Recombination as a driver of genome evolution: characterisation of biased gene conversion in mice

Maud Gautier

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de

l'Université Claude Bernard Lyon 1

École Doctorale ED341

Evolution Ecosystèmes Microbiologie Modélisation (E2M2)

Spécialité de doctorat : Génomique évolutive

Soutenue publiquement le 25 septembre 2019, par :

Maud GAUTIER

La recombinaison comme moteur de l'évolution des génomes : caractérisation de la conversion génique biaisée chez la souris

Devant le jury composé de :

Mme Dominique MOUCHIROUD, Professeur, Université Claude Bernard Lyon 1	Président du jury
Mme Valérie BORDE, Directrice de Recherche, CNRS/Institut Curie	Rapporteur
M. Adam EYRE-WALKER, Professeur, Université de Sussex (Royaume-Uni)	Rapporteur
Mme Gwenael PIGANEAU, Directrice de Recherche, CNRS/BIOM	Rapporteur
M. Bertrand LLORENTE, Directeur de Recherche, CNRS/CRCM	Examinateur
M. Laurent DURET, Directeur de Recherche, CNRS/LBBE	Directeur de thèse

À Loïc Rajjou,

qui, le premier, m'a donné goût à la recherche

et sans la rencontre duquel

l'idée d'un doctorat ne m'aurait pas même effleuré l'esprit.

'Ibergekumene tsores iz gut tsu dertseylin.'
(C'est un plaisir de raconter les ennuis passés.)

— Proverbe yiddish, placé en épigraphe de
Primo Levi, *Le système périodique* (1975)

Recombination as a driver of genome evolution: characterisation of biased gene conversion in mice

Abstract

During meiosis, recombination hotspots host the formation of DNA double-strand breaks (DSBs). DSBs are subsequently repaired through a process which, in a wide range of species, is biased towards the favoured transmission of G and C alleles: GC-biased gene conversion (gBGC). The intensity of this fundamental distorter of meiotic segregation strongly varies between species but the factors dictating its evolution are not known. We thus aimed at directly quantifying the transmission bias in mice and comparing the parameters on which it depends with other mammals.

Here, we coupled capture-seq and bioinformatic techniques to implement an approach that proved 100 times more powerful than current methods to detect recombination. With it, we identified 18,821 crossing-over (CO) and non-crossover (NCO) events at very high resolution in single individuals and could thus precisely characterise patterns of recombination in mice. In this species, recombination hotspots are targeted by PRDM9 and are therefore subject to a second type of biased gene conversion (BGC): DSB-induced BGC (dBGC). Quantifying both dBGC and gBGC with our data brought to light the fact that, in cases of structured populations, past gBGC from the parental lineages is hitchhiked by dBGC when the populations cross. We next observed that, in male mice, only NCOs — and more particularly single-marker NCOs — contribute to the intensity of gBGC. In contrast, in humans, both NCOs and at least a portion of COs (those with complex conversion tracts) distort allelic frequencies. This suggests that the DSB repair machinery leading to gBGC varies across mammals. Our findings are also consistent with the hypothesis of a selective pressure restraining the intensity of the deleterious gBGC process at the population-scale: this would materialise through a multi-level compensation of the effective population size by the recombination rate, the length of conversion tracts and the transmission bias.

Altogether, our work has allowed to better comprehend how recombination and biased gene conversion proceed in the mammalian clade.

Keywords: Recombination, Biased gene conversion, PRDM9, Hotspots, Genomics, Molecular evolution, Mammals, Sperm-typing.

Résumé en français

Au cours de la méiose, les points chauds de recombinaison sont le siège de la formation de cassures double-brin de l'ADN. Ces dernières sont ensuite réparées par un processus qui, chez de nombreuses espèces, favorise la transmission des allèles G et C : la conversion génique biaisée vers GC (gBGC). L'intensité de cet important distorteur de la ségrégation méiotique varie fortement entre espèces mais les facteurs déterminant son évolution sont toujours inconnus. Nous avons donc voulu quantifier directement le biais de transmission chez la souris et comparer les paramètres dont il dépend avec d'autres mammifères.

Dans cette étude, en couplant des développements bioinformatiques à une technique de capture ciblée d'ADN suivie de séquençage haut-débit (capture-seq), nous avons réussi à mettre au point une approche qui s'est révélée 100 fois plus performante pour détecter les événements de recombinaison que les méthodes existant actuellement. Ainsi, nous avons pu identifier 18 821 crossing-overs (COs) et non-crossovers (NCOs) à très grande résolution chez des individus uniques, ce qui nous a permis de caractériser minutieusement la recombinaison chez la souris. Chez cette espèce, les points chauds de recombinaison sont ciblés par la protéine PRDM9 et sont donc soumis à une deuxième forme de conversion génique biaisée (BGC) : le biais d'initiation (dBGC). La quantification du dBGC et du gBGC à partir de nos données nous a permis de mettre en lumière le fait que, au moment où des populations structurées s'hybrident, le gBGC des lignées parentales est propagé par un phénomène d'auto-stop génétique (genetic hitchhiking) provenant du dBGC. Nous avons ensuite pu observer que, chez les souris mâles, seuls les NCOs — et plus particulièrement les NCOs contenant un seul marqueur génétique— contribuent à l'intensité du gBGC. En comparaison, chez l'Homme, à la fois les NCOs et au moins une part des COs (ceux qui présentent des tracts de conversion complexes) distordent les fréquences alléliques. Ceci suggère que la machinerie de réparation des cassures double-brin qui induit le biais de conversion génique (BGC) présente des variations au sein des mammifères. Nos résultats sont aussi en accord avec l'hypothèse selon laquelle une pression de sélection limiterait l'intensité de ce processus délétère à l'échelle de la population. Cela se traduirait par une compensation de la taille efficace de population à de multiples niveaux : par le taux de recombinaison, par la longueur des tracts de conversion et par le biais de transmission.

Somme toute, notre travail a permis de mieux comprendre la façon dont la recombinaison et la conversion génique biaisée opèrent chez les mammifères.

Mots-clés: Recombinaison, Conversion génique biaisée, PRDM9, Points chauds, Génomique, Évolution moléculaire, Mammifères, Sperm-typing.

Résumé étendu en français

Lorsque l'on traite de l'évolution des génomes, trois forces sont classiquement invoquées : la mutation, la sélection naturelle et la dérive génétique. Toutefois, depuis une vingtaine d'année, une quatrième force a fait son entrée sur la scène évolutive : la conversion génique biaisée, que nous noterons 'BGC' (de l'anglais *biased gene conversion*). Ce phénomène est une conséquence directe du processus de recombinaison méiotique chez les espèces à reproduction sexuée.

Chez les mammifères en effet, après s'être fixée à certains loci cibles appelés 'points chauds de recombinaison', la protéine PRDM9 recrute la machinerie de formation de cassures double-brin et marque, de ce fait, l'initiation d'un événement de recombinaison (Baudat et al., 2010; Myers et al., 2010; Parvanov et al., 2010). Ce dernier doit ensuite être réparé en utilisant le chromosome homologue comme matrice, ce qui mène à ce qu'on appelle un événement de conversion génique, c'est-à-dire le transfert non-réciproque d'une information de séquence d'ADN.

Toutefois, si PRDM9 présente une plus grande affinité de liaison avec la séquence de l'un des deux chromosomes (que nous appellerons 'haplotype'), la cassure s'initiera préférentiellement sur cet haplotype, et l'événement de conversion génique se fera donc préférentiellement dans un sens donné : c'est ce qu'on appelle le biais d'initiation, aussi appelé conversion génique biaisée induite par cassure double brin et noté 'dBGC' (de l'anglais *double-strand break-induced biased gene conversion*). Du fait de ce phénomène, les points chauds finissent nécessairement par s'éroder : comme l'haplotype portant le motif ciblé par PRDM9 est le siège de la cassure, il est systématiquement converti par l'autre haplotype, et voué à disparaître (Boulton et al., 1997).

Il existe une deuxième forme de conversion génique biaisée : la conversion génique biaisée vers GC, que l'on notera 'gBGC' (de l'anglais *GC-biased gene conversion*). En effet, il a été observé chez plusieurs espèces de façon directe (Mancera et al., 2008; Si et al., 2015; Williams et al., 2015; Halldorsson et al., 2016; Keith et al., 2016; Smeds et al., 2016) ou indirecte (Escobar et al., 2011; Pessia et al., 2012; Figuet et al., 2014) que la réparation des cassures double-brin favorise les allèles G et C par rapport aux allèles A et T.

La quantification du coefficient de conversion génique biaisée à l'échelle des populations (B) chez un grand nombre de métazoaires (Galtier et al., 2018) a mis en évidence un résultat étonnant: l'intensité du gBGC ne varie que dans une gamme de valeurs très restreinte. Par exemple, chez les mammifères placentaires, B reste dans une fourchette de 0 à 7 (Lartillot, 2013b). Étant donné que B correspond

au produit de la taille efficace de population (N_e) par le coefficient de gBGC (b) et que la taille efficace peut varier sur plusieurs ordres de grandeurs parmi les métazoaires, b ne peut mécaniquement pas être identique chez toutes les espèces. Au contraire, un ou plusieurs des paramètres dont b dépend (le taux de recombinaison r, la longueur des tracts de conversion L et le biais de transmission b_0) varient nécessairement inversement à la taille efficace.

Cependant, peu de données sont disponibles pour comprendre la base de la dépendance entre N_e et b: le biais de transmission (b_0) n'a été mesuré que chez quelques espèces (Mancera et al., 2008; Si et al., 2015; Williams et al., 2015; Halldorsson et al., 2016; Keith et al., 2016; Smeds et al., 2016) et, parmi les mammifères, la seule espèce chez qui ce biais a été mesuré de façon directe (*Homo sapiens*) présente une très faible taille efficace d'environ 10,000 (Takahata, 1993; Erlich et al., 1996; Harding et al., 1997; Charlesworth, 2009; Yu et al., 2004).

Afin d'apporter un éclairage nouveau sur l'interaction entre b et N_e, nous avons donc voulu quantifier le gBGC chez une autre espèce de mammifères présentant une taille efficace beaucoup plus grande que celle de l'Homme (Geraldes et al., 2008; Phifer-Rixey et al., 2012; Davies, 2015): la souris *Mus musculus*.

Pour pouvoir quantifier précisément le gBGC, il est nécessaire de disposer d'un grand nombre d'événements de recombinaison. Or, la méthode généralement utilisée pour détecter ces événements — l'analyse de pedigrees — est extrêmement gourmande en ressources : elle requiert le séquençage de génomes complets d'un grand nombre d'individus et permet de détecter seulement un nombre limité de recombinants. Nous avons donc mis au point une nouvelle approche permettant de détecter plusieurs milliers de recombinants à très haute résolution chez des individus uniques.

Concrètement, notre approche repose sur deux étapes principales. Premièrement, puisque la recombinaison n'est identifiable qu'à partir du génotypage de marqueurs hétérozygotes, nous avons croisé deux lignées de souris (C57BL/6J que nous noterons 'B6' et CAST/EiJ que nous appellerons 'CAST') issues de deux sous-espèces (*Mus musculus domesticus* et *Mus musculus castaneus*) présentant un fort taux de polymorphisme de 0.74% (Keane et al., 2011; Yalcin et al., 2012). Les points chauds de recombinaison chez l'hybride F1 qui résulte de ce croisement (B6xCAST) ont déjà été identifiés par d'autres que nous (Baker et al., 2015a). Afin de maximiser le nombre de recombinants détectables, nous en avons donc sélectionné 1 018 qui sont particulièrement denses en marqueurs hétérozygotes. Nous avons ensuite enrichi l'ADN du sperme de cet hybride en fragments provenant de ces loci grâce à une technique de ciblage spécifique suivie de séquençage haut-débit (capture-seq).

La deuxième étape de notre procédure consiste à génotyper les molécules séquencées de façon individuelle, et d'identifier, parmi ces dernières, celles correspondant à des événements de recombinaison. Toute la difficulté de cette analyse réside dans le fait que les molécules sont uniques: dès lors, toute erreur de séquençage ou toute ambiguïté d'alignement peut devenir une source d'erreur à l'origine de faux positifs (i.e. de fragments détectés comme recombinants alors qu'ils ne le sont pas). Lors de la mise en œuvre de notre approche, nous nous sommes rendus compte que les anomalies les plus critiques à cet égard provenaient de l'étape d'alignement car celle-ci est biaisée vers le génome de référence. L'étape cruciale de notre méthode a donc été d'effectuer la procédure en utilisant successivement les deux génomes parentaux comme référence.

Au final, notre approche s'est révélée extrêmement performante. A titre de comparaison, les études récentes ayant obtenu des cartes de recombinaison à haute résolution chez l'Homme, la souris ou l'oiseau (Halldorsson et al., 2016; Smeds et al., 2016; Li et al., 2018) se sont montrées plus de cent fois moins puissantes que notre méthode pour détecter ces événements.

L'approche que nous avons mise au point nous a permis de détecter 18 821 événements de recombinaison chez la souris et donc de caractériser précisément la recombinaison sur environ un millier de points chauds (jusqu'alors, ceci n'avait été fait que sur une poignée de points chauds).

En premier lieu, nous avons pu observer l'étendue de la variation du taux de recombinaison entre les points chauds et identifier quelques uns de ses déterminants. En particulier, l'affinité de liaison entre la protéine PRDM9 et son motif cible est parfaitement proportionnelle à l'activité recombinationnelle du point chaud. Toutefois, les points chauds dont les deux haplotypes (celui venant de B6 et celui venant de CAST) présentent un différentiel d'affinité à PRDM9 important (les points chauds dits 'asymétriques') ont un taux de recombinaison fortement réduit (d'un facteur deux à quatre) par rapport à l'attendu basé sur l'intensité du signal PRDM9.

Un certain nombre d'événements de recombinaison (en particulier ceux dont le tract de conversion ne chevauche aucun marqueur polymorphe) ne sont pas détectables. Dès lors, les paramètres de recombinaison observés — comme la longueur des tracts de conversion, le taux de recombinaison et le ratio de COs et de NCOs — ne sont pas forcément représentatifs des paramètres de recombinaison réels. Pour pouvoir estimer ces paramètres réels, il est donc nécessaire de passer par des méthodes inférentielles telles que la méthode bayésienne approchée (*approximate bayesian computation*) qui consiste à simuler le processus biologiques avec différents paramètres et à sélectionner les simulations dont le résultat est proche

des observations biologiques. Par ce biais, nous avons pu estimer de façon indirecte les paramètres de recombinaison chez la souris : les tracts de conversion des COs mesurent 450 paires de bases en moyenne contre 35 pour les NCOs, et le taux de recombinaison moyen sur l'ensemble des points chauds que nous avons étudié est de 30 cM/Mb.

Ensuite, en cherchant à quantifier le biais de transmission (b_0) des allèles GC et donc l'intensité du gBGC (b) chez la souris, nous avons remarqué que, dans un dispositif expérimental tel que le nôtre, ce biais était affecté par l'autre forme de conversion génique: le biais d'initiation (dBGC). En effet, prenons le cas de deux populations possédant deux allèles *Prdm9* distincts évoluant donc de façon indépendante dans leurs lignées respectives. Dans chacune des lignées, les points chauds ciblés par l'allèle présent s'érodent sous l'effet du dBGC et s'enrichissent en même temps en allèles G et C sous l'effet du gBGC. Lorsque l'on croise deux individus issus de ces deux lignées, l'allèle *Prdm9* initie la cassure double-brin sur l'haplotype pour lequel il a la plus grande affinité, c'est-à-dire l'haplotype de la lignée avec laquelle il n'a *pas* co-évolué, puisque celle dans laquelle il se trouvait a vu ses points chauds s'éroder. Ainsi, c'est l'haplotype de sa lignée d'origine — qui est localement enrichi en GC — qui sera systématiquement le donneur lors de l'événement de conversion génique. De ce fait, le gBGC qui a eu lieu dans les lignées parentales est propagé par un phénomène d'auto-stop génétique (*genetic hitchhiking*) provenant du dBGC.

Pour pouvoir quantifier le gBGC correctement, il fallait donc contrôler pour cet effet d'auto-stop, ce que nous avons fait en sous-échantillonnant les tracts de conversion analysés pour égaliser le nombre d'événements de conversion ayant un donneur B6 à ceux ayant un donneur CAST. Dès lors, nous avons pu quantifier le gBGC et observer que le biais de transmission (b_0) est nul pour les COs et extrêmement faible chez les NCOs contenant plusieurs marqueurs génétiques (NCO-2+). En revanche, le biais est très élevé pour les NCOs contenant un seul marqueur (NCO-1) : l'intensité du biais est comparable à ce qui a été observé chez l'humain (Halldorsson et al., 2016).

A partir de là, nous avons pu comparer la relation entre l'intensité du gBGC (b) et la taille efficace de population (N_e) chez les deux espèces de mammifères pour lesquelles le biais de transmission (b_0) a été quantifié de façon directe : la souris et l'Homme. Nos analyses indiquent que le taux de recombinaison et la longueur des tracts de conversion participent tous deux à limiter l'intensité du gBGC (b) chez la souris par rapport à l'Homme et, bien que les données disponibles

à l'heure actuelle soient insuffisantes pour le confirmer, il semblerait que le biais de transmission des COs y participe également.

Globalement, ces observations sont compatibles avec l'hypothèse selon laquelle une pression de sélection limiterait l'intensité de ce processus délétère à l'échelle de la population par le biais d'une compensation de la taille efficace de population à de multiples niveaux : par le taux de recombinaison, par la longueur des tracts de conversion et, peut-être, par le biais de transmission des COs.

Enfin, la méthode de détection des recombinants à l'échelle d'individus uniques est tout indiquée pour étudier le rôle individuel de gènes impliqués dans le processus de recombinaison. Pour ce faire, il faut analyser des individus homozygotes pour une version inactivée du gène d'intérêt mais présentant tout de même un haut niveau d'hétérozygotie pour que la recombinaison soit détectable. Comme des individus F2 issus du croisement de trois lignées distinctes peuvent présenter de telles caractéristiques alors que des individus F1 issus d'un unique croisement ne le peuvent pas, il nous a fallu adapter notre méthode à un tel schéma de croisement.

Suite à cela, nous avons pu analyser le rôle du gène *Hfm1*, une hélicase d'ADN essentielle à la résolution des cassures double-brin en COs : nous avons observé que son inactivation menait à un taux de recombinaison plus élevé et à des tracts de conversion de COs sensiblement plus courts que chez les individus non mutants.

Somme toute, notre travail a mené à la mise au point d'une approche originale de détection de la recombinaison à haute résolution et à faible coût chez des individus uniques. Cette approche ouvre la voie à l'étude plus poussée des gènes impliqués dans le processus de recombinaison et nous a permis de mieux comprendre la façon dont la recombinaison et la conversion génique biaisée opèrent chez les mammifères.

Remerciements

Cela fait maintenant un peu plus de vingt-sept ans que, régulièrement, je m'étonne de ce que la chance me sourie si souvent et force est de constater que ce doctorat aura été une occasion de plus de la mesurer.

Pouvoir réaliser ma thèse sous la direction de Laurent DURET a en effet été une aubaine inouïe. Sa disponibilité manifestement infinie, son souci de me fournir un cadre idéal de travail et de vie, sa résolution à me prodiguer une formation complète et l'extraordinaire bienveillance dont il a fait preuve chaque jour sont allés bien au-delà de ce que j'aurais pu espérer. À travers son inspection minutieuse des détails méthodologiques, son intransigeance à l'égard des imprécisions, l'habitude qu'il a de décortiquer les notions complexes pour les rendre limpides et sa détermination à replacer systématiquement nos résultats dans une image plus globale, Laurent m'a littéralement appris à faire de la science. Mais, au delà de son encadrement scientifique exceptionnel, c'est son humanité que je voudrais saluer. Je n'arrive pas à m'imaginer comment j'aurais pu arriver au bout de cette thèse sans l'immense générosité dont il a fait preuve ni sans sa capacité à me redonner espoir dans les moments plus difficiles. Je voudrais donc lui dire ici toute l'admiration que je lui porte ainsi que toute ma gratitude pour les conditions privilégiées dont il m'a fait bénéficier au cours des trois années que j'ai passées à ses côtés.

Je voudrais aussi exprimer mes sincères remerciements aux personnes qui m'ont fait l'honneur d'accepter de lire et d'évaluer ce travail : Valérie BORDE, Adam EYRE-WALKER, Bertrand LLORENTE, Dominique MOUCHIROUD et Gwenael PIGANEAU ; ainsi qu'à ceux qui ont participé à mon comité de suivi pour leurs conseils tant sur les aspects scientifiques que personnels : Nicolas GALTIER, Annabelle HAUDRY, Tristan LEFÉBURE, Bernard DE MASSY et Jonathan ROMIGUIER.

Le travail qui est présenté dans ce manuscrit est le fruit de collaborations qui ont impliqué nombre de personnes envers qui je suis pleinement reconnaissante. Merci donc à Frédéric BAUDAT, Valérie BORDE, Corinne GREY et Bernard DE MASSY d'avoir réalisé la totalité du travail expérimental qui a été essentiel pour cette thèse, pour leur patience à mon endroit et pour leurs suggestions éclairées, à Nicolas LARTILLOT pour ses commentaires avisés, sa douceur et son indulgence à mon égard, ainsi qu'à Brice LETCHER pour le travail admirable qu'il a réalisé pendant son stage et dont je me suis allègrement servie pour rédiger cette thèse.

premier chapitre ; merci à M. BEAUX et à Mme MOLLIÈRE, professeurs de biologie en prépa, qui ont su faire de la génétique un sujet passionnant.

Finalement, pour m'avoir donné les clés nécessaires à la bonne réalisation de cette thèse, merci à mes encadrants de stage : Édouard BOVE, Claudia CHICA, Juliette DE MEAUX, Florine POIROUX, Loïc RAJJOU et François ROUDIER. Je tiens particulièrement à signifier à Loïc la profonde estime que j'ai pour lui et mon indéfectible reconnaissance pour m'avoir soutenue quand je me croyais abbatue, aiguillée quand je me sentais perdue et encouragée à entreprendre un doctorat quand je ne m'en savais pas capable.

Je voudrais terminer en remerciant ceux qui m'ont permis de garder une santé mentale à peu près normale en me rappelant que la vie ne s'arrête pas à l'enceinte du laboratoire.

Merci donc à mes amis plongeurs du COOLAPIC qui sont bien trop nombreux pour être mentionnés individuellement mais qui ont été une vraie bulle d'oxygène lors de nos sorties en mer. Je voudrais aussi dire toute ma gratitude à mes camarades de théâtre, en particulier Vérane LYON, Marie PINEL, Hannah SAMAMA et Estelle VALETTE, pour nos éclats de rires ainsi qu'à mes amis parisiens venus passer un ou plusieurs week-ends à mes côtés dans la contrée Lyonnaise : Alicia BERRET, Yann BOULESTREAU, Lucie GOMEZ, Mathieu LAUGIER, Ulysse LE GOFF, Jordane LELONG, Anne RABAULT et Anne SCHNEIDER.

Je n'ose imaginer la tournure qu'aurait pris ma vie si mes parents n'avaient pas été tels qu'ils sont. Ils ont toujours placé les intérêts de leurs quatre enfants bien avant les leurs et leur abnégation, leur amour et leur soutien inconditionnel font d'eux des réels modèles de vie. Merci donc, Papa, Maman : c'est un cadeau de vous avoir comme parents ! Mais ce cadre familial n'aurait pas été complet sans mes frères et leurs compagnes, Arthur et Marie-Élise, Ronan et Anna, ainsi que ma sœur Lévana que je me dois de remercier pour l'indulgence dont ils ont fait preuve à l'égard de mes moments d'anxiété, pour leur humour et pour la gaieté qu'ils apportent à mon quotidien. Merci en particulier à Arthur de m'avoir consolée et d'avoir soulagé mes difficultés en les prenant sur ses épaules, à Ronan dont l'espièglerie m'a allégé l'esprit et dont la curiosité m'a ouvert les yeux sur d'autres sciences et à Lévana de m'avoir toujours réconfortée et soutenue avec une constance et un dévouement qui forcent l'admiration.

Merci, enfin, à Alexandre CHAINTREUIL d'avoir su systématiquement transformer mes angoisses en rires, mes larmes en joie et mes craintes en doses de confiance en moi. Qu'il continue à le faire quelques décennies de plus me comblerait.

Villeurbanne, le 26 juillet 2019

Contents

II Objectives of the thesis 101

III Results 105

x

List of Figures

List of Tables

xvi

Abbreviations

AA . . . . Amino acid.

ABC . . . . . . Approximate bayesian computation.

ALT . . . . . . Alternate (*versus* REF: reference).

ATM (kinase) Ataxia telangiectasia mutated (kinase).

A . . . . Adenine.

B . . . . Population-scaled gBGC coefficient.

b . . . . gBGC coefficient.

b_0 . . . . Transmission bias.

b_{dBGC} . . . . . . dBGC coefficient.

B6 . . . . Mouse strain C57BL/6J.

BD . . . . B6 or DBA2.

BER . . . . . . Base excision repair.

BGC . . . . . . Biased gene conversion.

BLM . . . . . . Bloom syndrome RecQ helicase-like.

BQSR . . . . . Base quality score recalibration.

BWA . . . . . Burrows-Wheeler aligner (bioinformatic tool).

C2H2 . . . . . Cys_2-His_2.

CAST . . . . . Mouse strain CAST/EiJ.

CCO . . . . . . Complex CO.

CE . . . . Central element.

CI . . . . Confidence interval.

cM . . . . Centimorgan.

chB . . . . . . Cold haplotype: B6.

chC . . . . . . Cold haplotype: CAST.

COA . . . . . . Crossing-over assurance.

COH . . . . . . Crossing-over homeostasis.

COI . . . . . . Crossing-over interference.

CO . . . . Crossing-over (or crossover).

CT⋆ . . . . . . Observed (or inferred) conversion tract.

CT . . . . Conversion tract.

CUB . . . . . . Codon usage bias.

C . . . . Cytosine.

ChIP-seq . . . Chromatin immunoprecipitation followed by sequencing.

D-loop . . . . . Displacement loop.

DAFS . . . . . Derived allele frequency spectrum (a.k.a. site frequency spectrum, SFS).

DAF . . . . . . Derived allele frequency.

DBA2 . . . . . Mouse strain DBA/2J.

dBGC . . . . . DSB-induced biased gene conversion.

dHJ . . . . . . Double-Holliday junction.

DMC1 . . . . . DNA meiotic recombinase 1 (or Dosage Suppressor of Mck1 homologue).

DNA . . . . . . Deoxyribonucleic acid.

DNM . . . . . *De novo* mutation.

DSBR . . . . . Double-strand break repair.

DSB . . . . . . Double-strand break.

dsDNA . . . . Double-stranded DNA.

EME1 . . . . . Essential meiotic structure-specific endonuclease 1 (Yeast homologue: Mms4).

EM . . . . Expectation maximisation.

F1 hybrid . . . First filial generation of offspring of distinct parental types.

F2 . . . . Second filial generation. Results from a F1 × F1 cross.

F3, F4, etc . . Subsequent filial generations.

FIMO . . . . . Find individual motif occurrences (bioinformatic tool).

FP . . . . False positive.

GATK . . . . . Genome analysis toolkit (bioinformatic tool).

gBGC . . . . . GC-biased gene conversion.

GC₃ . . . . . . GC-content at third codon position.

GC* . . . . . . Stationary (or equilibrium) GC-content.

GRC . . . . . . Genome reference consortium.

GRCm38 . . . Genome reference consortium mouse build 38 (synonym: mm10).

G . . . . Guanine.

Gb . . . . Giga base pairs.

H3K36me3 . . Addition of 3 methyl groups to the lysine 36 on the histone H3 protein.

H3K4me3 . . . Addition of 3 methyl groups to the lysine 4 on the histone H3 protein.

H$_{n1}$K$_{n2}$me$_{n3}$ . . Addition of $n3$ methyl groups to the lysine $n2$ on the histone $n1$ protein.

HACNS . . . . Human accelerated conserved non-coding sequence (a.k.a. HAR).

HAR . . . . . . Human accelerated region (a.k.a. HACNS).

HFM1 . . . . . Helicase for meiosis 1 (yeast homologue: Mer3).

HJ . . . . Holliday junction.

HORMAD1 . HORMA domain-containing protein 1.

HRR . . . . . . Homologue recognition region.

HR . . . . Homologous recombination.

Hop1 . . . . . Homologue pairing 1 (mouse ortholog HORMAD1).

INM . . . . . . Inner nuclear membrane.

Indel . . . . . . Insertion or deletion.

kb . . . . Kilo base pairs.

KRAB . . . . . Krüppel-associated box.

L . . . . Conversion tract length.

LCA . . . . . . Last common ancestor.

LD . . . . Linkage disequilibrium.

LE . . . . Lateral element.

MCM8,9 . . . Minichromosome maintenance deficient 8, 9.

MEI1 . . . . . Meiosis inhibitor protein 1.

MEI4 . . . . . Meiosis inhibitor protein 4.

MEME . . . . Multiple EM for motif elicitation (bioinformatic tool).

MGP . . . . . Mouse genomes project.

MHC . . . . . Major histocompatibility complex.

MLH1 . . . . . MutL protein homologue 1.

mm10 . . . . . *Mus musculus* genome build 10 (synonym: GRCm38).

MMR . . . . . Mismatch repair.

MRE11 . . . . Meiotic Recombination 11 (yeast homologue: Mre11).

mRNA . . . . Messenger RNA.

MSH4,5 . . . . MutS protein homologue 4, 5.

MUS81 . . . . Crossover junction endonuclease 81.

M . . . . Morgan.

Mb . . . . Mega base pairs.

Mer2 . . . . . . Yeast recombination protein 107 (alias: Rec107).

MHC . . . . . Major histocompatibility complex.

Mms4 . . . . . Methyl methanesulfonate sensitivity 4 (Mouse homologue: EME1).

NBS1 . . . . . Nibrin (yeast homologue: Xrs2).

NCO-1 . . . . Single-marker NCO.

NCO-2+ . . . Multiple-marker NCO.

NCO . . . . . . Non crossing-over (or non-crossover).

NDR . . . . . . Nucleosome-depleted region.

Ne . . . . Effective population size.

NE . . . . Nuclear enveloppe.

NHEJ . . . . . Non-homologous end-joining.

NOV . . . . . . Novel.

ONM . . . . . Outer nuclear membrane.

PAR . . . . . . Pseudoautosomal region.

PCR . . . . . . Polymerase chain reaction.

PC . . . . Pairing centre.

PMS . . . . . . Post-meiotic segregation.

pot-NCO-1 . . Potential single-marker NCO.

PRDM9 . . . . Positive regulatory (PR) domain zinc finger protein 9.

r . . . . Recombination rate.

RAD50 . . . . RAD50 double-strand break repair protein (yeast homologue: Rad50).

RAD51 . . . . RAD51 double-strand break repair protein (yeast homologue: Rad51).

Rec114 . . . . Yeast recombination protein 114.

REC8 . . . . . Meiotic recombination protein 8.

REF . . . . . . Reference (*versus* ALT: alternate).

RFLP . . . . . Restriction fragment length polymorphism.

RNA . . . . . . Ribonucleic acid.

RNAi . . . . . RNA interference.

RNF212 . . . . RING finger protein 212 (yeast homologue: Zip4.)

RPA . . . . . . Replication protein A.

RR . . . . Recombination rate.

S . . . . Strong nucleotide (G or C).

SCP1 . . . . . Synaptonemal complex protein 1.

SCP2 . . . . . Synaptonemal complex protein 2.

SCP3 . . . . . Synaptonemal complex protein 3.

SC . . . . Synaptonemal complex.

SDSA . . . . . Synthesis-dependent strand annealing.

SEI . . . . Single-end invasion.

SFS . . . . . . Site frequency spectrum (a.k.a. derived allele frequency spectrum, DAFS).

SNP . . . . . . Single-nucleotide polymorphism.

SPO11 . . . . . SPO11 initiator of meiotic double-strand breaks.

ssDNA . . . . Single-stranded DNA.

SSRXD . . . . Synovial sarcoma X repression domain.

sym . . . . . . Symmetric.

SW (S → W) Mutation from a 'strong' (S) (i.e. G or C) to a 'weak' (W) nucleotide (i.e. A or T). Alternatively noted GC → AT.

SYCE1 . . . . Synaptonemal complex central element 1.

SYCE2 . . . . Synaptonemal complex central element 2.

Srs2 . . . . . . Yeast suppressor of Rad six 2.

tB . . . . Targeted by PRDM9$^{\text{Dom2}}$.

tC . . . . Targeted by PRDM9$^{\text{Cst}}$.

TEX11 . . . . Testis-expressed sequence 11 (yeast homologue: Zip3).

TE . . . . Transposable element.

TF . . . . Transverse filaments.

TSS . . . . . . Transcription start site.

T . . . . Thymine.

tRNA . . . . Transfer RNA.

UTR . . . . . . Untranslated region.

VQSR . . . . . Variant quality score recalibration.

W . . . . Weak nucleotide (A or T).

WS (W → S) Mutation from a 'weak' (W) (i.e. A or T) to a 'strong' (S) nucleotide (i.e. G or C). Alternatively noted AT → GC.

WT . . . . . . Wild-type.

Xrs2 . . . . . . DNA repair protein (Mouse homologue: NBS1).

Zip3,4 . . . . . Homologues of mammalian TEX11 and RNF212

Znf . . . . Zinc finger.

Prdm9^{Cst} . . . *Prdm9* allele carried by CAST mice.

Prdm9^{Dom2} . . *Prdm9* allele carried by B6 mice.

Definitions

Achiasmy The phenomenon where autosomal recombination is completely absent in one sex of a species.

Allele A variant form of a given gene.

Anaphase Third stage of mitosis, meiosis I and meiosis II.

Ascospore Reproductive cells of a certain class of fungi (ascomycetes).

Asymmetric hotspot Hotspot for which one of the two haplotypes is more likely to host the double-strand break.

Apoptosis Programmed cell death (from the Greek word ἀπόπτωσις: 'falling off').

Biased gene conversion Process by which gene conversion is biased towards a given outcome. It occurs when one haplotype has a higher probability of being the donor.

C-terminus End of an amino acid chain terminated by a free carboxyl group.

Centimorgan Unit of genetic distance: 1 cM corresponds to a frequency of crossing-overs of 1%.

ChIP-sequencing Method used to analyse protein interaction with DNA.

Chiasma (*pl.* chiasmata) An exchange (crossing-over) between paired chromatids, observed cytologically between diplotene and the first meiotic anaphase, from the Greek word χίασμα: 'X-shaped cross'.

Chromatid A DNA molecule associated to proteins and forming one half of the two identical copies of a replicated chromosome.

Codon usage bias Unequal frequency of the alternative codons that specify the same amino acid.

Codon Sequence of three nucleotides coding for a given amino acid.

Cold haplotype The haplotype that is most often the donor in the gene conversion event.

CpG (or CG) site Region of DNA where a cytosine nucleotide is followed by a guanine nucleotide in the 5'-to-3' direction.

CpG island Region with a high frequency of CpG sites.

Crossing-over Recombination event leading to the reciprocal exchange of the DNA sequences flanking the crossing-over point.

DNA capture Hybridisation-based targeted-DNA enrichment.

DSB-induced biased gene conversion The form of biased gene conversion due to the differential formation of double-strand breaks on the two haplotypes.

Diploid Organism (or phase) displaying a ploidy of 2 ($n = 2$), i.e. two sets of chromosomes (which are paired).

Ectopic gene conversion Gene conversion between copies of a gene family.

Effective population size The number of individuals in a population who contribute to the next generation.

GC-biased gene conversion The process by which the GC-content increases because of biased gene conversion.

GC-content The percentage of G or C nucleotidic bases in a DNA sequence.

Gamete Product of meiosis.

Gene conversion A non-reciprocal recombination process that results in one sequence being converted into the other.

Genetic drift The random fluctuation in allele frequencies due to random sampling of individuals.

Genetic distance Distance between DNA markers on a chromosome measured as the amount of crossing-overs between them.

Genetic interference The fact that the formation of a recombination event can affect that of others in adjacent regions.

Genetic linkage Non-independent assortment of genes.

Genetic marker A known site of heterozygosity.

Genotyping The process by which DNA is analyzed to determine which genetic variant (allele) is present for a given marker.

Haploid Organism (or phase) displaying a ploidy of 1 ($n = 1$), i.e. a single set of chromosomes.

Haplotype (In the context of this thesis, used to define the background of the PRDM9 motif)

Heterochiasmy The differential recombination rates between the sexes of a species.

Heteroduplex DNA A DNA portion where the two strands composing it contain different information for the segregating marker.

Holocentric Chromosome devoid of any major centromeric constriction.

Homologues A set of one paternal and one maternal chromosomes that pair up during meiosis (a.k.a. homologous chromosomes).

Homologous recombination The process through which segments of DNA are exchanged between two DNA duplexes with high sequence similarity.

Hot haplotype The haplotype that most often hosts the double-strand break.

in silico In a computing context.

in vitro Outside the normal biological context.

in vivo Inside the normal biological context.

Interphase Period of cell growth before cell division.

Locus (*pl.* loci) Fixed position of a genetic marker on a chromosome (from the Latin word *locus*: 'place').

Linkage disequilibrium Non-random associations between loci.

Meiosis Specialised cell division that reduces the chromosome number by half and leads to the formation of gametes.

Metaphase Second stage of mitosis, meiosis I and meiosis II.

***n*-fold degenerate codon** A position of a codon is said to be n-fold degenerate if n of the four nucleotides possibleat this position (A, T, C, G) end in the same amino acid (AA). By extension, a codon is said to be n-fold degenerate if n different three-nucleeotide sequences will code for the same AA.

N-terminus End of an amino acid chain terminated by a free amine group.

Non-crossover Recombination event without the exchange of flanking DNA sequences.

Nonself haplotype The haplotype that did not co-evolve with a given *Prdm9* allele.

Nonsynonymous substitution Substitution that does not modify the amino acid produced.

Outgroup Distantly related group of organisms that serves as the ancestral reference for the studied group (or ingroup).

Pedigree A family tree drawn with standard genetic symbols, showing inheritance patterns for specific phenotypic characters.

Phenotype The composite of observable traits.

Ploidy The number of complete sets of chromosomes (n) in a cell.

Polymerase chain reaction Molecular biology method used to make copies of a specific DNA fragment.

Polymorphic Which presents several forms. In other words: subject to inter-individual variability.

Post-meiotic segregation Segregation occuring after the end of meiosis, during the mitotic division (Figure 1.3).

Primer Short single-stranded nucleic acid used to initiate DNA synthesis.

Prophase First stage of mitosis, meiosis I and meiosis II.

Pseudogene, pseudogeneisation (pas utilise)

Purebred Bred from members of a recognised breed, strain, or kind without admixture of other blood over many generations.

Reciprocal cross Breeding experiment designed to test the role of parental sex on a given inheritance pattern.

Recombination hotspot Region of the genome with an elevated rate of recombination.

Recombination Exchange of DNA sequence information.

Self haplotype The haplotype that co-evolved with a given *Prdm9* allele.

Sister chromatids The two chromatids originating from the same chromosome (after a replication event).

Stationary GC-content (GC*) The GC-content that sequences would reach at equilibrium if patterns of substitution remained constant over time.

Symmetric hotspot Hotspot for which the two haplotypes are equally likely to host the double-strand break.

Synapsis Pairing of homologues.

Synonymous substitution Substitution that modifies the amino acid produced.

Telophase Fourth stage of mitosis, meiosis I and meiosis II.

Tetrad analysis Analysis of the four products (gametes) resulting from one single meiosis event.

Transition Mutation between two nucleotidic bases of the same family (purine or pyrimidine), i.e. either a A $\leftrightarrow$ G or a C $\leftrightarrow$ T mutation.

Transversion Mutation involving a change of nucleotidic family (from a purine to a pyrimidine or the other way round), i.e. either a A $\leftrightarrow$ C, a A $\leftrightarrow$ T, a G $\leftrightarrow$ C or a G $\leftrightarrow$ T mutation.

Variant-calling The process of identifying variant (a.k.a. polymorphic) sites on a genome.

ZMM complex A set of conserved yeast proteins Zip1, Zip2, Zip3, Zip4, Mer3, Msh4, Msh5 and Spo16 (a.k.a. synapsis initiation complex, SIC).

Preamble

While Charles Darwin (1809–1882) was advocating an evolutionary interpretation of vestigial structures[1] in his groundbreaking opus *On the Origin of Species* (1859), he drew a parallel between the work of linguists and that of evolutionary biologists:

> *'Rudimentary organs may be compared with the letters in a word, still retained in the spelling, but become useless in the pronunciation, but which serve as a clue in seeking for its derivation.'*

Nowadays, with the rise of sequencing technologies, the meaningfulness of his analogy is just as topical as ever: evolutionary biologists can now directly 'read' DNA and search for its 'etymology' by analysing the series of its 'letters'. Ultimately, their goal is to uncover the kinship ties between species, just like linguists would disclose the paths through which words have travelled by examining the remnants of unpronounced letters within them.

Indeed, the discovery of DNA in the mid-twentieth century (Franklin and Gosling, 1953; Watson and Crick, 1953; Wilkins et al., 1953) brought about a real revolution in the study of evolution and even led to the establishment of a new research field to which this thesis belongs: molecular evolution — now rather called evolutionary genomics for whole genomes, rather than single genes, get analysed. I will therefore open the introduction in Part I with Chapter 1 devoted to tracing back the scientific findings in genetics that directly led to the emergence of this research field aiming at understanding genome evolution.

---

[1]A vestigial structure is an anatomical feature or behaviour that has lost part or all of its initial function and that thus no longer seems to have a purpose in the current species. For instance, the human appendix and coccyx are two such vestigial organs.

But, precisely, why and how do genomes evolve? Three main evolutionary forces are classically invoked in this process: mutation, natural selection and genetic drift. Though, a couple of decades ago, a fourth force made an entrance in the evolutionary scene: biased gene conversion (BGC). This driver of genome evolution is a direct consequence of recombination — a process essential to meiotic cell division in sexually-reproducing organisms. I will thus review the mechanism of meiotic recombination in Chapter 2 and the sources of recombination rate variation in Chapter 3. This will lead me, in Chapter 4, to go over the knowledge acquired so far on the fourth evolutionary force of interest for this thesis.

From that point on, I will focus on the puzzling observation which laid the foundation for this work and will set, in Part II, the objectives we wanted to address.

The results presented in Part III will then be divided into four chapters. In Chapter 5, I will describe the unprecedentedly powerful approach we implemented to detect recombination events at high resolution in single individuals. Next, I will show how we used this method to precisely characterise mouse recombination patterns in Chapter 6 and to quantify biased gene conversion in Chapter 7. Last, in Chapter 8, I will detail how we adapted our method to other studies of recombination with more complex experimental designs involving several genomic introgressions. All the developments presented in this part are the result of a collaboration with Bernard de Massy and Frédéric Baudat, and those of Chapter 8 also involved Valérie Borde and Corinne Grey: the totality of the experimental work necessary for this study (mouse crosses and DNA extraction) was carried out by them. As for me, I contributed to this project by designing and implementing the bioinformatic procedures allowing to detect and quantify recombination and biased gene conversion and by analysing the ensuing results.

Finally, Part IV will be dedicated to discussing this work: I will first consider the scientific implications of our study in Chapter 9 and will then share ideas related to it in the broader fields of epistemology, philosophy of science and sociology of knowledge in Chapter 10.

Part I

Introduction

1

A geneticist's history of genetics

Contents

Grand scientific discoveries sometimes lead a research field to completely reorganise around new principles or axioms. This was the case with the comprehension of heredity. Up until the late nineteenth century, the inheritance of acquired characters — the idea that an organism can transmit features that it has acquired through use or disuse during its lifetime to its progeny — was a supposedly well-established fact that had been accepted by a plethora of philosophers and scientists, starting with Hippocrates (c. 460–c. 370 BC) (Zirkle, 1935). However, Mendel's pioneering work on hybridisation questioned the latter paradigm and shaked the scientific

community so well that it ended in the creation of a brand-new field in biology: genetics — which was first institutionalised in 1906 (Gayon, 2016).

In this chapter, I will review the main events of the genetics era that led to the concepts of recombination, gene conversion and genome evolution, which are of major interest for this thesis. A reader who is not familiar with the vocable of recombination (such as 'meiosis', 'gene conversion', 'post-meiotic segregation', 'interference', etc...) may find this chapter slightly difficult, as these denominations will not be fully detailed here. I therefore send them back to the definitions at the beginning of this thesis, or to the subsequent chapters of this introduction where the terms will be fully described, whenever they come across one of them.

The historical developments that one can appreciate are nothing but the result of what was transmitted to us by our predecessors and I therefore entitled this chapter *A geneticist's history of genetics* as a wink to what Richard Feynman (1918–1988), one of the most influential physicists of his time, wrote on this subject in his famous book on quantum physics *QED: The Strange Theory of Light and Matter* (2006):

> *'By the way, what I have just outlined is what I call a "physicist's history of physics," which is never correct. What I am telling you is a sort of conventionalized myth-story that the physicists tell to their students, and those students tell to their students, and is not necessarily related to the actual historical development, which I do not really know!'*

1.1   Emergence of a concept: recombination

1.1.1   An abstruse exception to Mendel's laws of heredity

Between 1857 and 1864, the Austrian monk Johann Gregor Mendel (1822–1884) undertook a series of hybridisation experiments on the garden pea plant *Pisum sativum*. This led him to describe the idea of an 'independent assortment of traits' (Mendel, 1865), thereby proving the existence of paired 'elementary units of heredity' (i.e. genes) and establishing the statistical laws governing them. His work remained unrecognised by the scientific community for several decades but was

finally rediscovered in the early twentieth century when three botanists (Hugo de Vries (1848–1935), Carl Correns (1864–1933) and Erich von Tschermak (1871–1962)) independently confirmed his findings (Dunn, 2003). Meanwhile, William Bateson (1861–1926) fiercely defended Mendel's thesis in *Mendel's Principles of Heredity: A Defense* (Bateson, 1902) against his contemporary biometricians (reviewed in Bateson, 2002), thus spreading Mendel's view into the scientific world.

A few years later, Bateson noticed exceptions to Mendel's principles of independent assortment: some crosses generated certain phenotypes in far excess from the expected Mendelian ratios (Bateson and Killby, 1905). This led him and his collaborators to propose that certain traits were somehow coupled with one another, although they did not know how (Bateson et al., 1905).

1.1.2 The chromosomal theory of inheritance

In the meantime, it had been understood that cells derived from other cells, but the exact process was unknown. To understand it, Walther Flemming (1843–1905) used stains to intensify the contrasts of cell contents observed through microscopy and identified a substance located within the nucleus, which he named 'chromatin' (from the Greek word $\chi\rho\widetilde{\omega}\mu\alpha$: 'color'). He described precisely the movements of chromosomes during cell division (which he termed 'mitosis'), thus providing a mechanism for the distribution of nuclear material into daughter cells during mitosis (Flemming, 1879).

Theodor Boveri (1862–1915) went one step further by demonstrating the individuality of chromosomes in the roundworm *Ascaris megalocephala*, which allowed him to suggest that the chromosomes of the germ cells are involved in heredity (Boveri, 1888). In addition, he showed that the egg and the spermatozoon contribute the same number of chromosomes to the new individual, thus providing the first descriptions of meiosis (Boveri, 1890). Walter Sutton (1877–1916) independently came to the same conclusion at about the same time: he enunciated the chromosomal

theory of inheritance with the following words closing his 1902 paper: 'I may finally call attention to the probability that the association of paternal and maternal chromosomes in pairs and their subsequent separation during the reducing division [. . .] may constitute the physical basis of the Mendelian law of heredity' (Sutton, 1902).

However, this theory was debated in the scientific community, because there was yet no direct proof of a link between the inheritance of traits and the segregation of chromosomes.

In parallel, based on cytological observations of chromosomes, Frans Janssens (1863–1924), a priest also known as the 'microscopy wizard' for he mastered the process, developed the idea that the chromosomes' 'filaments [chromatids] are involved in contacts that can modify their organisation from one segment to the next' which 'will generate new segmental combinations' in his *Chiasmatype Theory* (Janssens, 1909).

1.1.3   Morgan's theory of gene linkage and crossing-over

In 1909, Thomas Hunt Morgan (1856–1945) expressed his strong skepticism of the Mendelian theory of inheritance in his very derisive article *What are Factors in Mendelian Inheritance?* (Morgan, 1909) and doubted the chromosomal basis of heredity (reviewed in Koszul et al., 2012). Little did he know at the time that he was to become the main craftsman of the reconciliation of these two theories.

In his famous 'fly room' where he bred *Drosophila melanogaster* fruit flies, he found an unusual male white-eyed individual. Crossing it with purebred red-eyed females yielded red-eyed male and female F1 hybrids, — a typical result proving that the white eye color is a recessive trait. Unexpectedly, after inbreeding the heterozygous F1 progeny, he discovered that the traits of the F2 offspring did not assort independently: all white-eyed flies were males (Figure 1.1, left). However, when he crossed the white-eyed male with F1 daughters, he found both male

First cross

P

w^+ w^+

X X

Red female

×

w

X Y

White male

F$_1$ — Male gametes

w — $\frac{1}{2}$ — $\frac{1}{2}$

Female gametes

w^+

w^+ w — $\frac{1}{2}$ Red female

w^+ — $\frac{1}{2}$ Red male

F$_2$ — Male gametes

w^+ — $\frac{1}{2}$ — $\frac{1}{2}$

Female gametes

w^+ — $\frac{1}{2}$

w^+ w^+ — $\frac{1}{4}$ Red female

w^+ — $\frac{1}{4}$ Red male

w — $\frac{1}{2}$

w^+ w — $\frac{1}{4}$ Red female

w — $\frac{1}{4}$ White male

Second cross

P

w w

X X

White female

×

w^+

X Y

Red male

F$_1$ — Male gametes

w^+ — $\frac{1}{2}$ — $\frac{1}{2}$

Female gametes

w

w^+ w — $\frac{1}{2}$ Red female

w — $\frac{1}{2}$ White male

F$_2$ — Male gametes

w — $\frac{1}{2}$ — $\frac{1}{2}$

Female gametes

w^+ — $\frac{1}{2}$

w^+ w — $\frac{1}{4}$ Red female

w^+ — $\frac{1}{4}$ Red male

w — $\frac{1}{2}$

w w — $\frac{1}{4}$ White female

w — $\frac{1}{4}$ White male

Figure 1.1: Reciprocal crosses between red-eyed (red) and white-eyed (white)
Drosophila.
In the first cross (left), a red-eyed purebred female is crossed with a white-eyed male, resulting in F1 hybrids made of heterozygous red-eyed females bearing both the dominant (w^+) and the recessive (w) alleles and red-eyed males bearing only the dominant (w^+) allele. The inbreeding of F1 individuals results in a F2 generation with a 3:1 ratio of red-eyed:white-eyed individuals, all white-eyed individuals being males.
In the second cross (right), a white-eyed female is crossed with a red-eyed purebred male, resulting in F1 hybrids made of heterozygous red-eyed females bearing both the dominant (w^+) and the recessive (w) alleles and white-eyed males bearing only the recessive (w) allele. The inbreeding of F1 individuals results in a F2 generation with a 2:2 ratio of red-eyed:white-eyed individuals, half of white-eyed being males and half being females.
The results of these two crosses show that the gene coding for eye color is located on the female sexual chrosome (X). The fact that results in the F2 progeny differ according to the direction of the cross ($(\frac{w^+}{w^+}) \times$ (w) or $(\frac{w}{w}) \times (w^+)$) is a typical signature of linkage disequilibrium between the observed trait (eye color) and the sex chromosomes.
This figure was reproduced from Griffiths et al. (2015) (permission in Appendix B).

and female white-eyed flies (Figure 1.1, right), thus showing that the white eye color was not lethal for females.

He immediately hypothesised that eye color was connected to the sex determinant (Morgan, 1910) and, as these findings were consistent with the idea that genes were physical objects located on chromosomes, Morgan soon came up with the idea of genetic linkage, i.e. the fact that two genes closely associated on a chromosome do not assort independently (Morgan, 1911). He also suggested that this coupling dependended on the distance between genes: 'we find coupling in certain characters, and little or no evidence at all of coupling in other characters; the difference depending on the linear distance apart of the chromosomal material that represent the factors.'

With three of his students (Alfred Sturtevant (1891–1970), Hermann Muller (1890–1967) and Calvin Bridges (1889–1938)), he summarised all the evidence in *The Mechanism of Mendelian Heredity* which constitutes one of the most important books in the whole history of genetics (Gayon, 2016). There were two major propositions in that book.

First, the recognition that Mendelian factors — Morgan would soon call them 'genes' — are physical portions of chromosomes. This brought a mechanistic support to Mendel's 'law of segregation' (according to which the zygote inherits only one version of each gene from each parent) and to the so far unexplained exception to Mendel's 'law of independent assortment fo traits': when two genes are located on the same chromosome, they have to segregate together — and thus the law does not apply to this special case.

Second, they proposed that the linkage between genes located on the same chromosome could sometimes break, through the process of what Morgan called 'crossing-over' (Figure 1.2). This was to take place at the positions of the chiasmata previously observed by Janssens (Janssens, 1909). Later, Edgar Wilson (1908–1992) and Morgan crafted structures of crossing-overs with clay to materialise how the

crossing-over could physically form (Wilson and Morgan, 1920).

Altogether, with the ideas of recombination and crossing-over, Morgan had fused three theories: gene linkage (the major exception to Mendel's laws of heredity), the chromosomal theory of inheritance and the chiasmatype theory. This triggered a real revolution in biology and marked the commencement of genetics. His major contribution through his work on *Drosophila* won him the *Nobel Prize in Physiology or Medicine* in 1933.

It was only ten years later that Harriet Creighton (1909–2004) and Barbara McClintock (1902–1992) would bring the first proof of that theory by correlating cytological and genetic exchanges in maize (Creighton and McClintock, 1931).
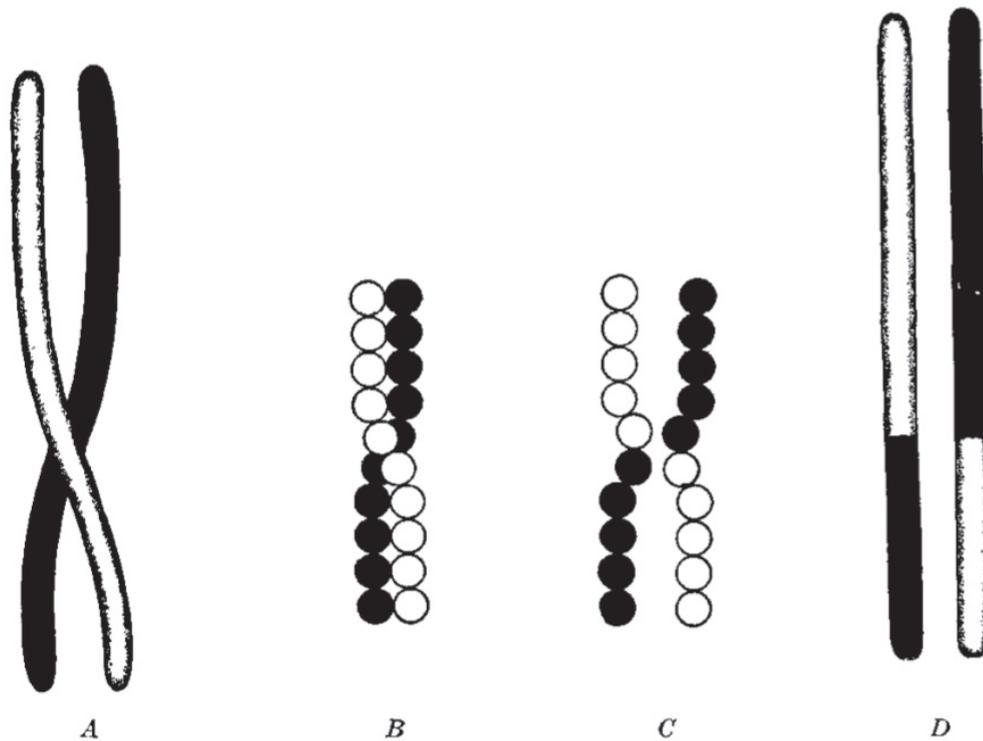
Figure 1.2: Original drawing of crossing over in *The Mechanism of Mendelian Heredity* (Morgan et al., 1915).
Original legend by the authors: 'At the level where the black and the white rod cross in A, they fuse and unite as shown in D. The details of the crossing over are shown in B and C.' This drawing symbolises the reconciliation between Mendel's and the chromosomal theories of inheritance.
This figure was reproduced from Morgan et al. (1915) (permission in Appendix B).

1.2 Emergence of a concept: gene conversion

1.2.1 The study of fungal products of meiosis

The next major advances on the comprehension of the recombination mechanism were to come through the study of fungi, soon adopted as model organisms for the multiple advantages they confer to genetics reseach. First, as they take up little space and are easy and cheap to propagate, they can be studied in very large numbers.

Second, it was reported early that they alternate haploid[1] and diploid[2] phases. Indeed, the Czech scientist Jan Šatava (1878–1938) managed to isolate the ascospores[3] of a yeast and saw that they germinated without fusing other ascopores, thus giving rise to haploid cultures (Šatava (1918), reviewed in Barnett, 2007). This feature, — haploidy of the progeny, — considerably facilitates the interpretation of the products of meiosis since the phenotype of each offspring is a direct manifestation of its genotype (contrary to diploid or higher-order of ploidy cases for which dominance and recessiveness may blur gene expression).

Third, in some fungi, the cells corresponding to the four products of meiosis remain grouped in a tetrad of four sexual spores, which makes the direct observation of a single meiosis possible. The first study of this type, — a 'tetrad analysis', — was achieved by Øjvind Winge (1886–1964), the founder of yeast genetics (Winge and Laustsen, 1937). In some ascomycetes, the meiotic products undergo one additional mitotic division, thus ending in 'octads' of four pairs of identical spores (Figure 1.3).

Last, in certain fungi, the spindles of the meiotic (and mitotic, if applicable) divisions are constrained in a tube-shaped ascus preventing them from overlapping, which leads the tetrads (or octads) to arrange linearly, and makes the interpretation of the behaviour of genes during meiosis (and mitosis) straightforward (Figure 1.3) (Casselton and Zolan, 2002).

---

[1]Single set of chromosomes
[2]Two sets of chromosomes
[3]Reproductive cells of a certain class of fungi (ascomycetes)
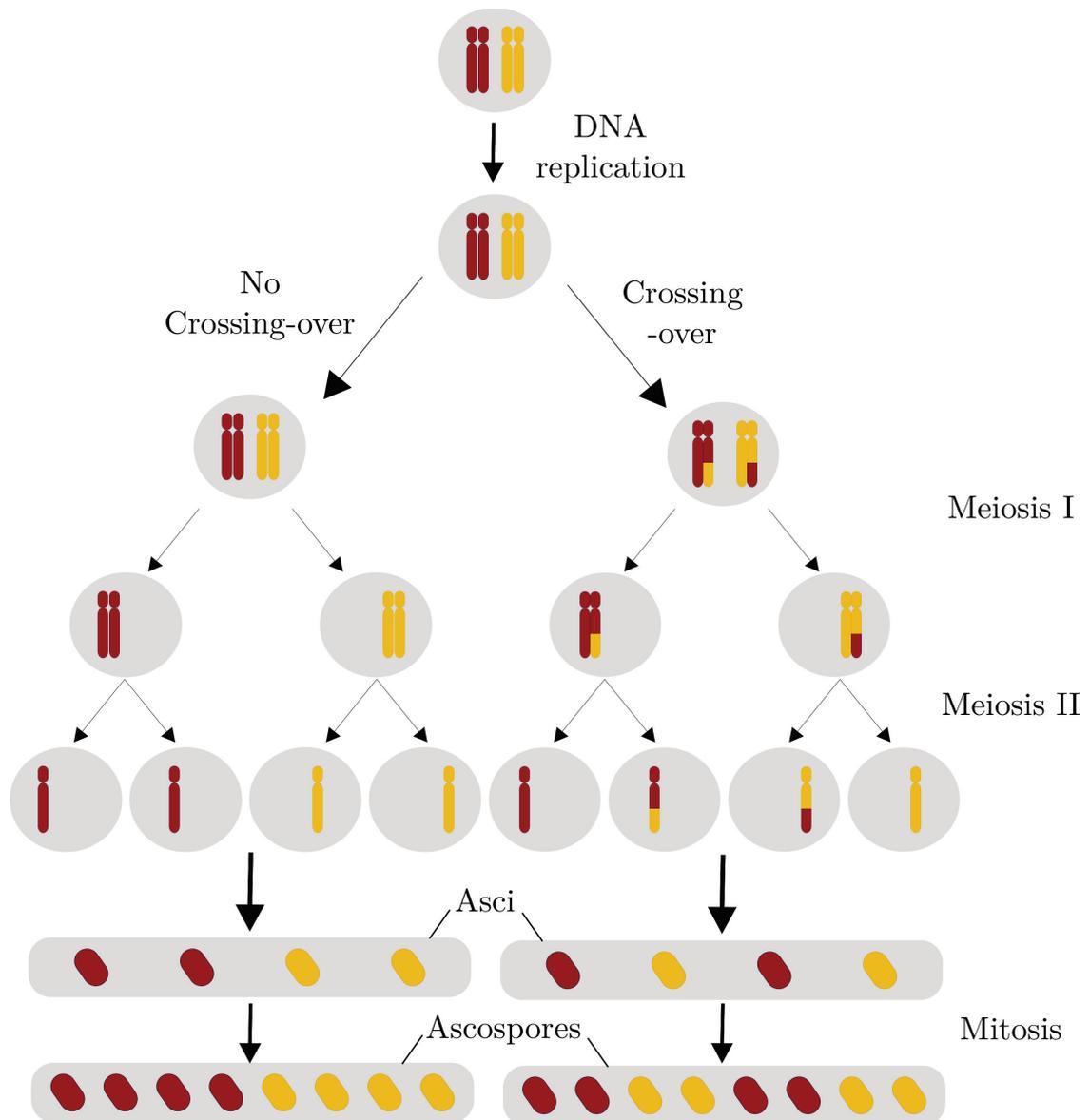
Figure 1.3: Meiotic and post-meiotic mitotic segragations of chromosomes in a linear ascomycete tetrad.
During the first meiotic segregation (Meiosis I), the homologous chromosomes either segregate with the occurrence of a crossing-over (right) or not (left). In absence of a crossing-over, the markers segregate at different nuclei at the end of first meiotic division and this results in ascopores displaying a sequence of four times the paternal allele and four times the maternal allele. In presence of a crossing-over, the markers segregate at different nuclei only at the end of the second meiotic division, which thus results in ascospores displaying an alternance of two times the paternal allele and two times the maternal allele.

All these attributes and technical achievements rendered fungi superior model organisms for the study of recombination. And there began the dawn of the fungal genetics era.

1.2.2   Four novel phenomena associated to recombination

Gene conversion

Using them, Hans Winkler (1877–1945) observed 3+:1- and 1+:3- departures from the expected Mendelian segragation among tetrads of +/- diploids (Winkler (1930), reviewed in Roman, 1985), which meant that the information present on one chromatid was replaced by that from another chromatid (Orr-Weaver and Szostak, 1985).

This observation was later confirmed by Carl Lindegren (1896–1987), a former student of Morgan's, who obtained similar irregular ratios with frequencies of about 1% in the budding yeast *Saccharomyces cerevisiae* (Lindegren, 1953) as well as by Mary Mitchell (fl. 1950–1965) who found 2:6 segregations[4] of wild-type:recessive phenotypes in *Neurospora* (Mitchell, 1955a,b).

Originally, Winkler had hypothesised that a mutational mechanism was at the origin of this replacement and invented the term 'gene conversion' to describe it. Although his interpretation turned out to be wrong (the mechanism is in fact purely recombinational, not mutational) and some authors suggested alternative nomenclature for it (e.g. Roman, 1986), the term he had come up with persisted over the years and is still used today.

Post-meiotic segregation

Soon after, Lindsay Olive (1917–1988) observed another type of aberrant segregation in the octads of *Sordaria fimicola*: 5:3 segregation ratios (Figure 1.4b) (Olive, 1959;

---

[4]A 2:6 segregation in the eight-spored *Neurospora* ascus is equivalent to a 1:3 segregation in the four-spored ascus of the budding yeast *Saccharomyces cerevisiae*.
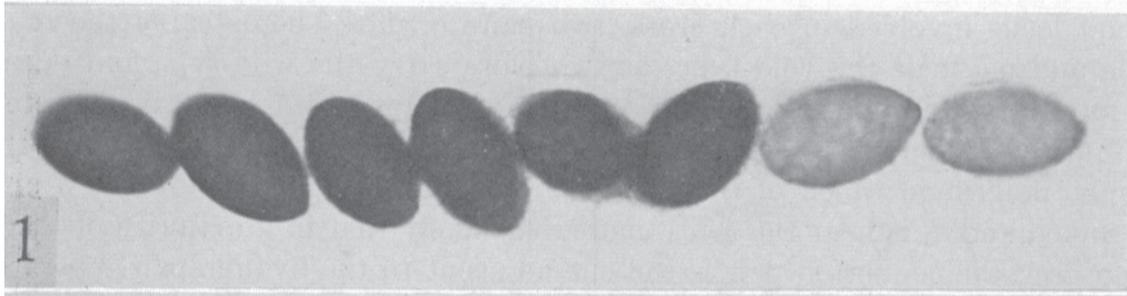
(a) Ascus containing 6 black and 2 white ascospores: gene conversion.
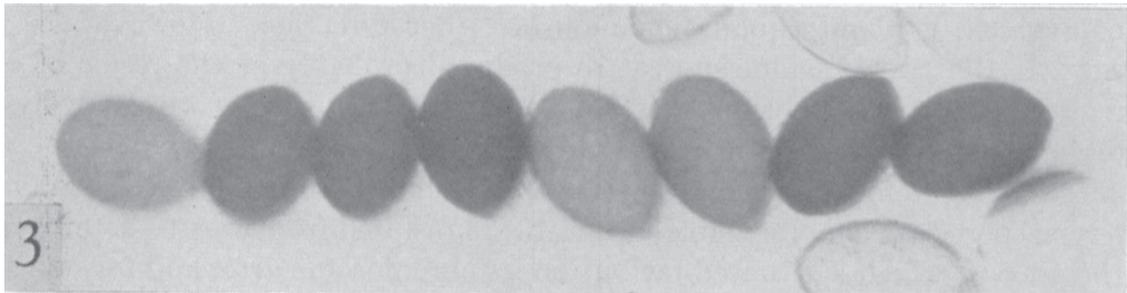
(b) Ascus containing 5 black and 3 white ascospores: post-meiotic segregation.

Figure 1.4: Original photographs of aberrant octads in *Sordaria fimicola*.
This figure was reproduced from Olive (1959) (permission in Appendix B).

Kitani et al., 1962). This result was puzzling, since it was not congruent with the models so far: 6:2 segregations were explainable on the basis of a non-directional transfer of information from one chromatid to another one, but this sole explanation could not account for the 5:3 segregation ratios. However, these results were totally reconciliable with the concept of a chromatid composed of two functional subunits, which had been proposed after autoradiographic studies on DNA (Taylor et al., 1957) in accordance with the Watson-Crick model of DNA (Watson and Crick, 1953).

This feature was again observed in *Neurospora crassa* concomitantly with the finding that several alleles were converted concertedly (Case and Giles, 1964). Such co-conversion of alleles was also found in *S. cerevisiae*, together with the finding that the frequency of co-conversion decreases with increasing distance between the alleles (Fogel and Mortimer (1969), reviewed in Orr-Weaver and Szostak, 1985).

Altogether, these findings indicated the presence of 'heteroduplex DNA', i.e. a DNA portion where the two strands composing it contain different information for the segregating marker. Such heteroduplex DNA cannot be detected genetically

until an additional round of DNA replication produces two duplexes, each expressing the information from one of the strands of the heteroduplex. These segregations, occuring after the end of meiosis, are called 'post-meiotic segregations' (PMS). The additional observation that markers are co-converted at frequencies dependent on their distance suggested that heteroduplex DNA (and thus, gene conversion) could span hundreds of nucleotides (Orr-Weaver and Szostak, 1985).

Conversion polarity

In addition, it was found that gene conversion frequencies vary linearly from one end of a gene to the other (reviewed in Nicolas and Petes, 1994): this discovery was made in both *Ascobolus immeraus* (Lissouba and Rizet, 1960; Lissouba et al., 1962) and in *Neurospora crassa* (Murray et al., 1960) at approximately the same time. This phenomenon was observed again in *Aspergillus nidulans* (Siddiqi, 1962) and in other mutants of *Neurospora* (Stadler and Towe, 1963), and was designated as 'conversion polarity' or 'polarised recombination'.

Later, one of its discoverers, Lady Noreen Murray (1935–2011) demonstrated that this polarity was due to elements located close to the gene, as opposed to being imposed by the orientation of the gene with respect to the centromere (Murray, 1968). This led to the idea that recombination initiates on 'pseudofixed sites', the erstwhile concept for what we now call 'recombination hotspots'.

Interference

One last important observation made during this decade came from a study on *Aspergillus nidulans* (Pritchard, 1955). The authors looked at four linked marker genes, whose recessive alleles will here be designated as 'y', '11', '8' and 'bi', and whose dominant alleles will here be designated as '+' in all four cases. They crossed a strain of genotype (y+8+) with a strain of genotype (+11+bi) to obtain a F1 hybrid of genotype $(\frac{y}{+} \frac{+}{11} \frac{8}{+} \frac{+}{bi})$ and found that the largest proportion of recombinants from this hybrid was of genotype (y++11), while all other combinations ((y+++), (+++bi) and (++++)) were under-represented

(reviewed in Whitehouse, 1965). Similar observations of this phenomenon were made in *Neurospora crassa* (Mitchell, 1956).

These findings suggested that recombination between alleles (in this case, between the second and third marker) are negatively associated with recombination in neighbouring regions (in this case, between the first and second, and between the third and fourth markers). This feature was designated as 'interference'.

1.2.3   The first theories on the recombinational mechanism

To sum up, over the course of the 1950's and of the early 1960's, numerous studies evidenced that crossing-over was associated with gene conversion, PMS, polarised recombination and interference.

It was soon proposed that all these processes were somehow mechanistically linked (Perkins, 1962) and from that point on, several scientists conjectured theories reuniting these observations. One important one, the 'copy-choice hypothesis', was postulated by Joshua Lederberg (1925–2008) (Lederberg, 1955). According to this (wrong) theory, the process of replication switches from copying one parental chromosome to the other — the switch occuring when both chromosomes are closely paired. An alternative hypothesis, 'the hybrid DNA hypothesis', was proposed (Whitehouse, 1963), allegedly inspired from the model of Robin Holliday (1932–2014) (Holliday, 2011) which the latter would publish the following year (Holliday, 1964).

The Holliday model (Holliday, 1964, 1968), which was in accordance with the then recent discovery of the double-stranded structure of DNA (Franklin and Gosling, 1953; Watson and Crick, 1953; Wilkins et al., 1953), happened to be the first widely accepted molecular explanation for the phenomena with which crossing-over had been found to be associated, namely aberrant segregation (i.e. gene conversion and PMS) and polarised recombination. Briefly, this model rested on the formation of two concomitant DNA breaks, the separation of the two DNA strands followed by base pairing between the complementary segments to form symmetric heteroduplex

DNA and the so-called 'Holliday junction', and last, the resolution of this junction by cutting either the originally crossed or the non-crossed strands.

Over the following two decades, the Holliday model was meticulously tested and revised (reviewed in Haber, 2008) and several other models were formulated to account for novel experimental observations. Notably, Matthew Meselson (born 1930) and Charles Radding (born 1946) proposed one according to which a Holliday structure would be generated by a single-strand nick in only one chromosome (Meselson and Radding, 1975). A few years later, their model was supplanted by one that is still used today: the double-strand break repair (DSBR) model (Szostak et al., 1983). According to the latter, recombination is initiated by a DNA double-strand break (DSB) on one chromosome and the resulting strand exchange leads to the formation of a double-Holliday junction (dHJ). This model, as well as all the other recombinational models used today, will be detailed in Chapter 2.

In addition to all these advances on the mechanistical aspects of heredity, the early twentieth century was marked by theoretical breakthroughs in the study of evolution, which I review in the upcoming section.

1.3 Emergence of a concept: genome evolution

1.3.1 The dawn of population genetics

Soon after the rediscovery of Mendel's laws of inheritance, a fierce debate opposed two groups of biologists: Mendelians who believed that evolution was driven by mutations transmitted by the discrete segregation of alleles (Bowler, 2003), and biometricians who claimed that variation was continuous. The first group, led by Bateson and de Vries, maintained that the variations measured by biometricians were too small to account for evolution while the second, led by Karl Pearson (1857–1936) and Walter Weldon (1860–1906), rejected Mendelian genetics on the basis that it would necessarily imply discontinuous evolutionary leaps (Provine, 2001).

It was only fifteen years later that the British statistician Ronald A. Fisher (1890–1962) reconciled both theories, first by proving mathematically that mutiple discrete loci could result in a continuous variation (Fisher, 1919) and then by showing in subsequent papers and in his book *The Genetical Theory of Natural Selection* (1930) that natural selection could change allele frequencies in a population and result in evolution. Soon after, in a series of ten papers named *A Mathematical Theory of Natural and Artificial Selection* (1927), another British geneticist — John B. S. Haldane (1892–1964) — derived equations of allele frequency change at a single locus under a broad range of conditions. This allowed him to re-establish natural selection as the major cause of evolution (Haldane, 1932). The contributions of the two of them, — together with that of Sewall Wright (1889–1988), a geneticist living across the Atlantic who worked out the mathematics for combinations of interacting genes, — laid the foundations for population genetics, a discipline which basically integrated Mendelism, Darwinism and biometry.

The emergence of this new field of study was the first step towards the development of a unified theory of evolution named the 'modern synthesis' (Huxley, 1942). Its founders — Theodosius Dobzhansky (1900–1975), George Ledyard Stebbins Jr. (1906–2000) and Ernst Mayr (1904–2005) — all defined it on the basis of natural selection acting on the heritable variation supplied by mutations (Mayr, 1959; Stebbins, 1966; Dobzhansky, 1974). But the exclusive contribution of this adaptive process to genome evolution was soon to be contested.

1.3.2 Neutralists *versus* selectionists: a conflictual story

One of Wright's main contributions to population genetics was the introduction of the concept of 'adaptive landscapes' according to which phenomena other than natural selection, — like genetic drift and inbreeding, — could push small populations away from adaptive peaks, thus propelling, in turn, natural selection to drive them towards different adaptive peaks (Wright, 1932). As such, the relative contributions

of neutral forces (like genetic drift) and adaptive forces (like natural selection) became a major subject of debate between Wright and Fisher (Plutynski, 2007).

But this controversy really intensified after Motoo Kimura (1924–1994) proposed the neutral theory of molecular evolution (Kimura, 1968, 1991; Kimura et al., 1986) and Tomoko Ohta (born 1933) adapted it as the nearly neutral theory (Ohta, 1973). For selectionists, most mutations are either beneficial or harmful and are thus either retained or purged by the action of natural selection, whereas supporters of the neutral theory claim that most mutations are adaptively neutral and thus become fixed in populations through the cumulative effect of sampling drift (Lewin, 1996).

As of today, it is widely accepted that both genetic drift and natural selection participate in the evolution of genomes: the controversy is no longer strictly dichotomous but rather concerns the quantitative contributions of adaptive and of non-adaptive evolutionary processes. Though, distinguishing between both types of processes may not be that simple, for selection also has important indirect effects directly due to the process of recombination, as detailed in the next subsection.

1.3.3 Recombination in the context of genome evolution

At approximately the same time, scientists suggested that other evolutionary processes may be linked to recombination (Maynard Smith and Haigh, 1974): theoretically, a gene undergoing a selective sweep could result in allele frequency changes of the loci in its vicinity, thus resulting in a local decrease of polymorphism. This phenomenon — later known as genetic hitchhiking or background selection depending on the direction of selection — was then empirically demonstrated by Begun and Aquadro (1992) when they put to light an apparent correlation between the level of genetic diversity and the recombination rate in flies.

As such, it became obvious that recombination plays a major role in genome evolution and that it should, in no case, be overlooked. But, to understand precisely the extent of its contribution to evolution, it is necessary to know more about its mechanisms: this will be reviewed in the following chapter.

2

Meiotic recombination, the essence of heredity

Contents

'Why all this silly rigmarole of sex? Why this gavotte of chromosomes? Why all these useless males, this striving and wasteful bloodshed, these grotesque horns, colors. . . and why, in the end, novels, like Cancer Ward, about love?'

— W. D. Hamilton, *Review of Ghiselin (1974) and Williams (1975)* (1975)

This is how the fanciful Bill Hamilton (1936—2000) sums up the mystery of sexual

reproduction (or simply, 'sex') that has been puzzling biologists for over a century and which, to this day, remains unanswered (de Visser and Elena, 2007; Otto, 2009).

This so-called 'paradox of sex' finds its roots in that most theoretical arguments plead an elevated cost of sex as compared to asexual modes of reproduction (Otto and Lenormand, 2002; Lehtonen et al., 2012). First, females invest half their reproductive resources in the production of males which, in turn, invest minimally into the progeny, as epitomised by the uncommonness of paternal care when it is not beneficial to the male (Maynard Smith, 1977; Fromhage et al., 2007) — a concept known as the 'twofold cost of sex' or 'cost of meiosis' (Bell, 1982). Second, the sexual act itself wastes time and energy to find and attract a sexual partner, and exposes the individual to the risks of contracting diseases and of being predated (sometimes by the mate itself), thus making sex a pearilous and unprofitable endeavour.

Nevertheless, only 80 (Vrijenhoek et al., 1989; Neaves and Baumann, 2011) of the 70,000 vertebrate species discovered so far (IUCN (International Union for Conservation of Nature), 2019) and as little as 0.1% of all named animals (Vrijenhoek, 1998) reproduce otherwise than sexually. Such pervasiveness of sex in nature constitutes indisputable proof of its evolutionary success.

But, given its considerable drawbacks, how come sex has superseded all other forms of reproduction? Over 20 theories have been put forward to answer this question (Kondrashov, 1993), but the most generally claimed advantages revolve around the idea that sex both eliminates deleterious mutations and brings up more favourable combinations of alleles (Normarck et al., 2003; Speijer, 2016). This defensibly profitable reshuffling of alleles is called 'recombination' and occurs during meiosis, the cellular process leading to the formation of gametes.

This chapter — named after a review on the subject (Hunter, 2015) — explores the cytological features of meiosis and the mechanistic principles of homologous recombination (HR), before venturing into the body of molecular actors enacting in this complex process and the reasons why their performance is critical for heredity.

2.1 Meiosis in the context of gametogenesis

2.1.1 A two-step division process to form gametes

Most sexually-reproducing organisms have diploid cells, i.e. cells counting two sets of chromosomes: one from each parent. The transmission of half this genetic material to the progeny goes through the formation of specialised haploid cells (i.e. cells encompassing a single set of chromosomes) called 'gametes'. Such transition from diploidy to haploidy occurs during a particular type of cell division called 'meiosis' (from the Greek word $\mu\varepsilon\acute{\iota}\omega\sigma\iota\varsigma$: 'lessening'). The evolutionary origin of meiosis is still a mystery (Lenormand et al., 2016) but its wide occurrence in eukaryotes suggests that their last common ancestor had already acquired it (Cavalier-Smith, 2002; Ramesh et al., 2005; Speijer et al., 2015) through a process that is still largely debated (Wilkins and Holliday, 2009; Bernstein and Bernstein, 2010; Bernstein et al., 2011). Despite its somewhat blurry origins, its cytological features are conserved.

Concretely, meiosis is preceded by a unique round of chromosome duplication occurring during the interphase of diploid germinal cells (ovocytes in females and spermatocytes in males). Thence, before entering meiosis, each homologous chromosome (or 'homologue') i.e. each parental copy, is formed of two identical double-helix DNA molecules called 'sister chromatids' which are physically attached at a point called the 'centromere'[1] and adjoined along their whole length by cohesins (Klein et al., 1999). Therefrom, the two successive cell divisions that compose meiosis will result in the distribution of the chromatids into four gametes.

The first meiotic division is also known as the 'reductional division' because it reduces ploidy by setting apart the homologues of each pair. It is classically divided into four stages: prophase, metaphase, anaphase and telophase (Figure 2.1, top). Prophase I, described more extensively in Subsection 2.1.2, stages the pairing of homologous chromosomes along with recombination. Next, the meiotic spindle bonds

---

[1]Except for species with holocentric chromosomes (i.e. chromosomes devoid of any major centromeric constriction), like Lepidoptera, aphids and nematodes.

Figure 2.1: Chromosome organisation diagrams and cytological pictures of the two meiotic divisions leading to the formation of four gametes.
During the first meiotic division (top), homologous chromosomes condense and pair up *via* chiasmata at prophase I, line up on the equatorial plate at metaphase I, segregate at anaphase I and decondense at telophase I, thus giving two secondary gametocytes. During the second meiotic division (bottom), the sister chromatids of the two secondary gametocytes go through the same steps. Chromosome segments of the same colour (blue or pink) come from the same parental origin and centrosomes are coloured in green. The micrographs show meiosis in a male lily while the diagrams show it in an animal cell. This figure was reproduced from Hillis et al. (2012) (permission in Appendix B).

the paired homologues and lines them up on the equatorial plate during metaphase I, before separating them during anaphase I. Such partition of homologues is achieved thanks to the existence of two opposite forces that stabilise the chromosomes until they are correctly oriented: first, the chiasmata that maintain the homologues attached and second, the meiotic spindle that creates a poleward tension (Petronczki et al., 2003). The co-segregation of sister chromatids is likely due to a physical jointure of their kinetochores (Nasmyth, 2015). Segregation *per se* terminates at telophase I during which the chromosomes decondense and a nuclear enveloppe (NE) forms around the nuclei. At the end of the first meiotic division, each of the two haploid daughter cells ('secondary gametocytes') contains one pair of sister chromatids corresponding either to the paternal or to the maternal homologue.

Following a short interkinesis during which DNA does not replicate, the second meiotic division splits sister chromatids in a manner much similar to a haploid mitosis. This division is termed 'equational' because the number of chromosomes stays equal before and after it. Like the first one, it is partitioned into four stages (Figure 2.1, bottom) executed synchroneously in the two secondary gametocytes. During prophase II, the NEs break down and the chromatids recondense. In the meantime, the centrosomes duplicated during interkinesis move towards opposite poles while a new meiotic spindle forms in between and starts to capture chromatids. The single chromosomes line up across the equational plates of each cell during metaphase II and sister chromatids segregate towards opposing poles during anaphase II. At telophase II, the chromosomes begin to decondense and new NEs form around them, thus producing the final set of four genetically-unique haploid gametes.
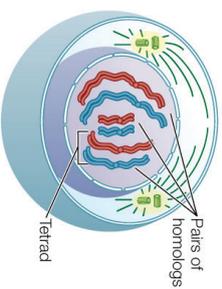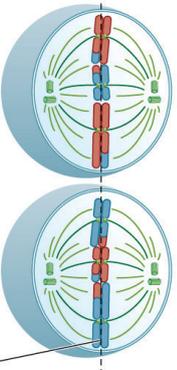
Albeit these general features of meiosis are shared, its timing and the products it forges are sexually dimorphic in mammals (reviewed in Handel and Schimenti, 2010). Indeed, male meiosis forms four gametes (spermatids) whereas female meiosis ends in a single functional gamete and three non-functional haploid cells called 'polar bodies'. As for the timing, spermatogonia mature into spermatocytes which initiate meiosis all along male adulthood, thus ensuring a continuous production

of sperm. In contrast, the common conception in females is that the integrality of oogonia mature into ovocytes during fetal development (Pearl and Schoppe, 1921; Zuckerman, 1951), even though recent findings suggest that oocyte production may be sustained in postnatal ovaries (Johnson et al., 2004, 2005). In any case, female meiotic prophase I — initiated and arrested right after the production of ovocytes — is resumed in small batches of ovocytes at periodic intervals during the reproductive lifespan. It halts once again at metaphase II, until fertilisation by a spermatozoid (if it ever occurs) triggers the completion of the process.

While the transition from plain cell cycle to meiotic entry is managed by a complex body of checkpoints (reviewed in Marston and Amon, 2005), the metronomic completion of meiotic subprocesses is abundantly warranted by the capacity of chromosomes to respond to cell cycle controls (reviewed in McKim and Hawley, 1995). But the most regulated — and perhaps most critical — meiotic step is the synapsis of homologous chromosomes which takes place during prophase I.

2.1.2   The synapsis of homologues during prophase I

Four differential degrees of synapsis

Prophase I is commonly subdivided into four stages (Figure 2.2): leptotene (or leptonema), zygotene (or zygonema), pachytene (or pachynema) and diplotene (or diplonema). Each is characterised by a particular chromosomal configuration that mirrors their degree of 'synapsis' i.e. pairing of homologues.

At leptotene, chromosome ends connect the cytoskeleton located outside the nucleus (Scherthan et al., 1996) *via* their binding a complex body of SUN-domain proteins of the inner nuclear membrane (INM) that have beforehand bridged KASH-domain proteins of the outer nuclear membrane (ONM) (Tzur et al., 2006; Yanowitz, 2010). This allows cytoplasmic forces to animate the motion of chromosome ends at the surface of the INM (Penkner et al., 2009) and ends at late leptotene by the

formation of a 'bouquet' (Figure 2.3) (Zickler and Kleckner, 1998) which constrains the chromosomes to a limited nuclear area (Zickler, 2006).

At zygotene, the homologous chromosomes begin to synapse, starting with the telomeric regions tethered in the bouquet (Pfeifer et al., 2003). By the end of pachytene, synapsis is complete for all pairs of chromosomes, with the notable exception of the non-homologous male X and Y chromosomes. Instead, the sex chromosomes are transcriptionnally inactivated ('meiotic sex chromosome inactivation': MSCI) by remodelling into heterochromatin (Fernandez-Capetillo et al., 2003) and are pushed to the periphery of the nucleus where they form the 'sex body' (Handel, 2004) (Figure 2.2.e.). Then, during diplotene, the homologous chromosomes desynapse but remain attached in pairs *via* their chiasmata.
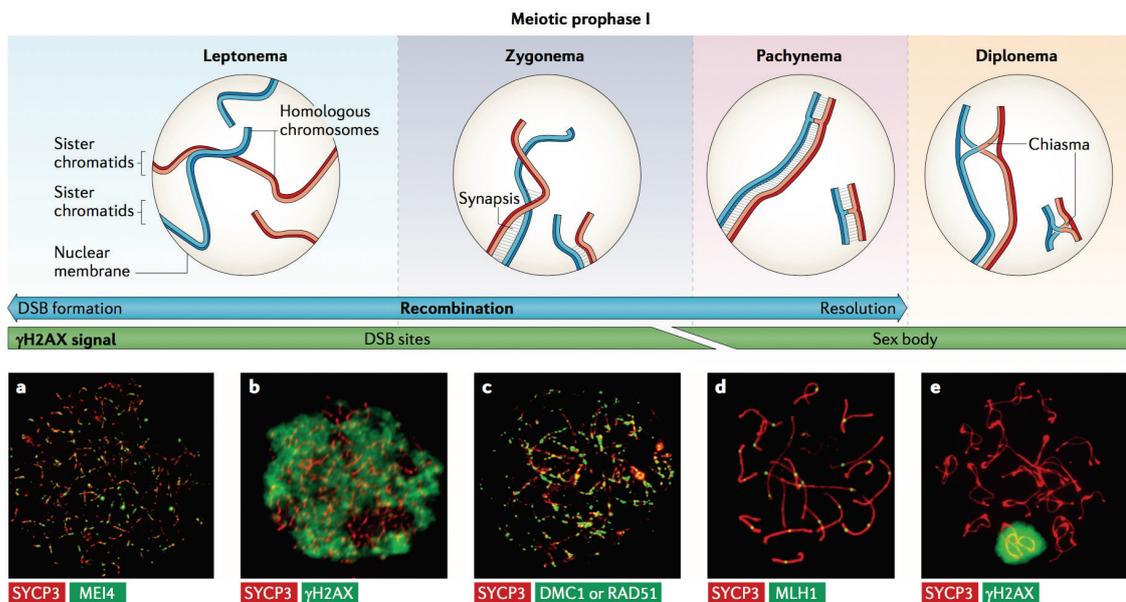
Figure 2.2: Chromosome organisation and cytology during prophase I.
Top: Two pairs of duplicated homologous chromosomes (red and blue) display different configurations in the four substages of meiotic prophase I. Double-strand break (DSB) formation at leptotene triggers both synapsis and the DSB resolution materialising as chiasmata during zygotene. Synapsis is completed at the onset of pachytene. Diplotene stages desynapsis, with homologues held together *via* chiasmata.
Bottom: immunofluorescence staining of synaptonemal complex protein 3 (SYCP3) and stage-specific signals on mouse spermatocyte spreads. **a |** Meiosis-specific MEI4-homologue (MEI4) colocalises with the synaptonemal complex (SC). **b |** H2AX is phosphorylated (γH2AX) following DSB formation. **c |** DNA recombinases DMC1 and RAD51 localise at DSB repair sites. **d |** MutL protein homologue 1 (MLH1) localises at DSB sites repaired as COs. **e |** Unrepaired DSB sites in the sex body are marked by γH2AX.
This figure was reproduced from Baudat et al. (2013) (permission in Appendix B).

Presynaptic pairing

Matching homologous chromosomes into pairs constitutes the most critical event of synapsis. This challenge is colossal: for human cells, it compares to finding a 20-cm stretch — other than the sister chromatid — throughout the London-Moscow distance, simulaneously for hundreds of sites and coordinately with higher-order cellular processes (Neale and Keeney, 2006).

This search is likely facilitated by the establishment of pre-meiotic physical contacts between homologues (reviewed in McKee, 2004; Zickler, 2006). Such presynaptic pairing was evidenced in mice (Boateng et al., 2013; Ishiguro et al., 2014) and, although its mechanism remains unknown, several theories wrestle to explain it.

According to one of them, presynaptic associations may occur through DNA-DNA duplexes (Danilowicz et al., 2009). This assumption relies on the observation that meiotic chromosomes pair only when they are transcriptionally active (Cook, 1997). DNA duplexes could thus momentarily form within the 'transcription factory' to which DNA loops are attached (Xu and Cook, 2008). Alternatively, these associations may be promoted by sequence-specific RNA molecules, in a manner similar to gene silencing in plants and fungi (Bender, 2004, cited in Zickler, 2006). A third scenario suggests a mechanism analogous to the 'pairing centres' (PC) or 'homologue recognition regions' (HRR) described in *Caenorhabditis elegans* (Villeneuve, 1994; MacQueen et al., 2005), *Drosophila melanogaster* (McKee, 1996) and *Saccharomyces cerevisiae* (Kemp et al., 2004). Namely, the *cis*-acting PCs (or HRRs) could initiate interactions between homologues (Gerton and Hawley, 2005).

In any case, demonstrating the existence of such presynaptic pairing in mice has driven Boateng et al. (2013) to propose a new model for homology search (Figure 2.3). With it, they challenge the commonly accepted view that homology search is triggered by the need to repair newly-formed DNA double-strand breaks (DSBs). Instead, they propose that DSBs occur after the pre-leptotene pairing and that their repair serves as a prophase checkpoint to proofread the initial connection.

If that were so, homology search for DSB repair would be restrained to a reduced territory and thus, much facilitated (Barzel and Kupiec, 2008; Mirny, 2011).

Whether or not this view is correct, chromosomal movements allow random collisions between chromosomes (Fung et al., 1998), thus creating opportunities for homologues to encounter and, more importantly, to disrupt unwanted (non-homologous) associations (Koszul and Kleckner, 2009). Yet, at this stage, the interstitial interactions between homologues are transient and reversible (Boateng et al., 2013). They thus need to be strengthened by a higher-order chromosomal structure: the synaptonemal complex (SC).
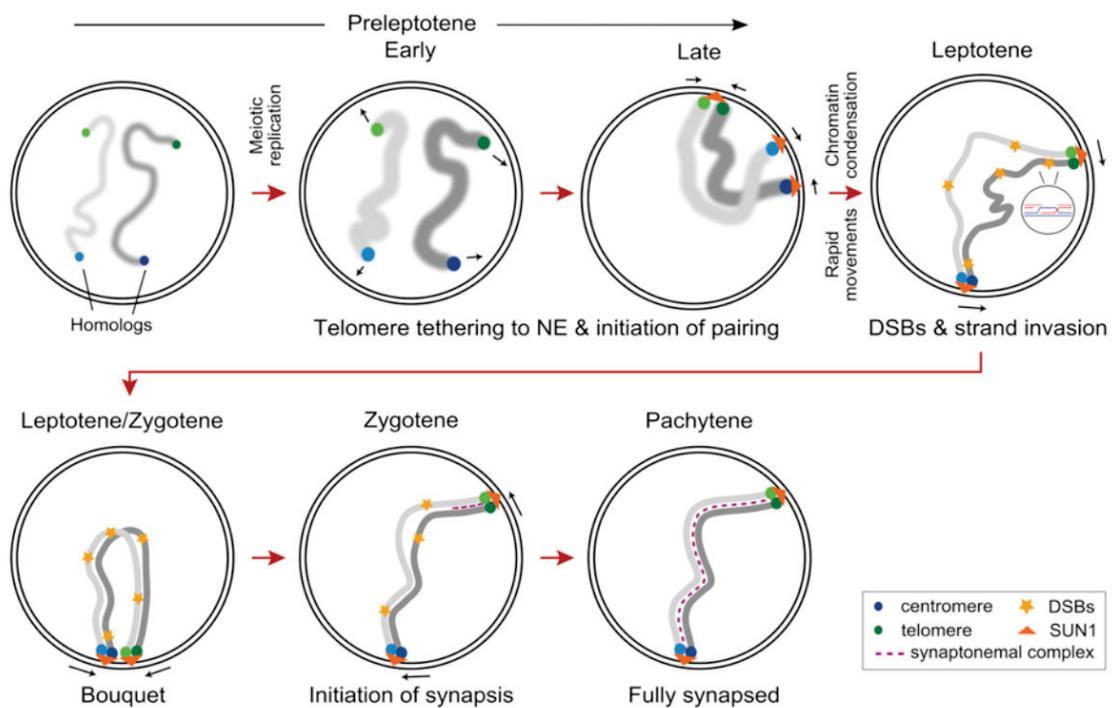
Figure 2.3: Mouse preleptotene DSB-independent pairing model proposed by Boateng et al. (2013).
Boateng et al. (2013)'s model stipulates that the tethering of telomeres (green points) to the NE in late preleptotene facilitates the initiation of synapsis at subtelomeric regions by simplifying the search for the homologous chromosome (light and dark grey lines). The authors also conjecture that, upon entry into prophase (leptotene), this DSB-independent pairing at non-telomeric sites is lost, but that telomeric pairing is maintained at least at one end until homologues recombine. Ultimately, DSB repair and synapsis at zygotene and pachytene would progressively restore pairing at non-telomeric sites.
This figure was reproduced from Boateng et al. (2013) (permission in Appendix B).

The synaptonemal complex (SC)

The synaptonemal complex (SC), discovered by Fawcett (1956) and Moses (1956), is a remarkably well-conserved ribbon-like proteinaceous structure composed of three units: two dense lateral (or axial) elements (LE) and — except in the green alga *Ulva* (Braten and Nordby, 1973) and *Chlamydomonas* (Storms and Hastings, 1977) — one less dense central element (CE) (Figure 2.4) (Schmekel and Daneholt, 1995).

LEs resemble axes along which the sister chromatids are loaded, binding short stretches of DNA to the LE and condensing the rest of it into long loops of tens to hundreds of kilo base pairs (kb). Generally, the loops closer to the telomeres are much shorter than the ones located elsewhere (Heng et al., 1996).

LE assembly begins at leptotene with the aggregation of both REC8 cohesins and axial proteins (SCP2 and SCP3 in mammals) into small fragments (Eijpe et al., 2003) which later fuse into full LEs (Schalk et al., 1998). At full synapsis, they are connected to the CE (formed of SYCE1 and SYCE2 proteins (Pera et al., 2013)) by transverse filaments (TFs), thus giving the SC a striated, zipper-like appearance. The main constituent of TFs — the SCP1 protein, in mammals — has homologues in worms (MacQueen et al., 2002; Colaiácovo et al., 2003), flies (McKim et al., 2002) and yeasts (reviewed in Zickler and Kleckner, 1999) that, despite little sequence conservation, display a similar structure: two head-to-head homodimers of an $\sim$80 nm coiled coil flanked by globular C and N termini (Meuwissen et al., 1992; Liu et al., 1996). The polymerisation of these central region proteins between paired homologue axes results in the tight pairing ($\sim$100 nm) of the bivalents[2] along their entire length at the end of pachytene (Page and Hawley, 2004), as compared to their $\sim$400-nm spacing during presynaptic alignement (Tessé et al., 2003).

Synapsis is indeed the most commonly acknowledged role of the SC, but it may also act to limit recombination with the sister chromatid. Avoiding the sister may seem a trivial problem given the 2:1 odds ratio in favour of homologue templates (Lao and Hunter, 2010). However, an important guarantee of genome stability is

---

[2]Homologous chromosomes
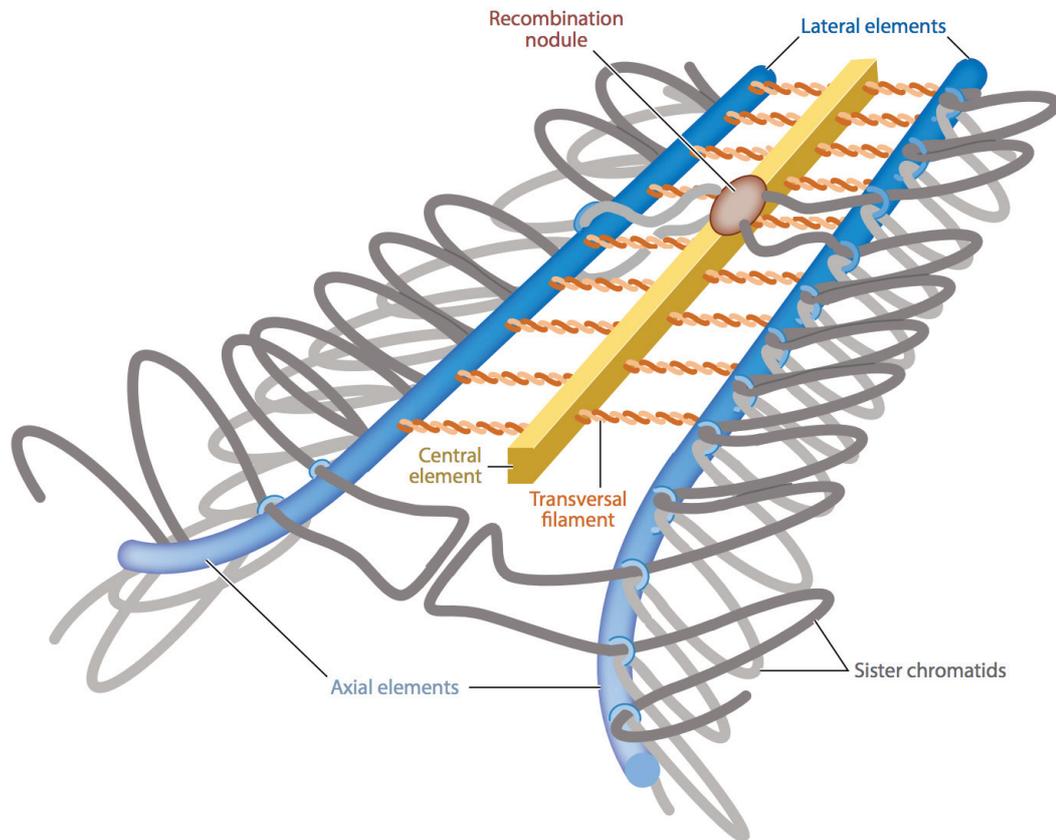
Figure 2.4: Structure of the synaptonemal complex (SC).
Original legend by the author: 'The SC consists of a pair of parallel strands, the lateral elements, that are linked by transversal filaments. The central element runs halfway between the lateral elements. Loops of sister chromatids are tethered both to each other and to a lateral element. Synapsis progresses along pairs of homologous chromosomes in a zipper-like fashion. The axes of unsynapsed portions are called axial elements. Initial homologous interactions may or may not need axial elements. The sites of crossing over are marked by recombination nodules, which are located between the axial elements.'
This figure was reproduced from Loidl (2016) (permission in Appendix B).

the preferential use of the sister chromatid in mitotically dividing cells (Kadyk and Hartwell, 1992; Bzymek et al., 2010) which is likely promoted by their cohesin-dependent proximity (Sjögren and Ström, 2010). Thus, switching this mitotic inter-sister bias to a meiotic inter-homologue bias is essential for synapsis. Even though this could be ensured by other features of meiosis (Schwacha and Kleckner, 1997; Goldfarb and Lichten, 2010; Hong et al., 2013, reviewed in Humphryes and Hochwagen, 2014), recent evidence points that the components of the CE are effectively involved in template choice (Kim et al., 2010) as was suggested in the

past (Haber, 1998).

Microscopy observation of the SC reveals dense nodules where recombination occurs ('recombination nodules') (Carpenter, 1975; Schmekel and Daneholt, 1998). Indeed, the formation of DSBs is a prerequisite for SC formation in many species including plants, mammals and fungi (Zickler and Kleckner, 1999; Henderson and Keeney, 2004). Yet, the meiotic program seems to vary for other species: SC formation is recombination-independent in species with holocentric chromosomes like *Caenorhabditis elegans* (Dernburg et al., 1998) and *Bombyx mori* (Rasmussen, 1977) but also in *Drosophila* females (McKim et al., 1998) (and recombination does not even occur in *Drosophila* males, as reviewed in Tsai and McKee, 2011) whereas *Schizosaccharomyces pombe* (Bahler et al., 1993) and *Aspergillus nidulans* (Egel-Mitani et al., 1982) recombinate but have no SC (reviewed in Zickler and Kleckner, 2015).

More generally, whenever SC is associated to recombination, it seems that its correct formation is important to facilitate stable DNA connections between homologues (Hunter and Kleckner, 2001, reviewed in Hunter, 2003). If, contrariwise, it builds improperly, the resulting asynapsis may have dramatic consequences on the fate of maturating gametes.

2.1.3   Impaired meiosis-associated diseases

Asynapsis

To prevent the formation of abnormal gametes, surveillance systems (a.k.a. 'checkpoints') chase after defects at several meiotic stages (reviewed in Handel and Schimenti, 2010). In particular, the 'pachytene checkpoint' (Roeder and Bailis, 2000) monitors chromosome synapsis in *Saccharomyces cerevisiae* (Wu and Burgess, 2006), *Drosophila melanogaster* (Ghabrial and Schüpbach, 1999; Abdu et al., 2002) and *Caenorhabditis elegans* (Bhalla and Dernburg, 2005). In mammals however,

this one in multiple surveillance systems (Barchi et al., 2005) seems to be associated to the completion of recombination rather than to synapsis *per se* (Li et al., 2007).

An early pachytene response to asynapsis in both mice (Baarends et al., 2005; Turner et al., 2005) and humans (Ferguson et al., 2008; Sciurano et al., 2007) is the meiotic silencing of unsynapsed chromatin (MSUC). In normal males, its specialisation, meiotic sex chromosome inactivation (MSCI), silences sex chromosomes in both mammals and birds (Schoenmakers et al., 2009) and leads to their compartmentalisation into the sex body (Figure 2.2).

MSUC of only one asynapsed chromosome (on top of the sex chromosomes) allows to escape apoptosis[3] (Mahadevaiah et al., 2008; Jaramillo-Lambert and Engebrecht, 2010), the normal response to asynapsis (Hochwagen and Amon, 2006).

Infertility

Regarding sex effects, chromosomal anomalies associated with asynapsis are found in 3% of infertile men (Vincent et al., Feb, cited in Burgoyne et al., 2009) and, more generally, mammalian males are more severely affected by asynapsis-dependent sterility than females (reviewed in Burgoyne et al., 2009 and Hunt and Hassold, 2002), likely because meiosis checkpoints are either less numerous or less efficient in females (Champion and Hawley, 2002).

The converse is true for aneuploidy: since female checkpoints interrupt a smaller proportion of abnormal meioses, they exhibit a higher rate of unbalanced conceptions.

Aneuploidy

In humans, aneuploidy is the primary cause of miscarriage and congenital birth defects (Hassold et al., 2007).

As one studied chromosome proved to transmit properly even in the absence of chiasma (Fledel-Alon et al., 2009), the incapacity to control for proper disjunction,

---

[3]Programmed cell death (from the Greek word ἀπόπτωσις: 'falling off')

— rather than the effective number of recombination events, — may cause these irregularities. These female-specific failures are likely due to the dictyate arrest: female chiasmata, formed at the fetal age, have to hold for decades until puberty resumes meiosis. Consequently, they may degrade over time (Hassold and Hunt, 2001). In accordance with this hypothesis, the frequency of Down Syndrome (a.k.a. trisomy 21) (Penrose, 2009) and other human trisomies (Morton et al., 1988, reviewed in Hassold et al., 1996 and Smith and Nicolas, 1998) are positively correlated with maternal age. In yeasts too, trisomies correlate with parental age (Boselli et al., 2009).

These aneuploidy defects are caused by segregation errors, 80% of which arising during the first meiotic division and many involving an achiasmate bivalent (Székvölgyi and Nicolas, 2010). Therefore, this suggests that one of the most crucial features of meiosis is that yielding chiasmata: homologous recombination (HR).

2.2 Models of homologous recombination (HR)

Ever since the unexpected observations on fungal products of meiosis (see Chapter 1), a few *aficionados* with a craving to understand the exchange of genetic information between chromosomes have come up with theoretical models of homologous recombination (HR).

The Holliday model (Holliday, 1964) was the first widely accepted molecular explanation of the relationship between aberrant segregation and crossing-over. It has since then been refuted by posterior discoveries but one of its concepts, the 'Holliday junction' (HJ), remains a key feature in all current models of HR.

2.2.1   The Holliday junction (HJ)

One central tenet of the Holliday model lies in the idea that DNA can break, thus allowing complementary sequences to pair in a cruciform structure that was later designated as the 'Holliday junction' (HJ). The HJ forms as a consequence of the single-end invasion (SEI) of a nicked DNA strand into the homologous, intact chromosome.

Double Holliday junctions (dHJs) have later been directly observed in recombination intermediates of yeasts (Schwacha and Kleckner, 1994, 1995). However, these studies, like prior works (Sun et al., 1989a; Cao et al., 1990), have shown that recombination does not start with single-strand nicks as enunciated in the Holliday model, but with double-strand breaks (DSBs) as posited in the DSB repair (DSBR) model.

2.2.2   Double-strand break repair (DSBR)

The double-strand break repair (DSBR) model (Szostak et al., 1983) was originally developed from yeast studies (Orr-Weaver et al., 1981; Orr-Weaver and Szostak, 1983) and postulates the formation of DSBs. The broken ends are then processed into two single-stranded DNA (ssDNA) tails. One of them invades the homologue by displacing one of its intact strands into a D-shaped loop designated as the 'D-loop'. This forms the prime HJ (Figure 2.6). Following DNA synthesis of the invading strand, the D-loop broadens sufficiently to anneal the opposite, free 5' end. This completes the formation of a second HJ, crisscrossed with the first one. According to this model, the newly formed dHJ is later resolved into a crossing-over (CO) or a non-crossover (NCO) with a 50:50 odds-ratio.

Many of the predictions of this model revealed true and, as such, it is still used today (see Subsection 2.3.3). But the prognosis regarding the equal number of COs and NCOs was never confirmed biologically (Bishop and Zickler, 2004) which suggested that a portion of NCOs were created *via* another mechanism.

2.2.3 Synthesis-dependent strand annealing (SDSA)

The synthesis-dependent strand annealing (SDSA) model (Resnick, 1976; Nassif et al., 1994; Ferguson and Holloman, 1996) shares its initial steps with the DSBR model: it begins with a DSB and involves a D-loop that extends along the recipient strand (reviewed in McMahill et al., 2007). Once it has elongated past the DSB site, the D-loop is disrupted and the invading strand anneals its original complementary ssDNA on the *vis-à-vis* side of the DSB. Last, the remaining gaps are filled in by DNA synthesis and ligation. This generates NCOs *prior* to the formation of dHJs in the DSBR pathway (Allers and Lichten, 2001).

In the past decades, many experimental studies have uncovered additional spatial and temporal features of meiotic recombination, many of which being in accordance with the aforementioned HR models. I review these findings in the upcoming section.

2.3 Molecular mechanisms of recombination

Homologous recombination (HR), which occurs during prophase I, leads to the formation of a (relatively) long-term connection that maintains the bivalents together until their separation at anaphase I.

It begins at leptotene with the formation of a DNA double-strand break (DSB) on one homologue. To repair properly, this crack needs a DNA strand to use as template. There begins a homology search accomplished at zygotene by the broken-strand invasion onto the mating chromosome. The template-based repair process creates a transient structure, subsequently resolved into either a crossing-over (CO) or a non-crossover (NCO) during late zygotene and pachytene.

In mammals, each of these actions is executed by a complex body of proteins summarised in Figure 2.5.

2.3.1 Initiation of recombination

The evolutionarily conserved SPO11 transesterase — observed in a wide range of species (Baudat et al., 2000; McKim and Hayashi-Hagihara, 1998; Romanienko and Camerini-Otero, 2000; Steiner et al., 2002; Bowring et al., 2006; Stacey et al., 2006) — catalyses the programmed formation of DSBs (Keeney et al., 1997; Bergerat et al., 1997) that marks the beginning of HR (Sun et al., 1989a). Of the two isoforms found in mice (Metzler-Guillemain and de Massy, 2000), SPO11β is the one responsible for DSB formation (Bellani et al., 2010). DNA cleavage by this homodimeric protein leaves a two-nucleotide 5' overhang (de Massy et al., 1995) onto which it remains trapped till the further processing of DSB ends (see Subsection 2.3.2) (reviewed in Cole et al., 2010b).

Several other proteins have been identified as essential for the correct formation of DSBs (extensively reviewed in Keeney, 2008 and de Massy, 2013). Among them, the yeast Mer2-Mei4-Rec114 complex (Li et al., 2006; Maleki et al., 2007) and two of its mouse homologues (MEI4 and REC114) have been identified as functional and required for double-strand break formation by SPO11 (Kumar et al., 2010, 2015), thus suggesting a conserved mechanism for recombination initiation. Nevertheless, the mammalian system has some specificities since MEI1 (Libby et al., 2002, 2003), which does not set forth any yeast homologue, has been uncovered as essential for normal DSB levels, along with HORMAD1 (yeast homologue: Hop1) (Shin et al., 2010; Daniel et al., 2011).

Once DSBs have been generated, the ataxia telangiectasia mutated (ATM) kinase both phosphorylates the 139th serine residue of histone H2AX variants located in their vicinity (then named γH2AX) (Rogakou et al., 1998; Burma et al., 2001) and thwarts further DSB formation (Lange et al., 2011; Lukaszewicz et al., 2018).

In mice and humans, ∼200—400 DSBs initiated in this manner at early leptotene are required to avoid defects in synapsis (Kauppi et al., 2013; Smagulova et al., 2013). From this point forward, they thus have to be repaired to secure the production of viable gametes.

2.3.2 Repair of double-strand breaks (DSBs)

DSB-end processing

The repair of DSBs begins with the processing of its ends: an endonucleolytic cleavage several nucleotides downstream of the 5' end (Neale et al., 2005) is executed by the Mre11/MRE11 complex both in yeasts (reviewed in Borde and Cobb, 2009) and mammals (reviewed in Borde, 2007). In *Saccharomyces cerevisiae* and *Caenorhabditis elegans*, Mre11/MRE11 acts collaboratively with Rad50/RAD50 and Xrs2/NBS1, two proteins required for DSB mending (reviewed in Lam and Keeney, 2015). Both have mammalian homologues, but their putative role in DSB repair (reviewed in Baudat et al., 2013) is hard to prove since knocking them out is lethal for mice (Luo et al., 1999; Zhu et al., 2001).

Single-end invasion (SEI)

As removal of SPO11 is paired with the 5'-to-3' end resection of the DSB, 3' single-stranded DNA (ssDNA) tails become accessible to the nuclear machinery (Figure 2.5.b.). As such, RPA proteins rapidly bind them (He et al., 1995) but are then displaced by RAD51 and/or DMC1 recombinases (Pittman et al., 1998; Yoshida et al., 1998) which catalyze the pairing and exchange between the ssDNA strand and the intact, homologous double-stranded DNA (dsDNA). Their relationship is complex: RPA is necessary both for RAD51 filament formation and for DMC1-catalysed strand exchange, but notwithstandingly, it also competes with them for ssDNA binding (Sung et al., 2003).

The proper functioning of DMC1 and RAD51 in strand invasion requires several other proteins that interact with either one or both of them: HOP2 and MND1 (Bugreev et al., 2014), BRCA1 (Scully et al., 1997) and BRCA2 (Thorslund et al., 2007). This complex process also requires other, less well-characterised actors that I will not describe here for they are of little interest for the scope of this thesis (but for review, see Neale and Keeney, 2006, and Figure 2.5.c.).

Next, the sensor proteins of the mismatch repair (MMR) system (MSH2-MSH3 and MSH2-MSH6 complexes in mammals) control the identity between the targeted
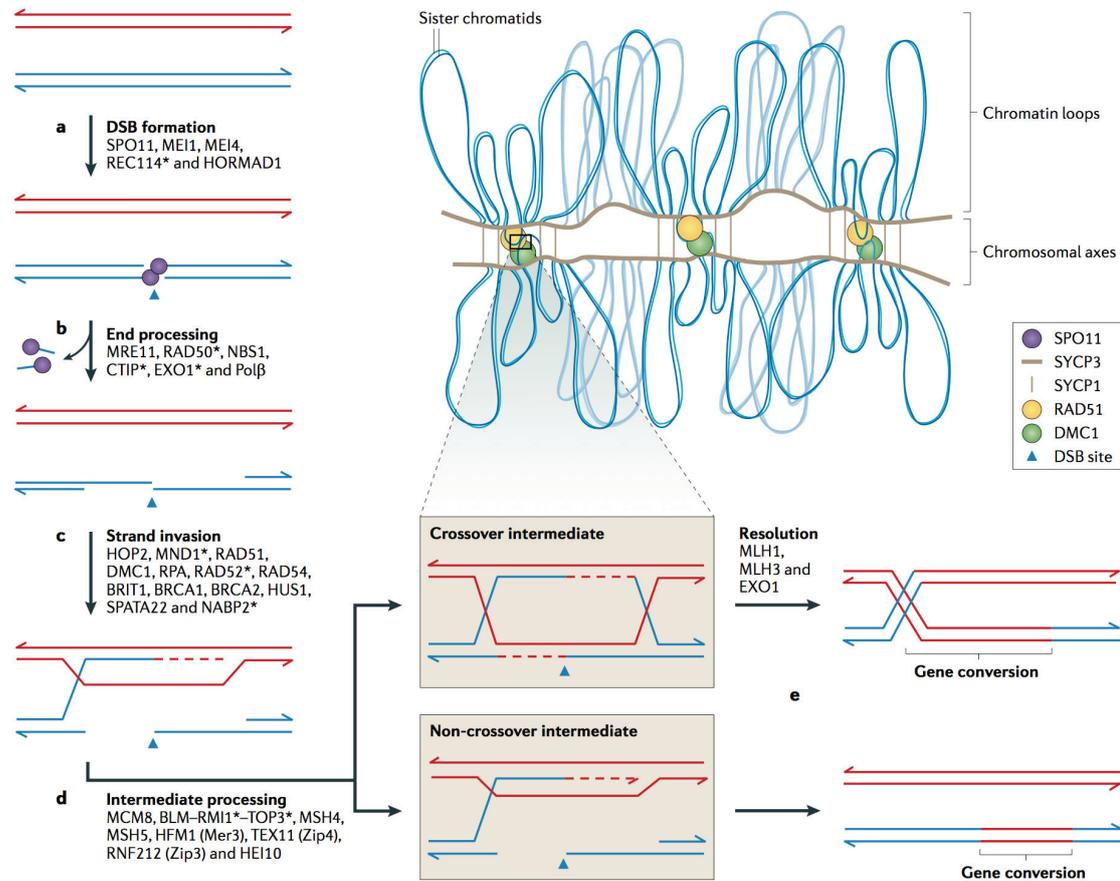
Figure 2.5: Proteins involved in mammalian meiotic recombination.
a | DNA double-strand break (DSB) formation (blue triangles) is catalysed by SPO11 (purple spheres) on the chromosome axes and requires MEI1, MEI4, REC114 and HORMAD1. **b** | Endonucleolytic cleavage of DSB ends by MRE11, RAD50, NBS1, CTIP and Pol β forms SPO11-oligonucleotide complexes of 12—36 nucleotides (purple spheres with tails). EXO11 further enacts a 5'-to-3' resection of DSB tails. **c** | Strand invasion is catalysed by DMC1 and RAD51 recombinases in the presence of several co-factors: HOP2, MND1, RAD52, RAD54, BRIT1, BRCA1 and BRCA2. RPA and NBPA2 bind recombination intermediates. At this stage (zygotene), homologous chromosomal axes are synapsed at DSB repair sites by proteins of the synaptonemal complex, including SYCP1 (brown segments). **d** | Recombination intermediates are either dismantled by BLM-RMI1-TOP3 to generate non-crossover intermediates, or stabilized by TEX11, MSH4-MSH5, RNF212, ZIP2, HFM1 and HEI10 to generate double Holliday junctions (CO intermediates). **e** | Resolution into crossovers requires MLH1, MLH3 and EXO1 while non-crossovers are formed after strand displacement and annealing. Non-crossovers formed *via* alternative pathways are not shown. Recombination products are generated at the end of pachytene. Gene conversion (unidirectional transfer of genetic information in the vicinity of DSB) is present in both products.
Proteins marked with an asterisk ([*]) are predicted to be involved, but not yet confirmed by experimental evidence. Chromatin loops and chromosome axes during zygotene are illustrated in the top right.
This figure was reproduced from Baudat et al. (2013) (permission in Appendix B).

strand and the invader. When it is insufficient, the latter is rejected and repaired using the sister chromatid instead, thus preventing any potentially deleterious ectopic recombination (reviewed in Surtees et al., 2004 and Goldfarb and Lichten, 2010).

Recombination-intermediate processing

The interaction between the invading strand and the homologue is subsequently stabilised by several proteins. Indeed, BLM, TEX11 (yeast homologue: Zip4) and RNF212 (yeast homologue: Zip3) appear at zygotene at recombination foci and progressively decrease until the end of pachytene, i.e. when DSBs are repaired (reviewed in Baudat et al., 2013). In addition, together with MCM8 and MCM9 proteins (Lutzmann et al., 2012), heterodimers of MSH4 and MSH5 (Scully et al., 1997) are required for synapsis stabilisation in both mice (de Vries et al., 1999; Kneitz et al., 2000) and humans (Snowden et al., 2004).

Though, the role of MSH4 continues beyond synapsis establishment. Indeed, the stabilisation of the interaction between the two homologues creates an intertwined recombination intermediate structure, and MSH4 participates in its resolution when it leads to COs, but also, as argued by Baudat and de Massy (2007), to NCOs.

2.3.3   Resolution of recombination intermediates

Recombination intermediate structures may be resolved *via* two main pathways (Figure 2.6). In the pathway leading to COs, the non-invading strand of the broken chromosome interacts with the displaced homologue strand which forms the D-loop. In constrast, in the pathway leading to NCOs, the non-invading strand anneals again the invading strand from the same chromatid, after the latter has elongated on the homologue and displaced from it. Assertedly, these two pathways presuppose the production of distinct recombination intermediates (Figure 2.5.d. and e.).
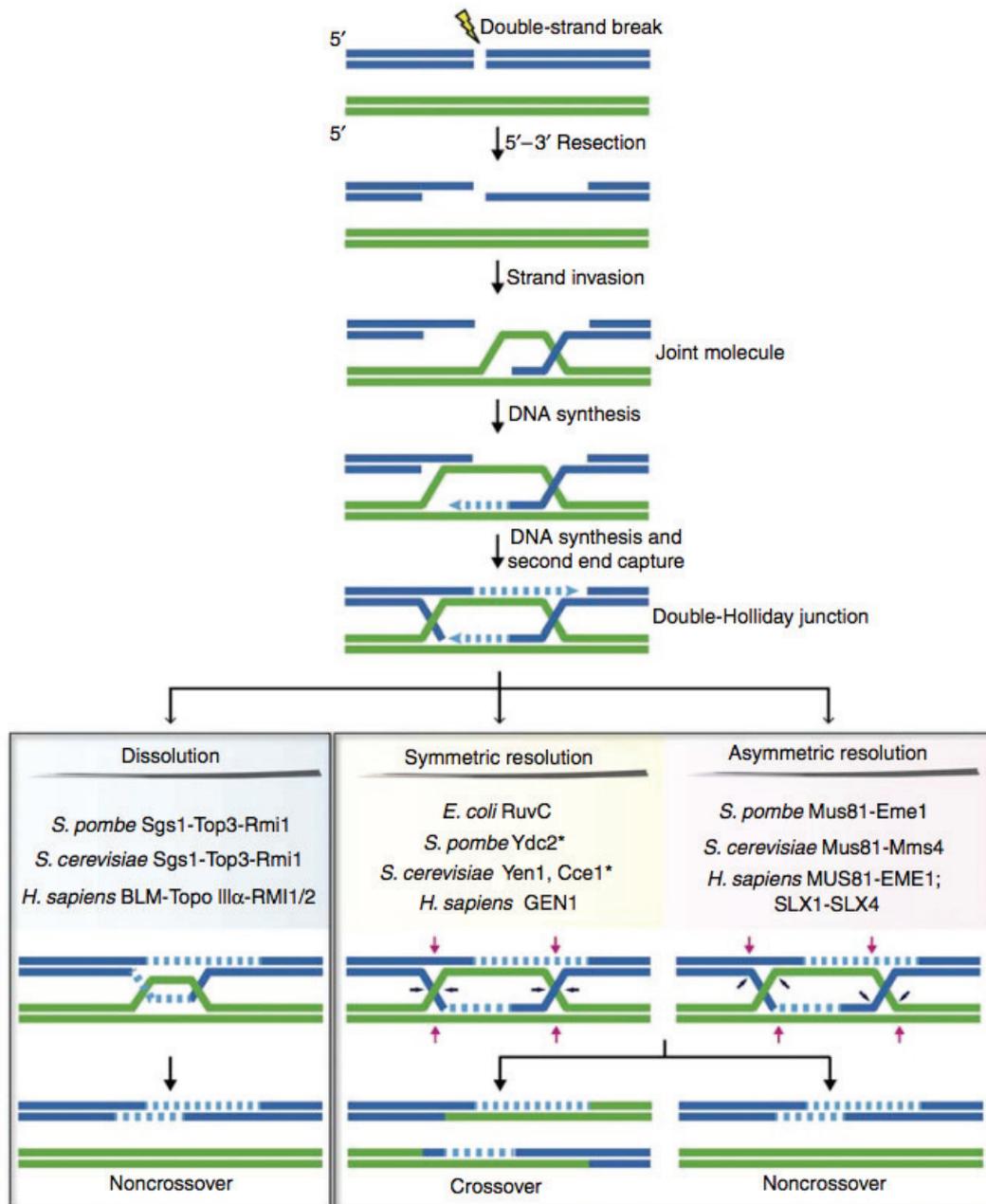
Figure 2.6: Molecular mechanism of pathways leading to crossing-overs (COs) and non-crossovers (NCOs).
Resected DSBs invade homologous duplex DNA to form a D-loop structure. The invading 3' end then serves as a primer for DNA synthesis, which leads to the capture of the second end and, ultimately, to the formation of a double Holliday junction. This junction is then either dissolved into a NCO (left panel), resolved by canonical Holliday junction resolvases introducing a pair of symmetrical nicks to generate nicked DNA duplexes that can be directly ligated (middle panel) or resolved by noncanonical resolvases introducing asymmetrical nicks to produce gapped and flapped DNA duplexes that require further processing prior to ligation (bottom right panel). If only two strands are cleaved, the outcome is necessarily a NCO while it is a CO if all four strands are cleaved.
This figure was reproduced from Wyatt and West (2014) (permission in Appendix B).

The CO pathway

In certain cases, the homologues are physically bound twice: one strand from each chromosome (the invading strand and the D-loop strand) displaces to bind the homologue, thus creating a double Holliday junction (dHJ) in step with the DSBR model. TEX11 (yeast homologue: Zip4), RNF212 (yeast homologue: Zip3) and HFM1 (yeast homologue: Mer3) — three of the eight proteins of the ZMM complex conserved between the budding yeast and mammals (reviewed in Pyatnitskaya et al., 2019) — are thought to play a role in processing the dHJ, since knocking one of them out leads to a diminished level of chiasmata and COs (Adelman and Petrini, 2008; Guiraldelli et al., 2013; Reynolds et al., 2013, reviewed in Baudat et al., 2013). In yeasts, Mer3 seems to stimulate heteroduplex extension, possibly to stabilise D-loop structures (Mazina et al., 2004).

The resolution of the dHJ *per se* is catalysed by resolvases, i.e. enzymes that slice the interwound strands. In mice, a pair of nicks is introduced across the helical branchpoint of most (90%) dHJs by the concerted action of the MLH1-MLH3 heterodimer (Baker et al., 1996; Edelmann et al., 1996; Lipkin et al., 2002) and of EXO1 (Wei et al., 2003).

Alternatively, the dHJ can be resolved by introducing two single-stranded incisions (Wyatt and West, 2014). In that case, the two nicks are asymmetric and can be located several nucleotides away from the branchpoint. This resolution is catalysed by MUS81 and EME1 (yeast homologue: Mms4). In *Schizosaccharomyces pombe* where it was first discovered, it is the only pathway to produce COs (Osman et al., 2003). However, in plants (Mercier et al., 2005), budding yeasts (de los Santos et al., 2003) and mice (Holloway et al., 2008), it coexists with the MLH1-dependent CO pathway.

Of the 200—400 recombination foci in mice, only ∼20 (approximately one per chromosome) lead to a CO (Baudat and de Massy, 2007). This implies the existence of another repair pathway: that leading to NCO events.

The NCO pathway

Instead of being resolved, the dHJ is sometimes dissolved by the BLM helicase together with a topoisomerase (Wu and Hickson, 2003). This pathway thus interferes with the formation of COs. Indeed, inactivating BLM leads to an increased number of chiasmata (Holloway et al., 2010).

Though, most NCOs are formed *via* another pathway that occurs before the resolution of dHJs: the synthesis-dependent strand annealing (SDSA) pathway (see Subsection 2.2.3). In *Saccharomyces cerevisiae*, it produces the large majority of NCOs (Martini et al., 2011) and the dissociation between the invading strand and the homologue is promoted by Sgs1 (De Muyt et al., 2012) while another helicase, Srs2, also promotes the SDSA pathway *via* a different mode of action (Ira et al., 2003). However, the latter helicase does not have any mammalian homologue (Spell and Jinks-Robertson, 2004). Therefore, the molecular operations of SDSA in mammals are still unclear.

Altogether, the resolution of a genetically programmed DSB into a CO *versus* a NCO outcome seems to be decided early: in most species, they arise from distinct intermediates (reviewed in Hunter, 2015). This intermediate structure involves the formation of a heteroduplex, which, in mammals, can spread over 500–2,000 bp for COs, but generally less than 300, and sometimes as little as tens of base pairs, for NCOs (Jeffreys and May, 2004; Ng et al., 2008). Heterozygous markers located within the heteroduplex are either all converted in the same direction (in that case, the conversion tract of the CO or NCO is said to be 'simple') or alternate converted and unconverted markers (in that case, the conversion tract is said to be 'complex') (Borts and Haber, 1989).

In contrast, non-programmed DSBs, which correspond to DNA lesions, can be repaired either by homologous recombination (reviewed in Sung and Klein, 2006) or by alternative processes. Indeed, such spontaneous DSBs are frequent in mitotic cells and mitotic breaks are mainly repaired by recombining with the genetically identical sister chromatid, or *via* one of two repair systems that are more error-prone

(Smith et al., 2001): non-homologous end-joining (NHEJ), which consits in directly ligating the broken strands of DNA (Weterings and van Gent, 2004) or single-strand annealing (both reviewed in Helleday, 2003 and Moynahan and Jasin, 2010).

Recombination may also occur between non-allelic sequences located at different genomic locations — generally low copy repeats resulting from duplication events (Bailey and Eichler, 2006). This is called non-allelic homologous recombination (NAHR) (or 'ectopic recombination') and proceeds similarly to HR (Sasaki et al., 2010).

Distinguishing between HR and NAHR implies knowing where recombination effectively takes place on the genome, which is the object of the next chapter.

'Intense selection favours a variable response to the environment. . . Were this not so, the world would be much duller than is actually the case.'

— John B. S. Haldane, *The Causes of Evolution*
(1932)

3
Causes and consequences of recombination rate evolution

Contents

The very mechanism of meiosis introduces genetic mixing in two separate ways. On the one hand, the paternal and maternal chromosomes are independently re-assorted during the first meiotic division. On the second hand, genetic content is exchanged during recombination at the points where homologues cross over (a.k.a. chiasmata).

Even if this phenomenon was not known in Charles Darwin's time, he had the intuition that genetic diversity — which meiosis participates in instilling —

was essential to the formation of new species:

> *'The principle, which I have designated by this term [ed. divergence of character], is of high importance on my theory, and explains, as I believe, several important facts. [. . .] according to my view, varieties are species in the process of formation, or are, as I have called them, incipient species. How, then, does the lesser difference between varieties become augmented into the greater difference between species?'*
>
> — Charles Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*
> (1859)

As enunciated by his theory, the transition from varieties to species requires 'a severe struggle for life [which] certainly cannot be disputed' (natural selection), the occurrence of 'variations useful to any organic being' (mutations) and 'the strong principle of inheritance' through which 'they will tend to produce offspring similarly characterised' (heredity). As such, the emergence of new species is tightly linked to the process of meiotic recombination since it is a major vector of genetic variation at the heart of the process of heredity.

Furthermore, the notion of biological species itself, formally defined by Ernst Mayr (1904–2005) as 'groups of interbreeding natural populations that are reproductively (genetically) isolated from other such groups' (Mayr, 1999), rests on the ability to sexually reproduce and thus, to meiotically recombine.

The relationship between these two concepts (further developed in Felsenstein, 1981 and Butlin, 2005) is such that, in the mammalian clade, the only speciation gene discovered so far (PRDM9) is the one that controls the localisation of double-strand breaks (DSBs) on the genome (Baudat et al., 2010; Myers et al., 2010; Parvanov et al., 2010).

I will come back to this essential gene and to its impact on the evolution of recombination rate in the third section of this chapter. But prior to that, I will review the existing methods to detect recombination genome-wide, and the multiple layers of recombination rate (RR) variation that have been observed along genomes and across species.

3.1   Genome-wide detection of recombination

3.1.1   Linkage maps *via* the analysis of crosses or pedigrees

The comprehension of genetic linkage by the group of Thomas Hunt Morgan (see Chapter 1) was the inaugural step towards the establishment of the first genetic map (a.k.a. linkage map) (Sturtevant, 1913). Basically, these maps abstractly represent the proportion of crossing-overs (COs) occurring between pairs of 'genetic markers', i.e. polymorphic[1] DNA sequences located at fixed genomic positions.

Initially, genetic markers exclusively comprised genes coding for visually discernable phenotypes. Since their relatively wide genomic spacing granted a poor resolution to detect recombination, they were eventually supplanted by other types of markers: restriction fragment length polymorphisms (RFLPs) i.e. sequences enzymatically shortenable first used for linkage analysis by Botstein et al. (1980); minisatellites and microsatellites (Hamada and Kakunaga, 1982) i.e. tandem repeats of short motifs highly variable in length (Ellegren, 2004) and widely spread in eukaryotes (Hamada et al., 1982); and single-nucleotide polymorphisms (SNPs) i.e. one-base sequence variations.

When the two parental chromosomes carry distinct alleles at these loci[2], one can track their transmission by genotyping the markers in the descendants. As such, the mosaic of paternal and maternal haplotypes — and thus, the positions of recombination exchange points — can be reconstituted using various statistical methods (Haldane, 1919; Kosambi, 1943, reviewed in Backström, 2009).

These kindred individuals are generally obtained by crossing members of highly divergent inbred populations (e.g. Rowe et al., 1994; Dietrich et al., 1996), one of which being, if possible, homozygous for the recessive alleles ('test cross') so as to disentangle the genotypes of the descendants (reviewed in Brown, 2002). Alternatively, in species that have long generation time or that cannot be manipulated

---

[1]Which presents several forms. In other words: subject to inter-individual variability.

[2]Fixed position of a genetic marker on a chromosome (from the Latin word *locus*: 'place')

genetically for ethical considerations, successive generations of existing families (a.k.a. pedigrees) can be examined (e.g. Kong et al., 2002, 2010; Cox et al., 2009).

Examining large numbers of individuals allows to estimate the genetic distance (measured in 'morgans' (M) as a tribute to its designer) between pairs of markers: one centimorgan (cM) expresses a frequency of 1 CO every 100 meioses. However, for high recombination frequencies (i.e. long distances), some experiments (e.g. Morgan, 1911; Morgan and Cattell, 1912) showed exceptions to additivity: the genetic distance between two polymorphic sites could be smaller than the sum of their distances with an in-between marker. Indeed, in cases of 'double crossing-overs' (i.e. two COs occurring within a given interval — which is more likely in wider stretches), the two loci are inherited together. Thus, the CO event is not detectable and, in the end, the recombination frequency is underestimated.

In addition, genetic distances are not proportional to physical remoteness, as stated by Hermann Muller (1890–1967) (Muller, 1920) in a response to William Castle (1867–1962) who disputed the graphical representation of these maps (Castle, 1919a,b, reviewed in Vorms, 2013):

> '[I]t has never been claimed, in the theory of linear linkage, that the per cents of crossing over are actually proportional to the map distances [ed. physical distances]: what has been stated is that the per cents of crossing overs are calculable from the map distances — or, to put the matter in more mathematical terms, that the per cents of crossing over are functions of the distances of points from each other along a straight line.'

Decades later, the complete sequencing of the *Saccharomyces cerevisiae* chromosome III (Oliver et al., 1992) confirmed this statement by enabling the first direct comparison between linkage and physical maps. The discrepancies between the two distances legitimised the introduction of a new measurement: the estimation of recombination rates (RRs) per physical distance (expressed in cM/Mb), useful to compare RRs across genomic regions, individuals or species.

Altogether, linkage maps directly measure recombination occurring in the offspring and thus allow to observe differences between sexes (e.g. Cheung et al., 2007; Coop et al., 2008) or among individuals (e.g. Broman et al., 1998). However, the resolution of these maps is restrained by the position of polymorphic sites and the number of meioses analysed. Consequently, in mammals, except for one very recent study (Halldorsson et al., 2019), the resolution has remained capped at tens to hundreds of kilo base pairs (kb) (Shifman et al., 2006; Billings et al., 2010; Kong et al., 2010). This limitation motivated the development of a population-genetic method to learn about RRs at a finer-scale: the linkage disequilibrium (LD) analysis.

3.1.2 Linkage disequilibrium (LD) analysis

Populations of unrelated beings can be analysed in a fashion similar to family members since kinship (or non-kinship) only conveys a *relative* sense: unrelated individuals are merely more distantly akin than traditional pedigrees (Nordborg and Tavaré, 2002).

Therefore, the principle remains the same for populations of unrelated individuals as for families: recombination breaks down linkage disequilibrium (LD) (Lewontin and Kojima, 1960), i.e. non-random associations between loci (materialised by non-random segregations of alleles), which results in the fragmentation of LD into blocks. Reciprocally, analysing patterns of LD (i.e. the positions of LD blocks) will allow to trace back the underlying recombination process.

Concretely, LD can be quantified using statistics of association between allelic states at pairs of loci (Lewontin, 1964; Hill and Robertson, 1968) and the recombination rates (RRs) further estimated through a myriad of methods (reviewed in Stumpf and McVean, 2003) which basically consist in using the allelic diversity of each LD block to reconstruct the genealogy (reviewed in Hinch, 2013). Indeed, patterns of LD do not account for recombination only (reviewed in Venn, 2013): they are also shaped by other forces such as population history (Golding, 1984), mutation

(Calafell et al., 2001) (though easily distinguishable from recombination (Hudson and Kaplan, 1985)), natural selection (Barton, 2000) and drift (Charlesworth et al., 1997). Modelling the underlying genealogical history of the population therefore allows to take the latter effects into account and thus, to estimate RR accurately from LD patterns (Stumpf and McVean, 2003).

Recombination events have been inferred by LD analysis in a plethora of mammalian orders including Artiodactyla (Farnir et al., 2000; McRae et al., 2002; Nsengimana et al., 2004), Carnivora (Menotti-Raymond et al., 1999; Sutter et al., 2004; Verardi et al., 2006), Lagomorpha (Carneiro et al., 2011), Rodentia (Brunschwig et al., 2012), Perissodactyla (Corbin et al., 2010; McCue et al., 2012) and Primates (Auton et al., 2012). Though, the resolution of recombination events is greatest in humans, where it has reached 1 to 2 kb (The International HapMap Consortium, 2007; Hinch et al., 2011; The 1000 Genomes Project Consortium, 2015). Such precision arises from the fact that there have had many oppportunities for recombination to take place between the last common ancestor (LCA) of a population of unrelated beings and its studied descendants. Since recombination decreases LD at every generation (Slatkin, 2008), the more ancient the LCA, the shorter the LD blocks and thus, the higher the resolution.

However, the recombination events identified with LD analysis sum up the whole recombination process that has occurred since the LCA: historical recombination, rather than current recombination, is uncovered. In addition, LD studies give a population average of recombination, with no possibility to extricate sex-specific nor individual recombination events. Third, both LD studies and linkage maps allow the detection of COs, but not NCOs.

Another method, — sperm-typing, — solves the three aforementioned caveats: it provides fine-scale mapping of current CO and NCO recombination events in separate individuals.

3.1.3 High-resolution sperm-typing studies

Sperm-typing consists in analysing the transmission of recombination events directly in the sperm of an individual. This was made possible by the development of a polymerase chain reaction[3] (PCR) method allowing to genotype single diploid and haploid cells (Li et al., 1988). Since PCR only allows the copy of size-limited DNA sequences and cannot be performed automatically, sperm-typing cannot be applied genome-wide (Coop et al., 2008), unless a microfluidic device is used (Fan et al., 2011; Wang et al., 2012a). Instead, sperm-typing is generally restricted to regions of high recombinational activity inferred from linkage or LD maps (see Subsections 3.1.1 and 3.1.2).

It can be applied either to single gametes or to total-sperm DNA (reviewed in Arnheim et al., 2003). In single-sperm typing, the PCR is performed on the lysed sperm of an individual gamete with the use of pairs of primers[4] flanking two polymorphic markers at the extremities of the locus of interest (Cui et al., 1989; Lien et al., 1993). This *modus operandi* has soon been used to construct linkage maps on highly recombining regions (Schmitt et al., 1994; Lien et al., 2000; Cullen et al., 2002) while others (Tusié-Luna and White, 1995; Jeffreys et al., 1998, 2001; Guillon and de Massy, 2002) have used the alternative approach with total-sperm DNA which requires allele-specific PCR to capture and amplify recombinant molecules (Wu et al., 1989).

In both cases, the precise CO exchange point can be mapped using the genetic markers internal to the selected locus. Sperm-typing thus offers the best resolution for recombination exchange points since it is only limited by SNP density — a resolution even sufficient to detect the difficult-to-access NCOs that only affect a few markers (Hellenthal and Stephens, 2006), as in Tusié-Luna and White (1995) and Guillon and de Massy (2002).

---

[3]Molecular biology method used to make copies of a specific DNA fragment.

[4]Short single-stranded nucleic acid used to initiate DNA synthesis.

However, even though some authors have managed allele-specific PCR in pooled ovaries (Guillon et al., 2005; Baudat and de Massy, 2007) and single oocytes (Cole et al., 2014), it has almost exclusively been used for the study of male products of meiosis.

The three methods described so far allow to detect the outcome of the recombination process: COs (and NCOs in the case of sperm-typing). To get insights into other stages of the recombination process, one can use chromatin-immunoprecipitation (ChIP) of proteins involved in a given recombination stage (see Chapter 2) to crosslink them on their DNA binding sites, followed by the identification of bound DNA sequences either with a microarray (ChIP-chip) or by direct sequencing of the fragments (ChIP-seq) (reviewed in Park, 2009). The sites of recombination initiation have been identified by using this technique with Spo11 proteins in yeasts (Gerton et al., 2000; Mieczkowski et al., 2007; Pan et al., 2011) and mice (Lange et al., 2016) and the repair sites with RPA proteins in yeasts (Borde et al., 2009) and RAD51 and DMC1 proteins in mice (Smagulova et al., 2011; Brick et al., 2012). Alternatively, sites of recombination initiation have been mapped by analysing the enrichment of single-stranded DNA (ssDNA) in yeasts (Blitzblau et al., 2007; Buhler et al., 2007) and mice (Khil et al., 2012).

These methods do not rely on the existence of polymorphic markers and, therefore, only depend on the size of the region bound by the protein. As such, the resolution reaches up to ∼500 bp for DMC1, ∼50 bp for PRDM9 and a few base pairs for SPO11.

All these approaches have contributed to a better understanding of recombination genome-wide. In particular, it was soon understood that COs do not appear at random locations on the genome. The reasons for this particular distribution became the object of many research works.

3.2 The landscape of recombination

3.2.1 The non-random distribution of crossing-overs (COs)

The number and distribution of crossing-overs (COs) along the genome are subject to a tight regulation (reviewed in Jones, 1984; Jones and Franklin, 2006): a minimum number of COs ('CO assurance'), evenly spaced ('CO interference') — including when few DSBs are generated ('CO homeostasis') — are formed preferentially with the homologous chromosome.

Crossing-over assurance (COA), or the 'obligatory crossing-over'

Together with sister chromatid cohesion, COs hold the homologous chromosomes joint until anaphase I (reviewed in Roeder, 1997) and are therefore essential to the proper disjunction of bivalents. Accordingly, in most sexually-reproducing organisms, the total number of COs ranges between one per chromosome and one per chromosome arm[5], irrespective of chromosome length (Dutrillaux, 1986; Pardo-Manuel de Villena and Sapienza, 2001; Dumas and Britton-Davidian, 2002; Hillers and Villeneuve, 2003; Hassold et al., 2004; Dumont, 2017). As such, mammalian genetic map lengths (which are proportional to CO numbers) can be predicted with the haploid number of chromosome arms (Figure 3.1).

The sexual chromosomes also comply to this phenomenon: they systematically have one CO on their pseudoautosomal region (PAR), a feature likely facilitated by the much higher DSB rate on the PAR than on the autosomes (Kauppi et al., 2011). However, this 'obligatory CO' rule suffers exceptions: *Drosophila melanogaster* females do not display any CO on their tiny 4th chromosome nor, in certain cases, on their X chromosome (Orr-Weaver, 1995; Koehler and Hassold, 1998) and neither do marsupial sex chromosomes (Sharp, 1982).

---

[5]With the notable exceptions of honey bees (Beye et al., 2006) and birds (Groenen et al., 2009) which display higher numbers of COs per chromosome, and of *Drosophila melanogaster* males who do not display any CO throughout their genome (McKee, 1998).
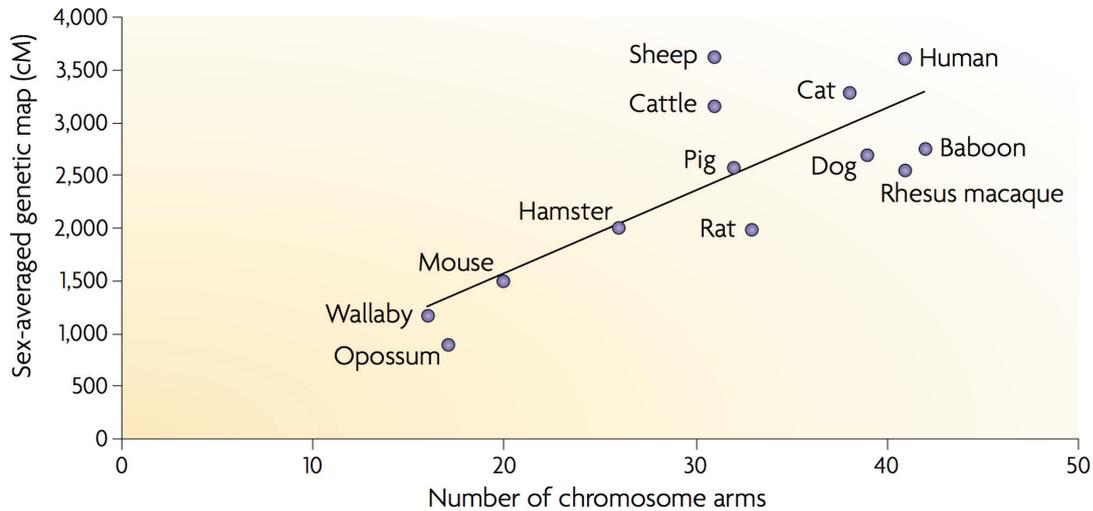
Figure 3.1: Correlation between genetic map lengths and the number of chromosomal arms in mammals.

The y-axis represents genetic map lengths from which the number of crossing-overs (COs) can be extrapolated. The x-axis represents the total number of chromosomal arms per species, excluding the small arms of acrocentric chromosomes and the sex chromosomes of baboon and rhesus macaques. The black line corresponds to the best fit between these two measures.

This figure was reproduced from Coop and Przeworski (2007) (permission in Appendix B).

In *Caenorhabditis elegans*, crossing-over assurance (COA) is so strong that only one DSB per pair of chromosome suffices to guarantee a CO (Rosu et al., 2011). Nevertheless, chromosome pairs holding only one DSB may be uncommon since the number and position of DSBs is also under tight control, at least in yeasts (Wu and Lichten, 1995; Fan et al., 1997; Robine et al., 2007; Anderson et al., 2015): the formation of a DSB reduces the likelihood for another to form nearby (Garcia et al., 2015). This phenomenon, called 'interference', applies to DSBs and another one, also called interference but applying this time to COs *via* a distinct mechanism, has also been reported, as reviewed in the upcoming paragraph.

Crossing-over interference (COI)

Early studies on recombination (Sturtevant, 1915; Muller, 1916) have shown that, when more than one CO appears on a given chromosome, the chiasmata they form tend to be evenly spaced (Jones, 1967, 1974, 1984; Jones and Franklin, 2006).

Indeed, the occurrence of a CO hampers the coincident formation of another one in the same pair of chromosomes (van Veen and Hawley, 2003; Hillers, 2004) — the physical length of prophase chromosomes, rather than the genomic (bp) or genetic (cM) distance, being the primary parameter (Zhang et al., 2014; Wang et al., 2015). So far, COI has been noted in several species including *Arabidopsis thaliana* (Drouaud et al., 2007), *Saccharomyces cerevisiae* (Shinohara et al., 2003), *Homo sapiens* (Laurie and Hultén, 1985; Broman and Weber, 2000) and *Mus musculus* (Lawrie et al., 1995; Anderson et al., 1999; Broman et al., 2002).

The mechanism of COI remains unclear but several models have been proposed (reviewed in Youds and Boulton, 2011). One early hypothesis, — the polymerisation model, — posits that the completion of a CO triggers the polymerisation of an inhibitor of recombination, thus preventing the formation of adjacent COs (Maguire, 1988; King and Mortimer, 1990). According to another one, — the stress model, — axis buckling converts the recombination intermediate into a CO, and this mechanical tension is released in the vicinity of established COs, thus making neighbouring DSBs repair into NCOs instead (Börner et al., 2004; Kleckner et al., 2004). The most recent pieces of evidence point that, in mice, COI may operate in two consecutive steps: at late zygotene and at pachytene (de Boer et al., 2006).

Correlations between the length of the synaptonemal complex (SC) and interference have been reported (Sym and Roeder, 1994; Lynn et al., 2002; Petkov et al., 2007), but others have found that COI does not depend on the SC (de Boer et al., 2007; Shodhan et al., 2014), which suggests that COI operates before SC formation: either prior to single-end invasion (SEI) (Hunter and Kleckner, 2001; Bishop and Zickler, 2004) or during the stabilisation of the SEI (Shinohara et al., 2008).

Whatever the mechanism at play, it may have a role in controlling the outcome of the repair (e.g. by preferentially recruiting the MUS81 repair machinery). Indeed, the COs formed *via* the DSBR pathway comply to COI whereas those repaired *via* the MUS81 pathway do not (de los Santos et al., 2003; Kohl and Sekelsky, 2013).

In particular, neither *Schizosaccharomyces pombe* for which all COs depend on the Mus81 pathway (Munz, 1994; Hollingsworth and Brill, 2004; Cromie et al., 2006) nor *Aspergillus nidulans* which lacks SC (Strickland, 1958, reviewed in Shaw and Moore, 1998 and Egel, 1995) show CO interference.

As for NCOs, their formation is undoubtedly promoted by COI to downregulate the number of COs (Rockmill et al., 2003; Youds et al., 2010; Crismani et al., 2012; Séguéla-Arnaud et al., 2015).

Crossing-over homeostasis (COH)

Even though it has been disputed (Shinohara et al., 2008), the mechanism that ensures COI may be responsible for another level of regulation: crossing-over homeostasis (COH) (Joshi et al., 2009; Zanders and Alani, 2009, reviewed in Youds and Boulton, 2011). COH promotes the formation of COs at the expense of NCOs when fewer DSBs than the wild-type level are generated. This phenomenon was initially observed in *Saccharomyces cerevisiae* (Martini et al., 2006; Chen et al., 2008), but also exists in *Caenorhabditis elegans* (Yokoo et al., 2012; Globus and Keeney, 2012), *Drosophila melanogaster* (Mehrotra and McKim, 2006) and *Mus musculus* (Cole et al., 2012).

Preference for the homologue over the sister chromatid in DSB repair

So that the homologous chromosomes disjoin properly, a fourth regulatory level applies to the repair of DSBs into COs: the promotion of interhomologue repair over intersister mending. Template choice must be regulated differently in mitosis and meiosis (Andersen and Sekelsky, 2010). Indeed, in mitosis, the sister chromatid is always favoured (Kadyk and Hartwell, 1992; Bzymek et al., 2010), whereas evidence in *Saccharomyces cerevisiae* suggests that, in meiosis, two thirds (Goldfarb and Lichten, 2010) to nearly all (Pan et al., 2011) DSBs are repaired using the homologue.

Cohesins and components of the SC seem to be implicated in template choice (Couteau et al., 2004; Kim et al., 2010, reviewed in Pradillo and Santos, 2011) but the

proteins that play a role in homology search are also adequate candidates for this endeavour (reviewed in Youds and Boulton, 2011). Indeed, in *Saccharomyces cerevisiae*, the phosphorylation of Hop1 (mouse homologue: HORMAD1) triggers a mechanism that prevents intersister repair of DSBs (Niu et al., 2005): it inhibits Rad51 (Niu et al., 2009), thus leaving homology search to Dmc1 which promotes interhomologue recombination more efficiently than Rad51 (Schwacha and Kleckner, 1997).

Elucidating these four layers of control on the formation and genome-wide distribution of COs was largely fostered by the immunodetection of the MLH1 protein (which is a marker of CO events) on meiotic chromosome spreads. Such maps have been obtained in multiple clades including primates (e.g. Sun et al., 2005; Codina-Pascual et al., 2006; Garcia-Cruz et al., 2011; Gruhn et al., 2013; Muñoz-Fuentes et al., 2015), rodents (e.g. Froenicke et al., 2002; Dumont and Payseur, 2011), ruminants (e.g. Vozdova et al., 2013; Sebestova et al., 2016) and other eutherians (e.g. Borodin et al., 2008; Segura et al., 2013; Mary et al., 2014, reviewed in Capilla et al., 2016).

Further analysis of maps like those has allowed to uncover both the large-scale and fine-scale patterns of recombination rate (RR) variation along the genomes, which are reviewed in the forthcoming subsection.

3.2.2 Intragenomic patterns of variation

Large-scale variations across genomic regions

When compared over the scale of megabases (Mb), recombination rates (RRs) vary by an order of magnitude in both humans (Figure 3.2.a.) (Nachman, 2002; Myers et al., 2005) and mice (Billings et al., 2010; Morgan et al., 2017).

These large-scale variations associate with certain elements of the genome (reviewed in de Massy, 2013 and Lam and Keeney, 2015). Centromeric regions, for instance, are generally associated with little or no recombination, like in mammals (Qiao et al., 2012) and yeasts: in *Schizosaccharomyces pombe*, components of the RNA interference (RNAi) pathway repress DSB formation around centromeres (Ellermeier et al., 2010) and in *Saccharomyces cerevisiae*, Spo11 relocalises onto chromosome arms at prophase, thus preventing the formation of DSBs adjacent to centromeres (Kugou et al., 2009). This feature likely aids in the proper disjunction of homologues, since centromere-proximal COs result in aneuploidy in yeasts (Rockmill et al., 2006), humans (Hassold and Hunt, 2001) and flies (Koehler et al., 1996).

A similar suppression is also observed at telomeric regions in yeasts (Blitzblau et al., 2007; Buhler et al., 2007), possibly because DSBs in repetitive sequences are likely to be repaired through the non-allelic homologous recombination (NAHR) pathway which can alter genome architecture *via* chromosomal rearrangements (Sasaki et al., 2010). However, recombination seems increased in the neighbouring (subtelomeric) regions of yeasts (Chen et al., 2008; Barton et al., 2008) albeit this was not observed in other genome-wide studies (Buhler et al., 2001; Pan et al., 2011). High RRs are also observed in the subtelomeric regions of mammals (Kong et al., 2002; Jensen-Seaman et al., 2004; Pratto et al., 2014) and plants (Giraut et al., 2011).

In lieu of occurring at centromeres and telomeres, recombination primarily localises within interstitial regions, themselves fragmented into DSB-rich and DSB-poor domains — of about 100 kb in *Saccharomyces cerevisiae* (Baudat and Nicolas, 1997; Borde et al., 1999). The DSB-rich domains are associated with higher GC-content in yeasts (Gerton et al., 2000; Petes, 2001; Marsolier-Kergoat and Yeramian, 2009), rodents (Jensen-Seaman et al., 2004) and mammals (Eyre-Walker, 1993; Fullerton et al., 2001). In humans and chimpanzees, these domains are further enriched in 5' and 3' untranslated regions (UTRs) and CpG islands (Kong et al., 2002; Auton et al., 2012).

It was suggested early that these highly-recombinant regions may correspond to structural genes (Thuriaux, 1977), which is indeed the case in maize (Nelson, 1959, 1962, 1975; Dooner and Martínez-Férez, 1997; Dooner and He, 2008, reviewed in Okagaki et al., 2018). Notwithstandingly, neither *Arabidopsis thaliana* (Kim et al., 2007; Horton et al., 2012), *Schizosaccharomyces pombe* (Cromie et al., 2007) nor mammals (McVean et al., 2004; Myers et al., 2005; Brick et al., 2012) share this characteristic: in humans and mice, recombination correlates negatively with both gene content (Kong et al., 2002; Jensen-Seaman et al., 2004) and gene transcription rate (McVicker and Green, 2010; Pouyet et al., 2017).

Recently, Halldorsson et al. (2019) argued that the mechanism guiding recombination away from genes may have emerged through evolution in order to reduce the deleterious effect of its inherent *de novo* mutations (DNMs) on coding sequences. The mutagenicity of recombination was indeed demonstrated in yeasts (Strathern et al., 1995; Rattray et al., 2015) and humans (Arbeithuber et al., 2015; Halldorsson et al., 2019) and explained, — together with Hill-Robertson effects, — the correlations found between recombination and genetic diversity in humans (Nachman, 2001; Lercher and Hurst, 2002; Hellmann et al., 2003, 2005; Spencer et al., 2006; Montgomery et al., 2013; Smith et al., 2018) and other species (Begun and Aquadro, 1992; Aquadro, 1997; Webster and Hurst, 2012; Cutter and Payseur, 2013).

More generally, sites of recombination initiation seem to correspond to regions of open chromatin: highly active sites present trimethylation of the 4th lysine of histone H3 (H3K4me3) marks in yeasts (Borde et al., 2009) and mice (Buard et al., 2009) and DNA hypomethylation in plants (Maloisel and Rossignol, 1998; Melamed-Bessudo and Levy, 2012; Mirouze et al., 2012). Curiously though, in mammals, long-range recombination rates seem to be associated to DNA hypermethylation rather than hypomethylation (Sigurdsson et al., 2009; Zeng and Yi, 2014).

Nucleosome-depleted regions (NDRs) are another typical feature of open chromatin and recombinational activity is stronger at these sites in mammals (Getun et al., 2010; Lange et al., 2016; Yamada et al., 2017, reviewed in Jabbari et al., 2019)
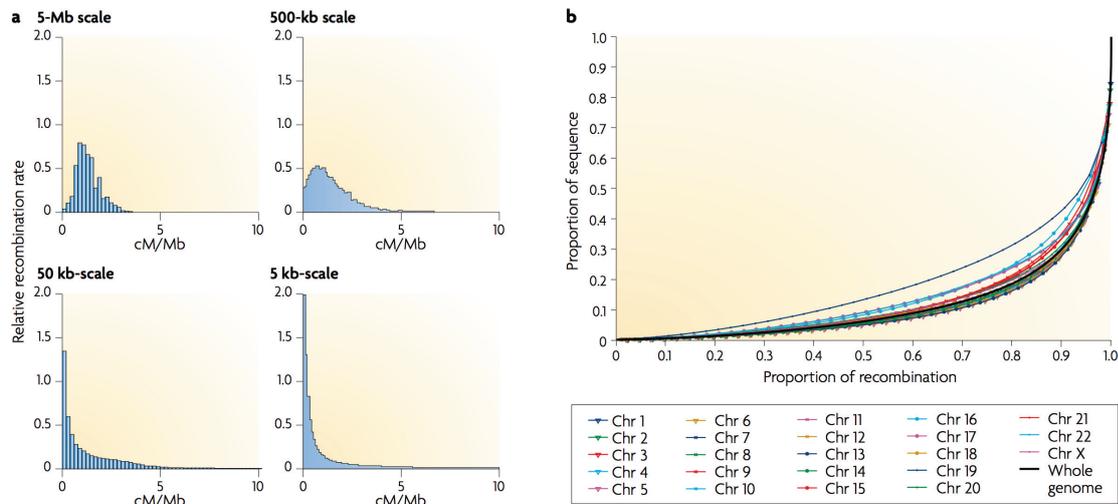
Figure 3.2: Heterogeneity in recombination rates along the human genome.
a | The shape of the distribution of recombination rates (RRs) depends on the level
of resolution. **b |** Most recombination events cluster in a small proportion of the total
genomic sequence.
This figure was reproduced from Coop and Przeworski (2007) and originally adapted from
(Myers et al., 2005) (permission in Appendix B).

as well as in *Schizosaccharomyces pombe* (de Castro et al., 2012) and *Saccharomyces
cerevisiae* (Wu and Lichten, 1994; Berchowitz et al., 2009) for which NDRs host
most DSBs. More precisely, recombination is found near transcription start sites
(TSSs) of gene promoters in budding yeasts (Baudat and Nicolas, 1997; Petes, 2001;
Mancera et al., 2008), dogs (Auton et al., 2013; Campbell et al., 2016), plants
(Hellsten et al., 2013; Choi et al., 2018) and birds (Singhal et al., 2015).

Recombination hotspots

The level of resolution matters tremendously when analysing patterns of RR
variation (reviewed in Smukowski and Noor, 2011). Indeed, at finer genomic
scales of 1–10 kb, recombination rates considerably vary (Figure 3.2.a.): in humans
(McVean et al., 2004; The 1000 Genomes Project Consortium, 2010) and other
eukaryotes (Mézard et al., 2015), 80% of recombination events gather in only 20%
of the genome (Figure 3.2.b.), primarily into 1—2-kb[6] regions called 'recombination
hotspots' (Myers et al., 2005).

---

[6]In mammals. But, in yeasts, recombination hotspots span several kilo base pairs.

Hotspots are generally defined as sequences that show a recombinational activity several times greater than the background rate (Crawford et al., 2004; Stapley et al., 2017). However, the activity of adjacent regions and the genome-wide average are alternately used as the comparative criterium (de Massy, 2013), which renders the delimitation and the number of hotspots slightly imprecise.

Nevertheless, apart from *Drosophila melanogaster* (Comeron et al., 2012; Manzano-Winkler et al., 2013), *Caenorhabditis elegans* (Kaur and Rockman, 2014) and *Apis mellifera* (Mougel et al., 2014; Wallberg et al., 2015) which lack them, recombination hotspots have been identified in a myriad of eukaryotes, including *Saccharomyces cerevisiae* (Sun et al., 1989b; Lichten and Goldman, 1995), *Schizosaccharomyces pombe* (Steiner and Smith, 2005; Cromie et al., 2007), *Arabidopsis thaliana* (Drouaud et al., 2006), *Zea mays* (Brown and Sundaresan, 1991; Dooner and Martínez-Férez, 1997; Yao et al., 2002; Fu et al., 2002), *Triticum aestivum* (Saintenac et al., 2011) and other plants (Mézard, 2006), *Canis lupus* (Axelsson et al., 2012), *Mus musculus* (Guillon and de Massy, 2002; Kauppi et al., 2007; Smagulova et al., 2011), *Pan troglodytes* (Winckler et al., 2005; Auton et al., 2012) and *Homo sapiens* (Jeffreys et al., 2001; Myers et al., 2005).

The first experimental evidence for hotspots was found serendipitously in the H2 region (i.e. major histocompatibility complex, MHC) of mouse chromosome 17 (Steinmetz et al., 1982). The first human hotspots were later identified in β-globin and insulin regions (Chakravarti et al., 1984, 1986). Since then, the list of recognised hotspots has grown extensively (reviewed in Arnheim et al., 2007; Paigen and Petkov, 2010) and many have been studied individually *via* sperm-typing studies (e.g. Hubert et al., 1994; Jeffreys et al., 2001; Schneider et al., 2002) (see Appendix A).

Later, genome-wide lists of hotspots — concordant with sperm-typing analyses (e.g. Tiemann-Boege et al., 2006) — have been achieved by analysing linkage disequilibrium in pedigrees or populations (see Subsections 3.1.1 and 3.1.2): about 30,000 have been uncovered in humans (Myers et al., 2005; The International

HapMap Consortium, 2007) and 47,000 in mice (Brunschwig et al., 2012).

Two additional layers of RR variation exist at the hotspot level in mammals. First, the recombinational activity of individual hotspots varies over orders of magnitude (Jeffreys et al., 2001; Kauppi et al., 2004; Paigen et al., 2008), with the number of hotspots per class of intensity following a negative exponential relationship (Paigen and Petkov, 2010). Second, the apparent[7] relative ratio of CO to NCO outcomes also varies between hotspots in flies (Singh, 2012), yeasts (Mancera et al., 2008), mice (Paigen et al., 2008) and humans (Jeffreys and May, 2004).

These relative differences in hotspot activity come from their both *cis-* and *trans-* regulations (reviewed in Paigen and Petkov, 2010) which also account for the differences in hotspot usage among individuals.

3.2.3   Inter-individual differences in hotspot usage

Sexual dimorphism

Sex differences in recombination were discovered over a century ago with the first linkage studies in *Drosophila melanogaster* (Morgan, 1912, 1914), *Bombyx mori* (Takana, 1914) and *Gammarus chevreuxi* (Huxley, 1928). Since then, several levels of sexual dimorphism have been unveiled.

First, as compared to males, the overall recombinational activity is greater in females[8] for most mammals (Dunn and Bennett, 1967) including mice (Shifman et al., 2006) and humans (Donis-Keller et al., 1987; Broman et al., 1998) — a result consistent with the fact that the genetic maps are longer in females than in males in these two species (Lynn et al., 2004; Cox et al., 2009) as well as in pigs (Mikawa et al., 1999), dogs (Neff et al., 1999) and thale cresses (Drouaud et al., 2007). In mammals, this observation could be partly due to the fact that female meiosis

---

[7]The density of polymorphic markers (which can vary across hotspots) affects the ability to detect NCOs. As such, the apparent CO:NCO ratio may differ from the genuine CO:NCO ratio.

[8]This feature (a species with different RRs in both sexes) is termed 'heterochiasmy'.

entails a dictyate arrest which can last for decades (from the fetal age to ovulation), thus leaving time for spontaneous DSBs to arise and to be repaired as complex COs (see Chapter 2). But it has also been argued that the synaptonemal complex (SC) length *per se* could play a major role in determining recombination rate differences, since the SC is much longer — and the DNA loops much shorter — in oocytes than in spermatocytes (Tease and Hultén, 2004). Of note, this effect is reversed in sheeps (Maddox et al., 2001), flycatchers (Backström et al., 2008) and most marsupials (Bennett et al., 1986; Hayman et al., 1988; Hayman and Rodger, 1990) and it not visible in one marsupial (Hayman et al., 1990) nor cattle (Kappes et al., 1997).

Second, sexual differences are regionalised: CO rates in men are several times lower near centromeres and higher near telomeres than in women (reviewed in Buard and de Massy, 2007), arguably because the SC is shorter in males (Tease and Hultén, 2004) and their synapsis preferentially initiates at subtelomeric regions (Brown et al., 2005). Contrariwise, females display more numerous interstitial initiation sites and their recombination landscape is thus generally flatter (Paigen et al., 2008).

Despite these sexual differences in hotspot usage — which can be so strong that a few hotspots are sometimes perceived as entirely sex-specific (Shiroishi et al., 1990, 1991), — nearly all hotspots are shared by both males and females (Bhérer et al., 2017).

Altogether, this sexual dimorphism mainly results from disparities in hotspot usage (Brick et al., 2018) possibly coming from haploid selection (Lenormand and Dutheil, 2005), imprinting (Lercher and Hurst, 2003) or sex-based differences in chromatin structure (Gerton and Hawley, 2005) and SC length (Petkov et al., 2007).

Heterogeneity between individuals

Hotspot usage is also variable between individuals of the same sex (reviewed in Popa, 2011 and Capilla et al., 2016).

In humans, fluctuations in recombination rates are greater between women than between men, but both sexes show inter-individual variation (Cheung et al., 2007). For instance, the major histocompatibility complex (MHC) shows a 2-fold

difference among 5 men (Yu et al., 1996), some hotspots are active in only a few men (Neumann and Jeffreys, 2006) and the CO:NCO ratio shows inter-individual disparities (Jeffreys and Neumann, 2005; Sarbajna et al., 2012).

As for mice, an inter-individual effect was also found in one strain (Koehler et al., 2002), but not in others. Thus, RRs vary not only between chromosomal regions and individuals, but also across populations and species, which indicates that they evolve with time, as reviewed in the following section.

3.3   Evolvability of recombination rates (RRs)

3.3.1   Intra- and inter-species comparison of fine-scale RRs

The comparison of human linkage disequilibrium (LD) maps has shown that LD blocks are highly correlated among populations (Gabriel et al., 2002), but the positions of the historical recombination hotspots they uncover are not entirely concordant with the one-generation recombination of genetic maps (Tapper et al., 2005). This non-concordance between historical and actual recombination was also observed independently at specific regions (Jeffreys et al., 2005; Kauppi et al., 2005) and suggests that the set of hotspots reorganises through time. Thus, discrepancies in the fine-scale RR should be found both within and among species.

On the one hand, recombination rates exhibit intra-species disparity. In mice, for instance, the number of MLH1 foci (a proxy for the number of COs) differs between strains (Koehler et al., 2002; Paigen et al., 2008; Baier et al., 2014) and, in humans, the use of recombination hostpots vary across populations (Berg et al., 2011; Hinch et al., 2011).

On the other hand, even though closely related species show similar average recombination rates (RRs) (Dumont and Payseur, 2008; Hassold et al., 2009; Garcia-Cruz et al., 2011; Auton et al., 2012) when compared over the scale of megabases (Mb), dissimilarities appear at finer scales, as was shown between humans and

macaques (Wall et al., 2003), between humans and chimpanzees (Ptak et al., 2004, 2005; Winckler et al., 2005) and between humans and great apes (Stevison et al., 2016).

The reasons for such a rapid turnover of recombination hotspots were understood about a decade ago with the discovery of the protein that determines the position of recombination hotspots in mammals: PRDM9.

3.3.2 *Prdm9*, the fast-evolving mammalian speciation gene

Discovery of the *Prdm9* gene

Positive regulatory (PR) domain zinc finger protein 9 (PRDM9) — encoded by a gene originally named *Meisetz* (for 'meiosis-induced factor containing PR/SET domain and zinc-finger motif') — was discovered in mouse germ cells as a histone H3 lysine 4 methyltransferase protein essential to the progression through meiotic prophase (Hayashi et al., 2005; Hayashi and Matsui, 2006). In 2010, three groups simultaneously identified it as responsible for the positioning of recombination hotspots in mice and humans (Baudat et al., 2010; Myers et al., 2010; Parvanov et al., 2010, reviewed in Cheung et al., 2010 and Hochwagen and Marais, 2010).

One of these groups had previously identified a degenerate 13-bp GC-rich motif (Myers et al., 2005) implicated in the activity of 40% of human hotspots (Myers et al., 2008; Webb et al., 2008) and had predicted that it was likely bound by a zinc finger protein of at least 12 units (Myers et al., 2008). Later, the computational analysis of all predicted zinc-finger DNA-binding proteins in the human genome yielded PRDM9 as both the only binding partner compatible with the observed degeneracy of the motif and the only candidate consistent with the lack of activity in chimpanzees (Myers et al., 2010).

The other two groups had previously independently identified a ∼5-Mb region on mouse chromosome 17 containing a *trans*-acting locus controlling the activation of

specific hotspots (Grey et al., 2009; Parvanov et al., 2009), respectively named *Dsbc1* and *Rcr1* at the time. Parvanov et al. (2010) used a mouse cross to narrow the interval down to 181 kb and argued that, among the four genes it comprised, *Prdm9* was the only relevant candidate that could explain the differences in hotspot usage.

Baudat et al. (2010) also reduced the interval with additional crosses to identify *Prdm9* as a relevant candidate. They further sequenced several human variants and found that the human *Prdm9* alleles were associated with hotspot usage, thus providing convincing evidence that it plays a major role in hotspot positioning, and demonstrated its sequence-specific binding to the 13-bp motif *in vitro*.

The dots were later reconnected with two past studies: one had found a haplotype associated with the control of recombination (Shiroishi et al., 1982) — this haplotype actually contained *Prdm9*; and in another, a protein binding a minisatellite motif had been partially purified (Wahls et al., 1991) — this protein turned out to be PRDM9 (Wahls and Davidson, 2011).

Since then, the role of PRDM9 in regulating the position of recombination hotspots has been confirmed multiple times in humans (Berg et al., 2010; Pratto et al., 2014) and observed in other primates (Groeneveld et al., 2012; Heerschop et al., 2016; Schwartz et al., 2014), rodents (Buard et al., 2014; Capilla et al., 2014; Kono et al., 2014), ruminants (Sandor et al., 2012; Ahlawat et al., 2016a,b, 2017) and equids (Steiner and Ryder, 2013).

Nevertheless, PRDM9 does not bind solely its specific binding motifs (Grey et al., 2017) and, in PRDM9-lacking mice, DSBs are located at functional sites (Brick et al., 2012). It has been proposed that DSB repair at such sites is inefficient and leads to sterilty (Brick et al., 2012) but a recent study proved that PRDM9 is *not* essential to fertility in male mice (Mihola et al., 2019). As for humans, a woman lacking a functional *Prdm9* allele was found to be fertile (Narasimhan et al., 2016). Hotspots are also defined independently of PRDM9 in canids (Axelsson

et al., 2012; Muñoz-Fuentes et al., 2011; Auton et al., 2013) and birds (Singhal et al., 2015) in which they instead locate at transcription start sites (TSSs) and are stable over evolutionary times.

Structure of the protein

PRDM9 determines the precise localisation of hotspots thanks to its carboxy-terminal tandem array of 8 to over 20 Cys_2-His_2 (C2H2) zinc fingers (Znf) (reviewed in Paigen and Petkov, 2018): the residues -1, +3 and +6 (relative to the alpha helix) of each Znf specify the DNA trinucleotide to bind and thus, altogether, the sequence target of the Znf array (Neale, 2010). A few fingers contribute preponderantly to the principal motif recognised (Figure 3.3.B.) and one Znf is separated from the rest of the array and closer to the central region (Figure 3.3.A.).

The central region also contains the histone methyltransferase PR/SET domain which is distantly related to the family of Suppressor of variegation 3–9, Enhancer of Zeste and Trithorax (SET) domains (reviewed in Grey et al., 2018). Thanks to this domain required for DSB formation (Diagouraga et al., 2018), PRDM9 can catalyse the mono-, di- and trimethylation of H3K4 and H3K36[9] (Wu et al., 2013; Powers et al., 2016) but also its own authomethylation (Koh-Stenta et al., 2017) which may help to regulate its activity by modulating the folding of the PR/SET domain.

The N-terminus hosts the Krüppel-associated box (KRAB)–related domain involved in protein:protein interactions (Parvanov et al., 2016, 2017; Imai et al., 2017), and a synovial sarcoma X repression domain (SSXRD). These two domains are also known to be involved in transcriptional repression (Margolin et al., 1994; Lim et al., 1998) but no such activity was identified in human PRDM9 (Born et al., 2014), and they both seem essential to the hotspot-targeting role of PRDM9 (Baker et al., 2017; Thibault-Sennett et al., 2018).

---

[9]H3K4, H3K36: Lysine 4 (resp. 36) of histone H3.
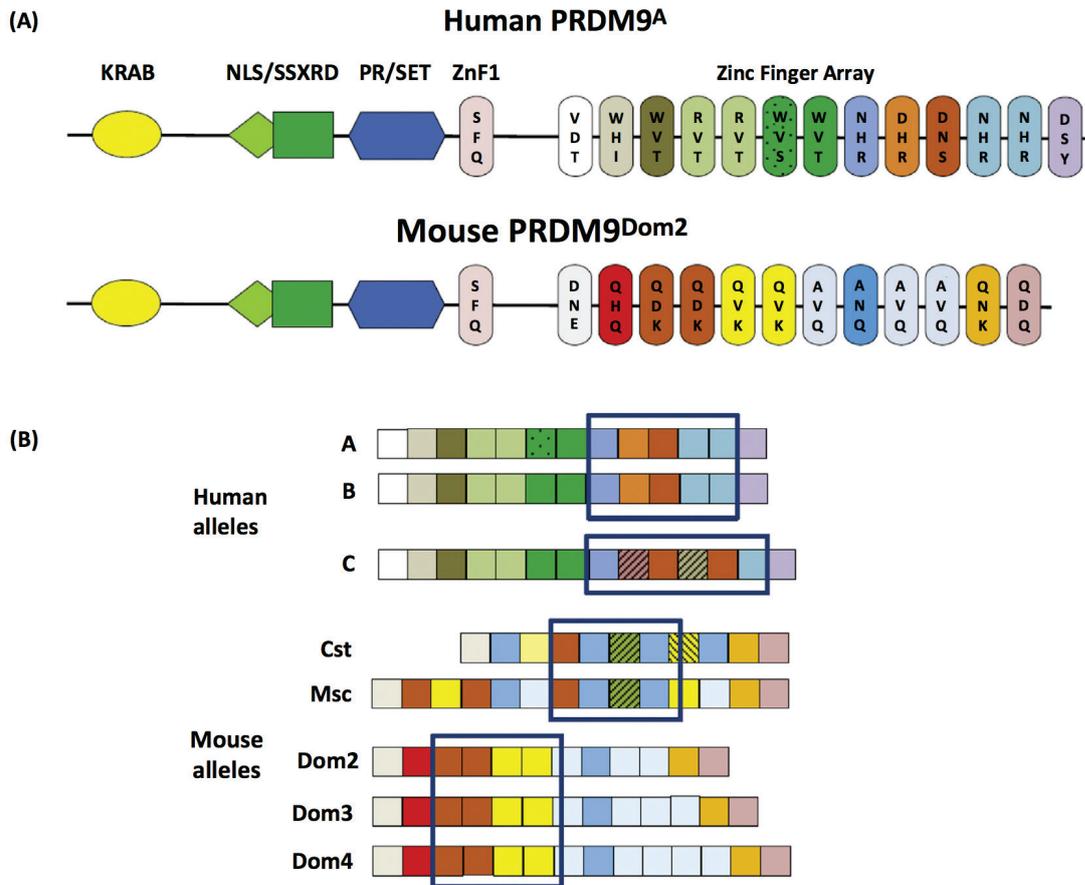
Figure 3.3: Molecular structure of PRDM9.
A | The PRDM9 protein consists of a KRAB-like, a SSXRD, a PR/SET and a zinc finger (Znf) array domains. The three residues (located at positions -1, +3 and +6 relative to the alpha helix) that are explicitely lettered specify the DNA target of each Znf. Human *A* and mouse *Dom2* alleles are shown. **B |** The composition of the tandem Znf array of the major human and mouse *Prdm9* alleles are represented as a sequence of squares, coloured based on the composition of residues at positions -1, +3 and +6. The boxes frame the fingers that contribute most to the principal motif of each allele.
This figure was reproduced from Paigen and Petkov (2018) (permission in Appendix B).

Multimerisation and hybrid sterility

PRDM9 has been proposed to act as a multimer (Baker et al., 2015b; Altemose et al., 2017; Schwarz et al., 2019) which may explain the dominance of certain alleles reported for human *C* over *A* (Pratto et al., 2014) and *I* over *A* alleles (Baudat et al., 2010), as well as mouse *13R* over *9R* (Brick et al., 2012) and *Cst* over *Dom2* alleles (Smagulova et al., 2011; Baker et al., 2015a,b).

Multimer formation certainly may play a role in PRDM9-mediated homologue pairing (Davies et al., 2016) and dominance may affect the dosage sensitivity of

PRDM9 (Flachs et al., 2012; Ségurel et al., 2011) and thus participate in both hybrid infertility — which was observed long before the known implication of *Prdm9* (Forejt and Iványi, 1974) — and in speciation (Mihola et al., 2009).

Given its critical role in fertility, one might expect PRDM9 to be under strong purifying selection and thus to be highly conserved. But, counterintuitively, it seems to evolve rapidly.

The rapid evolution of *Prdm9*

The Znf array forms a vast reservoir of variability since it may differ both in length (number of fingers) and composition, thus yielding extensive allelic possibilities for *Prdm9*.

Indeed, a large number of *Prdm9* alleles have been uncovered in primates (Groeneveld et al., 2012; Heerschop et al., 2016) and ruminants (Ahlawat et al., 2016a). As for mice, over 100 distinct alleles have been detected thus far (Buard et al., 2014; Kono et al., 2014). Most laboratory inbred strains derived from the *Mus musculus domesticus* subspecies carry either the *Dom2* or *Dom3* allele while those derived from *Mus musculus musculus* carry the *Msc* allele and those derived from *Mus musculus castaneus* the *Cst* allele (Figure 3.3.B.).

Human populations also vary in their PRDM9 allelic composition (Berg et al., 2010, 2011; Fledel-Alon et al., 2011): African populations have ∼50% of allele *A*, 13% of allele *C* and the rest composed of other minor alleles (Berg et al., 2011); non-African populations mainly encompass allele *A* and, to a smaller extent, allele *B* (Baudat et al., 2010; Berg et al., 2010; Hinch et al., 2011); and the Neanderthal and Denisovan samples studied so far exhibit yet other alleles (Schwartz et al., 2014; Lesecque et al., 2014).

Such great allelic diversity, which is associated with diversity in hotspot usage, is made possible by the high mutation rate of *Prdm9* (Jeffreys et al., 2013) and by the strong positive selection exerted on its decisive Znf residues (Oliver et al., 2009; Thomas et al., 2009; Ponting, 2011).

3.3.3 The Red Queen dynamics of hotspot evolution

DSB-induced biased gene conversion (dBGC) and the erosion of targets
Once PRDM9 has bound its allele-specific target, a DSB is initiated and subsequently
repaired as a CO or a NCO (see Chapter 2). In most hotspots studied, the
distribution of CO exchange points — which likely reflect the position of the
resolution of the transient Holliday junction rather than the DSB initiation site
(Smith, 2001) — decreases identically on the two sides (5' and 3') of the DSB
(Arnheim et al., 2007). However, a skewed CO exchange point distribution appeared
in a few hotspots (Jeffreys and Neumann, 2002, 2005; Yauk et al., 2003; Neumann
and Jeffreys, 2006) and was interpreted as a visible corollary of the differential DSB
initiation on the two homologues (Baudat and de Massy, 2007).

Indeed, PRDM9 can *a priori* bind its target on either homologue ('haplotype'
henceforth). However, if one haplotype has a higher PRDM9-binding affinity, it
hosts more DSBs and is thus 'hotter' (Zelazowski and Cole, 2016). Therefore, the
other, 'colder' haplotype is used as a template to repair the DSB, which results
in the hot haplotype being frequently converted by the cold one. This meiotic
initiation bias thus yields biased gene conversion (BGC) recombination events
and, since this phenomenon is induced by the preferential placement of DSBs on
one haplotype, I will henceforth call it 'DSB-induced BGC' (dBGC), as others
before (Lesecque, 2014; Grey et al., 2018).

A differential binding affinity between the two haplotypes arises when one
target motif acquires mutations: the more affinity-disruptive mutations the targeted
motif gains (i.e. the more eroded the hotspot), the more asymmetrically the DSBs
initiate (i.e. the more asymmetric the hotspot), and the stronger the dBGC effect
(reviewed in Tiemann-Boege et al., 2017).

The hotspot paradox
As just stated, during the repair of the DSB, the hot (recombination-activating)
haplotype is converted into the cold (recombination-suppressing) haplotype from the

other chromosome (Gutz, 1971; Schuchert and Kohli, 1988; Jeffreys and Neumann, 2009) and therefore suffers a meiotic drive against itself. Consequently, in the long-term, the very mechanism of recombination is expected to lead to the self-destruction of hotspots — a prediction that seems antipodal with the observation that hotspots are abundant in sexually active eukaryotes. This dilemma has been called the 'hotspot paradox' (Boulton et al., 1997): individually, hotspots are suicidal but, collectively, they are maintained.

Over the decade following the discovery of that paradox, several theoretical studies have been conducted to try and understand how hotspots are maintained despite their self-destruction (Boulton et al., 1997; Pineda-Krch and Redfield, 2005; Coop and Myers, 2007). Three main hypotheses were put forward by these studies to justify the maintenance of hotspots.

First, all three studies have proposed that recombination-activating back-mutations could arise in hotspots to counteract their extinction by dBGC. Though, all three conclude that the mutation rate required in face of the intensity of gene conversion would need to be unfeasibly large for them to be likely to be observed.

Second, the authors suggested that, given the benefits of recombination on fertility and viability, there could be a selective force opposing the spread of recombination-suppressing haplotypes: to ensure the correct segregation of alleles, recombination hotspot alleles could be directly selected for. However, for such a selective force to be strong enough to counterbalance hotspot extinction, DSBs would have to resolve into COs with a much higher probability than is observed in reality.

The third main hypothesis put forward was arguably the most plausible one: hotspots appear to compete for a finite amount of recombination with other adjacent hotspots. As such, it may be possible for them to increase their activity — and thus to start experiencing drive — only when nearby ones have been lost. This inter-hotspot competition drastically slowed down the expected rate of extinction. Still, it did not allow hotspots to persist indefinitely. As such, at that time, the mystery remained complete as to the way the paradox could be solved.

Determinants of the Red Queen dynamics

Further progress in solving the hotspot paradox came in 2010 with the discovery of PRDM9 as the determinant of hotspot localisation (Baudat et al., 2010; Myers et al., 2010; Parvanov et al., 2010). Indeed, it had been mentionned two years before that the hotspot paradox could theroretically be resolved if a *trans*-acting modifier (thus escaping gene conversion) had the ability to activate or inactivate the hotspots (Peters, 2008; Friberg and Rice, 2008).

Úbeda and Wilkins (2011) formally formulated the model involving PRDM9 as the *trans*-acting protein solving the paradox under the form of a race for evolution termed a 'Red Queen dynamics' (van Valen, 1973), after the words of the Red Queen in the *Through the Looking-Glass and What Alice Found There* book by Lewis Caroll (1871) (Figure 3.4).
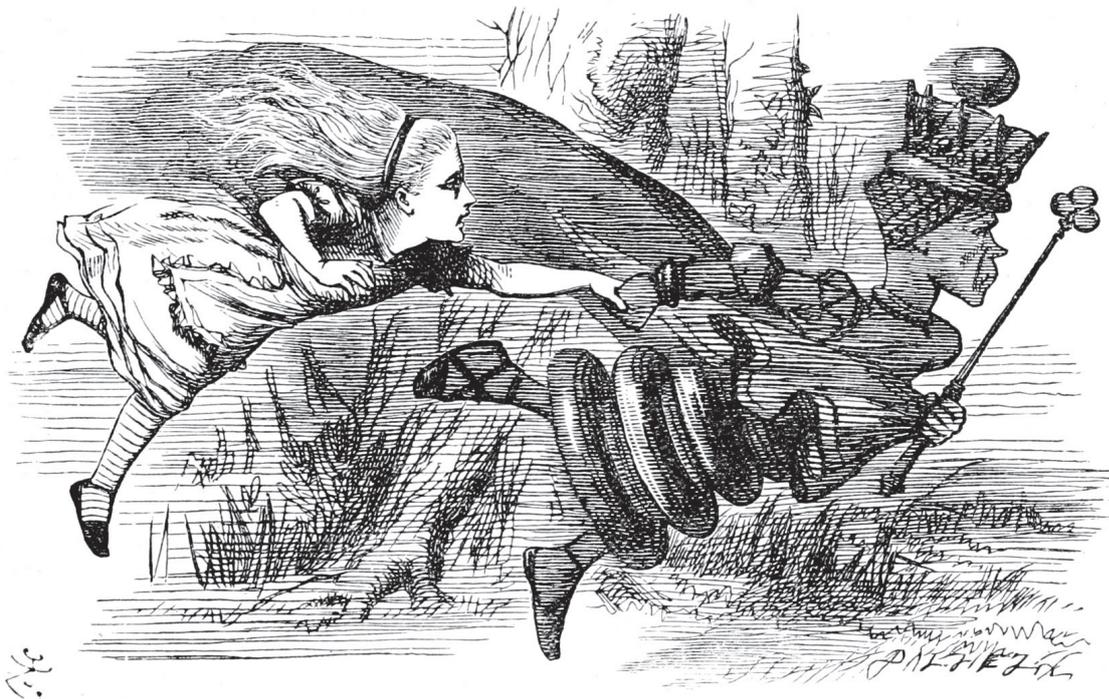
Figure 3.4: Original drawing of Alice and the Red Queen by John Tenniel.
The 'Red Queen dynamics' term is derived from a statement of the Red Queen in Lewis Caroll's *Through the Looking-Glass and What Alice Found There* (1871) about the nature of her world: *'Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!'.* This figure is free of rights and was reproduced from Carroll (1871).

In their model (Figure 3.5), owing to dBGC, the target loci lose their propensity to be bound by PRDM9, thereby reducing the overall recombination rate and creating a selective pressure for PRDM9 alleles to evolve and target a new set of binding sites. This intragenomic conflict leads to a never-ending situation where recombinogenic PRDM9 alleles continually chase their target motifs and evolve into other allelic variants as soon as their targeted sites are sufficiently eroded.

More recently, Latrille et al. (2017) formalised a quantitative population-genetic model accounting for all possible actors of the Red Queen model. Their mathematical developments led to the identification that both an extremely high mutation rate of PRDM9 and a strong dBGC eroding its target motifs are required for the model to be valid.

However, Ponting (2011) questioned this theory on the basis that the number of recombination hotspots ($\sim$25,000 in humans) far exceeds the number of chromosome arms ($\sim$40) and proposed four explanations justifying the strong and sustained positive selection on the DNA-binding determinant sites of PRDM9.

First, it could be that only a portion of the hotspots are bound by PRDM9 with strong affinity and that PRDM9 could evolve to keep a high binding affinity with a maximum number of these strong sites.

Second, since the PAR of sexual chromosomes is very short and is the only region where COs can form between these chromosomes, PRDM9 may be driven to evolve rapidly to ensure their correct segregation.

Third, if multiple weakly deleterious alleles accumulate in a non-recombining region, PRDM9 may be driven to evolve and target this particular region to break down the detrimental linkage in it.

Last, PRDM9 may evolve so as to prevent diseases, since increased CO rates in certain regions can lead individuals to certain diseases.
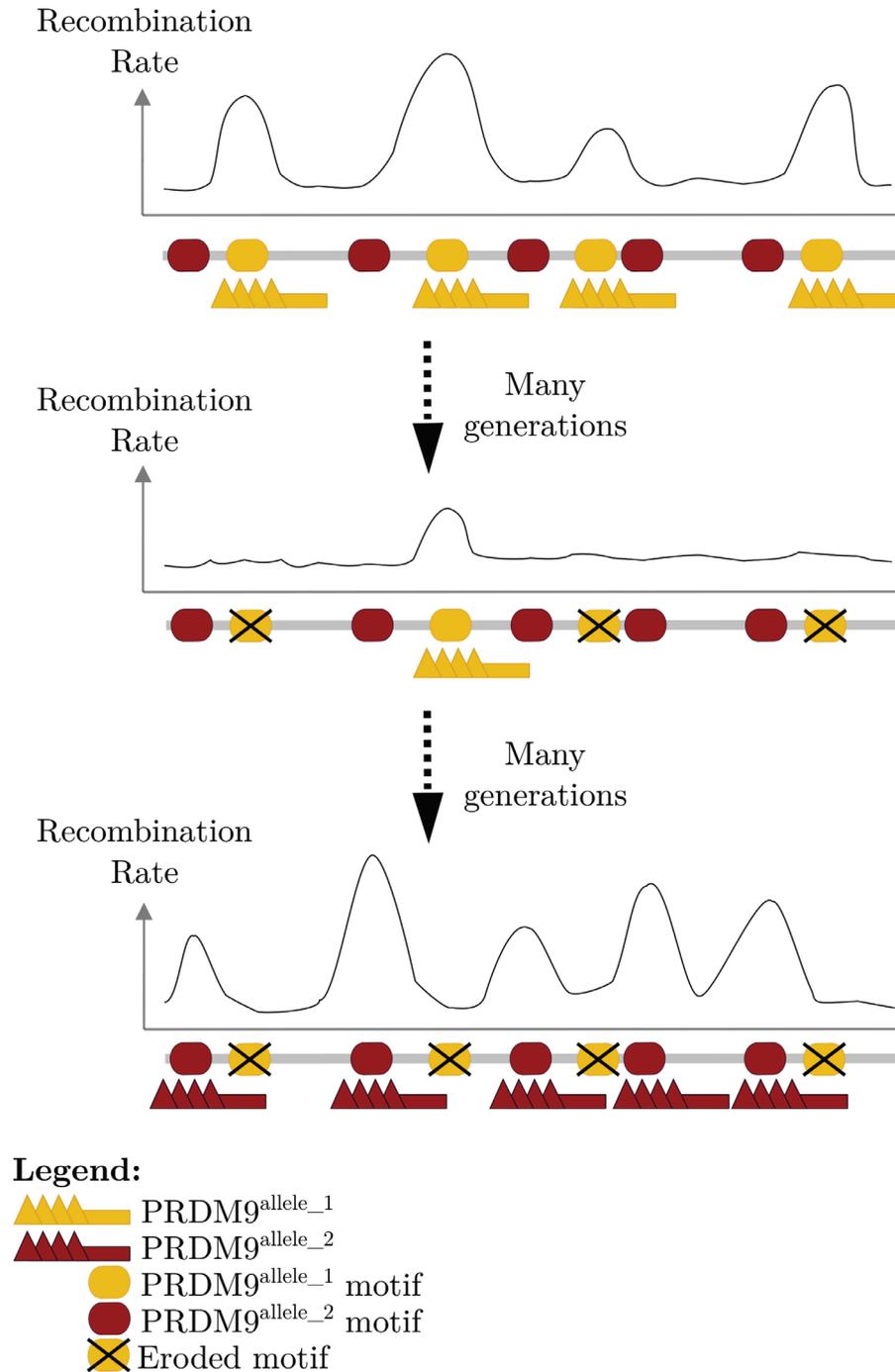
Figure 3.5: The Red Queen model of recombination hotspots.
In mice, the position of recombination hotspots, defined as regions of elevated recombination rate, is determined by PRDM9. At a given generation (top panel), one allelic variant of this protein, PRDM9$^{\text{allele\_1}}$, targets specifically its target motif (yellow square) thanks to its sequence-specific zing finger array (yellow triangles). Over time, because of double-strand break induced biased gene conversion (dBGC), the recombination-activating haplotypes carrying the target motif get eroded (crossed yellow square), which directly leads to a deprivation of hotspots as fewer sites are targeted by the PRDM9 allele present in the individual (middle panel). According to the Red Queen model of recombination hotspots, this creates a selective pressure for PRDM9 to evolve rapidly into a new allele, PRDM9$^{\text{allele\_2}}$, carrying a distinct zinc finger array (red triangles) targeting a new set of motifs (red square). As such, the recombination landscape with this new allele (bottom panel) is completely different from the one with the original allele (top panel).

Experimental proofs of the Red Queen model

Whichever the reason driving PRDM9 to evolve, all hypotheses rest on the following assumption of the Red Queen model: that the destruction of PRDM9 targets *via* dBGC is at the origin of the raise in frequency of new PRDM9 variants. Though, for this model to be plausible, dBGC must be strong enough to lead to a significant loss of hotspots genome-wide. Therefore, Lesecque et al. (2014) empirically quantified the dynamics of hotspot turnover by estimating the age and life expectancy of human hotspots. Their estimates showed that human hotspots were both much younger and much shorter-lived than had previsouly been suggested, and that dBGC was extremely high in certain hotspots. As such, they showed that dBGC was indeed sufficiently strong to explain the rapid loss of hotspots.

Further experimental testings of PRDM9 driving the evolutionary erosion of hotspots were carried in mice by Baker et al. (2015a). They indeed compared the activity of a *Prdm9* allele originating from the *Mus musculus castaneus* subspecies (*Prdm9*^{Cst}) in both *Mus musculus castaneus* and *Mus musculus domesticus*. They found that most variants affecting PRDM9^{Cst} binding had arisen specifically in the *Mus musculus castaneus* subspecies in which it had evolved and that hotspots had thus been greatly eroded in that lineage, which confirmed experimentally the predictions of the Red Queen model.

As a consequence of this haplotype difference, F1 hybrids between the two subspecies showed large haplotype biases in PRDM9 binding. The latter were sometimes so large that novel hotspots appeared in the hybrid, as a result of the interplay between one parent's *Prdm9* allele and the other parent's chromosome (for the hotspot on the 'self' chromosome had eroded).

Smagulova et al. (2016) further analysed the consequences of such sequence divergence generated by hotspot turnover in mouse hybrids and suggested that, because COs are disfavoured at the hotspots showing large haplotype biases, this may lead to reduced fertility and, ultimately, to speciation. The precise reasons why a shortage of symmetric hotspots can cause asynapsis remain to be elucidated, but it has been proposed that it may be due to a concomittant asymmetry in

PRDM9-dependent chromatin remodelling (Davies et al., 2016) or to an excessively high level of heterozygosity impeding recombination (Gregorova et al., 2018).

Altogether, DSB-induced biased gene conversion (dBGC) is an important driver for the evolution of the recombination landscape. Though, it is not the only one: another form of meiotic drive (GC-biased gene conversion, gBGC) also shapes the genome around recombination hotspots. I will review it in the following chapter.

4

Biased gene conversion, a major designer of genomic lanscapes

Contents

Gene conversion, i.e. the process through which one DNA sequence is cleaved and non-reciprocally replaced by another one (the homologue in the case of allelic gene conversion), leads to the non-Mendelian segregation of genetic information at the locus where it occured. If the two alleles are equally likely to be converted, this has no incidence at the population scale: allelic frequencies remain constant over generations. If, however, one homologue preferentially converts the other, it is more frequent in the pool of gametes and the transmission of alleles is necessarily

biased: the donor has an evolutionary advantage over the acceptor.

Such biased gene conversion (BGC) exists under two forms: DSB-induced BGC (dBGC) when the bias comes from a differential competency for homologues to host the double-strand break (see Chapter 3), and GC-biased gene conversion (gBGC) when it comes from the nature of the nucleotides involved. Indeed, the repair of the cut homologue involves the formation of heteroduplex DNA, i.e. a stretch of DNA where the two strands bear distinct alleles. These mismatches are either 'restored' if the original allele of the cut sequence is reinstated, or 'converted' if it is supplanted by the allele of the homologue. The position of these events delineate 'conversion tracts' (CTs) — which are designated as 'complex CTs' when they alternate conversion and restoration events (Borts and Haber, 1989) and 'simple CTs' otherwise.

In some species, whether through conversions or restorations, the repair favours GC over AT alleles (Mancera et al., 2008; Si et al., 2015; Williams et al., 2015; Halldorsson et al., 2016; Smeds et al., 2016), hence the term 'GC-biased gene conversion' (gBGC). Because its consequences on genome evolution ressemble those of natural selection, the very existence of this recently discovered phenomenon has been questioned by many in the more global context of the controversy opposing selectionists to neutralists (see Chapter 1). I will therefore start this chapter by reviewing the breakthrough of gBGC in the climate of this debate, then explore the similitudes of its implications for genome evolution with those of natural selection and finish by looking into the first studies that characterised it.

4.1   Discovery of GC-biased gene conversion (gBGC)

4.1.1   The debated origin of isochores

In double-stranded DNA, adenosine (A) and thymine (T) nucleotides pair up while cytosine (C) nucleotides mate guanine (G) bases (Chargaff, 1950, reviewed in Kresge et al., 2005). Therefore, when studying the composition of a stretch of DNA, it is conventional to measure its GC-content.

Originally, this was done *via* the ultra-centrifugation of DNA fragments (Meselson et al., 1957; Corneo et al., 1968). Using this technique, a few studies have characterised GC-content distribution in several eukaryotes (Filipski et al., 1973; Thiery et al., 1976; Macaya et al., 1976, 1978; Cortadas et al., 1977) and revealed that mammalian, avian and reptilian genomes — but not amphibians nor fishes (Bernardi and Bernardi, 1990) — display a long-range compositional heterogeneity (Figure 4.1). The long regions of 100 kb or more that carry a relatively homogeneous GC-content were later termed 'isochores' (Cuny et al., 1981).

GC-rich isochores are enriched in genes (Bernardi et al., 1985; Mouchiroud et al., 1991; Lander et al., 2001, reviewed in Bernardi, 2005) that are shorter and more compact than in GC-poor regions (Duret et al., 1995). Regional GC-content further correlates with the timing of DNA replication (Federico et al., 1998; Watanabe
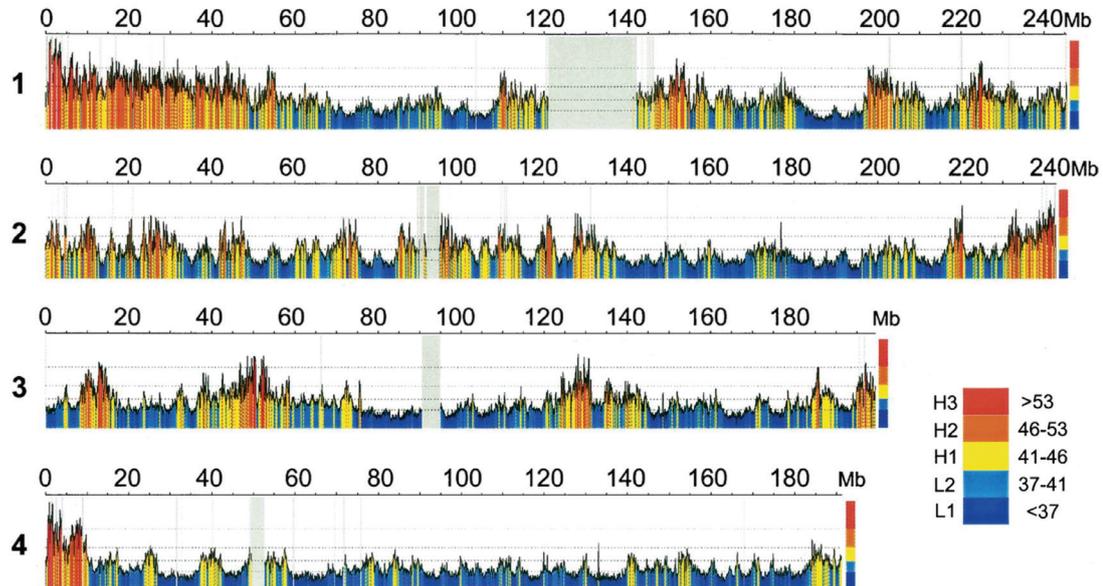
Figure 4.1: Overview of isochores on four human chromosomes.
Human chromosomes 1, 2, 3 and 4 are divided into 100-kb windows coloured according to their mean GC-content: the spectrum of GC-level was divided into five classes (indicated by broken horizontal lines) from ultra-marine blue (GC-poorest L1 isochores) to scarlet red (GC-richest H3 isochores). Grey vertical lines correspond to gaps present in the sequences and grey vertical regions to centromeres.
This figure was reproduced from Costantini et al. (2006) and corresponds to a subsample of the original figure (permission in Appendix B).

et al., 2002; Costantini and Bernardi, 2008), the density in transposable elements (TEs) (Smit, 1999; Lander et al., 2001; Mouse Genome Sequencing Consortium et al., 2002) and the recombinational activity (Fullerton et al., 2001; Kong et al., 2002).

Since base composition of homologous genomic regions correlate between the three amniotic lineages (mammals, birds and reptiles) (Kadi et al., 1993; Cacciò et al., 1994; Hughes et al., 1999), it is thought that isochores were inherited from their last common ancestor (LCA). Since then, certain lineages have undergone additional somehow steep changes. For instance, the isochore GC-content of mice is less variable than that of other mammals — a pattern that is in the derived state as compared to nonrodents (Galtier and Mouchiroud, 1998) and which likely reflects one (Mouchiroud et al., 1988) or two (Smith and Eyre-Walker, 2002) extra 'murid shifts' since the LCA.

Originally, two main hypotheses had been proposed as for the origin of isochores (reviewed in Duret and Galtier, 2009a). According to the mutational bias hypothesis, isochores would be caused by a variation along chromosomes in the mutational bias towards either AT or GC nucleotides (Filipski, 1988; Wolfe et al., 1989; Francino and Ochman, 1999; Fryxell and Zuckerkandl, 2000). If this were true, GC → AT and AT → GC mutations should have the same probability of fixation at neutral sites. The finding that this was not the case (Eyre-Walker, 1999; Smith and Eyre-Walker, 2001; Lercher et al., 2002; Webster and Smith, 2004; Spencer et al., 2006) ruled out this theory.

Another proposition involving natural selection has been thoroughly defended by one of the major discoverers of isochores (Bernardi, 2000, 2007, 2012). In his view, the fact that G and C bases are linked *via* three hydrogen bonds (instead of two for A and T bases) would compensate for the purportedly instable nature of DNA in warm-blooded animals. However, this does not explain why only a fraction of the genome is affected by higher GC-content (Duret and Galtier, 2009a). This theory was further invalidated by the facts that no correlation between body

temperature and GC-content was found (Belle et al., 2002; Ream et al., 2003) and that this isochore organisation also takes place in cold-blooded animals like reptiles (Hughes et al., 1999; Hamada et al., 2003; Costantini et al., 2016). In addition, a scenario according to which all sites are under selection has theoretical limitations: given the elevated rate of deleterious mutations in their protein-coding sequences (Eyre-Walker and Keightley, 1999; Keightley and Eyre-Walker, 2000), mammalian genomes would probably accumulate a mutation load too high to be coped with (Eyre-Walker and Hurst, 2001).

An alternative role for natural selection in causing isochore organisation would be its fine-tuning the expression of tissue-specific genes (Vinogradov, 2003, 2005). This hypothesis may not hold, though, since the correlation between GC-content and gene expression is extremely weak (Sémon et al., 2005, 2006; Pouyet et al., 2017, reviewed in Duret and Galtier, 2009a).

Since natural selection thus seems insufficient to explain, on its own, the bias towards the fixation of *GC* alleles, another track has been considered: GC-biased gene conversion (gBGC).

4.1.2 An alternative causation: the gBGC model

The excess of AT $\rightarrow$ GC substitutions in a context where GC $\rightarrow$ AT mutations are preponderant can be explained in two non-mutually exclusive ways: either because of non-stationarity (i.e. the GC-content in GC-rich isochores would still be decreasing) or because of GC-biased gene conversion (gBGC). This hypothesis, initially mentioned by Holmquist (1992) and Eyre-Walker (1993, 1999), has been promoted by Galtier et al. (2001).

The latter model originates from the observation that the mismatch repair (MMR) system — the main pathway active during recombination to correct base misalignments (Alani et al., 1994; Nicolas and Petes, 1994, reviewed in Evans and Alani, 2000 and Spies and Fishel, 2015) which is also involved in the mending of
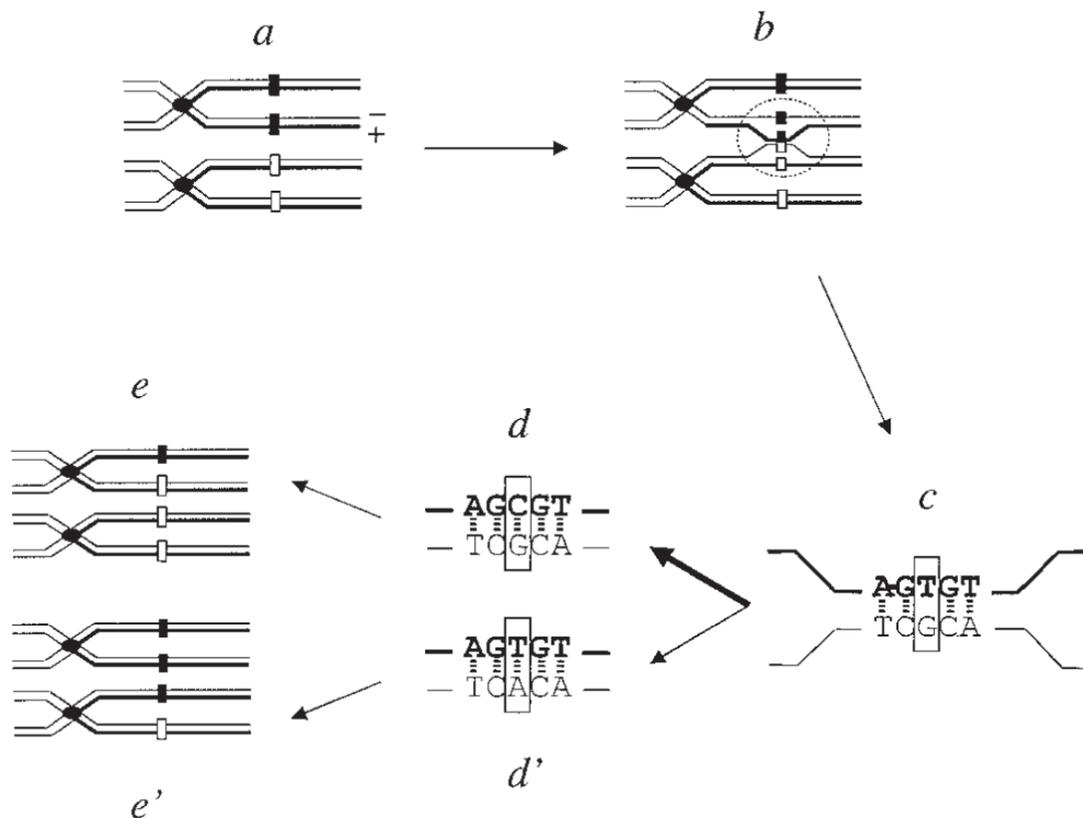
Figure 4.2: Gene conversion during a recombination event involving a strong (G or C) *versus* a weak (A or T) base mismatch.
A pair of homologous chromosomes displaying a heterozygous site with a G:C pair represented as a black rectangle and a A:T pair as a white rectangle (**a**) undergoes a recombination event which materialises as a heteroduplex (**b**) containing a T:G mismatch (**c**). The T:G mismatch is repaired and results either in a G:C (**d**) or a A:T (**d'**) pair which yield a non-Mendelian segregation of alleles (**e** and **e'**). This has an incidence at the population-scale if the repair towards G:C (**d**) or A:T (**d'**) is more frequent than the other one. It is called GC-biased gene conversion (gBGC) in the particular case where the repair towards G:C (**d**) occurs more often.
This figure was reproduced from Galtier et al. (2001) (permission in Appendix B).

base misincorporations during DNA replication (Surtees et al., 2004) — may favour

G and *C* alleles (Brown and Jiricny, 1988; Bill et al., 1998). (Figure 4.2).

   A predictable consequence of such alteration in the frequency of transmission of

G and *C* alleles is the long-term evolution of base composition in regions undergoing

gBGC. Though, at the time, one major argument against the gBGC model was

that there was only a one-order-of-magnitude range of parameters for which the

rate of biased gene conversion would be sufficiently high to alter polymorphism patterns significantly but remain sufficiently low not to induce an extreme base composition (Eyre-Walker, 1999). This objection was adressed by Duret and Arndt (2008) who found that the gBGC model explains well the relationship between recombination and substitution rates. Indeed, considering that gBGC acts only at recombination hotspots, the substitution rate increases greatly at these loci, but stops before their GC-content reaches 100%, because hotspots generally have a short lifespan (Ptak et al., 2005; Winckler et al., 2005). In particular, as soon as a hotspot gets inactivated, its GC-content should start decreasing, consistently with what has been observed in the GC-rich regions[1] of primates (Duret et al., 2002; Belle et al., 2004; Meunier and Duret, 2004; Duret, 2006).

gBGC also provides an explanation for the higher heterogeneity of GC-rich isochores (Clay et al., 2001; Clay and Bernardi, 2001): since recombination hotspots would locally display higher GC-levels than the genome-wide average, hotspot-dense regions would exhibit a particularly disparate GC-content.

Another objection to gBGC (Eyre-Walker, 1999) came with the observation that GC-content at the synonymous third position of codons (GC_3) is generally greater than intronic GC-content (Clay et al., 1996). But Duret and Hurst (2001) provided an explanation compatible with gBGC to this observation: assuming that transposons are GC-poorer than the GC-rich regions of the genome, their accumulation within introns (but not exons) would justify such difference between intronic GC-content and GC_3.

Altogether, the presence of isochores seems to fit theoretically with gBGC (Duret et al., 2006). But, under the gBGC hypothesis, a number of other consequences are expected and their footprints can be researched in genomes.

---

[1]According to the gBGC hypothesis, GC-rich regions are those that host the hotspots.

4.1.3   Footprints of gBGC in mammalian genomes

One strong prediction of the gBGC model is that highly recombining regions should be GC-rich, which happens to be the case in several instances.

For example, components of the genome that undergo ectopic gene conversion (i.e. conversion between copies of a gene family) — like transfer RNAs (tRNAs), introns of ribosomal RNAs (rRNAs) (Galtier et al., 2001), human and mouse major histocompatibility complex (MHC) regions (Högstrand and Böhme, 1999) and other gene families (Backström et al., 2005; Galtier, 2003; Kudla et al., 2004) — are all GC-richer than the rest of the genome.

The human pseudoautosomal region (PAR) of X and Y chromosomes — the only portion of male sexual chromosomes which has homology and therefore recombines — provides another example of the association between recombination and GC-content. Indeed, given its short size, the per-nucleotide recombination rate (RR) of the PAR is much higher than that of autosomes (Soriano et al., 1987), while the non-PAR sections of sex chromosomes recombine even less (X chromosome) or not at all (Y chromosome). Under the gBGC model, the average GC_3 of these four genomic domains is expected to increase with their RR — which, as a matter of fact, is the case (Galtier et al., 2001).

This relationship between recombination and GC-content is really impressive in the *Fxy* gene that has been translocated onto the boundary of the mouse PAR a few million years ago: as compared to its X-linked portion, the PAR-side part of *Fxy* has undergone an acceleration in substitution rates (Perry and Ashworth, 1999) together with a strong increase in GC-content at both coding and non-coding positions (Montoya-Burgos et al., 2003; Galtier and Duret, 2007) — a finding that is consistent with gBGC occuring at the highly recombining PAR-side of the gene. Surprisingly however, the *XG* gene overlapping the PAR boundary of primates does not show the same pattern (Yi et al., 2004). Nevertheless, this observation does not necessarily conflict with the gBGC model: if *XG* was wholly located within the PAR before displacing onto its boundary, — or rather, before the PAR boundary displaces onto the gene, since the mammalian PAR gradually erodes

(Lahn and Page, 1999; Marais and Galtier, 2003), — it would have accumulated a high GC-content and would now be undergoing a slow, mutation-driven decrease in GC-content that would not be detectable yet (Galtier, 2004).

At the genome-wide scale, GC-content correlates positively with recombination rate in many eukaryotes (Pessia et al., 2012) including yeasts (Gerton et al., 2000; Birdsell, 2002), nematodes and flies (Marais et al., 2001, 2003; Marais and Piganeau, 2002), birds (International Chicken Genome Sequencing Consortium, 2004; Mugal et al., 2013), turtles (Kuraku et al., 2006), paramecia (Duret et al., 2008), algae (Jancek et al., 2008), plants (Glémin et al., 2006) and humans (Fullerton et al., 2001; Yu et al., 2001; Meunier and Duret, 2004; Khelifi et al., 2006; Duret and Arndt, 2008).

But, since the evolution of GC-content is relatively slow as compared to that of recombination rates in mammalian clades, it has been claimed that these estimates should be measured on similar time scales to be correctly compared (Duret and Galtier, 2009a). To do this, the stationary GC-content (GC^*), i.e. the GC-content that sequences would reach at equilibrium if patterns of substitution remained constant over time, is generally used. Under the assumption that all sites evolve independently from one another (Sueoka, 1962), this statistic can be calculated as:

$$GC^* = \frac{u}{u+v}$$

where u and v represent respectively the AT $\rightarrow$ GC and the GC $\rightarrow$ AT substitution rates. But, because the latter assumption is not valid in vertebrates where the mutation rate of a given base depends on the nature of the neighbouring bases[2], Duret and Arndt (2008) used a maximum likelihood approach to improve the estimation of GC^* and showed that it correlated better with recombination rate than with the observed GC-content (Figure 4.3). This further suggests that recombination acts upon GC-content, and not the other way round, as was proposed by Gerton et al. (2000), Blat et al. (2002) and Petes and Merker (2002).

---

[2]For instance, CpG sites (i.e. CG dinucleotides) are hypermutable (Arndt et al., 2003).

In past primate lineages, GC* also correlates well with the historical recombination rate (Munch et al., 2014).

These correlations between GC-content and recombination appear to be greater in males than in females for several species including mice, dogs and sheeps (Popa et al., 2012) as well as humans (Webster et al., 2005; Dreszer et al., 2007; Duret and Arndt, 2008). Since chiasmata persist many years in females (Coop and Przeworski, 2007), it is possible that the repair of mismatches proceeds differently in the two sexes, which could explain the seemingly male-specific gBGC (Duret and Galtier, 2009a). Alternatively, the sex-specific strategies for the distribution of recombination events along chromosomes (and more specifically, as a distance to telomeres) seem to account for this difference between males and females (Popa et al., 2012).
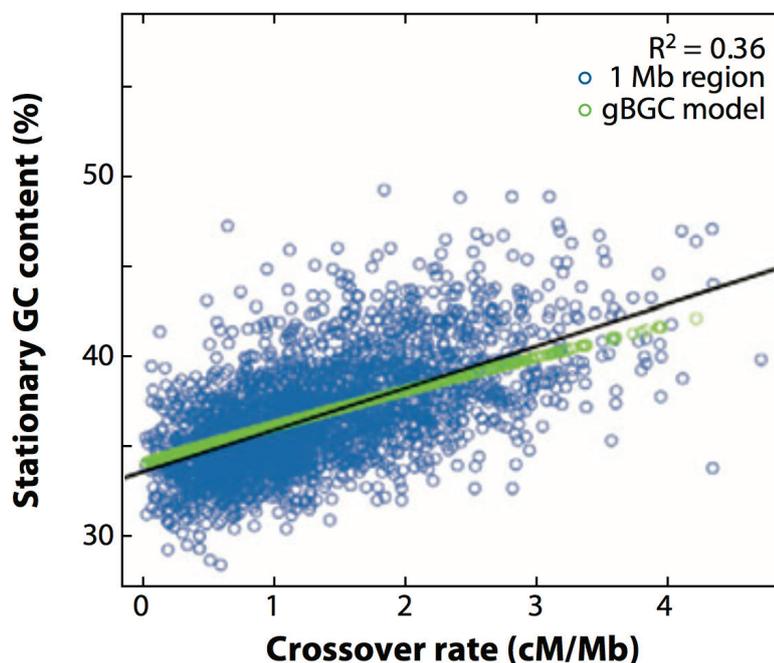
Figure 4.3: Correlation between the stationary GC-content (GC*) and the crossover rate (cM/Mb) in human autosomes.
Each dot corresponds to a 1-Mb-long genomic region. Green dots correspond to the predictions of the gBGC model.
This figure was reproduced from Duret and Galtier (2009a) and originally adapted from Duret and Arndt (2008) (permission in Appendix B).

4.2 Interference with natural selection

Several of the aforementioned observations supporting gBGC would also be predicted under a natural selection model. For instance, since linkage reduces the efficacy of selection (Hill and Robertson, 1966), a correlation between GC-content and recombination rate would be expected if there was a very high selection coefficient in favour of *GC* alleles (Galtier et al., 2001). More generally, the dynamics of the fixation process for one locus is identical no matter which of the two forces (biased gene conversion or natural selection) is responsible for it (Nagylaki, 1983), which explains why the first observations were initially interpreted as resulting from natural selection (e.g. Eyre-Walker, 1999). In this section, I review a few case studies in which such confounding patterns between gBGC and natural selection exist.

4.2.1 The case of codon usage bias (CUB)

Codon usage bias (CUB) corresponds to the observation that the frequency of use of synonymous codons (i.e. sequences of three nucleotides coding for the same amino acid (AA)) can vary across or within species (Fitch, 1976). Both adaptative (natural selection) and non-adaptative (mutation (Marais and Duret, 2001) or biased gene conversion) forces account for CUB (Bulmer, 1991; Sharp et al., 1993; Akashi and Eyre-Walker, 1998), but there remains a controversy about the quantitative contribution of each of these mechanisms to CUB (Pouyet, 2016).

In *Drosophila*, the CUB of each gene is correlated to transfer RNA (tRNA) content (Akashi, 1994; Duret and Mouchiroud, 1999; Bierne and Eyre-Walker, 2006; Behura and Severson, 2011), particularly for genes that are highly expressed (Chavancy et al., 1979; Shields et al., 1988; Moriyama and Powell, 1997; Hey and Kliman, 2002). This association between CUB and gene expression also holds true in *Caenorhabditis* (Duret and Mouchiroud, 1999; Castillo-Davis and Hartl, 2002; Marais and Piganeau, 2002), *Daphnia* (Lynch et al., 2017), *Arabidopsis* (Duret

and Mouchiroud, 1999; Wright et al., 2004), *Oryza* (Muyle et al., 2011) and single-celled organisms like *Giardia* (Lafay and Sharp, 1999), *Saccharomyces* (Bennetzen and Hall, 1982; Akashi, 2003; Harrison and Charlesworth, 2011), *Dictyostelium* (Sharp and Devine, 1989) and bacteria (Gouy and Gautier, 1982; Ikemura, 1985; Sharp and Li, 1987). This has been interpreted as 'translational selection': the coevolution of tRNA content with codon usage would increase either the accuracy or the efficiency of translation (Sharp et al., 1995; Duret, 2002). Though, other processes, like messenger RNA (mRNA) stability, protein folding, splicing regulation and robustness to translational errors could also play a role (Chamary et al., 2006; Cusack et al., 2011; Plotkin and Kudla, 2011, reviewed in Clément et al., 2017).

In contrast, in lowly recombining regions of *Drosophila* (Kliman and Hey, 1993) and in species with small effective population size (Ne) (Subramanian, 2008; Galtier et al., 2018), like mammals (Urrutia and Hurst, 2003; Comeron, 2004; Lavner and Kotlar, 2005), selection for codon usage remains weak. Instead, in mammals, codon usage is primarily governed by variations in GC-content (Sémon et al., 2006; Rudolph et al., 2016; Pouyet et al., 2017), which implies that gBGC could be one of the drivers of CUB in that clade. In *Drosophila* too, even if selection on codon usage predominates (Zeng and Charlesworth, 2009, 2010; Zeng, 2010), gBGC could also participate to CUB. Indeed, one peculiar feature of codon usage in this species is that, for all 20 amino acids (AAs), the preferred codon systematically ends with a G or a C nucleotide (reviewed in Duret and Galtier, 2009a). Even if the reason for this remains unknown, the finding that the base composition of the third position of 4-fold degenerate[3] codons is similar to that of non-coding regions (Clay and Bernardi, 2011) indicates that the patterns of CUB could (at least partly) come from evolutionary processes influencing base composition irrespectively of translational selection — such as gBGC (Duret, 2002; Galtier et al., 2006; Lynch, 2007, but see Jackson et al., 2017). A similar observation made in plants was also interpreted as the consequence of gBGC (Clément et al., 2017).

---

[3]A codon is said to be *n*-fold degenerate if *n* distinct three-nucleotide sequences result in the same amino acid (AA).

4.2.2   The case of human accelerated regions (HAR)

gBGC has also been mistaken for positive selection in fast-evolving regions specific to the human genome (reviewed in Duret and Galtier, 2009a). Such regions, — named human accelerated regions (HAR) or human accelerated conserved non-coding sequences (HACNS), — have been searched by several groups (Pollard et al., 2006a,b; Prabhakar et al., 2006; Bird et al., 2007; Bush and Lahn, 2008; Lindblad-Toh et al., 2011) in a quest to find the molecular adaptations that make the human genome distinct from other mammals.

HARs have first been interpreted as resulting from positive selection (reviewed in Hubisz and Pollard, 2014) but, because they harbour an excess of AT $\rightarrow$ GC substitutions, gBGC has been proposed as an alternative origin for these accelerated sequences (Galtier and Duret, 2007; Berglund et al., 2009; Duret and Galtier, 2009b; Katzman et al., 2010; Ratnakumar et al., 2010). And indeed, about one fifth of HARs seem to have evolved under gBGC alone (Kostka et al., 2012).

Thus, altogether, gBGC mimics natural selection in terms of consequences on the nucleotidic sequence (Bhérer and Auton, 2014), and this is likely to bring biases to molecular evolution and phylogenomics analyses (Berglund et al., 2009; Ratnakumar et al., 2010; Webster and Hurst, 2012; Romiguier et al., 2013, 2016; Romiguier and Roux, 2017; Bolívar et al., 2018, 2019; Rousselle et al., 2019). Consequently, prior to concluding that positive selection explains a given observation, one should check that the extended null hypothesis of molecular evolution (i.e. both the neutral and the gBGC models) has been rejected (Galtier and Duret, 2007; Duret and Galtier, 2009a). To check for this, three observations should be taken into consideration: first, whether AT $\rightarrow$ GC substitutions are preponderant; second, whether the studied locus is in a highly recombining region; and third, whether both functional and non-functional sites are affected. Whenever all three criteria are met, gBGC remains a likely explanation for any observed acceleration in substitution rates.

But, if gBGC interferes with natural selection, what happens when both forces drive evolution in the opposite direction?

4.2.3 The deleterious effects of gBGC

The AT → GC mutations whose fixation is favoured by gBGC can be either beneficial, inconsequential or detrimental to the fitness of the individual carrying it. To quantify the fate of all these categories of mutations in presence of gBGC, Duret and Galtier (2009a) performed simulations with characteristics close to those of human populations (in terms of effective population size and mutation rate) and showed that gBGC mainly favours the fixation of slightly deleterious and neutral AT → GC mutations.

Analysing the ratio ($\frac{d_N}{d_S}$) of the rate of nonsynonymous[4] (d_N) over that of synonymous[5] substitutions (d_S) in exon-specific episodes of accelerated amino acid (AA) evolution, Galtier et al. (2009) demonstrated that gBGC has been sufficiently strong to outdo the effect of purifying selection[6] and promote, instead, the fixation of deleterious AT → GC mutations within proteins. In wheat too, the accumulation of deleterious AT → GC mutations shown by the analysis of $\frac{d_N}{d_S}$ has been interpreted as originating from gBGC (Haudry et al., 2008). More generally, gBGC maintains deleterious mutations associated to human diseases (Necşulea et al., 2011; Capra et al., 2013; Lachance and Tishkoff, 2014; Xue et al., 2016).

But, if gBGC prejudices fitness, how come it has persisted over evolutionary times? This question remains open as of today, but it has been claimed that gBGC could somehow counterbalance the mutational load (Bengtsson, 1986; Marais, 2003; Glémin, 2010; Arbeithuber et al., 2015) which favours GC → AT mutations in both eukaryotes (Lynch, 2010) and procaryotes (Hershberg and Petrov, 2009). Alternatively, gBGC has been proposed to be a meiotical side-effect of the GC-biased base excision repair (BER) mechanism which is crucial in mitosis to reduce the number of somatic mutations (Marais and Galtier, 2003; Lesecque, 2014). Though, a study aiming at characterising gBGC in *Saccharomyces cerevisiae* ruled out the latter hypothesis for yeasts (Lesecque, 2014).

---

[4]A nonsynonymous substitution does not modify the amino acid (AA) produced.

[5]A synonymous substitution modifies the amino acid (AA) produced.

[6]Purifying selection (or negative selection) is the selective removal of deleterious alleles.

4.3 Characterisation of gBGC

Understanding the still-blurry reason for the evolutionary maintenance of gBGC requires to better quantify it in living beings and characterise its relationship with other parameters of genome evolution. I review the knowledge acquired so far on this topic in the last section of this chapter.

4.3.1 Quantification *via* site frequency spectra (SFS)

Fundamentally, gBGC shifts the allelic frequency of strong (S) (i.e. G and C) and weak (W) (i.e. A and T) bases, since it favours the fixation of the former and hinders that of the latter. Thus, comparing the distribution of the derived allele frequency (DAF) of S bases arising from W $\to$ S (WS) mutations and of W bases arising from S $\to$ W (SW) mutations can allow to estimate the intensity of gBGC.

In practice, this is done by analysing site frequency spectra (SFS), a.k.a. derived allele frequency spectra (DAFS). Indeed, because the SFS provides the number of SNPs for each class of frequency, it summarises the information in a much more detailed manner than any existing statistics (such as the GC_3 content in the case of gBGC, the ratio of non-synonymous over synonymous diversity $\left(\frac{\pi_N}{\pi_S}\right)$ in the case of polymorphism, or the ratio of non-synonymous over synonymous substitutions $\left(\frac{d_N}{d_S}\right)$ in the case of divergence) (Rousselle, 2018).

In the particular case of gBGC, the spectra for WS and SW mutations must be compared. This requires to polarise mutations from the ancestral to the derived state, thanks to an outgroup[7] giving the ancestral state. But, because the increased propensity for transitional[8] over transversional[9] mutations as well as the hypermutability of CpG sites and other context-dependent DNA replication

---

[7]An outgroup is a distantly related group of organisms that serves as the ancestral reference for the studied group (or ingroup).

[8]A transition is a mutation between two nucleotidic bases of the same family (purine or pyrimidine), i.e. either a A $\leftrightarrow$ G or a C $\leftrightarrow$ T mutation.

[9]A transversion is a mutation involving a change of nucleotidic family (from a purine to a pyrimidine or the other way round), i.e. either a A $\leftrightarrow$ C, a A $\leftrightarrow$ T, a G $\leftrightarrow$ C or a G $\leftrightarrow$ T mutation.
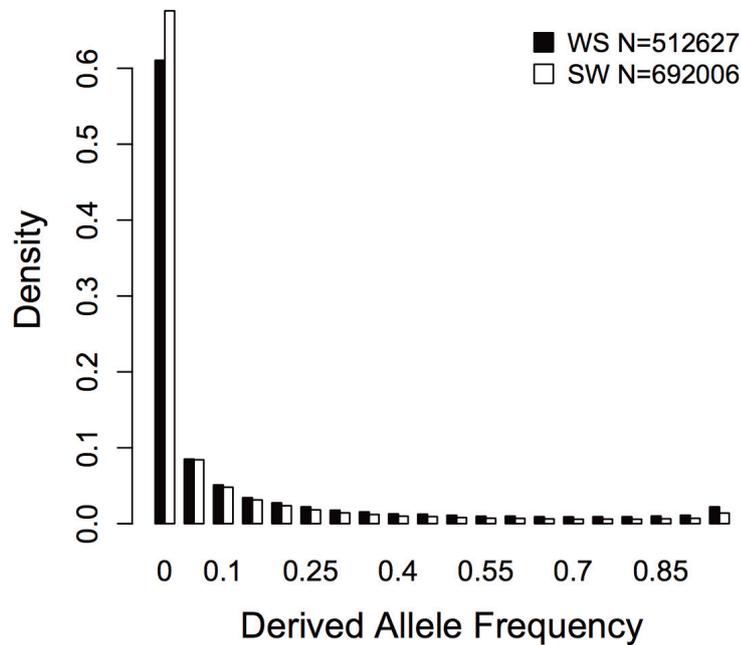
Figure 4.4: Example of a derived allele frequency spectrum (DAFS) separated for AT → GC (WS) and GC → AT (SW) mutations.
AT → GC (WS) mutations are coloured in black and GC → AT (SW) in white. The spectrum for WS mutations is shifted towards higher frequencies, as compared to the spectrum for SW mutations, as predicted in the gBGC model.
This figure was reproduced from Glémin et al. (2015) and corresponds to a subsample of the original figure (permission in Appendix B).

errors (Hwang and Green, 2004) are known to induce spurious signatures of gBGC (Hernandez et al., 2007), Glémin et al. (2015) developed a method correcting for such polarisation errors and thus allowing to better detect and quantify gBGC. Indeed, if gBGC participates in the evolution of the genome studied, the SFS will present WS mutations shifted towards higher frequencies than SW mutations (e.g. in Figure 4.4), and the intensity of the shift will reflect that of gBGC.

As an alternative to SFS, comparative genomics approaches exist to quantify gBGC. For instance, Lartillot (2013b) created a method based on the analysis of substitution patterns to quantify gBGC in a whole phylogeny and Capra et al. (2013) developed another one allowing to quantify gBGC along a given genome (reviewed in Mugal et al., 2015).

As such, gBGC has been quantified in several organisms *via* theoretical approaches. But empirical studies too have helped in better characterising this driver of genome evolution, as reviewed hereunder.

4.3.2   Empirical studies of gBGC

All in all, the gBGC model is in accordance with observations in countless metazoans (Capra and Pollard, 2011; Galtier et al., 2018) including vertebrates (Figuet et al., 2014), — among which mammals (Romiguier et al., 2010; Katzman et al., 2011; Lartillot, 2013b; Clément and Arndt, 2013; Glémin et al., 2015; Dutta et al., 2018), avians (Webster et al., 2006; Weber et al., 2014; Bolívar et al., 2016) and reptiles (Figuet et al., 2014), — and also some invertebrates like bees (Kent et al., 2012; Wallberg et al., 2015) and *Daphnia* (Keith et al., 2016). Though, not all invertebrates are subject to gBGC: *Drosophila*, except for its X chromosome (Galtier et al., 2006; Haddrill and Charlesworth, 2008), is not affected (Robinson et al., 2014). Plants — both angiosperms (Escobar et al., 2011; Glémin et al., 2014; Rodgers-Melnick et al., 2016; Clément et al., 2017; Niu et al., 2017, but see Liu et al., 2018) and gymnosperms (Serres-Giardi et al., 2012) — also show molecular characteristics compatible with gBGC. Thus, these eukaryotes, as well as numerous others (Escobar et al., 2011; Pessia et al., 2012) — but also certain prokaryotes (Lassalle et al., 2015; Long et al., 2018), — likely undergo gBGC.

Nevertheless, in all the aforementioned cases, gBGC was only observed indirectly — for instance *via* correlations between GC-content and recombination, or *via* the analysis of patterns of substitutions between closely related species. A decade ago though, gBCG has been confirmed experimentally in yeasts thanks to the creation of the first high-resolution recombination map (Mancera et al., 2008): this map allowed to precisely analyse conversion tracts (CTs) at the genome-wide scale and to demonstrate that S alleles are significantly overtransmitted, even if the effect is extremely weak (GC-bias: 50.065%). Further analyses of this dataset have revealed

that, in yeasts, gBGC is only associated with COs (but not NCOs), and solely affects the markers at the extremities of CTs (Lesecque et al., 2013).

In contrast, the first experimental evidence for gBGC in humans was restricted to a few hotspots (Odenthal-Hesse et al., 2014; Arbeithuber et al., 2015) and was found exclusively in NCOs. Nonetheless, gBGC remains a pervasive driver of human genome evolution since it has been estimated to affect about 15% of our genome (Pouyet et al., 2018).

The mechanism at the origin of gBGC may not be the same for these two species. Indeed, in yeasts, gBGC is primarily associated to simple CTs (Lesecque et al., 2013), which rules out the hypothesis of gBGC originating from the base excision repair (BER) machinery (according to which gBGC should be associated mainly with complex CTs) and instead suggests that it would originate from the mismatch repair (MMR) machinery. As for humans, Halldorsson et al. (2016) found that gBGC was stronger at CpG than at non-CpG sites, which argues in favour of the BER hypothesis.

Interestingly, the BER and the non-canonical MMR (i.e. MMR activated by DNA lesions) pathways have been shown to cooperate in the removal of mismatches in the context of DNA demethylation (Grin and Ishchenko, 2016), and a similar interplay between the two machineries in the context of meiotic repair of programmed DSBs could alternatively be conceived.

More recently, direct observations of gBGC at a larger scale in humans have been reported by two independent studies (Williams et al., 2015; Halldorsson et al., 2016). They confirmed that gBGC affects NCOs (GC-bias: 68%), but also COs displaying complex CTs (GC-bias: 70%). However, the framework used did not allow to test for gBGC in COs with simple CTs. This phenomenon was also directly observed in NCO CTs of birds (Smeds et al., 2016) and rice (Si et al., 2015) (GC-bias: 59% in both cases), but could not be tested either in CO CTs.

4.3.3   Relationship with parameters of genome evolution

Provided that no evolutionary force acts upon its transmission, the allelic frequency of a heterozygous locus in a pool of gametes should equal the Mendelian frequency of 50%. In presence of gBGC however, the allelic frequency of the favoured allele in the gametic pool (x) increases proportionately to the gBGC coefficient (b) according to the following relationship:

$$x = \frac{1}{2} \times (1 + b)$$

Since x is a proportion and is thus necessarily bounded between 0 and 1, b is bounded between -1 (when AT alleles are systematically transmitted) and 1 (when GC alleles are systematically transmitted).

The intensity of the gap between the observed transmission and the Mendelian frequency (and thus, the gBGC coefficient b) depends on the recombination rate r (including both COs and NCOs), the length of gene conversion tracts L and the transmission bias (a.k.a. mismatch repair bias) b_0, as such:

$$b = r \times L \times b_0$$

Finally, the spread of the favoured allele in the population is represented by the population-scaled gBGC coefficient (B), which itself depends on both b and the effective population size (N_e) in a fashion much similar to the probability of fixation under selection defined by Kimura (1962):

$$B = 4 \times N_e \times b$$

In human genomes, apart from recombination hotspots which display a mean B value of 3 (Glémin et al., 2015), the average B found in several independent studies circumscribes between 0.1 and 0.5 (Lartillot, 2013b; De Maio et al., 2013; Glémin et al., 2015), which is a low estimate as compared to other mammalian genomes (Lartillot, 2013b) (Figure 4.5).
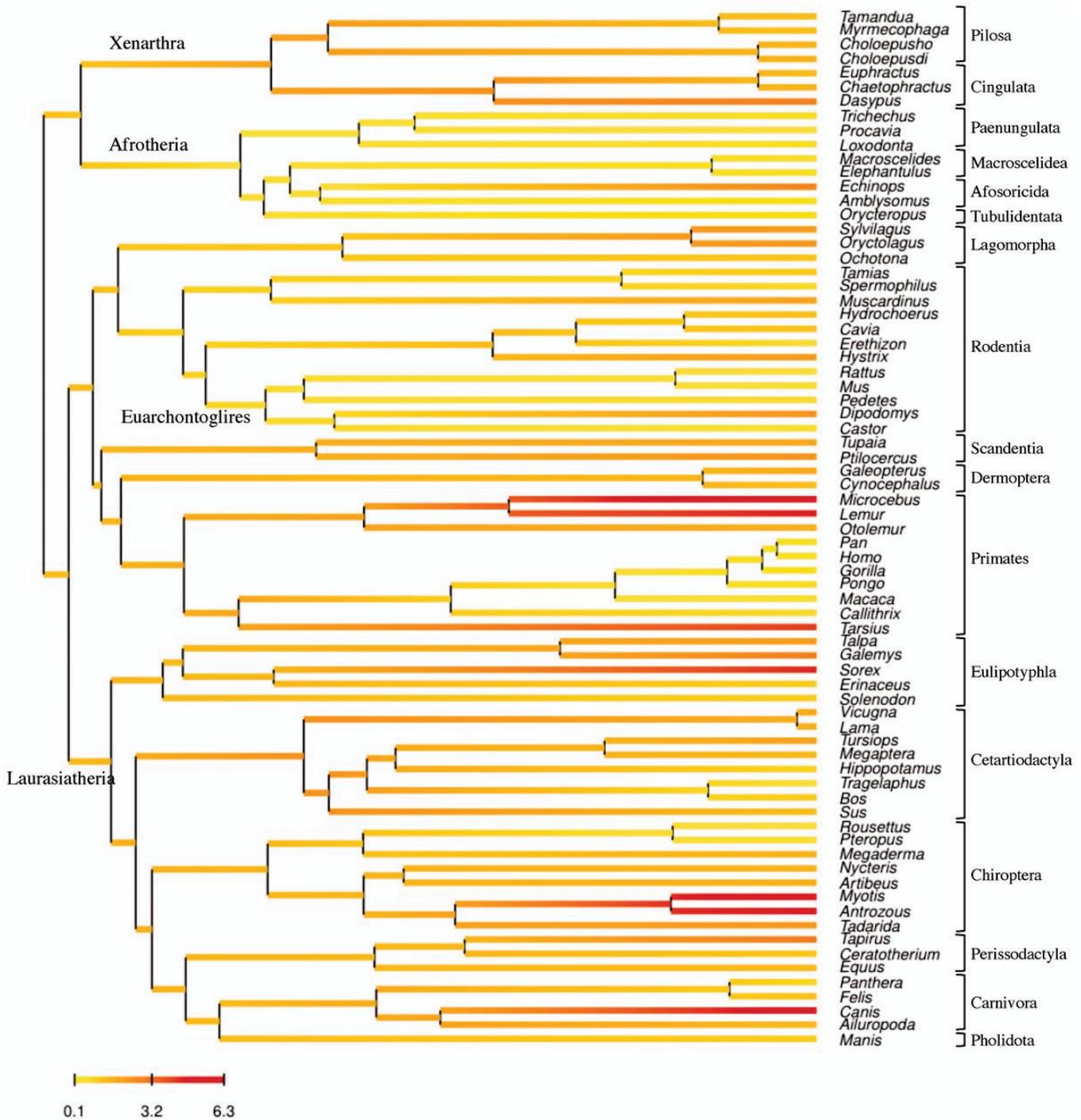
Figure 4.5: Reconstructed phylogenetic history of $B = 4 \times N_e \times b$ in placental mammals.

The names of orders are given on the right side of the tree and each branch is coloured according to its average genome-wide B.

This figure was reproduced from Lartillot (2013b) (permission in Appendix B).

It has also been found that B (approximated by the average GC_3 content) correlates with certain life history traits. Indeed, it correlates negatively with genome size in mammals (Romiguier et al., 2010), likely because the per-megabase recombination rate is greater in short chromosomes (Kaback et al., 1992; Lander et al., 2001; International Chicken Genome Sequencing Consortium, 2004).

B also correlates negatively with body mass, longevity and age of sexual maturity in mammals (Romiguier et al., 2010; Lartillot, 2013a) and birds (Weber et al., 2014; Figuet, 2015; Figuet et al., 2016), which was interpreted in terms of effective population size (N_e), since body mass negatively correlates with N_e in both mammals (Damuth, 1981; White et al., 2007) and birds (Nee et al., 1991).

Nevertheless, this relationship between life history traits and GC-content dynamics is not (or not entirely) mediated by N_e since no direct correlation between N_e and B has been observed among animals (Galtier et al., 2018). This unexpected observation has been interpreted by two non-mutually exclusive possibilities. One interpretation would be that, since gBGC is a deleterious process (Galtier et al., 2009; Necşulea et al., 2011; Lachance and Tishkoff, 2014), there may be a selective pressure to minimise b in species with large N_e.

Alternatively, there may be a 'dilution effect' if, as in yeasts (Lesecque et al., 2013), only the SNPs located at the extremities of conversion tracts (CTs) are converted: in that case where only one part of the CT markers are subject to gBGC, the mean b would decrease with N_e since polymorphism correlates positively with N_e (Tajima, 1996; Woolfit, 2009).

Part II

Objectives of the thesis

One striking result that came with the quantification of the population-scaled GC-biased gene conversion (gBGC) coefficient (B) across metazoans (Galtier et al., 2018) is that its intensity restricts to a very limited scope. For instance, in placental mammals, B settles in a [0; 7] range (Lartillot, 2013b). Given that B is nothing but the product of the effective population size (N_e) by the gBGC coefficient (b) (see Chapter 4) and that N_e fluctuates over orders of magnitude across metazoans, any theory according to which the intensity of gBGC (b) would be evolutionarily stable has to be ruled out (Galtier et al., 2018). Instead, one or several of the parameters on which b depends (the recombination rate r, the length of conversion tracts L and the transmission bias b_0) necessarily vary inversely with N_e.

However, data still lack to understand the basis of the dependency between N_e and b: the transmission bias (b_0) has only been measured in a handful of species (Mancera et al., 2008; Si et al., 2015; Williams et al., 2015; Halldorsson et al., 2016; Keith et al., 2016; Smeds et al., 2016) and, among mammals, the only species for which b_0 has been quantified is one with a very low N_e of 10,000 (Takahata, 1993; Erlich et al., 1996; Harding et al., 1997; Charlesworth, 2009; Yu et al., 2004): *Homo sapiens.*

In order to shed new light on the interplay between b and N_e, we thus aimed at quantifying gBGC in another mammalian species displaying an effective population size much larger than that of humans (Geraldes et al., 2008; Phifer-Rixey et al., 2012; Davies, 2015): *Mus musculus.*

Such endeavour calls for a large number of recombination events on which gBGC could be measured. Though, the method prominently used to detect recombination — pedigree analysis — is extremely resource-intensive: it requires a considerable number of individuals sequenced genome-wide and results in the detection of a limited amount of recombination events (see Chapter 3). Thus, we implemented a novel approach allowing to detect thousands of such events at high resolution in single individuals. I describe it in Chapter 5.

Then, in Chapter 6, I describe how these tens of thousands of events allowed us to precisely characterise recombination in over 1,000 autosomal recombination hotspots and how we could infer the genuine parameters of mouse meiotic recombination (in particular the recombination rate r and the length of conversion tracts L) through inferential methods.

Next, after distinguishing the effects of GC-biased gene conversion (gBGC) from those of DSB-induced biased gene conversion (dBGC) on the observable transmission of alleles, we managed to quantify the transmission bias (b_0) of GC alleles in the conversion tracts of our detected recombination events as well as the intensity of dBGC in several hundreds of recombination hotspots. I describe these findings in Chapter 7.

Last, because the approach presented in Chapter 5 showed unprecedented power to detect recombination events in a single individual, the logical follow-up was to re-use it in other studies involving the inactivation of genes essential to recombination. In Chapter 8, I describe the methodological adaptations of our procedure to such investigations and the preliminary results of our analysis.

The results described in the four aforementioned chapters will then be discussed in Chapter 9.

Part III

Results

'I don't claim to be a methodologist, but I act like one only because I do methodology to protect myself from crazy methodologists.'

— Ward Cunningham, *Geek Noise* (2004)

5

High-resolution detection of recombination in single individuals

Contents

This chapter in brief — *Because the existing approaches to study recombination at high resolution are extremely resource-intensive, we implemented a novel approach based on the unique-molecule genotyping of recombination-enriched sperm DNA from single highly heterozygous individuals. We found that the main source of errors when genotyping unique recombinant molecules of DNA did not come from sequencing errors, but from alignment ambiguities — for the aligners are biased towards the reference genome. Thus, searching for events after mapping fragments onto both parental genomes proved to be the most critical step of our pipeline. In the end, our approach proved 100 times more powerful than current methods to detect recombination: it allowed to identify several thousands of recombination events in single individuals, with a false positive rate below 5%.*

The existing approaches to study recombination events at high resolution are limited (see Chapter 3). For instance, the total number of events detectable with approaches like pedigree analyses is capped by the restricted number of meioses that can be analysed (one meiosis per individual sequenced). In addition, since whole genomes are sequenced to retrieve these events, the cost/benefit ratio is particularly elevated for species with large genomes, like mammals.

Here, we propose a different procedure which rests on the unique-molecule genotyping of recombination-enriched sperm DNA from single highly heterozygous individuals (Figure 5.1). In this chapter, I first describe how our experimental design led to an enrichment in detectable recombination events and how we implemented our unique-genotyping pipeline to identify such events and then discuss the impact of each component of our workflow onto the detectability of events.

5.1 Overview of the experimental design

5.1.1 Acquisition of highly polymorphic individuals

Detecting recombination events rests on one indispensable prerequisite: the presence of markers (i.e. polymorphic sites).

Therefore, we performed a cross between two subspecies of mice that present a high level of heterozygosity (1 SNP every 150 bp) (Keane et al., 2011; Yalcin et al., 2012) and that are known to hybridise naturally (Orth et al., 1998): *Mus musculus domesticus* (strain C57BL/6J, hereafter called B6) and *Mus musculus castaneus* (strain CAST/EiJ, hereafter called CAST). This cross resulted in F1 hybrid mice (B6xCAST), of which two males were selected. Sperm DNA was then collected from these two individuals and kindly given to us by D. Bourch'is (Institut Curie, Paris).

The extracted DNA from both biological replicates was then sonicated to produce fragments of a mean size of 350 bp.
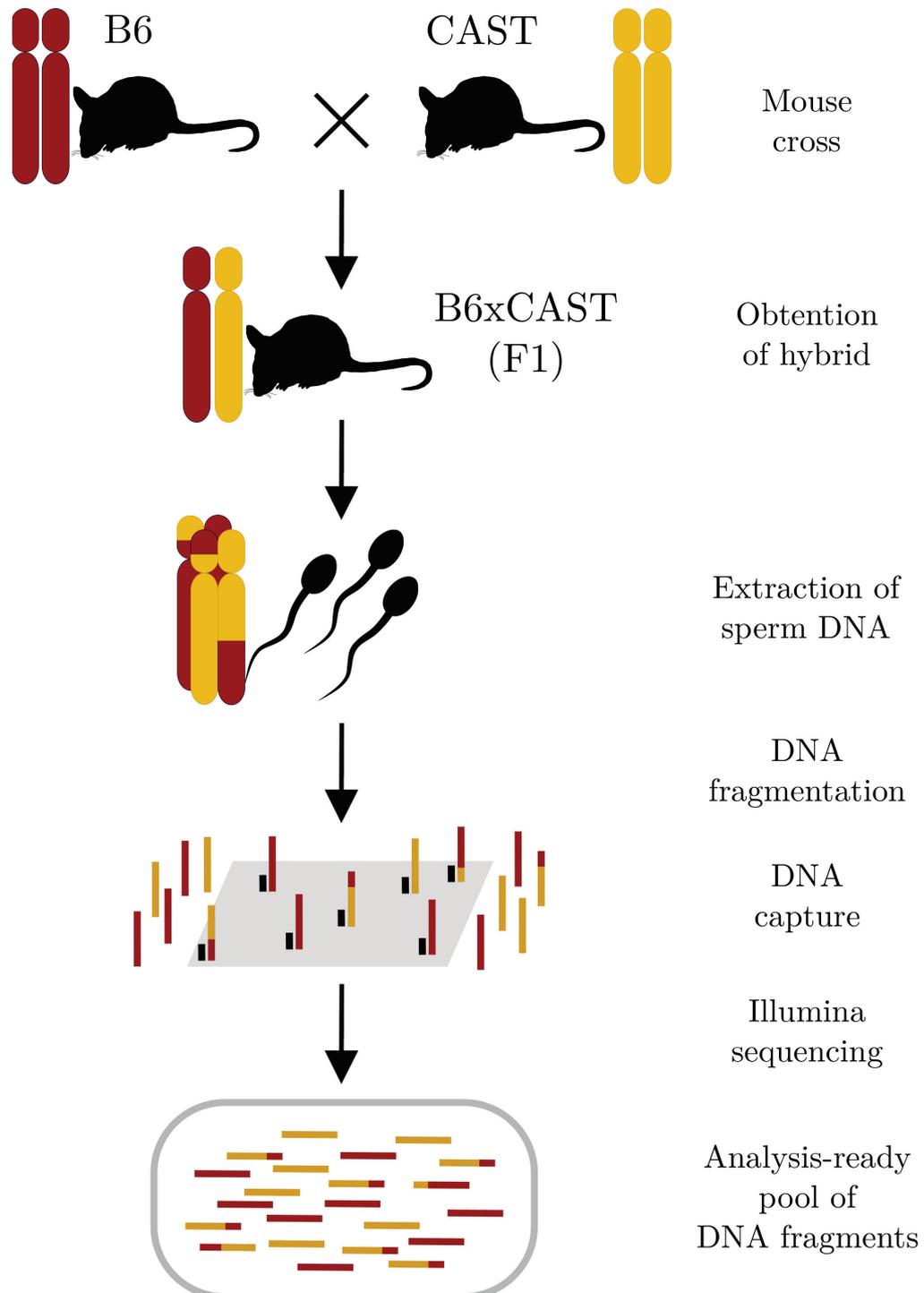
Figure 5.1: Overview of the experimental design.
We performed a cross between a *Mus musculus domesticus* (B6) and a *Mus musculus castaneus* (CAST) mouse individual to obtain a F1 hybrid, from which we extracted sperm DNA, i.e. the substrate of recombination products. We then performed two rounds of DNA capture to target the 1,018 hotspots and 500 control regions selected, and sequenced captured DNA with an Illumina device, using a 250-pb paired-end protocol. At the end of this process, the pool of DNA was enriched in recombination events. B6 chromosomes and fragments of DNA are coloured in red and CAST chromosomes and fragments of DNA in yellow.

5.1.2   Enrichment in detectable recombination events

Since the recombination rate is relatively weak genome-wide, we wanted to target specifically recombination hotspots, i.e. regions of the genome where recombination massively occurs. This required two steps: selecting hotspots, and performing DNA capture (i.e. hybridisation-based targeted-DNA enrichment) of these loci (Gnirke et al., 2009; Hodges et al., 2007, reviewed in Horn, 2012).

Selection of targets

The recombination hotspots of B6xCAST mice had previously been identified *via* chromatin immunoprecipitation followed by sequencing (ChIP-seq) of the PRDM9 protein (Baker et al., 2015a). We restricted this known list of 6,758 hotspots to those (1) displaying a high marker density in the vicinity of the PRDM9 binding site (so as to increase the chance of detecting recombination events) and (2) aligning on their whole length on both the CAST and the B6 reference genome (so as to restrain mapping artifacts).

In practice, the selection criterium on heterozygosity (a minimum of 4 SNPs in the 300-bp central region of the locus centred on the PRDM9 peak summit) was the most stringent: it cut down the original list of 6,758 hotspots to only 1,261 hotspots. The other two criteria on mappability (a strict maximum of 60 sites with low sequence quality in the 1-kb central region, and the absence of a large indel by ensuring that a minimum of 800 bp in the 1-kb from the B6 genome shared at least 90% identity with the CAST genome) respectively discarded 205 and 38 additional loci. Altogether thus, a total of 1,018 1-kb long regions centred on the summit of the PRDM9 ChIP-seq peaks were selected. These were positioned randomly both across and along chromosomes (Figure 5.2).

In addition, we selected 500 1-kb control regions which displayed genomic characteristics similar to those of the 1,018 hotspots (in terms of GC-content, SNP density, sequence quality and content in transposable elements) but which did not belong to the list of known recombination hotspots.
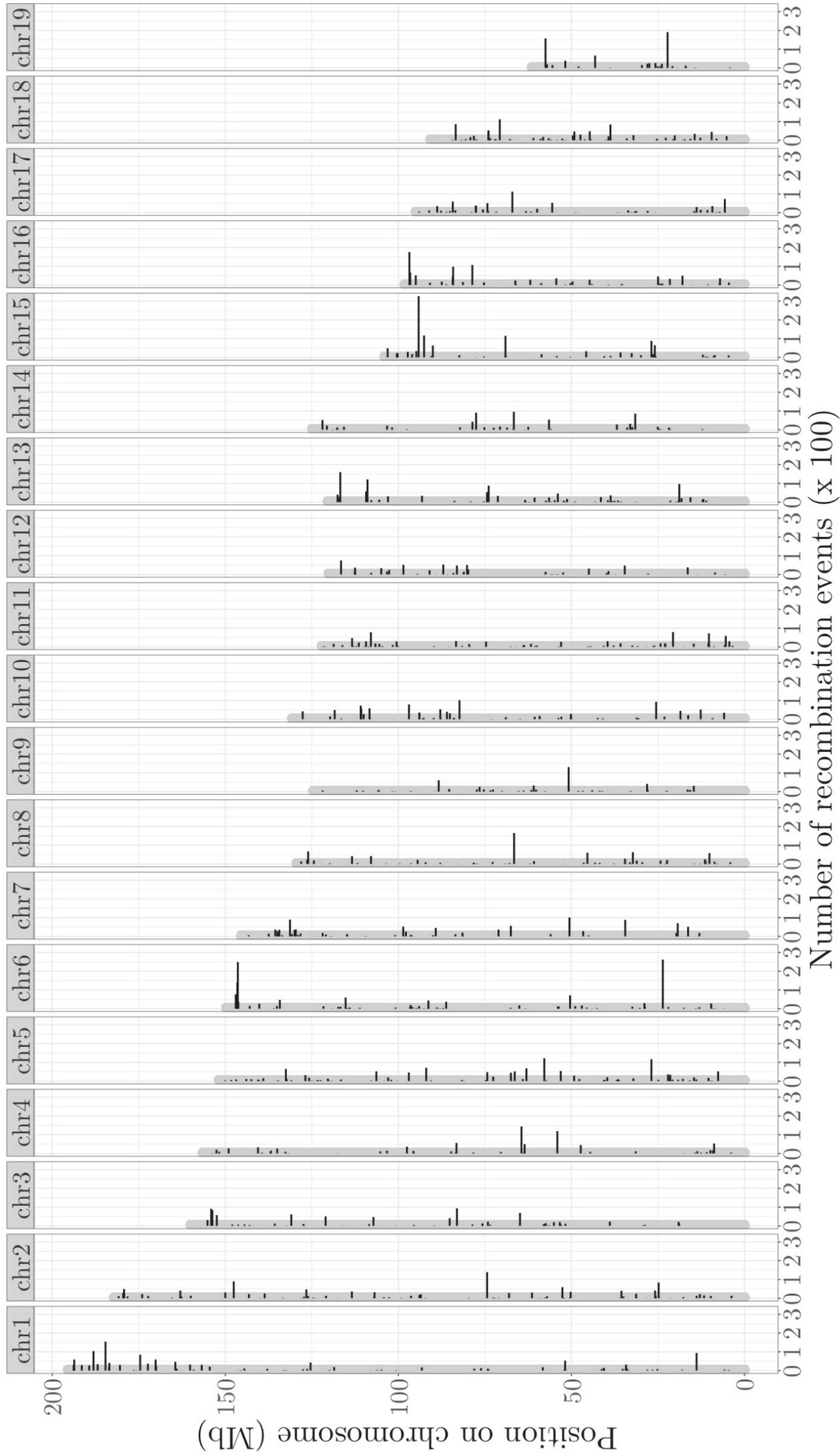
Figure 5.2: Distribution of recombination events across the 1,018 selected hotspots positioned randomly along chromosomes.
Chromosomes are represented in grey and oriented so that the centromere is on the bottom side of the figure (mouse chromosomes are acrocentric). The total number of recombination events identified is given by the length of the horizontal black bar at the position of each of the 1,018 selected hotspots.

DNA capture

To enrich the sequencing data in fragments coming from the 1,518 aforementioned loci (hereafter called targets), we performed either one or two rounds of DNA capture targeting them. Since our final aim was to detect recombination events, i.e. fragments carrying both a portion of the B6 haplotype and a portion of the CAST haplotype, it was essential that the efficiency of the capture be similar for both haplotypes. We thus designed two baits (one for each of the two haplotypes) for every target.

We next monitored the existence of any capture bias by looking at the origin of all the non-recombinant fragments. Indeed, as recombination is rare, the vast majority of sequenced fragments do not correspond to recombination events and
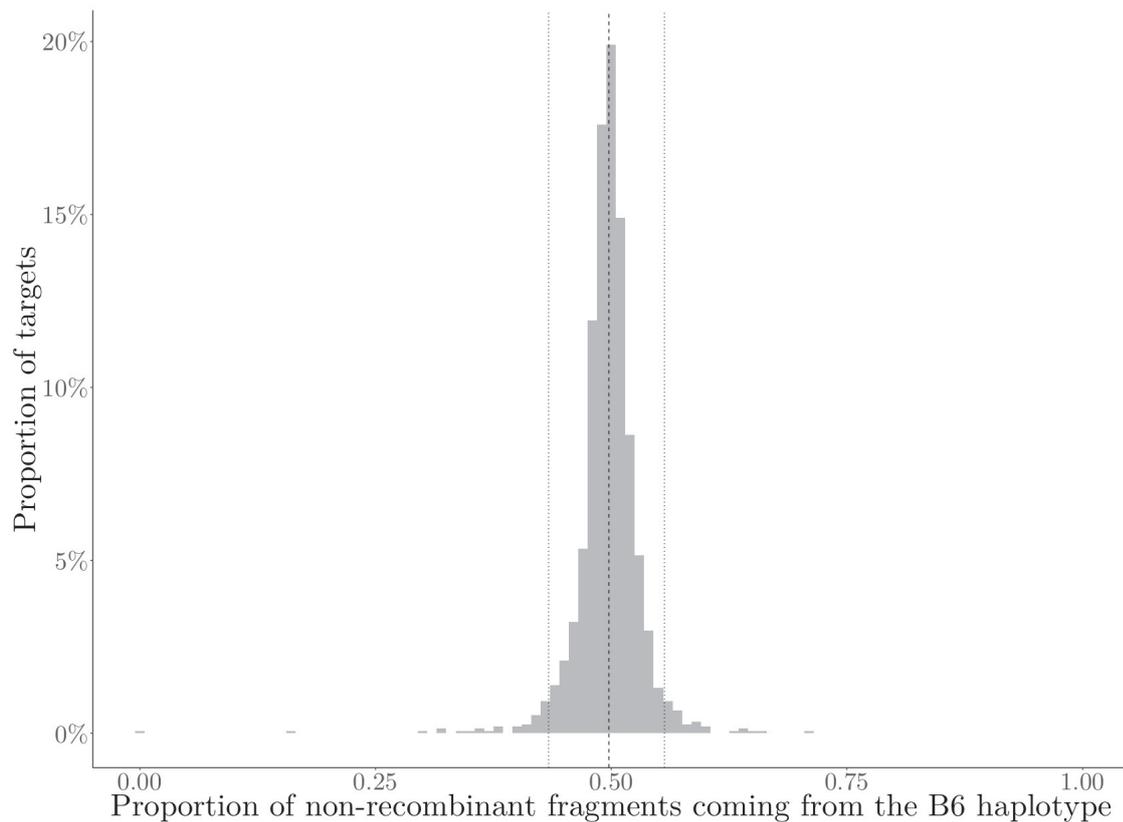
Figure 5.3: Absence of capture bias between the B6 and CAST haplotypes.
All fragments exclusively containing B6-typed markers were designed as non-recombinant fragments coming from the B6 haplotype. The distribution of the proportion of such fragments across targets is reported in this figure. The dashed line corresponds to the median proportion of B6-genotyped fragments across targets and the two dotted lines correspond to the 2.5 and 97.5 percentiles (i.e. the delimitation of the proportion for 95% of targets).

thus, half of all non-recombinant fragments should come from the B6 haplotype (and consequently, the other half from the CAST haplotype). We found that this was indeed the case since the proportion of fragments containing only *B6*-typed markers (i.e. coming from the B6 haplotype) revolved around 50% for nearly all targets: 95% of targets held in a [43.4%; 55.6%] range (Figure 5.3).

Thus, although a small fraction of hotspots displayed a haplotype bias (possibly because one of the two baits better matched one of the two haplotypes), overall, there was no systematic bias favouring the capture of one haplotype relative to the other.

5.1.3   Ultra deep-sequencing and mapping of captured DNA

Libraries were sequenced by an Illumina device using a 250-bp paired-end protocol, except for 4 small libraries (out of 18) which contributed to 6% of the total number of fragments and which were sequenced as a pilot experiment using a 100-bp paired-end protocol (Table 5.1).

We then mapped the sequenced reads to both the GRCm38/mm10 version of the B6 genome (`ftp://ftp-mouse.sanger.ac.uk/ref/`) and to the CAST/EiJ draft reference genome (`ftp://ftp-mouse.sanger.ac.uk/REL-1509-Assembly/`), using BWA-MEM (Li and Durbin, 2009; Li, 2013) with default parameters and marking shorter split hits as secondary. PCR duplicates were marked thanks to picardTools (version 1.98(1547)) (Broad Institute, 2018) and pairs of reads which were either marked as unmapped, as secondary alignment[1] or as mapping in an improper pair[2] were filtered out, for they were not likely to be real fragments.

Overall, sequenced reads mapped equally well to both the B6 and the CAST reference genome assemblies (Table 5.1). In addition, DNA capture was efficient

---

[1]BWA marks a read as secondary-aligned in cases where it can align at several locations. The best hit (i.e. location with the best alignment score) is marked as primary alignment, while all others are marked as secondary alignment.

[2]A proper pair flag is attributed by the aligner (here, BWA) to a pair of reads if the reads are oriented in an inward-facing direction and are mapped within 4 standard-deviations of the mean insert size of the block of 10^6 read pairs to which they belong.

since 72% of sequenced fragments mapped within the selected targets. This resulted in a substantial coverage of the targeted loci: the mean sequencing depth on 1-kb long targets was 97,177 x, and it raised up to above 941,737 x in certain regions.

The variation in coverage across hotspots was similar to that across control regions (data not shown) and it was rather limited as the variation in coverage of 80% of hotspots held in a 5:1 ratio ([10; 90] quantiles = [71,745; 385,412] reads). Nonetheless, we found that the variation in coverage across hotspots was highly correlated to the mean GC-content of the targets (Pearson correlation: $r = -0.641$; p-val $< 2.2 \times 10^{-16}$). Thus, apart from a GC-content effect, there was no large capture nor mapping bias across hotspots.

5.2   The unique-molecule genotyping pipeline

Since recombination ends in the juxtaposition of DNA from the two parental haplotypes, discerning recombination events comes back to spotting fragments presenting both *B6*-typed and *CAST*-typed genetic markers. This requires two steps: disclosing the position of polymorphic sites and identifying the allele carried by a given fragment at all the markers it overlaps.

5.2.1   Identification of polymorphic sites

We performed variant-calling (i.e. the process of identifying variant (a.k.a. polymorphic) sites on a genome) with GATK[3] (version 3.3) (McKenna et al., 2010).

Basically, GATK performs four main steps: local insertion/deletion (indel) realignment, base quality score recalibration (BQSR), variant-calling *per se* and variant quality score recalibration (VQSR). Briefly, local indel realignment consists

---

[3]Other routine manipulations of files and visualisation of alignments were performed using the following tools and versions: SAMTools (version 1.4) (Li et al., 2009), BEDTools (version 2.26.0) (Quinlan and Hall, 2010), JVarKit (Lindenbaum, 2015), the IGV interface (version 2.3_88) (Robinson et al., 2011).

	Sample and sequencing characteristics						Mapping (%)		Capture efficiency		
Library ID	Mouse ID	# DNA capture	Rep. ID	Lane ID	Read length	Library size	Ref. B6	Ref. CAST	# Filtered Fragments	% in targets	# in targets
A-1	1	2	-	-	100 bp	14,977,880	99.87	99.76	7,206,235	86.09	6,203,730
A-2	1	1	-	-	100 bp	12,457,816	99.75	99.27	5,813,649	26.86	1,561,461
A-3	2	2	-	-	100 bp	16,000,000	99.85	99.73	7,631,724	84.88	6,478,370
A-4	2	1	-	-	100 bp	13,291,526	99.74	99.24	6,110,086	24.29	1,484,199
B-1	1	2	-	-	250 bp	51,923,148	99.76	99.74	24,887,319	85.95	21,391,551
B-2	1	1	-	-	250 bp	64,260,092	99.72	99.66	29,732,709	26.67	7,927,136
B-3	2	2	-	-	250 bp	98,238,822	99.64	99.61	46,831,049	84.87	39,749,391
B-4	2	1	-	-	250 bp	130,482,992	99.60	99.52	59,942,764	24.52	14,700,518
C-1	1	2	1	1	250 bp	67,775,154	99.79	99.75	33,221,010	82.54	27,421,183
C-2	1	2	1	2	250 bp	69,564,748	99.78	99.75	34,102,906	82.65	28,185,446
C-3	1	2	2	1	250 bp	79,002,218	99.79	99.75	38,818,821	84.05	33,013,536
C-4	1	2	2	2	250 bp	81,074,012	99.77	99.74	39,841,630	85.14	33,921,167
C-5	2	2	1	1	250 bp	60,911,138	99.71	99.68	29,876,369	85.45	25,530,140
C-6	2	2	1	2	250 bp	62,042,362	99.71	99.67	30,437,699	85.55	26,039,414
C-7	2	2	2	1	250 bp	66,489,166	99.77	99.74	32,701,170	86.18	28,182,084
C-8	2	2	2	2	250 bp	68,382,888	99.76	99.73	33,636,524	86.28	29,022,692
C-9	-	2	1&2	1	250 bp	9,796,678	83.27	84.44	3,801,189	62.12	2,361,438
C-10	-	2	1&2	2	250 bp	10,156,096	82.77	84.93	3,920,630	62.04	2,432,278
Total	-	-	-	-	-	976,826,736	99.68	99.37	468,513,483	71.63	335,605,734

Table 5.1: Sequencing, mapping and capture-efficiency summary metrics.
The sperm from the two individuals was sequenced three times (horizontal panels): 'A-libraries' correspond to a pilot where reads were sequenced with a 2 × 100-bp protocol at a low coverage; 'B-libraries' were enriched in recombination events *via* either one or two rounds of DNA-capture and were then sequenced deeply with a 2 × 250-bp protocol; 'C-libraries' were sequenced *a posteriori* (after having analysed results of the 'B-libraries'), using only samples with two rounds of DNA-capture to increase the total number of recombination events. The first vertical panel recaps the biological and sequencing characteristics of the samples. The library size is the number of sequenced reads. The second vertical panel recaps the percentage of reads mapping on either of the two reference genomes (B6 and CAST). The third vertical panel recaps statistics on the efficiency of DNA-capture. Capture efficiency corresponds to the percentage of filtered fragments (i.e. fragments remaining after the removal of unmapped and secondary-aligned reads) that map within the targeted sites.

in transforming regions with indel-based misalignments into clean reads containing a
consensus indel; BQSR consists in applying a score correction accounting for sources
of systematic technical errors by modelling sequencing miscalls empirically; variant-
calling allows to call both SNPs and indels; and VQSR provides an estimate of the
probability that a called variant is a true genetic variant thanks to the establishment
of an empirical model linking the latter likelihood to metrics describing the variants.

For all these steps, the GATK team recommends a number of best practices
(DePristo et al., 2011; Van der Auwera et al., 2013) which, in many instances, require
several external datasets of 'true' (i.e. validated by several independent studies) and
'false' variants which are available for human genomes only. Therefore, we adapted
the variant-calling process to mouse genomes as described hereunder, by using other
types of datasets as close as possible to the recommendations from the GATK team.

First, to perform local indel realignment, our list of known indels was made of
all the indels found between the B6 strain and any of the other 35 strains of the
version 5 release (`ftp://ftp-mouse.sanger.ac.uk/REL-1505-SNPs_Indels/`) of
the mouse genomes project (MGP) (Keane et al., 2011).

Second, to perform base quality score recalibration (BQSR), our list of known
indels was the same as that used for local indel realignment, and that of known
SNPs was made of all the SNPs found between the B6 strain and any of the
other 35 strains of the MGP.

Third, to perform variant quality score recalibration (VQSR), our list of true
variant sites was made of all the sites (both SNPs and indels) found to vary between
the B6 and the CAST strains by the MGP, and our list of both true variants
and false positives of all the sites (both SNPs and indels) found to vary between
the B6 strain and any of the other 16 strains of the MGP. The annotations we
specified as covariates for the model were: the quality by read depth (QD), the
overall mapping quality of reads supporting the variants called (MQ), the rank
sum test for mapping qualities (MQRankSum), the rank sum test for the distance
from the end of the reads (ReadPosRankSum), and two measures of strand bias

(FS and SOR). Last, since VQSR classifies variants according to their confidence, we discarded all those which were marked as 'LowQuality'.

5.2.2   Genotyping of individual DNA fragments

Genotyping basically consists in comparing the allele carried by the analysed fragment at a given polymorphic site with those of the parental genomes. Though, the accuracy of genotyping is subject to two main types of errors.

First, the variant-calling step can output a small proportion of false positives (FPs), for instance because of mapping artifacts in the vicinity of indels. In that case, even if a fragment is correctly sequenced, a genotyping error may arise from these FP markers. Since recombination is rare ($\ll 1\%$), the allelic frequencies at genuine polymorphic sites should comply with the Mendelian transmission of alleles. To avoid any error coming from the aforementioned FPs, we thus applied a hard filter on these frequencies: only sites with allelic frequencies within the [36%; 64%][4] were retained. We additionally applied a hard filter on read coverage: any called variant supported by fewer than 100 reads was automatically discarded.

Second, sequencing errors directly lead to genotyping errors. To avoid that, one can use the information provided by the sequencer: the Phred quality score which is logarithmically related to the probability for the base call to be incorrect (Ewing et al., 1998; Ewing and Green, 1998). However, the Phred scores produced by the sequencing machines are subject to various sources of systematic technical error and sequencing machines generally underestimate the probability of error (GATK team, 2012). Thus, we used the base quality scores recalibrated by GATK to filter out base calls with high probabilities of error: all sites with a recalibrated quality below 20 (i.e. with a probability to be miscalled greater than 1%) were discarded.

We then genotyped each fragment at every of the remaining high-confidence variant sites that it overlapped by comparing its base call (or sequence of base calls

---

[4]The values of that range were deliberately set as relatively large to account for any difference in capture efficiency at individual hotspots (Figure 5.3)

in the case of indels) with that of the reference genome. Whenever the fragment carried an allele distinct from that of the reference genome, we checked that the allele carried was that of the other parental genome to avoid misgenotyping any remaining erroneous base call.

5.2.3   Identification of recombination events

A simple way to test for the accuracy of our genotyping was to monitor the polymophic sites overlapped by the two reads of a given fragment: in principle, such markers should have the same genotype call on both reads. In our data, only 0.3%[5] (97 out of 32114) of all such markers were genotyped discordingly. Even if this seems to be a low error rate, it is not negligible in view of the scarcity of recombination events. Therefore, to avoid false positives (FPs) due to genotyping errors, we identified fragments as recombination events when they bore a minimum of 2 *CAST*- and 2 *B6*-typed markers.

Last, since targets were sequenced deeply, a non-negligible portion of the fragments sequenced were likely to have arisen from PCR duplicates. Therefore, we discarded all events which showed an homologue both starting and ending at the same genomic position, so as to be sure to retain only one copy of any given recombination event in our dataset.

Finally, aside from sequencing errors, alignment ambiguities can lead to false positive calls (see Section 5.3). These depend on the aligner and its parameters, — among which the reference genome. Thus, we performed the whole procedure (mapping, variant-calling, marker selection, recombination event identification) twice: once using the B6 parental genome as reference, and once using the CAST parental genome as reference.

---

[5]Markers overlapped by both reads correspond to markers that are located at the end of reads. Since read extremities are more prone to both misalignments and sequencing errors (Kircher et al., 2009; Minoche et al., 2011; Abnizova et al., 2012; Wang et al., 2012b; Laehnemann et al., 2016), the genotyping error rate provided here is likely to be overestimated.

5.3 The determinants of sensitivity and specificity

5.3.1 An unprecedentedly powerful approach

Since none of the 500 control loci correspond to known recombination hotspots, they should host few — or no — recombination events. Therefore, the number of recombination events detected in these control regions provides an upper limit for the number of false positives (FP) and, as hotspots and controls share similar genomic characteristics, the FP rate is expected to be comparable in both backgrounds.

All in all, 18,821 recombination events were retrieved in the 1,018 selected hotspots, and we estimated the maximum FP rate to be 3.73% (Table 5.2).

Target category	Nb of targets	Nb of fragments	Nb of events	Event rate ($\times 10^{-6}$)
Hotspots	1,018	228,984,512	18,821	82.2
Controls	500	106,850,906	328	3.07
FP rate				**3.73 %**

Table 5.2: Number of events detected in hotspot and control targets.
Events (false positives (FPs) or genuine recombination events) were detected using the unique-molecule genotyping pipeline described in Section 5.2. All fragments or events overlapping at least 1 bp with a given target are counted in this table. The event rate corresponds to the ratio of candidate recombination events over the total number of fragments. The maximum false positive (FP) rate is the ratio of the event rate in control targets over that in hotspots.

Altogether, our approach displayed a much better efficiency/cost ratio than comparable methods to characterise recombination events at high resolution.

Indeed, in a recent study carried on mice by Li et al. (2018), the sequencing of 119 genomes of mice at a 12–30-x coverage (which corresponds to about 6,742 Gb sequenced) ended in the identification of 4,075 recombination events. In contrast, our approach required the sequencing of a total of 980 million 250-bp long reads (which corresponds to 244 Gb sequenced) to retrieve 18,821 recombination events. Thus, the number of recombination events detected per Gb sequenced was over 100 times superior with our method (77.1 events/Gb) than in that by Li et al. (2018) (0.604 events/Gb).

Despite the fact that humans and flycatchers have a recombination rate respectively twice and six times as high as mice (Kawakami et al., 2014, 2017), our approach remained largely more powerful than what two other recent studies achieved on these species *via* a pedigree analysis. Indeed, Halldorsson et al. (2016) sequenced 530 whole human genomes with a sequencing depth of over 30-x (which corresponds to approximately 50,000 Gb sequenced) and detected 485 recombination events (after applying very stringent selection criteria); and Smeds et al. (2016) sequenced the genomes of 11 birds at a mean 42-x coverage (which, since the flycatcher genome is 1.1 Gb long (Ellegren et al., 2012), corresponds to about 500 Gb sequenced) and identified 592 events. Therefore, their approaches respectively led to the detection of only 0.00970 and 1.18 events per Gb sequenced.

Altogether thus, our approach was indisputably much more powerful than comparable studies in detecting recombination events.

5.3.2   The critical step: mapping onto both genomes

Performing the whole procedure twice (once for each of the two parental genomes used as reference) was absolutely critical to the specificity of our approach.

Indeed, when several alignment alternatives exist for a given fragment to map at a particular genomic location, aligners (like BWA) are programmed to select the alternative with the best score. But, for the similarity between the mapped fragment and the reference genome to be maximal (Smith and Waterman, 1981), the penalty associated to opening a gap (i.e. for an indel) is generally higher than that associated to a sequence of several mismatches. As a consequence, read extremities, especially when they encompass an indel, are generally misaligned in a way that better matches the sequence of the reference genome than they truly do. In other words, mapping is biased towards the reference genome.

In principle, local realignment around indels corrects a large part of these reference-biased misalignments. Nevertheless, in view of the rarity of recombination events, a non-negligible portion of these misalignments remain and are likely to

lead to spurious detections of recombination events. Therefore, to counterbalance this mapping bias, we decided to perform the whole procedure using consecutively the two parental genomes as the mapping reference.

We found this *modus operandi* to be crucial to the specificity of our method. Indeed, all else being equal (same values for all the other filters), performing the procedure using only the B6 genome as a reference resulted in 1,088,237 events found in the hotspots. This meant that only 1.7% (18,821 out of 1,088,237) of all the events detected on the B6 genome were validated on the CAST genome and thus, that over 98% of the events detected with only one genome used as a reference likely corresponded to FPs.

In contrast, when the procedure was repeated onto the other reference genome (CAST), the FP rate dropped down to below 5% (Table 5.2). Thus, identifying events based on the mapping onto *both* parental genomes was, by far, the most crucial step to the specific detection of recombination events.

5.3.3   Impact of the filters on the false positive (FP) rate

The shortness of sequenced read pairs circumscribed the number of polymorphic sites accessible on each fragment (*median* = 7; *mean* = 7.66) and, to the difference of pedigree analyses where all fragments carry the same allele because they all arise from the same recombination event, the DNA fragments in the sperm we analysed originated from millions of distinct meioses and were thus to be genotyped individually. Therefore, any sequencing error made by the Illumina device — which occurs at low (Fox et al., 2014; Pfeiffer et al., 2018) but non-negligible rates as compared to that of recombination events — may be fatal to the accurate genotyping of unique molecules. The several filtering steps we added all along our pipeline to ensure genotyping a high accuracy all had an impact on the sensitivity and the specificity of our method, as discussed hereunder.

	Hotspot targets			Control targets			
Threshold	# Fragments	# Events	‰	# Fragments	# Events	‰	FP (%)
1+1	28,282,623	172,414	6.10	14,154,263	59,550	4.21	69
2+2	**19,965,597**	**11,059**	**0.554**	**10,563,573**	**75**	**0.0071**	**1.3**
3+3	12,105,005	3,101	0.256	6,734,622	6	0.00089	0.35
4+4	6,378,813	874	0.137	3,840,375	0	0.00	0.0

Table 5.3: Impact of the minimum requirement of *B6*- and *CAST*-typed markers on the FP rate.
A threshold $n_1 + n_2$ corresponds to a minimum requirement of n_1 *B6*-typed and n_2 *CAST*-typed markers for a fragment to be identified as an event (FP or genuine recombination event). The event rate corresponds to the ratio of candidate recombination events over the total number of fragments. The maximum false positive (FP) rate is the ratio of the event rate in control targets over that in hotspots. The line in bold corresponds to the selected threshold: 2 *B6*-typed and 2 *CAST*-typed markers. The values reported in this table correspond to the results obtained on a subset (100,000,000 fragments) of our whole dataset. At this stage of the process, PCR duplicates have not been removed.

First, the filters set for variant selection (no strong departure from the Mendelian transmission of alleles and a minimal number of reads supporting the variant) were not stringent: only 7% of all markers across all sequenced fragments were eliminated, and these corresponded principally to variants located at the extremities of target regions[6].

The filter on the base quality score had a greater impact on the specificity of our method. Indeed, all else being equal, not setting this filter resulted in a 32% rate of FPs (as compared to the 3.7% rate when the filter was on). This shows that most genotyping errors, aside from those originating from misalignments, arose from sequencing miscalls.

As for the removal of PCR duplicates, it divided the total number of events by a factor 2. We note that the fragments we discarded at this step (i.e. those starting and ending at the exact same genomic locations) were likely to be genuine PCR duplicates since the vast majority ($> 95\%$) of pairs of identically-located fragments were found inside the same sperm sample.

Last but not least, we considered that a fragment was a recombination event if at least two of its markers were *B6*-typed and two were *CAST*-typed. This was a

---

[6]The targets spanned 1 kb, but our analysis extended to an additional 1 kb on both the 5'- and the 3'-end of each target, so as to include all sequenced fragments overlapping at least 1 bp of the target (*NB*: the maximum fragment length was 800 bp).

minimum requirement, since applying a less stringent filter of only one *B6*-typed and one *CAST*-typed marker led to a much higher FP rate of 69% (Table 5.3). This means that, in spite of all the filters that were set, the genotyping error rate remained sufficiently high to call spurious recombination events and thus needed to be double-checked (by requiring at least two genotype calls).

However, the obvious limitation of this filter is that it prevents the retrieval of events with conversion tracts overlapping only one polymorphic site. More generally, the detectability of recombination events depends on the polymorphism: the lower the SNP density, the fewer events are detectable. Therefore, the recombination parameters that can be directly observed are likely to differ from the real recombination parameters. This limitation calls for the use of inferential methods to obtain the real (and undetectable) parameters of recombination, as will be described in the following chapter.

6

Characterisation of recombination in mouse autosomal hotspots

Contents

This chapter in brief — *Even if recombinational activity is known to vary by orders of magnitude across individual hotspots, the properties determining this variation are still poorly understood. Further progress in comprehending the basis of these fluctuations can only arise with the thorough examination of individual hotspots but, in mammals, only a handful of these have been directly characterised at high resolution. Here, thanks to the 18,821 events that we detected with the approach developed in Chapter 5, we identified some of the main factors governing the recombinational activity of individual hotspots, we precisely described recombination in over a thousand hotspots and we estimated the hidden biological parameters of recombination through an inferential approach. Overall, this study provides the first global picture of recombination patterns in mouse autosomal hotspots.*

Although many of the molecular details of the recombination process have been dissected (see Chapter 2) and the recombinational activity is known to vary by orders of magnitude across mouse individual hotspots (Paigen et al., 2008), the properties determining this variation are still poorly understood. So far, in mice, only a handful of recombination hotspots have been directly characterised by sequencing recombination products in sperm or oocytes (Yauk et al., 2003; Bois, 2007; Baudat and de Massy, 2007; Ng et al., 2008; Cole et al., 2010a, 2014).

Recently, several genome-wide hotspot maps have been obtained, either by ChIP-seq against either PRDM9 (Baker et al., 2015a), RAD51 or DMC1 (Smagulova et al., 2011) or thanks to the sequencing-based detection of DMC1-bound ssDNA (Khil et al., 2012; Brick et al., 2012) or SPO11 oligos (Lange et al., 2016). However, all these techniques give only indirect information on recombination: ChIP-seq against PRDM9 reflects its binding affinity to a given locus but does not indicate the associated recombination rate; ChIP-seq against DMC1 reveals both the DSB rate and the repair efficiency, but the two phenomena are indistinctable with this sole method; and the sequencing-based detection of SPO11 oligos requires an extremely large amount of material and, thus, so far, is only available for one dataset of *Mus musculus domesticus* mice. As such, all these approaches only provide information on the intermediary steps of recombination, but none at all on its outcome. Therefore, to characterise recombination, it appeared essential to use another method allowing to directly study its products.

Here, to better understand the extent of the variation in recombinational activity and the factors governing it, we precisely characterised recombination in 1,018 hotspots of a mouse F1 hybrid descended from a cross between *Mus musculus domesticus* (strain C57BL/6J, hereafter called B6) and *Mus musculus castaneus* (strain CAST/EiJ, hereafter called CAST). In this chapter, I show how the set of recombination events detected with the approach developed in Chapter 5 allowed us to describe some of the determinants of recombinational activity, better characterise recombination and infer its hidden parameters *via* inferential approaches.

6.1 Determinants of recombinational activity

6.1.1 A high-confidence set of recombination events

To determine whether the recombination rates we observed with our approach were quantitatively accurate, we aimed at comparing our results with those of more classical approaches. We thus used data from Paigen et al. (2008) who examined in detail the recombinational activity of chromosome 1 in a mouse
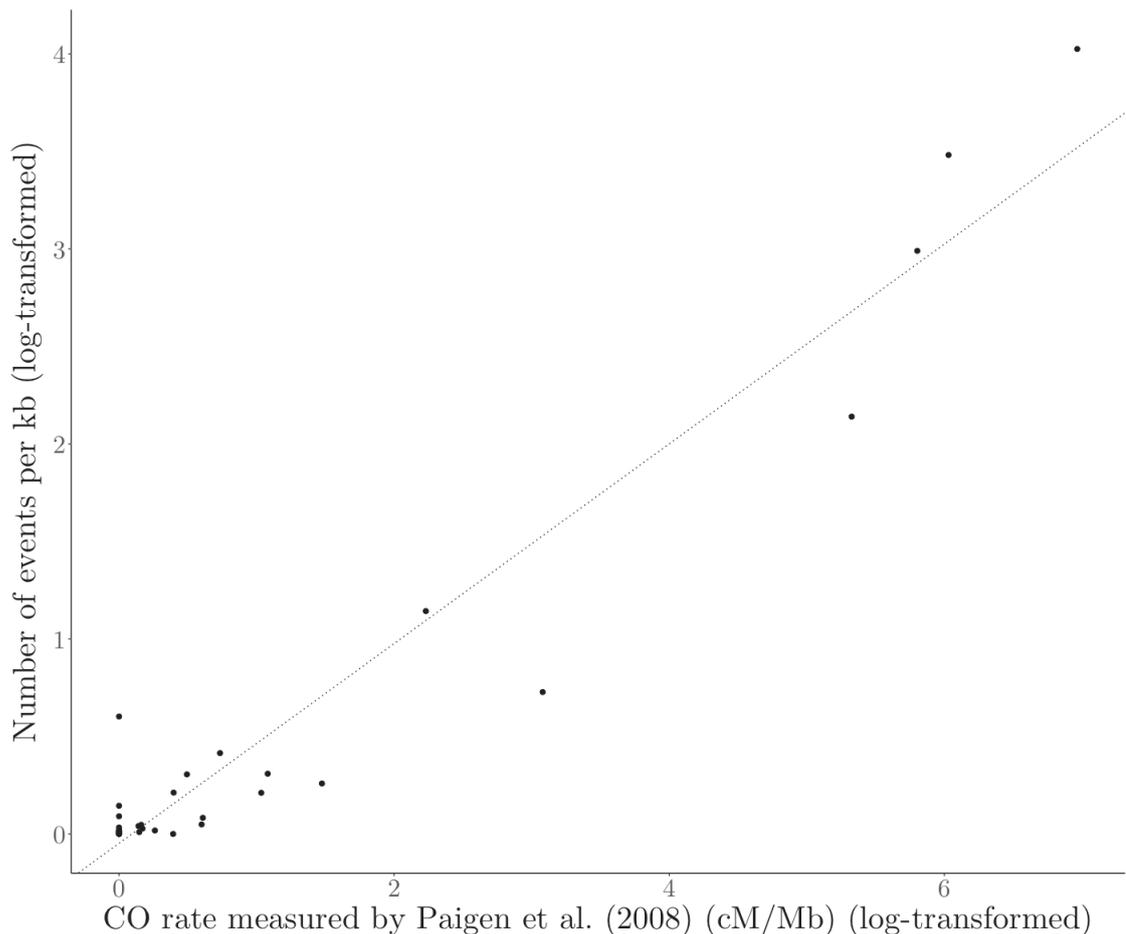
Figure 6.1: Correlation between the expected recombination rate and the observed number of events on the 33 intervals analysed by Paigen et al. (2008). We compared the recombination rates of 33 intervals defined by Paigen et al. (2008) to the total number of events we detected on these intervals, brought back to the length of each interval (see main text). The intervals selected were those which exclusively encompassed hotspots that were analysed in our study. The Pearson correlation between the two measures was extremely high both with raw ($R^2 = 0.974$; p-val $< 2.2 \times 10^{-16}$) and log-transformed ($R^2 = 0.934$; p-val $< 2.2 \times 10^{-16}$) measures.

exhibiting the same genetic background as ours (B6xCAST). They particularly focused on the telomere-proximal 24.7 Mb, which was cut into 65 intervals (*median length* = 205 kb) encompassing a total of 130 *Prdm9* hotspots. Under the asumption that the recombination rate is null outside hotspots, the recombination rate they measured should equal:

$$r_i = \frac{\sum_{h=1}^{n_h^i} (r_h \times L_h)}{L_i} \tag{6.1}$$

where r and L respectively represent the recombination rate and the length of the region considered, the subscripts i and h respectively stand for the interval defined by Paigen et al. (2008) and the 1-kb hotspots we defined, and n_h^i corresponds to the number of hotspots in interval i.

Among the 65 intervals of their study, there were 33 for which all *Prdm9* hotspots were included in our dataset. We thus compared the CO rates that Paigen et al. (2008) measured on these intervals to the total number of recombination events we observed in the 37 hotspots comprised in these intervals, brought back to the length of the interval as given in Equation 6.1 (Figure 6.1). We found that both measures correlated extremely well (Pearson correlation: $R^2 = 0.974$; p-val $< 2.2 \times 10^{-16}$). Therefore, the recombination events we detected are highly reliable since they concord exceptionally well with those identified by this independent study.

6.1.2 Predictors of hotspot intensity

Next, since PRDM9 (Baker et al., 2015a) and DMC1 (Smagulova et al., 2016) ChIP-seq data are available for mice exhibiting the same genetic background (B6xCAST) as ours, we analysed the relationship between these two signals on the one hand, and the recombinational activity of the hotspots we selected on the other hand (Figure 6.2).

Overall, we found that both PRDM9 and DMC1 binding affinity (both proxies of the propensity for a hotspot to form DSBs) are accurate predictors of recombinational activity. Of note, the relationship with PRDM9 binding affinity is linear (regression:
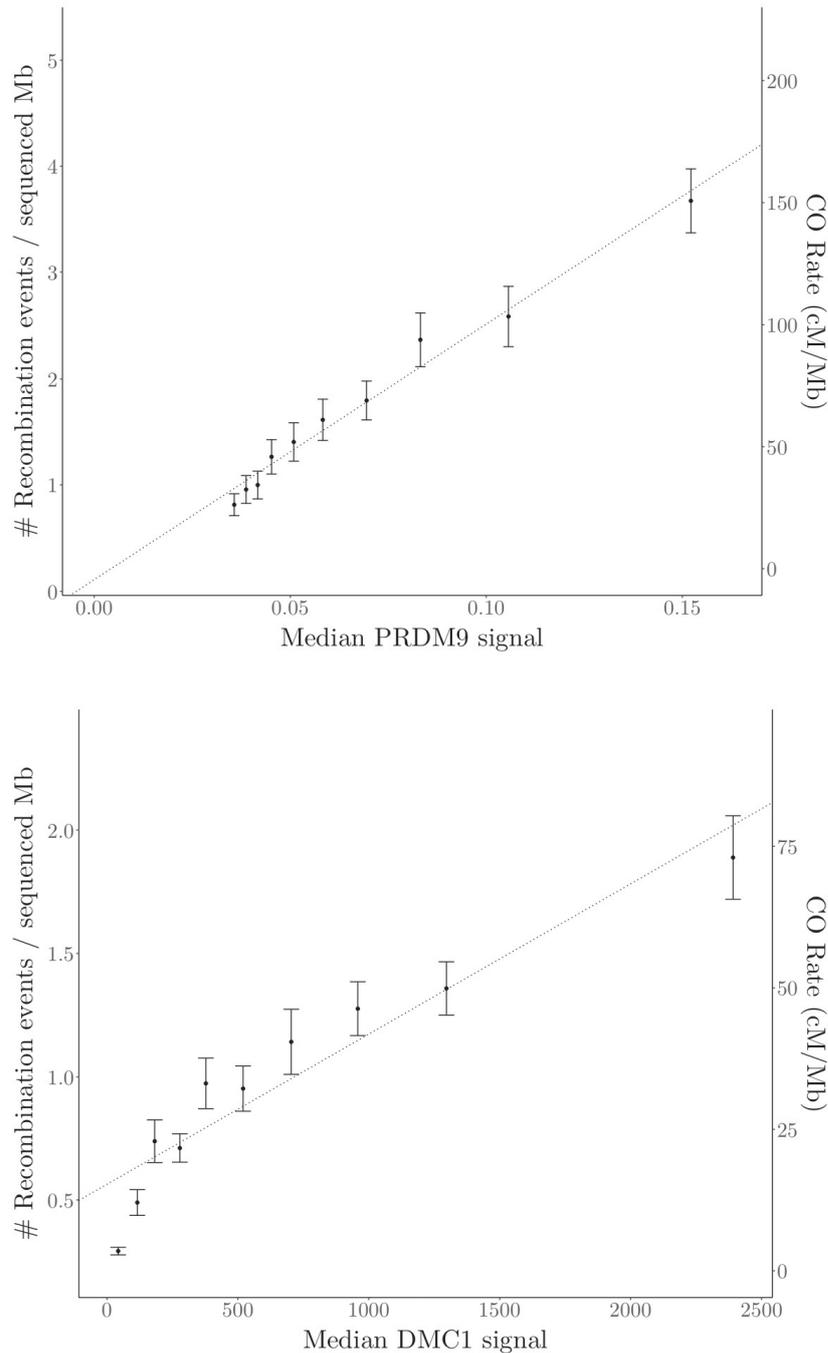
Figure 6.2: Proportionality between the recombination rate and either PRDM9 (top) or DMC1 (bottom) binding intensity.
All 1,018 hotspots were divided into 10 classes of either increasing PRDM9 signal (top), i.e. the number of PRDM9 ChIP-seq tags on each PRDM9 ChIP-seq peak, brought back to the width of the peak (PRDM9 ChIP-seq data on B6xCAST hybrid mice from Baker et al., 2015a), or increasing DMC1 signal (bottom), i.e. the number of DMC1 ChIP-seq tags on each PRDM9 ChIP-seq peak (DMC1 ChIP-seq data on B6xCAST hybrid mice from Smagulova et al., 2016). The observed number of recombination events identified per sequenced Mb (left y-axis) was converted into a CO rate (right y-axis) as detailed in Subsection 6.3.3. The points and error bars respectively represent the mean number of events (or CO rate) and the standard error on the mean for hotspots of each class. The linear regression model for PRDM9 (slope = 1047; intercept = 0; p-val = 4.21×10^{-8}) and DMC1 (slope = 0.027; intercept = 15; p-val = 3.8×10^{-5}) were drawn as dotted lines.

p-val $= 4.2 \times 10^{-8}$) but that with DMC1 signal is not (Figure 6.2.b.), possibly because DMC1 ChIP-seq reflects not only the DSB rate but also the efficiency of the repair of the DSB which varies among hotspots (Lange et al., 2016; Davies et al., 2016).

6.1.3   Lower recombination rate of asymmetric hotspots

In a B6xCAST hybrid, the PRDM9 target motif located on the homologous chromosome originating from the B6 parent (hereafter called B6 haplotype) may differ from that located on the homologue originating from the CAST parent (hereafter called CAST haplotype) because of the accumulation of mutations along the separate lineages (B6 or CAST) since their common ancestor (Davies et al., 2016; Smagulova et al., 2016). Consequently, in certain hotspots, PRDM9 may bind preferentially one of the two haplotypes while, in other hotspots, it may bind both haplotypes equally. The former class of hotspots is referred to as 'asymmetric' and the latter as 'symmetric'.

Li et al. (2018) previously identified that such variations in hotspot asymmetry explain part of the variations in recombination rates for a given DMC1 signal. To check whether this pattern was also observed in our dataset, we distinguished between symmetric and asymmetric hotspots based on the strand-specific PRDM9 ChIP-seq data from Baker et al. (2015a) (see Chapter 7). We found that, for a given PRDM9 or DMC1 signal, the number of recombination events was greater for symmetric than for asymmetric hotspots (two to four times greater) (Figure 6.3).

As hypothesised by Li et al. (2018), this relationship can be explained if these asymmetric hotspots are repaired using the sister chromatid instead of the homologue: since one haplotype is not bound by PRDM9 in such asymmetric hotspots, it is possible that the presence of PRDM9 on both homologues may play a role in homology search. Yet, hotspot asymmetry does not account for all the variation (see the width of boxplots in Figure 6.3): instead, the sampling variance (i.e. the limited number of events per hotspot) and most likely a biological factor not yet identified may explain the residual variation.
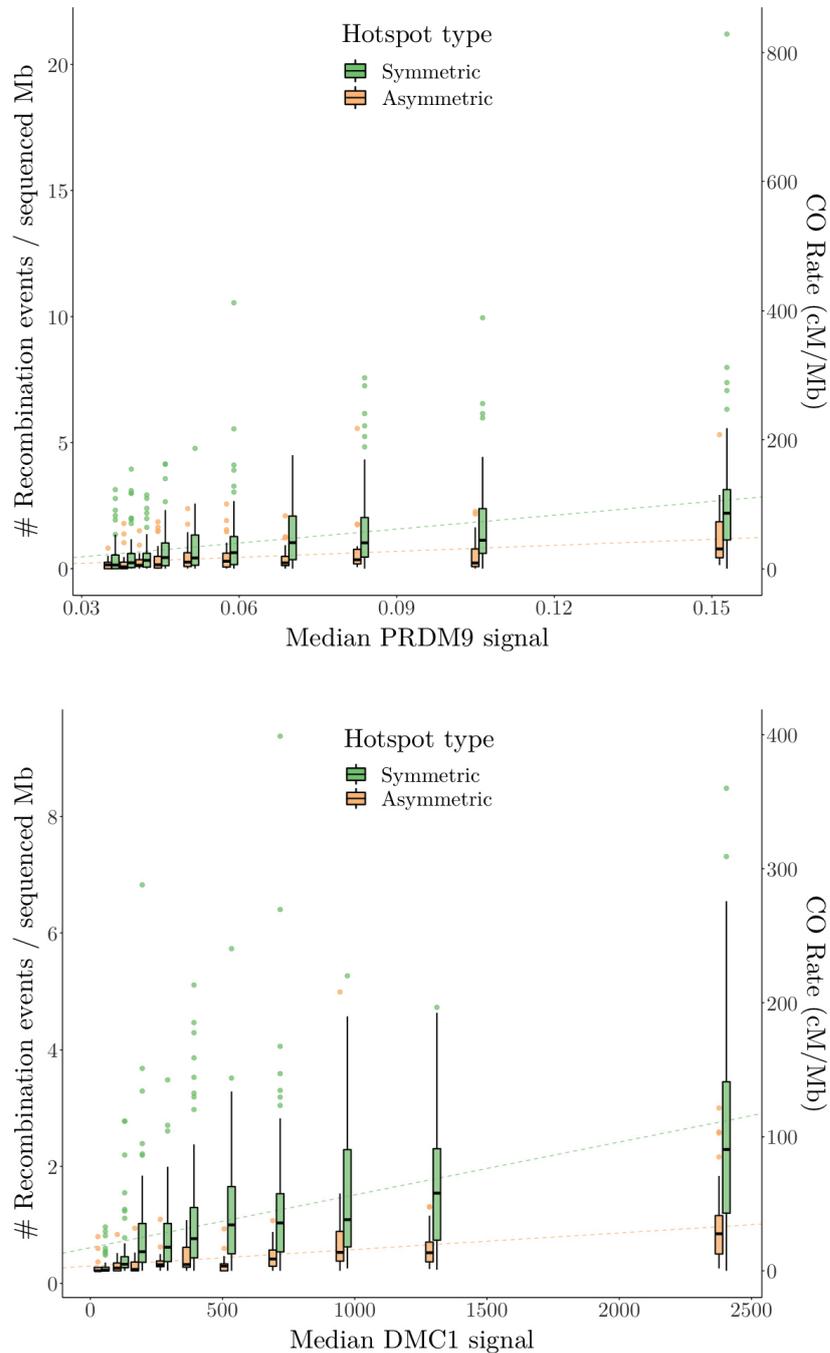
Figure 6.3: Asymmetric hotspots display lower recombinational activity than expected by their PRDM9 (top) or DMC1 (bottom) binding affinity.
All 1,018 hotspots were divided into 10 classes of increasing PRDM9 signal (top), i.e. the number of PRDM9 ChIP-seq tags brought back to the width of the peak (data from Baker et al., 2015a), or increasing DMC1 signal (bottom) (data from Smagulova et al., 2016). The observed number of recombination events identified per sequenced Mb (left y-axis) was converted into a CO rate (right y-axis) as detailed in Subsection 6.3.3. Symmetric hotspots (green, $N = 650$) were distinguished from asymmetric hotspots (orange, $N = 236$) as detailed in Chapter 7. The linear regression model with the PRDM9 signal for symmetric (slope $= 18$; intercept $= 0$; p-val $< 2.2 \times 10^{-16}$) and asymmetric (slope $= 7.9$; intercept $= 0$; p-val $= 9.21 \times 10^{-7}$) hotspots and with the DMC1 signal for symmetric (slope $= 1.8$; intercept $= 0$; p-val $< 2.2 \times 10^{-16}$) and asymmetric (slope $= 8.6$; intercept $= 0$; p-val $< 2.2 \times 10^{-16}$) hotspots are drawn as dotted lines.

6.2   Observable recombination parameters

6.2.1   Definition of observable conversion tracts

To characterise the recombination events we observed, we needed to localise the position of their conversion tracts (CTs). Though, the latter are not directly observable in the data, but they can be *inferred* from 'haplotype switches', i.e. changes of haplotype along a DNA fragment. To avoid any confusion between the real CT and the inferred CT, we decided to denote the latter CT*.

To infer the position of CTs*, we defined 'switch intervals' (black segments in Figure 6.4) as sequence segments delineated by two consecutive markers with distinct genotypes (*B6-CAST* or *CAST-B6*). We also defined the 'switch point' as
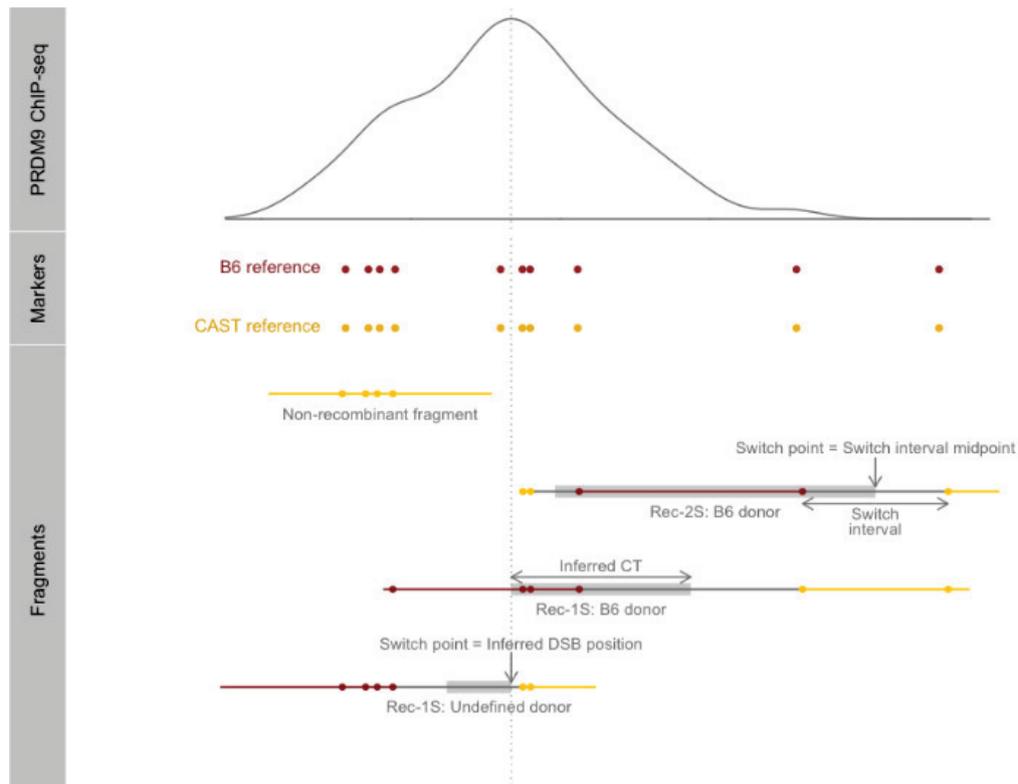
Figure 6.4: Terminology used to characterise recombination events.
The read coverage of the PRDM9 ChIP-seq peak (data from Baker et al., 2015a) is drawn in the top panel ('PRDM9 ChIP-seq'). *B6* (red) and *CAST* (yellow) reference alleles are reported in the middle panel ('Markers'). Examples of sequenced fragments that are or are not recombination events are drawn in the bottom panel ('Fragments'). See main text for the description of each annotation.

the midpoint of the switch interval.

We could distinguish three types of recombination events in our data: 11,665 events including strictly one switch point, which we called 'single-switch recombination events' ('Rec-1S'); 5,932 events including strictly two switch points, which we called 'double-switch recombination events' ('Rec-2S'); and 1,224 events including more than two switch points, which we called 'multiple-switch recombination events' ('Rec-MS').

Rec-2S events most probably correspond to NCO events. Rec-1S events may correspond to either CO events or to fragments that partially overlap a NCO. Rec-MS correspond to complex events (COs or NCOs) and represent only a small fraction (6.5%) of recombinant fragments.

For Rec-2S events, we simply defined the CT* as the region between its two switch points.

For Rec-1S events, only one edge of the CT can be detected (the one corresponding to the switch point). In principle, it is necessary to compare the four products of meiosis to be able to detect the extent of CO CTs. However, previous studies have shown that, in the vast majority of cases, CTs overlap the DSB site (Cole et al., 2014). Hence, the segment between the switch point and the DSB is, in most cases, included in the CT. Thus, for Rec-1S events, we defined the CT* as the region between the switch point and PRDM9 ChIP-seq peak, which colocates precisely with DSB sites (Lange et al., 2016). It should be noted that for Rec-1S events, the CT* corresponds to only one end of the CT (the edge located on the other side of the DSB cannot be detected). Furthermore, when the DSB site is located within the switch interval, the CT* cannot be inferred.

Finally, for Rec-MS fragments, the CT* cannot be inferred either.

6.2.2 Identification of the gene-conversion donor

The conversion tracts that we inferred (CT⋆) directly determine which haplotype (B6 or CAST) is the donor in the gene conversion event (Figure 6.4).

Smagulova et al. (2016) measured the relative proportion of DSB initiation on each haplotype of thousands of hotspots in a mouse exhibiting the same genetic background as ours (B6xCAST). Thus, to assess the accuracy of our inference, we wanted to compare the proportion of B6- and CAST-donor fragments we inferred in each of the hotspot we studied to what would be expected based on their DMC1 ssDNA-sequencing (SSDS) data (Figure 6.5).
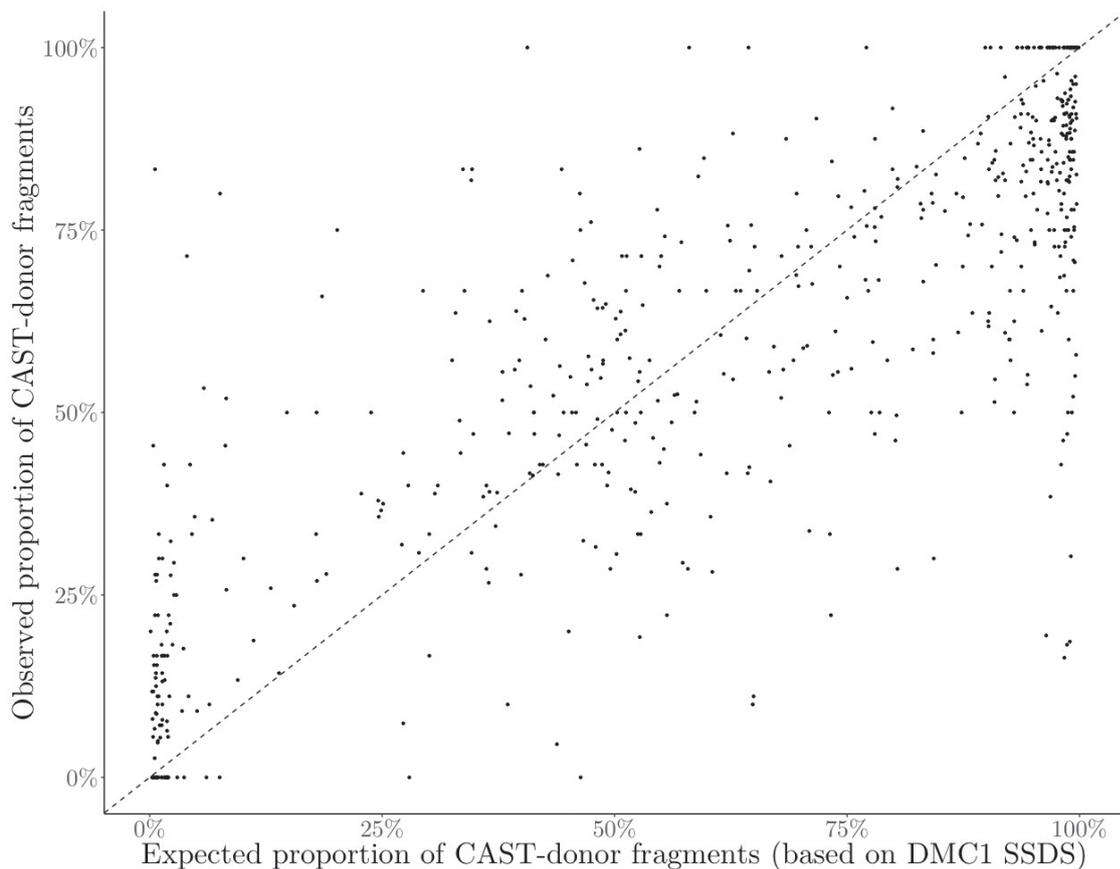
Figure 6.5: Correlation between the expected and the observed proportions of CAST-donor fragments across hotspots displaying at least 5 events.
The expected proportion of CAST-donor fragments (x-axis) was based on the probability that the DSB initiates on the B6 haplotype from DMC1 ssDNA-sequencing (SSDS) data by Smagulova et al. (2016) (see main text). Only the 582 hotspots displaying a minimum of 5 recombination events were reported in this figure. The Pearson correlation between the two measures gave: $R^2 = 0.66$; p-val $< 2.2 \times 10^{-16}$.

Overall, we found a strong positive correlation between the expected and the observed proportions of CAST-donor fragments (Pearson correlation: $R^2 = 0.66$; p-val $< 2.2 \times 10^{-16}$). Our simple and intuitive way of assigning its donor to each fragment was thus sufficient to explain, at the hotspot-scale, most of the variance due to DSB initiation bias.

6.2.3 Description of the recombination events

Among the 18,821 recombination events detected across 898 hotspots (median $= 10$; max $= 327$ events per hotspot), 11,665 corresponded to Rec-1S events, 5,932 to Rec-2S events and 1,224 to Rec-MS events. The CTs* of Rec-1S events (median $= 97$ bp; mean $= 142$ bp) were longer — and, consequently, somewhat more spread (Figure 6.6 and Appendix A) — than the CTs* of Rec-2S events (median $= 78$ bp; mean $= 90$ bp). These features of Rec-1S and Rec-2S are reminiscent of those of COs and NCOs, respectively.

However, our data revealed about twice as many Rec-1S as Rec-2S events — an observation much different from the expected CO:NCO ratio. Indeed, in mice, among the 200–300 DSBs formed per meiosis, 20 are expected to be repaired as COs and the remaining 180–280 as NCOs (Baudat and de Massy, 2007; Martinez-Perez and Colaiácovo, 2009). Since NCOs affect only one of four chromatids (while COs affect two), one would *a priori* expect to identify only one quarter of NCOs (i.e. 45–70) and half of COs ($\approx$10), hence a CO:NCO ratio ranging between 1:4.5 and 1:7.

Two non-mutually exclusive reasons justify the gap between the observed and the expected ratios. On the one hand, the Rec-1S:Rec-2S ratio does not directly reflect the CO:NCO ratio. Indeed, everytime one edge of a sequenced fragment falls into the middle of a NCO CT, this event is necessarily detected as a Rec-1S. Thus, Rec-1S events do not exclusively comprise COs: a portion of them correspond to NCOs. On the other hand, for NCOs to be detected with our approach, their CTs must be long enough to overlap at least two markers (see Chapter 5). Though, since NCO CTs are only a few base pairs to a few tens of base pairs long (Cole et al.,

2014), one would *a priori* expect a non-negligible (unknown at this stage, but see Subsection 6.3.3) proportion of them to be intrinsically undetectable, especially in regions with low SNP density.

Therefore, to characterise recombination regardless of these two limitations of direct observations, it appeared necessary to use inferential approaches to uncover the true[1] recombination parameters. This is developed in the following section.

6.3 Inferred recombination parameters

6.3.1 Approximate bayesian computation (ABC)

In order to discover which range of values of the biological parameters were compatible with our observations, we implemented an approximate bayesian computation (ABC) approach (Csilléry et al., 2010; Sunnåker et al., 2013). In short, an ABC consists in creating a simulator that reproduces at best the biological experiment, performing a large number of simulations with variable input parameters and assessing which range of values are biologically relevant by confronting the summary statistics representative of the output of the simulations to the biological observations.

Implementation of the simulator

We built a simulator that mimicked the formation of recombination events, their sequencing and their genotyping. Briefly, all simulated recombination events were distributed across the 1,018 hotspots proportionately to their predicted propensity to form DSBs, which we approximated by their PRDM9 signal intensity (i.e. the number of tags per kb on each PRDM9 ChIP-seq peak from Baker et al., 2015a). For each simulated hotspot, the ratio of CO over NCO recombination events was $r_{CO:NCO}$.

---

[1]I use 'true' as opposed to 'observed', but the parameters that are inferred correspond to the most likely ones and not necessarily to the exact real ones.
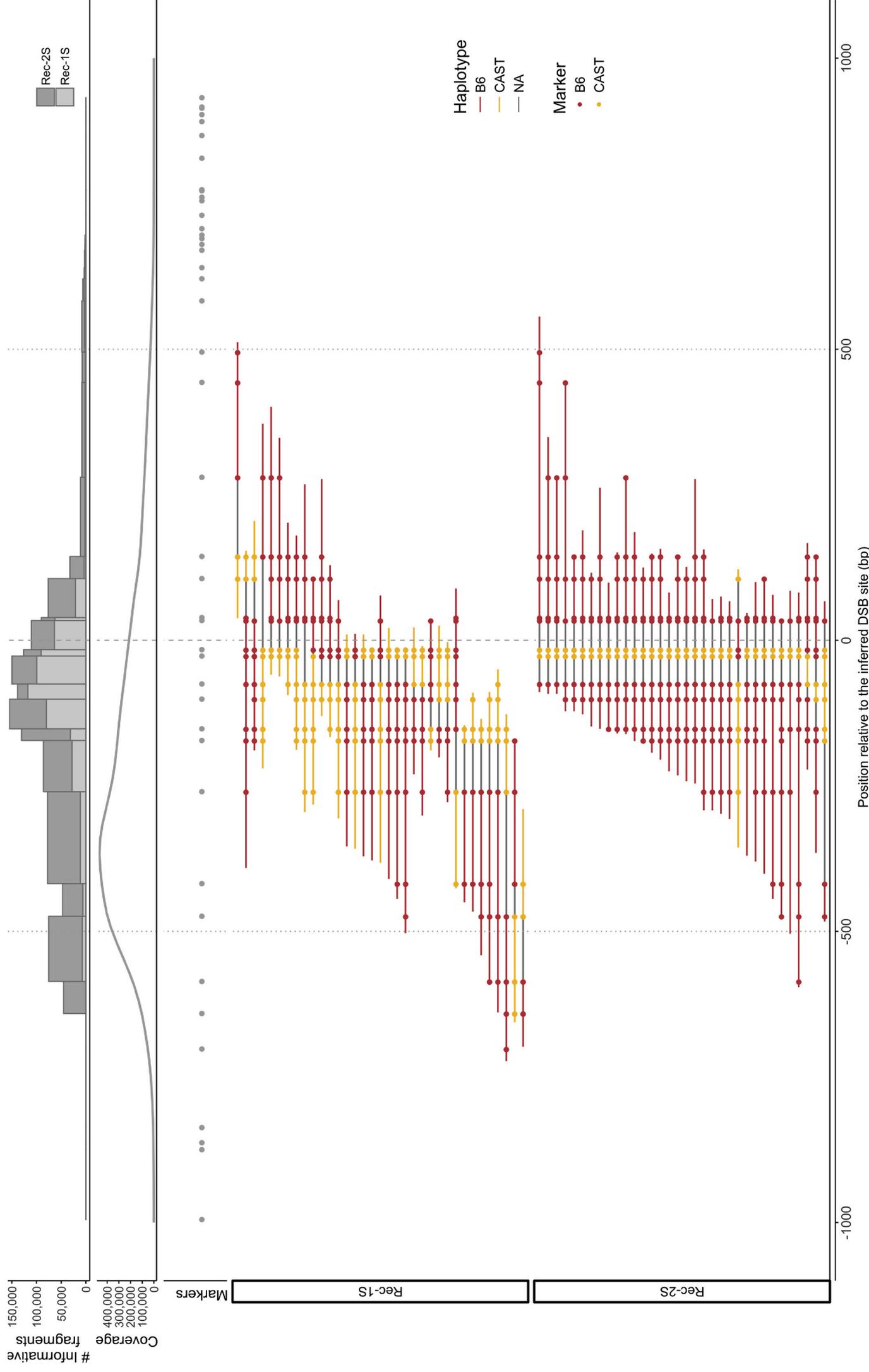
Figure 6.6: Recombination events in a PRDM9^{CAST}-targeted hotspot located on chromosome 11 (chr11:10175985–10177504). The figure is centred on the PRDM9 ChIP-seq peak summit. The top panel depicts the maximum number of detectable Rec-1S (light grey) and Rec-2S (dark grey) switch intervals. The two middle panels indicate variation in read coverage and the positions of markers (filled circles). The bottom panel pictures the detected Rec-1S (upper board) and Rec-2S (bottom board) events. *B6*-typed markers and intervals are coloured in red, *CAST*-typed markers and intervals in yellow, and switch intervals in grey.

SNPs, insertions and deletions were positioned along each hotspot at the exact locations were they were found by variant-calling on our sequencing data (see Chapter 5). CO CT lengths were drawn from a normal distribution of mean m_{CO} and standard deviation sd_{CO}, and NCO CT lengths were drawn from a gamma distribution of alpha $\frac{m_{NCO}^2}{sd_{NCO}^2}$ and beta $\frac{sd_{NCO}}{m_{NCO}}$ (i.e. a distribution with mean m_{NCO} and standard deviation sd_{NCO}). The middle point of the CT for both COs and NCOs was positioned at the inferred DSB site (i.e. the summit of the PRDM9 ChIP-seq peak) and each recombination event was assigned a donor (either B6 or CAST) under a binomial distribution with probability 0.5.

For each simulated recombination event, we randomly selected one of the two gametes involved in the recombination event and simulated $n_{fragments}$ sequenced fragment, whose start and end positions were drawn from the real positions of the fragments in our experimental dataset. We ran our unique-molecule genotyping pipeline (see Chapter 5) on all the simulated fragments to identify those that would be detected as recombination events.

Selection of the simulations compatible with the experimental data
In total, we simulated 100,000 datasets $\mathcal{D}^*$ by assigning a value taken from the following prior distributions to each of the input parameters:

- $m_{CO} \hookrightarrow \mathcal{U}([100, 1000])$ bp,

- $sd_{CO} \hookrightarrow \mathcal{U}([50, 300])$ bp,

- $m_{NCO} \hookrightarrow \mathcal{U}([1, 300])$ bp,

- $sd_{NCO} \hookrightarrow \mathcal{U}([1, 100])$ bp,

- and $r_{CO:NCO} = 10^r$ with $r \hookrightarrow \mathcal{U}([-2, 1])$,

where $\mathcal{U}$ represents the uniform distribution.

For each simulated dataset as well as for the experimental dataset, we summarised the results of the recombination events found with the following summary statistics:

the observed Rec-1S:Rec-2S ratio $r^{obs}_{Rec-1S:Rec-2S}$, the observed mean and quartiles of Rec-1S CT* lengths (l^{mean}_{Rec-1S}, $l^{0.25}_{Rec-1S}$, $l^{0.5}_{Rec-1S}$, $l^{0.75}_{Rec-1S}$) and the observed mean and quartiles of Rec-2S CT* lengths (l^{mean}_{Rec-2S}, $l^{0.25}_{Rec-2S}$, $l^{0.5}_{Rec-2S}$, $l^{0.75}_{Rec-2S}$).

We then used the R package 'abc' (Csilléry et al., 2012) to select the simulated datasets $\mathcal{D}^*$ that ended in summary statistics $\mathcal{S}^*$ close to the summary statistics $\mathcal{S}$ of the experimental dataset $\mathcal{D}$, with a tolerance threshold (ϵ) of 5% (i.e. $\mathcal{D}^*$ was retained if $d(\mathcal{S}^*, \mathcal{S}) \leq \epsilon$).

6.3.2   Comparison with direct observations

CO and NCO CT lengths had previously been measured in two hotspots *via* the analysis of mouse tetrads (Cole et al., 2014) and the mouse CO:NCO ratio had been determined *via* cytological estimates of DSB numbers: among about 250 DSBs arising in each meiosis, around 23 are repaired as COs (Baudat and de Massy,

Parameter	Literature	ABC approach
CO:NCO ratio	0.1[1]	0.119 [0.014–0.20]
CO CT length (bp)		
Mean	566[2]	447 [245–874]
Sd	277[2]	363 [92–471]
Detectable NCO markers CT* length (bp)		
Mean	94[2]	95 [74–110]
Sd	62[2]	49 [30–60]
Real NCO CT length (bp)		
Mean	-	36 [4–54]
Sd	-	45 [3–86]

Table 6.1: Consistency between the recombination parameters inferred *via* our ABC approach and those directly measured by independent studies. References from which the values were extracted are given inside superscript brackets: [1] corresponds to Cole et al. (2010a) and [2] corresponds to Cole et al. (2014). CT stands for 'conversion tract' and CT* for 'inferred (or observable) conversion tract'. Only a portion of NCOs are detectable by tetrad analyses (those whose CT* overlaps at least 1 marker). Thus, we report the CT* length of both this subset of detectable NCOs that have been analysed by tetrad analyses, but also report the mean CT length of all NCOs (both detectable and undetectable). For the ABC, the 95% confidence intervals are reported between brackets.

2007; Martinez-Perez and Colaiácovo, 2009), which, with the assumption that the remaining ones are repaired as NCOs, leads to a rough estimate for the CO:NCO ratio of 0.1.

To assess the correctness of the parameter ranges identified by the ABC, we compared them to the aforementioned estimates (Table 6.1). Altogether, we found that these two sets of parameter ranges were strikingly close. This adequacy was particularly impressive regarding the CO:NCO ratio, considering the fact that we did not set any prior constraint on any of the simulated parameters.

Similarly, the length of the CTs* of detectable NCOs (i.e. those with a CT overlapping at least one marker) estimated by the ABC was almost identical to that directly observed by tetrad analyses. However, this reported CT* length did not take into account that of undetectable NCO CTs (i.e. those too short to overlap any marker). As such, the actual mean CT length for all NCO events is necessarily shorter than that reported by direct observations and can only be provided by the ABC: we estimated it to be around 36 bp (Table 6.1).

As for COs, even if the 95% confidence interval from the ABC included it, the value reported in the literature was slightly higher than the punctual estimate from the ABC. This was likely due to the fact that the summary statistics were compared to the observations of CTs spreading onto a maximum of 500 bp (as the 1-kb hotspots were centred on the DSB site). If, instead, observations had been extended to larger regions, the estimated CTs would surely have been longer (as in Chapter 8).

All in all thus, the ABC allowed to estimate the mean recombination parameters for the 1,018 hotspots we had selected and thus provided a broad insight of recombination patterns in mice.

6.3.3 Extrapolation of recombination parameters

Next, we used the results of the ABC to extrapolate other pieces of information on recombination: the CO rate and the composition in COs and NCOs of the observed Rec-1S events.

Estimation of the average CO rate

Applying our unique-molecule genotyping pipeline on simulated recombination events (as was done with the ABC) allowed us to estimate the proportion of events that are detectable.

We defined the detectability (d) as the ratio of detected recombination events (n) over the total number of recombining gametes that were simulated (N_r):

$$d = \frac{n}{N_r} \tag{6.2}$$

As for the recombination rate (R), it corresponded to the proportion of recombining gametes (N_r) among all the gametes analysed (N_g):

$$R = \frac{N_r}{N_g} \tag{6.3}$$

Combining equations 6.2 and 6.3, we get:

$$R = \frac{n}{d \times N_g}$$

In the 4,997 simulations selected by the ABC, 6.7% of simulated recombination events were discovered, which gave us a direct estimate for d. As for n and N_g, we observed 18,821 recombination events out of 228,984,512 fragments analysed. Using these values, we found that the recombination rate in 1-kb long hotspots was 1.23×10^{-3}. Since 0.119 of all recombination events corresponded to COs (see Table 6.1), the CO rate in 1-kb long hotspots was 1.46×10^{-4} per gamete or also 2.92×10^{-4} per bivalent, i.e. an average recombination rate of 29.2 cM/Mb across all analysed hotspots.
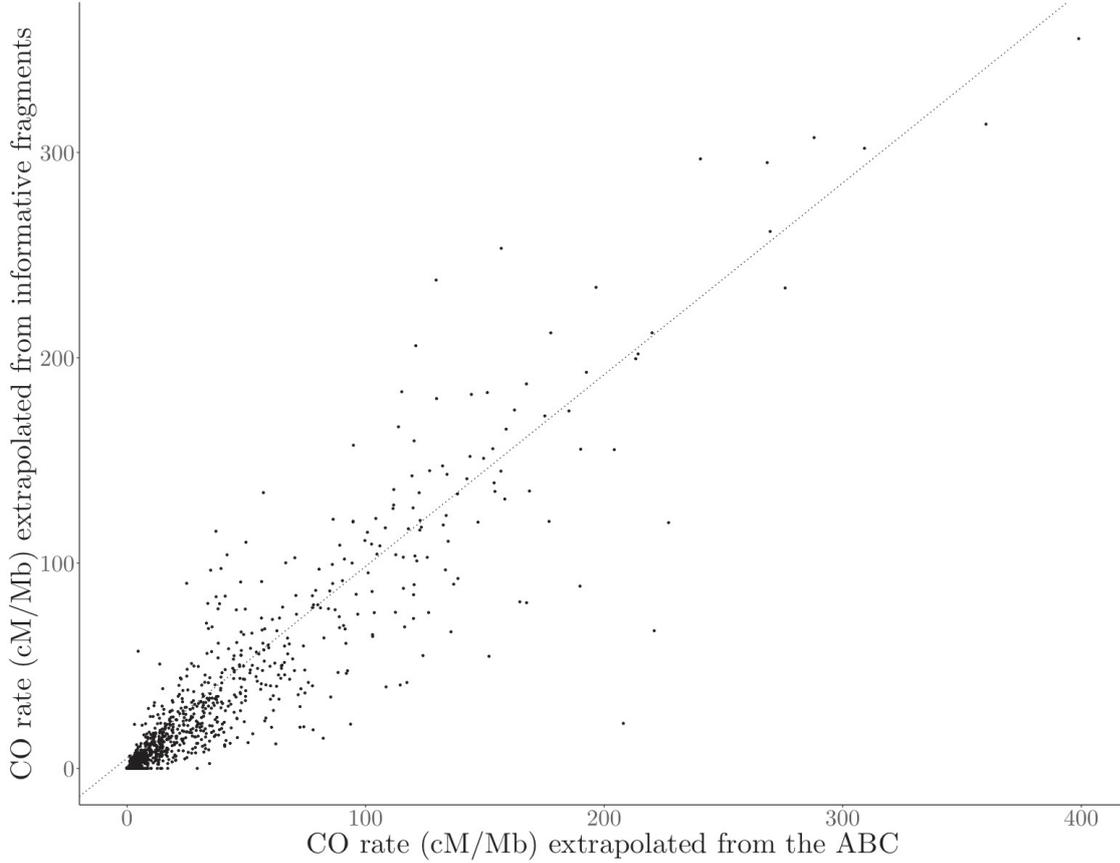
Figure 6.7: Adequacy between two independent manners of extrapolating the CO rate.
For each of the 1,018 hotspots, the CO rate was extrapolated based on the detectability inferred by the ABC (x-axis) and compared to the CO rate extrapolated from the number of informative fragments (y-axis). The calculus for the extrapolation of the CO rate *via* the ABC was the following: $R = \frac{n}{d \times N_g} \times f_{CO}$, where n is the number of recombination events detected, d the detectability (inferred by the ABC), N_g the total number of fragments analysed and f_{CO} the proportion of COs in all the recombination events. This extrapolated CO rate was converted into cM/Mb by multiplying it by $10^2 \times \frac{10^{-3}}{10^{-6}}$ (the 10^{-3} multiplying factor came from the fact that the CO rate was measured on 1-kb long hotspots). As for the extrapolation from the number of informative fragments, our calculus was the following: $R = \frac{n_{Rec-1S}}{L_{seq}^i \times N_f^i}$, where n_{Rec-1S} represents the number of Rec-1S events, L_{seq}^i the length sequenced on each fragment and N_f^i the number of informative fragments (i.e. those overlapping a minimum of 4 markers). This rate was then converted into cM/Mb by multiplying it by $10^2 \times 10^6$. The two measures correlated extremely well (Pearson correlation: $R^2 = 0.84$; p-val $< 2.2 \times 10^{-16}$) and the slope of the linear regression equalled 0.9342 (p-val $< 2 \times 10^{-16}$).

Using the simplifying assumption that the CO:NCO ratio is similar in all hotspots, we used the process just described to transform the number of events in a given hotspot to its recombination rate: this is how the right y-axis (CO rate in cM/Mb) was calculated for Figures 6.2 and 6.3.

Alternatively, the CO rate could be extrapolated independently of the results of the ABC: assuming that Rec-1S events mainly correspond to COs, the CO rate would equal the fraction of Rec-1S events per sequenced base pair where events are detectable. Indeed, recombination events can only be detected in 'informative' fragments, i.e. fragments overlapping at least 4 markers, and this irregularity should not be counted into the CO rate. Thus, when based on the number of informative fragments (N_f^i), their sequenced length (L_{seq}^i) and the number of Rec-1S detected (n_{Rec-1S}), the recombination rate R equals:

$$R = \frac{n_{Rec-1S}}{L_{seq}^i \times N_f^i}$$

We found a remarkable adequacy between those two independent ways of extrapolating the CO rate (Figure 6.7): the slope of the linear regression between the two was extremely close to 1 (slope = 0.9342; p-val $< 2 \times 10^{-16}$).

However, even if these two independent extrapolated estimates of CO rates concorded well, they were 10 times lower than those measured by Paigen et al. (2008) on chromosome 1 (data not shown). As such, our extrapolation may underestimate the actual CO rate by a factor 10. Though, if it is indeed the case, we do not know where the gap comes from.

CO:NCO composition of Rec-1S events

As estimated by the ABC (Table 6.1), for every 1,000 recombination events repaired as NCOs, 119 are repaired as COs and, since NCOs affect only one chromatid when COs affect two, only 500 NCOs are expected to be seen. In the simulations selected

by the ABC, the detectability for COs equalled 0.105 while that for NCOs equalled 0.0548. Based on these estimates, one would expect to detect 27.4 NCOs (out of the 500 chromatids affected) and 12.5 COs (out of the 119 chromatids affected).

Because they encompass only one switch point, all COs should be detected as Rec-1S events. NCOs, however, could be detected as either Rec-1S or Rec-2S events. In the simulations selected by the ABC, 49.9% of all the NCOs detected were detected as Rec-1S. Thus, among the 27.4 NCOs expected, 13.7 should be detected as Rec-1S events and 13.7 as Rec-2S events.

All in all thus, we would expect to detect 26.2 Rec-1S events (13.7 (52.3%) NCOs + 12.5 (47.7%) COs) and 13.7 Rec-2S (all NCOs), i.e. a Rec-1S:Rec-2S ratio of 1.91, thus very close the ratio found experimentally (1.96).

'In relation to any experiment we may speak of this hypothesis as the "null hypothesis," and it should be noted that the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis.'

— Ronald Fisher, *The design of experiments* (1935)

7

Quantification of biased gene conversion in mouse hotspots

Contents

This chapter in brief — *In order to shed new light into the relationship between the intensity of GC-biased gene conversion (gBGC) and the effective population size (N_e), we wanted to precisely quantify the transmission bias (b_0) in a mammalian species with relatively high N_e: mice. We first quantified DSB-induced biased gene conversion (dBGC) in autosomal hotspots and observed that, in our B6×CAST F1 hybrid, dBGC hitchhiked the past gBGC that had occurred in the parental lineages. We then controlled for this confounding effect to quantify gBGC in both COs and NCOs. We found that the transmission bias (b_0) was null for COs and very weak for multiple-marker NCOs. In contrast, single-marker NCOs exhibited a large transmission bias comparable with that observed in humans.*

Considering that the intensity of GC-biased gene conversion (gBGC) at the population-scale (B) is the product of the effective population size (N_e) by the gBGC coefficient (b) (see Chapter 4), the finding that B confines into a very small range of values — even across animals with considerably disparate N_e — was puzzling (Galtier et al., 2018). Logically thus, one or several of the parameters on which b depends — among which the transmission bias b_0 — should vary inversely with N_e.

Though, among mammals, the transmission bias (b_0) has only been measured in humans (Williams et al., 2015; Halldorsson et al., 2016) and consequently, the interplay between b and N_e remains unexplained. In this chapter, I describe how we managed to shed new insight into this relationship by quantifying gBGC in another mammalian species with larger N_e (mice). Since this quantification required to classify hotspots according to their PRDM9 target so as to control for the confounding effect of DSB-induced biased gene conversion (dBGC), I will start this chapter with two sections presenting this process.

7.1 Identification of the PRDM9 target

7.1.1 Methodology to classify hotspots

In a B6xCAST F1 hybrid, hotspots are either activated by the *Prdm9* allele originating from the B6 lineage (*Prdm9^Dom2*) or by that originating from the CAST lineage (*Prdm9^Cst*). To discriminate between these two scenarii, we classified all 1,018 hotspots based on two criteria (Table 7.1).

First, we used PRDM9 ChIP-seq data in the parental B6 and CAST strains from Baker et al. (2015a): when a peak was detected in one parental strain, the hotspot was necessarily targeted by the allele present in that strain, i.e. PRDM9^Dom2 (resp. PRDM9^Cst) when the peak was found in the B6 (resp. CAST) lineage.

When, however, no PRDM9 ChIP-seq peak was detected in either parent (i.e. for novel hotspots detected only in the F1 hybrid), knowing which allele targeted the hotspot was not straightforward. As a substitute, we used information from the

strand-specific detection of PRDM9 ChIP-seq reads (Baker et al., 2015a). Indeed, the proportion of PRDM9 ChIP-seq tags mapping onto one haplotype directly reflects its propensity to be bound by PRDM9 (relatively to that of the other haplotype). Using the assumption that the least bound haplotype had a lower affinity because it had co-evolved with the *Prdm9* allele targeting the hotspot and had thus undergone erosion in the parental lineage, we inferred that $Prdm9^{Dom2}$ (resp. $Prdm9^{Cst}$) was the target when at least 75% of PRDM9 ChIP-seq tags mapped preferentially onto the CAST (resp. B6) haplotype.

7.1.2 Symmetric *versus* asymmetric hotspots

We named the aforementioned class of hotspots displaying large haplotype biases (i.e. those with over 75% of PRDM9 ChIP-seq tags mapping onto one haplotype)

Hotspot category	Selected hotspots		All hotspots	
	Number	Percentage (%)	Number	Percentage (%)
$PRDM9^{Dom2}$-targeted				
tB.sym	24	2.36	181	2.68
tB.chB	63	6.19	267	3.95
NOV.tB.chB	80	7.86	245	3.63
$PRDM9^{Cst}$-targeted				
tC.sym	322	31.63	2,775	41.06
tC.chC	241	23.67	1,370	20.27
NOV.tC.chC	156	15.32	659	9.75
Unclassified	132	12.9	1,261	18.2
Total	**1,018**	**100**	**6,758**	**100**

Table 7.1: Distribution of hotspots into each category of our classification. All 6,758 PRDM9 ChIP-seq-defined hotspots identified by Baker et al. (2015a) and the subset of 1,018 that we selected were classified into 6 categories of hotspots, as described in the main text. Hotspot categories were labeled as follows. 'tB' (resp. 'tC') stands for a PRDM9 allele originating from the B6 (resp. CAST) strain. 'chB' (resp. 'chC') stands for the B6 (resp. CAST) haplotype being the cold one. 'sym' stands for symmetric hotspots, i.e. those having both haplotypes equally targeted by the two PRDM9 alleles. 'NOV' stands for novel hotspots, i.e. those for which no PRDM9 ChIP-seq peak was detected in either parent. Thus, the target (tB or tC) for 'NOV' hotspots was exclusively determined based on the strand-specific mapping of PRDM9 ChIP-seq tags (see main text).

'asymmetric' hotspots, and all the others 'symmetric' hotspots. As mentioned above, such asymmetry materialises the erosion of the target motif in the parental lineage. We further subdivided the group of asymmetric hotspots into two subgroups, based on the presence (or not) of a ChIP-seq peak in the parental strain: either no PRDM9 ChIP-seq peak was detected in either parental strain — in that case, the hotspot had undergone full erosion in the parental lineage and we classified it as a 'novel' hotspot; or a PRDM9 ChIP-seq peak was detected in one of the two parents — in that case, the hotspot had only been partially eroded in the parental strain.

Altogether thus, we could infer both the *Prdm9* allele (*Prdm9^{Dom2}* or *Prdm9^{Cst}*) and the level of asymmetry (symmetric, asymmetric or novel hotspot) and we used these two pieces of information to classify hotspots into six categories (Table 7.1).

All in all, 87% (886) of our 1,018 hotspots fell into one of these categories. Of these, 81% (719) were inferred to be targeted by PRDM9^{Cst} (322 symmetric, 241 partially eroded and 156 fully eroded) and 19% (167) by PRDM9^{Dom2} (24 symmetric, 63 partially eroded and 80 fully eroded).

We note that, as most polymorphic sites in F1 hybrid hotspots result from hotspot erosion in one parental lineage (Smagulova et al., 2016), the requirement of a minimum of 4 markers in the 300-bp central region that we set to select hotspots (see Chapter 5) led to a greater proportion of asymmetric hotspots in our selection (61%) than in the total list of 6,758 hotspots identified by Baker et al. (2015a) (35%).

7.1.3   Validation by detection of the target motifs

Identifying the hotspot-activating *Prdm9* allele at fully eroded hotspots (i.e. those for which no PRDM9 ChIP-seq peak was detected in the parental strains) was done by deduction in lieu of direct observations (see Subsection 7.1.1) and may thus entail errors. Since the accuracy of the inferred *Prdm9* target was critical to the posterior quantification of biased gene conversion, we wanted to make sure that

our predictions were correct by verifying that occurrences of the PRDM9$^{\text{Dom2}}$ (resp. PRDM9$^{\text{Cst}}$) target motif were found in hotspots predicted to be bound by it.

Discovery of the consensus target motifs

To do this, we first had to discover the motif targeted by PRDM9$^{\text{Dom2}}$ and PRDM9$^{\text{Cst}}$. Since the motifs targeted by PRDM9 are known to be located in the vicinity of the DSB site (Brick et al., 2012; Baker et al., 2014), we used the 300-bp central regions of the hotspots undoubtedly targeted by each of the *Prdm9* alleles to discover their consensus motif. For the PRDM9$^{\text{Dom2}}$ consensus motif, the hotspots we used were those for which a PRDM9 ChIP-seq peak had been found in the B6 strain: the B6 and CAST haplotypes of the symmetric PRDM9$^{\text{Dom2}}$-targeted hotspots (tB.sym) and the CAST haplotype of the partially eroded PRDM9$^{\text{Dom2}}$-targeted hotspots (tB.chB). Respectively, for the PRDM9$^{\text{Cst}}$ consensus motif, the hotspots we used were those for which a PRDM9 ChIP-seq peak had been found in the CAST strain: the B6 and CAST haplotypes of the symmetric PRDM9$^{\text{Cst}}$-targeted hotspots (tC.sym) and the B6 haplotype of the partially eroded PRDM9$^{\text{Cst}}$-targeted hotspots (tC.chC).

In practice, to search for the consensus motifs, we used the MEME motif discovery tool (Bailey et al., 2006) from the MEME Suite (version 4.11.2) (Bailey et al., 2009), in the any-number-of-repetitions mode and allowing up to 10 motifs of width comprised between 10 and 30 bp. For each *Prm9* allele, the consensus

(a) *Prdm9^{Dom2}* (Baker et al., 2015a)   **(b)** *Prdm9^{Cst}* (Baker et al., 2015a)

(c) *Prdm9^{Dom2}* (this study)   **(d)** *Prdm9^{Cst}* (this study)
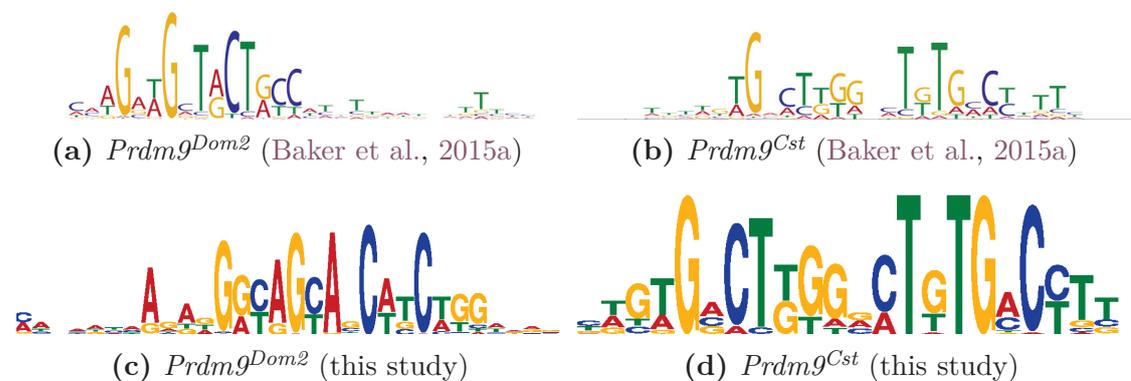
Figure 7.1: Comparison of consensus motifs for *Prdm9^{Dom2}* and *Prdm9^{Cst}*. The consensus motifs for *Prdm9^{Dom2}* (left) and *Prdm9^{Cst}* (right) alleles found by Baker et al. (2015a) are reported at the top and those found in our study at the bottom.

motif we retained was the one with the lowest E-value. We found that, in both cases, they were either identical to or the complement reverse of the ones published by Baker et al. (2015a) (Figure 7.1). We also verified that these consensus motifs were specific to the sequences we selected: we searched for them in control regions defined as sequences located 5-kb downstream of those used to discover the motifs and found that the consensus motif for $Prdm9^{Dom2}$ (resp. $Prdm9^{Cst}$) appeared 10 (resp. 7) times less in these control sequences than in the training set.

Occurrences of consensus motifs in the predicted hotspots

Next, we searched for occurrences of both these consensus motifs in the two haplotypes of each 1-kb long hotspot, using the FIMO tool (Grant et al., 2011) with default parameters. When more than one occurrence of the motif was found in a given hotspot, we retained solely the motif with the highest log-likelihood ratio score.

Altogether, we found that, in hotspots predicted to be targeted by PRDM9^{Dom2}, the majority (76.25%) of the haplotypes predicted to be hot (i.e. targeted by PRDM9) on the basis of the strand-specific detection of PRDM9 ChIP-seq reads from Baker et al. (2015a) (see Subsection 7.1.1) indeed contained a $Prdm9^{Dom2}$ motif. Reciprocally, in hotspots predicted to be targeted by PRDM9^{Cst}, most (72.44%) haplotypes predicted to be hot contained a $Prdm9^{Cst}$ motif.

More precisely, the distribution of motif occurrences along both haplotypes of the hotspots predicted to be targeted by either one of the two *Prdm9* alleles are reported in Figure 7.2. As expected, occurrences of the $Prdm9^{Dom2}$ consensus motif were specific to hotspots predicted to be targeted by PRDM9^{Dom2} and, conversely, occurrences of the $Prdm9^{Cst}$ consensus motif were specific to hotspots predicted to be targeted by PRDM9^{Cst}. In particular, motifs occurred more often in the 'nonself' haplotype (i.e. the B6 haplotype for PRDM9^{Cst}-targeted hotspots and the CAST haplotype for PRDM9^{Dom2}-targeted hotspots), most assuredly because the motif had undergone erosion in its 'self' lineage. Also, these motifs gathered in the close vicinity of the inferred DSB sites (i.e. the summits of the PRDM9 ChIP-seq peaks): 52% of them were located closer than 60 bp away from the DSBs.

(a) Predicted PRDM9^{Dom2}-targeted.

(b) Predicted PRDM9^{Cst}-targeted.

(c) Predicted PRDM9^{Dom2}-targeted.

(d) Predicted PRDM9^{Cst}-targeted.
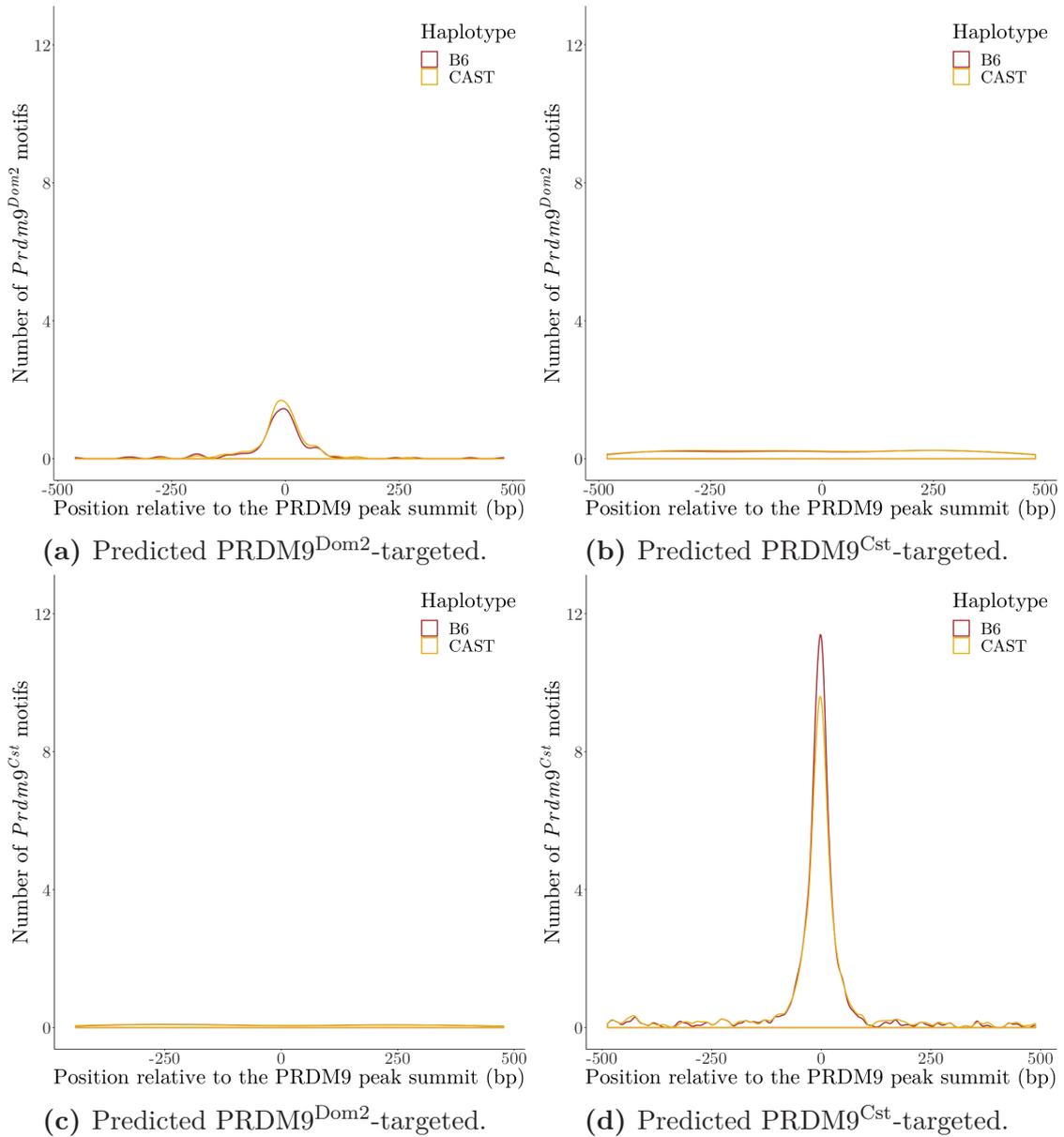
Figure 7.2: Occurrences of $Prdm9^{Dom2}$ and $Prdm9^{Cst}$ consensus motifs along hotspots predicted to be targeted by these alleles.
Occurrences of the consensus motifs for $Prdm9^{Dom2}$ (top) and $Prdm9^{Cst}$ (bottom) were searched in the B6 (red) and CAST (yellow) haplotypes of hotspots predicted to be targeted by $Prdm9^{Dom2}$ (left) or by $Prdm9^{Cst}$ (right). The numbers of hotspots for which the searched motif was found in at least one haplotype were: **(a)** $N = 158$, **(b)** $N = 242$, **(c)** $N = 78$, **(d)** $N = 698$.

7.2 dBGC hitchhiking of past gBGC

7.2.1 Direct quantification of dBGC

Next, we aimed at quantifying DSB-induced biased gene conversion (dBGC) for each hotspot. Thus, we directly extrapolated the dBGC coefficient (b_{dBGC}) from the observed frequency of CAST-donor fragments (x) which we measured in Chapter 6, from the equation of Nagylaki (1983):

$$x = \frac{1}{2} \times (1 + b_{dBGC})$$

We looked at the distribution of the dBGC coefficient across four categories of hotspots (Figure 7.3): on the one hand, the PRDM9$^{\text{Dom2}}$-targeted hotspots which were either fully eroded or still present in the B6 lineage; and, on the other hand, the PRDM9$^{\text{Cst}}$-targeted hotspots completely eroded or still present in the CAST lineage. As expected, we observed that hotspots that are eroded in the parental lineages were those for which the absolute dBGC coefficient was the greatest, while hotspots displaying a quasi-null dBGC coefficient corresponded to symmetric hotspots, i.e. targeted equally by PRDM9.

7.2.2 dBGC and the overtransmission of *GC* alleles

After quantifying the intensity of dBGC, we examined the allelic composition of conversion tracts (CTs) to measure that of gBGC. Among the 30,627 AT/GC (WS) polymorphic sites involved in the CTs of the recombination events detected, 17,876 (58.3%, $CI = [57.8\%; 58.9\%]$) carried the S (G or C) allele. This proportion was slightly lower for Rec-1S (54.8%, $CI = [54.1\%; 55.5\%]$) than for Rec-2S (64.0%, $CI = [63.2\%; 64.9\%]$) events.

However, this observed transmission bias was not solely due to gBGC, but could — in part — come from dBGC. Indeed, on the one hand, at PRDM9$^{\text{Dom2}}$-targeted
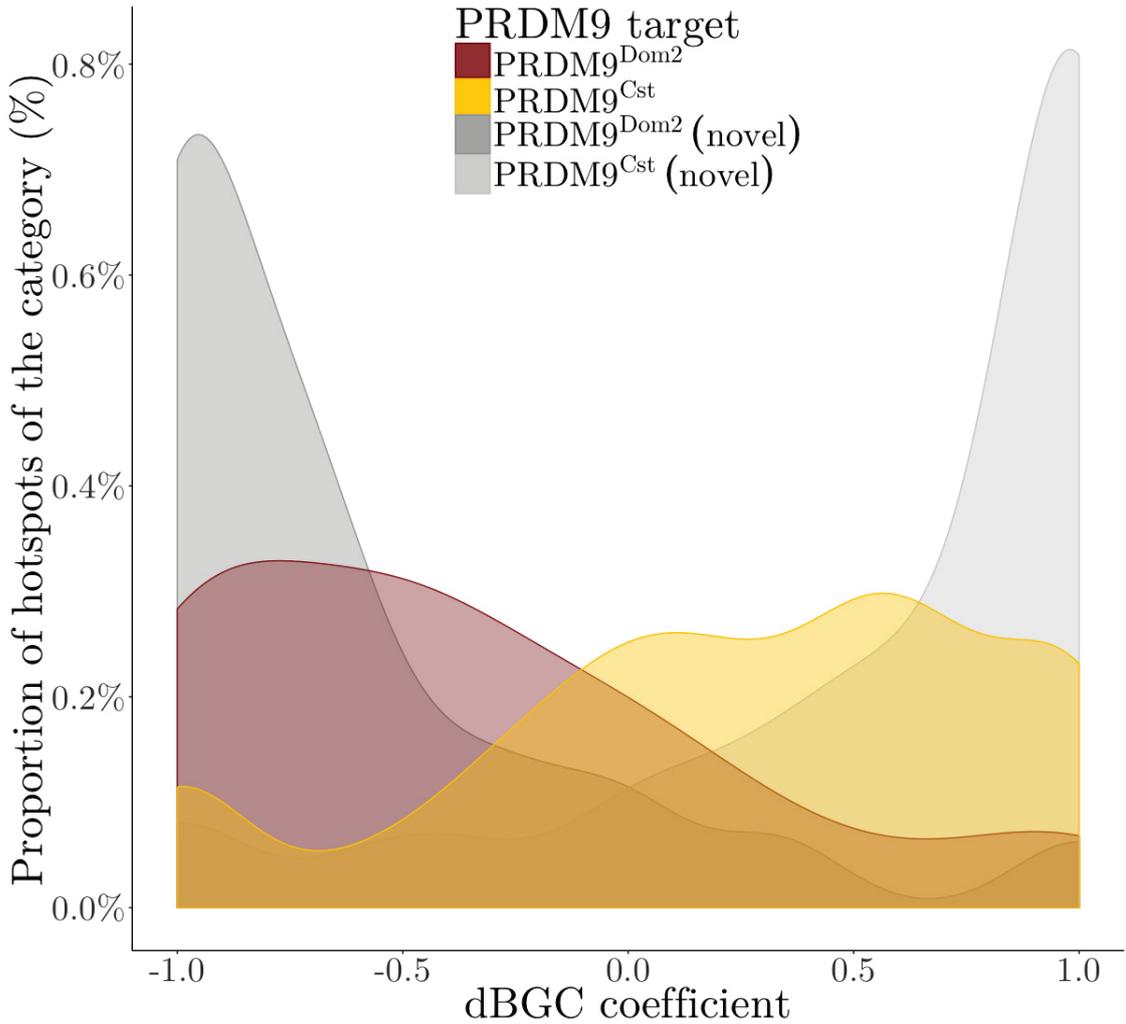
Figure 7.3: Distribution of the dBGC coefficient across categories of hotspots. The dBGC coefficient (b_{dBGC}) was directly extrapolated from the observed frequency of CAST-donor fragments in the pool of gametes (x) as follows: $b_{dBGC} = 2 \times x - 1$ (Nagylaki, 1983) (see main text). The distribution was reported for four groups of hotspots: those targeted by PRDM9^{Dom2} which were either completely eroded (NOV.tB.chB, dark grey) or not (tB.chB and tB.sym, red) in the B6 lineage, and those targeted by PRDM9^{Cst} which were either completely eroded (NOV.tC.chC, light grey) or not (tC.chC and tC.sym, yellow) in the CAST lineage. The frequencies were normalised to the total number of hotspots in each category.
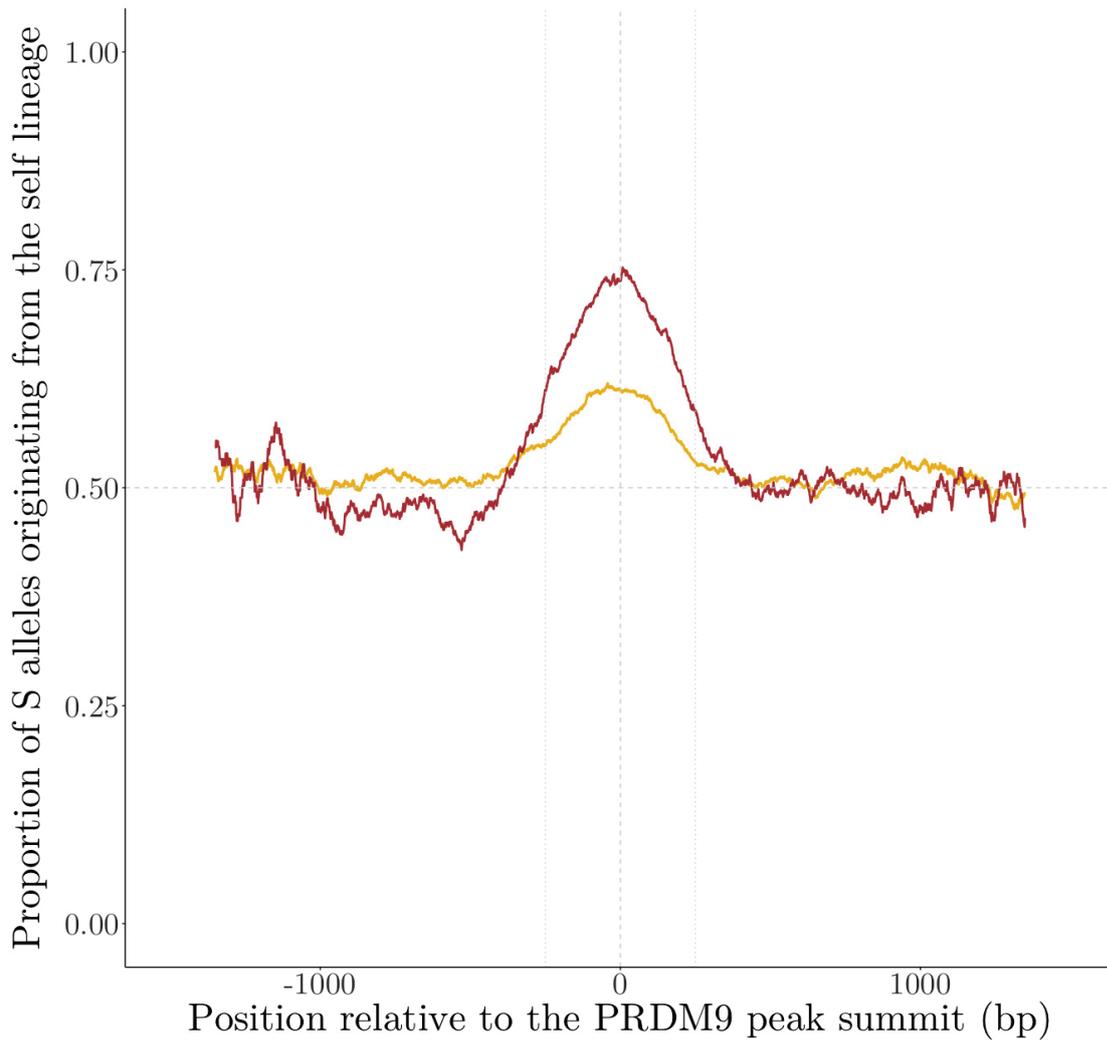
Figure 7.4: GC-profiles at AT/GC (WS) polymorphic sites of PRDM9$^{\text{Dom2}}$- and PRDM9$^{\text{Cst}}$-targeted hotspots.
The proportion of S (G or C) alleles originating from the B6 lineage for PRDM9$^{\text{Dom2}}$-targeted hotspots ($N_{hotspots} = 167$) (red curve) or from the CAST lineage for PRDM9$^{\text{Cst}}$-targeted hotspots ($N_{hotspots} = 719$) (yellow curve) was computed over 300-bp sliding windows.

(resp. PRDM9$^{\text{Cst}}$-targeted) hotspots, the B6 (resp. CAST) haplotype was GC-enriched in the vicinity of the DSB (Figure 7.4). Such local increase in GC-content was a clear signature of the past gBGC that occurred in the parental lineages. Interestingly, we note that this effect was stronger in PRDM9$^{\text{Dom2}}$-targeted than in PRDM9$^{\text{Cst}}$-targeted hotspots, which could be explained by two independent reasons. First, as suggested by Smagulova et al. (2016), the *Prdm9^{Cst}* allele may be younger than the *Prdm9^{Dom2}* one and, consequently, PRDM9$^{\text{Dom2}}$-targeted hotspots

may have undergone more gBGC than PRDM9$^{\text{Cst}}$-targeted hotspots. Alternatively, considering that PRDM9$^{\text{Dom2}}$-targeted hotspots are less numerous than PRDM9$^{\text{Cst}}$-targeted ones in the B6xCAST hybrid (likely because of dominance effects), the subset of PRDM9$^{\text{Dom2}}$-targeted hotspots active in the hybrid may correspond to hotspots with particularly high PRDM9-affinity, which would thus have undergone stronger gBGC in the past lineage.

On the other hand, for the majority (81%, resp. 75%) of conversion events occuring at PRDM9$^{\text{Dom2}}$-targeted (resp. PRDM9$^{\text{Cst}}$-targeted) hotspots, the B6 (resp. CAST) haplotype was the donor (Table 7.2).

In summary, the haplotype which was most often the donor (due to dBGC) was also the GC-richer (due to past gBGC). In other words, dBGC occuring in the hybrid somehow hitchhiked the gBGC that occurred in the past lineages, thus creating a confounding effect to estimate the intensity of gBGC at a single meiotic generation.

Consequently, at this point, cancelling the action of dBGC was absolutely critical to quantify gBGC precisely.

	Donor haplotype		
Hotspot category	**B6**	**CAST**	**NA**
PRDM9$^{\text{Dom2}}$-targeted			
tB.sym	177	164	68
tB.chB	553	165	142
NOV.tB.chB	515	87	170
PRDM9$^{\text{Cst}}$-targeted			
tC.sym	2400	3075	1868
tC.chC	1404	4329	1226
NOV.tC.chC	248	664	284

Table 7.2: Number of B6- and CAST-donor fragments per category of hotspots.
The donor haplotype for each fragment was identified as described in Chapter 6 in each of the six categories of hotspots reported in Table 7.1.

7.2.3 Controlling for dBGC to quantify gBGC

To control for the impact of dBGC onto the transmission bias, we equalised the number of fragments coming from B6-donor and from CAST-donor conversion events. Concretely, we counted, for each hotspot, the total number of B6-donor (n_{B6}) and of CAST-donor (n_{CAST}) fragments. If there were fewer B6- than CAST-donor fragments (resp. fewer CAST- than B6-donor fragments), all B6-donor (resp. CAST-donor) fragments as well as a random selection of n_{B6} fragments among the n_{CAST} CAST-donor (resp. n_{CAST} among the n_{B6} B6-donor) fragments were retained.

To check if this simple method functioned properly, we examined the portions of the fragments located outside the observed CTs (CTs*). By definition, gene conversion does not occur in these DNA chunks and, thus, the allelic frequencies are expected *not* to depart from a 1:1 transmission ratio. We found that the transmission of S and W alleles indeed abode by the Mendelian transmission of alleles (Table 7.3), which confirmed that our per-hotspot equalisation procedure allowed to efficiently control for the dBGC effect.

Category	# S	# W	% S	CI (min-max)	*p*-val	b_0
Inside CTs*						
Rec-1S	5408	5179	0.5108	0.5012–0.5204	0.0267	0.0216
Rec-2S	2261	2078	0.5211	0.5061–0.5360	0.0057	0.0422
Total	**7669**	**7257**	**0.5138**	**0.5057–0.5218**	**0.0007**	**0.0276**
Outside CTs*						
Rec-1S	9355	9433	0.4979	0.4907–0.5051	0.5743	-
Rec-2S	5051	5028	0.5011	0.4913–0.5109	0.8265	-
Total	**14406**	**14461**	**0.4990**	**0.4933–0.5048**	**0.7506**	**-**

Table 7.3: Transmission of the S alleles inside (upper board) and outside (lower board) observed conversion tracts (CTs*) after controlling for dBGC. Controlling for dBGC was operated by subsampling B6-donor and CAST-donor fragments in individual hotspots (see main text). The values reported in this table correspond to the results obtained after one round of random sampling representative of all sampling combinations. # S: Number of S (G or C) alleles in the fragments sampled. # W: Number of W (A or T) alleles in the fragments sampled. % S: Proportion of S alleles in the fragments sampled ($\frac{\#S}{\#S+\#W}$). CI: 95%-confidence interval (test of proportions). b_0: Transmission bias, calculated as $b_0 = 2 \times x - 1$, where x is the mean frequency of a S allele within a pool of gametes coming from a WS heterozygous context.

7.3   Quantification of GC-biased gene conversion

7.3.1   Null b_0 in COs and weak b_0 in multiple-marker NCOs

After controlling for dBGC, it became possible to measure the intensity of gBGC. Indeed, for both Rec-1S and Rec-2S events, the proportion of S alleles inside CTs* (x) was significantly — but weakly — above 50% (Table 7.3). The transmission bias (b_0) could then be calculated directly for both Rec-1S and Rec-2S events as $b_0 = 2 \times x - 1$ (Nagylaki, 1983).

Though, these estimates were not directly representative of the transmission bias in COs and NCOs, since we previously showed that, contrary to Rec-2S events which exclusively comprised NCOs, Rec-1S events were composed of about 52% of NCOs and 48% of COs (see Chapter 6). Assuming that the b_0 for the NCOs observed as Rec-2S events was representative of the b_0 for the NCOs identified as Rec-1S events, we could decompose the transmission bias as such:

$$b_0^{Rec2S} = (b_0^{Rec1S} \times 0.52) + (b_0^{CO} \times 0.48)$$

Using the b_0 values for Rec-1S and Rec-2S reported in Table 7.3, the latter formula resulted in b_0^{CO} equalling 0. Therefore, COs do not contribute to gBGC in mice.

As for NCOs, their contribution to gBGC could be directly extracted from that measured on Rec-2S events: $b_0^{NCO} = 0.0422$, i.e. a 52.11% transmission of S alleles.

All in all thus, the transmission bias was null (or too weak to be detectable) for COs and weak — albeit significant — for the NCOs we detected. One important limitation of our protocol is that we analysed only recombinant fragments overlapping at least two markers for each haplotype (to limit false positives). Hence, NCO events that overlap a single marker (NCO-1) were excluded from this analysis. Given the average length of NCO CTs (36 bp on average, see Chapter 6), NCO-1 events represent a large fraction of NCO events. Thus, we aimed at quantifying the transmission bias in single-marker NCOs as well, this time through an indirect approach that I describe in the following subsection.

7.3.2   Strong b_0 in single-marker NCOs

To fish NCO-1 events out, we mapped all reads on both the B6 and the CAST
reference genomes and checked, for all variants, (1) that the allele supporting the
genotype call with the mapping onto the B6 genome was identical to that based
on the mapping onto the CAST genome, (2) that the Phred quality score was
greater than 20 and (3) that the allelic frequencies did not show a strong departure
from the Mendelian transmission. We then designated all fragments containing
one *B6*-typed marker surrounded by *CAST*-typed markers on both sides (resp. one
CAST-typed marker surrounded by *B6*-typed markers on both sides) as potential
NCO-1 events ('pot-NCO-1'). We found 147,792 such pot-NCO-1 within hotspots
and 62,074 within control regions. Under the assumption that the recombination
rate in control regions is null, this implies that 90.0% of pot-NCO-1 events detected
within hotspots are false positives (Table 7.4), which meant that as few as 14,766
of the pot-NCO-1 events detected within hotspots corresponded to genuine NCO-1
events.

To investigate the origin of these FPs, we measured the base-specific sequenc-
ing error rate by analysing the frequency of *de novo* variants observed at non-
polymorphic sites, directly in our sequencing data (see Appendix A). The rate of
base-substitution sequencing errors (i.e. ignoring indels) varies among bases from

Target category	Nb of targets	Nb of fragments	Nb of events	Event rate ($\times$ 10^{-6})
Hotspots	1,018	228,984,512	147,792	645.4
Controls	500	106,850,906	62,074	580.9
FP rate				**90.0 %**

**Table 7.4:   Number of pot-NCO-1 events detected in hotspot and control
targets.**
Pot-NCO-1 events were detected as detailed in the main text. All fragments or events
overlapping at least 1 bp with a given target are counted in this table. The event rate
corresponds to the ratio of candidate recombination events over the total number of
fragments. The maximum false positive (FP) rate is the ratio of the event rate in control
targets over that in hotspots.

3×10^{-5} to 10^{-4} per bp. This source of error accounts for 66.7% (CI = [60%; 78%]) of detected FPs (see Appendix A). Among all pot-NCO-1 events, we observed an excess of S $\rightarrow$ W over W $\rightarrow$ S potential conversion events ($\frac{WS}{WS+SW} \sim 0.39$). This is in large part explained by the fact that the pattern of sequencing errors is biased: S bases were more often mistakenly sequenced as W bases than the other way round (see Appendix A).

Interestingly, we found that this ratio was significantly higher in hotspot regions (0.356, $CI_{95\%} = [0.3532, 0.3594]$) than in control regions (0.317, $CI_{95\%}$ = [0.3129, 0.3220]). Under the assumption that the pattern of sequencing errors is the same in hotspots as in control regions, the contribution of FPs and true NCO-1 events to the observed WS/WS+SW ratio can be expressed as:

$$n_{hotspots}^{NCO_1+FP} \times r_{hotspots}^{NCO_1+FP} = n_{hotspots}^{NCO_1} \times r_{hotspots}^{NCO_1} + n_{hotspots}^{FP} \times r_{control}^{FP}$$

where n_j^i corresponds to the counts of events i in regions j, and r_j^i to the observed ratio of $\frac{WS}{WS+SW}$ due to events i in regions j.

Using this formula, we predicted that $r_{hotspots}^{NCO_1}$ equalled 0.70 (i.e. that $b_0^{NCO_1}$ equalled 0.40). This estimate was much higher than what we found for both Rec-1S and Rec-2S events, but, interestingly, it was strikingly close to what had previously been found in humans (Halldorsson et al., 2016) and concorded with recent findings in mice (Li et al., 2018).

Altogether thus, we found that the transmission bias differed tremendously between multiple-marker NCOs (NCO-2+) for which b_0 was extremely weak, and single-marker NCOs (NCO-1) for which b_0 was as high as that of humans. Thus, the overall contribution of NCOs to gBGC depends on the relative proportion of NCO-1 and NCO-2+ events, and these must thus be estimated to finally quantify gBGC.

7.3.3   Global estimation of b_0 for NCOs

To estimate the overall contribution of NCOs to gBGC, we deconvoluted b_0^{NCO} as the sum of the intensity of gene conversion bias in NCO-1 ($b_0^{NCO_1}$) and NCO-2+ ($b_0^{NCO_{2+}}$) events, weighted by the chance for a given NCO CT marker to be involved in a NCO-1 (f^{NCO_1}) or in a NCO-2+ ($f^{NCO_{2+}}$) event:

$$b_0^{NCO} = b_0^{NCO_1} \times f^{NCO_1} + b_0^{NCO_{2+}} \times f^{NCO_{2+}}$$

The genome-wide level of polymorphism in natural populations of *Mus musculus domesticus* mice was estimated to be around 0.47% (Davies, 2015), a result similar to the 0.55% value previously found on a subset of the genome (Frazer et al., 2007). With such SNP density, we would expect 74.75% of NCO CT markers
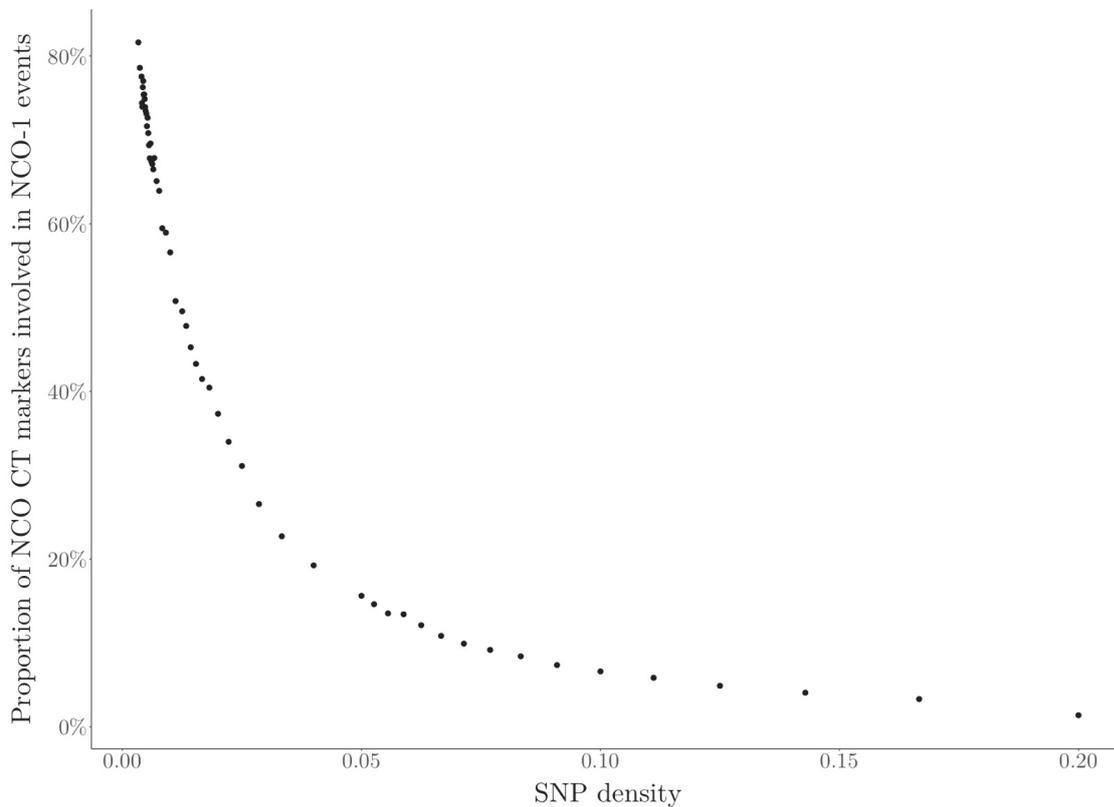
Figure 7.5: Relationship between the proportion of NCO CT markers involved in NCO-1 events and marker density.
We performed simulations to estimate the proportion of NCO-2+ and NCO-1 events given marker density, by distributing a given number of markers (x-axis) along each hotpot and counting the proportion of NCO-2+ and NCO-1 events, i.e. the number of NCO events whose CT overlapped at least two or strictly one marker, respectively.

to come from NCO-1 events (representing 86.23% of the NCOs overlapping at least one marker) and the remaining 25.25% from NCO-2+ events (representing 13.77% of the NCOs overlapping at least one marker) (Figure 7.5), which would result in an overall b_0^{NCO} of 0.310.

8

Methodological adaptations to other studies of recombination

Contents

This chapter in brief — *The method we previously implemented to detect recombination in single individuals can be used to study the role of genes essential to the process of recombination. This requires the use of individuals homozygous for the mutant version of the gene but nonetheless displaying a high level of heterozygosity for recombination to be detectable. As this can only be achieved with F2 individuals, we adapted the method we implemented for simple F1 hybrids to such design. Basically, we had to distinguish the polymorphic sites expressing variation between the two parental genomes from those originating from the third introgressed genome. This implementation was as powerful as the original method and we could thus study the role of the interaction between HFM1 and MLH1: we observed that impeding this interaction led to an increased recombination rate and shortened CO conversion tracts.*

As the method we implemented in Chapter 5 allows to detect recombination in single individuals, it can be used to study the individual role of genes involved in the process of recombination. In particular, Bernard de Massy and Valérie Borde are interested in the specific role of the mouse gene *Hfm1* whose yeast homologue (*MER3*) codes for a meiosis-specific DNA helicase (Nakagawa and Ogawa, 1999; Nakagawa and Kolodner, 2002b) that participates in CO control and in DNA heteroduplex extension (Mazina et al., 2004; Nakagawa and Kolodner, 2002a). This gene is also essential to CO formation in other fungi (Sugawara et al., 2009), plants (Mercier et al., 2005; Chen et al., 2005), humans (Tanaka et al., 2006) and mice (Guiraldelli et al., 2013).

It was recently shown that, in yeasts, Mer3 can connect the MutLβ heterodimer of Mlh1-Mlh2 and that this interaction limits CT lengths genome-wide (Duroc et al., 2017). In mice, the interplay between HFM1 and MLH1 is conserved, but whether or not its role in regulating CT length is also maintained remains a mystery. To find that out, the laboratories of Valérie Borde and Bernard de Massy introgressed a punctual mutation that impedes the interaction between HFM1 and MLH1 (*Hfm1^{KI}*) into F2 individuals, as I detail in the first section of this chapter. In this experimental design, the individuals studied contain three genetic backgrounds and thus, our method to detect recombination needs to be refactored. I describe in the last two sections of this chapter how we worked this out and what the preliminary results of this analysis were.

8.1 Experimental design

8.1.1 Introgression of the mutant *hfm1* allele

A mutant *Hfm1* allele (*Hfm1^{KI}*) was introduced in the zygote of a cross between two F1 mice deriving from hybridisations between two *Mus musculus domesticus* strains: strain C57BL/6J, hereafter called B6 and strain DBA/2J, hereafter called DBA2. The resulting founder mice (F0#2 and F0#3) were thus heterozygous for the *Hfm1*

(a) Ancestry of S28353 and S28355.

Mouse ID	Relationship	% B6	% DBA2	% CAST	HFM1	Mother	Father
39856	Maternal grandmother	0.0	0.0	100.0	WT/WT	N/A	N/A
28130	Maternal grandfather	75.0	25.0	0.0	KI/WT	72205	N/A
F0#2 (72205)	Paternal grandmother	50.0	50.0	0.0	KI/WT	N/A	N/A
N/A	Paternal grandfather	100.0	0.0	0.0	WT/WT	N/A	N/A
22228	Mother	37.5	12.5	50.0	KI/WT	39856	28130
28196	Father	75.0	25.0	0.0	KI/WT	72205	N/A
28353	Mutant analysed	56.25	18.75	25.0	KI/KI	22228	28196
28355	WT analysed	56.25	18.75	25.0	WT/WT	22228	28196

(b) Ancestry of S28367.

Mouse ID	Relationship	% B6	% DBA2	% CAST	HFM1	Mother	Father
F0#3 (72212)	Maternal grandmother	50.0	50.0	0.0	KI/WT	N/A	N/A
N/A	Maternal grandfather	100.0	0.0	0.0	WT/WT	N/A	N/A
28163	Paternal grandmother	75.0	25.0	0.0	KI/WT	72205	N/A
39978	Paternal grandfather	0.0	0.0	100.0	WT/WT	N/A	N/A
28172	Mother	75.0	25.0	0.0	KI/WT	72212	N/A
28238	Father	37.5	12.5	50.0	KI/WT	28163	39978
28371	Mutant analysed	56.25	18.75	25.0	KI/KI	28172	28238

(c) Ancestry of S28371.

Mouse ID	Relationship	% B6	% DBA2	% CAST	HFM1	Mother	Father
39856	Maternal grandmother	0.0	0.0	100.0	WT/WT	N/A	N/A
28130	Maternal grandfather	75.0	25.0	0.0	KI/WT	72205	N/A
F0#2 (72205)	Paternal grandmother	50.0	50.0	0.0	KI/WT	N/A	N/A
N/A	Paternal grandfather	100.0	0.0	0.0	WT/WT	N/A	N/A
28250	Mother	37.5	12.5	50.0	KI/WT	39856	28130
28198	Father	75.0	25.0	0.0	KI/WT	72205	N/A
28371	WT analysed	56.25	18.75	25.0	WT/WT	28250	28198

Table 8.1: Genealogy of the four mice analysed.
The genealogies (parents and grandparents) of each of the two mutant mice (IDs: 28353 and 28371) and of the two wild-type (WT) mice (IDs: 28355 and 28367) analysed in this study, as well as the characteristics (background composition in B6, CAST and DBA2 genomes, and the *Hfm1* alleles carried: either the mutant impeding the interaction between HFM1 and MLH1 (KI) or the wild-type (WT) allele) of all the individuals involved in the ancestry are reported in the subtables above: **(a)** 28353 and 28353; **(b)** 28367; **(c)** 28371.

gene ($Hfm1^{WT/KI}$) and their genetic backgrounds were composed of 50% DBA2 and 50% B6 genomes. Further crosses with other B6 and *Mus musculus castaneus* (strain CAST/EiJ, hereafter called CAST) mice resulted in individuals carrying either two mutant alleles for *Hfm1* ($Hfm1^{KI/KI}$), two WT alleles ($Hfm1^{WT/WT}$) or one allele of each ($Hfm1^{WT/KI}$). The genetic backgrounds for these mice were composed of a mixture of B6, DBA2 and CAST genomes (Table 8.1). Of these, two *hfm1* homozygous mutant (28353 and 28367) and two WT (28355 and 28371) male mice were selected for further analysis: their sperm DNA was extracted and sonicated to produce fragments of a mean size of 450 bp.

8.1.2   Target selection, DNA capture and sequencing

Like in Chapter 5, we selected hotspots from the list identified by Baker et al. (2015a) on the basis of PRDM9 ChIP-seq peak detection. We used the same criteria as before: a minimum of 4 SNPs in the 300-bp central region, a strict maximum of 60 sites with low sequence quality in the 1-kb central region and at least 90% of identity between the B6 and the CAST reference genome on at least 80% of the selected region.

Though, since the main aim of this analysis was to test for any effect of the *Hfm1* mutation on CO CT length, we extended the width of our selected hotspots to 3 kb. Thus, the third selection criterium discarded a larger number of candidate hotspots than in Chapter 5, since identity was required on 3 kb instead of 1 kb. In the end, 890 3-kb long hotspots were retained and, as in Chapter 5, 500 control regions were added to that list of targets.

For the efficiency of DNA capture to be identical in both haplotypes, two baits were designed for each of the 1,390 targets: one corresponding to the CAST haplotype and one to the B6 haplotype. We then performed two successive rounds of DNA capture on each of the four DNA samples from the four mice. Libraries were then sequenced by an Illumina device using a 250-bp paired-end protocol, and the sequenced reads were mapped onto the B6 and the CAST reference genomes

Sample		Mapping (%)		Capture efficiency		
Library ID	Library size	Ref. B6	Ref. CAST	# Filtered Fragments	% in targets	# in targets
28355	164,210,468	98.76	98.00	162,168,344	48.62	78,851,718
28371	171,930,499	98.25	98.20	170,081,808	48.63	82,713,025
Total WT	**336,140,967**	**98.84**	**98.10**	**332,250,152**	**48.63**	**161,564,743**
28353	161,294,272	99.15	98.35	159,920,297	48.62	78,851,718
28367	227,590,570	97.91	97.18	222,826,196	37.11	82,713,025
Total mutants	**388,884,842**	**98.42**	**97.67**	**382,746,493**	**48.63**	**186,150,465**

Table 8.2: Sequencing, mapping and capture-efficiency summary metrics.
Reads were mapped onto the B6 and CAST reference genomes, and fragments were filtered as described in Chapter 5. The lines in bold represent the totals for the two WT and the two mutant mice.

as described in Chapter 5. Overall, read mapping statistics and capture efficiency were similar to what was found in Chapter 5 (Table 8.2).

8.1.3 Expected genetic background composition

The point mutation on *Hfm1* originated from B6/DBA2-background founder mice (F0#2 and F0#3) and was introgressed into a B6xCAST hybrid *via* two consecutive crosses: on the one hand, the founder mice were crossed with B6/B6-background mice, thus yielding one 75%-B6/25%-DBA2 parent; on the other hand, other 75%-B6/25%-DBA2 mice were crossed with CAST mice to yield a second parent with a background composed of 37.5% B6, 12.5% DBA2 and 50% CAST genomes (Table 8.1). Each of the four selected mice (28353, 28355, 28367 and 28371) were then obtained by crossing the two aforementioned parents together. Thus, their background encompassed 56.25% B6, 18.75% DBA2 and 25% CAST genomes.

More precisely, the expected genomic proportion (and therefore, the expected proportion of targets) of each genetic background were those reported in Table 8.3. Overall, 68.75% of the targeted loci were expected to be heterozygous (either B6/DBA2, B6/CAST or DBA2/CAST) and could, in principle, be used to detect recombination events.

Detailed		Simplified	
Background	**% expected**	**Background**	**% expected**
B6/B6	28.125		
B6/DBA2	18.750	DOM/DOM	50.0
DBA2/DBA2	3.125		
B6/CAST	37.500	DOM/CAST	50.0
DBA2/CAST	12.500		
CAST/CAST	0.000	CAST/CAST	0.0

Table 8.3: Expected distribution of genetic backgrounds in the mice analysed. Because the B6 and DBA2 genomes present high sequence conservation (Davis et al., 2005), we regrouped them under the label 'DOM'. The expected genomic proportion (and thus proportion of targets) in each of the six possible 'detailed' backgrounds were reported on the left panel and the expected proportions in each of the three 'simplified' backgrounds were reported on the right panel.

However, the power to detect recombination depends on the density of heterozygous sites, and the latter is much lower at B6/DBA2-background targets than at B6/CAST- or DBA2/CAST-background loci. Indeed, the B6 and the DBA2 genomes present a low sequence divergence of 0.2% (Keane et al., 2011) because these two strains derive from the same mouse subspecies (*Mus musculus domesticus*) from which they inherited large genomic regions (Davis et al., 2005). We note that, since the latter two strains derive from the same subspecies, we will regroup the labels B6 and DBA2 under a more general notation: 'DOM'. In comparison, as the DOM (B6 or DBA2) and CAST strains derive from two distinct subspecies which diverged about 350,000 to 500,000 years ago (Geraldes et al., 2008), they present a much higher genome-wide divergence of 0.74% (Keane et al., 2011).

Therefore, in order to avoid any spurious fluctuation in detectability between individuals and to thus allow the comparison of recombination rates across samples, we chose to search for recombination events exclusively in one type of heterozygous background. And, so as to maximise the detectability of recombination events, we focused on the background displaying the highest rate of polymorphism: DOM/CAST-background targets. The following section will be dedicated to detailing the procedure we implemented to identify them specifically.

8.2 Detection of recombination in F2 individuals

8.2.1 Inference of the origin of polymorphic sites

Distinguishing the targets of interest (DOM/CAST-background targets) from others (DOM/DOM-background targets) comes back to genotyping the DOM-CAST markers (i.e. the polymorphic sites for which the CAST strain carries an allele different from that carried by the B6 and the DBA2 strains). Though, given that the F2 individuals carry a mosaic of three genomes, three types of polymorphic sites can occur: either the B6, the DBA2 or the CAST genome carries an allele different from that of the other two (Figure 8.1). Therefore, prior to genotyping targets, the DOM-CAST markers must be distinguished from the other (B6-DBA2) markers.

Given the crosses made, no portion of the genome of the F2 individuals could display a CAST/CAST background (Table 8.3). Therefore, if, at a given polymorphic site, at least one of the four individuals is homozygous for the allele carried by the CAST strain, the site necessarily corresponds to a B6-DBA2 marker (Figure 8.1). We distinguished between B6-DBA2 and DOM-CAST markers on this basis.
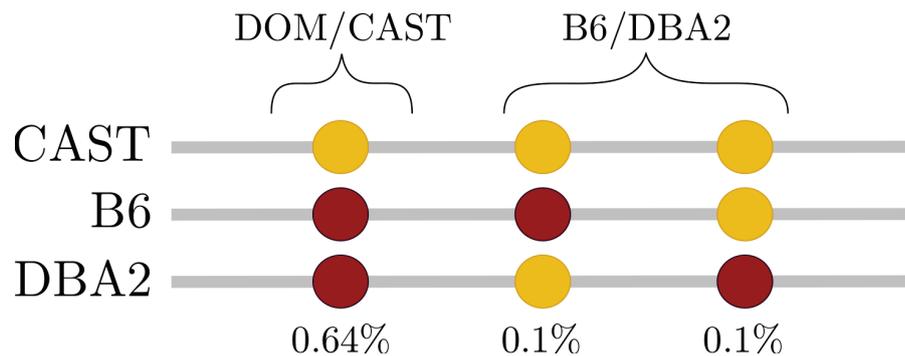
Figure 8.1: The three possible types of polymorphic sites.
According to the principle of parsimony, any polymorphic site (circle) should result, in most cases, in two of the strains carrying the same allele and one of them carrying a different one. In this example, the polymorphic site on the left corresponds to a DOM-CAST marker, where the B6 and the DBA2 haplotypes carry the same allele, different from that of the CAST haplotype. The polymorphic sites in the middle and on the right correspond to two B6-DBA2 markers, with either the B6 (middle) or the DBA2 (right) haplotype carrying the same allele as the CAST one. Given the divergence between strains (see main text), DOM-CAST markers occur more often than the B6-DBA2 markers.

8.2.2 Identification of the genetic background

Next, we inferred the genetic backgrounds using the following criteria: if more than 90% of the DOM-CAST markers of a given hotspot were genotyped as heterozygous in a given individual, a DOM/CAST background was inferred; if more than 90% of the DOM-CAST markers were genotyped as homozygous, a DOM/DOM background was inferred; in any other case, the background was not inferred.

Out of the 4×1390 targeted loci, 145 (2.6%) had a read coverage too low for the target to be genotyped. Aside from those, the aforementioned *modus operandi* allowed us to genotype 97.5% of all the targets presenting sufficient coverage and ended in a mosaic of DOM/DOM and DOM/CAST genetic backgrounds consistent with 0 or 1 (and sometimes 2) crossing-overs per chromosome (Figure 8.2 and Appendix A). This provided strong support that our inference was correct. Among the remaining 2.5% (135) ambiguous targets, 6 (4%) were flanked by DOM/DOM-background targets on one side and by DOM/CAST-background targets on the other side: these most likely corresponded to sites where recombination occurred in one of the parents. All other ambiguous targets (94%) were flanked on both sides by DOM/DOM-background targets: these were most likely erroneously inferred because some B6-DBA2 markers were erroneously classified as DOM-CAST markers.

All in all, across all 1,390 loci of the 4 mice, 7 were incongruent with the surrounding genetic background (either because they were subject to a double crossing-over, or because our inference was incorrect at these sites). We thus chose to remove them from the analysis. Altogether, the proportion of heterozygous DOM/CAST-background targets (Table 8.4) was close to the expected 50% (Table 8.3). To further verify that these observed proportions fitted what was expected, we simulated a DOM/CAST×DOM/DOM cross in which COs (number given by the sex-averaged genetic length) were drawn randomly along each chromosome. We found that the distribution of the expected proportion of heterozygous targets (data not shown) fitted with the observations (Table 8.4).

This genotyping map also allowed to control that all four mice were heterozygous for *Prdm9* since this gene was located in a DOM/CAST background in all samples.
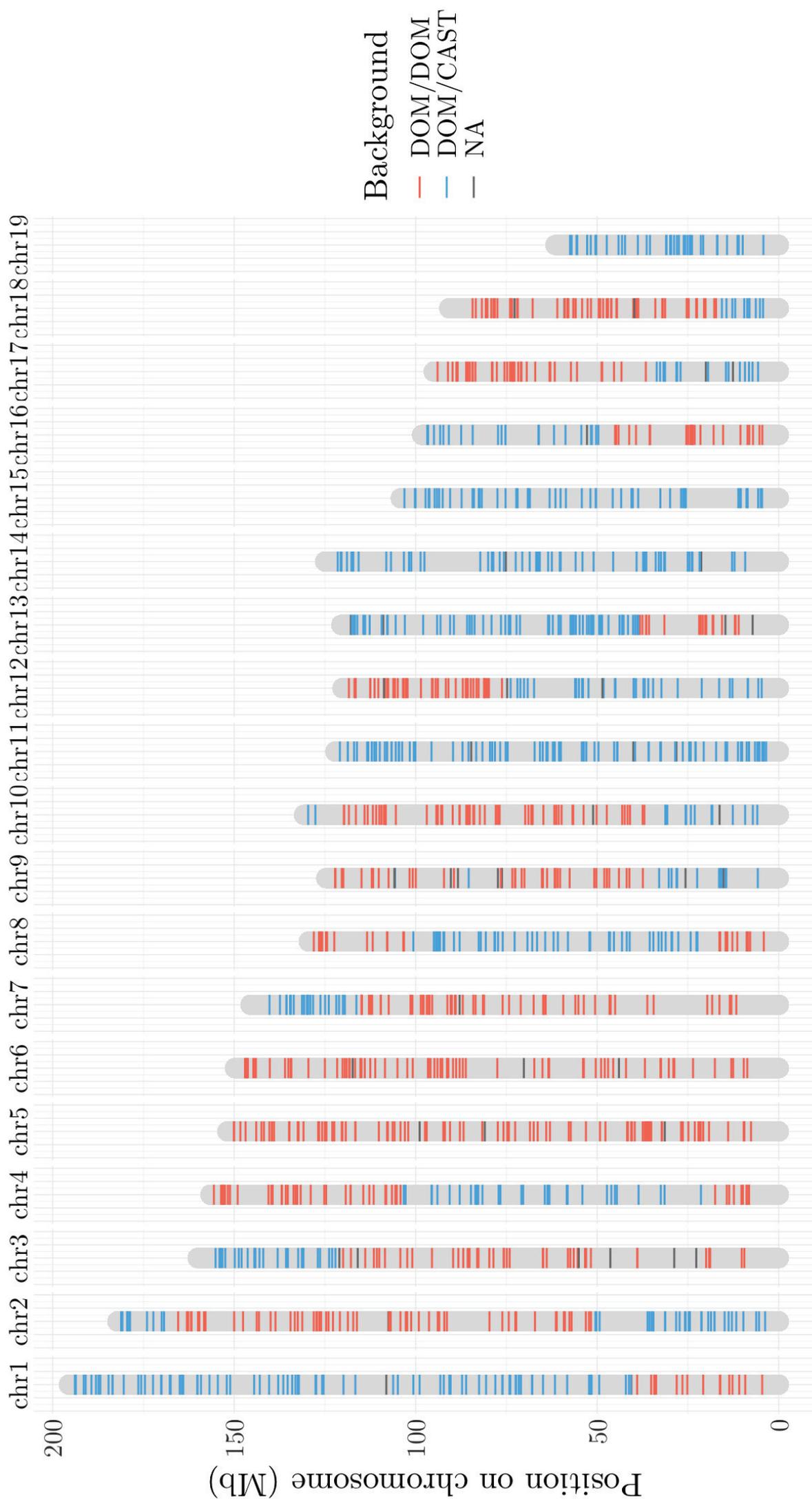
Figure 8.2: Mosaic of genetic backgrounds inferred at each target along the autosomes of mouse 28353. Chromosomes are represented in grey and oriented so that the centromere is on the bottom side of the figure (mouse chromosomes are acrocentric). Each segment corresponds to the position of a target (hotspot or control region) and was coloured in red when the background inferred was DOM/DOM (homozygous) and in blue when the background inferred was DOM/CAST (heterozygous). The corresponding figures for the three other mice (28355, 28367, 28371) are reported in Appendix A.

Category	Sample	# DOM/CAST background	# DOM/DOM background	% of Het. targets
WT	28355	764	561	57.66
	28371	845	461	64.70
Mutant	28353	663	669	49.77
	28367	624	693	47.38
	Total	**2896**	**2384**	**54.85**

Table 8.4: Observed proportion of heterozygous targets in the studied mice.
The background for each hotspot was inferred as described in the main text, for wild-type
(WT) mice (top panel) and mutant mice (bottom panel). The line in bold represents the
average across all four mice.

Target category	Sample	Nb of targets	Nb of fragments	Nb of events	Event rate ($\times 10^{-6}$)
Hotspots	28355	485	28,181,748	1,298	46.1
	28371	552	34,015,365	1,847	54.3
	Tot. WT	**1037**	**62,197,113**	**3,145**	**50.6**
	28353	429	25,598,721	3,486	136
	28367	390	30,863,121	2,082	67.4
	Tot. mutants	**819**	**56,461,842**	**5,568**	**98.6**
Controls	28355	279	15,206,411	34	2.24
	28371	293	16,997,729	58	3.41
	Tot. WT	**572**	**32,204,140**	**92**	**2.86**
	28353	234	13,658,994	33	2.42
	28367	234	17,565,253	25	1.42
	Tot. mutants	**468**	**31,224,247**	**58**	**1.86**
FP rate					**3.22 %**

Table 8.5: Number of events detected in hotspot and control targets.
Events (false positives (FPs) or genuine recombination events) were detected using the
unique-molecule genotyping pipeline described in Chapter 5. All fragments or events
overlapping at least 1 bp with a given target are counted in this table. The event rate
corresponds to the ratio of candidate recombination events over the total number of
fragments. The maximum false positive (FP) rate is the ratio of the event rate in control
targets over that in hotspots. The lines in bold represent the totals for the two WT and
the two mutant mice.

8.2.3   Detection of events in heterozygous hotspots

Finally, for each individual, we applied the unique-molecule genotyping pipeline described in Chapter 5 to all the heterozygous targets and we found that the maximum FP error rate for this re-adaptation of our approach (3.22%, Table 8.5) was similar to that from Chapter 5 (3.73%, Table 5.2).

Altogether thus, our procedure was as efficient to detect recombination events in F2 individuals containing three genetic backgrounds as it was for F1 hybrids. From this point on, we could thus assess the impact of the *hfm1* mutation on several aspects of recombination.

8.3   Impact of the mutation on recombination

8.3.1   Impact on the recombination rate (RR)

We observed that the recombination rate (RR) was, on average, almost twice as high for mutants as for WT mice (Table 8.5). This finding was unexpected since the only effect of the interaction between Mer3 and Mlh1 that was reported in yeasts concerned the length of gene conversion tracts, but not the recombination rate (Duroc et al., 2017).

In our case, this modification of the RR was majoritarily driven by the extremely high recombination rate of mouse 28353 (136 events per million of sequenced fragments, Table 8.5), which was over twice that of the other mutant mouse 28367 (67.4 events per million of sequenced fragments).

Though, if, say, the subset of heterozygous hotspots of mouse 28353 were more intense (i.e. displayed higher recombinational activity on average than other hotspots), this observation would not correspond to a genuine biological effect. In the following subsection, I describe how we thus controlled for such technical biases.

8.3.2   Pairwise comparison of the RR in shared hotspots

To test whether the variation in overall recombination rate (RR) across mice was due to the fact that the sets of hotspots analysed (i.e. heterozygous hotspots) were different between mice, we performed comparisons of the RR in shared hotspots for all pairs of mice (Figure 8.3 and Appendix A). We found that the difference in recombination rates between WT and mutant mice was observed even for shared hotspots, which proved that the effect was not due to a differential sampling of heterozygous loci.

In addition, to see whether the difference in RR applied specifically to one type of recombination products (either COs or NCOs), we reproduced the pairwise comparisons separately for Rec-1S and Rec-2S events (Appendix A). We found that the results were similar for both Rec-1S and Rec-2S events, which showed that both COs and NCOs were affected.

All in all, the RRs for the two WT mice were extremely close (Figure 8.3.a.): the slope of the linear regression was almost 1 (slope $= 1.03$; p-val $< 2 \times 10^{-16}$). However, the recombination rates between the two mutant mice was extremely variable (Table 8.5 and Figure 8.3.b.). What drives such variability among *hfm1* mutants remains, at this stage, unknown: to get more insight into this topic, it would be necessary to analyse the data from additional mutant mice displaying distinct mosaics of genetic backgrounds, and see, for instance, if the increased RR is associated to a given locus in the DOM/DOM or DOM/CAST background.

8.3.3   Impact on CO tract length

Finally, because the interaction between the HFM1 yeast homologue (Mer3) and the MLH1 yeast homologue (Mlh1) has been shown to play a role in DNA heteroduplex extension (Duroc et al., 2017), we wanted to assess whether tract lengths differed between the WT and the *hfm1* mutant mice.

(a) Between the two WT mice

(b) Between the two mutant mice
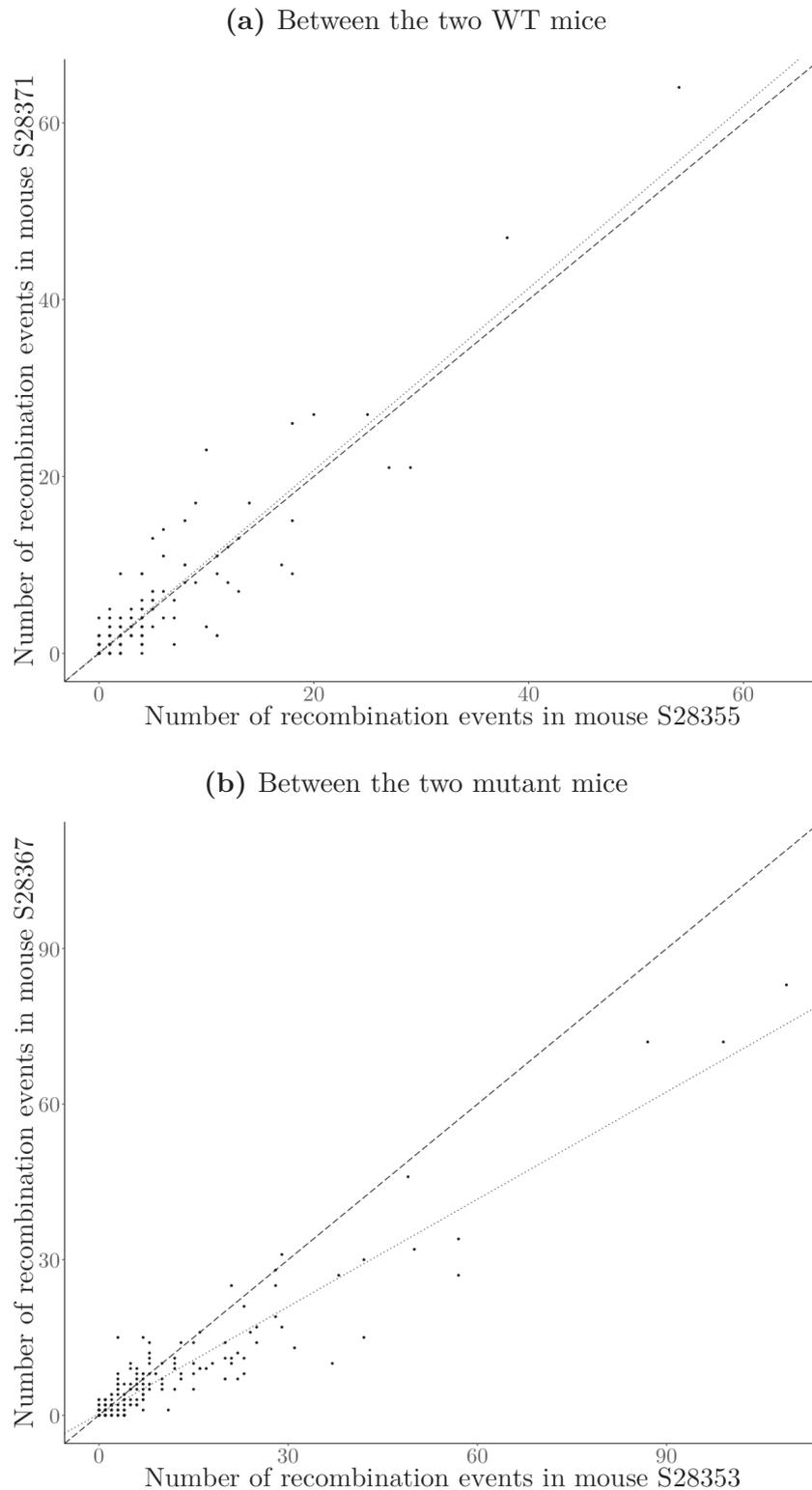
Figure 8.3: Correlation of the number of recombination events in shared hotspots for the two WT (a) and the two mutant (b) mice.
The linear regression was significant for the two WT mice (slope $= 1.03$; p-val $< 2 \times 10^{-16}$; $n_{hotspots} = 257$) and for the two mutant mice (slope $= 0.69$; p-val $< 2 \times 10^{-16}$; $n_{hotspots} = 241$). Figures for all other pairwise correlations are reported in Appendix A. Results of the linear correlation were similar for Rec-1S and Rec-2S events, as well as when controlling for the total number of events sequenced at each hotspot (data not shown).

Because CO and NCO CT lengths are not directly observable from the data, we performed an approximate bayesian computation (ABC) similar to what was described in Chapter 6, based on 50,000 simulations reproducing this experiment (thus, as compared to Chapter 6, we modified hotspot width, fragment start and stop positions and polymorphic sites to fit with this experiment).

Altogether, the CO:NCO ratio and the NCO CT length estimated for WT mice were strikingly close to those measured on the WT mice of Chapter 6 (Table 6.1). CO CTs were slightly longer (albeit *not* significantly) than those found in the previous ABC, which likely comes from the fact that the targeted regions were wider in this experiment whereas the maximum distance between DSB sites and CO switch points was limited in the previous one.

Interestingly, we found a clear CO CT length reduction in *hfm1* mutant mice as compared to WT mice (Table 8.6). Because the observed recombination rates varied greatly between the two mutants (see Subsection 8.3.1), we checked whether this effect was also visible in single individuals and found that, indeed, the inferred conversion tract lengths were stable, no matter the recombination rate. This observation was consistent with the idea that, in mice, the interaction between HFM1 and MLH1 plays a role in extending the DNA heteroduplex. Surprisingly,

Parameter	WT			Mutant		
	Both	28355	28371	Both	28353	28367
CO:NCO ratio	0.108 [0.009–0.189]	0.095	0.098	0.092 [0.0003–0.40]	0.051	0.166
CO CT length						
Mean	744 [219–2790]	539	654	236 [145–478]	238	253
Sd	582 [101–765]	514	759	292 [30–397]	416	232
NCO CT length						
Mean	34 [5–47]	35	31	30 [0.75–397]	49	32
Sd	43 [1–101]	38	57	108 [13–260]	130	66

Table 8.6: Recombination parameters inferred from an approximate bayesian computation for WT and mutant mice.
Parameters (CO:NCO ratio and CO and NCO conversion tract (CT) length reported in bp) were estimated for the two WT and the two mutant mice. 95% confidence intervals were reported between brackets. Because the observed recombination rate varied greatly between the two mutants (Subsection 8.3.1), we also reported than point estimates for all single individuals.

this effect was the oppposite to what had been previously observed in yeasts (Duroc et al., 2017) but the biological reason why the role of the interaction between HFM1 and MLH1 differs between these two species remains to be determined.

In summary, the method we implemented to detect recombination was adaptable to cases where other genomes had been introgressed into the hybrid and allowed to gain new insight into recombination in mice. Though, as any approach, it had inherent limitations, which I will discuss in the following chapter, together with the scientific implications of the whole work done in the context of this thesis.

Part IV

Discussion

*'I suppose the process of acceptance will pass through
the usual four stages:*
(i) this is worthless nonsense;
(ii) this is an interesting, but perverse, point of view;
(iii) this is true, but quite unimportant;
(iv) I always said so.'

— John B. S. Haldane, *The Truth About Death* (1963)

9

Implications for mammalian genome evolution

Contents

The work presented in this thesis allowed to better comprehend recombination and its impact on genome evolution: in brief, we precisely characterised patterns of recombination in mice, brought preliminary answers to the specific role of one gene essential to recombination, and quantified the contribution of all types of recombination products to GC-content evolution *via* biased gene conversion.

The progress we made on this topic principally rested on the analysis of the recombination events we could detect in mouse autosomal hotspots with the approach

we developed. In this chapter, I will first discuss both the significance and the limitations of our method in the context of studies on recombination. Next, I will try and give answers upon the original motivation for this work: figuring out the interplay between the effective population size (N_e) and the gBGC coefficient (b), by comparing our findings in mice to those of others in humans. Last, I will provide more speculative interpretations about this interplay and the evolution of biased gene conversion in general.

9.1 Significance and limitations of our method

9.1.1 Comparison with classical pedigree approaches

The method we implemented to detect recombination events exhibits several advantages as compared to the more classical approach of pedigree analysis.

First, it allows to quantify and precisely characterise recombination events in a single individual whereas the events identified by pedigree analysis span at least several tens — and sometimes a few hundred — members of a given family.

Second, because we specifically targeted recombination hotspots, we only needed to sequence ∼244 Gb of DNA (as compared to the sequencing of between 500 and 50,000 Gb in comparable studies) and identified several thousands of events, whereas pedigree approaches cap at several hundreds (Halldorsson et al., 2016; Smeds et al., 2016) or at a few thousands of events at best (Li et al., 2018). Thus, our method was much more powerful than pedigree analyses: the number of events detected per Gb sequenced with our method (77.1 events/Gb) was over 100 times as great as that of a recent study carried on mice by Li et al. (2018) (0.604 events/Gb). Despite the fact that the recombination rate is respectively twice and six times as high in humans and in flycatchers as in mice (Kawakami et al., 2014, 2017), the power in detecting events in these two species *via* pedigree analyses (0.00970 events/Gb and 1.18 events/Gb, respectively) was also largely lower than *via* our method (Halldorsson et al., 2016; Smeds et al., 2016).

In addition, even if our approach was originally designed to study F1 hybrids, we showed that it could be extended to more complex designs to study recombination (see Chapter 8). Indeed, we managed to deal with the incorporation of a third genome and, theoretically, this adaptation should be possible with any other number of genomic introgressions.

As such, our approach indisputably outperforms classical pedigree analyses in detecting recombination. This paves the way to study the individual role of genes that are essential to recombination, as we did in Chapter 8. Notwithstandingly, our method also encompasses a number of limitations, which will be discussed in the next two subsections.

9.1.2   A *prior* knowledge of recombination hotspots in males

To detect recombination events, one obvious prerequisite is the presence of polymorphic sites. We thus selected hotspots that displayed a minimum of 4 markers in the 300-bp central region. Yet, as many SNPs in F1 hybrid hotspots result from hotspot erosion in one parental lineage (Smagulova et al., 2016), this minimum-number-of-SNPs requirement led to a slightly greater proportion of asymmetric (i.e. eroded in one lineage) hotspots than would be expected with a random selection (Table 7.1).

As Li et al. (2018) pointed out, such asymmetric hotspots display, on average, lower recombinational activities than expected on the basis of PRDM9 ChIP-seq binding (Figure 6.2). Thus, the overall recombination rate we extrapolated from our data is likely to be slightly underestimated (see Chapter 6).

What's more, by definition, hotspot asymmetry implies a haplotype bias for PRDM9 binding. As such, our enrichment of targets in asymmetric hotspots likely amplified dBGC, and thus, the variations in dBGC intensity we observed may be somewhat more extreme than what would be expected on average hotspots.

Therefore, because of the hotspot selection step, the recombination events directly observable with our approach are those occurring in the close vicinity of highly polymorphic hotspots and, if their characteristics differ from those of the other, non-observable events, they may not be representative of the totality of recombination events.

In addition, our approach necessitates a large quantity of gametes to be analysed. As such, it is much better suited to the study of recombination in males and, thus, does not permit to give insight into the process of recombination in the other sex.

9.1.3   The issue of NCO detectability

To minimise the rate of false positive calls, we filtered out all fragments that did not include a minimum of two *B6*- and two *CAST*-typed variants. This implies that, for a NCO to be detected, its conversion tract (CT) must be long enough to overlap at least two variants. Since NCO CTs are only a few base pairs to a few tens of base pairs long (Cole et al., 2014), one would *a priori* expect a non-negligible proportion of them to be intrinsically undetectable, especially in regions with low marker density.

In particular, single-marker NCO (NCO-1) events cannot be detected directly with our approach. As for multiple-marker NCO (NCO-2+) events, their level of detectability depends on marker density, which can vary across — but also along — hotspots. To make this along-hotspot fluctuation visible, we added that information (the maximum number of Rec-1S and Rec-2S switch points detectable) for each existing marker-marker interval (Figure 6.6).

Given that many events are undetectable, we used an approximate bayesian computation (ABC) approach to estimate the genuine values of certain recombination parameters: the lengths of CO and NCO conversion tracts (CTs) and the CO:NCO ratio. Since the estimates that we obtained with the ABC were extremely close to the direct observations of CO and NCO CT lengths in a few mouse hotspots

and to the CO:NCO ratio predicted on the basis of cytological estimates of DSBs (see Chapter 6), we are confident that this approach was globally valid.

Still, it should be noted that the validity of the ABC rests on the assumptions that were made to simulate recombination events. Notably, we hypothesised that the CO:NCO ratio was identical for all hotspots, that both CO and NCO CTs were centred on the DSB and that their CT lengths were arranged according to a unimodal distribution, i.e. that the resolution process was *not* perceptibly different for any subclass of COs or NCOs. But, since the process of recombination has not been completely elucidated yet, we cannot know whether the latter assumptions were biologically accurate nor whether other hypotheses could be more relevant to simulate these events. Nevertheless, based on these assumptions, the ABC allowed us to assess that NCOs were approximately 3 times less detectable than COs.

As such, even if our sequencing fragments were relatively short (2×250 bp) and if NCOs were less detectable than COs, our method allowed to detect an unprecendentedly large number of both types of events, and the ABC permitted to extrapolate the genuine average recombination parameters.

Altogether thus, our approach provides exceptional power to detect recombination at high resolution and at low cost in single individuals. However, it it is only applicable to males and it requires a *prior* knowledge of the position of recombination hotspots. In addition, because the hotspots selected need to encompass multiple polymorphic sites, they may not be representative of the average hotspots, but even this high rate of heterozygous sites remains insufficient to totally erase NCO detectability issues.

Despite these few limitations, our approach was well suited to measure biased gene conversion in mice. To better understand how it evolved in mammals, we next compared our results with those found in another mammalian species: humans.

9.2   Evolution of gBGC in mammals

9.2.1   Measure of the population-scaled gBGC coefficient

Using the approach previously described by Glémin et al. (2015), Brice Letcher, an intern in our lab, measured the population-scaled gBGC coefficient (B) in humans and in the two subspecies from which the parents of the hybrid mice we studied originated. He found that B is 1.5 to 3.5 times lower in humans than in mice (Table 9.1).

Interestingly, the effective population size (N_e) is respectively 20- and 70-fold as high in *Mus musculus domesticus* and in *Mus musculus castaneus* as in humans (Charlesworth, 2009; Phifer-Rixey et al., 2012). Since $B = 4 \times N_e \times b$ (see Chapter 4), this implies that b is 6 to 10 times as high in humans as in the two mouse subspecies (Table 9.1).

Different factors may contribute to this b discrepancy between humans and mice. Indeed, the gBGC coefficient (b) can be decomposed as:

$$b = \qquad b^{CO} \qquad + \qquad b^{NCO}$$
$$= (b_0^{CO} \times r^{CO} \times L^{CO}) + (b_0^{NCO} \times r^{NCO} \times L^{NCO})$$

where r^i, L^i and b_0^i respectively represent the rate, conversion tract length and transmission bias on recombination events i (i corresponding either to CO or NCO events).

In the remaining portion of the discussion, I will go through all the parameters on which b depends to try and identify which contribute more to the difference in b between humans and mice.

	$B \: [CI_{95\%}]$	N_e [min.-max.]	Predicted b
Homo sapiens	0.355 [0.282–0.445]	15,000 [10,000–20,000][1]	5.9×10^{-6}
M. m. domesticus	0.465 [0.337–0.603]	129,000 [58,000–200,000][2]	0.90×10^{-6}
M. m. castaneus	1.21 [1.13–1.26]	466,500 [200,000–733,000][2]	0.65×10^{-6}

Table 9.1: Prediction of the gBGC coefficient (b) on the basis of the population-scaled gBGC coefficient (B) and the effective population size (N_e). A point estimate for the gBGC coefficient (b) was predicted based on the measurement of the population-scaled gBGC coefficient (B) using the approach described by Glémin et al. (2015) and a point estimate (arbitrarily chosen as the mid value between the minimum and the maximum reported values) for the effective population size (N_e). The sources providing the values reported in this table for N_e are given with the following numbered superscript brackets. [1]: Charlesworth (2009). [2]: Phifer-Rixey et al. (2012).

9.2.2 Variation in recombination rate and tract length

In this subsection, we will examine the contribution of the recombination parameters (r^{CO}, r^{NCO}, L^{CO} and L^{NCO}) by considering the $r \times L$ parameter for COs and NCOs separately.

On the one hand, the $r^{CO} \times L^{CO}$ parameter is twice as small in male mice as in men (Table 9.2). This directly comes from the 2-fold difference in r^{CO} between these two species, since the L^{CO} we estimated in mice in this study (447 bp) was almost identical to that measured by others in human sperm (460 bp). Therefore, the CO rate (r^{CO}) contributes to decreasing b by 2-fold in mice as compared to humans.

On the other hand, the $r^{NCO} \times L^{NCO}$ parameter is three times as small in male mice as in men (Table 9.2). Since we found mouse L^{NCO} to be at least 1.5 times as small as human L^{NCO}, the 3-fold difference on the $r^{NCO} \times L^{NCO}$ parameter is compatible with a 2-fold difference on the NCO rate (r^{NCO}) between male mice and humans. We note that, since the mouse CO rate (r^{CO}) too is twice as small as the human CO rate, the 2-fold difference on the NCO rate is compatible with a human CO:NCO rate of 0.10, i.e. close to the known mouse CO:NCO rate (Cole et al., 2010a).

Table 9.2: Estimation of biased gene conversion parameters in *Homo sapiens* and *Mus musculus*.

Parameter (unit)	*Mus musculus*	*Homo sapiens*		
	Male	*Male*	*Female*	*Sex-averaged*
r^{CO} (cM/Mb)	0.42[3,4]	0.812[2]	1.398[2]	1.13[1,2]
L^{CO} (bp)	447⋆ [245–874][b]	460[5] [300–1,019][a]	-	-
b_0^{CO}	1.88 [1.03–3.67][b]	3.73 [2.44–8.27][a]	-	-
$r^{CO} \times L^{CO}$ ($\times 10^{-6}$)	0.00⋆ [-0.0407–-0.00428][b]	0.5[6]	0.404[6] [0.262–0.572][b]	0.402[6] [0.262–0.576][b]
$r^{CO} \times L^{CO} \times b_0^{CO}$ ($\times 10^{-6}$)	**0.00 [-0.0529–0.0198]**[c]	**1.87 [1.22–4.135]**[b]	**2.60 [1.10–8.15]**[c]	**2.09 [0.882–6.63]**[c]
r^{NCO} (cM/Mb)	3.78‡	-	-	-
L^{NCO} (bp)	36⋆ [4–54][b]	55–290[5,§]	-	-
b_0^{NCO}	1.36 [0.151–2.04][b]	3.9[6] [3.5–4.4][b]	10.0[6] [8.5–11.6][b]	7.0[6] [6.0–8.0][b]
$r^{NCO} \times L^{NCO}$ ($\times 10^{-6}$)	0.310⋆	0.262[7] [0.142–0.382][b]	0.450[7] [0.382–0.522][b]	0.36[7] [0.16–0.56][b]
$r^{NCO} \times L^{NCO} \times b_0^{NCO}$ ($\times 10^{-6}$)	**0.422 [0.0469–0.633]**[b]	**1.02 [0.497–1.68]**[c]	**4.50 [3.25–6.06]**[c]	**2.52 [0.96–4.48]**[c]
b ($\times 10^{-6}$)	**0.422 [0.0469–0.633]**[b]	**2.89 [1.72–5.81]**[c]	**7.10 [4.35–14.21]**[c]	**4.61 [1.84–11.1]**[c]

The sources providing the values reported in this table are given with the following numbered superscript brackets. [1]: Dumont and Payseur (2008). [2]: Kong et al. (2002). [3]: Shifman et al. (2006). [4]: Paigen et al. (2008). [5]: Jeffreys and May (2004). [6]: Halldorsson et al. (2016). [7]: Williams et al. (2015). ⋆: Measured or estimated in this study. ‡: We assumed that the mouse r^{NCO} was about 9 times the mouse r^{CO}, since the CO:NCO ratio is 1:10 in the mouse (Handel and Schimenti, 2010). §: Human L^{NCO} values provided by Jeffreys and May (2004) correspond to the mean CT lengths of the two most extreme simulated distributions compatible with observed NCO events. Any value reported without any source was directly calculated by us on the basis of the other parameters in this table. Between brackets, we report uncertainty intervals on these values. As their types may differ according to the sources, we specify them explicitly with alphabetical characters: Minimal and maximal values (a); 95% confidence interval (b); 90% confidence interval (c).

9.2.3 Confidence in the estimation of b_0^{CO} and b_0^{NCO}

Next, we wanted to examine whether the last parameters on which b depends (the transmission bias b_0 of COs and NCOs) also changed in the same direction. But prior to assessing the contribution of the latter to the intensity of gBGC, it is important to authenticate the validity of our estimates for b_0^{CO} and b_0^{NCO}. This will be the object of this subsection, while the extent of their contributions to decreasing b in mice will be discussed in the last section of this chapter.

The estimates for b_0^{CO} and b_0^{NCO} were based on the direct observation of b_0 for Rec-1S, Rec-2S and NCO-1 events (see Chapter 7). The latter depend largely on the correctness in the identification of the donor in the gene conversion event. Indeed, if the inferred donor were *not* accurate, results for the defective fragment would be reversed: all polymorphic sites within the CT* would be designated as being outside CTs*, and conversely.

Regarding Rec-2S events (NCO-2+ and NCO-1 events), since both edges of its CTs* were directly observable, there should not have been any mistake on their orientation.

As for Rec-1S events, we reduced their genuine CT to the segment (CT*) located between the switch point and the PRDM9 ChIP-seq peak summit (see Chapter 6). But, if the DSB site were located outside this CT* (for example, in the portion of the CT on the opposite side of the unambiguous CT* edge), donor inference would be erroneous. It was previously shown that the position of the DSB may vary by up to 30 bp from the consensus motif (Lange et al., 2016) and we thus performed simulations in which the genuine position of the DSB was 30 bp away from its inference (the PRDM9 ChIP-seq peak summit). Using biologically realistic values for all other parameters, we found that the inferred donor was incorrect in fewer than 1% of all recombination events identified under that scenario (data not shown). Therefore, the procedure we used to infer the donor in the recombination event was robust to the inferred position of the DSB.

Whatsoever, even under a worst-case scenario where the donor would be erroneously inferred for most Rec-1S, this would not change results for Rec-2S events and, since NCOs are, by far, the main contributors to b (see below and in Table 9.2), our main conclusions regarding the quantification of gBGC would not change drastically.

9.3   Speculations on the evolution of BGC

9.3.1   Role of CO and NCO events in limiting B

Since the transmission bias on NCOs (b_0^{NCO}) is similar for humans and mice (Table 9.2), this parameter does not participate in the disparity regarding b between the two species.

However, the transmission bias on COs (b_0^{CO}) could explain the remaining difference on b. Indeed, we found in this study that, in mice, the transmission bias of COs is null (Table 9.2). In contrast, in humans, Halldorsson et al. (2016) observed that the transmission bias equals 0.5. It should be noted, however, that Halldorsson et al. (2016) measured b_0^{CO} only for COs displaying complex conversion tracts, which represent only about 0.31% and 1.33% of male and female COs, respectively (Webb et al., 2008; Halldorsson et al., 2016). As the repair mechanism which leads to the formation of these complex COs might be different from that leading to the formation of those with simple conversion tracts, whether or not simple COs display the same transmission bias remains unknown.

As such, aside from the recombination rate and conversion tract lengths, the factors explaining the b difference between mice and humans are still unclear.

9.3.2 A selective pressure restraining gBGC?

All in all, both qualitative and quantitative differences exist between humans and mice for males, and likely between men and women too (but data is lacking to verify this in mice). This suggests that the DSB repair machinery leading to gBGC proceeds differently in these two species and, thus, that this machinery evolved rapidly within the mammalian clade.

As gBGC is known to promote the fixation of G and C alleles even when they are deleterious (Galtier et al., 2009; Necşulea et al., 2011), the burden of this force at the population-scale should be higher in species with large N_e. Nonetheless, B remains in a small range, irrespective of the effective population size (N_e). It is thus tempting to suggest that there may be a selective pressure on the DSB repair machinery to minimise b in species with large N_e, as has already been proposed by Galtier et al. (2018).

Given our observations, it seems that several parameters would allow to restrain B in species — like mice — where the effective population size is high. Indeed, both the recombination rate and the lengths of NCO CTs are smaller in mice than in humans and thus participate in lessening b.

In addition, since $b_0^{NCO_1}$ is much greater than $b_0^{NCO_{2+}}$ in mice, the relative proportion of NCO-1 and NCO-2+ events — which depends on the level of polymorphism — has an impact on b (see Chapter 6): the more polymorphic, the greater proportion of NCO-2+ events, and thus the lower b.

Interestingly, in mice, the transmission bias on multiple-marker NCOs ($b_0^{NCO_{2+}} = 0.042$) is extremely weak as compared to that for single-marker NCOs ($b_0^{NCO_1} = 0.40$) (see Chapter 7). To find out whether this was the case in humans, we reanalysed data from Halldorsson et al. (2016) and found that the transmission bias on multiple-marker NCOs is similar to the transmission bias on single-marker NCOs (Table 9.3). This suggests that the repair mechanism leading to NCOs might differ between humans and mice. However, as, in *Homo sapiens*, most (84%) of the NCO-2+

CT class	All NCO events			
	# CTs	# W→S	# S→W	b_0
1 marker	1818	1192	554	0.37
2 markers	150	187	95	0.33
3 markers	53	106	47	0.39
>4 markers	119	731	357	0.35
All CTs	2140	2216	1053	0.36

Table 9.3: Transmission biases for all human NCOs.
The data used in this table correspond to the NCO events in the ChIP-seq dataset of
Halldorsson et al. (2016). Similar results were obtained for the NCO events coming from
the sequencing dataset of Halldorsson et al. (2016) (data not shown).

events come from women (Table 9.4), the difference between $b_0^{NCO_{2+}}$ and $b_0^{NCO_1}$
may reflect a sex-based rather than an interspecific discrepancy. But it is presently
impossible to discriminate between these two possible explanations since no data
is yet available in female mice.

More generally, since species with large N_e are more polymorphic and thus
entail more NCO-2+ events, the fact that $b_0^{NCO_{2+}}$ is much smaller than $b_0^{NCO_1}$ may
be interpreted as another manifestation of the existence of a selective pressure
acting to restrain B in large-N_e populations.

9.3.3 dBGC hitchhiking in structured populations

Finally, the other type of biased gene conversion — dBGC — also seems to play a
significant role in genome evolution, particularly in experimental designs such as
ours, and this should also be discussed. Indeed, the hybrid mice that we analysed
descended from crosses between two strains derived from subspecies which displayed
distinct *Prdm9* alleles. Thus, their respective hotspots specifically underwent gBGC
and got GC-enriched as compared to the genome of the other ('nonself') strain:
in the B6 (resp. CAST) lineage, PRDM9^{Dom2}-targeted (resp. PRDM9^{Cst}-targeted)
hotspots locally enriched in GC while these positions in the CAST (resp. B6)

CT class	Paternal NCO events				Maternal NCO events			
	# CTs	# W→S	# S→W	b_0	# CTs	# W→S	# S→W	b_0
1 marker	824	513	270	0.31	994	679	284	0.41
2 markers	34	34	23	0.19	116	153	72	0.36
3 markers	5	7	7	0.00	48	99	40	0.42
>4 markers	12	51	46	0.05	107	680	311	0.37
All CTs	875	605	346	0.27	1265	1611	707	0.39

Table 9.4: Transmission biases for human paternal and maternal NCOs.
The data used in this table correspond to the NCO events in the ChIP-seq dataset of Halldorsson et al. (2016). Similar results were obtained for the NCO events coming from the sequencing dataset of Halldorsson et al. (2016) (data not shown).

lineage did not. In parallel, the targeted hotspots got eroded in the 'self' lineage, as predicted by the hotspot conversion paradox (Boulton et al., 1997).

Consequently, when the two strains were crossed into a hybrid, each hotspot had been eroded in the locally GC-enriched (self) haplotype. Thus, the DSB initiated preferentially on the other (nonself), non-eroded and GC-poorer haplotype. In turn, this led the eroded, GC-richer (self) haplotype to be the donor during the gene conversion event and its *GC* alleles to be overtransmitted into the pool of gametes.

Such interplay between dBGC (targeting the non-eroded haplotype) and past gBGC (local enrichment in *GC* alleles) can be extended to any more general case of structured population: if two populations with distinct *Prdm9* alleles have evolved independently during a length of time sufficient for the hotspots targeted by each allele to erode specifically in their lineage, crossing them together will end in dBGC hitchhiking past gBGC (Figure 9.1).

This phenomenon of dBGC hitchhiking brought a confounding effect to quantify gBGC and we thus decoupled the two processes by equalising, at every hotspot, the number of B6- and CAST-donor fragments to cancel the dBGC effect. This allowed us to quantify the transmission bias (b_0) in mice, which was useful to comprehend how the gBGC coefficient (b) varies with the effective population size (N_e) and to show
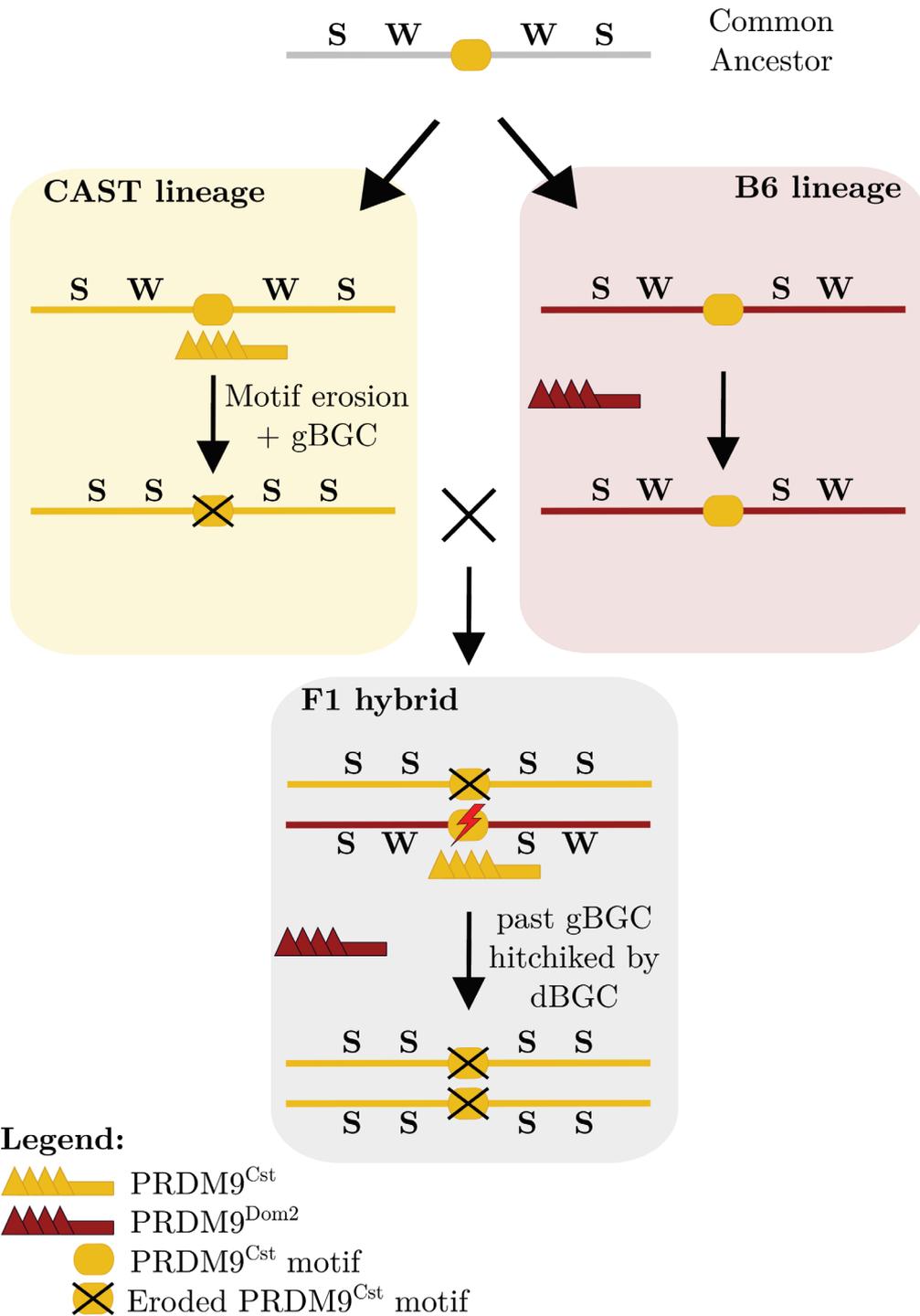
Figure 9.1: dBGC hitchhiking in structured populations.
In the CAST lineage (yellow box) where PRDM9$^{\text{Cst}}$ (yellow triangles) is present,
PRDM9$^{\text{Cst}}$-targeted motifs (yellow square) undergo erosion and, because of gBGC, weak
(W) bases (A or T) get supplanted by strong (S) bases (G or C) at WS polymorphic sites.
In contrast, in the B6 lineage (red box) where only PRDM9$^{\text{Dom2}}$ (red triangles) is present,
PRDM9$^{\text{Cst}}$-targeted motifs do *not* undergo biased gene conversion. As such, the CAST
haplotype (yellow segment) is locally enriched in S bases as compared to the B6 haplotype
(red segment). When populations cross into a hybrid (grey box), the non-eroded motif
from the B6 lineage is targeted by PRDM9$^{\text{Cst}}$ and the DSB initiates on the B6 haplotype
(red thunderbolt). Consequently, the CAST haplotype with both the eroded motif and
the local enrichment in S bases is the donor in the conversion event: past gBGC that
occurred in the CAST lineage is hitchhiked by dBGC occurring in the hybrid.

that the observations complied with the hypothesis of a selective pressure restraining gBGC at the individual-scale to limit its nefast consequences at the population-scale.

But, this hypothesis, — if it were true, — would surely bring other more conceptual questions like the following: how can effects on a population drive the evolution of a molecular mechanism in single individuals? And at what scale, — populational or individual, — should the concept of evolutionary forces be defined? Rather than answering them, I will try and provide food for thought on these open questions in the following — and final — chapter of this thesis.

10

A little bit of scientific philosophising

Contents

'Is there any knowledge in the world which is so certain that no reasonable man could doubt it?

This question, which was the first sentence of the book *The Problems of Philosophy* (1912) by Bertrand Russell (1872–1970), summarises rather well his life's quest: the search for truth — which he believed could be attained with logic. Russell spent all his life working on this topic, both in mathematics and in philosophy, and this made him one of the founding fathers of contemporary logic. We can get a small taste of his logical developments in the paradox he discovered in the domain

of set theory, and which he himself translated into 'ordinary language' (Russell, 1918) under the form of the barber's paradox:

> *'You can define the barber as "one who shaves all those, and those only, who do not shave themselves". The question is, does the barber shave himself?'*

Answering this question results in a contradiction: if he shaves himself, he cannot shave himself (because the barber shaves only those who do not shave themselves); and if he does not shave himself, he must shave himself (because the barber shaves all those who do not shave themselves). This is a typical 'logical paradox'.

For Russell, the solution to such contradictory phenomena is to break down each proposition (scientific or philosophic) into ultimate logical units (or atoms) which can be understood independently of other units: this is what he called 'logical atomism'. In his view, to know whether a proposition is true or false comes back to analysing the veracity of each atom and the relationship between them[1]. To further decide on the veracity of a given simple proposition, — which he redefines as the adequacy between a belief and a fact, — one must agree to hierarchise the degree of certainty of each 'known' fact. For instance, one can be absolutely certain of the things they directly experimented with their five senses ('sense-data'), — he calls that 'knowledge by acquaintance', — but the confidence one has in 'knowledge by induction' (i.e. the process of deriving a theory from the repeated observation of events) must be questioned. To borrow one of his own illustrations of that matter (Russell, 1912), we do not feel the slightest doubt that the sun will rise tomorrow because of the laws of motion. *'But the only reason for believing that the laws of motion will remain in operation is that they have operated hitherto, so far as our knowledge of the past enables us to judge. [...] But the real question is: do any number of cases of a law being fulfilled in the past afford evidence that it will be fulfilled in the future?'* It is, of course, highly unlikely that the laws of motion would stop tomorrow and that the sun would not rise; though, we cannot *prove* it is impossible and, thus, the degree of confidence we can have in such knowledge is

---

[1]This is, by the way, what led him to redemonstrate every simplistic principle of algebra (like the fact that $1 + 1 = 2$) in his *Principia mathematica* (1912).

lower than that for the things we are directly acquainted with, like the fact that the paper on which this text is written is white.

In the case of an evolutionary force, we are in a typical case of such knowledge learnt by induction: we infer its very existence on the basis of the observation of its consequences on genomes and, on top of that, it is not even tangible, but merely a concept useful to theorise how genomes evolve. To analyse such ideas, Russell systematically started with redefining precisely the terms. But, what is an evolutionary force, exactly? Is it even a cause (for genome evolution) or a consequence (of the molecular processes taking place in individuals)? And, consequently, at what scale — individual or populational — should it be studied? In the first section of this chapter, I will try and provide ideas to answer those questions. In the second section, I will dive into the more general notion of the way scientific knowledge can be obtained and the context in which it arises and last, I will focus on the particular and more recent role of bioinformatics in acquiring such knowledge.

10.1   About evolutionary forces

10.1.1   Forces as conceptual frameworks

By definition, a force represents an interaction which, if unopposed, can change the motion of an object. As such, forces are generally viewed as causes driving objects or phenomena in a certain direction and are commonly symbolised as vectors giving their direction and intensity. But are forces mere conceptual tools useful to better apprehend physical phenomena, or could they exist as real physical entities?

Gravitation, which ensures the mechanical movement of planets and other celestial bodies, is a most interesting case study to think of the aforementioned interrogation. Indeed, for over 200 years, the theory formulated by Isaac Newton (1642–1727) — the law of universal attraction stating that a 'gravitational force' leads masses to attract one another — had been widely accepted. But, in the early 1900's, Albert Einstein (1879–1955) established the theory of general relativity which

accounted for the physical effects unexplained by Newton's law and contradicted the idea that the gravitational force was even a force at all: instead, gravitational attraction would be the result of the warping of spacetime by large masses. Since then, gravitation has stopped being considered as a force, but its pictural representation as vectors has nonetheless persited, for it helps conceptualising the physical phenomena it explains.

10.1.2   Forces as emerging properties of individuals

Another way to regard forces consists in perceiving them as emerging properties of the individuals (particles, people, cells, etc...) which constitute them, i.e. as phenomena resulting from the intrinsic characteristics of their components, but *not* reductible to the latter. In other words, a force would be the consequence of the fundamental properties of its components, but somehow more than the mere sum of its parts.

To borrow once again an example taken from physics, pressure corresponds to the mean action of the collision of gas particles on a given area and, thus, arises from the intrinsic properties of its components. Though, each of these particles moves completely randomly ('Brownian motion') and does not cease bumping into other molecules or into the surfaces of the walls. As such, pressure *cannot* be seen in any particle by itself (for its trajectory is random and the force it exerts on an area is unpredictable) but it nonetheless *emerges* from the collective action of many.

In a totally different context, what is called peer pressure results from the individual choices of single people and can thus be seen as a consequence of the biological processes occurring inside their brains. When looked at it at the scale of a population though, this phenomenon becomes the root cause of the behaviour, attitude or values of other people to conform to the influencing group. As such, peer pressure — and the same would apply to other sociological phenomena, like consumer behaviour — can be seen both as a cause or as a consequence, depending

on the point of view.

Altogether thus, even if forces are most generally used as concepts useful to understand phenomena (whether physical, biological, sociological, or else), they are the result of more fundamental properties emerging from their individual components. With this in mind, at what scale, — populational or individual, — would it be most meaningful to study them in the context of evolutionary biology?

10.1.3 Processes *versus* patterns

In the 1930's and 1940's, the modern synthesis (a.k.a. neo-Darwinian synthesis), — which was formally defined by Dobzhansky (1937), Huxley (1942), Mayr (1942) and Simpson (1944) — reconciled Darwin's theory of evolution and Mendel's ideas on heredity (see Chapter 1).

Since then, the way of considering the objects of study in evolution and their relationships has considerably changed (reviewed in Paulin, 2015): a bipolarisation between *patterns* (i.e. the description of the results of evolution, as independently as possible from any explanatory theory) and *processes* (i.e. the mechanisms responsible for evolution) emerged. What is less well known is that this distinction was defensibly already present in Darwin's theory (Gayon and Petit, 2018) as the name he gave it — 'descent with modification by means of natural selection' — suggests: the 'descent with modification' part would correspond to the *patterns* of evolution and the 'by means of natural selection' part to the *processes* leading to it.

This distinction could arguably be applied to the study of evolutionary forces as well. In the case of the object of this thesis, — biased gene conversion (BGC), — the *process* would correspond to the functional study of the way the molecular machinery responsible for the repair of DNA mismatches results in BGC, and the *pattern* to describing its deleterious consequences on genomes and the extent to which it induces divergence between them. As such, the joint study of both aspects seems essential to describe this evolutionary force as a whole.

Though, the distinction between *patterns* and *processes* may be too simplistic, and it has been much criticised by Stephen Jay Gould (1946–2002) and Niles Eldredge (born 1943) from the 1970's on (reviewed in de Ricqlès and Padian, 2009). Their major objection concerned gradualism (i.e. the idea that all evolutionary changes are slow, gradual and cumulative) because this implied that there would be a nearly total determinism of micro-evolution (*processes*) onto macro-evolution (*patterns*) and that almost everything could be explained by the sole action of natural selection and adaptation (reviewed in Paulin, 2015). Instead, Gould put into perspective the extent to which such deterministic features contributed to macro-evolution by reintroducing historical contingency, i.e. the idea that the history of life also depends on a series of historical events that are often random or, at least, unpredictable (Gould, 1989).

As such, even though his view is still debated, Gould managed to question parts of a theory which was already widely accepted by the scientific community. The way through which such novel ideas can spread into the scientific world participates much in the progress of science and represents one of the main questions tackled by epistemologists. As such, I will focus on this issue in the following section.

10.2 About scientific advances

10.2.1 Scientific revolutions and paradigm shifts

To face gradualism in the modern synthesis of evolution, Gould and Eldredge put forward another thesis: the theory of punctuated equilibria, according to which periods of rapid change are followed by longer periods of relative stasis, i.e. states of little change (Gould and Eldredge, 1972).

We could draw a parallel between this new theory about evolution and that by Thomas Samuel Kuhn (1922–1996) about scientific progress. Indeed, when it began in the eighteenth century, history of science was written by scientists who presented

the discoveries of their time as the culmination of a long process of advancing knowledge. Thus, science was perceived as a progressive accumulation of cognition where true theories replaced false beliefs (Golinski, 2008).

In contrast, Kuhn portrayed scientific progress as a cyclic process involving paradigm shifts, i.e. fundamental changes in the basic principles of a scientific discipline (Kuhn, 1962). In his view, periods of 'normal science' where scientists work under a conceptual framework which works globally well alternate with shorter periods of 'revolutionary science' where the repeated detection of anomalies (i.e. observations unreconciliable with the paradigm of the time) leads to another paradigm under which the world that scientists perceive, as well as the principles, methods or even language they use, are different.

According to Kuhn, the transition from one paradigm to another does not rest solely on rational scientific reasons justifying that the new paradigm would be more accurate: he firmly believes that these major shifts also largely depend on external factors, like the sociological and ideological context of the time. I give examples of these in the following subsection.

10.2.2 The impact of external factors

Paul Forman (born 1937), a former student of Kuhn's, defended the thesis of a cultural conditioning of scientific knowledge. He developed his proposition with the example of the connection between the culture of Weimar Germany and the emergence of quantum mechanics in the 1920's (Forman, 1971). According to him, in the aftermath of the defeat of Germany in World War I, the dominant tendancy was characterised by intellectual revolts against causality, determinism and materialism and welcomed the rise of anti-rationalist movements such as existentialism, i.e. a philosophy of life claiming that individuals are faced with the absurdity of life and that the essence of their being lies in their own actions which are *not* predetermined by any kind of theological, philosophical or moral doctrine.

In Forman's view, the concept of quantum *a*causality could spread much more easily into this German scientific world marked by the rejection of determinism and analytical rationality than in other Western countries which did not undergo such crises, and explains why the most prominent advances in that field were made by Germans.

On top of the sociological, political and religious context, Barry Barnes (born 1943) argues that the personal interests of researchers also play a major role in determining their actions and, thus, in shaping scientific advances (Barnes, 1977). Interests at stake in scientific practice may include the use of techniques or theories specific to a given paradigm which they want to promote, or defined by their social, political or ideological position (Gingras, 2017). As such, 'inner' and 'outer' factors are not necessarily distinct.

For instance, in nowadays world where ecological awareness is growing, several scientists promote the creation of a new geological epoch — the so-called 'Anthropocene' — that would account for the impact of mankind on Earth's geology and ecosystems (Crutzen, 2002) and some geologists and mineralogists have already started doing research in this still unofficial field of investigation (Corcoran et al., 2014; Hazen et al., 2017).

In the case of Gould and Eldredge too, their challenging the modern synthesis was made possible thanks to the contemporary creation of additional fields of investigation — including developmental genetics, phylogenetic cladistics, the molecular clock and gene transfers: these provided novel findings or original ways of thinking, which participated a great deal in questioning parts of the modern synthesis (Lecointre, 2009).

Generally, the creation of new domains of study pairs up with the establishment of modern techniques which themselves play a significant role in advancing knowledge. I discuss this topic in the next subsection.

10.2.3 The contribution of modern techniques

It goes without saying that scientific knowledge has systematically considerably benefited from both technological advances and the expertise of scientists in using the latter. Cell biology, for one, would not have existed had microscopy not been invented (Bechtel, 2006) and chromosomes would not have been discovered if it had not been for Frans Janssens's mastery of cell staining (see Chapter 1).

Though, the very use of technologies for scientific progress can bring a set of questions of its own. Indeed, it has been argued that there is often a circular relationship between the pieces of evidence for a phenomenon of interest and the instruments detecting it (Collins, 1975, 1985, reviewed in Godin and Gingras, 2002): according to the words of the sociologist who developed this idea, '*we won't know if we have built a good detector until we have tried it and obtained the correct outcome. But we don't know what the correct outcome is until. . . and so on ad infinitum*' (Collins, 1985). He termed this pitfall the 'experimenter's regress'.

On top of that, the belief (or not) in the outcome and the acceptance (or not) of the value given by the instrument somehow depends on the researcher's interests: a scientist who believes in the existence of a phenomenon will be willing to accept the announcement of its detection, while one who does not would probably rather question the validity of either the apparatus or the method used (Gingras, 2017).

In genetics, the development of the first sequencing techniques in the 1970's have led to a major upheaval in the way research is carried. Indeed, the rise of '-omics' (genomics, transcriptomics, metabolomics, proteomics, etc. . .) as major fields of study, together with the large progresses in computing resources and data storage capacity, has led some to re-think of the interplay between data-driven and hypothesis-driven science (Kell and Oliver, 2004; Mazzocchi, 2015).

But, from now on, future advances in the field surely depend much more on the ability of bioinformaticians to analyse the deluge of data standing before them rather than on further technological leaps. In the last section, I thus share my vision on the way I believe bioinformaticians can best help scientific progress.

10.3 About bioinformaticians

10.3.1 Biologists before informaticians

The word 'bioinformatics' is a contraction of 'biology' and 'informatics' and both facets are of course required in this domain. Though, it seems to me that, in view of the colossal quantity of data that genomicians are supposed to deal with, it can be tempting to let the informatics side take over. On top of that, some bioinformaticians perceive results obtained purely by an automated process involving bioinformatic tools with little or no input from the experimenter as objective, and negatively regard as subjective any choice made by the biologist.

I would like to argue against that line of reasoning by taking an example from machine learning — a set of methods which has begun to be used by bioinformaticians in the last few years. Basically, machine learning is a subset of artificial intelligence aiming at 'learning' from data. In the vast majority of cases, these programs 'learn' on the basis of the correlations they find within the training sets they are provided with. Retracing how these associations have been made is actually a rather complex process but, in one study, after creating a classifier allowing to distinguish between dogs and wolves, Ribeiro et al. (2016) wanted to understand the reasons why their artificial-intelligence method was so outstandingly accurate. They analysed the associations made by the program and found out that the main feature used to distinguish between the two animals was the background in the training pictures: wolves were often standing on snow whereas dogs were rather standing on grass. As such, even if the classifier outputted the correct results, it became obvious that it could not be trusted. Nevertheless, such caveats originating from automated processes can easily be avoided by human knowledge.

In the context of this thesis, the method we implemented to detect recombination events from sequencing data rested on identifying and iteratively suppressing sources of error (see Chapters 5 and 8). In the process leading to it, a considerable amount of time was spent visually inspecting the candidate events and hypothesising on the origin of miscalls. Automation was only used in a second phase to assess the impact

of each possible adjustment to the final outcome. This is, by the way, together with the crucial role of negative controls, how we could identify that mapping biases explained most of the false positive miscalls.

As such, I firmly believe that the input from any savvy human can make analyses much more accurate than the sole work of bioinformatic tools.

I would also like to argue in favour of simplicity. Indeed, considering the extremely wide range of bioinformatic tools — but also statistical and mathematical methods — available today, it is often tempting to create sophisticated processes to tackle biological problems that are generally rather complex. Though, it seems to me that, except for some specific issues, aiming at the maximal simplicity carries many advantages, including a better reproducibility of analyses, a more straightforward detection of errors, greater smooth in adapting code or methods to other frameworks and much larger clarity in transmitting the ideas.

10.3.2 Training biologists in genomics

With the ever increasing amount of sequencing data available, one of the major limitations in genomics becomes the ability to process them. I argued in the previous subsection that the input from humans — biologists in the case of bioinformatics — was crucial to analyse the data correctly.

Though, it is not that easy for biologists to get trained in bioinformatics: to the extent of my knowledge, there is no free website that explains the basic know-how of next-generation sequencing data analysis. Therefore, I decided to create one (`https://gnomics.io/`) to account for this lack. In it, I try to provide biologists with a global overview of the major steps that one should follow to perform the most common genomic analyses, indicate the tools allowing to complete each of these and the way to use them concretely and, finally, explain the assumptions on which they are based and the way the outcome they render should be interpreted.

10.3.3   A genomician in evolutionary biology

According to the paleontologist Stephen Jay Gould, evolutionary biology is a kind of science somewhat special in the way that it creates knowledge. Indeed, in most research fields, the best way to know whether a hypothesis is true or false consists in experimentally testing for it and comparing the outcome it predicted to the real one: if they concord, the hypothesis may be true; otherwise, we can be sure that it is false. Though, this so-called scientific method is not adapted to the study of evolution because the objects of study cannot be reproduced experimentally[2]. Instead, the past is to be *inferred* and, arguably, if there was a past, remnants of it should persist in today's world. The whole work of the evolutionary biologist thus consists in searching for these relics — which, according to Gould, are often imperfections or incongruities — and to make sense of them in a more global picture of evolution (Gould, 1979).

In this context, a genomician working in evolutionary biology should scan genomes to try and find vestiges of the past which could help reconstruct indirectly the unobservable evolutionary history. The discovery of biased gene conversion was typically such a case of evolutionary inference based on unexplained incongruities seen in genomes: it all started with the strange observation that GC-content varies along genomes (see Chapter 4). Several hypotheses were then proposed to explain it — one of which being the existence of biased gene conversion. Since then, a lot of work — including that carried for this thesis, — has been done with the aim of providing evidence for this hypothesis.

Bioinformaticians generally have a training in either informatics, algorithmics, mathematics, statistics or any other field in which certainty is much more widespread than in biology, and especially more than in evolutionary biology. As such, for them to work in this research field, I would argue that one of the major difficulties may reside in fighting an inner struggle to make room for doubt in the middle of all the apparent objectivity of computer programs.

---

[2]Nevertheless, this is precisely what studies of so-called 'experimental evolution' aim to do.

All in all, science is not much different than a quest for truth and scientists generally try and pursue objectivity so as to get to it. Though, in this chapter where I gathered epistemological, philosophical and sociological thoughts, I showed that scientific progress also depends on the contingency of external events and on the subjective interests of researchers, no matter how neutral they are willing to be. In the particular case of bioinformatics applied to evolutionary genomics, I believe that the subjectivity of human expertise can be used as an advantage rather than as an obstacle to make further progress. It was with these thoughts in mind that the work useful to this thesis was carried. As for now, there is nothing left for me but to conclude about it all.

Conclusion

'L'ineptie consiste à vouloir conclure.'

— Gustave Flaubert, *Correspondance* (1889)

In summary, the aim of this thesis was to better understand the interplay between the intensity of GC-biased gene conversion and the effective population size (N_e) within the mammalian clade. We thus wanted to estimate the parameters on which the gBGC coefficient (b) depends — namely the recombination rate r, the length of conversion tracts L and the transmission bias b_0 — in a species with large N_e (mice) to compare them with those found in a species with lower N_e (humans).

To do this, we implemented a method that allowed to detect recombination events at high resolution in the recombination hotspots of single individuals. Our approach appeared unprecedentedly powerful in detecting such events and we showed that it could be adapted to practically any kind of experimental design, no matter the number of genomic introgressions it may involve.

In the course of our enterprise, we managed to quantify double-strand break-induced biased gene conversion (dBGC) in several hundreds of autosomal recombination hotspots and brought to light the fact that, in cases of structured populations, dBGC hitchhiked past gBGC, thus creating an intrincate interplay between the two forms of biased gene conversion occurring in PRDM9-dependent species.

Overall, we found that, in mouse autosomal hotspots, the transmission bias b_0 was similar to that measured in humans for single-marker non-crossover (NCO-1) events but extremely reduced for multiple-marker non-crossover (NCO-2+) events and null for crossing-overs (COs). As, in addition, the recombination rate r and

the length of conversion tracts L were smaller in mice, the gBGC coefficient (b) was globally much reduced in this species.

Altogether, the globally stable intensity of biased gene conversion at the population-scale (B) in *Homo sapiens* ($B = 0.355$), *Mus musculus domesticus* ($B = 0.465$) and *Mus musculus castaneus* ($B = 1.21$) was permitted by the joint decrease of all three parameters on which b depends (r, L and b_0) in the species with 20- to 70-fold larger N_e. We argued that such large differences in b between the two species in spite of their comparable B was consistent with the hypothesis of a selective pressure restraining gBGC at the population-scale and materialising under the form of an extremely rapid evolution of the molecular machinery leading to it.

If our hypothesis were to be correct, the way the information on the effective population size could be integrated by a selective force to constrain the evolution of the molecular machinery at the scale of single individuals remains a widely open question. I would have probably even ventured into saying that this conundrum might be indecipherable, it if were not for John Maynard Smith's observation that *'It is an occupational risk of biologists to claim, towards the end of their careers, that the problems which they have not solved are insoluble'* (Smith, 1988).

Appendices

A

Supplementary data and figures

A.1 Supplementary data

A.1.1 PRDM9$^{\text{Dom2/Cst}}$-targeted hotspots studied

The table below gives the list of mouse hotspots targeted by either PRDM9$^{\text{Dom2}}$ or PRDM9$^{\text{Cst}}$ that have been individually studied.

Name	Target allele	Chr.	Reference
A3	PRDM9$^{\text{Dom2}}$	1	Kelmenson et al. (2005); Cole et al. (2010a)
G7c	PRDM9$^{\text{Dom2}}$	17	Snoek et al. (1998)
E$_\beta$	PRDM9$^{\text{Dom2}}$	17	Steinmetz et al. (1982)
Esrrg1	PRDM9$^{\text{Cst}}$	1	Billings et al. (2013)
Hlx1	PRDM9$^{\text{Cst}}$	1	Ng et al. (2008); Billings et al. (2013)
HS9	PRDM9$^{\text{Dom2}}$	19	Bois (2007); Getun et al. (2010)
HS22	PRDM9$^{\text{Dom2}}$	19	Getun et al. (2010)
HS59.4	PRDM9$^{\text{Dom2}}$	19	Getun et al. (2010)
HS61.1	PRDM9$^{\text{Dom2}}$	19	Wu et al. (2010); Getun et al. (2010)
Pbx1	PRDM9$^{\text{Dom2}}$	1	Billings et al. (2013); Baker et al. (2015b)
Psmb9	PRDM9$^{\text{Cst}}$	17	Guillon and de Massy (2002); Baudat and de Massy (2007)

Table A.1: List of PRDM9$^{\text{Dom2}}$- and PRDM9$^{\text{Cst}}$-targeted hotspots individually studied.

A.1.2   Disclaimer for the resources used

This work was performed using the computing facilities of the CC LBBE/PRABI.

A.1.3   Erroneously called W → S and S → W events

Quantifying gBGC comes back to measuring the $\frac{WS}{WS+SW}$ ratio. However, since the large majority of pot-NCO-1 events corresponded to FPs, we had to distinguish the (potential) contribution of FPs to this ratio from that of genuine NCO-1 events. In particular, this ratio may depart from the expected 50% ratio if (1) a non-negligible proportion of FPs arise from sequencing miscalls and (2) W → S and S → W sequencing errors appear at different frequencies.

First, we thus wanted to quantify the proportion of FPs due to sequencing miscalls. To do this, we estimated the sequencing error rate directly in our sequencing data by monitoring the apparition of *de novo* variants: given that the mutation rate ($\sim10^{-8}$/bp) is much lower than the sequencing error rate ($\sim10^{-3}$/bp), we assumed that, outside the polymorphic sites identified by variant-calling, any base call that differed from the nucleotide of the reference genome was a sequencing error and counted them to compute the conditional frequency matrix of sequencing errors[1] (M):

$$M = \begin{bmatrix} \Pr(A \to A \mid A) & \Pr(A \to C \mid A) & \Pr(A \to G \mid A) & \Pr(A \to T \mid A) \\ \Pr(C \to A \mid C) & \Pr(C \to C \mid C) & \Pr(C \to G \mid C) & \Pr(C \to T \mid C) \\ \Pr(G \to A \mid G) & \Pr(G \to C \mid G) & \Pr(G \to G \mid G) & \Pr(G \to T \mid G) \\ \Pr(T \to A \mid T) & \Pr(T \to C \mid T) & \Pr(T \to G \mid T) & \Pr(T \to T \mid T) \end{bmatrix}$$

$\forall(i,j) \in \{A, C, G, T\}^2$, the number of NCO-1 FPs expected due to sequencing errors involving a genuine base i mistakenly called as a j base ($e_{i \to j}$) simply equalled the product of the number of central markers (i.e. markers *not* located

---

[1]Matrix M was computed based on the analysis of one chromosome (chromosome 10) for all of our 18 samples individually (because the sequencing errors may vary between the biological samples and sequencing runs). This matrix gives the probability of each erroneous base call, given the genuine nucleotide.

at the extremity of fragments) that were genuinely i in ij polymorphic sites (g_i^{ij}) by the conditional probability that a genuine i would mistakenly be called a j ($\Pr(i \rightarrow j \mid i)$):

$$e_{i \rightarrow j} = g_i^{ij} \times \Pr(i \rightarrow j \mid i) \tag{A.1}$$

g_i^{ij} was not directly accessible from the data because we could not know which base calls were correctly sequenced. Though, this number was linked to the number of central markers containing an i allele and involved in a polymorphic site ij (n_i^{ij}) through the following equation:

$$n_i^{ig} = g_i^{ij} \times (1 - \Pr(i \rightarrow j \mid i)) + g_j^{ij} \times \Pr(j \rightarrow i \mid j) \tag{A.2}$$

When we computed the M matrix, we found that the frequency of sequencing errors was very low ($\simeq 10^{-3}$). Thus, to approximate g_i^{ij}, we used the simplifying assumption that the frequency of wrong calls were close to zero and that of good calls close to 1:

$$\forall (i,j) \in \{A,C,G,T\}^2 \setminus i \neq j, \Pr(i \rightarrow j \mid i) \simeq 0, \tag{A.3a}$$

$$\forall i \quad \in \{A,C,G,T\}, \qquad \Pr(i \rightarrow i \mid i) \simeq 1 \tag{A.3b}$$

From equation A.3a, equation A.2 simplified to:

$$n_i^{ij} \simeq g_i^{ij} \tag{A.4}$$

And, by incorporating equation A.4 into equation A.1, we had:

$$e_{i \rightarrow j} = n_i^{ij} \times \Pr(i \rightarrow j \mid i)$$

Finally, the total number of FPs that were expected due to sequencing errors (E) was the total sum of each type of sequencing error:

Target category	Nb of targets	Nb of fragments	Nb of events	Event rate ($\times\ 10^{-6}$)
Hotspots	1,018	228,984,512	243,390	1062.9
Controls	500	106,850,906	110,615	1035.2
FP rate				**97.4 %**

Table A.2: Number of pot-NCO-1 events detected in hotspot and control targets without the sequencing error filter.
Pot-NCO-1 events were detected without the sequencing error filter controlling that the allele supporting the genotype call with the mapping onto the B6 genome is identical to that based on the mapping onto the CAST genome. All fragments or events overlapping at least 1 bp with a given target are counted in this table. The event rate corresponds to the ratio of candidate recombination events over the total number of fragments. The maximum false positive (FP) rate is the ratio of the event rate in control targets over that in hotspots.

$$E = \sum_{\substack{(i,j)\in\{A,C,G,T\}^2 \\ i\neq j}} e_{i\to j}$$

This allowed us to predict that, among the total 287,577,349 fragments overlapping 3 markers or more, 231,905 were expected to be discovered as NCO-1 FPs due to sequencing errors only. This represented 66.7% of the 347,652[2] NCO-1 FPs that we found in pot-NCO-1 events (110,615 in control regions + an estimate of 237,037 in hotspots, Table A.2).

We further evaluated the imprecision on this percentage by calculating, for each sample individually[3], the ratio between the latter number of FPs expected in the sample due to sequencing errors and the total number of fragments in the sample. We sequentially applied the multiplier of each sample to the total number of fragments and finally determined that the proportion of FPs due to sequencing errors capped between 60 and 78% of all FPs.

Therefore, the largest part (66.7%, CI = [60%; 78%]) of FPs arose from sequencing errors. The next step thus consisted in estimating the $\frac{WS}{WS+SW}$ ratio expected

---

[2]The sequencing error estimate was calculated upon all sequenced fragments, i.e. before setting the sequencing error filter, and thus had to be compared to the total number of NCO-1 FPs obtained without the filter (Table A.2).

[3]With the exception of the four samples which were lowly sequenced

because of these sequencing errors. To do this, we simply computed the total number of FPs containing an erroneous W → S base call ($E_{W \to S}$) and the number containing an erroneous S → W base call ($E_{S \to W}$) as follows:

$$E_{W \to S} = e_{A \to C} + e_{A \to G} + e_{T \to C} + e_{T \to G}, \tag{A.5}$$

$$E_{S \to W} = e_{C \to A} + e_{C \to T} + e_{G \to A} + e_{G \to T} \tag{A.6}$$

Importantly, we found that $E_{S \to W}$ was greater than $E_{W \to S}$, i.e. S bases were more oftenly mistakenly sequenced as W bases than the other way round. More precisely, we found that the $\frac{WS}{WS+SW}$ ratio expected with such FPs (i.e. $\frac{E_{W \to S}}{E_{W \to S}+E_{S \to W}}$) equalled 0.39.

We note that this estimate was slightly higher than the $\frac{WS}{WS+SW}$ observed in control regions (0.31), possibly because the non-negligible portion (33.3%) of FPs that did not originate from these sequencing errors may somehow also bias the ratio.

A.2 Supplementary figures for Chapters 6 and 7

A.2.1 Figures of recombination events per hotspot

The figures corresponding to the recombination events detected on all 889 recombination hotspots displaying at least one event will be accessible until the end of year 2019 at the following url: `https://drive.google.com/open?id=1d48R_npcqyWTCixwiMpo9DC2oyrLV4v_`.

Afterwards, they might be moved to another location online (unknown at the time this manuscript was written).

A.2.2   Distribution of switch points
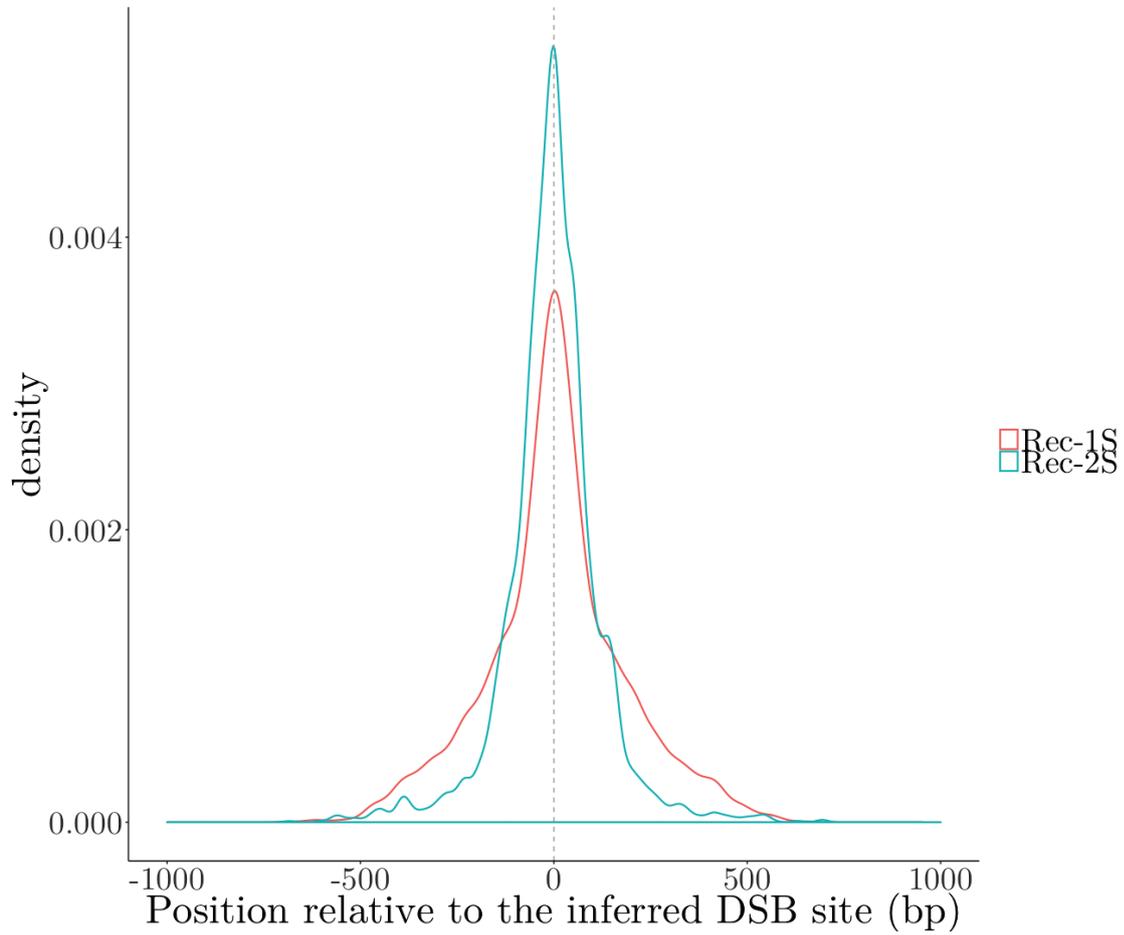
Figure A.1: Distribution of switch points along hotspots for Rec-1S and Rec-2S events.

A.2.3   Correlation between expected and observed donor
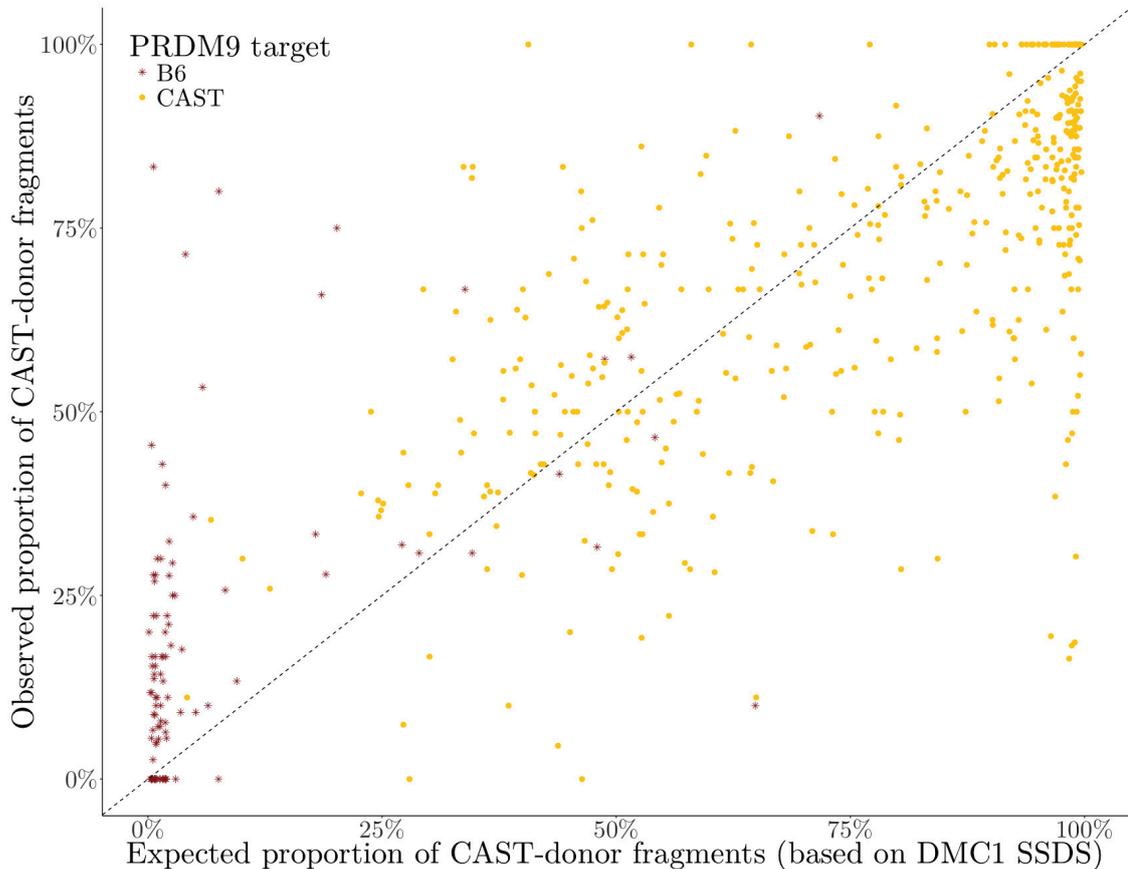
Figure A.2: Correlation between the expected and observed proportions of CAST-donor fragments across hotspots displaying at least 5 events, coloured per PRDM9 target.

The expected proportion of CAST-donor fragments (x-axis) was based on the probability that the DSB initiates on the B6 haplotype from DMC1 ssDNA-sequencing (SSDS) data by Smagulova et al. (2016) (see main text). Only the 582 hotspots displaying a minimum of 5 recombination events were reported in this figure. The Pearson correlation between the two measures gave: $R^2 = 0.66$; p-val $< 2.2 \times 10^{-16}$.

A.3   Supplementary figures for Chapter 8

A.3.1   Genetic background of all chromosomes
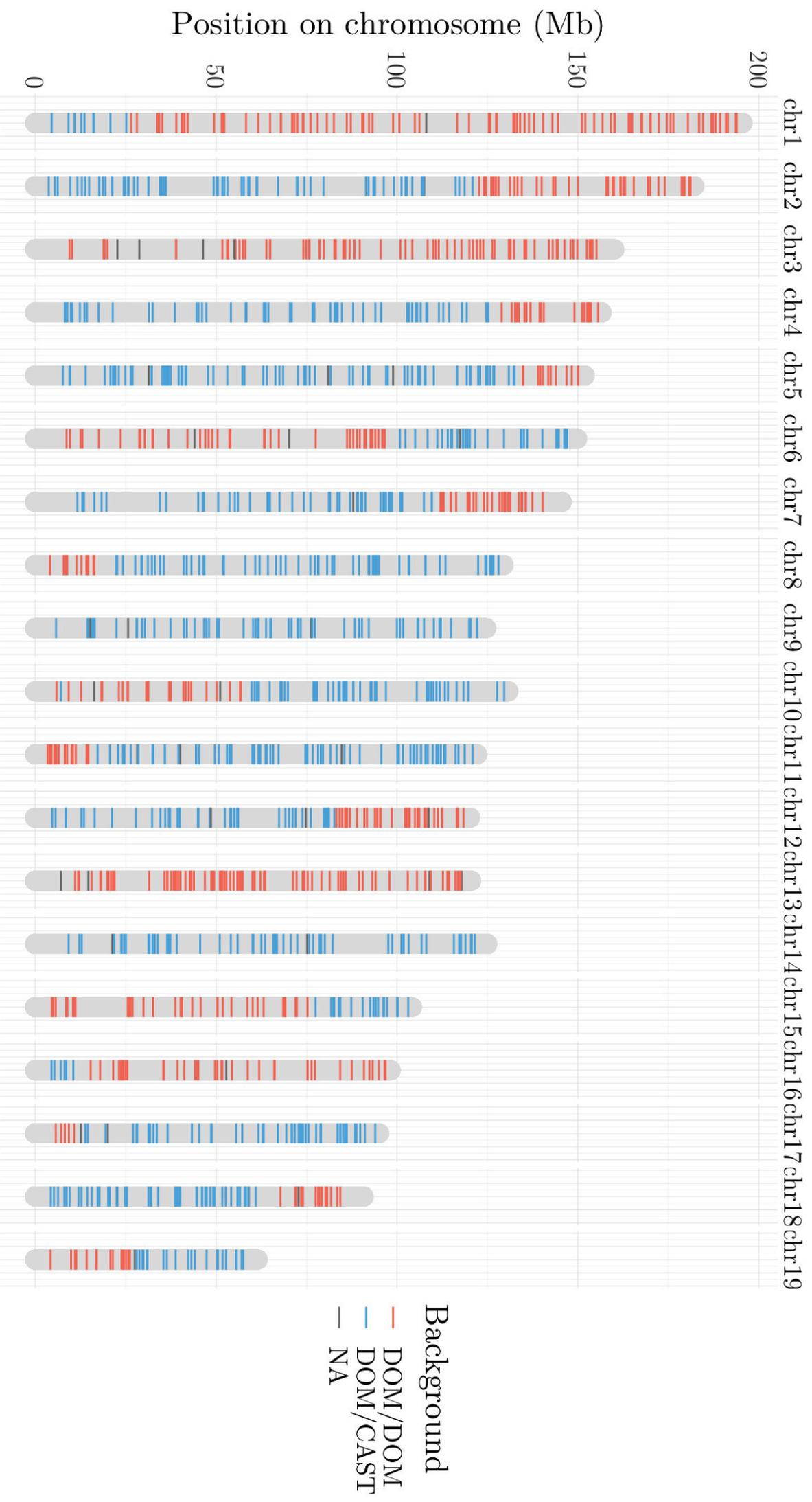
Figure A.3: Mosaic of genetic backgrounds inferred at each target along the autosomes of mouse 28355. Chromosomes are represented in grey and oriented so that the centromere is on the bottom side of the figure (mouse chromosomes are acrocentric). Each segment corresponds to the position of a target (hotspot or control region) and was coloured in red when the background inferred was BD/BD (homozygous) and in blue when the background inferred was BD/CAST (heterozygous).
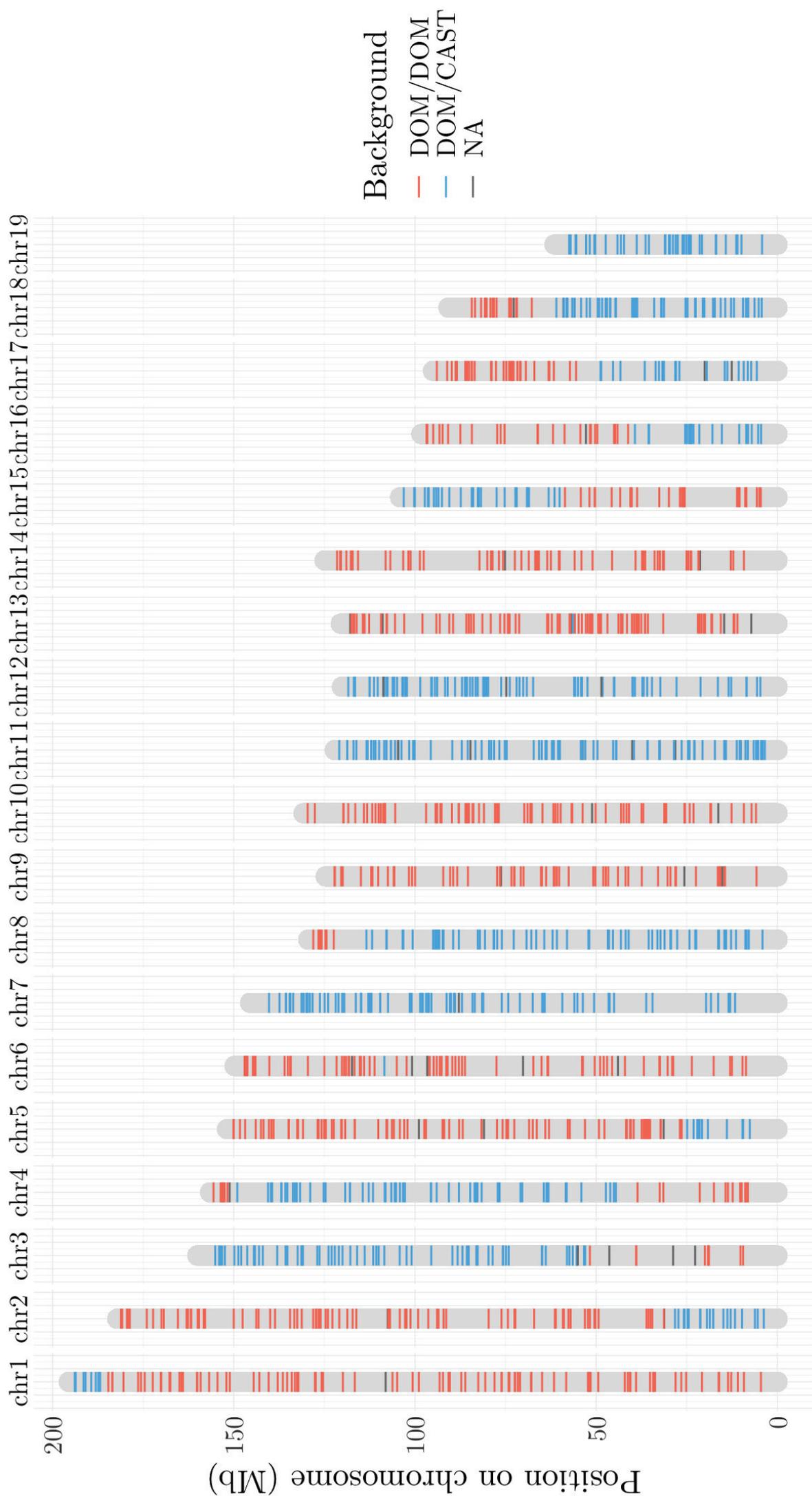
Figure A.4: Mosaic of genetic backgrounds inferred at each target along the autosomes of mouse 28367. Chromosomes are represented in grey and oriented so that the centromere is on the bottom side of the figure (mouse chromosomes are acrocentric). Each segment corresponds to the position of a target (hotspot or control region) and was coloured in red when the background inferred was BD/BD (homozygous) and in blue when the background inferred was BD/CAST (heterozygous).
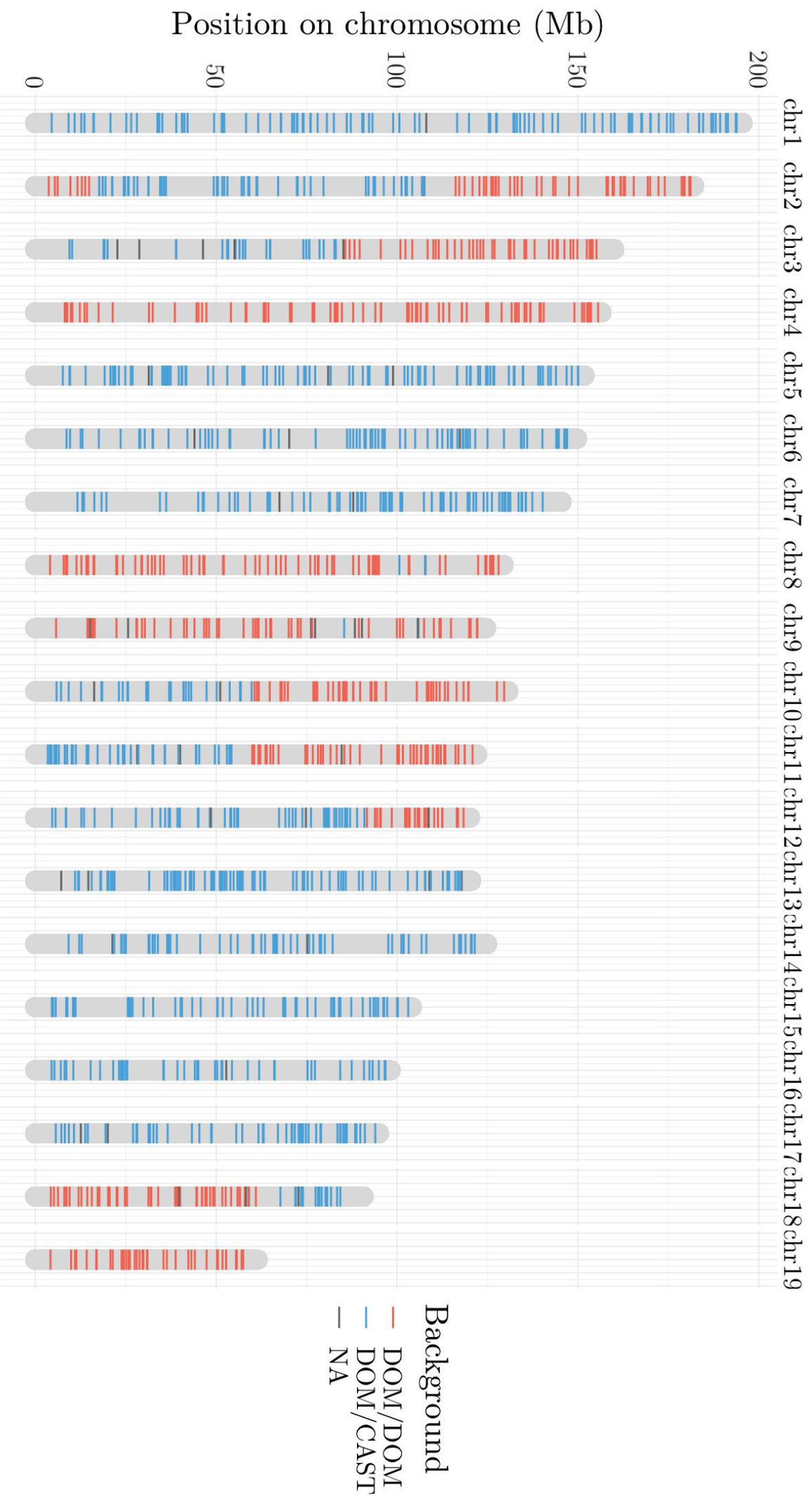
Figure A.5: Mosaic of genetic backgrounds inferred at each target along the autosomes of mouse 28371. Chromosomes are represented in grey and oriented so that the centromere is on the bottom side of the figure (mouse chromosomes are acrocentric). Each segment corresponds to the position of a target (hotspot or control region) and was coloured in red when the background inferred was BD/BD (homozygous) and in blue when the background inferred was BD/CAST (heterozygous).

A.3.2 Pairwise comparison of the RR in shared hotspots

(a) Between 28371 (WT) and 28353 (mutant)

(b) Between 28371 (WT) and 28367 (mutant)
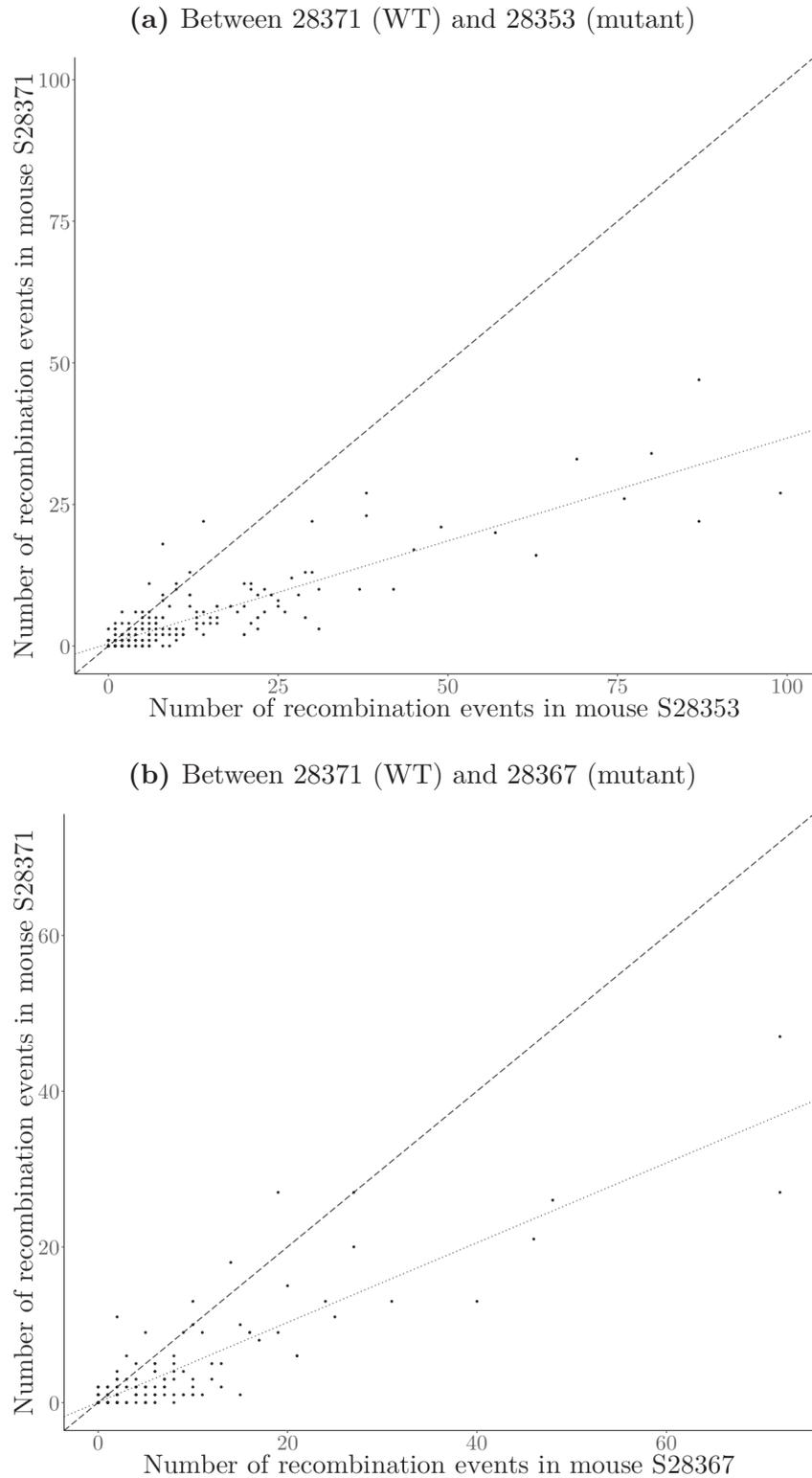
Figure A.6: Correlation of the number of recombination events in shared hotspots between the 28371 WT mouse and the two mutant mice.

(a) Between 28355 (WT) and 28367 (mutant)

(b) Between 28355 (WT) and 28353 (mutant)
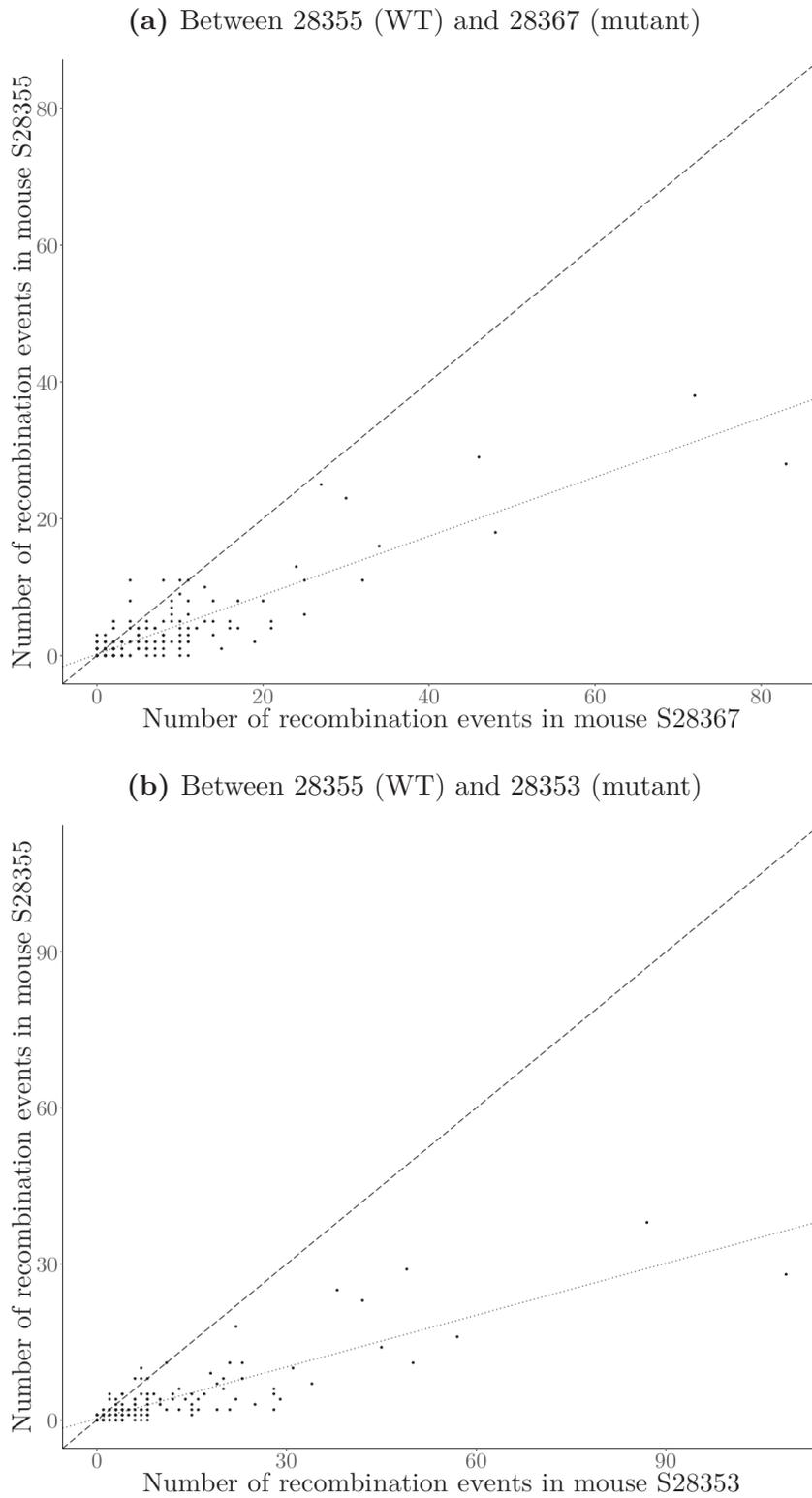
Figure A.7: Correlation of the number of recombination events in shared hotspots between the 28355 WT mouse and the two mutant mice.

A.3.3   Pairwise comparison of the rate of Rec-1S events

(a) Between the two WT mice

(b) Between the two mutant mice
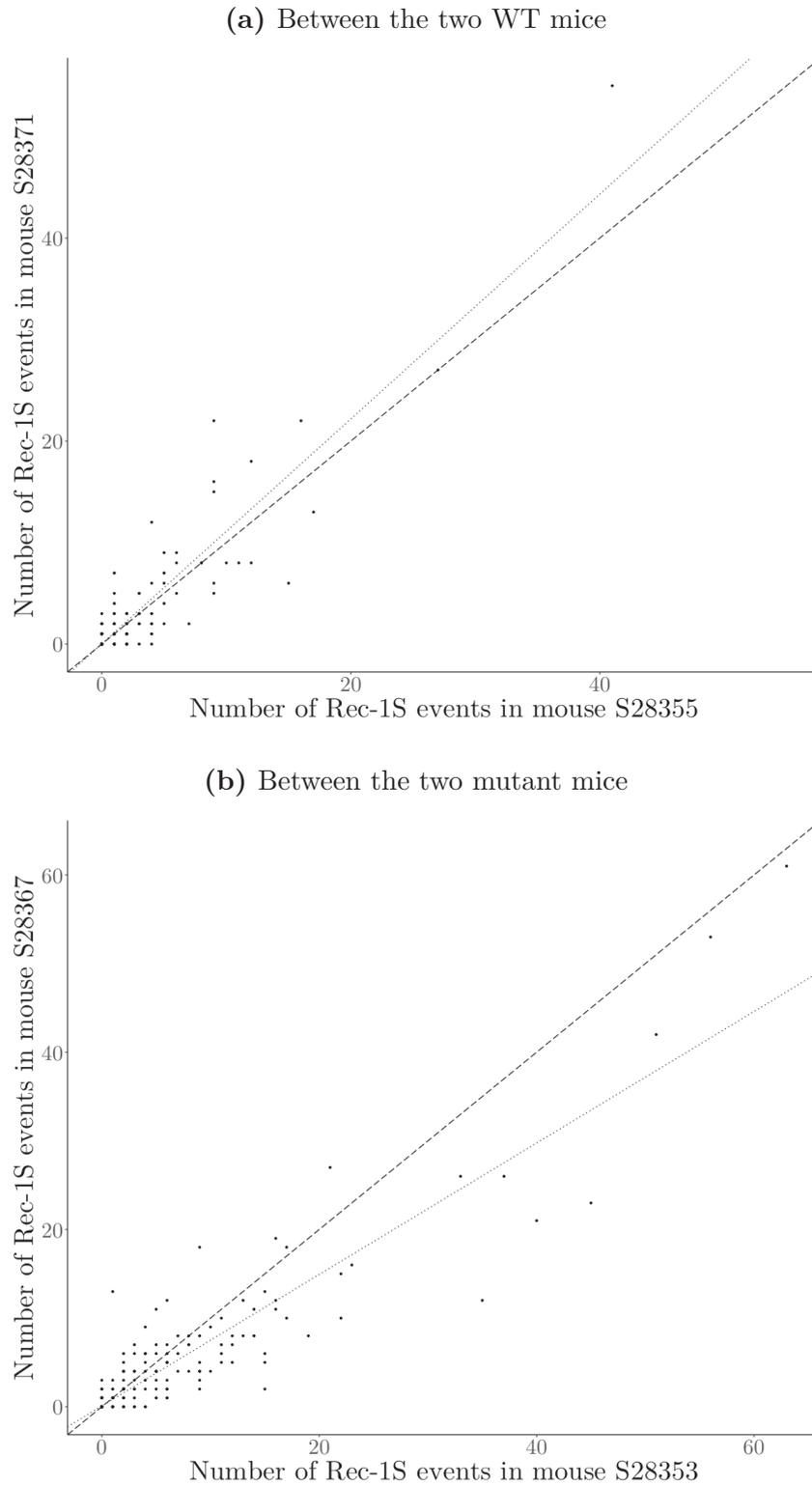
Figure A.8: Correlation of the number of Rec-1S events in shared hotspots for the two WT (a) and the two mutant (b) mice.

(a) Between 28371 (WT) and 28353 (mutant)

(b) Between 28371 (WT) and 28367 (mutant)
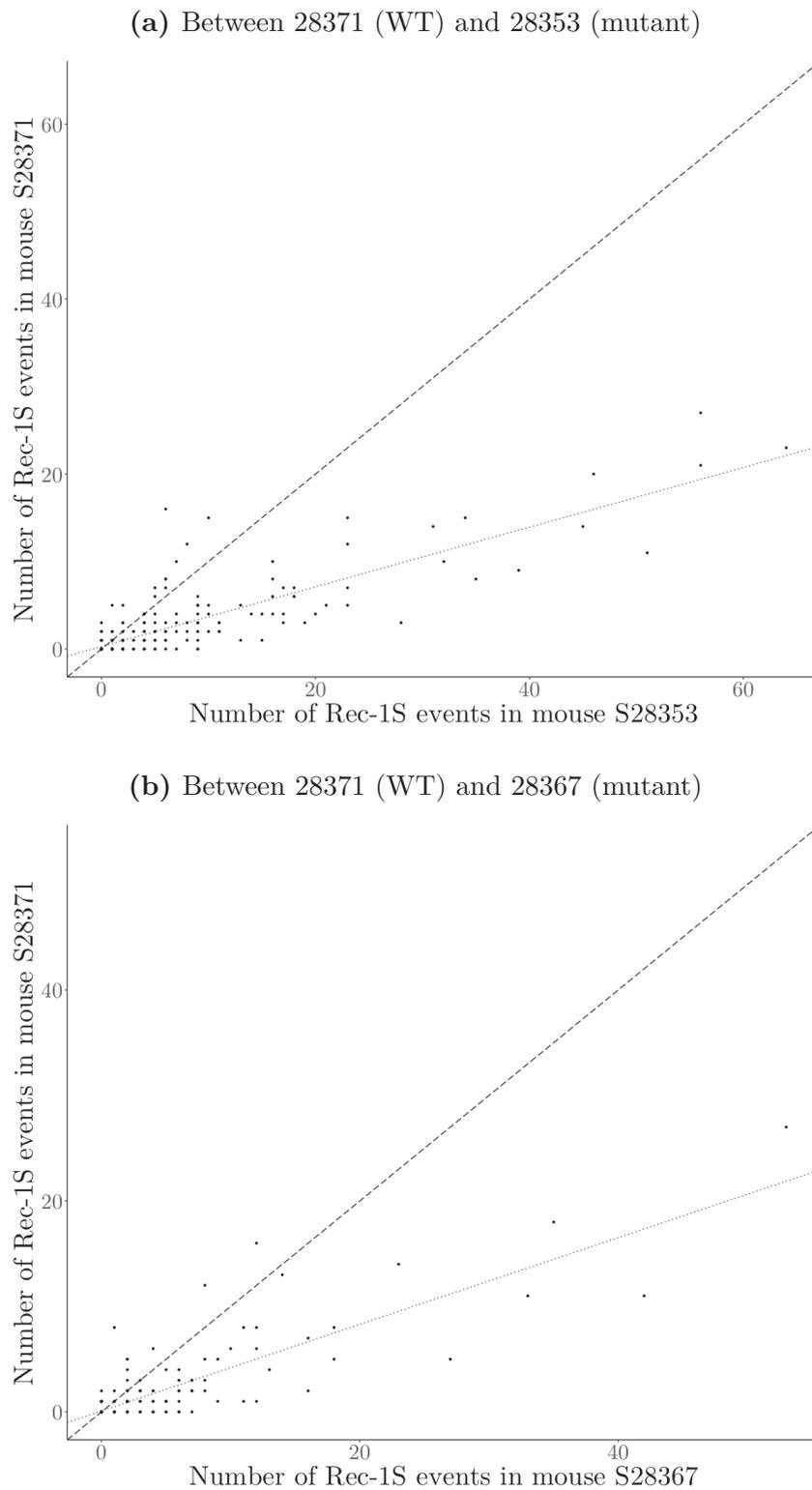
Figure A.9: Correlation of the number of Rec-1S events in shared hotspots between the 28371 WT mouse and the two mutant mice.

(a) Between 28355 (WT) and 28367 (mutant)

(b) Between 28355 (WT) and 28353 (mutant)
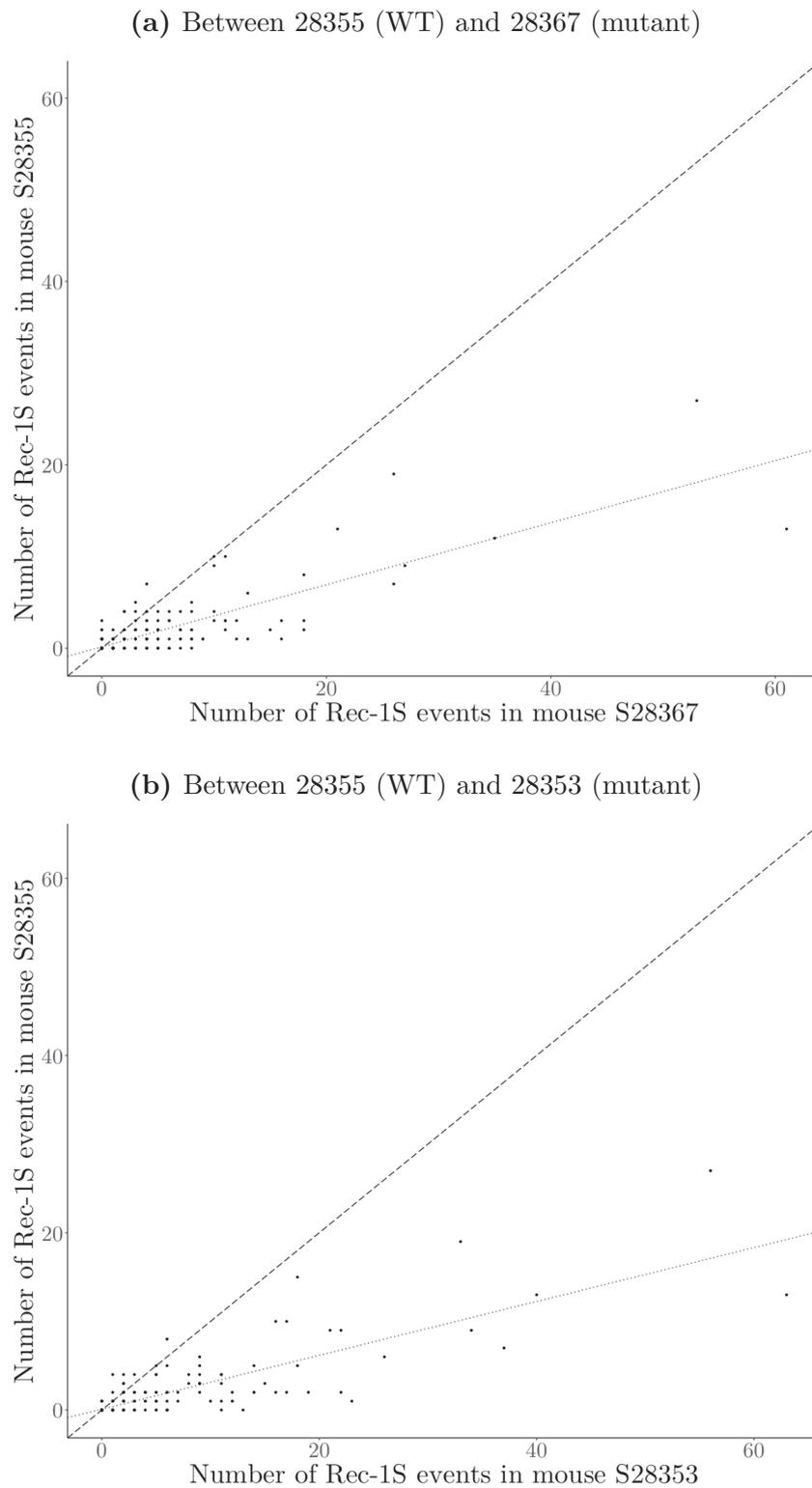
Figure A.10: Correlation of the number of Rec-1S events in shared hotspots between the 28355 WT mouse and the two mutant mice.

'Scientists are not dependent on the ideas of a single man, but on the combined wisdom of thousands of men, all thinking of the same problem and each doing his little bit to add to the great structure of knowledge which is gradually being erected.'

— Sir Ernest Rutherford, *Forty Years of Physics*
(1939)

B
Permissions to reproduce figures

The licence numbers for all the figures which were reproduced from other journals were obtained either *via* Copyright Clearance Center or *via* PLS Clear.

The licence numbers for these figures are the following:

- MG_Thesis_072519 for Figure 1.1 (email from Alexis Gargin of Macmillan Learning);

- 9781429257213 for Figure 2.1;

- 4606001226406 for Figures 2.2 and 2.5;

- 4606010096003 for Figure 2.3;

- 4606011369265 for Figure 2.4;

- 4606020741232 for Figures 3.1 and 3.2;

- 4606030051154 for Figure 3.3;

- 4606031234078 for Figure 4.2;

- 4606040190006 for Figure 4.3;

- 4606040523385 for Figure 4.5

Figure 2.6 was reproduced with permission from Cold Spring Harbor Laboratory Press (no licence number but a confirmation email from Carol Brown of the Cold Spring Harbor Laboratory Press received on July 26th, 2019).

Figures 4.1 and 4.4 were already under a licence format[1] authorising their full reproduction 6 months after the publication[2].

Figures 3.4[3] and 1.2[4] were both in the public domain and, thus, no permission was necessary to reproduce them.

No permission was necessary for Figure 1.4, because it was reproduced for a noncommercial use[5].

---

[1]See `https://creativecommons.org/licenses/by-nc/4.0/`.
[2]As mentioned here: `https://genome.cshlp.org/site/misc/permissions_landing.xhtml`.
[3]See `http://lewiscarrollsociety.org.uk/pages/eventspeopleplaces/LewisCarrollAndUKCopyright.pdf`.
[4]See `https://us.macmillan.com/henryholt/about/faq#p`.
[5]See `https://www.pnas.org/page/about/rights-permissions`.

'Of course, trying to keep an intellectual away from literature works about as well as recommending chastity to Homo sapiens, the sexiest primate of all.'

— Stephen Jay Gould, *The median isn't the message* (1985)

References

Abdu, U., Brodsky, M., and Schüpbach, T. (2002). Activation of a meiotic checkpoint during Drosophila oogenesis regulates the translation of Gurken through Chk2/Mnk. *Current biology: CB*, 12(19):1645–1651. *Cited at page 34*

Abnizova, I., Leonard, S., Skelly, T., Brown, A., Jackson, D., Gourtovaia, M., Qi, G., Te Boekhorst, R., Faruque, N., Lewis, K., and Cox, T. (2012). Analysis of context-dependent errors for illumina sequencing. *Journal of Bioinformatics and Computational Biology*, 10(02):1241005. *Cited at page 118*

Adelman, C. A. and Petrini, J. H. J. (2008). ZIP4H (TEX11) deficiency in the mouse impairs meiotic double strand break repair and the regulation of crossing over. *PLoS genetics*, 4(3):e1000042. *Cited at page 44*

Ahlawat, S., De, S., Sharma, P., Sharma, R., Arora, R., Kataria, R. S., Datta, T. K., and Singh, R. K. (2017). Evolutionary dynamics of meiotic recombination hotspots regulator PRDM9 in bovids. *Molecular genetics and genomics: MGG*, 292(1):117–131. *Cited at page 68*

Ahlawat, S., Sharma, P., Sharma, R., Arora, R., and De, S. (2016a). Zinc Finger Domain of the PRDM9 Gene on Chromosome 1 Exhibits High Diversity in Ruminants but Its Paralog PRDM7 Contains Multiple Disruptive Mutations. *PLoS ONE*, 11(5). *Cited at pages 68, 71*

Ahlawat, S., Sharma, P., Sharma, R., Arora, R., Verma, N. K., Brahma, B., Mishra, P., and De, S. (2016b). Evidence of positive selection and concerted evolution in the rapidly evolving PRDM9 zinc finger domain in goats and sheep. *Animal Genetics*, 47(6):740–751. *Cited at page 68*

Akashi, H. (1994). Synonymous codon usage in Drosophila melanogaster: Natural selection and translational accuracy. *Genetics*, 136(3):927–935. *Cited at page 89*

Akashi, H. (2003). Translational selection and yeast proteome evolution. *Genetics*, 164(4):1291–1303. *Cited at page 90*

Akashi, H. and Eyre-Walker, A. (1998). Translational selection and molecular evolution. *Current Opinion in Genetics & Development*, 8(6):688–693. *Cited at page 89*

Alani, E., Reenan, R. A., and Kolodner, R. D. (1994). Interaction between mismatch repair and genetic recombination in Saccharomyces cerevisiae. *Genetics*, 137(1):19–39. *Cited at page 83*

Allers, T. and Lichten, M. (2001). Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell*, 106(1):47–57. *Cited at page 38*

Altemose, N., Noor, N., Bitoun, E., Tumian, A., Imbeault, M., Chapman, J. R., Aricescu, A. R., and Myers, S. R. (2017). A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis. *eLife*, 6:e28383. *Cited at page 70*

Andersen, S. L. and Sekelsky, J. (2010). Meiotic versus Mitotic Recombination: Two Different Routes for Double-Strand Break Repair. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 32(12):1058–1066. *Cited at page 58*

Anderson, C. M., Oke, A., Yam, P., Zhuge, T., and Fung, J. C. (2015). Reduced Crossover Interference and Increased ZMM-Independent Recombination in the Absence of Tel1/ATM. *PLOS Genetics*, 11(8):e1005478. *Cited at page 56*

Anderson, L. K., Reeves, A., Webb, L. M., and Ashley, T. (1999). Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localization of MLH1 protein. *Genetics*, 151(4):1569–1579. *Cited at page 57*

Aquadro, C. F. (1997). Insights into the evolutionary process from patterns of DNA sequence variability. *Current Opinion in Genetics & Development*, 7(6):835–840. *Cited at page 61*

Arbeithuber, B., Betancourt, A. J., Ebner, T., and Tiemann-Boege, I. (2015). Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *Proceedings of the National Academy of Sciences*, 112(7):2109–2114. *Cited at pages 61, 92, 96*

Arndt, P. F., Petrov, D. A., and Hwa, T. (2003). Distinct changes of genomic biases in nucleotide substitution at the time of Mammalian radiation. *Molecular Biology and Evolution*, 20(11):1887–1896. *Cited at page 87*

Arnheim, N., Calabrese, P., and Nordborg, M. (2003). Hot and Cold Spots of Recombination in the Human Genome: The Reason We Should Find Them and How This Can Be Achieved. *American Journal of Human Genetics*, 73(1):5–16. *Cited at page 53*

Arnheim, N., Calabrese, P., and Tiemann-Boege, I. (2007). Mammalian meiotic recombination hot spots. *Annual Review of Genetics*, 41:369–399. *Cited at pages 63, 72*

Auton, A., Fledel-Alon, A., Pfeifer, S., Venn, O., Ségurel, L., Street, T., Leffler, E. M., Bowden, R., Aneas, I., Broxholme, J., Humburg, P., Iqbal, Z., Lunter, G., Maller, J., Hernandez, R. D., Melton, C., Venkat, A., Nobrega, M. A., Bontrop, R., Myers, S., Donnelly, P., Przeworski, M., and McVean, G. (2012). A Fine-Scale Chimpanzee Genetic Map from Population Sequencing. *Science*, 336(6078):193–198. *Cited at pages 52, 60, 63, 66*

Auton, A., Li, Y. R., Kidd, J., Oliveira, K., Nadel, J., Holloway, J. K., Hayward, J. J., Cohen, P. E., Greally, J. M., Wang, J., Bustamante, C. D., and Boyko, A. R. (2013). Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs. *PLOS Genetics*, 9(12):e1003984. *Cited at pages 62, 69*

Axelsson, E., Webster, M. T., Ratnakumar, A., Consortium, T. L., Ponting, C. P., and Lindblad-Toh, K. (2012). Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. *Genome Research*, 22(1):51–63. *Cited at pages 63, 68*

Baarends, W. M., Wassenaar, E., van der Laan, R., Hoogerbrugge, J., Sleddens-Linkels, E., Hoeijmakers, J. H. J., de Boer, P., and Grootegoed, J. A. (2005). Silencing of unpaired chromatin and histone H2A ubiquitination in mammalian meiosis. *Molecular and Cellular Biology*, 25(3):1041–1053. *Cited at page 35*

Backström, N. (2009). *Gene Mapping in Ficedula Flycatchers*. PhD thesis, Acta universitatis Upsaliensis : Universitetsbiblioteket [distributör], Uppsala. OCLC: 938017950. *Cited at page 49*

Backström, N., Ceplitis, H., Berlin, S., and Ellegren, H. (2005). Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. *Molecular Biology and Evolution*, 22(10):1992–1999. *Cited at page 86*

Backström, N., Karaiskou, N., Leder, E. H., Gustafsson, L., Primmer, C. R., Qvarnström, A., and Ellegren, H. (2008). A Gene-Based Genetic Linkage Map of the Collared Flycatcher (Ficedula albicollis) Reveals Extensive Synteny and Gene-Order Conservation During 100 Million Years of Avian Evolution. *Genetics*, 179(3):1479–1495. *Cited at page 65*

Bahler, J., Wyler, T., Loidl, J., and Kohli, J. (1993). Unusual nuclear structures in meiotic prophase of fission yeast: a cytological analysis. *J. Cell Biol.*, 121(2):241–256. *Cited at page 34*

Baier, B., Hunt, P., Broman, K. W., and Hassold, T. (2014). Variation in genome-wide levels of meiotic recombination is established at the onset of prophase in mammalian males. *PLoS genetics*, 10(1):e1004125. *Cited at page 66*

Bailey, J. A. and Eichler, E. E. (2006). Primate segmental duplications: Crucibles of evolution, diversity and disease. *Nature Reviews. Genetics*, 7(7):552–564. *Cited at page 46*

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, 37(suppl_2):W202–W208. *Cited at page 149*

Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, 34(suppl_2):W369–W373. *Cited at page 149*

Baker, C. L., Kajita, S., Walker, M., Saxl, R. L., Raghupathy, N., Choi, K., Petkov, P. M., and Paigen, K. (2015a). PRDM9 Drives Evolutionary Erosion of Hotspots in Mus musculus through Haplotype-Specific Initiation of Meiotic Recombination. *PLoS Genetics*, 11(1). *Cited at pages , 70, 77, 110, 126, 128, 129, 130, 131, 132, 136, 146, 147, 148, 149, 150, 166*

Baker, C. L., Petkova, P., Walker, M., Flachs, P., Mihola, O., Trachtulec, Z., Petkov, P. M., and Paigen, K. (2015b). Multimer Formation Explains Allelic Suppression of PRDM9 Recombination Hotspots. *PLOS Genetics*, 11(9):e1005512. *Cited at pages 70, 217*

Baker, C. L., Walker, M., Kajita, S., Petkov, P. M., and Paigen, K. (2014). PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Research*, 24(5):724–732. *Cited at page 149*

Baker, S. M., Plug, A. W., Prolla, T. A., Bronner, C. E., Harris, A. C., Yao, X., Christie, D. M., Monell, C., Arnheim, N., Bradley, A., Ashley, T., and Liskay, R. M. (1996). Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. *Nature Genetics*, 13(3):336–342.   *Cited at page 44*

Baker, Z., Schumer, M., Haba, Y., Bashkirova, L., Holland, C., Rosenthal, G. G., and Przeworski, M. (2017). Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife*, 6.   *Cited at page 69*

Barchi, M., Mahadevaiah, S., Di Giacomo, M., Baudat, F., de Rooij, D. G., Burgoyne, P. S., Jasin, M., and Keeney, S. (2005). Surveillance of different recombination defects in mouse spermatocytes yields distinct responses despite elimination at an identical developmental stage. *Molecular and Cellular Biology*, 25(16):7203–7215.
Cited at page 35

Barnes, B. (1977). *Interests and the Growth of Knowledge*. Routledge and Kegan Paul.   *Cited at page 204*

Barnett, J. A. (2007). A history of research on yeasts: foundations of yeast genetics. *Yeast*, 24(10):799–845.   *Cited at page 14*

Barton, A. B., Pekosz, M. R., Kurvathi, R. S., and Kaback, D. B. (2008). Meiotic Recombination at the Ends of Chromosomes in Saccharomyces cerevisiae. *Genetics*, 179(3):1221–1235.   *Cited at page 60*

Barton, N. H. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 355(1403):1553–1562.
Cited at page 52

Barzel, A. and Kupiec, M. (2008). Finding a match: How do homologous sequences get together for recombination? *Nature Reviews. Genetics*, 9(1):27–37.
Cited at page 31

Bateson, P. (2002). William Bateson: a biologist ahead of his time. *Journal of Genetics*, 81(2):49–58.   *Cited at page 9*

Bateson, W. (1902). *Mendel's Principles of Heredity: A Defense*. Cambridge University Press.   *Cited at page 9*

Bateson, W. and Killby, H. (1905). Experimental studies in the physiology of heredity: Peas (*Pisum sativum*). *Royal Society Reports to the Evolution Committee*, 2:55–80.   *Cited at page 9*

Bateson, W., Saunders, E., and Punnett, R. (1905). Experimental studies in the physiology of heredity. *Royal Society Reports to the Evolution Committee*, 2.
Cited at page 9

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010). PRDM9 is a Major Determinant of Meiotic Recombination Hotspots in humans and mice. *Science (New York, N.Y.)*, 327(5967):836–840.   *Cited at pages , 48, 67, 68, 70, 71, 74*

Baudat, F. and de Massy, B. (2007). Cis- and Trans-Acting Elements Regulate the Mouse Psmb9 Meiotic Recombination Hotspot. *PLOS Genetics*, 3(6):e100.
Cited at pages 54, 72, 126, 217

Baudat, F. and de Massy, B. (2007). Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 15(5):565–577. *Cited at pages 42, 44, 135, 139*

Baudat, F., Imai, Y., and De Massy, B. (2013). Meiotic recombination in mammals: localization and regulation. *Nature Reviews Genetics*, 14(11):794. *Cited at pages 29, 40, 41, 42, 44*

Baudat, F., Manova, K., Yuen, J. P., Jasin, M., and Keeney, S. (2000). Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Molecular Cell*, 6(5):989–998. *Cited at page 39*

Baudat, F. and Nicolas, A. (1997). Clustering of meiotic double-strand breaks on yeast chromosome III. *Proceedings of the National Academy of Sciences of the United States of America*, 94(10):5213–5218. *Cited at pages 60, 62*

Bechtel, W. (2006). *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge University Press. *Cited at page 205*

Begun, D. J. and Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. *Nature*, 356(6369):519–520. *Cited at pages 22, 61*

Behura, S. K. and Severson, D. W. (2011). Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in Aedes aegypti and Anopheles gambiae. *Insect Molecular Biology*, 20(2):177–187. *Cited at page 89*

Bell, G. (1982). *The Masterpiece of Nature:: The Evolution and Genetics of Sexuality*. CUP Archive. *Cited at page 24*

Bellani, M. A., Boateng, K. A., McLeod, D., and Camerini-Otero, R. D. (2010). The expression profile of the major mouse SPO11 isoforms indicates that SPO11beta introduces double strand breaks and suggests that SPO11alpha has an additional role in prophase in both spermatocytes and oocytes. *Molecular and Cellular Biology*, 30(18):4391–4403. *Cited at page 39*

Belle, E. M. S., Duret, L., Galtier, N., and Eyre-Walker, A. (2004). The decline of isochores in mammals: An assessment of the GC content variation along the mammalian phylogeny. *Journal of Molecular Evolution*, 58(6):653–660. *Cited at page 85*

Belle, E. M. S., Smith, N., and Eyre-Walker, A. (2002). Analysis of the phylogenetic distribution of isochores in vertebrates and a test of the thermal stability hypothesis. *Journal of Molecular Evolution*, 55(3):356–363. *Cited at page 83*

Bender, J. (2004). Dna methylation and epigenetics. *Annu. Rev. Plant Biol.*, 55:41–68. *Cited at page 30*

Bengtsson, B. O. (1986). Biased conversion as the primary function of recombination. *Genetics Research*, 47(1):77–80. *Cited at page 92*

Bennett, J. H., Hayman, D. L., and Hope, R. M. (1986). Novel sex differences in linkage values and meiotic chromosome behaviour in a marsupial. *Nature*, 323(6083):59. *Cited at page 65*

Bennetzen, J. L. and Hall, B. D. (1982). Codon selection in yeast. *Journal of Biological Chemistry*, 257(6):3026–3031. *Cited at page 90*

Berchowitz, L. E., Hanlon, S. E., Lieb, J. D., and Copenhaver, G. P. (2009). A positive but complex association between meiotic double-strand break hotspots and open chromatin in Saccharomyces cerevisiae. *Genome Research*, 19(12):2245–2257. *Cited at page 62*

Berg, I. L., Neumann, R., Lam, K.-W. G., Sarbajna, S., Odenthal-Hesse, L., May, C. A., and Jeffreys, A. J. (2010). PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature genetics*, 42(10):859–863. *Cited at pages 68, 71*

Berg, I. L., Neumann, R., Sarbajna, S., Odenthal-Hesse, L., Butler, N. J., and Jeffreys, A. J. (2011). Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. *Proceedings of the National Academy of Sciences*, 108(30):12378–12383. *Cited at pages 66, 71*

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P. C., Nicolas, A., and Forterre, P. (1997). An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature*, 386(6623):414–417. *Cited at page 39*

Berglund, J., Pollard, K. S., and Webster, M. T. (2009). Hotspots of Biased Nucleotide Substitutions in Human Genes. *PLOS Biology*, 7(1):e1000026. *Cited at page 91*

Bernardi, G. (2000). Isochores and the evolutionary genomics of vertebrates. *Gene*, 241(1):3–17. *Cited at page 82*

Bernardi, G. (2005). The distribution of genes in human genome. In *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics.* American Cancer Society. *Cited at page 81*

Bernardi, G. (2007). The neoselectionist theory of genome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 104(20):8385–8390. *Cited at page 82*

Bernardi, G. (2012). The genome: An isochore ensemble and its evolution. *Annals of the New York Academy of Sciences*, 1267(1):31–34. *Cited at page 82*

Bernardi, G. and Bernardi, G. (1990). Compositional transitions in the nuclear genomes of cold-blooded vertebrates. *Journal of Molecular Evolution*, 31(4):282–293. *Cited at page 81*

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., and Rodier, F. (1985). The mosaic genome of warm-blooded vertebrates. *Science*, 228(4702):953–958. *Cited at page 81*

Bernstein, H. and Bernstein, C. (2010). Evolutionary origin of recombination during meiosis. *BioScience*, 60(7):498–505. *Cited at page 25*

Bernstein, H., Bernstein, C., and Michod, R. E. (2011). Meiosis as an evolutionary adaptation for dna repair. In *DNA repair.* IntechOpen. *Cited at page 25*

Beye, M., Gattermeier, I., Hasselmann, M., Gempe, T., Schioett, M., Baines, J. F., Schlipalius, D., Mougel, F., Emore, C., Rueppell, O., Sirviö, A., Guzmán-Novoa, E., Hunt, G., Solignac, M., and Page, R. E. (2006). Exceptionally high levels of recombination across the honey bee genome. *Genome Research*, 16(11):1339–1344.
Cited at page 55

Bhalla, N. and Dernburg, A. F. (2005). A conserved checkpoint monitors meiotic chromosome synapsis in Caenorhabditis elegans. *Science (New York, N.Y.)*, 310(5754):1683–1686.
Cited at page 34

Bhérer, C. and Auton, A. (2014). Biased Gene Conversion and Its Impact on Genome Evolution. In *eLS*. American Cancer Society.
Cited at page 91

Bhérer, C., Campbell, C. L., and Auton, A. (2017). Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. *Nature Communications*, 8.
Cited at page 65

Bierne, N. and Eyre-Walker, A. (2006). Variation in synonymous codon use and DNA polymorphism within the Drosophila genome. *Journal of Evolutionary Biology*, 19(1):1–11.
Cited at page 89

Bill, C. A., Duran, W. A., Miselis, N. R., and Nickoloff, J. A. (1998). Efficient repair of all types of single-base mismatches in recombination intermediates in Chinese hamster ovary cells. Competition between long-patch and G-T glycosylase-mediated repair of G-T mismatches. *Genetics*, 149(4):1935–1943. *Cited at page 84*

Billings, T., Parvanov, E. D., Baker, C. L., Walker, M., Paigen, K., and Petkov, P. M. (2013). DNA binding specificities of the long zinc-finger recombination protein PRDM9. *Genome Biology*, 14(4):R35.
Cited at page 217

Billings, T., Sargent, E. E., Szatkiewicz, J. P., Leahy, N., Kwak, I.-Y., Bektassova, N., Walker, M., Hassold, T., Graber, J. H., Broman, K. W., and Petkov, P. M. (2010). Patterns of Recombination Activity on Mouse Chromosome 11 Revealed by High Resolution Mapping. *PLOS ONE*, 5(12):e15340. *Cited at pages 51, 59*

Bird, C. P., Stranger, B. E., Liu, M., Thomas, D. J., Ingle, C. E., Beazley, C., Miller, W., Hurles, M. E., and Dermitzakis, E. T. (2007). Fast-evolving noncoding sequences in the human genome. *Genome Biology*, 8(6):R118. *Cited at page 91*

Birdsell, J. A. (2002). Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Molecular Biology and Evolution*, 19(7):1181–1197. *Cited at page 87*

Bishop, D. K. and Zickler, D. (2004). Early decision; meiotic crossover interference prior to stable strand exchange and synapsis. *Cell*, 117(1):9–15.
Cited at pages 37, 57

Blat, Y., Protacio, R. U., Hunter, N., and Kleckner, N. (2002). Physical and Functional Interactions among Basic Chromosome Organizational Features Govern Early Steps of Meiotic Chiasma Formation. *Cell*, 111(6):791–802. *Cited at page 87*

Blitzblau, H. G., Bell, G. W., Rodriguez, J., Bell, S. P., and Hochwagen, A. (2007). Mapping of Meiotic Single-Stranded DNA Reveals Double-Strand-Break Hotspots near Centromeres and Telomeres. *Current Biology*, 17(23):2003–2012.
Cited at pages 54, 60

Boateng, K. A., Bellani, M. A., Gregoretti, I. V., Pratto, F., and Camerini-Otero, R. D. (2013). Homologous pairing preceding spo11-mediated double-strand breaks in mice. *Developmental cell*, 24(2):196–205. *Cited at pages xi, 30, 31*

Bois, P. R. J. (2007). A Highly Polymorphic Meiotic Recombination Mouse Hot Spot Exhibits Incomplete Repair. *Molecular and Cellular Biology*, 27(20):7053–7062. *Cited at pages 126, 217*

Bolívar, P., Guéguen, L., Duret, L., Ellegren, H., and Mugal, C. F. (2019). GC-biased gene conversion conceals the prediction of the nearly neutral theory in avian genomes. *Genome Biology*, 20(1):5. *Cited at page 91*

Bolívar, P., Mugal, C. F., Nater, A., and Ellegren, H. (2016). Recombination Rate Variation Modulates Gene Sequence Evolution Mainly via GC-Biased Gene Conversion, Not Hill-Robertson Interference, in an Avian System. *Molecular Biology and Evolution*, 33(1):216–227. *Cited at page 95*

Bolívar, P., Mugal, C. F., Rossi, M., Nater, A., Wang, M., Dutoit, L., and Ellegren, H. (2018). Biased Inference of Selection Due to GC-Biased Gene Conversion and the Rate of Protein Evolution in Flycatchers When Accounting for It. *Molecular Biology and Evolution*, 35(10):2475–2486. *Cited at page 91*

Borde, V. (2007). The multiple roles of the Mre11 complex for meiotic recombination. *Chromosome Research*, 15(5):551–563. *Cited at page 40*

Borde, V. and Cobb, J. (2009). Double functions for the Mre11 complex during DNA double-strand break repair and replication. *The International Journal of Biochemistry & Cell Biology*, 41(6):1249–1253. *Cited at page 40*

Borde, V., Robine, N., Lin, W., Bonfils, S., Géli, V., and Nicolas, A. (2009). Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *The EMBO Journal*, 28(2):99–111. *Cited at pages 54, 61*

Borde, V., Wu, T. C., and Lichten, M. (1999). Use of a recombination reporter insert to define meiotic recombination domains on chromosome III of Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 19(7):4832–4842. *Cited at page 60*

Born, N., Thiesen, H.-J., and Lorenz, P. (2014). The B-Subdomain of the Xenopus laevis XFIN KRAB-AB Domain Is Responsible for Its Weaker Transcriptional Repressor Activity Compared to Human ZNF10/Kox1. *PLOS ONE*, 9(2):e87609. *Cited at page 69*

Börner, G. V., Kleckner, N., and Hunter, N. (2004). Crossover/Noncrossover Differentiation, Synaptonemal Complex Formation, and Regulatory Surveillance at the Leptotene/Zygotene Transition of Meiosis. *Cell*, 117(1):29–45. *Cited at page 57*

Borodin, P. M., Karamysheva, T. V., Belonogova, N. M., Torgasheva, A. A., Rubtsov, N. B., and Searle, J. B. (2008). Recombination Map of the Common Shrew, Sorex araneus (Eulipotyphla, Mammalia). *Genetics*, 178(2):621–632. *Cited at page 59*

Borts, R. H. and Haber, J. E. (1989). Length and distribution of meiotic gene conversion tracts and crossovers in Saccharomyces cerevisiae. *Genetics*, 123(1):69–80. *Cited at pages 45, 80*

Boselli, M., Rock, J., Unal, E., Levine, S. S., and Amon, A. (2009). Effects of age on meiosis in budding yeast. *Developmental Cell*, 16(6):844–855. *Cited at page 36*

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3):314–331. *Cited at page 49*

Boulton, A., Myers, R. S., and Redfield, R. J. (1997). The hotspot conversion paradox and the evolution of meiotic recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 94(15):8058–8063. *Cited at pages , 73, 193*

Boveri, T. (1888). *Zellen-Studien: Die Befruchtung und Teilung des Eies von Ascaris megalocephala.* G. Fischer. *Cited at page 9*

Boveri, T. (1890). *Zellen-Studien: Über das Verhalten der chromatischen Kernsubstanz bei der Bildung der Richtungskörper und bei der Befruchtung.* Number 3. G. Fischer. *Cited at page 9*

Bowler, P. J. (2003). *Evolution: The History of an Idea.* University of California Press. *Cited at page 20*

Bowring, F. J., Yeadon, P. J., Stainer, R. G., and Catcheside, D. E. A. (2006). Chromosome pairing and meiotic recombination in Neurospora crassa spo11 mutants. *Current Genetics*, 50(2):115–123. *Cited at page 39*

Braten, T. and Nordby, O. (1973). Ultrastructure of meiosis and centriole behaviour in Ulva mutabilis Foyn. *J. Cell. Sci.*, 13(1):69–81. *Cited at page 32*

Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R. D., and Petukhova, G. V. (2012). Genetic recombination is directed away from functional genomic elements in mice. *Nature*, 485(7400):642–645. *Cited at pages 54, 61, 68, 70, 126, 149*

Brick, K., Thibault-Sennett, S., Smagulova, F., Lam, K.-W. G., Pu, Y., Pratto, F., Camerini-Otero, R. D., and Petukhova, G. V. (2018). Extensive sex differences at the initiation of genetic recombination. *Nature*, 561(7723):338. *Cited at page 65*

Broad Institute (2018). Picard tools. http://broadinstitute.github.io/picard/. Version 1.98(1547). *Cited at page 113*

Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L., and Weber, J. L. (1998). Comprehensive Human Genetic Maps: Individual and Sex-Specific Variation in Recombination. *The American Journal of Human Genetics*, 63(3):861–869. *Cited at pages 51, 64*

Broman, K. W., Rowe, L. B., Churchill, G. A., and Paigen, K. (2002). Crossover interference in the mouse. *Genetics*, 160(3):1123–1131. *Cited at page 57*

Broman, K. W. and Weber, J. L. (2000). Characterization of human crossover interference. *American Journal of Human Genetics*, 66(6):1911–1926. *Cited at page 57*

Brown, J. and Sundaresan, V. (1991). A recombination hotspot in the maize A1 intragenic region. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 81(2):185–188. *Cited at page 63*

Brown, P. W., Judis, L., Chan, E. R., Schwartz, S., Seftel, A., Thomas, A., and Hassold, T. J. (2005). Meiotic synapsis proceeds from a limited number of subtelomeric sites in the human male. *American Journal of Human Genetics*, 77(4):556–566. *Cited at page 65*

Brown, T. A. (2002). *Mapping Genomes.* Wiley-Liss.   *Cited at page 49*

Brown, T. C. and Jiricny, J. (1988). Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. *Cell*, 54(5):705–711.
Cited at page 84

Brunschwig, H., Levi, L., Ben-David, E., Williams, R. W., Yakir, B., and Shifman, S. (2012). Fine-Scale Maps of Recombination Rates and Hotspots in the Mouse Genome. *Genetics*, 191(3):757–764.   *Cited at pages 52, 64*

Buard, J., Barthès, P., Grey, C., and de Massy, B. (2009). Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. *The EMBO Journal*, 28(17):2616–2624.   *Cited at page 61*

Buard, J. and de Massy, B. (2007). Playing hide and seek with mammalian meiotic crossover hotspots. *Trends in Genetics*, 23(6):301–309.   *Cited at page 65*

Buard, J., Rivals, E., Dunoyer de Segonzac, D., Garres, C., Caminade, P., de Massy, B., and Boursot, P. (2014). Diversity of Prdm9 Zinc Finger Array in Wild Mice Unravels New Facets of the Evolutionary Turnover of this Coding Minisatellite. *PLoS ONE*, 9(1).   *Cited at pages 68, 71*

Bugreev, D. V., Huang, F., Mazina, O. M., Pezza, R. J., Voloshin, O. N., Daniel Camerini-Otero, R., and Mazin, A. V. (2014). HOP2-MND1 modulates RAD51 binding to nucleotides and DNA. *Nature Communications*, 5:4198.
Cited at page 40

Buhler, C., Borde, V., and Lichten, M. (2007). Mapping meiotic single-strand DNA reveals a new landscape of DNA double-strand breaks in Saccharomyces cerevisiae. *PLoS biology*, 5(12):e324.   *Cited at pages 54, 60*

Buhler, C., Lebbink, J. H. G., Bocs, C., Ladenstein, R., and Forterre, P. (2001). DNA Topoisomerase VI Generates ATP-dependent Double-strand Breaks with Two-nucleotide Overhangs. *Journal of Biological Chemistry*, 276(40):37215–37222.
Cited at page 60

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics*, 129(3):897–907.   *Cited at page 89*

Burgoyne, P. S., Mahadevaiah, S. K., and Turner, J. M. A. (2009). The consequences of asynapsis for mammalian meiosis. *Nature Reviews. Genetics*, 10(3):207–216.
Cited at page 35

Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., and Chen, D. J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *The Journal of Biological Chemistry*, 276(45):42462–42467.   *Cited at page 39*

Bush, E. C. and Lahn, B. T. (2008). A genome-wide screen for noncoding elements important in primate evolution. *BMC evolutionary biology*, 8:17.   *Cited at page 91*

Butlin, R. K. (2005). Recombination and speciation. *Molecular Ecology*, 14(9):2621–2635.   *Cited at page 48*

Bzymek, M., Thayer, N. H., Oh, S. D., Kleckner, N., and Hunter, N. (2010). Double Holliday junctions are intermediates of DNA break repair. *Nature*, 464(7290):937–941.   *Cited at pages 33, 58*

Cacciò, S., Perani, P., Saccone, S., Kadi, F., and Bernardi, G. (1994). Single-copy sequence homology among the GC-richest isochores of the genomes from warm-blooded vertebrates. *Journal of Molecular Evolution*, 39(4):331–339.
Cited at page 82

Calafell, F., Grigorenko, E. L., Chikanian, A. A., and Kidd, K. K. (2001). Haplotype Evolution and Linkage Disequilibrium: A Simulation Study. *Human Heredity*, 51(1-2):85–96.
Cited at page 52

Campbell, C. L., Bhérer, C., Morrow, B. E., Boyko, A. R., and Auton, A. (2016). A Pedigree-Based Map of Recombination in the Domestic Dog Genome. *G3: Genes, Genomes, Genetics*, 6(11):3517–3524.
Cited at page 62

Cao, L., Alani, E., and Kleckner, N. (1990). A pathway for generation and processing of double-strand breaks during meiotic recombination in S. cerevisiae. *Cell*, 61(6):1089–1101.
Cited at page 37

Capilla, L., Caldés, M. G., and Ruiz-Herrera, A. (2016). Mammalian Meiotic Recombination: A Toolbox for Genome Evolution. *Cytogenetic and Genome Research*, 150(1):1–16.
Cited at pages 59, 65

Capilla, L., Medarde, N., Alemany-Schmidt, A., Oliver-Bonet, M., Ventura, J., and Ruiz-Herrera, A. (2014). Genetic recombination variation in wild Robertsonian mice: On the role of chromosomal fusions and Prdm9 allelic background. *Proceedings of the Royal Society B: Biological Sciences*, 281(1786). *Cited at page 68*

Capra, J. A., Hubisz, M. J., Kostka, D., Pollard, K. S., and Siepel, A. (2013). A model-based analysis of GC-biased gene conversion in the human and chimpanzee genomes. *PLoS genetics*, 9(8):e1003684.
Cited at pages 92, 94

Capra, J. A. and Pollard, K. S. (2011). Substitution Patterns Are GC-Biased in Divergent Sequences across the Metazoans. *Genome Biology and Evolution*, 3:516–527.
Cited at page 95

Carneiro, M., Afonso, S., Geraldes, A., Garreau, H., Bolet, G., Boucher, S., Tircazes, A., Queney, G., Nachman, M. W., and Ferrand, N. (2011). The Genetic Structure of Domestic Rabbits. *Molecular Biology and Evolution*, 28(6):1801–1816.
Cited at page 52

Carpenter, A. T. (1975). Electron microscopy of meiosis in Drosophila melanogaster females: II. The recombination nodule–a recombination-associated structure at pachytene? *Proc. Natl. Acad. Sci. U.S.A.*, 72(8):3186–3189. *Cited at page 34*

Carroll, L. (1871). *Through the Looking-Glass and What Alice Found There*. Macmillan.
Cited at page 74

Case, M. E. and Giles, N. H. (1964). Allelic recombination in *Neurospora*: tetrad analysis of a three-point cross within the *pan-2* locus. *Genetics*, 49(3):529.
Cited at page 17

Casselton, L. and Zolan, M. (2002). The art and design of genetic screens: filamentous fungi. *Nature Reviews Genetics*, 3(9):683. *Cited at page 14*

Castillo-Davis, C. I. and Hartl, D. L. (2002). Genome evolution and developmental constraint in Caenorhabditis elegans. *Molecular Biology and Evolution*, 19(5):728–735.
Cited at page 89

Castle, W. E. (1919a). Are Genes Linear or Non-Linear in Arrangement? *Proceedings of the National Academy of Sciences of the United States of America*, 5(11):500–506. *Cited at page 50*

Castle, W. E. (1919b). Is the Arrangement of the Genes in the Chromosome Linear? *Proceedings of the National Academy of Sciences of the United States of America*, 5(2):25–32. *Cited at page 50*

Cavalier-Smith, T. (2002). Origins of the machinery of recombination and sex. *Heredity*, 88(2):125. *Cited at page 25*

Chakravarti, A., Buetow, K. H., Antonarakis, S. E., Waber, P. G., Boehm, C. D., and Kazazian, H. H. (1984). Nonuniform recombination within the human beta-globin gene cluster. *American Journal of Human Genetics*, 36(6):1239–1258. *Cited at page 63*

Chakravarti, A., Elbein, S. C., and Permutt, M. A. (1986). Evidence for increased recombination near the human insulin gene: Implication for disease association studies. *Proceedings of the National Academy of Sciences of the United States of America*, 83(4):1045–1049. *Cited at page 63*

Chamary, J. V., Parmley, J. L., and Hurst, L. D. (2006). Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics*, 7(2):98. *Cited at page 90*

Champion, M. D. and Hawley, R. S. (2002). Playing for half the deck: The molecular biology of meiosis. *Nature Cell Biology*, 4 Suppl:s50–56. *Cited at page 35*

Chargaff, E. (1950). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, 6(6):201–209. *Cited at page 80*

Charlesworth, B. (2009). Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nature Reviews. Genetics*, 10(3):195–205. *Cited at pages , 103, 186, 187*

Charlesworth, B., Nordborg, M., and Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research*, 70(2):155–174. *Cited at page 52*

Chavancy, G., Chevallier, A., Fournier, A., and Garel, J. P. (1979). Adaptation of iso-tRNA concentration to mRNA codon frequency in the eukaryote cell. *Biochimie*, 61(1):71–78. *Cited at page 89*

Chen, C., Zhang, W., Timofejeva, L., Gerardin, Y., and Ma, H. (2005). The Arabidopsis ROCK-N-ROLLERS gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation. *The Plant Journal*, 43(3):321–334. *Cited at page 164*

Chen, S. Y., Tsubouchi, T., Rockmill, B., Sandler, J. S., Richards, D. R., Vader, G., Hochwagen, A., Roeder, G. S., and Fung, J. C. (2008). Global Analysis of the Meiotic Crossover Landscape. *Developmental Cell*, 15(3):401–415. *Cited at pages 58, 60*

Cheung, V. G., Burdick, J. T., Hirschmann, D., and Morley, M. (2007). Polymorphic Variation in Human Meiotic Recombination. *The American Journal of Human Genetics*, 80(3):526–530. *Cited at pages 51, 65*

Cheung, V. G., Sherman, S. L., and Feingold, E. (2010). Genetic Control of Hotspots. *Science*, 327(5967):791–792. *Cited at page 67*

Choi, K., Zhao, X., Tock, A. J., Lambing, C., Underwood, C. J., Hardcastle, T. J., Serra, H., Kim, J., Cho, H. S., Kim, J., Ziolkowski, P. A., Yelina, N. E., Hwang, I., Martienssen, R. A., and Henderson, I. R. (2018). Nucleosomes and DNA methylation shape meiotic DSB frequency in Arabidopsis thaliana transposons and gene regulatory regions. *Genome Research*, 28(4):532–546. *Cited at page 62*

Clay, O. and Bernardi, G. (2001). Compositional heterogeneity within and among isochores in mammalian genomes: II. Some general comments. *Gene*, 276(1):25–31. *Cited at page 85*

Clay, O., Cacciò, S., Zoubak, S., Mouchiroud, D., and Bernardi, G. (1996). Human coding and noncoding DNA: Compositional correlations. *Molecular Phylogenetics and Evolution*, 5(1):2–12. *Cited at page 85*

Clay, O., Carels, N., Douady, C., Macaya, G., and Bernardi, G. (2001). Compositional heterogeneity within and among isochores in mammalian genomes: I. CsCl and sequence analyses. *Gene*, 276(1):15–24. *Cited at page 85*

Clay, O. K. and Bernardi, G. (2011). GC3 of genes can be used as a proxy for isochore base composition: A reply to Elhaik et al. *Molecular Biology and Evolution*, 28(1):21–23. *Cited at page 90*

Clément, Y. and Arndt, P. F. (2013). Meiotic recombination strongly influences GC-content evolution in short regions in the mouse genome. *Molecular Biology and Evolution*, 30(12):2612–2618. *Cited at page 95*

Clément, Y., Sarah, G., Holtz, Y., Homa, F., Pointet, S., Contreras, S., Nabholz, B., Sabot, F., Sauné, L., Ardisson, M., Bacilieri, R., Besnard, G., Berger, A., Cardi, C., Bellis, F. D., Fouet, O., Jourda, C., Khadari, B., Lanaud, C., Leroy, T., Pot, D., Sauvage, C., Scarcelli, N., Tregear, J., Vigouroux, Y., Yahiaoui, N., Ruiz, M., Santoni, S., Labouisse, J.-P., Pham, J.-L., David, J., and Glémin, S. (2017). Evolutionary forces affecting synonymous variations in plant genomes. *PLOS Genetics*, 13(5):e1006799. *Cited at pages 90, 95*

Codina-Pascual, M., Campillo, M., Kraus, J., Speicher, M. R., Egozcue, J., Navarro, J., and Benet, J. (2006). Crossover frequency and synaptonemal complex length: Their variability and effects on human male meiosis. *Molecular Human Reproduction*, 12(2):123–133. *Cited at page 59*

Colaiácovo, M. P., MacQueen, A. J., Martinez-Perez, E., McDonald, K., Adamo, A., La Volpe, A., and Villeneuve, A. M. (2003). Synaptonemal complex assembly in c. elegans is dispensable for loading strand-exchange proteins but critical for proper completion of recombination. *Developmental cell*, 5(3):463–474. *Cited at page 32*

Cole, F., Baudat, F., Grey, C., Keeney, S., de Massy, B., and Jasin, M. (2014). Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics. *Nature Genetics*, 46(10):1072–1080. *Cited at pages 54, 126, 133, 135, 139, 184*

Cole, F., Kauppi, L., Lange, J., Roig, I., Wang, R., Keeney, S., and Jasin, M. (2012). Homeostatic control of recombination is implemented progressively in mouse meiosis. *Nature cell biology*, 14(4):424–430. *Cited at page 58*

Cole, F., Keeney, S., and Jasin, M. (2010a). Comprehensive, fine-scale dissection of homologous recombination outcomes at a hotspot in mouse meiosis. *Molecular cell*, 39(5):700–710. *Cited at pages 126, 139, 187, 217*

Cole, F., Keeney, S., and Jasin, M. (2010b). Evolutionary conservation of meiotic DSB proteins: More than just Spo11. *Genes & Development*, 24(12):1201–1207. *Cited at page 39*

Collins, H. (1985). *Changing Order: Replication and Induction in Scientific Practice*. Beverley Hills & London: Sage. *Cited at page 205*

Collins, H. M. (1975). The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics. *Sociology*, 9(2):205–224. *Cited at page 205*

Comeron, J. M. (2004). Selective and mutational patterns associated with gene expression in humans: Influences on synonymous composition and intron presence. *Genetics*, 167(3):1293–1304. *Cited at page 90*

Comeron, J. M., Ratnappan, R., and Bailin, S. (2012). The Many Landscapes of Recombination in Drosophila melanogaster. *PLOS Genetics*, 8(10):e1002905. *Cited at page 63*

Cook, P. (1997). The transcriptional basis of chromosome pairing. *Journal of cell science*, 110(9):1033–1040. *Cited at page 30*

Coop, G. and Myers, S. R. (2007). Live Hot, Die Young: Transmission Distortion in Recombination Hotspots. *PLoS Genetics*, 3(3). *Cited at page 73*

Coop, G. and Przeworski, M. (2007). An evolutionary view of human recombination. *Nature Reviews Genetics*, 8(1):23–34. *Cited at pages 56, 62, 88*

Coop, G., Wen, X., Ober, C., Pritchard, J. K., and Przeworski, M. (2008). High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans. *Science*, 319(5868):1395–1398. *Cited at pages 51, 53*

Corbin, L. J., Blott, S. C., Swinburne, J. E., Vaudin, M., Bishop, S. C., and Woolliams, J. A. (2010). Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics*, 41(s2):8–15. *Cited at page 52*

Corcoran, P. L., Moore, C. J., and Jazvac, K. (2014). An anthropogenic marker horizon in the future rock record. *GSA Today*, pages 4–8. *Cited at page 204*

Corneo, G., Ginelli, E., Soave, C., and Bernardi, G. (1968). Isolation and characterization of mouse and guinea pig satellite deoxyribonucleic acids. *Biochemistry*, 7(12):4373–4379. *Cited at page 81*

Cortadas, J., Macaya, G., and Bernardi, G. (1977). An Analysis of the Bovine Genome by Density Gradient Centrifugation: Fractionation in Cs2SO4/3,6-Bis(acetatomercurimethyl)dioxane Density Gradient. *European Journal of Biochemistry*, 76(1):13–19. *Cited at page 81*

Costantini, M. and Bernardi, G. (2008). Replication timing, chromosomal bands, and isochores. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9):3433–3437. *Cited at page 82*

Costantini, M., Clay, O., Auletta, F., and Bernardi, G. (2006). An isochore map of human chromosomes. *Genome Research*, 16(4):536–541. *Cited at page 81*

Costantini, M., Greif, G., Alvarez-Valin, F., and Bernardi, G. (2016). The Anolis Lizard Genome: An Amniote Genome without Isochores? *Genome Biology and Evolution*, 8(4):1048–1055. *Cited at page 83*

Couteau, F., Nabeshima, K., Villeneuve, A., and Zetka, M. (2004). A Component of C. elegans Meiotic Chromosome Axes at the Interface of Homolog Alignment, Synapsis, Nuclear Reorganization, and Recombination. *Current Biology*, 14(7):585–592. *Cited at page 58*

Cox, A., Ackert-Bicknell, C. L., Dumont, B. L., Ding, Y., Bell, J. T., Brockmann, G. A., Wergedal, J. E., Bult, C., Paigen, B., Flint, J., Tsaih, S.-W., Churchill, G. A., and Broman, K. W. (2009). A New Standard Genetic Map for the Laboratory Mouse. *Genetics*, 182(4):1335–1344. *Cited at pages 50, 64*

Crawford, D. C., Bhangale, T., Li, N., Hellenthal, G., Rieder, M. J., Nickerson, D. A., and Stephens, M. (2004). Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nature Genetics*, 36(7):700–706. *Cited at page 63*

Creighton, H. B. and McClintock, B. (1931). A correlation of cytological and genetical crossing-over in *Zea mays*. *Proceedings of the National Academy of Sciences of the United States of America*, 17(8):492. *Cited at page 13*

Crismani, W., Girard, C., Froger, N., Pradillo, M., Santos, J. L., Chelysheva, L., Copenhaver, G. P., Horlow, C., and Mercier, R. (2012). FANCM limits meiotic crossovers. *Science (New York, N.Y.)*, 336(6088):1588–1590. *Cited at page 58*

Cromie, G. A., Hyppa, R. W., Cam, H. P., Farah, J. A., Grewal, S. I. S., and Smith, G. R. (2007). A Discrete Class of Intergenic DNA Dictates Meiotic DNA Break Hotspots in Fission Yeast. *PLOS Genetics*, 3(8):e141. *Cited at pages 61, 63*

Cromie, G. A., Hyppa, R. W., Taylor, A. F., Zakharyevich, K., Hunter, N., and Smith, G. R. (2006). Single Holliday Junctions Are Intermediates of Meiotic Recombination. *Cell*, 127(6):1167–1178. *Cited at page 58*

Crutzen, P. J. (2002). Geology of mankind. *Nature*, 415(6867):23. *Cited at page 204*

Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., and François, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, 25(7):410–418. *Cited at page 136*

Csilléry, K., François, O., and Blum, M. G. B. (2012). Abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3(3):475–479. *Cited at page 139*

Cui, X. F., Li, H. H., Goradia, T. M., Lange, K., Kazazian, H. H., Galas, D., and Arnheim, N. (1989). Single-sperm typing: Determination of genetic distance between the G gamma-globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. *Proceedings of the National Academy of Sciences of the United States of America*, 86(23):9389–9393. *Cited at page 53*

Cullen, M., Perfetto, S. P., Klitz, W., Nelson, G., and Carrington, M. (2002). High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *American Journal of Human Genetics*, 71(4):759–776. *Cited at page 53*

Cunningham, W. (2003). Collective ownership of code and text. 'A Conversation with Ward Cunningham, Part II', Interview with Bill Venners. *Cited at page 163*

Cunningham, W. (2004). Geek noise. Talk at MS (Microsoft) Resarch. *Cited at page 107*

Cuny, G., Soriano, P., Macaya, G., and Bernardi, G. (1981). The Major Components of the Mouse and Human Genomes. *European Journal of Biochemistry*, 115(2):227–233. *Cited at page 81*

Cusack, B. P., Arndt, P. F., Duret, L., and Crollius, H. R. (2011). Preventing Dangerous Nonsense: Selection for Robustness to Transcriptional Error in Human Genes. *PLoS Genetics*, 7(10). *Cited at page 90*

Cutter, A. D. and Payseur, B. A. (2013). Genomic signatures of selection at linked sites: Unifying the disparity among species. *Nature Reviews Genetics*, 14(4):262–274. *Cited at page 61*

Damuth, J. (1981). Population density and body size in mammals. *Nature*, 290(5808):699. *Cited at page 99*

Daniel, K., Lange, J., Hached, K., Fu, J., Anastassiadis, K., Roig, I., Cooke, H. J., Stewart, A. F., Wassmann, K., Jasin, M., Keeney, S., and Tóth, A. (2011). Meiotic homologue alignment and its quality surveillance are controlled by mouse HORMAD1. *Nature Cell Biology*, 13(5):599–610. *Cited at page 39*

Danilowicz, C., Lee, C., Kim, K., Hatch, K., Coljee, V. W., Kleckner, N., and Prentiss, M. (2009). Single molecule detection of direct, homologous, dna/dna pairing. *Proceedings of the National Academy of Sciences*, 106(47):19824–19829. *Cited at page 30*

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, volume 220. *Cited at pages 3, 48*

Davies, B., Hatton, E., Altemose, N., Hussin, J. G., Pratto, F., Zhang, G., Hinch, A. G., Moralli, D., Biggs, D., Diaz, R., Preece, C., Li, R., Bitoun, E., Brick, K., Green, C. M., Camerini-Otero, R. D., Myers, S. R., and Donnelly, P. (2016). Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature*, 530(7589):171–176. *Cited at pages 70, 78, 130*

Davies, R. W. (2015). *Factors Influencing Genetic Variation in Wild Mice*. http://purl.org/dc/dcmitype/Text, University of Oxford. *Cited at pages , 103, 160*

Davis, R. C., Schadt, E. E., Smith, D. J., Hsieh, E. W. Y., Cervino, A. C. L., van Nas, A., Rosales, M., Doss, S., Meng, H., Allayee, H., and Lusis, A. J. (2005). A genome-wide set of congenic mouse strains derived from DBA/2J on a C57BL/6J background. *Genomics*, 86(3):259–270. *Cited at page 168*

de Boer, E., Dietrich, A. J. J., Höög, C., Stam, P., and Heyting, C. (2007). Meiotic interference among MLH1 foci requires neither an intact axial element structure nor full synapsis. *Cited at page 57*

de Boer, E., Stam, P., Dietrich, A. J. J., Pastink, A., and Heyting, C. (2006). Two levels of interference in mouse meiotic recombination. *Proceedings of the National Academy of Sciences*, 103(25):9607–9612. *Cited at page 57*

de Castro, E., Soriano, I., Marín, L., Serrano, R., Quintales, L., and Antequera, F. (2012). Nucleosomal organization of replication origins and meiotic recombination hotspots in fission yeast. *The EMBO journal*, 31(1):124–137. *Cited at page 62*

de los Santos, T., Hunter, N., Lee, C., Larkin, B., Loidl, J., and Hollingsworth, N. M. (2003). The Mus81/Mms4 Endonuclease Acts Independently of Double-Holliday Junction Resolution to Promote a Distinct Subset of Crossovers During Meiosis in Budding Yeast. *Genetics*, 164(1):81–94. *Cited at pages 44, 57*

De Maio, N., Schlötterer, C., and Kosiol, C. (2013). Linking Great Apes Genome Evolution across Time Scales Using Polymorphism-Aware Phylogenetic Models. *Molecular Biology and Evolution*, 30(10):2249–2262. *Cited at page 97*

de Massy, B. (2013). Initiation of meiotic recombination: How and where? Conservation and specificities among eukaryotes. *Annual Review of Genetics*, 47:563–599. *Cited at pages 39, 60, 63*

de Massy, B., Rocco, V., and Nicolas, A. (1995). The nucleotide mapping of DNA double-strand breaks at the CYS3 initiation site of meiotic recombination in Saccharomyces cerevisiae. *The EMBO journal*, 14(18):4589–4598. *Cited at page 39*

De Muyt, A., Jessop, L., Kolar, E., Sourirajan, A., Chen, J., Dayani, Y., and Lichten, M. (2012). BLM Helicase Ortholog Sgs1 Is a Central Regulator of Meiotic Recombination Intermediate Metabolism. *Molecular Cell*, 46(1):43–53. *Cited at page 45*

de Ricqlès, A. and Padian, K. (2009). Quelques apports à la théorie de l'Évolution, de la « Synthèse orthodoxe » à la « Super synthèse évo-dévo » 1970–2009 : Un point de vue. *Comptes Rendus Palevol*, 8(2):341–364. *Cited at page 202*

de Visser, J. A. G. and Elena, S. F. (2007). The evolution of sex: empirical insights into the roles of epistasis and drift. *Nature Reviews Genetics*, 8(2):139. *Cited at page 24*

de Vries, S. S., Baart, E. B., Dekker, M., Siezen, A., de Rooij, D. G., de Boer, P., and te Riele, H. (1999). Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes & Development*, 13(5):523–531. *Cited at page 42*

DePristo, M., Banks, E., Poplin, R., Garimella, K., Maguire, J., Hartl, C., Philippakis, A., del Angel, G., Rivas, M., Hanna, M., McKenna, A., Fennell, T., Kernytsky, A., Sivachenko, A., Cibulskis, K., Gabriel, S., Altshuler, D., and Daly, M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, 43(5):491–498. *Cited at page 116*

Dernburg, A. F., McDonald, K., Moulder, G., Barstead, R., Dresser, M., and Villeneuve, A. M. (1998). Meiotic recombination in c. elegans initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell*, 94(3):387–398. *Cited at page 34*

Diagouraga, B., Clément, J. A. J., Duret, L., Kadlec, J., de Massy, B., and Baudat, F. (2018). PRDM9 Methyltransferase Activity Is Essential for Meiotic DNA Double-Strand Break Formation at Its Binding Sites. *Molecular Cell*, 69(5):853–865.e6. *Cited at page 69*

Dietrich, W. F., Miller, J., Steen, R., Merchant, M. A., Damron-Boles, D., Husain, Z., Dredge, R., Daly, M. J., Ingalls, K. A., O'Connor, T. J., Evans, C. A., DeAngelis, M. M., Levinson, D. M., Kruglyak, L., Goodman, N., Copeland, N. G., Jenkins, N. A., Hawkins, T. L., Stein, L., Page, D. C., and Lander, E. S. (1996). A comprehensive genetic map of the mouse genome. *Nature*, 380(6570):149. *Cited at page 49*

Dobzhansky, T. (1937). Genetic Nature of Species Differences. *The American Naturalist*, 71(735):404–420. *Cited at page 201*

Dobzhansky, T. (1974). Chance and Creativity in Evolution. In Ayala, F. J. and Dobzhansky, T., editors, *Studies in the Philosophy of Biology: Reduction and Related Problems*, pages 307–338. Macmillan Education UK, London. *Cited at page 21*

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T. P., Bowden, D. W., Smith, D. R., and Lander, E. S. (1987). A genetic linkage map of the human genome. *Cell*, 51(2):319–337. *Cited at page 64*

Dooner, H. K. and He, L. (2008). Maize genome structure variation: Interplay between retrotransposon polymorphisms and genic recombination. *The Plant Cell*, 20(2):249–258. *Cited at page 61*

Dooner, H. K. and Martínez-Férez, I. M. (1997). Recombination occurs uniformly within the bronze gene, a meiotic recombination hotspot in the maize genome. *The Plant Cell*, 9(9):1633–1646. *Cited at pages 61, 63*

Dreszer, T. R., Wall, G. D., Haussler, D., and Pollard, K. S. (2007). Biased clustered substitutions in the human genome: The footprints of male-driven biased gene conversion. *Genome Research*, 17(10):1420–1430. *Cited at page 88*

Drouaud, J., Camilleri, C., Bourguignon, P.-Y., Canaguier, A., Bérard, A., Vezon, D., Giancola, S., Brunel, D., Colot, V., Prum, B., Quesneville, H., and Mézard, C. (2006). Variation in crossing-over rates across chromosome 4 of Arabidopsis thaliana reveals the presence of meiotic recombination "hot spots". *Genome Research*, 16(1):106–114. *Cited at page 63*

Drouaud, J., Mercier, R., Chelysheva, L., Bérard, A., Falque, M., Martin, O., Zanni, V., Brunel, D., and Mézard, C. (2007). Sex-Specific Crossover Distributions and Variations in Interference Level along Arabidopsis thaliana Chromosome 4. *PLoS Genetics*, 3(6). *Cited at pages 57, 64*

Dumas, D. and Britton-Davidian, J. (2002). Chromosomal rearrangements and evolution of recombination: Comparison of chiasma distribution patterns in standard and robertsonian populations of the house mouse. *Genetics*, 162(3):1355–1366. *Cited at page 55*

Dumont, B. L. (2017). Variation and Evolution of the Meiotic Requirement for Crossing Over in Mammals. *Genetics*, 205(1):155–168. *Cited at page 55*

Dumont, B. L. and Payseur, B. A. (2008). Evolution of the genomic rate of recombination in mammals. *Evolution; International Journal of Organic Evolution*, 62(2):276–294.   *Cited at pages 66, 188*

Dumont, B. L. and Payseur, B. A. (2011). Genetic Analysis of Genome-Scale Recombination Rate Evolution in House Mice. *PLOS Genetics*, 7(6):e1002116.
Cited at page 59

Dunn, L. C. and Bennett, D. (1967). Sex differences in recombination of linked genes in animals. *Genetics Research*, 9(2):211–220.   *Cited at page 64*

Dunn, P. M. (2003). Gregor Mendel, OSA (1822–1884), founder of scientific genetics. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 88(6):F537–F539.
Cited at page 9

Duret, L. (2002). Evolution of synonymous codon usage in metazoans. *Current Opinion in Genetics & Development*, 12(6):640–649.   *Cited at page 90*

Duret, L. (2006). The GC content of primates and rodents genomes is not at equilibrium: A reply to Antezana. *Journal of Molecular Evolution*, 62(6):803–806.
Cited at page 85

Duret, L. and Arndt, P. F. (2008). The Impact of Recombination on Nucleotide Substitutions in the Human Genome. *PLOS Genetics*, 4(5):e1000071.
Cited at pages 85, 87, 88

Duret, L., Cohen, J., Jubin, C., Dessen, P., Goût, J.-F., Mousset, S., Aury, J.-M., Jaillon, O., Noël, B., Arnaiz, O., Bétermier, M., Wincker, P., Meyer, E., and Sperling, L. (2008). Analysis of sequence variability in the macronuclear DNA of Paramecium tetraurelia: A somatic view of the germline. *Genome Research*, 18(4):585–596.   *Cited at page 87*

Duret, L., Eyre-Walker, A., and Galtier, N. (2006). A new perspective on isochore evolution. *Gene*, 385:71–74.   *Cited at page 85*

Duret, L. and Galtier, N. (2009a). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual Review of Genomics and Human Genetics*, 10:285–311.   *Cited at pages 82, 83, 87, 88, 90, 91, 92*

Duret, L. and Galtier, N. (2009b). Comment on "Human-Specific Gain of Function in a Developmental Enhancer". *Science*, 323(5915):714–714.   *Cited at page 91*

Duret, L. and Hurst, L. D. (2001). The elevated GC content at exonic third sites is not evidence against neutralist models of isochore evolution. *Molecular Biology and Evolution*, 18(5):757–762.   *Cited at page 85*

Duret, L. and Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8):4482–4487.   *Cited at page 89*

Duret, L., Mouchiroud, D., and Gautier, C. (1995). Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. *Journal of Molecular Evolution*, 40(3):308–317.   *Cited at page 81*

Duret, L., Semon, M., Piganeau, G., Mouchiroud, D., and Galtier, N. (2002). Vanishing GC-rich isochores in mammalian genomes. *Genetics*, 162(4):1837–1847.
Cited at page 85

Duroc, Y., Kumar, R., Ranjha, L., Adam, C., Guérois, R., Md Muntaz, K., Marsolier-Kergoat, M.-C., Dingli, F., Laureau, R., Loew, D., Llorente, B., Charbonnier, J.-B., Cejka, P., and Borde, V. (2017). Concerted action of the MutLβ heterodimer and Mer3 helicase regulates the global extent of meiotic gene conversion. *eLife*, 6.
Cited at pages 164, 173, 174, 177

Dutrillaux, B. (1986). [Role of chromosomes in evolution: a new interpretation]. *Annales De Genetique*, 29(2):69–75.
Cited at page 55

Dutta, R., Saha-Mandal, A., Cheng, X., Qiu, S., Serpen, J., Fedorova, L., and Fedorov, A. (2018). 1000 human genomes carry widespread signatures of GC biased gene conversion. *BMC Genomics*, 19(1):256.
Cited at page 95

Edelmann, W., Cohen, P. E., Kane, M., Lau, K., Morrow, B., Bennett, S., Umar, A., Kunkel, T., Cattoretti, G., Chaganti, R., Pollard, J. W., Kolodner, R. D., and Kucherlapati, R. (1996). Meiotic pachytene arrest in MLH1-deficient mice. *Cell*, 85(7):1125–1134.
Cited at page 44

Egel, R. (1995). The synaptonemal complex and the distribution of meiotic recombination events. *Trends in Genetics*, 11(6):206–208.
Cited at page 58

Egel-Mitani, M., Olson, L. W., and Egel, R. (1982). Meiosis in Aspergillus nidulans: another example for lacking synaptonemal complexes in the absence of crossover interference. *Hereditas*, 97(2):179–187.
Cited at page 34

Eijpe, M., Offenberg, H., Jessberger, R., Revenkova, E., and Heyting, C. (2003). Meiotic cohesin rec8 marks the axial elements of rat synaptonemal complexes before cohesins smc1β and smc3. *The Journal of cell biology*, 160(5):657–670.
Cited at page 32

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews. Genetics*, 5(6):435–445.
Cited at page 49

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., and Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491(7426):756–760.
Cited at page 120

Ellermeier, C., Higuchi, E. C., Phadnis, N., Holm, L., Geelhood, J. L., Thon, G., and Smith, G. R. (2010). RNAi and heterochromatin repress centromeric meiotic recombination. *Proceedings of the National Academy of Sciences*, 107(19):8701–8705.
Cited at page 60

Erlich, H. A., Bergström, T. F., Stoneking, M., and Gyllensten, U. (1996). HLA sequence polymorphism and the origin of humans. *Science (New York, N.Y.)*, 274(5292):1552–1554.
Cited at pages , 103

Escobar, J. S., Glémin, S., and Galtier, N. (2011). GC-Biased Gene Conversion Impacts Ribosomal DNA Evolution in Vertebrates, Angiosperms, and Other Eukaryotes. *Molecular Biology and Evolution*, 28(9):2561–2575. *Cited at pages , 95*

Evans, E. and Alani, E. (2000). Roles for Mismatch Repair Factors in Regulating Genetic Recombination. *Molecular and Cellular Biology*, 20(21):7839–7844. *Cited at page 83*

Ewing, B. and Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, 8(3):186–194. *Cited at page 117*

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-Calling of Automated Sequencer Traces UsingPhred. I. Accuracy Assessment. *Genome Research*, 8(3):175–185. *Cited at page 117*

Eyre-Walker, A. (1993). Recombination and mammalian genome evolution. *Proceedings. Biological Sciences*, 252(1335):237–243. *Cited at pages 60, 83*

Eyre-Walker, A. (1999). Evidence of selection on silent site base composition in mammals: Potential implications for the evolution of isochores and junk DNA. *Genetics*, 152(2):675–683. *Cited at pages 82, 83, 85, 89*

Eyre-Walker, A. and Hurst, L. D. (2001). The evolution of isochores. *Nature Reviews Genetics*, 2(7):549. *Cited at page 83*

Eyre-Walker, A. and Keightley, P. D. (1999). High genomic deleterious mutation rates in hominids. *Nature*, 397(6717):344–347. *Cited at page 83*

Fan, H. C., Wang, J., Potanina, A., and Quake, S. R. (2011). Whole-genome molecular haplotyping of single cells. *Nature Biotechnology*, 29(1):51–57. *Cited at page 53*

Fan, Q. Q., Xu, F., White, M. A., and Petes, T. D. (1997). Competition between Adjacent Meiotic Recombination Hotspots in the Yeast Saccharomyces Cerevisiae. *Genetics*, 145(3):661–670. *Cited at page 56*

Farnir, F., Coppieters, W., Arranz, J.-J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D., and Georges, M. (2000). Extensive Genome-wide Linkage Disequilibrium in Cattle. *Genome Research*, 10(2):220–227. *Cited at page 52*

Fawcett, D. W. (1956). The fine structure of chromosomes in the meiotic prophase of vertebrate spermatocytes. *J Biophys Biochem Cytol*, 2(4):403–406. *Cited at page 32*

Federico, C., Saccone, S., and Bernardi, G. (1998). The gene-richest bands of human chromosomes replicate at the onset of the S-phase. *Cytogenetics and Cell Genetics*, 80(1-4):83–88. *Cited at page 81*

Felsenstein, J. (1981). Skepticism Towards Santa Rosalia, or Why Are There so Few Kinds of Animals? *Evolution*, 35(1):124–138. *Cited at page 48*

Ferguson, D. O. and Holloman, W. K. (1996). Recombinational repair of gaps in DNA is asymmetric in Ustilago maydis and can be explained by a migrating D-loop model. *Proceedings of the National Academy of Sciences of the United States of America*, 93(11):5419–5424. *Cited at page 38*

Ferguson, K. A., Chow, V., and Ma, S. (2008). Silencing of unpaired meiotic chromosomes and altered recombination patterns in an azoospermic carrier of a t(8;13) reciprocal translocation. *Human Reproduction (Oxford, England)*, 23(4):988–995. *Cited at page 35*

Fernandez-Capetillo, O., Mahadevaiah, S. K., Celeste, A., Romanienko, P. J., Camerini-Otero, R. D., Bonner, W. M., Manova, K., Burgoyne, P., and Nussenzweig, A. (2003). H2ax is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis. *Developmental cell*, 4(4):497–508. *Cited at page 29*

Feynman, R. (1964). The character of physical law. *Cornell University Messenger lectures.* Lecture 7: Seeking New Laws. *Cited at page 125*

Feynman, R. P. (2006). *QED: The Strange Theory of Light and Matter.* Princeton University Press. *Cited at page 8*

Figuet, E. (2015). *Impact génomique des stratégies d'histoire de vie et reconstruction de traits ancestraux chez les amniotes.* PhD thesis. *Cited at page 99*

Figuet, E., Ballenghien, M., Romiguier, J., and Galtier, N. (2014). Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. *Genome Biology and Evolution*, 7(1):240–250. *Cited at pages , 95*

Figuet, E., Nabholz, B., Bonneau, M., Mas Carrio, E., Nadachowska-Brzyska, K., Ellegren, H., and Galtier, N. (2016). Life History Traits, Protein Evolution, and the Nearly Neutral Theory in Amniotes. *Molecular Biology and Evolution*, 33(6):1517–1527. *Cited at page 99*

Filipski, J. (1988). Why the rate of silent codon substitutions is variable within a vertebrate's genome. *Journal of Theoretical Biology*, 134(2):159–164. *Cited at page 82*

Filipski, J., Thiery, J.-P., and Bernardi, G. (1973). An analysis of the bovine genome by Cs2SO4—Ag+ density gradient centrifugation. *Journal of Molecular Biology*, 80(1):177–197. *Cited at page 81*

Fisher, R. A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, 52(2):399–433. *Cited at page 21*

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* The Clarendon Press. *Cited at page 21*

Fisher, R. A. (1935). The design of experiments. *Cited at page 145*

Fisher, R. A. (1950). Contributions to mathematical statistics. *Cited at page 217*

Fitch, W. M. (1976). Is there selection against wobble in codon-anticodon pairing? *Science*, 194(4270):1173–1174. *Cited at page 89*

Flachs, P., Mihola, O., Šimeček, P., Gregorová, S., Schimenti, J. C., Matsui, Y., Baudat, F., de Massy, B., Piálek, J., Forejt, J., and Trachtulec, Z. (2012). Interallelic and Intergenic Incompatibilities of the Prdm9 (Hst1) Gene in Mouse Hybrid Sterility. *PLoS Genetics*, 8(11). *Cited at page 71*

Flaubert, G. (1889). *Correspondance.* Paris : E. Fasquelle. *Cited at page 213*

Fledel-Alon, A., Leffler, E. M., Guan, Y., Stephens, M., Coop, G., and Przeworski, M. (2011). Variation in Human Recombination Rates and Its Genetic Determinants. *PLOS ONE*, 6(6):e20321. *Cited at page 71*

Fledel-Alon, A., Wilson, D. J., Broman, K., Wen, X., Ober, C., Coop, G., and Przeworski, M. (2009). Broad-Scale Recombination Patterns Underlying Proper Disjunction in Humans. *PLoS Genetics*, 5(9).   *Cited at page 35*

Flemming, W. (1879). Contributions to the knowledge of the cell and its life phenomena. *Arch. Mikr. Anat*, 16:302–406.   *Cited at page 9*

Fogel, S. and Mortimer, R. (1969). Informational transfer in meiotic gene conversion. *Proceedings of the National Academy of Sciences*, 62(1):96–103.   *Cited at page 17*

Forejt, J. and Iványi, P. (1974). Genetic studies on male sterility of hybrids between laboratory and wild mice (Mus musculus L.). *Genetics Research*, 24(2):189–206.   *Cited at page 71*

Forman, P. (1971). Weimar Culture, Causality, and Quantum Theory, 1918-1927: Adaptation by German Physicists and Mathematicians to a Hostile Intellectual Environment. *Historical Studies in the Physical Sciences*, 3:1–115.   *Cited at page 203*

Fox, E. J., Reid-Bayliss, K. S., Emond, M. J., and Loeb, L. A. (2014). Accuracy of Next Generation Sequencing Platforms. *Next Generation, Sequencing & Applications*, 1.   *Cited at page 121*

Francino, M. P. and Ochman, H. (1999). Isochores result from mutation not selection. *Nature*, 400(6739):30–31.   *Cited at page 82*

Franklin, R. E. and Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, 171(4356):740.   *Cited at pages 3, 19*

Frazer, K. A., Eskin, E., Kang, H. M., Bogue, M. A., Hinds, D. A., Beilharz, E. J., Gupta, R. V., Montgomery, J., Morenzoni, M. M., Nilsen, G. B., Pethiyagoda, C. L., Stuve, L. L., Johnson, F. M., Daly, M. J., Wade, C. M., and Cox, D. R. (2007). A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature*, 448(7157):1050–1053.   *Cited at page 160*

Friberg, U. and Rice, W. R. (2008). Cut thy neighbor: Cyclic birth and death of recombination hotspots via genetic conflict. *Genetics*, 179(4):2229–2238.   *Cited at page 74*

Froenicke, L., Anderson, L. K., Wienberg, J., and Ashley, T. (2002). Male Mouse Recombination Maps for Each Autosome Identified by Chromosome Painting. *The American Journal of Human Genetics*, 71(6):1353–1368.   *Cited at page 59*

Fromhage, L., McNamara, J. M., and Houston, A. I. (2007). Stability and value of male care for offspring: is it worth only half the trouble? *Biology Letters*, 3(3):234–236.   *Cited at page 24*

Fryxell, K. J. and Zuckerkandl, E. (2000). Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Molecular Biology and Evolution*, 17(9):1371–1383.   *Cited at page 82*

Fu, H., Zheng, Z., and Dooner, H. K. (2002). Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. *Proceedings of the National Academy of Sciences of the United States of America*, 99(2):1082–1087.   *Cited at page 63*

Fullerton, S. M., Bernardo Carvalho, A., and Clark, A. G. (2001). Local Rates of Recombination Are Positively Correlated with GC Content in the Human Genome. *Molecular Biology and Evolution*, 18(6):1139–1142. *Cited at pages 60, 82, 87*

Fung, J. C., Marshall, W. F., Dernburg, A., Agard, D. A., and Sedat, J. W. (1998). Homologous chromosome pairing in drosophila melanogaster proceeds through multiple independent initiations. *The Journal of cell biology*, 141(1):5–20. *Cited at page 31*

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., and Altshuler, D. (2002). The Structure of Haplotype Blocks in the Human Genome. *Science*, 296(5576):2225–2229. *Cited at page 66*

Galtier, N. (2003). Gene conversion drives GC content evolution in mammalian histones. *Trends in genetics: TIG*, 19(2):65–68. *Cited at page 86*

Galtier, N. (2004). Recombination, GC-content and the human pseudoautosomal boundary paradox. *Trends in genetics: TIG*, 20(8):347–349. *Cited at page 87*

Galtier, N., Bazin, E., and Bierne, N. (2006). GC-biased segregation of noncoding polymorphisms in Drosophila. *Genetics*, 172(1):221–228. *Cited at pages 90, 95*

Galtier, N. and Duret, L. (2007). Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends in genetics: TIG*, 23(6):273–277. *Cited at pages 86, 91*

Galtier, N., Duret, L., Glémin, S., and Ranwez, V. (2009). GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends in Genetics*, 25(1):1–5. *Cited at pages 92, 99, 191*

Galtier, N. and Mouchiroud, D. (1998). Isochore evolution in mammals: A human-like ancestral structure. *Genetics*, 150(4):1577–1584. *Cited at page 82*

Galtier, N., Piganeau, G., Mouchiroud, D., and Duret, L. (2001). GC-Content Evolution in Mammalian Genomes: The Biased Gene Conversion Hypothesis. *Genetics*, 159(2):907–911. *Cited at pages 83, 84, 86, 89*

Galtier, N., Roux, C., Rousselle, M., Romiguier, J., Figuet, E., Glémin, S., Bierne, N., and Duret, L. (2018). Codon Usage Bias in Animals: Disentangling the Effects of Natural Selection, Effective Population Size, and GC-Biased Gene Conversion. *Molecular Biology and Evolution.* *Cited at pages , 90, 95, 99, 103, 146, 191*

Garcia, V., Gray, S., Allison, R. M., Cooper, T. J., and Neale, M. J. (2015). Tel1(ATM)-mediated interference suppresses clustered meiotic double-strand-break formation. *Nature*, 520(7545):114–118. *Cited at page 56*

Garcia-Cruz, R., Pacheco, S., Brieño, M. A., Steinberg, E. R., Mudry, M. D., Ruiz-Herrera, A., and Garcia-Caldés, M. (2011). A comparative study of the recombination pattern in three species of Platyrrhini monkeys (primates). *Chromosoma*, 120(5):521–530. *Cited at pages 59, 66*

GATK team (2012). Base quality score recalibration (BQSR). https://gatkforums.broadinstitute.org/gatk/discussion/44/base-quality-score-recalibration-bqsr. *Cited at page 117*

Gayon, J. (2016). From Mendel to epigenetics: History of genetics. *Comptes rendus biologies*, 339(7-8):225–230. *Cited at pages 8, 12*

Gayon, J. and Petit, V. (2018). *La connaissance de la vie aujourd'hui.* ISTE Group. *Cited at page 201*

Geraldes, A., Basset, P., Gibson, B., Smith, K. L., Harr, B., Yu, H.-T., Bulatova, N., Ziv, Y., and Nachman, M. W. (2008). Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*, 17(24):5349–5363. *Cited at pages , 103, 168*

Gerton, J. L., DeRisi, J., Shroff, R., Lichten, M., Brown, P. O., and Petes, T. D. (2000). Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 97(21):11383–11390. *Cited at pages 54, 60, 87*

Gerton, J. L. and Hawley, R. S. (2005). Homologous chromosome interactions in meiosis: diversity amidst conservation. *Nature Reviews Genetics*, 6(6):477. *Cited at pages 30, 65*

Getun, I. V., Wu, Z. K., Khalil, A. M., and Bois, P. R. J. (2010). Nucleosome occupancy landscape and dynamics at mouse recombination hotspots. *EMBO reports*, 11(7):555–560. *Cited at pages 61, 217*

Ghabrial, A. and Schüpbach, T. (1999). Activation of a meiotic checkpoint regulates translation of Gurken during Drosophila oogenesis. *Nature Cell Biology*, 1(6):354–357. *Cited at page 34*

Gingras, Y. (2017). Les déterminants sociaux des connaissances scientifiques. *Que sais-je?*, 2e éd.:87–121. *Cited at pages 204, 205*

Giraut, L., Falque, M., Drouaud, J., Pereira, L., Martin, O. C., and Mézard, C. (2011). Genome-Wide Crossover Distribution in Arabidopsis thaliana Meiosis Reveals Sex-Specific Patterns along Chromosomes. *PLoS Genetics*, 7(11). *Cited at page 60*

Glémin, S. (2010). Surprising Fitness Consequences of GC-Biased Gene Conversion: I. Mutation Load and Inbreeding Depression. *Genetics*, 185(3):939–959. *Cited at page 92*

Glémin, S., Arndt, P. F., Messer, P. W., Petrov, D., Galtier, N., and Duret, L. (2015). Quantification of GC-biased gene conversion in the human genome. *Genome Research*, 25(8):1215–1228. *Cited at pages 94, 95, 97, 186, 187*

Glémin, S., Bazin, E., and Charlesworth, D. (2006). Impact of mating systems on patterns of sequence polymorphism in flowering plants. *Proceedings. Biological Sciences*, 273(1604):3011–3019. *Cited at page 87*

Glémin, S., Clément, Y., David, J., and Ressayre, A. (2014). GC content evolution in coding regions of angiosperm genomes: A unifying hypothesis. *Trends in Genetics*, 30(7):263–270. *Cited at page 95*

Globus, S. T. and Keeney, S. (2012). The joy of six: How to control your crossovers. *Cell*, 149(1):11–12. *Cited at page 58*

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., and Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, 27(2):182–189. *Cited at page 110*

Godin, B. and Gingras, Y. (2002). The experimenters' regress: From skepticism to argumentation. *Studies in History and Philosophy of Science Part A*, 33(1):133–148. *Cited at page 205*

Goldfarb, T. and Lichten, M. (2010). Frequent and efficient use of the sister chromatid for dna double-strand break repair during budding yeast meiosis. *PLoS biology*, 8(10):e1000520. *Cited at pages 33, 42, 58*

Golding, G. B. (1984). The Sampling Distribution of Linkage Disequilibrium. *Genetics*, 108(1):257–274. *Cited at page 51*

Golinski, J. (2008). *Making Natural Knowledge: Constructivism and the History of Science, with a New Preface.* University of Chicago Press. *Cited at page 203*

Gould, S. J. (1979). Of turtles, vets, elephants and castles. *New Scientist.* *Cited at page 208*

Gould, S. J. (1985). The median isn't the message. *Discover*, 6(6):40–42. *Cited at page 235*

Gould, S. J. (1989). *Wonderful Life: The Burgess Shale and the Nature of History.* WW Norton \& Company. *Cited at page 202*

Gould, S. J. and Eldredge, N. (1972). Punctuated equilibria: An alternative to phyletic gradualism. pages 82–115. *Cited at page 202*

Gouy, M. and Gautier, C. (1982). Codon usage in bacteria: Correlation with gene expressivity. *Nucleic Acids Research*, 10(22):7055–7074. *Cited at page 90*

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, 27(7):1017–1018. *Cited at page 150*

Gregorova, S., Gergelits, V., Chvatalova, I., Bhattacharyya, T., Valiskova, B., Fotopulosova, V., Jansa, P., Wiatrowska, D., and Forejt, J. (2018). Modulation of Prdm9-controlled meiotic chromosome asynapsis overrides hybrid sterility in mice. *eLife*, 7:e34282. *Cited at page 78*

Grey, C., Baudat, F., and de Massy, B. (2009). Genome-Wide Control of the Distribution of Meiotic Recombination. *PLOS Biology*, 7(2):e1000035. *Cited at page 68*

Grey, C., Baudat, F., and de Massy, B. (2018). PRDM9, a driver of the genetic map. *PLOS Genetics*, 14(8):e1007479. *Cited at pages 69, 72*

Grey, C., Clément, J. A. J., Buard, J., Leblanc, B., Gut, I., Gut, M., Duret, L., and de Massy, B. (2017). In vivo binding of PRDM9 reveals interactions with noncanonical genomic sites. *Genome Research*, 27(4):580–590. *Cited at page 68*

Griffiths, A. J., Wessler, S. R., Caroll, S. B., and John, D. (2015). *An introduction to genetic analysis.* W. H. Freeman and Company, 11 edition. *Cited at page 11*

Grin, I. and Ishchenko, A. A. (2016). An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. *Nucleic Acids Research*, 44(8):3713–3727. *Cited at page 96*

Groenen, M. A. M., Wahlberg, P., Foglio, M., Cheng, H. H., Megens, H.-J., Crooijmans, R. P. M. A., Besnier, F., Lathrop, M., Muir, W. M., Wong, G. K.-S., Gut, I., and Andersson, L. (2009). A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Research*, 19(3):510–519. *Cited at page 55*

Groeneveld, L. F., Atencia, R., Garriga, R. M., and Vigilant, L. (2012). High Diversity at PRDM9 in Chimpanzees and Bonobos. *PLoS ONE*, 7(7). *Cited at pages 68, 71*

Gruhn, J. R., Rubio, C., Broman, K. W., Hunt, P. A., and Hassold, T. (2013). Cytological Studies of Human Meiosis: Sex-Specific Differences in Recombination Originate at, or Prior to, Establishment of Double-Strand Breaks. *PLOS ONE*, 8(12):e85075. *Cited at page 59*

Guillon, H., Baudat, F., Grey, C., Liskay, R. M., and de Massy, B. (2005). Crossover and Noncrossover Pathways in Mouse Meiosis. *Molecular Cell*, 20(4):563–573. *Cited at page 54*

Guillon, H. and de Massy, B. (2002). An initiation site for meiotic crossing-over and gene conversion in the mouse. *Nature Genetics*, 32(2):296–299. *Cited at pages 53, 63, 217*

Guiraldelli, M. F., Eyster, C., Wilkerson, J. L., Dresser, M. E., and Pezza, R. J. (2013). Mouse HFM1/Mer3 is required for crossover formation and complete synapsis of homologous chromosomes during meiosis. *PLoS genetics*, 9(3):e1003383. *Cited at pages 44, 164*

Gutz, H. (1971). Site Specific Induction of Gene Conversion in SCHIZOSACCHA-ROMYCES POMBE. *Genetics*, 69(3):317–337. *Cited at page 73*

Haber, J. E. (1998). Meiosis: Avoiding inappropriate relationships. *Current biology*, 8(23):R832–R835. *Cited at page 34*

Haber, J. E. (2008). Evolution of Models of Homologous Recombination. In Egel, R. and Lankenau, D.-H., editors, *Recombination and Meiosis: Models, Means, and Evolution*, Genome Dynamics and Stability, pages 1–64. Springer Berlin Heidelberg, Berlin, Heidelberg. *Cited at page 20*

Haddrill, P. R. and Charlesworth, B. (2008). Non-neutral processes drive the nucleotide composition of non-coding sequences in Drosophila. *Biology Letters*, 4(4):438–441. *Cited at page 95*

Haldane, J. (1919). The combination of linkage values, and the calculation of distance between the loci of linked factors. *J Genet*, 8:299–309. *Cited at page 49*

Haldane, J. B. S. (1927). *A Mathematical Theory of Natural and Artificial Selection*, volume 23. *Cited at page 21*

Haldane, J. B. S. (1932). *The Causes of Evolution.* Princeton University Press. *Cited at pages 21, 47*

Haldane, J. B. S. (1963). The truth about death. Book Review of 'The Chester Beatty Research Institute Serially Abridged Life Tables, England and Wales, 1841-1960'. *Cited at page 181*

Halldorsson, B. V., Hardarson, M. T., Kehr, B., Styrkarsdottir, U., Gylfason, A., Thorleifsson, G., Zink, F., Jonasdottir, A., Jonasdottir, A., Sulem, P., Masson, G., Thorsteinsdottir, U., Helgason, A., Kong, A., Gudbjartsson, D. F., and Stefansson, K. (2016). The rate of meiotic gene conversion varies by sex and age. *Nature Genetics*, 48(11):1377–1384. *Cited at pages , 80, 96, 103, 120, 146, 159, 182, 188, 190, 191, 192, 193*

Halldorsson, B. V., Palsson, G., Stefansson, O. A., Jonsson, H., Hardarson, M. T., Eggertsson, H. P., Gunnarsson, B., Oddsson, A., Halldorsson, G. H., Zink, F., Gudjonsson, S. A., Frigge, M. L., Thorleifsson, G., Sigurdsson, A., Stacey, S. N., Sulem, P., Masson, G., Helgason, A., Gudbjartsson, D. F., Thorsteinsdottir, U., and Stefansson, K. (2019). Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science*, 363(6425):eaau1043. *Cited at pages 51, 61*

Hamada, H. and Kakunaga, T. (1982). Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature*, 298(5872):396–398. *Cited at page 49*

Hamada, H., Petrino, M. G., and Kakunaga, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 79(21):6465–6469. *Cited at page 49*

Hamada, K., Horiike, T., Ota, H., Mizuno, K., and Shinozawa, T. (2003). Presence of isochore structures in reptile genomes suggested by the relationship between GC contents of intron regions and those of coding regions. *Genes & Genetic Systems*, 78(2):195–198. *Cited at page 83*

Hamilton, W. (1975). Review of ghiselin (1974) and williams (1975). *Quarterly Review of Biology*, 50:175–179. *Cited at page 23*

Hamilton, W. D. (1996). *Narrow Roads of Gene Land: Volume 2: Evolution of Sex*, volume 2. Oxford University Press. *Cited at page 23*

Handel, M. A. (2004). The xy body: a specialized meiotic chromatin domain. *Experimental cell research*, 296(1):57–63. *Cited at page 29*

Handel, M. A. and Schimenti, J. C. (2010). Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. *Nature Reviews Genetics*, 11(2):124. *Cited at pages 27, 34, 188*

Harding, R. M., Fullerton, S. M., Griffiths, R. C., Bond, J., Cox, M. J., Schneider, J. A., Moulin, D. S., and Clegg, J. B. (1997). Archaic African and Asian lineages in the genetic ancestry of modern humans. *American Journal of Human Genetics*, 60(4):772–789. *Cited at pages , 103*

Harrison, R. J. and Charlesworth, B. (2011). Biased gene conversion affects patterns of codon usage and amino acid usage in the Saccharomyces sensu stricto group of yeasts. *Molecular Biology and Evolution*, 28(1):117–129. *Cited at page 90*

Hassold, T., Abruzzo, M., Adkins, K., Griffin, D., Merrill, M., Millie, E., Saker, D., Shen, J., and Zaragoza, M. (1996). Human aneuploidy: Incidence, origin, and etiology. *Environmental and Molecular Mutagenesis*, 28(3):167–175. *Cited at page 36*

Hassold, T., Hall, H., and Hunt, P. (2007). The origin of human aneuploidy: Where we have been, where we are going. *Human Molecular Genetics*, 16 Spec No. 2:R203–208. *Cited at page 35*

Hassold, T., Hansen, T., Hunt, P., and VandeVoort, C. (2009). Cytological studies of recombination in rhesus males. *Cytogenetic and Genome Research*, 124(2):132–138. *Cited at page 66*

Hassold, T. and Hunt, P. (2001). To err (meiotically) is human: The genesis of human aneuploidy. *Nature Reviews. Genetics*, 2(4):280–291. *Cited at pages 36, 60*

Hassold, T., Judis, L., Chan, E. R., Schwartz, S., Seftel, A., and Lynn, A. (2004). Cytological studies of meiotic recombination in human males. *Cytogenetic and Genome Research*, 107(3-4):249–255. *Cited at page 55*

Haudry, A., Cenci, A., Guilhaumon, C., Paux, E., Poirier, S., Santoni, S., David, J., and Glémin, S. (2008). Mating system and recombination affect molecular evolution in four Triticeae species. *Genetical Research*, 90(1):97–109. *Cited at page 92*

Hayashi, K. and Matsui, Y. (2006). Meisetz, a novel histone tri-methyltransferase, regulates meiosis-specific epigenesis. *Cell Cycle (Georgetown, Tex.)*, 5(6):615–620. *Cited at page 67*

Hayashi, K., Yoshida, K., and Matsui, Y. (2005). A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature*, 438(7066):374. *Cited at page 67*

Hayman, D. L., Moore, H. D. M., and Evans, E. P. (1988). Further evidence of novel sex differences in chiasma distribution in marsupials. *Heredity*, 61(3):455–458. *Cited at page 65*

Hayman, D. L. and Rodger, J. C. (1990). Meiosis in male and female *Trichosurus vulpecula* (Marsupialia). *Heredity*, 64(2):251–254. *Cited at page 65*

Hayman, D. L., Smith, M. J., and Rodger, J. C. (1990). A comparative study of chiasmata in male and femaleBettongia penicillata (Marsupialia). *Genetica*, 83(1):45–49. *Cited at page 65*

Hazen, R. M., Grew, E. S., Origlieri, M. J., and Downs, R. T. (2017). On the mineralogy of the "Anthropocene Epoch". *American Mineralogist*, 102(3):595–611. *Cited at page 204*

He, Z., Henricksen, L. A., Wold, M. S., and Ingles, C. J. (1995). RPA involvement in the damage-recognition and incision steps of nucleotide excision repair. *Nature*, 374(6522):566–569. *Cited at page 40*

Heerschop, S., Zischler, H., Merker, S., Perwitasari-Farajallah, D., and Driller, C. (2016). The pioneering role of PRDM9 indel mutations in tarsier evolution. *Scientific Reports*, 6. *Cited at pages 68, 71*

Helleday, T. (2003). Pathways for mitotic homologous recombination in mammalian cells. *Mutation Research*, 532(1-2):103–115. *Cited at page 46*

Hellenthal, G. and Stephens, M. (2006). Insights into recombination from population genetic variation. *Current Opinion in Genetics & Development*, 16(6):565–572. *Cited at page 53*

Hellmann, I., Ebersberger, I., Ptak, S. E., Pääbo, S., and Przeworski, M. (2003). A neutral explanation for the correlation of diversity with recombination rates in humans. *American Journal of Human Genetics*, 72(6):1527–1535. *Cited at page 61*

Hellmann, I., Prüfer, K., Ji, H., Zody, M. C., Pääbo, S., and Ptak, S. E. (2005). Why do human diversity levels vary at a megabase scale? *Genome Research*, 15(9):1222–1231. *Cited at page 61*

Hellsten, U., Wright, K. M., Jenkins, J., Shu, S., Yuan, Y., Wessler, S. R., Schmutz, J., Willis, J. H., and Rokhsar, D. S. (2013). Fine-scale variation in meiotic recombination in Mimulus inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 110(48):19478–19482. *Cited at page 62*

Henderson, K. A. and Keeney, S. (2004). Tying synaptonemal complex initiation to the formation and programmed repair of DNA double-strand breaks. *Proc. Natl. Acad. Sci. U.S.A.*, 101(13):4519–4524. *Cited at page 34*

Heng, H., Chamberlain, J. W., Shi, X.-M., Spyropoulos, B., Tsui, L.-C., and Moens, P. B. (1996). Regulation of meiotic chromatin loop size by chromosomal position. *Proceedings of the National Academy of Sciences*, 93(7):2795–2800. *Cited at page 32*

Hernandez, R. D., Williamson, S. H., Zhu, L., and Bustamante, C. D. (2007). Context-Dependent Mutation Rates May Cause Spurious Signatures of a Fixation Bias Favoring Higher GC-Content in Humans. *Molecular Biology and Evolution*, 24(10):2196–2202. *Cited at page 94*

Hershberg, R. and Petrov, D. A. (10 juil. 2009). General Rules for Optimal Codon Choice. *PLOS Genetics*, 5(7):e1000556. *Cited at page 92*

Hey, J. and Kliman, R. M. (2002). Interactions between natural selection, recombination and gene density in the genes of Drosophila. *Genetics*, 160(2):595–608. *Cited at page 89*

Hill, W. G. and Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research*, 8(3):269–294. *Cited at page 89*

Hill, W. G. and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, 38(6):226–231. *Cited at page 51*

Hillers, K. J. (2004). Crossover interference. *Current Biology*, 14(24):R1036–R1037. *Cited at page 57*

Hillers, K. J. and Villeneuve, A. M. (2003). Chromosome-wide control of meiotic crossing over in C. elegans. *Current biology: CB*, 13(18):1641–1647. *Cited at page 55*

Hillis, D. M., Sadava, D. E., Heller, H. C., and Price, M. V. (2012). *Principles of life.* Sinauer Associates, Inc. *Cited at page 26*

Hinch, A. G. (2013). *The Landscape of Recombination in African Americans: Leveraging Human Population Variation to Investigate Homologous Recombination.* http://purl.org/dc/dcmitype/Text, University of Oxford. *Cited at page 51*

Hinch, A. G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C. D., Chen, G. K., Wang, K., Buxbaum, S. G., Akylbekova, E. L., Aldrich, M. C., Ambrosone, C. B., Amos, C., Bandera, E. V., Berndt, S. I., Bernstein, L., Blot, W. J., Bock, C. H., Boerwinkle, E., Cai, Q., Caporaso, N., Casey, G., Adrienne Cupples, L., Deming, S. L., Ryan Diver, W., Divers, J., Fornage, M., Gillanders, E. M., Glessner, J., Harris, C. C., Hu, J. J., Ingles, S. A., Isaacs, W., John, E. M., Linda Kao, W. H., Keating, B., Kittles, R. A., Kolonel, L. N., Larkin, E., Le Marchand, L., McNeill, L. H., Millikan, R. C., Murphy, Musani, S., Neslund-Dudas, C., Nyante, S., Papanicolaou, G. J., Press, M. F., Psaty, B. M., Reiner, A. P., Rich, S. S., Rodriguez-Gil, J. L., Rotter, J. I., Rybicki, B. A., Schwartz, A. G., Signorello, L. B., Spitz, M., Strom, S. S., Thun, M. J., Tucker, M. A., Wang, Z., Wiencke, J. K., Witte, J. S., Wrensch, M., Wu, X., Yamamura, Y., Zanetti, K. A., Zheng, W., Ziegler, R. G., Zhu, X., Redline, S., Hirschhorn, J. N., Henderson, B. E., Taylor Jr, H. A., Price, A. L., Hakonarson, H., Chanock, S. J., Haiman, C. A., Wilson, J. G., Reich, D., and Myers, S. R. (2011). The landscape of recombination in African Americans. *Nature*, 476(7359):170–175. *Cited at pages 52, 66, 71*

Hochwagen, A. and Amon, A. (2006). Checking your breaks: Surveillance mechanisms of meiotic recombination. *Current biology: CB*, 16(6):R217–228. *Cited at page 35*

Hochwagen, A. and Marais, G. A. B. (2010). Meiosis: A PRDM9 Guide to the Hotspots of Recombination. *Current Biology*, 20(6):R271–R274. *Cited at page 67*

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., Middle, C. M., Rodesch, M. J., Albert, T. J., Hannon, G. J., and McCombie, W. R. (2007). Genome-wide *in situ* exon capture for selective resequencing. *Nature Genetics*, 39(12):1522–1527. *Cited at page 110*

Högstrand, K. and Böhme, J. (1999). Gene conversion of major histocompatibility complex genes is associated with CpG-rich regions. *Immunogenetics*, 49(5):446–455. *Cited at page 86*

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genetics Research*, 5(2):282–304. *Cited at pages 19, 36*

Holliday, R. (1968). Genetic recombination in fungi. *Replication and recombination of genetic material*, pages 157–174. *Cited at page 19*

Holliday, R. (2011). The recombination, repair and modification of dna. *DNA repair*, 10(10):993–999. *Cited at page 19*

Hollingsworth, N. M. and Brill, S. J. (2004). The Mus81 solution to resolution: Generating meiotic crossovers without Holliday junctions. *Genes & development*, 18(2):117–125. *Cited at page 58*

Holloway, J. K., Booth, J., Edelmann, W., McGowan, C. H., and Cohen, P. E. (2008). MUS81 generates a subset of MLH1-MLH3-independent crossovers in mammalian meiosis. *PLoS genetics*, 4(9):e1000186. *Cited at page 44*

Holloway, J. K., Morelli, M. A., Borst, P. L., and Cohen, P. E. (2010). Mammalian BLM helicase is critical for integrating multiple pathways of meiotic recombination. *The Journal of Cell Biology*, 188(6):779–789. *Cited at page 45*

Holmquist, G. P. (1992). Chromosome bands, their chromatin flavors, and their functional features. *American Journal of Human Genetics*, 51(1):17–37. *Cited at page 83*

Hong, S., Sung, Y., Yu, M., Lee, M., Kleckner, N., and Kim, K. P. (2013). The logic and mechanism of homologous recombination partner choice. *Molecular cell*, 51(4):440–453. *Cited at page 33*

Horn, S. (2012). Target Enrichment via DNA Hybridization Capture. In Shapiro, B. and Hofreiter, M., editors, *Ancient DNA: Methods and Protocols*, Methods in Molecular Biology, pages 177–188. Humana Press, Totowa, NJ. *Cited at page 110*

Horton, M. W., Hancock, A. M., Huang, Y. S., Toomajian, C., Atwell, S., Auton, A., Muliyati, N. W., Platt, A., Sperone, F. G., Vilhjálmsson, B. J., Nordborg, M., Borevitz, J. O., and Bergelson, J. (2012). Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. *Nature Genetics*, 44(2):212–216. *Cited at page 61*

Hubert, R., MacDonald, M., Gusella, J., and Arnheim, N. (1994). High resolution localization of recombination hot spots using sperm typing. *Nature Genetics*, 7(3):420–424. *Cited at page 63*

Hubisz, M. J. and Pollard, K. S. (2014). Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Current Opinion in Genetics & Development*, 29:15–21. *Cited at page 91*

Hudson, R. R. and Kaplan, N. L. (1985). Statistical Properties of the Number of Recombination Events in the History of a Sample of Dna Sequences. *Genetics*, 111(1):147–164. *Cited at page 52*

Hughes, S., Zelus, D., and Mouchiroud, D. (1999). Warm-blooded isochore structure in Nile crocodile and turtle. *Molecular Biology and Evolution*, 16(11):1521–1527. *Cited at pages 82, 83*

Humphryes, N. and Hochwagen, A. (2014). A non-sister act: recombination template choice during meiosis. *Experimental Cell Research*, 329(1):53–60. *Cited at page 33*

Hunt, P. A. and Hassold, T. J. (2002). Sex matters in meiosis. *Science (New York, N.Y.)*, 296(5576):2181–2183. *Cited at page 35*

Hunter, N. (2003). Synaptonemal complexities and commonalities. *Molecular cell*, 12(3):533–535. *Cited at page 34*

Hunter, N. (2015). Meiotic recombination: the essence of heredity. *Cold Spring Harbor perspectives in biology*, 7(12):a016618. *Cited at pages 24, 45*

Hunter, N. and Kleckner, N. (2001). The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell*, 106(1):59–70. *Cited at pages 34, 57*

Huxley, J. (1942). Evolution. The Modern Synthesis. *Evolution. The Modern Synthesis.* *Cited at pages 21, 201*

Huxley, J. S. (1928). Sexual difference of linkage inGammarus chevreuxi. *Journal of Genetics*, 20(2):145–156. *Cited at page 64*

Hwang, D. G. and Green, P. (2004). Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 101(39):13994–14001. *Cited at page 94*

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, 2(1):13–34. *Cited at page 90*

Imai, Y., Baudat, F., Taillepierre, M., Stanzione, M., Toth, A., and de Massy, B. (2017). The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. *Chromosoma*, 126(6):681–695. *Cited at page 69*

International Chicken Genome Sequencing Consortium (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432(7018):695–716. *Cited at pages 87, 99*

Ira, G., Malkova, A., Liberi, G., Foiani, M., and Haber, J. E. (2003). Srs2 and Sgs1-Top3 suppress crossovers during double-strand break repair in yeast. *Cell*, 115(4):401–411. *Cited at page 45*

Ishiguro, K.-i., Kim, J., Shibuya, H., Hernández-Hernández, A., Suzuki, A., Fukagawa, T., Shioi, G., Kiyonari, H., Li, X. C., Schimenti, J., Höög, C., and Watanabe, Y. (2014). Meiosis-specific cohesin mediates homolog recognition in mouse spermatocytes. *Genes & Development*, 28(6):594–607. *Cited at page 30*

IUCN (International Union for Conservation of Nature) (2019). The IUCN red list of threatened species. Online; accessed 25-March-2019. *Cited at page 24*

Jabbari, K., Wirtz, J., Rauscher, M., and Wiehe, T. (2019). A common genomic code for chromatin architecture and recombination landscape. *PLOS ONE*, 14(3):e0213278. *Cited at page 61*

Jackson, B. C., Campos, J. L., Haddrill, P. R., Charlesworth, B., and Zeng, K. (2017). Variation in the Intensity of Selection on Codon Bias over Time Causes Contrasting Patterns of Base Composition Evolution in Drosophila. *Genome Biology and Evolution*, 9(1):102–123. *Cited at page 90*

Jancek, S., Gourbière, S., Moreau, H., and Piganeau, G. (2008). Clues about the genetic basis of adaptation emerge from comparing the proteomes of two Ostreococcus ecotypes (Chlorophyta, Prasinophyceae). *Molecular Biology and Evolution*, 25(11):2293–2300. *Cited at page 87*

Janssens, F. A. (1909). *La théorie de la chiasmatypie: nouvelle interpretation des cinèses de maturation.* Van In. *Cited at pages 10, 12*

Jaramillo-Lambert, A. and Engebrecht, J. (2010). A Single Unpaired and Transcriptionally Silenced X Chromosome Locally Precludes Checkpoint Signaling in the Caenorhabditis elegans Germ Line. *Genetics*, 184(3):613–628. *Cited at page 35*

Jeffreys, A. J., Cotton, V. E., Neumann, R., and Lam, K.-W. G. (2013). Recombination regulator PRDM9 influences the instability of its own coding sequence in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2):600–605. *Cited at page 71*

Jeffreys, A. J., Kauppi, L., and Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics*, 29(2):217–222. *Cited at pages 53, 63, 64*

Jeffreys, A. J. and May, C. A. (2004). Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nature Genetics*, 36(2):151–156. *Cited at pages 45, 64, 188*

Jeffreys, A. J., Murray, J., and Neumann, R. (1998). High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Molecular Cell*, 2(2):267–273. *Cited at page 53*

Jeffreys, A. J. and Neumann, R. (2002). Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nature Genetics*, 31(3):267–271. *Cited at page 72*

Jeffreys, A. J. and Neumann, R. (2005). Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Human Molecular Genetics*, 14(15):2277–2287. *Cited at pages 66, 72*

Jeffreys, A. J. and Neumann, R. (2009). The rise and fall of a human recombination hot spot. *Nature genetics*, 41(5):625–629. *Cited at page 73*

Jeffreys, A. J., Neumann, R., Panayi, M., Myers, S., and Donnelly, P. (2005). Human recombination hot spots hidden in regions of strong marker association. *Nature Genetics*, 37(6):601–606. *Cited at page 66*

Jensen-Seaman, M. I., Furey, T. S., Payseur, B. A., Lu, Y., Roskin, K. M., Chen, C.-F., Thomas, M. A., Haussler, D., and Jacob, H. J. (2004). Comparative Recombination Rates in the Rat, Mouse, and Human Genomes. *Genome Research*, 14(4):528–538. *Cited at pages 60, 61*

Johnson, J., Bagley, J., Skaznik-Wikiel, M., Lee, H.-J., Adams, G. B., Niikura, Y., Tschudy, K. S., Tilly, J. C., Cortes, M. L., Forkert, R., et al. (2005). Oocyte generation in adult mammalian ovaries by putative germ cells in bone marrow and peripheral blood. *Cell*, 122(2):303–315. *Cited at page 28*

Johnson, J., Canning, J., Kaneko, T., Pru, J. K., and Tilly, J. L. (2004). Germline stem cells and follicular renewal in the postnatal mammalian ovary. *Nature*, 428(6979):145. *Cited at page 28*

Jones, G. H. (1967). The control of chiasma distribution in rye. *Chromosoma*, 22(1):69–90. *Cited at page 56*

Jones, G. H. (1974). Correlated components of chiasma variation and the control of chiasma distribution in rye. *Heredity*, 32(3):375–387. *Cited at page 56*

Jones, G. H. (1984). The control of chiasma distribution. *Symposia of the Society for Experimental Biology*, 38:293–320. *Cited at pages 55, 56*

Jones, G. H. and Franklin, F. C. H. (2006). Meiotic crossing-over: Obligation and interference. *Cell*, 126(2):246–248. *Cited at pages 55, 56*

Joshi, N., Barot, A., Jamison, C., and Börner, G. V. (2009). Pch2 links chromosome axis remodeling at future crossover sites and crossover distribution during yeast meiosis. *PLoS genetics*, 5(7):e1000557. *Cited at page 58*

Kaback, D. B., Guacci, V., Barber, D., and Mahon, J. W. (1992). Chromosome size-dependent control of meiotic recombination. *Science (New York, N.Y.)*, 256(5054):228–232. *Cited at page 99*

Kadi, F., Mouchiroud, D., Sabeur, G., and Bernardi, G. (1993). The compositional patterns of the avian genomes and their evolutionary implications. *Journal of Molecular Evolution*, 37(5):544–551. *Cited at page 82*

Kadyk, L. C. and Hartwell, L. H. (1992). Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. *Genetics*, 132(2):387–402. *Cited at pages 33, 58*

Kappes, S. M., Keele, J. W., Stone, R. T., McGraw, R. A., Sonstegard, T. S., Smith, T. P., Lopez-Corrales, N. L., and Beattie, C. W. (1997). A second-generation linkage map of the bovine genome. *Genome Research*, 7(3):235–249. *Cited at page 65*

Katzman, S., Capra, J. A., Haussler, D., and Pollard, K. S. (2011). Ongoing GC-Biased Evolution Is Widespread in the Human Genome and Enriched Near Recombination Hot Spots. *Genome Biology and Evolution*, 3:614–626. *Cited at page 95*

Katzman, S., Kern, A. D., Pollard, K. S., Salama, S. R., and Haussler, D. (2010). GC-Biased Evolution Near Human Accelerated Regions. *PLOS Genetics*, 6(5):e1000960. *Cited at page 91*

Kauppi, L., Barchi, M., Baudat, F., Romanienko, P. J., Keeney, S., and Jasin, M. (2011). Distinct Properties of the XY Pseudoautosomal Region Crucial for Male Meiosis. *Science*, 331(6019):916–920. *Cited at page 55*

Kauppi, L., Barchi, M., Lange, J., Baudat, F., Jasin, M., and Keeney, S. (2013). Numerical constraints and feedback control of double-strand breaks in mouse meiosis. *Genes & Development*, 27(8):873–886. *Cited at page 39*

Kauppi, L., Jasin, M., and Keeney, S. (2007). Meiotic crossover hotspots contained in haplotype block boundaries of the mouse genome. *Proceedings of the National Academy of Sciences of the United States of America*, 104(33):13396–13401. *Cited at page 63*

Kauppi, L., Jeffreys, A. J., and Keeney, S. (2004). Where the crossovers are: Recombination distributions in mammals. *Nature Reviews Genetics*, 5(6):413. *Cited at page 64*

Kauppi, L., Stumpf, M. P. H., and Jeffreys, A. J. (2005). Localized breakdown in linkage disequilibrium does not always predict sperm crossover hot spots in the human MHC class II region. *Genomics*, 86(1):13–24. *Cited at page 66*

Kaur, T. and Rockman, M. V. (2014). Crossover Heterogeneity in the Absence of Hotspots in Caenorhabditis elegans. *Genetics*, 196(1):137–148. *Cited at page 63*

Kawakami, T., Mugal, C. F., Suh, A., Nater, A., Burri, R., Smeds, L., and Ellegren, H. (2017). Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Molecular Ecology*, 26(16):4158–4172. *Cited at pages 120, 182*

Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C. F., Olason, P., and Ellegren, H. (2014). A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology*, 23(16):4035–4058. *Cited at pages 120, 182*

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., Furlotte, N. A., Eskin, E., Nellåker, C., Whitley, H., Cleak, J., Janowitz, D., Hernandez-Pliego, P., Edwards, A., Belgard, T. G., Oliver, P. L., McIntyre, R. E., Bhomra, A., Nicod, J., Gan, X., Yuan, W., van der Weyden, L., Steward, C. A., Balasubramaniam, S., Stalker, J., Mott, R., Durbin, R., Jackson, I. J., Czechanski, A., Assunção, J. A. G., Donahue, L. R., Reinholdt, L. G., Payseur, B. A., Ponting, C. P., Birney, E., Flint, J., and Adams, D. J. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, 477(7364):289–294. *Cited at pages , 108, 116, 168*

Keeney, S. (2008). Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome dynamics and stability*, 2:81–123. *Cited at page 39*

Keeney, S., Giroux, C. N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, 88(3):375–384. *Cited at page 39*

Keightley, P. D. and Eyre-Walker, A. (2000). Deleterious Mutations and the Evolution of Sex. *Science*, 290(5490):331–333. *Cited at page 83*

Keith, N., Tucker, A. E., Jackson, C. E., Sung, W., Lucas Lledó, J. I., Schrider, D. R., Schaack, S., Dudycha, J. L., Ackerman, M., Younge, A. J., Shaw, J. R., and Lynch, M. (2016). High mutational rates of large-scale duplication and deletion in Daphnia pulex. *Genome Research*, 26(1):60–69. *Cited at pages , 95, 103*

Kell, D. B. and Oliver, S. G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays*, 26(1):99–105. *Cited at page 205*

Kelmenson, P. M., Petkov, P., Wang, X., Higgins, D. C., Paigen, B. J., and Paigen, K. (2005). A Torrid Zone on Mouse Chromosome 1 Containing a Cluster of Recombinational Hotspots. *Genetics*, 169(2):833–841. *Cited at page 217*

Kemp, B., Boumil, R. M., Stewart, M. N., and Dawson, D. S. (2004). A role for centromere pairing in meiotic chromosome segregation. *Genes & development*, 18(16):1946–1951. *Cited at page 30*

Kent, C. F., Minaei, S., Harpur, B. A., and Zayed, A. (2012). Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *Proceedings of the National Academy of Sciences*, 109(44):18012–18017. *Cited at page 95*

Khelifi, A., Meunier, J., Duret, L., and Mouchiroud, D. (2006). GC Content Evolution of the Human and Mouse Genomes: Insights from the Study of Processed Pseudogenes in Regions of Different Recombination Rates. *Journal of Molecular Evolution*, 62(6):745–752. *Cited at page 87*

Khil, P. P., Smagulova, F., Brick, K. M., Camerini-Otero, R. D., and Petukhova, G. V. (2012). Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. *Genome Research*, 22(5):957–965. *Cited at pages 54, 126*

Kim, K. P., Weiner, B. M., Zhang, L., Jordan, A., Dekker, J., and Kleckner, N. (2010). Sister cohesion and structural axis components mediate homolog bias of meiotic recombination. *Cell*, 143(6):924–937. *Cited at pages 33, 58*

Kim, S., Plagnol, V., Hu, T. T., Toomajian, C., Clark, R. M., Ossowski, S., Ecker, J. R., Weigel, D., and Nordborg, M. (2007). Recombination and linkage disequilibrium in Arabidopsis thaliana. *Nature Genetics*, 39(9):1151–1155. *Cited at page 61*

Kimura, M. (1962). On the Probability of Fixation of Mutant Genes in a Population. *Genetics*, 47(6):713–719. *Cited at page 97*

Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature*, 217(5129):624. *Cited at pages 22, 79*

Kimura, M. (1991). The neutral theory of molecular evolution: A review of recent evidence. *The Japanese Journal of Genetics*, 66(4):367–386. *Cited at page 22*

Kimura, M., Clarke, B. C., Robertson, A., and Jeffreys, A. J. (1986). DNA and the neutral theory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 312(1154):343–354. *Cited at page 22*

King, J. S. and Mortimer, R. K. (1990). A polymerization model of chiasma interference and corresponding computer simulation. *Genetics*, 126(4):1127–1138. *Cited at page 57*

Kircher, M., Stenzel, U., and Kelso, J. (2009). Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biology*, 10(8):R83. *Cited at page 118*

Kitani, Y., Olive, L., and El-Ani, A. S. (1962). Genetics of *Sordaria fimicola*. v. aberrant segregation at the *g* locus. *American Journal of Botany*, 49(7):697–706. *Cited at page 17*

Kleckner, N., Zickler, D., Jones, G. H., Dekker, J., Padmore, R., Henle, J., and Hutchinson, J. (2004). A mechanical basis for chromosome function. *Proceedings of the National Academy of Sciences*, 101(34):12592–12597. *Cited at page 57*

Klein, F., Mahr, P., Galova, M., Buonomo, S. B., Michaelis, C., Nairz, K., and Nasmyth, K. (1999). A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. *Cell*, 98(1):91–103. *Cited at page 25*

Kliman, R. M. and Hey, J. (1993). Reduced natural selection associated with low recombination in Drosophila melanogaster. *Molecular Biology and Evolution*, 10(6):1239–1258. *Cited at page 90*

Kneitz, B., Cohen, P. E., Avdievich, E., Zhu, L., Kane, M. F., Hou, H., Kolodner, R. D., Kucherlapati, R., Pollard, J. W., and Edelmann, W. (2000). MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes & Development*, 14(9):1085–1097. *Cited at page 42*

Koehler, K. E., Cherry, J. P., Lynn, A., Hunt, P. A., and Hassold, T. J. (2002). Genetic control of mammalian meiotic recombination. I. Variation in exchange frequencies among males from inbred mouse strains. *Genetics*, 162(1):297–306. *Cited at page 66*

Koehler, K. E. and Hassold, T. J. (1998). Human aneuploidy: Lessons from achiasmate segregation in Drosophila melanogaster. *Annals of Human Genetics*, 62(6):467–479. *Cited at page 55*

Koehler, K. E., Hawley, R. S., Sherman, S., and Hassold, T. (1996). Recombination and nondisjunction in humans and flies. *Human Molecular Genetics*, 5(Supplement_1):1495–1504. *Cited at page 60*

Koh-Stenta, X., Poulsen, A., Li, R., Wee, J. L. K., Kwek, P. Z., Chew, S. Y., Peng, J., Wu, L., Guccione, E., Joy, J., and Hill, J. (2017). Discovery and characterisation of the automethylation properties of PRDM9. *Biochemical Journal*, 474(6):971–982. *Cited at page 69*

Kohl, K. P. and Sekelsky, J. (2013). Meiotic and Mitotic Recombination in Meiosis. *Genetics*, 194(2):327–334. *Cited at page 57*

Kondrashov, A. (1993). Classification of hypotheses on the advantage of amphimixis. *Journal of Heredity*, 84(5):372–387. *Cited at page 24*

Kong, A., Gudbjartsson, D. F., Shlien, A., Palsson, S. T., Frigge, M. L., Thorgeirsson, T. E., Gulcher, J. R., Stefansson, K., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardsson, B., Sigurbardottir, S., Barnard, J., Hallbeck, B., and Masson, G. (2002). A high-resolution recombination map of the human genome. *A high-resolution recombination map of the human genome*, 31(3):241–247. *Cited at pages 50, 60, 61, 82, 188*

Kong, A., Thorleifsson, G., Gudbjartsson, D. F., Masson, G., Sigurdsson, A., Jonasdottir, A., Walters, G. B., Jonasdottir, A., Gylfason, A., Kristinsson, K. T., Gudjonsson, S. A., Frigge, M. L., Helgason, A., Thorsteinsdottir, U., and Stefansson, K. (2010). Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*, 467(7319):1099–1103. *Cited at pages 50, 51*

Kono, H., Tamura, M., Osada, N., Suzuki, H., Abe, K., Moriwaki, K., Ohta, K., and Shiroishi, T. (2014). Prdm9 Polymorphism Unveils Mouse Evolutionary Tracks. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 21(3):315–326. *Cited at pages 68, 71*

Kosambi, D. D. (1943). The Estimation of Map Distances from Recombination Values. *Annals of Eugenics*, 12(1):172–175. *Cited at page 49*

Kostka, D., Hubisz, M. J., Siepel, A., and Pollard, K. S. (2012). The Role of GC-Biased Gene Conversion in Shaping the Fastest Evolving Regions of the Human Genome. *Molecular Biology and Evolution*, 29(3):1047–1057. *Cited at page 91*

Koszul, R. and Kleckner, N. (2009). Dynamic chromosome movements during meiosis: a way to eliminate unwanted connections? *Trends in cell biology*, 19(12):716–724. *Cited at page 31*

Koszul, R., Meselson, M., Van Doninck, K., Vandenhaute, J., and Zickler, D. (2012). The centenary of Janssens's chiasmatype theory. *Genetics*, 191(2):309–317. *Cited at page 10*

Kresge, N., Simoni, R. D., and Hill, R. L. (2005). Chargaff's Rules: The Work of Erwin Chargaff. *Journal of Biological Chemistry*, 280(24):e21–e21. *Cited at page 80*

Kudla, G., Helwak, A., and Lipinski, L. (2004). Gene conversion and GC-content evolution in mammalian Hsp70. *Molecular Biology and Evolution*, 21(7):1438–1444. *Cited at page 86*

Kugou, K., Fukuda, T., Yamada, S., Ito, M., Sasanuma, H., Mori, S., Katou, Y., Itoh, T., Matsumoto, K., Shibata, T., Shirahige, K., and Ohta, K. (2009). Rec8 Guides Canonical Spo11 Distribution along Yeast Meiotic Chromosomes. *Molecular Biology of the Cell*, 20(13):3064–3076. *Cited at page 60*

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions: 50th Anniversary Edition.* University of Chicago Press. *Cited at page 203*

Kumar, R., Bourbon, H.-M., and de Massy, B. (2010). Functional conservation of Mei4 for meiotic DNA double-strand break formation from yeasts to mice. *Genes & Development*, 24(12):1266–1280. *Cited at page 39*

Kumar, R., Ghyselinck, N., Ishiguro, K.-i., Watanabe, Y., Kouznetsova, A., Höög, C., Strong, E., Schimenti, J., Daniel, K., Toth, A., and de Massy, B. (2015). MEI4 – a central player in the regulation of meiotic DNA double-strand break formation in the mouse. *Journal of Cell Science*, 128(9):1800–1811. *Cited at page 39*

Kuraku, S., Ishijima, J., Nishida-Umehara, C., Agata, K., Kuratani, S., and Matsuda, Y. (2006). cDNA-based gene mapping and GC3 profiling in the soft-shelled turtle suggest a chromosomal size-dependent GC bias shared by sauropsids. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 14(2):187–202. *Cited at page 87*

Lachance, J. and Tishkoff, S. A. (2014). Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. *American Journal of Human Genetics*, 95(4):408–420. *Cited at pages 92, 99*

Laehnemann, D., Borkhardt, A., and McHardy, A. C. (2016). Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. *Briefings in Bioinformatics*, 17(1):154–179. *Cited at page 118*

Lafay, B. and Sharp, P. M. (1999). Synonymous codon usage variation among Giardia lamblia genes and isolates. *Molecular Biology and Evolution*, 16(11):1484–1495. *Cited at page 90*

Lahn, B. T. and Page, D. C. (1999). Four evolutionary strata on the human X chromosome. *Science (New York, N.Y.)*, 286(5441):964–967. *Cited at page 87*

Lam, I. and Keeney, S. (2015). Mechanism and Regulation of Meiotic Recombination Initiation. *Cold Spring Harbor Perspectives in Biology*, 7(1). *Cited at pages 40, 60*

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C.,

Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., Szustakowki, J., and International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921. *Cited at pages 81, 82, 99*

Lange, J., Pan, J., Cole, F., Thelen, M. P., Jasin, M., and Keeney, S. (2011). ATM controls meiotic double-strand-break formation. *Nature*, 479(7372):237–240. *Cited at page 39*

Lange, J., Yamada, S., Tischfield, S. E., Pan, J., Kim, S., Zhu, X., Socci, N. D., Jasin, M., and Keeney, S. (2016). The landscape of mouse meiotic double-strand break formation, processing and repair. *Cell*, 167(3):695–708.e16. *Cited at pages 54, 61, 126, 130, 133, 189*

Lao, J. P. and Hunter, N. (2010). Trying to avoid your sister. *PLoS biology*, 8(10):e1000519. *Cited at page 32*

Lartillot, N. (2013a). Interaction between Selection and Biased Gene Conversion in Mammalian Protein-Coding Sequence Evolution Revealed by a Phylogenetic Covariance Analysis. *Molecular Biology and Evolution*, 30(2):356–368. *Cited at page 99*

Lartillot, N. (2013b). Phylogenetic patterns of GC-biased gene conversion in placental mammals and the evolutionary dynamics of recombination landscapes. *Molecular Biology and Evolution*, 30(3):489–502. *Cited at pages , 94, 95, 97, 98, 103*

Lassalle, F., Périan, S., Bataillon, T., Nesme, X., Duret, L., and Daubin, V. (2015). GC-Content Evolution in Bacterial Genomes: The Biased Gene Conversion Hypothesis Expands. *PLOS Genetics*, 11(2):e1004941. *Cited at page 95*

Latrille, T., Duret, L., and Lartillot, N. (2017). The Red Queen model of recombination hot-spot evolution: A theoretical investigation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1736). *Cited at page 75*

Laurie, D. A. and Hultén, M. A. (1985). Further studies on chiasma distribution and interference in the human male. *Annals of Human Genetics*, 49(3):203–214. *Cited at page 57*

Lavner, Y. and Kotlar, D. (2005). Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene*, 345(1):127–138. *Cited at page 90*

Lawrie, N. M., Tease, C., and Hultén, M. A. (1995). Chiasma frequency, distribution and interference maps of mouse autosomes. *Chromosoma*, 104(4):308–314. *Cited at page 57*

Lecointre, G. (2009). Après la théorie synthétique, quelle biologie ? *Textes et Documents pour la Classe*, (981):16–19. *Cited at page 204*

Lederberg, J. (1955). Recombination mechanisms in bacteria. *Journal of Cellular and Comparative Physiology*, 45(S2):75–107. *Cited at page 19*

Lehtonen, J., Jennions, M. D., and Kokko, H. (2012). The many costs of sex. *Trends in ecology & evolution*, 27(3):172–178. *Cited at page 24*

Lenormand, T. and Dutheil, J. (2005). Recombination difference between sexes: A role for haploid selection. *PLoS biology*, 3(3):e63. *Cited at page 65*

Lenormand, T., Engelstadter, J., Johnston, S. E., Wijnker, E., and Haag, C. R. (2016). Evolutionary mysteries in meiosis. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 371(1706). *Cited at page 25*

Lercher, M. J. and Hurst, L. D. (2002). Human SNP variability and mutation rate are higher in regions of high recombination. *Trends in Genetics*, 18(7):337–340. *Cited at page 61*

Lercher, M. J. and Hurst, L. D. (2003). Imprinted chromosomal regions of the human genome have unusually high recombination rates. *Genetics*, 165(3):1629–1632. *Cited at page 65*

Lercher, M. J., Smith, N. G. C., Eyre-Walker, A., and Hurst, L. D. (2002). The Evolution of Isochores: Evidence From SNP Frequency Distributions. *Genetics*, 162(4):1805–1810. *Cited at page 82*

Lesecque, Y. (2014). *La conversion génique biaisée : origine, dynamique et intensité de la quatrième force d'évolution des génomes eucaryotes.* PhD thesis, Université Claude Bernard - Lyon I. *Cited at pages 72, 92*

Lesecque, Y., Glémin, S., Lartillot, N., Mouchiroud, D., and Duret, L. (2014). The Red Queen Model of Recombination Hotspots Evolution in the Light of Archaic and Modern Human Genomes. *PLOS Genetics*, 10(11):e1004790. *Cited at pages 71, 77*

Lesecque, Y., Mouchiroud, D., and Duret, L. (2013). GC-Biased Gene Conversion in Yeast Is Specifically Associated with Crossovers: Molecular Mechanisms and Evolutionary Significance. *Molecular Biology and Evolution*, 30(6):1409–1419. *Cited at pages 96, 99*

Levi, P. (1975). *Le système périodique.* Albin Michel. *Cited at page 5*

Lewin, R. (1996). Patterns in evolution: The new molecular view. *Patterns in evolution: the new molecular view.* *Cited at page 22*

Lewis, R. (1960). *The Evolution Man: Or, How I Ate My Father.* Pantheon. *Cited at page 103*

Lewontin, R. C. (1964). The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics*, 49(1):49–67. *Cited at page 51*

Lewontin, R. C. and Kojima, K.-i. (1960). The Evolutionary Dynamics of Complex Polymorphisms,,. *Evolution*, 14(4):458–472. *Cited at page 51*

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio].* *Cited at page 113*

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14):1754–1760. *Cited at page 113*

Li, H., Gyllensten, U. B., Cui, X., Saiki, R. K., Erlich, H. A., and Arnheim, N. (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature*, 335(6189):414. *Cited at page 53*

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079. *Cited at page 114*

Li, J., Hooker, G. W., and Roeder, G. S. (2006). Saccharomyces cerevisiae Mer2, Mei4 and Rec114 Form a Complex Required for Meiotic Double-Strand Break Formation. *Genetics*, 173(4):1969–1981. *Cited at page 39*

Li, R., Bitoun, E., Altemose, N., Davies, R. W., Davies, B., and Myers, S. R. (2018). A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination. *bioRxiv*, page 428987. *Cited at pages , 119, 130, 159, 182, 183*

Li, X. C., Li, X., and Schimenti, J. C. (2007). Mouse pachytene checkpoint 2 (trip13) is required for completing meiotic recombination but not synapsis. *PLoS genetics*, 3(8):e130. *Cited at page 35*

Libby, B. J., De La Fuente, R., O'Brien, M. J., Wigglesworth, K., Cobb, J., Inselman, A., Eaker, S., Handel, M. A., Eppig, J. J., and Schimenti, J. C. (2002). The mouse meiotic mutation mei1 disrupts chromosome synapsis with sexually dimorphic consequences for meiotic progression. *Developmental Biology*, 242(2):174–187. *Cited at page 39*

Libby, B. J., Reinholdt, L. G., and Schimenti, J. C. (2003). Positional cloning and characterization of Mei1, a vertebrate-specific gene required for normal meiotic chromosome synapsis in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26):15706–15711. *Cited at page 39*

Lichten, M. and Goldman, A. S. H. (1995). Meiotic Recombination Hotspots. *Annual Review of Genetics*, 29(1):423–444.   *Cited at page 63*

Lien, S., Kamiński, S., Aleström, P., and Rogne, S. (1993). A simple and powerful method for linkage analysis by amplification of DNA from single sperm cells. *Genomics*, 16(1):41–44.   *Cited at page 53*

Lien, S., Szyda, J., Schechinger, B., Rappold, G., and Arnheim, N. (2000). Evidence for heterogeneity in recombination in the human pseudoautosomal region: High resolution analysis by sperm typing and radiation-hybrid mapping. *American Journal of Human Genetics*, 66(2):557–566.   *Cited at page 53*

Lim, F. L., Soulez, M., Koczan, D., Thiesen, H.-J., and Knight, J. C. (1998). A KRAB-related domain and a novel transcription repression domain in proteins encoded by SSX genes that are disrupted in human sarcomas. *Oncogene*, 17(15):2013.   *Cited at page 69*

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., Ward, L. D., Lowe, C. B., Holloway, A. K., Clamp, M., Gnerre, S., Alföldi, J., Beal, K., Chang, J., Clawson, H., Cuff, J., Di Palma, F., Fitzgerald, S., Flicek, P., Guttman, M., Hubisz, M. J., Jaffe, D. B., Jungreis, I., Kent, W. J., Kostka, D., Lara, M., Martins, A. L., Massingham, T., Moltke, I., Raney, B. J., Rasmussen, M. D., Robinson, J., Stark, A., Vilella, A. J., Wen, J., Xie, X., Zody, M. C., Broad Institute Sequencing Platform and Whole Genome Assembly Team, Baldwin, J., Bloom, T., Whye Chin, C., Heiman, D., Nicol, R., Nusbaum, C., Young, S., Wilkinson, J., Worley, K. C., Kovar, C. L., Muzny, D. M., Gibbs, R. A., Baylor College of Medicine Human Genome Sequencing Center Sequencing Team, Cree, A., Dihn, H. H., Fowler, G., Jhangiani, S., Joshi, V., Lee, S., Lewis, L. R., Nazareth, L. V., Okwuonu, G., Santibanez, J., Warren, W. C., Mardis, E. R., Weinstock, G. M., Wilson, R. K., Genome Institute at Washington University, Delehaunty, K., Dooling, D., Fronik, C., Fulton, L., Fulton, B., Graves, T., Minx, P., Sodergren, E., Birney, E., Margulies, E. H., Herrero, J., Green, E. D., Haussler, D., Siepel, A., Goldman, N., Pollard, K. S., Pedersen, J. S., Lander, E. S., and Kellis, M. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 478(7370):476–482.   *Cited at page 91*

Lindegren, C. C. (1953). Gene conversion in *Saccharomyces*. *Journal of Genetics*, 51(3):625–637.   *Cited at page 16*

Lindenbaum, P. (2015). JVarkit: Java-based utilities for Bioinformatics.   *Cited at page 114*

Lipkin, S. M., Moens, P. B., Wang, V., Lenzi, M., Shanmugarajah, D., Gilgeous, A., Thomas, J., Cheng, J., Touchman, J. W., Green, E. D., Schwartzberg, P., Collins, F. S., and Cohen, P. E. (2002). Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nature Genetics*, 31(4):385–390.   *Cited at page 44*

Lissouba, P., Mousseau, J., Rizet, G., and Rossignol, J. (1962). Fine structure of genes in the ascomcyete *Ascobolus immersus*. *Advances in Genetics*, 11:343–380.   *Cited at page 18*

Lissouba, P. and Rizet, G. (1960). On the existence of a polarized genetic unit which does not undergo non-reciprocal exchanges. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 250:3408.   *Cited at page 18*

Liu, H., Huang, J., Sun, X., Li, J., Hu, Y., Yu, L., Liti, G., Tian, D., Hurst, L. D., and Yang, S. (2018). Tetrad analysis in plants and fungi finds large differences in gene conversion rates but no GC bias. *Nature Ecology & Evolution*, 2(1):164. *Cited at page 95*

Liu, J.-G., Yuan, L., Brundell, E., Björkroth, B., Daneholt, B., and Höög, C. (1996). Localization of the n-terminus of scp1 to the central element of the synaptonemal complex and evidence for direct interactions between the n-termini of scp1 molecules organized head-to-head. *Experimental cell research*, 226(1):11–19. *Cited at page 32*

Loidl, J. (2016). Conservation and variability of meiosis across the eukaryotes. *Annual review of genetics*, 50:293–316. *Cited at page 33*

Long, H., Sung, W., Kucukyildirim, S., Williams, E., Miller, S. F., Guo, W., Patterson, C., Gregory, C., Strauss, C., Stone, C., Berne, C., Kysela, D., Shoemaker, W. R., Muscarella, M. E., Luo, H., Lennon, J. T., Brun, Y. V., and Lynch, M. (2018). Evolutionary determinants of genome-wide nucleotide composition. *Nature Ecology & Evolution*, 2(2):237–240. *Cited at page 95*

Lukaszewicz, A., Lange, J., Keeney, S., and Jasin, M. (2018). Control of meiotic double-strand-break formation by ATM: Local and global views. *Cell Cycle (Georgetown, Tex.)*, 17(10):1155–1172. *Cited at page 39*

Luo, G., Yao, M. S., Bender, C. F., Mills, M., Bladl, A. R., Bradley, A., and Petrini, J. H. (1999). Disruption of mRad50 causes embryonic stem cell lethality, abnormal embryonic development, and sensitivity to ionizing radiation. *Proceedings of the National Academy of Sciences of the United States of America*, 96(13):7376–7381. *Cited at page 40*

Lutzmann, M., Grey, C., Traver, S., Ganier, O., Maya-Mendoza, A., Ranisavljevic, N., Bernex, F., Nishiyama, A., Montel, N., Gavois, E., Forichon, L., de Massy, B., and Méchali, M. (2012). MCM8- and MCM9-deficient mice reveal gametogenesis defects and genome instability due to impaired homologous recombination. *Molecular Cell*, 47(4):523–534. *Cited at page 42*

Lynch, M. (2007). *The Origins of Genome Architecture*. Sinauer Associates. *Cited at page 90*

Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(3):961–968. *Cited at page 92*

Lynch, M., Gutenkunst, R., Ackerman, M., Spitze, K., Ye, Z., Maruki, T., and Jia, Z. (2017). Population Genomics of Daphnia pulex. *Genetics*, 206(1):315–332. *Cited at page 89*

Lynn, A., Ashley, T., and Hassold, T. (2004). Variation in human meiotic recombination. *Annual Review of Genomics and Human Genetics*, 5:317–349. *Cited at page 64*

Lynn, A., Koehler, K. E., Judis, L., Chan, E. R., Cherry, J. P., Schwartz, S., Seftel, A., Hunt, P. A., and Hassold, T. J. (2002). Covariation of Synaptonemal Complex Length and Mammalian Meiotic Exchange Rates. *Science*, 296(5576):2222–2225. *Cited at page 57*

Macaya, G., Cortadas, J., and Bernardi, G. (1978). An Analysis of the Bovine Genome by Density-Gradient Centrifugation. *European Journal of Biochemistry*, 84(1):179–188. *Cited at page 81*

Macaya, G., Thiery, J.-P., and Bernardi, G. (1976). An approach to the organization of eukaryotic genomes at a macromolecular level. *Journal of Molecular Biology*, 108(1):237–254. *Cited at page 81*

MacQueen, A. J., Colaiácovo, M. P., McDonald, K., and Villeneuve, A. M. (2002). Synapsis-dependent and-independent mechanisms stabilize homolog pairing during meiotic prophase in c. elegans. *Genes & development*, 16(18):2428–2442. *Cited at page 32*

MacQueen, A. J., Phillips, C. M., Bhalla, N., Weiser, P., Villeneuve, A. M., and Dernburg, A. F. (2005). Chromosome sites play dual roles to establish homologous synapsis during meiosis in c. elegans. *Cell*, 123(6):1037–1050. *Cited at page 30*

Maddox, J. F., Davies, K. P., Crawford, A. M., Hulme, D. J., Vaiman, D., Cribiu, E. P., Freking, B. A., Beh, K. J., Cockett, N. E., Kang, N., Riffkin, C. D., Drinkwater, R., Moore, S. S., Dodds, K. G., Lumsden, J. M., van Stijn, T. C., Phua, S. H., Adelson, D. L., Burkin, H. R., Broom, J. E., Buitkamp, J., Cambridge, L., Cushwa, W. T., Gerard, E., Galloway, S. M., Harrison, B., Hawken, R. J., Hiendleder, S., Henry, H. M., Medrano, J. F., Paterson, K. A., Schibler, L., Stone, R. T., and van Hest, B. (2001). An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Research*, 11(7):1275–1289. *Cited at page 65*

Maguire, M. P. (1988). Crossover site determination and interference. *Journal of Theoretical Biology*, 134(4):565–570. *Cited at page 57*

Mahadevaiah, S. K., Bourc'his, D., de Rooij, D. G., Bestor, T. H., Turner, J. M., and Burgoyne, P. S. (2008). Extensive meiotic asynapsis in mice antagonises meiotic silencing of unsynapsed chromatin and consequently disrupts meiotic sex chromosome inactivation. *The Journal of Cell Biology*, 182(2):263–276. *Cited at page 35*

Maleki, S., Neale, M. J., Arora, C., Henderson, K. A., and Keeney, S. (2007). Interactions between Mei4, Rec114, and other proteins required for meiotic DNA double-strand break formation in Saccharomyces cerevisiae. *Chromosoma*, 116(5):471–486. *Cited at page 39*

Maloisel, L. and Rossignol, J. L. (1998). Suppression of crossing-over by DNA methylation in Ascobolus. *Genes & Development*, 12(9):1381–1389. *Cited at page 61*

Mancera, E., Bourgon, R., Brozzi, A., Huber, W., and Steinmetz, L. M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*, 454(7203):479–485. *Cited at pages , 62, 64, 80, 95, 103*

Manzano-Winkler, B., McGaugh, S. E., and Noor, M. A. F. (2013). How Hot Are Drosophila Hotspots? Examining Recombination Rate Variation and Associations with Nucleotide Diversity, Divergence, and Maternal Age in Drosophila pseudoobscura. *PLoS ONE*, 8(8). *Cited at page 63*

Marais, G. (2003). Biased gene conversion: Implications for genome and sex evolution. *Trends in Genetics*, 19(6):330–338. *Cited at page 92*

Marais, G. and Duret, L. (2001). Synonymous codon usage, accuracy of translation, and gene length in Caenorhabditis elegans. *Journal of Molecular Evolution*, 52(3):275–280. *Cited at page 89*

Marais, G. and Galtier, N. (2003). Sex chromosomes: How X-Y recombination stops. *Current Biology*, 13(16):R641–R643. *Cited at pages 87, 92*

Marais, G., Mouchiroud, D., and Duret, L. (2001). Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10):5688–5692. *Cited at page 87*

Marais, G., Mouchiroud, D., and Duret, L. (2003). Neutral effect of recombination on base composition in Drosophila. *Genetical Research*, 81(2):79–87. *Cited at page 87*

Marais, G. and Piganeau, G. (2002). Hill-Robertson interference is a minor determinant of variations in codon bias across Drosophila melanogaster and Caenorhabditis elegans genomes. *Molecular Biology and Evolution*, 19(9):1399–1406. *Cited at pages 87, 89*

Margolin, J. F., Friedman, J. R., Meyer, W. K., Vissing, H., Thiesen, H. J., and Rauscher, F. J. (1994). Krüppel-associated boxes are potent transcriptional repression domains. *Proceedings of the National Academy of Sciences of the United States of America*, 91(10):4509–4513. *Cited at page 69*

Marsolier-Kergoat, M.-C. and Yeramian, E. (2009). GC Content and Recombination: Reassessing the Causal Effects for the Saccharomyces cerevisiae Genome. *Genetics*, 183(1):31–38. *Cited at page 60*

Marston, A. L. and Amon, A. (2005). Meiosis: cell-cycle controls shuffle and deal. *Nature Reviews Molecular Cell Biology*, 6(10):818. *Cited at page 28*

Martinez-Perez, E. and Colaiácovo, M. P. (2009). Distribution of meiotic recombination events: Talking to your neighbors. *Current Opinion in Genetics & Development*, 19(2):105–112. *Cited at pages 135, 140*

Martini, E., Borde, V., Legendre, M., Audic, S., Regnault, B., Soubigou, G., Dujon, B., and Llorente, B. (2011). Genome-wide analysis of heteroduplex DNA in mismatch repair-deficient yeast cells reveals novel properties of meiotic recombination pathways. *PLoS genetics*, 7(9):e1002305. *Cited at page 45*

Martini, E., Diaz, R. L., Hunter, N., and Keeney, S. (2006). Crossover Homeostasis in Yeast Meiosis. *Cell*, 126(2):285–295. *Cited at page 58*

Mary, N., Barasc, H., Ferchaud, S., Billon, Y., Meslier, F., Robelin, D., Calgaro, A., Loustau-Dudez, A.-M., Bonnet, N., Yerle, M., Acloque, H., Ducos, A., and Pinton, A. (2014). Meiotic Recombination Analyses of Individual Chromosomes in Male Domestic Pigs (Sus scrofa domestica). *PLOS ONE*, 9(6):e99123. *Cited at page 59*

Maynard Smith, J. (1977). Parental investment: A prospective analysis. *Animal Behaviour*. *Cited at page 24*

Maynard Smith, J. and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23(1):23–35. *Cited at page 22*

Mayr, E. (1942). *Systematics and the origin of species from the viewpoint of a zoologist.* Columbia University Press. *Cited at page 201*

Mayr, E. (1959). Where Are We? *Cold Spring Harbor Symposia on Quantitative Biology*, 24:1–14. *Cited at page 21*

Mayr, E. (1999). *Systematics and the Origin of Species, from the Viewpoint of a Zoologist.* Harvard University Press. *Cited at page 48*

Mazina, O. M., Mazin, A. V., Nakagawa, T., Kolodner, R. D., and Kowalczykowski, S. C. (2004). Saccharomyces cerevisiae Mer3 helicase stimulates 3'-5' heteroduplex extension by Rad51; implications for crossover control in meiotic recombination. *Cell*, 117(1):47–56. *Cited at pages 44, 164*

Mazzocchi, F. (2015). Could Big Data be the end of theory in science? *EMBO Reports*, 16(10):1250–1255. *Cited at page 205*

McCue, M. E., Bannasch, D. L., Petersen, J. L., Gurr, J., Bailey, E., Binns, M. M., Distl, O., Guérin, G., Hasegawa, T., Hill, E. W., Leeb, T., Lindgren, G., Penedo, M. C. T., Røed, K. H., Ryder, O. A., Swinburne, J. E., Tozaki, T., Valberg, S. J., Vaudin, M., Lindblad-Toh, K., Wade, C. M., and Mickelson, J. R. (2012). A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies. *PLOS Genetics*, 8(1):e1002451. *Cited at page 52*

McKee, B. D. (1996). The license to pair: Identification of meiotic pairing sites in *Drosophila. Chromosoma*, 105(3):135. *Cited at page 30*

McKee, B. D. (1998). Pairing sites and the role of chromosome pairing in meiosis and spermatogenesis in male Drosophila. *Current Topics in Developmental Biology*, 37:77–115. *Cited at page 55*

McKee, B. D. (2004). Homologous pairing and chromosome dynamics in meiosis and mitosis. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1677(1-3):165–180. *Cited at page 30*

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303. *Cited at page 114*

McKim, K. S., Green-Marroquin, B. L., Sekelsky, J. J., Chin, G., Steinberg, C., Khodosh, R., and Hawley, R. S. (1998). Meiotic synapsis in the absence of recombination. *Science*, 279(5352):876–878. *Cited at page 34*

McKim, K. S. and Hawley, R. S. (1995). Chromosomal control of meiotic cell division. *Science*, 270(5242):1595–1601. *Cited at page 28*

McKim, K. S. and Hayashi-Hagihara, A. (1998). Mei-W68 in Drosophila melanogaster encodes a Spo11 homolog: Evidence that the mechanism for initiating meiotic recombination is conserved. *Genes & Development*, 12(18):2932–2942. *Cited at page 39*

McKim, K. S., Jang, J. K., and Manheim, E. A. (2002). Meiotic recombination and chromosome segregation in drosophila females. *Annual review of genetics*, 36(1):205–232. *Cited at page 32*

McMahill, M. S., Sham, C. W., and Bishop, D. K. (2007). Synthesis-dependent strand annealing in meiosis. *PLoS biology*, 5(11):e299. *Cited at page 38*

McRae, A. F., McEwan, J. C., Dodds, K. G., Wilson, T., Crawford, A. M., and Slate, J. (2002). Linkage Disequilibrium in Domestic Sheep. *Genetics*, 160(3):1113–1122. *Cited at page 52*

McVean, G. A. T., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R., and Donnelly, P. (2004). The Fine-Scale Structure of Recombination Rate Variation in the Human Genome. *Science*, 304(5670):581–584. *Cited at pages 61, 62*

McVicker, G. and Green, P. (2010). Genomic signatures of germline gene expression. *Genome Research*, 20(11):1503–1511. *Cited at page 61*

Mehrotra, S. and McKim, K. S. (2006). Temporal Analysis of Meiotic DNA Double-Strand Break Formation and Repair in Drosophila Females. *PLoS Genetics*, 2(11). *Cited at page 58*

Melamed-Bessudo, C. and Levy, A. A. (2012). Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 109(16):E981–988. *Cited at page 61*

Mendel, G. (1865). Experiments in plant hybridization. *Verhandlungen des naturforschenden Vereins Brünn.* *Cited at page 8*

Menotti-Raymond, M., David, V. A., Lyons, L. A., Schäffer, A. A., Tomlin, J. F., Hutton, M. K., and O'Brien, S. J. (1999). A Genetic Linkage Map of Microsatellites in the Domestic Cat (Felis catus). *Genomics*, 57(1):9–23. *Cited at page 52*

Mercier, R., Jolivet, S., Vezon, D., Huppe, E., Chelysheva, L., Giovanni, M., Nogué, F., Doutriaux, M.-P., Horlow, C., Grelon, M., and Mézard, C. (2005). Two Meiotic Crossover Classes Cohabit in Arabidopsis: One Is Dependent on MER3,whereas the Other One Is Not. *Current Biology*, 15(8):692–701. *Cited at pages 44, 164*

Meselson, M., Stahl, F. W., and Vinograd, J. (1957). EQUILIBRIUM SEDIMENTA-TION OF MACROMOLECULES IN DENSITY GRADIENTS*. *Proceedings of the National Academy of Sciences of the United States of America*, 43(7):581–588. *Cited at page 81*

Meselson, M. S. and Radding, C. M. (1975). A general model for genetic recombination. *Proceedings of the National Academy of Sciences*, 72(1):358–361. *Cited at page 20*

Metzler-Guillemain, C. and de Massy, B. (2000). Identification and characterization of an SPO11 homolog in the mouse. *Chromosoma*, 109(1-2):133–138. *Cited at page 39*

Meunier, J. and Duret, L. (2004). Recombination drives the evolution of GC-content in the human genome. *Molecular Biology and Evolution*, 21(6):984–990. *Cited at pages 85, 87*

Meuwissen, R., Offenberg, H. H., Dietrich, A., Riesewijk, A., van Iersel, M., and Heyting, C. (1992). A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes. *The EMBO journal*, 11(13):5091–5100. *Cited at page 32*

Mézard, C. (2006). Meiotic recombination hotspots in plants. *Biochemical Society Transactions*, 34(4):531–534.   *Cited at page 63*

Mézard, C., Tagliaro Jahns, M., and Grelon, M. (2015). Where to cross? New insights into the location of meiotic crossovers. *Trends in Genetics*, 31(7):393–401.
                                                                        Cited at page 62

Mieczkowski, P. A., Dominska, M., Buck, M. J., Lieb, J. D., and Petes, T. D. (2007). Loss of a histone deacetylase dramatically alters the genomic distribution of Spo11p-catalyzed DNA breaks in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, 104(10):3955–3960.
                                                                        Cited at page 54

Mihola, O., Pratto, F., Brick, K., Linhartova, E., Kobets, T., Flachs, P., Baker, C. L., Sedlacek, R., Paigen, K., Petkov, P. M., Camerini-Otero, R. D., and Trachtulec, Z. (2019). Histone methyltransferase PRDM9 is not essential for meiosis in male mice. *Genome Research*, page gr.244426.118.   *Cited at page 68*

Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science (New York, N.Y.)*, 323(5912):373–375.   *Cited at page 71*

Mikawa, S., Akita, T., Hisamatsu, N., Inage, Y., Ito, Y., Kobayashi, E., Kusumoto, H., Matsumoto, T., Mikami, H., Minezawa, M., Miyake, M., Shimanuki, S., Sugiyama, C., Uchida, Y., Wada, Y., Yanai, S., and Yasue, H. (1999). A linkage map of 243 DNA markers in an intercross of Göttingen miniature and Meishan pigs. *Animal Genetics*, 30(6):407–417.   *Cited at page 64*

Minoche, A. E., Dohm, J. C., and Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology*, 12(11):R112.   *Cited at page 118*

Mirny, L. A. (2011). The fractal globule as a model of chromatin architecture in the cell. *Chromosome research*, 19(1):37–51.   *Cited at page 31*

Mirouze, M., Lieberman-Lazarovich, M., Aversano, R., Bucher, E., Nicolet, J., Reinders, J., and Paszkowski, J. (2012). Loss of DNA methylation affects the recombination landscape in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 109(15):5880–5885.   *Cited at page 61*

Mitchell, M. B. (1955a). Aberrant recombination of pyridoxine mutants of *Neurospora. Proceedings of the National Academy of Sciences of the United States of America*, 41(4):215.   *Cited at page 16*

Mitchell, M. B. (1955b). Further evidence of aberrant recombination in *Neurospora. Proceedings of the National Academy of Sciences of the United States of America*, 41(11):935.   *Cited at page 16*

Mitchell, M. B. (1956). A consideration of aberrant recombination in neurospora. *CR Lab. Carlsberg*, 26:285–298.   *Cited at page 19*

Montgomery, S. B., Goode, D. L., Kvikstad, E., Albers, C. A., Zhang, Z. D., Mu, X. J., Ananda, G., Howie, B., Karczewski, K. J., Smith, K. S., Anaya, V., Richardson, R., Davis, J., Consortium, T. . G. P., MacArthur, D. G., Sidow, A., Duret, L., Gerstein, M., Makova, K. D., Marchini, J., McVean, G., and Lunter, G. (2013). The origin, evolution, and functional impact of short insertion–deletion

variants identified in 179 human genomes. *Genome Research*, 23(5):749–761.
Cited at page 61

Montoya-Burgos, J. I., Boursot, P., and Galtier, N. (2003). Recombination explains isochores in mammalian genomes. *Trends in genetics: TIG*, 19(3):128–130.
Cited at page 86

Morgan, A. P., Gatti, D. M., Najarian, M. L., Keane, T. M., Galante, R. J., Pack, A. I., Mott, R., Churchill, G. A., and de Villena, F. P.-M. (2017). Structural Variation Shapes the Landscape of Recombination in Mouse. *Genetics*, 206(2):603–619.
Cited at page 59

Morgan, T. (1909). What are factors in Mendelian Inheritance? *American Breeders' Association Report*, 6:365–368.
Cited at page 10

Morgan, T., Sturtevant, A., Muller, H., and Bridges, C. (1915). *The Mechanism of Mendelian Heredity.* Henry Holt and Company.
Cited at pages xi, 13

Morgan, T. H. (1910). Sex limited inheritance in *Drosophila. Science*, 32(812):120–122.
Cited at page 12

Morgan, T. H. (1911). Random segregation versus coupling in Mendelian inheritance. *Science*, 34(873):384–384.
Cited at pages 12, 50

Morgan, T. H. (1912). Complete linkage in the second chromosome of the male Drosophila. *Science*, 36:719–720.
Cited at page 64

Morgan, T. H. (1914). No Crossing over in the Male of Drosophila of Genes in the Second and Third Pairs of Chromosomes. *Biological Bulletin*, 26(4):195–204.
Cited at page 64

Morgan, T. H. and Cattell, E. (1912). Data for the study of sex-linked inheritance in Drosophila. *Journal of Experimental Zoology*, 13(1):79–101.
Cited at page 50

Moriyama, E. N. and Powell, J. R. (1997). Codon Usage Bias and tRNA Abundance in Drosophila. *Journal of Molecular Evolution*, 45(5):514–523.
Cited at page 89

Morton, N. E., Jacobs, P. A., Hassold, T., and Wu, D. (1988). Maternal age in trisomy. *Annals of Human Genetics*, 52(3):227–235.
Cited at page 36

Moses, M. J. (1956). Chromosomal structures in crayfish spermatocytes. *J Biophys Biochem Cytol*, 2(2):215–218.
Cited at page 32

Mouchiroud, D., D'Onofrio, G., Aïssani, B., Macaya, G., Gautier, C., and Bernardi, G. (1991). The distribution of genes in the human genome. *Gene*, 100:181–187.
Cited at page 81

Mouchiroud, D., Gautier, C., and Bernardi, G. (1988). The compositional distribution of coding sequences and DNA molecules in humans and murids. *Journal of Molecular Evolution*, 27(4):311–320.
Cited at page 82

Mougel, F., Poursat, M.-A., Beaume, N., Vautrin, D., and Solignac, M. (2014). High-resolution linkage map for two honeybee chromosomes: The hotspot quest. *Molecular genetics and genomics: MGG*, 289(1):11–24.
Cited at page 63

Mouse Genome Sequencing Consortium, Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigó, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., McCarthy, M., McCombie, W. R., McLaren, S., McLay, K., McPherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M., Mott, R., Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M. J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J. P., Von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S.-P., Zdobnov, E. M., Zody, M. C., and Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915):520–562. *Cited at page 82*

Moynahan, M. E. and Jasin, M. (2010). Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nature Reviews. Molecular Cell Biology*, 11(3):196–207. *Cited at page 46*

Muñoz-Fuentes, V., Marcet-Ortega, M., Alkorta-Aranburu, G., Linde Forsberg, C., Morrell, J. M., Manzano-Piedras, E., Söderberg, A., Daniel, K., Villalba, A., Toth, A., Di Rienzo, A., Roig, I., and Vilà, C. (2015). Strong Artificial Selection in Domestic Mammals Did Not Result in an Increased Recombination Rate. *Molecular Biology and Evolution*, 32(2):510–523. *Cited at page 59*

Muñoz-Fuentes, V., Rienzo, A. D., and Vilà, C. (2011). Prdm9, a Major Determinant of Meiotic Recombination Hotspots, Is Not Functional in Dogs and Their Wild Relatives, Wolves and Coyotes. *PLOS ONE*, 6(11):e25498. *Cited at page 69*

Mugal, C. F., Arndt, P. F., and Ellegren, H. (2013). Twisted Signatures of GC-Biased Gene Conversion Embedded in an Evolutionary Stable Karyotype. *Molecular Biology and Evolution*, 30(7):1700–1712. *Cited at page 87*

Mugal, C. F., Weber, C. C., and Ellegren, H. (2015). GC-biased gene conversion links the recombination landscape and demography to genomic base composition. *BioEssays*, 37(12):1317–1326.   *Cited at page 94*

Muller, H. J. (1916). The Mechanism of Crossing-Over. *The American Naturalist*, 50(592):193–221.   *Cited at page 56*

Muller, H. J. (1920). Are the Factors of Heredity Arranged in a Line? *The American Naturalist*, 54(631):97–121.   *Cited at page 50*

Munch, K., Mailund, T., Dutheil, J. Y., and Schierup, M. H. (2014). A fine-scale recombination map of the human–chimpanzee ancestor reveals faster change in humans than in chimpanzees and a strong impact of GC-biased gene conversion. *Genome Research*, 24(3):467–474.   *Cited at page 88*

Munz, P. (1994). An Analysis of Interference in the Fission Yeast Schizosaccharomyces Pombe. *Genetics*, 137(3):701–707.   *Cited at page 58*

Murray, N. E. (1968). Polarized intragenic recombination in chromosome rearrangements of *Neurospora*. *Genetics*, 58(2):181.   *Cited at page 18*

Murray, N. E. et al. (1960). Complementation and recombination between *methionine-2* alleles in *Neurospora crassa*. *Heredity*, 15:207–17.   *Cited at page 18*

Muyle, A., Serres-Giardi, L., Ressayre, A., Escobar, J., and Glémin, S. (2011). GC-Biased Gene Conversion and Selection Affect GC Content in the Oryza Genus (rice). *Molecular Biology and Evolution*, 28(9):2695–2706.   *Cited at page 90*

Myers, S., Bottolo, L., Freeman, C., McVean, G., and Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science (New York, N.Y.)*, 310(5746):321–324.   *Cited at pages 59, 61, 62, 63, 67*

Myers, S., Bowden, R., Tumian, A., Bontrop, R. E., Freeman, C., MacFie, T. S., McVean, G., and Donnelly, P. (2010). Drive Against Hotspot Motifs in Primates Implicates the PRDM9 gene in Meiotic Recombination. *Science (New York, N.Y.)*, 327(5967).   *Cited at pages , 48, 67, 74*

Myers, S., Freeman, C., Auton, A., Donnelly, P., and McVean, G. (2008). A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genetics*, 40(9):1124–1129.   *Cited at page 67*

Nachman, M. W. (2001). Single nucleotide polymorphisms and recombination rate in humans. *Trends in Genetics*, 17(9):481–485.   *Cited at page 61*

Nachman, M. W. (2002). Variation in recombination rate across the genome: Evidence and implications. *Current Opinion in Genetics & Development*, 12(6):657–663.   *Cited at page 59*

Nagylaki, T. (1983). Evolution of a finite population under gene conversion. *Proceedings of the National Academy of Sciences of the United States of America*, 80(20):6278–6281.   *Cited at pages 89, 152, 153, 157*

Nakagawa, T. and Kolodner, R. D. (2002a). The MER3 DNA Helicase Catalyzes the Unwinding of Holliday Junctions. *Journal of Biological Chemistry*, 277(31):28019–28024.   *Cited at page 164*

Nakagawa, T. and Kolodner, R. D. (2002b). Saccharomyces cerevisiae Mer3 Is a DNA Helicase Involved in Meiotic Crossing Over. *Molecular and Cellular Biology*, 22(10):3281–3291. *Cited at page 164*

Nakagawa, T. and Ogawa, H. (1999). The Saccharomyces cerevisiae MER3 gene, encoding a novel helicase-like protein, is required for crossover control in meiosis. *The EMBO Journal*, 18(20):5714–5723. *Cited at page 164*

Narasimhan, V. M., Hunt, K. A., Mason, D., Baker, C. L., Karczewski, K. J., Barnes, M. R., Barnett, A. H., Bates, C., Bellary, S., Bockett, N. A., Giorda, K., Griffiths, C. J., Hemingway, H., Jia, Z., Kelly, M. A., Khawaja, H. A., Lek, M., McCarthy, S., McEachan, R., O'Donnell-Luria, A., Paigen, K., Parisinos, C. A., Sheridan, E., Southgate, L., Tee, L., Thomas, M., Xue, Y., Schnall-Levin, M., Petkov, P. M., Tyler-Smith, C., Maher, E. R., Trembath, R. C., MacArthur, D. G., Wright, J., Durbin, R., and van Heel, D. A. (2016). Health and population effects of rare gene knockouts in adult humans with related parents. *Science*, 352(6284):474–477. *Cited at page 68*

Nasmyth, K. (2015). A meiotic mystery: How sister kinetochores avoid being pulled in opposite directions during the first division. *Bioessays*, 37(6):657–665. *Cited at page 27*

Nassif, N., Penney, J., Pal, S., Engels, W. R., and Gloor, G. B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Molecular and Cellular Biology*, 14(3):1613–1625. *Cited at page 38*

Neale, M. J. (2010). PRDM9 points the zinc finger at meiotic recombination hotspots. *Genome Biology*, 11(2):104. *Cited at page 69*

Neale, M. J. and Keeney, S. (2006). Clarifying the mechanics of dna strand exchange in meiotic recombination. *Nature*, 442(7099):153. *Cited at pages 30, 40*

Neale, M. J., Pan, J., and Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature*, 436(7053):1053–1057. *Cited at page 40*

Neaves, W. B. and Baumann, P. (2011). Unisexual reproduction among vertebrates. *Trends in Genetics*, 27(3):81–88. *Cited at page 24*

Necşulea, A., Popa, A., Cooper, D. N., Stenson, P. D., Mouchiroud, D., Gautier, C., and Duret, L. (2011). Meiotic recombination favors the spreading of deleterious mutations in human populations. *Human Mutation*, 32(2):198–206. *Cited at pages 92, 99, 191*

Nee, S., Read, A. F., Greenwood, J. J. D., and Harvey, P. H. (1991). The relationship between abundance and body size in British birds. *Nature*, 351(6324):312. *Cited at page 99*

Neff, M. W., Broman, K. W., Mellersh, C. S., Ray, K., Acland, G. M., Aguirre, G. D., Ziegle, J. S., Ostrander, E. A., and Rine, J. (1999). A second-generation genetic linkage map of the domestic dog, Canis familiaris. *Genetics*, 151(2):803–820. *Cited at page 64*

Nelson, O. E. (1959). Intracistron Recombination in the Wx/wx Region in Maize. *Science (New York, N.Y.)*, 130(3378):794–795. *Cited at page 61*

Nelson, O. E. (1962). The Waxy Locus in Maize. I. Intralocus Recombination Frequency Estimates by Pollen and by Conventional Analyses. *Genetics*, 47(6):737–742.   *Cited at page 61*

Nelson, O. E. (1975). The Waxy Locus in Maize III. Effect of Structural Heterozygosity on Intragenic Recombination and Flanking Marker Assortment. *Genetics*, 79(1):31–44.   *Cited at page 61*

Neumann, R. and Jeffreys, A. J. (2006). Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. *Human Molecular Genetics*, 15(9):1401–1411.   *Cited at pages 66, 72*

Ng, S. H., Parvanov, E., Petkov, P. M., and Paigen, K. (2008). A quantitative assay for crossover and noncrossover molecular events at individual recombination hotspots in both male and female gametes. *Genomics*, 92(4):204–209.   *Cited at pages 45, 126, 217*

Nicolas, A. and Petes, T. (1994). Polarity of meiotic gene conversion in fungi: contrasting views. *Experientia*, 50(3):242–252.   *Cited at pages 18, 83*

Niu, H., Wan, L., Baumgartner, B., Schaefer, D., Loidl, J., and Hollingsworth, N. M. (2005). Partner Choice during Meiosis Is Regulated by Hop1-promoted Dimerization of Mek1. *Molecular Biology of the Cell*, 16(12):5804–5818.   *Cited at page 59*

Niu, H., Wan, L., Busygina, V., Kwon, Y., Allen, J. A., Li, X., Kunz, R. C., Kubota, K., Wang, B., Sung, P., Shokat, K. M., Gygi, S. P., and Hollingsworth, N. M. (2009). Regulation of Meiotic Recombination via Mek1-Mediated Rad54 Phosphorylation. *Molecular Cell*, 36(3):393–404.   *Cited at page 59*

Niu, Z., Xue, Q., Wang, H., Xie, X., Zhu, S., Liu, W., and Ding, X. (2017). Mutational Biases and GC-Biased Gene Conversion Affect GC Content in the Plastomes of Dendrobium Genus. *International Journal of Molecular Sciences*, 18(11).   *Cited at page 95*

Nordborg, M. and Tavaré, S. (2002). Linkage disequilibrium: What history has to tell us. *Trends in genetics : TIG*, 18(2):83–90.   *Cited at page 51*

Normarck, B. B., Judson, O. P., and Moran, N. A. (2003). Genomic signatures of ancient asexual lineages. *Biological Journal of the Linnean Society*, 79(1):69–84.   *Cited at page 24*

Nsengimana, J., Baret, P., Haley, C. S., and Visscher, P. M. (2004). Linkage Disequilibrium in the Domesticated Pig. *Genetics*, 166(3):1395–1404.   *Cited at page 52*

Odenthal-Hesse, L., Berg, I. L., Veselis, A., Jeffreys, A. J., and May, C. A. (2014). Transmission distortion affecting human noncrossover but not crossover recombination: A hidden source of meiotic drive. *PLoS genetics*, 10(2):e1004106.   *Cited at page 96*

Ohta, T. (1973). Slightly Deleterious Mutant Substitutions in Evolution. *Nature*, 246(5428):96.   *Cited at page 22*

Okagaki, R. J., Dukowic-Schulze, S., Eggleston, W. B., and Muehlbauer, G. J. (2018). A Critical Assessment of 60 Years of Maize Intragenic Recombination. *Frontiers in Plant Science*, 9.   *Cited at page 61*

Olive, L. S. (1959). Aberrant tetrads in *Sordaria fimicola*. *Proceedings of the National Academy of Sciences of the United States of America*, 45(5):727.
Cited at pages 16, 17

Oliver, P. L., Goodstadt, L., Bayes, J. J., Birtle, Z., Roach, K. C., Phadnis, N., Beatson, S. A., Lunter, G., Malik, H. S., and Ponting, C. P. (2009). Accelerated Evolution of the Prdm9 Speciation Gene across Diverse Metazoan Taxa. *PLOS Genetics*, 5(12):e1000753.
Cited at page 71

Oliver, S. G., van der Aart, Q. J. M., Agostoni-Carbone, M. L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J. P. G., Benit, P., Berben, G., Bergantino, E., Biteau, N., Bolle, P. A., Bolotin-Fukuhara, M., Brown, A., Brown, A. J. P., Buhler, J. M., Carcano, C., Carignani, G., Cederberg, H., Chanet, R., Contreras, R., Crouzet, M., Daignan-Fornier, B., Defoor, E., Delgado, M., Demolder, J., Doira, C., Dubois, E., Dujon, B., Dusterhoft, A., Erdmann, D., Esteban, M., Fabre, F., Fairhead, C., Faye, G., Feldmann, H., Fiers, W., Francingues-Gaillard, M. C., Franco, L., Frontali, L., Fukuhara, H., Fuller, L. J., Galland, P., Gent, M. E., Gigot, D., Gilliquet, V., Glansdorff, N., Goffeau, A., Grenson, M., Grisanti, P., Grivell, L. A., de Haan, M., Haasemann, M., Hatat, D., Hoenicka, J., Hegemann, J., Herbert, C. J., Hilger, F., Hohmann, S., Hollenberg, C. P., Huse, K., Iborra, F., Indje, K. J., Isono, K., Jacq, C., Jacquet, M., James, C. M., Jauniaux, J. C., Jia, Y., Jimenez, A., Kelly, A., Kleinhans, U., Kreisl, P., Lanfranchi, G., Lewis, C., vanderLinden, C. G., Lucchini, G., Lutzenkirchen, K., Maat, M. J., Mallet, L., Mannhaupet, G., Martegani, E., Mathieu, A., Maurer, C. T. C., McConnell, D., McKee, R. A., Messenguy, F., Mewes, H. W., Molemans, F., Montague, M. A., Falconi, M. M., Navas, L., Newlon, C. S., Noone, D., Pallier, C., Panzeri, L., Pearson, B. M., Perea, J., Philippsen, P., Pierard, A., Planta, R. J., Plevani, P., Poetsch, B., Pohl, F., Purnelle, B., Rad, M. R., Rasmussen, S. W., Raynal, A., Remacha, M., Richterich, P., Roberts, A. B., Rodriguez, F., Sanz, E., Schaaff-Gerstenschlager, I., Scherens, B., Schweitzer, B., Shu, Y., Skala, J., Slonimski, P. P., Sor, F., Soustelle, C., Spiegelberg, R., Stateva, L. I., Steensma, H. Y., Steiner, S., Thierry, A., Thireos, G., Tzermia, M., Urrestarazu, L. A., Valle, G., Vetter, I., van Vliet-Reedijk, J. C., Voet, M., Volckaert, G., Vreken, P., Wang, H., Warmington, J. R., von Wettstein, D., Wicksteed, B. L., Wilson, C., Wurst, H., Xu, G., Yoshikawa, A., Zimmermann, F. K., and Sgouros, J. G. (1992). The complete DNA sequence of yeast chromosome III. *Nature*, 357(6373):38.
Cited at page 50

Orr-Weaver, T. L. (1995). Meiosis in Drosophila: Seeing is believing. *Proceedings of the National Academy of Sciences of the United States of America*, 92(23):10443–10449.
Cited at page 55

Orr-Weaver, T. L. and Szostak, J. (1985). Fungal recombination. *Microbiological reviews*, 49(1):33.
Cited at pages 16, 17, 18

Orr-Weaver, T. L. and Szostak, J. W. (1983). Yeast recombination: The association between double-strand gap repair and crossing-over. *Proceedings of the National Academy of Sciences of the United States of America*, 80(14):4417–4421.
Cited at page 37

Orr-Weaver, T. L., Szostak, J. W., and Rothstein, R. J. (1981). Yeast transformation: A model system for the study of recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 78(10):6354–6358.
Cited at page 37

Orth, A., Adama, T., Din, W., and Bonhomme, F. (1998). [Natural hybridization between two subspecies of the house mouse, Mus musculus domesticus and Mus

musculus castaneus, near Lake Casitas, California]. *Genome*, 41(1):104–110.
Cited at page 108

Orwell, G. (1949). *Nineteen Eighty-Four*. Secker & Warburg.   *Cited at page 3*

Osman, F., Dixon, J., Doe, C. L., and Whitby, M. C. (2003). Generating crossovers
by resolution of nicked Holliday junctions: A role for Mus81-Eme1 in meiosis.
Molecular Cell, 12(3):761–774.   *Cited at page 44*

Otto, S. P. (2009). The evolutionary enigma of sex. *the american naturalist*,
174(S1):S1–S14.   *Cited at page 24*

Otto, S. P. and Lenormand, T. (2002). Evolution of sex: resolving the paradox of
sex and recombination. *Nature Reviews Genetics*, 3(4):252.   *Cited at page 24*

Page, S. L. and Hawley, R. S. (2004). The genetics and molecular biology of the
synaptonemal complex. *Annu. Rev. Cell Dev. Biol.*, 20:525–558. *Cited at page 32*

Paigen, K. and Petkov, P. (2010). Mammalian recombination hot spots:
Properties, control and evolution. *Nature reviews. Genetics*, 11(3):221–233.
Cited at pages 63, 64

Paigen, K. and Petkov, P. M. (2018). PRDM9 and Its Role in Genetic Recombination.
Trends in genetics: TIG, 34(4):291–300.   *Cited at pages 69, 70*

Paigen, K., Szatkiewicz, J. P., Sawyer, K., Leahy, N., Parvanov, E. D., Ng,
S. H. S., Graber, J. H., Broman, K. W., and Petkov, P. M. (2008). The
Recombinational Anatomy of a Mouse Chromosome. *PLoS Genetics*, 4(7).
Cited at pages xii, 64, 65, 66, 126, 127, 128, 143, 188

Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H. G., Tischfield,
S. E., Zhu, X., Neale, M. J., Jasin, M., Socci, N. D., Hochwagen, A., and
Keeney, S. (2011). A hierarchical combination of factors shapes the genome-
wide topography of yeast meiotic recombination initiation. *Cell*, 144(5):719–731.
Cited at pages 54, 58, 60

Pardo-Manuel de Villena, F. and Sapienza, C. (2001). Recombination is proportional
to the number of chromosome arms in mammals. *Mammalian Genome*, 12(4):318–
322.   *Cited at page 55*

Park, P. J. (2009). ChIP-Seq: Advantages and challenges of a maturing technology.
Nature reviews. Genetics, 10(10):669–680.   *Cited at page 54*

Parvanov, E. D., Ng, S. H. S., Petkov, P. M., and Paigen, K. (2009). Trans-
regulation of mouse meiotic recombination hotspots by Rcr1. *PLoS biology*,
7(2):e36.   *Cited at page 68*

Parvanov, E. D., Petkov, P. M., and Paigen, K. (2010). Prdm9 Controls Activation
of Mammalian Recombination Hotspots. *Science (New York, N.Y.)*, 327(5967):835.
Cited at pages , 48, 67, 68, 74

Parvanov, E. D., Tian, H., Billings, T., Saxl, R. L., Spruce, C., Aithal, R., Krejci,
L., Paigen, K., and Petkov, P. M. (2016). PRDM9 interactions with other proteins
provide a link between recombination hotspots and the chromosomal axis in
meiosis. *Molecular Biology of the Cell*, 28(3):488–499.   *Cited at page 69*

Parvanov, E. D., Tian, H., Billings, T., Saxl, R. L., Spruce, C., Aithal, R., Krejci, L., Paigen, K., and Petkov, P. M. (2017). PRDM9 interactions with other proteins provide a link between recombination hotspots and the chromosomal axis in meiosis. *Molecular Biology of the Cell*, 28(3):488–499. *Cited at page 69*

Paulin, F. (2015). *L'épistémologie contemporaine de la Théorie de l'évolution dans l'enseignement secondaire français : état des lieux et conséquences didactiques.* PhD thesis. *Cited at pages 201, 202*

Pearl, R. and Schoppe, W. F. (1921). Studies on the physiology of reproduction in the domestic fowl. xviii. further observations on the anatomical basis of fecundity. *Journal of Experimental Zoology*, 34(1):100–118. *Cited at page 28*

Penkner, A. M., Fridkin, A., Gloggnitzer, J., Baudrimont, A., Machacek, T., Woglar, A., Csaszar, E., Pasierbek, P., Ammerer, G., Gruenbaum, Y., et al. (2009). Meiotic chromosome homology search involves modifications of the nuclear envelope protein matefin/sun-1. *Cell*, 139(5):920–933. *Cited at page 28*

Penrose, L. S. (2009). The relative effects of paternal and maternal age in mongolism. 1933. *Journal of Genetics*, 88(1):9–14. *Cited at page 36*

Pera, R. R., Pellicer, A., and Sim, C. (2013). *Stem cells in reproductive medicine: Basic science and therapeutic potential.* Cambridge University Press. *Cited at page 32*

Perkins, D. D. (1962). The frequency in neurospora tetrads of multiple exchanges within short intervals. *Genetics Research*, 3(2):315–327. *Cited at page 19*

Perry, J. and Ashworth, A. (1999). Evolutionary rate of a gene affected by chromosomal position. *Current biology: CB*, 9(17):987–989. *Cited at page 86*

Pessia, E., Popa, A., Mousset, S., Rezvoy, C., Duret, L., and Marais, G. A. B. (2012). Evidence for widespread GC-biased gene conversion in eukaryotes. *Genome Biology and Evolution*, 4(7):675–682. *Cited at pages , 87, 95*

Peters, A. D. (2008). A Combination of cis and trans Control Can Solve the Hotspot Conversion Paradox. *Genetics*, 178(3):1579–1593. *Cited at page 74*

Petes, T. D. (2001). Meiotic recombination hot spots and cold spots. *Nature Reviews. Genetics*, 2(5):360–369. *Cited at pages 60, 62*

Petes, T. D. and Merker, J. D. (2002). Context Dependence of Meiotic Recombination Hotspots in Yeast: The Relationship Between Recombination Activity of a Reporter Construct and Base Composition. *Genetics*, 162(4):2049–2052. *Cited at page 87*

Petkov, P. M., Broman, K. W., Szatkiewicz, J. P., and Paigen, K. (2007). Crossover interference underlies sex differences in recombination rates. *Trends in genetics: TIG*, 23(11):539–542. *Cited at pages 57, 65*

Petronczki, M., Siomos, M. F., and Nasmyth, K. (2003). Un ménage à quatre: the molecular biology of chromosome segregation in meiosis. *Cell*, 112(4):423–440. *Cited at page 27*

Pfeifer, C., Scherthan, H., and Thomsen, P. D. (2003). Sex-specific telomere redistribution and synapsis initiationin cattle oogenesis. *Developmental biology*, 255(2):206–215. *Cited at page 29*

Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J. L., and Mayer, G. (2018). Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific Reports*, 8(1):10950.   *Cited at page 121*

Phifer-Rixey, M., Bonhomme, F., Boursot, P., Churchill, G. A., Piálek, J., Tucker, P. K., and Nachman, M. W. (2012). Adaptive Evolution and Effective Population Size in Wild House Mice. *Molecular Biology and Evolution*, 29(10):2949–2955.   *Cited at pages , 103, 186, 187*

Pineda-Krch, M. and Redfield, R. J. (2005). Persistence and Loss of Meiotic Recombination Hotspots. *Genetics*, 169(4):2319–2333.   *Cited at page 73*

Pittman, D. L., Cobb, J., Schimenti, K. J., Wilson, L. A., Cooper, D. M., Brignull, E., Handel, M. A., and Schimenti, J. C. (1998). Meiotic prophase arrest with failure of chromosome synapsis in mice deficient for Dmc1, a germline-specific RecA homolog. *Molecular Cell*, 1(5):697–705.   *Cited at page 40*

Plotkin, J. B. and Kudla, G. (2011). Synonymous but not the same: The causes and consequences of codon bias. *Nature reviews. Genetics*, 12(1):32–42.   *Cited at page 90*

Plutynski, A. (2007). Drift: A Historical and Conceptual Overview. *Biological Theory*, 2(2):156–167.   *Cited at page 22*

Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J. S., Bejerano, G., Baertsch, R., Rosenbloom, K. R., Kent, J., and Haussler, D. (2006a). Forces shaping the fastest evolving regions in the human genome. *PLoS genetics*, 2(10):e168.   *Cited at page 91*

Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J. S., Katzman, S., King, B., Onodera, C., Siepel, A., Kern, A. D., Dehay, C., Igel, H., Ares, M., Vanderhaeghen, P., and Haussler, D. (2006b). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, 443(7108):167.   *Cited at page 91*

Ponting, C. P. (2011). What are the genomic drivers of the rapid evolution of PRDM9? *Trends in Genetics*, 27(5):165–171.   *Cited at pages 71, 75*

Popa, A., Samollow, P., Gautier, C., and Mouchiroud, D. (2012). The Sex-Specific Impact of Meiotic Recombination on Nucleotide Composition. *Genome Biology and Evolution*, 4(3):412–422.   *Cited at page 88*

Popa, A.-M. (2011). *The Evolution of Recombination and Genomic Structures : A Modelling Approach.* PhD thesis, Université Claude Bernard - Lyon I.   *Cited at page 65*

Pouyet, F. (2016). *Étude Bioinformatique de l'évolution Du Code Génétique.* Thesis, Lyon.   *Cited at page 89*

Pouyet, F., Aeschbacher, S., Thiéry, A., and Excoffier, L. (2018). Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *eLife*, 7:e36317.   *Cited at page 96*

Pouyet, F., Mouchiroud, D., Duret, L., and Sémon, M. (2017). Recombination, meiotic expression and human codon usage. *eLife*, 6:e27344.   *Cited at pages 61, 83, 90*

Powers, N. R., Parvanov, E. D., Baker, C. L., Walker, M., Petkov, P. M., and Paigen, K. (2016). The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLOS Genetics*, 12(6):e1006146. *Cited at page 69*

Prabhakar, S., Noonan, J. P., Pääbo, S., and Rubin, E. M. (2006). Accelerated evolution of conserved noncoding sequences in humans. *Science (New York, N.Y.)*, 314(5800):786. *Cited at page 91*

Pradillo, M. and Santos, J. L. (2011). The template choice decision in meiosis: Is the sister important? *Chromosoma*, 120(5):447. *Cited at page 58*

Pratto, F., Brick, K., Khil, P., Smagulova, F., Petukhova, G. V., and Camerini-Otero, R. D. (2014). Recombination initiation maps of individual human genomes. *Science (New York, N.Y.)*, 346(6211):1256442. *Cited at pages 60, 68, 70*

Pritchard, R. (1955). The linear arrangement of a series of alleles of aspergillus nidulans. *Heredity*, 9(3):343. *Cited at page 18*

Provine, W. B. (2001). *The Origins of Theoretical Population Genetics: With a New Afterword.* University of Chicago Press. *Cited at page 20*

Ptak, S. E., Hinds, D. A., Koehler, K., Nickel, B., Patil, N., Ballinger, D. G., Przeworski, M., Frazer, K. A., and Pääbo, S. (2005). Fine-scale recombination patterns differ between chimpanzees and humans. *Nature Genetics*, 37(4):429–434. *Cited at pages 67, 85*

Ptak, S. E., Roeder, A. D., Stephens, M., Gilad, Y., Pääbo, S., and Przeworski, M. (2004). Absence of the TAP2 human recombination hotspot in chimpanzees. *PLoS biology*, 2(6):e155. *Cited at page 67*

Pyatnitskaya, A., Borde, V., and De Muyt, A. (2019). Crossing and zipping: Molecular duties of the ZMM proteins in meiosis. *Chromosoma.* *Cited at page 44*

Qiao, H., Chen, J. K., Reynolds, A., Höög, C., Paddy, M., and Hunter, N. (2012). Interplay between Synaptonemal Complex, Homologous Recombination, and Centromeres during Mammalian Meiosis. *PLOS Genetics*, 8(6):e1002790. *Cited at page 60*

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6):841–842. *Cited at page 114*

Ramesh, M. A., Malik, S.-B., and Logsdon Jr, J. M. (2005). A phylogenomic inventory of meiotic genes: evidence for sex in giardia and an early eukaryotic origin of meiosis. *Current Biology*, 15(2):185–191. *Cited at page 25*

Rasmussen, S. W. (1977). The transformation of the Synaptonemal Complex into the "elimination chromatin" in *Bombyx mori* oocytes. *Chromosoma*, 60(3):205–221. *Cited at page 34*

Ratnakumar, A., Mousset, S., Glémin, S., Berglund, J., Galtier, N., Duret, L., and Webster, M. T. (2010). Detecting positive selection within genomes: The problem of biased gene conversion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1552):2571–2580. *Cited at page 91*

Rattray, A., Santoyo, G., Shafer, B., and Strathern, J. N. (2015). Elevated Mutation Rate during Meiosis in Saccharomyces cerevisiae. *PLoS Genetics*, 11(1). *Cited at page 61*

Ream, R. A., Johns, G. C., and Somero, G. N. (2003). Base compositions of genes encoding alpha-actin and lactate dehydrogenase-A from differently adapted vertebrates show no temperature-adaptive variation in G + C content. *Molecular Biology and Evolution*, 20(1):105–110. *Cited at page 83*

Resnick, M. A. (1976). The repair of double-strand breaks in DNA: A model involving recombination. *Journal of Theoretical Biology*, 59(1):97–106. *Cited at page 38*

Reynolds, A., Qiao, H., Yang, Y., Chen, J. K., Jackson, N., Biswas, K., Holloway, J. K., Baudat, F., de Massy, B., Wang, J., Höög, C., Cohen, P. E., and Hunter, N. (2013). RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. *Nature Genetics*, 45(3):269–278. *Cited at page 44*

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Cited at page 206*

Robine, N., Uematsu, N., Amiot, F., Gidrol, X., Barillot, E., Nicolas, A., and Borde, V. (2007). Genome-Wide Redistribution of Meiotic Double-Strand Breaks in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 27(5):1868–1880. *Cited at page 56*

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1):24–26. *Cited at page 114*

Robinson, M. C., Stone, E. A., and Singh, N. D. (2014). Population Genomic Analysis Reveals No Evidence for GC-Biased Gene Conversion in Drosophila melanogaster. *Molecular Biology and Evolution*, 31(2):425–433. *Cited at page 95*

Rockmill, B., Fung, J. C., Branda, S. S., and Roeder, G. S. (2003). The Sgs1 helicase regulates chromosome synapsis and meiotic crossing over. *Current biology: CB*, 13(22):1954–1962. *Cited at page 58*

Rockmill, B., Voelkel-Meiman, K., and Roeder, G. S. (2006). Centromere-Proximal Crossovers Are Associated With Precocious Separation of Sister Chromatids During Meiosis in Saccharomyces cerevisiae. *Genetics*, 174(4):1745–1754. *Cited at page 60*

Rodgers-Melnick, E., Vera, D. L., Bass, H. W., and Buckler, E. S. (2016). Open chromatin reveals the functional maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, 113(22):E3177–3184. *Cited at page 95*

Roeder, G. S. (1997). Meiotic chromosomes: It takes two to tango. *Genes & Development*, 11(20):2600–2621. *Cited at page 55*

Roeder, G. S. and Bailis, J. M. (2000). The pachytene checkpoint. *Trends in genetics: TIG*, 16(9):395–403. *Cited at page 34*

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., and Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *The Journal of Biological Chemistry*, 273(10):5858–5868. *Cited at page 39*

Roman, H. (1985). Gene conversion and crossing-over. *Environmental mutagenesis*, 7(6):923–932.   *Cited at page 16*

Roman, H. (1986). The early days of yeast genetics: a personal narrative. *Annual review of genetics*, 20(1):1–14.   *Cited at page 16*

Romanienko, P. J. and Camerini-Otero, R. D. (2000). The mouse Spo11 gene is required for meiotic chromosome synapsis. *Molecular Cell*, 6(5):975–987.
Cited at page 39

Romiguier, J., Cameron, S. A., Woodard, S. H., Fischman, B. J., Keller, L., and Praz, C. J. (2016). Phylogenomics Controlling for Base Compositional Bias Reveals a Single Origin of Eusociality in Corbiculate Bees. *Molecular Biology and Evolution*, 33(3):670–678.   *Cited at page 91*

Romiguier, J., Ranwez, V., Delsuc, F., Galtier, N., and Douzery, E. J. P. (2013). Less Is More in Mammalian Phylogenomics: AT-Rich Genes Minimize Tree Conflicts and Unravel the Root of Placental Mammals. *Molecular Biology and Evolution*, 30(9):2134–2144.   *Cited at page 91*

Romiguier, J., Ranwez, V., Douzery, E. J., and Galtier, N. (2010). Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes. *Genome Research*, 20(8):1001–1009.
Cited at pages 95, 99

Romiguier, J. and Roux, C. (2017). Analytical Biases Associated with GC-Content in Molecular Evolution. *Frontiers in Genetics*, 8.   *Cited at page 91*

Rosu, S., Libuda, D. E., and Villeneuve, A. M. (2011). Robust Crossover Assurance and Regulated Interhomolog Access Maintain Meiotic Crossover Number. *Science (New York, N.y.)*, 334(6060):1286–1289.   *Cited at page 56*

Rousselle, M. (2018). *Estimation et Analyse Du Taux de Substitution Adaptatif Chez Les Animaux*. Thesis, Montpellier.   *Cited at page 93*

Rousselle, M., Laverré, A., Figuet, E., Nabholz, B., and Galtier, N. (2019). Influence of Recombination and GC-biased Gene Conversion on the Adaptive and Nonadaptive Substitution Rate in Mammals versus Birds. *Molecular Biology and Evolution*, 36(3):458–471.   *Cited at page 91*

Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J., and Birkenmeier, E. H. (1994). Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mammalian Genome*, 5(5):253–274.   *Cited at page 49*

Rudolph, K. L. M., Schmitt, B. M., Villar, D., White, R. J., Marioni, J. C., Kutter, C., and Odom, D. T. (2016). Codon-Driven Translational Efficiency Is Stable across Diverse Mammalian Cell States. *PLoS genetics*, 12(5):e1006024.
Cited at page 90

Russell, B. (1912). *The Problems of Philosophy*. Henry Holt and Company.
Cited at pages 197, 198

Russell, B. (1914). *The relation of sense-data to physics*. Nicola Zanichelli.
Cited at page 197

Russell, B. (1918). *The Philosophy of Logical Atomism.* The Monist. *Cited at page 198*

Rutherford, E. (1939). *Forty Years of Physics*, volume 14. Cambridge University Press. Background to Modern Science. Ten Lectures at Cambridge arranged by the History of Science Committee. *Cited at page 233*

Saintenac, C., Faure, S., Remay, A., Choulet, F., Ravel, C., Paux, E., Balfourier, F., Feuillet, C., and Sourdille, P. (2011). Variation in crossover rates across a 3-Mb contig of bread wheat (Triticum aestivum) reveals the presence of a meiotic recombination hotspot. *Chromosoma*, 120(2):185–198. *Cited at page 63*

Sandor, C., Li, W., Coppieters, W., Druet, T., Charlier, C., and Georges, M. (2012). Genetic Variants in REC8, RNF212, and PRDM9 Influence Male Recombination in Cattle. *PLoS Genetics*, 8(7). *Cited at page 68*

Sarbajna, S., Denniff, M., Jeffreys, A. J., Neumann, R., Soler Artigas, M., Veselis, A., and May, C. A. (2012). A major recombination hotspot in the XqYq pseudoautosomal region gives new insight into processing of human gene conversion events. *Human Molecular Genetics*, 21(9):2029–2038. *Cited at page 66*

Sasaki, M., Lange, J., and Keeney, S. (2010). Genome destabilization by homologous recombination in the germ line. *Nature Reviews. Molecular Cell Biology*, 11(3):182–195. *Cited at pages 46, 60*

Šatava, J. (1918). Polahvní formy kvasnic. *Pivoarské Listy*, 36:79–81. *Cited at page 14*

Schalk, J. A., Dietrich, A. J., Vink, A. C., Offenberg, H. H., van Aalderen, M., and Heyting, C. (1998). Localization of scp2 and scp3 protein molecules within synaptonemal complexes of the rat. *Chromosoma*, 107(8):540–548. *Cited at page 32*

Scherthan, H., Weich, S., Schwegler, H., Heyting, C., Härle, M., and Cremer, T. (1996). Centromere and telomere movements during early meiotic prophase of mouse and man are associated with the onset of chromosome pairing. *The Journal of Cell Biology*, 134(5):1109–1125. *Cited at page 28*

Schmekel, K. and Daneholt, B. (1995). The central region of the synaptonemal complex revealed in three dimensions. *Trends in cell biology*, 5(6):239–242. *Cited at page 32*

Schmekel, K. and Daneholt, B. (1998). Evidence for close contact between recombination nodules and the central element of the synaptonemal complex. *Chromosome Res.*, 6(3):155–159. *Cited at page 34*

Schmitt, K., Lazzeroni, L. C., Foote, S., Vollrath, D., Fisher, E. M., Goradia, T. M., Lange, K., Page, D. C., and Arnheim, N. (1994). Multipoint linkage map of the human pseudoautosomal region, based on single-sperm typing: Do double crossovers occur during male meiosis? *American Journal of Human Genetics*, 55(3):423–430. *Cited at page 53*

Schneider, J. A., Peto, T. E. A., Boone, R. A., Boyce, A. J., and Clegg, J. B. (2002). Direct measurement of the male recombination fraction in the human beta-globin hot spot. *Human Molecular Genetics*, 11(3):207–215. *Cited at page 63*

Schoenmakers, S., Wassenaar, E., Hoogerbrugge, J. W., Laven, J. S. E., Grootegoed, J. A., and Baarends, W. M. (2009). Female meiotic sex chromosome inactivation in chicken. *PLoS genetics*, 5(5):e1000466. *Cited at page 35*

Schuchert, P. and Kohli, J. (1988). The Ade6-M26 Mutation of Schizosaccharomyces Pombe Increases the Frequency of Crossing over. *Genetics*, 119(3):507–515. *Cited at page 73*

Schwacha, A. and Kleckner, N. (1994). Identification of joint molecules that form frequently between homologs but rarely between sister chromatids during yeast meiosis. *Cell*, 76(1):51–63. *Cited at page 37*

Schwacha, A. and Kleckner, N. (1995). Identification of double Holliday junctions as intermediates in meiotic recombination. *Cell*, 83(5):783–791. *Cited at page 37*

Schwacha, A. and Kleckner, N. (1997). Interhomolog bias during meiotic recombination: meiotic functions promote a highly differentiated interhomolog-only pathway. *Cell*, 90(6):1123–1135. *Cited at pages 33, 59*

Schwartz, J. J., Roach, D. J., Thomas, J. H., and Shendure, J. (2014). Primate evolution of the recombination regulator PRDM9. *Nature communications*, 5:4370. *Cited at pages 68, 71*

Schwarz, T., Striedner, Y., Horner, A., Haase, K., Kemptner, J., Zeppezauer, N., Hermann, P., and Tiemann-Boege, I. (2019). PRDM9 forms a trimer by interactions within the zinc finger array. *Life Science Alliance*, 2(4):e201800291. *Cited at page 70*

Sciurano, R., Rahn, M., Rey-Valzacchi, G., and Solari, A. J. (2007). The asynaptic chromatin in spermatocytes of translocation carriers contains the histone variant gamma-H2AX and associates with the XY body. *Human Reproduction (Oxford, England)*, 22(1):142–150. *Cited at page 35*

Scully, R., Chen, J., Plug, A., Xiao, Y., Weaver, D., Feunteun, J., Ashley, T., and Livingston, D. M. (1997). Association of BRCA1 with Rad51 in Mitotic and Meiotic Cells. *Cell*, 88(2):265–275. *Cited at pages 40, 42*

Sebestova, H., Vozdova, M., Kubickova, S., Cernohorska, H., Kotrba, R., and Rubes, J. (2016). Effect of species-specific differences in chromosome morphology on chromatin compaction and the frequency and distribution of RAD51 and MLH1 foci in two bovid species: Cattle (Bos taurus) and the common eland (Taurotragus oryx). *Chromosoma*, 125(1):137–149. *Cited at page 59*

Séguéla-Arnaud, M., Crismani, W., Larchevêque, C., Mazel, J., Froger, N., Choinard, S., Lemhemdi, A., Macaisne, N., Van Leene, J., Gevaert, K., De Jaeger, G., Chelysheva, L., and Mercier, R. (2015). Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM. *Proceedings of the National Academy of Sciences of the United States of America*, 112(15):4713–4718. *Cited at page 58*

Segura, J., Ferretti, L., Ramos-Onsins, S., Capilla, L., Farré, M., Reis, F., Oliver-Bonet, M., Fernández-Bellón, H., Garcia, F., Garcia-Caldés, M., Robinson, T. J., and Ruiz-Herrera, A. (2013). Evolution of recombination in eutherian mammals: Insights into mechanisms that affect recombination rates and crossover interference. *Proceedings of the Royal Society B: Biological Sciences*, 280(1771):20131945. *Cited at page 59*

Ségurel, L., Leffler, E. M., and Przeworski, M. (2011). The Case of the Fickle Fingers: How the PRDM9 Zinc Finger Protein Specifies Meiotic Recombination Hotspots in Humans. *PLoS Biology*, 9(12). *Cited at page 71*

Sémon, M., Lobry, J. R., and Duret, L. (2006). No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Molecular Biology and Evolution*, 23(3):523–529. *Cited at pages 83, 90*

Sémon, M., Mouchiroud, D., and Duret, L. (2005). Relationship between gene expression and GC-content in mammals: Statistical significance and biological relevance. *Human Molecular Genetics*, 14(3):421–427. *Cited at page 83*

Serres-Giardi, L., Belkhir, K., David, J., and Glémin, S. (2012). Patterns and Evolution of Nucleotide Landscapes in Seed Plants[W]. *The Plant Cell*, 24(4):1379–1397. *Cited at page 95*

Sharp, P. (1982). Sex chromosome pairing during male meiosis in marsupials. *Chromosoma*, 86(1):27–47. *Cited at page 55*

Sharp, P. M., Averof, M., Lloyd, A. T., Matassi, G., and Peden, J. F. (1995). DNA sequence evolution: The sounds of silence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 349(1329):241–247. *Cited at page 90*

Sharp, P. M. and Devine, K. M. (1989). Codon usage and gene expression level in Dictyosteiium discoidtum : Highly expressed genes do [prefer[optimal codons. *Nucleic Acids Research*, 17(13):5029–5040. *Cited at page 90*

Sharp, P. M. and Li, W. H. (1987). The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Molecular Biology and Evolution*, 4(3):222–230. *Cited at page 90*

Sharp, P. M., Stenico, M., Peden, J. F., and Lloyd, A. T. (1993). Codon usage: Mutational bias, translational selection, or both? *Biochemical Society Transactions*, 21(4):835–841. *Cited at page 89*

Shaw, P. and Moore, G. (1998). Meiosis: Vive la difference! *Current Opinion in Plant Biology*, 1(6):458–462. *Cited at page 58*

Shields, D. C., Sharp, P. M., Higgins, D. G., and Wright, F. (1988). "Silent" sites in Drosophila genes are not neutral: Evidence of selection among synonymous codons. *Molecular Biology and Evolution*, 5(6):704–716. *Cited at page 89*

Shifman, S., Bell, J. T., Copley, R. R., Taylor, M. S., Williams, R. W., Mott, R., and Flint, J. (2006). A High-Resolution Single Nucleotide Polymorphism Genetic Map of the Mouse Genome. *PLOS Biology*, 4(12):e395. *Cited at pages 51, 64, 188*

Shin, Y.-H., Choi, Y., Erdin, S. U., Yatsenko, S. A., Kloc, M., Yang, F., Wang, P. J., Meistrich, M. L., and Rajkovic, A. (2010). Hormad1 mutation disrupts synaptonemal complex formation, recombination, and chromosome segregation in mammalian meiosis. *PLoS genetics*, 6(11):e1001190. *Cited at page 39*

Shinohara, M., Oh, S. D., Hunter, N., and Shinohara, A. (2008). Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. *Nature Genetics*, 40(3):299–309. *Cited at pages 57, 58*

Shinohara, M., Sakai, K., Shinohara, A., and Bishop, D. K. (2003). Crossover interference in Saccharomyces cerevisiae requires a TID1/RDH54- and DMC1-dependent pathway. *Genetics*, 163(4):1273–1286.   *Cited at page 57*

Shiroishi, T., Hanzawa, N., Sagai, T., Ishiura, M., Gojobori, T., Steinmetz, M., and Moriwaki, K. (1990). Recombinational hotspot specific to female meiosis in the mouse major histocompatibility complex. *Immunogenetics*, 31(2):79–88.
             Cited at page 65

Shiroishi, T., Sagai, T., Hanzawa, N., Gotoh, H., and Moriwaki, K. (1991). Genetic control of sex-dependent meiotic recombination in the major histocompatibility complex of the mouse. *The EMBO journal*, 10(3):681–686.   *Cited at page 65*

Shiroishi, T., Sagai, T., and Moriwaki, K. (1982). A new wild-derived H–2 haplotype enhancing K–IA recombination. *Nature*, 300(5890):370.   *Cited at page 68*

Shodhan, A., Lukaszewicz, A., Novatchkova, M., and Loidl, J. (2014). Msh4 and Msh5 Function in SC-Independent Chiasma Formation During the Streamlined Meiosis of Tetrahymena. *Genetics*, 198(3):983–993.   *Cited at page 57*

Si, W., Yuan, Y., Huang, J., Zhang, X., Zhang, Y., Zhang, Y., Tian, D., Wang, C., Yang, Y., and Yang, S. (2015). Widely distributed hot and cold spots in meiotic recombination as shown by the sequencing of rice F2 plants. *New Phytologist*, 206(4):1491–1502.   *Cited at pages , 80, 96, 103*

Siddiqi, O. (1962). Mutagenic action of nitrous acid on aspergillus nidulans. *Genetics Research*, 3(2):303–314.   *Cited at page 18*

Sigurdsson, M. I., Smith, A. V., Bjornsson, H. T., and Jonsson, J. J. (2009). HapMap methylation-associated SNPs, markers of germline DNA methylation, positively correlate with regional levels of human meiotic recombination. *Genome Research*, 19(4):581–589.   *Cited at page 61*

Simpson, G. G. (1944). *Tempo and mode in evolution.* Number 15. Columbia University Press.   *Cited at page 201*

Singh, N. D. (2012). Classical Genetics Meets Next-Generation Sequencing: Uncovering a Genome-Wide Recombination Map in Drosophila melanogaster. *PLOS Genetics*, 8(10):e1003024.   *Cited at page 64*

Singhal, S., Leffler, E. M., Sannareddy, K., Turner, I., Venn, O., Hooper, D. M., Strand, A. I., Li, Q., Raney, B., Balakrishnan, C. N., Griffith, S. C., McVean, G., and Przeworski, M. (2015). Stable recombination hotspots in birds. *Science (New York, N.Y.)*, 350(6263):928–932.   *Cited at pages 62, 69*

Sjögren, C. and Ström, L. (2010). S-phase and DNA damage activated establishment of sister chromatid cohesion–importance for DNA repair. *Exp. Cell Res.*, 316(9):1445–1453.   *Cited at page 33*

Slatkin, M. (2008). Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9(6):477–485.
             Cited at page 52

Smagulova, F., Brick, K., Pu, Y., Camerini-Otero, R. D., and Petukhova, G. V. (2016). The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes & Development*, 30(3):266–280.
             Cited at pages 77, 128, 129, 130, 131, 134, 148, 154, 183, 223

Smagulova, F., Brick, K., Pu, Y., Sengupta, U., Camerini-Otero, R. D., and Petukhova, G. V. (2013). Suppression of genetic recombination in the pseudoauto-somal region and at subtelomeres in mice with a hypomorphic Spo11 allele. *BMC Genomics*, 14:493. *Cited at page 39*

Smagulova, F., Gregoretti, I. V., Brick, K., Khil, P., Camerini-Otero, R. D., and Petukhova, G. V. (2011). Genome-Wide Analysis Reveals Novel Molecular Features of Mouse Recombination Hotspots. *Nature*, 472(7343):375–378. *Cited at pages 54, 63, 70, 126*

Smeds, L., Mugal, C. F., Qvarnström, A., and Ellegren, H. (2016). High-Resolution Mapping of Crossover and Non-crossover Recombination Events by Whole-Genome Re-sequencing of an Avian Pedigree. *PLoS genetics*, 12(5):e1006044. *Cited at pages , 80, 96, 103, 120, 182*

Smit, A. F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Current Opinion in Genetics & Development*, 9(6):657–663. *Cited at page 82*

Smith, G. R. (2001). Homologous Recombination Near and Far from DNA Breaks: Alternative Roles and Contrasting Views. *Annual Review of Genetics*, 35(1):243–274. *Cited at page 72*

Smith, J., Baldeyron, C., De Oliveira, I., Sala-Trepat, M., and Papadopoulo, D. (2001). The influence of DNA double-strand break structure on end-joining in human cells. *Nucleic Acids Research*, 29(23):4783–4792. *Cited at page 46*

Smith, J. M. (1988). *Games, sex and evolution*. Prentice Hall. *Cited at page 214*

Smith, K. N. and Nicolas, A. (1998). Recombination at work for meiosis. *Current Opinion in Genetics & Development*, 8(2):200–211. *Cited at page 36*

Smith, N. G. and Eyre-Walker, A. (2001). Synonymous codon bias is not caused by mutation bias in G+C-rich genes in humans. *Molecular Biology and Evolution*, 18(6):982–986. *Cited at page 82*

Smith, N. G. C. and Eyre-Walker, A. (2002). The compositional evolution of the murid genome. *Journal of Molecular Evolution*, 55(2):197–201. *Cited at page 82*

Smith, T. C. A., Arndt, P. F., and Eyre-Walker, A. (2018). Large scale variation in the rate of germ-line de novo mutation, base composition, divergence and diversity in humans. *PLoS genetics*, 14(3):e1007254. *Cited at page 61*

Smith, T. F. and Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195–197. *Cited at page 120*

Smukowski, C. S. and Noor, M. a. F. (2011). Recombination rate variation in closely related species. *Heredity*, 107(6):496–508. *Cited at page 62*

Snoek, M., Teuscher, C., and van Vugt, H. (1998). Molecular analysis of the major MHC recombinational hot spot located within the G7c gene of the murine class III region that is involved in disease susceptibility. *Journal of Immunology (Baltimore, Md.: 1950)*, 160(1):266–272. *Cited at page 217*

Snowden, T., Acharya, S., Butz, C., Berardini, M., and Fishel, R. (2004). hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. *Molecular Cell*, 15(3):437–451. *Cited at page 42*

Soriano, P., Keitges, E. A., Schorderet, D. F., Harbers, K., Gartler, S. M., and Jaenisch, R. (1987). High rate of recombination and double crossovers in the mouse pseudoautosomal region during male meiosis. *Proceedings of the National Academy of Sciences of the United States of America*, 84(20):7218–7220. *Cited at page 86*

Speijer, D. (2016). What can we infer about the origin of sex in early eukaryotes? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1706):20150530. *Cited at page 24*

Speijer, D., Lukeš, J., and Eliáš, M. (2015). Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. *Proceedings of the National Academy of Sciences*, 112(29):8827–8834. *Cited at page 25*

Spell, R. M. and Jinks-Robertson, S. (2004). Examination of the Roles of Sgs1 and Srs2 Helicases in the Enforcement of Recombination Fidelity in Saccharomyces cerevisiae. *Genetics*, 168(4):1855–1865. *Cited at page 45*

Spencer, C. C. A., Deloukas, P., Hunt, S., Mullikin, J., Myers, S., Silverman, B., Donnelly, P., Bentley, D., and McVean, G. (2006). The Influence of Recombination on Human Genetic Diversity. *PLOS Genetics*, 2(9):e148. *Cited at pages 61, 82*

Spies, M. and Fishel, R. (2015). Mismatch Repair during Homologous and Homeologous Recombination. *Cold Spring Harbor Perspectives in Biology*, 7(3):a022657. *Cited at page 83*

Stacey, N. J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., Maxwell, A., Roberts, K., and Sugimoto-Shirasu, K. (2006). Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *The Plant Journal: For Cell and Molecular Biology*, 48(2):206–216. *Cited at page 39*

Stadler, D. and Towe, A. M. (1963). Recombination of allelic cysteine mutants in neurospora. *Genetics*, 48(10):1323. *Cited at page 18*

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., and Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1736). *Cited at page 63*

Stebbins, G. L. (1966). *Processes of Organic Evolution.[With Illustrations.].* Englewood Cliffs. *Cited at page 21*

Steiner, C. C. and Ryder, O. A. (2013). Characterization of Prdm9 in equids and sterility in mules. *PloS One*, 8(4):e61746. *Cited at page 68*

Steiner, W. W., Schreckhise, R. W., and Smith, G. R. (2002). Meiotic DNA breaks at the S. pombe recombination hot spot M26. *Molecular Cell*, 9(4):847–855. *Cited at page 39*

Steiner, W. W. and Smith, G. R. (2005). Natural Meiotic Recombination Hot Spots in the Schizosaccharomyces pombe Genome Successfully Predicted from the Simple Sequence Motif M26. *Molecular and Cellular Biology*, 25(20):9054–9062. *Cited at page 63*

Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Srelinger, J., Wake, C., Long, E., Mach, B., and Hood, L. (1982). A molecular map of the immune response region from the major histocompatibility complex of the mouse. *Nature*, 300(5887):35–42.   *Cited at pages 63, 217*

Stevison, L. S., Woerner, A. E., Kidd, J. M., Kelley, J. L., Veeramah, K. R., McManus, K. F., Great Ape Genome Project, Bustamante, C. D., Hammer, M. F., and Wall, J. D. (2016). The Time Scale of Recombination Rate Evolution in Great Apes. *Molecular Biology and Evolution*, 33(4):928–945.   *Cited at page 67*

Storms, R. and Hastings, P. J. (1977). A fine structure analysis of meiotic pairing in Chlamydomonas reinhardi. *Exp. Cell Res.*, 104(1):39–46.   *Cited at page 32*

Strathern, J. N., Shafer, B. K., and McGill, C. B. (1995). DNA synthesis errors associated with double-strand-break repair. *Genetics*, 140(3):965–972.
Cited at page 61

Strickland, W. N. (1958). An analysis of interference in Aspergillus nidulans. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 149(934):82–101.   *Cited at page 58*

Stumpf, M. P. H. and McVean, G. A. T. (2003). Estimating recombination rates from population-genetic data. *Nature Reviews Genetics*, 4(12):959. *Cited at pages 51, 52*

Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *Journal of Experimental Zoology*, 14(1):43–59.   *Cited at page 49*

Sturtevant, A. H. (1915). The behavior of the chromosomes as studied through linkage. *Zeitschrift für induktive Abstammungs- und Vererbungslehre*, 13(1):234–287.   *Cited at page 56*

Subramanian, S. (2008). Nearly Neutrality and the Evolution of Codon Usage Bias in Eukaryotic Genomes. *Genetics*, 178(4):2429–2432.   *Cited at page 90*

Sueoka, N. (1962). On the genetic basis of variation and heterogeneity of DNA base composition. *Proceedings of the National Academy of Sciences of the United States of America*, 48(4):582–592.   *Cited at page 87*

Sugawara, H., Iwabata, K., Koshiyama, A., Yanai, T., Daikuhara, Y., Namekawa, S. H., Hamada, F. N., and Sakaguchi, K. (2009). Coprinus cinereus Mer3 is required for synaptonemal complex formation during meiosis. *Chromosoma*, 118(1):127–139.   *Cited at page 164*

Sun, F., Trpkov, K., Rademaker, A., Ko, E., and Martin, R. H. (2005). Variation in meiotic recombination frequencies among human males. *Human Genetics*, 116(3):172–178.   *Cited at page 59*

Sun, H., Treco, D., Schultes, N. P., and Szostak, J. W. (1989a). Double-strand breaks at an initiation site for meiotic gene conversion. *Nature*, 338(6210):87–90.
Cited at pages 37, 39

Sun, H., Treco, D., Schultes, N. P., and Szostak, J. W. (1989b). Double-strand breaks at an initiation site for meiotic gene conversion. *Nature*, 338(6210):87.
Cited at page 63

Sung, P. and Klein, H. (2006). Mechanism of homologous recombination: Mediators and helicases take on regulatory functions. *Nature Reviews. Molecular Cell Biology*, 7(10):739–750. *Cited at page 45*

Sung, P., Krejci, L., Van Komen, S., and Sehorn, M. G. (2003). Rad51 recombinase and recombination mediators. *The Journal of Biological Chemistry*, 278(44):42729–42732. *Cited at page 40*

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. (2013). Approximate Bayesian Computation. *PLOS Computational Biology*, 9(1):e1002803. *Cited at page 136*

Surtees, J. A., Argueso, J. L., and Alani, E. (2004). Mismatch repair proteins: Key regulators of genetic recombination. *Cytogenetic and Genome Research*, 107(3-4):146–159. *Cited at pages 42, 84*

Sutter, N. B., Eberle, M. A., Parker, H. G., Pullar, B. J., Kirkness, E. F., Kruglyak, L., and Ostrander, E. A. (2004). Extensive and breed-specific linkage disequilibrium in Canis familiaris. *Genome Research*, 14(12):2388–2396. *Cited at page 52*

Sutton, W. S. (1902). On the morphology of the chromosome group in *Brachystola magna*. *Biological Bulletin*. *Cited at page 10*

Sym, M. and Roeder, G. S. (1994). Crossover interference is abolished in the absence of a synaptonemal complex protein. *Cell*, 79(2):283–292. *Cited at page 57*

Székvölgyi, L. and Nicolas, A. (2010). From meiosis to postmeiotic events: Homologous recombination is obligatory but flexible. *The FEBS Journal*, 277(3):571–589. *Cited at page 36*

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell*, 33(1):25–35. *Cited at pages 20, 37*

Tajima, F. (1996). The Amount of DNA Polymorphism Maintained in a Finite Population When the Neutral Mutation Rate Varies among Sites. *Genetics*, 143(3):1457–1465. *Cited at page 99*

Takahata, N. (1993). Allelic genealogy and human evolution. *Molecular Biology and Evolution*, 10(1):2–22. *Cited at pages , 103*

Takana, Y (田中), b. (1914). Sexual dimorphism of gametic series in the reduplication (生殖細胞式の雌雄ニ型). 札幌博物学会会報, 5(2):61–64. *Cited at page 64*

Tanaka, K., Miyamoto, N., Shouguchi-Miyata, J., and Ikeda, J.-E. (2006). HFM1, the human homologue of yeast Mer3, encodes a putative DNA helicase expressed specifically in germ-line cells. *DNA sequence: the journal of DNA sequencing and mapping*, 17(3):242–246. *Cited at page 164*

Tapper, W., Collins, A., Gibson, J., Maniatis, N., Ennis, S., and Morton, N. E. (2005). A map of the human genome in linkage disequilibrium units. *Proceedings of the National Academy of Sciences*, 102(33):11835–11839. *Cited at page 66*

Taylor, J. H., Woods, P. S., and Hughes, W. L. (1957). The organization and duplication of chromosomes as revealed by autoradiographic studies using tritium-labeled thymidine. *Proceedings of the National Academy of Sciences of the United States of America*, 43(1):122. *Cited at page 17*

Tease, C. and Hultén, M. A. (2004). Inter-sex variation in synaptonemal complex lengths largely determine the different recombination rates in male and female germ cells. *Cytogenetic and Genome Research*, 107(3-4):208–215. *Cited at page 65*

Tessé, S., Storlazzi, A., Kleckner, N., Gargano, S., and Zickler, D. (2003). Localization and roles of Ski8p protein in Sordaria meiosis and delineation of three mechanistically distinct steps of meiotic homolog juxtaposition. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22):12865–12870. *Cited at page 32*

The 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467(7319):1061–1073. *Cited at page 62*

The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74. *Cited at page 52*

The International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164):851–861. *Cited at pages 52, 63*

Thibault-Sennett, S., Yu, Q., Smagulova, F., Cloutier, J., Brick, K., Camerini-Otero, R. D., and Petukhova, G. V. (2018). Interrogating the Functions of PRDM9 Domains in Meiosis. *Genetics*, 209(2):475–487. *Cited at page 69*

Thiery, J.-P., Macaya, G., and Bernardi, G. (1976). An analysis of eukaryotic genomes by density gradient centrifugation. *Journal of Molecular Biology*, 108(1):219–235. *Cited at page 81*

Thomas, J. H., Emerson, R. O., and Shendure, J. (2009). Extraordinary molecular evolution in the PRDM9 fertility gene. *PloS One*, 4(12):e8505. *Cited at page 71*

Thorslund, T., Esashi, F., and West, S. C. (2007). Interactions between human BRCA2 protein and the meiosis-specific recombinase DMC1. *The EMBO Journal*, 26(12):2915–2922. *Cited at page 40*

Thuriaux, P. (1977). Is recombination confined to structural genes on the eukaryotic genome? *Nature*, 268(5619):460. *Cited at page 61*

Tiemann-Boege, I., Calabrese, P., Cochran, D. M., Sokol, R., and Arnheim, N. (2006). High-resolution recombination patterns in a region of human chromosome 21 measured by sperm typing. *PLoS genetics*, 2(5):e70. *Cited at page 63*

Tiemann-Boege, I., Schwarz, T., Striedner, Y., and Heissl, A. (2017). The consequences of sequence erosion in the evolution of recombination hotspots. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1736). *Cited at page 72*

Tsai, J. H. and McKee, B. D. (2011). Homologous pairing and the role of pairing centers in meiosis. *J. Cell. Sci.*, 124(Pt 12):1955–1963. *Cited at page 34*

Turner, J. M. A., Mahadevaiah, S. K., Fernandez-Capetillo, O., Nussenzweig, A., Xu, X., Deng, C.-X., and Burgoyne, P. S. (2005). Silencing of unsynapsed meiotic chromosomes in the mouse. *Nature Genetics*, 37(1):41–47. *Cited at page 35*

Tusié-Luna, M. T. and White, P. C. (1995). Gene conversions and unequal crossovers between CYP21 (steroid 21-hydroxylase gene) and CYP21P involve different mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, 92(23):10796–10800. *Cited at page 53*

Tzur, Y. B., Wilson, K. L., and Gruenbaum, Y. (2006). Sun-domain proteins:'velcro'that links the nucleoskeleton to the cytoskeleton. *Nature reviews Molecular cell biology*, 7(10):782. *Cited at page 28*

Úbeda, F. and Wilkins, J. F. (2011). The Red Queen theory of recombination hotspots. *Journal of Evolutionary Biology*, 24(3):541–553. *Cited at page 74*

Urrutia, A. O. and Hurst, L. D. (2003). The signature of selection mediated by expression on human genes. *Genome Research*, 13(10):2260–2264. *Cited at page 90*

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., and DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, 11(1110):11.10.1–11.10.33. *Cited at page 116*

van Valen, L. (1973). A New Evolutionary Law. *ResearchGate*. *Cited at page 74*

van Veen, J. E. and Hawley, R. S. (2003). Meiosis: When Even Two Is a Crowd. *Current Biology*, 13(21):R831–R833. *Cited at page 57*

Venn, O. C. (2013). *Inferring the Fine-Scale Structure and Evolution of Recombination from High-Throughput Genome Sequencing*. Ph.D., University of Oxford. *Cited at page 51*

Verardi, A., Lucchini, V., and Randi, E. (2006). Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (Canis lupus) by admixture linkage disequilibrium analysis. *Molecular Ecology*, 15(10):2845–2855. *Cited at page 52*

Villeneuve, A. M. (1994). A cis-acting locus that promotes crossing over between x chromosomes in caenorhabditis elegans. *Genetics*, 136(3):887–902. *Cited at page 30*

Vincent, M.-C., Daudin, M., De, M. P., Massat, G., Mieusset, R., Pontonnier, F., Calvas, P., Bujan, L., and Bourrouillout, G. (2002 Jan-Feb). Cytogenetic investigations of infertile men with low sperm counts: A 25-year experience. *Journal of Andrology*, 23(1):18–22; discussion 44–45. *Cited at page 35*

Vinogradov, A. E. (2003). Isochores and tissue-specificity. *Nucleic Acids Research*, 31(17):5212–5220. *Cited at page 83*

Vinogradov, A. E. (2005). Noncoding DNA, isochores and gene expression: Nucleosome formation potential. *Nucleic Acids Research*, 33(2):559–563. *Cited at page 83*

Vorms, M. (2013). Models of data and theoretical hypotheses: A case-study in classical genetics. *Synthese*, 190(2):293–319. *Cited at page 50*

Vozdova, M., Sebestova, H., Kubickova, S., Cernohorska, H., Vahala, J., and Rubes, J. (2013). A Comparative Study of Meiotic Recombination in Cattle (Bos taurus) and Three Wildebeest Species (Connochaetes gnou, C. taurinus taurinus and C. t. albojubatus). *Cytogenetic and Genome Research*, 140(1):36–45. *Cited at page 59*

Vrijenhoek, R. C. (1998). Animal clones and diversity: Are natural clones generalists or specialists? *BioScience*, 48(8):617–628. *Cited at page 24*

Vrijenhoek, R. C., Dawley, R. M., Cole, C. J., and Bogart, J. (1989). A list of the known unisexual vertebrates. *Evolution and ecology of unisexual vertebrates*, 466:19–23. *Cited at page 24*

Wahls, W. P. and Davidson, M. K. (2011). DNA Sequence-Mediated, Evolutionarily Rapid Redistribution of Meiotic Recombination Hotspots: Commentary on Genetics 182: 459–469 and Genetics 187: 385–396. *Genetics*, 189(3):685–694. *Cited at page 68*

Wahls, W. P., Swenson, G., and Moore, P. D. (1991). Two hypervarlable minisatellite DNA binding proteins. *Nucleic Acids Research*, 19(12):3269–3274. *Cited at page 68*

Wall, J. D., Frisse, L. A., Hudson, R. R., and Di Rienzo, A. (2003). Comparative linkage-disequilibrium analysis of the beta-globin hotspot in primates. *American Journal of Human Genetics*, 73(6):1330–1340. *Cited at page 67*

Wallberg, A., Glémin, S., and Webster, M. T. (2015). Extreme Recombination Frequencies Shape Genome Variation and Evolution in the Honeybee, Apis mellifera. *PLOS Genetics*, 11(4):e1005189. *Cited at pages 63, 95*

Wang, J., Fan, H. C., Behr, B., and Quake, S. R. (2012a). Genome-wide Single-Cell Analysis of Recombination Activity and De Novo Mutation Rates in Human Sperm. *Cell*, 150(2):402–412. *Cited at page 53*

Wang, S., Zickler, D., Kleckner, N., and Zhang, L. (2015). Meiotic crossover patterns: Obligatory crossover, interference and homeostasis in a single process. *Cell Cycle*, 14(3):305. *Cited at page 57*

Wang, X. V., Blades, N., Ding, J., Sultana, R., and Parmigiani, G. (2012b). Estimation of sequencing error rates in short reads. *BMC Bioinformatics*, 13(1):185. *Cited at page 118*

Watanabe, Y., Fujiyama, A., Ichiba, Y., Hattori, M., Yada, T., Sakaki, Y., and Ikemura, T. (2002). Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: Disease-related genes in timing-switch regions. *Human Molecular Genetics*, 11(1):13–21. *Cited at page 81*

Watson, J. D. and Crick, F. H. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738. *Cited at pages 3, 17, 19*

Webb, A. J., Berg, I. L., and Jeffreys, A. (2008). Sperm cross-over activity in regions of the human genome showing extreme breakdown of marker association. *Proceedings of the National Academy of Sciences of the United States of America*, 105(30):10471–10476. *Cited at pages 67, 190*

Weber, C. C., Boussau, B., Romiguier, J., Jarvis, E. D., and Ellegren, H. (2014). Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. *Genome Biology*, 15:549. *Cited at pages 95, 99*

Webster, M. T., Axelsson, E., and Ellegren, H. (2006). Strong regional biases in nucleotide substitution in the chicken genome. *Molecular Biology and Evolution*, 23(6):1203–1216. *Cited at page 95*

Webster, M. T. and Hurst, L. D. (2012). Direct and indirect consequences of meiotic recombination: Implications for genome evolution. *Trends in Genetics*, 28(3):101–109. *Cited at pages 61, 91*

Webster, M. T. and Smith, N. G. C. (2004). Fixation biases affecting human SNPs. *Trends in Genetics*, 20(3):122–126. *Cited at page 82*

Webster, M. T., Smith, N. G. C., Hultin-Rosenberg, L., Arndt, P. F., and Ellegren, H. (2005). Male-driven biased gene conversion governs the evolution of base composition in human alu repeats. *Molecular Biology and Evolution*, 22(6):1468–1474. *Cited at page 88*

Wei, K., Clark, A. B., Wong, E., Kane, M. F., Mazur, D. J., Parris, T., Kolas, N. K., Russell, R., Hou, H., Kneitz, B., Yang, G., Kunkel, T. A., Kolodner, R. D., Cohen, P. E., and Edelmann, W. (2003). Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. *Genes & Development*, 17(5):603–614. *Cited at page 44*

Weterings, E. and van Gent, D. C. (2004). The mechanism of non-homologous end-joining: A synopsis of synapsis. *DNA repair*, 3(11):1425–1435. *Cited at page 46*

Whewell, W. (1837). *History of Inductive Sciences: From the Earliest to the Present Times*, volume 1. London: John W. Parker, West Strand. *Cited at page 7*

White, E. P., Ernest, S. K. M., Kerkhoff, A. J., and Enquist, B. J. (2007). Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution*, 22(6):323–330. *Cited at page 99*

Whitehead, A. N. and Russell, B. (1912). *Principia mathematica*, volume 2. University Press. *Cited at page 198*

Whitehouse, H. L. (1963). A theory of crossing-over by means of hybrid deoxyribonucleic acid. *Nature*, 199(4898):1034. *Cited at page 19*

Whitehouse, H. L. (1965). Crossing-over. *Science Progress (1933-)*, pages 285–296. *Cited at page 19*

Wilkins, A. S. and Holliday, R. (2009). The evolution of meiosis from mitosis. *Genetics*, 181(1):3–12. *Cited at page 25*

Wilkins, M. H. F., Stokes, A. R., and Wilson, H. R. (1953). Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature*, 171(4356):738. *Cited at pages 3, 19*

Williams, A. L., Genovese, G., Dyer, T., Altemose, N., Truax, K., Jun, G., Patterson, N., Myers, S. R., Curran, J. E., Duggirala, R., Blangero, J., Reich, D., and Przeworski, M. (2015). Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *eLife*, 4:e04637. *Cited at pages , 80, 96, 103, 146, 188*

Wilson, E. and Morgan, T. H. (1920). Chiasmatype and crossing over. *The American Naturalist*, 54(632):193–219. *Cited at page 13*

Winckler, W., Myers, S. R., Richter, D. J., Onofrio, R. C., McDonald, G. J., Bontrop, R. E., McVean, G. A. T., Gabriel, S. B., Reich, D., Donnelly, P., and Altshuler, D. (2005). Comparison of Fine-Scale Recombination Rates in Humans and Chimpanzees. *Science*, 308(5718):107–111. *Cited at pages 63, 67, 85*

Winge, Ö. and Laustsen, O. (1937). *On Two Types of Spore Germination, and on Genetic Segregations in Saccharomyces: Demonstrated Through Single-spore Cultures*. imp. de Bianco Luno. *Cited at page 14*

Winkler, H. (1930). *Die konversion der gene*. G. Fischer. *Cited at page 16*

Wolfe, K. H., Sharp, P. M., and Li, W. H. (1989). Mutation rates differ among regions of the mammalian genome. *Nature*, 337(6204):283–285. *Cited at page 82*

Woolfit, M. (2009). Effective population size and the rate and pattern of nucleotide substitutions. *Biology Letters*, 5(3):417–420. *Cited at page 99*

Wright, S. (1932). *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*, volume 1. na. *Cited at page 21*

Wright, S. I., Yau, C. B. K., Looseley, M., and Meyers, B. C. (2004). Effects of gene expression on molecular evolution in Arabidopsis thaliana and Arabidopsis lyrata. *Molecular Biology and Evolution*, 21(9):1719–1726. *Cited at page 90*

Wu, D. Y., Ugozzoli, L., Pal, B. K., and Wallace, R. B. (1989). Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proceedings of the National Academy of Sciences*, 86(8):2757–2760. *Cited at page 53*

Wu, H., Mathioudakis, N., Diagouraga, B., Dong, A., Dombrovski, L., Baudat, F., Cusack, S., de Massy, B., and Kadlec, J. (2013). Molecular Basis for the Regulation of the H3K4 Methyltransferase Activity of PRDM9. *Cell Reports*, 5(1):13–20. *Cited at page 69*

Wu, H.-Y. and Burgess, S. M. (2006). Two distinct surveillance mechanisms monitor meiotic chromosome metabolism in budding yeast. *Current biology: CB*, 16(24):2473–2479. *Cited at page 34*

Wu, L. and Hickson, I. D. (2003). The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature*, 426(6968):870–874. *Cited at page 45*

Wu, T.-C. and Lichten, M. (1994). Meiosis-induced double-strand break sites determined by yeast chromatin structure. *Science*, 263(5146):515. *Cited at page 62*

Wu, T. C. and Lichten, M. (1995). Factors That Affect the Location and Frequency of Meiosis-Induced Double-Strand Breaks in Saccharomyces Cerevisiae. *Genetics*, 140(1):55–66. *Cited at page 56*

Wu, Z. K., Getun, I. V., and Bois, P. R. J. (2010). Anatomy of mouse recombination hot spots. *Nucleic Acids Research*, 38(7):2346–2354. *Cited at page 217*

Wyatt, H. D. and West, S. C. (2014). Holliday Junction Resolvases. *Cold Spring Harbor Perspectives in Biology*, 6(9).   *Cited at pages 43, 44*

Xu, M. and Cook, P. R. (2008). Similar active genes cluster in specialized transcription factories. *The Journal of cell biology*, 181(4):615–623.   *Cited at page 30*

Xue, C., Chen, H., and Yu, F. (2016). Base-Biased Evolution of Disease-Associated Mutations in the Human Genome. *Human Mutation*, 37(11):1209–1214.   *Cited at page 92*

Yalcin, B., Adams, D. J., Flint, J., and Keane, T. M. (2012). Next-generation sequencing of experimental mouse strains. *Mammalian Genome*, 23(9):490–498.   *Cited at pages , 108*

Yamada, S., Kim, S., Tischfield, S. E., Jasin, M., Lange, J., and Keeney, S. (2017). Genomic and chromatin features shaping meiotic double-strand break formation and repair in mice. *Cell Cycle*, 16(20):1870–1884.   *Cited at page 61*

Yanowitz, J. (2010). Meiosis: making a break for it. *Current opinion in cell biology*, 22(6):744–751.   *Cited at page 28*

Yao, H., Zhou, Q., Li, J., Smith, H., Yandeau, M., Nikolau, B. J., and Schnable, P. S. (2002). Molecular characterization of meiotic recombination across the 140-kb multigenic a1-sh2 interval of maize. *Proceedings of the National Academy of Sciences*, 99(9):6157–6162.   *Cited at page 63*

Yauk, C. L., Bois, P. R., and Jeffreys, A. J. (2003). High-resolution sperm typing of meiotic recombination in the mouse MHC Ebeta gene. *The EMBO journal*, 22(6):1389–1397.   *Cited at pages 72, 126*

Yi, S., Summers, T. J., Pearson, N. M., and Li, W.-H. (2004). Recombination Has Little Effect on the Rate of Sequence Divergence in Pseudoautosomal Boundary 1 Among Humans and Great Apes. *Genome Research*, 14(1):37–43. *Cited at page 86*

Yokoo, R., Zawadzki, K. A., Nabeshima, K., Drake, M., Arur, S., and Villeneuve, A. M. (2012). COSA-1 Reveals Separable Licensing and Reinforcement Steps and Efficient Homeostasis Governing Meiotic Crossovers. *Cell*, 149(1):75–87.   *Cited at page 58*

Yoshida, K., Kondoh, G., Matsuda, Y., Habu, T., Nishimune, Y., and Morita, T. (1998). The mouse RecA-like gene Dmc1 is required for homologous chromosome synapsis during meiosis. *Molecular Cell*, 1(5):707–718.   *Cited at page 40*

Youds, J. L. and Boulton, S. J. (2011). The choice in meiosis – defining the factors that influence crossover or non-crossover formation. *J Cell Sci*, 124(4):501–513.   *Cited at pages 57, 58, 59*

Youds, J. L., Mets, D. G., McIlwraith, M. J., Martin, J. S., Ward, J. D., ONeil, N. J., Rose, A. M., West, S. C., Meyer, B. J., and Boulton, S. J. (2010). RTEL-1 Enforces Meiotic Crossover Interference and Homeostasis. *Science*, 327(5970):1254–1258.   *Cited at page 58*

Yu, A., Zhao, C., Fan, Y., Jang, W., Mungall, A. J., Deloukas, P., Olsen, A., Doggett, N. A., Ghebranious, N., Broman, K. W., and Weber, J. L. (2001). Comparison of human genetic and sequence-based physical maps. *Nature*, 409(6822):951–953.   *Cited at page 87*

Yu, J., Lazzeroni, L., Qin, J., Huang, M. M., Navidi, W., Erlich, H., and Arnheim, N. (1996). Individual variation in recombination among human males. *American Journal of Human Genetics*, 59(6):1186–1192. *Cited at page 66*

Yu, N., Jensen-Seaman, M. I., Chemnick, L., Ryder, O., and Li, W.-H. (2004). Nucleotide Diversity in Gorillas. *Genetics*, 166(3):1375–1383. *Cited at pages , 103*

Zanders, S. and Alani, E. (2009). The pch2Delta mutation in baker's yeast alters meiotic crossover levels and confers a defect in crossover interference. *PLoS genetics*, 5(7):e1000571. *Cited at page 58*

Zelazowski, M. J. and Cole, F. (2016). X marks the spot: PRDM9 rescues hybrid sterility by finding hidden treasure in the genome. *Nature structural & molecular biology*, 23(4):267–269. *Cited at page 72*

Zeng, J. and Yi, S. V. (2014). Specific Modifications of Histone Tails, but Not DNA Methylation, Mirror the Temporal Variation of Mammalian Recombination Hotspots. *Genome Biology and Evolution*, 6(10):2918–2929. *Cited at page 61*

Zeng, K. (2010). A simple multiallele model and its application to identifying preferred-unpreferred codons using polymorphism data. *Molecular Biology and Evolution*, 27(6):1327–1337. *Cited at page 90*

Zeng, K. and Charlesworth, B. (2009). Estimating Selection Intensity on Synonymous Codon Usage in a Nonequilibrium Population. *Genetics*, 183(2):651–662. *Cited at page 90*

Zeng, K. and Charlesworth, B. (2010). Studying Patterns of Recent Evolution at Synonymous Sites and Intronic Sites in Drosophila melanogaster. *Journal of Molecular Evolution*, 70(1):116–128. *Cited at page 90*

Zhang, L., Liang, Z., Hutchinson, J., and Kleckner, N. (2014). Crossover Patterning by the Beam-Film Model: Analysis and Implications. *PLoS Genetics*, 10(1). *Cited at page 57*

Zhu, J., Petersen, S., Tessarollo, L., and Nussenzweig, A. (2001). Targeted disruption of the Nijmegen breakage syndrome gene NBS1 leads to early embryonic lethality in mice. *Current biology: CB*, 11(2):105–109. *Cited at page 40*

Zickler, D. (2006). From early homologue recognition to synaptonemal complex formation. *Chromosoma*, 115(3):158–174. *Cited at pages 29, 30*

Zickler, D. and Kleckner, N. (1998). The leptotene-zygotene transition of meiosis. *Annual review of genetics*, 32(1):619–697. *Cited at page 29*

Zickler, D. and Kleckner, N. (1999). Meiotic chromosomes: integrating structure and function. *Annual review of genetics*, 33(1):603–754. *Cited at pages 32, 34*

Zickler, D. and Kleckner, N. (2015). Recombination, Pairing, and Synapsis of Homologs during Meiosis. *Cold Spring Harb Perspect Biol*, 7(6). *Cited at page 34*

Zirkle, C. (1935). The inheritance of acquired characters and the provisional hypothesis of pangenesis. *The American Naturalist*, 69(724):417–445. *Cited at page 7*

Zuckerman, S. (1951). The number of oocytes in the mature ovary. *Recent Progress in Hormone Research*, 6:63–109. *Cited at page 28*