Genetics: Early Online, published on May 21, 2019 as 10.1534/genetics.119.302045

Early sex-chromosome evolution in the diploid dioecious plant *Mercurialis annua*

Paris Veltsos[*,†], Kate E. Ridout[*, §,‡], Melissa A. Toups[*,**,††], Santiago C. González-Martínez[‡‡], Aline Muyle[§§], Olivier Emery[***], Pasi Rastas[†††], Vojtech Hudzieczek[§§§], Roman Hobza[§§§], Boris Vyskot[§§§], Gabriel A.B. Marais[§§], Dmitry A. Filatov[§], and John R. Pannell[*]

[*] Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland

[†] Current address: Department of Biology, Jordan Hall, 1001 East Third Street, Indiana University, Bloomington, IN 47405, United States of America

[§] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, United Kingdom

[‡] Department of Oncology, John Radcliffe Hospital, Oxford, OX3 9DU, United Kingdom

[**] Current address: Institute of Science and Technology Austria, Am Campus 1A, Klosterneuburg 3400, Austria

[††] Department of Integrative Biology, University of Texas, Austin, Texas, United States of America

[‡‡] BIOGECO, INRA, Univ. Bordeaux, 33610 Cestas, France

[§§] Laboratoire Biométrie et Biologie Évolutive (UMR 5558), CNRS / Université Lyon 1, 69100, Villeurbanne, France.

[***] Current address: Department of Fundamental Microbiology, University of Lausanne, CH-1015 Lausanne, Switzerland.

[†††] University of Helsinki, Institute of Biotechnology, P.O.Box 56, 00014, Finland

[§§§] Department of Plant Developmental Genetics, Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 61200, Brno, Czech Republic.

The first two authors contributed equally to the research.

Copyright 2019.

The last three authors supervised the research.

Running title: Sex chromosome evolution in *Mercurialis annua*

Key words: sex chromosomes, whole genome sequencing, sex linkage, evolutionary strata, gene expression

Author for correspondence: John R. Pannell

Email: john.pannell@unil.ch

Telephone: +4121 692 41 70

Department of Ecology and Evolution

Le Biophore - Unil Sorge

University of Lausanne

Lausanne

1015

Switzerland

Abstract

Suppressed recombination allows divergence between homologous sex chromosomes and the functionality of their genes. Here, we reveal patterns of the earliest stages of sex-chromosome evolution in the diploid dioecious herb *Mercurialis annua* on the basis of cytological analysis, *de novo* genome assembly and annotation, genetic mapping, exome resequencing of natural populations, and transcriptome analysis. The genome assembly contained 34,105 expressed genes, of which 10,076 were assigned to linkage groups. Genetic mapping and exome resequencing of individuals across the species range both identified the largest linkage group, LG1, as the sex chromosome. Although the sex chromosomes of *M. annua* are karyotypically homomorphic, we estimate that about a third of the Y chromosome has ceased recombining, containing 568 transcripts and spanning 22.3 cM in the corresponding female map. Nevertheless, we found limited evidence for Y-chromosome degeneration in terms of gene loss and pseudogenization, and most X- and Y-linked genes appear to have diverged in the period subsequent to speciation between *M. annua* and its sister species *M. huetii* which shares the same sex-determining region. Taken together, our results suggest that the *M. annua* Y chromosome has at least two evolutionary strata: a small old stratum shared with *M. huetii*, and a more recent larger stratum that is probably unique to *M. annua* and that stopped recombining about one million years ago. Patterns of gene expression within the non-recombining region are consistent with the idea that sexually antagonistic selection may have played a role in favoring suppressed recombination.

Article summary Plants that evolved separate sexes (dioecy) recently are ideal models for studying the early stages of sex-chromosome evolution. Here, we use karyological, whole genome and transcriptome data to characterize the homomorphic sex chromosomes of the annual dioecious plant *Mercurialis annua*. Our analysis reveals many typical hallmarks of dioecy and sex-chromosome evolution, including sex-biased gene expression and high X/Y sequence divergence, yet few premature stop codons in Y-linked genes and very little outright gene loss, despite 1/3 of the Y chromosome having ceased recombination. Our results confirm that the *M. annua* species complex is a fertile system for probing early stages in the

evolution of sex chromosomes.

Introduction

The evolution of dioecy (separate sexes) from hermaphroditism, which has occurred repeatedly in independent lineages of flowering plants (Renner, 2014), is a prelude to the possible evolution of sex chromosomes. Early sex-chromosome evolution typically involves the accumulation of repetitive sequences in a non-recombining region (Wang *et al.*, 2012a; Hobza *et al.*, 2017), differences in codon use between homologues (Ono, 1939; Qiu *et al.*, 2010), different patterns of gene expression at sex-linked loci (Zemp *et al.*, 2016), pseudogenization and gene loss (Papadopulos *et al.*, 2015; Wu and Moore, 2015), and ultimately divergence in chromosome length between homologues (Puterova *et al.*, 2018). Extreme divergence is common in many animals, but it is also known in some plants in which the homologous chromosomes are heteromorphic and distinguishable by karyotype, e.g., in the plant species *Silene latifolia* (Ono, 1939; Krasovec *et al.*, 2018) and *Rumex hastatulus* (Smith, 1955; Hough *et al.*, 2014). In other plants, the sex chromosomes remain indistinguishable by karyotype, and gene function is only mildly compromised, e.g., *Asparagus officinalis* (Loeptien, 1979; Telgmann-Rauber *et al.*, 2007), *Spinacia oleracea* (Yamamoto *et al.*, 2014), *Diospyros lotus* (Akagi *et al.*, 2014), *Fragaria chiloensis* (Tennessen *et al.*, 2016), *Populus* (Geraldes *et al.*, 2015), *Carica papaya* (Horovitz and Jiménez, 1967; Liu *et al.*, 2004) and *Salix* (Pucholt *et al.*, 2015). Because of this variation, and because dioecy in plants has often evolved recently, plants with young homomorphic sex chromosomes provide particularly good models for studying the very earliest stages in sex-chromosome divergence (Charlesworth, 2016).

Two hypotheses have been proposed to explain the suppression of recombination in plants. First, if dioecy evolves through the spread of male- and female-sterility mutations, these mutations must become linked on opposite chromosomes to avoid the expression of either hermaphroditism, or both male and female sterility simultaneously (Charlesworth and Charlesworth, 1978). The main experimental support for this two-locus model comes from classic genetic studies in *Silene latifolia* (Westergaard, 1958) that

demonstrated the presence of two sex-determining factors on the Y-chromosome, the stamen-promoting factor (SPF) and gynoecium suppression factor (GSF). More recent work mapped the location of these genes on the *S. latifolia* Y-chromosome (Kazama *et al.*, 2016), although the actual GSF and SPF genes have yet to be identified. Nevertheless, although there is some support for it, the two-locus model does not explain why non-recombining regions on sex chromosomes often expand greatly (Bergero and Charlesworth, 2009), well beyond the region harboring the original sex-determining genes.

A second hypothesis invokes selection favoring suppressed recombination between a sex-determining locus and loci elsewhere on the sex chromosome that have different allelic effects on the fitness of males and females, i.e., alleles with sexually antagonistic effects (Rice, 1987; Charlesworth, 1991; Gibson *et al.*, 2002; Charlesworth *et al.*, 2005; Bergero and Charlesworth, 2009). The suppression of recombination is expected to extend consecutively to generate linkage between the sex-determining locus and more sexually antagonistic loci (Charlesworth, 2015), and these extensions can be identified as discrete 'strata', with greater X/Y divergence in strata that ceased recombining earliest. Evidence for strata has been found both in animals (Lahn and Page, 1999; Nam and Ellegren, 2008) and plants (Bergero *et al.*, 2007; Wang *et al.*, 2012b), but there is still little direct evidence for the role played by sexually antagonistic selection in bringing them about (Bergero *et al.*, 2019). A recent study of guppy sex chromosomes claimed evidence for the evolution of strata consistent with the sexual-antagonism hypothesis (Wright *et al.*, 2017), but subsequent work has shown that strata could actually not have been involved in the evolution of suppressed recombination (Bergero *et al.*, 2019). Although our understanding of the implications of suppressed recombination is well developed, evidence for its driver therefore remains weak.

Sexually antagonistic selection may be resolved either through differential gene expression between males and females, or through sex linkage of the responsible loci, with the phenotypic expression of sexual dimorphism being the ultimate result. Sexual dimorphism is known in dioecious species for a wide range of morphological (Eckhart, 1999), life-history (Delph, 1999) and physiological traits (Dawson and Geber,

1999), as well as at the level of gene expression (Baker *et al.*, 2007; Sharma *et al.*, 2014; Mohanty *et al.*, 2017; Sanderson *et al.*, 2018). In some cases, the differentially expressed genes are enriched on the sex chromosomes (Parisi *et al.*, 2003; Leder *et al.*, 2010; Albritton *et al.*, 2014; Pucholt *et al.*, 2017), likely as a result of the degeneration of one of the diploid copies following the suppression of recombination, or due to interactions between sex and gene expression at the loci concerned (reviewed in (Mank and Ellegren, 2009; Parsch and Ellegren, 2013). In the latter scenario, dosage compensation may subsequently evolve in response to selection to restore similar levels of expression in males and females. While dosage compensation is an important feature of gene expression in many animal lineages (Mank, 2013), its extent in plants is understudied. The clearest analysis remains that for *S. latifolia*, in which Papadopulos *et al.* (2015) confirmed gene loss, or lost expression, for many Y-linked genes, with associated incomplete dosage compensation from X-linked homologues, including full compensation at some loci. Further study of sequence evolution and patterns of gene expression in other species would be valuable, particularly those with homomorphic sex chromosomes at the very earliest stages of their evolution, as well as for loci close to the sex-determining locus.

Here, we identify the homomorphic sex chromosomes of the wind-pollinated dioecious annual plant *Mercurialis annua* (Euphorbiacae) using genetic mapping and sequence analysis of the genome and patterns of gene expression. Until recently, sex in dioecious *M. annua* was thought to be determined by allelic variation at three independent loci (Durand *et al.*, 1987; Durand and Durand, 1991), but it is now known to have a simple XY system (Khadka *et al.*, 2005; Russell and Pannell, 2015). YY males of *M. annua*, which lack an X chromosome, are viable but partially sterile (Li et al; Kuhn, 1939), indicating that the Y chromosome is only mildly degenerate. Our present study confirms this suggestion, finding very limited gene loss, yet suppressed recombination over a large portion of the sex chromosomes and substantial differences in gene expression between males and females at sex-linked loci. It would thus appear that the sex chromosomes of the species are at a particularly interesting and potentially revealing stage in their evolution (Veltsos *et al.*, 2018).

To assemble the genome of *M. annua*, we combined short-read sequencing on the Illumina platform with long-read technology developed by Pacific Biosciences. We also analyzed the karyotype of *M. annua* using current imaging technology in an attempt to identify heteromorphism at the cytological level, and we obtained SNPs from the transcriptomes of small families to construct a genetic map for *M. annua*, to identify non-recombining (sex-linked) genes and scaffolds, and to compare them with non-sex-linked regions. We then sought evidence for the evolution of evolutionary strata and genetic degeneration on the Y chromosome, including the fixation of pseudogenising mutations in the inferred fully Y-linked sequences, and the proportion of genes deleted from the Y. Finally, we examined sex-biased gene expression and assessed whether sex-biased genes might be enriched on the sex chromosomes, as expected by theory (Connallon and Clark, 2010; Meisel *et al.*, 2012). Our results suggest that the sex chromosomes of *M. annua* are about 1.5 million years old. Early stages of Y-chromosome degeneration are clearly apparent, but there are also signs that patterns of gene expression might have been affected by the accumulation of sexually antagonistic mutations.

Materials and Methods

Hairy root culture, chromosome preparation and cytogenetics

Mercurialis annua seeds were sterilized by incubation in 4% sodium hypochloride and subsequently washed in 50% ethanol and sterile water. Seeds were grown on MS medium and leaf discs were taken from 2-3 week old plants, which had been sexed using a previously developed Sequence Characterised Amplified Region (SCAR) marker (Khadka *et al.*, 2002). *Agrobacterium rhizogenes* strain ARqua1 (Quantdt *et al.*, 1993) was grown overnight at 28°C in LB medium supplemented with 300 µM acetosyringone to OD_600 = 0.6. The culture was resuspended in liquid 1/2 MS with 300 µM acetosyringone and used for direct inoculation of *Mercurialis annua* leaf discs. Explants were cocultivated at 28°C on MS medium with 300µM acetosyringone for 2 days. After cocultivation, explants were moved to MS medium supplemented with 300µg/L Timentin. Media were changed every 2 weeks.

To synchronize the hairy roots of *Mercurialis annua*, the DNA polymerase inhibitor aphidicolin was

added for 12 h. Mitoses were then accumulated in protoplasts using oryzalin treatment, transferred on chromic acid washed slides by dropping, and stored at -20 ^{o}C until use. For FISH experiments, slides were denatured in 7:3 (v/v) formamide: 2xSSC for 2 min at 72°C, immediately dehydrated through 50%, 70%, and 100% ethanol (-20°C), and air dried. The hybridization mixture (30 µl per slide) consisted of 200 ng of labeled probe, 15 µl formamide, 6 µl 50% dextrane sulfate, and 3 µl of 20x SSC. The volume was brought to 30 µl by adding TE, pH 8. The probes were denatured at 70°C for 10 min, and slides hybridized for 18 h at 37°C in a humid chamber. Slides were analyzed using an Olympus Provis AX70 microscope, and image analysis was performed using ISIS software (Metasystems). DNA was labeled with Fluorolink Cy3-dUTP (Amersham Pharmacia Biotech; red labeling) using a Nick Translation mix (Roche) or with SpectrumGreen direct-labeled dUTP (Vysis; green labeling) and a Nick Translation kit (Vysis).

Genome and transcriptome sequencing

DNA for genomic sequencing was taken from leaf tissue of a single male, M1. RNA samples were collected from leaves and flower buds of this individual as well as from three unrelated females, G1, G2 and G3, and two additional unrelated males, M2 and M3, all sampled from north-western France and grown together in a glasshouse. 65 F_1 and F_2 progeny (five families) were then produced by crossing G1xM1 and G2xM1 (F_1) as well as three pairs of G1xM1 progeny (F_2) (Table S1), which were also used for RNA extraction and transcriptome sequencing. DNA was extracted using a Qiagen Plant DNeasy kit (Qiagen, Hilden, Germany). Illumina paired-end and mate-pair sequencing was carried out by the Beijing Genomics Institute (BGI) using Illumina HiSeq 2000 technology (100 bp reads). Pacific Biosciences long-read sequencing was performed on the individual M1 by the Centre for Integrative Genomics hosted at the University of Lausanne. RNA was extracted from a mixture of flower buds and leaf tissues using the Qiagen plant RNAeasy kit (Qiagen, Hilden, Germany), and individual libraries were prepared for all 65 individuals (Table S1), which were sequenced on three lanes of Illumina HiSeq 2000 at the Wellcome Trust Centre for Human Genetics, Oxford.

Transcriptome assembly and annotation

The SEX-DETector pipeline (http://lbbe.univ-lyon1.fr/-SEX-DETector-.html?lang=fr) was used to produce the reference transcriptome (henceforth 'ORFs'), and to obtain allele-specific expression and SNPs for X and Y haplotypes. Sex linkage was inferred from genetic mapping (see below). To identify sex-linked haplotypes, we employed a probabilistic model based on maximum-likelihood inference, implemented in SEX-DETector (Muyle *et al.*, 2016). SEX-DETector, which assumes that X-linked and Y-linked haplotypes are passed only from fathers to daughters or from fathers to sons, respectively, is embedded into a Galaxy workflow pipeline that includes extra-assembly, mapping and genotyping steps prior to sex-linkage inference, which has been shown to have greater sensitivity, without an increased false positive rate, than without these steps (Muyle *et al.*, 2016). First, poly-A tails were removed from transcripts using PRINSEQ (Schmieder and Edwards, 2011), with parameters -trim_tail_left 5 -trim_tail_right 5. rRNA-like sequences were removed using riboPicker version 0.4.3 (Schmieder and Edwards, 2011) with parameters -i 90 -c 50 -l 50 and the following databases: SILVA Large subunit reference database, SILVA Small subunit reference database, the GreenGenes database and the Rfam database. Transcripts were then further assembled within Trinity components using cap3 (Huang and Madan, 1999), with parameter -p 90 and custom Perl scripts. Coding sequences were predicted using Trinity TransDecoder (Haas *et al.*, 2013), including PFAM domain searches as ORF retention criteria; these sequences were taken as our reference transcriptome. The RNAseq reads from parents and progeny were mapped onto the reference transcriptome using BWA (Li and Durbin, 2009). The alignments were analyzed using reads2snp, a genotyper for RNAseq data that gives better results than standard genotypers when X and Y transcripts have different expression levels (Tsagkogeorga *et al.*, 2012).

The SEX-DETector transcripts were mapped onto the genome using GMAP v2017-06-20 (Wu and Watanabe, 2005), disallowing chimeric genes, mapping 38,963 (99%) of them to 48,298 loci. We then collapsed the inferred ORFs into 34,105 gene models using gffread v0.9.9 (http://ccb.jhu.edu/software/stringtie/gff.shtml). In total, 26,702 (67.9%) of the ORFs could be annotated using fast blastx on the Swiss-Prot database and Blast2GO (Conesa *et al.*, 2005) in Geneious v9 (Kearse *et*

al., 2012). The resulting transcriptome annotation information, including GO terms, is presented in File S4. We characterized genes as candidate sex-determining genes by identifying those containing the following strings in their annotation: flower, auxin, cytokinin, pollen, kelch, ethylene, swi, retinol, calmodulin, short-root, jasmonic. This manual characterization allowed us to visualize the mapping locations, test for their enrichment in the non-recombining region, and identify whether any were among the outliers in various transcript-associated metrics.

Genome size estimation, assembly and annotation

The genome size was estimated from the distribution of k-mers (size 31) obtained from jellyfish v2.1.0 (Marçais and Kingsford, 2011) and analyzed by the web version of GenomeScope v1 (Vurture *et al.*, 2017), limiting max k-mer coverage to 1000.

Sliding-window trimming and adaptor removal were carried out using Trimmomatic v. 0.30, with the default parameters (Bolger *et al.*, 2014). Exact duplicate read pairs were collapsed using fastx-collapser from the Fastx-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Low complexity masking was carried out using DUST (Morgulis *et al.*, 2006), with the default parameters, and ambiguous reads were removed. Pacific Biosciences long reads were error-corrected using Bowtie2 version 2.1.0 (Langmead and Salzberg, 2012) in combination with LSC version 0.3.1 (Au *et al.*, 2012).

Filtered paired-end and mate-paired genomic reads were assembled using SOAPdenovo2 (Luo *et al.*, 2012), with k-mer values between 35 and 55 (odd values only). The best assembly was chosen using REAPR (Hunt *et al.*, 2013). GapCloser (Luo *et al.*, 2012) was run on the best assembly to correct false joins and fill gaps. Error-corrected Pacific Biosciences reads were then used to extend scaffolds, to fill gaps, and to join scaffolds using PBJelly2 (English *et al.*, 2012). Additional scaffolding was carried out using default parameters on SSPACE (Boetzer *et al.*, 2011), which revisits gaps using existing paired-end and mate-paired sequences. This step is intended to correct for any ambiguity introduced by the low-coverage of PacBio reads. Finally, L_RNA_Scaffolder (Xue *et al.*, 2013) was used to bridge genomic scaffolds using

the transcript assembly.

Transposable elements (TEs) and tandem repeats were predicted using a combination of Tandem Repeats Finder (Benson, 1999), RepeatModeler and RepeatMasker (Smit *et al.*, 2013). Repeat libraries from *Mercurialis* (from RepeatModeler), Euphorbiacae and *Vitis vinifera* were used for masking the genome before further analyses.

Summary of SNP calling pipelines

We called SNPs using four different pipelines. (1) X- and Y-specific SNPs were called and haplotypes with premature stop codons and allele-specific expression SNPs were inferred on the basis of RNAseq data from the two largest families by SEX-DETector. (2) SNPs were called from genome-capture data of multiple populations, for independent analysis of genomic regions of differing sex linkage, as inferred from the genetic map. (3) SNPs were called on all individuals from five families, to perform genetic mapping. (4) SNPs were called on transcripts from the three unrelated males (M1, M2 and M3) and the three unrelated females (G1, G2 and G3) for comparisons between ORFs of differing sex linkage and comparisons of pairwise diversity between species. The SNPs for which there was evidence of separate X and Y copies and also had homologues in *M. huetii* and *R. communis* were analyzed in a phylogenetic framework to estimate the (tree-based) evolution rate of the Y sequences. Specific details of each pipeline are given in the corresponding sections below.

Genetic map construction

We called SNPs independently of the SEX-DETector pipeline, using all individuals available over the three generations of our family crosses. We first mapped RNAseq reads to the reference transcriptome using Bowtie2 v2.3.1 (Langmead and Salzberg, 2012). The resulting sam files were converted to bam, sorted and converted to mpileup format using samtools v 1.3 (Li *et al.*, 2009). The resulting posterior file containing segregation information of transcripts in the five mapping families (>164,000 markers and their genotype likelihoods) was passed through the LepMap3 (LM3) pipeline (Rastas, 2017). First, we calculated the relatedness between individuals using the IBD module (of LM3) using a random subset of 3,000

markers. Three individuals were discarded because their relatedness to their putative father (M6) was < 0.2, i.e., they were likely the result of a different cross through contamination. The parental genotypes were then called using ParentCall2, with parameter halfSibs = 1, to take into account the half-sib family structure, in addition to the genotypic information of offspring, parents and grandparents. The 158,000 remaining markers were further filtered using the Filtering2 module, with parameter dataTolerance=0.001, to remove markers with distorted segregating. We then combined markers of each transcript into pseudo-markers, as they were not sufficiently informative on their own to separate into linkage groups. This was done by running OrderMarkers2 separately for each transcript without recombination (parameter values: recombination1 = 0, recombination2 = 0), and then obtaining genotype likelihoods for each transcript with parameter outputPhasedData = 4. This resulted in 13,261 markers, each consisting of a single transcript, which we re-formatted as input for LM3. We defined linkage groups by running SeparateChromosomes2 with default parameters (lodLimit = 10), adding additional markers to the resulting eight linkage groups using JoinSingles2 with lodLimit = 8 and lodDifference = 2. We were able to assign 10,076 transcripts into eight linkage groups, which were ordered using OrderMarkers2. The linkage maps were drawn with R/qtl v1.41 (Broman *et al.*, 2003) running in R 3.4.2 (R *et al.*, 2007).

Sequence divergence in *M. annua*, and in comparison with closely related species

We aligned reads from six unrelated individuals from France (M1, M2, M3, G1, G2, G3) to the 39,302 open reading frames from the reference transcriptome assembled via the SEX-DETector pipeline, based on the M1 male transcriptome and using the BWA-MEM algorithm of BWA v0.7.13 (Li and Durbin, 2009). Picard Tools v2.2.1 (http://broadinstitute.github.io/picard/) were used to mark duplicate read pairs. Local realignment around insertions and deletions (indels) was performed with GATK v3.7 (DePristo *et al.*, 2011), followed by SNP calling on each individual using the HaplotypeCaller module in GATK (McKenna *et al.*, 2010). Joint SNP calling was performed using GATK's GenotypeGVCFs module, overlapping SNPs and indels were filtered out with the VariantFiltration module, and SNPs and indels were separated and filtered to produce two high-quality variant sets with the following parameters in the VariantFiltration

module: 'QUAL < 30' 'DP < 30' 'MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1)' 'QD < 5.0'. The high-quality SNP set was used to perform variant quality score recalibration to filter the full SNP set. This set was used to calculate within-species nucleotide diversity (π) using vcftools v 0.1.13 (Danecek *et al.*, 2011).

Pairwise d_N and d_S values from *M. huetii* and *R. communis* were obtained using yn00 from PAML 4.9 (Yang, 2007). This analysis used 4,761 and 2,993 homologues from the *M. huetii* and *R. communis* (http://castorbean.jcvi.org/introduction.shtml) transcriptomes, based on a *de novo* assembly using default settings in Trinity ((Haas *et al.*, 2013); File S5). Sequences were aligned with LASTZ (Harris, 2007). Effective codon usage was calculated using ENCprime (https://jnpopgen.org/software/).

To estimate the rate of Y-sequence evolution, we identified one-to-one orthologs between the *M. annua* sex linked genes and the closely related species *M. huetii* and *R. communis*, through reciprocal best BLASTP, with an e-value of $1e^{-50}$, a culling limit of 1, and considering only transcripts that were not split across contigs. The 98 orthologs identified were aligned with LASTZ (Harris, 2007) and analyzed with PAML 4.9 (Yang, 2007) to obtain tree-based estimates of d_N and d_S values. We employed three models: the 'null model' (M_0), where all parts of the tree are assumed to have the same d_N/d_S values; the 'branch-specific' model (M_1), where each branch may have its own d_N/d_S value; and the 'two-ratio' model (M_2), where the branch of interest (Y) may have a d_N/d_S ratio that differs from the rest of the tree (Nielsen and Yang, 1998). Preliminary analysis revealed similar results for the M_1 and M_2 models (results not shown); we thus chose the M_2 as the simpler model for further analysis. We tested whether the Y branch evolved differently from the rest of the tree, using a likelihood ratio test (LRT) between the M_2 and M_0 models, with one degree of freedom and LRT = 2 x abs(l_2 - l_0), where l_2 and l_0 are the likelihoods for M_2 and M_0 respectively.

Expression analysis

Reads from the RNA libraries of all 30 females and 35 males were pseudoaligned to the reference transcriptome using Kallisto v0.43.1 (Bray *et al.*, 2016). The resulting raw count data were compiled in a

table, and differential expression analysis between male and female samples was conducted with edgeR v3.18.1 (Robinson *et al.*, 2010), running in R v.3.4.0 (R Development Core Team, 2007). Genes were filtered from the analysis if their average $\log_2$ count per million (as computed by edgeR's aveLogCPM function) was negative. This retained genes with an average count of 24 or more per sample. Two individual libraries were removed that were obvious outliers in the MDS plots of the filtered data. Libraries were normalized with the default (TMM) normalization. Dispersion was measured with default parameters using a negative binomial model. Sex-biased transcripts were defined statistically, i.e., only by a false discovery rate (FDR – Benjamini and Hochberg; 1995) of 5%. A statistical definition for sex-biased genes is justified because our RNAseq analysis was based on many samples, and because the sex-biased categories (male-, female- and unbiased) used in our analysis capture biological information in that they differed statistically in various measurements. Our results for effects of sex bias should be conservative, because our approach diluted 'real' sex-biased genes with unbiased genes. We note, however, that the lack of minimum $\log_2$FC in our definition of sex-bias means that we cannot differentiate between allometric differences in tissue composition or gene regulation differences between males and females (Montgomery and Mank, 2016).

SEX-DETector calculates X- and Y-allele expression from each male. X and Y read numbers were summed for each contig and individual separately and divided by the number of X/Y SNPs of the contig and adjusted for the library size of the respective individual. X and Y expression levels were then averaged among individuals, and the ratio of the means was computed. Gene ontology (GO) enrichment analysis was performed on the sex-biased genes, as well as separately for the male- and female-biased genes, using topGO with the weight01 algorithm, which accounts for GO topology (Alexa and Rahnenfuhrer, 2010).

Comparison of expression between sex-linked genes and autosomal genes

We categorized transcripts based on their genomic location into fully sex-linked (SL; transcripts in the non-recombining region of LG1 in males), pseudo-autosomal (PAR; transcripts in the rest of LG1 that showed recombination in both sexes) and autosomal (Au; transcripts that mapped to the remaining seven

linkage groups). We investigated the effects of sex linkage, sex-biased expression and their interactions through linear models on the following metrics, using R version 3.4.2: absolute male/female $\log_2$FC (fold-change, i.e., overall sex-bias intensity); nucleotide diversity (π); d_N/d_S from pairwise comparisons with *M. huetii* and *R. communis*; and Y/X allele-specific $\log_2$FC obtained from SEX-DETector. The data were not normally distributed and were thus Box-Cox-transformed after adding $1e^{-11}$ to all values to allow inclusion of zero values and estimating the transformation parameter using the command boxcoxnc with the Anderson-Darling method in the R package AID v2.3 (Dag *et al.*, 2014). Absolute male/female ($\log_2$FC) and π remained non-normal after transformation when unbiased genes were included in the analysis, because there were too many transcripts with ~ 0 values. Exclusion of these transcripts did not qualitatively affect the results (data not shown). We also analyzed the sex-biased genes on their own within the same statistical framework. This allowed a more similar distribution between the sex-bias categories, and successful tests of normality after Box-Cox transformation. The summary tables from the models were generated using the R package sjPlot version 2.4 (Lüdecke, 2017).

We also compared the proportions of sex-biased/unbiased genes and male-biased/female-biased genes between pairs of genomic regions of differing sex linkage using chisq tests (or Fishers's exact tests when the gene numbers were small), employing the sjt.xtab function of the in R package sjPlot v 2.4 (Lüdecke, 2017). The pairs of genomic regions compared were: PAR-SL; Au-LG1; and Au-SL. Finally, we looked for enrichment of candidate genes in the same genomic region comparisons using the same methodology.

Divergence between males and females from natural populations

To assess divergence between males and females, we examined F_{ST} and sex-biased heterozygosity from individuals sampled from across the species' range. Seeds obtained from multiple *M. annua* populations (Table S2) were germinated and grown in a glasshouse at the University of Lausanne. Twenty individuals of each sex were selected for genomic analysis to achieve a balance between representation of many populations and sufficient common sequences of good quality. High-quality genomic DNA was extracted from about 100 mg of frozen leaf material using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany),

following standard protocols. Based on the successful probes from a previous gene-capture experiment (see González-Martínez *et al.*, 2017), about 21 Mbp of the *M. annua* diploid genome (175,000 120-mer probes) were sequenced, with sequence capture using SureSelect DNA Capture technology (Agilent Technologies, (Santa Clara, CA) followed by sequencing using Illumina HiSeq 2500, outsourced to Rapid Genomics (Gainesville, FL). Our gene-capture experiment included both genic and intergenic regions and avoided repetitive regions and organelle genomes. Raw data was provided by Rapid Genomics as FASTQ files.

Sequencing quality was assessed using FASTQC v.10.5 (www.bioinformatics.babraham.ac.uk/ projects/fastqc). Adapter removal and trimming used Trimmomatic v.0.36 (Bolger *et al.*, 2014), with options ILLUMINACLIP: TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW: 4:15 MINLEN:36. After filtering, we retained an average of 13 million paired-end sequences per sample. We aligned sequences to the *M. annua* scaffolds using BWA-MEM v.0.7.12 (Li and Durbin, 2009) and processed the alignments with Picard Tools v.1.141 (http://broadinstitute.github.io/picard/). We ran Samtools mpileup (Li *et al.*, 2009) with the probabilistic alignment disabled, and called SNPs using Varscan (Koboldt *et al.*, 2009) with a minimum variant allele frequency of 0.15 and a minimum threshold for homozygotes of 0.85. We required a minimum of 10 reads per site and a Phred quality score > 20.

The gene capture ORFs were aligned to the genomic scaffolds using BLAST v.2.2.31 (Altschul *et al.*, 1990). We retained hits with a maximum e-value of e^{-20}, and a minimum identity of 95%. In the case of multiple ORFs mapping to the same genomic location, we selected the longest alignment, and removed all overlapping ORFs. This resulted in 21,203 transcripts aligning to the genome, 8,317 of which were also included in the linkage map. All scaffolds with a transcript were assigned a linkage group position in the map. 31 scaffolds contained transcripts assigned to more than one linkage group; in these cases, we assigned the scaffold to the linkage group with > 50% of the transcripts (in the case of equal assignment, we excluded the scaffold from the analysis). In total, we were able to assign 6,458 scaffolds to linkage-group positions.

We used the BLAST output to construct a bed file that defined the coordinates of each transcript on a genomic scaffold. We then extracted genetic information for the transcripts specified in the bed file. Using this information, we calculated two complementary measures of differentiation between males and females for transcripts exceeding 100 bp. First, we computed F_{ST} for each transcript using VCFtools v.0.1.15 (Danecek *et al.*, 2011). Similarly, we assessed sex-biased heterozygosity for each transcript (Toups *et al.*, 2019), which we defined as the $\log_{10}$ of the male heterozygosity:female heterozygosity, where heterozygosity is measured as the fraction of sites that are heterozygous. We expect this ratio to be zero for autosomal transcripts, and elevated on young sex chromosomes due to excess heterozygosity in males. For transcripts on older sex chromosomes, the X and Y alleles may have diverged sufficiently to prevent Y alleles from mapping to the X allele reference; in this case, we should expect X-chromosome hemizygosity in males and no inflated female heterozygosity.

For both F_{ST} and sex-biased heterozygosity in the two sexes, we assessed differences between the autosomes, PAR, and the sex-linked region using Wilcoxon signed rank tests. We then computed moving averages in sliding windows of 20 transcripts on all chromosomes using the rollmean function, from package zoo v.1.8-0 in R v3.2.2. To identify regions of elevated F_{ST} and sex-biased heterozygosity, we computed 95% confidence intervals on the basis of an average of 30 consecutive transcripts on the autosomes calculated 1,000 times. Finally, for both metrics, we assessed whether the top 5% of transcripts were overrepresented on LG1 using chi-squared tests.

Data Availability

All raw DNA and RNA sequence data generated in this study have been submitted to NCBI under accession SRP098613 BioProject ID PRJNA369310. Scripts, supplementary files and intermediate files are available at https://osf.io/a9wjb/. File S1 contains the genetic map data and other information on each mapped transcript. File S2 contains the Sanger sequence of clones containing the X and Y variants, where there is a premature stop codon on the Y, and the script used to make the phylogenetic tree. File S3 summarizes the GO results for sex-biased genes. File S4 contains transcriptome annotation information.

File S5 contains the assembled transcriptome and X and Y haplotype predictions from SEX-DETector. File S6 contains the genome and predicted genes on it. File S7 is the *M. annua* repeat library.

Results

Male karyotype

The M. annua karyotype has eight linkage groups, with no difference between males and females, i.e., the sex chromosomes are homomorphic (Figures 1 and S1). Idiogram data, chromosomal arm ratios and relative length measurements from 50 images (obtained from 5 individuals; Table S3) indicate that all *M. annua* chromosomes can be discriminated from one another. A partial distinction between the one large, two medium and five smaller chromosomes is possible on the basis of the two common ribosomal DNA probes (45S and 5S), with a secondary constriction often visible at the location of the 45S locus.

M. annua genome assembly

Using k-mer distribution, we estimate the haploid genome size to be 325 Mb (Figure S2, Table S4), which is similar to the previous estimate of 640 Mb for the diploid genome (Obbard *et al.*, 2006). Using a combination of short-read Illumina and long-read Pacific Biosciences DNA sequencing, we generated ~56.3 Gb of sequence data from the *M. annua* male M1, corresponding to a coverage of ~86.6x. After read filtering, genome coverage dropped to ~63.7x (Table S5). *De novo* assembly and scaffolding yielded a final assembly of 89% of the genome (78% without gaps), 65% of which was assembled in scaffolds > 1 kb, with an N_{50} of 12,808 across 74,927 scaffolds (Table S6). Assembly statistics were consistent with those for other species of the Malpighiales (Table S7), as was our estimate of total genomic GC content (34.7%; Smarda *et al.*, 2012). The *M. annua* assembly encompassed over 89% of the assembled transcriptome; most of the unassembled genome sequence data is therefore likely repetitive. Using BUSCO2 (Simão *et al.*, 2015), we recovered 76.1% of the 1,440 genes in the BUSCO embryophyte database (of which 3.9% were duplicated). A total of 10.3% of the genes were fragmented, and 13.6% were missing.

Repeat masking identified simple tandem repeats in over 10% of the assembly. This proportion is likely

to be an underestimate, given the difficulty in assembling microsatellites. We characterized 15% of the assembly on the basis of homology information and DNA transposon and retrotransposon masking, with an additional 33% that corresponds to 1,472 predicted novel transposable elements. The most frequent transposable repeats were *gypsy* LTR, *copia* LTR, and L1 LINE retrotransposons (Table S8), similar to findings for other plant genomes (Chan *et al.*, 2010; Sato *et al.*, 2011; Wang *et al.*, 2012a; Rahman *et al.*, 2013). Across all data, 58% of the ungapped *M. annua* assembly was repetitive (Table S8), corresponding to 44% of the genome. Given that our assembly covered 78% of the genome, and if we assume that the missing fraction comprises only repeats, up to 66% of the *M. annua* genome may be repetitive. This too would be consistent with what is known for other plants, which have a similarly high AT-rich repeat content and a similar number of unclassified repetitive elements (e.g. Chan *et al.*, 2010). The *M. annua* genome appears to be about 240 Mb larger than that of its relative *Ricinus communis* (see Table S7), perhaps reflecting on-going transposon activity.

Sex-linked transcripts and genetic map for diploid *M. annua*

Genetic mapping recovered eight linkage groups (LG1 through LG8), corresponding to the expected chromosome number for the diploid *M. annua* karyotype (Durand, 1963). LepMap3 (LM3) identified 568 sex-linked (SL) transcripts, of which 365 were also identified as sex-linked by SEX-DETector and phased into X and Y haplotypes. The 568 SL transcripts mapped to LG1, which is thus the sex chromosome. An additional 1,209 transcripts mapped to the two ends of LG1, representing two putative pseudoautosomal regions (PAR1 + PAR2; we refer to these as PAR and consider them together in all analyses); 8,299 transcripts mapped to the other (autosomal) linkage groups. In total, 641 markers could be resolved from one another and were thus informative. We inferred the sex-linked region to comprise of transcripts without recombination in males and in the same phase as that of their paternal grandfather. Assuming an equal transcript density across the genome, the sex-linked region represents 32.6% (about a third) of LG1 and 5.8% of the 320 Mb haploid genome (Obbard *et al.*, 2006), or 18.6 Mb. This Y-linked region maps to LG1 at male position 53.85 cM and spans 22.23 cM on the female recombination map (from positions

46.44 - 68.67 cM; Figure 2). The total male and female recombination maps were 700.43 and 716.06 cM, respectively. All marker names and their associated positions and metrics are provided in File S1.

Characterization of contigs by sex linkage and ORF localization

We characterized the genome assembly by mapping all ORFs using GMAP and dividing genomic contigs into three groups: 'sex-linked', 'autosomal', and 'expressed ORF' contigs (containing all contigs that mapped to a transcript, regardless of its inclusion on the genetic map; note that PAR contigs were included in the autosomal bin). There were 548 sex-linked contigs containing 1,641 genes and a total of 8.3 Mb of sequence; 7,579 autosomal contigs containing 23,502 genes and 106 Mb sequence data; and 15,392 contigs with 34,105 expressed ORFs distributed across 176 Mb (Table 1).

Sequence divergence, nucleotide diversity and codon usage of X and Y haplotypes

We used the SNP and haplotype calls from SEX-DETector to compare nucleotide diversity and divergence between the X and Y ORFs within *M. annua*, as well as between X, Y and autosomal ORFs of *M. annua* and closely related species. The mean pairwise d_N/d_S between X and Y haplotypes within *M.* annua was 0.182 (Table 2). For sex-linked transcripts, we excluded the few SNPs that did not show clear sex-linked segregation, but we could not apply such error correction for autosomes. Accordingly, π might be underestimated for sex-linked genes compared to autosomes (see Figure S3, Table 1), so that metrics associated with sex-linked sequences should be regarded as conservative.

The pairwise d_S and d_N/d_S values for all possible orthologs between *M. annua* and its relatives *M. huetti* and *R. communis,* and the X and Y haplotypes, are presented in Table 2 and Figure 3. For autosomal genes, d_N/d_S between *M. annua* and *M. huetii* and between *M. annua* and *R. communis*, respectively, was 0.149 and 0.140. Pairwise d_S was lower between X/Y gene pairs within *M. annua* than between orthologous autosomal genes in *M. annua* and *M. huetii* or *R. communis*.

Codon usage in *M. annua* did not differ significantly between sex-linked, pseudoautosomal and autosomal (SL, PAR and Au) ORFs (Nc = 54, Nc = 53 and Nc = 53, respectively; Figure S4).

We used PAML (Yang, 2007) to analyze the 98 sex-linked genes with orthologs in *M.annua*, *M. huetii*, and *R. communis*, using the 'two-ratio' M_2 model (which accommodates differences in d_N/d_S between the Y branch and other branches of the tree; Nielsen and Yang 1998). Our analysis identified 84 sequences with $0.001 < d_S < 2$ on both X and Y branches. Variation for 74 of these sequences was consistent with a tree topology corresponding to the species tree and recent X-Y divergence for *M. annua*, i.e. trees with topology (((*M. annua*-X, *M. annua*-Y), *M. huetii*), *R. communis*); we focused further analysis on these genes. For 52 of these 74 genes (21 after FDR correction at 5% level; Benjamini and Hochberg, 1995), the M_2 model provided a better fit than the M_0 model. Moreover, the Y lineage had more synonymous and more non-synonymous mutations compared to the X lineage, following divergence from *M. huetii*. Together, these results suggest that 28% of the Y-linked genes in our sample have been evolving faster than their corresponding X-linked orthologs, as inferred from the tree-based d_N and d_S values (Wilcoxon Rank Sum = 392 and 568, respectively; $p < 0.001$ for both tests; Figure 4), as well as under relaxed selection, as inferred from the tree-based d_N/d_S ratio (Wilcoxon Rank Sum = 267, $p < 0.001$; Figure 4). The seven sequences for which the Y variant clustered outside the clade (*M. annua*-X, *M. huetii*) were scattered across the genetic map of the non-recombining region of the Y (Figure 5), i.e. they did not indicate a clear ancestral/common sex-determining region between *M. annua* and *M. huetii*. We did not investigate these genes further.

Identification of a degenerated and duplicated sequence on the Y

The SEX-DETector pipeline detected putative premature stop codons in four Y and three X transcripts, although there was always a functional X haplotype for the same transcript. We confirmed one of the four Y.-haplotype stop codons using PCR, through cloning and sequencing from five males and nine females sampled across the geographic range of diploid *M. annua*, which only showed the stop codons on sequences from males. Briefly, a phylogenetic tree was constructed for this gene (Figure S5) using RAxML v.8.2.10 using 1,000 bootstraps (Stamatakis, 2014) on sequences aligned with mafft v7.310 (Katoh and Standley, 2013). Interestingly, two different Y-linked sequences could be obtained, sometimes from the same male, suggesting that the sequence has been duplicated in at least some males. All males and females

analyzed also contained a functional copy that had 89% identity with an F-box/kelch repeat protein in *Ricinus communis*. The transcript shows a lower Y/X expression ratio and a high π (marked "K_s" in Figure 5), suggesting that it represents a degenerated copy on the Y.

Identification and annotation of sex-biased genes

There were many more (1,385) male-biased than female-biased genes (325), based on 5% FDR. This difference increased when bias was determined on the basis of a minimum $\log_2$ fold-change ($\log_2$FC) threshold of one (1,141 vs. 140; Figure S6). Male-biased genes were enriched in biological functions such as anther-wall tapetum development, response to auxin, response to ethylene, cell-tip growth, pollen-tube growth, and floral-organ senescence, whereas female-biased genes were enriched for functions related to the maintenance of inflorescence meristem identity, jasmonic acid and ethylene-dependent systemic resistance, regulation of innate immune response, and seed maturation (File S3).

Identification of the ancestral evolutionary stratum on the Y

To identify a putative ancestral stratum in the non-recombining region, we investigated variation in the following metrics across the female recombination map (Figure 5): the magnitude of sex-biased expression; the nucleotide diversity (π) of each transcript; and the pairwise d_N/d_S ratio based on comparison with homologous genes in the closely related dioecious sister species *M. huetti*, and its monoecious more distant relative *R. communis* (for which we could identify 4,761 and 2,993 homologues, respectively); and the expression ratio of X and Y haplotypes inferred from SEX-DETector (Muyle *et al.,* 2016). We annotated Figure 5 with letters to indicate the position of candidate sex-determining genes, based on their blast hit descriptions. The sole obvious outlier for any of the metrics above was a transcript associated with auxin production, which had a high pairwise d_N/d_S value in the comparison with the close outgroup species, *M. huetii*.

Comparison of sex-linked, autosomal and pseudo-autosomal ORFs

We examined the effects of sex-bias (male-, female-, and unbiased genes) and genomic location (SL, PAR or Au), as well as their interactions, on the five metrics plotted in Figure 5. We repeated the analyses

after excluding unbiased transcripts for the four metrics with sufficient gene numbers (sex bias log_2FC, π, and pairise d_N/d_S using either *M. huetii* or *R. communis*) to directly compare male- and female-biased transcripts. The full models are presented in Tables S9-S11. Below, we summarize their significant results; Figure 6 shows the data distributions. The analyses revealed significantly higher differences in gene expression between the sexes (absolute log_2FC) in the sex-linked region ($p = 0.037$; Table S9). These differences were largely due to significantly higher expression of male- than female-biased transcripts(male-bias*SL interaction: $p=0.036$; Table S10).

The increased magnitude of male-biased gene expression in fully sex-linked genes is consistent with a recent masculinization as sex-linked genes had slightly lower nucleotide diversity (π) than those in the PAR or autosomal genes ($p = 0.039$; Table S9), but male-biased genes were not affected as much ($p = 0.017$; Table S9). The pairwise d_N/d_S between *M. annua* and *M. huetii* was higher for sex-linked genes ($p = 0.017$; Table S9), as well as for female-biased genes localizing to the PAR (PAR*female-bias interaction: $p = 0.005$; Table S9). Sex-biased genes in the *M. annua* PAR had a lower π (PAR: $p = 0.007$; Table S10), and the d_N/d_S for sex-biased PAR genes between the *M. annua* and *M. huetii* was also lower ($p = 0.018$; Table S10). d_N/d_S with respect to the more distantly related *R. communis* was only higher for male-biased genes ($p = 0.046$; Table S9), though power was limited by the availability of fewer identified homologues (4,761 in *M. huetti* compared to 2,993 in *R. communis*). The Y-linked copy of female-biased genes had significantly lower expression than its X-linked counterpart ($p = 0.035$; Table S11).

Finally, we used chi-squared tests to compare the relative proportions of sex-biased versus unbiased and male- versus female-biased genes in different pairs of genomic regions (PAR-SL, Au-LG1, Au-SL; Table 3). The sex-linked region had significantly more sex-biased genes (compared to either PAR or Au regions). This enrichment was largely responsible for the significant sex-biased gene enrichment compared with the entire LG1 and the autosomes, and was largely due to more female-biased genes on LG1. The sex chromosomes thus appear to be feminized in terms of number of female-biased genes, in contrast to the genome as a whole, which was masculinized in terms of number of male-biased gene numbers, and in

contrast to absolute expression, which was higher for male-biased genes. We did not detect any enrichment for candidate genes with functions with obvious potential sexually antagonistic effects in the sex-linked region (Table 3).

Sex-linked sequence analysis of individuals sampled across the species range

To confirm the lack of recombination in the sex-linked region, we estimated F_{ST} and sex-biased heterozygosity for genome capture data from 20 males and 20 females sampled from natural populations across the species' range (Table S2), based on 6,557 transcripts assigned to the linkage map. The sex-linked region showed clear differentiation between males and females using both metrics. F_{ST} between males and females was elevated in the sex-linked region relative to the PAR and the autosomes (Wilcoxon, $p = 4.79\times10^{-5}$, $p = 1.099\times10^{-7}$ respectively; Figure 7A), although the F_{ST} values were very small. Similarly, males were more heterozygous than females in the sex-linked region relative to the PAR and the autosomes (Wilcoxon, $p<2.2\times10^{-16}$ in both; Figure 7C). A sliding window analyses of F_{ST} and sex-biased heterozygosity identified a region of elevated F_{ST} at ~ 65 cM of LG1 in the female recombination map (Figure 7B), a value exceeding that for any autosomal region. We also identified a region between 50 and 65 cM on LG1 (Figure 7D) with higher heterozygosity in males than females (Figure 7E). Finally, both F_{ST} and male-biased heterozygosity were greatest in the sex-linked region of LG1 (Table S12).

Discussion

A genome assembly, annotation and genetic map for dioecious *M. annua*

Our study provides a draft assembly and annotation of diploid dioecious *M. annua* (2n = 16). Cytogenetic analysis has confirmed the haploid number of n = 8 chromosomes for diploid *M. annua* (Durand, 1963), with about 10,000 ORFs assigned to the corresponding eight linkage groups by genetic mapping. Similar to other plants, about 2/3 of the *M. annua* genome comprises repetitive sequences, mostly *gypsy* LTR, *copia* LTR, and L1 LINE retrotransposons (Chan *et al.*, 2010; Sato *et al.*, 2011; Wang *et al.*, 2012a; Rahman *et al.*, 2013). The number of genes in *M. annua* (34,105) is similar to that of diploid species such as *Arabidopsis thaliana* (> 27,000 genes) and *Ricinus communis* (> 31,000 genes) and is thus compatible with having a long recent history under diploidy rather than polyploidy; for comparison, the gene content of diploid *Gossypium* is > 40,000 genes, pointing to a polyploid history (Wang *et al.*, 2012a). The draft annotated genome of *M. annua* will be useful for understanding potential further links between sex determination and sexual dimorphism, as well as the consequences of genome duplication and hybridization (Obbard *et al.*, 2006) for sexual-system transitions in polyploid lineage of the species complex (Pannell *et al.*, 2008).

The size of the non-recombining Y-linked region of *M. annua*

Patterns of SNP segregation in small families from 568 ORFs suggests that a substantial length (1/3) of the *M. annua* Y chromosome has ceased recombining with the X. Lack of recombination on the sex-linked region is further supported by the existence of a Y-specific non-functional gene duplication, which could be amplified in populations ranging from Israel to Britain, as well as additional population-genetic data from males and females sampled from across the diploid species' range. The sex-linked region had slightly, but significantly, higher F_{ST}, a higher male/female SNP ratio, and higher heterozygote frequency in males than females, indicating a history of low recombination. Previous mapping of bacterial artificial chromosomes (BACs) found male-specific PCR products distributed over a region of between 52 and 66.82 cM of the sex-linked region, corresponding to 4.86% of the genome and implying a physical length

of between 14.5 and 19 Mb for the non-recombining genome (Veltsos *et al.*, 2018). The large non-recombining region of the *M. annua* Y chromosome contrasts with that of other plants with homomorphic sex chromosomes, which typically have < 1% of their Y not recombining, e.g., *Vitis vinifera* (Fechter *et al.*, 2012; Picq *et al.*, 2014), *Fragaria chiloensis* (Tennessen *et al.*, 2016) and *Populus* species (Paolucci *et al.*, 2010; Geraldes *et al.*, 2015).

Limited gene loss, pseudogenization, and mild purifying selection on the *M. annua* Y

We found only limited evidence for Y-chromosome degeneration within the non-recombining region of the *M. annua* Y chromosome, despite its substantial length. Very few genes have been lost or have become non-functional, and only one of the 528 X-linked genes corresponding to the non-recombining region of the Y (0.2%) did not have a Y-linked homologue. Our analysis may have overlooked genes that have been lost from the Y, because their detection from RNAseq data relies on the presence of polymorphisms on the X copy (see Blavet *et al.*, 2015) and requires their detectable expression in our sampled tissues. Nevertheless, the absence of evidence for substantial gene loss is consistent with other signatures of only mild Y-chromosome degeneration in *M. annua* and is likely reliable.

There was only slightly lower relative expression from the Y compared to the X allele. Allele-specific expression was inferred on the basis of mapping to the reference transcriptome, and we expect only limited mapping bias against expressed Y-specific genes. The fact that different levels of Y/X expression depended on sex-bias (it was reduced for female-biased genes) further suggests that the result is probably not an artifact of the analysis. We found some evidence for relaxed purifying selection on Y alleles compared to X alleles, as expected in genomic regions experiencing limited recombination (Charlesworth and Charlesworth, 2000). Specifically, pairwise comparisons of X and Y alleles (after excluding the few with in-frame stop codons) showed higher d_N/d_S than pairwise comparisons of *M. annua* orthologs with *M. huetii* or *R. communis*. Moreover, our tree-based analysis of common orthologs in all species indicated that 28% of the identified Y-linked alleles have experienced an accelerated rate of molecular evolution in the *M.*

annua lineage since its divergence from its sister species *M. huetii*, which is also dioecious. However, we found no evidence for differential codon usage bias between the Y-linked and other sequences. Codon usage bias is expected to be lower for non-recombining regions of the genome experiencing weaker purifying selection (Hill and Robertson, 1966), and there appears to have been a shift towards less preferred codon usage in *Rumex hastatulus* Y-linked genes, increasing in severity with time since the putative cessation of recombination between X and Y chromosomes (Hough *et al.*, 2014). Although our analysis involves multiple statistical tests of significance such that some results might represent Type 1 error, taken together they point to a very recent cessation of recombination for much of the Y chromosome. In contrast, up to 28% of *Rumex hastatulus* Y-linked genes have been lost (Hough *et al.*, 2014) and about 50% of the *Silene latifolia* Y-linked genes are dysfunctional (Papadopulos *et al.*, 2015; Krasovec *et al.*, 2018).

Sex-biased gene expression in *M. annua*

We found substantial differences in gene expression between males and females of *M. annua*, consistent with the observation of sexual dimorphism for a number of morphological, life-history and defense traits in *M. annua* (Hesse and Pannell, 2011; Labouche and Pannell, 2016). There were at least four times more male-biased than female-biased genes; this is probably conservative because we did not use a minimum log_2FC threshold in our analysis, which would have further increased the proportion of male-biased genes. The observed slight decrease in expression of Y- compared with X-linked alleles suggests that dosage compensation could have begun evolving. However, any such dosage compensation would likely be at a very early stage, not least because YY males (which lack an X chromosome) are viable (Li et al; Kuhn, 1939). Interestingly, sex-linkage sometimes influenced male- and female-biased genes differently, with male-biased genes having a higher nucleotide diversity (π) than female-biased genes when sex-linked, again perhaps reflecting a history of relaxed purifying selection. Male- and female-biased genes also differed in pairwise d_N/d_S relative to *M. huetii*, which was higher for female-biased genes, contrary to expectations under a hypothesis of faster male evolution (Wu and Davis, 1993).

Evidence for evolutionary strata on the *M. annua* Y?

We were unable to find any direct evidence for evolutionary strata with different levels of divergence on the *M. annua* Y chromosome, but phylogenetic comparisons point to the likely existence of at least two such strata on the *M. annua* Y. Dioecy evolved in a common ancestor to *M. annua* and *M. huetii* (Krähenbühl *et al.*, 2002; Obbard *et al.*, 2006), and both species share the same Y chromosome and sex-determining region, as revealed by crosses (Russell and Pannell, 2015) and the possession of common male-specific PCR markers (Veltsos *et al.*, 2018). The phylogeny of sex-linked genes within the shared sex-determining region should have the topology ((*M. annua*-Y, *M. huetii*), *M. annua*-X). In contrast, our PAML analysis clearly indicates that the phylogeny of most genes in the non-recombining region of the *M. annua* Y have the topology ((*M. annua*-Y, *M. annua*-X), *M. huetii*), indicating divergence between the X and Y copies of these genes in *M. annua* more recently than the species split between *M. annua* and *M. huetii*. These results suggest that *M. annua* and *M. huetii* share an old (and possibly small) non-recombining stratum that includes the sex-determining locus, and that a larger stratum has been added to it in the *M. annua* lineage. Another possible interpretation is gene conversion between X and Y copies within *M. annua*, as has been reported for humans and some plants (Rautenberg *et al.*, 2008; Wu *et al.*, 2010; Trombetta *et al.*, 2014). This would lead to an underestimate of the size of the ancestral stratum and may have erased evidence of a longer common shared history between the non-recombining regions of *M. annua* and *M. huetii*.

Although most Y-linked genes that we sampled in *M. annua* point to an X/Y divergence that postdates the species' split from *M. huetii*, we found seven genes with a phylogeny consistent with the hypothesized old stratum shared with *M. huetii*. If these genes are indeed within an older stratum, we might expect them to colocalize in the same part of the non-recombining region. However, we found that they were quite scattered on the female recombination map. We have no explanation for this pattern, but we note that it would be consistent with a break of synteny between the *M. annua* Y and X chromosomes (and thus potentially also between the *M. huetii* and *M. annua* Y chromosomes) in the region corresponding to the non-recombining region.

Patterns of divergence in terms of tree-based d_S and d_N between the X and Y chromosomes of *M. annua* and their corresponding orthologs in *M. huetii* and *R. communis* were largely consistent with the expected phylogenetic relationships between these species (Krähenbühl *et al.*, 2002; Obbard *et al.*, 2006). The increased d_S for the Y branch is consistent with a higher mutation rate in male meiosis. Nevertheless, the d_S branch length for *R. communis* was lower than expected for a simple molecular clock across the phylogeny. The recent transition to an annual life cycle in the clade that includes *M. annua* and *M. huetii* may be partially responsible for the larger number of mutations that have accumulated along their branches compared to the perennial *R. communis* (Smith and Donoghue, 2008).

Age of the non-recombining Y-linked region of *M. annua*

The age of the *M. annua* non-recombining region can be estimated on the basis of inferred mutation rates in plants. Ossowski *et al.* (2010) estimated the mutation rate for *Arabidopsis thaliana* to be between 7×10^{-9} to 2.2×10^{-8} per nucleotide per generation. More recently, Krasovec *et al.* (2018) estimated the mutation rate for *S. latifolia* to be 7.31×10^{-9}. The similarity of these estimates suggests that they might be broadly applicable. Accepting a mutation rate for plants of approximately 7.5×10^{-9} and using the synonymous site divergence between X- and Y-linked sequences of *M. annua* of 0.011 (Table 2), we infer that recombination between most of the X- and Y-linked genes in *M. annua* may have ceased as recently as 1.5 million generations ago (more recently if we adopt the upper end of the mutation-rate range estimated by Ossowski *et al.* (2010)). Given that *M. annua* is an annual plant, the putative second non-recombining stratum of the Y chromosome of diploid *M. annua* could be less than a million years old. Krasovec *et al.* (2018) used their mutation-rate estimate to infer an age of 11 million years for the oldest sex-linked stratum of *S. latifolia*, i.e., an order of magnitude older than the putatively expanded sex-linked region of *M. annua*. The few estimates for the time since recombination suppression in other plants with homomorphic sex chromosomes, and those with smaller non-recombining regions, range from 15 to 31.4 million years ago (reviewed in Charlesworth, 2016). *Mercurialis annua* thus appears to have particularly young sex chromosomes.

Concluding remarks

We conclude by speculating on a possible role for sexually antagonistic selection in favoring suppressed recombination on the *M. annua* Y. The canonical model for sex-chromosome evolution supposes that suppressed recombination originally evolves in response to selection to bring sexually antagonistic loci into linkage with the sex-determining locus, and that the Y chromosome begins to degenerate after recombination has ceased (Rice, 1987). Given this model, we might expect to see signatures of sexual antagonism in the sex-linked region before the onset of substantial Y-chromosome degeneration. While we do not observe obvious signatures of sexually antagonistic selection, some of the observed patterns in the sex-linked region are difficult to interpret as the outcome of degeneration. For example, while the Y-linked alleles showed slightly lower expression than X-linked alleles, this was not true for male-biased genes. One possible explanation is that the female-biased Y-linked alleles in males might have been driven to lower expression by sexually antagonistic selection. If female-biased expression is beneficial in females (Parisi *et al.*, 2003; Connallon and Clark, 2010), the observed enrichment of female-biased genes on the X would be consistent with sexually antagonistic selection because X-linked alleles occur twice as often in females than in males. Similarly, the magnitude of sex-biased gene expression (absolute $\log_2$FC) was higher in the sex-linked region, as might be expected if levels of gene expression were the result of an escalation of antagonistic gene expression between the sexes. The possibility that sexually antagonistic selection played a role in the evolution of suppressed recombination might be investigated further by seeking sex-linked QTL for sexually antagonistic traits (Delph *et al.*, 2010), or functional analysis of candidate genes. It would also be valuable to compare expression levels of sex-linked genes with those in a related species with ancestral levels of gene expression, such as orthologs in *M. huetii* that still lie outside the non-recombining region of the Y.

Acknowledgements

We thank John Russell for setting up the crosses, Jos Käfer for help with the Galaxy workflow for the SEX-DETector analysis, Jerome Goudet for statistical advice, and Guillaume Cossard, Wen-Juan Ma, Deborah Charlesworth, Stephen Wright and anonymous reviewers for valuable comments on the manuscript. KER was supported by grants 31003A_163384 and 31003A_141052 to JRP from the Swiss National Science Foundation, and by the University of Lausanne. PV was supported by Sinergia grant 26073998 from the Swiss National Science Foundation to JRP, Nicolas Perrin and Mark Kirkpatrick. JRP and DAF acknowledge support from the NERC and BBSRC, UK, which funded the early stages of this project. The computations were performed at the server at the department of Plant Sciences, University of Oxford, the Vital-IT Center for high-performance computing of the SIB Swiss Institute of Bioinformatics (http://www.vital-it.ch), and the Texas Advanced Computing Center. AM and GABM acknowledge support from ANR (grant number ANR-14-CE19-0021-01). VH, RH, BV were supported by the Czech Science Foundation (grant 18-06147S). MAT was supported by the National Institute of Health grant R01GM116853 (to Mark Kirkpatrick) and the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement number 715257). Conceived the project: JRP, DAF; wet lab work: DAF, SCGM; cytogenetics work: VH, RH, PV; genome assembly and annotation: KER; gene expression analysis: KER, OE, PV; segregation analysis and mapping: PV, AM, PR, GM; population genetic and divergence analysis: KER, PV; analysis of natural populations: SCGM, MAT: wrote the paper: JRP, KER, PV; commented and contributed to the final manuscript: all authors.

References

AKAGI, T., I. M. HENRY, R. TAO, and L. COMAI, 2014 A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. Science **346**: 646-650.

ALBRITTON, S. E., A. L. KRANZ, P. RAO, M. KRAMER, C. DIETERICH *et al.*, 2014 Sex-biased gene expression and evolution of the x chromosome in nematodes. Genetics **197**: 865-883.

ALEXA, A., and J. RAHNENFUHRER, 2010 topGO: enrichment analysis for gene ontology. R package version **2**:

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN, 1990 Basic local alignment search tool. J Mol Biol **215**: 403-410.

AU, K. F., J. G. UNDERWOOD, L. LEE, and W. H. WONG, 2012 Improving PacBio long read accuracy by short read alignment. PloS one **7**: e46679.

BAKER, D. A., L. A. MEADOWS, J. WANG, J. A. DOW, and S. RUSSELL, 2007 Variable sexually dimorphic gene expression in laboratory strains of *Drosophila melanogaster*. BMC Genomics **8**: 454.

BENJAMINI, Y., and Y. HOCHBERG, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B **57**: 289-300.

BENSON, G., 1999 Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res **27**: 573-580.

BERGERO, R., and D. CHARLESWORTH, 2009 The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol **24**: 94-102.

BERGERO, R., A. FORREST, E. KAMAU, and D. CHARLESWORTH, 2007 Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: Evidence from new sex-linked genes. Genetics **175**: 1945-1954.

BERGERO, R., J. GARDNER, B. BADER, L. YONG, and D. CHARLESWORTH, 2019 Exaggerated heterochiasmy in a fish with sex-linked male coloration polymorphisms. Proceedings of the National Academy of Sciences

Proc Natl Acad Sci USA **116**: 6924.

BLAVET, N., H. BLAVET, A. MUYLE, J. KÄFER, R. CEGAN *et al.*, 2015 Identifying new sex-linked genes through BAC sequencing in the dioecious plant *Silene latifolia*. BMC Genomics **16**: 546.

BOETZER, M., C. V. HENKEL, H. J. JANSEN, D. BUTLER, and W. PIROVANO, 2011 Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**: 587-579.

BOLGER, A. M., M. LOHSE, and B. USADEL, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114-2120.

BRAY, N. L., H. PIMENTEL, P. MELSTED, and L. PACHTER, 2016 Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol **34**: 525-527.

BROMAN, K. W., H. WU, S. SEN, and G. A. CHURCHILL, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19**: 889-890.

CHAN, A. P., J. CRABTREE, Q. ZHAO, H. LORENZI, J. ORVIS *et al.*, 2010 Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol **28**: 951-956.

CHARLESWORTH, B., 1991 The evolution of sex chromosomes. Science **251**: 1030-1033.

CHARLESWORTH, B., and D. CHARLESWORTH, 2000 The degeneration of Y chromosomes. Philosophical Transactions of the Royal Society B: Biological Sciences **355**: 1563.

CHARLESWORTH, B., and D. CHARLESWORTH, 1978 A model for the evolution of dioecy and gynodioecy. Am Nat **112**: 975-997.

CHARLESWORTH, D., 2015 Plant contributions to our understanding of sex chromosome evolution. New Phytol **208**: 52-65.

CHARLESWORTH, D., 2016 Plant sex chromosomes. Annu Rev Plant Biol **67**: 397-420.

CHARLESWORTH, D., B. CHARLESWORTH, and G. A. MARAIS, 2005 Steps in the evolution of heteromorphic sex chromosomes. Heredity **95**: 118.

CONESA, A., S. GOTZ, J. M. GARCIA-GOMEZ, J. TEROL, M. TALON *et al.*, 2005 Blast2GO: a universal tool for annotatioin, visualisatioin and analysis of functional genomics research. Bioinformatics **21**:

CONNALLON, T., and A. G. CLARK, 2010 Sex linkage, sex-specific selection, and the role of recombination in the evolution of sexually dimorphic gene expression. Evolution **64**: 3417-3442.

DAG, O., O. ASAR, and O. ILK, 2014 A Methodology to Implement Box-Cox Transformation When No Covariate is Available. Communications in Statistics - Simulation and Computation **43**: 1740-1759.

DANECEK, P., A. AUTON, G. ABECASIS, C. A. ALBERS, E. BANKS *et al.*, 2011 The variant call format and VCFtools. Bioinformatics **27**: 2156-2158.

DAWSON, T. E., and M. A. GEBER, 1999 Sexual dimorphism in physiology and morphology, pp. 175-215 in *Gender and sexual dimorphism in flowering plants*, edited by Geber, M. A., T. E. Dawson, and L. F. Delph. Springer, Heidelberg.

DELPH, D. F., 1999 Sexual dimorphism in life history, pp. 149-179 in *Gender and sexual dimorphism in flowering plants*, edited by Geber, M. A., T. E. Dawson, and L. F. Delph. Springer, Heidelberg.

DELPH, L. F., A. M. ARNTZ, C. SCOTTI-SAINTAGNE, and I. SCOTTI, 2010 The genomic architecture of sexual dimorphism in the dioecious plant *Silene latifolia*. Evolution **64**: 2873-2886.

DEPRISTO, M. A., E. BANKS, R. POPLIN, K. V. GARIMELLA, J. R. MAGUIRE *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet **43**: 491-498.

DURAND, B., 1963 Le complèxe Mercurialis annua L. s.l.: Une étude biosystématique. Annales des Sciences Naturelles, Botanique, Paris **12**: 579-736.

DURAND, B., and R. DURAND, 1991 Sex determination and reproductive organ differentiation in *Mercurialis*. Plant science **80**: 49-66.

DURAND, B., J. P. LOUIS, S. HAMDI, E. CABRE, L. X. YU *et al.*, 1987 Major regulator genes, phytohormone levels and specific gene-expression for reproductive organogenesis in *Mercurialis annua* L. (2n=16). J Cell Biochem 18-20.

ECKHART, V. M., 1999 Sexual dimorphism in flowers and inflorescences, pp. 123-148 in *Gender and sexual dimorphism in flowering plants*, edited by Geber, M. A., T. E. Dawson, and L. F. Delph. Springer, Heidelberg.

ENGLISH, A. C., S. RICHARDS, Y. HAN, M. WANG, V. VEE *et al.*, 2012 Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One **7**: e47768.

FECHTER, I., L. HAUSMANN, M. DAUM, T. R. SÖRENSEN, P. VIEHÖVER *et al.*, 2012 Candidate genes within a 143 kb region of the flower sex locus in *Vitis*. Molecular Genetics and Genomics **287**: 247-259.

GERALDES, A., C. A. HEFER, A. CAPRON, N. KOLOSOVA, F. MARTINEZ-NUÑEZ *et al.*, 2015 Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Mol Ecol **24**: 3243-3256.

GIBSON, J. R., A. K. CHIPPINDALE, and A. M. RICE, 2002 The X chromosome is a hot spot for sexually antagonistic fitness variation. Proc R Soc Lond, B **269**: 499-505.

GONZÁLEZ-MARTÍNEZ, S. C., K. RIDOUT, and J. R. PANNELL, 2017 Range expansion compromises adaptive evolution in an outcrossing plant. Curr Biol **27**: 1-8.

HAAS, B. J., A. PAPANICOLAOU, M. YASSOUR, M. GRABHERR, P. D. BLOOD *et al.*, 2013 *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc **8**: 1494-1512.

HARRIS, R. S., 2007 Improved pairwise alignment of genomic DNA. The Pennsylvania State University **PhD**:

HESSE, E., and J. PANNELL, 2011 Sexual dimorphism in a dioecious population of the wind-pollinated herb *Mercurialis annua*: the interactive effects of resource availability and competition.

HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet Res **8**: 269-294.

HOBZA, R., R. CEGAN, W. JESIONEK, E. KEJNOVSKY, B. VYSKOT *et al.*, 2017 Impact of repetitive elements on the Y chromosome formation in plants. Genes (Basel) **8**:

HOROVITZ, S., and H. JIMÉNEZ, 1967 Cruzamientos interespecíficos e intergenéricos en Caricaceas y sus implicaciones fitotécnicas. Agronomia Tropical **17**: 323-343.

HOUGH, J., J. D. HOLLISTER, W. WANG, S. C. H. BARRET, and S. I. WRIGHT, 2014 Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. Proc Natl Acad Sci USA **111**: 7713-7718.

HUANG, X., and A. MADAN, 1999 CAP3: A DNA sequence assembly program. Genome Res **9**: 868-877.

HUNT, M., T. KIKUCHI, M. SANDERS, C. NEWBOLD, M. BERRIMAN *et al.*, 2013 REAPR: a universal tool for genome assembly evaluation. Genome Biol **14**: R47.

KATOH, K., and D. M. STANDLEY, 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **30**: 772-780.

KAZAMA, Y., K. ISHII, W. AONUMA, T. IKEDA, H. KAWAMOTO *et al.*, 2016 A new physical mapping approach refines the sex-determining gene positions on the *Silene latifolia* Y-chromosome. Sci Rep **6**: 18917.

KEARSE, M., R. MOIR, A. WILSON, S. STONES-HAVAS, M. CHEUNG *et al.*, 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28**: 1647-1649.

KHADKA, D. K., A. NEJIDAT, M. TAL, and A. GOLAN-GOLDHIRSH, 2002 DNA markers for sex: Molecular evidence for gender dimorphism in dioecious *Mercurialis annua* L. Mol Breed **9**: 251-257.

KHADKA, D. K., A. NEJIDAT, M. TAL, and A. GOLAN-GOLDHIRSH, 2005 Molecular characterization of a gender-linked DNA marker and a related gene in *Mercurialis annua* L. Planta **222**: 1063-1070.

KOBOLDT, D. C., K. CHEN, T. WYLIE, D. E. LARSON, M. D. MCLELLAN *et al.*, 2009 VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics **25**: 2283-2285.

KRÄHENBÜHL, M., Y. M. YUAN, and P. KÜPFER, 2002 Chromosome and breeding system evolution of the genus *Mercurialis* (Euphorbiaceae): implications of ITS molecular phylogeny. Plant Syst Evol **234**: 155-169.

KRASOVEC, M., M. CHESTER, K. RIDOUT, and D. A. FILATOV, 2018 The mutation rate and the age of the sex chromosomes in *Silene latifolia*. Curr Biol **28**: 1832-1838.

KUHN, E., 1939 Selbstbestäubungen subdiöcischer Blütenpflanzen, ein neuer Beweis für die genetische Theorie der Geschlechtsbestimmung. Planta **30**: 457-470.

LABOUCHE, A. M., and J. R. PANNELL, 2016 A test of the size-constraint hypothesis for a limit to sexual dimorphism in plants. Oecologia **181**: 873-884.

LAHN, B. T., and D. C. PAGE, 1999 Four evolutionary strata on the human X chromosome. Science **286**: 964-957.

LANGMEAD, B., and S. L. SALZBERG, 2012 Fast gapped-read alignment with Bowtie 2. Nature Methods **9**: 357-359.

LEDER, E. H., J. M. CANO, T. LEINONEN, R. B. O'HARA, M. NIKINMAA *et al.*, 2010 Female-biased expression on the X chromosome as a key step in sex chromosome evolution in threespine sticklebacks. Mol Biol Evol **27**: 1495-1503.

LI, H., and R. DURBIN, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**: 1754-1760.

LI, H., B. HANDSAKER, A. WYSOKER, T. FENNELL, J. RUAN *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**: 2078-2079.

LI, X., P. VELTSOS, G. COSSARD, J. GERCHEN, and J. R. PANNELL, YY males of the dioecious plant *Mercurialis annua* are fully viable but produce largely infertile pollen. (in review)

LIU, Z., P. H. MOORE, H. MA, C. M. ACKERMAN, M. RAGLBA *et al.*, 2004 A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature **427**: 348-352.

LOEPTIEN, H., 1979 Identification of the sex chromosome pair in asparagus (*Asparagus officinalis L.*). Zeitschrift fur Pflanzenzüchtung **82**: 162-173.

LÜDECKE, D., 2017 sjPlot: Data visualization for statistics in social science. R package version 2.4.0, https://CRAN.R-project.org/package=sjPlot.

LUO, R., B. LIU, Y. XIE, Z. LI, W. HUANG *et al.*, 2012 SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. GigaScience **1**: 18.

MANK, J. E., 2013 Sex chromosome dosage compensation: definitely not for everyone. Trends Genet **29**: 677-683.

MANK, J. E., and H. ELLEGREN, 2009 Are sex-biased genes more dispensable. Biol Lett **5**: 409-412.

MARÇAIS, G., and C. KINGSFORD, 2011 A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics **27**: 764-770.

MCKENNA, A., M. HANNA, E. BANKS, A. SIVACHENKO, K. CIBULSKIS *et al.*, 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20**: 1297-1303.

MEISEL, R. P., J. H. MALONE, and A. G. CLARK, 2012 Disentangling the relationship between sex-biased gene expression and X-linkage. Genome Res **22**: 1255-1265.

MOHANTY, J. N., S. NAYAK, S. JHA, and R. K. JOSHI, 2017 Transcriptome profiling of the floral buds and discovery of genes related to sex-differentiation in the dioecious cucurbit *Coccinia grandis* (L.) Voigt. Gene **626**: 395-406.

MONTGOMERY, S. H., and J. E. MANK, 2016 Inferring regulatory change from gene expression: the confounding effects of tissue scaling. Mol Ecol **25**: 5114-5128.

MORGULIS, A., E. M. GERTZ, A. A. SCHÄFFER, and R. AGARWALA, 2006 A fast and symmetric DUST implementation to mask low-complexity DNA sequences. J Comput Biol **13**: 1028-1040.

MUYLE, A., J. KÄFER, N. ZEMP, S. MOUSSET, F. PICARD *et al.*, 2016 SEX-DETector: A probabilistic approach to study sex chromosomes in non-model organisms. Genome Biol Evol **8**: 2530-2543.

NAM, K., and H. ELLEGREN, 2008 The chicken (*Gallus gallus*) Z chromosome contains at least three nonlinear evolutionary strata. Genetics **180**: 1131-1136.

NIELSEN, R., and Z. YANG, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148**: 929-936.

OBBARD, D. J., S. A. HARRIS, R. J. A. BUGGS, and J. PANNELL, 2006 Hybridization, polyploidy, and the evolution of sexual systems in *Mercurialis* (Euphorbiaceae). Evolution **60**: 1801-1815.

ONO, T., 1939 Polyploidy and sex determination in *Melandrium*. I. Colcicine-induced polyploids of *Melandrium album*. Botanical Magazine **53**: 549-556.

OSSOWSKI, S., K. SCHNEEBERGER, J. I. LUCAS-LLEDÓ, N. WARTHMANN, R. M. CLARK *et al.*, 2010 The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science **327**: 92-94.

PANNELL, J., M. E. DORKEN, B. PUJOL, and R. BERJANO, 2008 Gender variation and transitions between sexual systems in *Mercurialis annua* (Euphorbiaceae). Int J Plant Sci **169**: 129-139.

PAOLUCCI, I., M. GAUDET, V. JORGE, I. BERITOGNOLO, S. TERZOLI *et al.*, 2010 Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. Tree Genetics & Genomes **6**: 863-875.

PAPADOPULOS, A. S., M. CHESTER, K. RIDOUT, and D. A. FILATOV, 2015 Rapid Y degeneration and dosage compensation in plant sex chromosomes. Proc Natl Acad Sci USA **112**: 13021-13026.

PARISI, M., R. NUTTALL, D. NAIMAN, G. BOUFFARD, J. MALLEY *et al.*, 2003 Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. Science **299**: 697-700.

PARSCH, J., and H. ELLEGREN, 2013 The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet **14**: 83-87.

PICQ, S., S. SANTONI, T. LACOMBE, M. LATREILLE, A. WEBER *et al.*, 2014 A small XY chromosomal region explains sex determination in wild dioecious *V. vinifera* and the reversal to hermaphroditism in domesticated grapevines. BMC Plant Biol **14**: 229.

PUCHOLT, P., A. C. RÖNNBERG-WÄSTLJUNG, and S. BERLIN, 2015 Single locus sex determination and female heterogamety in the basket willow (*Salix viminalis* l.). Heredity **114**: 575-583.

PUCHOLT, P., A. E. WRIGHT, L. LIU CONZE, J. E. MANK, and S. BERLIN, 2017 Recent sex chromosome divergence despite ancient dioecy in the willow *Salix viminalis*. Mol Biol Evol

PUTEROVA, J., Z. KUBAT, E. KEJNOVSKY, W. JESIONEK, J. CIZKOVA *et al.*, 2018 The slowdown of Y chromosome expansion in dioecious Silene latifolia due to DNA loss and male-specific silencing of retrotransposons. BMC Genomics **19**: 153.

QIU, S., R. BERGERO, A. FORREST, V. B. KAISER, and D. CHARLESWORTH, 2010 Nucleotide diversity in *Silene latifolia* autosomal and sex-linked genes. Proc Biol Sci **277**: 3283-3290.

QUANTDT, H. J., A. PÜHLER, and I. BROER, 1993 Transgenic root nodules of *Vicia hirsuta*: A fast and efficient system for the study of gene expression in indeterminate-type nodules. Mol Plant-Microbe Interact **6**: 699-706.

R, DEVELOPMENT, CORE, and Team, 2007 R: A Language and Environment for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org. R-project.org

RAHMAN, A. Y., A. O. USHARRAJ, B. B. MISRA, G. P. THOTTATHIL, K. JAYASEKARAN *et al.*, 2013 Draft genome sequence of the rubber tree *Hevea brasiliensis*. BMC Genomics **14**: 75.

RASTAS, P., 2017 Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics **33**: 3726-3732.

RAUTENBERG, A., D. FILATOV, B. SVENNBLAD, N. HEIDARI, and B. OXELMAN, 2008 Conflicting phylogenetic signals in the SlX1/Y1 gene in *Silene*. BMC Evol Biol **8**: 299.

RENNER, S. S., 2014 The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. Am J Bot **101**: 1588-1596.

RICE, W. R., 1987 The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution **41**: 911-914.

ROBINSON, M. D., D. J. MCCARTHY, and G. K. SMYTH, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139-140.

RUSSELL, J. R. W., and J. PANNELL, 2015 Sex determination in dioecious *Mercurialis annua* and its close diploid and polyploid relatives. Heredity **114**: 262-271.

SANDERSON, B. J., L. WANG, P. TIFFIN, Z. WU, and M. S. OLSON, 2018 Sex-biased gene expression in flowers, but not leaves, reveals secondary sexual dimorphism in *Populus balsamifera*. New Phytol

SATO, S., H. HIRAKAWA, S. ISOBE, E. FUKAI, A. WATANABE *et al.*, 2011 Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res **18**: 65-76.

SCHMIEDER, R., and R. EDWARDS, 2011 Quality control and preprocessing of metagenomic datasets. Bioinformatics **27**: 863-864.

SHARMA, E., A. KÜNSTNER, B. A. FRASER, G. ZIPPRICH, V. A. KOTTLER *et al.*, 2014 Transcriptome assemblies for studying sex-biased gene expression in the guppy, *Poecilia reticulata*. BMC Genomics **15**: 400.

SIMÃO, F. A., R. M. WATERHOUSE, P. IOANNIDIS, E. V. KRIVENTSEVA, and E. M. ZDOBNOV, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**: 3210-3212.

SMARDA, P., P. BUREŠ, J. SMERDA, and L. HOROVÁ, 2012 Measurements of genomic GC content in plant genomes with flow cytometry: a test for reliability. New Phytol **193**: 513-521.

SMIT, A. F. A., R. HUBLEY, and G. P, 2013 RepeatMasker Open-4.0. http://www.repeatmasker.org.

SMITH, B., 1955 Sex chromosomes and natural polyploidy in dioecious *Rumex*. J Hered **46**: 226-232.

SMITH, S. A., and M. J. DONOGHUE, 2008 Rates of molecular evolution are linked to life history in flowering plants. Science **322**: 86-89.

STAMATAKIS, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**: 1312-1313.

TELGMANN-RAUBER, A., A. JAMSARI, M. S. KINNEY, J. C. PIRES, and C. JUNG, 2007 Genetic and physical maps around the sex-determining M-locus of the dioecious plant *Asparagus*. Mol Genet Genomics **278**: 221-234.

TENNESSEN, J. A., R. GOVINDARAJULU, A. LISTON, and T. L. ASHMAN, 2016 Homomorphic ZW chromosomes in a wild strawberry show distinctive recombination heterogeneity but a small sex-determining region. New Phytol **211**: 1412-1423.

TOUPS, M., N. RODRIGUES, N. PERRIN, and M. KIRKPATRICK, 2019 A reciprocal translocation radically reshapes sex-linked inheritance in the common frog. Mol Ecol doi: 10.1111/mec.14990.

TROMBETTA, B., D. SELLITTO, R. SCOZZARI, and F. CRUCIANI, 2014 Inter- and intraspecies phylogenetic analyses reveal extensive X-Y gene conversion in the evolution of gametologous sequences of human sex chromosomes. Mol Biol Evol **31**: 2108-2123.

TSAGKOGEORGA, G., V. CAHAIS, and N. GALTIER, 2012 The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. Genome Biol Evol **4**: 740-749.

VELTSOS, P., G. COSSARD, E. BEAUDOING, G. BEYDON, D. SAVOVA BIANCHI *et al.*, 2018 Size and content of the sex-determining region of the Y chromosome in dioecious *Mercurialis annua*, a plant with homomorphic sex chromosomes. Genes (Basel) **9**: 277.

VURTURE, G. W., F. J. SEDLAZECK, M. NATTESTAD, C. J. UNDERWOOD, H. FANG *et al.*, 2017 GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics **33**: 2202-2204.

WANG, J., J. K. NA, Q. YU, A. R. GSCHWEND, J. HAN *et al.*, 2012a Sequencing papaya X and Y-h chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc Natl Acad Sci USA **109**: 13710-13715.

WANG, K., Z. WANG, F. LI, W. YE, J. WANG *et al.*, 2012b The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet **44**: 1098-1103.

WESTERGAARD, M., 1958 The mechanism of sex determination in dioecious splants. Adv Genet **9**: 217-281.

WRIGHT, A. E., I. DAROLTI, N. I. BLOCH, V. OOSTRA, B. SANDKAM *et al.*, 2017 Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. Nat Commun **8**: 14251.

WU, C. I., and A. W. DAVIS, 1993 Evolution of postmating reproductive isolation: the composite nature of Haldane’s rule and its genetic bases. Am Nat **142**: 187-212.

WU, M., and R. C. MOORE, 2015 The evolutionary tempo of sex chromosome degradation in *Carica papaya*. J Mol Evol **80**: 265-277.

WU, T. D., and C. K. WATANABE, 2005 GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics **21**: 1859-1875.

WU, X., J. WANG, J. K. NA, Q. YU, R. C. MOORE *et al.*, 2010 The origin of the non-recombining region of sex chromosomes in *Carica* and *Vasconcellea*. Plant J **63**: 801-810.

XUE, W., J. T. LI, Y. P. ZHU, G. Y. HOU, X. F. KONG *et al.*, 2013 L_RNA_scaffolder: scaffolding genomes with transcripts. BMC Genomics **14**: 604.

YAMAMOTO, K., Y. ODA, A. HASEDA, S. FUJITO, T. MIKAMI *et al.*, 2014 Molecular evidence that the genes for dioecism and monoecism in *Spinacia oleracea* L. are located at different loci in a chromosomal region. Heredity **112**: 317-324.

YANG, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol **24**: 1586-1591.

ZEMP, N., R. TAVARES, A. MUYLE, D. CHARLESWORTH, G. A. MARAIS *et al.*, 2016 Evolution of sex-biased gene expression in a dioecious plant. Nat Plants **2**: 16168.

Figure 1: Basic karyotype for a *Mercurialis annua* male individual. The top row shows DAPI counterstain. The middle row shows chromosomes counterstained with DAPI (blue) after bicolor FISH with 45S rDNA (red) and 5S rDNA (green). The bottom row shows an idiogram of the male chromosomes. The white bar in the middle row represents 5 µm.

Figure 2: Genetic map of the eight linkage groups (LG) for *M. annua*, with each LG represented by the male (left) and female (right) maps. Lines linking the two maps indicate the mapping position of groups of transcripts that segregated together (10,076 total transcripts). The sex-linked region is at 53.85 cM of chromosome 1 (LG1) on the male recombination map, and is highlighted by the gray vertical bar. The two pseudoautosomal regions (PAR) comprise the rest of LG1.

Figure 3: Density plots of synonymous site divergence (d_S) for: (A) all *M. annua* X/Y haplotypes, (B) autosomal *M. annua* / *M. huetii* homologues, (C) autosomal *M. annua* / *R. communis* homologues, (D) X-linked *M. annua* / *M. huetii* homologues, and (E) X-linked *M. annua* / *R. communis* homologues. The numbers on the bottom right indicate the sample size for each plot.

Figure 4: Boxplots of tree-based estimates of d_N/d_S, d_S and d_N for the X, Y, *M. huetii* and *R. communis* branches of the 72 gene trees that correspond to tree topology (((*M. annua*-X, *M. annua*-Y), *M. huetii*), *R. communis*). A Y-sequence d_N/d_S outlier of 2.7 and an *M. huetii*-sequence d_S outlier of 0.23 are not shown.

Figure 5: Summary of transcript metrics calculated for sequences across the female recombination map of the putative sex chromosome LG1, based on analysis of X and Y sequences combined. The non-recombining region in males is located between the vertical dotted lines. Orange points in the top panel indicate significant sex-bias (5% FDR). Potential candidate sex-determining genes are indicated by letters, with their putative function given in the inset legend. The only verified premature stop codon on the Y copy is indicated by k_s beside the non-recombining region (i.e., for a kelch protein gene). Transcripts for which the Y sequence diverged prior to the *M. annua*/*M. huetii* species split are labeled '2'.

Figure 6: Graphical summary of the analysis of genes based on their sex-bias (blue: male-biased, red: female-biased, gray: unbiased) and sex linkage (SL: sex linked, PAR: remaining LG1, Au: LG 2-7). A) absolute $\log_2$FC expression difference between the sexes, B) nucleotide diversity (π), C) pairwise d_N/d_S compared to *M. huetii* homologues, D) pairwise d_N/d_S compared to *R. communis* homologues, E) relative expression of Y over X alleles. The numbers to the right indicate the number of genes in each category. Data are plotted prior to Box-Cox transformation.

Figure 7: Summary of the effect of sex linkage on SNP metrics from genome capture data across the species range. (A) F_{ST} for sex linkage category, (B) distribution of F_{ST} across the female recombination map for LG1, (C) ratio of male/female heterozygosity for sex linkage category, (D) ratio of male/female heterozygosity across the female recombination map of LG1, (E) observed heterozygosity in males (blue) and females (red) across LG1. Asterisks in A and C indicate significant Wilcoxon tests ($p < 0.001$), the sex-linked region is indicated by vertical lines in B, D and E.

Table 1. Summary statistics of contigs containing sex-linked, autosomal or all expressed ORFs. Only ORFs supported by expression data were used for analysis of sex linkage (*de novo* predicted genes were excluded). Confidence intervals are one standard deviation.

	Whole genome	Expressed ORFs	Sex-linked	Autosomal
Number of contigs	720,537	15,392	538	7,579
Total base pairs	546,375,413	176,423,112	8,271,542	105,835,810
% GC	34.7	34.5	34.5	34.3
Average contig length	15,330 (±20,028)	11,462 (±17,207)	15,374 (±21,615)	13,964 (±19,678)
Total repeat density (%)	51.5	18.7	18.4	18.3
Final coding transcript number	34,105	34,105	1,641	23,502
Average transcript length	887 (±754)	810 (±702)	1,265 (±878)	1,242 (±853)
Mean effective number of codons	53.7	53.7	52.2	51.9
Number of SNPs	181,567	180,925	5,274	102,881
Number of ORFs transcripts with SNPs	16,667 (49%)	16,572 (49%)	545 (33%)	7,714 (33%)

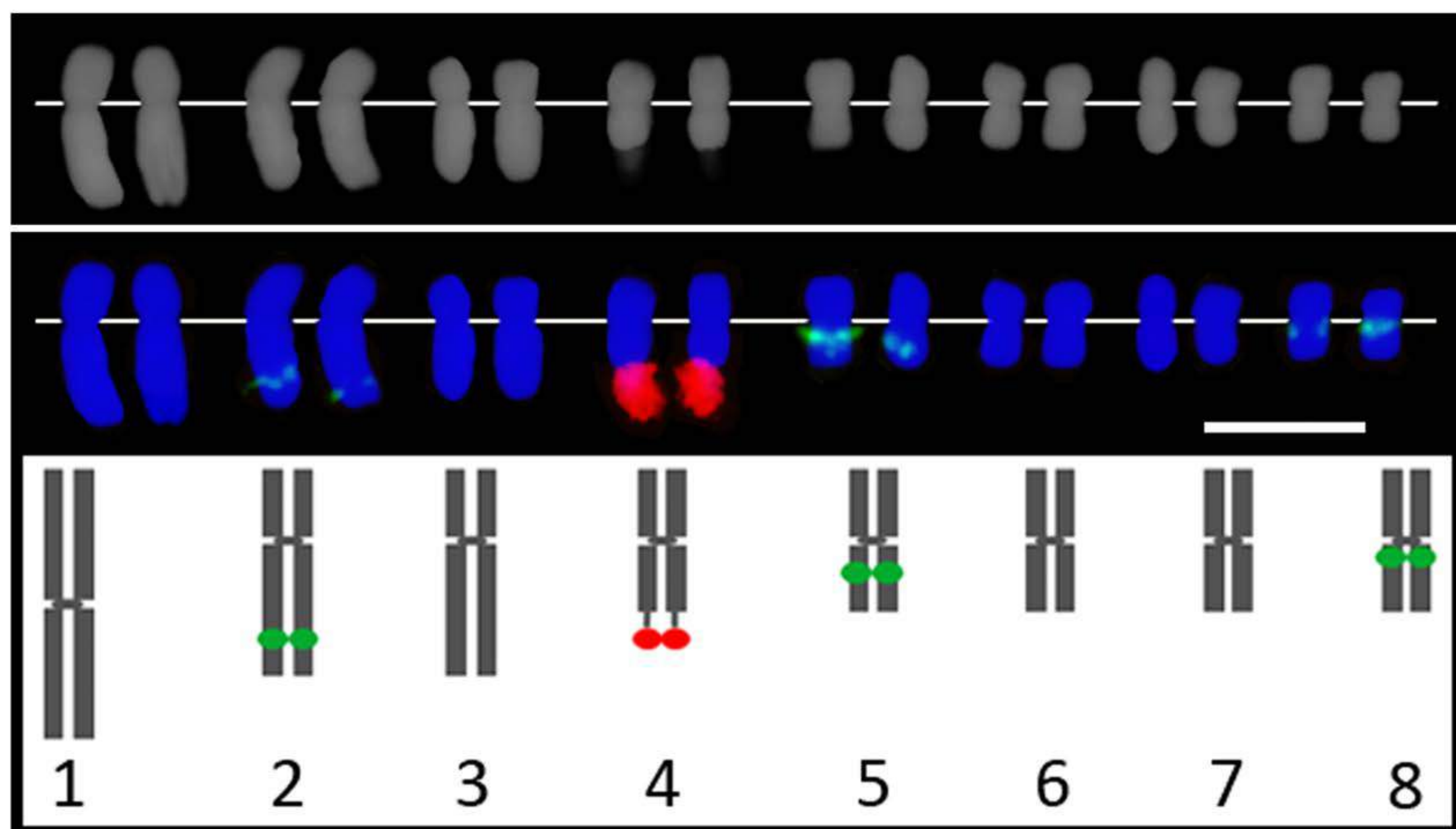
Median nucleotide diversity (π) / ORF	0.0035 (±0.0113)	0.0035 (±0.0113)	0.006647 (±0.008079)	0.007951 (±0.01016)

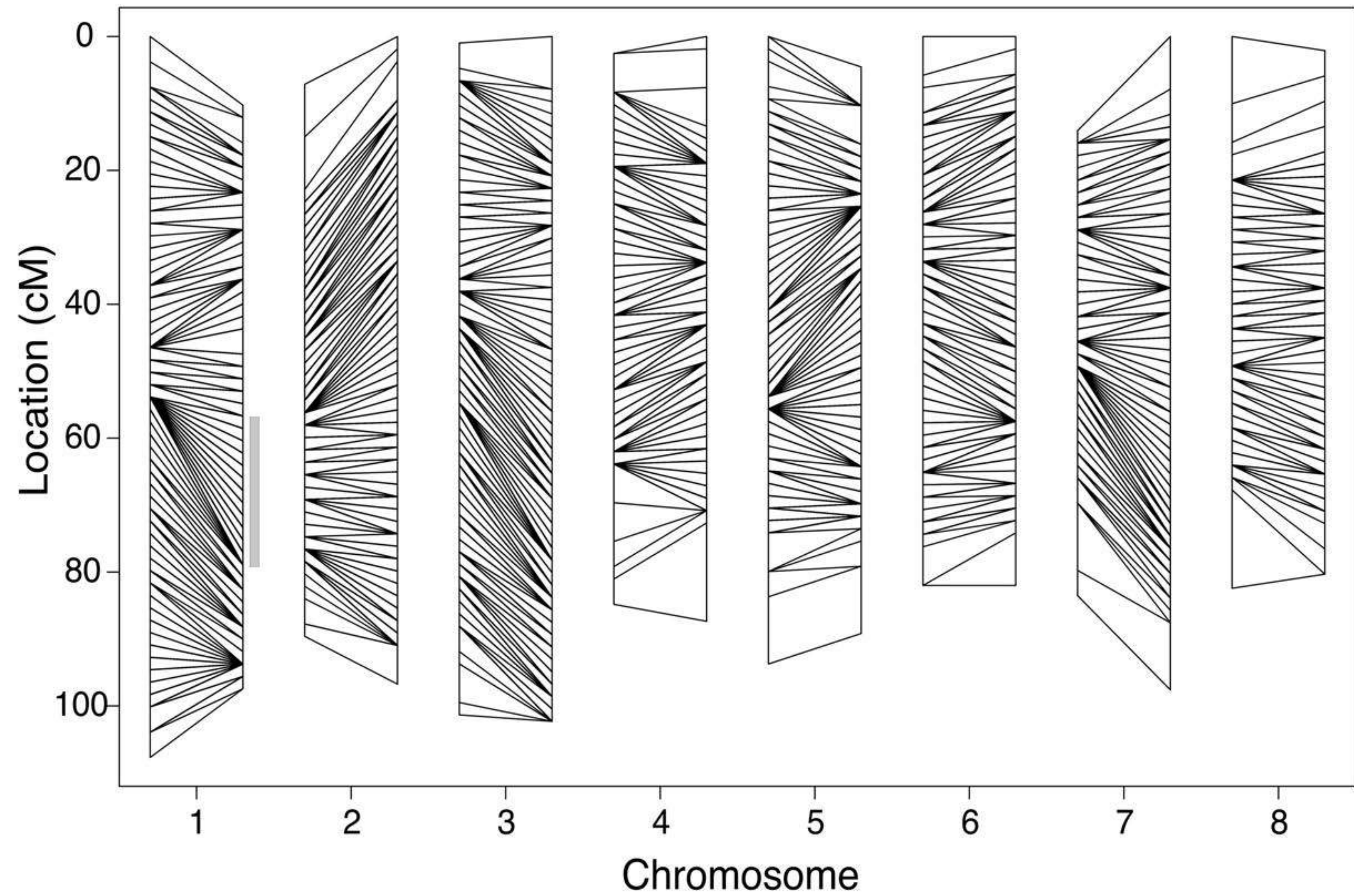
Table 2. Average pairwise d_S and d_N for ORFs of *M. annua* from X/Y haplotype comparison within *M. annua* or against *M. huetii* and *R. communis* sequences. Autosomal ORF comparisons with the other species are also reported. Haplotypes with premature stop codons were excluded.

	d_S	d_N	d_N/d_S
X/Y pairs	0.011	0.002	0.182
X / *M. huetii*	0.112	0.016	0.143
Y / *M. huetii*	0.113	0.017	0.150
X / *R. communis*	0.548	0.070	0.128
Y / *R. communis*	0.555	0.070	0.126
M. annua / *M. huetii*	0.106	0.014	0.149
M. annua / *R. communis*	0.565	0.079	0.140

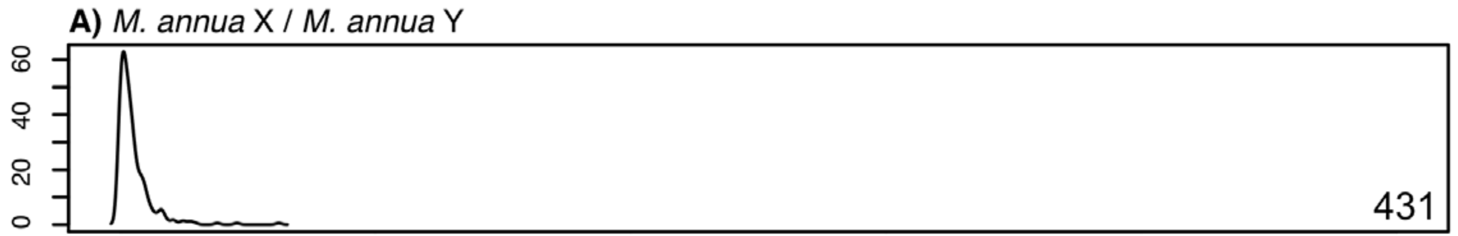
Table 3. Results of chi-squared tests for differences in the frequencies of sex-biased versus -unbiased genes, male- versus female-biased genes, and candidate sex-determining genes (CG) versus other genes (not CG) between the pseudoautosomal region (PAR), the sex-linked region (SL), linkage group 1 (LG1) and autosomes (Au).

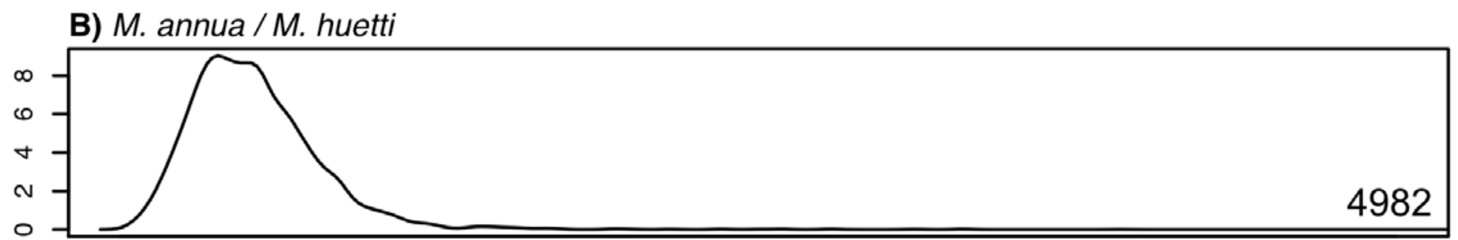
	Sex-biased vs unbiased			Female-biased vs male-biased			Candidate gene vs other gene		
	Unbiased	Sex-biased	Total	Female biased	Male biased	Total	Not CG	CG	Total
PAR	1168	41	1209	22	19	41	1177	32	1209
	96.6 %	3.4 %	100 %	53.7 %	46.3 %	100 %	97.4 %	2.6 %	100 %
SL	536	32	568	16	16	32	546	22	568
	94.4 %	5.6 %	100 %	50 %	50 %	100 %	96.1 %	3.9 %	100 %
Total	1704	73	1777	38	35	73	1723	54	1777
	95.9 %	4.1 %	100 %	52.1 %	47.9 %	100 %	97 %	3 %	100 %
Test	χ^2=4.381 df=1 φ=0.053 **p=0.036**			χ^2=0.006 df=1 φ=0.036 p=0.941			χ^2=1.578 df=1 φ=0.033 p=0.209		
Au	8056	243	8299	90	153	243	8094	207	8301
	97.1 %	2.9 %	100 %	37 %	63 %	100 %	97.5 %	2.5 %	100 %
LG1	1704	73	1777	38	35	73	1723	54	1777
	95.9 %	4.1 %	100 %	52.1 %	47.9 %	100 %	97 %	3 %	100 %
Total	9760	316	10076	128	188	316	9817	261	10078
	96.9 %	3.1 %	100 %	40.5 %	59.5 %	100 %	97.4 %	2.6 %	100 %
Test	χ^2=6.326 df=1 φ=0.026 **p=0.012**			X^2=4.649 df=1 φ=0.129 **p=0.031**			χ^2=1.515 df=1 φ=0.013 p=0.218		
Au	8056	243	8299	90	153	243	8094	207	8301
	97.1 %	2.9 %	100 %	37 %	63 %	100 %	97.5 %	2.5 %	100 %
SL	536	32	568	16	16	32	546	22	568
	94.4 %	5.6 %	100 %	50 %	50 %	100 %	96.1 %	3.9 %	100 %
Total	8592	275	8867	106	169	275	8640	229	8869
	96.9 %	3.1 %	100 %	38.5 %	61.5 %	100 %	97.4 %	2.6 %	100 %
Test	χ^2=12.066 df=1 φ=0.038 **p<0.001**			χ^2=1.496 df=1 φ=0.085 p=0.221			χ^2=3.493 df=1 φ=0.021 p=0.062		

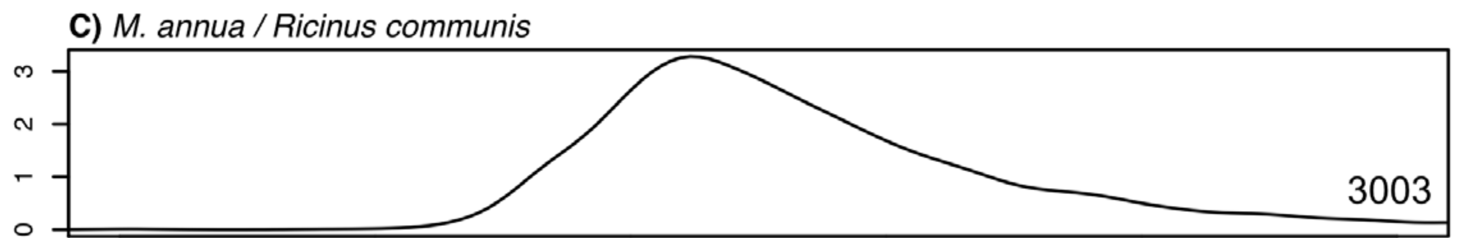

1
2
3
4
5
6
7
8

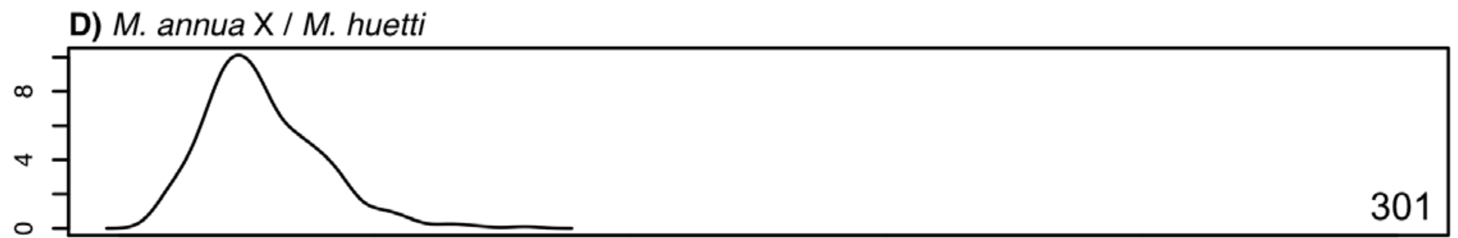


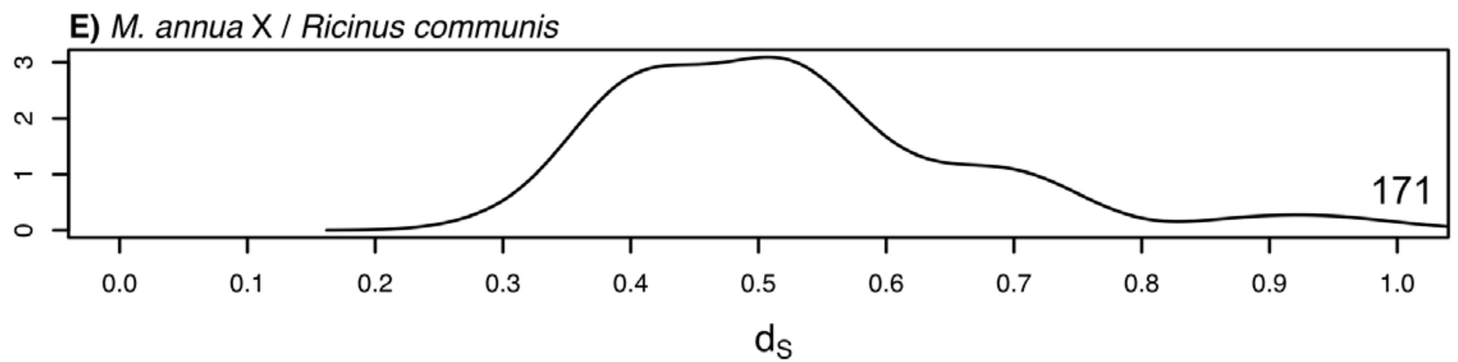
Location (cM)
0
20
40
60
80
100
1
2
3
4
5
6
7
8
Chromosome

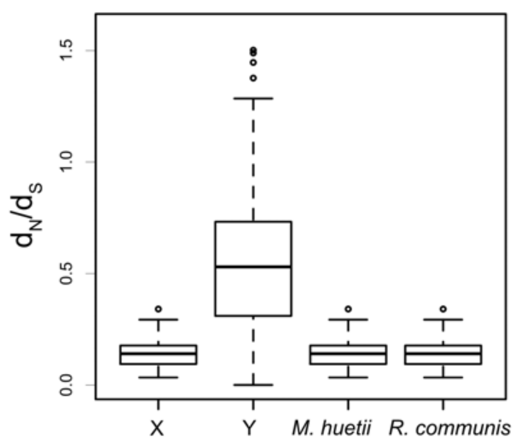

A) M. annua X / M. annua Y
0
20
40
60
431


B) M. annua / M. huetti
0
2
4
6
8
4982

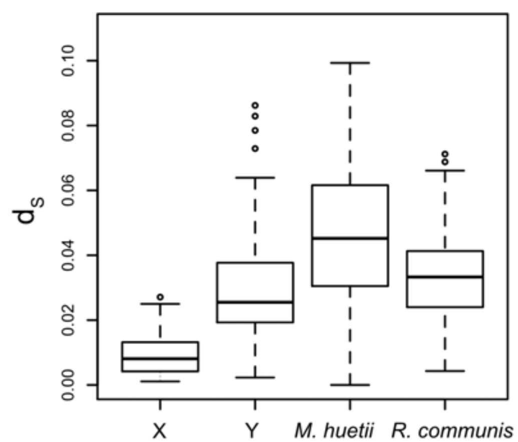

Frequency
C) M. annua / Ricinus communis
0
1
2
3
3003


D) M. annua X / M. huetti
0
4
8
301

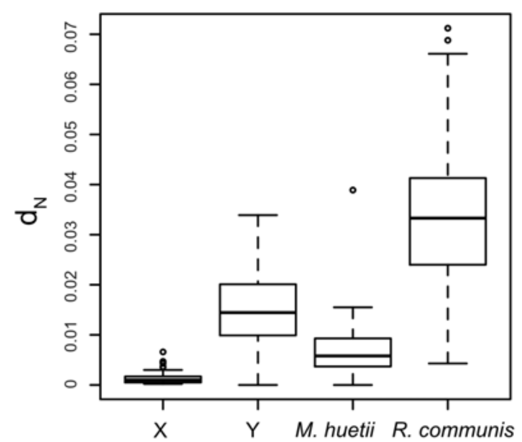

E) M. annua X / Ricinus communis
0
1
2
3
171
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
dS



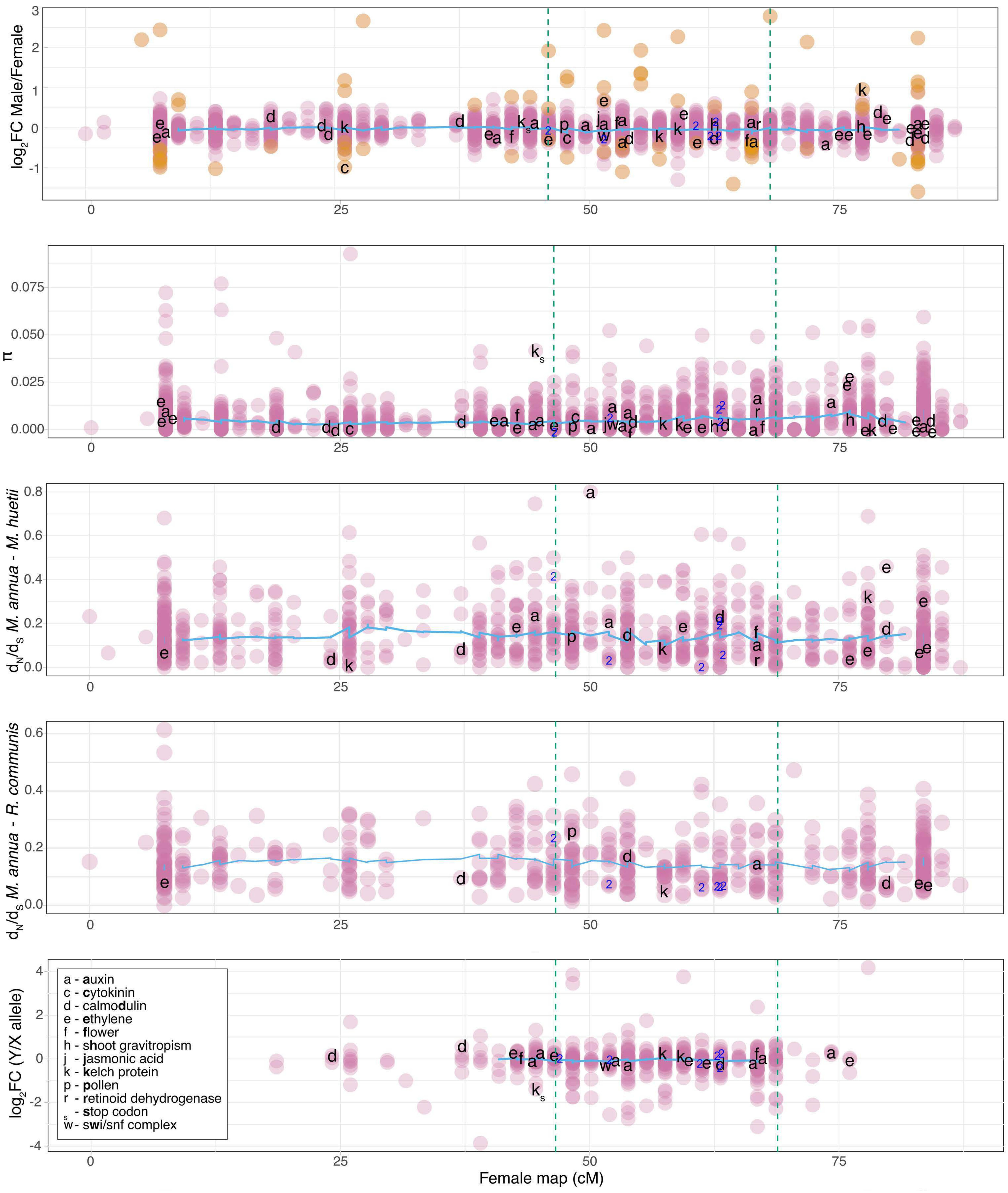
d_N/d_S
1.5
1.0
0.5
0.0
X
Y
M. huetii
R. communis



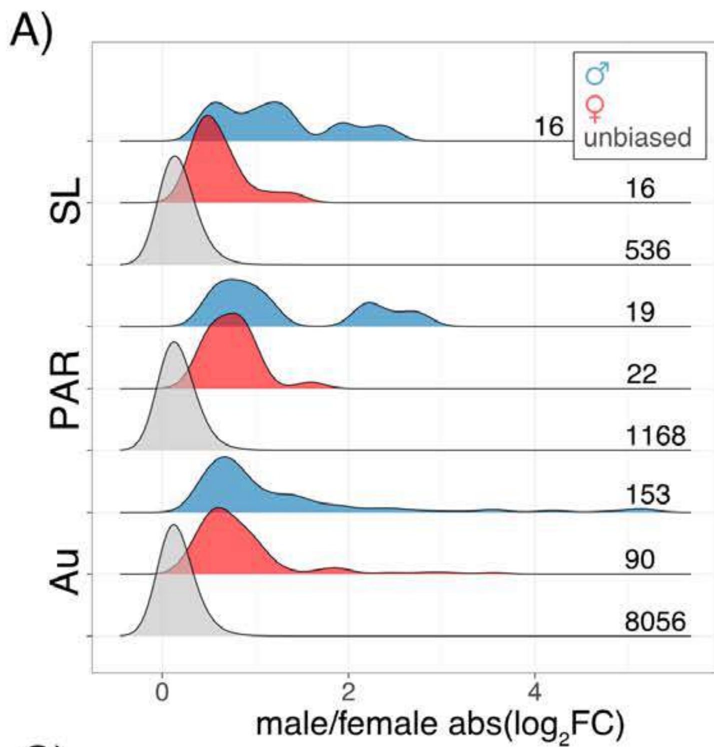
d_S
0.10
0.08
0.06
0.04
0.02
0.00
X
Y
M. huetii
R. communis

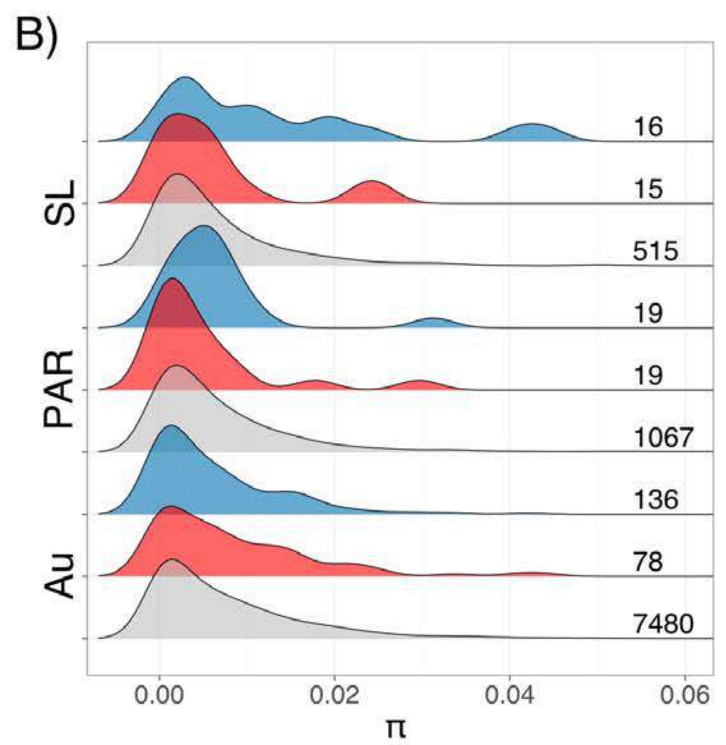


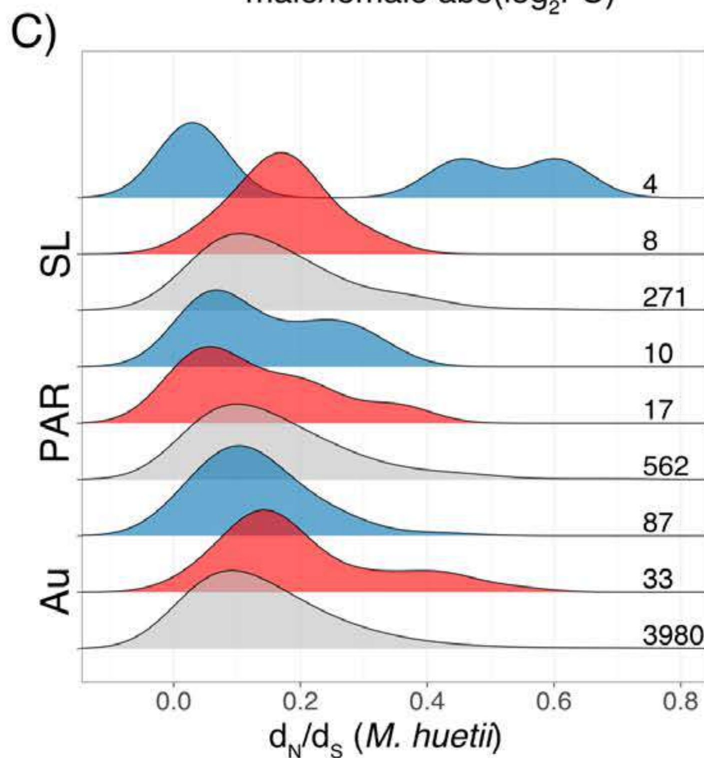
d_N
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
X
Y
M. huetii
R. communis

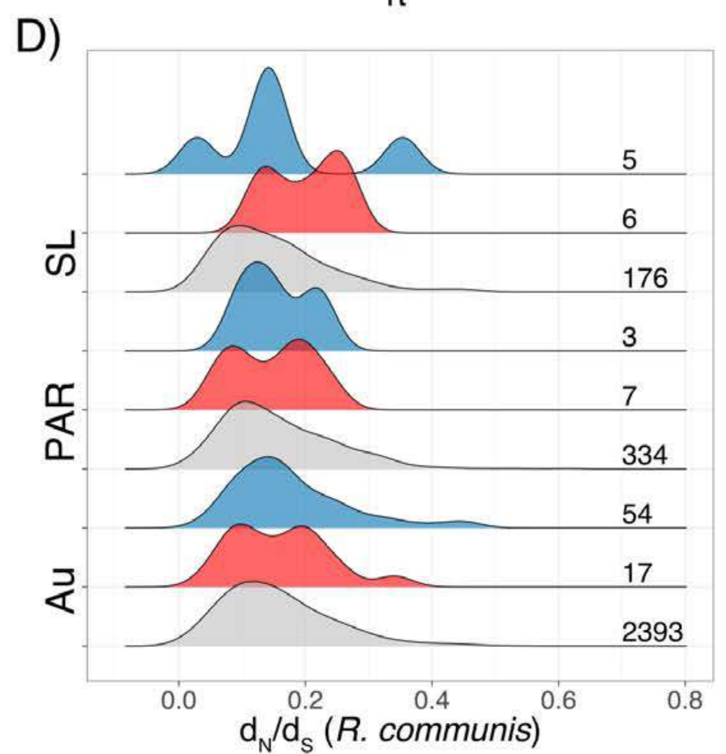


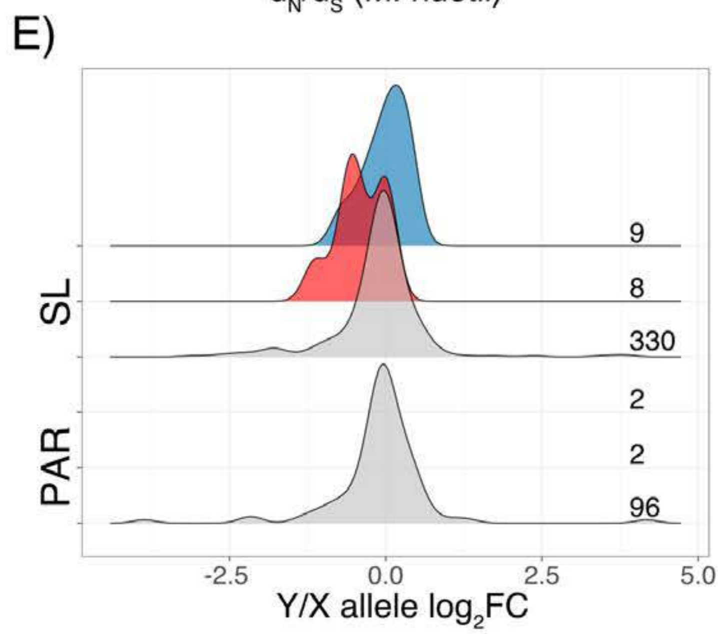
log2FC Male/Female
π
dN/dS M. annua - M. huetii
dN/dS M. annua - R. communis
log2FC (Y/X allele)
Female map (cM)
a - auxin
c - cytokinin
d - calmodulin
e - ethylene
f - flower
h - shoot gravitropism
j - jasmonic acid
k - kelch protein
p - pollen
r - retinoid dehydrogenase
s - stop codon
w - swi/snf complex

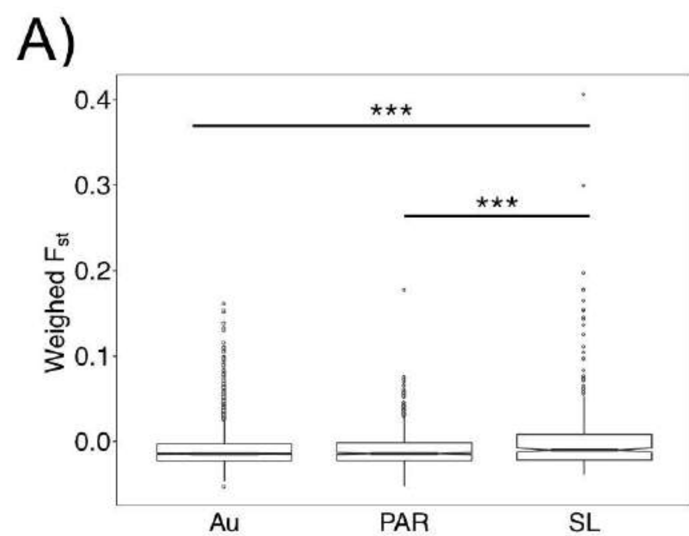

A)
♂
♀
unbiased
SL
PAR
Au
16
16
536
19
22
1168
153
90
8056
0
2
4
male/female abs(log_2FC)

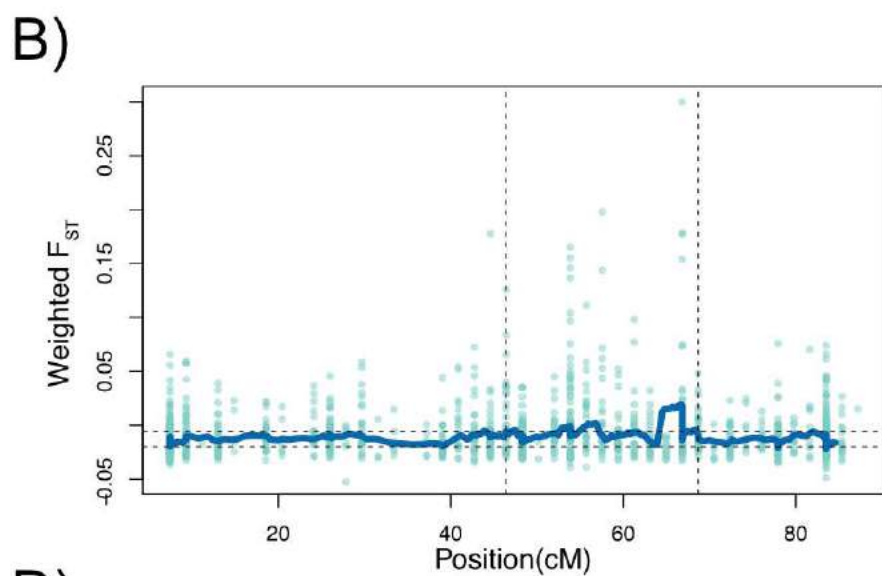

B)
SL
PAR
Au
16
15
515
19
19
1067
136
78
7480
0.00
0.02
0.04
0.06
π

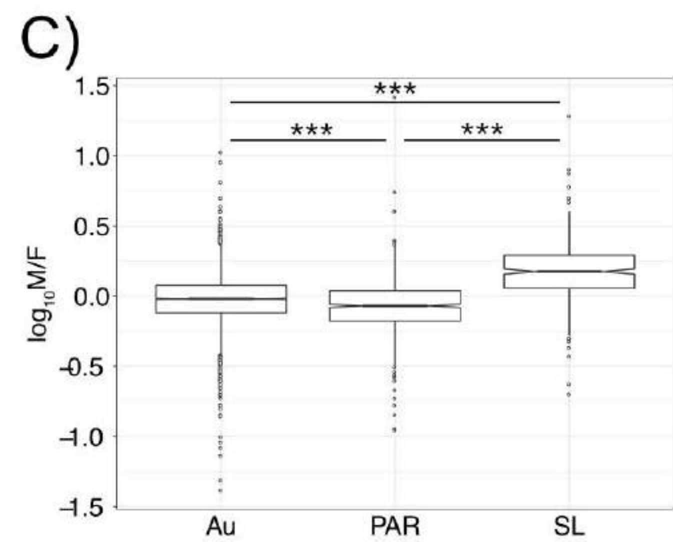

C)
SL
PAR
Au
4
8
271
10
17
562
87
33
3980
0.0
0.2
0.4
0.6
0.8
d_N/d_S (*M. huetii*)

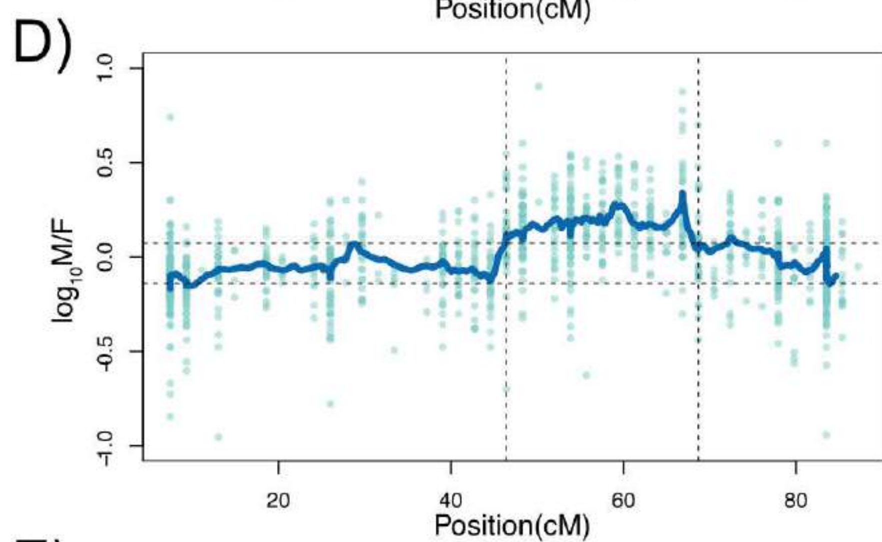

D)
SL
PAR
Au
5
6
176
3
7
334
54
17
2393
0.0
0.2
0.4
0.6
0.8
d_N/d_S (*R. communis*)

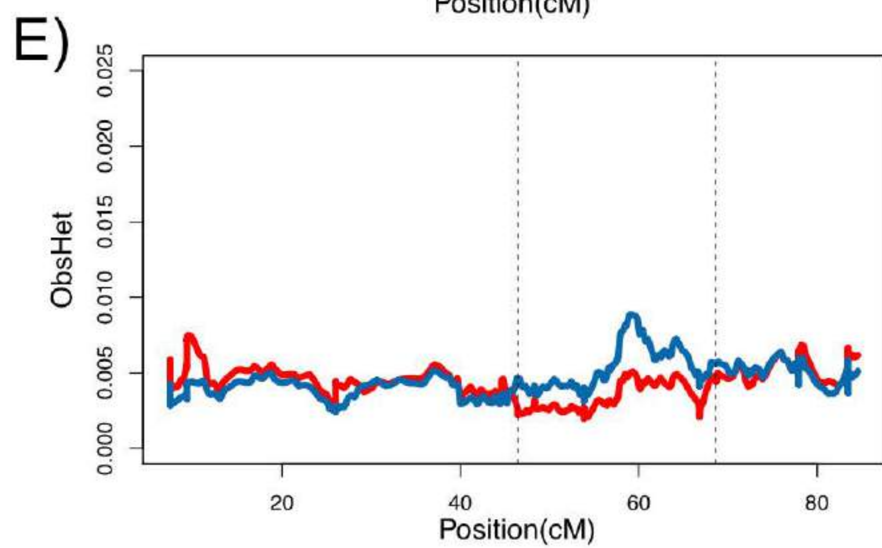

E)
SL
PAR
9
8
330
2
2
96
-2.5
0.0
2.5
5.0
Y/X allele log_2FC


A)
***
***
Weighed F_{st}
0.4
0.3
0.2
0.1
0.0
Au
PAR
SL


B)
Weighted F_{ST}
0.25
0.15
0.05
-0.05
20
40
60
80
Position(cM)


C)
***
***
***
$log_{10}M/F$
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
Au
PAR
SL


D)
$log_{10}M/F$
1.0
0.5
0.0
-0.5
-1.0
20
40
60
80
Position(cM)


E)
ObsHet
0.025
0.020
0.015
0.010
0.005
0.000
20
40
60
80
Position(cM)