Genet. Res., Camb. (2009), **91**, *pp.* 171–182. © Cambridge University Press 2009
doi:10.1017/S0016672309000147 Printed in the United Kingdom

Molecular signature of epistatic selection: interrogating genetic interactions in the *sex-ratio* meiotic drive of *Drosophila simulans*

LUIS-MIGUEL CHEVIN[1,2]*†, HÉLOÏSE BASTIDE[3], CATHERINE MONTCHAMP-MOREAU[3] AND FRÉDÉRIC HOSPITAL[4]
[1] *UMR de Génétique Végétale, Ferme du Moulon, 91190 Gif Sur Yvette, France*
[2] *Ecologie, Systématique et Evolution, UMR 8079, Université Paris-Sud, 91405 Orsay Cedex, France*
[3] *Laboratoire Evolution Génome et Spéciation, UPR9034, CNRS, 91198 Gif-sur-Yvette Cedex, France*
[4] *INRA, UMR 1313 "Génétique Animale et Biologie Intégrative", 78352 Jouy-en-Josas, France*

(*Received 25 July 2008 and in revised form 19 March 2009*)

Summary

Fine scale analyses of signatures of selection allow assessing quantitative aspects of a species' evolutionary genetic history, such as the strength of selection on genes. When several selected loci lie in the same genomic region, their epistatic interactions may also be investigated. Here, we study how the neutral polymorphism pattern was shaped by two close recombining loci that cause 'sex-ratio' meiotic drive in *Drosophila simulans*, as an example of strong selection with potentially strong epistasis. We compare the polymorphism data observed in a natural population with the results of forward stochastic simulations under several contexts of epistasis between the candidate loci for the drive. We compute the likelihood of different possible scenarios, in order to determine which configuration is most consistent with the data. Our results highlight that fine scale analyses of well-chosen candidate genomic regions provide information-rich data that can be used to investigate the genotype–phenotype–fitness map, which can hardly be studied in genome-wide analyses. We also emphasize that initial conditions and time of observation (here, time after the interruption of a partial selective sweep) are crucial parameters in the interpretation of real data, while these are often overlooked in theoretical studies.

1. Introduction

Understanding how selection operates at the gene level is one of the main goals of evolutionary genetics. Most of the current effort to identify positively selected genes involves searching for molecular signatures of selection on neutral polymorphism (Nielsen, 2005). Indeed, the growth experienced by a beneficial mutation partly affects patterns of polymorphism at linked neutral variants through genetic hitchhiking (Maynard-Smith & Haigh, 1974), so neutral loci linked to a locus under selection may be distinguished from loci that evolve under pure neutrality. A popular approach in this context consists of analysing the polymorphism pattern around a candidate region that was previously identified through quantitative trait locus (QTL) analysis or association mapping. In contrast to large-scale genome scans (Nielsen *et al.*, 2005; Williamson *et al.*, 2007), which provide a global picture of natural selection, fine-scale studies of this kind allow asking detailed questions about how selection affected peculiar regions, up to the order of the Mb. For instance, Kim & Stephan (2002) designed a method to jointly estimate the precise location of the target of selection and the selection coefficient involved, thus yielding more quantitative information than the simple presence of positive selection in a genomic region.

Another appealing possibility would be to use the pattern of polymorphism in a candidate region to investigate the relationship between the genotype, the phenotype and fitness. This step is crucial in order to get an integrated view of evolution and adaptation, since selection only operates at the level of phenotypes, not directly on genes. However, in most cases, the genotype–phenotype–fitness map is extremely complex, and cannot be modelled explicitly without

* Corresponding author: e-mail: l.chevin@imperial.ac.uk
† Present address: Division of biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL57PY, UK.

huge simplifying assumptions (Gavrilets, 2004, chapter 2). And even then, its underlying parameters are often difficult to estimate empirically.

The best examples of integrated investigations of selection, from the phenotype *in natura* to the molecular level, mainly focused on QTL with very strong additive effects (Rogers & Bernatchez, 2005; Hoekstra *et al*., 2006). In such studies, the complexity of the traits compels to use a reductive approach that neglects the interactions that may occur (i) between the focal QTL and other genes that contribute to the trait and (ii) between the focal trait and other traits that contribute to fitness. If there were phenotypic traits whose relationship to fitness was clearly characterized, and the interactions between loci were simple and biologically explicit, then we could address specific questions about the functional interactions of genes under selection using molecular signatures of selection.

Selfish genetic elements are very appealing candidates in that respect. They take profit of the genomic machinery in order to increase their own reproductive success, largely independently (and often at the expense) of the fitness of the host organism (Hurst & Werren, 2001). Hence, their prevailing phenotype is directly their own fitness (besides possible pleiotropic effects on fertility or viability). They thus allow emitting simple, biologically explicit and empirically testable hypotheses about the genotype–phenotype–fitness map. These hypotheses can in turn be tested by various methods, including molecular signatures of selection.

Segregation distorters (Lyttle, 1991) are among the best-studied examples of selfish genetic elements. They hijack the process of meiosis (meiotic drive) or gametogenesis such as to be found in more than half of the gametes produced by heterozygous individuals that carry them, thus violating Mendel's law of random segregation. This confers them a strong selective advantage, and hence they can affect neutral polymorphism through the hitchhiking effect (Chevin & Hospital, 2006). The molecular mechanisms underlying the drive are usually unknown but likely many. In males, the known driving elements kill or disable the alternative gamete (Lyttle, 1991). In females, they take advantage of the asymmetry of female meiosis to end up into the egg nucleus. When they act on sex chromosomes in the heterogametic sex, they also modify the sex ratio of the population, in which case they are sometimes called *sex-ratio* distorters (Jaenike, 2001).

Here, we study the *sex-ratio* drive in *Drosophila simulans*, which has been well characterized genetically. This meiotic drive favours distorter X chromosomes (X^{SR}) against susceptible Y chromosomes in males. At least three independent *sex-ratio* systems have been found in this species (Tao *et al*., 2007; Jaenike, 2008). In the most thoroughly analysed case (denoted the 'Paris' *sex-ratio* in TAO *et al*. (2007)), the X^{SR} chromosomes have reached high prevalence in southeast Africa and Madagascar (frequency up to 60%), but their effect is now completely suppressed by autosomal and Y-linked suppressors (Atlan *et al*., 1997). Montchamp-Moreau *et al*. (2006) investigated the genetic determinism of the 'Paris' drive. Using a reference X^{SR} chromosome in a suppressor-free genetic background, they showed that two close genomic regions were both necessary for the drive to occur in the lab, which points towards obligate interaction between the alleles involved in the drive. However, we do not know whether their interaction in the genetic background of the natural populations at the time of their spread was the same as in the drive sensitive genetic context used in the lab. This question can be investigated using molecular signatures of selection.

The polymorphism pattern in the driving region of X^{SR} chromosomes of *D. simulans* was investigated in two recent studies. Derome *et al*. (2004) first showed that the *Nrg* gene located close to the meiotic drive elements of *D. simulans* exhibits the signature of a selective sweep in the islands of Madagascar and La Réunion. A further study of the sample from Madagascar, using several intragenic markers in the same genomic region, allowed uncovering a spatial pattern consistent with incomplete selective sweeps at two loci (Derome *et al*., 2008). This pattern is reproduced in Fig. 1, including three new markers that were not in Derome *et al*. (2008). It has three notable features. First, along each of the two causative regions previously identified in the genetic study of Montchamp-Moreau *et al*. (2006), the diversity of X^{SR} chromosomes is dramatically reduced relative to that of standard (non-distorter) chromosomes (X^{ST}). This is consistent with a strong association of those two regions with the phenotype under study (meiotic drive). Together with the high frequency of the haplotypes associated with the drive, this is also indicative of positive selection (Sabeti *et al*., 2002; Voight *et al*., 2006). Second, the linkage disequilibrium (LD) within each of the candidate regions, and also most notably between them, is very strong as can be seen from both the D' values and the significance of the Fisher exact test. This feature may have emerged as a result of positive epistasis between the drive loci included in each region. Third, in spite of this (putatively strong epistatic) selection involving two loci only 1 cM apart, the diversity at markers located in between the two regions is high. To understand how selection has shaped this pattern, we need a model with two close loci under selection and varying levels of epistasis.

The effect of positive selection at two closely linked loci on neutral polymorphism has been investigated

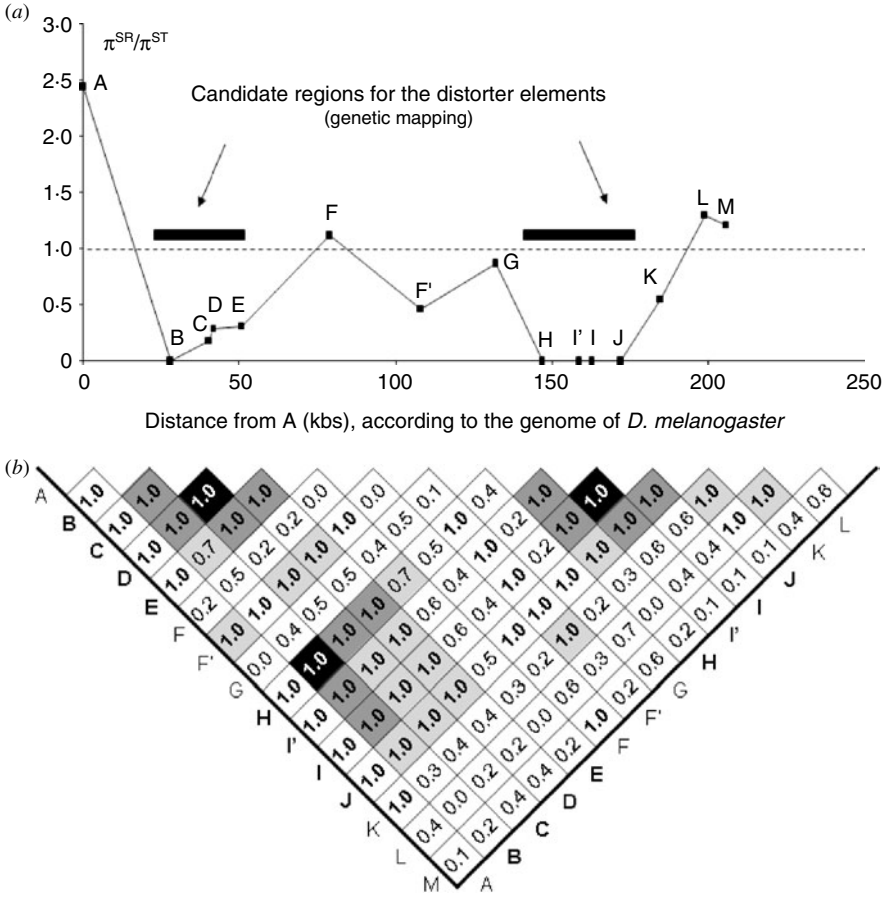

Fig. 1. Polymorphism pattern in the SR region of *D. simulans* in Madagascar. (*a*) Ratio of nucleotide diversities between distorter (SR) and standard (ST) X chromosomes. (*b*) LD between the markers (in letters), quantified by values of D' (in bold when equal to 1·0) and P values of Fisher's exact test (white: non-significant, light grey: $P<0{\cdot}05$, dark grey: $P<0{\cdot}01$, black: $P<0{\cdot}001$). Note the high linkage within and between the two candidate regions (letters in bold).

in several recent studies. Kim & Stephan (2003) showed that selective sweeps at two closely linked loci have on average less than additive effects on the reduction of heterozygosity at a neutral locus, because the selective interference between the selected loci slow down their dynamics. Chevin *et al.* (2008) further showed that the fine-scale polymorphism pattern around two partially linked loci with independent (multiplicative) effects on fitness can exhibit a spike in diversity in the interval delimited by the selected loci. This occurs because each beneficial mutation can hitchhike a different neutral allele. However, the pattern may be different if there is epistasis between the loci under selection. In the *sex-ratio* meiotic drive in *D. simulans*, epistasis is suggested by both the genetic analysis and the strong LD. This genetic system is thus a good example to study the signature of selection at two close loci in the presence of genetic interactions. Because meiotic drive systems are expected to evolve by recruiting interacting elements at linked loci (Crow, 1991; Palopoli & Wu, 1996), the case is particularly appealing.

In this paper, we investigate the effect of selective interactions on the pattern of neutral polymorphism, in the context of the *sex-ratio* meiotic drive of *D. simulans*. We first build a simple model of the interaction between the loci that cause the meiotic drive. This model is used to understand how epistasis between the driving loci influences their LD. Then, we run stochastic simulations using this quantitative framework, to compare the likelihoods of various levels of epistatic interactions between the SR loci, and various scenarios of introduction of the driving mutations on the X chromosome. We focus on the three features of the polymorphism pattern that were emphasized above, because they bear interpretable information. We show that, in a region where several genes have been identified as candidates for selection on a phenotype, the polymorphism pattern can provide more information than the simple presence/absence of selection, and can potentially be enlightening about the fundamental genotype–phenotype–fitness relationship.

2. Methods

Wherever possible we used deterministic numerical computations. However, the evolution of the

https://doi.org/10.1017/S0016672309000147 Published online by Cambridge University Press

polymorphism pattern at several neutral loci under the infinite site model of mutation and with complex selection is mathematically intractable, so our analysis mainly relies on stochastic simulations.

(i) *A general 2-locus model of sex-ratio meiotic drive*

We model sex chromosome meiotic drive favouring X chromosomes against Y chromosomes in males, and determined by two loci (SR_1 and SR_2) on the X chromosome. Alleles are noted in italics; SR_i denotes the mutant, potentially driving allele at locus SR_i, whereas sr_i denotes the wild allele. Recombination occurs at rate r between SR_1 and SR_2 in females, consistently with what was shown for *D. simulans*, but contrary to other meiotic drive systems that have evolved chromosomal inversions (Jaenike, 2001). The segregation coefficient k_i ($1/2 \leqslant k_i \leqslant 1$) is the proportion of X-bearing sperm produced by males that carry only one driving allele SR_i. Note that k_i is also the proportion of females in their progeny. If we note α_i ($0 \leqslant \alpha_i \leqslant 1$) the proportion of Y-bearing sperms eliminated by the meiotic drive due to SR_i (starting from equal amounts of X and Y chromosomes), then $k_i = 1/(2-\alpha_i)$. The phenotypic effect of SR_i is thus the elimination of a proportion $\alpha_i = (2k_i - 1)/k_i$ of Y-bearing sperm, since we assume no pleiotropic effect of meiotic drive genes on viability or fertility (see Discussion section). We note k_{12} the segregation coefficient of individuals carrying both driving alleles SR_1 and SR_2 (and α_{12} accordingly). In the absence of interaction between the SR loci, each allele SR_i eliminates a proportion α_i of sperms left intact by the other allele SR_j, so that the proportions of surviving Y-bearing sperm are multiplicative between the SR loci. When there is also functional interaction between the SR loci, we consider that the proportion of Y-bearing sperm is further reduced by a factor $1-e$ ($0 \leqslant e \leqslant 1$) such that in the general case, the fraction of the initial Y-bearing sperms that survive is

$$1-\alpha_{12} = (1-\alpha_1)(1-\alpha_2)(1-e) = \frac{(1-k_1)(1-k_2)(1-e)}{k_1 k_2} \tag{1}$$

and

$$k_{12} = \frac{1}{2-\alpha_{12}} = \frac{k_1 k_2}{k_1 k_2 + (1-e)(1-k_1)(1-k_2)}. \tag{2}$$

The term e quantifies the strength of the interaction between SR loci in an explicit way, related to their phenotypic effect (the destruction of Y-bearing sperm). The combined segregation coefficient k_{12} then varies from k_1 (for $e=0$ and $k_2 = 1/2$) to 1 (for $e=1$). This general framework allows considering different cases of interest regarding the genetic determinism of the meiotic drive, from independent effects of both loci ($k_1 > 1/2$, $k_2 > 1/2$, $e=0$) to obligate synergistic effects ($k_1 = 1/2$, $k_2 = 1/2$, $e>0$). An intermediate relevant case is the interaction of a driving locus with an enhancer locus otherwise neutral ($k_1 > 1/2$, $k_2 = 1/2$, $e>0$).

The dynamics at the SR loci can be calculated deterministically. We use labels a, b, c and d, for the two-locus haplotypes SR_1–SR_2, SR_1–sr_2, sr_1–SR_2 and sr_1–sr_2, respectively, and denote $x_{h\mathrm{f}}$ and $x_{h\mathrm{m}}$ the frequencies of any haplotype h among X chromosomes in male and female gametes, respectively. In eggs, the new frequency of haplotype h after one generation is

$$\begin{aligned} &x'_{h\mathrm{f}} = \overline{x_h} + 2\delta r(\overline{\overline{C}} - 1/2\overline{C}), \quad \delta = 1 \text{ if } h \in \{b, c\}, \\ &\delta = -1 \text{ if } h \in \{a, d\}. \end{aligned} \tag{3}$$

where $\overline{x_h} = (x_{h\mathrm{m}} + x_{h\mathrm{f}})/2$, $\overline{\overline{C}} = \overline{x_a}\,\overline{x_d} - \overline{x_c}\,\overline{x_b}$ and $\overline{C} = (C_\mathrm{m} + C_\mathrm{f})/2$. $C_\mathrm{m} = x_{a\mathrm{m}}x_{d\mathrm{m}} - x_{c\mathrm{m}}x_{b\mathrm{m}}$ is the LD in sperm and similarly C_f the LD in eggs. Note that the '$-$' symbol is used here for the sake of clarity of notation, but does not denote an average value in the population, since the sex ratio is not necessarily 1/2.

In males, the X chromosome is maternally inherited, so the genotypic frequencies in males are equal to those among females in the previous generation. Those frequencies are then affected by the meiotic drive, and the new frequencies of haplotypes a, b, c and d among X chromosomes in sperm after one generation are

$$\begin{aligned} x'_{a\mathrm{m}} &= \frac{2k_{12}x_{a\mathrm{f}}}{2k_{12}x_{a\mathrm{f}} + 2k_1x_{b\mathrm{f}} + 2k_2x_{c\mathrm{f}} + x_{d\mathrm{f}}}, \\ x'_{b\mathrm{m}} &= \frac{2k_1x_{b\mathrm{f}}}{2k_{12}x_{a\mathrm{f}} + 2k_1x_{b\mathrm{f}} + 2k_2x_{c\mathrm{f}} + x_{d\mathrm{f}}}, \\ x'_{c\mathrm{m}} &= \frac{2k_2x_{c\mathrm{f}}}{2k_{12}x_{a\mathrm{f}} + 2k_1x_{b\mathrm{f}} + 2k_2x_{c\mathrm{f}} + x_{d\mathrm{f}}}, \\ x'_{d\mathrm{m}} &= \frac{x_{d\mathrm{f}}}{2k_{12}x_{a\mathrm{f}} + 2k_1x_{b\mathrm{f}} + 2k_2x_{c\mathrm{f}} + x_{d\mathrm{f}}}. \end{aligned} \tag{4}$$

We iterated equations (3) and (4) to calculate the deterministic dynamics of the frequency of X^{SR} chromosomes, and that of the LD between the SR loci. The LD was calculated as D', which is the classical D (covariance between allelic states at two loci) divided by its maximum expected value based on allelic frequencies (Lewontin, 1995). Specifically, we calculated the expected D' in males at the generation when the frequency of X^{SR} chromosomes reached 0·6, which is the frequency observed in natural populations of Madagascar. This allowed us to evaluate the influence of the epistasis parameter e on the association between the SR loci.

(ii) *Simulation method*

To study the influence of selection and epistasis on the polymorphism pattern along the recombining region

https://doi.org/10.1017/S0016672309000147 Published online by Cambridge University Press

of the X chromosome that includes the SR_1 and SR_2 distorter loci, we used forward individual-based stochastic simulations. We used a modified version of the program used in Chevin *et al.* (2008), which can simulate several DNA sequence fragments ('markers') with mutation within fragments (under the infinite site model) and recombination within and between fragments. Two markers were placed such as to include each of the SR loci (the causative loci were considered to be restricted to a single nucleotide). Another marker was placed in the middle of the SR_1–SR_2 interval. For each marker, we generated the initial neutral polymorphisms for all the X chromosomes in the population by coalescence simulations using the program 'ms' (Hudson, 2002), since coalescence theory remains a good approximation when the sample size is close to the effective population size (Wakeley & Takahashi, 2003). We assumed that the sex ratio was unbiased before the introduction of the meiotic drive allele(s), so that the size of the population of X chromosomes was $N_X = 3N_e$, where N_e is the effective population size. For each fragment, we simulated $3N_e$ sequences using 'ms', with the mutation parameter θ and the recombination parameter ρ defined at the scale of the entire fragment (rather than per nucleotide), as is common practice when using the infinite site model (see for instance Hudson (2002) and Przeworski (2002)). Empirical estimates of ρ and θ suggest that ρ is roughly twice as large as θ in normally recombining genomic regions of *D. simulans* (Kliman *et al.*, 2000), so we used $\theta = 3$ and $\rho = 6$, which roughly corresponds to 300 bp long sequences. Then, the selective sweeps at the meiotic drive loci SR_1 and SR_2 were simulated forward in time. Recombination occurred in females only, at rate $r = \rho/(2N_e)$. Segregation distortion occurred in males, as described in the Model section. We assumed that the driving alleles had no deleterious effects on fertility or viability (such an effect would be mostly equivalent to decreasing the strength of the drive). Mutation occurred in both sexes, at a rate $\mu = \theta/3N_e$. We used an effective population size of $N_e = 10\,000$. This value is lower than the actual effective size usually reported for fruit flies, and was chosen because it was tractable in individual-based forward simulations. Nevertheless it may not affect our result strongly, since we used relevant values of the population parameters for recombination and mutation inside each fragment. Hence the main consequence of using a small population size in our context is to increase the amount of drift, thus limiting the strength and duration of signatures of selection. This could affect our results quantitatively to some extent, but not qualitatively.

The various simulations differed in the parameters of the meiotic drive (k_1, k_2 and e). The simulations also differed in the scenarios regarding driver alleles at SR_1 and SR_2, which could be introduced either (i) together on the same haplotype (which represents a migration event from another population), or (ii) separately in time (delayed). In all cases, simulations, where either of the alleles was lost, were discarded as in Chevin *et al.* (2008). When the driving alleles appeared by mutation, they were introduced in five copies in order to decrease computation time. This does not affect the generality of the results (see Chevin *et al.* (2008)), and relies on the fact that a beneficial mutation that is fated to fixation (i.e. conditional on ultimate fixation) rises quickly in frequency (Barton, 1998). In the case of delayed appearance, SR_2 was present but behaved neutrally before the introduction of the driving allele at SR_1. Hence SR_2 was taken to be one of the neutral polymorphic sites from the 'ms' simulation, at which the derived allele was chosen to be the driving allele.

We cancelled the meiotic drive effect when the pooled frequency of distorters – regardless of their quantitative effects – among X chromosomes reached 0·6, the value observed in the natural population of Madagascar (Atlan *et al.*, 1997). This was meant to represent the effect of rapidly invading drive suppressors on the Y chromosomes and/or on autosomes (Atlan *et al.*, 2003), or a frequency-dependent disadvantage of X^{SR} in fertility (Taylor & Jaenike (2002), see Discussion section). The population was then left to evolve for an additional 200 generations, and 25 samples were drawn from each simulation every 50 generations, from which statistical measures were made. Samples consisted of 10 X^{SR} and 5 X^{ST} chromosomes as in Derome *et al.* (2008).

(iii) *Likelihood of scenarios*

Our aim was to find which genetic scenario was the most consistent with the observed polymorphism pattern in the region. We chose to study two realistic cases of interest regarding the interaction between SR loci. The first case is obligate interaction of the meiotic drive elements, whereby none has an effect of its own ($k_1 = 0{\cdot}5$ and $k_2 = 0{\cdot}5$), as observed in the lab against standard Y chromosomes. In the second case, SR_1 is a meiotic drive locus, whose effect is possibly enhanced by SR_2 (otherwise neutral). In this scenario, we chose $k_1 = 0{\cdot}75$ (and still $k_2 = 0{\cdot}5$). This other scenario is consistent with the fact that many meiotic drive systems are thought to evolve by recruiting interacting elements at linked loci during their spread in a population (Crow, 1991). Within each of these qualitatively distinct genetic interaction schemes, several values of the epistasis parameter e were simulated (from $e = 0{\cdot}33$ to $e = 0{\cdot}89$, corresponding under obligate interaction to $k_{12} = 0{\cdot}6$ and $k_{12} = 0{\cdot}9$, respectively). Hence, we assessed both the qualitative and quantitative influences of epistasis on the likelihood

https://doi.org/10.1017/S0016672309000147 Published online by Cambridge University Press

of the data. We also assessed the importance of the time since the meiotic drive effect stopped. We contrasted the two scenarios of introduction of the *SR* alleles presented in the 'Simulation methods' section (either together on the same haplotype, or one after the other at different generations), since those correspond to two extreme cases regarding the initial LD between SR_1 and SR_2.

In analysing the polymorphism pattern, we focused on the three features described in the introduction. First, Derome *et al.* (2008) showed that diversity is very low in X^{SR} chromosomes relative to the standard chromosomes at two regions involved in the sex-ratio distortion (Fig. 1). Defining $R_\pi = \pi(\mathrm{SR})/\pi(\mathrm{ST})$ as the ratio of nucleotide diversities of distorter over standard X chromosomes, our first criterion was then $R_\pi = 0$ for the markers that include SR_1 or SR_2. The proportions of simulations satisfying the criterion for each locus yielded probabilities P_{SR_1} and P_{SR_2}, respectively.

The second feature was the LD between the SR loci, measured by D'. Derome *et al.* (2008) found strong LD between the two SR candidate regions ($D' = 1$, see Fig. 1). In our simulation results, we treated each of the two markers encompassing the SR loci as a biallelic locus, by considering the most frequent haplotype against all the others. This gave a measure of D'. Then we calculated P_{LD} as the proportion of simulations satisfying the criterion $D' = 1$.

The third feature was the relative nucleotide diversity at markers located between the SR loci, in the middle of the interval, which we denote $R_\pi(m)$. In the data $R_\pi(m) \approx 1$, so the third probability we calculated was P_m, the proportion of simulations with $R_\pi(m) \geqslant 1$.

Finally, we scored the likelihood as the joint probability of all the above-mentioned events, namely $\Lambda = \Pr((R_\pi(1) = 0) \text{ and } (R_\pi(2) = 0) \text{ and } (R_\pi(m) \geqslant 1) \text{ and } (D' \geqslant 1))$. Note that this is not equal to the product of the above-mentioned probabilities, since some of these events may be strongly correlated.

3. Results

To elucidate how selection and epistasis may have shaped the polymorphism pattern in X chromosomes of *D. simulans*, we proceeded in three steps. First, we used deterministic recursions to investigate how epistasis influences the expected LD between the selected loci in this context. This allowed us to assess what type of interaction between the SR loci could lead to high LD between them. Second, we simulated the evolution of neutral polymorphism across the region that encompasses the driving loci, in order to contrast two scenarios of introduction of the driving alleles: either simultaneously on the same haplotype, or at different generations on two distinct chromosomes. The polymorphism pattern was averaged over many replicate simulations for each scenario. This did not however account for the stochasticity of population genetics, which may produce outcomes different from the expected ones. Therefore in the third and final part, we calculated the probability of the observed polymorphism pattern for several values of epistasis and modes of introduction of the driving alleles, in order to identify which scenario is the most consistent with the molecular data.

(i) *Epistasis and LD*

The high LD between the regions involved in the *sex-ratio* drive of *D. simulans* in the natural population of Madagascar (Fig. 1*b*) suggests that epistatic interactions exist between the genes underlying the drive. However, this does not necessarily imply obligate interaction of the SR loci as observed in the lab against standard chromosomes (Montchamp-Moreau *et al.*, 2006). Obligate interaction would imply that SR_1 and SR_2 both drifted neutrally before they reached high frequency and recombined together, which may seem like an unlikely assumption. Alternatively, the high LD may have been caused by less extreme interaction, as occurs for instance when SR_2 is an enhancer allele. We used deterministic numerical recursions to assess how likely it is to observe high LD under various types of epistasis. Figure 2 plots the LD (measured as D') when the frequency of X^{SR} chromosomes reaches 0·6, the value observed in Madagascar, against the epistasis parameter e. The LD between SR loci increases logistically with e. Moreover for a given value of e (except at very low values), D' increases with decreasing k_1. Indeed, when k_1 is low, most of the effect of the meiotic drive is due to the genetic interaction, so it is a gene combination that is selected, which results in strong LD. As a result, a strong LD between the SR loci (i.e. $D' \sim 1$) can be achieved over a large range of epistasis values, not only when SR_1 does not drive on its owns (obligate interaction of SR loci) but also when its own drive effect is moderate (for instance when $k_1 = 0{\cdot}6$, for $0{\cdot}4 < e < 1$, which correspond to $k_{12} \geqslant 0{\cdot}7$, Fig. 2). Therefore in the following, we compare the case of an enhancer locus with that of obligate interaction between the SR loci.

(ii) *Average polymorphism pattern*

Beside the strong LD between regions involved in the *sex-ratio* distortion, the X polymorphism in the Malagasy populations of *D. simulans* exhibits a striking pattern, with reduced $R_\pi = \pi(\mathrm{SR})/\pi(\mathrm{ST})$ at the causative regions, but a high R_π in between. We investigated this pattern using forward stochastic simulations. Their results (averaged over 500 replicates) are shown in Fig. 3 in the case of obligate

https://doi.org/10.1017/S0016672309000147 Published online by Cambridge University Press

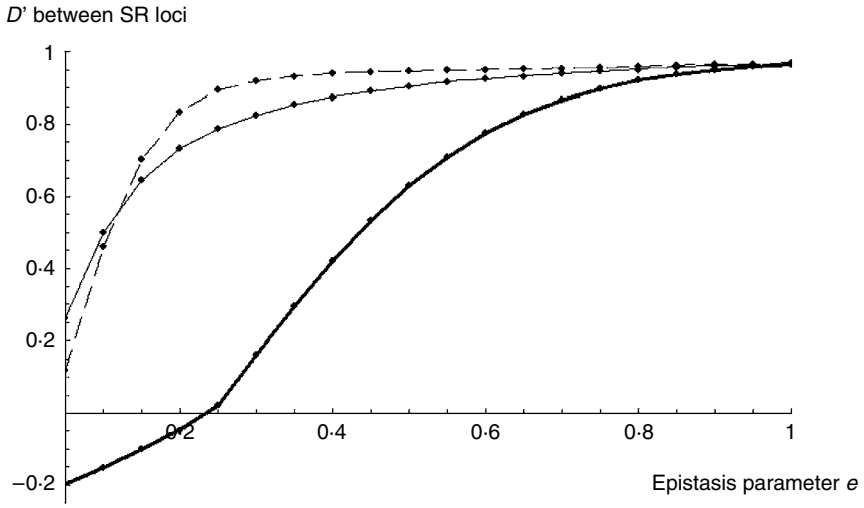

Fig. 2. Epistasis and LD: the expected LD between SR_1 and SR_2 (measured as D') when the frequency of X^{SR} chromosomes reaches 0·6 in the population is plotted against the value of the epistasis parameter e. Results were obtained by recursions of equations (3) and (4), with population size $N=10\,000$, and the driver alleles introduced simultaneously in one copy on different haplotypes. Dashed line: $k_1=0{\cdot}5$, $k_2=0{\cdot}5$ (obligate interaction), thin line: $k_1=0{\cdot}6$, $k_2=0{\cdot}5$ (SR_2 enhancer), heavy line: $k_1=0{\cdot}75$, $k_2=0{\cdot}5$ (SR_2 enhancer).

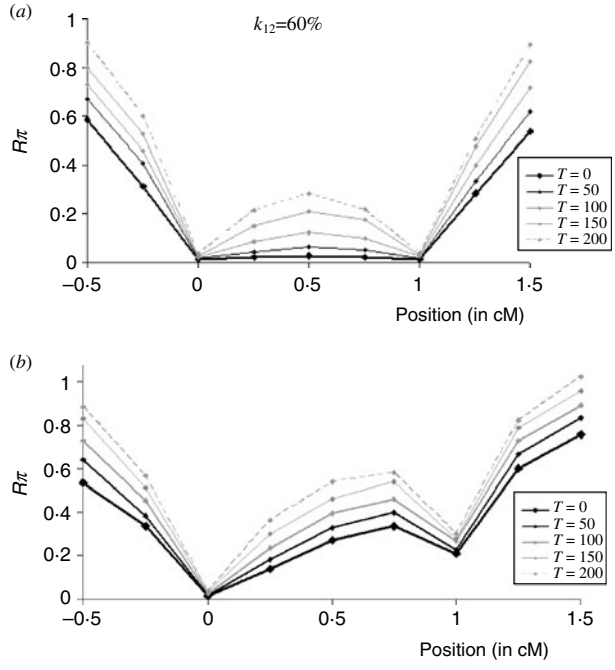

Fig. 3. Example of simulated polymorphism pattern close to the SR loci: the ratio R_π of nucleotide diversities between X^{SR} (driver) and X^{ST} (non-driver) chromosomes is plotted against the location along the chromosome (in cM, the location of the first driving locus SR_1 is set to 0, the second locus SR_2 is at 1 cM). Values are averaged over 500 simulations under obligate interaction ($k_1=k_2=0{\cdot}5$), and with low combined effect of the SR loci on the meiotic drive ($e=0{\cdot}33$, $k_{12}=0{\cdot}6$). (*a*) Both driving alleles introduced together on the same haplotype, (*b*) driving allele at SR_2 drawn from the standing genetic variation. T: time after the interruption of the selective sweeps.

interaction between the loci, with a low level of drive ($k_{12}=60\,\%$). Distinctive patterns emerge depending on the scenario of introduction of the driver alleles.

If both driver alleles are introduced together on the same haplotype (Fig. 3*a*) – thus mimicking migration – then the diversity in X^{SR} when the meiotic drive effect stops ($T=0$) is strongly reduced relative to X^{ST} chromosomes in the whole SR_1–SR_2 interval. This reflects the propensity of this region to behave functionally like a single locus as a consequence of the obligate interaction between the *SR* loci, even if the level of drive is moderate. After the meiotic drive effect has stopped, the frequency of X^{SR} chromosomes (i.e. SR_1–SR_2 genotypes) drifts in the population, and recombination shuffles neutral polymorphisms between X^{SR} and X^{ST} backgrounds, thus increasing the ratio R_π. Note however that with obligate interaction, at least two recombination events are required to introduce new polymorphisms on X^{SR} chromosomes in this region.

The pattern is quite different when SR_2 exists in the population and drifts neutrally before SR_1 is introduced by mutation (Fig. 3*b*). In such a case, selection of both SR_1 and SR_2 is triggered by the recombination event(s) that bring(s) them together. During the selective sweep, X^{SR} chromosomes (SR_1–SR_2 genotypes) can recombine with several of the sr_1–SR_2 chromosomes that are in non-negligible frequency in the population (frequency drawn from the neutral distribution under the infinite site model (Ewens, 2004)). These single recombination events introduce new polymorphisms in the SR_1–SR_2 genotypic class, thus increasing the ratio R_π within the

https://doi.org/10.1017/S0016672309000147 Published online by Cambridge University Press

[SR_1–SR_2] interval relative to the case in Fig. 3*a*. Similarly, R_π at SR_2 at the end of the selective sweep is also higher than in Fig. 3*a*, since several haplotypes carrying the SR_2 mutation may have swept to high frequency. The presence of several different haplotypes carrying the SR_2 mutation at the beginning of the sweep makes the selective sweep at SR_2 similar to a 'soft sweep' from the standing genetic variation (Innan & Kim, 2004; Hermisson & Pennings, 2005; Przeworski *et al*., 2005), a particular case of 'traffic' selection (Kirby & Stephan, 1996) where the alleles under selection are at the same site, and are identical by descent. Here, however, contrary to the usual 'soft sweep' models, selection of the SR_2 allele is controlled by its genetic background rather than by the environment: only the copies of SR_2 associated with a SR_1 allele become selected, while the others remain effectively neutral.

(iii) *Likelihood of evolutionary scenarios*

The simulation results described above are averaged over many replicates, and allow comparing models and data qualitatively, not quantitatively. They do not account for the stochasticity in the population genetics processes involved and in the sampling. Since stochasticity may cause the observed pattern to differ substantially from the expected one, we also calculated the probability to observe the polymorphism pattern under the various scenarios presented above and for various levels of epistasis.

The proportions of simulations (P) that satisfied each of the criteria defined in the Methods section are shown in Fig. 4, using a colour plot (with lighter colour for higher values). Maximum and minimum values are also indicated. As seen in the first row of Fig. 4, the probability of a high value of D' between SR_1 and SR_2 at the end of the sweep is high under the four situations presented, and decreases with time as recombination events accumulate. In the case of obligate interaction (columns 1 and 3), epistasis also influences the LD: higher values of e induce higher P_{LD} at the time when meiotic drive is interrupted, and even afterwards, as is apparent from the inclined iso-P_{LD} in Fig. 4 under obligate interaction.

The trend for P_{SR_1} is very similar to that for P_{LD}, but there is less variation between scenarios (columns) and between parameter values within each scenario. Indeed, SR_1 always experiences strong selection through its effect on the meiotic drive, so a strong signature of selection at SR_1 is a general outcome of all the scenarios considered here. For the reason explained above, the epistasis parameter e has also more influence on the signature at SR_1 under obligate interaction.

Consistent with the results of the qualitative approach, the high relative diversity between the SR loci is very difficult to obtain under all scenarios, as can be seen by the very low values of P_{m}. Indeed, in all the scenarios investigated, the epistatic interaction (and the strong initial LD in the 'coupling' case) leaves little opportunity for the selective sweeps at SR_1 and SR_2 to interfere in their hitchhiking effects, as described in Chevin *et al*. (2008). However in all cases, P_{m} increases sharply after the selective sweeps are interrupted (a 30+ fold increase from $T=0$ to $T=200$ in the first column), as a consequence of recombination with single mutant genotypes (SR_1–sr_2 or sr_1–SR_2). Higher P_{m} values are reached under the 'soft sweep' scenario, due to the traffic selection of several haplotypes. Comparing the modes of interaction between SR loci shows that P_{m} is higher when SR_2 is an enhancer of the drive than with obligate interaction. Indeed in such a case, recombination events that produce SR_1–sr_2 genotypes also increase R_π.

Diversity at SR_2 is the most variable feature of the polymorphism pattern. For instance, when the selective sweeps stop ($T=0$), $P_{SR_2}=0{\cdot}95$ under obligate interaction with initial coupling, whereas $P_{SR_2}=0{\cdot}17$ in the enhancer case with selection from the standing genetic variation. Besides, the influence of the epistasis parameter e on P_{SR_2} depends on the scenario: P_{SR_2} is most sensitive to e in the case of obligate interaction and selection from the standing variation (soft sweep or traffic selection). Those results can be interpreted as follows. SR_2 is necessary for the drive to occur only in the case of obligate epistasis. When SR_2 is an enhancer allele, the signature of selection at SR_2 can be strong only if SR_2 is recruited early in the course of the sweep at SR_1. If SR_2 is already present when SR_1 is introduced in the population, then it can combine earlier with a SR_1 background and become selected, which may strengthen the signature of selection at SR_2. However, the larger the initial number of SR_2 copies in the population, the higher the probability of trafficking, and so the weaker the signature of selection at SR_2. This latter effect dominates, and so the probability P_{SR_2} to observe no diversity at the marker located at SR_2 is low in this case.

The global likelihood Λ gives the probability to jointly observe all the features of the polymorphism pattern. Its behaviour depends on the scenario and on the type of interaction. In the case of obligate interaction between the SR loci, when SR_1 and SR_2 are initially on the same haplotype, Λ mainly reflects the behaviour of P_{m}, which indeed is the most variable feature in this context (column 1 in Fig. 4). In this case, the time T after the end of the meiotic drive mainly determines the probability to observe the data, since with increasing T more diversity is introduced by recombination between SR_1 and SR_2. In all other cases, the global likelihood Λ depends in a complex way on the interaction of all the previously described

https://doi.org/10.1017/S0016672309000147 Published online by Cambridge University Press

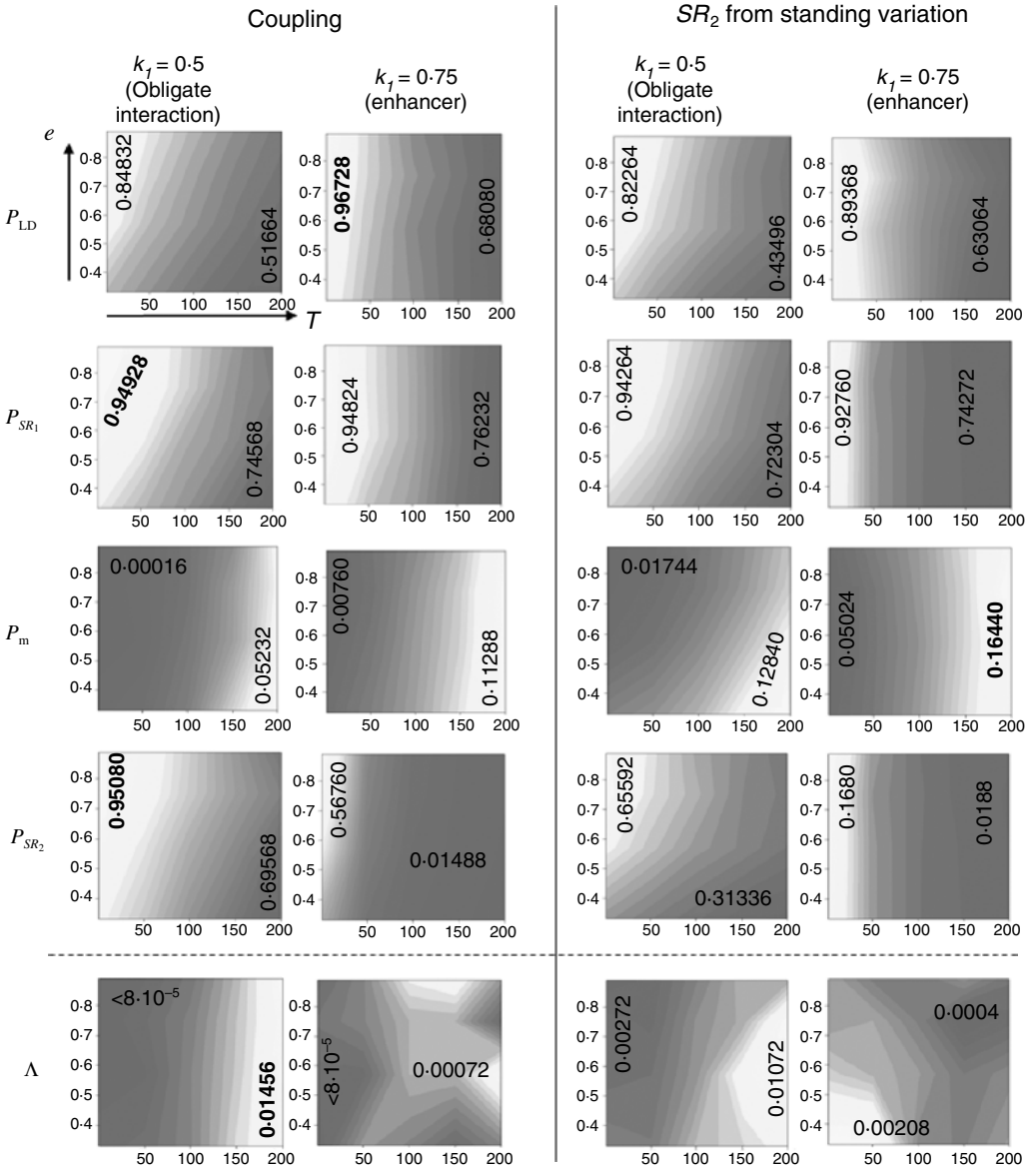

Fig. 4. Likelihood of the polymorphism pattern. Each figure shows the proportions of simulations (Probabilities P …) that satisfied each of the criteria defined in the Methods section, where the ordinate is the strength of the epistatic interaction e between SR_1 and SR_2, and the abscissa is the time T after the end of the meiotic drive effect. Lighter colours correspond to higher values, and the minimum and maximum values are indicated for each plot. On each line, the maximum value over all conditions is written in bold characters. Probabilities are that: P_{LD}: the scaled LD D' between the markers that carry the SR loci equals 1; P_{SR_1} or P_{SR_2}: the relative diversity R_π is zero at the haplotypes carrying the SR loci; P_m:R_π in the middle of the interval is greater than 1; or Λ: all those criteria are satisfied together. 'Coupling': introduction of the driver alleles at SR_1 and SR_2 on the same haplotype; 'SR_2 from standing variation': SR_2 is a neutral (derived) allele drawn from the background variation at the beginning of the simulation. Parameters are as described in the 'Simulation method' section.

probabilities, so it cannot be predicted directly by their product (not shown). For instance, for a 'soft' sweep at SR_2 under obligate interaction, Λ is maximized for intermediate values of epistasis and older sweeps. The pattern is even more complex in the 'enhancer' case: intermediate values of e and of T are favoured if SR_1 and SR_2 are introduced on the same haplotype, whereas low values of both T and e are favoured if SR_2 is introduced first. Indeed in this latter case P_{SR_2} is the most variable feature (fourth column in Fig. 3), such that T close to 0 is favoured. Among all the cases studied here, the most consistent with the data would be those where both driving loci are necessary for the drive (i.e. the 'obligate interaction'). The case where both driving alleles are introduced together on the same haplotype seems more likely, although we lack power to rule out the possibility that SR_2 was introduced first. Hence, the

https://doi.org/10.1017/S0016672309000147 Published online by Cambridge University Press

diversity between the SR loci on X^{SR} chromosomes seems to have been introduced by recombination events that occurred after the meiotic drive effect had stopped, rather than through the effect of two interfering selective sweeps as in Chevin *et al.* (2008). A purely enhancer SR_2 locus is not sufficient to cause substantial reduction of diversity at SR_2 in the context we studied here.

4. Discussion and conclusions

We studied the polymorphism pattern in the SR region of the X chromosome of *D. simulans*, where two genomic regions putatively interact to cause sex chromosome meiotic drive in the natural populations of Madagascar. Using stochastic forward simulations, we investigated what type and strength of epistatic interactions were the most consistent with the observed polymorphism data. We showed that, under simple scenarios, despite the high diversity between the candidate regions for the drive, the polymorphism pattern was best explained by strong obligate epistasis between the SR loci, which is consistent with what was otherwise observed in the lab using a peculiar X–Y pair of chromosomes (Montchamp-Moreau *et al.*, 2006).

That the diversity in-between the candidate regions for the meiotic drive is as high in X^{SR} as in X^{ST} chromosomes seems at first contradictory with strong positive epistasis between two close loci under selection. However, if the selective sweeps at the driver loci have been interrupted in their course, recombination has had opportunity to shuffle polymorphisms between X^{SR} and X^{ST} chromosomes after selection has stopped. Hence, the polymorphism pattern in this case may reflect recombination events that took place not only while the mutations were sweeping through the population, but also afterwards, contrary to cases of completed selective sweeps. Our result do not allow to conclusively decide which scenario is most likely between the introduction of both driver alleles on the same initial haplotype, or the presence of one of the driver alleles in a neutral state in the population before the introduction of the driver allele at the other locus. The presence of several haplotypes at intermediate frequencies between the candidate regions for the drive on X^{SR} chromosomes (Derome *et al.*, 2008) seems more consistent with this second possibility.

Here, we assumed that the selective sweeps at the SR loci had been stopped suddenly by suppressors (not explicitly modelled). Another possibility would be that the X^{SR} chromosomes have been held in check by a disadvantage in sperm competition against X^{ST} chromosomes, as modelled by Taylor & Jaenike (2002). Both in the case of suppressors and of sperm competition, the dynamics are expected to be frequency dependent, since opposing mechanisms are triggered by the spread of X^{SR} chromosomes, either through direct competition or as a by-product of the change in the sex ratio (which increases the disadvantage of X^{SR} in male fertility). This frequency dependence was overlooked here for the sake of clarity. On the one hand, the dynamics of a complete suppressor on the Y chromosome is expected to be very quick once the frequency of X^{SR} chromosomes is sufficiently high, such that the spread of the suppressor is well mimicked by instantaneous cancellation of the meiotic drive (results not shown). On the other hand, if the drive had been stopped as a consequence of frequency-dependent deleterious effects, it is not clear how quickly the dynamics of the SR loci would stop. We should, however, point out that sperm competition is expected to impact the dynamics of X^{SR} chromosomes in a restricted range of sex ratio values (combining high male and female mating rates), and to depend on frequency independent factors such as local population density, for instance. In any case, since the time after the end of the drive was expected to affect several features of the polymorphism pattern, we needed to assess the consequences of an interruption of the selective sweeps. That the selective sweeps may have stopped progressively to some extent, instead of suddenly, should not strongly affect our results, at least qualitatively.

We also chose to focus on simple scenarios regarding the type of interaction between the SR loci and the introduction of the driver alleles, which appeared consistent with the current knowledge of the biology of the species and the specificities of this genetic system (Atlan *et al.*, 1997, 2004; Montchamp-Moreau *et al.*, 2006). The history of the Paris *sex-ratio* meiotic drive may also have been governed by more complex dynamics. Indeed, meiotic drive elements induce genetic conflicts that can lead to arms races (Hurst *et al.*, 1996; Hurst & Werren, 2001), in which meiotic drive genes that sweep through the population are stopped by suppressors on the Y chromosomes and on autosomes, before being followed by another bout of meiotic drive and subsequent suppression. Such an antagonistic intragenomic coevolution seems to be confirmed by the existence of two other *sex-ratio* systems in the same species, which are apparently independent of the one studied here, and likely suppressed in natural populations (Tao *et al.*, 2007; Jaenike, 2008). Hence, one alternative to the scenarios considered here could be that a distorter allele at one of the loci was stopped in its course by (a) suppressor gene(s), and then restored in its drive ability by an enhancer allele at the other locus. The outcome of this scenario may be similar to some extent to the 'soft sweep from the standing genetic variation' (or 'traffic') scenarios considered here, except that the driving allele already present in the population would have

been selected in a recent past instead of being neutral. This would be consistent with the low relative diversity found around the two drive loci. It may also be possible that each locus had an effect of its own on meiotic drive, besides their combined effect through their genetic interaction. However in this case, there is only a small subset of parameter values that allows both driver alleles to reach similar frequencies simultaneously, and even then the likelihood of the data remains far below the values reported here (not shown). Finally, we could also envision less parsimonious scenarios involving balancing or negative selection at another locus between the SR loci to explain the high relative nucleotide diversity observed in this region. One interesting situation in that respect would be the one where a deleterious mutation linked to one of the distorter alleles would favour recombination to purge X^{SR} chromosomes from their load. This is particularly appealing since one of the two-drive locus in the Paris *sex-ratio* system is associated with a segmental duplication containing several genes (Montchamp-Moreau *et al*., 2006).

Finally here, the structure of the data allowed asking questions of interest, since it combined polymorphism information around two (previously demonstrated) selective sweeps that had not reached fixation (which allowed using the LD together with diversity measures), with phenotypic information about meiotic drive in the lab. Yet there were not enough degrees of freedom to estimate all possibly relevant parameters together (individual segregation coefficients k_i, epistatic interaction, possible delay between the introduction of the driving mutations, time since the end of the selective sweep). Instead, this study allowed comparing markedly different realistic scenarios and gave hints to guide further investigations in the lab or *in natura*. Indeed, the ideal data to study epistatic interaction with a population genetics perspective would imply serial measures of temporal variations in allelic frequencies and LD. While this is difficult to acquire, it may be easier to compare several populations in which it is suspected that the selected genes have been introduced at different times. This should be possible in the case of the Paris *sex-ratio* system of *D. simulans*, because a spread of distorter alleles is likely in progress from an African location in this worldwide species (Derome *et al*., 2004, Jutier *et al*., 2004).

We hope that this work illustrates how well chosen molecular data can indeed help deciphering gene by gene interactions, as already emphasized in recent papers (Caicedo *et al*., 2004; Derome *et al*., 2008), and guide further experimentation. Future work will tell whether it is possible to quantitatively estimate the strength of genetic interactions for fitness using molecular population genetic data, and to compare those estimates with measures obtained in the lab. The present paper is a first step in this direction, which will hopefully motivate other attempts in the future.

We thank Michel Cazemajor, Xavier Vekemans and Sylvain Billiard for helpful discussions throughout this project, and two anonymous reviewers for comments that helped improving the quality of this manuscript. L.-M.C., C.M.-M. and F.H. were supported by the ANR-06-BLAN-0128 grant from the Agence Nationale de la Recherche.

References

Atlan, A., Capillon, C., Derome, N., Couvet, D. & Montchamp-Moreau, C. (2003). The evolution of autosomal suppressors of sex-ratio drive in *Drosophila simulans*. *Genetica* **117**, 47–58.

Atlan, A., Joly, D., Capillon, C. & Montchamp-Moreau, C. (2004). Sex-ratio distorter of *Drosophila simulans* reduces male productivity and sperm competition ability. *Journal of Evolutionary Biology* **17**, 744–751.

Atlan, A., Mercot, H., Landre, C. & Montchamp-Moreau, C. (1997). The sex-ratio trait in *Drosophila simulans*: geographical distribution of distortion and resistance. *Evolution* **51**, 1886–1895.

Barton, N. H. (1998). The effect of hitch-hiking on neutral genealogies. *Genetical Research, Cambridge* **72**, 123–133.

Caicedo, A. L., Stinchcombe, J. R., Olsen, K. M., Schmitt, J. & Purugganan, M. D. (2004). Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences of the USA* **101**, 15670–15675.

Chevin, L. M., Billiard, S. & Hospital, F. (2008). Hitchhiking both ways: effect of two interfering selective sweeps on linked neutral variation. *Genetics* **180**, 301–316.

Chevin, L. M. & Hospital, F. (2006). The hitchhiking effect of an autosomal meiotic drive gene. *Genetics* **173**, 1829–1832.

Crow, J. (1991). Why is Mendelian segregation so exact? *Bioessays* **13**(6), 305–312.

Derome, N., Baudry, E., Ogereau, D., Veuille, M. & Montchamp-Moreau, C. (2008). Selective sweeps in a 2-locus model for sex-ratio meiotic drive in *Drosophila simulans*. *Molecular Biology and Evolution* **25**, 409–416.

Derome, N., Metayer, K., Montchamp-Moreau, C. & Veuille, M. (2004). Signature of selective sweep associated with the evolution of sex-ratio drive in *Drosophila simulans*. *Genetics* **166**, 1357–1366.

Ewens, W. J. (2004). *Mathematical Population Genetics – I. Theoretical Introduction*. New York: Springer-Verlag.

Gavrilets, S. (2004). *Fitness Landscapes and the Origin of Species*. Princeton, NJ: Princeton University Press.

Hermisson, J. & Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352.

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A. & Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101–104.

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338.

Hurst, G. D. D. & Werren, J. H. (2001). The role of selfish genetic elements in eukaryotic evolution. *Nature Reviews Genetics* **2**, 597–606.

Hurst, L. D., Atlan, A. & Bengtsson, B. O. (1996). Genetic conflicts. *Quarterly Review of Biology* **71**, 317–364.

Innan, H. & Kim, Y. (2004). Pattern of polymorphism after strong artificial selection in a domestication event. *Proceedings of the National Academy of Sciences of the USA* **101**, 10667–10672.

Jaenike, J. (2001). Sex chromosome meiotic drive. *Annual Review of Ecology and Systematics* **32**, 25–49.

Jaenike, J. (2008). X chromosome drive. *Current Biology* **18**, R508–R511.

Jutier, D., Derome, N. & Montchamp-Moreau, C. (2004). The *sex-ratio* trait in *Drosophila simulans*: a comparative approach. *Genetica* **120**, 87–99.

Kim, Y. & Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**, 765–777.

Kim, Y. & Stephan, W. (2003). Selective sweeps in the presence of interference among partially linked loci. *Genetics* **163**, 389–398.

Kirby, D. A. & Stephan, W. (1996). Multi-locus selection and the structure of variation at the white gene of *Drosophila melanogaster*. *Genetics* **144**, 635–645.

Kliman, R. M., Andolfatto, P., Coyne, J. A., Depaulis, F., Kreitman, M., Berry, A. J., McCarter, J., Wakeley, J. & Hey, J. (2000). The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**, 1913–1931.

Lewontin, R. C. (1995). The detection of linkage disequilibrium in molecular sequence data. *Genetics* **140**, 377–388.

Lyttle, T. W. (1991). Segregation distorters. *Annual Review of Genetics* **25**, 511–557.

Maynard Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research, Cambridge* **23**, 23–35.

Montchamp-Moreau, C., Ogereau, D., Chaminade, N., Colard, A. & Aulard, S. (2006). Organization of the sex-ratio meiotic drive region in *Drosophila simulans*. *Genetics* **174**, 1365–1371.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197–218.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G. & Bustamante, C. D. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research* **15**, 1566–1575.

Palopoli, M. F. & Wu, C. I. (1996). Rapid evolution of a coadapted gene complex: evidence from the segregation distorter (SD) system of meiotic drive in *Drosophila melanogaster*. *Genetics* **143**, 1675–1688.

Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179–1189.

Przeworski, M., Coop, G. & Wall, J. D. (2005). The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312–2323.

Rogers, S. M. & Bernatchez, L. (2005). Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology* **14**, 351–361.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffer, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R. & Lander, E. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837.

Tao, Y., Masly, J. P., Araripe, L., Ke, Y., Hartl, D. L. (2007). A sex-ratio meiotic drive system in *Drosophila simulans*. I: an autosomal suppressor. *PLoS Biology* **5**, 2560–2575.

Taylor, J. E. & Jaenike, J. (2002). Sperm competition and the dynamics of X chromosome drive: stability and extinction. *Genetics* **160**, 1721–1731.

Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology* **4**, e72.

Wakeley, J. & Takahashi, T. (2003). Gene genealogies when the sample size exceeds the effective size of the population. *Molecular Biology and Evolution* **20**, 208–213.

Williamson, S. H., Hubisz, M. J., Clark, A. G., Payseur, B. A., Bustamante, C. D. & Nielsen, R. (2007). Localizing recent adaptive evolution in the human genome. *PLoS Genetics* **3**, e90.