



**HAL**  
open science

# The interplay of local adaptation and gene flow may lead to the formation of supergenes

Paul Jay, Thomas G Aubier, Mathieu Joron

## ► To cite this version:

Paul Jay, Thomas G Aubier, Mathieu Joron. The interplay of local adaptation and gene flow may lead to the formation of supergenes. *Molecular Ecology*, 2024, 33 (24), pp.e17297. <10.1111/mec.17297>. <hal-04792556>

**HAL Id: hal-04792556**

**<https://hal.science/hal-04792556v1>**

Submitted on 20 Nov 2024

**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.



HAL Authorization

# The interplay of local adaptation and gene flow may lead to the formation of supergenes

Jay<sup>1,2</sup>  | Thomas G. Aubier<sup>3,4,5</sup>  | Mathieu Joron<sup>1</sup> 

Molecular Ecology (2024). e17297. <https://www.doi.org/10.1111/mec.17297>

Paul Jay and Thomas G. Aubier contributed equally to this work.

<sup>1</sup>Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), Université de Montpellier, CNRS, EPHE, IRD, Montpellier, France

<sup>2</sup>Center for GeoGenetics, University of Copenhagen, Copenhagen, Denmark

<sup>3</sup>Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

<sup>4</sup>Department of Biology, University of North Carolina, Chapel Hill, North Carolina, USA

<sup>5</sup>Centre de Recherche sur la Biodiversité et l'Environnement (CRBE), Université de Toulouse, CNRS, IRD, Toulouse INP, Université Toulouse 3 – Paul Sabatier (UT3), Toulouse, France

## Correspondence

Paul Jay, Center for GeoGenetics, University of Copenhagen, Copenhagen, Denmark.

Email: [paul.yann.jay@gmail.com](mailto:paul.yann.jay@gmail.com)

## Funding information

Agence Nationale de la Recherche, Grant/Award Number: ANR-18-CE02-0019-01; Human Frontier Science Program; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung

**Handling Editor:** Cristina Arias-Sardá

## Abstract

Supergenes are genetic architectures resulting in the segregation of alternative combinations of alleles underlying complex phenotypes. The co-segregation of alleles at linked loci is often facilitated by polymorphic chromosomal rearrangements suppressing recombination locally. Supergenes are involved in many complex polymorphisms, including sexual, colour or behavioural polymorphisms in numerous plants, fungi, mammals, fish, and insects. Despite a long history of empirical and theoretical research, the formation of supergenes remains poorly understood. Here, using a two-island population genetic model, we explore how gene flow and the evolution of overdominant chromosomal inversions may jointly lead to the formation of supergenes. We show that the evolution of inversions in differentiated populations, both under disruptive selection, leads to an increase in frequency of poorly adapted, immigrant haplotypes. Indeed, rare allelic combinations, such as immigrant haplotypes, are more frequently reshuffled by recombination than common allelic combinations, and therefore benefit from the recombination suppression generated by inversions. When an inversion capturing a locally adapted haplotype spreads but is associated with a fitness cost hampering its fixation (e.g. a recessive mutation load), the maintenance of a non-inverted haplotype in the population is enhanced; under certain conditions, the immigrant haplotype persists alongside the inverted local haplotype, while the standard local haplotype disappears. This establishes a stable, local polymorphism with two non-recombining haplotypes encoding alternative adaptive strategies, that is, a supergene. These results bring new light to the importance of local adaptation, overdominance, and gene flow in the formation of supergenes and inversion polymorphisms in general.

## KEYWORDS

chromosomal inversion, evolutionary theory, hybridization, introgression, population genetics, supergene

## 1 | INTRODUCTION

Discrete polymorphisms of traits composed of multiple phenotypic features are widespread in nature. Classic examples include

polymorphisms which combine variation in colour and behaviour in birds (Küpper et al., 2016; Lamichhaney et al., 2016; Tuttle et al., 2016), insects (Jay et al., 2021; Kunte et al., 2014), snails (Gonzalez et al., 2019; Murray & Clarke, 1976), and lizards (Sinervo

& Lively, 1996). Sexual and mating types in animals, plants, and fungi (Abbott et al., 2017; Bachtrog et al., 2014; Billiard et al., 2012; Hartmann et al., 2021; Kappel et al., 2017), and immune systems in vertebrates (Kaufman et al., 1999; The MHC Sequencing Consortium, 1999; Trowsdale & Knight, 2013), are other examples of polymorphisms involving the coordination of many phenotypic modalities. These particular polymorphisms have often been considered to be controlled by genetic architectures called supergenes, which are groups of multiple, linked functional genetic elements segregating together as single Mendelian loci (Thompson & Jiggins, 2014). While other genetic mechanisms, such as gene pleiotropy, may lead to such multi-dimensional polymorphism in some cases, several studies have shown that supergenes exist in nature (Hartmann et al., 2021; Jay, Leroy, et al., 2022; Takayama & Isogai, 2005). The genes participating in supergenes are often in linkage disequilibrium because they are captured by polymorphic chromosomal rearrangements, notably chromosomal inversions (Jay et al., 2018; Knief et al., 2017; Mérot, Oomen, et al., 2020; Tuttle et al., 2016; Wang et al., 2013). The coexistence of segments in alternative orientation which do not recombine with each other allows the maintenance of distinct combinations of phenotypic traits within populations. These alternative combinations of phenotype elements—sometimes called ‘syndromes’ or ‘morphs’—are often seen as alternative adaptive strategies; suppressed recombination prevents the formation of intermediate maladapted phenotypes and favours the maintenance of discrete phenotypic variation (Kirkpatrick, 2010; Schwander et al., 2014; Wellenreuther & Bernatchez, 2018).

The evolution of supergenes poses a longstanding puzzle in evolutionary biology (Charlesworth & Charlesworth, 1975; Llaurens et al., 2017; Schwander et al., 2014; Turner, 1967), primarily rooted in the conflicting impact of recombination on the evolution and maintenance of differentiated haplotypes. While suppressed recombination might seem to facilitate the evolution of distinct combinations of alleles when selection favours alternative evolutionary strategies (Thompson & Jiggins, 2014), it also prevents beneficial mutations that have arisen independently to be brought together on the same haplotype. This makes the formation of haplotypes with multiple beneficial alleles more challenging than when recombination occurs (Weinreich & Chao, 2005; Weissman et al., 2010). Seen from the opposite angle, recombination not only promotes the emergence of new adaptive allele combinations but also hinders their maintenance in a polymorphism due to its homogenizing effect. Further complicating the issue, the long-term coexistence of alternative evolutionary strategies in the same population requires specific mechanisms that balance their frequencies (Curtsinger et al., 1994; Johnston et al., 2013; Mérot, Oomen, et al., 2020; Roff & Fairbairn, 2007). Indeed, while many processes may lead to the emergence of new adaptive strategies in populations, even with epistasis [e.g. via genetic drift (Kopp & Hermisson, 2006; Weissman et al., 2010; Wright, 1932), varying selective conditions (Whitlock, 1997), major effect mutations (Charlesworth & Charlesworth, 1975), high mutation rate, and population size (Weinreich & Chao, 2005)], these processes are not expected to result in the maintenance of ancestral

and derived adaptive strategies in a polymorphism. To understand the evolution of supergenes, it is therefore crucial to elucidate jointly the processes by which non-recombining differentiated haplotypes arise and persist within a single population.

Recently, the evolution of differentiated non-recombining supergene haplotypes in different populations followed by gene flow was proposed as a viable scenario to account for the puzzle of supergene formation (Jay et al., 2018; Llaurens et al., 2017). Such a scenario would involve four main steps: (i) divergence of haplotypes in separate populations; (ii) accumulation of structural differences suppressing recombination (e.g. inversions) between these haplotypes; (iii) gene flow between these populations bringing structural variants carrying divergent haplotypes together in the same population; and (iv) maintenance of the different haplotypes by balancing selection within this population. This scenario is appealing because it involves a combination of mechanisms that occur frequently in natural populations and are each thoroughly studied theoretically. The first step, divergence in isolated populations, largely simplifies the problem of the formation of alternative supergene haplotypes, because divergence occurs much more readily if populations are at least partially isolated than under panmixia (Pinho & Hey, 2010; Raeymaekers et al., 2017). The second step involves structural variations (such as inversions) occurring in one or another population, whose evolution could be facilitated by selection, for instance, because it reduces the load induced by migration between differentially adapted populations (Connallon & Olito, 2021; Kirkpatrick & Barton, 2006; Yeaman, 2013). Once populations are characterized by ecologically and structurally differentiated haplotypes, gene flow may bring the haplotypes together (Dasmahapatra et al., 2012; Leroy et al., 2017; Roux et al., 2016). The last step is the stabilization of this haplotype polymorphism, for instance, via disassortative mate choice (Chouteau et al., 2017; Tuttle et al., 2016) or heterozygote advantage (Jay et al., 2021; Mérot, Llaurens, et al., 2020; Wellenreuther & Bernatchez, 2018), thereby resulting in the formation of a supergene. Such balancing selection often seems to result from an association between chromosomal inversions and a recessive mutation load in supergenes, an observation we will discuss below (Jay et al., 2021; Küpper et al., 2016; Mérot, Llaurens, et al., 2020; Tuttle et al., 2016; Wang et al., 2013).

Gene flow has recently been shown to be involved in the formation of a mimicry supergene in the butterfly *Heliconius numata* (Jay et al., 2018), in the formation of a new sex chromosome in the ninespine stickleback, *Pungitius pungitius* (Dixon et al., 2019), and is suspected in several other cases (e.g. Helleu et al., 2022; Stolle et al., 2022; Tuttle et al., 2016). Moreover, a similar process has been shown to be responsible for the diversification of mating types in *Arabidopsis* species (Castric et al., 2008) and in several fungi (Corcoran et al., 2016; Hesnauer et al., 2020; Sun et al., 2012).

The independent processes denoted as i, ii, iii, and iv, which could potentially jointly contribute to supergene formation, have been subject to thorough investigation (e.g. Berdan et al., 2021, 2023; Connallon & Olito, 2021; Kirkpatrick & Barton, 2006; Lenormand & Otto, 2000; Nei, 1967; Nei et al., 1967; Santos, 1986;

Tafreshi et al., 2022). For instance, Lenormand and Otto (2000) illustrated that modifiers reducing recombination between two loci with or without epistasis may be selectively advantageous in populations connected by migration. Kirkpatrick and Barton (2006) demonstrated that inversions capturing locally adapted alleles are likely to spread when two populations are exchanging migrants because it reduces the recombination-migration load. Yeaman (2013) further showed that chromosomal rearrangements, such as transpositions leading to a closer physical proximity between co-adapted loci, could be favoured for similar reasons. Kirkpatrick and Barton (2006) also posited that inversions could be stabilized at intermediate frequency if they capture one or several recessive deleterious mutations, creating a homozygous disadvantage and thereby preventing inversion fixation. Connallon and Olito (2021), Berdan et al. (2021) and Jay, Tezenas, et al. (2022) demonstrated theoretically that such homozygous disadvantage of inversions, resulting from the capture of recessive deleterious mutations, is likely to be a common phenomenon contributing to the maintenance of polymorphism.

However, while these models predict the spread (1) and, under certain conditions, the maintenance at intermediate frequency (2) of inversions, they do not provide insights into the formation of supergenes. Indeed, they do not account for the formation of a polymorphism involving two or more alternative haplotypes controlling alternative 'morphs', each characterized by different genome arrangements, within *a single population*. They only predict that populations exchanging migrants can exhibit different genome arrangements at specific genome locations, thereby maintaining alternative haplotypes controlling alternative morphs *in different populations* (1), and that *the same haplotype* can exist in two different genome arrangements (e.g. inverted and non-inverted) within a single population due to the presence of recessive deleterious mutations hindering fixation of the most beneficial arrangement (2). To date, models have therefore studied the condition governing the evolution of inversions, but not how it may lead to the evolution of a supergene, which are two different features. An inversion is indeed a *genetic variant* consisting of a switch of gene order and orientation at a genome location, while a supergene is a *genetic architecture* involving the segregation in a population of two or more haplotypes underlying alternative evolutionary strategies.

One can consider that if two connected populations are each associated with a given genome arrangement and haplotype, migration between these populations is likely to result in the formation of a polymorphism involving alternative haplotypes and arrangements within each population. This phenomenon aligns with observed instances in nature, such as in monkeyflower (Twyford & Friedman, 2015) or deer mice (Hager et al., 2022), but within-population processes do not strongly contribute to the maintenance of polymorphism in these cases. By contrast, other documented supergenes, exemplified by those controlling social polymorphism in the fire ant (Wang et al., 2013) and mating tactics in the ruff (Küpper et al., 2016), do not seem to be maintained polymorphic via a migration-selection balance. In other words, the maintenance of polymorphism does not seem to involve an exchange of migrants with

other populations fixed or polymorphic for alternative haplotypes in these cases. Consequently, existing models do not predict that such supergenes could form and be maintained.

Therefore, the formation of divergent functional inversion haplotypes needs to be investigated jointly with their maintenance in a stable polymorphism segregating within populations [supergenes, *sensu* (Thompson & Jiggins, 2014)]. We hypothesized that, under specific conditions, the interplay between migration, recombination suppression, local adaptation, and a recessive mutation load acting against the fixation of inversions, may lead to the formation of supergenes that persist even when migration ceases, that is, not caused by an influx of maladapted haplotypes due to migration. However, whether this may occur and under which condition remains unknown.

Here, we present a population genetic model to fill this gap and shed new light to the formation of supergenes owing to local adaptation and gene flow. We investigated the conditions under which an inversion polymorphism encoding a discrete polymorphism (i.e. a supergene) can evolve when adjacent populations experience different selection regimes. To this end, we constructed a diploid three-locus, two-island population genetic model. Among these three loci, we implemented a recombination modifier locus acting on heterokaryotypes (mimicking chromosomal inversions) and associated with a fitness cost when homozygous, which generates overdominance and balancing selection. Here, we are not only interested in understanding the evolution of inversions, which has been extensively studied (Berdan et al., 2023), but also the conditions under which a supergene forms (Thompson & Jiggins, 2014). Therefore, we tracked the frequencies not only of inverted segments but also of all allelic combinations in the two populations, allowing us to precisely pinpoint how inverted and non-inverted haplotypes are distributed across space [unlike in previous models such as (Kirkpatrick & Barton, 2006)]. Using stochastic simulations under a wide diversity of fitness landscapes, we show that the interplay between local adaptation and gene flow is likely to lead to the formation of supergenes. More generally, our study provides new insights into the consequences of the spread of inversions.

## 2 | METHOD

We investigate the conditions under which a supergene underlying the maintenance of a discrete polymorphism evolves when chromosomal inversions spread in neighbouring populations under divergent and disruptive selection. We define supergenes as genetic architectures maintaining, within a population, multiple non-recombining haplotypes (comprising multiple loci) encoding alternative strategies associated with different fitness optima. Polymorphisms at supergenes are therefore maintained in the long term by evolutionary forces balancing haplotype frequencies.

In a standard diploid population genetic model with drift, we implement chromosomal inversions as modifier alleles that control the recombination rate between two ecological loci. The expression of

these modifier alleles at homozygous state may also be associated with a fitness cost that reflects recessive deleterious effects associated with chromosomal inversions, as observed in many cases (Jay et al., 2021; Lamichhane et al., 2016; Mérot, Llaurens, et al., 2020; Tuttle et al., 2016; Wang et al., 2013). Simulations are divided into two phases. First, populations evolve under migration–selection–mutation–drift balance (migratory phase) in order to implement the role of local adaptation and gene flow in the formation of alternative haplotypes and their exchange between populations. Second, in order to isolate the influence of migration in the polymorphisms observed, we suppress migration and let populations evolve under a selection–mutation–drift balance only (non-migratory phase). With such simulation experiments, we aim at describing in detail the forces maintaining polymorphisms in populations and the conditions under which a supergene underlying a discrete polymorphism is formed and stably maintained even when gene flow is reduced or stops. While being a modelling ploy used to distinguish the influence of spatial and non-spatial processes on polymorphism, this two-phase scenario also corresponds to common situations in nature (Johannesson et al., 2020; Leroy et al., 2017; Roux et al., 2016). Indeed, the alternation of periods of gene flow and no gene flow between taxa is a fundamental process of generating and maintaining genetic diversity. Here, the two genetic compartments modelled may be taxa at any level of divergence and initially exchanging migrants until this exchange stops.

## 2.1 | Spatial structure and genotypes

We model a system of two diploid populations exchanging migrants (two-island model). This population genetic model postulates three autosomal diploid loci with alternative alleles represented by small versus capital letters. Two ecological loci, A and B, are subject to disruptive selection and can recombine. The third locus, C, is a recombination modifier. Locus C is physically linked to locus A (no recombination) and can impede recombination among loci A and B. More specifically, in heterozygotes with genotype Cc at the C locus, there is no recombination between loci A and B. The invasion of allele C at the modifier locus reflects the invasion of a chromosomal inversion, and we therefore denote the presence of the recombination modifier in a genotype by inverting the order of loci. For instance, the haplotype *c*-A-B is referred to as AB, and the haplotype C-A-B is referred to as BA. We refer to haplotypes with the C allele as ‘inverted’ haplotypes and haplotypes with the *c* allele as ‘standard’ haplotypes. There are thus  $2^3=8$  possible combinations of alleles that can be found on any chromosome (i.e. haplotypes): *c*AB, *c*Ab, *ca*B, *cab*, CAB, CAb, CaB, and Cab. In each population  $p \in \{1, 2\}$ , the frequency of each haplotype  $i \in \{1, \dots, 8\}$  of alleles is referred to as  $x_{p,i}$ .

Each individual carries two copies of the chromosome. As a result, our system is characterized by  $8 \times 8 = 64$  genotypes. In each population  $p \in \{1, 2\}$ , the frequency of each genotype  $i \in \{1, \dots, 64\}$  is referred to as  $y_{p,i}$ . Assuming discrete generations, we follow the

evolution of genotype frequencies  $\mathbf{y}(t)$  within a finite population over time.  $\mathbf{y} = \{y_{p,j}\}$  is a vector consisting of 128 elements referring to the frequencies of the 64 genotypes present in newborn offspring in the two populations. Haplotype frequencies  $x_{p,i}$  are easily calculated from  $\mathbf{y}(t)$ .

## 2.2 | Recursion equations

We run numerical simulations involving migration, selection, reproduction (including stochastic sampling and mutation), and census, in that order. To verify that the life cycle has no strong effects on the results, we also run simulations with different life cycles involving these processes in different orders in supplementary analyses (Figures S2 and S3).

### 2.2.1 | Migration

The life cycle begins with symmetric migration between populations at rate  $m$ , which describes the proportion of each population that consists of migrants right after the migration event. Thus, the frequency of genotype  $j$  in population  $p$  is

$$y_{p,j}^* = (1 - m)y_{p,j} + my_{k,j}$$

where  $k=2$  when  $p=1$ , and  $k=1$  when  $p=2$ .

### 2.2.2 | Selection

Migration is followed by natural selection in each population. The relative fitness associated with genotype  $j$  in population  $p$  is  $w_{p,j} \in [0, 1]$ . After natural selection, genotypic frequencies in population  $p$  are therefore:

$$y_{p,j}^{**} = \frac{w_{p,j} y_{p,j}^*}{\sum_{j'} w_{p,j'} y_{p,j'}^*}$$

The nature of the fitness landscape in each population (parameter values  $w_{p,j}$ ) is detailed in Section 2.3.

### 2.2.3 | Recombination, mating, and zygote formation

At this point, gametes are produced. Depending on the genotype at the modifier locus C, recombination between loci A and B may occur during gamete production. The recombination rate between loci A and B is equal to 0 for a genotype Cc at the C locus, and is equal to  $r$  for genotypes CC and cc. Many supergenes are formed by very large inversions spanning regions of several centimorgans (e.g. dozens of megabases in many animals). We were thus interested in the distance between A and B being of the scale of Mb or tens of Mb. In each population,  $r \in [0, 0.5]$  and mating is random.

As the allele C represents a chromosomal inversion capturing loci A and B, we assume that there is no recombination among loci C and A. Overall, the assumption of total physical linkage involving the recombination modifier and one of the other loci, as well as the assumption of recombination suppression induced by heterozygotes at the recombination modifier, makes that our model captures the main properties of an inversion and leads to the same recombination landscape as in Kirkpatrick and Barton (2006). Therefore, our model captures the main properties of an inversion, which are different from that of more classic recombination modifiers found in the literature (Barton, 1995; Nei, 1967; Otto & Barton, 1997). In the new generation, expected genotype frequencies  $y_{pj}^{\text{exp}}$  of zygotes are therefore calculated by summing the appropriate mating frequencies, assuming Mendelian segregation and accounting for recombination between loci A and B. As part of the reproduction step, we account for stochastic sampling and mutations, as described below.

### 2.2.4 | Stochastic sampling and mutations

It is necessary to account for the effect of genetic drift, because in deterministic models without drift, the disfavored haplotypes may reach frequencies closer and closer to zero, without actually reaching it (such as in Kirkpatrick & Barton, 2006). A new vector  $\mathbf{y}(t + 1)$  is therefore obtained by randomly sampling  $N$  offspring individuals from the expected genotype frequencies  $y_{pj}^{\text{exp}}$  of zygotes following a multinomial probability distribution.  $N$  therefore represents the population size. Stochasticity allows alleles to be totally lost (or conversely, fixed) in the population instead of remaining at extremely low but non-zero frequency even when they are disfavored, which is unrealistic when population sizes are finite. As we will see later, this feature proves crucial for the formation of supergenes. In each offspring individual, we assume that mutation occurs at a rate  $\mu$  at loci A and B involved in local adaptation. We also assume that at locus C (the recombination modifier with the same key properties as those of inversions), allele  $c$  can mutate to allele  $C$  at a rate  $\mu_{\text{inv}}$ . This reflects the rate of inversion occurrence. We assume unidirectional mutation at locus C (no reversion possible) since inversions are unlikely to affect the exact same portion of the genome twice.

Inversions that partially overlap a previously established inversion but that do not have the same breakpoints are unlikely to restore recombination since it would result in a complex reshuffling of gene order (as discussed in Jay, Tezenas, et al., 2022). However, we considered that both  $ab$  and  $AB$  haplotypes can be inverted (i.e.  $C\text{-}A\text{-}B$  and  $C\text{-}a\text{-}b$  can co-occur), and that these haplotypes recombine normally (i.e. at a rate  $r$ ).

## 2.3 | Simulation experiments

### 2.3.1 | Fitness landscape

In each population  $p$ , each genotype  $j$  is associated with its fitness  $w_{pj}$  due to selection imposed by the environment (leading to local adaptation involving the evolution at the A and B loci) and intrinsic cost associated with chromosomal rearrangements. Let us first focus on the case where allele C is absent and when fitness is only determined by genotypes at the A and B loci due to selection imposed by the environment. In that case, the fitness landscape is characterized in both environments by two peaks of fitness reached by genotype  $AB/AB$  and  $ab/ab$ . The main peak is reached by genotype  $AB/AB$  in population 1 and by genotype  $ab/ab$  in population 2, and the second peak is reached by genotype  $ab/ab$  in population 1 and by genotype  $AB/AB$  in population 2 (Figure 1). Therefore, there is divergent selection between the two populations (leading to local adaptation at the A and B loci). We consider a symmetric fitness landscape, such as

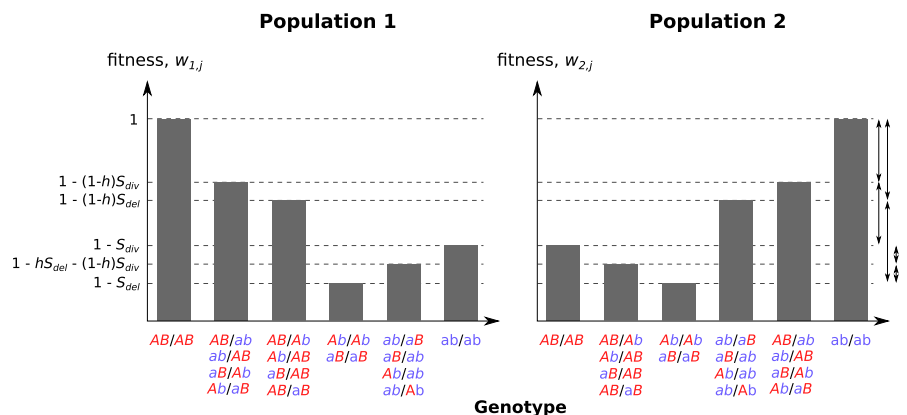
$$w_{1,AB/AB} = w_{2,ab/ab} = 1$$

and

$$w_{1,ab/ab} = w_{2,AB/AB} = 1 - S_{\text{div}}$$

where  $w_{p,AB/AB}$  (resp.  $w_{p,ab/ab}$ ) corresponds to the fitness associated with the genotype  $AB/AB$  (resp.  $ab/ab$ ) in population  $p$ , and where  $S_{\text{div}}$  reflects the fitness difference between the two local peaks (Figure 1). We refer to the best haplotype in a population as the ‘local’ haplotype ( $AB$  in population 1 and  $ab$  in population 2), and the haplotype associated with the second fitness peak as the ‘immigrant’ haplotype, that is, the haplotype that is favoured in the

**FIGURE 1** Genotype-to-fitness map in the two populations. In this example, we assume partial dominance at all loci ( $h=0.5$ ), and no recombination modifier (i.e. without allele C). We thus represent the fitness of genotypes at the A and B loci. Arrows highlight the partial dominance of genotypes involving at least one locus at heterozygous state (for a given genotype, the lengths of the two arrows are the same because we assume partial dominance with  $h=0.5$ ).



neighbouring population, even if this haplotype can eventually be maintained in the local population without gene flow. For instance, *ab* is considered as the ‘immigrant’ haplotype in population 1. We referred to *Ab* and *aB* haplotypes as ‘recombinant’ haplotypes, and we assume that they are associated with a fitness ‘valley’, so that there is fitness epistasis. Indeed, we assume that the genotypes resulting from the association of two copies of these haplotypes are associated with a low fitness, so that

$$w_{p,aB/aB} = w_{p,Ab/Ab} = 1 - S_{\text{del}}$$

where  $S_{\text{del}}$  reflects the reduction in fitness within the fitness ‘valley.’ We implement fitness landscapes such that  $0 < 1 - S_{\text{del}} < 1 - S_{\text{div}} < 1$  and, for simplicity, we consider that the fitness valley is the same in both populations (Figure 1).

For heterozygous genotypes, we assume that dominant alleles (alleles A and B, i.e. represented by capital letters) are expressed proportionally to a dominance factor  $h \in [0.5, 1]$  (see Figure 1 in the case of partial dominance, i.e.  $h=0.5$ ). For instance, the fitness associated with genotype *Ab/AB* in environment *p* is equal to

$$w_{p,Ab/AB} = w_{p,Ab/Ab} + h(w_{p,AB/AB} - w_{p,Ab/Ab})$$

We refer to the *AB* haplotype as ‘dominant’ haplotype (the locally adapted haplotype in population 1), and to the *ab* haplotype as ‘recessive’ haplotype (the locally adapted in population 2).

### 2.3.2 | Fitness cost associated to CC homozygotes

The allele C at the C locus may correspond to rearrangements like chromosomal inversions that may be associated with recessive fitness costs, as often observed, notably at supergenes. Inversions may be associated with such a recessive fitness cost because they carry recessive deleterious mutations, which generate associative overdominance (Faria et al., 2019; Jay et al., 2021; Jay, Tezenas, et al., 2022; Kirkpatrick, 2010), or because their breakpoints disrupt important genes (Küpper et al., 2016; Villoutreix et al., 2021). Therefore, we assume that the fitness of homozygotes CC is multiplied by a factor  $1 - \gamma_{\text{inv}}$ . For instance,

$$w_{p,BA/BA} = w_{p,AB/AB} \times (1 - \gamma_{\text{inv}})$$

Parameter  $\gamma_{\text{inv}} \in [0, 1]$  reflects the fitness cost in homozygotes at the C locus. For  $\gamma_{\text{inv}} = 0$ , CC homozygotes do not incur a fitness cost. For  $\gamma_{\text{inv}} = 1$ , CC homozygotes incur a strong fitness cost; they all die during the selection process. Since it affects only homozygotes, this fitness cost generates overdominance: at the C locus, heterozygotes perform better overall than homozygotes. We refer hereafter to this inversion fitness cost as inversion recessive load. Although we have not modelled explicitly deleterious mutations, this parameter  $\gamma_{\text{inv}}$  captures the effects of the expression of recessive deleterious

mutations on the fitness associated with the inversions (Berdan et al., 2021; Jay, Tezenas, et al., 2022). We acknowledge here that some inversions do seem to be underdominant in nature, notably because rare recombination events occurring between differentially oriented segments seem to cause unbalanced gene dosage (Faria et al., 2019; Wellenreuther & Bernatchez, 2018). Nevertheless, such a type of inversion is unexpected to be associated with local polymorphism and consequently is out of the scope of this study (Wellenreuther & Bernatchez, 2018).

### 2.3.3 | Initialization

Simulations start with only *AB* and *ab* homozygotes carrying alleles *c* at the C locus in the two populations, that is, chromosomal inversions are absent initially. The simulation rapidly reaches an equilibrium (typically, in fewer than 10 generations), so that the starting frequency has no effect on simulation outcome. We therefore started all simulations with the frequency of *ab* and *AB* being each 50%.

### 2.3.4 | Simulations

Each simulation is divided into two phases, each lasting 5000 generations. During the first phase, migration occurs at a rate  $m$ . Then, during the second phase, there is no migration between the two populations, that is,  $m=0$ . Those phases are referred to as ‘migratory phase’ and ‘non-migratory phase,’ respectively. The objective of the non-migratory phase is to estimate to what extent the maintenance of inversion polymorphism formed during the migratory phase relies on gene flow. Thereby, we aimed to study inversion polymorphisms not only maintained by migration–selection balance but also by balancing selection in isolated populations, thereby forming supergenes.

Table 1 sums up parameters notations and provides the list of values implemented for each parameter in our simulations. We performed 100 simulations for each of the 54,432 combinations of parameter values tested (Table 1). We also tested different mutation and inversion rates ( $10^{-8}$ ,  $10^{-7}$ ,  $10^{-6}$ ,  $10^{-5}$ ,  $10^{-4}$ ) and found that they have little influence on the simulations, as long as mutations and inversions occur several times during the simulation times. We therefore show the results with a mutation rate at the A and B loci equal to  $10^{-6}$  (‘mutation rate’ in Table 1) and a mutation rate at the C locus equal to  $10^{-6}$  (called ‘inversion rate,’ as the recombination modifier locus C has the same properties as an inversion). This means that a recombination modifier allele (i.e. a chromosomal inversion) is expected to arise around 50 times in a population of size  $N=10,000$  during 5000 generations. In addition, we implemented different population sizes ( $10^3$ ,  $10^4$ ,  $10^5$ ) and also found that it had little influence on the results, provided it was not infinite. We therefore set the population size to  $N=10,000$ . All simulations are performed using Julia (version 1.9.4).

TABLE 1 Summary of parameters implemented in simulations.

Parameter	Definition	Values
$m$	Migration between populations during the migratory phase	{0.05, 0.10, 0.15}
$r$	Recombination rate between loci A and B	{0.0, 0.1, 0.2, 0.3, 0.4, 0.5}
$\gamma_{inv}$	Inversion recessive load	{0.00, 0.05, ..., 0.95, 1.00}
$S_{div}$	Difference in fitness between local and immigrant genotypes	{0.2, 0.3, ..., 0.8, 0.9}
$S_{del}$	Difference in fitness between local and recombinant genotypes	{0.0, 0.1, 0.2, 0.3, 0.4, 0.5}
$h$	Dominance level	{0.50, 0.75, 1.00}
$\mu_{inv}$	Inversion rate (i.e. one-way mutation rate from allele $c$ to $C$ )	$10^{-6}$
$N$	Population size	10,000
$u$	Mutation rate at ecological loci (A and B)	$10^{-6}$

### 3 | RESULTS

Conducting simulations under a wide range of parameter values, we observed seven distinct scenarios regarding the fate of newly evolved chromosomal inversions (Figure 2; Figures S1–S3; the 7th scenario is shown only in Figure S1). Inversions may not spread at all (Figure 2b). When an inversion spreads during the migratory phase and carries either the dominant or the recessive haplotype, it can reach fixation (Figure 2c,d) or be maintained at intermediate frequency (Figure 2e–g). If migration is stopped, this polymorphism can either be lost (Figure 2e,f) or be maintained stably (Figure 2g). In this latter case, the polymorphism involves an inverted local haplotype and a standard (i.e. non-inverted) immigrant haplotype and is thus controlled by a supergene. Simulations conducted using different life cycles resulted in the same scenarios in similar proportions (Figures S2 and S3).

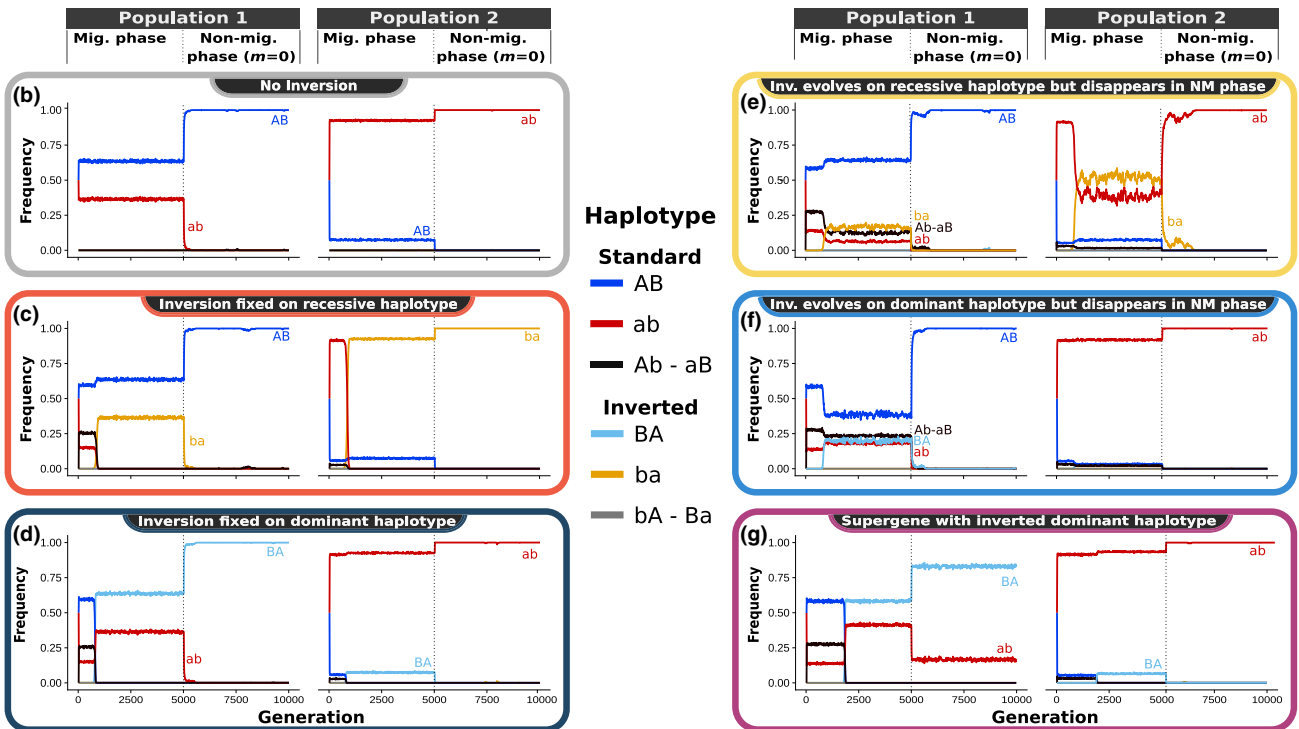
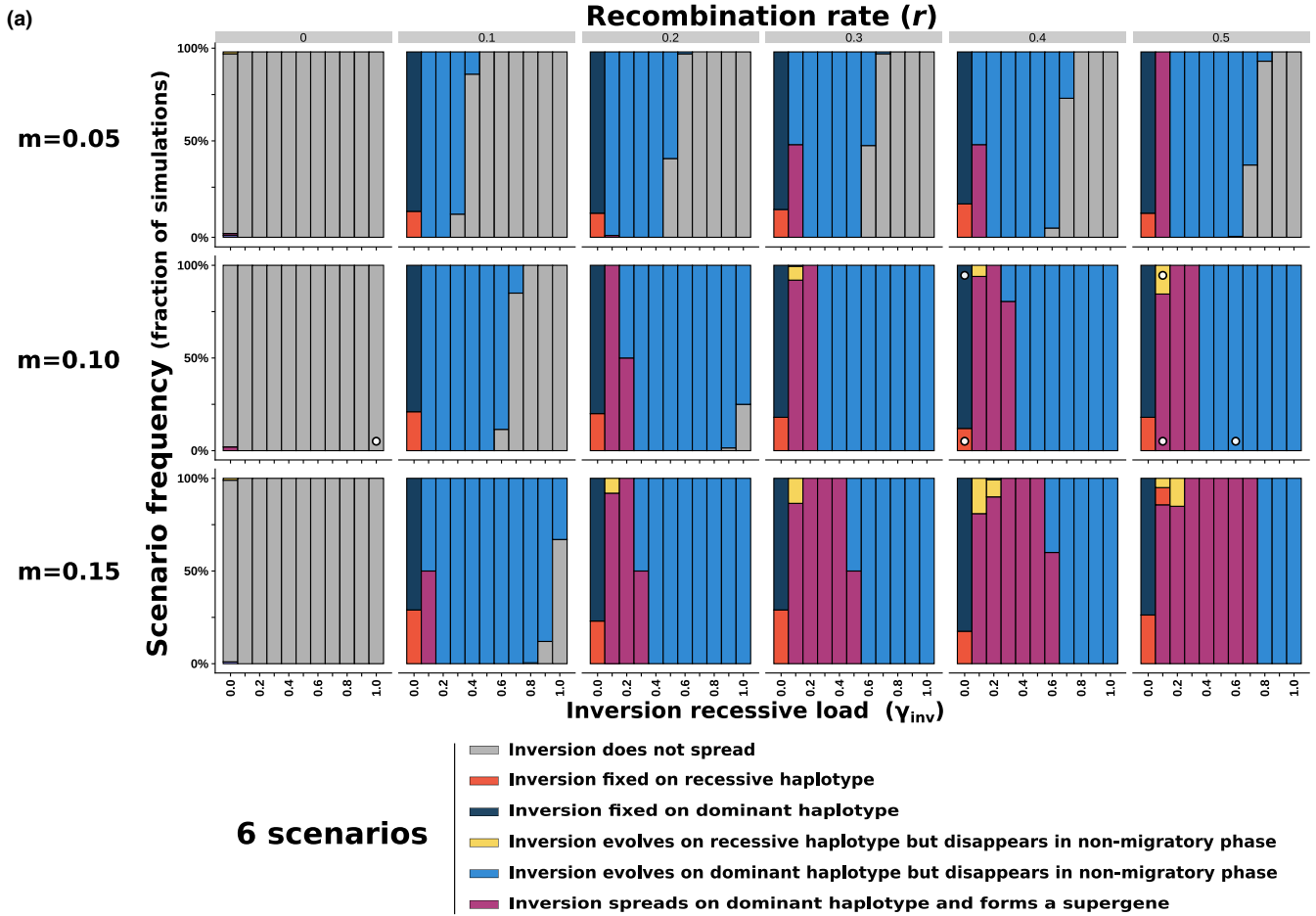
#### 3.1 | Conditions for the spread of an inversion

In simulations where loci A and B never recombine, inversions do not associate with any benefit, and therefore rarely spread (Figure 2a,b). When recombination is low and the recessive load carried by inversions is high, the low benefit of reducing recombination is offset by the cost brought by the recessive load, and inversions do not spread (Figure 2a). Finally, a low fitness difference between local and immigrant haplotypes leads the same alleles to fixate in both populations ('gene swamping,' Figure S4), suppressing the benefit associated with reduced recombination between loci and thus inhibiting the spread of inversions. Overall, as expected (Kirkpatrick & Barton, 2006; Yeaman, 2013), inversions reducing the recombination between epistatic alleles can easily spread in the population when the benefit associated with suppressed recombination among captured loci is higher than the cost brought by the recessive load (all scenarios that are not coloured in grey in Figure 2a).

#### 3.2 | Condition for the fixation of an inversion

Natural selection can indirectly favour the spread of inversions. Inversions suppress recombination and thereby reduce the fraction of recombinant genotypes with low fitness (Figure 3b). This effect is observed in the two populations, irrespective of the haplotype captured by the inversions (Figures S5 and S6). When inversions capture two epistatic alleles at recombining loci, inverted haplotypes rise in frequency and tend to replace standard (i.e. non-inverted) haplotypes (Figures 2 and 3). Nevertheless, standard local and immigrant haplotypes are not affected in the same way by the spread of inversions (Figure 3a–c). Indeed, before inversion spread, immigrant haplotypes segregate at lower frequency than local haplotypes in populations, and consequently they recombine more (they are more often paired with alternative haplotypes). Therefore, and perhaps surprisingly, the rise in frequency of inversions carrying locally adapted haplotypes (e.g. BA) leads to an increase in frequency of immigrant, maladapted haplotypes (e.g. ab) because it decreases the frequency of haplotypes they recombine with (notably the standard local haplotypes, e.g. AB). In other words, the inverted haplotype BA outcompetes the non-inverted haplotype AB because it does not recombine with haplotype ab, and the decreased frequency of the non-inverted haplotype AB indirectly benefits the immigrant haplotype ab by lowering the frequency of haplotypes it can recombine with. The frequency of standard immigrant haplotypes therefore increases when inversions spread (Figures 2c–g and 3c), and this occurs irrespective of immigrant fitness (as long as there is no gene swamping, Figure S7). Conversely, standard local haplotypes (e.g. AB) do not benefit from this effect and are not constantly brought by migration, and thus tend to disappear when the inversion reaches a high frequency (Figures 2c–g and 3a).

Overall, when inversions are associated with an absence of recessive load ( $\gamma_{inv}$ ), they tend to replace their standard counterparts and segregate with standard immigrant haplotypes at migration–selection balance during the migratory phase (Figure 2c,d). When migration stops, immigrant haplotypes are rapidly eliminated by



**FIGURE 2** The evolution of chromosomal inversions and consequences for the formation of supergenes. Simulations lead to six possible scenarios, depending on migration rate, recombination rate, and the strength of the inversion recessive load. (a) Effect of inversion recessive load, recombination rate, and migration rate on the frequency of occurrence of each scenario. Bar height represents the frequency of each scenario among 100 simulations for each combination of parameter values implemented. Other parameter values used  $h=1.0$ ,  $S_{div}=0.5$ ,  $S_{del}=0.8$ . White dots indicate parameters used for each scenario plot (b–g). The results of 19,800 simulations are displayed in total. For clarity, we plot the effect of inversion recessive load by intervals of 0.1. (b) Inversions do not spread. (c) Inversions fix, associated with recessive haplotypes (in population 2). (d) Inversions fix, associated with dominant haplotypes (in population 1). (e) Inversions evolve in association with recessive haplotypes, segregate at intermediate frequency with standard dominant and recessive haplotypes during the migratory phase, but disappear during the non-migratory phase. (f) Inversions evolve in association with dominant haplotypes, segregate at intermediate frequency with both standard haplotypes during the migratory phase, but disappear during the non-migratory phase. (g) Inversions evolve in association with dominant haplotypes and segregate at intermediate frequency with only the standard recessive haplotypes in both migratory and non-migratory phases. The evolutionary dynamics shown in panels b–g are characterized by abrupt changes in equilibrium states. This is because the inversion rate is low, and when an inversion occurs, it changes the equilibrium state.

selection and inversions reach fixation (Figure 2c,d). The situation is rather different if the inversion associates with an intermediate recessive load, as detailed below.

### 3.3 | The fate of overdominant inversions

When inversions associate with an intermediate recessive load ( $\gamma_{inv}$ ), they behave as overdominant and cannot fix in the populations. Recessive deleterious effects expressed by homozygous inversion genotypes indeed generate a homozygous disadvantage, which translates into negative frequency-dependent selection acting on inversions. A recessive load generates a cost that is low when inversions are rare but that becomes high when inversions reach higher frequencies. Therefore, increasing the inversion load  $\gamma_{inv}$  does not strongly affect the probability of spread of inversions (Figure S8) but determines their frequency at equilibrium (Figure S10; Nei & Roychoudhury, 1973). As a consequence, the inversion load  $\gamma_{inv}$  strongly determines whether inverted haplotypes replace their non-inverted counterparts and form supergenes or disappear when the migration ceases (Figure 2a; Figure S11).

The overdominant behaviour of inversions makes them prone to form a polymorphism with standard haplotypes and under certain conditions, this polymorphism persists in non-migratory phase and can be considered as forming a supergene. Two major types of behaviours can indeed be described when overdominant inversions evolve:

#### 3.3.1 | When overdominant inversions do not replace their standard counterparts

As described above, the rise of inverted haplotypes leads to a decrease in the frequency of their standard counterpart and an increase in the frequency of the immigrant haplotypes (Figures 2c–g and 3). When inversions remain at low frequency, they do not replace their standard counterparts and this leads to a coexistence of inverted local, standard local, and standard immigrant haplotypes during the migratory phase (Figures 2e,f and 3). During the non-migratory phase, immigrant haplotypes are rapidly lost, so inversions no longer

bring a fitness advantage and are impaired by their recessive fitness loads. Consequently, when inverted (e.g. BA) and standard (e.g. AB) haplotypes segregate at end of the migratory phase, the standard form always fixes in the population as soon as migration is stopped and inversions are lost (Figures 2e,f and 3; Figure S9). Notably, this outcome prevails in all deterministic simulations without genetic drift (i.e. without stochastic sampling; not shown), as long as inverted haplotypes can spread initially. This is because in the absence of drift, a haplotype cannot be lost; they just remain at extremely low frequency (and therefore can spread back if the migration or selection regime changes).

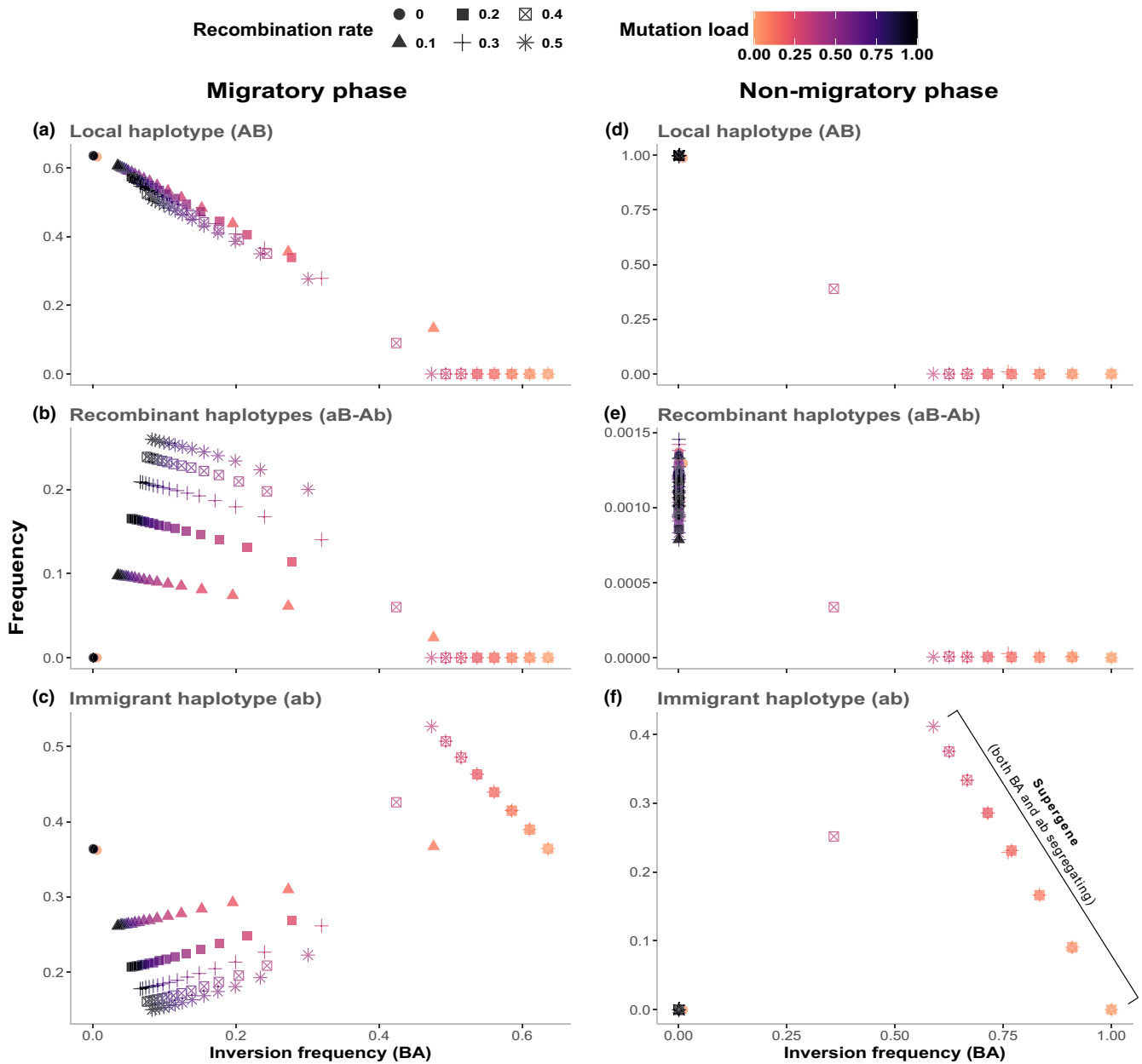
#### 3.3.2 | When overdominant inversions replace their standard counterparts

The rise of inversions (e.g. BA) at high frequency may lead to the crash of their standard equivalent (e.g. AB) but not of the standard immigrant haplotypes (e.g. ab). When inversions are overdominant, however, they are extremely unlikely to fix in the population during the non-migratory phase because of homozygote disadvantage (Figure S10). Therefore, when overdominant inversions have replaced their standard counterpart and segregate only with standard immigrant haplotypes at the end of the migratory phase, this polymorphism is stably maintained during the non-migratory phase (Figures 2g and 3; Figure S9). Since this polymorphism is maintained by balancing selection and is composed of non-recombining haplotypes coding for alternative phenotypes, the genetic architecture underlying this polymorphism can be considered a supergene.

Overall, the fate of an overdominant inverted haplotype when migration is stopped is therefore determined by whether its standard equivalent was lost or not during the migratory phase.

### 3.4 | The determinants of overdominant inversion equilibrium frequency

Our simulations account for a wide diversity of demographic scenarios, dominance hierarchies, and selective landscapes. We



**FIGURE 3** Consequences of the evolution of inversions on the frequencies of standard haplotypes. Plots show how the spread of inversions capturing the dominant haplotype (BA) in population 1 affects the frequency of other haplotypes for parameter values:  $m=0.1$ ,  $S_{del}=0.8$ ,  $S_{div}=0.5$ ,  $h=1.0$ . Each dot represents the average state of simulations for a given combination of parameter values (126 combinations in total, 100 simulations *per* combination) at the end of the migratory phase (a–c) and non-migratory phase (d–f), enough time for the equilibrium state to be reached. Only simulations where inversions spread in population 1 are shown (12,389 simulations among 12,600 for the combinations of parameters displayed). (a) Effect of the spread of inverted haplotypes on the frequency of standard local haplotypes at the end of the migratory phase. (b) Effect of the spread of inverted haplotypes on the frequency of recombinant haplotypes at the end of the migratory phase. (c) Effect of the spread of inverted haplotypes on the frequency of standard immigrant haplotype at the end of the migratory phase. (d) Effect of the spread of inverted haplotypes on the frequency of standard local haplotypes at the end of the non-migratory phase. (e) Effect of the spread of inverted haplotypes on the frequency of recombinant haplotypes at the end of the non-migratory phase. (f) Effect of the spread of inverted haplotypes on the frequency of standard immigrant haplotype at the end of the non-migratory phase. It shows that the more frequent is the inverted haplotype (BA) at the end of the migratory phase, the less frequent are its non-inverted counterpart (AB) and the recombinant haplotypes, as expected. It also shows that, in most cases, the more frequent is the inverted haplotype, the more frequent is the immigrant haplotype, that is that the spread of the inversion is usually associated with an increase of the immigrant haplotype frequency. This can also be observed in [Figure 2c–g](#) and [Figure S7](#). The relationship between haplotype frequencies is non-linear because each haplotype frequency cannot decrease below 0, thereby affecting other haplotype frequencies. Supergenes are formed and stably maintained during the non-migratory phase (simulations where the frequencies of BA and ab are not equal to zero on panel f) only when the local standard haplotype (AB) disappears during the migratory phase. Conditions for this to happen are shown in panel a (low-to-moderate inversion load and non-null initial recombination rate between loci A and B).

highlight here the factors determining the frequency of overdominant inversions in a migration–selection equilibrium and therefore the persistence of polymorphisms during the non-migratory phase.

First, the frequency reached by an inversion during the migratory phase strongly depends on the dominance of the haplotype captured by the inversion. Since heterozygotes have a higher fitness when dominance is strong (with a fitness closer to the dominant fitness peak), inversions reach higher frequencies when they capture dominant haplotypes (Figure S11). Recessive inversions are less likely to reach fixation than dominant inversions (for  $\gamma_{inv} = 0$ , Figure 2a), and overdominant inversions never reach high frequency or form a protected polymorphism when capturing recessive haplotypes (Figure 2a; Figures S2 and S11). The spread of an inversion capturing one haplotype also inhibits the spread of an inversion capturing another haplotype (Figure S12). Therefore, the dynamics of a simulation depend on which inversion spreads first, and this explains why, when the fitness load is low (e.g.  $\gamma_{inv} = 0.1$ ), the formation of supergenes appears slightly less likely (Figure 2a): by chance, an inversion capturing the recessive haplotype arises and spreads first, hampering the spread of other inversions capturing the dominant haplotype (Figure 2e).

The major determinants of inversion frequency are the strength of their associated recessive load ( $\gamma_{inv}$ ), the migration rate ( $m$ ), and the recombination rate ( $r$ ) (Figure 2a; Figures S1 and S10). Inversions are under indirect selection because they decrease the frequency of recombinant haplotypes between local and immigrant haplotypes and therefore inversions spread at higher frequency under high migration and recombination rates (Figure S6). The recessive load generating a fitness cost on inversions is expressed more often as they increase in frequency (because of the increasing frequency of homozygotes) and therefore determines their equilibrium frequency (Figure S10). This leads to a dramatic change in the frequency of standard and inverted haplotypes when this load increases (Figures 2 and 3; Figure S10). The interaction between recombination rate, migration rate, and recessive load thus ultimately determines whether an overdominant inversion is maintained in a protected polymorphism or disappears during the non-migratory phase.

When the recessive load is low relative to the recombination rate, inversions can spread easily (Figure 2a; Figures S1 and S2). When the recessive load is high relative to the recombination rate, the inverted haplotype never reaches a sufficient frequency to replace standard haplotypes, and thus the inversion is lost during the non-migratory phase (i.e. it is only maintained by migration; Figure 1f). By contrast, for intermediate recessive load, inversions can spread and drift can lead to the loss of the standard haplotype, and to the formation of a supergene (as depicted in Figures 2a and 3).

Interestingly, the fitness costs associated with the recombinant or immigrant genotypes appear as poor determinants of the frequency of inversions and the formation of supergenes (Figures S13 and S14). This suggests that supergenes can be formed by haplotypes with large variations in fitness components and under a wide diversity of fitness landscapes.

## 4 | DISCUSSION

Using a two-island population genetic model, we explored how a recessive fitness load associated with a chromosomal inversion influences the fate of the inversion and the formation of a protected polymorphism, maintaining multiple loci in full linkage and the co-variation of multiple phenotypic differences. We show that recombination and the recessive fitness load have opposing effects on the frequency of inversion, which leads to the formation of a supergene when these forces are balanced. Our results, showing the effect of migration and recombination rates on the invasion of a chromosomal inversion, are consistent with Kirkpatrick and Barton's (2006) analytical derivations, but by tracking the frequencies of all haplotypes in two diploid populations under different levels of (over)dominance, epistasis, and recombination, we were also able to determine how the invasion of inversions affects the frequencies of standard (i.e. non-inverted) haplotypes. Perhaps surprisingly, the spread of inversions does not only decrease the frequencies of local standard and recombinant haplotypes but also increase the frequency of immigrant haplotypes irrespective of their local fitness and dominance levels. When the initial recombination rate between epistatic loci is high and the recessive fitness cost of inversions is moderate, the spread of inversions is associated with not only the loss of local standard haplotypes but also the maintenance of locally maladapted, immigrant haplotypes. This leads to the formation of a supergene characterized by the maintenance of a polymorphism with a locally adapted inverted haplotype and a locally maladapted standard haplotype, even though these haplotypes display large differences in their degree of adaptation to the focal environment.

Here, we restricted our definition of supergenes to those cases of discrete polymorphisms that can be maintained in the absence of gene flow. This strict definition allowed us to untangle the forces contributing to the emergence of inversion polymorphisms within populations. In natural populations, however, it is difficult to precisely assess the effect of gene flow among other factors. Yet we believe it is useful to recognize the differences between polymorphisms resulting from an influx of migrants from those owing to frequency-dependent processes within populations. As shown in Figures 2 and 3, a continuum of situations can be observed, characterized by variations in the importance of migration in the maintenance of polymorphism. We consider the term supergene is most useful to describe the genetic architectures which facilitate the maintenance of polymorphism and the segregation of discrete forms within a population (Thompson & Jiggins, 2014), as in most of the classic cases on distyly, mimicry, shell morphologies, or behavioural syndromes (Gonzalez et al., 2019; Iijima et al., 2018; Joron et al., 2011; Kunte et al., 2014; Lamichane et al., 2016; Li et al., 2016; Tuttle et al., 2016). Those tend to be cases with known or suspected epistasis among linked loci as well as some level of overdominance and represent one end of the spectrum predicted by our model (case represented in Figure 2f). Other well-described cases of inversion polymorphisms such as those distributed along clines or ecotones (Ayala et al., 2014; Durmaz et al., 2018; Twyford

& Friedman, 2015) might represent cases at the other end of the spectrum, where gene flow plays a large role in the maintenance of a high level of polymorphism (cases represented in Figure 2c,d). Whether these inversions controlling ecotypes distributed along environmental gradients also tend to harbour recessive fitness cost that may contribute to their maintenance in the face of varying migration levels still remains to be elucidated. The major point made here is, however, that the evolution of supergenes maintained by local balancing selection owing to overdominant inversions may be facilitated by spatial segregation of selection.

Our model explicitly includes overdominant inversions, as it seems to be widespread in nature, and notably stands as a general explanation for supergene inheritance. For instance, in the social chromosome of the fire ant *Solenopsis invicta* (Wang et al., 2013) or in the supergene controlling plumage colour and mating behaviour in the white-throated sparrow *Zonotrichia albicollis* (Tuttle et al., 2016), inversion homozygotes are rarely found among reproductive adults in nature. Likewise, in the butterfly *Heliconius numata*, inversion homozygotes at the supergene controlling mimicry polymorphism display a very low larval survival (Jay et al., 2021) and are rare in nature (Chouteau et al., 2017). Overdominance of inversions may have two major causes. First, inversion breakpoints may directly generate deleterious effects, for instance, when they disrupt a gene (Küpper et al., 2016; Villoutreix et al., 2021). Second, chromosomal inversions may harbour recessive deleterious mutations (Berdan et al., 2021; Connallon & Olito, 2021; Jay et al., 2021; Kirkpatrick & Barton, 2006). This recessive mutation load causes homozygous disadvantage (Frydenberg, 1963; Kirkpatrick, 2010), which is silent when inversions are rare but becomes expressed as they reach higher frequencies and are more likely to form homozygotes, therefore promoting polymorphism (Berdan et al., 2021; Jay et al., 2021; Jay, Tezenas, et al., 2022; Kirkpatrick & Barton, 2006; Mérot, Llaurens, et al., 2020; Wellenreuther & Bernatchez, 2018). Such a load may result from the fact that inversions capture unique haplotypes, which by chance may include several deleterious recessive mutations that segregate in the population at mutation–selection–drift equilibrium (Jay, Tezenas, et al., 2022). Despite having on average the same number of deleterious mutations upon formation, standard haplotypes all carry distinct deleterious mutations, whereas inverted haplotypes all harbour the same set of deleterious mutations. Consequently, in contrast to standard haplotypes, when inverted haplotypes rise in frequency and form homozygous, they express all of their deleterious mutations simultaneously (Jay, Tezenas, et al., 2022). Since recombination is strongly reduced between differentially oriented segments, captured deleterious variants in inversions can hardly be purged.

In addition, since inverted segments do not recombine with standard segments, genomic regions with inversion polymorphism display reduced effective population size, which lowers the efficiency of selection and may favour the accumulation of further deleterious variants by drift (Berdan et al., 2021; Glémin, 2003; Jay, Tezenas, et al., 2022). Furthermore, inversions at low frequency are rarely homozygous (i.e. they have low opportunity of recombination) and

therefore tend to behave as asexual populations and accumulate deleterious variants by drift in a Muller's ratchet-like process (Berdan et al., 2021; Jay, Tezenas, et al., 2022). Taken together, these mechanisms probably explain why many inversions, notably those forming supergenes, behave as overdominant (Jay et al., 2021; Jay, Tezenas, et al., 2022; Wellenreuther & Bernatchez, 2018). It may be noted that because suppression of recombination leads to an accumulation of deleterious mutations, the fitness of alternative haplotypes at the supergene is expected to decrease with time, which might eventually lead to the loss of one of them (Berdan et al., 2021; Charlesworth, 2023; Jay, Tezenas, et al., 2022). The long-term maintenance of supergene therefore probably involves the evolution of further mechanisms maintaining the polymorphism in face of mutation accumulation. This could, for instance, implicate the evolution of disassortative mating (Maisonneuve et al., 2021), as in the case of the colour-pattern supergene in *H. numata*. The long-term maintenance of a supergene in the face of reversion (reverse inversion) and the accumulation of deleterious mutations merit further theoretical studies accounting for these additional mechanisms.

Our model shows that balancing selection due to inversion overdominance can favour the spread and the maintenance of supergene alleles with strong differences in fitness components. More specifically, an allele that is not locally favoured can be maintained alongside a locally adapted allele if the latter associates with a strong homozygous disadvantage. This may reflect the observation that many supergenes are characterized by the maintenance of several phenotypes associated with strong fitness differences. For instance, multiple-queen fire ant colonies seem to outcompete single-queen colonies in many aspects of their ecology but both strategies are maintained in nature (Ross & Keller, 1995). Likewise, mimetic and non-mimetic butterfly morphs coexist in several *Papilio* species while only mimetic morphs seem to benefit from reduced predation (Palmer et al., 2018). Supergenes are often assumed to be maintained because environmental factors generate alternative adaptive optima that are associated with equal fitness consequences (Schwander et al., 2014; Thompson & Jiggins, 2014). Our results suggest that supergenes could instead evolve because balancing selection caused by intrinsic features of chromosomal inversions leads to the maintenance of alternative haplotypes even if they bring highly unequal benefits in terms of local adaptation. We modelled the inversion overdominance as a parameter, and it is therefore not an emerging property due to the expression of deleterious recessive mutations. Implementing explicitly deleterious recessive mutations and investigating how they determine the inversion overdominance, and therefore the emergence of supergene architectures, could be an interesting avenue for future theoretical research.

Our model may apply in a wide diversity of demographic conditions and, as such, it may offer a parsimonious explanation to the puzzle posed by the formation of supergene involving highly differentiated alternative haplotypes. Considering separate taxa evolving in different conditions, as observed everywhere in nature, differentiated haplotypes could easily evolve and be affected by chromosomal rearrangements. We show that gene flow between these taxa

could result in the formation of supergenes and indeed, several supergenes are found in species that show evidence of gene flow with other species (e.g. Castric et al., 2008; Dixon et al., 2019; Dudek et al., 2019; Jay et al., 2018; Sun et al., 2012; Zhang et al., 2017), such as in the white-throated sparrow (Tuttle et al., 2016) and the fire ants (Helleu et al., 2022). However, this scenario does not imply that these taxa are categorized as distinct species and the mechanism proposed here can readily apply to differentiated populations from the same species. During the past decade, progress in sequencing technologies has revealed that gene flow between populations showing local adaptation is common in nature, including between species (Mallet et al., 2016). This could explain why structural polymorphisms are widespread across the tree of life (Mérot, Oomen, et al., 2020; Wellenreuther & Bernatchez, 2018) and indeed, several known supergenes display genomic signals suggesting that gene flow may have played a role in their formation (Dixon et al., 2019; Huang et al., 2020; Jay et al., 2018; Tuttle et al., 2016).

In summary, the recombination load resulting from migration between differentiated populations is expected to foster the spread of recombination modifiers such as chromosomal inversions (Kirkpatrick & Barton, 2006). Nevertheless, these recombination modifiers are prone to be associated with recessive fitness loads, which may translate into balancing selection hampering their fixation (Jay, Tezenas, et al., 2022). As shown by others, this fosters the maintenance of polymorphism (Berdan et al., 2021; Connallon et al., 2018; Connallon & Olito, 2021; Jay, Tezenas, et al., 2022). We show here that when the recombination load due to gene flow balances the inversion fitness load, the spread of an inversion may lead to the loss of the locally adapted standard haplotype (associated with a high recombination load) but not of the immigrant haplotype. Therefore, it results in the maintenance of differentially adapted, non-recombining haplotypes within a population, that is, to the formation of a supergene, even when gene flow ceases. Taken together, our results provide a realistic scenario for the evolution of supergenes and inversion polymorphisms in general and bring new light to the importance of gene flow in the formation of new genetic architectures.

#### AUTHOR CONTRIBUTIONS

PJ and TA conceived the study. MJ supervised the study. PJ and TA designed the models. PJ performed the simulations and created the figures. All authors analysed the results. All authors wrote, reviewed, and edited the manuscript.

#### ACKNOWLEDGEMENTS

This research was supported by grants from the Swiss National Science Foundation and the National Science Foundation (to T.G.A.), from an Agence Nationale de la Recherche Grant acronym Supergene (ANR-18-CE02-0019-01) (to M.J.) and from a Human Frontier Science Program (HFSP) fellowship (to P.J.).

#### CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

#### DATA AVAILABILITY STATEMENT

The R and Julia scripts used for this study are available from Zenodo (<https://doi.org/10.5281/zenodo.10526748>).

#### ORCID

Paul Jay  <https://orcid.org/0000-0001-5979-1263>

Thomas G. Aubier  <https://orcid.org/0000-0001-8543-5596>

Mathieu Joron  <https://orcid.org/0000-0003-1043-4147>

#### REFERENCES

- Abbott, J. K., Nordén, A. K., & Hansson, B. (2017). Sex chromosome evolution: Historical insights and future perspectives. *Proceedings of the Royal Society B: Biological Sciences*, 284(1854), 20162806. <https://doi.org/10.1098/rspb.2016.2806>
- Ayala, D., Ullastres, A., & González, J. (2014). Adaptation through chromosomal inversions in *Anopheles*. *Frontiers in Genetics*, 5, 129.
- Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J. C., & The Tree of Sex Consortium. (2014). Sex determination: Why so many ways of doing it? *PLoS Biology*, 12(7), e1001899. <https://doi.org/10.1371/journal.pbio.1001899>
- Barton, N. H. (1995). A general model for the evolution of recombination. *Genetical Research*, 65(2), 123–145. <https://doi.org/10.1017/s001672300033140>
- Berdan, E. L., Barton, N. H., Butlin, R., Charlesworth, B., Faria, R., Fragata, I., Gilbert, K. J., Jay, P., Kapun, M., Lotterhos, K. E., Mérot, C., Durmaz Mitchell, E., Pascual, M., Peichel, C. L., Rafajlović, M., Westram, A. M., Schaeffer, S. W., Johannesson, K., & Flatt, T. (2023). How chromosomal inversions reorient the evolutionary process. *Journal of Evolutionary Biology*, 36(12), 1761–1782. <https://doi.org/10.1111/jeb.14242>
- Berdan, E. L., Blanckaert, A., Butlin, R. K., & Bank, C. (2021). Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genetics*, 17(3), e1009411. <https://doi.org/10.1371/journal.pgen.1009411>
- Billiard, S., López-Villavicencio, M., Hood, M. E., & Giraud, T. (2012). Sex, outcrossing and mating types: Unsolved questions in fungi and beyond. *Journal of Evolutionary Biology*, 25(6), 1020–1038. <https://doi.org/10.1111/j.1420-9101.2012.02495.x>
- Castric, V., Bechsgaard, J., Schierup, M. H., & Vekemans, X. (2008). Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genetics*, 4(8), e1000168. <https://doi.org/10.1371/journal.pgen.1000168>
- Charlesworth, B. (2023). The effects of inversion polymorphisms on patterns of neutral genetic diversity (p. 2023.02.23.529778). *bioRxiv*. <https://doi.org/10.1101/2023.02.23.529778>
- Charlesworth, D., & Charlesworth, B. (1975). Theoretical genetics of Batesian mimicry II. Evolution of supergenes. *Journal of Theoretical Biology*, 55(2), 305–324. [https://doi.org/10.1016/s0022-5193\(75\)80082-8](https://doi.org/10.1016/s0022-5193(75)80082-8)
- Chouteau, M., Llaurens, V., Piron-Prunier, F., & Joron, M. (2017). Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures. *Proceedings of the National Academy of Sciences*, 114, 8325–8329. <https://doi.org/10.1073/pnas.1702482114>
- Connallon, T., & Olito, C. (2021). Natural selection and the distribution of chromosomal inversion lengths. *Molecular Ecology*, 31, 3627–3641. <https://doi.org/10.1111/mec.16091>
- Connallon, T., Olito, C., Dutoit, L., Papoli, H., Ruzicka, F., & Yong, L. (2018). Local adaptation and the evolution of inversions on sex chromosomes and autosomes. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 373(1757), 20170423. <https://doi.org/10.1098/rstb.2017.0423>

- Corcoran, P., Anderson, J. L., Jacobson, D. J., Sun, Y., Ni, P., Lascoux, M., & Johannesson, H. (2016). Introgression maintains the genetic integrity of the mating-type determining chromosome of the fungus *Neurospora tetrasperma*. *Genome Research*, 26(4), 486–498. <https://doi.org/10.1101/gr.197244.115>
- Curtsinger, J. W., Service, P. M., & Prout, T. (1994). Antagonistic pleiotropy, reversal of dominance, and genetic polymorphism. *The American Naturalist*, 144(2), 210–228.
- Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., Zimin, A. V., Hughes, D. S. T., Ferguson, L. C., Martin, S. H., Salazar, C., Lewis, J. J., Adler, S., Ahn, S.-J., Baker, D. A., Baxter, S. W., Chamberlain, N. L., Chauhan, R., Counterman, B. A., ... The Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405), 94–98. <https://doi.org/10.1038/nature11041>
- Dixon, G., Kitano, J., & Kirkpatrick, M. (2019). The origin of a new sex chromosome by introgression between two stickleback fishes. *Molecular Biology and Evolution*, 36(1), 28–38. <https://doi.org/10.1093/molbev/msy181>
- Dudek, K., Gaczorek, T. S., Zieliński, P., & Babik, W. (2019). Massive introgression of major histocompatibility complex (MHC) genes in newt hybrid zones. *Molecular Ecology*, 28(21), 4798–4810. <https://doi.org/10.1111/mec.15254>
- Durmaz, E., Benson, C., Kapun, M., Schmidt, P., & Flatt, T. (2018). An inversion supergene in *Drosophila* underpins latitudinal clines in survival traits. *Journal of Evolutionary Biology*, 31(9), 1354–1364. <https://doi.org/10.1111/jeb.13310>
- Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving inversions. *Trends in Ecology & Evolution*, 34(3), 239–248. <https://doi.org/10.1016/j.tree.2018.12.005>
- Frydenberg, O. (1963). Population studies of a lethal mutant in *Drosophila melanogaster*. *Hereditas*, 50(1), 89–116. <https://doi.org/10.1111/j.1601-5223.1963.tb01896.x>
- Glémin, S. (2003). How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution; International Journal of Organic Evolution*, 57(12), 2678–2687. <https://doi.org/10.1111/j.0014-3820.2003.tb01512.x>
- Gonzalez, D. R., Aramendia, A. C., & Davison, A. (2019). Recombination within the *Cepaea nemoralis* supergene is confounded by incomplete penetrance and epistasis. *Heredity*, 123(2), 153–161. <https://doi.org/10.1038/s41437-019-0190-6>
- Hager, E. R., Harringmeyer, O. S., Wooldridge, T. B., Theingi, S., Gable, J. T., McFadden, S., Neugeboren, B., Turner, K. M., Jensen, J. D., & Hoekstra, H. E. (2022). A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science*, 377(6604), 399–405. <https://doi.org/10.1126/science.abg0718>
- Hartmann, F. E., Duhamel, M., Carpentier, F., Hood, M. E., Foulongne-Oriol, M., Silar, P., Malagnac, F., Grognet, P., & Giraud, T. (2021). Recombination suppression and evolutionary strata around mating-type loci in fungi: Documenting patterns and understanding evolutionary and mechanistic causes. *New Phytologist*, 229(5), 2470–2491. <https://doi.org/10.1111/nph.17039>
- Helleu, Q., Roux, C., Ross, K. G., & Keller, L. (2022). Radiation and hybridization underpin the spread of the fire ant social supergene. *Proceedings of the National Academy of Sciences*, 119(34), e2201040119. <https://doi.org/10.1073/pnas.2201040119>
- Hessenaue, P., Fijarczyk, A., Martin, H., Prunier, J., Charron, G., Chapuis, J., Bernier, L., Tanguay, P., Hamelin, R. C., & Landry, C. R. (2020). Hybridization and introgression drive genome evolution of Dutch elm disease pathogens. *Nature Ecology & Evolution*, 4(4), 626–638. <https://doi.org/10.1038/s41559-020-1133-6>
- Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, 29(14), 2535–2549. <https://doi.org/10.1111/mec.15428>
- Iijima, T., Kajitani, R., Komata, S., Lin, C.-P., Sota, T., Itoh, T., & Fujiwara, H. (2018). Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*. *Science Advances*, 4(4), ea05416. <https://doi.org/10.1126/sciadv.aao5416>
- Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., & Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics*, 53(3), 288–293. <https://doi.org/10.1038/s41588-020-00771-1>
- Jay, P., Leroy, M., Le Poul, Y., Whibley, A., Arias, M., Chouteau, M., & Joron, M. (2022). Association mapping of colour variation in a butterfly provides evidence that a supergene locks together a cluster of adaptive loci. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 377(1856), 20210193. <https://doi.org/10.1098/rstb.2021.0193>
- Jay, P., Tezenas, E., Véber, A., & Giraud, T. (2022). Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. *PLoS Biology*, 20(7), e3001698. <https://doi.org/10.1371/journal.pbio.3001698>
- Jay, P., Whibley, A., Frézal, L., Rodriguez de Cara, M. Á., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. *Current Biology*, 28(11), 1839–1845.e3. <https://doi.org/10.1016/j.cub.2018.04.072>
- Johannesson, K., Le Moan, A., Perini, S., & André, C. (2020). A Darwinian laboratory of multiple contact zones. *Trends in Ecology & Evolution*, 35(11), 1021–1036. <https://doi.org/10.1016/j.tree.2020.07.015>
- Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, 502(7469), 93–95. <https://doi.org/10.1038/nature12489>
- Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., Whibley, A., Becuwe, M., Baxter, S. W., Ferguson, L., Wilkinson, P. A., Salazar, C., Davidson, C., Clark, R., Quail, M. A., Beasley, H., Glithero, R., Lloyd, C., Sims, S., ... ffrench-Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, 477(7363), 203–206. <https://doi.org/10.1038/nature10341>
- Kappel, C., Huu, C. N., & Lenhard, M. (2017). A short story gets longer: Recent insights into the molecular basis of heterostyly. *Journal of Experimental Botany*, 68(21-22), 5719–5730. <https://doi.org/10.1093/jxb/erx387>
- Kaufman, J., Jacob, J., Shaw, J., Walker, B., Milne, S., Beck, S., & Salomonsen, J. (1999). Gene organisation determines evolution of function in the chicken MHC. *Immunological Reviews*, 167(1), 101–117. <https://doi.org/10.1111/j.1600-065X.1999.tb01385.x>
- Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, 8(9), e1000501. <https://doi.org/10.1371/journal.pbio.1000501>
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1), 419–434. <https://doi.org/10.1534/genetics.105.047985>
- Knief, U., Forstmeier, W., Pei, Y., Ihle, M., Wang, D., Martin, K., Opatová, P., Albrechtová, J., Wittig, M., Franke, A., Albrecht, T., & Kempaenaers, B. (2017). A sex-chromosome inversion causes strong overdominance for sperm traits that affect siring success. *Nature Ecology & Evolution*, 1(8), 1177–1184. <https://doi.org/10.1038/s41559-017-0236-1>
- Kopp, M., & Hermisson, J. (2006). The evolution of genetic architecture under frequency-dependent disruptive selection. *Evolution; International Journal of Organic Evolution*, 60(8), 1537–1550.
- Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., Mullen, S. P., & Kronforst, M. R. (2014). *Doublesex* is a mimicry supergene. *Nature*, 507(7491), 229–232. <https://doi.org/10.1038/nature13112>

- Küpper, C., Stocks, M., Risse, J. E., dos Remedios, N., Farrell, L. L., McRae, S. B., Morgan, T. C., Karlionova, N., Pinchuk, P., Verkuil, Y. I., Kitaysky, A. S., Wingfield, J. C., Piersma, T., Zeng, K., Slate, J., Blaxter, M., Lank, D. B., & Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, 48(1), 79–83. <https://doi.org/10.1038/ng.3443>
- Lamichhane, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoepfner, M. P., Kerje, S., Gustafson, U., Shi, C., Zhang, H., Chen, W., Liang, X., Huang, L., Wang, J., Liang, E., Wu, Q., Lee, S. M.-Y., Xu, X., Höglund, J., ... Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics*, 48(1), 84–88. <https://doi.org/10.1038/ng.3430>
- Lenormand, T., & Otto, S. P. (2000). The evolution of recombination in a heterogeneous environment. *Genetics*, 156(1), 423–438. <https://doi.org/10.1093/genetics/156.1.423>
- Leroy, T., Roux, C., Villate, L., Bodénès, C., Romiguier, J., Paiva, J. A. P., Dossat, C., Aury, J.-M., Plomion, C., & Kremer, A. (2017). Extensive recent secondary contacts between four European white oak species. *New Phytologist*, 214(2), 865–878. <https://doi.org/10.1111/nph.14413>
- Li, J., Cocker, J. M., Wright, J., Webster, M. A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., van Oosterhout, C., & Gilmartin, P. M. (2016). Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nature Plants*, 2(12), 1–7. <https://doi.org/10.1038/nplants.2016.188>
- Llaurens, V., Whibley, A., & Joron, M. (2017). Genetic architecture and balancing selection: The life and death of differentiated variants. *Molecular Ecology*, 26(9), 2430–2448. <https://doi.org/10.1111/mec.14051>
- Maisonneuve, L., Chouteau, M., Joron, M., & Llaurens, V. (2021). Evolution and genetic architecture of disassortative mating at a locus under heterozygote advantage. *Evolution*, 75(1), 149–165. <https://doi.org/10.1111/evo.14129>
- Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays*, 38(2), 140–149. <https://doi.org/10.1002/bies.201500149>
- Mérot, C., Llaurens, V., Normandeau, E., Bernatchez, L., & Wellenreuther, M. (2020). Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nature Communications*, 11(1), 670. <https://doi.org/10.1038/s41467-020-14479-7>
- Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, 35(7), 561–572. <https://doi.org/10.1016/j.tree.2020.03.002>
- Murray, J., & Clarke, B. (1976). Supergenes in polymorphic land snails. *Heredity*, 37(2), 271–282. <https://doi.org/10.1038/hdy.1976.87>
- Nei, M. (1967). Modification of linkage intensity by natural selection. *Genetics*, 57(3), 625–641.
- Nei, M., Kojima, K.-I., & Schaffer, H. E. (1967). Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics*, 57(4), 741–750.
- Nei, M., & Roychoudhury, A. K. (1973). Probability of fixation and mean fixation time of an overdominant mutation. *Genetics*, 74(2), 371–380.
- Otto, S. P., & Barton, N. H. (1997). The evolution of recombination: Removing the limits to natural selection. *Genetics*, 147(2), 879–906.
- Palmer, D. H., Tan, Y. Q., Finkbeiner, S. D., Briscoe, A. D., Monteiro, A., & Kronforst, M. R. (2018). Experimental field tests of Batesian mimicry in the swallowtail butterfly *Papilio polytes*. *Ecology and Evolution*, 8(15), 7657–7666. <https://doi.org/10.1002/ece3.4207>
- Pinho, C., & Hey, J. (2010). Divergence with gene flow: Models and data. *Annual Review of Ecology, Evolution, and Systematics*, 41, 215–230. <https://doi.org/10.1146/annurev-ecolsys-102209-144644>
- Raeymaekers, J. A. M., Chaturvedi, A., Hablützel, P. I., Verdonck, I., Hellemans, B., Maes, G. E., de Meester, L., & Volckaert, F. A. M. (2017). Adaptive and non-adaptive divergence in a common landscape. *Nature Communications*, 8(1), 267. <https://doi.org/10.1038/s41467-017-00256-6>
- Roff, D. A., & Fairbairn, D. J. (2007). The evolution of trade-offs: Where are we? *Journal of Evolutionary Biology*, 20(2), 433–447. <https://doi.org/10.1111/j.1420-9101.2006.01255.x>
- Ross, K. G., & Keller, L. (1995). Ecology and evolution of social organization: Insights from fire ants and other highly eusocial insects. *Annual Review of Ecology and Systematics*, 26, 631–656.
- Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, 14(12), e2000234. <https://doi.org/10.1371/journal.pbio.2000234>
- Santos, M. (1986). The role of genic selection in the establishment of inversion polymorphism in *Drosophila subobscura*. *Genetica*, 69(1), 35–45. <https://doi.org/10.1007/BF00122932>
- Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, 24(7), R288–R294. <https://doi.org/10.1016/j.cub.2014.01.056>
- Sinervo, B., & Lively, C. M. (1996). The rock-paper-scissors game and the evolution of alternative male strategies. *Nature*, 380(6571), 240–243. <https://doi.org/10.1038/380240a0>
- Stolle, E., Pracana, R., López-Osorio, F., Priebe, M. K., Hernández, G. L., Castillo-Carrillo, C., Arias, M. C., Paris, C. I., Bollazzi, M., Priyam, A., & Wurm, Y. (2022). Recurring adaptive introgression of a supergene variant that determines social organization. *Nature Communications*, 13(1), 1180. <https://doi.org/10.1038/s41467-022-28806-7>
- Sun, Y., Corcoran, P., Menkis, A., Whittle, C. A., Andersson, S. G. E., & Johannesson, H. (2012). Large-scale introgression shapes the evolution of the mating-type chromosomes of the filamentous ascomycete *Neurospora tetrasperma*. *PLoS Genetics*, 8(7), e1002820. <https://doi.org/10.1371/journal.pgen.1002820>
- Tafreshi, A. G., Otto, S. P., & Chapuisat, M. (2022). Unbalanced selection: The challenge of maintaining a social polymorphism when a supergene is selfish. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 377(1856), 20210197. <https://doi.org/10.1098/rstb.2021.0197>
- Takayama, S., & Isogai, A. (2005). Self-incompatibility in plants. *Annual Review of Plant Biology*, 56, 467–489. <https://doi.org/10.1146/annurev.ev.arplant.56.032604.144249>
- The MHC Sequencing Consortium. (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature*, 401(6756), 921–923. <https://doi.org/10.1038/44853>
- Thompson, M. J., & Jiggins, C. D. (2014). Supergenes and their role in evolution. *Heredity*, 113(1), 1–8. <https://doi.org/10.1038/hdy.2014.20>
- Trowsdale, J., & Knight, J. C. (2013). Major histocompatibility complex genomics and human disease. *Annual Review of Genomics and Human Genetics*, 14, 301–323. <https://doi.org/10.1146/annurev-genom-091212-153455>
- Turner, J. R. G. (1967). On supergenes. I. The evolution of supergenes. *The American Naturalist*, 101(919), 195–221.
- Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., Stager, M., Betuel, A., Cheviron, Z. A., Warren, W. C., Gonser, R. A., & Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, 26(3), 344–350. <https://doi.org/10.1016/j.cub.2015.11.069>
- Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution*, 69(6), 1476–1486. <https://doi.org/10.1111/evo.12663>
- Villoutreix, R., Ayala, D., Joron, M., Gompert, Z., Feder, J. L., & Nosil, P. (2021). Inversion breakpoints and the evolution of supergenes. *Molecular Ecology*, 30(12), 2738–2755. <https://doi.org/10.1111/mec.15907>

- Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, 493(7434), 664–668. <https://doi.org/10.1038/nature11832>
- Weinreich, D. M., & Chao, L. (2005). Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution; International Journal of Organic Evolution*, 59(6), 1175–1182.
- Weissman, D. B., Feldman, M. W., & Fisher, D. S. (2010). The rate of fitness-valley crossing in sexual populations. *Genetics*, 186(4), 1389–1410. <https://doi.org/10.1534/genetics.110.123240>
- Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, 33(6), 427–440. <https://doi.org/10.1016/j.tree.2018.04.002>
- Whitlock, M. C. (1997). Founder effects and peak shifts without genetic drift: Adaptive peak shifts occur easily when environments fluctuate slightly. *Evolution*, 51(4), 1044–1048. <https://doi.org/10.1111/j.1558-5646.1997.tb03951.x>
- Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the XI International Congress of Genetics*, 8, 209–222.
- Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences*, 110(19), E1743–E1751. <https://doi.org/10.1073/pnas.1219381110>
- Zhang, W., Westerman, E., Nitzany, E., Palmer, S., & Kronforst, M. R. (2017). Tracing the origin and evolution of supergene mimicry in butterflies. *Nature Communications*, 8(1), 1269. <https://doi.org/10.1038/s41467-017-01370-1>

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Jay, P., Aubier, T. G., & Joron, M. (2024). The interplay of local adaptation and gene flow may lead to the formation of supergenes. *Molecular Ecology*, 00, e17297. <https://doi.org/10.1111/mec.17297>