Plant genera Cannabis and Humulus share the same pair of well-differentiated sex chromosomes

Djivan Prentout, Natasa Stajner, Andreja Cerenak, Theo Tricou, Celine Brochier-Armanet, Jernej Jakse, Jos Käfer, Gabriel Marais

New Phytologist

Research Article

Plant genera *Cannabis* and *Humulus* share the same pair of well-differentiated sex chromosomes

D Prentout[1], N Stajner[2], A Cerenak[3], T Tricou[1], C Brochier-Armanet[1], J Jakse[2], J Käfer[1]*, GAB Marais[1,4]*

1. Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR 5558, F-69622 Villeurbanne, France

2. Department of Agronomy, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, SI-1000 Ljubljana, Slovenia

3. Slovenian Institute of Hop Research and Brewing, Cesta Zalskega Tabora 2, SI-3310 Zalec, Slovenia

4. Current address: LEAF- Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa, Portugal

*** These authors contributed equally to this work**

Author for correspondence:

Djivan Prentout

djivan.prentout@univ-lyon1.fr

Word count

Total Main text: 7,316 (citations included) – 5 Figures (to be published in colour) – 3 Tables
Introduction: 1,126
Materials and Methods: 2,051 – 1 Figure
Results: 2,129 – 4 Figures – 3 Tables
Discussion: 2,007

Supporting Information: 1,679 – 10 Figures – 5 Tables

Summary

- We recently described, in *Cannabis sativa*, the oldest sex chromosome system documented so far in plants (12-28 Myo). Based on the estimated age, we predicted that it should be shared by its sister genus *Humulus*, which is known to also possess XY chromosomes.
- Here, we used transcriptome sequencing of a F1 family of *Humulus lupulus* to identify and study the sex chromosomes in this species using the probabilistic method SEX-DETector.
- We identified 265 sex-linked genes in *H. lupulus*, which preferentially mapped to the *C. sativa* X chromosome. Using phylogenies of sex-linked genes, we showed that a region of the sex chromosomes had already stopped recombining in an ancestor of both species. Furthermore, as in *C. sativa*, Y-linked gene expression reduction is correlated to the position on the X chromosome, and highly Y degenerated genes showed dosage compensation.
- We report, for the first time in Angiosperms, a sex chromosome system that is shared by two different genera. Thus, recombination suppression started at least 21-25 My ago, and then (either gradually or step-wise) spread to a large part of the sex chromosomes (~70%), leading to a degenerated Y chromosome.

Keywords

Cannabaceae; dioecy; dosage compensation; *Humulus lupulus*; sex chromosomes; Y degeneration

Introduction

Among more than 15,000 dioecious angiosperm species (*i.e.* species with separate sexes; Renner, 2014), less than twenty sex chromosome systems have been studied with genomic data (Ming *et al.*, 2011; Baránková *et al.*, 2020). Most plants with sex chromosomes exhibit male heterogamety, with XY chromosomes in males, and XX chromosomes in females (Westergaard, 1958; Charlesworth, 2016). The portion of the Y chromosome that never recombines with the X experiences reduced selection, which results in an accumulation of deleterious mutations and transposable elements (Charlesworth & Charlesworth, 2000). This accumulation of transposable elements initially leads to an increase of the size of the Y chromosome, which becomes larger than the X (Ming *et al.*, 2011). When Y degeneration progresses, genetic material can be lost without fitness costs and the Y may shrink (Ming *et al.*, 2011). Therefore, after sufficient time of divergence, we may observe chromosome heteromorphy, *i.e.* a Y chromosome larger or smaller than the X chromosome, depending on the progress of degeneration (Ming *et al.*, 2011). Classically, heteromorphy was determined using light microscopy, which is rather imprecise and size differences of about 10% could be considered homomorphic (see Divashuk *et al.*, 2014). While heteromorphy often corresponds to the later stages of sex chromosome evolution, it is nevertheless possible that sex chromosomes are homomorphic despite a large non-recombining region and strong degeneration of the Y chromosome (e.g. Prentout *et al.*, 2020). Moreover, some systems do not evolve large non-recombining region and stay homomorphic (Renner & Muller, 2021).

In plants, dioecy is often of recent origin (Renner, 2014; Käfer *et al.*, 2017), thus limiting the age of the sex chromosomes. Indeed, several rather recently evolved (less than 10 million years (My) old) homomorphic sex chromosome systems with small non-recombining regions have been described, as in *Carica papaya* and *Asparagus officinalis* (Wu & Moore, 2015; Harkess *et al.*, 2017). Heteromorphic sex chromosome systems are also found, with the Y being larger than the X, but recombination suppression happened also relatively recently (less than 20 My ago), as in *Silene latifolia* and *Coccinia grandis* (Sousa *et al.*, 2013; Krasovec *et al.*, 2018; Fruchard *et al.*, 2020).

A few cases in which dioecy evolved longer ago also exist (Käfer *et al.*, 2017), but no strongly degenerated sex chromosomes have been described so far (Renner & Muller, 2021). Pucholt *et al.* (2017) described very young sex chromosomes in *Salix viminalis* despite ancestral dioecy for the sister genera *Salix* and *Populus*. Thus, either the sex chromosomes evolved independently in different species, or there have been frequent turnovers. In the fully dioecious palm tree genus

Phoenix, a sex-linked region evolved before the speciation of the fourteen known species (Cherif *et al.*, 2016; Torres *et al.*, 2018). These sex chromosomes might be old, but do not appear to be strongly differentiated. A similar situation has been reported in the grapevine (*Vitis*) genus (Badouin *et al.*, 2020; Massonet *et al.*, 2020), possibly because sex chromosome evolution is slowed down in such perennials with long generation time (Muyle *et al.*, 2017).

Thus, while homologous sex chromosomes are sometimes shared between species belonging to the same genus (e.g. *Silene* sect. Melandrium, *Phoenix*) (Cherif *et al.*, 2016; Bacovsky *et al.*, 2020), homologous sex chromosomes between different genera have never been described in plants so far. This situation is in stark contrast to the situation in animals, for which several systems are more than 100 My old and are shared by whole classes, *e.g.* birds and mammals (Ohno, 1969; Fridolfsson *et al.*, 1998; Cortez *et al.*, 2014). Thus, although undoubtedly sex chromosomes have been less intensively studied in plants, there seem to be fundamental differences in the evolution of sex chromosomes in plants and animals (*e.g.* lack of strong sexual dimorphism in plants, discussed in Renner & Muller, 2021). However, the extent of the differences needs to be clarified and more plant sex chromosomes need to be studied.

Dioecy very likely evolved before the genera *Cannabis* and *Humulus* split, and might even be ancestral in the Cannabaceae family (Yang *et al.*, 2013; Zhang *et al.*, 2018). *Cannabis sativa* (marijuana and hemp) is a dioecious species with nearly homomorphic XY chromosomes (with homomorphy defined as above). These sex chromosomes have a large non-recombining region and are estimated to have started diverging between 12 and 28 My ago (Peil *et al.*, 2003; Divashuk *et al.*, 2014, Prentout *et al.*, 2020).

As for *C. sativa*, cytological analyses of *Humulus lupulus* (hop) found a XY chromosome system with a large non-recombining region, but the Y chromosome is smaller than the X (Shephard *et al.*, 2000; Karlov *et al.*, 2003; Divashuk *et al.*, 2011). The *H. lupulus* and *C. sativa* lineages split between 21 and 25 My old (Divashuk *et al.*, 2014; Jin *et al.*, 2020), which is more recently than our higher bound estimate of the age of the *C. sativa* sex chromosomes (28 Mya; Prentout *et al.*, 2020). It is thus possible that the sex chromosomes of *C. sativa* and *H. lupulus* evolved from the same pair that already stopped recombining in their common ancestor, a question we address here.

As in many cultivated dioecious species, only female hop plants are harvested. Hop is used in beer brewing for its bitterness, and its production is increasing worldwide (Neve, 1991; King & Pavlovic, 2017), mostly because of the craft beer revolution (Barth-Haas, 2019; Mackinnon & Pavlovic, 2019). The molecule responsible for hop flower bitterness, lupulin, is concentrated in

female ripe inflorescences, called cones (Okada & Ito, 2001). In pollinated cones, the presence of seed reduces their brewing quality; since *H. lupulus* is wind pollinated, a single male plant in the hop field or its vicinity can cause broad scale damage to the crop (Thomas & Neve, 1976).

Usually, hop is not grown from seeds, so female-only cultures are easy to obtain, and there is no need for large-scale early sexing as in *Cannabis sativa* (cf. Prentout *et al.*, 2020). However, for varietal improvement where controlled crosses are needed, knowing the sex early might be beneficial. In *H. lupulus*, sexing is reliable 1-2 years after the sewing (Conway and Snyder, 2008; Patzak *et al.*, 2002). A few markers have been developed, but the use of Y-specific coding sequences may increase marker quality (Patzak *et al.*, 2002, Cerenak *et al.*, 2019).

Here we sequenced the transcriptome of fourteen *H. lupulus* individuals. These individuals came from a cross, from which we sequenced the parents and six offspring of each sex. We used the probabilistic approach SEX-DETector, which is based on allele segregation analysis within a cross, to identify sex-linked sequences (Muyle *et al.*, 2016). From theses analyses on *H. lupulus* and our previous results on *C. sativa* (Prentout *et al.*, 2020) we describe for the first time well-differentiated sex chromosomes shared by two different genera in plants.

Materials and Methods

Biological material and RNA-sequencing

As indicated in Fig. **1a**, we conducted a controlled cross for sequencing. The *H. lupulus* parents, cultivar 'Wye Target' (WT; female) and the Slovenian male breeding line 2/1 (2/1), as well as 6 female and 6 male F1 siblings (Jakše *et al.*, 2013) were collected in July 2019 in the experimental garden of Slovenian Institute of Hop Research and Brewing, Žalec.

All offspring were phenotypically confirmed to carry either male or female reproductive organs and showed no anomalies in microsatellite genotyping data (Jakše *et al.*, 2013). Young leaves from the laterally developing shoots were picked, wrapped in aluminium foil and flash frozen *in situ* in liquid nitrogen. Later they were pulverized and stored at -80°C until RNA isolation.

Total RNA was isolated from 100 mg frozen tissue pulverized in liquid nitrogen according to the protocol of Monarch Total RNA Miniprep Kit, including removal of DNA from the column with DNase I (New England Biolabs). Total RNA was quantified with Qbit 3.0, and quality was verified with the Agilent RNA Nano 6000 Kit to confirm appropriate sample RIN numbers. The total RNA samples were sent to Novogen for mRNA sequencing using Illumina's 100 bp paired end service.

The data were submitted to the SRA database of the NCBI (BioSample accession SAMN17526021).

Mapping, genotyping and SEX-DETector

The bioinformatic pipeline is schematically described in Fig. **1b**. First, the RNA-seq data were mapped to two different references: (1) the transcriptome *H. lupulus* (obtained from the annotated genome; Padgitt-Cobb *et al.*, 2019) and (2) the transcriptome assembly of *C. sativa* that we also used for our previous *C. sativa* sex chromosome analysis (Supporting Information; Van Bakel *et al.*, 2011; Prentout *et al.*, 2020). For the mapping, we ran GSNAP (version 2019-09-12; Wu and Nacu, 2010; Wu *et al.*, 2016), an aligner that enables SNP-tolerant mapping, with 10% mismatches allowed. This approach, already used for *C. sativa* analysis, was iterated several times by adding Y-specific SNPs to the references (and *H. lupulus* specific SNPs while mapping on *C. sativa* reference; see Prentout *et al.*, 2020), which increased the number of mapped reads.

Then, SAMTOOLS (version 1.4; Li *et al.*, 2009) was used to remove unmapped reads and sort mapping output files for the genotyping. We genotyped individuals with reads2snp (version 2.0.64; Gayral *et al.*, 2013), as recommended for SEX-DETector (Muyle *et al.*, 2016), *i.e.*, by accounting for allelic expression biases, without filtering for paralogous SNPs, and only conserving SNPs supported by at least three reads for subsequent analysis.

We ran the XY model of SEX-DETector on the genotyping data, using the SEM algorithm and a threshold for an assignment of 0.8. SEX-DETector computes a posterior probability of being autosomal (P_A), XY (P_{XY}) and X-hemizygous ($P_{X\text{-hemi}}$) for each SNP and for each gene (Fig. **1c**). Thus, a gene with a P_A greater than or equal to 0.8 and at least one autosomal SNP without genotyping error is classified as “autosomal”; a gene with $P_{XY} + P_{X\text{-hemi}}$ greater than or equal to 0.8 and at least one sex-linked SNP without genotyping error is classified as “sex-linked”; otherwise, the gene is classified as “lack-of-information”. Among the sex-linked genes, we classified a gene as X-hemizygous if it fulfilled one of these two criteria: (1) the gene carried only X-hemizygous SNPs and at least one SNP without genotyping error, (2) the Y expression of the gene is detected only from positions with genotyping errors. A parameter that is important to optimize with SEX-DETector is the Y specific genotyping error rate (*p*; see Muyle *et al.*, 2016). However, the quantity of Y-linked reads that map on a female reference diminishes with X-Y divergence, therefore, old and highly divergent sex chromosomes are more susceptible to mapping errors and thus genotyping errors. *p* is expected to be close to the whole transcriptome genotyping error rate (ε), but could be

higher due to weak expression (resulting in less reads) of the Y-linked copies or to mapping on a divergent X reference. To reduce the gap between these two error rates, we ran 4 iterations with GSNAP, using at each time the SNPs file generated by SEX-DETector. This SNPs file contains *H. lupulus* specific polymorphisms, initially absent from the *C. sativa* reference transcriptome, and increased the quantity of mapped reads by adding these SNPs to the reference, and thus, fitting it with the *H. lupulus* RNA-seq.

As detailed in the Supporting Information, we retained the mapping on *C. sativa* transcriptome assembly for downstream analysis. Indeed, the mapping of Y-linked reads and SEX-DETector results obtained with a mapping on *C. sativa* reference transcriptome were more robust than those obtained with a mapping on *H. lupulus* reference transcriptome (Supporting Information).

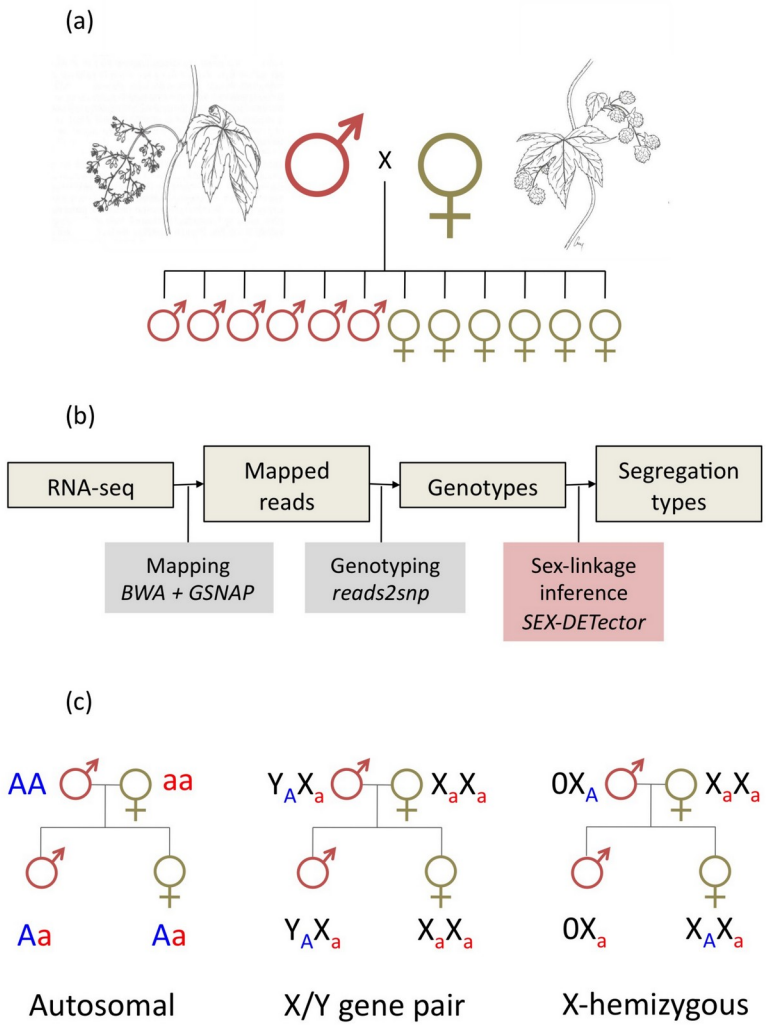

Figure 1. Schematic representation of the workflow used to detect sex-linkage. **(a)** Experimental design: six females and six males were obtained by a controlled cross, and all individuals (14) were sequenced. **(b)** Bioinformatic pipeline for the treatment of RNA-seq data. **(c)** Illustration of the underlying principles of the SEX-DETector segregation analysis.

Sex-linked gene positions on *C. sativa* genome

As the *H. lupulus* RNA-seq data were mapped on *C. sativa* transcriptome we determined the position of the transcript sequences from the *C. sativa* transcriptome assembly (van Bakel *et al*., 2011) on a chromosome-level assembly of the *C. sativa* genome (Grassa *et al.*, 2018) with blast (version 2.2.30+; Altschul *et al.*, 1990). We selected the best hit with an e-value lower than 10^{-4} to determine the position of the transcript on the genome. Then, we split each chromosome in windows of 2 Mb and computed the density of sex-linked genes and non-sex-linked genes per window using BEDTOOLS (version 2.26.0; Quinlan & Hall, 2010). Proportions of sex-linked genes were computed by dividing the number of sex-linked genes by the total number of genes (sex-linked, autosomal, and undetermined) in the same window. For *C. sativa,* densities were already available from our previous analysis (Prentout *et al.*, 2020).

Molecular clock and age of sex chromosomes

We used the translated reference transcripts (van Bakel *et al.*, 2011) to determine the X and Y Open Reading Frame (ORF) of nucleotide reference transcripts. For each XY gene pair, the *dS* values were estimated with codeml (PAML version 4.9; Yang, 2007) in pairwise mode. Then, we used two molecular clocks, derived from *Arabidopsis* species, to estimate the age of *H. lupulus* sex chromosomes (Koch *et al.*, 2000; Ossowski *et al.*, 2010). In the wild, *H. lupulus* flowers in the second or third year of development (Patzak *et al.*, 2002; Polley *et al.*, 1997), therefore, we took a generation time (GT) of 2 years, and used the molecular clocks as follows: $(dS)/rate = dS/(1.5\times10^{-8})$ using the molecular clock from Koch *et al.* (2000); $(\mathrm{GT}\times dS)/(2\times\mu) = dS/(7\times10^{-9})$ using the clock from Ossowski *et al.* (2010). Three different estimates of *dS* were used: the maximum *dS* value, the mean of the 5% highest *dS* values, and the mean of the 10% highest *dS* values.

X and Y allele-specific expression analysis

In addition to identifying X and Y alleles, SEX-DETector estimates their expression based on the number of reads (Muyle *et al.,* 2016). These estimates rely on counting reads spanning XY SNPs only and were normalized using the total read number in a library for each individual. We further normalized them by the median autosomal expression for each individual. *C. sativa* results presented here were generated in our previous analysis on *C. sativa* sex chromosomes (Prentout *et al.,* 2020).

Correction of Y read mapping bias.

The use of a female reference for the mapping of the reads might create mapping biases, resulting in the absence of Y reads in the most diverging parts of the genes. This issue may reduce the divergence detected and change the phylogenetic signal (Dixon *et al.,* 2019). If, within a same gene, regions that lack Y reads coexist with regions where the Y reads correctly mapped, we expect to see a signature similar to gene conversion, *i.e.* region-wise variation in divergence. Therefore, we ran geneconv (version 1.81a; Sawyer, 1999) in pairwise and group mode with the multiple alignments used for the phylogeny (on 85 gene alignments before Gblock filtering, see below) in order to identify and remove regions with reduced divergence. We defined two groups, one for X and Y sequences in *H. lupulus* and the other one for X and Y sequences in *C. sativa*. Then, we conserved only inner fragments and split the gene conversion regions from regions without gene conversion to obtain two subsets per gene. Thus, we obtained a subset of sequences corrected for the mapping bias, in addition to the set of genes not filtered with geneconv.

Phylogenetic analysis

We reconstructed gene families for genes identified as sex-linked in both *C. sativa* and *H. lupulus*. Then, we used blastp, filtering for the best hit (with an e-value threshold fixed at 10^{-4}), to find homologous sequences between *C. sativa* reference transcripts (the query sequence in blastp) (van Bakel *et al.,* 2011) and 4 outgroup transcriptomes (the subject sequence in blastp): *Trema orientalis* (Cannabaceae; van Velzen *et al.,* 2018), *Morus notabilis* (Moraceae; He *et al.,* 2013), *Fragaria vesca ssp. vesca* (Rosaceae; Shulaev *et al.,* 2011), and *Rosa chinensis* (Rosaceae; Raymond *et al.,* 2018). Gene families for which at least two outgroup sequences have been identified were kept, other gene families were discarded from subsequent analysis. Then, we added X and Y sequences

reconstructed by SEX-DETector to each gene family. To distinguish potential paralogous sequences or variants from alternative splicing, a blast of all sequences *vs* all sequences was realized. If two sequences from two distinct gene families matched with each other (with an e-value threshold fixed at 10^{-4}), then both families were removed from the dataset. Finally, we retrieved the corresponding nucleotide sequence of each protein, which constituted the dataset used for the phylogenetic analysis.

Using Macse (version 2.03; Ranwez *et al.*, 2011), and before alignment, non-homologous segments of at least 60 nucleotides within or 30 nucleotides at the extremity of a nucleotide sequence were trimmed if they displayed less than 30% of similarity with other sequences from the gene family. This step allowed to remove misidentified outgroup sequences. Then, gene families with no remaining outgroup sequences were discarded. Finally, remaining families were aligned with Macse, allowing sequences to be removed and realigned, one sequence at a time and over multiple iterations, to improve local alignment.

Nucleotide alignments were cleaned at the codon level using Gblocks (with default parameters) to conserve only codons shared by all sequences (version 0.91b; Castresana, 2000). For maximum-likelihood (ML) phylogenetic tree reconstruction, we used ModelFinder in IQ-TREE (version 1.639; Nguyen *et al.*, 2015; Kalyaanamoorthy *et al.*, 2017) to select the best-fit substitution model for each alignment. Those models were then used in RAxML-NG (version 1.0.0; Kozlov *et al.*, 2019) to reconstruct gene family trees. The number of bootstrap replicates was estimated using autoMRE (Pattengale *et al.*, 2010) criterion (maximum 2,000 bootstraps). The ML phylogenetic tree reconstruction was run on two datasets, one without removing potential mapping biases, and one with the potential mapping bias removed, as described above.

Bayesian phylogenies were built using Phylobayes (version 3.4; Lartillot *et al.*, 2009) with the site-specific profiles CAT and the CAT-GTR models with a gamma distribution to handle across site rate variations. Two chains were run in parallel for a minimum of 500 cycles. The convergence between the two chains was checked every 100 cycles (with a burn-in equal to one fifth of the total length of the chains). Chains were stopped once all the discrepancies were lower or equal to 0.1 and all effective sizes were larger than 50 and used to build a majority rule consensus tree.

Statistics and linear chromosome representations

The statistical analyses have been conducted with R (version 3.4.4; R Core Team, 2013). We report exact p-values when they are larger than 10^{-5}. The representation of phylogenetic topologies, *dS*

values on the first chromosome and the dosage compensation graphics have been done with ggplot2 (Wickham, 2011). For the circular representation of the sex-linked gene density along the *C. sativa* genome we used Circlize package in R (GU *et al.*, 2014). We calculated confidence intervals for the median of a dataset of *n* observations by resampling 5000 times *n* values from the dataset (with replacement). The confidence intervals are then given by the quantiles of the distribution of median values obtained by resampling.

Results

Identification of sex-linked genes in *H. lupulus*

As mentioned in the Materials and Methods section, we used the mapping of the *H. lupulus* RNA-seq data on the *C. sativa* transcriptome assembly for downstream analysis. Of the 30,074 genes in the *C. sativa* reference transcriptome, 21,268 had detectable expression in our *H. lupulus* transcriptome data. The difference of properly-paired mapped reads between males (mean: 32.3%) and females (mean: 34.9%) is slightly significant (Wilcoxon's test two-sided p-value = 0.038, see Supporting Information Table S1), which may be explained by a reduced mapping efficiency of Y-linked reads on the female reference.

The sex-linked sequences from *H. lupulus* transcriptome data were identified with SEX-DETector (Muyle *et al.*, 2016). It is important that genotyping error rate parameters ε *and* p have similar values (ε: whole transcriptome; p: Y chromosome) to obtain reliable SEX-DETector outputs. At the fourth iteration of GSNAP mapping on *C. sativa* reference transcriptome ε and p stabilized at 0.06 and 0.20, respectively (Supporting Information Table S2). Upon closer inspection, one *H. lupulus* male (#3) appeared to have many genotyping errors, as for some XY genes, this male was genotyped both heterozygous (XY) and homozygous (XX), which increased the error rate p. The identification of Y SNPs with this individual RNA-seq data discarded the hypothesis of a mislabelled female or a XX individual that developed male flowers. A particularly strong Y reads mapping bias in this male may explain these observations. After removal of this male, the error rate p dropped to 0.10 (Supporting Information Table S2). A total of 265 sex-linked genes were identified in *H. lupulus*, which represents 7.8% of all assigned genes (autosomal genes + sex-linked genes; Table 1).

Table 1. Summary of the SEX-DETector results.

	Number
All genes*	30,074
Expressed genes	21,268
Genes with SNPs used by SEX-DETector	4,472
Genes with undetermined segregation type class 1 **	462
Genes with undetermined segregation type class 2 ***	354
Autosomal genes	3391
Sex-linked genes	265
XY genes	265
X-hemizygous genes	0

*transcripts from gene annotation of the *C. sativa* reference genome (van Bakel *et al.*, 2011).

** Posterior probabilities < 0.8

***Posterior probabilities > 0.8 but absence of SNPs without error.

H. lupulus and _C. sativa_ sex chromosomes are homologous

Among 265 *H. lupulus* XY genes from the *C. sativa* transcriptome assembly (van Bakel *et al.*, 2011), 254 genes are present on the *C. sativa* chromosome-level genome assembly (Grassa *et al.*, 2018). As shown in Fig. **2a**, 192 of these genes (75.6%) map on *C. sativa* chromosome number 1, which is the chromosome we previously identified as the X chromosome in *C. sativa* (Prentout *et al.*, 2020). Of the 265 sex-linked genes in *H. lupulus*, 112 were also detected as sex-linked in *C. sativa*, while 64 were detected as autosomal and 89 had unassigned segregation type (Prentout *et al*., 2020).

The synonymous divergence (*dS*) between X and Y copies of the sex-linked genes of *H. lupulus* is distributed similarly along the *C. sativa* sex chromosome as the values for this latter species, as shown in Fig. **2b**. While the sampling variation of these *dS* values is large, as expected (cf Takahata & Nei, 1985), it can be observed that the larger values occur in the region beyond 65 Mb.

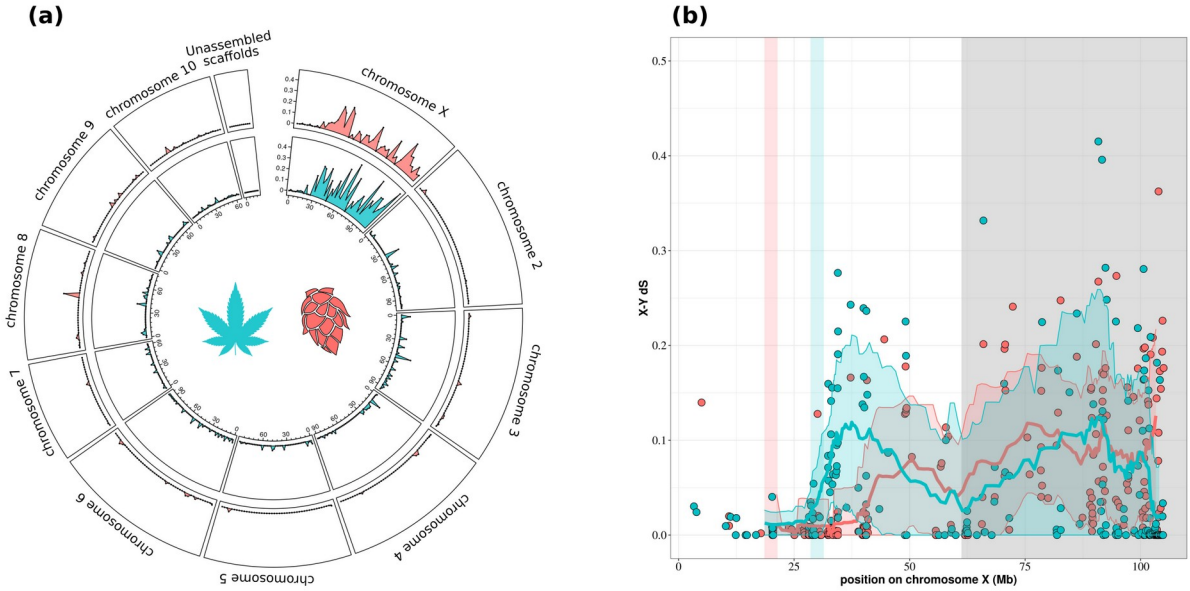

Figure 2. **(a)** *H. lupulus* sex-linked genes mapped on the *C. sativa* genome (Grassa *et al*., 2018). Inner graphs (in blue): *C. sativa* sex-linked gene density corrected by the total gene density in 2-Mb windows (from Prentout *et al.*, 2020). Outer graphs (in red): *H. lupulus* sex-linked gene density corrected by the total gene density in 2-Mb windows. Chromosome positions are given in Megabases. **(b)** Synonymous divergence (*dS*) between X and Y copies of *H. lupulus* sex-linked genes (red) and those of *C. sativa* (blue) along the X chromosome of *C. sativa*. The curves represent the average *dS* with sliding windows (windows of 20 points), for *H. lupulus* (red) and *C. sativa* (blue). Confidence intervals (average ± standard deviation) are indicated around the *H. lupulus* curve (red area) and the *C. sativa* curve (blue area). The vertical red bar represents the putative Pseudo-Autosomal Boundary (PAB) in *H. lupulus*, the vertical blue bar represents the putative PAB in *C. sativa*, the grey area represents the region that stopped recombining in a common ancestor.

X-Y recombination likely stopped before the *Cannabis* and *Humulus* genera split

We reconstructed phylogenetic trees of genes detected as sex-linked in both species, including outgroup sequences from the order Rosales. For 27 out of the 112 sex-linked genes present in both species, we could not identify any homologous sequences in the outgroup species and those genes were excluded from further analysis. For the remaining 85, we determined the topology of the gametologous sequences in the Cannabaceae, considering a node as well resolved when the bootstrap support exceeded 95%, or Bayesian support exceeded 0.95.

The three different methods for phylogenetic reconstruction provided consistent phylogenies (Table 2). More precisely, we observed three major topologies, as shown in Figure 3: X copies of both species form a clade separated from a clade of Y sequences (topology I, Fig. **3a**), the X and Y sequences of each species group together (topology II, Fig. **3b**), or a paraphyletic placement of the X and Y sequences of *H. lupulus*, relative to *C. sativa* sequences (topology III, Fig. **3c**). As shown in Table 2, we found that most genes had topology II, corresponding to recombination suppression after the split of the genera. A few genes, however, had topology I, which corresponds to genes for which recombination was already suppressed in a common ancestor of both species. As shown in Fig. **3d**, topologies I and III occurred mainly beyond 80 Mb, while topology II occurred all over the chromosome. Topology I is associated with higher synonymous divergence.

We identified 42 genes, out of the 85 genes used for the phylogeny, with at least one fragment in at least one species that displayed reduced divergence (with a p-value < 0.05 in geneconv output). Because this reduction of divergence may be caused by a mapping bias of Y reads, we ran the ML phylogenetic reconstruction method on regions with and without mapping bias (example in Supporting Information Fig. S7). As shown in Table 2 and Fig. **3e**, after mapping bias filter with geneconv, the proportion of genes displaying topology I, indicating recombination suppression in a common ancestor, increased, while less genes with topology II were mainly found in a restricted region corresponding to the region where recombination stopped independently in both species.

Table 2. Results of the phylogenetic reconstruction of sex-linked genes. Phylogenetic trees with a bootstrap value equal or greater than 95% (and posterior probabilities higher than 0.95 for Bayesian reconstructions) at the node separating *C. sativa* and *H. lupulus*, or Y and X sequences, are presented in the first four columns. Phylogenetic trees without such support are classified as "unresolved".

	Topology I $((X_c,X_h),(Y_c,Y_h))$	Topology II $((X_c,Y_c),(X_h,Y_h))$	Topology III $(Y_h,(X_h,(X_c,Y_c)))$	Other	Un-resolved	Total
Maximum Likelihood (ML)	7	44	7	1	26	85
GTR (bayesian)	4	45	4	8	24	85
CAT-GTR (bayesian)	4	44	7	7	23	85
ML after geneconv filtering	11	27	11	4	32	85

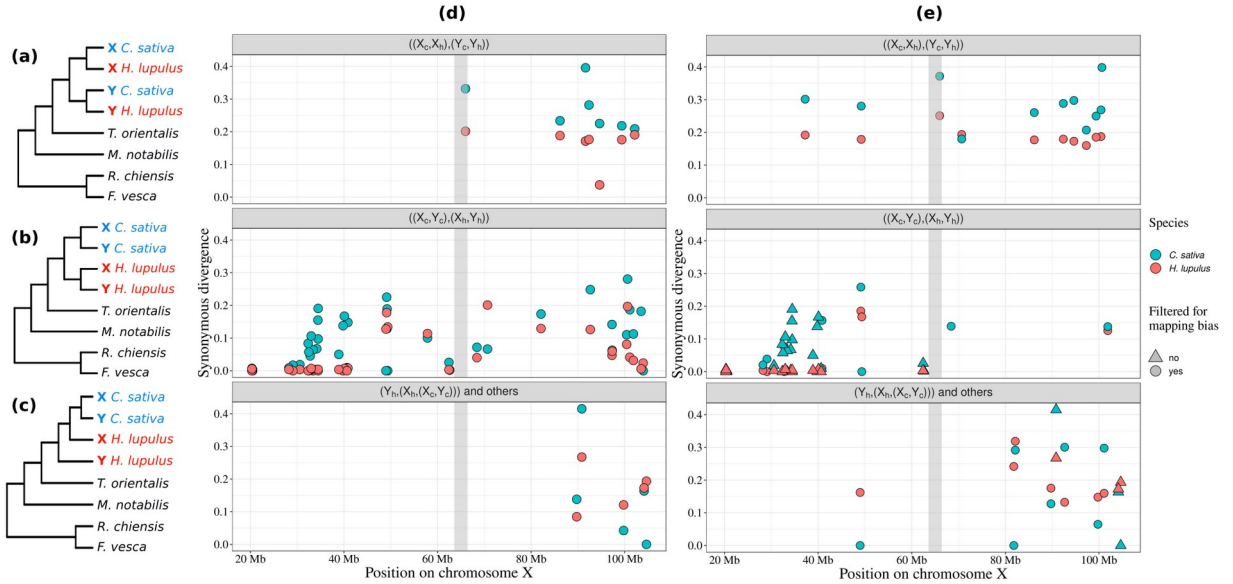

Figure 3. Distribution of the three topologies of the sex-linked genes on the X chromosome: **(a)** Topology I, XX-YY – arrest of recombination before the split of the two genera, **(b)** Topology II, XY-XY – arrest of recombination after the split of the two genera, **(c)** Topology III, Y-X-XY – *H. lupulus* X chromosome is closer to *C. sativa* sequences than its Y counterpart. **(d)** Distribution of the topologies along the *C. sativa* X chromosome ("other" topology is included in the Y-X-XY topology panel), using the full gene sequences. For each gene, dots represent the *dS* values in *C. sativa* (blue) and *H. lupulus* (red). **(e)** Distribution of the topologies after filtering out

possible mapping biases through geneconv. Triangles indicate that at least one segment was removed, dots indicate sequences for which no mapping bias was detected.
The vertical grey bar (panels **(d)** and **(e)**) represents the putative boundary between the region that stopped recombining in a common ancestor and the region that stopped recombining independently in the two species.

This leads us to define three regions on the X chromosomes of *C. sativa* and *H. lupulus* (with the *C. sativa* X chromosome as a reference). A region from ~65Mb to the end of the X chromosome that already stopped recombining in a common ancestor; from ~20-30Mb to ~65Mb, a part of the non-recombining region that evolved independently in the two species; and from the beginning of the chromosome to ~20-30Mb, the pseudo-autosomal region (PAR), where few sex-linked genes are found.

Age of *H. lupulus* sex chromosomes

To estimate the age of the sex chromosomes, we used the maximum synonymous divergence between X and Y sequences and two molecular clocks, which were both derived from *Arabidopsis*. Because the sampling variance in *dS* values can be large, we used three ways to calculate the maximum *dS* value: the single highest *dS* value; the average of the 5% highest values; and the average of the 10% highest values. Furthermore, we calculated these on the raw alignments as well as the alignments with possible mapping biases removed. The different estimates are given in Table 3, and yield values between 14.5 and 51.4 My. Minimum synonymous divergence between *C. sativa* and outgroup species *Morus notabilis* and *Rosa chinensis* is ~0.45 and ~0.65, respectively (Supporting Information Fig. S5 and Fig. S6), higher than the maximum synonymous divergence between sex-linked gene copies, indicating that the sex chromosomes probably evolved in the Cannabaceae family.

Table 3. Age estimates (in millions of years, My) with two molecular clocks and different maximum *dS* values. For each *dS* value, two ages were obtained using the molecular clocks of [1]Ossowski *et al.* (2010) and [2]Koch *et al.* (2000). Two alignment datasets were used, with or without filtering for possible mapping bias.

	No filtering			Mapping bias filtering		
	dS	age (My)[1]	age (My)[2]	*dS*	age (My)[1]	age (My)[2]
Highest *dS*	0.362	51.4	24.0	0.362	51.4	24.0
Mean highest 5%	0.249	35.6	16.6	0.274	39.1	18.3
Mean highest 10%	0.217	31.0	14.5	0.214	34.4	16.1

Y gene expression

The Y over X expression ratio is a standard proxy for the degeneration of the Y chromosome. A Y/X expression ratio close to 1 means no degeneration, a Y/X expression ratio close to 0.5 or below means strong degeneration. In *H. lupulus*, the median Y/X expression ratio is equal to 0.637 (Supporting Information Fig. S1), which is significantly different from 1 (99th percentile of median distribution with 5,000 samples in initial distribution = 0.673, see methods). The median is not different when considering all sex-linked genes (0.637) or only the sex-linked genes mapping on *C. sativa* X chromosome (0.639, p-value = 0.70, one-sided Wilcoxon rank sum test).

In both species, the reduced Y expression is correlated to the position on the X chromosome (linear regression: adjusted $R^2 = 0.134$, p-value $< 10^{-5}$ for *H. lupulus*; and adjusted $R^2 = 0.278$, p-value $< 10^{-5}$ for *C. sativa*). As shown in Figure 4, the Y/X expression ratio decreases while moving away from the PAR in *H. lupulus*, and this is also confirmed in *C. sativa.*

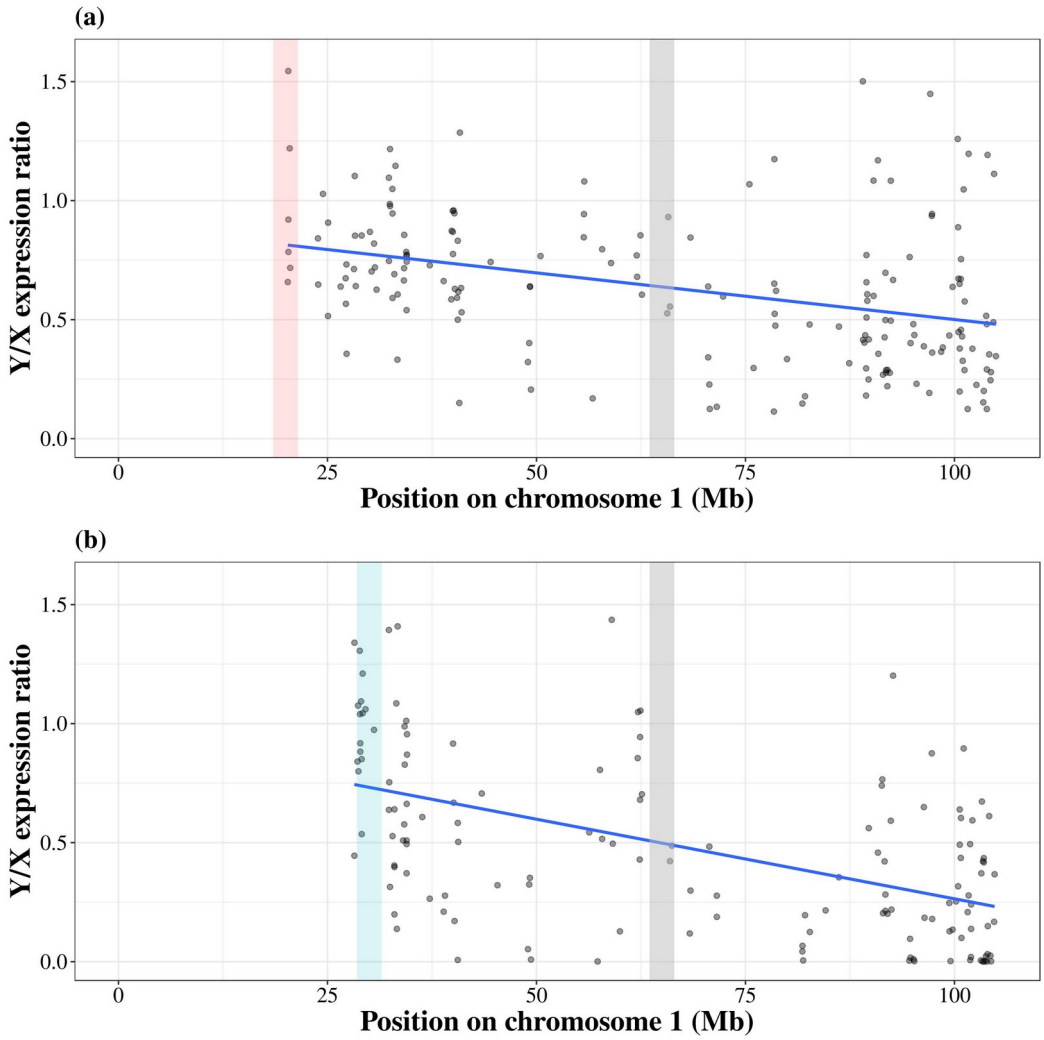

Figure 4. Y/X expression ratio along the X chromosome in *H. lupulus* **(a)**, and *C. sativa* **(b)**. Each dot is the Y/X expression ratio for one gene in the non-recombining region only (the linear regression result is indicated by the blue line). The vertical red bar represents the putative PAB in *H. lupulus*, the vertical blue bar represents the putative PAB in *C. sativa*, the vertical grey bar represents the putative boundary between the region that stopped recombining in a common ancestor and the region that stopped recombining independently in the two species.

Dosage compensation

We tested whether the expression of the X chromosome changed following degeneration of the Y chromosome, a phenomenon called dosage compensation (Muyle *et al.*, 2017). We used the ratio of the male X expression over the female XX expression as a proxy for dosage compensation (Muyle *et al.*, 2012) and Y/X expression ratio as a proxy for Y degeneration. Genes with strong degeneration (Y/X expression ratio close to zero) display an increased expression of the X in males

(linear regression: adjusted R^2=0.179, p-value < 10^{-5} and adjusted R^2=0.097, p-value < 10^{-5} for *H. lupulus* and *C. sativa* respectively), as shown in Figure 5. A dosage compensation pattern was found in both in *H. lupulus* and *C. sativa* in agreement with previous work (Prentout *et al.*, 2020).

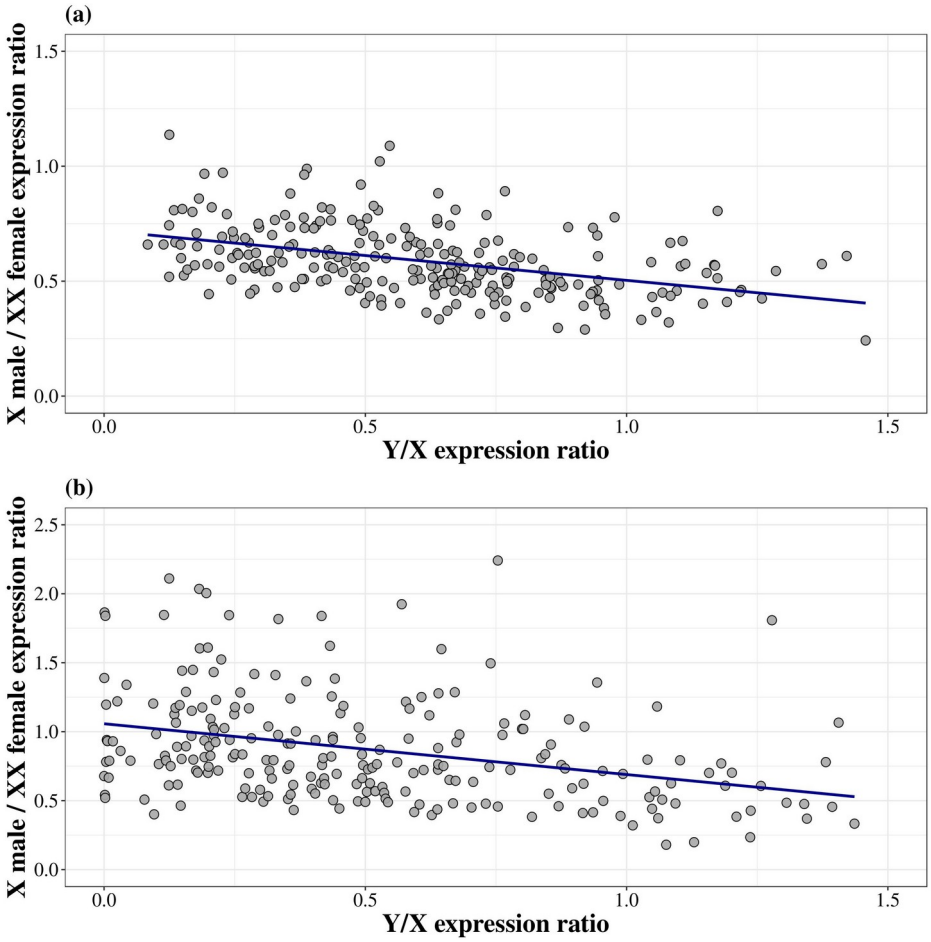

Figure 5. The male X expression over female XX expression versus Y/X expression ratio for *H. lupulus* **(a)** and *C. sativa* **(b)**. Each black dot represents one gene. The blue line represents a linear regression.

Discussion

We here identified the *H. lupulus* sex chromosomes, and found that they are homologous to those of *C. sativa* (Prentout *et al.*, 2020), and that a part of these chromosomes had already stopped recombining in a common ancestor of the two species. Performing a segregation analysis with SEX-DETector (Muyle *et al.*, 2016), we identified 265 XY genes in *H. lupulus*, among which 112 are also sex-linked in *C. sativa.* Mapping these genes on the chromosome-level assembly of *C. sativa* (Grassa *et al.*, 2018) suggested that the non-recombining region is large in *H. lupulus*, as proposed before, based on cytological studies (Divashuk *et al.*, 2011).

We identified three different regions on the sex chromosome, based on the distribution of sex-linked gene phylogenetic topologies and synonymous divergence between the X and Y copies on the *C. sativa* X chromosome: one region that had already stopped recombining in a common ancestor of *C. sativa* and *H. lupulus*, a region that independently stopped recombining in both species, and the pseudo-autosomal region. Our results suggest the pseudo-autosomal boundary (PAB) in *H. lupulus* may be located around position 20Mb, whereas we estimated a PAB around 30Mb in *C. sativa* (Prentout *et al.*, 2020); the non-recombining region may thus be larger in *H. lupulus* than in *C. sativa*. With this estimation of the size of the non-recombining region in *H. lupulus*, among the 3469 genes present on the X chromosome, 2045 genes would be located in this non-recombining region (which represents 59.1% of all the genes on the X chromosome). However, a chromosome-level assembly of the *H. lupulus* genome would be needed to determine the exact position of the PAB in this species, as synteny might not be fully conserved. In addition, because we used one single cross, it is possible that we overestimated the size of the non-recombining region due to linkage disequilibrium. Thus, genes around the PAB classified as sex-linked and for which we estimated a low *dS* value may still be recombining. An accurate estimation of the PAB, as has been done for example in *S. latifolia*, would require much more offspring and data from several populations (Krasovec *et al.*, 2020).

Several sex-linked genes had topologies that were not compatible with either recombination suppression in a common ancestor or in each of the species independently. Strikingly, most of these topologies placed the *H. lupulus* Y sequence as an outgroup to the other sex-linked gene sequences. Whether this is the result of errors (*e.g.* long branch attraction, mapping biases) remains to be investigated. Interestingly, genes with theses “unexpected” topologies all clustered (except for one gene) in a region of ~25Mb. This region is located at the extremity of the X chromosome, which, as

we suggested, stopped recombining first. It is likely to observe a high rate of unexpected phylogenetic results in the region that stopped recombining first since the X-Y divergence should be the highest in this region, which could increase the mapping bias. Our approach to correct for the Y read mapping relies on geneconv, which is known to have a high rate of false negatives (Lawson & Zhang, 2009). This could also explain the unexpected presence of some of the XY-XY genes in the older region.

X-Y gene conversion has been shown to affect only a few genes in animals (Katsura *et al.*, 2012; Trombetta *et al.*, 2014; Peneder *et al.*, 2017). Although we don't expect gene conversion for half of the genes that are sex-linked in both species, it is worth noting that a part of fragments identified by geneconv may correspond to real gene conversion rather than mapping biases. Here again, assemblies of Y and X chromosomes in both species are required to determine the presence of true X-Y gene conversion.

The highest *dS* values and the genes with a topology indicating that recombination was already suppressed in the common ancestor are located in the same region (65Mb to the end of the chromosome). These results suggest the presence of at least two strata in these sex chromosomes. We estimated that the youngest stratum in *H. lupulus* is 10.1-29.4 Myo, and is 15.9-19.8 Myo in *C. sativa* (Supporting Information Table S4). However, while recombination suppression clearly did not occur for all the sex-linked genes at the same time, we cannot determine the exact number of strata in the sex chromosomes of *C. sativa* and *H. lupulus*. It is also possible that recombination was suppressed gradually, with the recombination suppression starting before the split of both genera and continuing afterwards. To clearly determine the number of strata, an identification of chromosomal inversions or significative differences in *dS* values along the sex chromosomes are required (Nicolas *et al.*, 2004; Lemaitre *et al.*, 2009; Wang *et al.*, 2012; reviewed in Wright *et al.*, 2016). Thus, X and Y chromosome assemblies for both *H. lupulus* and *C. sativa* are needed to exactly determine the number (and location) of strata in both species. Moreover, a Y chromosome assembly will allow the identification of Y-specific genes, which is not possible with SEX-DETector and the data we used.

We did not find X-hemizygous genes in *H. lupulus*. This is striking as 218 X-hemizygous genes (38% of all sex-linked genes) were found in *C. sativa* using the same methodology (Prentout *et al.*, 2020). A very low level of polymorphism could result in the inability of SEX-DETector to identify X-hemizygous genes (Muyle *et al.*, 2016), but in that case SEX-DETector should also have problems identifying autosomal genes, which was not the case here. Non-random X-inactivation in

females could be an explanation, as the non-random expression of a single X allele in females would impede SEX-DETector to identify X-linkage and X-hemizygous genes (Muyle *et al.*, 2016). We ran an Allele-Specific Expression (ASE) analysis, which doesn't support this hypothesis (Supporting Information Fig. S2, Fig. S3, Fig. S4). *H. lupulus* is probably an ancient polyploid that reverted to the ancestral karyotype (Padgitt-Cobb *et al.*, 2019). It is thus possible, that the *H. lupulus* X chromosome is made of two copies of the ancestral X as some cytological data seem to suggest (Divashuk et al., 2011). In this case, SEX-DETector would manage to identify the XY gene pairs, but would fail to identify the X-hemizygous genes as these genes would exhibit unexpected allele transmission patterns (Supporting Information Fig. S10).

H. lupulus is a rare case of XY systems in plants in which the Y is smaller than the X (cf Ming *et al.*, 2011). In *C. sativa*, both sex chromosomes have similar sizes (Divashuk *et al.*, 2014). If the size difference is caused by deletions of parts of the *H. lupulus* Y chromosome, which is the hypothesized mechanism in many species (cf Ming *et al.*, 2011), we expect to observe that many XY gene pairs in *C. sativa* have missing Y copies in *H. lupulus*. As explained above, we did not detect any X-hemizygous genes. Furthermore, the XY gene pairs of *H. lupulus* are distributed uniformly on the *C. sativa* X chromosome, and no region appeared to be depleted in XY genes, which is not what we would observe if large deletions were present on the *H. lupulus* Y chromosome. The sex chromosome size differences observed in *H. lupulus* probably reflect complex dynamics, different from that of old animal systems with tiny Y chromosome due to large deletions (*e.g.* Skaletsky *et al.*, 2003; Ross *et al.*, 2005). The large size of the X chromosome in *H. lupulus* may be due to a full-chromosome duplication followed by a fusion (see above), whereas the Y chromosome has remained unchanged. Assemblies of the *H. lupulus* sex chromosomes will be needed to test these hypotheses.

Our estimates of the age of the *H. lupulus* sex chromosomes are larger than the estimates for *C. sativa*, although we found very similar X-Y maximum divergence in both species (higher bound age estimates are ~50My and ~28My; highest *dS* values are 0.362 and 0.415 in *H. lupulus* and *C. sativa* respectively, see Prentout *et al.*, 2020). Of course, the molecular clocks that we used are known to provide very rough estimates as they derive from the relatively distant *Arabidopsis* genus and are sensitive to potential differences in mutation rates between the annual *C. sativa* and the perennial *H. lupulus* (Neve, 1991; Petit & Hampe, 2006; Small, 2015; but see Krasovec *et al.*, 2018). Indeed, only one of these molecular clocks (which is based on the mutation rate) takes into account the generation time, (two years in *H. lupulus vs.* one year in *C. sativa*). This produced age estimates

approximately twice as high as the other clock (based on the substitution rate), which was not the case with *C. sativa* (Prentout *et al.*, 2020). It is not known, however, if the generation time influences the substitution rate (Petit and Hampe, 2006). Furthermore, the short generation time in *C. sativa* is probably a derived trait, not reflecting the long-term generation time of the *Cannabis-Humulus* lineage, as the *Cannabis* genus is the only herbaceous genus in the Cannabaceae family (Yang *et al.*, 2013). Thus, the remarkable similarity between the highest dS values in both species indicates that the *C. sativa* and *H. lupulus* sex chromosomes have a similar age, as expected if they derive from the same common ancestor. Although it's not possible to estimate their age exactly with the current data, initial recombination suppression at least predates the split between the genera, that occurred between 21 and 25 My ago (Divashuk *et al.*, 2014; Jin *et al.*, 2020), and might even be 50 My old. We thus confirmed here that the XY system shared by *C. sativa* and *H. lupulus* is among the oldest plant sex chromosome systems documented so far (Prentout *et al.*, 2020).

Dioecy was inferred as the ancestral sexual system for the Cannabaceae, Urticaceae and Moraceae (Zhang *et al.*, 2018; note however that many monoecious Cannabaceae were not included). We found that the synonymous divergence between the Cannabaceae species and *Morus notabilis* was about 0.45, higher than the maximum divergence of the X and Y copies in the Cannabaceae. It remains possible that the sex chromosomes evolved before the split of the Cannabaceae and Moraceae families, because the oldest genes might have been lost or were not detected in our transcriptome data. There is however no report of whether or not sex chromosomes exist in Urticaceae and Moraceae (Ming *et al.*, 2011).

To estimate the Y expression, we counted the number of reads with Y SNPs. Therefore, the impact of a potential Y reads mapping bias should be weaker on Y expression analysis than on X-Y divergence analysis. We validated this assumption by removing genes with detected mapping bias from the analysis, which didn’t change the signal of Y expression reduction and dosage compensation (Supporting Information Fig. S8, Fig. S9). Dosage compensation is a well-known phenomenon in animals (*e.g.* Gu & Walters, 2017). It has only been documented quite recently in plants (reviewed in Muyle *et al.*, 2017). Here we found evidence for dosage compensation in *H. lupulus*. This is not surprising as previous work reported dosage compensation in *C. sativa* and we showed here that both systems are homologous. *C. sativa* and *H. lupulus* add up to the list of plant sex chromosome systems with dosage compensation (see Muyle *et al.*, 2017 for a review and Prentout *et al.*, 2020; Fruchard *et al.*, 2020 for the latest reports of dosage compensation in plants).

Further analyses are needed to determine whether this dosage compensation has been selected or is an outcome of regulatory feedback (Malone *et al*., 2012; Krasovec *et al.*, 2019).

H. lupulus sex chromosomes, as those of *C. sativa*, are well-differentiated, with a large non-recombining region. Both species show similar patterns of Y degeneration and dosage compensation, despite the fact that a large part of the non-recombining region evolved independently in both species. These similarities, as well as the age of the chromosomes and the fact that they have been conserved since the most recent common ancestor of the two genera, a unique situation in plants so far, provide an exciting opportunity to test and elaborate hypotheses on sex chromosome evolution in plants.

673 # Acknowledgments

674 We thank Roberto Bacilieri for his help in setting up this collaboration and for discussions, Aline
675 Muyle for advice on SEX-DETector and Florian Bénitière for helpful suggestions regarding
676 graphical representations. This work was performed using the computing facilities of the CC LBBE/
677 PRABI; we thank Bruno Spataro and Stéphane Delmotte for cluster maintenance. Virtual machines
678 from the Institut Français de Bioinformatique were also used to perform this work. This work
679 received financial support from P4-0077 grant by ARRS (Slovenian Research Agency) to JJ and
680 from an ANR grant (ANR-14-CE19-0021-01) to GABM.
681
682

683 # Author Contribution

684 Conceptualization of the study: G.A.B.M., J.K. and D.P.; methodology: G.A.B.M., J.K., D.P., N.S.
685 and J.J.; software: D.P., T.T. and C.B.A.; formal analysis: D.P., T.T. and C.B.A.; investigation:
686 D.P., N.S., T.T., C.B.A., J.J., J.K., and G.A.B.M.; resources: A.C., N.S. and J.J.; writing—original
687 draft: D.P., G.A.B.M., J.K. and T.T.; writing—review and editing: all authors; visualization: D.P.
688 and T.T.; supervision: G.A.B.M., J.K.; project administration: G.A.B.M.; funding acquisition: N.S.,
689 J.J. and G.A.B.M.
690

691 # Data Availability

692 The sequence data were deposited under the Bioproject with accession number PRJNA694508,
693 BioSample SAMN17526021 (SRR13528971 ; SRR13528970; SRR13528969; SRR13528968;
694 SRR13528966; SRR13528965; SRR13528964; SRR13528967; SRR13528963; SRR13528962;
695 SRR13528961; SRR13528960; SRR13528959; SRR13528958)
696

697 # References

698 • **Regular research articles:**

699 **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. **1990**. Basic local alignment search
700 tool. *Journal of Molecular Biology* **215**: 403–410.
701

Bačovský V, Čegan R, Šimoníková D, Hřibová E, Hobza R. 2020. The Formation of Sex Chromosomes in Silene latifolia and S. dioica Was Accompanied by Multiple Chromosomal Rearrangements. *Frontiers in Plant Science* 11.

Badouin H, Velt A, Gindraud F, Flutre T, Dumas V, Vautrin S, Marande W, Corbi J, Sallet E, Ganofsky J, *et al.* 2020. The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. *Genome Biology* **21**: 223.

van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE. **2011**. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology* **12**: R102.

Baránková S, Pascual-Díaz JP, Sultana N, Alonso-Lifante MP, Balant M, Barros K, D'Ambrosio U, Malinská H, Peska V, Lorenzo IP, *et al.* 2020. Sex-chrom, a database on plant sex chromosomes. *New Phytologist* **227**: 1594–1604.

Bergero R, Charlesworth D. **2009**. The evolution of restricted recombination in sex chromosomes. *Trends in Ecology & Evolution* **24**: 94–102.

Castresana J. **2000**. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution* **17**: 540–552.

Čerenak A, Kolenc Z, Sehur P, Whittock SP, Koutoulis A, Beatson R, Buck E, Javornik B, Škof S, Jakše J. **2019**. New Male Specific Markers for Hop and Application in Breeding Program. *Scientific Reports* **9**: 14223.

Charlesworth B, & Charlesworth D. **2000**. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **355**: 1563–1572.

Charlesworth D. **2016**. Plant Sex Chromosomes. *Annual Review of Plant Biology* **67**: 397–420.

Cherif E, Zehdi-Azouzi S, Crabos A, Castillo K, Chabrillange N, Pintaud J-C, Salhi-Hannachi A, Glémin S, Aberlenc-Bertossi F. **2016**. Evolution of sex chromosomes prior to speciation in the dioecious *Phoenix* species. *Journal of Evolutionary Biology* **29**: 1513–1522.

Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grützner F, Kaessmann H. **2014**. Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**: 488–493.

Divashuk MG, Alexandrov OS, Kroupin PY, Karlov GI. **2011**. Molecular Cytogenetic Mapping of *Humulus lupulus* Sex Chromosomes. *Cytogenetic and Genome Research* **134**: 213–219.

Divashuk MG, Alexandrov OS, Razumova OV, Kirov IV, Karlov GI. **2014**. Molecular Cytogenetic Characterization of the Dioecious *Cannabis sativa* with an XY Chromosome Sex Determination System. *PLOS ONE* **9**: e85118.

Dixon G, Kitano J, Kirkpatrick M. **2019**. The Origin of a New Sex Chromosome by Introgression between Two Stickleback Fishes. *Molecular Biology and Evolution* **36**: 28–38.

Fridolfsson A-K, Cheng H, Copeland NG, Jenkins NA, Liu H-C, Raudsepp T, Woodage T, Chowdhary B, Halverson J, Ellegren H. **1998**. Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proceedings of the National Academy of Sciences* **95**: 8147–8152.

Fruchard C, Badouin H, Latrasse D, Devani RS, Muyle A, Rhoné B, Renner SS, Banerjee AK, Bendahmane A, Marais GAB. **2020**. Evidence for Dosage Compensation in *Coccinia grandis*, a Plant with a Highly Heteromorphic XY System. *Genes* **11**: 787.

Gayral P, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, Lourenco JM, Alves PC, Ballenghien M, Faivre N, *et al.* 2013. Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate–Invertebrate Gap. *PLOS Genetics* **9**: e1003457.

Gu L, Walters JR. **2017**. Evolution of Sex Chromosome Dosage Compensation in Animals: A Beautiful Theory, Undermined by Facts and Bedeviled by Details (K Makova, Ed.). *Genome Biology and Evolution* **9**: 2461–2476.

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics*, ***30*(19)**: 2811-2812.

Harkess A, Zhou J, Xu C, Bowers JE, Van der Hulst R, Ayyampalayam S, Mercati F, Riccardi P, McKain MR, Kakrana A, *et al.* 2017. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nature Communications* **8**: 1279.

He N, Zhang C, Qi X, Zhao S, Tao Y, Yang G, Lee T-H, Wang X, Cai Q, Li D, *et al.* 2013. Draft genome sequence of the mulberry tree *Morus notabilis*. *Nature Communications* **4**: 2445.

Jakse J, Cerenak A, Radisek S, Satovic Z, Luthar Z, Javornik B. 2013. Identification of quantitative trait loci for resistance to Verticillium wilt and yield parameters in hop (*Humulus lupulus L.*). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **126**: 1431–1443.

Jin J-J, Yang M-Q, Fritsch PW, Velzen R van, Li D-Z, Yi T-S. 2020. Born migrators: Historical biogeography of the cosmopolitan family Cannabaceae. *Journal of Systematics and Evolution* **58**: 461–473.

Käfer J, Marais GAB, Pannell JR. 2017. On the rarity of dioecy in flowering plants. *Molecular Ecology* **26**: 1225–1241.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Karlov GI, Danilova TV, Horlemann C, Weber G. 2003. Molecular cytogenetics in hop (*Humulus lupulus L.*) and identification of sex chromosomes by DAPI-banding. *Euphytica* **132**: 185–190.

Katsura Y, Iwase M, Satta Y. 2012. Evolution of Genomic Structures on Mammalian Sex Chromosomes. *Current Genomics* **13**: 115–123.

Kejnovsky E, Vyskot B. 2010. Silene latifolia: The Classical Model to Study Heteromorphic Sex Chromosomes. *Cytogenetic and Genome Research* 129: 250–262.

Koch MA, Haubold B, Mitchell-Olds T. 2000. Comparative Evolutionary Analysis of Chalcone Synthase and Alcohol Dehydrogenase Loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Molecular Biology and Evolution* **17**: 1483–1498.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. **2019**. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Krasovec M, Chester M, Ridout K, Filatov DA. **2018**. The Mutation Rate and the Age of the Sex Chromosomes in *Silene latifolia*. *Current Biology* **28**: 1832-1838.e4.

Krasovec M, Kazama Y, Ishii K, Abe T, Filatov DA. **2019**. Immediate Dosage Compensation Is Triggered by the Deletion of Y-Linked Genes in Silene latifolia. *Current Biology* **29**: 2214-2221.e4.

Krasovec M, Zhang Y, Filatov DA. 2020. The Location of the Pseudoautosomal Boundary in Silene latifolia. *Genes* 11: 610.

Lartillot N, Lepage T, Blanquart S. **2009**. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**: 2286–2288.

Lawson MJ, Zhang L. **2009**. Sexy gene conversions: locating gene conversions on the X-chromosome. *Nucleic Acids Research* **37**: 4570–4579.

Lemaitre C, Braga MDV, Gautier C, Sagot M-F, Tannier E, Marais GAB. 2009. Footprints of Inversions at Present and Past Pseudoautosomal Boundaries in Human Sex Chromosomes**.** *Genome Biology and Evolution* **1:** 56–66.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. **2009**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Mackinnon D, Pavlovič M. **2019**. Global Hop Market Analysis Within the International Hop Growers' Convention. *GLOBALNA ANALIZA HMELJSKEGA TRGA V OKVIRU SVETOVNE HMELJARSKE ORGANIZACIJE.*: 99–108.

Malone JH, Cho D-Y, Mattiuzzo NR, Artieri CG, Jiang L, Dale RK, Smith HE, McDaniel J, Munro S, Salit M, *et al.* 2012. Mediation of Drosophila autosomal dosage effects and compensation by network interactions. *Genome Biology* 13: R28.

Massonnet M, Cochetel N, Minio A, Vondras AM, Lin J, Muyle A, Garcia JF, Zhou Y, Delledonne M, Riaz S, *et al.* 2020. The genetic basis of sex determination in grapes. *Nature Communications* **11**: 2902.

Ming R, Bendahmane A, Renner SS. **2011**. Sex Chromosomes in Land Plants. *Annual Review of Plant Biology* **62**: 485–514.

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. **2012**. Rapid De Novo Evolution of X Chromosome Dosage Compensation in *Silene latifolia*, a Plant with Young Sex Chromosomes. *PLOS Biology* **10**: e1001308.

Muyle A, Käfer J, Zemp N, Mousset S, Picard F, Marais GA. **2016**. SEX-DETector: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. *Genome Biology and Evolution* **8**: 2530–2543.

Muyle A, Shearn R, Marais GA. **2017**. The Evolution of Sex Chromosomes and Dosage Compensation in Plants. *Genome Biology and Evolution* **9**: 627–645.

Natri HM, Shikano T, Merilä J. 2013. Progressive Recombination Suppression and Differentiation in Recently Evolved Neo-sex Chromosomes. *Molecular Biology and Evolution* 30: 1131–1144.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. **2015**. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Negrutiu I, Charlesworth D, Monéger F. 2004. A Gradual Process of Recombination Restriction in the Evolutionary History of the Sex Chromosomes in Dioecious Plants. *PLOS Biology* **3**: e4.

Ohno S. **1969**. Evolution of Sex Chromosomes in Mammals. *Annual Review of Genetics* **3**: 495–524.

Okada Y, Ito K. **2001**. Cloning and Analysis of Valerophenone Synthase Gene Expressed Specifically in Lupulin Gland of Hop (*Humulus lupulus L.*). *Bioscience, Biotechnology, and Biochemistry* **65**: 150–155.

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. **2010**. The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. *Science* **327**: 92–94.

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. **2010**. How Many Bootstrap Replicates Are Necessary? *Journal of Computational Biology* **17**: 337–354.

Patzak J, Nesvadba V (Chmelarsky I, Vejl P, Skupinova S. **2002**. Identification of sex in F1 progenies of hop (*Humulus lupulus*) by molecular marker. *Rostlinna Vyroba - UZPI (Czech Republic)*.

Peil A, Flachowsky H, Schumann E, Weber WE. **2003**. Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (*Cannabis sativa L.*). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **107**: 102–109.

Peneder P, Wallner B, Vogl C. 2017. Exchange of genetic information between therian X and Y chromosome gametologs in old evolutionary strata. *Ecology and Evolution* **7**: 8478–8487.

Petit RJ, Hampe A. **2006**. Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics* **37**: 187–214.

Polley A, Ganal MW, Seigner E. **2011**. Identification of sex in hop (*Humulus lupulus*) using molecular markers. *Genome*.

Prentout D, Razumova O, Rhoné B, Badouin H, Henri H, Feng C, Käfer J, Karlov G, Marais GAB. **2020**. An efficient RNA-seq-based segregation analysis identifies the sex chromosomes of *Cannabis sativa*. *Genome Research* **30**: 164–172.

Pucholt P, Wright AE, Conze LL, Mank JE, Berlin S. **2017**. Recent Sex Chromosome Divergence despite Ancient Dioecy in the Willow *Salix viminalis*. *Molecular Biology and Evolution* **34**: 1991–2001.

Quinlan AR, Hall IM. **2010**. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Ranwez V, Harispe S, Delsuc F, Douzery EJP. **2011**. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. *PLOS ONE* **6**: e22594.

Raymond O, Gouzy J, Just J, Badouin H, Verdenaud M, Lemainque A, Vergne P, Moja S, Choisne N, Pont C, *et al.* 2018. The Rosa genome provides new insights into the domestication of modern roses. *Nature Genetics* **50**: 772–777.

Renner SS. **2014**. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany* **101**: 1588–1596.

Renner SS, Müller NA. **2021**. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nature Plants*: 1–11.

Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, *et al.* 2005. The DNA sequence of the human X chromosome. *Nature* **434**: 325–337.

Shephard HL, Parker JS, Darby P, Ainsworth CC. **2000**. Sexual development and sex chromosomes in hop. *New Phytologist* **148**: 397–411.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, *et al.* 2011. The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* **43**: 109–116.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, *et al.* 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**: 825–837.

Small E. **2015**. Evolution and Classification of *Cannabis sativa* (Marijuana, Hemp) in Relation to Human Utilization. *The Botanical Review* **81**: 189–294.

Sousa A, Fuchs J, Renner SS. **2013**. Molecular Cytogenetics (FISH, GISH) of *Coccinia grandis*: A ca. 3 myr-Old Species of Cucurbitaceae with the Largest Y/Autosome Divergence in Flowering Plants. *Cytogenetic and Genome Research* **139**: 107–118.

Takahata N, Nei M. **1985**. Gene genealogy and variance of interpopulational nucleotide differences. *Genetics* **110**: 325–344.

Team, R. C. 2013. R: A language and environment for statistical computing.

Thomas GG, Neve RA. **1976**. Studies on the Effect of Pollination on the Yield and Resin Content of Hops (*humulus Lupulus L.*). *Journal of the Institute of Brewing* **82**: 41–45.

Torres MF, Mathew LS, Ahmed I, Al-Azwani IK, Krueger R, Rivera-Nuñez D, Mohamoud YA, Clark AG, Suhre K, Malek JA. **2018**. Genus-wide sequencing supports a two-locus model for sex-determination in *Phoenix*. *Nature Communications* **9**: 3969.

Trombetta B, Sellitto D, Scozzari R, Cruciani F. **2014**. Inter- and intraspecies phylogenetic analyses reveal extensive X-Y gene conversion in the evolution of gametologous sequences of human sex chromosomes. *Molecular Biology and Evolution* **31**: 2108–2123.

Velzen R van, Holmer R, Bu F, Rutten L, Zeijl A van, Liu W, Santuari L, Cao Q, Sharma T, Shen D, *et al.* 2018. Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proceedings of the National Academy of Sciences* **115**: E4700–E4709.

Wang J, Na J-K, Yu Q, Gschwend AR, Han J, Zeng F, Aryal R, VanBuren R, Murray JE, Zhang W, *et al.* 2012. Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences* **109**: 13710–13715.

Westergaard M. **1958**. The Mechanism of Sex Determination in Dioecious Flowering Plants. In: Demerec M, ed. Advances in Genetics. Academic Press, 217–281.

Wickham, H. 2011. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180-185.

Wright AE, Dean R, Zimmer F, Mank JE. **2016**. How to make a sex chromosome. *Nature Communications* **7**: 12087.

Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. **2016**. GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. In: Mathé E, Davis S, eds. Methods in Molecular Biology. Statistical Genomics: Methods and Protocols. New York, NY: Springer, 283–334.

Yang Z. **2007**. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Yang M-Q, Velzen R van, Bakker FT, Sattarian A, Li D-Z, Yi T-S. **2013**. Molecular phylogenetics and character evolution of Cannabaceae. *TAXON* **62**: 473–485.

Zhang Q, Onstein RE, Little SA, Sauquet H. **2019**. Estimating divergence times and ancestral breeding systems in Ficus and Moraceae. *Annals of Botany* **123**: 191–204.

- **Preprint repository**:

Padgitt-Cobb, L. K., Kingan, S. B., Wells, J., Elser, J., Kronmiller, B., Moore, D., et al. 2019. A phased, diploid assembly of the Cascade hop (*Humulus lupulus*) genome reveals patterns of selection and haplotype variation. *BioRxiv*, 786145.

Grassa, C. J., Wenger, J. P., Dabney, C., Poplawski, S. G., Motley, S. T., Michael, T. P., et al 2018. A complete Cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. *BioRxiv*, 458083.

- **Web Document:**

Barth-Haas GmbH & Co. KG. 2019. Barth Report (1950-2019). Nuremberg

Conway Sean and Snyder Reid 2008. Humulus lupulus– Hops. College Seminar 235 Food for Thought: The Science, Culture, & Politics of Food.

King M, Pavlovic M. **2018**. Analysis of Hop Use in Craft Breweries in Slovenia. **3**: 21–26.

Neve, R. A. 1991. Hops. Chapman and Hall. *London, England*.

Supporting information legends

Table S1.
Statistics of mapping on *H. lupulus* and *C. sativa* references.

Table S2.
Summary of SEX-DETector genotyping errors and inferences

Table S3.
Expression analysis statistics summary.

Table S4.
Age estimates of the youngest strata in *H. lupulus* and *C. sativa*.

Table S5.
Summary of chromosome name in Fig. 2A and assembly fasta file.

Table S6
SEX-DETector assignment file output

Figure S1.
Histogram of the Y/X expression ratio.

Figure S2.
Histogram of the Allele-specific expression analysis for the parents.

Figure S3.
Histogram of the Allele-specific expression analysis for the daughters.

Figure S4.
Histogram of the Allele-specific expression analysis for the sons.

Figure S5.
Histogram of synonymous divergence *(dS)* between *C. sativa* and *M. notabilis*.

Figure S6.
Histogram of synonymous divergence *(dS)* between *C. sativa* and *R. chinensis*.

Figure S7.
Example of genes which topology changed with the mapping bias filtering.

Figure S8.
Y/X expression ratio along the sex chromosome without genes with a detected mapping bias.

Figure S9.
Dosage compensation analysis without genes with a detected mapping bias.

Figure S10.
SEX-DETector inference errors due to Whole Genome Duplication in *H. lupulus.*