



HAL
open science

Détection de processus sexuellement antagonistes dans le génome humain

Elise Lucotte

► **To cite this version:**

Elise Lucotte. Détection de processus sexuellement antagonistes dans le génome humain. Génétique. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066507 . tel-01308195

HAL Id: tel-01308195

<https://theses.hal.science/tel-01308195>

Submitted on 27 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

École Doctorale 227 – « Science de la Nature et de l'Homme : évolution et écologie »

UMR7206 – Éco-anthropologie et Ethnobiologie

Équipe Anthropologie Évolutive

Détection de processus sexuellement antagonistes dans le génome humain

par **ÉLISE LUCOTTE**

THÈSE DE DOCTORAT DE GÉNÉTIQUE DES POPULATIONS

Dirigée par Bruno Toupance et Évelyne Heyer

Soutenance prévue le 23 novembre 2015 devant un jury composé de :

Dominique Higué	Professeur – Université Paris VI	Président
Emmanuelle Génin	DR – INSERM	Rapporteur
Catherine Montchamp-Moreau	DR – CNRS	Rapporteur
Mark Kirkpatrick	Professeur – University of Texas, Austin	Examineur
Patricia Balaesque	CR – CNRS	Examineur
Évelyne Heyer	Professeur – MNHN	Directeur
Bruno Toupance	MCU – Université Paris VII	Directeur

Remerciements

Je tiens tout d'abord à remercier Bruno pour son encadrement et son soutien tout au long de ma thèse et pour m'avoir guidée dans mes premiers pas dans la recherche ainsi que dans l'enseignement. Je le remercie également de m'avoir fait confiance, même lors de mes trébuchements, et de m'avoir poussée à faire de mon mieux avec patience. Je m'excuse d'avance pour ses futurs étudiants, qui devront probablement passer un scanner complet avant de signer leur contrat.

Je remercie aussi Évelyne de m'avoir donné l'opportunité de travailler sur ce sujet passionnant, ainsi que de son enthousiasme sans faille et communicatif.

Je tiens à remercier Laure pour avoir rejoint l'aventure avec enthousiasme lors de l'écriture et ré-écriture(s) de mon premier article. Son expertise et son professionnalisme furent d'une aide précieuse et je tâcherai d'en prendre exemple pour la suite.

Je voudrais également remercier Fred pour m'avoir formée aux réflexions philosophiques, Paul pour m'avoir montré que tout plan doit avoir 17 étapes et Marie-France pour avoir apporté une touche féminine aux réunions du lundi soir. Je les remercie surtout pour leur soutien et leurs nombreux conseils, toujours exprimés avec convivialité.

Merci à Romain qui m'a soutenue depuis le début de ma thèse, a toujours été présent et surtout m'a supportée pendant ma rédaction. Merci aussi de m'avoir montré que la procrastination fait partie d'un travail bien fait, et que Google est mon ami.

Je remercie évidemment mes co-bureau qui m'ont accompagnée tout au long de ces trois années, plus particulièrement merci à Nina pour sa vision de l'ornithologie et ses anecdotes croustillantes, et à Julie pour avoir été notre vigie et avoir partagé mes déboires et mes fous rires. Merci à Goki pour ses interrogations philosophiques sur la vie et les éléments, à Valentin pour son goût prononcé pour les débats et le houblon, à Christophe pour avoir été un leader mais aussi aux autres étudiants qui participent ou ont participé à l'ambiance chaleureuse de cette équipe. Merci à Marie pour ses nombreux « pep talks », à Bérénice qui ne m'a jamais laissé muriner et à Claire pour son plateau à fromage.

Je remercie aussi les membres de l'unité d'Éco-anthropologie et Ethnobiologie et plus particulièrement toute l'équipe d'Anthropologie Génétique pour leur bonne humeur et leur convivialité, qui ont construit un environnement de travail motivant et stimulant.

J'espère sincèrement pouvoir refaire partie de l'équipe dans le futur. Un grand merci à Céline pour les séances de remise en forme et potinage, à Priscille pour son soutien ainsi qu'à Samuel et Raphaëlle pour leurs encouragements. Je remercie également Aline, Flora, Angeles et Aude pour leur convivialité et Franz pour être Franz. Merci à Taouès et Florence pour m'avoir guidée dans les méandres de l'administration, toujours avec bonne humeur. Merci également à Liliana, Véronique et Aurélie, qui m'ont accompagnée dans mes premiers pas au sein de l'unité.

Je remercie bien sûr mes amis pour leur humour face à mes recherches : Cathy pour sa promesse d'empañadas qui m'a fait tenir pendant la rédaction, Laura pour ses messages de soutien, mais aussi Xavier, Thomas, Guillaume, Jonathan, Victor et Erwan pour leur participation : des cailloux à la semence, il n'y a qu'un pas.

Merci à mes compagnons de galère depuis six ans, les magistériens P16. Plus particulièrement, merci à Élodie, mon acolyte qui m'a remotivée de nombreuses fois, n'hésitant pas à user de noisettes, de crêpes et de cidre. Je remercie aussi Amélie, qui partage mon incompréhension face aux techniques de laboratoire (oui, nous avons réussi l'extraction d'ADN de banane, mais combien de bananes ont été sacrifiées?), et Julia pour ses méthodes infaillibles pour me remonter le moral.

Enfin, je voudrais remercier ma famille pour leur soutien, plus particulièrement mes parents pour leur compréhension et pour m'avoir toujours donné les moyens de réussir, ma sœur pour ses illusions sur mon futur salaire, et mes tantes Micette et Mane pour leurs nombreux encouragements.

Liste des Abréviations

ESD *Environmental Sex Determination*, détermination environnementale du sexe

FDR *False Discovery Rate*, taux de faux positifs

GSD *Genetic Sex Determination*, détermination génétique du sexe

IRSC *InterLocus Sexual Conflict*, conflit sexuel interlocus

IASC *IntraLocus Sexual Conflict*, conflit sexuel intralocus

LD *Linkage Disequilibrium*, déséquilibre de liaison

MSY *Male Specific Y*, région du Y mâle-spécifique

OSR *Operational Sex Ratio*, sexe-ratio opérationnel

PAR *Pseudo-Autosomal Region*, région pseudo-autosomale

PRR *Potential Rate of Reproduction*, taux potentiel de reproduction

TD *Transmission Distortion*, distorsion de transmission

TDT *Transmission Disequilibrium Test*, test du déséquilibre de transmission

SA Sexuellement Antagoniste

SD *Sexually Differentiated*, différencié en fonction du sexe

SDR *Sex Determining Region*, région déterminant le sexe

SDT *Sexually Distorted Transmission*, distorsion de transmission en fonction du sexe

SNP *Single Nucleotide Polymorphism*, polymorphisme à un nucléotide

Définitions

Fertilité et fécondité La fertilité correspond à la capacité biologique de l'individu à se reproduire dans son environnement tandis que la fécondité correspond à la réalisation de cette capacité, c'est-à-dire au nombre de descendants que l'individu a effectivement eu. Un individu peut être capable d'avoir des descendants mais ne pas en avoir, donc être fertile mais non fécond. En anglais, la fertilité peut être traduite par « fecundity » et la fécondité par « fertility », ce qui peut mener à des confusions de traduction.

Polygynie Lorsque les mâles d'une population se reproduisent avec plusieurs femelles.

Monogamie Lorsque chaque individu d'une population ne se reproduit qu'avec un seul partenaire. Un individu peut être monogame pour une ou plusieurs saisons reproductrices ou pour toute sa vie.

Polyandrie Lorsque les femelles d'une population se reproduisent avec plusieurs mâles.

Fitness Correspond à la valeur sélective d'un individu du point de vue de sa survie et de son succès reproducteur.

Succès reproducteur Peut être mesuré par le nombre de descendants d'un individu, par le nombre de descendants survivants jusqu'à l'âge adulte ou par le nombre de descendants fertiles.

Pedigrees Des individus appartenant à une même famille, dont on connaît les liens de parenté, sur au moins une génération.

Table des matières

1	Introduction Générale	4
1.1	Les dimorphismes sexuels	5
1.1.1	L'anisogamie à l'origine de conflits entre les sexes	5
1.1.2	Les dimorphismes sexuels liés à la sélection naturelle	6
1.1.3	Les dimorphismes sexuels liés à la sélection sexuelle	8
1.1.4	Les conflits sexuels interlocus	14
1.1.5	L'architecture génétique des dimorphismes sexuels	15
1.2	Les conflits sexuels intralocus	19
1.2.1	Émergence des IASC	19
1.2.2	Résolution des conflits sexuels intralocus	20
1.2.3	Implications évolutives des IASC	25
1.3	Détection de sélection sexuellement antagoniste	28
2	Détection de sélection sexuellement antagoniste dans le génome humain	31
2.1	Introduction	31
2.1.1	La localisation génomique des locus SA prédite par le modèle de Rice	31
2.1.2	Critiques de la prédiction du modèle de Rice	35
2.1.3	Examen empirique de la localisation génomique des locus SA	41
2.1.4	La sélection SA à l'origine de différences de fréquences alléliques entre les sexes	42
2.2	Détection d'une signature de sélection sexuellement antagoniste dans les données de HapMap3.3	46
2.2.1	Matériels et méthodes	46
2.2.2	Résultats	48
2.3	Discussion	49
2.4	Article	55
2.5	Annexes de l'article	76

3	Distorsion de transmission sexe-spécifique	117
3.1	Introduction	117
3.1.1	La distorsion de transmission et les IASC	121
3.1.2	Distorsion de transmission sexe-spécifique	122
3.2	Méthodes	124
3.2.1	La base de données Genome of the Netherlands	124
3.2.2	Détection de signatures de sélection sexuellement antagoniste . . .	127
3.2.3	Détection de distorsions de transmission	128
3.2.4	Simulation d'un chromosome avec un locus sous sélection SA . . .	131
3.2.5	Évaluation du F_{ST} intersexuel chez les enfants : est-il extrême comparé aux génotypes de leurs parents?	134
3.3	Résultats	135
3.3.1	Détection de signatures de sélection sexuellement antagoniste dans Genome of the Netherlands	135
3.3.2	Détection de distorsions de transmission sur des chromosomes simulés	138
3.3.3	Détection de distorsions de transmission dans les données de GoNL	149
3.4	Discussion	170
4	Conclusion Générale	180
	Références bibliographiques	191
	Annexes du chapitre 3	203

Chapitre 1

Introduction Générale

Les dimorphismes sexuels regroupent les différences morphologiques entre mâles et femelles d'une même espèce, et sont très répandus dans la nature. On peut distinguer deux types de dimorphismes sexuels : les dimorphismes sexuels primaires, c'est-à-dire les organes sexuels, et les dimorphismes sexuels secondaires, qui ne sont pas directement impliqués dans la reproduction. Darwin proposa dans « *The Descent of Man* » que l'émergence de ces dimorphismes entre les sexes serait due soit à l'action de la sélection naturelle, qui agit sur la capacité intrinsèque d'un individu à survivre et se reproduire, soit à l'action de la sélection sexuelle, qui « dépend de l'avantage que possèdent certains individus sur d'autres de même sexe et de même espèce, uniquement en ce qui concerne la reproduction »¹ (Darwin 1871). Les traits impliqués dans la sélection sexuelle n'augmentent donc pas nécessairement la survie des individus qui les portent, mais favorisent leur accès à la reproduction par rapport aux autres individus du même sexe. La sélection sexuelle s'opère de deux façons : par une compétition entre individus du même sexe pour l'accès aux partenaires, la sélection intra-sexuelle, ou par un choix du partenaire par le sexe opposé, la sélection inter-sexuelle (Darwin 1871). La sélection intra-sexuelle prend souvent la forme de combats entre individus du même sexe, à l'issue desquels les vainqueurs s'accouplent avec les individus réceptifs du sexe opposé, et mène à l'évolution d'armements. La sélection inter-sexuelle mène quant à elle à l'évolution d'ornements. Darwin avait remarqué que les mâles arborent plus souvent des traits exagérés, ornements ou armements, que les femelles. L'évolution de bois plus développés chez les cerfs mâles par le biais de la sélection intra-sexuelle, et l'évolution de plumages colorés ainsi que de comportements de parade chez les oiseaux de paradis mâles, par le biais de la sélection

1. « depends on the advantage which certain individuals have over others of the same sex and species solely in respect of reproduction. » Ici, le terme « reproduction » désigne la fécondité tandis que pour la sélection naturelle le terme reproduction désigne la fertilité.

tion inter-sexuelle sont des exemples emblématiques de l'action de la sélection sexuelle (Darwin 1871).

Au cours de cette introduction, nous allons tout d'abord détailler les mécanismes menant à l'émergence de dimorphismes sexuels en nous concentrant sur les animaux, bien que les dimorphismes sexuels soient aussi répandus chez les plantes (voir Barrett et Hough 2013 pour une revue). Ensuite, nous allons nous intéresser à la base génétique des dimorphismes sexuels dont l'évolution met en jeu un conflit entre les sexes.

1.1 Les dimorphismes sexuels

1.1.1 L'anisogamie à l'origine de conflits entre les sexes

Chez les espèces sexuées isogames, la fécondation se déroule entre gamètes de même taille, mais des types sexuels peuvent tout de même exister. En général, les deux types d'individus sont souvent désignés par « + » et « - ». Les espèces isogames sont relativement rares dans la nature, on les trouve notamment chez certaines algues, champignons, organismes unicellulaires ou plantes.

Chez les espèces sexuées anisogames, certains individus produisent des gamètes de grande taille, les ovules, contenant une importante quantité de ressources, tandis que les autres produisent des gamètes plus petits, les spermatozoïdes, contenant peu de ressources. Les individus produisant les gamètes les plus gros sont appelés femelles et les individus produisant les gamètes les plus petits sont appelés mâles.

La théorie historique sur l'évolution de l'anisogamie à partir de l'isogamie (Parker *et al.* 1972), considérée comme l'état ancestral de la reproduction sexuée, invoque un conflit entre les sexes résultant d'une stratégie égoïste des producteurs de petits gamètes (voir Randerson et Hurst 2001 pour une revue des autres hypothèses). En effet, les producteurs des plus gros gamètes sont avantagés si la fitness du zygote augmente avec la taille de l'oeuf. Néanmoins, si un mutant produisant des gamètes plus petits et en plus grand nombre émerge, celui-ci pourra féconder de nombreux gamètes tout en investissant peu de ressources dans le futur zygote. Une spécialisation d'une partie des individus dans la production de petits gamètes générés en grand nombre et investissant peu de ressources dans le zygote, et de l'autre partie des individus dans la production de gros gamètes, permettant l'apport nécessaire en ressources au zygote et générés en plus petit nombre, aurait ainsi été avantageuse. Les gamètes de taille intermédiaires auraient été contre-sélectionnés car ils ne possédaient ni les avantages des gros gamètes ni ceux des petits

gamètes.

Chez les animaux, on observe en grande majorité de l'oogamie, la forme la plus marquée de l'anisogamie, où les gamètes femelles ont une taille beaucoup plus importante que les gamètes mâles et sont immobiles tandis que les gamètes mâles sont très mobiles. Dans ce cas, l'ovule contient la quasi-totalité des ressources nécessaires au développement de l'embryon tandis que les spermatozoïdes sont mobiles et se déplacent vers les ovules pour transmettre l'information génétique du mâle sans donner de ressources énergétiques au futur embryon.

Produire les gamètes de grande taille est coûteux et leur nombre est limité par la quantité d'énergie disponible chez les femelles tandis que les gamètes petits peuvent être produits en grand nombre (Bateman 1948 ; Trivers 1972). Ce dimorphisme gamétique a été invoqué pour expliquer l'asymétrie fréquente des coûts et des gains associés à la reproduction entre les sexes, et de la qualité de « sexe » limitant de la femelle, impliquant que l'accès aux femelles est une ressource limitante pour les mâles. Cette asymétrie serait à l'origine de l'émergence de conflits d'intérêt entre les sexes au niveau de l'acquisition de ressources et de la reproduction, menant à l'évolution de dimorphismes sexuels.

1.1.2 Les dimorphismes sexuels liés à la sélection naturelle

Les dimorphismes sexuels peuvent évoluer en réponse à la sélection naturelle. En effet, si les sexes ont des traits d'histoire de vie différents, l'apparition de dimorphismes sexuels en réponse aux pressions de sélection sexe-spécifiques peut se révéler avantageuse. Ces traits ont souvent un rapport avec la reproduction, mais de façon indirecte.

Un des dimorphismes sexuels les plus étudié est le dimorphisme de taille. Chez la majorité des espèces, notamment chez les insectes, les reptiles et les poissons, les femelles ont une taille plus importante que les mâles tandis que chez de nombreuses espèces de mammifères et d'oiseaux, on observe une taille plus importante chez le mâle (Shine 1989). Une plus grande taille chez les mâles serait favorisée par sélection sexuelle lors de la compétition intra-sexuelle pour l'accès aux femelles sous forme de combat (Darwin 1871). En revanche, une plus grande taille corporelle serait favorisée chez les femelles par sélection naturelle car cela augmenterait leur fertilité, les femelles les plus grandes pouvant produire plus d'œufs (Darwin 1871).

Shine (1989) dresse dans sa revue un inventaire détaillé des différentes formes de dimorphisme sexuel lié à l'appareil trophique, c'est-à-dire aux organes et membres qui permettent la préhension, la consommation et la digestion de la nourriture. Deux catégories sont recensées : les dimorphismes dus à l'occupation de niches écologiques différentes entre les deux sexes et ceux dus à d'autres facteurs.

La première catégorie regroupe les dimorphismes liés à une différence de fréquence de prise de nourriture et ceux liés à une différence de régime alimentaire.

Chez les espèces avec un dimorphisme de taille marqué, la fréquence de prise de nourriture est souvent différente entre les mâles et les femelles. Les femelles subissent des coûts plus élevés associés à la reproduction que les mâles qui vont avoir tendance à moins se nourrir que les femelles en l'absence de compétition intra-sexuelle (Shine 1989). Les cas les plus extrêmes sont représentés chez les espèces où les mâles ne se nourrissent pas du tout et possèdent des systèmes digestifs et des pièces buccales réduites voire absentes. Par exemple chez le poisson *Cryptosaras couesii*, le mâle s'attache à la femelle qui lui fournit des apports nutritifs, devenant ainsi un appendice parasitaire produisant des spermatozoïdes.

Une autre catégorie de dimorphisme sexuel lié à des niches écologiques différentes regroupe les cas où les deux sexes se nourrissent de ressources différentes, menant à l'évolution de pièces buccales sexuellement dimorphiques. Ce type de dimorphisme peut évoluer pour augmenter les capacités collectives à trouver des ressources ou peut être dû à la nécessité qu'ont les femelles à avoir une alimentation plus riche. Par exemple, chez les moustiques (famille des *Culicidae*), les femelles se nourrissent de nectar de fleur, comme les mâles, mais sont les seules à être hématophages car l'hématophagie est essentielle à la ponte (Darwin 1871). Des différences de dentition chez les mammifères herbivores pourraient aussi être expliquées par des différences en besoins nutritionnels (Shine 1989).

Parmi les facteurs non attribuables à une différence de niche écologique entre les sexes, on retrouve les activités spécifiques aux rôles des sexes durant la reproduction. Celles-ci peuvent entraîner le développement de certains caractères chez un seul sexe. Par exemple, chez les guêpes *Vespa vulgaris* les femelles creusent un nid et possèdent un appareil trophique différent des mâles, facilitant cette tâche. Les appareils trophiques peuvent également être modifiés dans le cas de soins parentaux particuliers, comme chez les mâles poisson-chat (*Loricariidae*) qui couvent les œufs oralement grâce à une morphologie buccale particulière qui remplit cette fonction (Shine 1989).

Enfin, Rice (1984) soulève que de nombreux autres traits, plus difficilement observables, pourraient être sujets à des pressions de sélection différentes entre les sexes. Par exemple des traits physiologiques comme la croissance, la thermorégulation, les taux métaboliques, les rythmes biologiques ou les modalités sensorielles pourraient avoir des optimums différents entre les sexes (Glucksmann 1981).

La théorie prédit que de nombreux traits dimorphiques pourraient résulter de diffé-

rences de niches écologiques (Slatkin 1984), et de nombreuses espèces montrent d'importantes divergences écologiques entre les sexes. Pourtant, cette hypothèse a souvent été éclipsée au profit de la sélection sexuelle pour expliquer les dimorphismes sexuels. En effet, cette hypothèse a un pouvoir prédictif plus élevé que l'hypothèse de la niche écologique, et a été beaucoup plus intensément étudiée.

1.1.3 Les dimorphismes sexuels liés à la sélection sexuelle

Un mécanisme majeur à l'origine de l'émergence de dimorphismes sexuels est la sélection sexuelle. Celle-ci agirait majoritairement chez les mâles (mais voir Clutton-Brock 2009 pour un revue sur la sélection sexuelle chez les femelles et voir plus bas) sous la forme de compétition intra-sexuelle pour l'accès à la femelle, ou de sélection inter-sexuelle sous la forme de choix du partenaire effectué par les femelles. L'hypothèse historique pour expliquer l'action de la sélection sexuelle est qu'elle serait une conséquence de l'anisogamie et de l'asymétrie de l'investissement dans les soins parentaux entre les sexes.

Asymétrie du succès reproducteur et des soins parentaux

Une conséquence de l'anisogamie serait que le nombre de descendants d'un mâle augmenterait avec le nombre de femelles avec lequel il s'accouple, tandis que la fécondité d'une femelle n'augmenterait pas si elle s'accouple avec plus qu'un mâle (figure 1.1), hypothèse formulée par Bateman (1948). Chez les mâles, il serait donc avantageux de s'accoupler plusieurs fois, menant ainsi à une compétition intense entre mâles pour l'accès aux femelles. Dans cette configuration, les femelles ne tirent aucun avantage à s'accoupler de multiple fois, et ont tendance à limiter les risques liés à l'accouplement². En conséquence, la force de la sélection sexuelle chez le sexe considéré dépendrait de la relation entre le nombre de partenaires et le nombre de descendants, défini comme le gradient de Bateman par Andersson et Iwasa (1996). Cette relation a été initialement décrite chez la drosophile (Bateman 1948).

En raison de l'anisogamie, le système de reproduction favorisé chez les mâles serait donc la polygynie, c'est-à-dire quand les mâles s'accouplent avec plusieurs femelles.

Chez les espèces polygynes, la variance du succès reproducteur est souvent différente entre les mâles et les femelles. Chaque enfant ayant un père et une mère, le succès reproducteur moyen des mâles et des femelles est égal (Fisher 1958). Néanmoins, une importante proportion de mâles ne se reproduit pas tandis que les autres se reproduisent de nombreuses fois. La variance du succès reproducteur serait donc beaucoup plus élevée

2. Les risques liés à la reproduction comprennent notamment la transmission de parasites et, dans le cas d'IRSC (voir section 1.1.4), à des coûts associés aux interactions avec les mâles lors de l'accouplement.

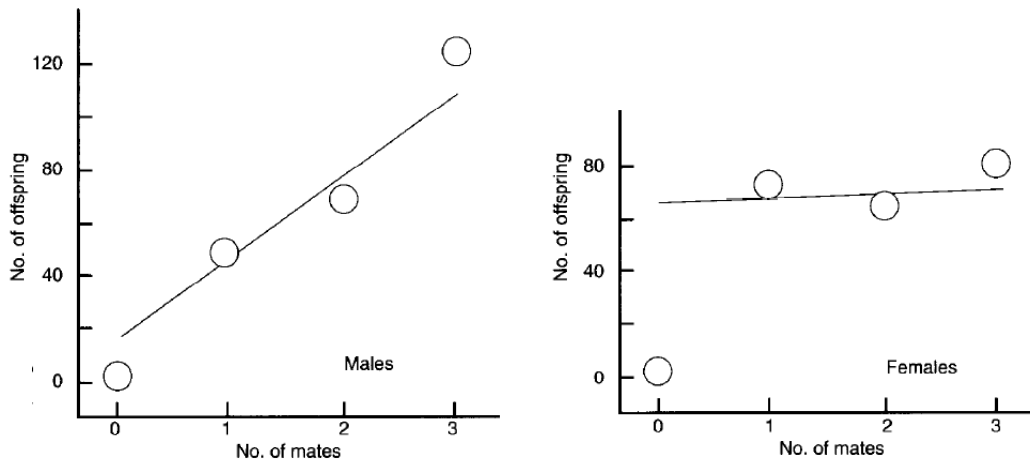


Figure 1.1 – Le gradient de Bateman. Dans cette expérience menée chez la drosophile, le nombre de descendants augmente avec le nombre de partenaires chez les mâles, mais pas chez les femelles. Ce gradient est donc élevé uniquement chez les mâles. Adapté de la revue de Andersson et Iwasa (1996)

chez les mâles que chez les femelles (Bateman 1948).

Dans ce cas, l'évolution de soins parentaux effectués par les mâles est contre sélectionnée car les mâles vont avoir un intérêt à chercher d'autres partenaires directement après l'accouplement, afin d'augmenter leur probabilité de paternité. De plus, la femelle, investissant plus de ressources dans sa descendance dès la production des gamètes, aurait un coût plus élevé associé à la mort de ses descendants en comparaison aux mâles. Si le mâle abandonne la femelle afin de chercher d'autres partenaires, il est très désavantageux pour la femelle d'abandonner à son tour ses petits à une mort certaine, elle va donc avoir tendance à prodiguer les soins parentaux (Dawkins 1976). Cela serait à l'origine d'une asymétrie entre les sexes au niveau des soins parentaux, plus souvent prodigués par les femelles. Enfin, s'il est impossible pour les mâles d'identifier leurs descendants, donc s'ils ne peuvent être certains de leur paternité, ce phénomène est amplifié (Queller 1997).

Le taux potentiel de reproduction et le sexe-ratio opérationnel

Le taux potentiel de reproduction (Potential Rate of Reproduction, PRR) correspond au nombre maximum de descendants qu'un individu peut produire par unité de temps (Clutton-Brock et Vincent 1991). Ce taux est influencé par le gradient de Bateman chez le sexe considéré, ainsi que par l'asymétrie des soins parentaux. Le PRR influence à son tour le sexe-ratio opérationnel (Operational Sex Ratio, OSR), qui correspond au rapport entre le nombre de mâles actifs sexuellement et le nombre de femelles réceptives à un instant donné (Emlen et Oring 1977). Si l'on introduit ces principes dans le cadre des espèces polygynes, un investissement en ressources plus faible des mâles dans les gamètes et

dans les soins parentaux conduit à un PRR plus élevé chez les mâles que chez les femelles, c'est-à-dire que les mâles ont potentiellement plus de descendants que les femelles. Les femelles ayant tendance à effectuer les soins parentaux, elles sont disponibles plus tard pour se reproduire à nouveau tandis que les mâles peuvent être sexuellement actifs directement après l'accouplement. Le nombre de mâles disponibles est donc supérieur au nombre de femelles disponibles et une compétition entre les mâles se produit pour l'accès aux femelles. Ces deux facteurs, l'OSR et le PRR, influencent la direction et l'intensité de la sélection sexuelle.

Mécanismes de sélection sexuelle

Comme expliqué ci-dessus, la sélection sexuelle va plus souvent s'opérer chez les mâles, nous considérerons donc ce cas pour la suite du raisonnement. Chez les espèces polygynes, les mâles ont un intérêt à s'accoupler avec plusieurs femelles. À l'inverse, la fécondité des femelles n'augmentant pas avec le nombre d'accouplements, les femelles ont donc un intérêt à s'accoupler avec les mâles ayant la fitness la plus haute. Les ornements et armements émergeant de la sélection sexuelle sont désavantageux pour la survie des mâles car ils peuvent non seulement être coûteux à produire, mais ils peuvent aussi diminuer l'agilité des mâles et augmenter leur visibilité face aux prédateurs. La sélection de ces traits peut néanmoins être favorisée s'ils procurent aux mâles un avantage pour l'accès aux femelles. Il est facile de comprendre pourquoi des armements exagérés évoluent car ils sont avantageux lors d'une compétition directe entre mâles et peuvent aussi servir dans la défense contre les prédateurs. Dans le cadre du choix du partenaire, il est plus difficile d'expliquer pourquoi des préférences pour des traits coûteux pour les mâles ont évolué chez les femelles. Cela s'explique s'il y a un effet bénéfique de ce choix chez les femelles, qu'il soit direct ou indirect. Nous allons ici présenter trois grandes classes de mécanismes : le modèle de bénéfices directs et deux modèles de bénéfices indirects, l'emballement de Fisher et le principe du handicap. La revue de Kirkpatrick et Ryan (1991) dresse une liste de références d'études proposant d'autres mécanismes.

Les modèles de bénéfices directs

Ces modèles supposent que les femelles reçoivent un bénéfice direct à choisir un mâle, qui lui procure par exemple des cadeaux nuptiaux, la défense du territoire ou des soins parentaux à ses descendants (voir Jones et Ratterman 2009 pour une revue). Chez le criquet *Requena verticalis*, il a été montré que le cadeau nuptial offert à la femelle par le mâle, un spermatophore contenant des nutriments que la femelle va consommer, augmente la fitness de la femelle en augmentant le nombre et la taille des œufs produits (Gwynne 1984), apportant ainsi à la femelle un bénéfice précopulatoire. Les femelles ont donc

tendance à choisir le mâle qui leur offre le cadeau nuptial le plus important. Dans le cas de bénéfices postcopulatoires, une corrélation entre le trait mâle choisi par la femelle et le bénéfice apporté par le mâle est nécessaire afin qu'il y ait un choix de la part de la femelle. Par exemple, chez l'oiseau *Turdus merula*, les femelles choisissent de s'accoupler avec un mâle notamment en fonction de l'intensité de la couleur orange de leurs becs, positivement corrélé avec les soins parentaux qu'il effectuera (Préault *et al.* 2005).

Les modèles de bénéfices indirects – l'emballement de Fisher

L'une des hypothèses historiques pour expliquer l'évolution des préférences des femelles pour des traits coûteux chez le mâle est « l'emballement de Fisher » (ou la « sexy-son hypothesis » (Fisher 1958)) qui propose que ces préférences évoluent si elles sont associées génétiquement avec les gènes produisant le trait sous sélection sexuelle chez le mâle, de façon à ce que les gènes codant pour la préférence et le trait exagéré soient transmis ensemble à la descendance. Les femelles ayant une préférence pour ce trait se reproduiront avec des mâles le possédant, le trait sera ainsi transmis à leur descendance. Si la préférence est répandue dans la population, les fils auront eux même un succès reproducteur plus élevé que les mâles sans ce trait. En parallèle, les allèles codant la préférence sont eux aussi transmis par les fils et par les filles, permettant à la préférence d'envahir peu à peu la population. Ce processus peut mener à l'exagération du trait sans que ce trait n'augmente la survie ni des femelles, ni des mâles, ni de leurs descendants, voire la diminue. Cela s'apparente à un mécanisme de gène égoïste (Dawkins 1976). Néanmoins, les descendants mâles possédant ce trait vont avoir un avantage du point de vue de l'accès à la reproduction.

Les modèles de bénéfices indirects – la théorie du handicap

Une autre hypothèse fut proposée par Zahavi (1975), celle dite « du handicap ». Elle propose qu'un trait ornemental doit être nécessairement coûteux pour envoyer un « signal honnête » aux femelles sur la fitness du mâle. En effet, un individu capable de produire ces traits coûteux et de survivre jusqu'à la saison de reproduction doit avoir une fitness élevée. Les femelles développent ainsi une préférence pour ces traits car leurs descendants auront également une fitness élevée, puisqu'ils hériteront des « bons gènes » paternels. L'hypothèse du parasite développée par Hamilton et Zuk (1982) découle directement de l'idée de signal honnête. Celle-ci propose qu'une coévolution entre un hôte et ses parasites pourrait mener à l'évolution de préférences pour des traits mâles corrélés avec la résistance aux parasites, menant à l'exagération de ces traits.

Ces deux derniers processus rentrent dans le cadre des mécanismes de bénéfices indirects, car ce n'est pas la femelle qui reçoit un avantage en fitness en choisissant un mâle

avec ces traits, mais ses descendants.

Lien entre la sélection sexuelle et les systèmes de reproduction

En raison de l'anisogamie, on s'attend à ce que la polygynie soit favorisée chez les mâles. Sa présence ou son absence sont donc conditionnées par l'avantage ou le désavantage que ce système procure aux femelles. Plusieurs facteurs, souvent environnementaux et spécifiques aux traits d'histoire de vie d'une espèce, peuvent agir sur le gradient de Bateman et engendrer des situations alternatives à celle illustrée en figure 1.1 :

- si les gradients de Bateman sont inversés entre les sexes, le système de reproduction favorisé chez les femelles est la polyandrie et la sélection sexuelle agit préférentiellement chez les femelles. Le chevalier grivelé (*Actitis macularius*) est un exemple d'espèce polyandre (Oring et Lank 1982). On s'attend à ce que la polyandrie puisse être avantageuse sous certaines conditions environnementales (voir Clutton-Brock 2009), mais elle est rarement observée (Orrians 1969).
- si les gradients de Bateman sont élevés pour les deux sexes, le système de reproduction favorisé est multi-mâles/multi-femelles et on s'attend à ce que la sélection sexuelle soit forte chez les deux sexes. C'est ce qui est notamment observé chez les chimpanzés *Pan troglodytes* et les bonobos *Pan paniscus* (Dixson 2015).
- si les gradients de Bateman sont faibles pour les deux sexes, le système de reproduction favorisé est la monogamie. Dans ce cas, on s'attend à ce que la sélection sexuelle soit faible et à ce que les dimorphismes sexuels résultant de la sélection sexuelle soient peu marqués. Par exemple, la corneille des clochers (*Coloeus monedula*) est une espèce strictement monogame qui ne présente aucun dimorphisme sexuel visible (Henderson *et al.* 2000). Néanmoins, si la variance de la fécondité est plus élevée chez un sexe, cela peut mener à l'évolution de sélection sexuelle sous la forme de choix du partenaire.

Un des facteurs qui conditionne l'évolution de différents systèmes de reproduction est la nécessité de soins parentaux pour la survie des descendants. En effet, la monogamie est avantageuse dans le cas où un soin parental de la part du mâle augmente fortement la survie de ses descendants.

Chez les mammifères, la présence de mamelles chez les femelles les spécialise pour assurer l'apport nutritionnel de leurs descendants, et suggère un rôle mineur des mâles dans les soins parentaux. De fait, les mammifères sont majoritairement polygynes (Orrians 1969). Les exemples d'espèces monogames, représentant 3% des espèces de mammifères, sont en majorité des carnivores terrestres, comme par exemple les renards et les blaireaux, chez qui la collecte de nourriture est moins aisée que chez les herbivores. Les mâles auront

alors un rôle important : tuer les proies pour nourrir soit la femelle afin qu'elle produise du lait, soit les juvéniles suffisamment âgés pour être capables de se nourrir de viande.

Chez les oiseaux, environ 90% des espèces sont monogames (Lack 1968). Contrairement aux mammifères, seule la production d'œufs est une capacité strictement associée aux femelles. La monogamie sociale permettrait aux femelles de protéger leur nid tandis que les mâles cherchent de la nourriture. Un système polygyne aurait une influence néfaste sur ce processus, car le mâle devrait diviser son temps et les ressources rapportées entre ses nids (Orians 1969). Un paradoxe apparaît dans le fait que les oiseaux, en majorité monogames, montrent souvent des ornements très développés chez les mâles tandis que la sélection sexuelle est moins intense chez les espèces monogames. On s'attendrait donc à observer peu de dimorphismes sexuels. Néanmoins, une monogamie sociale n'implique pas une monogamie génétique, où chaque individu se reproduit uniquement avec l'autre membre de la paire. À ce jour, une monogamie génétique stricte a été avérée chez moins de 25% des oiseaux socialement monogames (Griffith *et al.* 2002). Le fait que les oiseaux ne soient que rarement monogames génétiquement implique que l'action de la sélection sexuelle est plus intense que chez des espèces strictement monogames.

Limites du modèle

Les théories développées ci-dessus expliquent la grande majorité des variations dans le développement de traits exagérés chez les mâles (Clutton-Brock 2007). Il est néanmoins important de noter que la sélection sexuelle agit aussi chez les femelles sous ses deux formes, la sélection intra- et inter-sexuelle.

Comme pour les mâles, l'intensité de la sélection sexuelle chez les femelles est influencée par le système de reproduction. En effet, chez les espèces polyandres, les OSR sont biaisés vers les femelles et la sélection sexuelle s'opèrent de façon plus intense chez celles-ci, résultant en des caractères sexuels secondaires plus développés chez les femelles (Clutton-Brock 2007).

Néanmoins, l'intensité de la sélection sexuelle chez les femelles est plus influencée par la répartition des ressources nécessaires à la reproduction que par le système de reproduction, contrairement aux mâles (Clutton-Brock 2009). En effet, la reproduction étant plus coûteuse pour les femelles, une compétition intra-sexuelle pour les ressources nécessaires à la reproduction et à la survie de leurs descendants émerge si celles-ci sont limitées. De fait, chez certaines espèces, on observe qu'un petit nombre de femelles dominantes sont les seules à se reproduire, créant de fortes variations du succès reproducteur entre individus femelles. Cette compétition pour l'accès au statut social de dominance va favoriser l'évolution de traits augmentant le succès compétiteur des femelles face aux autres individus du même sexe.

Par exemple, chez le suricate *Suricata suricatta*, bien que ce soient les femelles qui prodiguent les soins parentaux, moins de femelles se reproduisent que de mâles en réponse à une compétition intense entre les femelles pour l'accès au statut de dominance. Elles ont en effet besoin de l'aide de subordonnées pour élever leur petits. Chez cette espèce, on observe que les femelles sont souvent plus agressives que les mâles, que les individus ayant le poids le plus élevé dans la population sont les femelles dominantes et que celles-ci possèdent également un niveau plus élevé de testostérone dans le sang en comparaison aux femelles subordonnées (Clutton-Brock 2009).

Une compétition intra-sexuelle pour l'accès aux ressources mène généralement à de fortes variations individuelles en fertilité chez les femelles. Les mâles ayant un avantage à s'accoupler avec les femelles les plus fertiles, celles-ci vont développer des signaux honnêtes de fertilité (Clutton-Brock 2007). Par exemple, chez des primates sociaux et certains oiseaux, on observe un gonflement du périnée durant la période d'ovulation. Chez l'Homme, la poitrine et les fesses plus développées chez les femmes pourraient avoir évoluées comme indicateurs de fertilité (Clutton-Brock 2007).

De plus, si l'OSR est biaisé vers les femelles, c'est-à-dire si plus de femelles sont disponibles pour la reproduction que de mâles, cela peut mener à une inversion des rôles décrits jusqu'à présent en dehors des espèces polyandres. Par exemple, chez l'hippocampe monogame *Hippocampus subelongatus*, les mâles couvent les oeufs et la compétition pour l'accès aux mâles est forte chez les femelles (Clutton-Brock 2009).

1.1.4 Les conflits sexuels interlocus

Les conflits sexuels interlocus (Interlocus Sexual Conflict, IRSC) peuvent également être à l'origine de l'évolution de dimorphismes sexuels sur des traits impliqués dans les interactions mâles-femelles lors de la reproduction. Un IRSC émerge lorsqu'un nouveau trait avantageux pour la reproduction chez les mâles est désavantageux pour la survie ou la reproduction des femelles. Souvent, ces traits mâles sont impliqués dans la sélection sexuelle, notamment dans la compétition intra-sexuelle pour l'accès aux femelles. Les femelles auraient un avantage à se reproduire avec les mâles possédant des traits exagérés afin d'avoir des fils compétitifs, néanmoins le coût associé à ce trait mâle va favoriser l'évolution d'une résistance à ce trait chez les femelles. Les IRSC sont ainsi à l'origine de la coévolution de stratégies visant chez les mâles à diminuer le succès des compétiteurs ou à pratiquer la coercition sexuelle et chez les femelles à résister à l'accouplement et à éviter les mâles.

Le premier exemple formel d'IRSC a été décrit chez *Scathophaga stercoraria* (Parker 1979). Lors de la compétition intra-sexuelle entre les mâles pour l'accès aux femelles, les

mâles engagent un combat durant lequel la femelle a de forte chance de subir des blessures colatérales. Ce phénomène a pour effet de sélectionner chez la femelle des adaptations sexuellement antagonistes pour contrecarrer les effets délétères de la compétition entre mâles (Chapman *et al.* 2003).

Le fluide séminal produit par les glandes accessoires des mâles chez *Drosophila melanogaster* est un autre exemple d'IRSC. Ce fluide séminal est à l'origine d'une augmentation du taux de ponte des femelles, d'une réduction de la réceptivité des femelles à de futurs accouplements et d'une destruction du sperme des mâles s'étant récemment accouplés avec la femelle, augmentant ainsi la compétitivité du mâle. Néanmoins, ce fluide serait composé d'une substance toxique pour les femelles dans la mesure où il a été montré qu'une exposition à ce fluide est corrélée à une diminution de la survie chez les femelles. Ce trait a donc un effet positif sur la reproduction des mâles, mais a en parallèle un effet néfaste sur la survie des femelles (Chapman *et al.* 1995), donnant lieu à un IRSC.

De nombreux autres exemples peuvent être interprétés comme des IRSC. Chez l'éléphant de mer *Mirounga leonina*, les mâles ont un comportement de harcèlement envers les femelles pour la copulation, ce qui diminue la survie et le succès reproducteur de celles-ci. Les femelles synchronisent leur activité de copulation et se reproduisent rarement de façon isolée, ce qui diminue le risque de harcèlement (Galimberti *et al.* 2000). Ces comportements pourraient avoir évolué en réponse à un IRSC.

Les IRSC prennent ainsi souvent la forme d'une « course à l'armement » entre les sexes et peuvent mener à l'évolution de phénotypes extrêmes, comme des parties génitales à épines chez les mâles et des épines abdominales défensives chez les femelles (voir Arnqvist et Rowe 2005 pour une revue).

1.1.5 L'architecture génétique des dimorphismes sexuels

Nous avons jusqu'à présent détaillé les mécanismes favorisant l'émergence de dimorphismes sexuels, nous allons maintenant nous intéresser plus spécifiquement aux mécanismes génétiques sous-jacents à leur évolution. En effet, un paradoxe apparaît dans la mesure où l'information génétique est partagée entre les sexes et que chaque descendant reçoit la moitié de son matériel génétique de son père et l'autre moitié de sa mère. Comment expliquer l'émergence puis le maintien de ces dimorphismes sexuels ?

Si le trait dimorphique est codé génétiquement, il est nécessaire que les gènes codant pour ce trait aient une expression sexe-spécifique afin qu'il soit présent uniquement chez un sexe.

Selon Rice (1984), un trait sexuellement dimorphique peut évoluer de deux façons différentes à partir d'un trait monomorphe :

- via l’augmentation en fréquence des allèles du gène codant pour la nouvelle version du trait, qui sera exprimée uniquement chez un sexe. Ce mécanisme requiert l’apparition simultanée d’une mutation produisant la nouvelle version du trait et limitant son expression chez un sexe.
- via l’augmentation en fréquence des gènes codant pour le trait, qui sera dans un premier temps exprimé chez les deux sexes, suivie par l’évolution de gènes qui restreindront l’expression du trait à un sexe.

Dans la mesure où il est rare d’observer des mutations ayant un effet complètement sexe-spécifique (Rice 1984), il est parcimonieux de considérer que la majorité des variations disponibles pour l’évolution des dimorphismes sexuels sont initialement exprimées chez les deux sexes. Le deuxième mécanisme évoqué ci-dessus serait donc plus plausible dans le cadre de l’émergence de traits sexuellement dimorphiques.

Il est raisonnable d’en déduire que durant la phase d’émergence des dimorphismes sexuels, les deux versions du trait considéré sont exprimées chez les deux sexes.

Si chaque version de ce trait est avantageuse chez un sexe mais désavantageuse chez l’autre sexe, on s’attend à ce que ces versions soient sélectionnées de façon sexe-spécifique. Ce trait sera alors sous sélection sexuellement antagoniste (SA) : chaque version du trait sera sélectionnée positivement chez un sexe mais négativement chez l’autre sexe. Si ce trait est codé par les mêmes gènes chez les deux sexes, un génome unique ne peut coder pour les deux optimums sexe-spécifiques. Ce trait est sous conflit sexuel intralocus (*Intralocus Sexual Conflict*, IASC).

Ce type de conflit est très désavantageux pour la valeur sélective moyenne au niveau de l’espèce, car les femelles se reproduisant avec des mâles ayant une fitness élevée auront des filles avec une fitness plus basse que si ces femelles s’étaient accouplées avec des mâles ayant une fitness basse (et *vice-versa* pour les fils). La fitness chez les descendants est donc en moyenne diminuée.

Les IRSC et IASC représentent les deux types de conflits sexuels émergeant de différences d’optimums en fitness entre les sexes, et peuvent interagir de plusieurs façons (voir encadré 1.1 et figure 1.2 pour plus de détails).

Au contraire des IRSC, l’étude des IASC a été longtemps négligée car ils sont plus difficilement détectables. Néanmoins, depuis la publication des bases théoriques de ces conflits (Rice 1984) et les premières études empiriques de l’action des IASC (par exemple Chippindale *et al.* 2001), ce sujet suscite de plus en plus d’intérêt dans la communauté scientifique (voir Bonduriansky et Chenoweth 2009).

Au cours de cette thèse, nous nous sommes intéressée à ces conflits sexuels intralocus, à leur signature génétique et à leur implication dans l’évolution des dimorphismes sexuels chez l’humain.

Jusqu'ici, nous avons décrit les principales causes de l'évolution des dimorphismes sexuels, résultant d'optimums en fitness différents entre les sexes au niveau de la survie et de la reproduction. Ces divergences d'intérêt sont à la source de conflits entre les sexes. Deux types de conflit entre les sexes existent : les conflits sexuels interlocus (IRSC) et les conflits sexuels intralocus (IASC). Ces deux types de conflits ne sont pas mutuellement exclusifs, mais on peut les différencier sur un critère simple : les traits impliqués dans les conflits (voir figure 1.2). Les IRSC agissent si un trait s'exprimant chez un seul sexe a un effet négatif sur la fitness de l'autre sexe lors de leurs interactions, déclenchant l'évolution d'une résistance chez le sexe opposé, généralement sur un trait différent. Quant aux IASC, ils agissent si un trait partage la même architecture génétique chez les deux sexes et est sélectionné de façon sexuellement antagoniste. Ces deux types de conflits peuvent interagir de plusieurs façons.

Prenons l'exemple d'un IRSC agissant sur un trait mâle augmentant la fréquence d'accouplement et engendrant l'évolution d'un trait femelle de résistance à l'accouplement (exemple tiré de la revue de Bonduriansky et Chenoweth 2009). La fitness mâle augmente avec la fréquence d'accouplement tandis que celle des femelles diminue, notamment à cause de l'investissement énergétique plus élevé des femelles pendant la gestation et les soins parentaux. Le premier type d'interaction possible entre les IASC et les IRSC se produit si les traits augmentant la résistance à l'accouplement partagent la même architecture génétique chez les deux sexes. Si c'est le cas, les mâles et les femelles ayant un certain génotype vont exprimer une résistance à l'accouplement, avantageuse chez les femelles mais désavantageuse chez les mâles, menant à l'émergence d'IASC sur le trait de résistance. Les IRSC peuvent donc accentuer les IASC. De façon symétrique, si le trait de résistance à l'accouplement des femelles est sous IASC, cela va influencer la co-évolution des traits sous IRSC. En effet, un IASC limite la réponse évolutive pour ce trait chez les femelles, les contraignant à demeurer en deçà de leur optimum en fitness tandis que le trait mâle pourra évoluer vers son optimum. Ceci pourrait expliquer pourquoi on observe souvent une absence d'évolution de contre-adaptations des femelles aux traits nocifs mâles.

Un troisième aspect de ces interactions est une conséquence de la résolution des IASC, aboutissant à une évolution indépendante entre les sexes des traits sous conflit. Si le trait augmentant la fréquence d'accouplement, premièrement sous IASC, évolue une expression sexe-spécifique, cela permet alors à la version mâle du trait de devenir exagérée au point d'être nocive chez la femelle, menant à l'émergence d'IRSC.

Encadré 1.1 – Les conflits sexuels

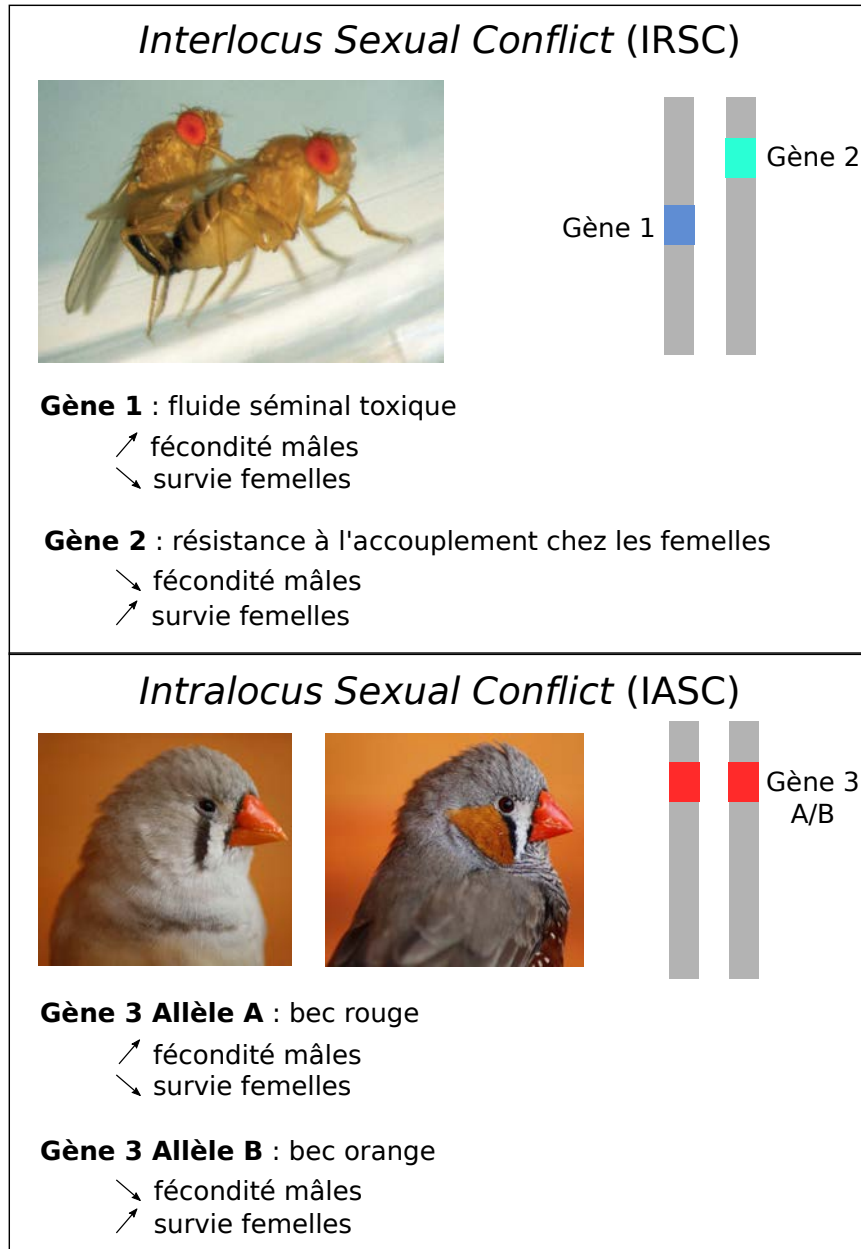


Figure 1.2 – Schéma simplifié d'un conflit sexuel interlocus (IRSC) et d'un conflit sexuel intralocus (IASC). IRSC- L'exemple illustré est l'IRSC sur le fluide séminal de *Drosophila melanogaster* (Chapman *et al.* 1995). Les IRSC agissent sur des traits différents (la toxicité du fluide séminal chez le mâle et la résistance à l'accouplement chez les femelles), impliqués dans les interactions mâles-femelles (ici l'accouplement). Le fluide séminal toxique est avantageux chez les mâles mais diminue la fitness des femelles, tandis que la résistance à l'accouplement des femelles est avantageux pour celles-ci mais diminue la fécondité des mâles (Crédit photographique : Chapman (2008)). IASC- L'exemple illustré est l'IASC agissant sur la couleur du bec chez *Taeniopygia guttata* (Price et Burley 1993, 1994). Les IASC agissent sur un seul trait, codé par les mêmes gènes chez les deux sexes, pour lequel chaque version sera sélectionnée de façon sexuellement antagoniste. Une couleur rouge du bec est avantageuse chez les mâles (à droite) mais désavantageuse chez les femelles (à gauche), et *vice versa* pour la couleur orange du bec (Crédit photographique : Maurice van Bruggen. License : CC-BY-SA).

1.2 Les conflits sexuels intralocus

1.2.1 Émergence des IASC

Les IASC sont une conséquence du partage d'un même génome entre les deux sexes associé à l'évolution d'antagonismes sélectifs entre les sexes. Ils se produisent lorsque les optimums en fitness sont différents entre mâles et femelles pour un trait dont l'architecture génétique est partagée entre les mâles et les femelles. Les IASC peuvent émerger par le biais de deux types de processus :

- **la sélection sexuelle** : l'émergence d'un trait avantageux pour un sexe lors de la sélection sexuelle, si elle engendre une réponse phénotypique similaire chez les deux sexes, peut avoir un effet désavantageux sur la survie ou la fertilité chez l'autre sexe. Dans ce cas, ce trait sera sous sélection sexuellement antagoniste (SA). Par exemple, la couleur rouge du bec chez le mandarin diamant (*Taeniopygia guttata*) est un trait ornemental avantageux chez le mâle car il favorise son accès aux femelles, mais désavantageux chez la femelle car il diminue son nombre de descendants et sa survie (Price et Burley 1993, 1994). De la même façon, la taille des cornes, un trait impliqué dans la compétition pour l'accès aux femelles, est sous IASC chez le mouton de Soay (*Ovis aries*). Des cornes plus grandes augmentent le succès reproducteur des mâles mais diminuent leur longévité, tandis que chez les femelles, elles diminuent le succès reproducteur et la longévité (Robinson *et al.* 2006). Cela suggère qu'il y aurait un coût à produire de longues cornes chez les deux sexes, contre-balançé chez les mâles par une augmentation du succès reproducteur.
- **la sélection naturelle** : comme expliqué auparavant, les mâles et les femelles ont des rôles reproductifs différents pouvant conduire à une divergence de leurs traits d'histoire de vie. Si les deux sexes ont des optimums différents pour des traits influençant leur survie ou leur fertilité, ce trait sera sous sélection SA. Par exemple, la taille corporelle chez le gobemouche à collier (*Ficedula albicollis*) est sous sélection SA, car une petite taille est avantageuse chez les mâles tandis qu'elle est désavantageuse chez les femelles (Merila *et al.* 1998). Une raison de cet antagonisme, avancée par les auteurs de cette étude, serait que les individus plus petits ont des besoins en ressources moins élevés pour survivre. Or, la période de retour de migration au printemps est différente entre les mâles et les femelles : les mâles reviennent plus d'une semaine avant les femelles en moyenne et entrent en compétition pour l'emplacement du nid. Il est possible que durant cette période stressante pour les mâles, les ressources disponibles étant plus faibles, des mâles de petite taille soient avantagés. Au contraire, les femelles revenant de migration plus tard, une taille corporelle

plus petite est désavantageuse.

1.2.2 Résolution des conflits sexuels intralocus

Les conflits sexuels intralocus ne peuvent être résolus que si les optimums sexe-spécifiques sont atteints chez les deux sexes par l'établissement d'une expression sexe-spécifique des gènes sous conflit. Cette résolution mène à l'émergence de dimorphismes sexuels (figure 1.3).

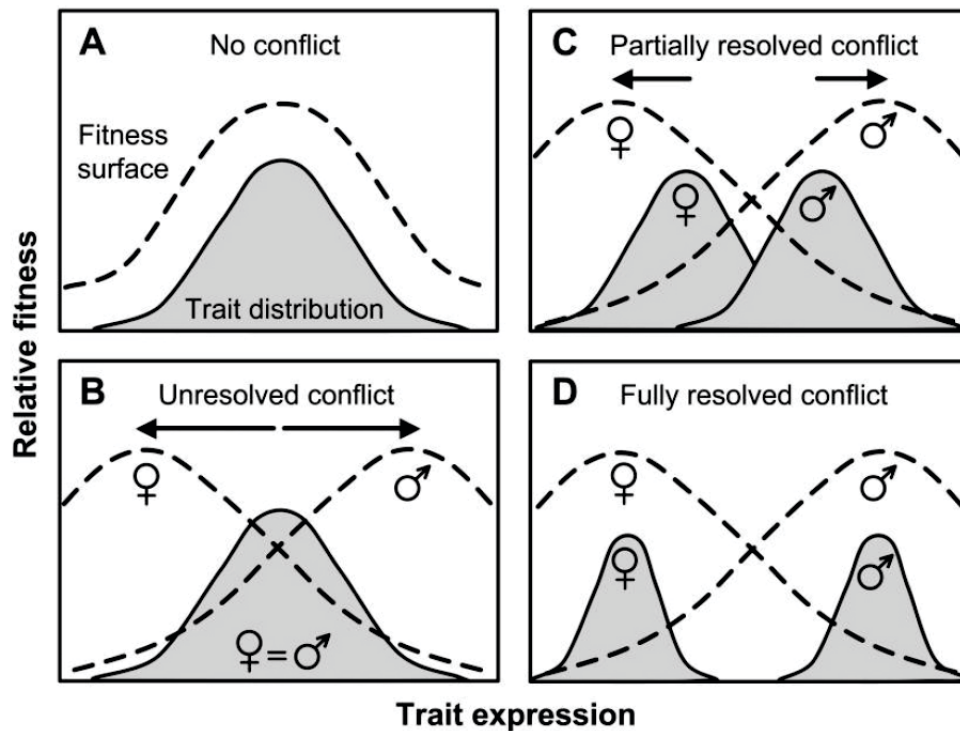


Figure 1.3 – Relation entre les conflits sexuels intralocus (IASC) et l'émergence des dimorphismes sexuels. A- Lorsqu'il n'y a pas de conflit, la distribution phénotypique du trait suit la distribution de la fitness associée à ce trait. Les deux distributions sont les mêmes chez les deux sexes. B- Si un IASC agit sur ce trait et n'est pas résolu, des optimums en fitness sexe-spécifiques existent mais la distribution phénotypique du trait est la même entre mâles et femelles, le trait est monomorphe. C- Au cours de la résolution du conflit, la distribution phénotypique du trait va se découpler et se déplacer vers les optimums sexe-spécifiques en fitness. Néanmoins, il existe toujours des phénotypes désavantageux partagés chez les deux sexes. Le trait est partiellement dimorphique entre les sexes. D- Quand le conflit est complètement résolu, les distributions phénotypiques sont sexe-spécifiques et suivent les distributions sexe-spécifiques en fitness. Le trait est sexuellement dimorphique. Tiré de Cox et Calsbeek (2009).

La facilité de résolution des IASC dépend majoritairement de la corrélation génétique inter-sexuelle pour le trait considéré (voir encadré 1.2) (Lande 1980). La résolution,

partielle ou complète, des IASC peut s'opérer de différentes façons :

- **la relocalisation physique du locus sexuellement antagoniste sur un chromosome sexuel** (Rice 1984). Le chromosome Y est caractérisé par le gain et la perte de nombreux gènes. Des études d'alignement de séquence chez l'humain ont montré que les plus proches paralogues de sept familles de gènes situés sur le chromosome Y sont localisés sur des autosomes, suggérant une relocalisation de ces gènes depuis les autosomes vers le chromosome Y (Bachtrog 2006). La plupart de ces gènes étant exprimés uniquement dans les testicules, leur localisation sur le chromosome Y aurait un avantage sélectif, permettant une évolution indépendante de ces gènes chez les mâles (Bachtrog 2006). Néanmoins, la localisation de certains gènes sous IASC sur le chromosome Y peut être désavantageuse s'ils ont également une fonction importante chez les femelles, ce processus ne peut donc pas être largement répandu.
- **l'addition d'un site de fixation d'un facteur sexe-spécifique trans dans un élément de régulation en cis du gène** (Williams et Carroll 2009). Ces facteurs sexe-spécifiques trans peuvent prendre la forme d'une hormone sexuelle ou d'une autre protéine codée par un gène qui aurait lui-même une expression sexe-spécifique (Rice 1984 ; Kopp *et al.* 2003). De fait, l'expression des caractères sexuels secondaires impliqués dans un processus de « signaux honnêtes » est régulée par la testostérone chez les vertébrés (Blas *et al.* 2006).
- **la duplication du gène sous IASC suivie de l'évolution d'une expression sexe-spécifique** (Ellegren et Parsch 2007). Chez *Drosophila*, il a été montré que les gènes exprimés uniquement chez les mâles présentent plus souvent des paralogues que les gènes exprimés chez les deux sexes et ont donc probablement été plus souvent dupliqués (Gnad et Parsch 2006).
- **l'évolution d'épissages alternatifs entre les sexes**. Par exemple, la part importante de gènes épissés de façon sexe-spécifique, 12% chez *Drosophila melanogaster* (McIntyre *et al.* 2006), pourrait refléter de nombreux cas d'IASC résolus.
- **l'évolution d'empreintes épigénétiques provenant des parents**, menant à une expression différente du gène considéré selon l'origine, paternelle ou maternelle, de l'allèle (Patten et Haig 2009). Ceci pourrait permettre de réduire chez l'enfant l'expression des allèles provenant du parent du sexe opposé pour les gènes sous IASC. Des empreintes épigénétiques sexe-spécifiques ont été trouvées chez la souris (Hager *et al.* 2008). Néanmoins, pour que ce mécanisme soit à l'origine d'une résolution d'IASC, l'empreinte épigénétique devrait être différente entre gamètes selon le chromosome sexuel qu'ils portent. Ainsi, les spermatozoïdes X et Y devraient avoir des empreintes différentes, ce qui n'a encore jamais été mis en évidence de façon

empirique (Pennell et Morrow 2013).

- **l'évolution d'un ajustement du sexe-ratio.** Un exemple de ce type de mécanisme a été décrit chez le lézard *Uta stansburiana*, chez qui la femelle opère un choix cryptique des spermatozoïdes du mâle en fonction de son phénotype (Calsbeek et Sinervo 2003). En effet, une grande taille corporelle est avantageuse chez les mâles mais désavantageuse chez les femelles, et il a été observé que les femelles se reproduisant avec des grands mâles avaient plus de fils tandis que les femelles se reproduisant avec des petits mâles avaient plus de filles. La nature de ce choix est cryptique et le mécanisme sous-jacent inconnu. D'autres exemples de mécanismes similaires ont été reportés chez plusieurs espèces (voir Pennell et Morrow 2013 pour une revue).
- **l'évolution d'une condition-dépendance sexe-spécifique** (Bonduriansky 2007). La condition-dépendance implique que l'exagération phénotypique du trait sous IASC sera dépendante de la condition physique de l'individu, elle-même fonction de la disponibilité des ressources dans l'environnement et de la qualité des gènes de l'individu impliqués dans l'acquisition et l'utilisation des ressources. Les individus avec la meilleure fitness vont développer les traits les plus exagérés, mais si les ressources sont faibles lors du développement de ces traits, ce trait sera moins exagéré. Cela permet de diminuer l'expression d'un trait coûteux à produire dans des conditions environnementales défavorables. Si le niveau de condition-dépendance est différent entre les sexes, il est possible que la femelle exprime toujours la version la moins extrême du trait, tandis que le mâle sera libre d'exprimer des versions plus extrêmes de façon indépendante. Il s'agit néanmoins d'une hypothèse qu'aucune étude empirique n'a encore étayée.

L'évolution d'un mécanisme de résolution des IASC peut être rapide, notamment l'évolution d'une expression sexe-spécifique des gènes codant pour les traits sous conflit. Il a en effet été montré chez de nombreux organismes qu'il existe une proportion importante de gènes avec une expression différente entre les sexes (voir van Doorn 2009 pour une revue). De plus, parmi les gènes ayant une expression différente entre *Drosophila simulans* et *Drosophila melanogaster*, 83% sont des gènes montrant un gain, une perte, une inversion ou une exagération d'expression sexe-spécifique (Ranz *et al.* 2003), ce qui indique que des mécanismes d'expression sexe-spécifique peuvent émerger rapidement.

Néanmoins, ces processus de résolution peuvent être considérablement ralentis, notamment si les gènes ont une expression ubiquitaire chez les deux sexes, donc potentiellement une fonction importante. Par exemple, chez la souris et le poulet, il a été montré que les gènes ayant une expression différente entre les sexes sont exprimés dans un plus petit panel de tissus que les gènes ayant une expression non biaisée (Mank *et al.* 2008).

Des gènes sous IASC avec une expression ubiquitaire et des fonctions multiples pourraient ainsi être plus lents à acquérir une expression sexe-spécifique. La résolution peut également être ralentie si l'architecture génétique codant pour les traits sous IASC est complexe. De plus, un niveau plus faible d'IASC pourrait subsister même après l'action d'un mécanisme de résolution. Par exemple, dans le cas d'une duplication du gène sous IASC suivie de l'évolution de son expression sexe-spécifique, des mutations délétères sur ce gène ne seraient pas contre-sélectionnées chez le sexe n'exprimant pas le trait sous IASC, et persisteraient plus longtemps. Cela n'aura pas d'effet sur la corrélation génétique inter-sexuelle (voir encadré 1.2) mais pourrait contribuer à la persistance d'IASC en éloignant le sexe exprimant le trait de son optimum en fitness.

Enfin, la fluctuation des pressions de sélection au cours du temps, due à un environnement changeant, est susceptible de ralentir l'ensemble de ces processus de résolution des IASC. Les modèles de co-évolution sexuelle prédisent des cycles évolutifs rapides donnant lieu à de nouveaux caractères sexuels secondaires (Bonduriansky et Chenoweth 2009). Par exemple, un IRSC peut être le moteur de nouvelles pressions de sélection sexuellement antagonistes et générer des IASC (voir encadré 1.1). Il a été montré que des antagonismes sexuels agissent sur des traits sexuellement dimorphiques (Pischedda et Chippindale 2006 ; Long et Rice 2007 ; Cox et Calsbeek 2009), ce qui est attendu si le niveau de dimorphisme n'est pas suffisant pour atteindre les optimums en fitness sexe-spécifiques (Cox et Calsbeek 2009). Il est donc possible que de nombreux IASC puissent être seulement résolus de façon partielle, conservant un niveau d'antagonisme sexuel (Pennell et Morrow 2013). Ainsi, l'observation de traits sous IASC non résolus est attendue chez les espèces sexuées.

La **corrélacion génétique inter-sexuelle** (r_{mf}) mesure la corrélation entre les effets d'un allèle chez un sexe et les effets de ce même allèle chez l'autre sexe. Cet estimateur permet d'évaluer la contrainte génétique agissant sur l'évolution indépendante d'un trait entre les sexes, qui dépend du degré auquel les deux sexes partagent la même architecture génétique pour ce trait. r_{mf} correspond au ratio entre la covariance additive entre les sexes (COV_{Amf}) et la moyenne géométrique de la variance additive des mâles (V_{Am}) et des femelles (V_{Af}) pour le trait :

$$r_{mf} = \frac{COV_{Amf}}{\sqrt{V_{Am} \times V_{Af}}}$$

r_{mf} est estimé à l'aide de pedigrees, et peut être calculé sur un trait phénotypique ou sur un estimateur de fitness. S'il est **calculé sur un trait phénotypique**, $r_{mf} = 1$ reflète une contrainte absolue à l'évolution de dimorphisme sexuel sur ce trait, car l'architecture et les effets additifs des gènes codants pour ce trait sont partagés entre les sexes (figure 1.4A). De plus, si des gradients de sélection dans des directions opposées entre les sexes sont observés pour le trait considéré, cela indique que ce trait est sous IASC. Lorsque $0 < r_{mf} < 1$, la contrainte n'est pas absolue et plus r_{mf} se rapproche de 0, plus l'évolution d'un dimorphisme sexuel sera facilitée.

Lorsque r_{mf} est **calculé sur un estimateur de fitness**, si $r_{mf} = 1$, la corrélation entre la fitness chez les mâles et les femelles est totale pour un même génotype, on conclut qu'il y a absence d'IASC. Si $r_{mf} < 0$, la corrélation est négative, un même génotype produit une réponse antagoniste sur la fitness entre les sexes, ce qui est concordant avec un processus d'IASC (figure 1.4B). Si $0 \leq r_{mf} < 1$, on ne peut pas conclure car une sélection chez un seul des deux sexes peut être à l'origine de valeurs de r_{mf} comprises dans cet intervalle.

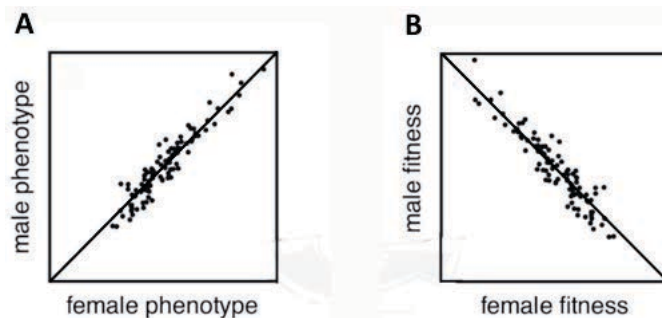


Figure 1.4 Représentation de r_{mf} pour **A-** Un trait phénotypique : $r_{mf} \approx 1$ **B-** un estimateur de fitness : $r_{mf} \approx -1$. Chaque point représente un génotype. Adapté de van Doorn (2009).

Encadré 1.2 – La corrélation génétique inter-sexuelle. Texte adapté d'après Bonduriansky et Chenoweth (2009) et figure adaptée d'après van Doorn (2009).

1.2.3 Implications évolutives des IASC

Évolution des chromosomes sexuels

Chez les mammifères, les chromosomes sexuels ont divergé depuis une paire autosomale il y a 300 millions d'années grâce à la spécialisation d'un gène pour la détermination du sexe mâle, SRY, sur le chromosome Y (Ohno 1967 ; Graves et Shetty 2001). L'évolution d'un autre système de détermination du sexe, le système ZW, a eut lieu indépendamment et est trouvé notamment chez les oiseaux, des reptiles et les papillons. Dans ce système la femelle est le sexe hétérogamétique (ZW) tandis que le mâle est le sexe homogamétique (ZZ), ce qui entraîne des différences majeures avec le système XY pour l'évolution des dimorphismes sexuels (voir ci-dessous). Une accumulation d'allèles sexuellement antagonistes sur le chromosome Y (respectivement W) serait à l'origine de la suppression progressive de la recombinaison entre le X et le Y (respectivement Z et W). En effet, une liaison au locus déterminant le sexe favoriserait l'accumulation d'allèles sexuellement antagonistes favorables au sexe hétérogamétique (Fisher 1958 ; Charlesworth et Charlesworth 1980). Une suppression de la recombinaison entre ces locus et le locus déterminant le sexe est ainsi avantageuse (Bull 1983 ; Rice 1987). L'accumulation de locus sous sélection SA serait donc à l'origine de l'élargissement de la zone non recombinante des chromosomes sexuels. Les régions recombinantes des chromosomes sexuels sont appelées les régions pseudo-autosomales (*Pseudo-autosomal regions*, PAR), et sont situées aux extrémités des chromosomes sexuels. Des PAR sont observées chez de nombreuses espèces, et ont différentes tailles selon les espèces (Otto *et al.* 2011). Leur maintien s'explique par la nécessité de l'appariement des deux chromosomes sexuels pendant la méiose, afin que la ségrégation se déroule ensuite normalement. Chez l'Homme, il existe deux PAR qui sont situées aux deux extrémités des chromosomes X et Y, et qui constituent de petites régions d'environ 2,7 Mb et 0,33 Mb chacune (d'après les coordonnées hg19). La structure des chromosomes sexuels humains est représentée en figure 1.5. Tandis que PAR1 est partagée par la majorité des euthériens, seul l'Homme possède PAR2 (Graves 2006).

L'accumulation de locus SA sur les chromosomes sexuels est aussi à l'origine d'une spécialisation fonctionnelle des chromosomes sexuels. Pour les espèces XY, on s'attend en effet à ce que le chromosome Y accumule des allèles avantageux pour les mâles et désavantageux pour les femelles, et que le chromosome X accumule des allèles dominants avantageux pour les femelles mais désavantageux pour les mâles, ainsi que des allèles récessifs avantageux pour les mâles mais désavantageux pour les femelles (Rice 1984) (voir plus de détails section 2.1.1). Ceci expliquerait pourquoi les chromosomes sexuels sont enrichis en fonctions impliquées dans la reproduction. En effet, chez l'Homme, les quelques gènes spécifiques au chromosome Y sont en majorité impliqués dans la spermatogénèse

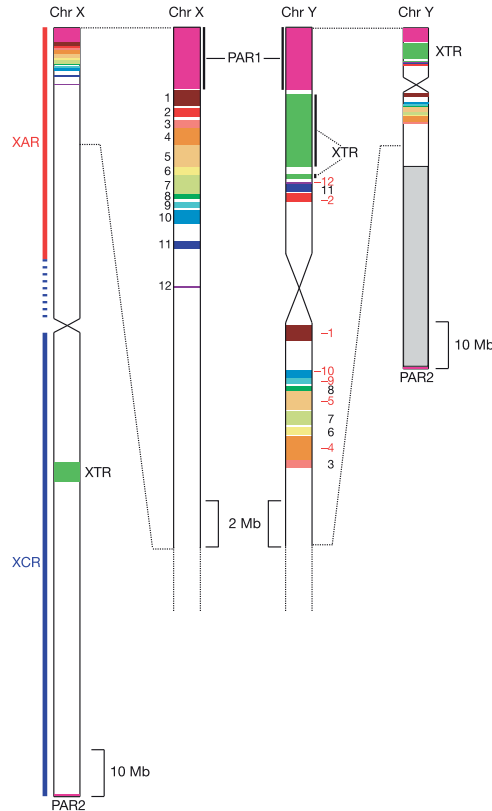


Figure 1.5 – Structure des chromosomes sexuels chez l’Homme. PAR1 et PAR2 correspondent aux régions pseudo-autosomales. La région XCR (*X Conserved Region*) est homologue chez tous les mammifères. La région XAR (*X Added Region*) est homologue à une région autosomale chez les marsupiaux et les monotrèmes, et a été ajoutée par translocation depuis un autosome avant la radiation des euthériens. Les régions d’homologie entre le chromosome X et le chromosome Y sont indiquées par les mêmes couleurs. La région XTR (*X transposed Region*) est le résultat d’une récente transposition d’une région du chromosome X vers le chromosome Y. Tiré de Ross *et al.* (2005).

(Lahn *et al.* 2001), tandis que le chromosome X est enrichi en gènes impliqués dans la reproduction et les fonctions cérébrales (Ross *et al.* 2005). Une hypothèse formulée par Graves (2006) propose que l’enrichissement du chromosome X en gènes impliqués dans les fonctions cérébrales soit le résultat d’un processus de sélection sexuelle, en l’occurrence le choix d’hommes intelligents par les femmes.

Évolution du déterminisme du sexe

La sélection sexuellement antagoniste joue aussi un rôle dans les transitions entre mécanismes de détermination du sexe. Une transition depuis une détermination environnementale du sexe (ESD) vers une détermination génétique du sexe (GSD) peut être favorisée par l’action de la sélection sexuellement antagoniste. En effet, les modèles théo-

riques prédisent qu'un allèle déterminant le sexe augmente en fréquence dans la population s'il est lié à un gène ayant un effet bénéfique chez le sexe que ce locus détermine. Les transitions entre ESD et GSD, ou entre deux systèmes GSD différents, sont répandus dans plusieurs genres (voir van Doorn 2009 pour une revue).

Évolution des dimorphismes sexuels en réponse au choix du conjoint

L'évolution de traits impliqués dans la sélection sexuelle, les ornements ou armements, implique souvent l'émergence d'IASC. En effet, ces traits sont coûteux à produire et n'apportent pas de bénéfice direct chez les femelles, ce qui les rend désavantageux, tandis qu'ils apportent un bénéfice au niveau de l'accès à la reproduction chez les mâles.

Selon l'hypothèse de Rice (1984), le chromosome X est enrichi en locus sous sélection sexuellement antagoniste. Or, les femelles transmettent leurs chromosomes X à leurs fils et à leurs filles, tandis que les mâles ne transmettent leur chromosome X qu'à leurs filles. Cela signifie qu'un chromosome X entraînant une fitness élevée pour les mâles ne sera pas transmis de père en fils. Si l'on considère que les locus sexuellement antagonistes sont plus souvent localisés sur le chromosome X, cela a une influence considérable sur la sélection sexuelle, notamment sur le choix du conjoint effectué par la femelle (Albert et Otto 2005). En effet, une femelle choisissant un mâle avec une fitness élevée, présentant ainsi des allèles avantageux pour les mâles mais désavantageux pour les femelles préférentiellement localisés sur le chromosome X, va désavantager ses filles sans que ses fils ne reçoivent les bénéfices associés au chromosome X de leur père.

Chez les oiseaux et autres animaux possédant un système de chromosome sexuel ZW, le mâle est le sexe homogamétique et transmet donc ses chromosomes Z à ses fils et à ses filles. Le processus de sélection sexuelle est facilité dans cette situation. En effet, il a été montré que les préférences des femelles mènent plus facilement à une accumulation de traits sexuellement antagonistes avantageux pour les mâles dans les systèmes ZW en comparaison aux systèmes XY (Albert et Otto 2005). Cela pourrait expliquer pourquoi, chez les oiseaux, le degré de dimorphisme sexuel est plus important que chez des espèces XY, notamment lorsque l'on considère les dimorphismes sexuels d'apparat (quantifié par Reeve et Pfennig 2003). Au contraire, dans les systèmes XY, cela pourrait mener à une évolution des préférences des femelles pour des mâles avec des fitness basses (Albert et Otto 2005). La présence d'IASC peut apporter des explications au maintien de la diversité génétique dans le cas d'un choix des femelles pour des mâles avec un certain phénotype, processus qui devrait mener à une fixation de ce phénotype dans la population si cette préférence est maintenue de générations en générations (« the lek paradox ») (Taylor et Williams 1982 ; Kirkpatrick et Ryan 1991).

Spéciation

Les IASC pourraient aussi être impliqués dans des processus de spéciation. Les processus évolutifs opérant dans les cas de sélection sexuellement antagoniste étant très dynamiques, des différences de pressions de sélection entre populations pourraient émerger et mener à une évolution de traits sexuellement dimorphiques différents. De plus, on s'attend à ce que les mécanismes de résolution de ces conflits soient perturbés chez les hybrides, menant à un isolement reproducteur post-zygotique (Parker et Partridge 1998 ; Rice et Chippindale 2002). Enfin, une interaction entre IASC et IRSC (voir encadré 1.1), menant à une course à l'armement entre les sexes et donc à une co-évolution rapide des caractères sexuels secondaires entre mâles et femelles, pourrait mener à un événement de spéciation entre deux populations isolées.

1.3 Détection de sélection sexuellement antagoniste

De façon classique, la sélection sexuellement antagoniste est détectée au sein d'une population grâce à l'observation de corrélation négative pour des indicateurs de différentes composantes de fitness entre mâles et femelles, c'est-à-dire la survie et le succès reproducteur. Deux types de méthodes sont utilisés :

- le calcul de la corrélation génétique inter-sexuelle (r_{mf}) sur une composante de fitness (voir encadré 1.2), permettant de déterminer si un même génotype a un effet antagoniste entre les sexes. Ce type de méthode ne peut être utilisé que sur des populations de laboratoire, grâce à l'utilisation d'hémiclones. Des hémiclones ont la moitié de leur génome en commun avec les autres hémiclones et l'autre moitié est sélectionnée au hasard. Cette méthode a été développée par Chippindale *et al.* (2001) chez *Drosophila melanogaster*, et permet de quantifier l'effet d'un même génotype sur la fitness chez les deux sexes.
- la comparaison de composantes de fitness chez les parents et les enfants dans des populations naturelles ou de laboratoire. Une corrélation négative entre la fitness chez les parents et chez les enfants, selon leur sexe, c'est-à-dire l'observation d'une fitness diminuée chez les filles d'individus mâles avec une fitness haute, et d'une fitness plus élevée chez les filles d'individus mâles avec une fitness basse, et *vice versa* chez les fils, est un signe d'IASC.

En utilisant ces méthodes, de nombreuses études empiriques ont démontré que la sélection sexuellement antagoniste agissait sur des populations naturelles et expérimentales, notamment chez *Drosophila melanogaster* (Chippindale *et al.* 2001 ; Innocenti et Morrow 2010) et *Drosophila serrata* (Delcourt *et al.* 2009) à l'aide de la première méthode, le cerfs élaphe (*Cervus elaphus*) (Foerster *et al.* 2007) et le gobemouche à collier (*Ficedula albi-*

collis) (Brommer *et al.* 2007) à l'aide de la seconde méthode (voir aussi Cox et Calsbeek 2009 pour une revue).

Ces méthodes permettent de savoir si des IASC agissent dans le génome des populations étudiées, néanmoins elles ne permettent pas de connaître précisément la cible de la sélection sexuellement antagoniste.

Ces conflits étant à l'origine de l'évolution de dimorphismes sexuels, l'observation de phénotypes différents entre mâles et femelles peut être un indicateur de l'action, passée ou présente, d'IASC. Néanmoins, un dimorphisme sexuel peut évoluer par le biais d'autres mécanismes, par exemple si le trait considéré partage des contraintes développementales avec d'autres traits sexuellement dimorphiques, s'il est codé par des gènes situés sur des chromosomes présents chez un sexe uniquement (le chromosome Y ou W), ou s'il n'existe pas de corrélation génétique inter-sexuelle ($r_{mf} = 0$) pour ce trait. Dans les cas de traits pour lesquels le conflit n'est pas résolu, la distribution phénotypique du trait est la même chez les deux sexes (figure 1.3), il est donc impossible de les détecter. En conséquence, identifier des traits spécifiques sous IASC soulève plusieurs difficultés et la seule observation d'un dimorphisme sexuel ne permet pas de conclure que le trait est sous IASC.

La méthode classiquement utilisée pour détecter les traits sous IASC associe l'observation d'une corrélation génétique inter-sexuelle positive pour ce trait, r_{mf} , et l'observation de gradients de sélection dans des directions opposées entre les sexes pour le trait considéré (voir encadré 1.2). Des exemples de traits sous IASC ont été trouvés chez *Drosophila melanogaster* pour le comportement locomoteur (Long et Rice 2007), chez le mandarin diamant (*Taeniopygia guttata*) pour la couleur du bec (Price et Burley 1993, 1994) et chez le gobemouche à collier (*Ficedula albicollis*) pour la taille corporelle (Merila *et al.* 1998). Chez l'Homme, des preuves de l'existence d'IASC ont été trouvées pour la taille (Stulp *et al.* 2012), l'expression de la masculinité/féminité anthropométrique et les niveaux endogènes d'œstrogène et de testostérone (Garver-Apgar *et al.* 2011). Ces études ont mis en évidence les cibles phénotypiques des IASC, néanmoins les locus spécifiques sous sélection sexuellement antagoniste restent inconnus.

Une étude a permis de mettre en évidence des gènes présentant une expression biaisée entre les sexes en analysant le transcriptome d'une population expérimentale de *Drosophila melanogaster* (Innocenti et Morrow 2010). En réalisant une régression linéaire entre l'expression génique, la fitness, le sexe, et l'interaction entre la fitness et le sexe, les auteurs ont montré que seuls 8,5% des gènes exprimés différemment entre les sexes montraient un signal d'IASC, c'est-à-dire une interaction significative entre le sexe et

la fitness. Cela suggère qu'une expression sexe-spécifique reflète en majorité des conflits résolus plutôt qu'un signal d'IASC en cours. Néanmoins, les IASC détectés dans cette étude sont déjà partiellement résolus par l'établissement d'une expression biaisée entre les sexes et ne représentent qu'une partie des IASC potentiellement en cours dans le génome de *Drosophila melanogaster*.

Au cours de cette thèse, nous nous sommes intéressée à la détection d'IASC chez l'Homme en utilisant des bases de données publiques de génotypage et de séquençage. En effet, ces conflits sont susceptibles d'être détectables dans le génome au niveau de la population car ils génèrent des différences de fréquences alléliques entre les sexes. Il est intéressant de travailler sur l'Homme pour ce type de problématiques car c'est, à notre connaissance, l'espèce pour laquelle le plus de données génétiques sont disponibles pour de nombreuses populations. Ceci nous a permis de mettre en évidence une liste de locus potentiellement sous sélection SA dans des populations humaines. Nous avons ensuite effectué une étude fonctionnelle de ces locus afin d'identifier les fonctions sous sélection SA chez l'Homme. Dans un second temps, nous avons étudié un mécanisme pouvant être à l'origine de différences de fréquences alléliques entre les sexes au sein d'une population : une distorsion de transmission sexe-spécifique. Pour cela, nous avons développé une nouvelle approche, combinant plusieurs méthodes de génétique des populations, sur des données de génomes complets à l'aide de trios parents-enfant.

Chapitre 2

Détection de sélection sexuellement antagoniste dans le génome humain

2.1 Introduction

Pour résoudre les IASC, une variabilité génétique à l'origine de différences phénotypiques conduisant à l'expression des deux versions antagonistes du trait considéré peut être nécessaire. Il est donc attendu qu'au cours de la phase de résolution de ces conflits les variabilités génétiques conduisant aux différentes versions du trait sexuellement antagoniste ségrègent chez les deux sexes. Des études théoriques ont montré que, sous certaines conditions, un polymorphisme au niveau d'un locus sous IASC peut persister et atteindre l'équilibre dans une population (Rice 1984 ; Kidwell *et al.* 1977).

La localisation génomique des locus SA est sujette à débat dans la littérature. Un modèle classique de génétique des populations développé par Rice (1984) prédit que le chromosome X offre un environnement plus favorable à l'accumulation de locus sous sélection SA que les autosomes. Dans cette introduction, nous allons détailler le modèle de Rice (1984), ainsi que ses limitations, discutées par Fry (2010). Dans un second temps, nous allons rassembler les résultats des études empiriques visant à mettre en évidence la localisation des locus sous sélection SA. Enfin, nous allons nous intéresser à un signal de l'action de la sélection SA, des différences de fréquences alléliques entre les sexes.

2.1.1 La localisation génomique des locus SA prédite par le modèle de Rice

Rice (1984), à l'aide d'un modèle de génétique des populations, a établi les conditions d'émergence et de maintien d'un polymorphisme au niveau d'un gène sexuellement antagoniste dans une population et a comparé ces conditions entre le chromosome X (PAR

exclus) et les autosomes.

Le modèle de Rice considère un locus bi-allélique sous sélection sexuellement antagoniste : A_1 est un allèle rare ayant un effet sexuellement antagoniste sur la fitness et A_2 est l'allèle établi et a le même effet sur la fitness chez les deux sexes.

Plusieurs paramètres sont utilisés :

- deux paramètres de sélection, S et T , compris entre 0 et 1. S correspond au gain en fitness que procure l'allèle A_1 à un sexe, et T correspond au coût en fitness que procure l'allèle A_1 au sexe opposé.
- un paramètre de dominance, h , compris entre 0 et 1. Si $h = 0$, l'allèle A_1 est complètement récessif et si $h = 1$, l'allèle A_1 est complètement dominant.

La fréquence de l'allèle A_1 est notée p , et celle de l'allèle A_2 est notée q , tel que $p + q = 1$. On fait la distinction entre la fréquence de A_1 et A_2 chez les mâles et les femelles grâce aux indices m et f .

Le cas autosomal

Ici, on considère que l'allèle A_1 est bénéfique chez les mâles et est désavantageux chez les femelles, mais les résultats sont applicables à l'autre cas. Le modèle de sélection pour un autosome est résumé table 2.1.

Sexe	Mâles			Femelles		
Génotype	A_1A_1	A_1A_2	A_2A_2	A_1A_1	A_1A_2	A_2A_2
Fréquence du zygote	$p_m p_f$	$p_m q_f + p_f q_m$	$q_m q_f$	$p_m p_f$	$p_m q_f + p_f q_m$	$q_f q_m$
Fitness (W)	$1 + S$	$1 + hS$	1	$1 - T$	$1 - hT$	1

Table 2.1 – Modèle de sélection sexuellement antagoniste pour un locus autosomal. Tiré de Rice (1984).

La fitness moyenne chez chaque sexe peut être exprimée par :

$$\begin{aligned}\overline{Wm} &= 1 + S(p_m p_f + h(p_m q_f + p_f q_m)) \\ \overline{Wf} &= 1 - T(p_m p_f + h(p_m q_f + p_f q_m))\end{aligned}$$

La fréquence p de A_1 dans la population correspond à la moyenne arithmétique de p_m et p_f . Les variations de p par génération sont exprimées par Δp dont l'équation est donnée ci dessous (équation 2.1). Si Δp est positif, cela signifie que l'allèle A_1 augmente

en fréquence.

$$\Delta p = \frac{p_m p_f (1 + S) + \left[\frac{p_m q_f + p_f q_m}{2} \right] (1 + hS)}{\overline{W}_m} + \frac{p_m p_f (1 - T) + \left[\frac{p_m q_f + p_f q_m}{2} \right] (1 - hT)}{\overline{W}_f} \quad (2.1)$$

Quand p est petit, $\overline{W}_m \approx \overline{W}_f \approx 1$ et l'équation 2.1 peut être simplifiée :

$$\Delta p = \frac{S - T}{2} [p_m p_f + (h/2)(p_m q_f + p_f q_m)] \quad (2.2)$$

D'après l'équation 2.2, Δp est positif uniquement si $S > T$. Par conséquent, un allèle sexuellement antagoniste augmentera en fréquence, quand il est rare, sur un autosome seulement si le gain en fitness chez un sexe excède le coût en fitness chez l'autre sexe.

La liaison à l'X

Nous allons ici considérer le cas du système XY, mais le modèle est applicable au système ZW. On considère que le chromosome Y ne contient pas les gènes trouvés sur le chromosome X, et que les deux chromosomes X sont actifs chez le sexe homogamétique, c'est-à-dire que les deux chromosomes contribuent au phénotype de façon à ce que la dominance entre allèles puisse être exprimée au niveau de l'organisme. Le cas des régions pseudo-autosomales n'est pas décrit dans ce chapitre.

Sexe	Mâles		Femelles		
Génotype	A_1	A_2	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$
Fréquence du zygote	p_f	q_f	$p_m p_f$	$p_m q_f + p_f q_m$	$q_f q_m$
Fitness (W1)	$1 + S$	1	$1 - T$	$1 - hT$	1
Fitness (W2)	$1 - T$	1	$1 + S$	$1 + hS$	1

Table 2.2 – Modèle de sélection sexuellement antagoniste pour un locus lié au chromosome X. W1 correspond aux cas où A_1 est avantageux chez les mâles, et W2 correspond au cas où A_1 est avantageux chez les femelles. Tiré de Rice (1984).

A_1 est avantageux chez les mâles

Dans ce cas, l'allèle A_1 est avantageux chez les mâles, et désavantageux chez les femelles. Le modèle de sélection correspondant est résumé table 2.2 (ligne W1).

Si A_1 est avantageux chez les mâles, l'allèle sexuellement antagoniste augmentera en fréquence quand il est rare si l'inégalité suivante est vraie :

$$S > \frac{2hT}{1 - hT} \quad (2.3)$$

Cela signifie que :

- si A_1 est au moins partiellement dominant ($h > 0.5$), les conditions pour l'augmentation en fréquence de A_1 sont plus restrictives que si le locus se trouve sur un autosome,
- si A_1 est suffisamment récessif ($h < 1/(2 + T)$), A_1 pourra augmenter en fréquence même si le coût en fitness chez les femelles est plus élevé que le gain en fitness chez les mâles.

Les fréquences à l'équilibre de A_1 pour chaque sexe dépendent de S , T et h :

$$\hat{p}_f = \frac{S - hT(2 + S)}{2T[(1 + S) - (hT(2 + S))]} \quad (2.4)$$

$$\hat{p}_m = \frac{S^2(1 - hT) + S(1 - 3hT) - 2hT}{S^2(1 - hT) + ST(2 - 4h) + T(2 - 4h)} \quad (2.5)$$

A_1 est avantageux chez les femelles

Dans ce cas, l'allèle A_1 est avantageux chez les femelles, et désavantageux chez les mâles. Le modèle de sélection correspondant est résumé table 2.2 (ligne W2).

Les conditions requises pour que A_1 augmente en fréquence quand il est rare sont exprimées par l'inégalité suivante :

$$S > \frac{T}{h(2 - T)} \quad (2.6)$$

Ceci indique que :

- si A_1 est suffisamment dominant ($h > 1/(1 - T)$), les conditions pour l'augmentation en fréquence de A_1 sont moins stringentes que sur un autosome,
- si A_1 est complètement dominant ($h = 1.0$), et T et S sont petits, l'inégalité 2.6 peut être approximée par $S > T/2$. Quand S et T sont petits, A_1 peut donc augmenter en fréquence même si le désavantage chez les mâles est presque deux fois plus grand que l'avantage chez les femelles.

Les valeurs à l'équilibre des fréquences de A_1 chez chaque sexe sont données par les équations suivantes :

$$\hat{p}_f = \frac{2 - T - \left[\frac{2}{1 + hS} \right]}{4 - 2T - \left[\frac{2 + 2(1 - T)(1 + S)}{1 + hS} \right]} \quad (2.7)$$

$$\hat{p}_m = \frac{(1 - T) \left(2 - T - \left[\frac{2}{1 + hS} \right] \right)}{(1 - T) \left(2 - T - \left[\frac{2}{1 + hS} \right] \right) + 2 - T - \left[\frac{2(1 + S)(1 - T)}{1 + hS} \right]} \quad (2.8)$$

En résumé, ce modèle prédit qu'un nouvel allèle sous sélection SA augmentera en fréquence, quand il est rare, si :

- **pour les autosomes** : l'avantage d'un sexe est supérieur au coût chez l'autre sexe,
- **pour le chromosome X** : il est avantageux pour les femelles et dominant ($h > 1/(2 - T)$); ou s'il est avantageux pour les mâles et récessif ($h < 1/(2 + T)$); et cela même si le coût en fitness chez un sexe est plus élevé que le gain en fitness chez l'autre sexe.

Ce modèle prédit donc que le chromosome X devrait être un *hotspot* pour l'accumulation de locus sous sélection SA en comparaison aux autosomes, c'est-à-dire que le chromosome X devrait être enrichi en locus sous sélection SA.

2.1.2 Critiques de la prédiction du modèle de Rice

Une étude discute l'hypothèse de Rice et conclut qu'il est prématuré de considérer le chromosome X comme un *hotspot* pour l'accumulation de locus sous sélection SA (Fry 2010). En effet, l'auteur met en avant une limitation du modèle de Rice : une des hypothèses sur laquelle le modèle a été construit est que la dominance chez les deux sexes est la même. Si l'on considère que la dominance entre les sexes peut différer et que l'allèle bénéfique chez chaque sexe est dominant chez celui-ci, un polymorphisme autosomal est possible pour une gamme de paramètres plus étendue que si le locus est lié au chromosome X.

Réévaluation du modèle de Rice par Fry

En utilisant les notations de Kidwell *et al.* (1977), Fry (2010) a réécrit le modèle de Rice. Deux paramètres de dominance sont utilisés, h_m qui représente la dominance de l'allèle désavantageux chez le mâle, et h_f qui représente la dominance de l'allèle désavantageux chez la femelle. De plus, les paramètres de fitness sont différents : s_m et s_f sont

les coefficients de sélection contre l'allèle délétère pour chaque sexe, et la fitness est comprise entre 0 et 1. Si l'allèle désavantageux chez un sexe s'exprime à l'état hétérozygote, la fitness sera diminuée selon la valeur du paramètre de sélection et de dominance chez chaque sexe ($1 - h_m s_m$ chez les mâles et $1 - h_f s_f$ chez les femelles). La correspondance entre les notations de Rice et de Kidwell est donnée en table 2.3. Le modèle de sélection pour les autosomes et le chromosome X sont représentés table 2.4 et 2.5 respectivement.

Kidwell <i>et al.</i> (1977)	Rice (1984)	
	Allèle rare avantageux chez les mâles	Allèle rare avantageux chez les femelles
s_f	T	$S/(1 + S)$
s_m	$S/(1 + S)$	T
h_f	h	$1 - h$
h_m	$1 - h$	h

Table 2.3 – Correspondance entre les notations de Kidwell *et al.* (1977) et Rice (1984). Adapté de Fry (2010).

Sexe	Mâles		Femelles		
Génotype	A_1	A_2	A_1A_1	A_1A_2	A_2A_2
Fitness	$1 - s_m$	1	1	$1 - h_f s_f$	$1 - s_f$

Table 2.4 – Modèle de sélection sexuellement antagoniste pour un locus lié au chromosome X, dans le cas où A_1 est avantageux chez les femelles et désavantageux chez les mâles, et où A_2 est avantageux chez les mâles et désavantageux chez les femelles. Adapté de Fry (2010).

Sexe	Mâles			Femelles		
Génotype	A_1A_1	A_1A_2	A_2A_2	A_1A_1	A_1A_2	A_2A_2
Fitness	$1 - s_m$	$1 - h_m s_m$	1	1	$1 - h_f s_f$	$1 - s_f$

Table 2.5 – Modèle de sélection sexuellement antagoniste pour un locus autosomal, dans le cas où A_1 est avantageux chez les femelles et désavantageux chez les mâles, et où A_2 est avantageux chez les mâles et désavantageux chez les femelles. Adapté de Fry (2010).

Dans son modèle, Rice (1984) considère que, dans le cas autosomal, la dominance d'un allèle donné est la même chez les deux sexes, se traduisant par $h_m = 1 - h_f$ (et $h_f = 1 - h_m$) avec les notations de Kidwell *et al.* (1977). Dans ces conditions, un polymorphisme à un locus sexuellement antagoniste autosomal se produit quand :

$$\frac{1}{1 + s_f} < \frac{s_m}{s_f} < \frac{1}{1 - s_f} \quad (2.9)$$

La première partie de l'inégalité donne les conditions d'augmentation en fréquence de l'allèle A_2 quand il est avantageux chez les mâles et désavantageux chez les femelles tandis que la deuxième partie de l'inégalité donne les conditions d'augmentation de l'allèle A_1 quand il est avantageux chez les femelles et désavantageux chez les mâles. Ces deux inégalités ne dépendent pas de la dominance. De la même façon, pour un locus sexuellement antagoniste lié au chromosome X, un polymorphisme se produit quand :

$$\frac{2h_f}{1 + h_f s_f} < \frac{s_m}{s_f} < \frac{2(1 - h_f)}{1 - h_f s_f} \quad (2.10)$$

Cette inégalité dépend de la dominance chez les femelles, h_f .

Les conditions de maintien d'un polymorphisme à un locus SA autosomal et à un locus SA lié au chromosome X sous le modèle de Rice sont représentées en figure 2.1.

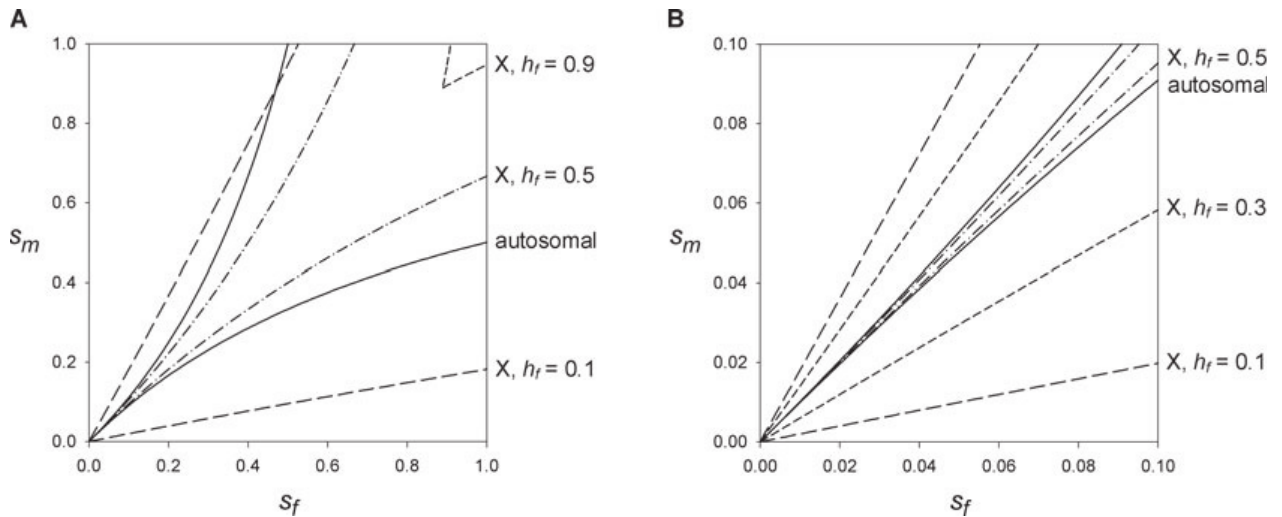


Figure 2.1 – Condition de maintien d'un polymorphisme sexuellement antagoniste à un locus lié au chromosome X et à un locus autosomal, sous l'hypothèse que la dominance est la même chez les mâles et les femelles. Le polymorphisme est maintenu dans la région entre les lignes identiques. La région sous la ligne supérieure correspond aux combinaisons de paramètres permettant l'augmentation en fréquence quand rare de l'allèle A_1 (avantageux chez les femelles, désavantageux chez les mâles), tandis que la région au dessus de la ligne inférieures correspond aux combinaisons de paramètres permettant l'augmentation en fréquence quand rare de l'allèle A_2 (avantageux chez les mâles et désavantageux chez les femelles). Les lignes pleines correspondent au cas autosomal tandis que les lignes en pointillées correspondent aux cas liés à l'X. **A-** prend en compte des paramètres de sélection compris entre 0.0 et 1.0, **B-** prend en compte des paramètres de sélection plus réalistes, compris entre 0.0 et 0.1, et correspond à l'agrandissement d'une partie de **A**. Tiré de Fry (2010).

Sous l'hypothèse que la dominance est la même chez les deux sexes ($h_f = 1 - h_m$), comme décrit en section 2.1.1, dans les cas où le locus est lié à l'X et où l'allèle bénéfique aux femelles est partiellement dominant (l'allèle désavantageux est partiellement récessif), la gamme de paramètres permettant le maintien d'un polymorphisme est plus étendue que pour le cas autosomal (figure 2.1).

Les cas où la gamme de paramètres permettant le maintien d'un polymorphisme est la plus étendue sont, par ordre décroissant (figure 2.1A) :

1. les cas où le locus est lié à l'X et où l'allèle bénéfique aux femelles est partiellement dominant (l'allèle désavantageux est partiellement récessif),
2. les cas où le locus est autosomal ,
3. les cas où le locus est lié à l'X et où les deux allèles ont des effets additifs ($h_f = 0.5$),
4. les cas où le locus est lié à l'X et où l'allèle délétère chez les femelles est presque complètement dominant ($h_f = 0.9$).

Pour des coefficients de sélection réalistes (s_m et $s_f < 0.1$), les cas qui permettent le maintien d'un polymorphisme à un locus SA sont les cas liés à l'X avec un h_f faible (figure 2.1B).

Ces prédictions sont valides si l'on considère que la dominance est la même chez les deux sexes, ce qui est observé dans le cas de la plupart des mutations les plus visibles documentées dans la littérature et dans le cas de maladies mendéliennes chez l'Homme (Fry 2010). Néanmoins, Fry (2010) critique cette hypothèse car les allèles ayant des effets sexuellement antagonistes ségrégeant dans la population devraient avoir des effets moins drastiques.

Fry (2010) adapte le modèle en permettant aux paramètres de dominance de varier entre les sexes pour un locus autosomal et compare les conditions de maintien d'un polymorphisme SA entre les autosomes et le chromosome X. D'après Kidwell *et al.* (1977), un polymorphisme SA à un locus autosomal sera maintenu si :

$$\frac{h_f}{1 - h_m + h_f s_f} < \frac{s_m}{s_f} < \frac{1 - h_f}{h_m(1 - s_f)} \quad (2.11)$$

Si l'allèle bénéfique est dominant chez chaque sexe, c'est-à-dire quand h_m et $h_f < 0.5$, un polymorphisme est maintenu pour une gamme de paramètres plus importante si le locus est autosomal (figure 2.2).

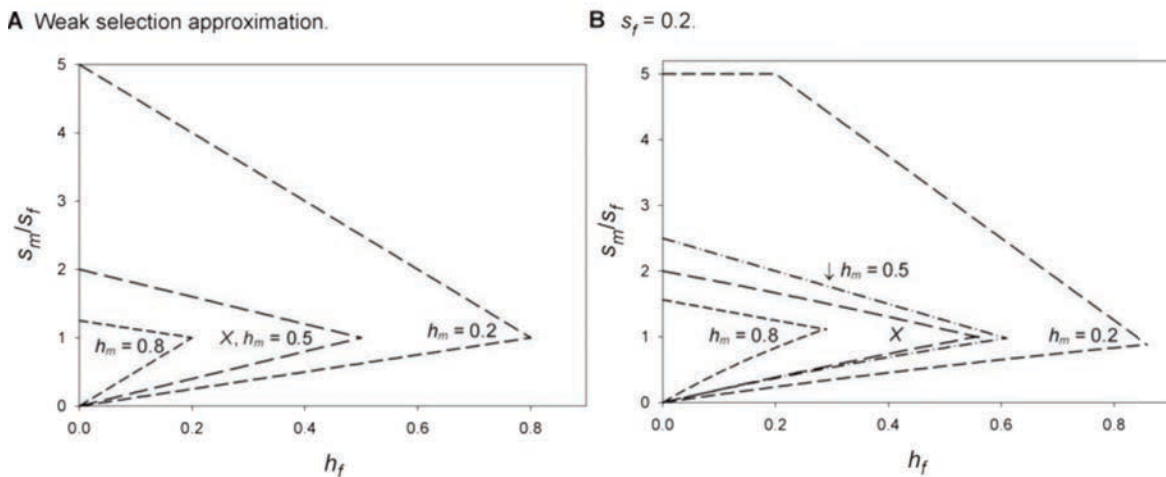


Figure 2.2 – Condition de maintien d'un polymorphisme sexuellement antagoniste à un locus autosomal, sous l'hypothèse que la dominance peut être différente chez les mâles et les femelles. Le polymorphisme est maintenu dans la région entre les lignes identiques (voir légende figure 2.1). Le cas pour le chromosome X est aussi indiqué. **A-** prend en compte des coefficients de sélection faibles de sorte que les termes en s_f des équations 2.10 et 2.11 puissent être négligés **B-** pour $s_f = 0.2$. Tiré de Fry (2010).

Fry (2010) donne un exemple théorique de traits pour lesquels la dominance pourrait être différente chez les deux sexes. Il considère un gène pour lequel les sexes ont un optimum différent au niveau de l'activité de la protéine codée par ce gène, et il suppose que les deux allèles ont un effet additif sur le trait. L'hétérozygote aura un niveau d'activité intermédiaire par rapport aux deux homozygotes. Néanmoins, si l'hétérozygote a une fitness plus élevée que la moyenne entre les fitness basses (par exemple A_1A_1 chez le mâle) et hautes (A_2A_2 chez le mâle), alors les sexes n'auront pas les mêmes taux de dominance pour la fitness (voir figure 2.3 pour un schéma explicatif). Ainsi, bien que les allèles aient un effet additif au niveau de l'activité de la protéine codée par ce gène, la non-additivité des fitness correspondantes sera favorable aux hétérozygotes chez les deux sexes.

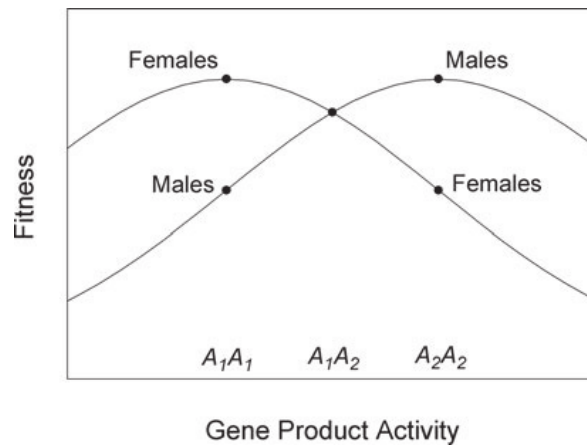


Figure 2.3 – Illustration d'un cas où les allèles sexuellement antagonistes ont un effet additif sur l'activité de la protéine codée par le gène considéré, mais où les fitness ne sont pas additives. Tiré de Fry (2010).

Autres études évaluant la prédiction de Rice

Plusieurs études théoriques vont dans le sens de la prédiction de Rice (Mullon *et al.* 2012 ; Patten et Haig 2009). Vicoso et Charlesworth (2009) discutent néanmoins cette hypothèse en soulevant qu'il est attendu que le chromosome X soit sujet à une dérive génétique plus importante que les autosomes en raison de son effectif efficace plus réduit. En conséquence, la sélection sur le chromosome X pourrait être peu efficace face à la dérive génétique, conduisant à une faible proportion de locus SA sur le chromosome X. Néanmoins, Mullon *et al.* (2012) ont montré que la taille efficace du chromosome X était augmentée par la variance élevée du succès reproducteur chez les mâles, et concluent que le chromosome X devrait être un *hotspot* pour l'accumulation de locus SA.

2.1.3 Examen empirique de la localisation génomique des locus SA

À ce jour, les études empiriques qui ont permis de déterminer la localisation chromosomique (autosomale ou liée à l’X) des variations antagonistes donnant lieu à des fitness sexe-spécifiques ont donné des résultats contradictoires, certaines ont mis en évidence une localisation liée au chromosome X et d’autres liée aux autosomes. La méthode la plus utilisée pour déterminer la localisation des variations génétiques à l’origine d’une sélection sexuellement antagoniste consiste à estimer les corrélations entre la fitness des parents et celle de leurs descendants. Chez les espèces XY, s’il existe une corrélation négative entre la fitness des femelles et de leurs fils, ainsi que des mâles et de leurs filles, et une corrélation positive entre la fitness des femelles et de leurs filles mais pas de corrélation entre la fitness des mâles et de leur fils, cela signifie que les variations sont situées sur le chromosome X. En effet, celui-ci n’est pas transmis de père en fils contrairement aux autosomes.

Études montrant une liaison à l’X de locus SA

Trois études ont été effectuées chez des populations expérimentales de *Drosophila melanogaster*. L’étude de Gibson *et al.* (2002), utilisant des hémiclones, montre que 97% des variations sexuellement antagonistes sont situées sur le chromosome X. Ces observations ont été confirmées sur la même population par Pischedda et Chippindale (2006), qui ont montré que les mères avec une fitness haute produisaient des fils avec une fitness basse tandis que les pères avec une fitness haute produisaient des filles avec une fitness basse, mais que la fitness mâle n’était pas héritée par les fils, signe que les gènes codant pour les traits sexuellement antagonistes étaient localisés sur le chromosome X. Les mêmes observations ont été effectuées sur une population différente de *Drosophila melanogaster* (Connallon et Jakubowski 2009).

Dans une étude d’une population naturelle de cerfs élaphe (*Cervus elaphus*), une corrélation significativement négative entre les mesures de fitness des pères et des filles a été observée, ainsi que l’absence de corrélation entre les fitness des pères et des fils (Foerster *et al.* 2007), indiquant une liaison à l’X des variations génétiques sexuellement antagonistes chez cette population.

Dean *et al.* (2012) ont montré, grâce à une expérience d’évolution expérimentale effectuée chez *Drosophila serrata*, qu’après l’introduction d’un nouvel allèle sexuellement antagoniste, dans ce cas avantageux pour le mâle et désavantageux pour la femelle, un polymorphisme pouvait être maintenu grâce à une liaison à l’X.

Enfin, une étude du transcriptome d’une population expérimentale de *Drosophila*

melanogaster a permis de mettre en évidence que le chromosome X était enrichi en gènes montrant un signal d'IASC en cours (Innocenti et Morrow 2010).

Études montrant une localisation autosomale de locus SA

Plusieurs autres études ont mis en évidence une localisation autosomale de variations sexuellement antagonistes. Chez le criquet *Allonemobius socius* (Fedorka et Mousseau 2004), une corrélation négative entre la fitness des pères et des filles, ainsi qu'entre celle des pères et des fils, a été observée. Des observations équivalentes ont été réalisées chez *Drosophila serrata* (Delcourt *et al.* 2009) et le lézard *Uta stansburiana* (Calsbeek et Sinervo 2003).

Chez *Drosophila melanogaster*, l'utilisation d'hémiclones a permis d'évaluer la contribution de deux autosomes et du chromosome X aux variations sexuellement antagonistes en fitness pour le comportement locomoteur, trait sous IASC, et a montré que les autosomes et le chromosome X contribuent de façon substantielle à ces variations (Long et Rice 2007).

2.1.4 La sélection SA à l'origine de différences de fréquences alléliques entre les sexes

Développement du modèle de Rice

La sélection SA est une forme de sélection balancée. En effet, la sélection balancée peut agir si des environnements contrastés créent des conditions de sélection différentes entre des groupes au sein d'une même population, menant à la sélection d'allèles différents. Dans le cas de la sélection SA, on peut considérer que les hommes et les femmes sont deux groupes, et que les forces de sélection agissant sur ces groupes sont antagonistes. Les signatures génétiques de la sélection balancée incluent une forte diversité dans la région où la sélection agit, ainsi qu'un excès d'allèle à des fréquences intermédiaires et des temps de coalescence plus longs que sous neutralité. Dans le cas d'un locus sous sélection SA présentant un polymorphisme sexuellement antagoniste, on s'attend également à observer des différences de fréquences alléliques entre les sexes au sein d'une population si la sélection SA agit sur la viabilité. Nous nous sommes particulièrement intéressée à cette signature.

Balaresque *et al.* (2004) ont modélisé les différences de fréquences alléliques attendues entre hommes et femmes sur le chromosome X pour un locus sous sélection SA, en utilisant le modèle de Rice. A_1 est désavantageux pour les mâles et avantageux pour les femelles, et A_2 est avantageux pour les mâles et désavantageux pour les femelles. Sur le chromosome

X, avec les notations de Kidwell *et al.* (1977), les équations 2.7 et 2.8 sont modifiées de la façon suivante :

Fréquence de A_1 à l'équilibre chez les mâles :

$$\hat{p}_f = \frac{s_f(2 - h_f(2 - s_m)) - s_m}{2s_f(1 - h_f(2 - s_m))} \quad (2.12)$$

Fréquence de A_1 à l'équilibre chez les femelles :

$$\hat{p}_m = \frac{(1 - s_m)(s_m - s_f(2 - h_f(2 - s_m)))}{s_f(h_f(2 - s_m)^2 - 2(1 - s_m)) - s_m^2} \quad (2.13)$$

La différence en fréquence de A_1 à l'équilibre entre les femmes et les hommes peut donc être exprimée par :

$$\Delta = \hat{p}_f - \hat{p}_m = \frac{s_m(s_m - s_f(2 - h_f(2 - s_m)))(s_m - h_f s_f(2 - s_m))}{2s_f(1 - h_f(2 - s_m))(s_f(h_f(2 - s_m)^2 - 2(1 - s_m)) - s_m^2)} \quad (2.14)$$

Les paramètres pour lesquels un polymorphisme est en équilibre stable sont représentés en figure 2.4.

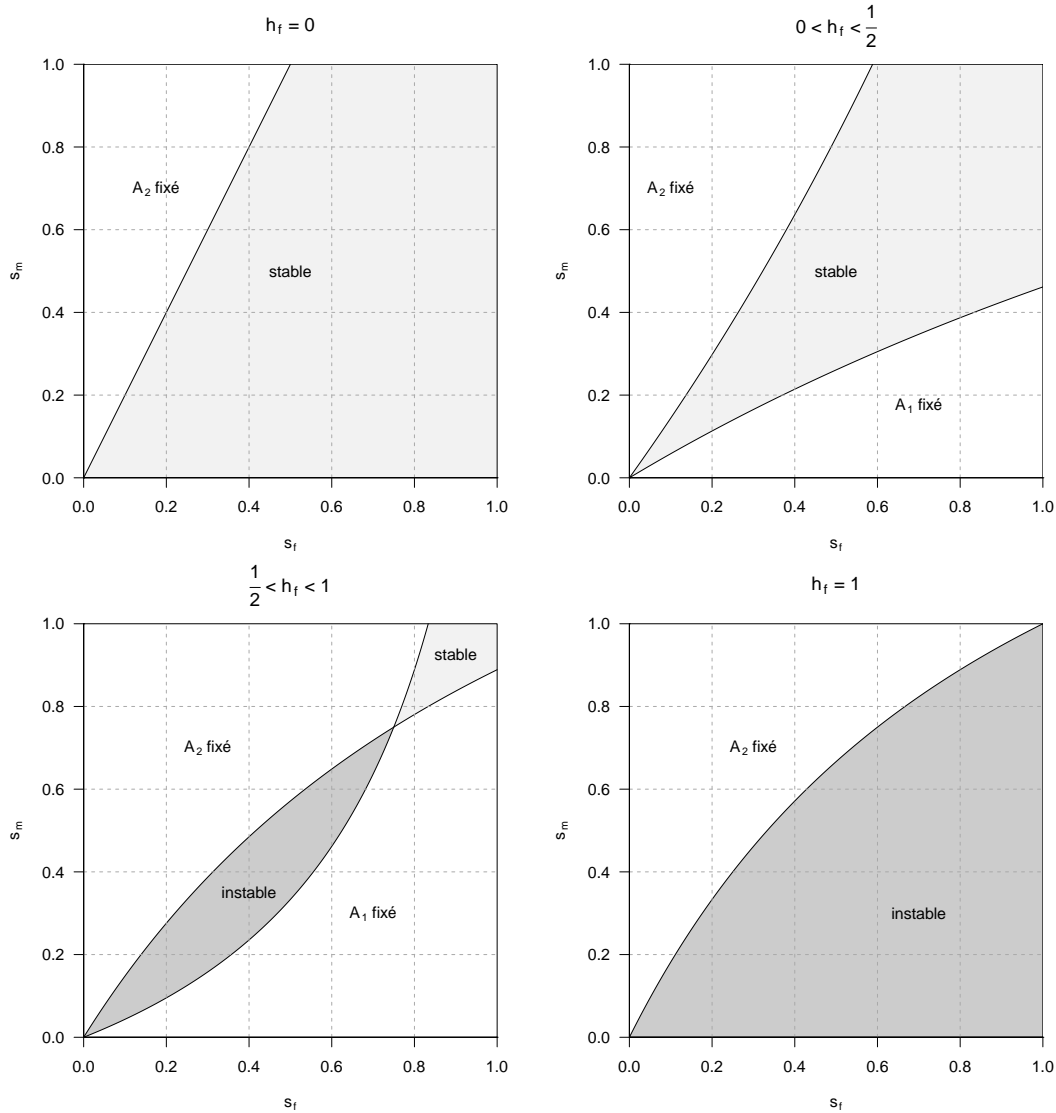


Figure 2.4 – Régions d'équilibre pour un locus lié au chromosome X sous sélection sexuellement antagoniste, A_1 étant avantageux chez les femelles et désavantageux chez les mâles et A_2 étant avantageux chez les mâles et désavantageux chez les femelles. Les régions en blanc correspondent aux équilibres monomorphes (A_1 ou A_2 est fixé), les régions en gris clair correspondent aux polymorphismes en équilibre stable, les régions en gris foncé correspondent aux polymorphismes en équilibre instable. Un équilibre stable est possible quand $(1 - h_f s_f)(2 - s_m) > 2(1 - s_m)$ et $(1 - h_f s_f)(2 - s_f) > 2(1 - s_f)$ **(a)** Quand A_2 est récessif / A_1 est dominant chez les femelles ($h_f = 0$), deux équilibres sont possibles : un polymorphisme stable et une fixation de A_2 . **(b)** Quand A_2 est partiellement récessif / A_1 est partiellement dominant chez les femelles ($0 < h_f < 0.5$), trois équilibres sont possibles : deux fixations, une où A_1 est fixé et l'autre où A_2 est fixé, et un polymorphisme stable. **(c)** Quand A_2 est partiellement dominant / A_1 est partiellement récessif chez les femelles ($0.5 < h_f < 1.0$), une région de polymorphisme en équilibre instable apparaît. **(d)** Quand A_2 est complètement dominant / A_1 est complètement récessif chez les femelles ($h_f = 1$), aucune région de polymorphisme en équilibre stable n'est présente, et on prédit une région de polymorphisme instable, donc A_2 est toujours fixé. Adapté de Balaresque *et al.* (2004).

Δ correspond à la différence entre la fréquence de A_1 chez les femmes et la fréquence de A_1 chez les hommes, et nous indique l'intensité des différences de fréquences alléliques entre les sexes pour le locus considéré. Les valeurs de Δ ont été modélisées en utilisant l'équation 2.14 pour $h_f = 0$, c'est-à-dire le cas le plus favorable au maintien d'un polymorphisme stable, et sont représentées figure 2.5. On remarque que, pour les gammes de paramètres où A_2 n'est pas fixé, les différences de fréquences alléliques entre hommes et femmes (dont l'intensité est modélisée par Δ), augmentent lorsque les taux de sélection s_f et s_m augmentent. Cette figure montre qu'après sélection SA suffisamment forte, un locus présentera des différences de fréquences alléliques entre les hommes et les femmes, bien que leur détection nécessite des taux de sélection élevés.

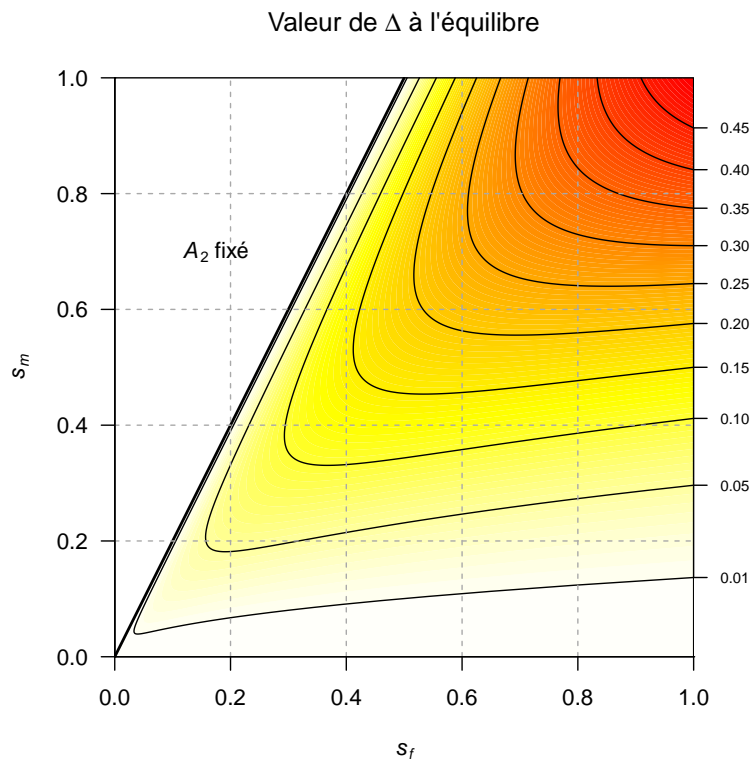


Figure 2.5 – Valeur à l'équilibre de $\Delta = \hat{p}_f - \hat{p}_m$ pour les paramètres de sélection s_m et s_f et dans le cas où A_1 complètement dominant ($h_f = 0$). Les valeurs de Δ sont indiquées sur la droite, et sont comprises entre 0.0 et 0.45. Adapté de Balaesque *et al.* (2004).

Utilisation de données génomiques pour la détection de locus sous sélection SA

Nous avons vu que des différences de fréquences alléliques entre les sexes peuvent émerger à un locus si celui-ci est sous sélection sexuellement antagoniste et si cette sélection implique une survie différentielle entre les mâles et les femelles.

Une étude de la diversité génétique d'un microsatellite lié à l'X dans 10 populations humaines a permis de mettre en évidence une signature d'IASC chez l'Homme pour ce locus, c'est-à-dire une différence significative de fréquences alléliques entre les hommes et les femmes (Balaesque *et al.* 2004). Ce locus est lié génétiquement au gène VCX10r, membre de la famille des VCX, qui s'expriment uniquement dans les testicules.

Nous avons adapté cette méthode afin de détecter des signaux d'IASC dans le cadre d'un balayage du génome, en utilisant des données de génotypage de génomes complets de 11 populations humaines (Altshuler *et al.* 2010). Le but de cette étude était de mettre en évidence une liste de SNP (*Single Nucleotide Polymorphism*) enrichie en locus sous sélection sexuellement antagoniste, afin d'évaluer si le chromosome X présente une accumulation de ce type de locus, et ainsi de tester l'hypothèse de Rice chez l'Homme. Nous avons ensuite étudié les locus se trouvant dans des gènes, afin d'effectuer des analyses d'enrichissement fonctionnel pour déterminer les fonctions biologiques potentiellement sous sélection sexuellement antagoniste chez l'Homme.

2.2 Détection d'une signature de sélection sexuellement antagoniste dans les données de HapMap3.3

Dans cette partie, nous avons effectué un balayage du génome afin de détecter une signature de sélection sexuellement antagoniste, des différences de fréquences alléliques entre les hommes et les femmes, en utilisant un balayage du génome. Dans cette partie, nous allons résumer les principaux résultats de cette analyse, suivi de l'article original.

2.2.1 Matériels et méthodes

Nous avons utilisé la base de données HapMap3.3 (Altshuler *et al.* 2010), qui comporte le génotypage d'approximativement 1.5 millions de SNP répartis sur l'ensemble du génome chez 1397 individus provenant de 11 populations différentes.

Nous avons dans un premier temps effectué des contrôles sur les données, incluant la suppression des individus apparentés, des enfants, et des SNP présentant une fréquence de

l'allèle mineur inférieure à 5%. Nous avons également supprimé les SNP dont les séquences flanquantes ($\pm 30b$) avaient 100% de similarité avec des séquences sur le chromosome Y à l'aide du logiciel BLAST (Altschup *et al.* 1990). Une population a été retirée de notre analyse, LWK (Luhya de Webuye, Kenya) car elle présentait une structure génétique sexe-spécifique, c'est-à-dire que les hommes et les femmes ne représentaient pas un échantillon homogène de la population.

Nous avons calculé les différences de fréquences alléliques entre les sexes SNP par SNP, au sein de chaque population indépendamment. De plus, nous avons utilisé le F_{ST} calculé entre hommes et femmes (F_{ST} intersexuel) comme estimateur des différences de fréquences alléliques entre les sexes, à l'aide de la formule de Weir (1996). La significativité des différences entre les fréquences alléliques des hommes et des femmes a été évaluée avec un test exact de Fisher. Les SNP ayant une p-value significative pour le test exact de Fisher ont été désignés par le terme *Sexually Differentiated* SNP (*SD* SNP).

Le nombre de tests étant très élevé, de nombreux faux positifs ont potentiellement été détectés. Nous avons donc effectué une correction pour tests multiples (FDR, *False Discovery Rate*) (Benjamini et Hochberg 1995) sur les p-values du test exact de Fisher à différents niveaux. Le premier niveau, correspondant à la correction la plus stringente, consistait à corriger la p-value du test exact de Fisher pour le nombre de SNP contenus dans le génome. Le second niveau de correction consistait à corriger pour le nombre de SNP contenu au sein de chaque chromosome.

Nous avons ensuite concentré notre étude sur les SNP géniques, c'est-à-dire les SNP dont les positions se trouvaient dans un intervalle $\pm 5kb$ des positions des gènes. Dans ce but, une troisième correction, moins stringente, a été utilisée sur les SNP géniques uniquement : une correction FDR pour le nombre de SNP par gène. En effet, les gènes les plus longs ont plus de chance de contenir des SNP significatifs, ce qui constitue un biais pour nos analyses. Les SNP ayant une p-value significative après correction pour tests multiples par gène ont été désignés par le terme *genic SD* SNP.

Nous avons ensuite effectué une analyse d'enrichissement fonctionnel sur la liste de gènes comportant au moins un *genic SD* SNP dans leur intervalle $\pm 5kb$, désignés par le terme *SD gene*, pour tous les chromosomes et toutes les populations confondues. Nous avons pour cela utilisé le logiciel DAVID (Huang *et al.* 2009a,b), qui permet de détecter si, au sein de la liste de *SD gene*, certaines fonctions sont plus souvent représentées que dans des listes de gènes construites au hasard.

2.2.2 Résultats

Détection de différences de fréquences alléliques extrêmes entre les hommes et les femmes

A l'aide des deux premières corrections FDR pour le nombre de SNP à l'échelle du génome et pour le nombre de SNP par chromosome, nous avons identifié des différences de fréquences alléliques extrêmes entre les sexes. Nous avons effectué un contrôle qualité de ces résultats en comparant les fréquences alléliques obtenues dans HapMap 3.3 avec celles obtenues dans 1000 Genomes (The 1000 Genome Project Consortium 2010), une base de données de reséquençage des populations de HapMap 3.3. De nombreux SNP ont été supprimés de l'analyse car ils ne présentaient pas les mêmes fréquences alléliques dans la base de données de 1000 Genomes. Nous avons retenu 2 SNP significatifs après correction à l'échelle du génome, tous deux présents sur le chromosome X et significatifs dans la population GIH (Gujarati Indians from Houston). Néanmoins, la population GIH n'étant pas présente dans la base de données de 1000 Genomes, nous n'avons pas pu contrôler pour des erreurs de génotypage. Après correction par chromosome, nous avons retenu 13 SNP, dont les 2 SNP identifiés avec la première correction.

Localisation des polymorphismes SA dans les régions géniques

La distribution des proportions de *genic SD* SNP (nombre de *genic SD* SNP / nombre de SNP géniques) pour les 10 populations a ensuite été comparée entre le chromosome X et chaque autosome à l'aide d'un test de Wilcoxon unilatéral. Le chromosome X présentait une proportion de *genic SD* SNP plus importante que la majorité des autosomes, à l'exception des chromosomes 17 et 19. Cette observation est conservée lorsque l'on supprime les SNP géniques en déséquilibre de liaison (*Linkage Disequilibrium*, LD) pour un $r^2 > 0.1$ et lorsque l'on inclut uniquement les SNP montrant des fréquences alléliques concordantes entre HapMap 3.3 et 1000 Genomes. Cela signifie que ni un excès de SNP en LD sur le chromosome X, ni des erreurs de génotypage ne sont à l'origine de la tendance observée. Nous avons également effectué cette comparaison après avoir enlevé aléatoirement une copie autosomale chez l'homme afin que les autosomes aient la même taille d'échantillonnage que le chromosome X. Les proportions de *genic SD* SNP étaient fortement réduites sur les autosomes, indiquant que le test exact de Fisher est sensible à la taille de l'échantillon et que la puissance statistique est moins élevée sur le chromosome X, renforçant nos résultats.

Nous avons ensuite évalué s'il existait une hétérogénéité entre les 10 populations étudiées, en comparant les proportions de *genic SD* SNP entre les chromosomes X, 10 et

22 pour chacune d’entre elles. Ces deux autosomes ont été choisis car ils ont respectivement le même nombre de gènes et le même nombre de SNP que le chromosome X. Les proportions de *genic* SD SNP étaient significativement plus élevées sur le chromosome X que sur les chromosomes 10 et 20 pour six et quatre populations, respectivement (test exact de Fisher unilatéral). Pour deux populations, une population européenne (CEU, résidents de l’Utah avec des origines d’Europe de l’ouest et du nord) et une population japonaise (JPT, japonais de Tokyo), bien que les différences entre le chromosome X et les autosomes n’étaient pas significatives, la même tendance était observée. Pour les deux populations métisses incluses dans HapMap 3.3, ASW (individus d’origines africaine, résidents au sud-ouest des Etats-Unis) et MEX (individus d’origines mexicaine, résidents à Los Angeles en Californie), les proportions de *genic* SD SNP présentaient une tendance différente : non seulement les proportions de *genic* SD SNP sur le chromosome X n’étaient pas significativement plus élevées que celles du chromosome 10 et 22, mais de plus les proportions pour les trois chromosomes étaient plus faibles que pour toutes les autres populations.

Analyse d’enrichissement fonctionnel des gènes présentant un signal de sélection SA

Dans un second temps, nous avons effectué une analyse d’enrichissement fonctionnel sur la liste des *SD gene*. Les fonctions suivantes étaient significativement enrichies : des fonctions impliquées dans les processus développementaux, le métabolisme du sucre et l’énergie, la régulation de l’expression des gènes, le système immunitaire et la reproduction.

2.3 Discussion

Dans cette étude, nous avons détecté des différences de fréquences alléliques entre les sexes dans une base de données de génotypage en incluant des échantillons de 10 populations. Un indice de différenciation, le F_{ST} intersexuel a été utilisé SNP par SNP comme indicateur de l’intensité des différences de fréquences alléliques entre hommes et femmes, et le test de significativité de cette différence a été effectué sur les comptes d’allèles à l’aide de tests exacts de Fisher.

Des différences de fréquences alléliques peuvent être causées par une sélection SA sur la survie entre la gaméto-genèse et l’âge adulte. Chez l’Homme, l’action de la sélection sur la viabilité au cours du développement est potentiellement élevée puisque moins de 50% des fécondations arrivent à terme (Benagiano *et al.* 2010). Par ailleurs, des différences

de fréquences alléliques ont été observées chez des nouveaux nés humains (Ucisik-Akkaya *et al.* 2010). Cela indique que des processus sexuellement antagonistes pourraient avoir lieu au cours de la gestation.

D'autres mécanismes peuvent mener à l'observation de différences de fréquences alléliques entre les sexes. Premièrement, si les individus mâles et femelles de HapMap 3.3 ont été échantillonnés à des âges différents et qu'une sélection âge-spécifique a lieu dans ces populations, cela mènerait à des différences de fréquences alléliques entre mâles et femelles. Nous n'avons pas accès à l'âge des individus de HapMap 3.3. Néanmoins, cinq des populations incluses dans notre analyse incluent des couples s'étant reproduits. La différence d'âge entre les couples chez l'Homme étant en moyenne de 3,5 ans (Fenner 2005), il est peu probable que cette faible différence d'âge affecte nos résultats. Deuxièmement, des mécanismes sélectifs non sexuellement antagonistes tels qu'une sélection limitée à un sexe ou une sélection dans le même sens chez les deux sexes mais agissant avec une plus forte intensité chez l'un des sexes pourraient mener à des différences de fréquences alléliques entre les hommes et les femmes. Néanmoins, ces mécanismes engendreraient des différences de fréquences alléliques entre les sexes plus difficilement détectables que la sélection SA. En effet, les écarts entre les fréquences alléliques chez les hommes et chez les femmes seraient non seulement plus faibles mais aussi transitoires dans la mesure où il est attendu que l'allèle bénéfique soit rapidement fixé. Il est possible que la liste de SNP montrant des différences de fréquences alléliques entre les sexes que nous avons détectée contienne des signaux dus à ces deux derniers mécanismes, mais cette liste devrait être enrichie pour des SNP sous sélection SA.

Nous avons effectué plusieurs types de corrections pour tests multiples sur la p-value du test exact de Fisher utilisé pour estimer la significativité des différences de fréquences alléliques entre les sexes. Le premier type corrigeait les p-values en prenant en compte le nombre de tests effectués dans le génome, et le second corrigeait les p-values en prenant en compte le nombre de tests effectués par chromosome. Ces deux corrections sont très stringentes et peu de SNP étaient toujours significatifs. Cependant, ce résultat n'est pas surprenant dans le cadre d'une sélection SA dans la mesure où l'on ne s'attend pas à observer des différences de fréquences alléliques très élevées entre les sexes car elles nécessiteraient une force de sélection très importante. Ici, les SNP identifiés montrent des différences de fréquences alléliques très élevées, entre 0,263 et 0,447, qui sont difficiles à expliquer par la sélection SA et par les autres mécanismes sélectifs présentés plus haut. Néanmoins, pour deux SNP situés sur le chromosome X nous n'avons pas pu contrôler pour des erreurs de génotypage en comparant les fréquences alléliques avec celles de la base de données 1000 Genomes car la population dans laquelle ils sont significatifs n'est

pas présente dans 1000 Genomes. Il est donc possible que ces résultats soient causés par des erreurs de génotypage. De plus, dans la région autour de ces deux SNP, aucun autres SNP n'a un F_{ST} élevé. En revanche, nous avons pu contrôler dans 1000 Genomes les fréquences alléliques des autres SNP significatifs après correction pour tests multiples par chromosome. Un d'entre eux est localisé sur le chromosome 2 et les dix autres forment un cluster sur le chromosome 20. Pour ces deux régions, le signal de F_{ST} intersexuel s'étend aux SNP voisins, en LD avec les SNP focaux. Bien que ces résultats soient intéressants, des analyses plus précises de ces SNP dans des bases de données de séquençage seraient nécessaires pour déterminer l'origine des différences de fréquences alléliques entre les sexes observées.

Dans un second temps, nous nous sommes particulièrement intéressés aux signaux géniques. Nous avons effectué une correction pour tests multiples prenant en compte le nombre de tests effectués par gène. Les SNP montrant des différences de fréquences alléliques significatives entre les sexes après cette correction ont été appelés des *genic SD* SNP. Cette correction pour tests multiples étant moins stringente que les précédentes, les SNP identifiés présentent des différences fréquences alléliques intermédiaires.

Nous avons détecté un excès de *genic SD* SNP sur le chromosome X par rapport aux autosomes. Cela pourrait signifier que le chromosome X est une région propice à l'accumulation de locus sous sélection SA, résultat attendu dans le cadre de l'hypothèse de Rice (1984). Afin de vérifier que des variations de LD entre chromosomes et populations ne biaisaient pas nos résultats, nous avons comparé les proportions de *genic SD* SNP après suppression d'un SNP par paire de SNP en LD (pour un $r^2 > 0,1$). Le chromosome X était toujours enrichi en *genic SD* SNP après ce contrôle. Néanmoins, l'observation d'une proportion plus élevée de *genic SD* SNP sur le chromosome X par rapport aux autosomes pourrait également signifier que les différences de fréquences alléliques entre les sexes peuvent être plus facilement détectées sur le chromosome X. En effet, il est possible qu'en raison de l'hémizygotie du mâle, les F_{ST} intersexuels soient plus élevés sur le chromosome X que sur les autosomes pour une même force de sélection. Des simulations effectuées dans le chapitre 3 nous permettront de répondre partiellement à cette question. Cependant, le chromosome X a une taille d'échantillon deux fois plus faible chez les mâles en comparaison aux autosomes, ce qui influence fortement la puissance du test de Fisher utilisé pour détecter les différences de fréquences alléliques. Nous avons en effet montré qu'après haploïdisation des autosomes, c'est-à-dire suppression d'une copie autosomale de chaque SNP chez les mâles, les proportions de *genic SD* SNP étaient drastiquement diminuées sur les autosomes. Ceci indique que la puissance statistique du test exact de Fisher est beaucoup moins élevée sur le chromosome X en comparaison aux autosomes.

S'il existe effectivement un biais dans l'intensité des différences de fréquences alléliques sur le chromosome X, il est possible que celui-ci soit contre-balancé par une perte de puissance statistique pour détecter ces différences.

Nos résultats indiquent que les *genic* SD SNP sont en majorité spécifiques aux populations, ce qui pourrait être expliqué par une sélection récente. Pour évaluer si une sélection récente pouvait être à l'origine du signal détecté, nous avons calculé le F_{ST} intersexuel dans les régions voisines des *genic* SD SNP et nous avons comparé cette distribution avec celle du F_{ST} intersexuel autour de SNP géniques tirés au hasard. Les résultats montrent que les F_{ST} intersexuels sont significativement plus élevés dans les régions voisines des *genic* SD SNP en comparaison à l'attendu sous neutralité dans une région d'environ 120 kb, ce qui est compatible avec une sélection récente.

Malgré ce manque de concordance entre les populations, un enrichissement en *genic* SD SNP sur le chromosome X en comparaison aux chromosomes 10 et 22 a été observé chez la majorité des populations, à l'exception des deux populations métisses. Celles-ci montrent une tendance différente et des proportions de *genic* SD SNP moins élevées que les autres populations sur les 3 chromosomes considérés. Cela pourrait être expliqué par le processus de métissage ayant eu lieu dans ces populations. En effet, on s'attend à ce que, si le signal est en majorité spécifique aux populations, celui-ci soit perturbé lors du métissage de deux populations auparavant isolées. Cela pourrait aussi être dû à la petite taille de ces échantillons, qui comportent seulement 41 et 53 individus (chez ASW et MEX, respectivement) tandis que les autres échantillons comportent au moins le double d'individus.

Afin de vérifier que l'enrichissement en *genic* SD SNP sur le chromosome X est bien dû à un processus sexe-spécifique, c'est-à-dire qu'il dépend de la comparaison des fréquences alléliques entre les sexes, nous avons construit au hasard des groupes d'individus de sexe-ratio variables, du plus équilibré au moins équilibré. Les résultats indiquent que l'intensité de la différence entre les proportions de *genic* SD SNP sur le chromosome X et celles sur les autosomes est fortement corrélée au sexe-ratio des groupes. Cette différence est en effet plus élevée quand un groupe est enrichi en femelles et l'autre en mâles.

Un excès d'erreurs de génotypage sur le chromosome X pourrait également biaiser nos résultats. Pour tester ce biais, nous avons utilisé la base de données 1000 Genomes, qui consiste en un re-séquençage des populations de HapMap 3.3. En considérant uniquement les SNP montrant des fréquences alléliques concordantes entre les deux bases de données, le chromosome X montrait toujours une proportion de *genic* SD SNP plus élevée que celles des autosomes.

En prenant en compte l'ensemble des *genic* SD SNP, les moyennes des F_{ST} intersexuels étaient de 0.099 sur le chromosome X et de 0.065 sur les autosomes, sans les populations

admixtes. Dans le cadre d'un scénario n'impliquant que l'action d'une sélection SA dans une population à l'équilibre, ces F_{ST} intersexuels sont hauts, et nécessiteraient des forces de sélection très élevées. Néanmoins, notre étude étant basée sur un balayage du génome, ce phénomène peut être expliqué par un effet de *winner's curse*. Le *winner's curse* consiste à identifier des SNP montrant des signaux extrêmes par effet d'échantillonnage. Nous avons montré que la puissance du test exact de Fisher est fortement dépendante de la taille de l'échantillon. Si l'échantillon est petit, les différences de fréquences alléliques entre les sexes doivent être élevées pour être détectées. De plus, avec les corrections pour tests multiples que nous avons effectuées, nous capturons les signaux les plus extrêmes. Les F_{ST} intersexuels observés sont donc probablement des estimations artificiellement hautes des valeurs de F_{ST} intersexuels des populations en globalité. L'estimation des coefficients de sélection nécessaires pour obtenir ces différences de fréquences alléliques entre les sexes est dès lors difficile, car elle ne reflèterait pas la sélection en cours dans la population. Des simulations effectuées en chapitre 3 nous permettront de discuter des taux de sélection nécessaires à l'observation de différences de fréquences alléliques de cet ordre pour des échantillons d'une population.

L'effet de *winner's curse* est aggravé chez les populations métisses, ASW et MEX, qui sont les deux plus petits échantillons de HapMap 3.3. En effet, ces populations, bien que montrant des proportions de *genic* SD SNP plus faibles que les autres populations, montraient des F_{ST} intersexuels moyens beaucoup plus élevés. Pour ASW, le F_{ST} intersexuel des *genic* SD SNP sur le chromosome X était de 0.224 et sur les autosomes de 0.185 en moyenne. Pour MEX, le F_{ST} intersexuel des *genic* SD SNP sur le chromosome X était de 0.180 et sur les autosomes de 0.129 en moyenne. Les échantillons étant plus petits, le signal doit être plus important afin de le détecter.

Après suppression des SNP géniques en LD, nous avons trouvé en moyenne 118,5 *genic* SD SNP par population. Notre hypothèse étant que ces SNP présentent des différences de fréquences alléliques entre les sexes par le biais d'une sélection sexuellement antagoniste agissant sur la survie des embryons, cela impliquerait un lourd fardeau génétique. Cependant, il convient de noter que cette liste correspond à une liste enrichie pour des SNP sous sélection SA et non à une liste exclusive. De plus, dans le contexte de cette hypothèse, il est plus probable qu'il existe plusieurs combinaisons d'allèles permettant à un embryon de survivre plutôt qu'une seule, menant ainsi à des différences de fréquences alléliques détectables mais pas extrêmes, comme observées pour les *genic* SD SNP.

Nous avons ensuite effectué une analyse d'enrichissement de fonctions sur la liste de gènes comportant au moins un *genic* SD SNP. La première catégorie de fonction enrichie

concerne les processus développementaux, ce qui est étonnant car on s'attend à ce que les IASC agissent plutôt à l'âge adulte (Cox et Calsbeek 2009 ; Chippindale *et al.* 2001). Néanmoins, certains dimorphismes sexuels pourraient nécessiter de découpler les voies de développement entre les sexes, comme pour les dimorphismes de taille (Badyaev 2002). De plus, des différences entre les sexes au niveau du développement ont été observées chez de nombreux mammifères, notamment chez l'Homme (voir Aiken et Ozanne 2013 pour une revue). Nous avons également trouvé de nombreux termes référant à la régulation de l'expression génique. Or, on sait qu'une expression différente entre les sexes des gènes sous IASC est une façon de résoudre les IASC (voir section 1.2.2). Des différences de fréquences alléliques entre les sexes pour des gènes régulant l'expression d'autres gènes pourraient être un intermédiaire à l'évolution d'un mécanisme d'expression sexe-spécifique plus stable. Enfin, les catégories fonctionnelles incluant la reproduction, le système immunitaire et le métabolisme réfèrent à des traits qui sont sexuellement dimorphiques chez l'Homme (Nookaew *et al.* 2013 ; Shi et Clegg 2009 ; Klein 2012 ; Marriott et Huet-Hudson 2006). Les résultats de l'étude d'enrichissement fonctionnel que nous avons effectuée sont en accord avec l'attendu : lorsque l'on étudie des traits sous IASC, on devrait trouver des traits sexuellement dimorphiques (Bonduriansky et Chenoweth 2009).

Dans ce chapitre, nous avons détecté des différences de fréquences alléliques entre les sexes dans une base de données de génotypage. Bien que nos résultats suggèrent que la liste de SNP identifiés soit enrichie pour des SNP sous sélection SA, le mécanisme à l'origine de ces différences restent néanmoins inconnu. Dans le chapitre suivant, nous allons tester une de nos hypothèses pour expliquer l'émergence de différences de fréquences alléliques : un phénomène de distorsion de transmission sexe-spécifique durant le développement.

Insights into Sexually Antagonistic Selection in the Human Genome

Authors: E.A. Lucotte¹, R. Laurent^{1,2}, E. Heyer¹, L. Ségurel^{1,#}, B. Toupance^{1,#}

[#] Co-supervised the work

Affiliations:

¹Eco-anthropologie et Ethnobiologie, UMR 7206 CNRS - MNHN - Univ Paris Diderot, Sorbonne Paris Cité, Paris, France

²Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, 04103, Germany

Keywords: Sexually antagonistic selection, intralocus sexual conflict, sexual dimorphism, genome scan, X chromosome

Corresponding author:

Elise A. Lucotte: Musée de l'Homme, Département HNS, 17 place du Trocadéro, 75016 Paris.
Email: elucotte@mnhn.fr

INTRODUCTION

Sexually antagonistic (SA) selection can occur in species with two separate sexes if males and females have different fitness optima for a trait. Furthermore, if this trait is encoded by the same set of genes in the two sexes and if each genotype induces the same phenotype in males and females (i.e., a positive intersexual genetic correlation exists for this trait), a single genome cannot encode both sex-specific optima, leading to an intralocus sexual conflict (IASC) [1]. For instance, in humans, adult height is suspected to be a trait with ongoing IASC [2].

To resolve such conflicts, the sex-specific optima must be reached in the two sexes through the establishment of sex-specific gene expression. Hence, the resolution of IASCs leads to the evolution of stable sexual dimorphisms [3]. Such a process can take a long time to achieve [4] because of the complex genetic architecture of some traits. Indeed, pleiotropic effects of genes encoding such traits and the fluctuation of selective pressures across time and environment are expected to impede the resolution of IASCs [1].

Consequently, during the initial phase of the evolution of sexual dimorphism, genetic variability is expressed in the two sexes [5] and the difference in fitness between males and females induces sex-differences in allelic frequencies. Theoretical studies [5,6] have shown that, under some assumptions, a sex-specific polymorphism at a locus undergoing unresolved IASC could persist and reach equilibrium in the genome. Such loci, named SA polymorphisms, are therefore under balancing selection.

A classic model comparing sex-linked to autosomal loci, proposed by Rice [5], predicts that the X chromosome is especially prone to the accumulation of loci under IASC because of male hemizyosity. Indeed, because of the asymmetry of transmission and expression of the X chromosome between the sexes, partially recessive male-advantageous alleles and partially dominant female-advantageous alleles can increase in frequency under more lenient conditions if X-linked than if autosomal. While this theoretical prediction has been extensively discussed over the last 30 years, the studies investigating this hypothesis have provided conflicting results [7]: some in favor of an X-linked location of loci under IASC, in accordance with Rice's model prediction [8–10], and others providing evidence for an autosomal location of such loci [11–13]. Importantly, no studies attempted to map loci under SA selection at the genome-wide level in natural populations.

Using HapMap III, a dataset of approximately 1.5 million SNPs genotyped in over 1000 individuals from 11 worldwide human populations [14], we developed a framework to detect at the genome-wide scale a signature of SA polymorphisms, namely significant differences in allelic frequencies between males and females from the same population, as previously highlighted by Balaesque et al. for one locus [15]. Our genome-wide scan aimed to address two major questions: **i)** Is the X chromosome a hotspot for the accumulation of loci under SA selection? **ii)** In which biological functions are SA loci involved?

MATERIAL AND METHODS

Polymorphism dataset in humans

The HapMap III.3 dataset consists of approximately 1.5 million genome-wide SNPs [14] typed in 1397 individuals sampled in 11 worldwide human populations.

We only included unrelated individuals from the HAP1161 set proposed by Pemberton et al. (2010) in which a member of every first degree relative pair was excluded. The number of males and females after quality controls were overall similar in each population (Table S1). SNPs with a minor allele frequency lower than 5% were removed. Because we focused our analysis on differences between the sexes, any SNP mapping on the Y chromosome would be critical, as it would skew allelic frequency in a sex-specific manner. Therefore, all X chromosome SNPs located in the pseudo-autosomal regions (PARs), which are homologous to regions on the Y chromosome, were removed. Likewise, the X-linked SNPs outside of the PARs that were reported as heterozygous in males were excluded. To control whether other SNPs, including autosomal SNPs, could map on the Y chromosome and bias our results, we performed a systematic sequence similarity search using BLAST [17] of their flanking sequences (± 30 bp around the SNP) against the Y chromosome sequence (using the hg18 human genome assembly). Any SNP whose flanking sequences had a perfect match with Y-linked sequences was removed (36 SNPs, Table S2). After quality control, the mean number of SNPs per population was 1,122,087 SNPs.

We focused most of our analyses on genic SNPs. The positions of all genes known to date were downloaded from the UCSC database (build NCBI36/hg18, [18]), and genic SNPs were defined as being within ± 5 kb of a gene's coordinates. The mean number of genic SNPs per population was 556,558 SNPs.

Detection of Sexually Differentiated SNPs

Within each population, we considered males and females as two distinct samples. A Fisher's exact test was performed for each SNP to assess if the differences in allelic frequencies between males and females were significant. SNPs with a significant Fisher's exact test p-value were named Sexually Differentiated (SD) SNPs.

To correct the Fisher's test p-values for multiple testing, we performed a false discovery rate (FDR) correction [19] in each population at three different levels independently. First, we focused our study on genic SNPs and corrected the genic p-values for the number of tests performed (i.e., of SNPs) per gene to correct for gene length bias (Figure S1). SD SNPs with a significant q-value after per-gene FDR correction were named genic SD SNPs. Second, we corrected for the number of tests performed at the genome-wide level, and finally for the number of tests performed per chromosome. The SD SNPs with a significant q-value after genome-wide and chromosomal FDR correction were named genome-wide SD SNPs and chromosomal SD SNPs, respectively.

Potential factors biasing the X versus autosomes comparison

To investigate the effect of potential confounding factors on the difference in significance between the X chromosome and the autosomes, we applied different filters to our dataset and verified we obtained similar results on these reduced datasets.

Simulated haploidization of autosomes in males

To assess the effect of smaller sample sizes in males for X-linked loci compared to autosomal loci on Fisher's exact tests, we simulated a haploidization of males' autosomes by randomly removing one autosomal copy in males.

Correction for linkage disequilibrium (LD)

To remove any bias due to differences in LD between chromosomes, we LD-pruned the genic SNPs in each population using the "--indep-pairwise" option in PLINK [20] with an r^2 threshold of 0.1 within a window of 50 SNPs and a step of 10 SNPs. The r^2 chosen is very stringent so that the retained SNPs are at most slightly linked.

Concordance between the 1000 Genomes project and the HapMap dataset

We compared the allelic frequencies of the SD SNPs (genic, chromosomal and genome-wide) in HapMap III.3 [14], a genotyping dataset, with those obtained in the 1000 Genomes [21], a re-sequencing dataset, i.e., obtained with a different molecular technology. For this

comparison, we used the 1000 Genomes Phase 1 SNP calls dataset, which contains the genotype of 1092 individuals for 14 populations, including 8 populations shared with the HapMap dataset. This allowed us to detect genotypes with contrasted patterns of frequency between the two molecular technologies and which therefore are likely to present technical issues. To assess significance, two Fisher's exact tests were used between both allelic distributions, in males and females separately. The SNPs for which either one of the Fisher's exact tests was significant (i.e., $p\text{-val} < 0.05$) were considered as discordant between the datasets. When possible, if the population was absent from the 1000 Genomes project, we compared the allelic frequencies with close populations from the same geographical area.

Functional enrichment analysis

A gene was defined as Sexually Differentiated (SD) if it contained at least one genic SD SNP in one population within ± 5 kb of its coordinates. Each SD gene was associated with the lowest p-value of its genic SD SNPs. For this analysis, we did not use the list of LD-pruned genic SD SNPs because the SNPs were randomly pruned, which could result in removing SNPs with functional significance. We did not restrict our analysis to the list of SNPs with concordant allelic frequencies with the 1000 Genomes dataset because all the populations studied were not represented. We performed a genome-wide functional enrichment analysis on the SD gene list using the DAVID functional annotation chart tool [22]. A subset of the Gene Ontology (GO) database, the GO FAT set, was used to calculate the functional enrichment, allowing a better readability of the results by keeping only specific terms. The functional enrichment analysis was performed on four different gene lists: the complete SD gene list (SDG, 5705 genes), the 3000 SD genes with the lowest p-values (SDG₃₀₀₀), the 1000 SD genes with the lowest p-values (SDG₁₀₀₀), and the SD genes shared by at least two populations (SDG_{sh}, 717 genes, Table S3). A functional enrichment analysis was performed for the SDG, SDG₃₀₀₀ and SDG₁₀₀₀ lists using the "Functional Annotation Chart" tool available in DAVID. The redundant terms with significant enrichment p-values were then manually curated into the following functional categories: developmental process, sugar metabolism and energy, response to stimulus, reproduction, immune system, epidermis development, regulation of gene expression and muscle.

We used the "Functional Annotation Clustering" tool to rank the biological significance of groups of genes and to assign them an enrichment score (a cluster being usually considered as

significantly enriched if its score is 1.3 or more). As this analysis is only possible for input gene lists with less than 3000 genes, we only analyzed the SDG₃₀₀₀, SDG₁₀₀₀ and SDG_{sh} lists with this tool.

RESULTS

Our study aims at detecting differences in allelic frequencies between males and females, a signature of SA selection. For each locus in each population, we assessed significance of such differences by using Fisher's exact tests and obtained a set of Sexually Differentiated (SD) SNPs. To correct the Fisher's exact tests p-values for the large number of tests performed, we performed FDR corrections.

However, sex-specific population structure could also potentially result in differences in allelic frequencies between males and females. To test whether sex-specific demographic events influence our results, we verified if such structures exist in each of the 11 populations from HapMap. We performed a multidimensional scaling (MDS) analysis, based on pairwise allele sharing distances among individuals, computed either for the autosomes or the X chromosome and tested if the genetic distance between males and females was significant (see Text S1, Figure S2-S3, Table S4-S5). Following this analysis, we excluded the LWK population (Luhya from Webuye, Kenya) because males and females were significantly structured when considering the autosomes (Table S4), and therefore did not represent a random sample of the population. We retained the other ten HapMap populations.

SA polymorphisms in genic regions

To evaluate the functional significance of SD SNPs, we first focused our analysis on genic SNPs. Because long genes are more likely to contain an excess of significant SNPs by chance compared to shorter genes, we corrected the Fisher's exact tests p-values of genic SNPs for the number of SNPs per gene with a FDR correction. The SNPs with a significant q-value were defined as genic SD SNPs. We found a mean of 2,441 genic SD SNPs per population.

The distributions of the proportion of genic SD SNPs per chromosome over the 10 HapMap populations included in the study are shown in Figure 1A. The X chromosome presents a significantly higher proportion of genic SD SNPs than most autosomes (using one-sided Wilcoxon–Mann–Whitney test, at a significance level of 5%), except for chromosomes

17, 19 and 22.

Effect of potentially confounding factors

Next, we assessed the effect of various potentially confounding factors on the proportions of genic SD SNPs.

Because Fisher's exact tests, used here to assess the significance of differentiation between the sexes, are sensitive to sample sizes [23], and because autosomes have a sample size twice as large as that of the X chromosome in males, our test on the X chromosome might lack statistical power. To evaluate this effect, we simulated a haploidization of autosomes in males by randomly removing one of the two autosomal copies for each SNP. As expected, the distributions of proportions of genic SD SNPs on autosomes were shifted towards lower values after haploidization (Figure S4), and the signal of enrichment in genic SD SNP of the X-chromosome was now highly significant with respect to all autosomes (one-sided Wilcoxon–Mann–Whitney tests with $p < 10^{-4}$).

Linkage Disequilibrium (LD) patterns can vary greatly across chromosomal regions, and the X chromosome is known to carry SNPs in higher LD than other chromosomes [24]. If not corrected, this bias could lead to a larger number of SNPs showing a signal of SA selection on the X chromosome. To correct for potential LD effects, we LD-pruned the genic SNPs (i.e., randomly removed one SNP from each pair of SNPs showing an $r^2 > 0.1$) and calculated the new proportions of genic SD SNPs over the 10 populations. After LD-pruning, the number of loci identified as potential SA polymorphisms were strongly reduced, with a mean of 118.5 genic SD SNPs per population. For all autosomes but one, chromosome 19, the mean proportions of genic SD SNPs were significantly lower than that of the X chromosome (one-sided Wilcoxon–Mann–Whitney test at a significance level of 5%, Figure 1B). Therefore, differences in LD between chromosomes do not lead to a bias of significant SNPs in favor of the X chromosome. Chromosome 19 tends to show higher proportions of genic SD SNPs compared to the other autosomes. This chromosome is known to have a high gene density compared to the other chromosomes [29]; which could explain the pattern observed.

To evaluate whether genotyping errors in the HapMap dataset could lead to an excess of genic SD SNPs on the X chromosome, notably because of genotyping quality issues owing to male hemizyosity, we also performed the analysis exclusively on SNPs that had concordant allelic frequencies between the genotyping dataset used here (HapMap III) and a re-

sequencing dataset, the 1000 Genomes project [21]. Four autosomes that were previously significant when comparing the proportions of their genic SD SNPs with the X chromosome did not reach significance in this analysis (Figure S5). However, the pattern observed in this reduced dataset was consistent with our initial analysis and the differences might largely be due to a lack of statistical power in the reduced dataset.

Heterogeneity among populations

To verify whether patterns of enrichment on the X chromosome are concordant among populations, we then compared the proportions of genic SD SNPs between the X chromosome and two autosomes for each of the 10 populations (Figure 2). We used the set of LD-pruned genic SD SNPs because LD patterns vary among populations. We chose chromosomes 10 and 22 for this comparison because the former contains approximately the same number of genes and the latter the same number of SNPs after LD-pruning as the X chromosome. The proportions of genic SD SNPs were significantly greater for the X chromosome compared to chromosome 10 and 22 in 6 populations and 4 populations, respectively (one-sided Fisher's exact test at a significance level of 5%). Although the differences were not significant in the CEU (Utah residents with Northern and Western Europe ancestry) and JPT (Japanese from Tokyo) populations, they showed the same tendency. Thus, the pattern observed is consistent with the previous results in most populations. In two populations, namely ASW (African ancestry in Southwest USA) and MEX (Mexican ancestry in Los Angeles, California), the proportions of genic SD SNPs on the X chromosome did not show the same tendency: they were not significantly different from that of both autosomes and the proportions appeared to be overall lower. These two populations are both known to be recently admixed [16] and have approximately half the number of individuals as those in other HapMap populations (Table S1), two characteristics that could influence their proportions of genic SD SNPs (see Text S2, Figure S6A-B). In the case of ASW, though, this does not seem to fully explain the observed pattern (Text S2). When comparing the mean proportions of genic SD SNPs between the X chromosome and the autosomes for all populations but the two admixed ones, the proportion on the X chromosome became significantly higher than that of all autosomes, including chromosome 19 (Figure S7). This is mainly due to the removal of the ASW population, which harbored no genic SD SNPs after LD-pruning.

Only a few genic SD SNPs are shared between populations (Table S3), indicating that the

detected signal is mostly population-specific and suggesting that selection is recent, consistent with unresolved IASC. If so, we expect the signal around candidate SNPs to extend in the neighboring region. To test whether this is the case, we computed the F_{ST} (a standardized measure of frequency differences between males and females) for every SNPs neighboring each LD-pruned genic SD SNPs in a window of ± 300 kb (Text S3, Figure S8). The F_{ST} signal spreads in regions neighboring genic SD SNPs, and is significantly higher than for random SNPs regions in a window of approximately 100 kb, which is consistent with hitchhiking of neighboring loci due to recent selection.

Pattern of genic SD SNPs proportions for random groups of individuals

To further ascertain that the higher proportion of genic SD SNPs observed on the X chromosome is a consequence of IASC rather than, for example, inherent non-sex-specific demographic or genomic differences between the X chromosome and the autosomes, we asked whether we still find the same pattern of signal enrichment on the X when performing our analyses on random groups of individuals in each population (Text S4, Figure S10-11). The X chromosome's proportions were significantly higher than that of the autosomes even when the two groups are homogenous with respect to sex-ratio (see Text S4 for further details). However, we found the difference of signal between the X chromosome and the autosomes to be strongly correlated with the sex-ratio disequilibrium between the groups, i.e., the proportion of significant allelic frequency differences between two groups of individuals increases on the X chromosome relative to autosomes as the groups are more and more skewed for their sex-ratio.. The results supports that the pattern observed on Figure 1 is mostly due to differences in allelic frequencies between males and females, a signature of IASC.

Strong signals of SA polymorphisms at the chromosomal and genome-wide levels

We were also interested in characterizing the strongest signals in each population, including all SNPs, whether genic or non-genic. We performed a FDR correction both at the genome-wide level (correcting for the total number of SNPs included in the dataset), and at the chromosomal level (correcting for the number of SNPs per chromosome). The latter analysis aimed to highlight and contrast the SNPs showing strong signals of SA polymorphisms between chromosomes. We found a total of 10 significant SNPs with this

genome-wide correction, named genome-wide SD SNPs, and 23 significant SNPs with the chromosomal correction, named chromosomal SD SNPs (Table S6). To prevent potential biases due to genotyping errors, we compared the allelic frequencies of these SNPs between the HapMap and 1000 Genomes datasets and retained only concordant SNPs (Table S6). Many SNPs (8 out of 10 genome-wide SD SNPs and 10 out of 23 chromosomal SD SNPs) were excluded using this criterion, highlighting technical errors associated with one dataset or the other.

We retained a set of 13 chromosomal SD SNPs, including two that were also significant at the genome-wide level, which showed strong signals of allelic frequency differentiation between the sexes. Two X-linked SNPs, both found in the GIH population (Gujarati Indians in Houston, Texas), were significant after both genome-wide and chromosomal FDR corrections (Table S6). Even though this signal on the X chromosome is potentially functionally relevant- because both SNPs map in a gene which causes different neuronal pathologies between the sexes when defective (see Text S5) - the GIH population is not present in the 1000 Genomes dataset, so we were unable to control for genotyping errors and we cannot rule out that the extreme difference in allelic frequencies observed for these SNPs is due to technical issues.

Are loci under IASC involved in specific functions?

To assess which functions were undergoing IASC in humans, we performed a functional enrichment analysis using the DAVID annotation chart tool [22]. We used the genome-wide list of Sexually Differentiated Genes (SDG list, 5706 genes) defined as the list of genes containing at least one genic SD SNP in their vicinity (± 5 kb).

The analysis was performed at three nested levels with increasing stringency: the complete SDG list, the 3000 SD genes with the lowest p-values (SDG₃₀₀₀) and the 1000 SD genes with the lowest p-values (SDG₁₀₀₀). Our approach was to manually classify redundant functions enriched in the SDG, the SDG₃₀₀₀ and the SDG₁₀₀₀ lists (Table S7). Eight categories were created: functions involved in developmental processes, sugar metabolism and energy, response to stimulus, regulation of gene expression, immune system, epidermis development, reproduction and muscle. The genes included in “epidermis development” were however found to be located in two clusters of genes that are involved in the same functions. These genes might display the same signal for SA selection as an artifact of linkage disequilibrium

and bias the functional enrichment; we therefore did not further consider this category. Additionally, we note that most of the genes included in the category “immune system” can also be found in the category “response to stimulus.”

The percentage of enriched GO terms in each category was compared between the three tested sets of genes (Figure 3). For some categories (immune system and reproduction), the percentage of enriched GO terms tended to increase as the stringency in the gene list increases. This trend suggests that these functional categories are under stronger selection, and is expected for categories composed of genes with strong effects. However, a category for which the percentage of term decreases may encompass a large number of genes with small effects, harboring a signal not strong enough to be included in the more stringent lists of genes. This could explain why we did not find terms referring to height, a trait suspected to be under IASC in humans but with a highly polygenic determination [2]. We therefore equally considered the categories that were significant in all or part of the three lists.

We also used a systematic clustering method proposed by DAVID on the SDG_{3000} , SDG_{1000} and SDG_{sh} (SD genes shared by at least two populations) lists (Figure 3, Table S8). Two categories, “response to stimulus” and “immune system”, were found to be enriched in the SDG_{3000} and SDG_{1000} lists. Moreover, the category “glycolysis”, referring to “sugar metabolism and energy”, is found enriched in the SDG_{3000} list, as well as the category “reproduction” in the SDG_{1000} list. Additionally, one of the two clusters highlighted in SDG_{sh} referred to terms present in the manually curated category “developmental process.”

DISCUSSION

In this study, we detected differences in allelic frequencies between the sexes, which may reflect sexually antagonistic selection that occurs between gametogenesis and adulthood, when the individuals were sampled and genotyped. Such sexually-differentiated loci may indeed exist in adults, as severe selection is suspected to occur in humans during early development. The probability of fetus survival from fertilization to term could indeed be lower than 50% [25]. We propose that the polymorphisms that we highlight here are mainly due to SA selection occurring during early development, resulting in a sex-specific transmission distortion. Further support of this hypothesis comes from a study in which significant differences in allelic frequencies between the sexes were observed in newborn humans [26].

Alternative selective mechanisms can lead to differences in allelic frequencies between the sexes: sex-specific (but not antagonistic) selection, where a given locus will have a selective effect in one sex but will be neutral in the other, or cases where the selective pressures act in the same direction in both sexes but not with the same intensity. However, these cases will likely lead to smaller (and less detectable) differences in allelic frequencies between the sexes than sexually antagonistic selection. Furthermore, polymorphisms induced by these two types of selection are highly transient and, as a result, should be rarely observed. Thus, it is likely that most of the signal of differentiation between the sexes identified by our method reflects more stable polymorphisms, as those expected under SA selection.

Additionally, when the phenotypic effect of an allele differs between the sexes but a same phenotype is favored by selection in the two sexes, SA selection will arise on the genotypes, leading to differences in allelic frequencies between the sexes. However, such SA loci will not undergo IASC because there is no positive intersexual genetic correlation for the trait considered. As differences in allelic frequencies between the sexes are the signals we aimed at detecting, it is possible that the list of SA loci we identified contains SNPs undergoing this kind of SA selection.

It is also important to note that a difference in age between the males and females sampled could lead to differences in allelic frequencies between the sexes, reflecting selective pressures occurring at specific ages but in both sexes. It is unfortunately impossible to ascertain if such bias exists in our data, as the age of the participants is unknown. However, to our knowledge, such bias in the HapMap dataset has not been reported. Moreover, 5 out of the 10 populations incorporated in the study include couples, and the difference in age within couples is known to be, on average, only 3.5 years [27], which would only slightly affect our result. Importantly, such loci are not expected to show enrichment on the X chromosome. It would however be interesting to explore this hypothesis with an adequate dataset.

When considering genic SD SNPs to test Rice's hypothesis, which predicts that the X chromosome is a hotspot for the accumulation of loci under SA selection, we found that the X chromosome exhibits a significantly higher mean proportion of genic SD SNPs than most autosomes. The F_{ST} signal observed is not restricted to the genic SD SNPs identified, but spreads in the neighboring regions, suggesting that recent selection has played an important role in the signal we detect. Moreover, we showed that this pattern is not influenced by LD or

population effects and specifically occurs when males and females are compared. The presence of a difference in statistical power due to differences in sample sizes and in favor of the autosomes further supports our conclusions.

Although we found that the X chromosome tended to be a hotspot for the accumulation of SA polymorphism, it is not the only chromosome that yields genic SD SNPs. This indicates that sexually antagonistic polymorphisms could also be maintained on the autosomes. We may speculate that the complete loss of loci under IASC from the autosomes could be partially prevented by complex pleiotropic mechanisms of some autosomal genes [28].

The strongest signals for SA polymorphisms were found on chromosomes X, 2 and 20. The 11 autosomal loci identified showed concordant patterns of allele frequencies in the 1000 Genomes and HapMap datasets, which indicate that the detected signals were not due to a technical issue. However, the maintenance of allelic frequency differences between the sexes as high as observed for these SNPs (with differences in frequency between 0.263 and 0.447, see Table S6) implies very strong selection coefficients. Although severe selection occurs during development, it is still difficult to fully explain the extreme patterns observed. The high differences observed here could result from a winner's curse effect, which could lead to the detection of artificially high values of differences in allelic frequencies due to the method of detection and the effect of sampling. Although this effect precludes any attempt to confidently estimate the underlying selection coefficients, our results suggest that we can detect strong differences in allelic frequencies between the sexes in the human genome. Further analyses and modeling would be needed to untangle the underlying process at the origin of these patterns.

The last aim of our analysis was to perform a functional enrichment of the genome-wide SD genes list. The category "regulation of gene expression", highly enriched in the manual clustering method, is of particular interest given that the resolution of IASC requires the establishment of sex-specific gene expression. Indeed, the majority of sexually dimorphic traits result from the differential expression of genes present in both sexes [30], achieved for example through the action of a modifier gene that regulates the expression of the effector gene in a sex-specific manner [5]. A sex-specific polymorphism at a modifier locus could be an intermediate state to the evolution of a more stable sex-specific gene expression, which can be slow and may require complex mutation event to occur [4].

Another category of interest that is significantly over-represented in our analysis is “developmental process.” Sexual dimorphisms are thought to be favored because of the different roles that the sexes have to play at the adult stage [31,32]. Yet, dissociating developmental pathways between the sexes could be necessary to achieve morphological sexual dimorphism, for example size dimorphism [3], and many examples of developmental differences between the sexes exist, as reviewed in Aiken et al. [33]. This suggests that substantial SA selection can occur during development and could be one cause of the high number of spontaneous abortions in humans. Moreover, genes shared by at least two populations (SDG_{sh}) are functionally enriched in GO terms involved in developmental processes, indicating that SA selection during development could occur in several human populations.

Additionally, several functional categories identified as being enriched in this study encompass traits known to show at least minor sexual dimorphisms in humans. Using the manual and the systematic clustering methods, we found several terms referring to “reproduction”, covering highly dimorphic traits, and to “sugar metabolism and energy” for which several studies reported sexually dimorphic traits [34,35]. Likewise, the category “immune system” encompasses terms that can be linked to sexually dimorphic traits; some examples can be found in Klein [36] and Marriot et al. [37]. These findings are consistent with the expectation that when studying traits under IASC, one should find sexually dimorphic traits [38].

Another region of particular interest to study SA polymorphism would be the Pseudo Autosomal Regions (PARs), and it has recently been the subject of several theoretical and empirical studies [39–41]. Indeed, they are the only regions of the sex chromosomes to exhibit autosomal features for recombination and inheritance, but their evolutionary dynamics are also influenced by their sex linkage. Under some conditions, the maintenance of SA polymorphisms on the PARs is facilitated compared to the sex-specific regions of the X chromosome and to the autosomes. The emergence of SA polymorphisms in PARs can cause female-beneficial alleles to be preferentially carried by the X chromosome and male-beneficial alleles by the Y chromosome, potentially leading to the loss of recombination in the PARs [42]. However, to study this complex region of the genome, the quality of the data needs to be thoroughly controlled. The genotyping of the PARs in the HapMap dataset is not

of high enough quality to further study the PARs. It would be interesting to investigate signatures for SA selection in these regions in the future with sequencing datasets.

CONCLUSION

In this study, for the first time, we used a genome-wide scan to identify genetic signatures of SA selection. We detected multiple loci that potentially undergo IASC in natural human populations, providing a list of SNPs enriched for targets of SA selection. Our results support a preferential location of SA loci on the X chromosome in humans. The use of trio datasets, including genome-wide sequencing of parents and children, would enable studying sex-specific transmission distortions to assess the strength of SA selection during embryonic development and childhood.

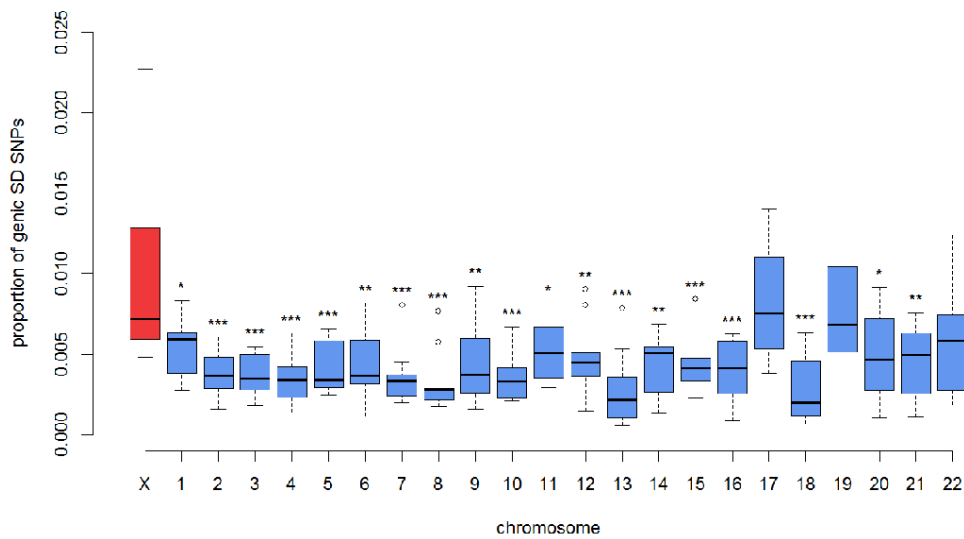
ACKNOWLEDGMENTS

We thank Paul Verdu, Frederic Austerlitz and Patricia Balaesque for helpful discussions and comments on an earlier version of the manuscript. EL was financed by a PhD grant from the French Ministry of Higher Education and Research.

FIGURES LEGENDS

Figure 1- Proportions of genic SD SNPs per chromosome (A) before LD pruning and (B) after LD pruning (filtering of genic SNPs with a $r^2 > 0.1$). For each chromosome, the distribution of these proportions (number of genic SD SNPs after FDR correction at the gene level / number of genic SNPs) over the 10 HapMap populations is represented by a boxplot. Levels of significance of one-sided Wilcoxon–Mann–Whitney tests between the distributions of proportions on the X chromosome (in red) and each of the autosomes (in blue) are reported (p-value: * < 0.05 , ** $< 10^{-2}$, *** $< 10^{-3}$).

A



B

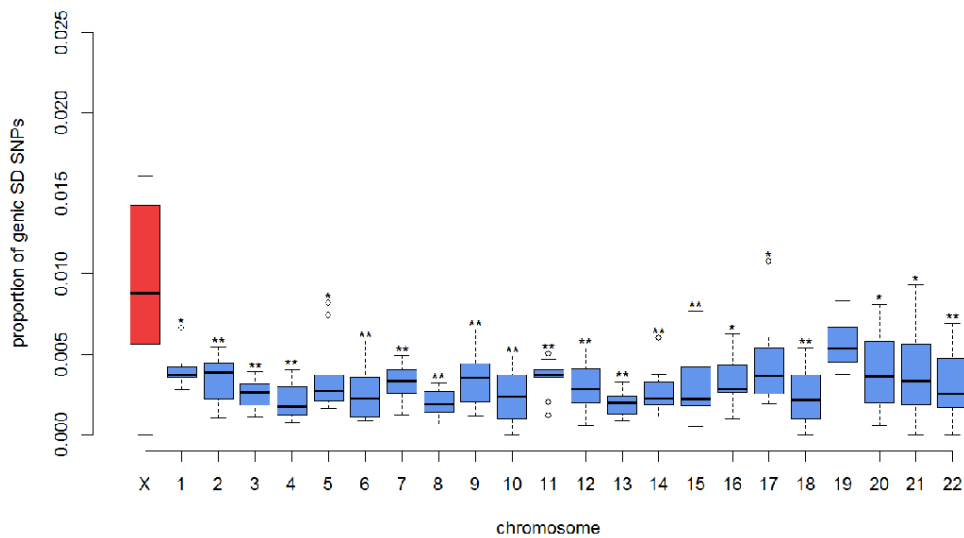


Figure 2- Proportions of genic SD SNPs per population for chromosomes X, 10 and 22. For each population, these proportions (number of genic SD SNPs after FDR correction at the gene level / number of genic SNPs) is shown for chromosomes X, 10 and 22 after LD-pruning ($r^2 > 0.1$). A Fisher's exact test was performed to compare the proportions between the X chromosome and each autosome (p-value: * <0.05 , ** $<10^{-2}$, *** $<10^{-3}$).

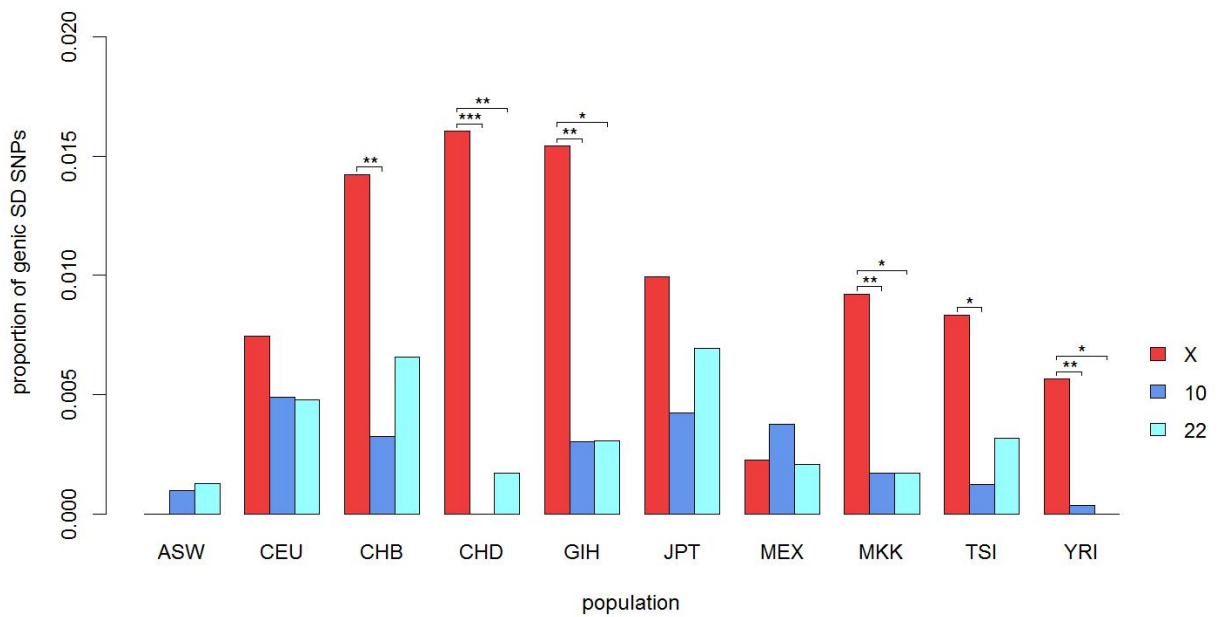
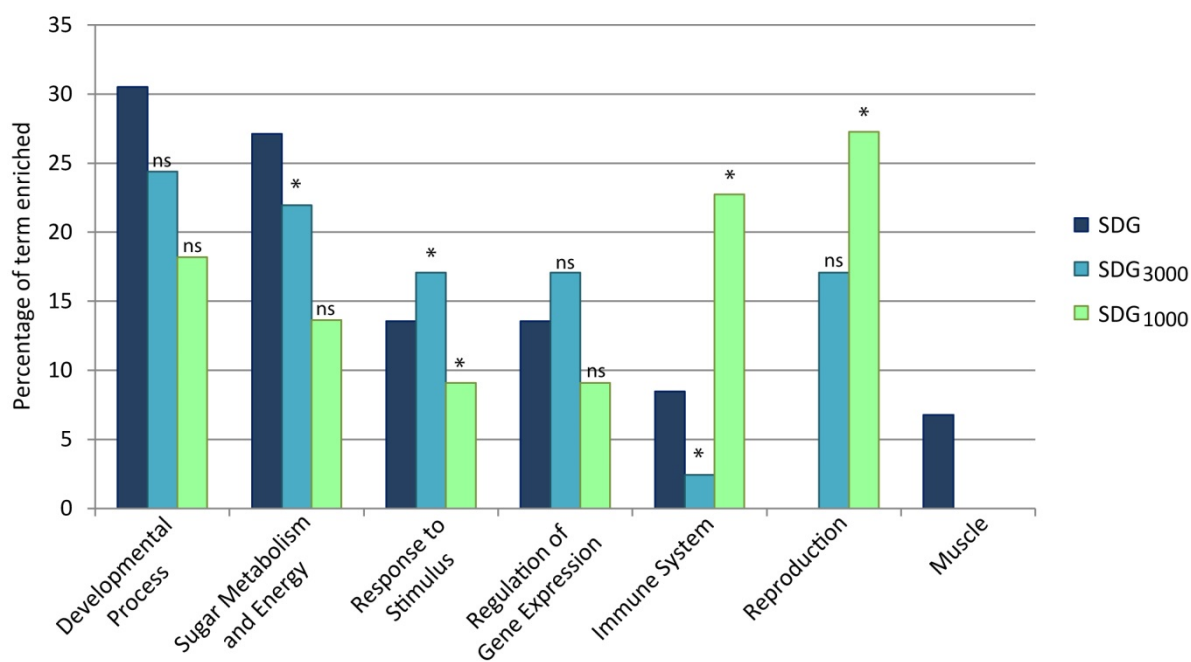


Figure 3- Percentage of GO terms significantly enriched in genes containing SD SNPs classified into seven manually curated categories. The results are presented for the whole set of SD genes (SDG, 5705 genes) and for the sets of SD genes containing the 3000 and 1000 genes with the lowest p-values (SDG₃₀₀₀ and SDG₁₀₀₀). A star means that a similar category was found using the automated DAVID clustering approach with the corresponding list, and “ns” means that it was not significant. As the SDG list contains more than 3000 genes, it was not possible to perform the automated method on this list.



REFERENCES

1. Van Doorn, G. S. 2009 Intralocus sexual conflict. *Ann. N. Y. Acad. Sci.* **1168**, 52–71. (doi:10.1111/j.1749-6632.2009.04573.x)
2. Stulp, G., Kuijper, B., Buunk, A. P., Pollet, T. V & Verhulst, S. 2012 Intralocus sexual conflict over human height. *Biol. Lett.* **8**, 976–8. (doi:10.1098/rsbl.2012.0590)
3. Badyaev, A. V. 2002 Growing apart: an ontogenetic perspective on the evolution of sexual size dimorphism. *Trends Ecol. Evol.* **17**, 369–378. (doi:10.1016/S0169-5347(02)02569-7)
4. Stewart, A. D., Pischedda, A. & Rice, W. R. 2010 Resolving intralocus sexual conflict: genetic mechanisms and time frame. *J. Hered.* **101**, S94–9. (doi:10.1093/jhered/esq011)
5. Rice, W. R. 1984 Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735–742.
6. Kidwell, J. F., Clegg, M. T., Stewart, F. M. & Prout, T. 1977 Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics* **85**, 171–183.
7. Fry, J. D. 2010 The genomic location of sexually antagonistic variation: some cautionary comments. *Evolution* **64**, 1510–6. (doi:10.1111/j.1558-5646.2009.00898.x)
8. Gibson, J. R., Chippindale, A. K. & Rice, W. R. 2002 The X chromosome is a hot spot for sexually antagonistic fitness variation. *Proc. Biol. Sci.* **269**, 499–505. (doi:10.1098/rspb.2001.1863)
9. Foerster, K., Coulson, T., Sheldon, B. C., Pemberton, J. M., Clutton-Brock, T. H. & Kruuk, L. E. B. 2007 Sexually antagonistic genetic variation for fitness in red deer. *Nature* **447**, 1107–10. (doi:10.1038/nature05912)
10. Innocenti, P. & Morrow, E. H. 2010 The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol.* **8**, e1000335. (doi:10.1371/journal.pbio.1000335)
11. Fedorka, K. M. & Mousseau, T. A. 2004 Female mating bias results in conflicting sex-specific offspring fitness. *Nature* **234**, 65–67.
12. Delcourt, M., Blows, M. W. & Rundle, H. D. 2009 Sexually antagonistic genetic variance for fitness in an ancestral and a novel environment. *Proc. R. Soc.* **276**, 2009–14. (doi:10.1098/rspb.2008.1459)
13. Calsbeek, R. & Sinervo, B. 2003 Within-clutch variation in offspring sex determined by differences in sire body size: cryptic mate choice in the wild. *J. Evol. Biol.* **17**, 464–470. (doi:10.1046/j.1420-9101.2003.00665.x)
14. Altshuler, D. M. et al. 2010 Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–8. (doi:10.1038/nature09298)
15. Balaesque, P., Toupance, B., Quintana-Murci, L., Crouau-Roy, B. & Heyer, E. 2004 Sex-specific selection on the human X chromosome? *Genet. Res.* **83**, 169–176. (doi:10.1017/S0016672304006858)
16. Pemberton, T. J., Wang, C., Li, J. Z. & Rosenberg, N. A. 2010 Inference of unexpected genetic relatedness among individuals in HapMap Phase III. *Am. J. Hum. Genet.* **87**, 457–64. (doi:10.1016/j.ajhg.2010.08.014)
17. Altschup, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990 Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410.
18. Karolchik, D. et al. 2008 The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* **36**, D773–9. (doi:10.1093/nar/gkm966)

19. Benjamini, Y. & Hochberg, Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300.
20. Purcell, S. et al. 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75. (doi:10.1086/519795)
21. The 1000 Genome Project Consortium 2010 A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–73. (doi:10.1038/nature09534)
22. Huang, D. W., Sherman, B. T. & Lempicki, R. A. 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57. (doi:10.1038/nprot.2008.211)
23. Crans, G. & Shuster, J. 2008 How conservative is Fisher’s exact test? A quantitative evaluation of the two-sample comparative binomial trial. *Stat. Med.* **27**, 3598–3611. (doi:10.1002/sim)
24. Pritchard, J. K. & Przeworski, M. 2001 Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**, 1–14. (doi:10.1086/321275)
25. Benagiano, G., Farris, M. & Grudzinskas, G. 2010 Fate of fertilized human oocytes. *Reprod. Biomed. Online* **21**, 732–41. (doi:10.1016/j.rbmo.2010.08.011)
26. Ucisik-Akkaya, E., Davis, C. F., Do, T. N., Morrison, B. a, Stemmer, S. M., Amadio, W. J. & Dorak, M. T. 2010 Examination of genetic polymorphisms in newborns for signatures of sex-specific prenatal selection. *Mol. Hum. Reprod.* **16**, 770–7. (doi:10.1093/molehr/gaq047)
27. Fenner, J. N. 2005 Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–23. (doi:10.1002/ajpa.20188)
28. Fitzpatrick, M. J. 2004 Pleiotropy and the genomic location of sexually selected genes. *Am. Nat.* **163**, 800–8. (doi:10.1086/386297)
29. Grimwood, J. et al. 2004 The DNA sequence and biology of human chromosome 19. *Nature* **428**, 529–35. (doi:10.1038/nature02399)
30. Ellegren, H. & Parsch, J. 2007 The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.* **8**, 689–98. (doi:10.1038/nrg2167)
31. Chippindale, A. K., Gibson, J. R. & Rice, W. R. 2001 Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1671–5. (doi:10.1073/pnas.041378098)
32. Cox, R. M. & Calsbeek, R. 2009 Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *Am. Nat.* **173**, 176–87. (doi:10.1086/595841)
33. Aiken, C. E. & Ozanne, S. E. 2013 Sex differences in developmental programming models. *Reproduction* **145**, R1–13. (doi:10.1530/REP-11-0489)
34. Nookaew, I. et al. 2013 Adipose tissue resting energy expenditure and expression of genes involved in mitochondrial function are higher in women than in men. *J. Clin. Endocrinol. Metab.* **98**, E370–8. (doi:10.1210/jc.2012-2764)
35. Shi, H. & Clegg, D. J. 2009 Sex differences in the regulation of body weight. *Physiol. Behav.* **97**, 199–204. (doi:10.1016/j.physbeh.2009.02.017)
36. Klein, S. L. 2012 Immune cells have sex and so should journal articles. *Endocrinology* **153**, 2544–50. (doi:10.1210/en.2011-2120)
37. Marriott, I. & Huet-Hudson, Y. M. 2006 Sexual dimorphism in innate immune responses to infectious organisms. *Immunol. Res.* **34**, 177–92.

(doi:10.1385/IR:34:3:177)

38. Bonduriansky, R. & Chenoweth, S. F. 2009 Intralocus sexual conflict. *Trends Ecol. Evol.* **24**, 280–8. (doi:10.1016/j.tree.2008.12.005)
39. Qiu, S., Bergero, R. & Charlesworth, D. 2013 Testing for the footprint of sexually antagonistic polymorphisms in the pseudoautosomal region of a plant sex chromosome pair. *Genetics* **194**, 663–72. (doi:10.1534/genetics.113.152397)
40. Charlesworth, B., Jordan, C. Y. & Charlesworth, D. 2014 The evolutionary dynamics of sexually antagonistic mutations in pseudoautosomal regions of sex chromosomes. *Evolution* **68**, 1339–50. (doi:10.1111/evo.12364)
41. Kirkpatrick, M. & Guerrero, R. F. 2014 Signatures of Sex-Antagonistic Selection on Recombining Sex Chromosomes. *Genetics* **197**, 531–541. (doi:10.1534/genetics.113.156026)
42. Otto, S. P. et al. 2011 About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet.* **27**, 358–67. (doi:10.1016/j.tig.2011.05.001)

Supplementary Materials

Table of content

Supplementary Text.....	2
Text S1- Sex-specific population structure.....	2
Text S2- Influence of admixture and sample size on genic SD SNPs proportions in the ASW population.....	2
Text S3- Analysis of the signal distribution close to the candidate SNPs.....	3
Text S4- Pattern of genic SD SNPs proportions for random groups.....	4
Text S5- Functional analysis of the SD SNPs with extreme signals of F_{ST} after correction for multiple testing at the genome-wide and chromosomal level	5
References	7
Supplementary Figures.....	8
Figure S1	8
Figure S2	9
Figure S3	11
Figure S4	13
Figure S5	14
Figure S6	15
Figure S7	16
Figure S8	17
Figure S9	19
Figure S10	21
Figure S11	22
Figure S12	23
Figure S13	25
Figure S14	27
Supplementary Tables	28
Table S1.....	28
Table S2.....	29
Table S3.....	30
Table S4.....	31
Table S5.....	32
Table S6.....	33
Table S7.....	34
Table S8.....	40

Supplementary Text

Text S1- Sex-specific population structure

Sex-specific structure at the genome-wide level (i.e., genome-wide differences in allelic frequencies between the sexes) might occur for various reasons, including sex-biased sampling in a population with fine-scale genetic structure or sex-specific admixture processes (i.e., if mating occurs preferentially between males and females originating from genetically differentiated populations). In the case of a single event of sex-specific admixture, the differences in allelic frequencies between sexes would disappear after one generation on the autosomes, while it would take longer to vanish on the X chromosome, leading to a higher differentiation between males and females as compared to autosomes (Balaesque et al. 2004). However, recurrent sex-specific admixture events, with constant migrations biased toward one sex, would affect allelic frequencies both on the autosomes and the X chromosome.

To assess whether such structures exist in our dataset, and to measure their potential influence on the allelic frequencies of the autosomes and the X chromosome, we performed a multidimensional scaling (MDS) analysis on each of the 11 HapMap populations and tested whether they were sex-specifically structured. We calculated the Allele Sharing Distance (ASD) (Bowcock et al. 1994) between individuals using the software [asd](http://szpiech.com/software.html) (<http://szpiech.com/software.html>) developed by Zachary Szpiech and performed MDS on these genetic distances (Figure S2-S3). Because missing data result in skewed ASD values, we included only individuals harboring less than 0.5% of missing data. Additionally, we LD-pruned the data to keep only independent SNPs ($r^2 > 0.25$). For the X chromosome, to circumvent the problem of male hemizyosity when calculating ASD, we considered only one X chromosome per female by randomly picking one allele for each SNP and calculated a haploid ASD. We repeated this sampling 30 times.

To test if the observed mean genetic distance between males and females was significant, we randomly permuted 10,000 times the ‘male’ and ‘female’ labels in the distance matrix, and calculated a new mean distance between males and females for each permutation. We then tested whether our observed value was extreme (in the top 5%) compared to the distance distribution obtained after permutations. On the autosomes (Table S4), one population, LWK, showed a significant mean genetic distance between males and females and was therefore removed from the analysis. For the other ten populations, we showed that there is no sex-specific genetic structuration on the autosomes. On the X chromosome, the CHB population presented significant genetic distances between males and females for all 30 repeats (Table S5). This structure was found to be driven by two individuals, NA18544 and NA18558, two outliers on the MDS. When removed, the genetic distances observed were not significant anymore. The permutations method we used to assess significance is therefore greatly influenced by outliers. As it is unlikely that two males over 53 would affect the overall differences in allelic frequencies between males and females, we did not remove these individuals from the study. We found no X-linked sex-specific structuration in the other populations.

We can therefore consider that there are no substructures linked to the sexes in the ten populations further included in the study for the X chromosome or the autosomes.

Text S2- Influence of admixture and sample size on genic SD SNPs proportions in the ASW population

In order to explore the possible effect of recent (non sex-specific) admixture on genic SD SNPs proportions, we simulated ten admixed populations (ADP₁₋₁₀) by randomly sampling individuals half from CEU and half from YRI and merging them into new

populations with the same sample size and sex-ratio as ASW. The source populations, CEU and YRI, have been chosen because of the known admixed history of the ASW population. Because one of the source populations of the MEX population was not available here (i.e., Amerindian populations), we did not perform this analysis on this population.

ASW population has also a much smaller sample size than CEU and YRI populations; therefore the exact Fisher's test is less powerful to detect genetic differentiation between the sexes in this population. To discriminate the admixture effect from a potential sample size bias, we therefore also compared the proportions of genic SD SNPs between ASW and ten subsamples of CEU (CEUR₁₋₁₀) and ten subsamples of YRI (YRIR₁₋₁₀) with the same number of individuals than ASW and the same sex ratio.

We computed the 99.9% confidence intervals of the mean proportions of genic SD SNPs over the simulated admixed populations ADP₁₋₁₀ and over the subsamples CEUR₁₋₁₀ and YRIR₁₋₁₀. We then aimed to compare the proportions observed in the original HapMap populations and those observed in the simulated populations.

For the X chromosome, the proportion of genic SD SNPs in ASW is not included in the confidence interval of neither the CEUR₁₋₁₀ and YRIR₁₋₁₀ populations, nor the ADP₁₋₁₀ populations (Figure S7A-B). On the contrary, the proportions of genic SD SNPs in ASW on autosomes are always included in at least one of the CEUR₁₋₁₀ or YRIR₁₋₁₀ confidence intervals, except for chromosome 3, and most of them are included in the ADP₁₋₁₀ confidence intervals. Therefore, the mean proportions of genic SD SNP over the resized populations and the simulated admixed populations are overall equal to the autosomal genic SD SNP proportions observed in ASW, while they are significantly higher than the X chromosome genic SD SNP proportion observed in ASW.

These results suggest that, for the autosomes, the lower proportion of genic SD SNPs observed in ASW compared to CEU and YRI could be due to a sample size bias. On the contrary, the lower proportion of genic SD SNPs in ASW observed for the X chromosome cannot be explained by a sample size and/or recent admixture effect. Moreover, it is interesting to see that the effect of admixture combined to sample size and the effect of sample size alone on the proportions of genic SD SNP are the same; therefore it is difficult to discriminate the effect of admixture.

Text S3- Analysis of the signal distribution close to the candidate SNPs

We have shown that the signal of differentiation in allelic frequencies between the sexes is mostly population-specific (Table S4), suggesting that selection is recent. If so, we expect the signal around candidate SNPs to extend in the neighboring region. To test if it is the case, we studied every SNPs neighboring each genic SD SNP in a window of ± 300 kb. To quantify the signal of differentiation in allelic frequencies between the sexes, we used the fixation index F_{ST} , a commonly used measure of divergence between groups of chromosomes. The haploid formula in Weir (Weir 1996) was used to take into account the differences in sample sizes between the sexes for the X chromosome. We used the per-population genome-wide list of genic SD SNPs after LD-pruning. The mean number of genic SD SNPs per population is therefore of 118.5 SNPs. We then compared the results with those obtained when sampling random genic SNPs as focal SNPs, chosen to match the allelic diversity of the genic SD SNPs. We considered intervals of 1 kb for the first 50 kb, and then of 5 kb between 50 kb and 300 kb. The intervals are non-overlapping, and each SNP is present at most once. Any SNP located in the vicinity of several focal SNPs was associated with the closest. The F_{ST} was first computed per SNP. For each focal SNP, we then computed a mean F_{ST} for every interval. Finally, F_{ST} was averaged per interval over all focal SNPs. The 95% quantiles were calculated for each interval. The result for the CEU, JPT and YRI populations are shown in Figure S9A and the results for the admixed populations, ASW and MEX are shown in Figure S9B. The other populations are not shown, but the results are consistent with those of CEU,

JPT and YRI. For every population, the signal observed at the genic SD SNPs spreads in the neighboring regions, and is significantly higher than the signal observed around random genic SNPs, until approximately ± 100 kb, depending on the population (one sided Wilcoxon–Mann–Whitney test at a significance level of 5%, Figure S10). The F_{ST} averaged over the genic SD SNPs is always included in the 95% quantile of the neighboring intervals. This suggests that, even if the high F_{ST} observed for genic SD SNPs is due to a winner’s curse effect, i.e., a value artificially high due to the method of detection, the neighboring SNPs show a lower, but significantly higher than random, F_{ST} signal.

Interestingly, while the proportions of genic SD SNPs are lower in the admixed populations (ASW and MEX) than for the other populations (Figure 2), we find that the F_{ST} signal is higher. As developed in Text S2, it is likely to be due to fluctuations accounting for the smaller sample size of these populations, leading to an increased winner’s curse effect.

Overall, these results are consistent with recent and population-specific selection events.

Text S4- Pattern of genic SD SNPs proportions for random groups

In this study, we calculated the proportion of genic SNPs showing a significant difference in allelic frequencies between males and females (genic SD SNPs) for each population. We found that the X chromosome harbors a significantly higher proportion of genic SD SNPs than most autosomes. To further ascertain that this higher proportion is due to IASC, we ask whether we still find the same pattern when performing our analyses on 2 random groups of individuals sampled in each population.

For this analysis, we removed the admixed populations (ASW and MEX) because they do not present the same pattern than the other populations (Figure 2, see discussion in Text S3).

We generated N independent bipartitions for each population by randomly dividing it into sets of 2 groups of individuals. To measure the difference in sex ratio between the 2 groups, we used the absolute value of the difference between the proportion of females in one group and the proportion of females in the other group (Δ_p). The N bipartitions were constructed to obtain all possible values for Δ_p , which ranks from 0 when both groups have the same proportion of males and females, to 1 when there are only females in one group and only males in the other. The aim of this analysis was to test whether the difference between the X chromosome’s and the autosomes’ proportions of SNPs showing significant differences in allelic frequencies between the 2 groups increases with the intergroup sex-ratio differential (Δ_p). This will give us clues about the nature of the signal we detected, and whether it is sex-specific.

The number of individuals in each group was determined by the number of males and females in the focal population. For the first set, group A was exclusively composed of males, and group B of females. For the autosomes, between each set, group A gained a female and lost a male, and group B gained a male and lost a female. For the X chromosome, between each set, group A lost 2 males and gained 1 female, while group B gained 2 males and lost 1 female. The framework was different between the autosomes and the X chromosome to preserve the same number of chromosomes in each group across the different possible bipartitions, and consequently the same statistical power. We repeated the resampling for each value of Δ_p 10 times.

This analysis was performed on genic SNPs only (within 5 kb from gene boundaries). For each set, after performing per-SNP Fisher’s exact tests, a correction for multiple testing (FDR) of the Fisher’s exact test p-values was implemented per gene (see material and methods). The set of genic SNPs was then LD-pruned as explained in the materials and methods. The proportion of genic SD SNPs (genic SNPs with a significant p-value after FDR per gene) was calculated for each chromosome.

To compare the result of this analysis with the observed data, we associated sets with the closest Δ_p for each populations and chromosomes. The difference Δ_{X-A} between the X chromosome's and the autosomes' proportions of genic SD SNPs averaged over the 8 populations considered in this analysis was calculated for each set. The relation between the pooled values of Δ_{X-A} and the values of Δ_p was then investigated (Figure S10A). Δ_{X-A} is significantly and positively correlated with Δ_p (Spearman correlation, $\rho = 0.92$, $p = 4.04e-6$). However, the p-values of these differences are always significant (one sided Wilcoxon test), even when Δ_p is close to 0.

The X chromosome has a concentration of SNPs per gene more than 2-fold lower than the other chromosomes (2.2 SNPs/kb for the X chromosome compared to a mean of 4.9 SNPs/kb for the autosomes). This can be due to the lower diversity inside genic regions observed on the X compared to autosomes (Arbiza et al. 2014), or to an ascertainment bias. The FDR correction used here corrects for the number of SNPs per gene, therefore the X chromosome would be less corrected for multiple testing, leading to an increase of genic SD SNPs compared to the autosomes. Additionally, as we can see on Figure S1, the p-values of the Spearman correlation between the number of SD SNPs and the gene length are overall less significant for the X chromosome compared to the autosomes.

However, we have seen that Δ_{X-A} is strongly correlated with Δ_p , which means that the signal we consider increases when the groups are composed of males and females only. To test whether the increase of Δ_{X-A} is an artifact of the FDR correction, we computed the proportion of SD SNPs, without FDR correction, per chromosome (Figure S01B). The Δ_{X-A} shows the same pattern than the previous analysis, and is also significantly and positively correlated with Δ_p (Spearman correlation, $\rho = 0.93$, $p = 4.11e-6$). However, the proportions of SD SNPs on the X chromosome become significantly higher than those of the autosomes only when $\Delta_p \geq 0.72$, reaching a maximum of 12 autosomes over 22 showing a significantly lower proportion of SD SNPs than the X chromosome for $\Delta_p = 1$ (Figure S11).

The FDR correction per gene increases the proportion of genic SD SNPs on the X chromosome compared to autosomes. However, when we consider the Fisher's exact tests p-values without FDR correction and after LD pruning, we observe that Δ_{X-A} is strongly correlated with Δ_p , and that the comparison between the X chromosome's and autosomes' proportions are never significant when the 2 groups are homogenous. This is consistent with the hypothesis that the signal we detect is mostly sex-specific. Moreover, when $\Delta_p = 1$, the distribution of Δ_{X-A} before FDR correction per gene (0.0007-0.0105, median=0.0068) overlaps strongly with the distribution of Δ_{X-A} after correction (0.0045-0.0087, median=0.0076), suggesting that for groups composed of only males and only females, the Δ_{X-A} is only slightly affected by the FDR correction per gene.

Text S5- Functional analysis of the SD SNPs with extreme signals of F_{ST} after correction for multiple testing at the genome-wide and chromosomal level

To further investigate the functional significance of chromosomal and genome-wide SD SNPs, we searched for potential regulatory regions in their vicinity using the UCSC genome browser (Karolchik et al. 2008), also considering neighboring SNPs in high Linkage Disequilibrium (LD).

The linkage disequilibrium (LD) was calculated as a r^2 value between every possible pair of SNPs from each set including the extreme SD SNP and ± 50 SNPs using the formula from the R package genetics version 1.3.8 (Warnes et al. 2012) re-implemented in Python.

The SD SNPs located on chromosomes 2 and 20 are in LD with several adjacent SNPs, also displaying low F_{ST} p-values (Figure S12, Figure S13C). On the contrary, both X-linked SNPs identified are in high LD with each other ($r^2 > 0.8$), but are isolated from their neighbouring SNPs which do not show any sex-specific differences in allelic frequencies

(Figure S13A-B). These patterns are consistent with the expectation that loci undergoing SA selection should be in LD with each other (Ubeda et al. 2011).

We examined the chromatin state segmentation based on epigenetics markers (Broad ChromHMM, integrating ChIP-seq data for eight common histone modifications) which indicates whether the region might be transcribed or not. We also looked at the Transcription Factor Binding Sites (TFBS conserved sites track) identified by ChIP-seq assays and the DNaseI hypersensitivity sites (ENCODE DNase Cluster track). The DNaseI cuts DNA in unoccupied regions; therefore a region of DNaseI hypersensitivity is likely to be a binding site of a transcription factor. Finally, we looked for signals of Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE, UNC FAIRE track) assays that determine if the region is associated with a regulatory activity by revealing which regions are depleted of nucleosome. We also looked for potential eQTLs in the region on the University of Chicago QTL Browser (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>).

We found no signal of FAIRE or eQTLs in any of the regions studied.

For the chromosomal SD SNPs showing extreme signals of SA polymorphisms on the X chromosome, rs5943145 maps in a repressed region for one cell type and is located in an intron of DCX, whose protein product directs neuronal migration. Interestingly, defects in this gene cause different diseases between the sexes: a “double cortex” syndrome in females and a “smooth brain” syndrome in males, suggesting that the protein has a different role in both sexes. The extreme SD SNP rs7053144 maps in an intron of ZCCHC16, a zinc finger protein.

Concerning chromosome 20, a functional analysis of the region containing extreme signals of SA polymorphism reveals that several SNPs in LD with the focal chromosomal SD SNPs are located in regions of interest (Figure S14). Indeed, two SNPs (rs1201897 and rs1201989) are located in a strong enhancer including one that is also located in a region of weak DNaseI hypersensitivity (rs1201989). Three other SNPs (2: rs1201901, 18: rs1201987, 19: rs1201988) map in regions of weak enhancer activity and two (15: rs1201966, 16: rs1201973) in weakly transcribed regions. Several of the SNPs that map in region of enhancer activity or transcription also map in repressed region in other cell types. It suggests that the chromosomal SD SNPs identified here might have a functional significance.

The chromosomal SD SNPs on chromosome 2 (rs2249737) maps in a region of transcription elongation for all cell types studied. Several SNPs in LD with the focal chromosomal SD SNPs map in genic regions including OSBPL6, an intracellular lipid receptor, PRKRA, involved in the interferon response to viral infection, DFNB59, which has a role in auditory pathway neurons functions, FKBP7, a molecular chaperon and PLECKHA3 which contains a Pleckstrin homology domain (Figure S12). Inside the SNP cluster comprising rs2114582, rs10207436, rs2059691 and rs4141635, 3 SNPs in LD with rs2249737 map in several region of transcription elongation (rs2114582, rs10207436, rs2059691) while rs4141635 map in a region of DNase hypersensitivity, in two transcription factor binding sites and in a region with a chromatin state defined as insulator (Figure S12).

This latter SNP, rs4141635, is of particular interest. Indeed, it is both in high LD with the focal chromosomal SD SNP ($r^2 > 0.6$) and located inside a region showing multiple overlapping signals of regulatory functions (Figure S12). In addition, one of the transcription factors presenting a binding site in this region, CTCF, is known as a chromatin insulator (Bell et al. 1999). If the SNP disrupts the binding site of CTCF, this could create a difference in expression of the set of genes on which the insulator acts according to the SNP allele.

References

- Arbiza L, Gottipati S, Siepel A, Keinan A. 2014. Contrasting X-linked and autosomal diversity across 14 human populations. *Am J Hum Genet* **94**: 827–844.
- Balaresque P, Toupance B, Quintana-Murci L, Crouau-Roy B, Heyer E. 2004. Sex-specific selection on the human X chromosome? *Genet Res* **83**: 169–176.
- Bell AC, West AG, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**: 387–96.
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**.
- Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Hartel R, Hinrichs S, Hsu F, et al. 2008. The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res* **36**: D773–9.
- Ubeda F, Haig D, Patten MM. 2011. Stable linkage disequilibrium owing to sexual antagonism. *Proc Biol Sci* **278**: 855–62.
- Warnes G, Gorjanc G, Leisch F, Man M. 2012. genetics: Population Genetics.
- Weir BS. 1996. *Genetic Data Analysis II*.

Supplementary Figures

Figure S1- Effect of FDR corrections at the gene level in assessing SD SNPs. P-values of Spearman correlation tests between the length of a gene and its number of SD SNPs **(A)** before and **(B)** after FDR correction at the gene level. P-values lower than 10^{-5} were assigned to the darkest blue and p-values greater than 0.05 are shown in white.

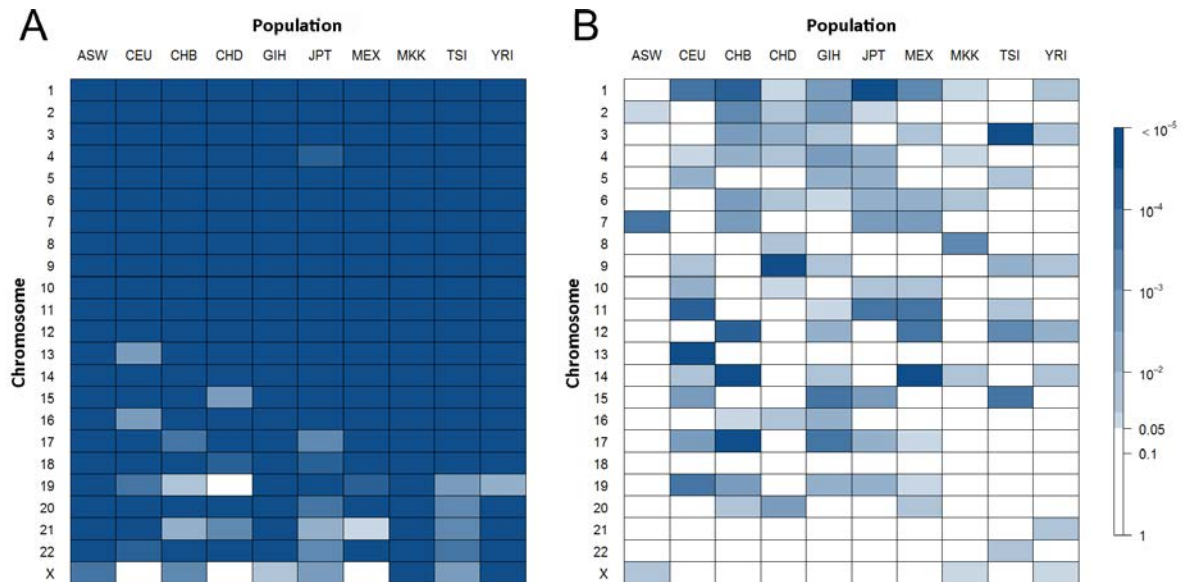
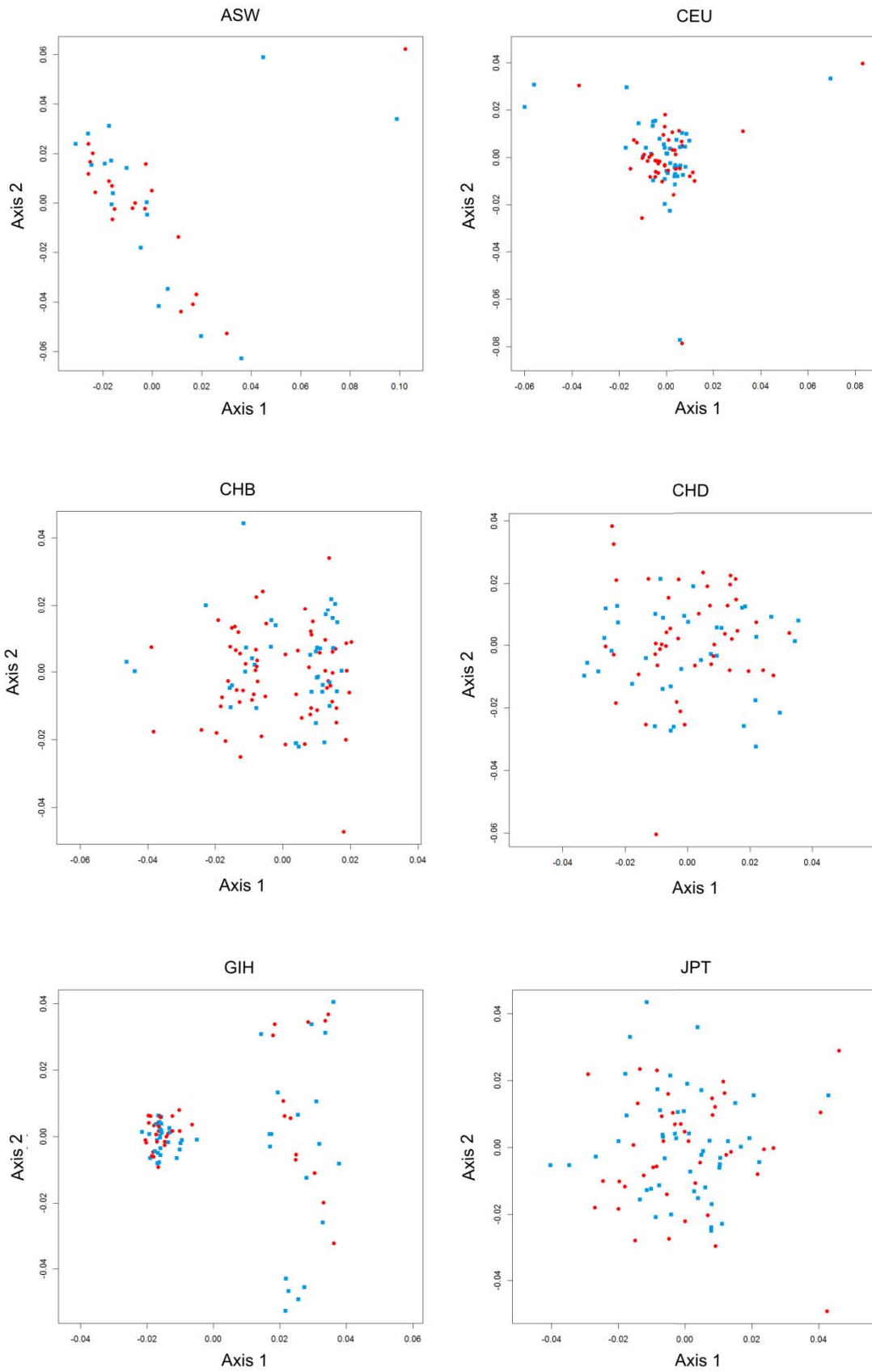


Figure S2- MDS plots computed from autosomal genetic data. Blue dots: male individuals, red dots: female individuals.



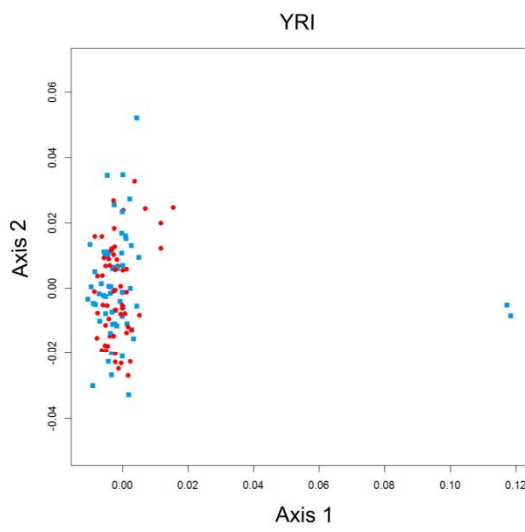
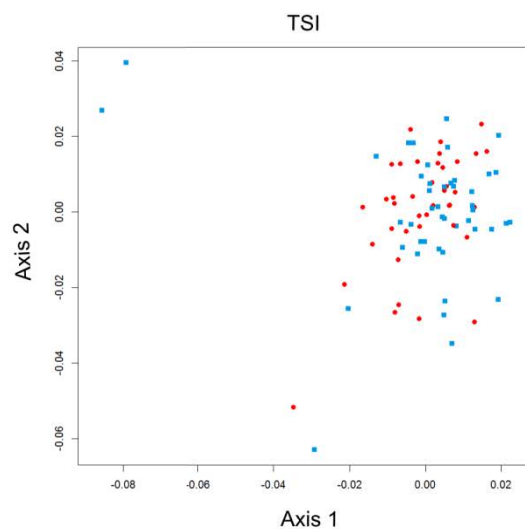
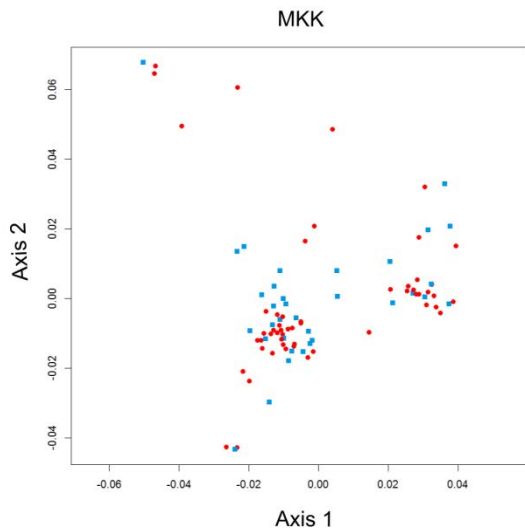
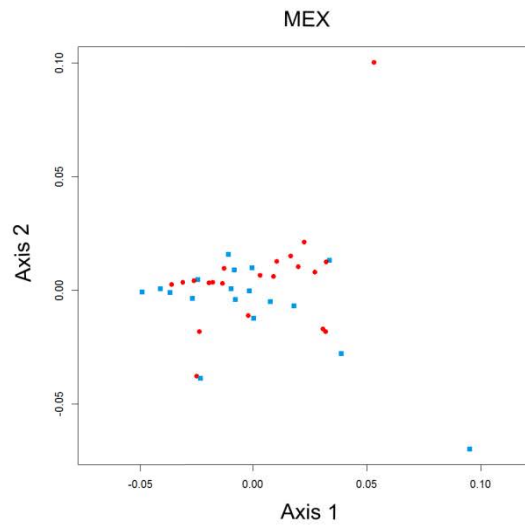
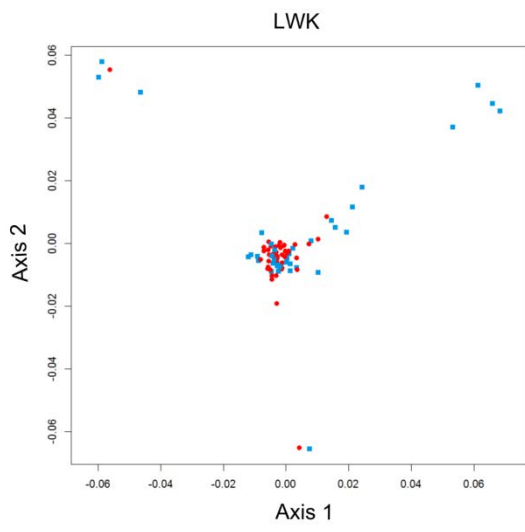
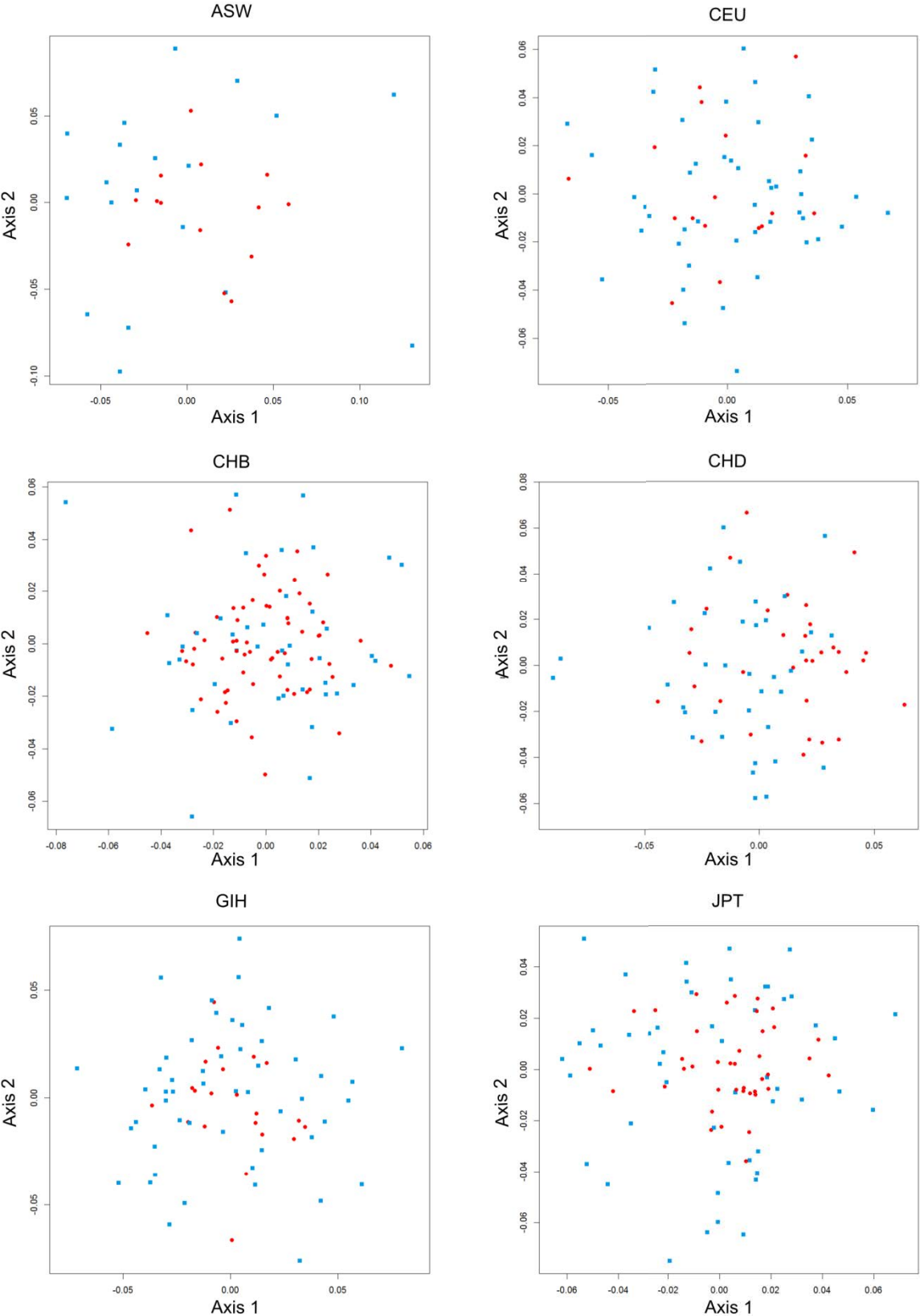


Figure S3- MDS plots computed from X-linked genetic data for the first female subsampling (out of 30, see Text S1). Blue dots: male individuals, red dots: female individuals.



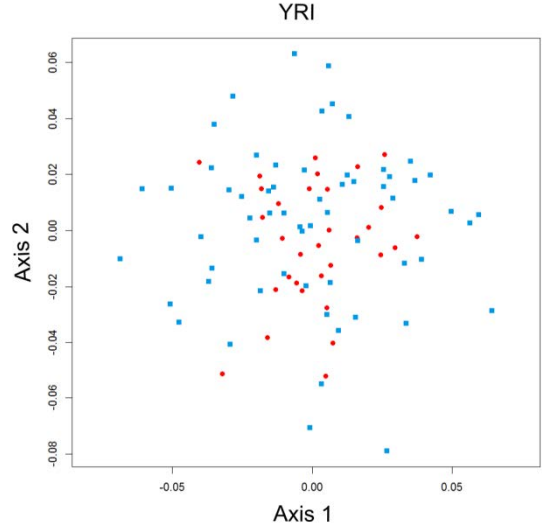
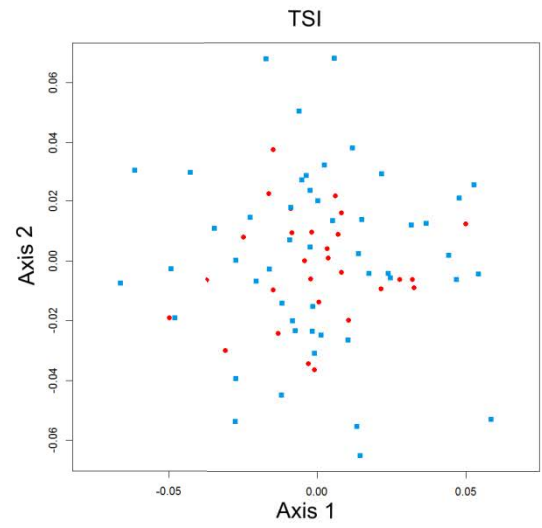
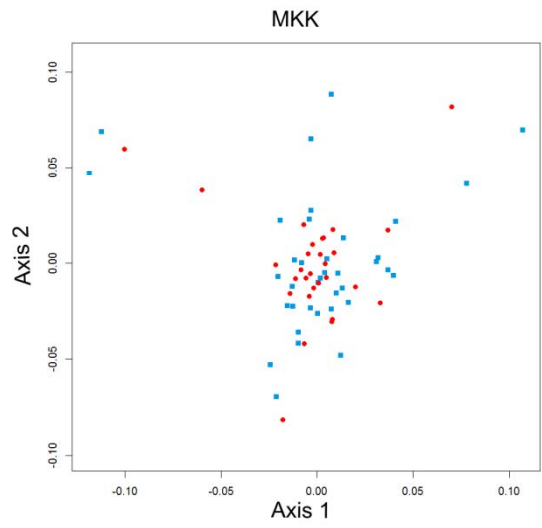
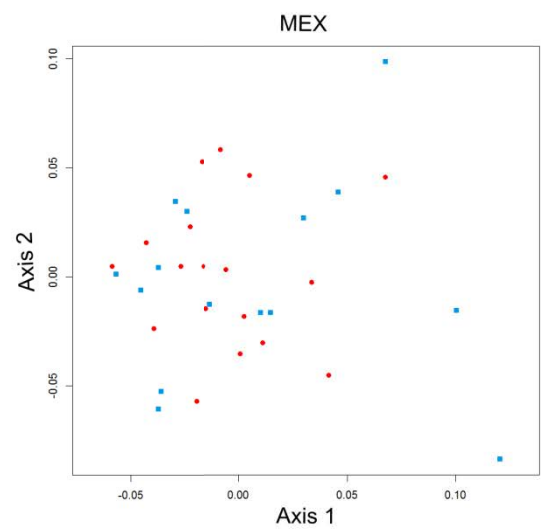
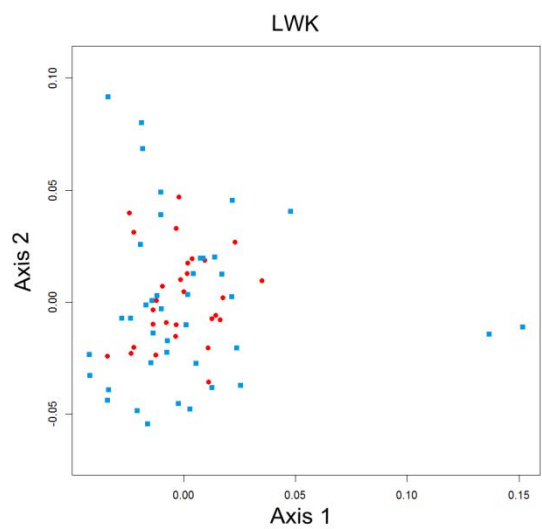


Figure S4- Effect of haplodisation of autosomes in males on the mean proportion of genic SD SNPs per chromosome. The distribution of proportions of genic SD SNPs (number of genic SD SNPs after FDR correction at the gene level / number of genic SNPs) over the 10 HapMap populations is shown per chromosome. Levels of significance of one-sided Wilcoxon–Mann–Whitney tests between the mean proportions of the X chromosome and each of the autosomes are reported (p-value: * < 0.05, ** < 10^{-3} , *** < 10^{-4}).

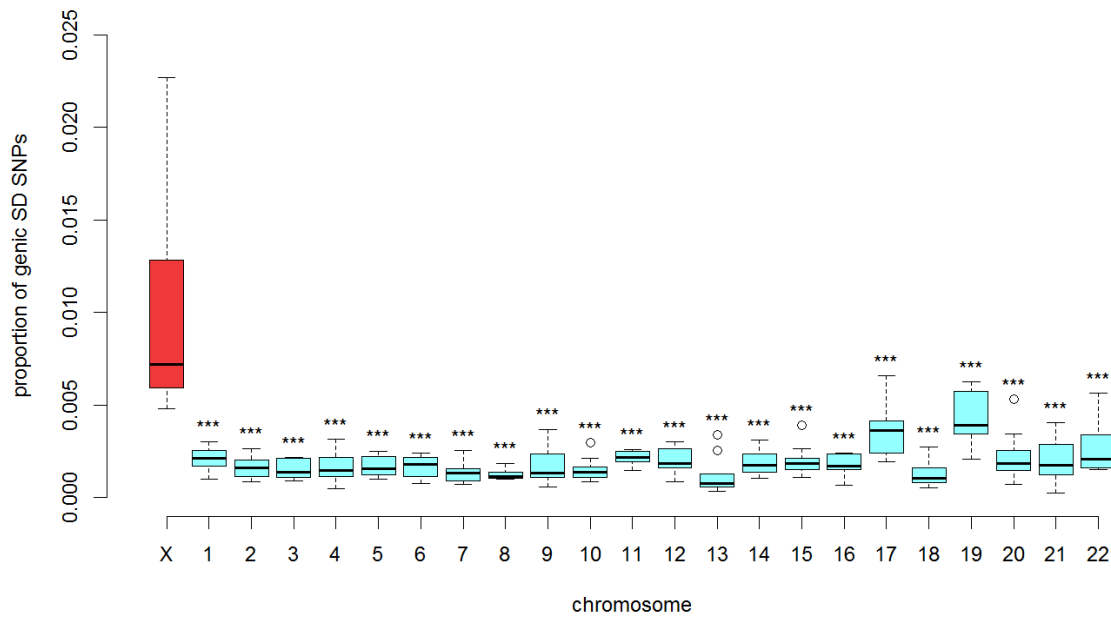


Figure S5- Concordance between the HapMap and 1000 Genomes datasets. Only genic SD SNPs showing similar frequency patterns between the two datasets are included in the analysis. The distribution of proportions of genic SD SNPs (number of genic SD SNPs after FDR correction at the gene level / number of genic SNPs) over the 10 HapMap populations is shown per chromosome. Levels of significance of one-sided Wilcoxon–Mann–Whitney tests between the mean proportions of the X chromosome and each of the autosomes are reported (p-value: * < 0.05, ** < 10^{-3} , *** < 10^{-4}).

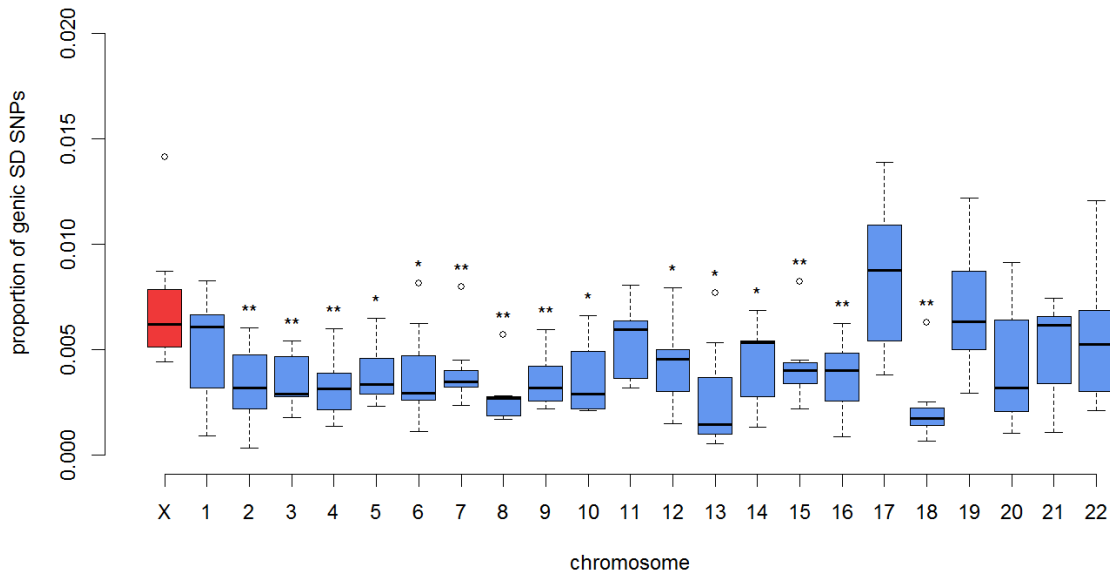


Figure S6- Influence of admixture and sample size on genic SD SNP proportions.

(A) Simulation of admixture: For each chromosome, mean proportions of genic SD SNPs per chromosome (number of genic SD SNPs / number of genic SNPs) over 10 simulated admixed populations (ADP) and proportion of genic SD SNP in ASW, YRI and CEU populations after FDR correction at the gene level. The error bars represent the 99.9% confidence intervals of the expected genic SD SNP proportion in simulated admixed populations.

(B) Resized populations: For each chromosome, mean proportions of genic SD SNPs per chromosome (number of genic SD SNPs / number of genic SNPs) in simulated resized populations (10 CEUR and 10 YRIR) and proportion of genic SD SNPs in ASW, YRI and CEU, after FDR correction at the gene level. Error bars represent the 99.9% confidence intervals of the expected genic SD SNP proportion in resized populations.

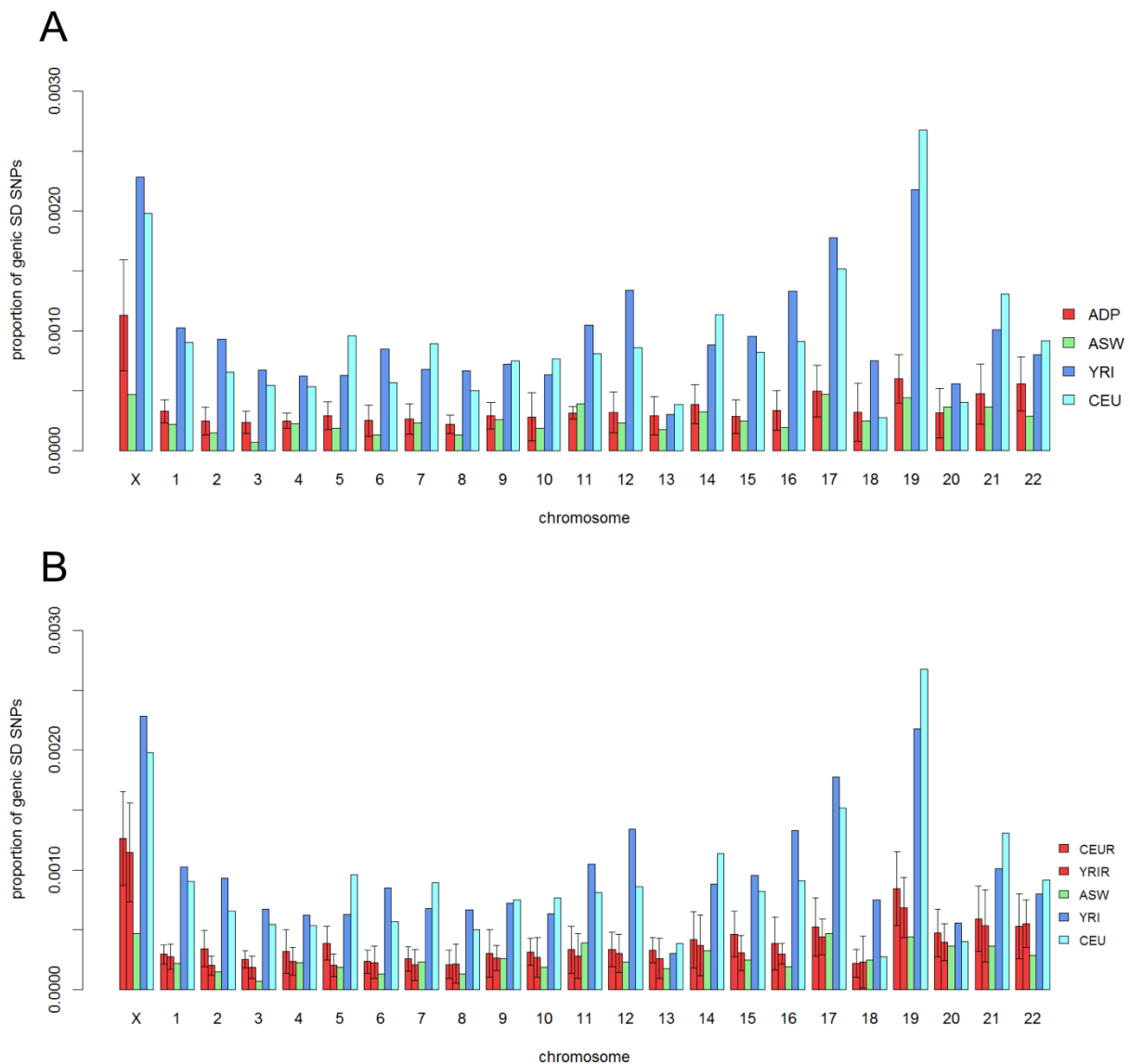


Figure S7- Proportions of genic SD SNPs per chromosome after LD pruning (filtering of genic SNPs with a $r^2 > 0.1$) **without the admixed populations ASW and MEX**. For each chromosome, the distribution of these proportions (number of genic SD SNPs after FDR correction at the gene level / number of genic SNPs) over the 10 HapMap populations is represented by a boxplot. A one-sided Wilcoxon–Mann–Whitney test was performed to compare these distributions between the chromosome X (in red) and each autosome (in blue) (p-value: * < 0.05 , ** $< 10^{-2}$, *** $< 10^{-3}$).

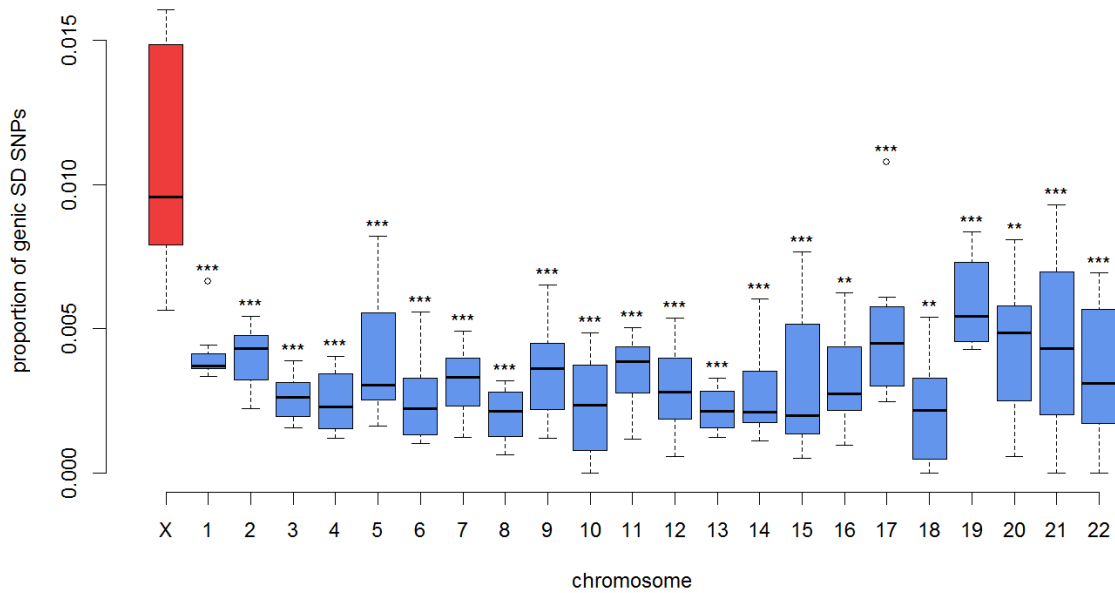
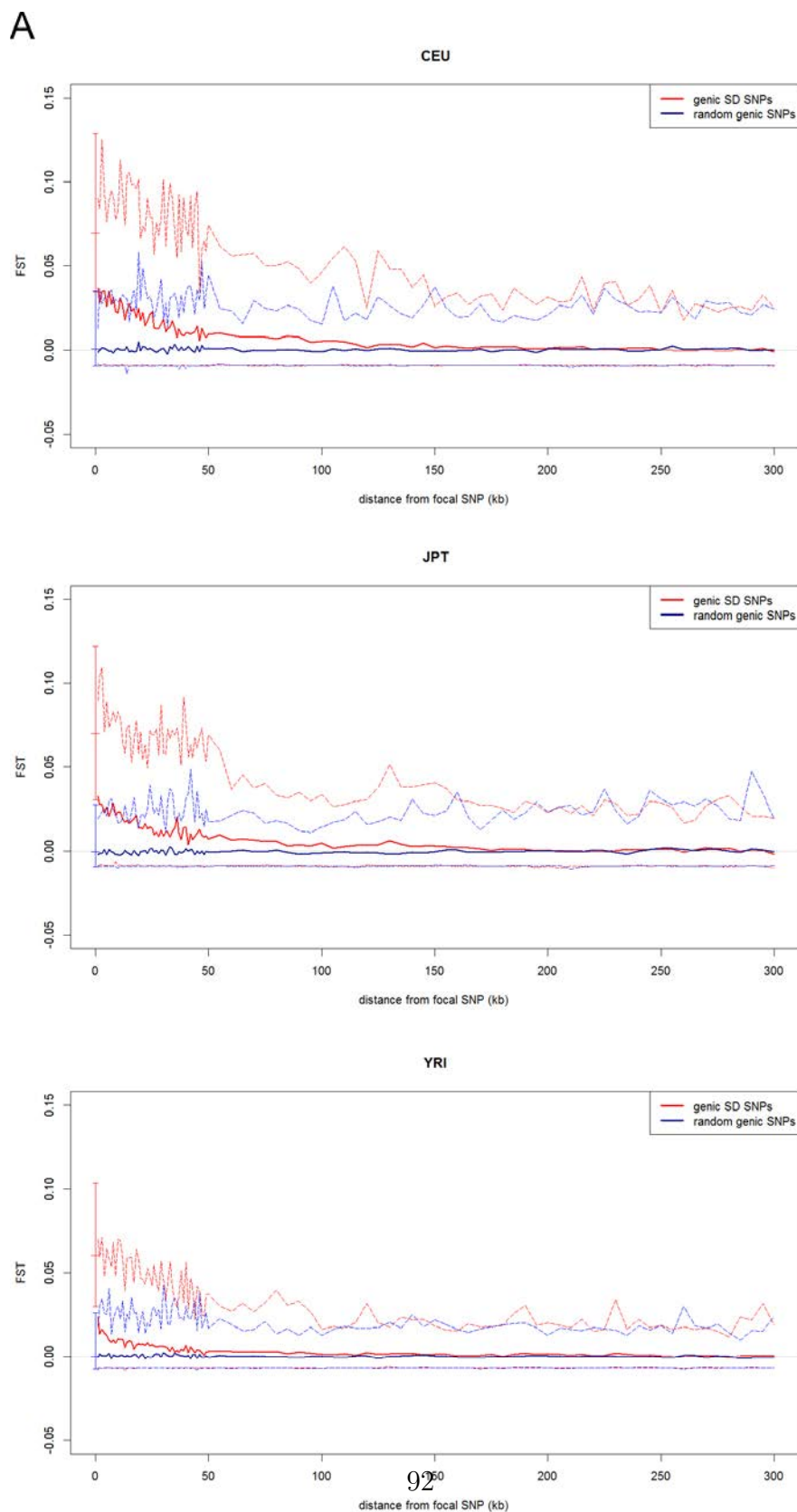


Figure S8- F_{ST} distribution in region neighboring the genic SD SNPs after LD pruning for (A) the CEU, JPT and YRI populations and (B) the admixed populations ASW and MEX. The F_{ST} were computed per SNPs, and then averaged per interval. We considered intervals of 1 kb for the first 50 kb, and then of 5 kb between 50 kb and 300 kb. The mean F_{ST} around genic SD SNPs are shown in red, the mean F_{ST} around random genic SNPs are shown in blue. The dotted lines represent the 95% quantiles of the mean F_{ST} per interval.



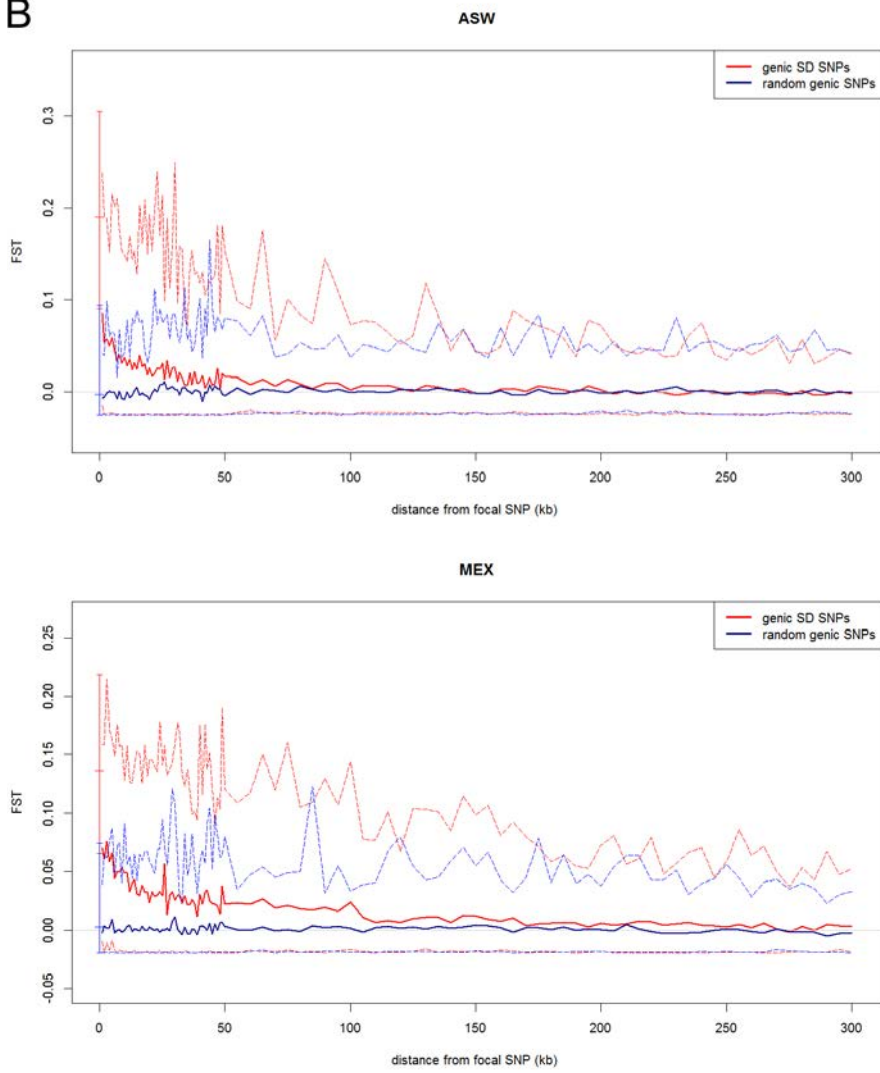
B

Figure S9- Significance of the F_{ST} signal around genic SD SNPs as compared to random genic SNPs. A one-sided Wilcoxon–Mann–Whitney test was performed for each interval to assess whether the F_{ST} signal in the vicinity of genic SD SNPs was higher than the one observed around random genic SNPs. Blue boxes and white boxes represent conditions where the p-value was significant ($p < 0.05$) and non-significant, respectively, for the interval considered. From 0kb until 50kb, the step between each interval is of 1kb, while from 50kb until 300kb, the step is of 5kb.

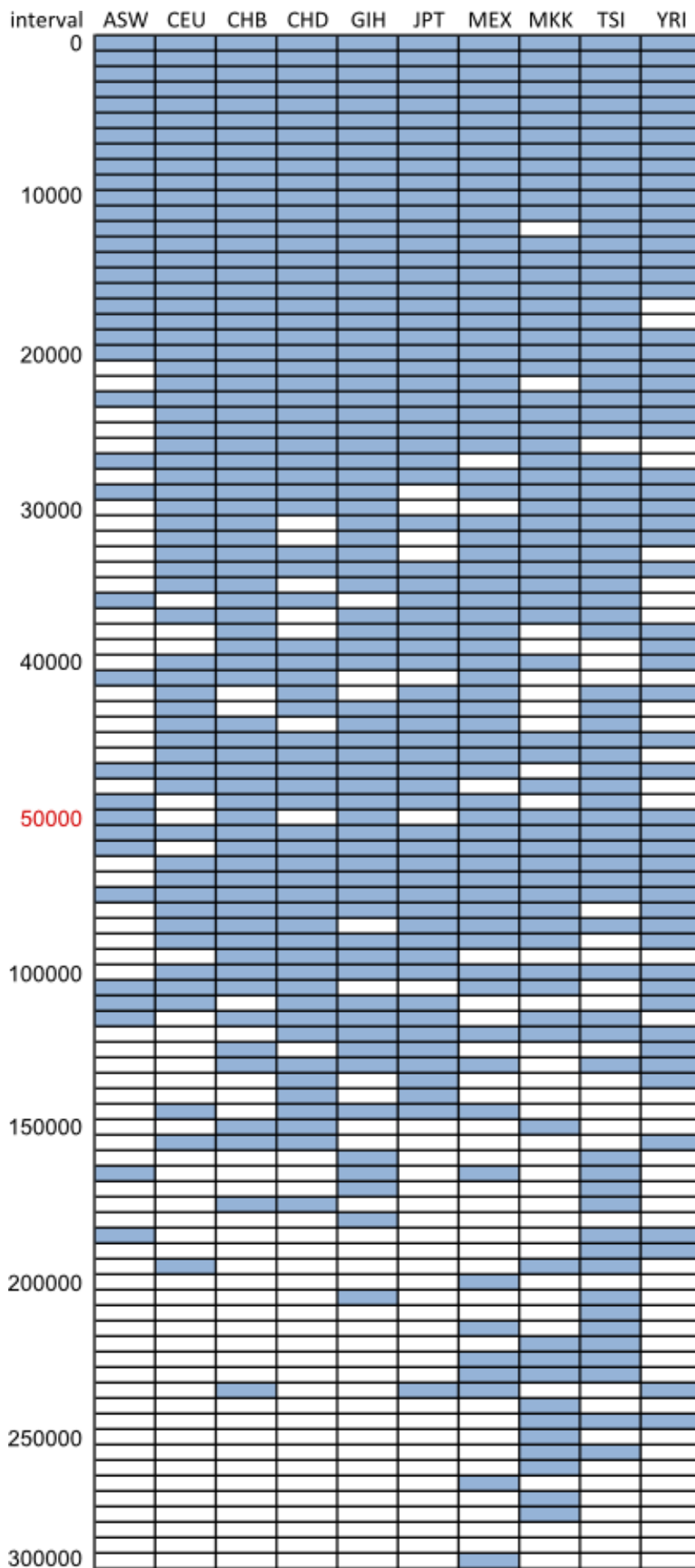
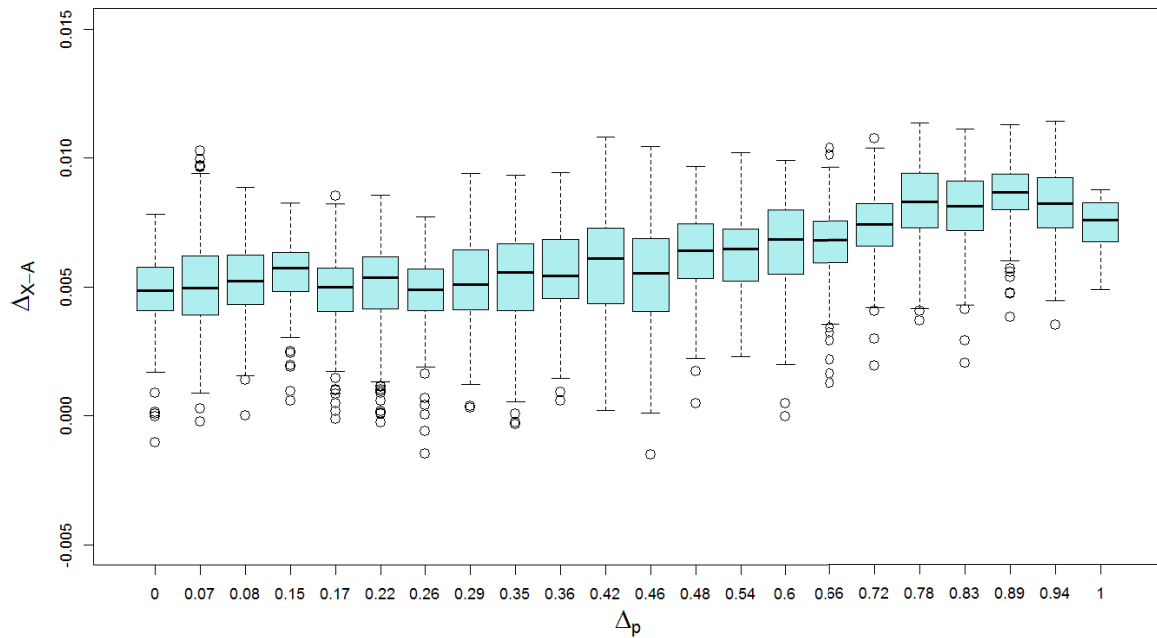


Figure S10- Effect of Δ_p on Δ_{X-A} , which is the difference between the proportion of (A) genic SD SNPs, (B) SD SNPs, on the X chromosome and on the autosomes. The 10 repetitions are pooled together. The number of runs is 22, corresponding to the number of runs for the smallest population, MKK.

A



B

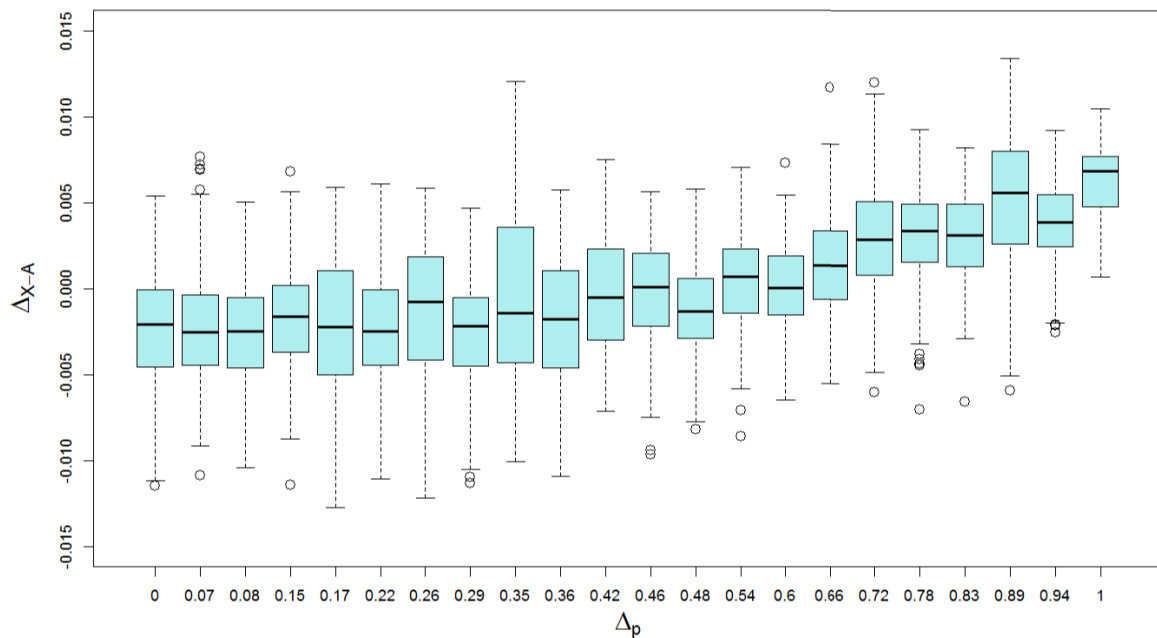


Figure S11- Effect of Δ_p on the one-sided Wilcoxon–Mann–Whitney test’s p-values between the proportion of SD SNPs (before FDR correction) of the X and those of the autosomes. The configurations where the X chromosome’s proportion is significantly higher than the autosomes’ proportions are marked in blue.

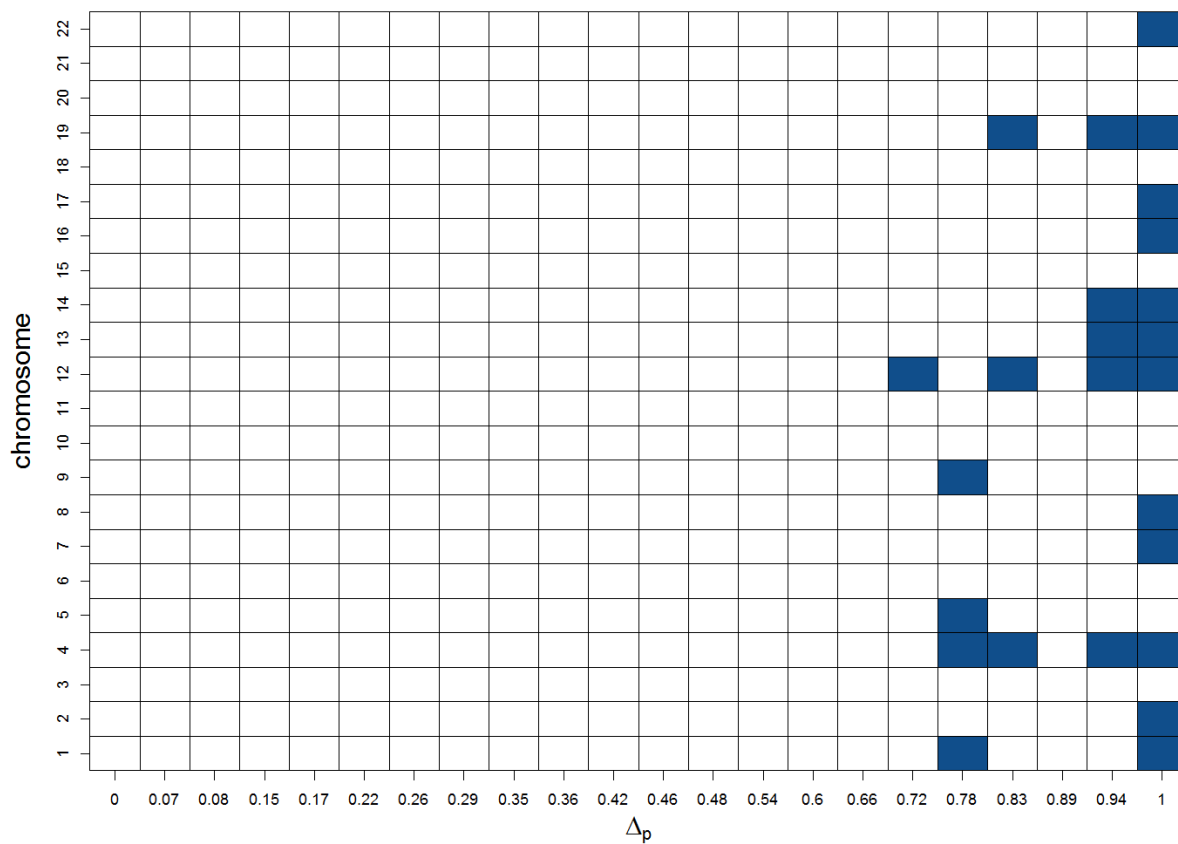
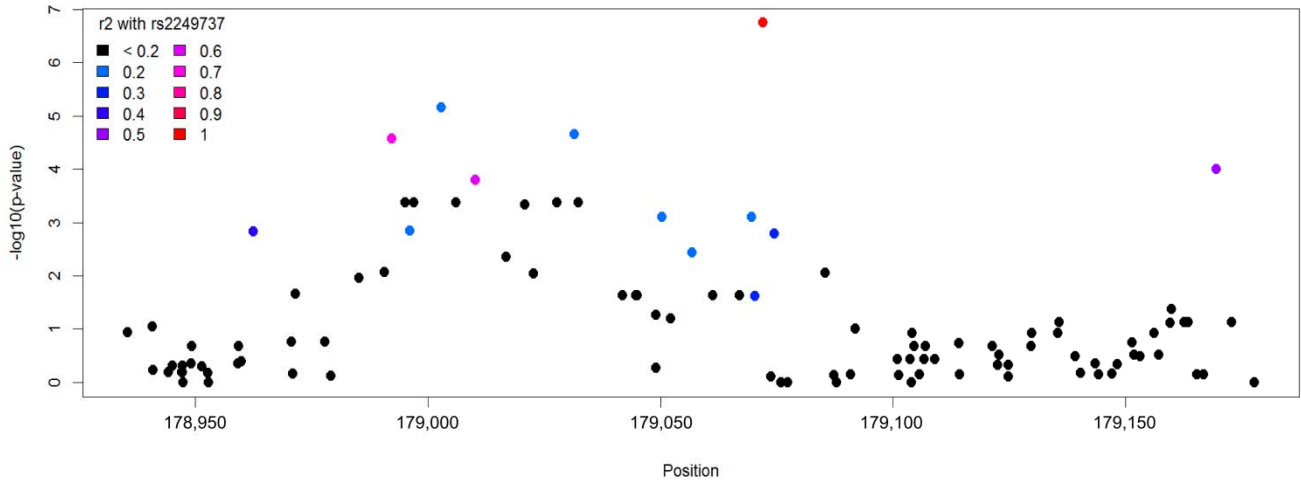


Figure S12- Functional analysis of the region surrounding chromosomal SD SNP rs2249737 (chromosome 2).

(A) F_{ST} p-values of the SNPs located in the vicinity of the chromosomal SD SNP found on chromosome 2 (rs2249737) and their Linkage Disequilibrium with this SNP (as calculated by r^2).

(B) Chromosomal features around the SNPs in LD with rs2249737 and zoom on a cluster of 4 SNPs in high LD with the focal chromosomal SD SNP. The colors of the bars representing each SNP show the LD value (r^2) calculated above. Shown are the locations of genes (RefSeq genes), DNaseI hypersensitivity clusters, transcription factor binding sites identified by ChIP-seq and chromatin state segmentation obtained by HMM. **Legend for the “DNaseI hypersensitivity clusters” track-** above the boxes are indicated the number of cell types in which a signal was found. **Legend for the “Transcription factor ChIP-seq” track-** the black and grey boxes represent two clusters of transcription factor binding sites, respectively very dense in TFBS and mildly dense in TFBS. **Legend for the “Chromatin state segmentation” track-** dark green: transcriptional elongation, light green: weakly transcribed, yellow: weak/poised enhancer, bright red: active promoter, blue: insulator. The positions are indicated in Mb and were extracted from build NCBI36/hg18.

A



B

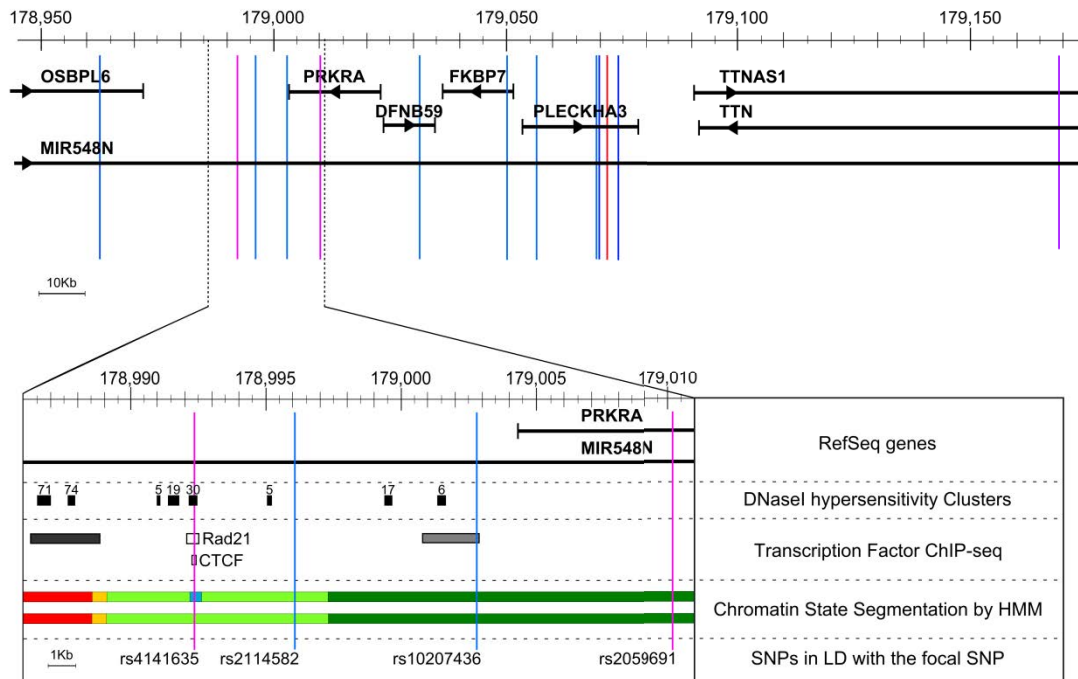


Figure S13- Analysis of the neighboring SNPs of the SD SNPs found on chromosome X and 20 with extreme signals of F_{ST} after correction for multiple testing at the genome-wide or/and chromosomal level. P-values of the F_{ST} between males and females for SNPs in the neighborhood of focal SD SNPs (in the boxes): **(A)** rs5943145 (chromosome X – GIH population), **(B)** rs7053144 (chromosome X – GIH population) and **(C)** rs1201998-rs1202030 (chromosome 20 – YRI population). Colors (from dark blue to red) indicate increasing LD (as quantified by r^2) with the focal SNP. Positions were extracted from build NCBI36/hg18.

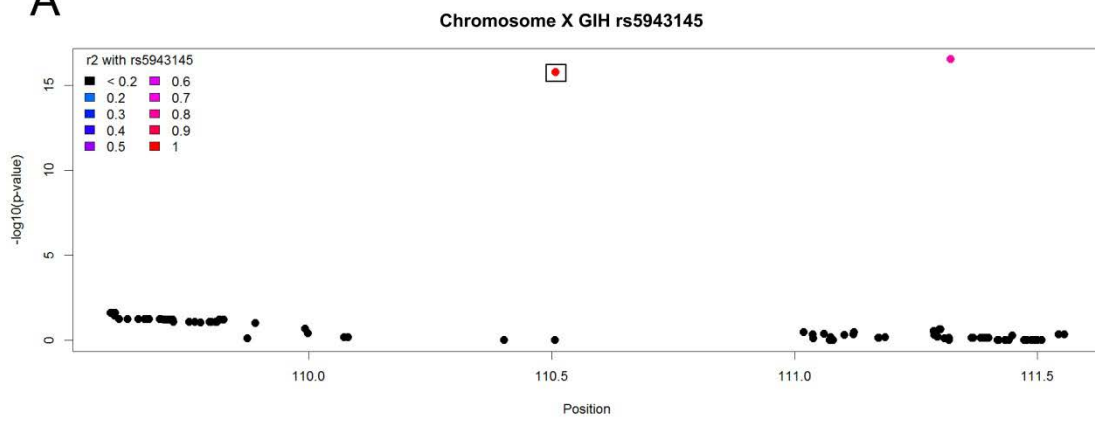
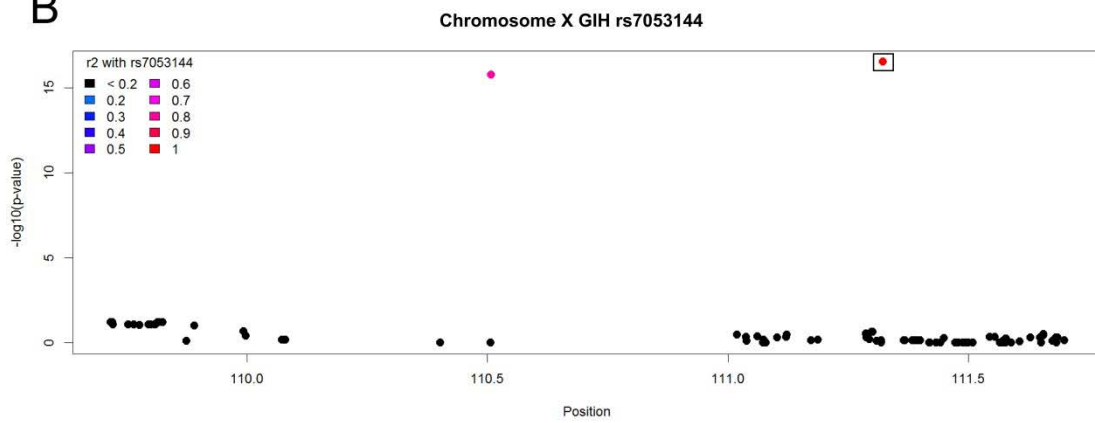
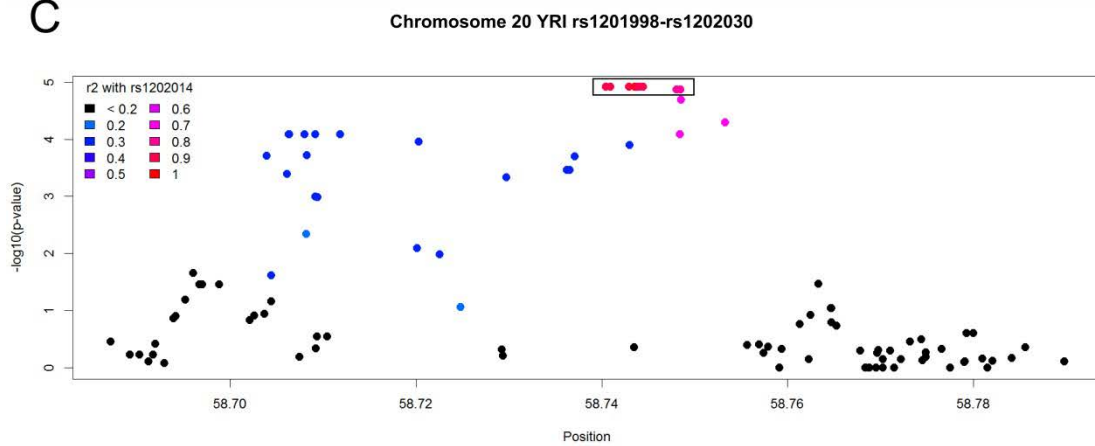
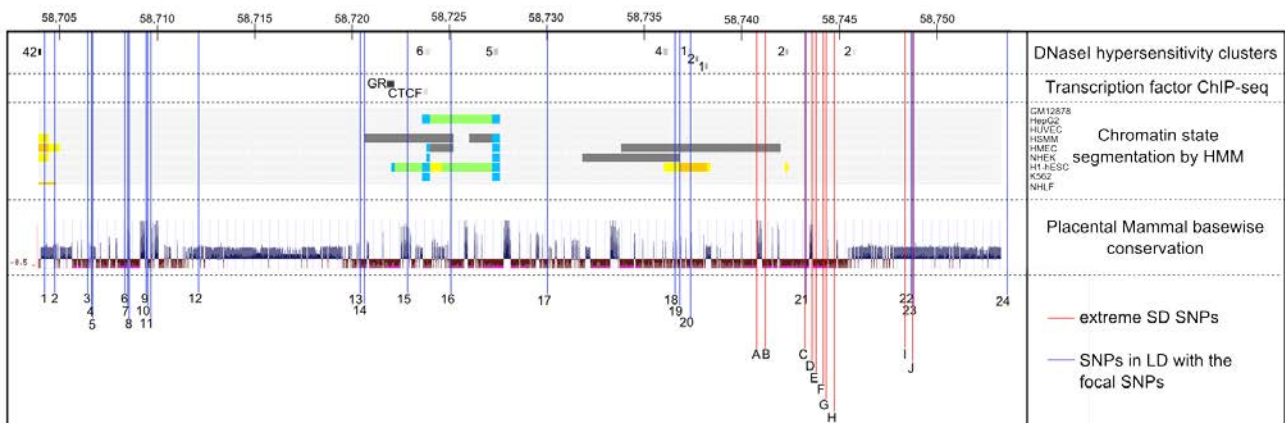
A**B****C**

Figure S14- Details on the functional analysis of the genomic regions neighboring the chromosomal SD SNPs found on chromosome 20. Functional information obtained on UCSC Genome Browser for the adjacent region of the 10 chromosomal SD SNPs found on chromosome 20. The 10 chromosomal SD SNPs are named from A to J, according to their position order (see Table 1). Similarly, the 24 SNPs in LD with them are numbered from 1 to 24 according to their position order: rs1201897, rs1201901, rs1201904, rs1201905, rs1201906, rs1201914, rs1201915, rs1201916, rs1000552, rs1000551, rs1201919, rs1201928, rs1201958, rs1201959, rs1201966, rs1201973, rs1201982, rs1201987, rs1201988, rs1201989, rs1742016, rs1202029, rs1202031, rs1202042. **Legend for the “DNaseI hypersensitivity clusters” track-** next to the clusters is indicated the number of cell types in which a signal was found. The intensity of the box (from white to black) is proportional to the ratio between the number of cell types that showed a signal and the number of cell types tested. **Legend for the “Transcription factor ChIP-seq” track-** The intensity of the box (from white to black) is proportional to the ratio between the number of cell types where the binding site was detected and the number of cell types tested. **Legend for the “Chromatin state segmentation” track-** dark green: transcriptional elongation, light green: weakly transcribed, orange: strong enhancer, yellow: weak/poised enhancer, bright red: active promoter, light red: weak promoter, blue: insulator, purple: inactive/poised promoter, grey: repressed. Positions were extracted from build NCBI36/hg18.



Supplementary Tables

Table S1. Number of individuals in the HapMap HAP1161 dataset in each population after quality control.

Population		Number of individuals		
		Total	Male	Female
ASW	African ancestry in Southwest USA	41	19	22
CEU	Utah residents with Northern and Western Europe ancestry	106	54	52
CHB	Han Chinese in Beijing, China	137	53	84
CHD	Chinese in Denver, Colorado	106	48	58
GIH	Gujarati Indians in Houston, Texas	97	56	41
JPT	Japanese in Tokyo, Japan	113	57	56
LWK	Luhya in Webuye, Kenya	99	45	54
MEX	Mexican ancestry in Los Angeles, California	53	26	27
MKK	Maasai in Kinaywa, Kenya	105	43	62
TSI	Toscans in Italia	102	53	49
YRI	Yoruba in Ibadan, Nigeria	137	68	69
TOTAL		1096	522	574

Table S2- SNPs removed from the dataset after BLAST filtering. These SNPs were removed from the dataset after performing a sequence similarity search using BLAST of the flanking sequences of autosomal SNPs (± 30 bp around the SNP) against the Y chromosome sequence. Any SNP with flanking sequences having a perfect match with Y-linked sequences was removed. The sequences were downloaded from UCSC (hg18 assembly).

Chr	SNP code	Flanking sequence -30bp	SNP	Flanking sequence +30bp
1	rs10915767	TCTCTTTTGATCTTTGTTGGTTAAAGTCT	G/A	TTTTATCAGAGACTAGGATTGCAACCCCTG
1	rs11810671	AGCTTGCCTGGCAGAAGTCTAAGCAAAGGG	G/T	CACAACACTCTTGGCAAGTTGAGAGTGGCA
1	rs12403708	AGAACTTGCCAAATATCCAAGCCCCAAAAA	C/A	CGCACCTGCTGCAATCCCCAAAAGAATAGC
1	rs12743401	AAGAAAAAATTGTCATAACCATCCATAAC	T/C	ATACTTTTCAGAAGAAATTATTAACATTCT
1	rs4658379	ATCATGTCCTTTGCAAGAATATAAATGGAG	C/A	TGCAGGCCATTATCCTTAGCAAATAATGC
1	rs7364930	CGGGAGCGAACTACCAACTTTGGGAGGTGA	A/C	ATATCTTACAGAAGGAGTGTGGGAGAAGAA
1	rs7533316	CCACACTGGGACCTTCTGGCTCCTGGATGT	G/A	ACTAGTTTCTACTTCAGGGTCTCCACATAT
1	rs7549760	ATCACAACCTTTCATAACTCTTGGTTTGGC	A/G	GGATACAACCTATCACTGATGCAGTGGTAC
1	rs805909	CCTAGGCCTTCTTTCTCTCCTATTGAATT	T/C	ACTGACCATTCTTCTATTGAATTTAAGTA
2	rs10874458	TATAACAGTCATAAGACTGTCATTCTTAGA	A/C	AGGCCTACATGCAAACTGGCCCTTTGCTG
4	rs4962914	CAATTAACAACAAAAAGAAATAAGAAAAA	A/G	TTAGCTGGGCATGGTAGTATGCATCTGTAG
4	rs6448544	GGAGCACCCAGATTATAAAGCAAGTCTCTG	C/A	GTGACCTACAAAGAGACAATGCTTATTTT
4	rs9761692	TGATTGAATTTACTTCACTTAACAAAT	A/G	TTCATTACATGCCCACTGTTTGTGAGATAT
5	rs36019094	ATAATCTTCTGATTGGATTGATCAAGAAAG	C/A	CATTAACTTGGTTGAAGAAGGTAGTCAGG
6	rs9346294	AAACTCTCAATAAATTAGGTATTGATGGGA	C/T	GTATCTCAAATAATAAGAGCTATTTATGA
7	rs10236123	AACAGATGCTGGAGAGGATGTGGAGAAATA	G/A	GAATGCTTTTACTGTTGGTGGGAGTGTA
8	rs11786942	CTCAGCTTTGTTCTTTGGCTTAGGATTGT	C/G	TTGGCAATGCAGGCTCTTTTTGGTTCAT
9	rs12551092	CATACATGCCCTGCTCTTGTTTACTGCC	A/G	GTTTACTGTTTCTCCAAGCCATCACAGC
9	rs3121249	GATGACCTCATGGGACTTGCCTTTGAGCA	G/A	TTGACACAGGAAGGGGAGGACTAGGGCCTGG
10	rs10794037	AGTAGGATCTGTGACTTGTCTAACTGAC	G/A	TAATATTGCAAATGTGATGGGTTGTTACTC
11	rs12271307	AGAAGATGGCTTCCAATAGAAACCAGTTGC	C/T	TCAAACTGAAAATAAGTTTGCAGACAGC
11	rs7294016	ATCCATGAGCATGGAATGTTTTCCATTTG	G/C	TTGTGCTCTCTTATTTCTTCTGAGCAGTG
12	rs12425187	GTTCTTCCATTTGTTTGTGCTCTTTTAT	T/G	TCATTGAGCAGTGGTTTGTAGTTCTCCTTG
12	rs9706434	TTTACAGACAAGCAAATGCTGAGAGATTTT	G/A	TCACCACCAGGCTGCCTTACAAGAGCTCC
15	rs11638630	TAAAATATCCATCACCCAGGAACTTAAC	C/T	TTGAGCACAACTCTACAACATGTTCAATG
15	rs11857958	AAACAAGCAATGGGAAAGGATTCCCTATT	T/G	AATAAATGGTGTGGGAAAAGTGGCTAGCT
15	rs150963	GGGGTTTGAAATAGACTCATACTGAATTCT	G/A	TGCTATCGTGGAGTCCATGGTGGTATGAAC
15	rs17861107	GCCACTGCACTCCAGCCTGGGTGACAGAGC	G/A	AGACTCCATCTCAAATAAAGACTGCAACT
15	rs4486824	AATTCACACATAACAATATTAACCTAAAT	G/A	TAAATGGGCTAAATGCTCCATTGTTGGTTT
16	rs391728	AGTTGAACCTTCTTTGATTGAGCAGTTT	T/G	GAAACACTCTTTTGTAGAATCTGCAAGTG
18	rs11661664	TCTCTGATGACCAGTGATGATGAGCATTTT	T/C	TCATGTGTCTGTTGGCTGCATAAATGCTCT
18	rs12967815	TAGGACCCTCCGAGCCAGGTGTGGGATATA	G/A	TCTCGTGGTGCCTGTTTTTAAGCCGGTC
18	rs4939894	GGGAGCTGTAGACCAGAGCTGTTCTATTCT	G/A	GCCATCTTGGCTCCTCGATTTCCAATTTA
19	rs35387182	TCAGCTCCCAAAGTGTGGGATTACAGGC	A/G	TGAGCCACCACATCTGCTTCTCCCAATTT
20	rs34165968	AGCTCTTAGTTTAATTAGATCCATTTGT	C/A	AATTTTGTCTTTGTTGCCATTGCTTTTGG
22	rs5993671	TATTTCTTCTTTGTGTTTATAAGTCCTTA	G/T	GATTTAGCTACCCTGATAGGTGAGAACAT

Table S3- Number of genes and SNPs that are sexually differentiated in a various number of populations (from 0 to 4 populations).

Number of populations	0	1	2	3	4	total
genes	17504	4988	672	43	2	23209
SNPs	689592	22754	812	11	0	713169

Table S4. P-values of the observed genetic distances between males and females computed on the autosomes for the 11 HapMap populations.

population	Autosomal p-values
ASW	0.84
CEU	0.77
CHB	0.61
CHD	0.35
GIH	0.57
JPT	0.41
LWK	0.03
MEX	0.78
MKK	0.82
TSI	0.39
YRI	0.77

Table S5. The p-value of the observed genetic distances between males and females on the X chromosome for the first subsampling of females (over 30) and the number of repetitions with a p-value < 5%.

population	p-value (first repetition)	Number of repetition with p-value < 5%
ASW	1	0
CEU	0.35	0
CHB	0.005	30
CHD	0.26	1
GIH	0.99	0
JPT	0.78	0
LWK	0.95	0
MEX	0.45	0
MKK	0.99	0
TSI	0.98	0
YRI	1.0	0

Table S6- List of chromosomal and genome-wide SD SNPs and comparison of their allelic frequencies in the HapMap 3.3 and the 1000 Genome (phase 1) datasets. The physical positions are taken from the human genome build NCBI36/hg18. The p-values of Fisher’s exact tests of differentiation between males and females are reported after FDR correction at the chromosomal scale: q-value (1), and at the genome-wide scale: q-value (2). Frequencies of the reference allele (allele 0) are given in males and females for the two datasets. The corresponding numbers of chromosomes are given in parenthesis. Inconsistent allelic frequencies are indicated in grey. SNPs in grey boxes were removed. For the CHD population, we compared the HapMap allelic frequencies with the 1000 Genome frequencies in CHB. SNPs with ‘suspected’ status in dbSNP were removed and are not shown in the table. For rs17315996 (in MKK), we found that CHB showed a high F_{ST} between the sexes as well and compared the allelic frequencies between the two datasets for CHB. We found that they were not concordant and therefore removed this SNP from further analysis.

Chrom	SNP	position	Pop	HapMap				1000 genomes		
				F_{ST}	q-value (1)	q-value (2)	Freq 0 Male	Freq 0 Female	Freq 0 Male	Freq 0 Female
X	rs12394407	51868390	CHB	0,39	6,1E-05	2,2E-03	0.33 (51)	0.02 (164)	0.00 (44)	0.00 (106)
			JPT	0,38	3,8E-05	1,4E-03	0.33 (55)	0.01 (110)	0.00 (50)	0.00 (78)
X	rs16980138	15711203	CHB	0,43	5,4E-06	1,5E-04	0.73 (51)	1.00 (168)	1.00 (44)	1.00 (106)
			JPT	0,40	1,7E-05	4,7E-04	0.69 (55)	1.00 (112)	1.00 (50)	1.00 (78)
X	rs10429791	10620289	CHD	0,41	7,0E-05	3,1E-03	0.29 (45)	0.00 (116)	0.00 (44)*	0.01 (106)*
X	rs2961384	40217230	CHD	0,41	7,0E-05	3,1E-03	0.67 (46)	0.99 (116)	0.95 (44)*	0.99 (106)*
X	rs5973231	34636210	CHD	0,71	1,5E-15	4,2E-14	0.40 (48)	1.00 (116)	1.00 (44)*	1.00 (106)*
X	rs5973480	35717997	CHD	0,71	1,6E-14	5,9E-13	0.59 (44)	0.00 (116)	0.00 (44)*	0.00 (106)*
X	rs5943145	110508468	GIH	0,63	1,8E-12	6,4E-11	0.59 (54)	0.00 (78)	NA	NA
X	rs7053144	111322009	GIH	0,64	6,3E-13	2,6E-11	0.39 (56)	1.00 (78)	NA	NA
X	rs2522993	16157942	LWK	0,33	2,9E-02	6,8E-01	0.23 (43)	0.00 (106)	0.00 (48)	0.00 (98)
X	rs7059594	56060125	LWK	0,27	2,9E-02	6,8E-01	0.84 (45)	0.43 (104)	0.83 (48)	0.76 (98)
X	rs17315996	41631945	MKK	0,44	3,6E-04	8,6E-03	0.64 (42)	0.98 (122)	NA	NA
2	rs2249737	179071931	ASW	0,45	1,8E-02	1,6E-01	0.45 (38)	0.00 (44)	0.40 (48)	0.11 (74)
3	rs6442608	16563277	YRI	0,20	1,9E-03	1,4E-02	0.47 (136)	0.80 (138)	0.76 (86)	0.80 (90)
20	rs1201998	58740471	YRI	0,13	4,1E-02	7,3E-01	0.38 (136)	0.64 (138)	0.36 (86)	0.64 (90)
20	rs1202001	58740936	YRI	0,13	4,1E-02	7,3E-01	0.38 (136)	0.64 (138)	0.36 (86)	0.64 (90)
20	rs1202009	58742939	YRI	0,13	4,1E-02	7,3E-01	0.63 (136)	0.36 (138)	0.64 (86)	0.36 (90)
20	rs1202013	58743543	YRI	0,13	4,1E-02	7,3E-01	0.38 (136)	0.64 (138)	0.36 (86)	0.64 (90)
20	rs1202014	58743557	YRI	0,13	4,1E-02	7,3E-01	0.63 (136)	0.36 (138)	0.64 (86)	0.36 (90)
20	rs1202016	58743888	YRI	0,13	4,1E-02	7,3E-01	0.38 (136)	0.64 (138)	0.36 (86)	0.64 (90)
20	rs1202019	58744029	YRI	0,13	4,1E-02	7,3E-01	0.63 (136)	0.36 (138)	0.64 (86)	0.36 (90)
20	rs1202021	58744481	YRI	0,13	4,1E-02	7,3E-01	0.63 (136)	0.36 (138)	0.64 (86)	0.36 (90)
20	rs1202027	58748073	YRI	0,12	4,1E-02	7,3E-01	0.66 (136)	0.40 (138)	0.67 (86)	0.42 (90)
20	rs1202030	58748486	YRI	0,12	4,1E-02	7,3E-01	0.34 (136)	0.60 (138)	0.33 (86)	0.58 (90)

*Data extracted from CHB population

Table S7- Enriched terms and their p-values in each manually curated category for the three lists of SD genes. Eight categories were created (in bold). The terms that could not be classified into these categories were labeled “other”. The p-values correspond to the significance of the functional enrichment for the corresponding term.

Enriched term	p-value		
	SDG	SDG ₃₀₀₀	SDG ₁₀₀₀
EPIDERMIS DEVELOPMENT			
cornified envelope	1.92E-02	1.56E-02	1.26E-02
ectoderm development	1.39E-02	2.49E-03	3.21E-03
epidermal cell differentiation	3.99E-02	2.57E-03	6.39E-04
epidermis development	8.82E-03	3.66E-03	1.40E-03
hair follicle morphogenesis			4.01E-02
keratin filament			1.18E-02
keratinization	2.41E-03	4.71E-04	2.96E-04
keratinocyte differentiation	2.64E-02	2.38E-03	3.13E-04
DEVELOPMENTAL PROCESS			
angiogenesis	3.54E-02		
atrial cardiac muscle morphogenesis	4.36E-02		
cardiac muscle tissue morphogenesis	4.70E-03		
developmental cell growth		3.83E-02	
endothelial cell differentiation	1.04E-03		
epithelial cell development	9.97E-04		
epithelial cell differentiation	1.21E-04	6.11E-04	3.90E-03
epithelial to mesenchymal transition	4.24E-02		
epithelium development	5.04E-03	3.40E-02	4.50E-02
forebrain development		3.49E-02	
heart morphogenesis	3.51E-03		
hippocampus development		4.72E-02	
lens morphogenesis in camera-type eye	2.65E-02	3.83E-02	
limbic system development		3.84E-02	
mesoderm formation		2.89E-02	
mesoderm morphogenesis		3.99E-02	
midbrain development		3.98E-02	
muscle tissue morphogenesis	4.70E-03		
odontogenesis	3.73E-02		
odontogenesis of dentine-containing tooth	2.77E-02		
positive regulation of axonogenesis	2.00E-02		
regulation of cell growth	1.89E-03		
regulation of growth	4.35E-03		
sprouting angiogenesis			4.93E-02
tissue morphogenesis	1.99E-02		3.27E-02
ventricular cardiac muscle morphogenesis	2.03E-02		
SUGAR METABOLISM AND ENERGY			
carbohydrate binding	2.93E-03	3.81E-03	2.23E-02
carbohydrate kinase activity	2.57E-03	1.89E-03	
galactose binding	4.42E-02		
glucose transmembrane transporter activity	4.43E-02		

hexose metabolic process		2.07E-02	
hexose transmembrane transporter activity	4.36E-02		
monosaccharide binding	1.65E-02		
monosaccharide metabolic process		1.58E-02	
oligosaccharide metabolic process		3.22E-02	
regulation of carbohydrate catabolic process	5.05E-03	3.98E-02	
regulation of carbohydrate metabolic process	4.76E-03		
regulation of catabolic process	6.17E-03	1.35E-02	4.18E-02
regulation of cellular carbohydrate catabolic process	5.05E-03		
regulation of cellular carbohydrate metabolic process	3.43E-03	3.98E-02	
regulation of cellular catabolic process	4.54E-03		3.72E-02
regulation of glucose metabolic process	9.39E-03		
regulation of glycolysis	4.25E-02		
regulation of insulin secretion	1.58E-02		
sugar binding	2.22E-02	1.79E-02	
REGULATION OF GENE EXPRESSION	SDG	SDG₃₀₀₀	SDG₁₀₀₀
gene silencing by miRNA		1.34E-02	
gene silencing by RNA		1.66E-02	4.38E-02
negative regulation of gene expression		1.22E-02	
negative regulation of gene-specific transcription	2.21E-02		4.47E-02
positive regulation of transcription factor activity	7.79E-03		
posttranscriptional gene silencing		1.09E-02	
posttranscriptional gene silencing by RNA		1.09E-02	
regulation of gene-specific transcription	2.54E-02	4.11E-02	
regulation of RNA metabolic process	3.73E-02		
regulation of transcription	3.74E-02		
regulation of transcription factor activity	1.13E-02		
rRNA processing	3.18E-02		
transcription factor activity	4.73E-02		
transcription repressor activity		3.46E-02	
MUSCLE	SDG	SDG₃₀₀₀	SDG₁₀₀₀
muscle adaptation	4.25E-02		
positive regulation of muscle contraction	1.40E-02		
positive regulation of smooth muscle contraction	4.24E-02		
regulation of smooth muscle contraction	2.17E-02		
IMMUNE SYSTEM	SDG	SDG₃₀₀₀	SDG₁₀₀₀
antigen processing and presentation of peptide antigen	4.77E-02		
leukocyte migration	3.42E-02		
positive regulation of innate immune response			3.82E-02
regulation of lymphocyte differentiation			1.38E-02
regulation of inflammatory response	4.26E-02		
regulation of interleukin-1 beta production	2.93E-02		
regulation of T cell activation			4.93E-02
regulation of T cell differentiation		4.61E-02	1.69E-02
regulation of T cell proliferation			3.99E-02
response to virus	4.38E-02		
RESPONSE TO STIMULUS	SDG	SDG₃₀₀₀	SDG₁₀₀₀
detection of biotic stimulus	4.24E-02		

positive regulation of cell communication	4.05E-03		
positive regulation of response to external stimulus	1.89E-03	1.66E-03	4.28E-03
positive regulation of response to stimulus	6.13E-04	6.24E-03	3.68E-03
positive regulation of signal transduction	2.63E-03		
regulation of response to external stimulus	1.49E-02		
response to corticosteroid stimulus	3.27E-02	1.57E-02	
response to glucocorticoid stimulus	1.87E-02	6.40E-03	
REPRODUCTION	SDG	SDG₃₀₀₀	SDG₁₀₀₀
female gonad development		4.92E-02	
gamete generation			1.11E-02
gonad development		3.48E-02	
male gamete generation			4.80E-02
multicellular organism reproduction		2.07E-02	1.64E-02
ovulation cycle process		3.93E-02	
reproductive developmental process		2.29E-02	
reproductive process in a multicellular organism		2.07E-02	1.64E-02
reproductive structure development		3.71E-02	
sexual reproduction			1.24E-02
spermatogenesis			4.80E-02
OTHER	SDG	SDG₃₀₀₀	SDG₁₀₀₀
3'-phosphoadenosine 5'-phosphosulfate binding	3.16E-02	2.94E-02	
acyl-CoA dehydrogenase activity	4.43E-02		
adenyl nucleotide binding		1.76E-02	2.26E-02
adenyl ribonucleotide binding		1.80E-02	2.56E-02
asymmetric protein localization	4.25E-02		
ATPase activity	2.68E-02		
ATPase activity, coupled	1.83E-02		
ATPase activity, coupled to movement of substances	2.05E-02		
ATPase activity, coupled to transmembrane movement of ions	3.30E-02		
ATPase activity, coupled to transmembrane movement of substances	1.78E-02		
ATP binding		2.87E-02	3.55E-02
behavior		1.76E-02	
biogenic amine catabolic process	4.31E-02		
BMP signaling pathway	1.89E-02		
calcium-mediated signaling		1.23E-03	
calcium-transporting ATPase activity	4.01E-02		
cation-transporting ATPase activity	3.11E-02		
cell migration	4.28E-02		
cell motility	4.25E-02		
cell soma		4.96E-02	
cGMP-mediated signaling	3.92E-02		
coreceptor activity	3.02E-02		
cysteine metabolic process	3.10E-02		
cytolysis	4.08E-02	3.82E-02	
cytosolic part	3.20E-02		
dsRNA fragmentation			1.42E-02
eating behavior		1.84E-02	

endosome transport	2.14E-02	2.38E-02	
exonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters	3.02E-02		
external side of plasma membrane	4.29E-02		
feeding behavior		4.27E-02	
galactose binding	4.42E-02		
glycosaminoglycan binding	3.60E-02		
growth factor binding	4.49E-02		
heterocycle catabolic process	1.57E-02		
hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	1.41E-02		
insulin-like growth factor I binding		2.94E-02	
integral to plasma membrane	2.05E-02		
intracellular signaling cascade		2.93E-02	
intrinsic to plasma membrane	3.71E-02		
localization of cell	4.25E-02		
lysosomal transport		2.89E-02	
microtubule cytoskeleton organization		2.38E-02	
modification by symbiont of host morphology or physiology	3.92E-02		
nucleoside binding		3.35E-02	2.69E-02
nucleotide diphosphatase activity			4.07E-02
nucleoside diphosphate biosynthetic process		4.57E-02	
nucleotide binding		1.22E-02	
NADPH oxidase complex	3.80E-02		
neuron projection	4.56E-02		
organ regeneration		3.54E-02	
organelle localization			3.37E-02
palmitoyltransferase activity			1.83E-02
peptide hormone receptor binding			4.68E-02
peptide transport	3.73E-02		
peptidoglycan binding	7.70E-03		
phosphoinositide-mediated signaling		4.19E-02	
phosphoprotein binding	3.79E-02		
polyamine metabolic process	2.65E-02		
positive regulation of axon extension	3.92E-02		
positive regulation of behavior	1.77E-02	2.02E-02	4.81E-02
positive regulation of binding	3.35E-02		
positive regulation of cell cycle	1.75E-02		
positive regulation of cell migration	3.47E-02		
positive regulation of cell motion	2.68E-02		
positive regulation of cell projection organization	3.66E-02		
positive regulation of cellular amine metabolic process		4.26E-02	
positive regulation of cellular component organization	4.69E-02		
positive regulation of chemotaxis			2.88E-02
positive regulation of cytokine production	2.36E-02		
positive regulation of DNA binding	7.82E-03		
positive regulation of homeostatic process	3.49E-02		
positive regulation of locomotion	1.57E-02		4.63E-02
positive regulation of MAPKKK cascade	4.45E-02		

positive regulation of multicellular organismal process	1.12E-03		4.17E-02
positive regulation of NF-kappaB transcription factor activity	4.34E-02		
positive regulation of organelle organization	4.13E-02	4.82E-02	
positive regulation of oxidoreductase activity	9.87E-03		
positive regulation of positive chemotaxis			4.01E-02
positive regulation of protein kinase cascade	3.63E-02		
positive regulation of steroid metabolic process	2.88E-02		
P-P-bond-hydrolysis-driven transmembrane transporter activity	1.48E-02		
primary active transmembrane transporter activity	1.48E-02		
protein binding, bridging	4.46E-02		
protein heterooligomerization	4.90E-02		
protein phosphorylated amino acid binding	4.35E-02		
protein tyrosine kinase activity			2.07E-02
proteoglycan biosynthetic process	9.87E-03		
purine nucleoside binding		3.08E-02	3.10E-02
purine nucleotide binding		3.26E-02	
purine ribonucleotide binding		3.65E-02	
purine ribonucleotide catabolic process	2.85E-02		
receptor complex	2.34E-02		
regeneration		4.27E-02	
regulation of binding	6.82E-03		
regulation of cell cycle process	3.26E-02		
regulation of cell migration	8.38E-03		
regulation of cell motion	1.07E-02		
regulation of cell size	1.96E-02		
regulation of cellular ketone metabolic process			2.78E-02
regulation of chemotaxis			3.59E-02
regulation of DNA binding	1.17E-02		
regulation of excretion	3.92E-02		
regulation of generation of precursor metabolites and energy	3.43E-02		
regulation of homeostatic process	2.01E-02		
regulation of hormone secretion	4.77E-02		
regulation of locomotion	6.12E-03		
regulation of multi-organism process	4.42E-02		
regulation of neuron apoptosis	1.37E-02	1.37E-02	
regulation of organelle organization	2.74E-02	2.81E-02	
regulation of oxidoreductase activity	1.58E-02		
regulation of positive chemotaxis			4.01E-02
regulation of steroid metabolic process	2.98E-02	2.89E-02	
regulation of system process	7.46E-03		
response to activity	4.25E-02	1.10E-02	
response to axon injury	2.93E-02	3.04E-02	
response to estrogen stimulus			2.73E-02
response to ethanol	1.82E-02		
response to inorganic substance		2.00E-02	
response to metal ion	7.89E-03	1.55E-02	
response to oxygen radical		3.83E-02	
response to superoxide		2.70E-02	

ribonucleotide binding		3.65E-02	
ribonucleoprotein complex	2.82E-03		
RNA binding		3.05E-02	
sodium channel activity		4.02E-02	
STAT protein nuclear translocation	4.36E-02		
sulfur compound biosynthetic process	3.36E-02		
superoxide metabolic process	6.75E-03		
transmembrane transport	4.69E-02		
tRNA (guanine) methyltransferase activity		1.62E-02	
vacuolar transport		3.84E-02	

Table S8- Enriched terms and their p-values in each functional gene cluster for SDG₃₀₀₀, SDG₁₀₀₀ and SDG_{sh}. Details of the functional enrichment analysis results from the clustering tool performed on DAVID for the 3000 and 1000 SD genes with the lowest p-values (SDG₃₀₀₀ and SDG₁₀₀₀) and the SD genes shared by at least two populations (SDG_{sh}).

	SDG ₃₀₀₀	SDG ₁₀₀₀	SDG _{sh}
Enriched term	p-value		
EPIDERMIS DEVELOPMENT			
Enrichment Score	2.51	2.65	
Cornified envelope	1.56E-02	1.26E-02	
Ectoderm development	2.49E-03	3.21E-03	
Epidermal cell differentiation	2.57E-03	6.39E-04	
Epidermis development	3.66E-03	1.40E-03	
Epithelial cell differentiation	6.11E-04	3.90E-03	
Epithelium development	3.40E-02	4.50E-02	
Keratinization	4.71E-04	4.50E-02	
Keratinocyte differentiation	2.38E-03	3.13E-04	
RESPONSE TO STIMULUS AND IMMUNE SYSTEM			
Enrichment Score	1.38	1.37	
Positive regulation of defense response	1.15E-01		
Positive regulation of inflammatory response	1.62E-01		
Positive regulation of response to external stimulus	1.66E-03	4.28E-03	
Positive regulation of response to stimulus	6.24E-03		
Regulation of inflammatory response	4.76E-01		
Regulation of response to external stimulus	5.66E-02	6.04E-02	
Positive regulation of behavior		4.81E-02	
Positive regulation of chemotaxis		2.88E-02	
Positive regulation of locomotion		4.63E-02	
Positive regulation of positive chemotaxis		4.01E-02	
Regulation of behavior		1.16E-01	
Regulation of chemotaxis		3.59E-02	
Regulation of locomotion		1.61E-01	
Regulation of positive chemotaxis		4.01E-02	
BINDING TO NUCLEOSIDE, NUCLEOTIDE AND RIBO-NUCLEOTIDE			
Enrichment Score	1.59		
Adenyl nucleotide binding	1.76E-02		
Adenyl ribonucleotide binding	1.80E-02		
ATP binding	2.87E-02		
Nucleoside binding	3.35E-02		
Nucleotide binding	1.22E-02		
Purine nucleoside binding	3.08E-02		
Purine nucleotide binding	3.26E-02		
Purine ribonucleotide binding	3.65E-02		
Ribonucleotide binding	3.65E-02		
GLYCOLYSIS			

Enrichment Score		1.43	
Carbohydrate kinase activity	1.89E-03		
Glycolysis	2.24E-01		
Hexokinase activity	1.22E-01		
REPRODUCTIVE PROCESS			
Enrichment Score		1.68	
Gamete generation		1.11E-02	
Male gamete generation		4.80E-02	
Multicellular organism reproduction		1.64E-02	
Reproductive process in a multicellular organism		1.64E-02	
Sexual reproduction		1.24E-02	
Spermatogenesis		4.80E-02	
FILAMENT CYTOSKELETON			
Enrichment Score		1.33	
Intermediate filament		8.74E-02	
Intermediate filament cytoskeleton		9.86E-02	
Keratin filament		1.18E-02	
VACUOLE			
Enrichment Score		1.55	
Vacuolar membrane			5.85E-03
Vacuolar part			1.11E-02
Tubulin binding			7.59E-02
Vacuole			1.25E-01
ODONTOGENESIS			
Enrichment Score		1.48	
Odontogenesis of dentine-containing tooth			5.72E-03
Odontogenesis			2.19E-02
Tissue morphogenesis			2.90E-01
MICROTUBULE AND CYTOSKELETON			
Enrichment Score		1.46	
Microtubule			7.21E-03
Microtubule cytoskeleton organization			1.24E-02
Microtubule-based process			1.92E-02
Microtubule cytoskeleton			8.81E-02
Cytoskeleton organization			3.15E-01
PEPTIDASE INHIBITOR ACTIVITY			
Enrichment Score		1.42	
Endopeptidase inhibitor activity			2.96E-02
Enzyme inhibitor activity			3.16E-02
Peptidase inhibitor activity			3.97E-02
Serine-type endopeptidase inhibitor activity			5.67E-02

Chapitre 3

Distorsion de transmission sexe-spécifique

3.1 Introduction

Les lois de Mendel prédisent que les allèles des deux parents sont chacun transmis avec une probabilité de 0,5 à leurs descendants. La distorsion de transmission est un phénomène de déviation de ces lois, et se produit si un des allèles des parents est transmis de façon préférentielle à ses descendants. La distorsion de transmission peut émerger de plusieurs mécanismes biologiques se déroulant entre la production des gamètes et le développement embryonnaire.

1. **La sélection de la lignée germinale** : au cours de la production de la lignée germinale par mitose, divers phénomènes comme des mutations, des recombinaisons et des conversions géniques peuvent mener des cellules avec un certain génotype à être représentées en plus grand nombre que les autres. Néanmoins, un allèle apportant un avantage sélectif à une cellule au sein de la lignée germinale pourrait être désavantageux chez l'adulte (Hastings 1989). Par exemple, chez l'Homme, plusieurs études ont montré que des mutations *de novo* provoquant le syndrome d'Apert chez les enfants, qui engendre d'importantes déformations crâniennes et un retard mental, seraient sélectionnées positivement dans la lignée germinale des pères (Goriely *et al.* 2005 ; Choi *et al.* 2008). Ces mutations dominantes sont responsables d'un gain de fonction de la protéine qui conférerait un avantage au niveau de l'expansion clonale au sein de la lignée germinale. En conséquence, cette maladie est observée beaucoup plus souvent qu'attendu si cet allèle était représenté dans la lignée germinale uniquement en fonction du taux de mutation *de novo* (Goriely *et al.* 2005).

2. **La distorsion de ségrégation** : pendant la méiose femelle, l'ovogénèse, ou la méiose mâle, la spermatogénèse, une distorsion de ségrégation a lieu si des gamètes avec un certain génotype sont produits en plus grandes proportions que les autres. L'ovogénèse étant asymétrique, une seule chromatide sur les quatre devient un gamète fonctionnel tandis que les autres deviennent des globules polaires et sont détruits. Un haplotype ayant un avantage durant la méiose sera représenté en plus grande proportion que les autres dans les ovocytes. Cette asymétrie de la méiose chez les femelles implique qu'une distorsion de ségrégation est plus facilement réalisable que chez les mâles (Huang *et al.* 2013). Un exemple très étudié de distorsion de ségrégation chez le mâle, appelé *SD* pour « Segregation Distorter », est trouvé chez les populations sauvages de *Drosophila melanogaster* (voir Larracuente et Presgraves 2012 pour une revue). Le complexe génique *SD*, situé sur un autosome, provoque une perturbation lors de l'empaquetage de la chromatine des spermatides *SD+* pendant la spermatogénèse, les rendant non fonctionnels. En conséquence, les mâles hétérozygotes *SD/SD+* ont presque exclusivement des descendants *SD*. Néanmoins, les homozygotes *SD/SD* n'étant pas viables ou étant stériles, la fréquence d'équilibre de *SD* dans la population est limitée à moins de 5%.

La distorsion de ségrégation des chromosomes sexuels est un autre type de distorsion de ségrégation qui peut survenir lors de la spermatogénèse, et mener à une sur-représentation de spermatozoïdes portant un chromosome X (ou un chromosome Y) particulier. La distorsion de ségrégation d'un chromosome sexuel s'accompagne donc d'une distorsion de sexe-ratio car le sexe de la descendance est biaisé. L'exemple le plus frappant de ce type de distorsion a été observé chez plusieurs espèces de drosophiles. Ces distorsions de transmission peuvent être très fortes et mener à une composition exclusivement femelle de la descendance. Elles sont souvent associées à un dysfonctionnement du chromosome Y pendant la deuxième phase de la méiose qui cause une mauvaise ségrégation des chromatides sœurs du Y (voir Helleu *et al.* 2015 pour une revue).

3. **La sélection gamétique** : la compétition entre les spermatozoïdes peut mener des gamètes porteurs d'un certain génotype à être plus souvent impliqués dans la fécondation. Ce phénomène a une composante pré-copulatoire et une composante post-copulatoire. Une survie différentielle entre les spermatozoïdes selon leur génotype constitue la composante pré-copulatoire de la sélection gamétique. Après la copulation, les spermatozoïdes se déplacent vers l'ovule afin de le féconder. Les spermatozoïdes les plus rapides ou ceux qui se dirigent le mieux vers l'ovule se-

ront avantagés. Quand les spermatozoïdes ont rejoint l'ovule, il est nécessaire qu'ils perforent la membrane entourant l'ovule, la zone pellucide, afin de le féconder. Ce mécanisme met en jeu des systèmes de reconnaissance de protéines membranaires qui influence quel spermatozoïde féconde l'ovule. Un exemple connu de sélection gamétique post-copulatoire est l'haplotype t chez la souris. L'haplotype t est un variant d'une région du chromosome 17 et contient des gènes causant une distorsion de transmission, une infertilité chez les mâles, et une létalité des embryons (voir Lyon 2003 pour une revue). Les mâles possédant cet haplotype à l'état homozygote (t/t) sont stériles tandis que les mâles hétérozygotes ($+/t$) sont fertiles et transmettent le chromosome porteur de l'haplotype t à plus de 50% (99% au maximum) de leur descendance. De plus, on observe une suppression de la recombinaison au niveau de la région déterminante de l'haplotype t . Cette distorsion de transmission serait due à une diminution de la mobilité des spermatozoïdes porteurs de l'haplotype $+$ chez les mâles hétérozygotes par rapport aux spermatozoïdes porteurs de l'haplotype t (Schimenti 2000).

4. **La sélection post-fécondation** : une mortalité différentielle selon le génotype des embryons peut également mener à une distorsion de transmission. Le facteur *Medea*, pour « Maternal Effect Dominant Embryonic Arrest » décrit chez le coléoptère *Tribolium castaneum* est un exemple de ce type de distorsion de transmission. Si une femelle est hétérozygote pour le facteur *Medea* et que ses descendants ne reçoivent ce facteur ni de leur mère, ni de leur père, ils ne seront pas viables (Beeman *et al.* 1992). Cela serait dû à un processus de type poison/antidote. Le facteur *Medea* serait en effet constitué de deux gènes liés, l'un produisant une toxine et l'autre produisant l'antidote à cette toxine (Beeman et Friesen 1999). Les descendants d'une femelle porteuse du facteur *Medea* possèderaient donc tous cette toxine, transmis par leur mère *via* l'ovule, mais seuls ceux ayant hérité du facteur *Medea*, donc de l'antidote, survivraient.

Il est important de distinguer les distorsions de transmission liées à la sélection naturelle, où un allèle est trouvé plus souvent chez les enfants car il est avantageux pour la fitness, des distorsions de transmission dues à des éléments égoïstes qui seront transmis plus souvent sans effets bénéfiques de l'allèle sur la fitness. Néanmoins, ces deux types de distorsion de transmission sont difficilement différenciables.

Afin de détecter un phénomène de distorsion de transmission, il est nécessaire que le locus où la distorsion se produit soit polymorphe. Or, si un distorteur n'est pas désavantageux, il envahira la population rapidement et ne sera pas détectable. Pour qu'il existe

un polymorphisme stable à un locus sous distorsion de transmission, il est nécessaire que le distorateur soit contre sélectionné à l'échelle de la population, comme les distorateurs de sexe-ratio pour lesquels il existe une forte sélection négative pour rétablir un sexe-ratio équilibré, ou soit contre sélectionné chez certains individus, comme les individus homozygotes pour le distorateur.

Méthodes de détection de la distorsion de transmission à l'aide de données génomiques chez l'Homme

Une étude de Zöllner *et al.* (2004) suggère que des processus de distorsion de transmission seraient répandus dans le génome humain. En effet, ils ont observé dans une base de données de 148 familles provenant de la *Framingham Heart Study* (Cupples *et al.* 2007), à l'aide de 888 marqueurs génétiques (microsatellites et SNP), que des frères et des sœurs étaient identiques par descendance pour 50,43% de leur génome autosomal, valeur significativement différente de l'attendu (50%). Cette différence n'a néanmoins pas été observée sur le chromosome X.

Après l'étude de Zöllner *et al.* (2004), d'autres études se sont intéressées à la détection de distorsion de transmission à l'aide de données génomiques. Trois d'entre elles ont utilisées un test de déséquilibre de transmission (Transmission Disequilibrium Test, TDT) afin de détecter des distorsions de transmission dans le génome humain. Le TDT requiert l'utilisation de données génomiques sur des familles, c'est-à-dire comportant le séquençage ou le génotypage d'au moins deux parents et un de leurs enfants (trio). Le TDT est effectué locus par locus et permet de détecter si certains d'entre eux présentent une distorsion de transmission. Ce test consiste en un test de McNemar dont l'hypothèse nulle est qu'un des allèles est transmis par des parents hétérozygotes dans les mêmes proportions que l'allèle alternatif au locus considéré (Spielman *et al.* 1993).

Le TDT a historiquement été développé afin de détecter des associations entre un génotype et une maladie. L'utilisation de trios composés d'enfants atteints de la maladie considérée permet de détecter à l'aide d'un TDT si un allèle a plus souvent été transmis aux enfants atteints de cette maladie. Ce test apporte des avantages par rapport à une étude d'association cas-contrôles car le TDT n'est pas biaisé par la structure de la population (Huang *et al.* 2013).

Si le but premier de l'étude est d'identifier des locus subissant une distorsion de transmission, que nous appellerons désormais TD (*Transmission Distortion*), au sein d'une population, les trios doivent être choisis indépendamment du phénotype des enfants qui seront *a priori* sains (Huang *et al.* 2013). Les distorsions de transmission détectées seraient alors dues à l'un des mécanismes décrits ci-dessus. La première étude qui a utilisé cette

méthode sur des données de génotypage chez l'Homme est celle de Santos *et al.* (2009), qui ont étudié le bras court du chromosome 6 à l'aide de la base de données HapMap II (The HapMap International Consortium 2005). Cette base de données comporte soixante familles provenant de deux populations différentes (CEU, résidents de l'Utah avec des origines Européennes, et YRI, Yoruba de Ibadan, Nigeria). Santos *et al.* (2009) ont mis en évidence un signal de TD pour une région contenant 3 gènes, SUPT3H, RUNX2 et MIRN586, uniquement chez les pères de la population CEU. Cela signifie que, pour cette région, les pères hétérozygotes vont préférentiellement transmettre un de leurs allèles à leurs descendants.

Ensuite, Paterson *et al.* (2009) ont utilisé la base de données de génotypage de la *Framingham Heart Study* (Cupples *et al.* 2007) qui comporte plusieurs milliers de familles. Cette étude a permis d'identifier 8 locus potentiellement sous forte TD à l'échelle du génome.

L'étude la plus récente a été effectuée par Meyer *et al.* (2012) et inclut la *Framingham Heart Study*, une base de données de familles huttérites (Ober *et al.* 2001) et la cohorte de l'*Autism Genetic Resource Exchange* (Geschwind *et al.* 2001). Chacune de ces bases de données comporte au minimum 800 familles. Plusieurs régions candidates ont été identifiées, dont la même région du bras court du chromosome 6 que celle détectée par Santos *et al.* (2009) qui présente dans les deux études une TD chez les pères.

Enfin, une cinquième étude a utilisé un test exact de Fisher, et non un TDT, afin de détecter des distorsions de transmission entre les parents et les enfants (Deng *et al.* 2009). Deng *et al.* (2009) ont également utilisé la base de données de génotypage de HapMap II (The HapMap International Consortium 2005). En prenant en compte le génome complet, ils ont mis en évidence 1161 SNP et 224 gènes potentiellement sous TD. Les SNP identifiés étaient exclusivement spécifiques à la population considérée et seuls 3 gènes étaient partagés entre les deux populations, suggérant une restriction géographique de la distorsion de transmission.

3.1.1 La distorsion de transmission et les IASC

La distorsion de transmission et les IASC sont deux types de conflits génomiques qui partagent un certain nombre de caractéristiques. Premièrement, la distorsion de transmission et les IASC ont des effets sur les fréquences alléliques dans la population (Patten 2014). Un polymorphisme stable à un locus sous sélection SA est possible sous certaines conditions (Kidwell *et al.* 1977 ; Rice 1984), et un allèle bénéfique pour les femelles mais coûteux pour les mâles sera plus souvent trouvé chez les femelles et *vice versa* pour un allèle bénéfique chez les mâles et coûteux chez les femelles. Dans le cas d'un distorateur de transmission, un allèle sera transmis plus de 50% du temps à la descendance, créant

des différences de fréquences alléliques chez les enfants par rapport à ce qui est attendu compte-tenu des génotypes des parents. Deuxièmement, il est attendu que le chromosome X soit enrichi en locus sous IASC (Rice 1984 ; Patten et Haig 2009), et en distor-teur de transmission (Hurst et Pomiankowski 1991 ; Jaenike 2001). En effet, un modèle de génétique des populations développé par Hurst et Pomiankowski (1991) montre que les conditions de persistance d'un locus distor-teur de ségrégation sont plus stringentes sur les autosomes que sur le chromosome X. Cela est dû au fait qu'un distor-teur doit être associé à un locus d'insensibilité à la distorsion afin de ne pas perturber sa propre transmission. L'absence de recombinaison entre le chromosome X et le chromosome Y favoriserait l'émergence de distor-teur de ségrégation, sur le X (ou sur le Y), qui favoriserait sa propre transmission grâce à une liaison avec le locus d'insensibilité, au dépend du chromosome Y (ou du chromosome X).

Burt et Trivers (2006) ont émis l'hypothèse que les haplotypes subissant une distorsion de transmission seraient enrichis pour des locus ayant des effets bénéfiques chez les mâles ou chez les femelles. En effet, une étude théorique récente a examiné les conditions de maintien d'un polymorphisme d'un locus sexuellement antagoniste lié à un locus provoquant une distorsion de ségrégation (Patten 2014). Dans ce cas, la distorsion de ségrégation étudiée a lieu pendant la spermatogénèse, mais les résultats sont les mêmes si celle-ci a lieu pendant l'ovogenèse.

Cette étude montre que la liaison d'un locus SA à un locus distor-teur augmente la gamme de paramètres pour lesquels un polymorphisme stable sera observé au niveau du locus SA. À chaque génération, les spermatozoïdes sont enrichis pour l'haplotype comportant l'allèle distor-teur de ségrégation et l'allèle avantageux chez les mâles, tandis que les ovules sont enrichis pour l'haplotype comportant l'allèle sensible à la distorsion de transmission et l'allèle avantageux chez les femelles. Le même raisonnement s'applique pour les distor-teurs de ségrégation spécifiques à l'ovogenèse. Le maintien d'un polymorphisme SA stable entre mâles et femelles est ainsi facilité lorsque le locus SA est lié à un distor-teur de ségrégation. Les distor-teurs de ségrégation auraient donc tendance à recruter des locus SA (Úbeda et Haig 2005). L'inverse est aussi vrai : le maintien d'un nouveau locus distor-teur est facilité si celui-ci est lié à un locus SA.

3.1.2 Distorsion de transmission sexe-spécifique

Dans le chapitre précédent, nous avons mis en évidence des signaux potentiels d'IASC en cours au sein du génome humain, des différences de fréquences alléliques entre les hommes et les femmes, dans les populations humaines de HapMap 3.3. Ces différences de fréquences alléliques pourraient être notamment dues à une viabilité différente entre

les embryons mâles et femelles en fonction de certaines combinaisons génotypiques, correspondant à une distorsion de transmission sexe-spécifique au niveau de la sélection post-fécondation. En effet, pendant la gestation, plus de 50% des grossesses ne seraient pas menées à terme chez l'Homme (Benagiano *et al.* 2010), et des différences de fréquences alléliques entre les sexes chez des nouveaux-nés ont été mises en évidence à certains locus (Ucisik-Akkaya *et al.* 2010). Une partie de ces nombreuses fausses couches pourrait donc être due à des mécanismes sexuellement antagonistes. De plus, comme suggéré par l'étude de Patten (2014) pour la distorsion de ségrégation, d'autres mécanismes de distorsion de transmission pourrait interagir avec les IASC. L'ensemble de ces processus se traduirait par une distorsion de transmission sexe-spécifique.

À notre connaissance, peu d'études se sont intéressées à ce type de mécanisme. Une étude chez la souris a mis en évidence une distorsion de transmission sexe-spécifique pour un locus sur le chromosome 2, se traduisant par des différences sexe-spécifiques de transmission des allèles (Siracusa *et al.* 1991). Une autre étude chez la souris a montré qu'un allèle à un locus situé sur le chromosome X était transmis préférentiellement aux filles mais pas aux fils, ce qui serait dû à la mort d'embryons possédant certaines combinaisons génotypiques (Casa-esperon *et al.* 2000).

Chez l'Homme, une TD sexe-spécifique a été mise en évidence sur la région du chromosome 7, impliqué dans la mucoviscidose (Kitzis *et al.* 1988). Par ailleurs, lors d'une étude sur la maladie de Crohn, Liu *et al.* (2012) ont observé des différences de fréquences alléliques entre les hommes et les femmes au niveau d'un SNP, mais uniquement chez les individus contrôles. Ce SNP est localisé dans un gène de susceptibilité à la maladie de Crohn, ATG16L1. Ce résultat a été validé dans la même étude en utilisant les données de HapMap 3, au sein desquelles une distorsion de transmission a été détectée pour ce SNP grâce à un TDT : les mères transmettaient plus souvent l'allèle T à leurs descendants. De plus, bien que la différence ne soit pas significative, les filles avaient deux fois plus de chance de recevoir l'allèle C que l'allèle T, tandis que les fils avaient plus de chance de recevoir l'allèle T que l'allèle C. Liu *et al.* (2012) concluent que dans les populations saines, des transmissions différentes en fonction du sexe du parent et du sexe de l'enfant pourraient être à l'origine d'un maintien de ces différences de fréquences alléliques entre les sexes. Un variant associé au risque d'avoir la maladie de Crohn pourrait perturber ce processus chez les individus malades.

Enfin, l'étude de Zöllner *et al.* (2004) a montré qu'en moyenne le pourcentage d'identité par descendance entre les soeurs était plus élevé qu'entre les frères, 50.78% contre 50.12%, suggérant un effet sexe-spécifique de certains locus impliqués dans la TD.

Afin d'évaluer si le signal d'IASC en cours que nous avons détecté est associé à des processus de distorsion de transmission, nous avons mis au point une méthode de dé-

tection de TD sexe-spécifique. Cette méthode nous a permis d'identifier des locus pour lesquels un allèle est préférentiellement transmis aux filles tandis qu'un autre allèle est préférentiellement transmis aux fils. De plus, nous avons pu tester si des processus de distorsion de transmission sexe-spécifique pouvaient être à l'origine des différences de fréquences alléliques entre les sexes et de leur maintien de génération en génération.

Dans ce chapitre, nous avons utilisé cette méthode statistique afin de détecter des locus sous TD sexe-spécifique sur une base de données de trios. Cette base de données a l'avantage de comporter un nombre important de trios provenant tous de la même population. De plus, à notre connaissance, aucune étude sur la TD n'a été effectuée sur cette base de données. Dans un premier temps, nous avons répété l'analyse effectuée sur HapMap 3.3 (voir section 2.2) afin d'identifier des locus montrant un signal d'IASC. Dans un second temps, nous avons identifié des locus montrant des signaux de TD chez les deux sexes, limitée à un sexe ou sexe-spécifique et nous nous sommes intéressée aux fonctions dans lesquelles ces locus étaient impliqués. Enfin, nous avons comparé le signal d'IASC avec le signal de TD sexe-spécifique, afin d'évaluer l'association entre ces deux mécanismes.

3.2 Méthodes

3.2.1 La base de données Genome of the Netherlands

« Genome of the Netherlands » (GoNL) est une base de données de séquençage de génomes entiers d'individus des Pays-Bas à accès limité (The Genome of the Netherlands Consortium 2014 ; Boomsma *et al.* 2014). Cette base de données a été construite par la collaboration de la bio-banque Néerlandaise BBMRI-NL. La particularité de cette base de données est qu'elle comporte 250 trios (les deux parents et un enfant), non apparentés entre eux. Environ 20,4 millions de SNP ont été détectés dans GoNL. La couverture génomique est d'environ 13x. Cette base de données a permis de mettre en évidence de nouveaux variants à basse fréquence, soulignant l'importance d'étudier de larges échantillons provenant d'une même population, afin d'étudier la démographie de la population des Pays-Bas. Ces variants pourraient être informatifs dans le cadre d'études d'association sur des maladies, mais aussi d'études démographiques des populations de façon locale et globale (The Genome of the Netherlands Consortium 2014 ; Boomsma *et al.* 2014).

Contrôle qualité de GoNL

Détermination du sexe des enfants

Le sexe des enfants n'étant pas renseigné, nous avons calculé la proportion de SNP hétérozygotes sur le chromosome X chez tous les individus. Les individus de sexe féminin devraient avoir une proportion de SNP hétérozygotes plus élevée que les individus de sexe masculin et former un cluster (voir figure 3.1). Nous pouvons voir en figure 3.1 que les deux sexes ont en effet des proportions de SNP hétérozygotes très différentes. Pour les enfants, on distingue deux clusters d'individus, l'un avec des proportions de SNP hétérozygotes supérieures à 6%, les individus de ce cluster ont été assignés au sexe féminin, et les autres avec des proportions de SNP hétérozygotes inférieures à 2 %, les individus de ce cluster ont été assignés au sexe masculin. Chez les parents, 2 couples avaient été mal référencés, les labels ayant été échangés entre mères et pères (couples n° 78 et n° 244).

Deux mères n'avaient pas été incluses dans la base de données à cause d'une qualité insuffisante de leur séquençage (n° 223 et n° 200). La base de données comporte donc 248 trios, dont 98 fils et 150 filles, et 2 duos père-fils.

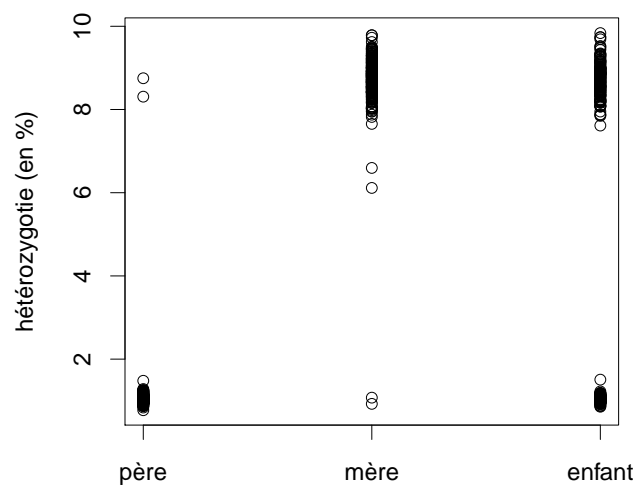


Figure 3.1 – Niveau d'hétérozygotie moyen sur le chromosome X. Pour les pères, les mères et les enfants. Les points aberrants au niveau des pères et des mères (avec des hétérozygoties supérieures à 8% et inférieures à 2% respectivement) représentent les 2 couples pour lesquels les labels de sexes ont été échangés.

Filtres sur les SNP

Seuls les SNP bi-alléliques ayant passés les contrôles qualité de GoNL ont été gardés.

Les régions pseudo-autosomales du chromosome X (build hg19) ont été mises de côté. Le terme chromosome X sera donc synonyme de chromosome X non recombinant chez l'homme dans cette partie. Les SNP liés au chromosome X qui présentaient des hommes hétérozygotes ont été supprimés. La présence de trios nous a permis d'effectuer un contrôle supplémentaire, nous avons pu vérifier s'il existait des erreurs mendéliennes au sein des trios, c'est-à-dire si les génotypes des parents concordent avec les génotypes de leur enfant. Les SNP montrant plus de 2 erreurs mendéliennes ont été supprimés. Pour les autres SNP, les erreurs mendéliennes ont été remplacées par des données manquantes.

Structure génétique sexe-spécifique dans GoNL

Comme pour les populations de HapMap 3.3 (section 2.2), nous avons vérifié que GoNL ne présentait pas de structure génétique par rapport au sexe, c'est-à-dire que les deux sexes formaient un groupe homogène. Pour cela, nous avons effectué des MDS (« Multi Dimensional Scaling ») chez les parents uniquement, pour les autosomes et le chromosome X indépendamment. Les SNP en déséquilibre de liaison avec un $r^2 > 0.25$ ont été supprimés. Pour chaque analyse indépendamment, sur les autosomes et le chromosome X, les individus ayant plus de 0.5% de données manquantes ont été supprimés. Pour les autosomes, 1 million de SNP ont été tirés au hasard, et cette opération a été répétée 10 fois. Pour le chromosome X, nous avons pris en compte tous les SNP après ces filtres. Un chromosome X a été gardé au hasard chez les femmes, tirage effectué 30 fois indépendamment. Des matrices de distances ont été ensuite calculées entre les individus en utilisant l'ASD (« allele sharing distance »). Des MDS ont été construites à partir de ces matrices de distances. Afin de déterminer si les distances entre femmes et hommes étaient significativement différentes de 0, nous avons utilisé un test de Mantel entre la matrice des distances entre hommes et femmes et une matrice où les distances entre les hommes et les femmes sont égales à 1 et les distances entre individus du même sexe sont égales à 0. Pour chacune des répétitions pour les autosomes et le chromosome X, la corrélation entre les deux matrices n'était significative pour aucune répétition (figure 3.2).

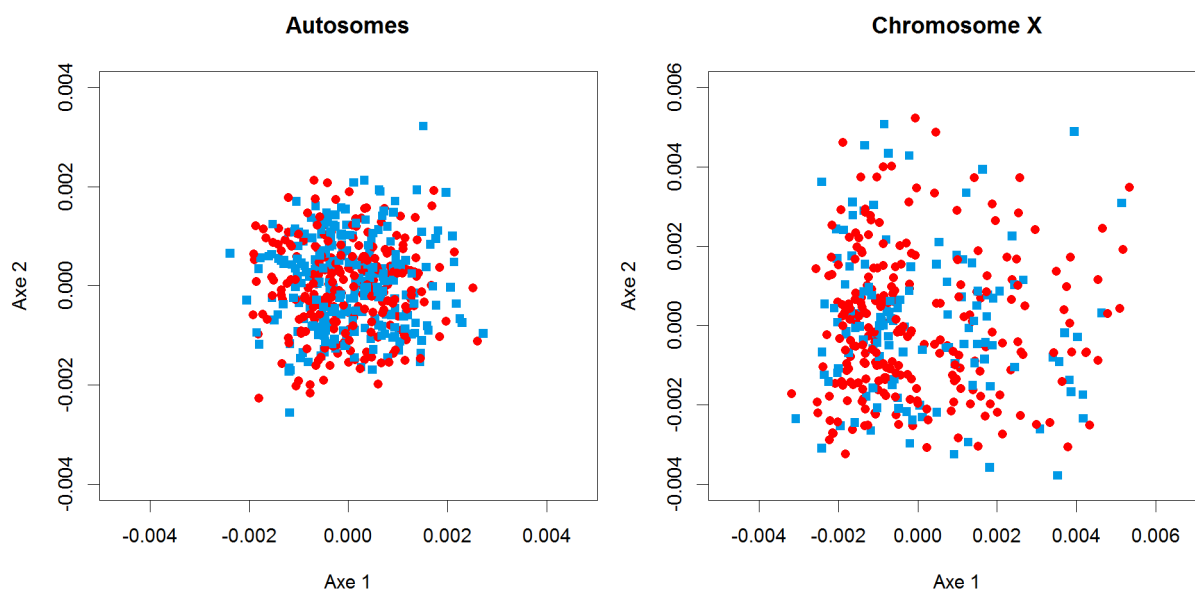


Figure 3.2 – Représentation des deux premiers axes des MDS sur les distances génétiques inter-individuelles (ASD) chez GoNL pour les autosomes (1 million de SNP au hasard, tirage n° 1) et pour le chromosome X (répétition n° 1). Les carrés bleus représentent les hommes et les ronds rouges représentent les femmes.

3.2.2 Détection de signatures de sélection sexuellement antagoniste

Dans un premier temps, nous avons effectué la même analyse que sur les populations d’HapMap 3.3 (section 2.2) sur les parents afin de détecter des signatures de sélection sexuellement antagoniste, des différences de fréquences alléliques entre les sexes. La même analyse a été effectuée sur les enfants (voir section 3.3.3). Les SNP avec une fréquence de l’allèle mineur (MAF) inférieure à 5% chez les parents ont été supprimés. Afin de déterminer quels SNP se trouvaient dans des gènes, nous avons utilisé une méthode différente de celle mise en œuvre en section 2.2. En effet, nous avons retenu les SNP se trouvant dans un intervalle de -5kb des limites du gène uniquement, en prenant en compte le brin sur lequel les gènes sont localisés (+ ou -) afin d’inclure les régions régulatrices. Pour les gènes dont le brin était indéterminé, nous avons inclus les SNP compris dans entre ± 5 kb des limites du gène. Pour chaque SNP, nous avons effectué un test exact de Fisher afin de déterminer si les fréquences alléliques étaient significativement différentes entre les sexes. Le calcul du F_{ST} intersexuel a été effectué par SNP en utilisant les fréquences alléliques chez les hommes et les femmes, à l’aide de la formule de Weir et Cockerham (1984). Une correction FDR (False Discovery Rate) a ensuite été utilisée sur les SNP géniques afin de corriger pour le nombre de SNP par gène. Les SNP ayant

une p-value significative après cette correction ont été appelés des *genic* SD SNP. Les SNP en déséquilibre de liaison au sein des listes de SNP géniques avec un $r^2 > 0.1$ ont été supprimés. Une analyse d'enrichissement fonctionnel sur les gènes comportant au moins un *genic* SD SNP (appelés *SD* gènes) a été effectuée à l'aide du logiciel DAVID (voir section 2.2) (Huang *et al.* 2009a,b). Seules les annotations provenant de la base de données d'annotation « Gene Ontology FAT » ont été utilisées.

3.2.3 Détection de distorsions de transmission

Afin de détecter une distorsion de transmission (TD) sexe-spécifique, nous avons mis au point une nouvelle approche de balayage du génome. Cela nous a permis d'identifier dans un premier temps, pour chaque SNP, s'il existait une distorsion de transmission chez les fils, puis dans un second temps s'il existait une distorsion de transmission chez les filles, et enfin si les allèles préférentiellement transmis aux deux sexes étaient, ou non, les mêmes.

Détection de distorsion de transmission chez un sexe

Pour cette analyse, nous avons utilisé un test classique, le *Transmission/Disequilibrium Test* (TDT) (Spielman *et al.* 1993). Cette analyse a été effectuée SNP par SNP pour chaque sexe indépendamment, c'est-à-dire en considérant les trios avec des fils et les trios avec des filles séparément. Les SNP retenus étant bi-alléliques, les deux allèles ont été identifiés comme l'allèle A et B.

Cas autosomal

Seuls les couples où au moins un des parents est hétérozygote ont été sélectionnés. Pour chaque parent hétérozygote, l'allèle transmis a été identifié, et le nombre de fois où B a été transmis (T) a été comparé au nombre de fois où A a été transmis (NT) chez les filles et chez les fils indépendamment (voir tableau 3.1 pour un détail des génotypes possibles chez les individus des trios).

On a donc :

$$T = BA_{AB-AA} + AB_{AB-AB} + BA_{AB-AB} + 2BB_{AB-AB} + BB_{AB-BB}$$

$$NT = AA_{AB-AA} + AB_{AB-AB} + BA_{AB-AB} + 2AA_{AB-AB} + AB_{AB-BB}$$

Où, par exemple, BA_{AB-AA} est le nombre d'enfants BA dans les trios où les parents ont pour génotypes AB-AA.

Génotypes des parents		AB-AA	AB-AB	AB-BB
Génotype	si B transmis	BA	AB, BA, BB	BB
de l'enfant	si A transmis	AA	AB, BA, AA	AB

Table 3.1 – Génotype possible de l'enfant compte-tenu des génotypes de ses parents si le parent hétérozygote a transmis A ou s'il a transmis B, pour un locus autosomal. Le premier allèle du genotype de l'enfant correspond à l'allèle transmis par le premier parent.

Un test de McNemar a ensuite été utilisé pour évaluer si T et NT étaient significativement différents, c'est-à-dire si un allèle est préférentiellement transmis par le ou les parents hétérozygotes chez leurs enfants, en séparant les fils et les filles.

Sous l'hypothèse nulle d'une absence de TD, le critère du test est :

$$\frac{(T - NT)^2}{T + NT} \text{ et suit une loi de } \chi^2$$

Ces tests seront appelés TD_{fils} et TD_{fille} .

Cas du chromosome X

La méthode utilisée est équivalente à celle des autosomes, mais est adaptée à l'hémizygotie des mâles pour le chromosome X. Seuls les couples où la mère est hétérozygote ont été sélectionnés. L'allèle transmis par la mère a été identifié, et le nombre de fois où B a été transmis (T) a été comparé au nombre de fois où A a été transmis (NT) chez les fils et chez les filles indépendamment (voir tableau 3.2 pour un détail des génotypes possibles chez les individus des trios).

Génotypes des parents (mère-père)		Filles		Fils	
		AB-A	AB-B	AB-A	AB-B
Génotype	si B transmis	BA	BB	B	B
de l'enfant	si A transmis	AA	AB	A	A

Table 3.2 – Génotype possible de l'enfant compte-tenu des génotypes de ses parents si la mère a transmis A ou si elle a transmis B, pour un locus lié à l'X. Les génotypes des filles et des fils sont indiqués. Le premier allèle du genotype de l'enfant correspond à l'allèle transmis par la mère.

On a donc, pour les filles :

$$T = BA_{AB-A} + BB_{AB-B}$$

$$NT = AA_{AB-A} + AB_{AB-B}$$

Pour les fils :

$$T = B_{AB-A} + B_{AB-B}$$

$$NT = A_{AB-A} + A_{AB-B}$$

Où, par exemple, BA_{AB-A} est le nombre de filles BA dans les trios où les parents ont pour génotypes AB-A. L'approche statistique utilisée est identique à celle utilisée pour les autosomes et repose sur un test de McNemar. Ces tests seront également appelés TD_{fils} et TD_{fille} .

Comparaison de l'allèle le plus transmis chez les fils et les filles

Après avoir effectué l'analyse de distorsion de transmission chez les fils et les filles, nous avons évalué si l'allèle le plus transmis était le même chez les deux sexes. Pour cela, nous avons comparé le nombre de fois où A et B ont été transmis chez les filles avec le nombre de fois où A et B ont été transmis chez les fils. Un tableau de contingence a été construit entre T et NT chez les fils et les filles (table 3.3). Un test de Fisher a été utilisé afin d'évaluer la significativité de la différence entre le nombre de fois où A et B ont été transmis chez chaque sexe. Ce test sera nommé pour la suite le test SDT (pour *Sexually Distorted Transmission*).

	Fils	Filles
T	T_{fils}	T_{filles}
NT	NT_{fils}	NT_{filles}

Table 3.3 – Tableau de contingence utilisé pour tester si l'allèle le plus transmis est le même chez les fils et chez les filles.

Interprétation des résultats

Le tableau 3.4 résume les interprétations possibles des résultats des 3 différents tests. Ces tests vont nous permettre de différencier les cas où l'on détecte :

- pas de TD (cas 1)
- une TD limitée à un sexe, un seul des sexes va subir une TD (cas 2, 3, 6 et 7)
- une TD chez les deux sexes dans le même sens, qui est la définition de la TD classique (cas 4)
- une TD sexe-spécifique, qui a lieu chez les deux sexes mais dans des directions opposées (cas 5, 6, 7 et 8)

La puissance des tests utilisés est très dépendante de la taille de l'échantillon. De fait, les cas 6 et 7 peuvent être interprétés de différentes façons : soit il s'agit d'une TD sexe-spécifique car le test SDT (test de Fisher) est significatif, et il y a suffisamment de puissance pour détecter une TD chez un sexe mais pas chez l'autre sexe, soit il s'agit

d'une TD très importante limitée à un sexe, qui crée des écarts entre les T et NT entre les sexes suffisamment grands pour que le test SDT soit significatif. Ce dernier scénario devrait être rare car il impliquerait une TD très importante limitée à un seul sexe.

Cas	TD_{fils}	TD_{fille}	SDT	Interprétations
1				Pas de TD
2	×			TD limitée à un sexe
3		×		TD limitée à un sexe
4	×	×		TD
5			×	TD sexe-spécifique
6	×		×	TD limitée à un sexe ou TD sexe-spécifique
7		×	×	TD limitée à un sexe ou TD sexe-spécifique
8	×	×	×	TD sexe-spécifique

Table 3.4 – Interprétations des résultats des 3 tests effectués : le test TD_{fils} , le test TD_{filles} et le test SDT. Une croix signifie que le test est significatif pour le SNP considéré.

3.2.4 Simulation d'un chromosome avec un locus sous sélection SA

Nous avons ensuite simulé une région sous sélection SA sur la viabilité, qui par définition sera sous TD sexe-spécifique. Le but de ces simulations était de déterminer si on pouvait, à l'aide de notre méthode, détecter communément les deux signaux. Nous avons effectué des simulations sur un autosome et un chromosome X, chaque chromosome possédant un site sous sélection sexuellement antagoniste affectant la survie, pour différents paramètres de dominance et de sélection.

Génération d'une population initiale

Tout d'abord, les deux types de chromosomes ont été simulés en utilisant le logiciel *ms*, qui génère des chromosomes sous neutralité en utilisant le modèle de coalescence avec recombinaison (logiciel *ms*, Hudson 2002). Dans un premier temps, la généalogie d'un échantillon aléatoire de copies de gènes est déterminée puis des mutations sont placées aléatoirement sur la généalogie. Nous avons simulé 10 000 chromosomes en utilisant :

- un taux de mutation $\mu = 2.5 \times 10^{-8}$ par paire de bases par génération
- un effectif $N = 10000$, avec une population de taille constante
- une longueur de 200000 paires de base
- un taux de recombinaison $r = 1 \times 10^{-8}$ recombinaison entre deux paires de bases adjacentes par génération

A partir de ces 10000 chromosomes, 3000 individus ont été créés en tirant deux chromosomes au hasard sans remise dans le cas de l'autosome, et un chromosome pour les mâles et deux chromosomes pour les femelles dans le cas du chromosome X. Pour les autosomes, un sexe a été assigné au hasard à chaque individu.

Simulation « forward » de la sélection SA

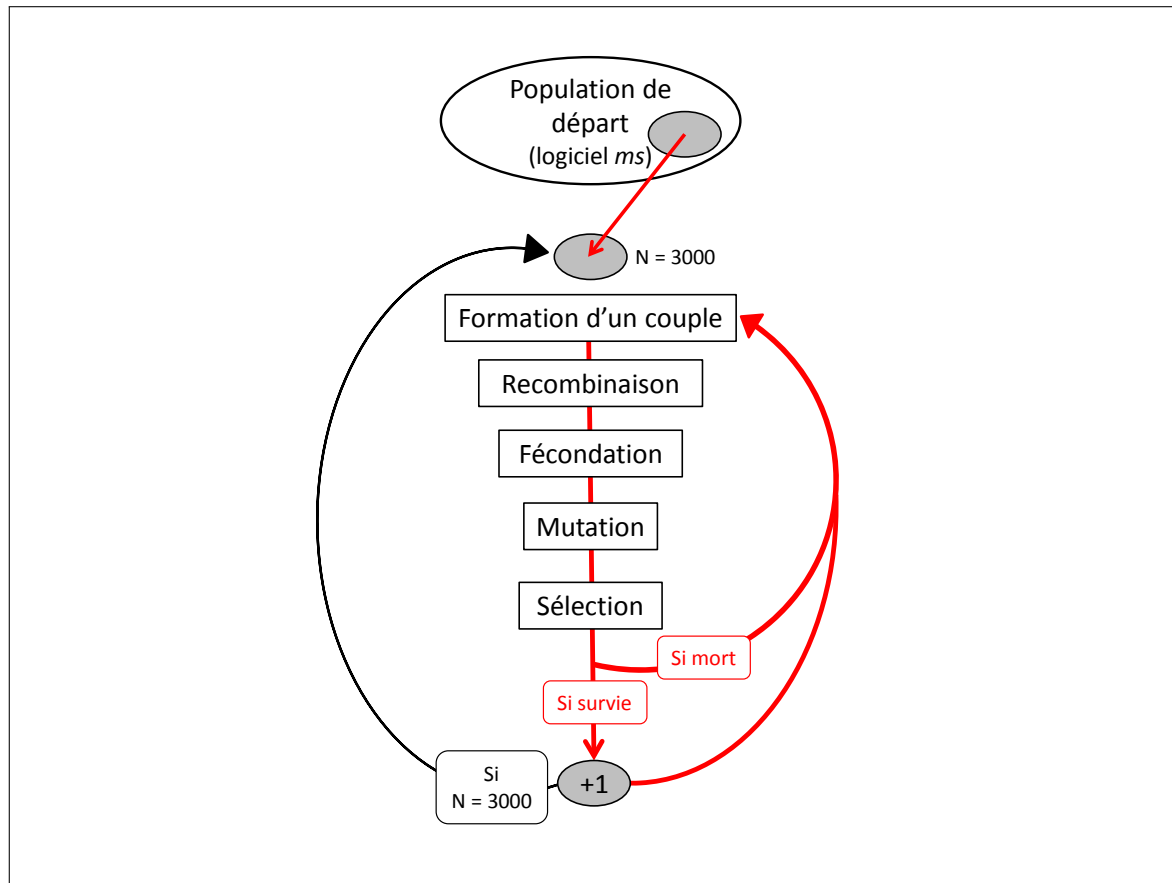


Figure 3.3 – Schéma des étapes des simulations en « forward » de la sélection SA sur un chromosome ségrégeant dans une population de taille constante.

Nous avons ensuite simulé en *forward* l'évolution de ces chromosomes, en désignant un locus polymorphe comme sexuellement antagoniste. Ce locus a été tiré au hasard parmi ceux ayant une fréquence allélique intermédiaire (entre 0.4 et 0.6) et ayant une position centrale sur le chromosome (comprise entre 90kb et 110kb). Nous avons ensuite simulé 21 générations de la façon suivante (les étapes sont résumées figure 3.3) :

1. **Accouplement aléatoire** : un individu mâle et un individu femelle sont tirés au hasard pour former un couple,
2. **Recombinaison** : les chromosomes des parents subissent des événements de recombinaison, avec un taux de recombinaison $r = 1 \times 10^{-8}$ recombinaison entre deux

paires de bases adjacentes par génération,

3. **Fécondation** : un chromosome est tiré au hasard chez le père et chez la mère pour créer un enfant dont le sexe est également tiré au hasard,
4. **Mutation** : les chromosomes choisis pour la fécondation sont soumis à des événements de mutation selon un taux de mutation $\mu = 2.5 \times 10^{-8}$ par paire de bases par génération,
5. **Sélection** : une valeur de fitness w est attribuée à l'enfant en fonction du génotype qu'il porte au locus SA, et de son sexe (voir le modèle de sélection table 3.5 et 3.6) : w est compris entre 0 et 1 et représente la probabilité de survie de l'individu. Le processus est répété jusqu'à ce que le nombre d'enfants vivants soit égal au nombre d'individus du départ (N=3000).

À la fin de cette simulation, les enfants deviennent des parents potentiels et la simulation reprend à l'étape 1.

Modèle de sélection

Pour construire les modèles de sélection, nous nous sommes inspirée du modèle développé par Rice (1984) et des notations de Fry (2010) (voir section 2.1.1). On considère un locus bi-allélique (A/B), A étant avantageux chez les mâles et désavantageux chez les femelles et B étant avantageux chez les femelles et désavantageux chez les mâles. Les paramètres de ce modèle comporte :

- un paramètre de dominance : h_f pour la dominance de A et h_m pour la dominance de B. Ce paramètre est compris entre 0 et 1, 0 correspondant à une récessivité totale et 1 à une dominance totale.
- un paramètre de sélection : s_m , le coefficient de sélection chez les mâles et s_f , le coefficient de sélection chez les femelles. s_m et s_f sont compris entre 0 et 1. Plus ces coefficients sont élevés, plus le coût en fitness est haut.

Pour simplifier nos simulations, nous avons supposé que la dominance des allèles A et B était la même chez les deux sexes (c'est-à-dire $h_m = 1 - h_f$) comme dans le modèle de Rice (voir section 2.1.1), ainsi que la sélection (c'est-à-dire $s_m = s_f$).

Les modèles de sélection utilisés sont représentés en table 3.5 pour les autosomes et en table 3.6 pour le chromosome X.

Les simulations ont été effectuées sur toutes les combinaisons possibles de paramètres de dominance et de sélection avec un pas de 0,1, et chaque combinaison a été répétée 10 fois. Cette analyse nous a permis de simuler un locus sous sélection sexuellement antagoniste ainsi que sa région environnante pour différents taux de sélection et de dominance.

Génotype	AA	AB	BB
Mâle	1	$1 - h_m s_m$	$1 - s_m$
Femelle	$1 - s_f$	$1 - h_f s_f$	1

Table 3.5 – Modèle de sélection sexuellement antagoniste pour un autosome. Les valeurs indiquées représentent la fitness de l'individu considéré, comprise entre 0 et 1, chez les deux sexes selon leur génotype.

Génotype	AY / AA	\emptyset / AB	BY / BB
Mâle	1		$1 - s_m$
Femelle	$1 - s_f$	$1 - h_f s_f$	1

Table 3.6 – Modèle de sélection sexuellement antagoniste pour le chromosome X. Il n'y a pas de mâles hétérozygotes. Les valeurs indiquées représentent la fitness de l'individu considéré, comprise entre 0 et 1, chez les deux sexes selon leur génotype.

Nous avons simulé un type de sélection sexuellement antagoniste agissant directement sur la survie des embryons, et créant des taux de mortalité différents entre les enfants mâles et femelles. En conséquence, nous nous attendons à détecter des différences de fréquences alléliques entre les sexes ainsi qu'une TD sexe-spécifique au niveau de ce locus.

3.2.5 Évaluation du F_{ST} intersexuel chez les enfants : est-il extrême comparé aux génotypes de leurs parents ?

Afin de pouvoir comparer les F_{ST} intersexuels observés chez les enfants avec ceux des parents, nous avons dû déterminer si les F_{ST} intersexuels observés chez les enfants sont extrêmes compte tenu des génotypes de leur parents. En effet, les enfants ne constituant pas un ré-échantillonnage aléatoire de la population, nous avons dû prendre en compte les génotypes de la génération précédente. Pour chaque SNP et chaque couple, nous avons construit tous les enfants possibles compte tenu des génotypes des parents. À partir de ce nouvel ensemble d'individus, nous avons tiré au hasard le même nombre de filles et de fils que dans GoNL (respectivement 150 et 98) et le F_{ST} intersexuel a été calculé entre les deux sexes. Ce tirage a été répété 10000 fois afin d'obtenir une distribution attendue des F_{ST} intersexuels chez les enfants. Le F_{ST} intersexuel observé a été comparé à cette distribution nulle, la p-value de ce F_{ST} intersexuel correspond au pourcentage de F_{ST} plus élevés, dans la distribution nulle, que le F_{ST} observé.

3.3 Résultats

3.3.1 Détection de signatures de sélection sexuellement antagoniste dans Genome of the Netherlands

Dans un premier temps, nous avons effectué la même analyse que sur les populations de HapMap 3.3 (voir section 2.2) sur les données de GoNL, uniquement chez les parents. La proportion de *genic* SD SNP par chromosome a été calculée avant et après suppression des SNP géniques en déséquilibre de liaison (Linkage Disequilibrium, LD) pour un $r^2 > 0,1$. Dans les deux cas, un test exact de Fisher a été effectué pour évaluer la significativité de la différence entre les proportions du chromosome X et de chaque autosome. Comme observé pour les populations de HapMap 3.3, la proportion de *genic* SD SNP sur le chromosome X est généralement plus élevée que sur les autosomes (figure 3.4 et 3.5). Avant suppression des SNP en LD, le chromosome X montre une proportion de *genic* SD SNP significativement plus élevée que celle de 15 autosomes sur 22. Après suppression des SNP en LD, cette différence n'est significative que pour 4 autosomes. Le résultat obtenu dans la population néerlandaise de GoNL est similaire à celui observé dans la population européenne de HapMap 3.3 (CEU), pour laquelle on observait la même tendance mais la différence entre les proportions de *genic* SD SNP sur le chromosome X et les chromosomes 10 et 22 n'était pas significative après suppression des SNP en LD.

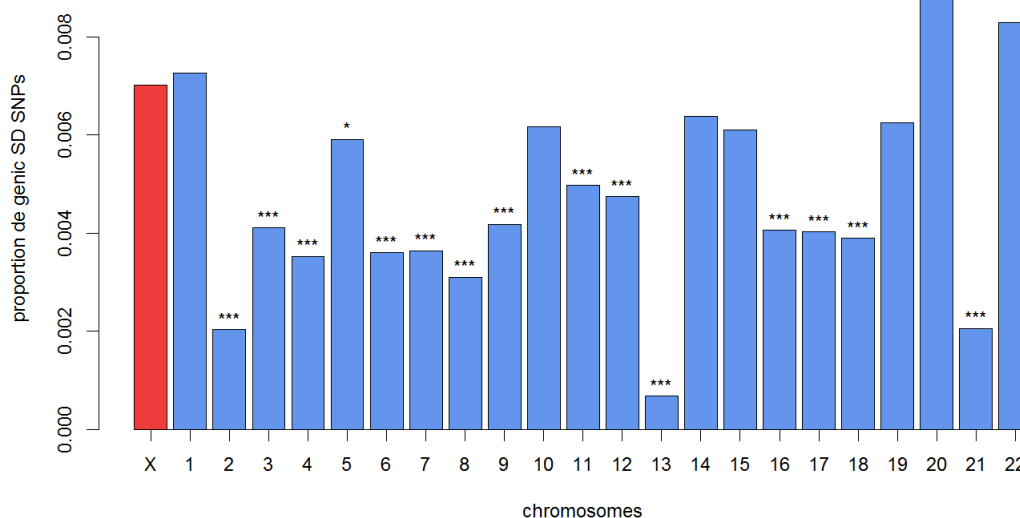


Figure 3.4 – Proportion de *genic* SD SNP par chromosome. Les niveaux de significativité des tests exacts de Fisher effectués entre les proportions de *genic* SD SNP du chromosome X (en rouge) et celles des autosomes (en bleu) sont représentés (p-value : * < 0.05, ** < 10^{-2} , *** < 10^{-3}).

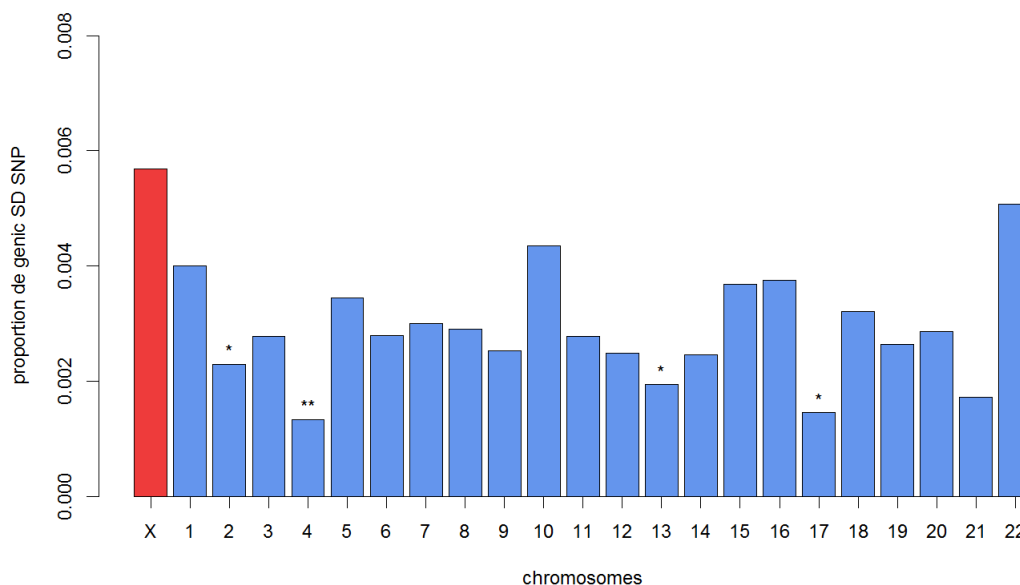


Figure 3.5 – Proportion de *genic* SD SNP par chromosome après suppression des SNP en déséquilibre de liaison pour un $r^2 > 0,1$. Les niveaux de significativité des tests exacts de Fisher effectués entre les proportions de *genic* SD SNP du chromosome X (en rouge) et celles des autosomes (en bleu) sont représentés (p-value : * < 0.05, ** < 10^{-2}).

Termes	p-value
One-carbon metabolic process	0.003
Vesicle targeting, to, from or within Golgi	0.006
G-protein coupled receptor protein signaling pathway	0.008
RPTP-like protein binding	0.010
Neuropeptide signaling pathway	0.010
G-protein signaling, coupled to cyclic nucleotide second messenger	0.012
Sensory perception of taste	0.013
Cell cycle phase	0.016
Histamine receptor activity	0.020
Regulation of fatty acid metabolic process	0.021
Cyclic-nucleotide-mediated signaling	0.025
Positive regulation of fatty acid metabolic process	0.029
Vesicle localization	0.031
Cell surface receptor linked signal transduction	0.032
Mitotic cell cycle	0.032
Vesicle targeting	0.033
Kidney development	0.036
Regulation of cellular ketone metabolic process	0.037
COPI coating of Golgi vesicle	0.040
Golgi vesicle budding	0.040
Golgi transport vesicle coating	0.040
Second-messenger-mediated signaling	0.042
Meiotic cell cycle	0.043
Endoplasmic reticulum lumen	0.045
Cell cycle process	0.045
Meiosis I	0.046
Regulation of cholesterol storage	0.047

Table 3.7 – Termes enrichis non inclus dans le cluster « Chromatin assembly and organization », classés selon leurs p-values. Les termes ont été colorés s'ils référaient : aux processus métaboliques (en vert), à l'appareil de Golgi (en bleu), au cycle cellulaire (en orange) et à la méiose (en rose).

Nous avons ensuite réalisé une étude d'enrichissement fonctionnel sur la liste de SD gènes du génome complet avec le logiciel DAVID. Tout d'abord, nous avons utilisé le regroupement de fonctions (*Clustering*) proposé par DAVID. Le cluster fonctionnel « Chromatin assembly and organization » est significativement enrichi avec un score d'enrichissement de 1,38. Il contient des termes référant aux nucléosomes, à la chromatine et à l'organisation du chromosome (tableau 4.1 en annexe).

Dans un second temps, nous avons étudié la liste complète des termes fonctionnels enrichis (tableau 3.7). Plusieurs termes réfèrent à des fonctions du métabolisme (en vert), des fonctions méiotiques (en rose), des fonctions impliquant l'appareil de Golgi (en bleu)

et des fonctions du cycle cellulaire (en orange). De plus, un terme référant aux fonctions neuronales est significatif (« Neuropeptide signaling pathway »), ainsi qu'un terme en rapport avec des fonctions sensorielles (« Sensory perception of taste ») et le développement du rein (« Kidney development »).

3.3.2 Détection de distorsions de transmission sur des chromosomes simulés

Nous avons simulé une région sous sélection SA, créant de la TD sexe-spécifique. À l'aide de ces données simulées, nous avons pu évaluer si la méthode de détection de TD sexe-spécifique que nous avons mis au point identifie ce signal au bon endroit. Nous avons également pu observer l'influence de différents taux de sélection et de dominance sur les F_{ST} intersexuels et la détection de TD sexe-spécifique. Les résultats présentés pour les calculs de F_{ST} intersexuels correspondent à la dernière génération simulée, et la distorsion de transmission a été calculée entre l'avant dernière génération et la dernière génération. La plupart des résultats sont montrés pour un s_f compris entre 0,0 et 0,5 car des valeurs de coefficient de sélection plus élevées sont irréalistes.

Dans un premier temps, nous avons étudié le comportement de la moyenne du F_{ST} intersexuel au niveau du locus sous sélection SA au cours des générations (figure 3.6). On remarque qu'à la dernière génération les F_{ST} sont du même ordre entre l'autosome et le chromosome X, sauf pour des s_f élevés, pour lesquels on observe des F_{ST} plus élevés sur l'autosome pour des valeurs de h_f élevées ou basses. Le chromosome X présente une plus forte variation du F_{ST} que l'autosome, particulièrement importantes entre les générations 1 à 5, puis les valeurs de F_{ST} se stabilisent. L'autosome présente des variations intergénérationnelles moins importantes, mais le F_{ST} continue d'augmenter pour des valeurs de h_f élevées et basses. Ceci indique que le F_{ST} , pour une même force de sélection, n'est pas plus élevé sur le chromosome X que sur l'autosome, sauf lors des premières générations.

Le modèle de Rice prédit qu'un locus sous sélection SA augmente en fréquence s'il est rare sur les autosomes si le coût en fitness chez un sexe est plus faible que le gain en fitness chez l'autre sexe ($s_m > s_f$ ou $s_f > s_m$) et sur le chromosome X si l'allèle bénéfique aux femelles est dominant ($h_m > 0,5$) et l'allèle bénéfique au mâle est récessif ($h_f < 0,5$). Dans les simulations effectuées ici, $s_m = s_f$ et $h_f = 1 - h_m$. Le modèle de Rice prédit les conditions d'invasion d'un locus sous sélection SA, mais nos simulations ne nous permettent pas de tester cela. Nous avons néanmoins voulu déterminer quelles sont les conditions les moins coûteuses pour le maintien d'un polymorphisme SA au sein d'une population à l'équilibre, donc pour la dernière génération. Nous avons ensuite comparé les conditions les plus favorables au maintien du polymorphisme aux conditions les plus

favorables à l'augmentation en fréquence d'un locus SA telles que décrites par Rice. Pour cela, nous avons divisé le F_{ST} intersexuel par le nombre de mort, donc par le coût induit par le maintien du polymorphisme SA.

Figure 3.7, on remarque que le F_{ST} corrigé pour le nombre de mort augmente quand la sélection est forte, et qu'il est plus haut lorsque h_f est proche de 0 ou de 1 pour l'autosome et lorsque h_f est plus proche de 1 pour le chromosome X. Cela signifie que le ratio entre le F_{ST} et le nombre de morts est plus élevé pour le chromosome X lorsque l'allèle bénéfique aux mâles est récessif et l'allèle bénéfique aux femelles est dominant, ce qui correspond à la prédiction du modèle de Rice pour l'émergence d'un locus SA. Pour l'autosome lorsqu'un des allèles est dominant, le coût de maintien d'un polymorphisme SA semble plus faible. Or, selon le modèle de Rice, la dominance n'intervient pas dans les conditions de maintien d'un polymorphisme SA. Néanmoins il semblerait que la dominance ait un effet sur le coût du maintien du polymorphisme SA.

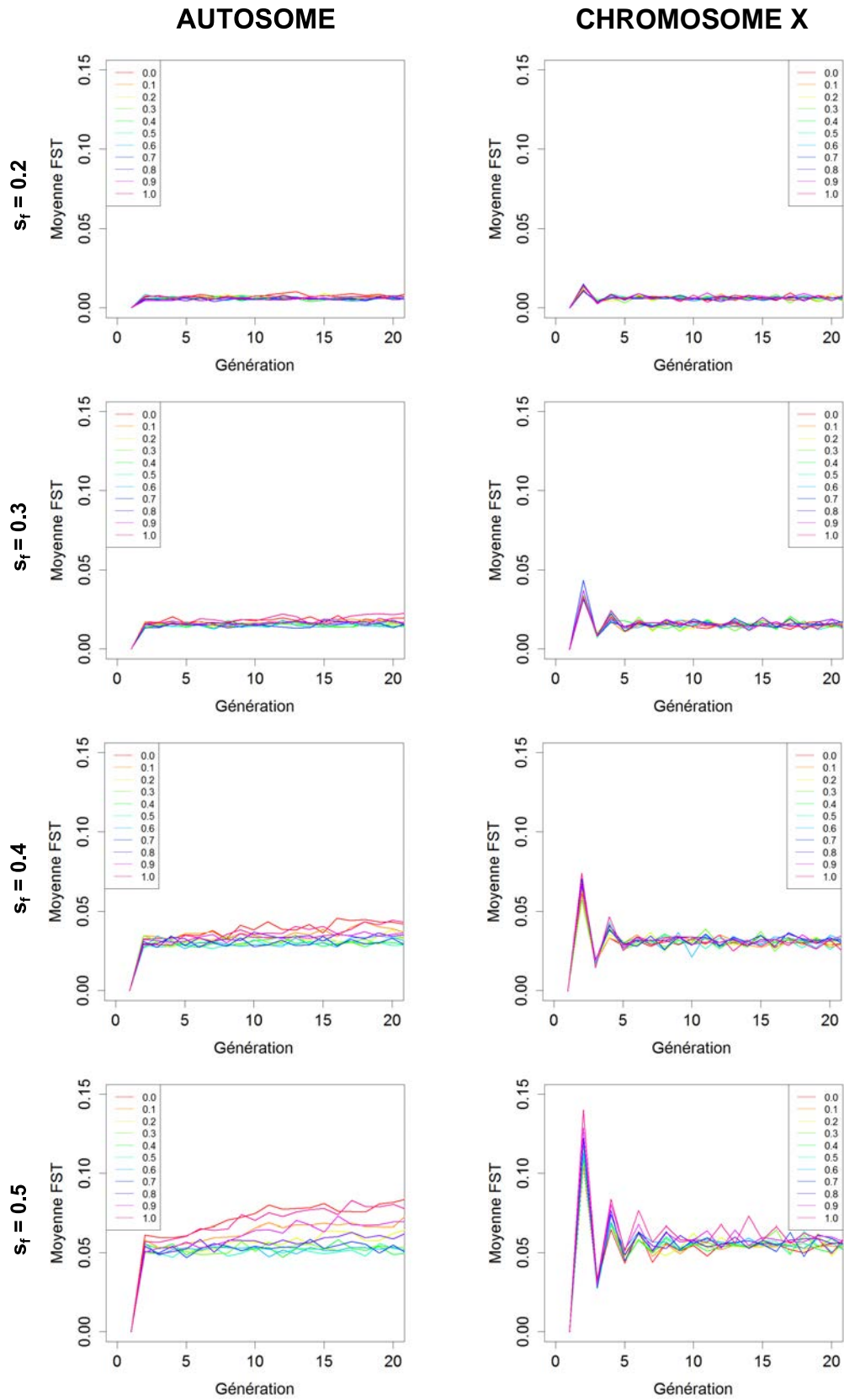


Figure 3.6 – Évolution des valeurs de F_{ST} intersexuel par génération moyennées pour les 10 simulations en fonction de s_f et de h_f (échelle de couleur).

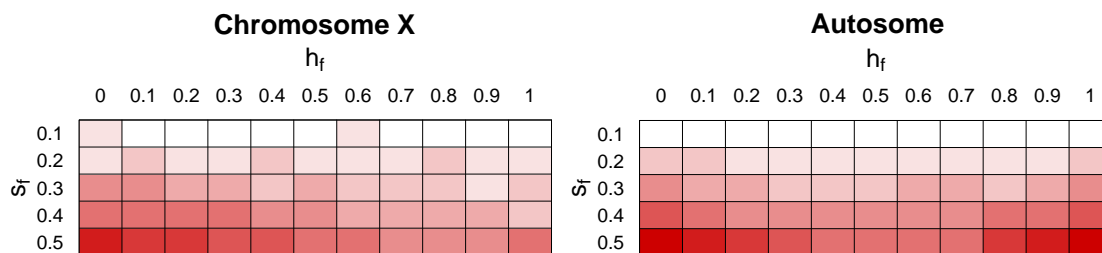


Figure 3.7 – Moyenne du F_{ST} intersexuel corrigé par le nombre de morts en fonction de h_f , s_f (10 répétitions par combinaison). Plus une case est rouge, plus le coût de maintien est faible. (Voir matrice de valeur en annexe table 4.3.)

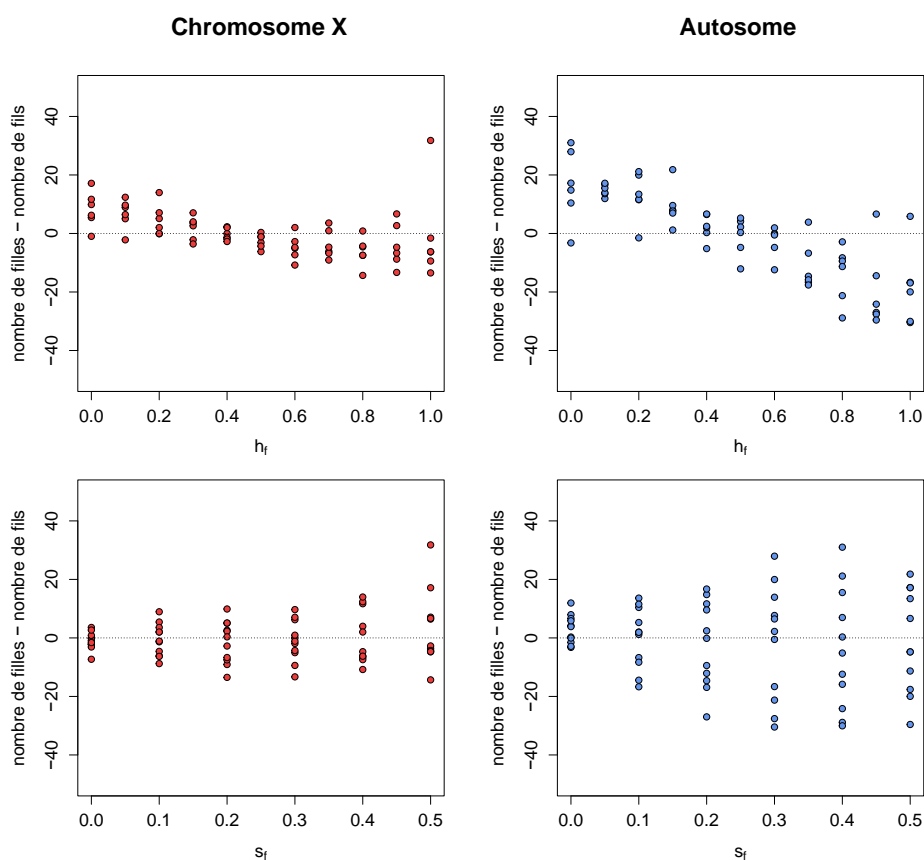


Figure 3.8 – Différence entre le nombre de filles et le nombre de fils pour des valeurs de h_f comprises entre 0,0 et 1,0 et des valeurs de s_f comprises entre 0,0 et 0,5. Chaque point représente la moyenne de la différence entre le nombre de filles et de fils pour la génération 21 entre les 10 répétitions pour une même combinaison de valeur de s_f et de h_f .

Nous avons ensuite évalué si les taux de mortalité sont différents entre filles et fils. Pour cela, nous avons calculé le nombre de filles et de fils présents à la dernière géné-

ration (figure 3.8), sachant qu'un sexe-ratio équilibré devrait être observé si les taux de mortalité étaient les mêmes chez les deux sexes. On remarque que, pour le chromosome X et l'autosome, plus de fils meurent lorsque h_f est proche de 0, c'est-à-dire quand l'allèle bénéfique au mâle est récessif, et plus de filles meurent lorsque h_f est proche de 1, c'est-à-dire quand l'allèle bénéfique au mâle est dominant. Cette tendance est plus marquée sur l'autosome. En résumé, si l'allèle délétère au sexe considéré est dominant, ce sexe aura tendance à avoir un taux de mortalité plus élevé. Si l'on considère le paramètre de sélection s_f , les points sont en majorité uniformément répartis pour l'autosome tandis que pour le chromosome X, une plus forte valeur de s_f semble avoir une influence sur le sexe-ratio, la mortalité semblant plus élevée chez les fils.

La figure 3.9 représente le F_{ST} intersexuel pour une région autosomale et la figure 3.10 pour une région liée à l'X, pour une répétition de chaque paramètre. Pour les deux chromosomes, on observe que plus le taux de sélection augmente, plus le F_{ST} intersexuel est élevé au niveau du locus sous sélection SA, et plus ce signal s'étend aux locus voisins. Par ailleurs, si on applique la méthode de détection de TD sur ces données de simulation, on observe que les locus montrant un signal de TD sexe-spécifique sont concentrés autour du locus sous sélection SA. Les locus avec les plus forts F_{ST} intersexuels sont aussi en majorité ceux qui montrent le signal le plus fort de TD sexe-spécifique, c'est-à-dire pour lesquels les 3 tests (TD_{fille} , TD_{fils} et SDT) sont significatifs. Les locus montrant un signal intermédiaire, c'est-à-dire pour lesquels uniquement le test SDT est significatif, sont également enrichis dans la région proche du locus SA.

Nous avons ensuite calculé les distributions des valeurs de F_{ST} intersexuel au niveau des locus SA et les distributions du nombre de locus montrant un fort signal de TD sexe-spécifique (3 tests significatifs) pour l'autosome (figure 3.11) et le chromosome X (figure 3.12). Les F_{ST} intersexuels au niveau du locus SA ont été calculés sur 10 ré-échantillonnages de 500 individus, correspondant au nombre de parents de GoNL. Les distributions du F_{ST} intersexuel pour les 10 ré-échantillonnages au sein des 10 répétitions pour chaque combinaison de valeurs des paramètres sont montrées. Pour les autosomes, on observe que les F_{ST} augmentent pour des valeurs de h_f proche de 0 ou de 1 comparées aux valeurs de h_f intermédiaires, ce qui est en accord avec les observations de la figure 3.7. Pour le chromosome X, h_f ne semble pas avoir d'effet sur la distribution des F_{ST} , sauf lorsque $s_f = 0,5$ où l'on observe des valeurs plus basses de F_{ST} pour $h_f = 1$. Le nombre de SNP avec un signal de TD sexe-spécifique augmente lorsque s_f augmente pour les deux chromosomes. Les distributions du nombre de SNP montrant un signal de TD sexe-spécifique pour les autosomes suivent celles du F_{ST} tandis que les distributions sur

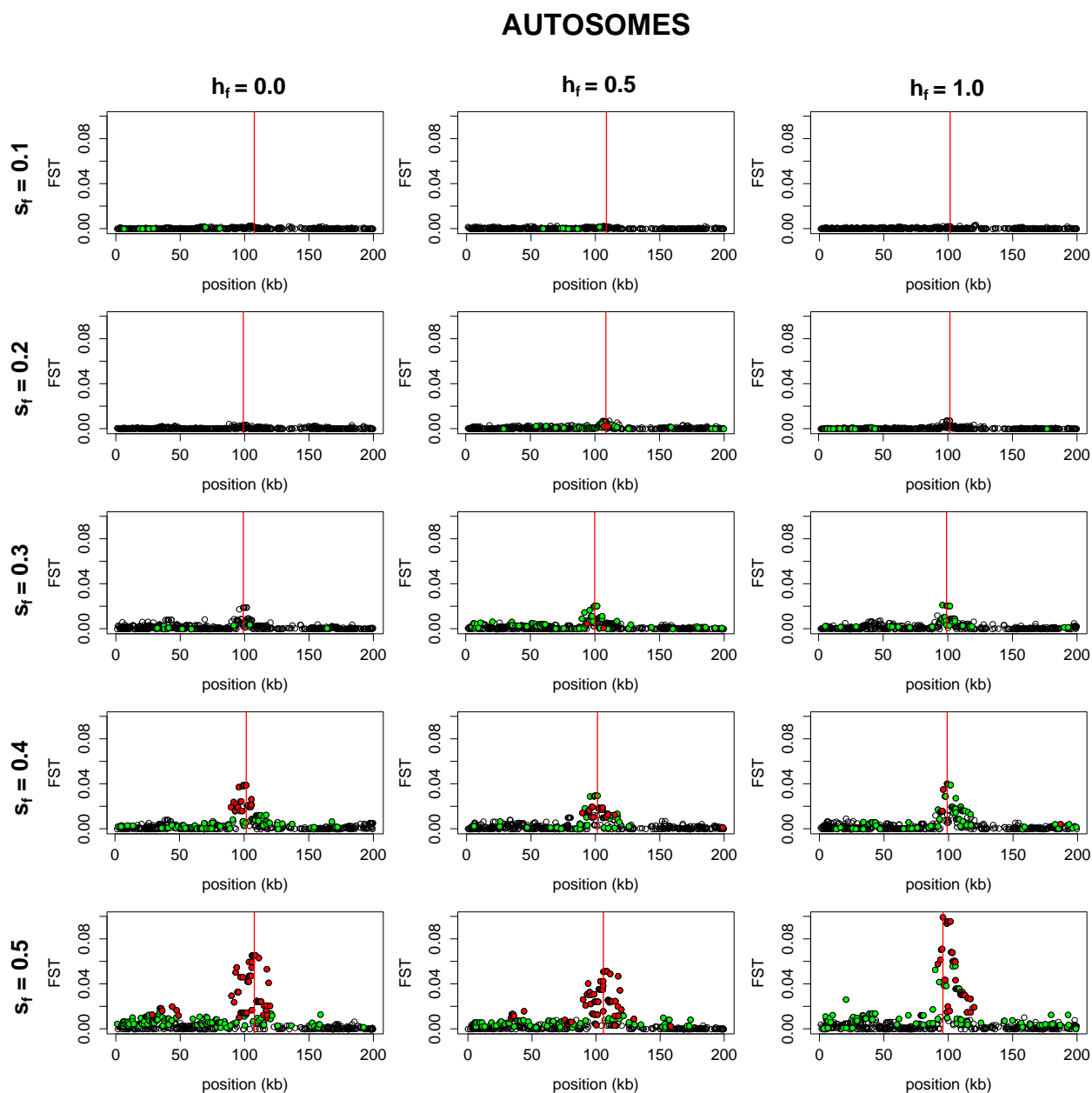


Figure 3.9 – F_{ST} intersexuel en fonction de la position sur le chromosome simulé, ici un autosome, pour différents paramètres de sélection et de dominance et pour la répétition 0. Le trait rouge représente la position du locus sous sélection SA. Les points verts sont les SNP ayant une p-value significative pour le test SDT uniquement. Les points rouges sont les SNP ayant une p-value significative pour les tests de TD_{fille} , TD_{fils} et SDT.

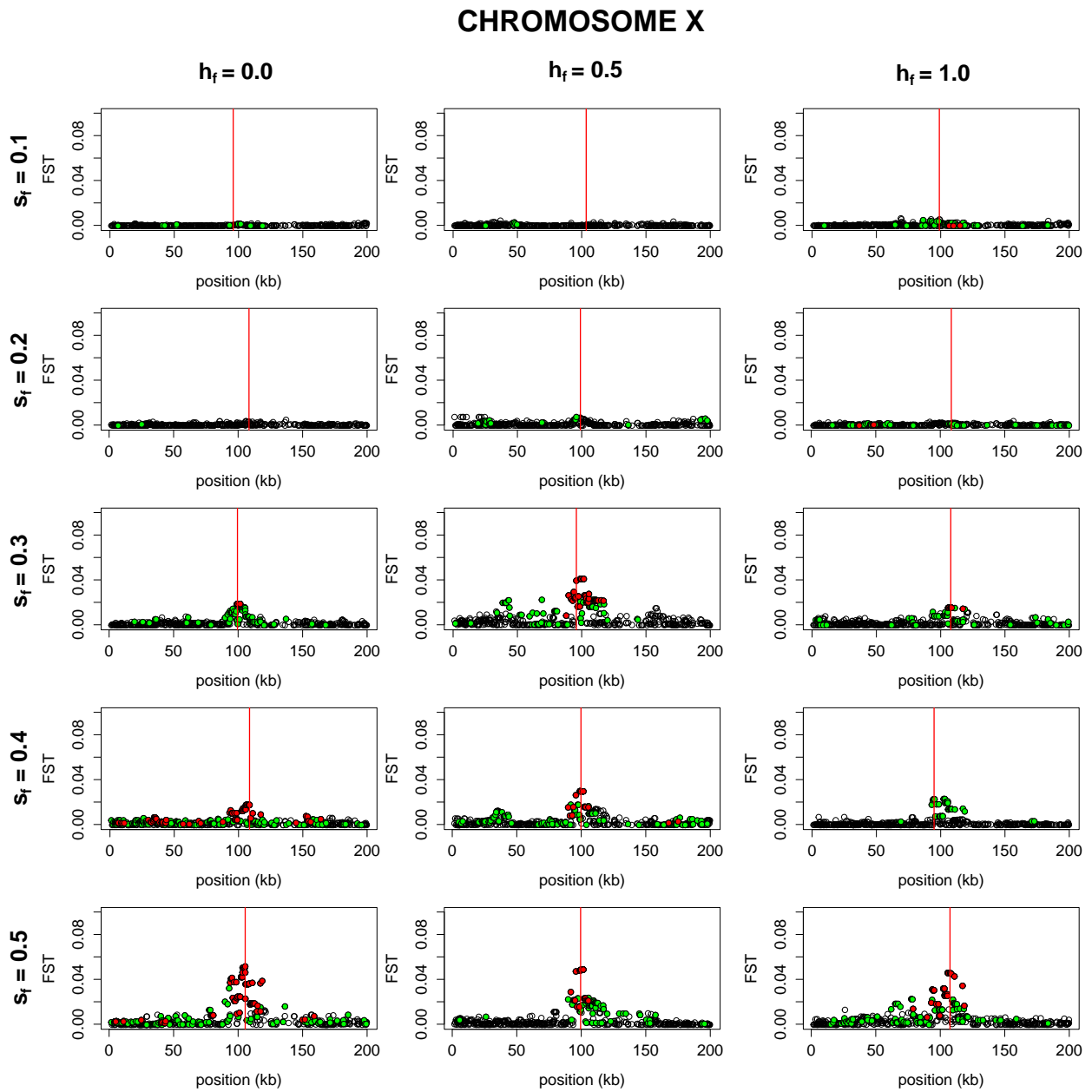


Figure 3.10 – F_{ST} intersexuel en fonction de la position sur le chromosome simulé, ici le chromosome X, pour différents paramètres de sélection et de dominance et pour la répétition 0. Le trait rouge représente la position du locus sous sélection SA. Les points verts sont les SNP ayant une p-value significative pour le test SDT uniquement. Les points rouges sont les SNP ayant une p-value significative pour les tests de TD_{fille} , TD_{fils} et SDT.

le chromosome X semblent plus affectées par les valeurs de h_f que celles du F_{ST} , mais montrent une tendance similaire.

La forte baisse du nombre de SNP avec un signal de TD sexe-spécifique pour $h_f = 1$ et $s_f = 0,5$ sur le chromosome X pourrait s'expliquer par le taux de mortalité différentiel entre les sexes observé figure 3.8. En effet, on observe une différence entre le nombre de fils et de filles plus important pour ces valeurs de paramètres. La puissance des tests sur le chromosome X étant globalement moins élevée que sur les autosomes en raison de l'hémizygotie des mâles, une baisse du nombre de mâles par rapport au nombre de femelles pourrait avoir un effet plus drastique sur le chromosome X que sur l'autosome, et diminuer la puissance de notre méthode pour détecter de la TD sexe-spécifique.

Nous avons également étudié la distribution des F_{ST} dans la population simulée sans ré-échantillonnage (voir en annexe figure 4.1) et avec un ré-échantillonnage de 100 individus (voir en annexe figure 4.2), correspondant au nombre moyen d'individus dans les populations de HapMap étudiées en chapitre 2. Si l'on compare les valeurs de F_{ST} entre ces différents échantillonnages, on remarque que plus l'échantillon pris en compte est petit, plus la variance des valeurs de F_{ST} est élevée, ce qui se traduit par des distributions plus étendue et des valeurs extrêmes plus hautes. De plus, lorsque l'on compare les distributions du chromosome X avec celles de l'autosome dans trois les types d'échantillon, on observe peu de différences au niveau des F_{ST} moyens.

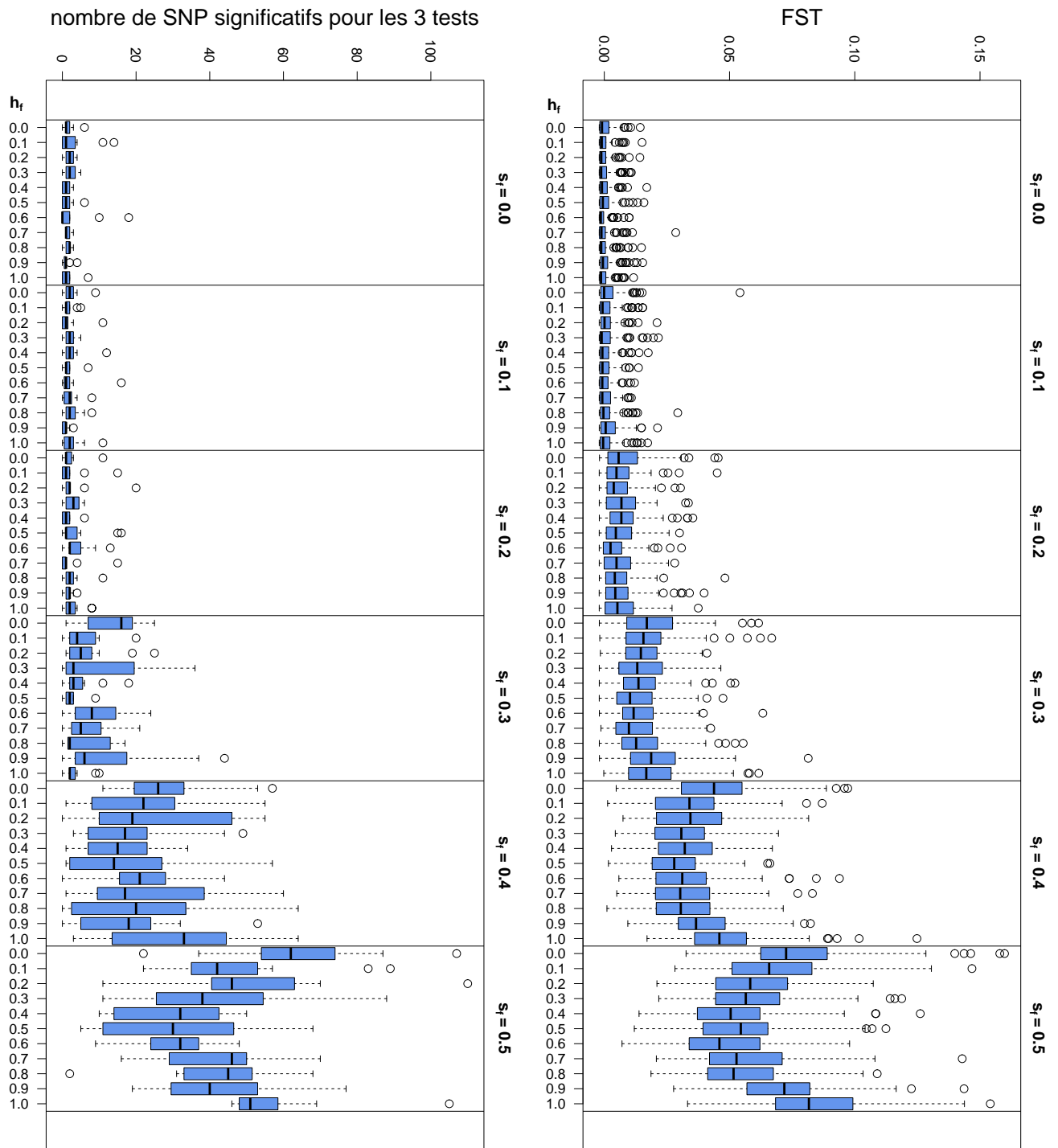


Figure 3.11 – Distributions du F_{ST} intersexuel pour le locus SA et du nombre de SNP montrant un fort signal de TD sexe-spécifique (c'est-à-dire une p-value significative pour les 3 tests effectués) pour l'**autosome**. Ces distributions sont construites sur les 10 répétitions, pour des valeurs de s_f comprises entre 0,0 et 0,5 et des valeurs de h_f comprise entre 0,0 et 1,0.

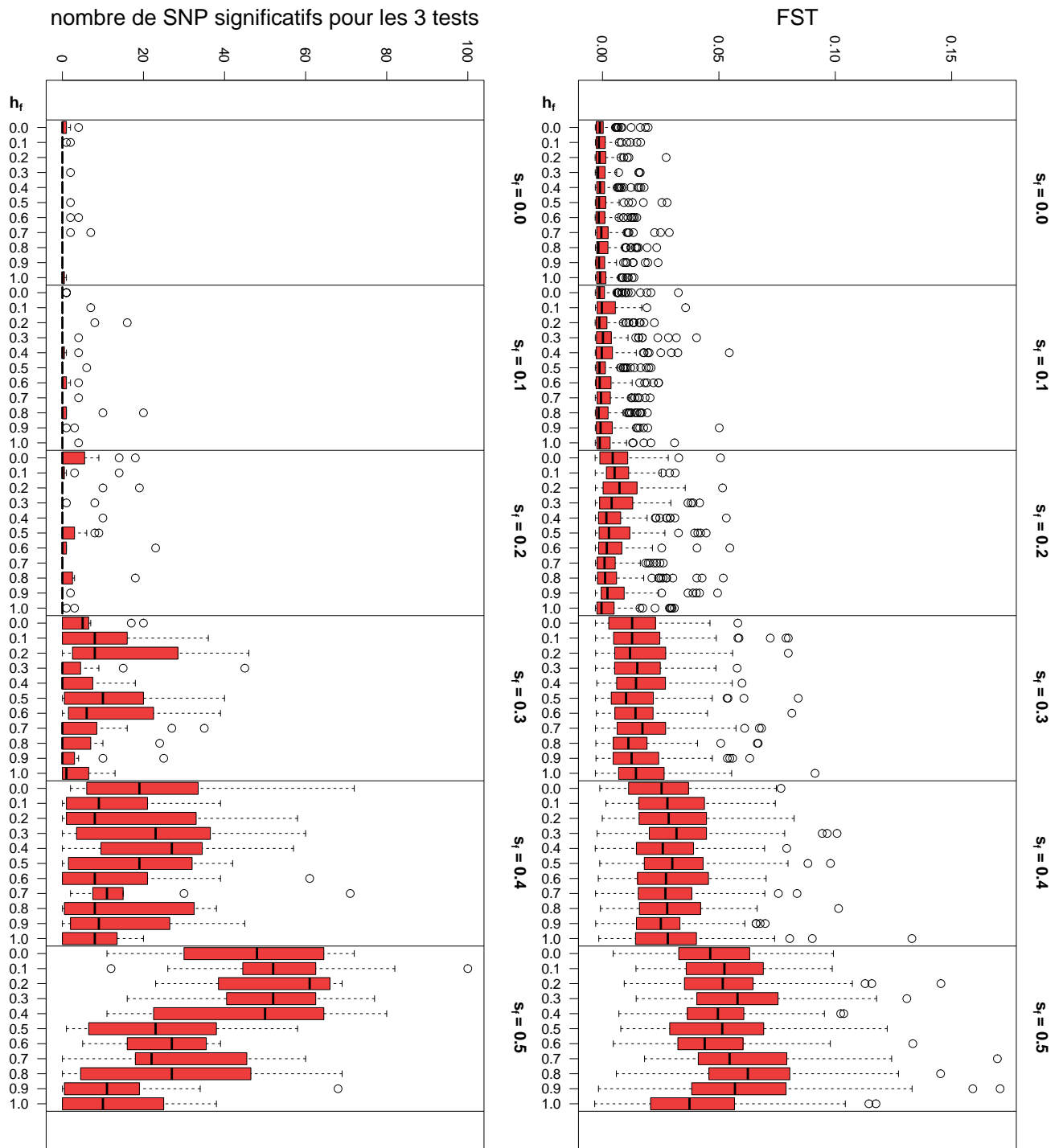


Figure 3.12 – Distributions du F_{ST} intersexuel pour le locus SA et du nombre de SNP montrant un fort signal de TD sexe-spécifique dans la région simulée (c'est-à-dire une p-value significative pour les 3 tests effectués) pour le **chromosome X**. Ces distributions sont construites sur les 10 répétitions, pour des valeurs de s_f comprises entre 0,0 et 0,5 et des valeurs de h_f comprise entre 0,0 et 1,0.

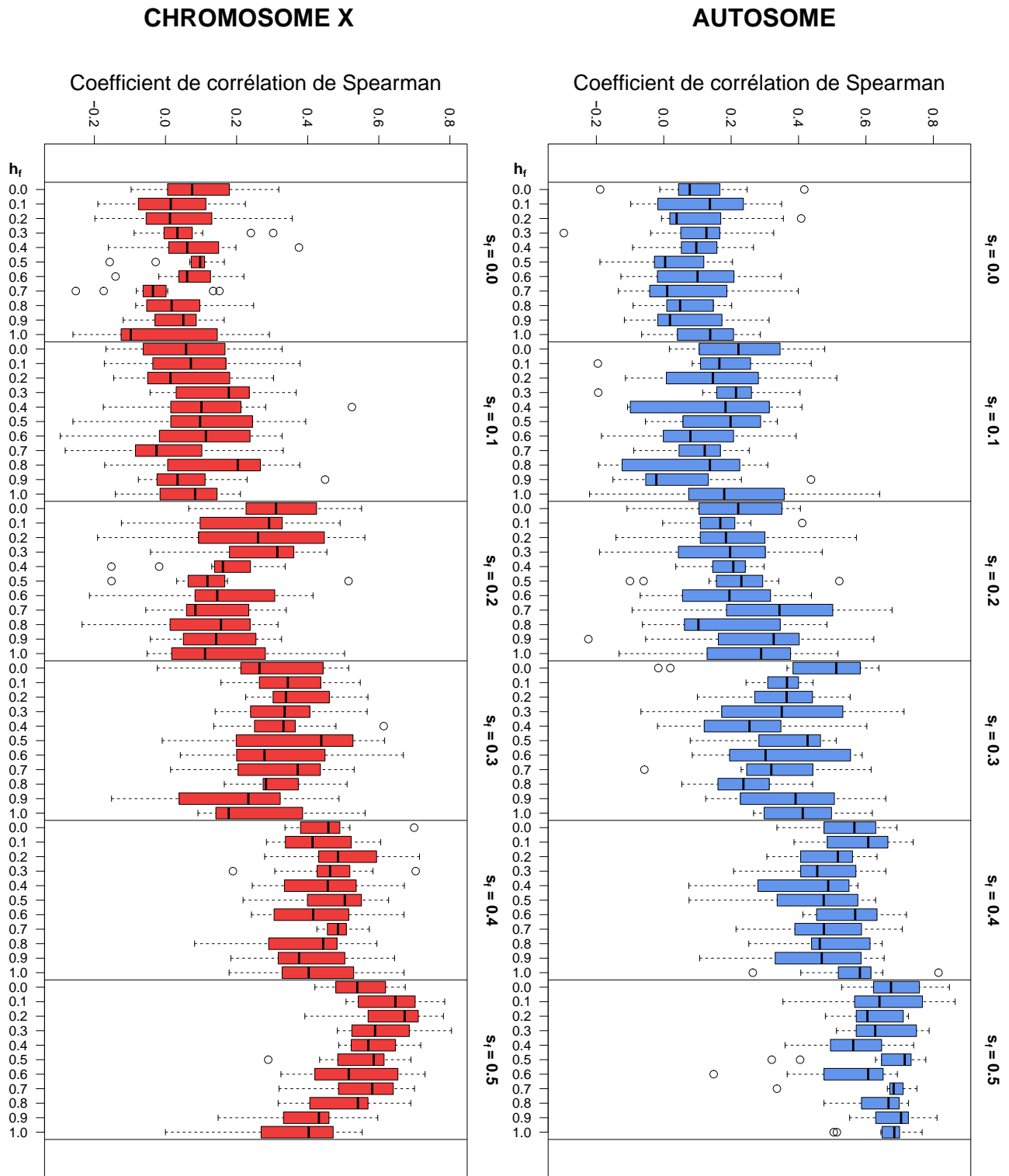


Figure 3.13 – Distributions des coefficients de corrélation de Spearman entre le ΔTD et le F_{ST} intersexuel construites sur les 10 répétitions pour des valeurs de s_f comprises entre 0,0 et 0,5 et des valeurs de h_f comprise entre 0,0 et 1,0 pour le chromosome X (en rouge) et l'autosome (en bleu).

Enfin, nous avons voulu décrire plus précisément la relation entre le F_{ST} intersexuel et la TD sexe-spécifique. Pour cela nous avons calculé un estimateur de l'intensité de la distorsion de transmission sexe-spécifique correspondant à la différence entre la proportion de A transmis chez les filles et la proportion de A transmis chez les fils :

$$\Delta TD = \left| \frac{NT_{fille}}{NT_{fille} + T_{fille}} - \frac{NT_{fils}}{NT_{fils} + T_{fils}} \right|$$

Les distributions des coefficients de corrélation de Spearman entre ΔTD et le F_{ST} intersexuel pour l'autosome et le chromosome X sont montrées figure 3.13. La corrélation a été calculée entre le F_{ST} intersexuel et le ΔTD de l'ensemble des locus de la région simulée. La corrélation entre ces deux statistiques est en grande majorité positive et augmente avec s_f . Elle est globalement toujours significative à partir de $s_f = 0,1$. En ce qui concerne la dominance, on observe la même tendance que précédemment pour l'autosome et le chromosome X : les coefficients de corrélations augmentent pour un h_f proche de 0 ou 1 sur l'autosome, et diminuent pour un h_f proche de 1 sur le chromosome X.

En résumé, nous avons construit à l'aide de ces simulations une région comportant un locus SA avec une sélection sur la viabilité et donc une distorsion de transmission sexe-spécifique. L'observation d'une forte concordance entre des F_{ST} intersexuels élevés et un signal de distorsion de transmission sexe-spécifique au niveau du locus SA nous indique que nous détectons les deux types de signaux au bon endroit, c'est-à-dire autour du locus SA. De plus, nous observons une corrélation positive entre le F_{ST} intersexuel et le ΔTD sur l'ensemble de la région. Cette corrélation, ainsi que l'intensité du signal de sélection SA et de TD sexe-spécifique, augmentent avec la force de sélection. Ainsi, le comportement des signaux des deux mécanismes modélisés est en accord avec les attendus. Ces résultats indiquent donc que notre méthode est adaptée pour détecter un signal de distorsion de transmission sexe-spécifique.

3.3.3 Détection de distorsions de transmission dans les données de GoNL

Analyse des résultats de distorsion de transmission

Nous avons utilisé notre méthode de détection de distorsion de transmission sur les données de Genome of the Netherland (GoNL). Les nombres de SNP significatifs pour chaque combinaisons de tests (SDT , TD_{fille} et TD_{fils}) sont reportés en annexe table 4.4.

Dans un premier temps, nous avons voulu vérifier que notre méthode capturerait bien l'information voulue sur des données réelles. Pour cela, nous avons utilisé deux indices, TD_{div1} et TD_{div2} . L'indice TD_{div1} estime la différence entre la déviation de la valeur attendue des proportions de A ou B, c'est-à-dire 1/2, chez les filles et celle chez les fils et permet de discriminer les deux cas de TD limitée à un sexe. La formule de TD_{div1} est reportée si dessous, et le signe de cet indice pour les différents cas de TD sont décrits en table 3.8.

$$\underline{TD_{div1}}$$

Si $Max(NT_{fils}, T_{fils}) = T_{fils}$:

$$TD_{div1} = \left| \frac{T_{fille}}{NT_{fille} + T_{fille}} - \frac{1}{2} \right| - \left| \frac{T_{fils}}{NT_{fils} + T_{fils}} - \frac{1}{2} \right|$$

Si $Max(NT_{fils}, T_{fils}) = NT_{fils}$:

$$TD_{div1} = \left| \frac{NT_{fille}}{NT_{fille} + T_{fille}} - \frac{1}{2} \right| - \left| \frac{NT_{fils}}{NT_{fils} + T_{fils}} - \frac{1}{2} \right|$$

Cas	$\left \frac{T_{fille}}{NT_{fille} + T_{fille}} - \frac{1}{2} \right $	$\left \frac{T_{fils}}{NT_{fils} + T_{fils}} - \frac{1}{2} \right $	Signe de TD_{div1}
Pas de TD	0	0	0
TD limitée aux filles	+++ ou ---	0	+++
TD limitée aux fils	0	+++	---
TD	+++	+++	0
TD sexe-spécifique	---	+++	0

Table 3.8 – Signe de TD_{div1} pour les différents cas si $Max(NT_{fils}, T_{fils}) = T_{fils}$.

L'indice TD_{div2} permet de discriminer les cas où la TD est dans la même direction chez les deux sexes et ceux où la TD est sexe-spécifique. La formule de TD_{div2} est reportée si dessous, et le signe de cet indice pour les différents cas de TD sont décrits en table 3.9.

$$\underline{TD_{div2}}$$

Si $Max(NT_{fils}, T_{fils}) = T_{fils}$:

$$TD_{div2} = \left(\frac{T_{fille}}{NT_{fille} + T_{fille}} - \frac{1}{2} \right) \times \left(\frac{T_{fils}}{NT_{fils} + T_{fils}} - \frac{1}{2} \right)$$

Si $Max(NT_{fils}, T_{fils}) = NT_{fils}$:

$$TD_{div2} = \left(\frac{NT_{fille}}{NT_{fille} + T_{fille}} - \frac{1}{2} \right) \times \left(\frac{NT_{fils}}{NT_{fils} + T_{fils}} - \frac{1}{2} \right)$$

Cas	$\left(\frac{T_{fille}}{NT_{fille} + T_{fille}} - \frac{1}{2} \right)$	$\left(\frac{T_{fils}}{NT_{fils} + T_{fils}} - \frac{1}{2} \right)$	Signe de TD_{div2}
Pas de TD	0	0	0
TD limitée aux filles	+++ ou ---	0	0
TD limitée aux fils	0	+++	0
TD	+++	+++	+++
TD sexe-spécifique	---	+++	---

Table 3.9 – Signe de TD_{div2} pour les différents cas si $Max(NT_{fils}, T_{fils}) = T_{fils}$.

En combinant ces deux indices, on peut discriminer les quatre cas suivants : une TD dans le même sens chez les deux sexes, une TD limitée aux filles, une TD limitée aux fils et une TD sexe-spécifique. Les attendus sont représentés figure 3.14.

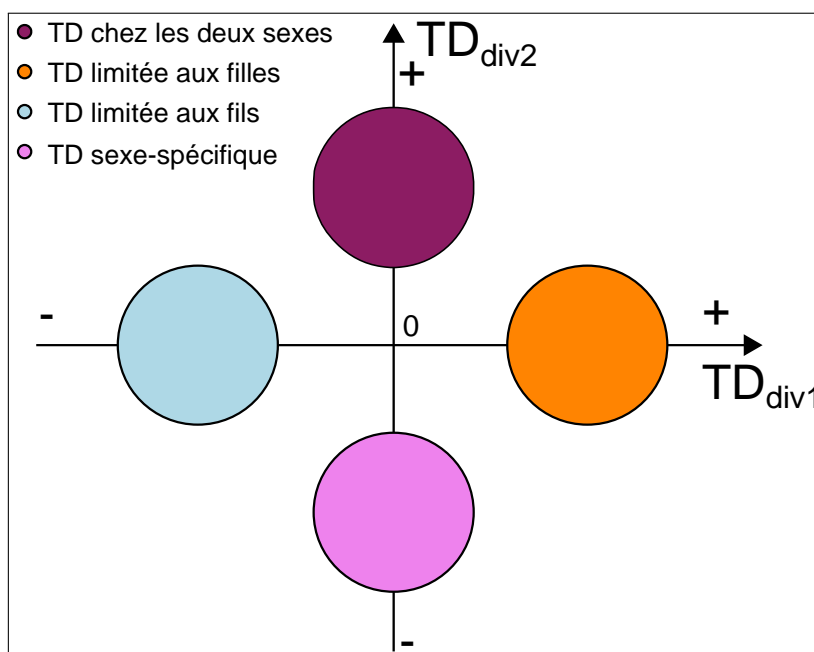


Figure 3.14 – Positions attendues pour des locus présentant les différents cas de TD, en fonction de TD_{div1} et TD_{div2}

Nous avons calculé ces deux indices sur les résultats de TD de GoNL. Les valeurs des deux indices par SNP sont représentées pour le chromosome X et le chromosome 1 en figure 3.15. On remarque que, pour les SNP de GoNL retenus pour cette analyse, les SNP montrant soit une TD limitée aux filles, soit une TD limitée aux fils, soit une TD chez les deux sexes, soit une TD sexe-spécifique (3 tests significatifs) se trouvent aux positions attendues (figure 3.14). Les SNP pour lesquels seul le test SDT est significatif, donc montrant un signal plus faible de TD sexe-spécifique, se trouvent au même niveau que les SNP pour lesquels les 3 tests sont significatifs (SDT, TD_{fille} et TD_{fils}), ce qui confirme que le même signal est capturé dans les deux cas. De plus, on remarque que les SNP ayant le test SDT et un des deux tests de TD significatifs ont en majorité un TD_{div2} inférieur à 0, ce qui indique une TD sexe-spécifique. Néanmoins, les SNP significatifs pour le test SDT et le test TD_{fille} sont distinguables des SNP significatifs pour le test SDT et le test TD_{fils} sur la figure. Cela pourrait indiquer que le TD sexe-spécifique pour ces SNP est plus fort chez un des sexes, bien qu'existant dans des directions opposées chez les deux sexes. Les tests effectués pour détecter de la TD étaient adaptés à l'hémizygotie des mâles pour le chromosome X, et nous observons des résultats très similaires entre le chromosome 1 et le chromosome X ce qui indique que notre méthode détecte le même signal sur les autosomes et le chromosome X. De plus, la même distribution est observée pour les autres autosomes. Le fait que les résultats soient concordants entre chromosomes nous indique que nous détectons bien le signal attendu.

Analyse fonctionnelle

Nous avons ensuite effectué une analyse fonctionnelle sur les gènes comportant au moins un SNP des différentes catégories : une TD dans le même sens chez les deux sexes (tests TD_{fils} et TD_{fille} significatifs), une TD limitée aux filles (test TD_{fille} significatif), une TD limitée aux fils (test TD_{fils} significatif) et une TD sexe-spécifique forte (les 3 tests significatifs) et intermédiaire (au moins le test SDT est significatif). Une correction FDR par gène pour le nombre de SNP contenus dans les gènes a été effectuée.

TD sexe-spécifique forte (les 3 tests significatifs)

Aucun cluster n'est significatif pour cette liste de gènes (9 gènes). Les termes enrichis présentent des fonctions impliquées dans la GTPase (table 3.10).

TD sexe-spécifique intermédiaire (au moins le test SDT est significatif)

Pour cette catégorie (755 gènes), 3 clusters sont significatifs, et regroupent des fonctions impliquées dans la perception sensorielle et la cognition (score d'enrichissement de 1,97), l'activité de la lipase (1,47) et la réponse au stress oxydatif (1,37) (voir en annexe

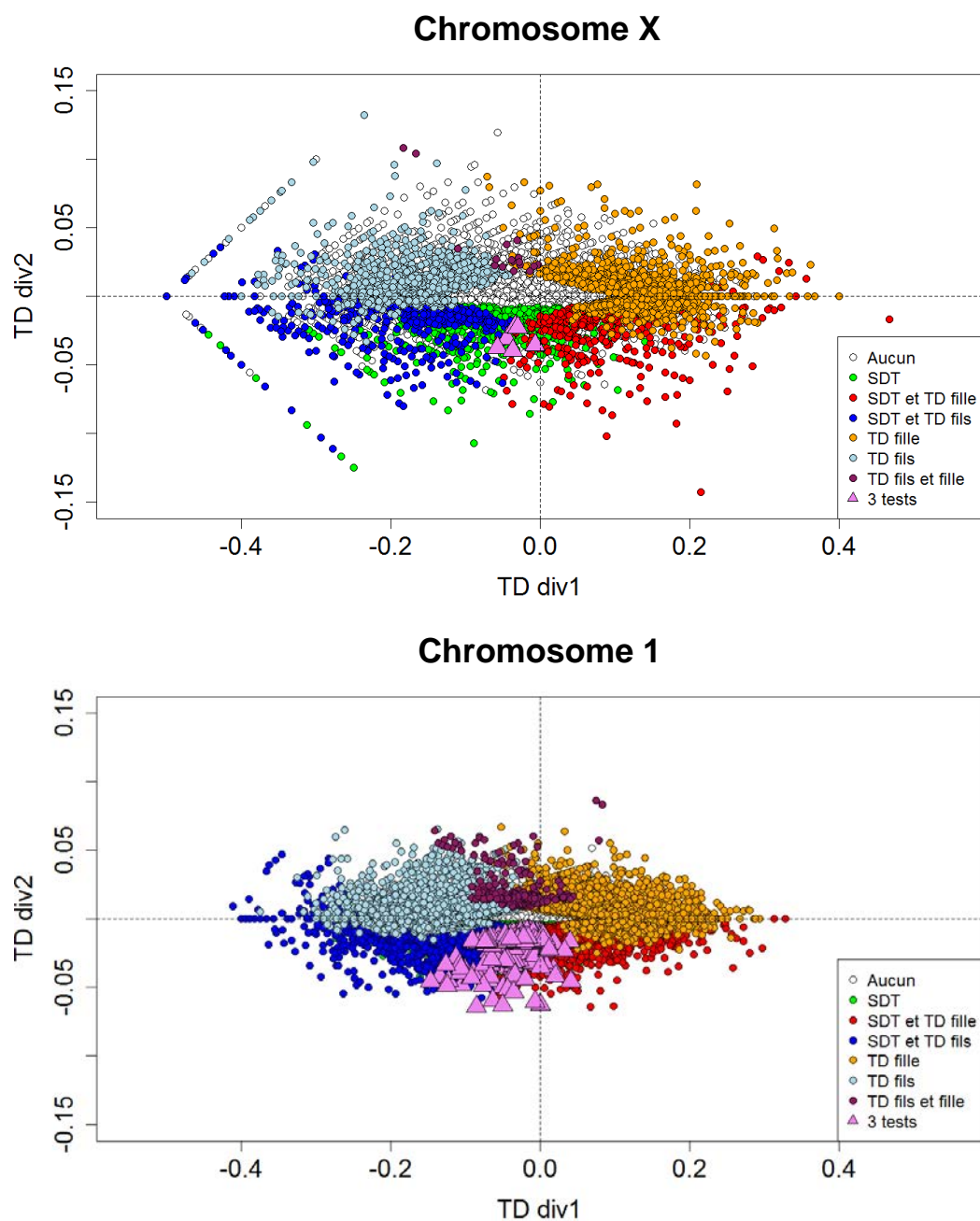


Figure 3.15 – Valeurs des deux indices, TD_{div1} et TD_{div2} , pour chaque SNP pour le chromosome X et le chromosome 1. La couleur des SNP correspond à la catégorie dans laquelle se trouve le SNP, dépendant de quel(s) test(s) (SDT, TD_{fille} et TD_{fils}) étai(en)t significatif(s) pour ce SNP.

Termes	p-value
Regulation of Ras protein signal transduction	0.031
GTPase activator activity	0.034
Regulation of small GTPase mediated signal transduction	0.037
Small GTPase regulator activity	0.042

Table 3.10 – Termes enrichis pour les gènes comportant au moins un SNP montrant une **TD sexe-spécifique forte** (significatif après correction FDR par gène pour les tests SDT, TD_{fille} et TD_{fils}), classés selon leur p-value (9 gènes).

table 4.5). D'autres termes non inclus dans les clusters sont significatifs, regroupés dans la table 3.11. De façon intéressante, deux termes concernent la perception du goût et deux autres la réponse immunitaire.

TD limitée aux fils (test TD_{fils} significatif)

Trois clusters sont significatifs pour cette catégorie (431 gènes), et réfèrent à l'organisation et l'assemblage de la chromatine (score d'enrichissement de 3,16), à la régulation et traduction des ARN messagers (1,43), et aux vésicules membranaires (1,41) (voir en annexe table 4.6). Les termes enrichis, excluant les termes des clusters, réfèrent à des fonctions potentiellement intéressantes dans un contexte de distorsion de transmission : 4 d'entre eux sont impliqués dans la gamétogénèse, dont 2 plus particulièrement dans la spermatogénèse. On trouve aussi deux termes impliqués dans la régulation de la transcription, un terme dans les processus cataboliques des lipides et un terme dans la régulation négative de l'expression génique (table 3.12).

TD limitée aux filles (test TD_{fille} significatif)

Au sein de cette catégorie (527 gènes), 2 clusters regroupant des fonctions associées aux membranes (score d'enrichissement de 1.62) et au repliement des protéines après la traduction (1.33) sont significatifs (voir en annexe table 4.7). De plus, 6 termes en dehors des clusters sont impliqués dans le développement embryonnaire. Deux autres termes sont impliqués dans les membranes mitochondriales, un terme dans le Complexe Majeur d'Histocompatibilité (CMH), et un terme dans la réponse comportementale à la douleur (table 3.13).

TD chez les deux sexes (les tests TD_{fille} et TD_{fils} significatifs)

Aucun cluster significatif n'a été trouvé pour cette catégorie. Néanmoins, malgré le peu de gènes identifiés (3 gènes), de nombreux termes sont significativement enrichis. Ces termes concernent l'organisation de la chromatine, des nucléosomes et la compaction de l'ADN (voir en annexe table 4.8).

Termes	p-value
Sensory perception of taste	0.001
Cation :amino acid symporter activity	0.004
Taste receptor activity	0.008
Glycine transport	0.011
Glycine transmembrane transporter activity	0.011
Symporter activity	0.014
Solute :cation symporter activity	0.017
Amine transmembrane transporter activity	0.017
5'-nucleotidase activity	0.019
Amino acid transmembrane transporter activity	0.024
Nucleotidase activity	0.024
Cyclic nucleotide binding	0.025
Protein-DNA complex	0.026
Proton transport	0.027
Hydrogen transport	0.030
Plasma membrane	0.031
Regulation of leukocyte chemotaxis	0.036
NAD or NADH binding	0.041
Positive regulation of acute inflammatory response	0.042
Cofactor binding	0.045
Nucleotide binding	0.047

Table 3.11 – Termes enrichis pour les gènes comportant au moins un SNP significatif après correction FDR par gène pour au moins le test SDT, montrant une **TD sexe-spécifique**, classés selon leur p-value (755 gènes). Les termes présents dans les clusters (voir en annexe table 4.5) ont été enlevés de la liste. Deux termes impliqués dans la perception du goût (en bleu) et deux termes impliqués dans la réponse immunitaire (en rose) ont été surlignés.

Termes	p-value
Negative regulation of macromolecule metabolic process	0.011
Cell surface	0.012
DNA binding	0.016
Voltage-gated channel activity	0.017
Voltage-gated ion channel activity	0.017
External side of plasma membrane	0.021
Oogenesis	0.025
Eukaryotic cell surface binding	0.032
Gamete generation	0.033
RNA binding	0.033
Male gamete generation	0.034
Spermatogenesis	0.034
Negative regulation of transcription, DNA-dependent	0.037
Voltage-gated cation channel activity	0.038
Cellular lipid catabolic process	0.038
Transcription repressor activity	0.041
Negative regulation of RNA metabolic process	0.041
Identical protein binding	0.042
Negative regulation of gene expression	0.044
piRNA binding	0.050
Hydrogen :amino acid symporter activity	0.050

Table 3.12 – Termes enrichis pour les gènes comportant au moins un SNP significatif après correction FDR par gène pour le test TD_{fils} , montrant une **TD limitée aux fils**, classés selon leur p-value (431 gènes). Les termes présents dans les clusters (voir annexe table 4.6) ont été enlevés de la liste. Les termes référant à la gamétogénèse (en rose), dans la régulation de la transcription (en vert), et le terme dans les processus cataboliques des lipides (en bleu) et dans la régulation négative de l’expression génique (en orange) ont été surlignés.

Termes	p-value
Actin cytoskeleton	0.003
Glutamate receptor activity	0.004
Chordate embryonic development	0.005
Embryonic development ending in birth or egg hatching	0.006
Mitochondrial outer membrane translocase complex	0.008
Behavioral response to pain	0.012
Protein transport	0.013
Establishment of protein localization	0.015
Dopamine receptor binding	0.015
MHC protein binding	0.016
Nerve development	0.016
Protein targeting	0.018
Cranial nerve morphogenesis	0.022
Protein localization	0.024
Calcium-dependent protein binding	0.030
Channel activity	0.031
Ionotropic glutamate receptor complex	0.032
Macromolecular complex subunit organization	0.032
Passive transmembrane transporter activity	0.032
Response to mechanical stimulus	0.032
Embryonic skeletal system morphogenesis	0.034
Gated channel activity	0.034
Calcium ion binding	0.036
Mitochondrial outer membrane	0.036
In utero embryonic development	0.036
Ion channel activity	0.038
L-glutamine transmembrane transporter activity	0.042
Glutamine transport	0.042
Cellular macromolecular complex subunit organization	0.044
Substrate specific channel activity	0.047
Sarcomere	0.049
Macromolecular complex assembly	0.049

Table 3.13 – Termes enrichis pour les gènes comportant au moins un SNP significatif après correction FDR par gène pour le test TD_{fille} , montrant une **TD limitée aux filles**, classés selon leur p-value (527 gènes). Les termes présents dans les clusters (voir en annexe table 4.7) ont été enlevés de la liste. Les termes impliqués dans le développement embryonnaire (en bleu), les membranes mitochondriales (en vert), le CMH (en orange), et dans la réponse comportementale à la douleur (en rose) ont été surlignés.

Analyse de la corrélation entre le F_{ST} intersexuel et la distorsion de transmission sexe-spécifique

Comparaison entre les différents signaux de TD et le F_{ST} chez les enfants

Nous avons ensuite calculé la corrélation entre le ΔTD (voir section 3.3.2) et le F_{ST} intersexuel chez les enfants pour chacun des chromosomes (figure 3.16). Pour la grande majorité des chromosomes (16 sur 22), la corrélation est positive et significative. De plus, la corrélation sur le chromosome X est deux fois plus élevée que pour les autosomes. Seuls deux chromosomes, le 15 et le 18, montrent une corrélation négative significative.

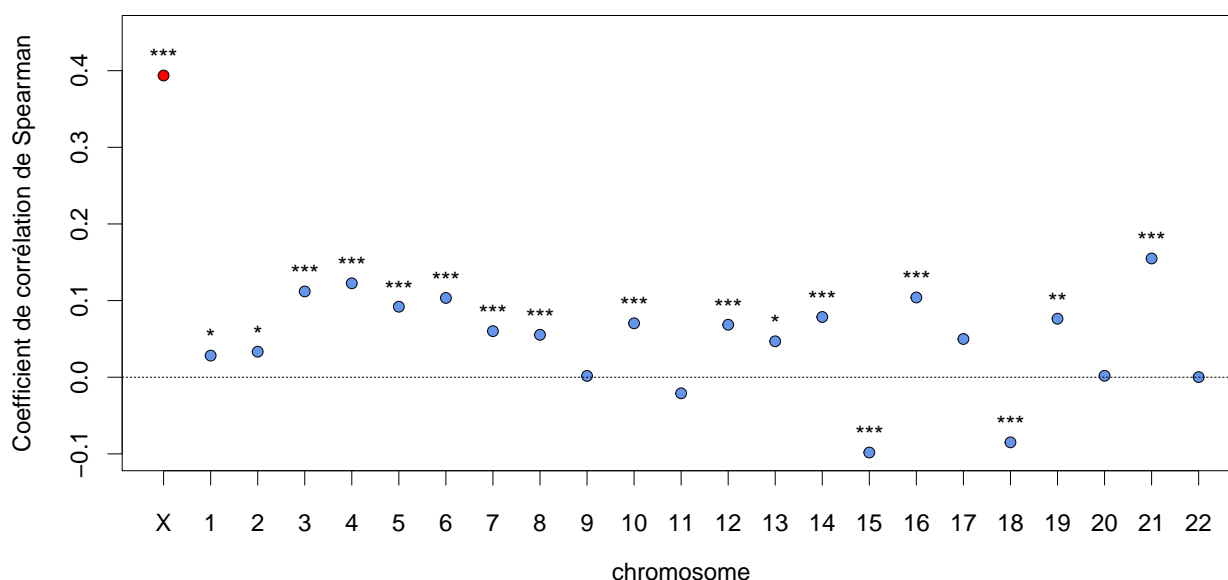


Figure 3.16 – Coefficient de corrélation de Spearman entre le ΔTD et le F_{ST} intersexuel chez les enfants pour le chromosome X (en rouge) et les autosomes (en bleu). Les niveaux de significativité de la corrélation sont représentés (p-value : * < 0.05, ** < 10^{-2} , *** < 10^{-3}).

Dans un second temps, afin d'étudier plus en détail la relation entre les différences de fréquences alléliques entre les sexes et la distorsion de transmission sexe-spécifique, nous avons comparé les distributions de F_{ST} intersexuel dans les différents cas de distorsion de transmission (figure 3.17). Pour les deux chromosomes représentés, le chromosome X et le chromosome 1, on remarque que les SNP montrant un signal de TD sexe-spécifique présentent des distributions de F_{ST} déviées vers des valeurs plus élevées de F_{ST} que les SNP montrant un signal de TD limitée à un sexe ou chez les deux sexes dans le même

sens. De plus, les distributions de F_{ST} pour les SNP de la catégorie SDT et les SNP de la catégorie SDT- TD_{fille} ont des distributions similaires, bien que ces premiers montrent un pic de densité pour un F_{ST} de 0,05. Les SNP de la catégorie SDT- TD_{fils} ne montrent pas la même distribution que les SNP de la catégorie SDT et les SNP de la catégorie SDT- TD_{fille} pour le chromosome X, contrairement au chromosome 1, ce qui pourrait être dû à une différence de puissance statistique entre le chromosome X et les autosomes. Cette tendance est la même pour tous les autosomes.

Nous avons ensuite comparé la liste de SNP significatifs pour le F_{ST} intersexuel chez les enfants, avant et après correction FDR par gène, avec les listes de SNP significatifs pour les différents tests de TD (table 3.14). Parmi les SNP montrant un F_{ST} significatif, 29% d'entre eux en moyenne sont également significatifs pour un TD sexe-spécifique (significatif pour au moins pour le test SDT). Après correction FDR par gène sur la p-value du test de significativité du F_{ST} intersexuel, 51% des SNP avec un F_{ST} significatif montrent également un signal de TD sexe-spécifique. Moins de 5% des SNP avec un F_{ST} significatif, avant et après FDR, montrent un signal pour chacun des autres cas de TD, dont 0% pour le TD chez les deux sexes (table 3.14).

Cas	proportion avant FDR	proportion après FDR
TD sexe-spécifique	0.293	0.510
TD limitée aux filles	0.040	0.037
TD limitée aux fils	0.049	0.041
TD chez les deux sexes	0.000	0.000
Aucun	0.610	0.406

Table 3.14 – Proportion de SNP montrant un F_{ST} intersexuel significatif chez les enfants avant et après FDR, ainsi qu'un signal de TD spécifique, TD limitée aux filles, TD limitée aux fils, TD chez les deux sexes, et aucun signal de TD.

Comparaison entre les différents signaux de TD et le F_{ST} intersexuel chez les parents

Nous avons ensuite effectué les mêmes analyses sur le F_{ST} intersexuel chez les parents. La corrélation entre le ΔTD et le F_{ST} intersexuel n'est significativement positive que pour 5 autosomes. De plus, un très faible pourcentage de SNP ayant un F_{ST} intersexuel significatif chez les parents montrent également un signal de TD sexe-spécifique (3,6% en moyenne avant FDR). Cette tendance se confirme en regardant les distributions de F_{ST} dans les différents cas de distorsion de transmission (figure 3.18). En effet, les SNP montrant un signal de TD sexe-spécifique n'ont pas des F_{ST} plus élevés chez les parents

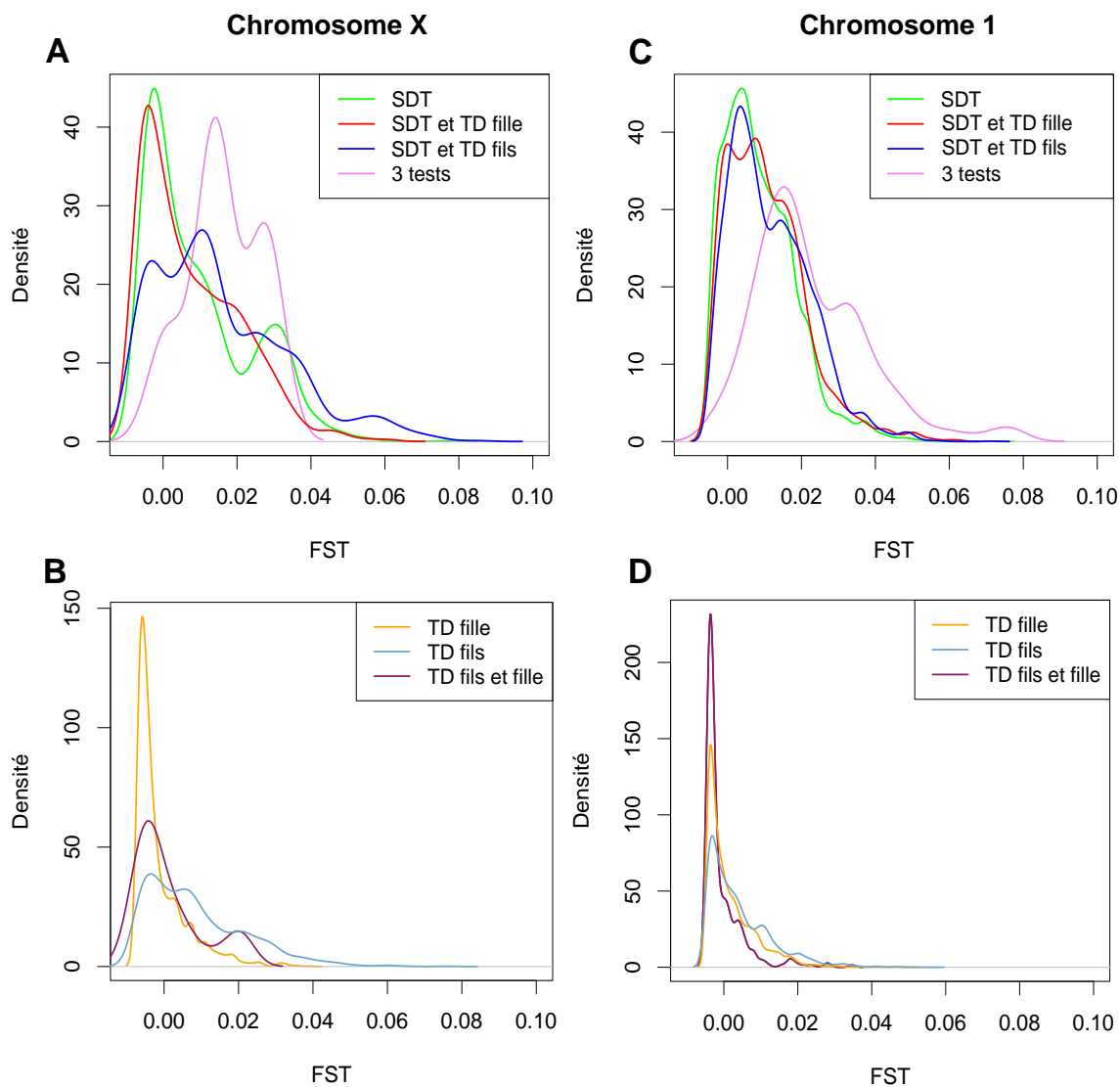


Figure 3.17 – Distributions des valeurs de F_{ST} intersexuel pour les différentes catégories de SNP montrant des signaux de distorsion de transmission. Les distributions des F_{ST} intersexuels pour les SNP montrant un signal de TD sexe-spécifique sont représentées en **A-** pour le chromosome X et **C-** pour le chromosome 1 et pour les SNP montrant un signal de TD limitée à un sexe ou chez les deux sexes, les distributions sont représentées en **B-** pour le chromosome X et en **D-** pour le chromosome 1.

comparés aux SNP montrant d'autres signaux de TD.

Le F_{ST} intersexuel chez les enfants est-il extrême compte tenu des génotypes de leurs parents ?

Afin de pouvoir comparer les F_{ST} intersexuels des parents et des enfants, nous avons calculé une nouvelle p-value pour les F_{ST} intersexuels observés chez les enfants, tenant compte des fréquences alléliques observées chez les parents.

Étude de la distribution du signal de F_{ST} intersexuel chez les enfants

Tout d'abord, nous avons comparé la liste des SNP ayant un F_{ST} significatif chez les enfants à l'aide de la méthode utilisée dans les analyses précédentes (c'est-à-dire un test exact de Fisher sur les comptes alléliques), avec la liste des SNP ayant un F_{ST} significatif chez les enfants compte-tenu du génotype de leurs parents. Plus de 80% des SNP ayant un F_{ST} significatif chez les enfants compte-tenu des génotypes de leur parents ont aussi un F_{ST} significatif selon le test exact de Fisher, avec pour exception le chromosome 2 (63%) et le chromosome X (60%).

Nous avons ensuite effectué une correction FDR par gène sur cette nouvelle p-value, afin d'obtenir les proportions de *genic* SD SNP par chromosome (figure 3.19). On remarque que le chromosome X a une proportion de *genic* SD SNP significativement plus élevée que 9 autosomes. La tendance observée chez les parents en figure 3.4 est donc un peu moins marquée chez les enfants.

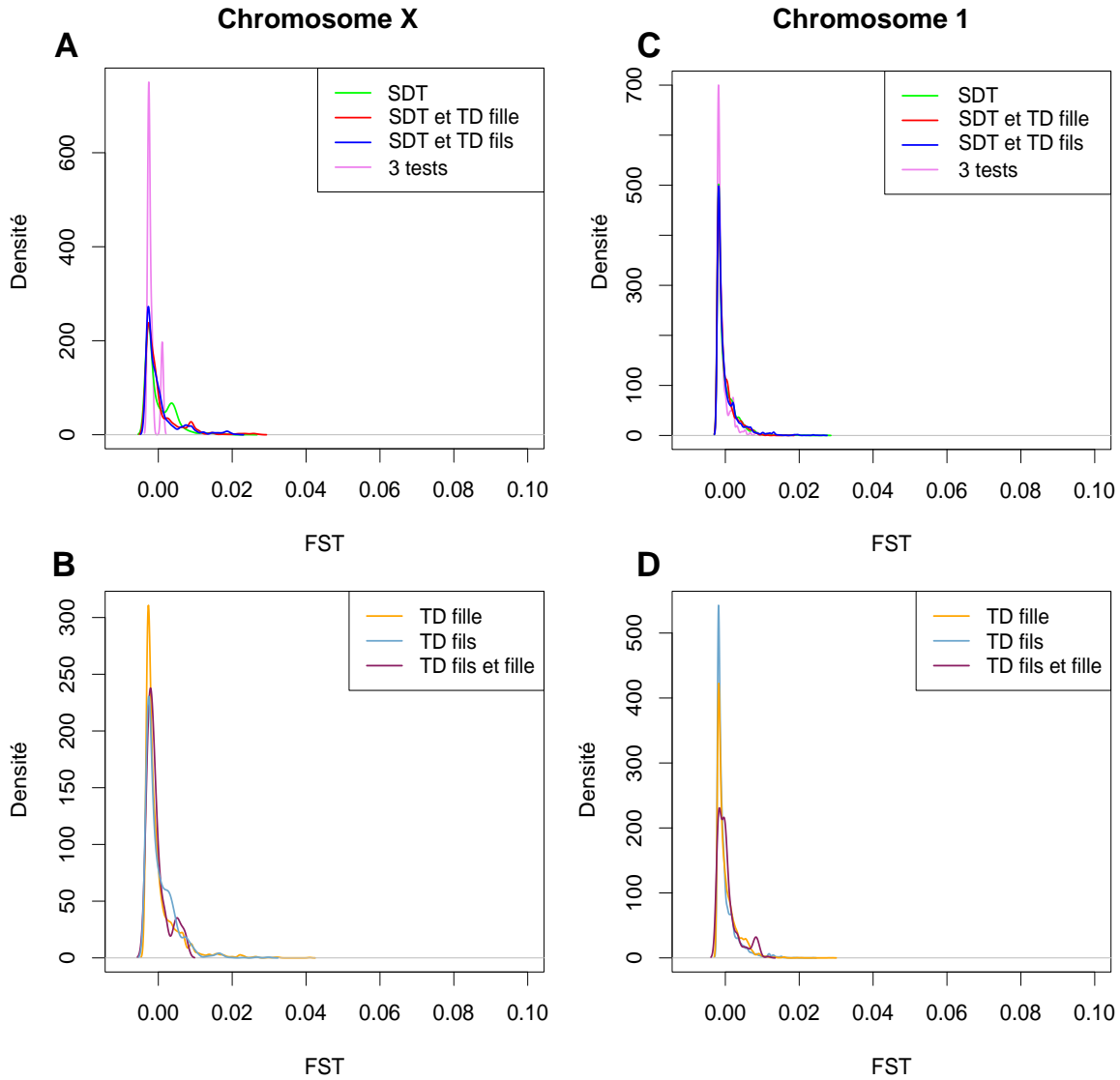


Figure 3.18 – Distributions des valeurs de F_{ST} intersexuel chez les parents pour les différentes catégories de SNP montrant des signaux de distorsion de transmission. Les distributions des F_{ST} intersexuels chez les parents pour les SNP montrant un signal de TD sexe-spécifique sont représentées en **A-** pour le chromosome X et **C-** pour le chromosome 1 et pour les SNP montrant un signal de TD limitée à un sexe ou chez les deux sexes, les distributions sont représentées en **B-** pour le chromosome X et en **D-** pour le chromosome 1.

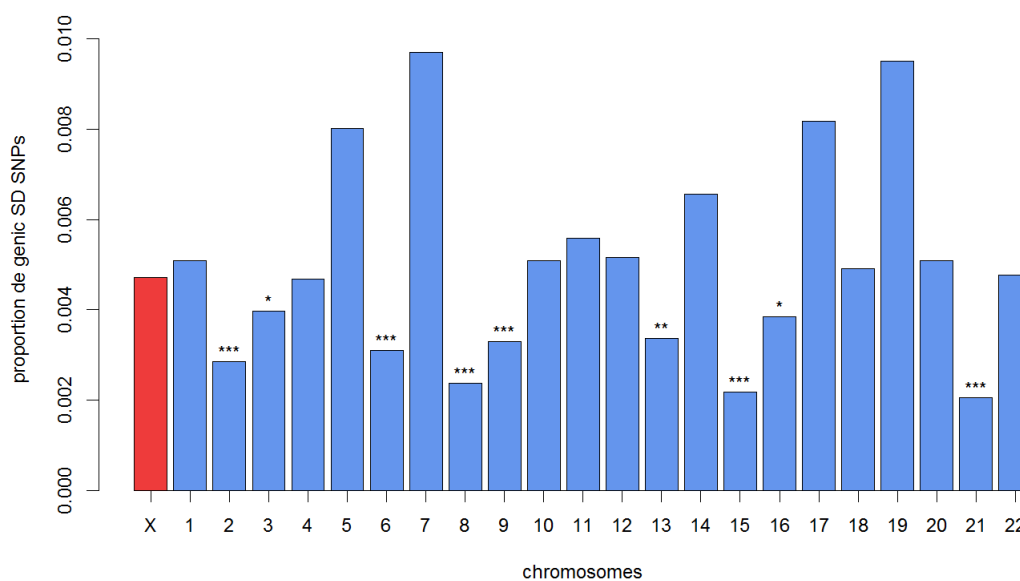


Figure 3.19 – Proportion de *genic* SD SNP par chromosome des enfants, la significativité des différences de fréquences alléliques a été calculée en prenant en compte le génotype des parents. Les niveaux de significativité du test exact de Fisher effectué entre le chromosome X (en rouge) et les autosomes (en bleu) sont représentés (p-value : * < 0.05, ** < 10^{-2} , *** < 10^{-3}).

Analyse d'enrichissement fonctionnel

Nous avons ensuite effectué une analyse d'enrichissement fonctionnel sur la liste de gènes comportant au moins un *genic* SD SNP. Trois clusters fonctionnels significatifs ont été trouvés : l'activité des chemokines (score d'enrichissement de 2,29), la fraction cellulaire (1,69) et la perception sensorielle des odeurs (1,57) (voir en annexe table 4.9). Parmi les autres termes significativement enrichis (table 3.15), nous avons trouvés plusieurs termes référant à la régulation de l'activité des kinases, aux processus métaboliques des catécholamines, et à la croissance des neurones.

Nous avons ensuite comparé la liste de SNP ayant un F_{ST} intersexuel significatif chez les enfants après notre méthode de ré-échantillonnage des enfants, avec la liste de SNP ayant un F_{ST} intersexuel significatif chez les parents. En moyenne pour chaque chromosome, seuls 4,6% des SNP présents dans la première liste le sont aussi dans la deuxième, ce qui indiquerait qu'il y a peu de correspondance entre les F_{ST} intersexuels des enfants et ceux des parents.

Nous avons ensuite comparé la liste des gènes qui comportent au moins un SNP avec un F_{ST} intersexuel significatif chez les enfants, avec notre méthode de ré-échantillonnage, et

Termes	p-value
Anchored to plasma membrane	0.006
Behavioral response to ethanol	0.008
Regulation of protein kinase activity	0.017
Regulation of transferase activity	0.020
Regulation of kinase activity	0.024
Cyclic nucleotide-gated ion channel activity	0.024
Intracellular cyclic nucleotide activated cation channel activity	0.024
SH3/SH2 adaptor activity	0.032
Catecholamine metabolic process	0.035
Catechol metabolic process	0.035
Diol metabolic process	0.035
Dopamine metabolic process	0.036
Growth cone	0.037
Phenol metabolic process	0.039
Site of polarized growth	0.040
Oxidoreductase activity, oxidizing metal ions	0.040
ncRNA metabolic process	0.046
Plasma membrane	0.048

Table 3.15 – Termes enrichis pour les gènes comportant au moins un SNP significatif après correction FDR par gène pour la nouvelle p-value du F_{ST} chez les enfants, classés selon leur p-value (935 gènes). Les termes présents dans les clusters (voir en annexe table 4.9) ont été enlevés de la liste. Les termes impliqués dans la régulation de l’activité des kinases (en orange), les processus métaboliques des catécholamines (en vert), la croissance des neurones (en rose) ont été surlignés.

la liste des gènes qui comportent au moins un SNP avec un F_{ST} intersexuel significatif chez les parents. En effet, même si les SNP significatifs sont différents, les mêmes gènes peuvent être impliqués dans des différences de fréquences alléliques entre les sexes. Nous avons trouvés 39 gènes en commun, et avons effectué une analyse d’enrichissement fonctionnel sur cette liste. Un cluster fonctionnel, la liaison à l’ATP, est significativement enrichi (score d’enrichissement de 1,56, voir en annexe table 4.9). Les autres termes enrichis, présentés en table 3.16, comprennent notamment l’activité des hormones.

Parmi les 39 gènes montrant un F_{ST} intersexuel significatif chez les parents et chez les enfants après ré-échantillonnage, 25 gènes possèdent au moins un SNP avec une p-value significative pour le test SDT. Néanmoins, seuls 12 d’entre eux ont toujours une p-value significative pour le test SDT après correction FDR par gène.

Ces 12 gènes sont très proches les uns des autres, et forment un cluster dans une région du chromosome 17 située entre 61,7 Mb et 61,95 Mb (figure 3.20). Parmi ces gènes, 2 d’entre eux, MAP3K3 et STRADA, sont impliqués dans une activité de kinase.

Termes	p-value
Hormone activity	0.011
Positive regulation of protein modification process	0.028
ncRNA metabolic process	0.041
Positive regulation of cellular protein metabolic process	0.042
Positive regulation of protein metabolic process	0.045

Table 3.16 – Termes enrichis pour les gènes en commun entre la liste de SD genes chez les enfants et celle chez les parents, classés selon leur p-value (39 gènes). Les termes présents dans les clusters (voir en annexe table 4.9) ont été enlevés de la liste. Un terme impliqué dans l’activité des hormones (en bleu) a été surligné.

DDX42 est un membre de la famille DEAD box, qui serait impliquée dans l’embryogenèse, la spermatogenèse, la croissance et la division cellulaire. FTSJ3 serait impliqué dans l’embryogenèse, PSMC5 code un protéasome et SMARCD2 est associé à une activité d’hélicase, et serait impliqué dans la régulation de la transcription de certains gènes en agissant sur la structure de la chromatine. TCAM1P est devenu un pseudogène suite à de nombreuses duplications, l’homologue chez la souris a une expression spécifique aux testicules et jouerait un rôle dans les interactions entre les cellules souches et les cellules de Sertoli.

De façon intéressante, cette région contient un cluster de 5 gènes membres de la famille des somatotrophines et prolactines. Trois d’entre eux, CSH2, CSHL1 et GH1, sont significatifs pour le F_{ST} intersexuel chez les enfants et chez les parents, ainsi que pour le test SDT. Les gènes CSH2 et CSHL1 codent pour des hormones de croissance et sont exprimés majoritairement dans le placenta tandis que le gène GH1 est exprimé dans l’hypophyse. Le gène CSH1, qui fait partie de cette famille, présente un signal de TD sexe-spécifique et un F_{ST} intersexuel significatif pour les enfants, mais pas pour les parents. Ces gènes sont responsables de l’enrichissement de la fonction de l’activité des hormones (table 3.16).

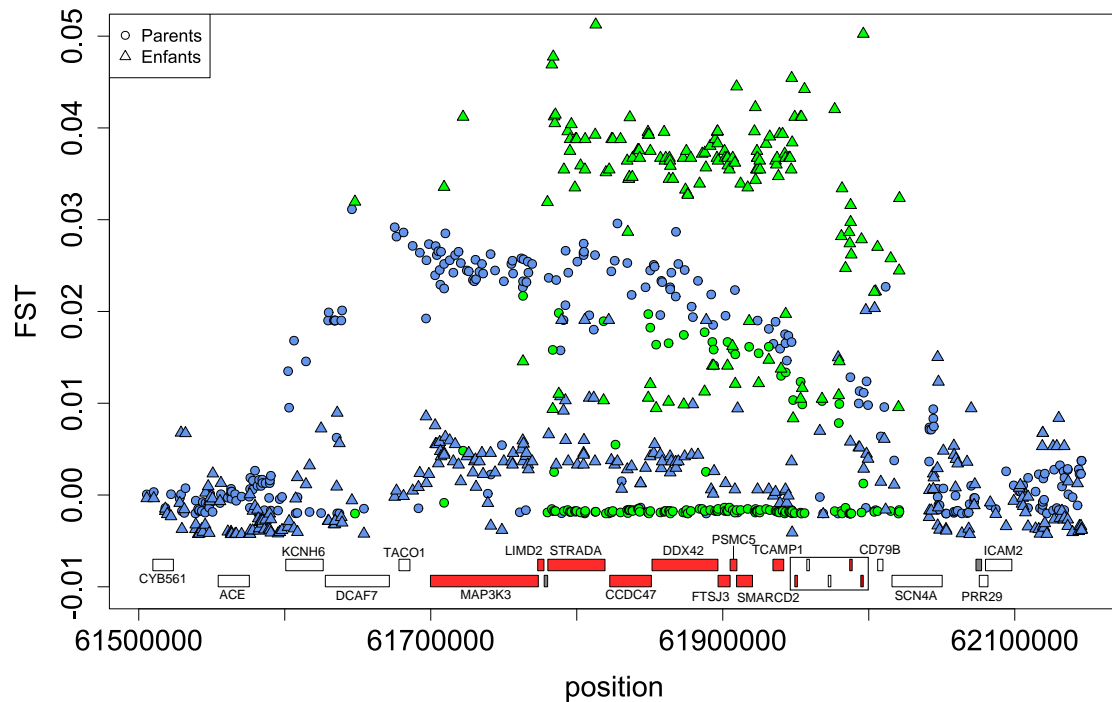


Figure 3.20 – F_{ST} intersexuel chez les parents et chez les enfants pour la région du chromosome 17 comprise entre 61,7 Mb et 61,95 Mb. Chaque locus est représenté deux fois, une fois en fonction de son F_{ST} chez les parents (les cercles), et une fois en fonction de son F_{ST} chez les enfants (les triangles). Les locus en vert ont une p-value significative pour le test SDT, les locus en bleu ont une p-value non significative pour le test SDT. Les positions des gènes sont renseignées par des rectangles en bas de la figure. Dans l’encadré se trouvent, dans l’ordre : CSH2, GH2, CSH1, CSHL1 et GH1. Les gènes en rouge présentent des locus significatifs pour le F_{ST} chez les parents et le F_{ST} chez les enfants.

Sur la figure 3.20, on remarque que les locus pour lesquels le test SDT est significatif présentent en majorité un F_{ST} faible chez les parents, mais ont un F_{ST} élevé chez les enfants. Par contre, les SNP ayant un F_{ST} élevé chez les parents ont un F_{ST} intermédiaire chez les enfants, et ne sont pas significatifs pour le test SDT. Cette observation est en accord avec le peu de concordance entre le F_{ST} chez les parents et le F_{ST} chez les enfants. Cette configuration particulière pourrait être due à la présence d’haplotype ségrégeant dans la population.

Pour déterminer si cela est le cas, nous avons divisé la région en deux haplotypes, présentés en figure 3.21. Pour définir les haplotypes, nous avons dans un premier temps différencié les SNP des deux haplotypes avec à une limite de $F_{ST} \geq 0.08833$ chez les parents pour l’haplotype 1 et $F_{ST} < 0.08833$ chez les parents pour l’haplotype 2. Des

analyses complémentaires seraient nécessaires pour déterminer des haplotypes précis. Le LD a ensuite été calculé chez les parents grâce au r^2 sur l'intégralité de la région (figure 3.22), puis sur l'haplotype 1 (figure 3.23a) et l'haplotype 2 (figure 3.23b). On observe que le LD est beaucoup moins élevé lorsque la région entière est prise en compte, en comparaison au LD pour chaque haplotype. Cela nous indique qu'il y aurait effectivement plusieurs haplotypes ségrégeant dans la population pour cette région.

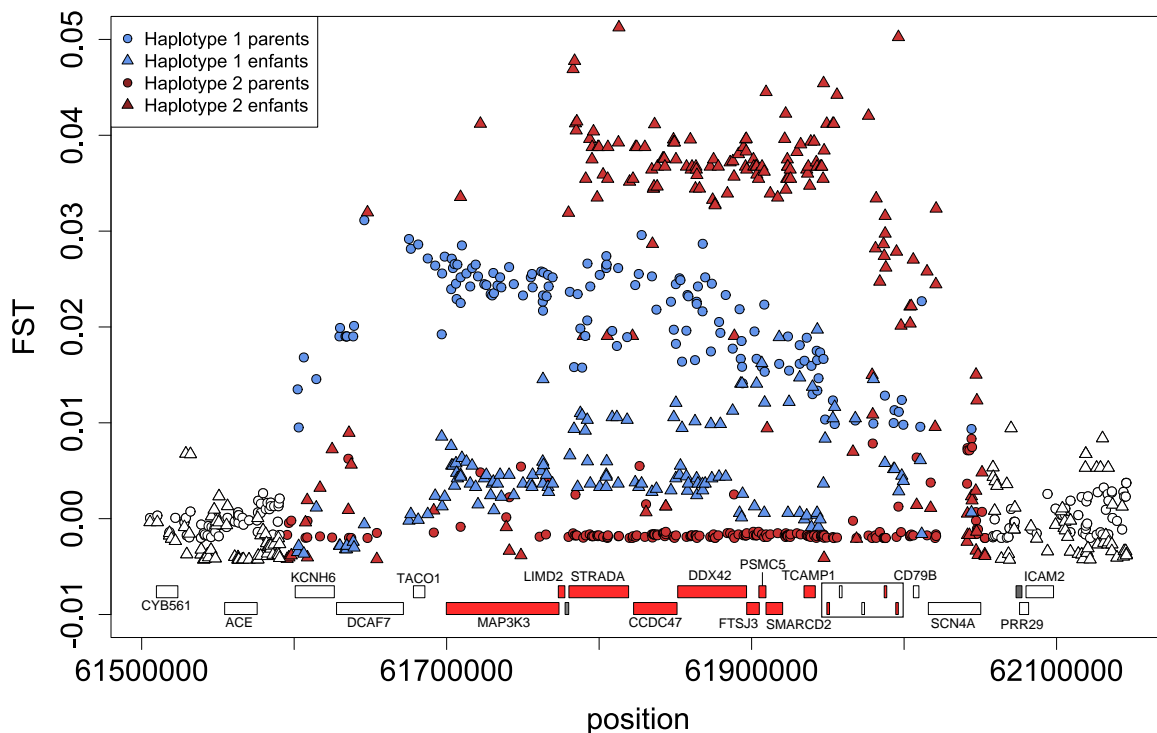


Figure 3.21 – Haplotypes testés pour la région du chromosome 17 comprise environ entre 61,50 Mb et 62,15 Mb. L'haplotype 1 est représenté en bleu et l'haplotype 2 en rouge.

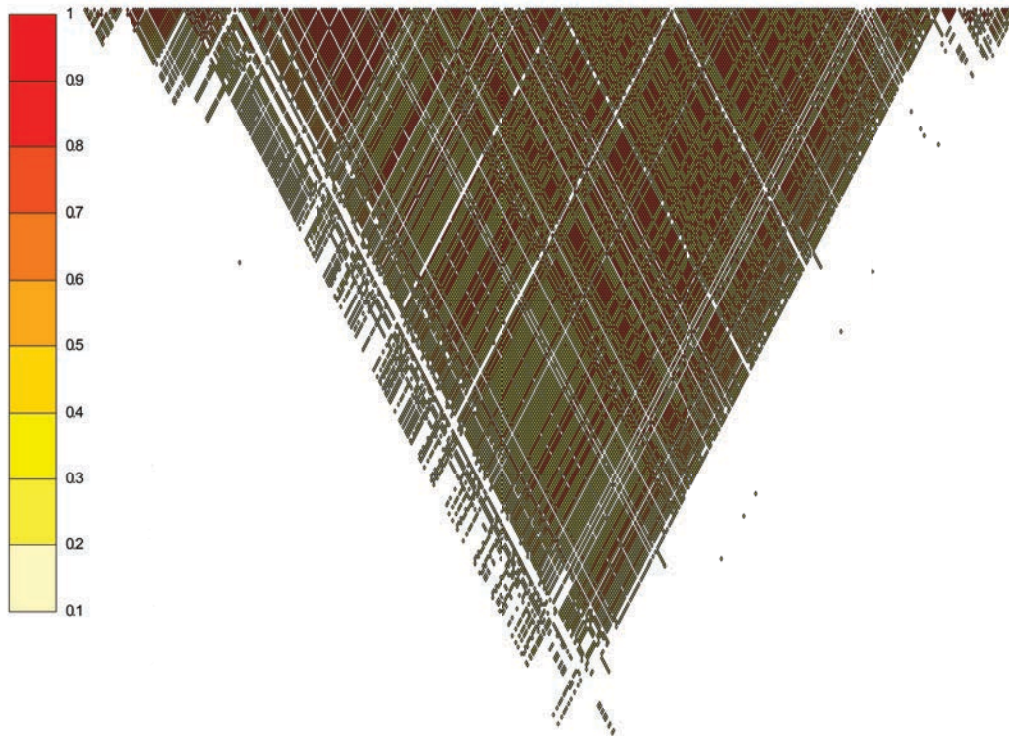
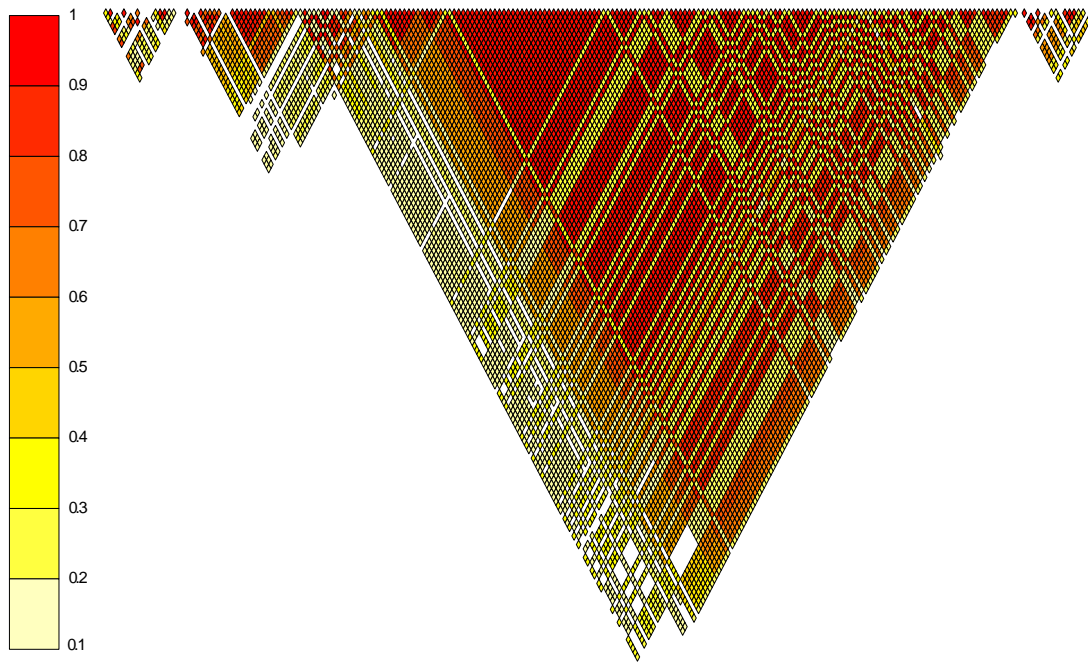
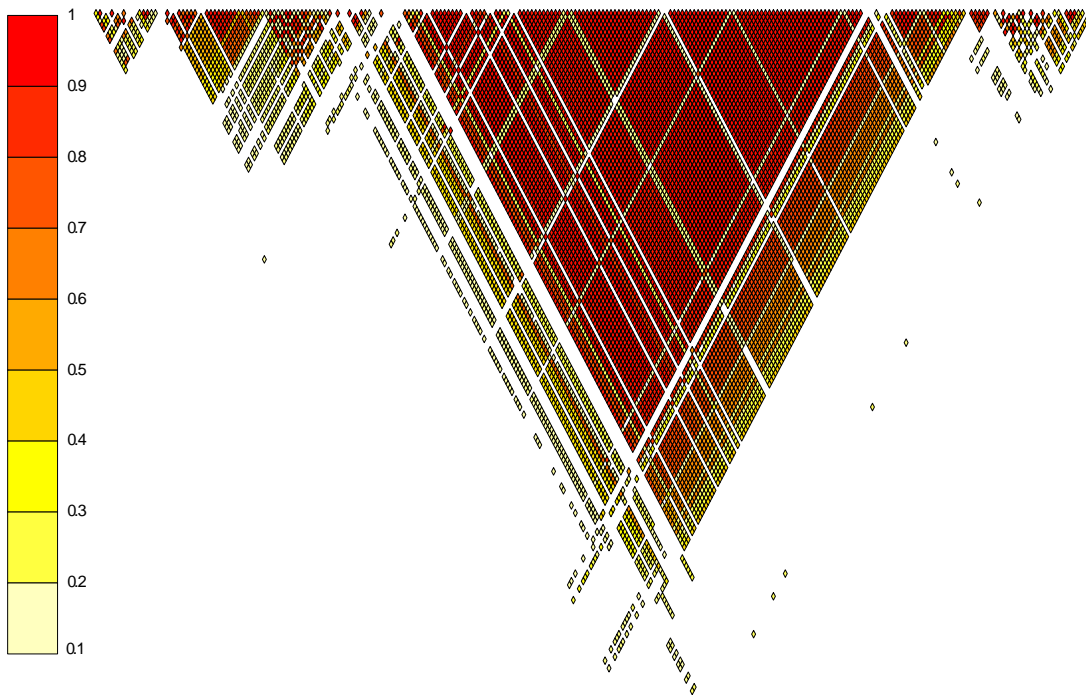


Figure 3.22 – Valeurs de r^2 pour la région entière calculées par paire de SNP. Chaque carré représente la valeur de r^2 entre les deux SNP considérés.



(a) Valeurs de r^2 pour l'haplotype 1 calculées par paire de SNP. Chaque carré représente la valeur de r^2 entre les deux SNP considérés.



(b) Valeurs de r^2 pour l'haplotype 2 calculées par paire de SNP. Chaque carré représente la valeur de r^2 entre les deux SNP considérés.

Figure 3.23

3.4 Discussion

Dans le chapitre 2, nous avons détecté dans HapMap 3.3 des différences de fréquences alléliques entre les sexes, un signal de sélection sexuellement antagoniste (SA). Dans cette partie, nous avons dans un premier temps effectué cette même analyse en adaptant notre méthode à GoNL. Les bases de données HapMap 3.3 et GoNL ont des caractéristiques différentes sur plusieurs points : tout d’abord au niveau de la méthode d’acquisition des données, génotypage pour HapMap 3.3 et séquençage pour GoNL, mais aussi au niveau de la taille des populations, de 41 à 137 individus par échantillon pour HapMap 3.3 tandis que GoNL comporte un échantillon de 498 parents et 250 enfants provenant de la même population. Effectuer la même analyse sur ces deux bases de données nous a permis de déterminer si les résultats obtenus dans HapMap 3.3 étaient reproductibles dans une base de données de séquençage avec une taille d’échantillon plus importante.

Dans un second temps, nous avons mis en évidence des signaux génomiques de distorsion de transmission (TD) en prenant en compte le sexe des enfants en utilisant les données de trio parents-enfant au sein de GoNL. Cela nous a permis d’étudier la relation entre la TD sexe-spécifique et les signaux de sélection SA et ainsi de mieux comprendre l’origine des différences de fréquences alléliques entre les sexes.

Nous avons donc tout d’abord effectué la même analyse que dans le chapitre 2 sur GoNL, ce qui nous a permis de confirmer les résultats obtenus précédemment sur la base de données HapMap 3.3. En effet, chez les parents de GoNL, nous trouvons également une proportion de *genic* SD SNP significativement plus élevée sur le chromosome X que sur la majorité des autosomes. La même tendance est observée après suppression des SNP en LD, bien que celle-ci soit moins significative. Ce résultat est cohérent avec ceux obtenus dans la population européenne CEU de HapMap 3.3. En effet nous avons observé, après suppression des SNP en LD, une proportion plus élevée de *genic* SD SNP sur le chromosome X par rapport aux chromosomes 10 et 22, mais ces différences de proportions n’étaient pas significatives. Nous retrouvons cette tendance chez les enfants de GoNL, ce qui indique qu’un enrichissement en SNP montrant des différences de fréquences alléliques entre hommes et femmes subsiste sur le chromosome X de génération en génération. Néanmoins, l’écart entre les proportions de *genic* SD SNP du chromosome X et celles des autosomes est moins important chez les enfants en comparaison aux parents. Cela pourrait être dû à une plus faible puissance statistique chez les enfants, le nombre d’enfants étant deux fois plus faible que le nombre de parents.

Les F_{ST} intersexuels moyens au sein les *genic* SD SNP sont moins élevés pour GoNL

que pour les populations de HapMap 3.3. En effet, chez les parents, la moyenne des F_{ST} intersexuels pour les *genic* SD SNP est de 0,020 sur le chromosome X et de 0,016 sur les autosomes tandis que la population CEU montrait un F_{ST} intersexuel moyen de 0,080 sur le chromosome X et de 0,066 sur les autosomes, soit approximativement quatre fois plus. La taille des échantillons des populations de HapMap 3.3, qui sont beaucoup plus petits que GoNL, pourraient influencer les valeurs de F_{ST} intersexuel par le biais d'un effet de *winner's curse*. Les résultats des simulations nous indiquent en effet que plus l'échantillon considéré est petit par rapport à la population globale, plus la variance du F_{ST} intersexuel est importante et plus les valeurs extrêmes sont hautes. Cela expliquerait pourquoi les F_{ST} intersexuels sont plus élevés en moyenne dans les populations de HapMap 3.3 en comparaison à GoNL et indique qu'un effet de *winner's curse* serait à l'origine de cette augmentation. Le nombre d'individus inclus dans GoNL étant plus important, l'effet de *winner's curse* est moins fort au sein de cet échantillon. Néanmoins, selon nos simulations, pour obtenir les valeurs de F_{ST} intersexuel observées pour les *genic* SD SNP de GoNL sous l'hypothèse d'une sélection SA, les coefficients de sélection devraient être compris entre 0,2 et 0,3, ce qui représente une forte sélection.

Par ailleurs, les résultats des simulations montrent que, lorsqu'un locus SA apparaît dans une population, le F_{ST} intersexuel fluctue de façon importante, notamment sur le chromosome X. Donc, si la population n'est pas à l'équilibre, cela mènerait à l'identification de locus dont le F_{ST} à l'équilibre serait en réalité beaucoup plus faible. Or, dans la base de données de HapMap 3.3, nos résultats indiquent que la sélection agissant sur les *genic* SD SNP serait récente. De plus, des fluctuations de F_{ST} peuvent subsister ou resurgir si des événements de migrations éloignent les populations de l'équilibre.

Ces deux facteurs, l'effet de *winner's curse* et le fait que les populations ne soient probablement pas à l'équilibre, pourraient conduire à l'identification de locus avec des différences de fréquences alléliques artificiellement augmentées et correspondraient à des pressions de sélection moins extrêmes que celles inférées en appliquant directement un modèle théorique de population à l'équilibre.

Une hypothèse pour expliquer l'observation d'un enrichissement en *genic* SD SNP sur le chromosome X est que les différences de fréquences alléliques entre les sexes pourraient être plus élevées sur le chromosome X en comparaison aux autosomes, de part l'hémi-zygotie des mâles. Si cela est le cas, l'enrichissement en *genic* SD SNP observé sur le chromosome X pourrait signifier que nous arrivons mieux à détecter des différences de fréquences alléliques sur le chromosome X, et non pas que celui-ci est enrichi en SNP potentiellement sous sélection SA. Dans nos simulations, on n'observe pas de différences notables entre les F_{ST} intersexuels du chromosome X et les F_{ST} intersexuels des auto-

somes pour les locus SA, ce qui indique que l'hémizygotie chez les mâles ne mène pas en-soi à des différences de fréquences alléliques plus élevées sur le chromosome X. Le fait qu'on observe des F_{ST} intersexuels moyens plus importants sur le chromosome X en comparaison aux autosomes dans GoNL et HapMap 3.3 est donc probablement dû à la diminution de la puissance statistique du test de Fisher pour les locus liés au chromosome X, comme montré en chapitre 2. En d'autres termes, pour être détectées, les différences de fréquences alléliques doivent être plus importantes sur le chromosome X en comparaison aux autosomes.

Une étude d'enrichissement fonctionnel des gènes comportant au moins un *genic* SD SNP (les SD gènes) a été effectuée chez les parents et chez les enfants et a permis de mettre en évidence des fonctions communes avec celles trouvées dans HapMap 3.3. En effet, chez les parents, une des fonctions enrichies est impliquée dans l'organisation de la chromatine qui est connue pour jouer un rôle majeur dans la régulation de l'expression génique, fonction enrichie dans HapMap 3.3. De plus, des fonctions impliquées dans des processus métaboliques étaient enrichies dans les deux bases de données. Enfin, des termes relatifs à la méiose ont été identifiés chez les parents de GoNL, et nous avons détecté des fonctions impliquées dans la production des gamètes dans HapMap 3.3. Chez les enfants de GoNL, la liste de SD gènes est enrichie pour des termes référant à la croissance des neurones, et des fonctions impliquées dans le développement du cerveau avaient été détectées dans HapMap 3.3. Une autre fonction enrichie chez les enfants réfère aux processus métaboliques des catécholamines. Ces molécules, comportant notamment l'adrénaline et la dopamine, peuvent agir comme des hormones ou des neurotransmetteurs et modifient le métabolisme de l'organisme lors d'un stress. Elles sont notamment impliquées dans la régulation de la masse grasseuse car elles enclenchent ou inhibent la lipolyse selon les récepteurs auxquels elles se fixent. Des études ont montré que l'activité des catécholamines est sexuellement dimorphique chez l'Homme (Shi et Clegg 2009). De plus, chez le rat, la distribution des récepteurs aux catécholamines est différente chez les mâles et les femelles (Shi et Clegg 2009).

Nous avons jusqu'à présent détecté des différences de fréquences alléliques entre les sexes dans deux bases de données. Bien que ce signal puisse être causé par d'autres mécanismes, discutés en chapitre 2, il est attendu que la liste de locus identifiés grâce à notre méthode soit enrichie pour des locus sous sélection SA. Dans le cadre de cette hypothèse, des différences de fréquences alléliques entre les sexes seraient dues à une sélection SA sur la survie des individus. Dans une étude récente, Ucisik-Akkaya *et al.* (2010) ont détecté des différences de fréquences alléliques entre les sexes chez des nouveaux-nés dans une

population humaine. Si ces différences sont dues à de la sélection SA, cela indiquerait que des processus sexuellement antagonistes ont lieu durant le développement embryonnaire. Ce type de mécanisme se traduirait par une distorsion de transmission sexe-spécifique et pourrait se produire entre la production de la lignée germinale et la naissance, menant à des différences de fréquences alléliques entre les sexes. Dans ce chapitre, nous avons mis au point une méthode de détection de distorsion de transmission en fonction du sexe des enfants par balayage du génome en utilisant des données génomiques de trio parents-enfant. Cette méthode nous a permis d'évaluer si une distorsion de transmission sexe-spécifique pouvait être à l'origine de différences de fréquences alléliques entre les sexes.

Pour tester notre méthode, nous avons effectué des simulations en *forward* individuelles centrées d'une population de 3000 individus possédant un chromosome (chromosome X ou autosome) sur lequel nous avons placé en position médiane un locus sous sélection SA, pour une large gamme de coefficients de sélection et de dominance. À chaque génération, un accouplement aléatoire entre deux individus ainsi que des événements de recombinaison et de mutation ont été simulés. La survie de chaque embryon formé par un couple a été déterminée par son sexe et son génotype au locus sous sélection SA. Les résultats de nos tests sur les simulations indiquent que notre méthode de détection de distorsion de transmission capture, comme attendu, un signal de TD sexe-spécifique et que celui-ci est associé au signal de sélection SA.

Nous avons donc utilisé notre méthode de détection de TD sur les données de GoNL et nous avons mis en évidence des locus montrant des signaux de TD chez les deux sexes, de TD limitée à un sexe et de TD sexe-spécifique. Nous n'avons pas effectué cette analyse sur les données de HapMap 3.3 car la taille d'échantillons pour les données de trio, lorsqu'ils existent, est faible. Notre méthode nécessitant de diviser l'échantillon des enfants en deux selon leur sexe, nous aurions manqué de puissance pour détecter une TD sexe-spécifique dans ces données.

Tout d'abord, nous avons comparé les signaux de TD avec le F_{ST} intersexuel chez les enfants afin de déterminer si les différences de fréquences alléliques entre les filles et les fils pouvaient être dues à une TD sexe-spécifique. Nos résultats indiquent que les signaux de TD sexe-spécifique et de sélection SA sont fortement corrélés. En effet, l'intensité de la différence de transmission entre les sexes et le F_{ST} intersexuel sont corrélés positivement et de façon significative pour 21 chromosomes sur 23. La corrélation sur le chromosome X, environ égale à 0,4, est deux fois plus élevée que celles des autosomes. De plus, une analyse des distributions du F_{ST} intersexuel chez les enfants en fonction des différents signaux de TD montre que les locus ayant un signal de TD sexe-spécifique ont des F_{ST} intersexuels plus élevés que les locus ayant un autre signal de TD ou pas de signal TD.

Ces résultats sont confirmés par l'estimation du pourcentage de *genic* SD SNP au sein des SNP montrant les différents signaux de TD : après correction pour tests multiples sur les signaux de TD, 51% des SNP avec un F_{ST} intersexuel significatif chez les enfants montrent également un signal de TD sexe-spécifique.

Grâce à notre méthode, nous avons pu détecter des signaux de TD sexe-spécifique, et nos résultats suggèrent que ce processus est à l'origine d'une grande partie des différences de fréquences alléliques entre les sexes chez les enfants.

Cependant, on observe que 40% des locus ayant un F_{ST} significatif chez les enfants ne présentent pas de signal de TD. Il est possible que les différences de fréquences alléliques observées pour ces locus soient tout de même dues à une TD sexe-spécifique, mais que notre méthode ne nous permette pas d'avoir assez de puissance statistique pour la détecter. En effet le nombre de parents hétérozygotes pour un locus détermine la puissance du TDT et du test SDT. Une hypothèse alternative serait que des processus démographiques soient à l'origine des F_{ST} intersexuels. Néanmoins, nous avons montré que la population n'était pas structurée en fonction du sexe au niveau du génome, il est donc peu probable que ces différences soient dues à des processus démographiques sexe-spécifiques.

L'hypothèse que nous avons formulée dans le chapitre 2 est que la TD sexe-spécifique résulte d'une sélection post-fécondation agissant sur la survie des embryons en fonction de leur sexe et de leur génotype. Néanmoins, un signal de TD sexe-spécifique pourrait aussi être dû à des processus de distorsion de transmission se produisant *via* une sélection gamétique ou une distorsion de ségrégation. D'ailleurs, comme présenté en introduction, il a été montré qu'un processus de distorsion de ségrégation pouvait favoriser le maintien d'un polymorphisme entre les sexes à un locus sous sélection SA (Patten 2014). De plus, une étude chez *Drosophila melanogaster* (Connallon et Jakubowski 2009) et une autre étude chez le lézard *Uta stansburiana* (Calsbeek et Sinervo 2003) ont mis en avant la possibilité que les femelles effectuent des choix cryptiques des spermatozoïdes fécondants et biaisent ainsi le sexe-ratio de leur descendance. En effet, lors d'un accouplement avec des mâles ayant une fitness basse, les femelles produisaient plus de filles chez les deux espèces tandis que lors d'un accouplement avec des mâles avec une fitness élevée, les femelles avaient tendance à produire plus de fils chez le lézard, et une proportion égale de filles et de fils chez la drosophile. Ce phénomène pourrait être dû à un choix cryptique des femelles qui ajusteraient le sexe-ratio de leur descendance en fonction de la qualité de leur partenaire, donc de leur génotype. Dans le cas de la drosophile, ce phénomène pourrait également être expliqué par une distorsion de ségrégation du chromosome X chez les mâles avec une fitness basse. Un choix cryptique de la femelle en fonction du chromosome sexuel porté par le spermatozoïde pourrait être à l'origine d'une TD sexe-

spécifique et d'une distorsion de sexe-ratio dans la descendance. Ces différents processus se dérouleraient pendant le développement embryonnaire et mèneraient à une TD sexe-spécifique et à des différences de fréquences alléliques entre les sexes.

Néanmoins, le signal de TD sexe-spécifique observé pourrait également résulter d'une sélection SA sur la survie pendant l'enfance. En effet, il est probable que les enfants de notre base de données ne soient pas des nouveaux-nés et il est donc possible que des épisodes de sélection aient eu lieu après leur naissance. Cependant, la mortalité infantile aux Pays-Bas est inférieure à 1% depuis les années 1980, selon le département de l'économie et des affaires sociales des nations unies (esa.un.org/unpd/wpp/). Même si l'on suppose qu'une grande part de la mortalité infantile serait due à des processus sexuellement antagonistes, elle participerait à une faible proportion des différences de fréquences alléliques entre les sexes observées. D'autre part, le taux d'avortement spontané chez l'Homme est d'au moins 50% (Benagiano *et al.* 2010), il est donc plus parcimonieux de considérer que des processus de sélection SA sur la survie agiraient au cours du développement.

Si la population est à l'équilibre, dans le cadre d'une sélection SA, on s'attend à ce qu'un polymorphisme SA entre les sexes soit stable entre deux générations. Nous avons donc testé si les locus montrant des F_{ST} intersexuels élevés chez les parents montraient un signal de TD sexe-spécifique. Les résultats indiquent qu'il y a peu de concordance entre le signal de TD sexe-spécifique et le F_{ST} intersexuel chez les parents. Nous trouvons également peu de concordance entre le F_{ST} intersexuel des enfants et des parents : 4,6% des SNP montrant un F_{ST} intersexuel significatif chez les parents montrent aussi un F_{ST} intersexuel significatif chez les enfants. Une explication pour le peu de concordance entre le F_{ST} intersexuel chez les parents et les enfants pourrait être une sélection SA plus tardive dans la vie des individus qui diminuerait le signal de concordance. Pour vérifier si une sélection SA spécifique à certains âges agit, il faudrait prendre en compte l'âge des individus dans nos analyses, en comparant les différences de fréquences alléliques entre les sexes pour différentes classes d'âge.

Une région sur le chromosome 17 présente un fort signal de TD sexe-spécifique, ainsi que des F_{ST} significatifs chez les parents et les enfants. Cette région contient notamment un cluster de 5 gènes codant pour des hormones de croissance. Trois de ces gènes, CSH2, CSHL1 et GH1, montrent une concordance entre le signal de TD sexe-spécifique et de sélection SA chez les enfants et les parents. Ces gènes font partie de la famille des somatotrophines et prolactines, et sont importants pour le développement embryonnaire et la croissance. Ce signal est donc particulièrement intéressant dans le contexte d'une sélection SA. Nous avons mis en évidence un comportement particulier des F_{ST} intersexuels entre parents et enfants dans cette région. D'une part, les locus présentant un signal de TD

sexe-spécifique ont majoritairement un F_{ST} bas chez les parents mais un F_{ST} élevé chez les enfants. D'autre part, les locus ne présentant pas un signal de TD sexe-spécifique ont majoritairement un F_{ST} élevé chez les parents et présentent un F_{ST} faible chez les enfants. Un F_{ST} bas chez les parents pourrait être dû à la présence de nombreux hétérozygotes chez les pères et chez les mères. Or, un excès d'hétérozygotes augmenterait la puissance de notre méthode de détection de TD sexe-spécifique dans cette région. Par contre, si les parents ont des F_{ST} élevés, cela pourrait signifier qu'il y a moins d'hétérozygotes, et donc moins de puissance pour détecter un signal de TD sexe-spécifique. Un mécanisme possible pouvant expliquer ces observations est le suivant : considérons une région sous sélection SA composée de locus présentant de nombreux hétérozygotes, à la génération suivante, les différences de fréquences alléliques entre les sexes pourront être plus élevées car le pool génétique de départ est plus divers chez chaque sexe. Au contraire, si l'on considère une région sous la même force de sélection SA composée de locus où il y a peu d'hétérozygotes et où les différences de fréquences alléliques entre les sexes sont élevées, à la génération suivante ces différences seront atténuées car on observera un excès d'hétérozygotes.

En conséquence, si les parents sont en majorité hétérozygotes pour un locus sous sélection SA, on s'attend à ce que la différence entre les fréquences alléliques chez les filles et chez les fils soient plus importantes, pour une même intensité de sélection, que s'il y a peu d'hétérozygotes chez les parents et que ceux-ci présentent d'importantes différences de fréquences alléliques entre les sexes. Un manque de puissance pourrait expliquer le peu de concordance entre les F_{ST} intersexuels chez les enfants et chez les parents.

Afin de tester si ce mécanisme pourrait être à l'origine de ces résultats, une perspective serait d'évaluer le taux d'hétérozygotes chez les parents et chez les enfants et de tester si celui-ci est différent de l'attendu. Il a été montré que les tests de Hardy-Weinberg sont soit trop sensibles, soit pas assez sensibles aux écarts à l'attendu car la puissance statistique de ce test dépend de la fréquence de l'allèle mineur ainsi que de la taille de l'échantillon (Wigginton *et al.* 2005). L'utilisation de ce type de tests est donc peu conseillée dans notre cas. Par ailleurs, comme le suggèrent nos résultats, une analyse des signaux de TD et de F_{ST} intersexuel par fenêtre glissante serait plus adaptée, car cela augmenterait notre puissance pour détecter une concordance entre les régions ayant un F_{ST} intersexuel élevé chez les parents, les enfants et un signal de TD sexe-spécifique.

Dans cette région du chromosome 17, nos résultats suggèrent qu'il y aurait au moins deux haplotypes ségrégeant dans la population. En effet, les locus montrant des F_{ST} intersexuels élevés chez les parents sont en fort LD et les locus montrant des F_{ST} intersexuels faibles chez les parents sont également en fort LD, tandis que, lorsqu'on prend en compte l'ensemble des locus de la région, le LD est plus faible. Ce résultat est retrouvé pour les mêmes locus chez les enfants. Il est donc possible qu'il y ait plusieurs haplotypes

sous sélection SA dans cette région. Il est néanmoins difficile d'en déduire le mécanisme précis à l'origine de ces résultats car, à notre connaissance, nous n'avons aucun attendu théorique sur ce type de sélection SA. Une perspective serait donc de mettre au point des simulations avec plusieurs haplotypes sous sélection SA afin de déterminer si ce type de processus pourrait être à l'origine des résultats observés. Nous avons détecté cette région car elle montre des signaux extrêmes, mais il est possible que des mécanismes similaires soient en jeu dans le reste du génome.

Dans un dernier temps, une analyse d'enrichissement fonctionnel a été effectuée sur les locus montrant différents signaux de TD.

Nous avons trouvé 755 gènes présentant un signal de TD sexe-spécifique intermédiaire, et 9 gènes présentant un signal de TD sexe-spécifique fort. Ces gènes sont enrichis pour des fonctions impliquées dans l'activité de la GTPase, la perception sensorielle et la cognition, ainsi que l'activité de la lipase, la réponse au stress oxydatif et le système immunitaire. Or, un terme référant aux fonctions sensorielles était enrichi dans l'analyse sur les *SD* gènes chez les parents et chez les enfants. Dans HapMap 3.3, nous avons par ailleurs trouvé un enrichissement fonctionnel pour des gènes impliqués dans le système immunitaire.

De nombreux signaux de TD limitée à un sexe ont été identifiés, 431 gènes pour la TD limitée aux fils et 527 gènes pour la TD limitée aux filles. Les gènes présentant un signal de TD limitée aux fils sont enrichis pour des fonctions impliquées dans l'organisation de la chromatine, la régulation de l'expression génique, la transcription et la gamétogénèse. Les fonctions associées à la gamétogénèse suggèrent que des processus de distorsion de ségrégation ou de sélection gamétique pourraient avoir lieu. Cela signifierait qu'un allèle localisé dans un gène impliqué dans la spermatogénèse serait plus souvent transmis aux fils qu'aux filles. Les gènes pour lesquels un allèle est préférentiellement transmis aux filles uniquement seraient quant à eux impliqués dans le développement embryonnaire, les fonctions mitochondriales et le système immunitaire. Les différences entre les fonctions trouvées dans le cas d'une TD limitée au fils et d'une TD limitée aux filles sont potentiellement intéressantes, car elles indiquent dans un cas une TD dans des fonctions liées à la spermatogénèse et dans l'autre dans des fonctions liées à la gestation. Néanmoins, il est relativement étonnant de mettre en évidence chez les enfants des fonctions sur lesquelles la sélection n'agirait qu'à l'âge adulte. On peut éventuellement imaginer qu'un locus distorateur soit associé à un locus affectant la spermatogénèse (la gestation) et ait été sélectionné car confère un avantage durant la reproduction.

Enfin, nous n'avons trouvé que 3 gènes montrant une TD chez les deux sexes. Ce faible nombre est cohérent avec les précédentes études sur la TD, qui avaient trouvé

peu de signaux. Les gènes identifiés ici sont enrichis pour des fonctions impliquées dans l'organisation de la chromatine et des nucléosomes, comme les gènes présentant une TD limitée aux fils. Nos résultats indiquent que les gènes pour lesquels un allèle est préférentiellement transmis, indépendamment du sexe de l'enfant, sont enrichis en fonctions impliquées dans l'expression génique et le marquage épigénétique.

Nous avons trouvé peu de concordance entre nos résultats et les résultats des précédentes études sur la TD. Cela pourrait notamment être dû à la sensibilité du TDT aux erreurs de génotypage, qui mènerait à de nombreux faux positifs (Mitchell *et al.* 2003). De plus, les résultats de Deng *et al.* (2009) suggèrent que le signal de TD n'est pas partagé entre populations. Nous avons utilisé une base de données de séquençage avec une taille de population importante et nous avons pu corriger une partie au moins des erreurs de séquençage grâce à la présence de trios. De plus, cette population, à notre connaissance, n'a jamais été utilisée dans la détection de signaux de TD.

Comme perspective de cette étude, il serait intéressant d'analyser l'origine parentale des allèles présentant un signal de TD, afin de détecter si le père ou la mère est plus souvent impliqué dans la distorsion de transmission. En effet, nous n'avons pas trouvé de signal de TD sur le bras court du chromosome 6, comme identifié par Santos *et al.* (2009) et Meyer *et al.* (2012), mais ces TD se produisaient chez les pères et notre méthode de détection de TD ne prend pas en compte le sexe du parent. Dans nos analyses, les signaux de TD observés sur le chromosome X non recombinant avec le Y ne peuvent provenir que de la mère car le TDT ne prend en compte que les parents hétérozygotes. Néanmoins, sur les autosomes les signaux de TD que nous avons détectés pourraient provenir des deux parents. Il est attendu qu'une distorsion de ségrégation soit plus courante chez les femmes à cause du caractère asymétrique de la gaméto-genèse, il serait intéressant de vérifier si cela est le cas dans les données de GoNL.

Dans notre base de données, nous n'avons qu'un seul enfant par couple, il est donc impossible de savoir s'il y a une distorsion de sexe-ratio chez les descendants en comparant le nombre de fils et de filles par couple. De plus, il est difficile de détecter des distorsions de sexe-ratio dans les populations humaines car, généralement, la puissance statistique pour détecter une déviation de l'attendu est très faible à cause du petit nombre d'enfants par couple. Dans l'objectif de mettre en évidence de potentielles distorsions de sexe-ratio, il serait intéressant d'étudier les régions pseudo-autosomales (PAR) afin de détecter un effet du sexe des parents dans la TD sur le chromosome X. En effet, sur les PAR, une TD chez les pères pourrait mener à une distorsion de sexe-ratio. Par exemple, si un allèle partiellement associé au chromosome Y est plus souvent représenté dans les gamètes du père, cela pourrait mener à un excès de garçons dans sa descendance. L'utilisation de lo-

cus liés aux PAR pourraient être intéressante dans la détection de distorsion de sexe-ratio.

Une limite de notre méthode est que le test SDT que nous avons utilisé pour détecter une TD sexe-spécifique est basé sur des comptes alléliques, comme le test exact de Fisher utilisé pour évaluer la significativité des différences de fréquences alléliques entre les sexes. Le fait qu'il y ait une corrélation entre le F_{ST} intersexuel et la TD sexe-spécifique est donc attendu. Néanmoins, le test SDT prend en compte un sous échantillon des individus dans la population car il considère les comptes d'allèles transmis par des parents hétérozygotes. De plus, les résultats de TD sexe-spécifique nous apportent une information supplémentaire par rapport au F_{ST} intersexuel car le SDT prend en compte indirectement le génotype des parents, et propose un mécanisme à l'origine des différences de fréquences alléliques entre les sexes. Une perspective serait de développer une méthode statistique se basant sur des modèles de maximum de vraisemblance, qui prendraient en compte le fait que les comptes d'allèle transmis suivent une loi binomiale ayant pour paramètre une probabilité de $1/2 + \varepsilon$, où ε représente la déviation de l'attendu selon les lois de Mendel. Il serait ainsi possible de tester la vraisemblance des différents modèles de TD. Cette méthode aurait un avantage par rapport à celle que nous avons développée dans ce chapitre car, bien que celle-ci soit adaptée à la détection de TD, l'utilisation de 3 tests en parallèle pourrait engendrer une perte de puissance statistique.

À notre connaissance, ce chapitre constitue la première étude qui s'intéresse à des signaux de distorsion de transmission en prenant en compte le sexe des enfants. En utilisant la méthode développée dans ce chapitre sur des données de séquençage de trios, nous avons détecté des signaux de TD dans des directions opposées chez les deux sexes, de TD limitée à un sexe ainsi que de TD dans la même direction chez les deux sexes. Nous avons trouvé plus de locus présentant une TD limitée à un sexe ou une TD sexe-spécifique que de signaux de TD dans la même direction chez les deux sexes. Cela indique qu'à l'avenir il serait intéressant de prendre en compte le sexe de l'enfant lors d'étude sur la TD. Néanmoins, notre méthode a une puissance plus faible pour détecter des processus de TD dans la même direction chez les deux sexes qu'un test de TDT classique.

Chapitre 4

Conclusion Générale

Au cours de cette thèse, nous avons dans un premier temps étudié une signature génétique de la sélection sexuellement antagoniste : une différence de fréquences alléliques entre les sexes, calculée sur des SNP, et nous avons mis en évidence des locus potentiellement sous sélection SA à l'aide d'un balayage du génome sur une base de données de génotypage de populations humaines.

Nos résultats suggèrent que le chromosome X est enrichi en SNP montrant des différences de fréquences alléliques entre les sexes en comparaison aux autosomes, ce qui est en accord avec l'hypothèse de Rice (1984). Le test statistique que nous avons utilisé pour évaluer la significativité des différences de fréquences alléliques entre les sexes, le test exact de Fisher, est sensible à la taille de l'échantillon et a donc moins de puissance pour des locus liés au chromosome X que pour des locus autosomaux de part l'hémizygotie des mâles sur le chromosome X, renforçant ainsi nos résultats. De plus, nous avons montré que des effets de déséquilibre de liaison ou d'erreurs de génotypage n'induisaient pas un biais dans nos résultats. Enfin, nous avons effectué une analyse afin de vérifier que l'enrichissement en SNP présentant des différences de fréquences alléliques entre les sexes sur le chromosome X était bien dû à une comparaison entre les hommes et les femmes, et non à un biais statistique. Pour cela, nous avons calculé, pour plusieurs sets, les différences de fréquences alléliques entre deux groupes ayant des sexe-ratios du plus équilibré, c'est-à-dire composés d'autant de mâles que de femelles, au moins équilibré, c'est-à-dire avec un groupe composé exclusivement de mâles et l'autre groupe composé exclusivement de femelles, qui correspondent aux groupes étudiés dans nos analyses. Les résultats montrent que l'enrichissement du chromosome X pour des SNP montrant des différences de fréquences alléliques entre les sexes est fortement corrélé au déséquilibre de sexe-ratio des groupes comparés. Cela indique que le signal que nous détectons est dû à des différences entre les mâles et les femelles.

Les différences de fréquences alléliques détectées entre les sexes étaient en grande

majorité spécifiques aux populations étudiées, ce qui suggèrent que la sélection est récente. Une étude de la distribution du F_{ST} intersexuel autour des SNP identifiés nous a permis de mettre en évidence un signal significativement plus élevé dans les régions voisines des SNP potentiellement sous sélection SA en comparaison à des SNP pris au hasard, ce qui est en accord avec l'hypothèse d'une sélection récente.

Ensuite, une étude d'enrichissement fonctionnel sur les gènes présentant au moins un SNP avec des différences de fréquences alléliques entre les sexes (SD gènes) nous a permis de mettre en évidence des fonctions impliquées dans la réalisation des dimorphismes sexuels : la régulation de l'expression des gènes et les processus développementaux, et des fonctions sexuellement dimorphiques chez l'Homme, notamment la reproduction, le système immunitaire et le métabolisme énergétique.

Cette même analyse a été effectuée sur une base de données de séquençage comportant 248 trios parents-enfant. Nous avons trouvé, chez les parents et chez les enfants indépendamment, un enrichissement en SNP présentant des différences de fréquences alléliques entre les sexes sur le chromosome X en comparaison aux autosomes. Cela indique que cette tendance est cohérente entre populations mais aussi entre méthodes d'acquisition des données, dans notre cas séquençage et génotypage. De plus, cette tendance est retrouvée sur deux générations successives, suggérant qu'un mécanisme permettrait le maintien d'un enrichissement en locus présentant des différences de fréquences alléliques entre les sexes sur le chromosome X.

Une étude d'enrichissement fonctionnel sur les SD gènes chez les parents nous a permis de mettre en évidence des fonctions impliquées dans l'organisation de la chromatine, dans les processus métaboliques et dans la gamétogénèse, également enrichies dans la base de données de HapMap 3.3. Chez les enfants, les fonctions enrichies réfèrent à la croissance des neurones, ce qui est cohérent avec des fonctions enrichies dans HapMap 3.3. Une autre fonction détectée chez les enfants est impliquée dans le métabolisme des catécholamines, qui agissent notamment dans la régulation de la masse graisseuse et dont l'activité est sexuellement dimorphique chez l'Homme (Shi et Clegg 2009).

Des simulations de chromosomes (X ou autosomes) comportant un locus sous sélection SA agissant sur la survie nous ont permis de mettre en évidence que les chromosomes X simulés ne présentent pas de différences de fréquences alléliques entre les sexes plus élevées que les autosomes simulés, pour une même intensité de sélection. L'enrichissement que nous avons détecté ne serait donc pas dû à des différences de fréquences alléliques intrinsèquement plus élevées sur le chromosome X, causées par l'hémizygotie des mâles, en comparaison aux autosomes.

Une perspective intéressante de cette étude serait de simuler plusieurs chromosomes

(autosomes et chromosome X) au sein d'un même génome, comportant des locus sous sélection sexuellement antagoniste. En fixant le nombre de locus sous sélection SA sur chaque chromosome, cela nous permettrait de calculer les différences de proportions de locus sous sélection SA entre les autosomes et le chromosome X et de déterminer plus précisément si notre méthode est plus sensible à l'intensité des différences de fréquences alléliques entre les sexes ou au nombre de locus présents sur les chromosomes.

De plus, ces simulations nous ont permis de mettre en évidence un effet de *winner's curse* dans nos analyses. En effet, en réduisant la taille des échantillons tirés au hasard au sein de la population simulée, les valeurs extrêmes ainsi que la variance des F_{ST} intersexuels augmentent. Les F_{ST} intersexuels moyens que nous avons observés dans les deux bases de données sont cohérents avec cette hypothèse : le F_{ST} intersexuel est près de 4 fois plus élevé dans les populations d'HapMap 3.3 en comparaison à GoNL. Il est donc important de considérer que les différences de fréquences alléliques détectées entre les sexes dans les bases de données utilisées sont des estimations hautes des différences de fréquences alléliques de la population globale. Par ailleurs, nos simulations ont montré que, durant les premières générations de ségrégation d'un locus SA, le F_{ST} intersexuel varie fortement et continue de varier de façon substantielle, notamment sur le chromosome X. Si un locus SA est récent, il est possible que la population n'ait pas atteint l'équilibre et que nous observions des F_{ST} intersexuels particulièrement élevés. Il est donc difficile d'estimer le taux de sélection nécessaire pour aboutir à des différences de fréquences alléliques entre les sexes de l'ordre de celles que nous avons observées dans nos analyses.

Il serait intéressant d'étudier le déséquilibre gamétique entre les locus que nous avons identifiés, afin de déterminer si ceux-ci sont en plus forte association statistique qu'attendu. En effet, une autre étude théorique de Ubeda *et al.* (2011) suggère que la sélection sexuellement antagoniste serait à l'origine d'un maintien de déséquilibre gamétique entre des locus, même s'ils sont impliqués dans la détermination de traits différents et qu'ils sont éloignés physiquement.

En introduction, nous avons discuté du fait que la résolution d'IASC devrait être plus longue si le gène sous conflit a une expression ubiquitaire en comparaison à une expression tissu-spécifique. Il serait intéressant de contraster les différences d'expression génique au niveau des gènes comportant des locus avec un signal de sélection SA que nous avons détectés afin de déterminer s'il y a un enrichissement en gène ayant une expression ubiquitaire ou tissus-spécifique.

L'émergence de différences de fréquences alléliques entre les sexes pourrait être due

à une sélection SA agissant sur la viabilité durant le développement embryonnaire, se traduisant par une distorsion de transmission sexe-spécifique. Pour tester cette hypothèse, nous avons mis au point une méthode de détection de locus pour lesquels les parents ne transmettent pas leurs allèles de façon aléatoire selon le sexe de leurs enfants. Un test de cette méthode a été effectué sur des données simulées et a montré que nous détectons bien une distorsion de transmission sexe-spécifique associée au signal de sélection SA sur la viabilité. Nous avons donc appliqué cette méthode sur une base de données de séquençage de 248 trios (GoNL) et nous avons détecté des locus montrant un signal de distorsion de transmission chez les deux sexes dans la même direction, chez un seul sexe et chez les deux sexes dans des directions opposées.

Nos résultats indiquent qu'une TD sexe-spécifique serait à l'origine d'une grande partie des différences de fréquences alléliques chez les enfants. Bien que notre hypothèse de départ est que la TD sexe-spécifique soit due à une sélection SA durant le développement embryonnaire, celle-ci pourrait aussi être due à une distorsion de ségrégation ou à une sélection gamétique. En effet, une sélection gamétique pourrait être à l'origine de différence de fréquences alléliques entre les sexes si un choix cryptique de la femelle pour des spermatozoïdes a lieu en fonction du chromosome sexuel que celui-ci porte, impactant le sexe-ratio de sa descendance. Par ailleurs, dans la mesure où la présence d'un locus distorateur de ségrégation favorise le maintien d'un locus SA lié génétiquement (Patten 2014), il est attendu que ces deux types de locus soient souvent proche physiquement.

Nous avons ensuite comparé les F_{ST} intersexuels observés chez les enfants et chez les parents. Dans un scénario de sélection SA, on s'attend à ce qu'à l'équilibre les différences de fréquences alléliques entre les sexes soient maintenues d'une génération à l'autre. Nos résultats indiquent néanmoins qu'il y a peu de concordance entre les F_{ST} intersexuels des parents et ceux des enfants. De plus, nous avons trouvé qu'il y a peu de concordance entre le F_{ST} intersexuel chez les parents et les signaux de TD sexe-spécifique. Cependant, une région du chromosome 17 présente un signal de TD sexe-spécifique associé à des F_{ST} intersexuels significatifs chez les enfants et chez les parents. L'analyse de cette région indique qu'on manquerait de puissance pour détecter de la TD sexe-spécifique lorsque les parents ont des différences importantes de fréquences alléliques entre les sexes, probablement à cause d'un déficit en hétérozygotes.

La région du chromosome 17 identifiée contient un cluster de 5 gènes codant pour des hormones de croissance. Ces hormones participent à la mise en place des caractères sexuels secondaires durant la puberté chez les deux sexes (Ogilvy-Stuart et Shalet 1992) et à des fonctions de reproduction pendant la vie adulte comme la gamétogenèse et l'ovulation (voir Hull et Harvey 2001 pour une revue). Il est donc particulièrement intéressant que cette région, impliquée dans le développement de dimorphismes sexuels chez l'Homme et

dans la reproduction, montre des signaux de sélection SA chez les parents et les enfants, et que les parents transmettent différents allèles à leurs fils et à leurs filles. Les hormones de croissance sont également impliquées dans le déterminisme de la taille, une déficience des gènes codant pour ces hormones pouvant provoquer des pathologies telles que le nanisme ou le gigantisme. Chez l'Homme, la taille est un trait avec une forte héritabilité, estimée à 80% (Visscher *et al.* 2006), et il a été montré qu'au sein de fratries comportant des individus de petite taille, les femmes ont un succès reproducteur plus élevé que leurs frères tandis qu'au sein de fratries comportant des individus de taille intermédiaire, les hommes ont un succès reproducteur plus élevé que leurs sœurs (Stulp *et al.* 2012). Ces éléments indiquent qu'un IASC agit sur la taille chez l'Homme (Stulp *et al.* 2012). Bien que nous ayons détecté un signal de sélection SA dans la région comportant des gènes codant pour des hormones de croissance, la taille est un trait complexe. En effet, une étude récente de Lango Allen *et al.* (2010) a identifié plusieurs centaines de locus potentiellement impliqués dans le déterminisme de la taille chez l'Homme, expliquant environ 20% des variations de taille au sein de leur échantillon. Il est attendu que des locus sous sélection SA soient en déséquilibre gamétique (Ubeda *et al.* 2011), il serait donc intéressant d'étudier l'association statistique au niveau de la transmission entre les locus identifiés comme impliqués dans le déterminisme de la taille. Cependant, le nombre important de locus identifiés dans l'étude de Lango Allen *et al.* (2010) pourrait être dû au fait que des locus codant pour des traits également sous IASC, autres que la taille, soient en déséquilibre gamétique avec les locus déterminant la taille, comme proposé par Ubeda *et al.* (2011).

Nos résultats suggèrent la présence de plusieurs haplotypes au sein de cette région du chromosome 17. Or, les études théoriques de Patten *et al.* (2010) et Ubeda *et al.* (2011) proposent qu'il existerait des haplotypes sexuellement antagonistes ségrégeant au sein des populations, certains enrichis pour des allèles bénéfiques pour les mâles et d'autres enrichis pour des allèles bénéfiques aux femelles. Il serait donc intéressant d'étudier plus en détail cette région au niveau haplotypique afin de déterminer si cela est le cas. Une perspective majeure de cette thèse serait d'étudier des haplotypes plutôt que des SNP indépendamment à l'aide de données phasées. En effet, nous avons pu mettre en évidence cette région particulière car elle présente des signaux extrêmes, néanmoins il est possible que ce type de phénomène se produise dans le reste du génome mais que nous ayons manqué de puissance statistique pour les détecter en utilisant des méthodes monocus. Nous prévoyons donc de développer une méthode pour détecter des haplotypes SA à l'aide d'un balayage du génome.

À l'aide de notre méthode, nous avons également pu mettre en évidence des locus

montrant un signal de TD limitée à un sexe et un signal de TD dans le même sens chez les deux sexes. Nous avons effectué une analyse d'enrichissement fonctionnel sur les gènes montrant les différents signaux de TD. Les gènes présentant un signal de TD sexe-spécifique sont notamment enrichis en fonctions impliquées dans le système immunitaire, fonction également enrichie dans HapMap 3.3, et dans les fonctions sensorielles, enrichie chez les parents et chez les enfants dans GoNL. De façon intéressante, un enrichissement en fonctions impliquées dans la gamétogenèse a été trouvé dans les gènes montrant un signal de TD limitée au fils, tandis qu'un enrichissement en fonctions impliquées dans la gestation a été trouvé dans les gènes montrant un signal de TD limitée aux filles. Cela indique qu'un processus de distorsion de transmission limitée à un sexe agirait au niveau de gènes différents, impliqués dans des fonctions importantes pour la reproduction. Enfin, les gènes montrant un signal de TD chez les deux sexes sont enrichies pour des fonctions référant à l'organisation de la chromatine et des nucléosomes, qui sont des fonctions importantes dans la régulation de l'expression génique.

Lors de l'étude des SD gènes sur la base de données de HapMap 3.3 et de GoNL, nous avons mis en évidence un enrichissement en fonction sexuellement dimorphiques chez l'Homme, comme le système immunitaire, la reproduction ou le métabolisme, ou impliquées dans la réalisation des dimorphismes sexuels, comme les processus développementaux et la régulation de l'expression génique. Or, cela est cohérent avec l'attendu : si l'on étudie des traits sous IASC, nous devrions trouver des traits au moins partiellement dimorphiques entre les sexes (Bonduriansky et Chenoweth 2009). Bien que nous ayons trouvé dans nos différentes analyses des fonctions enrichies qui sont pertinentes par rapport à la problématique posée, il est important de rester critique face aux résultats des analyses d'enrichissement fonctionnel. En effet, une étude de Pavlidis *et al.* (2012) critique la tendance à élaborer des scénarios évolutifs à partir de résultats d'enrichissements fonctionnels au sein de gènes candidats détectés par balayage du génome. Les auteurs ont effectué des tests de sélection positive sur des simulations d'une population évoluant sous neutralité et ont trouvé des gènes candidats enrichis dans des fonctions qui pourraient être interprétées dans le cadre d'un scénario évolutif. La volonté de corroborer des résultats à l'aide d'hypothèses basées sur des enrichissements de fonctions pourrait ainsi mener à une sur-interprétation des résultats. De plus, approximativement 7,2% des gènes chez l'Homme sont organisés en cluster fonctionnel, et un signal de sélection va souvent s'étendre sur les régions génomiques voisines d'un locus candidat. Cela peut mener à la sur-représentation d'une fonction au sein d'une liste de candidat. Dans l'analyse d'enrichissement fonctionnel de HapMap 3.3, c'était le cas pour une catégorie référant au développement de l'épiderme, et nous avons supprimé cette fonction de nos analyses.

Néanmoins, il est possible que cela se produise pour d'autres fonctions, et il serait nécessaire de corriger pour ce type d'effet systématiquement. Nous avons également mis en évidence un autre biais dans les analyses d'enrichissement fonctionnel. Par le seul fait du hasard, dans le cadre d'analyse de balayage du génome, les gènes longs ont tendance à arborer plus de SNP significatifs que les gènes courts. Nous avons mis cet effet en évidence dans nos analyses car le nombre de SNP montrant des différences significatives de fréquences alléliques entre les sexes par gène était fortement corrélé avec la longueur des gènes. Cela peut considérablement biaiser les analyses d'enrichissement fonctionnel car les gènes longs ont tendance à être impliqués dans des fonctions particulières, comme les fonctions neuronales chez l'Homme (Kraft et Raychaudhuri 2009). Dans nos analyses, nous avons corrigé pour ce biais en effectuant une correction FDR pour le nombre de tests effectués au sein de chaque gène avant de procéder à l'analyse d'enrichissement fonctionnel. Après correction, la corrélation mise en évidence était largement diminuée, mais la possibilité de biais résiduels non corrigés amène à considérer les analyses d'enrichissement fonctionnel avec précaution.

Une sélection SA ne peut mener à des différences de fréquences alléliques entre les sexes au sein d'une population que si celle-ci agit sur la survie. La sélection sexuelle seule, bien qu'elle puisse être à l'origine d'IASC, ne peut être à l'origine de différences de fréquences alléliques entre les mâles et les femelles au sein d'une population. Néanmoins, si le trait sélectionné est sous IASC, une sélection sexuelle sur ce trait peut modifier les fréquences alléliques entre les mâles et les femelles dans le sous-groupe des individus qui se sont reproduits. En effet, si les femelles préfèrent un mâle possédant une version d'un trait mais que ce trait est sous IASC, on s'attend à observer une plus grande proportion de mâles se reproduisant ayant ce trait, et une plus faible proportion de femelles se reproduisant ayant ce trait.

En conséquence, les différences de fréquences alléliques que nous avons observées dans GoNL chez les parents pourraient en partie être dues à des processus de choix du conjoint ou de compétition intra-sexuelle. Il serait intéressant d'étudier si l'on peut détecter des gènes qui seraient impliqués dans le choix du conjoint et qui montrent des différences de fréquences alléliques entre les sexes dans cette base de données. En effet, chez trois populations comportant des couples dans la base de données de HapMap 3.3, Laurent *et al.* (2012) ont identifiés des gènes qui étaient plus différents au sein des couples qu'au hasard. Ces gènes sont susceptibles d'entrer en jeu dans le choix du conjoint chez l'Homme.

Par ailleurs, les parents de GoNL et les individus en couple dans HapMap 3.3 sont fertiles. Si l'on considère uniquement au sein d'une population des couples qui se sont reproduits, cela pourrait mener à l'observation de différences de fréquences alléliques

entre les sexes au niveau de gènes sous sélection SA impliqués dans la fertilité. Or, au sein de la liste de gènes comportant au moins un SNP avec des différences de fréquences alléliques entre les sexes dans HapMap 3.3 et chez les parents de GoNL, nous avons trouvé un enrichissement dans des fonctions impliquées dans la production de gamètes et la reproduction.

Au contraire, les différences de fréquences alléliques observées chez les enfants dans GoNL sont plus probablement dues à des processus se déroulant entre la gamétogenèse et la petite enfance. Cela pourrait expliquer le peu de concordance entre les locus montrant un signal de sélection SA chez les parents et chez les enfants. Les parents ont en effet été recrutés dans la base de données parce qu'ils étaient fertiles et en couples, et les enfants parce qu'ils avaient jusqu'à présent survécu et sont *a priori* en bonne santé.

Il est donc possible qu'une partie des différences de fréquences alléliques entre les sexes que nous avons détectées chez les parents soit due à la sélection sexuelle, en particulier le choix du conjoint. L'étude théorique de Albert et Otto (2005) prédit que, dans un système XY, les femelles auront un avantage à préférer les mâles avec une fitness basse si les traits sous IASC sont codés par des gènes situés sur le chromosome X. En effet, si une femelle se reproduit avec un mâle ayant une fitness basse, leurs filles auront une fitness haute et leurs fils une fitness basse, mais si une femelle se reproduit avec un mâle ayant une fitness haute, leurs filles et leurs fils auront une fitness basse car les pères ne transmettent pas leur chromosome X à leurs fils. Cela serait très défavorable pour les mâles, qui recevraient plus souvent la version bénéfique aux femelles. Ce paradoxe pourrait expliquer l'émergence et le maintien de locus SA sur les autosomes. En effet, dans nos résultats, nous observons que le chromosome X semble enrichi en locus montrant un signal de sélection SA, mais les autosomes montrent également des proportions non négligeables de ces locus.

En introduction, nous avons discuté de l'effet du système de reproduction sur l'intensité de la sélection sexuelle. Selon le principe de Bateman (1948), l'intensité de la sélection sexuelle serait plus élevée dans les populations polygynes et polyandres que dans les populations monogames. On s'attend à ce que le gradient de Bateman varie entre les populations humaines, car il est probable qu'elles diffèrent au niveau des variations des valeurs sélectives des partenaires, du sexe-ratio à l'âge adulte, donc de l'OSR (voir introduction section 1.1.3) et de la densité de population (Brown *et al.* 2009). Chez l'Homme, on estime que 17% des populations sont monogames tandis que 51% sont faiblement polygynes, 31% sont fortement polygynes et 1% sont polyandres (Marlowe 2000). Une étude de Brown *et al.* (2009) a analysé, chez des populations humaines, la relation entre les différences de succès reproducteur (mesuré par le nombre de descendants) entre les sexes et le nombre d'accouplement. Les résultats montrent que les hommes ont une variation moyenne du

succès reproducteur plus élevée que les femmes, ce qui est en accord avec l'attendu de Bateman. Le ratio entre le succès reproducteur des hommes et celui des femmes varie entre les populations considérées, et ces variations sont corrélées avec le système de reproduction : les ratios sont plus élevés chez les populations polygynes que chez les populations monogames. Les populations montrant une monogamie sérielle, c'est-à-dire où les individus forment des couples monogames mais changent de partenaires au cours de leur vie, montrent un ratio de succès reproducteur plus proche des populations polygynes (Brown *et al.* 2009). Les données disponibles sur le nombre d'accouplements suggèrent qu'il existe une variance plus élevée chez les hommes en comparaison aux femmes. Par exemple, dans les sociétés occidentales, les hommes ont une probabilité de remariage plus élevée que les femmes, ce qui est probablement dû à l'absence de ménopause chez l'homme (Buckle *et al.* 1996). Néanmoins, il est difficile d'estimer les différences de nombre de partenaires entre les hommes et les femmes des populations humaines à l'aide des données disponibles. Chez l'Homme, il est donc compliqué de déterminer la relation entre le nombre de descendants et le nombre de partenaires, et donc d'étudier le gradient de Bateman. L'étude de Brown *et al.* (2009) suggère néanmoins que la sélection sexuelle agirait de façon différente entre les populations humaines. Comme la sélection sexuelle peut mener à l'émergence de conflits sexuels, notamment d'IASC, et que les populations humaines ont des systèmes de reproduction divers, il serait intéressant d'étudier en parallèle dans des populations humaines les mécanismes de sélection sexuelle et les signaux génétiques d'IASC au sein de couples afin de déterminer s'il existe des différences entre les populations monogames et polygynes. Néanmoins, nous avons montré durant notre étude que les locus potentiellement sous sélection SA étaient majoritairement spécifiques aux populations. Il serait donc difficile de conclure si d'éventuelles différences sont dues aux régimes de sélection sexuelle si les populations étudiées sont géographiquement éloignées.

Par ailleurs, des processus de sélection sexuelle différents entre les populations pourraient expliquer le peu de concordance des signaux d'IASC entre les populations d'HapMap 3.3. En effet, 5 des 10 populations de HapMap 3.3 comportent des couples faisant partie de trios parents-enfant, ce qui signifie qu'au sein de ces populations des sous-groupes d'individus se sont reproduits au moins une fois et que leurs enfants sont *a priori* en bonne santé. Il a été montré que des processus de choix du conjoint en fonction du génotype ont lieu dans au moins trois de ces populations (Laurent *et al.* 2012) et les gènes impliqués dans les processus de choix du conjoint mis en évidence par Laurent *et al.* (2012) étaient différents entre les 3 populations étudiées. Cela suggèrent que, au moins pour un sous groupe au sein de 5 populations, des processus de choix du conjoint pourraient être à l'origine de différences de fréquences alléliques entre les sexes. Si les traits sélectionnés sont différents, les signaux de sélection SA ne seraient pas concordants

entre populations.

De nombreuses études ont montré que la sélection SA agissait chez les adultes dans des populations naturelles et expérimentales (notamment Chippindale *et al.* 2001 ; Innocenti et Morrow 2010 ; Delcourt *et al.* 2009 ; Foerster *et al.* 2007 ; Brommer *et al.* 2007 ; Prasad *et al.* 2007). Une méta-analyse des données présentes dans la littérature pour 34 espèces d’animaux a montré que, chez l’adulte, la sélection SA est plus faible sur la viabilité que sur la reproduction (Cox et Calsbeek 2009). Par ailleurs, l’existence d’IASC chez les juvéniles, agissant sur la viabilité, est controversée dans la littérature : chez *Drosophila melanogaster*, l’étude de Chippindale *et al.* (2001) suggère qu’il n’y a pas ou peu d’IASC chez les juvéniles tandis que l’étude de Prasad *et al.* (2007) suggère l’inverse. Aucune étude ne s’est intéressée à la sélection SA se déroulant avant la naissance. À partir de ces observations et de nos résultats, nous pouvons spéculer qu’une sélection SA sur la viabilité agirait entre la gamétogénèse et la naissance, suivie d’une plus faible sélection SA sur la viabilité pendant l’enfance. Enfin, à l’âge adulte, une sélection SA agirait sur la viabilité et, de façon plus importante, sur la reproduction.

Bien que nous ne les ayons pas considérées au cours de cette thèse, les régions pseudo-autosomales sont particulièrement intéressantes pour l’étude de la sélection SA. Les PAR sont des régions où la recombinaison est maintenue entre les chromosomes sexuels X et Y (ou Z et W). Elles ont donc une transmission proche de celles des régions autosomales mais sont néanmoins fortement liées à la région déterminant le sexe (SDR *Sex Determining Region*) et leur dynamique évolutive est très différente de celle d’un autosome. De plus, l’accumulation de locus SA sur les chromosomes sexuels pourrait être à l’origine de l’élargissement de la zone non recombinante entre le X et le Y (Bull 1983 ; Rice 1987) (voir section 1.2.3). Il a été montré qu’un polymorphisme SA à un locus lié aux PAR et proche de la SDR est maintenu pour une gamme de paramètres plus vaste que pour un locus lié à l’X non recombinant avec le Y ou à un locus autosomal (Jordan et Charlesworth 2012). Une étude récente de Kirkpatrick et Guerrero (2014), a permis de prédire des signatures de sélection SA au sein des PAR à l’aide de modélisations de régions pseudo-autosomales comportant ou non un locus SA. Trois indicateurs ont été considérés : le temps de coalescence, le F_{ST} (entre des locus du X et du Y et entre les mâles et les femelles) et le LD. Cette étude prédit une asymétrie du signal de LD et de F_{ST} entre le locus sous sélection SA et la SDR. Dans nos analyses, nous avons développé des méthodes de calcul du F_{ST} et du LD, il serait donc réalisable d’appliquer ces méthodes sur les PAR dans GoNL. Néanmoins, une méthode pour évaluer l’asymétrie du signal devra être mise au point, ainsi qu’un phasage des données si nous utilisons le F_{ST} entre le chromosome X et le

chromosome Y.

À notre connaissance, cette thèse constitue la première étude s'intéressant à des signaux génomiques de sélection SA à l'aide d'un balayage du génome, ainsi qu'à des processus de distorsion de transmission sexe-spécifiques. Il serait intéressant de continuer à étudier ces signaux afin d'en apprendre plus sur les mécanismes génétiques à l'origine de l'émergence des dimorphismes sexuels.

Références bibliographiques

- Aiken C.E., Ozanne S.E. (2013) Sex differences in developmental programming models. *Reproduction*, **145**, R1–13.
- Albert A.Y.K., Otto S.P. (2005) Sexual selection can resolve sex-linked sexual antagonism. *Science*, **310**, 119–21.
- Altschup S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **215**, 403–410.
- Altshuler D.M., Gibbs R.A., Peltonen L. *et al.* (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52–8.
- Andersson M., Iwasa Y. (1996) Sexual selection. *Trends in Ecology and Evolution*, **11**, 53–58.
- Arnqvist G., Rowe I. (2005) *Sexual Conflict*. Princeton University Press.
- Bachtrog D. (2006) A dynamic view of sex chromosome evolution. *Current opinion in genetics & development*, **16**, 578–85.
- Badyaev A.V. (2002) Growing apart: an ontogenetic perspective on the evolution of sexual size dimorphism. *Trends in Ecology and Evolution*, **17**, 369–378.
- Balaresque P., Toupance B., Quintana-Murci L., Crouau-Roy B., Heyer E. (2004) Sex-specific selection on the human X chromosome? *Genetical Research*, **83**, 169–176.
- Barrett S.C.H., Hough J. (2013) Sexual dimorphism in flowering plants. *Journal of Experimental Botany*, **64**, 67–82.
- Bateman A. (1948) Intra sexual selection in *Drosophila*. *Heredity*, **2**, 349–368.
- Beeman R.W., Friesen K.S. (1999) Properties and natural occurrence of maternal-effect selfish genes ('Medea' factors) in the Red Flour Beetle, *Tribolium castaneum*. *Heredity*, **82**, 529–534.

- Beeman R.W., Friesen K.S., Denell R.E. (1992) Maternal-effect selfish genes in flour beetles. *Science*, **256**, 89–92.
- Benagiano G., Farris M., Grudzinskas G. (2010) Fate of fertilized human oocytes. *Reproductive biomedicine online*, **21**, 732–41.
- Benjamini Y., Hochberg Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.
- Blas J., Pérez-Rodríguez L., Bortolotti G.R., Viñuela J., Marchant T.A. (2006) Testosterone increases bioavailability of carotenoids: insights into the honesty of sexual signaling. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 18633–18637.
- Bonduriansky R. (2007) The genetic architecture of sexual dimorphism: the potential roles of genomic imprinting and condition-dependence. In *Sex, Size & Gender Role* (edited by DJ Fairbairn, WU Blanckenhorn, T Székely), pp. 176–184. Oxford University Press, Oxford.
- Bonduriansky R., Chenoweth S.F. (2009) Intralocus sexual conflict. *Trends in ecology and evolution*, **24**, 280–8.
- Boomsma D.I., Wijmenga C., Slagboom E.P. *et al.* (2014) The Genome of the Netherlands: design, and project goals. *European journal of human genetics*, **22**, 221–7.
- Brommer J.E., Kirkpatrick M., Qvarnström A., Gustafsson L. (2007) The intersexual genetic correlation for lifetime fitness in the wild and its implications for sexual selection. *PloS one*, **2**, e744.
- Brown G.R., Laland K.N., Mulder M.B. (2009) Bateman’s principles and human sex roles. *Trends in Ecology and Evolution*, **24**, 297–304.
- Buckle L., Gallup G.G., Rodd Z.A. (1996) Marriage as a reproductive contract: Patterns of marriage, divorce, and remarriage. *Ethology and Sociobiology*, **17**, 363–377.
- Bull J. (1983) *Evolution of Sex Determining Mechanisms*. Benjamin/Cummings Publ. Co., Menlo Park, California.
- Burt A., Trivers R.L. (2006) *Genes in Conflict: The Biology of Selfish Genetic Elements*. Belknap Press of Harvard University Press, Cambridge, MA.

- Calsbeek R., Sinervo B. (2003) Within-clutch variation in offspring sex determined by differences in sire body size: cryptic mate choice in the wild. *Journal of Evolutionary Biology*, **17**, 464–470.
- Casa-esperon E., De Villena F.P.M., Verner A.E. *et al.* (2000) Sex-Of-Offspring-Specific Transmission Ratio Distortion on Mouse Chromosome X. *Genetics*, **154**, 343–350.
- Chapman T. (2008) The soup in my fly: Evolution, form and function of seminal fluid proteins. *PLoS Biology*, **6**, 1379–1382.
- Chapman T., Arnqvist G., Bangham J., Rowe L. (2003) Sexual conflict. *Trends in Ecology and Evolution*, **18**, 41–47.
- Chapman T., Liddle L.F., Kalb J.M., Wolfner M.F., Partridge L. (1995) Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature*, **373**, 241–244.
- Charlesworth D., Charlesworth B. (1980) Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genetical research*, **35**, 205–214.
- Chippindale A.K., Gibson J.R., Rice W.R. (2001) Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 1671–5.
- Choi S.K., Yoon S.R., Calabrese P., Arnheim N. (2008) A germ-line-selective advantage rather than an increased mutation rate can explain some unexpectedly common human disease mutations. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 10143–10148.
- Clutton-Brock T.H. (2007) Sexual Selection in Males and Females. *Science*, **318**, 1882–1885.
- Clutton-Brock T.H. (2009) Sexual selection in females. *Animal Behaviour*, **77**, 3–11.
- Clutton-Brock T.H., Vincent A. (1991) Sexual selection and the potential reproductive rates of males and females. *Nature*, **354**, 56–58.
- Connallon T., Jakubowski E. (2009) Association between sex ratio distortion and sexually antagonistic fitness consequences of female choice. *Evolution*, **63**, 2179–2183.
- Cox R.M., Calsbeek R. (2009) Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *The American naturalist*, **173**, 176–87.

- Cupples L.A., Arruda H.T., Benjamin E.J. *et al.* (2007) The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. *BMC medical genetics*, **8**(Suppl 1), S1.
- Darwin C. (1871) *The Descent of Man, and Selection in Relation to Sex*. John Murray, London.
- Dawkins R. (1976) *The selfish gene*. Oxford University Press, Oxford.
- Dean R., Perry J.C., Pizzari T., Mank J.E., Wigby S. (2012) Experimental evolution of a novel sexually antagonistic allele. *PLoS genetics*, **8**, e1002917.
- Delcourt M., Blows M.W., Rundle H.D. (2009) Sexually antagonistic genetic variance for fitness in an ancestral and a novel environment. *Proceedings of the Royal Society B*, **276**, 2009–14.
- Deng L., Zhang D., Richards E. *et al.* (2009) Constructing an initial map of transmission distortion based on high density HapMap SNPs across the human autosomes. *Journal of Genetics and Genomics*, **36**, 703–709.
- Dixson A.F. (2015) *Primate sexuality : comparative studies of the prosimians, monkeys, apes, and human beings*. Oxford University Press, Oxford, 2nd edn..
- van Doorn G.S. (2009) Intralocus sexual conflict. *Annals of the New York Academy of Sciences*, **1168**, 52–71.
- Ellegren H., Parsch J. (2007) The evolution of sex-biased genes and sex-biased gene expression. *Nature reviews Genetics*, **8**, 689–98.
- Emlen S.T., Oring L.W. (1977) Ecology , Sexual Selection , and the Evolution of Mating Systems. *Science*, **197**, 215–223.
- Fedorka K.M., Mousseau T.A. (2004) Female mating bias results in conflicting sex-specific offspring fitness. *Nature*, **234**, 65–67.
- Fenner J.N. (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *American journal of physical anthropology*, **128**, 415–23.
- Fisher R. (1958) *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, 2nd edn..

- Foerster K., Coulson T., Sheldon B.C., Pemberton J.M., Clutton-Brock T.H., Kruuk L.E.B. (2007) Sexually antagonistic genetic variation for fitness in red deer. *Nature*, **447**, 1107–10.
- Fry J.D. (2010) The genomic location of sexually antagonistic variation: some cautionary comments. *Evolution*, **64**, 1510–6.
- Galimberti F., Boitani L., Marzetti I. (2000) Female strategies of harassment reduction in southern elephant seals. *Ethology Ecology & Evolution*, **12**, 367–388.
- Garver-Apgar C.E., Eaton M.A., Tybur J.M., Emery Thompson M. (2011) Evidence of intralocus sexual conflict: physically and hormonally masculine individuals have more attractive brothers relative to sisters. *Evolution and Human Behavior*, **32**, 423–432.
- Geschwind D.H., Sowiński J., Lord C. *et al.* (2001) The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *American journal of human genetics*, **69**, 463–466.
- Gibson J.R., Chippindale A.K., Rice W.R. (2002) The X chromosome is a hot spot for sexually antagonistic fitness variation. *Proceedings of the Royal Society B*, **269**, 499–505.
- Glucksmann A. (1981) *Sexual Dimorphism in Human and Mammalian Biology and Pathology*. Academic Press, London.
- Gnad F., Parsch J. (2006) Sebida: A database for the functional and evolutionary analysis of genes with sex-biased expression. *Bioinformatics*, **22**, 2577–2579.
- Goriely A., McVean G.A.T., van Pelt A.M.M. *et al.* (2005) Gain-of-function amino acid substitutions drive positive selection of FGFR2 mutations in human spermatogonia. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6051–6056.
- Graves J. (2006) Sex chromosome specialization and degeneration in mammals. *Cell*, **124**, 901–914.
- Graves J., Shetty S. (2001) Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes. *The Journal of experimental zoology*, **290**, 449–62.
- Griffith S.C., Owens I.P.F., Thuman K.A. (2002) Extra pair paternity in birds: A review of interspecific variation and adaptive function. *Molecular Ecology*, **11**, 2195–2212.

- Gwynne D.T. (1984) Courtship feeding increases female reproductive success in bushcrickets. *Nature*, **307**, 361–363.
- Hager R., Cheverud J.M., Leamy L.J., Wolf J.B. (2008) Sex dependent imprinting effects on complex traits in mice. *BMC evolutionary biology*, **8**, 303.
- Hamilton W.D., Zuk M. (1982) Heritable true fitness and bright birds: a role for parasites? *Science*, **218**, 384–387.
- Hastings I.M. (1989) Potential germline competition in animals and its evolutionary implications. *Genetics*, **123**, 191–197.
- Helleu Q., Gérard P.R., Montchamp-Moreau C. (2015) Sex Chromosome Drive. *Cold Spring Harbor Perspectives in Biology*, **7**, a017616.
- Henderson I.G., Hart P.J.B., Burke T. (2000) Strict monogamy in a semi-colonial passerine: The Jackdaw *Corvus monedula*. *Journal of Avian Biology*, **31**, 177–182.
- Huang D.W., Sherman B.T., Lempicki R.A. (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*, **37**, 1–13.
- Huang D.W., Sherman B.T., Lempicki R.A. (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, **4**, 44–57.
- Huang L.O., Labbe A., Infante-Rivard C. (2013) Transmission ratio distortion: Review of concept and implications for genetic association studies. *Human Genetics*, **132**, 245–263.
- Hudson R.R. (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.
- Hull K.L., Harvey S. (2001) Growth hormone: Roles in female reproduction. *Journal of Endocrinology*, **168**, 1–23.
- Hurst L.D., Pomiankowski A. (1991) Causes of sex ratio bias may account for unisexual sterility in hybrids: A new explanation of Haldane's rule and related phenomena. *Genetics*, **128**, 841–858.
- Innocenti P., Morrow E.H. (2010) The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS biology*, **8**, e1000335.
- Jaenike J. (2001) Sex Chromosome Meiotic Drive. *Annual Review of Ecology and Systematics*, **32**, 25–49.

- Jones A.G., Ratterman N.L. (2009) Mate choice and sexual selection: what have we learned since Darwin? *Proceedings of the National Academy of Sciences of the United States of America*, **106** (Suppl, 10001–10008.
- Jordan C.Y., Charlesworth D. (2012) The potential for sexually antagonistic polymorphism in different genome regions. *Evolution*, **66**, 505–16.
- Kidwell J.F., Clegg M.T., Stewart F.M., Prout T. (1977) Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics*, **85**, 171–183.
- Kirkpatrick M., Guerrero R.F. (2014) Signatures of Sex-Antagonistic Selection on Recombining Sex Chromosomes. *Genetics*, **197**, 531–541.
- Kirkpatrick M., Ryan M.J. (1991) The evolution of mating preferences and the paradox of the lek. *Nature*, **350**, 33–38.
- Kitzis A., Chomel J.C., Kaplan J.C. *et al.* (1988) Unusual segregation of cystic fibrosis allele to males. *Nature*, **333**, 215.
- Klein S.L. (2012) Immune cells have sex and so should journal articles. *Endocrinology*, **153**, 2544–50.
- Kopp A., Graze R.M., Xu S., Carroll S.B., Nuzhdin S.V. (2003) Dimorphic Traits in *Drosophila melanogaster*. *Genetics*, **787**, 771–787.
- Kraft P., Raychaudhuri S. (2009) Complex diseases, complex genes: keeping pathways on the right track. *Epidemiology (Cambridge, Mass.)*, **20**, 19–23.
- Lack D.L. (1968) *Ecological Adaptations for Breeding in Birds*, vol. 163. Methuen & Co. Ltd, London.
- Lahn B.T., Pearson N.M., Jegalian K. (2001) The human Y chromosome, in the light of evolution. *Nature reviews Genetics*, **2**, 207–216.
- Lande R. (1980) Sexual Dimorphism, Sexual Selection, and Adaptation in Polygenic Characters. *Evolution*, **34**, 292–305.
- Lango Allen H., Estrada K., Lettre G. *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, **467**, 832–838.
- Larracuenta A.M., Presgraves D.C. (2012) The selfish Segregation Distorter gene complex of *Drosophila melanogaster*. *Genetics*, **192**, 33–53.

- Laurent R., Toupance B., Chaix R. (2012) Non-random mate choice in humans: insights from a genome scan. *Molecular ecology*, **21**, 587–96.
- Liu L.Y., Schaub M.A., Sirota M., Butte A.J. (2012) Transmission distortion in Crohn's disease risk gene ATG16L1 leads to sex difference in disease association. *Inflammatory Bowel Diseases*, **18**, 312–322.
- Long T.A.F., Rice W.R. (2007) Adult locomotory activity mediates intralocus sexual conflict in a laboratory-adapted population of *Drosophila melanogaster*. *Proceedings of the Royal Society B*, **274**, 3105–12.
- Lyon M.F. (2003) Transmission Ratio Distortion in Mice. *Annual Review of Genetics*, **37**, 393–408.
- Mank J.E., Hultin-Rosenberg L., Zwahlen M., Ellegren H. (2008) Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *The American naturalist*, **171**, 35–43.
- Marlowe F. (2000) Paternal investment and the human mating system. *Behavioural Processes*, **51**, 45–61.
- Marriott I., Huet-Hudson Y.M. (2006) Sexual dimorphism in innate immune responses to infectious organisms. *Immunologic research*, **34**, 177–92.
- McIntyre L.M., Bono L.M., Genissel A. *et al.* (2006) Sex-specific expression of alternative transcripts in *Drosophila*. *Genome biology*, **7**, R79.
- Merila J., Sheldon B., Ellegren H. (1998) Quantitative genetics of sexual size dimorphism in the collared flycatcher, *Ficedula albicollis*. *Evolution*, **52**, 870–876.
- Meyer W.K., Arbeithuber B., Ober C. *et al.* (2012) Evaluating the evidence for transmission distortion in human pedigrees. *Genetics*, **191**, 215–232.
- Mitchell A.A., Cutler D.J., Chakravarti A. (2003) Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. *American journal of human genetics*, **72**, 598–610.
- Mullon C., Pomiankowski A., Reuter M. (2012) The effects of selection and genetic drift on the genomic distribution of sexually antagonistic alleles. *Evolution*, **66**, 3743–53.
- Nookaew I., Svensson P.A., Jacobson P. *et al.* (2013) Adipose tissue resting energy expenditure and expression of genes involved in mitochondrial function are higher in women than in men. *The Journal of clinical endocrinology and metabolism*, **98**, E370–8.

- Ober C., Abney M., McPeck M.S. (2001) The genetic dissection of complex traits in a founder population. *American journal of human genetics*, **69**, 1068–1079.
- Ogilvy-Stuart A.L., Shalet S.M. (1992) Growth hormone and puberty. *Journal of Endocrinology*, **135**, 405–406.
- Ohno S. (1967) *Sex Chromosomes and Sex Linked Genes*. Springer-Verlag, Berlin.
- Orians G. (1969) On the evolution of mating systems in birds and mammals. *The American naturalist*, **103**, 589–603.
- Oring L.W., Lank D.B. (1982) Sexual selection, arrival times, philopatry and site fidelity in the polyandrous spotted sandpiper. *Behavioral Ecology and Sociobiology*, **10**, 185–191.
- Otto S.P., Pannell J.R., Peichel C.L. *et al.* (2011) About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends in Genetics*, **27**, 358–67.
- Parker G.A. (1979) Sexual selection and sexual conflict. In *Sexual Selection and reproductive competition in insects* (edited by M Blum, N Blum), pp. 123–166. Academic Press, London.
- Parker G.A., Baker R.R., Smith V.G. (1972) The origin and evolution of gamete dimorphism and the male-female phenomenon. *Journal of theoretical biology*, **36**, 529–553.
- Parker G.A., Partridge L. (1998) Sexual conflict and speciation. *Philosophical Transaction of the Royal Society of London B*, **353**, 261–273.
- Paterson A.D., Waggott D., Schillert A. *et al.* (2009) Transmission-ratio distortion in the Framingham Heart Study. *BMC proceedings*, **3 Suppl 7**, S51.
- Patten M.M. (2014) Meiotic drive influences the outcome of sexually antagonistic selection at a linked locus. *Journal of Evolutionary Biology*, **27**, 2360–2370.
- Patten M.M., Haig D. (2009) Maintenance or loss of genetic variation under sexual and parental antagonism at a sex-linked locus. *Evolution*, **63**, 2888–95.
- Patten M.M., Haig D., Úbeda F. (2010) Fitness variation due to sexual antagonism and linkage disequilibrium. *Evolution*, **64**, 3638–3642.
- Pavlidis P., Jensen J.D., Stephan W., Stamatakis A. (2012) A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution*, **29**, 3237–3248.

- Pennell T.M., Morrow E.H. (2013) Two sexes, one genome: the evolutionary dynamics of intralocus sexual conflict. *Ecology and Evolution*, **3**, 1819–1834.
- Pischedda A., Chippindale A.K. (2006) Intralocus sexual conflict diminishes the benefits of sexual selection. *PLoS biology*, **4**, e356.
- Prasad N.G., Bedhomme S., Day T., Chippindale a.K. (2007) An evolutionary cost of separate genders revealed by male-limited evolution. *The American naturalist*, **169**, 29–37.
- Préault M., Chastel O., Cézilly F., Faivre B. (2005) Male bill colour and age are associated with parental abilities and breeding performance in blackbirds. *Behavioral Ecology and Sociobiology*, **58**, 497–505.
- Price D.K., Burley N.T. (1993) Constraints on the evolution of attractive traits: genetic (co)variance of zebra finch bill colour. *Heredity*, **71**, 405–12.
- Price D.K., Burley N.T. (1994) Constraints on the evolution of attractive traits: selection in male and female zebra finches. *American Naturalist*, **144**, 908–934.
- Queller D.C. (1997) Why do females care more than males? *Proceedings of the Royal Society B*, **264**, 1555–1557.
- Randerson J.R., Hurst L.D. (2001) The uncertain evolution of the sexes. *Trends in Ecology and Evolution*, **16**, 571–579.
- Ranz J.M., Castillo-Davis C.I., Meiklejohn C.D., Hartl D.L. (2003) Sex-dependent gene expression and evolution of the Drosophila transcriptome. *Science*, **300**, 1742–1745.
- Reeve H.K., Pfennig D.W. (2003) Genetic biases for showy males: are some genetic systems especially conducive to sexual selection? *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 1089–1094.
- Rice W.R. (1984) Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, **38**, 735–742.
- Rice W.R. (1987) The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. *Evolution*, **41**, 911–914.
- Rice W.R., Chippindale A.K. (2002) The evolution of hybrid infertility: Perpetual coevolution between gender-specific and sexually antagonistic genes. *Genetica*, **116**, 179–188.

- Robinson M.R., Pilkington J.G., Clutton-Brock T.H., Pemberton J.M., Kruuk L.E.B. (2006) Live fast, die young: trade-offs between fitness components and sexually antagonistic selection on weaponry in Soay sheep. *Evolution*, **60**, 2168–2181.
- Ross M., Grafham D.V., Coffey A.J., Scherer S. (2005) The DNA sequence of the human X chromosome. *Nature*, **434**, 325–37.
- Santos P.S.C., Höhne J., Schlattmann P. *et al.* (2009) Assessment of transmission distortion on chromosome 6p in healthy individuals using tagSNPs. *European journal of human genetics*, **17**, 1182–1189.
- Schimenti J. (2000) Segregation distortion of mouse t haplotypes - The molecular basis emerges. *Trends in Genetics*, **16**, 240–243.
- Shi H., Clegg D.J. (2009) Sex differences in the regulation of body weight. *Physiology & behavior*, **97**, 199–204.
- Shine R. (1989) Ecological Causes for the Evolution of Sexual Dimorphism : A Review of the Evidence. *The Quarterly Review of Biology*, **64**, 419–461.
- Siracusa L.D., Alvord W.G., Bickmore W.A., Jenkins N.A., Copeland N.G. (1991) Interspecific backcross mice show sex-specific differences in allelic inheritance. *Genetics*, **128**, 813–821.
- Slatkin M. (1984) Ecological causes of sexual dimorphism. *Evolution*, **38**, 622–630.
- Spielman R.S., McGinnis R.E., Ewens W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American journal of human genetics*, **52**, 506–516.
- Stulp G., Kuijper B., Buunk A.P., Pollet T.V., Verhulst S. (2012) Intralocus sexual conflict over human height. *Biology letters*, **8**, 976–8.
- Taylor P.D., Williams G.C. (1982) The lek paradox is not resolved. *Theoretical Population Biology*, **22**, 392–409.
- The 1000 Genome Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–73.
- The Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature genetics*, **46**, 818–828.

- The HapMap International Consortium (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–320.
- Trivers R.L. (1972) Parental investment and sexual selection. In *Sexual selection and the descent of man*, vol. 12, pp. 136–179. Aldine Publishing Company, b.campbell edn..
- Úbeda F., Haig D. (2005) On the evolutionary stability of Mendelian segregation. *Genetics*, **170**, 1345–1357.
- Ubeda F., Haig D., Patten M.M. (2011) Stable linkage disequilibrium owing to sexual antagonism. *Proceedings of the Royal Society B*, **278**, 855–62.
- Ucisik-Akkaya E., Davis C.F., Do T.N. *et al.* (2010) Examination of genetic polymorphisms in newborns for signatures of sex-specific prenatal selection. *Molecular human reproduction*, **16**, 770–7.
- Vicoso B., Charlesworth B. (2009) Effective population size and the faster-X effect: An extended model. *Evolution*, **63**, 2413–2426.
- Visscher P.M., Medland S.E., Ferreira M.A.R. *et al.* (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genetics*, **2**, 0316–0325.
- Weir B.S. (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, Massachusetts.
- Weir B.S., Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wigginton J.E., Cutler D.J., Abecasis G.R. (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American journal of human genetics*, **76**, 887–93.
- Williams T.M., Carroll S.B. (2009) Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nature reviews. Genetics*, **10**, 797–804.
- Zahavi A. (1975) Mate selection—a selection for a handicap. *Journal of theoretical biology*, **53**, 205–214.
- Zöllner S., Wen X., Hanchard N.A., Herbert M.A., Ober C., Pritchard J.K. (2004) Evidence for extensive transmission distortion in the human genome. *American journal of human genetics*, **74**, 62–72.

Annexes du chapitre 3

Cluster fonctionnel - score d'enrichissement : 1.38	
Termes	p-value
Nucleosome	0.004
Nucleosome assembly	0.006
Protein-DNA complex	0.007
Chromatin assembly	0.007
Protein-DNA complex assembly	0.009
Nucleosome organization	0.010
DNA packaging	0.014
Chromatin assembly or disassembly	0.054
Chromatin organization	0.108
Chromosome organization	0.156
Cellular macromolecular complex assembly	0.230
Chromosome	0.240
Cellular macromolecular complex subunit organization	0.269
Chromosomal part	0.276
Chromatin	0.324

Table 4.1 – Cluster fonctionnel pour les gènes comportant au moins un *genic* SD SNP classés selon leur score d'enrichissement.

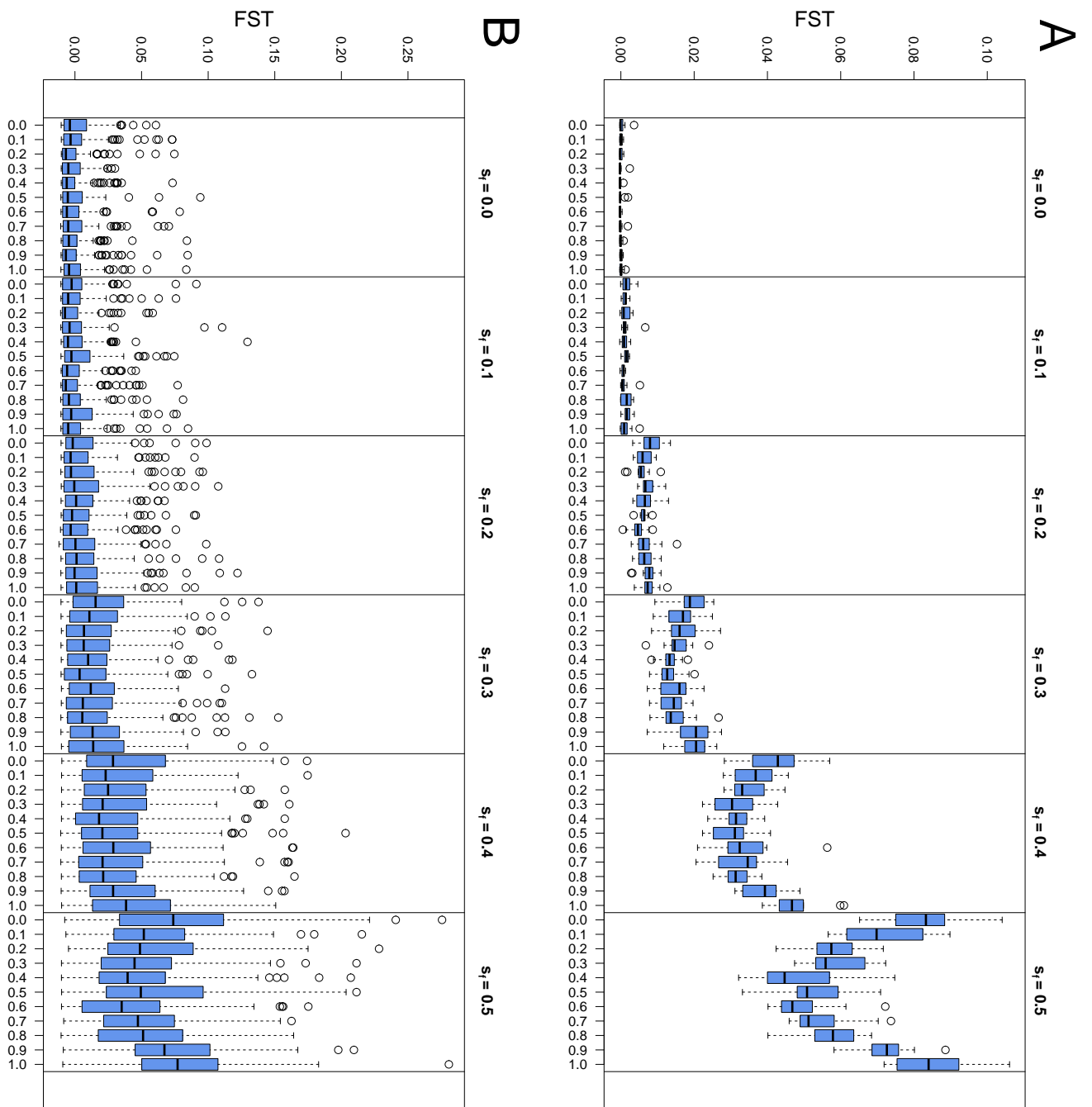


Figure 4.1 – Distributions du F_{ST} intersexuel pour le locus SA pour l'autosome. **A** - sur l'ensemble des 3000 individus simulés. **B**- Sur 10 ré-échantillonnages de 100 individus. Ces distributions sont construites sur les 10 répétitions, pour des valeurs de s_f comprises entre 0,0 et 0,5 et des valeurs de h_f comprise entre 0,0 et 1,0.

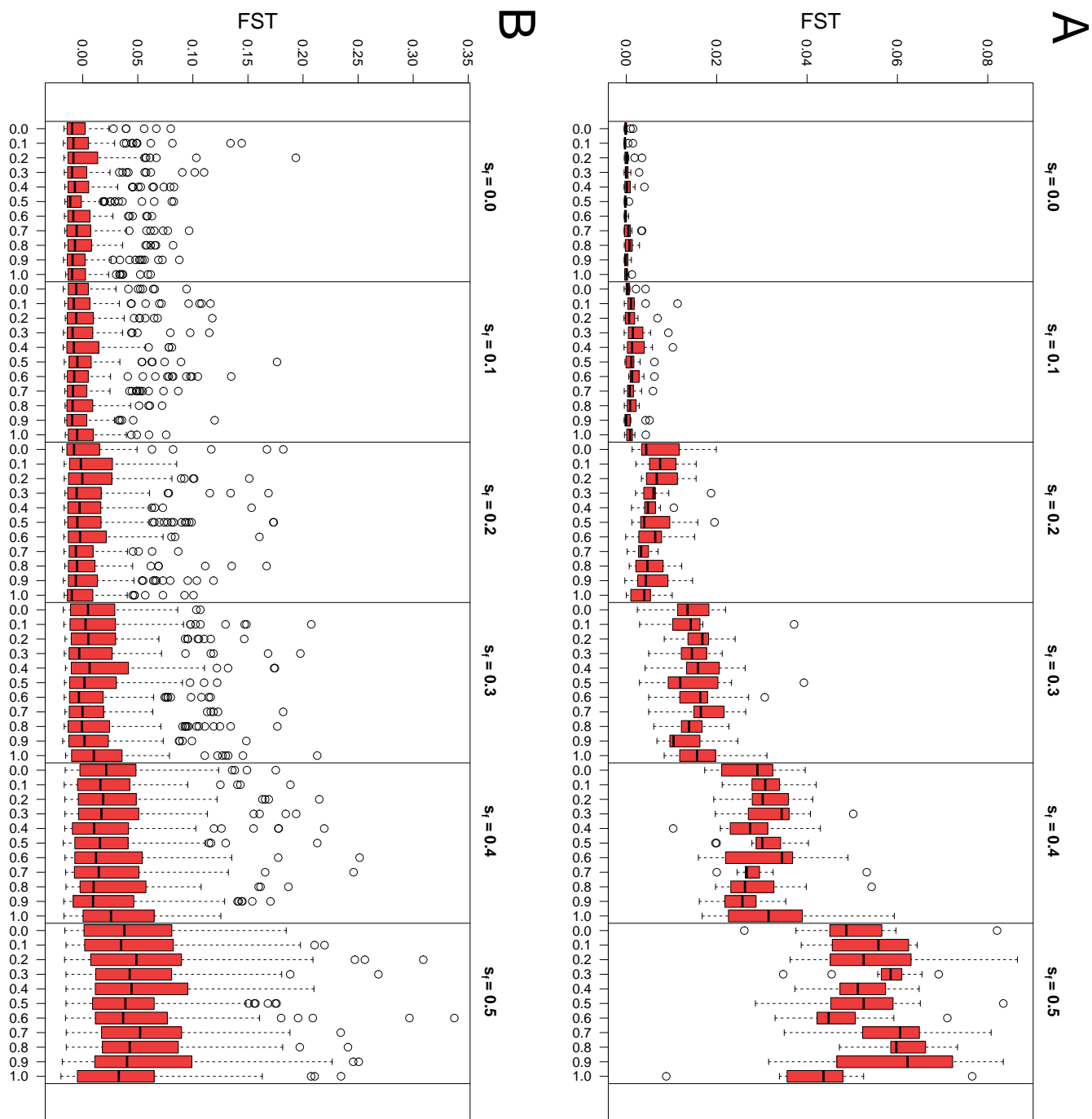


Figure 4.2 – Distributions du F_{ST} intersexuel pour le locus SA pour le **chromosome X**. **A** - sur l'ensemble des 3000 individus simulés. **B**- Sur 10 ré-échantillonnages de 100 individus. Ces distributions sont construites sur les 10 répétitions, pour des valeurs de s_f comprises entre 0,0 et 0,5 et des valeurs de h_f comprise entre 0,0 et 1,0.

s_f/h_f	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	1.20	1.07	1.12	0.77	0.95	0.61	1.17	0.92	0.62	1.02	0.69
0.2	2.73	2.36	1.81	2.00	1.71	1.69	1.91	1.59	1.70	1.68	2.24
0.0	3.81	3.09	2.96	2.81	2.69	2.62	3.03	3.08	2.75	3.19	4.27
0.4	5.62	4.99	4.15	4.43	3.83	3.87	4.15	3.83	4.73	4.88	5.67
0.5	8.38	7.41	6.56	5.26	4.92	5.22	5.02	4.88	6.30	6.85	7.70

Table 4.2 – Matrice de valeurs du $F_{ST}/\text{nombre de mort}$ multiplié par 10^5 pour les autosomes.

s_f/h_f	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	1.76	0.95	0.52	0.65	1.11	0.55	1.44	0.60	1.16	0.74	0.58
0.2	1.84	2.32	1.88	1.39	2.22	1.81	1.91	1.61	2.21	1.86	1.99
0.3	3.73	3.68	3.47	3.20	2.35	3.35	2.42	2.76	2.41	1.73	2.19
0.4	5.14	4.74	5.08	4.52	4.35	4.07	3.62	3.24	3.32	3.52	2.33
0.5	7.17	6.11	6.41	5.30	5.47	4.61	4.89	4.29	4.22	4.07	4.46

Table 4.3 – Matrice de valeurs du $F_{ST}/\text{nombre de mort}$ multiplié par 10^5 pour le chromosome X.

chr	total	3 tests	SDT	TD_{fille} + SDT	TD_{fils} + SDT	TD_{fils}	TD_{fille}	$TD_{fille} +$ TD_{fils}
X	95551	6	1571	738	928	1617	2518	22
1	333021	264	5707	3106	5248	7069	8709	301
2	375757	262	5704	3208	5153	7406	9307	195
3	326544	143	5841	2460	3751	6113	8698	196
4	332710	87	5215	3271	4236	6386	8600	224
5	293889	133	5146	2813	5519	6746	6861	251
6	297987	301	4661	2277	4622	5805	7894	131
7	258595	139	3896	2549	3256	5217	7173	207
8	253431	191	3953	2864	2783	4996	6617	255
9	190411	139	3006	2041	2368	3086	5080	87
10	226825	170	3231	2181	3192	4740	5350	142
11	224548	121	3934	1830	2807	3924	5490	100
12	212128	132	2620	1925	2431	4078	6154	205
13	167672	89	2846	1434	2348	4123	5120	276
14	145487	58	2616	1286	2106	2775	4078	65
15	123676	101	1885	881	1919	3124	2997	68
16	126705	70	1838	886	1329	2323	2671	73
17	100378	49	1383	933	1607	1797	3106	60
18	131215	103	2337	1335	2021	2363	3182	67
19	72893	14	1231	1040	1383	1868	2421	113
20	95371	98	1381	633	1431	1938	2145	38
21	61851	23	644	341	654	1120	1615	88
22	49136	14	650	453	692	901	1417	39

Table 4.4 – Compte du nombre de SNP significatifs pour les différentes combinaisons des tests effectués : SDT, TD_{fille} et TD_{fils} , pour chaque chromosome.

« Sensory perception and cognition »- score d'enrichissement : 1.98	
Termes	p-value
Sensory perception of chemical stimulus	1.61E-04
Sensory perception	0.003
Sensory perception of smell	0.009
Cognition	0.012
Olfactory receptor activity	0.018
G-protein coupled receptor protein signaling pathway	0.019
Neurological system process	0.054
Cell surface receptor linked signal transduction	0.143
« Lipase activity »- score d'enrichissement : 1.47	
Termes	p-value
Lipase activity	0.002
Phospholipase activity	0.028
Phosphoinositide phospholipase C activity	0.031
Phospholipase C activity	0.054
Phosphoric diester hydrolase activity	0.084
Lipid catabolic process	0.205
« Response to oxydative stress »- score d'enrichissement : 1.37	
Termes	p-value
Response to reactive oxygen species	0.005
Response to hydrogen peroxide	0.019
Response to oxidative stress	0.039
Response to inorganic substance	0.117
Response to cAMP	0.325

Table 4.5 – Clusters fonctionnels pour les gènes comportant au moins un SNP significatif après correction FDR par gène pour au moins le test SDT, montrant une TD sexe-spécifique, classés selon leur score d'enrichissement.

« **Chromatine assembly and organization** »- score d'enrichissement : 3.16

Termes	p-value
Protein-DNA complex	1.27E-06
Nucleosome	7.72E-06
Nucleosome assembly	1.07E-05
Chromatin assembly	1.43E-05
Protein-DNA complex assembly	2.06E-05
Nucleosome organization	2.46E-05
DNA packaging	1.50E-04
Chromatin assembly or disassembly	2.78E-04
Cellular macromolecular complex assembly	9.29E-04
Cellular macromolecular complex subunit organization	0.003
Chromatin	0.005
Chromosomal part	0.011
Macromolecular complex subunit organization	0.029
Macromolecular complex assembly	0.030
Chromosome	0.038
Chromatin organization	0.053
Chromosome organization	0.189

« **mRNA binding and regulation** »- score d'enrichissement : 1.43

Termes	p-value
mRNA binding	0.005
Polysome	0.034
Regulation of translation	0.08
Posttranscriptional regulation of gene expression	0.143

« **Response to oxydative stress** »- score d'enrichissement : 1.41

Termes	p-value
Cytoplasmic membrane-bounded vesicle	0.018
Membrane-bounded vesicle	0.023
Vesicle	0.046
Cytoplasmic vesicle	0.059
Secretory granule	0.074

Table 4.6 – Clusters fonctionnels pour les gènes comportant au moins un SNP significatif après correction FDR par gène pour le test TD_{fils} , montrant une TD limitée à un sexe, classés selon leur score d'enrichissement.

« Membrane »- score d'enrichissement : 1.62	
Termes	p-value
Plasma membrane	0.011
Intrinsic to membrane	0.035
Integral to membrane	0.036
« Post-translational protein folding »- score d'enrichissement : 1.33	
Termes	p-value
Chaperone mediated protein folding requiring cofactor	0.018
De novo posttranslational protein folding	0.031
De novo protein folding	0.045
Protein folding	0.181

Table 4.7 – Clusters fonctionnels pour les gènes comportant au moins un SNP significatif après correction FDR par gène pour le test TD_{fille} , montrant une TD limitée à un sexe, classés selon leur score d'enrichissement.

Termes	p-value
Nucleosome assembly	0.006
Chromatin assembly	0.006
Protein-DNA complex assembly	0.007
Nucleosome organization	0.007
DNA packaging	0.009
Chromatin assembly or disassembly	0.009
Nucleosome	0.010
Protein-DNA complex	0.013
Cellular macromolecular complex assembly	0.024
Cellular macromolecular complex subunit organization	0.026
Chromatin organization	0.028
Chromatin	0.031
Chromosome organization	0.036
Macromolecular complex assembly	0.049

Table 4.8 – Termes enrichis pour les gènes comportant au moins un SNP significatif après correction FDR par gène pour les tests TD_{fille} et TD_{fils} , montrant une TD dans le même sens chez les deux sexes, classés selon leur p-value (3 gènes).

« **ATP binding** »- score d'enrichissement : 1.56

Termes	p-value
ATP binding	0.016
Adenyl ribonucleotide binding	0.017
Adenyl nucleotide binding	0.021
Purine nucleoside binding	0.023
Nucleoside binding	0.023
Nucleotide binding	0.034
Ribonucleotide binding	0.042
Purine ribonucleotide binding	0.042
Purine nucleotide binding	0.050

Table 4.9 – Cluster fonctionnel pour les gènes en commun entre la liste de SD genes chez les enfants et celle chez les parents (39 gènes).