MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 2084–2094 doi: 10.1111/mec.13154

INVASION GENETICS: THE BAKER AND STEBBINS LEGACY

Expansion load: recessive mutations and the role of standing genetic variation

STEPHAN PEISCHL*† and LAURENT EXCOFFIER*†

*Institute of Ecology and Evolution, University of Berne, Berne 3012, Switzerland, †Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland

Abstract

Expanding populations incur a mutation burden – the so-called expansion load. Previous studies of expansion load have focused on codominant 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 codominant 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.

Keywords: hybridization, invasive species, molecular evolution, mutation load, population genetics – theoretical, recessive

Received 18 November 2014; revision received 11 March 2015; accepted 11 March 2015

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 behaviour 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.* 2015).

Strong genetic drift at the margins of expanding populations allows some neutral genetic variants that are on the wave front to strongly increase in frequencies and spread over large territories in newly colonized habitats (Edmonds *et al.* 2004), a phenomenon called 'gene surfing' (Klopfstein *et al.* 2006). Gene surfing of neutral variation has been investigated both theoretically (Hallatschek & Nelson 2008; Excoffier *et al.* 2009; Slatkin & Excoffier 2012) and empirically (Hallatschek

Correspondence: Stephan Peischl, Fax: +41 31 631 48 88; E-mail: stephan.peischl@iee.unibe.ch

& 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 & Travis 2008; Lehe *et al.* 2012; Peischl *et al.* 2013, 2015). 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.* 2015).

Previous studies of expansion load assumed that mutations were codominant. 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, 2015). 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.

Empirical evidence for expansion load may come from humans, where a proportional excess of deleterious mutations in non-African populations has been found (Lohmueller *et al.* 2008; Subramanian 2012; Torkamani *et al.* 2012; Peischl *et al.* 2013; Fu *et al.* 2014; Lohmueller 2014). Importantly, when focusing on mutations that occurred during or after the out-of-Africa expansion, the excess of deleterious variants is not restricted to rare variants (Peischl *et al.* 2013). This suggests that proportionally more deleterious mutations have risen to high frequencies in human populations located in newly settled habitats. In contrast to what would be expected from expansion-load theory, recent analyses found no significant differences in the average allele frequency of predicted deleterious alleles (Fu *et al.* 2014; Simons *et al.* 2014; Do *et al.* 2015). The average number of predicted deleterious mutations carried by an individual is, however, slightly but significantly larger in non-Africans (Fu *et al.* 2014). In addition, non-African individuals have significantly more loci homozygous for predicted deleterious alleles than African individuals (Lohmueller *et al.* 2008; Subramanian 2012; Fu *et al.* 2014). Whether and how human past demography affected the efficacy of selection and the spatial distribution of mutation load is thus still ongoing, and the interested readers are referred to the recent review of Lohmueller (2014) who provides a constructive attempt at reconciling views on this subject.

It is an old observation that deleterious mutations tend to be (partially) recessive (Morton *et al.* 1956; Mukai *et al.* 1972). More recently, it has been shown that most deleterious mutations have small effects and that their effect size correlates negatively with recessiveness (Garcia-Dorado & Caballero 2000; Peters *et al.* 2003; Eyre-Walker & Keightley 2007; Agrawal & Whitlock 2011). Importantly, if mutations are recessive, the number of deleterious mutations per individual alone is not informative about the mutation load (Kimura *et al.* 1963). For instance, if deleterious mutations are completely recessive, mutation load is determined by sites that are homozygous for deleterious alleles. Thus, if mutations are even partially recessive, the genotypic partitioning of deleterious variation is more important than the total number of deleterious mutations carried by an individual.

Past demographic events have been shown to affect the genotypic composition of standing genetic variation and therefore the mutation load, to an extent that is still debated (Kirkpatrick & Jarne 2000; Lohmueller *et al.* 2008; Simons *et al.* 2014; Do *et al.* 2015). Kirkpatrick & Jarne (2000) studied analytically the effect of a single-generation bottleneck on the mutation load and showed that a severe bottleneck would always increase the load relative to a population at mutation–selection equilibrium for partially recessive variants and that effect is stronger for more recessive alleles. In their model, most deleterious variants are lost, but others sharply increase in frequency and contribute proportionally more to the load. Recently, these results were confirmed and extended to more complex bottleneck scenarios that were estimated from human genomic data (Gravel 2014). Interestingly, it seems that changes in the distribution of the number of deleterious variants carried by individuals after a bottleneck can be used to infer whether selection is predominantly recessive or additive (Balick *et al.* 2014).

Similar to bottlenecks, range expansions are also known to affect the genotypic composition of neutral standing genetic variation (Excoffier *et al.* 2009), but seem to have a larger effect than bottlenecks on the mutation load (Peischl *et al.* 2013). The role of standing genetic variation in models of expansion load remains however unclear when mutations are recessive. We investigate here the effect of recessive mutations on the dynamics of expansion load. In particular, we use individual-based simulations to investigate the role of standing genetic variation, the width of the habitat and the composition of expansion load with respect to allele frequencies and mutational effects.

Model and Results

Model

We model a population of diploid monoecious individuals that occupy discrete demes located on a one- or

two-dimensional grid (Kimura & Weiss 1964). Generations are discrete and nonoverlapping, and mating within each deme is random. Adult individuals migrate to adjacent demes with probability m per generation. Migration is homogeneous and isotropic, except that the boundaries of the habitat are reflecting, that is individuals cannot migrate out of the habitat.

Population size grows logistically within demes. The expected number of offspring in the next generation produced by the N_j adults in deme j is

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

where R_0 is the fundamental (geometric) growth rate and K is the deme's carrying capacity (Beverton & Holt 1957). To model demographic stochasticity, the actual number of offspring, N'_j is then drawn from a Poisson distribution with mean N^*_j. Mating pairs are formed by randomly drawing individuals (with replacement) according to their relative fitness, and each mating pair produces a single offspring. The process is repeated N'_j times, leading to approximately Poisson-distributed numbers of offspring per individual.

The relative fitness of individuals is determined by n independently segregating biallelic loci. The alleles at locus i are denoted a_i (wild type) and A_i (derived). Mutations occur in both directions and the genomewide mutation rate is u; in each new gamete k, randomly chosen sites change their allelic state, where k is drawn from a Poisson distribution with mean u. The fitness contributions of the genotypes a_ia_i a_iA_i and A_iA_i at locus i are 1, $1–hs_i$ and $1–s_i$, respectively. Here, s_i denotes the strength of selection at locus i and h is the dominance coefficient. Fitness effects are multiplicative across loci, such that the fitness of an individual is given by $w = \prod_i w_i$, where w_i is the fitness effect of the ith locus of the focus individual, that is there is no epistasis. In contrast to absolute fitness, relative fitness is density independent in our model. This assumption seems conservative, because if the fitness of individuals was density dependent (i.e. individuals would have more similar fitness at low densities), neutral processes at the expansion front would become even more important. In the following, we will focus on codominant ($h = 0.5$) or fully recessive ($h = 0$) mutations and refer to the Fig. S6 (Supporting information) for results obtained for intermediate degrees of recessiveness. We assume that mutation effects are drawn from the same distribution of fitness effects (DFE) for all individuals (independently from their current fitness).

We perform individual-based simulations of the above-described model in 1D or 2D habitats. Our simulations start from ancestral populations located in 10 leftmost (rows of) demes of the range. After a burn-in phase that ensures that the ancestral populations are at mutation–selection–drift balance, the population expands from left to right until the habitat is filled. Because we are mainly interested in the role of standing genetic variation, we focus on relatively short expansions, that is colonization of a 1×50 (1D) or a 20×50 (2D) deme habitat. The long-term dynamics of expansion load have been studied elsewhere (Peischl *et al.* 2013, 2015).

Impact of standing genetic variation on expansion load

For simplicity, we first consider expansions along a one-dimensional habitat and assume that all mutations have the same effect, that is we set $s_i = s$, and investigate 2D habitats and more complex distributions of fitness effects in later sections. We mainly focus here on mildly deleterious mutations with effects on the order of $N_S = 1$, because these mutations contribute most to mutation load from standing genetic variation, and they have been shown to behave essentially like neutral mutations on the wave front during range expansions (Peischl *et al.* 2013). In Peischl *et al.* (2013), we derived an analytical approximation for the rate of change of mean fitness at the expansion front due to the establishment of new mutations. This approximation can be modified in a straightforward way to account for recessive mutations, and the numerical evaluation of these results is shown in Fig. 1.

If mutations are codominant ($h = 0.5$), expansion load is caused exclusively by the establishment of new mutations occurring during the expansion, and standing genetic variation has a negligible effect on the dynamics of mean fitness (Peischl *et al.* 2013). Mean fitness at the wave front decreases at a constant rate over time (Fig. 1), and the rate at which mean fitness decreases per generation is proportional to the number of new mutations entering the population per generation (Peischl *et al.* 2013).

The dynamics of expansion load changes dramatically if mutations are recessive (Fig. 1). The analytical approximation obtained in Peischl *et al.* (2013), which ignores standing genetic variation, is a poor fit to the observed dynamics of mean fitness (Fig. 1). In the first few generations, mean fitness decreases much faster than predicted by analytical theory for the accumulation of new mutations (cf. solid and dashed black lines in Fig. 1). Over the course of the expansion, the rate at which expansion load is created slows down and gradually approaches the analytical prediction. Then, changes in expected mean fitness arise exclusively from new mutations (cf. solid and dashed black lines for $t > 50$ in Fig. 1, see also Fig. S1, Supporting information). This shows that standing genetic variation plays

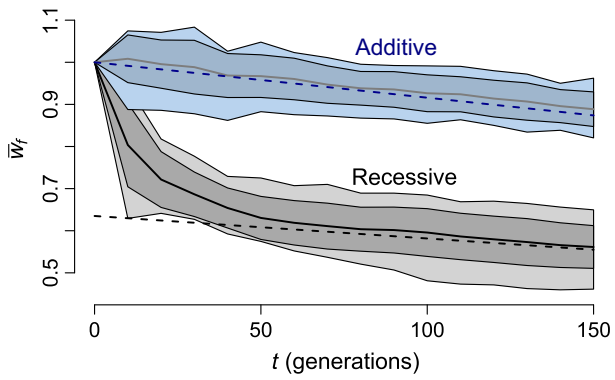


Fig. 1 Evolution of mean fitness at the wave front. Dashed lines show analytical predictions for the evolution of the mean fitness due to de novo mutations initially derived in Peischl et al. (2013), and adjusted here to account for the dominance coefficient of deleterious mutations by simply changing the selection coefficients of heterozygotes from $1-s/2$ to $1-hs$. In the codominant case, the analytical prediction is adjusted to match the mean fitness at the onset of the expansion, whereas in the recessive case it is adjusted to the mean fitness observed at generation t = 150. Simulations show results for the combination of standing and new genetic variation. Grey shaded areas and black lines show results for recessive mutations (h = 0), and blue shaded areas and lines show results for additive codominant mutations (h = 0.5). Solid lines indicate the average mean fitness from 50 simulations, and dark and light shaded areas indicate ± one standard deviation and the minimum and maximum of mean fitness, respectively. Other parameter values are n = 1000, K = 100, u = 0.1, m = 0.1, s = 0.01, R = 2 .

an important role in the establishment of expansion load if mutations are recessive, especially during early phases of expansions. This result is qualitatively similar to that of Kirkpatrick & Jarne (2000), who showed that load was always increasing after a single-generation bottleneck. The main novelty here is to consider the effects of recurrent bottlenecks such as those occurring during range expansions. For additional examples for the evolution of mean fitness during range expansions, including different migration rates (Fig. S2, Supporting information), carrying capacities (Fig. S3, Supporting information) and distribution of fitness effects (Fig. S4, Supporting information), we refer to the Supporting Information.

We next investigate the evolution of the genotypic composition of standing genetic variation on the front of an expansion along a single dimension. In general, we find that the change in levels of diversity is very similar in the neutral, codominant and recessive case. In all cases, strong drift at the expansion front leads to increased inbreeding, a reduction of heterozygosity and higher homozygosity (Fig. 2). Indeed, the average number of heterozygous loci per individual decreases during the expansion (Fig. 2A), whereas the number of loci that are homozygous for the derived allele increases (Fig. 2B). Because we simulated a fixed number of loci, the derived allele frequency shown in Fig. 2C is proportional to the average number of mutations carried by an individual. Note that for a given selection coefficient, the initial allele frequency in the core depends on the mode of selection, and is generally higher if mutations are recessive. However, Fig. 2C shows that the total number of mutations per individual remains nearly constant during the expansion. Therefore, range expansions (and bottlenecks) should have a relatively weak effect on the individual fitness component that is due to codominant mutations, in agreement with several recent observations (Fu *et al.* 2014; Simons *et al.* 2014; Do *et al.* 2015). Also, at any given locus, mutations are either lost or fixed over the course of the expansion, and the probability of fixation of a given mutation is close to its initial frequency (Peischl *et al.* 2013), suggesting that (mildly and moderately) deleterious mutations are behaving like neutral mutations on the wave front (dashed lines in Fig. 2A–C). In 2D expansions, the dynamics of genotype frequencies are qualitatively very similar to 1D expansions (Fig. S5, Supporting information). Strong genetic drift is therefore the major force driving the evolution of genotype frequencies at the wave front.

The nearly neutral evolution of allele frequencies on the expansion front reveals a critical role of the degree of dominance on the build-up of the expansion load. If mutations are codominant, the fitness of an individual is determined by the total number of mutations it carries (Wright 1930). Thus, Fig. 2C shows that in the codominant case, standing genetic variation would have a negligible impact on fitness. In contrast, if mutations are recessive, the fitness of an individual is determined by its number of loci homozygous for the derived allele. Because the number of derived homozygous loci per individual rapidly increases at the front of the expansions, standing genetic variation has a severe effect on fitness if mutations are recessive (Figs 1 and 2B).

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

In the following section, we focus on completely recessive mutations ($h = 0$). Figure 3 shows an example of the evolution of the mean fitness during an expansion in a 2D habitat (20 × 50 demes). As in 1D expansions, the mean fitness drops to low levels on the expansion front within the first few (≈30) generations and then continues to gradually decreases at a slower rate. There is however a considerable variation in fitness across the

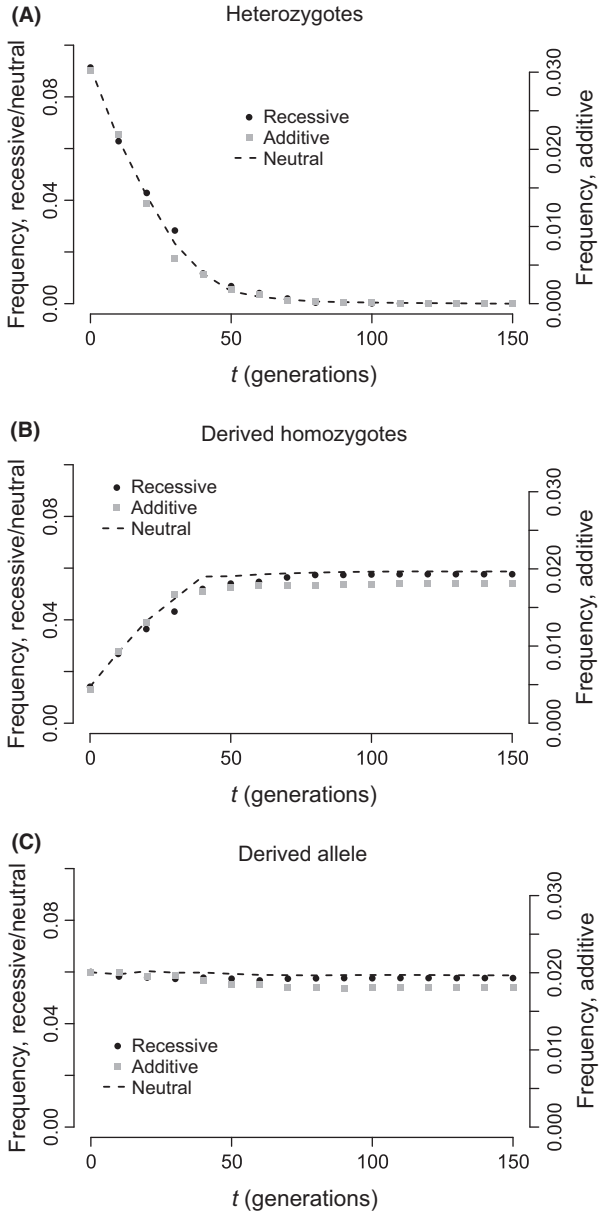


Fig. 2 Evolution of standing genetic variation on the wave front of a range expansion in one dimension. Each panel shows results for codominant, recessive and neutral mutations. For better comparison, neutral mutations were assumed recessive deleterious during the burn-in phase of the simulations, but neutral (in all demes) after the onset of the expansion. Parameter values are as in Fig. 1. Note the different scales on the y-axis for recessive (or neutral) and additive mutations in each plot.

wave front of 2D expansions (fitness differences of more than 40%, Figs 3 and 4). At the end of the expansion (Fig. 3, $t = 150$), we find a high-fitness ridge along the expansion axis in the central part of the newly settled species range, surrounded by sectors of low fitness on the lateral edges of the species range (see also Fig. 4A). This is partially caused by the lack of immigrants at the lateral edge of the species range (boundary effect). However, the location of the high-fitness ridge varies across simulation runs, suggesting that a boundary effect alone cannot explain the observed patterns (Fig. 4B).

Figure 4A shows the variation in fitness, heterozygosity and derived allele frequency across the wave front at the end of the expansion shown in Fig. 3. We find that the average number of mutations per individual is uniform across the expansion front, which means that the variation in fitness across the expansion front is not driven by a differential accumulation of mutations. Contrastingly, variation in heterozygosity across demes is substantial, ranging from demes with almost zero heterozygosity to demes with heterozygosity as high as before the onset of the expansion (cf. Figs 2A and 4A, C). Genetic variation is quickly lost along the expansion axis during the expansion, but gene flow between nearby demes having established different mutations at high frequency typically restores heterozygosity, especially after the expansion. Interestingly, diversity is lower in populations close to the lateral edges of the expansions due to a border effect translating into reduced gene flow, which explains the occurrence of lateral regions of low heterozygosity and mean fitness (Fig. 4B). Because the deleterious effects of recessive mutations are masked in heterozygotes, heterozygosity correlates strongly with mean fitness across the wave front (cf. solid and dashed line in Fig. 4A, and see Fig. 4C).

Using a generalized linear mixed model, we estimate the effects of several key quantities that can be measured in experimental set-ups (deme coordinates, simulation number (random effect) and average heterozygosity as independent variables, and mean fitness as dependent variable), to identify whether we find different heterozygosity–fitness correlations (HFC) in core and in front populations. We indeed find a strongly positive HFC at the front of the expansion (Fig. 4C, regression slope ≈ 0.4, $P \approx 10^{-5}$, see Table S1, Supporting information), but not in the ancestral population (Fig. 4D, $P \approx 0.7$, see Table S2, Supporting information). The contributions of all other parameters to mean fitness are not significantly different from zero (see Table S1 and S2, Supporting information). Interestingly, weaker but similar correlations are found at the individual level within demes, suggesting that HFC created after range expansions could be detected in samples from single populations. We performed linear regression of heterozygosity and fitness for 50 individuals sampled from the same deme. Repeating this across demes at the wave front (at generation $t = 150$) and across simulation replicates, we found an average

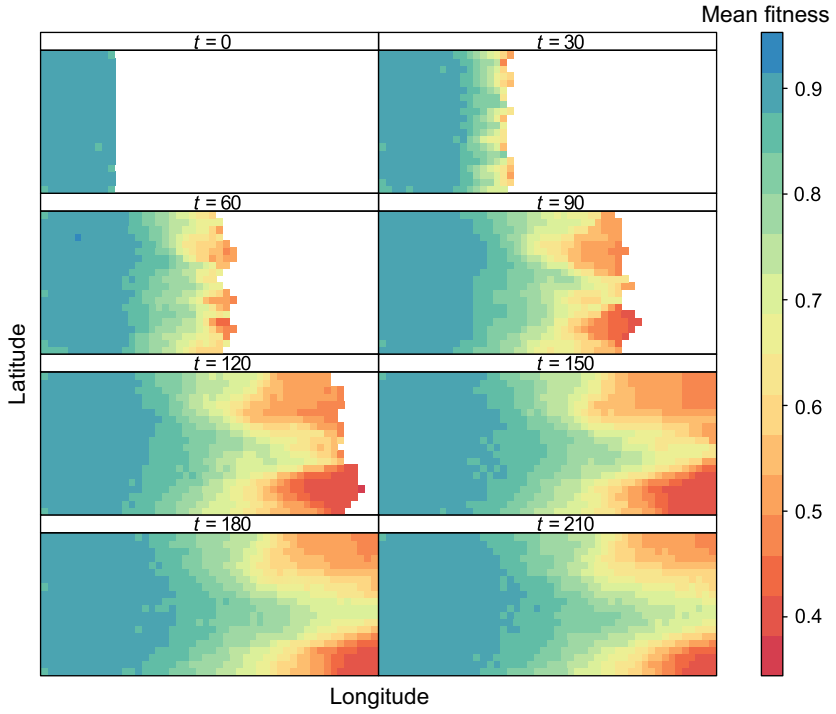


Fig. 3 Evolution of mean fitness during a range expansion. The simulated grid is 20 × 50 demes. Mutations are recessive and parameter values are as in Fig. 1.

regression slope of ≈0.1 ($P < 0.05$ in 78% of demes, Figs 5 and S8, Supporting information). Furthermore, we found that latitudinal position of demes has no significant effect on the strength of within deme HFC (Table S3, Supporting information).

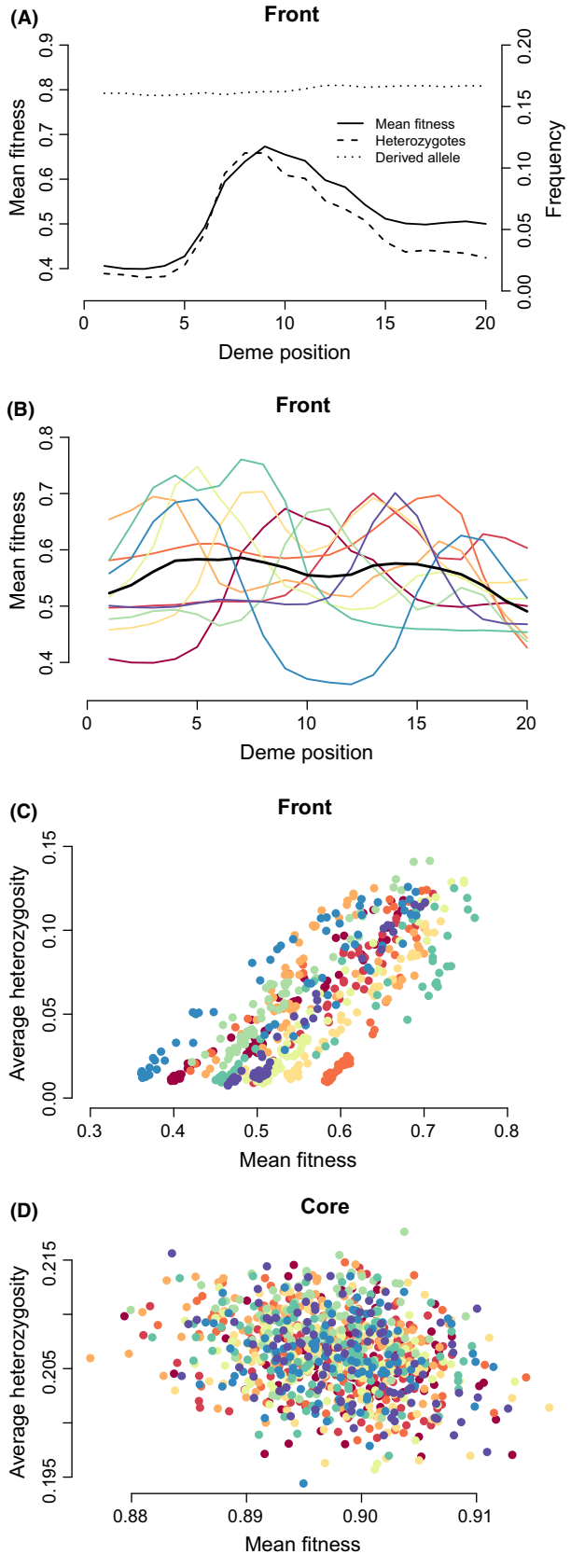
Expansion load is driven by a few mutations occurring at high frequency

So far we assumed that all mutations had the same effect s. To investigate the composition of expansion load with respect to mutation fitness effects, we now consider the case where mutation fitness effects are drawn from an exponential distribution with mean s. Figure 6A,B shows the site frequency spectrum (SFS) observed in core and front populations, respectively. In core populations, the SFS shows the pattern expected for sites under negative selection (Bustamante *et al.* 2001), with a large excess of low-frequency variants. On the wave front, the total number of segregating sites is reduced in marginal populations (cf. Fig. 6A,B). More interestingly, as compared to core populations, we see a markedly different SFS on the front, with a clear deficit of rare and intermediate-frequency variants and an increase in high-frequency variants (Fig. 6B). Thus, even though fewer polymorphic sites with deleterious variants are found in more recently colonized areas than in the ancestral region, the alleles at polymorphic sites tend to be at higher frequency in more recently colonized populations.

Figure 6C,D shows the distribution of polymorphic loci stratified according to their mutation effect sizes. The eight mutation effect classes have been defined such that they represent the eight quantiles of the DFE, that is the rate at which mutations of a given category enter the population are equal for all categories. As expected, we find that the number of polymorphic loci generally decreases with increasing mutation effect size and that large-effect mutations tend to be present at lower frequencies than low-effect mutations (see Fig. 6C,D). Compared to core populations, the allele frequencies at polymorphic sites on the wave front tend to be larger across all mutational effect categories (cf. yellow and red coloured areas in Fig. 6C,D). Furthermore, the increase in allele frequency is most pronounced for small effect mutations. Thus, expansion load is driven mainly by standing deleterious mutations of small to moderate effect (i.e. up to $N_S < 2$ for the parameter values used in Fig. 6, see also Fig. S7 (Supporting information) which shows analogous results for larger mean s) that rise to high frequency during the expansion.

Discussion

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



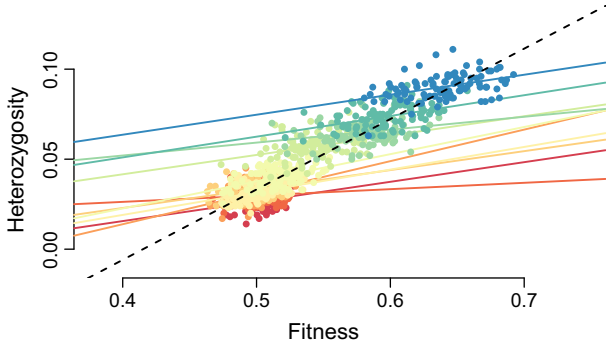


Fig. 5 Within deme heterozygosity–fitness correlations (HFC). Each point corresponds to the fitness and heterozygosity of a single individual, as inferred from a single simulation on a 20 by 50 grid. Colours correspond to different demes at the wave front (at generation t = 150). Coloured lines show the linear regression lines of heterozygosity and fitness within demes. The dashed black line shows the overall (between deme) HFC for comparison.

individual-based simulations, we have shown that shifts in the genotypic composition of standing genetic variation can lead to a rapid drop of mean fitness at the onset of an expansion (see Figs 1 and 2, and Figs S2–S4, Supporting information) without necessarily affecting the total number of deleterious alleles per individuals (see Figs 2, 4, and S5, Supporting information). Figure 2 shows that genotype frequencies evolve almost neutrally at the expansion front and that strong genetic drift at the expansion front increases the number of derived homozygote sites per individual. The derived homozygote frequency at the expansion front approaches the initial frequency of the derived allele over the course of the expansion (see Fig. 2B). The total expansion load due to standing genetic variation, which is proportional to the number of derived homozygous

Fig. 4 Genetic properties of demes at the wave front and heterozygosity–fitness correlations (HFC) across demes in different parts of the species range. (A) Example of the mean fitness, the heterozygosity and the derived allele frequency at different latitudinal positions at the expansion front when the habitat has just been fully colonized (t = 150, simulation shown in Fig. 3). The deme mean fitness on the wave front correlates with heterozygosity, but not with derived allele frequency. (B) Mean fitness at the front of the expansion from 10 distinct simulation runs (coloured lines) and the average over all simulation runs (solid black line). (C) HFC on the expansion front at generation t = 150. (D) No significant HFC in core populations before the onset of the expansion (t = 0). In (C) and (D), each point represents the mean fitness and average heterozygosity of a single deme. Different colours in panels B–D correspond to 10 distinct simulation replicates. Parameter values are as in Fig. 3.

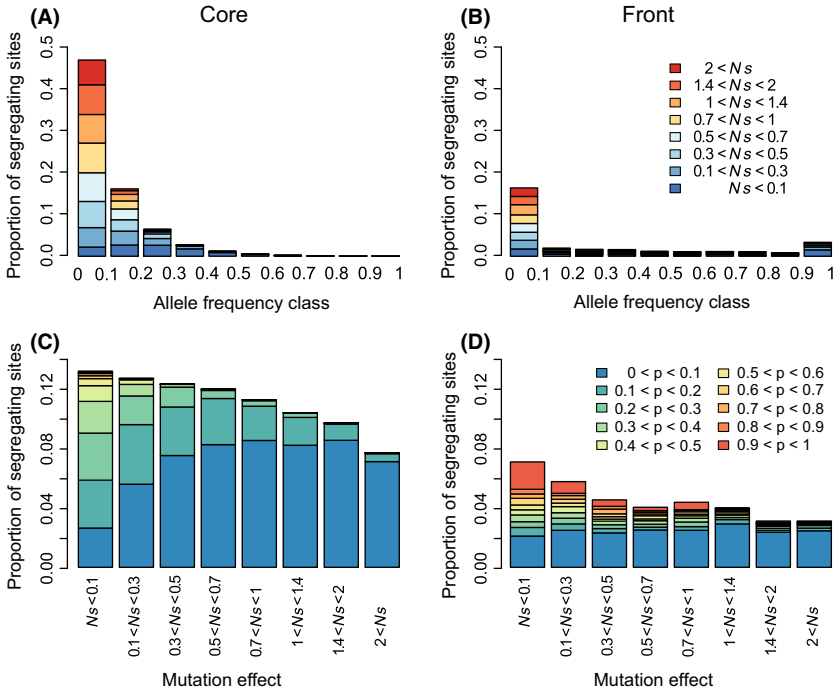


Fig. 6 Distributions of segregating sites in core and front populations after 2D range expansions. The distributions are stratified according to allele frequencies (site frequency spectrum, top row) and mutation effects (bottom row). The plots are normalized with respect to the total number of sites that segregating in any of the colonized demes. In the top row, allele frequencies (p) are binned in decimal intervals, and the boundaries of these intervals are indicated on the x-axis. Results were recorded 150 generations after the onset of the expansion, which is shortly after the habitat was colonized completely (mean time to colonization ≈130 generations, see also Fig. 3). Panels (A) and (C) show results for a core population [XY coordinates (5, 10)], (B) and (D) for a front population [XY *coordinates* (45, 10)]. Mutations are recessive and their effects are drawn from an exponential distribution with mean s = 0.01. Other parameter values are as in Fig. 1.

sites, is therefore limited by the initial frequency of deleterious mutations. Thus, if many loci are polymorphic for deleterious variants at the onset of the expansion, the (recessive) expansion load from standing genetic variation can dominate the total mutation load (see Fig. 1). A similar phenomenon, although of lesser magnitude, occurs if only some of the mutations would be fully or if they would be partially recessive (see Fig. S6, Supporting information).

Even though we find that our results are robust with respect to changes in the migration rate (Fig. S2, Supporting information) and carrying capacity (Fig. S3, Supporting information), it would be interesting to further explore the parameter space. Unfortunately, individual-based simulations are computationally intensive, especially, if both N and m are large, which prevents an exhaustive exploration of the parameter space. Theoretical results suggest, however, that the effective population size at the front of expanding populations depends only weakly on the local carrying capacities (Hallatschek & Nelson 2008). Similarly, migration rates seem to have a weak effect on the effective population size at the expansion front (Peischl *et al.* 2013; see also Fig. S2, Supporting information). Allee effects, that is a reduction in fitness when conspecific density is low, strongly decrease the strength of drift at expansion fronts (Hallatschek & Nelson 2008; Roques *et al.* 2012) and could therefore mitigate expansion load. Importantly, Allee effects appear to be important in many examples of range expansions or invasive species (Green 1997; Taylor & Hastings 2005).

The effect of range expansions on deleterious genetic diversity is also reflected in the site frequency spectrum (SFS, see Fig. 6). As compared to stationary populations in the core of the species range, populations from more recently colonized areas have fewer segregating sites, but proportionally more high- and low-frequency variants (cf. Fig. 6A,B). These differences in the SFS of core and front populations should provide an opportunity to evidence expansion load from sequence data and to infer important quantities such as the distribution of fitness effects (Keightley & Eyre-Walker 2007; Boyko *et al.* 2008; Racimo & Schraiber 2014). The development of statistical and computational methods able to infer parameters under spatially explicit models including range expansions and selection remains, however, a major challenge (Sousa *et al.* 2014).

Interestingly, human genomic data are consistent with our predictions for genomic signatures of expansion load. In particular, the number of segregating sites is higher in African populations than in non-African populations (Lohmueller *et al.* 2008), non-African populations show an excess of low-frequency and high-frequency deleterious alleles (Lohmueller *et al.* 2008; Fu *et al.* 2014), the average number of sites that are homozygous for predicted deleterious variants sites is larger in non-African individuals (Fu *et al.* 2014), and the average number of predicted deleterious mutations per individual is slightly, but significantly, larger in non-Africans (Fu *et al.* 2014). Determining mutation load (or, alternatively, fitness) from genomic variation data is, however, an intrinsically difficult problem

because mutation load depends on many unknown parameters (selection coefficients that may vary over space and time, epistatic interactions, dominance relationships, etc.), and the relevance of comparing a population with deleterious mutations to a theoretical population free of such mutations is questionable (Lesecque *et al.* 2012). Testing theoretical predictions of the effect of a range expansion, or other demographic scenarios (e.g. Simons *et al.* 2014; Do *et al.* 2015), on functional diversity with human genomic data might nevertheless be extremely useful to substantially increase our understanding of the complex interactions of demography and selection.

We assumed here that fitness is density independent, but one could imagine that fitness differences between individuals could be decreased at low densities, making drift relatively stronger than selection on range margins, promoting expansion load even further than in the cases studied here. We also assumed here that selection is soft, that is demographic parameters are independent of fitness (Wallace 1975), but it would be interesting to extend our results to models of hard selection, where mutation load on the front can stop an expansion and even drive parts of the species range to extinction (Peischl *et al.* 2015). Our results suggest that admixture during range expansions, or secondary contact between expanding lineages, could mitigate expansion load and prevent marginal populations from collapsing. A previous study of range expansions under an additive model with hard selection has shown that suppressing recombination at the wave front can have beneficial effects for the spread of high-fitness lineages (Peischl *et al.* 2015). Recombination modifiers, such as inversions, could have a similar effect if mutations are recessive and facilitate the spread of admixed lineages. An interesting example for studying the potentially beneficial role of admixture and suppressed recombination during range expansions is from the clam genus *Corbicula*, which includes both sexual and asexual (androgenetic diploid) lineages. Sexual populations are restricted to their native Asian areas, but the androgenetic lineages are widely distributed and extend as far as in America and Europe where they are invasive (Pigneur *et al.* 2014). Intriguingly, the invasive lineages also show an excess of heterozygosity, which is preserved through clonal reproduction. No such excess of heterozygosity is found in the native range, suggesting that the combination of asexual reproduction and high heterozygosity may have been key drivers of the invasion.

An interesting prediction of our model is that if a given proportion of deleterious mutations are recessive, then heterozygosity–fitness correlations (HFC) should naturally occur in populations that have recently expanded their range (see Fig. 4A,C). Importantly, the positive correlation between heterozygosity and fitness in recently colonized areas can be observed at both the individual level and the population level (see Figs 4 and 5). Even though our simulations modelled a single expansion in a 2D habitat, we would expect similar HFCs if there was a secondary contact between expanding populations from different areas (e.g. from different LGM refuge areas). The HFC should be even stronger in the case of a secondary contact, because the isolation between expanding lineages should be larger and different recessive alleles could have fixed in different refugia or during the expansion from these refugia. HFC have been observed in many cases of natural range expansions and invasive species (Chapman *et al.* 2009), but their underlying mechanisms and their role during range expansions and invasions are still unclear (Szulkin *et al.* 2010; Rius & Darling 2014). A particular interesting example of HFC is found in the invasive weed *Silene vulgaris*, where, as predicted by our model (see Figs 4 and 5), HFC correlations are observed in the recently invaded North American range, but not in their native European range (Keller *et al.* 2014). It remains, however, unclear whether admixture between divergent lineages has indeed a causal role in range expansions. A combination of transplantation experiments and genomic data analyses could certainly be used to test the predictions of our model.

In summary, we have investigated here the evolution of standing genetic variation during range expansions, the dynamics of mean fitness on the expansion front if mutations are recessive, and the genomic signature of range expansions. Importantly, our results make predictions that can be tested in natural populations. Empirical validation of our results would increase our understanding of the interactions of demography and selection (Lohmueller 2014) and could help us identifying key drivers of range expansions and biological invasions (Rius & Darling 2014).

Acknowledgements

SP was supported by a Swiss NSF Grant No. 31003A-143393 to LE. We thank Vitor Sousa and Isabelle Duperret for stimulating discussions on this topic. We are grateful for the helpful comments from four anonymous reviewers.

References

Agrawal AF, Whitlock MC (2011) Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics*, **187**, 553–566.

Balick DJ, Do R, Reich D, Sunyaev SR (2014) Response to a population bottleneck can be used to infer recessive selection. *bioRxiv*, 003491.

Beverton R, Holt S (1957) On the dynamics of exploited fish populations. Fisheries Investigation Series 2 (19). London: Ministry of Agriculture. Fisheries and Food.

Boyko AR, Williamson SH, Indap AR *et al.* (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genetics*, **4**, e1000083.

Burton OJ, Travis JMJ (2008) The frequency of fitness peak shifts is increased at expanding range margins due to mutation surfing. *Genetics*, **179**, 941–950.

Bustamante CD, Wakeley J, Sawyer S, Hartl DL (2001) Directional selection and the site-frequency spectrum. *Genetics*, **159**, 1779–1788.

Chapman J, Nakagawa S, Coltman D, Slate J, Sheldon B (2009) A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology*, **18**, 2746–2765.

Darwin C (1859) *On the Origins of Species by Means of Natural Selection*. Murray, London.

Do R, Balick D, Li H *et al.* (2015) No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nature Genetics*, **47**, 126–131.

Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 975–979.

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, **8**, 610–618.

Fu W, Gittelman RM, Bamshad MJ, Akey JM (2014) Characteristics of neutral and deleterious protein-coding variation among individuals and populations. *The American Journal of Human Genetics*, **95**, 421–436.

Garcia-Dorado A, Caballero A (2000) On the average coefficient of dominance of deleterious spontaneous mutations. *Genetics*, **155**, 1991–2001.

Graciá E, Botella F, Anadón JD *et al.* (2013) Surfing in tortoises? Empirical signs of genetic structuring owing to range expansion. *Biology Letters*, **9**, 20121091.

Gravel S (2014) When is selection effective? bioRxiv, 010934.

Green RE (1997) The influence of numbers released on the outcome of attempts to introduce exotic bird species to New Zealand. *Journal of Animal Ecology*, **66**, 25–35.

Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. *Theoretical Population Biology*, **73**, 158–170.

Keightley PD, Eyre-Walker A (2007) Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics*, **177**, 2251–2261.

Keller S, Fields P, Berardi A, Taylor D (2014) Recent admixture generates heterozygosity–fitness correlations during the range expansion of an invading species. *Journal of Evolutionary Biology*, **27**, 616–627.

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, **49**, 561–576.

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

Kirkpatrick M, Jarne P (2000) The effects of a bottleneck on inbreeding depression and the genetic load. *American Naturalist*, **155**, 154–167.

Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, **23**, 482–490.

Lehe R, Hallatschek O, Peliti L (2012) The rate of beneficial mutations surfing on the wave of a range expansion. *PLoS Computational Biology*, **8**, e1002447.

Lesecque Y, Keightley PD, Eyre-Walker A (2012) A resolution of the mutation load paradox in humans. *Genetics*, **191**, 1321–1330.

Lindström T, Brown GP, Sisson SA, Phillips BL, Shine R (2013) Rapid shifts in dispersal behavior on an expanding range edge. *Proceedings of the National Academy of Sciences*, **110**, 13452–13456.

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

Lohmueller KE, Indap AR, Schmidt S *et al.* (2008) Proportionally more deleterious genetic variation in European than in African populations. *Nature*, **451**, U994–U995.

MacArthur RH (1972) *Geographical Ecology: Patterns in the Distribution of Species*. Princeton University Press, Oxfordshire, UK.

Moreau C, Bherer C, Vezina H *et al.* (2011) Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science*, **334**, 1148–1150.

Morton NE, Crow JF, Muller HJ (1956) An estimate of the mutational damage in man from data on consanguineous marriages. *Proceedings of the National Academy of Sciences of the United States of America*, **42**, 855.

Mukai T, Chigusa SI, Mettler L, Crow JF (1972) Mutation rate and dominance of genes affecting viability in *Drosophila* melanogaster. *Genetics*, **72**, 335–355.

Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, **22**, 5972–5982.

Peischl S, Kirkpatrick M, Excoffier L (2015) Expansion load and the evolutionary dynamics of a species range. *American Naturalist*, **185**, 1–13.

Peters A, Halligan D, Whitlock M, Keightley P (2003) Dominance and overdominance of mildly deleterious induced mutations for fitness traits in *Caenorhabditis elegans*. *Genetics*, **165**, 589–599.

Phillips BL, Brown GP, Webb JK, Shine R (2006) Invasion and the evolution of speed in toads. *Nature*, **439**, 803.

Phillips BL, Brown GP, Shine R (2010) Life-history evolution in range-shifting populations. *Ecology*, **91**, 1617–1627.

Pigneur LM, Etoundi E, Aldridge DC, Marescaux J, Yasuda N, Van Doninck K (2014) Genetic uniformity and long-distance clonal dispersal in the invasive androgenetic Corbicula clams. *Molecular Ecology*, **23**, 5102–5116.

Racimo F, Schraiber JG (2014) Approximation to the distribution of fitness effects across functional categories in human segregating polymorphisms. *PLoS Genetics*, **10**, e1004697.

Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology and Evolution*, **29**, 233–242.

Roques L, Garnier J, Hamel F, Klein EK (2012) Allee effect promotes diversity in traveling waves of colonization. *Proceedings of the National Academy of Sciences*, **109**, 8828–8833.

Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009) Evolution and ecology of species range limits. *Annual Review of Ecology Evolution and Systematics*, **40**, 415–436.

Shine R, Brown GP, Phillips BL (2011) An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences*, **108**, 5708.
Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. *Nature Genetics*, **46**, 220–224.
Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics*, **191**, 171–181.
Sousa V, Peischl S, Excoffier L (2014) Impact of range expansions on current human genomic diversity. *Current Opinion in Genetics and Development*, **29**, 22–30.
Subramanian S (2012) The abundance of deleterious polymorphisms in humans. *Genetics*, **190**, 1579–1583.
Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: a time for reappraisal. *Evolution*, **64**, 1202–1217.
Taylor CM, Hastings A (2005) Allee effects in biological invasions. *Ecology Letters*, **8**, 895–908.
Torkamani A, Pham P, Libiger O *et al.* (2012) Clinical implications of human population differences in genome-wide rates of functional genotypes. *Frontiers in Genetics*, **3**, 1–19.
Travis JMJ, Munkemuller T, Burton OJ *et al.* (2007) Deleterious mutations can surf to high densities on the wave front of an expanding population. *Molecular Biology and Evolution*, **24**, 2334–2343.
Wallace B (1975) Hard and soft selection revisited. *Evolution*, **29**, 465–473.
Wright S (1930) The genetical theory of natural selection. A review. *Journal of Heredity*, **21**, 349–356.

L.E. and S.P. conceived and designed the study. S.P. wrote the simulation code and analyzed the simulation output. L.E. and S.P. wrote the paper.

Data accessibility

The code used for the simulations is available on GitHub: https://github.com/CMPG/ADMRE.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Evolution of mean fitness on the front of a 1D expansion.

Fig. S2 Evolution of mean fitness at the front of a 2D expansion, in a habitat of 10 × 50 demes.

Fig. S3 Evolution of mean fitness at the front of a 2D expansion.

Fig.S4 Evolution of mean fitness at the front of a 2D expansion.

Fig. S5 Evolution of genotype frequencies at the front of a two-dimensional expansion, in a habitat of 20 × 50 demes.

Fig. S6 Evolution of mean fitness at the front of the expansion for varying degrees of dominance.

Fig. S7 Distributions of segregating sites in core and front population after a 2D range expansion.

Fig. S8 Slope and *P*-values of linear regression of individual heterozygosity and fitness.

Table S1 HFC on the expansion front.

Table S2 HFC in core populations.

Table S3 Within-deme HFC.