Yeaman & Whitlock, 2011; Yeaman et al., 2016).
723
724 *Conclusions*

725 Overall, our work suggests that sex-limited hemizygosity and recombination, both of
726 which are maximized in Hymenoptera and other haplodiploid clades, can have substantial effects
727 on genomic differentiation in wild populations. One potential implication of this work is that
728 haplodiploid taxa are more likely to undergo sympatric speciation and can withstand greater
729 levels of gene flow during divergence, which may ultimately give rise to higher rates of local
730 adaptation and speciation. A comparative analysis of divergence history between diploids and
731 haplodiploids would be an informative step in testing this hypothesis. More generally, there are
732 potentially numerous evolutionary consequences of haplodiploidy that may shed light on patterns
733 of biodiversity and have implications that extend to non-haplodiploid taxa.

734

735 **Acknowledgements**

736 We thank members of the Linnen lab for assistance with collection and rearing of pine sawflies,
737 and we thank Kim Vertacnik for providing data from an outgroup taxon. This work was
738 supported by the National Science Foundation (DEB-CAREER-1750946 and DEB-1257739 to
739 C.R.L); USDA-NIFA pre-doctoral fellowship (2015-67011-22803 to R.K.B.); Portuguese
740 National Science Foundation ("Fundação para a Ciência e a Tecnologia" – FCT;
741 UIDB/00329/2020, individual grants CEECIND/02391/2017 and
742 CEECINST/00032/2018/CP1523/CT0008 to V.C.S.); and by EU H2020 program (Marie

Skłodowska-Curie grant number 799729 to V.C.S.). For computing resources, we thank the University of Kentucky Center for Computational Sciences and the Lipscomb High Performance Computing Cluster, as well as the INCD (https://incd.pt/), for which access was funded through FCT Advanced Computing Projects (CPCA/A0/7303/2020 to V.C.S.).

References

Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences*, *114*(27), 7061–7066. doi: 10.1073/pnas.1616755114

Aron, S., de Menten, L., van Bockstaele, D. R., Blank, S. M., & Roisin, Y. (2005). When hymenopteran males reinvented diploidy. *Current Biology*, *15*(9), 824–827. doi: 10.1016/J.CUB.2005.03.017

Avery, P. J. (1984). The population genetics of haplo-diploids and X-linked genes. *Genetical Research*, *44*(3), 321–341. doi: 10.1017/S0016672300026550

Bagley, R. K., Sousa, V. C., Niemiller, M. L., & Linnen, C. R. (2017). History, geography and host use shape genome wide patterns of genetic variation in the redheaded pine sawfly (*Neodiprion lecontei*). *Molecular Ecology*, *26*(4), 1022–1044. doi: 10.1111/mec.13972

Barton, N. H., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *56*(3), 357–376. doi: 10.1038/hdy.1986.135

[dataset] Bendall, E.E., Bagley, R.K., Sousa, V.C. & Linnen, C.R.; 2021a; Simulation data for: Faster-haplodiploid evolution under divergence-with-gene-flow: simulations and empirical data from pine-feeding hymenopterans; Github; https://github.com/vsousa/EG_cE3c

[dataset] Bendall, E.E., Bagley, R.K., Sousa, V.C. & Linnen, C.R.; 2021b; Short-read data for: Faster-haplodiploid evolution under divergence-with-gene-flow: simulations and empirical data from pine-feeding hymenopterans; NCBI SRA; TBD

[dataset] Bendall, E.E., Bagley, R.K., Sousa, V.C. & Linnen, C.R.; 2021c; SNP data for: Faster-haplodiploid evolution under divergence-with-gene-flow: simulations and empirical data from pine-feeding hymenopterans; DRYAD; TBD

Bendall, E. E., Mattingly, K. M., Moehring, A. J., & Linnen, C. R. (2020). Lack of intrinsic postzygotic isolation in haplodiploid male hybrids despite high genetic distance. *BioRxiv*, 2020.01.08.898957. doi: 10.1101/2020.01.08.898957

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. doi: 10.1186/s12862-017-0872-8

Benjamin, D. M. (1955). *The Biology and Ecology of the Red-headed Pine Sawfly*. U.S. Department of Agriculture.

Berlocher, S., & Feder, J. (2002). Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology*, *47*, 773–815. doi: 10.1146/ANNUREV.ENTO.47.091201.145312

Betancourt, A. J., Kim, Y., & Orr, H. A. (2004). A pseudohitchhiking model of X vs. autosomal diversity. *Genetics*, *168*(4), 2261–2269. doi: 10.1534/genetics.104.030999

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting FST: the impact of rare variants. *Genome Research*, *23*(9), 1514–1521. doi: 10.1101/gr.154831.113

Blackmon, H., Ross, L., & Bachtrog, D. (2017). Sex determination, sex chromosomes, and karyotype evolution in insects. *Journal of Heredity*, *108*(1), 78–93. doi: 10.1093/jhered/esw047

Bolnick, D. I., & Fitzpatrick, B. M. (2007). Sympatric speciation: Models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics*, *38*, 459–487. doi: 10.1146/annurev.ecolsys.38.091206.095804

Boulton, R. A., Collins, L. A., & Shuker, D. M. (2015). Beyond sex allocation: the role of mating systems in sexual selection in parasitoid wasps. *Biological Reviews*, *90*(2), 599–627. doi: 10.1111/BRV.12126

Burford Reiskind, M. O., Coyle, K., Daniels, H. v., Labadie, P., Reiskind, M. H., Roberts, N. B., … Vargo, E. L. (2016). Development of a universal double-digest RAD sequencing approach for a group of nonmodel, ecologically and economically important insect and fish taxa. *Molecular Ecology Resources*, *16*(6), 1303–1314. doi: 10.1111/1755-0998.12527

Bush, G. L. (1975a). Modes of animal speciation. *Annual Review of Ecology and Systematics*, *6*(1), 339–364. doi: 10.1146/annurev.es.06.110175.002011

Bush, G. L. (1975b). Sympatric speciation in phytophagous parasitic insects. In P. W. Price (Ed.), *Evolutionary Strategies of Parasitic Insects and Mites* (pp. 187–206). Springer, Boston, MA. doi: 10.1007/978-1-4615-8732-3_9

Charlesworth, B., Coyne, J. A., & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist*, *130*(1), 113–146. doi: 10.1086/284701

Charlesworth, B, Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*(4), 1289–1303.

Charlesworth, B, Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, *70*(2), 155–174.

Charlesworth, B. (2012). The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the *Drosophila X* chromosome. *Genetics*, *191*(1), 233–246. doi: 10.1534/genetics.111.138073

Charlesworth, B., Campos, J. L., & Jackson, B. C. (2018). Faster-X evolution: Theory and evidence from *Drosophila*. *Molecular Ecology*, *27*(19), 3753–3771. doi: 10.1111/mec.14534

Chen, H., Rangasamy, M., Tan, S. Y., Wang, H., & Siegfried, B. D. (2010). Evaluation of five methods for total DNA extraction from western corn rootworm beetles. *PLoS ONE*, *5*(8), e11963. doi: 10.1371/journal.pone.0011963

Codella, S. G., & Raffa, K. F. (2002). Desiccation of *Pinus* foliage induced by conifer sawfly oviposition: effect on egg viability. *Ecological Entomology*, *27*(5), 618–621. doi: 10.1046/j.1365-2311.2002.00447.x

Coppel, H., & Benjamin, D. (1965). Bionomics of the nearctic pine-feeding diprionids. *Annual Review of Entomology*, *10*, 69–96.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. doi: 10.1111/mec.12796

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., … Project, G. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi: 10.1093/bioinformatics/btr330

Davies, N. G., & Gardner, A. (2014). Evolution of paternal care in diploid and haplodiploid populations. *Journal of Evolutionary Biology*, *27*(6), 1012–1019. doi: 10.1111/jeb.12375

de la Filia, A. G., Bain, S. A., & Ross, L. (2015). Haplodiploidy and the reproductive ecology of Arthropods. *Current Opinion in Insect Science*, *9*, 36–43. doi: 10.1016/j.cois.2015.04.018

Dearden, P. K., Wilson, M. J., Sablan, L., Osborne, P. W., Havler, M., McNaughton, E., … Beye, M. (2006). Patterns of conservation and change in honey bee developmental genes. *Genome Research*, *16*(11), 1376–1384. doi: 10.1101/gr.5108606

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905. doi: 10.1371/journal.pgen.1003905

Feder, J L, Egan, S. P., & Nosil, P. (2012a). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*(7), 342–350.

Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012b). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 461–474. doi: 10.1098/rstb.2011.0256

Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, *64*(6), 1729–1747. doi: 10.1111/j.1558-5646.2009.00943.x

Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution*, *35*(1), 124–138.

Flaxman, S. M., Feder, J. L., & Nosil, P. (2012). Spatially explicit models of divergence and genome hitchhiking. *Journal of Evolutionary Biology*, *25*(12), 2633–2650. doi: 10.1111/JEB.12013

Flaxman, S. M., Feder, J. L., & Nosil, P. (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*, *67*(9), 2577–2591. doi: 10.1111/evo.12055

Foote, A. D. (2018). Sympatric speciation in the genomic era. *Trends in Ecology and Evolution*, *33*(2), 85–95. doi: 10.1016/j.tree.2017.11.003

Forbes, A. A., Bagley, R. K., Beer, M. A., Hippee, A. C., & Widmayer, H. A. (2018). Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. *BMC Ecology 2018 18:1*, *18*(1), 1–11. doi: 10.1186/S12898-018-0176-X

Fraïsse, C., & Sachdeva, H. (2021). The rates of introgression and barriers to genetic exchange between hybridizing species: sex chromosomes vs autosomes. *Genetics*, *217*(2), doi: 10.1093/genetics/iyaa025

Frank, S. A. (1991). Divergence of meiotic drive-suppression systems as an explanation for sex-biased hybrid sterility and inviability. *Evolution*, *45*(2), 262–267. doi: 10.1111/j.1558-5646.1991.tb04401.x

Fusco, D., & Uyenoyama, M. K. (2011). Sex-specific incompatibility generates locus-specific rates of introgression between species. *Genetics*, *189*(1), 267–288. doi: 10.1534/genetics.111.130732

Gardner, A. (2012). Evolution of maternal care in diploid and haplodiploid populations. *Journal of Evolutionary Biology*, *25*(8), 1479–1486. doi: 10.1111/j.1420-9101.2012.02551.x

Ghenu, A.-H., Blanckaert, A., Butlin, R. K., Kulmuni, J., & Bank, C. (2018). Conflict between heterozygote advantage and hybrid incompatibility in haplodiploids (and sex chromosomes). *Molecular Ecology*, *27*(19), 3935–3949. doi: 10.1111/mec.14482

Glastad, K. M., Hunt, B. G., Yi, S. v., & Goodisman, M. A. D. (2014). Epigenetic inheritance and genome regulation: is DNA methylation linked to ploidy in haplodiploid insects? *Proceedings of the Royal Society B: Biological Sciences*, *281*(1785). doi: 10.1098/RSPB.2014.0411

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637. doi: 10.1093/molbev/msy228

Hamilton, W. D. (1964a). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, *7*(1), 1–16. doi: 10.1016/0022-5193(64)90038-4

Hamilton, W. D. (1964b). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, *7*(1), 17–52. doi: 10.1016/0022-5193(64)90039-6

Hamilton, W. D. (1967). Extraordinary sex ratios. *Science*, *156*(3774), 477–488. doi: 10.1126/science.156.3774.477

Hamilton, W. D. (1972). Altruism and related phenomena. *Annual Review of Ecological Systems*, *3*, 193–232.

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. doi: 10.1038/hdy.2016.46

Hedrick, P. W., & Parker, J. D. (1997). Evolutionary genetics and genetic variation of haplodiploids and X-linked genes. *Annual Review of Ecology and Systematics*, *28*, 55–83. doi: 10.1146/annurev.ecolsys.28.1.55

Hey, J., & Pinho, C. (2012). Population genetics and objectivity in species diagnosis. *Evolution*, *66*(5), 1413–1429. doi: 10.1111/j.1558-5646.2011.01542.x

Hitchcock, T., Gardner, A. & Ross, L. (2021). Sexual antagonism in haplodiploids. *biorXiv*, doi: 10.1101/2021.03.26.437233.

Hoban, S., Bertorelle, G., & Gaggiotti, O. E. (2012). Computer simulations: tools for population and evolutionary genetics. *Nature Reviews Genetics*, *13*(2), 110–122.

Hölldobler, B., & Wilson, E. O. (1990). *The Ants. Cambridge*. Harvard University Press. doi: 10.1007/978-3-662-10306-7

Hurst, L. D., & Pomiankowski, A. (1991). Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics*, *128*(4), 841–858.

Irwin, D. E. (2018). Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology*, *27*(19), 3831–3851. doi: 10.1111/mec.14537

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*(1), 419–434. doi: 10.1534/genetics.105.047985

Kirkpatrick, M., & Hall, D. W. (2004). Sexual selection and sex linkage. *Evolution*, *58*(4), 683–691. doi: 10.1111/j.0014-3820.2004.tb00401.x

Knerer, G., & Atwood, C. E. (1973). Diprionid sawflies: Polymorphism and speciation. *Science*, *179*(4078), 1090–1099. doi: 10.1126/science.179.4078.1090

Koevoets, T., & Beukeboom, L. W. (2008). Genetics of postzygotic isolation and Haldane's rule in haplodiploids. *Heredity 2009 102:1*, *102*(1), 16–23. doi: 10.1038/hdy.2008.44

Kong, A., Gudbjartsson, D. F., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardsson, B., … Stefansson, K. (2002). A high-resolution recombination map of the human genome. *Nature Genetics*, *31*(3), 241–247. doi: 10.1038/ng917

Kraaijeveld, K. (2009). Male genes with nowhere to hide; Sexual conflict in haplodiploids. *Animal Biology*, *59*(4), 403–415. doi: 10.1163/157075509X12499949744225

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. doi: 10.1038/nmeth.1923

Lasne, C., Sgrò, C. M., & Connallon, T. (2017). The relative contributions of the X chromosome and autosomes to local adaptation. *Genetics*, *205*(3), 1285–1304. doi: 10.1534/genetics.116.194670

Lester, L. J., & Selander, R. K. (1979). Population genetics of haplodiploid insects. *Genetics*, *92*(4), 1329–1345. doi: 10.1093/genetics/92.4.1329

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Lindstedt, C., Bagley, R., Calhim, S., Jones, M., & Linnen, C.R. (2020). The impact of life stage and pigment source on the evolution of novel signal traits. *bioRxiv*, doi: 10.1101/2020.10.20.346700

Linnen, C. R., & Farrell, B. D. (2007). Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution*, *61*(6), 1417–1438. doi: 10.1111/j.1558-5646.2007.00114.x

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. doi: 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. doi: 10.1098/rspb.2010.0577

Linnen, C. R., O'Quin, C. T., Shackleford, T., Sears, C. R., & Lindstedt, C. (2018). Genetic basis of body color and spotting pattern in redheaded pine sawfly larvae (*Neodiprion lecontei*). *Genetics*, *209*(1), 291–305. doi: 10.1534/genetics.118.300793

Lohse, K., & Ross, L. (2015). What haplodiploids can teach us about hybridization and speciation. *Molecular Ecology*, *24*(20), 5075–5077. doi: 10.1111/mec.13393

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads *EMBnet.Journal*, *1*, 10–12.

Meiklejohn, C. D., Landeen, E. L., Gordon, K. E., Rzatkiewicz, T., Kingan, S. B., Geneva, A. J., … Presgraves, D. C. (2018). Gene flow mediates the role of sex chromosome meiotic drive during complex speciation. *eLife*, *7*, e35468. doi: 10.7554/eLife.35468

Meisel, R. P., & Connallon, T. (2013). The faster-X effect: Integrating theory and data. *Trends in Genetics*, Vol. 29, pp. 537–544. doi: 10.1016/j.tig.2013.05.009

Moran, P. (1959). The theory of some genetical effects of population subdivision. *Australian Journal of Biological Sciences*, *12*(2), 109. doi: 10.1071/bi9590109

Muirhead, C. A., & Presgraves, D. C. (2016). Hybrid incompatibilities, local adaptation, and the genomic distribution of natural introgression between species. *The American Naturalist*, *187*(2), 249–261. doi: 10.1086/684583

Normark, B. B. (2003). The evolution of alternative genetic systems in insects. *Annual Review of Entomology*, *48*, 397–423. doi: 10.1146/annurev.ento.48.091801.112703

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 332–342.

Nouhaud, P., Blanckaert, A., Bank, C., & Kulmuni, J. (2020). Understanding admixture: haplodiploidy to the rescue. *Trends in Ecology and Evolution*, *35*, 34–42. doi: 10.1016/j.tree.2019.08.013

Owen, R. E. (1986). Gene frequency clines at X-linked or haplodiploid loci. *Heredity*, *57*(2), 209–219. doi: 10.1038/hdy.1986.111

Owen, R. E. (1988). Selection at two sex-linked loci. *Heredity*, *60*(3), 415–425. doi: 10.1038/hdy.1988.59

Patten, M. M. (2018). Selfish X chromosomes and speciation. *Molecular Ecology*, *27*(19), 3772–3782. doi: 10.1111/mec.14471

Patten, M. M., Carioscia, S. A., & Linnen, C. R. (2015). Biased introgression of mitochondrial and nuclear genes: A comparison of diploid and haplodiploid systems. *Molecular Ecology*, *24*(20), 5200–5210. doi: 10.1111/mec.13318

Patterson, N. J., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., … Reich, D. (2012). ancient admixture in human history. *Genetics*, *192(1)*, 1065–1093. doi: 10.1534/genetics.112.145037

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5), e37135. doi: 10.1371/journal.pone.0037135

Phung, T. N., Huber, C. D., & Lohmueller, K. E. (2016). Determining the Effect of Natural Selection on Linked Neutral Divergence across Species. *PLOS Genetics*, *12*(8), e1006199. doi: 10.1371/JOURNAL.PGEN.1006199

Pinho, C., & Hey, J. (2010). Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 215–230.

Presgraves, D. C. (2018). Evaluating genomic signatures of "the large X-effect" during complex speciation. *Molecular Ecology*, *27*(19), 3822–3830. doi: 10.1111/mec.14777

Rauf, A., & Benjamin, D. M. (1980). The biology of the white pine sawfly, *Neodiprion pinetum* (Hymenoptera: Diprionidae) in Wisconsin. *The Great Lakes Entomologist*, *13*(4), 219–224.

Rautiala, P., Helanterä, H., & Puurtinen, M. (2019). Extended haplodiploidy hypothesis. *Evolution Letters*, *3*(3), 263–270. doi: 10.1002/evl3.119

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., … Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. doi: 10.1111/jeb.13047

Reeve, H. K., & Pfennig, D. W. (2003). Genetic biases for showy males: Are some genetic systems especially conducive to sexual selection? *Proceedings of the National Academy of Sciences*, *100*(3), 1089–1094. doi: 10.1073/PNAS.0337427100

Rice, W., & Hostert, E. (1993). Laboratory experiments on speciation: what have we learned in 40 years? *Evolution*, *47*(6), 1637–1653. doi: 10.1111/J.1558-5646.1993.TB01257.X

Schweyen, H., Rozenberg, A., & Leese, F. (2014). Detection and Removal of PCR Duplicates in Population Genomic ddRAD Studies by Addition of a Degenerate Base Region (DBR) in Sequencing Adapters. *The Biological Bulletin*, *227*(2), 146–160. doi: 10.1086/BBLv227n2p146

Servedio, M. R., Doorn, G. S. van, Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: 'magic' but not rare? *Trends in Ecology & Evolution*, *26*(8), 389–397. doi: 10.1016/J.TREE.2011.04.005

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews Genetics*, *14*(6), 404–414. doi: 10.1038/nrg3446

Staab, P. R., Zhu, S., Metzler, D., & Lunter, G. (2015). scrm: efficiently simulating long sequences using the approximated coalescent with recombination. *Bioinformatics*, *31*(10), 1680–1682. doi: 10.1093/BIOINFORMATICS/BTU861

Vertacnik, K. (2020). Predicting patterns of gene family evolution in taxa with similar ecological niches (University of Kentucky). University of Kentucky. doi: https://doi.org/10.13023/etd.2020.371

Vertacnik, K. L., & Linnen, C. R. (2015). *Neodiprion lecontei genome assembly Nlec1.0. NCBI/GenBank.* doi: http://dx.doi.org/10.15482/USDA.ADC/1235565

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, *16*(7), 381–390.

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, *106*(1), 9939–9946.

Via, S, & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*(19), 4334–4345. doi: 10.1111/J.1365-294X.2008.03921.X

Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 451–460. doi: 10.1098/rstb.2011.0260

Vicoso, B., & Charlesworth, B. (2006). Evolution on the X chromosome: Unusual patterns and processes. *Nature Reviews Genetics*, Vol. 7, pp. 645–653. Nature Publishing Group. doi: 10.1038/nrg1914

Vicoso, B., & Charlesworth, B. (2009). Effective population size and the Faster-X effect: an extended model. *Evolution*, *63*(9), 2413–2426. doi: 10.1111/j.1558-5646.2009.00719.x

1037 Werren, J. H. (1993). The evolution of inbreeding in haplodiploid organisms. In N. Thornhill
1038 (Ed.), *The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical*
1039 *Perspectives*. University of Chicago Press.

1040 White, N. J., & Butlin, R. K. (2021). Multidimensional divergent selection, local adaptation, and
1041 speciation. *Evolution*, *75*(9), 2167–2178. doi: 10.1111/evo.14312

1042 Wilfert, L., Gadau, J., & Schmid-Hempel, P. (2007). Variation in genomic recombination rates
1043 among animal taxa and the case of social insects. *Heredity*, *98*(4), 189–197. doi:
1044 10.1038/sj.hdy.6800950

1045 Wilson, L. F., Wilkinson, R. C., & Averill, R. D. (1992). *Redheaded pine sawfly : its ecology*
1046 *and management*. U.S. Dept. of Agriculture, Forest Service.

1047 Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification
1048 using exact alignments. *Genome Biology*, *15*(3), R46. doi: 10.1186/gb-2014-15-3-r46

1049 Wright, A. E., Harrison, P. W., Zimmer, F., Montgomery, S. H., Pointer, M. A., & Mank, J. E.
1050 (2015). Variation in promiscuity and sexual selection drives avian rate of Faster-Z
1051 evolution. *Molecular Ecology*, *24*(6), 1218–1235. doi: 10.1111/mec.13113

1052 Yeaman, S, & Whitlock, M. (2011). The genetic architecture of adaptation under migration-
1053 selection balance. *Evolution*, *65*(7), 1897–1911. doi: 10.1111/J.1558-5646.2011.01269.X

1054 Yeaman, S., Aeschbacher, S., & Bürger, R. (2016). The evolution of genomic islands by
1055 increased establishment probability of linked alleles. *Molecular Ecology*, *25*(11), 2542–
1056 2558. doi: 10.1111/mec.13611

1057 Yeaman, Sam, & Otto, S. P. (2011). Establishment and maintenance of adaptive genetic
1058 divergence under migration, selection, and drift. *Evolution*, *65*(7), 2123–2129. doi:
1059 10.1111/j.1558-5646.2011.01277.x

1060

1061 **Data accessibility statement**

1062 All the scripts used for SLiM v3 simulations (bash and Slim input files) and analysis of
1063 SLiM results (R scripts) are available on VCS's GitHub page
1064 (https://github.com/vsousa/EG_cE3c) ([dataset] Bendall et al., 2021a). Upon acceptance for
1065 publication, *Neodiprion* sequencing reads will be deposited in the NCBI SRA ([dataset] Bendall
1066 et al., 2021b). All VCF files, custom scripts, and input files for the analysis of sawfly data will be
1067 deposited in DRYAD ([dataset] Bendall et al., 2021c).

1068

1069

1070 **Author contributions**

1071 EEB, VCS, and CRL designed the research; EEB, VCS, RKB, and CRL performed the research;
1072 EEB, VCS, and CRL analyzed the data; EEB, VCS, and CRL wrote the paper, with input from
1073 all authors.

1074

1075 **Table 1.** Maximum likelihood values and maximum-likelihood parameter estimates for the five
1076 demographic models tested.

Parameters	No migration	Secondary contact - one burst of migration	Continuous migration that starts after divergence	Continuous migration that ends before present day	Continuous migration
# of parameters estimated	7	7	7	7	6
Ancestral population size	2075281	1962052	1971660	2006404	1982187
N. pinetum population size	645779	369032	322897	337579	328311
N. lecontei population size	1057278	1052921	1104528	1075336	1093739
Time since divergence (generations)	1010864	1255392	1542789	1553120	1548690
N. pinetum bottleneck size	717	NA	NA	NA	NA
N. lecontei bottleneck size	470	NA	NA	NA	NA
Time since bottle neck	1010854	NA	NA	NA	NA
Admixture proportion from *N. pinetum* to *N. lecontei*	NA	0.0060941	NA	NA	NA
Admixture proportion from *N. lecontei* to *N. pinetum*	NA	0.118292	NA	NA	NA
Time since admixture	NA	115046	NA	NA	NA
Migration rate from *N. lecontei* to *N. pinetum*	NA	NA	3.94E-07	3.63E-07	3.65E-07
Migration rate from *N. pinetum* to *N. lecontei*	NA	NA	1.64E-08	1.83E-08	1.71E-08
Number of *N. pinetum* migrants	NA	NA	0.1273259	0.1226254	0.1196984
Number of *N. lecontei* migrants	NA	NA	0.0181423	0.0196645	0.018665
Time since migration started	NA	NA	901891	NA	NA
Time since migration ended	NA	NA	NA	1906	NA
Estimated maximum likelihood (log 10)	-33034.2	-32805.5	-32794.6	-32795	-32794.3
Maximum observed likelihood (log 10)	-32727.9	-32727.9	-32727.9	-32727.9	-32727.9
Maximum Likelihood (est-obs)	**-306.3**	**-77.6**	**-66.7**	**-67.1**	**-66.4**

1077

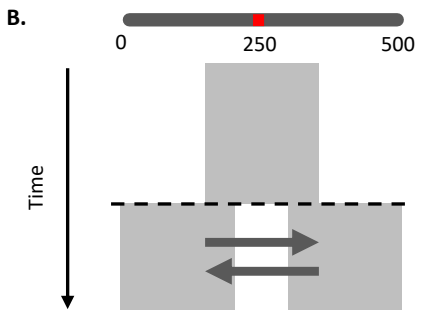
1078

A. Fitness Model of Diploid Chromosomes

Fitness	**AA**	**Aa**	**aa**
Population 1	1	1+hs	1+s
Population 2	1+s	1+(1-h)s	1

Fitness Model of Haploid Chromosomes

Fitness	**A**	**a**
Population 1	1	1+s
Population 2	1+s	1



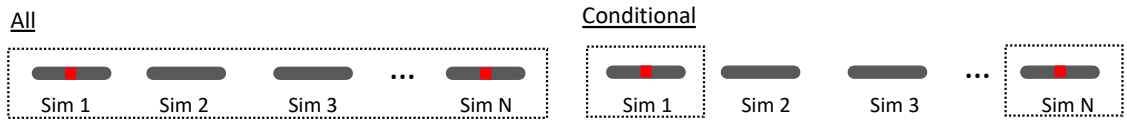


Figure 1. Overview of simulation approach for evaluating faster-haplodiploid effects on genomic differentiation. A. Fitness models for diploid chromosomes (diploid autosomes and haplodiploid females) and haploid chromosomes (haplodiploid males). This is a parallel dominance model in which the fitness of heterozygotes depends on the dominance of allele *a*, which is assumed to be the same in both populations. Population 1 is the population where the derived allele is beneficial. **B**. Overview of stochastic simulations under an isolation-with-migration model. An ancestral population with an effective size of $2N_e$=1500 gene copies (i.e., a haplodiploid locus with 500 females and 500 males or a diploid locus with 375 females and 375 males) of a 500-kb chromosome (dark grey bar) evolves for 10,000 generations to reach mutation-drift equilibrium. This population then splits into two equally sized populations. A divergently selected site at position 250-kb (red line) with two alleles (*A* and *a*) is introduced at the time of split, with an initial frequency q_0 of allele *a* in both populations. The populations evolve for T_{div}=2000 generations, experiencing symmetric migration at a constant rate. For each parameter combination, 1000 simulations were run. **C**. Two approaches were used to summarize simulation results. In the "all"-simulations approach, mean F_{ST} is computed across all simulations (i.e., including simulations in which the derived allele *a* was lost due to drift). In the "conditional" approach, mean F_{ST} is computed only using simulations for which the derived allele *a* was retained.

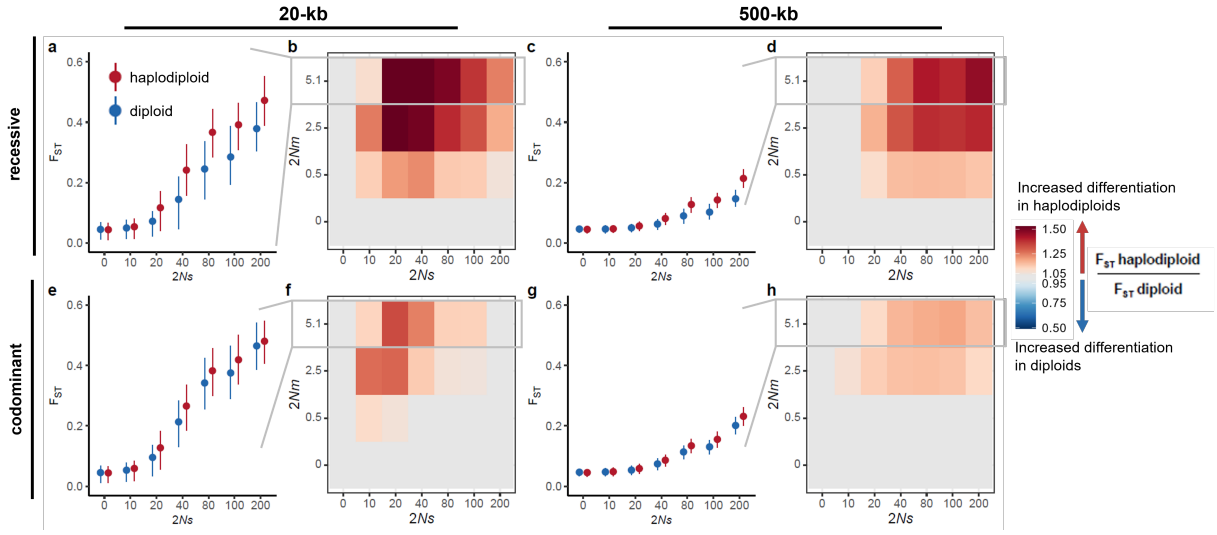


Figure 2. Faster-haplodiploid effects as a function of strength of divergent selection, migration rate and dominance. (a, e, c, g) Differentiation (F_{ST}) for haplodiploids and diploids with *2Nm*=5.1 and varying selective coefficients (*2Ns*) for different window sizes with selected site in the middle [**(a, e)** 20-kb, **(c, g)** 500-kb], and different dominance coefficients [**(a, c)** recessive (*h*=0.01) and **(e, g)** codominant (*h*=0.50)]. The points correspond to mean F_{ST} and the whiskers to the interquantile range based on 1000 simulations. **(b, d, f, h)** Heatmap of the ratio of haplodiploid to diploid (H/D) mean F_{ST} for different combinations of selective coefficients and migration rates for different window sizes [**(b, f)** 20-kb and **(d, h)** 500-kb], and dominance coefficients [**(b, d)** recessive (*h*=0.01) and **(f, h)** codominant (*h*=0.50)]. Results were obtained with 1000 simulations for each parameter combination with an initial frequency q_0=0.10, sampling 20 females from each population.

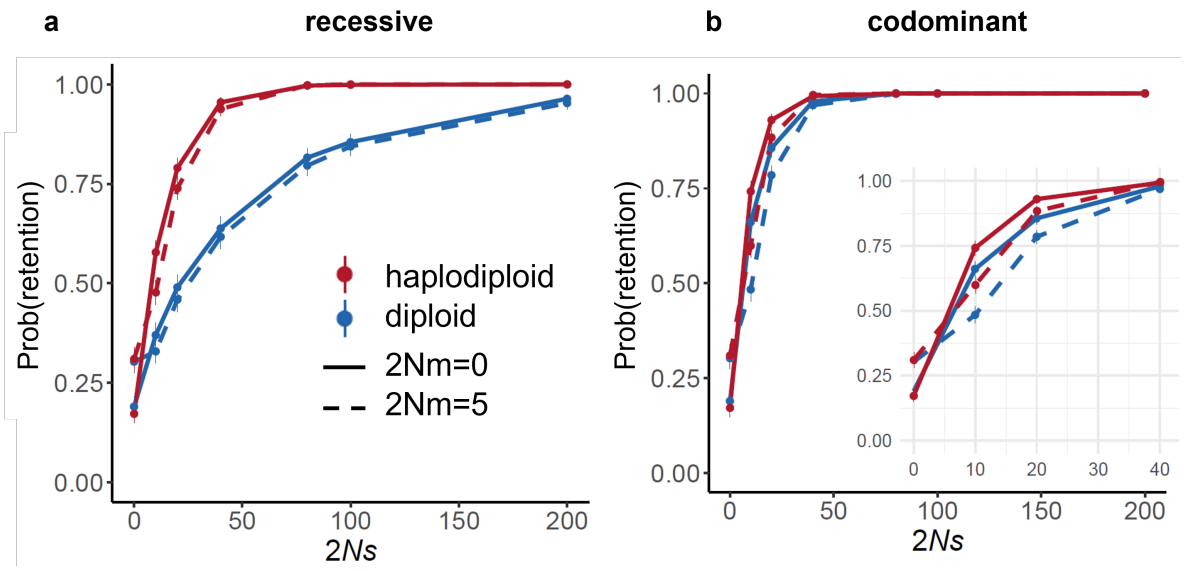


Figure 3. Effect of haplodiploidy on retaining the derived allele at the site under divergent selection. Probability of retaining - Prob(retention) - the derived allele *a* for haplodiploid and diploid simulations at the two extremes of migration rates considered: no migration (2*Nm*=0; solid lines) and high migration (2*Nm*=5.1, dotted lines), for **a)** recessive (h=0.01), and **b)** codominant (h=0.50) mutations. In **b)**, inset shows a zoom for *2Ns* values between 0 and 40. Probability of allele retention is calculated as the proportion out of 1000 simulations that retained the derived allele *a* in population 1 (where *a* is favored). 95% confidence intervals are Clopper-Pearson CI for proportions. These results are for an initial allele frequency of 0.10 (q_0=0.10).

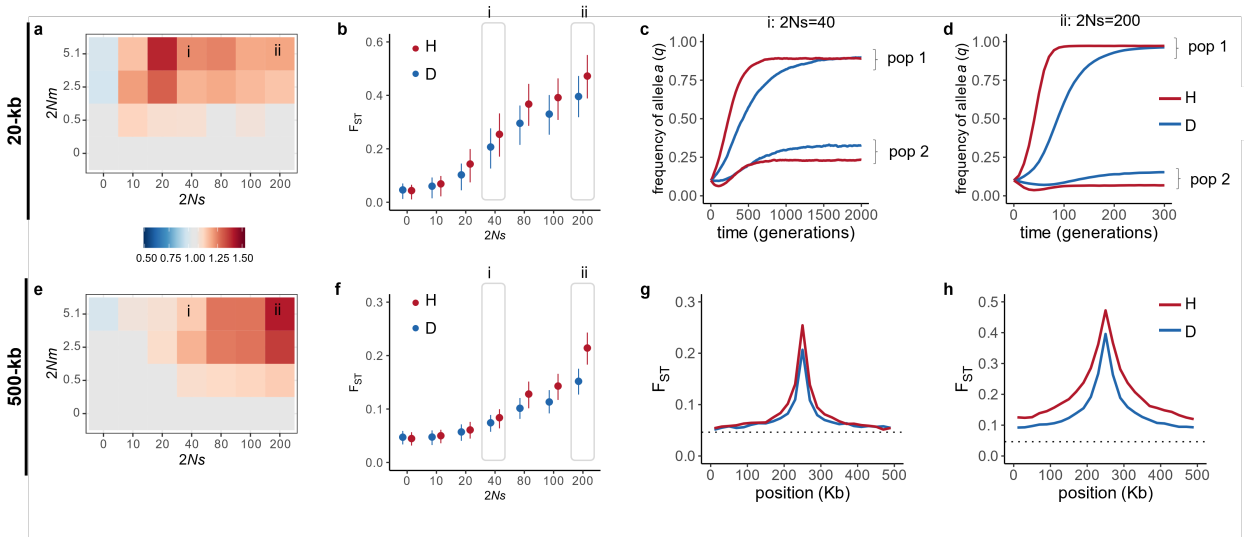


Figure 4. Faster-haplodiploid effects for a recessive (*h*=0.01) derived allele after removing the effect of differential allele loss. (a, e) Heatmap of ratio of haplodiploid (H) to diploid (D) mean F_{ST} for a combination of selective coefficients and migration rates for different window sizes with the selected site at the middle: **(a)** 20-kb and **(e)** 500-kb. Labels **i** and **ii** indicate the two cases selected to illustrate allele trajectories and scans of differentiation. **(b, f)** F_{ST} for haplodiploids and diploids with high migration (2*Nm*=5.1) and varying selective coefficients (2*Ns*), for different window sizes: **(b)** 20-kb and **(f)** 500-kb. Points correspond to mean F_{ST} and whiskers to interquantile ranges. Labels **i** and **ii** indicate the two cases selected to illustrate allele trajectories and scans of differentiation. **(c, d)** Trajectories of allele frequencies at the site under divergent selection in both populations, for: **(c)** moderate selection (2*Ns*=40), and **(d)** strong selection (2*Ns*=200). Note that the time scale is different because equilibrium differentiation is reached faster under strong selection. **(g, h)** Scan of mean F_{ST} along the 500-kb chromosome in non-overlapping 20-kb windows, obtained for: **(g)** moderate selection (2*Ns*=40), and **(h)** strong selection (2*Ns*=200). Mean and interquantile F_{ST} are based on the simulations out of 1000 that kept the derived allele *a* in population 1 at the site under divergent selection. Results are for simulations with an initial frequency q_0=0.10, sampling 20 females from each population.

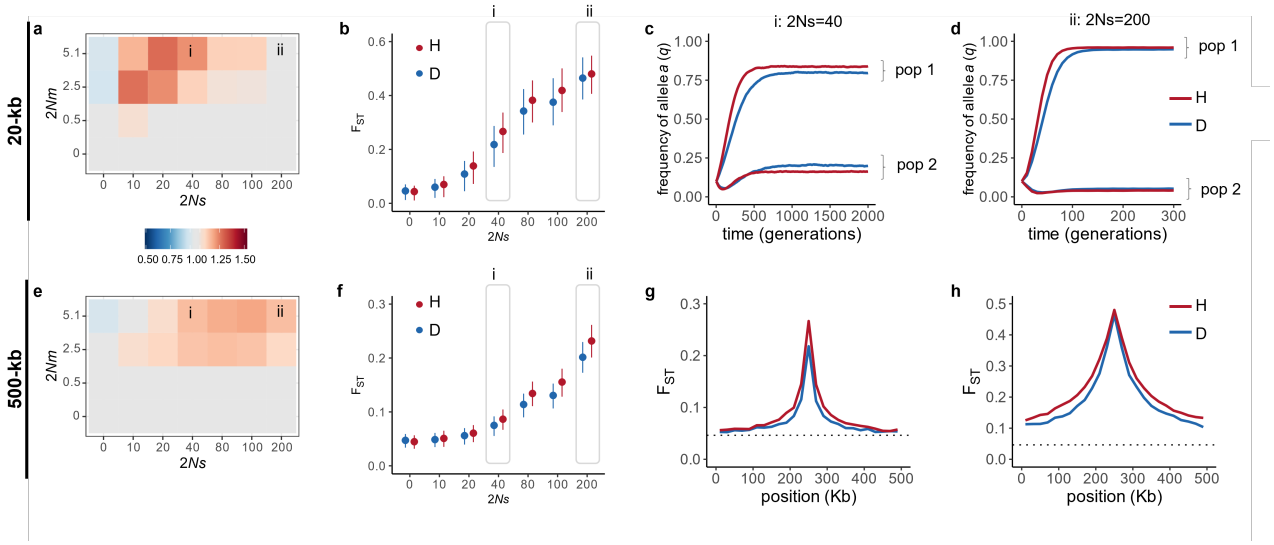


Figure 5. Faster-haplodiploid effects for a codominant (*h*=0.50) derived allele after removing the effect of differential allele loss. (a, e) Heatmap of ratio of haplodiploid (H) to diploid (D) mean F_{ST} for a combination of selective coefficients and migration rates for different window sizes with the selected site at the middle: **(a)** 20-kb and **(e)** 500-kb. Labels **i** and **ii** indicate the two cases selected to illustrate allele trajectories and scans of differentiation. **(b, f)** F_{ST} for haplodiploids and diploids with high migration (2*Nm*=5.1) and varying selective coefficients (2*Ns*), for different window sizes: **(b)** 20-kb and **(f)** 500-kb. Points correspond to mean F_{ST} and whiskers to interquantile ranges. Labels **i** and **ii** indicate the two cases selected to illustrate allele trajectories and scans of differentiation. **(c, d)** Trajectories of allele frequencies at the site under divergent selection in both populations, for: **(c)** moderate selection (2*Ns*=40), and **(d)** strong selection (2*Ns*=200). Note that the time scale is different because equilibrium differentiation is reached faster under strong selection. **(g, h)** Scan of mean F_{ST} along the 500-kb chromosome in non-overlapping 20-kb windows, obtained for: **(g)** moderate selection (2*Ns*=40), and **(h)** strong selection (2*Ns*=200). Mean and interquantile F_{ST} are based on the simulations out of 1000 that kept the derived allele *a* in population 1 at the site under divergent selection. Results are for simulations with an initial frequency q_0=0.10, sampling 20 females from each population.

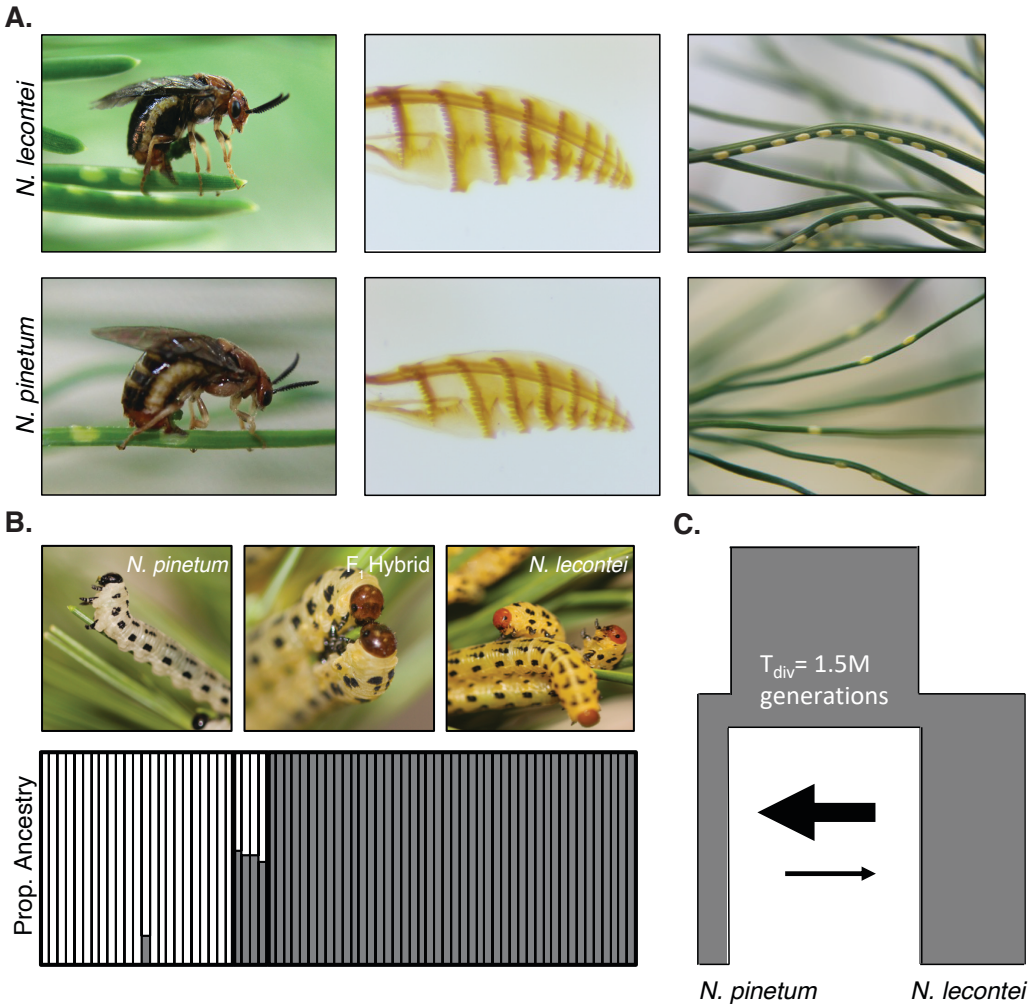
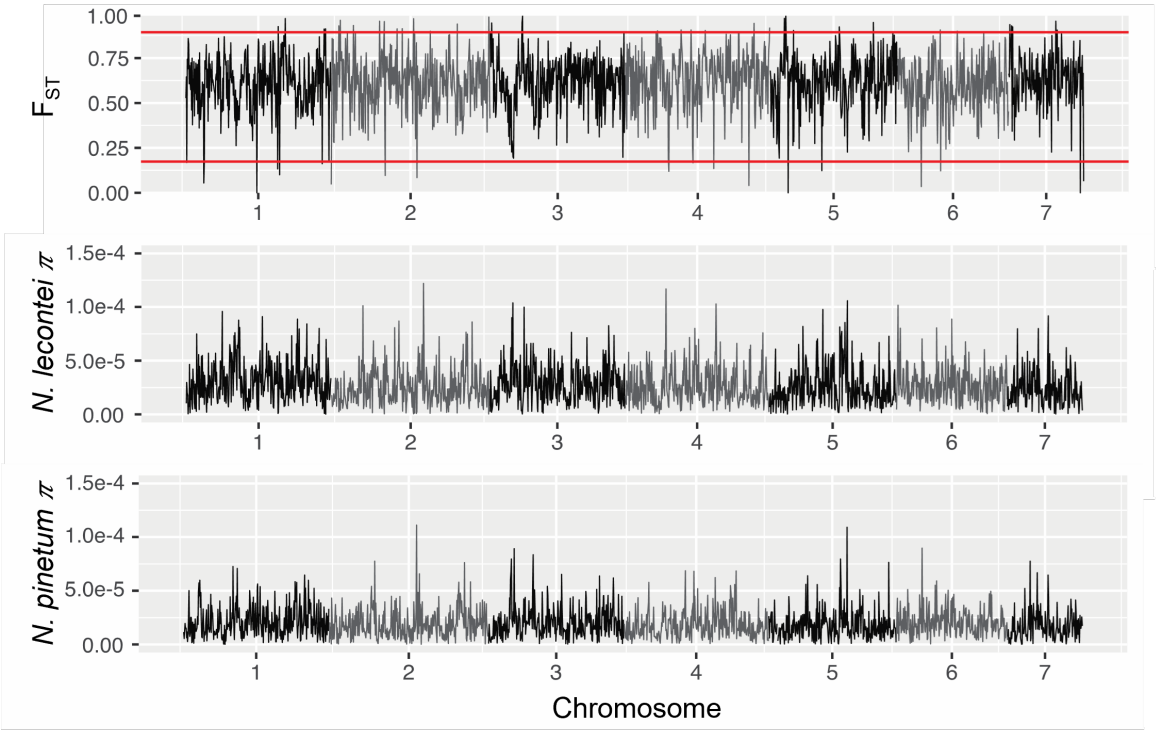


Figure 6. Divergent selection and divergence-with-gene-flow in pine sawflies (*N. pinetum* and *N. lecontei*). A. *N. pinetum* and *N. lecontei* differ in multiple oviposition traits, such as host preference, oviposition stance, ovipositor morphology, and oviposition pattern resulting in strong extrinsic postzygotic isolation. These oviposition traits along with additional host-use adaptations likely result in multiple independent regions of the genome experiencing divergent selection. **B.** Representative pictures of *N. pinetum*, *N. lecontei*, and F_1 hybrid larvae above an Admixture plot (K=2) of individuals sampled from Kentucky. *N. pinetum* ancestry is in white; *N. lecontei* ancestry is in grey. Lab-reared (N=1) and field-caught (N=3) hybrids are genetically admixed with approximately half of their ancestry coming from each species. **C.** *N. pinetum* and *N. lecontei* have diverged with continuous but asymmetric gene flow. Diagram is based on an estimated demographic model for this species pair, with width of boxes proportional to population size and width of arrows is proportional to migration rate (see Table 1).

1174



1175

1176 **Figure 7. Genome scans of differentiation and diversity for *N. pinetum* and *N. lecontei.***
1177 Genetic differentiation (F_{ST}) and genetic diversity (nucleotide diversity π) for *N. lecontei* and *N.*
1178 *pinetum* calculated in 100-kb windows. The red lines mark the 95% confidence interval obtained
1179 from data simulated under neutrality and the demographic model estimated for this species pair.

1180

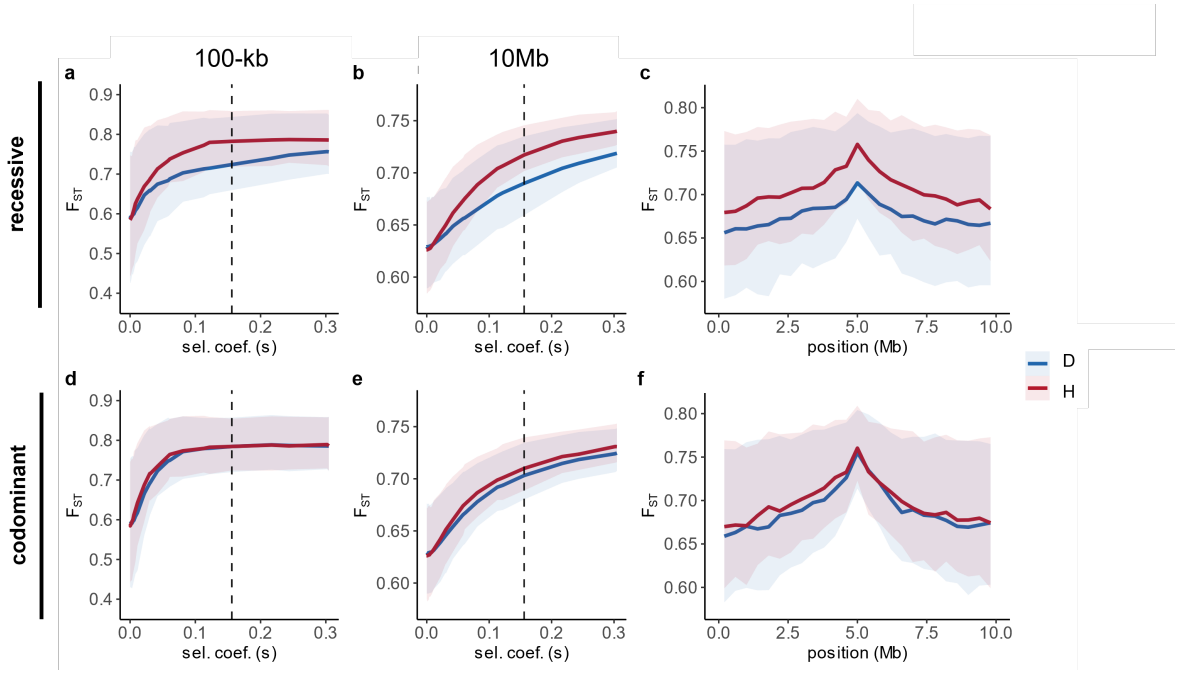


Figure 8. Effect of haplodiploidy and divergent selection on differentiation under inferred demographic history of *Neodiprion* sawflies. Results from simulations performed assuming a sex-ratio with a proportion of 0.7 females and 0.3 males. **(a-b, d-e)** Mean F_{ST} and interquantile range for diploid and haplodiploid populations as a function of selective coefficient for recessive **(a-b)** and codominant **(d-e)** mutations, for different window sizes centered at the selected site **(a,d)** 100Kb, **(b,e)** 10Mb. **(c, f)** Genome scan of F_{ST} for diploid and haplodiploid for recessive **(c)** and codominant **(f)** mutations. Results for $s = 0.16$ with initial frequency of 0.1. Solid line corresponds to mean F_{ST} and shaded area indicates interquantile 0.25-0.75 range. Dashed lines in (**a-b**, **d-e**) indicate the selective coefficient used in genome scan shown in **c, f**.

1198