OXFORD
UNIVERSITY PRESS | Molecular Biology and Evolution

Deleterious variation mimics signatures of genomic incompatibility and adaptive introgression

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **Deleterious variation mimics signatures of genomic incompatibility and adaptive**

2 **introgression**

3

4 Bernard Y. Kim[1], Christian D. Huber[1], Kirk E. Lohmueller[1,2,3,]*

5

6 **Affiliations:**

7 [1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA

8 90095, USA.

9 [2]Interdepartmental Program in Bioinformatics, University of California, Los Angeles, CA 90095,

10 USA.

11 [3]Department of Human Genetics, David Geffen School of Medicine, University of California,

12 Los Angeles, CA 90095, USA.

13 *Corresponding author: E-mail: klohmueller@ucla.edu

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

Abstract

While it is appreciated that population size changes can impact patterns of deleterious variation in natural populations, less attention has been paid to how population admixture affects the dynamics of deleterious variation. Here we use population genetic simulations to examine how admixture impacts deleterious variation under a variety of demographic scenarios, dominance coefficients, and recombination rates. Our results show that gene flow between populations can temporarily reduce the genetic load of smaller populations, especially if deleterious mutations are recessive. Additionally, when fitness effects of new mutations are recessive, between-population differences in the sites at which deleterious variants exist creates heterosis in hybrid individuals. This can lead to an increase in introgressed ancestry, particularly when recombination rates are low. Under certain scenarios, introgressed ancestry can increase from an initial frequency of 5% to 30-75% and fix at many loci, even in the absence of beneficial mutations. Further, deleterious variation and admixture can generate correlations between the frequency of introgressed ancestry and recombination rate or exon density, even in the absence of other types of selection. The direction of these correlations is determined by the specific demography and whether mutations are additive or recessive. Therefore, it is essential that null models include both demography and deleterious variation before invoking reproductive incompatibilities or adaptive introgression to explain unusual patterns of genetic variation.

Introduction

There is tremendous interest in quantifying the effects that demographic history has had on the patterns and dynamics of deleterious variation and genetic load (Kirkpatrick and Jarne 2000; Gazave et al. 2013; Lohmueller 2014a; Lohmueller 2014b; Henn et al. 2015; Brandvain and Wright 2016; Simons and Sella 2016). Several studies have suggested that recent human demography has had little impact on load (Simons et al. 2014; Do et al. 2015) while others have suggested weak, but subtle differences between human populations (Casals et al. 2013; Fu et al. 2014; Gravel 2016; Henn et al. 2016; Pedersen et al. 2017). All of these studies have typically focused on how population size changes, such as expansions and bottlenecks, have affected deleterious variation. Other types of complex demography, however, have received considerably less attention.

In particular, gene flow may be important for shaping patterns of deleterious variation. Population admixture, or hybridization between closely related species, appears to be quite common in nature (Payseur and Rieseberg 2016) and has a significant role in shaping human genomes (Wall and Yoshihara Caldeira Brandt 2016). Gene flow alone can subtly change the

69   effects of selection on deleterious variation (Gravel 2016), but should have notable fitness

70   consequences if deleterious variation is distributed differently between admixing populations. For

71   example, Neanderthals likely had a higher genetic load than coincident human populations due to

72   the former's smaller long-term population size (Do et al. 2015; Harris and Nielsen 2016; Juric et

73   al. 2016). As a result, gene flow from Neanderthals into the ancestors of modern humans could

74   have increased the genetic load of some human populations by 0.5% (Harris and Nielsen 2016),

75   and selection should have removed deleterious Neanderthal ancestry over time. In contrast,

76   domesticated species likely have increased genetic load due to domestication bottlenecks and

77   hitchhiking with artificially selected variants (Marsden et al. 2016; Liu et al. 2017; Moyers et al.

78   2017). Then, gene flow from their wild counterparts should alleviate the genetic load of

79   domesticated species and increased levels of introgression should be observed (e.g. Wang L,

80   Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J, Hufford M, unpublished data,

81   https://www.biorxiv.org/content/early/2017/03/07/114579, last accessed Nov. 13, 2017).

82   If effects of gene flow can be modulated by the consequences of deleterious variation,

83   selection on deleterious variants may provide an alternative explanation for patterns of

84   introgression that are usually attributed to processes such as the evolution of reproductive

85   incompatibility or adaptive introgression. Patterns such as the depletion of introgressed ancestry

86   in regions of low recombination (Sankararaman et al. 2014) might be instead partially explained

87   by purifying selection removing deleterious variation and partially by reproductive

88   incompatibilities (Sankararaman et al. 2014; Harris and Nielsen 2016; Juric et al. 2016; Vernot et

89   al. 2016). Rampant adaptive introgression may create the opposite pattern where there is

90   increased amounts of introgressed ancestry in regions of low recombination (e.g. Pool 2015;

91   Corbett-Detig and Nielsen 2017). Alternatively, this pattern could be due to selection favoring

92   introgressed haplotypes from a larger population that carries fewer deleterious variants. The

93   general extent of these effects in nature and their magnitude given realistic parameter

94   combinations remains to be studied.

95   Another significant issue with many models incorporating deleterious variation is the

96   assumption that fitness effects are strictly additive. A considerable proportion of strongly

97   deleterious new mutations are likely to be fully recessive or partially recessive (Simmons and

98   Crow 1977; Agrawal and Whitlock 2011; Huber CD, Durvasula A, Hancock AM, Lohmueller

99   KE, unpublished data, https://www.biorxiv.org/content/early/2017/08/31/182865, last accessed

100   Nov. 13, 2017). In addition, if some proportion of deleterious recessive variants is private to a

101   population, admixed populations should experience heterosis since deleterious variants are more

102   likely to be found in a heterozygous state (Crow 1948). As a result, heterosis can increase

3

103 effective migration rates between populations (Ingvarsson and Whitlock 2000) and may play a

104 significant role in determining the fitness of highly structured populations (Whitlock et al. 2000).

105 Heterosis may also increase introgression and the probability that introgressed ancestry will

106 persist in an admixed population, even if the introgressed ancestry contains more deleterious

107 variation (Harris and Nielsen 2016). The extent to which heterosis contributes to increases in the

108 frequency of introgressed ancestry and confounds inference of adaptive introgression is also not

109 well understood.

110   The objective of this study is to develop a clearer picture of the effect of deleterious

111 variants on fitness and the dynamics of introgression in admixing populations. Further, we aim to

112 understand how inferences of selection on introgressed ancestry are impacted by deleterious

113 variation. Previous simulation and empirical work have shown that for at least some systems,

114 deleterious variation is a significant factor modulating gene flow (Harris and Nielsen 2016; Juric

115 et al. 2016; Wang et al. 2017), but few studies have investigated these questions outside of

116 demographic models specific to a species or system. Our present study fills this void by

117 presenting a series of simulations utilizing demographic models that generalize biological

118 scenarios of interest. We include a realistic distribution of fitness effects for both additive and

119 recessive mutation models. In addition, we examine how the relationship between introgressed

120 ancestry and recombination rates, or functional content, is determined by the underlying

121 demography.

122

123 **Results**

124 *Forward simulations*

125   We used SLiM 4.2.2 (Haller and Messer 2017) to simulate admixture in the presence of

126 deleterious variation. In general, the underlying demographic model was a population split model

127 with an ancestral population size (N_a) of 10,000 diploids, where a single pulse of admixture

128 occurs at an initial proportion of 5%, in one direction and for one generation, at some time after

129 the split (**Figure 1, Table 1**). Throughout, we will refer to the subpopulation from which gene

130 flow occurs as the *source* subpopulation, and the subpopulation which receives gene flow as the

131 *recipient* subpopulation. Furthermore, we will refer to ancestry in the recipient subpopulation that

132 originated in the source subpopulation as *introgression-derived ancestry*. We tracked the

133 frequency of introgression-derived ancestry in simulations by placing marker mutations in one

134 subpopulation, and use p_I to denote the estimated total proportion of ancestry in the recipient

135 subpopulation that is introgression-derived. The sizes of the subpopulations were varied as

136 described in the forthcoming sections and in **Figure 1** and **Table 1**.

4

137   Unless specified otherwise, we simulated approximately 5 Mb of sequence, with

138 randomly generated genic structure (see Methods). The mutation rate (μ) was set at a constant rate

139 of 1.5×10^{-8} per bp per generation. All simulated mutations were either neutral or deleterious, and

140 only nonsynonymous mutations had nonzero selection coefficients. Deleterious mutations had

141 selection coefficients (s) were all drawn from the same distribution of fitness effects (Kim et al.

142 2017). In other words, the selection coefficients did not depend on the population in which

143 mutations occurred. No positively selected mutations were simulated in any of our models. We

144 simulated additive (h=0.5) and recessive (h=0.0) mutations separately. Additive fitness effects

145 were computed by multiplicatively calculating fitness at each locus. Recessive fitness effects

146 were computed additively, but only at homozygous loci. All fitness effects were computed

147 multiplicatively across loci. To assess the effect of recombination rate, we also varied the per-

148 base pair per chromosome recombination rate, r, between sets of simulations ($r \in \{10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}\}$).

149 $^8, 10^{-9}\}$). See **Methods** for additional details on the simulations and calculation of fitness.

150   In each simulation, we recorded the fitness of each subpopulation, the proportion of

151 ancestry that is introgression-derived, and other measures of genetic load (**Figures S1-S3**) at

152 different time points. Differences in subpopulation fitness are presented as (w_R/w_S), which

153 represents the relative differences in load between the recipient and source subpopulation.

154 Subpopulation fitnesses are presented separately in **Figure S1**.

155

The effect of admixture on deleterious variation

157   Our first step was to look at only the effects of gene flow on deleterious variation when

158 the two subpopulations had identical sizes. Given that the populations were of identical size, they

159 both should contain similar amounts of deleterious variation and similar genetic loads. Here, we

160 simulated under a population split model with a size of 10,000 diploids in the ancestral and both

161 divergent subpopulations (Model 0, **Figure 1**), where 20,000 ($2N_A$) generations after the

162 population split a single pulse of admixture occurred.

163   Under the additive fitness model, the fitness of the recipient subpopulation does not

164 change due to gene flow (Model 0, **Figure 2**). At the time point of admixture, the two

165 subpopulations have similar fitnesses due to identical population sizes ($w_S \approx w_R$, **Figures 2** and

166 **S1**), although many deleterious variants will be private to each subpopulation. In addition, gene

167 flow does not change the mean number of deleterious variants per haplotype, since any

168 introgression-derived haplotypes carry, on average, the same number of derived variants as

169 background haplotypes (**Figure S2**).

170   When deleterious mutations are recessive instead of additive, admixture is predicted to

171 generate heterosis in hybrid individuals, particularly when some proportion of segregating

172 variants is private to each subpopulation (Crow 1948; Whitlock et al. 2000; Harris and Nielsen

173 2016). Indeed, 20,000 generations after the population split, many of the deleterious variants in

174 our simulations should be private to one subpopulation (**Figure S4**). At the time of gene flow,

175 both subpopulations have declined in fitness to a similar degree ($w_R \approx w_S$, Model 0, **Figure 2**).

176 After admixture, homozygosity in the recipient population is immediately decreased (**Figure S3**),

177 with a corresponding increase in the fitness of the recipient population (**Figures 2** and **S1**).

178 Importantly, the number of derived deleterious variants per haplotype in the recipient

179 subpopulation is unaffected by gene flow (**Figure S2**). After the initial increase in fitness due to

180 admixture, the decline in fitness following admixture is slow, such that fitness will be greater than

181 the pre-admixture value for many generations following admixture (Model 0, h=0, **Figure S1**).

182 Notably, the fitness increase conferred by heterosis is the most pronounced ($\approx$2.5% increase) and

183 lasts the longest (>10,000 generations after admixture) in simulations with low amounts of

184 recombination (r=10^{-9}). This occurs for two reasons. First, fitness declines more quickly when

185 recombination rates are low (Muller 1964), resulting in stronger selection against non-admixed

186 individuals. Second, introgression-derived haplotypes remain intact, maximizing the heterotic

187 advantage of admixed individuals.

188   The effect of selection on deleterious variation also influences the fraction of ancestry

189 that is introgressed in the recipient population. Considering that immediately following

190 admixture, 5% of the ancestry in the recipient population is introgression-derived (p_I =5%), it

191 follows that in the absence of allele frequency changes due to selection the expected proportion of

192 introgression-derived ancestry should remain E[p_I]=5%. Our simulations show that this is the

193 case when fitness effects are additive and both subpopulations have similar genetic loads (Model

194 0 in **Figure 3**). Therefore, selection does not favor haplotypes of a particular ancestry, and the

195 long-term $p_I \approx$ 5%.

196   If deleterious mutations are recessive, introgression-derived ancestry increases in

197 frequency as protective haplotypes rise in frequency through heterosis (p_I > 5%; Model 0, h=0,

198 **Figure 3**). The increase in p_I is inversely related to the recombination rate. Specifically, the

199 increase in p_I is greatest (average p_I~35% at 10,000 generations after the split, **Figure 3**) for the

200 lowest recombination rate r=10^{-9}, (**Figures 2** and **S1**). This effect is still observed even when the

201 simulated recombination rate is greater than 10^{-9}, but the magnitude of the effect is far less

202 pronounced, with increases to $p_I \approx$ 6-9%. These results also corroborate studies which showed that

203 heterosis can increase effective migration rates (Ingvarsson and Whitlock 2000) or enhance the

6

204 introgression of linked deleterious variants (Harris and Nielsen 2016). However, our results

205 additionally show that heterosis should be greater when recombination rates are low.

206

207 *Short-term bottlenecks have little influence on the dynamics of introgression*

208 Next, we investigated how short-term bottlenecks might affect fitness and patterns of

209 introgression by adding a short bottleneck into one subpopulation of the split model (Model 1,

210 **Figure 1**). Specifically, we added a bottleneck of size 1,000 diploids, or a 10-fold reduction in

211 population size, in the recipient subpopulation for the 50 generations immediately preceding the

212 admixture event. All population sizes remained at 10,000 diploids otherwise.

213 In this model, the additive genetic load was insensitive to the short bottleneck (Model 1,

214 **Figures 2** and **S1**). Although some proportion of the deleterious variants is lost in the bottleneck,

215 the average number of deleterious variants per haplotype is unchanged (Lohmueller 2014b;

216 Simons et al. 2014; Simons and Sella 2016), and the additive load in each population is the same

217 ($w_S \approx w_R$, **Figures 2** and **S1**). Therefore, gene flow has no effect on the additive load or the

218 number of deleterious variants per haplotype (**Figure S2**) in the recipient population, and there is

219 no selection on introgressed ancestry in this model.

220 We found broadly similar patterns when deleterious mutations were recessive, with some

221 important differences. Our simulations show that the bottleneck causes an additional ~2% decline

222 in the recipient population's mean fitness prior to admixture (Model 1, **Figure S1**) due to a small

223 increase in homozygosity (**Figure S3**). However, the average number of deleterious variants per

224 haplotype is unaffected by any selection against the increased proportion of homozygotes (Model

225 1, **Figure S2**). Fitness increases quickly after admixture, but consistently remains slightly less

226 than the model without a bottleneck, suggesting that any change in load due to an increased

227 number of homozygous sites is mostly cancelled out by the increased heterozygosity that results

228 from admixture (**Figure S3**). The magnitude of the fitness increase from admixture is again

229 inversely related to the recombination rate, in a manner similar to that in the model without a

230 bottleneck (Model 0).

231 The frequency of introgression-derived ancestry was largely unaffected by the short

232 bottleneck. When fitness effects were additive, the average frequency of introgressed ancestry

233 remained at the initial admixture proportion of $p_I = 5\%$, 10,000 generations after the admixture

234 event (Model 1, **Figure 3**). When fitness effects were recessive, introgression-derived ancestry

235 increased in frequency by carrying protective alleles, similar to the simulations with identical

236 subpopulation size. The same inverse relationship to the recombination rate was also observed. In

237 the case of recessive mutations, the long-term frequency of introgression-derived ancestry (e.g. p_I

7

238    $\approx 33\%$ for $r=10^{-9}$, Model 1 in **Figure 3**) was similar but slightly lower than the model without a

239   bottleneck (e.g. $p_I \approx 35\%$ for $r=10^{-9}$, Model 0 in **Figure 3**).

240

241   *Long-term population contractions greatly influence the dynamics of introgression*

242   At equilibrium, smaller populations will have a greater reductions of fitness due to

243   deleterious variation than larger populations (Kimura et al. 1963). Therefore, a long-term

244   reduction in population size should have different implications for fitness and the fate of

245   introgression-derived ancestry than the short bottleneck described above.

246   To investigate the effect of long-term differences in population size and subsequent gene

247   flow on patterns of deleterious alleles, we simulated a split model with the addition of a long-term

248   reduction in population size. Immediately following the split, the size of one subpopulation was

249   reduced 10-fold to 1,000 diploids (Models 2 and 3, **Figure 1**). After 20,000 additional

250   generations, gene flow occurred in a single generation at an admixture proportion of 5%. In

251   Model 2, the direction of admixture is from the small into the large population, and in Model 3

252   the direction of admixture is from the large into the small population.

253   As a consequence of long-term differences in population size, the additive fitness of the

254   small subpopulation is 7-10% less than that of the large subpopulation (Models 2 and 3 in

255   **Figures 2** and **S1**) at the time of admixture. In the additive fitness model, gene flow from the

256   small to the large subpopulation (Model 2) resulted in a small fitness decrease (<1%) in the

257   recipient subpopulation's fitness, but had little effect on the average number of derived

258   deleterious variants per haplotype (**Figure S2**). Gene flow from the large to the small

259   subpopulation (Model 3) resulted in small increases (<1%) in fitness in the recipient

260   subpopulation's fitness, except (~5% increase) when recombination was low (**Figure S1**). In this

261   case, selection for less deleterious haplotypes resulted in an overall decrease of the average

262   number of deleterious variants per haplotype, but because the recipient population remained at a

263   small size after admixture, load continued to accumulate afterward at the same rate (**Figures 2**

264   **and S1**).

265   When deleterious mutations were recessive, the effect of admixture on the recipient

266   population's fitness was determined both by differences in genetic load between populations and

267   heterosis from admixture. Immediately prior to admixture, the recipient population's fitness was

268   approximately 10-30% greater (in Model 2) or less (in Model 3) than the fitness of the source

269   subpopulation (**Figures 2** and **S1**) due to long-term differences in population size. Gene flow

270   from the small to the large population (Model 2) only increased the recipient subpopulation's

271   fitness by 1-2% ($h=0$, **Figure S1**), and thus did not drastically affect the trajectory of w_R/w_S

272   (**Figure 2**). However, gene flow from the large to the small population (Model 3) drastically and

273   immediately increased the fitness of the recipient population (h=0, **Figures 2** and **S1**). Because

274   the population size of the recipient subpopulation remained small in Model 3, fitness continued to

275   decline quickly after admixture. In both models, increased fitness in the recipient population after

276   admixture is a consequence of a decrease in the mean number of homozygous deleterious sites

277   per individual due to admixture. These effects are more pronounced for Model 3 because of the

278   higher number of homozygous sites per individual in the large subpopulation (on average about

279   110; **Figure S3**) compared to the small subpopulation (on average about 50-60; **Figure S3**). In

280   other words, more private recessive deleterious variants are masked by introgressing haplotypes

281   in Model 3, despite the fact that introgressing haplotypes carry a slightly larger number of

282   deleterious variants. Importantly, gene flow had little effect on the mean number of deleterious

283   variants per haplotype (**Figure S2**). Again, our simulations show that the fitness changes from

284   admixture should be largest when recombination rates are low.

285   Due to these differences in fitness between the small and large subpopulations, the

286   frequency of introgression-derived ancestry in the recipient population changed noticeably for

287   both the additive and recessive models (Models 2 and 3 in **Figure 3**). When fitness effects were

288   additive, these changes were directly linked to selection on introgressed variation. If introgressed

289   haplotypes carried a larger deleterious burden (i.e. came from the smaller population as in Model

290   2), introgressed ancestry linked to deleterious alleles was removed by selection (long-term p_I≈0-

291   4%). On the other hand, if introgressed haplotypes had a smaller deleterious burden (i.e. came

292   from the larger population as in Model 3), linked introgression-derived ancestry increased in

293   frequency due to selection for haplotypes with fewer deleterious variants (long-term p_I≈6-38%).

294   Again, we observe that the magnitude of this effect is inversely related to the recombination rate.

295   Specifically, the proportion of introgression-derived ancestry decreased or increased at the

296   greatest magnitude in simulations with low recombination, and the least in simulations with a

297   high recombination rate (Models 2 and 3 in **Figure 3**).

298   When fitness effects were recessive, the frequency of introgression-derived ancestry in

299   the recipient subpopulation were determined by heterosis and differences in genetic load between

300   subpopulations. Gene flow from the large to the small subpopulation (Model 3) resulted in slight

301   (p_I=7%, r=10^{-6}) to drastic increases (p_I=51%, r=10^{-9}) in the average proportion of introgression-

302   derived ancestry (Model 3, **Figure 3**). However, admixture from the small to the large (Model 2)

303   population resulted in smaller (p_I≈6-17%) proportions of introgression-derived ancestry (Model

304   2, **Figure 3**). Additionally, the increase in frequency from selection on recessive variation

305   opposes the effect of selection on additive variation. In the case of Model 3, heterosis and

9

306   differences in load drive the frequency of introgression-derived ancestry in the same direction,

307   but in the case of Model 2, these factors work in opposite directions. The rate of change of the

308   proportion of introgression-derived ancestry was greatest in simulations with low recombination

309   rates ($r=10^{-9}$).

310

311   *Long-term population contractions with subsequent population recovery greatly influence the*

312   *dynamics of introgression*

313   To investigate what occurs when differences in genetic load exist between two

314   subpopulations, but where the strength of selection is increased in the recipient population post-

315   admixture, we simulated another split model where the recipient subpopulation is subjected to a

316   long-term bottleneck but recovers to its original size (Model 4, **Figure 1**). Specifically, the model

317   we simulated was a split model where the recipient subpopulation is reduced to 1,000 diploids

318   immediately after the population split and for 20,000 generations afterwards. At that time point,

319   gene flow occurred in a single generation at an admixture proportion of 5%. Immediately

320   following the admixture event, the recipient subpopulation was restored to its original size

321   (10,000 diploids).

322   Like the model where the recipient subpopulation size remained small after the split

323   (Model 3), immediately prior to admixture, the recipient population's fitness was approximately

324   7-10% less than the fitness of the source population when mutations were additive and 10-30%

325   less than the fitness of the source subpopulation when mutations were recessive (**Figures 2** and

326   **S1**). When mutations were additive, admixture and population recovery result in a gradual

327   increase in the recipient subpopulation's fitness (h=0.5, **Figures 2** and **S1**). When mutations were

328   recessive, admixture rapidly increases the fitness of the recipient subpopulation to at least 90%

329   and up to 95% of its pre-bottleneck fitness, then continues to recover slowly thereafter (h=0.0,

330   **Figures 2** and **S1)**. In both cases, gene flow results in substantial decreases in the number of

331   deleterious variants per haplotype (**Figure S2**) as well as the number of homozygous deleterious

332   sites per individual (**Figure S3**). The quickest increase in fitness is again observed in the

333   simulations with the lowest recombination rate ($r=10^{-9}$; **Figures 2** and **S1**).

334   Large changes in subpopulation fitness are tied to the largest changes in the frequency of

335   introgression-derived ancestry in the recipient subpopulation (Model 4 in **Figure 3**). When

336   mutations are additive, the fraction of introgression-derived ancestry quickly increases from an

337   initial p_I=5% to p_I≈7-68%, with higher long-term fractions of introgression-derived ancestry

338   occurring at lower recombination rates. When mutations are recessive, the fraction of

339   introgression-derived ancestry has increases to p_I≈10-74%. Thus, the population expansion after

340   the bottleneck drives a rapid increase in the frequency of introgressed ancestry in the recipient

341   population.

342

343   *Increasing population split times enhances the effect of heterosis*

344   In models where fitness effects are recessive, heterosis, although modulated by

345   subpopulation differences in genetic load, determines the fitness effects of admixture and the

346   direction of selection on introgression-derived ancestry. However, in these simulations, we have

347   fixed the split time before admixture at $2N$ generations, a substantially long time for private

348   deleterious variation to accumulate within each subpopulation. To further examine the

349   relationship between split time and selection on introgression-derived ancestry, we simulated split

350   models while varying the split time before gene flow. Specifically, we simulated gene flow

351   between two populations of equal size; one where the recipient population experienced a brief

352   bottleneck of 1,000 diploids for the 50 generations immediately before the admixture event; and

353   one where the recipient population's size was reduced to 1,000 diploids immediately after the

354   split until a single pulse of admixture at 5%, where it then recovered to the original size of 10,000

355   diploids. These models are analogous to Models 0 and 4 (**Figure 1**), respectively, the only

356   difference being the split time before admixture. The recombination rate was set to $r=10^{-9}$ in these

357   simulations.

358   **Figure 4** depicts the long-term proportion of introgressed ancestry, p_I, 10,000 generations

359   after the single pulse of admixture for these three models varying the amount of time between the

360   split and admixture (t_s). First, we found that across our range of simulated t_s, p_I always increases

361   monotonically with t_s regardless of the underlying demography (**Figure 4**). This can be attributed

362   to the fact that longer split times result in more deleterious variation being unique to each

363   population (**Figure S4**), enhancing heterosis after admixture. Second, we found as a bottleneck

364   increases in duration, differences in genetic load become a significant contributor to long-term p_I.

365   At a split time and thus bottleneck time of 20,000 generations, the effects of heterosis and the

366   bottleneck increase long-term p_I nearly 2-fold relative to heterosis alone (compare Model 0 to

367   Model 4 in **Figure 4**).

368

369   *A realistic map of chromosomal structure and recombination rates*

370   Thus far we have shown how selection on deleterious variation can affect the dynamics

371   of introgression-derived ancestry when the recombination rate was set to a single value in each

372   set of our simulations. The correlation between introgression-derived ancestry and genomic

373   features such as local recombination rates or exon density are often considered potential evidence

374   for selection against introgression-derived ancestry due to genomic incompatibility or

375   maladaptive alleles (Brandvain et al. 2014; Sankararaman et al. 2014; Pool 2015; Aeschbacher et

376   al. 2017; Corbett-Detig and Nielsen 2017) or adaptive introgression (Corbett-Detig and Nielsen

377   2017). Selection on deleterious variation is one possible confounder of these patterns (Harris and

378   Nielsen 2016; Juric et al. 2016; Wang et al. 2017).

379   To investigate how the correlation of introgression-derived ancestry with genomic

380   features is influenced by deleterious variation under different demographic models, we simulated

381   a 100 Mb segment of human chromosome 1, using realistic exon definitions and a recombination

382   map defined on a 10 kb scale (see **Methods**). Unlike the simulations we described previously, we

383   fixed the exon definitions and recombination map to be the same in all simulations. We simulated

384   under three of the split models described previously (**Figure 1**): Model 0, Model 2, and Model 4,

385   separately for recessive and additive fitness effects. Only deleterious mutations were simulated.

386   At the end of each simulation, we split the chromosome into non-overlapping 100 kb windows

387   and computed the frequency of introgression-derived ancestry, exon density, and the average

388   recombination rate in each window.

389   The average genomic landscape of introgression for 100 simulation replicates varied

390   widely across demographic models and dominance coefficients (**Figure 5** or see **Figure S5** for a

391   representative single simulation replicate). In general, simulations with recessive mutations

392   always showed a genome-wide increase in the frequency of introgressed ancestry, and

393   simulations with additive fitness were dependent upon the demographic model. In the model with

394   equal population sizes (Model 0), we observed no average change in the frequency of

395   introgression-derived ancestry when mutations were additive, but when mutations were recessive

396   we observed a large overall genome-wide increase in the frequency of introgression-derived

397   ancestry (**Figure 5**), with several regions that are at high frequency or fixed in a single simulation

398   replicate (**Figure S5**). Importantly, this increase in frequency is only due to selection on recessive

399   mutations and local variation in recombination rate, since no positively selected mutations were

400   simulated. In the model where introgressed haplotypes contained a larger deleterious burden

401   (Model 2), we observed an overall depletion of introgressed ancestry consistent with the effects

402   of purifying selection upon introgressed ancestry (**Figure 5**). However, for the model with

403   recessive mutations, the effects of heterosis were strong enough such that many genomic regions

404   showed average increases in frequency of 5-10% in our simulations. Importantly, Harris and

405   Nielsen (2016) predicted that heterosis would increase the frequency of introgressed ancestry by

406   only a few percent, but our simulations with a similar demographic model show that low

407   recombination rates can greatly enhance the increase in frequency of introgressed ancestry.

408   Finally, when we simulated the introgression of haplotypes from a population with lower genetic

409   load (Models 3 and 4), we observed drastic, genome-wide increases in the average frequency of

410   introgressed ancestry in the recipient subpopulations (**Figure 5**) as well as many fixed loci in

411   individual simulations (**Figure S5**), regardless of whether fitness effects of mutations were

412   additive or recessive. For example, local regions of the simulated chromosome showed an

413   increase in introgressed ancestry from an initial frequency of 5% up to 70-80% frequency. This is

414   the type of signature that would be unlikely to be generated under neutral demographic models

415   and could be attributed to adaptive introgression.

416   It is also notable that the frequency of introgression-derived ancestry (p_I) in each window

417   appears to be driven by exon density, or the local concentration of sites at which deleterious

418   mutations can occur. For recessive mutations, p_I is greatly increased on the left-hand side of the

419   simulated chromosome, which tends to be more gene-rich than the right-hand side of the

420   chromosome (**Figures 5** and **S5**). For additive mutations, the pattern is not as straightforward.

421   When introgressed ancestry is depleted in the recipient population (Model 2), this depletion is

422   strongest in the left-hand side of the figure compared to the right-hand side. In Model 4, where

423   introgressed ancestry increases in the recipient population, the increase is greatest in the most

424   gene-dense part of the chromosome.

425   We show the correlations that deleterious mutations generate between genomic features

426   and the frequency of introgressed ancestry, measured in 100 kb windows, in **Table 2**. In the

427   model of equal population sizes (Model 0), the frequency of introgression-derived ancestry is not

428   significantly related to the recombination rate or exon density when fitness is additive, but is

429   positively correlated to exon density when fitness effects are recessive. When introgressed

430   ancestry comes from the population with higher load (Model 2), the frequency of introgression-

431   derived ancestry is positively correlated to the recombination rate and negatively correlated to

432   exon density when fitness is additive. When fitness effects are recessive in this model, the

433   frequency of introgressed ancestry is only positively correlated to exon density. Lastly, when

434   introgressed ancestry comes from a larger population having a lower deleterious burden than the

435   recipient population (Model 4), a negative correlation is observed between the frequency of

436   introgression-derived ancestry and recombination rate, and a positive correlation between the

437   frequency of introgression-derived ancestry and exon density. For the last model, these

438   correlations are observed for both models of additive and recessive mutations.

439

440

441

Discussion

442

443   We have shown, through simulations, that deleterious variation can greatly influence the

444 dynamics of introgression between admixing populations, in markedly different directions for

445 different demographic scenarios and modes of dominance. In addition, the recombination rate is a

446 key parameter that determines the way in which deleterious variants accumulate between

447 populations, how selection acts on admixed individuals, how selection acts at a locus with

448 admixed ancestry, and ultimately the dynamics of introgression-derived ancestry.

449   Our work further demonstrates how demography can shape patterns of deleterious

450 variation in different populations. Previous studies have examined the role of population size

451 changes (Kirkpatrick and Jarne 2000; Lohmueller et al. 2008; Simons et al. 2014; Balick et al.

452 2015; Gravel 2016) and serial founder effect models (Peischl and Excoffier 2015; Henn et al.

453 2016) on deleterious variation. Interpreting how differences in the distribution of deleterious

454 variation impact fitness has been a contentious issue (Fu et al. 2014; Lohmueller 2014b; Do et al.

455 2015; Henn et al. 2015; Simons and Sella 2016). In this study, we observed that admixture can

456 increase the fitness of the recipient population, sometimes drastically if the source population is

457 of larger long-term effective size and thus carries lower genetic load. Generally, gene flow is

458 observed to drive smaller, subtle changes in fitness. Nevertheless, the influx of new alleles can

459 result in a rearrangement of deleterious variants in an admixed population (**Figures S2** and **S3**),

460 and even subtle or no changes in fitness or the number of derived alleles per individual can lead

461 to significant shifts in the frequency of introgressed ancestry (e.g. see Model 0, *h=0*, in **Figure 3)**.

462 These effects can be long lasting, persisting for thousands of generations in some of our

463 simulations (**Figures 2, 3, S1**). If gene flow or hybridization is a significant feature of a study

464 population, studies concerning load should consider the fitness consequences of admixture as

465 well as population size changes.

466   That dynamics of introgression-derived ancestry can thus be driven by deleterious

467 variation also is particularly relevant for the study of gene flow between populations or species.

468 Patterns of introgression between hybridizing species are often asymmetric, vary across the

469 genome, and can be driven by demography at expansion fronts (Currat et al. 2008), dispersal

470 processes (Amorim et al. 2017), or by natural selection. However, when natural selection is

471 implicated as driving changes in introgression-derived ancestry, processes such as genomic

472 incompatibility or adaptive introgression are usually invoked to explain variation in introgression

473 across the genome. We have shown that standing deleterious variation, rather than differences in

474 selection on alleles transplanted onto a new genomic background or new environment, has the

475 potential to explain some of these patterns.

476   To the best of our knowledge, only a few studies have considered the contribution of

477   selection on deleterious variation to observed patterns of introgression (Ingvarsson and Whitlock

478   2000; Gravel 2016), and mostly in specific systems (Harris and Nielsen 2016; Juric et al. 2016;

479   Wang et al. 2017; Schumer M, Xu C, Powell D, Durvasula A, Skov L, Holland C, Sankararaman

480   S, Andolfatto P, Rosenthal G, Przeworski M, unpublished data,

481   https://www.biorxiv.org/content/early/2017/11/01/212407, accessed Nov. 11, 2017). It is possible

482   that deleterious variation may play a similar role in other species, in particular those that have

483   experienced reductions in population size or range expansions associated with human

484   demography. For example, work by Pool (2015) and Corbett-Detig and Nielsen (2017) on

485   mapping the introgression of African ancestry into North American populations of *Drosophila*

486   *melanogaster* show that the frequency of introgressed African ancestry is negatively correlated

487   with the recombination rate. This is particularly notable because the opposite relationship is

488   observed for Neanderthal ancestry in humans (Sankararaman et al. 2014) and in hybridizing

489   populations of swordtail fish (Schumer M, Xu C, Powell D, Durvasula A, Skov L, Holland C,

490   Sankararaman S, Andolfatto P, Rosenthal G, Przeworski M, unpublished data,

491   https://www.biorxiv.org/content/early/2017/11/01/212407, accessed Nov. 11, 2017), where

492   regions of the genome with high recombination rate are enriched for introgressed ancestry. Our

493   simulations show that selection on deleterious variants can plausibly explain these opposing

494   patterns in different species.

495   Originating in Africa, *D. melanogaster* has serially colonized the world in association

496   with humans (Stephan and Li 2006; Duchen et al. 2013), including parts of North America

497   approximately a hundred years ago (Keller 2007). If derived populations of *D. melanogaster* have

498   accumulated differences in deleterious variation due to bottlenecks or increased drift at the

499   expansion front (Peischl and Excoffier 2015; Henn et al. 2016), selection in North American

500   populations may favor introgressed African ancestry simply because this ancestry came from a

501   population with a larger long-term effective size, thus carrying fewer deleterious variants. If

502   recessive deleterious variation also creates heterosis in admixed individuals of North American

503   populations of *D. melanogaster*, the effects of heterosis and population size will be synergistic,

504   further enhancing introgression in genomic regions of low recombination.

505   Importantly, we do not claim that selection on standing deleterious variation explains all

506   the patterns of introgression in *D. melanogaster* or any other species, but rather that it is a

507   plausible alternative explanation that is important to consider when testing hypotheses about the

508   nature of selection on gene flow. In swordtail fish, recombination rates are positively correlated

509   with the frequency of introgressed ancestry even when the source population has a smaller

510 effective population size than the recipient population (Schumer M, Xu C, Powell D, Durvasula

511 A, Skov L, Holland C, Sankararaman S, Andolfatto P, Rosenthal G, Przeworski M, unpublished

512 data, https://www.biorxiv.org/content/early/2017/11/01/212407, accessed Nov. 11, 2017). This

513 pattern is not explained by models that only include deleterious variation.

514   Because selection on additive and recessive variation can act in complementary or

515 opposing directions, our study also highlights the fundamental importance of understanding the

516 distribution of selection coefficients and the relationship to dominance coefficients in natural

517 populations (the *h-s* relationship). In this study, we simulated additive and recessive mutations

518 separately, using the same distribution of fitness effects, so that we could demonstrate the effect

519 of only changing the dominance coefficient. However, strongly deleterious new mutations are

520 more likely to be partially or fully recessive (Simmons and Crow 1977; Agrawal and Whitlock

521 2011; Huber CD, Durvasula A, Hancock AM, Lohmueller KE, unpublished data,

522 https://www.biorxiv.org/content/early/2017/08/31/182865, last accessed Nov. 13, 2017) and so

523 new mutations are likely to have varying degrees of recessivity.

524   The underlying demographic model will determine how these additive and recessive new

525 mutations should interact after gene flow. For example, the introgression of deleterious

526 haplotypes should be assisted by recessive deleterious mutations but impeded by additive ones,

527 leading to uncertainty about the overall contribution of the effects of deleterious variation in

528 certain scenarios, such as Neanderthal to human admixture (Harris and Nielsen 2016). In other

529 scenarios, selection on additive and recessive variants should operate in the same direction. There

530 is evidence for this type of introgression of wild teosinte into maize (Wang et al. 2017), but it is

531 difficult to disentangle the effects of selection on additive versus recessive variation.

532   Our simulations reveal that the recombination rate also influences the dynamics of

533 introgressed ancestry in the presence of deleterious variation. Models of Hill-Robertson

534 interference (Hill and Robertson 1966; Keightley and Otto 2006) predict that deleterious

535 mutations will not be removed as effectively in regions of the genome with low recombination

536 rates because they may be linked to the non-deleterious alleles at other sites. We observed the

537 opposite effect in our simulated admixed populations. Specifically, in our simulations, the fitness

538 in the admixed population increased the most for the lowest recombination rates, suggesting that

539 deleterious mutations were most effectively eliminated when recombination rates were the

540 lowest. To understand this effect, it is important to realize that selection for a haplotype will be

541 most effective when all alleles on a haplotype have fitness effects in the same direction. For

542 example, if introgression-derived ancestry carries fewer deleterious variants than the other

543 haplotypes in the recipient population, selection will act to increase the frequency of the

544   protective alleles contained within the introgressed ancestry. This applies directly to our

545   simulations of admixture since immediately following an admixture event, all the protective or

546   deleterious variants are found on the same haplotype. Higher rates of recombination will result in

547   the selected variants being shuffled onto different haplotypes, decreasing the efficacy of selection.

548   An important objective of genomic studies of hybridization is to identify loci that are

549   adaptively introgressed and to ascertain the overall importance of introgression to adaptive

550   evolution (Racimo et al. 2015). Genomic regions that contain introgressed haplotypes at high

551   frequency are considered likely candidates for adaptive introgression (Huerta-Sánchez et al. 2014;

552   Gittelman et al. 2016; Racimo et al. 2017; Richards and Martin 2017), but we have shown that

553   deleterious variation can generate similar patterns, even in the absence of new beneficial

554   mutations and local adaptation. Model-based statistical approaches that compare summary

555   statistics computed for a particular window of the genome to a simulated null distribution that

556   only accounts for demography may thus be misled by deleterious variation. Again, it may be

557   difficult to differentiate heterosis due to the masking of deleterious recessive alleles from

558   heterozygote advantage at introgressed loci, despite the fact that these are two very different

559   evolutionary processes with dramatically different biological interpretations.

560   Our results argue that new null models are needed in studies seeking to identify

561   candidates of adaptive introgression. These new null models should include deleterious genetic

562   variation, as well as complex demography. In order for these models to accurately capture the

563   dynamics of deleterious variation, they should also include realistic parameters for the DFE of

564   deleterious mutations and the relationship between dominance coefficients and selection

565   coefficients. Lastly, the new null models should also include realistic models of the variation in

566   recombination rate across the genome, as recombination rate is a key determinant of the dynamics

567   of introgression (**Figure 3**). Failure to consider deleterious variation in a realistic way in studies

568   of admixing populations or hybridizing species can mislead inferences about evolutionary

569   processes acting on the genome.

570

571   **Materials and Methods**

572   *Simulation details*

573   All simulations were performed with SLiM 4.2.2 (Haller and Messer 2017). The

574   sequences from simulations with randomly generated chromosome structure were approximately

575   5 Mb in length (see below). The simulated sequences generated using features from human

576   chromosome 1 were fixed to be exactly 100 Mb in length. Every simulation contained intergenic,

577   intronic, and exonic regions, but only nonsynonymous new mutations experienced natural

17

578 selection. The per base pair mutation rate was constant and set to $\mu=1.5\times10^{-8}$. Within coding

579 sequences, we set nonsynonymous and synonymous mutations to occur at a ratio of 2.31:1 (Huber

580 et al. 2017). The selection coefficients (*s*) of new nonsynonymous mutations were drawn from a

581 gamma-distributed DFE with shape parameter 0.186 and expected selection coefficient E[*s*] = -

582 0.01314833 (Kim et al. 2017). We chose to simulate additive (*h*=0.5) and recessive (*h*=0) fitness

583 separately, using the same DFE for *s* for each simulation, to allow the effects of dominance to be

584 directly compared.

585 Importantly, we chose to discard from our simulations, and therefore from calculations of

586 fitness, mutations that were fixed in the ancestral or both subpopulations. Although fixed

587 deleterious variants contribute to the overall genetic load of finite populations, they will have no

588 effect on the relative differences between admixing subpopulations and no effect on the dynamics

589 of introgression-derived ancestry. Therefore, each fitness calculation does not reflect the true

590 fitness, but rather the fitness components that are relevant during gene flow.

591 An admixture event in SLiM is handled by modifying the way the parents in each

592 generation are chosen (SLiM manual 5.2.1). For example, at an admixture proportion of 5% the

593 recipient population reproduces as follows. Five percent of the parents of the recipient population,

594 in that generation, are chosen from the source population, and 95% of the parents are chosen from

595 the recipient population.

596

597 *Simulations with randomly generated chromosomal structure*

598 Unless specified otherwise, the chromosomal structure of each simulation was randomly

599 generated by drawing exon lengths from $Lognormal(\mu = \log(50), \sigma^2 = \log(2))$, intron lengths

600 from $Lognormal(\mu = \log(100), \sigma^2 = \log(1.5))$, and the length of noncoding regions from

601 $unif(100,5000)$, following the specification in the SLiM 4.2.2 manual (7.3), which is modeled

602 after the distribution of intron and exon lengths in Deutsch and Long (1999). The per-base pair

603 recombination rate (*r*) was fixed in each simulation, but we varied *r* between different sets of

604 simulations where $r \in \{10^{-6},10^{-7},10^{-8},10^{-9}\}$. Lastly, we simulated 200 replicates for each set of

605 simulations, of each specific *h* and *r*.

606

607 *Simulations with fixed, realistic chromosomal structure*

608 In simulations with fixed chromosomal structure (**Figure 6**), we fixed the structure to 100

609 Mb randomly chosen from human genome build GRCh37, chromosome 1 (chr1:5,005,669-

610 105,005,669). The exon ranges were defined by the GENCODE v14 annotations (Harrow et al.

611   2012) and the sex-averaged recombination map by Kong et al. (2010), averaged over a 10 kb

612   scale.

613

614   *Avoiding heterosis in the additive fitness model*

615   Computing fitness as additive at a locus but multiplicative across loci creates artificial

616   heterosis in admixed individuals. This occurs because the product of a fitness decrease will

617   reduce fitness less than the sum of a fitness decrease. As an example, assume two deleterious

618   alleles are in a single individual, each with selection coefficient *s* where *s*<0. If they are found as

619   a single homozygous site, the fitness decrease is usually computed as (1+2*s*). If they are found in

620   two heterozygous sites, the fitness would be computed as (1+*s*)². This second quantity is larger

621   than the first by *s*². Because admixed individuals are more likely to carry deleterious alleles as

622   heterozygotes than non-admixed individuals, the fitness of the admixed individuals will always

623   be higher than a non-admixed individual in the above computation of fitness even when the

624   number of deleterious variants per individual is the same.

625   Because our intent with an additive fitness model was to make the fitness effect of each

626   variant independent of its genotypic state, we computed additive fitness as purely multiplicative

627   across all deleterious variants, such that an individual *j* carrying *i* variants each with selection

628   coefficient s_i has fitness w_i:

629   $$w_j = \prod_i (1 + s_i)$$

630   This computation of fitness is approximately equivalent to additive fitness. Recessive fitness

631   effects were computed in the standard manner, that is, as 1+2s_i when homozygous for the

632   deleterious allele and as 1 otherwise.

633

634   *Data availability*

635   All scripts necessary for reproducing the simulations we have presented are available at

636   https://github.com/LohmuellerLab/admixture_load_scripts.

637

638   **Acknowledgements**

639   We thank Jacqueline Robinson, Annabel Beichman, and other members of the

640   Lohmueller Lab for helpful comments throughout the project. This work was supported by the

641   National Institutes of Health (R35 GM119856 to K.E.L.).

642

643
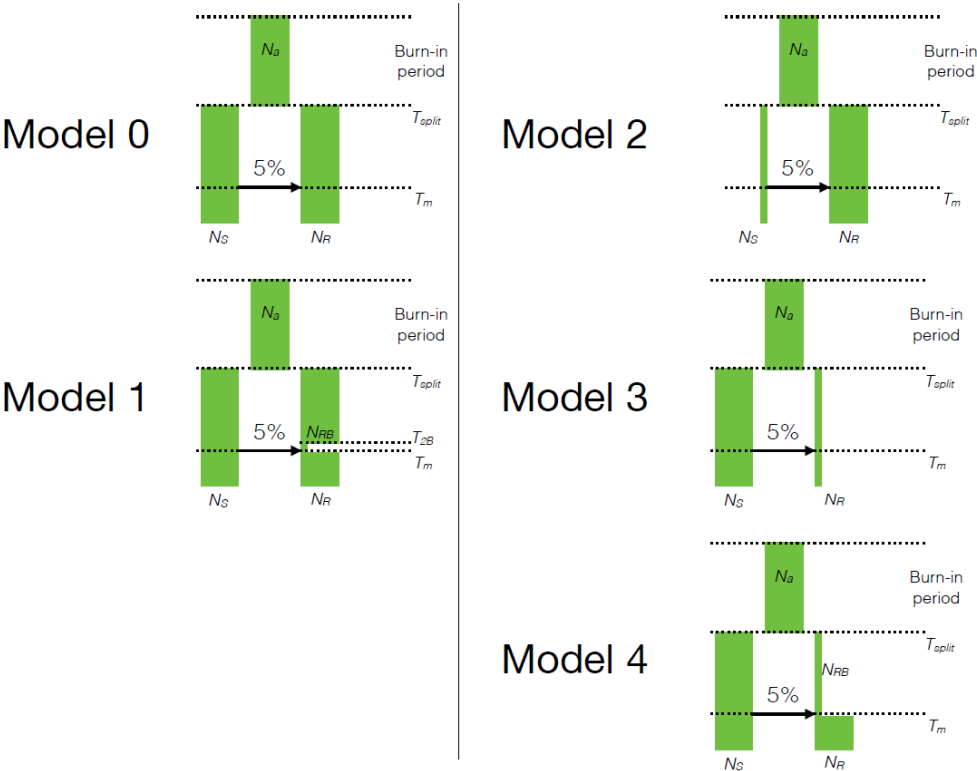
644   **Figures and Tables**

645
646   **Figure 1**. The demographic models used for the simulations. After a burn-in period of 100,000

647   generations, a single population diverged into two subpopulations. The demography of the

648   subpopulations was modified in ways that changed the distribution of deleterious variation. 2,000

649   generations after the split, a single pulse of admixture occurred such that 5% of the ancestry of

650   the recipient population came from the donor population. Arrows in each panel denote the

651   direction of gene flow. The simulation was run for 10,000 additional generations after admixture.

652   Population sizes were changed as shown for each model. See **Table 1** for specific parameter
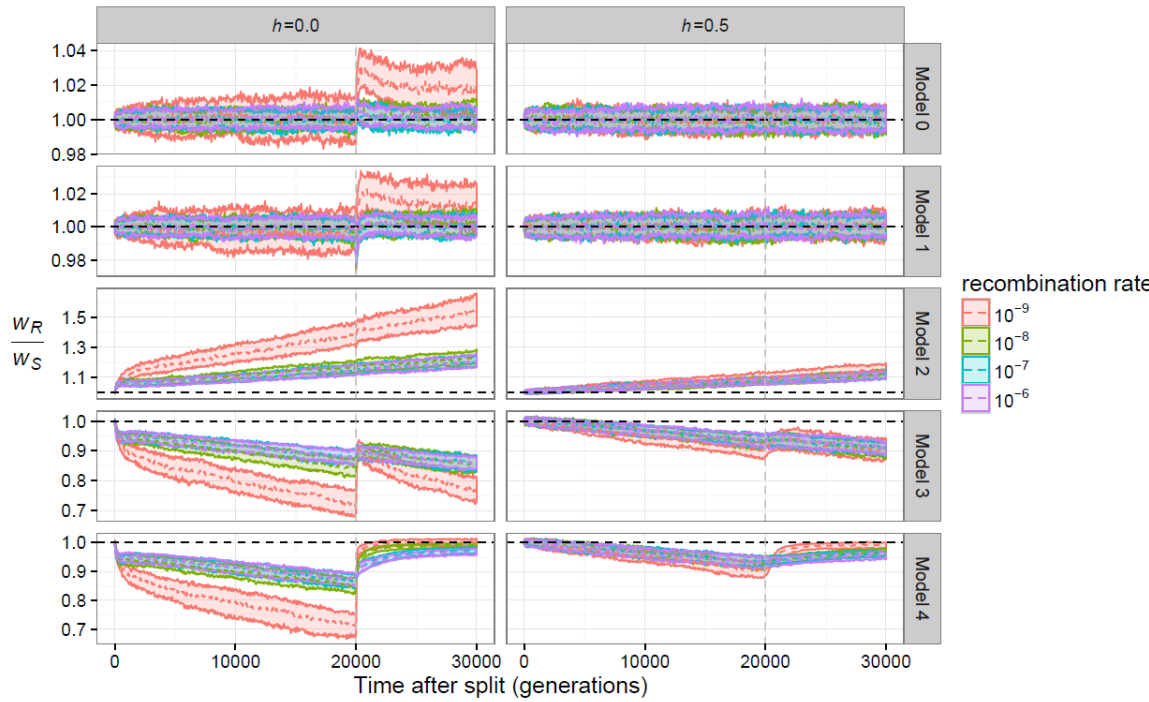
653   values for each model.

654

Figure 2. The change in mean fitness over time due to demography. Each individual plot depicts the ratio of the relative mean fitness of the recipient population (w_R) to the source population (w_S) for the demographic models shown in **Figure 1**. The median (dotted line) and the 25th to 75th percent quantiles are shown for 200 simulation replicates. The vertical grey line depicts the time of gene flow, and the horizontal black line depicts w_R/w_S=1. Different colors denote distinct recombination rates used in the simulations. Left panel denotes recessive mutations (h=0) while the right panel shows additive mutations (h=0.5).
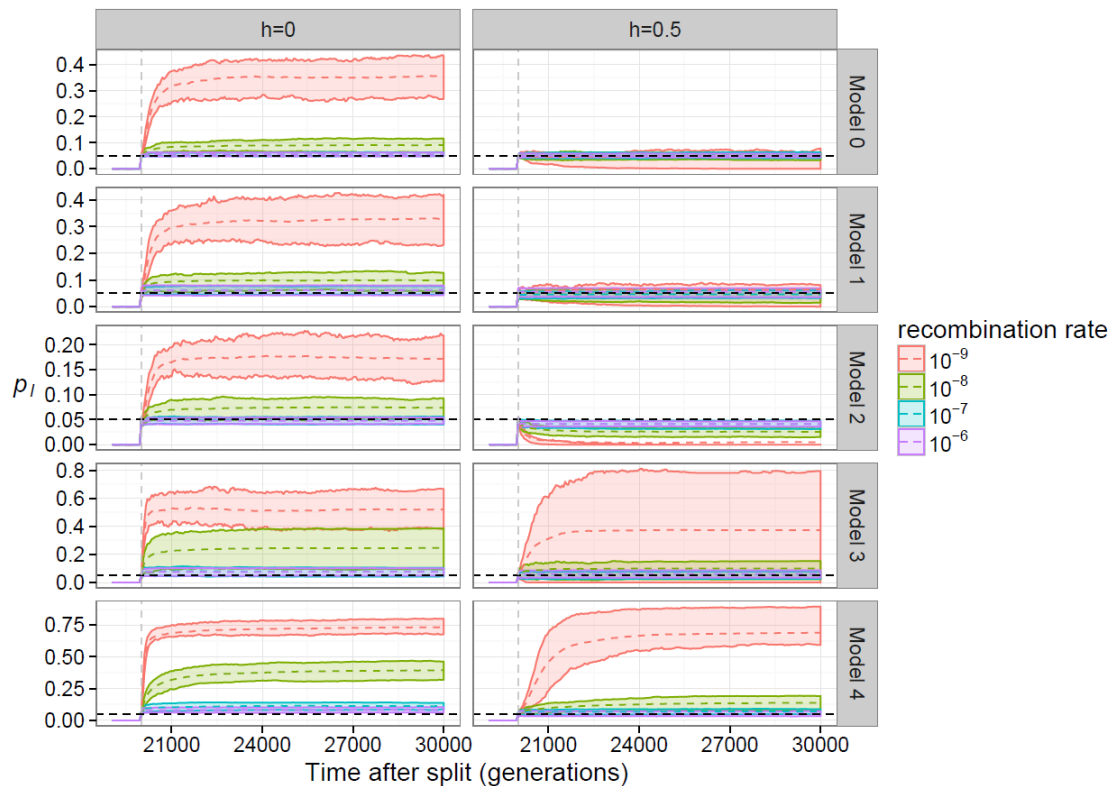
Figure 3. The frequency of introgression-derived ancestry (p_I) in each model. Earlier generations are not shown since p_I=0 prior to admixture. The mean (dotted line) and the 25th to 75th percent quantiles are shown for 200 simulation replicates. The vertical red line depicts the time of gene flow, and the horizontal black line depicts the initial admixture proportion of 0.05. Different colors denote distinct recombination rates used in the simulations. Left panel denotes recessive mutations (h=0) while the right panel shows additive mutations (h=0.5).
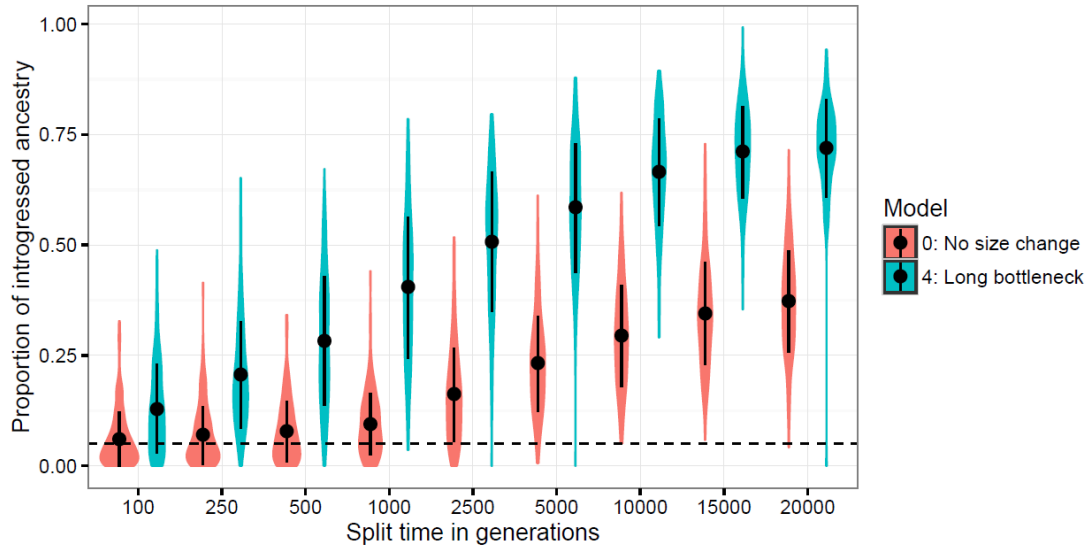
681
682 **Figure 4**. Population split time and population size impact the amount of introgressed ancestry

683 when mutations are recessive. The proportion of ancestry that is introgression-derived, p_I, is

684 shown for 200 simulation replicates and two demographic models (Model 0 and Model 4, refer to

685 **Figure 1**). The recombination rate in all simulations is $r=10^{-9}$ per base pair. Violin plots represent

686 the density while dot and whiskers represent the mean and one standard deviation to either side.

687 The horizontal black line represents the initial admixture proportion of 0.05. Note that as the split

688 time increases, p_I also increases.

689
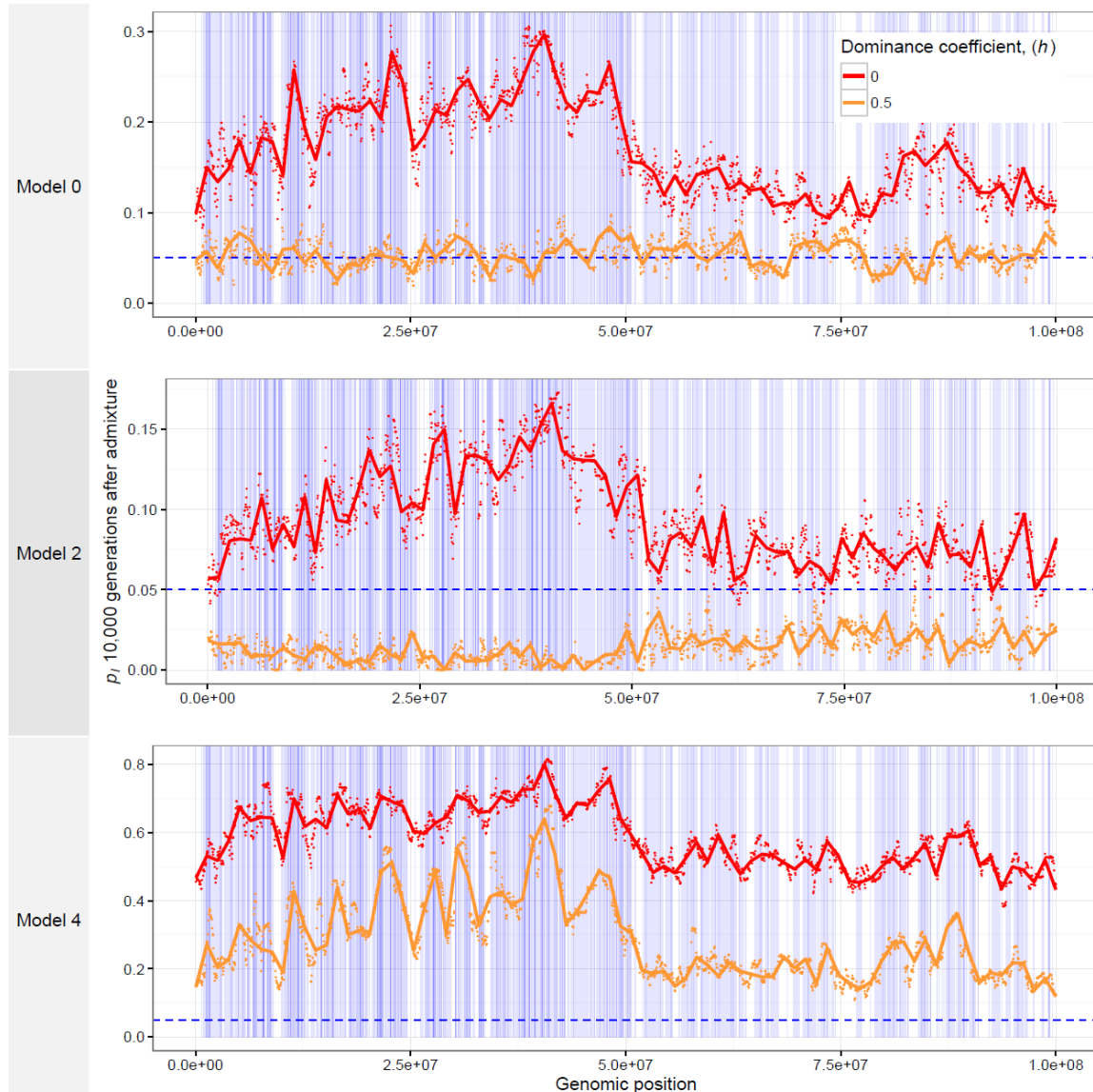
690

691

692

693

694

695

Figure 5. The average genomic landscape of introgression for three demographic models. The frequency of ancestry that is introgression-derived is shown for non-overlapping 100 kb windows in a simulated 100 Mb region of human chromosome 1. The model numbers refer to the models shown in **Figure 1**. Points represent a single value for each 100 kb window and lines are loess curves fitted to the data. The horizontal, blue dashed line represents the initial frequency of introgression-derived ancestry, p_I=0.05. Vertical blue bars represent genes in which deleterious mutations can occur. Red curves denote the results for recessive mutations while orange curves show the results for additive mutations.

708 **Table 1. Demographic parameters of the simulated models shown in Figure 1.**

Model	N_a	N_S	N_{RB}	N_R	T_m	T_{RB}	T_{split}
Model 0	10,000	10,000	--	10,000	1,000	--	3,000
Model 1	10,000	10,000	1,000	10,000	1,000	1,050	3,000
Model 2	10,000	1,000	--	10,000	1,000	--	3,000
Model 3	10,000	10,000	--	1,000	1,000	--	3,000
Model 4	10,000	10,000	1,000	10,000	1,000	--	3,000

709 NOTE.--All population sizes are in diploids and all times in generations from the present.

710 Parameters are defined as follows. N_a: ancestral population size, N_S: size of the source

711 subpopulation, N_{RB}: size of the bottleneck in the recipient population, N_R: size of the recipient

712 population, T_m: time of migration, T_{RB}: time at which the bottleneck in the recipient population

713 began, T_{split}: time at which the subpopulations diverged

714

715

716

717

718 **Table 2. The correlation of recombination rate or exon density with the long-term**
719 **frequency of introgression-derived ancestry in 100 kb windows is affected by dominance**
720 **and demography.**

h	Genomic feature	Mean Spearman's ρ	p-value [a]
		Model 0 [b]	
0.5	recombination rate	0.003601	0.486
0.5	exon density	0.01190	0.0797
0.0	recombination rate	-0.009629	0.0859
0.0	exon density	0.09109	0
		Model 2 [b]	
0.5	recombination rate	0.01186	0.0135
0.5	exon density	-0.02972	0
0.0	recombination rate	0.0002355	0.9635
0.0	exon density	0.06164	0
		Model 4 [b]	
0.5	recombination rate	-0.06495	0
0.5	exon density	0.1743	0
0.0	recombination rate	-0.03889	0
0.0	exon density	0.1190	0

721 [a]approximate p-values indicate the significance of H_1: $\rho \neq 0$ and were estimated with a single

722 sample permutation test. One hundred simulation replicates were permuted 10,000 times.

723 [b]model numbers reference **Figure 1**.

724

725

726 **References**

727 Aeschbacher S, Selby JP, Willis JH, Coop G. 2017. Population-genomic inference of the strength

728 and timing of selection against gene flow. Proc. Natl. Acad. Sci. 114:7061–7066.

729 Agrawal AF, Whitlock MC. 2011. Inferences about the distribution of dominance drawn from

730 yeast gene knockout data. Genetics 187:553–566.

731 Amorim CEG, Hofer T, Ray N, Foll M, Ruiz-Linares A, Excoffier L. 2017. Long-distance

732 dispersal suppresses introgression of local alleles during range expansions. Heredity

733 118:135–142.

734 Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR. 2015. Dominance of Deleterious Alleles

735 Controls the Response to a Population Bottleneck. PLOS Genet. 11:e1005436.

736 Brandvain Y, Kenney AM, Flagel L, Coop G, Sweigart AL. 2014. Speciation and Introgression

737 between *Mimulus nasutus* and *Mimulus guttatus*. PLOS Genet. 10:e1004410.

738 Brandvain Y, Wright SI. 2016. The Limits of Natural Selection in a Nonequilibrium World.

739 Trends Genet. 32:201–210.

740 Casals F, Hodgkinson A, Hussin J, Idaghdour Y, Bruat V, Maillard T de, Grenier J-C, Gbeha E,

741 Hamdan FF, Girard S, et al. 2013. Whole-Exome Sequencing Reveals a Rapid Change in

742 the Frequency of Rare Functional Variants in a Founding Population of Humans. PLOS

743 Genet. 9:e1003815.

744 Corbett-Detig R, Nielsen R. 2017. A Hidden Markov Model Approach for Simultaneously

745 Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data

746 in Samples of Arbitrary Ploidy. PLOS Genet. 13:e1006529.

747 Crow JF. 1948. Alternative Hypotheses of Hybrid Vigor. Genetics 33:477–487.

748 Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive

749 introgression by local genes. Evol. Int. J. Org. Evol. 62:1908–1920.

750 Deutsch M, Long M. 1999. Intron-exon structures of eukaryotic model organisms. Nucleic Acids

751 Res. 27:3219–3228.

752 Do R, Balick D, Li H, Adzhubei I, Sunyaev S, Reich D. 2015. No evidence that selection has

753 been less effective at removing deleterious mutations in Europeans than in Africans. Nat.

754 Genet. 47:126–131.

755 Duchen P, Živković D, Hutter S, Stephan W, Laurent S. 2013. Demographic Inference Reveals

756 African and European Admixture in the North American *Drosophila melanogaster*

757 Population. Genetics 193:291–301.

758 Fu W, Gittelman RM, Bamshad MJ, Akey JM. 2014. Characteristics of Neutral and Deleterious

759 Protein-Coding Variation among Individuals and Populations. Am. J. Hum. Genet.

760 95:421–436.

761 Gazave E, Chang D, Clark AG, Keinan A. 2013. Population Growth Inflates the Per-Individual

762 Number of Deleterious Mutations and Reduces Their Mean Effect. Genetics 195:969–

763 978.

764 Gittelman RM, Schraiber JG, Vernot B, Mikacenic C, Wurfel MM, Akey JM. 2016. Archaic

765 Hominin Admixture Facilitated Adaptation to Out-of-Africa Environments. Curr. Biol.

766 26:3375–3382.

767 Gravel S. 2016. When Is Selection Effective? Genetics 203:451–462.

768 Haller BC, Messer PW. 2017. SLiM 2: Flexible, Interactive Forward Genetic Simulations. Mol.

769 Biol. Evol. 34:230–240.

770 Harris K, Nielsen R. 2016. The Genetic Cost of Neanderthal Introgression. Genetics 203:881–

771 891.

772 Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell

773 D, Zadissa A, Searle S, et al. 2012. GENCODE: The reference human genome annotation

774 for The ENCODE Project. Genome Res. 22:1760–1774.

775 Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S. 2015. Estimating the mutation load

776 in human genomes. Nat. Rev. Genet. 16:333–343.

777 Henn BM, Botigué LR, Peischl S, Dupanloup I, Lipatov M, Maples BK, Martin AR, Musharoff

778 S, Cann H, Snyder MP, et al. 2016. Distance from sub-Saharan Africa predicts mutational

779 load in diverse human genomes. Proc. Natl. Acad. Sci. 113:E440–E449.

780 Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet. Res.

781 8:269–294.

782 Huber CD, Kim BY, Marsden CD, Lohmueller KE. 2017. Determining the factors driving

783 selective effects of new nonsynonymous mutations. Proc. Natl. Acad. Sci. 114:4465–

784 4470.

785 Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M,

786 Somel M, et al. 2014. Altitude adaptation in Tibetans caused by introgression of

787 Denisovan-like DNA. Nature 512:194–197.

788 Ingvarsson PK, Whitlock MC. 2000. Heterosis increases the effective migration rate. Proc. R.

789 Soc. Lond. B Biol. Sci. 267:1321–1326.

790 Juric I, Aeschbacher S, Coop G. 2016. The Strength of Selection against Neanderthal

791 Introgression. PLOS Genet. 12:e1006340.

792   Keightley PD, Otto SP. 2006. Interference among deleterious mutations favours sex and
793   recombination in finite populations. Nature 443:89–92.

794   Keller A. 2007. Drosophila melanogaster's history as a human commensal. Curr. Biol. 17:R77–
795   R81.

796   Kim BY, Huber CD, Lohmueller KE. 2017. Inference of the Distribution of Selection
797   Coefficients for New Nonsynonymous Mutations Using Large Samples. Genetics
798   206:345–361.

799   Kimura M, Maruyama T, Crow JF. 1963. The Mutation Load in Small Populations. Genetics
800   48:1303–1312.

801   Kirkpatrick M, Jarne P. 2000. The Effects of a Bottleneck on Inbreeding Depression and the
802   Genetic Load. Am. Nat. 155:154–167.

803   Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB,
804   Jonasdottir A, Gylfason A, Kristinsson KT, et al. 2010. Fine-scale recombination rate
805   differences between sexes, populations and individuals. Nature 467:1099–1103.

806   Liu Q, Zhou Y, Morrell PL, Gaut BS. 2017. Deleterious Variants in Asian Rice and the Potential
807   Cost of Domestication. Mol. Biol. Evol. 34:908–924.

808   Lohmueller KE. 2014a. The Impact of Population Demography and Selection on the Genetic
809   Architecture of Complex Traits. PLOS Genet 10:e1004379.

810   Lohmueller KE. 2014b. The distribution of deleterious genetic variation in human populations.
811   Curr. Opin. Genet. Dev. 29:139–146.

812   Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White
813   TJ, Sunyaev SR, Nielsen R, et al. 2008. Proportionally more deleterious genetic variation
814   in European than in African populations. Nature 451:994–997.

815   Marsden CD, Vecchyo DO-D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, Marques-Bonet T,
816   Schnabel RD, Wayne RK, Lohmueller KE. 2016. Bottlenecks and selective sweeps
817   during domestication have increased deleterious genetic variation in dogs. Proc. Natl.
818   Acad. Sci. 113:152–157.

819   Moyers BT, Morrell PL, McKay JK. 2017. Genetic costs of domestication and improvement. J.
820   Hered. esx069, https://doi.org/10.1093/jhered/esx069

821   Muller HJ. 1964. The relation of recombination to mutational advance. Mutat. Res. Mol. Mech.
822   Mutagen. 1:2–9.

823   Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. Mol.
824   Ecol. 25:2337–2360.

825   Pedersen C-ET, Lohmueller KE, Grarup N, Bjerregaard P, Hansen T, Siegismund HR, Moltke I,

826   Albrechtsen A. 2017. The Effect of an Extreme and Prolonged Population Bottleneck on

827   Patterns of Deleterious Variation: Insights from the Greenlandic Inuit. Genetics 205:787–

828   801.

829   Peischl S, Excoffier L. 2015. Expansion load: recessive mutations and the role of standing genetic

830   variation. Mol. Ecol. 24:2084–2094.

831   Pool JE. 2015. The Mosaic Ancestry of the Drosophila Genetic Reference Panel and the D.

832   melanogaster Reference Genome Reveals a Network of Epistatic Fitness Interactions.

833   Mol. Biol. Evol. 32:3236–3251.

834   Racimo F, Marnetto D, Huerta-Sánchez E. 2017. Signatures of Archaic Adaptive Introgression in

835   Present-Day Human Populations. Mol. Biol. Evol. 34:296–317.

836   Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. 2015. Evidence for archaic adaptive

837   introgression in humans. Nat. Rev. Genet. 16:359–371.

838   Richards EJ, Martin CH. 2017. Adaptive introgression from distant Caribbean islands contributed

839   to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes.

840   PLOS Genet. 13:e1006919.

841   Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D.

842   2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature

843   507:354–357.

844   Simmons MJ, Crow JF. 1977. Mutations affecting fitness in Drosophila populations. Annu. Rev.

845   Genet. 11:49–78.

846   Simons YB, Sella G. 2016. The impact of recent population history on the deleterious mutation

847   load in humans and close evolutionary relatives. Curr. Opin. Genet. Dev. 41:150–158.

848   Simons YB, Turchin MC, Pritchard JK, Sella G. 2014. The deleterious mutation load is

849   insensitive to recent population history. Nat. Genet. 46:220–224.

850   Stephan W, Li H. 2006. The recent demographic and adaptive history of *Drosophila*

851   *melanogaster*. Heredity 98:65–68.

852   Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M, Grote S,

853   McCoy RC, Norton H, et al. 2016. Excavating Neandertal and Denisovan DNA from the

854   genomes of Melanesian individuals. Science:aad9416.

855   Wall JD, Yoshihara Caldeira Brandt D. 2016. Archaic admixture in human history. Curr. Opin.

856   Genet. Dev. 41:93–97.

857   Whitlock MC, Ingvarsson PK, Hatfield T. 2000. Local drift load and the heterosis of

858   interconnected populations. Heredity 84 (Pt 4):452–457.