Expansion load: recessive mutations and

the role of standing genetic variation

Stephan Peischl[1,2], Laurent Excoffier[1,2]

[1] Institute of Ecology and Evolution, University of Berne, 3012 Berne, Switzerland
[2] Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

Email addresses of authors:
stephan.peischl@iee.unibe.ch, laurent.excoffier@iee.unibe.ch

Corresponding author:
Stephan Peischl
Institute of Ecology and Evolution
University of Bern
Baltzerstrasse 6
CH-3012 Bern
Switzerland

Phone: +41 31 631 30 36 Fax: +41 31 631 48 88

Email: stephan.peischl@iee.unibe.ch

Abstract

Expanding populations incur a mutation burden – the so-called expansion load. Previous studies of expansion load have focused on co-dominant mutations. An important consequence of this assumption is that expansion load stems exclusively from the accumulation of new mutations occurring in individuals living at the wave front. Using individual-based simulations we study here the dynamics of standing genetic variation at the front of expansions, and its consequences on mean fitness if mutations are recessive. We find that deleterious genetic diversity is quickly lost at the front of the expansion, but the loss of deleterious mutations at some loci is compensated by an increase of their frequencies at other loci. The frequency of deleterious homozygotes therefore increases along the expansion axis whereas the average number of deleterious mutations per individual remains nearly constant across the species range. This reveals two important differences to co-dominant models: (i) mean fitness at the front of the expansion drops much faster if mutations are recessive, and (ii) mutation load can increase during the expansion even if the total number of deleterious mutations per individual remains constant. We use our model to make predictions about the shape of the site frequency spectrum at the front of range expansion, and about correlations between heterozygosity and fitness in different parts of the species range. Importantly, these predictions provide opportunities to empirically validate our theoretical results. We discuss our findings in the light of recent results on the distribution of deleterious genetic variation across human populations, and link them to empirical results on the correlation of heterozygosity and fitness found in many natural range expansions.

Introduction

Identifying and understanding the ecological and evolutionary processes that cause range expansions, range shifts, or contractions has a long tradition in evolutionary biology (Darwin 1859; MacArthur 1972; Sexton *et al.* 2009). More recently, the growing appreciation of the consequences of dynamic range margins on the ecology, population genetics, and behavior of species has changed our views about several evolutionary processes, such as the evolution of dispersal (Phillips *et al.* 2006; Shine *et al.* 2011; Lindström *et al.* 2013), life-history traits (Phillips *et al.* 2010), and species range limits (Peischl *et al.* 2014).

The evolutionary processes at the margins of expanding populations allow neutral genetic variants to quickly spread into new territories (Klopfstein *et al.* 2006), a phenomenon called "gene surfing". Gene-surfing of neutral variation has been investigated both theoretically (Hallatschek and Nelson 2008; Excoffier *et al.* 2009; Slatkin and Excoffier 2012) and empirically (Hallatschek and Nelson 2008; Moreau *et al.* 2011; Graciá *et al.* 2013). Gene surfing can also affect the spread of selected variants (Travis *et al.* 2007; Burton and Travis 2008; Lehe *et al.* 2012; Peischl *et al.* 2013; Peischl *et al.* 2014). Population-genetics models of range expansions predict that expanding populations incur a mutation burden – the "expansion load" (Peischl *et al.* 2013). Expansion load is a transient phenomenon, but it can persist for several hundreds to thousands of generations, and may limit the ability of a species to colonize new habitats (Peischl *et al.* 2014).

Previous studies of expansion load assumed that mutations were co-dominant. An important consequence of this assumption is that standing genetic variation has no effect on the dynamics of mean fitness at the front of expanding populations (Peischl *et al.* 2013). In particular, the total number of mutations per individual, and hence the individual's fitness, remains approximately constant if new mutations are ignored (Peischl *et al.* 2013; Peischl *et al.* 2014). In additive models, expansion load thus stems exclusively from the accumulation of new mutations that occur in individuals living at the front of the expansion.

3

70   Empirical evidence for expansion load may come from humans, where a proportional excess

71   of deleterious mutations in non-African populations has been found (Lohmueller *et al.* 2008;

72   Subramanian 2012; Torkamani *et al.* 2012; Peischl *et al.* 2013; Fu *et al.* 2014; Lohmueller 2014).

73   Importantly, when focusing on mutations that occurred during or after the out-of Africa expansion,

74   the excess of deleterious variants is not restricted to rare variants (Peischl *et al.* 2013). This suggests

75   that proportionally more deleterious mutations have risen to high frequencies in human populations

76   located in newly settled habitats. In contrast to what would be expected from expansion-load theory,

77   a recent analysis found no significant differences in the average allele frequency of predicted

78   deleterious alleles (Simons *et al.* 2014). The average number of predicted deleterious mutations

79   carried by an individual is, however, significantly larger in non-Africans (Fu *et al.* 2014). In addition,

80   non-African individuals have significantly more loci homozygous for predicted deleterious alleles than

81   African individuals (Lohmueller *et al.* 2008; Subramanian 2012; Fu *et al.* 2014; Lohmueller 2014). The

82   debate whether human past demography affected the efficacy of selection and the spatial

83   distribution of mutation load is thus still ongoing (Lohmueller 2014).

84   There is mounting evidence that deleterious mutations tend to be recessive (Agrawal and

85   Whitlock 2011). Importantly, if mutations are completely recessive the number of deleterious

86   mutations per individual is not informative about the mutation load (Kimura *et al.* 1963). Instead,

87   mutation load is determined by sites that are homozygous for deleterious alleles. Thus, if mutations

88   are (partially) recessive, the genotypic composition of deleterious genetic variation is more important

89   than the total number of deleterious mutations carried by an individual. Range expansions are

90   known to affect the genotypic composition of neutral standing genetic variation (Excoffier *et al.*

91   2009). The role of standing genetic variation in models of expansion load remains, however, unclear

92   if mutations are recessive.

93   We investigate here the effect of recessive mutations on the dynamics of expansion load. In

94   particular, we use individual based-simulations to investigate the role of standing genetic variation,

4

95   the width of the habitat, and the composition of expansion load with respect to allele frequencies

96   and mutational effects.

Model and Results

Model

99   We model a population of diploid monoecious individuals that occupy discrete demes

100   located on a one- or two-dimensional grid (Kimura and Weiss 1964). Generations are discrete and

101   non-overlapping, and mating within each deme is random. Mating pairs are formed by randomly

102   drawing individuals (with replacement) according to their relative fitness, and each mating pair

103   produces a single offspring. The process is repeated N times, where N is the total number of

104   offsprings of the parental generation, leading to approximately Poisson-distributed numbers of

105   offspring per individual. Individuals then migrate to adjacent demes with probability m per

106   generation. Migration is homogeneous and isotropic, except that the boundaries of the habitat are

107   reflecting, i.e., individuals cannot migrate out of the habitat.

108   Population size grows logistically within demes. The expected number of offspring in the next

109   generation produced by the N_j adults in deme j is

110
$$N_j^* = \frac{R_0}{1 + \left(R_0 - 1 \right) N_j / K} N_j \,,$$

111   where R_0 is the fundamental (geometric) growth rate and K is the deme's carrying capacity

112   (Beverton and Holt 1957). To model demographic stochasticity, the actual number of offspring, N_j',

113   is then drawn from a Poisson distribution with mean N_j^*.

114   The relative fitness of individuals is determined by n independently segregating biallelic loci.

115   The alleles at locus i are denoted a_i (wildtype) and A_i (derived). Mutations occur in both directions

116   and the genome wide mutation rate is u; in each new gamete k randomly chosen sites change their

5

117   allelic state, where k is drawn from a Poisson-distribution with mean u. The fitness contributions of

118   the genotypes $a_i a_i$, $a_i A_i$ and $A_i A_i$ at locus i are 1, $1 - hs_i$, and $1 - s_i$, respectively. Here s_i

119   denotes the strength of selection at locus i and h is the dominance coefficient. Fitness effects are

120   multiplicative across loci, such that the fitness of an individual is given by $w = \prod_i w_i$, where w_i is the

121   fitness effect of the ith locus of the focus individual, i.e., there is no epistasis. In the following we will

122   focus on co-dominant ($h = 0.5$) or recessive ($h = 0$) mutations. We assume that mutation effects

123   are drawn from the same distribution of fitness effects for all individuals (independently from their

124   current fitness).

125   We perform individual-based simulations of the above described model in 1D or 2D habitats.

126   Our simulations start from ancestral populations located in 10 leftmost (rows of) demes of the range.

127   After a burn-in phase that ensures that the ancestral populations are at mutation-selection-drift

128   balance, the population expands from left to right until the habitat is filled. Because we are mainly

129   interested in the role of standing genetic variation, we focus on relatively short expansions, i.e.,

130   colonization of a 1x50 (1D) or a 20x50 (2D) deme habitat. The long-term dynamics of expansion load

131   have been studied elsewhere (Peischl *et al.* 2013; Peischl *et al.* 2014).

Impact of standing genetic variation on expansion load

133   For simplicity, we first consider expansions along a one-dimensional habitat and assume that

134   all mutations have the same effect, i.e., we set $s_i = s$. If mutations are co-dominant ($h = 0.5$),

135   expansion load is caused exclusively by the establishment of new mutations occurring during the

136   expansion, and standing genetic variation has a negligible effect on the dynamics of mean fitness

137   (Peischl *et al.* 2013). Mean fitness at the wave front decreases at a constant rate over time (Figure 1),

138   and the rate at which mean fitness decreases per generation is proportional to the number of new

139   mutations entering the population per generation (Peischl *et al.* 2013).

140   The dynamics of expansion load changes dramatically if mutations are recessive (Figure 1).

141   The analytical approximation obtained in Peischl *et al.* (2013), which ignores standing genetic

6

142   variation, is a poor fit to the observed dynamics of mean fitness (Figure 1). In the first few

143   generations mean fitness decreases much faster than predicted by analytical theory for the

144   accumulation of new mutations (cf. solid and dashed black lines in Figure 1). The rate at which

145   expansion load is created then slows down and gradually approaches the analytical prediction.Then,

146   changes in expected mean fitness arise exclusively from new mutations (cf. solid and dashed black

147   lines for $t > 50$ in Figure 1). This shows that standing genetic variation plays an important role in the

148   establishment of expansion load if mutations are recessive, especially during early phases of

149   expansions.

150
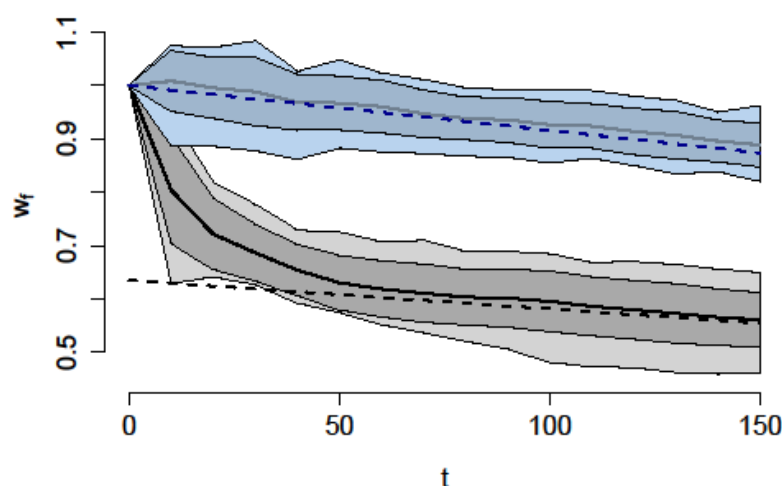
151   **Figure 1**: Evolution of mean fitness at the wave front. Dashed lines show analytical prediction for the

152   evolution of the mean fitness due to de-novo mutations (see Peischl et al. 2013). Simulations show

153   results for the combination of standing and new genetic variation. Gray shaded areas and black lines

154   show results for recessive mutations ($h = 0$), and blue shaded areas and lines show results for additive

155   co-dominant mutations ($h = 0.5$). Solid lines indicate the average mean fitness from 50 simulations,

156   and dark and light shaded areas indicate $\pm$ one standard deviation and the minimum and maximum

157   of mean fitness, respectively. Other parameter values are $n = 1000$, $K = 100$, $u = 0.1$, $m = 0.1$,

158    $s = 0.01$, $R = 2$.

159

7

160   We next investigate the evolution of the genotypic composition of standing genetic variation

161   on the expansion front. In general, we find that the average number of heterozygous loci per

162   individual decreases during the expansion, whereas the number of loci that are homozygous for the

163   derived allele increases (Figure 2). Because we simulated a fixed number of loci, the derived allele

164   frequency shown in Figure 2 is proportional to the average number of mutations carried by an

165   individual. Thus, Figure 2 shows that the total number of mutations per individual remains nearly

166   constant during the expansion (Figure 2). Strong genetic drift is therefore the major force driving the

167   evolution of genotype frequencies at the wave front. At any given locus, mutations are either lost or

168   fixed over the course of the expansion, and the probability of fixation of a given mutation is close to

169   its initial frequency (Peischl *et al.* 2013), suggesting that deleterious mutations are behaving like

170   neutral mutations on the wave front. In 2D expansions, the dynamics of genotype frequencies are

171   qualitatively very similar to 1D expansions (Figure S1).
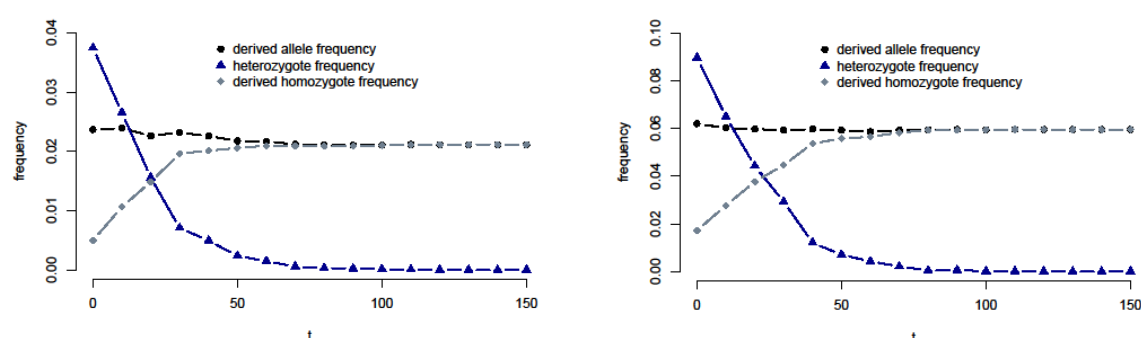
172   **Figure 2**: Evolution of standing genetic variation on the wave front. Panel A shows results for co-

173   dominant mutations and panel B for recessive mutations. Parameter values are as in Figure 1.

174

175   The nearly neutral evolution of allele frequencies on the expansion front reveals a critical role

176   of the degree of dominance on the build-up of the expansion load. If mutations are co-dominant, the

177   fitness of an individual is determined by the total number of mutations it carries (Wright 1930). Thus,

178   Figure 2 shows that standing genetic variation has a negligible impact on fitness (Figure 2A). In

179   contrast, if mutations are recessive, the fitness of an individual is determined by its number of loci

8

180   homozygous for the derived allele. Because the number of derived homozygous loci per individual

181   rapidly increases at the front of the expansions, standing genetic variation has a severe effect on

182   fitness if mutations are recessive (Figures 1 and 2B).

183   Gene flow on the wave front of 2D expansions restores diversity and fitness

184   In the following section, we focus on completely recessive mutations ($h = 0$). Figure 3 shows

185   an example of the evolution of the mean fitness during an expansion in a 2D habitat (20x50 demes).

186   As in 1D expansions, the mean fitness drops to low levels on the expansion front within the first few (

187   ≈ 30) generations and then continues to gradually decreases at a slower rate. There is however a

188   considerable variation in fitness across the wave front of 2D expansions (fitness-differences of more

189   than 40%, Figures 3 and 4). At the end of the expansion (Figure 3, $T = 150$), we find a high-fitness

190   ridge along the expansion axis in the central part of the newly settled species range, surrounded by

191   sectors of low fitness on the lateral edges of the species range. This is partially caused by the lack of

192   immigrants at the lateral edge of the species range (boundary effect). However, the location of the

193   high-fitness ridge varies across simulation runs, suggesting that a boundary effect alone cannot

194   explain the observed patterns (Figure S2).

195
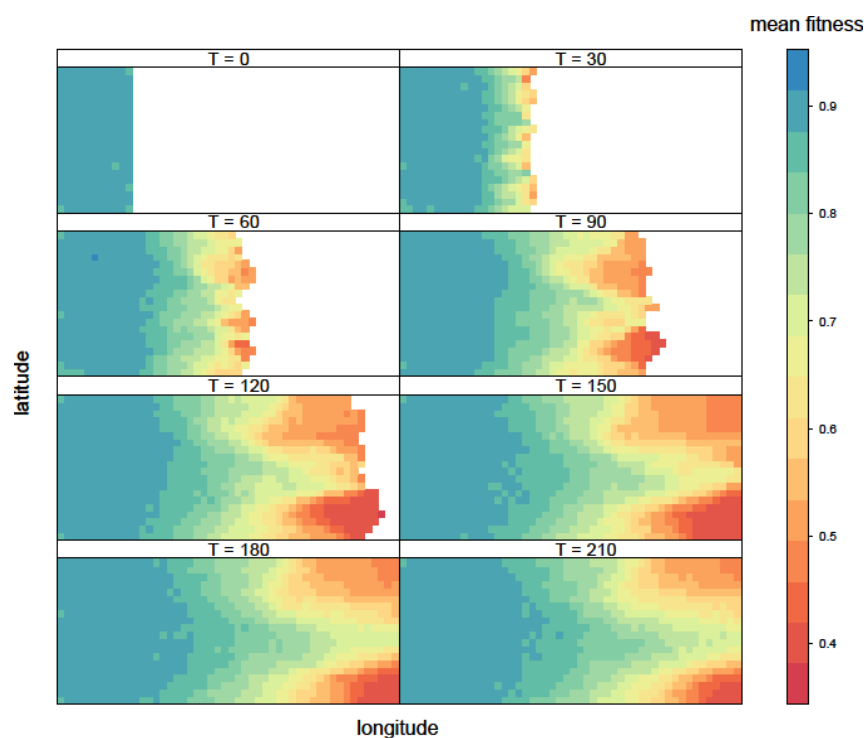
196   **Figure 3**: Evolution of mean fitness during range expansion. The simulated grid is 20x50 demes.

197   Mutations are recessive and parameter values are as in Figure 1.

198

199
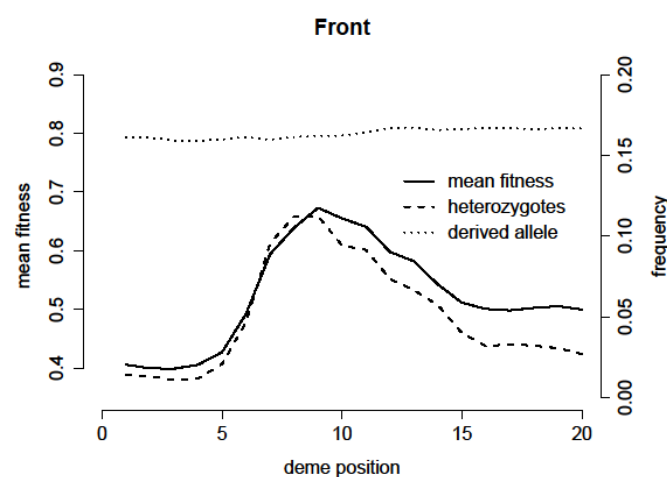
200   **Figure 4**. Genetic properties of the demes located on the wave. The figure shows the mean fitness,

201   the heterozygosity, and the derived allele frequency at the front of the expansion when the habitat

202   has just been fully colonized. The deme mean fitness on the wave front correlates with

203   heterozygosity, but not with derived allele frequency. Statistics were computed from the simulation

204   shown in Figure 3 at generation 150.

205   Figure 4 shows the variation in fitness, heterozygosity, and derived allele frequency across

206   the wave front at the end of the expansion shown in Figure 3. We find that the average number of

207   mutations per individual is uniform across the expansion front, which means that the variation in

208   fitness across the expansion front is not driven by a differential accumulation of mutations.

209   Heterozygosity, on the other hand, correlates strongly with mean fitness (cf. solid and dashed line in

210   Figure 4). This observation suggests that different mutations establish in different parts of the wave

211   front, and that gene flow between demes restores heterozygosity, which masks the effect of

212   deleterious recessive mutations. These results show that heterozygosity-fitness correlations (HFC)

213   are readily created during range expansions. We indeed find a strongly positive HFC at the front of

214   the expansion (Figure 5A, $R^2 = 0.526$., *slope = 0.32,* $p < 10^{-16}$), but not in the ancestral population

215   (Figure 5B, $R^2 = 0.001$, *slope = 0.03*, $p > 0.07$). Interestingly, weaker but similar correlations are

216   found at the individual level within demes (mean slope ≈ 0.1, $p < 0.05$ in 81% of all simulated

217   demes).

218
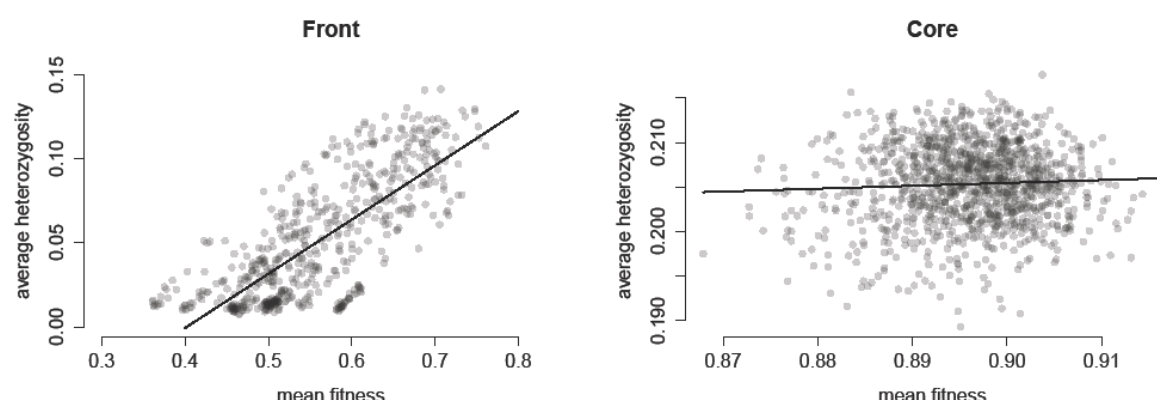
Figure 5. Heterozygosity-Fitness correlations (HFC). (A) HFC on the expansion front at generation

$T = 120$ ($R^2 = 0.526$, slope = 0.32, $p < 10^{-16}$). (B) No significant HFC in core populations before the

222   onset of the expansion ($R^2 = 0.001$, slope = 0.03 , $p > 0.07$). Each point represents a single deme.

223   The results from 10 simulation replicates are shown. Parameter values are as in Figure 3.

224

225   ## Expansion load is driven by a few mutations occurring at high frequency

226   So far we assumed that all mutations had the same effect s . To investigate the composition

227   of expansion load with respect to mutation fitness effects, we now consider the case where mutation

228   fitness effects are drawn from an exponential distribution with mean s . Figure 6A and B show the

229   site frequency spectrum (SFS) observed in core and front populations, respectively. In core

230   populations, the SFS shows the pattern expected for sites under negative selection (Bustamante *et*

231   *al.* 2001), with a large excess of low frequency variants. On the wave front, the total number of

232   segregating sites is reduced in marginal populations (cf. Figures 6 A and B). More interestingly, we

233   see a markedly different SFS, with, as compared to neutral expectation, a clear deficit of rare and

234   intermediate frequency variants and an increase in high frequency variants (Figure 6B). Thus, even

235   though fewer polymorphic sites with deleterious variants are found in more recently colonized areas

236   than in the ancestral region, the alleles at polymorphic sites tend to be at higher frequency in more

237   recently colonized populations.

238   Figure 6 C and D show the distribution of polymorphic loci stratified according to their

239   mutation effect sizes. The eight mutation effect classes have been defined such that they represent

240   the 8-quantiles of the DFE, i.e., the rate at which mutations of a given category enter the population

241   are equal for all categories. As expected, we find that the number of polymorphic loci generally

242   decreases with increasing mutation effect size, and that large effect mutations tend to be present at

243   lower frequencies than low effect mutations (Figure 6 C and D). Compared to core populations, the

244   allele frequencies at polymorphic sites on the wave front tend to be larger across all mutational

245   effect categories. Furthermore, the increase in allele frequency is most pronounced for small effect

246   mutations. Thus, expansion load is driven mainly by mildly and moderately standing deleterious

247   mutations (i.e., up to $Ns < 2$ for the parameter values used in Figure 6) that rise to high frequency

248   during the expansion.

249
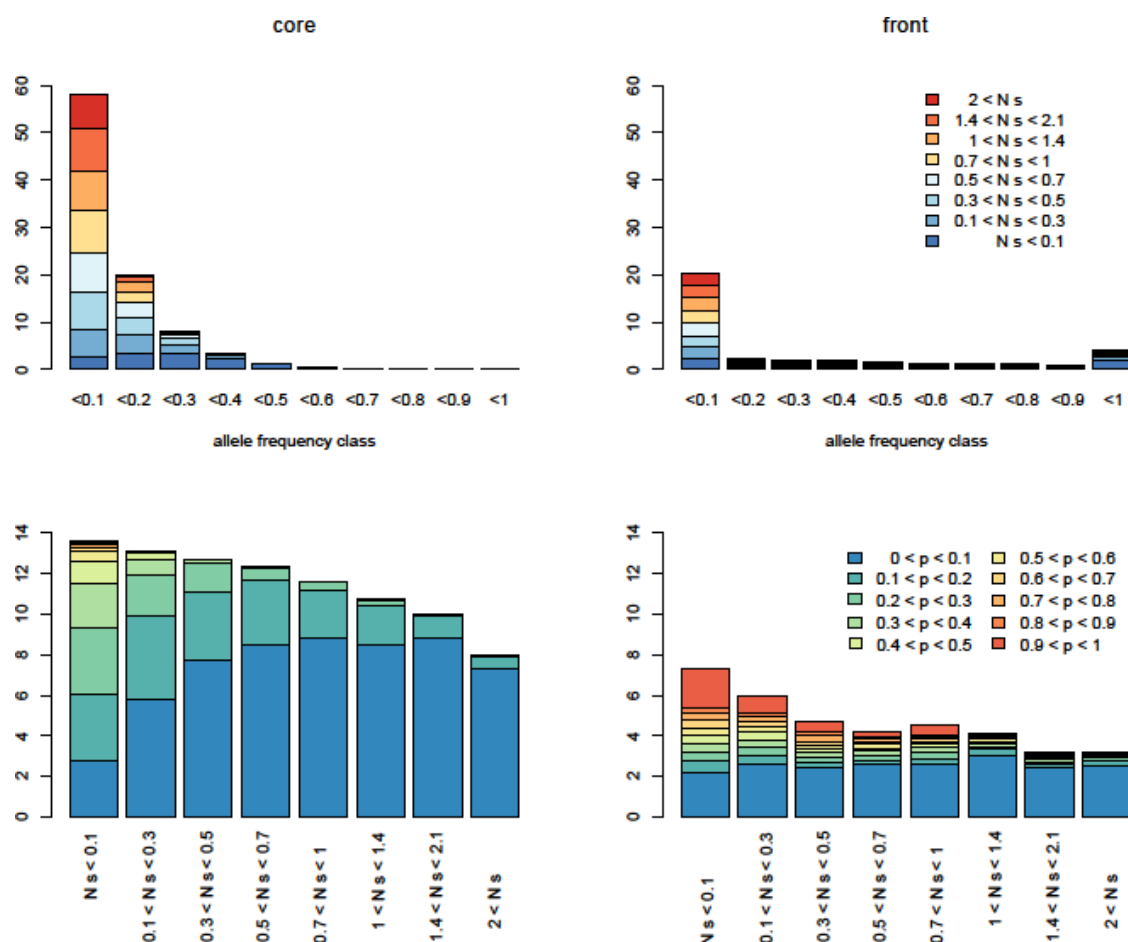
250   **Figure 6:** Distribution of average number of polymorphic loci per individual. The distribution is

251   stratified for allele frequencies (top row) and mutation effects (bottom row). Results were recorded

252   150 generations after the onset of the expansion, which is shortly after the habitat was colonized

253   completely (mean time to colonization ≈ 130 generations, see also Figure 3). Panels (A) and (C)

254   show results for a core population (coordinates 10, 5), (B) and (D) front population (coordinates

255   10,45). Mutations are recessive and their effects are drawn from an exponential distribution with

256   mean s = 0.01. Other parameter values are as in Figure 1.

Discussion

258   We have investigated here the dynamics of an expansion load caused by recessive mutations.

259   Using individual-based simulations we have shown that shifts in the genotypic composition of

260   standing genetic variation can lead to a rapid drop of mean fitness at the onset of an expansion (see

261   Figures 1 and 2) without necessarily affecting the total number of deleterious alleles per individuals

262   (see Figure 2). The total expansion load resulting from standing genetic variation is limited by the

263   initial frequency of deleterious mutations (see Figure 2). Thus, if many loci are polymorphic for

264   deleterious variants at the onset of the expansion, the (recessive) expansion load from standing

265   genetic variation can be the dominating the total mutation load (see Figure 1). Even though these

266   results have been inferred by assuming that all deleterious mutations were recessive, we would

267   predict that a similar phenomenon, though of lesser amplitude, would occur if only some of the

268   mutations would be fully or partially recessive.

269   The effect of range expansions on deleterious genetic diversity is also reflected in the site

270   frequency spectrum (SFS, see Figure 6). As compared to stationary populations in the core of the

271   species range, populations from more recently colonized areas have fewer segregating sites, but

272   proportionally more high and low frequency variants (cf. Figure 6A and B). These differences in the

273   SFS of core and front populations should provide an opportunity to evidence expansion load from

274   sequence data, and to infer important quantities such as the distribution of fitness effects (Keightley

275   and Eyre-Walker 2007; Boyko *et al.* 2008 ; Racimo and Schraiber 2014). The development of

276   statistical and computational methods able to infer parameters under spatially explicit models

277   including range expansions and selection remains, however, a major challenge (Sousa *et al.* 2014).

278   Interestingly, human genomic data are consistent with our predictions for genomic

279   signatures of expansion load. In particular, the number of segregating sites is higher in African

280   populations than in non-African populations (Lohmueller *et al.* 2008), non-African populations show

281   an excess of low-frequency and high-frequency deleterious alleles (Lohmueller *et al.* 2008; Fu *et al.*

282   2014), the average number of sites that are homozygous for predicted deleterious variants sites is

283   larger in non-African individuals (Fu *et al.* 2014), and the average number of predicted deleterious

284   mutations per individual is slightly, but significantly, larger in non-Africans (Fu *et al.* 2014) .

285   Determining mutation load (or, alternatively, fitness) from genomic variation data is, however, an

14

286   intrinsically difficult problem because mutation load depends on many unknown parameters

287   (selection coefficients that may vary over space and time, epistatic interactions, dominance

288   relationships, etc.), and the relevance of comparing a population with deleterious mutations to a

289   theoretical population free of such mutations is questionable (Lesecque *et al.* 2012). Testing

290   theoretical predictions of the effect of a range expansion on functional diversity with human genomic

291   data might nevertheless be extremely useful to substantially increase our understanding of the

292   complex interactions of demography and selection.

293   We assumed here that selection was soft, i.e., demographic parameters are independent of

294   fitness (Wallace 1975), but it would be interesting to extend our results to models of hard selection,

295   where mutation load on the front can stop an expansion and even drive parts of the species range to

296   extinction (Peischl *et al.* 2014). Our results suggest that admixture during range expansions, or

297   secondary contact between expanding lineages, could mitigate expansion load and prevent marginal

298   populations from collapsing. A previous study of range expansions under an additive model with hard

299   selection has shown that suppressing recombination at the wave front can have beneficial effects for

300   the spread of high fitness lineages (Peischl *et al.* 2014). Recombination modifiers, such as inversions,

301   could have a similar effect if mutations are recessive and facilitate the spread of admixed lineages.

302   An interesting example for studying the potentially beneficial role of admixture and suppressed

303   recombination during range expansions is from the clam genus *Corbicula,* which includes both sexual

304   and asexual (androgenetic diploid) lineages. Sexual populations are restricted to their native Asian

305   areas, but the androgenetic lineages are widely distributed and extend as far as in America and

306   Europe where they are invasive (Pigneur *et al.* 2014). Intriguingly, the invasive lineages also show an

307   excess of heterozygosity, which is preserved through clonal reproduction. No such excess of

308   heterozygosity is found in the native range, suggesting that the combination of asexual reproduction

309   and high heterozygosity may have been key drivers of the invasion.

310   An interesting prediction of our model is that if a given proportion of deleterious mutations

311   are recessive, then heterozygosity-fitness correlations (HFC) should naturally occur in populations

15

312   that have recently expanded their range (see Figures 4 and 5A). Importantly, the positive correlation

313   between heterozygosity and fitness in recently colonized areas can be observed at both the

314   individual and the population level (Fig. 5). Even though our simulations modeled a single expansion

315   in a 2D habitat, we would expect similar HFCs if there was a secondary contact between expanding

316   populations from different areas (e.g., from different LGM refuge areas). The HFC should be even

317   stronger in the case of a secondary contact, because the isolation between expanding lineages

318   should be larger and different recessive alleles could have fixed in different refugia or during the

319   expansion from these refugia. HFC have been observed in many cases of natural range expansions

320   and invasive species (Chapman *et al.* 2009) but their underlying mechanisms and their role during

321   range expansions and invasions are still unclear (Szulkin *et al.* 2010; Rius and Darling 2014)- A

322   particularly interesting example of HFC is found in the invasive weed *Silene vulgaris,* where, as

323   predicted by our model (see Figure 5), HFC correlations are observed in the recently invaded North

324   American range, but not in their native European range. It remains however unclear whether

325   admixture between divergent lineages has indeed a causal role in range expansions. A combination

326   of transplantation experiments and genomic data analyses could certainly be used to test the

327   predictions of our model.

328   In summary, we have investigated here the evolution of standing genetic variation during

329   range expansions, the dynamics of mean fitness on the expansion front if mutations are recessive,

330   and the genomic signature of range expansions. Importantly, our results make predictions that can

331   be tested in natural populations. Empirical validation of our results would increase our understanding

332   of the interactions of demography and selection (Lohmueller 2014), and could help us identifying key

333   drivers of range expansions and biological invasions (Rius and Darling 2014).

16

Supplementary Figures
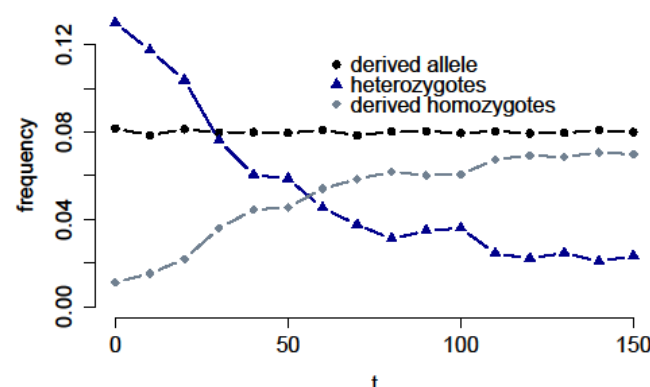
Figure S1. Evolution of genotype frequencies at the front of an expansion in a 2D habitat of 20x50 demes. Parameter values are as in Figure 1B.
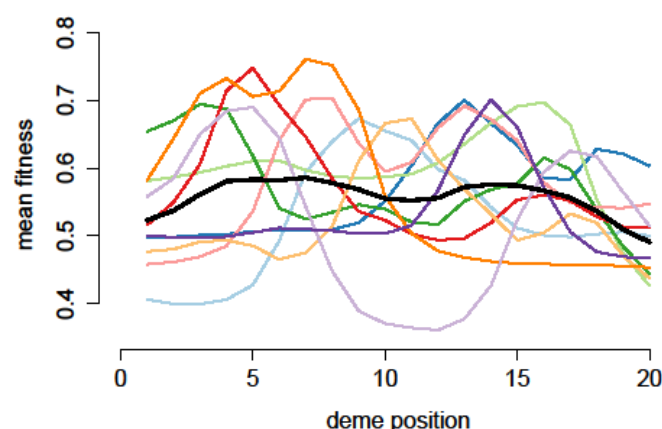
Figure S2. Mean fitness at the front of the expansion. The figure shows the mean fitness at the front of the expansion (at generation 150). Colored lines are the results from 10 simulation runs, solid black line shows the average over all simulation runs. Parameter values are as in Figure 3.

Acknowledgements

References

347 Agrawal AF, Whitlock MC (2011) Inferences about the distribution of dominance drawn from yeast

348   gene knockout data. *Genetics* **187**, 553-566.

349 Beverton R, Holt S (1957) On the dynamics of exploited fish populations. Fisheries Investigation

350   Series 2 (19). London: Ministry of Agriculture. *Fisheries and Food*.

351 Boyko AR, Williamson SH, Indap AR*, et al.* (2008) Assessing the evolutionary impact of amino acid

352   mutations in the human genome. *PLoS Genet* **4**, e1000083.

353 Burton OJ, Travis JMJ (2008) The frequency of fitness peak shifts is increased at expanding range

354   margins due to mutation surfing. *Genetics* **179**, 941-950.

355 Bustamante CD, Wakeley J, Sawyer S, Hartl DL (2001) Directional selection and the site-frequency

356   spectrum. *Genetics* **159**, 1779-1788.

357 Chapman J, Nakagawa S, Coltman D, Slate J, Sheldon B (2009) A quantitative review of

358   heterozygosity–fitness correlations in animal populations. *Molecular ecology* **18**, 2746-2765.

359 Darwin C (1859) On the origins of species by means of natural selection. *London: Murray*.

360 Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annual Review of*

361   *Ecology, Evolution, and Systematics* **40**, 481-501.

362 Fu W, Gittelman RM, Bamshad MJ, Akey JM (2014) Characteristics of Neutral and Deleterious

363   Protein-Coding Variation among Individuals and Populations. *The American Journal of Human*

364   *Genetics* **95**, 421-436.

365 Graciá E, Botella F, Anadón JD*, et al.* (2013) Surfing in tortoises? Empirical signs of genetic structuring

366   owing to range expansion. *Biology letters* **9**, 20121091.

367 Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. *Theor Popul Biol* **73**, 158-

368   170.

369 Keightley PD, Eyre-Walker A (2007) Joint inference of the distribution of fitness effects of deleterious

370   mutations and population demography based on nucleotide polymorphism frequencies.

371   *Genetics* **177**, 2251-2261.

372   Kimura M, Maruyama T, Crow JF (1963) The mutation load in small populations. *Genetics* **48**, 1303.

373   Kimura M, Weiss GH (1964) The Stepping Stone Model of Population Structure and the Decrease of

374   Genetic Correlation with Distance. *Genetics* **49**, 561-576.

375   Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range

376   expansion. *Mol Biol Evol* **23**, 482-490.

377   Lehe R, Hallatschek O, Peliti L (2012) The rate of beneficial mutations surfing on the wave of a range

378   expansion. *PLoS Comput Biol* **8**, e1002447.

379   Lesecque Y, Keightley PD, Eyre-Walker A (2012) A resolution of the mutation load paradox in

380   humans. *Genetics* **191**, 1321-1330.

381   Lindström T, Brown GP, Sisson SA, Phillips BL, Shine R (2013) Rapid shifts in dispersal behavior on an

382   expanding range edge. *Proceedings of the National Academy of Sciences* **110**, 13452-13456.

383   Lohmueller KE (2014) The distribution of deleterious genetic variation in human populations. *bioRxiv*.

384   Lohmueller KE, Indap AR, Schmidt S*, et al.* (2008) Proportionally more deleterious genetic variation in

385   European than in African populations. *Nature* **451**, 994-U995.

386   MacArthur RH (1972) *Geographical ecology: patterns in the distribution of species* Princeton

387   University Press.

388   Moreau C, Bherer C, Vezina H*, et al.* (2011) Deep Human Genealogies Reveal a Selective Advantage

389   to Be on an Expanding Wave Front. *Science* **334**, 1148-1150.

390   Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious

391   mutations during range expansions. *Molecular ecology* **22**, 5972-5982.

392   Peischl S, Kirkpatrick M, Excoffier L (2014) Expansion load and the evolutionary dynamics of a species

393   range. *American Naturalist, in press*.

394   Phillips BL, Brown GP, Shine R (2010) Life-history evolution in range-shifting populations. *Ecology* **91**,

395   1617-1627.

396   Phillips BL, Brown GP, Webb JK, Shine R (2006) Invasion and the evolution of speed in toads. *Nature*

397   **439**, 803-803.

398   Pigneur LM, Etoundi E, Aldridge DC, *et al.* (2014) Genetic uniformity and long‑distance clonal

399   dispersal in the invasive androgenetic Corbicula clams. *Molecular ecology*.

400   Racimo F, Schraiber JG (2014) Approximation to the distribution of fitness effects across functional

401   categories in human segregating polymorphisms. *PLoS genetics* **10**, e1004697.

402   Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of

403   colonising populations? *Trends in ecology & evolution* **29**, 233-242.

404   Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009) Evolution and Ecology of Species Range Limits.

405   *Annual Review of Ecology Evolution and Systematics* **40**, 415-436.

406   Shine R, Brown GP, Phillips BL (2011) An evolutionary process that assembles phenotypes through

407   space rather than through time. *Proceedings of the National Academy of Sciences* **108**, 5708.

408   Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to

409   recent population history. *Nature genetics*.

410   Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: a spatial analog of genetic

411   drift. *Genetics* **191**, 171-181.

412   Sousa V, Peischl S, Excoffier L (2014) Impact of range expansions on current human genomic

413   diversity. *Current opinion in genetics & development* **29**, 22-30.

414   Subramanian S (2012) The abundance of deleterious polymorphisms in humans. *Genetics* **190**, 1579-

415   1583.

416   Szulkin M, Bierne N, David P (2010) HETEROZYGOSITY‑FITNESS CORRELATIONS: A TIME FOR

417   REAPPRAISAL. *Evolution* **64**, 1202-1217.

418   Torkamani A, Pham P, Libiger O*, et al.* (2012) Clinical implications of human population differences in

419   genome-wide rates of functional genotypes. *Frontiers in genetics* **3**.

420   Travis JMJ, Munkemuller T, Burton OJ*, et al.* (2007) Deleterious mutations can surf to high densities

421   on the wave front of an expanding population. *Mol Biol Evol* **24**, 2334-2343.

422   Wallace B (1975) Hard and soft selection revisited. *Evolution*, 465-473.

423   Wright S (1930) THE GENETICAL THEORY OF NATURAL SELECTION A Review. *Journal of Heredity* **21**,

424   349-356.

425