HAL
open science

Heterogeneous gene duplications can be adaptive because they permanently associate overdominant alleles

Pascal Milesi, Mylène Weill, Thomas Lenormand, Pierrick Labbe

To cite this version:

Pascal Milesi, Mylène Weill, Thomas Lenormand, Pierrick Labbe. Heterogeneous gene duplications can be adaptive because they permanently associate overdominant alleles. Evolution Letters, 2017, 1 (3), pp.169-180. 10.1002/evl3.17 . hal-01938023

HAL Id: hal-01938023

https://hal.science/hal-01938023

Submitted on 28 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

doi:10.1002/evl3.17

Open Access

Heterogeneous gene duplications can be adaptive because they permanently associate overdominant alleles

Pascal Milesi,[1] Mylène Weill,[1] Thomas Lenormand,[2] and Pierrick Labbé[1,3]

[1]*Institut des Sciences de l'Evolution de Montpellier (UMR 5554, CNRS-Université de Montpellier-IRD-EPHE), Campus Université de Montpellier, Place Eugène Bataillon 34095, Montpellier, CEDEX 05, France*

[2]*Centre d'Ecologie Fonctionnelle et Evolutive (UMR 5175, CNRS-Université de Montpellier-Université Paul-Valéry Montpellier-EPHE) 1919 route de Mende, F-34293 Montpellier, CEDEX 05, France*

[3]*E-mail: pierrick.labbe@umontpellier.fr*

Received April 5, 2017
Accepted June 23, 2017

Gene duplications are widespread in genomes, but their role in contemporary adaptation is not fully understood. Although mostly deleterious, homogeneous duplications that associate identical repeats of a locus often increase the quantity of protein produced, which can be selected in certain environments. However, another type exists: heterogeneous gene duplications, which permanently associate two (or more) alleles of a single locus on the same chromosome. They are far less studied, as only few examples of contemporary heterogeneous duplications are known. Haldane proposed in 1954 that they could be adaptive in situations of heterozygote advantage, or overdominance, but this hypothesis was never tested. To assess its validity, we took advantage of the well-known model of insecticide resistance in mosquitoes. We used experimental evolution to estimate the fitnesses associated with homozygous and heterozygous genotypes in different selection regimes. It first showed that balanced antagonist selective pressures frequently induce overdominance, generating stable polymorphic equilibriums. The frequency of equilibrium moreover depends on the magnitude of two antagonistic selective pressures, the survival advantage conferred by the resistant allele versus the selective costs it induces. We then showed that heterogeneous duplications are selected over single-copy alleles in such contexts. They allow the fixation of the heterozygote phenotype, providing an alternative and stable intermediate fitness trade-off. By allowing the rapid fixation of divergent alleles, this immediate advantage could contribute to the rarity of overdominance. More importantly, it also creates new material for long-term genetic innovation, making a crucial but underestimated contribution to the evolution of new genes and gene families.

KEY WORDS: genetic polymorphism, gene duplication, genome evolution, overdominance, balancing selection, insecticide resistance.

Impact Summary

Understanding the maintenance of polymorphism in natural populations despite the erosion due to natural selection and genetic drift has been one of the challenges of the early 20th century. One of the propositions, rapidly put aside due to the lack of examples was overdominance: when the heterozygote phenotype is the fittest, none of the alleles can fix and only half of the progeny of two heterozygotes carry the best phenotype (this is called segregation burden). However, in 1954 Haldane suggested that heterogeneous duplications associating the two alleles on the same chromosome could be selected, but could also allow the fixation of the heterozygote phenotype. This hypothesis was however never tested, because both overdominance and gene duplications were considered too rare. Recent genome-wide studies showed that duplications are actually more common than substitutions. However, they

© 2017 The Author(s). *Evolution Letters* published by Wiley Periodicals, Inc. on behalf of Society for the Study of Evolution (SSE) and European Society for Evolutionary Biology (ESEB).
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Evolution Letters

usually associate identical repeats of the same allele, which can increase the protein production and be selected for (a quantitative advantage). The recent discovery of several heterogeneous duplications implicated in insecticide resistance in mosquitoes provided means to finally test Haldane's hypothesis. Insecticide-resistant individuals are advantaged in the presence of insecticide (they survive) but endure severe selective costs (their general performances are reduced) compared to susceptible individuals. Using experimental evolution and manipulating the selective pressures (insecticide dose for the advantage and the rearing conditions for the cost), we first showed that balanced antagonist pressures generate overdominance. Introducing a duplicated allele latter showed that it was indeed selected over single copies, as predicted by Haldane. Our study thus shows that heterogeneous duplications associating two already divergent alleles can be selected when the environmental conditions result in trade-offs favoring overdominance. However, they also create new and already divergent material for genetic innovation and could play a crucial role in the evolution of gene families.

Introduction

Recent genomic studies have shown that gene duplications are widespread, but mostly deleterious (Schrider and Hahn 2010). Some are nevertheless adaptive, and many examples of quantitative advantages (i.e., increased protein production) have been reported associated with identical repeats of a locus (homogeneous duplications) (reviewed in Innan and Kondrashov 2010; Katju and Bergthorsson 2013). Another kind of duplication associates two or more alleles of a single locus (i.e., divergent copies) on the same chromosome (heterogeneous duplications, Fig. 1). Their adaptive role is much harder to assess (Lenormand et al. 1998; Labbé et al. 2007a; Innan and Kondrashov 2010), as only a few examples of recent heterogeneous duplications have been described to date (Labbé et al. 2007a; Djogbénou et al. 2008; Remnant et al. 2013; Sonoda et al. 2014). In 1954, Haldane suggested that they could be selected for by enabling the permanent association of overdominant heterozygous alleles, with no segregation burden (Haldane 1954). In the early 20th century, overdominance (i.e., the situation in which the heterozygote is fitter than either of the homozygotes) was originally put forward as one of several mechanisms underlying the hybrid vigor observed in crosses between inbred strains, and as the only condition allowing a polymorphic equilibrium with constant fitnesses in population genetics equations (Fisher 1922; East 1936; Dobzhansky 1952). Sickle cell anemia is an iconic and much cited example (Lewontin 1974). However, empirical evidence for overdominance in natural populations remains scarce (Hedrick 2012; Llaurens et al. 2017), and its contribution to the selection of heterogeneous duplications has never been assessed.

In several mosquito species, including the malaria vector *Anopheles gambiae* and the West Nile virus vector *Culex pipiens*, the *ace-1* R resistance allele encodes an acetylcholinesterase protein differing from that encoded by the susceptible S allele by a single amino-acid substitution (G119S): this substitution prevents organophosphate insecticides (OPs) from binding to their target, resulting in resistance (Weill et al. 2003). However, the G119S substitution also greatly decreases the activity of the protein (Alout et al. 2008), probably accounting for the high selective cost associated with homozygous RR individuals: higher larval mortality, lower fecundity, and lower mating success for RR males than for homozygous SS males (Duron et al. 2006; Assogba et al. 2015, 2016). This situation results in a fitness trade-off, with the selection of RR individuals in the presence of insecticides, and selection against these individuals in the absence of pesticides. RS individuals have intermediate characteristics, with lower resistance than RR individuals, at a lower cost (Labbé et al. 2014; Assogba et al. 2015). This resistance/cost trade-off, as well as the migration/selection balance resulting from the alternating treated and nontreated areas, promote *ace-1* polymorphism (Lenormand et al. 1999; Labbé et al. 2007b).

Gene duplications (D alleles) bringing a susceptible S copy and a resistant R copy together in the same haplotype, on a single chromosome, have been identified in both *A. gambiae* and *C. pipiens* (Lenormand et al. 1998; Labbé et al. 2007a; Djogbénou et al. 2008). In *C. pipiens*, several independent duplications have been identified, and some may fix in natural populations (Labbé et al. 2007a; Alout et al. 2010; Osta et al. 2012). Both R and S copies are identical to single-copy S and R alleles found in the same populations, so that the D alleles probably result from unequal crossing-overs in heterozygotes (Labbé et al. 2007a). Most importantly these heterogeneous duplications confer a heterozygote-like phenotype [RS] (Labbé et al. 2014).

This model thus provides us with a unique opportunity to test Haldane's hypothesis: moderate doses of insecticide could result in balanced selective pressures, where the advantage/cost trade-off could favor the intermediate performances of heterozygotes, and thus the duplicated alleles.

We used experimental evolution to test these hypotheses. Our study showed that antagonist selective pressures could lead to overdominance, in which case the ultimate allele frequencies depended on the balance between the relative intensities of these selective pressures. It also showed that heterogeneous duplications were indeed selected in these contexts. Half a century later, our study thus provides strong experimental support for a theoretical claim made by one of the fathers of modern evolutionary theory that has long remained untested.

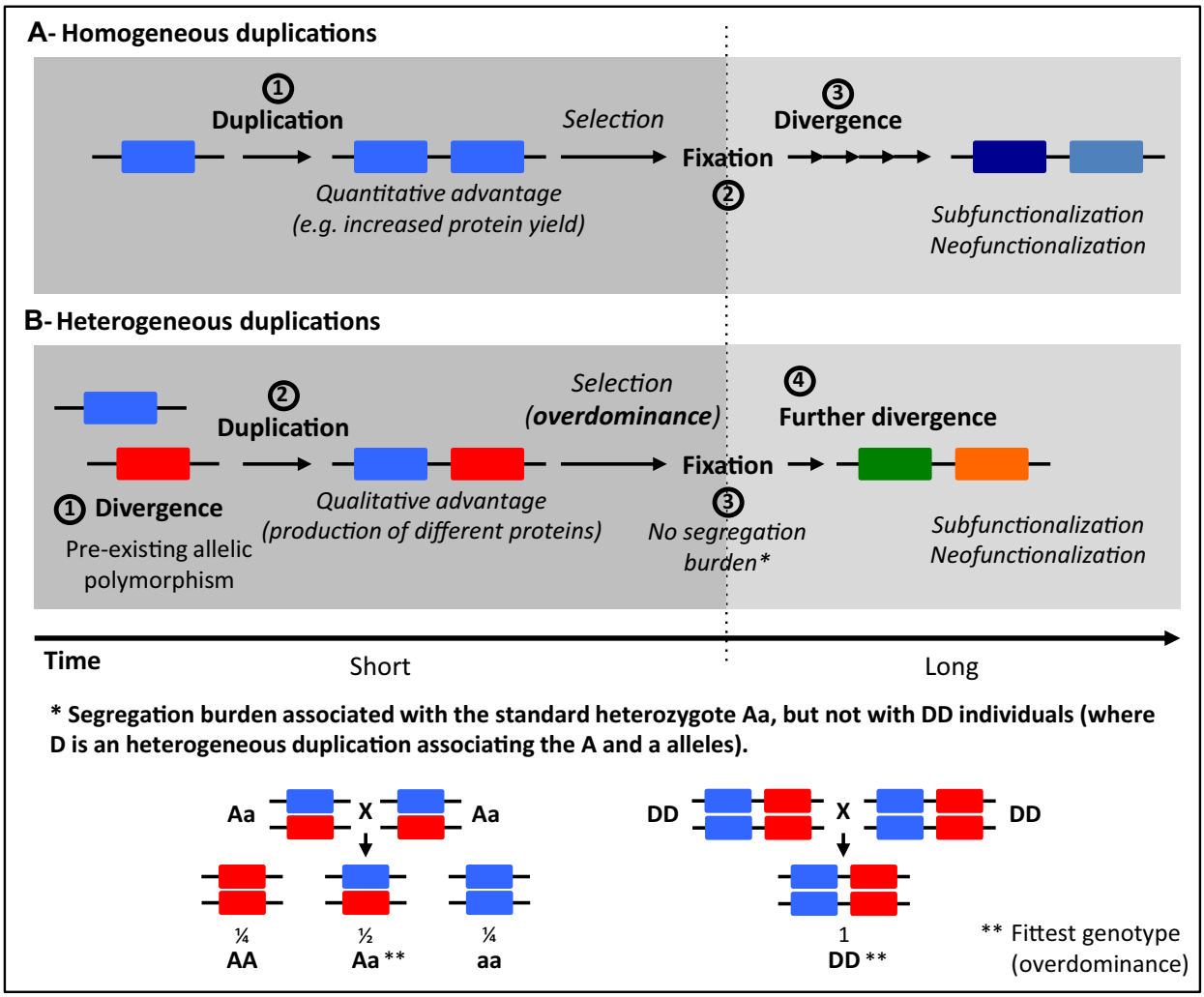

Figure 1. Origins and adaptive fates of the different types of gene duplications.
Each rectangle represents one copy of a given locus; the colors indicate that two copies/alleles differ in sequence and are associated with different functions. The timing of the various processes is indicated by the long black arrow, and the background colors differentiate the short and longer term processes.
(A) Homogeneous duplications associate two identical copies of a gene (1). They can be selected, for example, for quantitative advantages due to increased protein yield, and get fixed (2). On the long term, they can diverge and acquire new functions (3).
(B) In the case of heterogeneous duplications, the allelic divergence (1) precedes the duplication (2). They can be selected in case of overdominance because they allow the production of different proteins, and get fixed because they do not endure the segregation burden associated with standard heterozygotes (the * diagram illustrates the progenies of the two crosses and their proportions) (3). They carry different functions before fixation, but can then further diverge and acquire new functions (4).

Methods

MOSQUITO STRAINS

Three *C. pipiens* mosquito laboratory lines sharing the same genetic background (>99%) were used in this study: Slab (Georghiou et al. 1966), SR (Berticat et al. 2002), and Ducos-DFix (Labbé et al. 2014). All share the Slab genetic background, but they carry different *ace-1* alleles: the single-copy susceptible S (isolated in California), the single-copy resistance R (isolated in Southern France), and the heterogeneous resistance duplication D (D_1 allele, isolated in Martinique), respectively. D copies are identical except for the G119S mutation, and differ from the S and R alleles by a few synonymous mutations (Labbé et al. 2007a).

EXPERIMENTAL EVOLUTION IN POPULATION CAGES

At the start of each experiment replicate, we mixed 500 second-instar larvae (L_2) of the different strains in various proportions, to control for any effect of initial conditions or genetic drift. The larvae were reared until the emergence of the adult. The adults were allowed to mate freely and the females were provided with a blood meal (chicken). Thereafter, the experiment was conducted

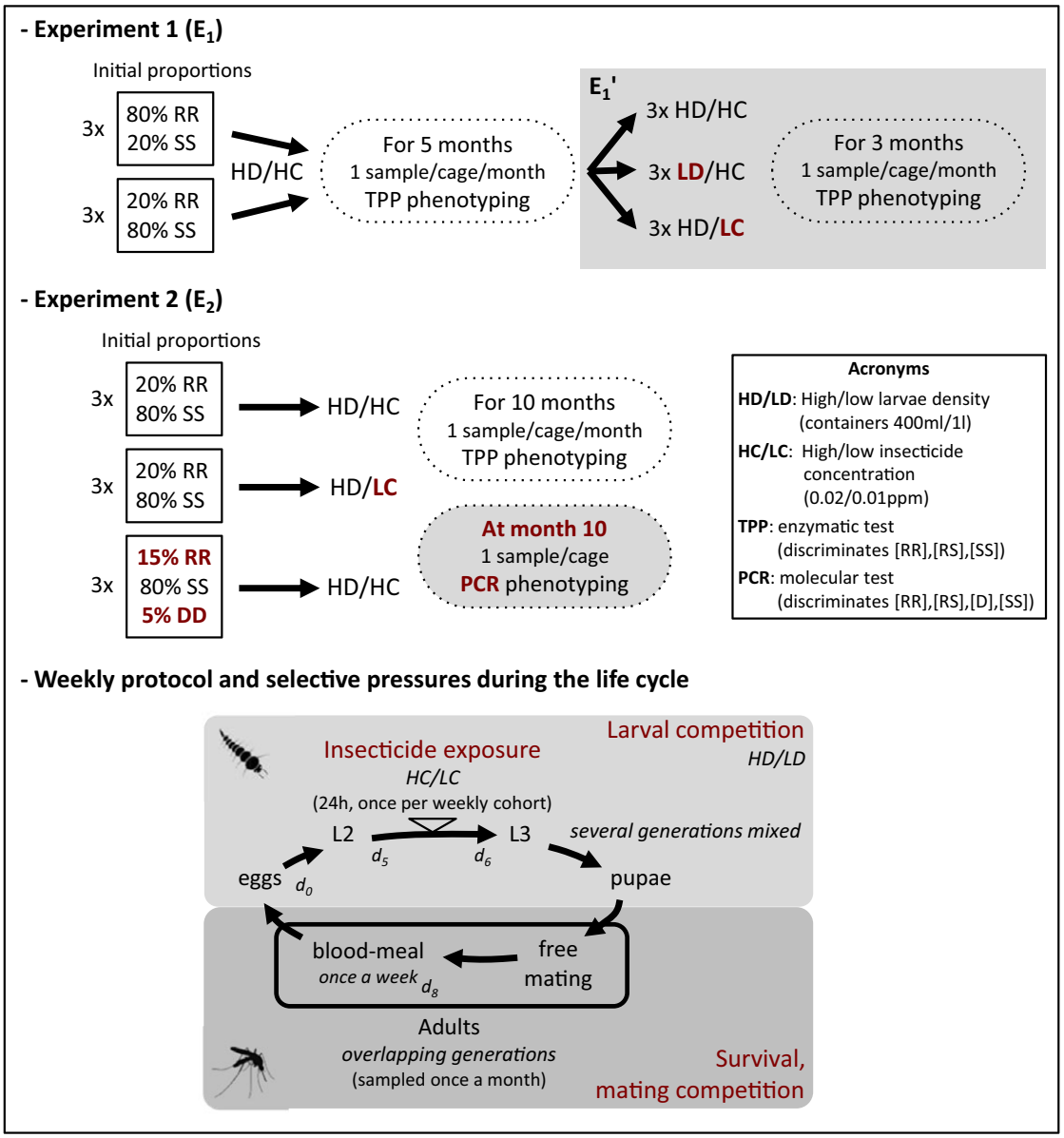

Figure 2. Experimental design.

The experimental design of the two experiments (E_1 and E_2) is presented (see text also): number of replicates, initial genotype proportions, experimental conditions, duration, sampling design, and phenotyping methods (the acronyms are defined in the box).

For experiment E_1, the replicates were split at month 4 (E_1′) and the experimental conditions were modified (reference = HD/HC) in the different subsets; the modifications are bolded in red. For experiment E_2, the conditions differing from the reference (HD/HC) and the alterations to the initial genotype proportions are bolded in red. A specific phenotyping method was used only for the 10th month (bolded in red).

The weekly protocol is also represented (see also text); the timing is indicated in days from the egg rafts' collection (d_i). The various selective pressures are indicated in red and their origins are italicized.

in weekly cycles, under standard conditions (25°C, >60% humidity, 12-h light:12-h dark). On the first day (d_0), the egg rafts were collected and placed together in a single container with ~400 ml of water, resulting in a high larval density (high density or HD, >1000 larvae/l). On day five (d_5), larvae (~L_2) were exposed for 24 h to 0.02 ppm of temephos (OP insecticide, Bayer®), the high insecticide concentration (or HC), estimated from bioassays (Labbé et al. 2014). On day six (d_6), the survivors were collected and placed together in a single container in the cage, to allow the adults to emerge; the adults mated freely. On day eight (d_8), the females were provided with a new blood meal. These weekly cycles generated overlapping generations (adults remained in the

cage until their death, two to six weeks later) and ensured that all individuals were exposed to the insecticide once in their lives. The control conditions were HD/HC. These conditions were modified by reducing larval density (low density or LD) during rearing, through the use of larger containers (~1 l), or by reducing the concentration of insecticide (low concentration or LC = 0.01 ppm). The experimental design is summarized in Figure 2. Each month, we sampled an average of 48 adults at random from each cage for the estimation of phenotypic frequencies (total number of individuals analyzed = 6547; Table A1).

PHENOTYPING

For each sample, individual phenotypes were established with the TPP test (Bourguet et al. 1996) (Table A1), based on the activity of the acetylcholinesterase AChE1 in the presence and absence of insecticide (propoxur, Baygon®). This test discriminates between three phenotypes: [SS], [RR], and [RS]. The first two are always unambiguous and correspond to the homozygous SS and RR genotypes, respectively. When only the R and S alleles are present, the last phenotype corresponds to the standard heterozygote RS (i.e., the phenotypic frequencies correspond to the genotypic frequencies). However, when the D allele is present, the [RS] phenotype becomes ambiguous, as it can result from four genotypes: RS, DD, DS, and DR.

In the last month of the experiment E_2 (Fig. 2), we distinguished standard heterozygotes (RS) from individuals carrying at least one D allele (DD, DS, and DR; these three genotypes cannot be distinguished), using a PCR test specific for the susceptible copy of the D allele (DucosEx3dir – DucosEx3rev; Labbé *et al.* 2014), directly on second-instar larvae (no DNA extraction was required, the larvae dissolve in the buffer during the first PCR 95°C denaturation step) (Table A2).

ESTIMATION OF ALLELE FREQUENCIES

When only the R and S alleles were present, the R allele frequency $f(R)$ was calculated directly from the phenotypic frequencies for each generation of each replicate.

However, this was not possible when the D allele was present (see, section "Phenotyping"). In this case, we calculated the apparent R frequency $f^*(R) = f([RR]) + 0.5f([RS])$, that is, as if only R and S were present (D carriers then appear as heterozygotes [RS]).

For the last month of experiment E_2 (Fig. 2), we inferred the D, R, and S allele frequencies from the phenotypic frequencies (Table A2), assuming panmixia and using the maximum-likelihood approach developed by Lenormand et al. (1998): for each replicate, we calculated the log-likelihood L of observing the phenotypic data as

$$L = \sum_i n_i ln(f_i) ,$$

with n_i and f_i the observed number and the predicted frequency of individuals with phenotype i, respectively. L was then maximized (L_{max}) with a simulated annealing algorithm (Lenormand et al. 1998). For each allele frequency, the support limits (SL) were then calculated as the minimum and maximum values that this frequency could take without significantly decreasing the likelihood (i.e., L_{max} −1.96, roughly equivalent to 95% confidence intervals).

ESTIMATION OF SELECTION COEFFICIENTS

We estimated the relative fitness of the various phenotypes, using a simple deterministic (i.e., no drift) population genetics model considering infinite populations and discrete generations (two per month). This model does not completely reflect our experiments. First, drift did play a role in the allele dynamics, but it had a much smaller impact than selection, which could lead to more dispersion than expected in the observed data; this was handled statistically by controlling for overdispersion in the likelihood model (see below). Second, the generations in the experiment were overlapping, so that the actual generation number is probably below two, which should, conservatively, lead to an underestimation of fitness differences. Hence, our approach was statistically robust to these simplifications.

We used a two-step model: (1) Reproduction: the frequency f_{gi} of each genotype g among the larvae in generation i was calculated from the allele frequencies (f_R and f_S for the R and S alleles, respectively) in the gametes of the previous generation ($i - 1$), assuming panmixia ($f_{RRi} = f_{R(i-1)}{}^2$, $f_{SSi} = f_{S(i-1)}{}^2$ and $f_{RSi} = 2f_{R(i-1)}f_{S(i-1)}$); (2) Selection was taken into account between the larval stage and the adult stage, to calculate the frequency f'_{gi} of each genotype g in the adults of generation i ($f'_{gi} = (f_{gi}w_g)/\Sigma(f_{gi}w_g)$, where w_g is the fitness of the genotype g). The fitness of heterozygous individuals, [RS], $w_{RS} = 1$ was used as a reference. The relative fitnesses of the SS and RR genotypes were set as $w_{SS} = 1 + s_{SS}$ and $w_{RR} = 1 + s_{RR}$, respectively, where s_{SS} and s_{RR} are the corresponding selection coefficients. The allelic frequencies in the gametes produced by the surviving adults of generation i were then calculated from the genotypic frequencies of these individuals ($f_{Ri} = f'_{RRi} + 0.5f'_{RSi}$, $f_{Si} = f'_{SSi} + 0.5f'_{RSi}$).

This model was adjusted to the data (phenotypic frequencies) through a maximum likelihood approach (R software version 2.15.1 https://www.r-project.org/ package optim, method L-BFGS-B). The SL associated with the selection coefficients s_{SS} and s_{RR} were calculated from the likelihood profile (L_{max} −1.96) established from 10^6 simulations. For each model, we calculated the percentage of the total deviance explained as $\%TD = (D_{max} - D_{mod}) / (D_{max} - D_{min})$ (where D_{min}, D_{max}, and D_{mod}, are the minimal, maximal, and model deviances, respectively, with $D = -2L$), and the overdispersion as $od = D_{res} / df_{res}$ (where D_{res} and df_{res} are the residual deviance and the residual degrees of freedom,

respectively). We assessed the significance of differences in selection coefficients between rearing conditions (HC/HD, HC/LD, and LC/HD), by adjusting a complete model with two parameters for each set of conditions considered (s_{SS} and s_{RR}). A simplified model, with only one s_{SS} or s_{RR} parameter for all conditions, was then computed. The two models were then compared using likelihood ratio tests corrected for overdispersion (LRT_{od}): when significant, the simplified model was rejected, and the tested selection parameter was considered to differ significantly between the conditions tested.

Results and Discussion

OVERDOMINANCE RESULTS FROM BALANCED ANTAGONIST SELECTIVE PRESSURES

We first investigated the possibility that intermediate insecticide doses could result in overdominance by setting up six population evolution experiments in cages containing a mixture of *C. pipiens* RR and SS genotypes, with three replicates having an initial R frequency $f_0(R) = 0.8$ and three replicates having an initial $f_0(R) = 0.2$ (experiment E_1, Fig. 2). Each new generation was exposed to a high concentration (HC) of insecticide that killed almost all SS individuals, a few RS but no RR individuals, thereby favoring selection of the R allele. The larvae were also reared at HD, to increase competition, resulting in higher levels of larval mortality. The emerging adults were released into the same cage: they mated freely and were able to reproduce every week until their death (overlapping generations), favoring the selection of individuals with longer lifespans and higher mating success, that is, those carrying the less costly S allele. We monitored changes in R allele frequency $f(R)$ every month (about two overlapping generations), by genotyping adults randomly sampled from each of the six cages (there are only RR, SS, and RS individuals; Fig. 2 and Table A1). If these conditions result in overdominance, it is possible to predict the dynamics of the alleles: (1) $f(R)$ should reach a stable equilibrium ($\hat{f}(R)$): even when RS is the fittest genotype, when two heterozygotes mate, only half of their progeny carries the RS genotype (i.e., there is a segregation burden), the other half carrying the less fit genotypes RR or SS; the polymorphism is thus stable; (2) the frequency at equilibrium should depend on the relative fitness of each genotype in the environmental conditions considered, but should be independent of the initial frequencies of the alleles (f_0).

This is what we observed, despite variations due to drift and sampling: the frequency of R in all replicates ultimately converged around $\hat{f}(R) = 0.68 \pm 0.08$ (mean $\pm$ standard deviation), regardless of its initial frequency (Fig. 3A1). Estimates based on a population genetics model confirmed that the relative fitness of the heterozygotes (w_{RS}) was significantly higher than those of the two homozygotes (confirming overdominance) and that selection for heterozygotes was strong (E_1, Table 1): RR individuals were fitter than SS individuals (as expected from $\hat{f}(R) > 0.5$), but this resistance advantage did not compensate for the fitness cost relative to RS mosquitoes.

We then confirmed that the frequencies at equilibrium were constrained by the evolutionary trade-offs of the various genotypes by altering the environmental conditions. After four months, we kept three of the cages in the original conditions (controls: HD, HC), but the other three cages were each split in two (experiments E_1', Table 1, Fig. 2). In one half of these three cages, we reduced larval rearing density (low density, LD), but maintained the HC of insecticide treatment (Fig. 2). These conditions were expected to reduce the competition between larvae, and thus, the cost of the R allele. As expected, $f(R)$ increased (Fig. 3A2, brown stars). Accordingly, the relative fitness ranking changed significantly (LRT_{od}, $F = 10.1$; $\Delta df = 2$; $P < 0.001$): RR became the fittest genotype (E_1', Table 1), but RS remained fitter than SS. In the other half of these three cages, we reduced the insecticide concentration (LC), but maintained the high larval density (HD) (Fig. 2). These conditions were expected to reduce the selective advantage of the R allele. As expected, $f(R)$ decreased (Fig. 3A2, blue squares). Due to the shorter duration, equilibriums were not reached: the selective coefficient estimates obtained are less precise, RS remained the fittest genotype, but SS became slightly fitter than RR (LRT_{od}, $F = 32.7$; $\Delta df = 2$; $P < 0.001$; E_1', Table 1). These experiments thus provided evidence that environmental modifications affecting the selective advantage or the cost can alter the frequency equilibrium.

We assessed the robustness of these conclusions by replicating the study described above. Three cages were set up as controls (HD, HC), with an initial $f_0(R) = 0.2$ (experiment E_2, Fig. 2): after 10 months, the mean R frequency had reached an equilibrium at $\hat{f}(R) = 0.73 \pm 0.04$ (Fig. 3B1, red triangles), similar to that in E_1 ($\hat{f}(R) = 0.68 \pm 0.08$; Student's t test, $P = 0.34$). This robustness (nine cages stabilizing at about the same frequency, regardless of their initial frequency) confirmed that the allele dynamics in our experiments were mostly driven by selection. In parallel, three other cages were set up with a low insecticide concentration (HD, LC), again with an initial $f_0(R) = 0.2$ (E_2, Fig. 2). The mean R frequency had also reached equilibrium by 10 months, but at a lower value, $\hat{f}(R) = 0.24 \pm 0.14$ (Fig. 3B1, blue squares). As expected, model estimates confirmed that RS individuals were the fittest in both conditions (Table 1). However, fitness ranking changed according to the different conditions (LRT_{od}, $F = 172.7$; $\Delta df = 2$; $P < 0.001$; Table 1): in control (HD, HC) conditions, RR individuals were fitter than SS individuals ($w_{RS} > w_{RR} > w_{SS}$), whereas SS individuals were fitter than RR individuals in LC conditions ($w_{RS} > w_{SS} > w_{RR}$), indicating that RR resistance relative advantage cannot compensate its cost at lower insecticide concentrations. E_2 thus confirmed the conclusions of E_1:

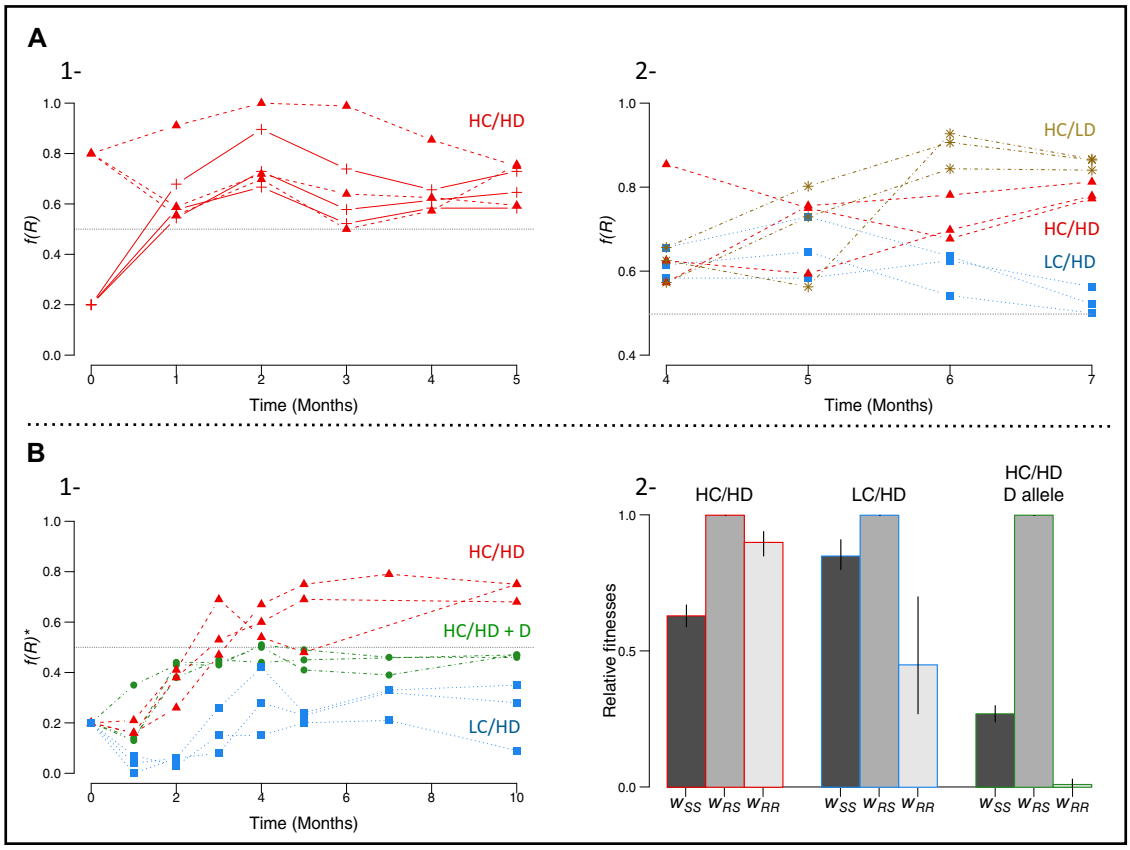

Figure 3. R allele dynamics in the various experimental evolution studies and fitness estimations
(A) Changes in R allele frequency *f(R)* in E_1 (Table 1). (A1) Original rearing conditions at high density (HD) and high insecticide concentration (HC), R and S alleles only, with $f_0(R) = 0.8$ (triangles, dashed lines) or $f_0(R) = 0.2$ (plus signs, solid lines). (A2) Altered rearing conditions after four months (E_1', Table 1): controls (HD/HC; red triangles, dashed lines), reduced density (LD/HC; brown stars, dotted-dashed lines), reduced insecticide concentration (HD/LC; blue squares, dotted lines).
(B1) Changes in *f(R)* in E_2 (Table 1): controls (HD/HC, R and S; red triangles, dashed lines), reduced insecticide concentration (HD/LC, R and S; blue squares, dotted lines), duplicated allele assay, that is, control rearing conditions (HD/HC), R, S, and D alleles, with $f_0(D) = 0.05$, $f_0(R) = 0.15$, and $f_0(S) = 0.8$ (*NB: in this set up, as all genotypes carrying D confer a heterozygote phenotype, *f(R)* is estimated as if only R and S were present, see text). (b2) The relative fitnesses of the [SS] (w_{SS}, dark gray bars), [RS] ($w_{RS} = 1$, medium gray bars), and [RR] (w_{RR}, light gray bars) phenotypes (see "Estimation of selection coefficients" in section Methods) are presented for the different conditions of the E_2 experiment (B1 plot colors are conserved).

different overdominance equilibria may exist, depending on selective pressure intensities, which can alter the relative fitness trade-offs of the different genotypes (providing a clear alternative example of overdominance to sickle cell anemia; Lewontin 1974).

OVERDOMINANCE FAVORS THE HETEROGENEOUS DUPLICATED ALLELE

We then tested whether the heterogeneous *ace-1* duplication (D) was favored when the environmental conditions result in overdominance. Previous studies suggested that D alleles could confer fitness trade-offs on their carriers similar to those of standard RS heterozygotes (Labbé et al. 2014; Assogba et al. 2015). Moreover, DD individuals should not suffer from the segregation burden associated with the RS genotype. We set up three replicates in control conditions (HC, HD), but introduced a D allele, such that $f_0(D) = 0.05$, $f_0(R) = 0.15$, $f_0(S) = 0.80$ (E_2, Fig. 2). If D is indeed favored, then most individuals should be phenotypic heterozygotes [RS], with an apparent R frequency at equilibrium $\hat{f}(R)^*$ close to 0.5 (see, section Methods).

This is precisely what we observed: after 10 months, most of the individuals were indeed [RS] and no [RR] individuals were observed ($\hat{f}(RS) = 0.93 \pm 0.02$), with $\hat{f}(R)^* = 0.47 \pm 0.01$ (Fig. 3B1, green circles). The estimated relative fitnesses of [SS] and [RR] were both close to 0 (E_2, Table 1 and Fig. 3B2). The persistence of a few [SS] individuals while all [RR] disappeared suggests a slightly asymmetrical trade-off: as phenotyping preceded selection, it could result from a higher fitness of DS than DR individuals, thereby still generating a few new [SS]

Table 1. Experimental design and relative fitness estimations.

	Allele	Conditions	W_{SS} (SL)	W_{RR} (SL)	%TD	od
E_1	*S / R*	*HC/HD*	***0.20 (0.16–0.24)***	***0.58 (0.54–0.62)***	*0.66*	*5.55*
E_1'	*S / R*	*HC/HD*	***0.10 (0.04–0.21)***	***0.69 (0.62–0.76)***	*0.85*	*2.8*
	S / R	HC/**LD**	**0.59 (0.35–0.90)**	**1.22 (1.09–1.39)**	0.89	2.73
	S / R	**LC**/HD	0.91 (0.68–1.04)	**0.78 (0.67–0.89)**	0.84	2.16
E_2	*S /R*	*HC/HD*	***0.63 (0.59–0.67)***	***0.90 (0.85–0.94)***	*0.76*	*3.04*
	S / R	**LC**/HD	**0.85 (0.80–0.91)**	**0.45 (0.27–0.70)**	0.73	5.41
	S / R / **D**	HC/HD	**0.27 (0.24–0.30)**	**0.01 (0.00–0.03)**	0.89	3.54

For the various evolution experiments (E_i), the alleles in competition are indicated (single-copy susceptible S, single-copy resistant R, and heterogeneous duplication D). The conditions in which the larvae were reared are also indicated: high or low insecticide concentration (HC = 0.02 ppm and LC = 0.01 ppm temephos, respectively) and high or low larval density (HD or LD, respectively); controls are indicated in italics; conditions differing from the controls are shown in bold. For each set of conditions, a population genetics model was used to calculate the fitnesses of the single-copy susceptible (WSS) and resistance (WRR) homozygotes relative to that of the heterozygote (WRS = 1). When the associated support limits (SL, in brackets) include 1, w_{SS} and/or w_{RR} are not significantly different from w_{RS}; significant differences are shown in bold. The percentage of the total deviance explained by each model (*%TD*) and its overdispersion (*od*) are also indicated.

individuals. D invasion was further confirmed with a specific molecular test applied to about 90 individuals from each replicate: the frequency of this allele increased from 0.05 to an estimated $\hat{f}(D) = 0.72 \pm 0.07$ (Table A2). All the [RS] individuals carried D (i.e., no standard RS heterozygotes were found in the cages); the [RS] phenotype fitness therefore corresponded to genotypes DS, DR, or DD. Heterogeneous duplications, by conferring the heterozygous phenotype without the associated segregation burden, can thus be fixed when selective trade-off favors overdominance, that is, when antagonist selective pressures are balanced.

OVERDOMINANCE IS PROBABLY COMMON BUT TRANSITORY

Overdominance is almost certainly more widespread than generally thought (Hedrick 2012; Llaurens et al. 2017): (i) multivariate stabilizing selection probably frequently leads to overdominance for new mutations that are beneficial in the heterozygous state (Manna et al. 2011), and (ii) recent adaptation usually involves trade-offs (Orr 2005), so balanced antagonist selective pressures would result in overdominance; some studies measuring the fitness associated to new mutations in the laboratory have confirmed these expectations (Peters et al. 2003). Our study contributes to explain the discrepancy between the prevalence of overdominance in newly arising mutations and the relative rarity of segregating overdominant alleles in the field (Manna et al. 2011): it shows that overdominance may not be robust, as limited modifications of the environment can favor one allele over the other; more importantly, we showed that, as predicted by Haldane (1954), overdominance can be rapidly abolished by the occurrence and selective spread of a heterogeneous duplication, a situation that has probably contributed to the scarcity of persistent cases of overdominance in natural populations.

HETEROGENEOUS DUPLICATIONS CAN BE IMMEDIATELY ADAPTIVE AND COULD FUEL FUTURE EVOLUTION

Overdominance (and more generally heterozygote advantage) has been proposed to explain the diversification of several multigenic families, such as MHC genes in vertebrates (Spurgin and Richardson 2010), R genes in plants (Michelmore and Meyers 1998; Panchy et al. 2016), and MAT genes in basidiomycetes (May et al. 1999): heterozygotes display a higher fitness because they can resist to more pathogens (MHC and R genes) or mate with more sexual types (MAT). However, these duplications are ancient and it is difficult to determine whether copy sequence polymorphism existed before the duplications (i.e., heterogeneous duplications associating existing alleles) or resulted from postduplication divergence (i.e., originally homogeneous duplications; Fig. 1). The difficulty to identify ancient heterogeneous duplications probably explains why their potentially crucial adaptive role has been overlooked.

Fortunately, a handful of contemporaneous heterogeneous duplications have been described that allow assessing how and why they are selected. Interestingly, they all concern insecticide target genes, probably because these duplications are recent (occurring in response to a few decades of anthropic environmental modification), associated with irreducible trade-offs, and highly scrutinized due to their impact on vector control and public health (Labbé et al. 2007a; Djogbénou et al. 2008; Remnant et al. 2013; Sonoda et al. 2014). The *ace-1* duplications in *C. pipiens* remain however the most deeply studied: so far, at least 13 duplicated alleles have been identified, most of them resulting from independent duplication events (Labbé et al. 2007a; Alout et al. 2010; Osta et al. 2012). This suggests that the conditions for overdominance are probably quite frequent: insecticide treatment practices typically

result in a patchy environment, with alternating treated and non-treated areas: if the grain of the environment is smaller than the dispersal distance of the mosquito, it could result in marginal overdominance (Lenormand et al. 1998; Labbé et al. 2007b), or even full overdominance if field conditions result in low insecticide doses (this study). Moreover, the treatment intensities typically vary in time, as they are applied usually only in some periods of the year (Lenormand et al. 1999): these fluctuations of antagonist selective pressures could also on average favor the heterozygote phenotype (i.e., marginal overdominance resulting from fluctuating selection). This frequent selection of D alleles could seriously hinder mosquito control: because they display a lower cost than R, these resistance alleles could make OP and CX insecticides virtually obsolete and threaten control strategies based on insecticide alternation. This is particularly pressing in the case of *A. gambiae*, the malaria vector, where these insecticides have been suggested as replacements for the widely used pyrethroids that face high and widespread resistance (Assogba et al. 2015, 2016).

This unique example shows that heterogeneous duplications can result in an immediate qualitative advantage in fluctuating or patchy environments, or more generally in environments with balanced antagonistic selective pressures. As predicted by Haldane in 1954, we indeed demonstrated that these duplications can be selected because they allow the permanent association of overdominant heterozygous alleles, with no segregation burden (Haldane 1954). These properties could prove useful for breeders (plant breeders in particular), as new genome-editing tools (e.g., CRISPR-cas9; Sander and Joung 2014) could be used to generate heterogeneous duplications to create stable lines displaying specific heterosis otherwise found only in hybrids (Fu et al. 2014).

However, heterogeneous duplications should also be studied in more detail in terms of their role in long-term evolution, as they probably bear witness to ancestral situations of transitory overdominance. As homogeneous duplications, they create new material for genetic innovation (Lynch and Force 2000; Osada and Innan 2008; Innan and Kondrashov 2010). However, as they result from the association of two already divergent alleles, their dissimilar copies immediately carry different functions, that is, copy functional divergence precedes fixation (Fig. 1). They are thus more likely to evolve further through subfunctionalization and neofunctionalization, and should do it more rapidly than homogeneous duplications; these can be first selected, for example, for increased protein quantity, but would diverge later (Lenormand et al. 1998; Labbé et al. 2007a; Innan and Kondrashov 2010). These heterogeneous duplications could thus play a major, albeit yet underestimated, role in the evolution of gene families.

Author Contributions

PL, PM, MW, and TL conceived and designed the study; PL, PM, and MW performed the experiments; PL, PM, MW, and TL analyzed and interpreted the data; PL and PM drafted the manuscript and all authors contributed to later versions of the manuscript.

ACKNOWLEDGMENTS

We would like to thank N. Pasteur, H. Alout, M. Sicard, P. H. Gouyon and M. Raymond for helpful comments on the manuscript, and P. Makoundou and S. Unal for technical support. This work was funded by French ANR program (project "SilentAdapt", BIOADAPT 2013–2015, Grant number ANR-13-ADAP-0016). This study is the contribution 2017–150 of the *Institut des Sciences de l'Evolution de Montpellier* (UMR 5554, CNRS-UM-IRD-EPHE).

DATA ARCHIVING

DATA ACCESSIBILITY

All the data reported in this article are in the Tables A1 and 2.

LITERATURE CITED

Alout, H., L. Djogbénou, C. Berticat, F. Chandre, and M. Weill. 2008. Comparison of *Anopheles gambiae* and *Culex pipiens* acetycholinesterase 1 biochemical properties. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 150:271–277.

Alout, H., P. Labbé, N. Pasteur, and M. Weill. 2010. High incidence of *ace-1* duplicated haplotypes in resistant *Culex pipiens* mosquitoes from Algeria. Insect Biochem. Mol. Biol. 41:29–35.

Assogba, B. S., L. S. Djogbénou, P. Milesi, A. Berthomieu, J. Perez, D. Ayala, F. Chandre, M. Makoutodé, P. Labbé, and M. Weill. 2015. An *ace-1* gene duplication resorbs the fitness cost associated with resistance in *Anopheles gambiae*, the main malaria mosquito. Sci. Rep. 5:1–12.

Assogba, B. S., P. Milesi, L. S. Djogbénou, A. Berthomieu, P. Makoundou, L. S. Baba-Moussa, A.-S. Fiston-Lavier, K. Belkhir, P. Labbé, and M. Weill. 2016. The *ace-1* locus is amplified in all resistant *Anopheles gambiae* mosquitoes: fitness consequences of homogeneous and heterogeneous duplications. PLoS Biol. 14:e2000618.

Berticat, C., G. Boquien, M. Raymond, and C. Chevillon. 2002. Insecticide resistance genes induce a mating competition cost in *Culex pipiens* mosquitoes. Genet. Res. 79:41–47.

Bourguet, D., N. Pasteur, J. A. Bisset, and M. Raymond. 1996. Determination of *Ace.1* genotypes in single mosquitoes: toward an ecumenical biochemical test. Pestic. Biochem. Physiol. 55:122–128.

Djogbénou, L., F. Chandre, A. Berthomieu, R. K. Dabiré, A. Koffi, H. Alout, and M. Weill. 2008. Evidence of introgression of the *ace-1R* mutation and of the *ace-1* duplication in west African *Anopheles gambiae s. s*. PLoS One 3:e2172:1–7.

Dobzhansky, T. 1952. Nature and origin of heterosis. Pp. 218–223 in J. W. Gowan, ed. Heterosis. Iowa State College Press, Ames, IA.

Duron, O., P. Labbé, C. Berticat, F. Rousset, S. Guillot, M. Raymond, and M. Weill. 2006. High *Wolbachia* density correlates with cost of infection for insecticide resistant *Culex pipiens* mosquitoes. Evolution 60:303–314.

East, E. M. 1936. Heterosis. Genetics 21:375–397.

Fisher, R. A. 1922. On the dominance ratio. Proc. R. Soc. London B 42:321–341.

Fu, D., M. Xiao, A. Hayward, Y. Fu, G. Liu, G. Jiang, and H. Zhang. 2014. Utilization of crop heterosis: a review. Euphytica 197:161–173.

Georghiou, G. P., R. L. Metcalf, and F. E. Gidden. 1966. Carbamate-resistance in mosquitoes: selection of *Culex pipiens fatigans* Wied (= *Culex quinquefasciatus*) for resistance to Baygon. Bull. World Heal. Organ. 35:691–708.

Haldane, J. B. S. 1954. The biochemistry of genetics. George Allen and Unwin, Ltd., Lond.

Hedrick, P. W. 2012. What is the evidence for heterozygote advantage selection? Trends Ecol. Evol. 27:698–704.

Innan, H., and F. Kondrashov. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11:97–108.

Katju, V., and U. Bergthorsson. 2013. Copy-number changes in evolution: rates, fitness effects and adaptive significance. Front. Genet. 4:1–12.

Labbé, P., A. Berthomieu, C. Berticat, H. Alout, M. Raymond, T. Lenormand, and M. Weill. 2007a. Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito *Culex pipiens*. Mol. Biol. Evol. 24:1056–1067.

Labbé, P., C. Berticat, A. Berthomieu, S. Unal, C. Bernard, M. Weill, and T. Lenormand. 2007b. Forty years of erratic insecticide resistance evolution in the mosquito *Culex pipiens*. PLoS Genet. 3:e205.

Labbé, P., P. Milesi, A. Yébakima, N. Pasteur, M. Weill, and T. Lenormand. 2014. Gene-dosage effects on fitness in recent adaptive duplications: *ace-1* in the mosquito *Culex pipiens*. Evolution 68:2092–2101.

Lenormand, T., T. Guillemaud, D. Bourguet, and M. Raymond. 1998. Appearance and sweep of a gene duplication: adaptive response and potential for new functions in the mosquito *Culex pipiens*. Evolution 52:1705–1712.

Lenormand, T., D. Bourguet, T. Guillemaud, and M. Raymond. 1999. Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. Nature 400:861–864.

Lewontin, R. C. 1974. The genetic basis of evolutionary change. Columbia Univ. Press, New York.

Llaurens, V., A. Whibley, and M. Joron. 2017. Genetic architecture and balancing selection: the life and death of differentiated variants. Mol. Ecol. 26:2430–2448.

Lynch, M., and A. Force. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics, 154, 459–473.

Manna, F., G. Martin, and T. Lenormand. 2011. Fitness landscapes: an alternative theory for the dominance of mutation. Genetics 189:923–937.

May, G., F. Shaw, H. Badrane, and X. Vekemans. 1999. The signature of balancing selection: fungal mating compatibility gene evolution. Proc. Natl. Acad. Sci. U S A 96:9172–9177.

Michelmore, R. W., and B. C. Meyers. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res. 8:1113–1130.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Osada, N., and H. Innan. 2008. Duplication and gene conversion in the *Drosophila melanogaster* genome. PLoS Genet. 4:e1000305.

Osta, M., Z. Rizk, P. Labbé, M. Weill, and K. Knio. 2012. Insecticide resistance to organophosphates in *Culex pipiens* complex from Lebanon. Parasit. Vectors 5:1–6.

Panchy, N., M. Lehti-shiu, and S. Shiu. 2016. Evolution of gene duplication in plants. Plant Physiol. 171:2294–2316.

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

Remnant, E. J., R. T. Good, J. M. Schmidt, C. Lumb, C. Robin, P. J. Daborn, and P. Batterham. 2013. Gene duplication in the major insecticide target site, *Rdl*, in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 110:14705–14710.

Sander, J. D., and J. K. Joung. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. 32:347–355.

Schrider, D. R., and M. W. Hahn. 2010. Gene copy-number polymorphism in nature. Proc. Biol. Sci. 277:3213–3221.

Sonoda, S., X. Shi, D. Song, P. Liang, X. Gao, Y. Zhang, J. Li, Y. Liu, M. Li, M. Matsumura, et al. 2014. Duplication of acetylcholinesterase gene in diamondback moth strains with different sensitivities to acephate. Insect Biochem. Mol. Biol. 48:83–90.

Spurgin, L. G., and D. S. Richardson. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc. R. Soc. B Biol. Sci. 277:979–988.

Weill, M., G. Lutfalla, K. Mogensen, F. Chandre, A. Berthomieu, C. Berticat, N. Pasteur, A. Philips, P. Fort, and M. Raymond. 2003. Insecticide resistance in mosquito vectors. Nature 423:423–426.

Associate Editor:
Handling Editor:

Appendix

Table A1. Population cages phenotypic data.

			Replicate 1				Replicate 2				Replicate 3			
	Alleles and conditions	Month	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot
E_1	S / R HC / HD	1	0	41	4	45	5	17	20	42	4	30	11	45
		3	1	24	23	48	0	10	38	48	1	30	17	48
		3	5	28	12	45	1	21	22	44	6	31	8	45
		4	5	27	16	48	2	29	17	48	3	34	11	48
		5	6	22	20	48	0	26	22	48	4	32	12	48
			Replicate 4				Replicate 5				Replicate 6			
		Month	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot
		1	1	38	6	45	3	31	11	45	0	8	37	45
		3	4	21	23	48	3	21	24	48	0	0	48	48
		3	3	19	3	25	3	25	15	43	0	1	43	44
		4	9	23	16	48	0	36	12	48	1	12	35	48
		5	1	19	23	43	0	39	9	48	2	12	18	32

			Replicate 1				Replicate 2				Replicate 3			
	Alleles and conditions	Month	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot
E_1'	S / R HC / HD	4	9	23	16	48	1	12	35	48	0	36	12	48
		5	1	19	23	43	2	12	18	32	0	39	9	48
		6	0	21	27	48	1	29	18	48	1	27	20	48
		7	0	18	30	48	0	5	6	11	0	19	24	43
	S / R HC / LD	4	9	23	16	48	2	29	17	48	0	36	12	48
		5	5	16	27	48	0	19	29	48	8	26	14	48
		6	1	13	34	48	0	9	39	48	0	7	41	48
		7	0	15	32	47	0	13	35	48	0	12	33	45
	S / R LC / HD	4	5	27	16	48	2	29	17	48	3	34	11	48
		5	6	22	20	48	0	26	22	48	4	32	12	48
		6	4	14	6	24	8	19	21	48	6	24	18	48
		7	13	22	13	48	8	28	10	46	8	26	14	48

			Replicate 1				Replicate 2				Replicate 3			
	Alleles and conditions	Month	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot
E_2	S / R HC / HD	1	39	19	0	58	39	19	0	58	28	20	0	48
		2	12	36	0	48	26	19	3	48	11	35	2	48
		3	6	33	9	48	11	29	8	48	3	24	21	48
		4	1	36	11	48	1	26	16	43	4	36	8	48
		5	3	24	21	48	1	22	25	48	5	11	4	20
		6	–	–	–	–	–	-	–	–	–	–	–	–
		7	–	–	–	–	0	14	20	34	–	–	–	–
		8	–	–	–	–	–	–	–	–	–	–	–	–
		9	–	–	–	–	–	–	–	–	–	–	–	–
		10	2	57	37	96	2	44	50	96	0	48	48	96

(*Continued*)

Table A1. Continued.

		Replicate 1				Replicate 2				Replicate 3			
Alleles and conditions	Month	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot	[SS]	[RS]	[RR]	Tot
S / R HC / LD	1	48	2	1	51	47	0	0	47	41	7	0	48
	2	42	6	0	48	42	6	0	48	45	3	0	48
	3	27	17	4	48	40	8	0	48	35	12	1	48
	4	16	24	8	48	21	27	0	48	34	14	0	48
	5	25	23	0	48	15	13	0	28	28	19	0	47
	6	–	–	–	–	–	–	–	–	–	–	–	–
	7	9	14	1	24	17	31	0	48	27	19	0	46
	8	–	–	–	–	–	–	–	–	–	–	–	–
	9	–	–	–	–	–	–	–	–	–	–	–	–
	10	30	64	2	96	42	52	0	96	79	17	0	96
S / R / D* HD / HC	1	35	13	0	48	15	34	0	49	36	12	0	48
	2	12	36	0	48	7	41	0	48	5	26	1	32
	3	5	43	0	48	7	41	0	48	6	42	0	48
	4	3	23	0	26	0	57	1	58	1	46	1	48
	5	6	40	1	47	1	47	0	48	9	39	0	48
	6	–	–	–	–	–	–	–	–	–	–	–	–
	7	–	–	–	—-	4	44	0	48	11	37	0	48
	8	–	–	–	–	–	–	–	–	–	–	–	–
	9	–	–	–	–	–	–	–	–	–	–	–	–
	10	9	130	0	139	12	128	0	139	7	129	0	136

The number of individuals for each phenotype ([RR], [RS], [SS], TPP test) is indicated for each replicate and each month. The total number of individuals tested (Tot) was usually 48, but varied due to technical issues.
*Due to the presence of the D allele, the [RS] phenotype corresponds to RS, DS, DR, or DD genotypes in these replicates.

Table A2. Phenotypic data used to estimate D frequency at month 10 in E_2.

Phenotype*	Replicate 1	Replicate 2	Replicate 3
[RR]	0	0	0
[RS]	0	0	0
[D]	83	80	86
[SS]	7	11	4
Tot	90	91	90

The number of individuals for each phenotype ([RR], [RS], [D], [SS]) and the total number of tested individuals (Tot) are indicated for each replicate.
*All phenotypes correspond to only one genotype, except [D] that corresponds to DS, DR, or DD genotypes.