UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
http://serval.unil.ch

Year : 2021

Comparative genomics of transitions between combined and separate sexes in the plant genus Mercurialis

Gerchen Jörn Frederik

Gerchen Jörn Frederik, 2021, Comparative genomics of transitions between combined and separate sexes in the plant genus Mercurialis

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_F14E6EFF159A0

Droits d'auteur
L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright
The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département d'Ecologie et d'Evolution

Comparative genomics of transitions between combined and separate sexes in the plant genus *Mercurialis*

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Jörn Frederik GERCHEN

Master de Organismic Biology and Evolution de la Humboldt Universität zu Berlin

Jury

Prof. Stephan Gruber, Président
Prof. John R. Pannell, Directeur de thèse
Prof. Barbara Mable, Experte
Prof. Nicolas Salamin, Expert

Lausanne
2021

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département d'Ecologie et d'Evolution

Comparative genomics of transitions between combined and separate sexes in the plant genus *Mercurialis*

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Jörn Frederik GERCHEN

Master de Organismic Biology and Evolution de la Humboldt Universität zu Berlin

Jury

Prof. Stephan Gruber, Président
Prof. John R. Pannell, Directeur de thèse
Prof. Barbara Mable, Experte
Prof. Nicolas Salamin, Expert

Lausanne
2021

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Stephan	**Gruber**
Directeur·trice de thèse	Monsieur	Prof.	John	**Pannell**
Expert·e·s	Madame	Prof.	Barbara	**Mable**
	Monsieur	Prof.	Nicolas	**Salamin**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Jörn Frederik Gerchen

Master of Science, Humboldt-Universität zu Berlin, Allemagne

intitulée

Comparative genomics of transitions between combined and separate sexes in the plant genus *Mercurialis*

Date de l'examen : 16 août 2021

Date d'émission de l'imprimatur : Lausanne, le 27 août 2021

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Niko GELDNER
Directeur de l'Ecole Doctorale

Acknowledgements

First, I would like to thank my supervisor John Pannell. John has been an amazing supervisor, who always gave me the space to follow my own ideas in my own time while always being available to discuss ideas or help with issues. I have always been impressed and inspired by both his broad interest, but also by his dedication to critical scientific reasoning. At the same time, I really appreciated his effort to create a welcoming and collaborative athmosphere in his lab. I cannot imagine many other supervisors with whom a PhD would have been such an interesting and pleasant experience.

In addition, I would like to thank all the members of the Pannell group, who have been incredibly friendly, open and helpful. In particular, I would like to thank Xinji Li, Guillaume Cossard, Paris Veltsos, Dan Jeffries and Mathias Scharmann, with whom I had very productive collaborations on various projects. In addition, I would like to thank Nicolas Ruch and Aline Revel, without whom I would have never been able to grow the thousands of plants that are the basis of this thesis. I would like to thank Yves Cuenot and Dessislava Savova Bianchi, without whom I surely would have gotten lost in the wetlab. I'd also like to thank my first step student Ludovic Rüedi, and Elvire Sarton-Lohèac, who did a project with me during an internship in our lab. Working with both of them has been a great experience, which tought me a lot about how to supervise and teach highly interested students. A big thanks also goes out to my student assistants Solenn Sarton-Lohèac, Anouck Brodard, Apiha Shanmuganathan and Alexandre Jann, who spend countless hours with phenotyping plants. The results from Chapter 4 of this thesis would not have been possible without their hard work. An special thanks goes out to Alexandre Jann and Mélissa Monn, for last-minute french translation support for the abstracts for this thesis.

I also would like to thank the members of the Department of Ecology and Evolution, for making it such a lively and interesting place. I want to thank many other scientists, who took time to discuss my projects with me, among them Jeffrey Jensen and Martin Kapun, who took a lot of time to discuss various naive questions during earlier stages of my PhD. Finally, I would like to thank the members of my thesis committee for taking the time to read and comment on this thesis and for giving valuable advice during my midthesis defense.

Abstract

While the majority of plant species have combined sexes (hermaphroditism or monoecy), evolutionary transitions to separate sexes (dioecy) have occured numerous times independently. In addition, there have been multiple reversions from dioecy towards combined sexes. Little is known about the underlying genetics of these convergent evolutionary transitions. In this thesis I explore genomic datasets to understand several evolutionary genomic processes that can be involved in the evolution of separate sexes and other transitions between plant sexual systems, focusing on the genus *Mercurialis*, in which dioecy is ancestral. Transitions from dioecy to monoecy and androdioecy (coocurrence of males and monoecious individuals) have occured in its annual clade and different lineages show variation in ploidy, and all monoecious and androdioecious lineages are polyploids. I analyze several datasets, which are related to the evolution of sex chromosomes in *Mercurialis*. As in animals, sex chromosomes determine sex in many dioecious plant species. However, it is unclear if sex chromosomes in plants commonly follow the same evolutionary trajectories, which involve the evolution of recombination suppression and sequence degeneration. Using a number of phylogenetic markers, I show that annual and several related perennial lineages of *Mercurialis* share an ancestral XY system of sex determination. In addition, I improve the understanding of phylogenetic relationships between lineages and find evidence for several previously unknown allopolyploidization events. Also I show that males in androdioecious populations of hexaploid *M. annua* lineages are the result of introgression of a Y chromosome from a distantly related perennial lineage. This pattern is at odds with the strong contribution of sex chromosomes to reproductive isolation, which has been observed in animals, but it may be explained by the limited degree of recombination suppression, similarities in ploidy between both lineages and with a strong outcrossing advantage of males in a monoecious background, which would favour the rapid spread of a Y chromosome in populations in which dioecy had previously been lost. Using further genomic datasets, I locate the Y chromosome in diploid, dioecious *M. annua* in a pericentromeric region with low recombination. In addition, I find evidence for two independent extensions of the non-recombining region in annual lineages of *Mercurialis*. These results show that plant sex chromosomes may evolve repressed recombination in a similar fashion as is known in animals. Finally, I explore the genetic architecture underlying transitions towards monoecy in experimental populations of diploid *M. annua*. In these populations, removal of males resulted in a dramatic increase in females producing male flowers. I identify two independent regions in which loci with strong antagonistic effects on male and female flower production are colocalized. This implies that the genetic architecture underlying transitions towards monoecy can be based on few major effect loci.

Résumé

Alors que la majorité des espèces végétales ont des sexes combinés (hermaphrodisme ou monoécie), des transitions évolutives vers des sexes séparés (dioécie) se sont produites de nombreuses fois indépendamment. De plus, il y a eu de multiples réversions de la dioécie vers les sexes combinés. On sait peu de choses sur la génétique sous-jacente de ces transitions évolutives convergentes. En me concentrant sur le genre *Mercurialis* chez qui la dioécie est ancestrale, j'explore, dans cette thèse, des ensembles de données génomiques pour comprendre plusieurs processus évolutifs qui peuvent être impliqués dans l'évolution des sexes séparés ainsi que dans d'autres transitions entre les systèmes sexuels des plantes. Des transitions de la dioécie vers la monoécie et l'androdioécie (coexistance de mâles et d'individus monoïques) ont eu lieu dans son clade annuel et différentes lignées montrent des variations de ploïdie. Toutes les lignées monoïques et androdioïques sont polyploïdes. J'analyse plusieurs ensembles de données qui sont liés à l'évolution des chromosomes sexuels chez *Mercurialis*. Comme chez les animaux, les chromosomes sexuels déterminent le sexe chez de nombreuses espèces végétales dioïques. Cependant, on ne sait pas avec certitude si les chromosomes sexuels des plantes suivent généralement les mêmes trajectoires évolutives qui impliquent l'évolution de la suppression de recombinaison et la dégénérescence des séquences. En utilisant un certain nombre de marqueurs phylogénétiques, je montre que les lignées annuelles ainsi que plusieurs lignées pérennes apparentées au genre *Mercurialis*, partagent un système ancestral XY de détermination du sexe. En outre, j'améliore la compréhension des relations phylogénétiques entre les lignées et je trouve des preuves de plusieurs événements d'allopolyploïdie inconnus auparavant. Je montre également que les mâles des populations androdioïques des lignées hexaploïdes de M. annua sont le résultat de l'introgression d'un chromosome Y provenant d'une lignée pérenne lointainement apparentée. Ce schéma est en contradiction avec la forte contribution des chromosomes sexuels à l'isolement reproductif qui a été observée chez les animaux. Ce dernier peut être expliqué par le degré limité de suppression de recombinaison, les similitudes de ploïdie entre les deux lignées et un fort avantage d'outcrossing des mâles dans un contexte monoïque qui favoriserait la propagation rapide d'un chromosome Y dans les populations ayant précédemment perdu la dioécie. En utilisant d'autres ensembles de données génomiques, je localise le chromosome Y chez M. annua diploïde et dioïque dans une région péricentromérique à faible recombinaison. En outre, je trouve des preuves de deux extensions indépendantes de la région de non-recombinaison dans les lignées annuelles de *Mercurialis*. Ces résultats suggèrent que les chromosomes sexuels des plantes peuvent évoluer vers une recombinaison réprimée de manière similaire à ce qui est connu chez les animaux. Finalement, j'explore l'architecture génétique qui sous-tend les transitions vers la monoécie dans des populations expérimentales de M. annua diploïdes. Dans ces populations, l'élimination des mâles a entraîné une

augmentation spectaculaire du nombre de femelles produisant des fleurs mâles. J'identifie deux régions indépendantes dans lesquelles les loci ayant de forts effets antagonistes sur la production de fleurs mâles et femelles sont colocalisés. Cela implique que l'architecture génétique qui est responsable des transitions vers la monoécie peut être basée sur quelques loci à effet majeur.

Contents

1 Introduction **4**

1.1 The evolution of plant reproductive diversity 4

1.2 Evolutionary theory of sexual system evolution 5

1.3 Genomics of sex determination 9

1.3.1 The evolution of sex chromosomes 10

1.3.2 Genetics of polygenic and quantitative sex determination systems 17

1.4 The effects of introgression, hybridization and polyploidy on sexual system evolution . . . 18

1.4.1 Introgression and hybridization 18

1.4.2 Polyploidy 20

1.5 Transitions between sexual systems in *Mercurialis* 22

1.6 Aims and outline of the thesis 26

2 Phylogenetic analysis using sex-linked, sex chromosomal and autosomal phylogenetic markers **45**

2.1 Introduction 45

2.2 Material and Methods 51

2.2.1 Development of a sex-specific PCR marker 51

2.2.2 Development of phylogenetic markers located on autosomes and sex chromosomes 53

2.3 Results . . . 56

2.3.1 Sex-specific PCR marker . . . 56

2.3.2 Phylogenetic markers on autosomes and sex chromosomes . . . 56

2.4 Discussion . . . 59

3 Sex chromosome evolution in the genus *Mercurialis* 72

3.1 Introduction . . . 72

3.2 Material and Methods . . . 77

3.2.1 Genome assembly . . . 77

3.2.2 Linkage mapping . . . 78

3.2.3 Scaffolding of genome assemblies . . . 80

3.2.4 Identification of sex-linked genomic regions . . . 81

3.2.5 Identification of Y- and X-hemizygous regions . . . 82

3.2.6 Identification of sex-linked genetic variation . . . 83

3.2.7 Genetic correlations on linkage group 1 in *M. annua* and *M. huetii* . . . 83

3.3 Results . . . 84

3.3.1 Genome assembly . . . 84

3.3.2 Linkage mapping . . . 84

3.3.3 Identification of Y- and X-hemizygous regions . . . 85

3.3.4 Identification of sex-linked genetic differentiation . . . 88

3.3.5 Genetic correlations on linkage group 1 in *M. annua* and *M. huetii* . . . 89

3.4 Discussion . . . 92

4 QTL analyis of the genetic basis of enhanced leakiness in experimental populations of *M. annua* **107**

4.1 Introduction 107

4.2 Material and Methods 113

4.2.1 Crosses 113

4.2.2 Genotyping 115

4.2.3 QTL analysis 116

4.3 Results 117

4.4 Discussion 120

5 Conclusions **130**

5.0.1 Perspectives for further research 136

6 Supplementary Material **142**

Chapter 1

Introduction

1.1 The evolution of plant reproductive diversity

Reproduction is one of the essential hallmarks of life, and all living organisms are the result of either asexual or sexual reproduction. Sexual reproduction involves meiosis, which allows recombination of genetic variation and fusion of gametes, which can be of similar size (isogamy) or which differ in size and motility (anisogamy), with male gametes being smaller and motile and female gametes being larger and stationary. Generally, recombination via sexual reproduction can both bring together alleles at different loci that arose in different lineages, and it can break down associations between alleles at different loci. This has fundamental consequences for evolution on the organismic and the genomic level, as it facilitates both adaptive evolution and purifying selection against deleterious variants and allele combinations (Charlesworth 1993; Betancourt et al. 2009). In addition, recombination has a range of secondary effects, affecting gene conversion (Marais 2003), the evolution of repetitive elements (Kent et al. 2017) and the genetic diversity at a large proportion of the genome through linked selection (Slotte 2014). However, compared with asexual reproduction, sexual reproduction can have a strong transmission disadvantage as male gametes do not produce offspring (Smith 1978). In addition, recombination can break down coadapted allele complexes, and sexual reproduction may come with additional fitness costs through sexual selection and sexually transmitted diseases and parasites (Otto and Lenormand 2002).

The question of which of these factors are most important at maintaining sexual reproduction has been the subject of controversial discussion for a long time (Felsenstein 1974; Barton and Charlesworth 1998; Otto 2009). Nonetheless, sexual reproduction is extremely widespread among eukaryotes (Beukeboom and

Perrin 2014). Meiosis and fusion of gametes, which underly sexual reproduction at the cellular level, likely come from a single common ancestor (Dacks and Roger 1999) and its underlying molecular machinery is remarkably conserved across sexually reproducing eukaryotes (Villeneuve and Hillers 2001). In contrast, there is a striking diversity in the way sex is expressed at the organismic level (Bachtrog et al. 2014).

This diversity is particularly great in plants, where extensive variation of sex expression is found in the distribution of reproductive organs of each sex within and between individuals. Individuals can be either male, female or express both sexes simultaneously, either in the same (hermaphroditism) or separate flowers (monoecy). Individuals that express both sexes can vary in their relative expression of each sex, and both combinations of male and hermaphrodite flowers (andromonoecy) and female and hermaphrodite flowers (gynomonoecy) can be found on the same individual. There is also variation of the sexual system within natural populations. Besides hermaphrodites, combinations of either males and females (dioecy), females and hermaphrodites (gynodioecy) or, rarely, males and hermaphrodites (androdioecy) exist (Barrett 2002).

The distribution of these sexual systems and their mode of sex determination varies among different clades of eukaryotes. While most vertebrates and a majority of invertebrates have separate sexes, the majority of angiosperm species are hermaphrodites (The Tree of Sex Consortium 2014). This difference between plants and animals can be traced back to differences in lifeform (Eppley and Jesson 2008) and evolutionary history. Hermaphroditism has likely been the ancestral state of angiosperms (Sauquet et al. 2017), while basal animals have separate sexes and the occurrence of separate sexes in plants has been the result of evolutionary transitions. Despite the dominance of hermaphroditism in angiosperms, separate sexes are found in about half of all plant families, and it has been estimated that separate sexes evolved up to 5000 times independently (Renner 2014). This high degree of evolutionary convergence makes the study of these transitions a major question in evolutionary biology, and understanding which evolutionary forces underly them is of general interest, as it touches on a wide range of questions and concepts.

1.2 Evolutionary theory of sexual system evolution

Evolutionary theory makes clear predictions about the ratio of male and female gametes in a population based on the application of game theory to evolutionary genetics (Smith 1982; West 2009). As each female gamete has to be fertilized by a male gamete, there is strong frequency-based selection resulting in an equal investment in production and dispersal of male and female gametes. Genetic variants that bias the investment into gametes of one sex will be selected against by variants of the opposite effect, because the fitness of gametes of the opposite sex will increase as they become rarer (Fisher 1930). The

investment into male and female gametes is thus expected to converge to an evolutionary stable strategy where both sexes have equal fitness. This logic applies both to dioecious sexual systems, where it defines the investment in sons or daughters, and to cosexual individuals, where it defines the reproductive effort invested into male and female function by each individual.

Individuals have a limited amount of resources available for reproduction and, as a result, individuals that express both sexual functions will allocate a subset of available resources to each of them, while unisexual individuals will invest all resources into only one. The evolution and maintenance of different sexual systems can be understood in this context: hermaphroditism will be an evolutionary stable strategy if individuals maximize their fitness by investing into both sexual functions. In contrast, if fitness is maximized by allocating resources into just one sexual function, dioecy will be stable. Conditions for the maintenance of gynodioecy or androdioecy would be based on intermediate scenarios, where fitness is maximized if resources are invested in one of the sexes and hermaphrodites (Charnov et al. 1976). These fitness optima are determined by how fitness returns change for each sex function with increased allocation into them. Sex allocation theory uses fitness gain curves to describe these fitness functions and it depends on the interplay of fitness gain curves for both sexual functions if a specific sexual system will be favored. Based on the combined effect of both fitness gain curves (the fitness set), the relative fitness contributions for male and female functions determine if a specific sexual system is stable, or if it will be susceptible to invasion by mutants that cause transitions (Charnov et al. 1976).

There is a complex interplay of different factors that shape fitness gain curves, and distinguishing them experimentally can be challenging (Campbell 2000), but in general the shape of fitness gain curves is determined by sexual specialization and selection for outcrossing or selfing (Charlesworth and Charlesworth 1981).

Sexual specialization is often related to sex-specific resource requirements. These can be either direct, as production of male and female gametes has different resource requirements (e.g., pollen and seeds tend to differ in their carbon-nitrogen ratio (Obeso 2002)) or indirect (there are often additional morphological adaptations that favor reproduction of one sex (Lloyd and Webb 1977)). If resource requirements show little overlap, the maintenance of hermaphroditism can be favored (Lloyd 1982), but not all investments into one sex function will directly translate into the production of gametes, as there may be sex-specific fixed resource requirements that have to be invested before gametes can be produced, and unisexual individuals can be favored when they only have to invest fixed costs for one sex function (Charnov 1979).

In animals indirect sex-specific resource requirements can be substantial, as they are often related to sexual dimorphism that may result from sexual selection. This can be understood in terms of Bateman's

Principle, which states that reproductive fitness is limited by resources for females, but by access to mates for males (Bateman 1948). This may be less apparent in hermaphroditic animal-pollinated plants, where pollinators are required for successful reproduction through both male and female functions. However, an increased investment in attractive structures should largely result in accelerating fitness gains for the male function if pollinators are abundant, because female fitness will not increase after all seeds are pollinated (Charnov 1979). In wind-pollinated plants, very different structures facilitate male and female reproduction. Male reproductive organs often are exposed and extruded to help spread pollen in the wind, while female structures are optimized to capture pollen, which often involves morphological adaptations that differ from male flowers (Culley et al. 2002).

In addition, the shape of fitness gain curves can be influenced by interactions between offspring. Sex-specific gain curves will show diminishing returns if there is local competition among offspring of one sex for resources or mates. In plants, competition can either happen at the seedling stage, where increased pollen dispersal could cause siblings related through their male function to experience lower competition among siblings (Lloyd 1982), or at the seed stage, when multiple seeds grow up in the same fruit, which can have different fathers, but will always have the same mother (Bawa 2016). In principle, fitness gain curves could also accelerate by differential cooperation among relatives of one sex. This effect has been shown in animals (West 2009), but conclusive evidence from plants is lacking.

The other driver in sexual-system evolution is selection for outcrossing or selfing. It has been known for a long time that offspring produced by relatives often have lower fitness than those of unrelated individuals, and these differences in fitness are even stronger in offspring sired by selfing of cosexual plants, which will be inbred to an even stronger degree than the offspring of immediate relatives (Darwin 1876). These negative fitness effects can be explained by the decrease of heterozygosity because the probability of inheriting two alleles identical by state will increase with the degree of inbreeding. There are two ways in which the decrease in heterozygosity can affect offspring fitness. The overdominance hypothesis explains decreases in fitness by the loss of variation at loci where it is advantageous to have different alleles (heterosis). The partial dominance hypothesis explains fitness effects by recessive deleterious mutations being expressed in a homozygous state (Charlesworth and Charlesworth 1987). Overall, empirical data shows that it's mostly recessive deleterious mutations that contribute to inbreeding depression, which is largely in favor of the partial dominance hypothesis (Charlesworth and Willis 2009). Besides these direct fitness effects, selfing can further hinder adaptive evolution by reducing effective population sizes and genetic diversity (Charlesworth and Wright 2001) and result in more extended linkage disequilibrium (Nordborg 2000).

Besides these genetic reasons, selfing can also affect reproduction on the phenotypic level: transfer of

self pollen between anther and style of the same flower (autogamy) or between different flowers of the same plant (geitonogamy) can have additional fitness costs, as it can limit the amount of pollen available for outcrossing, and it can reduce the rate of germination and seed set from outcrossed pollen by blocking styles even if the self pollen does not germinate (Jong et al. 1993). There is a number of morphological and genetic adaptations that reduce selfing and self-pollination. Anthers and stigma can be separated spatially, either within the same flower (herkogamy) or in different types of flowers on the same (monoecy) or on different individuals (Barrett 2002), or the presentation of pollen and receptivity of stigmas on the same plant can be separated in time (Lloyd and Webb 1986). Genetic systems of self-incompatibility have evolved multiple times independently and can involve different genetic mechanism of recognizing and rejecting inbred gametes (Charlesworth, Vekemans, et al. 2005), which can be associated with reciprocal differences in style and anther morphology (heterostyly) (Barrett 2019). Of all of these adaptations, only complete dioecy can prevent all forms of these interactions between pollen and pistil on the same individual.

However not all effects of selfing are detrimental, and while only 10-15% of plant species are predominantly selfing, a much larger proportion of plant species show mixed mating systems, where some proportion of gametes is sired through selfing (Goodwillie et al. 2005). There are two selective hypotheses that explain transitions to and the maintenance of selfing (Busch and Delph 2012): the transmission advantage hypothesis is based on the fact that variants that favor selfing have a strong transmission advantage, because they will be propagated through both sexual functions of the same individual (Fisher 1941). The reproductive assurance hypothesis states that selfing can be advantageous when pollinators or mates are absent or rare, which is often the case in colonizing species or in species where populations are subject to frequent extinction and recolonization. This higher colonization ability is also a likely explanation for Baker's law (Baker 1955), which accounts for the higher frequency of self-compatible species found on islands than on the mainland (Grossenbacher et al. 2017). In addition, the deleterious effects of inbreeding through recessive mutations, due to which selfing would have severe fitness effects in outcrossing plants, tend to be purged from predominantly selfing populations (Lande and Schemske 1985). Selfing also tends to result in stronger genetic differentiation between populations, by which it could favor speciation (Wright et al. 2013).

Most of the theory underlying sex determination in plants is well understood, but still very little is known about how transitions in sexual systems happen in detail. It is still unclear whether inbreeding avoidance or sexual specialization are more important for transitions towards, and maintenance of, separate sexes, and it is also not clear how common transitions from separate towards combined sexes are, and which the selective factors causing them are.

One kind of evidence that can be used to infer the evolutionary pathways underlying transitions is the distribution of sexual systems across the tree of life (Weller and Sakai 1999). These types of studies show broad scale patterns of the frequency and directionality of sexual system transitions. A general result is the extreme rarity of androdioecy, which is in accordance with evolutionary models that predict that it would require a very strong male outcrossing advantage to evolve (Pannell 2002). Another well-known result is the scattered distribution of dioecy and the small size of dioecious clades (Heilbuth 2000), which resulted in dioecy often being considered an evolutionary dead-end. However, analyses using newly-developed tests for sister clade comparisons showed that dioecy is actually associated with slightly elevated diversification rates (Käfer, Boer, et al. 2014), and the rarity of dioecy may in fact be related to frequent reversions towards monoecy (Käfer, Marais, et al. 2017). There is a strong association of both dioecy and monoecy at the family level (Renner and Ricklefs 1995), and dioecy and gynodioecy at the genus level (Dufay et al. 2014), which has been interpreted as evidence of dioecy evolving frequently both from either monoecy or gynodioecy. However, an alternative explanation is that the association of monoecy and dioecy could have resulted from a general clustering of labile sexual systems (Weiblen et al. 2000). In addition, associations between traits and sexual systems can be indicative of the underlying selective pressures that cause sexual-system transitions. Traits that are associated with dioecy are abiotic pollination, climbing growth form (Renner and Ricklefs 1995) and animal dispersal through fleshy fruits (Vamosi and Vamosi 2004). These observations are compatible with transitions towards dioecy caused by sexual specialization.

Both evolutionary theory and phylogenetic associations can provide insights into the selective pressures causing changes in sexual systems. For a truly integrated understanding of these evolutionary processes, it is necessary to understand how the variance, on which such selective processes act, is generated by molecular changes underlying the genetic architecture that determines sex expression.

1.3 Genomics of sex determination

Individual sex can be determined by environmental cues or individuals can change sex during the course of their lives (Vega-Frutis et al. 2014), but in most cases differences in sex expression have a strong heritable genetic component (Bachtrog et al. 2014). When individuals in a population can express two clearly distinct sexes, determined genetically, they are usually based on a single locus, while polygenic systems of sex determination are rare for bisexual organisms (Moore and Roberts 2013). Much less is known in cases where there is continuous, genetically determined variation in sex expression; in this case it is possible that multiple loci across the genome can have a quantitative effect on sex expression.

A common theme in animals is that orthologs of the same genes become involved in sex determination independently. In vertebrates, a set of genes located at different positions in the genetic network involved in sex determination acquired the role of master sex determination gene in diverged lineages (Capel 2017). In holometabolous insects sex determination shares a conserved set of core genes, which can be activated by a number of different mechanisms and can have a variety of downstream effects (Bopp et al. 2014).

1.3.1 The evolution of sex chromosomes

A common feature of single locus sex determination systems is the evolution of sex chromosomes. Sex chromosomes were first discovered over a hundred years ago (Stevens 1906; Brush 1978) and have been described in clades as diverse as animals (Kaiser and Bachtrog 2010; Miura 2017; Zhou et al. 2014; Graves 2006), plants (Ming et al. 2011), algae (Coelho et al. 2019) and fungi (Branco et al. 2017), and can be found in organisms where the sexual phase is diploid as well as in organisms where the sexual phase is haploid. Sex chromosomes are defined as the pair of chromosomes on which the sex-determining locus is located, and in diploid sex-determination systems there is usually a heterogametic sex, where either maleness or femaleness are determined by either a dominant sex determining locus (XY or ZW system) or based on the copy number of the X chromosome (X0 system), while in haploid sex determination systems sex is determined by two alleles at the same locus (UV system). In this thesis I will adopt the convention to always refer to X and Y chromosomes when I discuss general patterns of sex chromosome evolution, which would also apply for Z and W chromosomes. I will refer to Z and W chromosomes when I describe specific sex chromosome systems or when I want to emphasize general differences between the evolution of male and female heterogametic sex chromosome systems.

Sex chromosomes evolve from autosomes through two different mechanisms. In bisexual organisms, which already have separate sexes, an autosomal locus can take over sex determination, a process called sex chromosome turnover (Vicoso 2019). In plants, sex chromosomes often evolve from autosomes when there is a transition from combined towards separate sexes (Charlesworth 2002). The rates of sex-chromosome turnover can differ strongly between clades. While some old clades like mammals or birds have conserved sex chromosomes, turnover is rapid in many groups of fish and amphibians. There is evidence in multiple systems that not all autosomes have the same likelihood of becoming a sex chromosome, but that particular autosomes repeatedly became sex chromosomes, likely as a result of the same locus repeatedly taking over the top of the sex determination cascade (Jeffries, Lavanchy, et al. 2018).

Many sex chromosomes share features that are not required for sex determination. One of these is

the evolution of recombination suppression. Here, recombination between both sex chromosomes in the heterogametic sex is often suppressed around the sex determining locus, while both sex chromosomes can freely recombine in the homogametic sex. The loss of recombination usually results in the accumulation of sequence differences between the sex chromosomes. These can be caused by a variety of genetic processes such as the accumulation of single nucleotide polymorphisms (SNPs), tandem repeats and transposable elements, and often also on the chromosomal scale due to structural variants (Bachtrog 2013). As a result, the gene content in the region of suppressed recombination diverges between both copies of the sex chromosomes, either through loss of gene sequences or through the accumulation of loss-of function mutations. Importantly, recombination suppression often happens in a stepwise fashion, which can result in so-called evolutionary strata, discrete regions of the non-recombining part of the sex chromosome with different degrees of sequence divergence and gene losses (Lahn and Page 1999).

The degree of recombination suppression and degeneration can vary strongly between lineages and can range from homomorphic (microscopically indistinguishable) sex chromosomes, with a single SNP that has no discernable recombination suppression around it determining sex (e.g., in tiger pufferfish (Kamiya et al. 2012)), through strongly heteromorphic (microscopically distinguishable) sex chromosomes (e.g., in mammals), to complete loss of one of the sex chromosomes resulting in an X0 system of sex determination (e.g., in *Caenorhabditis elegans* (Madl and Herman 1979)). While these differences in recombination suppression and sex-chromosome degeneration are often related to the relative age of the sex chromosomes, there are examples of very old sex-chromosome systems that show little recombination suppression, e.g., in rhatite birds (Yazdi and Ellegren 2014) and sturgeons (Kuhl et al. 2020).

These processes of sex-chromosome degeneration can have strong fitness consequences for the heterogametic sex. Large-scale patterns of reduced adult sex ratio (Pipoly et al. 2015) in tetrapods and lifespan (Xirocostas et al. 2020) as well as differences in sex ratio in dioecious plants (Field et al. 2013) in the heterogametic sex may be related to these deleterious effects. These fitness effects can be the consequence of a number of processes related to sex-chromosome degeneration. The loss of functional gene copies on one of the sex chromosomes can result in the reduction of expression levels of dosage sensitive genes. There has been convergent evolution of dosage compensation mechanisms in different clades through a variety of mechanisms (Gu and Walters 2017), pointing to the importance of this effect. Furthermore sex chromosomes are ideal targets for meiotic drive loci, which bias the sex ratio to favor their own transmission (Jaenike 2001). Often the evolution of meiotic drivers selects for other loci that compensate the sex ratio bias. In the absence of such compensation, the sex ratio biases can have severe fitness consequences on the population level (including extinction). There is evidence for rampant gene duplications of loci likely involved in these effects on the sex chromosomes of different *Drosophila* species (Ellison and Bachtrog

2019). In addition, increased expression of repeats and accumulation of heterochromatin on the sex chromosome can have fitness consequences for the rest of the genome (Nguyen and Bachtrog 2020).

These deleterious effects raise the question of why recombination suppression evolves in the first place. The most common hypothesis for recombination suppression on sex chromosomes assumes some initial selective advantage, which is caused by loci that are under sexually antagonistic selection (Charlesworth and Charlesworth 1980; Rice 1987). When expressed, these loci would have positive fitness effects in one of the sexes, but negative fitness effects in the other. Linkage between a sexually antagonistic locus and the sex-determining locus could be one way to resolve the resulting sexual conflict over the expression of these loci by ensuring that sexually antagonistic alleles co-segregate with the heterogametic sex. This idea is conceptually related to so-called 'supergenes', clusters of tightly linked loci that underly a variety of adaptive phenomena in a wide range of species (Schwander et al. 2014).

While the theoretical predictions of sexual antagonistic selection seem clear, there has been limited evidence that sexually antagonistic loci contribute to recombination suppression on sex chromosomes. One promising system to find evidence for this effect has been the guppy fish, where male coloration loci increase male fitness through sexual selection but decrease female fitness by making them more conspicuous to predators; the alleles are inherited in a male-specific manner (Gordon et al. 2012). However, recent research shows that recombination rates in males are limited on all chromosomes, which would allow for male-specific inheritance of sexually antagonistic alleles without sex-chromosome evolution (Bergero, Gardner, et al. 2019). In addition, there appears to be only a small region of completely suppressed recombination on the male Y chromosome, which precludes any effect of sexually antagonistic selection being involved in the evolution of recombination suppression (Kirkpatrick et al. 2020). Another line of evidence comes from naturally occurring variation in the degree of recombination suppression on sex chromosomes in the common frog *Rana temporaria*. Unlike the prediction for sexually anatagonistic selection, males with different sex chromosomes show no difference in fitness (Veltsos, Rodrigues, et al. 2019).

In light of the limited support for sexual antagonism as a cause of recombination suppression on sex chromosomes, a number of alternative explanations have been proposed (Ponnikas et al. 2018). The meiotic drive hypothesis proposes that a new sex determiner could arise if it is initially linked to a meiotic drive locus. The initial region of suppressed recombination between driver and sex determining loci would persist after an unlinked suppressor locus, which restores sex ratios, evolves (Ubeda et al. 2015). The heterozygote advantage theory proposes that heterosis could select for recombination suppression by linking the sex determination locus to loci that are advantageous when heterozygous. As the initial degree of heterozygote advantage has to be higher than is usually expected through evolution by single mutational

steps, this model is thought to apply mostly to the fusion between sex chromosomes and autosomes (Charlesworth and Wall 1999). A third hypothesis proposes that genetic drift can cause, or at least facilitate, the evolution of recombination suppression. This is because Y or W chromosomes have a lower effective population size than autosomes and the X or Z chromosomes, and, at least in Y chromosomes, there may be additional reductions due to variance in male reproductive success. This could increase the probability of fixation of either inversions, or variants that increase sequence divergence between sex chromosomes around the sex determining locus, both of which could decrease local recombination rates (Jeffries, Gerchen, et al. 2021). In addition, repeats and transposable elements could play an significant role in the evolution of recombination suppression, because they can increase the probability of inversions and can induce sequence divergence between sex chromosomes (Kejnovsky et al. 2009). Most sex chromosomes harbor a significant proportion of transposable elements (TEs) and repeats, and there appears to be a link between TE accumulation and the emergence of young sex chromosomes in fish, though it is still unclear if this is a cause or a consequence of recombination suppression (Chalopin et al. 2015).

In plants, models for the evolution of separate sexes from hermaphroditism add an additional role for recombination suppression during sex-chromosome evolution. The canonical gynodioecy model proposes that the evolution of dioecy from hermaphroditism requires three mutational steps. The first step in this model is the evolution of gynodioecy through a recessive loss-of-male-function mutation, which results in females when homozygous. In a second mutation step, a dominant suppressor of female function evolves at a different locus, which converts hermaphrodites into males. Importantly, in a third step, selection against neuters results in tight linkage between both loci, which results in a young sex chromosome with a region of suppressed recombination (Charlesworth and Charlesworth 1978a). This model specifically requires gynodioecy as an intermediate step. If androdioecy would be the intermediate step, the relative outcrossing advantage of males compared to hermaphrodites (either through increased pollen production or pollination efficiency, or through reduction of inbreeding depression) would have to be implausibly high (Charlesworth 1984).

An alternative pathway for the evolution of separate sexes is the monoecy or paradioecy model (Lloyd 1980). Here, the separation of male and female reproductive organs into different flowers in the same individual precedes the evolution of separate sexes, and, under some conditions, bisexuality may evolve through a series of small effect mutations that reduce male and female fertility. This model is less explicit about the number and types of mutations involved, but it proposes that linkage between loci involved in male and female fertility could facilitate the evolution of dioecy (Charlesworth and Charlesworth 1978b).

Thanks to recent advances in next-generation sequencing (NGS) technology, an increasing number of

sex-determination genes is being identified, allowing to test how often dioecy evolved via the gynodioecy or monoecy pathway. Consistent with the gynodioecy pathway, sex is determined by two Y-linked genes, that cause female sterility and male function in kiwifruit (Akagi, Pilkington, et al. 2019) and date palms (Torres et al. 2018). In the case of the Y chromosome of asparagus, two putative sex determination genes have been identified (Harkess, Zhou, et al. 2017; Harkess, Huang, et al. 2020), but the relative timing of sterility mutations suggests that it did not evolve via gynodioecy, but rather via androdioecy, which does not conform with theoretical expectations (Pannell and Gerchen 2018). Other species show some evidence for a two-locus model, but identification or functional validation of sex-determination genes are still lacking. There is a limited number of candidate genes for both recessive male sterility mutations and dominant suppressor of female function in wild grape (Badouin et al. 2020). In *Silene latifolia*, deletion mapping identified both a gene with stamen promoting and a gene with gynoecioum suppressing function located on different parts of the Y chromosome (Kazama et al. 2016). In the wild strawberry *Fragara virginia*, genetic mapping suggests the presence of two loci involved in sex determination. Importantly, in this system hermaphrodites and neuters are produced at low frequencies as a result of incomplete recombination suppression between both loci, consistent with expectations for the earliest stages of sex-chromosome evolution (Spigler et al. 2008).

In contrast, several examples are known, in which plant sex is determined by a single gene (Renner 2016). In persimmon, the male sex-determining gene encodes a small RNA, named *OGI*, which represses an autosomal gene, *MeGI*, which downregulates anther fertility (Akagi, Henry, et al. 2014). In monoecious cucumber and melon species, a pathway for the evolution of a single locus dioecious sex determination system has been shown experimentally (Boualem et al. 2015). Here, dioecy could be artificially induced by a two-step mutational pathway, which is based on epistatic interactions between the gene *acs11*. This gene is involved in ethylene biosynthesis, which inhibits the expression of *wip11*, a gene that in turn inhibits carpel development. Dioecy could evolve when a loss-of-function mutation becomes fixed at *acs11*, and a similar loss-of-function mutation keeps segregating at *wip11*, where it determines maleness when heterozygous and femaleness when homozygous. More examples of single-locus sex determination systems are found in Salicacea, where orthologs and partial copies of the type-A response regulator gene have evolved sex-determination functions independently resulting in multiple different systems of sex determination (Yang et al. 2020; Müller et al. 2020).

So far it is unclear how these different sex-chromosome systems relate to the evolution of recombination suppression. Most plant sex chromosomes are homomorphic, which is likely due to their relatively young evolutionary age (Ming et al. 2011), but there is increasing evidence that recombination suppression does happen at the sequence level in similar ways as in animal sex chromosomes. In most plant systems in

which sex chromosomes have been characterized at a molecular level, at least some region of repressed recombination has been found around the sex determination gene(s), both in systems with single-gene sex-determination (Akagi, Henry, et al. 2014; Yang et al. 2020) and around the two-gene sex-determination system of kiwifruit (Pilkington et al. 2019; Akagi, Pilkington, et al. 2019), where it extends beyond the range required to link both genes. In this context, it is also worth noting that preexisting regions of low recombination may facilitate sex-chromosome evolution in plants, as both the sex chromosomes of kiwifruit and papaya are found in low-recombination regions close to centromeres (Yu et al. 2007; Pilkington et al. 2019). In *Rumex hastatulus*, the sex-linked regions on both XY and XYY chromosomes are located in regions of low recombination, which make up large parts of the genome (Rifkin et al. 2021). There is also evidence for the presence of evolutionary strata on the Y chromosomes of both *C. papaya* (Wang et al. 2012) and *Spinacia oleracea* (She et al. 2020). In both of these cases, chromosomal inversions are found in the region of suppressed recombination, and variation in the synonymous divergence of X- and Y-linked genes in *Silene latifolia* across the length of the X chromosome may be indicative of evolutionary strata as well (Bergero, Forrest, et al. 2007). Furthermore, there is evidence for the accumulation of repetitive elements (Kejnovsky et al. 2009) and sequence degeneration on the sex chromosomes of almost all of these systems, and pseudogenes are found in multiple species as well (Bergero and Charlesworth 2011; Wang et al. 2012).

Similar to animals, it is unclear to what degree regions of suppressed recombination on plant sex chromosomes affect plant fitness, and how selection affects the evolution of recombination suppression. Unlike in animals, where sperm tend to express only a limited number of genes (Joseph and Kirkpatrick 2004), between 30 and 70% of genes are expressed in plant pollen (Sandler et al. 2018; Honys and Twell 2004), and pollen competition is thought to play a major role in plant evolution (Delph 2019). This allows for extensive haploid selection at the pollen level, which is thought to have a number of consequences for sex chromosome evolution (Sandler et al. 2018): First haploid selection may restrain the rate of sex-chromosome degeneration, because purifying selection will be efficient at maintaining functional gene copies on the Y chromosome required for pollen function. This is because due to the haploid nature of pollen there will be no masking effect of functional copies on the X chromosome. Besides the young age of many plant sex chromosome systems, this may be another main reason for the limited degree of plant sex chromosome degeneration compared to many animal groups. Evidence consistent with this process has been found in *Silene latifolia* (Chibalina and Filatov 2011) and in two species of *Rumex* (Sandler et al. 2018), where genes on the Y chromosome with high expression in pollen show slower rates of sequence degeneration and reduced loss of expression. Second, the presence of genes on the Y chromosome that contribute to pollen competition can favor the evolution of recombination suppression on the Y

chromosome. Linkage between these genes and the sex determining locus (Scott and Otto 2017) can be favored because the Y chromosome is always present in pollen. In contrast, the X chromosome is present in female gametes half of the time, where it does not experience competition. Third, these processes could further favor the enrichment of genes that experience haploid selection on the Y chromosome, thereby increasing both sequence and gene expression divergence between X and Y chromosomes. This process of "pollenization" has been found in *Rumex* (Sandler et al. 2018). Also, translocation and duplication of autosomal genes to the Y linked region, which could be involved in this process, have been found on the Y chromosomes of *Spinacia oleracea*, with several of these genes being related to pollen tube growth (She et al. 2020).

Haploid selection on pollen function is expected to result in male-biased sex ratios (Scott and Otto 2017), but later stages of sex-chromosome degeneration may compromise pollen fitness and result in female-biased sex ratios. Comparison of sex ratios in flowering plants do indeed show that male-biased sex ratios are more common, while there is a significant association of female-biased sex ratios with possessing heteromorphic sex chromosomes (Field et al. 2013). Further experimental evidence of the effect of sex-chromosome degeneration on pollen competitive ability comes from *Rumex nivalis*, where sex ratios became more female biased with increased pollination intensity (Stehlik and Barrett 2006), a process termed 'certation'. Certation is likely a consequence of sex chromosome-degeneration on the XY_1Y_2 sex-chromosome system of *R. nivalis*.

These examples show that while sex-chromosome evolution in plants is shaped by many of the same forces as in animals, there is likely also a strong contribution of plant-specific processes like haploid selection. Disentangling these factors will require both a better understanding of the genetics of sex determination in plants as well as the characterization of size, gene content and degeneration of non-recombining regions on plant sex chromosomes. The rapid development of genomic sequencing technologies is making it feasible to quickly and cost-efficiently generate these kinds of datasets for non-model species. The high degree of evolutionary convergence with likely hundreds of independent origins of sex-chromosome systems in plants will make these upcoming analyses interesting, allowing us to address questions beyond those relevant to the evolution of sexual systems, such as the evolution of transposable elements, recombination and the repeatability of genome evolution.

1.3.2 Genetics of polygenic and quantitative sex determination systems

Even though multiple genes may be involved in plant sex determination, the evolution of recombination suppression on sex chromosomes ensures that sex-determination systems act as a single bi-allelic Mendelian locus. Such a system is an efficient and simple way to achieve equal sex ratios in bisexual systems. However, in systems were sex is not completely binary but where heritable variation in sex expression within individuals exists, more complex genetic systems can be involved in determining sex expression.

Examples for these kind of sexual systems are found in the stinging nettle, *Urtica dioica* and in the castor bean, *Ricinus communis*. In *U. dioica*, male, female and monoecious individuals can be found in the same population. Sex ratios from crosses between plants of all sexes from a single population of *U. dioica* revealed sex ratios that largely conform with a hypothetical model where sex is determined by a single locus with four alleles. However several crossing families showed deviations from the expected sex ratios, suggesting a more complex mode of sex determination, likely involving multiple loci (Glawe and Jong 2009). A larger crossing study, which also involved the completely monoecious subspecies *U. gracilis*, suggested that sex expression of males, females, and monoecious individuals is determined by multiple genetic mechanisms, probably also with a strong maternally inherited component (Shannon and Holsinger 2007). In *R. communis*, sex expression varies in both the sex of inflorescences and their relative position on the plant and branches, and crossing experiments indicate that this is under polygenic control (George Jr and Shifriss 1967). In addition, sex reversal from femaleness towards monoecy is found in some lineages, where multiple genes control the timing of sex reversal (Shifriss 1956).

In many plants, sex expression is even more continuous. Continuous variation in sex expression is particularly common in hermaphroditic and monoecious species, but also between individuals of the same sex in dioecious species, where males or females can often express a varying degree of flowers of the opposite sex (Ehlers and Bataillon 2007). If this kind of variation is heritable, it could both be involved in transitions to dioecy (Charlesworth and Charlesworth 1978b) and in reversions from dioecy towards combined sexes, possibly as a result of selection for reproductive assurance. A heritable genetic component of variation in sex expression has been shown experimentally in *Hebe subalpina* (Delph and Lloyd 1991) and in *Solanum carolinense* (Elle 1998), but the genetic architecture underlying these traits is unknown, and further studies of genetics of polygenic and quantitative sex expression are required for a better understanding of transitions of sexual systems.

1.4 The effects of introgression, hybridization and polyploidy on sexual system evolution

Besides the evolution of the genetics of sex determination themselves, two additional genomic and population genetic processes, polyploidy and introgression, are very common in plants and have the potential to influence the evolution of sexual systems.

1.4.1 Introgression and hybridization

Introgression is defined as gene flow across species barriers through hybridization and repeated backcrossing with one of the parental species (Anderson and Hubricht 1938). Although hybridization has been estimated to occur in as many as 25% of plant species (Mallet 2005), its evolutionary importance has been underappreciated for a long time because allopatric speciation has been the dominant model of speciation. However, the development of next generation sequencing technology has challenged this view by uncovering genomic evidence of introgression events in a wide range of species, which show that low rates of hybridization can allow gene flow across species boundaries while maintaining species identity (Harrison and Larson 2014). Hybridization and introgression are being recognized as part of the speciation process, in which they shape species boundaries and can form new hybrid species (Abbott et al. 2013).

This modern paradigm of speciation views different genomic regions as being differently permeable to introgression (Wu 2001). Depending on species divergence, large parts of the genome may introgress across species boundaries, while some loci will be affected by incompatibilities resulting in hybrid offspring are either inviable, infertile or have other fitness disadvantages. In many species putative speciation loci have been identified by their genomic signature of increased differentiation between hybridizing species (e.g. Ellegren et al. 2012; Poelstra et al. 2014; Malinsky et al. 2015), although care has to be taken to distinguish them from signatures of divergent selection (Cruickshank and Hahn 2014). In contrast, other introgressed loci may be adaptive in their new genomic environment, and while on average 50% of introgressed genome content gets lost at every generation of backcrossing, directional selection can cause maintenance of these loci in the population. Prominent examples of such adaptive introgression events are found between modern humans, Neaderthals and Denisovans (Racimo et al. 2015), as well as in butterflies of the genus *Heliconius*, where a strong signal of introgression has been found at a wing color locus involved in Müllerian mimicry (Pardo-Diaz et al. 2012). There are also numerous examples for adaptive introgression in plants (Suarez-Gonzalez et al. 2018), where introgression has favored adaptation

to serpentine barrens (Arnold et al. 2016), and where it has been both involved in domestication of crops (Baute et al. 2015) and in the establishment of agricultural weeds (Le Corre et al. 2020).

The sexual system and the underlying genetics of sex determination can influence interspecific introgression in a variety of ways. The degree of selfing or outcrossing can affect introgression rates (Pickup et al. 2019). If a genetic system of self-incompatibility exists in both species, it may facilitate introgression via strong frequency dependent selection of novel, interspecific self-incompatibility alleles (Schierup and Vekemans 2008), but it may also result in asymmetries if one of the species rejects interspecific pollen. In the absence of such genetic systems, introgression rates in predominantly selfing or outcrossing species may be influenced by multiple factors. In general, outcrossers are expected to be more efficient at pollen transfer, which should greatly facilitate introgression. However, subsequent selection may select against hybrid ancestry if it comes with a high genetic load, which in turn can strongly depend on the mating system (Pickup et al. 2019).

Sex chromosomes tend to play an outstanding role in the genomic introgression landscape. It has long been recognized that it is usually the heterogametic sex that is more likely to be inviable or sterile in interspecific crosses (Haldane 1922). This pattern, named Haldane's Rule, has been found in a large number of animal species (Schilthuizen et al. 2011), as well as in dioecious plants (Brothers and Delph 2010; Kasjaniuk et al. 2019). It is now thought that the sterility and decreased viability described by Haldane's Rule is most likely caused by sex chromosome evolution, as genetic data from both genome scans and natural hybrid zones often show strongly reduced introgression patterns at markers located on the non-recombining parts of the sex chromosomes (e.g. Saetre et al. 2003; Gowen et al. 2014; Geraldes et al. 2008). There have been multiple, non-mutually exclusive explanations of the causes of Haldane's Rule (Delph and Demuth 2016), most of them related to sex chromosome evolution and recombination suppression. The theory with the most empirical support, the dominance hypothesis, is based on the evolution of partially recessive Bateson-Dobzhansky-Muller incompatibilities: sex-chromosome degeneration results in the loss of functional and expressed gene copies on the Y, which cannot restore the function of incompatible alleles on the hybrid X copy, resulting in sterility or inviability (Turelli and Orr 1995). This is concordant with patterns from animal crosses, which indicate that sterility and inviability are associated with increased sex-chromosome divergence (Lima 2014), but further evidence is required to test which features resulting from sex-chromosome evolution and degeneration are associated with incompatibilities in the heterogametic sex, and dioecious plants with their many independently evolved young sex chromosome systems have a great potential for further insights into these questions.

Hybridization can also affect the evolution of sexual systems through introgression of genes that affect

sex expression. Evidence for this "hybridization hypothesis" is found in the aquatic plant *Sagittaria latifolia*, where both subdioecy and androdioecy are associated with strong signals of hybridization and introgression between populations with dioecious and monoecious sexual phenotypes (Yakimowski and Barrett 2016). As analysis of *S. latifolia* was performed on the basis of neutral genetic markers, the underlying genetics of the phenotypic variation in sexual system remain unclear. Nonetheless, the example shows that hybridization could be an important driver in sexual-system evolution, which has been largely overlooked and will require further genomic analysis of hybridizing species with different sexual systems.

1.4.2 Polyploidy

Another important genomic feature that can substantially influence the evolution of sexual systems is polyploidy. Polyploids have more than the usual two genome copies found in diploids, as a result of one or multiple whole genome duplications. Polyploidy is very common in plants (Otto and Whitton 2000), and there has been a rising appreciation of its evolutionary importance through the use of next-generation sequencing. While polyploidy has been previously considered an evolutionary dead end (Soltis et al. 2014), comparative studies using genomic and transcriptomic sequencing data have shown evidence of multiple ancestral polyploidization events across the phylogeny of flowering plants (Jiao et al. 2011). Similarly, ancient polyploidization has been inferred in many animal clades, and it is estimated that polyploidy has been involved in about 15% of speciation events in plants (Wood et al. 2009). Generally, polyploids are distinguished into autopolyploids, where duplicated chromosome sets originate from a single species, and allopolyploids, where whole genome duplications coincide with hybridization and as a result genome copies from different species are combined in the same individual (Stebbins Jr 1947). Auto- and allopolyploids differ strongly in their recombination patterns, with autopolyploids tending to recombine freely between homeologous loci (resulting in tetrasomic inheritance of chromosomes in progeny), while recombination tends to be restricted between homeologs with a common origin in allopolyploids (resulting in disomic inheritance). However, these two modes of polyploidy represent two extremes along a continuum, and tetrasomic inheritance or intermediates are found in some genomic regions in allotetraploids.

Polyploidization can influence both genomic and phenotypic evolution, either through immediate effects of polyploidization or by shaping their subsequent evolutionary trajectories. Immediate genomic effects of polyploidization are often described with 'genomic shock' and can include gene expression changes, mobilization of transposable elements and meiotic irregularities. Most of these outcomes are probably maladaptive factors, and have been overcome in the successful establishment of neopolyploids (Hollister 2015). Neopolyploidization can sometimes also result in immediate phenotypic and physiological changes

(Chao et al. 2013), but mostly phenotypic changes are considered to be the consequence of subsequent evolutionary processes, which can facilitate adaptation (Selmecki et al. 2015). Whole genome duplications are thought to provide raw material for the evolution of novelty by providing redundant gene copies that are free to acquire novel functions or allow the novel copies to diverge in functionality; these two processes are called neo- and sub-functionalization, respectively. In allopolyploids, there can be additional effects due to the combination of genomes from divergent lineages in a permanent heterozygous state, potentially resulting in immediate transgressive gene expression patterns (Buggs, Zhang, et al. 2011). Such effects can influence subsequent evolutionary trajectories, e.g. through chromosomal variation (Chester et al. 2012) and differential regulation of transposable elements (Parisod et al. 2010).

Polyploidization also affects hybridization and reproductive isolation. While allopolyploidization results in the permanent combination of hybrid genotypes, polyploidy is often considered to result in immediate reproductive isolation between cytotypes because the offspring of crosses between polyploids and their diploid progenitors, or with relatives of different ploidy levels, usually show reduced fertility due to meiotic mis-segregation (Ramsey and Schemske 1998). Alternatively, such hybrids may be inviable due to dosage imbalance problems, for example during endosperm cellularization (Lafon-Placette and Köhler 2016). However, recent studies suggest that the offspring of these interploidy crosses may not be completely infertile, and may act as a so-called 'triploid-bridge', allowing for gene flow between ploidy levels (Husband 2004). There is recent evidence of interploidy gene flow in natural populations (Zohren et al. 2016; Monnahan et al. 2019), and there is even some evidence for adaptive introgression between ploidy levels (Chapman and Abbott 2010).

Polyploidization can also affect the sexual system in various ways. A broad association between polyploidy and transitions towards dimorphic sexual systems has been found in plants (Glick et al. 2016), although there is substantial variation between clades, and there are also examples of clades where polyploidy is associated with transitions from dioecy towards monomorphic sexual systems (Miller and Venable 2000; Volz and Renner 2008). This variation reflects the many potential cause-effect relationships between polyploidy and sexual system evolution. There may be either direct effects of whole-genome duplications, with differences between auto- and allopolyploids, or subsequent selection for sexual-system transitions as a result of polyploidization. Alternatively, changes in sexual systems may favor the establishment of polyploids (Pannell, Obbard, et al. 2004; Ashman et al. 2013). Immediate effects of polyploidization on the sexual system may be caused by dosage effects, or by the combination of two or more genomes with divergent history and potentially divergent evolution of their sexual systems in allopolyploids. This could result in male or female sterility, which could be a first step in the evolution of separate sexes (Ashman et al. 2013). Polyploidy can also cause the breakdown of genetic systems of self-incompatibility, resulting

in transitions to selfing, which in turn could select for the evolution of separate sexes as a new means of inbreeding avoidance (Miller and Venable 2000). Subsequent effects of polyploidization can facilitate the evolution of transitions of sexual systems, e.g., polyploidy can effect inbreeding depression and the genetic load by facilitating the sheltering of recessive deleterious variants through the presence of multiple redundant gene copies. These effects can be different between autopolyploids and allopolyploids. In the former, inbreeding depression is low upon initial polyploid formation, but can increase in subsequent generations of selfing (Ronfort 1999; Husband et al. 2008). In the latter, the lack of recombination between homeologs can result in permanent masking of deleterious variants and reduction of inbreeding depression (Vandepitte et al. 2011). It has been hypothesized that reproductive assurance through selfing can favor the establishment of newly formed polyploids (Spoelhof et al. 2020), because it can prevent neopolyploids from mating primarily with their more numerous diploid relatives, which would result in interploidy offspring with low fitness (Levin 1975). However, it is unclear how often this has actually happened, as there is no significant statistical association between polyploidy and loss of self-incompatibility at the species and family level (Mable 2004): In general more research is needed to understand the complex associations between polyploidy and sexual systems transitions and especially the underlying genetics.

1.5 Transitions between sexual systems in *Mercurialis*

The genus *Mercurialis* (Euphorbiaceae) consists of ten nominal species of perennial or annual herbs and bushes. These species are found mostly in central Europe and around the Mediterranian Sea (Fig. 1.1), with one species in southeast Asia (Krähenbühl et al. 2002). Dioecious *M. annua* was one of the first plant species in which the Mendelian inheritance of sex was shown experimentally (Yampolsky 1919), and the diversity of sexual systems in related lineages make *Mercurialis* an interesting clade to study the evolution of sexual systems. While dioecy is ancestral in *Mercurialis* (Krähenbühl et al. 2002), within the annual clade there are lineages in which populations contain either monoecious plants with both male and female flowers (functional hermaphrodites), or monoecious plants and males (functional androdioecy) (Durand 1963, Tab. 1.1). Like in many other species in this area, the distribution of these annual lineages, which are found in central Europe and across the Mediterranean Sea in North Africa and the Middle East (Fig. 1.1), is the result of range expansions from ice-age refugia after the last Pleistocene glacial maximum (Obbard, Harris, and Pannell 2006). Patterns of genetic diversity suggest that different lineages recolonized Europe from different refugia: dioecious *M. annua* recolonized central and eastern Europe from the middle east, which resulted in strong signatures of genetic bottlenecks, an increase in the number of deleterious mutations and limitations to adaptive evolution (Gonzalez-Martinez et al. 2017). Androdioecious *M. annua*

colonized the Iberian peninsula from Northern Africa, which resulted in a strong population bottleneck, reflected in a lower genetic diversity in Spanish populations (Obbard, Harris, and Pannell 2006), as well as a reduction of inbreeding depression (Pujol, Zhou, et al. 2009) and a reduction in the response to selection (Pujol and Pannell 2008). There is evidence that this range expansion is an ongoing process: in northeastern and nortwestern Spain, the dioecious lineage of *M. annua* is rapidly replacing populations of the monoecious lineage in two independent contact zones, with estimates suggestting that their contact zones have moved by 80 and 200 km respectively over a period of about 40 years (Buggs and Pannell 2006).

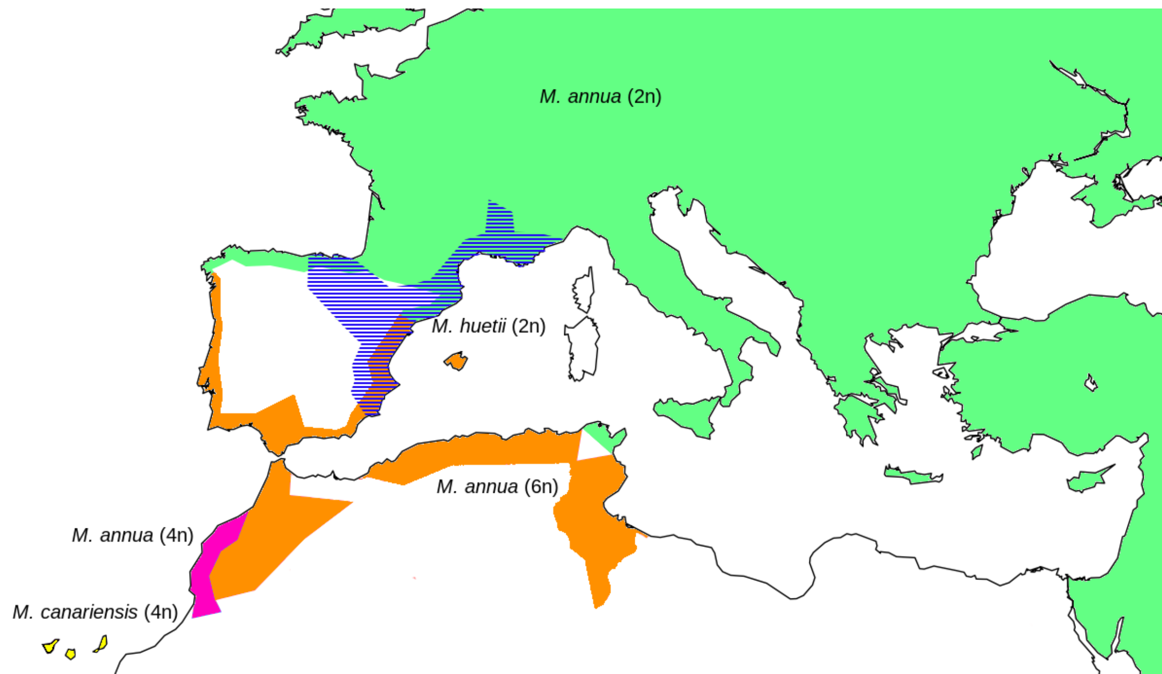


Figure 1.1: Distribution of lineages of annual *Mercurialis* species across Europe, northern Africa and the Middle East. Green: diploid *M. annua*; purple: tetraploid *M. annua*; orange: hexaploid *M. annua*; yellow: *M. canarienesis*; blue: *M. huetii*

Table 1.1: *Mercurialis* lineages used in this study. Two different chromosome counts were published for *M. elliptica*, we assume the lower count to be more likely to be correct.

Lineage	Sexual system	Life history	Chromosome count	Ploidy
M. annua	Dioecy	annual	16	Diploid
M. annua	Monoecy/Androdioecy	annual	32	Tetraploid
M. annua	Monoecy/Androdioecy	annual	48	Hexaploid
M. canariensis	Dioecy	annual	32	Tetraploid
M. huetii	Dioecy	annual	16	Diploid
M. reverchonii	Dioecy	perennial	26	Tetraploid
M. tomentosa	Dioecy	perennial	26	Tetraploid
M. elliptica	Dioecy	perennial	42 (220)	Hexaploid

Probably the most apparent feature of sex expression in *Mercurialis* is the location of male flowers on

elongated stalks, so-called peduncules (Durand and Durand 1991), in males. This structure facilitates outcrossing, and in experimental populations the presence of peduncules alone conferred a 60% outcrossing advantage on pollen compared with pollen produced by monoecious plants (Eppley and Pannell 2007a), in addition to the fact that males produce substantially more pollen than monoecious plants (Pannell 1997c). In contrast, male flowers on monoecious plants are usually located in a similar axillary position as female flowers in both females and monoecious plants. However, monoecious populations with male flowers also located on peduncules have recently been documented in southeastern Spain. These populations exist in close proximity to populations in which monoecious plants do not show this trait, but crossing experiments showed that both forms are reproductively isolated (Ma et al. 2019). While pedunculate males usually make up about 50% of individuals in dioecious populations of diploid *M. annua*, *M. huetii* and *M. canariensis*, their proportion can vary strongly when they co-occur with monoecious plants. In these populations, there is also considerable variation in sex expression, and the monoecious plants can display a wide range of sexual phenotypes, which can be almost completely male or female in extreme cases (Pannell 1997b; Pannell, Eppley, et al. 2014). In addition, in both sexes of dioecious *M. annua* females often produce small numbers of male flowers and vice versa (Cossard and Pannell 2019). This kind of 'leakiness' in sex expression is also found in females of *M. canariensis* (personal observation).

An essential factor for the establishment and maintenance of variation in sex expression in annual *Mercurialis* species is its population structure. Populations are often found in spatially isolated patches of disturbed habitats, which are characterized by frequent extinctions and recolonizations. In these kind of metapopulations, plants that can self fertilize by producing gametes of both sexes can have a selective advantage over strictly unisexual individuals. Specifically, unlike pure females, females that also produce male flowers can self-fertilize in the absence of pollen produced by males (Hesse and Pannell 2011), which allows them to colonize empty habitat patches. This ability of selfing individuals to colonize habitat patches is similar to Baker's Law (Baker 1955), and theoretical models predict that selection for reproductive assurance will be greatest if patch abundance is low (Pannell and Barrett 1998). In addition to providing a selective explanation for the evolution of monoecy from females, these metapopulation processes, in combination with floral architecture of males, can also explain the evolution of functional androdioecy (Pannell 2000), a mating system that would otherwise be difficult to evolve (Pannell 2002). As males have a strong outcrossing advantage over monoecious plants (Eppley and Pannell 2007a), they should easily invade populations consisting only of monoecious plants that emphasise their female function if population densities are high (and if the population are thus largely outcrossing). According to this model, after females evolve monoecy as a result of selection for reproductive assurance, males can still persist in a dynamic state, where empty habitat patches are first colonized by few, largely selfing, monoecious individuals. As

these populations grow in size, their increased population density will favor the establishment of males. Populations of *M. annua* tend to turnover regularly, and modelling predicts that the maintenance of males in the metapopulation is sensitive to the relative rates of immigration and population turnover (Pannell 1997a).

Natural populations of monoecious or androdioecious *M. annua* show variation in the proportion of males, both on a local metapopulation scale and on a regional scale (Pannell, Eppley, et al. 2014). Thus, while males are found in varying proportions over large parts of the range of monoecious populations in Spain and northern Africa, they are largely absent at the edges of the species range in northeastern and western Spain (Obbard, Harris, and Pannell 2006) as well as in a different lineage on the Atlantic coast in Morocco (Durand 1963). This variation allowed to test some predictions of the meta-population model. First, patterns of genetic variation within and between populations confirmed predictions that population differentiation is higher and within-population diversity is lower in populations in which males are absent (Obbard, Harris, and Pannell 2006). Second, populations consisting only of monoecious *M. annua* are generally smaller and less abundant than those in which males and monoecious plants co-occur, in line with predictions of relative rates of immigration and population turnover required for the maintenance of males in a metapopulation (Eppley and Pannell 2007b). These patterns point to the importance of metapopulation processes for the selection of sexual diversity in *Mercurialis*. They also suggest that such processes might be of greater general importance for the evolution of sexual systems than has been previously appreciated, as the canonical models that explain the evolution of sexual systems assume large, unstructured populations. Despite the progress in understanding sexual system tansitions, it is still unclear how such transitions occur in detail, and what specific role might be played by different historical biogeographic, genetic, genomic and population genetic factors involved in these transitions (Pannell, Obbard, et al. 2004).

It is not known, for example, what role polyploidy or hybridization might play. The chromosome numbers of lineages with different sexual systems suggest an association between polyploidy and transitions in sexual systems within the annual clade of *Mercurialis*: the dioecious lineage of *M. annua* and dioecious *M. huetii* are diploid with a chromosome number of 2x=16 and only *M. canariensis* is tetraloid with double the number of chromosomes (2n=32) (Obbard, Pannell, et al. 2006). In contrast, all hermaphrodite and androdioecious lineages of *M. annua* are polyploids with chromosome numbers from 2n=32 to 2n=96 (Durand 1963) (Tab. 1.1). Early experiments, in which artificially induced whole genome duplications in diploid, dioecious *M. annua*, using the mitosis inhibitor colchicine, caused the expression of male flowers on females, indicated that whole genome duplications could be directly related to changes in the sexual system (Durand 1963). However, it remains unclear to what degree these changes are the direct result of polyploidization or the colchicine treatment. In addition, evidence from phylogenetic markers indicates

that at least two of the polyploid lineages (tetraploid *M. canariensis* and monoecious and androdioecious hexaploid *M. annua*) are allopolyploids (Obbard, Harris, Buggs, et al. 2006). In this case, sexual-system evolution may not only be affected by polyploidy, but by the combination of genomic information from different lineages in the same individual as a result of historical hybridization events.

We are also beginning to build a picture of the genetics of sex determination in *Mercurialis*. Maleness is determined by an XY system of sex determination in diploid *M. annua*, and crosses between annual lineages resulted in mostly 50/50 sex ratios, which is consistent with an ancestrally shared system of sex determination in these lineages (Russell and Pannell 2015). Although the sex chromosomes of *M. annua* are homomorphic, analysis of transcriptome data and BAC clones identified a nonrecombining region on the Y chromosome, with an estimated size between 14.5 and 19 mb (Veltsos, Cossard, et al. 2018), a region that includes 500 genes (Veltsos, Ridout, et al. 2019). In addition, there is evidence of additional genetic and plastic mechanisms that determine sex expression in monoecious plants that coexist with male determining sex chromosomes in androdioecious populations. In experimental populations, monoecious plants from androdioecious populations evolved a stronger female biased sex expression when males where absent, while sex expression in plants did not change in the presence of males (Dorken and Pannell 2009). These results imply that there is quantitative genetic variation in sex allocation in *M. annua* on which selection might act. In addition, there is evidence for a plastic response of sex allocation to density, with hermaphrodites producing more pollen when grown at lower densities in experimental populations (Dorken and Pannell 2008). Recent experiments show that both quantitative genetic variation in sex expression (Cossard et al. 2021) and a plastic response to pollen availability (Cossard and Pannell 2021) also apply for leaky sex expression in diploid, dioecious *M. annua*.

1.6 Aims and outline of the thesis

In the following chapters I analyse genomic datasets from several species of the genus *Mercurialis* with the aim of exploring several questions in plant sex chromosome evolution and in the genomics of transitions between sexual systems. In Chapter 2 and Chapter 3, I address several questions in sex chromosome evolution. Understanding the evolution of sex chromosomes is essential for understanding transitions in sexual systems, since they are involved in the evolution of separate sexes, but they can also play essential roles in transitions from dioecy towards combined sexes. In addition, their unique evolutionary trajectories can have other important population genetic consequences, which can have secondary effects on sexual system evolution. However, most of the expectations on plant sex chromosome evolution are based on

animal systems and more data on plant sex chromosomes are needed. In Chapter 2, I want to test if plant sex chromosomes contribute to reproductive isolation in the same way as has often been shown for animal sex chromosomes. In addition, I also address the role of hybridization and polyploidization for sexual system evolution. Both of these processes are very common in plants and they can have fundamental influences on population genetics and genome evolution, and can influence the evolution of sexual systems in different ways. Sex chromosome evolution, hybridization and polyploidy are all involved in transitions towards androdioecy in hexaploid *M. annua* and I attempt to improve our understanding of their individual roles. Specifically, I address the following questions:

- Are sex chromosomes always involved in reproductive isolation or are there circumstances where they can introgress between lineages?
- What role can allopolyploidization and introgression play in transitions between sexual systems?
- What is the origin of the Y chromosome in androdioecious hexaploid *M. annua*?

I use a set of newly-developed phylogenetic markers to improve our understanding of the phylogeny and history of hybridization and allopolyploidization of both autosomes and the sex chromosomes of annual *Mercurialis* lineages and their perennial sister species. I discuss the results in relation to the origin and maintenance of males in androdioecious populations of *M. annnua*.

In Chapter 3 I examine the role of recombination for the evolution of plant sex chromosomes. Recombination is essential for understanding the evolution of sex chromosomes because recombination suppression often evolves on animal sex chromosomes. However, it is unclear what role it plays for the evolution of separate sexes and if it does evolve on plant sex chromosomes after the evolution of separate sexes. In addition, recent studies show that specifically plant sex chromosomes often evolve in regions with reduced recombination and further studies are needed to understand how this relates to different genetic architectures of bisexuality. I address the following questions in this chapter:

- What role does variation of recombination rates between sexes and on different parts of the chromosome play for the evolution of sex chromosomes and separate sexes?
- Can recombination suppression evolve on plant sex chromosomes after the evolution of separate sexes?

I use new genome assemblies, crosses and an exon-capture dataset to further explore the origin of the Y chromosome of *M. annnua*. I use linkage maps to compare the recombination landscape on both sex

chromosomes and the autosomes. In addition I test the extent of recombination suppression on the Y chromosome of annual and perennial lineages of *Mercurialis*. These results are relevant for understanding sex-chromosome evolution in *M. annnua*, but also in a wider sense, because they add an additional dataset on the evolution of plant sex chromosomes. This dataset is one of the first plant datasets which compares recombination suppression between diverged lineages which share the same ancestral sex chromosome system.

In Chapter 4 I explore the genetic architecture of transitions towards combined sexes. Recent studies show that transitions from separate towards combined sexes could be common in plants, but very little is known about the underlying genetic architecture. In this chapter I analyze the genetics underlying transitions from dioecy to monoecy, which were caused by selection on naturally occurring leakiness in females in an experimental evolution setup. Selection on such leakiness is a plausible way how transitions towards combined sexes could happen and I use this system to address the following questions:

- What kind of genetic architecture could be involved in transitions between separate and combined sexes? Are many small effect loci or can transitions be caused by few loci with large effects?
- Do these loci effect the expression of one sexual function or do they regulate both sex functions at the same time?

I use crosses between control and selection lines to perform QTL mapping with the aim of identifying genomic regions involved in increased leakiness. This can help to better understand the genetics of sexual-system transitions that involve quantitative genetic variance in sex expression, but it is also of wider interest because little is known about the genetic architecture of major evolutionary transitions in general. The final chapter provides a synthesis and general discussion of the key results of this thesis.

References

Abbott, Richard, Dirk Albach, Stephen Ansell, Jan W Arntzen, Stuart JE Baird, Nicolas Bierne, Jenny Boughman, Alan Brelsford, C Alex Buerkle, Richard Buggs, et al. (2013). "Hybridization and speciation". In: *Journal of evolutionary biology* 26.2, pp. 229–246.

Akagi, Takashi, Isabelle M Henry, Ryutaro Tao, and Luca Comai (2014). "A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons". In: *Science* 346.6209, pp. 646–650.

Akagi, Takashi, Sarah M Pilkington, Erika Varkonyi-Gasic, Isabelle M Henry, Shigeo S Sugano, Minori Sonoda, Alana Firl, Mark A McNeilage, Mikaela J Douglas, Tianchi Wang, et al. (2019). "Two Y-chromosome-encoded genes determine sex in kiwifruit". In: *Nature plants* 5.8, pp. 801–809.

Anderson, Edgar and Leslie Hubricht (1938). "Hybridization in Tradescantia. III. The evidence for introgressive hybridization". In: *American Journal of Botany*, pp. 396–402.

Arnold, Brian J, Brett Lahner, Jeffrey M DaCosta, Caroline M Weisman, Jesse D Hollister, David E Salt, Kirsten Bomblies, and Levi Yant (2016). "Borrowed alleles and convergence in serpentine adaptation". In: *Proceedings of the National Academy of Sciences* 113.29, pp. 8320–8325.

Ashman, T-L, A Kwok, and BC Husband (2013). "Revisiting the dioecy-polyploidy association: alternate pathways and research opportunities". In: *Cytogenetic and genome research* 140.2-4, pp. 241–255.

Bachtrog, Doris (2013). "Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration". In: *Nature Reviews Genetics* 14.2, pp. 113–124.

Bachtrog, Doris, Judith E Mank, Catherine L Peichel, Mark Kirkpatrick, Sarah P Otto, Tia-Lynn Ashman, Matthew W Hahn, Jun Kitano, Itay Mayrose, Ray Ming, et al. (2014). "Sex determination: why so many ways of doing it?" In: *PLoS biology* 12.7, e1001899.

Badouin, Hélène, Amandine Velt, François Gindraud, Timothée Flutre, Vincent Dumas, Sonia Vautrin, William Marande, Jonathan Corbi, Erika Sallet, Jérémy Ganofsky, et al. (2020). "The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication". In: *Genome biology* 21.1, pp. 1–24.

Baker, Herbert G (1955). "Self-compatibility and establishment after'long-distance'dispersal". In: *Evolution* 9.3, pp. 347–349.

Barrett, Spencer C. H. (2002). "The evolution of plant sexual diversity". In: *Nature Reviews Genetics* 3.4, pp. 274–284.

Barrett, Spencer CH (2019). "'A most complex marriage arrangement': recent advances on heterostyly and unresolved questions". In: *New Phytologist* 224.3, pp. 1051–1067.

Barton, Nicholas H and Brian Charlesworth (1998). "Why sex and recombination?" In: *Science* 281.5385, pp. 1986–1990.

Bateman, Angus J (1948). "Intra-sexual selection in Drosophila". In: *Heredity* 2.3, pp. 349–368.

Baute, Gregory J, Nolan C Kane, Christopher J Grassa, Zhao Lai, and Loren H Rieseberg (2015). "Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives". In: *New Phytologist* 206.2, pp. 830–838.

Bawa, Kamaljit S (2016). "Kin selection and the evolution of plant reproductive traits". In: *Proceedings of the Royal Society B: Biological Sciences* 283.1842, p. 20160789.

Bergero, Roberta and Deborah Charlesworth (2011). "Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system". In: *Current Biology* 21.17, pp. 1470–1474.

Bergero, Roberta, Alan Forrest, Esther Kamau, and Deborah Charlesworth (2007). "Evolutionary strata on the X chromosomes of the dioecious plant Silene latifolia: evidence from new sex-linked genes". In: *Genetics* 175.4, pp. 1945–1954.

Bergero, Roberta, Jim Gardner, Beth Bader, Lengxob Yong, and Deborah Charlesworth (2019). "Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms". In: *Proceedings of the National Academy of Sciences* 116.14, pp. 6924–6931.

Betancourt, Andrea J, John J Welch, and Brian Charlesworth (2009). "Reduced effectiveness of selection caused by a lack of recombination". In: *Current biology* 19.8, pp. 655–660.

Beukeboom, Leo W and Nicolas Perrin (2014). *The evolution of sex determination*. Oxford University Press, USA.

Bopp, D, G Saccone, and M Beye (2014). "Sex determination in insects: variations on a common theme". In: *Sexual Development* 8.1-3, pp. 20–28.

Boualem, Adnane, Christelle Troadec, Céline Camps, Afef Lemhemdi, Halima Morin, Marie-Agnes Sari, Rina Fraenkel-Zagouri, Irina Kovalski, Catherine Dogimont, Rafael Perl-Treves, et al. (2015). "A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges". In: *Science* 350.6261, pp. 688–691.

Branco, Sara, Hélene Badouin, Ricardo C Rodrîguez de la Vega, Jérôme Gouzy, Fantin Carpentier, Gabriela Aguileta, Sophie Siguenza, Jean-Tristan Brandenburg, Marco A Coelho, Michael E Hood, et al. (2017).

"Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism". In: *Proceedings of the National Academy of Sciences* 114.27, pp. 7067–7072.

Brothers, Amanda N and Lynda F Delph (2010). "Haldane's rule is extended to plants with sex chromosomes". In: *Evolution* 64.12, pp. 3643–3648.

Brush, Stephen G (1978). "Nettie M. Stevens and the discovery of sex determination by chromosomes". In: *Isis* 69.2, pp. 163–172.

Buggs, Richard JA and John R Pannell (2006). "Rapid displacement of a monoecious plant lineage is due to pollen swamping by a dioecious relative". In: *Current Biology* 16.10, pp. 996–1000.

Buggs, Richard JA, Linjing Zhang, Nicholas Miles, Jennifer A Tate, Lu Gao, Wu Wei, Patrick S Schnable, W Brad Barbazuk, Pamela S Soltis, and Douglas E Soltis (2011). "Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant". In: *Current Biology* 21.7, pp. 551–556.

Busch, Jeremiah W and Lynda F Delph (2012). "The relative importance of reproductive assurance and automatic selection as hypotheses for the evolution of self-fertilization". In: *Annals of botany* 109.3, pp. 553–562.

Campbell, Diane R (2000). "Experimental tests of sex-allocation theory in plants". In: *Trends in Ecology & Evolution* 15.6, pp. 227–232.

Capel, Blanche (2017). "Vertebrate sex determination: evolutionary plasticity of a fundamental switch". In: *Nature Reviews Genetics* 18.11, p. 675.

Chalopin, Domitille, Jean-Nicolas Volff, Delphine Galiana, Jennifer L Anderson, and Manfred Schartl (2015). "Transposable elements and early evolution of sex chromosomes in fish". In: *Chromosome research* 23.3, pp. 545–560.

Chao, Dai-Yin, Brian Dilkes, Hongbing Luo, Alex Douglas, Elena Yakubova, Brett Lahner, and David E Salt (2013). "Polyploids exhibit higher potassium uptake and salinity tolerance in Arabidopsis". In: *Science* 341.6146, pp. 658–659.

Chapman, Mark A and Richard J Abbott (2010). "Introgression of fitness genes across a ploidy barrier". In: *New Phytologist* 186.1, pp. 63–71.

Charlesworth, B and JD Wall (1999). "Inbreeding, heterozygote advantage and the evolution of neo–X and neo–Y sex chromosomes". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266.1414, pp. 51–56.

Charlesworth, Brian (1993). "Directional selection and the evolution of sex and recombination". In: *Genetics Research* 61.3, pp. 205–224.

Charlesworth, Brian and Deborah Charlesworth (1978a). "A model for the evolution of dioecy and gynodioecy". In: *The American Naturalist* 112.988, pp. 975–997.

Charlesworth, D (1984). "Androdioecy and the evolution of dioecy". In: *Biological Journal of the Linnean Society* 22.4, pp. 333–348.

Charlesworth, D (2002). "Plant sex determination and sex chromosomes". In: *Heredity* 88.2, pp. 94–101.

Charlesworth, D and B Charlesworth (1978b). "Population genetics of partial male-sterility and the evolution of monoecy and dioecy". In: *Heredity* 41.2, pp. 137–153.

Charlesworth, D and B Charlesworth (1980). "Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes". In: *Genetics Research* 35.2, pp. 205–214.

Charlesworth, D and B Charlesworth (1981). "Allocation of resources to male and female functions in hermaphrodites". In: *Biological Journal of the Linnean Society* 15.1, pp. 57–74.

Charlesworth, Deborah and Brian Charlesworth (1987). "Inbreeding depression and its evolutionary consequences". In: *Annual review of ecology and systematics* 18.1, pp. 237–268.

Charlesworth, Deborah, Xavier Vekemans, Vincent Castric, and Sylvain Glémin (2005). "Plant self-incompatibility systems: a molecular evolutionary perspective". In: *New Phytologist* 168.1, pp. 61–69.

Charlesworth, Deborah and John H Willis (2009). "The genetics of inbreeding depression". In: *Nature reviews genetics* 10.11, pp. 783–796.

Charlesworth, Deborah and Stephen I Wright (2001). "Breeding systems and genome evolution". In: *Current opinion in genetics & development* 11.6, pp. 685–690.

Charnov, Eric L (1979). "Simultaneous hermaphroditism and sexual selection". In: *Proceedings of the National Academy of Sciences* 76.5, pp. 2480–2484.

Charnov, Eric L, James J Bull, and J Maynard Smith (1976). "Why be an hermaphrodite?" In: *Nature* 263.5573, pp. 125–126.

Chester, Michael, Joseph P Gallagher, V Vaughan Symonds, Ana Veruska Cruz da Silva, Evgeny V Mavrodiev, Andrew R Leitch, Pamela S Soltis, and Douglas E Soltis (2012). "Extensive chromosomal variation in a recently formed natural allopolyploid species, Tragopogon miscellus (Asteraceae)". In: *Proceedings of the National Academy of Sciences* 109.4, pp. 1176–1181.

Chibalina, Margarita V and Dmitry A Filatov (2011). "Plant Y chromosome degeneration is retarded by haploid purifying selection". In: *Current Biology* 21.17, pp. 1475–1479.

Coelho, Susana M, Laure Mignerot, and J Mark Cock (2019). "Origin and evolution of sex-determination systems in the brown algae". In: *New Phytologist* 222.4, pp. 1751–1756.

Cossard, Guillaume G and John R Pannell (2019). "A functional decomposition of sex inconstancy in the dioecious, colonizing plant Mercurialis annua". In: *American journal of botany* 106.5, pp. 722–732.

Cossard, Guillaume G and John R Pannell (2021). "Enhanced leaky sex expression in response to pollen limitation in the dioecious plant Mercurialis annua". In: *Journal of Evolutionary Biology* 34.2, pp. 416–422.

Cruickshank, Tami E and Matthew W Hahn (2014). "Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow". In: *Molecular ecology* 23.13, pp. 3133–3157.

Culley, Theresa M, Stephen G Weller, and Ann K Sakai (2002). "The evolution of wind pollination in angiosperms". In: *Trends in Ecology & Evolution* 17.8, pp. 361–369.

Dacks, Joel and Andrew J Roger (1999). "The first sexual lineage and the relevance of facultative sex". In: *Journal of Molecular Evolution* 48.6, pp. 779–783.

Darwin, Charles (1876). *The effects of cross and self fertilisation in the vegetable kingdom*. J. Murray.

Delph, Lynda F (2019). "Pollen competition is the mechanism underlying a variety of evolutionary phenomena in dioecious plants". In: *New Phytologist* 224.3, pp. 1075–1079.

Delph, Lynda F and Jeffery P Demuth (2016). "Haldane's rule: genetic bases and their empirical support". In: *Journal of Heredity* 107.5, pp. 383–391.

Delph, Lynda F and David G Lloyd (1991). "Environmental and genetic control of gender in the dimorphic shrub Hebe subalpina". In: *Evolution* 45.8, pp. 1957–1964.

Dorken, Marcel E and John R Pannell (2008). "Density-dependent regulation of the sex ratio in an annual plant". In: *The American Naturalist* 171.6, pp. 824–830.

Dorken, Marcel E and John R Pannell (2009). "Hermaphroditic sex allocation evolves when mating opportunities change". In: *Current Biology* 19.6, pp. 514–517.

Dufay, Mathilde, P Champelovier, J Käfer, Jean-Pierre Henry, Sylvain Mousset, and GAB Marais (2014). "An angiosperm-wide analysis of the gynodioecy–dioecy pathway". In: *Annals of botany* 114.3, pp. 539–548.

Durand, B. and R. Durand (1991). "Sex determination and reproductive organ differentiation in Mercurialis". In: *Plant science* 80.1, pp. 49–65.

Durand, Bernard (1963). *Le complexe Mercurialis annua LSL: une étude biosystématique*. Vol. 12, pp. 579–736.

Ehlers, Bodil K and Thomas Bataillon (2007). "'Inconstant males' and the maintenance of labile sex expression in subdioecious plants". In: *New Phytologist* 174.1, pp. 194–211.

Elle, Elizabeth (1998). "The quantitative genetics of sex allocation in the andromonoecious perennial, Solanum carolinense (L.)" In: *Heredity* 80.4, pp. 481–488.

Ellegren, Hans, Linnea Smeds, Reto Burri, Pall I Olason, Niclas Backström, Takeshi Kawakami, Axel Künstner, Hannu Mäkinen, Krystyna Nadachowska-Brzyska, Anna Qvarnström, et al. (2012). "The genomic landscape of species divergence in Ficedula flycatchers". In: *Nature* 491.7426, pp. 756–760.

Ellison, Christopher and Doris Bachtrog (2019). "Recurrent gene co-amplification on Drosophila X and Y chromosomes". In: *PLoS genetics* 15.7, e1008251.

Eppley, Sarah M and John R Pannell (2007a). "Density-dependent self-fertilization and male versus hermaphrodite siring success in an androdioecious plant". In: *Evolution* 61.10, pp. 2349–2359.

Eppley, Sarah M and John R Pannell (2007b). "Sexual systems and measures of occupancy and abundance in an annual plant: testing the metapopulation model". In: *The American Naturalist* 169.1, pp. 20–28.

Eppley, SM and LK Jesson (2008). "Moving to mate: the evolution of separate and combined sexes in multicellular organisms". In: *Journal of evolutionary biology* 21.3, pp. 727–736.

Felsenstein, Joseph (1974). "The evolutionary advantage of recombination". In: *Genetics* 78.2, pp. 737–756.

Field, David L, Melinda Pickup, and Spencer CH Barrett (2013). "Comparative analyses of sex-ratio variation in dioecious flowering plants". In: *Evolution* 67.3, pp. 661–672.

Fisher, Ronald Aylmer (1930). *The Genetical Theory of Natural Selection*. Oxford University Press, pp. 141–143.

Fisher, Ronald Aylmer (1941). "Average excess and average effect of a gene substitution". In: *Annals of Eugenics* 11.1, pp. 53–63.

George Jr, William L and Oved Shifriss (1967). "Interspersed sexuality in Ricinus". In: *Genetics* 57.2, p. 347.

Geraldes, A, M Carneiro, M Delibes-Mateos, R Villafuerte, MW Nachman, and N Ferrand (2008). "Reduced introgression of the Y chromosome between subspecies of the European rabbit (Oryctolagus cuniculus) in the Iberian Peninsula". In: *Molecular Ecology* 17.20, pp. 4489–4499.

Glawe, Grit A and Tom J de Jong (2009). "Complex sex determination in the stinging nettle Urtica dioica". In: *Evolutionary ecology* 23.4, p. 635.

Glick, Lior, Niv Sabath, Tia-Lynn Ashman, Emma Goldberg, and Itay Mayrose (2016). "Polyploidy and sexual system in angiosperms: Is there an association?" In: *American journal of botany* 103.7, pp. 1223–1235.

Gonzalez-Martinez, Santiago C, Kate Ridout, and John R Pannell (2017). "Range expansion compromises adaptive evolution in an outcrossing plant". In: *Current Biology* 27.16, pp. 2544–2551.

Goodwillie, Carol, Susan Kalisz, and Christopher G Eckert (2005). "The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence". In: *Annu. Rev. Ecol. Evol. Syst.* 36, pp. 47–79.

Gordon, Swanne P, Andrés López-Sepulcre, and David N Reznick (2012). "Predation-associated differences in sex linkage of wild guppy coloration". In: *Evolution* 66.3, pp. 912–918.

Gowen, Fiona C, James M Maley, Carla Cicero, A Townsend Peterson, Brant C Faircloth, T Caleb Warr, and John E McCormack (2014). "Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact". In: *BMC evolutionary biology* 14.1, p. 135.

Graves, Jennifer A Marshall (2006). "Sex chromosome specialization and degeneration in mammals". In: *Cell* 124.5, pp. 901–914.

Grossenbacher, Dena L, Yaniv Brandvain, Josh R Auld, Martin Burd, Pierre-Olivier Cheptou, Jeffrey K Conner, Alannie G Grant, Stephen M Hovick, John R Pannell, Anton Pauw, et al. (2017). "Self-compatibility is over-represented on islands". In: *New Phytologist* 215.1, pp. 469–478.

Gu, Luiqi and James R Walters (2017). "Evolution of sex chromosome dosage compensation in animals: a beautiful theory, undermined by facts and bedeviled by details". In: *Genome biology and evolution* 9.9, pp. 2461–2476.

Haldane, John BS (1922). "Sex ratio and unisexual sterility in hybrid animals". In: *Journal of genetics* 12.2, pp. 101–109.

Harkess, Alex, Kun Huang, Ron van der Hulst, Bart Tissen, Jeffrey L Caplan, Aakash Koppula, Mona Batish, Blake C Meyers, and James Leebens-Mack (2020). "Sex Determination by Two Y-linked Genes in Garden Asparagus". In: *The Plant Cell*.

Harkess, Alex, Jinsong Zhou, Chunyan Xu, John E Bowers, Ron Van der Hulst, Saravanaraj Ayyampalayam, Francesco Mercati, Paolo Riccardi, Michael R McKain, Atul Kakrana, et al. (2017). "The asparagus genome sheds light on the origin and evolution of a young Y chromosome". In: *Nature communications* 8.1, pp. 1–10.

Harrison, Richard G and Erica L Larson (2014). "Hybridization, introgression, and the nature of species boundaries". In: *Journal of Heredity* 105.S1, pp. 795–809.

Heilbuth, Jana C (2000). "Lower species richness in dioecious clades". In: *The American Naturalist* 156.3, pp. 221–241.

Hesse, Elze and John R Pannell (2011). "Density-dependent pollen limitation and reproductive assurance in a wind-pollinated herb with contrasting sexual systems". In: *Journal of Ecology* 99.6, pp. 1531–1539.

Hollister, Jesse D (2015). "Polyploidy: adaptation to the genomic environment". In: *New Phytologist* 205.3, pp. 1034–1039.

Honys, David and David Twell (2004). "Transcriptome analysis of haploid male gametophyte development in Arabidopsis". In: *Genome biology* 5.11, R85.

Husband, Brian C (2004). "The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations". In: *Biological Journal of the Linnean Society* 82.4, pp. 537–546.

Husband, Brian C, Barbara Ozimec, Sara L Martin, and Lisa Pollock (2008). "Mating consequences of polyploid evolution in flowering plants: current trends and insights from synthetic polyploids". In: *International Journal of Plant Sciences* 169.1, pp. 195–206.

Jaenike, John (2001). "Sex chromosome meiotic drive". In: *Annual Review of Ecology and Systematics* 32.1, pp. 25–49.

Jeffries, Daniel L., Jörn F. Gerchen, Mathias Scharmann, and John R. Pannell (2021). "A neutral model for the loss of recombination on sex chromosomes". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1832, p. 20200096.

Jeffries, Daniel L, Guillaume Lavanchy, Roberto Sermier, Michael J Sredl, Ikuo Miura, Amael Borzée, Lisa N Barrow, Daniele Canestrelli, Pierre-André Crochet, Christophe Dufresnes, et al. (2018). "A rapid rate of sex-chromosome turnover and non-random transitions in true frogs". In: *Nature communications* 9.1, pp. 1–11.

Jiao, Yuannian, Norman J Wickett, Saravanaraj Ayyampalayam, André S Chanderbali, Lena Landherr, Paula E Ralph, Lynn P Tomsho, Yi Hu, Haiying Liang, Pamela S Soltis, et al. (2011). "Ancestral polyploidy in seed plants and angiosperms". In: *Nature* 473.7345, pp. 97–100.

Jong, Tom J de, Nickolas M Waser, and Peter GL Klinkhamer (1993). "Geitonogamy: the neglected side of selfing". In: *Trends in Ecology & Evolution* 8.9, pp. 321–325.

Joseph, Sarah B and Mark Kirkpatrick (2004). "Haploid selection in animals". In: *Trends in Ecology & Evolution* 19.11, pp. 592–597.

Käfer, Jos, Hugo J de Boer, Sylvain Mousset, Anneleen Kool, Mathilde Dufay, and GAB Marais (2014). "Dioecy is associated with higher diversification rates in flowering plants". In: *Journal of evolutionary biology* 27.7, pp. 1478–1490.

Käfer, Jos, Gabriel AB Marais, and John R Pannell (2017). "On the rarity of dioecy in flowering plants". In: *Molecular Ecology* 26.5, pp. 1225–1241.

Kaiser, Vera B and Doris Bachtrog (2010). "Evolution of sex chromosomes in insects". In: *Annual review of genetics* 44, pp. 91–112.

Kamiya, Takashi, Wataru Kai, Satoshi Tasumi, Ayumi Oka, Takayoshi Matsunaga, Naoki Mizuno, Masashi Fujita, Hiroaki Suetake, Shigenori Suzuki, Sho Hosoya, et al. (2012). "A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, Takifugu rubripes (fugu)". In: *PLoS genet* 8.7, e1002798.

Kasjaniuk, Magdalena, Aleksandra Grabowska-Joachimiak, and Andrzej J Joachimiak (2019). "Testing the translocation hypothesis and Haldane's rule in Rumex hastatulus". In: *Protoplasma* 256.1, pp. 237–247.

Kazama, Yusuke, Kotaro Ishii, Wataru Aonuma, Tokihiro Ikeda, Hiroki Kawamoto, Ayako Koizumi, Dmitry A Filatov, Margarita Chibalina, Roberta Bergero, Deborah Charlesworth, et al. (2016). "A new physical mapping approach refines the sex-determining gene positions on the Silene latifolia Y-chromosome". In: *Scientific reports* 6.1, pp. 1–10.

Kejnovsky, E, R Hobza, Tomas Cermak, Z Kubat, and B Vyskot (2009). "The role of repetitive DNA in structure and evolution of sex chromosomes in plants". In: *Heredity* 102.6, pp. 533–541.

Kent, Tyler V, Jasmina Uzunović, and Stephen I Wright (2017). "Coevolution between transposable elements and recombination". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1736, p. 20160458.

Kirkpatrick, Mark, Groves Dixon, Jason M Sardell, Manfred Schartl, and CL Peichel (2020). "Evolution of the canonical sex chromosomes of the guppy and its relatives". In: *bioRxiv*.

Krähenbühl, M., Y.-M. Yuan, and Philippe Küpfer (2002). "Chromosome and breeding system evolution of the genus Mercurialis (Euphorbiaceae): implications of ITS molecular phylogeny". In: *Plant Systematics and Evolution* 234.1-4, pp. 155–169.

Kuhl, Heiner, Yann Guiguen, Christin Hoehne, Eva Keuz, Christophe Klopp, Kang Du, Celine Lopez-Roques, Elena Santidrian Yebra-Pimentel, Mitica Ciorpac, Joern Gessner, et al. (2020). "A 180 My-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes". In: *bioRxiv*.

Lafon-Placette, Clément and Claudia Köhler (2016). "Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers". In: *Molecular Ecology* 25.11, pp. 2620–2629.

Lahn, Bruce T and David C Page (1999). "Four evolutionary strata on the human X chromosome". In: *Science* 286.5441, pp. 964–967.

Lande, Russell and Douglas W Schemske (1985). "The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models". In: *Evolution* 39.1, pp. 24–40.

Le Corre, Valérie, Mathieu Siol, Yves Vigouroux, Maud I Tenaillon, and Christophe Délye (2020). "Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe". In: *Proceedings of the National Academy of Sciences*.

Levin, Donald A (1975). "Minority cytotype exclusion in local plant populations". In: *Taxon* 24.1, pp. 35–43.

Lima, Thiago G (2014). "Higher levels of sex chromosome heteromorphism are associated with markedly stronger reproductive isolation". In: *Nature Communications* 5.1, pp. 1–8.

Lloyd, David G (1980). "The distributions of gender in four angiosperm species illustrating two evolutionary pathways to dioecy". In: *Evolution* 34.1, pp. 123–134.

Lloyd, David G (1982). "Selection of combined versus separate sexes in seed plants". In: *The American Naturalist* 120.5, pp. 571–585.

Lloyd, David G and CJ Webb (1977). "Secondary sex characters in plants". In: *The botanical review* 43.2, pp. 177–216.

Lloyd, David G and CJ Webb (1986). "The avoidance of interference between the presentation of pollen and stigmas in angiosperms I. Dichogamy". In: *New Zealand journal of botany* 24.1, pp. 135–162.

Ma, Wen-Juan, Luis Santos del Blanco, and John R Pannell (2019). "A new biological species in the Mercurialis annua polyploid complex: functional divergence in inflorescence morphology and hybrid sterility". In: *Annals of botany* 124.1, pp. 165–178.

Mable, Barbara K (2004). "Polyploidy and self-compatibility: is there an association?" In: *New phytologist* 162.3, pp. 803–811.

Madl, James E and Robert K Herman (1979). "Polyploids and sex determination in Caenorhabditis elegans". In: *Genetics* 93.2, pp. 393–402.

Malinsky, Milan, Richard J Challis, Alexandra M Tyers, Stephan Schiffels, Yohey Terai, Benjamin P Ngatunga, Eric A Miska, Richard Durbin, Martin J Genner, and George F Turner (2015). "Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake". In: *Science* 350.6267, pp. 1493–1498.

Mallet, James (2005). "Hybridization as an invasion of the genome". In: *Trends in ecology & evolution* 20.5, pp. 229–237.

Marais, Gabriel (2003). "Biased gene conversion: implications for genome and sex evolution". In: *TRENDS in Genetics* 19.6, pp. 330–338.

Miller, JS and David L Venable (2000). "Polyploidy and the evolution of gender dimorphism in plants". In: *Science* 289.5488, pp. 2335–2338.

Ming, Ray, Abdelhafid Bendahmane, and Susanne S Renner (2011). "Sex chromosomes in land plants". In: *Annual review of plant biology* 62, pp. 485–514.

Miura, Ikuo (2017). "Sex determination and sex chromosomes in Amphibia". In: *Sexual Development* 11.5-6, pp. 298–306.

Monnahan, Patrick, Filip Kolár, Pierre Baduel, Christian Sailer, Jordan Koch, Robert Horvath, Benjamin Laenen, Roswitha Schmickl, Pirita Paajanen, Gabriela Sràmkovà, et al. (2019). "Pervasive population genomic consequences of genome duplication in Arabidopsis arenosa". In: *Nature ecology & evolution* 3.3, pp. 457–468.

Moore, Emily C and Reade B Roberts (2013). "Polygenic sex determination". In: *Current Biology* 23.12, R510–R512.

Müller, Niels A, Birgit Kersten, Ana P Leite Montalvão, Niklas Mähler, Carolina Bernhardsson, Katharina Bräutigam, Zulema Carracedo Lorenzo, Hans Hoenicka, Vikash Kumar, Malte Mader, et al. (2020). "A single gene underlies the dynamic evolution of poplar sex determination". In: *Nature Plants*, pp. 1–8.

Nguyen, Alison H and Doris Bachtrog (2020). "Toxic Y chromosome: increased repeat expression and age-associated heterochromatin loss in male Drosophila with a young Y chromosome". In: *bioRxiv*.

Nordborg, Magnus (2000). "Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization". In: *Genetics* 154.2, pp. 923–929.

Obbard, Darren J, Stephen A Harris, Richard JA Buggs, and John R Pannell (2006). "Hybridization, polyploidy, and the evolution of sexual systems in Mercurialis (Euphorbiaceae)". In: *Evolution* 60.9, pp. 1801–1815.

Obbard, Darren J, Stephen A Harris, and John R Pannell (2006). "Sexual systems and population genetic structure in an annual plant: testing the metapopulation model". In: *The American Naturalist* 167.3, pp. 354–366.

Obbard, Darren J, John R Pannell, and Stephen A Harris (2006). "Mercurialis canariensis (Euphorbiaceae), a new endemic to the Canary Islands". In: *Kew Bulletin*, pp. 99–106.

Obeso, José Ramón (2002). "The costs of reproduction in plants". In: *New phytologist* 155.3, pp. 321–348.

Otto, Sarah P (2009). "The Evolutionary Enigma of Sex". In: *The American Naturalist* 174.S1, S1–S14.

Otto, Sarah P and Thomas Lenormand (2002). "Resolving the paradox of sex and recombination". In: *Nature Reviews Genetics* 3.4, pp. 252–261.

Otto, Sarah P and Jeannette Whitton (2000). "Polyploid incidence and evolution". In: *Annual review of genetics* 34.1, pp. 401–437.

Pannell, John (1997a). "The maintenance of gynodioecy and androdioecy in a metapopulation". In: *Evolution* 51.1, pp. 10–20.

Pannell, John (1997b). "Variation in sex ratios and sex allocation in androdioecious Mercurialis annua". In: *Journal of Ecology*, pp. 57–69.

Pannell, John (1997c). "Widespread functional androdioecy in Mercurialis annua L.(Euphorbiaceae)". In: *Biological Journal of the Linnean Society* 61.1, pp. 95–116.

Pannell, John R (2000). "A hypothesis for the evolution of androdioecy: the joint influence of reproductive assurance and local mate competition in a metapopulation". In: *Evolutionary Ecology* 14.3, pp. 195–211.

Pannell, John R (2002). "The evolution and maintenance of androdioecy". In: *Annual Review of Ecology and Systematics* 33.1, pp. 397–425.

Pannell, John R and Spencer CH Barrett (1998). "Baker's law revisited: reproductive assurance in a metapopulation". In: *Evolution* 52.3, pp. 657–668.

Pannell, John R and Jörn Gerchen (2018). "Sex determination: sterility genes out of sequence". In: *Current Biology* 28.2, R80–R83.

Pannell, John R, Darren J Obbard, and Richard JA Buggs (2004). "Polyploidy and the sexual system: what can we learn from Mercurialis annua?" In: *Biological Journal of the Linnean Society* 82.4, pp. 547–560.

Pannell, JR, SM Eppley, ME Dorken, and R Berjano (2014). "Regional variation in sex ratios and sex allocation in androdioecious Mercurialis annua". In: *Journal of Evolutionary Biology* 27.7, pp. 1467–1477.

Pardo-Diaz, Carolina, Camilo Salazar, Simon W Baxter, Claire Merot, Wilsea Figueiredo-Ready, Mathieu Joron, W Owen McMillan, and Chris D Jiggins (2012). "Adaptive introgression across species boundaries in Heliconius butterflies". In: *PLoS Genet* 8.6, e1002752.

Parisod, Christian, Karine Alix, Jérémy Just, Maud Petit, Véronique Sarilar, Corinne Mhiri, Malika Ainouche, Boulos Chalhoub, and Marie-Angele Grandbastien (2010). "Impact of transposable elements on the organization and function of allopolyploid genomes". In: *New Phytologist* 186.1, pp. 37–45.

Pickup, Melinda, Yaniv Brandvain, Christelle Fraisse, Sarah Yakimowski, Nicholas H Barton, Tanmay Dixit, Christian Lexer, Eva Cereghetti, and David L Field (2019). "Mating system variation in hybrid zones: facilitation, barriers and asymmetries to gene flow". In: *New Phytologist* 224.3, pp. 1035–1047.

Pilkington, SM, J Tahir, E Hilario, SE Gardiner, D Chagné, A Catanach, J McCallum, L Jesson, LG Fraser, MA McNeilage, et al. (2019). "Genetic and cytological analyses reveal the recombination landscape of a partially differentiated plant sex chromosome in kiwifruit". In: *BMC plant biology* 19.1, p. 172.

Pipoly, Ivett, Veronika Bókony, Mark Kirkpatrick, Paul F Donald, Tamás Székely, and András Liker (2015). "The genetic sex-determination system predicts adult sex ratios in tetrapods". In: *Nature* 527.7576, pp. 91–94.

Poelstra, Jelmer W, Nagarjun Vijay, Christen M Bossu, Henrik Lantz, Bettina Ryll, Inge Müller, Vittorio Baglione, Per Unneberg, Martin Wikelski, Manfred G Grabherr, et al. (2014). "The genomic landscape underlying phenotypic integrity in the face of gene flow in crows". In: *Science* 344.6190, pp. 1410–1414.

Ponnikas, Suvi, Hanna Sigeman, Jessica K Abbott, and Bengt Hansson (2018). "Why do sex chromosomes stop recombining?" In: *Trends in Genetics* 34.7, pp. 492–503.

Pujol, Benoit and John R Pannell (2008). "Reduced responses to selection after species range expansion". In: *Science* 321.5885, pp. 96–96.

Pujol, Benoit, Shu-Rong Zhou, Julia Sanchez Vilas, and John R Pannell (2009). "Reduced inbreeding depression after species range expansion". In: *Proceedings of the National Academy of Sciences* 106.36, pp. 15379–15383.

Racimo, Fernando, Sriram Sankararaman, Rasmus Nielsen, and Emilia Huerta-Sànchez (2015). "Evidence for archaic adaptive introgression in humans". In: *Nature Reviews Genetics* 16.6, pp. 359–371.

Ramsey, Justin and Douglas W Schemske (1998). "Pathways, mechanisms, and rates of polyploid formation in flowering plants". In: *Annual review of ecology and systematics* 29.1, pp. 467–501.

Renner, Susanne S (2014). "The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database". In: *American Journal of botany* 101.10, pp. 1588–1596.

Renner, Susanne S (2016). "Pathways for making unisexual flowers and unisexual plants: moving beyond the "two mutations linked on one chromosome" model". In: *American Journal of Botany* 103.4, pp. 587–589.

Renner, Susanne S and Robert E Ricklefs (1995). "Dioecy and its correlates in the flowering plants". In: *American journal of botany* 82.5, pp. 596–606.

Rice, William R (1987). "The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes". In: *Evolution* 41.4, pp. 911–914.

Rifkin, Joanna L, Felix EG Beaudry, Zoë Humphries, Baharul I Choudhury, Spencer CH Barrett, and Stephen I Wright (2021). "Widespread recombination suppression facilitates plant sex chromosome evolution". In: *Molecular biology and evolution* 38.3, pp. 1018–1030.

Ronfort, Joelle (1999). "The mutation load under tetrasomic inheritance and its consequences for the evolution of the selfing rate in autotetraploid species". In: *Genetics Research* 74.1, pp. 31–42.

Russell, JRW and JR Pannell (2015). "Sex determination in dioecious Mercurialis annua and its close diploid and polyploid relatives". In: *Heredity* 114.3, pp. 262–271.

Saetre, Glenn–Peter, Thomas Borge, Katarina Lindroos, Jon Haavie, Ben C Sheldon, Craig Primmer, and Ann–Christine Syvänen (2003). "Sex chromosome evolution and speciation in Ficedula flycatchers". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270.1510, pp. 53–59.

Sandler, George, Felix EG Beaudry, Spencer CH Barrett, and Stephen I Wright (2018). "The effects of haploid selection on Y chromosome evolution in two closely related dioecious plants". In: *Evolution letters* 2.4, pp. 368–377.

Sauquet, Hervé, Maria von Balthazar, Susana Magallon, James A Doyle, Peter K Endress, Emily J Bailes, Erica Barroso de Morais, Kester Bull-Hereñu, Laetitia Carrive, Marion Chartier, et al. (2017). "The ancestral flower of angiosperms and its early diversification". In: *Nature communications* 8.1, pp. 1–10.

Schierup, Mikkel Heide and Xavier Vekemans (2008). "Genomic consequences of selection on self-incompatibility genes". In: *Current opinion in plant biology* 11.2, pp. 116–122.

Schilthuizen, M, MCWG Giesbers, and LW Beukeboom (2011). "Haldane's rule in the 21st century". In: *Heredity* 107.2, pp. 95–102.

Schwander, Tanja, Romain Libbrecht, and Laurent Keller (2014). "Supergenes and complex phenotypes". In: *Current Biology* 24.7, R288–R294.

Scott, Michael F and Sarah P Otto (2017). "Haploid selection favors suppressed recombination between sex chromosomes despite causing biased sex ratios". In: *Genetics* 207.4, pp. 1631–1649.

Selmecki, Anna M, Yosef E Maruvka, Phillip A Richmond, Marie Guillet, Noam Shoresh, Amber L Sorenson, Subhajyoti De, Roy Kishony, Franziska Michor, Robin Dowell, et al. (2015). "Polyploidy can drive rapid adaptation in yeast". In: *Nature* 519.7543, pp. 349–352.

Shannon, Robynn K and Kent E Holsinger (2007). "The genetics of sex determination in stinging nettle (Urtica dioica)". In: *Sexual Plant Reproduction* 20.1, pp. 35–43.

She, Hongbing, Zhiyuan Liu, Zhaosheng Xu, Helong Zhang, Feng Cheng, Xiaowu Wang, and Wei Qian (2020). "The female (XX) and male (YY) genomes provide insights into the sex determination mechanism in spinach". In: *bioRxiv*.

Shifriss, Oved (1956). "Sex instability in Ricinus". In: *Genetics* 41.2, p. 265.

Slotte, Tanja (2014). "The impact of linked selection on plant genomic variation". In: *Briefings in functional genomics* 13.4, pp. 268–275.

Smith, John Maynard (1978). *The evolution of sex*. Cambridge University Press.

Smith, John Maynard (1982). *Evolution and the Theory of Games*. Cambridge University Press.

Soltis, Douglas E, Clayton J Visger, and Pamela S Soltis (2014). "The polyploidy revolution then... and now: Stebbins revisited". In: *American journal of botany* 101.7, pp. 1057–1078.

Spigler, RB, KS Lewers, DS Main, and TL Ashman (2008). "Genetic mapping of sex determination in a wild strawberry, Fragaria virginiana, reveals earliest form of sex chromosome". In: *Heredity* 101.6, pp. 507–517.

Spoelhof, Jonathan P, Rachel Keeffe, and Stuart F McDaniel (2020). "Does reproductive assurance explain the incidence of polyploidy in plants and animals?" In: *New Phytologist* 227.1, pp. 14–21.

Stebbins Jr, G Ledyard (1947). "Types of polyploids: their classification and significance". In: *Advances in genetics*. Vol. 1. Elsevier, pp. 403–429.

Stehlik, Ivana and Spencer CH Barrett (2006). "Pollination intensity influences sex ratios in dioecious Rumex nivalis, a wind-pollinated plant". In: *Evolution* 60.6, pp. 1207–1214.

Stevens, Nettie Maria (1906). *Studies in spermatogenesis: with especial reference to the "accessory chromosome". A comparative study of the heterochromosomes in certain species of coleoptera, Hemiptera and Lepidoptera, with especial reference to sex determination*. Carnegie Institution of Washington.

Suarez-Gonzalez, Adriana, Christian Lexer, and Quentin CB Cronk (2018). "Adaptive introgression: a plant perspective". In: *Biology letters* 14.3, p. 20170688.

The Tree of Sex Consortium (2014). "Tree of Sex: A database of sexual systems". In: *Scientific Data* 1.1, pp. 1–8.

Torres, Maria F, Lisa S Mathew, Ikhlak Ahmed, Iman K Al-Azwani, Robert Krueger, Diego Rivera-Nuñez, Yasmin A Mohamoud, Andrew G Clark, Karsten Suhre, and Joel A Malek (2018). "Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix". In: *Nature communications* 9.1, pp. 1–9.

Turelli, Michael and H Allen Orr (1995). "The dominance theory of Haldane's rule." In: *Genetics* 140.1, pp. 389–402.

Ubeda, Francisco, Manus M Patten, and Geoff Wild (2015). "On the origin of sex chromosomes from meiotic drive". In: *Proceedings of the Royal Society B: Biological Sciences* 282.1798, p. 20141932.

Vamosi, Jana C and Steven M Vamosi (2004). "The role of diversification in causing the correlates of dioecy". In: *Evolution* 58.4, pp. 723–731.

Vandepitte, Katrien, Hans Jacquemyn, Isabel Roldàn-Ruiz, and Olivier Honnay (2011). "The consequences of mating over a range of parental genetic similarity in a selfing allopolyploid plant species". In: *Journal of evolutionary biology* 24.12, pp. 2750–2758.

Vega-Frutis, Rocio, ROGELIO Macias-Ordonez, R Guevara, and L Fromhage (2014). "Sex change in plants and animals: a unified perspective". In: *Journal of evolutionary biology* 27.4, pp. 667–675.

Veltsos, Paris, Guillaume Cossard, Emmanuel Beaudoing, Genséric Beydon, Dessislava Savova Bianchi, Camille Roux, Santiago C Gonzàlez-Martinez, and John R Pannell (2018). "Size and content of the sex-determining region of the Y chromosome in dioecious Mercurialis annua, a plant with homomorphic sex chromosomes". In: *Genes* 9.6, p. 277.

Veltsos, Paris, Kate E Ridout, Melissa A Toups, Santiago C Gonzàlez-Martinez, Aline Muyle, Olivier Emery, Pasi Rastas, Vojtech Hudzieczek, Roman Hobza, Boris Vyskot, et al. (2019). "Early sex-chromosome evolution in the diploid dioecious plant Mercurialis annua". In: *Genetics* 212.3, pp. 815–835.

Veltsos, Paris, Nicolas Rodrigues, Tania Studer, Wen-Juan Ma, Roberto Sermier, Julien Leuenberger, and Nicolas Perrin (2019). "No evidence that Y-chromosome differentiation affects male fitness in a Swiss population of common frogs". In: *Journal of evolutionary biology*.

Vicoso, Beatriz (2019). "Molecular and evolutionary dynamics of animal sex-chromosome turnover". In: *Nature ecology & evolution*, pp. 1–10.

Villeneuve, Anne M and Kenneth J Hillers (2001). "Whence meiosis?" In: *Cell* 106.6, pp. 647–650.

Volz, Stefanie M and Susanne S Renner (2008). "Hybridization, polyploidy, and evolutionary transitions between monoecy and dioecy in Bryonia (Cucurbitaceae)". In: *American Journal of Botany* 95.10, pp. 1297–1306.

Wang, Jianping, Jong-Kuk Na, Qingyi Yu, Andrea R Gschwend, Jennifer Han, Fanchang Zeng, Rishi Aryal, Robert VanBuren, Jan E Murray, Wenli Zhang, et al. (2012). "Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution". In: *Proceedings of the National Academy of Sciences* 109.34, pp. 13710–13715.

Weiblen, George D, Ryan K Oyama, and Michael J Donoghue (2000). "Phylogenetic analysis of dioecy in monocotyledons". In: *The American Naturalist* 155.1, pp. 46–58.

Weller, Stephen G and Ann K Sakai (1999). "Using phylogenetic approaches for the analysis of plant breeding system evolution". In: *Annual Review of Ecology and Systematics* 30.1, pp. 167–199.

West, Stuart (2009). *Sex allocation*. Vol. 44. Princeton University Press.

Wood, Troy E, Naoki Takebayashi, Michael S Barker, Itay Mayrose, Philip B Greenspoon, and Loren H Rieseberg (2009). "The frequency of polyploid speciation in vascular plants". In: *Proceedings of the national Academy of sciences* 106.33, pp. 13875–13879.

Wright, Stephen I, Susan Kalisz, and Tanja Slotte (2013). "Evolutionary consequences of self-fertilization in plants". In: *Proceedings of the Royal Society B: Biological Sciences* 280.1760, p. 20130133.

Wu, Chung-I (2001). "The genic view of the process of speciation". In: *Journal of evolutionary biology* 14.6, pp. 851–865.

Xirocostas, Zoe A, Susan E Everingham, and Angela T Moles (2020). "The sex with the reduced sex chromosome dies earlier: a comparison across the tree of life". In: *Biology Letters* 16.3, p. 20190867.

Yakimowski, Sarah B and Spencer CH Barrett (2016). "The role of hybridization in the evolution of sexual system diversity in a clonal, aquatic plant". In: *Evolution* 70.6, pp. 1200–1211.

Yampolsky, Cecil (1919). "Inheritance of sex in Mercurialis annua". In: *American Journal of Botany*, pp. 410–442.

Yang, Wenlu, Zhiyang Zhang, Deyan Wang, Yiling Li, Shaofei Tong, Mengmeng Li, Xu Zhang, Lei Zhang, Liwen Ren, Xinzhi Ma, et al. (2020). "A general model to explain repeated turnovers of sex determination in the Salicaceae". In: *bioRxiv*.

Yazdi, Homa Papoli and Hans Ellegren (2014). "Old but not (so) degenerated—slow evolution of largely homomorphic sex chromosomes in ratites". In: *Molecular biology and evolution* 31.6, pp. 1444–1453.

Yu, Qingyi, Shaobin Hou, Roman Hobza, F Alex Feltus, Xiue Wang, Weiwei Jin, Rachel L Skelton, Andrea Blas, Cornelia Lemke, Jimmy H Saw, et al. (2007). "Chromosomal location and gene paucity of the male specific region on papaya Y chromosome". In: *Molecular Genetics and Genomics* 278.2, pp. 177–185.

Zhou, Qi, Jilin Zhang, Doris Bachtrog, Na An, Quanfei Huang, Erich D Jarvis, M Thomas P Gilbert, and Guojie Zhang (2014). "Complex evolutionary trajectories of sex chromosomes across bird taxa". In: *Science* 346.6215.

Zohren, Jasmin, Nian Wang, Igor Kardailsky, James S Borrell, Anika Joecker, Richard A Nichols, and Richard JA Buggs (2016). "Unidirectional diploid–tetraploid introgression among British birch trees with shifting ranges shown by restriction site-associated markers". In: *Molecular ecology* 25.11, pp. 2413–2426.

Chapter 2

Phylogenetic analysis using sex-linked, sex chromosomal and autosomal phylogenetic markers

2.1 Introduction

Compared to the autosomal genome, sex chromosomes stand out by the frequent evolution of recombination suppression and the uniparental mode of inheritance of the non-recombining segment. Both of these features result in a specific molecular signature of sex chromosome evolution. Recombination suppression in combination with its male-specific mode of inheritance can cause reduced levels of genetic diversity and a reduction in the efficacy of purifying selection on the Y chromosome. This affects the gene content on the Y chromosome through loss of functional gene copies and increased accumulation of repetitive elements (Bachtrog 2013). The evolutionary trajectories of X chromosomes can also differ from the Y chromosomes and from autosomes. X chromosomes can differ in effective population size and mutation rates and selection on recessive variants may be more effective in the heterogametic sex (Vicoso and Charlesworth 2006). Taken together, both of these effects cause an accumulation of sequence differences between the X and the Y chromosome. Like in maternally inherited mitochondrial and chloroplast genes in plants, the sex-specific mode of inheritance of sex chromosomes causes the distribution of these variants across species to recapture the genealogical history of the non-recombining region, which can differ from the history of the autosomal genome.

Such differences in ancestry at the species level can have multiple causes. One main cause can be sex-specific differences in reproductive isolation and hybridization between species. This can happen if one of the sexes is more likely to hybridize, for example through sex-specific differences in dispersal (Brunelli et al. 2010) or if hybrids of one sex are more likely to be affected by negative fitness effects, as predicted by Haldane's Rule (Haldane 1922). Another general cause for differences in ancestry between loci is incomplete lineage sorting , which is less likely to happen in regions with a low effective population size and low recombination rates (Pamilo and Nei 1988), like sex chromosomes. As a result, these regions are thought to be more likely to reflect the true species tree in comparison with autosomal loci and in the absence of introgression (Pease and Hahn 2013). In plants there can be additional processes that cause divergent ancestries of cytoplasmic genes. Horizontal gene transfer, for example between parasitic plants and their hosts (Davis and Xi 2015), often involves mitochondrial genes (Richardson and Palmer 2007) and chloroplast capture can result in divergent ancestries of chloroplast genes without leaving an autosomal signal of introgression (Stegemann et al. 2012).

On sex chromosomes, such differences in ancestry can also be caused by the process of recombination suppression. Recombination suppression between X and Y chromosomes causes divergent patterns of ancestry between the sex chromosomes, because all of the copies of the Y chromosome are descendents of a single ancestral sequence, while recombination reshuffles X-linked and pseudoautosomal sequences (Dixon et al. 2019). These patterns can be found by comparing phylogenetic trees based on genetic markers based on homologous loci located on both sex chromosomes, so-called gametologs (Garcia-Moreno and Mindell 2000), and the autosomes of multiple species. Differences in topology can inform about the timing of recombination suppression relative to the underlying speciation events. If phylogenetic markers are located on parts of the sex chromosomes that stopped recombining in a common ancestor, gametologs of different species will form monophyletic groups. In contrast, if Y-linked markers are located in regions where recombination suppression happened after a speciation event, they are expected to cluster with X-linked markers from the same species (Dixon et al. 2019). Such phylogenetic approaches helped to estimate the relative timing of recombination surpression events on sex chromosomes in a number of animal species like mammals (Marais 2003), birds (Handley et al. 2004) and snakes (Matsubara et al. 2016).

While most of these studies have focused on animal systems, phylogenetic analyses are of particular interest in plants, where dioecy has evolved repeatedly from hermaphroditism. Phylogenetic analyses can show if sex chromosomes are ancestral or the result of multiple independent transitions to dioecy in larger dioecious clades (Yang et al. 2020). In addition, phylogenetic studies have the potential to test hypotheses concerning the evolution of dioecy itself, because the canonical gynodioecy model for the evolution of

dioecy assumes that the evolution of male sterility in females predates the evolution of female sterility in males (Charlesworth and Charlesworth 1978), and in some situations this should be reflected in the phylogenetic relationships of these genes between dioecious, hermaphrodite or gynodioecious relatives (Pannell and Gerchen 2018).

Analysing the patterns of origins of sex chromosomes in dioecious clades has the potential to understand sex chromosome evolution in a broader evolutionary context. The question whether different species with separate sexes use the same or different sex chromosomes is important in the context of the repeatability of evolution. While separate sexes evolved many times in plants (Renner 2014), it is unclear to what degree this convergence is reflected in the underlying genetic architecture that determines separate sexes. In the Salicacea, homologues of a single gene have been repeatedly and independently involved in the evolution of sex chromosomes (Yang et al. 2020; Müller et al. 2020). This result suggests that at least in this clade the situation may be similar to some animal groups, where similar genes have been coopted to take over the sex determination cascade (Capel 2017), however it is unclear to what degree the same applies within and between other clades where dioecy evolved independently. In addition, such analyses could also help to better understand the evolution of recombination suppression on plant sex chromosomes. Here comparative analysis between lineages that share an ancestral sex chromosome system may distinguish whether recombination suppression evolved once in an ancestor or if non-recombining regions evolved repeatedly in the same region on the sex chromosomes. In the latter case, comparative genomic analyses of syntenic regions where recombination suppression evolved in some lineages, but not in others could help to test if specific molecular features predispose genomic regions to evolve recombination suppression.

Phylogenetic studies have also been used to study the origin of genome copies in polyploids (Oxelman et al. 2017). Distinguishing between autopolyploid and allopolyploid origins of polyploids using cytogenetic approaches can be difficult, as chromosome counts and sizes are often similar between parental species in allopolyploids. However, when individual copies of phylogenetic markers on two or more homeologs can be distinguished, phylogenetics can help to disentangle complex origins of allopolyploids. If different copies of the same locus in a polyploid individual are located on diverged branches in the resulting phylogenetic tree it is a clear sign of an allopolyploid origin, and the location of other diploid species or homeologous copies from other polyploids on the same branch indicate parental genome donors or their relatives. This approach helped to understand the history of polyploidization in species like bread wheat (Marcussen, Sandve, et al. 2014), Shephert's purse (Douglas et al. 2015) and strawberry (Tennessen, Govindarajulu, et al. 2014) and has been especially insightful in more complex cases, where the network-like relationships between high level allopolyploids can be resolved using phylogenetics (Marcussen, Heier, et al. 2015).

Most research on sex chromosome evolution has focussed on diploid plants and animals, and little is known how polyploidy and the evolution of sexual systems and sex chromosomes are related. The contrasting pattern of the prevalences of polyploidy and bisexuality in animals and plants has lead to the hypothesis that there may be a negative relationship between polyploidization and bisexuality (Muller 1925). One explanation implied that polyploidy would be less likely to evolve in organisms with sex chromosomes, because it impede the evolution of dosage compensation mechanisms on sex chromosomes (Orr 1990). However, dosage compensation mechanisms can differ strongly between animal groups (Gu and Walters 2017), and it is unlikely that this explanation applies for all animal groups (Mable 2004). In addition, a broad association between polyploidy and transitions towards dimorphic sexual systems has been discovered in plants (Glick et al. 2016), which does not support a negative association between polyploidization and the evolution of separate sexes.

Polyploidy can be found in several clades of plants, in which the evolution of sex chromosomes has been studied in dioecious lineages, like in *Salix* (Wagner et al. 2020), *Silene* (Popp and Oxelman 2007) and *Rumex* (Singh and Smith 1971). However, little research has been done in these lineages to understand how sex chromosome evolution and polyploidy are related. One example where sex chromosome evolution in polyploids was addressed is octoploid strawberries. Here, genomic analyses in combination with phylogenetics showed that sex chromosomes evolved via translocation of a sex determining region between homeologous copies of the same chromosome (Tennessen, Wei, et al. 2018). With successive translocation events, the sex determining region captured parts of adjacent sequences, which resulted in an extension of the non-recombing region. This shows that polyploidization can cause unusual dynamics when sex chromosome evolution is involved and that it can affect the evolution of the non-recombining region of the sex chromosomes.

The genus *Mercurialis* is a promising system to further understand the role of polyploidization and hybridization in the context of the evolution of sex chromosomes and transitions between sexual systems. Separate sexes are ancestral in *Mercurialis* (Krähenbühl et al. 2002), but there have been several transitions towards monoecy and androdioecy (Obbard, Harris, Buggs, et al. 2006). In addition, there appears to be an association between polyploidy and transitions between sexual systems in the annual clade, where all monoecious and androdioecious lineages are polyploids (Tab. 1.1). Polyploidization and hybridization are very common in plants, but the relevance of both of these processes for adaptive evolution and speciation has only started to be appreciated in recent years (Jiao et al. 2011; Abbott et al. 2013; Suarez-Gonzalez et al. 2018). Understanding the role of polyploidization and transitions between sexual systems in *Mercurialis* may improve our comprehension of how these processes could be involved in major evolutionary transitions in general. Crosses between annual *Mercurialis* lineages suggest that a common mechanism can determine

maleness in hybrids between dioecious, monoecious and androdioecious lineages (Russell and Pannell 2015). However, it is unclear if these are based on an ancestral sex chromosome system shared by all lineages, or if different sex determiners can cause maleness in multiple lineages. In the first case, sex chromosomes could both be involved in determining maleness in dioecious and androdioecious lineages, and may have played an important role in the transitions between these sexual systems. Identifying the origin of sex chromosomes in *Mercurialis* could thus help to understand how sex chromosomes can be involved in transitions between sexual systems.

Previous studies aimed to infer the phylogeny of subgenomes in polyploid *Mercurialis* species. The genomic internally transcribed spacer (ITS) locus, which is a commonly used genetic marker in plant phylogenetics (Alvarez and Wendel 2003), was amplified and cloned with the aim of understanding the origin of annual and perennial *Mercurialis* species (Obbard, Harris, Buggs, et al. 2006), and additional information was generated using two chloroplast markers, which can help to assign putative paternal origins to allopolyploid lineages. An additional study used the same set of phylogenetic markers, but extended the dataset to include a newly described lineage of hexaploid *M. annua* with male-like inflorescences (Ma et al. 2019).

Based on these results, a model for the evolution of annual *Mercurialis* lineages has been developed (Fig. 2.1). This model places all annual *Mercurialis* in a monophyletic clade, which is the sister clade to perennial *M. reverchonii*, *M. elliptica* and *M. tomentosa*. The origin of tetraploid *M. annua* within the annual clade could not be inferred with certainty. Only a single sequence could be cloned for the genomic ITS marker, which would indicate that it's an autopolyploid, derived from diploid *M. annua*. However, the chloroplast marker does not place tetraploid *M. annua* as a direct relative to diploid *M. annua*, which could be indicative of an allopolyploid origin of tetraploid *M. annua*, with diploid *M. annua* being the paternal genome donor. In addition, approximate Bayesian computation using an exon-capture dataset resulted in allopolyploidization as the preferred model explaining the origin of tetraploid *M. annua* (C. Roux, unpublished results). In tetraploid *M. canariensis*, two divergent ITS sequences could be cloned, one of which is closely related to diploid *M. annua* and the other belongs to an unknown and likely extinct lineage, which is placed as sister lineage to all extant annual *Mercurialis* species. This is clear evidence for an allotetraploid origin of *M. canariensis*, and chloroplast phylogenies indicated that diploid *M. annnua* is the maternal pollen donor. Two genomic ITS sequences could also be cloned from hexaploid *M. annua*, one closely related to tetraploid *M. annua* and the other related to *M. huetii*. This suggests that hexaploid *M. annua* originated from an additional allopolyploidization event between these two lineages. Evidence from the chloroplast phylogeny suggests that tetraploid *M. annua* is the maternal genome donor. Following this logic, the Y chromosome, which determines maleness in androdioecious populations of hexaploid *M.*

annua, would come from the paternal genome donor, *M. huetii*. The ITS phylogeny for newly described hexaploid *M. annua* with male-like inflorescences agrees with that of other hexaploids, which would be indicative of speciation from a single ancestral hexaploid lineage. However, the two lineages are placed in divergent clades on the chlorplast phylogeny, which could mean that they are the result of independent allopolyploid origins.

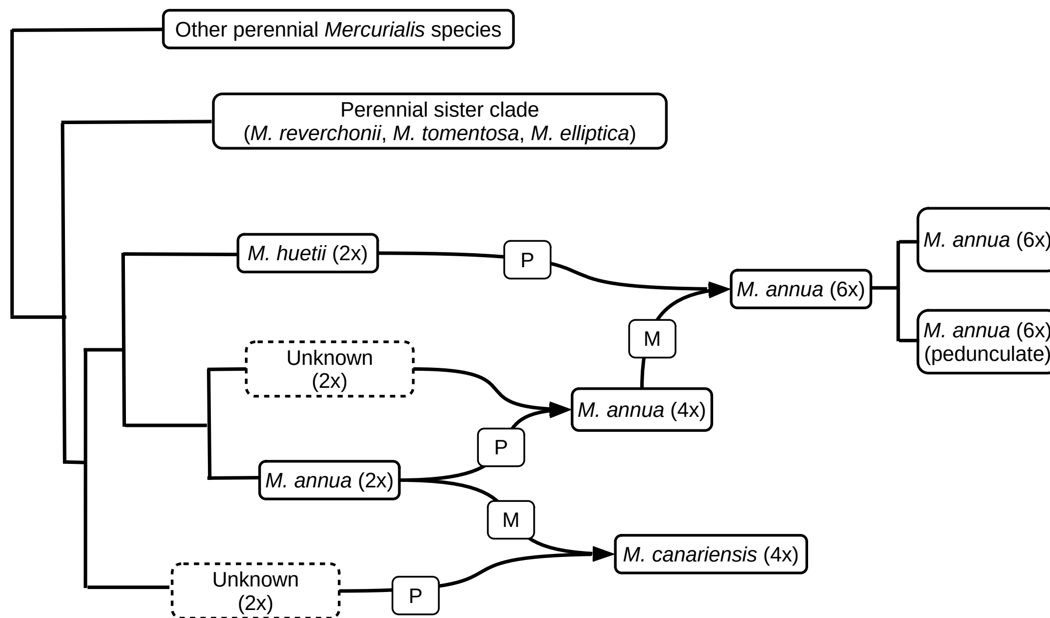


Figure 2.1: Model for the evolution of polyploid *Mercurialis* lineages (Obbard, Harris, Buggs, et al. 2006; Ma et al. 2019). Arrows indicate polyploidization events, with M indicating putative maternal and P indicating putative paternal genome donors in allopolyploids. Bars indicate homoploid speciation events.

However, these models leave many open questions, especially about how this evolutionary history relates to transitions between sexual systems. The main question is where the Y chromosome in androdioecious populations in *M. annua* comes from. Androdioecy is a rare sexual system, which is found in only a few plant species (Pannell 2002) and canonical models of sexual system evolution predict that androdioecy should be hard to evolve (Charlesworth 1984). However, metapopulation models suggest that males in androdioecious populations could be maintained by a combination of local mate competition and selection for reproductive assurance (Pannell 2000). There is strong empirical evidence that supports the role of these metapopulation processes in maintaining androdioecy in hexaploid *M. annua* (Obbard, Harris, and Pannell 2006; Eppley and Pannell 2007b), and knowing the origin of the Y chromosome will help to understand how the complex history of allopolyploidization in annual lineages of *Mercurialis* is related to the evolution of androdioecy. The assumption that the Y chromosome comes from *M. huetii* is solely based

on the fact that the chloroplast phylogeny places it it the same clade as tetraploid *M. annua*. However, this is no evidence for the origin of the Y chromosome, which could have come from an additional, so far undetected, allopolyploidization event. Also the origin of the Y chromosome in allotetraploid *M. canariensis* is not clear, because it was inferred based on similar indirect evidence. In addition, we recently discovered the presence of males in some tetraploid populations of *M. annua* in Morocco, which had been previously assumed to be exclusively monoecious (Durand 1963; Obbard, Harris, Buggs, et al. 2006). Although we do not know if maleness in tetraploid *M. annua* is determined by the same XY system of sex determination as in diploid *M. annua*, no sequences that were related to *M. huetii* were found in monoecious plants in the previous phylogenetic analysis (Obbard, Harris, Buggs, et al. 2006). Thus, an origin of the Y chromosome from allopolyploidzation with *M. huetii*, which has been proposed for hexaploid *M. annua* (Obbard, Harris, Buggs, et al. 2006), seems unlikely in tetraploid *M. annua*. This makes the origin of the Y chromosome in tetraploid *M. annua* very interesting, because it could provide evidence for the independent evolution of androdioecy in this lineage. Finally, the differences in the nuclear and chloroplast phylogenies of *M. huetii* are not explained by the current models. Such patterns could be the result of chloroplast capture, but they could also be the result of unrecognized allopolyploidization events.

In this study we address some of these open questions by developing a new set of phylogenetic markers, located on all autosomes and the sex chromosomes. In addition we present a new phylogenetic marker which shows sex-specific amplification. Using this dataset we uncover additional, previously unrecognized, allopolyploidization events in both annual and perennial *Mercurialis* species and a divergent phylogeny of the sex-linked marker, which we interpret as a sign of interspecific introgression of the Y chromosome. We propose an evolutionary scenario to explain this unusual pattern, where the Y-linked male phenotype confers a strong outcrossing advantage in populations consisting of monoecious plants.

2.2 Material and Methods

2.2.1 Development of a sex-specific PCR marker

Based on an exon-capture dataset of diploid *M. annua* (Gonzalez-Martinez et al. 2017), we previously identified 17 exonic markers with Y-linked inheritance throughout the diploid species range and confirmed them with PCR and sanger sequencing (Veltsos, Cossard, et al. 2018). We attempted to amplify them by PCR using DNA from male and female/monoecious samples of other lineages of the *M. annua* complex. Of the 17 markers, only one (g3639/gm56331), located on contig56631 of the *M. annua* genome assembly,

was successfully amplified in a male-specific fashion in all annual *Mercurialis* lineages with separate males (*M. huetii*, *M. canariensis*, androdioecious hexaploid populations; Fig. S1). While only a short (200 bp) PCR product could be successfully amplified in all lineages, this is the first evidence of a common Y chromosome in the *M. annua* species complex.

For this thesis, we developed a second exon-capture dataset, which is a subset of the first dataset and includes all exon-capture probes that aligned to linkage group 1, which has been identified as the sex chromosome (Veltsos, Cossard, et al. 2018), all probes that showed a male-specific pattern of sequencing coverage, and a subset of probes located on the other autosomal linkage groups. Two overlapping 80 bp probes were placed on each of the sites of the 120 bp probes of the initial dataset, resulting in a total of 39998 sites. This probeset was applied to a total of 128 samples of annual and perennial *Mercurialis* species and sequenced with 150 bp paired end reads using Illumina NovaSeq S4. Probe synthesis, enrichment and sequencing were done by Arbor Biosciences (Ann Arbor, USA).

We trimmed the raw exon-capture data using trimmomatic 0.38 (Bolger et al. 2014) with standard parameters, aligned the reads to v1.3 of the *M. annua* reference genome assembly (Gonzalez-Martinez et al. 2017) using BWA mem 0.7.17 (Li 2013) with standard parameters and marked PCR duplicates using the MarkDuplicates function from PicardTools 2.90 (*Picard toolkit* 2018). We then used a custom Python script to generate consensus sequences based on aligned exon-capture reads at the sex-linked contig for males from annual and perrenial lineages of *Mercurialis* by doing a pileup using the Python implementation of Samtools 1.9 (Li et al. 2009), and marking variable sites with Ns. We used these sequences to search for PCR primers using primer3plus (Untergasser, Nijveen, et al. 2007).

To test for sex specificity we amplified the markers in males and females or monoecious plants of annual and perennial *Mercurialis* species using 1x Hotstart PCR Buffer (Qiagen), 0.2 µm dNTPS, 0.2 µm forward and reverse primer and 0.001 U/µl Hotstart Taq Polymerase (Qiagen). PCR protocol was 15 min of initial denaturation at 95 °C, followed by 35 cycles of amplification with initial denaturation for 30 s at 94 °C, Annealing at 60 °C for 30 s and elongation at 73 °C for 1 min. Final elongation was at 73 °C for 5 min. Afterwards we visualized bands using gel electrophoresis and sent a subset of PCR products of males that amplified successfully for Sanger sequencing to Microsynth (Balgach, Switzerland).

We aligned the *Mercurialis* sequences using MAFFT 7.475 (Katoh et al. 2002) and did a blast search of one diploid *M. annua* sequence against the NCBI nucleotide collection (NCBI 2018) using blastn (Altschul et al. 1990). We added the sequence of the resulting best blast hit as outgroup sequence to the alignment, realigned the sequences using MAFFT and generated a phylogenetic tree by doing a full search with the generalised time reversible (GTR) model and 100 bootstrap replicates using RAxML 8.2.12 (Stamatakis

2014).

2.2.2 Development of phylogenetic markers located on autosomes and sex chromosomes

Development of phylogenetic markers in allopolyploids poses particular challenges, because there are often multiple homeologous copies of the same sequence, which have to be distinguished to infer origins of allopolyploid subgenomes (Rothfels 2021). In this study we approach this problem by making use of recent advances in longread sequencing, which allows us to obtain long reads, which span the complete amplified region of PCR products without the need for time consuming cloning (Rothfels et al. 2017).

First we used aligned exon capture sequences to find genomic regions, which were conserved between species. This allowed us to develop general PCR primer pairs, which amplify the regions in between. For this we combined data from three different exon-capture datasets. For diploid *M. annua* samples, we used the published dataset from Gonzalez-Maritinez et al. (2017), for two samples of monoecious hexaploid *M. annua* with male-like inflorescences (Ma et al. 2019) we used an additional, unpublished dataset, in which the probes that worked in the first dataset where applied to additional *Mercurialis* lineages, and for the rest of the samples we used the previously described, newly-developed exon-capture dataset. Reads were aligned against v1.3 of the *M. annua* genome assembly using the previously described pipeline. Then we used a custom Python script which identified conserved regions and placed PCR primers in them.

This script uses the Python implementation of Samtools 1.9 (Li et al. 2009), Biopython 1.73 (Cock et al. 2009) and the Python implementation of Primer3 2.3.7 (Untergasser, Cutcutache, et al. 2012). First it does a pileup and builds new reference sequences, where all positions that are not conserved are masked with Ns. In a second step Primer3 finds suitable primer pairs with at most one degenerate base in both forward and reverse primers. From the output we selected a total of 24 primer pairs, with 15 primer pairs located on linkage group 1, which is the sex chromosome (Veltsos, Ridout, et al. 2019), and one or two primer pairs on each of the autosomes (Fig. 2.2, Tab. S1).

We amplified these primer pairs in 24 samples of perennial and annual *Mercurialis* species (Tab. S2) using 1x Qiagen Hotstart PCR Buffer, 0.2 µm dNTPS, 0.2 µm forward and reverse primer and 0.001 U/µl Hotstart Taq Polymerase (Qiagen). PCR protocol was 15 min of initial denaturation at 95 °C, followed by 35 cycles of amplfication with initial denaturation for 30 s at 94 °C, Annealing at the primer-specific annealing temperature (TA, Tab. S1) for 30 s and elongation at 73 °C for 1 min. Final elongation was at 73 °C for 5 min.

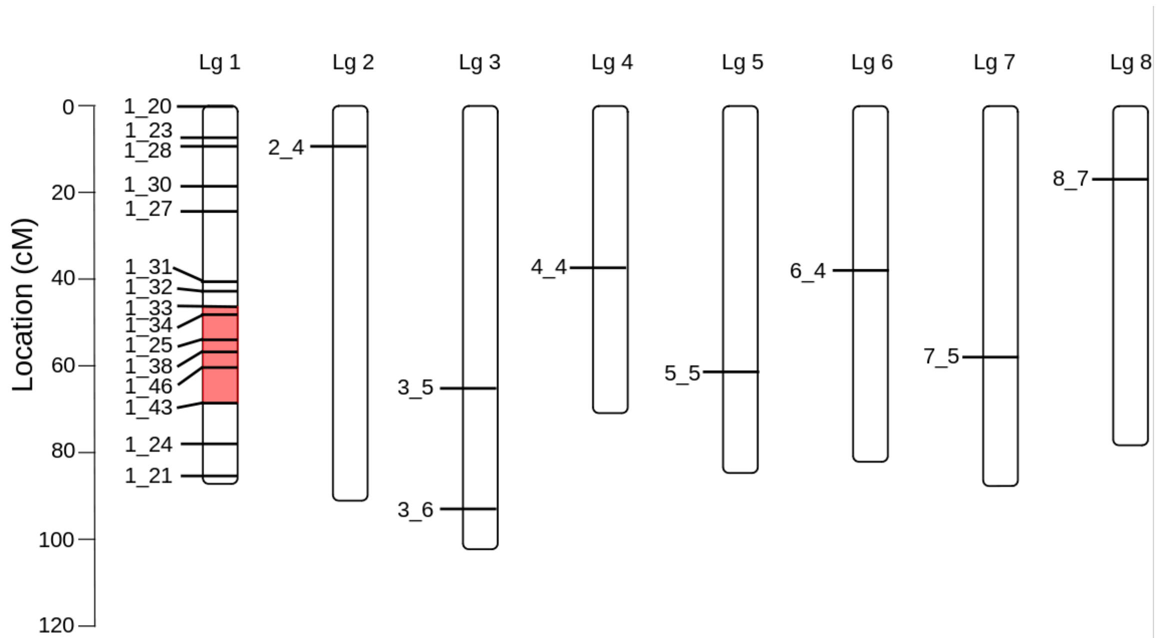


Figure 2.2: Location of primer pairs on the female linkage map of *M. annua*. The region marked in red has been determined to be the non recombining region.

After amplification, we pooled PCR products for each sample in approximately equimolar amounts estimated based on the density of bands on agarose gels. Afterwards we cleaned pooled PCR products using cleanNGS beads (Labgene) according to manufacturers instructions and quantified DNA concentrations using a Qubit fluorometer (Fisher Scientific). In a second step we used these samples to build a barcoded Oxford Nanopore sequencing library. We followed the recommendations for native barcoding using genomic DNA using the ligation sequencing kit and the native barcoding extension 1-12 and 13-24 from Oxford Nanopore: First we did an DNA repair and end prep step. We adjusted the pooled PCR products for each sample to a volume of 24 µl and added 1.75 µl of NEBNext FFPE DNA repair buffer, 1 µl of NEBNext FFPE DNA repair mix, 1.75 µl Ultra II End-prep reaction buffer and 1.5 µl Ultra II End-prep enzyme mix. We incubated this mixture at 20 °C for 5 min and then at 60 °C for 5 min. We then added 30 µl of AMPure XP beads, incubated the mixture for 5 min at room temperature, pelleted the beads on a magnetic stand, washed the sample two times with 500 µl of 70% Ethanol and eluted the cleaned DNA in 13 µl of DNAse free water. Next we ligated an individual bardcode from the Native barcoding extensions 1-12 and 13-24 to each sample. For each sample we added 1 µl of an individual barcode solution and 10 µl of Blunt TA/Ligase Master mix (NEB) to 9 µl of DNA, incubated it for 10 minutes at room temperature and cleaned it using AMPure XP beads as described in the previous step. Afterwards samples were quantified using Qubit, and pooled with volumes adjusted according to DNA concentrations and ploidy. For ligation of sequencing adaptors, about 100 pmol of library were diluted to 65 µl in water and 5 µl Adaptor Mix II, 20 µl NEBNext quick ligation buffer and 10 µl Quick T4 ligase were added and incubated at room

temperature for 10 min. The library was purified using 50 µl AMPure XP beads, which were washed twice using 250 µl short fragment buffer and resuspended in 15 µl elution buffer. Afterwards, about 100 fmol of the final library, 34 µl sequencing buffer and 25.5 µl were loaded onto an Oxford Nanopore 9.4.1 Flowcell and sequenced. Raw reads were basecalled and demultiplexed using Guppy 4.0.15 (Oxford Nanopore).

In polyploid samples, the resulting long reads represent a mixture of alleles from different homeologs. Due to the high error rates of Oxford Nanopore reads, final alignments have to be based on a consensus of multiple reads representing unique template sequences, which requires clustering of raw reads. This problem has been addressed by the PURC pipeline, which was developed for doing polyploid phylogenetics using Pacific Biosciences long reads (Rothfels et al. 2017). However, this pipeline requires optimization of multiple parameters for each marker, which becomes unfeasible for the larger number of markers used in this study and we developed a new approach, which clusters reads at each sample and locus. The main parameters are the ploidy of the sample, p, the minimal proportion of reads supporting a variant, m, which is based on sample ploidy, and the number of reads at the locus, c. We used Minimap2 2.17-r941 (Li 2018) to align longreads against the *M. annua* genome assembly. Then we used a custom Python script to cluster reads and generate consensus sequences for alignments. For this, we first did a pileup using the Python implementation of Samtools 1.9 (Li et al. 2009) and identified heterozygous variants at each site based on whether the relative proportion of each nucleotide is larger than m. We then built a distance matrix based on the relative proportion of shared variants between all reads and clustered reads using kmeans clustering implemented in scikitlearn 0.23.2 (Pedregosa et al. 2011) with the number of clusters equal to sample ploidy. This approach distinguishes diverged sequences well, but it tends to build additional small clusters of reads, which have either multiple sequencing errors at variable sites or which are chimeric reads, which often get produced in small proportions during PCR. To address this, we discarded all clusters that have a smaller number of reads than $c*m$. We built consensus sequences using freebayes 1.3.1 (Garrison and Marth 2012) to call variants on the final clusters with ploidy set to one and then modified the original sequence from the *M. annua* genome assembly based on these variant calls. Finally, we concatenated consensus sequences for all samples at each locus and built phylogenetic trees using RaxML 8.2.12 (Stamatakis 2014) by doing a full search with the GTR model and 100 bootstrap replicates and analyzed and plotted phylogenies using the online tool iTOL (Letunic and Bork 2019).

2.3 Results

2.3.1 Sex-specific PCR marker

We developed individual primer pairs for diploid *M. annua*, *M. huetii*, tetraploid *M. canariensis* and a combined primer pair for hexaploid *M. annua*, *M. reverchonii*, *M. elliptica* and *M. tomentosa*. These primer pairs amplified only in male samples and never in females or monoecious plants, including two hexaploid monoecious plants with male like inflorescences. Amplification was sucessful in all males of diploid *M. annua*, hexaploid *M. annua* and *M. tomentosa*. No or very weak amplification was observed in a minority of male samples of *M. huetii* (1 sample), *M. canariensis* (3 samples), *M reverchonii* (2 samples) and *M. elliptica* (1 sample). In tetraploid *M. annua* no primer pair amplified in any male or monoecious samples (Fig. S2).

The alignment of sequenced PCR amplicons had a total length of 707 bp shared between all *Mercurialis* sequences and it included two deletions and one insertion which were specific to hexaploid *M. annua*, *M. reverchonii*, *M. elliptica* and *M. tomentosa* and absent in the other annual lineages (Fig. S3). The best blast hit (E=1e-117, 73% identity) of the male-specific *Mercurialis* sanger sequence was a mRNA (XM_002528314.3) from *Ricinus communis*, with its function predicted as serine-rich adhesin for platelets. The resulting phylogenetic tree placed the sex-linked sequence of hexaploid *M. annua* in one clade with perennial *Mercurialis* species, while the other annual lineages formed a distinct clade (Fig. 2.3).

2.3.2 Phylogenetic markers on autosomes and sex chromosomes

We generated a total 2.2 gb of Oxford nanopore longreads. Upon visual inspection of alignments of demultiplexed reads to the *M. annua* genome, we noticed that a significant proportion of reads for primer pair 1_24 were much shorter than expected based on the location of primers at the *M. annua* genome assembly. We interpreted this as a result of non-specific PCR amplification and excluded this marker from further analyses. For the remaining markers, we excluded a total of 13 (2.4%) low coverage samples, which had a sequencing depth smaller than ten times its ploidy at the target locus.

Many of the phylogenetic trees based on markers located on the autosomes and sex chromosomes had low bootstrap support at the nodes defining the deeper relationships between perennial and annual lineages of *Mercurialis*. However, there were several clades that were well supported in the majority of trees. We coloured these clades in the plotted trees (Fig. S4). All sequences of *M. perennis* formed a

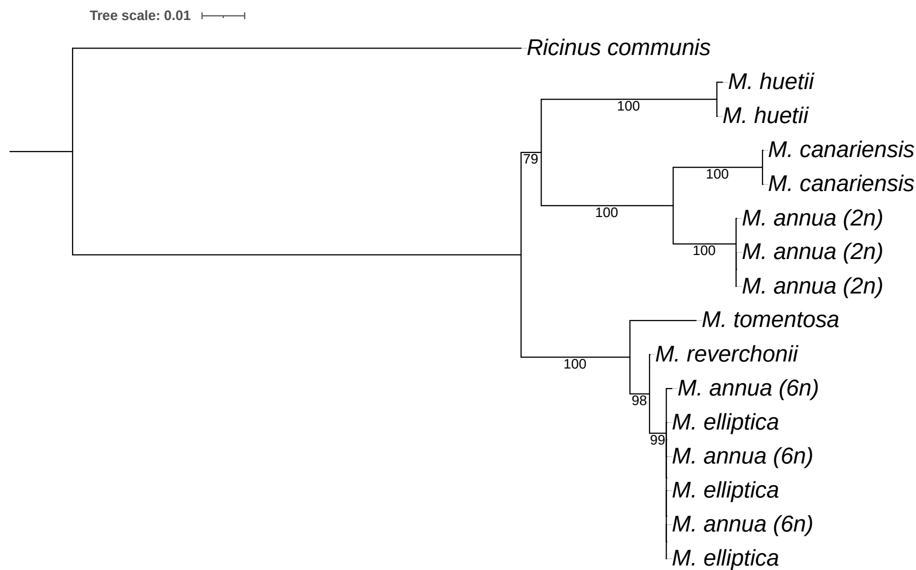


Figure 2.3: Maximum-likelihood phylogeny of the sex-linked PCR marker in males of annual and perennial *Mercurialis* species. Numbers below branches indicate bootstrap support.

single monophyletic group, except for a single sequence for marker 1_22. Based on previous analysis, which placed *M. perennis* as basal relative to our focal lineages, we rooted our phylogenies on the branch of this perennis clade (coloured red in Fig. S4). Sequences of annual *Mercurialis* lineages formed three distinct clades. The *huetii* clade (coloured blue in Fig. S4) contained sequences from *M. huetii*, and hexaploid *Mercurialis* lineages and formed a well supported clade in 13 trees. In addition, there were five trees located on linkage group 1 in which only a single male sequence fell outside the huetii clade, which was otherwise monophyletic. In both of these cases, in 14 trees this clade also contained sequences of *M. elliptica*, which were either nested within the clade or formed a monophyletic group with it. The *annua* clade (coloured green) contained sequences of diploid *M. annua*, tetraploid *M. annua*, hexaploid *M. annua* and *M. canariensis*. It formed a monophyletic group in 12 trees, and in four cases it was paraphyletic with sequences of either *M. huetii* or perennial lineages nested within it. The *canariensis* clade (coloured orange), which contained sequences from *M. canariensis*, tetraploid *M. annua* and hexaploid *M. annua* lineages, formed a monophyletic group in 17 trees. In addition, in one case it was paraphyletic with perennial lineages nested within it. Perennial *M. reverchonii*, *M. elliptica* and *M. tomentosa* were represented by two distinct monophyletic clades, each of which contained sequences from all three lineages, in 15 trees. In *M. elliptica*, additional sequences were found, which either fell within the *M. huetii* clade or formed a monophyletic group with it. All of these clades were coloured purple in Fig.S4.

While these clades were well suppported in a majority of trees, their deeper relationships relative to each other were much less clear. In many cases the nodes defining these relationships had low bootstrap support and in cases where relationships were well supported they were often contradictory between loci. There were three likely scenarios for the relationships between the annual *huetii*, *canariensis* and *annua* clades (Fig. 2.4). Scenario a, which placed the *canariensis* clade as outgroup to the *huetii* and *annua* clades and was also supported by the ITS marker (Obbard, Harris, Buggs, et al. 2006; Ma et al. 2019), had the most support, as there was a well supported monophyletic clades containing all samples from the *annua* and *huetii* clade and none from the *canariensis* clade in 5 loci.

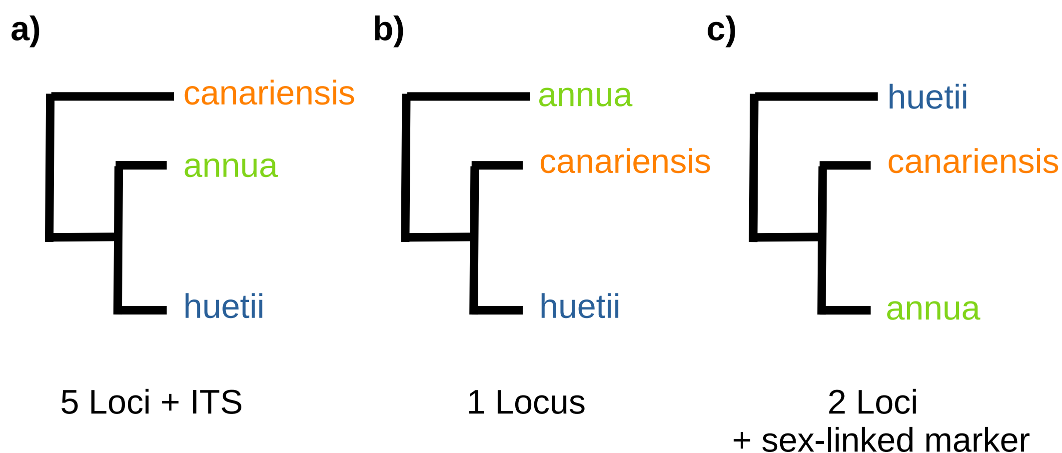


Figure 2.4: Potential relationships between the three main clades of annual *Mercurialis* lineages based on different phylogenetic datasets. The number of loci under each topology indicates the number of loci from the 23 newly developed genetic markers located on the sex chromosomes and autosomes that support them. ITS indicates the topology supported by the ITS locus (Obbard, Harris, Buggs, et al. 2006; Ma et al. 2019), sex-linked marker indicates the topology of the male-specific phylogenetic marker developed in this study.

For the two clades containing perennial *M. reverchonii*, *M. elliptica* and *M. tomentosa* also multiple scenarios were possible. Similar to the ITS phylogeny, one of the clades was supported as being the outgroup of the annual lineages and the other perennial clade was supported by 5 loci. The location of the second clade within the perennial lineage varied substantially between loci, the most common pattern was that this clade formed a monophyletic group with the *huetii* clade.

2.4 Discussion

The male-specific pattern of amplification of the newly-developed sex-linked marker in annual and perennial lineages of *Mercurialis* (Fig. S1, Fig. S2) is evidence for a common XY sex determination system. The only lineage in which the sex-specific locus had neither sequencing coverage in the exon-capture data nor showed any amplification with sequencing primers was tetraploid *M. annua* (Fig. S2). One possibility is that tetraploid *M. annua* share the same Y chromosome as the other lineages, on which the region that contains the sex-linked marker is missing. Alternatively, males could have evolved independently in this lineage. A first approach to test this would be to cross these males into other lineages to see if maleness is inherited in a Mendelian fashion.

We also generated additional phylogenetic markers located on the sex chromosomes and autosomes (Fig. 2.2, Tab. S1). Our aim was to generate phylogenetic markers, which can distinguish homeologous sequences in polyploids and use them to generate phylogenies. We generated our phylogenetic markers by identifying conserved regions in an exon-capture dataset, where we placed PCR primers. Our preliminary tests showed that we could only generate PCR primers that had good amplification success across lineages when there was continuous sequencing coverage between prime-regions in all samples. The reasons for this limitation are not completely clear, but they may involve variation in the content of repetitive elements in intronic regions between exon-capture probes or even structural rearangements relative to the diploid *M. annua* reference assembly against which we aligned our sequencing reads. As a result, many of our phylogenetic markers were limited in length (Tab. S1). This may have limited the amount of phylogenetically informative polymorphism, which likely contributed to the limited bootstrap support at deeper nodes and to the disagreement between some nodes with high bootstrap support (Fig. S4). Such disagreements could be either the result of homoplasy and incomplete lineage sorting. Alternatively, differences between loci at deeper nodes could represent biologically meaningful differences in ancestry across the genome, which could be caused by introgression or gene conversion. Since we cannot clearly distinguish between these hypotheses, we will focus mostly on the relationships within clades that have high bootstrap support in a majority of phylogenetic markers. We also propose hypotheses about the deeper relationships between these clades, but we want to stress that they should be interpreted with caution. Our results show that unambiguosly resolving phylogenetic relationships in polyploids remains challenging. Approaches that could help to improve phylogenetic inference in these scenarios could use longread sequencing of larger genomic fragments than those in our study, which were limited by the length of amplicons, for example by using whole genome sequencing data, or by applying them to exon-capture datasets. Alternatively, phylogenetic approaches could be developed that explicitely take allopolyploidization

into account and estimate phylogenies based on genomewide polyploid SNP data without requiring phasing of homeologous haplotypes (Rothfels 2021).

Despite these limitations, our newly-developed phylogenetic markers substantially change our model of the origin of annual and perennial *Mercurialis* species (Fig. 2.5). We found additional sequences in perennial *M. reverchonii* and *M. tomentosa* and in *M. elliptica*. All three species are found in two monophyletic clades. We think that the most likely scenario involved a single alloploidization event, which happened in a common ancestor prior to speciation. The additional sequence found in *M. elliptica*, would be the result of an additional allopolyploidization event between an allotetraploid progenitor to *M. elliptica* and a diploid species closely related to *M. huetii*.

We also showed that two diverged sequences are found in tetraploid *M. annua*, which are closely related to sequences also found in hexaploid lineages of *M. annua*, but also to both sequences found in allotetraploid *M. canariensis*. This data is consistent with an allopolyploid origin of tetraploid *M. annua*, similar to that of *M. canariensis*. One implication of this additional allopolyploidization event concerns genome size evolution in tetraploid *M. canariensis*. The genome size of *M. canariensis* has been estimated to be 25% larger than that of tetraploid *M. annua* (Obbard, Harris, Buggs, et al. 2006). Our new phylogenetic analyses places the ancestral genome donors of tetraploid *M. annua* and *M. canariensis* in the same clades. This makes it likely that this increase in genome size happened after the allopolyploidization event in *M. canariensis*, probably as a result of the expansion of repetitive elements (Michael 2014). Like in the previous model, the hexaploid lineages of *M. annua* could be the result of an additional allopolyploidization event between tetraploid *M. annua* and *M. huetii*.

Despite the limited capacity of our dataset to resolve deeper relationships between clades, we discovered clear evidence of hitherto unsuspected allopolyploidization events. Previous analyses missed allopolyploidization events that affected 6 out of 7 polyploid lineages of interest. This shows that the ITS locus, which is a commonly used genomic marker in plant phylogenetics (Alvarez and Wendel 2003), is not suitable for reliable inference of the origin of polyploids, because it tends to show patterns of concerted evolution in allopolyploids (Wendel et al. 1995), which, among other factors, can confound phylogenetic inference (Alvarez and Wendel 2003).

Our data also sheds new light on the origin of Y chromosomes in males of *M. canariensis* and hexaploid *M. annua*. Based on the fact that five genomic loci as well as the ITS locus agreed with scenario a, we determined that it the most support for the origin of the subgenomes found in the three main clades of annual *Mercurialis* lineages (Fig. 2.4). This scenario disagrees with the phylogeny of the sex-linked marker (Fig. 2.3), which would imply that the Y chromosome in *M. canariensis* comes from the *annua* clade. This

in turn seems to disagree with the previous inference of diploid annua also being the maternal pollen donor of *M. canariensis* (Obbard, Harris, Buggs, et al. 2006). However, in the chloroplast tree *M. huetii* is not placed in the annual clade, which means that the sistergroup relation of the annual and canariensis clades are insufficient to infer maternal origins. We found a divergent phylogeny of the sex-specific PCR marker, where the sequence of hexaploid *M. annua* is nested within the perennial clade, with *M. elliptica* being its closest relative, a pattern which is confirmed by two shared deletions and one insertion (Fig. S3). There is no evidence for such a pattern in any of the other markers located on autosomes and sex chromosomes. We therefore think that the most likely explanation for this pattern is introgression of the sex determining region of the Y chromosome from one of the perennial lineages (most likely a close relative of *M. elliptica*) into hexaploid *M. annua*.

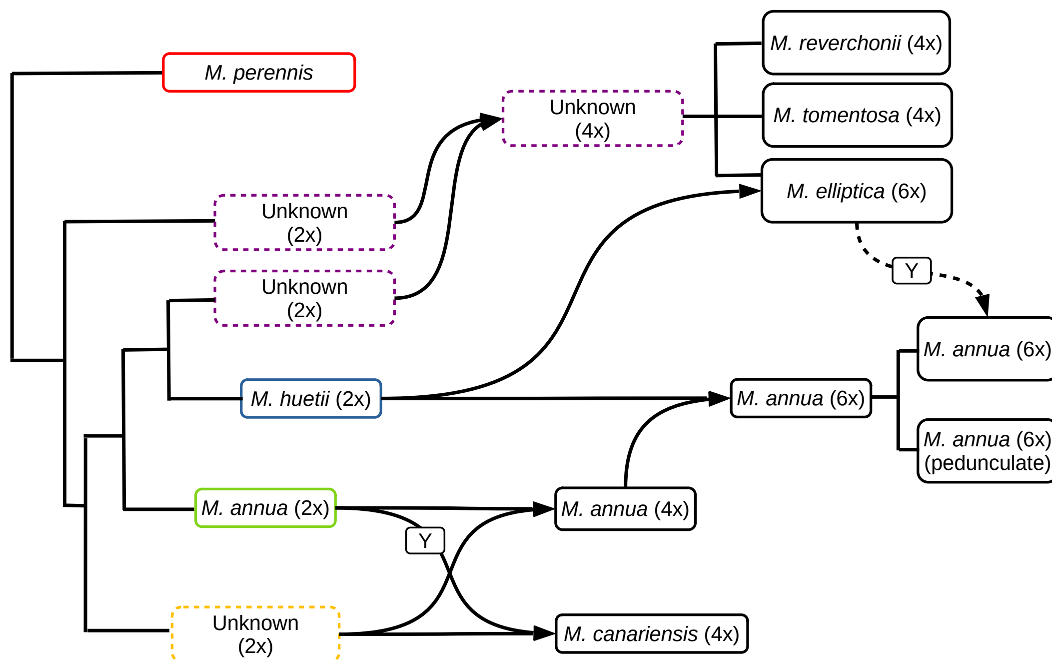


Figure 2.5: New model for the evolution of polyploid *Mercurialis* lineages. Arrows indicate polyploidization events, dotted arrow indicates Y chromosome introgression. Lines indicate homoploid speciation events.

These new results also have implications for our understanding of transitions of sexual systems in *Mercurialis*. One previous hypothesis was that either hybridization or polyploidization itself could have resulted in transitions towards monoecy in polyploid lineages of *M. annua* (Pannell, Obbard, et al. 2004). Here we showed that tetraploid *M. annua* does have an allotetraploid origin, which is similar to that of *M. canariensis*. However, despite this similar evolutionary history, there has been no transition towards monoecy in dioecious *M. canariensis*, which makes it unlikely that polyploidization or hybridization itself were the

main cause of the transition to monoecy in tetraploid *M. annua*. Besides the effect of polyploidization, a recent experimental evolution setup shows that transitions between sexual systems in *Mercurialis* can happen quickly in diploids as well (Cossard et al. 2021), which further suggests that transitions towards monoecy may not have been related to changes in ploidy, but rather be caused by selection for reproductive assurance under pollen limitation.

In addition to these new allopolyploidization events, we discovered a pattern of introgression at a locus which shows a clear pattern of male linkage in annual and perennial *Mercurialis* lineages. No such pattern could be found at any other of the markers located on autosomes and sex chromosomes. Our interpretation is that this is the result of a hybridization event between a male from one of the perennial lineages (likely *M. elliptica*) and a monoecious hexaploid *M. annua*, followed by backcrossing with hexaploid *M. annua*. As far as we can detect with our data, only the sex determining region of the perennial Y chromosome was retained in hexaploid *M. annua*. Such a pattern is surprising, because the Y chromosome is thought to play a disproportional role in the evolution of reproductive isolation (Qvarnström and Bailey 2009).

There are few examples where comparable patterns of sex chromosome introgression are found. In the European house mouse hybrid zone unidirectional introgression of the Y chromosome of the eastern house mouse *Mus musculus musculus* into the range of the western house mouse *M. musculus domesticus* has been found across large parts of Europe (Macholan et al. 2019). Recent research showed negative effects of having an introgressed Y chromosome on several fitness related traits in both species (Bimova et al. 2020). However, *M. musculus domesticus* males that carry an introgressed *M. musculus musculus* Y chromosome showed increased sperm quality, agressiveness and shorter generation cycles, which could be a sufficient fitness advantage to overcome other negative fittness effects and promote introgression. A second example comes from the ninespine stickleback, *Pungitius pungitius*. Here, a neo-Y chromosome originated from introgression from the Amur stickleback, *Pungitius sinensis* (Dixon et al. 2019). However, this case is different from the one found in *Mercurialis*, because the introgressed region was autosomal in the Amur stickleback and only acquired sex determining function after it introgressed, which means that it was not expected to show any specific patterns of sex chromosome evolution, which are thought to contribute to reproductive isolation. A pattern of increased introgression has also been found on plant sex chromosomes, between *Populus alba* and *Populus tremula*, but the data does not unambiguosly show if the signal also affects the sex determining region of the W chromosome or only the pseudoautosomal region immediately surounding it (Stölting et al. 2013).

In the case of *Mercurialis*, the signal of introgression was found at a single Y-linked marker, but not on any of the markers located along the length of the *M. annua* sex chromosome, which indicates that the

introgressed region is smaller than the region of repressed recombination in diploid *M. annua*. Evidence from crosses in animals showed an increase of reproductive isolation with increased sex chromosome differentiation (Lima 2014), and a small sex determining region could have a limited effect on reproductive isolation in *Mercurialis* as well. Another factor that may have facilitated gene flow between hexaploid *M. annua* and *M. elliptica* in general could be similar ploidies and shared allopolyploidization events. The sex-linked sequence in hexaploid *M. annua* is most similar to *M. elliptica*. The number of chromosomes differs between hexaploid *M. annua* (2*n*=48) and *M. elliptica* for which two different chromosome counts were published (2*n*=220, Krähenbühl et al. 2002, 2*n*=42 Obbard, Harris, Buggs, et al. 2006). A chromosome number of 2*n*=42 would be concordant with an allopolyploidization event between an allotetraploid ancestor of *M. elliptica* (2*n*=26) and *M. huetii* (2*n*=16), which would make *M. elliptica* allohexaploid. This is supported by the fact that the sequences of both *M. elliptica* and hexaploid *M. annua* from our study fall into three diverged clades in both lineages, which is indicative of allohexaploid origins. In addition, both species share an allopolyploidization event with *M. huetii* or a close relative, which could further facilitate hybridization.

The main difference between hexaploid *M. annua* and most other sex chromosome systems is the presence of monoecious plants instead of females. This means that the Y chromosome likely introgressed into populations where no other males and Y chromosomes were present, and in which seeds were predominantly produced by selfing. If such populations became more outcrossed, then male immigrants might have a large siring advantage, and their Y chromosome could spread. Such immigrants could come from a different species through hybridization, and backcrossing with selection for maleness would retain the SDR only. In this case, there would be no selection required to exlain the replacement of the old Y chromosome with the newly introgressed one, but a selective advantage of maleness itself would be sufficient to explain the spread of an introgressed Y chromosome. This advantage would just have to be great enough to overcome potential initial fitness definciencies of hybrids. Previous studies showed that males indeed have a strong outcrossing advantage over monoecious plants in hexaploid *M. annua*, due to their strongly increased pollen production (Pannell 1997) and because of their specialized male inflorescences, which in itself confers an outcrossing advantage over monoecious plants of approximately 60% per pollen grain produced (Eppley and Pannell 2007a). Additional fitness advantages could be conferred by avoidance of inbreeding depression through increased outcrossing by males, which has been found to be stronger in the North-African populations, which is likely the origin of hexaploid *M .annua* (Pujol et al. 2009). It would be interesting to test the degree of reproductive isolation by crossing male *M. elliptica* into monoecious hexaploid *M. annua* plants and see if these crosses result in viable and fertile male offspring.

Sex chromosome introgression would be an extension of the model of the evolution of androdioecy in a metapopulation (Pannell 2000). In this model, selection for reproductive assurance during frequent extinction and recolonization of patches would favor the establishments of cosexual individuals. Males could invade these cosexual populations after local mate competition caused selection for female biased sex allocation in monoecious populations. This would increase the outcrossing rates of invading males and give them a selective advantage. Importantly, this model assumes that males persist somewhere in the metapopulation at all times. In contrast, introgression of a Y chromosome from a different species would allow males to become reestablished after they have been lost, which could happen during an early phase when individuals are present at low densities. Another system where the evolution of androdioecy via introgression seems possible is *Sagittaria latifolia*, where transitions to androdioecy are associated with signals of introgression (Yakimowski and Barrett 2016). However, so far no evidence from sex-linked markers is available for this system.

References

Abbott, Richard, Dirk Albach, Stephen Ansell, Jan W Arntzen, Stuart JE Baird, Nicolas Bierne, Jenny Boughman, Alan Brelsford, C Alex Buerkle, Richard Buggs, et al. (2013). "Hybridization and speciation". In: *Journal of evolutionary biology* 26.2, pp. 229–246.

Altschul, Stephen F, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman (1990). "Basic local alignment search tool". In: *Journal of molecular biology* 215.3, pp. 403–410.

Alvarez, IJFW and Jonathan F Wendel (2003). "Ribosomal ITS sequences and plant phylogenetic inference". In: *Molecular phylogenetics and evolution* 29.3, pp. 417–434.

Bachtrog, Doris (2013). "Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration". In: *Nature Reviews Genetics* 14.2, pp. 113–124.

Bimova, Barbora Voslajerova, Milov Macholan, L'udovit Dureje, Katerina Berchova Bimova, Iva Martincova, and Jaroslav Pialek (2020). "Sperm quality, aggressiveness and generation turnover may facilitate unidirectional Y chromosome introgression across the European house mouse hybrid zone". In: *Heredity* 125.4, pp. 200–211.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30.15, pp. 2114–2120.

Brunelli, Joseph P, Craig A Steele, and Gary H Thorgaard (2010). "Deep divergence and apparent sex-biased dispersal revealed by a Y-linked marker in rainbow trout". In: *Molecular phylogenetics and evolution* 56.3, pp. 983–990.

Capel, Blanche (2017). "Vertebrate sex determination: evolutionary plasticity of a fundamental switch". In: *Nature Reviews Genetics* 18.11, p. 675.

Charlesworth, Brian and Deborah Charlesworth (1978). "A model for the evolution of dioecy and gynodioecy". In: *The American Naturalist* 112.988, pp. 975–997.

Charlesworth, D (1984). "Androdioecy and the evolution of dioecy". In: *Biological Journal of the Linnean Society* 22.4, pp. 333–348.

Cock, Peter JA, Tiago Antao, Jeffrey T Chang, Brad A Chapman, Cymon J Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, et al. (2009). "Biopython: freely available Python tools for computational molecular biology and bioinformatics". In: *Bioinformatics* 25.11, pp. 1422–1423.

Cossard, Guillaume G, Jörn F Gerchen, Xinji Li, Yves Cuenot, and John R Pannell (2021). In: *Current Biology* 31.6, pp. 1277–1283.

Davis, Charles C and Zhenxiang Xi (2015). "Horizontal gene transfer in parasitic plants". In: *Current Opinion in Plant Biology* 26, pp. 14–19.

Dixon, Groves, Jun Kitano, and Mark Kirkpatrick (2019). "The origin of a new sex chromosome by introgression between two stickleback fishes". In: *Molecular biology and evolution* 36.1, pp. 28–38.

Douglas, Gavin M, Gesseca Gos, Kim A Steige, Adriana Salcedo, Karl Holm, Emily B Josephs, Ramesh Arunkumar, J Arvid Agren, Khaled M Hazzouri, Wei Wang, et al. (2015). "Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid Capsella bursa-pastoris". In: *Proceedings of the National Academy of Sciences* 112.9, pp. 2806–2811.

Durand, Bernard (1963). *Le complexe Mercurialis annua LSL: une étude biosystématique*. Vol. 12, pp. 579–736.

Eppley, Sarah M and John R Pannell (2007a). "Density-dependent self-fertilization and male versus hermaphrodite siring success in an androdioecious plant". In: *Evolution* 61.10, pp. 2349–2359.

Eppley, Sarah M and John R Pannell (2007b). "Sexual systems and measures of occupancy and abundance in an annual plant: testing the metapopulation model". In: *The American Naturalist* 169.1, pp. 20–28.

Garcia-Moreno, Jaime and David P Mindell (2000). "Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): a case study using avian CHD". In: *Molecular biology and evolution* 17.12, pp. 1826–1832.

Garrison, Erik and Gabor Marth (2012). "Haplotype-based variant detection from short-read sequencing". In: *arXiv preprint arXiv:1207.3907*.

Glick, Lior, Niv Sabath, Tia-Lynn Ashman, Emma Goldberg, and Itay Mayrose (2016). "Polyploidy and sexual system in angiosperms: Is there an association?" In: *American journal of botany* 103.7, pp. 1223–1235.

Gonzalez-Martinez, Santiago C, Kate Ridout, and John R Pannell (2017). "Range expansion compromises adaptive evolution in an outcrossing plant". In: *Current Biology* 27.16, pp. 2544–2551.

Gu, Luiqi and James R Walters (2017). "Evolution of sex chromosome dosage compensation in animals: a beautiful theory, undermined by facts and bedeviled by details". In: *Genome biology and evolution* 9.9, pp. 2461–2476.

Haldane, John BS (1922). "Sex ratio and unisexual sterility in hybrid animals". In: *Journal of genetics* 12.2, pp. 101–109.

Handley, Lori-Jayne Lawson, Helene Ceplitis, and Hans Ellegren (2004). "Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution". In: *Genetics* 167.1, pp. 367–376.

Jiao, Yuannian, Norman J Wickett, Saravanaraj Ayyampalayam, André S Chanderbali, Lena Landherr, Paula E Ralph, Lynn P Tomsho, Yi Hu, Haiying Liang, Pamela S Soltis, et al. (2011). "Ancestral polyploidy in seed plants and angiosperms". In: *Nature* 473.7345, pp. 97–100.

Katoh, Kazutaka, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata (2002). "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform". In: *Nucleic acids research* 30.14, pp. 3059–3066.

Krähenbühl, M., Y.-M. Yuan, and Philippe Küpfer (2002). "Chromosome and breeding system evolution of the genus Mercurialis (Euphorbiaceae): implications of ITS molecular phylogeny". In: *Plant Systematics and Evolution* 234.1-4, pp. 155–169.

Letunic, Ivica and Peer Bork (2019). "Interactive Tree Of Life (iTOL) v4: recent updates and new developments". In: *Nucleic acids research* 47.W1, W256–W259.

Li, Heng (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM". In: *arXiv preprint arXiv:1303.3997*.

Li, Heng (2018). "Minimap2: pairwise alignment for nucleotide sequences". In: *Bioinformatics* 34.18, pp. 3094–3100.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin (2009). "The sequence alignment/map format and SAMtools". In: *Bioinformatics* 25.16, pp. 2078–2079.

Lima, Thiago G (2014). "Higher levels of sex chromosome heteromorphism are associated with markedly stronger reproductive isolation". In: *Nature Communications* 5.1, pp. 1–8.

Ma, Wen-Juan, Luis Santos del Blanco, and John R Pannell (2019). "A new biological species in the Mercurialis annua polyploid complex: functional divergence in inflorescence morphology and hybrid sterility". In: *Annals of botany* 124.1, pp. 165–178.

Mable, Barbara K (2004). "Polyploidy and self-compatibility: is there an association?" In: *New phytologist* 162.3, pp. 803–811.

Macholan, Milos, Stuart JE Baird, Alena Fornuskova, Iva Martincov, Pavel Rubik, Ludovit Dureje, Emanuel Heitlinger, and Jaroslav Pialek (2019). "Widespread introgression of the Mus musculus musculus Y chromosome in Central Europe". In: *bioRxiv*.

Marais, Gabriel (2003). "Biased gene conversion: implications for genome and sex evolution". In: *TRENDS in Genetics* 19.6, pp. 330–338.

Marcussen, Thomas, Lise Heier, Anne K Brysting, Bengt Oxelman, and Kjetill S Jakobsen (2015). "From gene trees to a dated allopolyploid network: insights from the angiosperm genus Viola (Violaceae)". In: *Systematic biology* 64.1, pp. 84–101.

Marcussen, Thomas, Simen R Sandve, Lise Heier, Manuel Spannagl, Matthias Pfeifer, Kjetill S Jakobsen, Brande BH Wulff, Burkhard Steuernagel, Klaus FX Mayer, Odd-Arne Olsen, et al. (2014). "Ancient hybridizations among the ancestral genomes of bread wheat". In: *science* 345.6194.

Matsubara, Kazumi, Chizuko Nishida, Yoichi Matsuda, and Yoshinori Kumazawa (2016). "Sex chromosome evolution in snakes inferred from divergence patterns of two gametologous genes and chromosome distribution of sex chromosome-linked repetitive sequences". In: *Zoological letters* 2.1, pp. 1–16.

Michael, Todd P (2014). "Plant genome size variation: bloating and purging DNA". In: *Briefings in functional genomics* 13.4, pp. 308–317.

Muller, HJ (1925). "Why polyploidy is rarer in animals than in plants". In: *The American Naturalist* 59.663, pp. 346–353.

Müller, Niels A, Birgit Kersten, Ana P Leite Montalvão, Niklas Mähler, Carolina Bernhardsson, Katharina Bräutigam, Zulema Carracedo Lorenzo, Hans Hoenicka, Vikash Kumar, Malte Mader, et al. (2020). "A single gene underlies the dynamic evolution of poplar sex determination". In: *Nature Plants*, pp. 1–8.

NCBI (2018). "Database resources of the national center for biotechnology information". In: *Nucleic acids research* 46.D1, pp. D8–D13.

Obbard, Darren J, Stephen A Harris, Richard JA Buggs, and John R Pannell (2006). "Hybridization, polyploidy, and the evolution of sexual systems in Mercurialis (Euphorbiaceae)". In: *Evolution* 60.9, pp. 1801–1815.

Obbard, Darren J, Stephen A Harris, and John R Pannell (2006). "Sexual systems and population genetic structure in an annual plant: testing the metapopulation model". In: *The American Naturalist* 167.3, pp. 354–366.

Orr, H Allen (1990). "" Why polyploidy is rarer in animals than in plants" revisited". In: *The American Naturalist* 136.6, pp. 759–770.

Oxelman, Bengt, Anne Krag Brysting, Graham R Jones, Thomas Marcussen, Christoph Oberprieler, and Bernard E Pfeil (2017). "Phylogenetics of allopolyploids". In: *Annual Review of Ecology, Evolution, and Systematics* 48, pp. 543–557.

Pamilo, Pekka and Masatoshi Nei (1988). "Relationships between gene trees and species trees." In: *Molecular biology and evolution* 5.5, pp. 568–583.

Pannell, John (1997). "Widespread functional androdioecy in Mercurialis annua L.(Euphorbiaceae)". In: *Biological Journal of the Linnean Society* 61.1, pp. 95–116.

Pannell, John R (2000). "A hypothesis for the evolution of androdioecy: the joint influence of reproductive assurance and local mate competition in a metapopulation". In: *Evolutionary Ecology* 14.3, pp. 195–211.

Pannell, John R (2002). "The evolution and maintenance of androdioecy". In: *Annual Review of Ecology and Systematics* 33.1, pp. 397–425.

Pannell, John R and Jörn Gerchen (2018). "Sex determination: sterility genes out of sequence". In: *Current Biology* 28.2, R80–R83.

Pannell, John R, Darren J Obbard, and Richard JA Buggs (2004). "Polyploidy and the sexual system: what can we learn from Mercurialis annua?" In: *Biological Journal of the Linnean Society* 82.4, pp. 547–560.

Pease, James B and Matthew W Hahn (2013). "More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting". In: *Evolution* 67.8, pp. 2376–2384.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Picard toolkit (2018). `http://broadinstitute.github.io/picard/`.

Popp, Magnus and Bengt Oxelman (2007). "Origin and evolution of north american polyploid silene (caryophyllaceae)". In: *American Journal of Botany* 94.3, pp. 330–349.

Pujol, Benoit, Shu-Rong Zhou, Julia Sanchez Vilas, and John R Pannell (2009). "Reduced inbreeding depression after species range expansion". In: *Proceedings of the National Academy of Sciences* 106.36, pp. 15379–15383.

Qvarnström, Anna and Richard I Bailey (2009). "Speciation through evolution of sex-linked genes". In: *Heredity* 102.1, pp. 4–15.

Renner, Susanne S (2014). "The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database". In: *American Journal of botany* 101.10, pp. 1588–1596.

Richardson, Aaron O and Jeffrey D Palmer (2007). "Horizontal gene transfer in plants". In: *Journal of experimental botany* 58.1, pp. 1–9.

Rothfels, Carl J (2021). "Polyploid phylogenetics". In: *New Phytologist* 230.1, pp. 66–72.

Rothfels, Carl J, Kathleen M Pryer, and Fay-Wei Li (2017). "Next-generation polyploid phylogenetics: Rapid resolution of hybrid polyploid complexes using PacBio single-molecule sequencing". In: *New Phytologist* 213.1, pp. 413–429.

Russell, JRW and JR Pannell (2015). "Sex determination in dioecious Mercurialis annua and its close diploid and polyploid relatives". In: *Heredity* 114.3, pp. 262–271.

Singh, RB and Ben W Smith (1971). "The mechanism of sex determination in Rumex acetosella". In: *Theoretical and Applied Genetics* 41.8, pp. 360–364.

Stamatakis, Alexandros (2014). "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies". In: *Bioinformatics* 30.9, pp. 1312–1313.

Stegemann, Sandra, Mandy Keuthe, Stephan Greiner, and Ralph Bock (2012). "Horizontal transfer of chloroplast genomes between plant species". In: *Proceedings of the National Academy of Sciences* 109.7, pp. 2434–2438.

Stölting, Kai N, Rick Nipper, Dorothea Lindtke, Celine Caseys, Stephan Waeber, Stefano Castiglione, and Christian Lexer (2013). "Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species". In: *Molecular ecology* 22.3, pp. 842–855.

Suarez-Gonzalez, Adriana, Christian Lexer, and Quentin CB Cronk (2018). "Adaptive introgression: a plant perspective". In: *Biology letters* 14.3, p. 20170688.

Tennessen, Jacob A, Rajanikanth Govindarajulu, Tia-Lynn Ashman, and Aaron Liston (2014). "Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps". In: *Genome biology and evolution* 6.12, pp. 3295–3313.

Tennessen, Jacob A, Na Wei, Shannon CK Straub, Rajanikanth Govindarajulu, Aaron Liston, and Tia-Lynn Ashman (2018). "Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries". In: *PLoS biology* 16.8, e2006062.

Untergasser, Andreas, Ioana Cutcutache, Triinu Koressaar, Jian Ye, Brant C Faircloth, Maido Remm, and Steven G Rozen (2012). "Primer3—new capabilities and interfaces". In: *Nucleic acids research* 40.15, e115–e115.

Untergasser, Andreas, Harm Nijveen, Xiangyu Rao, Ton Bisseling, René Geurts, and Jack AM Leunissen (2007). "Primer3Plus, an enhanced web interface to Primer3". In: *Nucleic acids research* 35.suppl_2, W71–W74.

Veltsos, Paris, Guillaume Cossard, Emmanuel Beaudoing, Genséric Beydon, Dessislava Savova Bianchi, Camille Roux, Santiago C Gonzàlez-Martinez, and John R Pannell (2018). "Size and content of the sex-determining region of the Y chromosome in dioecious Mercurialis annua, a plant with homomorphic sex chromosomes". In: *Genes* 9.6, p. 277.

Veltsos, Paris, Kate E Ridout, Melissa A Toups, Santiago C Gonzàlez-Martinez, Aline Muyle, Olivier Emery, Pasi Rastas, Vojtech Hudzieczek, Roman Hobza, Boris Vyskot, et al. (2019). "Early sex-chromosome evolution in the diploid dioecious plant Mercurialis annua". In: *Genetics* 212.3, pp. 815–835.

Vicoso, Beatriz and Brian Charlesworth (2006). "Evolution on the X chromosome: unusual patterns and processes". In: *Nature Reviews Genetics* 7.8, pp. 645–653.

Wagner, Natascha D, Li He, and Elvira Hörandl (2020). "Phylogenomic Relationships and Evolution of Polyploid Salix Species Revealed by RAD Sequencing Data". In: *Frontiers in plant science* 11, p. 1077.

Wendel, Jonathan F, Andrew Schnabel, and Tosak Seelanan (1995). "Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium)". In: *Proceedings of the National Academy of Sciences* 92.1, pp. 280–284.

Yakimowski, Sarah B and Spencer CH Barrett (2016). "The role of hybridization in the evolution of sexual system diversity in a clonal, aquatic plant". In: *Evolution* 70.6, pp. 1200–1211.

Yang, Wenlu, Zhiyang Zhang, Deyan Wang, Yiling Li, Shaofei Tong, Mengmeng Li, Xu Zhang, Lei Zhang, Liwen Ren, Xinzhi Ma, et al. (2020). "A general model to explain repeated turnovers of sex determination in the Salicaceae". In: *bioRxiv*.

Chapter 3

Sex chromosome evolution in the genus *Mercurialis*

3.1 Introduction

Recombination suppression is a common pattern found on sex chromosomes (Charlesworth 2017) and has evolved repeatedly in diverse lineages including animals, plants (Ming et al. 2011), algae (Coelho et al. 2019) and fungi (Branco et al. 2017). The causes of recombination suppression are still a matter of scientific discussion (Charlesworth 2017), but due to recent advances in NGS technology, an increasing number of independently evolved sex chromosome systems are being characterized at the molecular level. The majority of these datasets represent animal sex chromosomes, where studies revealed several chracteristics that are commonly found in the non-recombining region (Bachtrog 2013). In many cases they show a strong enrichment in repetitive elements (Chalopin et al. 2015) and a loss of functional gene copies (Skaletsky et al. 2003). Another pattern that has been found in multiple sex chromosome systems are 'evolutionary strata'. Evolutionary strata are discreet regions within the non-recombining region, characterized by their differences in their molecular signatures of genetic differentiation between both sex chromosome copies (Lahn and Page 1999; Charlesworth 2017). These differences are interpreted as the result of a stepwise cessation of recombination across the length of the Y chromosome and molecular signatures are indicative of the time of recombination suppression.

Several hypothesis have been put forward to explain the evolution of recombination suppression on animal sex chromosomes (Ponnikas et al. 2018). Recombination suppression has been explained by either

adaptive processes, by which linkage of sexually antagonistic alleles to the sex determiner increases the fitness of the sexes (Charlesworth and Charlesworth 1980; Rice 1987). Alternative explanations proposed that recombination suppression could be the consequence of an evolutionary arms race between meiotic drive loci and their suppressors (Ubeda et al. 2015), or recombination suppression has been interpreted as largely neutral process, driven by reduced purifying selection due to reductions of effective population size and recombination on the sex chromosomes (Jeffries et al. 2021).

Sex chromosome evolution in plants differs from animals because in many dioecious plants, separate sexes and sex chromosomes evolved from combined sexual systems (Renner and Ricklefs 1995), while this is rare in animals where new sex chromosomes most commonly evolve through sex chromosome turnover (Vicoso 2019). This difference lead to the idea that recombination suppression on plant sex chromosomes may be intrinsically linked to the evolution of separate sexes, and several evolutionary models of the evolution of separate sexes include recombination suppression (Charlesworth and Charlesworth 1978a; Charlesworth and Charlesworth 1978b). An additional difference in plants is that unlike animal sperm, plant pollen expresses a large proportion of genes during its haploid pollen growth state, which results in a strong potential for haploid selection, which could also effect the evolution of sex chromosomes in different ways (Delph 2019; Scott and Otto 2017; Sandler et al. 2018; Chibalina and Filatov 2011).

These differences lead to question of whether processes of sex chromosome evolution that are commonly found on animal sex chromosomes also happen on plant sex chromosomes. As dioecy is often found at the tips of phylogenetic trees, plant sex chromosomes are often young (Charlesworth 2019). This makes it difficult to distinguish if plant sex chromosomes are homomorphic because they had less time to evolve recombination suppression or if they are less prone to evolve recombination suppression. In a review, Müller and Renner argued that plant sex chromosomes do not evolve recombination suppression in the same way as in animals after the initial establishment of separate sexes (Renner and Müller 2021). Their argument is based on a lack of correspondence between the estimated extent of recombination suppression and the age of plant sex chromosome systems that have been characterized on a molecular level. However, there are several problems with this argument. Comparing the degree of recombination between independently evolved sex chromosome systems assumes that recombination suppression evolves at a constant or predictable rate in different clades. However, there is no evidence that this is the case in animal sex chromosomes. There are multiple examples of old animal sex chromosomes that show limited recombination suppression (Yazdi et al. 2020; Kuhl et al. 2020) and the presence of evolutionary strata suggests that recombination suppression can evolve in discrete steps and not at a constant rate (Lahn and Page 1999; Charlesworth 2017). Also their analysis is based on the limited number of plant systems for which sufficient molecular data are available. It is technically easier to sequence sex chromosomes with a

more limited degree of recombination suppression. As a result, there may be a bias in this dataset which excludes sex chromosomes with a greater degree of recombination suppression, which have not yet been sufficiently characterized on a molecular scale.

A better way to test hypotheses about the evolution of recombination suppression on plant sex chromosomes would be to compare sex chromosomes of multiple related species from dioecious clades that share an ancestral sex chromosome system. This would allow to distuinguish between situations where recombination suppression evolved as a consequence of the evolution of separate sexes and where it evolved or extended afterwards. Alternatively, if sex chromosome turnover events are identified in dioecious clades of plants, it could be of particular interest to test if their evolutionary trajectories differ from sex chromosomes that originate from transitions towards separate sexes. However, despite the fact that there are likely hundreds of independently evolved sex chromosome systems in plants (Renner 2014), the number of such datasets is very limited so far (Torres et al. 2018; Yang et al. 2020).

The molecular analysis of sex chromosomes has been met by particular challenges that in many systems have still not been completely overcome. The high repeat content of sex-linked regions makes their assembly difficult (Simpson and Pop 2015) and many genome assembly projects focussed on the homogametic sex, thereby excluding information on sex chromosome evolution (Tomaszkiewicz, Medvedev, et al. 2017). In addition, when sequencing genomic DNA of the heterogametic sex, genomic reads will contain reads from both sex chromosomes. Since the sex chromosomes usually have a common autosomal ancestor, many parts of both sex chromosomes share ancestral homology and it is difficult to assign reads to either sex chromosome. As a result, assemblies of sex chromosomes often include chimeric contigs that are based on reads from both gametologs that can confound inferences based on them.

There are different approaches for studying sex chromosomes in the face of these difficulties (Tomaszkiewicz, Medvedev, et al. 2017). It is not always necessary to generate perfect assemblies of sex chromosomes to study sex chromosome evolution and a number of bioinformatics approaches that use short read data from individuals or populations to identify sex-specific reads and variants have been developed (Palmer et al. 2019). In addition, segregation analysis in sexed offspring of crosses can be used to identify sex-linked variants (Muyle et al. 2016). Also there are technical approaches that can help to selectively sequence a single degenerate sex chromosome copy. For the single haplotype iterative mapping and sequencing approach genomic DNA is cloned into bacterial artificial chromosomes (BACs), and BAC clones that contain sex-specific sequences are identified using genetic markers. These regions can then be iteratively extended by identifying additional genetic variants in overlapping BAC sequences (Hughes et al. 2010). Alternatively, if sex chromosomes differ strongly in their size, repeat content or GC content from each other

and from autosomes, it is possible to use flow sorting to enrich and sequence specific sex chromosomes (Tomaszkiewicz, Rangavittal, et al. 2016).

In microscopically indistinguishable (homomorphic) sex chromosomes, in which recombination suppression is limited, different approaches for generating complete sex chromosome assemblies are possible. If the non-recombining region is small and has a limited repeat content, modern long read technology can be sufficient to generate complete assemblies of both sex chromosomes (Müller et al. 2020). Additional technologies like HiC mapping (Van Berkum et al. 2010) or optical mapping (Neely et al. 2011) can help to scaffold contigs in fragmented longread assemblies. In some cases it is also possible to generate individuals that have only the sex chromosome that is specific to the heterogametic sex. This is the case if either haploid individuals or tissues can be generated or, in diploids, if it is possible to generate viable offspring with two copies of the Y or W chromosome (Harkess et al. 2017; She et al. 2020). Viable YY males can sometimes be generated in species with a limited degree of sex chromosome degeneration, although data on this is sparse (Ming et al. 2011).

Previous analyses used several of these techniques to infer the extent and gene- and repeat content of the non-recombining region in diploid *M. annua* (Veltsos, Cossard, et al. 2018; Veltsos, Ridout, et al. 2019). In a first study, Veltsos et al. (2018) used a combination of RNA seq data and a previously published exon-capture dataset (Gonzalez-Martinez et al. 2017) to identify potential candidate genes and to identify sex-specific PCR markers. These markers were then used to screen BACs, which were sequenced and their gene and repeat content was analyzed. Overall, these results pointed to an SDR in *M. annua* with low gene density that is strongly enriched for repetitive elmenents. In an additional study, Veltsos et al. (2019) used a *de novo* genome assembly and transcriptomic data from crosses to further characterize the sex determining region using linkage maps in combination with SexDetector, a maximum-likelihood approach to identify sex-linked transcripts from cross data (Muyle et al. 2016). Combined data from both analyses estimated the size of the SDR to be between 14.5 and 19 Mb and to contain about 500 genes. However, a much smaller region on the SDR showed molecular patterns usually associated with sex chromosome evolution, like an elevated F_{ST} between male and female samples. These patterns were interpreted as evidence for the existence of two evolutionary strata on the Y chromosome of *M. annua*. The old stratum showed a signature of genetic differentiation, while the younger stratum around it was mostly defined by the absence of recombination in genetic crosses (Veltsos, Ridout, et al. 2019).

Both of these studies on *M. annua* had limitations. The BAC sequences represented only a subset of the estimated size of the sex determining region, which is insufficient to make clear inferences about the gene content in the non-recombining region. In contrast, the segregation-based analysis can only identify

regions for which an expressed gene copy exists on either both the X and the Y chromosome or on the X chromosome, but not for genes that are completely Y-linked and have no copy on the X chromosome. In addition, the size of the non-recombining region was estimated based on a segregation analysis of three crosses, which had a limited number of offspring. Estimating the extent of the SDR in this way can be problematic as the number of recombination events will be limited by the number of offspring. Sex differences in recombination rates are common in animals and plants and may affect different regions of the genome (Sardell and Kirkpatrick 2020). If the nonrecombining region is located in a region that has low recombination in the heterogametic sex, which has been shown to be the case for several plant sex chromosomes (Zhang et al. 2008; Pilkington et al. 2019; Kersten et al. 2014), rare recombination events could be missed and the extent of the SDR could be overestimated. A better measure for estimating the size of the SDR may be the presence of sex-specific variants inferred from population data. This has also been done in the same study and there appears to be a much smaller region that shows significant differences in male vs. female F_{ST}.

In the present study we improve the information on sex chromosome evolution in diploid *M. annua* and study the evolution of the non-recombining region in a phylogenetic context. We use new *M. annua* genome assemblies for both a YY male and an XX *M. annua* female. We use population data and comparative whole genome sequencing data between YY and XX individuals to infer molecular signatures of sex chromosome evolution and combine them with the result from a newly-developed linkage map of diploid *M. annua* males and with a linkage map based on a selfed F_2 cross from Chapter 4 of this thesis. We then use these newly-developed genomic resources to align additional population genomic data of diploid and polyploid lineages to determine the extent of the non-recombining region in other annual and perennial *Mercurialis* clades.

We find that the non-recombining region of the sex chromosomes in diploid *M. annua* is likely located in a pericentromeric region. In combination with a lack of molecular evidence of differentiation between males and females across large parts of this region, this challenges the conclusion by Veltsos et al. (2019) that a second evolutionary stratum is present on the sex chromosome of *M. annua*. However, our data from other lineages of *Mercurialis* suggests that there have been substantial independent extensions of the non-recombing region in *M. canariensis* and *M. huetii*, which indicates that recombination suppression did evolve after the evolution of separate sexes in these lineages.

3.2 Material and Methods

3.2.1 Genome assembly

The plants used for whole genome sequencing were generated by selfing a single *M. annua* male, which was feminized by external application of cytokinin (Li, Veltsos, et al. 2019) and was placed in a closed crossing box. We used a restriction enzyme based assay to identify YY males (Li, Veltsos, et al. 2019) and used a single YY male and XX female for further analyses. We did two rounds of high molecular weight DNA extraction from each individual using two different protocols.

First we extracted high molecular weight DNA from extracted nuclei. For this we first ground 5 grams of fresh leaves in liquid nitrogen and added 50 ml isolation buffer (15 mM Tris, 10 mM EDTA, 130 mM KCl, 20 mM NaCl, 8% (m/V) PVP-10, 0.1% Triton X-100, 7.5% (v/v) 2-Mercaptoethanol, pH 9.4). Nuclei suspension was filtered through 50 µl mira-cloth, 500 µl Triton X-100 were added and pelleted by centrifuging at 2000g at 4 °C for 10 min. Ten ml of Carlson lysis buffer (100 mM Tris pH 9.5, 2% CTAB, 1.4 M NaCl, 1% PEG 6000, 20 mM EDTA, prewarmed to 74 °C) and 25 µl of 2-Mercaptoethanol were added and the solution was incubated at 74 °C for 2 h. Samples were extracted twice by adding one volume of 24:1 chloroform/isoamyl alcohol to the aqueous phase and centrifuging at 5000 g for 10 min at 4 °C. 0.1 volume of 3 M NaOAc were added and DNA was precipitated by adding 1 volume of isopropanol and incubating overnight. After centrifuging at 4500 g for 90 min, pellets were washed by adding 10 ml of 70% EtOH and centrifuging at 4500 g for 90 min. Supernatant was discarded and pellets were air dried for 10 min and resuspended in Qiagen G2 buffer. We added 2 µl/ml RNase A (Qiagen, 100 mg/ml) to the resuspended DNA pellets, mixed and incubated for 5 min at room temperature and added 10 µl/ml Proteinase K. We incubated the sample at 50 °C for 1 h and cleaned it using Qiagen Genomic Tip 500/G columns following standard instructions. We added 10.5 ml of room temperature isopropanol to the eluted DNA and precipitated DNA overnight. We then centrifuged at 4500 g for 90 min at 4 °C, discarded the supernatant and added 40 ml of ice cold 70% EtOH, centrifuged at 4500 g for 30 min at 4 °C, discarded the supernatant, let the pellet air dry for 10 min and resuspended the DNA in 10mM TrisCl (pH 8).

For a second high molecular weight DNA extraction, 5 grams of fresh leaves were ground in liquid nitrogen and 40 ml of Carlson lysis buffer with 2.5 % (v/v) 2-Mercaptoethanol and 80 µl of RNase A (Qiagen, 100 mg/ml) were added. Samples were incubated at 65 °C for 1 h and after cooling to room temperature samples were extracted by adding 1 volume of 24:1 chloroform/isoamyl alcohol to the aqueous phase and centrifuging at 3500 g for 15 min at 4 °C and retaining the aqeuous phase. DNA was precipitated by adding 0.7 volumes of isopropanol and incubating at -80 °C for 15 min. The sample was centrifuged at

3500 g for 15 min at 4 °C and the supernatant discarded. The pellet was resuspended in Qiagen G2 buffer and cleaned using Qiagen Genomic Tip 500/G columns following standard instructions. 0.7 volumes of isopropanol were added to the eluted DNA and DNA was precipitated at -20 °C overnight and centrifuged at 3500 g for 45 min at 4 °C. The supernatant was discarded and the pellet was washed by adding 10 ml of ice cold 70% EtOH and centrifuging at 3500 g for 10 min at 4 °C. The supernatant was discarded and the pellet was air dried for 10 min after being resuspended in 10 mM TrisCl (pH 8).

We sequenced genomic DNA on Oxford Nanopore flowcells (v. 9.4.1) using an Oxford Nanopore Minion. Libraries were prepared using either ligation sequencing kit (SQK-LSK109) following standard instructions or the rapid sequencing kit (SQK-RAD004) from Oxford Nanopore following standard instructions. In addition, genomic DNA was sequenced on the Illumina NovaSeq S6 platform with 150 bp paired-end reads by Novogene UK.

Raw Oxford Nanopore reads were base-called using Guppy 4.2.2 (Oxford Nanopore). We assembled genomes using Flye 2.8.1 (Kolmogorov et al. 2019). We used the internal polishing function in Flye to do a first round of polishing with the raw Nanopore longreads. We then aligned short reads to our assembly using Minimap2 2.17-r941 (Li 2018) and sorted and indexed alignments using Samtools 1.3.1 (Li, Handsaker, et al. 2009) and polished our genome assemblies using Hypo 1.0.3 (Kundu et al. 2019). We generated a coverage histogram using PurgeHaplotigs 1.0.4 (Roach et al. 2018) to determine if a substantial number of allelic duplicate contigs are present in the assemblies, in which case we identified and removed them using the PurgeHaplotigs pipeline. We built a *de novo* repeat library from the male assembly using RepeatModeler2 2.01 (Flynn et al. 2020) and used RepeatMasker 4.1.1 to mask repetitive regions in both our genome assemblies and benchmarked completeness of our genome assemblies using Busco 5.1.2 (Seppey et al. 2019) using the eudicots_odb10 lineage dataset.

3.2.2 Linkage mapping

We built a new combined male linkage map for *M. annua* from three interspecific crosses between a male *M. annua* and a female *M. huetii*. DNA was extracted using either a robotic BioSprint 96 workstation (Qiagen) and the BioSprint DNA Plant kit (Qiagen) or using the DNeasy Plant Mini kit (Qiagen). Extracted DNA was quantified using the Qubit dsDNA HS Assay Kit on a Qubit fluorometer.

We generated Genotype data using a modified ddRad protocol (Peterson et al. 2012). For each sample we first digested about 100 ng DNA diluted to 22 µl by adding 2.5 µl Smartcut buffer (NEB), 0.4 µl EcoRI-HF (NEB) and 0.4 µl Taq1 (NEB) restriction enzymes and incubating at 37 °C for 30 min, 65 °C for

30 min and 80 °C for 20 min. We then ligated P1 and P2 adapters described in Peterson et al. (2012), which were diluted to 40 µM in 1x annealing buffer (50mM NaCl, 10 mM). We added 3 µl rATP (10 mM), 2 µl annealed P2 adapter (3 µM), 0.8 µl 10x T4 ligase buffer (NEB) 1 µl T4 ligase (400 U/µl, NEB), 22 µl digested DNA and 2 µl annealed P1 adapter (0.3 µM) and incubated for 20 min at 23 °C followed by 10 min at 65 °C. We then size selected 300 µl of pooled samples for each library to 550 bp. For this we first added 0.57x volume of Ampure XP beads incubated for 10 min at room temperature and saved the supernantant, to which we added 0.12x volume of Ampure XP beads (Beckman Coulter), incubated on magnetic stand, washed the beads with 70% ethanol and eluted DNA in 30 µl water for 2 min. We selected for biotin-labeled P2 adapters using M-270 Dynabeads (Invitrogen). We washed 15 µl Dynabeads 3 times in 1x bind and wash buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1M NaCl), resuspended them in 30 µl 2x bind and wash buffer, added the size selected DNA and incubated for 15 min at room temperature. We discarded the supernatant, washed the beads three times using 1x bind and wash buffer and resuspended the beads in 45 µl water. We then amplified size selected fragments by PCR. We used single indexed primers recommended by the modified ddRAD protocol (Peterson et al. 2012). We added 45 µl bead suspension, 3 µl forward primer (10 µM), 3 µl reverse primer (10 µM) and 50 µl KAPA HiFi Hotstart Ready Mix (Roche). PCR program was 2 min at 95 °C initial denaturation and 11 cycles at 98 °C for 20 s, 65 °C for 20 s and 72 °C for 30 s. We cleaned PCR products in 0.7x Ampure XP beads according to manufactorers instructions and eluted PCR products in 20 µl water. Libraries were sequenced using 150 bp single end reads by the Genomics Technologies Facility (GTF) at the University of Lausanne.

We demultiplexed and quality filtered raw illumina reads using Stacks 2.55 (Catchen et al. 2013). For this study, we were interested in the male *M. annua* linkage map and we wanted to only reconstruct alleles that were specific to the *M. annua* fathers. To this end, we ran a modified version of the Stacks pipeline. We first ran the ustacks program on all samples, allowing a maximum of four nucleotides distance between stacks. Then we ran cstacks and build the allele catalogue using only the fathers from all families, allowing a maximum of four mismatches between samples. We then ran the sstacks, tsv2bam and gstacks scripts on all samples with standard parameters. We first ran the populations program and removed low coverage samples in which fewer than 50% of loci could be genotyped. Then we ran the populations program on samples from each cross separately and kept loci, which were found in at least 60% of samples. We ran a principal component analysis using adegenet 2.1.3 (Jombart 2008) to identify and remove outlier samples, which are likely the result of pollen contamination. We then ran populations a second time for all crosses (removing previously filtered samples), keeping markers that were found in 60% of samples in at least one of the crosses. Since we built the catalogue of alleles only from male *M. annua* samples, all genotype calls in the offspring should be homozygous for alleles inherited from their fathers and we allowed a maximum

heterozygosity of 0.1 in the populations script. We then used a custom Python script that further filters the output of the populations program and prepares the dataset for construction of linkage maps. This script filtered offspring genotypes that were heterozygous or contained alleles that were not one of the two major alles in the family and infered missing paternal genotypes from the offspring genotypes and it removed loci from families where the minor allele frequency was lower than 0.2. It added a dummy maternal sample genotype that was homozygous for the reference allele at all loci and also replaced one of the alleles of homozygous offspring genotypes with the reference allele. LepMap3 0.2 (Rastas 2017) was used to build the linkage map. We separated loci into linkage groups by adjusting the lodLimit parameter of the SeparateChromosomes2 module until markers formed eight major linkage groups. We added single markers using the JoinSingles2All module, with the distortionLod and iterate parameters activated and a LodLimit of five. We then ordered markers on linkage groups by running 20 runs of the OrderMarkers2 module for each linkage group and chose the result with the highest log likelihood.

3.2.3 Scaffolding of genome assemblies

We scaffolded our genomic contigs by using linkage map information from the new male *M. annua* linkage map developed in this chapter and the F_2 linkage map from cross 2 described in Chapter 4 of this thesis. We determined the genomic location of markers from the linkage maps by aligning either loci built by stacks for the male *M. annua* map or 100 bp genomic sequences around each marker extracted from the unscaffolded XX genome assembly that was used to build the female linkage map in Chapter 4, using Minimap2 2.17-r941 (Li 2018) with the 'sr' mode. We kept markers which unambiguously aligned to a single contig and had an alignment quality larger than 40. We combined the information from both linkage maps and built a scaffolded assembly using ALLMAPS 1.1.8 (Tang et al. 2015). ALLMAPS allows to set different weights for the linkage maps used for scaffolding, so that in case of conflicts between linkage maps, the ordering supported by the linkage map with the greatest weight is chosen. For both linkage maps we set the weight of the F_2 linkage map to 2 and the weight of the new male map to 1, because we wanted to order contigs within the non-recombining region based on the female F_2 cross. For comparison with the linkage maps from Veltsos et al. (2019) we aligned assembled transcripts against the scaffolded genome assembly using the 'splice' mode in Minimap2 2.17-r941 (Li 2018) and kept transcripts with a minimum alignment quality of 20. We used these scaffolded assemblies to test for associations between repeat density and recombination frequencies across all linkage groups. We fitted a loess regression model of the position of genetic markers on the linkage map against their position on the scaffolded YY genome assembly on both the newly-developed male *M. annua* linkage map and the F_2 linkage map using the

'loess' function in R 3.6.1 (R Core Team 2019) with the smoothing parameter (span) set to 0.1 and the degree of the polynomial set to 1. We then determined the proportion of sequence that was marked as repetitive in 100 kb windows and used the 'predict' function to determine the genetic position on the beginning and end of each window based on the loess model. We fitted linear models of repeat density against the genetic distance in each window for each linkage group.

3.2.4 Identification of sex-linked genomic regions

We used two types of data to identify sex-linked regions in diploid *M. annua*. We used whole genome resequencing data from the two individuals used for genome assembly, which has the advantage of covering the whole genome assembly, and since we are comparing a YY and an XX individual, there will always be only one of the sex chromosomes present in either dataset. In principle it should be possible to identify regions that are specific to the Y or X chromosomes by comparing genomic data from both individuals. However, since this analysis relies only on genomic data from two individuals, such an analysis is expected to have a high rate of false positives. These false positives occur because it is not possible to distinguish between regions that are truly sex-linked and regions that just show differences between both samples. We combined this information with a second, already published exon-capture dataset, which includes 20 males and 20 females across the *M. annua* species range (Gonzalez-Martinez et al. 2017). This dataset is based on more than 400000 hybrid capture probes that represent large parts of the non-repetitive fraction of an older male *M. annua* genome assembly. We first aligned the sequences of the exon-capture probes against our male and female genome assemblies using the 'sr' mode of Minimap2 2.17-r941 (Li 2018). We filtered our alignments and kept alignments with an alignment quality greater than 20.

We used additional population genetic data to estimate the extent of the non-recombining region in hexaploid *M. annua* as well as in *M. huetii*, *M. canariensis* and in the perennial lineages *M. reverchonii*, *M. tomentosa* and *M. elliptica*. To this end we used the newly-developed exon-capture dataset described in Chapter 2. It included between 5 and 11 samples from each sex from each lineage (Tab. S3).

We trimmed and quality filtered the raw Illumina reads of the exon-capture datasets using trimmomatic 0.38 (Bolger et al. 2014) with standard parameters, aligned the reads against the male and female genome assemblies using BWA mem 0.7.17-r1188 (Li 2013) and marked PCR duplicates using PicardTools 2.90 (*Picard toolkit* 2018). In addition, we aligned both male and female whole-genome resequencing reads against both genome assemblies using BWA mem 0.7.17-r1188 (Li 2013).

3.2.5 Identification of Y- and X-hemizygous regions

We identified putative Y-hemizygous and X-hemizygous regions based on the mean sequencing depth of exon-capture data and of genomic resequencing data at the location of exon-capture probes in both genome assemblies using the bedcov function of samtools 1.11 (Li, Handsaker, et al. 2009) with a quality threshold of 20. We then used a custom Python script that estimates the mean sequencing depth of male and female samples for the exon-capture probeset. Only probes with a mean sequencing depth larger than 4 in at least 75% of the probes of at least one of the sexes were included in further analyses. In addition, the relative sequencing depth of male (YY) and female (XX) genomic resequencing reads was calculated at each of these probe locations for diploid *M. annua*. For this we first calculated the mean sequencing depths $\bar{d}_{XX}$ and $\bar{d}_{YY}$ across all probes in the XX and the YY datasets respectively. We then calculated the relative sequencing depths d_{YY} and d_{XX} for the YY and XX dataset for each probe by dividing the observed sequencing depth at the probe by $\bar{d}_{YY}$ or $\bar{d}_{XX}$. We then calculated the relative female genomic coverage as $c_f = \frac{d_{XX}}{d_{XX}+d_{YY}}$.

We identified putative Y-hemizygous probes for all lineages. Probes located in these regions should be found exclusively in males. We determined probes to be male-specific if the mean male coverage based on the exon-capture data was larger than 4 and the mean female coverage was smaller than 0.1. For diploid *M. annua* we used additional information from the genomic resequencing data and we only determined probes to be sex-linked if their relative female genomic coverage was smaller than 0.1 in addition to being male-specific, based on the exon-capture data.

For diploid *M. annua*, we also identified putative X-hemizygous probes. These should have half the sequencing depth in males compared to females in the exon-capture data, but should be completely missing in the genomic resequencing data based on the YY male. We determined probes to be female-specific if mean females coverage was at least 1.9 times the coverage of males in the exon-capture dataset and the relative female genomic coverage was 1. We did not attempt to identify female-specific probes in the other *Mercurialis* lineages, because we expect reduced statistical power to detect these probes for multiple reasons. First, we don't have any genomic resequencing data available for any species other than diploid *M. annua*, which we could use as an additional source of evidence. Second, the lower number of samples would result in a reduced power to detect these differences in general. Third, the relative difference in the number of X copies between males and females or monoecious individuals decreases with increasing ploidy, and all other lineages except for *M. huetii* are polyploids.

3.2.6 Identification of sex-linked genetic variation

For diploid *M. annua* and *M. huetii*, we called variants of exon-capture reads aligned to the male (YY) *M. annua* genome assembly using freebayes 1.0.2 (Garrison and Marth 2012) with standard parameters. We used a custom Python script which filters variants. We first discarded variants with a QUAL score smaller than 20. We then filtered genotypes with a sequencing depth smaller than 20 and kept variants with at least 75% of genotypes of both sexes remaining and we excluded low frequency variants by discarding sites where the frequency of any allele was greater than 0.7. At each of these sites we estimated Weir and Cockerham's F_{ST} (Weir and Cockerham 1984) between male and female samples using the wcFst module from vcflib 1.0 (https://github.com/vcflib/vcflib). For diploid *M. annua* we estimated median F_{ST} on both the sex chromsome and autosomal linkage groups in 100 kb sliding windows with at least five variants and we determined empirical percentiles from the distribution of F_{ST} values on autosomes. For the allopolyploid lineages, we estimated F'_{ST}, an alternative measurement of differentiation. F'_{ST} is based on the average number of alleles by which pairs of individuals differ, which accounts for the presence of fixed differences between homeologs in allopolyploids (Obbard et al. 2006). We filtered genotypes with a sequencing depth smaller than five times its ploidy and kept variants with at least 75% of genotypes of both sexes remaining and we excluded low frequency genotypes by discarding sites where the frequency of any genotype was greater than 0.7. We then estimated F'_{ST} for each site as described in Obbard et al. (2016). The exon-capture dataset for other lineages than diploid *M. annua* included a much lower density of autosomal exon-capture probes, so that calculating median F_{ST} or F'_{ST} in sliding windows on the autosomes was not possible. We estimated the percentiles of F_{ST} or F'_{ST} for autosomal variants from the distribution of 10000 replicate estimates of the median of randomly sampled autosomal variants instead. Since the number of variants in each window can differ, we drew the number of variants from the empirical distribution of the number of variants in sliding windows on linkage group 1 for each replicate.

3.2.7 Genetic correlations on linkage group 1 in *M. annua* and *M. huetii*

Sex chromosomes are expected to show an increased signature of linkage disequilibrium in and around the sex-linked region (Kirkpatrick, Guerrero, et al. 2010). The classical definition of linkage disequilibrium requires phased genotypes (Lewontin and Kojima 1960). However, our short read exon-capture data does not allow to phase genotypes on a chromosomal scale. We used pairwise correlations between the number of minor allele copies for bi-allelic variants, as an alternative measurement in diploid *M. annua* and *M. huetii*. For each sex, we filtered bi-allelic variants, which had a sequencing depth of at least 20 in 75% of the samples and which had a minor allele frequency greater than 0.3. We then calculated the squared

Pearson correlation coefficient (R^2) between the number of minor allele copies for all pairs of variants on linkage group 1 and estimated the median R^2 in 100 kb sliding windows across the length of linkage group 1.

3.3 Results

3.3.1 Genome assembly

We generated a total of 16.1 Gb (approximately 25.15x sequencing depth) and 15 Gb (approximately 23.44x sequencing depth) of Oxford Nanopore sequencing data for the XX and YY samples respectively. The raw assemblies had a total length of 744.61 Mb and 525.53 Mb and an N50 of 112406 bp and 155799 bp for the XX and YY assemblies respectively. On the coverage histogram from PurgeHaplotigs, a peak indicative of the presence of duplicated sequences was evident in the XX but not in the YY assembly. After running the PurgeHaplotigs pipeline on the XX assembly, the total length of the assembly was reduced to 494.59 mb and the N50 increased to 183581 bp. In comparison with v. 1.4 of the *M. annua* assembly (Veltsos, Ridout, et al. 2019), both our newly-developed genome assemblies have a higher contiguity and a greater number of complete universal single-copy orthologs used for benchmarking (BUSCOs) (Tab. 3.1). A total of 67.62% of the male and 65.94% of the female genome assembly were marked as repetitive.

Table 3.1: Comparison of the new *M. annua* genome assemblies with v1.4 of the *M. annua* genome assembly (Veltsos, Ridout, et al. 2019).

Assembly	New YY assembly	New XX Assembly	Old XY assembly
Total size [Mb]	523.23	494.59	546.38
N contigs	20358	5488	720537
N contigs (>1kb)	16602	5372	74927
N50 [bp]	155684	183581	6398
Largest contig [bp]	3898311	4716000	329358
Complete BUSCOs	97.8%	97.5%	77.7%
Complete and single-copy BUSCOs	90.4%	89.1%	73.3%
Complete and duplicated BUSCOs	7.4%	8.4%	4.4%
Fragmented BUSCOs	0.7%	0.6%	11.3%
Missing BUSCOs	1.5%	1.9%	11%

3.3.2 Linkage mapping

We generated a total of 330 Gb of RadSeq data. From a total of 469 F_1 offspring we removed 82 individuals that had low sequencing depth. We then identified 11 individuals (1 individual in Cross 1, 1

individual in Cross 2 and 9 individuals in Cross 3) that were outliers in the principal component analysis and likely the result of pollen contamination (Fig. S5). We built a combined linkage map from the remaining 376 F_1 offspring. Our newly-developed male *M. annua* linkage map consisted of 5915 genetic markers, with a total linkage map size of 521.34 cM. We could align 2924 and 3002 of these markers and 1804 and 2180 markers from the female F_2 *M. annua* map against the YY and XX assemblies respectively. For the YY male genome assembly, ALLMAPS anchored 1003 contigs with a total length of 257174394 bp (48.9% of the total genome assembly size) to chromosomal scaffolds. Fourhundred of these contigs with a total length of 163391262 bp (31.1% of the total genome assembly size) were oriented. For the XX female genome assembly, ALLMAPS anchored 1294 contigs with a total length of 259410467 bp (52.4% of the total genome assembly size) to chromosomal scaffolds. 488 of these contigs with a total length of 137707992 bp (27.8% of the total genome assembly size) were oriented. We aligned 9293 and 9487 transcripts from Veltsos et al. (2019), which were associated with male and female linkage map positions, against the scaffolded YY and XX assemblies respectively.

On our male *M. annua* linkage map, we identified an extended region of low recombination on linkage group 1, which spans approximately 19 Mb on the scaffolded YY genome assembly (Fig. 3.1 a). This is slightly smaller than the low-recombination region on the male linkage map from Veltsos et al. (2019) (Fig. 3.1 c), which spans approximately 22 Mb on the scaffolded male genome assembly. In addition, we found variation in recombination rates of both male and female linkage maps across the length of autosomes and regions of reduced recombination are found in the centers of all autosomes except for linkage group 4, where a region of reduced recombination is found at the beginning of the linkage group (Fig. 3.1 a-d). We also found variation in the proportion of sequences, which were masked as repetitive, along the length of both the X and the Y chromosomes and across all autosomes (Fig. 3.1 e). There was a highly significant negative association between the proportion of repetitive sequences with recombination rate ($p<0.001$) on all linkage groups for both the new male *M. annua* linkage map and for the F_2 linkage map (Fig. S6).

3.3.3 Identification of Y- and X-hemizygous regions

Diploid *M.annua*

We could align a total of 334869 and 340755 of the original exon-capture probe sequences against the YY and XX genome assemblies respectively. Based on the aligned exon-capture reads, we retained 262521 and 265403 of these probes that had sufficient sequencing depth of reads in at least one of the sexes in the YY and XX assemblies respectively. We identified 519 male-specific probes, located on 60 contigs with a

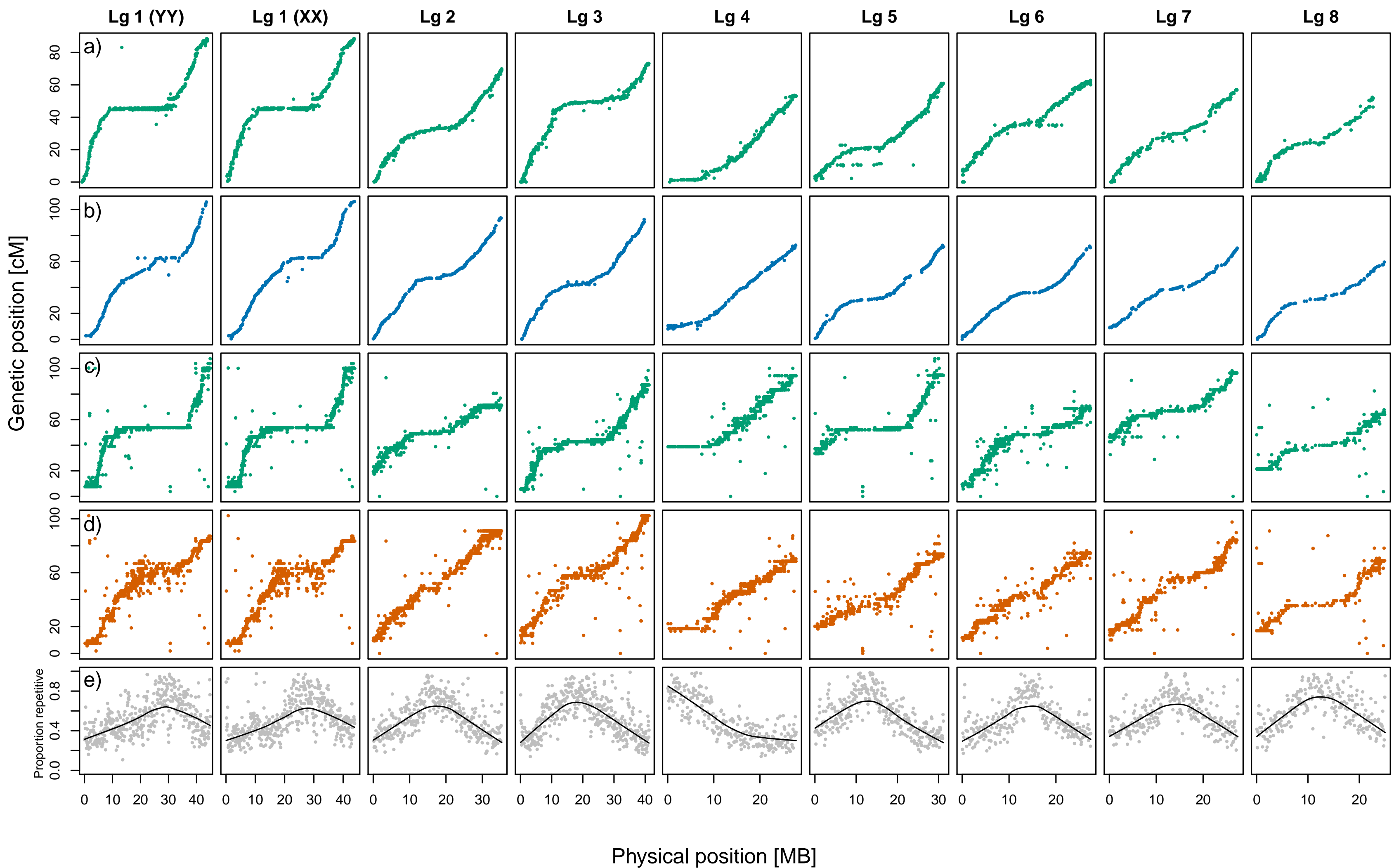


Figure 3.1: Linkage maps of (a) male *M. annua*, (b) female F_2 linkage map, (c) male and (d) female linkage maps from Veltsos et al. (2019) aligned to the newly-developed *M.annua* genome assemblies. (e) is the proportion of the genome assembly marked as repetitive.

total size 10267342 bp. None of these contigs were anchored to the scaffolded genome assembly. We also identified 330 female-specific probes, which were located on 8 contigs with a total length of 1421383 bp. 4 of these contigs were anchored to the XX assembly, two of which were located close to the center of linkage group 1 (contig_18631: 32 probes, contig_4957: 233 probes), while one was placed on linkage group 5 (contig_21714: 15 probes) and one on linkage group 6 (contig_4169: 2 probes) (Fig. 3.2).

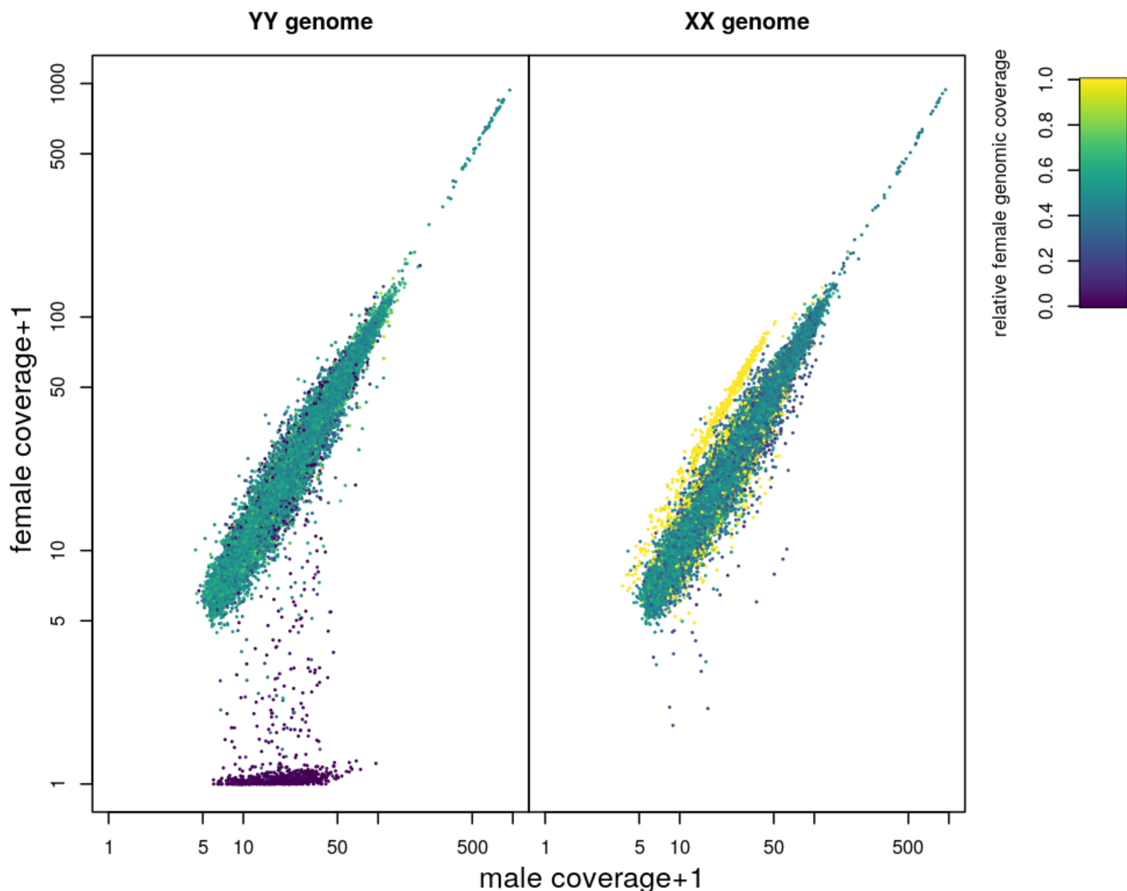


Figure 3.2: Male versus female sequencing depth of exon-capture data aligned against in the male (YY) and female (XX) genome assemblies of diploid *M. annua*. Color shading indicates relative female sequencing depth of genomic sequencing reads from the male (YY) and the female (XX) individual used for whole genome sequencing.

Other *Mercurialis* lineages

The exon-capture dataset in the other lineages consisted of a smaller number of probes than in diploid *M. annua*, with 40787 probes in *M. huetii*, 48655 probes in *M. canariensis*, 46730 in hexaploid *M. annua*, 43263 probes in *M. reverchonii*, 42516 probes in *M. tomentosa* and 45877 probes in *M. elliptica*. We determined a total of 206 probes to be male-specific in *M. huetii*. Out of these probes, 173 probes were anchored to linkage group 1 on the scaffolded genome assembly, 7 probes were anchored to other linkage

groups (one on linkage group 2, one on linkage group 3, two on linkage group 6, two on linkage group 7 and one on linkage group 8). In *M. canariensis*, 37 probes were determined to be male-specific. Four of these probes were anchored to linkage group 1. We identified a single male-specific exon-capture probe in hexaploid *M. annua*, six male-specific exon-capture probes *M. reverchonii* and three male-specific exon-capture probes in *M. tomentosa*, none of which were anchored to the scaffolded *M. annua* genome assembly. In *M. elliptica*, three male-specific probes were found, one of which was anchored to linkage group 4 of the *M. annua* genome assembly. A single exon-capture probe (located on contig_3334) was found to be male-specific in all lineages, except for *M. huetii*, where it was missing.

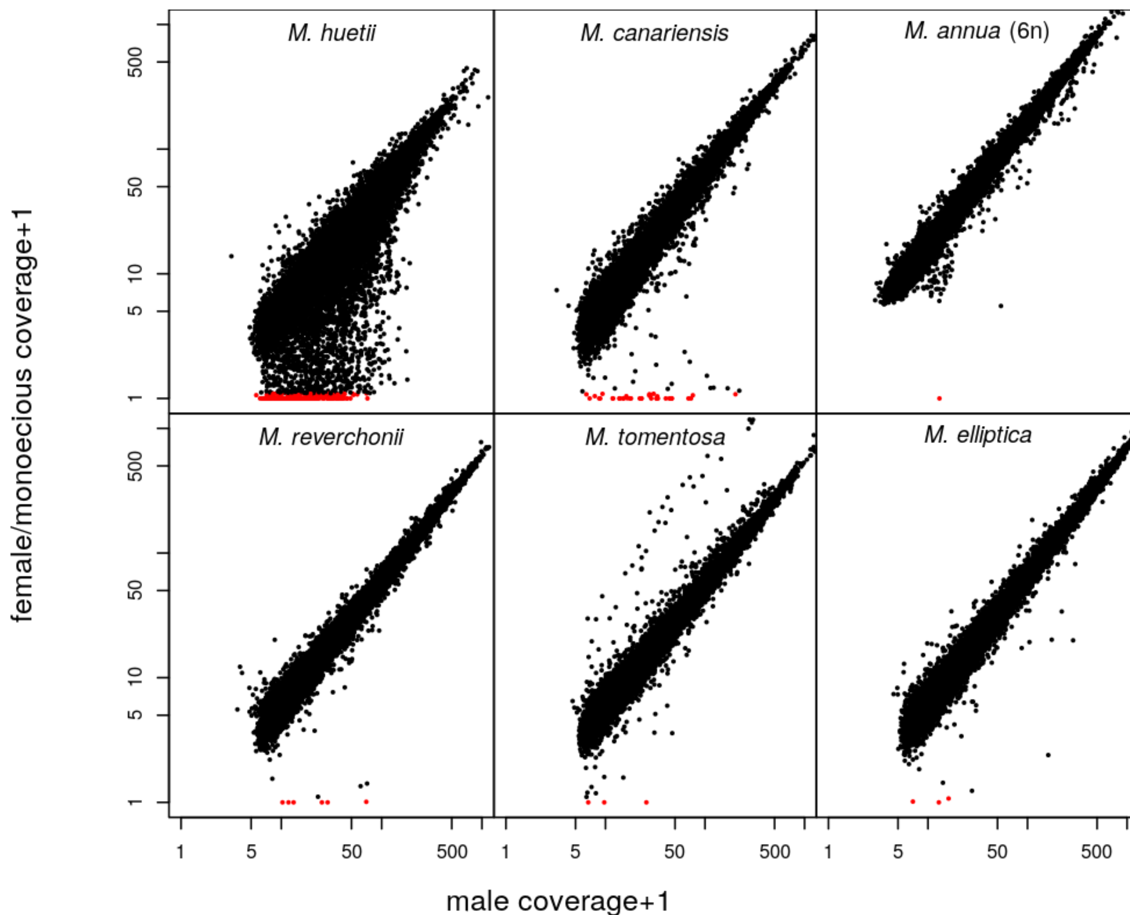


Figure 3.3: Male versus female/monoecious coverage of exon-capture from different lineages of annual and perennial *Mercurialis* species aligned against in the male (YY) genome assembly of *M. annua*. Probes colored in red were determined to be male-specific.

3.3.4 Identification of sex-linked genetic differentiation

A single peak of differentiation between male and female samples of diploid *M. annua* was found at 23.7 Mb at the center of the scaffolded chromosome 1 assembly. In *M. huetii*, large regions with a strong

signal of differentiation between males and females were found along a 29.6 Mb region (67.6% of the total length of the scaffolded chromosome assembly) of chromosome 1 of the *M. annua* genome assembly. In this region, also 173 male-specific exon-capture probes were located. In *M. canariensis*, four peaks of differentiation were found in a 19.7 Mb region across chromosome 1 (45% of the total length of the scaffolded chromosome assembly). In this region, also four male-specific exon-capture probes were found. In contrast, no clear peaks of differentiation were found on chromosome 1 in hexaploid *M. annua*. In *M. reverchonii*, a single narrow peak of differentiation was found at 17.9 Mb of the *M. annua* genome assembly. No peaks of differentiation were found for *M. tomentosa* and *M. elliptica* (Fig. 3.4).

3.3.5 Genetic correlations on linkage group 1 in *M. annua* and *M. huetii*

In diploid *M. annua* males and females, a small region with high median genetic correlations was found at approximately 19 Mb (Fig. 3.5). This region coincided with the peak of genetic differentiation between male and female samples (Fig. 3.4)

In *M. huetii* males large regions with strong median genetic correlations were found across large parts of linkage group 1 (Fig. 3.6). These regions coincided with regions where strong genetic differentiation between male and female samples were found (Fig. 3.4) In contrast, no strong genetic correlations were found in the females of *M. huetii*.

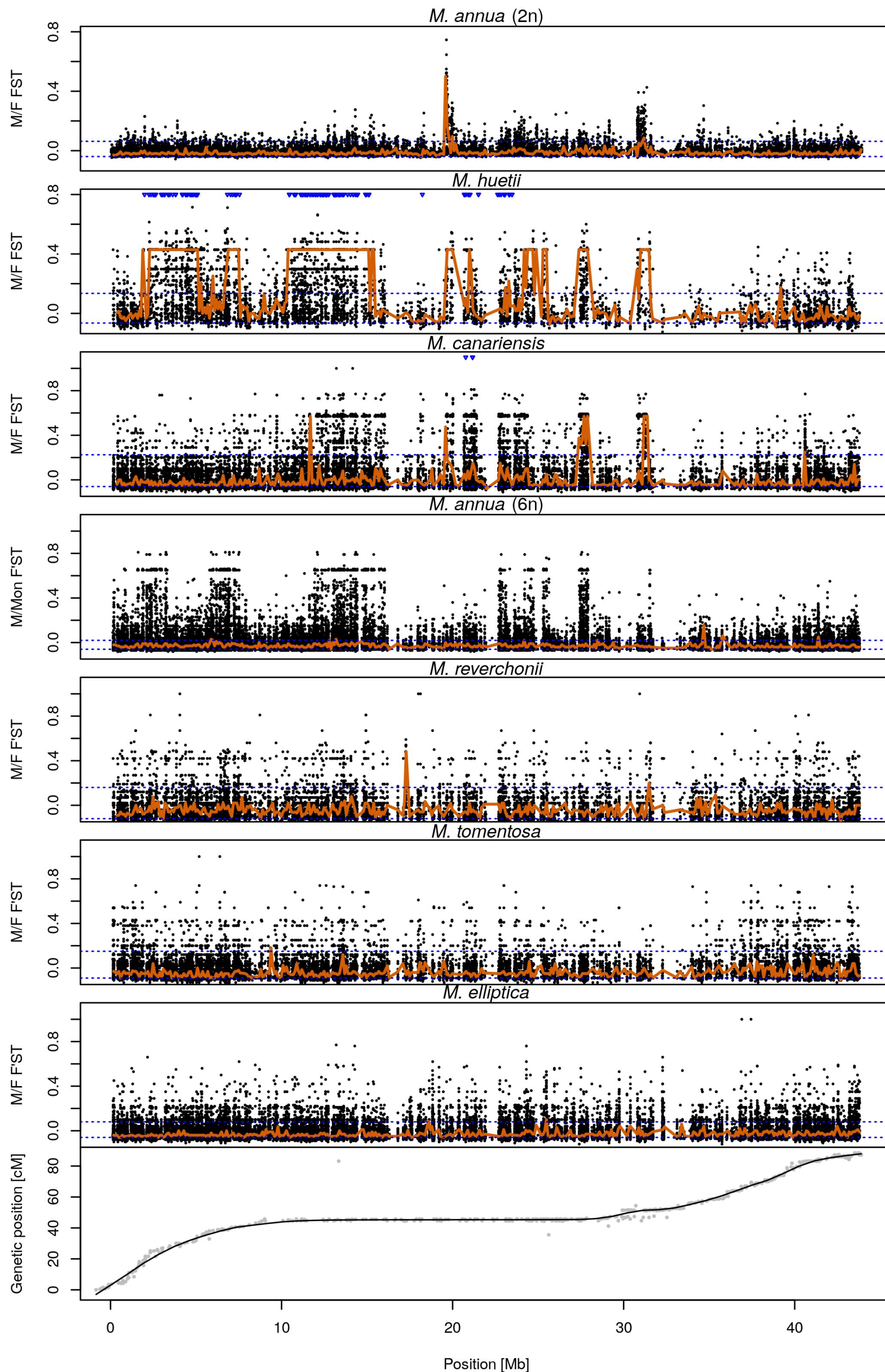


Figure 3.4: F_{ST} between males and females in *M. annua* and *M. huetii* and F'_{ST} for polyploid *Mercurialis* lineages on linkage group 1. Orange lines indicate the median F_{ST} or F'_{ST} across 100 kb sliding windows and dotted blue lines indicate the percentiles of these estimates based on autosomal markers. Blue triangles on the top of the plots of *M. huetii* and *M. canariensis* indicate the location of male-specific exon-capture probes. The bottom panel indicates genetic positions of the male *M. annua* linkage map.

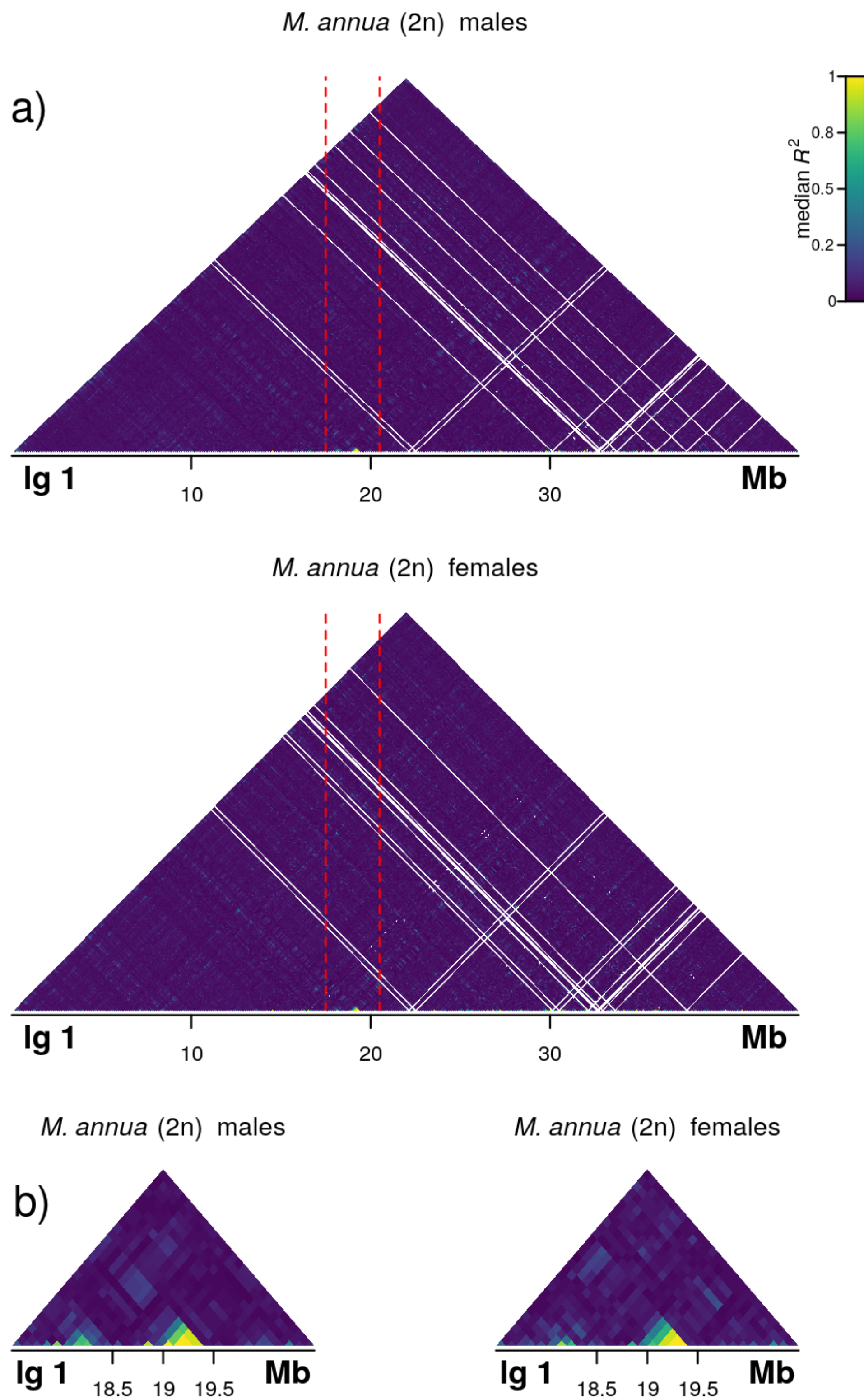


Figure 3.5: Median pairwise genetic correlations in 100 kb windows for male and female diploid *M. annua* across the whole length of linkage group 1 (a). Dashed red lines indicate the region where a peak of high genetic differentiation between male and female *M. annua* is found, which is enlarged in subfigure b.

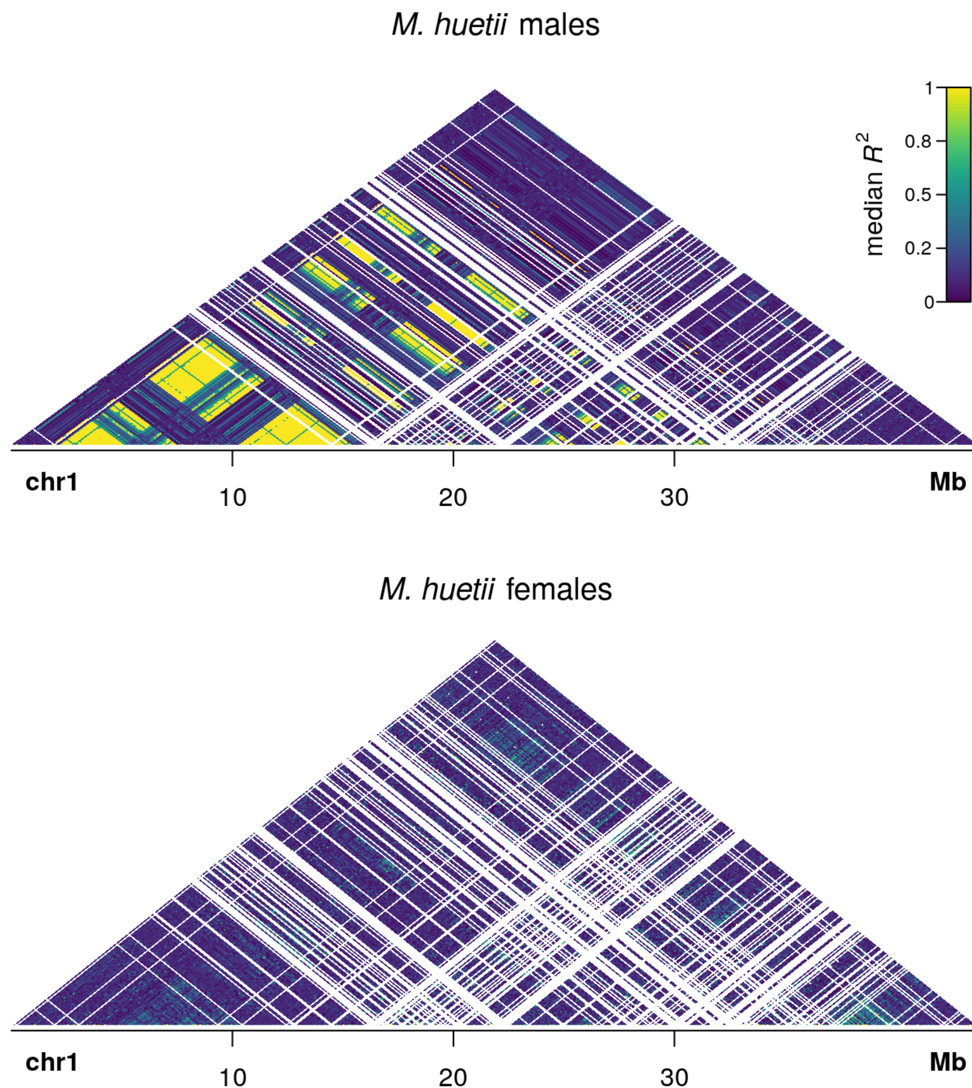


Figure 3.6: Median pairwise genetic correlations in 100 kb windows across linkage group 1 for male and female *M. huetii*.

3.4 Discussion

Our newly-developed male *M. annua* linkage map shows an extended region of reduced recombination on linkage group 1 (Fig. 3.1), while our analysis of an exon-capture dataset identifies a much smaller region of genetic differentiation between males and females within this region (Fig. 3.4). These results are concordant with the main results from Veltsos et al. (2019), who interpreted this as evidence for two evolutionary strata, one older stratum in which a clear signal of genetic differentiation is found and a surrounding younger stratum, where recombination is suppressed, but no substantial accumulation of

genetic differences between X and Y sequences has taken place yet (Veltsos, Ridout, et al. 2019). We used additional genome assemblies of an XX female and a YY males to analyze the density of repetitive sequences on sex chromosomes and autosomes (Fig. 3.1 e, Fig. S6). The accumulation of repetitive sequences is one of the hallmarks of sex chromosome degeneration (Kejnovsky et al. 2009) and we found a region of increased repeat density, which coincides with the region of reduced recombination on the Y chromosome. However, we found regions of increased repeat density, which coincide with reduced recombination, on all other autosomes and on the X chromosome (Fig. 3.1).

Such a pattern of reduced recombination (Anderson et al. 2003; Jensen-Seaman et al. 2004) and increased density of repeats (Willard 1990; Ma et al. 2007) is commonly found in pericentromeric regions of plants and animals. In *M. annua*, there is a similar increase at the center of the XX and YY assemblies (Fig 3.1), which has been identified as the sex determining region (Veltsos, Cossard, et al. 2018; Veltsos, Ridout, et al. 2019). In addition, the correlation between repeat density and recombination rates is greater in the F_2 map than in the male *M. annua* map (Fig. S6). This suggests that the majority of repeats in this region did not accumulate as a result of sex chromosome evolution, but rather that the sex determining region is located in a pericentromeric region, where the increased repeat density predates the evolution of the sex chromosomes. In many plant and animal species, there are sex-specific differences in recombination rates around the centromeres (Sardell and Kirkpatrick 2020) and linkage maps may not be reliable indicators of the size of the non-recombining region of sex chromosomes that are located in these regions. More robust estimators of sex determining regions are peaks of differentiation between male and female samples and the buildup of genetic correlations between variants in males, which are expected in and around the sex-linked region (Kirkpatrick, Guerrero, et al. 2010).

We found a single peak of genetic differentiation between males and females at 19.6 Mb on the scaffolded Y chromosome assembly of *M. annua*, and it makes up only a minor proportion of the region of reduced recombination on the male *M. annua* linkage map (Fig. 3.4). In addition, increased levels of pairwise genetic correlations, which are expected to build up in non-recombining regions, are limited to a small region around this peak of differentiation as well (Fig. 3.5). However, the presence of sex-specific exon-capture probes on unscaffolded contigs suggests that there is likely additional sex-linked sequence in diploid *M. annua* that is not included in the scaffolded genome assembly. Based on the total size of unscaffolded contigs, on which male-specific exon-capture probes were found, we estimate that this additional region has a size of approximately 10 Mb. However, this could be an underestimate, since there could be a substantial amounts of additional repetitive sequence that is not included in these contigs. In addition, we detected a smaller number of female-specific exon-capture probes, indicative of further differentiation between X and Y chromosomes at the sequence level.

Using additional exon-capture data, we tested for signatures of differentiation between male and female or monoecious samples in other lineages of *Mercurialis* as well. We identified several regions that show both a strong signal of differentiation between males and females (Fig. 3.4) and a strong signal of genetic correlations in *M. huetii* (Fig. 3.6). These regions cover a major part of the scaffolded assembly of linkage group 1, which is larger than the region of reduced recombination on the male *M. annua* linkage map. At the same time there is a large number of male-specific exon-capture probes located in or close to these regions of differentiation (Fig 3.4). These patterns suggest that there has been a substantial extension of the non-recombining region on the sex chromosomes of *M. huetii*. The high degree of differentiation suggests that the sex chromosomes are differentiated at the sequence level. Most of the exon-capture probes that are male-specific on linkage group 1 of *M. huetii* show no sex bias in sequencing depth in diploid *M. annua*, which suggests that copies were lost from the X chromosome. Due to limited statistical power to detect female-specific probes in our *M. huetii* samples, we can't test if a similar or greater number of exon capture probes were also lost from the Y chromosome. Usually genetic degeneration on sex chromosomes is assumed to affect the Y chromosome (Bachtrog 2013), but due to its hemizygosity in males the X chromosome can also show unusual evolutionary patterns compared to autosomal sequences. However these are usually thought to cause accumulation of additional sex-biased genes (Vicoso and Charlesworth 2006) and large scale gene loss on the X chromosome would be unexpected. Further analysis of genomic data could be used to test if such an unusual pattern does exist on the X chromosome of *M. huetii*.

In *M. canariensis* we found four peaks of differentiation between males and females as well as four male-specific exon-capture probes, located on linkage group 1 of the scaffolded genome assembly (Fig. 3.4). These signals of differentiation are found in a region, which is approximately equivalent to the region of reduced recombination on the male *M. annua* linkage map. These data suggest that there has been an extension of the non-recombining region on the Y chromosome of *M. canariensis* as well, although both the extent of the non-recombining region and the level of sequence divergence are smaller than in *M. huetii*, suggesting a more recent extension.

In contrast, we did not find a peak of differentiation between males and monoecious plants on linkage group 1 of hexaploid *M. annua*. Similarly, we did not find a peak of differentiation between males and females in perennial *M. tomentosa* and *M. elliptica*. In perennial *M. reverchonii*, a single peak was found at 17.3 Mb of linkage group 1 (Fig. 3.4). In addition, we found a low number of male-specific exon-capture probes in hexaploid *M. annua* and the perennial lineages, but we found a single exon-capture probe which is male-specific in all lineages except *M. huetii*, where it is missing (Fig. 3.3). All of this points to a much smaller extent of recombination suppression around a common sex determining region in perennial

Mercurialis lineages and in hexaploid *M. annua*.

When interpreting these patterns, it is important to take the limitations of our datasets into account. Our genome assemblies are a substantial improvement both in terms of contiguity and of completeness over the previous *M. annua* genome assembly (Tab. 3.1). Nonetheless, the size of contigs is still limited and our assemblies are far from being chromosome scale assemblies. With our scaffolding approach we could only anchor about half of the sequence in both genome assemblies to linkage map markers. From these we can infer the orientation for an even lower proportion of contigs, on which more than one marker is present. For the region of reduced recombination on the male linkage map, the order of contigs in the assembly is based on the female map, even though it is possible that the true order of contigs on the male map may differ. As only non-repetitive markers can be used to anchor contigs to linkage maps, contigs with a higher proportion of repeats will be underrepresented in the scaffolded genome assembly. As a consequence, our assembly likely underestimates the size of the repeat rich pericentromeric regions in which the sex determining region in diploid *M. annua* is located. As a result, only a small proportion of the sex-linked region is included in the scaffolded genome assembly, while the majority of male-specific exon-capture probes are located on unscaffolded contigs. In addition, such a scaffolding approach strongly relies on the correctness of the linkage maps. Our scaffolding uses information from two linkage maps. We used the F_2 linkage map as reference and markers that disagree between it and the male *M. annua* linkage map are located in positions outside the curve of markers. This can be seen in Fig. 3.1 a and b. However, the degree of disagreement between linkage maps is much smaller than with the old linkage maps developed by Veltsos et al. (2019), where a much larger number of markers is located in positions beyond the curve (Fig. 3.1 c and d). Such a level of disagreement cannot only be explained by the smaller number of offspring in the crosses used by Veltsos et al. and is probably caused by less thorough data filtering and lack of manual curation of markers on linkage maps. These results illustrate the fact that linkage maps are not perfect and genome assemblies that are scaffolded using linkage maps can inherit these errors. In addition, there may be errors in the genome assembly itself, which may be the case for the two contigs on which we identified female-specific exon-capture probes, but which were anchored to the autosomes and not the sex chromosomes.

The diploid *M. annua* exon-capture dataset contains a large part of the non-repetitive proportion of the *M. annua* genome (Gonzalez-Martinez et al. 2017) and with a total of over 260000 probes that remained after filtering, on average one probe is located at every 2.5 kb of the genome. With such a dense dataset it seems unlikely that our analysis missed large portions of the non-repetitive fraction of the sex chromosomes and a sample size of 20 males and females should give sufficient statistical power for detecting sex-linkage of probes and variants. When designing our new exon-capture dataset for other lineages of *Mercurialis*, we

intended to cover multiple lineages of *Mercurialis*, which limited both the number of exon-capture probes and the number of samples for each sex in each of the lineages. As a consequence, we used only a subset of exon-capture probes and it is likely that we missed sex-specific exon-capture probes that were not included in this probeset. In addition, the lower number of samples from each sex increases the background level of differentiation between sexes, which can obscure a true signal of sex-linkage. This may be the case in *M. canariensis*, where we only found peaks of differentiation in a larger region that is likely sex-linked as well (Fig. 3.4). Similarly, despite a strong signal of differentiation in *M. huetii* across a large proportion of linkage group 1, there are still smaller interspersed regions, which show a low signal of differentiation as well as low genetic correlations in males (Fig. 3.4, Fig. 3.6). Alternatively, such a fragmented landscape of differentiation could also be the result of structural rearrangements, like inversions, translocations or chromosomal fusions, on the sex chromosomes of *M. canariensis* and *M. huetii*. Such rearrangements have been shown to be associated with the evolution of recombination suppression on sex chromosomes in other systems (Wang et al. 2012; Pennell et al. 2015), however it is not possible to test this with our current dataset.

Despite these limitations, several clear inferences about sex chromosome evolution in *Mercurialis* can nevertheless be made. Our analyses show that the sex chromosomes of diploid *M. annua* are likely located in a pericentromeric region with reduced recombination and increased repeat density (Fig. 3.1). Several other plant sex chromosome systems, where the sex determining region is located in such pericentromeric regions, are known. In papaya, the SDR, which makes up only about 15% of the size of the sex chromosome, includes the centromeres (Zhang et al. 2008) and the SDR in kiwifruit has been shown to be located in a pericentromeric region as well (Pilkington et al. 2019). Similarly, the sex-linked region in *Populus tremuloides* is located in a pericentromeric region (Kersten et al. 2014), and in *Rumex hastatulus*, recombination is restricted to the tips of chromosomes, and the SDR is located in the central, non-recombining part of the sex chromosome (Rifkin et al. 2021).

One explanation for why sex chromosomes should be located in pericentromeric regions has been that limited recombination could facilitate the evolution of separate sexes and the establishment of sex chromosomes (Charlesworth 2019). In the canonical gynodioecy model of plant sex chromosome evolution, the evolution of recombination suppression between a suppressor of female function and a male promoting variant is an essential step, as it prevents the formation of unfit recombinants (Charlesworth and Charlesworth 1978a). This model is plausible for kiwifruit, where it has been shown that sex is indeed determined by two loci with reduced recombination between them (Akagi et al. 2019). In papaya, the genetic architecture of sex determination is not known yet, but the presence of modified Y chromosomes, that result in hermaphrodites instead of males (Wang et al. 2012), could also be concordant with such

a model. The genetic architecture underlying sex determination in annual *Mercurialis* species is not known yet, but both the presence of leaky individuals of both sexes and the fact that species in the family Euphorbiaceae are either dioecious or monoecious (always lacking bisexual flowers) suggests that separate sexes evolved via monoecy in an ancestor of *Mercurialis*. In such a situation where successive mutational steps shift quantitative sex expression towards male and femaleness, reduced recombination may facilitate sex chromosome evolution as well (Charlesworth and Charlesworth 1978b). Alternatively, reduced recombination around sex determining regions could also facilitate the evolution of sexually antagonistic variants (Charlesworth and Charlesworth 1980; Rice 1987). Generally, such sexually antagonistic variants could have helped to maintain separate sexes after sex chromosomes evolved. In addition, if the evolution of sexually antagonistic variants predates the evolution of sex chromosomes, reduced recombination could also favor the evolution of separate sexes by linking sexually antagonistic variants with sterility mutations (Olito and Connallon 2019).

Our data also allows us to critically reevaluate the conclusion by Veltsos et al. (2019), that the largest part of the region of reduced recombination on the Y chromosome represents a second evolutionary stratum. Recombination rates differ between sexes in many species of plants and animals (Sardell and Kirkpatrick 2020) and the presence of an extended region of reduced recombination in males does not have to be caused by sex chromosome evolution, even if the sex determining region is located within this region. Recent discussions about the extent of recombination suppression on the guppy sex chromosome highlighted the importance of critical data analysis for identifying regions of suppressed recombination (Kirkpatrick, Dixon, et al. 2020). Based on whole genomic resequencing data, Wright et al. (2017) proposed the presence of two evolutionary strata with lengths of 3 and 7 Mb on the guppy sex chromosome. Further analysis of crosses showed the presence of strong heterochiasmy, with low recombination across the length of the chromosome (Bergero et al. 2019). However, the original results by Wright et al. were questioned by the presence of rare recombination events in a location which has previously been described as belonging to the non recombining region by Wright et al. (Charlesworth, Bergero, et al. 2020). In a recent study, Kirkpatrick et al. revisited the available evidence and concluded that the sex determining region is likely much smaller (Ĩ Mb). They argued that the statistical method used, which compares male vs. female SNP ratios with the autosomal average, is unsuitable to unambiguosly identify non-recombining regions on sex chromosomes. In addition they cautioned against overinterpreting the lack of crossing overs in mapping studies (Kirkpatrick, Dixon, et al. 2020).

The main evidence for the presence of a second evolutionary stratum on the *M. annua* Y chromosome has largely been based on the lack of recombination in a cross and on an increase of male vs. female heterozygosity. This is a similar evidence to that used by Wright et al. (2017). At the same time, a signal

of increased genetic differentiation between males and females of *M. annua* is limited to a much smaller region (Veltsos, Ridout, et al. 2019). We focused on a more robust analysis that used patterns of genetic differentiation between males and females and genetic correlations between male and female samples. Our evidence points to a much smaller region in which a clear signal of differentiation was found (Fig. 3.4, Fig. 3.5). These results are consistent with the buildup of genetic correlations in males, which were found in a similarly limited region. Unlike in *M. huetii* (Fig. 3.6), we also found a similar buildup of genetic correlations in females of diploid *M. annua* (Fig. 3.5). Such a result could be explained by the reduced effective population size on the X-linked region compared with autosomes. Overall, besides the lack of recombination in the genetic cross, our results do not find any evidence for recombination suppression across the previously proposed second evolutionary stratum and, in line with the critical approach outlined by Kirkpatrick et al. (2020), we think that it is more likely that the region of suppressed recombination is limited to a smaller region defined by molecular signatures of differentiation.

On a phylogenetic scale our results point to an extension of the non-recombining region on the Y chromosomes of *M. huetii* and *M. canariensis*. In Chapter 2, we showed that the Y chromosome of *M. canariensis* is derived from a close relative of diploid *M. annua*, which implies that the loss of recombination happened independently in *M. huetii* and *M. canariensis*. Such a pattern of repeated evolution of recombination suppression in species that share the same sex chromosome system, has rarely been shown in plants. A comparable pattern has been found in the date palm genus *Phoenix* (Torres et al. 2018). However, with approximately 5.5 Mb, the total increase in the size of the non-recombining region in *Phoenix* has been smaller than our estimates for *M. canariensis* and *M. huetii*. In addition, substantial recombination suppression may have evolved after the evolution of sex chromosomes in hemp and hops, where there is strong evidence for a shared sex chromosome with a large non-recombining region (Divashuk et al. 2014; Prentout et al. 2021). A recent study proposed the presence of an additional region of suppressed recombination, which independently evolved in both species (Prentout et al. 2021). However, this region was determined to be sex-linked based on evidence from genetic crosses, which could have resulted in an overestimate of the size of the SDR in the same way as we proposed for diploid *M. annua*. Additional population genetic data will be required to confirm the presence of evolutionary strata in this system. To our knowledge, the only other studies on plant sex chromosomes where a similar or larger extension of the SDR after the evolution of sex chromosomes has been shown are systems with fusions between sex chromosomes and autosomes (Barlow and Wiens 1976; Rifkin et al. 2021).

These results imply that recombination suppression on plant sex chromosomes can evolve in similar ways as in animal sex chromosomes after the initial evolution of separate sexes. The different *Mercurialis* lineages are a promising system to better understand the reasons for the evolution of recombination

suppression. An additional genome assembly in *M. huetii* in combination with linkage maps could allow for more detailed comparisons between the sex chromosomes in diploid *M. annua* and *M. huetii*. Such an analysis could help to understand how the large extension of the SDR evolved in *M. huetii* and it could show whether structural rearrangements or an increase of repetitive elements were involved in recombination suppression. In addition, *M. canariensis* could be useful for testing hypotheses regarding the role of haploid selection in plant sex chromosome evolution, because the pollen of tetraploid *M. canariensis* is diploid and may not underly the same selective pressures as the haploid pollen in diploid plants. In addition, hexaploid *M. annua* could be helpful to understand how sex chromosome evolution acts as a barrier to gene flow. Additional genomic data could be used to confirm the limited size of the non-recombining region in this lineage, which could have facilitated introgression of the Y chromosome from perennial lineages into hexaploid *M. annua*, which we showed in Chapter 2 of this thesis. Furthermore, such a study could help to narrow down the genes that are common to sex chromosomes of all *Mercurialis* species, and identify candidates for master sex determining genes.

References

Akagi, Takashi, Sarah M Pilkington, Erika Varkonyi-Gasic, Isabelle M Henry, Shigeo S Sugano, Minori Sonoda, Alana Firl, Mark A McNeilage, Mikaela J Douglas, Tianchi Wang, et al. (2019). "Two Y-chromosome-encoded genes determine sex in kiwifruit". In: *Nature plants* 5.8, pp. 801–809.

Anderson, Lorinda K, Gregory G Doyle, Brian Brigham, Jenna Carter, Kristina D Hooker, Ann Lai, Mindy Rice, and Stephen M Stack (2003). "High-resolution crossover maps for each bivalent of Zea mays using recombination nodules". In: *Genetics* 165.2, pp. 849–865.

Bachtrog, Doris (2013). "Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration". In: *Nature Reviews Genetics* 14.2, pp. 113–124.

Barlow, BA and D Wiens (1976). "Translocation heterozygosity and sex ratio in Viscum fischeri". In: *Heredity* 37.1, pp. 27–40.

Bergero, Roberta, Jim Gardner, Beth Bader, Lengxob Yong, and Deborah Charlesworth (2019). "Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms". In: *Proceedings of the National Academy of Sciences* 116.14, pp. 6924–6931.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30.15, pp. 2114–2120.

Branco, Sara, Hélene Badouin, Ricardo C Rodrîguez de la Vega, Jérôme Gouzy, Fantin Carpentier, Gabriela Aguileta, Sophie Siguenza, Jean-Tristan Brandenburg, Marco A Coelho, Michael E Hood, et al. (2017). "Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism". In: *Proceedings of the National Academy of Sciences* 114.27, pp. 7067–7072.

Catchen, Julian, Paul A Hohenlohe, Susan Bassham, Angel Amores, and William A Cresko (2013). "Stacks: an analysis tool set for population genomics". In: *Molecular ecology* 22.11, pp. 3124–3140.

Chalopin, Domitille, Jean-Nicolas Volff, Delphine Galiana, Jennifer L Anderson, and Manfred Schartl (2015). "Transposable elements and early evolution of sex chromosomes in fish". In: *Chromosome research* 23.3, pp. 545–560.

Charlesworth, Brian and Deborah Charlesworth (1978a). "A model for the evolution of dioecy and gynodioecy". In: *The American Naturalist* 112.988, pp. 975–997.

Charlesworth, D and B Charlesworth (1978b). "Population genetics of partial male-sterility and the evolution of monoecy and dioecy". In: *Heredity* 41.2, pp. 137–153.

Charlesworth, D and B Charlesworth (1980). "Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes". In: *Genetics Research* 35.2, pp. 205–214.

Charlesworth, Deborah (2017). "Evolution of recombination rates between sex chromosomes". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1736, p. 20160456.

Charlesworth, Deborah (2019). "Young sex chromosomes in plants and animals". In: *New Phytologist* 224.3, pp. 1095–1107.

Charlesworth, Deborah, Roberta Bergero, Chay Graham, Jim Gardner, and Lengxob Yong (2020). "Locating the sex determining region of linkage group 12 of guppy (Poecilia reticulata)". In: *G3: Genes, Genomes, Genetics* 10.10, pp. 3639–3649.

Chibalina, Margarita V and Dmitry A Filatov (2011). "Plant Y chromosome degeneration is retarded by haploid purifying selection". In: *Current Biology* 21.17, pp. 1475–1479.

Coelho, Susana M, Laure Mignerot, and J Mark Cock (2019). "Origin and evolution of sex-determination systems in the brown algae". In: *New Phytologist* 222.4, pp. 1751–1756.

Delph, Lynda F (2019). "Pollen competition is the mechanism underlying a variety of evolutionary phenomena in dioecious plants". In: *New Phytologist* 224.3, pp. 1075–1079.

Divashuk, Mikhail G, Oleg S Alexandrov, Olga V Razumova, Ilya V Kirov, and Gennady I Karlov (2014). "Molecular cytogenetic characterization of the dioecious Cannabis sativa with an XY chromosome sex determination system". In: *PloS one* 9.1, e85118.

Flynn, Jullien M, Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G Clark, Cédric Feschotte, and Arian F Smit (2020). "RepeatModeler2 for automated genomic discovery of transposable element families". In: *Proceedings of the National Academy of Sciences* 117.17, pp. 9451–9457.

Garrison, Erik and Gabor Marth (2012). "Haplotype-based variant detection from short-read sequencing". In: *arXiv preprint arXiv:1207.3907*.

Gonzalez-Martinez, Santiago C, Kate Ridout, and John R Pannell (2017). "Range expansion compromises adaptive evolution in an outcrossing plant". In: *Current Biology* 27.16, pp. 2544–2551.

Harkess, Alex, Jinsong Zhou, Chunyan Xu, John E Bowers, Ron Van der Hulst, Saravanaraj Ayyampalayam, Francesco Mercati, Paolo Riccardi, Michael R McKain, Atul Kakrana, et al. (2017). "The asparagus genome sheds light on the origin and evolution of a young Y chromosome". In: *Nature communications* 8.1, pp. 1–10.

Hughes, Jennifer F, Helen Skaletsky, Tatyana Pyntikova, Tina A Graves, Saskia KM Van Daalen, Patrick J Minx, Robert S Fulton, Sean D McGrath, Devin P Locke, Cynthia Friedman, et al. (2010). "Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content". In: *Nature* 463.7280, pp. 536–539.

Jeffries, Daniel L., Jörn F. Gerchen, Mathias Scharmann, and John R. Pannell (2021). "A neutral model for the loss of recombination on sex chromosomes". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1832, p. 20200096.

Jensen-Seaman, Michael I, Terrence S Furey, Bret A Payseur, Yontao Lu, Krishna M Roskin, Chin-Fu Chen, Michael A Thomas, David Haussler, and Howard J Jacob (2004). "Comparative recombination rates in the rat, mouse, and human genomes". In: *Genome research* 14.4, pp. 528–538.

Jombart, Thibaut (2008). "adegenet: a R package for the multivariate analysis of genetic markers". In: *Bioinformatics* 24.11, pp. 1403–1405.

Kejnovsky, E, R Hobza, Tomas Cermak, Z Kubat, and B Vyskot (2009). "The role of repetitive DNA in structure and evolution of sex chromosomes in plants". In: *Heredity* 102.6, pp. 533–541.

Kersten, B, B Pakull, K Groppe, J Lueneburg, and M Fladung (2014). "The sex-linked region in Populus tremuloides corresponds to a pericentromeric region of about two million base pairs on P. trichocarpa chromosome 19". In: *Plant Biology* 16.2, pp. 411–418.

Kirkpatrick, Mark, Groves Dixon, Jason M Sardell, Manfred Schartl, and CL Peichel (2020). "Evolution of the canonical sex chromosomes of the guppy and its relatives". In: *bioRxiv*.

Kirkpatrick, Mark, Rafael F Guerrero, and Samuel V Scarpino (2010). "Patterns of neutral genetic variation on recombining sex chromosomes". In: *Genetics* 184.4, pp. 1141–1152.

Kolmogorov, Mikhail, Jeffrey Yuan, Yu Lin, and Pavel A Pevzner (2019). "Assembly of long, error-prone reads using repeat graphs". In: *Nature biotechnology* 37.5, pp. 540–546.

Kuhl, Heiner, Yann Guiguen, Christin Hoehne, Eva Keuz, Christophe Klopp, Kang Du, Celine Lopez-Roques, Elena Santidrian Yebra-Pimentel, Mitica Ciorpac, Joern Gessner, et al. (2020). "A 180 My-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes". In: *bioRxiv*.

Kundu, Ritu, Joshua Casey, and Wing-Kin Sung (2019). "Hypo: super fast & accurate polisher for long read genome assemblies". In: *bioRxiv*.

Lahn, Bruce T and David C Page (1999). "Four evolutionary strata on the human X chromosome". In: *Science* 286.5441, pp. 964–967.

Lewontin, RC and Ken-ichi Kojima (1960). "The evolutionary dynamics of complex polymorphisms". In: *Evolution*, pp. 458–472.

Li, Heng (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM". In: *arXiv preprint arXiv:1303.3997*.

Li, Heng (2018). "Minimap2: pairwise alignment for nucleotide sequences". In: *Bioinformatics* 34.18, pp. 3094–3100.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin (2009). "The sequence alignment/map format and SAMtools". In: *Bioinformatics* 25.16, pp. 2078–2079.

Li, Xinji, Paris Veltsos, Guillaume G Cossard, Jörn Gerchen, and John R Pannell (2019). "YY males of the dioecious plant Mercurialis annua are fully viable but produce largely infertile pollen". In: *New Phytologist* 224.3, pp. 1394–1404.

Ma, Jianxin, Rod A Wing, Jeffrey L Bennetzen, and Scott A Jackson (2007). "Plant centromere organization: a dynamic structure with conserved functions". In: *TRENDS in Genetics* 23.3, pp. 134–139.

Ming, Ray, Abdelhafid Bendahmane, and Susanne S Renner (2011). "Sex chromosomes in land plants". In: *Annual review of plant biology* 62, pp. 485–514.

Müller, Niels A, Birgit Kersten, Ana P Leite Montalvão, Niklas Mähler, Carolina Bernhardsson, Katharina Bräutigam, Zulema Carracedo Lorenzo, Hans Hoenicka, Vikash Kumar, Malte Mader, et al. (2020). "A single gene underlies the dynamic evolution of poplar sex determination". In: *Nature Plants*, pp. 1–8.

Muyle, Aline, Jos Käfer, Niklaus Zemp, Sylvain Mousset, Franck Picard, and Gabriel AB Marais (2016). "SEX-DETector: a probabilistic approach to study sex chromosomes in non-model organisms". In: *Genome biology and evolution* 8.8, pp. 2530–2543.

Neely, Robert K, Jochem Deen, and Johan Hofkens (2011). "Optical mapping of DNA: Single-molecule-based methods for mapping genomes". In: *Biopolymers* 95.5, pp. 298–311.

Obbard, DJ, SA Harris, and JR Pannell (2006). "Simple allelic-phenotype diversity and differentiation statistics for allopolyploids". In: *Heredity* 97.4, pp. 296–303.

Olito, Colin and Tim Connallon (2019). "Sexually antagonistic variation and the evolution of dimorphic sexual systems". In: *The American Naturalist* 193.5, pp. 688–701.

Palmer, Daniela H, Thea F Rogers, Rebecca Dean, and Alison E Wright (2019). "How to identify sex chromosomes and their turnover". In: *Molecular ecology* 28.21, pp. 4709–4724.

Pennell, Matthew W, Mark Kirkpatrick, Sarah P Otto, Jana C Vamosi, Catherine L Peichel, Nicole Valenzuela, and Jun Kitano (2015). "Y fuse? Sex chromosome fusions in fishes and reptiles". In: *PLoS Genet* 11.5, e1005237.

Peterson, Brant K, Jesse N Weber, Emily H Kay, Heidi S Fisher, and Hopi E Hoekstra (2012). "Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species". In: *PloS one* 7.5, e37135.

Picard toolkit (2018). http://broadinstitute.github.io/picard/.

Pilkington, SM, J Tahir, E Hilario, SE Gardiner, D Chagné, A Catanach, J McCallum, L Jesson, LG Fraser, MA McNeilage, et al. (2019). "Genetic and cytological analyses reveal the recombination landscape of a partially differentiated plant sex chromosome in kiwifruit". In: *BMC plant biology* 19.1, p. 172.

Ponnikas, Suvi, Hanna Sigeman, Jessica K Abbott, and Bengt Hansson (2018). "Why do sex chromosomes stop recombining?" In: *Trends in Genetics* 34.7, pp. 492–503.

Prentout, Djivan, Natasa Stajner, Andreja Cerenak, Theo Tricou, Celine Brochier-Armanet, Jernej Jakse, Jos Käfer, and Gabriel AB Marais (2021). "Plant genera Cannabis and Humulus share the same pair of well-differentiated sex chromosomes". In: *New Phytologist*.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rastas, Pasi (2017). "Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data". In: *Bioinformatics* 33.23, pp. 3726–3732.

Renner, Susanne S (2014). "The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database". In: *American Journal of botany* 101.10, pp. 1588–1596.

Renner, Susanne S and Niels A Müller (2021). "Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration". In: *Nature Plants*, pp. 1–11.

Renner, Susanne S and Robert E Ricklefs (1995). "Dioecy and its correlates in the flowering plants". In: *American journal of botany* 82.5, pp. 596–606.

Rice, William R (1987). "The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes". In: *Evolution* 41.4, pp. 911–914.

Rifkin, Joanna L, Felix EG Beaudry, Zoë Humphries, Baharul I Choudhury, Spencer CH Barrett, and Stephen I Wright (2021). "Widespread recombination suppression facilitates plant sex chromosome evolution". In: *Molecular biology and evolution* 38.3, pp. 1018–1030.

Roach, Michael J, Simon A Schmidt, and Anthony R Borneman (2018). "Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies". In: *BMC bioinformatics* 19.1, pp. 1–10.

Sandler, George, Felix EG Beaudry, Spencer CH Barrett, and Stephen I Wright (2018). "The effects of haploid selection on Y chromosome evolution in two closely related dioecious plants". In: *Evolution letters* 2.4, pp. 368–377.

Sardell, Jason M and Mark Kirkpatrick (2020). "Sex differences in the recombination landscape". In: *The American Naturalist* 195.2, pp. 361–379.

Scott, Michael F and Sarah P Otto (2017). "Haploid selection favors suppressed recombination between sex chromosomes despite causing biased sex ratios". In: *Genetics* 207.4, pp. 1631–1649.

Seppey, Mathieu, Mosè Manni, and Evgeny M Zdobnov (2019). "BUSCO: assessing genome assembly and annotation completeness". In: *Gene prediction*, pp. 227–245.

She, Hongbing, Zhiyuan Liu, Zhaosheng Xu, Helong Zhang, Feng Cheng, Xiaowu Wang, and Wei Qian (2020). "The female (XX) and male (YY) genomes provide insights into the sex determination mechanism in spinach". In: *bioRxiv*.

Simpson, Jared T and Mihai Pop (2015). "The theory and practice of genome sequence assembly". In: *Annual review of genomics and human genetics* 16, pp. 153–172.

Skaletsky, Helen, Tomoko Kuroda-Kawaguchi, Patrick J Minx, Holland S Cordum, LaDeana Hillier, Laura G Brown, Sjoerd Repping, Tatyana Pyntikova, Johar Ali, Tamberlyn Bieri, et al. (2003). "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes". In: *Nature* 423.6942, pp. 825–837.

Tang, Haibao, Xingtan Zhang, Chenyong Miao, Jisen Zhang, Ray Ming, James C Schnable, Patrick S Schnable, Eric Lyons, and Jianguo Lu (2015). "ALLMAPS: robust scaffold ordering based on multiple maps". In: *Genome biology* 16.1, pp. 1–15.

Tomaszkiewicz, Marta, Paul Medvedev, and Kateryna D Makova (2017). "Y and W chromosome assemblies: approaches and discoveries". In: *Trends in genetics* 33.4, pp. 266–282.

Tomaszkiewicz, Marta, Samarth Rangavittal, Monika Cechova, Rebeca Campos Sanchez, Howard W Fescemyer, Robert Harris, Danling Ye, Patricia CM O'Brien, Rayan Chikhi, Oliver A Ryder, et al. (2016). "A time-and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y". In: *Genome research* 26.4, pp. 530–540.

Torres, Maria F, Lisa S Mathew, Ikhlak Ahmed, Iman K Al-Azwani, Robert Krueger, Diego Rivera-Nuñez, Yasmin A Mohamoud, Andrew G Clark, Karsten Suhre, and Joel A Malek (2018). "Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix". In: *Nature communications* 9.1, pp. 1–9.

Ubeda, Francisco, Manus M Patten, and Geoff Wild (2015). "On the origin of sex chromosomes from meiotic drive". In: *Proceedings of the Royal Society B: Biological Sciences* 282.1798, p. 20141932.

Van Berkum, Nynke L, Erez Lieberman-Aiden, Louise Williams, Maxim Imakaev, Andreas Gnirke, Leonid A Mirny, Job Dekker, and Eric S Lander (2010). "Hi-C: a method to study the three-dimensional architecture of genomes." In: *JoVE (Journal of Visualized Experiments)* 39, e1869.

Veltsos, Paris, Guillaume Cossard, Emmanuel Beaudoing, Genséric Beydon, Dessislava Savova Bianchi, Camille Roux, Santiago C Gonzàlez-Martinez, and John R Pannell (2018). "Size and content of the

sex-determining region of the Y chromosome in dioecious Mercurialis annua, a plant with homomorphic sex chromosomes". In: *Genes* 9.6, p. 277.

Veltsos, Paris, Kate E Ridout, Melissa A Toups, Santiago C Gonzàlez-Martinez, Aline Muyle, Olivier Emery, Pasi Rastas, Vojtech Hudzieczek, Roman Hobza, Boris Vyskot, et al. (2019). "Early sex-chromosome evolution in the diploid dioecious plant Mercurialis annua". In: *Genetics* 212.3, pp. 815–835.

Vicoso, Beatriz (2019). "Molecular and evolutionary dynamics of animal sex-chromosome turnover". In: *Nature ecology & evolution*, pp. 1–10.

Vicoso, Beatriz and Brian Charlesworth (2006). "Evolution on the X chromosome: unusual patterns and processes". In: *Nature Reviews Genetics* 7.8, pp. 645–653.

Wang, Jianping, Jong-Kuk Na, Qingyi Yu, Andrea R Gschwend, Jennifer Han, Fanchang Zeng, Rishi Aryal, Robert VanBuren, Jan E Murray, Wenli Zhang, et al. (2012). "Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution". In: *Proceedings of the National Academy of Sciences* 109.34, pp. 13710–13715.

Weir, Bruce S and C Clark Cockerham (1984). "Estimating F-statistics for the analysis of population structure". In: *evolution*, pp. 1358–1370.

Willard, Huntington F (1990). "Centromeres of mammalian chromosomes". In: *Trends in Genetics* 6, pp. 410–416.

Yang, Wenlu, Zhiyang Zhang, Deyan Wang, Yiling Li, Shaofei Tong, Mengmeng Li, Xu Zhang, Lei Zhang, Liwen Ren, Xinzhi Ma, et al. (2020). "A general model to explain repeated turnovers of sex determination in the Salicaceae". In: *bioRxiv*.

Yazdi, Homa Papoli, Willian TAF Silva, and Alexander Suh (2020). "Why do some sex chromosomes degenerate more slowly than others? The odd case of ratite sex chromosomes". In: *Genes* 11.10, p. 1153.

Zhang, Wenli, Xiue Wang, Qingyi Yu, Ray Ming, and Jiming Jiang (2008). "DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya". In: *Genome research* 18.12, pp. 1938–1943.

Chapter 4

QTL analyis of the genetic basis of enhanced leakiness in experimental populations of *M. annua*

4.1 Introduction

Dioecy has long been considered an evolutionary dead end (reviewed in Käfer et al. 2017). This view was based on its low frequency (Renner and Ricklefs 1995) and its scattered distribution across the plant tree of life. In addition, phylogenetic analyses found a lower species richness in dioecious clades using sister group comparisons between clades with different sexual systems (Heilbuth 2000). However, recent studies using newly-developed statistical methods challenged these results and showed that dioecious clades actually had higher diversification rates (Käfer et al. 2017) and no consistent negative effect of dioecy on diversification rates was observed in genera in which both dioecious and nondioecious species were present (Sabath et al. 2016).

Frequent reversions from dioecy towards cosexuality are a plausable mechanism that could help explain these seemingly contradictory patterns (Käfer et al. 2017; Renner 2014). Such reversions have been shown in several clades of plants (Lloyd 1975; Schaefer and Renner 2010; Volz and Renner 2008). The selective pressures that cause these reversions may be very different than those that caused transitions towards separate sexes in the first place. While separate sexes are thought to evolve because of selection for

inbreeding avoidance (Charlesworth and Charlesworth 1978a) and sexual specialization (Charnov et al. 1976), reversions towards combined sexes are caused by the capability to self fertilize, which can be promoted by either selection for reproductive assurance or automatic selection of selfing variants (Busch and Delph 2012).

The genetic changes that underly these transitions can differ substantially depending on how dioecy evolved in the first place. In the case of the canonical gynodioecy model, dioecy is determined by two loci, one of which is a recessive loss of male function mutation in females and the other is a dominant suppressor of female function in males (Charlesworth and Charlesworth 1978a). Transitions of males to hermaphroditism would require a single loss of function mutation or a recombination event with the X chromosome at the dominant suppressor of female function on the Y chromosome. In contrast, male function is not expected to be easily restored in females, and reversions are thought to happen mostly via the male sex (Crossman and Charlesworth 2014). Evidence for the gynodioecy model has been shown experimentally in *Silene latifolia*, where deletion mutations on the Y chromosome resulted in either hermaphrodites or neuters (Kazama et al. 2016). Similarly, in grapevine naturally occuring hermaphrodite lineages show evidence of either loss of expression through accumulation of transposable elements upstream of the putative suppressor of female function, or of loss of this allele through recombination with the X chromosome (Badouin et al. 2020) and in papaya, domestication selected for the Yh haplotype, which is derived from the Y chromosome, but determines plants to be hermaphrodites instead of males (VanBuren et al. 2015).

If separate sexes evolved from monoecy, different pathways for reversions to monoecy are conceivable. In persimmon, where separate sexes evolved from monoecy, maleness is determined by a Y-linked small RNA, which represses the function of an autosomal active suppressor of male function (Akagi et al. 2014). While it is not clear yet how often separate sexes evolve via such epistatic interactions, such systems could revert to combined sexes via loss of function mutations at the dominant suppressor of male function in females. Alternatively, multiple small effect mutations, which affect relative expression of flowers of both sexes, could be involved in transitions from monoecy to dioecy (Charlesworth and Charlesworth 1978b), but they could also be involved in the reverse direction and sex expression of either males or females could be shifted towards combined sexes in the same way. Such shifts may be caused by selection on variants that cause males or females to produce small numbers of flowers of the opposite sex. So-called leaky sex expression is found frequently in plants with dimorphic sexual systems, where males are more often the inconstant sex (Ehlers and Bataillon 2007; Cossard and Pannell 2019).

The theoretical framework that explains how selection could act on variation in sex expression and

causes transitions in sexual systems is sex allocation theory (West 2009). Sex allocation theory describes the conditions under which specific sexual systems are expected to be stable and under which they are susceptible to invasion of variants that influence the expression of the phenotypic sex (Charnov et al. 1976; Charlesworth and Charlesworth 1981). One of the fundamental assumptions of sex allocation theory is the existence of tradeoffs between sex functions (Charnov 1982). Most theories for the evolution of plant sexual systems include such tradeoffs, among them theories for evolution of separate sexes (Charnov et al. 1976) and those explaining the maintenance and relative contributions of male and female function in hermaphrodites (Charlesworth 1991). However, the empirical evidence supporting the presence of such tradeoffs has been limited. One of the predictions has been a negative genetic corellation between sex functions in hermaphrodites (Charlesworth 1991), but this has been shown in a limited number of plant species (Campbell 2000; Ashman 2003). So far, most of this evidence has been phenotypic and little is known about what the genetic architecture of these tradeoffs could look like. Sex allocation theory is not explicit about the genetics underlying transitions between sexual systems, but this information could help to integrate these models with population genetics.

An increasing number of loci underlying adaptive traits are being identified in a wide range of species. This allows to address a number of general questions including the maintenance of standing genetic variation and its role in adaptation, the role of convergence and pleiotropy, and that of genetic and genomic architecture in adaptive evolution (Lee et al. 2014). These questions are also important for understanding the evolutionary pathways that result in transitions between sexual systems. A major question concerning the genetic architecture of these transitions is if single large effect loci can cause large shifts in sex expression, as during the evolution of plant sex chromosomes (Ming et al. 2011), or if selection favors the evolution of multiple loci with a smaller, quantitative effect on sex expression that affect each sex function independently. In addition, dominance relationships can be important to understand how new mutations with effects on sex expression could spread and be maintained in natural populations. This may be particularly important in metapopulations like in *M. annua* (Pannell et al. 2005).

Identifying the genetic changes underlying transitions between sexual systems can be challenging. Transitions that involve mutations on the sex chromosomes can leave remainders of the previously evolved non-recomining region in the genome, which can be identified using comparative analysis with dioecious relatives (VanBuren et al. 2015; Badouin et al. 2020). However, transitions that involve autosomal loci may be much harder to identify in natural populations, especially if transitions are caused by many small effect loci. One alternative approach for both determining the effects of different selective factors and the underlying genetics is experimental evolution. Experimental evolution allows us to observe phenotypic changes under controlled conditions on contemporary timescales (Kawecki et al. 2012) and comparative

population genomic approaches between selection and control populations can be used to identify candidate loci associated with different selection regimes (Schlötterer et al. 2015). Most commonly, model organisms like *Escherichia coli* (Lenski et al. 1991), *Saccharomyces cerevisiae* (Zeyl 2006) and *Drosophila melanogaster* (Burke et al. 2010) are used for these types of experiments. Their short generation times allow to observe many generations in a single experiment and in addition their well developed genomic resources facilitate genomic analyses. However, recent experiments showed that experimental evolution can also result in significant effects on plant mating systems and related traits, despite the fact that the generation time of the plant species limited the experiments to few generations (Voillemot et al. 2019; Ramos and Schiestl 2019).

Such an experiment has been set up for diploid *M. annua* with the aim to test if pollen limitation could cause changes in the distribution of male flower production on leaky females. Leaky sex expression is also commonly found in males and females of diploid *M. annua* (Cossard and Pannell 2019) and experimental populations were set up in private gardens and on the campus of the University of Lausanne in a way that they were spatially separated from each other and naturally occuring populations of *M. annua*. The first generation of these populations was grown from seeds sampled from a natural metapopulation of *M. annua* from Spain. The experiment consisted of two experimental treatments. Control lines were made up of males and females at a 50/50 sex ratio, while selection lines consisted exclusively of genotypic females. This treatment ensured that all offspring in the selection lines were sired by pollen produced by leaky females, while the offspring in control populations was almost exclusively sired by males. Plants were left to openly pollinate, and bulk seeds were collected and used to set up the next generation in the following year (Cossard, Gerchen, et al. 2021).

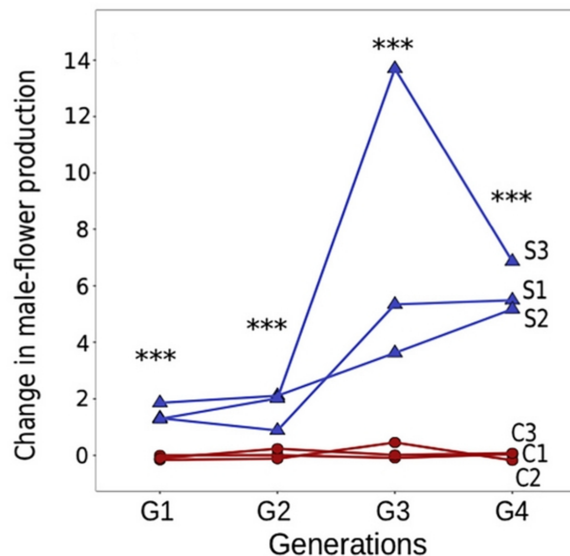


Figure 4.1: Fold change in male flower production after the first four generations in selection (S1-S3, blue) and control (C1-C3, red) lines of the experimental evolution setup. Stars indicate significant differences (*** p <0.001) in male flower production in selection lines. Figure modified from Cossard et al. (2021).

The effects of this treatments on sexual phenotypes were highly significant (Fig. 4.1). In the first generation of the experiment, leaky females in the selection lines produced a greater number of male flowers than females from the control lines (Cossard and Pannell 2021). Such an effect suggests a plastic response of sex expression on the mating environment. A similar effect has been observed in experimental populations of monoecious hexaploid *M. annua*, where plants changed their sex expression depending on population density (Dorken and Pannell 2008). In the following generations, both the number of leaky females and the number of male flowers on leaky females increased. After four generations of selection, there was a 23 fold increase in the selection lines, but no significant change in the control populations (Fig. 4.1). This is indicative of a response to selection on sex expression in a similar way as selection for reproductive assurance is expected to act under natural conditions. In an additional mating experiment, where evolved females were placed in a mating array with males, female biased offspring sex ratios showed that these evolved females were able to sire significant amounts of seeds through selfing in the presence of males (Cossard, Gerchen, et al. 2021).

These results show experimentally how reproductive assurance under pollen limitation could select for increased leakiness in females, resulting in a rapid breakdown of dioecy. In this chapter I report on a study in which we use plants derived from both control and selection lines of the previously described experiment, which has been continued for four further generations, to study the underlying genetics. A common approach to determine the genomics underlying different selection regimes in experimental evolution studies are so-called evolve and resequence setups (Kofler and Schlötterer 2014). In these studies, genomic data

of control and selection lines are sequenced at multiple timepoints. Targets of selection are identified by comparing allele frequency trajectories between selection and control lines. However, this approach works best when a large number of generations and replicates are included, which are limited in the experimental setup with *M. annua*. Due to these limitations, we chose to do a QTL mapping approach instead (Lynch and Walsh 1998).

QTL analyses identify molecular markers that correllate with variation of a quantitative trait (Mauricio 2001). However, QTLs usually explain only a subset of the phenotypic variance observed in a cross. An important source of unexplained variance can be environmental factors and phenotypic plasticity. This may be important in *M. annua*, because previous analyses suggest that leakiness can depend on the mating environment (Cossard and Pannell 2021). Furthermore, the amount of variance that can be explained also depends on the statistical power of the QTL analysis. Even with large crossing populations there is a lower limit to the effect size of QTLs that can be detected (Lynch and Walsh 1998), and QTL analyses have been critized for focussing on large effect loci, while small effect loci are not accounted for (Rockman 2012). An additional source of error that can reduce the amount of variance explained by QTL analyses are experimental errors that can occur both during phenotyping and genotyping. Also, QTL analyses can only detect variants that are present within the cross. In the context of our experimental evolution setup it is likely that these crosses do not represent all variants with effects on sex expression that are segregating within experimental populations. Despite these limitations, QTL analyses are able to directly identify genomic regions associated with phenotypic changes. This is an advantage over evolve and resequence studies, where associations are only indicative of a locus being affected by selection.

Based on a QTL analyses of two crosses between individuals from control and selection populations, we identify multiple QTLs involved in enhanced production of male flowers on leaky females. At two different loci, QTLs with antagonistic effects on male and female flower production were colocalized, which could be indicative of a tradeoff between male and female flower production. We identify additional QTLs with smaller effect sizes, which affect only male reproductive effort, without having an effect on female reproductive effort.

4.2 Material and Methods

4.2.1 Crosses

For a QTL mapping approach, we generated F_2 crosses between plants from the selection and control populations. We grew our parental plants from seeds collected from plants from the seventh generation of the experiment. Besides regular selection populations, where the next generation grown from pooled seeds from all individuals in the population, additional artificial selection populations have been established in the previous generation. Plants from the artificial selection populations were grown from seeds, which were among the top 20% of the pollen producers. The plants in this treatment showed an even stronger increase in pollen production than the regular selection populations (Xinji Li, unpublished results). As pollen donors, we chose to use plants with high pollen production from an artificial selection population (cross 1) and from a regular selection population (cross 2). As seed donors we chose female plants from one of the control populations, which did not show any leakiness (Fig. 4.2). Seeds were planted in horticultural soil and plants were grown in greenhouses until they started producing inflorescences. A single selected female as pollen donor and multiple females as ovule producers from the control lines were transferred to closed crossing boxes. In these boxes, plants were left to openly pollinate for several weeks after which seeds and tissue samples for DNA extraction were collected.

We grew F_1 plants from seeds and made crosses the same way as we did with their parents with the difference that only single individuals were transferred to crossing boxes, so that all seeds were the result of selfing. We grew F_2 plants from the resulting seeds in the greenhouse until they reached maturity. We collected, dried and weighed all male (W_{m}) and female flowers including seeds (W_{f}) and also collected, dried and weighed the remaining plant material (W_{p}). However, after phenotyping the first plants from cross 1 we realized that this phenotyping protocol would be too time intense and we changed the protocol and collected male and female flowers on every second branch. For these samples we multiplied W_{m} and W_{f} times two in the denominator of the calculations of *MRE* and *FRE*. In addition we collected tissue samples, extracted DNA using a TECAN extraction robot and quantified it using a Cybr Green protocol on a TECAN Plate reader.

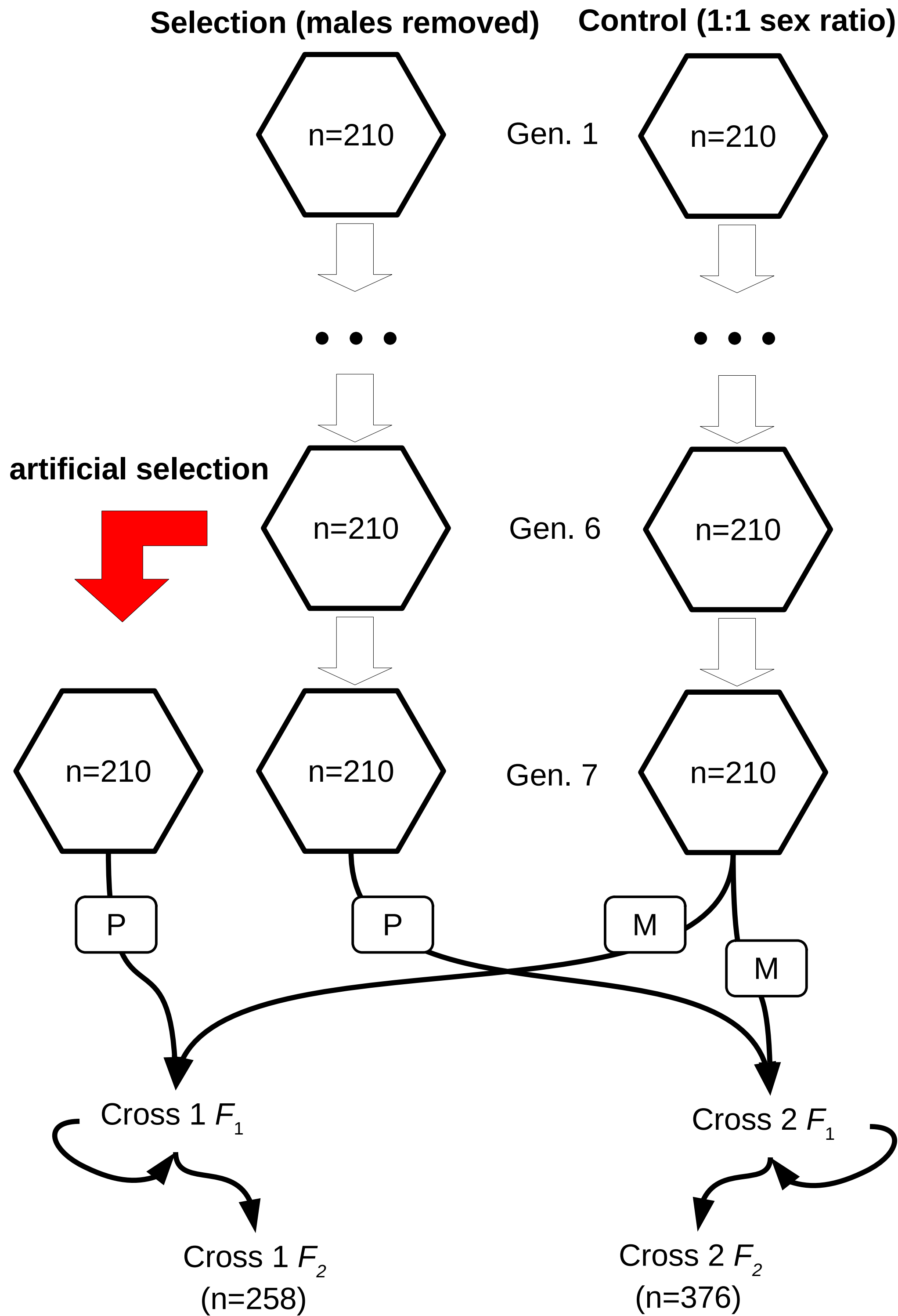


Figure 4.2: Schematic origin of the two crosses used in this study. Hexagonal shapes represent populations from the selection lines, large white arrows indicate generations founded by bulk seeds from the previous generation, the large red arrow indicates artificial selection (only seeds from the top 20 % of pollen producing females were used). Curved arrows indicate the directions of crosses, with P indicating pollen donors and M indicating maternal seed donors. Arrows pointing back on the same cross indicate selfing.

For each plant we calculated male reproductive effort, *MRE*, as the proportion of male flower biomass relative to total plant biomass,

$$MRE = \frac{W_m}{W_m + W_f + W_p}$$

female reproductive effort, *FRE*, as the proportion of female flower biomass relative to total plant biomass,

$$FRE = \frac{W_f}{W_m + W_f + W_p}$$

and relative male allocation, *rMA*, as the proportion of male reproductive effort relative to the total reproductive effort.

$$rMA = \frac{MRE}{MRE + FRE}$$

.

4.2.2 Genotyping

We generated genotype data using a modified ddRad protocol (Peterson et al. 2012). For each sample we first digested about 100 ng DNA diluted to 22 µl by adding 2.5 µl Smartcut buffer (NEB), 0.4 µl EcoRI-HF (NEB) and 0.4 µl Taq1 (NEB) restriction enzymes and incubating at 37 °C for 30 min, 65 °C for 30 min and 80 °C for 20 min. We then ligated P1 and P2 adapters described in Peterson et al. (2012), which were diluted to 40 µM in 1x annealing buffer (50mM NaCl, 10 mM). We added 3 µl rATP (10 mM), 2 µl annealed P2 adapter (3 µM), 0.8 µl 10x T4 Ligase buffer (NEB) 1 µl T4 Ligase (400 U/µl, NEB), 22 µl digested DNA and 2 µl annealed P1 adapter (0.3 µM) and incubated for 20 min at 23 °C followed by 10 min at 65 °C. We then size selected 300 µl of pooled samples for each library to 550 bp. For this we first added 0.57x volume of Ampure XP beads incubated for 10 min at room temperature and saved the supernantant, to which we added 0.12x volume of Ampure XP beads (Beckman Coulter), incubated on a magnetic stand, washed the beads with 70% ethanol and eluted DNA in 30 µl water for 2 min. We selected for biotin-labeled P2 adapters using M-270 Dynabeads (Invitrogen). We washed 15 µl Dynabeads 3 times in 1x bind and wash buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1M NaCl), resuspended them in 30 µl 2x bind and wash buffer, added the size selected DNA and incubated for 15 min at room temperature. We discarded the supernatant, washed the beads three times using 1x bind and wash buffer and resuspended the beads in 45 µl water. We then amplified size selected fragments by PCR, using dual indexed primer pairs. We added 45 µl bead suspension, 3 µl forward primer (10 µM), 3 µl reverse primer (10 µM) and 50 µl KAPA HiFi Hotstart Ready Mix (Roche). PCR program was 2 min at 95 °C initial denaturation and 11 cycles at 98 °C for 20 s, 65 °C for 20 s and 72 °C for 30 s. We cleaned PCR products in 0.7x Ampure XP

beads according to manufactorers instructions and eluted PCR products in 20 µl water. Cleaned libraries were sequenced using 150 bp paired-end reads on the Illumina Novaseq S6 platform by Novogene UK.

We demultiplexed and quality filtered raw illumina reads using Stacks 2.55 (Catchen et al. 2013) and aligned reads against the diploid female *M. annua* genome assembly using BWA mem 0.7.17-r1188 (Li 2013) and used the gstacks module from Stacks 2.55 to generate reference guided variant calls, from which we kept variants that had a minor allele frequency greater than 0.2 and that were present in 60% of all samples. We identified unrelated offspring individuals, which are likely the result of pollen contamination, by doing do a principal component analysis in adegenet 2.1.3 (Jombart 2008). We plotted the first two principal components and visually identified and removed outlier samples.

4.2.3 QTL analysis

We wrote a custom Python script, which identified alleles for which unambiguos genotype phase of the parental plants could be inferred and generated input files for R/Qtl 1.47-9 (Broman, Wu, et al. 2003) and LepMap3 0.2 (Rastas 2017). We used functions implemented in R/Qtl to further filter problematic markers, which were found in less than 50 % of samples or showed a strong signal of segregation distortion ($p<1e-10$), and individuals, which had more than 50% of missing genotypes. We then performed an initial round of separation of markers into linkage groups by adjusting minimum LOD values until markers separated into 8 major linkage groups. R/Qtl infers if the phase of the parents has been flipped due to genotyping errors. We used this information to correct genotypes of affected markers, formed new linkage groups and discarded all remaining markers that did not fall into the 8 major linkage groups. Since R/Qtl cannot efficiently determine the marker order on linkage groups containing hundreds of markers, as in our dataset, we used LepMap3 (Broman, Wu, et al. 2003) to determine marker order on the final linkage maps. We did 100 replicate estimates of marker orders for each linkage group, selected the ordering with the highest likelihood and estimated sex averaged linkage map positions. We then identified problematic markers, which are typically located at the ends of linkage groups and have a much larger estimated recombination frequency than the rest of the linkage map. We removed these types of markers if their distance to the closest marker was greater than 10 cM.

We scanned these datasets for QTLs for *MRE*, *FRE* and *rMA* using R/QTl2 0.24 (Broman, Gatti, et al. 2019). We used Haley-Knott regression to establish genotype/phenotype associations (Haley and Knott 1992) and ran a permutation test with 1000 replicates to estimate significance thresholds for LOD scores. We estimated QTL effects and identified the position with the maximum LOD peak and inferred

95% Bayesian credible intervals for each QTL. We inferred genotypes with maximum marginal probability for QTL peaks. Based on these genotypes, we estimated the mean values of phenotypes for genotypes that were heterozygotes (μ_{SC}) or homozygotes for alleles inherited from the control and the selection lines (μ_{CC} and μ_{SS}). In addition, we calculated additive and dominant effects (a and d) using the scan1coef function from R/Qtl2 and did an anova on genotypes and phenotypes to calculate the proportion of variance explained by each QTL (R^2).

4.3 Results

A total of 258 and 376 F_2 plants were used for the analyses of cross 1 and cross 2 respectively. There was substantial variation in *MRE* and *FRE* in both crosses. In addition, relative male allocation reached from 0 (completely female) to 1 (completely male) in both crosses (Fig. 4.3).

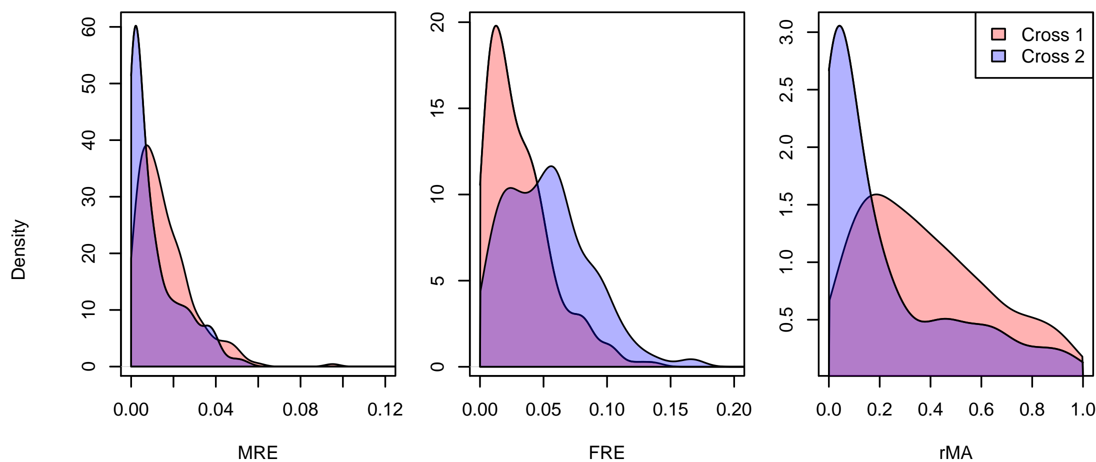


Figure 4.3: Distribution of male reproductive effort (*MRE*), female reproductive effort (*FRE*) and relative male allocation (*rMA*) of F_2 crosses.

We generated a total of 265.4 Gb and 385.4 Gb of Illumina sequencing data for cross 1 and cross 2 respectively. The results of the PCA on variant calls showed 4 and 3 outlier samples in cross 1 and cross 2 respectively (Fig. S7). We considered these samples to be the result of pollen contamination and removed them from further analyses. From each of the crosses we removed one suspicious marker from the linkage map, because of its location and distance to other markers. The final linkage maps had a total length of 640.32 cM and 653.3 cM and included 2267 and 4009 genetic markers and we retained 239 and 332 individuals for QTL analyses in cross 1 and cross 2 respectively.

For cross 1, we estimated the $p<0.05$ significance level to be at LOD scores >3.64 for all phenotypes measured. We identified QTLs with LOD scores which are clearly above this level on linkage groups 4 and 5 as well as a marginally significant QTL on linkage group 7 (Fig. 4.4). On linkage group 5, large effect QTLs were discovered for all three phenotypic measurements that we analyzed. At these three QTLs alleles inherited from the selection line had a positive and largely additive effect on male reproductive effort (*MRE*), a negative and largely recessive effect on female reproductive effort (*FRE*) and a largely additive effect on relative male allocation (*rMA*). The credible Bayesian intervals of these three QTLs overlapped (Tab. 4.1). For the additional QTLs on linkage groups 4 and 7 alleles inherited from the selection line had a negative and largely additive effect on *MRE*.

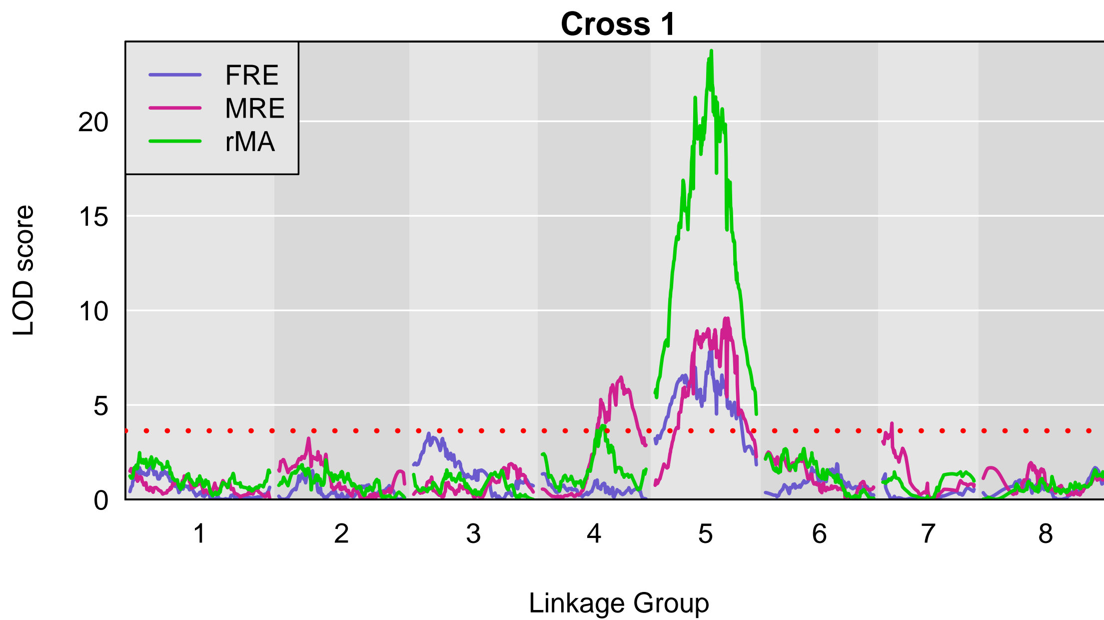


Figure 4.4: Results of the QTL analysis of male reproductive effort (*MRE*), female reproductive effort (*FRE*) and relative male allocation (*rMA*). The dotted red line indicates the $p<0.05$ significance threshold.

Table 4.1: QTLs found in cross 1. LG: linkage group on which the QTL is located. CI: Bayesian credible intervals of QTL location. μ_{SS}, μ_{SC}, μ_{CC}: Mean trait values for genotypes SS, SC and CC. Values in brackets are standard errors. *a*: additive effects. *d*: dominance effects. R^2 proportion of variance explained.

LG	Trait	Location [cM]	CI [cM]	μ_{SS}	μ_{SC}	μ_{CC}	*a*	*d*	R^2
4	*MRE*	54.86	40.75 - 62.04	0.010 (0.002)	0.015 (0.001)	0.023 (0.001)	-0.0062	-0.0018	0.12
5	*FRE*	39.26	15.29 - 47.66	0.013 (0.003)	0.033 (0.002)	0.041 (0.003)	-0.013	0.006	0.15
5	*MRE*	50.80	27.71 - 54.16	0.023 (0.002)	0.017 (0.001)	0.007 (0.002)	0.008	0.002	0.17
5	*rMA*	39.26	36.74 - 39.26	0.644 (0.028)	0.361 (0.018)	0.195 (0.028)	0.226	-0.054	0.37
7	*MRE*	6.33	0 - 16.1	0.012 (0.002)	0.016 (0.001)	0.026 (0.003)	-0.007	-0.003	0.07

For cross 2, we estimated the $p<0.05$ significance level at LOD scores >3.72 for all phenotypes measured. We found QTLs with significantly higher LOD scores on linkage groups 4 and 7 (Fig. 4.5, Tab. 4.2). On linkage group 4, we identified marginally significant QTLs, where alleles inherited from the selection line had mostly additive positive effects on *rMA* and and a mostly dominant positive effect on *MRE*. On linkage group 7 we identified major QTLs for *FRE*, *MRE* and *rMA* , all of which are in relative proximity, and the credible intervals between *rMA* and *FRE* as well as between *rMA* and *MRE*, but not between *MRE* and *FRE* overlap. Alleles inherited from the selection line had an additive negative effect on *FRE*, a partly recessive positive effect on *MRE* and a partly recessive positive effect on *rMA* (Tab. 4.2).

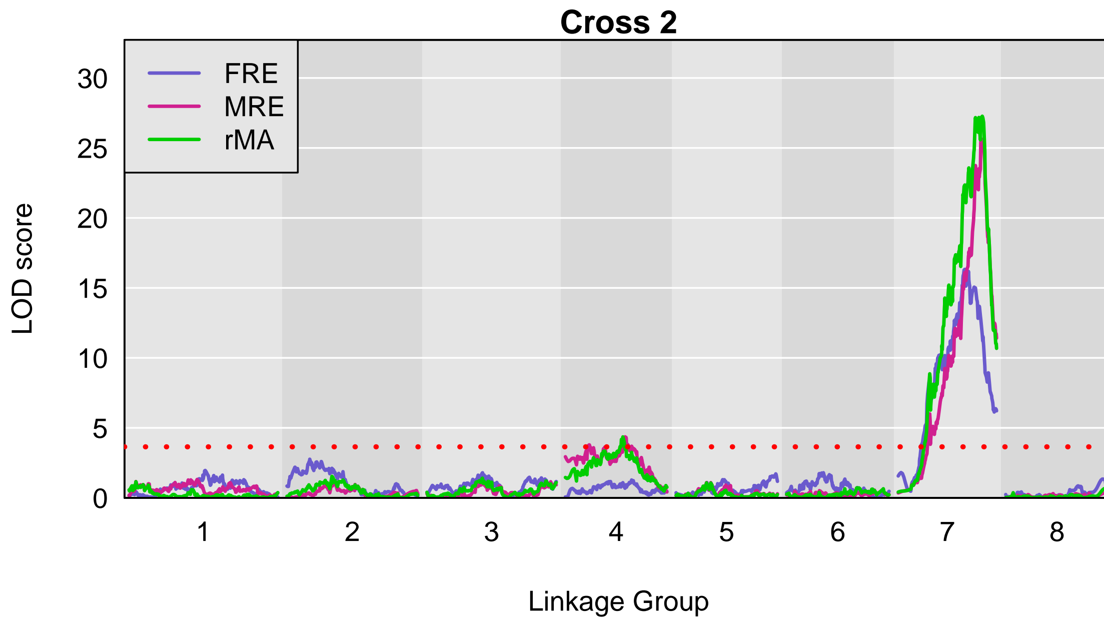


Figure 4.5: Results of the QTL analysis of male reproductive effort (*MRE*), female reproductive effort (*FRE*) and relative male allocation (*rMA*). The dotted red line indicates the $p<0.05$ significance threshold.

Table 4.2: QTLs found in cross 2. LG: linkage group on which the QTL is located. CI: Bayesian credible intervals of QTL location. μ_{SS}, μ_{SC}, μ_{CC}: Mean trait values for genotypes SS, SC and CC. Values in brackets are standard errors. *a*: additive effects. *d*: dominance effects. R^2 proportion of variance explained.

LG	Trait	Location [cM]	CI [cM]	μ_{SS}	μ_{SC}	μ_{CC}	a	d	R^2
4	*MRE*	42.55	0 - 54.55	0.014 (0.001)	0.012 (0.001)	0.006 (0.001)	0.004	0.002	0.06
4	*rMA*	41.03	17.19 - 51.20	0.313 (0.028)	0.209 (0.022)	0.132 (0.030)	0.090	-0.014	0.06
7	*FRE*	47.53	45.70 - 55.15	0.030 (0.004)	0.053 (0.002)	0.073 (0.003)	-0.022	0.002	0.21
7	*MRE*	60.00	58.94 - 61.84	0.023 (0.001)	0.008 (0.001)	0.005 (0.001)	0.009	-0.006	0.31
7	*rMA*	60.00	54.99 - 61.84	0.499 (0.026)	0.159 (0.017)	0.098 (0.026)	0.197	-0.143	0.33

4.4 Discussion

We observed substantial variation in male and female reproductive effort (*MRE* and *FRE*) as well as in relative male allocation (*rMA*) in both F_2 crossing lines analyzed in this study (Fig. 4.4, Tab. 4.1, Fig. 4.5, Tab. 4.2). We identified 10 QTLs which showed associations with this variation in sexual phenotypes. In both crosses, major QTLs with antagonistic effects on female and male reproductive effort were located on the same linkage groups. Unsurprisingly, this was also true for relative male allocation, however this trait is based on the relative proportion of *MRE* and *FRE*, which means that it is expected to be colocalized with both of these traits. In cross 1, the Bayesian credible intervals of QTLs for *FRE* and *MRE* overlapped (Tab. 4.1). This was not the case for *FRE* and *MRE* in cross 2, but the distance between both confidence intervals was just 3.79 cM (Tab. 4.2). These results show that the major QTLs for *FRE* and *MRE* in both crosses could comprise either single loci that have an antagonistic effect on both *FRE* and *MRE* or multiple loci that affect both traits, which are in close linkage.

Besides these main effect QTLs we detected additional QTLs, that affected male flower production without having a significant effect on female flower production. In both crosses, one of these minor QTLs was located on linkage group 4 (Tab. 4.1, Tab. 4.2). In addition, their Bayesian credible intervals overlapped, however this result has to be interpreted with caution, as this location refers to two independent linkage maps. Both additional QTLs identified in Cross 1 had a negative association with *MRE*, and individuals that inherited both alleles from the control population had a greater *MRE* than those that inherited it from the selection populations (Tab. 4.1). One interpretation would be that loci with a negative effect on male flower production would have been segregating in the selection population, from which the pollen donor in Cross 1 originated. This interpretation seems counterintuitive, since selection is expected to favor variants with an opposite effect in the experimental evolution setup. However, despite a substantial increase in male flower production in leaky females, there were still pure females (that produced no male flowers) present in the experimental populations after four generations (Cossard and Pannell 2021). Alleles that affect male flower production would only be visible to selection via the male function and given the limited number of generations in the experiment it seems conceivable that such alleles could still be segregating in the experimental populations. Alternatively, alleles that have positive effects on *MRE* could have been inherited from the control population. Low frequencies of leaky females were found in control populations, and the alleles underlying leakiness could have been sampled by chance. We did not observe any leakiness in the maternal seed donors from the control population, but leakiness did show a plastic response to pollen limitation in the initial populations of the experimental evolution setup (Cossard and Pannell 2021). Given that the pollen donors in the crosses were individuals that produced large amounts

of pollen, alleles that favor leakiness may not have been expressed in these individuals. Our experimental design did not allow direct comparisons of the effects of phenotypes of parents and offspring, and as a consequence we could not distinguish between these interpretations.

Individual QTLs identified in our study explained between 6 and 31% of the observed variance in *MRE*, between 15 and 21% of the observed variance in *FRE* and between 6 and 37% of the observed variance in *rMA*. These results are in the range of other QTLs discovered in plants, although individual QTLs explaining more than 20% of the phenotypic variance can be considered to be located in the upper tail of the distribution (Kearsey and Farquhar 1998). A particularly large effect size is observed for the major QTLs for *rMA*. The mean difference in *rMA* between individuals that where homozygous for alleles inherited from the control population and those that were homozygous for alleles inherited from the selection populations at the large effect loci was greater than 0.4. This means that in these individuals on average more than 40% of the total drymass invested in reproduction are invested in male instead of female reproduction. This indicates that each of the two main QTLs for *rMA* would be sufficient to cause substantial shifts in the sexual systems by itself.

While these results imply that leakiness has a strong, heritable genetic basis our analysis still leaves a large proportion of residual variance unexplained. Taken together, the QTLs identified in this study explain between 21% and 39 % of the variation of a specific trait in the offspring from a single cross. A likely reason is that further QTLs of smaller effect exist, but our analysis does not have sufficient statistical power to detect them. Potential examples of this effect can be found on linkage groups 2 and 3 in cross 1, where we did not consider smaller peaks in *MRE* and *FRE*, which could be indicative of additional QTLs, because they did not reach the threshold of statistical significance (Fig. 4.4). An additional likely source of unaccounted variation in our study are environmental factors and developmental plasticity. Male flower production in leaky females has been shown to have a strong plastic component, which depends on the degree of pollen limitation (Cossard and Pannell 2021). In addition, other environmental factors could contribute to variation in sexual phenotypes in *Mercurialis*. For example, simulated herbivory has been shown to increase sex inconstancy in both males and females of *Mercurialis* (Villamil et al. 2021), and personal observations suggest that the expression of leaky phenotypes can differ between plants that were grown in greenhouses or under artificial light.

Our results allow to connect phenotypic measures of sex expression with the underlying genetics, which can be understood in the context of sex allocation. In our experimental populations, sex allocation refers to the absolute amounts of reproductive effort invested in inflorescences of each sex. In addition, the relative allocation of male and female reproductive effort indicates if an increase in the number of inflorescences of

one sex will result in a decrease in the number of flowers of the opposite sex. We identified a number of QTLs and their effect sizes and dominances, which are associated with changes in sex allocation in *M. annua*. The patterns of antagonistic effects on male and female flower production at the major effect loci, which cause a shift in relative sex allocation, are consistent with the presence of tradeoffs predicted by sex allocation theory (Charnov et al. 1976). However, we also find additional QTLs with an effect on *MRE*, but without apparent effects on *FRE* (Tab. 4.1), which does not seem to agree with a tradeoff model. The resolution of our dataset is limited both by the number of recombination events within the crosses and by the density of Rad-loci used for genotyping. As a result, the QTLs identified in our study could also contain multiple loci with effects on sexual phenotypes. These could be either few large effect loci with effects on *FRE*, *MRE* or both, or many small effect loci with cumulative effects on *FRE* and *MRE*, that are enriched in genomic regions represented by the QTLs found in our study. Due to these limitations, we can't precisely identify the genes involved in these changes in sexual phenotypes with our current dataset and different scenarios are possible.

The first scenario for the genes underlying these QTLs is that they could be controlling which proportion of meristems develops into male and female inflorescences. This could happen by affecting the relative ratios of auxin and cytokinin, which have been proposed to be involved in the sexual differentiation of inflorescences in *M. annua* (Khadka et al. 2019). Such a genetic architecture would be a parsimonious explanation for the function of the major effect loci with antagonistic effects on male and female sex expression. However, we also found additional QTLs that only affect *MRE* without significant effects on *FRE*, which seems to contradict this hypothesis. The effect size of these loci is smaller, and it is conceivable that the relationship of the antagonistic effects between *FRE* and *MRE* could be non-linear, so that at smaller effect sizes an increase of male flower production could happen without affecting production of female flowers. This could be concordant with results from the experimental evolution setup that found no consistent evidence of a tradeoff between male and female flower production on the population level (Cossard and Pannell 2021; Cossard, Gerchen, et al. 2021). A second scenario implies that different molecular pathways could be affecting male flower production, some of which have antagonistic effects on *FRE*, while others don't. For example, the loci with antagonistic effects on *MRE* and *FRE* could be replacing female flowers with male flowers, while the loci that only have an effect on *MRE* could increase male flower production in locations where no other female flowers would have grown otherwise. A third scenario would imply that loci that affect *FRE* and *MRE* are different types of loci, which are in close linkage at the major effect sites and have a combined effect on *rMA*. Such a scenario could be comparable with the model for the evolution of separate sexes in plants, where suppressed recombination links variants with antagonistic effects on male and female sex expression (Charlesworth and Charlesworth

1978a; Charlesworth and Charlesworth 1978b). However, the selective forces that cause selection and linkage of these variants are expected to be different than those that cause the breakdown of dioecy. The evolution of separate sexes is expected to be driven by evolution for outcrossing (Charlesworth and Charlesworth 1981) or sexual specialization (Charnov et al. 1976). In contrast, transitions towards combined sexual systems are expected to be driven by selection for reproductive assurance or by automatic selection of variants that increase selfing (Busch and Delph 2012). In the absence of any antagonistic effects on *MRE* it seems unlikely that selection would favor alleles that decrease *FRE*, since both selection for reproductive assurance and automatic selection would favor some degree of selfing, which would imply selection for increased investment in female function (Lloyd 1987). Distinguishing between these scenarios would require additional fine mapping of these loci and the use of improved genomic resources. A promising approach to find potential candidate genes in could be to generate additional genomic sequencing data or gene expression data for individuals which are homozygous for either alleles inherited from control or selection lines. Sequence variation or significant differences in gene expression in the regions surrounding the inferred QTL peaks of main effect loci could be indicatitive of potential candidate loci. While QTL analyses are generally considered to have more limited power than other genomic analysis for precisely locating genes of interest, studies where the location of genes was known showed a high concordance with inferred QTL locations (Price 2006) and the credible intervals around our identified QTLs with effects on relative male allocation could be small enough to make such an approach feasible.

The fact that these main QTLs identified in both crosses are located on different chromosomes suggests that multiple independent large effect QTLs are segregating in natural metapopulations of *M. annua*. It seems surprising that such large effect variants should be maintained in dioecious populations, where it is unlikely that the limited amount of pollen produced on leaky females could compete with the strong outcrossing advantage of males, while the negative effect on female flower production would decrease reproductive success through female function. A possible explanation could be that females of diploid *M. annua* frequently experience pollen limitation in low density situations, which happen due to metapopulation processes especially during range expansions (Gonzalez-Martinez et al. 2017). In these situations, leakiness could provide reproductive assurance through selfing. In addition, the phenotypic effects of these two main QTLs on the reduction of female reproductive effort are largely recessive in cross 1 (Tab. 4.1) and additive in cross 2 (Tab. 4.2). This implies that the negative effect on female flower production may be moderate when these loci are heterozygous, which could facilitate the maintenance of these alleles in dioecious populations. However, in situations of pollen limitation leaky individuals can sire a great number of offspring through their male function either through selfing or by outcrossing with other females. As a result, homozygosity at these loci will increase and populations could rapidly shift towards a selfing

sexual system, similar as in the experimental populations. This genetic architecture suggests that the sex expression of these leaky females could also revert quickly if pollen from males from dioecious populations starts invading these populations. Individual based modeling could help to show under which conditions loci with the observed effect sizes and dominance could be maintained in natural metapopulations.

Our results also have implications for our understanding of the evolution of sex chromosomes and separate sexes in *M. annua*. There are no significant QTLs located on linkage group 1 in either cross, which are the sex chromosomes (Fig. 4.4, Fig. 4.5). This means that the sex chromosomes do not play a major role in the transition towards monoecy in the crosses analyzed in this study. In other plant species, reversions towards combined sexes have been shown to be involved with mutations or recombination events on the Y chromosomes (VanBuren et al. 2015; Kazama et al. 2016; Badouin et al. 2020). These reversions are largely concordant with the canonical model of plant sex chromosome evolution, in which a recessive male sterility mutations becomes linked with a dominant suppressor of male function (Charlesworth and Charlesworth 1978a). In contrast, we showed that QTLs located on the autosomes can have a major effect on sex expression in *M. annua*. At the same time we observe substantial leakiness in both sexes of *M. annua* (Cossard and Pannell 2019). These patterns disagree with the presence of male sterility mutations on the sex chromosomes. However, such a system could be concordant with the monoecy model, where selection on variants with quantitative effects on male and female sex expression can cause a gradual shift towards separate sexes (Charlesworth and Charlesworth 1978b). Our results suggests that some of these variants could have rather large efffect sizes (Fig. 4.1, Fig. 4.2), and may have antagonistic effects on male and female sex expression. It is conceivable that similar variants with a greater effect size or linkage of multiple variants with additive effect could cause a complete shift to maleness in *M. annua*. However, the fact that these variants are either recessive or codominant suggests that the molecular pathways underlying the production of male flowers on leaky females could be different than those underlying dominant male determining genes located on Y chromosomes in males of *M. annua*.

References

Akagi, Takashi, Isabelle M Henry, Ryutaro Tao, and Luca Comai (2014). "A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons". In: *Science* 346.6209, pp. 646–650.

Ashman, Tia-Lynn (2003). "Constraints on the evolution of males and sexual dimorphism: field estimates of genetic architecture of reproductive traits in three populations of gynodioecious Fragaria virginiana". In: *Evolution* 57.9, pp. 2012–2025.

Badouin, Hélène, Amandine Velt, François Gindraud, Timothée Flutre, Vincent Dumas, Sonia Vautrin, William Marande, Jonathan Corbi, Erika Sallet, Jérémy Ganofsky, et al. (2020). "The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication". In: *Genome biology* 21.1, pp. 1–24.

Broman, Karl W, Daniel M Gatti, Petr Simecek, Nicholas A Furlotte, Pjotr Prins, Śaunak Sen, Brian S Yandell, and Gary A Churchill (2019). "R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations". In: *Genetics* 211.2, pp. 495–502.

Broman, Karl W, Hao Wu, Śaunak Sen, and Gary A Churchill (2003). "R/qtl: QTL mapping in experimental crosses". In: *Bioinformatics* 19.7, pp. 889–890.

Burke, Molly K, Joseph P Dunham, Parvin Shahrestani, Kevin R Thornton, Michael R Rose, and Anthony D Long (2010). "Genome-wide analysis of a long-term evolution experiment with Drosophila". In: *Nature* 467.7315, pp. 587–590.

Busch, Jeremiah W and Lynda F Delph (2012). "The relative importance of reproductive assurance and automatic selection as hypotheses for the evolution of self-fertilization". In: *Annals of botany* 109.3, pp. 553–562.

Campbell, Diane R (2000). "Experimental tests of sex-allocation theory in plants". In: *Trends in Ecology & Evolution* 15.6, pp. 227–232.

Catchen, Julian, Paul A Hohenlohe, Susan Bassham, Angel Amores, and William A Cresko (2013). "Stacks: an analysis tool set for population genomics". In: *Molecular ecology* 22.11, pp. 3124–3140.

Charlesworth, Brian and Deborah Charlesworth (1978a). "A model for the evolution of dioecy and gynodioecy". In: *The American Naturalist* 112.988, pp. 975–997.

Charlesworth, D and B Charlesworth (1978b). "Population genetics of partial male-sterility and the evolution of monoecy and dioecy". In: *Heredity* 41.2, pp. 137–153.

Charlesworth, D and B Charlesworth (1981). "Allocation of resources to male and female functions in hermaphrodites". In: *Biological Journal of the Linnean Society* 15.1, pp. 57–74.

Charlesworth, Deborah (1991). "Allocation of resources to sex functions in flowering plants". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 332.1262, pp. 91–102.

Charnov, Eric L (1982). "The theory of sex allocation. Princeton Monograph". In: *Population Biology* 18.

Charnov, Eric L, James J Bull, and J Maynard Smith (1976). "Why be an hermaphrodite?" In: *Nature* 263.5573, pp. 125–126.

Cossard, Guillaume G, Jörn F Gerchen, Xinji Li, Yves Cuenot, and John R Pannell (2021). In: *Current Biology* 31.6, pp. 1277–1283.

Cossard, Guillaume G and John R Pannell (2019). "A functional decomposition of sex inconstancy in the dioecious, colonizing plant Mercurialis annua". In: *American journal of botany* 106.5, pp. 722–732.

Cossard, Guillaume G and John R Pannell (2021). "Enhanced leaky sex expression in response to pollen limitation in the dioecious plant Mercurialis annua". In: *Journal of Evolutionary Biology* 34.2, pp. 416–422.

Crossman, Allan and Deborah Charlesworth (2014). "Breakdown of dioecy: models where males acquire cosexual functions". In: *Evolution* 68.2, pp. 426–440.

Dorken, Marcel E and John R Pannell (2008). "Density-dependent regulation of the sex ratio in an annual plant". In: *The American Naturalist* 171.6, pp. 824–830.

Ehlers, Bodil K and Thomas Bataillon (2007). "'Inconstant males' and the maintenance of labile sex expression in subdioecious plants". In: *New Phytologist* 174.1, pp. 194–211.

Gonzalez-Martinez, Santiago C, Kate Ridout, and John R Pannell (2017). "Range expansion compromises adaptive evolution in an outcrossing plant". In: *Current Biology* 27.16, pp. 2544–2551.

Haley, Chris S and Sarah A Knott (1992). "A simple regression method for mapping quantitative trait loci in line crosses using flanking markers". In: *Heredity* 69.4, pp. 315–324.

Heilbuth, Jana C (2000). "Lower species richness in dioecious clades". In: *The American Naturalist* 156.3, pp. 221–241.

Jombart, Thibaut (2008). "adegenet: a R package for the multivariate analysis of genetic markers". In: *Bioinformatics* 24.11, pp. 1403–1405.

Käfer, Jos, Gabriel AB Marais, and John R Pannell (2017). "On the rarity of dioecy in flowering plants". In: *Molecular Ecology* 26.5, pp. 1225–1241.

Kawecki, Tadeusz J, Richard E Lenski, Dieter Ebert, Brian Hollis, Isabelle Olivieri, and Michael C Whitlock (2012). "Experimental evolution". In: *Trends in ecology & evolution* 27.10, pp. 547–560.

Kazama, Yusuke, Kotaro Ishii, Wataru Aonuma, Tokihiro Ikeda, Hiroki Kawamoto, Ayako Koizumi, Dmitry A Filatov, Margarita Chibalina, Roberta Bergero, Deborah Charlesworth, et al. (2016). "A new physical mapping approach refines the sex-determining gene positions on the Silene latifolia Y-chromosome". In: *Scientific reports* 6.1, pp. 1–10.

Kearsey, MJ and AGL Farquhar (1998). "QTL analysis in plants; where are we now?" In: *Heredity* 80.2, pp. 137–142.

Khadka, Janardan, Narendra Singh Yadav, Micha Guy, Gideon Grafi, and Avi Golan-Goldhirsh (2019). "Epigenetic aspects of floral homeotic genes in relation to sexual dimorphism in the dioecious plant Mercurialis annua". In: *Journal of experimental botany* 70.21, pp. 6245–6259.

Kofler, Robert and Christian Schlötterer (2014). "A guide for the design of evolve and resequencing studies". In: *Molecular biology and evolution* 31.2, pp. 474–483.

Lee, Young Wha, Billie A Gould, and John R Stinchcombe (2014). "Identifying the genes underlying quantitative traits: a rationale for the QTN programme". In: *AoB Plants* 6.

Lenski, Richard E, Michael R Rose, Suzanne C Simpson, and Scott C Tadler (1991). "Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations". In: *The American Naturalist* 138.6, pp. 1315–1341.

Li, Heng (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM". In: *arXiv preprint arXiv:1303.3997*.

Lloyd, David G (1975). "Breeding systems in Cotula IV. Reversion from dioecy to monoecy". In: *New Phytologist* 74.1, pp. 125–145.

Lloyd, DG (1987). "Allocations to pollen, seeds and pollination mechanisms in self-fertilizing plants". In: *Functional Ecology*, pp. 83–89.

Lynch, Michael and Bruce Walsh (1998). *Genetics and analysis of quantitative traits*. Vol. 1. Sinauer Sunderland, MA.

Mauricio, Rodney (2001). "Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology". In: *Nature Reviews Genetics* 2.5, pp. 370–381.

Ming, Ray, Abdelhafid Bendahmane, and Susanne S Renner (2011). "Sex chromosomes in land plants". In: *Annual review of plant biology* 62, pp. 485–514.

Pannell, John R, Marcel E Dorken, and Sarah M Eppley (2005). "'Haldane's Sieve'in a metapopulation: sifting through plant reproductive polymorphisms". In: *Trends in ecology & evolution* 20.7, pp. 374–379.

Peterson, Brant K, Jesse N Weber, Emily H Kay, Heidi S Fisher, and Hopi E Hoekstra (2012). "Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species". In: *PloS one* 7.5, e37135.

Price, Adam H (2006). "Believe it or not, QTLs are accurate!" In: *Trends in plant science* 11.5, pp. 213–216.

Ramos, Sergio E and Florian P Schiestl (2019). "Rapid plant evolution driven by the interaction of pollination and herbivory". In: *Science* 364.6436, pp. 193–196.

Rastas, Pasi (2017). "Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data". In: *Bioinformatics* 33.23, pp. 3726–3732.

Renner, Susanne S (2014). "The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database". In: *American Journal of botany* 101.10, pp. 1588–1596.

Renner, Susanne S and Robert E Ricklefs (1995). "Dioecy and its correlates in the flowering plants". In: *American journal of botany* 82.5, pp. 596–606.

Rockman, Matthew V (2012). "The QTN program and the alleles that matter for evolution: all that's gold does not glitter". In: *Evolution* 66.1, pp. 1–17.

Sabath, Niv, Emma E Goldberg, Lior Glick, Moshe Einhorn, Tia-Lynn Ashman, Ray Ming, Sarah P Otto, Jana C Vamosi, and Itay Mayrose (2016). "Dioecy does not consistently accelerate or slow lineage diversification across multiple genera of angiosperms". In: *New Phytologist* 209.3, pp. 1290–1300.

Schaefer, Hanno and Susanne S Renner (2010). "A three-genome phylogeny of Momordica (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia". In: *Molecular phylogenetics and evolution* 54.2, pp. 553–560.

Schlötterer, Christian, Robert Kofler, E Versace, R Tobler, and SU Franssen (2015). "Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation". In: *Heredity* 114.5, pp. 431–440.

VanBuren, Robert, Fanchang Zeng, Cuixia Chen, Jisen Zhang, Ching Man Wai, Jennifer Han, Rishi Aryal, Andrea R Gschwend, Jianping Wang, Jong-Kuk Na, et al. (2015). "Origin and domestication of papaya Yh chromosome". In: *Genome research* 25.4, pp. 524–533.

Villamil, Nora, Xinji Li, Emily Seddon, and John Pannell (2021). "Herbivory enhances leaky sex expression in the dioecious herb Mercurialis annua". In: *bioRxiv*.

Voillemot, Marie, Francisco Encinas-Viso, and John R Pannell (2019). "Rapid loss of self-incompatibility in experimental populations of the perennial outcrossing plant Linaria cavanillesii". In: *Evolution* 73.5, pp. 913–926.

Volz, Stefanie M and Susanne S Renner (2008). "Hybridization, polyploidy, and evolutionary transitions between monoecy and dioecy in Bryonia (Cucurbitaceae)". In: *American Journal of Botany* 95.10, pp. 1297–1306.

West, Stuart (2009). *Sex allocation*. Vol. 44. Princeton University Press.

Zeyl, Clifford (2006). "Experimental evolution with yeast". In: *FEMS yeast research* 6.5, pp. 685–691.

Chapter 5

Conclusions

In this thesis I adressed questions regarding the genetics underlying transitions between sexual systems in plants. The majority of angiosperms have combined sexes (The Tree of Sex Consortium 2014) and this is also their most likely ancestral state (Sauquet et al. 2017). Only about 6% of angiosperms are dioecious, however these are found in more than 40% of families and it has been estimated that they evolved up to 5000 times independently (Renner 2014). In addition, there is evidence that transitions from separate towards combined sexes are common as well (Käfer et al. 2017). Currently we know little about the genomics underlying these transitions, however this knowledge will be essential to understand in detail how major evolutionary transitions, like those between sexual systems, work. The aim of this thesis was to better understand the genomics underlying sex determination in dioecious plants, but also those underlying transitions from dioecy towards monoecy and androdioecy. In addition I explored how two other important genomic and population genomic processes in plants, polyploidization and hybridization, can be involved in these processes. I used plants from the genus *Mercurialis* as a model system, because while dioecy is ancestral in *Mercurialis*, there have been multiple transitions towards monoecy and androdioecy in this clade (Krähenbühl et al. 2002; Obbard et al. 2006). I analyzed a number of different genomic datasets to further understand the genomics underlying these transitions.

One important genomic component involved in transitions between sexual systems in plants are sex chromosomes (Ming et al. 2011). Unlike in most animals, where sex chromosomes commonly evolve via sex chromosome turnover (Vicoso 2019), the evolution of sex chromosomes in plants is often a direct consequence of the evolution of separate sexes (Charlesworth 2002). Besides their role in the evolution of separate sexes, sex chromosomes have also been shown to be involved in the breakdown of dioecy (Crossman and Charlesworth 2014; Wang et al. 2012; Zou et al. 2021). In addition, their haploid mode of

sex-specific inheritance can cause the sex-linked regions to differ in their evolutionary trajectories from the rest of the genome, which can have a number of effects on individual fitness and genetic diversity within and between populations (Bachtrog 2013). Understanding the genomic and population genetic consequences of the evolution of sex chromosomes is essential for understanding the evolution of separate sexes as well as reversions towards combined sexes. I adressed several questions regarding sex chromosome evolution in Chapter 2 and Chapter 3.

In Chapter 2, I asked whether there are circumstances under which plant sex chromosomes can introgress between lineages. In addition, I wanted to explore what role allopolyploidization and introgression can play in transitions between sexual systems. To this end I developed a new set of phylogenetic markers for annual and perennial lineages of *Mercurialis*. Based on these markers I showed that the two main annual and perennial clades share the same ancestral sex chromosomes (Fig. S1, Fig. S2). I also found that the history of polyploid *Mercurialis* lineages is more complex than previously thought. I discovered multiple previously unknown allopolyploidization events and I showed that all polyploid lineages in this dataset are allopolyploids (Fig. 2.5). In addition, I found a signal of Y chromosome introgression from one of the perennial lineages (most likely *M. elliptica*) into hexaploid *M. annua* (Fig. 2.3, Fig. 2.5).

Such a pattern of sex chromosome introgression has rarely been shown, and it is usually assumed that sex chromosomes are disproportionately involved in reproductive isolation (Coyne 2018). In this case there may be multiple factors which facilitated introgression of the Y chromosome. The fact that only a single male-specific marker, but none of the markers located across the length of the sex chromosome, showed a signal of introgression implies that the size of the non-recombining region could be very small. This observation is concordant with the observations from Chapter 3 of this thesis, where I found only a single male-specific exon-capture probe in hexaploid *M. annua* (Fig. 3.3). Empirical observations suggest that the degree of reproductive barriers is greater on heteromorphic sex chromosomes (Lima 2014), which implies that sex chromosomes with a small non-recombining region, like in hexaploid *M. annua*, could be less likely to be involved in reproductive isolation.

Another important factor in this system may be polyploidy. The dominance hypothesis explains the commonly observed reduction of fitness in offspring of the heterogametic sex in interspecific crosses (Haldane's Rule (Haldane 1922)) with the effect of recessive deleterious mutations, which are expressed if a functional copy on one of the sex chromosomes is missing (Turelli and Orr 1995). In polyploids, multiple redundant functional copies can be expressed (Comai 2005). This means that additional genome copies could shelter recessive deleterious variants even if a functional copy is missing on one of the sex chromosomes. Differences in ploidy are also thought to cause reproductive isolation by itself. This may

be important to understand how the Y chromosome introgressed from perennial *Mercurialis* lineages into hexaploid *M. annua*. The perennial lineage, which is the most likely origin of the Y chromosome in hexaploid *M. annua*, is perennial *M. elliptica*. Both of these lineages are hexaploids, and in addition both of them contain a subgenome that is closely related to diploid *M. huetii* (Fig. 2.5). Taken together polyploidy may have either facilitated introgression in hexaploid *M. annua*, or at least similarities in ploidy and genome content may have avoided the formation of reproductive barriers commonly observed in interploid hybrids.

The introgression of the Y chromosome into hexaploid *M. annua* probably caused a transition from monoecy to androdioecy. Due to the strong outcrossing advantage of males in monoecious populations (Eppley and Pannell 2007), Y chromosomes are expected to be under strong positive selection. This would make Y chromosome introgression into these populations more likely than into dioecious populations, where introgressing Y chromosomes would have to replace other male determiners already present in the population. It has been previously hypothesized that hybridization may have played an important role in the evolution of the few known androdioecious plants (Yakimowski and Barrett 2016). My results show that this has likely been the case in hexaploid *M. annua*.

Overall, Chapter 2 shows that there are a number of factors that can potentially cause plant sex chromosomes to behave differently than what is expected from observations based on diploid, bisexual animal systems. Both the small size of the non-recombining region and polyploidy may have contributed to facilitate introgression. However, it was possibly positive selection involved with a transition from monoecy towards androdioecy that had the greatest effect in causing the pattern of sex chromosome introgression in this system. The importance of introgression for adaptive evolution is being increasingly recognized. Introgression has been shown to be involved in adaptive evolution a number of animal species (Hedrick 2013), including humans (Racimo et al. 2015), and in plants (Suarez-Gonzalez et al. 2018). The discovery of the introgression of part of a Y chromosome in hexaploid *M. annua* adds a novel and unusual example to these cases, and shows that interspecific gene flow can play an important role in evolutionary transitions between sexual systems.

In Chapter 3, I wanted to further explore the role of recombination in the evolution of plant sex chromosomes. Recombination is thought to play an outstanding role in the evolution of sex chromosomes (Charlesworth 2017). On animal sex chromosomes recombination suppression often evolves around the sex determining region of sex chromosomes, which can have a number of consequences on the molecular scale, like the accumulation of repetitive elements and the loss of functional gene copies, up to the evolution of microscopically observable differences between sex chromosomes on the chromosomal scale (Bachtrog 2013). Regions of suppressed recombination have also been observed on plant sex chromosomes. However,

in plants recombination suppression may also be involved in the initial evolution of separate sexes, as proposed by theoretical models (Charlesworth and Charlesworth 1978a; Charlesworth and Charlesworth 1978b), and it is not clear to what degree recombination suppression can evolve on plant sex chromosomes after that. In addition, it has been shown that in many cases sex chromosomes evolve in regions of reduced recombination (Charlesworth 2019), which can make it difficult to determine if regions of low recombination around the sex determiner are actually the result of sex chromosome evolution.

In Chapter 2, I showed that several annual and prennial lineages of *Mercurialis* share a common XY system of sex determination (Fig. S2). This makes this genus very useful to study the evolution of recombination suppression on plant sex chromosomes, because it allowed me to identify regions of repressed recombination that must have evolved after the evolution of separate sexes. In Chapter 3, I wanted to explore what role variation of recombination rates between sexes and on different parts of the chromosome play for the evolution of sex chromosomes and separate sexes and I wanted to test if recombination suppression can evolve on plant sex chromosomes after the evolution of separate sexes. To this end I developed a set of new genomic resources to build new genome assemblies and linkage maps for diploid *M. annua*. I combined information from these datasets to improve our understanding of the structure and origin of the sex chromosomes of diploid *M. annua*. Based on the patterns of repeat abundance and recombination rates on the sex chromosomes and autosomes I showed that the Y chromosome of *M. annua* likely evolved in a pericentromeric region, where limited recombination and increased repeat density predated the evolution of sex chromosomes (Fig. 3.1, Fig. S6). This pattern has been found in several other plant sex chromosomes (Zhang et al. 2008; Pilkington et al. 2019; Kersten et al. 2014), which implies that there may be selective reasons why sex chromosomes should evolve in these particular regions (Charlesworth 2019).

I also reevaluated a conclusion from a previous publication that claimed that a region of limited recombination on the *M. annua* Y chromosome is evidence for the presence of a second region of suppressed recombination with more limited divergence between sex chromosomes, a so-called evolutionary stratum (Veltsos et al. 2019). I did not find any evidence for such a stratum in diploid *M. annua*, which challenges this claim and stresses the fact that the lack of recombination in a genetic cross is not equivalent to complete recombination suppression (Kirkpatrick et al. 2020). This observation is in line with the idea that even very rare recombination events are sufficient to stop degeneration commonly associated with recombination suppression.

I used additional datasets to test if recombination suppression did evolve in any of the other *Mercurialis* lineages. I found molecular signatures commonly associated with the evolution of recombination suppression

on a major part of the sex chromosomes of *M. huetii* and *M. canariensis* (Fig. 3.4). Based on the phylogenies of the sex-linked marker from Chapter 2, I could also show that recombination suppression evolved independently in both of these lineages (Fig. 2.5). These results challenge the idea that evolution of recombination suppression does not evolve in the same way as on animal sex chromosomes (Müller et al. 2020) and they suggest that the region of suppressed recombination on plant sex chromosomes can extend after the evolution of separate sexes.

The results from Chapter 2 and Chapter 3 show that sex chromosomes played a dynamic role in the evolution of sexual systems in the genus *Mercurialis*. They were involved in transitions towards androdioecy in hexaploid *M. annua* through introgression. They also evolved recombination suppression over wide parts of the sex chromosome, while reduced recombination may have played a major role in the initial evolution of sex chromosomes. However, not all changes in sexual systems are related to the evolution of sex chromosomes. While sex chromosomes can also be involved in reversions towards combined sexes (Crossman and Charlesworth 2014), it is also likely that loci located on other genomic regions can cause the expression of the opposite sex in either males or females in dioecious plants. Such leakiness has been observed in many dioecious plant species (Ehlers and Bataillon 2007; Cossard and Pannell 2019), however almost nothing is known about the underlying genetic architecture.

In Chapter 4 I used plants from an experimental evolution setup in diploid *M. annua*, where removal of males caused a strong increase in leaky sex expression on females after a limited number of generations (Cossard, Gerchen, et al. 2021). I wanted to test what kind of genetic architecture could be involved in such an increase of leakiness, which has been strong enough to constitute a transition from separate towards combined sexes. Specifically, I wanted to test if these effects were caused by many small effect loci or by few loci with large effects and if these loci effect the expression of one sexual function or if they regulate both sex functions at the same time.

I identified QTLs for expression of both sexual functions based on F_2 crosses between control and selection lines and identified major effect QTLs in both crosses (Fig. 4.4, Fig. 4.5). QTLs that increased the expression of male flowers and decreased the expression of female flowers, were colocalized and had recessive or codominant effects (Tab. 4.1, Tab. 4.2). In each cross the major effect loci were located on a different linkage group. In addition I identified minor effect loci in both crosses, which explained a lower amount of the variance in male flower production observed in the cross (Tab. 4.1, Tab. 4.2). These results imply that leakiness in *M. annua* has a strong heritable genetic component, which is based on few large effect loci. The fact that different major effect loci were found in each cross implies that multiple of these loci segregate in natural metapopulations of diploid *M. annua*. The presence of loci with such

a strong effect on leakiness poses the question how they are maintained in natural metapopulations of diploid *M. annua*. Theoretical studies predict that selection for reproductive assurance could select for leakiness during recurrant periods of low population densities (Crossman and Charlesworth 2014). Such a pattern would be concordant with both the strong metapopulation dynamics in *M. annua* and its history of colonization of central Europe after the last glacial maximum (Obbard et al. 2006; Cossard and Pannell 2021; Cossard, Gerchen, et al. 2021). The fact that large effect loci with antagonistic effects on male and female sex expression are found on the same linkage groups implies that they could cause a tradeoff between both sex functions. Such tradeoffs are predicted by sex-allocation theory (Charnov et al. 1976; West 2009; Campbell 2000) and Chapter 4 of this thesis is one of the first studies that shows what their genetic architecture might look like.

I showed that transitions between separate sexes in *Mercurialis* are largely based on Mendelian loci with large effects on sex expression. Maleness is determined through dominant male determiners located on the Y chromosome. Inrogression of these Y chromosomes also caused transitions to androdioecy in hexaploid *M. annua*. In experimental populations, transitions from femaleness towards combined sexes are determined by different autosomal loci, which may be recessive or codominant, but have a strong effect on the expression of both sexes. This information adds a genetic component to our understanding of variation in sexual systems in *Mercurialis*, which has been largely based on phenotypic data from natural and experimental populations. Such a combination of phenotypic and genetic information can be a powerful tool to understand the evolution of transitions between sexual systems in other plant species as well. With a further decrease in the costs of genomic analyses, more such datasets can be expected in the future. This will allow comparative analyses between species, which can address more general questions, such as whether the same genes and gene networks become involved in transitions in sexual systems in different lineages.

My results touch on questions in evolutionary biology that go beyond transitions of plant sexual systems. One important question concerns whether adaptive changes are caused by novel mutations or by standing genetic variation (Barrett and Schluter 2008). In the case of transitions towards combined sexes in *Mercurialis*, my results suggest that transitions from separate to combined sexes can be based on standing genetic variation maintained in natural metapopulations of *M. annua*. In addition, introgression and hybridization are being increasingly recognized as sources of adaptive genetic variation (Suarez-Gonzalez et al. 2018). My example of Y chromosome introgression is an example of how geneflow between species can be the cause of substantial evolutionary change.

5.0.1 Perspectives for further research

My results pointed to several questions in sex chromosome evolution which could be adressed in the genus *Mercurialis*, but also in other species. I found evidence for a common XY sex chromosome system in annual and perennial *Mercurialis* lineages. However, I was not able to amplify the male-specific PCR marker in recently discovered tetraploid *M. annua* males. Due to limited sampling it is unclear if these males do not share the same XY system of sex determination, or if the male-specific PCR marker did not amplify due to mutations in the prime region, or because it is located on a region of the Y chromosome that is missing in these samples. Further sampling and genomic data analyses will help to test if maleness has the same genetic basis as in other lineages of *Mercurialis*, or if it evolved independently, which would be of high interest. Also my dataset is limited to the three perennial lineages that are the closests relatives to annual *Mercurialis* species (Krähenbühl et al. 2002). Genomic analysis of other more distantly related dioecious perennial lineages of *Mercurialis* could test if the XY system is ancestral in the whole clade.

I showed that sex chromosomes are able to introgress between lineages under certain circumstances. This observation has been contrary to previous conceptions of sex chromosomes contributing disproportionally to reproductive isolation. I proposed that either the small size of the non-recombing region, similarities in ploidy, or positive selection through increased outcrossing success of males are potential factors that could have contributed to this pattern. The effect of each of these factors can be further explored by artificial introgression of sex chromosomes between lineages. Since there is variation in ploidy, sexual system and in the extent of the non-recombining region on Y chromosomes, it may be possible to further narrow down which of these factors is most relevant for allowing or promoting introgression of sex chromosomes using a full factorial experimental design.

In Chapter 3 I showed that the sex chromosomes of *M. annua* likely evolved in a region of reduced recombination. It has been proposed that this reduction of recombination may have facilitated the evolution of separate sexes (Charlesworth 2019), based on models for the evolution of separate sexes in plants (Charlesworth and Charlesworth 1978a; Charlesworth and Charlesworth 1978b). These models propose that recombination suppression will link loci with effects on male and female sex expression, which seems plausible for several plant species, whose sex chromosomes have been found to be located in the pericentromeric regions as well (Wang et al. 2012; Pilkington et al. 2019). In order to test this hypothesis in *Mercurialis*, it would be necesseary to identify the genes responsible for sex determination. My new estimate of the size of the sex-linked region on the Y chromosome in diploid *Mercurialis annua* is approximately 10 Mb, which means that it could still contain dozens or hundreds of genes and it will be necessary to narrow down this region further. One approach that could be used to do this would be deletion

mapping of Y chromosomes (Kazama et al. 2016), however, this can be time and labor intensive. A promising alternative could involve further analysis of the sex-linked content of hexaploid *M. annua*, or the perennial lineages, since my results suggest that the size the non-recombining region may be smaller than in diploid *M. annua*. Comparing male-specific genes that are shared between Y chromosomes of related lineages has been shown to be an efficient way to narrow down candidate genes for sex determination (Torres et al. 2018). In addition, it will be of high interest to generate further genomic data for *M. huetii* and *M. canariensis*, since I showed that there have been independent extensions of the sex-linked region (Fig. 3.4). Based on this data, patterns of molecular evolution can be compared between sex chromosomal regions that are shared between all *Mercurialis* lineages and those that evolved recombination suppression in specific lineages. It will be of particular interest to test if regions that evolved recombination suppression at a later stage will follow the same evolutionary trajectories as non-recombining regions on animal sex chromosomes. Main questions will be to what degree the region suppressed recombination *Mercurialis* accumulated repetitive elements, if there has been genetic degeneration and if chromosomal translocations or inversions were involved in the evolution of recombination suppression.

In Chapter 4 I analyzed two crosses, each representing one experimental line, and identified large effect loci with a strong effect on leaky sex expression in females. I plan to extend these analyses with samples from further crosses between the two remaining experimental lines from the experimental evolution setup. This can show if further loci with different effect sizes were involved in changes in sexual phenotypes. In addition, it would be interesting to generate further genomic data, which, in combination with improved genome assemblies, could help to narrow down the number of potential candidate genes responsible for changes in sexual phenotypes. While it seems unlikely that it will be possible to identify single candidate genes based on the current crosses, further crosses could be generated for fine scale mapping and characterization of variants. If these variants are better characterized it would also be possible to study their distribution in natural populations of *Mercurialis*, which could give hints to their maintenance. Additional genomic analyses of these regions in tetraploid and hexaploid *M. annua* can show if similar variants could have been involved in their transition to monoecy. Similar leakiness has been observed in a large number of plant species (Ehlers and Bataillon 2007; Cossard and Pannell 2019), and it would be interesting to explore the underlying genetics in these species as well.

References

Bachtrog, Doris (2013). "Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration". In: *Nature Reviews Genetics* 14.2, pp. 113–124.

Barrett, Rowan DH and Dolph Schluter (2008). "Adaptation from standing genetic variation". In: *Trends in ecology & evolution* 23.1, pp. 38–44.

Campbell, Diane R (2000). "Experimental tests of sex-allocation theory in plants". In: *Trends in Ecology & Evolution* 15.6, pp. 227–232.

Charlesworth, Brian and Deborah Charlesworth (1978a). "A model for the evolution of dioecy and gynodioecy". In: *The American Naturalist* 112.988, pp. 975–997.

Charlesworth, D (2002). "Plant sex determination and sex chromosomes". In: *Heredity* 88.2, pp. 94–101.

Charlesworth, D and B Charlesworth (1978b). "Population genetics of partial male-sterility and the evolution of monoecy and dioecy". In: *Heredity* 41.2, pp. 137–153.

Charlesworth, Deborah (2017). "Evolution of recombination rates between sex chromosomes". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1736, p. 20160456.

Charlesworth, Deborah (2019). "Young sex chromosomes in plants and animals". In: *New Phytologist* 224.3, pp. 1095–1107.

Charnov, Eric L, James J Bull, and J Maynard Smith (1976). "Why be an hermaphrodite?" In: *Nature* 263.5573, pp. 125–126.

Comai, Luca (2005). "The advantages and disadvantages of being polyploid". In: *Nature reviews genetics* 6.11, pp. 836–846.

Cossard, Guillaume G, Jörn F Gerchen, Xinji Li, Yves Cuenot, and John R Pannell (2021). In: *Current Biology* 31.6, pp. 1277–1283.

Cossard, Guillaume G and John R Pannell (2019). "A functional decomposition of sex inconstancy in the dioecious, colonizing plant Mercurialis annua". In: *American journal of botany* 106.5, pp. 722–732.

Cossard, Guillaume G and John R Pannell (2021). "Enhanced leaky sex expression in response to pollen limitation in the dioecious plant Mercurialis annua". In: *Journal of Evolutionary Biology* 34.2, pp. 416–422.

Coyne, Jerry A (2018). "'Two rules of speciation' revisited". In: *Molecular ecology* 27.19, pp. 3749–3752.

Crossman, Allan and Deborah Charlesworth (2014). "Breakdown of dioecy: models where males acquire cosexual functions". In: *Evolution* 68.2, pp. 426–440.

Ehlers, Bodil K and Thomas Bataillon (2007). "'Inconstant males' and the maintenance of labile sex expression in subdioecious plants". In: *New Phytologist* 174.1, pp. 194–211.

Eppley, Sarah M and John R Pannell (2007). "Density-dependent self-fertilization and male versus hermaphrodite siring success in an androdioecious plant". In: *Evolution* 61.10, pp. 2349–2359.

Haldane, John BS (1922). "Sex ratio and unisexual sterility in hybrid animals". In: *Journal of genetics* 12.2, pp. 101–109.

Hedrick, Philip W (2013). "Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation". In: *Molecular ecology* 22.18, pp. 4606–4618.

Käfer, Jos, Gabriel AB Marais, and John R Pannell (2017). "On the rarity of dioecy in flowering plants". In: *Molecular Ecology* 26.5, pp. 1225–1241.

Kazama, Yusuke, Kotaro Ishii, Wataru Aonuma, Tokihiro Ikeda, Hiroki Kawamoto, Ayako Koizumi, Dmitry A Filatov, Margarita Chibalina, Roberta Bergero, Deborah Charlesworth, et al. (2016). "A new physical mapping approach refines the sex-determining gene positions on the Silene latifolia Y-chromosome". In: *Scientific reports* 6.1, pp. 1–10.

Kersten, B, B Pakull, K Groppe, J Lueneburg, and M Fladung (2014). "The sex-linked region in Populus tremuloides corresponds to a pericentromeric region of about two million base pairs on P. trichocarpa chromosome 19". In: *Plant Biology* 16.2, pp. 411–418.

Kirkpatrick, Mark, Groves Dixon, Jason M Sardell, Manfred Schartl, and CL Peichel (2020). "Evolution of the canonical sex chromosomes of the guppy and its relatives". In: *bioRxiv*.

Krähenbühl, M., Y.-M. Yuan, and Philippe Küpfer (2002). "Chromosome and breeding system evolution of the genus Mercurialis (Euphorbiaceae): implications of ITS molecular phylogeny". In: *Plant Systematics and Evolution* 234.1-4, pp. 155–169.

Lima, Thiago G (2014). "Higher levels of sex chromosome heteromorphism are associated with markedly stronger reproductive isolation". In: *Nature Communications* 5.1, pp. 1–8.

Ming, Ray, Abdelhafid Bendahmane, and Susanne S Renner (2011). "Sex chromosomes in land plants". In: *Annual review of plant biology* 62, pp. 485–514.

Müller, Niels A, Birgit Kersten, Ana P Leite Montalvão, Niklas Mähler, Carolina Bernhardsson, Katharina Bräutigam, Zulema Carracedo Lorenzo, Hans Hoenicka, Vikash Kumar, Malte Mader, et al. (2020). "A single gene underlies the dynamic evolution of poplar sex determination". In: *Nature Plants*, pp. 1–8.

Obbard, Darren J, Stephen A Harris, Richard JA Buggs, and John R Pannell (2006). "Hybridization, polyploidy, and the evolution of sexual systems in Mercurialis (Euphorbiaceae)". In: *Evolution* 60.9, pp. 1801–1815.

Pilkington, SM, J Tahir, E Hilario, SE Gardiner, D Chagné, A Catanach, J McCallum, L Jesson, LG Fraser, MA McNeilage, et al. (2019). "Genetic and cytological analyses reveal the recombination landscape of a partially differentiated plant sex chromosome in kiwifruit". In: *BMC plant biology* 19.1, p. 172.

Racimo, Fernando, Sriram Sankararaman, Rasmus Nielsen, and Emilia Huerta-Sànchez (2015). "Evidence for archaic adaptive introgression in humans". In: *Nature Reviews Genetics* 16.6, pp. 359–371.

Renner, Susanne S (2014). "The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database". In: *American Journal of botany* 101.10, pp. 1588–1596.

Sauquet, Hervé, Maria von Balthazar, Susana Magallon, James A Doyle, Peter K Endress, Emily J Bailes, Erica Barroso de Morais, Kester Bull-Hereñu, Laetitia Carrive, Marion Chartier, et al. (2017). "The ancestral flower of angiosperms and its early diversification". In: *Nature communications* 8.1, pp. 1–10.

Suarez-Gonzalez, Adriana, Christian Lexer, and Quentin CB Cronk (2018). "Adaptive introgression: a plant perspective". In: *Biology letters* 14.3, p. 20170688.

The Tree of Sex Consortium (2014). "Tree of Sex: A database of sexual systems". In: *Scientific Data* 1.1, pp. 1–8.

Torres, Maria F, Lisa S Mathew, Ikhlak Ahmed, Iman K Al-Azwani, Robert Krueger, Diego Rivera-Nuñez, Yasmin A Mohamoud, Andrew G Clark, Karsten Suhre, and Joel A Malek (2018). "Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix". In: *Nature communications* 9.1, pp. 1–9.

Turelli, Michael and H Allen Orr (1995). "The dominance theory of Haldane's rule." In: *Genetics* 140.1, pp. 389–402.

Veltsos, Paris, Kate E Ridout, Melissa A Toups, Santiago C Gonzàlez-Martinez, Aline Muyle, Olivier Emery, Pasi Rastas, Vojtech Hudzieczek, Roman Hobza, Boris Vyskot, et al. (2019). "Early sex-chromosome evolution in the diploid dioecious plant Mercurialis annua". In: *Genetics* 212.3, pp. 815–835.

Vicoso, Beatriz (2019). "Molecular and evolutionary dynamics of animal sex-chromosome turnover". In: *Nature ecology & evolution*, pp. 1–10.

Wang, Jianping, Jong-Kuk Na, Qingyi Yu, Andrea R Gschwend, Jennifer Han, Fanchang Zeng, Rishi Aryal, Robert VanBuren, Jan E Murray, Wenli Zhang, et al. (2012). "Sequencing papaya X and Yh

chromosomes reveals molecular basis of incipient sex chromosome evolution". In: *Proceedings of the National Academy of Sciences* 109.34, pp. 13710–13715.

West, Stuart (2009). *Sex allocation*. Vol. 44. Princeton University Press.

Yakimowski, Sarah B and Spencer CH Barrett (2016). "The role of hybridization in the evolution of sexual system diversity in a clonal, aquatic plant". In: *Evolution* 70.6, pp. 1200–1211.

Zhang, Wenli, Xiue Wang, Qingyi Yu, Ray Ming, and Jiming Jiang (2008). "DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya". In: *Genome research* 18.12, pp. 1938–1943.

Zou, Cheng, Mélanie Massonnet, Andrea Minio, Sagar Patel, Victor Llaca, Avinash Karn, Fred Gouker, Lance Cadle-Davidson, Bruce Reisch, Anne Fennell, et al. (2021). "Multiple independent recombinations led to hermaphroditism in grapevine". In: *Proceedings of the National Academy of Sciences* 118.15.

Chapter 6

Supplementary Material

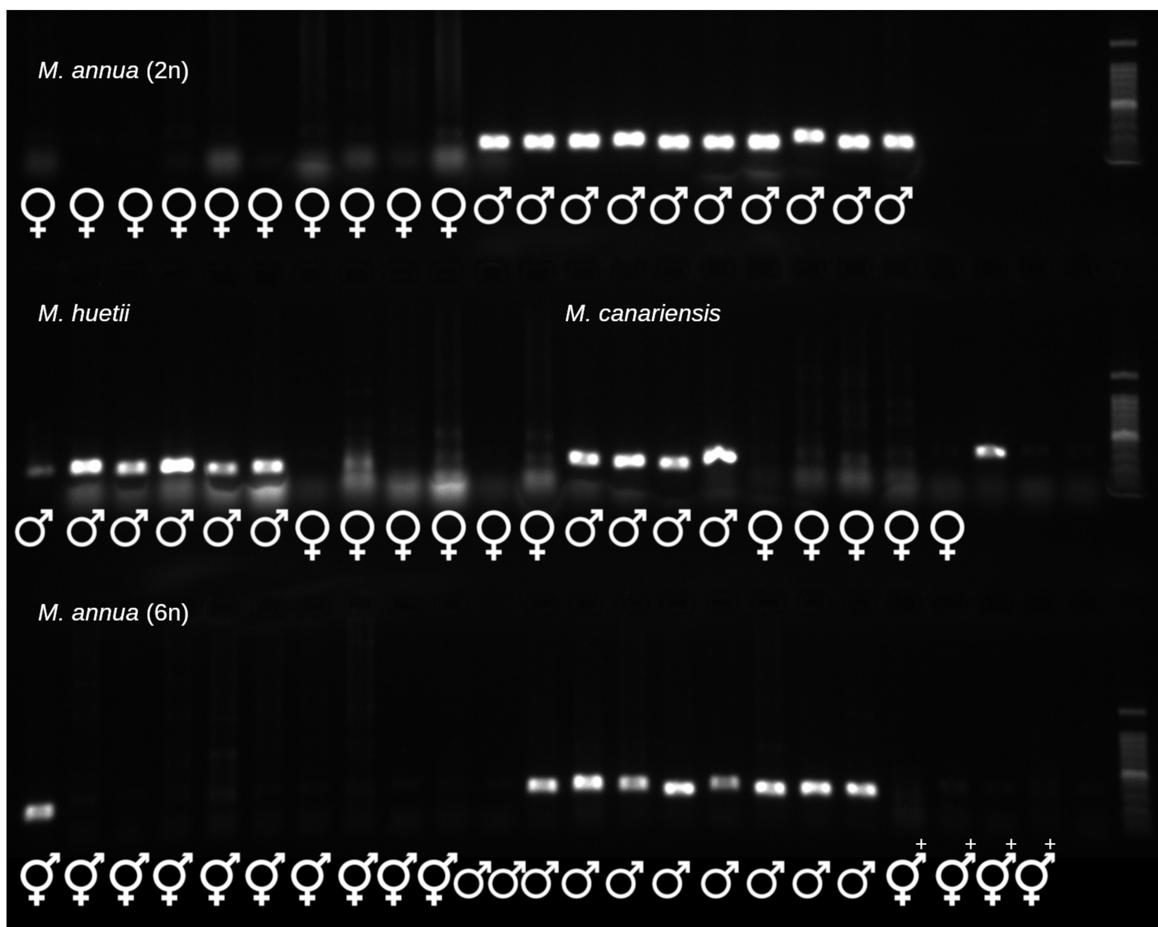


Figure S1: Agarose gel of a 200 bp sex-linked PCR marker in diploid *M. annua*, *M. huetii*, *M. canariensis* and hexaploid *M. annua*, including a hexaploid lineage with male-like inflorescences. In the rightmost lanes is a BenchTop 100 bp DNA ladder (Promega). Symbols under lanes indicate sample sex, hermaphrodite symbols with a small plus indicate monoecious plants with male-like inflorescences.

Table S1: Primers used for phylogenetic analyses.

Marker	Linkage Group	Position [cM]	Size [bp]	TA [C]	Sequence Forward Primer	Sequence Reverse Primer
sl_an	Y	-	780	60	CAAGCCCTCCTAGTGCTCAA	ATGATTTTCACAAGCATACCTCAA
sl_hue	Y	-	780	60	CAAGCCCTCCTAGTGCTCAA	AAGATTTCCACGAGCATACCTC
sl_can	Y	-	780	55	GGATGATGTCTCAAGCCCTTC	TTTTCACGAGCATACCTCAATG
sl_6n_per	Y	-	780	55	GATGCCTCAAGCCCTCCTA	AAGATTTCCACGAGCATACCTC
1_20	1	0	783	60	HGCTATGGTTGTCAATAAATTCA	TTGCACTTGAACCCACTCCC
1_23	1	7.48	2234	60	ATCCGTGGCCAYATTTTGCA	AATGGTGCCTCTTCCTCTGC
1_28	1	9.33	940	60	ACCTTGAGCTGAYCCTACTGCA	MATTCAGAACTTGCAGAAGC
1_30	1	18.59	848	60	YCAAGCTTGTTCTGTGCAAA	ACATTCCCTTTTGTYCTTAAGGC
1_27	1	24.3	1451	60	RGCTTTCTACAAGCAGCTCA	TGCTGTAGGAACTAGTGGCC
1_31	1	40.59	1031	60	TATGGTGGACACTTGGCAGC	RCCTCATAGCCCAATCCAA
1_32	1	42.74	717	60	TGAGGGRAATCCAAGAGCAC	AGAAGCTGGTCWGGAATGGACT
1_22	1	44.14	1442	60	AACTGCATGGGAGTGGGAAG	RCATATTGCACCAAGCCCAT
1_33	1	46.44	847	60	CCTCTCTARCCACCCACCGG	KCTTGTTCAGATCCCCATGTG
1_34	1	48.3	880	60	TCACTTCACTGCARAAATCAGCA	CTTCGCTGCAACATTGTGGT
1_25	1	53.85	1825	60	CCCCTCCAATTGCTGTCGAA	CTGGCCCTAGCAAGTCCTTC
1_38	1	56.76	924	60	TCAGGCACRTAAGGCATTAAGTGAACG	GGGCAAAGAACACAAGGYTGCTT
1_46	1	60.33	985	60	ATGCAGGTGGAAGCAGAGTG	CTGCGAGCCTCTTTTTGCAR
1_43	1	68.67	1266	60	GAGCAAGTCCTTGARACCCAG	CCCCATTTGTTCAGCAATYCCC
1_24	1	77.93	1839	60	RAGTCCAAATGAAATAGCCAGT	GCAAGTCTMTTGAGTGCTTGTGC
1_21	1	85.34	1052	60	RCTGGCTTCTACAATGGCAC	GAAGGGAGGTTCARGTTGGC
2_4	2	9.48	864	60	RTGGAAAGAAGTCAAAGCCCA	TCAATGCCAAAGCTTCATGCA
3_5	3	65.27	718	60	TGATGTTACTTTCTTGAGCTTGCT	SATGGCTACATGACAATCACT
3_6	3	93.05	727	60	AGGACKTGTGTTGGACAGTGT	AGCTTCTCTGTCAAGTCCACGT
4_4	4	37.48	700	60	AGATGGTTGATGTGGTTTGTGG	ATGAAATCRCTGGTTATCCCAACAT
5_5	5	61.56	1012	60	TGTTGCACACTCWGGAGCCATCA	AGTGCATGATTTGTCAAAGAAGGG
6_4	6	37.11	1005	60	ACCCTTRAGTCCTTTCCGGA	ATGGTGCTGCAYAAGTTTGG
7_5	7	57.94	875	60	TTCCTYGCTCTCACATCCCC	RGTGACACTAGTAATGCTTTTGA
8_7	8	16.95	876	60	AGAATGCTCCTYCTCCCAAGAA	RGGGATTCACTAGTCGATCG

Table S2: Samples used for phylogenetic analyses in Chapter 2. Ped indicates monoecious plants with male-like inflorescences.

Sample	Lineage	Sex	Origin	Ploidy
2nm1	M.annua (2n)	male	Spain	2x
2nf1	M.annua (2n)	female	Spain	2x
2nm2	M.annua (2n)	male	Turkey	2x
2nf2	M.annua (2n)	female	Turkey	2x
4nm	M. annua (4x)	male	Morocco	4x
4nf	M. annua (4x)	monoecious	Morocco	4x
6nm1	M. annua (6x)	male	Spain	6x
6nh1	M. annua (6x)	monoecious	Spain	6x
6nm2	M. annua (6x)	male	Spain	6x
6nh2	M. annua (6x)	monoecious	Spain	6x
6nped1	M. annua (6x) Ped	monoecious, ped	Spain	6x
6nped2	M. annua (6x) Ped	monoecious, ped	Spain	6x
Canm	M. canariensis	male	Teneriffe	4x
Canf	M. canariensis	female	Teneriffe	4x
Huem	M. huetii	male	Spain	2x
Huef	M. huetii	female	Spain	2x
Revm	M. reverchonii	male	Spain	4x
Revf	M. reverchonii	female	Spain	4x
Tomm	M. tomentosa	male	Spain	4x
Tomf	M. tomentosa	female	Spain	4x
Ellm	M. elliptica	male	Spain	6x
Ellf	M. elliptica	female	Spain	6x
Perm	M. perennis	male	Switzerland	6x
Perf	M. perennis	female	Switzerland	6x

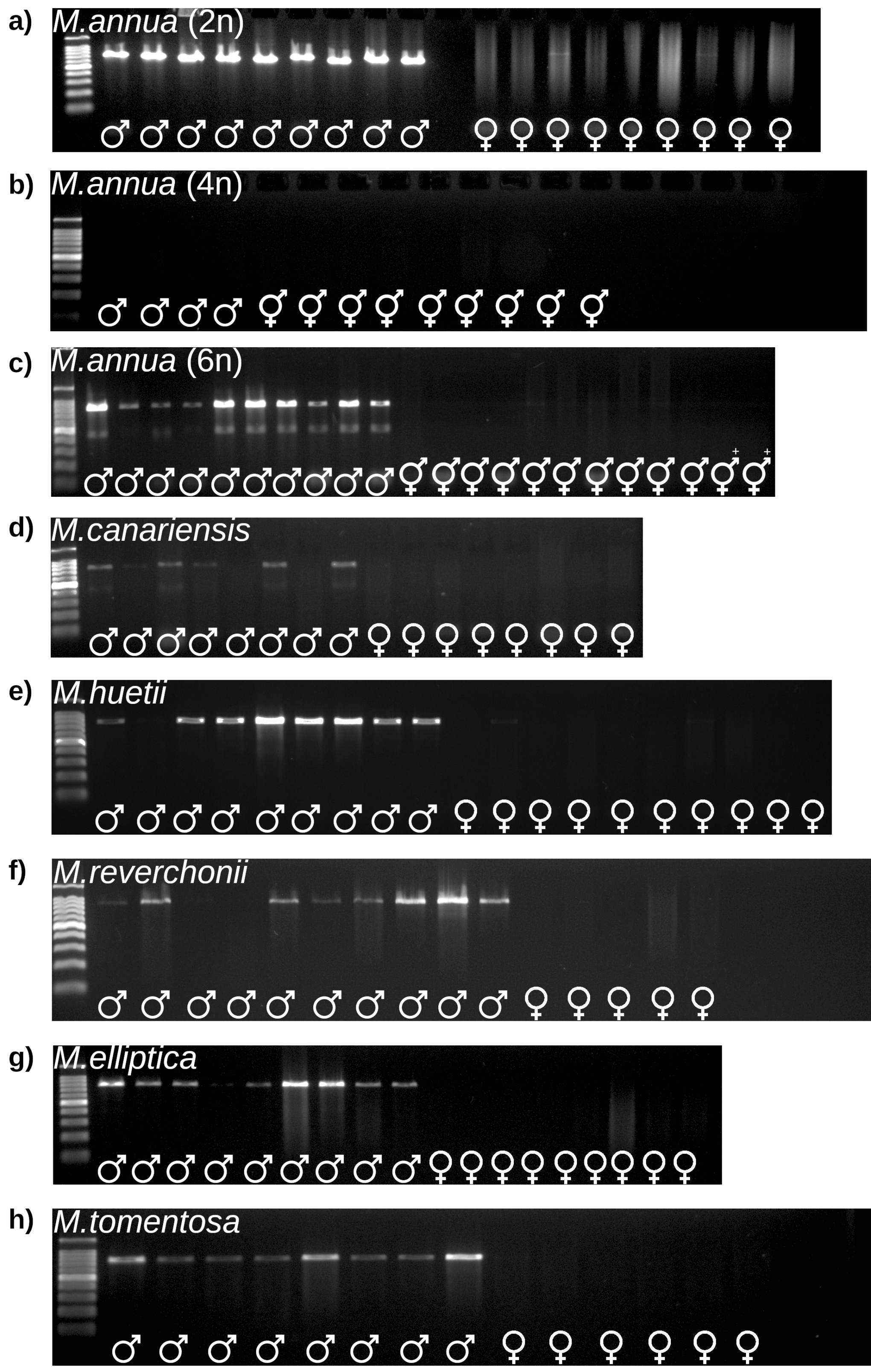


Figure S2: Agarose gel of a 700 bp sex-linked PCR marker in annual and perennial lineages of *Mercurialis*. Symbols under lanes indicate sample sex, hermaphrodite symbols with a small plus indicate monoecious plants with male-like inflorescences. In the leftmost lanes is a BenchTop 100 bp DNA ladder (Promega).

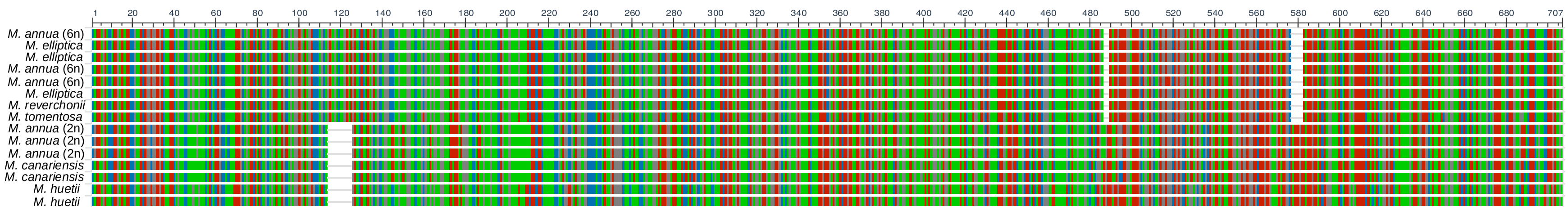
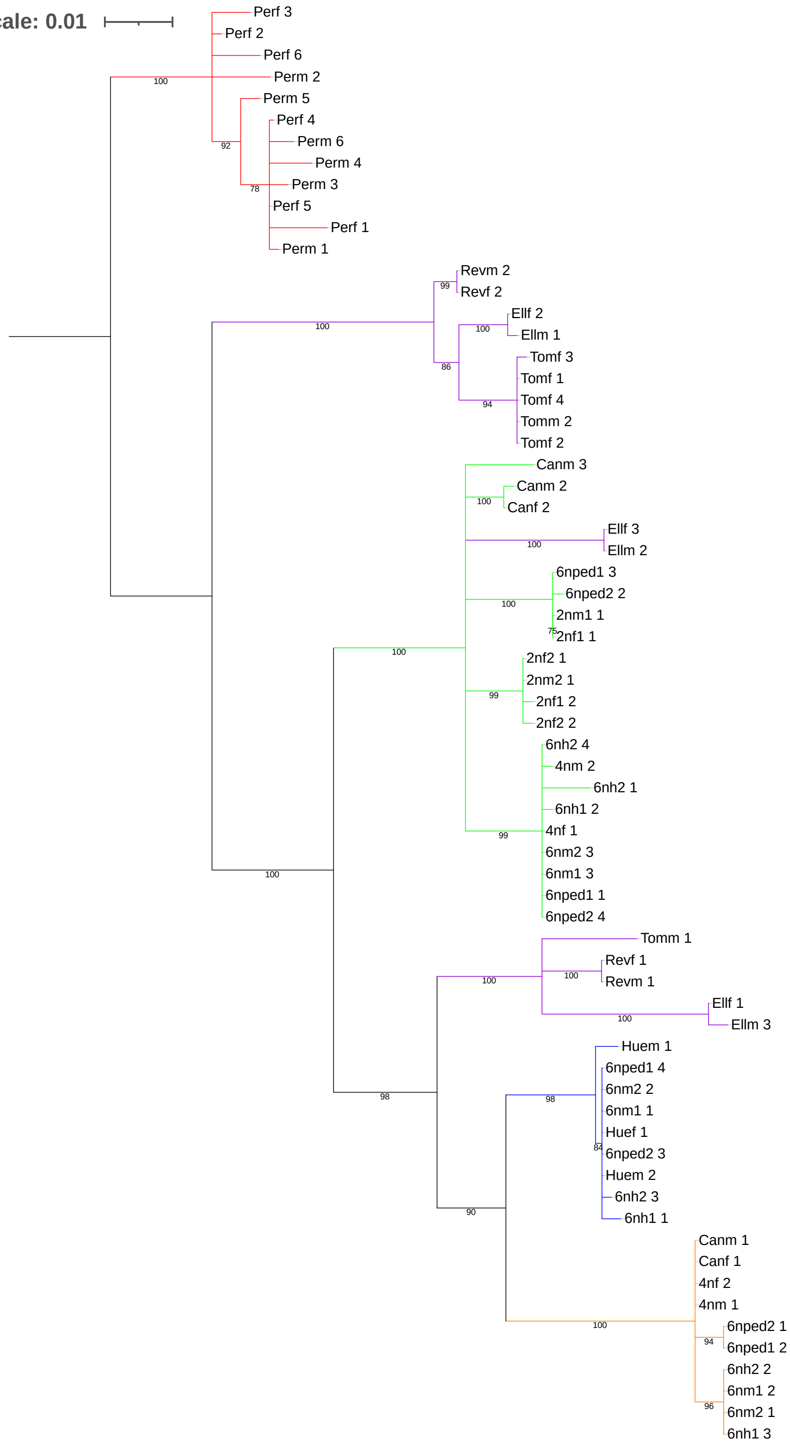
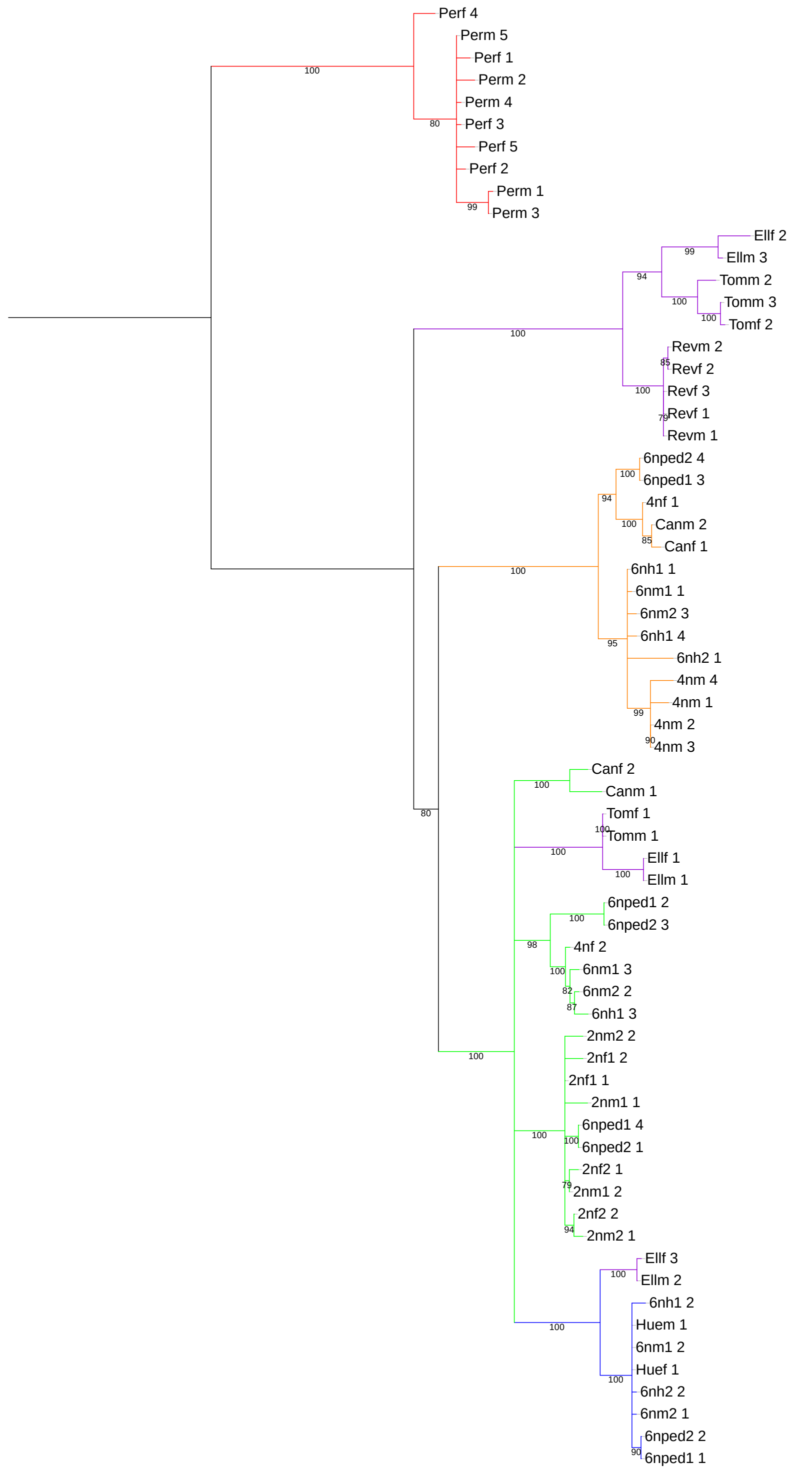


Figure S3: Alignmnent of sequenced sex-linked marker in annual and perennial *Mercurialis* lineages.


1_20
Tree scale: 0.01
Perf 3
Perf 2
Perf 6
Perm 2
Perm 5
Perf 4
Perm 6
Perm 4
Perm 3
Perf 5
Perf 1
Perm 1
Revm 2
Revf 2
Ellf 2
Ellm 1
Tomf 3
Tomf 1
Tomf 4
Tomm 2
Tomf 2
Canm 3
Canm 2
Canf 2
Ellf 3
Ellm 2
6nped1 3
6nped2 2
2nm1 1
2nf1 1
2nf2 1
2nm2 1
2nf1 2
2nf2 2
6nh2 4
4nm 2
6nh2 1
6nh1 2
4nf 1
6nm2 3
6nm1 3
6nped1 1
6nped2 4
Tomm 1
Revf 1
Revm 1
Ellf 1
Ellm 3
Huem 1
6nped1 4
6nm2 2
6nm1 1
Huef 1
6nped2 3
Huem 2
6nh2 3
6nh1 1
Canm 1
Canf 1
4nf 2
4nm 1
6nped2 1
6nped1 2
6nh2 2
6nm1 2
6nm2 1
6nh1 3

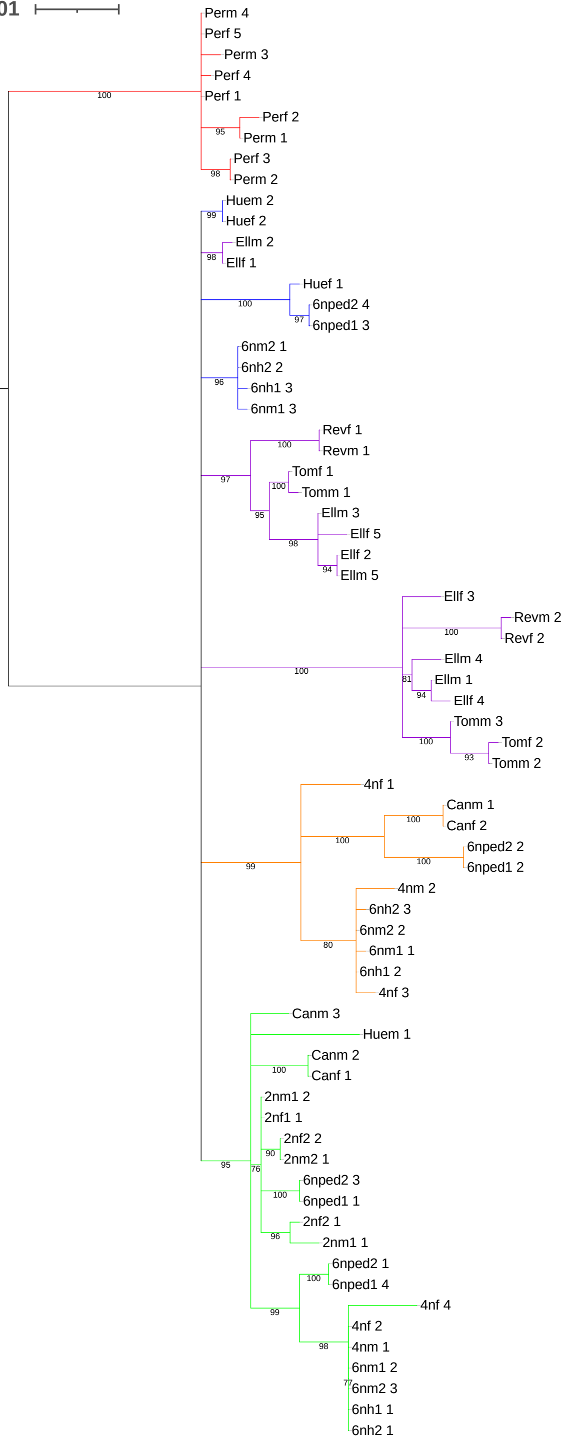
1_23

Tree scale: 0.01



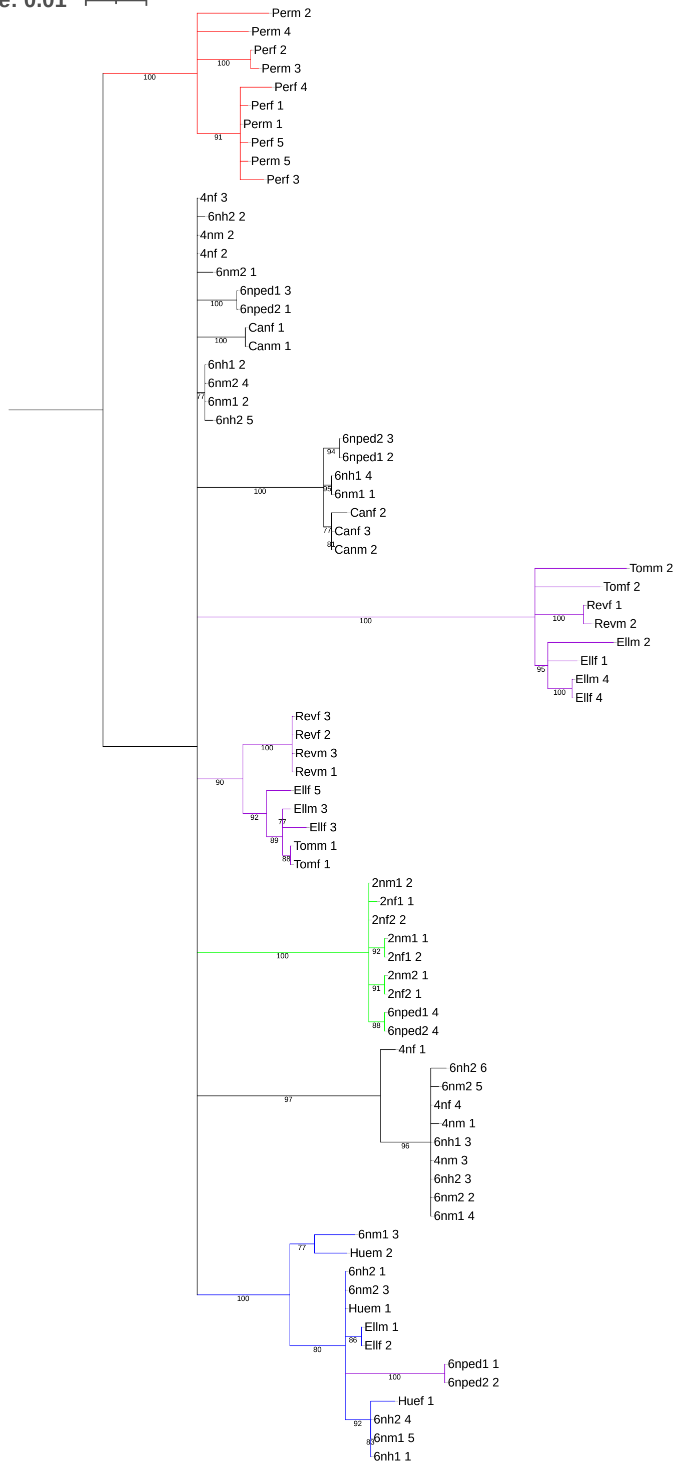
1_28

Tree scale: 0.01

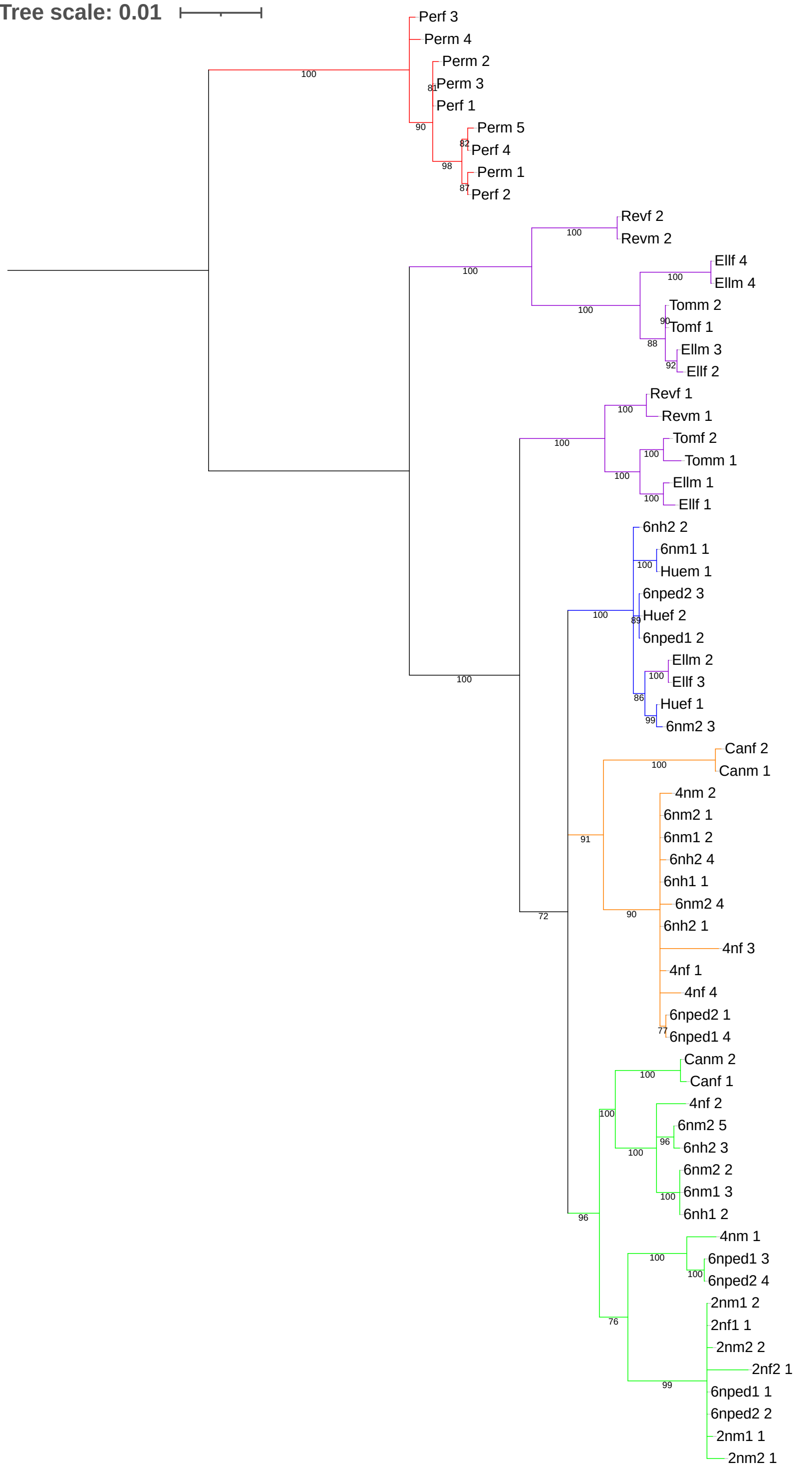


1_30

Tree scale: 0.01

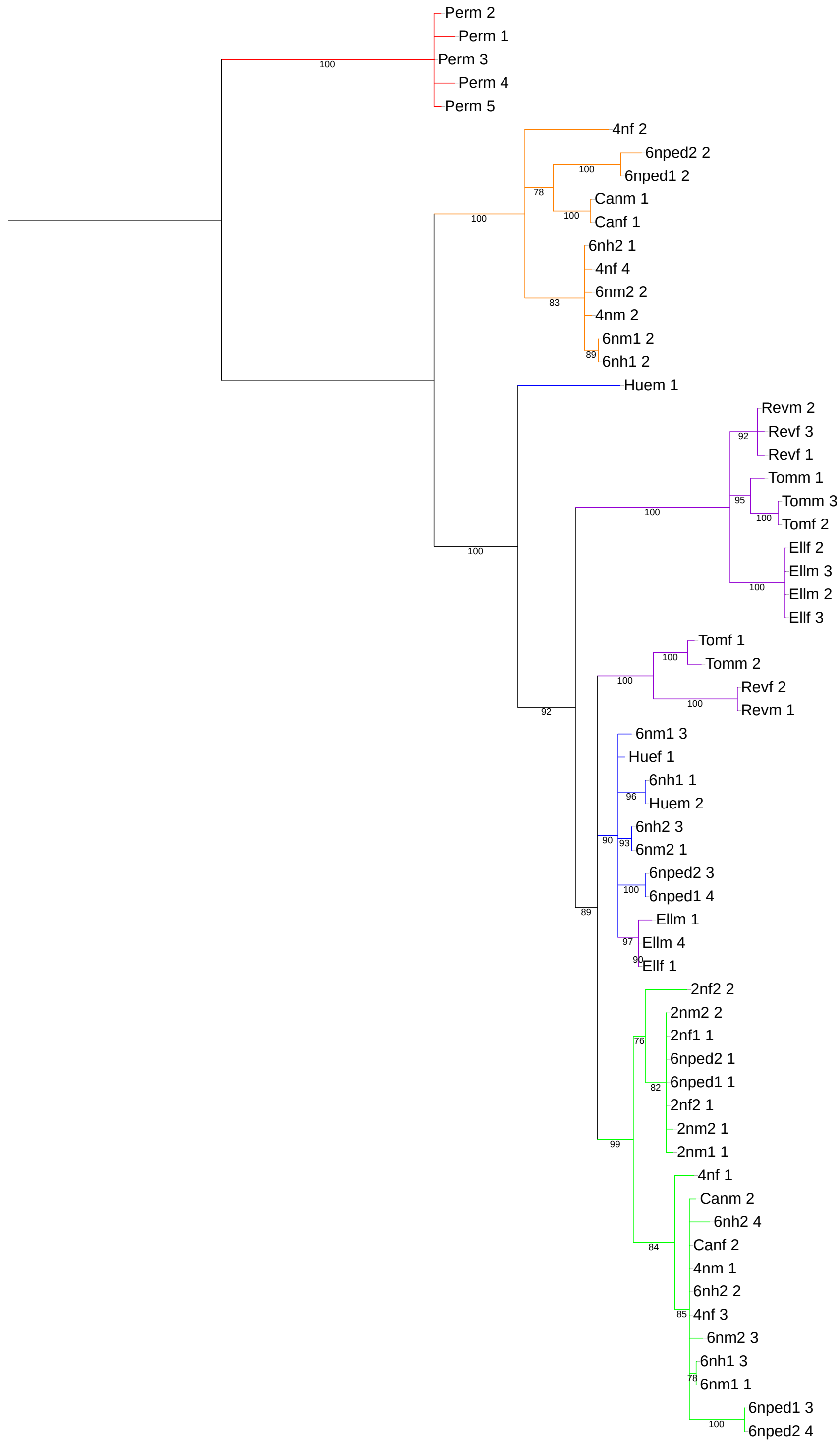


1_27



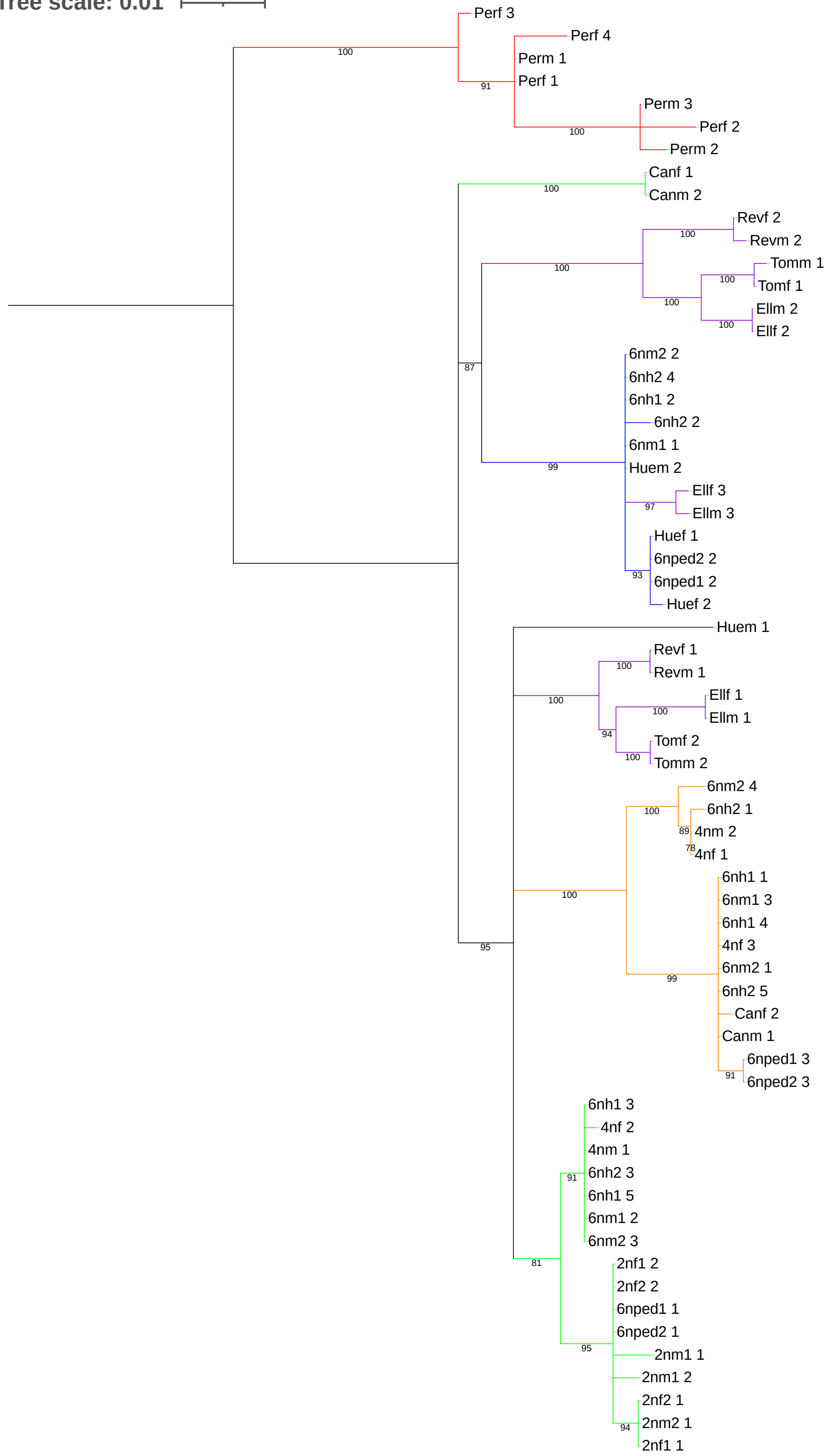
1_31

Tree scale: 0.01



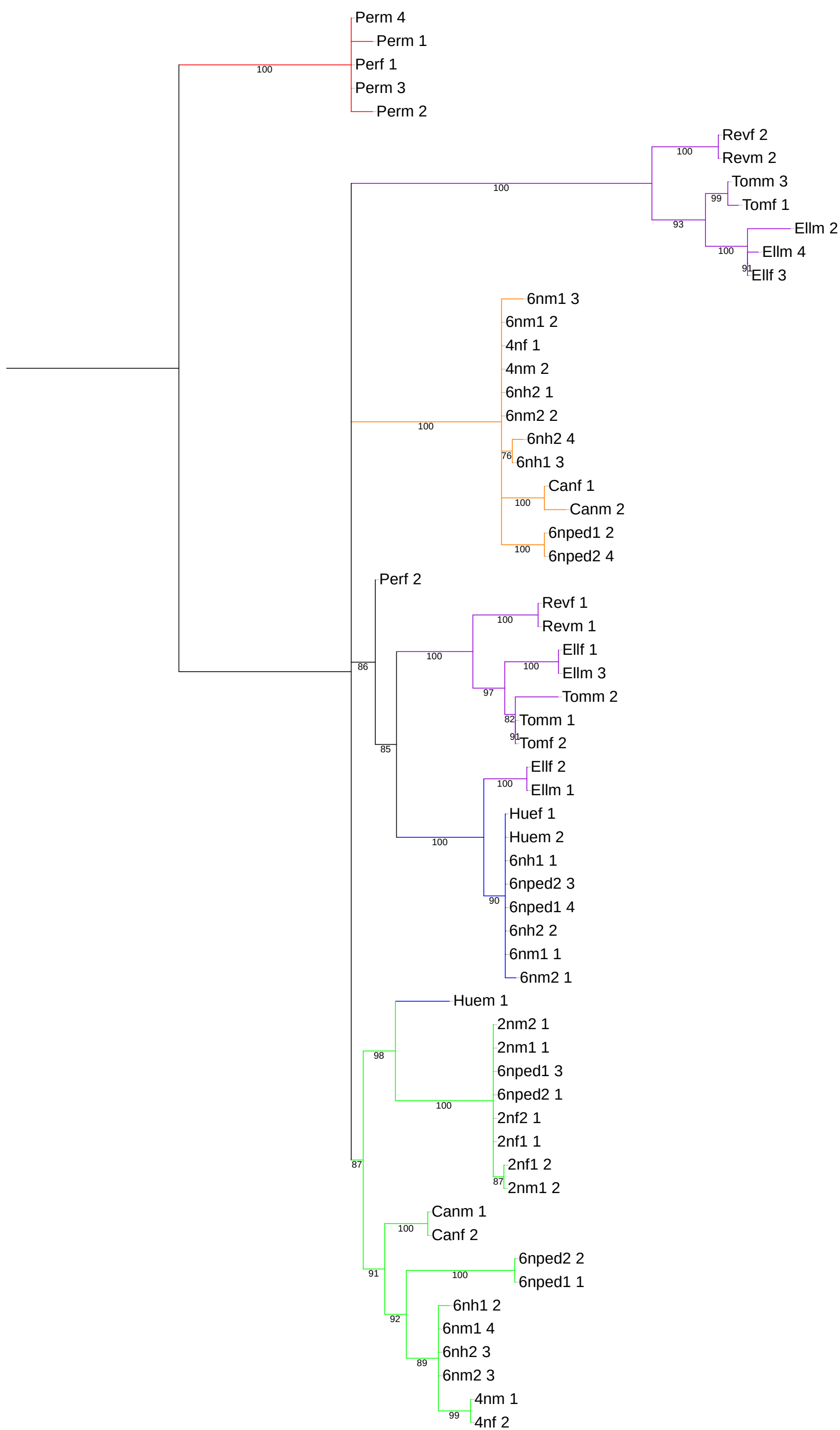
1_32

Tree scale: 0.01



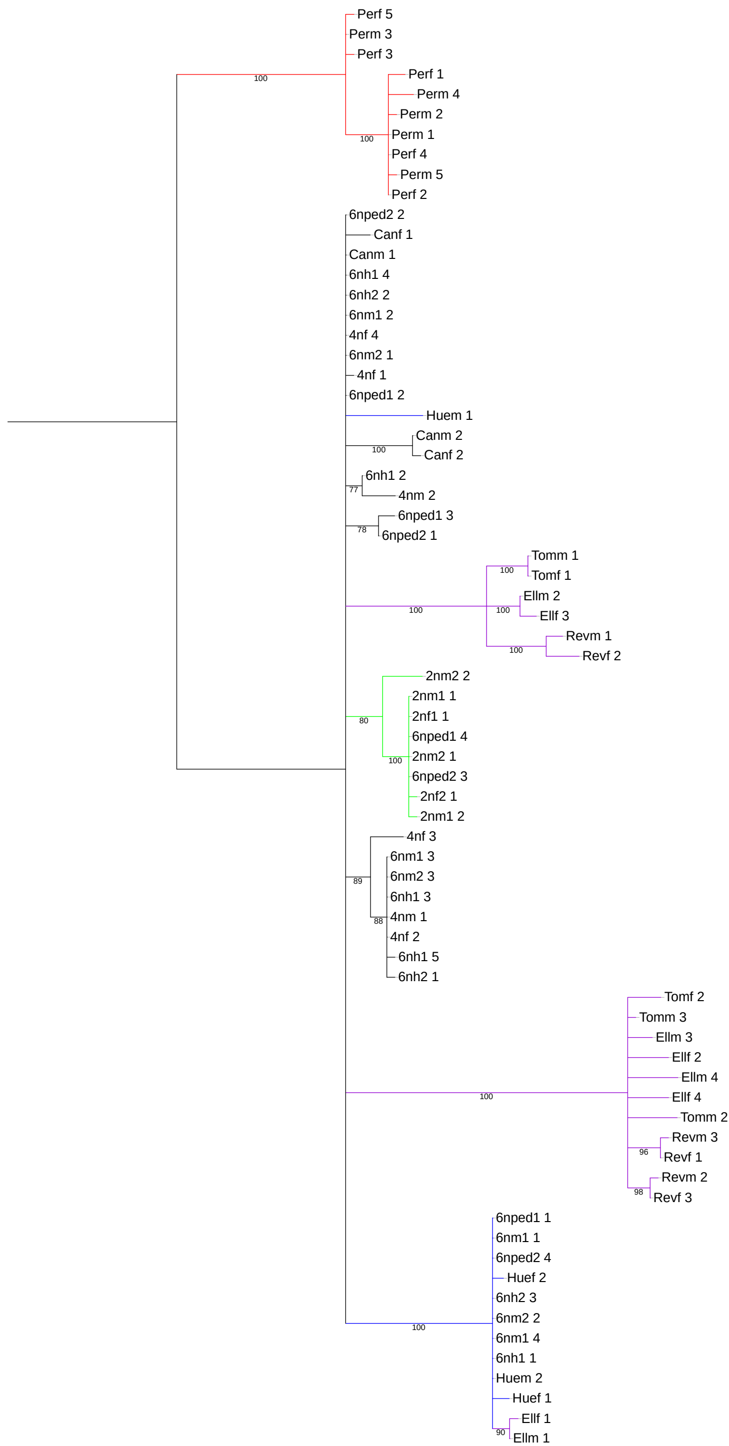
1_22

Tree scale: 0.01



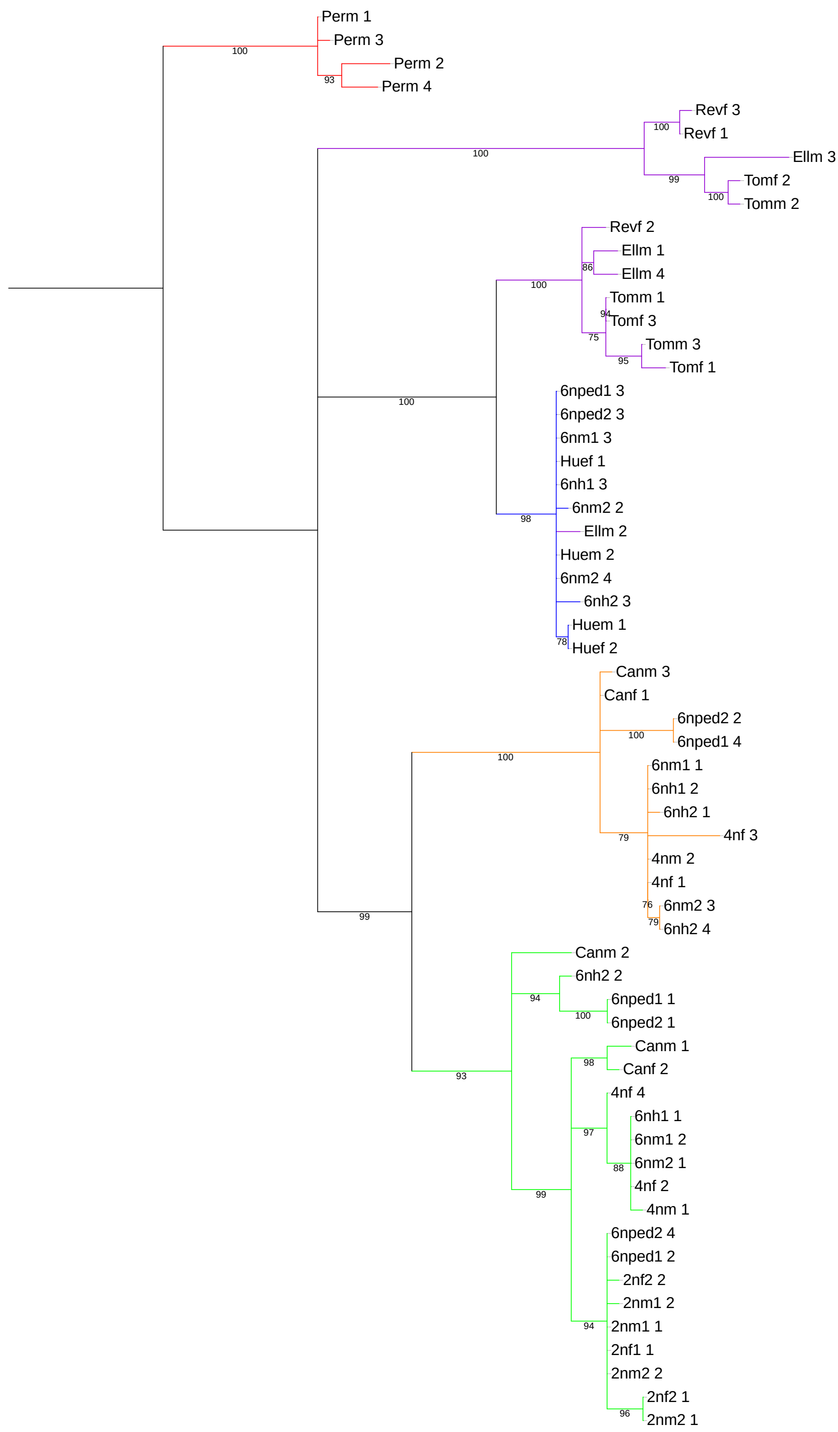
1_33

Tree scale: 0.01



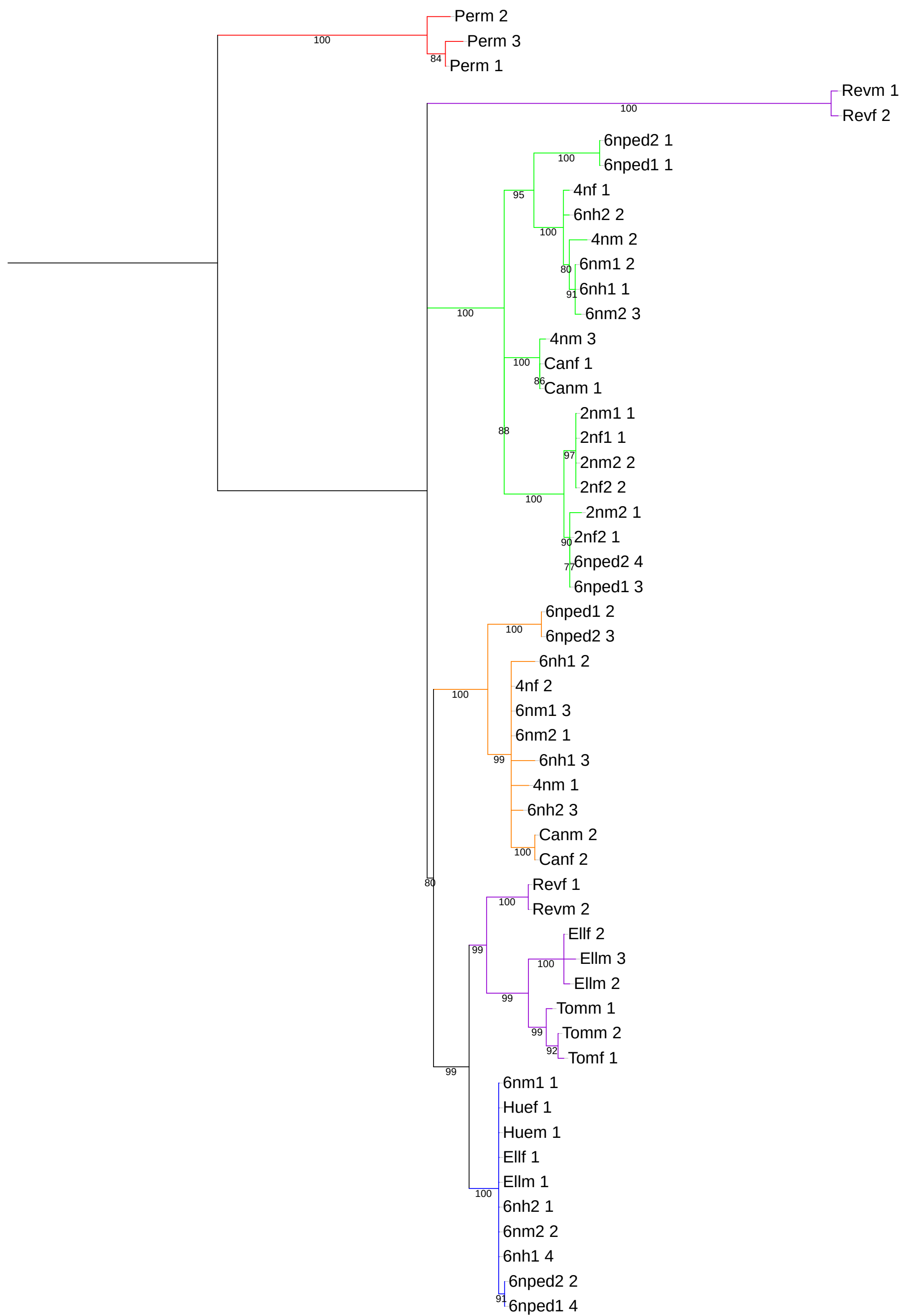
1_34

Tree scale: 0.01

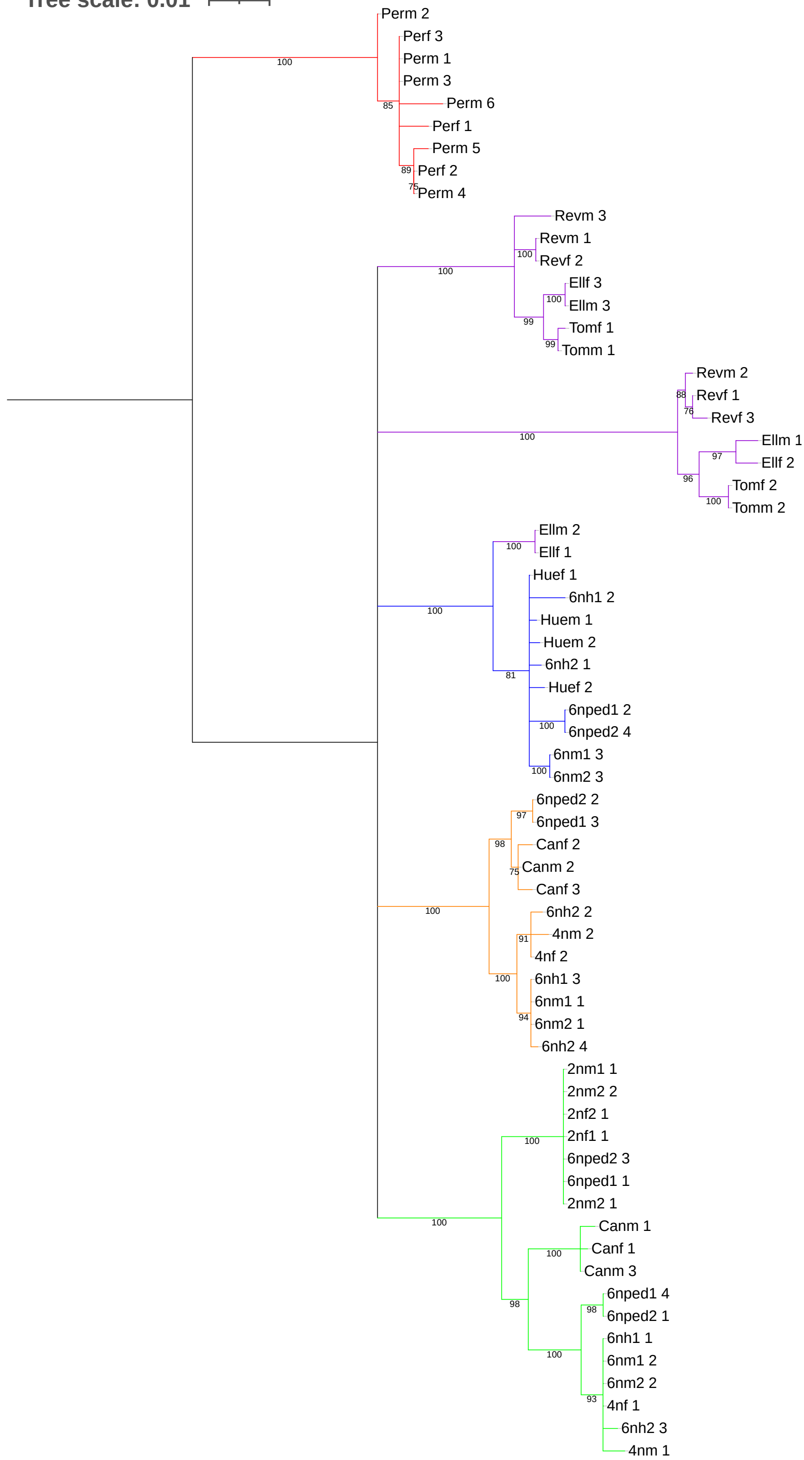


1_25

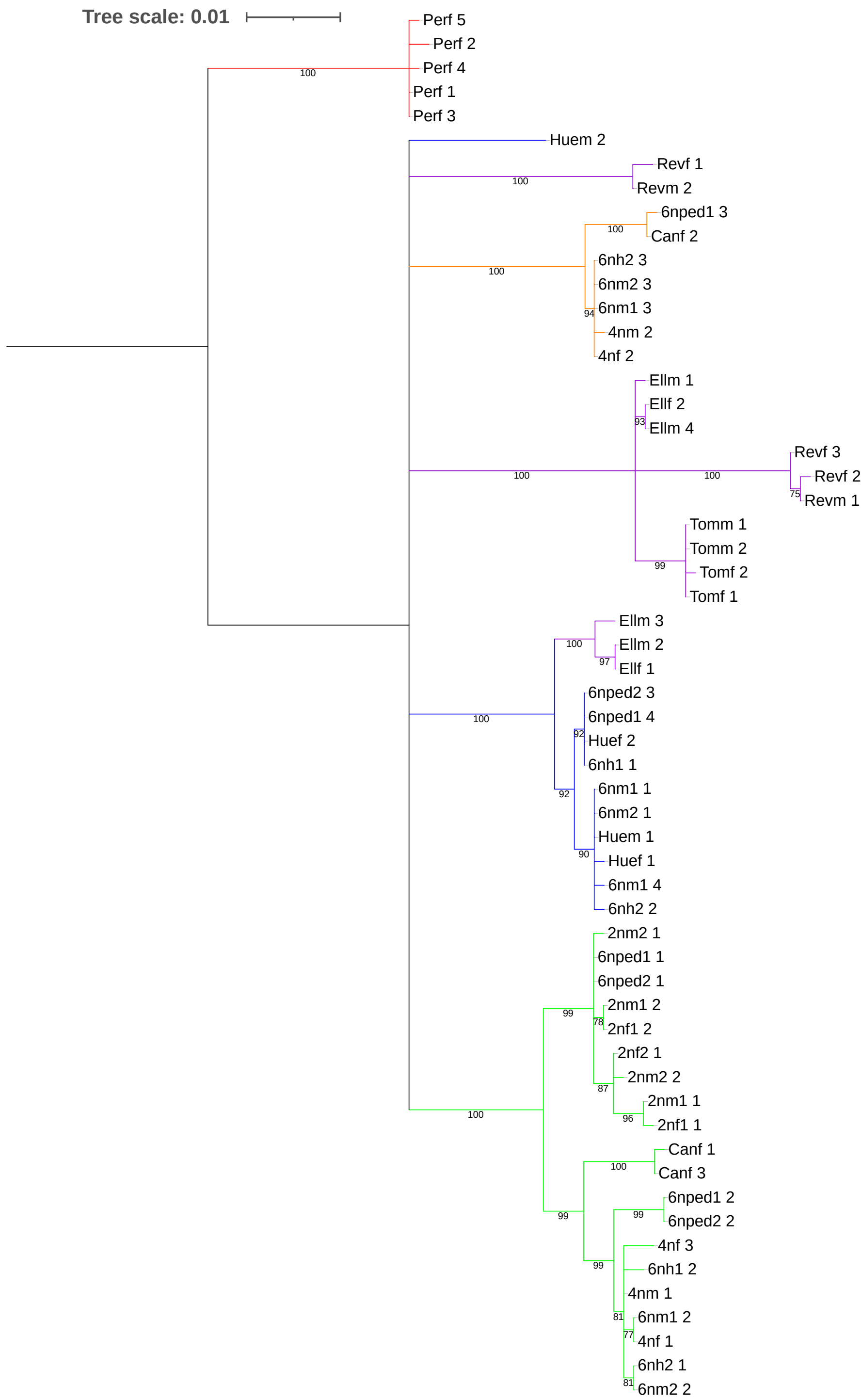
Tree scale: 0.01



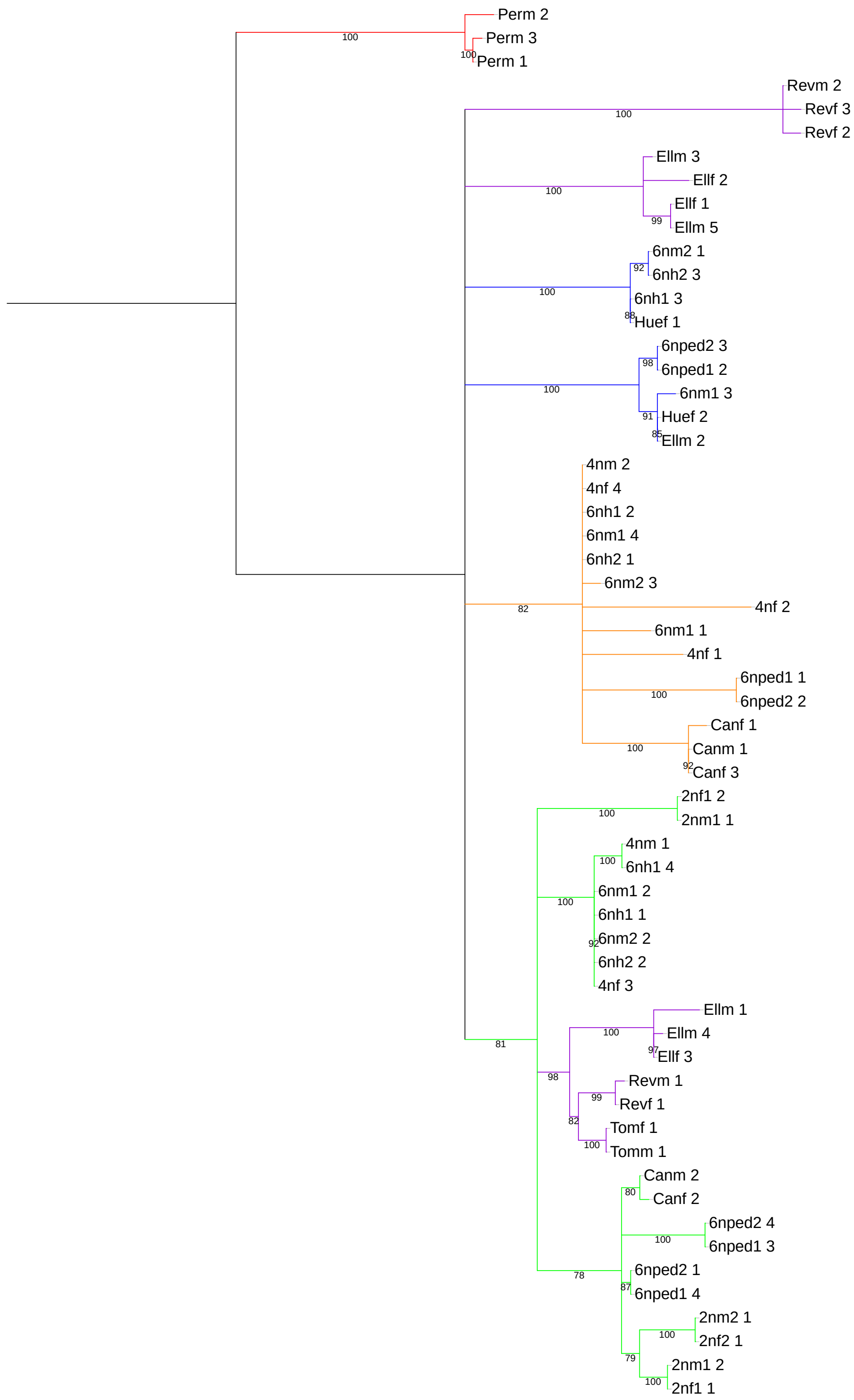
1_38



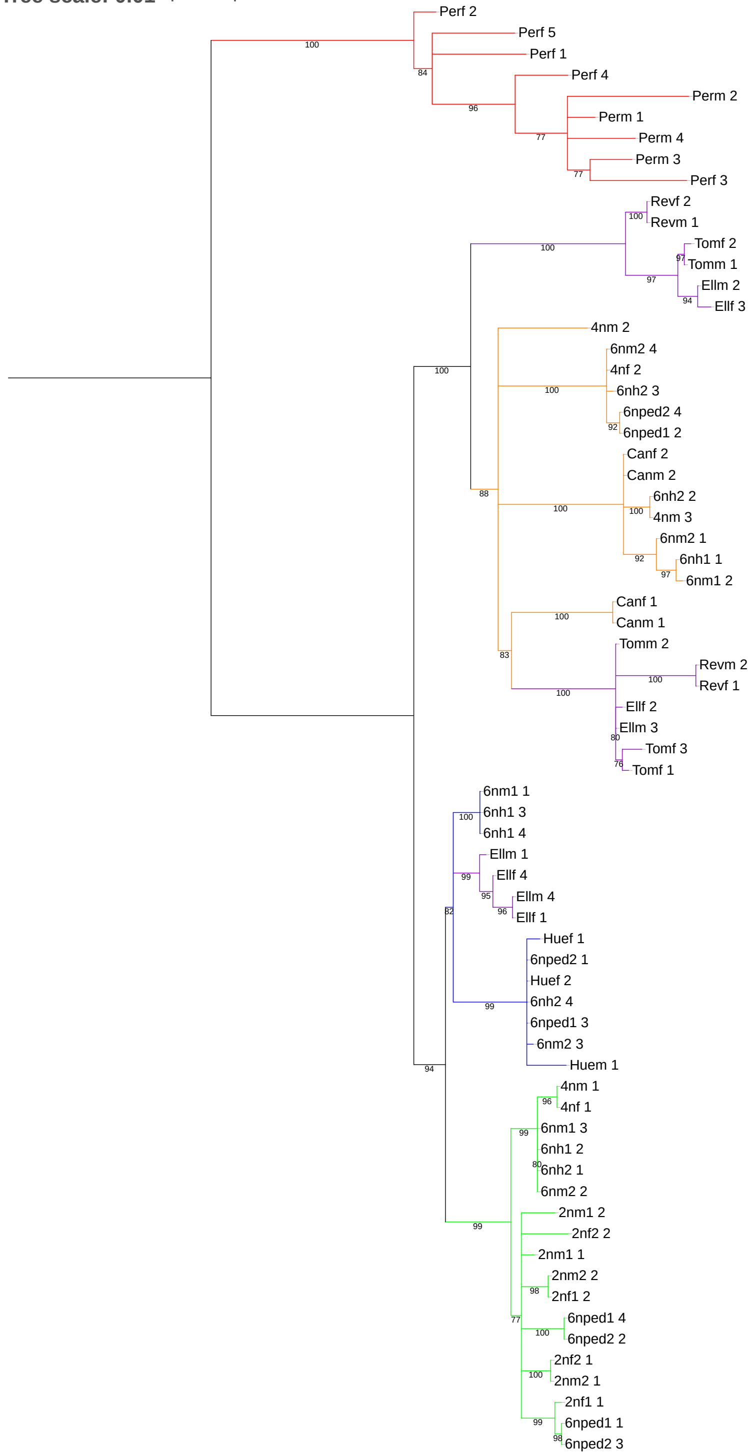
1_46



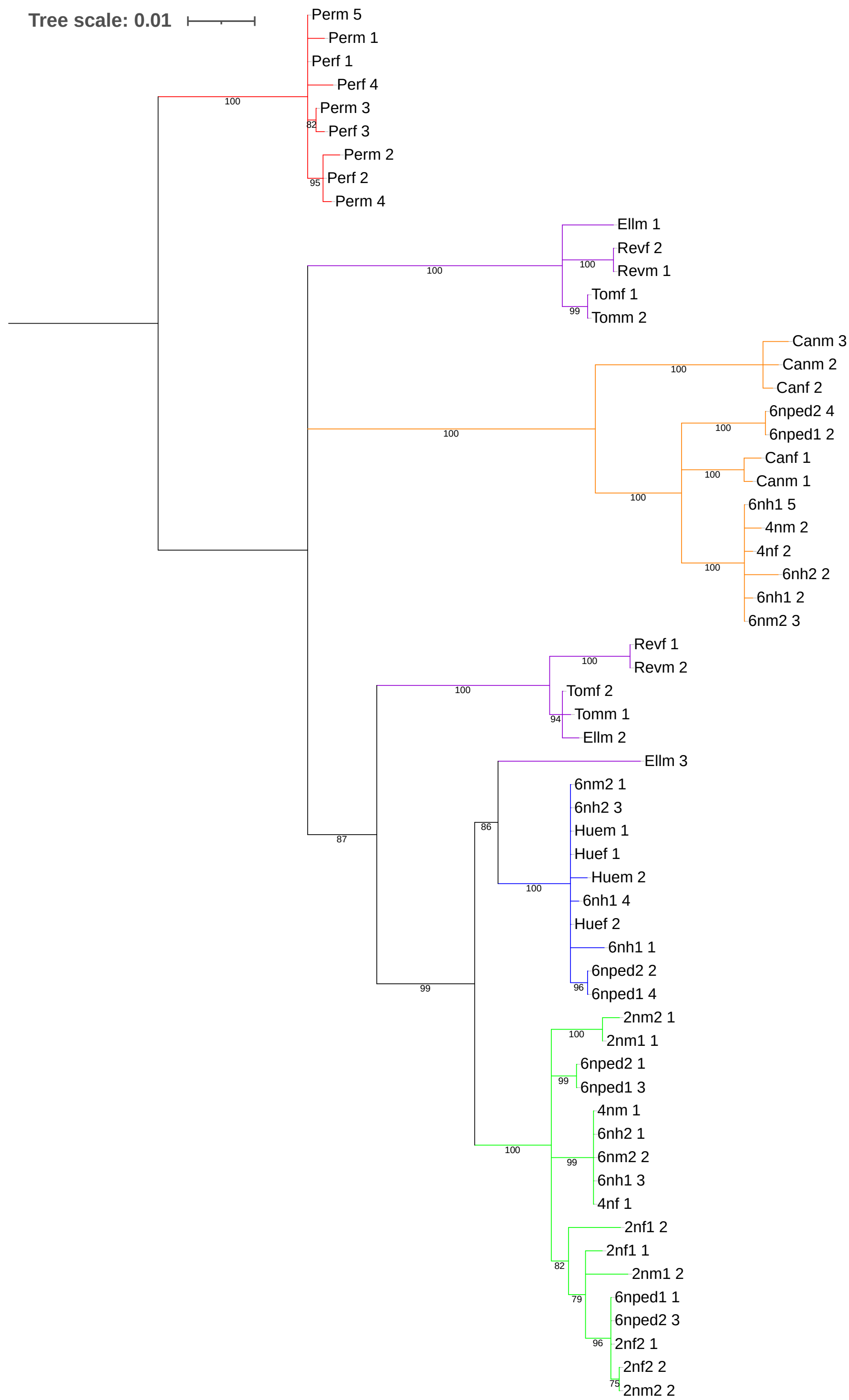
1_43



1_21

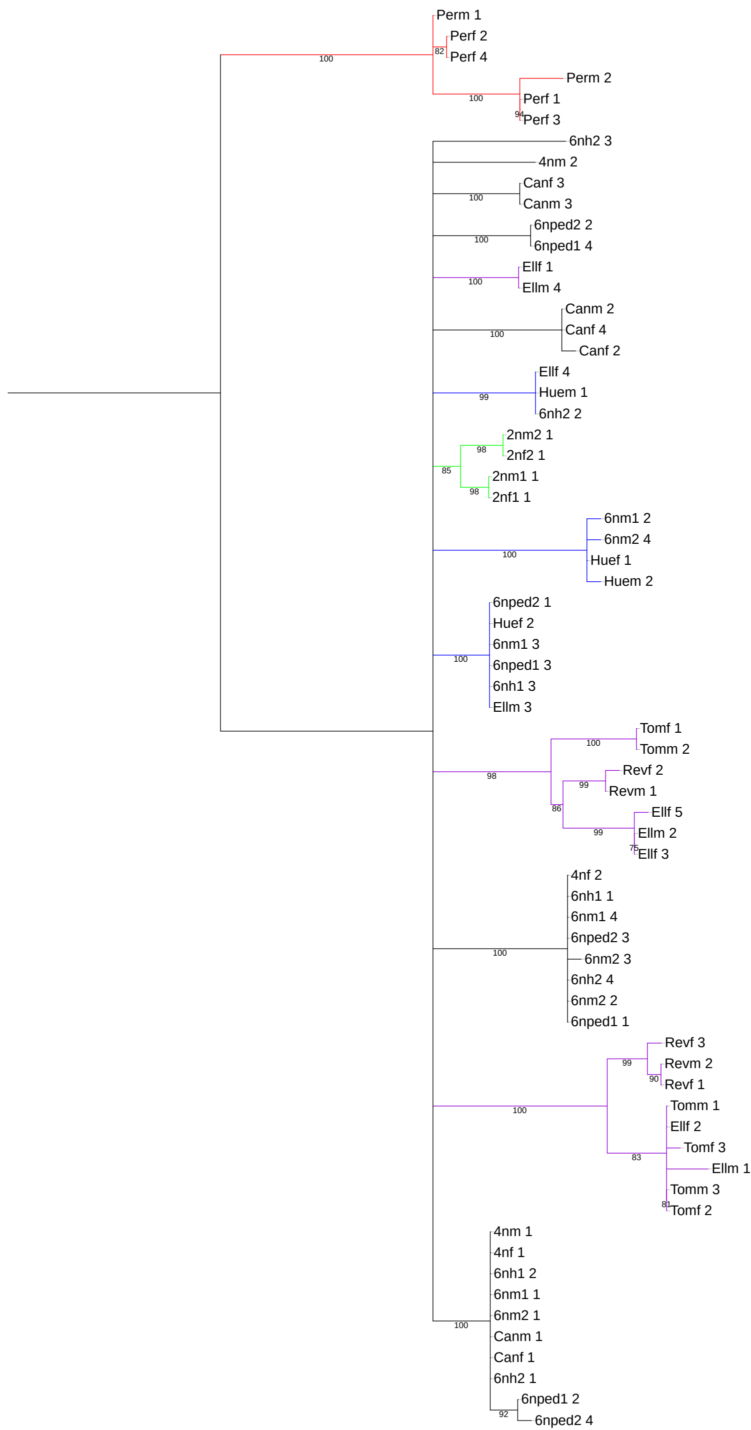


2_4



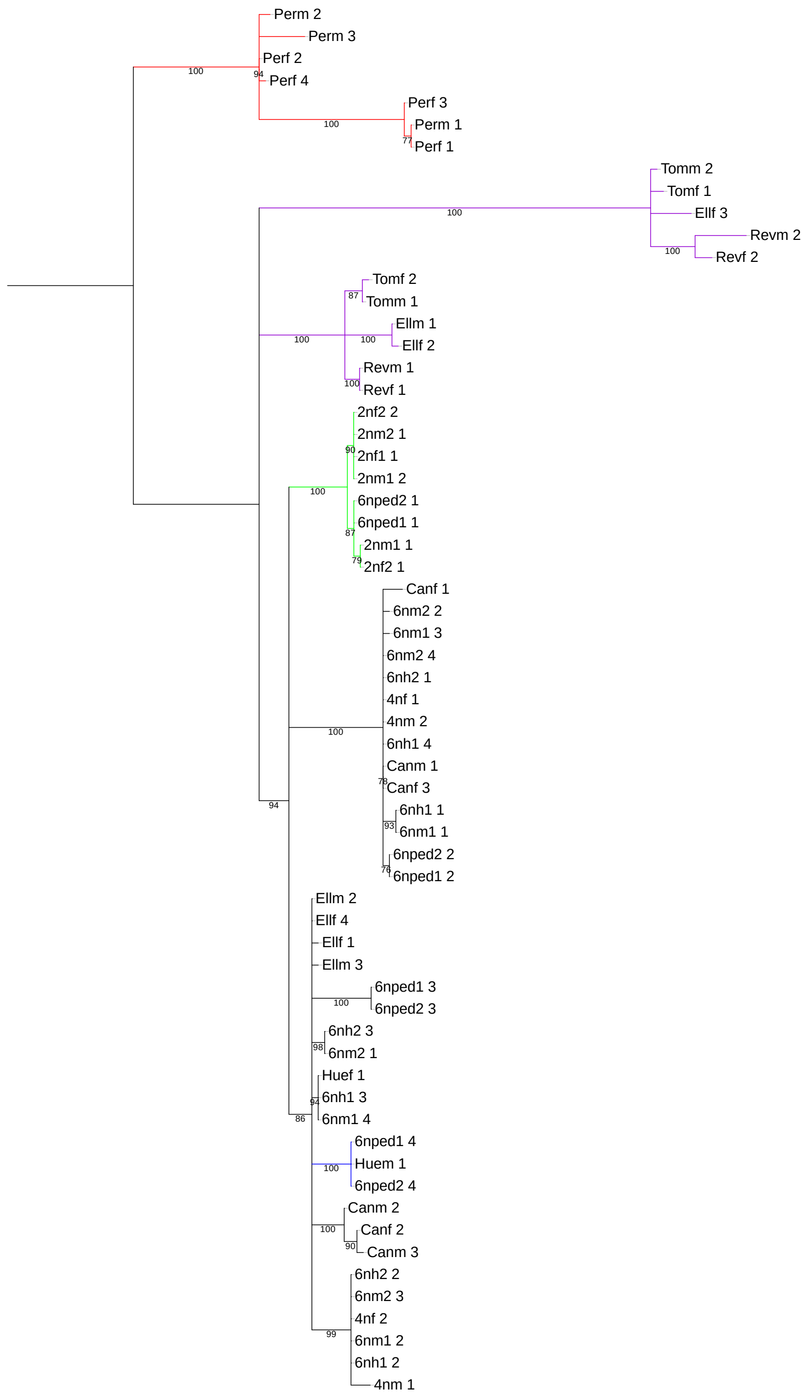
3_5

Tree scale: 0.01

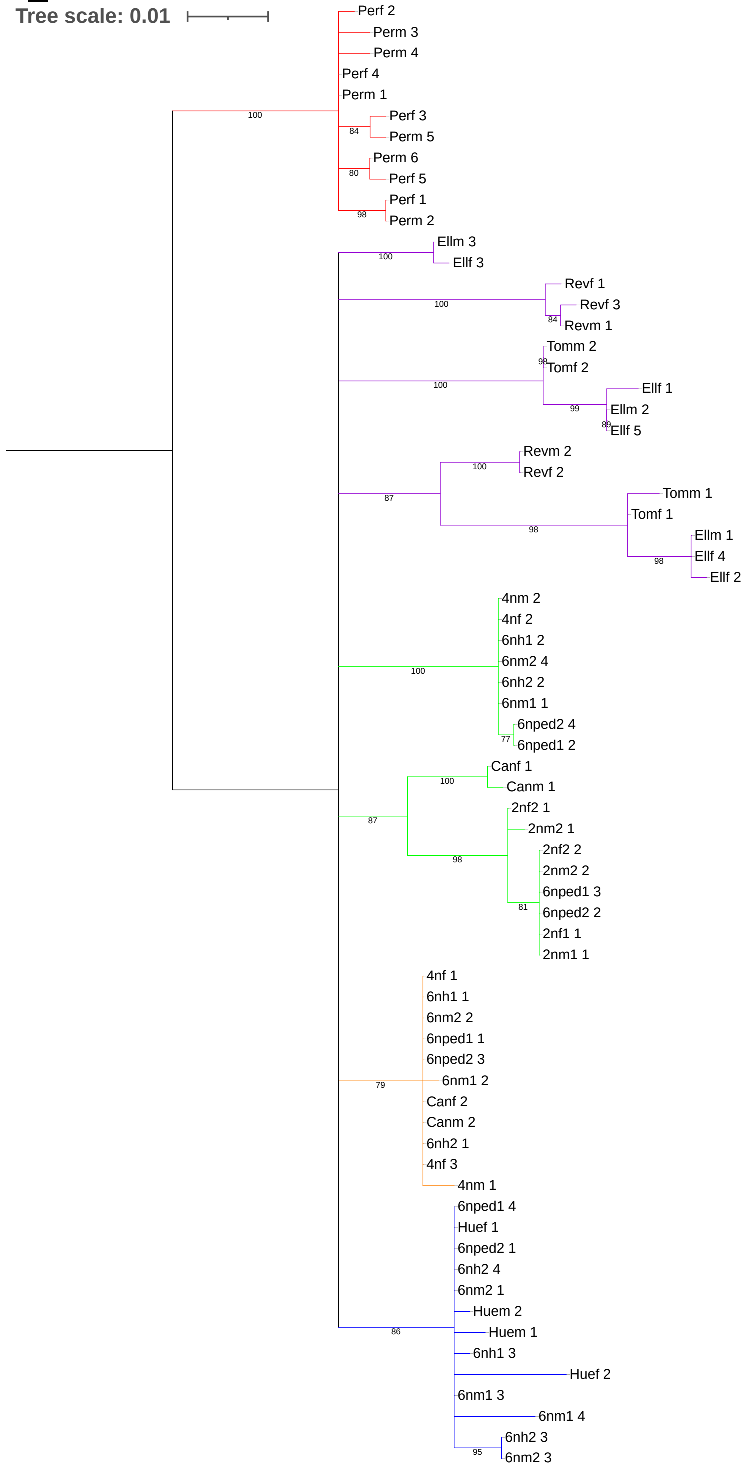


3_6

Tree scale: 0.1

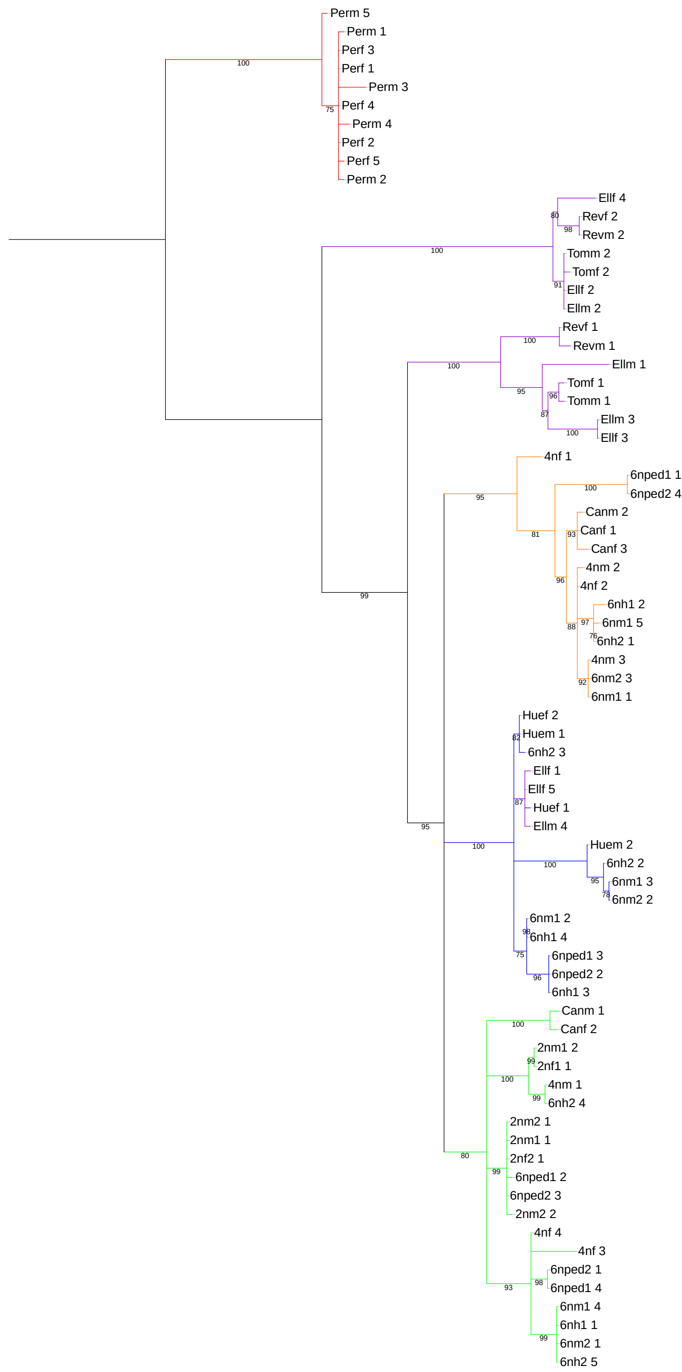


4_4



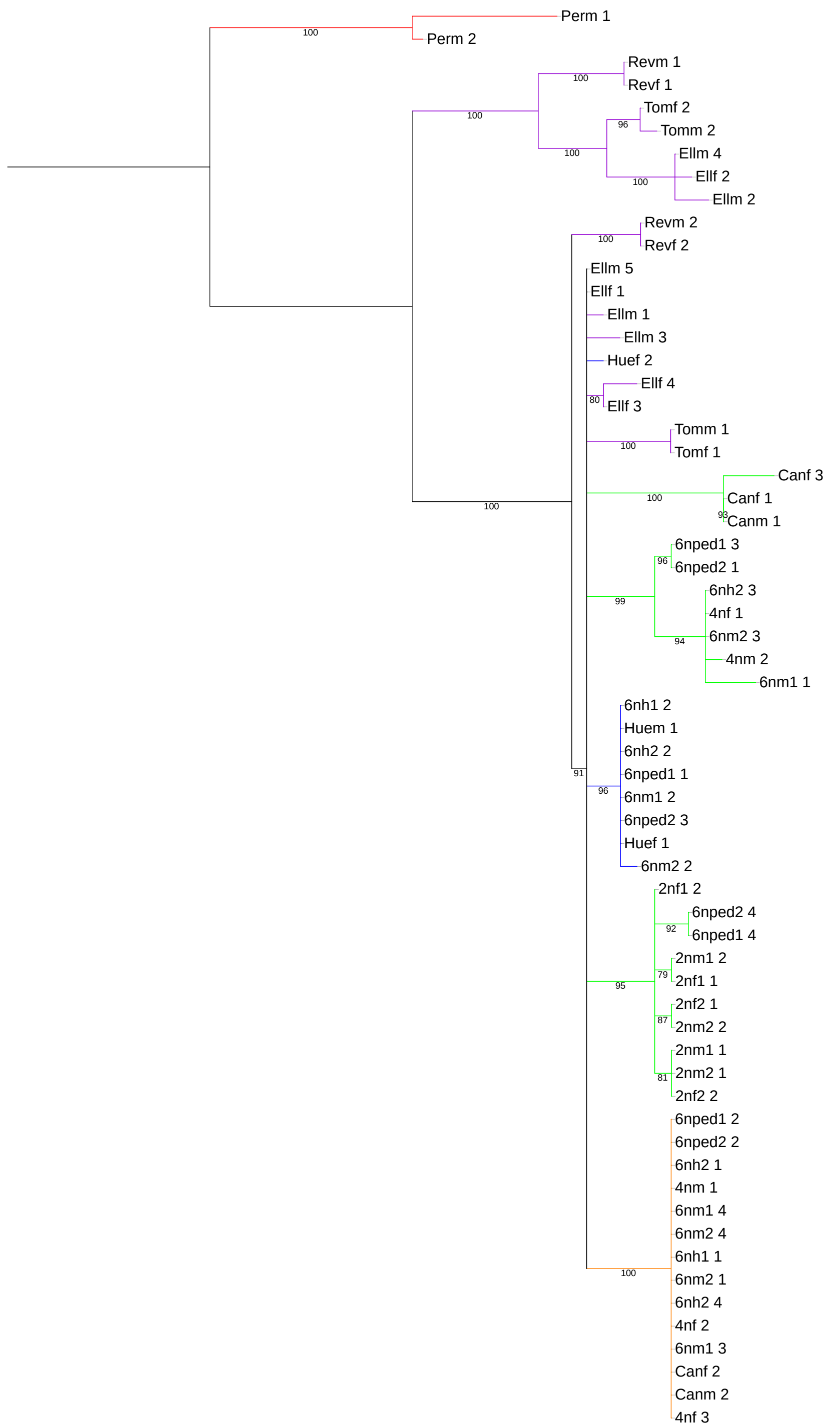
5_5

Tree scale: 0.1



6_4

Tree scale: 0.01



7_5

Tree scale: 0.01

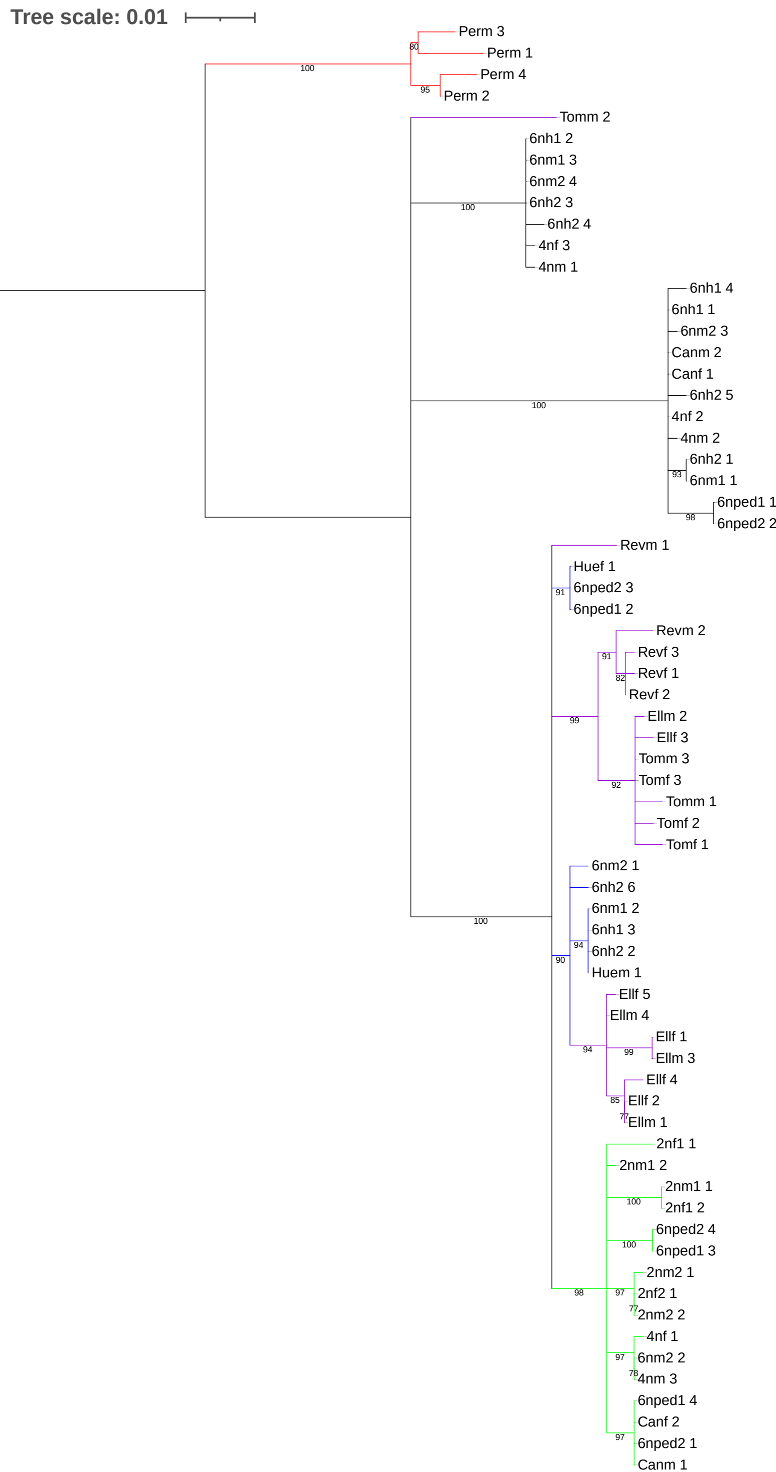


Figure S4: Maximum-likelihood phylogenetic trees based on phylogenetic markers located on the sex chromosomes and autosomes. Numbers below branches indicate bootstrap support. Samples belonging to the *perennis* clade are coloured red, samples belonging to the *annua* clade are coloured green, samples belonging to the *canariensis* clade are coloured orange and samples belonging to the other perennial lineages are coloured purple.

Table S3: Samples used for exon-capture data in Chapter 3. For sex, M indicates male, F indicates female, Mon indicates monoecious individuals.

Lineage	Sex	Sample ID	Origin	Ploidy
M. annua (2x)	M	1B	France	2x
M. annua (2x)	M	13A	Spain	2x
M. annua (2x)	M	21A	Greece	2x
M. annua (2x)	M	33A	Greece	2x
M. annua (2x)	M	38A	Turkey	2x
M. annua (2x)	M	43A	United Kingdom	2x
M. annua (2x)	M	49A	Spain	2x
M. annua (2x)	M	56A	Spain	2x
M. annua (2x)	M	7B	Turkey	2x
M. annua (2x)	F	4B	France	2x
M. annua (2x)	F	16A	Spain	2x
M. annua (2x)	F	24A	Greece	2x
M. annua (2x)	F	34A	Greece	2x
M. annua (2x)	F	40A	Turkey	2x
M. annua (2x)	F	46B	United Kingdom	2x
M. annua (2x)	F	52A	Spain	2x
M. annua (2x)	F	58A	Spain	2x
M. annua (2x)	F	10A	Turkey	2x
M. huetii	M	H16_1_1M	Spain	2x
M. huetii	M	H16_1_3M	Spain	2x
M. huetii	M	H16_2_3M	Spain	2x
M. huetii	M	H16_2_4M	Spain	2x
M. huetii	M	H16_3_1M	Spain	2x
M. huetii	M	H16_3_3M	Spain	2x
M. huetii	M	H16_3_8M	Spain	2x
M. huetii	M	H16_4_1M	Spain	2x
M. huetii	M	H16_4_4M	Spain	2x
M. huetii	F	H16_2_1F	Spain	2x
M. huetii	F	H16_2_2F	Spain	2x

M. huetii	F	H16_2_6F	Spain	2x
M. huetii	F	H16_3_1F	Spain	2x
M. huetii	F	H16_3_9F	Spain	2x
M. huetii	F	H16_3_10F	Spain	2x
M. huetii	F	H16_4_1F	Spain	2x
M. huetii	F	H16_4_2F	Spain	2x
M. huetii	F	H16_4_6F	Spain	2x
M. canariensis	M	Can212M	Teneriffe	4x
M. canariensis	M	Can213M	Teneriffe	4x
M. canariensis	M	Can212_19M	Teneriffe	4x
M. canariensis	M	Can209m4	Teneriffe	4x
M. canariensis	M	Can209m3	Teneriffe	4x
M. canariensis	M	Can209_M2	Teneriffe	4x
M. canariensis	M	Can209m1	Teneriffe	4x
M. canariensis	M	Can210M1	Teneriffe	4x
M. canariensis	F	Can212f3	Teneriffe	4x
M. canariensis	F	Can212f2	Teneriffe	4x
M. canariensis	F	Can212f1	Teneriffe	4x
M. canariensis	F	Can209f3	Teneriffe	4x
M. canariensis	F	Can209f2	Teneriffe	4x
M. canariensis	F	Can207_2f	Teneriffe	4x
M. canariensis	F	Can207_42f	Teneriffe	4x
M. canariensis	F	Can209_26f	Teneriffe	4x
M. annua (6x)	M	A30_28	Spain	6x
M. annua (6x)	M	A30_30	Spain	6x
M. annua (6x)	M	A29_28	Spain	6x
M. annua (6x)	M	A29_29	Spain	6x
M. annua (6x)	M	A20_29	Spain	6x
M. annua (6x)	M	A20_30	Spain	6x
M. annua (6x)	M	A11_28	Spain	6x
M. annua (6x)	M	A11_30	Spain	6x
M. annua (6x)	M	A5_29	Spain	6x
M. annua (6x)	M	A5_30	Spain	6x

M. annua (6x)	Mon	A30_1	Spain	6x
M. annua (6x)	Mon	A30_2	Spain	6x
M. annua (6x)	Mon	A29_2	Spain	6x
M. annua (6x)	Mon	A29_3	Spain	6x
M. annua (6x)	Mon	A20_4	Spain	6x
M. annua (6x)	Mon	A20_5	Spain	6x
M. annua (6x)	Mon	A11_2	Spain	6x
M. annua (6x)	Mon	A11_3	Spain	6x
M. annua (6x)	Mon	A5_2	Spain	6x
M. annua (6x)	Mon	A5_3	Spain	6x
M. tomentosa	M	Tom_M1	Spain	4x
M. tomentosa	M	Tom_M2	Spain	4x
M. tomentosa	M	Tom_M3	Spain	4x
M. tomentosa	M	Tom_M4	Spain	4x
M. tomentosa	M	Tom_M5	Spain	4x
M. tomentosa	M	Tom_M7	Spain	4x
M. tomentosa	M	Tom_M8	Spain	4x
M. tomentosa	M	Tom_M9	Spain	4x
M. tomentosa	F	Tom_F1	Spain	4x
M. tomentosa	F	Tom_F2	Spain	4x
M. tomentosa	F	Tom_F3	Spain	4x
M. tomentosa	F	Tom_F4	Spain	4x
M. tomentosa	F	Tom_F5	Spain	4x
M. tomentosa	F	Tom_F7	Spain	4x
M. reverchonii	M	Rev1_m1	Spain	4x
M. reverchonii	M	Rev1_m2	Spain	4x
M. reverchonii	M	Rev1_m3	Spain	4x
M. reverchonii	M	Rev1_m4	Spain	4x
M. reverchonii	M	Rev2_m1	Spain	4x
M. reverchonii	M	Rev2_m2	Spain	4x
M. reverchonii	M	Rev2_m3	Spain	4x
M. reverchonii	M	Rev2_m4	Spain	4x
M. reverchonii	M	Rev2_m5	Spain	4x

M. reverchonii	M	Rev2_m6	Spain	4x
M. reverchonii	F	Rev1_f1	Spain	4x
M. reverchonii	F	Rev1_f2	Spain	4x
M. reverchonii	F	Rev2_f1	Spain	4x
M. reverchonii	F	Rev2_f2	Spain	4x
M. reverchonii	F	Rev2_f3	Spain	4x
M. elliptica	M	Ell_M1	Spain	6x
M. elliptica	M	Ell_M2	Spain	6x
M. elliptica	M	Ell_M4	Spain	6x
M. elliptica	M	Ell_M6	Spain	6x
M. elliptica	M	Ell_M7	Spain	6x
M. elliptica	M	Ell_M8	Spain	6x
M. elliptica	M	Ell_M9	Spain	6x
M. elliptica	M	Ell_M10	Spain	6x
M. elliptica	M	Ell_M11	Spain	6x
M. elliptica	F	Ell_F1	Spain	6x
M. elliptica	F	Ell_F3	Spain	6x
M. elliptica	F	Ell_F4	Spain	6x
M. elliptica	F	Ell_F5	Spain	6x
M. elliptica	F	Ell_F6	Spain	6x
M. elliptica	F	Ell_F7	Spain	6x
M. elliptica	F	Ell_F9	Spain	6x
M. elliptica	F	Ell_F10	Spain	6x
M. elliptica	F	Ell_F11	Spain	6x

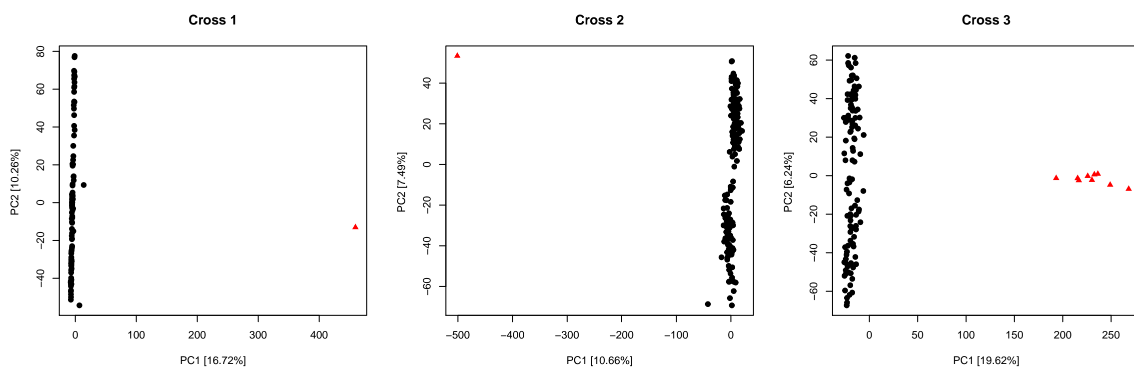


Figure S5: Plots of the first two principal components from a principal component analyses based on male *M. annua* genotypes of three crosses between male *M. annua* and female *M. huetii* used in Chapter 3. Black points represent individual samples, red triangles indicate outlier samples that were removed from the dataset. Values in square brackets indicate the proportion of variance explained by each principal component.

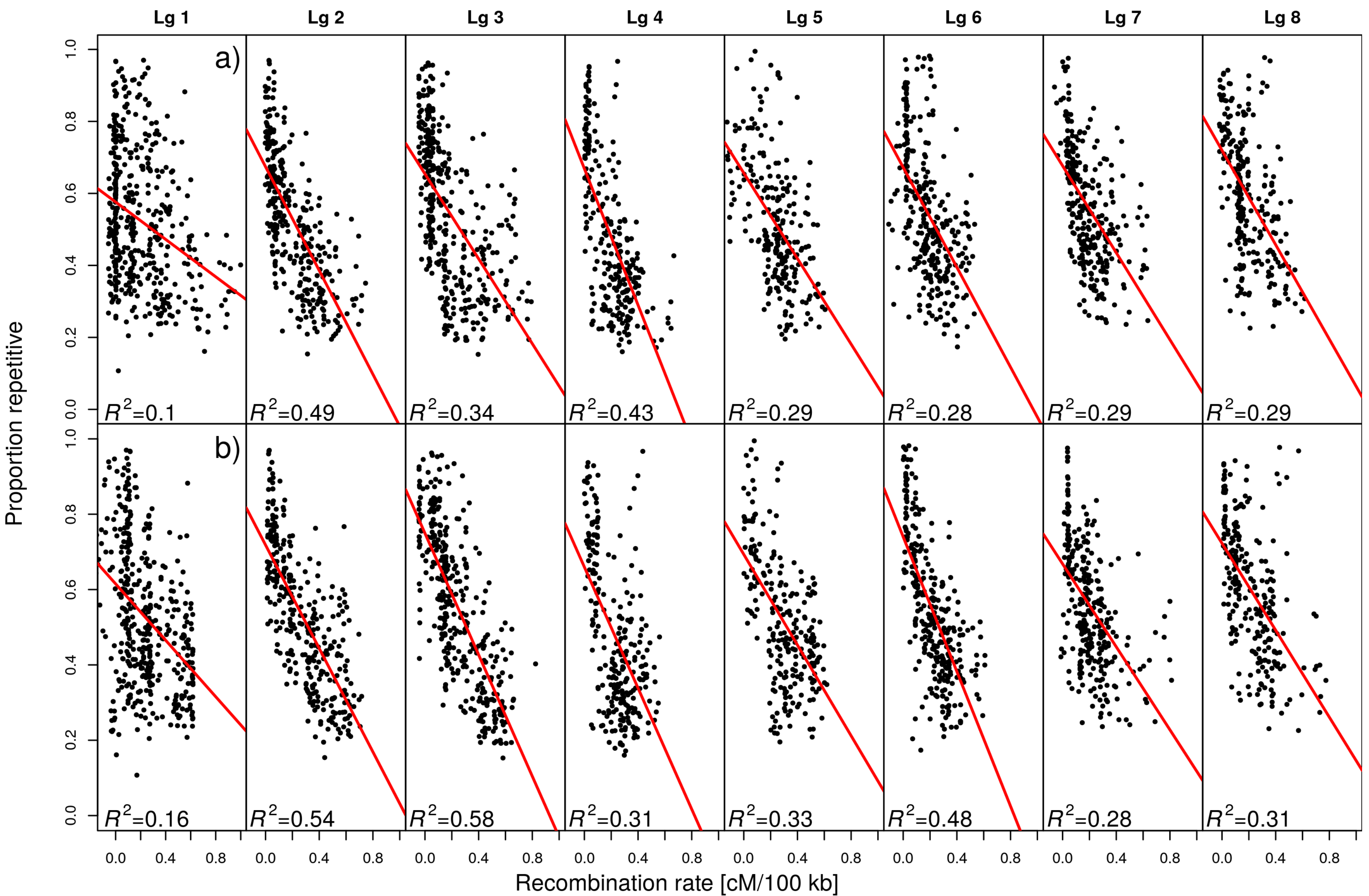


Figure S6: Proportion of sequence marked as repetitive in 100 kb windows plotted against the recombination rate per window for the new male *M. annua* linkage map (a) and the F_2 linkage map (b). Red lines are linear regression models, the value at the bottom of each plot is the proportion of variance explained by the model (R^2).

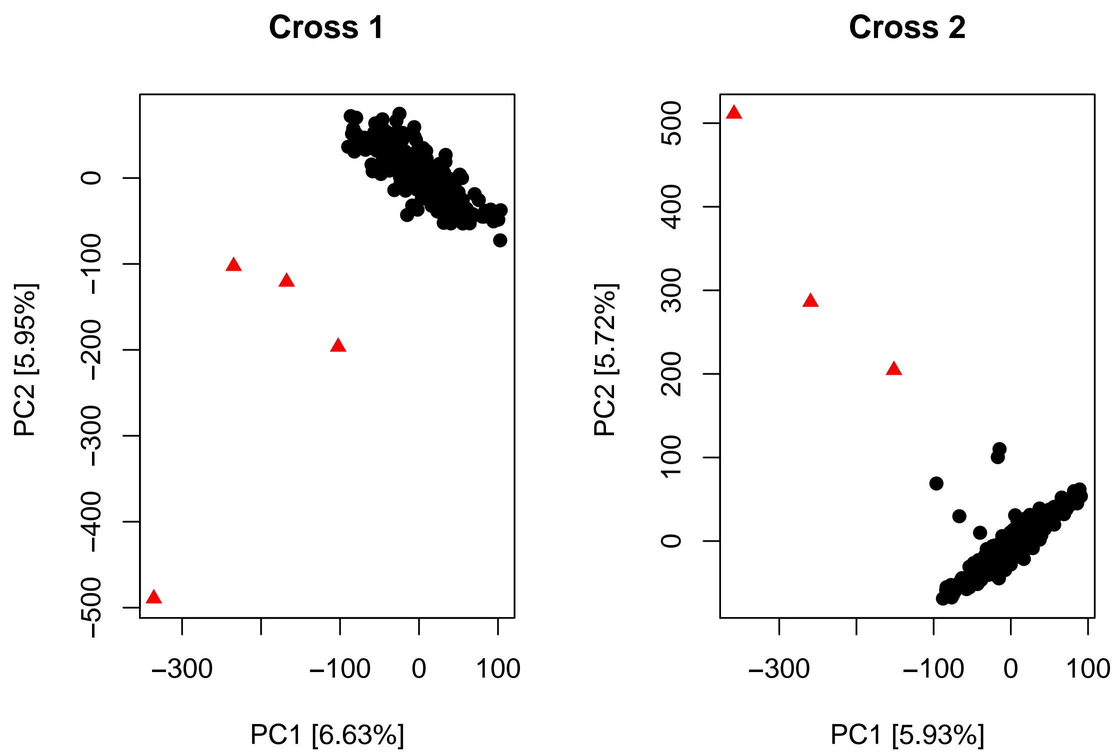


Figure S7: Plots of the first two principal components from a principal component analyses based on the genotypes of F_2 offspring of both crosses used in Chapter 4. Black points represent individual samples, red triangles indicate outlier samples that were removed from the dataset. Values in square brackets indicate the proportion of variance explained by each principal component.