HAL
open science

A deleterious mutation-sheltering theory for the evolution of sex chromosomes and supergenes

Paul Jay, Emilie Tezenas, Tatiana E Giraud

To cite this version:

Paul Jay, Emilie Tezenas, Tatiana E Giraud. A deleterious mutation-sheltering theory for the evolution of sex chromosomes and supergenes. 2021. hal-03365741

HAL Id: hal-03365741

https://hal.science/hal-03365741v1

Preprint submitted on 5 Oct 2021

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A deleterious mutation-sheltering theory for the evolution of sex chromosomes and supergenes

Paul Jay[1*], Emilie Tezenas[1], Tatiana Giraud[1]

[1] Ecologie Systématique Evolution, Bâtiment 360, CNRS, AgroParisTech, Université Paris-Saclay, 91400 Orsay, France

*corresponding author: paul.yann.jay@gmail.com

Key words: sex chromosomes, evolutionary strata, mating-type chromosome, supergenes, sexual antagonism, inversions, model, deleterious mutation, recombination suppression expansion

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Many organisms have sex chromosomes with large non-recombining regions having expanded stepwise, the reason why being still poorly understood. Theories proposed so far rely on differences between sexes but are poorly supported by empirical data and cannot account for the stepwise suppression of recombination around sex-determining loci in organisms without sexual dimorphism. We show here, by mathematical modeling and stochastic simulations, that recombination suppression in sex chromosomes can evolve simply because it shelters recessive deleterious mutations, which are ubiquitous in genomes. The permanent heterozygosity of sex-determining alleles protects linked chromosomal inversions against expression of their recessive mutation load, leading to an accumulation of inversions around these loci, as observed in nature. We provide here a testable and widely applicable theory to explain the evolution of sex chromosomes and of supergenes in general.

Introduction

Sex chromosomes, mating-type chromosomes and supergenes in general are widespread in nature. These structures are defined by extensive regions of recombination suppression encompassing multiple genes, and they control iconic polymorphisms, such as sexual dimorphism or color polymorphism, in many organisms, including humans[1–5]. Supergenes, and sex chromosomes in particular, often display a stepwise extension of non-recombining regions, generating "evolutionary strata" of differentiation[4,6–10]. It is generally thought that the regions of recombination suppression on sex chromosomes gradually expand because selection favors the linkage of sex-determining genes to sexually antagonistic loci (i.e. with alleles advantageous in one sex but disadvantageous in the other[7,11,12]). However, in addition to theoretical issues with this idea[13], the responsibility of this mechanism for evolutionary strata in natural populations has never been demonstrated[6,14,15] (but see ref. 16). More critically, the gradual expansion of recombination suppression has been observed around many fungal mating-type loci, despite the lack of male or female functions and hence of sexual antagonism in these species [8,17], as well as around other supergenes[9,10]. Alternative explanations for the evolution of non-recombining sex chromosomes have been proposed, such as meiotic drive[18] and genetic drift[19], but these models are also subject to similar limitations: they apply only in specific conditions and are poorly supported by empirical data. It therefore seems likely that other mechanisms may underlie the evolution of sex chromosomes.

We propose here a new theory explaining sex chromosome and supergene evolution. This theory is based on the observation that genomes carry numerous deleterious recessive variants, as demonstrated by the pervasiveness of inbreeding depression in nature[20,21]. The basic idea is simple (Figure 1). Consider a diploid population experiencing only recessive deleterious mutations and, thus, harboring many variants at low frequency. Through the combined action of recombination, mutation, selection and drift, individuals vary in the number of deleterious variants carried genome-wide and within any particular genomic region. Chromosomal inversions may occur at any position, cover a region of any size and suppress recombination when heterozygous, thereby capturing a single combination of deleterious variants (i.e. a haplotype).

Inversions capturing fewer deleterious variants than the population average in this region have a fitness advantage and should, therefore, increase in frequency. Unlike non-inverted segments, inverted

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

segments cannot recombine with haplotypes carrying more deleterious mutations, and they therefore retain this advantage during increases in frequency. However, as the frequency of an inversion increases, homozygotes for this inversion become more common. Homozygotes are at a strong disadvantage due to the deleterious recessive variants carried by these inversions, and selection against homozygotes therefore prevents such inversions from reaching high frequencies. Now, consider an inversion that, by chance, is in perfect linkage with a permanently heterozygous allele, such as the male-determining allele in a XY system. If this sex-linked inversion captures fewer deleterious variants than the population average, it should increase in frequency without ever suffering the deleterious consequences of having its load expressed. The recessive deleterious mutations captured by the sex-linked inversion are indeed fully linked to the permanently heterozygous, male-determining allele, and will, therefore, never occur as homozygotes. They are therefore sheltered from selection. Hence, such an inversion would be expected to spread, becoming perfectly associated with the male-determining allele, resulting in the suppression of recombination between the X and Y chromosomes in the region covered by the inversion. The successive appearance of additional inversions linked to this Y-fixed inversion will cause the non-recombining region to further expand by the same process, thereby leading to the formation of a chromosome with a large non-recombining region around the sex-determining locus —i.e. a sex chromosome with evolutionary strata of differentiation. This process does not require sexual antagonism or drift. It occurs simply through the selection of non-recombining genetic segments with fewer deleterious variants than the population average, linked to a permanently heterozygous allele. This process is general, and should apply to any genomic region with permanently heterozygous alleles, including fungal mating-type chromosomes and some supergenes[8,22,23], and to any mechanism suppressing recombination in *cis*, not only inversions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

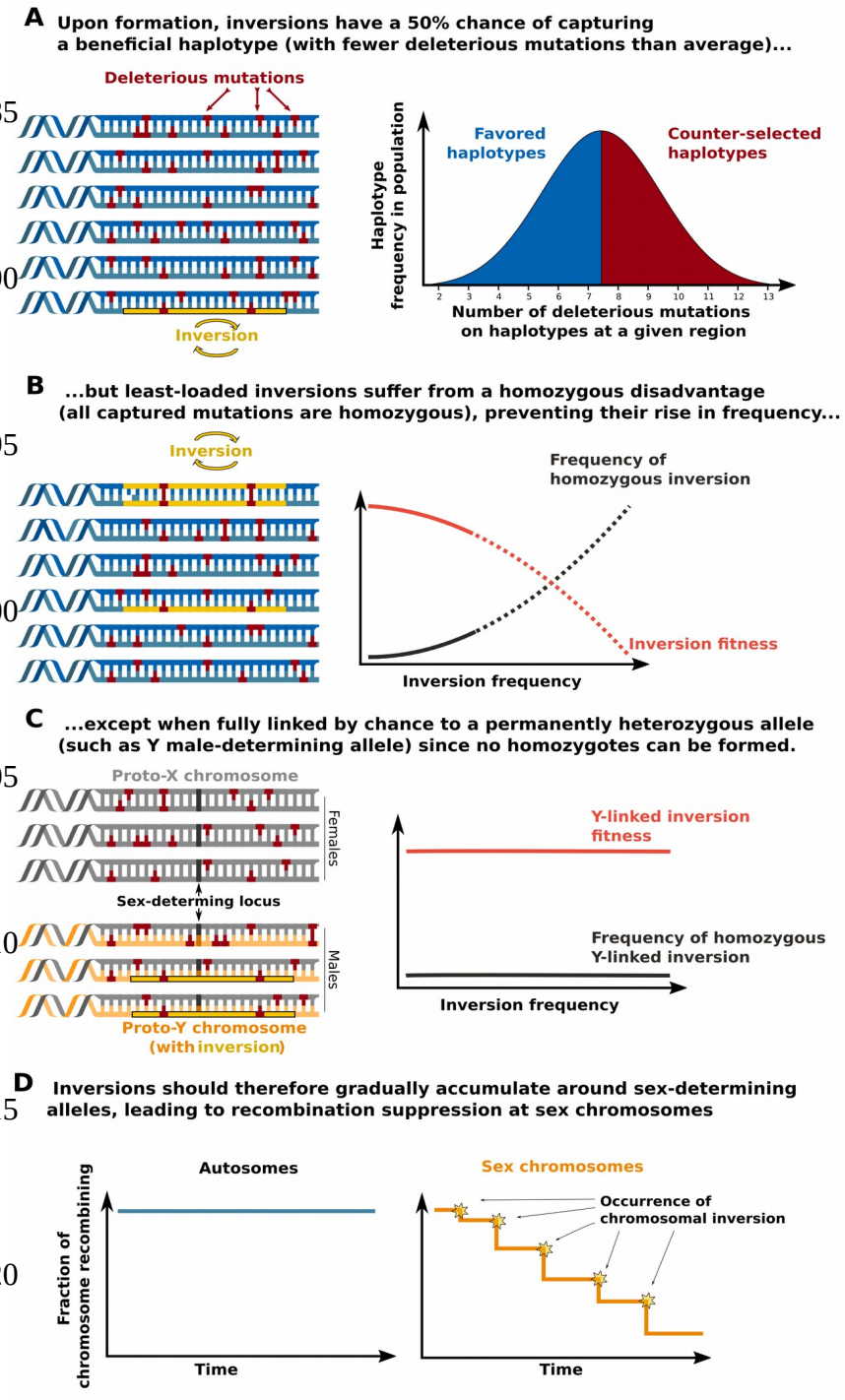

Figure 1 | Schematic presentation of the theory

A, Within any population and for any genomic region, diploid individuals (here represented by two strands of DNA) and haplotypes (i.e. combinations of mutations, here one strand of DNA) vary in terms of the number of recessive deleterious mutations they carry. About half the haplotypes have fewer deleterious mutations than the average and should be favored by selection. Chromosomal inversions therefore have a 50% chance of capturing beneficial haplotypes. **B,** The increase in frequency of a beneficial inversion is associated with an increase in the frequency of homozygous inversions, which have a fitness disadvantage because they are homozygous for all the deleterious recessive mutations carried by the inversion. The fitness of the inversion therefore decreases with increasing inversion frequency, maintaining the inversion to low frequencies. **C.** Permanent heterozygosity at a Y-chromosome sex-determining allele protects linked inversions from this disadvantage of homozygosity. Hence, beneficial inversions (carrying fewer deleterious mutations than average) should spread and become perfectly associated with the sex-determining allele. **D.** Unlike autosomes, on which such inversions cannot spread because of their homozygous disadvantage, inversions should accumulate on proto-Y chromosomes, leading to a suppression of recombination between the X and Y chromosomes. A similar mechanism should operate for any recombination suppressor acting in *cis* (not only chromosomal inversions) and any locus with permanently heterozygous alleles.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results and discussion

We formally tested this idea with infinite population deterministic models and individual-based simulations, the latter allowing an exploration of situations involving genetic drift. We modeled diploid populations experiencing only recessive deleterious mutations, with heterozygous and homozygous individuals suffering from a *1-hs* and *1-s* reduction in fitness, respectively. Mutations had multiplicative effects. Individuals were considered to have two pairs of chromosomes, one of which harbored a locus with at least one allele permanently or almost permanently heterozygous (see Methods and Data S1 for details). Several situations were considered, mimicking those encountered in XY sex-determination systems, in fungal mating-type systems, or in overdominant supergenes. We analyzed the evolution of recombination modifiers suppressing recombination across a fragment in which they also reside (i.e. *cis*-modifiers), either exclusively in heterozygotes (e.g. chromosomal inversions), or in both heterozygotes and homozygotes, (e.g. histone modifications). Each of these recombination modifiers appeared in a single sequence, and was thus in linkage disequilibrium with a random set of mutations, but was purely neutral in itself, such that its fitness depended exclusively on the number of deleterious alleles within the segment captured. Inversions (or recombination suppressors), denoted "*I*", appeared in a population with non-inverted segments (or without recombination suppressors) denoted "*N*". In stochastic simulations, inversions were introduced after allowing the population to evolve under mutation, selection and drift until an equilibrium in the number of mutations segregating within the genome was reached. We first describe the dynamics of an inversion fully linked to an XY sex-determining locus (males are XY and females are XX, the Y allele being permanently heterozygous), and we then consider other types of recombination modifiers and heterozygosity rules.

As noted by Nei (1967), inversions capturing fewer deleterious mutations than the population average should increase in frequency. Let q be the frequency of mutations at mutation-selection equilibrium in an infinite population. The number of mutations harbored by individuals over a large genomic segment of size n follows an approximate Gaussian distribution centered on nq. As this distribution is symmetric, half the inversions of a given size capture fewer deleterious mutations than the population average. We refer to such inversions hereafter as "least-loaded inversions".

If m is the number of mutations captured by the inversion, the mean fitness of individuals homozygous for the inversion (*II*), heterozygous for the inversion (*NI*) or lacking the inversion (*NN*) can be expressed as a function of n, q, h, s, and m, as follows:

$$\overline{W_{NN}} \approx e^{-nqsh-nqs(h+q(1-2h))}$$

$$\overline{W_{NI}} \approx e^{-nqsh-ms(h+q(1-2h))}$$

$$W_{II} \approx e^{-sm}$$

(see Methods and Data S1 for details). Once formed, inversions should increase in frequency if heterozygotes for the inversion are fitter than homozygotes without the inversion ($W_{NI} > W_{NN}$), which is true if $m<nq$, i.e. if the inversion carries fewer mutations than the population average. This deterministic result is readily found with stochastic simulations (Figures S1-3). Drift results in the frequencies of mutations fluctuating in finite populations, resulting in some mutations being rarer than others. We observed, in simulations, that inversions were also favored if they carried mutations that

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were rarer than average (Figures S1-3). Thus, once formed (i.e. when rare enough to never occur as homozygotes), about half of inversions occurring in genomes are beneficial, and this is true regardless of the selection coefficient (s) and dominance coefficient (h) of the mutations they capture.

The fitness advantage of the least-loaded inversions decreases rapidly with increasing frequency of the inversion. Indeed, inversions including at least one recessive deleterious mutation tend to suffer from a homozygous disadvantage (overdominance) such that $W_{NI} > W_{NN} > W_{II}$. The deterministic increase in frequency of an inversion (I) on an autosome or linkage to an XY sex-determining locus can easily be determined with a two-locus two-allele model. For instance, the change in frequency of an inversion on the proto-Y chromosome can be expressed as:

$$\Delta p_Y = \frac{p_Y(p_{Xf}W_{II} + (1-p_{Xf})W_{NI} - \overline{W_m}) \pm rDW_{NI}}{\overline{W_m}}$$

where p_{Xf} is the frequency of the inversion on proto-X chromosomes in females, *r* is the rate of recombination between the inversion and the sex-determining locus, *D* is their linkage disequilibrium and $\overline{W_m}$ is the mean male fitness (see Methods and Data S1 for details). Figure 2a and Figures S4-6 illustrate that inversions on autosomes (i.e. with *r*=0.5) or linked to the female sex-determining allele on the proto-X have a very low equilibrium frequency in infinite populations, even when they capture far fewer mutations than the population average. Inversions on autosomes or X chromosomes are, therefore, frequently favored when rare, but can never reach very high frequencies. They are, thus, often lost by drift in finite populations, as shown in stochastic simulations (Figures 2b-c and Figures S8-10). In the absence of linkage to a permanently heterozygous allele, only inversions capturing a mutation-free haplotype invaded populations (Figure 2b and Figure S8-10), as shown previously[24,25].

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

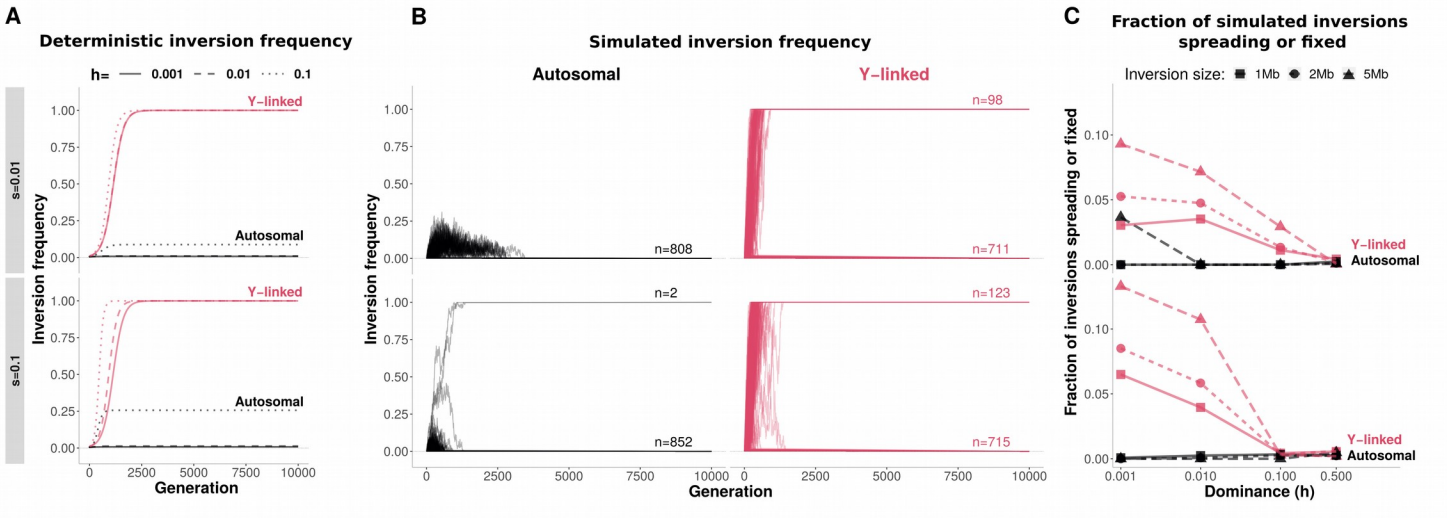

Figure 2 | Least-loaded inversions linked to the heterozygous sex-determining allele on the proto-Y chromosome spread to become perfectly associated with this allele

Deterministic and stochastic simulations of inversions fully linked or completely unlinked to an XY-like sex-determining locus (two alleles, one of which is permanently heterozygous). **A,** Deterministic change in inversion frequency, for the case of 2 Mb inversions fully linked or completely unlinked (i.e. on an autosome) to the Y-chromosome sex-determining allele. For clarity, the frequency displayed for Y-linked inversions is the frequency of inversions on the proto-Y chromosomes. The figure illustrates the case of inversions carrying a number of mutations 20% lower than the population average (*m=0.8*nq*), as frequently observed in simulations (Figures S1-S2). Mutations were considered to be at mutation-selection equilibrium frequencies with a mutation rate of 10^{-8}. Figures S4-7 illustrate the fate of X-linked inversions, with intermediate degrees of linkage to the Y male-determining allele or with various numbers of mutations. **B,** Change in inversion frequency in stochastic simulations of 1000 individuals experiencing only recessive deleterious mutations, with h=0.01. For each parameter combination, 10,000 inversions of 2 Mb were simulated, appearing on a random haplotype. Only inversions not lost after 20 generations are displayed. The figure displays the frequency of inversions on the autosomes and proto-Y chromosomes. The number of inversions fixed and lost at the end of the simulation are indicated above the lines ("n=..."). Overall, 13.4% of Y-linked inversions spread in the two cases illustrated here and became perfectly associated with the Y chromosome, whereas all but two inversions on the autosome were lost. Both of the inversions fixed on the autosome were free of mutations when they initially formed, whereas only eight out of the 221 Y-linked fixed inversions were initially mutation-free. The evolutionary trajectories of inversions of different sizes and with other parameter combinations are displayed in Figures S8-10. **C,** Proportions of the inversions in stochastic simulations that were still segregating or fixed after 10,000 generations, for different dominance coefficients and inversion sizes. Large inversions capturing highly recessive mutations were the most likely to escape drift and spread.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

By contrast, least-loaded inversions that are, by chance, fully linked to the permanently heterozygous sex-determining allele on the proto-Y chromosome spread until they become perfectly associated with this allele (Figure 2a and Figure S4-7). Through linkage to a permanently heterozygous allele, inversions avoid the consequences of their deleterious mutations being expressed in a homozygous state. Permanent heterozygosity effectively leads to directional selection for the least-loaded inversions, as confirmed by stochastic simulations (Figure 2b and Figure S8-10). Similar results were obtained when: i) two or more permanently heterozygous alleles segregated at the sex-determining-like locus, modeling plant self-incompatibility or fungal mating-type systems[8,26] (Figure S5-6,11-12); ii) the alleles were not permanently heterozygous but were strongly overdominant, therefore mostly occurring in the heterozygous state, as for several supergenes controlling color polymorphism[9,23,27] (Figure S13); or iii) the inversion was in strong but incomplete linkage with the permanently heterozygous allele (e.g. 0.1 cM away from the allele, Figures S4-7). Our model can thus account for the existence of inversions very close, but not fully linked to a mating-type locus, as reported in the chestnut blast fungus[28,29]. Similar results were also obtained when we considered recombination modifiers suppressing recombination even when homozygous, as for histone modifications, rather than solely when heterozygous, as for inversions (Figure S14). Provided that a region with multiple segregating mutations is considered, a propensity of neutral recombination modifiers to spread when linked to permanently heterozygous alleles was observed regardless of mutation rate, population size or the effect of mutations on fitness (Figure 2 and Figures S4-15). However, for the inversions to benefit from a heterozygote advantage, they must capture recessive mutations (i.e., with low dominance coefficients; Figure 2c and Figure S15). Furthermore, as for any variant, even inversions benefiting from a selective advantage can be lost by drift, the probability of such loss being inversely correlated with the relative fitness of the inversion, which depends on the number and type of mutations initially captured. Compared to inversions in mutation-poor region, inversions occuring in regions in which more mutations segregate have a wider fitness distribution, and are therefore more likely to benefit from a strong advantage. The probability of inversions being lost by drift hence decreases with increasing inversion size, mutation recessiveness and mutation rate (Figure 2c, Figure S15). In summary, because of their propensity to capture deleterious recessive variants, inversions (or any recombination suppressor) are much more likely to spread when tightly linked to a permanently heterozygous allele.

Inversions would therefore be expected to accumulate successively around sex-determining loci, forming large non-recombining regions, leading to the formation of sex chromosomes with a typical pattern of evolutionary strata. Figure 3 illustrates this process: we simulated, over 100,000 generations, two 100 Mb chromosomes, one of which carried a mammalian-type sex-determining locus (XY males and XX females). Individuals experienced only recessive deleterious mutations, with fitness effects similar to those observed in humans and occurring at similar rates. Each mutation dominance coefficient was chosen at random (see Methods for details). We allowed neutral inversions to occur at any position and at any time. Despite the complete absence of advantageous mutations, complex selection regimes, epistasis or particular demographic scenarios, the Y chromosome progressively stopped recombining with the X chromosome as it accumulated inversions linked to the male sex-determining allele. The non-recombining region thus extended around the sex-determining locus in a stepwise manner, due to the successive occurrence of chromosomal inversions on the Y chromosome. This situation perfectly reflects the evolution of sex chromosomes with evolutionary strata.

These results show that recombination suppression on sex chromosomes can evolve simply because genomes harbor many recessive deleterious variants. Our novel theory for the evolution of sex

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chromosomes, and supergenes generally, is based on simple and widespread phenomena: i) inversions (or any recombination suppressor acting in *cis*) may be favored solely because they contain fewer deleterious mutations than the population average, a situation occurring in about half the inversions formed; ii) such inversions tend to behave as overdominant (i.e. beneficial in the heterozygous state but with a homozygous disadvantage), which hampers them from reaching high frequencies in autosomes; iii) when linked, by chance, to a permanently heterozygous allele, inversions do not suffer from homozygous disadvantage, and they are therefore able to spread until they are perfectly associated with the permanently heterozygous allele. These three phenomena have been reported independently in several studies, but, to our knowledge, never in interaction (see references 24,25 for i, 30,31 for ii, and 8,32 for iii). Following the evolution of recombination suppression by inversions, further deleterious mutations should accumulate within the non-recombining regions, in particular due to Müller's ratchet-like processes, leading to the degeneration of Y chromosomes[33,34]. The degeneration of Y chromosomes has been well studied, but the mechanism proposed here implies that deleterious mutations not only follow, but also drive, recombination suppression.

Given its simplicity and wide scope of application, our theory of sex chromosome evolution is a powerful alternative to other explanations, although the various theories are not mutually exclusive. The strength of our model lies in the absence of specific assumptions, such as sexually antagonistic selection or small population sizes. This theory is instead based on the often overlooked observation that recessive deleterious mutations are widespread in genomes within natural populations, as shown by the ubiquitous nature of inbreeding depression[21]. Furthermore, our novel theory can also explain why some supergenes, such as those in butterflies and ants, display evolutionary strata[9,10], and why many fungal mating-type loci display a stepwise extension of non-recombining regions despite the absence of sexually antagonistic selection in these organisms[8,17,35]. Our model therefore provides a general and unified framework for understanding the evolution of non-recombining regions. More generally, our results provide a new framework for understanding the mechanisms underlying the maintenance of inversion polymorphisms in nature. We show that, because of their recessive mutation load, inversions are compelled to segregate at low frequency on autosomes and are prone to loss through drift. Under some conditions, they may nevertheless reach intermediate frequencies, if, for example, they capture, by chance, a segment with very few mutations relative to the population average (Figs. S8-10). This could explain why many inversions are polymorphic in nature, as reported for example in humans[36,37]. Consistent with this prediction, many studies have reported overdominant chromosomal rearrangements[22,27,38–41]. The propensity of inversions to display overdominance may also account for the evolution of some iconic polymorphisms in nature, as suggested by a recent study on the maintenance of a color polymorphism in a butterfly[9].

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

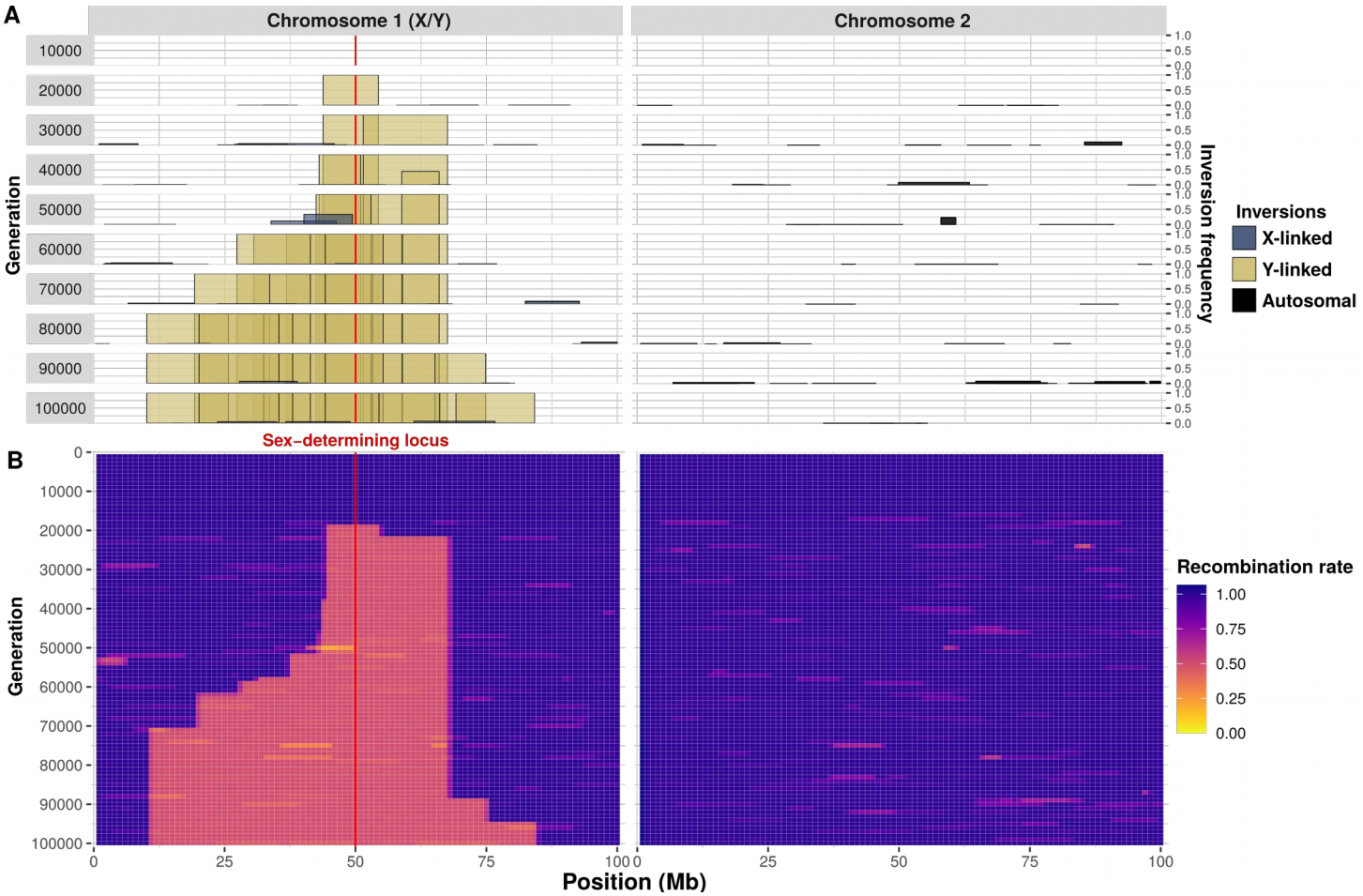

Figure 3 | Successive accumulation of inversions around a sex-determining locus leading to the formation of a non-recombining Y sex-chromosome.

A sampled simulation of *N*=1000 individuals bearing two pairs of 100 Mb chromosomes during 100,000 generations. Chromosome 1 harbors an X/Y sex-determining locus at 50 Mb (individuals are XX or XY). Each generation, an inversion of random size (maximum 20 Mb) appears in a randomly sampled individual, at a random position. **A,** Overview of chromosomal inversion frequency and position for 10 different generations. Square width represents inversion position and square height inversion frequency. Inversions appearing on the Y chromosome are depicted in yellow, those appearing on the X chromosomes are depicted in gray. The colors are not entirely opaque, so that regions with overlapping inversions appear darker. **B,** Changes in the relative rate of recombination over the entire course of the simulation. The numbers of recombination events occurring at each position (binned in 1Mb windows) are recorded at the formation of each offspring, across all homologous chromosomes in the population. As X chromosomes can still recombine in females, the recombination rate falls to 0.5 for chromosome 1 when all Y chromosomes bear an inversion at the focal position (recombination only in females, which account for 50% of the population). Unlike chromosome 1, chromosome 2 harbors no permanently heterozygous allele. All inversions on this chromosome suffer from homozygosity disadvantage and no inversion spread is therefore observed on chromosome 2. Other simulations are presented in Figures S16-18.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our proposed theory requires only that, within a population, individuals carry different numbers of recessive deleterious mutations in the focal region. Provided that sufficiently large regions are considered, this is probably true for every diploid, dikaryotic or heterokaryotic organism. For instance, the mean distance between two heterozygous positions is ~1000 bp in primates[42], and most of these mutations are probably recessive and deleterious[20,43]. Chromosomal rearrangements spanning hundreds of kilobases are frequently observed[36,44], and would therefore be expected to capture several deleterious variants[45]. We show that inversions are more prone to spreading in regions harboring many segregating deleterious mutations, in which they have a greater chance of capturing very advantageous haplotypes. Variation in the density of deleterious mutations segregating in species, due to the variation of population size, mutation rates, and mating system[21], for example, may account for the large number of different sex-chromosome structures in nature: species harboring many deleterious recessive variants may be more prone to the evolution of large non-recombining regions with evolutionary strata on sex chromosomes. In plants, for example, selection at the haploid stage seems to result in the efficient purging of recessive deleterious mutations[46], potentially accounting for the smaller non-recombining region observed on the sex chromosomes of plants and algae than on those of animals[47,48]. In fungi, evolutionary strata around mating-type genes have been reported only in species with an extended dikaryotic stage, and not those with only a haploid phase[8]. This is consistent with our model, in which the stepwise recombination suppression evolves through the sheltering of deleterious mutations, which cannot occur in haploid organisms. There are many promising possibilities for testing our theory, for example by estimating, in a large number of species, the correlation between the size of the sex-chromosome non-recombining region and inbreeding depression (serving as a proxy for the number of deleterious mutations segregating genome-wide).

In summary, the ubiquitous nature of recessive deleterious mutations in genomes tends to result in inversions being overdominant, and this may play a major role in shaping genome structures. We show here that regions of recombination suppression can readily expand on sex chromosomes simply because permanent heterozygosity at a sex-determining allele makes recombination modifiers associated with fewer-than-average deleterious mutations purely beneficial rather than overdominant. The novel theory we propose applies across a broad range of evolutionary parameter values, and provides a powerful mechanism to account for the stepwise extension of recombination suppression on sex chromosomes, mating-type chromosomes and around supergenes. Our model also provides a parsimonious explanation for the diversity of sex-chromosome architectures observed in nature. This work thus paves the way for further empirical and theoretical research in the rich and active field of genome evolution.

Methods

Simulations

We used SliM V3.2[49] to simulate the evolution of a single panmictic population of 1000 or 10000 individuals in a Wright-Fisher model. Individuals have two pairs of 10Mb chromosomes on which mutations occur at a rate u, with u ranging from 10^{-6} to 10^{-9} per bp, their dominance coefficient h ranges from 0 to 0.5 (0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.5) and their selection coefficient s from -0.5 to 0 (0, -0.001, -0.005, -0.01, -0.05, -0.1, -0.2, -0.5). The main and supplementary figures show the results

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for u=10^{-8} unless otherwise stated. We considered recombination rates of 10^{-6} and 10^{-5} per bp, which gave similar results. Results are presented for analyses in which the recombination rate was 10^{-6}. In the 5,000,000 bp position on chromosome 1, there is a locus with several alleles under balancing selection. Several situations were considered: (i) the locus has two permanently heterozygous alleles, mimicking, for instance, the situation encountered at most fungal mating-type loci, ii) the locus has three (or more) permanently heterozygous alleles, mimicking, for instance, the situation encountered in plant self-incompatibility systems and mushroom (*Agaromycotina*) mating-type loci, (iii) only one of the two alleles is permanently heterozygous, mimicking a classical XY (or ZW) determining system, or (iv) individuals can be either homozygous or heterozygous, with two alleles segregating, but homozygotes have a lower fitness, mimicking loci under balancing selection because of heterozygote advantage, as observed for supergenes in several species. For each parameter combination (u, h, s, heterozygosity rule at the balanced locus), a simulation was run for 15,000 generations, to allow the population to reach an equilibrium for the number of segregating mutations; the population state was saved at the end of this initialization phase. These saved states (one for each parameter combination) were repeatedly used as initial states for studying the dynamics of recombination modifiers. Recombination modifiers mimicking inversions of 500kb, 1000kb, 2000kb and 5000kb were then introduced at positions 0, 0.1, 1, 5, 10 or 50 centimorgans from the locus with permanently (or almost permanently) heterozygous alleles (0, 10000, 50000, 100000 or 500000bp, respectively, with a recombination rate of 10^{-6}), reflecting situations ranging from fully linked to completely unlinked inversions. Recombination modifiers were also introduced on chromosome 2 (with no locus bearing permanently heterozygous alleles), yielding results similar to those for completely unlinked recombination modifiers, as expected. For each parameter combination (h, s, u, size of the region affected by the recombination modification and its linkage with the permanently heterozygous alleles), 10,000 independent simulations were run starting with the introduction of the recombination modifier in the same saved initial population. These inversion-mimicking, recombination modifier mutations were introduced on a single, randomly selected chromosome and, when heterozygous, they suppressed recombination across the region in which they reside (i.e., a *cis*-recombination modifier). We surveyed the frequency of these inversions-mimicking mutations during 10,000 generations (during which all evolutionary processes —mutation, recombination, mating, etc. — remained unchanged). Under the same assumptions and parameters, we also studied the dynamics of recombination modifiers suppressing recombination also when homozygous and not only when heterozygous, again across a fragment in which they reside.

For the evolution of sex chromosomes (Figure 3), we also simulated, over 115,000 generations, the evolution of two 100 Mb chromosomes, one of which carried an XY sex-determining locus at the position 50 Mb (individuals could be either XX or XY and could only mate with individuals of a different genotype). Randomly mating populations of N=1000 and N=10,000 individuals were considered. Point mutations appeared at a rate of 10^{-8} per bp, and their individual selection coefficients were determined by sampling a gamma distribution with a mean ranging from -0.0001 to -0.01 (-0.0001, -0.001, -0.05, -0.03, -0.01) and shape=0.2 (according to observations in humans[50]). The chosen mean of the gamma distribution of the selection coefficient had no qualitative effect on simulation outcome, and we therefore present results only for a mean of -0.03. For each new mutation, a

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dominance coefficient was chosen from the following values, each considered to have an equal probability: 0, 0.001, 0.01, 0.1, 0.25, 0.5. After 15,000 generations, to allow populations to reach an equilibrium in terms of the number of segregating mutations, an inversion was introduced, at each generation, at a random position on a single chromosome. The size of each inversion was determined by sampling from a uniform distribution with a maximum of 50 Mb, 20 Mb, 10 Mb or 1 Mb (the maximum inversion size had no qualitative effect on the process; it simply changed the length of time required for the sex chromosomes to stop recombining; see Figure S17-19). Inversions could overlap. For each parameter combination (population size, mean of the selection gamma distribution, maximum size of inversion), five simulations were run. Simulations were parallelized with GNU Parallel[51].

Mathematical model

We consider the discrete-time evolution of an infinite size, randomly mating population experiencing only deleterious recessive mutations, with heterozygotes and homozygotes suffering from a *1-hs* and *1-s* reduction in fitness, respectively. At all sites, mutations were at the exact same mutation-selection equilibrium frequency, denoted *q*. We used non-approximated values of *q* as defined by ref. 52:

$$q = \frac{h(1+u)}{2(2h-1)}\left[1-\sqrt{1-\frac{4(2h-1)u}{sh^2(1+u)^2}}\right]$$

We follow the frequency of an inversion *I* of size *n*, considering that it captures *m* mutations and that it appears in a population carrying non-inverted segments N. Details of the mathematical analyses are available in Data S1 and summarized below. The marginal fitness of a deleterious mutation at a given position can be defined as:

$$W_i^* = q(1-s) + (1-q)(1-hs)$$

The marginal fitness of a segment with m mutations can therefore be defined as:

$$W_I^* = (q(1-s) + (1-q)(1-hs))^m$$

Considering a segment of size *n*, the frequency of a haplotype with *m* mutations follows a Poisson distribution and $P_m = e^{-nq}(nq)^m/m!$. Provided that *nq* is not too close to 0 (e.g. when a large region is considered), this distribution can be approximated by a Gaussian distribution of mean *nq*. The mean fitness of a non-inverted homozygote NN can be defined as follows:

$$\overline{W_{NN}} = (1-2q(1-q)hs - q^2s)^n$$

Considering small *h* and *s* values, this gives:

$$\overline{W_{NN}} \approx e^{-nqsh-nqs(h+q(1-2h))}$$

Similarly, an individual heterozygous for an inversion I with m mutations has a mean fitness of:

$$\overline{W_{NI}} = (q(1-s) + (1-q)(1-hs))^m(1-hs)^{nq-mq}$$

$$\overline{W_{NI}} \approx e^{-nqsh-ms(h+q(1-2h))}$$

An individual homozygous for an segment *I* with *m* mutations is homozygous for all these mutations. Its fitness can therefore be expressed as

$$\mathrm{W}_{II} = (1-s)^m$$

$$\mathrm{W}_{II} \approx e^{-sm}$$

As indicated in the main text, and considering small *h* and *s* values, the inversion increases in frequency if $W_{NI} > W_{NN}$, and thus if $nq > m$. We follow the frequency of such inversions, assuming a rate of recombination *r* with a locus carrying permanently heterozygous alleles. Four different situations are considered, depending on the possible heterozygosities at the locus with permanently

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

heterozygous alleles. Data S1 presents in detail the estimation of inversion frequency dynamics in these four situations. Here, we briefly describe the results for inversions linked to a permanently heterozygous locus with two alleles, denoted *1* and *2* .

We denote p_j the frequency of the inversion on the chromosome bearing the permanently heterozygous allele *j* (allele *1* or *2*) such that *p*, the population frequency of the inversion, is $p = \frac{p_1 + p_2}{2}$. We can express the marginal fitness of haplotypes as:

$$W^*_{1I} = (1 - p_2)\overline{W_{NI}} + p_2 W_{II}$$

$$W^*_{2I} = (1 - p_1)\overline{W_{NI}} + p_1 W_{II}$$

$$W^*_{1N} = (1 - p_2)\overline{W_{NN}} + p_2 W_{NI}$$

$$W^*_{2N} = (1 - p_1)\overline{W_{NN}} + p_1 W_{NI}$$

The population mean fitness is:

$$\overline{W} = p_1 p_2 W_{II} + p_1(1 - p_2)W_{NI} + (1 - p_1)p_2 W_{NI} + (1 - p_1)(1 - p_2)W_{NN}$$

This system behaves like a system with two loci and two alleles, and the increase in the frequency of the inversion is given by

$$\Delta p_j = \frac{p_j(1 - p_j)(W^*_{jI} - W^*_{jN}) \pm rDW_{NI}}{\overline{W}}$$

with D corresponding to their linkage disequilibrium such that $D = p_1(1 - p_2) - (1 - p_1)p_2$.

The deterministic change in haplotype frequency is highly dependent on the initial state, and, particularly, on the initial level of linkage disequilibrium. As inversions occur on single haplotypes, we present here the results of deterministic simulations with the inversion initially in linkage with only one of the alleles at the permanently or almost permanently heterozygous locus (therefore starting with $D \neq 0$ and not at Hardy-Weinberg equilibrium). Results for various initial states are presented in Figure S7. Unless otherwise stated, the deterministic simulations presented here were performed with *D*=-0.01 or *D*=0.01, depending on the allele with which the inversion appeared to be linked.

Data availability

The Slim and R scripts used for this study are available from GitHub (https://github.com/PaulYannJay/SexChromosomeTheory). No new data were generated for this study. Details concerning the mathematical modeling are available in Data S1.

Acknowledgment

This work was supported by the European Research Council (ERC) EvolSexChrom (832352) grant and a Louis D. Foundation (Institut de France) prize to TG. We thank Ricardo Rodriguez de la Vega, Fanny Hartmann, Jacqui Shykoff, Sylvain Billiard, Amandine Veber, Janis Antonovics and Olivier Tenaillon for comments on a draft version of the manuscript. PJ thanks the authors of SLiM for their outstanding software and manual.

Author contributions

Original ideas, PJ and TG; deterministic model conception, PJ and E.T.; simulations and data analyses, PJ.; interpretation, PJ, TG and ET; manuscript writing, PJ and TG.; editing, ET; project management and funding, TG

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

435 **References**

1. Abbott, J. K., Nordén, A. K. & Hansson, B. Sex chromosome evolution: historical insights and future perspectives. *Proc. Biol. Sci.* **284**, (2017).
2. Cocker, J. M. *et al.* Primula vulgaris (primrose) genome assembly, annotation and gene expression, with comparative genomics on the heterostyly supergene. *Sci. Rep.* **8**, 17942 (2018).
3. Charlesworth, D. Plant contributions to our understanding of sex chromosome evolution. *New Phytol.* **208**, 52–65 (2015).
4. Bergero, R. & Charlesworth, D. The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* **24**, 94–102 (2009).
5. Schwander, T., Libbrecht, R. & Keller, L. Supergenes and complex phenotypes. *Curr. Biol. CB* **24**, R288-294 (2014).
6. Ponnikas, S., Sigeman, H., Abbott, J. K. & Hansson, B. Why Do Sex Chromosomes Stop Recombining? *Trends Genet. TIG* **34**, 492–503 (2018).
7. Wright, A. E., Dean, R., Zimmer, F. & Mank, J. E. How to make a sex chromosome. *Nat. Commun.* **7**, 12087 (2016).
8. Hartmann, F. E. *et al.* Recombination suppression and evolutionary strata around mating-type loci in fungi: documenting patterns and understanding evolutionary and mechanistic causes. *New Phytol.* **229**, 2470–2491 (2021).
9. Jay, P. *et al.* Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nat. Genet.* **53**, 288–293 (2021).
10. Yan, Z. *et al.* Evolution of a supergene that regulates a trans-species social polymorphism. *Nat. Ecol. Evol.* **4**, 240–249 (2020).
11. Ruzicka, F. *et al.* The search for sexually antagonistic genes: Practical insights from studies of local adaptation and statistical genomics. *Evol. Lett.* **4**, 398–415 (2020).
12. Rice, W. R. The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. *Evolution* **41**, 911–914 (1987).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13. Cavoto, E., Neuenschwander, S., Goudet, J. & Perrin, N. Sex-antagonistic genes, XY recombination and feminized Y chromosomes. *J. Evol. Biol.* **31**, 416–427 (2018).
14. Ironside, J. E. No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *BioEssays* **32**, 718–726 (2010).
15. Beukeboom, L. & Perrin, N. *The Evolution of Sex Determination*. (Oxford University Press, 2014).
16. Wright, A. E. *et al.* Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nat. Commun.* **8**, 14251 (2017).
17. Branco, S. *et al.* Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc. Natl. Acad. Sci.* **114**, 7067–7072 (2017).
18. Úbeda, F., Patten, M. M. & Wild, G. On the origin of sex chromosomes from meiotic drive. *Proc. Biol. Sci.* **282**, 20141932 (2015).
19. Charlesworth, B., Coyne, J. A. & Barton, N. H. The Relative Rates of Evolution of Sex Chromosomes and Autosomes. *Am. Nat.* **130**, 113–146 (1987).
20. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).
21. Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
22. Tuttle, E. M. *et al.* Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Curr. Biol.* **26**, 344–350 (2016).
23. Wang, J. *et al.* A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).
24. Nei, M., Kojima, K.-I. & Schaffer, H. E. Frequency Changes of New Inversions in Populations under Mutation-Selection Equilibria. *Genetics* **57**, 741–750 (1967).
25. Olito, C. & Abbott, J. K. The evolution of suppressed recombination between sex chromosomes by chromosomal inversions. *bioRxiv* 2020.03.23.003558 (2020) doi:10.1101/2020.03.23.003558.
26. Takayama, S. & Isogai, A. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56**, 467–489 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

27. Küpper, C. *et al.* A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2016).

28. Stauber, L., Badet, T., Feurtey, A., Prospero, S. & Croll, D. Emergence and diversification of a highly invasive chestnut pathogen lineage across southeastern Europe. *eLife* **10**, e56279 (2021).

29. Kubisiak, T. L. & Milgroom, M. G. Markers linked to vegetative incompatibility (vic) genes and a region of high heterogeneity and reduced recombination near the mating type locus (MAT) in Cryphonectria parasitica. *Fungal Genet. Biol. FG B* **43**, 453–463 (2006).

30. Kirkpatrick, M. & Barton, N. Chromosome Inversions, Local Adaptation and Speciation. *Genetics* **173**, 419–434 (2006).

31. Ohta, T. Associative overdominance caused by linked detrimental mutations*. *Genet. Res.* **18**, 277–286 (1971).

32. Antonovics, J. & Abrams, J. Y. Intratetrad mating and the evolution of linkage relationships. *Evol. Int. J. Org. Evol.* **58**, 702–709 (2004).

33. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).

34. Lenormand, T., Fyon, F., Sun, E. & Roze, D. Sex Chromosome Degeneration by Regulatory Evolution. *Curr. Biol.* **30**, 3001-3006.e5 (2020).

35. Bazzicalupo, A. L., Carpentier, F., Otto, S. P. & Giraud, T. Little Evidence of Antagonistic Selection in the Evolutionary Strata of Fungal Mating-Type Chromosomes (Microbotryum lychnidis-dioicae). *G3 Bethesda Md* **9**, 1987–1998 (2019).

36. Giner-Delgado, C. *et al.* Evolutionary and functional impact of common polymorphic inversions in the human genome. *Nat. Commun.* **10**, 4222 (2019).

37. Todesco, M. *et al.* Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**, 602–607 (2020).

38. Mérot, C., Llaurens, V., Normandeau, E., Bernatchez, L. & Wellenreuther, M. Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nat. Commun.* **11**, 1–11 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

39. Lindtke, D. *et al.* Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Mol. Ecol.* **26**, 6189–6205 (2017).

40. Dobzhansky, T. *Genetics of the Evolutionary Process: Columbia Classics edition*. 505 Pages (Columbia University Press, 1972).

41. Guan, J. *et al.* Genome structure variation analyses of peach reveal population dynamics and a 1.67 Mb causal inversion for fruit shape. *Genome Biol.* **22**, 13 (2021).

42. Osada, N. Genetic diversity in humans and non-human primates and its evolutionary consequences. *Genes Genet. Syst.* **90**, 133–145 (2015).

43. Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nat. Rev. Genet.* **16**, 333–343 (2015).

44. Wellenreuther, M. & Bernatchez, L. Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends Ecol. Evol.* **33**, 427–440 (2018).

45. Connallon, T. *et al.* Local adaptation and the evolution of inversions on sex chromosomes and autosomes. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170423 (2018).

46. Chibalina, M. V. & Filatov, D. A. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol. CB* **21**, 1475–1479 (2011).

47. Filatov, D. A. Homomorphic plant sex chromosomes are coming of age. *Mol. Ecol.* **24**, 3217–3219 (2015).

48. Coelho, S., Gueno, J., Lipinska, A., Cock, J. & Umen, J. UV chromosomes and haploid sexual systems. *Trends Plant Sci.* **23**, 794–807 (2018).

49. Haller, B. C. & Messer, P. W. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Mol. Biol. Evol.* **36**, 632–637 (2019).

50. Kim, B. Y., Huber, C. D. & Lohmueller, K. E. Inference of the Distribution of Selection Coefficients for New Nonsynonymous Mutations Using Large Samples. *Genetics* **206**, 345–361 (2017).

51. Tange, O. Gnu parallel-the command-line power tool. *login: The USENIX Magazine* 42–47 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.17.444504; this version posted May 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

52. Chasnov, J. R. Mutation-Selection Balance, Dominance and the Maintenance of Sex. *Genetics* **156**, 1419–1425 (2000).