Mathematical models to study the interaction between recombination suppression and deleterious mutations near a mating-type locus

Emilie Tezenas Du Montcel

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2023IPPAX078

Thèse de doctorat

Modèles mathématiques pour l'étude de l'interaction entre suppression de recombinaison et mutations délétères au voisinage d'un locus de type sexuel

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École polytechnique

École doctorale n°574 Ecole doctorale de Mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Lille, le 27 octobre 2023, par

EMILIE TEZENAS DU MONTCEL

Composition du Jury :

Vincent Bansaye Professeur, Ecole Polytechnique (CMAP)	Président
Bastien Mallein Professeur, Université Toulouse 3 Paul Sabatier (IMT)	Rapporteur
Laurent Serlet Professeur, Université Clermont Auvergne (LMBP)	Rapporteur
Sophie Pénisson Maître de conférences, Université Paris-Est Créteil (LAMA)	Examinatrice
Solenn Stoeckel Chargé de Recherche, INRAE Rennes (IGEPP)	Examinateur
Amandine Véber Directrice de Recherche, Université Paris-Cité (MAP5)	Directrice de thèse
Sylvain Billiard Professeur, Université de Lille (EEP)	Co-directeur de thèse
Tatiana Giraud Directrice de recherche, Université Paris Saclay (ESE)	Co-directrice de thèse

European Research Council

Established by the European Commission

Résumé

Cette thèse propose et développe plusieurs modèles stochastiques permettant d'avancer dans la compréhension de l'évolution de la suppression de recombinaison sur les chromosomes sexuels et de type sexuel. La recombinaison est un mécanisme d'échange de parties de chromosomes, qui crée de nouvelles combinaisons d'allèles. Cependant, de larges régions de suppression de recombinaison ont été observées chez de nombreuses espèces englobant des gènes impliqués dans la compatibilité lors de la reproduction sexuée (gènes déterminant le sexe ou le type sexuel). Les mécanismes induisant l'extension de la zone sans recombinaison au-delà de ces gènes impliqués dans la compatibilité sexuelle font encore débat. Dans cette thèse, on met en place différentes approches mathématiques pour étudier l'interaction entre la dynamique des mutations délétères et celle d'un suppresseur de recombinaison. Le premier chapitre a permis, grâce à l'analyse d'un modèle déterministe simple et à des simulations d'un modèle stochastique plus complexe, de montrer l'effet de la présence de mutations délétères à plusieurs étapes de l'évolution de la suppression de recombinaison. Le deuxième chapitre se concentre sur la dynamique des mutations délétères au voisinage d'un locus en permanence hétérozygote et dans des populations d'individus pouvant se reproduire par allo- ou auto-fécondation. On modélise l'évolution initiale de mutations délétères au moyen d'un processus de branchement multitype, dont on étudie la criticité et la distribution du temps d'extinction. Enfin, le troisième chapitre compare l'effet de l'accumulation de mutations délétères chez des individus autofécondants selon qu'ils peuvent recombiner ou non. On utilise un processus de sauts à valeurs mesures sur un espace de traits à trois dimensions pour étudier l'évolution de la charge mutationnelle d'individus possédant un locus de type sexuel toujours hétérozygote.

Abstract

This PhD manuscript presents the development of several stochastic models that contribute to our understanding of recombination suppression evolution on sex and mating-type chromosomes. Recombination is a mechanism that exchanges parts of chromosomes, which creates novel allelic combinations. However, there have been reports in a wide range of organisms of large regions with suppressed recombination that encompass genes involved in mating compatibility (i.e., genes determining sex or mating type). The nature of the mechanisms that induce the extension of the non-recombining zone beyond the genes involved in sex compatibility remains debated. In this PhD thesis, we use various mathematical approaches to study the dynamics of deleterious mutations and recombination suppressors. The first chapter shows, with the analysis of a simple deterministic model and the simulation of a more complex stochastic one, that deleterious mutations impact the evolution of a recombination suppressor at several stages. The second chapter focuses on deleterious mutations dynamics near a permanently heterozygous locus in selfing or outcrossing populations. We model the initial evolution of deleterious mutations with a multitype branching process and study its criticality and its extinction time distribution. In the third chapter, we compare the effect of deleterious mutation accumulation in selfing populations of recombining or non-recombining individuals. We use a measure-valued branching process on a three-dimensional trait space to study the evolution of the mutational load for populations carrying an always-heterozygous mating-type locus.

Remerciements

Tout d'abord, merci au CNRS et à l'European Research Council pour avoir financé cette thèse. Merci à Tatiana Giraud de m'avoir fait confiance et sélectionnée sur ce financement.

Merci ensuite à toutes les structures qui m'ont accueillie pendant ces trois ans, pour des missions et des raisons diverses et variées : les laboratoires EEP et Paul Painlevé de l'Université de Lille, le laboratoire ESE et l'équipe GEE de l'université Paris-Saclay (feu Orsay...), le MAP5 de Paris-Cité, le CMAP, la Chaire MMB...

Merci encore à Bastien Mallein et Laurent Serlet d'avoir accepté de rapporter cette thèse, et à Vincent Bansaye, Sophie Pénisson et Solenn Stoeckel d'avoir accepté de faire partie du jury.

Ensuite, une thèse ne se fait pas en solo, loin de là. J'ai eu la chance pendant la mienne de croiser le chemin de beaucoup de belles personnes, qui m'ont aidée à traverser ces trois ans, et à faire de cette thèse, j'en suis convaincue, la meilleure expérience de la recherche que j'aurais pu avoir. Alors, même si mon coeur et mon envie sont définitivement du côté de l'enseignement, merci à vous de m'avoir accompagnée pendant ce petit bout de chemin, professionnel et personnel.

Les premiers remerciements vont bien sûr à mes trois encadrants, Amandine, Sylvain et Tatiana. Merci de m'avoir fait confiance sur ce projet, et de m'avoir embarquée avec vous dans cette aventure multi-disciplinaire. Merci pour votre présence et votre bienveillance, pour votre tri-encadrement si complémentaire et riche, tant scientifiquement qu'humainement. Merci pour votre patience, votre pédagogie, votre disponibilité, votre rigueur scientifique et votre bonne humeur constante.

C'était un honneur pour moi d'être encadrée par vous, j'ai énormément appris à votre contact, et me suis nourrie de votre passion pour la recherche. Certes, cela ne suffira pas à me détourner des salles de classe, mais je pense sincèrement que je n'aurais pas pu bénéficier d'un meilleur encadrement qu'avec vous. Alors un très grand MERCI.

Merci ensuite à tout le laboratoire EEP pour avoir fait de ces 3 années lilloises une chouette expérience du grand Nord ! Merci pour votre accueil chaleureux de l'extraterrestre mathématique qui pose des questions comme "Mais une plante peut avoir deux sexes ? ?" devant vos sourires amusés. Je n'aurais pas beaucoup mis les mains sur les paillasses, mais je suis ravie d'avoir passé ces années dans votre milieu. Merci à vous pour les discussions scientifiques, mais aussi pour tout ce qui fait qu'il fait bon vivre dans ce labo : les discussions au détour d'un couloir sur les manip qui marchent (ou pas) et les prochaines vacances prévues, les énigmes du tableau blanc et les gâteaux de la salle café, les repas partagés, au RU ou en salle de pause. Le débat sur la meilleure manière de faire bouillir de l'eau n'est pas encore résolu, mais la conclusion s'impose : on s'est bien marrés.

Merci également à vous, l'équipe GEE, pour avoir eu autant de patience et de petits sourires amusés que les lillois devant mes questions naïves, et avoir toujours pris le temps de m'expliquer vos travaux. Je garderai un excellent souvenir des apéros fromages, et aussi de vos jolis dessins dans les boites de Petri. Merci pour votre accueil toujours chaleureux.

Un merci aussi à tous les membres de la chaire MMB, avec lesquels il était toujours agréable de discuter, en montagne ou ailleurs... C'était une chance pour moi de pouvoir vivre dans cette communauté pendant ma thèse,

merci à tous. Un merci particulier pour Denis, pour sa bienveillance permanente, et pour Sepideh, Michael et Diala pour des discussions scientifiques et humaines.

Merci à Sophie Pénisson, Aline Marguet et Chi Tran pour avoir écrit des manuscrits de thèse très détaillés qui sont vite devenus des références indispensables pour mon sujet !

Merci aussi aux enseignants-chercheurs du laboratoire Paul Painlevé, avec qui j'ai pu partager des services d'enseignement. Mylène, Laurence, Christine, c'était un plaisir de travailler avec vous ! Merci aussi à Sophie Grivaux pour avoir pris le temps de me recevoir et de discuter de mon sujet.

Merci à Clément Mazoyer pour son aide précieuse sur la mise en place des simulations, et merci à lui et à Matthieu Genete pour leur travail impressionnant sur tous les outils numériques, c'est une chance de vous avoir ! Merci également à Sarah Kakkai et collaborateurs pour leur superbe outil de simulation de processus de sauts.

Merci à tous les intervenants administratifs, grâce à qui j'ai navigué dans les dédales administratifs de 5 (!) laboratoires. Cela multiplie d'autant le nombre de personnes qui ont dû s'arracher les cheveux devant la complexité de mon dossier... Merci pour votre patience !

Merci à mes grandes soeurs et grands frères de thèse. Céline pour ton soutien et ta présence, Paul pour m'avoir embarquée avec toi sur un petit bout de ton brillant chemin, Djivan pour nos discussions scientifiques sous le Catalpa du labo, Thibault pour ta bonne humeur et tes conseils avisés. Vous avez été de vrais guides pour moi, merci d'avoir été là.

Merci bien sûr à la fine équipe des "jeunes EE", qui ont rendu cette thèse aussi fun qu'elle pouvait l'être ;) On aura pas eu le temps de cocher toutes les cases sur les tableaux tant que j'étais encore là, mais j'ai confiance pour que le défi soit relevé avant le renouvellement complet de l'équipe ;)

Un merci particulier à François, pour tes questions toujours stimulantes, et surtout ta présence et ton soutien constants. Merci aussi Flavia pour ton grain de folie qui rend la vie plus piquante, et ta présence. Sans vous deux, nos pauses cafés, nos séances d'escalade, nos p'tites bouffes dans Lille et autres joyeusetés, ça aurait été beaucoup moins marrant...

Merci aussi à tous ceux qui m'ont soutenue de plus ou moins loin : ma famille de Lille, des Alpes, et d'ailleurs, le groupe Comuze, les copains de Rennes, de Lyon, de Lille, et partout ailleurs. Vous savoir derrière moi m'a permis de tenir, jusqu'au bout ! Merci à la SNCF au fait, pour m'avoir permis de parcourir la France en long en large et en travers pour rendre visite auxdits copains, jusqu'à cumuler assez de km pour faire au moins une fois le tour de la Terre... Et pour permettre auxdits copains et familles d'être là aujourd'hui !

Merci en particulier à Ségo, qui m'a suivie depuis plus de 10 ans dans toutes mes péripéties. Ségo Ségo, merci pour ton amitié constante, te savoir là me rassure <3

Merci aussi à Pierre, le maestro de LaTeX. Merci pour être encore là, pour tes conseils et ton aide précieuse. Après l'aventure Comuze, l'aventure agreg, et l'aventure thèse, je suis contente qu'on partage ensemble l'aventure prépa !

Et merci mon Djo, pour avoir embarqué avec moi dans une autre aventure du quotidien.

Je voudrais terminer ces remerciements en dédiant cette thèse à ma grand-mère Manou, qui fut elle aussi enseignante et chercheuse, et avec qui j'aurais aimé partager cette occasion.

Table des matières

Introduction

Cette thèse se situe au carrefour entre mathématiques et biologie de l'évolution. La problématique biologique est l'étude de l'extension de la suppression de recombinaison sur les chromosomes sexuels ou de type sexuel. Ces termes, et la motivation biologique derrière cette problématique, sont expliqués dans la première partie d'introduction. On s'intéresse ensuite aux méthodes mathématiques existantes pour étudier cette problématique. Dans un premier temps, nous nous concentrons sur un état de l'art de la littérature de biologie et modélisation. Les résumés des trois chapitres qui constituent cette thèse proposés ensuite contiennent davantage de détails mathématiques et un état de l'art plus précis des modèles-clés sur lesquels les analyses présentées dans cette thèse s'appuient. Un glossaire est proposé en fin d'introduction, contenant des termes biologiques et mathématiques. Dans les deux cas, le but est de donner une définition qui permette d'utiliser le mot donné dans le contexte de cette thèse. Les mots s'y trouvant sont repérés par une étoile dans le texte. Un dernier chapitre de discussion et de perspectives est proposé en fin de manuscrit.

1 Motivations biologiques

1.1 La sélection

Pour comprendre le travail réalisé, il est essentiel de comprendre ce qu'est la sélection naturelle, et comment elle agit. Au départ, un caractère apparaît de façon aléatoire chez un individu dans une population. Ce caractère peut conférer un avantage. Cela peut être un avantage reproductif direct (plus de facilité à trouver des partenaires, reproduction plus rapide, demandant moins de ressources...), un avantage pour la survie (survie plus longue donc plus d'opportunités de reproduction, meilleure capacité à se nourrir donc meilleure survie et plus d'énergie à consacrer à la reproduction,...), etc... Quelle que soit la forme de l'avantage, la capacité de reproduction des individus portant ce caractère est améliorée, donc le caractère est davantage transmis de génération en génération que d'autres caractères. On dit que ce caractère est *sélectionné* par rapport aux autres variants présents dans la population. Si le caractère finit par envahir toute la population, on dit qu'il est *fixé*.

Un nouveau caractère peut aussi être *délétère**. Dans ce cas, ce sont les variants pré-existants qui seront sélectionnés par rapport à ce nouveau caractère. On dit que le caractère délétère est *contre-sélectionné*. S'il finit par ne plus être porté par aucun individu, on dit qu'il a été *purgé*.

1.2 La recombinaison

La recombinaison est un phénomène observé chez la plupart des êtres vivants eucaryotes (c'est-à-dire possédant un noyau) passant par une phase *diploïde** au cours de leur cycle de vie. Elle consiste en l'échange de deux parties de chromosomes homologues, c'est-à-dire de chromosomes d'une même paire qui contiennent les mêmes gènes, et se produit durant la méiose, c'est-à-dire la production de gamètes (voir Figure 1). Lorsque deux chromosomes recombinent, on dit qu'un *crossing-over* se produit. Pour qu'un crossing-over soit possible, il faut que deux chromosomes homologues s'apparient, ce qui nécessite que la disposition des gènes soit similaire sur les deux chromosomes homologues. En particulier, si une mutation perturbe l'ordre des gènes (par exemple, une *inversion**), ou la distance entre deux gènes (par exemple, une *insertion** ou une *délétion**), l'appariement ne peut être effectué, et la recombinaison est bloquée. Dans cette thèse on parlera surtout d'inversion pour désigner un suppresseur de recombinaison.

La recombinaison peut être sélectionnée pour des avantages mécanistiques (ou proximaux). Entre autres, les crossing-over permettent d'assurer le bon appariement et donc la séparation correcte des chromosomes

durant la méiose, et la recombinaison peut permettre la réparation des chromosomes endommagés (Barton et Charlesworth, 1998, discuté dans Otto et Lenormand, 2002).

Par ailleurs, la recombinaison rend la sélection plus efficace. En effet, lorsque la recombinaison est possible entre deux gènes, ou deux *loci**, de nouvelles combinaisons d'*allèles** peuvent être créées chez les descendants. De façon générale, cela augmente la diversité génétique au sein d'une population, et donc la variance de la distribution de *valeur sélective** (ou *fitness**). Cela permet une sélection plus efficace vers l'optimum de valeur sélective, en particulier dans un environnement changeant (Barton et Charlesworth, 1998, Otto, 2009). Plus précisément, la recombinaison a des effets directs sur la composition génétique des individus. Elle permet par exemple de regrouper des allèles bénéfiques sur deux loci différents qui seraient apparus sur deux chromosomes différents, ce qui accélère l'adaptation de la population (Barton et Charlesworth, 1998, Otto, 2009). Par ailleurs, de la même façon que la recombinaison peut regrouper des allèles bénéfiques, elle peut regrouper des allèles délétères, et ainsi reformer des chromosomes peu chargés en mutations délétères à partir de chromosomes portant des mutations délétères à différents loci. Cela permet de stopper la course du "cliquet de Muller", qui désigne la perte par dérive génétique des chromosomes les moins chargés en mutation délétères dans une population. Si la recombinaison est possible, des chromosomes peu chargés peuvent être re-créés à partir de chromosomes plus chargés. Si la recombinaison est bloquée, le minimum de mutations délétères portées par un *haplotype** augmente irréversiblement (Muller, 1964). En résumé, la recombinaison permet de rendre la sélection plus efficace en créant des génotypes combinant différentes mutations bénéfiques, et peu chargés en mutations délétères (Bachtrog, 2006).

Cependant, la recombinaison peut être défavorable lorsqu'elle rompt des associations d'allèles bénéfiques. Ainsi, la suppression de recombinaison peut être sélectionnée, par exemple pour lier de façon permanente plusieurs allèles formant un complexe favorable, appelé *supergène**. Les allèles sont alors transmis en bloc, et l'avantage sélectif est conservé. De tels supergènes ont été observés pour des gènes impliqués dans la détermination du sexe ou du type sexuel (Hartmann et al., 2021) ainsi que pour des gènes codant pour des caractères phénotypiques (Schwander et al., 2014, Jay et al., 2022a).

1.3 Interactions entre mode de reproduction et recombinaison

Un individu diploïde est produit par auto-fécondation lorsque les *gamètes** qui fusionnent proviennent du même parent. On l'oppose à l'allo-fécondation, qui désigne une reproduction entre deux individus distincts. L'auto-fécondation a tendance à augmenter l'*homozygotie** et donc à exposer les mutations délétères, ce qui favorise leur purge. Mais ce mode de reproduction réduit aussi la *taille efficace** de la population, et donc l'efficacité de la sélection (Hartfield et Glémin, 2016). Plus largement, l'auto-fécondation agit sur la composition génétique des populations : voir Glémin et Galtier, 2012 pour une revue comparative des effets de l'autofécondation, l'allo-fécondation, et la reproduction asexuée, Roze et Rousset, 2004 pour une étude de l'impact de l'autofécondation dans des populations structurées, Roze, 2016 pour une étude sur l'effet de l'autofécondation sur la sélection d'arrière plan, Hartfield et Glémin, 2016 pour l'effet de l'autofécondation sur la fixation de mutations bénéfiques. La reproduction par auto-fécondation peut donc avoir un impact sur la sélection pour ou contre la recombinaison.

Roze et Lenormand, 2005 décrivent les effets opposés de la recombinaison entre une reproduction via auto-fécondation et une reproduction via allo-fécondation. Lorsque les individus d'une population se reproduisent en allo-fécondation, la recombinaison diminue la valeur sélective moyenne des descendants mais augmente la variance. C'est l'effet connu d'augmentation de l'efficacité de la sélection par la recombinaison. Au contraire, l'inverse se produit lorsque les individus se reproduisent par auto-fécondation : la recombinaison diminue la variance de la valeur sélective mais augmente la valeur sélective moyenne des descendants. En fonction du paysage de sélection, la recombinaison peut être sélectionnée ou contre-sélectionnée dans l'un ou l'autre des modes de reproduction. Par exemple, Charlesworth et Wall, 1999 répertorient des situations dans lesquelles un avantage hétérozygote est présent. La suppression de recombinaison est alors favorisée pour maintenir les allèles à l'état hétérozygote, et contrer la production accrue d'homozygotes induite par l'auto-fécondation.

1.4 Les chromosomes sexuels ou de type sexuel

Les chromosomes sexuels sont les chromosomes qui portent les gènes impliqués dans la détermination du sexe d'un individu. L'exemple le plus connu est celui des mammifères, dont le sexe est déterminé par un chromosome

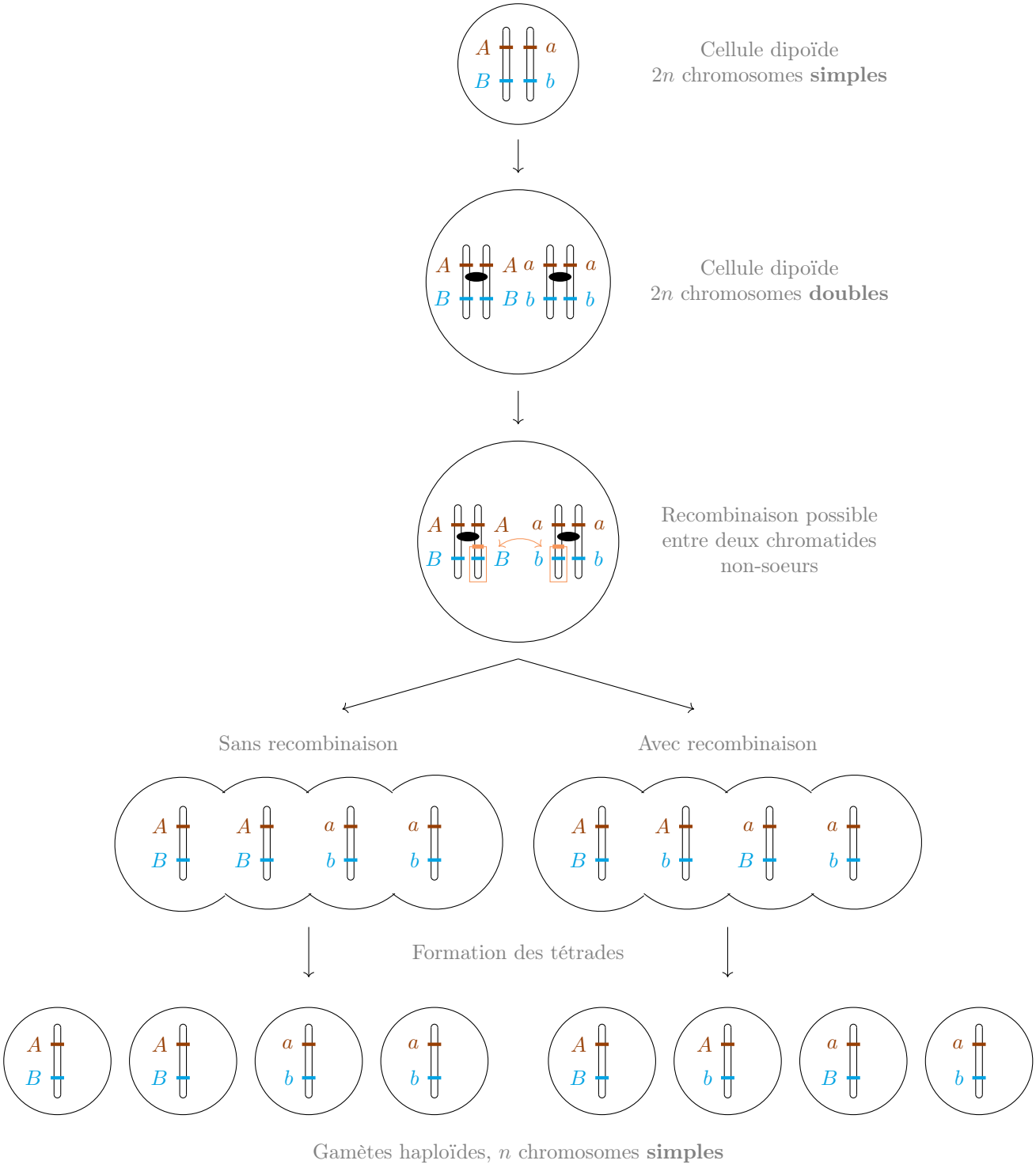


FIGURE 1 – Schéma représentant les étapes de la méiose. On part d'une cellule diploïde contenant des paires de chromosomes simples. Les chromosomes se dupliquent pour donner des chromosomes doubles, composées de chromatides soeurs reliées par le centromère (ellipse noire). À ce stade, un (ou plusieurs) crossing-over peut (peuvent) se produire entre deux chromatides non-soeurs. Les chromatides soeurs sont ensuite séparées en quatre cellules haploïdes contenant des chromosomes simples, les gamètes. On appelle *tétrade* l'ensemble des quatre gamètes produits lors d'un évènement de méiose. La tétrade de gauche présente le résultat de la méiose sans recombinaison, celle de droite présente le résultat de la méiose avec recombinaison. La recombinaison fait apparaître deux nouvelles combinaisons d'allèles : A/b et a/B.

X et un chromosome Y (il existe d'autres systèmes de détermination du sexe, voir par exemple Coelho et al., 2018). Dans cette thèse, on parlera également de chromosomes de type sexuel, présents chez des espèces qui ne comportent pas de sexe séparés (pas d'individus femelles ou mâles). Ces chromosomes de type sexuel ne sont pas associés à des sexes différents mais déterminent une compatibilité sexuelle : deux gamètes haploïdes ne peuvent fusionner et former un nouvel individu que s'ils portent des types sexuels différents (Billiard et al., 2012, Hartmann et al., 2021).

Les chromosomes sexuels, ou de type sexuel, ont une évolution particulière, qui diffère des autres chromosomes, et s'observe chez de nombreuses espèces (Otto et Lenormand, 2002, Bergero et Charlesworth, 2009, Wright et al., 2016, Hartmann et al., 2021). Un chromosome sexuel, ou de type sexuel, commence à se former lorsqu'un gène de déterminisme du sexe, ou du type sexuel, apparaît sur un chromosome. D'autres gènes impliqués dans la détermination du sexe ou du type sexuel sont susceptibles d'apparaître et de se fixer dans la population. L'arrêt de la recombinaison entre ces différents gènes déterminant le sexe ou le type sexuel peut être sélectionné pour lier ensemble les allèles déterminant le même sexe ou le même type sexuel. Ce supergène ainsi créé assure le bon fonctionnement du mécanisme de détermination du sexe ou du type sexuel. Ces étapes sont bien comprises, autant dans le cas des chromosomes sexuels que dans le cas des chromosomes de type sexuel (Bachtrog, 2006, Charlesworth et Charlesworth, 1987, Wright et al., 2016, Hartmann et al., 2021). Par exemple, dans le cas des champignons, deux loci sont impliqués dans la compatibilité de deux gamètes (les locus HD et PR). Deux gamètes ne peuvent fusionner que s'ils portent des allèles différents à chaque locus. La suppression de recombinaison entre les locus HD et PR a permis de fixer certaines combinaisons d'allèles entre ces deux loci, et d'augmenter les chances de compatibilité entre deux gamètes lors d'une reproduction par auto-fécondation, fréquente chez les champignons (voir la figure 2 de Hartmann et al., 2021 pour plus de détails sur les rôles des loci HD et PR).

Il a cependant été observé que la suppression de recombinaison s'étendait le long des chromosomes sexuels, au-delà des gènes impliqués dans la détermination du sexe ou du type sexuel (Bachtrog, 2006, Bergero et Charlesworth, 2009, Hartmann et al., 2021, Wright et al., 2016). Cette extension de suppression de recombinaison entraîne l'accumulation de mutations délétères sur le chromosome se retrouvant en permanence à l'état hétérozygote (typiquement, le Y dans un système XY), et, finalement, la dégénérescence de ce chromosome. Au contraire, le chromosome sexuel pouvant se retrouver à l'état homozygote continue de recombiner dans cet état, et purge ainsi les mutations délétères. On observe aujourd'hui chez l'humain un chromosome Y beaucoup plus petit que le X, et des analyses génétiques ont permis de montrer qu'il contenait beaucoup moins de gènes (Figure 2, et Wright et al., 2016).

Hypothèses pour expliquer l'extension de la suppression de recombinaison

Vu les avantages apportés par la recombinaison et l'impact sur l'évolution des chromosomes sexuels que son arrêt provoque, la sélection pour l'arrêt de recombinaison au-delà des gènes impliqués dans le déterminisme du sexe ou du type sexuel est mal comprise et fait l'objet de nombreux travaux récents (Ironside, 2010, Wright et al., 2016, Jeffries et al., 2021, Lenormand et Roze, 2022, voir aussi le numéro spécial des *Philosophical Transactions in Biodiversity*, Kratochvíl et Stöck, 2021). L'hypothèse longtemps admise pour expliquer l'extension de suppression de recombinaison sur les chromosomes sexuels est celle de l'antagonisme sexuel : la recombinaison peut être supprimée entre un allèle favorable pour seulement un des sexes et l'allèle de détermination du sexe (Rice, 1987, Charlesworth et al., 2005, Bergero et Charlesworth, 2009, Ruzicka et al., 2020). Un exemple classique pour illustrer ce mécanisme est celui de la coloration des mâles chez des espèces de poisson (Winge, 1927, Wright et al., 2016). La coloration d'un individu le rend plus exposé à la prédation, mais les mâles les plus colorés disposent d'un avantage à la reproduction. Le caractère "coloré" est donc bénéfique pour les mâles mais délétère pour les femelles. La liaison entre l'allèle responsable de la coloration d'un individu et l'allèle responsable de la détermination du sexe mâle peut alors être sélectionnée. Cependant, les nombreux travaux empiriques menés depuis plusieurs décennies n'ont pas permis de mettre en évidence le rôle systématique de l'antagonisme sexuel dans l'extension de la suppression de recombinaison sur les chromosomes sexuels (Ironside, 2010). De plus, cette hypothèse ne permet pas d'expliquer l'extension de la suppression de recombinaison chez les espèces ne présentant pas de sexes séparés ni une autre forme de sélection antagoniste. C'est le cas des champignons, dont les chromosomes de type sexuels présentent pourtant une extension de suppression de recombinaison par étapes (Bazzicalupo et al., 2019, Hartmann et al., 2021).

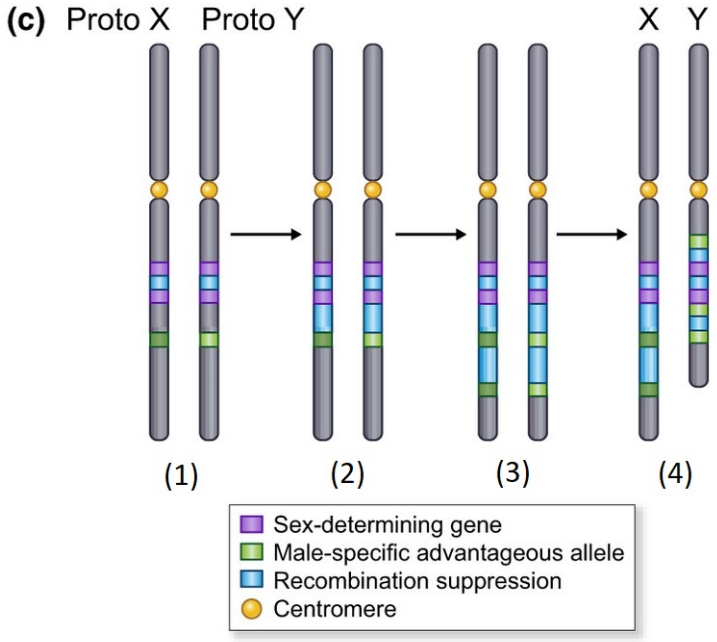


FIGURE 2 – Schéma de l'évolution des chromosomes sexuels X et Y, sous l'hypothèse d'antagonisme sexuel. Tiré de HARTMANN et al., 2021. Les zones bleues représentent l'arrêt de la recombinaison. Le plus petite taille du chromosome Y à droite témoigne du début de sa dégénérescence suite à l'arrêt de la recombinaison. Dans cette thèse, on s'intéresse à une hypothèse alternative permettant d'expliquer l'extension de la suppression de recombinaison au-delà des gènes de déterminisme du sexe, c'est-à-dire aux étapes (2) et (3) présentées sur ce schéma.

Plusieurs hypothèses ont été depuis proposées pour expliquer l'extension de la suppression de recombinaison au-delà des gènes de détermination du sexe ou du type sexuel, comme l'accumulation de divergence neutre (JEFFRIES et al., 2021) ou l'interaction avec les éléments transposables (KENT et al., 2017).

Rôle des mutations délétères

Une autre hypothèse est présentée sous le nom d'"abritement d'allèles délétères" (HARTMANN et al., 2021). Elle repose sur la présence de mutations récessives et délétères sur les chromosomes, qui est une caractéristique partagée par tous les génomes. Si un caractère supprime la recombinaison entre des mutations délétères récessives et le locus en permanence hétérozygote, les mutations sont également maintenues à l'état hétérozygote. Elles n'expriment alors pas tout leur potentiel délétère et sont dites "abritées". Cela peut conférer un avantage sélectif au fragment non-recombinant, et conduire à la fixation de la suppression de recombinaison dans la population (figure 3).

Le fragment non-recombinant a alors bénéficié d'un avantage hétérozygote, c'est-à-dire qu'il constituait un supergène *super-dominant**. Des modèles simples considérant un locus en permanence hétérozygote et un seul allèle super-dominant ont montré que la suppression de recombinaison entre ces deux loci pouvait être sélectionnée lorsque les individus se reproduisent par auto-fécondation intra-tétrade exclusivement ou avec un mélange d'allo et d'autofécondation (CHARLESWORTH et WALL, 1999, ANTONOVICS et ABRAMS, 2004). Les résultats de ces modèles sont cependant limités par la méthode choisie (les modèles sont déterministes, voir partie 2.1), et par la restriction à un seul locus sous sélection dans le voisinage du locus en permanence hétérozygote.

Le but de cette thèse est d'étendre les résultats de ces modèles et de poursuivre l'étude de l'impact des mutations délétères sur la sélection pour la suppression de recombinaison au moyen de modèles mathématiques. En particulier, on s'intéressera à deux points importants concernant l'effet des mutations délétères. Le premier est la distribution non-uniforme des mutations délétères dans la population, ce qui crée de la variance dans la valeur sélective des régions flanquantes du locus en permanence hétérozygote. La suppression de recombinaison pourra être sélectionnée si l'haplotype sur lequel elle apparaît initialement possède une valeur sélective meilleure que la moyenne de la population. Cela peut se produire par exemple si l'haplotype porte moins de mutations délétères que la moyenne dans la population (NEI et al., 1967). On cherchera donc à prendre en compte la

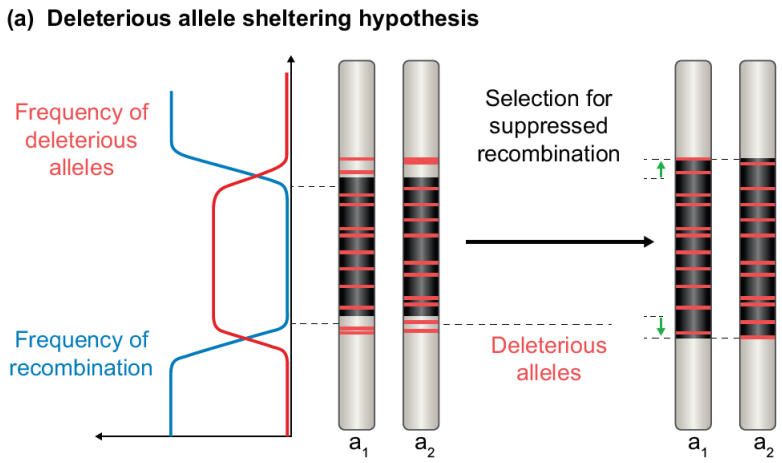


FIGURE 3 – Schéma explicatif de l'hypothèse d'abritement d'allèles délétères présentée dans HARTMANN et al., 2021 pour expliquer l'extension de la suppression de recombinaison sur les chromosomes de type sexuel. Deux chromosomes de type sexuel a_1 et a_2 sont représentés. La zone noire symbolise la zone de suppression de recombinaison. Les traits rouges symbolisent des mutations délétères.

possibilité de multiples mutations délétères dans les modèles étudiés. Le deuxième point est la présence d'un allèle trouvé en permanence à l'état hétérozygote sur les chromosomes sexuels ou de type sexuel. C'est le cas par exemple de l'allèle Y du locus de détermination du sexe chez les mammifères, ou des allèles de type sexuel chez les champignons. En effet, quand l'haplotype non-recombinant augmente en fréquence, ses mutations délétères pourront se retrouver à l'état homozygote. La présence d'un allèle en permanence hétérozygote sur l'haplotype non recombinant garantit alors son maintien à l'état hétérozygote, ce qui permet l'abritement des mutations délétères qui lui sont liées. L'haplotype non recombinant pourrait alors aller jusqu'à la fixation.

On continuera d'autre part à tester l'effet de différents modes de reproduction comme l'allo-fécondation ou l'auto-fécondation, fortement représentée chez les champignons. En effet, le mode de reproduction est susceptible de jouer un rôle non-négligeable dans la sélection pour ou contre la suppression de recombinaison, puisqu'il agit sur la composition génétique des populations (voir partie 1.3). En particulier, l'autofécondation limite le brassage allélique à l'échelle de la population et crée du *déséquilibre de liaison** entre allèles. Les mutations délétères peuvent alors être plus souvent associées à un des allèles du locus hétérozygote en particulier, ce qui peut favoriser la sélection pour la suppression de recombinaison pour abriter ces allèles délétères (CHARLESWORTH et WALL, 1999, ANTONOVICS et ABRAMS, 2004). La présence d'un locus de type sexuel en permanence hétérozygote peut de plus perturber l'interaction entre mode de reproduction et recombinaison. En effet, la contrainte d'hétérozygotie à un locus augmente l'hétérozygotie aux loci qui lui sont liés. La recombinaison peut au contraire rétablir l'homozygotie. Dans une population d'individus se reproduisant par auto-fécondation, ce qui augmente l'homozygotie, mais possédant un locus en permanence hétérozygote, ce qui augmente l'hétérozygotie, l'effet et l'avantage potentiel de la recombinaison n'ont pas été encore bien étudiés.

2 Quelques modèles mathématiques pour la génétique des populations

Dans cette section, on décrit les différents types de modèles utilisés dans cette thèse pour étudier l'évolution de la composition génétique d'une population. Dans un premier temps, on décrit des approches déterministes, qui permettent d'étudier le comportement moyen d'un système. On s'intéresse ensuite à des méthodes stochastiques, qui permettent d'étudier les fluctuations d'un système autour de ce comportement moyen. Enfin, on décrit un type d'approche stochastique particulier qui permet de détailler la dynamique fondamentalement aléatoire d'un caractère lors de sa phase initiale d'évolution, juste après son apparition.

2.1 Modèles déterministes

Le premier type de modèle introduit par Fisher, Wright et Haldane suit l'évolution génomique d'une population de taille infinie via la variation déterministe des fréquences alléliques (Crow et Kimura, 2010, Rice, 2004). L'idée est que les individus diploïdes produisent un "pool" infini de gamètes, dans lequel sont piochés les gamètes formant les individus de la génération suivante. Les objets mathématiques suivis sont les fréquences alléliques. La contribution de chaque génotype présent à la génération n au pool de gamètes, et la proportion d'individus diploïdes de chaque type formés pour la génération $n+1$ dépendent des paramètres du modèle tels que la sélection, qui peut avoir lieu à l'état haploïde sur les gamètes ou à l'état diploïde sur les parents, de la mutation, ou encore des différents modes de reproduction considérés (auto-fécondation, ou allo-fécondation). Ce genre de modèle permet d'étudier le comportement moyen d'un système. Les systèmes d'équations obtenus sur les fréquences alléliques ne sont pas toujours solubles analytiquement, surtout lorsqu'un grand nombre d'allèles ou de loci est considéré, mais ils permettent de réaliser des calculs numériques, ou d'en étudier des approximations linéaires.

On utilise un modèle déterministe dans le chapitre 1 pour étudier le comportement moyen d'une inversion (*i.e.* d'un suppresseur de recombinaison). On considère plusieurs systèmes sexuels, dont des systèmes reposant sur l'hétérozygotie permanente à un locus (les systèmes de type sexuel, ou un système XX/XY, voir la partie 3.1).

L'utilisation de tels modèles a permis d'obtenir des résultats sur les conditions pour la sélection pour ou contre la recombinaison, résumés et discutés dans Otto, 2009. Les premières générations de modèles font l'hypothèse d'une population infinie à l'équilibre sélectif et montrent que, dans ce cas, la recombinaison est défavorisée. En effet, les combinaisons d'allèles optimales sont déjà atteintes et les mélanger serait défavorable. En relâchant cette hypothèse d'équilibre et en rajoutant des termes d'épistasie (c'est-à-dire d'interaction entre allèles de plusieurs loci sur la valeur sélective), une deuxième génération de modèles montre que la recombinaison est favorisée lorsque l'épistasie est négative (c'est-à-dire lorsque l'effet de plusieurs allèles délétères est plus délétère que la somme des effets individuels). Cependant, la gamme de paramètres sur laquelle un tel comportement peut être observé est restreinte. Une troisième génération de modèles relâche l'hypothèse de population de taille infinie tout en restant dans un cadre déterministe, en prenant en compte la *dérive génétique* produite par l'aléatoire de la reproduction dans les populations de taille finie. En effet, contrairement au cas infini, la transmission des allèles dans une population finie est soumise à la dérive génétique, c'est-à-dire à des phénomènes aléatoires. Le choix aléatoire d'un nombre fini de génotypes composant une nouvelle génération peut avoir comme effet que certaines associations d'allèles sont sur-représentées par rapport au cas infini, ce qui crée du *desequilibre de liaison*. Cette troisième génération de modèles montre que la présence de déséquilibre de liaison dans les populations de taille finie induit une sélection plus forte pour la recombinaison, afin de reformer des associations favorables.

Des études récentes étendent ces résultats dans le cas d'individus pouvant se reproduire par auto-fécondation (Stetsenko et Roze, 2022, comme extension de Roze et Lenormand, 2005 basé sur des méthodes développées par Barton et Turelli, 1991). Ces études considèrent un modèle déterministe qui incorpore l'effet moyen des associations aléatoires générées par une taille de population finie et montrent que, dans des populations avec un système de reproduction mixte (auto- et allo-fécondation), la présence d'épistasie négative a un fort effet sur la sélection pour la recombinaison. Plus l'épistasie est négative, plus la recombinaison est favorisée, surtout pour des taux d'auto-fécondation intermédiaires. Ces études suggèrent donc que l'accumulation de mutations délétères favorise la sélection pour la recombinaison (et non pas sa suppression) dans des populations avec des systèmes mixtes de reproduction. Il serait alors intéressant d'étudier d'une part le comportement des fluctuations aléatoires de systèmes similaires, en adoptant un point de vue stochastique plutôt que déterministe, et d'autre part de tester si ces résultats persistent en présence d'un locus en permanence hétérozygote, qui contre l'effet d'homozygotie augmentée sous auto-fécondation.

2.2 Introduire de la stochasticité

Une extension classique et naturelle consiste à modéliser la stochasticité des reproductions, pour pouvoir notamment étudier les déviations d'une population par rapport à son comportement moyen. Deux modèles sont historiquement utilisés pour cela : le modèle de Moran, où on remplace un individu à la fois dans la population, et le modèle de Wright-Fisher, où toute la population est remplacée en une fois (Etheridge, 2011). Ces modèles considèrent une population de taille finie et fixe, et permettent donc de rendre compte de la stochasticité

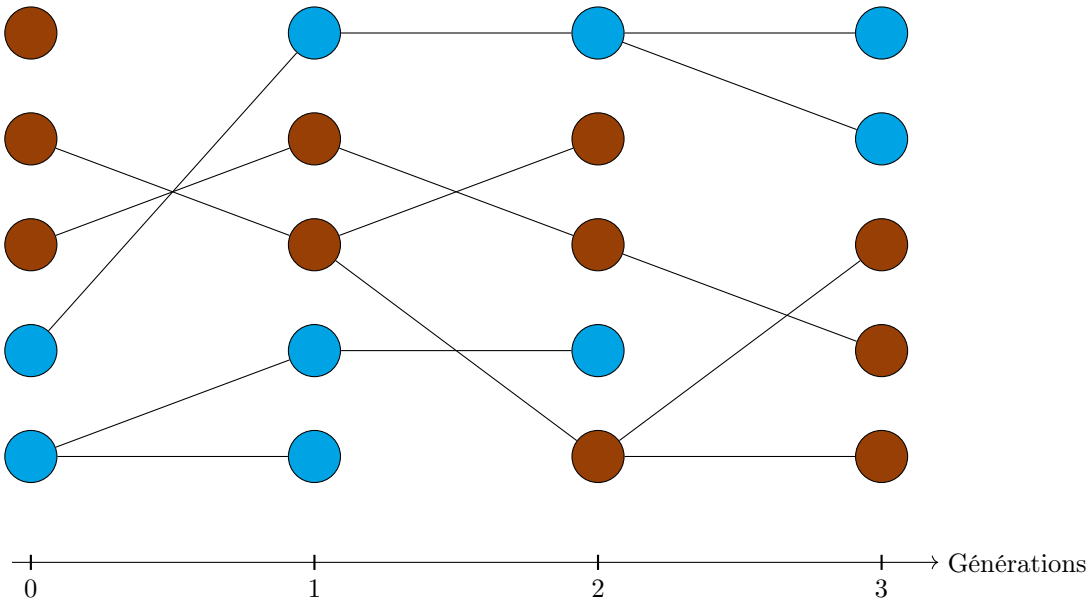


FIGURE 4 – Schéma d'évolution sur 4 générations d'une population de $N = 5$ individus, comportant deux allèles représentés par des couleurs (l'allèle bleu et l'allèle marron), et suivant la dynamique d'un processus de Wright-Fisher. À chaque génération, le parent de chaque individu est pioché uniformément au hasard parmi les individus de la génération précédente. L'enfant hérite alors de l'allèle de son parent.

inhérente à l'aléatoire de la reproduction. Ils suivent l'évolution de fréquences alléliques dans la population. On utilise un modèle de Wright-Fisher pour les simulations du chapitre 1, et un modèle de Moran au chapitre 2.

Modèle de Wright-Fisher et simulations au chapitre 1

Le modèle de Wright-Fisher considère une population de taille finie et fixe N et des individus haploïdes portant un gène qui possède deux allèles que l'on note A et a. Pour construire la population à la génération $n+1$, partant de la génération n, on considère que les parents de la génération n produisent un nombre infini de descendants potentiels, auxquels ils transmettent leur allèle, parmi lesquels on pioche uniformément au hasard les N individus constituant la génération $n+1$. De manière équivalente, on peut aussi considérer que pour chaque individu de la génération $n+1$, on choisit uniformément au hasard un parent dans la génération n. L'individu de la génération $n+1$ hérite alors de l'allèle du parent choisi (voir figure 4). Ainsi, conditionnellement au nombre x_n d'individus portant l'allèle A à la génération n, la distribution du nombre d'individus portant l'allèle A à la génération $n+1$, x_{n+1}, suit une loi binomiale de paramètres N et $\frac{x_n}{N}$. Ce point de vue permet d'étudier de nombreuses quantités d'intérêt, comme la probabilité et le temps moyen de fixation de l'allèle A dans la population, ou l'évolution de la variance de la fréquence des allèles (ETHERIDGE, 2011, MÉLÉARD, 2016).

De nombreuses extensions naturelles de ce modèle ont été développées pour incorporer de la sélection, de la mutation, ou encore de la reproduction sexuée et de la recombinaison. La complexité accrue du modèle empêche de calculer la plupart des quantités d'intérêt dans ces cas, mais le principe algorithmique du modèle de Wright-Fisher est encore utilisé pour des simulations individu-centrées.

Dans le chapitre 1, on utilise le logiciel SLiM (HALLER et MESSER, 2019) qui se base sur un algorithme de Wright-Fisher pour simuler une population d'individus diploïdes représentés par leurs chromosomes sexuels qui peuvent recombiner à un certain taux. Des mutations délétères sont présentes sur un nombre fini mais grand de sites, et une inversion (ou plus généralement un suppresseur de recombinaison) apparaît uniformément au hasard sur un haplotype. Les simulations permettent de suivre l'évolution de l'inversion dans la population, et de confronter les résultats aux attendus du modèle déterministe.

Modèle de Moran au chapitre 2 et processus Markovien de sauts du chapitre 3

Le modèle de Moran en temps discret suit la même logique que le modèle de Wright-Fisher, mais en remplaçant un seul individu à la fois. Un individu est choisi uniformément au hasard dans la population pour se reproduire et son descendant remplace un autre individu choisi uniformément au hasard. Dans le cas simple

sans sélection ni mutation, la probabilité de fixation d'un allèle peut être calculée explicitement et est la même que dans le cas du modèle de Wright-Fisher (Durrett, 2008).

Dans le chapitre 2, on utilise un modèle de Moran en temps continu. Cela signifie que les changements dans la population ont lieu à des temps aléatoires, et non plus à des générations prédéfinies. Formellement, chaque individu possède une "horloge", qui sonne au bout d'un temps aléatoire suivant une loi exponentielle de paramètre qui dépend du type de l'individu. Lorsqu'une horloge sonne, un évènement de reproduction se produit comme dans le cas du modèle discret : l'individu dont l'horloge a sonné se reproduit, et un individu est choisi uniformément au hasard dans la population pour être remplacé par le descendant produit. On étudie l'évolution de la composition génétique d'une population comportant des individus diploïdes représentés par une paire de chromosomes de type sexuel. On considère un locus de type sexuel toujours hétérozygote, et un locus de fardeau pouvant porter une mutation délétère. On introduit trois modes de reproduction possible, et on considère que des crossing-over peuvent se produire entre les deux loci. Un parent avec un certain génotype peut donc produire un descendant avec un génotype différent. Le faible nombre de génotypes possibles (quatre à cause de la contrainte d'hétérozygotie du locus de type sexuel) permet de calculer à la main les lois de reproduction de chaque génotype.

Ce modèle rentre dans la catégorie des processus markoviens de sauts en temps continu : en des temps aléatoires, la composition de la population change, donc la trajectoire décrivant la composition génétique de la population "saute". La distribution de la composition génétique après changement dépend uniquement de la composition de la population à l'instant du changement (propriété de Markov).

On suit le formalisme décrit dans Méléard, 2016 (partie 3.2) pour définir les taux de sauts et les lois de reproduction de notre processus. On utilise ensuite une approximation par processus de branchement (voir section suivante de cette introduction) pour analyser notre modèle.

Une des principales limites du modèle de Moran utilisé au chapitre 2 est de ne considérer qu'une seule mutation délétère possible, alors que la sélection pour ou contre la recombinaison est influencée par la présence de multiples mutations (Stetsenko et Roze, 2022, Jay et al., 2022b). Prendre en compte la possibilité d'évènements de mutation invite à considérer un nombre de types (ou génotypes) infini, voire continu. Dans le chapitre 3, on considère un espace de *traits** (ou caractères) continu multidimensionnel pour étudier l'interaction entre accumulation de mutations délétères et recombinaison. On utilise une fois de plus le formalisme des processus markoviens de sauts, dans le cadre particulier de processus à valeurs dans un espace de *mesures**. Une des premières définitions rigoureuses de ce type de processus pour répondre à une question biologique provient de Fournier et Méléard, 2004. Les auteurs considèrent une population d'individus repérés par leur position géographique, ce qui constitue donc une caractéristique continue. La composition de la population à chaque instant est repérée par une mesure sur l'espace des caractéristiques, de la forme

$$M_t = \sum_{i \in V_t} \delta_{x_t^i},$$

où V_t est l'ensemble des indices des individus présents dans la population au temps t et x_t^i la valeur au temps t de la caractéristique de l'individu i (dans cet exemple, c'est sa position géographique). $\delta_{x_t^i}$ est la *masse de Dirac** en la caractéristique x_t^i (voir le glossaire pour plus d'explications). Ce type de processus est donc appelé *processus de sauts à valeurs mesures**.

Dans le cadre du modèle de Fournier et Méléard, 2004, les individus peuvent se reproduire ou mourir. Le taux auquel les individus se reproduisent est constant, mais la caractéristique du descendant produit (c'est-à-dire sa position géographique) peut différer de la caractéristique du parent. On parle de reproduction *non locale*. Le taux auquel un individu meurt dépend de la caractéristique de l'individu, et du reste de la population. De nombreuses autres situations ont été explorées avec le même formalisme. Citons par exemple Tran, 2006 pour une population structurée par un trait phénotypique et prenant en compte le vieillissement, Champagnat et Méléard, 2007 pour un processus prenant en compte un trait phénotypique, une localisation spatiale, de la compétition et de la mutation, Coron, 2016 pour l'étude de l'accumulation de mutations délétères, Lambert, 2006 pour des diffusions branchantes, Collet et al., 2013 pour des populations se reproduisant sexuellement, Barton et al., 2013 pour des populations structurées spatialement, Etheridge et al., 2017 pour l'étude de zones hybrides. Une introduction complète à ce type de processus peut être trouvée dans Méléard et Bansaye,

2015.

Dans le chapitre 3, on se base sur Fournier et Méléard, 2004 et Tran, 2006 pour définir notre processus à valeurs mesures, calculer son générateur et obtenir une propriété classique de martingale.

On cherche ensuite à étudier le comportement en temps long de ce processus. Dans un premier temps, on cherche à appliquer des résultats de type "Many-to-One" décrits dans Marguet, 2019b, qui permettent de lier l'évolution moyenne de la population totale à l'évolution du trait moyen d'une lignée "typique" (voir le corps du chapitre 3 pour davantage de détails sur ces notions, et Marguet, 2017 pour une introduction et une définition plus complète de ce type résultats). Grâce à ces formules, on capture de l'information sur un processus à valeurs dans un espace de mesures (celui que l'on définit) par un processus à valeurs dans un espace de traits, donc plus simple. Cependant, la complexité de notre processus nous empêche de conclure avec ces résultats.

Dans un second temps, on cherche à appliquer un résultat récent de convergence démontré dans Bansaye et al., 2022, et appliqué dans Tomasevic et al., 2022, qui permettrait d'obtenir le comportement moyen du processus en temps long. Plus précisément, le résultat très général de Bansaye et al., 2022 porte sur la convergence en temps long de semi-groupes non-conservatifs, et peut être appliqué au semi-groupe de premier moment du processus markovien à valeurs mesures étudié (ces notions sont davantage détaillées dans le corps du chapitre 3). Ce semi-groupe de premier moment est non-conservatif car la taille de la population peut varier. Tomasevic et al., 2022 décrivent comment l'application de ce résultat permet d'obtenir un taux de croissance moyen et une distribution limite sur l'espace des traits pour leur processus. Dans leur cas, les individus sont décrits par un trait unidimensionnel. On tente d'appliquer la même méthode pour un trait multidimensionnel.

2.3 Processus de branchement pour l'étude d'un caractère nouvellement apparu dans une population

La dynamique des caractères nouvellement apparus dans une population parcourt trois phases (Durrett, 2008, partie 6.1.3). La première phase a lieu tant que le caractère est présent en très faible proportion dans la population. Ainsi, sa fréquence est fortement soumise à la dérive génétique. Ensuite, quand le caractère atteint une fréquence non négligeable et que la taille de la population est suffisamment grande, sa dynamique peut être approchée par un modèle déterministe. Une troisième phase, de nouveau stochastique, a lieu si le caractère approche la fixation. Cette phase est la symétrique de la phase 1, le caractère soumis à la dérive génétique étant le caractère sauvage présent au préalable dans la population.

On s'intéresse ici à la première phase stochastique d'un caractère nouvellement apparu dans une population, typiquement un suppresseur de recombinaison, ou une mutation délétère. Puisque le nombre de porteurs de ce caractère est faible dans cette première phase, on peut faire l'hypothèse qu'ils se reproduisent indépendamment les uns des autres. Cela fait naturellement apparaître un processus de branchement. Le processus de branchement permet de suivre l'évolution d'une population de taille non-fixée, dans laquelle les individus se reproduisent indépendamment les uns des autres. Cela permet notamment d'étudier les conditions de persistence ou d'extinction des populations.

Le plus simple de ces processus est celui de Galton-Watson, qui suit l'évolution de la taille d'une population composée d'individus d'un seul type en temps discret. Le passage de la génération n à la génération $n+1$ se déroule ainsi : les individus de la génération n produisent un nombre aléatoire de descendants, suivant une loi de reproduction aléatoire qui est la même pour tous les individus. La taille de la population à la génération $n+1$ est la somme des nombres de descendants de chaque individu de la génération n. Autrement dit, on a la relation de récurrence suivante :

$$Y_{n+1} = \sum_{k=1}^{Y_n} \xi_k,$$

où Y_{n+1} (resp. Y_n) est la taille de la population à la génération $n+1$ (resp. n) et les $(\xi_k)_{k\geq 1}$ sont des variables aléatoires indépendantes identiquement distribuées selon la loi de reproduction commune à tous les individus. La variable ξ_k représente le nombre de descendants du k^{ieme} individu présent à la génération n. De nombreuses extensions de ce processus classique ont été étudiées, comme par exemple la prise en compte de structure en âge, d'un environnement changeant qui agit sur la loi de reproduction, du répérage du type des individus par un trait continu (Harris, 1964, Athreya et Ney, 1972, Haccou et al., 2005), ou encore l'ajout de termes d'immigration (Heathcote, 1965), ou l'introduction de reproduction sexuée ou préférentielle (Karlin et Kaplan, 1973, Molina, 2010, Molina et al., 2014, Fritsch et al., 2022). Dans le chapitre 2, on considère

des individus de plusieurs génotypes possibles, et on souhaite suivre l'évolution du nombre d'individus de chaque génotype. La simplicité des génotypes considérés (deux loci, deux allèles) permet de ne suivre que trois types d'individus. On est donc en présence d'un processus de branchement multitype. La différence avec le processus de Galton-Watson unitype réside dans la loi de reproduction : Le nombre d'enfants de chaque type (ou génotype) produit par un individu dépend de son type. La formule de récurrence en temps discret devient donc, si on considère un nombre p de types,

$$\boldsymbol{Y}_{n+1} = \sum_{i=1}^{p} \sum_{k=1}^{Y_{n,i}} \boldsymbol{\xi}_{k,i},$$

où $\boldsymbol{Y}_{n+1}$ (resp. $\boldsymbol{Y}_n$) est maintenant un vecteur décrivant le nombre d'individus de chaque type à la génération $n+1$ (resp. n), $Y_{n,i}$ le nombre d'individus de type i présents à la génération n, et les $(\boldsymbol{\xi}_{k,i})_{1\leq k, 1\leq i\leq p}$ des variables indépendantes dont la distribution dépend du type i. La variable $\boldsymbol{\xi}_{k,i}$ est un vecteur qui donne le nombre d'enfants de chaque type produit par le k^{ieme} individu de type i présent dans la population à la génération n. On choisit par ailleurs dans le chapitre 2 de se placer dans le cadre du temps continu. Les évènements de reproduction se produisent alors un par un à des temps aléatoires, suivant le même principe d'horloge décrit pour le modèle de Moran en temps continu. Les processus de branchement multitype en temps continu sont bien connus (Harris, 1964, Kesten et Stigum, 1966, Mode, 1971, Athreya et Ney, 1972, Sewastjanow, 1975). Les résultats appliqués dans le chapitre 2 sont énoncés dans l'introduction de Pénisson, 2010, vers laquelle on renvoie le lecteur pour une définition précise et complète des processus de branchement multitype en temps continu. On s'intéresse en particulier à la probabilité et au temps d'extinction du processus, c'est-à-dire à la probabilité et au temps de purge de la mutation délétère.

On adopte également un point de vue de processus de branchement en temps continu au chapitre 3, pour étudier le maintien de certains génotypes dans une population. On s'intéresse cette fois à des processus de branchement unitypes, desquels on compare la criticité.

3 Résumé des chapitres

3.1 Chapitre 1 : Sélection de la suppression de recombinaison dans une population avec reproduction panmictique, et locus sexuel toujours hétérozygote

Ce chapitre a fait l'objet d'une publication dans le journal PLOS Biology : Jay *et al., 2022b*

Dans ce chapitre, on cherche à étudier si la suppression de recombinaison peut être sélectionnée au-delà des loci impliqués dans la détermination du sexe ou du type sexuel. Pour cela, on met en place deux modèles mathématiques sur une même base. On considère des individus diploïdes, avec un locus sexuel ou de type sexuel. On considère également un nombre grand mais fini de sites non liés pouvant porter des mutations récessives et délétères. Les mutations ont un effet multiplicatif sur la fitness (ou valeur sélective) des individus. Un site homozygote pour les mutations délétères contribue d'un facteur $1-s$, un site hétérozygote contribue d'un facteur $1-hs$, où s est le coefficient de sélection et h le coefficient de dominance. Un individu possédant m_1 sites hétérozygotes et m_2 sites homozygotes pour les mutations délétères a donc une valeur sélective égale à $(1-hs)^{m_1}(1-s)^{m_2}$. Les individus se reproduisent de manière *panmictique**. On modélise l'apparition aléatoire d'une inversion, donc un suppresseur de recombinaison, sur un nombre n de sites. L'inversion n'est soumise à la sélection que via les mutations délétères qu'elle porte, elle n'a pas de valeur sélective intrinsèque. Le but est d'étudier la possibilité de fixation de telles inversions, donc de suppression de recombinaison.

Modèle déterministe. Dans un premier temps, on utilise un modèle déterministe, qui permet d'étudier la tendance moyenne attendue pour la dynamique de la suppression de recombinaison. On se base sur les modèles de génétique de populations classiques décrits à la partie 2.1 de cette introduction. On suit l'évolution de la fréquence de l'haplotype "Inversé", noté I, par rapport à celle de l'haplotype "Non inversé", noté N. On considère que tous les sites sont indépendants et que chaque site porte une mutation délétère avec probabilité q, qui correspond à la fréquence de l'équilibre mutation-sélection calculée pour un locus (voir la section Methods de l'article pour le calcul de cette quantité q). On suppose de plus que de nouvelles mutations ne peuvent pas apparaître ultérieurement. La fitness de l'haplotype inversé dépend donc seulement du nombre de mutations

m	0	$nq\frac{h}{1-h}$	nq	n
Fréquence d'équilibre F_I	1	$\frac{W_{NN}-W_{NI}}{W_{II}+W_{NN}-2W_{NI}}$		0
Fréquence d'équilibre F_{YI}	1			0

FIGURE 5 – Fréquences d'équilibre de l'haplotype inversé en fonction du nombre m de mutations délétères portées par l'inversion (première ligne), et de la liaison avec un allèle en permanence hétérozygote, noté ici Y (F_I pour des inversions non-liées au Y, F_{YI} pour des inversions liées au Y). n est la taille totale du segment recouvert par l'inversion (en nombre de paires de bases) ; q la probabilité de trouver une mutation délétère sur un site lorsque la population est à l'équilibre mutation/sélection ; h le coefficient de dominance ; W_{NN} la fitness d'un individu diploïde ne portant pas l'inversion ; W_{NI} la fitness d'un individu diploïde hétérozygote pour l'inversion ; W_{II} la fitness d'un individu diploïde homozygote pour l'inversion.

délétères englobées par l'inversion au moment de son apparition, que l'on note m. Cette hypothèse forte traduit l'idée que l'on considère dans un premier temps que l'échelle de temps sur laquelle apparaissent les mutations délétères est beaucoup plus grande que celle sur laquelle une inversion peut se fixer dans la population. On relâchera cette hypothèse dans la deuxième partie de ce chapitre, lorsqu'on utilisera des simulations individu-centrées.

On considère différents systèmes de reproduction, détaillés dans l'annexe de l'article. On se concentre ici et dans le corps de l'article sur le cas du système XY avec individus mâles et femelles : pour former un nouvel individu, un gamète produit par un individu mâle fusionne avec un gamète produit par un individu femelle. On souhaite en particulier comparer l'évolution de la fréquence d'une inversion liée à l'allèle Y, notée F_{YI}, à l'évolution de la fréquence d'une inversion non-liée à l'allèle Y, notée F_I. Dans le premier cas, l'inversion est liée à un allèle en permanence hétérozygote et donc maintenue à l'état hétérozygote.

Les variations de fréquence entre la génération t et la génération $t+1$ sont données par

$$\Delta F_{YI} = \frac{F_{YI}(1-F_{YI})(W_{NI}-W_{NN})}{\overline{W_m}}, \tag{1}$$

et

$$\Delta F_I = \frac{F_I}{\overline{W}} P(F_I), \tag{2}$$

où W_{NI}, W_{II}, W_{NN} représentent les fitness moyennes des génotypes (diploïdes) hétérozygotes pour l'inversion, homozygotes pour l'inversion, et ne portant pas l'inversion, $\overline{W_m}$ et $\overline{W}$ la fitness moyenne des individus mâles et de la population totale, et P est un polynôme de degré 2 (voir la section Methods de l'article).

Les valeurs de F_{YI} et F_I annulant les second-membres respectifs de ces équations donnent les fréquences d'équilibre. La convergence en temps long de chaque fréquence vers l'un de ces équilibres est obtenue en étudiant le signe de ΔF_{YI} et ΔF_I. On obtient des conditions sur la charge initiale de l'inversion en mutations délétères pour qu'elle soit favorisée, et puisse être maintenue dans la population à une fréquence d'équilibre. Les résultats sont résumés dans la Figure 5, et montrent qu'une inversion liée au locus en permanence hétérozygote (haplotype YI) atteint une fréquence d'équilibre de 1 dans la sous-population des chromosomes Y lorsqu'elle porte moins de mutations que la moyenne (nq) sur le même segment dans la population. En revanche, une inversion non-liée au Y atteint une fréquence d'équilibre intermédiaire, qui correspond à la fréquence d'équilibre d'un allèle super-dominant. En fait, l'inversion est maintenue à l'état hétérozygote lorsqu'elle est liée à l'allèle Y, ce qui limite l'expression des mutations délétères portées. Si la fitness de l'inversion à l'état hétérozygote est meilleure que la fitness moyenne des individus non-inversés, ce qui est le cas lorsque l'inversion porte moins de mutations que la moyenne, alors l'inversion se fixe à l'état hétérozygote. Si l'inversion n'est pas liée au Y, elle peut se retrouver à l'état homozygote, et les mutations qu'elle porte sont alors également exprimées. Ces résultats préliminaires suggèrent donc que la présence d'un allèle en permanence hétérozygote peut favoriser l'arrêt de la recombinaison à son voisinage.

Simulations stochastiques. Un segment qui ne recombine pas purge moins efficacement les mutations délétères qu'il contient. Une inversion est donc vouée à accumuler des mutations délétères suivant un cliquet de Muller, ce qui peut réduire sa fitness et empêcher sa fixation. On utilise alors des simulations stochastiques basées sur un modèle de Wright-Fisher pour étudier la dynamique d'une inversion dans le même cadre que le précédent modèle déterministe, mais en rajoutant la possibilité d'accumulation de mutations délétères. En d'autres termes, on relâche l'hypothèse précédente de comparaison d'échelles de temps, et on considère maintenant que l'apparition de mutations se déroule sur une échelle de temps plus courte que la fixation d'une inversion. La population est simulée avec le logiciel SLiM (Haller et Messer, 2019). Les individus sont toujours diploïdes, représentés par une paire de chromosomes dont l'un porte un allèle en permanence hétérozygote. Des mutations délétères ségrégent sur un nombre fini de sites. Une inversion d'une certaine taille (pré-définie, mais que l'on fait varier suivant les simulations) apparaît au hasard sur un chromosome dans la population. Dans un premier temps, on simule des trajectoires avec l'introduction d'une seule inversion, et on étudie la proportion d'inversions fixées ou perdues parmi ces trajectoires. On compare la dynamique des inversions liées au locus en permanence hétérozygote (Y) à celle des inversions non-liées au Y (qu'on appelle autosomales). On observe deux scenarii possibles : soit l'inversion se fixe à l'état hétérozygote (par exemple, en étant liée à l'allèle Y) avant que l'accumulation de mutations délétère ne fasse trop diminuer sa fitness, soit l'accumulation de mutations délétères est trop lourde, et l'inversion est purgée de la population avant d'avoir pu se fixer. On observe en particulier que les inversions autosomales (c'est-à-dire non liées au Y) finissent toujours par être purgées, alors que certaines inversions liées au Y parviennent à se fixer dans la sous-population des chromosomes portant le Y. Contrairement aux résultats du modèle déterministe, il n'y a plus de fréquence intermédiaire pour les inversions autosomales : si l'inversion ne se fixe pas (fréquence strictement plus petite que 1), elle continue à accumuler des mutations et finit par être contre-sélectionnée.

Ces simulations montrent que les inversions liées au Y parvenant à la fixation bénéficiaient d'une valeur sélective élevée à l'état hétérozygote au moment de leur apparition, par exemple parce qu'elles portaient moins de mutations délétères que la moyenne sur le même segment, ou que les mutations portées étaient plus rares (et donc moins susceptibles de se retrouver à l'état homozygote). De plus, les simulations montrent qu'augmenter la taille de l'inversion augmente ses chances de fixation. En effet, en augmentant le nombre de sites englobés, on augmente la variance de fitness sur cette région dans la population, et l'inversion a ainsi plus de chances de capturer un haplotype favorisé par rapport à la moyenne. Enfin, un petit coefficient de dominance h induit une fixation plus probable des inversions liées au locus Y. En effet, réduire la valeur du coefficient de dominance h réduit l'expression délétère des mutations à l'état hétérozygote, et augmente encore le différentiel de fitness en faveur d'une inversion maintenue à l'état hétérozygote.

Dans un second temps, on simule des trajectoires avec possibilité d'apparition successive de plusieurs inversions, afin d'étudier l'extension de la suppression de recombinaison. En particulier, on introduit la possibilité de réversion (c'est-à-dire de rétablissement de la recombinaison) après fixation. Un exemple de résultat est présenté dans la figure 4 de l'article du chapitre 1, et montre que des fixations successives d'inversions autour de l'allèle Y sont observables. Cela suggère donc que l'extension de la suppression de recombinaison au-delà des loci sexuels ou de type sexuels peut être provoquée par la dynamique des mutations délétères.

En résumé : La suppression de recombinaison peut être étendue par étapes successives autour d'un locus en permanence hétérozygote, si l'haplotype non-recombinant i) possède un avantage intrinsèque (dû à un faible nombre de mutations délétères) et ii) est favorisé à l'état hétérozygote (par l'abritement des mutations délétères). Cette sélection pour la suppression de recombinaison est d'autant plus forte que de nombreuses mutations délétères ségrègent dans la population.

3.2 Chapitre 2 : Dynamique des mutations délétères au voisinage d'un locus de type sexuel pour des individus se reproduisant par auto-fécondation

Ce chapitre a fait l'objet d'une recommendation PCI EvolBiol (https://doi.org/10.24072/pci.evolbiol.100635) et d'une publication dans le Peer Community Journal : Tezenas *et al., 2022*

Les résultats obtenus dans le chapitre 1 montrent que le sort d'une inversion dépend beaucoup du paysage de mutations délétères dans lequel elle apparaît. Or, de nombreux facteurs peuvent impacter la dynamique

des mutations délétères, comme le système de reproduction : l'auto-fécondation augmente l'homozygotie et permet donc une purge plus efficace des allèles délétères. Mais la présence d'un locus de type sexuel toujours hétérozygote augmente l'hétérozygotie dans son voisinage, et peut donc venir contrer l'effet de l'auto-fécondation sur la dynamique des mutations délétères.

Modèle. Dans ce deuxième chapitre, on se concentre donc sur la dynamique des mutations délétères au voisinage d'un locus de type sexuel (donc toujours hétérozygote), dans une population d'individus pouvant se reproduire par auto-fécondation ou allo-fécondation. La plupart des espèces de champignons chez lesquelles une extension de suppression de recombinaison a été observée présentent deux modes d'auto-fécondation : l'auto-fécondation intra-tétrade, lors de laquelle les deux gamètes nécessaires à la formation d'un nouvel individu sont piochés dans une seule tétrade (*i.e.* un seul produit de méiose), et l'auto-fécondation inter-tétrade, lors de laquelle les deux gamètes sont piochés dans des tétrades différentes, produites par le même individu. On modélise ces trois modes de reproduction (un schéma est présent en Annexe B de l'article de ce chapitre).

On considère des individus diploïdes, portant deux loci avec chacun deux allèles : un locus de type sexuel, toujours hétérozygote, et un locus de fardeau, pouvant porter un allèle délétère récessif ou super-dominant. La fitness d'un individu est déterminée par son génotype à ce locus de fardeau. Les individus peuvent donc se reproduire en allo-fécondation, ou auto-fécondation intra- ou inter-tétrade, et un évènement de recombinaison peut se produire à un certain taux fixe entre les deux loci au moment de la production d'une tétrade (au moment de la méiose, voir Figure 1 de cette introduction). Le taux de recombinaison est donc un paramètre du modèle et non une variable. Ce modèle permet donc d'étudier la dynamique d'une mutation délétère en fonction de sa proximité génétique à un locus de type sexuel, mais pas d'étudier l'évolution du taux de recombinaison, contrairement aux modèles du chapitre 1.

On utilise un modèle de Moran (donc stochastique) en temps continu et multitype. La taille de la population N est fixe. Du fait de la contrainte d'hétérozygotie au locus de type sexuel, seulement quatre génotypes sont observables, et on peut calculer à la main les lois de reproduction de chaque type. On suit le nombre d'individus de chaque type contenus dans la population au temps t via le vecteur $g(t) = (g_1(t), g_2(t), g_3(t), g_4(t))$. On considère que les individus se reproduisent à taux 1 (donc au bout d'un temps aléatoire suivant une loi exponentielle de paramètre 1). Le taux $T_g(+G_i)$ auquel un individu de type i est produit dans une population décrite par g tient compte des trois modes de reproduction (allo-fécondation et auto-fécondation intra- ou inter-tétrade) et de la possibilité de recombinaison. La sélection opère au stade juvénile : le descendant produit a une certaine probabilité de survie S_i, qui est déterminée par son génotype. S'il survit, un individu est pioché au hasard dans la population pour être remplacé par le nouvel individu. Un changement dans la composition de la population consiste donc en un gain d'un individu de type i et la perte d'un individu de type j, ce qui se produit à taux

$$Q_{i,j}(g) = T_g(+G_i)S_i\frac{g_j}{N}.$$

Approximation par un processus de branchement. Comme expliqué en partie 2.3 de cette introduction, lorsque la mutation délétère apparaît dans la population, elle est portée par très peu d'individus. On peut alors considérer que le nombre d'individus porteurs de la mutation délétère, c'est-à-dire de types 1, 2 et 3, est très petit devant la taille de population. Au contraire, la nombre d'individus non porteurs de la mutation délétère, c'est-à-dire de type 4, est de même ordre de grandeur que la taille de la population. Plus précisément, on suppose que

$$g_4 \approx N, \quad \text{et que } g_i \ll N \text{ pour } i = 1, 2, 3.$$

En injectant cette approximation dans les taux $Q_{i,j}$ du processus de Moran précédemment défini, les termes d'allo-fécondation entre individus porteurs de l'allèle délétère deviennent négligeables. Cela rend les lois de reproduction des génotypes porteurs de l'allèle délétère indépendantes de la composition de la population. On peut alors approcher la dynamique de la mutation délétère par un processus de branchement multitype en temps continu, $(Z_t)_{t\geq 0} := \left(Z_{t,1}, Z_{t,2}, Z_{t,3}\right)_{t\geq 0}$. Ce processus de branchement nouvellement obtenu décrit le nombre d'individus de trois types, les trois génotypes portant la mutation délétère (deux hétérozygotes et un homozygote, correspondant aux types 1, 2, et 3). Les taux et lois de reproduction de ce nouveau processus sont obtenus à partir des taux $Q_{i,j}$ du processus de Moran en faisant tendre la taille de la population N vers l'infini.

Le nombre fini de types permet d'utiliser des méthodes matricielles pour l'étude de ce processus de bran-

chement (Mode, 1971, Pénisson, 2010). Si on note z_0 la condition initiale, alors il existe une matrice C, indépendante de z_0, telle que

$$\mathbb{E}[Z_t|Z_0 = z_0] = z_0 \mathrm{e}^{Ct} \tag{3}$$

pour tout $t \geq 0$. Les coefficients de cette matrice C sont calculés à partir des lois de reproduction de chaque type. L'étude des valeurs propres de cette matrice C (de dimension 3 dans notre cas) permet d'obtenir des informations sur l'extinction du processus, donc sur la purge de la mutation délétère. Formellement, plus la valeur propre dominante est grande, et plus la mutation a des chances de ne pas être purgée, et plus le temps de purge est long lorsqu'elle est purgée.

Etude analytique du processus de branchement. Nous obtenons une expression explicite pour la valeur propre dominante de la matrice C. Lorsque la mutation est super-dominante, c'est-à-dire si la fitness des hétérozygotes est plus élevée que la fitness des deux homozygotes (homozygotes pour la mutation et homozygotes pour l'allèle sauvage), le grand nombre de paramètres du modèle empêche d'étudier son signe analytiquement. On montre cependant numériquement qu'il existe des régions de valeurs de paramètres pour lesquels la valeur propre dominante est positive. Cela signifie que la probabilité que le processus ne s'éteigne pas est non nulle, et donc que la mutation peut ne pas être purgée de la population. Cela confirme les résultats du chapitre 1 : si une inversion possède un avantage hétérozygote, elle est favorisée dans la phase initiale de son évolution. Lorsque la mutation est délétère et récessive, nous montrons analytiquement que la valeur propre dominante est toujours strictement négative, sauf lorsque la mutation est neutre, ou qu'elle est totalement récessive et que le taux de recombinaison est nul (auquel cas la mutation est comme neutre, puisque maintenue à l'état hétérozygote et neutre dans cet état). Sauf dans ces cas extrêmes peu intéressants biologiquement, cela signifie que le processus de branchement est sous-critique, et donc que la probabilité d'extinction est égale à 1. Une mutation délétère récessive est donc purgée avec probabilité 1.

On raffine l'analyse en étudiant l'impact de la présence du locus de type sexuel et du mode de reproduction. On montre que la valeur propre augmente (en restant négative) lorsque la recombinaison diminue. Donc augmenter la liaison entre les deux loci augmente le maintien de la mutation délétère dans la population, on peut parler d'*effet d'abritement* du locus de type sexuel (voir le paragraphe 3.2 de l'article). On montre également que cet effet est d'autant plus fort que le taux d'auto-fécondation intra-tétrade est élevé.

Etude numérique. On complète l'analyse en réalisant des simulations du processus de branchement.

Dans un premier temps, on étudie le temps d'extinction dans des cas sous-critiques, c'est-à-dire le temps de purge d'une mutation délétère et récessive. Les simulations montrent que la plupart des trajectoires s'éteignent rapidement, mais que certaines (de l'ordre de 1%) persistent très longtemps : la mutation est toujours purgée, mais la purge peut être parfois très longue (Figure 4). Le temps d'extinction est par ailleurs rallongé lorsque la liaison entre les loci est augmentée (lorsqu'on diminue le taux de recombinaison). Cela suggère que des mutations délétères peuvent s'accumuler au voisinage d'un locus en permanence hétérozygote, créant un contexte favorable à la suppression de recombinaison comme observée au chapitre 1.

Une dernière analyse des simulations nous permet d'obtenir la probabilité qu'une mutation soit maintenue assez longtemps dans la population pour qu'une seconde mutation ait le temps d'apparaître, constituant ainsi une première étape vers l'accumulation de mutations délétères au voisinage d'un locus de type sexuel. On teste en particulier l'impact de l'auto-fécondation sur cette accumulation de mutations délétères. Les résultats montrent d'abord que la présence d'un locus de type sexuel a peu d'impact pour des populations avec un faible taux d'auto-fécondation. Ensuite, on observe que la présence d'un locus de type sexuel neutralise les effets de l'auto-fécondation à son voisinage. En effet, pour des régions peu liées au locus de type sexuel (taux de recombinaison élevé), la probabilité d'accumulation de mutations est plus faible en auto-fécondation qu'en allo-fécondation. Cependant, augmenter la liaison entre les deux loci (*i.e.* réduire le taux de recombinaison) ramène la probabilité d'accumulation sous auto-fécondation au même niveau que la probabilité d'accumulation sous allo-fécondation (Figure 5).

Ces résultats suggèrent donc que la présence d'un locus de type sexuel en permanence hétérozygote contre l'effet de purge de mutations délétères induit par l'auto-fécondation, ce qui induit l'existence d'évènements rares d'accumulation de mutations délétères au voisinage de ces loci, ce qui peut ensuite provoquer la sélection pour la suppression de recombinaison afin de limiter l'effet délétère des mutations moins bien purgées. En particulier,

l'utilisation de méthodes stochastiques permet d'observer des évènements rares que l'on ne peut pas détecter avec des méthodes déterministes.

En résumé : La présence d'un locus en permanence hétérozygote réduit l'effet de l'autofécondation en limitant la purge des mutations délétères récessives dans son voisinage.

3.3 Chapitre 3 : Dynamique de la suppression de recombinaison sous auto-fécondation et avec possible accumulation de mutations délétères

Une des principales limitations du modèle développé au chapitre 2 est qu'il ne prend en compte qu'une seule mutation délétère, alors que la sélection pour ou contre la suppression de recombinaison semble dépendre de la présence de multiples mutations. La méthode mise en oeuvre au chapitre 2 (calcul à la main des lois de reproduction) n'est pas pertinente pour élargir l'étude à de multiples mutations. En effet, chaque ajout de site augmente le nombre de génotypes possibles et nécessite de recalculer toutes les lois de reproduction à la main.

On développe donc dans ce troisième chapitre, exploratoire, un modèle permettant de prendre en compte l'accumulation de mutations délétères sur des chromosomes de type sexuel. On cherche à obtenir des résultats analytiques pour appuyer les résultats des simulations stochastiques du chapitre 1. Le but est donc de construire un modèle permettant d'une part de suivre un grand nombre de mutations, tout en différenciant si elles se trouvent à l'état hétérozygote ou homozygote, et d'autre part de modéliser une reproduction sexuée avec la contrainte d'un type sexuel toujours hétérozygote et des évènements de recombinaison possibles.

Modèle. On considère des individus diploïdes représentés par une paire de chromosomes de type sexuel. On distingue deux types sexuels : un individu est toujours constitué d'un chromosome portant le type sexuel a et d'un chromosome portant le type sexuel b. On se concentre sur le cas où les individus ne se reproduisent que via auto-fécondation intra-tétrade, puisque les résultats du chapitre 2 ont montré que c'était sous ce régime de reproduction que la présence d'un locus de type sexuel avait le plus d'impact sur la dynamique des mutations délétères. Cela nous permet de considérer les reproductions des individus de façon indépendante. On suppose que chaque individu se reproduit à taux 1. Pour suivre les mutations délétères, on choisit d'adopter un point de vue continu : on considère un nombre de sites infini, et on suit la proportion de sites occupés par une mutation délétère. Ainsi, on décrit le génotype d'un individu diploïde par un vecteur $y = (x_a, x_b, \omega)$ de trois coordonnées dans $[0, 1]$: deux coordonnées donnant la proportion de chaque chromosome de type sexuel occupé par des mutations délétères (x_a et x_b), et une coordonnée donnant la proportion de mutations à l'état homozygote entre ces deux chromosomes (ω). Cette modélisation permet de suivre l'évolution de la charge mutationnelle hétérozygote et homozygote des individus, ce qui permet de détecter un éventuel avantage hétérozygote comme celui observé pour les inversions du chapitre 1. Implicitement, on suppose que les mutations sont réparties uniformément sur le chromosome, et on ne peut donc pas suivre la position exacte des mutations sur le génome. On ne cherchera donc pas à étudier la dynamique des mutations délétères en fonction de leur proximité au locus de type sexuel, mais plutôt à comparer cette dynamique entre des populations où les individus peuvent recombiner ou non.

On considère deux catégories d'individus : ceux pouvant recombiner, et ceux ne le pouvant pas. Afin de les distinguer, on rajoute une coordonnée R à la description du trait d'un individu. Si $R = 0$ la recombinaison est bloquée entre les chromosomes de cet individu. Si $R = 1$, elle peut se produire. Un individu est donc totalement décrit par son trait $x = (R, y) = (R, x_a, x_b, \omega)$, qui contient l'information sur sa capacité à recombiner ou non, et sur sa charge mutationnelle. On note $\mathcal{X} = 0, 1 \times [0, 1]^3$ l'espace des traits.

Lorsque la recombinaison est bloquée chez un individu, les mutations délétères sont complètement liées entre elles et au locus de type sexuel. Lorsque la recombinaison est autorisée chez un individu, elle se produit lors de la méiose avec une probabilité r. Encore une fois, dans notre point de vue continu, on raisonne en termes de proportion plutôt qu'en termes de localisation précise : deux chromatides peuvent échanger une proportion θ aléatoire de leur génome. Puisqu'on suppose que les mutations délétères sont réparties uniformément le long du chromosome, on calcule alors la charge en mutation délétères des chromosomes recombinés en sommant les contributions de chaque chromosome recombinant (voir Figure 1 du chapitre 3). Cette façon de modéliser la recombinaison reste proche de la réalité biologique : en auto-fécondation, la recombinaison ne peut pas faire diminuer l'homozygotie d'un descendant par rapport à son parent, mais peut faire diminuer l'hétérozygotie.

On considère que la sélection s'exerce au moment de la mort des individus. La fitness d'un individu est modélisée par son taux de mort $d(y)$, qui dépend du génotype, c'est-à-dire de la charge en mutations délétères à l'état hétérozygote et homozygote (voir équation (1) du chapitre 3).

On inclut également la possibilité que des mutations délétères continuent à s'accumuler, en considérant des évènements de mutation. Toujours dans le cadre de notre point de vue continu, on suppose qu'un évènement de mutation se produit à taux μ, et ajoute une petite proportion de mutations délétères à un des deux chromosomes de type sexuel, augmentant alors la valeur de la coordonnée correspondante, ainsi que la valeur de la coordonnée d'homozygotie (voire Figure 2 du chapitre 3).

La dynamique mathématique est alors la suivante :

- Chaque individu se reproduit à taux 1.
  - Si l'individu ne peut pas recombiner ($R = 0$), alors la reproduction est clonale et un individu de même génotype est ajouté à la population.
  - Si l'individu peut recombiner, alors un évènement de recombinaison a lieu pendant la méiose avec probabilité r. Le descendant produit a ensuite l'un des génotypes suivants avec probabilité 1/4 pour chacun : (x_a, x_b, ω), $(x_a, \theta x_b + (1-\theta)x_a, \theta\omega + (1-\theta)x_a)$, $(\theta x_a + (1-\theta)x_b, x_b, \theta\omega + (1-\theta)x_b)$, ou $(\theta x_a + (1-\theta)x_b, \theta x_b + (1-\theta)x_a, \omega)$, où θ représente la proportion de chromatides échangée lors de la recombinaison (voir Figure 1 du chapitre 3).
- Chaque individu peut muter à taux μ. Son génotype devient alors $(R, x_a + \varepsilon(1-x_a), x_b, \omega + \varepsilon(x_b - \omega))$ ou $(R, x_a, x_b + \varepsilon(1-x_b), \omega + \varepsilon(x_a - \omega))$, chacun avec probabilité 1/2. Le paramètre ε quantifie la quantité de mutations délétères ajoutée à chaque évènement de mutation.
- Chaque individu meurt à taux $d(x_a, x_b, \omega) := s_0 + s_{hom}\omega + s_{het}(x_a + x_b - 2\omega)$, où s_0 est le taux de mort intrinsèque, s_{hom} le coefficient de sélection quantifiant l'effet délétère des mutations à l'état homozygote, et s_{het} le coefficient de sélection quantifiant l'effet délétère des mutations à l'état hétérozygote.

Le but est de comparer les sous-populations d'individus recombinants et non-recombinants. On s'intéresse d'une part à la démographie de chaque sous-population, et d'autre part à l'évolution au cours du temps de la charge en mutations délétères, à l'état hétérozygote et homozygote, dans chaque sous-population. Si un caractère (recombinant ou non-recombinant) est plus favorable que l'autre, on s'attend à observer une différence dans la démographie (par exemple, un taux de croissance différent). La distribution de la charge mutationnelle peut également différer au sein de chaque sous-population. Par exemple, on pourrait observer l'effet d'abritement du type sexuel dans la population non recombinante, qui induirait une forte proportion de mutations à l'état hétérozygote, mais une faible proportion de mutations à l'état homozygote.

On décrit l'évolution de la population par un processus de sauts à valeurs mesures en temps continu noté $(\mathcal{Z}_t)_{t\geq 0}$. Á chaque instant $t > 0$, la population est décrite par une somme de *masses de Dirac** situées en les traits des individus présents dans la population à cet instant :

$$\mathcal{Z}_t = \sum_{i \in V_t} \delta_{x_t^i},$$

où V_t est l'ensemble des indices des individus présents dans la population au temps t, et x_t^i le trait de l'individu i au temps t.

En particulier, puisque notre hypothèse de reproduction exclusivement par auto-fécondation nous permet de considérer des reproductions indépendantes, notre processus rentre dans le cadre de processus de branchement à valeurs mesures dont la théorie est aujourd'hui bien développée (Méléard et Bansaye, 2015, Fournier et Méléard, 2004, Tran, 2006). On utilise dans un premier temps ce formalisme classique pour montrer la bonne définition de notre processus, calculer son générateur, et obtenir une propriété de martingale (partie 2 du chapitre 3).

Etude analytique du modèle : convergence de la mesure moyenne. Dans un deuxième temps, on cherche à étudier la démographie des sous-populations et la distribution des mutations délétères en temps long. Une approche naturelle consiste à se concentrer sur la *mesure moyenne*, n_t, vérifiant pour toute fonction $f : \mathcal{M}_F(\mathcal{X}) \to \mathbb{R}$ mesurable et bornée sur l'espace des mesures finies sur $\mathcal{X}$ (l'espace des traits) :

$$\langle n_t, f \rangle = \mathbb{E}[\langle \mathcal{Z}_t, f \rangle],$$

où on utilise la notation $\langle \nu, f \rangle = \int_{\mathcal{X}} f(x)\nu(\mathrm{d}x)$ pour l'intégrale de la fonction f contre une mesure ν sur l'espace des traits $\mathcal{X}$. Cette mesure moyenne rend compte du comportement moyen du processus. Plus précisément, on cherche à appliquer le résultat de Bansaye et al., 2022 pour montrer que la mesure n_t se comporte asymptotiquement au premier ordre comme

$$C\mathrm{e}^{\lambda t} N(\mathrm{d}x),$$

où C est une constante que l'on sait expliciter, λ un réel et N une mesure de probabilité sur l'espace des traits $\mathcal{X}$. Cette mesure dans laquelle le temps (t) et l'espace $(x$, le trait, qui contient l'information sur la charge en mutations délétères) sont décorrélés nous donne de l'information sur la démographie, via le taux de croissance λ, et sur la distribution asymptotique de la charge en mutations délétères, via la mesure N. On peut ensuite comparer les éléments propres λ et N pour les sous-populations d'individus recombinants et non-recombinants. Plusieurs limitations nous empêchent cependant de conclure avec cette approche (décrites dans le chapitre 3).

Au-delà des limitations techniques rencontrées, une limitation structurelle apparaît. En effet, la charge mutationnelle homozygote d'un génome ne peut qu'augmenter (via la mutation ou la recombinaison), et pas diminuer. Cela s'apparente à un cliquet de Muller en version continue : la charge mutationnelle homozygote minimale augmente au fur et à mesure que les génotypes les moins chargés sont perdus par dérive génétique. Mathématiquement, cela signifie que le trait des individus d'une population (la proportion des chromosomes occupés par des mutations délétères) est tiré vers le point $(x_a, x_b, \omega) = (1, 1, 1)$. Or, une des hypothèses du théorème de convergence de Bansaye et al., 2022 requiert l'existence d'un ensemble compact dans lequel les trajectoires sont récurrentes. Autrement dit, quand on part d'un point de ce compact, l'hypothèse demande que la trajectoire issue de ce point et suivant la dynamique du processus $(\mathcal{Z}_t)_{t\geq 0}$ revienne dans le compact avec probabilité strictement positive. Le comportement particulier de notre processus impose a minima qu'un ensemble vérifiant cette hypothèse contienne le point $(1, 1, 1)$, mais pourrait même empêcher l'existence d'un tel ensemble. Nous n'avons pas élucidé cette question au cours de cette thèse, mais ce comportement particulier suggère un dernier axe d'analyse pour étudier la dynamique de mutations délétères en interaction avec la suppression de recombinaison.

Processus de branchement unitype pour étudier la persistence d'un génotype. Puisque la population subit cette sorte de cliquet de Muller continu, on s'attend à observer deux cas. Dans le premier cas, les individus faiblement chargés en mutations se reproduisent suffisamment vite pour empêcher le cliquet de cliquer, et maintiennent une population d'individus ayant leur trait. Cette sous-population *locale** (au sens de localisée en un trait particulier) peut alors devenir une source pour la population globale en contribuant aux sous-populations d'individus plus chargés en mutations, via la mutation ou la recombinaison. Dans le second cas, les individus faiblement chargés en mutations meurent ou accumulent trop vite de nouvelles mutations, et la population s'éteint localement.

On cherche à étudier cette dynamique d'extinction ou de persistance locale, et donc à suivre l'évolution du nombre d'individus ayant un trait fixé. Une fois de plus, on adopte un point de vue de processus de branchement. Pour chaque trait $x_0 := (R_0, y_0)$ (chaque génotype) possible, on construit un processus de branchement unitype en temps continu $(U_t^{x_0})_{t\geq 0}$ dont la loi de reproduction est obtenue à partir des taux de saut du processus à valeurs mesures. Si un individu de trait x_0 meurt ou mute, son nombre d'enfants pour le processus de branchement est zéro ; s'il se reproduit mais que le descendant produit n'a pas le même trait (reproduction non-clonale à cause de la recombinaison), son nombre d'enfants est un ; s'il se reproduit clonalement, son nombre d'enfants est deux. On néglige les contributions de populations de traits différents de x_0 à la sous-population de trait x_0, puisqu'on s'intéresse à la capacité d'une sous-population de trait donné à se maintenir localement.

Mathématiquement, on obtient la dynamique suivante : un individu de trait (R_0, y_0) est remplacé par sa descendance à taux $1 + \mu + d(y_0)$.

- Avec probabilité $\frac{\mu + d(y_0)}{1+\mu+d(y_0)}$, le parent meurt sans produire de descendance ;
- Avec probabilité $\frac{R_0 \frac{3}{4} r}{1+\mu+d(y_0)}$, où r est la probabilité qu'un évènement de recombinaison ait lieu pendant la reproduction, le parent est remplacé par un descendant de trait x_0 ;
- Avec probabilité $\frac{1 - R_0 \frac{3}{4} r}{1+\mu+d(y_0)}$, le parent est remplacé par deux descendants de trait x_0.

On étudie, pour chaque trait x_0, la criticité du processus de branchement associé, ainsi que la probabilité d'extinction lorsque le processus est sur-critique, et le temps d'extinction moyen lorsque le processus est

sous-critique. En comparant ces quantités entre populations d'individus recombinants et non-recombinants, on observe que la zone de sur-criticité, donc la zone sur laquelle la population locale a une probabilité non nulle de croître exponentiellement, est plus large lorsque la recombinaison est bloquée (Figure 3). Cela suggère que la suppression de recombinaison peut être maintenue dans un cadre d'auto-fécondation intra-tétrade. Cependant, la probabilité d'extinction des processus sur-critiques reste très élevée.

Simulations individu-centrées. On pousse l'exploration un peu plus loin en réalisant des simulations individu-centrées du processus de sauts à valeurs mesures. L'observation de quelques trajectoires semble suggérer que, pour le jeu de paramètres choisi, le taux de croissance est le même entre individus recombinants et non-recombinants, et que la charge mutationnelle peut atteindre une distribution stationnaire dans les deux cas. Cette distribution stationnaire diffère cependant, et on observe en particulier que les individus non-recombinants montrent un taux d'hétérozygotie plus élevée que les individus recombinants. Le tracé du paysage de fitness (c'est-à-dire de la distribution du taux de mort, figure 5) montre une distribution de fitness similaire pour les individus recombinants et non-recombinants. Cela suggère encore que la suppression de recombinaison ne serait pas défavorisée dans un cadre d'autofécondation intra-tétrade, et pourrait ainsi être maintenue dans une telle population.

Conclusions. L'analyse des processus de branchement locaux et des simulations individu-centrées suggère que la suppression de recombinaison peut être maintenue dans une population d'individus se reproduisant exclusivement par auto-fécondation intra-tétrade, et en présence d'un locus de type sexuel en permanence hétérozygote. Ces analyses restent cependant très préliminaires. La limite des capacités de calcul des serveurs utilisés ne nous ayant pas permis de simuler les trajectoires pendant des temps très longs, il ne faut pas exclure la possibilité que le taux de croissance ou la distribution de la charge mutationnelle varie pour une sous-population ou l'autre. Si c'est le cas, cela implique que dans le cas de tailles de population bornées (*i.e.* dans le cas réel où les ressources sont limitées), la compétition induite entre les individus peut amener à la sélection pour un caractère ou l'autre (pour la suppression de recombinaison ou au contraire, son rétablissement). Une analyse plus poussée via simulations, ou la résolution des limitations mathématiques, permettrait d'apporter plus de précisions sur le comportement asymptotique de chaque sous-population.

En résumé : La suppression de recombinaison semble pouvoir être maintenue dans une population constituée d'individus portant un locus de type sexuel en permanence hétérozygote et se reproduisant exclusivement par autofécondation, malgré l'accumulation de mutations délétères.

Glossaire

allèle Version d'un gène.

diploïde qui contient deux copies de l'information génétique, c'est-à-dire des paires de chromosomes. S'oppose à haploïde.

délétion suppression de matériel génétique dans un chromosome.

délétère qui apporte un désavantage.

déséquilibre de liaison deux allèles sont dits en déséquilibre de liaison lorsque la fréquence à laquelle on les retrouve dans un même génome est différente du produit de leur fréquences alléliques respectives. Plus précisément, si on considère deux allèles A et B de fréquences alléliques respectives p_A et p_B, on dit que A et B sont en déséquilibre de liaison si $p_{AB} \neq p_A p_B$, où p_{AB} est la fréquence à laquelle ces deux allèles sont observés dans le même génome.

fitness terme anglais pour *valeur sélective*.

gamète cellule reproductrice haploïde. Deux gamètes fusionnent pour former un nouvel individu. Chez l'Homme par exemple, les gamètes sont les ovules et les spermatozoïdes.

haplotype contraction de *haploid genotype*, désigne une information génétique donnée par un génome haploïde. On utilise souvent "génotype" pour désigner l'information génétique donnée par un génome diploïde.

homozygotie Un gène ou un locus est dit homozygote au sein d'un génome diploïde s'il porte le même allèle sur les deux chromosomes. Opposé à hétérozygote.

insertion ajout de matériel génétique dans un chromosome.

inversion type particulier de mutation, ou réarrangement chromosomique, qui renverse l'ordre des gènes sur une partie d'un chromosome. Cela empêche les chromosomes de s'apparier correctement et bloque la recombinaison.

locale on désigne par sous-population *locale* la sous-population formée par les individus d'un trait donné. L'adjectif *local* est ici utilisé en référence à l'espace des traits : on se concentre sur la sous-population d'individus ayant un trait dans une certaine zone de l'espace des traits.

loci pluriel de *locus*, position sur un chromosome.

masse de Dirac type particulier de mesure (voir *mesure*), qui donne un poids 1 à un point particulier de l'espace et un poids 0 ailleurs. Une population décrite par une masse de Dirac en un point x_0 est composée d'un seul individu de trait (ou caractère) x_0.

mesure une mesure peut être vue comme une généralisation d'une distribution de probabilité. Lorsque l'on décrit des populations par une mesure, on rend compte de la façon dont les individus se répartissent sur l'espace qui leur est accessible (qui peut être l'espace géographique, un espace de traits ou de caractères, un espace d'âge...).

panmictique de *panmixie*, désigne une population dans laquelle les individus se reproduisent entre eux uniformément au hasard.

processus de sauts à valeurs mesures type de processus stochastique dont l'objet suivi est une *mesure**. Un processus à valeurs mesures trace l'évolution de cette mesure au cours du temps, et permet ainsi de décrire l'évolution de la population.

récessive une mutation est récessive (ou un allèle est récessif) lorsque son l'impact sur la valeur sélective est moins ou pas exprimé lorsqu'elle se trouve à l'état hétérozygote. Opposé à "dominant". Pour quantifier la réduction d'expression d'un allèle récessif, on utilise un coefficient de dominance souvent noté h.

super-dominant lorsqu'un génotype hétérozygote à un locus a une meilleure valeur sélective que les génotypes homozygotes associés.

supergène bloc de gènes ou loci complètement liés entre eux.

taille efficace il existe de multiples définitions de la taille efficace, qui dépendent du contexte dans lequel on travaille, et des données que l'on souhaite étudier. L'idée générale est de donner un indicateur de la diversité génétique au sein d'une population réelle de taille finie. La taille efficace donne la taille d'une population "idéale" (c'est-à-dire sans sélection, sans mutation, sans migration, sans structure...) dans laquelle on observerait le même niveau de diversité que dans la population réelle à l'étude. Un des effets d'une petite taille efficace est de rendre la population plus sujette aux fluctuations stochastiques. La sélection directionnelle est donc moins efficace. On renvoie le lecteur intéressé vers Waples, 2022 pour un panorama plus complet.

trait terme général qui désigne une caractéristique d'un individu, le plus souvent utilisé pour désigner un phénotype.

tétrade regroupement des quatre gamètes produits par un évènement de méiose à partir d'une cellule diploïde.

valeur sélective ou fitness en anglais, quantifie la capacité d'un individu à se reproduire et donc à transmettre son génotype. Il s'agit d'une valeur relative aux autres individus de la même population.

Chapitre 1

Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes

Ce chapitre a fait l'objet d'une publication accessible librement en ligne (Jay et al., 2022b, https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001698).

Cet article est le fruit d'un travail de collaboration mené par Paul Jay et dirigé par Tatiana Giraud et Amandine Véber, dans le cadre du projet ERC EvolSexChrom. J'ai pour ma part contribué à l'élaboration de la modélisation utilisée pour le modèle déterministe et pour les simulations stochastique, et effectué les calculs d'évolution de fréquences alléliques pour le modèle déterministe (présentés dans le document "Supplementary methods" joint à ce manuscrit). L'objectif principal était de modéliser des individus diploïdes, possédant un locus sexuel ou de type sexuel (permettant de considérer plusieurs systèmes de reproduction), et un grand nombre de sites sur lesquels pouvaient se trouver des mutations délétères. Le but était ensuite d'étudier la dynamique d'un suppresseur de recombinaison, en fonction de sa position par rapport au locus sexuel ou de type sexuel, et du nombre de mutations délétères portées. Les simulations stochastiques réalisées par Paul Jay avec le logiciel SLiM (Haller et Messer, 2019) ont été implémentées dans ce cadre. En revanche, pour l'analyse du modèle déterministe, nous avons fait des hypothèses supplémentaires qui limitent le nombre d'équations à calculer et permettent de suivre la fréquence de l'inversion dans la population. Nous avons choisi de modéliser la présence ou l'absence de mutations délétères sur un site par des variables aléatoires indépendantes et identiquement distribuées selon une loi de Bernoulli de paramètre q (la fréquence d'équilibre mutation-sélection). En particulier, le nombre de mutations délétères portées par l'inversion est fixe dans le temps.

Le premier chapitre de ce manuscrit est donc constitué du texte principal de l'article, suivi de l'annexe "Supplementary methods", tels qu'on peut les trouver en ligne. Les figures supplémentaires, également accessibles en ligne, sont regroupées à la suite de ces documents. Les légendes de ces figures se trouvent dans le texte principal.

PLOS BIOLOGY

RESEARCH ARTICLE

Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes

Paul Jay[1]*, **Emilie Tezenas**[1,2,3], **Amandine Véber**[3], **Tatiana Giraud**[1]

1 Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, 91190, Gif-sur-Yvette, France, **2** Univ. Lille, CNRS, UMR 8198 –Evo-Eco-Paleo, F-59000 Lille, France, **3** Université Paris Cité, CNRS, MAP 5, F-75006 Paris, France

* paul.yann.jay@gmail.com

OPEN ACCESS

Citation: 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 Biol 20(7): e3001698. https://doi.org/10.1371/journal.pbio.3001698

Academic Editor: Laurence D. Hurst, University of Bath, UNITED KINGDOM

Received: February 14, 2022

Accepted: June 3, 2022

Published: July 19, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pbio.3001698

Data Availability Statement: The SLiM and R scripts used to produce all main and supplementary figures are available on GitHub (https://github.com/PaulYannJay/

Abstract

Many organisms have sex chromosomes with large nonrecombining regions that have expanded stepwise, generating "evolutionary strata" of differentiation. The reasons for this remain poorly understood, but the principal hypotheses proposed to date are based on antagonistic selection due to differences between sexes. However, it has proved difficult to obtain empirical evidence of a role for sexually antagonistic selection in extending recombination suppression, and antagonistic selection has been shown to be unlikely to account for the evolutionary strata observed on fungal mating-type chromosomes. We show here, by mathematical modeling and stochastic simulation, that recombination suppression on sex chromosomes and around supergenes can expand under a wide range of parameter values simply because it shelters recessive deleterious mutations, which are ubiquitous in genomes. Permanently heterozygous alleles, such as the male-determining allele in XY systems, protect linked chromosomal inversions against the expression of their recessive mutation load, leading to the successive accumulation of inversions around these alleles without antagonistic selection. Similar results were obtained with models assuming recombination-suppressing mechanisms other than chromosomal inversions and for supergenes other than sex chromosomes, including those without XY-like asymmetry, such as fungal mating-type chromosomes. However, inversions capturing a permanently heterozygous allele were found to be less likely to spread when the mutation load segregating in populations was lower (e.g., under large effective population sizes or low mutation rates). This may explain why sex chromosomes remain homomorphic in some organisms but are highly divergent in others. Here, we model a simple and testable hypothesis explaining the stepwise extensions of recombination suppression on sex chromosomes, mating-type chromosomes, and supergenes in general.

Introduction

Sex chromosomes, mating-type chromosomes, and supergenes in general are widespread in nature and control striking polymorphisms, such as sexual dimorphism or color

MutationShelteringTheory). The numerical outputs of the simulations used to produce this manuscript figures are available on Figshare (doi: 10.6084/m9.figshare.19961033). Details concerning the mathematical modelling are available in the S1 Appendix.

Funding: This work was supported by the European Research Council (ERC) EvolSexChrom (832352) grant and a Louis D. Foundation (Institut de France) prize to TG. ET and AV acknowledge support from the chaire program « Mathematical modeling and biodiversity » (Ecole Polytechnique, Museum National d Histoire Naturelle, Veolia Environnement, Fondation X). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

polymorphism, in many organisms, including humans [1–5]. Supergenes display puzzling genomic architectures, defined by extensive regions lacking recombination, encompassing multiple genes [3,4]. Sex chromosomes, in particular, often appear to have evolved by stepwise extension of nonrecombining regions, generating "evolutionary strata" of differentiation between haplotypes [4,6,7]. It is generally thought that the reason for this gradual expansion of recombination suppression on sex chromosomes is that 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 [7–9]. However, theoretical concerns have been raised about this idea [10], and it has proved difficult to obtain empirical support for a role of antagonistic selection in the extension of recombination suppression [6,11,12]. Furthermore, a gradual expansion of recombination suppression has been observed around many fungal mating-type loci [5] and at other supergenes, such as those controlling wing color in butterflies or social structure in ants [13,14]. Some type of antagonistic selection may exist between morphs determined by color or social supergenes, but antagonistic selection between fungal mating types is highly unlikely, given their life cycles and the results of genomic investigations [15,16]. For example, there has been repeated stepwise extensions of recombination suppression in anther-smut fungi despite the lack of a haploid phase consisting of cells of different mating types potentially expressing contrasted life history traits, leaving little room for antagonistic selection [15,16]. Alternative explanations have been put forward for the expansion of nonrecombining regions on sex chromosomes [11], such as meiotic drive [17], genetic drift [18], deleterious-mutation sheltering [19,20], and the neutral accumulation of sequence divergence [21], but the conditions in which such mechanisms could apply also appear to be restricted.

We develop here a general model for testing the idea that recombination suppression may extend stepwise around sex-determining or mating-type loci because it provides the fitness advantage of sheltering deleterious mutations segregating in nearby regions. Related hypotheses have been explored before, but in the restricted specific context of inbreeding [19,20]. We model here a more general hypothesis, based on the idea that genomes carry numerous deleterious recessive variants, as suggested by studies in a wide range of biological systems and by the pervasiveness of inbreeding depression in nature [22–27]. We use mathematical modeling and stochastic simulations to test the hypothesis that permanently heterozygous alleles, such as male-determining alleles in XY systems, protect linked chromosomal inversions against the expression of their recessive mutation load, potentially leading to an accumulation of inversions around permanently heterozygous alleles, generating evolutionary strata.

The rationale behind this model is illustrated in Fig 1. Consider a diploid population carrying partially recessive, deleterious mutations. The combined effects of recombination, mutation, selection, and drift result in individuals carrying different numbers of deleterious variants genome-wide and within particular genomic regions (Fig 1A). Chromosomal inversions may occur at any position and suppress recombination when heterozygous, thereby capturing a specific combination of deleterious variants (i.e., a haplotype). Inversions capturing fewer deleterious variants than the population average for the region concerned have a fitness advantage and should, therefore, increase in frequency. Such inversions are advantageous due to associative overdominance, i.e., the inversion itself is neutral but it captures a combination of alleles that is advantageous when heterozygous [28,29]. However, as the frequency of an inversion increases, homozygotes for this inversion become more common. Homozygotes are at a strong disadvantage due to the recessive deleterious variants carried by these inversions, and selection against

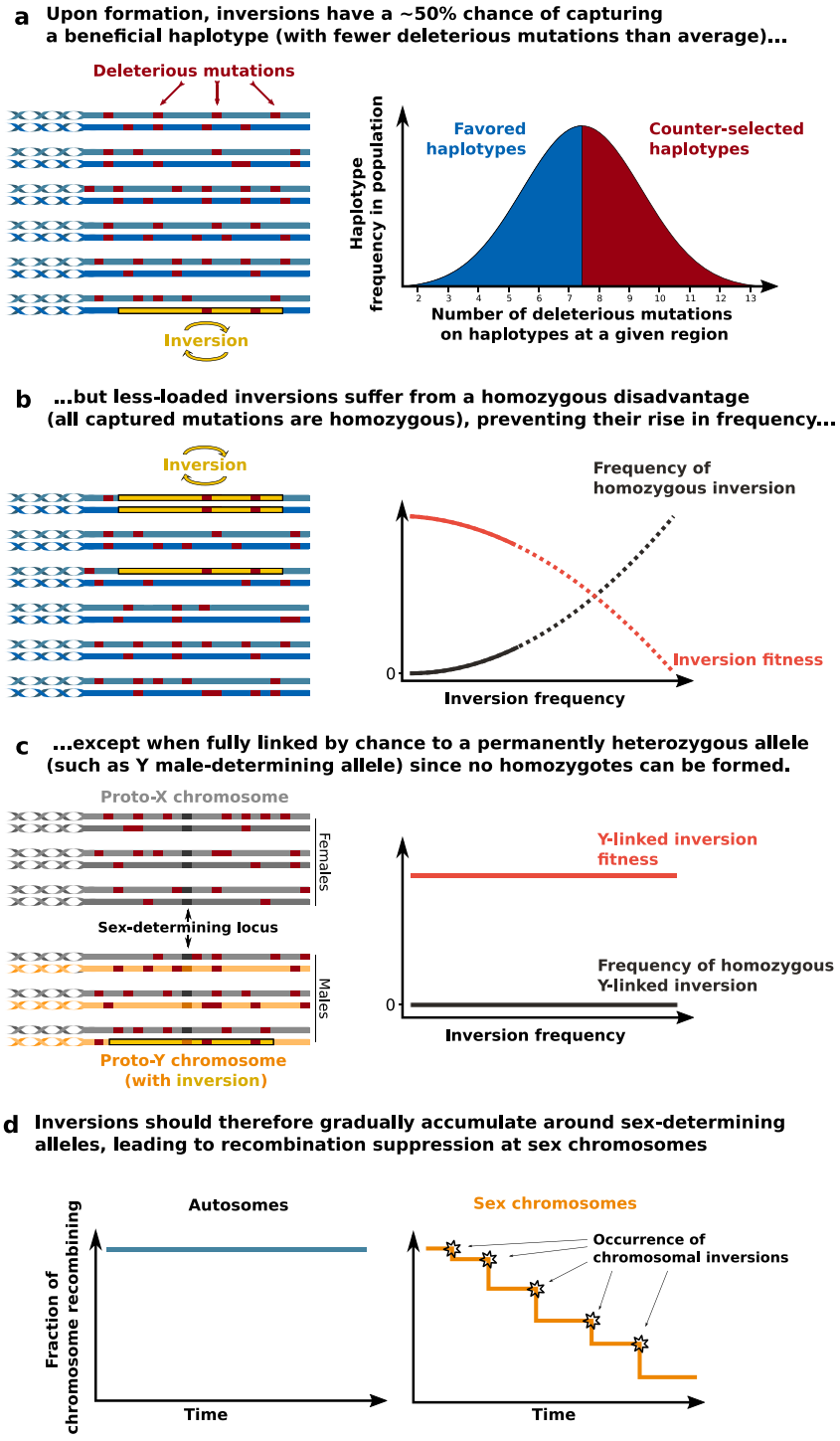


Fig 1. Schematic diagram of the model. (**A**) Within any population and for any genomic region, diploid individuals (here represented by 2 homologous chromosomes) and haplotypes (i.e., combinations of mutations, here one chromosome) carry variable numbers of partially recessive, deleterious mutations. A substantial fraction (about 50%) of haplotypes have fewer deleterious mutations than the average and should be favored by selection. Chromosomal inversions therefore have a significant chance of capturing beneficial haplotypes. (**B**) The increase in frequency of a

beneficial inversion leads to an increase in the frequency of homozygotes for the inversion, 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, keeping the frequency of the inversion low. (**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 fixed in the population of Y chromosomes. (**D**) Unlike those on autosomes, inversions capturing the Y sex-determining allele never suffer from homozygous disadvantage and should therefore 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 just chromosomal inversions) and any locus with permanently heterozygous alleles.

https://doi.org/10.1371/journal.pbio.3001698.g001

homozygotes would therefore be expected to prevent such inversions from reaching high frequencies (Fig 1B).

Now, consider an inversion that, by chance, captures a permanently heterozygous allele, such as the male-determining allele in an XY system. If this Y-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 associated with the permanently heterozygous, male-determining allele, and will, therefore, never occur as homozygotes. Unlike autosomal inversions, Y-linked inversions retain their fitness advantage with increasing frequency (Fig 1C). Hence, Y-linked inversions with a lower load than average would be expected to spread, becoming fixed in the population of Y chromosomes, resulting in a suppression of recombination between the X and Y chromosomes in the region covered by the inversion.

The successive fixation of additional inversions linked to this Y-fixed inversion should cause the nonrecombining region to expand further, by the same process, thereby leading to the formation of a chromosome with a large nonrecombining region around the sex-determining locus—i.e., a sex chromosome with evolutionary strata of differentiation (Fig 1D). We use the term "chromosomal inversions" here for simplicity, but the proposed mechanism would hold for any mechanism suppressing recombination, and we model several types of recombination suppressors. The proposed mechanism would be expected to apply to any genomic region with a permanently or near-permanently heterozygous allele, such as mating-type loci in fungi or supergenes in many organisms. The accumulation of deleterious mutations following recombination suppression has been extensively studied [30–33], but we investigate here its converse: that deleterious mutations could be a cause, and not only a consequence, of recombination suppression. We also investigate whether this mechanism of deleterious mutation sheltering can explain the fusions sometimes observed between sex chromosomes and autosomes [20,34–36] and we explore the impact of various parameters on the probability of recombination suppression.

Results

We explored this general hypothesis of stepwise recombination suppression around permanently heterozygous alleles with infinite population deterministic models and individual-based simulations. Very little is known about the dynamics of deleterious mutations in genomic regions under particular recombination regimes, such as those generated by polymorphic inversions. We therefore had to use simulations to explore realistic scenarios involving various levels of drift. In both the infinite population model and individual-based simulations, we modeled diploid populations with only partially recessive deleterious mutations, occurring at a rate u, with heterozygous and homozygous individuals suffering from $1\text{-}hs$ and $1\text{-}s$ reductions in fitness, respectively, and with multiplicative effects of mutations on fitness. We first

considered that all mutations occurring in genomes had the same dominance (h) and selection (s) coefficients, and then relaxed this hypothesis. Individuals were considered to have 2 pairs of chromosomes, one of which harbored a locus with at least 1 allele permanently or almost permanently heterozygous (see Methods for details). Several situations were considered, mimicking those encountered in XY sex-determination systems, in fungal mating-type systems and in overdominant supergenes. We analyzed the evolution of recombination modifiers suppressing recombination across the fragment in which they reside (i.e., *cis*-modifiers), either exclusively in heterozygotes (mimicking for example chromosomal inversions), or in both heterozygotes and homozygotes (e.g., histone modifications). Each of these recombination modifiers, which was assumed to be neutral in itself, appeared in a single haplotype, and was, thus, in linkage disequilibrium with a specific set of mutations, such that its fitness was exclusively dependent on the number of deleterious alleles within the segment captured. We first compared the dynamics of inversion-mimicking mutations in an autosome with those capturing a male-determining allele in an XY system (males are XY and females are XX, the male-determining allele being permanently heterozygous), and we then considered other types of recombination modifiers and heterozygosity rules.

Inversions less loaded than average are frequent in genomes

As noted by several authors [28,37,38], inversions capturing fewer deleterious mutations than the population average should increase in frequency. In infinite populations, the number of mutations harbored by individuals within a genomic segment with n sites follows a binomial distribution of parameters n and q, with q the mean frequency of mutations at mutation-selection equilibrium (Figs 2A and S1). On average, individuals harbor nq mutations on each of their chromosomal segments of size n. Under realistic parameter values, the vast majority of large chromosomal regions therefore carry several deleterious mutations. For example, considering $s = 0.001$, $h = 0.1$, $\mu = 10^{-9}$, and $n = 2$ Mb, more than 99.999% of chromosomal fragments carry a least 1 mutation, the mean number of mutations being $nq = 20$ (S1 Fig). If m is the number of recessive deleterious mutations captured by a given inversion, the mean fitness of individuals homozygous for the inversion (W_{II}), heterozygous for the inversion (W_{NI}), or lacking the inversion (W_{NN}) in an infinite population can be easily expressed as a function of n, q, h, s, and m (see Methods; [28,37]). 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 the case if the inversion carries fewer mutations than the population average [$m < nq$; [28]].

Repeated sampling from binomial distributions with a wide range of parameters showed that more than half the inversions occurring in genomes captured fewer deleterious mutations than the population average (Fig 2B; hereafter referred to as "less-loaded inversions"). Indeed, the distribution of mutation number across individuals is almost symmetric when nq is high (as the binomial distribution converges to a normal distribution; e.g., S1 Fig), but, when nq is low, the distribution is zero inflated, increasing the probability of inversions being less loaded. Therefore, a substantial fraction of inversions occurring in genomes are beneficial when they form (i.e., when rare enough not to occur as homozygotes). For example, with h values ranging from 0 to 0.5 and s values ranging from 0.001 to 0.25, between 36% and 98% (mean = 70%) of the 2-Mb inversions occurring in the genome are beneficial, carrying fewer recessive deleterious mutations than average (Fig 2B). Simulations in finite populations of different sizes confirmed that most inversions (mean = 66% in the range of parameter values studied) had a fitness advantage upon formation. The simulations also showed that inversions could be favored if they captured mutations that were rarer than average (S2 and S3 Figs).

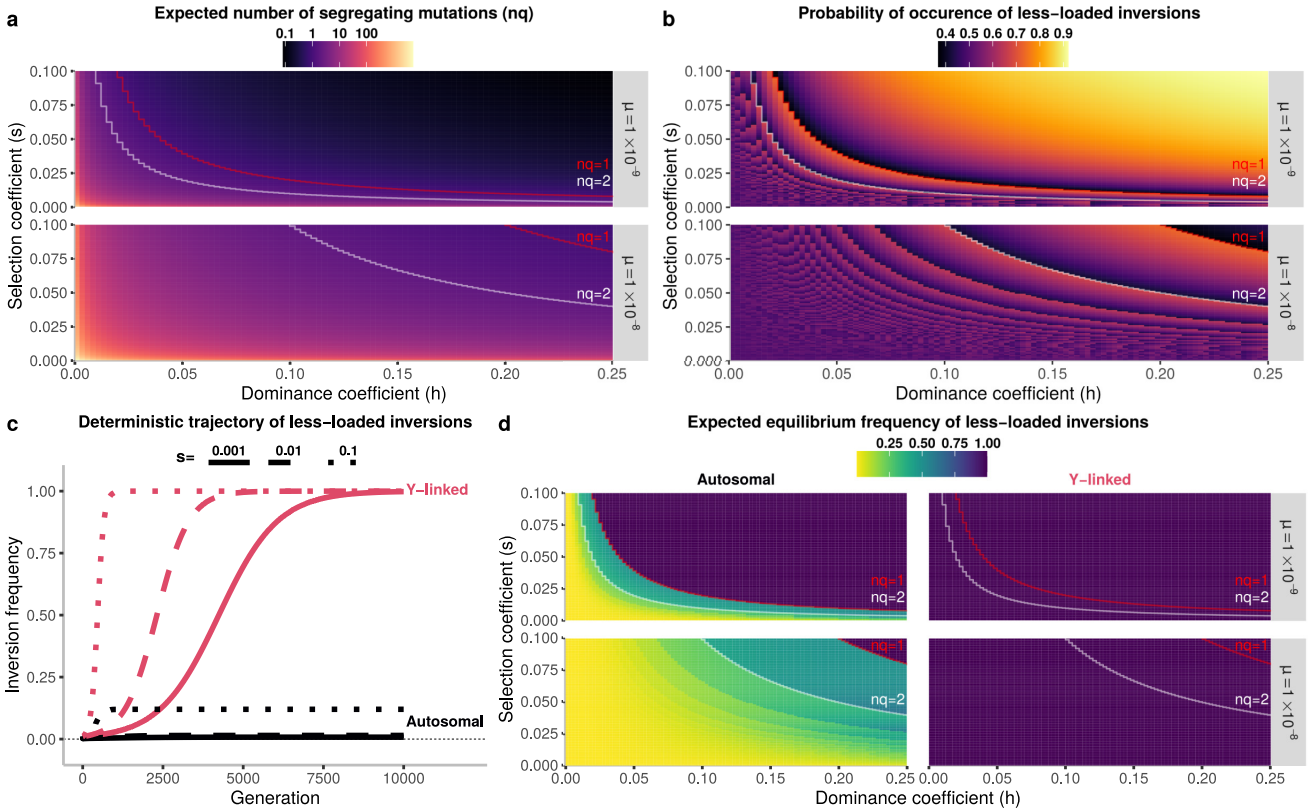


Fig 2. In infinite populations, all less-loaded chromosomal inversions become fixed on the Y sex chromosome, whereas most remain at a low frequency on the autosome. (**a**) Expected number of segregating recessive deleterious mutations within a segment of $n = 2$ Mb as a function of the dominance coefficient (h), selection coefficient (s), and mutation rate (u). Mutations were considered to be at mutation-selection equilibrium frequency (see Methods). Parameters resulting in 1 and 2 expected mutations are highlighted by red and white lines, respectively, and are also shown on panels b and d. (**b**) Probability of occurrence of a less-loaded inversion covering a segment of $n = 2$ Mb as a function of the dominance coefficient (h), selection coefficient (s), and mutation rate (u). Less-loaded inversions are defined as those for which $m<nq$, where m is the number of mutations in the inversion (i.e., fewer than the average for the population). The number of mutations in inversions can take only integer values ($m = 0, 1, 2$, etc.), and the transition between these values affects the probability of occurrence of less-loaded inversions, yielding the noncontinuous graphs in panels b and d. (**c**) Deterministic change in inversion frequency, for the case of 2-Mb inversions capturing the male-determining allele on the Y-chromosome or inversions on an autosome. For clarity, the frequency displayed for Y-linked inversions is the frequency of inversions in the population of Y chromosomes. The figure illustrates the case of inversions carrying a number of mutations 5% lower than the population average ($m = 0.95 \times nq$), mutations being at their mutation-selection equilibrium frequency with $\mu = 10^{-8}$ and $h = 0.1$. S4 to S8 Figs illustrate the fate of inversions with intermediate degrees of linkage to the male-determining allele, with various numbers of mutations relative to the population average, or inversions linked to a locus with variable heterozygosity rules or in the X chromosome population. (**d**) Expected equilibrium frequency of less-loaded 2-Mb inversions (i.e., with $m = 0,1, 2, \ldots, nq$) capturing the male-determining allele, or on an autosome, weighted by their probability of occurrence (Methods, Eqs 5, 6 and 7). This panel extends the result from panel c for a range of dominance coefficients (h), selection coefficients (s), and mutation rates (u). As above, the frequency displayed for Y-linked inversions is the frequency of inversions on the population of Y chromosomes. Above the line $nq = 1$, the mean number of segregating mutations in a given inversion (of length n) is less than 1, indicating that all less-loaded inversions are mutation-free. Such inversions can also become fixed on autosomes, leading to an expected equilibrium frequency of 1 for both autosomes and the Y chromosome. S9 Fig shows the overall frequencies of all inversions, whether less or more loaded than average, and over a larger range of parameters. The scripts used to produce this figure are available on GitHub.

https://doi.org/10.1371/journal.pbio.3001698.g002

Less-loaded inversions are much more likely to fix when they capture a Y-like male determining locus

The deterministic increase in frequency of an inversion (I) on an autosome or capturing an XY-like sex-determining locus can easily be determined with a 2-locus 2-allele model. For example, the change in frequency of an inversion on the proto-Y chromosome can be

expressed as:

$$\Delta F_{YI} = \frac{F_{YI}(F_{XIf}W_{II} + (1 - F_{XIf})W_{NI} - \overline{W_m}) \pm rW_{NI}D}{\overline{W_m}},$$

where F_{XIf} is the frequency of the inversion on the proto-X chromosome 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 mean male fitness (Fig 2C; see Methods and S1 Appendix for details). Based on this model, and initially assuming that inverted and noninverted segments no longer accumulate deleterious mutations after their formation, i.e., W_{II}, W_{NI}, and W_{NN} are fixed parameters (this strong assumption is relaxed latter), we simulated the evolutionary trajectory of inversions on autosomes and of inversions capturing the male-determining allele on the Y chromosome under a wide range of parameter values. We found that the frequency of less-loaded inversions tended to remain low in autosomes, whereas these inversions became fixed in the population of Y chromosomes (Fig 2C), as expected according to our hypothesis.

We therefore expressed the equilibrium frequency of inversions ($F_{equ,m}$, Fig 2D) as a function of m, the number of deleterious mutations carried by the inversion, to calculate the threshold values at which autosomal and Y-linked inversions could be maintained or fixed (see S1 Appendix). Assuming that $q \ll 1$ and $s \ll 1$, we found that inversions on autosomes become fixed when $m < qhn/(1-h)$ whereas they should stabilize at a frequency of $(W_{NI} - W_{NN})/(2W_{NI} - W_{NN} - W_{II})$, the equilibrium frequency of an overdominant locus, when $qhn/(1-h) < m < nq$ (see Methods). Inversions on autosomes capturing more than $qhn/(1-h)$ recessive deleterious mutations do, indeed, suffer from a homozygote disadvantage ($W_{NI} > W_{II}$, S1 Fig), preventing them from reaching high frequencies (Fig 2C). This is the case for most autosomal inversions under realistic parameter values (e.g., $m > qhn/(1-h)$ for more than 99.99% of inversions when $n = 2$ Mb, $\mu = 10^{-9}$, $h = 0.1$, and $s = 0.001$, S1 Fig). By contrast, permanently heterozygous alleles protect inversions from this homozygote disadvantage, allowing these inversions to become fixed. All inversions capturing the male-determining allele become, therefore, fixed if they carry fewer than average mutations (i.e., $m < nq$). Thus, contrary to the argument proposed in a previous study that only mutation-free inversions can become fixed [35], we found that inversions can carry deleterious mutations and nevertheless become fixed in the population, provided that they carry fewer than $qhn/(1-h)$ mutations if they are located on autosomes and fewer than nq mutations if they capture a permanently heterozygous allele (on the Y chromosome, for example).

The expected equilibrium frequency of all inversions occurring in a given genomic region can be expressed as:

$$[F_{equ}] = \sum_{m=0}^{n} P_m \times F_{equ,m}$$

with P_m being the probability of occurrence of inversions with m mutations and $F_{equ,m}$ the equilibrium frequency of inversions with m mutations (Fig 2D; see Methods). For permanently heterozygous inversions, we, thus, have:

$$[F_{equ}] = \sum_{m=0}^{\lfloor nq \rfloor} \binom{n}{m} q^m (1-q)^{n-m},$$

and for autosomal inversions:

$$\left[F_{equ}\right] = \sum_{m=0}^{\lfloor nq\frac{h}{1-h}\rfloor}\binom{n}{m}q^m(1-q)^{n-m} + \sum_{m=\lfloor nq\frac{h}{1-h}\rfloor+1}^{\lfloor nq\rfloor}\binom{n}{m}q^m(1-q)^{n-m}\frac{W_{NI}-W_{NN}}{2W_{NI}-W_{NN}-W_{II}}.$$

With a realistic range of parameters, inversions are much more likely to become fixed if they capture the male-determining allele on the Y chromosome than if they are unlinked to this allele (e.g., on an autosome; Figs 2C and 2D and S4 to S9). For example, with $\mu = 10^{-9}$, $h = 0.1$, $s = 0.001$, and $n = 2$ Mb, 47% of inversions occurring on the Y chromosome would be expected to become fixed, versus only 0.000045% of inversions on autosomes.

Drift and mutation accumulation do not prevent Y-linked inversion fixation

If deleterious mutations continue to arise after the formation of the inversion, the dynamics of the inversion become harder to predict deterministically. In infinite populations, the accumulation of mutations after the formation of the inversion can be approximated by $P_t = \frac{\mu}{sh} + \left(P_0 - \frac{\mu}{sh}\right)e^{-sht}$, where P_t is the number of mutations in the inversion at time t ([28,37]; see also [39]). This assumes that inverted segments evolve under mutation-selection dynamics with recombination. However, such assumptions do not hold in many situations, in contrast to what has been assumed in a previous study [39]. Indeed, low-frequency inversions in finite populations should almost never occur as homozygotes and should therefore evolve with almost no recombination. This is also true for inversions capturing a permanently heterozygous allele, which never undergoes recombination. We confirmed by simulations that the above approximation for an infinite population strongly departs from the situation observed in finite populations (e.g., S13 Fig). In finite populations, low-frequency or permanently heterozygous inversions tend to evolve under a Muller's ratchet-like dynamic, with the mean fitness of inversions decreasing in a stepwise manner due to the sequential loss, by drift, of the inverted haplotypes with the lowest mutational load. Following their formation, autosomal and Y-linked inversions tend to accumulate more mutations than the population average, contrasting with predictions for infinite populations (S13 Fig). Only inverted segments reaching relatively high frequencies in autosomes eventually recombine when homozygous, and their dynamics of mutation accumulation therefore involve a mixture of a Muller's ratchet-like regime (when rare, at the start of their spread) and a mutation-selection-drift regime with recombination (when they reach intermediate frequencies). Little is known about the transition between these regimes [40,41]. We therefore used individual-based simulations to study the fate of inversions accumulating deleterious mutations. This also made it possible to relax the previous assumption [38] that the time to inversion fixation is much shorter than the time taken for the inversions to accumulate deleterious mutations. Individual-based simulations also made it possible to relax the assumption that all genomic positions are independent.

Simulations of the spread of inversions in finite populations with recurrent mutation confirmed the tendency identified in the deterministic model without mutation accumulation: Over most of the parameter space, inversions are much more likely to spread if they capture the sex-determining allele on the Y chromosome than if they are located on autosomes (Figs 3 and S10 to S16). Many autosomal inversions carrying a mutation load segregated for hundreds of generations. For example, with $N = 1{,}000$, $s = 0.01$, $h = 0.1$, and $\mu = 10^{-8}$, 73 of 10,000 inversions of 2 Mb in length continued to segregate after 500 generations. However, all these autosomal inversions were lost at the end of the simulations (i.e., after 10,000 generations, Fig 3A). These inversions initially spread because, by chance, they had a lower-than-average mutation

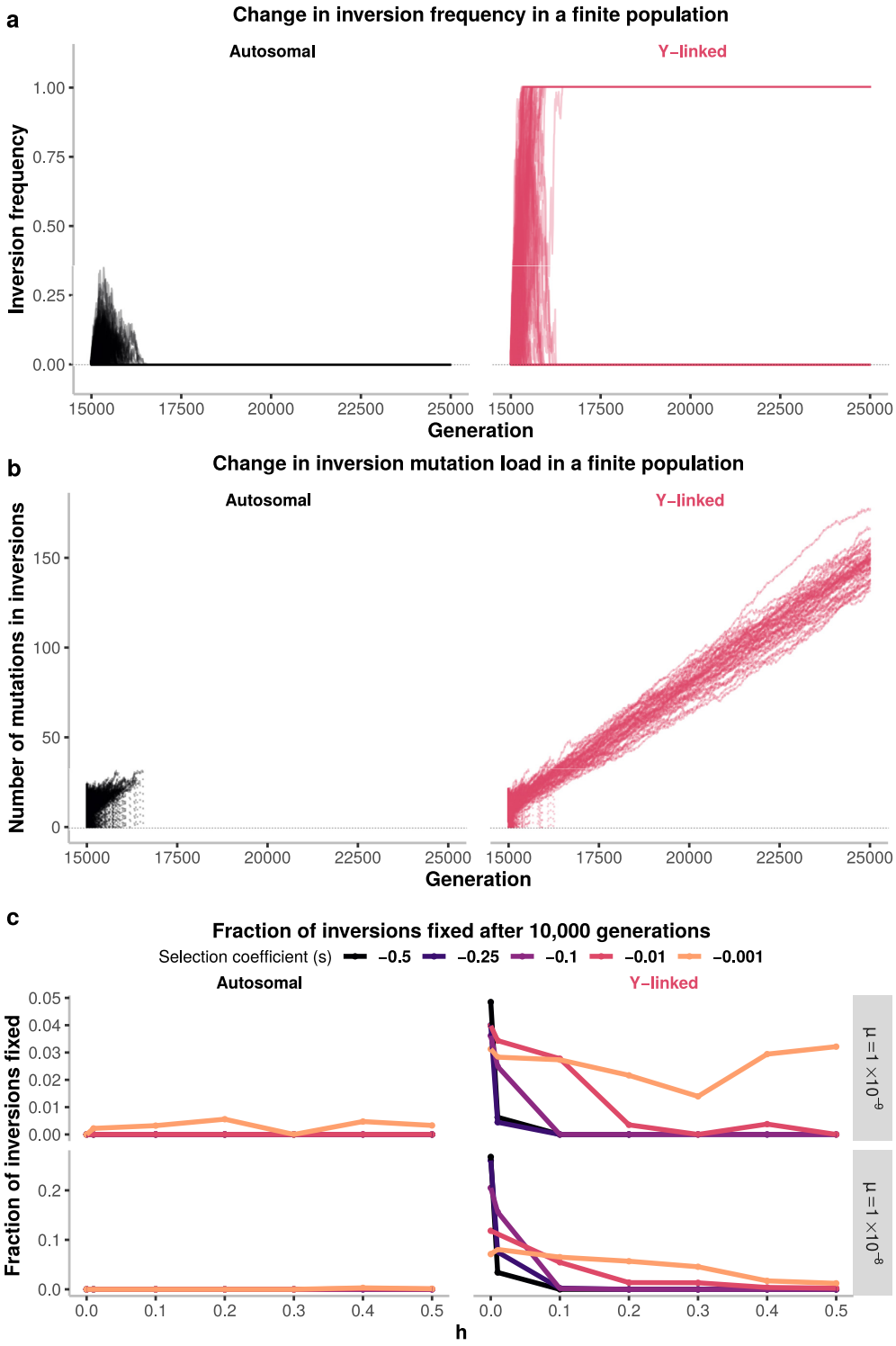


Fig 3. In finite populations, a substantial fraction of less-loaded chromosomal inversions on the Y sex chromosome become fixed, whereas all less-loaded inversion on autosomes are lost. (**a**) Change in inversion frequency in stochastic simulations of 1,000 individuals experiencing recessive deleterious mutations at a rate $\mu = 10^{-8}$, all mutations having the same dominance and selection coefficients (here, $h = 0.1$ and $s = 0.01$). The figure displays the frequency of 10,000 independent inversions on each of an autosome and a proto-Y chromosome, with each line representing a specific

inversion (i.e., a simulation). The evolutionary trajectories of inversions of different sizes, in larger populations, on the X-chromosome or with other parameter combinations, are displayed in S10–S14 Figs. (**b**) Change in the mean number of mutations carried by the inverted segments in each of the 10,000 simulations. Each line represents 1 simulation. A zoom-view of the autosomal inversion dynamics is shown in S13 Fig. (**c**) Proportions, in stochastic simulations, of 2-Mb inversions that were fixed after 10,000 generations for different parameter value combinations. This extends the result from panel a to other values of dominance coefficient (h), selection coefficient (s), and mutation rate (u). For each parameter combination, we simulated 10,000 inversions of 2 Mb capturing a random genomic fragment. Only inversions not lost after 20 generations are considered here. Note that Y axes have different scales. Results for inversions of different sizes are shown in S15 Fig and results for different population sizes are shown in S16 Fig. The datasets and scripts used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.

https://doi.org/10.1371/journal.pbio.3001698.g003

load, but their homozygote disadvantage prevented them from reaching fixation. They were eventually lost because they accumulated further mutations relative to noninverted segments (Figs 3B and S13).

By contrast, substantial fractions of less-loaded inversions capturing the permanently heterozygous sex-determining allele on the Y chromosome spread until they became fixed in the Y chromosome population; this was the case even for inversions that were not mutation-free (Figs 3 and S10 to S16). For example, for $s = 0.01$, $h = 0.1$, $\mu = 10^{-8}$, and $N = 1{,}000$, 49 of the 10,000 Y-linked inversions (2 Mb) became fixed in the Y chromosome population, whereas all inversions on the autosome were lost. New mutations occurred on Y-linked inversions, but they did not accumulate rapidly enough to prevent these 49 inversions from spreading and reaching fixation (Figs 3A and 3B and S13). Permanent heterozygosity effectively results in directional selection for the less-loaded inversions, and not only for those free from mutations (Fig 3), leading to rapid fixation before the accumulation of too many new deleterious mutations.

Other systems with permanently heterozygous alleles, other recombination modifiers, and sex chromosome–autosome fusion

Similar results were obtained when: (i) 2 or more permanently heterozygous alleles segregated at a mating incompatibility locus, modeling plant self-incompatibility or fungal mating-type systems (S5 and S14 Figs; [5,42]); (ii) the alleles were not permanently heterozygous, but were strongly overdominant, thus occurring mostly in the heterozygous state, as for several supergenes controlling color polymorphism (S8 Fig; [13,43,44]); (iii) the fitness effects of mutations occurring across the genome were drawn from a gamma distribution (i.e., the mutations segregating in populations had different fitness effects; S17 Fig); (iv) recombination modifiers suppressed recombination even when homozygous, as for histone modifications or methylation [45], rather than solely when heterozygous, as for inversions (S14 Fig); and (v) the inversion was in strong but incomplete linkage with the permanently heterozygous allele (e.g., 0.1 cM away from the allele, S4 to S8 Figs). 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 [46,47].

In addition, we found that the sheltering of deleterious recessive mutations could also lead to the fusion of the permanently heterozygous sex chromosome with an autosome when the fusion was associated with an extension of the nonrecombining region to the newly fused autosome (S18 Fig).

Population parameters impacting the probability of inversion spread and fixation: population size, mutation rate, recessiveness, cost of heterozygous inversion, and life cycle

As shown above, we found that neutral recombination modifiers spread in a very large range of conditions if they captured permanently heterozygous alleles (Figs 3C, S15 and S16). As

expected, inversions were more easily fixed on the Y chromosome when the deleterious mutations segregating in the genome were more recessive (Figs 3C, S15 and S16). Furthermore, as for any variant, even beneficial inversions benefiting from a selective advantage could be lost by drift, the probability of such loss depending on population size and the relative fitness of the inversion, in terms of the number and type of mutations initially captured relative to the average for the population (Figs 3, S15 and S16). Inversions occurring in regions in which large numbers of mutations segregate are more likely than those in mutation-poor regions to capture many fewer mutations than average, and therefore to have a stronger relative advantage. In other words, inversions in mutation-dense regions have a wider fitness distribution. The probability of Y-linked inversion fixation thus increases with increasing inversion size, mutation recessiveness, and mutation rate (Figs 3C, S15, S16 and S19).

As expected [48], we found that, with increasing population size, beneficial inversions took more generations to spread to fixation (S11 and S12 Figs). This longer time favored the further accumulation of deleterious mutations in inversions, lowering their fitness, and, in some cases, preventing their fixation. The probability of inversion fixation in the Y chromosome population was, therefore, inversely correlated with population size, but was always much higher than that of fixation on autosomes (S15 and S16 Figs).

Less-loaded inversions linked to a permanently heterozygous allele also spread when we assumed that inversion heterozygotes suffered from a fitness cost, for example, because of a decrease in fertility due to segregation issues during meiosis (S20 and S21 Figs). However, with a heterozygous cost, only inversions with much fewer mutations than average were beneficial and could spread, and not all less-loaded inversions (see Methods and S1 Appendix); large inversions were thus more likely to spread because, compared to small inversions, they were more likely to carry much fewer mutations than average (S20 and S21 Figs).

Less-loaded inversions linked to a permanently heterozygous allele could also spread and fix in simulations assuming a haplodiplontic life cycle (i.e., an alternation of haploid and diploid phases) (S22 and S23 Figs). However, the occurrence of a haploid phase enhanced the purge of deleterious recessive mutations, leading to a lower population mutation load and thereby to a narrower fitness distribution of inversions upon formation (S23 Fig; [49,50]). A lower proportion of inversions were therefore fixed in haplodiplontic populations than in purely diploid populations (S22 Fig).

Evolution of nonrecombining sex chromosomes with evolutionary strata despite possible reversions

We have shown above that inversions are likely to fix when capturing a permanently heterozygous allele, which should lead to their accumulation around such alleles and, for example, to the formation of nonrecombining sex chromosomes with a typical pattern of evolutionary strata. We studied the formation of such strata by the sequential occurrence of multiple inversions by simulating the evolution of large chromosomes experiencing the occurrence of multiple chromosomal inversions that can overlap with each other, under parameter values typical of those observed in mammals (Figs 4, 5 and S24). We simulated, over 100,000 generations, populations of $N = 10{,}000$ or $N = 1{,}000$ individuals, carrying two 100-Mb chromosomes, one of which harbored a mammalian-type sex-determining locus (XY males and XX females). Individuals experienced only deleterious or weakly deleterious mutations, with mutation rates and fitness effects similar to those observed in humans (fitness effect being drawn from a gamma distribution). The dominance coefficient of each mutation was chosen uniformly at random from a wide set of values (see Methods for details). At the start of each simulation, we randomly sampled k genomic positions that could be used as inversion breakpoints, with k

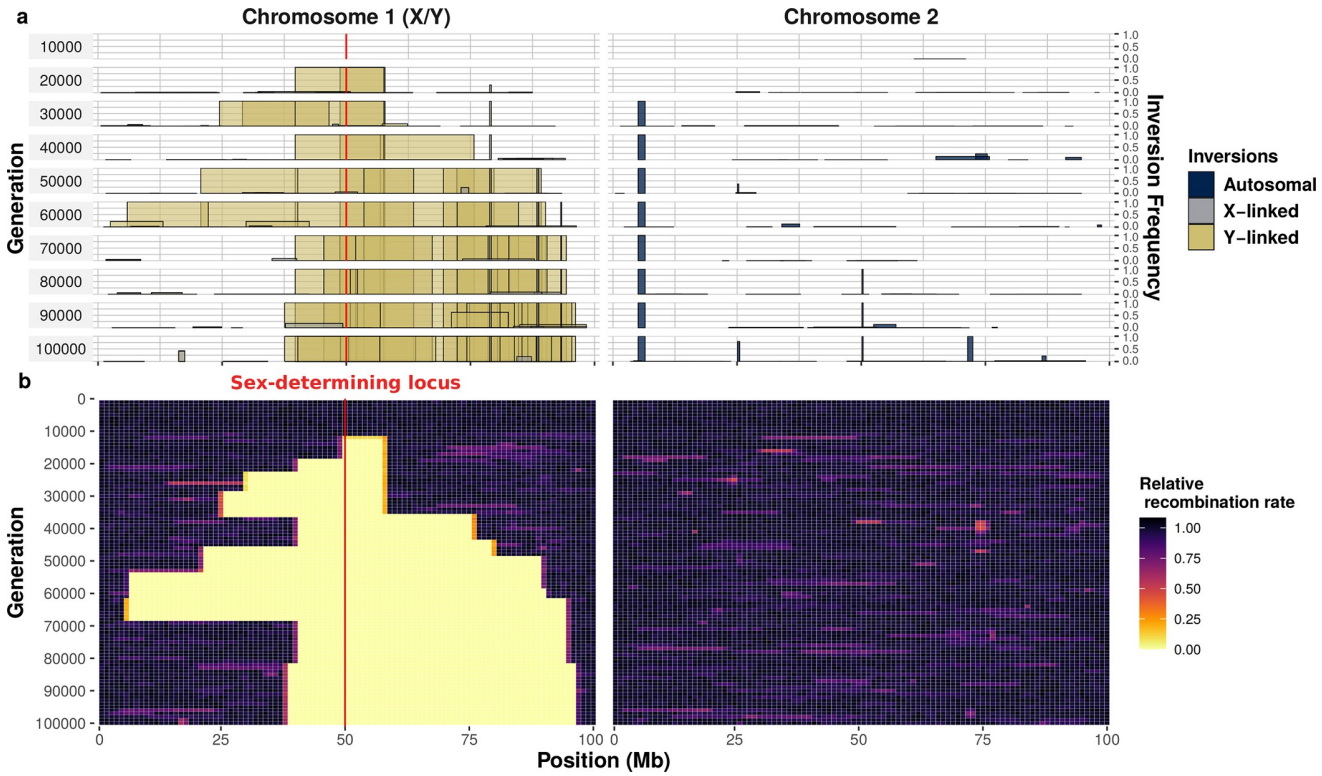


Fig 4. Successive accumulation of inversions around a male-determining allele in an XY system, leading to the formation of nonrecombining sex chromosomes. A simulation of $N = 1{,}000$ individuals, each with 2 pairs of 100-Mb chromosomes, over 100,000 generations. Chromosome 1 harbors an X/Y sex-determining locus at 50 Mb (individuals are XX or XY). In each generation, one inversion appears, on average, in the whole population, in a randomly sampled individual, with the 2 recombination breakpoints sampled uniformly at random from $k = 100$ potential breakpoints. (**a**) Overview of chromosomal inversion frequency and position for 10 different generations. The width of the box represents the position of the inversion, and the height of the box indicates 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. Previously fixed inversions may be lost due to the occurrence of beneficial reversions and selection. (**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 1-Mb windows) are recorded at the formation of each offspring, across all homologous chromosomes in the population. Only recombination events between the X and the Y chromosomes are shown for chromosome 1 (i.e., recombination events between the two X chromosomes in females are not shown). Unlike chromosome 1, chromosome 2 harbors no permanently heterozygous alleles. All inversions on this chromosome suffer from homozygote disadvantage and very few inversions therefore become fixed on chromosome 2. See S24 Fig for a simulation with $N = 10{,}000$ individuals. The datasets and scripts used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.

https://doi.org/10.1371/journal.pbio.3001698.g004

being 10, 100, 1,000, or 10,000. These genomic positions represent inversion hotspots, such as those that can be generated by repeats in genomes [51]. In each generation, we introduced j inversions, j being sampled from a Poisson distribution. Simulation times were limited by using inversion rates such that one inversion, on average, occurred in each generation, throughout the whole population. The 2 breakpoints of each inversion were chosen, at random, from the k positions. It was, therefore, possible for 2 independent inversions to appear at the same position, allowing, in particular, the reversion of an inversion to its ancestral orientation, thereby restoring recombination. Reversion may be favored if an inversion has accumulated enough deleterious mutations after its fixation on the Y chromosome to carry a larger number of mutations than the population average [52]. We assumed that inversions partially overlapped by another inversion or that captured a smaller inversion could not be reversed, i.e., recombination could not be restored in such situations even if subsequent inversions reused the same breakpoints. Indeed, reversions of partially overlapped inversions do not

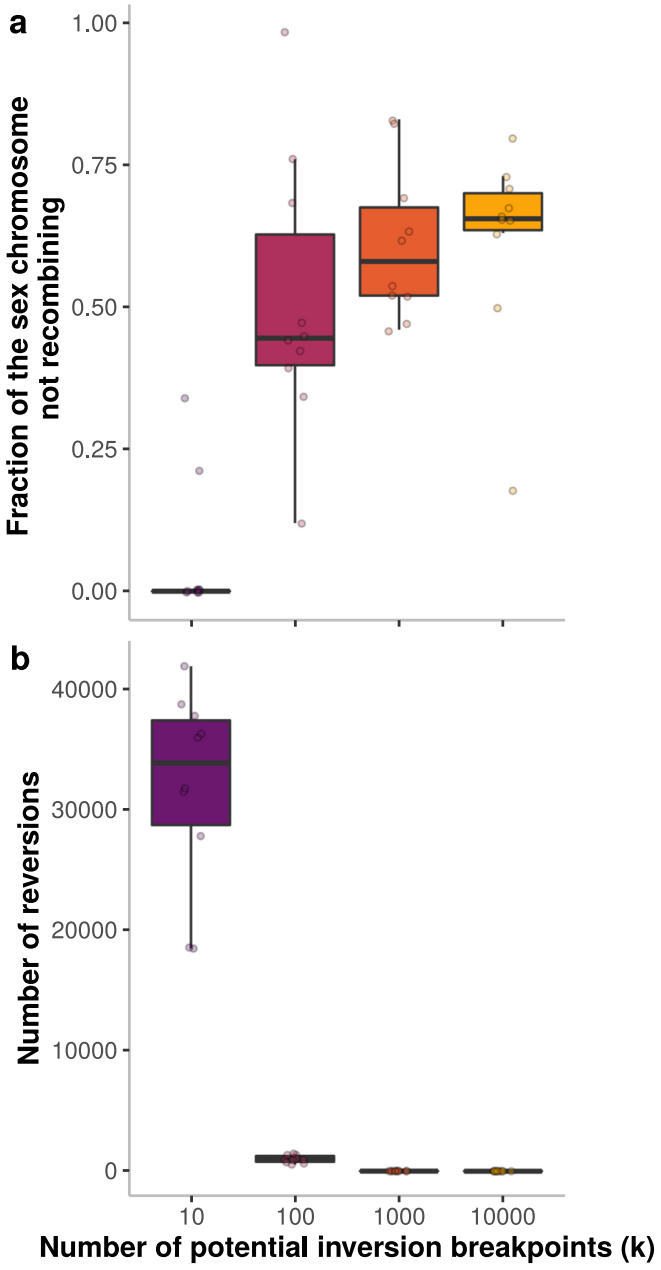


Fig 5. Effect of the number of potential inversion breakpoints on the evolution of recombination suppression at sex chromosomes. For each number of breakpoints, 10 simulations were conducted. Each dot represents the result of a simulation with $N = 1{,}000$ individuals. See Fig 4 for an example of such a simulation. (**a**) Fraction of the length of the Y sex chromosome not recombining after 100,000 generations. (**b**) Number of reversions occurring over the course of the 100,000 generations. Boxplot elements: central line: median, box limits: 25th and 75th percentiles, whiskers: 1.5× interquartile range. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.

https://doi.org/10.1371/journal.pbio.3001698.g005

restore ancestral arrangements but instead result in complex reshufflings of gene order and orientation and are therefore unlikely to restore recombination. We ran 10 simulations for each set of parameters.

In all simulations assuming relatively large numbers of potential inversion breakpoints (k = 100, 1,000, 10,000), the Y chromosome progressively stopped recombining with the X chromosome as it accumulated inversions fully linked to the male sex-determining allele (Figs 4, 5 and S24). After the occurrence of an initial inversion capturing the male sex-determining allele, multiple inversions partially overlapping this inversion or other Y-fixed inversions were selected, thereby generating a growing chaos of overlapping chromosomal rearrangements. The nonrecombining region, thus, extended around the sex-determining locus in a stepwise manner, perfectly reflecting the evolution of sex chromosomes and other supergenes with evolutionary strata (Fig 4). Some events in the gradual extension of recombination suppression were reversed, due to the occasional occurrence of beneficial reversions (Figs 4 and 5). The accumulation of overlapping inversions was, however, more rapid than the occurrence of beneficial reversions, leading to a progressive extension of the nonrecombining region (Figs 4 and 5). By contrast, when we assumed a smaller number of potential breakpoints (k = 10), recombination suppression between sex chromosomes evolved in only 2 of the 10 simulations, and across only a small genomic region (Fig 5), owing to the more frequent occurrence of beneficial reversions.

Discussion

Evolution of recombination suppression around permanently heterozygous alleles

Our results show that recombination suppression on sex chromosomes and other supergenes can evolve simply because genomes harbor many partially recessive, deleterious variants. Our model for the evolution of sex chromosomes, and supergenes in general, is based on simple and widespread phenomena: (i) inversions (or any recombination suppressor) can be favored solely because they contain fewer deleterious mutations than the population average, a situation applying to a substantial fraction of the inversions formed; (ii) such inversions tend to display overdominance: They are beneficial in the heterozygous state but suffer from a homozygote disadvantage, which prevents them from reaching high frequencies and becoming fixed on autosomes; and (iii) when, by chance, inversions capture a permanently heterozygous allele, they do not suffer from this homozygote disadvantage and are therefore able to spread until they are fully associated with the permanently heterozygous allele (e.g., they become fixed in the Y chromosome population). These 3 phenomena have been reported independently in several studies, but, to our knowledge, never in interaction (see references [28,38] for (i), [29,53] for (ii), and [5,19] for (iii)). The combined influence of the mechanisms related to (ii) and (iii) has been shown to promote sex chromosome–autosome fusion in highly inbred populations [20]. We show here that such mechanisms can readily lead to the stepwise extension of the nonrecombining region on sex chromosomes themselves, without the need for inbreeding. Moreover, unlike previous studies (e.g., [18,38,54]), we show that the higher probability of inversion fixation on Y chromosomes is not restricted to mutation-free inversions, but applies to any inversion capturing fewer deleterious mutations than the average. The proposed mechanism of deleterious mutation sheltering can also explain the fusions sometimes observed between sex chromosomes and autosomes, with the recombination suppression extending to a part of the fused autosome [20,34–36].

The theory proposed here to explain the stepwise evolution of recombination suppression applies to any locus with at least 1 permanently or nearly permanently heterozygous allele. It only requires that, within a population, individuals carry different numbers of partially recessive deleterious mutations in a genomic region that can be subjected to recombination suppression. This situation is probably frequent in diploid, dikaryotic, or heterokaryotic

organisms. For example, the mean distance between 2 heterozygous positions is approximately 1,000 bp in primates [55], and most mutations are likely partially recessive and deleterious [22,56]. Chromosomal rearrangements spanning hundreds of kilobases are frequently observed [51,57,58], and would, therefore, be expected to capture several deleterious variants.

Mutation accumulation and inversion reversions are unlikely to prevent recombination suppression extensions

On autosomes, inversions maintained at low frequencies because of their homozygote disadvantage tend to be lost rapidly because they accumulate further deleterious mutations. On the Y chromosome, contrary to previous suggestions [28], we show that the accumulation of further deleterious mutations following the formation of an inversion is generally too slow to prevent the fixation of less-loaded inversions. Deleterious mutations nevertheless accumulate following fixation of the inversion, leading to Y-linked inversion degeneration, as observed on many nonrecombining sex chromosomes [30]. This may lead to selection for reversion of the inversion, thereby restoring recombination [52]. The question as to whether this can actually occur is common to all mechanisms explaining evolutionary strata, including sexual antagonism.

We found inversion reversion to be rare, unless there were no more than 10 inversion breakpoints over the two 100-Mb chromosomes. Indeed, for reversions to invade the population and restore recombination, the following conditions must be met: (i) sufficient deleterious mutations must accumulate in the initial inversion to render reversion beneficial; and (ii) a partially overlapping inversion should not have invaded the population before the occurrence of a beneficial reversion. Indeed, partially overlapping inversions result in a complex reshuffling of gene order and orientation, preventing the restoration of recombination even in situations in which reversion could be selected. When the number of potential breakpoints is relatively high, the probability of partially overlapping inversions occurring is higher than the probability of a reversion occurring, regardless of the relative rates of inversions and deleterious mutations.

The reversion of inversions has been reported only in rare cases in which inversions occur at specific breakpoints rich in repeated elements [59,60], as in our simulations when assuming only a few possible breakpoints in the genome. The number of genomic positions at which inversions can occur in natural conditions is unknown, but this number is likely to be high given the chaos of rearrangements observed in some sex and mating-type chromosomes with hundreds of different breakpoints [16,51,61,62]. A recent study reported not only the recurrent appearance of inversions at the same positions, but also the existence of numerous potential breakpoints of inversions in human genomes [51]. Moreover, chromosomal rearrangements rapidly accumulate in recently established regions of nonrecombination [61] and overlapping inversions are observed on many sex chromosomes, which should prevent the restoration of recombination by reversions [16,62–65]. Therefore, over a wide range of realistic parameter values, the reversion of inversions should not prevent the stepwise formation of nonrecombining sex chromosomes. When there are few potential inversion breakpoints in the genome, the mechanism stabilizing inversions through dosage compensation considered in a recent model [52] may act in addition to the mechanism of sex-chromosome evolution studied here, as these 2 mechanisms are not mutually exclusive. However, by contrast to the framework of our model, the dosage compensation mechanism requires a form of sexual antagonism arising from XY-like asymmetry (i.e., one sex being heterogametic and the other homogametic); this dosage compensation mechanism [52] cannot, therefore, explain evolutionary strata on fungal mating-type chromosomes or bryophyte UV sex chromosomes [5,62].

Parameters restricting the extension of recombination suppression

Our model showed that recombination suppression could also evolve near supergenes carrying more than two permanently heterozygous alleles, such as plant self-incompatibility or mushroom (Agaricomycotina) mating-type loci. Such multi-allelic loci are however not known to often display extensions of recombination suppression beyond incompatibility loci [66]. This may be because the existence of multiple alleles allows the loss of degenerated alleles: If a permanently heterozygous allele evolves a large nonrecombining region, for example, because of an inversion fixation, and then degenerates through Muller's ratchet-like processes, it can be lost by selection, in contrast to alleles in biallelic compatibility systems. In addition, recessive self-incompatibility alleles can be homozygous in sporophytic systems [67]. In multi-allelic systems, we therefore expect to observe extensions of recombination cessation only around permanently heterozygous alleles, i.e., dominant self-incompatibility alleles, and only rare, recent and nondegenerated inversions; long-read polymorphism sequencing data would allow testing this hypothesis.

Assuming a reduced fertility in heterozygotes for the inversion due to segregation issues possibly arising during meiosis [68] decreased the probability of inversion fixation, as expected. We found that large inversions were more likely to spread and fix than small inversions as they were more likely to capture haplotypes with much fewer mutations than average, thereby compensating the heterozygous cost; larger inversions may however be also more likely to induce segregation issues, although little data is available to date.

Given the lack of knowledge about inversion rates in natural conditions and the computation challenge represented by the simulation of complex patterns of recombination, we used reasonably high inversion rates in our sex-chromosome evolution simulations, making it possible to observe the stepwise extension of a nonrecombining region within 100,000 generations. Recent studies suggested that inversions may not be too rare and that inversion breakpoints may be widely distributed throughout the genome [51,69–71]. The use of different inversion rates might result in much shorter or much longer times for the stepwise extension of the nonrecombining region, but should not change the final outcome. Of course, lower inversion rates and higher costs of inversions in terms of fertility in heterozygotes would lower the rate of inversion fixation. In nature, the stepwise extension of the nonrecombining region between sex or mating-type chromosomes often occurs over time scales of the order of tens of millions of generations (e.g., about 250 M years in human [50]), suggesting that the fixation of inversions may be relatively rare events.

Predictions regarding the variability of sex-chromosome structure across species

We show that inversions are more likely to spread in regions harboring many segregating deleterious mutations, in which they have a greater chance of capturing highly advantageous haplotypes. Our model therefore predicts that species harboring a large number of deleterious recessive variants, due to their small population size, short haploid phase, outcrossing mating system, high mutation rate, and mutation recessiveness, for example, will be more prone to the evolution of large nonrecombining regions with evolutionary strata on sex chromosomes than species with a low mutation load. In addition, in species with large population sizes, the time required for an inversion to become fixed may exceed the time for which the number of deleterious mutations in the inversion remains below average. This would prevent some inversions from becoming fixed, potentially lowering the rate of expansion of nonrecombining regions on sex chromosomes in species with large population sizes. Depending on mutation effects and on the relative rates of inversions and mutations, this could however be compensated by the occurrence of a higher number of inversions each generation in large populations. Variations in population

size, mutation rate, length of haploid phase, and mating system (outcrossing versus selfing or inbreeding) across lineages may, therefore, account for the large number of different sex-chromosome structures in nature, with some organisms maintaining homomorphic sex chromosomes and others evolving highly differentiated sex chromosomes with multiple strata [1].

The more efficient purging of recessive deleterious mutations in species with an extended haploid phase [49,50] could therefore potentially account for the smaller nonrecombining regions observed on the sex chromosomes of plants and algae than in animals [72,73]. In fungi, evolutionary strata around mating-type genes have been reported mostly in species with biallelic mating systems and whose main life stage is dikaryotic, with a very brief or inexistent haploid stage [5]. So far, an extension of recombination suppression beyond the mating-type locus in a fungus with an extended haploid phase has only be reported in *Cryphonectria parasitica*, but dikaryotic individuals are also relatively frequently found in this species [46,74,75].

A test of our model could thus look for association between the size of nonrecombining regions or the number of evolutionary strata on sex chromosomes and estimates of the number of deleterious mutations segregating in genomes, or parameters predicting the accumulation of deleterious mutations, such as population size, mating systems, length of the haploid phase, and mutation rates.

Conclusions

Given its simplicity and wide scope of application, our model of sex chromosome evolution is a powerful alternative to other explanations [6], although the various theories are not mutually exclusive. The strength of our model lies in the absence of strong assumptions, such as sexually antagonistic selection, XY heterogamety asymmetry, or small population sizes. Our model, based on the often overlooked observation that recessive deleterious mutations are widespread in genomes within natural populations, can explain the evolution of stepwise recombination suppression over a wide range of realistic parameter values. Furthermore, it can explain why some supergenes, such as those in butterflies and ants, display evolutionary strata [13,14], and why many biallelic mating-type loci in dikaryotic fungi display a stepwise extension of nonrecombining regions despite the absence of sexually antagonistic selection or XY-like asymmetry in these organisms [5,16]. Our model can also explain why meiotic drivers, which are often permanently heterozygous, are often associated with large nonrecombining region involving polymorphic chromosomal inversions [76,77]. Our model therefore provides a general and simple framework for understanding the evolution of nonrecombining regions around many kinds of loci carrying permanently heterozygous alleles.

Methods

Infinite population deterministic 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 n sites, mutations are at the same mutation-selection equilibrium frequency, denoted by q. We used nonapproximated values of q as derived by [78]:

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

We assume that the sites are independent. The number of mutations carried by a chromosomal segment of length n then follows a binomial distribution of parameters (n, q). 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 only noninverted segments. The mean fitness of a noninverted homozygote can be computed as follows (see S1 Appendix, section 2):

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

Note that in the parameter regimes where we can make the approximation $q \approx u/hs$ with $q \ll 1$, we have $W_{NN} \approx (1-2u)^n$, in accordance with mutation load theory [37,79]. Similarly, an individual who is heterozygous for an inversion with m mutations has a mean fitness of:

$$W_{NI} = (q(1-s) + (1-q)(1-hs))^m (q(1-hs) + 1 - q)^{n-m}.$$

Assuming that $q \ll 1$, Nei and colleagues [28] considered that individuals heterozygous for an inversion had no homozygous mutations, so that $W_{NI} \approx (1-hs)^m (1-hs)^{nq} = (1-hs)^{nq+m}$. Note that we use nonapproximated values in our computations. An individual who is homozygous for a segment I with m mutations is homozygous for all these mutations. Its fitness can therefore be expressed as:

$$W_{II} = (1-s)^m.$$

The inversion frequency trajectory can be determined with a simple 2-locus 2-allele model. We considered 4 different situations, depending on the possible heterozygosity at the locus with permanently heterozygous alleles. The S1 Appendix, sections 4 to 8, presents the evolution in time of the frequency of the inversion in detail in these 4 situations. Here, we briefly describe the results for inversions more or less linked to a permanently heterozygous allele in a XY system. The change in frequency of inversions on the Y chromosome, on the X chromosome, or on autosomes between generations t and $t+1$ (Fig 2C) is described by

$$F_{XIm}^{t+1} = \frac{F_{XIf}^t (W_{II} F_{YI}^t + W_{NI} F_{YN}^t) + r W_{NI} D^t}{\overline{W_m^t}},$$

$$F_{XNm}^{t+1} = \frac{F_{XNf}^t (W_{NN} F_{YN}^t + W_{NI} F_{YI}^t) - r W_{NI} D^t}{\overline{W_m^t}},$$

$$F_{XIf}^{t+1} = \frac{F_{XIm}^t F_{XIf}^t W_{II} + \frac{1}{2} W_{NI} \left(F_{XIm}^t F_{XNf}^t + F_{XIf}^t F_{XNm}^t \right)}{\overline{W_f^t}},$$

$$F_{XNf}^{t+1} = \frac{F_{XNm}^{t} F_{XNf}^{t} W_{NN} + \frac{1}{2} W_{NI}\left(F_{XNm}^{t} F_{XIf}^{t} + F_{XNf}^{t} F_{XIm}^{t}\right)}{\overline{W_f^t}},$$

$$F_{YI}^{t+1} = \frac{F_{YI}^{t}(W_{II} F_{XIf}^{t} + W_{NI} F_{XNf}^{t}) - r W_{NI} D^{t}}{\overline{W_m^t}},$$

$$F_{YN}^{t+1} = \frac{F_{YN}^{t}(W_{NN} F_{XNf}^{t} + W_{NI} F_{XIf}^{t}) + r W_{NI} D^{t}}{\overline{W_m^t}},$$

$$F_{XI} = \frac{2}{3} F_{XIf} + \frac{1}{3} F_{XIm},$$

$$F_{I} = \frac{1}{4} F_{YI} + \frac{3}{4} F_{XI}, \qquad \text{(Eq 1)}$$

where F_{YI}^{t} is the frequency of the inversion in the population of Y chromosomes at time t, F_{XIf}^{t} is the frequency of the inversion on the X chromosome in females (respectively, F_{XIm}^{t} in males), r is the rate of recombination between the inversion and the sex-determining locus, D is their linkage disequilibrium such that $D^{t} = F_{XNf}^{t} F_{YI}^{t} - F_{XIf}^{t} F_{YN}^{t}$, and $\overline{W_m^t}$ is the mean male fitness (Fig 2C; see S1 Appendix, section 8 for details). When $r = 0.5$, this system of equations describes the evolution of inversions on autosomes. Unless otherwise stated, the deterministic simulations presented here (Figs 2C and S4 to S8) were performed with an initial $D = -0.01$ or $D = 0.01$, depending on the allele which the inversion appeared linked to. Results for various initial linkage disequilibrium values are presented in S6 Fig. When the inversion captures the male-determining allele, $r = 0$, $F_{XIm} = F_{XIf} = 0$, and $F_{XNf} = F_{XNm} = 1$ at any time t. The equations for F_{YI}^{t}, the frequency of inversions capturing the male-determining allele then reduces to

$$F_{YI}^{t+1} = F_{YI}^{t} \frac{W_{NI}}{\overline{W_m^t}}$$

$$F_{YN}^{t+1} = F_{YN}^{t} \frac{W_{NN}}{\overline{W_m^t}}$$

with $\overline{W_m^t} = F_{YN}^{t} W_{NN} + F_{XI}^{t} W_{NI}$. Substituting F_{YN} by $1-F_{YI}$ and subtracting F_{YI}^{t}, we have $\Delta F_{YI} = \frac{F_{YI}^{t}(1-F_{YI}^{t})(W_{NI}-W_{NN})}{\overline{W_m^t}}$.

The search for F_{YI} such that $\Delta F_{YI} = 0$ readily gives 2 equilibria: 0 and 1. Since $\overline{W_m^t} > 0$ and $0 \leq F_{YI} \leq 1$, we can conclude that:

If $W_{NI} > W_{NN}$, $F_{YI} \rightarrow 1$,

if $W_{NI} < W_{NN}$, $F_{YI} \rightarrow 0$.

In the case of an inversion appearing on an autosome ($r = 0.5$), $F_{XIm} = F_{XIf} = F_{YI}$ and $\overline{W_m} = \overline{W_f} = \overline{W}$, so that the change in inversion frequency in the population is:

$$F_{I}^{t+1} = \frac{(F_{I}^{t})^{2} W_{II} + F_{I}^{t} F_{N}^{t} W_{NI}}{\overline{W_t}}$$

and

$$\Delta F_I = \frac{F_I}{\overline{W}}\left(F_I^2(2\,W_{NI} - W_{II} - W_{NN}) + F_I(W_{II} + 2W_{NN} - 3W_{NI}) + W_{NI} - W_{NN}\right)$$
$$= \frac{F_I}{\overline{W}} \times P(F_I),$$

with $P(X) = X^2(2\,W_{NI} - W_{II} - W_{NN}) + X(W_{II} + 2W_{NN} - 3W_{NI}) + W_{NI} - W_{NN}$.

The equilibria are 0 and the roots of the polynomial P, which are 1 and $\frac{W_{NI}-W_{NN}}{2W_{NI}-W_{II}-W_{NN}}$. We thus have:

If $W_{NI} > W_{NN}$ and $W_{II} > W_{NI}$, then $F_I \to 1$,

If $W_{NI} > W_{NN}$ and $W_{II} < W_{NI}$, then $F_I \to \frac{W_{NI}-W_{NN}}{2W_{NI}-W_{II}-W_{NN}}$,

If $W_{NI} < W_{NN}$, then $F_I \to 0$.

Inversion equilibrium frequencies therefore depend, as expected, on the relative fitness of homozygotes and heterozygotes for the inversion and of the noninverted homozygotes. We thus derive the conditions for the inversion to be favored or disfavored as a function of the number of mutations captured by the inversion (m). A straightforward computation gives (see the S1 Appendix, section 9):

$W_{II} > W_{NI}$ if and only if $m < \beta_1(q, h, s) \times n$,

$W_{NI} > W_{NN}$ if and only if $m < \beta_2(q, h, s) \times n$,with

$$\beta_1(q,h,s) = \frac{\ln(q(1-hs)+1-q)}{\ln\left(\frac{(1-s)(q(1-hs)+1-q)}{q(1-s)+(1-q)(1-hs)}\right)} = \frac{\ln(1-qhs)}{\ln\left(\frac{(1-s)(1-qhs)}{1-hs-qs(1-h)}\right)}, \quad \text{(Eq 2)}$$

$$\beta_2(q,h,s) = \frac{\ln\left(\frac{1-2q(1-q)hs-q^2s}{q(1-hs)+1-q}\right)}{\ln\left(\frac{q(1-s)+(1-q)(1-hs)}{q(1-hs)+1-q}\right)}. \quad \text{(Eq 3)}$$

Assuming $q \ll 1$ and $s \ll 1$, these quantities can be approximated by $\beta_1 \approx q\frac{h}{1-h}$ and $\beta_2 \approx q$ (the latter in accordance with a previous study [28]). When $\frac{qhn}{1-h} < m < nq$, inversions should thus go to fixation on Y chromosomes and stabilize at intermediate frequency on autosomes. When $m < \frac{qhn}{1-h}$, inversions should go to fixation on autosomes and on the Y chromosome. Observe that the closer h is to ½ (i.e., the scenario without dominance), the smaller the difference between the thresholds β_1 and β_2 is. When $h = 0.5$, $\beta_1 = \beta_2$, inversions should therefore go to fixation on autosome when they have fewer mutations than average, as they have then no homozygous disadvantage preventing their fixation. In contrast, when h is small, β_1 is significantly smaller than β_2, showing that the condition for heterozygotes to be favored over noninverted homozygotes ($W_{NI} > W_{NN}$) is much easier to meet than for inversion homozygotes to be favored over inversion heterozygotes ($W_{II} > W_{NI}$): inversions are therefore much likely to be maintained at intermediate frequencies on autosomes than to fix. See S1 Fig for a graphical representation of these results and the S1 Appendix, section 9, for derivation details.

We can use these equilibrium frequencies as functions of m to compute the expected equilibrium frequency of inversions that can occur in the genome (F_{equ}, Figs 2D and S9). To do so, we sum for all $m \in \{0,\ldots,n\}$ the equilibrium frequencies of inversions ($F_{equ,m}$) weighted by their occurrence probability (P_m). For inversions on the Y chromosome, we therefore have:

$$[F_{equ}] = \sum_{m=0}^{n} P_m \times F_{equ,m} = \sum_{m=0}^{n} \binom{n}{m} q^m (1-q)^{n-m} F_{equ,m} = \sum_{m=0}^{\lfloor nq \rfloor} \binom{n}{m} q^m (1-q)^{n-m}. \quad \text{(Eq 4)}$$

For inversions on autosomes, we obtain, using the approximate values for β_1 and β_2 derived earlier in the case $q \ll 1$ and $s \ll 1$:

$$\left[F_{equ}\right] = \sum_{m=0}^{\lfloor nq\frac{h}{1-h}\rfloor} \binom{n}{m} q^m (1-q)^{n-m} + \sum_{m=\lfloor nq\frac{h}{1-h}\rfloor+1}^{\lfloor nq\rfloor} \binom{n}{m} q^m (1-q)^{n-m} \frac{W_{NI} - W_{NN}}{2W_{NI} - W_{NN} - W_{II}}. \quad \text{(Eq 5)}$$

The expected equilibrium frequency of less-loaded inversions (Fig 2D) is therefore the expected equilibrium frequency of all inversions divided by the probability of occurrence of less-loaded inversions:

$$\left[F_{equ}^{less\ loaded}\right] = \frac{\left[F_{equ}\right]}{\sum_{m=0}^{\lfloor nq\rfloor} P_m} = \frac{\left[F_{equ}\right]}{\sum_{m=0}^{\lfloor nq\rfloor} \binom{n}{m} q^m (1-q)^{n-m}}. \quad \text{(Eq 6)}$$

The fate of inversions associated with a heterozygote fitness cost is described in section 10 of the S1 Appendix.

Individual-based simulations

We used SLiM V3.2 [80] to simulate the evolution of a single panmictic population of $N = 1{,}000$ or $N = 10{,}000$ individuals in a Wright–Fisher model. To assess the fate of inversions under various conditions (Fig 3), we simulated individuals with 2 pairs of 10-Mb chromosomes on which mutations occurred at a rate u, with u ranging from 10^{-6} to 10^{-9} per bp, their dominance coefficient h ranged from 0 to 0.5 (0, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5) and their selection coefficient s from −0.5 to 0 (0, −0.001, −0.01, −0.1, −0.25, −0.5). The main and supplementary figures show the results for $\mu = 10^{-8}$ and $\mu = 10^{-9}$ unless otherwise stated. We also simulated populations where each occurring mutation had its selection coefficient s drawn from a gamma distribution with a shape of 0.2, and its dominance coefficient h randomly sampled among 0, 0.001, 0.01, 0.1, 0.25, 0.5 with uniform probabilities (S17 Fig). 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}. Among the 5,000,000 sites on chromosome 1, a single segregating locus was subject to balancing selection, for which several situations were considered: (i) the locus had 2 alleles, only one of which was permanently heterozygous, mimicking a classical XY (or ZW) determining system (Fig 3); (ii) the locus had 2 permanently heterozygous alleles, mimicking, for example, the situation encountered at most fungal mating-type loci; (iii) the locus had 3 (or more) permanently heterozygous alleles, mimicking, for example, the situation encountered in plant self-incompatibility systems and mushroom (Agaromycotina) mating-type loci.

For each parameter combination (u, h, s, N, heterozygosity rule at the locus with a permanently heterozygous allele), a simulation was run for 15,000 generations to allow the population to reach an equilibrium for the number of segregating mutations. S25 Fig shows that each population had reached equilibrium by the end of the burn-in period. 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 500 kb, 1,000 kb, 2,000 kb, and 5,000 kb were then introduced on chromosome 1 around the locus under balancing selection (X-linked or Y-linked inversions) or on chromosome 2 (autosomal inversions). For each parameter combination (h, s, u, N, heterozygosity rule, size of the region affected by the

recombination modification and position on the genome), we ran 10,000 independent simulations starting with the introduction of a single 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., as a *cis*-recombination modifier). We monitored the frequency of these inversion-mimicking mutations during 10,000 generations, during which all evolutionary processes (such as point mutation, recombination, and mating) remained unchanged, e.g., mutations were still appearing on inversions following their formation. 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 (S14 Fig).

To study the effect of the existence of a haploid phase on the accumulation of deleterious mutations and the spread of inversions on autosomes and sex chromosomes, we performed additional simulations, involving 15,000 generations of burn-in and the introduction of 10,000 inversions under each combination of parameters in these initial populations, as above. The populations were considered to harbor a single locus with 2 permanently heterozygous alleles (similarly to the previous situation (ii)). Every x generation, populations experimented a haploid generation, with x taking the values 2, 3, 5, 10, or 100. For simulating haploidy, all mutations from one chromosome of each pair ("genome2" in SLiM) were removed, and the dominance coefficient of mutations on the other chromosome ("genome1" of each pair) was set to 1. The recombination rate was set to 0, and mating could only occur between gametes that were derived from the first chromosome of each pair. Therefore, during haploid generations, selection acted only on the first chromosome of each pair, and the second chromosome had no contribution to the following generation. These modifications allowed simulating the occurrence of haploid phases without changing most parameters or model behavior. Note that, during the haploid phase, the number of individuals remained unchanged, but the number of haploid genomes was divided by 2, because the second haploid genome of each pair had no contribution to the following generation.

To study the evolution of chromosomal fusion (S18 Fig), we simulated populations with no recombination between X and Y chromosomes (chromosome 1) between the position 1 Mb and 9 Mb during 15,000 generations (burn-in), mimicking the evolution of a population with old sex chromosomes and small pseudoautosomal regions (1 Mb on each chromosome edge). Then, we introduced a fusion-mimicking mutation resulting in the linkage of 1 sex chromosome (chromosome 1, X or Y) and 1 autosome (chromosome 2), and suppressing the recombination when heterozygous over 1 Mb of the fused side of each chromosome (see S18 Fig for a graphical representation). Therefore, these mutations behave like 2-Mb inversions that would also lead to chromosome fusion and result in the extension of the size of the nonrecombining region from 8 Mb to 10 Mb. We tracked the frequency of 10,000 X-autosome and Y-autosome fusion-mimicking mutations for each parameter combination (as done before for inversion-mimicking mutations).

To study more specifically the formation of evolutionary strata on sex chromosomes (Figs 4 and 5), we also simulated the evolution of two 100-Mb chromosomes, one of which carried an XY sex-determining locus at the 50-Mb position, over 115,000 generations (including an initial burn-in of 15,000 generations): individuals could be either XX or XY and could only mate with individuals of a different genotype at this locus. We simulated randomly mating populations of $N = 1{,}000$ and $N = 10{,}000$ individuals. Point mutations appeared at a rate of $\mu = 10^{-9}$ per bp, and their individual selection coefficients were determined by sampling a gamma distribution with a mean of −0.03 and with a shape of 0.2; these parameter values were set according to observations in humans [22,81]. For each new mutation, a dominance coefficient was

chosen from the following values, considered to have uniform probabilities: 0, 0.001, 0.01, 0.1, 0.25, 0.5. At the beginning of each simulation, we randomly sampled k genomic positions over the 2 chromosomes that could be used as inversion breakpoints, with k being 10, 100, 1,000, or 10,000. After the 15,000 generations of the burn-in period allowing populations to reach an equilibrium in terms of the number of segregating mutations, we introduced each generation j inversions in the population, j being sampled from a Poisson distribution of parameter λ, with $\lambda = N \times k \times u_i$, u_i being the inversion rate. In order to keep the simulation time tractable, we used inversion rates allowing 1 inversion to occur on average each generation in the population (i.e., $N \times k \times u_i = 1$). For each inversion, the first breakpoint was randomly chosen among the k positions, and the second among the potential breakpoint positions less than 20 Mb apart on the same chromosome (considering therefore only a subset of the k positions for the second breakpoints). Two independent inversions could use the same breakpoints, allowing in particular inversion reversion restoring recombination. If 2 independent inversions occurred with the same breakpoints on different haplotypes (for example, on the X and on the Y chromosome), we assumed that recombination was restored between these haplotypes, as a reversion would do. The occurrence of partially overlapping inversions (i.e., with different breakpoints) on different haplotypes did not restore recombination between these haplotypes. We assumed that inversions subsequently partially overlapped by another inversion or that captured a smaller inversion could not reverse, i.e., the restoration of recombination by further inversions, even if using the same breakpoints, was prevented. (For $N = 1{,}000$, for each values of k (i.e., 10, 100, 1,000, 10,000), we ran 10 simulations (Fig 5). For $N = 10{,}000$, because of computing limitation (each simulation taking about 3 weeks to run), we only ran 1 simulation per number of breakpoints.

Simulations were parallelized with GNU Parallel [82].

Supporting information

S1 Fig. Distribution of the occurrence probability of inversions depending on the number of mutations they capture. This figure represents the binomial distribution of the number of mutations on a segment of $n = 2{,}000{,}000$ sites, with mutations occurring at a rate $\mu = 10^{-8}$ and having selective effect of $s = 0.01$ and $h = 0.1$. The filling color of the distribution represents the expected state of the inversions on autosomes based on Eqs 2 and 3 (Methods). The probability of occurrence of an inversion carrying fewer than $m = qhn/(1-h)$ mutations, and therefore fixing on the autosome (dark blue filling), is so low that it is not visible on this plot. The script used to produce this figure is available on GitHub.
(PDF)

S2 Fig. Fate of neutral chromosomal inversions on autosomes depending on the relative number of mutations they carry compared to population average and the relative frequency of these mutations. Each dot represents a 2-Mb inversion on an autosome. For each parameter combination, the fate (i.e., being lost or still segregating) of 10,000 inversions after 1,000 generations in stochastic simulations of a population of $N = 1{,}000$ individuals is displayed depending on the number of mutations they captured upon formation relative to the population average. Only inversions not linked to a permanently heterozygous allele were considered. Results for sex-linked inversions are displayed in S3 Fig. The Y axis represents the number of mutations captured by the inversions upon formation relative to the mean number of mutations within the same region in noninverted segments. Dot color represents the mean of the frequencies (in the entire population) of the mutations captured by inversions relative to the mean of the frequencies (in the entire population) of mutations in noninverted segments: A blue dot represents an inversion capturing mutations rarer than average, whereas a red dot

represents an inversion capturing mutations more frequent than average. Boxplot elements: central line: median, box limits: 25th and 75th percentiles, whiskers: 1.5× interquartile range. The vast majority of inversions still segregating after 1,000 generations carried fewer or rarer mutations than the population average. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S3 Fig. Fate of neutral chromosomal inversions linked to a permanently heterozygous allele depending on the relative number of mutations they carry compared to population average and the relative frequency of these mutations. Similar to S2 Fig but with inversions linked to a permanently heterozygous allele (such as a Y sex-determining allele). For each parameter combination, the fate (i.e., being lost or still segregating) of 10,000 inversions after 1,000 generations in stochastic simulations is displayed depending on the number of mutations captured upon formation relative to the population average. Dot color represents the mean of the frequencies (in the entire population) of the mutations captured by inversions relative to the mean of the frequencies (in the entire population) of mutations in noninverted segments: A blue dot represents an inversion capturing mutations rarer than average, whereas a red dot represents an inversion capturing mutations more frequent than average. Boxplot elements: central line: median, box limits: 25th and 75th percentiles, whiskers: 1.5× interquartile range. The vast majority of inversions still segregating after 1,000 generations carried fewer or rarer mutations than the population average. The dataset and script used to produce this figure are available on Fig share (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S4 Fig. Deterministic trajectories of inversion frequency when in linkage with the sex-determining allele on the Y chromosome or on the X chromosome. Similar to Fig 2C, but considering both X-linked and Y-linked inversions, and various recombination rates between the sex-determining locus and inversions. Trajectory of inversions is represented for different values of the selection and dominance coefficients of their mutations and the linkage of these inversions to an X or to a Y sex-determining allele. Mutation frequencies are at mutation-selection equilibrium (see Methods) with a mutation rate of $\mu = 10^{-8}$. Inversions of 2 Mb are introduced with an initial linkage disequilibrium of $D = 0.01$ or $D = -0.01$, depending on whether they appear linked to the male-determining allele on the Y chromosome or to the female-determining allele on the X chromosome. The figure illustrates the case of inversions carrying a number of mutations 20% lower than the population average ($m = \lfloor 0.8 \times nq \rfloor$), as frequently observed (S1–S3 Figs). (**a**) Inversions appearing in linkage with a sex-determining allele on the Y chromosome (permanently heterozygous). The frequency of the inversion in the population of Y chromosomes is followed during the first 10,000 generations. Linkage strength is indicated by line colors (distance in cM between inversions and the sex-determining allele). Inversions at 50 cM from the sex-determining locus behave like inversions on autosomes. There is a nearly perfect overlap between curves representing the 1, 5, and 50 cM linkage cases, so that only 50 cM curves are visible. (**b**) Inversions appearing in linkage with the sex-determining allele on the X chromosome. The frequency of the inversion in the population of X chromosomes is followed during the first 10,000 generations. Linkage strength is indicated by line colors (distance in cM between inversions and the sex-determining allele). See S1 Appendix, section 8, for modeling details. The script used to produce this figure is available on GitHub.
(PDF)

S5 Fig. Deterministic trajectory of inversion frequency depending on their initial mutation load and their linkage to a locus with 2 permanently heterozygous alleles. Similar to Figs 2C and S4 but considering inversions in linkage to a locus with 2 permanently heterozygous alleles (and not only 1 as in the XY systems), and various mutation loads associated with the inversions upon formation. Trajectory of inversions in an infinite population depending on the selection coefficient of the segregating mutations, the dominance coefficient of these mutations, and the linkage of these inversions to a permanently heterozygous locus with 2 alleles. Mutations are at mutation-selection equilibrium frequencies (see Methods) with a mutation rate of $\mu = 10^{-8}$. Inversions of 2 Mb are introduced with an initial linkage disequilibrium of $D = 0.01$. Inversions are linked to only 1 of the 2 alleles at the permanently heterozygous locus, and so even fully linked inversions can increase up to a maximum frequency of 50%. Linkage strength is indicated by line colors (distance in cM between inversions and permanently heterozygous locus). Inversions at distances larger than 50 cM from the sex-determining locus behave like inversions on autosomes. The figure illustrates the case of inversions carrying varying numbers of mutations compared to the population average. (**a**) Case of inversions with 20% fewer mutations than the population average ($m = \lfloor 0.8 \times nq \rfloor 00$). (**b**) Case of inversions with 10% fewer mutations than the population average ($m = \lfloor 0.9 \times nq \rfloor$). (**c**) Case of inversions with 1% fewer mutations than the population average ($m = \lfloor 0.99 \times nq \rfloor$). In all cases, there is a nearly perfect overlap between curves representing the 1, 5, and 50 cM linkage cases, so that only 50-cM curves are visible. Individual-based simulations of the same system are shown in S14 Fig. See S1 Appendix, section 6, for modeling details. The script used to produce this figure is available on GitHub.
(PDF)

S6 Fig. Deterministic trajectory of inversion frequency depending on their linkage to a permanently heterozygous locus with 2 alleles (in centimorgan) and on their initial association (initial linkage disequilibrium). Trajectory of inversions in infinite populations as a function of the selection coefficient of the segregating mutations, their linkage to a permanently heterozygous locus with 2 alleles, and the initial level of linkage disequilibrium (D) between the inversion locus and the permanently heterozygous locus. Mutations are at mutation-selection equilibrium frequencies. Inversions of 2 Mb are introduced at different frequencies in the population, and with different levels of association with 1 of the 2 permanently heterozygous alleles, which therefore generates various initial linkage disequilibrium values (D). Inversions appear more or less linked to a locus with 2 permanently heterozygous alleles; completely linked inversions can therefore only increase up to a maximum frequency of 50%. Linkage strength is indicated by line colors, representing the distance in cM between inversions and the permanently heterozygous alleles/locus. Inversions at more than 50 cM from the permanently heterozygous locus behave like inversions on autosomes. The figure illustrates the case of inversions carrying a number of mutations 20% lower than the population average ($m = \lfloor 0.8 \times nq \rfloor$), as frequently observed (S1–S3 Figs). A dominance coefficient of $h = 0.01$ was used for these deterministic simulations. (**A**) Inversion change in frequency. Similar to Figs 2c and S5 but considering inversions in linkage with a locus with 2 permanently heterozygous alleles, and various initial levels of linkage disequilibrium between the inversion locus and the permanently heterozygous locus. (**B**) Change in the level of linkage disequilibrium (D) between the inversion locus and the permanently heterozygous locus. If there is no initial linkage disequilibrium (i.e., inversions are introduced at the same frequency in linkage with each of the 2 permanently heterozygous alleles), the inversion does not spread. As soon as there is an initial linkage disequilibrium, even very small, this linkage disequilibrium increases and the inversion spreads. Here, the initial linkage disequilibrium is artificially generated by

introducing the inversions with different levels of association with a permanently heterozygous allele. In finite populations, this non-null initial linkage disequilibrium can easily be generated by founder-like effects (an inversion typically appearing on a single haplotype) or by the random fluctuations of inversion frequency induced by drift. See S1 Appendix, section 6, for modeling details. The script used to produce this figure is available on GitHub.
(PNG)

S7 Fig. Deterministic trajectory of inversion frequency when in linkage to a permanently heterozygous locus with 3 alleles. Similar to Figs 2c and S3–S5 but considering inversions in linkage to a locus with 3 permanently heterozygous alleles. Trajectory of inversions depending on the selection and dominance coefficients of their mutations and their degree of linkage to the permanently heterozygous locus. Mutation frequencies are at mutation-selection equilibrium (see Methods) with a mutation rate of $\mu = 10^{-8}$. Inversions of 2 Mb are introduced at a 0.01 frequency in linkage with 1 of the 3 permanently heterozygous alleles. The figure illustrates the case of inversions carrying a number of mutations 20% lower than the population average ($m = \lfloor 0.8 \times nq \rfloor$), as frequently observed in simulations (S1–S3 Figs). Linkage strength is indicated by line colors, representing the distance in cM between inversions and permanently heterozygous alleles. When fully linked to the permanently heterozygous locus (0 cM), inversions can increase up to a maximal frequency of 1/3, meaning that they are fully associated to 1 of the 3 permanently heterozygous alleles. See S1 Appendix, section 7, for modeling details. The script used to produce this figure is available on GitHub.
(PDF)

S8 Fig. Deterministic trajectory of inversion frequency when in linkage with an overdominant locus with no permanently heterozygous alleles. Similar to Fig 2C but with an overdominant locus with 2 alleles. Trajectory of inversions depending on the selection and dominance coefficients of their mutations and the strength of selection acting on the overdominant locus (with no permanently heterozygous alleles). Mutation frequencies are at mutation-selection equilibrium (see Methods) with a mutation rate of $\mu = 10^{-8}$. The figure illustrates the case of 2-Mb inversions carrying a number of mutations 20% lower than the population average ($m = \lfloor 0.8 \times nq \rfloor$). Inversions appear in full linkage ($r = 0$) with 1 of the 2 alleles at an overdominant locus where both homozygotes have reduced fitness ($1-s_2$; when s_2 is close to 1, homozygotes suffer from a strong reduction in fitness and are therefore rare). See S1 Appendix, section 5, for modeling details. The script used to produce this figure is available on GitHub.
(PDF)

S9 Fig. Expected equilibrium frequency of 2-Mb inversions as a function of selection and dominance coefficients. Expected equilibrium frequency of all 2-Mb inversions that can occur capturing the male-determining allele in an XY system (right column) or on an autosome (left column). Therefore, this is similar to Fig 2D but considering all inversions, i.e., inversions less loaded and more loaded than average. The expected equilibrium frequency was calculated using Eqs 4 and 5 (Methods). The frequency displayed for Y-linked inversions is the frequency of inversions in the population of Y chromosomes. Y-linked inversions become fixed (equilibrium frequency = 1) when $m<nq$ and are lost (equilibrium frequency = 0) when $m>nq$ (see Methods and main text). These 2 events have nearly equal probabilities when nq is high (Figs 2B and S1), and therefore the expected equilibrium frequency of Y-linked inversion is often around 0.5. The script used to produce this figure is available on GitHub.
(PDF)

S10 Fig. Evolution of 2-Mb inversions in stochastic simulations during 10,000 generations. Similar to Fig 3A but with other parameter values (s, h). For each parameter combination, 10,000 inversions of 2,000 kb were simulated in a population harboring an XY locus (2 alleles, one being permanently heterozygous). Populations of 1,000 individuals were simulated. We considered inversions fully linked to the Y sex-determining allele, fully linked to the X sex-determining allele or on an autosome. Only inversions not lost after 20 generations are displayed. In contrast to Fig 3A, the overall frequency of Y-linked and X-linked inversion is displayed (instead of the frequency of inversion in the Y chromosome population). Inversions fixed on the Y chromosome have therefore a 0.25 frequency (and 0.75 for X-linked inversions). For each parameter combination, the number of inversions lost and segregating or fixed at the end of the simulation is indicated above lines (« c =. . . »). Because $N = 1{,}000$, mutations with $s = 0.001$ (i.e., 1/N) are nearly neutral, which explains that X-linked or autosomal inversions can spread in the population, in contrast to what is observed with $s > 0.001$. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S11 Fig. Evolution of 2-Mb inversions in stochastic simulations over 10,000 generations in populations of $N = 10{,}000$ individuals with $s = 0.01$. Similar to Fig 3A and 3B but with $N = 10{,}000$. (**a**) Change in inversion frequency in stochastic simulations with 10,000 individuals experiencing recessive deleterious mutations at a rate $\mu = 10^{-8}$, including following inversion occurrence, all mutations having the same dominance and selection coefficients (here, $h = 0.1$ and $s = 0.01$). The figure displays the frequency of inversions on an autosome and on a proto-Y chromosome, each line representing a specific inversion. (**b**) Fluctuations in the mean number of mutations carried by the inverted segments in each of the 10,000 simulations, each line representing 1 simulation. In contrast to what happens with $N = 1{,}000$ (Fig 3A and 3B), some inversions rise in frequency on the Y chromosome; however, the high population size renders the time for inversion fixation too long, they accumulate deleterious mutations before they can fix and therefore all inversions were lost at the simulation end. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S12 Fig. Evolution of 2-Mb inversions in stochastic simulations over 10,000 generations in populations of $N = 10{,}000$ individuals with $s = 0.001$. Similar to S11 Fig but with $s = 0.001$. (**a**) Change in inversion frequency in stochastic simulations with 10,000 individuals experiencing recessive deleterious mutations at a rate $\mu = 10^{-8}$, including after inversion occurrence, all mutations having the same dominance and selection coefficients (here, $h = 0.1$ and $s = 0.001$). The figure displays the frequency of 10,000 inversions on an autosome and on a proto-Y chromosome, each line representing a specific inversion. (**b**) Fluctuations in the mean number of mutations carried by the inverted segments in each of the 10,000 simulations, each line representing 1 simulation. In contrast to the S11 Fig, several inversions fix on the Y chromosome before they accumulate too many mutations. Inversions take more time to fix than when $N = 1{,}000$ (Figs 3 and S10). The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S13 Fig. The accumulation of mutations in autosomal and sex-linked inversions in finite populations strongly departs from infinite population approximation predictions. (**a–b**) Change in inversion frequency in stochastic simulations with $N = 1{,}000$ individuals

experiencing recessive deleterious mutations at a rate $\mu = 10^{-8}$ or $= 10^{-9}$, all mutations having the same dominance and selection coefficients (here, $h = 0.1$ and $s = 0.01$). The figure displays the frequency of 2-Mb inversions on an autosome and on a proto-Y chromosome. Each line represents the trajectory of one of these 10,000 simulations under each set of parameters. The lines end when no more inverted segments segregate in the population (i.e., when all inversions are lost or fixed). For $\mu = 10^{-8}$, the same plots as in Fig 3A are presented but with a different scale for autosomal inversions. (**c–d**) Minimum number of mutations found in inverted segments in stochastic simulations. Note the different scales of the x axes in panels c and d: as shown in panel a–b, no inversion segregated until the end of the simulation in the autosome (10,000 generations), thereby explaining the reduced x axis. (**e–f**) Same as panel c–d but displaying the mean number of mutations in inverted segments (instead of the minimum number). For $\mu = 10^{-8}$, the same plots as in Fig 3B are presented (but with different scales for autosomal inversions). (**g**) Deterministic change in inversion mutation number in infinite populations for a mutation rate of $\mu = 10^{-8}$ or $\mu = 10^{-9}$, as defined by Nei and colleagues (1967), and by Connallon and Olito (2021). All mutations were considered to have the same dominance and selection coefficients (here, $h = 0.1$ and $s = 0.01$). We considered here that inverted segments were initially mutation free. This shows that the observed accumulation of mutations in inversions on autosomes or on the Y chromosome strongly departs from the infinite population expectations.
(PDF)

S14 Fig. Evolution of 2-Mb recombination suppressors in linkage with a locus with 2 permanently heterozygous alleles in stochastic simulations during 10,000 generations. Similar to Fig 3A but with a locus with 2 permanently heterozygous alleles (instead of one in the case of the XY system) and with other recombination suppressors than chromosomal inversions. For each parameter combination, 10,000 recombination suppressors of 2 Mb were simulated. These recombination modifiers suppress recombination within the segment in which they reside, both when heterozygous and when homozygous. These recombination suppressors were either fully linked ($r = 0$) or fully unlinked ($r = 0.5$) to a locus with 2 permanently heterozygous alleles; recombination suppressors fully linked to 1 of the 2 permanently heterozygous alleles can spread to a maximum frequency of 50%. Only recombination suppressors not lost after 20 generations were considered for this figure. At the end of the simulation, all recombination suppressors display either a 0.0 or a 0.5 frequency, meaning they are either lost or fixed to a given permanently heterozygous allele. The number of recombination suppressors in each state at the end of the simulation is indicated above lines (« c =. . . »). Populations of $N = 1{,}000$ individuals were simulated. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S15 Fig. Proportions of inversions in stochastic simulations that were fixed after 10,000 generations for different parameter value combinations and with $N = 1{,}000$. Similar to Fig 3 but considering different sizes of inversions. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S16 Fig. Proportions of inversions in stochastic simulations that were fixed after 10,000 generations for different parameter value combinations with $N = 10{,}000$. Similar to Figs 3 and S15 but with $N = 10{,}000$ and for different sizes of inversions. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and

GitHub.
(PDF)

S17 Fig. Proportions of inversions in stochastic simulations that were fixed after 10,000 generations for different parameter value combinations with N = 1,000. Similar to Figs 3 and S15 but considering that mutations segregating in the genome have their fitness effects drawn from a gamma distribution with a shape of 0.2, and their dominance coefficient h randomly sampled among 0, 0.001, 0.01, 0.1, 0.25, 0.5 with uniform probabilities. Simulations were run considering that the mean of the selection coefficient values (mean of the gamma distribution) was either 0.001, 0.01, 0.1, or 0.5. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S18 Fig. Evolution of fusion-mimicking mutations in stochastic simulations during 10,000 generations. (**a**) Graphical representation of the chromosome fusions simulated. The fusion-mimicking mutations result in the linkage of 2 chromosomes of each pair and in the suppression of recombination in heterozygotes over 1 Mb in the fused side of each chromosome. Therefore, these mutations behave as 2-Mb inversions that would also lead to chromosome fusion. (**b**) Similar to Fig 3A but considering the fusion-mimicking mutations instead of inversions. Simulations where performed with N = 1,000, $\mu = 5\times10^{-8}$, and $s = -0.01$. For each parameter combination, 10,000 fusion-mimicking mutations were simulated in a population harboring a pair of autosomes and a pair of XY-like nonrecombining sex chromosomes (females being XX and males XY) with a 1-Mb recombining region on each sex-chromosome side (mimicking pseudoautosomal regions). Only fusions not lost after 20 generations are displayed. The displayed frequency of the fused chromosomes is their frequency among the chromosomes of the same type, e.g., a Y-autosome fusion at 1.0 frequency means that all Y chromosomes are fused with an autosome. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S19 Fig. Fraction of Y-linked inversions spreading or fixed after 10,000 generations depending on the mean number of mutations segregating in the region where they appear. For each parameter combination, 10,000 inversions fully linked to a Y sex-determining allele (permanently heterozygous) were simulated. The number of mutations displayed here represents the mean number of mutations in haplotypes of the whole population in the region covered by inversions and not the number of mutations in the inversions themselves. Inversions have higher chances of spreading when they appear in mutation-dense regions. Populations of 1,000 individuals were simulated. The X axis is log-scaled. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S20 Fig. Fraction of 1-Mb less-loaded inversions that went to fixation depending on the heterozygous fitness cost they are associated with. The heterozygous fitness cost is defined in such a way that $W_{NI}^{\eta} = W_{NI}(1-\eta)$, with W_{NI}^{η} the fitness of individual heterozygotes for the inversion with the cost (see S1 Appendix, section 10). Simulations were performed with N = 1,000, $\mu = 5\times10^{-9}$, and $s = -0.01$. A Y-linked inversion was considered fixed if it was present on all Y chromosomes. Y-linked inversions become fixed (equilibrium frequency = 1) when $m<nq-\ln(1-\eta)/(-hs)$ and are lost (equilibrium frequency = 0) when $m > nq - \frac{\ln(1-\eta)}{-hs}$. Autosomal inversions become fixed (equilibrium frequency = 1) when $m<nqh/(1-h)+\ln(1-\eta)/s(h-1)$ and are lost (equilibrium frequency = 0) when $m>nq-\ln(1-\eta)/(-hs)$ (see S1

Appendix, section 10). Parameters resulting in nq = 1,2,3,...,6 expected mutations are highlighted by red lines (top line, nq = 1; bottom line, nq = 6). The script used to produce this figure is available on GitHub.
(PDF)

S21 Fig. Fraction of 5 Mb less-loaded inversions that went to fixation depending on the heterozygous fitness cost they are associated with. The heterozygous fitness cost is defined in such a way that $W_{NI}^{\eta} = W_{NI}(1-\eta)$, with W_{NI}^{η} the fitness of individual heterozygotes for the inversion with the cost (see S1 Appendix, section 10). Simulations were performed with N = 1,000, $\mu = 5\times10^{-9}$, and $s = -0.01$. A Y-linked inversion was considered fixed if it was present on all Y chromosomes. Y-linked inversions become fixed (equilibrium frequency = 1) when $m<nq-\ln(1-\eta)/(-hs)$ and are lost (equilibrium frequency = 0) when $m > nq - \frac{\ln(1-\eta)}{-hs}$. Autosomal inversions become fixed (equilibrium frequency = 1) when $m<nqh/(1-h)+\ln(1-\eta)/s(h-1)$ and are lost (equilibrium frequency = 0) when $m>nq-\ln(1-\eta)/(-hs)$ (see S1 Appendix, section 10). Parameters resulting in nq = 3, 4, 5, or 6 expected mutations are highlighted by red lines (top line, nq = 3; bottom line, nq = 6). The script used to produce this figure is available on GitHub.
(PDF)

S22 Fig. Proportions of inversions in stochastic simulations that were fixed after 10,000 generations for different parameter value combinations with N = 1,000 and in populations with haplodiplontic life cycles. Similar to Fig 3C but considering species with different life cycles. Four life cycles were considered: fully diploid (as used throughout this study) and haplodiplontic with occurrence of a haploid phase every x generation, with x being 2, 3, or 10. Simulations were performed with N = 1,000, $\mu = 5\times10^{-9}$, and $s = -0.01$. This shows that inversions are less likely to spread and fix in species with extended haploid phases, as such life cycles lead to more efficient purge of deleterious recessive mutations. See S23 Fig for the variation in mutation load in populations with different life cycles. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S23 Fig. Mutation load evolution in populations with different life cycles during the burn-in phase. For the simulations of the effect of different life cycle on the fate of inversions (S22 Fig), for each set of parameter values, one simulation was run and used as the initial state after a burn-in period of 15,000 generations (see Methods). We display here the evolution of the mutation load during the burn-in phase for populations with different life cycles. Four haplodiplontic life cycles were considered, with a haploid phase every x generation, with x being 2, 3, 10, or 100. Each trajectory represents 1 simulation. (**a**) Total number of mutations segregating across the genome in the population. (**b**) Mean number of segregating mutations in the genome in diploid individuals. (**c**) Mean frequency of mutations in the population. This shows that populations with an extended or frequent haploid phase tend to have a reduced mutation load compared to populations with rare or brief haploid phases. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S24 Fig. Successive accumulation of inversions around a male-determining allele in an XY system, leading to the formation of nonrecombining sex chromosomes. Similar to Fig 4 but displaying the result of another simulation and with a population of N = 10,000 individuals. A simulation of N = 10,000 individuals, each with 2 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, 1 inversion appears on average in the whole population, in an individual sampled uniformly at random, with the 2 recombination breakpoints sampled uniformly at random among $k = 100$ potential breakpoints. (**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 in gray. The colors are not entirely opaque, so that regions with overlapping inversions appear darker. Loss of previously fixed inversions are due to beneficial reversion occurrence and selection. (**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 1-Mb windows) are recorded at the formation of each offspring, across all homologous chromosomes in the population. To illustrate the evolution of recombination suppression between the sex chromosomes, only recombination events between the X and the Y chromosomes are shown for chromosome 1 (i.e., not the recombination events between the 2 X chromosomes in females). Unlike chromosome 1, chromosome 2 harbors no permanently heterozygous allele. All inversions on this chromosome suffer from homozygosity disadvantage and very few inversions therefore become fixed on chromosome 2. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S25 Fig. Population evolution during the burn-in phase. For the simulations of the effect of different parameters (notably h and s) on the fate of inversions (Figs 3 and S10–S13, notably), for each set of parameter values, 1 simulation was run and used as the initial state after a burn-in period of 15,000 generations (see Methods). We display here the evolution of the population during the burn-in phase. Each trajectory represents the state of one simulation. (**a**) Number of mutations segregating in the genome. (**b**) Mean frequency of segregating mutations in the genome. This shows that each population had reached an equilibrium state at the end of the burn-in period. The dataset and script used to produce this figure are available on Figshare (doi: 10.6084/m9.figshare.19961033) and GitHub.
(PDF)

S1 Appendix. Supplementary methods.
(PDF)

Acknowledgments

We thank Ricardo Rodriguez de la Vega, Fanny Hartmann, Jacqui Shykoff, Sylvain Billiard, Janis Antonovics, Laurent Keller, and Olivier Tenaillon for comments on a draft version of the manuscript. ET thanks Denis Roze for insightful discussions. ET and AV acknowledge support from the chaire program « Mathematical modeling and biodiversity » (Ecole Polytechnique, Museum National d'Histoire Naturelle, Veolia Environnement, Fondation X). PJ thanks the authors of SLiM for their outstanding software and manual.

Author Contributions

Conceptualization: Paul Jay, Tatiana Giraud.

Formal analysis: Paul Jay, Emilie Tezenas.

Investigation: Paul Jay, Emilie Tezenas.

Methodology: Paul Jay.

Supervision: Paul Jay, Amandine Véber, Tatiana Giraud.

Validation: Paul Jay, Amandine Véber, Tatiana Giraud.

Visualization: Paul Jay.

Writing – original draft: Paul Jay, Tatiana Giraud.

Writing – review & editing: Paul Jay, Emilie Tezenas, Amandine Véber, Tatiana Giraud.

References

1. Abbott JK, Nordén AK, Hansson B. Sex chromosome evolution: historical insights and future perspectives. Proc Biol Sci. 2017; 284. https://doi.org/10.1098/rspb.2016.2806 PMID: 28469017
2. Charlesworth D. Plant contributions to our understanding of sex chromosome evolution. New Phytol. 2015; 208:52–65. https://doi.org/10.1111/nph.13497 PMID: 26053356
3. Schwander T, Libbrecht R, Keller L. Supergenes and complex phenotypes. Curr Biol. 2014; 24:R288–R294. https://doi.org/10.1016/j.cub.2014.01.056 PMID: 24698381
4. Bergero R, Charlesworth D. The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol. 2009; 24:94–102. https://doi.org/10.1016/j.tree.2008.09.010 PMID: 19100654
5. Hartmann FE, Duhamel M, Carpentier F, Hood ME, Foulongne-Oriol M, Silar P, et al. Recombination suppression and evolutionary strata around mating-type loci in fungi: documenting patterns and understanding evolutionary and mechanistic causes. New Phytologist. 2021; 229:2470–2491. https://doi.org/10.1111/nph.17039 PMID: 33113229
6. Ponnikas S, Sigeman H, Abbott JK, Hansson B. Why do sex chromosomes stop recombining? Trends Genet. 2018; 34:492–503. https://doi.org/10.1016/j.tig.2018.04.001 PMID: 29716744
7. Wright AE, Dean R, Zimmer F, Mank JE. How to make a sex chromosome. Nat Commun. 2016; 7:12087. https://doi.org/10.1038/ncomms12087 PMID: 27373494
8. Ruzicka F, Dutoit L, Czuppon P, Jordan CY, Li X-Y, Olito C, et al. The search for sexually antagonistic genes: Practical insights from studies of local adaptation and statistical genomics. Evol Lett. 2020; 4:398–415. https://doi.org/10.1002/evl3.192 PMID: 33014417
9. Rice WR. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution. 1987; 41:911–914. https://doi.org/10.1111/j.1558-5646.1987.tb05864.x PMID: 28564364
10. Cavoto E, Neuenschwander S, Goudet J, Perrin N. Sex-antagonistic genes, XY recombination and feminized Y chromosomes. J Evol Biol. 2018; 31:416–427. https://doi.org/10.1111/jeb.13235 PMID: 29284187
11. Ironside JE. No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. BioEssays. 2010; 32:718–726. https://doi.org/10.1002/bies.200900124 PMID: 20658710
12. Beukeboom L, Perrin N. The Evolution of Sex Determination. Oxford, New York: Oxford University Press; 2014.
13. Jay P, Chouteau M, Whibley A, Bastide H, Parrinello H, Llaurens V, et al. Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. Nat. Genet. 2021; 53:288–293. https://doi.org/10.1038/s41588-020-00771-1 PMID: 33495598
14. Yan Z, Martin SH, Gotzek D, Arsenault SV, Duchen P, Helleu Q, et al. Evolution of a supergene that regulates a trans-species social polymorphism. Nat Ecol Evol. 2020; 4:240–249. https://doi.org/10.1038/s41559-019-1081-1 PMID: 31959939
15. Bazzicalupo AL, Carpentier F, Otto SP, Giraud T. Little evidence of antagonistic selection in the evolutionary strata of fungal mating-type chromosomes (*Microbotryum lychnidis-dioicae*). G3. 2019; 9:1987–1998. https://doi.org/10.1534/g3.119.400242 PMID: 31015196
16. Branco S, Badouin H, de la Vega RCR, Gouzy J, Carpentier F, Aguileta G, et al. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. Proc Natl Acad Sci U S A. 2017; 114:7067–7072. https://doi.org/10.1073/pnas.1701658114 PMID: 28630332
17. Úbeda F, Patten MM, Wild G. On the origin of sex chromosomes from meiotic drive. Proc Biol Sci. 2015; 282:20141932. https://doi.org/10.1098/rspb.2014.1932 PMID: 25392470
18. Charlesworth B, Coyne JA, Barton NH. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 1987; 130:113–146.

19. Antonovics J, Abrams JY. Intratetrad mating and the evolution of linkage relationships. Evolution. 2004; 58:702–709. https://doi.org/10.1111/j.0014-3820.2004.tb00403.x PMID: 15154546

20. Charlesworth B, Wall JD. Inbreeding, heterozygote advantage and the evolution of neo–X and neo–Y sex chromosomes. Proc Biol Sci. 1999; 266:51–56.

21. Jeffries DL, Gerchen JF, Scharmann M, Pannell JR. A neutral model for the loss of recombination on sex chromosomes. Philos Trans R Soc Lond B Biol Sci. 2021; 376:20200096. https://doi.org/10.1098/rstb.2020.0096 PMID: 34247504

22. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat. Rev. Genet. 2007; 8:610–618. https://doi.org/10.1038/nrg2146 PMID: 17637733

23. Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009; 10:195–205.

24. Charlesworth B, Charlesworth D, Morgan MT. Genetic loads and estimates of mutation rates in highly inbred plant populations. Nature. 1990; 347:380–382.

25. Deng HW, Lynch M. Inbreeding depression and inferred deleterious-mutation parameters in Daphnia. Genetics. 1997; 147:147–155. https://doi.org/10.1093/genetics/147.1.147 PMID: 9286675

26. Charlesworth D, Charlesworth B. Inbreeding depression and its evolutionary consequences. Annu Rev Ecol Syst. 1987; 18:237–268.

27. Bertorelle G, Raffini F, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, et al. Genetic load: genomic estimates and applications in non-model animals. Nat Rev Gen. 2022;1–12. https://doi.org/10.1038/s41576-022-00448-x PMID: 35136196

28. Nei M, Kojima K-I, Schaffer HE. Frequency changes of new inversions in populations under mutation-selection equilibria. Genetics. 1967; 57:741–750. https://doi.org/10.1093/genetics/57.4.741 PMID: 6082619

29. Ohta T. Associative overdominance caused by linked detrimental mutations. Genet Res. 1971; 18:277–286. PMID: 5158298

30. Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet. 2013; 14:113–124. https://doi.org/10.1038/nrg3366 PMID: 23329112

31. Nei M. Accumulation of nonfunctional genes on sheltered chromosomes. Am Nat. 1970; 104:311–322.

32. Blaser O, Neuenschwander S, Perrin N, Gardner AEA, Day ET. Sex-chromosome turnovers: the hot-potato model. Am Nat. 2014; 183:140–146. https://doi.org/10.1086/674026 PMID: 24334743

33. Grossen C, Neuenschwander S, Perrin N. The evolution of XY recombination: sexually antagonistic selection versus deleterious mutation load. Evolution. 2012; 66:3155–3166. https://doi.org/10.1111/j.1558-5646.2012.01661.x PMID: 23025605

34. Pennell MW, Kirkpatrick M, Otto SP, Vamosi JC, Peichel CL, Valenzuela N, et al. Y Fuse? Sex Chromosome Fusions in Fishes and Reptiles. PLoS Genet. 2015; 11:e1005237. https://doi.org/10.1371/journal.pgen.1005237 PMID: 25993542

35. Ferchaud A-L, Mérot C, Normandeau E, Ragoussis J, Babin C, Djambazian H, et al. Chromosome-level assembly reveals a putative Y-autosomal fusion in the sex determination system of the Greenland Halibut (*Reinhardtius hippoglossoides*). G3. 2022; 12:jkab376. https://doi.org/10.1093/g3journal/jkab376 PMID: 34791178

36. Yashiro T, Tea Y-K, Van Der Wal C, Nozaki T, Mizumoto N, Hellemans S, et al. Enhanced heterozygosity from male meiotic chromosome chains is superseded by hybrid female asexuality in termites. Proc Natl Acad Sci U S A. 2021; 118:e2009533118. https://doi.org/10.1073/pnas.2009533118 PMID: 34903643

37. Connallon T, Olito C. Natural selection and the distribution of chromosomal inversion lengths. Mol Ecol. 2021. https://doi.org/10.1111/mec.16091 PMID: 34297880

38. Olito C, Abbott JK. The evolution of suppressed recombination between sex chromosomes by chromosomal inversions. bioRxiv. 2020. https://doi.org/10.1101/2020.03.23.003558

39. Olito C, Ponnikas S, Hansson B, Abbott JK. Consequences of partially recessive deleterious genetic variation for the evolution of inversions suppressing recombination between sex chromosomes. Evolution. 2022 Jun; 76(6):1320–1330. https://doi.org/10.1111/evo.14496 PMID: 35482933

40. Neher RA, Shraiman BI. Fluctuations of fitness distributions and the rate of Muller's ratchet. Genetics. 2012; 191:1283–1293. https://doi.org/10.1534/genetics.112.141325 PMID: 22649084

41. Glémin S. How are deleterious mutations purged? Drift versus nonrandom mating. Evolution. 2003; 57:2678–2687. https://doi.org/10.1111/j.0014-3820.2003.tb01512.x PMID: 14761049

42. Takayama S, Isogai A. Self-incompatibility in plants. Annu Rev Plant Biol. 2005; 56:467–489. https://doi.org/10.1146/annurev.arplant.56.032604.144249 PMID: 15862104

43. Küpper C, Stocks M, Risse JE, dos Remedios N, Farrell LL, McRae SB, et al. A supergene determines highly divergent male reproductive morphs in the ruff. Nat Genet. 2016; 48:79–83. https://doi.org/10.1038/ng.3443 PMID: 26569125

44. Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, Shoemaker D, et al. A Y-like social chromosome causes alternative colony organization in fire ants. Nature. 2013; 493:664–668. https://doi.org/10.1038/nature11832 PMID: 23334415

45. Boideau F, Richard G, Coriton O, Huteau V, Belser C, Deniot G, et al. Epigenomic and structural events preclude recombination in *Brassica napus*. New Phytologist. 2021. https://doi.org/10.1111/nph.18004 PMID: 35092024

46. Stauber L, Badet T, Feurtey A, Prospero S, Croll D. Emergence and diversification of a highly invasive chestnut pathogen lineage across southeastern Europe. Elife. 2021; 10:e56279. https://doi.org/10.7554/eLife.56279 PMID: 33666552

47. Kubisiak TL, Milgroom MG. 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. 2006; 43:453–463. https://doi.org/10.1016/j.fgb.2006.02.002 PMID: 16554177

48. Kimura M, Ohta T. The average number of generations until fixation of a mutant gene in a finite population. Genetics. 1969; 61:763–771. https://doi.org/10.1093/genetics/61.3.763 PMID: 17248440

49. Scott MF, Rescan M. Evolution of haploid–diploid life cycles when haploid and diploid fitnesses are not equal. Evolution. 2017; 71:215–226. https://doi.org/10.1111/evo.13125 PMID: 27859032

50. Kondrashov AS, Crow JF. Haploidy or diploidy: which is better? Nature. 1991; 351:314–315. https://doi.org/10.1038/351314a0 PMID: 2034273

51. Porubsky D, Höps W, Ashraf H, Hsieh P, Rodriguez-Martin B, Yilmaz F, et al. Haplotype-resolved inversion landscape reveals hotspots of mutational recurrence associated with genomic disorders. bioRxiv. 2021. https://doi.org/10.1101/2021.12.20.472354

52. Lenormand T, Roze D. Y recombination arrest and degeneration in the absence of sexual dimorphism. Science. 2022; 375:663–666. https://doi.org/10.1126/science.abj1813 PMID: 35143289

53. Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics. 2006; 173:419–434. https://doi.org/10.1534/genetics.105.047985 PMID: 16204214

54. Connallon T, Olito C, Dutoit L, Papoli H, Ruzicka F, Yong L. Local adaptation and the evolution of inversions on sex chromosomes and autosomes. Philos Trans R Soc Lond B Biol Sci. 2018; 373:20170423. https://doi.org/10.1098/rstb.2017.0423 PMID: 30150221

55. Osada N. Genetic diversity in humans and non-human primates and its evolutionary consequences. Genes Genet Syst. 2015; 90:133–145. https://doi.org/10.1266/ggs.90.133 PMID: 26510568

56. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. Nat Rev Genet. 2015; 16:333–343. https://doi.org/10.1038/nrg3931 PMID: 25963372

57. Giner-Delgado C, Villatoro S, Lerga-Jaso J, Gayà-Vidal M, Oliva M, Castellano D, et al. Evolutionary and functional impact of common polymorphic inversions in the human genome. Nat Commun. 2019; 10:4222. https://doi.org/10.1038/s41467-019-12173-x PMID: 31530810

58. Wellenreuther M, Bernatchez L. Eco-evolutionary genomics of chromosomal inversions. Trends Ecol Evol. 2018; 33:427–440. https://doi.org/10.1016/j.tree.2018.04.002 PMID: 29731154

59. Cui L, Neoh H, Iwamoto A, Hiramatsu K. Coordinated phenotype switching with large-scale chromosome flip-flop inversion observed in bacteria. Proc Natl Acad Sci U S A. 2012; 109:E1647–E1656. https://doi.org/10.1073/pnas.1204307109 PMID: 22645353

60. Hanson SJ, Byrne KP, Wolfe KH. Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus *Saccharomyces cerevisiae* system. Proc Natl Acad Sci U S A. 2014; 111:E4851–E4858. https://doi.org/10.1073/pnas.1416014111 PMID: 25349420

61. Badouin H, Hood ME, Gouzy J, Aguileta G, Siguenza S, Perlin MH, et al. Chaos of rearrangements in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. Genetics. 2015; 200:1275–1284. https://doi.org/10.1534/genetics.115.177709 PMID: 26044594

62. Carey SB, Jenkins J, Lovell JT, Maumus F, Sreedasyam A, Payton AC, et al. Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. Sci Adv. 2021; 7:eabh2488. https://doi.org/10.1126/sciadv.abh2488 PMID: 34193417

63. Skinner BM, Sargent CA, Churcher C, Hunt T, Herrero J, Loveland JE, et al. The pig X and Y Chromosomes: structure, sequence, and evolution. Genome Res. 2016; 26(1):130–139. https://doi.org/10.1101/gr.188839.114 PMID: 26560630

64. Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature. 2014; 508:494–499. https://doi.org/10.1038/nature13206 PMID: 24759411

65. Lemaitre C, Braga MDV, Gautier C, Sagot M-F, Tannier E, Marais GAB. Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes. Genome Biol Evol. 2009; 1:56–66. https://doi.org/10.1093/gbe/evp006 PMID: 20333177

66. Veve AL, Burghgraeve N, Genete M, Lepers-Blassiau C, Takou M, Meaux JD, et al. Long-term balancing selection and the genetic load linked to the self-incompatibility locus in *Arabidopsis halleri* and *A. lyrata*. bioRxiv. 2022. https://doi.org/10.1101/2022.04.12.487987

67. Llaurens V, Billiard S, Castric V, Vekemans X. Evolution of Dominance in Sporophytic Self-Incompatibility Systems: I. Genetic Load and Coevolution of Levels of Dominance in Pollen and Pistil. Evolution. 2009; 63:2427–2437. https://doi.org/10.1111/j.1558-5646.2009.00709.x PMID: 19473398

68. Kirkpatrick M. How and why chromosome inversions evolve. PLoS Biol. 2010; 8:e1000501. https://doi.org/10.1371/journal.pbio.1000501 PMID: 20927412

69. Hämälä T, Wafula EK, Guiltinan MJ, Ralph PE, dePamphilis CW, Tiffin P. Genomic structural variants constrain and facilitate adaptation in natural populations of Theobroma cacao, the chocolate tree. Proc Natl Acad Sci U S A. 2021; 118:e2102914118. https://doi.org/10.1073/pnas.2102914118 PMID: 34408075

70. Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, Beridze T, et al. The population genetics of structural variants in grapevine domestication. Nat Plants. 2019; 5:965–979. https://doi.org/10.1038/s41477-019-0507-8 PMID: 31506640

71. Todesco M, Owens GL, Bercovich N, Légaré J-S, Soudi S, Burge DO, et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature. 2020; 584:602–607. https://doi.org/10.1038/s41586-020-2467-6 PMID: 32641831

72. Coelho S, Gueno J, Lipinska A, Cock J, Umen J. UV chromosomes and haploid sexual systems. Trends Plant Sci. 2018; 23:794–807. https://doi.org/10.1016/j.tplants.2018.06.005 PMID: 30007571

73. Filatov DA. Homomorphic plant sex chromosomes are coming of age. Mol Ecol. 2015; 24:3217–3219. https://doi.org/10.1111/mec.13268 PMID: 26113024

74. McGuire IC, Marra RE, Milgroom MG. Mating-type heterokaryosis and selfing in *Cryphonectria parasitica*. Fungal Genet Biol. 2004; 41:521–533. https://doi.org/10.1016/j.fgb.2003.12.007 PMID: 15050541

75. McGuire IC, Davis JE, Double ML, MacDonald WL, Rauscher JT, McCawley S, et al. Heterokaryon formation and parasexual recombination between vegetatively incompatible lineages in a population of the chestnut blight fungus, *Cryphonectria parasitica*. Mol Ecol. 2005; 14:3657–3669. https://doi.org/10.1111/j.1365-294X.2005.02693.x PMID: 16202087

76. Dyer KA, Charlesworth B, Jaenike J. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. Proc Natl Acad Sci U S A. 2007; 104:1587–1592. https://doi.org/10.1073/pnas.0605578104 PMID: 17242362

77. Reinhardt JA, Brand CL, Paczolt KA, Johns PM, Baker RH, Wilkinson GS. Meiotic Drive Impacts Expression and Evolution of X-Linked Genes in Stalk-Eyed Flies. PLoS Genet. 2014; 10:e1004362. https://doi.org/10.1371/journal.pgen.1004362 PMID: 24832132

78. Chasnov JR. Mutation-selection balance, dominance and the maintenance of sex. Genetics. 2000; 156:1419–1425. https://doi.org/10.1093/genetics/156.3.1419 PMID: 11063713

79. Agrawal A, Whitlock M. Mutation load: the fitness of individuals in populations where deleterious alleles are abundant. Annu Rev Ecol Evol Syst. 2012; 43:115–135.

80. Haller BC, Messer PW. SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. Mol Biol Evol. 2019; 36:632–637. https://doi.org/10.1093/molbev/msy228 PMID: 30517680

81. Kim BY, Huber CD, Lohmueller KE. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics. 2017; 206:345–361. https://doi.org/10.1534/genetics.116.197145 PMID: 28249985

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

Appendix : Supplementary methods for "**Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes** "

Paul Jay, Emilie Tezenas, Amandine Véber, Tatiana Giraud

Correspondence to: paul.yann.jay@gmail.com

1 Prerequisite

We consider an infinite size, randomly mating population of diploid individuals with discrete, nonoverlapping generations. Each individual is represented by a single pair of chromosomes which carries a locus under balancing selection that may either represent a mating-type locus, a sex-determining locus or any overdominant locus, and n other sites on which deleterious mutations segregate. Mutations have multiplicative effect, reducing fitness by $1-hs$ when heterozygous, and by $1-s$ when homozygous, s and h being the selection and dominance coefficients. For example, if a diploid individual has m_1 heterozygous sites and m_2 homozygous deleterious sites, with $m_1+m_2 \leq n$, its fitness is

$$W = (1-hs)^{m_1}(1-s)^{m_2}.$$

Selection acts on individuals based on their genotype and individuals contribute to the next generation depending on their relative fitness. In each generation, diploid individuals are formed by randomly sampling two gametes (haploid) produced by individuals from the previous generation.

2 Evolution of inversions (or any recombination suppressor acting in *cis*)

Throughout this document, we will talk about inversions but the rationale applies equally to any recombination suppressor acting in *cis* (i.e. suppressing recombination in a segment in which it resides). At a given time, an inversion appears in the population, on one chromosome. This inversion captures a segment with a number of deleterious mutations that we denote by m.

We are interested in studying the fate of this inversion in the population depending on m, s, h and its linkage with the locus with permanently or nearly-permanently heterozygous allele(s).
Following [1], we denote the presence of an inversion on a chromosome by I, its absence (ancestral state) by N.
The fitness of an individual depends on the number of mutations it carries and therefore on its genotype for the inversion.

All mutations are considered to have the same coefficient of selection and the same coefficient of dominance and the n sites are supposed to be independent from one another. At mutation-selection balance, all mutations are therefore at the same frequency, denoted q. Hence, conditionally on an individual sequence carrying a mutation at a given site, the probability that it is paired with a sequence carrying the same mutation is q (the individual being therefore homozygous for this mutation), and the probability that it is not is $1-q$ (the individual being therefore heterozygous). Consider a single locus i with two alleles at mutation-selection balance. The average population fitness taking only this locus into account is:

$$\begin{aligned} \overline{W_i} &= (1-q)^2 + 2q(1-q)(1-hs) + q^2(1-s) \\ &= 1 - 2q(1-q)hs - q^2 s. \end{aligned} \tag{1}$$

Since the n sites are assumed independent of one another, the average fitness of an individual considering n loci is:

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

which is the same as the average fitness of a diploid individual without inversion :

$$W_{NN} = (1 - 2q(1-q)hs - q^2 s)^n. \tag{3}$$

Considering small h and small s, this can be approximated by

$$W_{NN} \approx e^{-nqsh - nqs(h+q(1-2h))}. \tag{4}$$

We can express the average fitness of an individual heterozygous for the inversion by:

$$W_{NI} = (q(1-s) + (1-q)(1-hs))^m (q(1-hs) + 1 - q)^{n-m}, \tag{5}$$

which can again be approximated for small h and s by:

$$W_{NI} \approx e^{-nqsh - ms(h+q(1-2h))}. \tag{6}$$

Indeed, there are $n - m$ sites where the inversion have the ancestral allele. Each of these sites have a probability q of being heterozygous and $(1 - q)$ of being homozygous for the ancestral allele.

Individuals homozygous for the inversion are homozygous for all mutations carried by this inversion. They have all the same fitness, which can therefore be expressed by

$$W_{II} = (1 - s)^m, \tag{7}$$

and approximated by

$$W_{II} \approx e^{-sm} \tag{8}$$

when s is small.

3 Inversion evolutionary trajectory

We considered multiple scenarios to study the fate of an inversion depending on its linkage with an allele following various heterozygosity rule. We tracked the frequency of the inversion, considering that inversions recombine at a rate r with a locus with permanently or nearly-permanently heterozygous allele(s). When $r = 0.5$, the inversion is completely unlinked with this locus and behaves as if it were on an autosome. We considered four situations:

1. The locus under balancing selection has two alleles, X and Y, and is overdominant such that homozygotes XX and YY have their fitness reduced by a factor $(1 - s_2)$.

2. The locus under balancing selection has two alleles, X and Y, and individuals are permanently heterozygotes.

3. The locus under balancing selection has three alleles, X, Y and Z, and individuals are permanently heterozygotes.

4. The locus under balancing selection has two alleles, X and Y, and individuals can be either XX or XY and must mate with a different genotype.

For each situation, we provide expressions for the evolution of haplotype frequencies. In each generation, genotypes frequencies are computed assuming random mating among the haplotypes. Each genotype then contributes to the next generation of haplotypes relatively to its fitness. The equations for the haplotypes frequencies are used to illustrate the deterministic change in inversion frequencies in the figures displayed in the main text and in supplementary material.

4 Trajectory of inversions appearing on an autosome

4.1 Allelic structure

We first consider an inversion that appears on an autosome, the easiest case. The possible haplotypes are :

$$\begin{array}{rl} Haplotype & Frequency \\ N & F_N \\ I & F_I = 1 - F_N \end{array} \tag{9}$$

Then, there are three possible genotypes, which frequencies are computed assuming random mating among haplotypes.

Genotype	Frequency	Fitness
II	F_I^2	W_{II}
IN	$2F_IF_N$	W_{NI}
NN	F_N^2	W_{NN}

The mean fitness of the population can then be computed from the genotypes frequencies as follows :

$$\overline{W} = F_N^2 W_{NN} + 2F_N F_I W_{NI} + F_I^2 W_{II}.$$

4.2 Evolution of frequencies

Considering that each genotype contributes relatively to its fitness to the next generation, we compute the evolution of the frequency of the inverted haplotype after one generation :

$$F_I^{t+1} = \frac{(F_I^t)^2 W_{II} + F_I^t F_N^t W_{NI}}{\overline{W^t}}.$$

Substituing $F_N = 1 - F_I$, we obtain the variation of F_I over one generation, at time t :

$$\Delta F_I^t = \frac{F_I^t}{\overline{W^t}}\Big(\left(F_I^t\right)^2 (2W_{NI} - W_{II} - W_{NN}) + F_I^t \left(W_{II} + 2W_{NN} - 3W_{NI}\right) + W_{NI} - W_{NN}\Big).$$

This equation is used in Methods to derive the equilibrium frequency of an inversion that appears on an autosome.

5 Trajectory of inversion appearing in linkage with an overdominant locus (S8 Fig)

5.1 Allelic structure

At a given locus, we consider that individuals can be either homozygotes XX or YY, or heterozygotes XY, but that homozygotes suffer from a lowered fitness compared to heterozygotes (overdominance). We define s_2 as the selection coefficient such that homozygotes have their fitness decreased by a factor $(1 - s_2)$. For example, $W_{XI/XI} = (1 - s_2)W_{II}$. This locus being under balancing selection because of overdominance, it is hereafter referred as the "OD" locus.

Considering the evolution of an inversion I in linkage with this "OD" locus, there are four different haplotypes:

$$\begin{array}{rl} Haplotype & Frequency \\ XN & F_{XN} \\ XI & F_{XI} \\ YN & F_{YN} \\ YI & F_{YI} \end{array} \tag{10}$$

The possible genotypes and their corresponding frequencies and fitnesses are given in the following table.

OD genotype	Inversion genotype	Frequency	Fitness
XX	II	$F_{XI}F_{XI}$	$(1-s_2)W_{II}$
	IN	$2F_{XI}F_{XN}$	$(1-s_2)W_{NI}$
	NN	$F_{XN}F_{XN}$	$(1-s_2)W_{NN}$
XY	II	$2F_{XI}F_{YI}$	W_{II}
	IN	$2F_{XI}F_{YN}$	W_{NI}
	NI	$2F_{XN}F_{YI}$	W_{NI}
	NN	$2F_{XN}F_{YN}$	W_{NN}
YY	II	$F_{YI}F_{YI}$	$(1-s_2)W_{II}$
	IN	$2F_{YI}F_{YN}$	$(1-s_2)W_{NI}$
	NN	$F_{YN}F_{YN}$	$(1-s_2)W_{NN}$

The frequencies are computed assuming random mating among haplotypes, and considering that $F_{XI}+F_{XN}+F_{YI}+F_{YN}=1$.

The mean fitness of the population in a given generation can then be computed from the genotypes frequencies as follows :

$$\begin{aligned}\overline{W} = {} & W_{II}\Big((1-s_2)\big({F_{XI}}^2+{F_{YI}}^2\big)+2F_{XI}F_{YI}\Big)\\ & + W_{NN}\Big((1-s_2)\big({F_{XN}}^2+{F_{YN}}^2\big)+2F_{XN}F_{YN}\Big)\\ & + 2W_{NI}\Big((1-s_2)\big(F_{XI}F_{XN}+F_{YI}F_{YN}\big)+F_{XI}F_{YN}+F_{XN}F_{YI}\Big).\end{aligned} \tag{11}$$

5.2 Evolution of frequencies

Considering a recombination rate r between the overdominant locus and the inversion locus, and a generation t, we compute the frequency of XI in the next generation :

$$\begin{aligned}F_{XI}^{t+1} = {} & {F_{XI}^t}^2\frac{W_{II}(1-s_2)}{\overline{W^t}}+F_{XI}^tF_{XN}^t\frac{W_{NI}(1-s_2)}{\overline{W^t}}+F_{XI}^tF_{YI}^t\frac{W_{II}}{\overline{W^t}}\\ & + F_{XI}^tF_{YN}^t\frac{W_{NI}}{\overline{W^t}}(1-r)+F_{YI}^tF_{XN}^t\frac{W_{NI}}{\overline{W^t}}r.\end{aligned} \tag{12}$$

Re-arranging the terms, we arrive at :

$$F_{XI}^{t+1} = \frac{F_{XI}^t\Big(W_{II}\Big[(1-s_2)F_{XI}^t+F_{YI}^t\Big]+W_{NI}\Big[(1-s_2)F_{XN}^t+F_{YN}^t\Big]\Big)}{\overline{W^t}}-\frac{rW_{NI}D^t}{\overline{W^t}}, \tag{13}$$

with $D^t=F_{XI}^tF_{YN}^t-F_{XN}^tF_{YI}^t$, the linkage desequilibrium.

We can then compute the variation of F_{XI} over one generation :

$$\begin{aligned}\Delta F_{XI}^t &= F_{XI}^{t+1}-F_{XI}^t\\ &= F_{XI}^{t+1}-F_{XI}^t\frac{\overline{W^t}}{\overline{W^t}}\\ &= F_{XI}^tW_{II}\frac{F_{YI}^t(1-2F_{XI}^t)+(1-s_2)\Big(F_{XI}^t(1-F_{XI}^t)-(F_{YI}^t)^2\Big)}{\overline{W^t}}\\ &\quad + F_{XI}^tW_{NI}\frac{(1-s_2)\Big(F_{XN}^t(1-2F_{XI}^t)+2F_{YI}^tF_{YN}^t\Big)+F_{YN}^t(1-2F_{XI}^t)-2F_{XN}^tF_{YI}^t}{\overline{W^t}}\\ &\quad - F_{XI}^tW_{NN}^t\frac{(1-s_2)\Big((F_{XN}^t)^2+(F_{YN}^t)^2\Big)+2F_{XN}^tF_{YN}^t}{\overline{W^t}}\\ &\quad - \frac{rW_{NI}D^t}{\overline{W^t}}.\end{aligned} \tag{14}$$

To compute the same quantities for F_{YI}, we can interchange the roles of X and Y in the expression for F_{XI}, since X and Y have symmetrical roles:

$$F_{YI}^{t+1} = \frac{F_{YI}^t\Big(W_{II}\Big[(1-s_2)F_{YI}^t + F_{XI}^t\Big] + W_{NI}\Big[(1-s_2)F_{YN}^t + F_{XN}^t\Big]\Big)}{\overline{W^t}} + \frac{rW_{NI}D^t}{\overline{W^t}}. \quad (15)$$

From those expressions, we derive F_{XN}^{t+1} by interverting I and N in the expression for F_{XI}^{t+1}, and F_{YN}^{t+1} by interverting X and Y in the expression for F_{XN}^{t+1} :

$$\begin{aligned} F_{XN}^{t+1} &= \frac{F_{XN}^t\Big(W_{NN}\Big[(1-s_2)F_{XN}^t + F_{YN}^t\Big] + W_{NI}\Big[(1-s_2)F_{XI}^t + F_{YI}^t\Big]\Big)}{\overline{W^t}} + \frac{rW_{NI}D^t}{\overline{W^t}}, \\ F_{YN}^{t+1} &= \frac{F_{YN}^t\Big(W_{NN}\Big[(1-s_2)F_{YN}^t + F_{XN}^t\Big] + W_{NI}\Big[(1-s_2)F_{YI}^t + F_{XI}^t\Big]\Big)}{\overline{W^t}} - \frac{rW_{NI}D^t}{\overline{W^t}}. \end{aligned} \quad (16)$$

S8 Fig shows a numerical solving of these equations.

6 Trajectory of inversions in linkage with a permanently heterozygous locus with two alleles (S5-6 Figs)

6.1 Allelic structure

At a given locus with two alleles, X and Y, we suppose that individuals are all heterozygotes XY.

Considering an inversion I segregating in this population, there are still the same four possible haplotypes in the population:

$$\begin{array}{rl} Haplotype & Frequency \\ XN & F_{XN} \\ XI & F_{XI} \\ YN & F_{YN} \\ YI & F_{YI} \end{array} \quad (17)$$

Since a diploid genotype can only be formed by two chromosomes that have different alleles (X or Y), the frequencies of X-chromosomes and of Y-chromosomes are equal. We therefore consider independently the frequency of inversions in proto-X chromosomes and in proto-Y chromosomes, and we can make the assumption that $F_{XN} + F_{XI} = 1$ and $F_{YN} + F_{YI} = 1$ to ease the calculations.

There are four possible genotypes, which frequencies are computed assuming random mating among haplotypes.

Genotype	Frequency	Fitness
XN/YN	$F_{XN}F_{YN}$	W_{NN}
XI/YN	$F_{XI}F_{YN}$	W_{NI}
XN/YI	$F_{XN}F_{YI}$	W_{NI}
XI/YI	$F_{XI}F_{YI}$	W_{II}

The mean fitness of the population is thus

$$\overline{W} = F_{XN}F_{YN}W_{NN} + \big(F_{XI}F_{YN} + F_{XN}F_{YI}\big)W_{NI} + F_{XI}F_{YI}W_{II}.$$

6.2 Evolution of frequencies

Considering a recombination rate r between the permanently heterozygous locus and the inversion locus, and a generation t, we can compute the frequency of the haplotype XI in the next generation:

$$\begin{aligned} F_{XI}^{t+1} &= \frac{F_{XI}^t F_{YI}^t W_{II} + F_{XI}^t F_{YN}^t W_{NI}(1-r) + F_{XN}^t F_{YI}^t W_{NI} r}{\overline{W^t}} \\ &= \frac{F_{XI}^t (F_{YI}^t W_{II} + F_{YN}^t W_{NI}) - r W_{NI}(F_{XI}^t F_{YN}^t - F_{XN}^t F_{YI}^t)}{\overline{W^t}} \\ &= \frac{F_{XI}^t W_{XI}^* - r W_{NI} D^t}{\overline{W^t}}, \end{aligned} \tag{18}$$

with $D^t = F_{XI}^t F_{YN}^t - F_{XN}^t F_{YI}^t$, the linkage disequilibrium, and $W_{XI}^* = F_{YN} W_{NI} + F_{YI} W_{II}$, the marginal fitness of haplotype XI.

The change in frequency in each generation is then

$$\begin{aligned} \Delta F_{XI}^t &= F_{XI}^{t+1} - F_{XI}^t \\ &= \frac{F_{XI}^t W_{XI}^* - r W_{NI} D^t}{\overline{W^t}} - \frac{F_{XI}^t \overline{W^t}}{\overline{W^t}} \\ &= \frac{F_{XI}^t (W_{XI}^* - \overline{W^t}) - r W_{NI} D^t}{\overline{W^t}} \\ &= \frac{F_{XI}^t F_{XN}^t (W_{XI}^* - W_{XN}^*) - r W_{NI} D^t}{\overline{W^t}} \\ &= \frac{F_{XI}^t (1 - F_{XI}^t)(W_{XI}^* - W_{XN}^*) - r W_{NI} D^t}{\overline{W^t}}. \end{aligned} \tag{19}$$

As before, we can determine the frequencies of XN, YI or YN in generation $t+1$ by interchanging the roles of X and Y, or of I and N in the equation for F_{XI}^{t+1} to obtain

$$\begin{aligned} F_{XN}^{t+1} &= \frac{F_{XN}^t (F_{YN}^t W_{NN} + F_{YI}^t W_{NI}) + r W_{NI} D^t}{\overline{W^t}}, \\ F_{YN}^{t+1} &= \frac{F_{YN}^t (F_{XN}^t W_{NN} + F_{XI}^t W_{NI}) - r W_{NI} D^t}{\overline{W^t}}, \\ F_{YI}^{t+1} &= \frac{F_{YI}^t (F_{XI}^t W_{II} + F_{XI}^t W_{NI}) + r W_{NI} D^t}{\overline{W^t}}. \end{aligned} \tag{20}$$

We observe that these equations are the same as in the overdominant case when $s_2 = 1$, *i.e.* when no homozygous are formed.

7 Trajectory of inversions in linkage with a permanently heterozygous locus with three alleles (S7 Fig)

7.1 Allelic structure

At a given locus with three alleles, X, Y and Z, we consider that individuals are all heterozygotes XY, XZ or YZ.

There are six possible haplotypes :

$Haplotype$	$Frequency$
XN	F_{XN}
XI	F_{XI}
YN	F_{YN}
YI	F_{YI}
ZN	F_{ZN}
ZI	F_{ZI}

(21)

The derivation of the frequency of each genotype is a bit more challenging than in the case of two permanently heterozygous alleles.

We denote the frequency of chromosomes with allele X among all chromosomes by F_X. We thus have $F_X = F_{XI} + F_{XN}$, and likewise for F_Y and F_Z. We also have $F_X + F_Y + F_Z = 1$.

We can compute the frequency of each genotype (e.g F_{XIYN}, the frequency of the genotype formed by a haplotype XI and a haplotype YN) assuming random mating. For instance, to compute F_{XIYN}, we have first to pick one of the two haplotypes among all haplotypes (either XI with probability F_{XI} or YN with probability F_{YN}), and then pick the second among the remaining compatible haplotypes. If XI was picked first, the second haplotype has to carry either the Y allele or the Z allele to ensure the heterozygosity. Among those compatible haplotypes, the YN haplotype has a probability of $\frac{F_{YN}}{F_Y+F_Z}$ to be picked. We have thus :

$$\begin{aligned}
F_{XIYN} &= \mathbb{P}(\text{choose } XI \text{ first})\mathbb{P}(\text{choose } YN \text{ after a } XI) + \mathbb{P}(\text{choose } YN \text{ first})\mathbb{P}(\text{choose } XI \text{ after a } YN) \\
&= \mathbb{P}(\text{choose } XI \text{ first})\mathbb{P}(\text{choose } YN \text{ among the } Y \text{ and } Z \text{ chrom.}) \\
&\qquad + \mathbb{P}(\text{choose } YN \text{ first})\mathbb{P}(\text{choose } XI \text{ among the } X \text{ and } Z \text{ chrom.}) \\
&= F_{XI}\frac{F_{YN}}{F_Y + F_Z} + F_{YN}\frac{F_{XI}}{F_X + F_Z} \\
&= F_{XI}\frac{F_{YN}}{1 - F_X} + F_{YN}\frac{F_{XI}}{1 - F_Y}.
\end{aligned}$$

We reduce the fractions to a common denominator, and use the equation $F_X + F_Y + F_Z = 1$ to obtain :

$$\begin{aligned}
F_{XIYN} &= F_{XI}F_{YN}\left(\frac{1 - F_Y}{(1 - F_X)(1 - F_Y)} + \frac{1 - F_X}{(1 - F_X)(1 - F_Y)}\right) \\
&= F_{XI}F_{YN}\left(\frac{1 + 1 - F_Y - F_X}{(1 - F_X)(1 - F_Y)}\right) \\
&= F_{XI}F_{YN}\frac{1 + F_Z}{(1 - F_X)(1 - F_Y)}.
\end{aligned}$$

The possible genotypes and their corresponding frequencies are therefore:

OD genotype	Inversion genotype	Frequency	Fitness
XY	II	$F_{XI}F_{YI}\frac{1 + F_Z}{(1 - F_X)(1 - F_Y)}$	W_{II}
	IN	$F_{XI}F_{YN}\frac{1 + F_Z}{(1 - F_X)(1 - F_Y)}$	W_{IN}
	NI	$F_{XN}F_{YI}\frac{1 + F_Z}{(1 - F_X)(1 - F_Y)}$	W_{IN}
	NN	$F_{XN}F_{YN}\frac{1 + F_Z}{(1 - F_X)(1 - F_Y)}$	W_{NN}
XZ	II	$F_{XI}F_{ZI}\frac{1 + F_Y}{(1 - F_X)(1 - F_Z)}$	W_{II}
	IN	$F_{XI}F_{ZN}\frac{1 + F_Y}{(1 - F_X)(1 - F_Z)}$	W_{IN}
	NI	$F_{XN}F_{ZI}\frac{1 + F_Y}{(1 - F_X)(1 - F_Z)}$	W_{IN}
	NN	$F_{XN}F_{ZN}\frac{1 + F_Y}{(1 - F_X)(1 - F_Z)}$	W_{NN}
YZ	II	$F_{YI}F_{ZI}\frac{1 + F_X}{(1 - F_Y)(1 - F_Z)}$	W_{II}
	IN	$F_{YI}F_{ZN}\frac{1 + F_X}{(1 - F_Y)(1 - F_Z)}$	W_{IN}
	NI	$F_{YN}F_{ZI}\frac{1 + F_X}{(1 - F_Y)(1 - F_Z)}$	W_{IN}
	NN	$F_{YN}F_{ZN}\frac{1 + F_X}{(1 - F_Y)(1 - F_Z)}$	W_{NN}

The mean fitness of the population is thus

$$\begin{aligned}\overline{W} =& F_{XIYI}W_{II} + F_{XIYN}W_{NI} + F_{XNYI}W_{NI} + F_{XNYN}W_{NN} + \\ & F_{XIZI}W_{II} + F_{XIZN}W_{NI} + F_{XNZI}W_{NI} + F_{XNZN}W_{NN} + \\ & F_{YIZI}W_{II} + F_{YIZN}W_{NI} + F_{YNZI}W_{NI} + F_{YNZN}W_{NN}.\end{aligned} \tag{22}$$

7.2 Evolution of frequencies

Considering a recombination rate r between the permanently heterozygous locus and the inversion, and a generation t, we can compute the frequency of each haplotype in the next generation. We write $F_X^t := F_{XN}^t + F_{XI}^t$ (likewise for F_Y^t and F_Z^t) and set $\widehat{F}^t := (1-F_X^t)(1-F_Y^t)(1-F_Z^t)$. We have

$$\begin{aligned}F_{XI}^{t+1} = & \frac{W_{II}}{2\overline{W^t}}\frac{F_{XI}^t}{\widehat{F}^t}\left[F_{YI}^t\left(1-(F_Z^t)^2\right) + F_{ZI}^t\left(1-(F_Y^t)^2\right)\right] \\ & + \frac{W_{IN}}{2\overline{W^t}}\frac{F_{XI}^t}{\widehat{F}^t}\left[F_{YN}^t\left(1-(F_Z^t)^2\right) + F_{ZN}^t\left(1-(F_Y^t)^2\right)\right] \\ & + r\frac{W_{IN}}{2\overline{W^t}}\left(\frac{F_{XN}^t}{\widehat{F}^t}\left[F_{YI}^t\left(1-(F_Z^t)^2\right) + F_{ZI}^t\left(1-(F_Y^t)^2\right)\right] - \frac{F_{XI}^t}{\widehat{F}^t}\left[F_{YN}^t\left(1-(F_Z^t)^2\right) + F_{ZN}^t\left(1-(F_Y^t)^2\right)\right]\right)\end{aligned} \tag{23}$$

$$\begin{aligned}F_{XN}^{t+1} = & \frac{W_{NN}}{2\overline{W^t}}\frac{F_{XN}^t}{\widehat{F}^t}\left[F_{YN}^t\left(1-(F_Z^t)^2\right) + F_{ZN}^t\left(1-(F_Y^t)^2\right)\right] \\ & + \frac{W_{IN}}{2\overline{W^t}}\frac{F_{XN}^t}{\widehat{F}^t}\left[F_{YI}^t\left(1-(F_Z^t)^2\right) + F_{ZI}^t\left(1-(F_Y^t)^2\right)\right] \\ & + r\frac{W_{IN}}{2\overline{W^t}}\left(\frac{F_{XI}^t}{\widehat{F}^t}\left[F_{YN}^t\left(1-(F_Z^t)^2\right) + F_{ZN}^t\left(1-(F_Y^t)^2\right)\right] - \frac{F_{XN}^t}{\widehat{F}^t}\left[F_{YI}^t\left(1-(F_Z^t)^2\right) + F_{ZI}^t\left(1-(F_Y^t)^2\right)\right]\right)\end{aligned} \tag{24}$$

$$\begin{aligned}F_{YI}^{t+1} = & \frac{W_{II}}{2\overline{W^t}}\frac{F_{YI}^t}{\widehat{F}^t}\left[F_{XI}^t\left(1-(F_Z^t)^2\right) + F_{ZI}^t\left(1-(F_X^t)^2\right)\right] \\ & + \frac{W_{IN}}{2\overline{W^t}}\frac{F_{YI}^t}{\widehat{F}^t}\left[F_{XN}^t\left(1-(F_Z^t)^2\right) + F_{ZN}^t\left(1-(F_X^t)^2\right)\right] \\ & + r\frac{W_{IN}}{2\overline{W^t}}\left(\frac{F_{YN}^t}{\widehat{F}^t}\left[F_{XI}^t\left(1-(F_Z^t)^2\right) + F_{ZI}^t\left(1-(F_X^t)^2\right)\right] - \frac{F_{YI}^t}{\widehat{F}^t}\left[F_{XN}^t\left(1-(F_Z^t)^2\right) + F_{ZN}^t\left(1-(F_X^t)^2\right)\right]\right)\end{aligned} \tag{25}$$

$$\begin{aligned}F_{YN}^{t+1} = & \frac{W_{NN}}{2\overline{W^t}}\frac{F_{YN}^t}{\widehat{F}^t}\left[F_{XN}^t\left(1-(F_Z^t)^2\right) + F_{ZN}^t\left(1-(F_X^t)^2\right)\right] \\ & + \frac{W_{IN}}{2\overline{W^t}}\frac{F_{YN}^t}{\widehat{F}^t}\left[F_{XI}^t\left(1-(F_Z^t)^2\right) + F_{ZI}^t\left(1-(F_X^t)^2\right)\right] \\ & + r\frac{W_{IN}}{2\overline{W^t}}\left(\frac{F_{YI}^t}{\widehat{F}^t}\left[F_{XN}^t\left(1-(F_Z^t)^2\right) + F_{ZN}^t\left(1-(F_X^t)^2\right)\right] - \frac{F_{YN}^t}{\widehat{F}^t}\left[F_{XI}^t\left(1-(F_Z^t)^2\right) + F_{ZI}^t\left(1-(F_X^t)^2\right)\right]\right)\end{aligned} \tag{26}$$

$$\begin{aligned}F_{ZI}^{t+1} = & \frac{W_{II}}{2\overline{W^t}}\frac{F_{ZI}^t}{\widehat{F}^t}\left[F_{XI}^t\left(1-(F_Y^t)^2\right) + F_{YI}^t\left(1-(F_X^t)^2\right)\right] \\ & + \frac{W_{IN}}{2\overline{W^t}}\frac{F_{ZI}^t}{\widehat{F}^t}\left[F_{XN}^t\left(1-(F_Y^t)^2\right) + F_{YN}^t\left(1-(F_X^t)^2\right)\right] \\ & + r\frac{W_{IN}}{2\overline{W^t}}\left(\frac{F_{ZN}^t}{\widehat{F}^t}\left[F_{XI}^t\left(1-(F_Y^t)^2\right) + F_{YI}^t\left(1-(F_X^t)^2\right)\right] - \frac{F_{ZI}^t}{\widehat{F}^t}\left[F_{XN}^t\left(1-(F_Y^t)^2\right) + F_{YN}^t\left(1-(F_X^t)^2\right)\right]\right)\end{aligned} \tag{27}$$

$$
\begin{aligned}
F_{ZN}^{t+1} = & \frac{W_{NN}}{2\overline{W^t}} \frac{F_{ZN}^t}{\widehat{F}^t} \left[F_{XN}^t \left(1 - (F_Y^t)^2\right) + F_{YN}^t \left(1 - (F_X^t)^2\right) \right] \\
& + \frac{W_{IN}}{2\overline{W^t}} \frac{F_{ZN}^t}{\widehat{F}^t} \left[F_{XI}^t \left(1 - (F_Y^t)^2\right) + F_{YI}^t \left(1 - (F_X^t)^2\right) \right] \\
& + r \frac{W_{IN}}{2\overline{W^t}} \left(\frac{F_{ZI}^t}{\widehat{F}^t} \left[F_{XN}^t \left(1 - (F_Y^t)^2\right) + F_{YN}^t \left(1 - (F_X^t)^2\right) \right] - \frac{F_{ZN}^t}{\widehat{F}^t} \left[F_{XI}^t \left(1 - (F_Y^t)^2\right) + F_{YI}^t \left(1 - (F_X^t)^2\right) \right] \right)
\end{aligned}
\tag{28}
$$

The expression for the change in frequency over one generation can be obtained from the above formulae. Since the expression is complex and not particularly informative, we do not detail it here.

If we set $F_{ZI} = F_{ZN} = 0$, we obtain the same equations as in the case of two alleles for the permanent heterozygous locus.

8 Trajectory of inversions in linkage with an XY mammal-like sex-determining locus (Figs 2, 3 & S4)

8.1 Allelic structure

At a given locus with two alleles, X and Y, we suppose that individuals can be either XX or XY and must mate with a different genotype.

There are still the same four possible haplotypes in the population, but, following [2], it is useful to distinguish the frequency of inverted and non-inverted segments in proto-X chromosomes in gametes produced by males (F_{XIm}) and by females (F_{XIf}).

Haplotype	Frequency in male	Frequency in female
XN	F_{XNm}	F_{XNf}
XI	F_{XIm}	F_{XIf}
YN	F_{YN}	0
YI	F_{YI}	0

We consider independently the frequency of inversions in proto-X chromosomes and proto-Y chromosomes so that we can assume $F_{XNf} + F_{XIf} = 1$, $F_{XNm} + F_{XIm} = 1$ and $F_{YN} + F_{YI} = 1$ to ease the calculations.

There are seven possible genotypes which frequencies can be computed assuming random mating.

Genotype	Frequency	Fitness
XN/XN	$F_{XNm}F_{XNf}$	W_{NN}
XN/XI	$F_{XNm}F_{XIf} + F_{XNf}F_{XIm}$	W_{NI}
XI/XI	$F_{XIm}F_{XIf}$	W_{II}
XN/YN	$F_{XNf}F_{YN}$	W_{NN}
XI/YN	$F_{XIf}F_{YN}$	W_{NI}
XN/YI	$F_{XNf}F_{YI}$	W_{NI}
XI/YI	$F_{XIf}F_{YI}$	W_{II}

We consider independently the fitness of males ($\overline{W_m}$) and females ($\overline{W_f}$). These quantities are given by

$$\overline{W_m} = F_{XNf}F_{YN}W_{NN} + F_{XNf}F_{YI}W_{NI} + F_{XIf}F_{YN}W_{NI} + F_{XIf}F_{YI}W_{II}$$

and

$$\overline{W_f} = F_{XNf}F_{XNm}W_{NN} + F_{XNm}F_{XIf}W_{NI} + F_{XNf}F_{XIm}W_{NI} + F_{XIf}F_{XIm}W_{II}.$$

8.2 Evolution of frequencies

Considering a recombination rate r between the sex-determining locus and the inversion, and a generation t, we can follow the frequency of inverted and non-inverted segments in proto-X chromosomes independently in gametes produced by males (F_{XIm}) and by females (F_{XIf}):

$$\begin{aligned}
F_{XIm}^{t+1} &= \frac{F_{XIf}^t F_{YI}^t W_{II} + F_{XIf}^t F_{YN}^t W_{NI}(1-r) + F_{XNf}^t F_{YI}^t W_{NI} r}{\overline{W_m^t}} \\
&= \frac{F_{XIf}^t \left(W_{II} F_{YI}^t + W_{NI} F_{YN}^t\right) + r W_{NI} D^t}{\overline{W_m^t}}, \\
F_{XNm}^{t+1} &= \frac{F_{XNf}^t F_{YN}^t W_{NN} + F_{XNf}^t F_{YI}^t W_{NI}(1-r) + F_{XIf}^t F_{YN}^t W_{NI} r}{\overline{W_m^t}} \\
&= \frac{F_{XNf}^t \left(W_{NN} F_{YN}^t + W_{NI} F_{YI}^t\right) - r W_{NI} D^t}{\overline{W_m^t}}, \\
F_{XIf}^{t+1} &= \frac{F_{XIm}^t F_{XIf}^t W_{II} + \frac{1}{2} W_{NI}[F_{XIm}^t F_{XNf}^t + F_{XIf}^t F_{XNm}^t]}{\overline{W_f^t}}, \\
F_{XNf}^{t+1} &= \frac{F_{XNm}^t F_{XNf}^t W_{NN} + \frac{1}{2} W_{NI}[F_{XNm}^t F_{XIf}^t + F_{XNf}^t F_{XIm}^t]}{\overline{W_f^t}},
\end{aligned} \tag{29}$$

with $D^t = F_{XNf}^t F_{YI}^t - F_{XIf}^t F_{YN}^t$.

Since two third of the X chromosomes are in females and one third are in male, the frequency of the inversion in proto-X chromosomes in the whole population can be obtained by:

$$F_{XI} = \frac{2F_{XIf}}{3} + \frac{F_{XIm}}{3}.$$

The variation in the frequency of XI over one generation is then :

$$\begin{aligned}
\Delta F_{XI}^t &= F_{XI}^{t+1} - F_{XI}^t \\
&= \frac{2}{3}\left(F_{XIf}^{t+1} - F_{XIf}^t\right) + \frac{1}{3}\left(F_{XIm}^{t+1} - F_{XIm}^t\right) \\
&= \frac{2}{3}\frac{1}{\overline{W_f^t}}\Big[F_{XIf}^t F_{XNf}^t \left(W_{II} F_{XIm}^t - W_{NN} F_{XNm}^t\right) \\
&\qquad + W_{NI}\left(\frac{1}{2} - F_{XIf}^t\right)\left(F_{XIm}^t F_{XNf}^t + F_{XIf}^t F_{XNm}^t\right)\Big] \\
&+ \frac{1}{3}\frac{1}{\overline{W_m^t}}\Big[W_{II} F_{YI}^t F_{XIf}^t F_{XNm}^t - W_{NN} F_{XIm}^t F_{XNf}^t F_{YN}^t \\
&\qquad + W_{NI}\Big(F_{YN}^t F_{XIf}^t F_{XNm}^t - F_{XIm}^t F_{XNf}^t F_{YI}^t\Big) + r W_{NI} D^t\Big]
\end{aligned} \tag{30}$$

The changes in frequency of inverted and non-inverted segments in proto-Y chromosomes are calculated as before:

$$\begin{aligned}
F_{YI}^{t+1} &= \frac{F_{YI}^t (F_{XIf}^t W_{II} + F_{XNf}^t W_{NI}) - r W_{NI} D^t}{\overline{W_m^t}}, \\
F_{YN}^{t+1} &= \frac{F_{YN}^t (F_{XNf}^t W_{NN} + F_{XIf}^t W_{NI}) + r W_{NI} D^t}{\overline{W_m^t}}.
\end{aligned} \tag{31}$$

9 Deterministic assessment of inversion fate

The fate of inversions can be determined as a function of their fitness. As summarized in the Methods sections, inversions linked to a Y allele or on autosomes can initially increase in frequency only if heterozygotes for the inversion have a better fitness than homozygotes for the absence of inversion (i.e. $W_{NI} > W_{NN}$). This condition is sufficient for a Y-linked inversion to go to fixation, whereas

on autosomes, inversion can go to fixation only if homozygotes for the inversion must have a better fitness than heterozygotes for the inversion ($W_{II} > W_{NI}$). We can easily derive the conditions on the number of mutations carried by the inversion (m) for these situations to occur. We have

$$\begin{aligned} & W_{II} > W_{NI} \\ \text{if and only if} \quad & (1-s)^m > (q(1-s)+(1-q)(1-hs))^m (q(1-hs)+1-q)^{n-m}, \\ \text{i.e.} \quad & \left[\frac{\big(1-s\big)\big(q(1-hs)+1-q\big)}{q(1-s)+(1-q)(1-hs)}\right]^m > \left[q(1-hs)+1-q\right]^n, \\ \text{i.e.} \quad & m \ln\left[\frac{\big(1-s\big)\big(q(1-hs)+1-q\big)}{q(1-s)+(1-q)(1-hs)}\right] > n \ln\left[q(1-hs)+1-q\right]. \end{aligned}$$

Since $1-s < 1-hs < 1$, we have $\big(1-s\big)\big(q(1-hs)+1-q\big) = q(1-s)(1-hs)+(1-q)(1-q) < q(1-s)+(1-q)(1-hq)$, and thus $\ln\left[\frac{\big(1-s\big)\big(q(1-hs)+1-q\big)}{q(1-s)+(1-q)(1-hs)}\right] < 0$.

Therefore,

$$W_{II} > W_{NI} \quad \text{if and only if} \quad m < n \times \frac{\ln\left[q(1-hs)+1-q\right]}{\ln\left[\frac{\big(1-s\big)\big(q(1-hs)+1-q\big)}{q(1-s)+(1-q)(1-hs)}\right]} = n \times \beta_1(q,h,s).$$

Similarly, we have

$$\begin{aligned} & W_{NI} > W_{NN} \\ \text{if and only if} \quad & (q(1-s)+(1-q)(1-hs))^m (q(1-hs)+1-q)^{n-m} > (1-2q(1-q)hs-q^2s)^n, \\ \text{i.e.} \quad & \left[\frac{q(1-s)+(1-q)(1-hs)}{q(1-hs)+1-q}\right]^m > \left[\frac{1-2q(1-q)hs-q^2s}{q(1-hs)+1-q}\right]^n, \\ \text{i.e.} \quad & m \ln\left[\frac{q(1-s)+(1-q)(1-hs)}{q(1-hs)+1-q}\right] > n \ln\left[\frac{1-2q(1-q)hs-q^2s}{q(1-hs)+1-q}\right], \\ \text{i.e.} \quad & m < n \frac{\ln\left[\frac{1-2q(1-q)hs-q^2s}{q(1-hs)+1-q}\right]}{\ln\left[\frac{q(1-s)+(1-q)(1-hs)}{q(1-hs)+1-q}\right]} = n \times \beta_2(q,h,s). \end{aligned}$$

Assuming $q \ll 1$ and $s \ll 1$, β_1 and β_2 can be approximated by $\beta_1 \approx q\frac{h}{1-h}$ and $\beta_2 \approx q$. This approximate value of β_2 is the same that the one defined by [1] using similar assumptions, as expected. On autosomes, inversions should therefore increase in frequency when $m < nq$ and go to fixation when $m < \frac{nqh}{1-h}$. When permanently heterozygous (for instance if capturing the male-determining allele on the Y chromosome), inversions should go to fixation when $m < nq$, *i.e.* when they capture fewer mutations than the population average

10 Inversions with of an heterozygous fitness cost

Crossover events occurring between the breakpoints of differentially-oriented segments (between inverted and non-inverted haplotype) can produce unbalanced gametes, which could reduce the fertility of individual heterozygous for inversions. Here we assess the effect of such a heterozygous fitness cost on the fate of the inversion.

We write $\eta \in [0,1]$ the heterozygous cost, and W_{NI}^{η} the fitness of an individual heterozygous for the inversion, so that $W_{NI}^{\eta} = W_{NI} \times (1-\eta)$. As before, we derive the conditions on the number of mutations carried by the inversion for it to spread or fix.

We have

$$\begin{aligned} & W_{II} > W_{NI}^{\eta} \\ & \quad \text{iff} \quad (1-s)^m > (q(1-s)+(1-q)(1-hs))^m (q(1-hs)+1-q)^{n-m} \times (1-\eta). \end{aligned}$$

As in the case without heterozygous cost, we write the above inequality as a condition on the number m of mutations carried by the inversion and obtain

$$
\begin{aligned}
&W_{II} > W_{NI}^{\eta} \\
&\quad \text{iff} \quad m < n \times \frac{\ln\left[q(1-hs)+1-q\right]}{\ln\left[\dfrac{\big(1-s\big)\big(q(1-hs)+1-q\big)}{q(1-s)+(1-q)(1-hs)}\right]} + \frac{\ln(1-\eta)}{\ln\left[\dfrac{\big(1-s\big)\big(q(1-hs)+1-q\big)}{q(1-s)+(1-q)(1-hs)}\right]} \\
&\quad = n \times \left(\beta_1(q,h,s) + \frac{1}{n} \frac{\ln(1-\eta)}{\ln\left[\dfrac{\big(1-s\big)\big(q(1-hs)+1-q\big)}{q(1-s)+(1-q)(1-hs)}\right]} \right) \\
&\quad = n \times \left(\beta_1(q,h,s) + \frac{1}{n}\delta_1^{\eta} \right), \\
&\text{with } \delta_1^{\eta} = \frac{\ln(1-\eta)}{\ln\left[\dfrac{\big(1-s\big)\big(q(1-hs)+1-q\big)}{q(1-s)+(1-q)(1-hs)}\right]} \underset{q \ll 1}{\approx} \frac{\ln(1-\eta)}{s(h-1)}.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&W_{NI}^{\eta} > W_{NN} \\
&\quad \text{iff} \quad (q(1-s)+(1-q)(1-hs))^m (q(1-hs)+1-q)^{n-m} \times (1-\eta) > (1-2q(1-q)hs-q^2 s)^n, \\
&\quad \text{iff} \quad m < n \frac{\ln\left[\dfrac{1-2q(1-q)hs-q^2 s}{q(1-hs)+1-q}\right]}{\ln\left[\dfrac{q(1-s)+(1-q)(1-hs)}{q(1-hs)+1-q}\right]} - \frac{\ln(1-\eta)}{\ln\left[\dfrac{q(1-s)+(1-q)(1-hs)}{q(1-hs)+1-q}\right]} \\
&\quad = n \times \left(\beta_2(q,h,s) - \frac{1}{n} \frac{\ln(1-\eta)}{\ln\left[\dfrac{q(1-s)+(1-q)(1-hs)}{q(1-hs)+1-q}\right]} \right) \\
&\quad = n \times \left(\beta_2(q,h,s) - \frac{1}{n}\delta_2^{\eta} \right), \\
&\text{with } \delta_2^{\eta} = \frac{\ln(1-\eta)}{\ln\left[\dfrac{q(1-s)+(1-q)(1-hs)}{q(1-hs)+1-q}\right]} \underset{q \ll 1}{\approx} \frac{\ln(1-\eta)}{-hs}.
\end{aligned}
$$

The new thresholds satisfy $\tilde{\beta}_1 := \beta_1(q,h,s) + \frac{1}{n}\delta_1^{\eta} > \beta_1(q,h,s)$ and $\tilde{\beta}_2 := \beta_2(q,h,s) - \frac{1}{n}\delta_2^{\eta} < \beta_2(q,h,s)$.

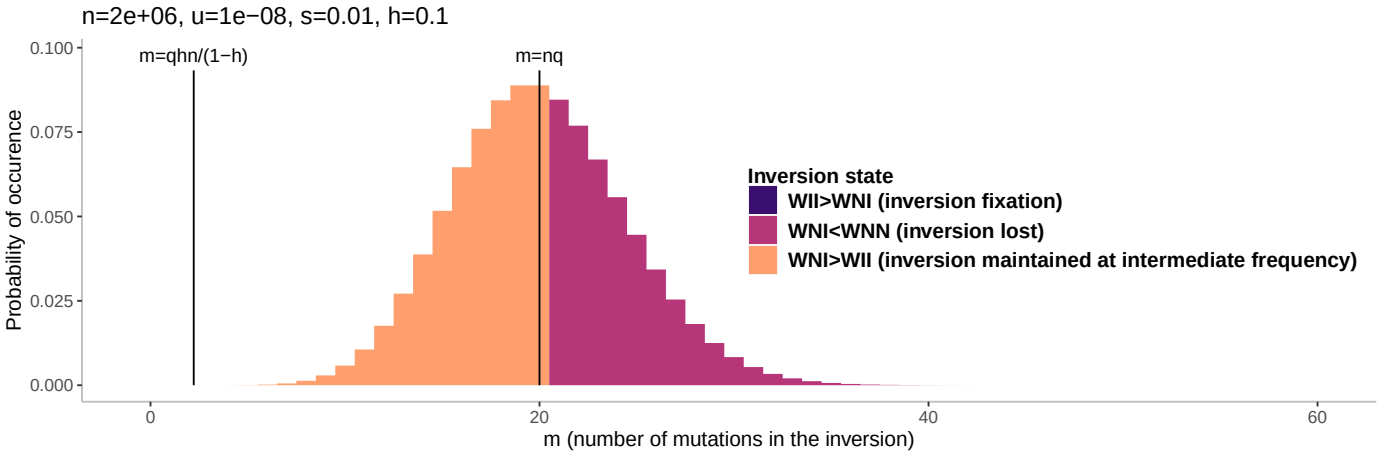
Thus, inversions capturing a permanently heterozygous allele are less likely to fix when suffering from a heterozygous cost than without such a cost ($\tilde{\beta}_2 < \beta_2$), whereas inversions on autosomes are less likely to segregate at intermediate frequencies ($\tilde{\beta}_2 < \beta_2$), but more likely to fix when suffering from a heterozygous cost than without such a cost ($\tilde{\beta}_1 > \beta_1$).

References

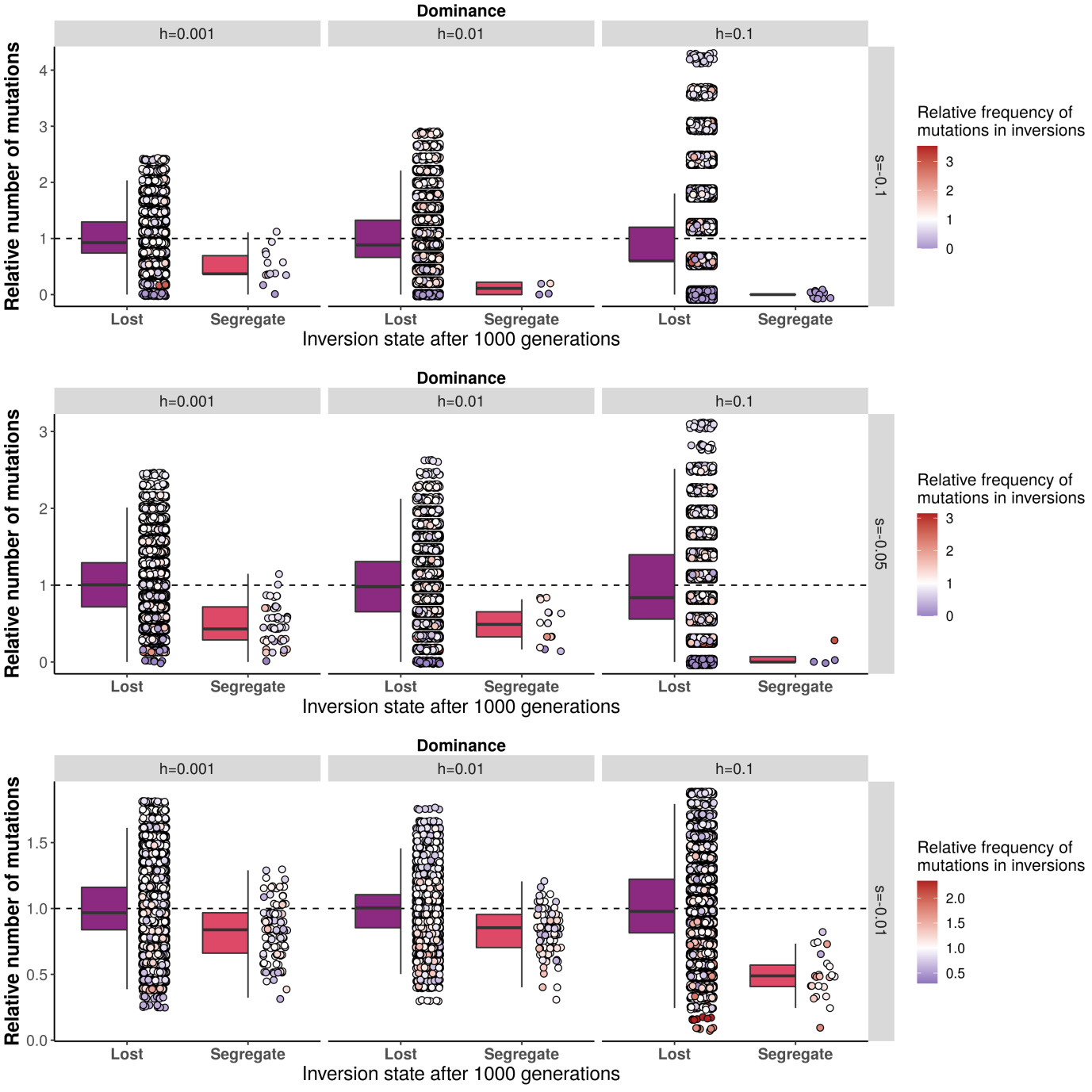
[1] Masatoshi Nei, Ken-Ichi Kojima, and Henry E. Schaffer. Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics*, 57(4):741–750, 1967.

[2] A. G. Clark. The Evolution of the Y Chromosome with X-Y Recombination. *Genetics*, 119(3):711–720, 1988.

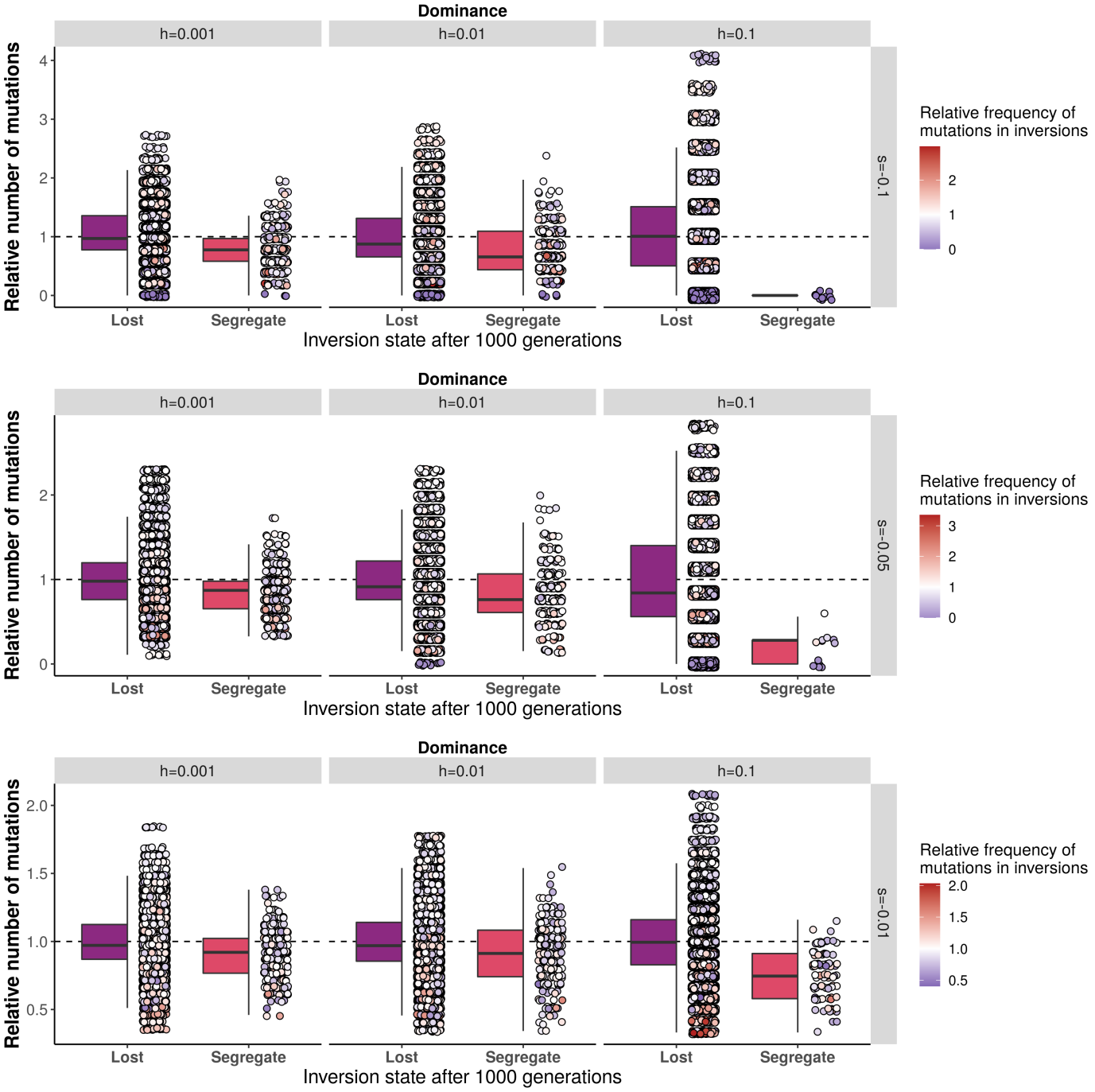
Supplementary figures



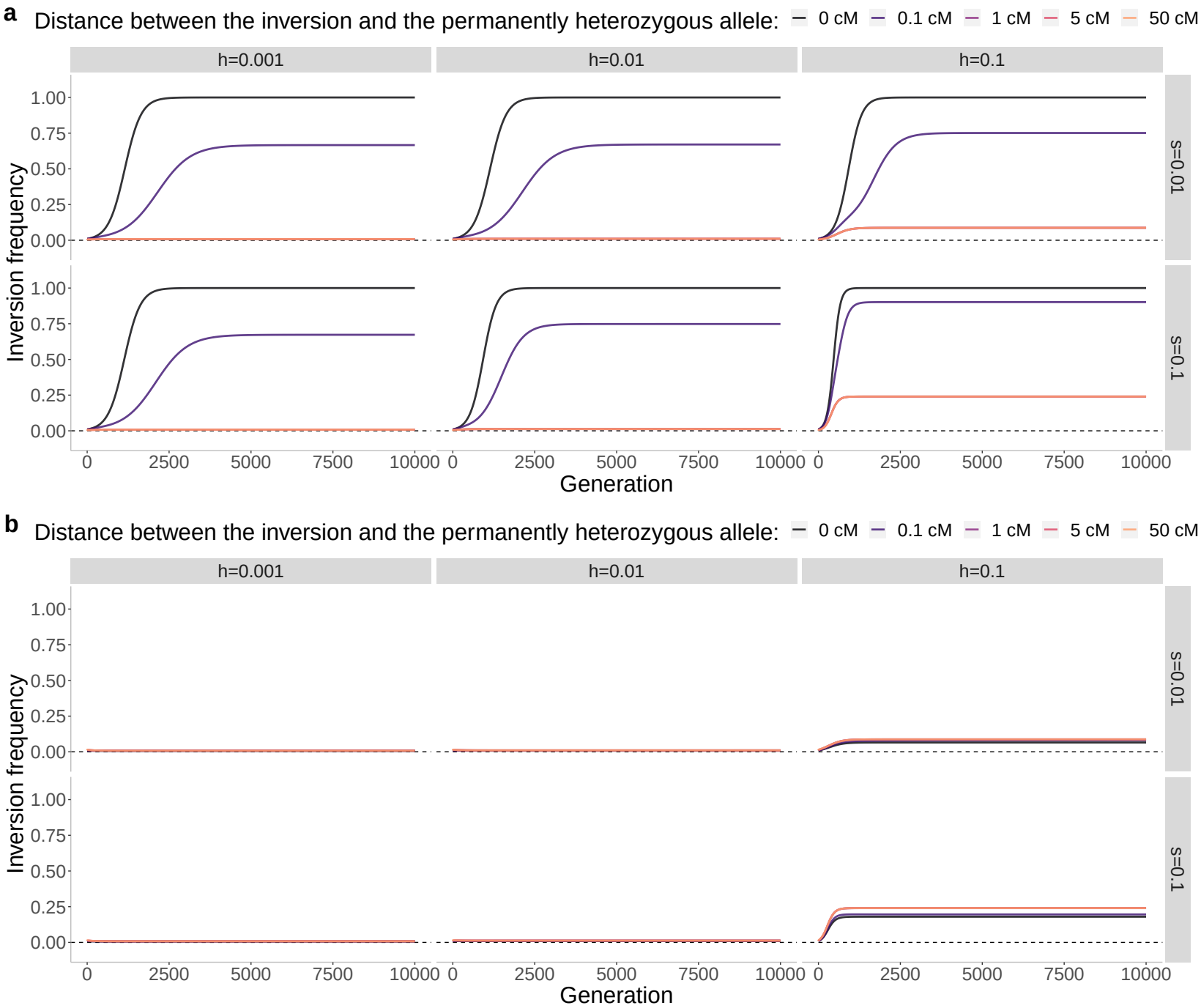
S1 Fig


Dominance
h=0.001
h=0.01
h=0.1
Relative number of mutations
s=-0.1
s=-0.05
s=-0.01
Lost
Segregate
Inversion state after 1000 generations
Relative frequency of mutations in inversions

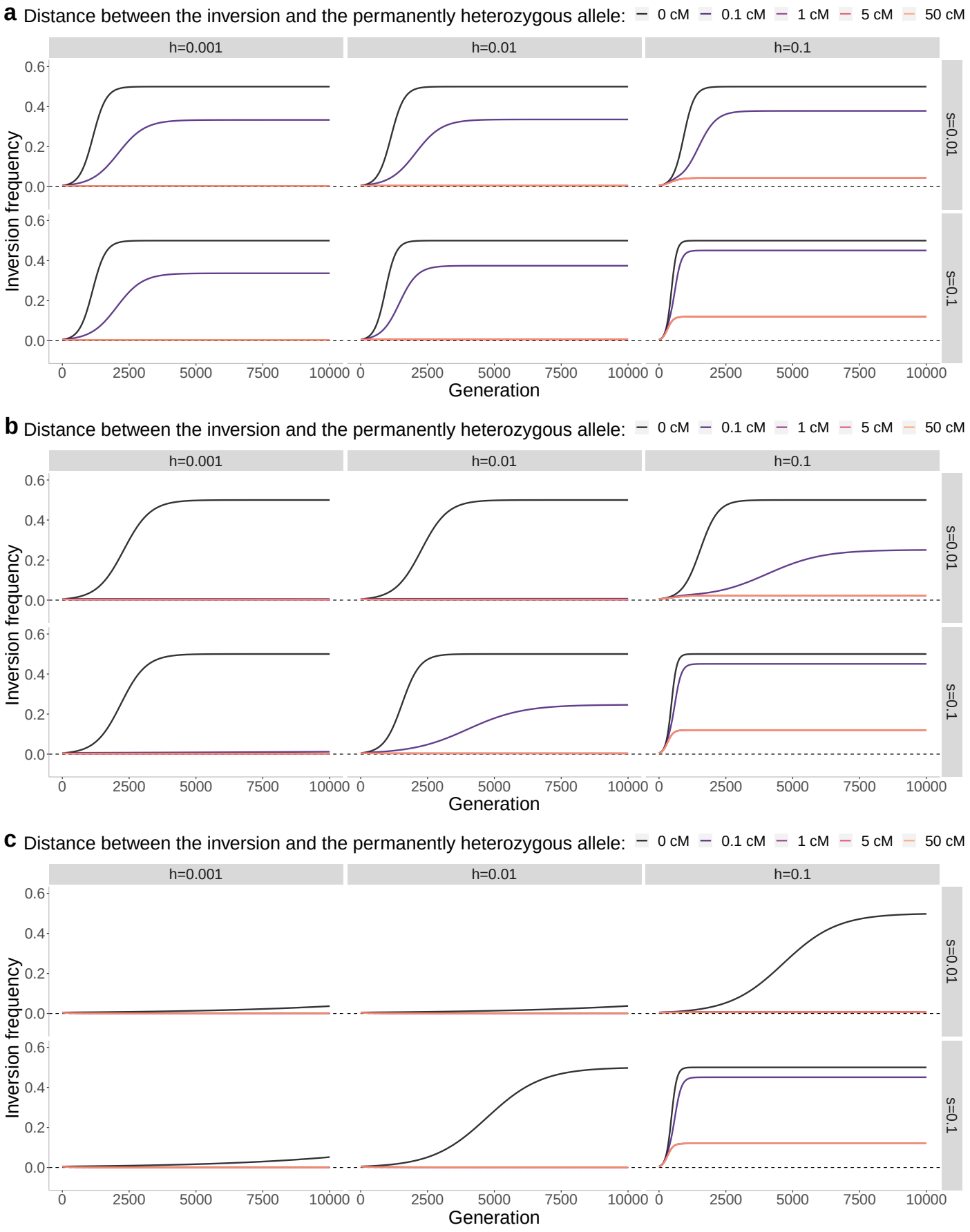
S2 Fig


Dominance
h=0.001
h=0.01
h=0.1
Relative number of mutations
s=-0.1
s=-0.05
s=-0.01
Lost
Segregate
Inversion state after 1000 generations
Relative frequency of mutations in inversions

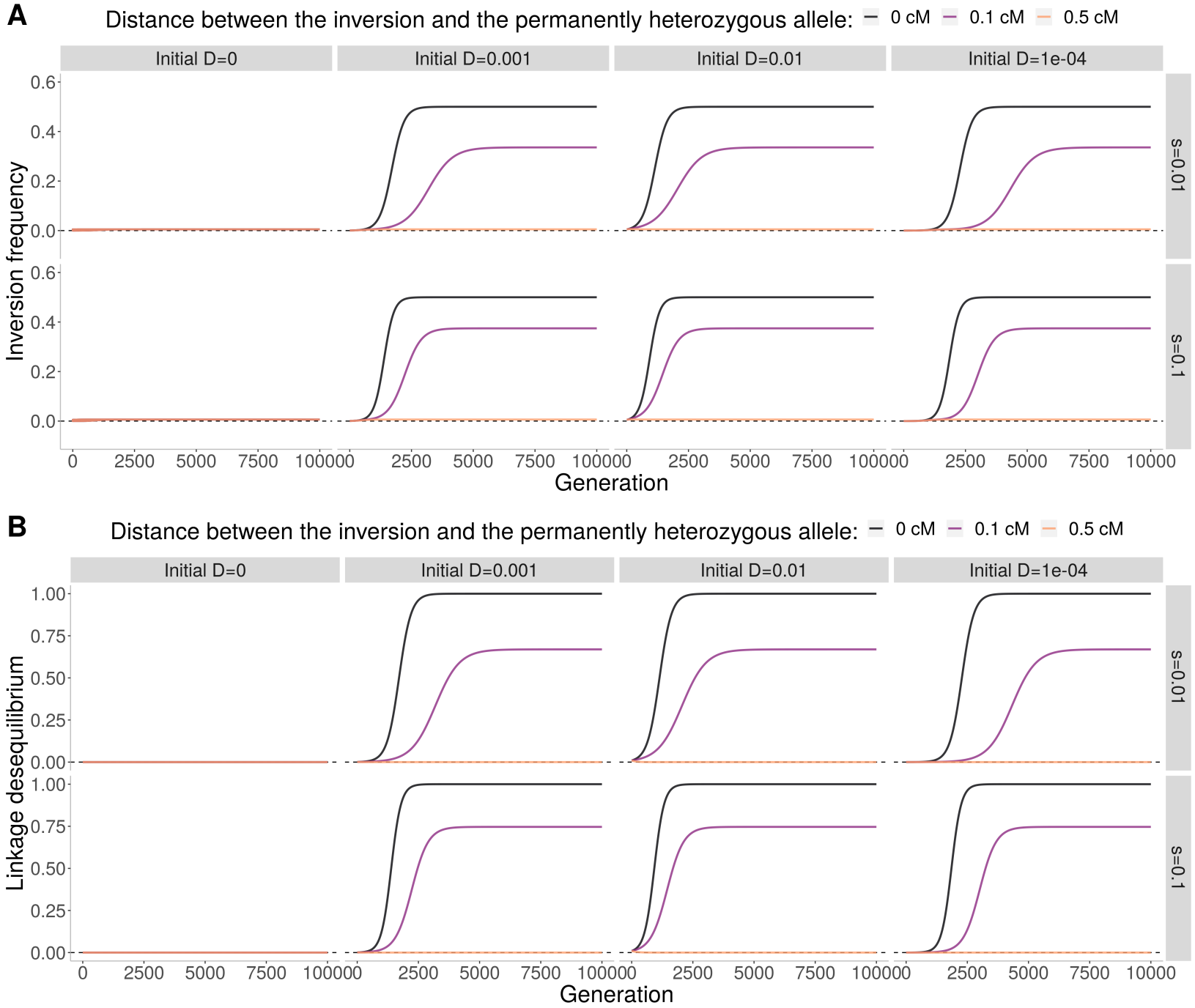
S3 Fig


a
Distance between the inversion and the permanently heterozygous allele:
0 cM
0.1 cM
1 cM
5 cM
50 cM
h=0.001
h=0.01
h=0.1
s=0.01
s=0.1
Inversion frequency
1.00
0.75
0.50
0.25
0.00
0
2500
5000
7500
10000
Generation
b
Distance between the inversion and the permanently heterozygous allele:
0 cM
0.1 cM
1 cM
5 cM
50 cM
h=0.001
h=0.01
h=0.1
s=0.01
s=0.1
Inversion frequency
1.00
0.75
0.50
0.25
0.00
0
2500
5000
7500
10000
Generation

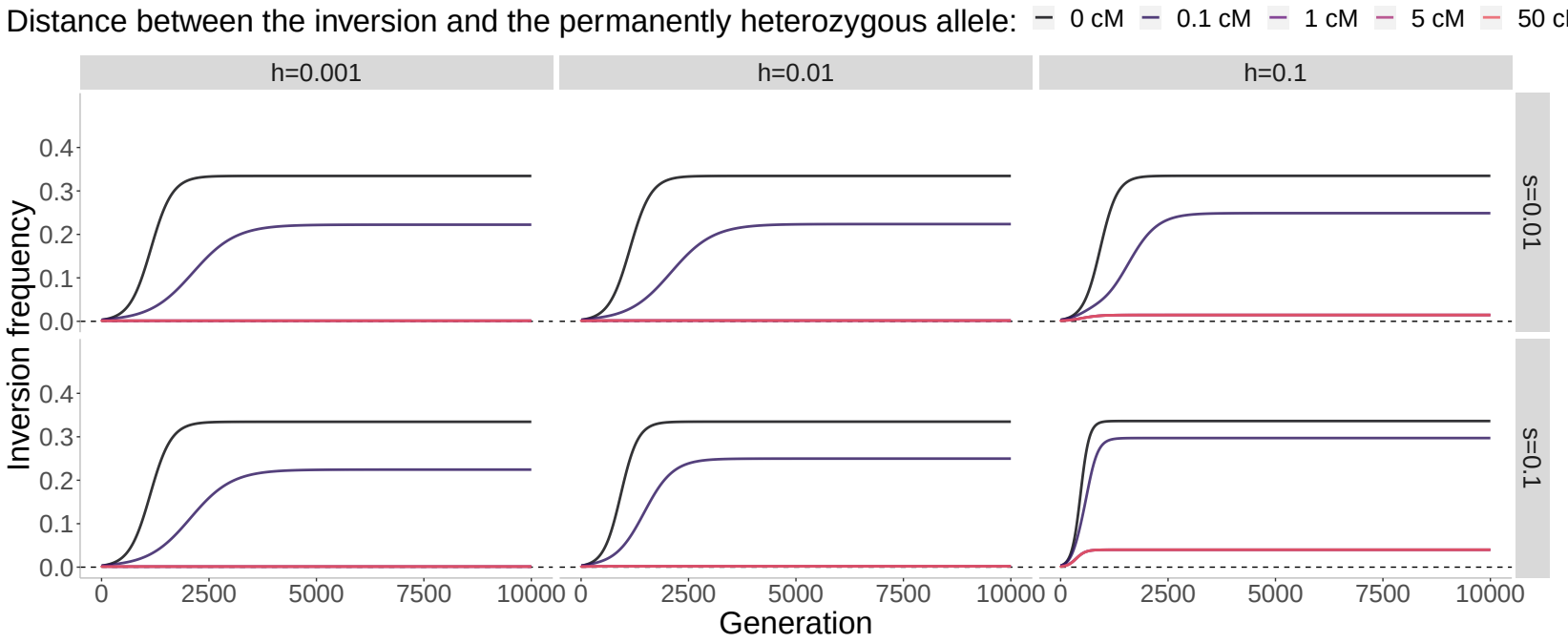
S4 Fig.


a Distance between the inversion and the permanently heterozygous allele: 0 cM 0.1 cM 1 cM 5 cM 50 cM
h=0.001
h=0.01
h=0.1
s=0.01
s=0.1
Inversion frequency
0.0
0.2
0.4
0.6
0
2500
5000
7500
10000
Generation
b Distance between the inversion and the permanently heterozygous allele: 0 cM 0.1 cM 1 cM 5 cM 50 cM
h=0.001
h=0.01
h=0.1
s=0.01
s=0.1
Inversion frequency
Generation
c Distance between the inversion and the permanently heterozygous allele: 0 cM 0.1 cM 1 cM 5 cM 50 cM
h=0.001
h=0.01
h=0.1
s=0.01
s=0.1
Inversion frequency
Generation

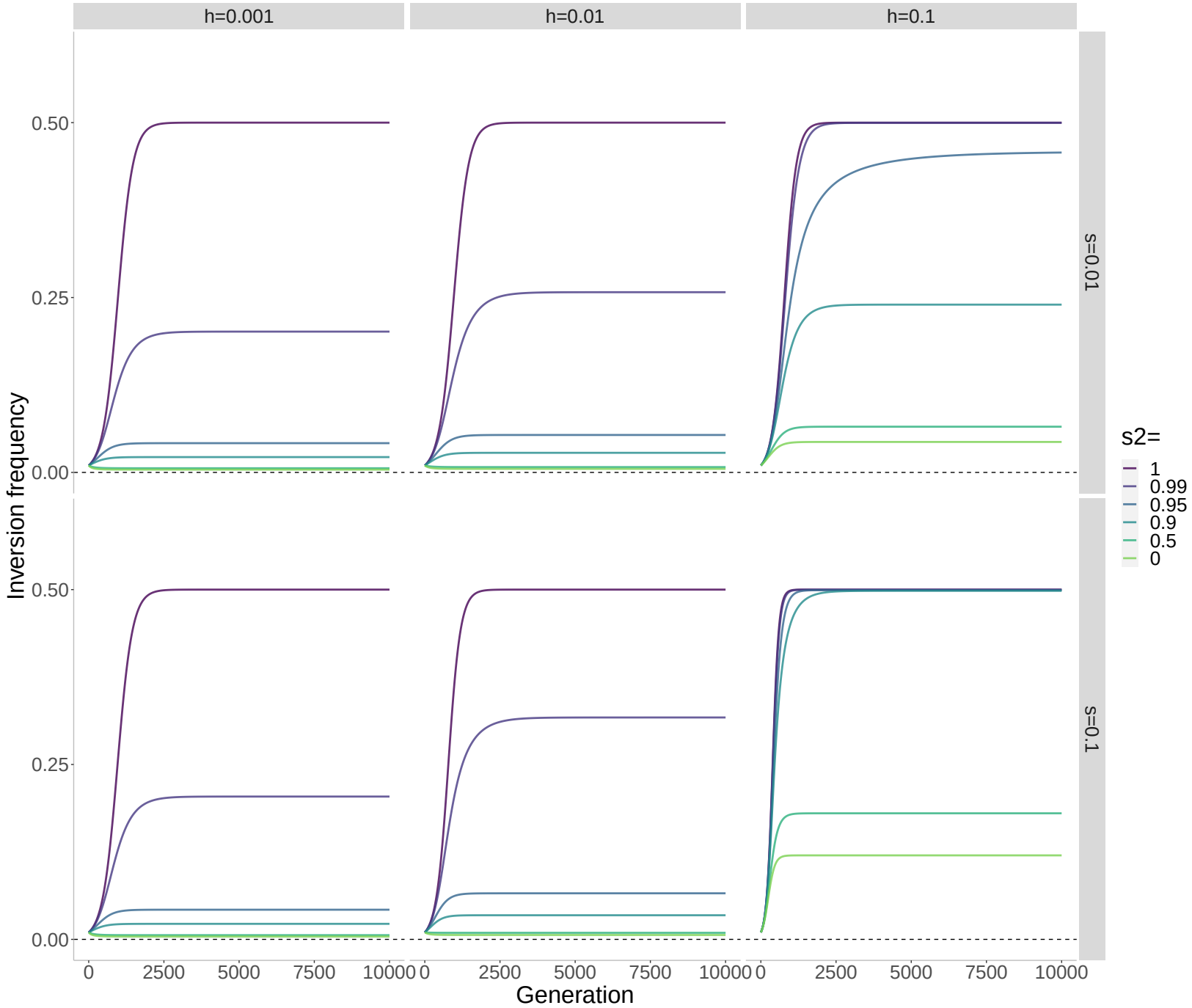
S5 Fig


A
Distance between the inversion and the permanently heterozygous allele:
0 cM
0.1 cM
0.5 cM
Initial D=0
Initial D=0.001
Initial D=0.01
Initial D=1e-04
s=0.01
s=0.1
Inversion frequency
0.0
0.2
0.4
0.6
0
2500
5000
7500
100000
10000
Generation
B
Distance between the inversion and the permanently heterozygous allele:
0 cM
0.1 cM
0.5 cM
Initial D=0
Initial D=0.001
Initial D=0.01
Initial D=1e-04
s=0.01
s=0.1
Linkage desequilibrium
0.00
0.25
0.50
0.75
1.00
0
2500
5000
7500
100000
10000
Generation

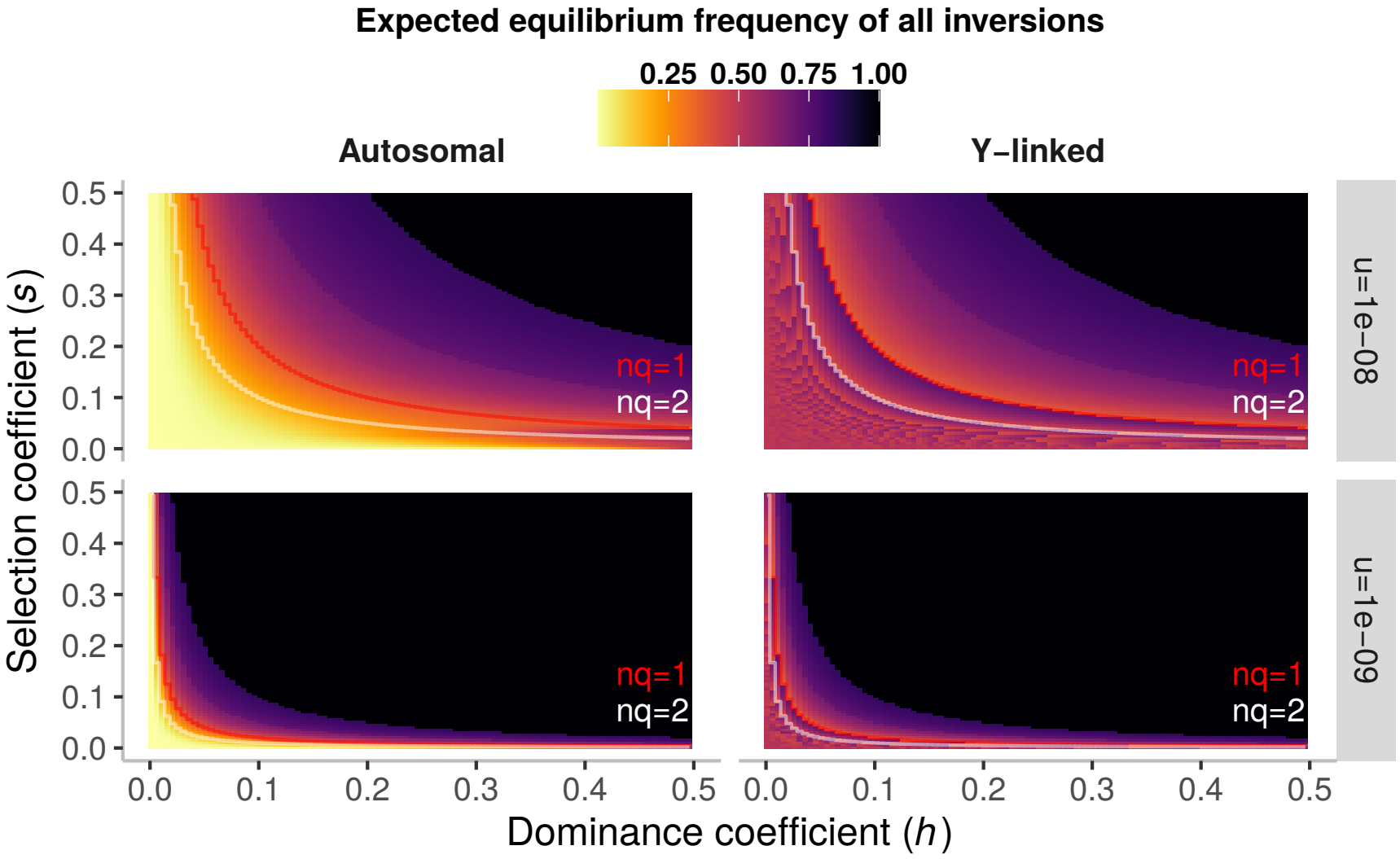
S6 Fig.


Distance between the inversion and the permanently heterozygous allele:
0 cM
0.1 cM
1 cM
5 cM
50 cM
h=0.001
h=0.01
h=0.1
s=0.01
s=0.1
Inversion frequency
0.0
0.1
0.2
0.3
0.4
0
2500
5000
7500
10000
Generation

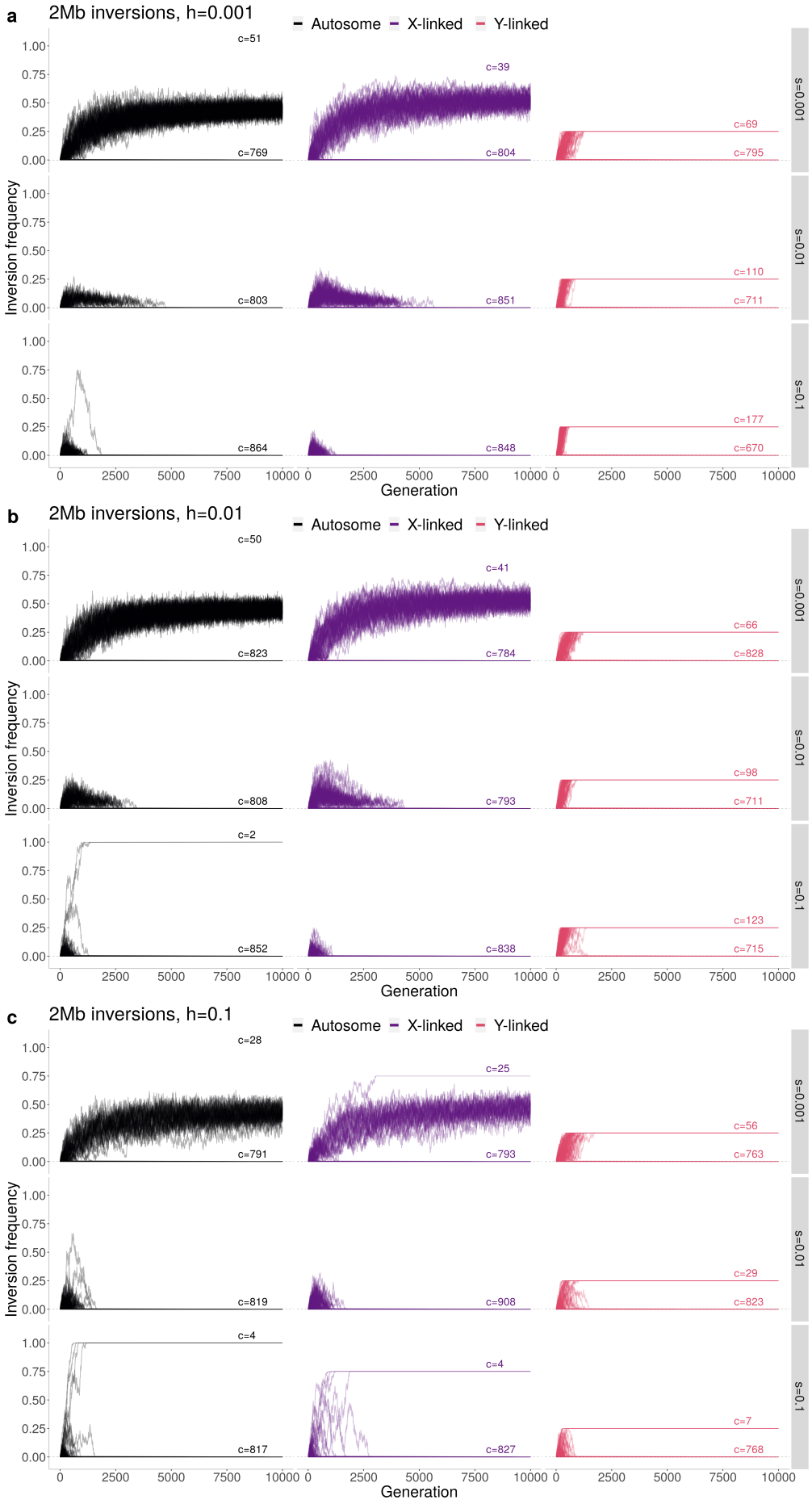
S7 Fig


h=0.001
h=0.01
h=0.1
s=0.01
s=0.1
Inversion frequency
0.00
0.25
0.50
0
2500
5000
7500
10000
Generation
s2=
1
0.99
0.95
0.9
0.5
0

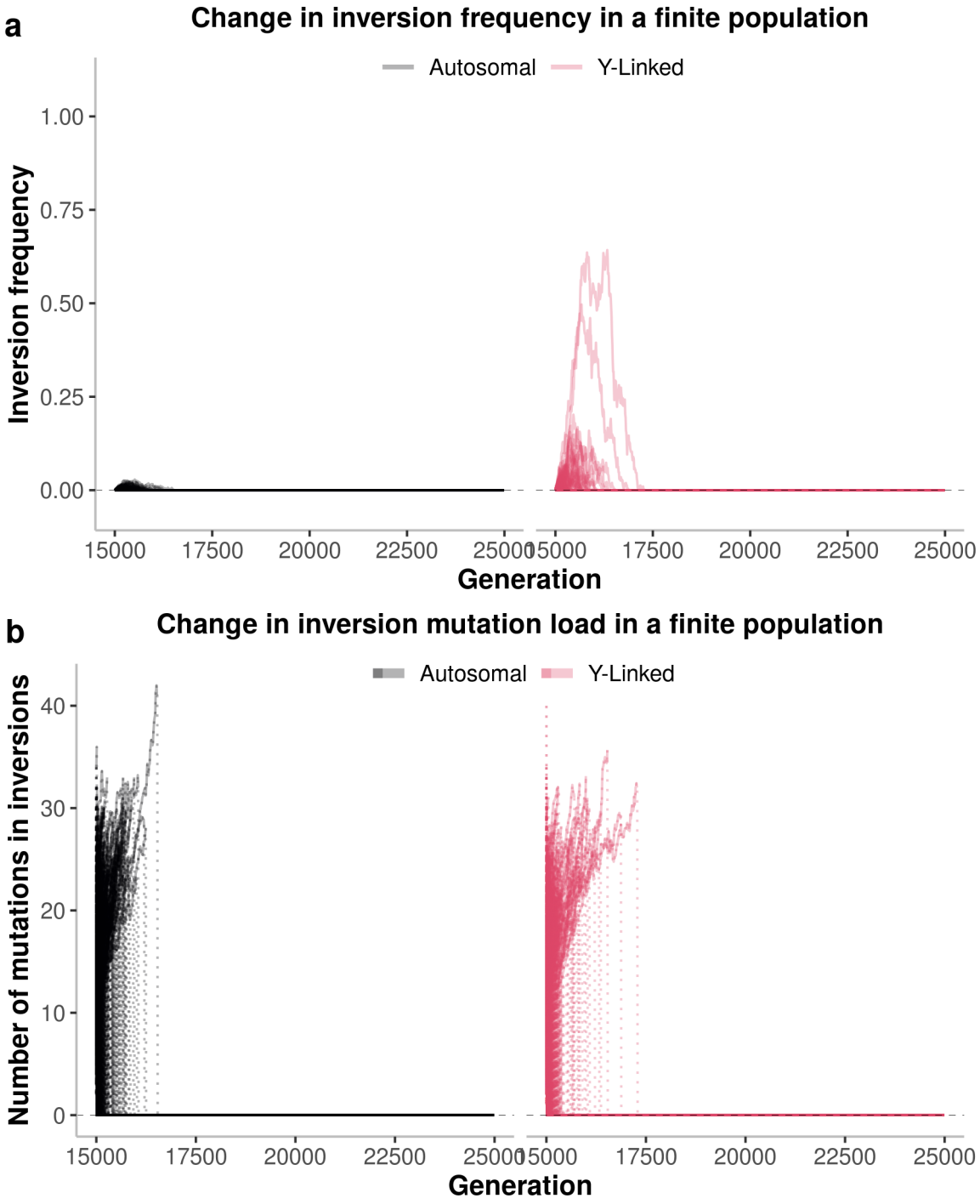
S8 Fig


Expected equilibrium frequency of all inversions
0.25 0.50 0.75 1.00
Autosomal
Y−linked
u=1e−08
u=1e−09
nq=1
nq=2
Selection coefficient (s)
Dominance coefficient (h)
0.0
0.1
0.2
0.3
0.4
0.5

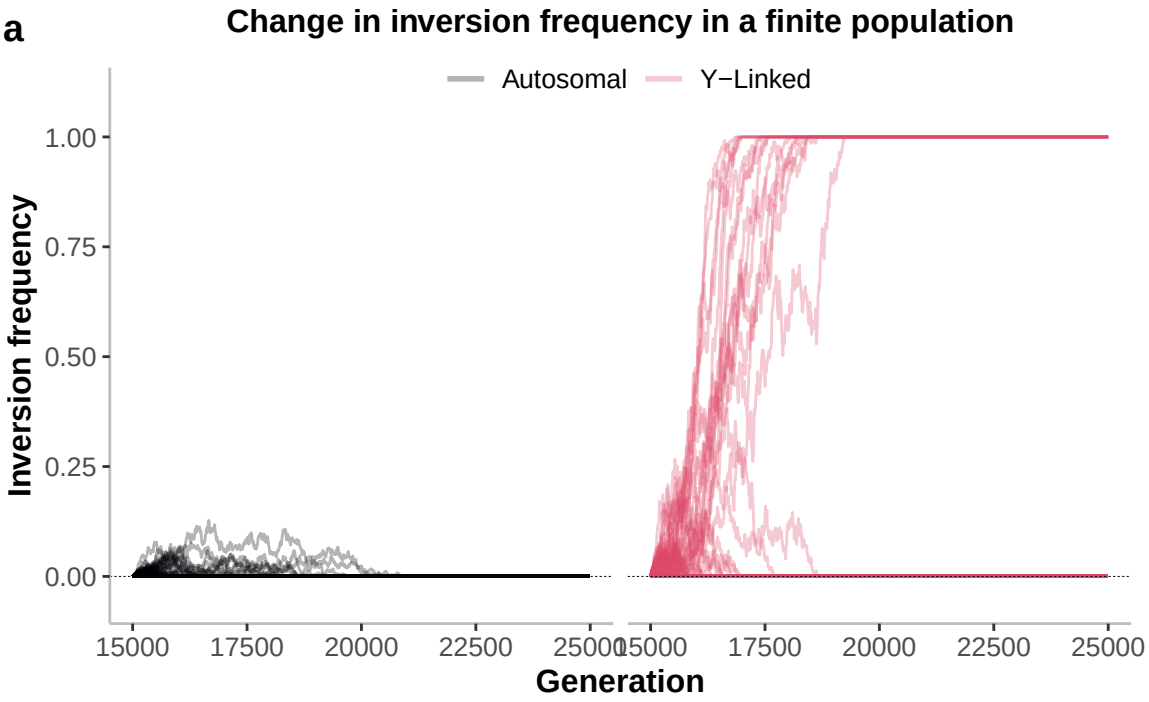
S9 Fig


a
2Mb inversions, h=0.001
Autosome
X-linked
Y-linked
c=51
c=39
c=69
c=769
c=804
c=795
c=803
c=851
c=110
c=711
c=864
c=848
c=177
c=670
b
2Mb inversions, h=0.01
c=50
c=41
c=66
c=823
c=784
c=828
c=808
c=793
c=98
c=711
c=2
c=852
c=838
c=123
c=715
c
2Mb inversions, h=0.1
c=28
c=25
c=56
c=791
c=793
c=763
c=819
c=908
c=29
c=823
c=4
c=4
c=7
c=817
c=827
c=768
Inversion frequency
Generation
s=0.001
s=0.01
s=0.1
0.00
0.25
0.50
0.75
1.00
0
2500
5000
7500
10000

S10 Fig

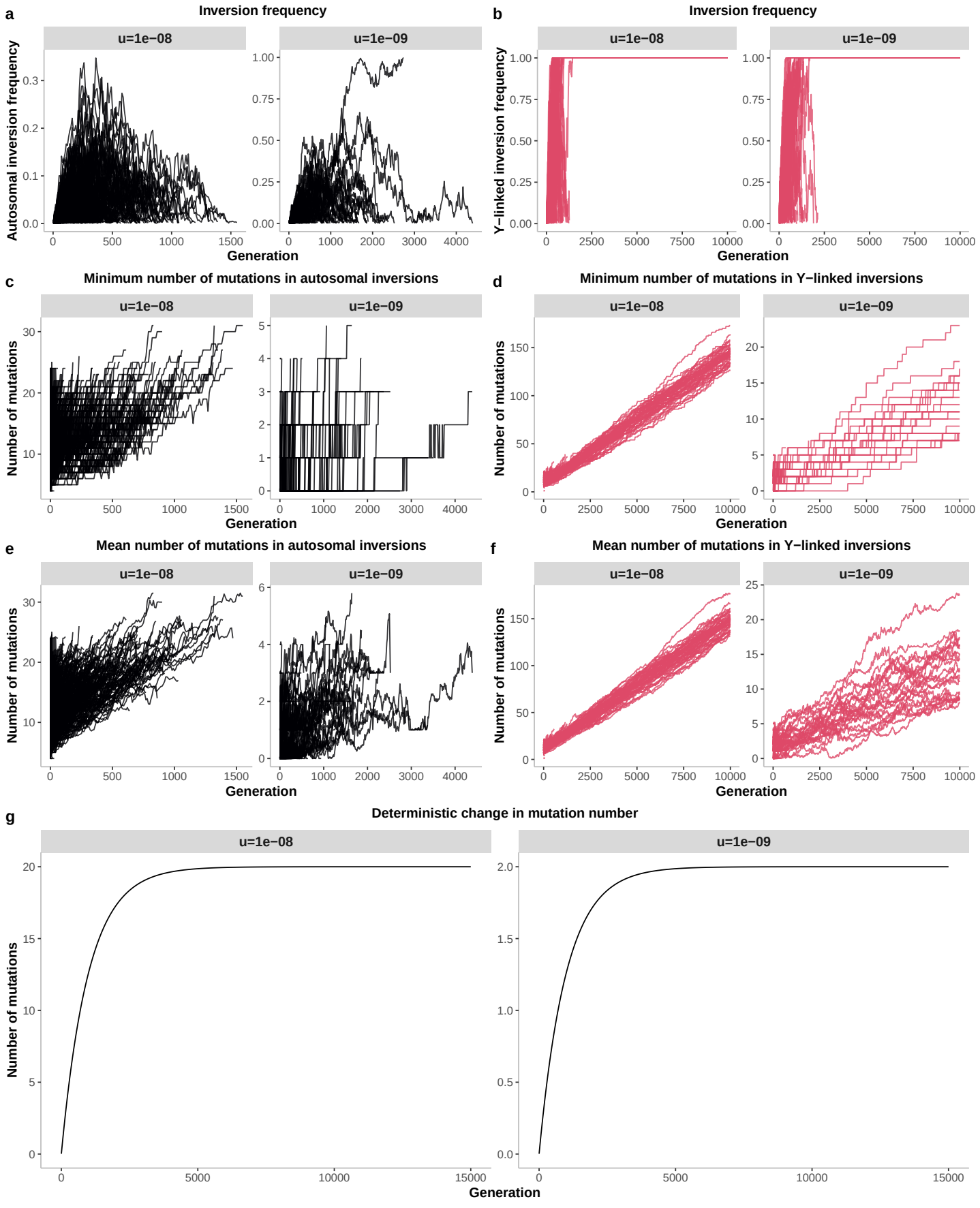

a
Change in inversion frequency in a finite population
Autosomal
Y-Linked
Inversion frequency
1.00
0.75
0.50
0.25
0.00
15000
17500
20000
22500
25000
Generation
b
Change in inversion mutation load in a finite population
Autosomal
Y-Linked
Number of mutations in inversions
40
30
20
10
0
15000
17500
20000
22500
25000
Generation

S11 Fig

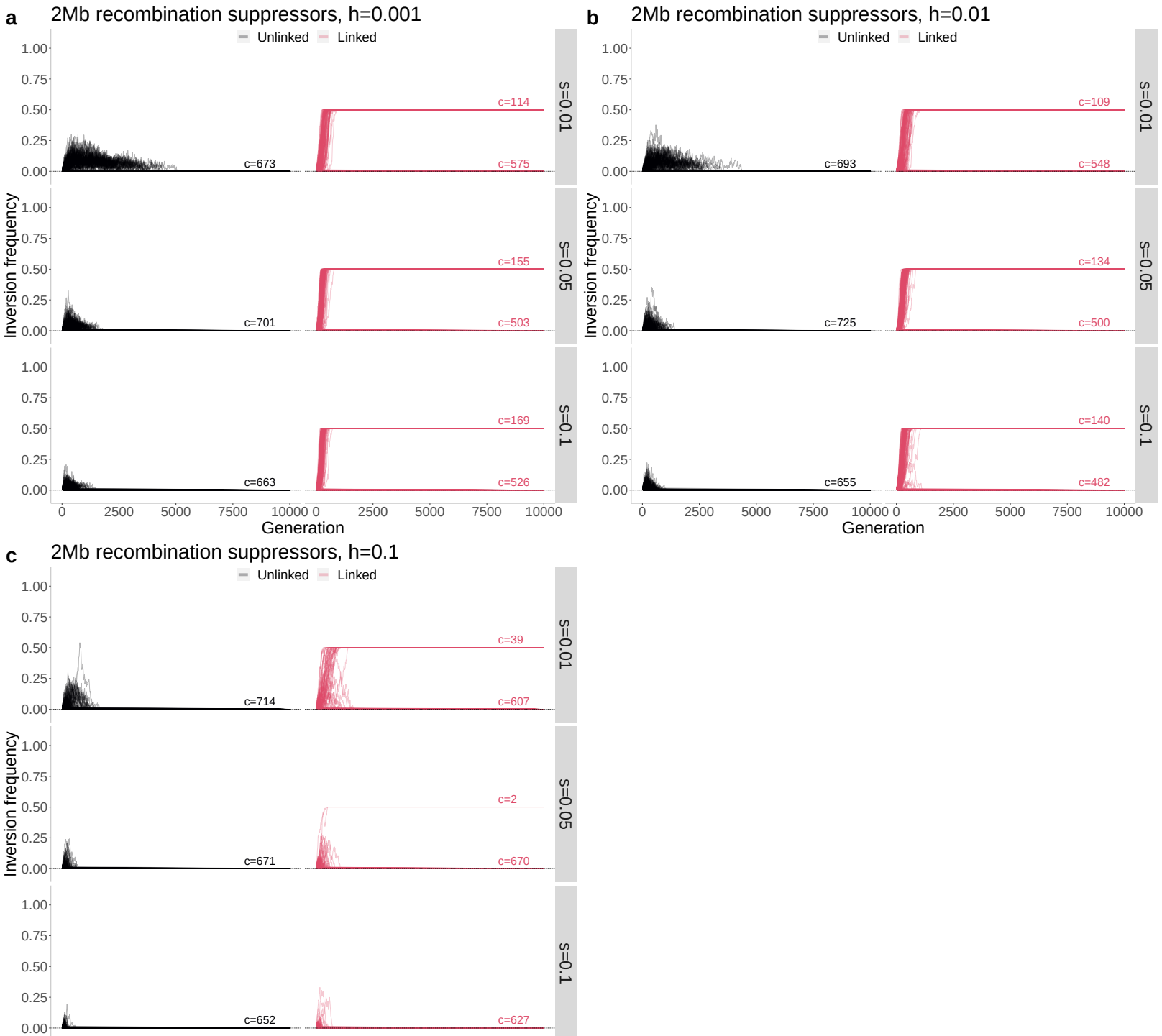

a
Change in inversion frequency in a finite population
Autosomal
Y−Linked
Inversion frequency
1.00
0.75
0.50
0.25
0.00
15000
17500
20000
22500
25000
Generation

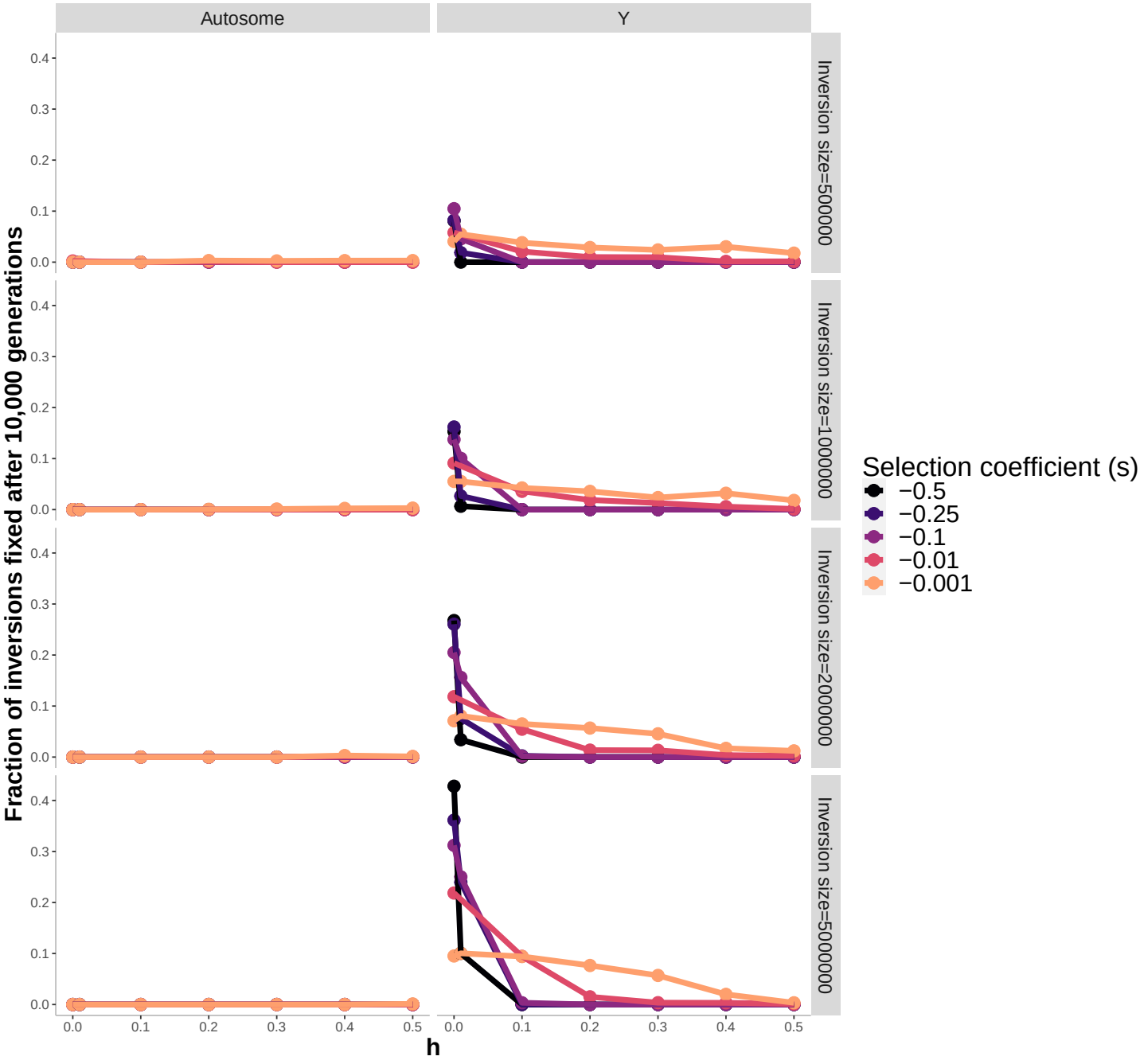

b
Change in inversion mutation load in a finite population
Autosomal
Y−Linked
Number of mutations in inversions
300
200
100
0
15000
17500
20000
22500
25000
Generation

S12 Fig

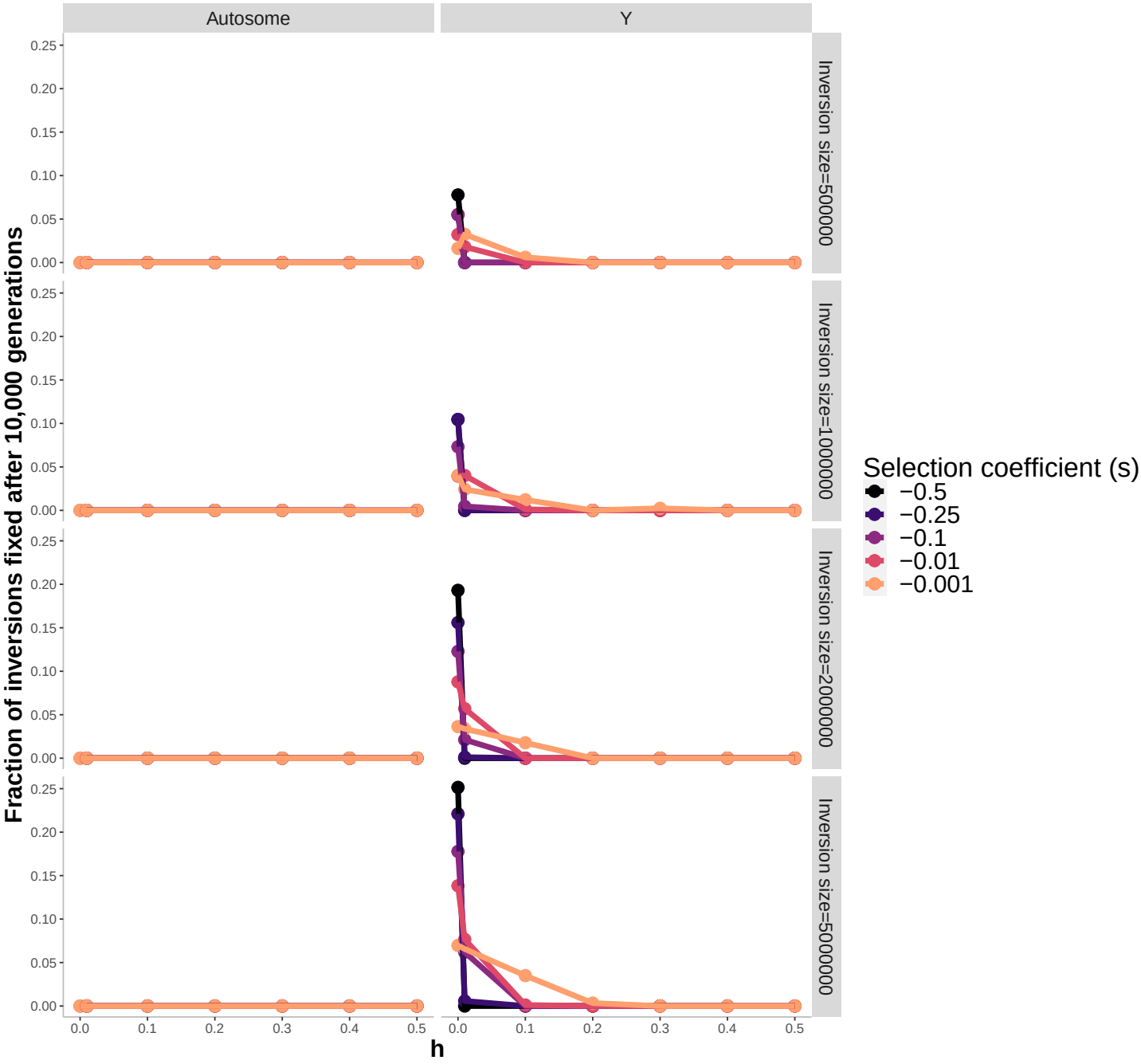

a
Inversion frequency
u=1e−08
u=1e−09
Autosomal inversion frequency
Generation
b
Inversion frequency
u=1e−08
u=1e−09
Y−linked inversion frequency
Generation
c
Minimum number of mutations in autosomal inversions
u=1e−08
u=1e−09
Number of mutations
Generation
d
Minimum number of mutations in Y−linked inversions
u=1e−08
u=1e−09
Number of mutations
Generation
e
Mean number of mutations in autosomal inversions
u=1e−08
u=1e−09
Number of mutations
Generation
f
Mean number of mutations in Y−linked inversions
u=1e−08
u=1e−09
Number of mutations
Generation
g
Deterministic change in mutation number
u=1e−08
u=1e−09
Number of mutations
Generation

S13 Fig

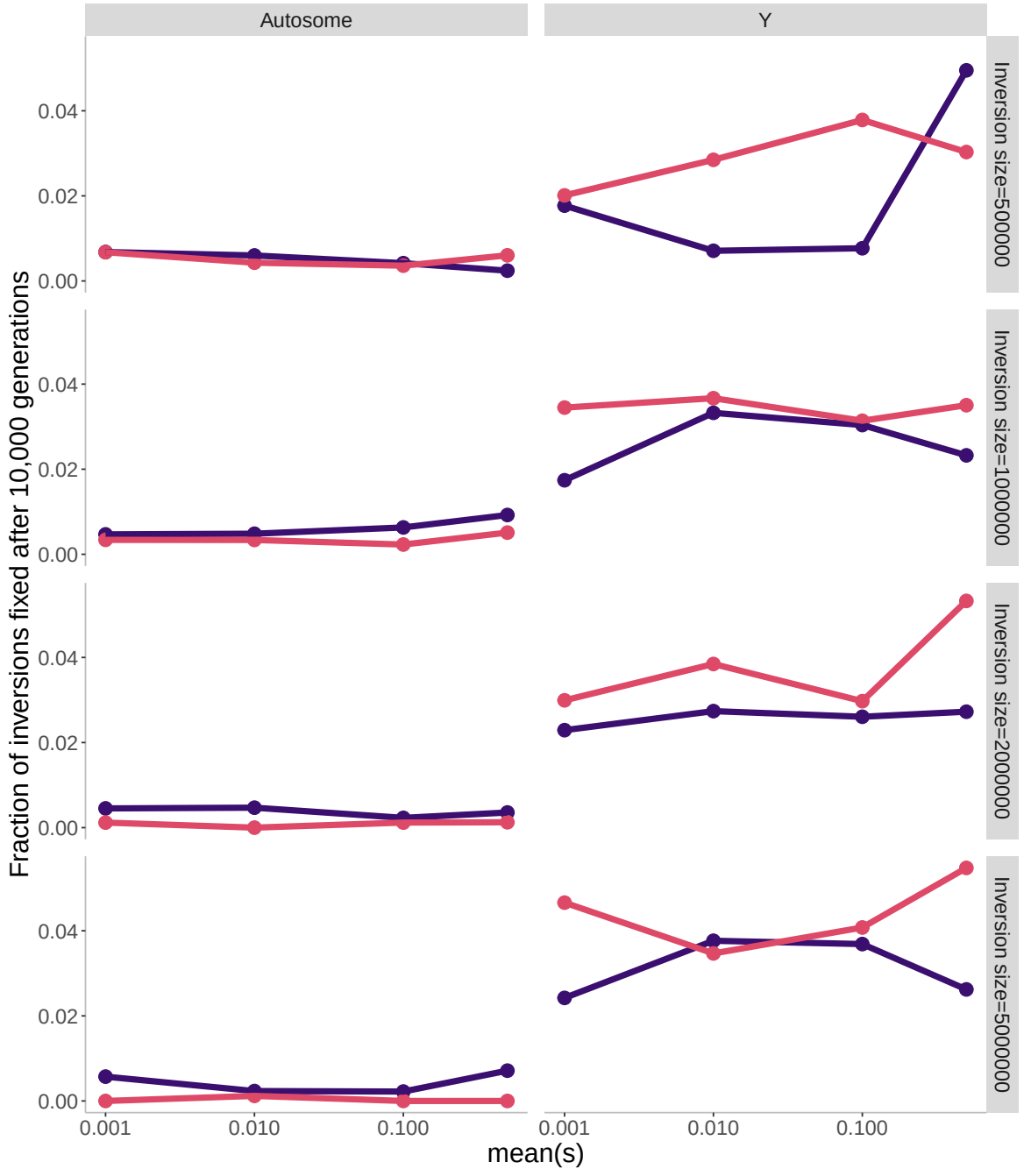

a
2Mb recombination suppressors, h=0.001
b
2Mb recombination suppressors, h=0.01
c
2Mb recombination suppressors, h=0.1
Unlinked
Linked
Inversion frequency
Generation
s=0.01
s=0.05
s=0.1
c=673
c=114
c=575
c=701
c=155
c=503
c=663
c=169
c=526
c=693
c=109
c=548
c=725
c=134
c=500
c=655
c=140
c=482
c=714
c=39
c=607
c=671
c=2
c=670
c=652
c=627


Autosome
Y
Inversion size=500000
Inversion size=1000000
Inversion size=2000000
Inversion size=5000000
Fraction of inversions fixed after 10,000 generations
h
0.0
0.1
0.2
0.3
0.4
0.5
Selection coefficient (s)
−0.5
−0.25
−0.1
−0.01
−0.001

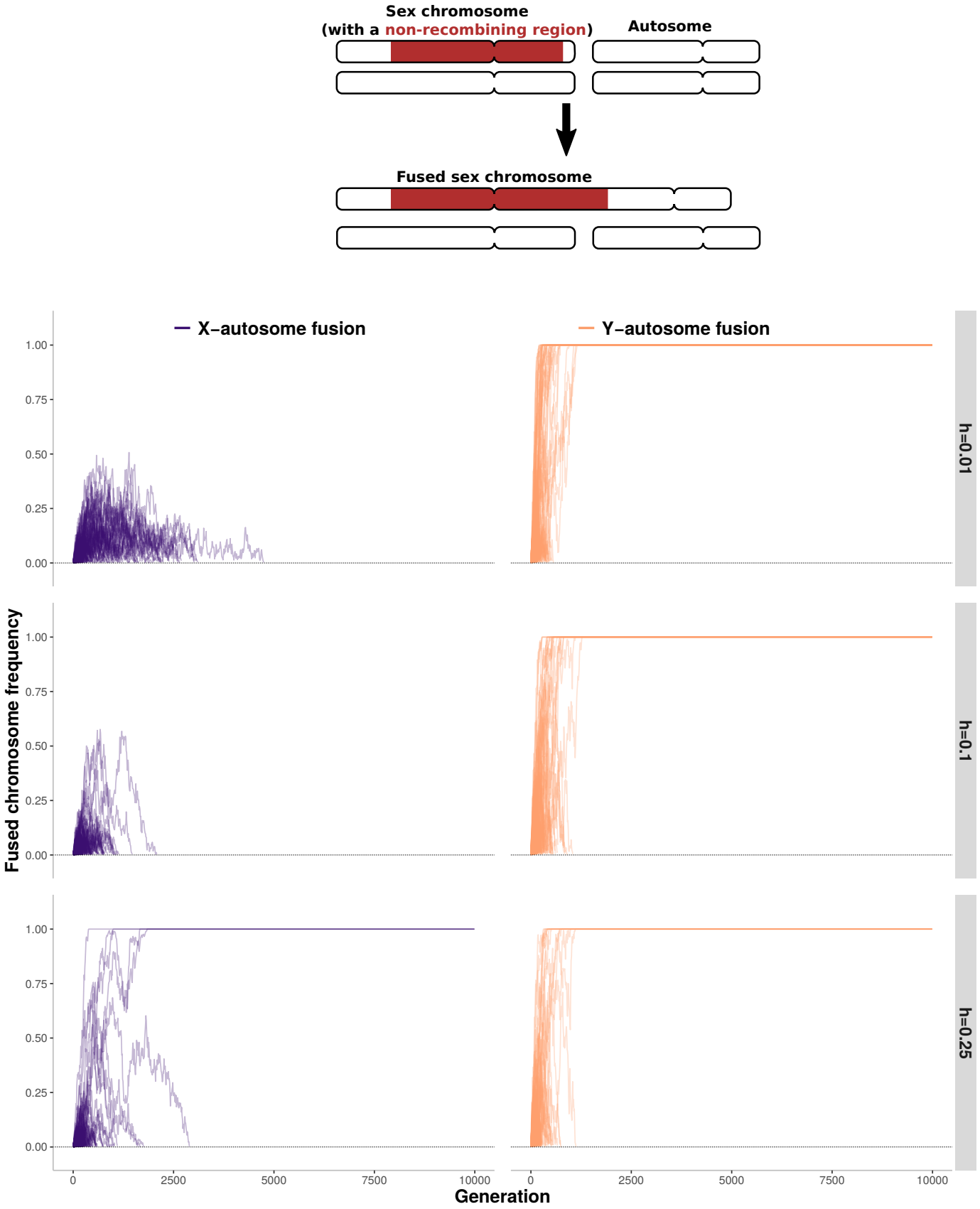
S15 Fig


Autosome
Y
Inversion size=500000
Inversion size=1000000
Inversion size=2000000
Inversion size=5000000
Fraction of inversions fixed after 10,000 generations
h
0.00
0.05
0.10
0.15
0.20
0.25
0.0
0.1
0.2
0.3
0.4
0.5
Selection coefficient (s)
−0.5
−0.25
−0.1
−0.01
−0.001

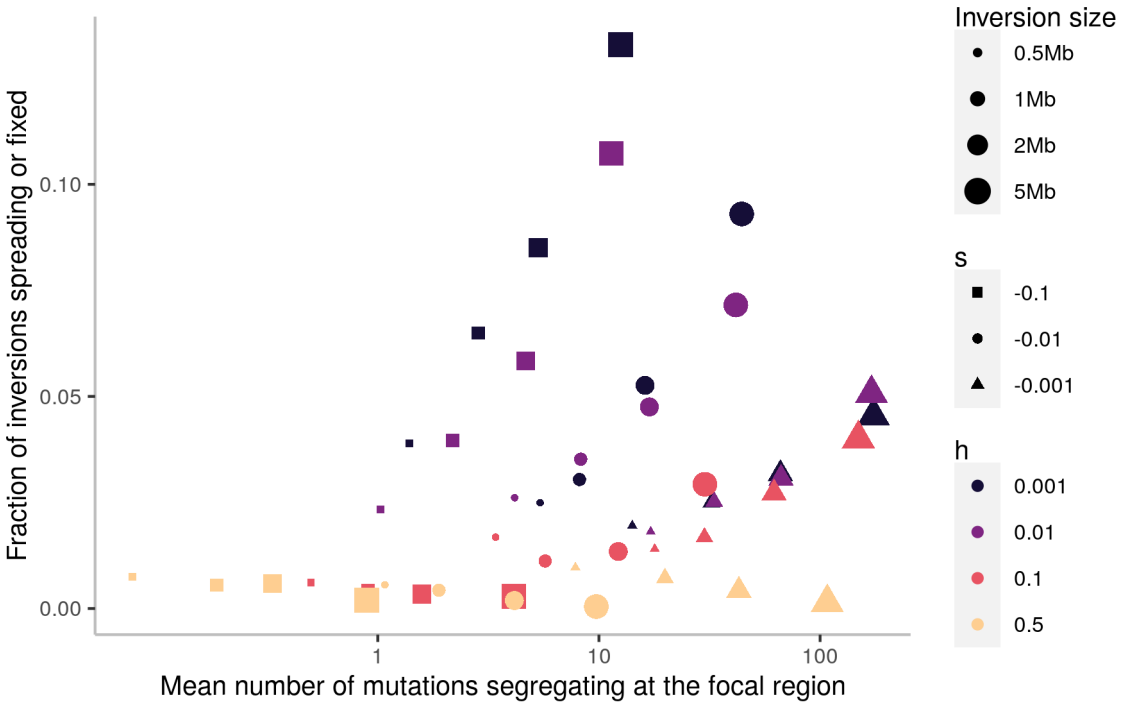
S16 Fig


Autosome
Y
Inversion size=500000
Inversion size=1000000
Inversion size=2000000
Inversion size=5000000
Fraction of inversions fixed after 10,000 generations
0.04
0.02
0.00
0.001
0.010
0.100
mean(s)
mutation rate (u)
1e−09
1e−08

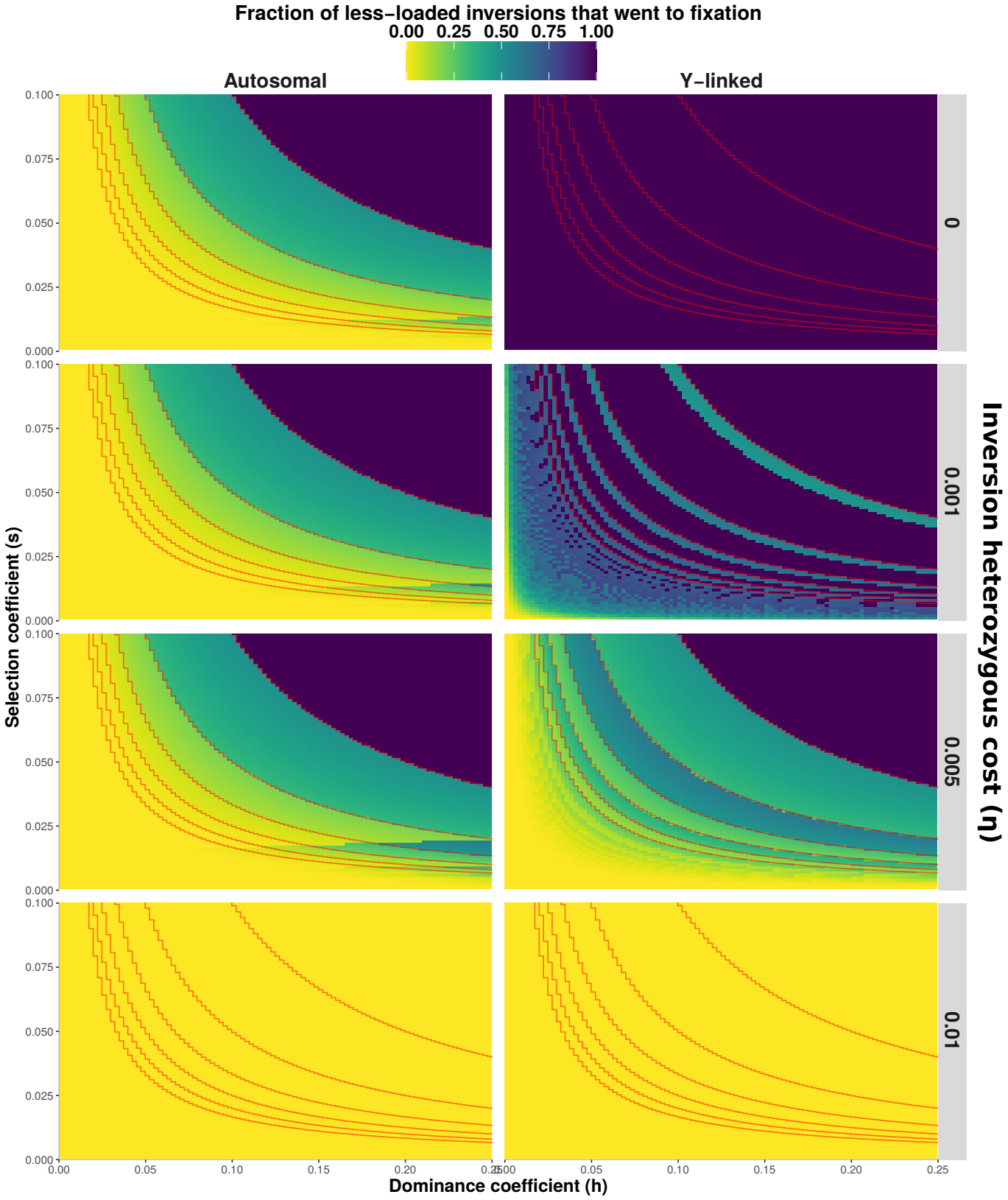
S17 Fig


Sex chromosome
(with a non-recombining region)
Autosome
Fused sex chromosome
X–autosome fusion
Y–autosome fusion
Fused chromosome frequency
1.00
0.75
0.50
0.25
0.00
h=0.01
h=0.1
h=0.25
0
2500
5000
7500
10000
Generation

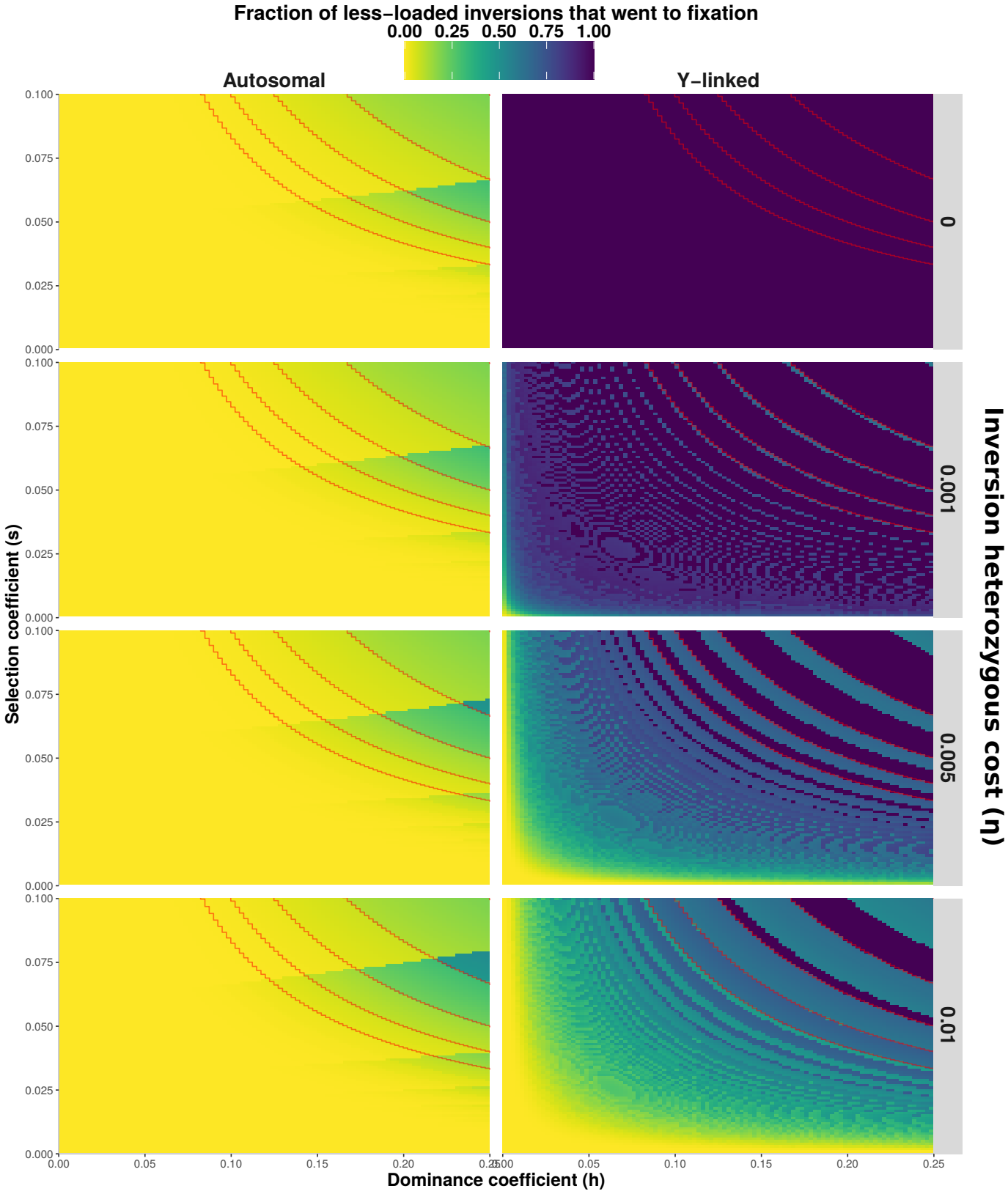
S18 Fig


Fraction of inversions spreading or fixed
0.10
0.05
0.00
1
10
100
Mean number of mutations segregating at the focal region
Inversion size
0.5Mb
1Mb
2Mb
5Mb
s
-0.1
-0.01
-0.001
h
0.001
0.01
0.1
0.5

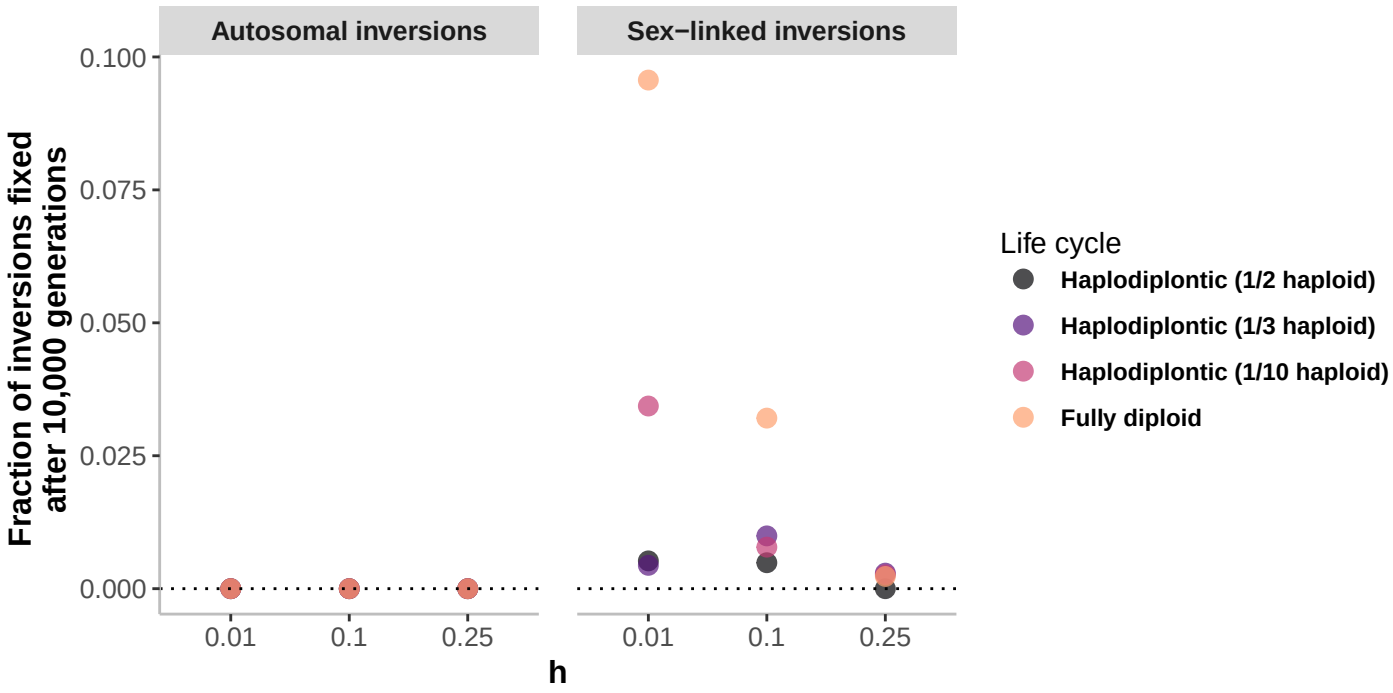
S19 Fig


Fraction of less−loaded inversions that went to fixation
0.00 0.25 0.50 0.75 1.00
Autosomal
Y−linked
Selection coefficient (s)
Dominance coefficient (h)
Inversion heterozygous cost (η)
0
0.001
0.005
0.01
0.000
0.025
0.050
0.075
0.100
0.00
0.05
0.10
0.15
0.20
0.25

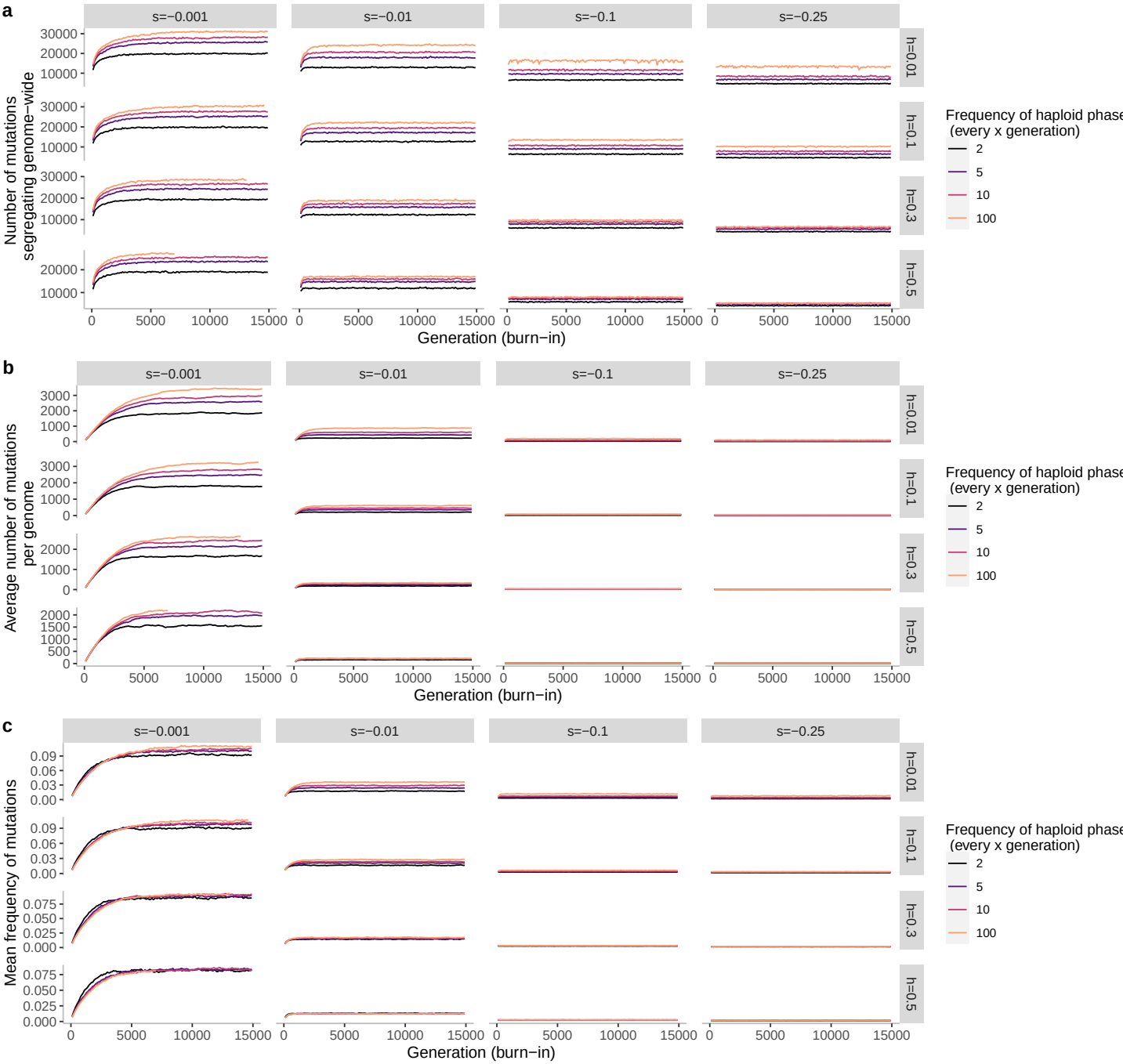
S20 Fig


Fraction of less−loaded inversions that went to fixation
0.00 0.25 0.50 0.75 1.00
Autosomal
Y−linked
Selection coefficient (s)
Dominance coefficient (h)
Inversion heterozygous cost (η)
0
0.001
0.005
0.01
0.000
0.025
0.050
0.075
0.100
0.00
0.05
0.10
0.15
0.20
0.25

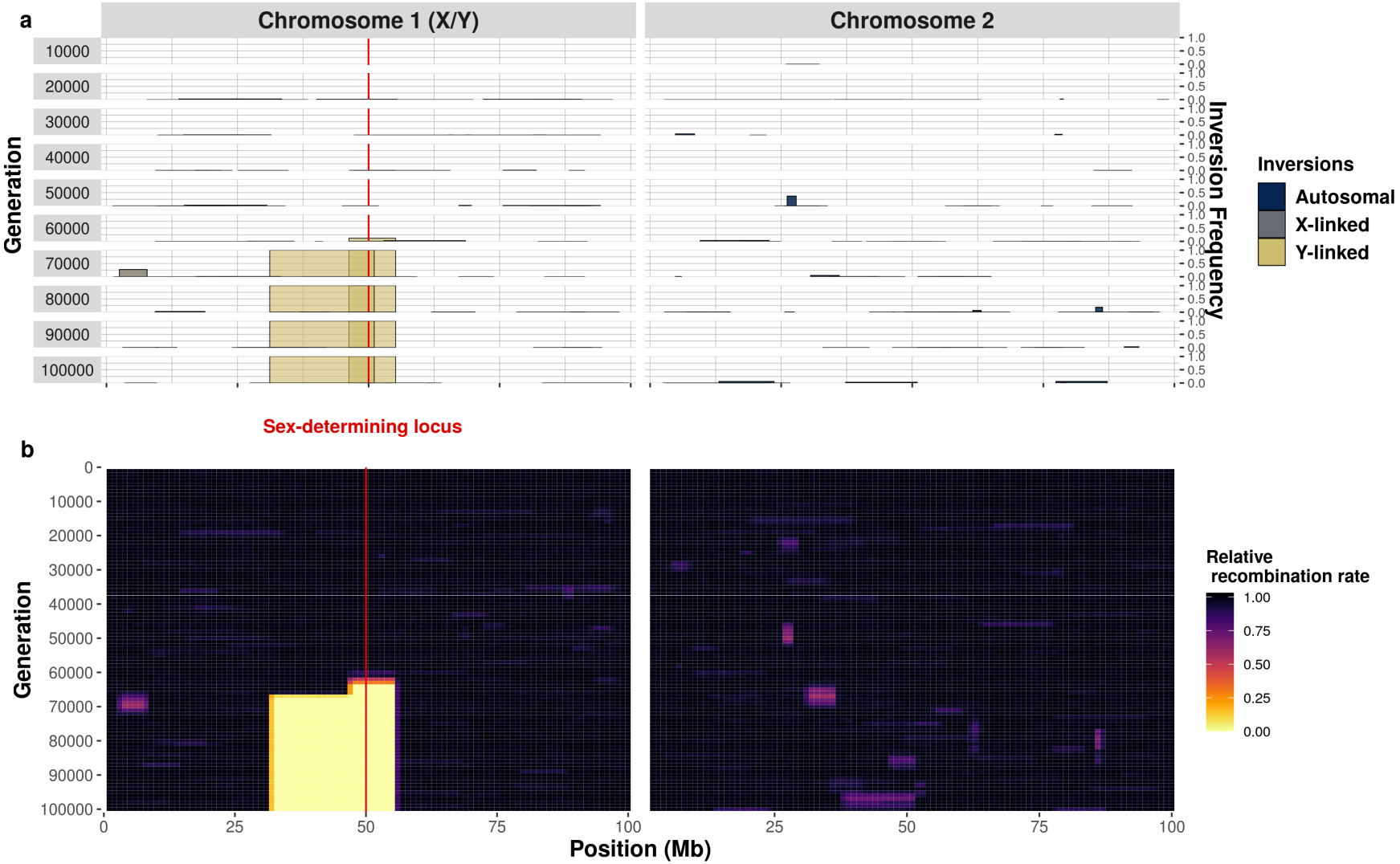
S21 Fig


Autosomal inversions
Sex−linked inversions
Fraction of inversions fixed after 10,000 generations
0.100
0.075
0.050
0.025
0.000
0.01
0.1
0.25
h
Life cycle
Haplodiplontic (1/2 haploid)
Haplodiplontic (1/3 haploid)
Haplodiplontic (1/10 haploid)
Fully diploid

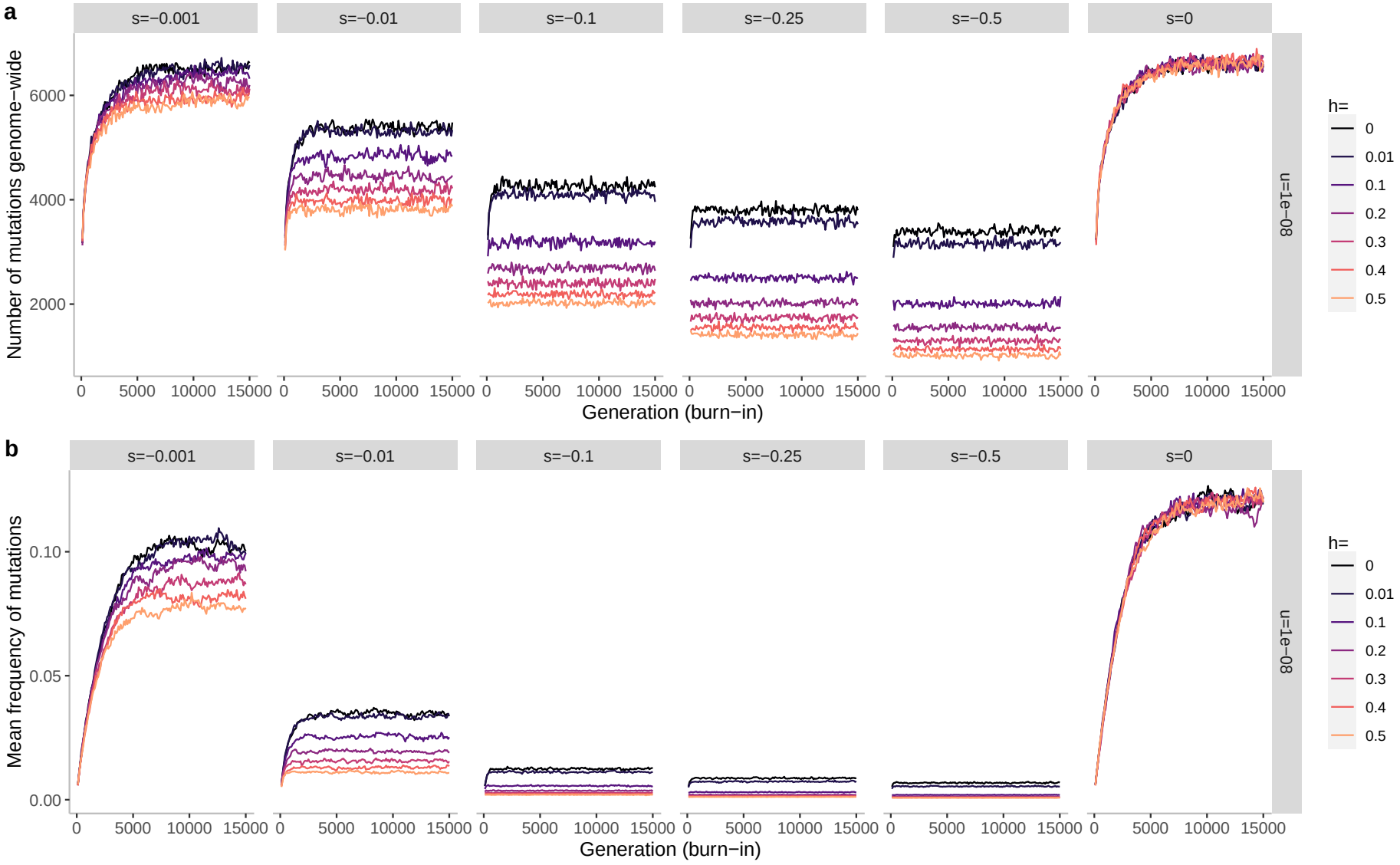
S22 Fig


a
s=−0.001
s=−0.01
s=−0.1
s=−0.25
h=0.01
h=0.1
h=0.3
h=0.5
Number of mutations segregating genome−wide
Generation (burn−in)
Frequency of haploid phase (every x generation)
2
5
10
100
b
Average number of mutations per genome
c
Mean frequency of mutations

S23 Fig


a
Chromosome 1 (X/Y)
Chromosome 2
Generation
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
Inversion Frequency
1.0
0.5
0.0
Inversions
Autosomal
X-linked
Y-linked
Sex-determining locus
b
Generation
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
0
25
50
75
100
Position (Mb)
Relative recombination rate
1.00
0.75
0.50
0.25
0.00

S24 Fig


a
s=−0.001
s=−0.01
s=−0.1
s=−0.25
s=−0.5
s=0
Number of mutations genome−wide
6000
4000
2000
0
5000
10000
15000
Generation (burn−in)
u=1e−08
h=
0
0.01
0.1
0.2
0.3
0.4
0.5
b
Mean frequency of mutations
0.10
0.05
0.00

S25 Fig

Chapitre 2

The fate of recessive deleterious or overdominant mutations near mating-type loci under partial selfing

Ce chapitre a fait l'objet d'une recommendation PCI EvolBiol (https://doi.org/10.24072/pci.evolbiol.100635) et d'une publication dans le Peer Community Journal : Tezenas et al., 2022.

Cet article est le fruit d'un travail commencé lors de mon stage de M2 et terminé pendant ma thèse, sous la direction de Sylvain Billiard, Amandine Véber, et Tatiana Giraud. Le principal objectif était d'étudier la dynamique de mutations délétères au voisinage d'un locus en permanence hétérozygote, et dans des populations d'individus pouvant se reproduire par auto-fécondation. En effet, ce mode de reproduction est très fréquent chez le champignon *Microbotryum violaceum*, chez qui on observe une extension successive de suppression de recombinaison sur les chromosomes de type sexuel. Nous nous sommes inspirés du travail présenté dans Antonovics et Abrams, 2004, en adoptant un modèle stochastique. On ne considère pas de locus modifieur de recombinaison, mais on implémente un mode de reproduction de plus (l'autofécondation inter-tétrade).

Le deuxième chapitre de ce manuscrit est constitué de l'article publié tel qu'il peut être trouvé en ligne.

Peer Community Journal

Section: Evolutionary Biology

RESEARCH ARTICLE

Published
2023-01-27

Cite as
Emilie Tezenas, Tatiana Giraud, Amandine Véber and Sylvain Billiard (2023) *The fate of recessive deleterious or overdominant mutations near mating-type loci under partial selfing*, Peer Community Journal, 3: e14.

Correspondence
emilie.tezenas-du-montcel@ens-rennes.fr

Peer-review
Peer reviewed and recommended by PCI Evolutionary Biology, https://doi.org/10.24072/pci.evolbiol.100635

The fate of recessive deleterious or overdominant mutations near mating-type loci under partial selfing

Emilie Tezenas[1,2,3], Tatiana Giraud[#,1], Amandine Véber[#,3], and Sylvain Billiard[#,2]

Volume **3** (2023), article e14

https://doi.org/10.24072/pcjournal.238

Abstract

Large regions of suppressed recombination having extended over time occur in many organisms around genes involved in mating compatibility (sex-determining or mating-type genes). The sheltering of deleterious alleles has been proposed to be involved in such expansions. However, the dynamics of deleterious mutations partially linked to genes involved in mating compatibility are not well understood, especially in finite populations. In particular, under what conditions deleterious mutations are likely to be maintained for long enough near mating-compatibility genes remains to be evaluated, especially under selfing, which generally increases the purging rate of deleterious mutations. Using a branching process approximation, we studied the fate of a new deleterious or overdominant mutation in a diploid population, considering a locus carrying two permanently heterozygous mating-type alleles, and a partially linked locus at which the mutation appears. We obtained analytical and numerical results on the probability and purging time of the new mutation. We investigated the impact of recombination between the two loci and of the mating system (outcrossing, intra and inter-tetrad selfing) on the maintenance of the mutation. We found that the presence of a fungal-like mating-type locus (*i.e.* not preventing diploid selfing) always sheltered the mutation under selfing, *i.e.* it decreased the purging probability and increased the purging time of the mutations. The sheltering effect was higher in case of automixis (intra-tetrad selfing). This may contribute to explain why evolutionary strata of recombination suppression near the mating-type locus are found mostly in automictic (pseudo-homothallic) fungi. We also showed that rare events of deleterious mutation maintenance during strikingly long evolutionary times could occur, suggesting that deleterious mutations can indeed accumulate near the mating-type locus over evolutionary time scales. In conclusion, our results show that, although selfing purges deleterious mutations, these mutations can be maintained for very long times near a mating-type locus, which may contribute to promote the evolution of recombination suppression in sex-related chromosomes.

[1]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systematique et Evolution, 91190, Gif-sur-Yvette, France, [2]Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France, [3]Université Paris Cité, CNRS, MAP 5, F-75006 Paris, [#]Equal contribution

Contents

1. Introduction

The evolution of sex chromosomes, and more generally of genomic regions lacking recombination, is widely studied in evolutionary biology as it raises multiple, unresolved questions (Ironside, 2010, Yan et al., 2020, Hartmann, Ament-Velásquez, et al., 2021, Kratochvíl and Stöck, 2021, Jay et al., 2022). A striking feature of many sex and mating-type chromosomes is the absence of recombination in large regions around the sex-determining genes. Recombination suppression indeed evolved in various groups of plants and animals in several steps beyond the sex-determining genes, generating evolutionary strata of differentiation between sex chromosomes (Nicolas et al., 2004, Bergero and Charlesworth, 2009, Hartmann, Duhamel, et al., 2021, Kratochvíl and Stöck, 2021). The reasons for the gradual expansion of recombination cessation beyond sex-determining genes remain debated (Ironside, 2010, A. E. Wright et al., 2016, Ponnikas et al., 2018, Hartmann, Duhamel, et al., 2021). Recombination suppression has extended

progressively with time not only on many sex chromosomes but also on mating-type chromosomes in fungi (Hartmann, Duhamel, et al., 2021) and other supergenes (Yan et al., 2020, Jay et al., 2021).

The main hypothesis to explain such stepwise extension of recombination cessation on sex chromosomes has long been sexual antagonism (D. Charlesworth et al., 2005, Bergero and Charlesworth, 2009). Theoretical studies have indeed shown that the suppression of recombination may evolve to link alleles that are beneficial in only one sex to the sex-determining genes (W. R. Rice, 1987, D. Charlesworth et al., 2005, Ruzicka et al., 2020). However, this hypothesis has received little evidence from empirical studies despite decades of research (Ironside, 2010, Dagilis et al., 2022). Moreover, the sexual antagonism hypothesis cannot explain the evolutionary strata found on fungal mating-type chromosomes. Indeed, in many fungi, two gametes can form a new individual only if they carry different mating types, but there is no sexual antagonism or other form of antagonistic selection between cells of opposite mating types; the cells of different mating types do not show contrasted phenotypes or footprints of diversifying selection (Bazzicalupo et al., 2019). Yet, evolutionary strata have been documented on the mating-type chromosomes of multiple fungi, with recombination suppression extending stepwise beyond mating-type determining genes (Fraser et al., 2004, Menkis et al., 2008, Branco et al., 2017, Branco et al., 2018, Hartmann, Duhamel, et al., 2021, Hartmann, Ament-Velásquez, et al., 2021, Vittorelli et al., 2023). Evolutionary strata have also been reported around other supergenes, *i.e.*, large genomic regions encompassing multiple genes linked by recombination suppression, such as in ants and butterflies (Yan et al., 2020, Jay et al., 2021). Several hypotheses alternative to sexual antagonism have been proposed and explored to explain the stepwise extension of recombination suppression on sex-related chromosomes (Ironside, 2010, Hartmann, Duhamel, et al., 2021). Theoretical models suggested that recombination suppression could be induced by a divergence increase in regions in linkage disequilibrium with a sex-determining locus (Jeffries et al., 2021) or that inversions could be stabilized by dosage compensation on asymmetric XY-like sex chromosomes (Lenormand and Roze, 2022).

A promising, widely applicable hypothesis is the sheltering of deleterious alleles by inversions carrying a lower load than average in the population (B. Charlesworth and Wall, 1999, Antonovics and Abrams, 2004, Hartmann, Duhamel, et al., 2021, Jay et al., 2022). Inversions (or any suppressor of recombination in *cis*) can indeed behave as overdominant: inversions with fewer recessive deleterious mutations than average are initially beneficial and increase in frequency, but can then occur in a homozygous state where they express their load, unless they are linked to a permanently heterozygous allele. In this case, they remain advantageous, and can reach fixation in the sex-related chromosome on which they appeared (Jay et al., 2022). The suppression of recombination is thereby selected for, and recessive deleterious mutations are permanently sheltered. The process can occur repeatedly, leading to evolutionary strata. Importantly, this is one of the few hypotheses able to explain the existence of evolutionary strata on fungal mating-type chromosomes and it can apply to any supergene with a permanently heterozygous allele (Llaurens et al., 2017, Jay et al., 2022).

A key point for the recombination suppressor to invade is that it must appear in populations where recessive deleterious mutations segregate near the mating-compatibility genes (Olito et al., 2022, Jay et al., 2022). We therefore need to understand whether such mutations can persist in the vicinity of permanently heterozygous alleles (such as those occurring at mating-type loci) and under what conditions. In particular, it is usually considered that selfing purges deleterious mutations (Glémin, 2007, Abu Awad and Billiard, 2017), while most evolutionary strata on fungal mating-type chromosomes have been reported in selfing (automictic) fungi (Branco et al., 2017, Branco et al., 2018, Hartmann, Ament-Velásquez, et al., 2021, Vittorelli et al., 2023). Indeed, because mating types are determined at the haploid stage in fungi, mating types do not prevent selfing when considering diploid individuals (Billiard et al., 2012). Some particular forms of selfing associated with a permanently heterozygous mating-type locus such as intra-tetrad mating (*i.e.* automixis, mating among gametes from the same meiosis) can however favor the maintenance of heterozygosity (Hood and Antonovics, 2000). Indeed, mating can only occur between haploid cells carrying different mating-type alleles, which maintains heterozygosity at the mating-type

locus, and to some extent at flanking regions, thereby possibly sheltering deleterious alleles. We therefore need to study whether deleterious or overdominant mutations can be maintained near mating-type compatibility loci, even under selfing, to assess whether the mechanism of sheltering deleterious mutations can drive extensions of recombination suppression.

The dynamics of deleterious mutation frequencies in genomes have been extensively studied independently of the presence of a permanently heterozygous locus. Deterministic models and diffusion approximations have been used to study the dynamics of deleterious mutations in a one locus-two allele setup (Kimura, 1980, Ewens, 2004, S. H. Rice, 2004), with the addition of sexual reproduction and in particular selfing (Ohta and Cockerham, 1974, Caballero and Hill, 1992, Abu Awad and Roze, 2018). Extensions of these models exist to cover the two locus-two allele case (Karlin, 1975) and multilocus systems (reviewed in Bürger, 2020), or to take stochastic fluctuations into account (Coron et al., 2013, Coron, 2014). However, the dynamics of deleterious mutations in genomic regions near a permanently heterozygous allele have been little studied. A deterministic model showed that a lethal allele can be sheltered in an outcrossing population only when it is completely linked to a self-incompatibility locus (Leach et al., 1986). Another deterministic model introduced selfing and showed with simulations that a lethal allele can be sheltered when it is completely linked to a mating-type allele, favored in a heterozygote state, and if there is intra-tetrad selfing (Antonovics et al., 1998). Assuming a variable recombination rate between the two loci, Antonovics and Abrams, 2004 showed that an overdominant allele lethal in a homozygous state could be maintained if recombination was twice as low as the selection for heterozygotes and mating occurred via intra-tetrad selfing. Stochastic simulations additionnally showed that a recessive deleterious allele could be maintained completely linked to a self-incompatibility allele, especially when it is highly recessive, and when the number of self-incompatibility alleles in the population is large (Llaurens et al., 2009), and that codominant weakly deleterious alleles could be maintained near loci under balancing selection in the major histocompatibility complex (MHC) in humans (Lenz et al., 2016).

Here, building on the work of Antonovics and Abrams, 2004, we use a similar though simplified two locus-two allele framework, taking into account the non-negligible reproductive stochasticity during the early stage of the dynamics of the mutant subpopulation, until it becomes extinct or reaches some appreciable fraction of the total population. More precisely, we consider a permanently heterozygous mating-type locus and a genetic load locus, and we assume that the recombination rate between the two loci is a fixed parameter. Individuals can reproduce via outcrossing, or via either one of two types of selfing, intra-tetrad mating or inter-tetrad mating. The two types of selfing depend on whether a given gamete mates with another gamete produced during the same meiosis event (within a tetrad) or with a gamete from a different meiosis (from another tetrad, App. B). The distinction is important because intra-tetrad mating maintains more heterozygosity in some genomic regions than inter-tetrad mating (Hood and Antonovics, 2000). Starting with a continuous-time Moran process, we derive the rates at which individuals of each genotype are produced. Then, as a new mutation is carried by very few individuals at the beginning of its evolution, a branching process naturally arises. Indeed, in this initial phase two individuals carrying the mutant allele have an extremely low probability to mate with each other. Mutant-carrier individuals can thus be assumed to reproduce independently of each other, leading to an approximation of the dynamics of the subpopulation of mutant carriers by a branching process.

The use of branching processes has shown its utility to account for the dynamics of a newly arised mutant allele in a population. Many estimates of the fixation or purging time of mutants in stochastic models (Champagnat and Méléard, 2011, Collet et al., 2013) relied on the use of branching processes to approximate the dynamics of a newly appeared mutant allele and of a nearly-fixed one. A branching-process approximation was used to study a two locus-two allele model, with individual fitness depending on the allelic state at both loci (Ewens, 1967, Ewens, 1968). For the diploid case, the framework of a seven-type branching process that can be used to study the fate of a deleterious mutation has been described, without deriving any analytical result (Pollard, 1966, Pollard, 1968). A similar branching process approximation was used to study the fate of a beneficial mutation with selfing (Pollak, 1987, Pollak and Sabran, 1992). Here, we

use a similar framework but consider deleterious mutations and a permanently heterozygous locus. Modeling multiple loci suggests the use of multitype branching processes, which have been widely studied (Harris, 1964, Kesten and Stigum, 1966 , Mode, 1971, Athreya and Ney, 1972, Sewastjanow, 1975, Pénisson, 2010). However, the multiplicity of types renders the derivation of analytical results on probabilities of extinction and on extinction times difficult (Heinzmann, 2009). We therefore use an analytical approach to study the probability that a new mutation is purged from the population, and a numerical approach to study the purging time (when purging occurs) to assess how long a deleterious or overdominant mutation remains in a population. We study in particular the impact of the mating system and of the level of linkage to a permanently heterozygous locus on the long-term maintenance of deleterious mutations near a fungal-like mating-type locus (*i.e.* not preventing diploid selfing).

2. Methods and Models

All parameters which will be needed below are listed in App. A.

2.1. Population and stochastic dynamics

We consider diploid (or dikaryotic) individuals, represented by their mating-type chromosomes, that harbor two biallelic loci: one mating-type locus, with alleles A and a, and one load locus, with a wild allele B and a mutant allele b. We model a fungal-like mating-type locus, so that mating is only possible between haploid cells carrying different alleles at the mating-type locus (this does not prevent diploid selfing as each diploid individual is heterozygous at the mating-type locus). Consequently, only four genotypes are admissible, denoted by $G_1, \dots, G_4$ in Figure 1. We follow the evolution of $(g(t))_{t\geq 0} = (g_1(t), \dots, g_4(t))_{t\geq 0}$, where $g_i(t)$ is the number of individuals of genotype G_i in the population at time t. We suppose that the reproduction dynamics is given by a biparental Moran model with selection. In this continous-time model, a single individual is replaced successively and the total population size, denoted by N, remains constant. A change in the population state g occurs in three steps.

The first step is the production of an offspring. After a random time following an exponential law of parameter N, an individual is chosen uniformly at random to reproduce. This means in particular that all individuals have the same probability to reproduce. Mathematically speaking, this formulation is equivalent to saying that each individual reproduces at rate 1. The chosen diploid individual produces haploid gametes, via meiosis, during which recombination takes place between the two loci with probability r (see Figure 1 (a)). The product of a meiosis is a tetrad that contains four haploid gametes (Figure 1 (b)). Mating can then occur through three modalities, illustrated in App. B (recall that two gametes can fuse only if they carry different mating-type alleles): (i) Intra-tetrad selfing, with probability fp_{in}: the two gametes are picked from the same tetrad, only one parent is involved; (ii) Inter-tetrad selfing, with probability $f(1-p_{in})$: the two gametes are picked from two different tetrads produced by the same individual, only one parent is involved; (iii) Outcrossing, with probability $1-f$: the two gametes are picked from tetrads produced by two different parents. In this case, the second parent is chosen uniformly at random in the remaining population, and produces haploid gametes via meiosis with the same recombination rate r. An offspring is produced following the chosen mating system, its genotype thus depending on the genotypes of the parents involved and on the occurrence of a recombination event in the tetrads.

The second step is the offspring survival. We assume that the fitness of a genotype G_i is the probability that an offspring with that genotype survives, and we denote it by S_i. We consider two selection scenarii (Figure 1, left): (i) The partial dominance case, where the mutant allele b is always deleterious and recessive. Homozygotes bb and heterozygotes Bb at the load locus have fitness values (*i.e.* a probability of survival) of $1-s$ and $1-hs$, respectively. Homozygotes BB have fitness 1; (ii) The overdominance case, where heterozygotes Bb are favored over BB and bb individuals. In this case, the fitness of Bb, bb and BB juveniles are respectively 1, $1-s_3$ and $1-s_4$, with $s_3 > s_4$ so that the fitness of bb individuals is lower than the fitness of wild-type individuals BB. The mating-type locus is considered neutral regarding survival.

The third step occurs if the offspring survives, in which case an individual chosen uniformly at random in the extant population is chosen to die and to be replaced by the offspring. If the offspring does not survive, the population state (g_1, g_2, g_3, g_4) does not change.

A jump in the stochastic process is thus an increase by one of the number of genotype G_i individuals in the population, when an offspring of genotype G_i is produced and survives, and a concomitant decrease by one of the number of genotype G_j individuals in the population, when an adult of genotype G_j dies. If $i = j$, *i.e.* if the surviving offspring and the individual chosen to die have the same genotype, the composition of the population does not change. We denote the jump rate from g to $g + e_i - e_j$ by $Q_{i,j}(g)$, where e_i is the vector with a 1 in position i and zeros everywhere else. $Q_{i,j}(g)$ is equal to the product of the rate at which an offspring of genotype G_i is produced (first step), which we denote by $T_g(+G_i)$, of the probability that it survives (S_i, second step), and of the probability that the adult chosen to die is of genotype G_j (third step). Thus, we have

$$Q_{i,j}(g) = T_g(+G_i) \times S_i \times \frac{g_j}{N}.$$

The total rates at which individuals of different genotypes are produced are given in App. C.1. For example, the rate at which an offspring of genotype G_1 is produced when the current state of the population is $g = (g_1, g_2, g_3, g_4)$ is given by

$$\begin{aligned} T_g(+G_1) = {} & f g_1 \left(1 - r\left(1 - \frac{1}{4}p_{in}\right) + \frac{1}{4}(1 - p_{in})\, r^2\right) + f g_2 \frac{r}{4}(p_{in} + r\,(1 - p_{in})) \\ & + \frac{1-f}{N-1}\Big[g_1\left(1 - \frac{r}{2}\right)\left((g_1 - 1)\left(1 - \frac{r}{2}\right) + g_3 + g_4\right) + g_2 r\left(g_2\frac{r}{4} + \frac{1}{2}(g_3 + g_4)\right) \\ & \qquad + g_1 g_2 r\left(1 - \frac{r}{2}\right) + g_3 g_4\Big]. \end{aligned}$$

The first two terms on the right-hand side, with a factor f, correspond to reproduction events by selfing. The third term, with a factor $1-f$, corresponds to reproduction events by outcrossing. Each subterm then encompasses the rate at which each genotype is involved in the reproduction event, and the probability that the offspring produced is of genotype G_1, taking into account possible recombinations. For example, the subterm $(1-f)/(N-1) \times g_1(g_1-1)(1-r/2)^2$ is the product of the total rate $g_1 \times 1$ at which an individual of genotype 1 reproduces, of the probability $1-f$ that reproduction happens by outcrossing, of the probability $(g_1-1)/(N-1)$ that the second parent is chosen among the other individuals of genotype G_1, and of the probability $(1 - r/2)^2$ that their offspring has genotype G_1.

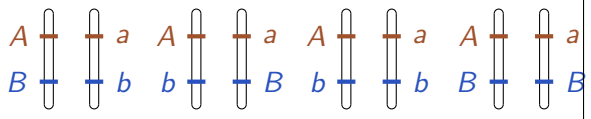
2.2. Branching-Process approximation

Let us now consider that the population size N is very large. When a mutation appears at the load locus, it is carried by a single individual. Hence, during the initial phase of the dynamics of the mutation b, the number of individuals who carry the mutation remains small compared to the number of wild-type individuals. The number of wild-type individuals is of the same order of magnitude as the total population size N, and the number of mutation-carrier individuals is negligible. More precisely, we assume that, when N is large,

$$g_4 \approx N, \quad \text{and } g_i \ll N \quad \text{for } i = 1, 2, 3. \tag{1}$$

Under this assumption, the jump rates $Q_{i,j}(g)$ of the process can be approximated by neglecting the terms of the form $1/(N-1) \times g_i \times g_j$, with $i, j \in \{1, 2, 3\}$, as they are of order $1/N$. This means that mating by outcrossing between individuals carrying the mutation b can be neglected. As a consequence, the birth rates of the different genotypes are linear in g_i, and a reproduction law for each genotype that is independent of the number of individuals of all other mutant-carrier genotypes can be derived. The Moran process can then be approximated by a branching process that follows the change in genotype counts for the mutation-carrier genotypes only.

We denote this branching process by $(Z_t)_{t\geq 0}$, where for each $t \geq 0$, we have $Z_t = (Z_{t,1}, Z_{t,2}, Z_{t,3})$, with $Z_{t,i}$ the number of individuals of genotype G_i in the population at

Genotype	G_1	G_2	G_3	G_4
				
Fitness S_i under Partial Dominance	$1-hs$	$1-hs$	$1-s$	1
Fitness S_i under Overdominance	1	1	$1-s_3$	$1-s_4$

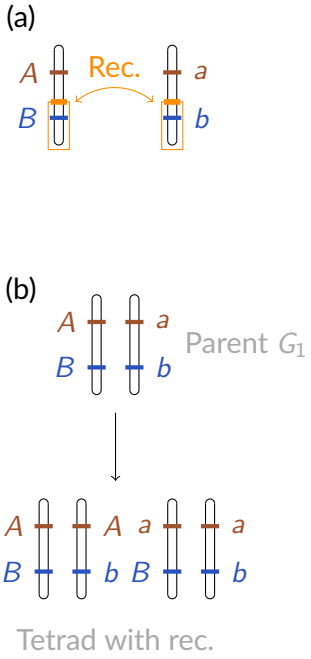


Figure 1 – Schematic drawings of the genotypes considered and their parameters. (Left) Description of the possible genotypes in the population and their fitness S_i for the two selection scenarii considered (partial dominance and overdominance). (Right) (a) Position of a putative event of recombination between the mating-type locus and the load locus. (b) Example of a tetrad that can be obtained after a meiosis of an individual of genotype G_1, with recombination. Four gametes are produced, two of each mating type. In the second and third gamete from the left, combinations of alleles that did not exist in the parent are observed (A with b and a with B).

time t. To each genotype is associated a reproduction law, that is, a probability distribution on $\mathbb{N}^3$ (vectors with three integer-valued coordinates) that gives the probability for an individual of that genotype to produce a given number of descendants of each genotype when it reproduces. Note that the rationale behind the branching process is different from the one for the Moran process. Indeed, each *replacement event* in the Moran model that involves an individual carrying the mutant allele b will be seen in the branching process as a *reproduction event*, in which the *offspring* is the mutant individual that is possibly produced during the first step of the Moran jump, and the *parent* is another mutant individual that is either one of the two actual parents in the replacement event, or the individual chosen to be replaced by the offspring in the Moran replacement event. A *reproduction event* of the branching process consists in the replacement of the parent by its descendants, which will be made of the mutant *offspring* when there is one, and of the mutant *parent* when it remains in the population. More precisely, we will encode three situations as follows: (i) when the replacement event in the Moran model corresponds to the reproduction of an individual of genotype G_i, $i \in \{1,2,3\}$ (via selfing or outcrossing with an individual of genotype G_4), that this reproduction event generates a mutant offspring of genotype $G_j, j \in \{1,2,3\}$, and the mutant parent is not chosen to die, we will see the *reproduction event* of the branching process as being an individual of genotype G_i having descendance vector $e_i + e_j$; (ii) When the Moran replacement event leads to the reproduction of an individual of genotype G_i, $i \in \{1,2,3\}$ (via selfing or outcrossing with an individual of genotype G_4), that this reproduction event generates an offspring of genotype G_4, and the mutant parent is not chosen to die, we will see the *reproduction event* as being an individual of genotype G_i having descendance vector e_i (as non-mutant individuals are not accounted for in the branching process approximation). Note that this reproduction event will imply no change in the population state, but for the sake of completeness we indicate here all Moran replacement events that have non-vanishing rates as N tends to infinity; (iii) When the Moran replacement event only involves non-mutant parents

and an individual of genotype G_i, $i \in \{1,2,3\}$, is chosen to die, we will see the *reproduction event* as being an individual of genotype G_i having descendance vector 0 (corresponding to the *parent* being removed from the branching process and no mutant offspring being produced). Other possible Moran replacement events occur at rates that vanish as N tends to infinity, and therefore do not contribute to the *reproduction events* of the branching process. The rates at which *reproduction events* described above occur are directly derived from the rates $Q_{i,j}(g)$ of the Moran model, under the approximation stated in Eq.(1). They are summarized in the matrices A, T, and D defined as follows:

$$A = \begin{pmatrix} (fa(r)+(1-f)d(r))\,S_1 & (fc(r)+(1-f)\frac{r}{2})\,S_1 & (1-f)S_1 \\ (fc(r)+(1-f)\frac{r}{2})\,S_2 & (fa(r)+(1-f)d(r))\,S_2 & (1-f)S_2 \\ fb(r)S_3 & fb(r)S_3 & fS_3 \end{pmatrix},$$

$$T = \begin{pmatrix} (fb(r)+\frac{1-f}{2})S_4 & 0 & 0 \\ 0 & (fb(r)+\frac{1-f}{2})S_4 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$D = \begin{pmatrix} S_4 & 0 & 0 \\ 0 & S_4 & 0 \\ 0 & 0 & S_4 \end{pmatrix},$$

with

$$a(r) = 1 - r + \frac{r}{4}\Big(1-(1-p_{in})(1-r)\Big), \quad b(r) = \frac{r}{4}\Big(1+(1-p_{in})(1-r)\Big), \tag{2}$$

$$c(r) = \frac{r}{4}\Big(1-(1-p_{in})(1-r)\Big), \quad \text{and} \quad d(r) = 1-\frac{r}{2}. \tag{3}$$

The entries A_{ij} of matrix A, T_{ij} of matrix T and D_{jj} of matrix D give the rates at which each individual of genotype j reproduces and gives rise to a descendance vector respectively equal to $e_i + e_j$ (situation (i)), e_i (situation (ii)), and 0 (situation (iii)). An example of derivation of the matrix coefficients is given in App. C.2.

2.3. Probability of purge and purging time

Under the assumption that the mutation is initially rare (after a mutation or migration event for example), we can use the branching process approximation described in Section 2.2 to derive the probability and purging time of the mutation from the population. In particular, our goal is to analyze the effect of the presence of a mating-type locus near the load locus on the purge of the deleterious mutant b, *i.e.* on the extinction of the mutant-carrier population described by the branching process.

Extinction Probability

The probability of extinction of the branching process can be determined by looking at the eigenvalues of the matrix C such that $\mathbb{E}[Z_t|Z_0 = z_0] = z_0 e^{Ct}$ for $t \geq 0$, where $z_0 \in \mathbb{N}^3$ is the initial state of the branching process $(Z_t)_{t\geq 0}$ (Sewastjanow, 1975 in German, and Pénisson, 2010 for a statement of these results in English). Under the assumption of irreducibility of the matrix C, results relying on the theory of Perron-Froebenius (see for example Athreya and Ney, 1972) state that the process almost surely dies out (*i.e.* the mutation is purged with probability 1) if and only if ρ, the maximum eigenvalue of C, satisfies $\rho \leq 0$. When C is not irreducible, which occurs for example if $f = 0$ or $f = 1$, the result still holds but requires the use of the theory of final classes (Sewastjanow, 1975, cited in Pénisson, 2010). Details are given in App. C.4.

We follow a method described in Bacaër, 2018, to compute the matrix C mentioned above and obtain

$$C_{ij} = \begin{cases} A_{ij} + T_{ij} & \text{if } i \neq j, \\ A_{jj} - \sum_{k\neq j} T_{kj} - D_{jj} & \text{if } i = j. \end{cases}$$

This gives

$$C = \begin{pmatrix} (fa(r) + (1-f)d(r))\,S_1 - S_4 & (fc(r) + (1-f)\frac{r}{2})\,S_1 & (1-f)S_1 \\ (fc(r) + (1-f)\frac{r}{2})\,S_2 & (fa(r) + (1-f)d(r))\,S_2 - S_4 & (1-f)S_2 \\ fb(r)S_3 & fb(r)S_3 & fS_3 - S_4 \end{pmatrix},$$

where the functions a, b, c, d were defined in Eqs. 2 and 3 (see details in App. C.3).

We derived the dominant eigenvalue using Mathematica (Wolfram Research, 2015) and study its sign analytically when possible, or numerically otherwise.

Comparison with previous results

Our results can be compared to the work of Ewens, 1967, who used a similar framework to study a random-mating population with two biallelic loci under selection, one of which carried a new allele. Assuming that the frequency of the gametes that carried a new allele was negligible compared to the frequencies of wild-type gametes, he used a branching process approximation to study the probability that the new allele was purged from the population. He considered a recombination rate R between the two loci, and fitnesses w_{ij} for each genotype (where i and j take the value 1 or 3 when loci are homozygous, and the value 2 when heterozygous). Setting $w_{i1} = w_{i3} = 0$ for $i = 1, 2, 3$ allows to force heterozygosity at the locus that does not carry the new allele in his model, and to compare his findings with our results on the fate of a new allele appearing near a permanently heterozygous locus. The dominant eigenvalue of the matrix driving the dynamics of the new allele in Ewens, 1967, is

$$\lambda_1 = \frac{w_{22}}{w_{32}} \tag{4}$$

with w_{22} being the fitness of individuals heterozygous for the new allele, and w_{32} the fitness of homozygous wild-type individuals. As Ewens, 1967, considered a discrete-time branching process, this dominant eigenvalue must be compared to one to deduce information on the new allele survival probability.

Sheltering effect of the mating-type locus

We investigate now to the potential effect of the presence of a mating-type locus on the maintenance of a mutant allele in a population: as mating-type alleles are always heterozygous, any mutation appearing completely linked to one mating-type allele is maintained in a heterozygous state as well. The load of the mutant allele is then less expressed when the mutation is recessive, and the mutation is said to be "sheltered".

This potential *sheltering effect* can be explored by looking at the variation of the dominant eigenvalue ρ when the recombination rate r is close to 0.5. Indeed, the quantity $|\rho|$ can be seen as the rate of decay of the deleterious mutant subpopulation (see the results on the probability of survival of a multitype branching process, Th. 3.1 of Heinzmann, 2009), and its value gives a rough approximation of the inverse of the mean time to extinction of this subpopulation, *i.e.* of the mean purging time of the mutant allele b. Moreover, setting the recombination rate to $r = 0.5$ in our model allows us to consider a load locus completely unlinked to the mating-type locus, while decreasing the value of r introduces some loose linkage between the two loci. We thus look at the derivative $\frac{\partial \rho}{\partial r}\,|_{r=0.5}$ to obtain the variation of the dominant eigenvalue of C when departing from this unlinked state.

The sign of the derivative gives information on the existence of a sheltering effect due to the mating-type locus: if $\frac{\partial \rho}{\partial r}\,|_{r=0.5} < 0$, then when r decreases from 0.5 to lower values, *i.e.* when linkage between the two loci appears, the (negative) value of ρ increases, which means that the purging of the mutation becomes slower. In this case, the mating-type locus has a sheltering

effect. Otherwise, if $\frac{\partial \rho}{\partial r} |_{r=0.5} > 0$, the presence of a mating-type locus accelerates the purging of a deleterious allele.

The absolute value of the derivative also gives information on the strength of the sheltering effect of the mating-type locus. The closer to 0 the derivative is, the smaller the impact of the mating-type locus. We compute the derivative and study its sign analytically. We then study the values of the derivative numerically in order to identify the impact of each parameter on the sheltering effect of the mating-type locus.

We also look at the strength of the sheltering effect on mutations close to the mating-type locus, by studying the eigenvalue variation around $r = 0$. Setting the recombination rate to $r = 0$ models a situation where the load locus is completely linked to the mating-type locus. Hence, the mutation is completely linked to one mating-type allele, and maintained in a heterozygous state. Looking at the derivative $\frac{\partial \rho}{\partial r} |_{r=0}$ allows us to quantify the impact of departing from this situation by loosening the linkage between the two loci. We study the difference between the derivative at $r = 0.5$ and the derivative at $r = 0$ to compare the effect of adding a small amount of linkage between completely unlinked loci ($r = 0.5$) and the effect of adding a small amount of recombination between completely linked loci ($r = 0$).

Extinction time

The mean time to extinction in a multitype branching process is finite for a subcritical process (that is, when the principal eigenvalue ρ of C is less than 0), and infinite for a critical process (*i.e.* when $\rho = 0$, see Pötscher, 1985, for the proof of existence and finiteness of extinction time moments). Previous work, in particular Theorem 4.2 in Heinzmann, 2009, showed that a Gumbel law gives a good approximation of the law of the extinction time, provided that the initial number of individuals in the branching process and the absolute value of the dominant eigenvalue are both large. In our case, however, the mutation appears in a single individual, and the dominant eigenvalue is close to zero, which prevents the use of the Gumbel law approximation. Therefore, we performed computer simulations to study the empirical distribution of the time to extinction of the process, *i.e.* the purging time of the b mutant allele.

The branching process was simulated with a Gillespie algorithm to obtain an empirical distribution for the time to extinction. More precisely, the Gillespie algorithm produces realizations of the stochastic process by iteratively updating the number of individuals of each genotype within the multitype branching process (Gillespie, 1976). To circumvent the problem of exponential increase of the population size in the supercritical case, the parameters were chosen so that the branching process was subcritical. The probability of extinction was thus equal to 1 and the mean time to extinction was finite. For each scenario, we looked at different values of the recombination rate r, in order to study the impact of linkage between the load locus and the mating-type locus on the purging time of the mutant allele. We also chose different values for the selfing rate f in order to assess the impact of the mating system on the purging time of the mutant allele. For each set of parameters, 100,000 independent simulation runs were performed with the same initial condition (a single individual heterozygous at the load locus was introduced). The scripts used to simulate the process and display the figures are available at Tezenas, 2022.

Probability of a new mutation apparition before the first one is purged

As a first step towards the study of the accumulation of deleterious mutations near a mating-type locus, we studied the probability that the deleterious mutation can be maintained long enough in the population so that a second mutation can appear before the first one is purged. We considered that a second mutation could appear during a reproduction event occurring in the population of mutation carriers (described by the branching process), on a region of a given length $d = 10^6$ base pairs, at a rate of $\mu = 10^{-8}$ mutations per base pairs per reproduction event. The mean number of reproduction events needed for a new mutation to appear in a region of length d, $\bar{n}_{ev}$, is the inverse of the mutation rate μ multiplied by the length d:

$$\bar{n}_{ev} = \frac{1}{\mu \times d} = 10^2.$$

We then estimated the probability that a new mutation appears in such a genomic region before the first one is purged by counting the number of independent simulations in which the number of reproduction events exceeded $\bar{n}_{ev}$ before the branching process went extinct (*i.e.* before the purging of the first mutation), over 100,000 simulation runs. Note that we did not take into account the genotype of the individual on which the second mutation appears, and therefore we did not distinguish whether the second mutation appears on a chromosome that carries the first one or not. Our estimate thus does not exactly equals the probability to have two mutations on the same chromosome, but this gives an order of magnitude of the probability of deleterious mutation accumulation and of the impact of the mating system. The length of the genomic region on which a second mutation can appear was chosen arbitrarily, and changing it can also change the probability. However, the important point for the deleterious-mutation mechanism to work is that there exists a size for regions flanking mating-type loci that allows both inversions to appear and mutations to accumulate, so that inversions can trap several deleterious mutations when suppressing recombination. The value $d = 10^6$ chosen here allows to cover such flanking regions.

We computed our estimate of the probability of deleterious mutations accumulation for $r = 0.001$ (the two loci are close, strongly linked), $r = 0.01$, $r = 0.1$, and $r = 0.5$ (the two loci are distant, unlinked). We considered several values of selfing and intra-tetrad mating rates f and p_{in} in order to assess the impact of the mating system on the probability of deleterious mutation accumulation near a mating-type locus.

3. Results

3.1. Deleterious mutations are almost surely purged in the partial dominance case, and can escape purge in the overdominance case

Partial Dominance scenario

Under partial dominance, we find that the dominant eigenvalue ρ of the matrix C is always negative or null (see App. E.1 and E.2 for more details on the proof and computations). Previous theoretical results on branching processes state that, when $\rho < 0$, the probability that the deleterious mutation is purged from the population before it reaches a substantial frequency is one, and the mean time of purging is finite (see the Methods section). In particular, the probability of purging does not depend on the mating system ($\rho < 0$ for any value of intratetrad, intertetrad and outcrossing rates), nor on the recombination probability, selection and dominance coefficients. The only exceptions are when the deleterious mutation is neutral ($s = 0$) or behaves as neutral ($h = 0$ and $r = 0$, the mutation is neutral when heterozygous and completely linked to one mating-type allele), in which case the dominant eigenvalue is 0. The mutation is still purged from the population but previous theoretical results on branching processes state that this can take a much longer time compared to the case where $\rho < 0$, as the mean purging time would be infinite (see the Methods section).

Taking $w_{22} = 1 - hs$ and $w_{32} = 1$ in the model of Ewens, 1967, to mirror our partial dominance scenario, the dominant eigenvalue becomes $1 - hs$. It is always smaller than one, except when $h = 0$ or $s = 0$, *i.e.* when the mutation is neutral in the heterozygous state. Except in those cases, the mutation is purged from the population with probability one. We therefore find the same results as Ewens, 1967, and we extend these results in the case where mating is not random among gametes. In particular, the mutation being neutral in the heterozygous case ($h = 0$) is not sufficient to prevent the purging probability to be one when mating is not random: the mutation has to be completely linked to a permanently heterozygous locus ($h = 0$ and $r = 0$).

Overdominance scenario

Under overdominance, the dominant eigenvalue ρ can take positive or negative values. When ρ is positive, the probability that the mutation escapes purging and that the number of mutation-carriers increases exponentially fast is strictly positive. The general conditions on the parameters for ρ to be positive in our model are given in App. F.2, but they are difficult to interpret. Below, we describe a few simple cases in order to elucidate the role of each parameter, and then we complement the analysis with a numerical approach.

Similarly to the partial dominance case, the dominant eigenvalue is 0 when the mutation is neutral ($s_3 = 0$, which implies $s_4 = 0$ as well). The dynamics of the b-subpopulation (*i.e.* mutation-carriers) is then critical, which means that the mutant is purged with probability 1 but the mean purging time can be arbitrarily long (as the average extinction time of a critical branching process is infinite, see the Methods section).

When the mutation is not neutral ($s_3 \neq 0$) but with no disadvantage to BB homozygotes ($s_4 = 0$), we prove that $\rho < 0$ (see App. F.2), which means that the dynamics of the b subpopulation is subcritical and that the mutant allele is purged with probability 1. This shows that the overdominant mutant allele is not maintained in the population when wild-type homozygotes are not disfavored compared to heterozygotes at the load locus. This corresponds to a completely recessive mutation, and is in agreement with the results for the partial dominance case with $h = 0$.

When the mutant allele is completely linked to a mating-type allele ($r = 0$), or under complete outcrossing ($f = 0$), the dominant eigenvalue is equal to s_4, the selection coefficient for the fitness reduction of the BB wild-type homozygotes. The dynamics of the b subpopulation is then supercritical, which means that there is a non-zero probability that the mutant allele is not purged and, instead, reaches a significant number of carriers. Moreover, the mutant allele is more favored in this case when selection against BB homozygotes is stronger as it induces a stronger advantage of the Bb heterozygotes. A similar result can be derived from the work of Ewens, 1967. Taking $w_{22} = 1$ and $w_{32} = 1 - s_4$ in his model to mirror our overdominance scenario, the dominant eigenvalue of Eq. 4 becomes $1/(1 - s_4)$. As long as $s_4 > 0$, this eigenvalue is always greater than one, and its value increases as the selection against wild-type homozygotes increases. This shows that the dynamics of an overdominant allele under random gamete mating is similar as under complete outcrossing.

In the case of complete intra-tetrad selfing ($f = 1$, $p_{in} = 1$), we find that $\rho \geq 0$ if $r \leq 2s_4$, in agreement with the results of Antonovics and Abrams, 2004. These results mean that the overdominant mutation can be maintained under complete selfing if it is tightly linked to the mating-type locus (r small) or if the heterozygote advantage over wild-type homozygotes is strong (s_4 large).

In the case of complete selfing ($f = 1$), we find that $\rho = s_4 - s_3 \leq 0$ when $r(2 - r - p_{in}(1 - r)) - 2s_3 \geq 0$. This shows that the dominant eigenvalue depends only on the selection coefficients when the recombination rate r exceeds a certain threshold (visible on the bottom panels of Figure 2). This means that, if the recombination rate is larger than the strength of the selection against deleterious homozygotes, the mutation is purged with probability one. Moreover, the purging time is shorter when the difference in fitness between the two homozygotes is larger. The threshold on recombination increases as p_{in} increases, which means that the strength of the linkage between the mating-type locus and the mutation has the highest effect under intra-tetrad selfing.

Figure 2 shows more generally that the mating system affects the purging of deleterious mutations. On Figure 2, the probability of purging is one in blue areas (the dominant eigenvalue is negative), and positive but smaller than one in red areas (the dominant eigenvalue is positive). The lines below which the mutation has a non-zero survival probability under the framework of Antonovics and Abrams, 2004, *i.e.* $r = 2s_4$ under complete intra-tetrad selfing, are displayed as well. Comparing the panels for different values of intratetrad mating rate (p_{in}) and selfing rate (f) shows that selfing favors the purging of the mutant allele (the blue area becomes larger as f increases), whereas intratetrad mating favors the maintenance of the deleterious allele (the

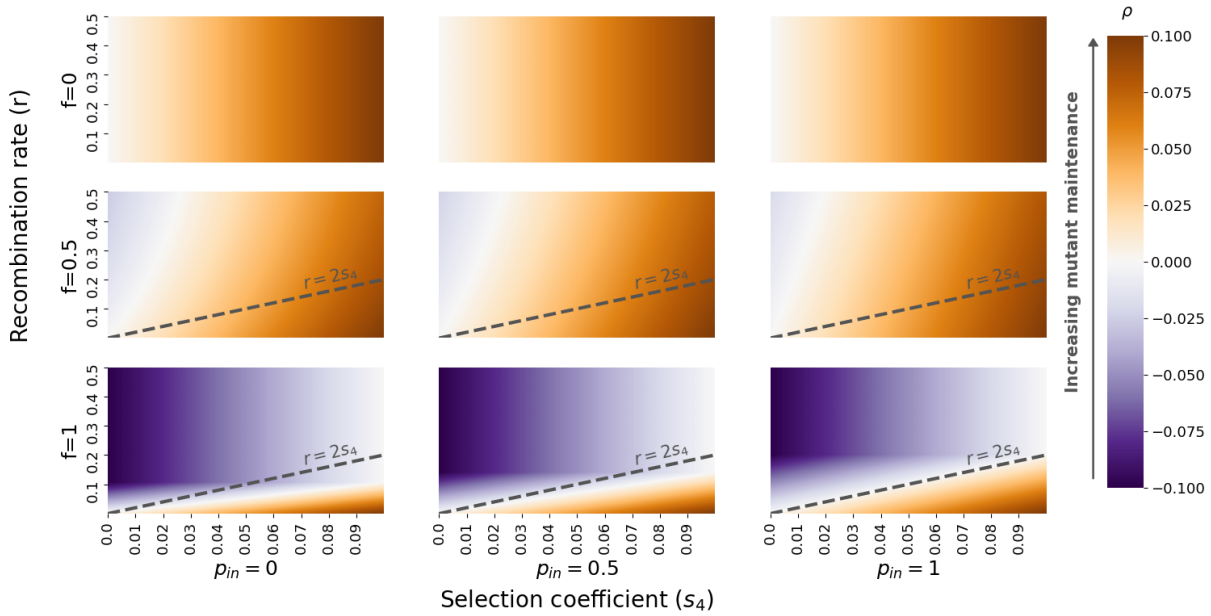


Figure 2 – Dominant eigenvalue ρ for the overdominance scenario. When $\rho \leq 0$ (blue areas), the mutation is purged with probability 1. When $\rho > 0$ (red areas), the mutation has a non-zero probability to escape purging. The mutation is maintained longer in the population as ρ increases. All panels have the same axes. x-axis: s_4, selection coefficient for wild-type *BB* homozygotes. y-axis: r, recombination rate between the two loci. Each column corresponds to a value of p_{in} (intra-tetrad rate, $0, 0.5, 1$), and each row to a value of f (selfing rate, $0, 0.5, 1$). The selection coefficient for *bb* homozygotes is set to $s_3 = 0.1$. The line $r = 2s_4$ is displayed for comparison with the findings in Antonovics and Abrams, 2004.

blue area become smaller as p_{in} increases). Indeed, selfing favors the creation of homozygous individuals, which are disfavored, and intra-tetrad selfing favors the creation of heterozygous individuals, which are favored, compared to inter-tetrad selfing: the probability that a heterozygous individual *Bb* produces a heterozygous offspring *Bb* is higher under intra-tetrad selfing (probability $1 - r/2$) than under inter-tetrad selfing (probability $1 - r + r^2/2$).

3.2. The presence of a mating-type locus has a sheltering effect under partial selfing

Looking at the derivative of the dominant eigenvalue at $r = 0.5$, we find that the presence of a mating-type locus near the mutation has a sheltering effect on the deleterious mutation, under partial selfing and in both selection scenarii. Indeed, the derivative $\frac{\partial \rho}{\partial r} |_{r=0.5}$ is always negative, except when the mutation is neutral ($s = 0$ under partial dominance or $s_3 = 0$ under overdominance), when it is lethal ($s = 1$) or dominant ($h = 1$) under partial dominance, or under complete outcrossing ($f = 0$) in both scenarii, in which cases the derivative is zero and there is no sheltering effect. Under complete selfing ($f = 1$), the derivative is also null when the intratetrad coefficient p_{in} is below a certain threshold (see App E.3 and F.3 for the proof). As explained in the Methods section, this analysis shows that, in a wide range of situations, the rate of decay of the mutant subpopulation is lower when the mutation is linked to a mating-type locus, even loosely (*i.e.* as soon as $r < 0.5$), than when recombination is free between the two loci. Hence, except for the particular cases cited above, the mating-type locus always has a sheltering effect on the deleterious mutation maintenance under partial selfing, independently of the mating system coefficients (f and p_{in}) and of the selection and dominance coefficients (s and h, or s_3 and s_4).

Figure 3 shows that, under both partial dominance or overdominance, the variation of the derivative at $r = 0.5$ is stronger when the selfing rate f (x-axis) or the intratetrad selfing probability p_{in} (y-axis) are high. This means that the sheltering effect of the mating-type locus is stronger under high selfing or high intratetrad mating. Two forces oppose here: increasing selfing induces a

greater production of homozygotes, which are disfavored, whereas increasing intra-tetrad selfing rate or increasing the linkage with a mating-type locus favors the production of heterozygotes, which are favored. The sheltering effect of the mating-type locus that counters the purging effect of selfing is higher when selfing is higher, and this countering effect is reinforced by a high intra-tetrad mating rate. Moreover, when approaching $f = 1$, the derivative decreases to 0. Indeed, the selection and dominance coefficients s, s_3 and h are here sufficiently small for the condition to have $\frac{\partial \rho}{\partial r}\Big|_{r=0.5} = 0$ when $f = 1$ to be met, for both selection scenarii (see App. E.3 and F.3 for the derivation of this condition). This means that the dynamics of the deleterious mutation is independent of the presence of a mating-type locus under complete selfing and weak selection.

We explore the impact of other parameters in the Supplementary materials. Figure S2 shows that, under partial dominance, the sheltering effect of a mating-type locus is stronger when the dominance coefficient h is lower (Bb heterozygotes, which are more prone to be created in the presence of a mating-type locus, are more favored) or when the selection coefficient s is high (the differential in fitness between Bb heterozygotes and bb homozygotes is higher). Similarly, Figure S3 shows that, under overdominance, the sheltering effect of the mating-type locus is stronger when the selection against bb homozygotes is higher (s_3 coefficient), whereas the selection against BB homozygotes does not impact the strength of the sheltering effect, suggesting that the dynamics of the deleterious allele is mostly driven by the difference in fitness between the favored heterozygotes and the disfavored deleterious homozygotes.

Looking at the derivative at $r = 0$, we show in App. E.3 and App. F.3 that it is also negative in both selection scenarii. This means that the eigenvalue decreases, *i.e.* that the mutation is less maintained in the population as soon as the two loci are no longer completely linked. Figure S4 shows that the difference $\Delta\left(\frac{\partial \rho}{\partial r}\right) = \frac{\partial \rho}{\partial r}\,|_{r=0.5} - \frac{\partial \rho}{\partial r}\,|_{r=0}$ is always positive, which means that the absolute value of the derivative at $r = 0$ is larger than the absolute value of the derivative at $r = 0.5$. This shows that the sheltering effect is stronger on mutations closely linked to the mating-type locus : adding a small chance of recombination on previously completely linked loci ($r = 0$) has a greater impact on the maintenance of deleterious mutations than adding a small amount of linkage between two previously completely unliked loci ($r = 0.5$). The largest difference between the two derivatives occurs for selfing rates close to one, the derivative being then zero at $r = 0.5$, while the derivative at $r = 0$ approaches -1. This shows that the linkage to the mating-type locus particularly impacts the strength of its sheltering effect under high selfing.

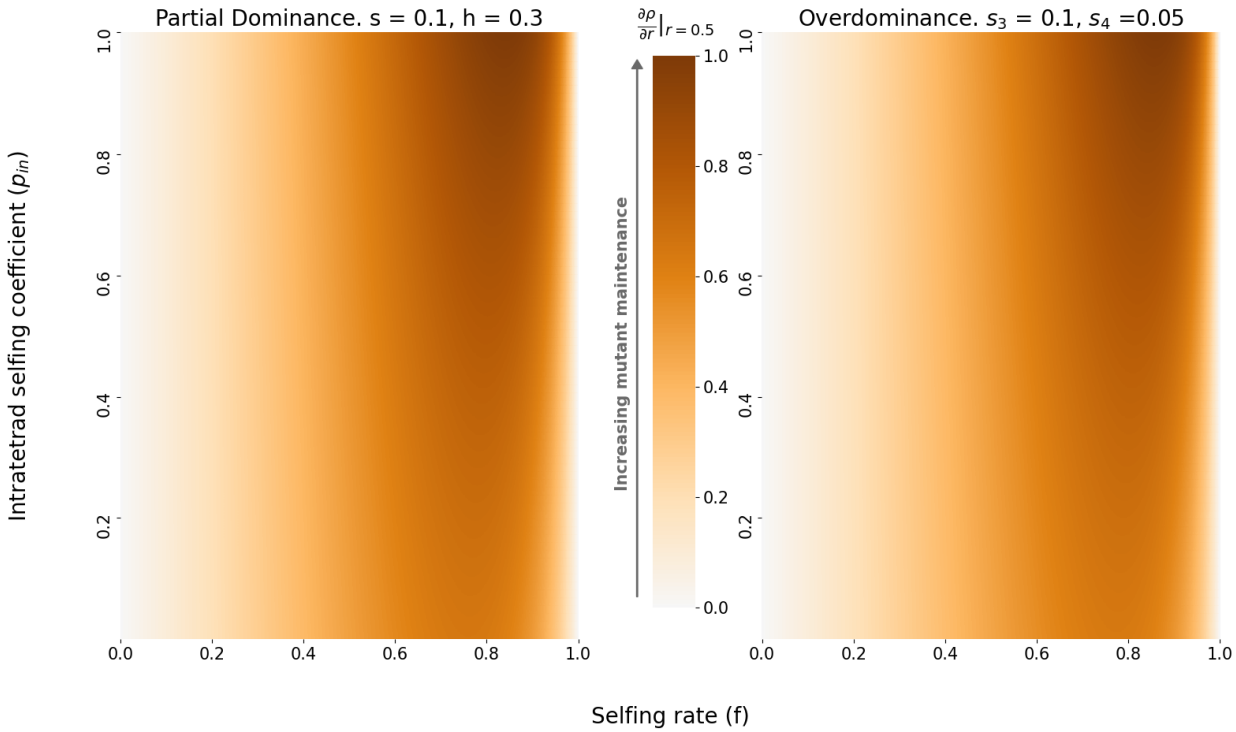


Figure 3 – Relative variation of the derivative of the dominant eigenvalue in the partial dominance case (left) and the overdominance case (right). For each panel, the values of $\frac{\partial \rho}{\partial r}|_{r=0.5}$ range from a minimal value, which is negative, to zero. We divided each value of the derivative by this minimum in order to plot values between 0 and 1 for every panel. This enables us to compare the effect of the presence of a mating-type locus on the same scale for both selection scenarii. x-axis: selfing rate f. y-axis: intratetrad selfing rate p_{in}. The darker the color, the more the mating-type locus shelters the mutation, thus promoting its maintenance.

3.3. Rare events of maintenance of the deleterious mutation occur in both selection scenarii, paving the way for an accumulation of mutations

The empirical distribution of the purging time of the deleterious mutation in the partial dominance case is shown on Figure 4: for ca. 75% of the independent runs, the mutation was rapidly purged, while in some rare cases (ca. 1%), the purge took very long (several orders of magnitude longer than the 75% percentile empirically obtained from the 100,000 runs). Note that the approximation of the distribution of the time to extinction by a Gumbel law (Th. 4.1 of Heinzmann, 2009) falls short here, because the initial number of individuals (one) and the absolute value of ρ (given in the caption) are too small.

Consistently with our results that $\frac{\partial \rho}{\partial r} < 0$, the sheltering effect of the mating-type locus implies that the purging time increases when the recombination rate decreases (Figure 4, and Figure S6 for the overdominant case). We also consistently find that increasing selfing decreases the purging time (Figures S5 and S7). In each case, the closer ρ is to zero, the more extreme the rare events are : the distribution of the 1% longest purging times is stretched towards higher values when ρ gets closer to zero, while the distributions of the 75% shortest remain similar.

Figure 5 displays the probability that the mutation can be maintained long enough in the population for another mutation to appear in a region of 10^6 bp near the mating-type locus. This probability is nonnegligible (of the order of 1% to 10%), which shows that accumulation events are rare but still occur near mating-type loci. This is true even under selfing as the sheltering effect of the mating-type locus can counter the purging effect of selfing. Indeed, when the recombination rate between the first mutation and the mating-type locus is high ($r = 0.5$ or $r = 0.1$), modeling a situation where the distance between the two loci is large, the probability that a

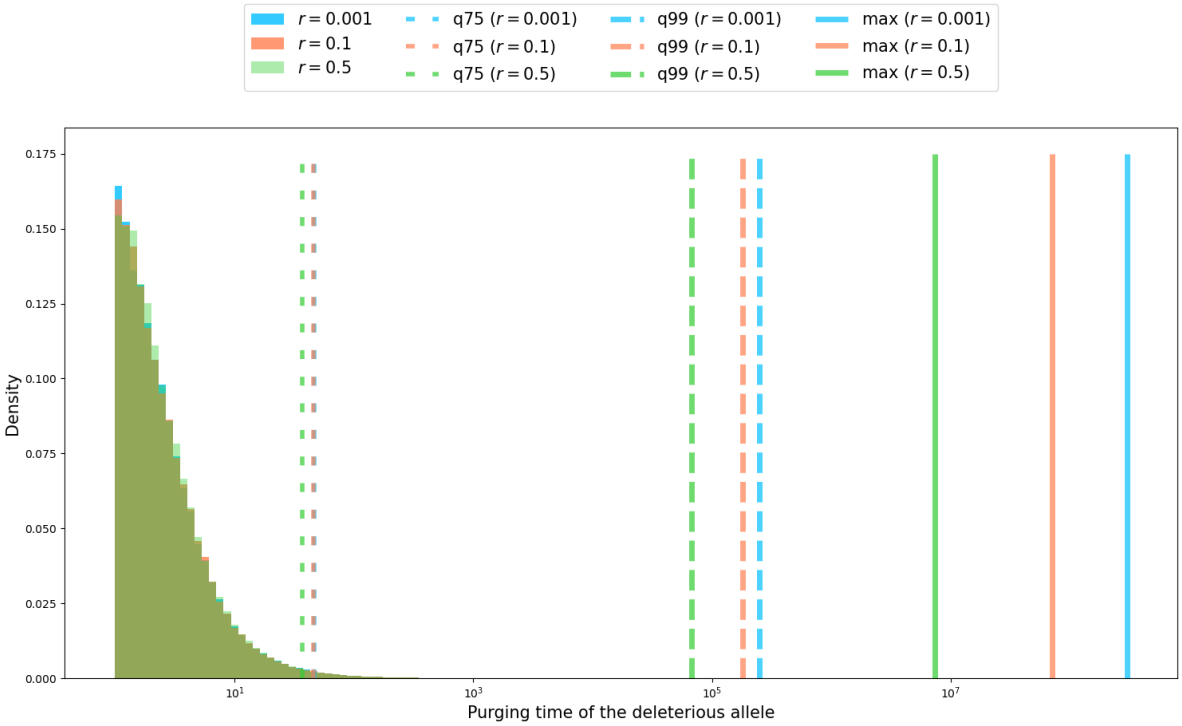


Figure 4 – Empirical distribution of the deleterious allele purging time for the partial dominance scenario. A total of 100,000 simulations were run, with $s = 0.1$, $h = 0.1$, $f = 0.5$, $p_{in} = 0.5$, starting from one heterozygous individual ($X_0 = (1, 0, 0)$), and for three values of the recombination rate ($r = 0.001$ in blue, $r = 0.1$ in red and $r = 0.5$ in green). The respective values for ρ are $\rho = -0.0101$, $\rho = -0.0106$ and $\rho = -0.0307$. The x-axis is log-scaled. The large-dotted lines represent the 75^{th} percentile (*q75*), the dashed lines indicate the 99^{th} percentile (*q99*), and solid lines the maximum value (*max*) of the purging time. Maximum values are several order of magnitudes higher than the 75th percentile of the empirical distribution of the purging time.

second mutation appears before the first one is purged decreases with increasing selfing, even with high intra-tetrad selfing rates. However, when the first mutation is closer to the mating-type locus (lower recombination rates), the probability that a second mutation appears before the first one is purged under selfing is similar to the probability under complete outcrossing. The presence of a mating-type locus can thus facilitate the accumulation of deleterious mutations in its flanking regions, especially in highly selfing populations.

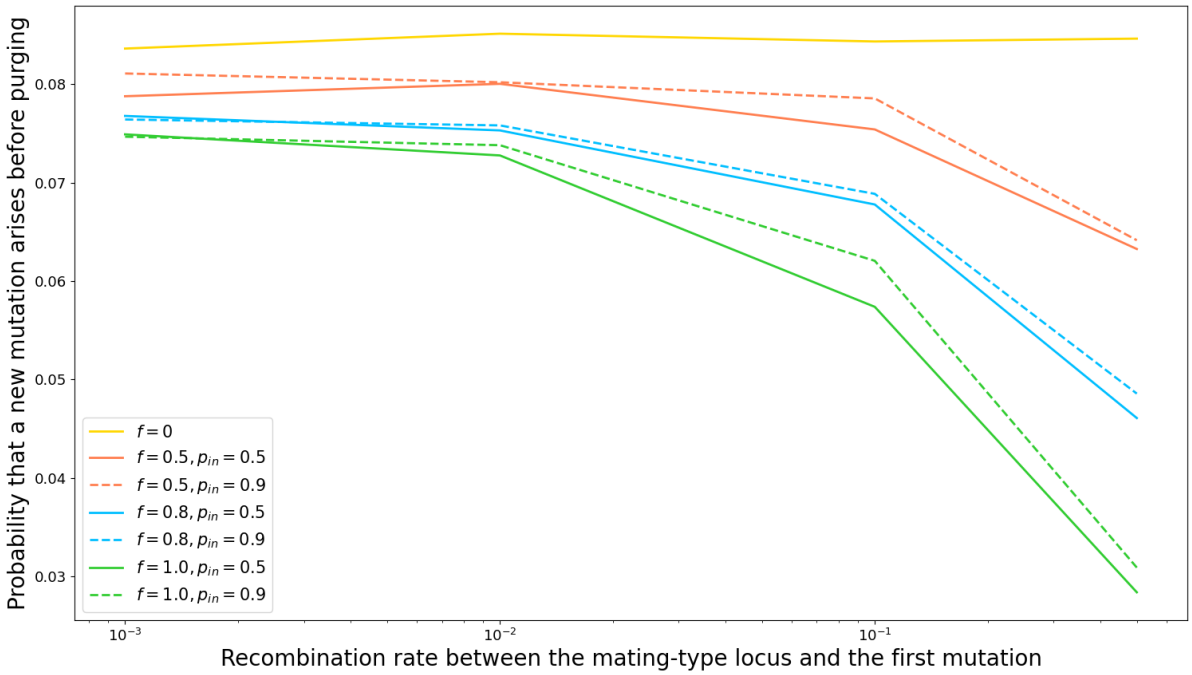


Figure 5 – Probability that a new mutation appears in a region of length 10^6 bp before the first mutation is purged from the population, under the partial dominance scenario, depending on the recombination rate between the first mutation and the mating-type locus. We considered a mutation rate per base pair per reproduction event of 10^{-8}. Here, the reproduction events are those of the branching process, that change the composition of the mutant-carriers subpopulation. The probability that a new mutation appears before the purge of the first one is approximated by the proportion of simulation runs for which the number of reproduction events exceeds the expected number of events needed for a new mutation to appear (see text). For each set of parameters (r, f, p_{in}), 100,000 independent simulations were run. Colors correspond to different values of the selfing rate f, and line styles to different values of the intra-tetrad selfing rate p_{in}. When $f = 0$, a single curve is displayed, as the value of p_{in} has no impact under complete outcrossing. For all simulations, we set $s = 0.1$ and $h = 0.1$.

4. Discussion

Partially recessive deleterious mutations are almost surely purged in finite time while overdominant mutations can persist

We have shown that partially recessive deleterious mutations close to a fungal-like mating-type locus (*i.e.* that does not prevent diploid selfing) are almost surely purged in finite time, except when they are neutral or behave as neutral. In the overdominance case, the probability of purge depends on parameter values. Low selfing rates, high intra-tetrad selfing rates or tight linkage to the mating-type locus increases both the maintenance probability and persistence of the overdominant allele, whereas a high selfing rate favors its purge.

In particular, if linkage is complete (corresponding to $r = 0$ here, or to the case where the inversion encompasses a permanently heterozygous locus in Jay et al., 2022), an overdominant allele may be maintained in a population and even sweep to fixation with non-zero probability, which confirms previous findings (Antonovics et al., 1998, Antonovics and Abrams, 2004, Jay et al., 2022). This means that, although selfing purges deleterious mutations, a mating-type locus can have a sheltering effect in its flanking regions.

In general, the overdominant allele is maintained longer and with a higher probability in the population when the fitness advantage of heterozygotes over homozygotes is higher, in line with previous simulation results (Antonovics and Abrams, 2004). This conclusion is sensible: if the mutant is strongly favored in a heterozygous state, it can be maintained in this state in the population.

The presence of the mating-type locus has a sheltering effect under selfing

We found that, in both selection scenarii, the presence of the mating-type locus had no effect on the maintenance of deleterious mutations under outcrossing, but always had a sheltering effect under selfing, which strengthened as the selfing rate increased. Indeed, selfing increases homozygosity and thus accelerates the purge of a deleterious allele, whereas the presence of a permanently heterozygous mating-type locus induces more heterozygosity in its flanking regions, that counters the purging effect of selfing. The sheltering effect of a mating-type locus is thus all the more tangible as it counters the strong purging effect induced by selfing. Increasing intra-tetrad selfing also induces more heterozygosity and thus slightly reinforces the sheltering effect of the mating-type locus. This is consistent with the findings that, in fungi, ascomycetes that reproduce via outcrossing and live as haploids do not show evolutionary strata (Skinner et al., 1993, Zhong et al., 2002, Phan et al., 2003, Kuhn et al., 2006, Jin et al., 2007, Malkus et al., 2009) whereas pseudo-homothallic ascomycete fungi, living as dikaryotic and undergoing mostly intra-tetrad selfing, are those with evolutionary strata around their mating-type locus (Menkis et al., 2008, Hartmann, Duhamel, et al., 2021, Hartmann, Ament-Velásquez, et al., 2021, Vittorelli et al., 2023). In basidiomycetes also, the species with evolutionary strata are dikaryotic and automictic, *e.g. Microbotryum fungi* and *Agaricus bisporus* var. *bisporus* (Branco et al., 2017, Branco et al., 2018, Foulongne-Oriol et al., 2021). This may be explained by the fact that intra-tetrad selfing favors the accumulation of deleterious alleles near the mating-type locus, which in turn can promote selection for recombination suppression because there will be more variability in the number of mutations present in a genomic region close to the mating-type locus, and therefore more fragments having a much lower number of deleterious mutations than average in the population (Jay et al., 2022).

Additionally, we found that the sheltering effect of a mating-type locus was stronger when the mutation was more strongly recessive. Indeed, the purging effect of selfing on partially recessive mutations is stronger for more recessive mutations (D. Charlesworth and Charlesworth, 1987, Caballero and Hill, 1992, Arunkumar et al., 2015), in which case the opposite force of the sheltering effect of a mating-type locus is strenghtened. This is in agreement with the results of studies on the sheltered load linked to a self-incompatibility locus, showing that completely recessive deleterious mutations are more easily fixed than partially recessive ones (Llaurens et al., 2009). This also confirms results on the fixation of inversions encompassing recessive deleterious mutations and linked to a permanently heterozygous locus (Olito et al., 2022, Jay et al., 2022). These results showed that inversions became fixed with a higher probability when segregating deleterious mutations were more strongly recessive.

Rare events of long maintenance of deleterious mutations in the population can occur

We further found that rare events of long maintenance of deleterious mutations in the population occurred under both selection scenarii. This shows that some deleterious mutations can persist in the population for an extended period of time before being purged, especially near the mating-type locus: in approximately 1% of our simulations, the purge of the deleterious mutation took several orders of magnitude longer than the 75% percentile empirically obtained from the 100,000. These surprisingly long purging times are likely to be due to the dynamics of the mutant being almost critical (the dominant eigenvalue in the branching process approximation is negative, but close to zero). However, from a modeling perspective very little is currently known about these trajectories, and more generally about the extinction time of multitype branching processes. Studying the extinction time of a deleterious allele in a one locus-two allele setting with a unitype branching process approximation and a diffusion approximation showed that the standard deviation of the mean extinction time was higher than the mean itself (Nei, 1971), which is

a feature that was also found in our simulations of multitype branching processes. These results show that the extinction time of deleterious alleles is highly variable, producing long-lasting mutations that may induce an accumulation of deleterious alleles near a mating-type locus, which is a prerequisite for recombination suppression to extend away from this locus (Jay et al., 2022).

The dynamics of deleterious mutations heavily relies on the mating system

Our results show that the mating system, and selfing in particular, is a prevailing force impacting the dynamics of deleterious mutations. Indeed, we found that a mating-type locus shelters mutations and thus favors their maintenance, but increasing selfing reduces the maintenance of mutations with a stronger effect. This result is congruent with previous studies showing that an increase in the selfing rate induces i) a reduction of the mutational load at a given locus or at multiple non-interacting loci far from mating-type compatibility loci (D. Charlesworth et al., 1990, for a deterministic model, Abu Awad and Roze, 2018, for diffusion approximation), and ii) a reduction of the purging time of deleterious mutations (Caballero and Hill, 1992).

However, we observed a particular behavior when the population reproduced only via selfing. Under complete selfing in our setting, the existence of a sheltering effect of a mating-type locus strongly depended on the values of the intra-tetrad selfing rate: the sheltering effect of the mating-type locus was detectable only when the intra-tetrad selfing coefficient exceeded a certain threshold, that depended on the dominance and selection coefficients. This strong effect of departing from complete selfing had previously been noted: introducing a small amount of outcrossing in a selfing population can lead to sharp changes in the dynamics of a deleterious mutation, whereas adding a small amount of selfing in an outcrossing population induces a smoother change (Holsinger and Feldman, 1985).

Limits of the methods

Our results are limited to the case of a single load locus, in interaction with a heterozygous mating-type locus, and may not apply when considering different frameworks, such as multiple epistatic loci or with additional beneficial mutations, especially regarding the impact of the mating system. Indeed, selfing has a non-monotonous effect depending on the tightness of linkage between multiple interacting loci (Abu Awad and Roze, 2018): at low selfing rates, increasing linkage between loci increases the mutation load, whereas the opposite effect is observed at high selfing rates. Selfing also has a non-monotonous effect on genetic variation in populations under stabilizing selection (Lande and Porcher, 2015, Clo and Opedal, 2021). In addition, selfing can enhance the fixation chances of a deleterious allele when it hitchhikes during a selective sweep (Hartfield and Otto, 2011, Hartfield and Glémin, 2014). Moreover, the impact of the mating system on the maintenance of deleterious mutations may be different if the number of individuals carrying the mutant allele exceeds a certain threshold. In this case, the branching process approximation does not hold anymore, and a deterministic model in large population may be used to further describe the dynamics of the deleterious allele (Durrett and Schweinsberg, 2004, Durrett, 2008 Section 6.1.3). The impact of the mating system then remains unclear: in large populations, selfing reduces the effective population size, which impairs the efficiency of selection and increases the mutational load of the population, but it also bolsters homozygosity, which favors the purge of deleterious mutations (Pollak, 1987, Caballero and Hill, 1992, D. Charlesworth and Wright, 2001, S. I. Wright et al., 2008).

Another limitation of our approach is that we considered a fixed recombination rate for simplicity, but allowing this rate to vary would allow us to test whether recombination suppression could evolve. Such an outcome may depend on the strength of selection against the deleterious mutation, as well as on the mating system (Antonovics and Abrams, 2004, Abu Awad and Roze, 2018). In some previous models, the impact of a modifier of recombination in the form of a multi-allelic locus was studied by simulations, but no analytical results were obtained (Feldman, 1972, Palsson, 2002, Antonovics and Abrams, 2004, Lenormand and Roze, 2022). The multitype branching process framework developed here would also be an interesting approach to obtain numerical results on this more complex situation, but analytical results would probably be out of reach because of the increase in complexity of the model.

Conclusion and Perspectives

In conclusion, our findings show that a mating-type locus has a sheltering effect on nearby deleterious mutations, especially in case of selfing and automixis, which can then play a role in the evolution of recombination suppression near mating-compatibility loci (Antonovics and Abrams, 2004, Jay et al., 2022). This may contribute to explain why evolutionary strata of recombination suppression near the mating-type locus are found mostly in automictic (pseudo-homothalic) fungi (Menkis et al., 2008, Branco et al., 2017, Branco et al., 2018, Hartmann et al., 2020, Hartmann, Ament-Velásquez, et al., 2021, Foulongne-Oriol et al., 2021, Vittorelli et al., 2023).

The results obtained here on the accumulation of deleterious mutations should apply, beyond fungal-like mating-type loci, to other permanently heterozygous loci, such as supergenes (Llaurens et al., 2017). In contrast, sporophytic or gametophytic plant self-incompatibility loci prevent diploid selfing, leading to a completely different evolutionary scenario in their flanking regions as imposed by complete outcrossing. The diversity of observed patterns regarding the presence or absence, length and number of evolutionary strata around these regions (Uyenoyama, 2005) may be explained, in addition to the mating system, by other factors controlling the long-term behavior of deleterious mutations which are not studied here, such as the number of alleles at supergenes, the length of the haploid phase (Jay et al., 2022), or the presence of multiple load loci that are possibly physically linked and with epistatic interactions (Abu Awad and Roze, 2018, Lenormand and Roze, 2022). The questions of the genome-wide impact of a mating-type locus, and of the interaction between a permanently heterozygous locus and background mutations, are currently debated (Abu Awad and Waller, 2023). The branching process framework developed here could be applied to diploid individuals carrying a load locus with two alleles, undergoing selfing or outcrossing, in order to investigate the dynamics of a new deleterious mutation in a population with or without a mating-type locus.

Our results showing the long maintenance of deleterious mutations in the vicinity of permanently heterozygous loci pave the way for future investigations on the accumulation of deleterious mutations. Previous studies (Coron et al., 2013, Coron, 2014) on mutational meltdown, showing that deleterious mutations accumulate faster when other mutations are already fixed, also encourage future work in this direction.

Acknowledgements

We thank Paul Jay and Denis Roze for insightful discussions, Aurélien Tellier for handling the recommendation process, and three anonymous reviewers for their useful and very constructive comments. Preprint version 2 of this article has been peer-reviewed and recommended by Peer Community In Evolutionary Biology (https://doi.org/10.24072/pci.evolbiol.100635).

Fundings

This work was supported by the European Research Council (ERC) EvolSexChrom (832352) grant to TG. ET, SB and AV acknowledge support from the chaire program « Mathematical modeling and biodiversity » (Ecole Polytechnique, Museum National d'Histoire Naturelle, Veolia Environnement, Fondation X).

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article. The authors declare the following non-financial conflict of interest: TG and SB are recommenders at PCIEvolBiol.

Data, script, code, and supplementary information availability

Script and codes are available online: https://doi.org/10.5281/zenodo.7464662

References

Abu Awad, D., & Billiard, S. (2017). The double edged sword: the demographic consequences of the evolution of self-fertilization. *Evolution*, *71*(5), 1178–1190. https://doi.org/10.1111/evo.13222

Abu Awad, D., & Roze, D. (2018). Effects of partial selfing on the equilibrium genetic variance, mutation load, and inbreeding depression under stabilizing selection. *Evolution*, *72*(4), 751–769. https://doi.org/10.1111/evo.13449

Abu Awad, D., & Waller, D. (2023). Conditions for maintaining and eroding pseudo-overdominance and its contribution to inbreeding depression. *Peer Community Journal*, *3*(e8). https://doi.org/10.24072/pcjournal.224

Antonovics, J., O'Keefe, K., & Hood, M. E. (1998). Theoretical population genetics of mating-type linked haplo-lethal alleles. *International Journal of Plant Sciences*, *159*(2), 192–198. https://doi.org/10.1086/297538

Antonovics, J., & Abrams, J. Y. (2004). Intratetrad mating and the evolution of linkage relationships. *Evolution*, *58*(4), 702–709. https://doi.org/10.1111/j.0014-3820.2004.tb00403.x

Arunkumar, R., Ness, R. W., Wright, S. I., & Barrett, S. C. H. (2015). The evolution of selfing is accompanied by reduced efficacy of selection and purging of deleterious mutations. *Genetics*, *199*(3), 817–829. https://doi.org/10.1534/genetics.114.172809

Athreya, K. B., & Ney, P. E. (1972). *Branching processes*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-65371-1

Bacaër, N. (2018). Sur l'extinction des populations avec plusieurs types dans un environnement aléatoire. *Comptes Rendus Biologies*, *341*(3), 145–151. https://doi.org/10.1016/j.crvi.2018.01.009

Bazzicalupo, A. L., Carpentier, F., Otto, S. P., & Giraud, T. (2019). Little evidence of antagonistic selection in the evolutionary strata of fungal mating-type chromosomes (*Microbotryum lychnidis-dioicae*). *G3 Genes|Genomes|Genetics*, *9*(6), 1987–1998. https://doi.org/10.1534/g3.119.400242

Bergero, R., & Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends in Ecology & Evolution*, *24*(2), 94–102. https://doi.org/10.1016/j.tree.2008.09.010

Billiard, S., López-Villavicencio, M., Hood, M. E., & Giraud, T. (2012). Sex, outcrossing and mating types: unsolved questions in fungi and beyond: sexy fungi. *Journal of Evolutionary Biology*, *25*(6), 1020–1038. https://doi.org/10.1111/j.1420-9101.2012.02495.x

Branco, S., Badouin, H., Rodríguez de la Vega, R. C., Gouzy, J., Carpentier, F., Aguileta, G., Siguenza, S., Brandenburg, J.-T., Coelho, M. A., Hood, M. E., & Giraud, T. (2017). Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proceedings of the National Academy of Sciences*, *114*(27), 7067–7072. https://doi.org/10.1073/pnas.1701658114

Branco, S., Carpentier, F., Rodríguez de la Vega, R. C., Badouin, H., Snirc, A., Le Prieur, S., Coelho, M. A., de Vienne, D. M., Hartmann, F. E., Begerow, D., Hood, M. E., & Giraud, T. (2018). Multiple convergent supergene evolution events in mating-type chromosomes. *Nature Communications*, *9*(1), 2000. https://doi.org/10.1038/s41467-018-04380-9

Bürger, R. (2020). Multilocus population-genetic theory. *Theoretical Population Biology*, *9*. https://doi.org/https://doi.org/10.1016/j.tpb.2019.09.004

Caballero, A., & Hill, W. G. (1992). Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics*, *131*(2), 493–507. https://doi.org/10.1093/genetics/131.2.493

Champagnat, N., & Méléard, S. (2011). Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields*, *151*(1), 45–94. https://doi.org/10.1007/s00440-010-0292-9

Charlesworth, B., & Wall, J. D. (1999). Inbreeding, heterozygote advantage and the evolution of neo–X and neo–Y sex chromosomes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1414), 51–56. https://doi.org/10.1098/rspb.1999.0603

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, *18*, 237–268. https://doi.org/10.1146/annurev.es.18.110187.001321

Charlesworth, D., Morgan, M. T., & Charlesworth, B. (1990). Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution*, *44*(6), 1469–1489. https://doi.org/10.1111/j.1558-5646.1990.tb03839.x

Charlesworth, D., & Wright, S. I. (2001). Breeding systems and genome evolution. *Current Opinion in Genetics & Development*, *11*(6), 685–690. https://doi.org/10.1016/S0959-437X(00)00254-9

Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, *95*(2), 118–128. https://doi.org/10.1038/sj.hdy.6800697

Clo, J., & Opedal, Ø. H. (2021). Genetics of quantitative traits with dominance under stabilizing and directional selection in partially selfing species. *Evolution*, *75*(8), 1920–1935. https://doi.org/10.1111/evo.14304

Collet, P., Méléard, S., & Metz, J. A. J. (2013). A rigorous model study of the adaptive dynamics of Mendelian diploids. *Journal of Mathematical Biology*, *67*(3), 569–607. https://doi.org/10.1007/s00285-012-0562-5

Coron, C., Méléard, S., Porcher, E., & Robert, A. (2013). Quantifying the mutational meltdown in diploid populations. *The American Naturalist*, *181*(5), 623–636. https://doi.org/10.1086/670022

Coron, C. (2014). Stochastic modeling of density-dependent diploid populations and the extinction vortex. *Advances in Applied Probability*, *46*(2), 446–477. https://doi.org/10.1239/aap/1401369702

Dagilis, A. J., Sardell, J. M., Josephson, M. P., Su, Y., Kirkpatrick, M., & Peichel, C. L. (2022). Searching for signatures of sexually antagonistic selection on stickleback sex chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1856), 20210205. https://doi.org/10.1098/rstb.2021.0205

Durrett, R., & Schweinsberg, J. (2004). Approximating selective sweeps. *Theoretical Population Biology*, *66*(2), 129–138. https://doi.org/10.1016/j.tpb.2004.04.002

Durrett, R. (2008). *Probability models for DNA sequence evolution*. Springer New York. https://doi.org/10.1007/978-0-387-78168-6

Ewens, W. J. (1967). The probability of fixation of a mutant : two-locus case. *Evolution*, *21*(3), 532–540. https://doi.org/10.1111/j.1558-5646.1967.tb03409.x

Ewens, W. J. (1968). Some applications of multiple-type branching processes in population genetics. *Journal of the Royal Statistical Society: Series B (Methodological)*, *30*(1), 164–175. https://doi.org/10.1111/j.2517-6161.1968.tb01515.x

Ewens, W. J. (2004). *Mathematical population genetics* (Vol. 27). Springer New York. https://doi.org/10.1007/978-0-387-21822-9

Feldman, M. W. (1972). Selection for linkage modification: i. Random mating populations. *Theoretical Population Biology*, *3*(3), 324–346. https://doi.org/10.1016/0040-5809(72)90007-X

Foulongne-Oriol, M., Taskent, O., Kües, U., Sonnenberg, A. S. M., van Peer, A. F., & Giraud, T. (2021). Mating-type locus organization and mating-type chromosome differentiation in the bipolar edible button mushroom *Agaricus bisporus*. *Genes*, *12*(7), 1079. https://doi.org/10.3390/genes12071079

Fraser, J. A., Diezmann, S., Subaran, R. L., Allen, A., Lengeler, K. B., Dietrich, F. S., & Heitman, J. (2004). Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biology*, *2*(12), e384. https://doi.org/10.1371/journal.pbio.0020384

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, *22*(4), 403–434. https://doi.org/10.1016/0021-9991(76)90041-3

Glémin, S. (2007). Mating systems and the efficacy of selection at the molecular level. *Genetics*, *177*(2), 905–916. https://doi.org/10.1534/genetics.107.073601

Harris, T. E. (1964). *The theory of branching processes* [OCLC: 931719544]. Springer-Verlag.
Hartfield, M., & Otto, S. P. (2011). Recombination and hitchhiking of deleterious alleles: recombination and the undesirable hitchhiker. *Evolution*, *65*(9), 2421–2434. https://doi.org/10.1111/j.1558-5646.2011.01311.x
Hartfield, M., & Glémin, S. (2014). Hitchhiking of deleterious alleles and the cost of adaptation in partially selfing species. *Genetics*, *196*(1), 281–293. https://doi.org/10.1534/genetics.113.158196
Hartmann, F. E., Rodríguez de la Vega, R. C., Gladieux, P., Ma, W.-J., Hood, M. E., & Giraud, T. (2020). Higher gene flow in sex-related chromosomes than in autosomes during fungal divergence. *Molecular Biology and Evolution*, *37*(3), 668–682. https://doi.org/10.1093/molbev/msz252
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
Hartmann, F. E., Ament-Velásquez, S. L., Vogan, A. A., Gautier, V., Le Prieur, S., Berramdane, M., Snirc, A., Johannesson, H., Grognet, P., Malagnac, F., Silar, P., & Giraud, T. (2021). Size variation of the nonrecombining region on the mating-type chromosomes in the fungal *Podospora anserina* species complex. *Molecular Biology and Evolution*, *38*(6), 2475–2492. https://doi.org/10.1093/molbev/msab040
Heinzmann, D. (2009). Extinction times in multitype markov branching processes. *Journal of Applied Probability*, *46*(1), 296–307. https://doi.org/10.1239/jap/1238592131
Holsinger, K. E., & Feldman, M. W. (1985). Selection in complex genetic systems. VI. Equilibrium properties of two locus selection models with partial selfing. *Theoretical Population Biology*, *28*(1), 117–132. https://doi.org/10.1016/0040-5809(85)90024-3
Hood, M. E., & Antonovics, J. (2000). Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum* (=*Ustilago violacea*). *Heredity*, *85*(3), 231–241. https://doi.org/10.1046/j.1365-2540.2000.00748.x
Ironside, J. E. (2010). No amicable divorce? Challenging the notion that sexual antagonism drives sex chromosome evolution. *BioEssays*, *32*(8), 718–726. https://doi.org/10.1002/bies.200900124
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., 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
Jeffries, D. L., Gerchen, J. F., Scharmann, M., & Pannell, J. R. (2021). A neutral model for the loss of recombination on sex chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1832), 20200096. https://doi.org/10.1098/rstb.2020.0096
Jin, Y., Allan, S., Baber, L., Bhattarai, E. K., Lamb, T. M., & Versaw, W. K. (2007). Rapid genetic mapping in *Neurospora crassa*. *Fungal Genetics and Biology*, *44*(6), 455–465. https://doi.org/10.1016/j.fgb.2006.09.002
Karlin, S. (1975). General two-locus selection models: some objectives, results and interpretations. *Theoretical Population Biology*, *7*(3), 364–398. https://doi.org/10.1016/0040-5809(75)90025-8
Kesten, H., & Stigum, B. P. (1966). A limit theorem for multidimensional Galton-Watson processes. *The Annals of Mathematical Statistics*, *37*(5), 1211–1223. https://doi.org/10.1214/aoms/1177699266

Kimura, M. (1980). Average time until fixation of a mutant allele in a finite population under continued mutation pressure: studies by analytical, numerical, and pseudo-sampling methods. *Proceedings of the National Academy of Sciences*, *77*(1), 522–526. https://doi.org/10.1073/pnas.77.1.522

Kratochvíl, L., & Stöck, M. (2021). Preface. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1832), 20200088. https://doi.org/10.1098/rstb.2020.0088

Kuhn, M.-L., Gout, L., Howlett, B. J., Melayah, D., Meyer, M., Balesdent, M.-H., & Rouxel, T. (2006). Genetic linkage maps and genomic organization in *Leptosphaeria maculans*. *European Journal of Plant Pathology*, *114*(1), 17–31. https://doi.org/10.1007/s10658-005-3168-6

Lande, R., & Porcher, E. (2015). Maintenance of quantitative genetic variance under partial self-fertilization, with implications for evolution of selfing. *Genetics*, *200*(3), 891–906. https://doi.org/10.1534/genetics.115.176693

Leach, C. R., Mayo, O., & Morris, M. M. (1986). Linkage disequilibrium and gametophytic self-incompatibility. *Theoretical and Applied Genetics*, *73*(1), 102–112. https://doi.org/10.1007/BF00273726

Lenormand, T., & Roze, D. (2022). Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science*, *375*(6581), 663–666. https://doi.org/10.1126/science.abj1813

Lenz, T. L., Spirin, V., Jordan, D. M., & Sunyaev, S. R. (2016). Excess of deleterious mutations around HLA genes reveals evolutionary cost of balancing selection. *Molecular Biology and Evolution*, *33*(10), 2555–2564. https://doi.org/10.1093/molbev/msw127

Llaurens, V., Gonthier, L., & Billiard, S. (2009). The sheltered genetic load linked to the *S* locus in plants: new insights from theoretical and empirical approaches in sporophytic self-incompatibility. *Genetics*, *183*(3), 1105–1118. https://doi.org/10.1534/genetics.109.102707

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

Malkus, A., Song, Q., Cregan, P., Arseniuk, E., & Ueng, P. P. (2009). Genetic linkage map of *Phaeosphaeria nodorum*, the causal agent of stagonospora nodorum blotch disease of wheat. *European Journal of Plant Pathology*, *124*(4), 681–690. https://doi.org/10.1007/s10658-009-9454-y

Menkis, A., Jacobson, D. J., Gustafsson, T., & Johannesson, H. (2008). The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *PLoS Genetics*, *4*(3), e1000030. https://doi.org/10.1371/journal.pgen.1000030

Mode, C. J. (1971). *Multitype branching processes: theory and applications*. American Elsevier Pub. Co.

Nei, M. (1971). Extinction time of deleterious mutant genes in large populations. *Theoretical Population Biology*, *2*(4), 419–425. https://doi.org/10.1016/0040-5809(71)90030-X

Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., Mouchiroud, D., Negrutiu, I., Charlesworth, D., & Monéger, F. (2004). A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biology*, *3*(1), e4. https://doi.org/10.1371/journal.pbio.0030004

Ohta, T., & Cockerham, C. C. (1974). Detrimental genes with partial selfing and effects on a neutral locus. *Genetical Research*, *23*(2), 191–200. https://doi.org/10.1017/S0016672300014816

Olito, C., Ponnikas, S., Hansson, B., & Abbott, J. K. (2022). Consequences of partially recessive deleterious genetic variation for the evolution of inversions suppressing recombination between sex chromosomes. *Evolution*, *76*(6), 1320–1330. https://doi.org/10.1111/evo.14496

Palsson, S. (2002). Selection on a modifier of recombination rate due to linked deleterious mutations. *Journal of Heredity*, *93*(1), 22–26. https://doi.org/10.1093/jhered/93.1.22

Pénisson, S. (2010). *Conditional limit theorems for multitype branching processes and illustration in epidemiological risk analysis* (Doctoral dissertation). Universität Potsdam, Université Paris-Sud.

Phan, H. T. T., Ford, R., & Taylor, P. W. J. (2003). Mapping the mating type locus of *Ascochyta rabiei*, the causal agent of ascochyta blight of chickpea. *Molecular Plant Pathology*, *4*(5), 373–381. https://doi.org/10.1046/j.1364-3703.2003.00185.x

Pollak, E. (1987). On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics*, *117*(2), 353–360. https://doi.org/10.1093/genetics/117.2.353

Pollak, E., & Sabran, M. (1992). On the theory of partially inbreeding finite populations. III. Fixation probabilities under partial selfing when heterozygotes are intermediate in viability. *Genetics*, *131*(4), 979–985. https://doi.org/10.1093/genetics/131.4.979

Pollard, J. H. (1966). On the use of the direct matrix product in analysing certain stochastic population models. *Biometrika*, *53*(3), 397. https://doi.org/10.1093/biomet/53.3-4.397

Pollard, J. H. (1968). The multi-type galton-watson process in a genetical context. *Biometrics*, *24*(1), 147. https://doi.org/10.2307/2528466

Ponnikas, S., Sigeman, H., Abbott, J. K., & Hansson, B. (2018). Why do sex chromosomes stop recombining? *Trends in Genetics*, *34*(7), 492–503. https://doi.org/10.1016/j.tig.2018.04.001

Pötscher, B. M. (1985). Moments and order statistics of extinction times in multitype branching processes and their relation to random selection models. *Bulletin of Mathematical Biology*, *47*(2), 263–272. https://doi.org/10.1007/BF02460035

Rice, S. H. (2004). *Evolutionary theory: mathematical and conceptual foundations*. Sinauer Associates.

Rice, W. R. (1987). The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution*, *41*(4), 911–914. https://doi.org/10.1111/j.1558-5646.1987.tb05864.x

Ruzicka, F., Dutoit, L., Czuppon, P., Jordan, C. Y., Li, X.-Y., Olito, C., Runemark, A., Svensson, E. I., Yazdi, H. P., & Connallon, T. (2020). The search for sexually antagonistic genes: practical insights from studies of local adaptation and statistical genomics. *Evolution Letters*, *4*(5), 398–415. https://doi.org/10.1002/evl3.192

Sewastjanow, B. (1975). *Verzweigungsprozesse* (R. Oldenbourg Verlag).

Skinner, D. Z., Budde, A. D., Farman, M. L., Smith, J. R., Leung, H., & Leong, S. A. (1993). Genome organization of *Magnaporthe grisea*: genetic map, electrophoretic karyotype, and occurrence of repeated DNAs. *Theoretical and Applied Genetics*, *87*(5), 545–557. https://doi.org/10.1007/BF00221877

Tezenas, E. (2022). *Mutations near mating-type locus codes and data (mut_near_mat_v2)*. Zenodo. https://doi.org/10.5281/zenodo.7464662

Uyenoyama, M. K. (2005). Evolution under tight linkage to mating type. *New Phytologist*, *165*(1), 63–70. https://doi.org/10.1111/j.1469-8137.2004.01246.x

Vittorelli, N., Snirc, A., Levert, E., Gautier, V., Lalanne, C., De Filippo, E., Rodríguez de la Vega, R. C., Gladieux, P., Guillou, S., Zhang, Y., Tejomurthula, S., Grigoriev, I. V., Debuchy, R., Silar, P., Giraud, T., & Hartmann, F. E. (2023). Stepwise recombination suppression around the mating-type locus in the fungus *Schizothecium tetrasporum* (ascomycota, sordariales). *Plos Genetics (in press)*. https://doi.org/10.1101/2022.07.20.500756

Wolfram Research, I. (2015). *Mathematica* (Version 10.1). Champaign, IL, 2015.

Wright, A. E., Dean, R., Zimmer, F., & Mank, J. E. (2016). How to make a sex chromosome. *Nature Communications*, *7*(1), 12087. https://doi.org/10.1038/ncomms12087

Wright, S. I., Ness, R. W., Foxe, J. P., & Barrett, S. C. H. (2008). Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences*, *169*(1), 105–118. https://doi.org/10.1086/523366

Yan, Z., Martin, S. H., Gotzek, D., Arsenault, S. V., Duchen, P., Helleu, Q., Riba-Grognuz, O., Hunt, B. G., Salamin, N., Shoemaker, D., Ross, K. G., & Keller, L. (2020). Evolution of a supergene that regulates a trans-species social polymorphism. *Nature Ecology & Evolution*, *4*(2), 240–249. https://doi.org/10.1038/s41559-019-1081-1

Zhong, S., Steffenson, B. J., Martinez, J. P., & Ciuffetti, L. M. (2002). A molecular genetic map and electrophoretic karyotype of the plant pathogenic fungus *Cochliobolus sativus*. *Molecular Plant-Microbe Interactions*, *15*(5), 481–492. https://doi.org/10.1094/MPMI.2002.15.5.481

Appendix A. Table of notation

Table S1

N	Population size
$G_1, \dots, G_4$	Genotypes
$(g_1, \dots, g_4)$	Number of individuals of each genotype
f	Selfing probability
p_{in} and $p_{out} = 1 - p_{in}$	Intra- and Inter-tetrad selfing probabilities
r	Recombination rate
S_i	Probability of survival of an offspring of genotype $i \in \{1, 2, 3, 4\}$ (see Figure 1)
s	Selection coefficient in the partial dominance case
h	Dominance coefficient in the partial dominance case
s_3, s_4	Selection coefficients in the overdominance case

Appendix B. Intra-, Inter-tetrad selfing and outcrossing

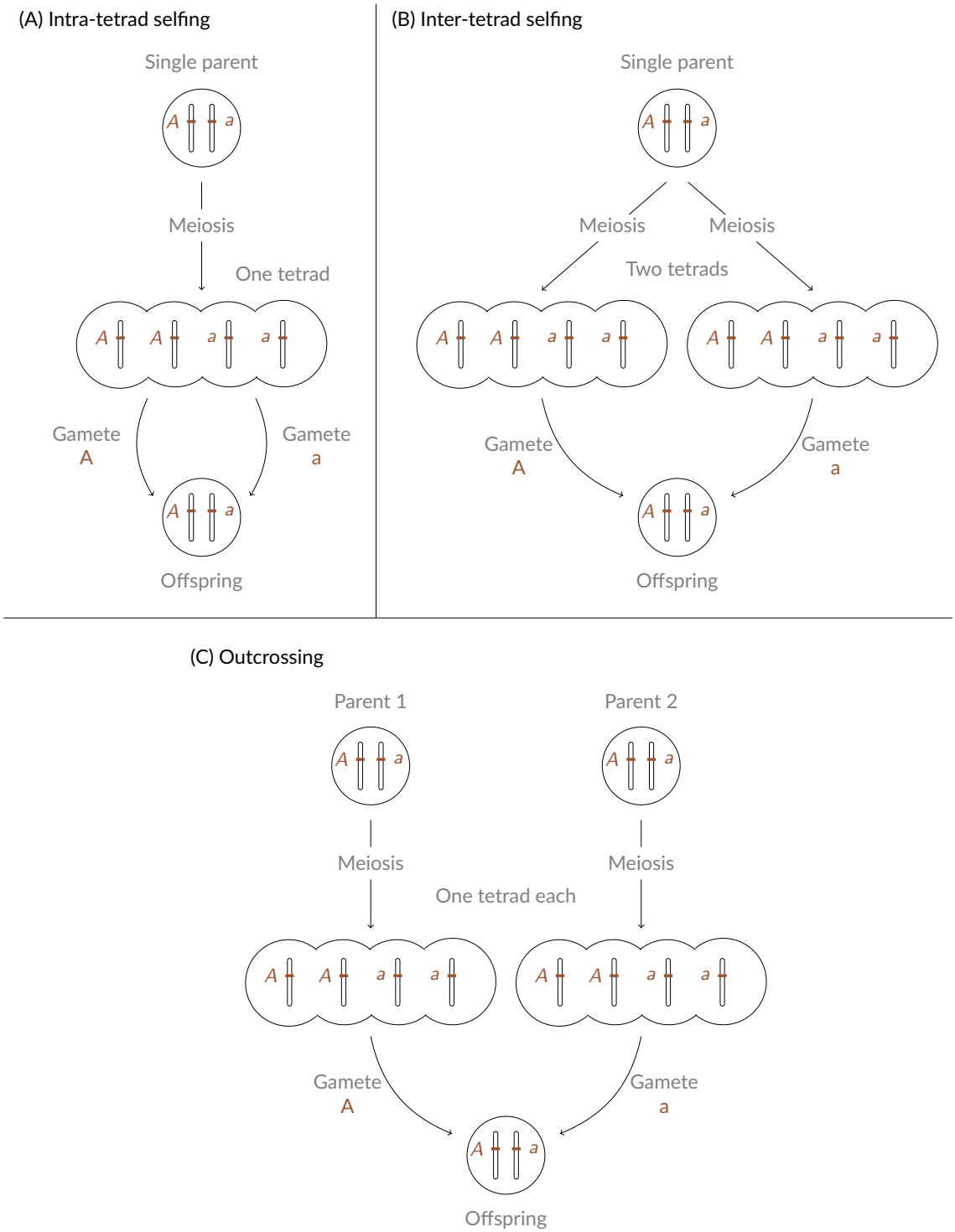


Figure S1 – Schematic representation of the three mating systems considered in the model. Individuals are represented by a pair of mating-type chromosomes, with the mating-type locus displayed. A diploid offspring is generated by the fusion of two gametes carrying different mating-type alleles (A and a). (A) Under intra-tetrad selfing, both gametes are picked from the same tetrad; only one parent is involved. (B) Under inter-tetrad selfing, the two gametes are picked from two different tetrads (meioses) produced by the same diploid parent; only one parent is involved. (C) Under outcrossing, the two gametes are picked in tetrads produced by different parents.

Appendix C. Appendices for the Method section

C.1. Rates of creation of offspring with given genotypes (Moran process)

Parental Geno-types	Intra /Inter - Tetrad	Recombination	Genotype of offspring			
			G_1	G_2	G_3	G_4
G_1 fg_1	Intra p_{in}	(A) $(1-r)$	$fg_1 p_{in}(1-r)$	0	0	0
		(A) r	$\frac{1}{4} fg_1 p_{in} r$	$\frac{1}{4} fg_1 p_{in} r$	$\frac{1}{4} fg_1 p_{in} r$	$\frac{1}{4} fg_1 p_{in} r$
	Inter p_{out}	(AA) $(1-r)^2$	$fg_1 p_{out}(1-r)^2$	0	0	0
		(AP) $2(1-r)r$	$\frac{1}{2} 2 fg_1 p_{out}(1-r)r$	0	$\frac{1}{4} 2 fg_1 p_{out}(1-r)r$	$\frac{1}{4} 2 fg_1 p_{out}(1-r)r$
		(PP) r^2	$\frac{1}{4} fg_1 p_{out} r^2$	$\frac{1}{4} fg_1 p_{out} r^2$	$\frac{1}{4} fg_1 p_{out} r^2$	$\frac{1}{4} fg_1 p_{out} r^2$
G_2 fg_2	Intra p_{in}	(A) $(1-r)$	0	$fg_2 p_{in}(1-r)$	0	0
		(P) r	$\frac{1}{4} fg_2 p_{in} r$	$\frac{1}{4} fg_2 p_{in} r$	$\frac{1}{4} fg_2 p_{in} r$	$\frac{1}{4} fg_2 p_{in} r$
	Inter p_{out}	(AA) $(1-r)^2$	0	$fg_2 p_{out}(1-r)^2$	0	0
		(AP) $2(1-r)r$	0	$\frac{1}{2} 2 fg_2 p_{out}(1-r)r$	$\frac{1}{4} 2 fg_2 p_{out}(1-r)r$	$\frac{1}{4} 2 fg_2 p_{out}(1-r)r$
		(PP) r^2	$\frac{1}{4} fg_2 p_{out} r^2$	$\frac{1}{4} fg_2 p_{out} r^2$	$\frac{1}{4} fg_2 p_{out} r^2$	$\frac{1}{4} fg_2 p_{out} r^2$
G_3 fg_3	Same tetrad (homozyg.)		0	0	fg_3	0
G_4 fg_4	Same tetrad (homozyg.)		0	0	0	fg_4

Table S2 – Table summarizing the rates of production of an offspring of each genotype (last four columns) in case of **selfing**. *Parental Genotype*: The genotype of the individual involved in the mating event; *Intra/Inter-tetrad*: Mating through intra- of inter-tetrad selfing (see section 2.1 for definitions); *Recombination*: Occurrence of a recombination event in the tetrads from which gametes are picked. "A" stands for "Absence" in one tetrad, "P" stands for "Presence" in one tetrad. We use only one letter when the two gametes come from the same tetrad or when one of the genotypes involved is homozygous at the load locus. For example, (AP) indicates that recombination occured in one tetrad but not in the other. G_i: the rate at which an offspring of genotype G_i is produced, due to the scenario of parental genotype, intra/inter tetrad selfing and presence/absence of recombination considered. The total rate $T_g(+G_i)$ at which a new offspring of genotype G_i is created when the population state is $g = (g_1, g_2, g_3, g_4)$ is then the sum of all the rates appearing in column G_i in this Table, Table S3 and Table S4.

Parental Genotypes	Recombination		Genotype of offspring			
			G_1	G_2	G_3	G_4
$G_1\,G_1$ $(1-f)\times g_1\frac{g_1-1}{N-1}$	(AA)	$(1-r)^2$	$(1-f)g_1\frac{g_1-1}{N-1}(1-r)^2$	0	0	0
	(AP)	$2(1-r)r$	$\frac{1}{2}2(1-f)g_1\frac{g_1-1}{N-1}(1-r)r$	0	$\frac{1}{4}2(1-f)g_1\frac{g_1-1}{N-1}(1-r)r$	$\frac{1}{4}2(1-f)g_1\frac{g_1-1}{N-1}(1-r)r$
	(PP)	r^2	$\frac{1}{4}(1-f)g_1\frac{g_1-1}{N-1}r^2$	$\frac{1}{4}(1-f)g_1\frac{g_1-1}{N-1}r^2$	$\frac{1}{4}(1-f)g_1\frac{g_1-1}{N-1}r^2$	$\frac{1}{4}(1-f)g_1\frac{g_1-1}{N-1}r^2$
$G_1\,G_2$ $2(1-f)\times g_1\frac{g_2}{N-1}$	(AA)	$(1-r)^2$	0	0	$\frac{1}{2}2(1-f)g_1\frac{g_2}{N-1}(1-r)^2$	$\frac{1}{2}2(1-f)g_1\frac{g_2}{N-1}(1-r)^2$
	(AP)	$(1-r)r$	$\frac{1}{2}2(1-f)g_1\frac{g_2}{N-1}(1-r)r$	0	$\frac{1}{4}2(1-f)g_1\frac{g_2}{N-1}(1-r)r$	$\frac{1}{4}2(1-f)g_1\frac{g_2}{N-1}(1-r)r$
	(PA)	$r(1-r)$	0	$\frac{1}{2}2(1-f)g_1\frac{g_2}{N-1}r(1-r)$	$\frac{1}{4}2(1-f)g_1\frac{g_2}{N-1}r(1-r)$	$\frac{1}{4}2(1-f)g_1\frac{g_2}{N-1}r(1-r)$
	(PP)	r^2	$\frac{1}{4}2(1-f)g_1\frac{g_2}{N-1}r^2$	$\frac{1}{4}2(1-f)g_1\frac{g_2}{N-1}r^2$	$\frac{1}{4}2(1-f)g_1\frac{g_2}{N-1}r^2$	$\frac{1}{4}2(1-f)g_1\frac{g_2}{N-1}r^2$
$G_1\,G_3$ $2(1-f)\times g_1\frac{g_3}{N-1}$	(A)	$1-r$	$\frac{1}{2}2(1-f)g_1\frac{g_3}{N-1}(1-r)$	0	$\frac{1}{2}2(1-f)g_1\frac{g_3}{N-1}(1-r)$	0
	(P)	r	$\frac{1}{4}2(1-f)g_1\frac{g_3}{N-1}r$	$\frac{1}{4}2(1-f)g_1\frac{g_3}{N-1}r$	$\frac{1}{2}2(1-f)g_1\frac{g_3}{N-1}r$	0
$G_1\,G_4$ $2(1-f)\times g_1\frac{g_4}{N-1}$	(A)	$1-r$	$\frac{1}{2}2(1-f)g_1\frac{g_4}{N-1}(1-r)$	0	0	$\frac{1}{2}2(1-f)g_1\frac{g_4}{N-1}(1-r)$
	(P)	r	$\frac{1}{4}2(1-f)g_1\frac{g_4}{N-1}r$	$\frac{1}{4}2(1-f)g_1\frac{g_4}{N-1}r$	0	$\frac{1}{2}2(1-f)g_1\frac{g_4}{N-1}r$

Table S3 – Part 1 of the table summarizing the rates of production of an offspring of each genotype (last four columns) in case of **outcrossing**. *Parental Genotype*: The genotype of the individuals involved in the mating event; *Recombination*: Occurrence of a recombination event in the tetrads from which gametes are picked. "A" stands for "Absence" in one tetrad, "P" stands for "Presence" in one tetrad. We use only one letter when the two gametes come from the same tetrad or when one of the genotypes involved is homozygous at the load locus. For example, (AP) indicates that recombination occured in one tetrad but not the other. G_i: the rate at which an offspring of genotype G_i is produced, due to the scenario of parental genotype, intra/inter tetrad selfing and presence/absence of recombination considered. The total rate $T_g(+G_i)$ at which a new offspring of genotype G_i is created when the population state is $g = (g_1, g_2, g_3, g_4)$ is then the sum of all the rates appearing in column G_i in this Table, Table S2 and Table S4.

Parental Genotypes	Recombination	Genotype of offspring			
		G_1	G_2	G_3	G_4
G_2G_2 $(1-f)\times g_2\frac{g_2-1}{N-1}$	(AA) $(1-r)^2$	0	$(1-f)g_2\frac{g_2-1}{N-1}(1-r)^2$	0	0
	(AP) $2(1-r)r$	0	$\frac{1}{2}2(1-f)g_2\frac{g_2-1}{N-1}(1-r)r$	$\frac{1}{4}2(1-f)g_2\frac{g_2-1}{N-1}(1-r)r$	$\frac{1}{4}2(1-f)g_2\frac{g_2-1}{N-1}(1-r)r$
	(PP) r^2	$\frac{1}{4}(1-f)g_2\frac{g_2-1}{N-1}r^2$	$\frac{1}{4}(1-f)g_2\frac{g_2-1}{N-1}r^2$	$\frac{1}{4}(1-f)g_2\frac{g_2-1}{N-1}r^2$	$\frac{1}{4}(1-f)g_2\frac{g_2-1}{N-1}r^2$
G_2G_3 $2(1-f)\times g_2\frac{g_3}{N-1}$	(A) $1-r$	0	$\frac{1}{2}(1-f)g_2\frac{g_3}{N-1}(1-r)$	$\frac{1}{2}2(1-f)g_2\frac{g_3}{N-1}(1-r)$	0
	(P) r	$\frac{1}{4}2(1-f)g_2\frac{g_3}{N-1}r$	$\frac{1}{4}2(1-f)g_2\frac{g_3}{N-1}r$	$\frac{1}{2}2(1-f)g_2\frac{g_3}{N-1}r$	0
G_2G_4 $2(1-f)\times g_2\frac{g_4}{N-1}$	(A) $1-r$	0	$\frac{1}{2}2(1-f)g_2\frac{g_4}{N-1}(1-r)$	0	$\frac{1}{2}2(1-f)g_2\frac{g_4}{N-1}(1-r)$
	(P) r	$\frac{1}{4}2(1-f)g_2\frac{g_4}{N-1}r$	$\frac{1}{4}2(1-f)g_2\frac{g_4}{N-1}r$	0	$\frac{1}{2}2(1-f)g_2\frac{g_4}{N-1}r$
G_3G_3 $(1-f)\times g_3\frac{g_3-1}{N-1}$	Same tetrads	0	0	$(1-f)g_3\frac{g_3-1}{N-1}$	0
G_3G_4 $2(1-f)\times g_3\frac{g_4}{N-1}$	Same tetrads	$\frac{1}{2}2(1-f)g_3\frac{g_4}{N-1}$	$\frac{1}{2}2(1-f)g_3\frac{g_4}{N-1}$	0	0
G_4G_4 $(1-f)\times g_4\frac{g_4-1}{N-1}$	Same tetrads	0	0	0	$(1-f)g_4\frac{g_4-1}{N-1}$

Table S4 – Part 2 of the table summarizing the rates of production of an offspring of each genotype (last four columns) in case of **outcrossing**. *Parental Genotype*: The genotype of the individuals involved in the mating; *Recombination*: Occurrence of a recombination event in the tetrads from which gametes are picked. "A" stands for "Absence" in one tetrad, "P" stands for "Presence" in one tetrad. We use only one letter when the two gametes come from the same tetrad or when one of the genotypes involved is homozygous at the load locus. For example, (AP) indicates that recombination occured in one tetrad but not the other. G_i: the rate at which an offspring of genotype G_i is produced, due to the scenario of parental genotype, intra/inter tetrad selfing and presence/absence of recombination considered. The total rate $T_g(+G_i)$ at which a new offspring of genotype G_i is created when the population state is $g=(g_1,g_2,g_3,g_4)$ is then the sum of all the rates appearing in column G_i in this Table, Table S2 and Table S3.

The total rate at which an offspring of a given genotype is produced is then obtained by summing the rates along each column G_i in Tables S2, S3 and S4. This gives:

$$\begin{aligned}T_g(+G_1) = {} & fg_1\left(1-r+\frac{r}{4}\Big(1-(1-p_{in})(1-r)\Big)\right)+fg_2\frac{r}{4}\Big(1-(1-p_{in})(1-r)\Big)\\ & +\frac{1-f}{N-1}\Bigg[g_1\left(1-\frac{r}{2}\right)\left((g_1-1)\left(1-\frac{r}{2}\right)+g_3+g_4\right)+g_2r\left((g_2-1)\frac{r}{4}+\frac{1}{2}(g_3+g_4)\right)\\ & +g_1g_2r\left(1-\frac{r}{2}\right)+g_3g_4\Bigg],\end{aligned}$$

$$\begin{aligned}T_g(+G_2) = {} & fg_1\frac{r}{4}\Big(1-(1-p_{in})(1-r)\Big)+fg_2\left(1-r+\frac{r}{4}\Big(1-(1-p_{in})(1-r)\Big)\right)\\ & +\frac{1-f}{N-1}\Bigg[g_2\left(1-\frac{r}{2}\right)\left((g_2-1)\left(1-\frac{r}{2}\right)+g_3+g_4\right)+g_1r\left((g_1-1)\frac{r}{4}+\frac{1}{2}(g_3+g_4)\right)\\ & +g_1g_2r\left(1-\frac{r}{2}\right)+g_3g_4\Bigg],\end{aligned}$$

$$\begin{aligned}T_g(+G_3) = {} & fg_1\frac{r}{4}\Big(1+(1-p_{in})(1-r)\Big)+fg_2\frac{r}{4}\Big(1+(1-p_{in})(1-r)\Big)+fg_3\\ & +\frac{1-f}{N-1}\Bigg[g_1(g_1-1)\frac{r}{2}\left(1-\frac{r}{2}\right)+g_2(g_2-1)\frac{r}{2}\left(1-\frac{r}{2}\right)+g_1g_2\left(1-r+\frac{r^2}{2}\right)\\ & +g_3(g_1+g_2+(g_3-1))\Bigg],\end{aligned}$$

$$\begin{aligned}T_g(+G_4) = {} & fg_1\frac{r}{4}\Big(1+(1-p_{in})(1-r)\Big)+fg_2\frac{r}{4}\Big(1+(1-p_{in})(1-r)\Big)+fg_4\\ & +\frac{1-f}{N-1}\Bigg[g_1(g_1-1)\frac{r}{2}\left(1-\frac{r}{2}\right)+g_2(g_2-1)\frac{r}{2}\left(1-\frac{r}{2}\right)+g_1g_2\left(1-r+\frac{r^2}{2}\right)\\ & +g_4(g_1+g_2+(g_4-1))\Bigg].\end{aligned}$$

C.2. Reproduction law for the branching process

We give here an example of how the reproduction laws for the branching process are derived from the rates of the Moran process, using the approximate regime (1).

Let us derive the coefficient A_{12} of the matrix A, which is the rate at which an individual of genotype G_2 generates an offspring of genotype G_1 and survives. Equivalently, this is the rate at which an individual of genotype G_2 generates a descendance vector equal to e_1+e_2.

Using the rates obtained for the Moran model, the rate at which an individual of genotype G_2 produces an offspring of genotype G_1 is:

$$(5)\quad f\frac{r}{4}\Big(1-(1-p_{in})(1-r)\Big)+(1-f)\left[r\left(\frac{g_2-1}{N-1}\times\frac{r}{4}+\frac{1}{2}\left(\frac{g_3}{N-1}+\frac{g_4}{N-1}\right)\right)+\frac{g_1}{N-1}r\left(1-\frac{r}{2}\right)\right].$$

The first term, with a factor f, is the rate at which an individual of genotype G_2 produces an offspring of genotype G_1 by selfing. The second term, with a factor $1-f$, is the rate at which an individual of genotype G_2 produces an offspring of genotype G_1 by outcrossing. In this term, the fractions of the form $\frac{g_i}{N-1}$ represent the probabilities that an individual of genotype G_i is chosen to mate with the G_2 parent.

Using the approximation (1), *i.e.* assuming that $g_4 \approx N$ and $g_i \ll N$ for $i=1,2,3$, we obtain that the quantity in Eq. (5) can be approximated by:

$$f\frac{r}{4}\Big(1-(1-p_{in})(1-r)\Big)+(1-f)\frac{r}{2}.$$

To obtain A_{12}, it remains to multiply this rate by the probability that the offspring survives, S_1, and the probability that the parent G_2 is not chosen to die, $\frac{N-1}{N}$. As the population size N is considered large, the latter probability is approximately equal to 1.

This gives:

$$A_{12}=\left[f\frac{r}{4}\Big(1-(1-p_{in})(1-r)\Big)+(1-f)\frac{r}{2}\right]\times S_1.$$

C.3. Equation for the expected value of the size of the mutant population

This appendix gives the details of the derivation of the coefficients of the matrix C defined by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_{Z_0}[Z_t]=\mathbb{E}_{Z_0}[Z_t]C,$$

following Bacaër, 2018. Note that this is the same matrix defined in Athreya and Ney, 1972, Eq. 9, part. V.7.2., or in Pénisson, 2010, Eq. 1.1.16, but here we use the methodology described by Bacaër, 2018 to derive its coefficients.

In the following, type j refers to the genotype G_j. We will use the standard notation $s^z := s_1^{z_1}s_2^{z_2}\dots s_d^{z_d}$ for s and z two vectors of the same dimension d.

C.3.1. Notation. For all $t\geq 0$, let us denote the expected value of the process at time t by $E(t)$:

$$E(t)=\begin{pmatrix}E_1(t)\\E_2(t)\\E_3(t)\end{pmatrix}=\begin{pmatrix}\mathbb{E}[Z_{t,1}]\\\mathbb{E}[Z_{t,2}]\\\mathbb{E}[Z_{t,3}]\end{pmatrix}.$$

For $z\in\mathbb{N}^3$ and $t\geq 0$, we let $p(t,z)=\mathbb{P}(Z_t=z)$ be the probability that the system is found in state z at time t. Let $f(t,.)$ be the generating function of the variable Z_t: for all $s\in[0,1]^3$,

$$f(t,s):=\sum_{z\in\mathbb{N}^3}p(t,z)s^z=\mathbb{E}\Big[s_1^{Z_{t,1}}s_2^{Z_{t,2}}s_3^{Z_{t,3}}\Big].$$

Recalling that Y^j stands for the random vector of number of descendants of each type generated by the reproduction of a type j individual, we also define $\pi_j(z)=\mathbb{P}(Y^j=(z_1,z_2,z_3))$. As indicated in the main text, the rates at which an individual of type j reproduces and gives rise to a descendance vector e_i+e_j, e_i or 0 are respectively A_{ij}, T_{ij} and D_{jj}. We denote the total rate at which a reproduction event occurs for a parent of type j by $c_j:=\sum_i A_{ij}+\sum_i T_{ij}+D_{jj}$.

The reproduction law of type j individuals is then given by, for every $i\in\{1,2,3\}$,

$$\mathbb{P}(Y^j=e_i+e_j)=\frac{A_{ij}}{c_j},\quad \mathbb{P}(Y^j=e_i)=\frac{T_{ij}}{c_j},\quad \mathbb{P}(Y^j=0)=\frac{D_{jj}}{c_j}.$$

Finally, let h_j be the generating function of the reproduction law of type j individuals, for $j\in\{1,2,3\}$. That is, for $s\in[0,1]^3$,

$$h_j(s)=\sum_{z\in\mathbb{N}^3}\pi_j(z)s^z=\mathbb{E}\Big[s_1^{Y_1^j}s_2^{Y_2^j}s_3^{Y_3^j}\Big].$$

C.3.2. Ordinary differential edecompoquation (ODE) satisfied by $(E(t))_{t\geq 0}$. The reproduction law of each type has finite moments of all order, because the number of descendants produced can not exceed 2. That garantees that there is no explosion of the population in finite time. Hence, standard results on multi-dimensional random variables (see for example Athreya and Ney, 1972) give us that, for all types j and all $t \geq 0$,

$$E_j(t) = \frac{\partial f}{\partial s_j}(t, \mathbb{1}),$$

with $\mathbb{1} = (1, 1, 1)$, which gives

$$\frac{\mathrm{d}E_j(t)}{\mathrm{d}t} = \frac{\partial^2 f}{\partial s_j \partial t}(t, \mathbb{1}) = \frac{\partial}{\partial s_j}\left(\sum_{z\in\mathbb{N}^3} \frac{\partial p}{\partial t}(t, z)s^z\right)\Big|_{s=\mathbb{1}}.$$

The variation of p over time $\frac{\partial p(t,z)}{\partial t}$ can be decomposed into two terms. For $z \in \mathbb{N}^3$,

$$\frac{\partial p}{\partial t}(t, z) = -\sum_{j=1}^{3} z_j c_j p(t, z) + \sum_{j=1}^{3} \sum_{\substack{u,v\in\mathbb{N}^3 \\ u+v=z}} (u_j + 1)c_j p(t, u + e_j)\pi_j(v).$$

The first term is the rate at which the population departs from state z, and is given by the sum over all types j of the rate at which individuals of type j reproduce. The second term is the rate at which the population arrives in state z from another state, and can be decomposed according to the individual type whose reproduction changes the population state. Note that the descendance vector generated during the reproduction event (v) counts the parent when it does not die, implying that the population is formally decreased by one individual of type j and increased by a vector v during the reproduction event. In other words, if the population starts from a state $u + e_j$ and an individual of type j reproduces by creating a vector v of descendants, the final state of the population is $u + e_j - e_j + v = u + v$.

Back to the derivative of f with respect to t, we use the fact that the rates c_j are independent of the current state of the population to re-arrange the sums and obtain:

$$\begin{aligned}
\frac{\partial f}{\partial t}(t, s) &= \sum_{z\in\mathbb{N}^3} \frac{\partial p}{\partial t}(t, z)s^z \\
&= \sum_{z\in\mathbb{N}^3} \left(-\sum_{j=1}^{3} z_j c_j p(t, z)s^z + \sum_{j=1}^{3} \sum_{\substack{u,v\in\mathbb{N}^3 \\ u+v=z}} (u_j + 1)c_j p(t, u + e_j)\pi_j(v)s^z \right) \\
&= \sum_j c_j \left(-\sum_{z\in\mathbb{N}^3} z_j p(t, z)s^z + \sum_{z\in\mathbb{N}^3} \sum_{\substack{u,v\in\mathbb{N}^3 \\ u+v=z}} (u_j + 1)p(t, u + e_j)\pi_j(v)s^z \right) \\
&= \sum_j c_j \left(-s_j \sum_{z\in\mathbb{N}^3} z_j p(t, z)s^{z-e_j} + \sum_{v\in\mathbb{N}^3} \sum_{u\in\mathbb{N}^3} (u_j + 1)p(t, u + e_j)\pi_j(v)s^{u+v} \right) \\
&= \sum_j c_j \left(-s_j \sum_{z\in\mathbb{N}^3} z_j p(t, z)s^{z-e_j} + \sum_{v\in\mathbb{N}^3} \pi_j(v)s^v \sum_{u\in\mathbb{N}^3} (u_j + 1)p(t, u + e_j)s^u \right) \\
&= \sum_j c_j \left(-s_j \frac{\partial f}{\partial s_j}(t, s) + h_j(s)\frac{\partial f}{\partial s_j}(t, s) \right) \\
&= \sum_j c_j (h_j(s) - s_j)\frac{\partial f}{\partial s_j}(t, s).
\end{aligned}$$

Writing $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise, we then obtain for the expected value:

$$\begin{aligned}\frac{\mathrm{d}E_i}{\mathrm{d}t} &= \frac{\partial}{\partial s_i}\frac{\partial f}{\partial t}(t,s)\big|_{s=\mathbb{1}} \\ &= \sum_j c_j \left(\frac{\partial h_j}{\partial s_i}(\mathbb{1}) - \delta_{i,j}\right)\frac{\partial f}{\partial s_j}(t,\mathbb{1}) + \sum_j c_j(h_j(\mathbb{1})-1)\frac{\partial^2 f}{\partial s_i \partial s_j}(t,\mathbb{1}) \\ &= \sum_j c_j \left(\frac{\partial h_j}{\partial s_i}(\mathbb{1}) - \delta_{i,j}\right) E_j(t) + \sum_j c_j(h_j(\mathbb{1})-1)\frac{\partial^2 f}{\partial s_i \partial s_j}(t,\mathbb{1}) \\ &= \sum_j c_j \left(\frac{\partial h_j}{\partial s_i}(\mathbb{1}) - \delta_{i,j}\right) E_j(t),\end{aligned}$$

where the last equality arises from the fact that, because h_j is a generating function,

$$h_j(\mathbb{1}) - 1 = \sum_{z\in\mathbb{N}^3} \pi_j(z) - 1 = 0.$$

The matrix C we are looking for is thus defined by $C_{ij} = c_j\left(\frac{\partial h_j}{\partial s_i}(\mathbb{1}) - \delta_{ij}\right)$ for $1 \le i, j \le 3$.

Furthermore, we have, for all j,

$$h_j(s) = \frac{1}{c_j}\left(\sum_i A_{ij}s_i s_j + \sum_i T_{ij}s_i + D_{jj}\right).$$

Combining the above, we arrive at

$$C_{ij} = \begin{cases} A_{ij} + T_{ij} & \text{if } i \neq j, \\ A_{jj} - \sum\limits_{k\neq j} T_{kj} - D_{jj} & \text{if } i = j. \end{cases}$$

In conclusion, the matrix C is given by

$$C = \begin{pmatrix} (fa(r) + (1-f)d(r))\,S_1 - S_4 & (fc(r) + (1-f)\frac{r}{2})\,S_1 & (1-f)S_1 \\ (fc(r) + (1-f)\frac{r}{2})\,S_2 & (fa(r) + (1-f)d(r))\,S_2 - S_4 & (1-f)S_2 \\ fb(r)S_3 & fb(r)S_3 & fS_3 - S_4 \end{pmatrix}, \tag{6}$$

with

$$a(r) = 1 - r + \frac{r}{4}\Big(1 - (1-p_{in})(1-r)\Big), \quad b(r) = \frac{r}{4}\Big(1 + (1-p_{in})(1-r)\Big),$$
$$c(r) = \frac{r}{4}\Big(1 - (1-p_{in})(1-r)\Big), \quad \text{and} \quad d(r) = 1 - \frac{1}{2}r.$$

C.4. Reducibility of the matrix C and probability of extinction of the branching process

We will use the standard notation $s^z := s_1^{z_1}s_2^{z_2}\dots s_d^{z_d}$ for s and z two vectors of the same dimension d.

Assessing the type of branching process at hand (super-, sub-, or critical) relies on the study of the eigenvalues of the matrix C. We use results of Sewastjanow, 1975 detailed in Pénisson, 2010 to obtain conditions on the almost-sure extinction of the process. When the matrix C is irreducible, the Perron-Froebenius theory of positive matrices states that it has a unique dominant eigenvalue. The branching process is then super-, sub-, or critical when this dominant eigenvalue is respectively positive, negative, or zero (Athreya and Ney, 1972, V.7.2.). In particular, the probability of extinction is equal to 1 when $\rho \le 0$.

In our case, the matrix C can be reducible (for example, when $f = 0$). In order to obtain a result on the probability of extinction in the subcritical case, we use the theory of sub-processes and

of final classes. We recall below useful definitions and the principal result used (Sewastjanow, 1975).

Let $(Z_t)_{t>0}$ be a multitype branching process, with types in a finite set K. The equivalence relation of *communication* is defined by: for all states $k_i, k_j \in K$, we say that k_i and k_j *communicate*, if and only if there exist $s, t > 0$ such that

$$\mathbb{P}_{e_{k_i}}(Z_{s,k_j} > 0) > 0 \quad \text{and} \quad \mathbb{P}_{e_{k_j}}(Z_{t,k_i} > 0) > 0.$$

This means that there exists a time at which the probability that the population described by a branching process initiated with a single individual of type k_i contains an individual of type k_j is positive, and a time at which the probability that the population described by a branching process initiated with a single individual of type k_j contains an individual of type k_i is positive as well. If a subset $\tilde{K} = \{k_1, \dots, k_p\}$ is a class for the communication equivalence relation (meaning that each state of $\tilde{K}$ communicates with all the others but communicates with none of the states in $\tilde{K}^c$), the $\tilde{K}$-subprocess is the process defined for all $t > 0$ by

$$\tilde{Z}_t := \left(Z_{t,k_1}, \dots, Z_{t,k_p}\right),$$

which is the vector Z_t from which only the coordinates of the types in the class $\tilde{K}$ are kept. $(\tilde{Z}_t)_{t\geq 0}$ is still a branching process, and is by definition irreducible.

Let $F_{t,k_i} : s \in [0,1]^d \longmapsto \mathbb{E}_{e_{k_i}}\left[s^{Z_t}\right]$ be the generating function of the process $(Z_t)_{t\geq 0}$ at time t, starting with one individual of type k_i. $\tilde{K} = \{k_1, \dots, k_p\}$ is then said to be a *final class* if it is non-empty, and satisfies the property that there exists $t > 0$ such that for all $k_i \in \tilde{K}$ and $s \in [0,1]^d$, $F_{t,k_i}(s)$ is of the form

$$F_{t,k_i}(s) = \alpha_{k_i,1}(t,s)s_{k_1} + \dots + \alpha_{k_i,p}(t,s)s_{k_p},$$

where the coefficients $\alpha_{k_i,j}$ can be expressed using the coordinates s_k of s such that $k \notin \tilde{K}$. In other words, $F_{t,k_i}(s)$ is linear in s_k for all $k \in \tilde{K}$. The interpretation of this property is that whenever the population starts from a single individual of type $k_i \in \tilde{K}$, at any time $t \geq 0$ there is one, and only one, individual of a type $k_j \in \tilde{K}$ (and potentially other individuals with types in $\tilde{K}^c$). The following result gives a condition for the almost sure extinction of the process $(Z_t)_{t\geq 0}$ in the general case where the matrix C is not necessarily irreducible. Recall that the Perron's root ρ of a process, when it exists, is a real eigenvalue of the matrix associated with the process such that all real parts of other eigenvalues are smaller than ρ (see Pénisson, 2010 Th. 1.1.7 and the following ones for a more detailed definition).

Theorem C.1 (Prop. 1.1.22 in Pénisson, 2010). *Let $(Z_t)_{t>0}$ be a continuous time Galton-Watson process, and let $\rho = \max_{\tilde{K}} \rho_{\tilde{K}}$ be the maximal value of the Perron's roots of all the possible $\tilde{K}$-subprocesses. Then the process $(Z_t)_{t>0}$ almost surely dies out if and only if there are no final classes and $\rho \leqslant 0$.*

Let us verify that our branching process does not contain a final class. For that, we show that the generating function of the process starting from any state has a non-zero coefficient of degree zero, and thus cannot be linear.

For any $t > 0$, $r \in [0,1]^3$ and any $j \in \{1,2,3\}$, we can decompose the generating function into

$$F_{t,j}(r) = \mathbb{E}_{e_j}(r^{Z_t}) = \mathbb{P}_{e_j}(Z_t = \mathbf{0}) + \sum_{z \in \mathbb{N}^3 \setminus \{\mathbf{0}\}} \mathbb{P}_{e_j}(Z_t = z) r^z.$$

Let us prove that $\mathbb{P}_{e_j}(Z_t = 0) > 0$ for every $j \in \{1,2,3\}$. This will prove that the generating function cannot be linear for any initial type, and thus that the process does not contain any final classes.

Let $j \in \{1,2,3\}$, τ_1 be the time of the first reproduction event, and Y_1^j be the descendance vector created at that time. We have

$$\begin{aligned}\mathbb{P}_{e_j}(Z_t = 0) &\geq \mathbb{P}_{e_j}\Big(\{\tau_1 \leq t\} \cap \{Y_1^j = \mathbf{0}\}\Big)\\ &= \mathbb{P}_{e_j}\Big(Y_1^j = \mathbf{0} | \tau_1 \leq t\Big)\mathbb{P}_{e_j}\Big(\tau_1 \leq t\Big)\\ &= \mathbb{P}_{e_j}(\tau_1 \leq t) \times \mathbb{P}(Y^j = \mathbf{0})\\ &= (1 - \mathrm{e}^{-c_j t})\frac{D_{jj}}{c_j},\end{aligned}$$

where c_j is the total rate of reproduction of an individual of type j, and D_{jj} is the rate at which an individual of type j reproduces and gives rise to a null vector of descendants. Hence, $\mathbb{P}_{e_j}(Z_t = 0) > 0$ when $D_{jj} > 0$.

For both selection scenarii, $D_{11} = D_{22} = D_{33} = S_4$. In the partial dominance selection scenario, $S_4 = 1$, and in the overdominant selection scenario, $S_4 = 1 - s_4$. Having $D_{jj} = 0$ for any j is impossible in the first scenario and requires $s_4 = 1$ in the second scenario, which means that the wild allele is lethal, which is not a reasonable assumption. We thus take $s_4 < 1$. As a consequence, the generating function cannot be linear, and the process does not contain any final class.

The result of Proposition C.1 then applies here, and the sign of the dominant eigenvalue of matrix C gives a condition on the almost-sure extinction of the process.

Appendix D. Supplementary figures

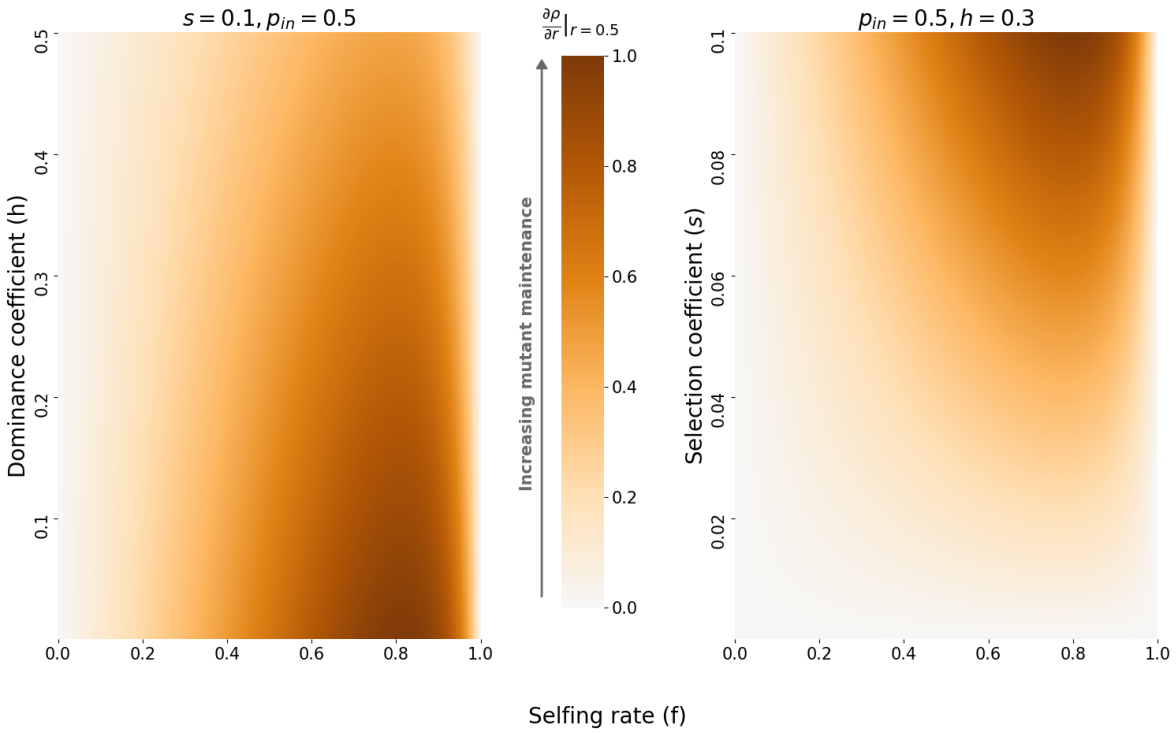


Figure S2 – Relative variation of the derivative of the eigenvalue in the partial dominance case, for varying selfing rate f (x-axis), dominance coefficient h (y-axis, left) and selection coefficient s (y-axis, right). For each panel, the values of $\frac{\partial \rho}{\partial r}|_{r=0.5}$ range from a minimal value, which is negative, to zero. We divided each value of the derivative by this minimum in order to plot values between 0 and 1 for every panel. This enables us to compare the impact of different parameters (h, s and f) on the sheltering effect of the mating-type locus. The darker the color, the more the mating-type locus shelters the mutation, thus promoting its maintenance.

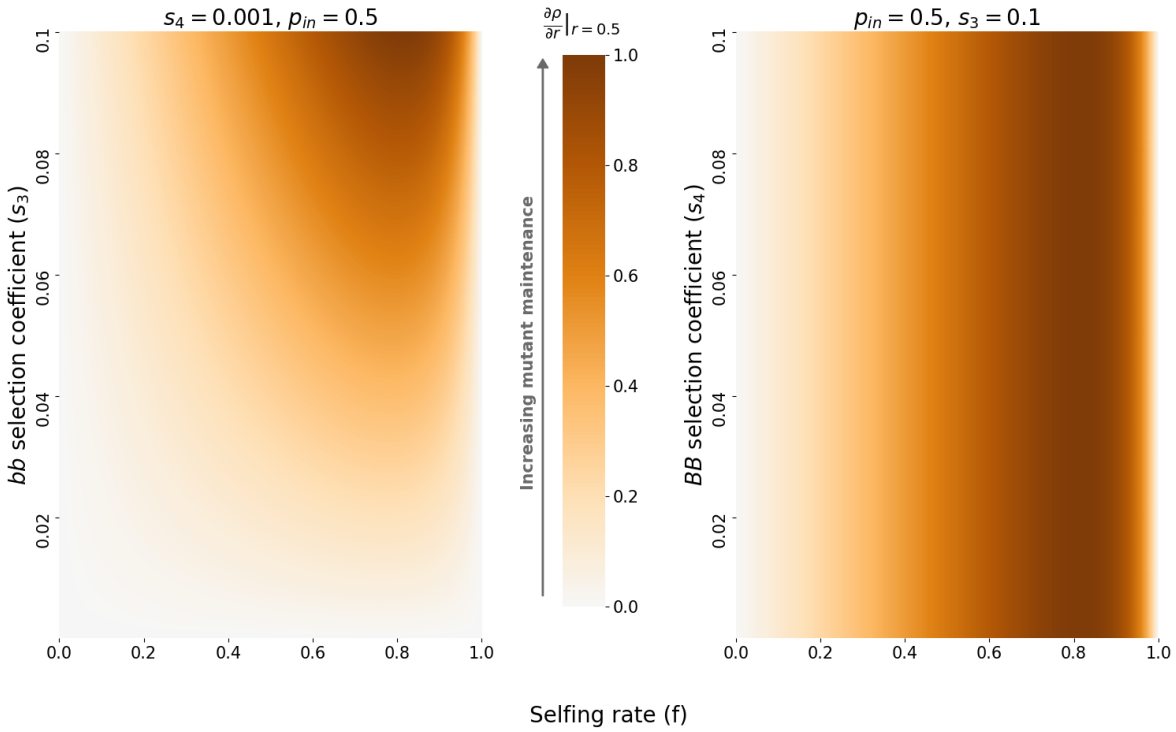


Figure S3 – Relative variation of the derivative of the eigenvalue in the overdominance case, for varying selfing rate f (x-axis), and selection coefficients s_3 (y-axis, left) and s_4 (y-axis, right), with $s_3 > s_4$. For each panel, the values of $\frac{\partial \rho}{\partial r}|_{r=0.5}$ range from a minimal value, which is negative, to zero. We divided each value of the derivative by this minimum in order to plot values between 0 and 1 for every panel. This enables us to compare the impact of different parameters (s_3, s_4 and f) on the sheltering effect of the mating-type locus. The darker the color, the more the mating-type locus shelters the mutation, thus promoting its maintenance.

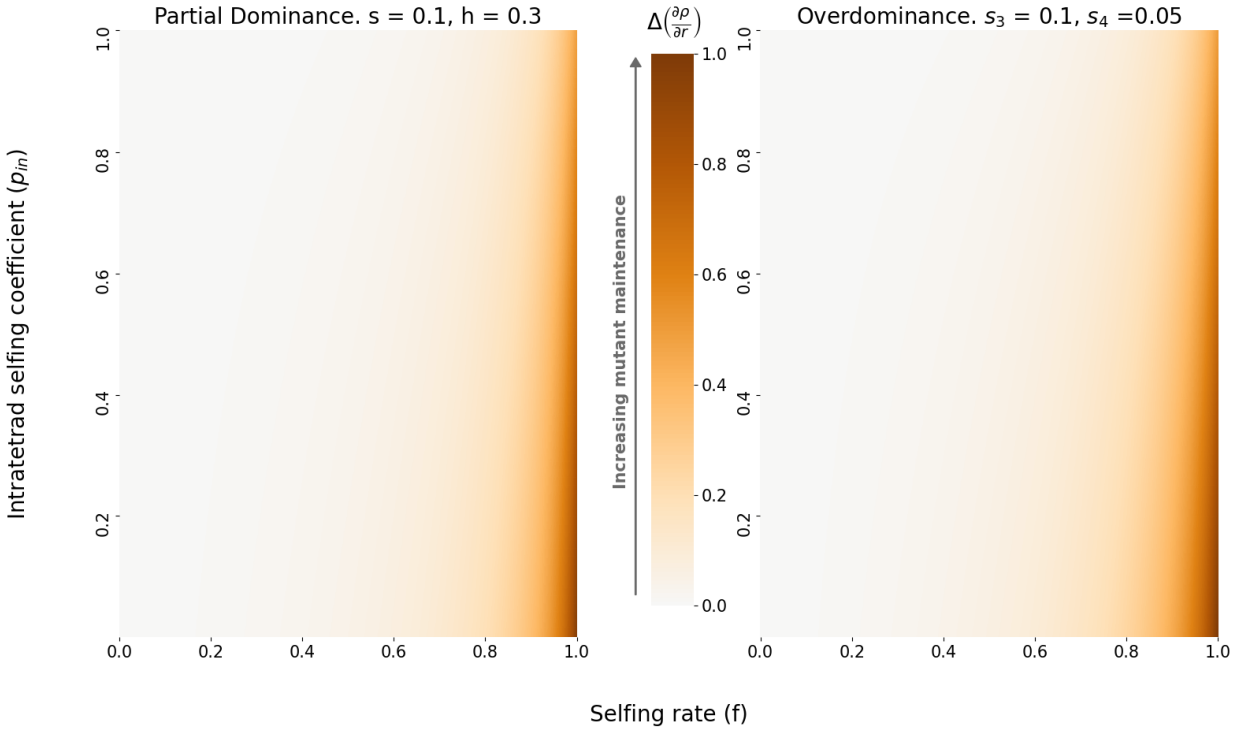


Figure S4 – Difference between the dominant eigenvalue derivative at $r = 0.5$ and at $r = 0$, $\Delta\left(\frac{\partial\rho}{\partial r}\right) = \frac{\partial\rho}{\partial r}\,|_{r=0.5} - \frac{\partial\rho}{\partial r}\,|_{r=0}$. The left panel shows the partial dominance case, the right panel shows the overdominance case, for varying selfing rate f (x-axis), and intratetrad selfing rate (y-axis). The difference is always positive, with both derivative being negative (see App. E.3 and App. F.3). This means that the absolute value of the derivative at $r = 0$ is always greater than the absolute value of the derivative at $r = 0.5$. The darker the color, the larger the difference between the two derivatives.

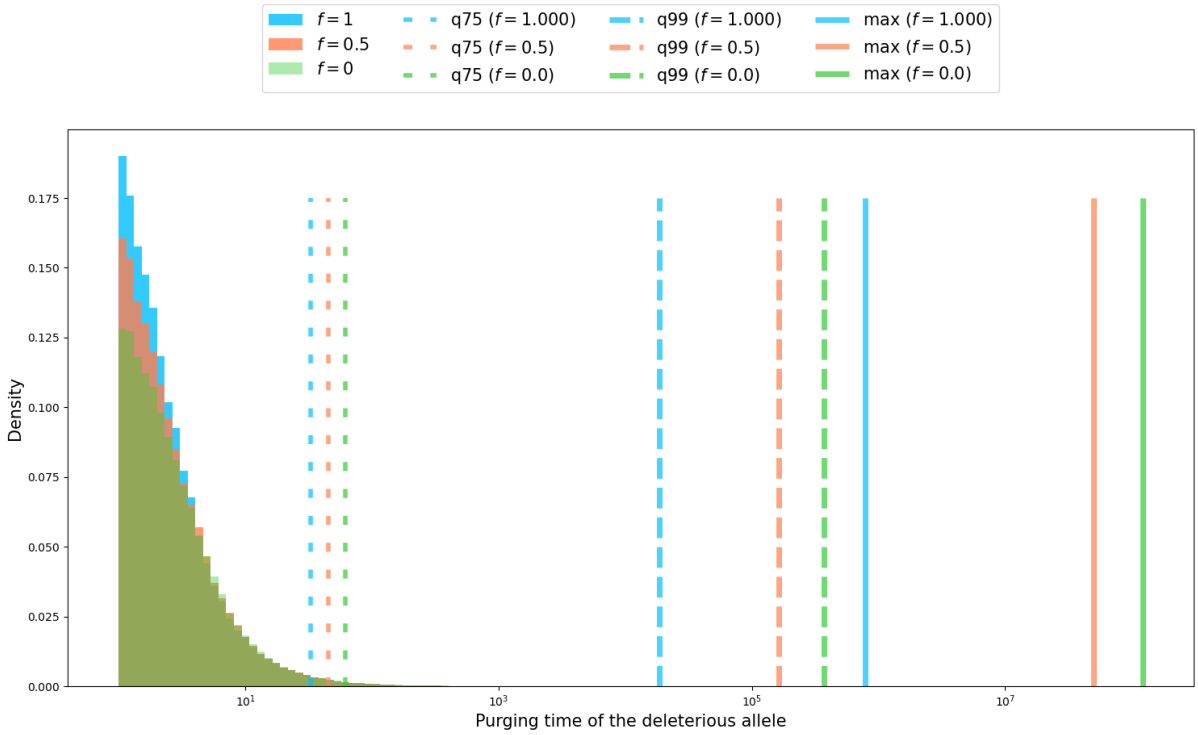


Figure S5 – Empirical distribution of the deleterious allele purging time for the partial dominance scenario. A total of 100,000 simulations were run, with $s = 0.1$, $h = 0.1$, $r = 0.1$, $p_{in} = 0.5$, starting from one heterozygous individual ($X_0 = (1, 0, 0)$), and for three values of the selfing rate ($f = 0$ in green, $f = 0.5$ in red and $f = 1$ in blue). The respective values for ρ are $\rho = -0.0100$, $\rho = -0.0157$ and $\rho = -0.0818$. The x-axis is log-scaled. The large-dotted lines represent the 75^{th} percentile (*q75*), the dashed lines indicate the 99^{th} percentile (*q99*), and solid lines the maximum value (*max*) of the purging time. Maximum values are several order of magnitudes higher than the 75th percentile of the empirical distribution of the purging time.

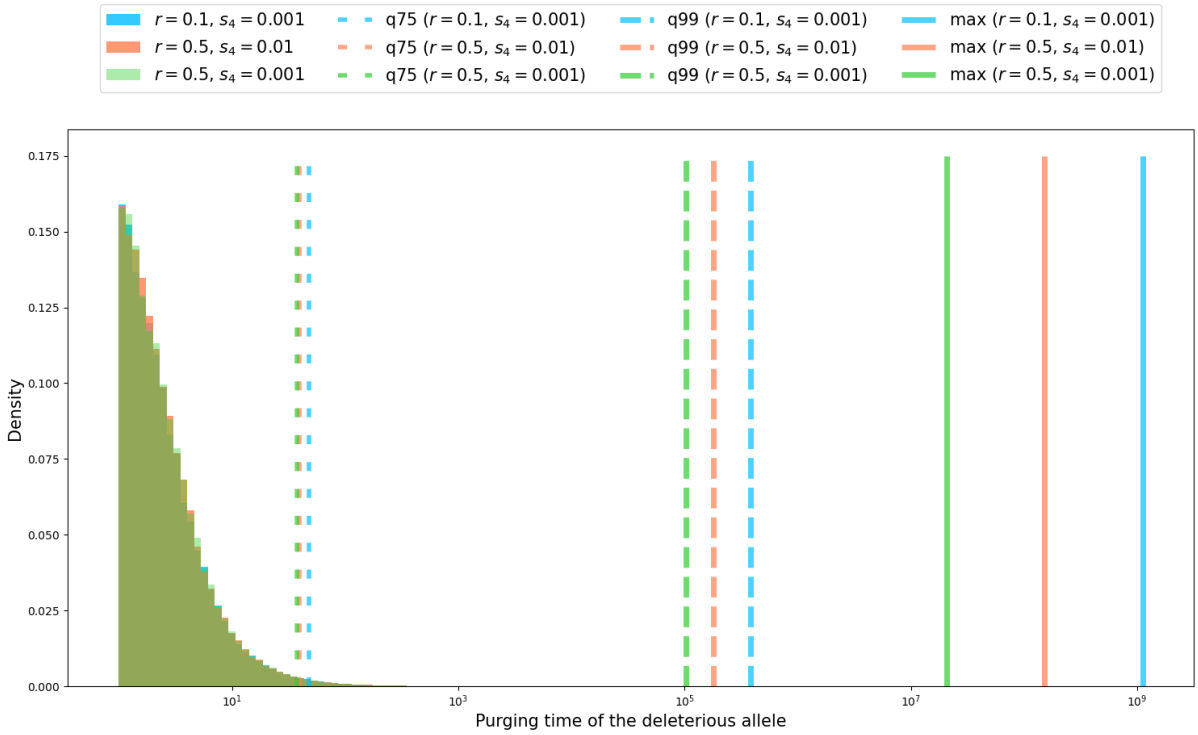


Figure S6 – Empirical distribution of the deleterious allele purging time for the over-dominance scenario. A total of 100,000 simulations were run, with $s_3 = 0.1$, $f = 0.5$, $p_{in} = 0.5$, starting from one heterozygous individual ($X_0 = (1, 0, 0)$), for several values of the recombination rate and of the selection coefficient s_4 ($r = 0.1, s_4 = 0.001$ in blue, $r = 0.5, s_4 = 0.01$ in red, and $r = 0.5, s_4 = 0.001$ in green). The respective values for ρ are $\rho = -0.0052$, $\rho = -0.0129$ and $\rho = -0.0219$. The parameters were chosen so that the process is sub-critical and thus the purging time is almost surely finite. The x-axis is log-scaled. The large-dotted lines represent the 75^{th} percentile (*q75*), the dashed lines indicate the 99^{th} percentile (*q99*), and solid lines the maximum value (*max*) of the purging time. Maximum values are several order of magnitudes higher than the 75th percentile of the empirical distribution of the purging time.

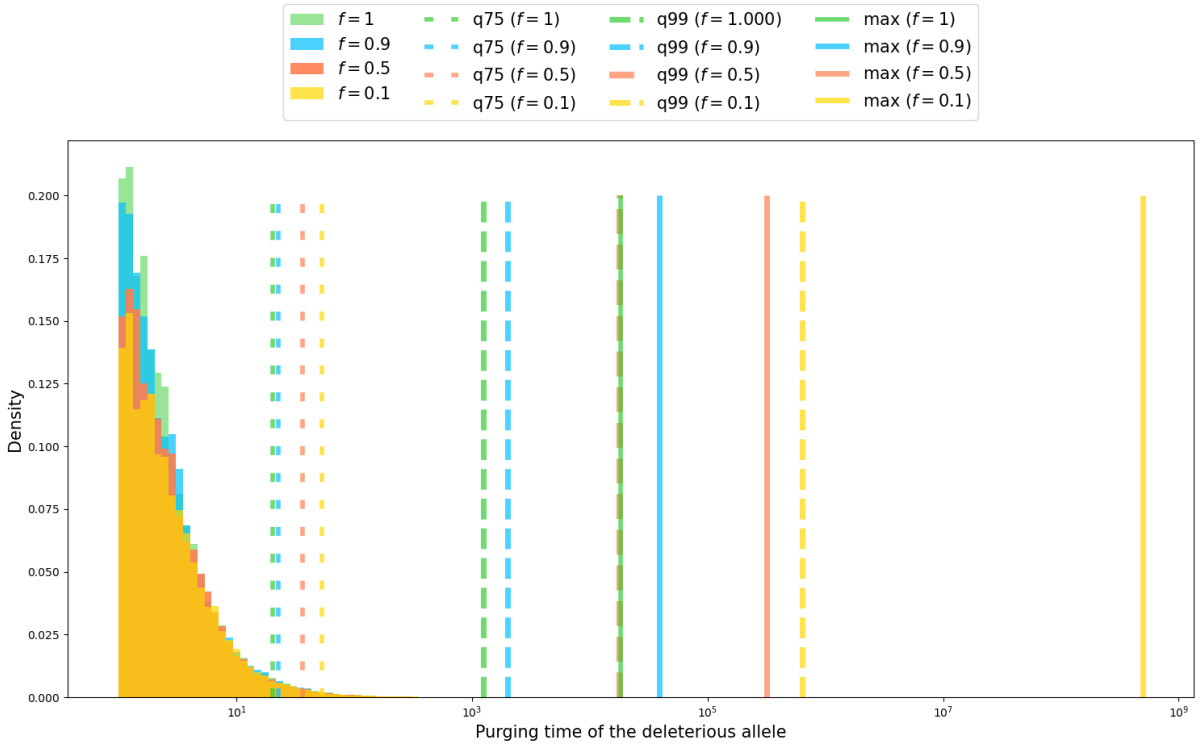


Figure S7 – Empirical distribution of the deleterious allele purging time for the overdominant scenario. A total of 100,000 simulations were run, with $s_3 = 0.5$, $p_{in} = 0.5$, $s_4 = 0.01$, $r = 0.4$, starting from one heterozygous individual ($X_0 = (1, 0, 0)$), for four values of the selfing rate ($f = 0.1$ in yellow, $f = 0.5$ in red, $f = 0.9$ in blue, $f = 1$ in green). The respective values for ρ are $\rho = -0.0035$, $\rho = -0.0713$, $\rho = -0.1905$ and $\rho = -0.25$. Parameters were chosen so that the process is sub-critical and thus the purging time is almost surely finite. The x-axis is log-scaled. The large-dotted lines represent the 75^{th} percentile (*q75*), dashed lines indicate the 99^{th} percentile (*q99*), and solid lines the maximum value (*max*) of the purging time. Maximum values are several order of magnitudes higher than the 75th percentile of the empirical distribution of the purging time. Note that the selection coefficient for *bb* homozygotes is high ($s_3 = 0.5$).

Appendix E. The dominant eigenvalue, its sign and its derivative: partial dominance scenario

E.1. Determination of the dominant eigenvalue

The eigenvalues computed with Mathematica (Wolfram Research, 2015) are, for the partial dominance case,

$$\lambda_0 = -r - hs(1-r), \quad \lambda_+ = \frac{1}{4}\Big(\beta + \sqrt{\Delta}\Big), \quad \lambda_- = \frac{1}{4}\Big(\beta - \sqrt{\Delta}\Big),$$

where

$$\beta = f\Big(-r(1-hs)\alpha + 2(1-s)\Big) - 2(1+hs), \qquad \alpha = 2 - r - p_{in}(1-r), \tag{7}$$

and

$$\Delta = (\beta + 4hs)^2 - 8fsr\alpha(1-h)(1-hs). \tag{8}$$

It is straightforward to see that $\lambda_+ > \lambda_-$.

Let us prove that we also have $\lambda_+ > \lambda_0$. We used Geogebra to assist us in the calculations.

$$\lambda_+ > \lambda_0 \quad \text{if and only if (iff)} \quad \frac{1}{4}\Big(\beta + \sqrt{\Delta}\Big) \geq \lambda_0 \quad \text{iff} \quad \sqrt{\Delta} \geq 4\lambda_0 - \beta.$$

If $4\lambda_0 - \beta \leq 0$, the last inequality if straightforward, as $\sqrt{\Delta} \geq 0$.
Let us study the sign of $4\lambda_0 - \beta$. We define $P(r) := 4\lambda_0 - \beta = a_2 r^2 + a_1 r + a_0$, with

$$a_2 = -f(1-hs)(1-p_{in}) < 0, \quad a_1 = f(1-hs)(2-p_{in}) - 4(1-hs) < 0,$$
$$\text{and } a_0 = -2f(1-s) + 2(1-hs) > 0.$$

P is a second-order polynomial, with negative quadratic coefficient and positive coefficient of order zero (because $1 - hs > 1 - s > f(1-s)$). Thus, P admits two roots, one negative and one positive. We denote the positive root by r_P. For $r \in [0, r_P]$, we have $P(r) \geq 0$, and for $r > r_P$, we have $P(r) < 0$. Consequently, we readily obtain that when $r > r_P$, $\lambda_+ > \lambda_0$.
Let us now consider the case $r \in [0, r_P]$. For such an r, using that $4\lambda_0 - \beta \geq 0$, we can write that

$$\sqrt{\Delta} \geq 4\lambda_0 - \beta \quad \text{iff} \quad \Delta \geq (4\lambda_0 - \beta)^2.$$

Let us write $Q(r) := (4\lambda_0 - \beta)^2 - \Delta = b_2 r^2 + b_1 r + b_0$, with

$$b_2 = f(1-hs)(1-p_{in}) > 0, \quad b_1 = 2(1-hs) - f(1-hs)(2-p_{in}) - fs(1-h)(1-p_{in}),$$
$$\text{and } b_0 = -2(1-hs)(1-f) - p_{in}fs(1-h) < 0.$$

Q is a second-order polynomial, with positive quadratic coefficient, and negative coefficient of order 0. Hence, Q admits two roots, one negative, and one positive. We denote the positive root by r_Q. In order to prove that $\lambda_+ > \lambda_0$, we have to prove that $Q(r) \leq 0$ when $P(r) > 0$, *i.e.* when $r \in [0, r_P]$. As $Q(0) < 0$ and Q has only one positive root, proving that $Q(r_P) < 0$ will imply that $Q(r) \leq 0$ for $r \in [0, r_P]$. Let us prove that $Q(r_P) < 0$. Noting that the quadratic coefficients of P and Q are the opposites of one another, we use the equation $P(r_P) = 0$ to obtain

$$Q(r_P) = r_P\Big(-2(1-hs) - fs(1-h)(1-p_{in})\Big) + fs(1-h)(2-p_{in}).$$

Seeing $Q(r_P)$ as an affine function of r_P, we obtain that the function $r_P \longmapsto Q(r_P)$ admits a unique root, which is positive, and that we will denote by r_P^0 :

$$r_P^0 = \frac{fs(1-h)(2-p_{in})}{2(1-hs)+fs(1-h)(1-p_{in})}.$$

We wish to prove that $r_P \geq r_P^0$, as it implies that $Q(r_P) \leq 0$. Having $r_P \geq r_P^0$ is equivalent to having $P(r_P^0) \geq 0$, as r_P is the unique positive root of P and $P(0) \geq 0$. Consequently, it only remains to prove that $P(r_P^0) \geq 0$, which is equivalent to

$$\Big(2(1-hs)+fs(1-h)(1-p_{in})\Big)^2 P(r_P^0) \geq 0.$$

To obtain this result, an efficient way is to consider the left-hand term as a polynomial in p_{in}. Let us write $K(p_{in}) = (2(1-hs)+fs(1-h)(1-p_{in}))^2 P(r_0) = c_2 p_{in}^2 + c_1 p_{in} + c_0$, with

$$c_2 = f^2 s(1-s)(1-fs)(1-h) > 0, \quad c_1 = 2(1-s)f^3 s^2 (1-h)^2 > 0$$
$$\text{and } c_0 = (1-h)^2 f^2 s^2 (1-hs+f(s-1)) + 4(1-f)(1-hs)^2 > 0.$$

K is thus a second-degree polynomial in p_{in}, with a positive quadratic coefficient, and a minimum reached for a negative value (minimum reached at $-c_1/(2c_2) < 0$). K is thus monotonic for positive abscissa, and the coefficient of order zero is positive. Consequently, for all $p_{in} \geq 0$, we have $K(p_{in}) \geq 0$. We have then $P(r_0) \geq 0$, which concludes the proof that $\lambda_+ \geq \lambda_0$.

Based on the result we just obtained, from now on we write $\rho = \lambda_+$.

E.2. Sign of the dominant eigenvalue

We prove that $\rho < 0$, except when $s = 0$, or when $h = 0$ and $r = 0$, in which cases $\rho = 0$. Recall the notation α, β from (7) and Δ from (8).

First, considering that $r \in [0,1]$ and $p_{in} \in [0,1]$, we have $0 < \alpha < 2$, which leads to $\beta < 0$.

When $s = 0$ or $(r,h) = (0,0)$, $\Delta = \beta^2$, which gives, as $\beta < 0$, $\sqrt{\Delta} = \sqrt{\beta^2} = |\beta| = -\beta$. We then have $\rho = \frac{1}{4}(\beta - \beta) = 0$.

Let us now consider the case where $s \neq 0$ and $(r,h) \neq (0,0)$. We have

$$\beta + \sqrt{\Delta} > 0 \quad \text{iff} \quad 0 > \beta > -\sqrt{\Delta} \quad \text{iff} \quad 0 < \beta^2 < \Delta \quad \text{iff} \quad \beta^2 - \Delta < 0.$$

Moreover, $\beta^2 - \Delta = -16f(1-s)hs + 8fr\alpha(1-hs)s + 16hs$. The sign of ρ is thus the sign of $fh(2-r\alpha)s + 2h(1-f) + r\alpha f$, which is an affine function of s. The slope and intercept of this function are both non-positive when $(r,h) \neq 0$ or $s \neq 0$, which gives $\rho < 0$ in those cases.

E.3. Derivative of the dominant eigenvalue

The derivative of ρ is

$$\frac{\partial \rho}{\partial r} = \frac{1}{4}\left(\beta'(r) + \frac{1}{2}\frac{\Delta'(r)}{\sqrt{\Delta}}\right),$$

Evaluating this derivative at $r = 0.5$, and using that $\beta'(0.5) = -f(1-hs)$, we obtain

$$\frac{\partial \rho}{\partial r}\Big|_{r=0.5} = -\frac{1}{4}f(1-hs)\left(1 + \frac{\beta(0.5)+4s}{\sqrt{\Delta}}\right).$$

Simple calculations lead to $\frac{\partial \rho}{\partial r}|_{r=0.5} = 0$ when $f = 0$, or $s = 0$, or $s = 1$, or $h = 1$, or $f = 1$ and $p_{in} \leq 3 - \frac{8s(1-h)}{1-hs}$. In the latter case, whether the inequality is verified or not determines the sign of Δ and therefore the value of $\sqrt{\Delta}$, which is either equal to $\beta(0.5)+4s$ or $-(\beta(0.5)+4s)$. The derivative is then either equal to zero or strictly negative.

For the rest of this paragraph, we study the sign of the derivative when none of the above cases is met.

Let us write $\gamma = \beta(0.5) + 4s$. If $\gamma \geq 0$, we readily obtain $\frac{\partial \rho}{\partial r}|_{r=0.5} < 0$. Let us then assume that $\gamma < 0$. In this case, we have

$$\frac{\partial \rho}{\partial r}\Big|_{r=0.5} < 0 \quad \text{iff} \quad 1 + \frac{\gamma}{\sqrt{\Delta}} > 0 \quad \text{iff} \quad \sqrt{\Delta} > -\gamma.$$

As $-\gamma > 0$, this comes down to

$$\frac{\partial \rho}{\partial r}\Big|_{r=0.5} < 0 \text{ iff } \Delta > \gamma^2 \quad \text{iff } (1-s)(f-1) < 0,$$

which is indeed satisfied.

In conclusion, we have shown that, in the general case,

$$\frac{\partial \rho}{\partial r}\Big|_{r=0.5} < 0.$$

We also compute the derivative at $r = 0$. We have $\beta(0) = 2f(1-s) - 2(1+hs)$, $\beta'(0) = -f(1-hs)(2-p_{in})$, $\Delta(0) = (2f(1-s) - 2(1-hs))^2$, and $\Delta'(0) = 2\beta'(0)(\beta(0) + 4hs) - 8fs(1-h)(1-hs)(2-p_{in})$.

After simplification, this gives

$$\frac{\partial \rho}{\partial r}\Big|_{r=0} = -\frac{f(1-hs)(2-p_{in})}{4}\left(1 + \operatorname{sgn}(f(1-s) - (1-hs)) + \frac{2s(1-h)}{|f(1-s) - (1-hs)|}\right),$$

with $\operatorname{sgn}(f(1-s) - (1-hs))$ is equal to 1 (respectively to -1) when $f(1-s) - (1-hs)$ is positive (resp. negative).

We obtain immediately that

$$\frac{\partial \rho}{\partial r}\Big|_{r=0} \leq 0.$$

Appendix F. The dominant eigenvalue, its sign and its derivative: overdominance scenario

F.1. Determination of the dominant eigenvalue

The eigenvalues computed with Mathematica (Wolfram Research, 2015) for the overdominant case are

$$\lambda_0 = s_4 - r, \quad \lambda_+ = \frac{1}{4}\left(\beta + \sqrt{\Delta}\right), \quad \lambda_- = \frac{1}{4}\left(\beta - \sqrt{\Delta}\right)$$

with

$$\beta = f\left(r\left[p_{in}(1-r) + r - 2\right] - 2s_3 + 2\right) + 4s_4 - 2, \tag{9}$$

and

$$\Delta = (\beta - 4s_4)^2 + 8frs_3(p_{in}(1-r) + r - 2) \tag{10}$$

Here again, we obviously have $\lambda_+ > \lambda_-$.

We follow the same method as in the partial dominance case to prove that $\lambda_+ > \lambda_0$.

We have

$$\lambda_+ > \lambda_0 \quad \text{if and only if (iff)} \quad \frac{1}{4}\left(\beta + \sqrt{\Delta}\right) \geq \lambda_0 \quad \text{iff} \quad \sqrt{\Delta} \geq 4\lambda_0 - \beta.$$

If $4\lambda_0 - \beta \leq 0$, the last inequality if straightforward, as $\sqrt{\Delta} \geq 0$. Let us thus study the sign of $4\lambda_0 - \beta$. Let us define the function P by $P(r) := 4\lambda_0 - \beta = a_2 r^2 + a_1 r + a_0$, with

$$a_2 = -f(1 - p_{in}) < 0, \quad a_1 = f(2 - p_{in}) - 4 < 0, \quad \text{and } a_0 = 2(1 - f(1 - s_3)) > 0.$$

P is a second-order polynomial, with a negative quadratic coefficient and a positive coefficient of order 0. Hence P admits two roots, one which is negative and one which is positive. We denote the positive root by r_P. For $r \in [0, r_P]$, we have $P(r) \geq 0$, and for $r > r_P$, we have $P(r) < 0$. Consequently, we readily obtain that when $r > r_P$, the conclusion follows.

Let us now consider $r \in [0, r_P]$. For such an r, as $4\lambda_0 - \beta \geq 0$, again we have

$$\sqrt{\Delta} \geq 4\lambda_0 - \beta \quad \text{iff} \quad \Delta \geq (4\lambda_0 - \beta)^2.$$

Let us define the function Q by $Q(r) := (4\lambda_0 - \beta)^2 - \Delta = b_2 r^2 + b_1 r + b_0$, with

$$b_2 = f(1 - p_{in}) > 0, \quad b_1 = f(1 + s_3)(p_{in} - 1) - f + 2, \quad \text{and } b_0 = -fs_3 p_{in} - 2(1 - f) < 0.$$

Q is a second-order polynomial, with positive quadratic coefficient, and negative coefficient of order 0. Hence, Q admits two roots, one negative and one positive. We denote the positive root by r_Q. In order to prove that $\lambda_+ > \lambda_0$, we have to prove that $Q(r) \leq 0$ when $P(r) > 0$, *i.e.* when $r \in [0, r_P]$. As $Q(0) < 0$ and Q has only one positive root, proving that $Q(r_P) < 0$ will imply that $Q(r) \leq 0$ for $r \in [0, r_P]$. Let us prove that $Q(r_P) < 0$. Noting that the quadratic coefficients of P and Q are the opposites of one another, we use the equation $P(r_P) = 0$ to obtain

$$Q(r_P) = r_P\Big(-2 - fs_3(1 - p_{in})\Big) + fs_3(2 - p_{in}).$$

Seeing $Q(r_P)$ as an affine function of r_P, we obtain that the function $r_P \longmapsto Q(r_P)$ admits a unique root, which is positive, and that we will denote by r_P^0:

$$r_P^0 = \frac{fs_3(2 - p_{in})}{2 + fs_3(1 - p_{in})}.$$

We wish to prove that $r_P \geq r_P^0$, as it implies that $Q(r_P) \leq 0$. Having $r_P \geq r_P^0$ is equivalent to having $P(r_0) \geq 0$, as r_P is the unique positive root of P and $P(0) \geq 0$. There is thus left to prove that $P(r_P^0) \geq 0$, which is equivalent to

$$\Big(2 + fs_3(1 - p_{in})\Big)^2 P(r_P^0) \geq 0.$$

To obtain this result, an efficient way is to consider the left-hand term as a polynomial in p_{in}. Let us write $K(p_{in}) = (2 + fs_3(1 - p_{in}))^2 P(r_P^0) = c_2 p_{in}^2 + c_1 p_{in} + c_0$, with

$$c_2 = f^2 s_3(1 - s_3)(1 - fs_3) > 0, \quad c_1 = 2(1 - s_3)f^3 s_3^2 > 0$$
$$\text{and } c_0 = f^2 s_3^2((1 - f)(1 - s_3) + s_3) + 4(1 - f) > 0.$$

K is thus a second-degree polynomial in p_{in}, with a positive quadratic coefficient and positive coefficient of order 0, that reaches its minimum for a negative value (minimum reached at $-c_1/(2c_2) < 0$). Consequently, for all $p_{in} \geq 0$, we have $K(p_{in}) \geq 0$. We have then $P(r_P^0) \geq 0$, which concludes the proof that $\lambda_+ \geq \lambda_0$.

Based on the result we just obtained, from now on we write $\rho = \lambda_+$.

F.2. Sign of the dominant eigenvalue

In this selection scenario, ρ is not of constant sign.
The condition for $\rho \geq 0$ is

$$\sqrt{f^2\Big[r(p_{in}(1 - r) + r - 2) - 2s_3 + 2\Big]^2 + 4f(2(s_3 - 1) - r(2s_3 - 1)((p_{in} - 1)r - p_{in} + 2)) + 4}$$
$$+ f\Big[r(p_{in}(1 - r) + r - 2) - 2s_3 + 2\Big] + 4s_4 \geq 2.$$

We compute the dominant eigenvalue and study its sign for simple cases, and then use a numerical approach to complete the analysis (Figure 2). Under complete intra-tetrad selfing ($f = 1, p_{in} = 1$), we have $\rho = s_4 - r/2$ if $s_3 \geq r/2$ and $\rho = s_4 - s_3$ if $s_3 < r/2$. As $s_4 \leq s_3$, the condition to have $\rho \geq 0$ reduces to $r \leq 2s_4$. This is consistent with the results of Antonovics and Abrams, 2004, as the authors set $s_3 = 1$ and thus obtain $\rho = s_4 - r/2$. Under complete selfing ($f = 1$), if $r(2 - r - p_{in}(1 - r)) - 2s_3 \geq 0$, then $\rho = s_4 - s_3 \leq 0$. This shows that the value of the dominant eigenvalue, and thus the dynamics of the process, depends only on the selection strength when the recombination rate r exceeds a certain threshold. Moreover, this threshold depends only on the selection coefficient for homozygous deleterious (s_3), and on the probability of intra-tetrad mating (p_{in}). This threshold appears on the bottom panels of Figure 2. Under complete outcrossing ($f = 0$), we have $\rho = s_4 \geq 0$. When the mutation is completely linked to a mating-type allele ($r = 0$), we have $\rho = s_4 \geq 0$. When the mutation is neutral ($s_3 = 0$, implying $s_4 = 0$ as well), we have $\rho = 0$. Finally, when BB homozygotes are not disfavored ($s_4 = 0$), we have $\rho < 0$. Indeed, in this case, $\beta = fr[p_{in}(1 - r) + r - 2] - 2fs_3 + 2(f - 1) < 0$. We thus have

$$\rho \geq 0 \quad \text{iff} \quad \sqrt{\Delta} \geq -\beta \geq 0 \quad \text{iff} \quad \Delta \geq \beta^2 \quad \text{iff} \quad 8frs_3(p_{in}(1 - r) + r - 2) \geq 0.$$

But we trivially have $p_{in}(1 - r) + r - 2 \leq 0$, and so the condition is not met and $\rho < 0$.

F.3. Derivative of the dominant eigenvalue

The derivative of the largest eigenvalue ρ is

$$\frac{\partial \rho}{\partial r} = \frac{1}{4}\left(\beta'(r) + \frac{1}{2}\frac{\Delta'(r)}{\sqrt{\Delta}}\right).$$

Moreover, we have

$$\beta'(r) = 2f(1 - p_{in})r + f(p_{in} - 2)$$

and

$$\Delta'(r) = 2\beta'(r)(\beta(r) - 4s_4) + 8fs_3(2(1 - p_{in})r + p_{in} - 2).$$

Evaluating these quantities at $r = 0.5$, we obtain $\beta'(0.5) = -f$, and so

$$\frac{\partial \rho}{\partial r}\Big|_{r=0.5} = -\frac{1}{4} f \left(1 + \frac{\beta(0.5) - 4s_4 + 4s_3}{\sqrt{\Delta}}\right).$$

Simple calculations lead to $\frac{\partial \rho}{\partial r}|_{r=0.5} = 0$ when $f = 0$, or $s_3 = 0$, or $f = 1$ and $p_{in} \leq 3 - 8s_3$.

For the rest of this paragraph, we study the sign of the derivative when none of the above cases is met.

Let us write $\gamma = \beta(r = 0.5) - 4s_4 + 4s_3$. If $\gamma \geq 0$, we readily obtain that $\frac{\partial \rho}{\partial r}|_{r=0.5} < 0$. Let us then assume that $\gamma < 0$. In this case, we have

$$\frac{\partial \rho}{\partial r}\Big|_{r=0.5} < 0 \quad \text{iff} \quad 1 + \frac{\gamma}{\sqrt{\Delta}} > 0 \quad \text{iff} \quad \sqrt{\Delta} > -\gamma.$$

As $-\gamma > 0$,

$$\frac{\partial \rho}{\partial r}\Big|_{r=0.5} < 0 \text{ iff } \Delta > \gamma^2 \text{ iff } (\beta - 4s_4)^2 + 8frs_3(p_{in}(1 - r) + r - 2) > (\beta - 4s_4)^2 + 8s_3(\beta - 4s_4) + 16s_3^2.$$

After some simplifications, we obtain

$$\frac{\partial \rho}{\partial r}\Big|_{r=0.5} < 0 \quad \text{iff} \quad f < \frac{8}{3},$$

which is always satisfied as $f \in [0, 1]$.

In conclusion, we have shown that

$$\frac{\partial \rho}{\partial r}\Big|_{r=0.5} < 0.$$

We also compute the derivative at $r = 0$. We have $\beta(0) = 2f(1 - s_3) + 4s_4 - 2$, $\beta'(0) = -f(2 - p_{in})$, $\Delta(0) = (2f(1 - s_3) - 2)^2$, and $\Delta'(0) = 2\beta'(0)(\beta(0) - 4s_4) - 8fs_3(2 - p_{in})$.

After simplification, this gives

$$\frac{\partial \rho}{\partial r}\Big|_{r=0} = -\frac{f(2 - p_{in})}{4} \left(1 + \operatorname{sgn}(2f(1 - s_3) - 2) + \frac{4s_4}{|2f(1 - s_3) - 4s_4|}\right),$$

with $\operatorname{sgn}(2f(1 - s_3) - 2)$ is equal to 1 (respectively to -1) when $2f(1 - s_3) - 2$ is positive (resp. negative).

We obtain immediately that

$$\frac{\partial \rho}{\partial r}\Big|_{r=0} \leq 0.$$

Chapitre 3

Accumulation of recessive deleterious mutations near a mating-type locus in selfing recombining and non-recombining individuals

Ce troisième chapitre, exploratoire, est le fruit d'un travail réalisé en fin de thèse. Le but principal était de poursuivre l'étude de la dynamique de mutations délétères liées ou non à un locus en permanence hétérozygote en considérant la possibilité d'accumulation de mutations délétères et en restant dans un cadre stochastique. Pour cela, l'idée directrice était d'adapter le travail récent de Tomasevic et al., 2022, basé sur le résultat de Bansaye et al., 2022, dans un cadre multidimensionnel afin de considérer plusieurs sites porteurs de mutations délétères. Le cadre de processus de branchement apparait naturellement avec l'hypothèse d'individus autofécondants. J'ai finalement proposé d'adopter un point de vue continu pour représenter l'accumulation de mutations délétères sous forme hétérozygote ou homozygote, ce qui permettait de limiter la dimension du trait individuel considéré.

1 Heuristic of the model and motivations

Simulations showed that an inversion, that suppress recombination, could fix in a population if it was favored in a heterozygous state and if it encompassed a permanently heterozygous locus, the latter garanteeing that the inversion was kept in a heterozygous state (Jay et al., 2022b). This advantage could be due to deleterious mutations carried by the inverted haplotype being less numerous, rarer, or more heterozygous compared to the population average on the same region. We thus needed information on the dynamics of deleterious mutations in the vicinity of a permanently heterozygous mating-type locus. A first step is presented in Tezenas et al., 2022, where we studied the fate of a single mutation near a mating-type locus. A stochastic model and simulations showed that the presence of a mating-type locus had a sheltering effect on a single recessive deleterious or overdominant mutation, inducing a longer maintenance of such mutations in the population. Moreover, this effect was strongest for highly selfing populations. Here, we try to extend these previous studies and obtain analytical results on the dynamics of multiple deleterious mutations segregating near a mating-type locus. We compare these dynamics for recombining and non-recombining individuals to test the conditions for suppressed recombination to be maintained around a permanently heterozygous locus. Since the impact of a permanently heterozygous mating-type locus was strongest for high rates of selfing, we focus on populations reproducing only via selfing. In particular, recombining (resp. non-recombining) individuals will only produce recombining (resp. non-recombining) individuals. Details on the reproduction, recombination, mutation and selection dynamics are given in the next section.

We consider a population of diploid individuals represented by two homologous chromosomes that carry deleterious mutations and a permanently heterozygous mating-type locus that is always heterozygous. The two

mating-type alleles will be denoted by a and b. We describe an individual by a vector of 4 coordinates :

— The first coordinate (R) indicates whether the chromosome can recombine or not during meiosis, and can take values 0 or 1. We assume that the suppression of recombination affects the entire chromosome, or, equivalently, we focus our attention on the region of the chromosome that shows recombination suppression.

— The second and third coordinates (x_a and x_b) indicate the "mutation load" carried by chromosome carrying each mating type, represented by a number in $[0, 1]$. In order to model a large number of deleterious mutations, we chose to use a continuous setting rather than a discrete one. The chromosome of interest is supposed to be very long (with an infinite number of sites), and x_a can be seen as the proportion of sites occupied by deleterious mutations on chromosome a. x_b is the same proportion for chromosome b.

— The fourth and last coordinate (ω) quantifies the homozygosity for deleterious mutations of the diploid individual. It is also a number in $[0, 1]$ and represents the fraction of sites that are homozygous for deleterious mutations. The proportion of sites heterozygous for deleterious mutations in the diploid individual can then be obtained as $x_{het} = x_a + x_b - 2\omega$.

We introduce a measure-valued branching process to mathematically describe the population, and try to derive results using branching process techniques. We compare recombining and non-recombining populations, and are particularly interested in the impact of recombination on population dynamics, such as the population extinction probability and extinction time, the population growth rate, and potential limits for the distribution of deleterious mutations in non-extinct individuals.

Section 2 is dedicated to the pathwise description of the process, and to the proof of some basic properties (existence, uniqueness, martingale property). We then discuss how this model fits a previously described framework (Marguet, 2019b, Marguet, 2019a, Bansaye et al., 2022) and which properties could be obtained with this framework (section 3). We describe the limitations we encountered, which prevented us from concluding with those mathematical tools, but which we think raise interesting questions for future work. In section 4, we propose a last point of view which emerged from the previous limitations and which is based on unitype branching processes to study local extinction. We conclude with some perspectives in section 5.

2 Mathematical definition of the model

This section is focused on the mathematical definition of the process (Section 2.1), and its elementary properties (Section 2.2).

2.1 Pathwise description of the process

We use the framework of Fournier et Méléard, 2004, Champagnat et Méléard, 2007 and Marguet, 2019b.

Each individual is described by a vector with four coordinates, (R, x_a, x_b, ω), taking values in $\mathcal{X} := \{0, 1\} \times [0, 1]^3$. We will use the notation $E = [0, 1]^3$.

As explained earlier, R is the recombination trait : if $R = 0$, the two homologous chromosomes cannot recombine during meiosis. If $R = 1$, recombination can occur (see the paragraph describing how recombination is modelled). x_a and x_b are the proportions of sites carrying a deleterious mutation on the chromosome of respectively mating type a and b. ω is the proportion of sites that are homozygous. The following equations hold : $x_a = \omega + x_{a,het}$ and $x_b = \omega + x_{b,het}$, with $x_{a,het}$ and $x_{b,het}$ being the proportion of sites that are heterozygous for the mutation, for each chromosome. In other words, ω represents the proportion of deleterious sites that are shared between the two chromosomes, $x_{a,het}$ and $x_{b,het}$ the proportions of deleterious sites that are exclusive to each chromosome. We also denote by $x_{het} = x_{a,het} + x_{b,het} = x_a + x_b - 2\omega$ the total proportion of deleterious mutations in a heterozygous state in an individual genotype.

In particular, we have

$$\max(0, x_a + x_b - 1) \leq \omega \leq \min(x_a, x_b).$$

Indeed, a site homozygous for a deleterious mutation is counted in both x_a and x_b, which gives the upper bound. The lower bound is obtained by noting that the total proportion of sites occupied by mutations (in a heterozygous or homozygous state) cannot exceed one. We have $x_{a,het} + x_{b,het} + \omega \leq 1$, which we can also write $x_a + x_b - \omega \leq 1$ to obtain the lower bound.

Because we are interested only in the mutation load of individuals, we will use the terms of individual "homozygosity" (resp. "heterozygosity") to refer to the proportion of sites that are homozygous (resp. heterozygous) for deleterious mutations in an individual genotype.

The dynamics of the population is described by a measure-valued process with values in $\mathcal{M}_F(\mathcal{X})$, the set of all finite counting measures on $\mathcal{X}$, the trait space. More precisely, at each time $t > 0$, we set

$$\mathcal{Z}_t := \sum_{i \in V_t} \delta_{x_t^i},$$

where V_t is the index set of individuals present at time t in the population, and x_t^i the trait in $\mathcal{X}$ of individual i at time t.

We consider a birth-and-death process constructed as follows.

Reproduction

All individuals reproduce at rate 1. If an individual with trait $x = (R, y)$, with $y \in [0,1]^3$, does not recombine ($R = 0$), its reproduction is clonal. The offspring produced has then the same genotype as its parent, *i.e.* (R, y).

If an individual $(R, x_a, x_b, \omega) \in \mathcal{X}$ can recombine ($R = 1$), it does so with probability r. With probability $1 - r$, its reproduction is clonal. With probability r, the individual produces a tetrad composed of four haploid cells containing one chromosome each. Two chromosomes have recombined and exchanged a proportion $1 - \theta$ of their length (see Figure 3.1). The random proportion θ is sampled uniformly at random in $[0,1]$. A chromosome with mating type a and a chromosome with mating type b are then chosen uniformly at random among the chromosomes of the tetrad to form an offspring, for which the heterozygosity at the mating-type locus is preserved. The third coordinate is computed using the deleterious mutation densities on the selected chromosomes (see Figure 3.1 for a schematic representation). The offspring genotype is then (x_a, x_b, ω), $(x_a, \theta x_b + (1-\theta)x_a, \theta\omega + (1-\theta)x_a)$, $(\theta x_a + (1-\theta)x_b, x_b, \theta\omega + (1-\theta)x_b)$, or $(\theta x_a + (1-\theta)x_b, \theta x_b + (1-\theta)x_a, \omega)$, each with probability $1/4$ (Figure 3.1).

Several underlying assumptions are made here. First, we assume that there can be only one event of recombination in a tetrad. Second, we do not keep track of the mutation positions, and assume that they are uniformly distributed along the genome. The homozygosity of an offspring is thus a mean homozygosity. These assumptions allow to develop a mathematical model that is amenable to analysis.

Mutation

Individuals mutate at rate μ. A mutation event occurs on one chromosome, and loads a proportion ε of the previously mutation-free part of the chromosome with deleterious mutations. For example, if a mutation event occurs on the a chromosome of an individual, a proportion $\varepsilon(1 - x_a)$ is added to the second coordinate, describing the a-chromosome load (cf Figure 3.2). The homozygosity (fourth coordinate) is changed as well, the amount of homozygosity added being equal to the additional load $\varepsilon(1 - x_a)$ multiplied by the probability that this new load faces an already present load on the b-chromosome. This probability is equal to the proportion of the b chromosome that is occupied by heterozygous deleterious mutations, divided by the length on which the additional load could have appeared on the a chromosome : $\frac{x_b - \omega}{1 - x_a}$. The additional homozygosity is then equal to $\varepsilon(1-x_a) \times (x_b - \omega)/(1-x_a) = \varepsilon(x_b - \omega)$ (Figure 3.2). If the additional load initially appears on the b chromosome, the change in the individual genotype is obtained by inverting the roles of a and b in the previous derivations. More precisely, when an individual mutates, its genotype becomes $(R, x_a + \varepsilon(1 - x_a), x_b, \omega + \varepsilon(x_b - \omega))$ with probability $1/2$, or $(R, x_a, x_b + \varepsilon(1 - x_b), \omega + \varepsilon(x_a - \omega))$ with probability $1/2$.

Death

An individual (R, x_a, x_b, ω) dies at rate

$$d(x_a, x_b, \omega) := s_0 + s_{hom}\omega + s_{het}(x_a + x_b - 2\omega), \tag{3.1}$$

where s_0 is the intrinsic death rate, $s_{hom} \geq 0$ is the selection coefficient describing the deleterious effect of mutations in a homozygous state, and $s_{het} \geq 0$ the coefficient describing the deleterious effect of mutations in a heterozygous state. The proportion of heterozygous sites is $x_{het} = x_a + x_b - 2\omega$.

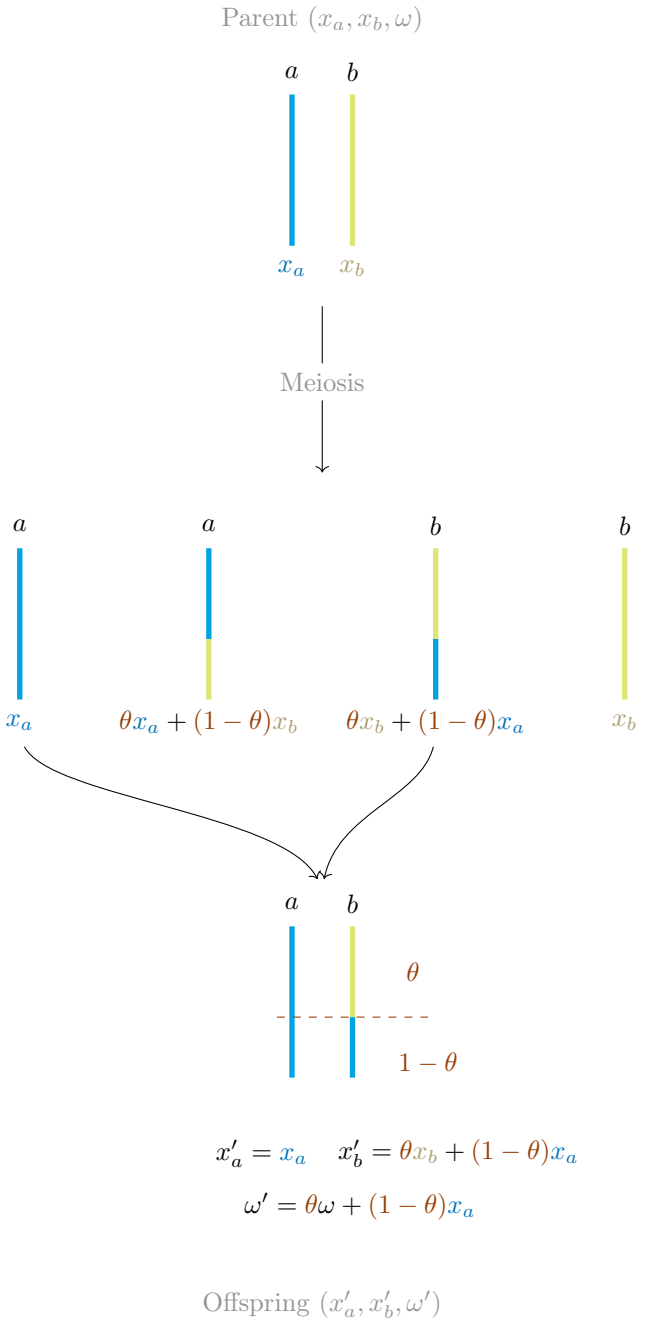


FIGURE 3.1 – Example of an offspring production with recombination. A parent with genotype (x_a, x_b, ω) produces a tetrad through a meiosis event during which recombination occurs. The recombination event exchanges a portion $1-\theta$ of the two central chromosomes. One chromosome of each mating type (a and b) is chosen to form the offspring (x'_a, x'_b, ω'), respecting the heterozygosity condition at the mating-type locus. In the example presented in this figure, the homozygosity of the offspring is the same as in the parent on a proportion θ of the offspring chromosomal pair, each chromosome carrying a proportion x_a (resp. x_b) of sites with deleterious mutations. On the remaining length, *i.e.* on a proportion $1-\theta$ of the offspring chromosomes, the two mating-type chromosomes carry the same proportion of sites with deleterious mutations : x_a. The homozygosity in this section is thus equal to x_a. The offspring homozygosity is then computed by adding the two contributions. Note that this is a schematic representation of the chromosomes that does not represent their spatial composition but rather illustrates the modelization through proportions.

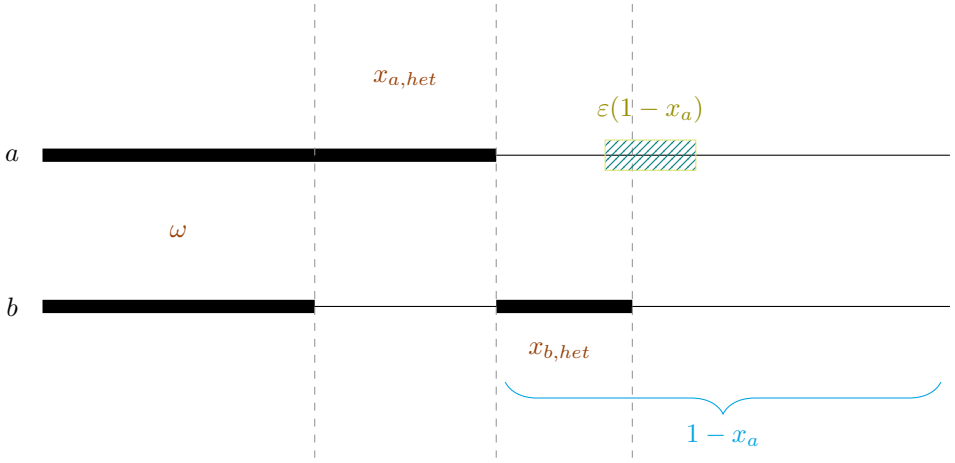


FIGURE 3.2 – Two chromosomes (a and b) of an individual are represented. Each black segment models the proportions of each chromosome that harbors deleterious mutations. A mutation event can occur and add a proportion $\varepsilon(1-x_a)$ to the a chromosome (represented by a green zone). The added homozygosity is then computed with the probability that this new load faces an already present load on the b chromosome. This probability is equal to the proportion of the b chromosome that is occupied by deleterious mutations, divided by the length on which the additionnal load could have appeared on chromosome a (represented with a blue bracket) : $x_{b,het}/(1-x_a) = (x_b-\omega)/(1-x_a)$. The additional homozygosity is then equal to $\varepsilon(1-x_a)\times(x_b-\omega)/(1-x_a) = \varepsilon(x_b-\omega)$. Note that this is a schematic representation of the chromosomes that does not represent their spatial composition but rather illustrates the modelization through proportions.

We now detail the mathematical construction of the process.

Let N_1 be a Poisson point measure on $\mathbb{R}_+\times\mathbb{N}^*\times\mathbb{R}_+\times[0,1]$ with intensity measure $ds\otimes\sum_{k\geq 1}\delta_k(di)\otimes du\otimes d\theta$, and N_2 and N_3 be two independent Poisson point measures on $\mathbb{R}_+\times\mathbb{N}^*\times\mathbb{R}_+$, with intensity measures $ds\otimes\sum_{k\geq 1}\delta_k(di)\otimes du$, N_2 and N_3 being independent from N_1.

We introduce, for $R\in\{0,1\}$, the following thresholds defined from the rates at which each event can occur to a single individual. We define

$$m_1(R) := 1-R+R\left((1-r)+\frac{1}{4}r\right) = 1-R\frac{3}{4}r \tag{3.2}$$

which is based on the rate at which a clonal reproduction event occurs (at rate $1-R$ for individuals with chromosomes that cannot recombine, and at rate $R\left(1-r+\frac{1}{4}r\right)$ for individuals with chromosomes that can recombine with probability r). We write

$$m_2(R) = m_1(R)+R\frac{3}{4}r = 1, \tag{3.3}$$

where we have added to m_1 the rate at which a reproduction with recombination occurs for a recombining individual, and

$$m_3(R) = m_2(R)+\mu = 1+\mu, \tag{3.4}$$

where we have added to m_2 the rate at which a mutation event occurs.

We also write $(R_i, y_{i,s})$ for the trait of the individual i at time s.

Following the description of the dynamics given above, for every $t \geq 0$ we define $\mathcal{Z}_t$ by :

$$\begin{aligned}\mathcal{Z}_t = \mathcal{Z}_0 &+ \int_0^t \int_{\mathbb{N}^* \times \mathbb{R}_+ \times [0,1]} \Bigg[\delta_{(R_i, y_{i,s^-})} \mathbb{1}_{\{i \leq \langle \mathcal{Z}_{s^-}, 1 \rangle, u \leq m_1(R_i)\}} \\ &\qquad + \sum_{k=1}^{3} \delta_{(R_i, f_{k,\theta}(y_{i,s^-}))} \mathbb{1}_{\{i \leq \langle \mathcal{Z}_{s^-}, 1 \rangle, m_1(R_i) + (k-1)\frac{r}{4} R_i < u \leq m_1(R_i) + k\frac{r}{4} R_i\}} \Bigg] N_1(ds, di, du, d\theta) \\ &+ \int_0^t \int_{\mathbb{N}^* \times \mathbb{R}_+} \sum_{j=1}^{2} \left[\delta_{(R_i, g_{j,\varepsilon}(y_{i,s^-}))} - \delta_{(R_i, y_{i,s^-})} \right] \mathbb{1}_{\{i \leq \langle \mathcal{Z}_{s^-}, 1 \rangle, m_2(R_i) + (j-1)\frac{\mu}{2} < u \leq m_2(R_i) + j\frac{\mu}{2}\}} N_2(ds, di, du) \\ &- \int_0^t \int_{\mathbb{N}^* \times \mathbb{R}_+} \delta_{(R_i, y_{i,s^-})} \mathbb{1}_{\{i \leq \langle \mathcal{Z}_{s^-}, 1 \rangle, m_3(R_i) < u \leq m_3(R_i) + d(y_{i,s^-})\}} N_3(ds, di, du),\end{aligned} \tag{3.5}$$

with

$$\begin{aligned} f_{1,\theta} \quad : \quad [0,1]^3 &\longrightarrow [0,1]^3 \\ y = (x_a, x_b, \omega) &\longmapsto (x_a, \theta x_b + (1-\theta) x_a, \theta\omega + (1-\theta) x_a), \\ f_{2,\theta} \quad : \quad [0,1]^3 &\longrightarrow [0,1]^3 \\ y = (x_a, x_b, \omega) &\longmapsto (\theta x_a + (1-\theta) x_b, x_b, \theta\omega + (1-\theta) x_b), \\ f_{3,\theta} \quad : \quad [0,1]^3 &\longrightarrow [0,1]^3 \\ y = (x_a, x_b, \omega) &\longmapsto (\theta x_a + (1-\theta) x_b, \theta x_b + (1-\theta) x_a, \omega), \end{aligned}$$

the functions describing the possible genotypes for an offspring genome created with recombination, and

$$\begin{aligned} g_{1,\varepsilon} \quad : \quad [0,1]^3 &\longrightarrow [0,1]^3 \\ y = (x_a, x_b, \omega) &\longmapsto (x_a + \varepsilon(1 - x_a), x_b, \omega + \varepsilon(x_b - \omega)), \end{aligned}$$

and

$$\begin{aligned} g_{2,\varepsilon} \quad : \quad [0,1]^3 &\longrightarrow [0,1]^3 \\ y = (x_a, x_b, \omega) &\longmapsto (x_a, x_b + \varepsilon(1 - x_b), \omega + \varepsilon(x_a - \omega)), \end{aligned}$$

the functions describing the possible genotypes obtained after mutation.

The first integral on the right hand side of (3.5) describes the possible jumps of the system due to reproduction at a time comprised in $[0, t]$. The second integral describes the occurence of mutation events, and the third integral encodes the death events occurring between times 0 and t.

Remark : Because of the assumption of selfing, it is possible to completely decorrelate the population of recombining individuals from the population of non-recombining individuals, since there are no interactions between the two populations. We could then have written the previous equations separately for each population, simplifying the expressions by replacing R with 0 or 1. However, a possible extension of this work is to consider rare outcrossing events, seen as deviations from the typical mating system in the population. For the sake of generality, inwhat follows we chose to keep the recombination factor R as a trait and not as a parameter in the global population.

2.2 Existence, Uniqueness, and Martingale property

In this section, we explain why a process defined as a solution to (3.5) exists, and lay out some basic properties. We rely on results proved in Fournier et Méléard, 2004, Tran, 2006 and derived works. We start by detailing how a solution to (3.5) can be constructed recursively. We then justify the existence of the process (*i.e.* the non-accumulation of jump times in a finite time interval), and compute its infinitesimal generator, from which we obtain a useful martingale property.

Algorithmic construction of the process

As explained in Tran, 2006 section 2.3.1, or Champagnat et Méléard, 2007, starting from an initial condition Z_0, a process solution of (3.5) can be constructed recursively. We now detail how the sequence of jump times and the associated states of the process can be constructed, adapting the neatly written work of

Tran, 2006. Note that the iterations are simpler in our case, because the trait of an individual does not vary between jumps. To shorten the notation, we write $N_t := \langle \mathcal{Z}_t, 1 \rangle$.

Let $\mathcal{Z}_0$ be a random variable with values in $\mathcal{M}_F(\mathcal{X})$ and such that $\mathbb{E}[N_0] < \infty$. By convention, $T_0 = 0$.

Let us proceed by induction. Let $n \in \mathbb{N}$ be fixed, and suppose we have constructed the sequence $\left(\mathcal{Z}_{T_k}, T_k\right)_{k \in \mathbb{N}}$ up to the n^{th} step. The process $(\mathcal{Z}_t)_{t \geq 0}$ is thus well defined on $[0, T_n[$, as individual traits do not evolve between jumps.

We use an acceptance-rejection method. The rate at which an event occurs to a single individual is bounded by $C := 1 + \mu + s_0 + s_{hom} + s_{het}$. The total rate at which an event occurs in a population of size N is then bounded by $C \times N$. Recall the definition of the thresholds $m_i(R)$ given in (3.2), (3.3) and (3.4). Initiate with $\tau_n := T_n$, and $\ell = 0$ (which will be the number of rejections at each step). The iterations over $\ell \in \mathbb{N}$ are then as follows.

1. Let $\epsilon_{n+\ell+1}$ be a random variable independent of all other variables, which follows an exponential distribution of parameter 1.

   We then set $\tau_{n+\ell+1} = \tau_{n+\ell} + \epsilon_{n+\ell+1}/CN_{T_n}$.

   This step defines the waiting time until an event potentially occurs in the population.

2. Let I_n be the realization of a uniformly distributed variable on $[\![1, N_{T_n}]\!]$. We denote by (R_{I_n}, y_{I_n}) the trait of the I_n^{th} individual in the population $\mathcal{Z}_{T_n}$.

   This is the random choice of the individual who will be involved in a possible jump.

3. Let then Θ_n be a random variable uniformly distributed over $[0, 1]$.

   — If $0 \leq \Theta_n \leq \dfrac{m_1(R_{I_n})}{C}$, the following $n+1^{st}$ event occurs : the I_n^{th} individual of the population at time T_n reproduces clonally. Then we set $T_{n+1} = T_n + \tau_{n+\ell+1}$ and $\mathcal{Z}_{T_{n+1}} = \mathcal{Z}_{T_n} + \delta_{(R_{I_n}, y_{I_n})}$.

   — If $\dfrac{m_1(R_{I_n}) + (k-1)\frac{r}{4}R_{I_n}}{C} < \Theta_n \leq \dfrac{m_1(R_{I_n}) + k\frac{r}{4}R_{I_n}}{C}$, for some $k \in \{1, 2, 3\}$, the following $n+1^{st}$ event occurs : the I_n^{th} individual of the population at time T_n reproduces with recombination. Let then θ be a random variable uniformly distributed over $[0, 1]$. We set $T_{n+1} = T_n + \tau_{n+\ell+1}$ and $\mathcal{Z}_{T_{n+1}} = \mathcal{Z}_{T_n} + \delta_{(R_{I_n}, f_{k,\theta}(y_{I_n}))}$.

   — If $\dfrac{m_2(R_{I_n}) + (j-1)\frac{\mu}{2}}{C} < \Theta_n \leq \dfrac{m_2(R_{I_n}) + j\frac{\mu}{2}}{C}$, for one $j \in \{1, 2\}$, the following $n+1^{st}$ event occurs : the I_n^{th} individual of the population at time T_n mutates. We set $T_{n+1} = T_n + \tau_{n+\ell+1}$ and $\mathcal{Z}_{T_{n+1}} = \mathcal{Z}_{T_n} - \delta_{(R_{I_n}, y_{I_n})} + \delta_{(R_{I_n}, g_{j,\varepsilon}(y_{I_n}))}$.

   — If $\dfrac{m_3(R_{I_n})}{C} < \Theta_n \leq \dfrac{m_3(R_{I_n}) + d(y_{I_n})}{C}$, the following $n+1^{st}$ event occurs : the I_n^{th} individual of the population at time T_n dies. We set $T_{n+1} = T_n + \tau_{n+\ell+1}$ and $\mathcal{Z}_{T_{n+1}} = \mathcal{Z}_{T_n} - \delta_{(R_{I_n}, y_{I_n})}$.

   — Finally, if $\dfrac{m_3(R_{I_n}) + d(y_{I_n})}{C} < \Theta_n \leq 1$, nothing happens and we increment ℓ by 1, coming back to step 1.

   This step determines if a jump happens or not in the population, based on an acceptance/rejection method. All cases but the last one result in a jump. Note that the thresholds are directly computed from the pathwise definition (3.5).

Existence and uniqueness

The previous algorithm gives an explicit construction of a solution to (3.5). There is left to prove that such a construction does not lead to the accumulation of jumps in a finite time to guarantee that the process is well defined. We obtain the following result by a straightforward adaptation of Theorem 3.1 in Fournier et Méléard, 2004 :

Proposition 1
Let $\mathcal{Z}_0 \in \mathcal{M}_F(\mathcal{X})$ such that $\mathbb{E}[\langle \mathcal{Z}_0, 1 \rangle] < +\infty$.

1. There exists a unique solution to (3.5) with initial condition $\mathcal{Z}_0$.
2. If furthermore there exists $p \geq 1$ such that $\mathbb{E}(\langle \mathcal{Z}_0, 1 \rangle^p) < +\infty$, then for any $T < +\infty$,

$$\mathbb{E}\left(\sup_{t \in [0,T]} \langle \mathcal{Z}_t, 1 \rangle^p \right) < +\infty.$$

Proof. The proof of Theorem 3.1 in Fournier et Méléard, 2004 relies on Assumption **A** of the paper, which requires that the trait-dependent rates at which an event can occur to an individual are uniformly bounded on the state space, and that the probability kernel that describes the distribution of the trait of a newly produced individual is bounded by a measure absolutely continuous with respect to Lebesgue measure on the trait space. Even if the framework of our model slightly differs from the model in Fournier et Méléard, 2004 (besides intrinsic death and birth, we consider mutation, and they consider competition between individuals), a similar assumption is satisfied. Indeed, in our model, the birth and mutation rates are constant, and the death rate is uniformly bounded on the state space as a continuous function on the compact set E. Moreover, our reproduction and mutation laws are probabilities, that can be bounded by the uniform distribution on $[0,1]^3$ multiplied by a constant, which is a measure absolutely continuous with respect to Lebesgue measure. This allows to perform the same kind of computations as in Fournier et Méléard, 2004, and obtain the existence of the process and item 2. of Proposition 1.

Uniqueness relies on the algorithmic construction. As shown by induction in Prop. 2.2.6 of Tran, 2006, the sequence of jump times and associated states $(T_k, \mathcal{Z}_{T_k})_{k\in\mathbb{N}}$ is uniquely defined by the initial condition and the Poisson measure used during the algorithmic construction. This completes the proof of item 1. □

Remark : We will not look into the dependence of the population dynamics in its initial size. We will thus assume in the rest of this chapter that the initial population is of deterministic finite size. The initial traits however can be chosen randomly. More precisely, $\mathcal{Z}_0$ will be a random variable with values in $\mathcal{M}_F(\mathcal{X})$, but its size $N_0 = \langle \mathcal{Z}_0, 1\rangle$ will be deterministic and finite. An example is given in section 4.3 where the stochastic simulations conducted to illustrate this chapter are described. Consequently, the assumptions on the boundedness of the mean population size required in Proposition (1) are satisfied.

Martingale property

Using Proposition 2.6 of Fournier et Méléard, 2004, we now prove that a solution to (3.5) satisfies the Markov property, and compute its infinitesimal generator.

Proposition 2
Consider a solution $(\mathcal{Z}_t)_{t\geq 0}$ to (3.5). Then $(\mathcal{Z}_t)_{t\geq 0}$ is a Markov process.
Its extended generator L is defined for all continuous functions $F : \mathbb{R} \to \mathbb{R}$, all bounded and measurable functions $f : \mathcal{X} \longrightarrow \mathbb{R}$, and all measures $\nu \in \mathcal{M}_F(\mathcal{X})$ by

$$\begin{aligned}
LF_f(\nu) := &\int_{\mathcal{X}} (1-R)\Big(F\Big(\langle \nu + \delta_{(R,y)}, f\rangle\Big) - F\Big(\langle \nu, f\rangle\Big)\Big)\nu(dR,dy) \\
&+ \int_0^1 \int_{\mathcal{X}} R\Big((1-\frac{3}{4}r)F\Big(\langle \nu + \delta_{(R,y)}, f\rangle\Big) \\
&\qquad + \frac{r}{4}\Big[F\Big(\langle \nu + \delta_{(R,f_{1,\theta}(y))}, f\rangle\Big) + F\Big(\langle \nu + \delta_{(R,f_{2,\theta}(y))}, f\rangle\Big) + F\Big(\langle \nu + \delta_{(R,f_{3,\theta}(y))}, f\rangle\Big)\Big] \\
&\qquad\qquad - F\Big(\langle \nu, f\rangle\Big)\Big)\nu(dR,dy)d\theta \\
&+ \mu \int_{\mathcal{X}} \Big(\frac{1}{2}\Big[F\Big(\langle \nu + \delta_{(R,g_{1,\varepsilon}(y))} - \delta_{(R,y)}, f\rangle\Big) + F\Big(\langle \nu + \delta_{(R,g_{2,\varepsilon}(y))} - \delta_{(R,y)}, f\rangle\Big)\Big] - F\Big(\langle \nu, f\rangle\Big)\Big)\nu(dR,dy) \\
&+ \int_{\mathcal{X}} d(y)\Big(F\Big(\langle \nu - \delta_{(R,y)}, f\rangle\Big) - F\Big(\langle \nu, f\rangle\Big)\Big)\nu(dR,dy).
\end{aligned} \tag{3.6}$$

Proof. The algorithmic construction given above guarantees that the process is Markovian.
The derivation of the generator is a direct adaptation of part 2.2.2. of Tran, 2006. The first step is to integrate (3.5) against f, and apply Itô's formula with F (Ikeda et Watanabe, 1981 Th. 5.1, p66). The second step is to take expectations, and compute the derivative with respect to t to identify the generator. We can exchange differentiation and expectation as all functions are bounded on the compact set E. □

We now obtain a martingale property, which is an important step in the study of the process dynamics. We apply Proposition 3.4 of Fournier et Méléard, 2004. Their Assumption A is satisfied in our case, as

explained before (the individual rates of events are uniformly bounded, as are the production kernels). Moreover, we assumed that the initial population size was deterministic and finite. We can then apply the result of Prop 3.4. (iii) in Fournier et Méléard, 2004 with $p = 2$ and $q = 1$. This yields the following result :

Proposition 3
For any bounded and measurable function f on $\mathcal{X}$, the process

$$\left(M_t^f\right)_{t\geq 0} := \left(\langle \mathcal{Z}_t, f\rangle - \langle \mathcal{Z}_0, f\rangle - \int_0^t L\mathcal{I}d_f\left(\mathcal{Z}_s\right) ds\right)_{t\geq 0}$$

is a mean-zero martingale.

3 Branching point of view and Many-to-One formula

In this section, we describe the attempts to analyze the previously defined model using the framework of Marguet, 2019b and derived works. The aim is to study the dynamics of the process in the long term. We are specifically interested in the population growth rate and the existence of a limiting distribution for the traits. We detail the limitations encountered that restrained our ability to conclude with those mathematical tools.

3.1 Many-to-One fomulae

The process defined in the previous section is a continuous time branching process, with trait-dependent rates and reproduction laws. Marguet, 2019b provides an adequate framework to study the evolution in time of such processes. More precisely, the author proves several Many-to-One formulae that allow to link the mean dynamics of the total population to the mean dynamics of a "typical" lineage, whose reproduction law is biased by the size of the subpopulation it generated in the present population. In this section, we first describe how our model can fit the framework of Marguet, 2019a, and then focus on the first Many-to-One formula proved by the author.

To fit the framework of Marguet, 2019a, we need to adopt a "branching" point of view, in which the population evolves by replacing individuals with their offspring. More precisely, we now consider that death, mutation and reproduction are all branching events, that result in the death of the parent and the production of zero, one or two offspring respectively. The number of offspring follows a law that depends on the trait of the reproducting individual, and the offspring traits can be different from the parent trait (the reproduction law is said to be *non local*). More precisely, we have :

- The trait X^u of an individual u does not evolve during its lifetime. The associated generator, denoted by $\mathcal{G}$ in Marguet, 2019b is thus equal to zero.
- An individual with trait (or genotype) $(R, y) \in \{0, 1\} \times [0, 1]^3$ dies at rate $1 + \mu + d(y)$.
- At death, the individual is replaced by $0, 1$ or 2 offspring with respective probabilities $p_0(R, y) = \frac{d(y)}{1+\mu+d(y)}$, $p_1(R, y) = \frac{\mu}{1+\mu+d(y)}$ and $p_2(R, y) = \frac{1}{1+\mu+d(y)}$.
- The traits of the offspring of a parent with trait (R, y) are characterized by the following distributions :
  - If the number of offspring is one (mutation event), the offspring trait distribution is $P_1^{(1)}(R, y, .) = \frac{1}{2}\delta_{(R,g_{1,\varepsilon}(y))} + \frac{1}{2}\delta_{(R,g_{2,\varepsilon}(y))}$;
  - If the number of offspring is 2 (reproduction event), one of them is considered as the parent that does not die. Its trait does not change, and thus follows the distribution $P_1^{(2)}(R, y, .) = \delta_{(R,y)}$. The trait of the second, new, offspring then follows the distribution $P_2^{(2)}(R, y, .) = (1 - R)\delta_{(R,y)} + R(1 - \frac{3}{4}r)\delta_{(R,y)} + R\frac{r}{4}\sum_{j=1}^{3}\int_0^1 \delta_{(R,f_{j,\theta}(y))}\mathrm{d}\theta$. The first term corresponds to the reproduction of a non-recombining individual (clonal reproduction). The second term corresponds the reproduction of a recombining individual for which recombination did not occur (this happens with probability $1 - r$) or when the chromosomes selected to form the offspring genotype were not the ones that recombined (probability $r/4$). The last term describes a reproduction with recombination.

We emphasize that here we have chosen to consider mutational events as branching events, resulting in $p_1 \neq 0$. However, we could also have chosen to see them as an independent process that acts between "real" birth and death events. Of course, the two points of view give the same results for the Many-to-One formulae.

An important quantity for the following Many-to-One results is the mean population size at time $t > 0$ starting from a single individual of type $x \in \mathcal{X}$ at time $s \in [0,t]$, denoted by $m(x,s,t) := \mathbb{E}\left[N_t | \mathcal{Z}_s = \delta_x\right]$.

In the examples provided in Marguet, 2019b, an explicit formula can be derived for this mean population size, by integrating the pathwise definition of the process against the function equal to 1. This leads to a closed differential equation on m, which can be explicitely solved. In our case, we obtain, for $x \in \mathcal{X}$ and $0 < s < t$,

$$m(x,s,t) = 1 + \int_s^t \Big(m(x,s,u) - \mathbb{E}\left[\int_E d(y) \mathcal{Z}_u(\mathrm{d}R, \mathrm{d}y) | \mathcal{Z}_s = \delta_x \right] \Big) \mathrm{d}u.$$

The trait dependence of the death rate thus prevents us from solving the differential equation satisfied by the expected population size. This is the first limitation encountered.

To study the evolution of the process through time, we aimed to apply the result of Theorem 3.1 of Marguet, 2019b to our process to obtain the following Many-to-One formula : for all $t > 0$, all $x_0 \in \mathcal{X}$, and any non-negative measurable function F from the set of càdlàg measure-valued processes on $[0,t]$ to $\mathbb{R}_+$,

$$\mathbb{E}_{\delta_{x_0}} \left[\sum_{i \in V_t} F(\mathcal{Z}_s^i, s \leq t) \right] = m(x_0, 0, t) \mathbb{E}_{x_0} \left[F(Y_s^{(t)}, s \leq t) \right]. \tag{3.7}$$

We use again V_t, the index set of individuals present in the population at time t, and denote by Z_s^i the trait at time s of the i^{th} individual alive at time t. This formula links the mean dynamics of the measure-valued process trajectories between 0 and t (left term) to the mean dynamics of a trajectory of a "typical" trait described by $(Y_t)_{t \geq 0}$, a time-inhomogeneous Markov process on the trait space. It is a powerful result, as it sums up the mean dynamics of a measure-valued process into the dynamics of a Markov process on a subset of $\mathbb{R}^d$. The infinitesimal generator of this "typical" auxiliary process is, for any non-negative measurable function f, $(R,y) \in \mathcal{X}$, and $0 \leq s < t$,

$$\begin{aligned} A_s^{(t)} f(R,y) = \Bigg(& \frac{\mu}{2} \sum_{j=1}^{2} \frac{m((R, g_{j,\varepsilon}(y)), s, t)}{m((R,y), s, t)} \Big(f(R, g_{j,\varepsilon}(y)) - f(R,y) \Big) \\ & + R \frac{r}{4} \sum_{k=1}^{3} \int_0^1 \frac{m((R, f_{k,\theta}(y)), s, t)}{m((R,y), s, t)} \Big(f(R, f_{k,\theta}(y)) - f(R,y) \Big) \mathrm{d}\theta \Bigg). \end{aligned} \tag{3.8}$$

The precise knowledge of the "typical" process generator allows in particular to operate effective simulations, as simulating the evolution of a process on a subset of $\mathbb{R}^d$ is less time and memory consuming than simulating the evolution of a measure-valued process.

In addition, Marguet, 2019a proves a law of large numbers for the empirical distribution of ancestral trajectories of the measure-valued process. Under the assumption of ergodicity of the auxiliary process, the empirical distribution of ancestral trajectories converges in $\mathbb{L}_2$, as time goes to infinity, towards a deterministic variable that depends on the asymptotic dynamics of the "typical" auxiliary process (Marguet, 2019a). Again, this result is powerful as it links the asymptotic dynamics of a measure-valued process to the (simpler) dynamics of a Markov process on a subset of $\mathbb{R}^d$, whose generator is known.

However in our case, the absence of precise results on the average population size induces two limitations :

— Theorem 3.1 of Marguet, 2019b that gives the Many-to-One formula (3.7) relies on assumptions regarding the mean population size. Assumption **C** in particular requires that the ratio of mean population sizes of the form $\frac{m(x_1,s,t)}{m(x_0,s,t)}$, where x_1 is a possible trait of an offspring produced by a parent of type x_0, is controlled uniformly over the trait space $\mathcal{X}$, and uniformly in $s \in [0,t]$. Obtaining this type of control without an explicit formula for the mean population size seems difficult.

— The infinitesimal generator of the "typical" process depends on the previously discussed ratios of the form $\frac{m(x_1,s,t)}{m(x_0,s,t)}$. The absence of an explicit expression for the mean population size prevents the use of this generator to simulate the associated process.

These limitations prevented us from using the very practical results brought by the Many-to-One formula. We now turn to a more general result to study the dynamics of the measure-valued process in long time.

3.2 Law of Large Numbers on the first moment semigroup

A key element to study the dynamics of a continuous-time Markov process $(\mathcal{Z}_t)_{t\geq 0}$ is its first moment semigroup, which describes the mean behavior of the process and is defined by

$$M_t f(x) := \mathbb{E}[\langle \mathcal{Z}_t, f\rangle | \mathcal{Z}_0 = \delta_x], \tag{3.9}$$

for bounded measurable functions f on $\mathcal{M}_F(\mathcal{X})$ and $x \in \mathcal{X}$. The Law of Large Numbers result in Marguet, 2019a is based on assumptions satisfied by the first moment semi-group of the auxiliary process, which is time-dependent and non-conservative. Similar assumptions are given in Bansaye et al., 2022 to prove a more general result on the convergence of non-conservative semigroups. Tomasevic et al., 2022 applied this result to a particular case of growth-fragmentation processes, and were able to derive explicit formulae for the growth rate and limiting trait distribution of the process they consider. We first present the result of Bansaye et al., 2022 and describe how we tried to conduct a similar approach as Tomasevic et al., 2022 to apply it to our process. We highlight the limitations we encountered along the way.

Convergence theorem for semigroups

We present here the framework and result of Bansaye et al., 2022 briefly and highlight why it can be of interest in our case.

Let us consider a measurable functions $V : \mathcal{X} \to \mathbb{R}_+^*$. Let $\mathcal{B}(V)$ be the space of measurable functions $f : \mathcal{X} \to \mathbb{R}$ that are dominated by V, *i.e.* such that $\|f\|_{\mathcal{B}(V)} := \sup_{x\in\mathcal{X}} \frac{|f(x)|}{V(x)} < \infty$. Let us also introduce the space $\mathcal{M}(V)$ which is the set of signed measures on $\mathcal{X}$ that integrate V. For a rigorous construction of this set, see Bansaye et al., 2022. Let $(M_t)_{t\geq 0}$ be a positive semigroup of kernel operators acting on $\mathcal{B}(V)$ (on the right) and $\mathcal{M}(V)$ (on the left), and satisfying the duality relation : for any $f \in \mathcal{B}(V)$ and $\mu \in \mathcal{M}(V)$, $(\mu M_t)(f) = \mu(M_t f)$. Let $\psi : \mathcal{X} \to \mathbb{R}_+^*$ be a measurable function such that $\psi \leq V$. We detail here the main assumption for the theorem.

Assumption A : There exist $\tau > 0$, $\beta > \alpha > 0$, $\theta > 0$, $0 < c, d \leq 1$, a subset $K \subset \mathcal{X}$ and a probability measure on $\mathcal{X}$ supported by K such that $\sup_K V/\psi < \infty$ and

1. $M_\tau V \leq \alpha V + \theta \mathbb{1}_K \psi$,
2. $M_\tau \psi \geq \beta \psi$,
3. $\inf_{x\in K} \dfrac{M_\tau(f\psi)(x)}{M_\tau\psi(x)} \geq c\nu(f)$ for all $f \in \mathcal{B}_+(V/\psi)$,
4. $\nu\left(\dfrac{M_{n\tau}\psi}{\psi}\right) \geq d \sup_{x\in K} \dfrac{M_{n\tau}\psi(x)}{\psi(x)}$ for all positive integers n.

The idea behind these assumptions is to introduce appropriate conservative operators, to apply results known in the conservative frameworks in a case where eigenelements are not known. We refer to Bansaye et al., 2022 and Tomasevic et al., 2022 for a detailed interpretation of these assumptions.

The result is stated as follows (Theorem 2.1 in Bansaye et al., 2022) :

Theorem 1
Let $V : \mathcal{X} \to]0, \infty[$ be measurable and $(M_t)_{t\geq 0}$ be a positive semigroup of kernel operators on $\mathcal{B}(V)$ and $\mathcal{M}(V)$ such that $t \mapsto \|M_t V\|_{\mathcal{B}(V)}$ is locally bounded on $[0, \infty)$.

— Let $\psi : \mathcal{X} \to (0, \infty)$ be a measurable function such that assumption **A** is satisfied. Then, there exists a unique triplet $(\gamma, h, \lambda) \in \mathcal{M}_+(V) \times \mathcal{B}_+(V) \times \mathbb{R}$ of eigenelements of M with $\gamma(h) = \|h\|_{\mathcal{B}(V)} = 1$ satisfying for all $t \geq 0$

$$\gamma M_t = \mathrm{e}^{\lambda t}\gamma \qquad \text{and} \qquad M_t h = \mathrm{e}^{\lambda t} h.$$

Moreover, there exist $C, \omega > 0$ such that, for all $t \geq 0$ and $\mu \in \mathcal{M}(V)$,

$$\|\mathrm{e}^{-\lambda t}\mu M_t - \mu(h)\gamma\|_{\mathcal{M}(V)} \leq C\|\mu\|_{\mathcal{M}(V)} e^{-\omega t}. \tag{3.10}$$

— Assume that there exist a triplet $(\gamma, h, \lambda) \in \mathcal{M}_+(V) \times \mathcal{B}_+(V) \times \mathbb{R}$ and constants $C, \omega > 0$ such that the previous equations hold. Then Assumption **A** is satisfied by the function $\psi = h$.

In our model, the semigroup that naturally arises is the first moment semigroup defined in equ.(3.9), which is non-conservative, as individuals can have zero or two offspring. This theorem thus provides a natural framework to study the long-time behavior of our process. A strength of this theorem is that it does not require previous knowledge of eigenelements, but rather proves their existence and provides estimation for them. However, the second point highlights the fact that eigenelements are good candidates to satisfy assumption **A**. In the following, we focus our attention on the search for those eigenelements in the case where $(M_t)_{t\geq 0}$ is the first moment semigroup defined in equ.(3.9).

Following Tomasevic et al., 2022, we define for $t \geq 0$ the *mean measure* as the measure $n_t \in \mathcal{M}_F(\mathcal{X})$ which satisfies, for bounded measurable function $f : \mathcal{X} \to \mathbb{R}$,

$$\langle n_t, f\rangle = M_t f = \mathbb{E}[\langle \mathcal{Z}_t, f\rangle]. \tag{3.11}$$

This mean measure is well defined thanks to point 2. of Proposition 1, and depends on the initial condition. We lighten the notation by making this dependence explicit only when necessary.

To identify good candidates for the triplet describing the asymptotic behaviour of $(n_t)_{t\geq 0}$, the idea is to look for a mean measure of the form $n_t^i(\mathrm{d}x) = \mathrm{e}^{-\lambda_i t} N_i(\mathrm{d}x)$, where λ_i is the growth rate of the subpopulation i (in our case, the recombining or non-recombining subpopulation), and N_i is a stationary distribution for the mean measure. The measure N_i naturally arises as a solution to an integro-differential system satisfied by n_t^i, and λ_i naturally arises as an eigenvalue of the adjoint operator associated to the mean measure. The convergence result is then obtained in the form of equation (3.10) with $\lambda = \lambda_i$ and $\gamma = N_i$ (Theorem 1.3 of Tomasevic et al., 2022). We now detail the strategy used by Tomasevic et al., 2022 and the limitations we encountered when trying to apply it to our model.

A differential system satisfied by the mean measure and the associated spectral problem

We focus first on the stationary profiles N_i. Taking the expectation in the expression for the martingale M_t^f defined in Proposition 3, we obtain

$$\langle n_t, f\rangle - \langle n_0, f\rangle = \int_0^t \mathbb{E}\big(L\mathcal{I}d_f(\mathcal{Z}_s)\big)ds,$$

which can be rewritten

$$\langle n_t, f\rangle - \langle n_0, f\rangle = \int_0^t \Big(\langle n_s, f_0\rangle + \int_0^1 \langle n_s, f_\theta\rangle d\theta + \langle n_s, f_\varepsilon\rangle - \langle n_s, f_d\rangle\Big)ds, \tag{3.12}$$

with $f_0(R,y) = (1-R)f(R,y)$ corresponding to the birth of non-recombining individuals, $f_\theta(R,y) = R\Big[\frac{r}{4}\Big(f(R,f_{1,\theta}(y)) + f(R,f_{2,\theta}(y)) + f(R,f_{3,\theta}(y))\Big) + (1-\frac{3}{4}r)f(R,y)\Big]$ corresponding to the birth of recombining individuals, $f_\varepsilon(R,y) = \mu\Big[\frac{1}{2}\Big(f(R,g_{1,\varepsilon}(y)) + f(R,g_{2,\varepsilon}(y))\Big) - f(R,y)\Big]$ corresponding to mutations, and $f_d(R,y) = d(y)f(R,y)$ corresponding to deaths.

Let us decompose n_t into the marginal measures n_t^1 and n_t^0 on $[0,1]^3$ defined by :

$$\langle n_t, f\rangle = \int_{[0,1]^3} f(0,y)n_t^0(dy) + \int_{[0,1]^3} f(1,y)n_t^1(dy), \quad \forall f \in \mathcal{C}_b(\mathcal{X}).$$

By evaluating equation (3.12) for test functions such that $f(1,.) \equiv 0$ and then test functions such that $f(0,.) \equiv 0$, we obtain integro-differential equations on n_t^0 and n_t^1 respectively, written as follows for bounded measurable

functions $\varphi : [0,1]^3 \to \mathbb{R}$:

$$\begin{aligned}\int_{[0,1]^3} \varphi(y) n_t^1(dy) - \int_{[0,1]^3} \varphi(y) n_0^1(dy) \\ = \int_0^t \int_{[0,1]^3} \mu\Big[\frac{1}{2}\Big(\varphi(g_{1,\varepsilon}(y)) + \varphi(g_{2,\varepsilon}(y))\Big) - \varphi(y)\Big] n_s^1(dy)ds - \int_0^t \int_{[0,1]^3} d(y)\varphi(y) n_s^1(dy)ds \\ + \int_0^t \int_0^1 \int_{[0,1]^3} \Big(\frac{r}{4}\Big(\varphi(f_{1,\theta}(y)) + \varphi(f_{2,\theta}(y)) + \varphi(f_{3,\theta}(y))\Big) + (1 - \frac{3}{4}r)\varphi(y)\Big) n_s^1(dy)d\theta ds\end{aligned} \tag{3.13}$$

and

$$\begin{aligned}\int_{[0,1]^3} \varphi(y) n_t^0(dy) - \int_{[0,1]^3} \varphi(y) n_0^0(dy) = \int_0^t \int_{[0,1]^3} \varphi(y) n_s^0(dy)ds - \int_0^t \int_{[0,1]^3} d(y)\varphi(y) n_s^0(dy)ds \\ + \int_0^t \int_{[0,1]^3} \mu\Big[\frac{1}{2}\Big(\varphi(g_{1,\varepsilon}(y)) + \varphi(g_{2,\varepsilon}(y))\Big) - \varphi(y)\Big] n_s^0(dy)ds.\end{aligned} \tag{3.14}$$

At this point, Tomasevic et al., 2022 prove a result that ensures that, if the initial distributions n_0^i admit densities with respect to Lebesgue measure, then the mean measures at any time $t \geq 0$, n_t^i, do too (Proposition 1.2. See also Prop 3.3 of Fournier et Méléard, 2004 for a proof in a case with no growth term). They then perform a change of variables between the time and their unidimensional trait to obtain an integro-differential system of which the densities n_t^i are weak solutions (Equation (1.12)). Finally, they use the form $n_t^i(x) = \mathrm{e}^{\lambda_i t} N_i(x)$ in the differential equation and are able to solve it to find an explicit expression for N_i.

First limitation

The fact that the time and the trait are bijectively linked in Tomasevic et al., 2022 is a crucial technical element that allows the authors to both prove the preservation of the density with respect to Lebesgue measure, and obtain the differential system weakly solved by the densities n_t^i. In our case however, we consider a multi-dimensionnal trait, which prevents such a bijection to exist as explained in Tran, 2008, Remark 3.5. Moreover, the proof of the preservation of a density with respect to Lebesgue measure in Fournier et Méléard, 2004 relies on the fact that the reproduction kernel admits a density with respect to Lebesgue measure. In our case, our mutation and reproduction kernels contain atoms, which prevents us from applying the same strategy of proof. We thus did not prove that the property of density with respect to Lebesgue measure was preserved, and were not able to derive explicitly the stationary profile of the mean measure. However, there is still hope that studying the integro-differential equations themselves (even without having an explicit solution) will bring some elements of understanding of the asymptotic behaviour of their solutions.

The associated spectral problem

We now focus on the growth rate λ_i. This rate can be found as the dominant eigenvalue of the adjoint problem associated to the integro-differential equation satisfied by n_t^i (Proposition 4.1 of Tomasevic et al., 2022). Equivalently, it is the dominant eigenvalue of the operator $\mathcal{L}$ satisfying, for any $t \geq 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\langle \mathcal{Z}_t, f\rangle] = \mathbb{E}[\langle \mathcal{Z}_t, \mathcal{L}f\rangle]. \tag{3.15}$$

Using the martingale property (Prop. 3) gives us that, for any non-negative measurable function $\varphi : \mathcal{X} \to \mathbb{R}$, the expression for $\mathcal{L}\varphi$ is the following :

$$\mathcal{L}\varphi(R,y) = (1 - d(y) - \mu)\varphi(R,y) + \frac{\mu}{2}\sum_{j=1}^{2}\varphi(R, g_{j,\varepsilon}(y)) + R\int_0^1 \Big(\frac{r}{4}\sum_{k=1}^{3}\varphi(R, f_{k,\theta}(y)) - \frac{3}{4}r\varphi(R,y)\Big)d\theta.$$

The functions $g_{j,\varepsilon}$ and $f_{k,\theta}$ were introduced after equ.(3.5). Recall that, in the case where the sub-populations of recombining and non-recombining individuals do not interact, it is possible to write the previous equations separately for each subpopulation by seeing R as a parameter rather than as a trait. In this case, the expressions are simplified.

The first part of the operator $((1-d(.)-\mu)\varphi(.))$ is a multiplication operator whose spectrum is easy to derive. However, the change of variables brought by mutation and recombination (visible through the functions $g_{j,\varepsilon}$ and $f_{k,\theta}$) renders the spectrum of the second part of the operator more difficult to analyse. An idea would be to see whether this second part is a compact operator, in which case the Fredholm alternative can be used to pursue the search for eigenvalues.

Second limitation

Unfortunately, we could not find the dominant eigenvalue and corresponding eigenvector for the adjoint operator, which prevented us from pushing this analysis further.

Verifying the convergence theorem assumptions

Assuming that we dispose of eigenelements for the adjoint operator $\mathcal{L}$, *i.e.* $\lambda \in \mathbb{R}_+$ and φ a measurable function from $\mathcal{X}$ to $\mathbb{R}$ such that $\mathcal{L}\varphi = \lambda\varphi$, we looked at whether the assumptions of the convergence theorem of Bansaye et al., 2022 were satisfied in our case.

Assumption A.2 is verified for any $\tau > 0$ with $\psi = \varphi$ and $\beta = \mathrm{e}^{\lambda\tau}$. Indeed, we have, for all $t > 0$,

$$\frac{\mathrm{d}}{\mathrm{d}t} M_t\varphi = M_t\mathcal{L}\varphi = \lambda M_t\varphi,$$

which leads to $M_t\varphi = \mathrm{e}^{\lambda t}\varphi$, and therefore $M_\tau\varphi \geq \mathrm{e}^{\lambda\tau}\varphi$.

Assumption A.4 is also easy to satisfy, setting $d = 1$. Using the previous equation, we have, for any probability measure ν,

$$\langle \nu, \frac{M_{n\tau}\varphi}{\varphi} \rangle = \langle \nu, \frac{\mathrm{e}^{\lambda n\tau}\varphi}{\varphi} \rangle = \langle \nu, \mathrm{e}^{\lambda n\tau} \rangle = \mathrm{e}^{\lambda n\tau}\nu(\mathcal{X}) = \mathrm{e}^{\lambda n\tau} = \sup_{z\in\mathcal{X}} \frac{M_{n\tau}\varphi(z)}{\varphi(z)}.$$

The two remaining points (A.1 and A.3) are more difficult to check as we do not have access to explicit expressions for the eigenelements, which was key to derive the results in Tomasevic et al., 2022.

Third limitation

Apart from the technical difficulties that can arise from the complexity of our process, it is possible that Assumption A.3 cannot be satisfied in our case. Indeed, Assumption A.3 can be interpretated as a mixing property, similar to irreducibility and aperiodicity for a Markov chain on a discrete state space (see the detailed interpretation of those assumptions in Bansaye et al., 2022). However, in our case, for non-recombining individuals, mutation events increase the values of the individual trait in $[0,1]^3$, without any possibility to decrease the value afterwards. Some area of the trait space $[0,1]^3$ may thus be unreachable after a certain time, impairing the possibility of "mixing".

The limitations listed above thus prevented us to conclude with these powerful mathematical tools. In the next section, we turn to a final simpler point of view to derive analytical results, and use individual-based simulations to complement our findings.

4 Unitype branching process point of view

Because of the limitations described above, we were unable to conclude on the persistence and long-term composition of populations of recombining and non-recombining individuals with the mathematical tools detailed in the previous section. Apart from the technical difficulties that we could not overcome for the moment, a limitation intrinsic to the model dynamics appeared : because individuals can only reproduce by selfing, the homozygosity of an offspring (*i.e.* the last coordinate of an individual trait in $\mathcal{X}$) can only be higher or equal to the homozygosity of its parent. In addition, mutation events also increase the proportion of sites homozygous for deleterious mutations. The process may thus lack the *local mixing* property necessary to apply the convergence result (Assumption A.3), because once the population minimum deleterious homozygosity exceeds a certain threshold, individuals with lower level of homozygosity than this threshold cannot be produced anymore. This forms a kind of Muller's ratchet in continuous state (Haigh, 1978), in the sense that genotypes with a low mutation load can be permanently lost.

The occurrence of the ratchet will therefore depend on the ability of subpopulations of individuals with low levels of homozygosity to self-maintain. More precisely, consider a region $B \subset \mathcal{X}$ of the trait space. Individuals with trait in B contribute to the subpopulation of this region by reproducing clonally. However, they also die, and are removed from the global population, or mutate or reproduce with recombination, and contribute to populations with traits outside of B (in particular, traits with higher homozygosity). If the rate at which individuals contribute to the local subpopulation of individuals with trait in B is higher than the rate at which they are removed or contribute to subpopulations outside of B, then the subpopulation of individuals with trait in B is said to be *locally maintained*. This subpopulation then acts as a source for the global population.

In this section, instead of looking at the global dynamics of the population, we therefore focus on the capacity of individuals of a fixed genotype to produce individuals of the same genotype. In other words, we are interested in the *local persistence* of the population. To study this question, we adopt a final point of view relying on a unitype branching process. In section 4.1, we describe this new process, and compute quantities of interest. Then, we analyze these quantities, and confront them with the results of individual-based simulations (section 4.4).

4.1 Unitype branching processes

Let $x_0 := (R_0, y_0)$ be a fixed trait in $\mathcal{X}$. We want to study the persistence of the subpopulations of individuals with trait x_0. We follow the number of individuals of trait x_0 with a unitype continuous-time branching process $(U_t^{x_0})_{t\geq 0}$, branching at rate $1 + \mu + d(y_0)$, with reproduction law :

$$p_0 = \frac{\mu + d(y_0)}{1 + \mu + d(y_0)}, \quad p_1 = \frac{R_0 \frac{3}{4} r}{1 + \mu + d(y_0)}, \quad p_2 = \frac{1 - R_0 \frac{3}{4} r}{1 + \mu + d(y_0)}, \quad p_k = 0 \text{ for } k \geq 3, \tag{3.16}$$

where p_k is the probability of having k offspring. This reproduction law is a direct consequence of the dynamics introduced above : an individual with trait x_0 is replaced by zero offspring in the subpopulation of individuals with trait x_0 when it dies (at rate $d(y_0)$), or mutates (at rate μ). An individual with trait x_0 is replaced by one offspring in the subpopulation of individuals with trait x_0 when it reproduces (at rate 1) and its offspring does not have the same genotype. This occurs with probability $R_0 \frac{3}{4} r$, when a recombination event renders reproduction non-clonal. The offspring is then not counted in the population of individuals with trait x_0, but the parent, who survives, is. An individual with trait x_0 is replaced by two offspring in the subpopulation of individuals with trait x_0 when it reproduces (still at rate 1), and its reproduction is clonal, which occurs with probability $1 - R_0 + R_0(1 - \frac{3}{4} r) = 1 - R_0 \frac{3}{4} r$. The reproduction event then adds an extra individual to the population of individuals with trait x_0.

4.2 Quantities of interest

Using this new branching process, we explore the conditions for *local persistence*, in the sense defined in the introduction of section 4.1. The idea is simple : if the process is subcritical, there is local extinction (be it because all individuals end up having a different trait than x_0, or because the population as a whole goes extinct). If the process is supercritical, then the population can be maintained locally with positive probability. We use Haccou et al., 2005 or Athreya et Ney, 1972 to derive the quantities of interest.

The mean of the reproduction law is

$$m(x_0) = \frac{2 - R_0 \frac{3}{4} r}{1 + \mu + d(y_0)}. \tag{3.17}$$

Immediately, we see that the process is supercritical if and only if

$$1 > \mu + d(y_0) + R_0 \frac{3}{4} r, \tag{3.18}$$

which can be rewritten

$$1 - R_0 \frac{3}{4} r > \mu + d(y_0),$$

that is, the rate at which the number of individuals with trait x_0 increases by 1 is greater than the rate at which this number decreases by 1.

We now look for the probability of extinction. Classical results show that it is the smallest root of $f(h)-h$ in $[0,1]$, where f is the generating function of the reproduction law.
We have, for $h\in[0,1]$,

$$f(h)=\frac{\mu+d(y_0)}{1+\mu+d(y_0)}+\frac{R_0\frac{3}{4}r}{1+\mu+d(y_0)}h+\frac{1-R_0\frac{3}{4}r}{1+\mu+d(y_0)}h^2.$$

The roots of $f(h)-h$ are then

$$\frac{\mu+d(y_0)}{1-R_0\frac{3}{4}r}\quad\text{and } 1.$$

In the supercritical case, *i.e.* in the case where $1>\mu+d(y_0)+R_0\frac{3}{4}r$, the first root is smaller than 1, which gives

$$\mathbb{P}_{ext}(x_0):=\frac{\mu+d(y_0)}{1-R_0\frac{3}{4}r}. \tag{3.19}$$

We now turn to the extinction time in the subcritical case. We introduce the generating function of the process at time t : for $h\in[0,1]$,

$$F(h,t):=\mathbb{E}[h^{U_t}]=\sum_{k=0}^{\infty}\mathbb{P}(U_t=k)h^k.$$

We have removed the dependence in x_0 of U_t to alleviate the notation. The Kolmogorov backward equation gives

$$\frac{\partial}{\partial t}F(h,t)=u(F(h,t)),$$

with $u(h)=(1+\mu+d(y_0))(f(h)-h)$. $F(h,.)$ is then solution to the Cauchy problem

$$\begin{cases}g'(t)=\mu+d(y_0)+\Big(R_0\frac{3}{4}r-1-\mu-d(y_0)\Big)f(t)+\Big(1-R_0\frac{3}{4}r\Big)f(t)^2\\ g(0)=h.\end{cases} \tag{3.20}$$

This equation with boundary condition can be solved, and the solution is, for $s\in[0,1[$,

$$F(s,t)=1+\frac{1-\mu-d(y_0)-R_0\frac{3}{4}r}{R_0\frac{3}{4}r-1+\Big(\frac{1-\mu-d(y_0)-R_0\frac{3}{4}r}{s-1}-R_0\frac{3}{4}r+1\Big)\,\mathrm{e}^{-(1-\mu-d(y_0)-R_0\frac{3}{4}r)t}}.$$

If we write $T:=\inf\{t\geq 0, U_t=0\}$ for the extinction time of the branching process, we have

$$\mathbb{P}(T>t)=1-F(0,t)=\frac{\mu+d(y_0)+R_0\frac{3}{4}r-1}{R_0\frac{3}{4}r-1+(\mu+d(y_0))\,\mathrm{e}^{-(1-\mu-d(y_0)-R_0\frac{3}{4}r)t}}. \tag{3.21}$$

We then compute the expected extinction time thanks to the formula

$$\mathbb{E}[T]=\int_0^{+\infty}\mathbb{P}(T>t)\mathrm{d}t,$$

which gives, observing that the ratio in equ. (3.21) is of the form u'/u,

$$\begin{aligned}\mathbb{E}[T]&=\int_0^t\frac{-\big(1-\mu-d(y_0)-R_0\frac{3}{4}r\big)\mathrm{e}^{(1-\mu-d(y_0)-R_0\frac{3}{4}r)t}}{\Big(R_0\frac{3}{4}r-1\Big)\mathrm{e}^{(1-\mu-d(y_0)-R_0\frac{3}{4}r)t}+\mu+d(y_0)}\mathrm{d}t\\&=-\frac{1}{R_0\frac{3}{4}r-1}\int_0^t\frac{\big(R_0\frac{3}{4}r-1\big)\big(1-\mu-d(y_0)-R\frac{3}{4}r\big)\mathrm{e}^{(1-\mu-d(y_0)-R_0\frac{3}{4}r)t}}{\Big(R_0\frac{3}{4}r-1\Big)\mathrm{e}^{(1-\mu-d(y_0)-R_0\frac{3}{4}r)t}+\mu+d(y_0)}\mathrm{d}t\\&=-\frac{1}{R_0\frac{3}{4}r-1}\left[\ln\left(\Big(R_0\frac{3}{4}r-1\Big)\mathrm{e}^{(1-\mu-d(y_0)-R_0\frac{3}{4}r)t}+\mu+d(y_0)\right)\right]_0^{\infty}.\end{aligned}$$

Finally, we obtain

$$\mathbb{E}[T]=-\frac{1}{1-R_0\frac{3}{4}r}\ln\left(1-\frac{1-R_0\frac{3}{4}r}{\mu+d(y_0)}\right). \tag{3.22}$$

Figure 3.3 compares the local process criticality for recombining and non-recombining populations ($R_0=0$

versus $R_0 = 1$). As criticality depends on the local genotype only via the death rate (condition (3.18)), which is a function of the homozygosity ω and the heterozygosity $x_{het} = x_a + x_b - 2\omega$, we illustrate the supercriticality condition on the space $\{(\omega, x_{het}) \in [0,1]^2, \text{ s.t. } \omega + x_{het} \leq 1\}$. For each pair (x_{het}, ω) in this space, the value of $\mu + R\frac{3}{4}r + d(x_{het}, \omega)$ is computed, with $d(x_{het}, \omega) = s_0 + s_{hom}\omega + s_{het}x_{het}$. We set $s_{hom} = s$ and $s_{het} = hs$ for a given $s \in [0,1]$, that stands for the selection coefficient, and $h \in [0, 1/2]$ that stands for the dominance coefficient. If $\mu + R\frac{3}{4}r + d(x_{het}, \omega)$ is smaller than 1, *i.e.* if condition (3.18) is met, the local process is supercritical. The extinction probability $\mathbb{P}_{ext} = \frac{\mu + d(x_{het},\omega)}{1 - R\frac{3}{4}r}$ is then showed (purple heatmap). If condition (3.18) is not met, the local process is subcritical, and the mean extinction time $\mathbb{E}[T] = -\frac{1}{1-R\frac{3}{4}r} \ln\left(1 - \frac{1-R\frac{3}{4}r}{\mu + d(x_{het},\omega)}\right)$ is showed (golden heatmap).

Condition (3.18) is trivially not met if μ or s_0 is large. In order to avoid these trivial cases and assess the effect of recombination on population maintenance, our parameters were chosen so that $\frac{\mu}{s}$, $\frac{1-s_0}{s}$ and $\frac{r}{s}$ are of the same order of magnitude as ω and hx_{het}. In that case, all the terms are of the same order of magnitude in the rewritten condition (3.18) :

$$\omega + hx_{het} + R\frac{3}{4}\frac{r}{s} < \frac{1 - s_0}{s} - \frac{\mu}{s}.$$

4.3 Individual-based simulations

We performed individual-based simulations to study the evolution of populations of non-recombining and recombining individuals that followed the dynamics described in equ.(3.5), and confront the results obtained with the unitype branching process point of view described in the previous section. The idea is to follow the algorithm described in section 2.2, with an acceptance-rejection method. In particular, our populations are not limited in size.

We used the neatly implemented and very effective *IBMPopSim* package (Giorgi et al., 2023), that provides an interface in *R* but uses the computation efficiency of *C++* to run the simulations. Codes used are avalaible at https://github.com/EmilieTezenas/Accum_Mut_Del.git.

We simulated separately recombining and non-recombining individuals, using the same parameters as in Figure 3.3. We simulated the process evolution on the longest time-scale we could with the computation power that was available. With the chosen parameters, the population size increases exponentially, and eventually exceeds the core capacity. The results are presented Figure 3.4.

4.4 Results

We analyze the results given by the unitype-branching processes point of view of section 4.1 and by the individual-based simulations through three problematics. The first two are the ones we tried to answer with the mathematical exploration of section 3. The last comes back to the question of recombination suppression maintenance.

Is there a difference in growth rates between recombining and non-recombining populations ?

Indiviudual-based simulations (Figure 3.4) show that the total population sizes are quite similar for populations of recombining and non-recombining individuals. Moreover, the exponential growth rate computed with a linear regression of the log-size with respect to time is the same for the recombining and non-recombining populations. This suggests that the growth rate is the same for the two sub-populations, even if figure 3.4 only displays one trajectory for each recombination case (the similarity in growth rates seemed to be consistent over several independent runs). We did not, however, perform enough simulation experiments to test whether this results holds. Moreover, as our simulations are limited in time, there can be a decrease in the growth rate of one or both populations in the long run.

Is there a difference in asymptotic mutation loads between recombining and non-recombining individuals ?

Figure 3.4 displays the trait distributions up to 180 units of time. These distributions may still vary in longer times, but there already are significant differences between populations of recombining and non-recombining

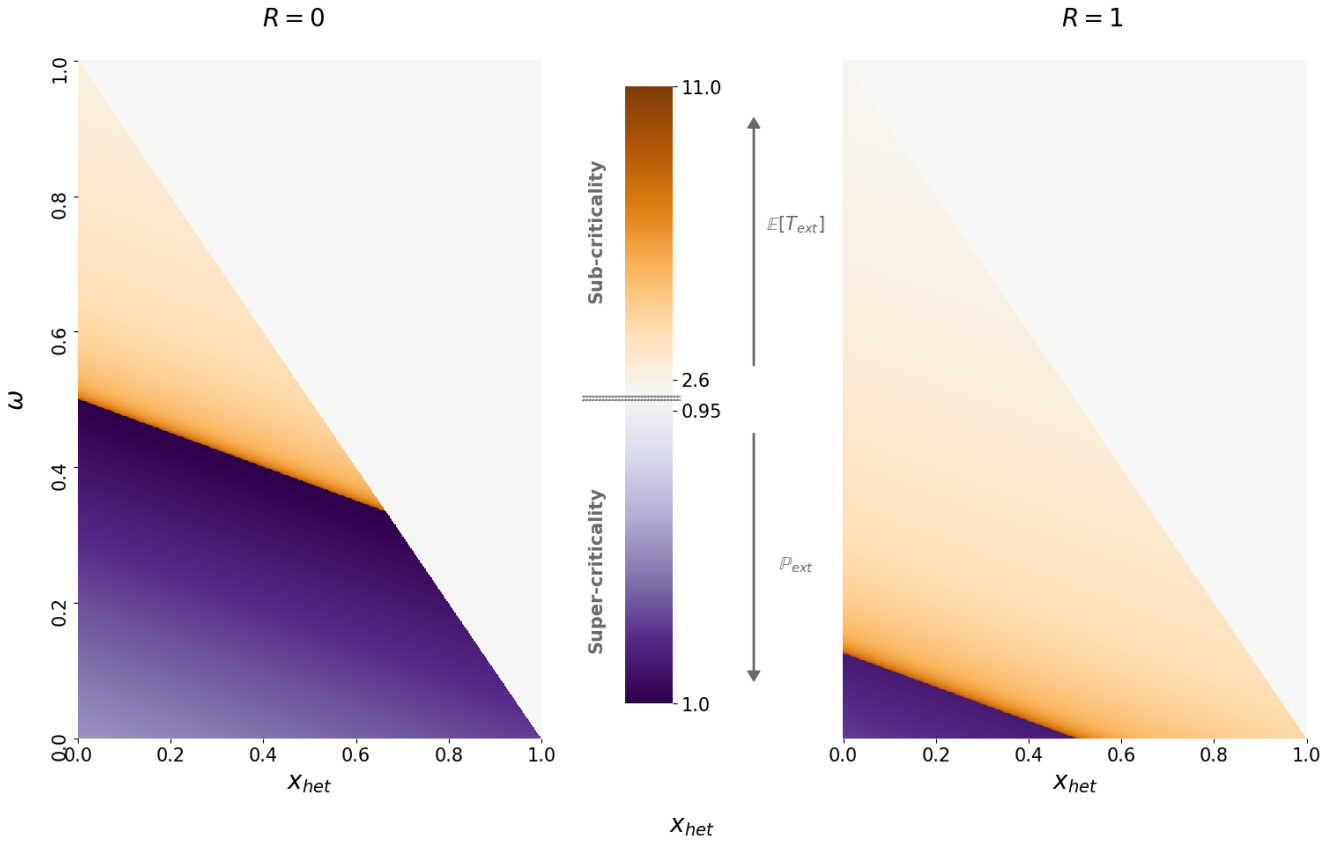


FIGURE 3.3 – Illustration of the super- or sub-criticality of a local process, with varying values for the heterozygosity level (x_{het}, x-axis) and for the homozygosity level (ω, y-axis). For each point (x_{het}, ω) such that $\omega + x_{het} \leq 1$, the value of the threshold for supercriticality given in (3.18) is computed. If the condition for supercriticality is met, the extinction probability is computed and its value is represented by a purple color at (x_{het}, ω). If the process is locally subcritical, then the value of the mean extinction time is computed and represented by a golden color at (x_{het}, ω). The left panel shows the non-recombining case ($R = 0$), and the right panel shows the recombining case ($R = 1$). The maximum and minimum values obtained for the mean extinction time (gold) and the extinction probability (purple) are indicated alongside the corresponding color on the colorbar, in the central part of the figure. The parameters used here are $\mu = 0.05$, $r = 0.05$, $s0 = 0.9$, $s_{hom} = 0.1$, $h = 0.25$ and thus $s_{het} = hs_{hom} = 0.025$.

individuals.

The population of recombining individuals is more widely distributed over the trait space than the population of non-recombining individuals. This is expected, as recombination shuffles alleles. In particular, an offspring produced after a reproduction event with recombination can have a genotype with lower level of heterozygosity than its parent. On the contrary, the heterozygosity of non-recombining individuals can only vary through mutation events, which increases both heterozygosity and homozygosity. The shape of the location of the highest density among the non-recombining population (left column) suggests that heterozygosity increases first, up to a maximum (along the $\omega = 1 - x_{het}$ border), which leads to the subsequent increase in homozygosity. Note however that the speed of heterozygosity increase and the shape of the trajectory heavily depend on the value chosen for ε, which quantifies the added load in a mutation event. Here, $\varepsilon = 0.1$, which is high. In particular, we see that the local-mixing assumption necessary to apply the semi-group convergence theorem on the process mean measure (section 3.2) is clearly not verified in the non-recombining case.

In both cases, the fact that we observe densely occupied regions suggests that the (normalised) distribution of traits may converge towards a stationary distribution. In particular, for the recombining population (right bottom panel), the increasing concentration of individuals shown by the red zone suggests that this area is somehow optimal : the rate at which individuals with these traits are created (via reproduction without recombination or contribution from less charged genotypes) is higher than the rate at which individuals with these traits are removed from the area (through death, mutation, or reproduction with recombination).

In conclusion, both recombining and non-recombining populations show a global increase in mutation load, but a stationary distribution around optimal traits may exist.

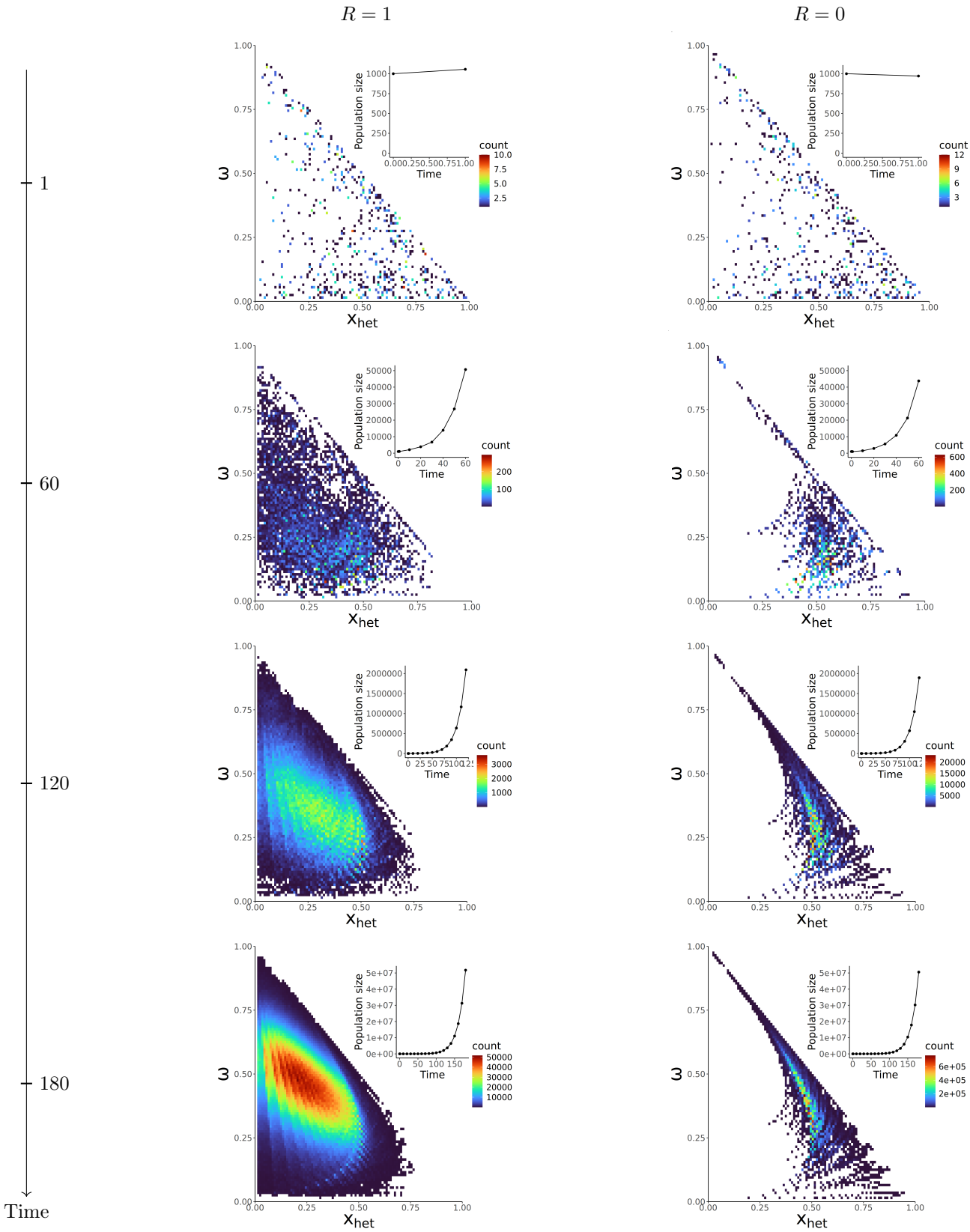


FIGURE 3.4 – Trait distribution of a trajectory of non-recombining (left column) and recombining (right column) individuals on the trait space $\{(x_{het}, \omega) \in [0,1]^2, x_{het} + \omega \leq 1\}$ at times 1, 60, 120, and 180 (one time for each row). Both trajectories (recombining or not) were initiated with 1000 individuals, with traits x_a and x_b picked uniformly at random in $[0,1]$ and ω picked uniformly at random in $[\max(0, x_a + x_b - 1), \min(x_a, x_b)]$. The number of individuals in each point of the space (x_{het}, ω) is displayed with the heatmap. An additionnal graph shows the evolution of the total population size up to the time at which the heatmap is plotted. The estimated growth rate of the population, computed with a simple regression, is 0.062 in both cases.

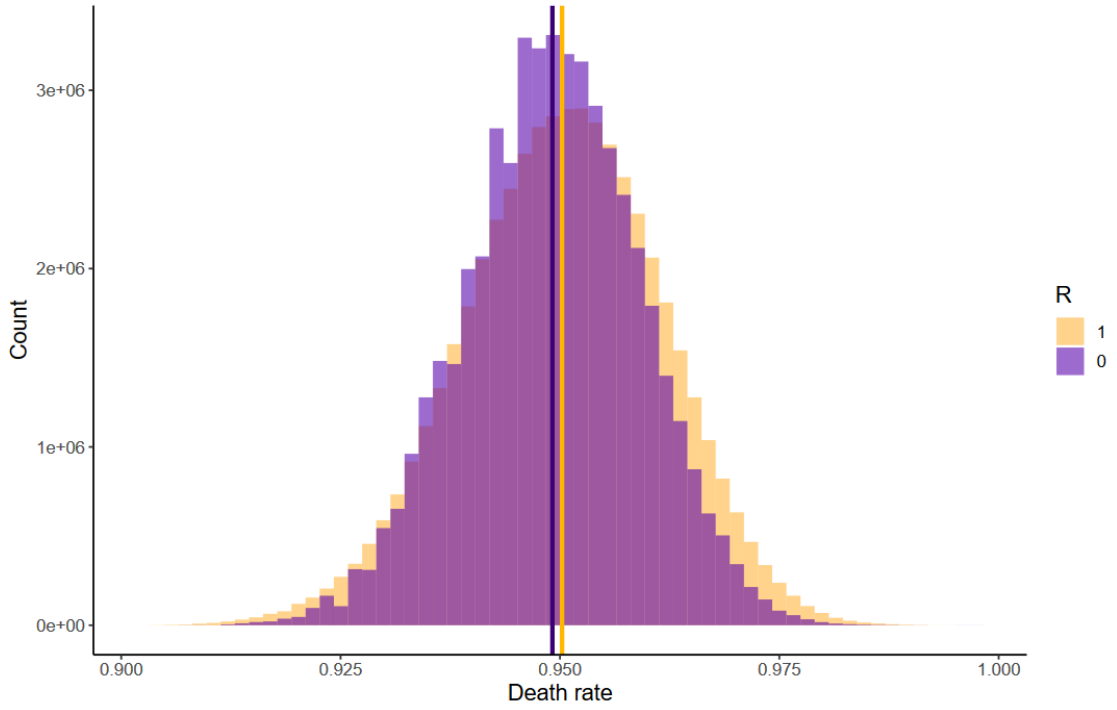


FIGURE 3.5 – Distribution of the death rates of recombining (gold) and non-recombining (purple) individuals at time 180. Vertical lines show the mean of each ditribution.

Can suppressed recombination be maintained in the long term ?

Figure 3.3 shows that the region in which a local unitype branching process is supercritical, *i.e.* where the local population can self-maintain (purple region), is significantly larger when recombination is not allowed (left panel). This suggests that recombination suppression can be maintained under complete selfing : the area of traits that can be at the origin of an exponentially large population is wider in the non-recombining case. Moreover, for a fixed genotype, suppressing recombination may change the criticality of the associated branching process from subcritical to supercritical (from the golden region in the recombining case to the purple region in the non-recombining case). In particular, in the suppressed recombination case, the local process is still supercritical for high levels of heterozygosity, as long as homozygosity is not too high (the region is purple for x_{het} between 0.7 and 1, and ω between 0 and 0.3 approximately). This suggests that individuals with a high density of mutations can be maintained in the population when recombination is suppressed, as long as they carry mostly heterozygous mutations. This suppports previous results where non-recombining fragments that fixed in a populations had an initial heterozygote advantage (JAY et al., 2022b).

Apart from the criticality property in each location, the extinction probability (equ.(3.19)) is higher for recombining individuals than non-recombining individuals, and the mean extinction time (equ.(3.22)) is shorter for recombining individuals than for non-recombining individuals. This again shows an enhanced possibility for local persistence of the population in the non-recombining case. Note that the extinction probability is high even in regions where the branching process is supercritical, for both recombining and non-recombining individuals (at least 0.95 with the parameters of Figure 3.3). This means that a population initiated with one individual with a certain trait has only 5% chance not to go extinct. This may explain why such traits are not present after a certain time in the population in the individual-based simulations (Figure 3.4).

Moreover, the fact that the population size of both recombining and non-recombining individuals increases with the same rate (Figure 3.4) suggests that suppressed recombination can be maintained over a long time-scale. In particular, the non-recombining population does not go extinct, despite the accumulation of deleterious mutations in a large subpopulation. Moreover, plotting the distribution of the death rate in each population shows that they are quite similar (Figure 3.5) which suggests that neither characteristic ($R = 0$ or $R = 1$) would be disfavored against the other.

5 Discussion and perspective

This model aims at studying the differences in dynamics of a recombining versus non-recombining population for individuals that carry a permanently heterozygous mating-type locus and can accumulate recessive deleterious mutations. The selfing assumption isolates one sub-population from the other.

The results obtained with the simple unitype branching process point of view suggest that, in a parameter range, there can be local persistence of the population in some regions of the trait space, and not in others. This typically occurs when the sum of the mutation rate and the death rate is lower than the birth rate. Subpopulations of individuals with traits that allow self-maintenance then act as source for the global population, contributing to subpopulations with higher proportions of deleterious mutations. Two behaviors may then be observed : either the global population goes extinct, or a stationary distribution around an optimal trait is reached. The first scenario may occur if the individual traits keep increasing as in Muller's ratchet. At some point, the individual death rates may be too high for the global population to persist. The second scenario can occur when an equilibrium is reached between the possible local persistence in some region of the trait space and the increase in the values of the trait. The individual-based simulations run and presented in this manuscript suggest that for the parameters used, a stationary distribution may be reached. In particular, this suggests that suppressed recombination can be maintained over long periods of time, despite the accumulation of deleterious mutations. This idea is supported by the similarity between the death rate distributions, that represents the fitness of individuals, in the two sub-populations.

To summarize, non-recombining individuals are not expected to be disfavored compared to recombining individuals in this particular framework of purely selfing individuals with a permanently heterozygous mating-type.

The assumption of selfing can however be relaxed to model occasional events of outcrossing. Outcrossing events can have a significant impact on a population dynamics, as an outcrossed offspring can have lower homozygosity than its parents. The enhanced local population persistence for non-recombining populations compared to recombining populations can then be lost if occasionnal outcrossing events occur.

Apart from the study of growth rates and limiting distributions for the model presented in this chapter, other theoretical investigations could be conducted to better understand the dynamics of recombination suppression in selfing populations.

First, it could be interesting to deepen the exploration through the spinal point of view described in section 3.1. Indeed, typical trajectories for the evolution of the proportion of deleterious mutations seem to be visible in the results of the individual-based simulations, especially for non-recombining populations, in which heterozygosity increases first, eventually leading to increased homozygosity. The speed of accumulation of mutations in a heterozygous or homozygous state should depend on the mutation and reproduction parameters.

Second, the exponential growth of populations of recombining and non-recombining individuals invites us to look at a limit in large population sizes. A large body of work is available in this field (Fournier et Méléard, 2004, Champagnat et Lambert, 2007 and derived works), and this could give insights on the relative impact of the mutation and recombination parameters on the trait distribution and population growth.

Third, the occurrence of densely occupied regions in the trait space invites us to explore the existence of stationary, or quasi-stationary distributions (see Champagnat et Villemonais, 2016 and derived works).

These mathematical tools could help disentangle the effect of the various parameters considered here, but also tackle the important question of the time scales on which each mechanism acts, which we do not consider here.

Notation

Notation	*Description*
$R \in \{0,1\}$	First coordinate of an individual trait ; indicate whether recombination can occur during meiosis ($R=1$) or not ($R=0$)
$x_a,\, x_b \in [0,1]$	Second and third coordinate of an individual trait ; indicate the proportion of deleterious mutations carried by each mating-type chromosome
$\omega \in [0,1]$	Last coordinate of an individual trait ; indicate the proportion of sites homozygous for deleterious mutations
$\mathcal{X} = \{0,1\} \times [0,1]^3$	Trait space
$E = [0,1]^3$	Genotype space
$\mathcal{M}_F(\mathcal{X})$	Set of all finite counting measures on $\mathcal{X}$
$(\mathcal{Z}_t)_{t\geq 0}$	Birth and death process that describes individuals alive at time t
r	Recombination probability, for individuals that can recombine ($R=1$)
$\theta \in [0,1]$	Variable that quantifies the position of a crossing-over during a recombination event
μ	Mutation rate
ε	Quantification of the mutation load added to a chromosome during a mutation event
$d(y)$	Death function (defined in equ.(3.1))
s_0	Intrinsic selection coefficient
$s_{hom}, s_{het} \in [0,1]$	Selection coefficient for sites homozygous (resp. heterozygous) for deleterious mutations
$s,\, h \in [0,1]$	Selection and dominance coefficients used in section (4)
$N_t = \langle \mathcal{Z}_t, 1 \rangle$	Population size at time t
$(U_t^{x_0})_{t\geq 0}$	Unitype continuous-time branching process approximating the numbers of individuals of trait x_0 in section (4.1)

TABLE 3.1 – Table of notation

Discussion Générale

Les résultats présentés dans cette thèse contribuent à la compréhension de l'extension de la suppression de recombinaison sur les chromosomes sexuels et de type sexuels, en montrant que les mutations délétères peuvent jouer un rôle important à plusieurs étapes de cette évolution.

1 - Accumulation de mutations délétères au voisinage d'un locus en permanence hétérozygote. Le chapitre 2 de cette thèse montre que la présence d'un locus en permanence hétérozygote impacte la dynamique des mutations délétères dans son voisinage, surtout sous auto-fécondation. La purge des mutations délétères proches d'un locus de type sexuel est ralentie, ce qui peut permettre leur accumulation.

2 - Sélection pour la suppression de recombinaison. Le chapitre 1 montre qu'un *haplotype* sur lequel apparaît une inversion peut être favorisé lorsqu'il porte moins de mutations délétères que la moyenne dans la population. Cela a d'autant plus de chances de se produire qu'un grand nombre de mutations délétères sont présentes dans le génome à l'échelle de la population, créant ainsi une forte variabilité dans la valeur sélective d'un fragment d'ADN de longueur donnée. Le fragment non-recombinant possède alors un avantage intrinsèque, et peut augmenter en fréquence dans la population.

3 - Fixation de la suppression de recombinaison. Lorsqu'une inversion favorisée par son faible nombre de mutations délétères augmente en fréquence dans la population, elle peut se retrouver à l'état homozygote. Son fardeau génétique est alors exposé, et sa valeur sélective diminue, ce qui bloque son augmentation en fréquence. Cependant, si l'inversion englobe le locus en permanence hétérozygote, alors elle est maintenue à l'état hétérozygote, tout comme les mutations délétères qu'elle porte : ces mutations sont "abritées". Le fardeau génétique de l'inversion est peu exprimé, et celle-ci continue alors à être avantagée et à augmenter en fréquence dans la population. Les simulations du chapitre 1 montrent qu'une inversion qui englobe l'allèle en permanence hétérozygote peut se fixer dans la population des chromosomes portant cet allèle.

4 - Maintien de la suppression de recombinaison en temps long. Une fois qu'une inversion (ou un suppresseur de recombinaison) est fixée dans la population, la question de son maintien sur le long terme se pose. En effet, sans recombinaison, les mutations délétères s'accumulent, pouvant mener à la dégénérescence de la zone concernée et à la baisse de valeur sélective des individus portant un chromosome non-recombinant. Dans ce contexte, un rétablissement de la recombinaison pourrait être sélectionné. Cependant, les conditions favorisant et permettant une restoration de la recombinaison ont été peu explorées.

Le chapitre 3 apporte un début de réponse concernant la baisse de valeur sélective des individus portant un chromosome non-recombinant. Pour cela, on s'est intéressés à la différence de vitesse de croissance exponentielle de la taille de population et à la distribution de valeur sélective dans des populations d'individus recombinant ou non le long d'un chromosome de type sexuel. Lorsque les individus se reproduisent exclusivement par auto-fécondation intra-tétrade, il semblerait que l'accumulation de mutations délétères sur une zone sans recombinaison ne crée pas de réduction particulière de la valeur sélective dans un premier temps. L'effet d'abritement d'un locus de type sexuel en permanence hétérozygote semble persister : en présence d'un tel locus, les mutations délétères s'accumulent d'abord à l'état hétérozygote, ce qui a un impact limité sur la diminution de la valeur sélective. Les résultats analytiques obtenus dans ce chapitre restent cependant très préliminaires, et les simulations individu-centrées sont limitées dans le temps. Par ailleurs, nous ne prenons pas en compte de régulation de la taille de population due à des ressources limitées. Une étude plus approfondie est nécessaire afin de conclure sur le maintien de la suppression de recombinaison en temps long.

Cette question du maintien de la suppression de recombinaison sur un temps long sous l'hypothèse d'abritement d'allèles délétères est activement discutée (Lenormand et Roze, 2023). Dans un système XY, où seul le Y est en permanence hétérozygote, l'accumulation de mutations délétères suite à l'arrêt de recombinaison sur le Y pourrait mener à une baisse de la valeur sélective des mâles dans la population, et même, à terme, à l'extinction de la population. Dans ce contexte, une mutation qui rétablit la recombinaison pourrait être favorisée, si elle permet de créer à court terme des descendants avec moins de mutations délétères et donc une plus haute valeur sélective. La possibilité de ré-inversion (ou réversion) est prise en compte dans les simulations stochastiques du chapitre 1. On considère un nombre fini de points de rupture symbolisant des positions du génome auxquelles un évènement de recombinaison a plus de chance d'avoir lieu que la moyenne du reste du génome. Dans notre modèle, on considère qu'une inversion peut se former entre deux points de rupture. Si une inversion se forme entre les mêmes points de rupture qu'une précédente inversion, alors la zone est considérée réversée, et la recombinaison rétablie. La principale contrainte est que la recombinaison ne peut pas être immédiatement rétablie sur une zone si cette zone est recouverte par plusieurs inversions, en raison de la complexité des réarrangements chromosomiques créés. Cela implique que le taux de réversion effectif est plus faible que le taux d'apparition d'inversions. Le nombre de points de rupture impacte également la probabilité qu'une réversion se produise : plus le nombre de points de rupture est élevé, moins une inversion a de chances de se former exactement entre les mêmes points de rupture qu'une précédente inversion. Les résultats des simulations montrent que des réversions ont bien lieu, rétablissant la recombinaison sur de larges portions du chromosome Y. On observe néanmoins une extension progressive de la suppression de recombinaison lorsque le nombre de points de rupture n'est pas trop petit, c'est-à-dire lorsque que la probabilité d'apparition d'une réversion n'est pas trop forte. La principale question soulevée par cette étude est alors celle du taux effectif de réversion dans les populations réelles. Des études empiriques, peu disponibles à ce jour, pourraient permettre de mesurer ce taux de réversion et ainsi de déterminer si les réarrangements peuvent contribuer à maintenir à long terme les régions sans recombinaison.

Une autre possibilité que celle des contraintes liées aux réarrangements pour expliquer le maintien de la suppression de recombinaison sur le long terme est l'évolution de compensation de dosage (Lenormand et Roze, 2022). L'idée est de réguler l'expression des gènes portés par les chromosomes sexuels pour maintenir un niveau égal d'expression des gènes, malgré la dégénérescence du chromosome qui arrête de recombiner. Par exemple, dans un système XY (*i.e.* dans lequel les individus mâles portent un chromosome X et un chromosome Y, et les femelles deux chromosomes X), un gène porté par le X chez les mâles peut être exprimé deux fois pour compenser une baisse d'expression du même gène sur le Y. Ce mécanisme permet donc de limiter l'effet de l'accumulation de mutations délétères suite à l'arrêt de la recombinaison sur le chromosome Y, et ainsi de maintenir la valeur sélective des individus portant un fragment non recombinant. Il reste cependant à déterminer si un tel mécanisme peut se produire pour des chromosomes de type sexuel, donc pour lesquels la suppression de recombinaison sur les deux chromosomes réduit l'efficacité de la sélection (contrairement au système XY dans lequel seul le Y arrête de recombiner et donc pour lequel la sélection reste efficace sur le X pour évoluer une compensation de dosage).

Perspectives. Une première ouverture pour de futurs travaux se situe dans la considération des échelles temporelles. Biologiquement, les différentes échelles de temps sur lesquelles peuvent agir les mécanismes en jeu (apparition, sélection et fixation d'un suppresseur de recombinaison, accumulation de mutations délétères dans un fragment non-recombinant, apparition, sélection pour un rétablissement de la recombinaison, évolution de la compensation de dosage) semblent constituer un point clé pour comprendre l'évolution de la suppression de recombinaison sur les chromosomes de type sexuel. Mathématiquement, des techniques développées pour l'étude des processus markoviens de sauts sous une hypothèse de grande taille de population permettraient d'étudier ces différentes échelles de temps (Fournier et Méléard, 2004, Champagnat et Lambert, 2007). Par ailleurs, les résultats du chapitre 2 et des simulations du chapitre 1 montrent que la dynamique des mutations délétères, et la fixation d'une inversion, reposent sur des évènements rares. Mathématiquement, les méthodes de grandes déviations permettraient peut-être d'obtenir des résultats analytiques sur la probabilité d'extension de suppression de recombinaison (Varadhan, 2008).

En outre, les résultats des simulations du chapitre 3 pour des sous-populations non-recombinantes suggèrent l'existence de "trajectoires typiques" pour l'évolution de la charge mutationnelle des individus. Ainsi, l'utilisation de méthodes d'étude de processus de branchement via les trajectoires spinales semble justifiée (Marguet, 2019b, Marguet, 2019a). L'approfondissement de l'étude des conditions d'application du théorème de conver-

gence présenté au chapitre 3 (Bansaye et al., 2022) dans le cadre d'un trait multidimensionnel permettrait également d'obtenir une avancée dans l'étude de la dynamique d'un fragment non-recombinant avec accumulation de mutations délétères. Les simulations pour les deux sous-populations (d'individus recombinant ou non) suggèrent également la possible existence de distributions stationnaires ou quasi-stationnaires (Champagnat et Villemonais, 2016). Biologiquement, cela permettrait d'étudier les conditions (ou jeux de paramètres) sous lesquelles la suppression de recombinaison peut être maintenue malgré l'accumulation de mutations délétères.

Enfin, nous n'avons pas considéré d'épistasie dans les modèles stochastiques. Or, de précédents travaux décrits dans l'introduction (Stetsenko et Roze, 2022) montrent que l'épistasie peut influencer la sélection pour ou contre la recombinaison, et probablement son maintien. La prise en compte des interactions d'allèles sur la fitness peut donc également constituer une piste d'étude intéressante.

Bibliographie

Antonovics, J., & Abrams, J. Y. (2004). Intratetrad mating and the evolution of linkage relationships. *Evolution*, *58*(4), 702–709. https://doi.org/10.1111/j.0014-3820.2004.tb00403.x

Athreya, K. B., & Ney, P. E. (1972). *Branching processes*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-65371-1

Bachtrog, D. (2006). A dynamic view of sex chromosome evolution. *Current Opinion in Genetics & Development*, *16*(6), 578–585. https://doi.org/10.1016/j.gde.2006.10.007

Bansaye, V., Cloez, B., Gabriel, P., & Marguet, A. (2022). A non-conservative harris ergodic theorem. *Journal of the London Mathematical Society*, *106*(3), 2459–2510. https://doi.org/10.1112/jlms.12639

Barton, N. H., Etheridge, A. M., & Véber, A. (2013). Modelling evolution in a spatial continuum. *Journal of Statistical Mechanics: Theory and Experiment*, *2013*(1), P01002. https://doi.org/10.1088/1742-5468/2013/01/P01002

Barton, N. H., & Turelli, M. (1991). Natural and sexual selection on many loci. *Genetics*, *127*(1), 229–255. https://doi.org/10.1093/genetics/127.1.229

Barton, N., & Charlesworth, B. (1998). Why sex and recombination? *Science*, *281*(5385), 1986–1990.

Bazzicalupo, A. L., Carpentier, F., Otto, S. P., & Giraud, T. (2019). Little evidence of antagonistic selection in the evolutionary strata of fungal mating-type chromosomes (*Microbotryum lychnidis-dioicae*). *G3 Genes|Genomes|Genetics*, *9*(6), 1987–1998. https://doi.org/10.1534/g3.119.400242

Bergero, R., & Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends in Ecology & Evolution*, *24*(2), 94–102. https://doi.org/10.1016/j.tree.2008.09.010

Billiard, S., López-Villavicencio, M., Hood, M. E., & Giraud, T. (2012). Sex, outcrossing and mating types: unsolved questions in fungi and beyond: sexy fungi. *Journal of Evolutionary Biology*, *25*(6), 1020–1038. https://doi.org/10.1111/j.1420-9101.2012.02495.x

Champagnat, N., & Lambert, A. (2007). Evolution of discrete populations and the canonical diffusion of adaptive dynamics. *The Annals of Applied Probability*, *17*(1). https://doi.org/10.1214/105051606000000628

Champagnat, N., & Méléard, S. (2007). Invasion and adaptive evolution for individual-based spatially structured populations. *Journal of Mathematical Biology*, *55*(2), 147–188. https://doi.org/10.1007/s00285-007-0072-z

Champagnat, N., & Villemonais, D. (2016). Exponential convergence to quasi-stationary distribution and q-process. *Probability Theory and Related Fields*, *164*(1), 243–283. https://doi.org/10.1007/s00440-014-0611-7

Charlesworth, B., & Wall, J. D. (1999). Inbreeding, heterozygote advantage and the evolution of neo–x and neo–y sex chromosomes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1414), 51–56. https://doi.org/10.1098/rspb.1999.0603

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, *18*(1987), 237–268.

Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, *95*(2), 118–128. https://doi.org/10.1038/sj.hdy.6800697

Coelho, S. M., Gueno, J., Lipinska, A. P., Cock, J. M., & Umen, J. G. (2018). UV chromosomes and haploid sexual systems. *Trends in Plant Science*, *23*(9), 794–807. https://doi.org/10.1016/j.tplants.2018.06.005

Collet, P., Méléard, S., & Metz, J. A. J. (2013). A rigorous model study of the adaptive dynamics of mendelian diploids. *Journal of Mathematical Biology*, *67*(3), 569–607. https://doi.org/10.1007/s00285-012-0562-5

Coron, C. (2016). Slow-fast stochastic diffusion dynamics and quasi-stationarity for diploid populations with varying size. *Journal of Mathematical Biology*, *72*(1), 171–202. https://doi.org/10.1007/s00285-015-0878-z

Crow, J. F., & Kimura, M. (2010). *An introduction to population genetics theory*. The Blackburn Press.

Durrett, R. (2008). *Probability models for DNA sequence evolution*. Springer New York. https://doi.org/10.1007/978-0-387-78168-6

Etheridge, A. (2011). *Some mathematical models from population genetics: école d'été de probabilités de saint-flour XXXIX-2009* (Vol. 2012). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-16632-7

Etheridge, A., Freeman, N., & Penington, S. (2017). Branching brownian motion, mean curvature flow and the motion of hybrid zones. *Electronic Journal of Probability*, *22*. https://doi.org/10.1214/17-EJP127

Fournier, N., & Méléard, S. (2004). A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *The Annals of Applied Probability*, *14*(4). https://doi.org/10.1214/105051604000000882

Fritsch, C., Villemonais, D., & Zalduendo, N. (2022). The multi-type bisexual galton-watson branching process. http://arxiv.org/abs/2206.09622

Giorgi, D., Kaakai, S., & Lemaire, V. (2023). *IBMPopSim: Individual Based Model Population Simulation*. https://github.com/DaphneGiorgi/IBMPopSim

Glémin, S., & Galtier, N. (2012). Genome evolution in outcrossing versus selfing versus asexual species. *Evolutionary genomics* (pp. 311–335). https://doi.org/10.1007/978-1-61779-582-4_11

Haccou, P., Jagers, P., & Vatutin, V. A. (2005). *Branching processes: variation, growth, and extinction of populations* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511629136

Haigh, J. (1978). The accumulation of deleterious genes in a population— Muller's Ratchet. *Theoretical Population Biology*, *14*(2), 251–267. https://doi.org/10.1016/0040-5809(78)90027-8

Haller, B. C., & Messer, P. W. (2019). SLiM 3: forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution*, *36*(3), 632–637. https://doi.org/10.1093/molbev/msy228

Harris, T. E. (1964). *The theory of branching processes*. Springer-Verlag.

Hartfield, M., & Glémin, S. (2016). Limits to adaptation in partially selfing species. *Genetics*, *203*(2), 959–974. https://doi.org/10.1534/genetics.116.188821

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

Heathcote, C. R. (1965). A branching process allowing immigration. *Journal of the Royal Statistical Society: Series B (Methodological)*, *27*(1), 138–143. https://doi.org/10.1111/j.2517-6161.1965.tb00596.x

Ikeda, N., & Watanabe, S. (1981). *Stochastic differential equations and diffusion processes*.

Ironside, J. E. (2010). No amicable divorce? challenging the notion that sexual antagonism drives sex chromosome evolution. *BioEssays*, *32*(8), 718–726. https://doi.org/10.1002/bies.200900124

Jay, P., Leroy, M., Le Poul, Y., Whibley, A., Arias, M., Chouteau, M., & Joron, M. (2022a). 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. (2022b). 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

Jeffries, D. L., Gerchen, J. F., Scharmann, M., & Pannell, J. R. (2021). A neutral model for the loss of recombination on sex chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1832), 20200096. https://doi.org/10.1098/rstb.2020.0096

Karlin, S., & Kaplan, N. (1973). Criteria for extinction of certain population growth processes with interacting types. *Advances in Applied Probability*, *5*(2), 183–199. https://doi.org/10.2307/1426032

Kent, T. V., Uzunović, J., & Wright, S. I. (2017). Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1736), 20160458. https://doi.org/10.1098/rstb.2016.0458

Kesten, H., & Stigum, B. P. (1966). A limit theorem for multidimensional galton-watson processes. *The Annals of Mathematical Statistics*, *37*(5), 1211–1223. https://doi.org/10.1214/aoms/1177699266

Kratochvíl, L., & Stöck, M. (2021). Preface. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1832), 20200088. https://doi.org/10.1098/rstb.2020.0088

Lambert, A. (2006). Probability of fixation under weak selection: a branching process unifying approach. *Theoretical Population Biology*, *69*(4), 419–441. https://doi.org/10.1016/j.tpb.2006.01.002

Lenormand, T., & Roze, D. (2022). Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science*, *375*(6581), 663–666. https://doi.org/10.1126/science.abj1813

Lenormand, T., & Roze, D. (2023). Revisiting the shelter theory: can deleterious mutations alone explain the evolution of sex chromosomes? https://doi.org/10.1101/2023.02.17.528909

Marguet, A. (2017). *Processus de branchement pour des populations structurées et estimateurs pour la division cellulaire* (Doctoral dissertation). Université Paris-Saclay, préparée à l'Ecole Polytechnique.

Marguet, A. (2019a). A law of large numbers for branching markov processes by the ergodicity of ancestral lineages. *ESAIM: Probability and Statistics*, *23*, 638–661. https://doi.org/10.1051/ps/2018029

Marguet, A. (2019b). Uniform sampling in a structured branching population. *Bernoulli*, *25*(4). https://doi.org/10.3150/18-BEJ1066

Méléard, S. (2016). *Modèles aléatoires en ecologie et evolution*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49455-4

Méléard, S., & Bansaye, V. (2015). *Stochastic models for structured populations: scaling limits and long time behavior*. Springer International Publishing. https://doi.org/10.1007/978-3-319-21711-6

Mode, C. J. (1971). *Multitype branching processes: theory and applications*. American Elsevier Pub. Co.

Molina, M. (2010). Two-sex branching process literature. *Workshop on branching processes and their applications* (pp. 279–293). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11156-3_20

Molina, M., Mota, M., & Ramos, A. (2014). Stochastic modeling in biological populations with sexual reproduction through branching models: application to coho salmon populations. *Mathematical Biosciences*, *258*, 182–188. https://doi.org/10.1016/j.mbs.2014.10.007

Muller, H. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *1*(1), 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

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. https://doi.org/10.1093/genetics/57.4.741

Otto, S. P. (2009). The evolutionary enigma of sex. *The American Naturalist*, *174*, S1–S14. https://doi.org/10.1086/599084

Otto, S. P., & Lenormand, T. (2002). Resolving the paradox of sex and recombination. *Nature Reviews Genetics*, *3*(4), 252–261. https://doi.org/10.1038/nrg761

Pénisson, S. (2010). *Conditional limit theorems for multitype branching processes and illustration in epidemiological risk analysis* (Doctoral dissertation). Universität Potsdam, Université Paris-Sud.

Rice, S. H. (2004). *Evolutionary theory: mathematical and conceptual foundations*. Sinauer Associates.

Rice, W. R. (1987). The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution*, *41*(4), 911–914. https://doi.org/10.1111/j.1558-5646.1987.tb05864.x

Roze, D. (2016). Background selection in partially selfing populations. *Genetics*, *203*(2), 937–957. https://doi.org/10.1534/genetics.116.187955

Roze, D., & Lenormand, T. (2005). Self-fertilization and the evolution of recombination. *Genetics*, *170*(2), 841–857. https://doi.org/10.1534/genetics.104.036384

Roze, D., & Rousset, F. (2004). Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. *Genetics*, *167*(2), 1001–1015. https://doi.org/10.1534/genetics.103.025148

Ruzicka, F., Dutoit, L., Czuppon, P., Jordan, C. Y., Li, X.-Y., Olito, C., Runemark, A., Svensson, E. I., Yazdi, H. P., & Connallon, T. (2020). The search for sexually antagonistic genes: practical insights from studies of local adaptation and statistical genomics. *Evolution Letters*, *4*(5), 398–415. https://doi.org/10.1002/evl3.192

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

Sewastjanow, B. (1975). *Verzweigungsprozesse* (R. Oldenbourg Verlag).

Stetsenko, R., & Roze, D. (2022). The evolution of recombination in self-fertilizing organisms (A. Agrawal, Ed.). *Genetics*, *222*(1), iyac114. https://doi.org/10.1093/genetics/iyac114

Tezenas, E., Giraud, T., Véber, A., & Billiard, S. (2022). The fate of recessive deleterious or overdominant mutations near mating-type loci under partial selfing. https://doi.org/10.1101/2022.10.07.511119

Tomasevic, M., Bansaye, V., & Véber, A. (2022). Ergodic behaviour of a multi-type growth-fragmentation process modelling the mycelial network of a filamentous fungus. *ESAIM: Probability and Statistics*. https://doi.org/10.1051/ps/2022013

Tran, V. C. (2006). *Modèles particulaires stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques* (Doctoral dissertation). Université de Nanterre - Paris X. https://theses.hal.science/tel-00125100

Tran, V. C. (2008). Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probability and Statistics*, *12*, 345–386. https://doi.org/10.1051/ps:2007052

Varadhan, S. R. S. (2008). Large deviations. *The Annals of Probability*, *36*(2). https://doi.org/10.1214/07-AOP348

Waples, R. S. (2022). What is N e, anyway? *Journal of Heredity*, *113*(4), 371–379. https://doi.org/10.1093/jhered/esac023

Winge, Ö. (1927). The location of eighteen genes inLebistes reticulatus. *Journal of Genetics*, *18*(1), 1–43. https://doi.org/10.1007/BF03052599

Wright, A. E., Dean, R., Zimmer, F., & Mank, J. E. (2016). How to make a sex chromosome. *Nature Communications*, *7*(1), 12087. https://doi.org/10.1038/ncomms12087

ECOLE
DOCTORALE
DE MATHEMATIQUES
HADAMARD

Titre : Modèles mathématiques pour l'étude de l'interaction entre suppression de recombinaison et mutations délétères au voisinage d'un locus de type sexuel

Mots clés : Processus de branchement multitype, Processus markovien de sauts, Simulations stochastiques, Recombinaison, Mutations délétères, Locus de type sexuel

Résumé : Cette thèse propose et développe plusieurs modèles stochastiques permettant d'avancer dans la compréhension de l'évolution de la suppression de recombinaison sur les chromosomes sexuels et de type sexuel. La recombinaison est un mécanisme d'échange de parties de chromosomes, qui crée de nouvelles combinaisons d'allèles. Cependant, de larges régions de suppression de recombinaison ont été observées chez de nombreuses espèces englobant des gènes impliqués dans la compatibilité lors de la reproduction sexuée (gènes déterminant le sexe ou le type sexuel). Les mécanismes induisant l'extension de la zone sans recombinaison au-delà de ces gènes impliqués dans la compatibilité sexuelle font encore débat. Dans cette thèse, on met en place différentes approches mathématiques pour étudier l'interaction entre la dynamique des mutations délétères et celle d'un suppresseur de recombinaison. Le premier chapitre a permis, grâce à l'analyse d'un modèle déterministe simple et à des simulations d'un modèle stochastique plus complexe, de montrer l'effet de la présence de mutations délétères à plusieurs étapes de l'évolution de la suppression de recombinaison. Le deuxième chapitre se concentre sur la dynamique des mutations délétères au voisinage d'un locus en permanence hétérozygote et dans des populations d'individus pouvant se reproduire par allo- ou auto-fécondation. On modélise l'évolution initiale de mutations délétères au moyen d'un processus de branchement multitype, dont on étudie la criticité et la distribution du temps d'extinction. Enfin, le troisième chapitre compare l'effet de l'accumulation de mutations délétères chez des individus autofécondants selon qu'ils peuvent recombiner ou non. On utilise un processus de sauts à valeurs mesures sur un espace de traits à trois dimensions pour étudier l'évolution de la charge mutationnelle d'individus possédant un locus de type sexuel toujours hétérozygote.

Title : Mathematical models to study the interaction between recombination suppression and deleterious mutations near a mating-type locus

Keywords : Multitype branching process, Markovian jump processes, Stochastic simulations, Recombination, Deleterious mutations, Mating-type locus

Abstract : This PhD manuscript presents the development of several stochastic models that contribute to our understanding of recombination suppression evolution on sex and mating-type chromosomes. Recombination is a mechanism that exchanges parts of chromosomes, which creates novel allelic combinations. However, there have been reports in a wide range of organisms of large regions with suppressed recombination that encompass genes involved in mating compatibility (i.e., genes determining sex or mating type). The nature of the mechanisms that induce the extension of the non-recombining zone beyond the genes involved in sex compatibility remains debated. In this PhD thesis, we use various mathematical approaches to study the dynamics of deleterious mutations and recombination suppressors. The first chapter shows, with the analysis of a simple deterministic model and the simulation of a more complex stochastic one, that deleterious mutations impact the evolution of a recombination suppressor at several stages. The second chapter focuses on deleterious mutations dynamics near a permanently heterozygous locus in selfing or outcrossing populations. We model the initial evolution of deleterious mutations with a multitype branching process and study its criticality and its extinction time distribution. In the third chapter, we compare the effect of deleterious mutation accumulation in selfing populations of recombining or non-recombining individuals. We use a measure-valued branching process on a three-dimensional trait space to study the evolution of the mutational load for populations carrying an always-heterozygous mating-type locus.

Institut Polytechnique de Paris
91120 Palaiseau, France