An ancient X chromosomal region harbours three genes potentially controlling sex determination in *Cannabis sativa*

Matteo Toscani[1], Afsheen Malik[1], Ainhoa Riera-Begue[1], Caroline Dowling[1], Quentin Rougemont[2], Ricardo C Rodríguez de la Vega[2], Tatiana Giraud[2], Susanne Schilling[1] and Rainer Melzer[1]

[1] UCD School of Biology and Environmental Science and Earth Institute, University College Dublin, Ireland

[2] Ecologie Société et Evolution, CNRS, Universite Paris-Saclay, AgroParisTech, 91198 Gif-sur-Yvette, France

Author emails and ORCID:

matteo.toscani@ucdconnect.ie 0009-0006-8216-6814

afsheen.malik@ucd.ie 0000-0001-8686-9740

ainhoa.rierabegue@ucdconnect.ie 0009-0005-1519-5626

caroline.a.dowling.1@gmail.com 0000-0002-1184-7798

susanne.schilling@ucd.ie 0000-0002-1142-6520

quentin.rougemont@universite-paris-saclay.fr 0000-0003-2987-3801

tatiana.giraud@universite-paris-saclay.fr 0000-0002-2685-6478

ricardo.rodriguez@cnrs.fr 0000-0002-9852-6654

rainer.melzer@ucd.ie 0000-0003-0631-9023 20

For correspondence: rainer.melzer@ucd.ie

Tables and Figures: 6 Figures, 17 Supplementary Figures, 11 Supplementary Tables

Key words: Cannabis, dioecy, hemp, monoecy, sex chromosomes, sex determination

Summary

- Sex determination mechanisms in dioecious plants remain poorly understood yet offer an excellent model system to study genetic changes underlying morphological evolution.
- We investigated the genetic basis of sex determination in *Cannabis sativa*, combining QTL mapping in a segregating population, comparative transcriptomics between monoecious and dioecious cultivars, and a genomic analysis of X-Y chromosome divergence.
- QTL mapping identified *Monoecy1*, a locus on the X chromosome putatively controlling the monoecy-dioecy trait. This locus resides in the most ancient and diverged region of the sex chromosomes and contains three genes within 60,000 bp (*CsREM16*, *lncREM16*, and *CsKAN4*) with distinct sex-specific and monoecy-specific expression patterns.
- *Monoecy1* harbours genes for male-female as well as monoecious-dioecious sex determination. We propose that the combinatorial interaction of *CSREM16*, *lncREM16* and *CsKAN4* provides a unifying genetic framework for understanding male-female and monoecious-dioecious sex determination in *Cannabis sativa*.

Introduction

The evolution of sex chromosomes remains a major evolutionary puzzle because it involves the emergence of highly differentiated chromosomes from initially homologous autosomes through recombination suppression, a process that is generally disadvantageous in the long term, notably due to reduced selection efficiency and gene loss. The current paradigm suggests that flowering plant sex determination evolved from hermaphroditic ancestors through sequential mutations causing male and female sterility, localized on proto-sex chromosomes (Charlesworth, 2013). In the chromosomal regions harboring the sterility loci, recombination ceases, to avoid producing infertile offspring that would have both sterility alleles. The non-recombining region then often expands over time along the sex chromosomes, generating evolutionary strata of differentiation between sex chromosomes (Bergero & Charlesworth, 2009). Indeed, sequence differences accumulate between sex chromosomes as a function of time since recombination suppression. Furthermore, the recombination suppression leads to less efficient selection, and therefore often degeneration, including gene losses (Bachtrog, 2013).

The size of non-recombining regions varies considerably among species. In asparagus, these regions are relatively small, containing only a very few genes (Harkess *et al.*, 2017), facilitating the identification of candidate sex-determining genes. By contrast, *Cannabis sativa* (hemp) features a vast non-recombining region spanning approximately 55Mb (Lynch *et al.*, 2025), encompassing the majority of the sex chromosomes and thousands of genes, complicating the identification of sex-determining candidates.

Cannabis is a multipurpose crop with applications in the textile industry due to its fiber production, food industry with hemp seeds and hemp seed oil, and medicinal compounds present in its inflorescences such as CBD (Andre *et al.*, 2016; Adhikary *et al.*, 2021; Schilling *et al.*, 2021). Unlike the vast majority of flowering plants that have bisexual flowers with male and female reproductive organs in one flower, *Cannabis* is mainly a dioecious crop, with male and female flowers appearing on separate individuals. However, monoecious hemp cultivars with male and female flowers on the same individual do exist as well. The monoecy trait is agriculturally significant, as monoecious cultivars are preferred for fiber production due to their uniformity because of the absence of male plants that flower and senesce earlier than female plants in dioecious cultivars (Pancaldi *et al.*, 2025).

In *Cannabis* the sex is primarily determined by sex chromosomes: male plants possess heterogametic XY chromosomes, while female as well as monoecious plants possess

homogametic XX sex chromosomes (Divashuk *et al.*, 2014; Faux *et al.*, 2014; Riera-Begue *et al.*, 2025). However, considerably plasticity exists in the system, the genetic sex can be overridden by environmental cues and hormone treatments and monoecious plants with an XY chromosome set have been identified (Lubell & Brand, 2018; Garcia-de Heer *et al.*, 2024; Timoteo Junior & Oswald, 2024; Monthony *et al.*, 2026).

In recent years several groups have attempted to identify sex determination genes in *Cannabis*. Those endeavours have led to several candidate genes: Akagi et al. (2025) proposed an *ETR1-like* ethylene receptor gene (*CsETR1*), Carey et al. (2025) proposed a homolog of an aminocyclopropane-1-carboxylate synthase (*CsACS3*) involved in ethylene biosynthesis and Shi et al. (2025) proposed a transcription factor gene belonging to the reproductive meristem gene family (*CsREM16*). All those sex determination candidate genes have in common that they are located on the X chromosome, and an X chromosome-autosome balance system has been suggested as the predominant sex determination system (Akagi *et al.*, 2025; Carey *et al.*, 2025).

Despite the clear agricultural and theoretical interest in understanding monoecy in hemp, very little is known about its genetic basis. Previous studies found QTLs for monoecy on unplaced scaffolds (Petit *et al.*, 2020). A more recent study suggested *CsACS3* as a candidate gene also for monoecy vs dioecy (Carey *et al.*, 2025)..

Moreover, despite our previous transcriptomics analysis identifying candidate genes responsible for male/female sex determination (Shi *et al.*, 2025), and transcriptomics analysis on hormone-induced male flower development in female plants (Adal *et al.*, 2021; Garcia-de Heer *et al.*, 2025), comprehensive studies comparing gene expression between female and monoecious hemp cultivars remain scarce. Yet, such studies have been successfully used in other species to pinpoint genes related to the monoecy trait (Zhao *et al.*, 2025). Additionally, gene mapping has proven effective in identifying candidate loci related to monoecy in species such as spinach and *Salix purpurea*, which exhibit both dioecious and monoecious lines (Hyden *et al.*, 2023; Yamano *et al.*, 2024).

Hemp, displaying both monoecy and dioecy, is an ideal candidate to identify their genetic basis and could further advance the more general question of plant sex determination, with sex-determining genes currently identified only in a handful of dioecious and monoecious plant species (Marais *et al.*, 2025). Notably, the Cannabaceae hemp and hop (*Humulus lupulus*) possess some of the oldest known pair of sex chromosomes in flowering plants, evolved before the divergence of the two species around 28 million years ago (Prentout *et al.*, 2021). Thus, elucidating the sex determination mechanism and the genetic network underlying monoecy

may offer critical insights into the evolution of plant sex chromosomes. To pinpoint the genetic basis of sex determination and monoecy in *Cannabis*, we employed three complementary methods: (1) QTL mapping in a segregating population, (2) comparative gene expression analyses between monoecious and dioecious cultivars (and their F_2progeny), and (3) genomic comparison between the X and Y chromosomes, to assess their degree of differentiation, the possible presence of evolutionary strata and of gene losses.

We show here that the X chromosome underpins not only male-female sex determination but also the monoecy trait. Our methods consistently converge at the most diverged region of the X chromosome, containing both male-female sex determination candidates and the monoecy locus, as central to these processes.

These findings support a model in which a mutated female (XX) genotype, through disruption of the femaleness-determining pathway, gained the capacity to produce male flowers in monoecious cultivars. This study thus provides a unifying genetic framework for understanding the evolution and regulation of sex in *Cannabis sativa*, with implications for breeding programs and for understanding how sex determination pathways evolve on plant sex chromosomes.

Materials and Methods

Plant Material and Growth Conditions

The 'Felina 32'x'FINOLA' mapping population was described previously (Dowling *et al.*, 2024). Briefly, an individual of the dioecious 'FINOLA' cultivar was used as pollen donor and the monoecious 'Felina 32' cultivar as pollen acceptor. The F_2 mapping population was generated by crossing a single male F_1 as a pollinator with four F_1 female plants. The F_2 population was grown in a glasshouse under natural light conditions in Dublin, Ireland from June to September 2020. Plants were phenotyped once per week during the observation period and analysed for the presence of male and female flowers. The individuals with both male and female flowers were phenotyped as monoecious while plants with entirely male and female flowers were marked as male and female, respectively.

Leaf tissue was harvested from the F_2 population and stored at -80°C until further processed. The DNA from all F_2 individuals, the F_1 individuals, and the parents was extracted using DNeasy® Plant Mini Kit (Qiagen GmbH, Germany) as per manufacturer's instructions. The quality and quantity of DNA was checked by agarose gel electrophoresis with a NanoDrop spectrophotometer (ND-1000). The samples were prepared to a final concentration of ≥ 35

149 ng/µl and sent to BMKGene (Beijing, China) for SLAF-Seq (Specific Locus Amplified
150 Fragment Sequencing).
151 The F_4 population was grown under glasshouse conditions from June to September 2024 under
152 glasshouse conditions in Dublin, Ireland. The floral meristem and axillary inflorescence
153 meristem were sampled for RNA extraction. The RNA extraction was carried out using
154 RNeasy® Plant Mini Kit (Qiagen GmbH, Germany) as per manufacturer's instructions. The
155 RNA concentration was determined using Nanodrop spectrophotometer (ND-1000), while
156 RNA integrity was determined using Bioanalyzer (Agilent). The samples with RIN > 5 were
157 further processed for RNA Sequencing, which was carried out by Novogene (Cambridge, UK).
158

159 **SLAF-seq and quality control**

160 A reduced representation genome sequencing approach, specific locus amplified fragment
161 sequencing (SLAF-seq) (Sun *et al.*, 2013), was used to sequence the F_2 population. SALF-seq
162 was carried out by BMKgene. The service provider (BMKGene) conducted a simulation
163 experiment for selecting the most appropriate restriction enzymes against the hemp reference
164 genome assembly (GF_029168945.1_ASM2916894v1) followed by their testing of selected
165 samples in a pilot experiment. From the simulation and pilot experiment results, the restriction
166 enzyme set RsaI and HaeIII was selected as the most appropriate set generating a substantial
167 number of fragments (SLAF-tags) that are randomly and uniformly distributed across the
168 genome and avoid repetitive regions. SLAF-tags were about 364 - 444 bp in size.
169 On average 12640 SLAF tags were produced per sample with a mean sequencing depth of 20X.
170 Libraries generated from SLAF-tags after a series of steps were processed on the Illumina
171 platform and about 51.60 Mb reads were generated. The quality of SLAF-seq fastq files was
172 checked using fastQC and summarised using MultiQC softwares (Ewels *et al.*, 2016).
173 All reads exceeded quality 20 and consequently no filtering was applied at this stage. The
174 adapter sequences were trimmed by the BMKgene sequencing company and consequently no
175 further trimming was performed.
176

177 **Mapping and genotyping**

178 After quality control, the paired-end reads from female and monoecious samples were mapped
179 against the Pink pepper reference genome (GCF_029168945.1_ASM2916894v1) using BWA-
180 mem2 (Vasimuddin *et al.*, 2019). The reads from male samples were mapped against the
181 reference genome to which the non-recombining region of the Y chromosome from the
182 'Kompolti' assembly (Lynch *et al.*, 2025) (*KOMPb*) was appended, to avoid reads from the Y

183 chromosome getting mapped to similar regions in the X chromosome and falsely genotyping
184 heterozygous SNPs in male X chromosome positions. We then measured for each sample the
185 proportion of reads mapping to the Y chromosome.
186 The resulting BAM files were filtered to keep reads of mapping quality MAPQ > 20. The
187 analysis was then repeated for the reads that mapped against the chromosome X with the higher
188 threshold filtering MAPQ > 35. Due to SLAF-seq reads being repetitive by design, duplicate
189 reads were not removed. The genotyping was then done using GATK 4.6.0 functions
190 HaplotypeCaller and GenotypeGVCFs (DePristo *et al.*, 2011).
191
192 **QTL analysis**
193 The resulting VCF file from GenotypeGVCFs, containing SNPs from the mapping population,
194 was filtered using VCFtools 0.1.17 (Danecek *et al.*, 2011) to keep biallelic markers that were
195 genotyped in at least 80 % of the samples, with quality greater than 30 and with depth of at
196 least 10, with the following parameters: --min-alleles 2 --max-alleles 2 --max-missing 0.8 --
197 minQ 30 --minGQ 30 --minDP 10.
198 The genotypes from the filtered VCF file were then imported in R using the package onemap
199 (Margarido *et al.*, 2007), which allowed also to convert the VCF file into a genotype table
200 coded by the parental origin: markers from F_2 samples which were homozygous for the
201 'FINOLA' parent were coded as AA, markers homozygous for the monoecious 'Felina 32'
202 parent were coded as BB, heterozygous markers from both parents were coded as AB. To this
203 genotype table, the phenotype data was then manually added, with the plant sex encoded as 1
204 for female and 0 for monoecious, in columns as required by R/qtl (Broman *et al.*, 2003) cross
205 format, and the cross object was imported with the read.cross() function. For the QTL analysis,
206 female and monoecious plants were used, as male plants inherit the whole Y chromosome.
207 Markers from the autosomes which show significantly distorted segregation (P.value < 0.05
208 after applying a Bonferroni correction for multiple testing) from the expected AA:AB:BB 1:2:1
209 ratio were removed. For each chromosome, markers were reordered based on their
210 recombination frequency using the R/qtl orderMarkers() function. The genetic map in
211 centimorgans is then estimated with the est.map() function, with parameters map.function =
212 "kosambi",error.prob = 0.01 , with average and median chromosomal lengths of 105.5 cM and
213 97.67 cM, respectively. The average marker spacing across chromosomes was <1 cM. Low
214 recombination fractions for markers on the same chromosomes , together with high Logarithm
215 of the Odds (LOD) scores, were used as indicators to confirm that markers assigned to the same

linkage groups were closely linked, as expected for markers within the same chromosome, to evaluate the accuracy of our genetic map and its concordance with the reference genome.

For each phenotypic trait the QTL LOD is calculated with the scanone function and method “mr-argmax”, and the 5 % significance threshold is generated by executing 10,000 permutations. To refine the significant QTL to a more precise genomic interval, we additionally applied the composite interval mapping method using the cim() function in the R/qtl package, using the same genetic map and phenotypic dataframe, improving resolution by fitting multiple marker covariates outside the testing interval to control for genetic background noise, thereby increasing the power to detect QTL, and potentially distinguish closely linked loci. In our case, we used the default model settings, with the default window of 10 cM, without specifying additional covariates, relying on the automatic forward selection procedure from the r/qtl package to select cofactors from the genome scan. The automatic cim forward selection procedure selected and used for the analysis 3 marker covariates: X_81454470 , 4_42194662 and 1_43452032.

The same QTL analysis with equal parameters was then repeated after remapping the reads against both haplotypes SAN2a and SAN2b of the phased ‘Santhica’ assembly (Lynch et al., 2025).

Gene loss ratio estimate

The degree of degeneration of different segments of the Y chromosome was estimated using the gene loss ratio as a proxy for divergence with the X chromosome. The X chromosome was divided into 10 bins of maximum 10 million nucleotides. From each bin all the proteins encoded by genes in that chromosomal section were BLASTed against the genomic sequences of the Y chromosome assemblies of 3 different cultivars (Lynch *et al.*, 2025): Kompolti (KOMPa), BooneCounty (BCMa), and a landrace cultivar from Germany (GRMa), using the tblastn algorithm. If no match with e-value < 0.05 was found in any of the assemblies, a gene was considered to have been lost on the Y chromosome. The percentage of X chromosomal genes with no counterpart on the Y chromosome was calculated for each bin and plotted.

We consider this approach a conservative estimate of gene loss due to the fact that a gene present on the X chromosome belonging to a gene family can have a match on the Y chromosome if a close member of this family is encoded on the Y chromosome, hence the "true" gene loss ratio segment is likely to be higher than the numbers estimated using this approach.

249 For each X chromosome section bin, additionally, the percentage of female biased genes was
250 calculated based on previously published transcriptomics analyses (Shi *et al.*, 2025). Genes
251 that were overexpressed in females at either the vegetative or flowering stage were counted as
252 female-biased.
253
254 **Differentiation between sex chromosomes: synonymous divergence (d_S) and evolutionary**
255 **strata**
256 From the available phased Kompolti assembly (KOMPa, KOMPb) (Lynch *et al.*, 2025), the
257 haplotypes containing X and Y chromosomes were selected and synonymous divergence (d_S)
258 values for single-copy orthologs calculated and the detection of evolutionary strata conducted
259 through the EASYstrata pipeline (Rougemont *et al.*, 2025). The pipeline identifies single-copy
260 orthologs between X and Y chromosomes using OrthoFinder (Emms & Kelly, 2019). Then it
261 aligns coding sequences for each orthologous pair accounting for frameshifts and stop codons.
262 In the next step, per-gene synonymous divergence (d_S) and non-synonymous divergence (d_N)
263 are calculated using the codeml program implemented in PAML (Yang, 2007). The resulting
264 d_S values are plotted along the X chromosome gene order to visualize patterns of divergence
265 along an ancestral-like gene order. Evolutionary strata are then inferred using Bayesian
266 changepoint analysis implemented in the MCP R package (Lindeløv, 2020), which detects
267 discontinuous changes in mean d_S values along the chromosome, for different change-point
268 values.
269 We then aimed to identify the evolutionary strata older than the *Cannabis-Humulus* speciation.
270 When recombination suppression precedes speciation, substitutions remain associated with the
271 same sex chromosome through speciation. Therefore, genes in those strata would be predicted
272 to contain trans-specific polymorphisms, i.e. in a phylogenetic tree X chromosomal genes from
273 *Cannabis* would be more closely related to X chromosomal genes from *Humulus* than to Y
274 chromosomal genes from *Cannabis*. We identified a subset of 313 genes that were single-copy
275 orthologs between X and Y chromosomes in both *Cannabis* and *Humulus*. We then constructed
276 single-gene genealogies to identify those with trans-specific polymorphisms based on the tree
277 topology.
278
279 **Transcriptomics analysis**
280 RNA-seq reads were generated previously (Dowling *et al.*, 2024) from leaves samples from
281 monoecious 'Felina 32' and dioecious 'FINOLA' females. They were now mapped against the
282 transcriptome of the reference genome 'Pink pepper' using Salmon, gene expression was

283 quantified and a comparative analysis of differential gene expression was performed using the
284 R Deseq2 package (Love *et al.*, 2014). The same comparison was performed between 3 F_2
285 female and 3 F_2 monoecious individuals with leaf samples that were obtained from the mapping
286 population. An F_3 and F_4 population was generated from the F_2 population by random
287 intercrossing. From the F_4 population 8 monoecious and 8 female dioecious samples from floral
288 meristem and axillary inflorescence meristem were taken and analyzed with the same
289 aforementioned pipeline. In DESeq2, adjusted p-values (padj) are set to NA for genes with
290 extremely low expression, zero variance across samples, or in cases where the statistical model
291 fails to converge, often due to a small sample size causing high dispersion. Our F_2 transcriptome
292 dataset, comprising three biological replicates per group, may contribute to increased
293 variability and dispersion, leading to a substantial number of genes with NA padj, including
294 some with extremely low unadjusted p-values. To identify genes consistently downregulated
295 in monoecious plants across generations, we applied a filtering threshold of $p < 0.05$ and $|\log_2$
296 fold change$| > 1$, to retain differentially expressed genes regardless of padj availability.
297 We analyzed the expression of three candidate sex-determination genes (*CsKAN4*, *CsREM16*,
298 *lncREM16*) in the 64 male, female, and monoecious samples from the cultivars 'Felina 32' and
299 'FINOLA's part of the previously published experiment mentioned above (Dowling *et al.*,
300 2024) and publicly available under BioProject PRJNA956491.
301 Expression of homologs of candidate sex determining genes *CsKAN4* and *CsREM16* was
302 evaluated in *Humulus* using published RNA-seq data from five male and five female hop plants
303 available under BioProject PRJNA694508. Homologs were identified through a BLAST
304 search, the gene expression was quantified though Salmon and the statistical significance
305 evaluated through the R Deseq2 package.
306 To compare expression patterns of our candidate sex-determining genes with previously
307 proposed candidates, we used RNA-Seq data from our earlier experiment (Shi *et al.,* 2025)
308 (BioProject accession PRJNA1126191). Briefly, directional RNA-Seq (PE150, Novogene,
309 UK) was performed on 20 samples comprising four biological replicates each of two-leaf stage
310 male (L2M) and female (L2F), four-leaf stage male (L4M) and female (L4F), and nine-leaf
311 stage female (L9F) plants. Moreover, to investigate whether the expression of candidate sex-
312 determining genes is affected by phytohormones treatment, we analysed published RNA-seq
313 data from *Cannabis sativa* male (XY), female (XX), and silver thiosulfate-induced male (XX)
314 flowers (Adal et al., 2021; BioProject PRJNA669389). Expression was quantified using
315 Salmon, and differential expression analysis was performed using DESeq2, with sex (male,
316 female, silver-induced male) as the experimental factor. Pairwise contrasts between all three

317 groups were computed and p-values were adjusted using the Benjamini-Hochberg method.
318
319 ***CsKAN4* expression analysis**
320 'FINOLA' and 'Felina 32' seeds were sown on a mixture of compost, John Innes no2,
321 vermiculite and perlite (2:1:1:1). After two days of incubation in darkness to promote
322 germination, lights (red and blue LEDs) were turned on with 18:6 photoperiod during 2 weeks
323 for vegetative growth, and then photoperiod changed to 14:10 to induce flowering.
324 To study the expression of *CsKAN4* in 'FINOLA' and Felina32, apical meristem samples were
325 taken at 1st leaf, 2nd leaf and 4th leaf stage. In the case of the ''FINOLA', from a part of the
326 sample DNA was isolated and genotyping was performed using the *CsPDS5*-CAPS protocol
327 described in Riera-Begue et al. (2025) to determine if the plants were male or female.
328 From the remaining sample RNA isolation was carried out using the RNeasy® Plant Mini Kit
329 (Qiagen, Germany) according to the manufacturer's instructions.
330 For qPCR, 200 ng of the isolated RNA was used to synthesize cDNA following the
331 SuperScript™ IV Reverse Transcriptase (Invitrogen™) protocol as per supplier's instructions.
332 The primer sequences used for the qPCR are the following:
333

Primer name	Nucleotide sequence
23_KAN4_fwd	TGAGCATGAGCATGTAGATGGA
23_KAN4_rv	TGGGACGAGTACTTACATGGG

334
335 qPCRs were conducted in a final volume of 10 µL with: 5 µL Fast SYBR™ Green Master Mix
336 (ThermoFisher Scientific), 0.2 µL of 23_KAN4_fwd (10 µM), 0.2 µL of 23_KAN4_rv (10
337 µM), 0.6 µL DEPC water and 4 µL of the cDNA template (2 ng/µL).
338 The QuantStudio™ 5 Real-Time PCR System (ThermoFisher Scientific) was used for the RT-
339 qPCR. After an initial hold stage at 95 °C for 20 s, 40 cycles at 95 °C for 10 s, at 58 °C annealing
340 temperature for 30 s were performed before a melt curve stage which consisted in 95 °C for 15
341 s, 60 °C for 80 s and 95 °C for 15s.
342 Expression levels were normalized to *CsPP2A* and *CsUBQ* as housekeeping genes (Guo et al.,
343 2018).
344
345

346 **Results**

347

348 **Investigation of evolutionary strata and shared recombination suppression between**
349 ***Cannabis* and *Humulus***

350 To identify the region of the sex chromosomes likely to harbor the sex determining genes, we
351 searched for the oldest area of recombination suppression between the X and Y chromosomes,
352 as this region would have been the first to cease recombining and would therefore show the
353 highest sequence divergence. To search for evolutionary strata, we plotted synonymous
354 divergence (d_S) for single-copy X-Y orthologs in the *Cannabis* cultivar 'Kompolti' (Fig. 1a,b,
355 Supplementary Table S1) and in *Humulus* along the X chromosome (Fig. 1c,d), as d_S increases
356 with time since recombination suppression. For both species, the changepoint analysis inferred
357 a single evolutionary stratum for the most supported model (Supplementary Fig. S1). A less
358 supported model with two changepoints inferred a small younger evolutionary stratum
359 expanding the non-recombining region towards the pseudo-autosomal region.

360 By anchoring the gene order to the same *Cannabis* physical map, we found that the
361 evolutionary strata inferred in the two species seemed concordant, suggesting that
362 recombination cessation occurred prior to the speciation event separating *Cannabis* from
363 *Humulus*. This was subsequently confirmed though single tree phylogenies identifying genes
364 with trans-specific polymorphisms between *Cannabis* and *Humulus* sex chromosomes
365 (highlighted in red in Fig. 1). The genes with trans-specific polymorphisms were contained in
366 an area spanning from 36 Mb towards the end of the chromosome X, again pointing to the
367 oldest region of recombination suppression being in this area (Supplementary Data S1).

368 We further identified specific genes exhibiting saturation of synonymous sites (dS>1),
369 indicating multiple synonymous mutations per site on average. The genes displaying the
370 highest divergence, LOC115700360 and LOC115699965, were located at 81.2 Mb and 82.3
371 Mb on the *Cannabis* X chromosome, respectively.

372

373 **Inheritance of monoecy reveals likely X-linked control system**

374 Monoecious and dioecious hemp cultivars are interfertile, which offers the opportunity to
375 analyse the monoecy-dioecy sex determination system genetically using QTL analyses. To
376 elucidate the genetic basis of monoecy, we analyzed segregation patterns across three
377 generations of a previously described cross between the monoecious cultivar 'Felina 32' acting
378 as pollen receiver and the dioecious 'FINOLA' cultivar acting as pollen donor (Dowling *et al.*,
379 2024). The F_1 generation consisted exclusively of dioecious plants (49% male, 44.9% female,

380 with the remaining 6.1% not flowering) with no monoecious individuals among 98 phenotyped
381 plants (Fig. 2a, Supplementary Table S2), demonstrating complete dominance of dioecy.
382 In the F_2 generation, monoecy reappeared with complex segregation patterns. Among 242
383 flowering individuals, we observed 43 % females, 29 % monoecious, and only 20.1 % males,
384 significantly deviating from the expected 50 % male percentage ($\chi^2 = 18.3$, $p < 0.001$; Fig. 2a).
385 To produce the F_2 population, several F_1 female individuals were pollinated using the same F_1
386 male plant. Sex frequencies in the F_2 progeny varied significantly depending on the identity of
387 the F_1 mother plant. Despite using the same F_1 male pollinator, male offspring frequency ranged
388 from 6.8% to 35.9% across different F_1 mothers ($\chi^2 = 23.65$, df = 6, $p = 0.0006$; Fig. 2b),
389 indicating maternal influence on sex inheritance.
390 252 F_2 individuals were sequenced using SLAF-seq (see below), and F_1 plants used for crossing
391 were subjected to whole genome sequencing (Dowling et al., 2024). The correlation between
392 the F_1 and F_2 plant sex phenotype and genotype was confirmed by measuring the percentage
393 of sequencing reads mapping to the Y chromosome for each plant, considered here a proxy for
394 the presence of the Y chromosome. The percentage of reads mapped against the Y chromosome
395 was > 6 % in every male sample and ≤ 3 % in female and monoecious plants (Supplementary
396 Table S3), indicating a strong correlation between genetic and phenotypic sex.
397 Moreover, the sex genotype was confirmed by a previously described PCR assay (Riera-Begue
398 et al., 2025) for 137 of the F_2 plants, confirming a 100% agreement between the phenotyped
399 and genotyped sex of the tested plants. In particular, phenotyped females and monoecious
400 plants were always genotyped as having XX chromosomes (Riera-Begue et al., 2025).
401 The F_3 generation, produced from the F_2 via random intercrossing, consisted overall of 26 %
402 female, 20 % monoecious, 31 % male individuals and 22 % of plants which did not flower in
403 the observation period. It confirmed a recessive inheritance type of the monoecy trait:
404 monoecious F_2 mothers produced 34 % monoecious offspring versus only 15.1 % monoecious
405 offspring from female F_2 mothers, a 2.3-fold difference ($\chi^2 = 11.783$, df = 2, $p = 0.0028$; Fig.
406 2c, Supplementary Fig. S2).
407
408 **A high-resolution genetic map identifies the *Monoecy1* locus on the X chromosome**
409 We constructed a high-density genetic map using SLAF-seq analysis of 252 F_2 individuals.
410 After filtering for quality and segregation distortion, we retained 1400 markers spanning 1,055
411 centiMorgans across all chromosomes, with an average spacing of <1 cM. The physical map
412 covered approximately 86 Mb of the X chromosome, providing sufficient resolution for QTL
413 mapping (Fig. 2d,e, Supplementary Figs. S3, S4).

414 QTL analysis using 189 female and monoecious F_2 plants revealed no significant associations
415 between the monoecy trait and any autosome. However, almost the entire X chromosome
416 showed a strong signal with a peak LOD score of 6.97 at marker X_80375543 (80.38 Mb),
417 exceeding the genome-wide significance threshold of 3.4 (10,000 permutations; Fig. 3a,
418 Supplementary Table S4).
419 As QTL mapping showed a broad peak covering almost the entire X chromosome (Fig. 3a, c),
420 we used Composite Interval Mapping to further refine the signal. This resulted in the
421 identification of a 2 Mb region (80-82 Mb) on the X chromosome (Fig. 3b,d, Supplementary
422 Table S5) strongly associated with monoecy and also peaking at marker X_80375543. This
423 locus, which we designate *Monoecy1*, explains 15% of phenotypic variance for the
424 monoecy/dioecy trait. Notably, *Monoecy1* was located within the area on the X chromosome
425 where the genes displaying the highest d_S divergence are contained (Figure 1a).
426 Further genotype-phenotype association analysis (Supplementary Fig. S5), comparing the
427 genotype at the identified marker X_80375543 (X_{803} hereafter) with the monoecy/dioecy
428 phenotype of the F_2 plants, provided support for the *Monoecy1* locus (Supplementary Table
429 S6). Individuals homozygous for allele X_{803_D} (which is the parental dioecious 'FINOLA'
430 genotype) exhibited an almost exclusively dioecious female phenotype (88 %, although the
431 $X_{803_D}X_{803_D}$ genotype was unexpected in F_2 plants, see Supplementary Text 1, Supplementary
432 Fig. S6, S7). In contrast, genotype X_{803_M}, the parental monoecious Felina 32 genotype, was
433 predominantly associated with monoecy (68 % homozygous F_2 plants being monoecious).
434 Individuals heterozygous for the parental genotypes ($X_{803_D}X_{803_M}$) showed an intermediate
435 distribution (62 % female, 38 % monoecious). An analysis of recombination breakpoints
436 confirmed the quantitative nature of *Monoecy1* but did not allow to narrow down the QTL
437 further (Supplementary Figure S8). Repeating the QTL analysis on chromosome X after
438 filtering the reads with a stricter mapping quality threshold of MAPQ > 35 resulted in fewer
439 markers retained but the peak marker remained the same (Supplementary Fig. S9).
440 To avoid the possibility that a genomic region only present on monoecious cultivars was being
441 missed when using the reference 'Pink Pepper' genome (Ryu *et al.*, 2024) (which is based on
442 a female plant of a dioecious cultivar), we repeated the QTL analysis after mapping the SLAF-
443 seq reads and genotyping against both haplotypes of the monoecious 'Santhica' cultivar
444 genome (Lynch *et al.*, 2025). The same pattern was obtained (Supplementary Fig. S10), with
445 most of the non-recombining region being associated with the monoecy trait, and the markers
446 with peak LOD significance being X_3257050 and X_3100042, respectively, for the SANa

447 and the SAN2b. By BLASTing the surrounding genomic region, both markers were found to
448 be placed in a region situated in the reference 'Pink Pepper' X chromosome at 82.79 Mb.
449
450 **Genomic comparisons point to the end of the X chromosome as the most divergent section**
451 To investigate the degeneration of the Y chromosome, as expected following recombination
452 suppression, we estimated the proportion of protein-coding genes that were present on the X
453 chromosome but absent on the Y chromosome. Overall, 418 out of 3,446 (12.1 %) of protein
454 coding genes on chromosome X did not have a counterpart on chromosome Y. The positions
455 of those hemizygous genes were not evenly distributed along the chromosome. The terminal
456 region spanning from 80 Mb to the end of the X chromosome displayed the highest levels of
457 divergence with the Y chromosome, with a conservative estimate of 40 % of X-encoded
458 proteins lacking a Y counterpart (Fig. 3e, Supplementary Table S7). This result is seen in the
459 reference genome but was also confirmed for the alternative 'Kompolti' cultivar genome
460 (Supplementary Fig. S11).
461 Furthermore, there is a clear pattern of the gene loss ratio increasing from one end of the X
462 chromosome to the opposite end, being near zero in the section from 0-30 Mbp. This is in
463 agreement with the position of the previously described pseudo-autosomal region (Prentout et
464 al., 2020) which is still actively recombining between X and Y chromosomes.
465 An inverse analysis was also performed measuring the genes on the Y chromosome that are
466 not present on the X chromosome. This showed a much lower ratio of Y-exclusive genes
467 (Supplementary Fig. S12), with only around 6% of proteins from the Y exclusive region not
468 found on the X, consistent with the general observation that X chromosomes are typically not
469 actively degenerating like Y chromosomes (Charlesworth, 2013).
470 Together, our observations may be taken as indication that the sex chromosome evolution in
471 *C. sativa* was initiated in a region towards the end of the X chromosome. This conclusion is
472 further supported by the previous synonymous substitution rate (d_S) analysis between X and Y
473 chromosome homologs (Prentout et al., 2020, 2021), where a peak in d_S is also observed in the
474 terminal section of chromosome X. Interestingly, *Monoecy1*, i.e. the locus controlling monoecy
475 vs. dioecy, is located in the same terminal section of the X chromosome We therefore next
476 analysed data from our previous transcriptomics experiments (Shi *et al.*, 2025), in which we
477 compared male with female apical meristem samples at different developmental stages to
478 identify genes with sex-biased expression. When analysing the percentage of X chromosomal
479 genes that showed a female-male biased expression we detected a pattern similar to the gene
480 loss analysis: the greatest percentage of female biased genes was found at the end of the X

481 chromosome, in the region where *Monoecy1* is located (Fig. 3f), with as much as 8 % of the
482 genes in this region being overexpressed in females as compared to males.
483
484 **A *KANADI* transcription factor gene is located in *Monoecy1* and differentially expressed**
485 **between monoecious and female dioecious plants**
486 We next reasoned that candidate genes controlling monoecy might be differentially expressed
487 between monoecious and dioecious cultivars. Initial comparisons between dioecious female
488 'FINOLA' and monoecious 'Felina 32' plants revealed 3,420 differentially expressed genes
489 (Fig. 4a, Supplementary Table S8, S8). The large number of differentially expressed genes in
490 the parental generation comprises both sex-related genes and cultivar-specific differences.
491 Genetic recombination in the F_2 population randomized cultivar-specific traits unrelated to sex
492 expression, allowing us to identify genes more directly associated with the monoecy trait.
493 Consequently, to refine our transcriptomics analysis, we compared gene expression profiles
494 between monoecious and dioecious female F_2 plants, which reduced the number of
495 differentially expressed genes to 90 (Fig. 4b, Supplementary Table S8, S9).
496 However, none of these 90 genes were located within the *Monoecy1* QTL when applying the
497 standard adjusted p-value filter (padj < 0.05) from DESeq2. This discrepancy likely reflects the
498 limited sample size in our F_2 transcriptomic analysis (n=3 per group), which resulted in high
499 dispersion estimates and consequently many genes with NA adjusted p-values in DESeq2,
500 despite some having extremely low unadjusted p-values.
501 To identify candidate genes within *Monoecy1* that might contribute to the monoecy trait, we
502 employed a complementary approach: we examined genes that were both (1) significantly
503 differentially expressed between the parental cultivars 'FINOLA' and 'Felina 32' (padj < 0.05,
504 $|\log_2 FC| > 1$), and (2) showed consistent differential expression in the F_2 generation using
505 instead a p-value threshold ($p < 0.01$, $|\log_2 FC| > 1$). This intersection identified LOC115699949
506 as the only gene within *Monoecy1* meeting both criteria. LOC115699949 bears homologies to
507 the *Arabidopsis thaliana* transcription factor gene *KANADI4* (*KAN4*/*ATS*) (Izhaki & Bowman,
508 2007; Gomez *et al.*, 2016) and was therefore termed *CsKAN4* (Supplementary Fig. S13).
509 *CsKAN4* is expressed more highly in female 'FINOLA' than in monoecious 'Felina 32' plants,
510 and also more highly in female F_2 than in monoecious F_2 plants (Fig. 3d). In the d_S analysis,
511 *CsKAN4* was found to be present exclusively on the X chromosome with no corresponding
512 homolog found on the Y chromosome.
513 We also extended our transcriptomic analysis to the F_4 monoecious and dioecious samples of
514 our mapping population, generated though random intercrossing of the F_2 and F_3 generation.

515 Here, the number of genes differentially expressed between monoecious and female dioecious
516 plants was just 9 (Fig. 4c, Supplementary Table S8, S9).
517 We further identified nine genes consistently downregulated across all generations of our
518 mapping population in monoecious vs. female dioecious plants when filtering for $p < 0.05$ and
519 $|\log_2 FC| > 1$ (Supplementary Table S10), one of them being likely involved in the gibberellic
520 acid (GA) pathway: *LOC115706670* (encoding a putative gibberellin 2-beta-dioxygenase 1),
521 located on chromosome 1 (Fig. 4d). This enzyme is presumably involved in GA catabolism,
522 suggesting that its reduced expression in monoecious plants may contribute to higher GA
523 catabolism and consequently more elevated GA levels compared to dioecious females, thus
524 promoting male flower development.
525 We did not detect differential expression of *CsKAN4* in the F_4 generation (Fig. 4e), but F_4
526 samples were isolated from flowering tissue whereas F_2 and parental samples were from leaf
527 tissue.To obtain a more detailed view of *CsKAN4*, we performed qPCR on apical meristem
528 samples from monoecious and dioecious individuals at different vegetative stages and young
529 flowering stages, using the *CsPP2A* housekeeping gene to normalize expression. This analysis
530 confirmed significant downregulation of *CsKAN4* in monoecious 'Felina 32' plants compared
531 to dioecious female 'FINOLA' plants (Fig. 4f, Supplementary Fig. S14), supporting the
532 hypothesis that reduced *CsKAN4* expression is a molecular signature of the monoecious
533 phenotype. Notably, *CsKAN4* differential expression was restricted to pre-flowering stages,
534 consistent with the absence of differential expression observed in RNA-seq data from the
535 inflorescences and flowers in F_4 plants. Furthermore, *CsKAN4* expression did not differ
536 significantly between male and female flowers of monoecious 'Felina 32' plants (Fig. 4f,
537 (Supplementary Fig. S14).
538
539 ***Monoecy1* contains putative male vs. female sex-determination genes**
540 Strikingly, a detailed analysis of the *CsKAN4* gene locus revealed that it is just 60,000 bp away
541 from the ortholog of the *Arabidopsis* gene AT4G33280, *CsREM16*. In *Arabidopsis*, some
542 genes of the REM (Reproductive Meristem) family are implicated in reproductive
543 development (Mantegazza *et al.*, 2014). For *Cannabis*, we previously reported that *CsREM16*
544 was strongly upregulated in females as compared to males and proposed that it constitutes a
545 molecular switch between male and female plant development (Shi *et al.*, 2025). Like *CsKAN4*,
546 *CsREM16* was found to have no corresponding Y chromosome counterpart.
547 We next analysed the region containing *CsREM16* and *CsKAN4* in more detail. Between
548 *CsREM16* and *CsKAN4* only two other genes are located, a putative long noncoding (lnc) RNA

549 (LOC133032448) and a putative pseudogene (LOC115699946) (Fig. 5a). Sequence analysis
550 further revealed that the lncRNA contains a region of ~300 nt with high similarity to *CsREM16*
551 (Supplementary Fig. S15, Supplementary Data S2, Supplementary Text 2). This might indicate
552 a possible regulatory function and we therefore termed LOC133032448 *lncREM16.*
553 To establish further evidence of the sex related expression pattern of the genes in *Monoecy1*,
554 we integrated 64 RNA-seq samples from previous experiments, encompassing monoecious and
555 dioecious samples from different developmental stages and growth conditions from the
556 'FINOLA' and' 'Felina 32' cultivars (Dowling *et al.*, 2024; Shi *et al.*, 2025) (Supplementary
557 Fig. S16a, Supplementary Table S11). This confirmed the dioecious-biased expression of
558 *CsKAN4*, the high expression of *CsREM16* in female and monoecious individuals, and showed
559 that *lncREM16* is largely silenced in monoecious and female dioecious individuals and
560 remarkably, despite being located on the non recombining region of the X chromosome,
561 *lncREM16* is expressed in male samples (Supplementary Fig. S16a, Supplementary Table S11).
562 Notably, LOC115699965, which is among the genes with the highest d_S values and is
563 homologous to the *Arabidopsis* pollen-function gene *JINGUBANG* (Ju *et al.*, 2016) is situated
564 approximately only 150kb downstream of our primary sex determination candidate *CsREM16*
565 (Shi et al., 2025) and 90kb downstream of *CsKAN4*, further supporting the ancient origin of
566 this locus. Furthermore, genome sequencing data from the monoecious 'Felina 32' cultivar
567 revealed an insertion containing a transposable element domain ca. 8,000 bp upstream the start
568 of the transcription site of *CsKAN4*, absent in the dioecious 'FINOLA' cultivar (Supplementary
569 Fig. S17, Supplementary Data S3) as a possible disruptor of its expression in the monoecious
570 cultivar.
571
572 **The candidate sex-determining genes are conserved between *Cannabis* and *Humulus***
573 Given that *Cannabis* and *Humulus* species separated after the evolution of their shared XY sex
574 chromosomes and sex determination system in their ancestor (Prentout *et al.*, 2021), we
575 investigated whether the candidate sex-determining genes identified in *Cannabis* are conserved
576 in hop (*Humulus lupulus*). Analysis of the hop X chromosome revealed that both *HlREM16*
577 (LOC133803719) and *HlKAN4* (LOC133804030) are located in the terminal section of the
578 non-recombining region (189.9-190 Mb), maintaining their close proximity as observed in
579 *Cannabis* (Fig. 5b). Protein sequence comparison showed high conservation, with *HlREM16*
580 sharing 83.3 % amino acid identity with *CsREM16*, while *HlKAN4* displayed a lower 59.3 %
581 protein level identity with *CsKAN4* (Supplementary Fig. S18). The full *lncREM16* transcript
582 was not identified in hop. Expression analysis using published RNA-seq data from five male

583 and five female hop plants demonstrated that *HlREM16* exhibits the same female-biased
584 expression pattern observed in hemp, with significantly higher ($p < 0.05$) expression in females.
585 Similar to *CsKAN4*, *HlKAN4* showed a similar expression level ($p > 0.05$) in females compared
586 to males (Supplementary Fig. S16b). In contrast to *Cannabis*, monoecy is uncommon in hops
587 (Haunold, 1991) and thus a female vs monoecious *HlKAN4* expression comparison is not
588 included here.
589
590 **Evaluation of alternative sex determination candidates**
591 Given recent proposals of ethylene signaling components as sex determinants in *Cannabis* and
592 *Humulus* (Akagi *et al.*, 2025; Carey *et al.*, 2025), we evaluated the expression patterns of the
593 two key proposed sex determination candidates: the ethylene biosynthesis gene *Cannabis*
594 *sativa 1-aminocyclopropane-1-carboxylic acid synthase 3* (*CsACS3*, *LOC115696400*) (Carey
595 *et al.*, 2025) and the putative ethylene receptor gene *Cannabis sativa Ethylene Receptor 1*
596 (*CsETR1*, *LOC115702329*) (Akagi *et al.*, 2025). Both genes reside within the terminal non-
597 recombining region of the X chromosome, positioning them as plausible sex determination
598 factors according to our model. Both are, however, not in the *Monoecy1* QTL, though *CsACS3*
599 is located very close to the QTL, at 79.96 Mb, while the QTL is positioned from 80.8Mb to
600 82.1 Mb.
601 To enable direct comparison of gene expression dynamics, we examined our three primary
602 candidates (*CsREM16*, *CsKAN4*, and *lncREM16*) across the same developmental stages and
603 tissues as the ethylene pathway genes (*CsACS3* and *CsETR1*). In vegetative apical meristem
604 samples as well as in flowers *CsREM16* showed consistent female-biased expression (Fig. 6a).
605 *lncREM16* displayed the opposite pattern, with male-specific expression in vegetative and
606 flowering samples (Fig. 6b). *CsKAN4* showed similar expression in both sexes during
607 vegetative and flowering stages with no significant male-female difference, consistent with its
608 proposed role in the monoecy/dioecy distinction rather than male/female determination (Fig.
609 6c).
610 Analysis of *CsACS3* expression across our dataset revealed tissue and developmental
611 specificity. The gene shows no or very low expression at the vegetative stage and in flowering
612 males, but expression is detectable in late-stage vegetative female and flowering female
613 samples, albeit at low levels (TPM ~ 1 to 4) (Fig. 6d). *CsACS3* has also been proposed as a
614 candidate gene for controlling monoecy vs. dioecy, with two indels in exon 4 being associated
615 with monoecy (Carey et al., 2025). Sequence analysis of the *CsACS3* locus revealed no
616 polymorphisms between the monoecious 'Felina 32' and dioecious 'FINOLA' plants used to

generate our mapping population across the entire protein coding regions (Supplementary Fig. S19).

The proposed sex determining ethylene receptor gene *CsETR1* displayed a different expression dynamic. Unlike the tissue- and sex-specific *CsACS3*, *CsETR1* maintained similar expression across all developmental stages and sex phenotypes examined (Fig. 6e, Supplementary Fig. S16), with mean TPM values ranging from 18 to 27.

Similarly, comparison between dioecious female and males 'FINOLA' and monoecious 'Felina 32' across a variety of tissues and sampling conditions revealed no significant expression differences for either *CsACS3* and *CsETR1* (adjusted p-value > 0.05) (Supplementary Fig. S16c,d).

We next reasoned that a master switch for sex determination would be expected not to change in expression even if the phenotypic sex is altered by phytohormone/hormone inhibitor application. To further analyse this we leveraged a transcriptomic data set by (Adal *et al.*, 2021) that provides expression data from relatively mature male (from XY plants) and female (XX plants) flowers but also male flowers that were induced by silver thiosulfate on female plants (i.e. male flowers on XX plants). Expression of *CsREM16, CsACS3* and *lncREM16* is associated with the genotypic sets in those datasets (Supplementary Fig. S20), i.e. *CsREM16* and *CsACS3* are expressed in female (XX) and induced male (XX) flowers at higher levels than in male (XY) flowers (for *CsETR1* the pattern was similar though not statistically significant, potentially due to high dispersion). For *lncREM16* expression in induced male flowers (XX) remains zero, similar to female flowers (XX), whereas some expression is detected in male flowers (XY). *CsKAN4* showed low expression in all three groups, indicating again low expression in the late flowering stage (Supplementary Fig. S20).

Discussion

Taken together, our data indicate that one of the oldest parts of the non-recombining region of the *Cannabis* X chromosome contains the *Monoecy1* QTL which contributes to the development of monoecious vs. dioecious individuals. *Monoecy1* contains three genes in the space of just 60,000 bp, the combinatorial expression of which predicts the sex of *Cannabis* plants: in monoecious individuals *CsREM16* is expressed, in female individuals *CsREM16* and *CsKAN4* are expressed and in male individuals *lncREM16* and *CsKAN4* are expressed (Fig. 6f).

We propose that *CsREM16, CsKAN4* and *lncREM16* constitute molecular determinants of both monoecy/dioecy and male/female sex expression in *Cannabis*. CsREM16 and CsKAN4 are

651 transcription factors belonging to the B3 and GARP family respectively, and homologs of those
652 proteins in *Arabidopsis* have been demonstrated to be involved in reproductive development
653 (Mantegazza *et al.*, 2014; Safi *et al.*, 2017). Previous studies have indicated that sex
654 determination in *Cannabis* is under X chromosome gene dosage control (Akagi *et al.*, 2025),
655 which is consistent with the X chromosomal location and male or female biased expression of
656 the genes identified here. *CsREM16* and *lncREM16* are not affected in their expression by the
657 masculinizing chemical silver thiosulfate, consistent with the idea that they are very high in the
658 hierarchy of the sex determination cascade, possibly upstream of phytohormone pathway.
659 Furthermore, that several genes controlling sex-determination are found on the oldest
660 evolutionary stratum on the *Cannabis* X chromosome supports the canonical model for the
661 evolution of plant sex chromosomes (Charlesworth, 2013), as also found recently in *Silene*
662 *latifolia* (Moraga *et al.*, 2025).
663 Though the exact molecular mechanisms of how the three genes control sex determination
664 remains elusive, it is interesting to note that *kan4* mutants in *Arabidopsis* show an increased
665 expression of GA biosynthesis genes and an increased level of GA (Gomez *et al.*, 2016).
666 Likewise, our transcriptomic data indicate that in monoecious individuals a gene encoding a
667 putative gibberellin 2-beta-dioxygenase 1 (*GA2ox1*) is downregulated in conjunction with
668 *CsKAN4*. GA2ox1 enzymes are involved in GA catabolism. Intriguingly, exogenous GA
669 applications induce the formation of male flowers on female plants in *Cannabis* (Ram &
670 Jaiswal, 1972) and a gene involved in the GA pathway has previously been implicated in sex
671 determination in *Cannabis* (Petit *et al.*, 2020). A reduction of *CsKAN4* expression in
672 monoecious plants may thus lead to an increase in GA, which in turn may foster the
673 development of male flowers. We detected differential *CsKAN4* expression mainly in
674 vegetative tissue, including leaves. Long distance transport of GA has been demonstrated in
675 other plants (Eriksson *et al.*, 2006; Regnault *et al.*, 2015b,a). It thus appears conceivable that
676 transcriptional changes in *CsKAN4* in vegetative organs can exert effects on floral
677 development.
678 The molecular role of *CsREM16* in sex determination is more difficult to pinpoint, but our data
679 indicate that its expression is strongly associated with XX as opposed to XY individuals across
680 different developmental stages and cultivars (Fig. 4b) (Shi *et al.*, 2025). In *Glycine max* the
681 *REM16* ortholog *GmREM16* is involved in controlling flowering time (Singh Yadav, 2025;
682 Wang *et al.*, 2025). In *Arabidopsis*, some *REM* mutants display strong female gametophytic
683 defects (Matias-Hernandez *et al.*, 2010; Mendes *et al.*, 2016; Caselli *et al.*, 2019). Furthermore,
684 recent data from *Arabidopsis* indicate that REM transcription factors recruit RNA polymerase

IV to activate expression of small interfering RNAs that in turn initiate DNA methylation of specific loci (Wu *et al.*, 2025; Xu *et al.*, 2025). Taken together, this may suggest that *CsREM16* has been co-opted to control sex determination of the sporophyte via changing the DNA methylation landscape.

In contrast to other plant sex determination genes, which are predominantly expressed in reproductive tissues (Leite Montalvão *et al.*, 2021; Masuda & Akagi, 2023; Charlesworth & Harkess, 2024; Marais *et al.*, 2025), *CsREM16* is expressed in all analysed female tissues. This is arguably unusual for a plant sex determination gene but a similar system exists e.g. in *Drosophila* where every somatic cell in females expresses the sex determination gene *SEX LETHAL* which is not expressed in males (Salz & Erickson, 2010). In *Drosophila*, sex is specified in a cell autonomous manner, and it will be fascinating to see whether a similar concept may apply to *Cannabis*.

Interestingly, *HlREM16*, the ortholog of *CsREM16* in *Humulus* (hops), is also located on the X chromosome and upregulated in female vs. male individuals. *Humulus* is the closest extant relative of *Cannabis* and also dioecious, and the sex chromosomes in the two genera originated before the separation of the lineages leading to hop and *Cannabis* (Prentout *et al.*, 2021). It is thus conceivable that the two genera share a similar sex determination system and that *HlREM16* and *CsREM16* are key regulators of female identity in both species. The conservation of gene location, sequence, and expression patterns between *Cannabis* and *Humulus* provides evidence that this sex-determining system predates their divergence, supporting the ancient origin of these sex chromosomes and the fundamental role of these genes in Cannabaceae sex determination.

Interestingly, in *Cannabis* we detected with *lncREM16* a long non-coding RNA with an expression pattern opposite to that of *CsREM16* and with sequence similarities to *CsREM16.* We consider it plausible that there is a regulatory relationship between the two genes, in which one may repress the other or they repress each other in a feedback loop.

We were not able to detect the long noncoding RNA *lncREM16* in hop, which may be taken as indication that the details of the sex determination system in hop and *Cannabis* diverged over time. Similar observations have been made in poplar, where two closely related species share the same gene regulatory circuits governing sex determination but the molecular details of gene regulation differ (Müller *et al.*, 2020).

719 **Quo vadis, sex determination?**

720 A number of different (not necessarily mutually exclusive) models have been proposed recently
721 on how sex expression and sex determination in the Cannabaceae is controlled (Shi *et al.*, 2025;
722 Akagi *et al.*, 2025; Carey *et al.*, 2025; Monthony *et al.*, 2026). Each of them makes testable
723 predictions about the role of different X and Y chromosomal genes.

724 Two of the proposed sex determination genes are involved in ethylene biosynthesis or
725 perception: *CsETR1* and *CsACS3* (Akagi *et al.*, 2025; Carey *et al.*, 2025). While ethylene
726 clearly plays an important role in *Cannabis* sex expression, with male plants producing female
727 flowers after ethephon treatment and female plants producing male flowers after silver nitrate
728 (an ethylene inhibitor) treatment (Ram & Jaiswal, 1972; Mohan Ram & Sett, 1982; Monthony
729 *et al.*, 2026), it is noteworthy that ethylene seems to change the sex expression of the flowers
730 but not the overall difference in architecture of male vs female plants (Shi *et al.*, 2024; Garcia-
731 de Heer *et al.*, 2025; Monthony *et al.*, 2026). Furthermore, whereas *CsACS3* is upregulated in
732 female plants starting at the four leaf pair stage, broad transcriptomic differences between male
733 and female plants are already detected earlier, with tens of genes overexpressed in females, and
734 hundreds overexpressed in males already at the 2-leaf pair stage (Shi *et al.*, 2025). *CsETR1* did
735 show differential expression between male and female plants only in mature flowers at
736 relatively late developmental stages, but not earlier during development (Figure 6,
737 Supplementary Figure S20). Taken together, we hypothesize that *CsETR1* and *CsACS3* might
738 well be involved in controlling sex expression and determining male vs. female floral identity.
739 However, they may not be the master switches controlling sex determination. The phenotypic
740 difference between male and female *Cannabis* plants extends beyond floral morphology, and
741 we predict sex determination genes like the putative candidate *CsREM16* act upstream of
742 *CsETR1* and *CsACS3*. Functional analyses will likely be required to elucidate the regulatory
743 relationships between those genes and to understand whether they act independently from each
744 other or in one regulatory cascade.

745 A polymorphism, specifically deletions in the fourth exon of *CsACS3,* was also proposed as a
746 candidate for the distinction between monoecy and dioecy. However, the amino acid sequence
747 of *CsACS3* in the monoecious 'Felina 32' and dioecious 'FINOLA' plants used in our mapping
748 population are identical, indicating that, though CsACS3 might be involved in sex
749 determination or sex expression, an amino acid sequence polymorphism in this gene may not
750 be the sole causal mutation for the monoecy/dioecy trait in 'FINOLA' vs 'Felina 32'. More
751 broadly, the discrepancies between our and previous findings (Carey *et al.*, 2025) may reflect
752 distinct genetic origins of monoecy across cultivars.

Moreover, the incomplete penetrance observed in our segregation analysis supports a polygenic model of monoecy/dioecy sex determination. *Monoecy1* explains approximately 15 % of phenotypic variance for monoecy/dioecy, suggesting additional genetic factors influence sex expression. These may include multiple loci selected through breeding, environmental interactions, and epigenetic regulation. There is a high degree of phenotypic variation in sex expression even within individual monoecious cultivars (Faux *et al.*, 2014; Trubanová *et al.*, 2025), supporting the notion that monoecy is under intricate genetic and environmental control. Further exploration of the link between GA homeostasis and sex expression may shed additional light on this complex phenotype and its genetic control.

In conclusion, all of the sex determination candidates discussed here (*CsREM16*, *lncREM16*, *CsKAN4*, *CsACS3* and *CsETR1*) may therefore be involved in controlling sex expression and sex determination to some extent. It will be fascinating to see how future experiments shape our understanding of sex determination in *Cannabis* and beyond.

Acknowledgements

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under grant number IRCLA/2022/3294. CAD was supported by an Irish Research Council–Environmental Protection Agency Government of Ireland Postgraduate Scholarship (grant no. GOIPG/2019/1987).

Competing interests

None declared.

Author contributions

MT conducted the study performing the bioinformatics analyses and generated figures and text. AM and ARB conducted sampling, DNA, and RNA extractions and molecular analyses. CD designed and created the original mapping population used as the basis for the experiment. SS and RM supervised the work. QR analyzed the strata with help from RCRLV. TG provided input in the interpretations. RM and MT wrote the manuscript with input from SS. RCRLV QR and TG commented on the manuscript. All authors read and approved the manuscript.

784 **Data availability**

785 All relevant data are contained in the manuscript and supplementary files.

786

787 **Supplementary Material**

788

789 **Supplementary Tables**

790 **Table S1. d_s and d_n values for the X-Y single copy orthologs in the Kompolti' cultivar**

791 **Table S2. Sex phenotype segregation across three generations of *Cannabis sativa* crosses**
792 **and maternal inheritance patterns**

793 **Table S3. Genetic linkage map and quantitative trait locus (QTL) analysis for sex**
794 **determination using standard interval mapping**

795 **Table S4. Refined QTL mapping for sex determination using composite interval mapping**
796 **analysis**

797 **Table S5. Association between marker X_80375543 genotypes and sex phenotype**
798 **distribution in F_2 female and monoecious plants**

799 **Table S6. Validation of male sex phenotype through Y chromosome read mapping**
800 **analysis**

801 **Table S7. Gene content and sex-biased expression patterns in the oldest evolutionary**
802 **stratum of the X chromosome (80-86 Mb)**

803 **Table S8. Differential gene expression analysis between monoecious and dioecious**
804 ***Cannabis* plants across parental and segregating generations**

805 **Table S9. Significantly differentially expressed genes (adjusted p-value < 0.05, |log2FC| >**
806 **1) between monoecious and dioecious plants across generations. For the F2 a p-value**
807 **<0.01 filter was used instead of the adjusted p-value due to the amount of NAs values.**

808 **Table S10. Core set of genes consistently downregulated in monoecious plants across**
809 **multiple generations**

810 **Table S11. Expression validation of sex determination candidate genes across 64 RNA-**
811 **seq samples from parental cultivars (BioProject PRJNA956491)**

812

813 **Supplementary Figures**

814 **Figure S1. Changepoint analysis for inferring evolutionary strata based on synonymous**
815 **divergence (d_S) between X and Y chromosomes in *Cannabis* and *Humulus*.**
816
817 **Figure S2. Progeny sex distribution by maternal phenotype.**
818
819 **Figure S3. F_2 population genotype matrix.**
820
821 **Figure S4. Genetic map and marker linkage analysis.**
822
823 **Figure S5. Genotype distribution by sex phenotype at X chromosome markers.**
824
825 **Figure S6. Genetic map after removal of unexpected genotypes.**
826
827 **Figure S7. QTL analysis with filtered genotype data.**
828
829 **Figure S8. QTL analysis after filtering the mapped reads for MAPQ >35.**
830
831 **Figure S9 . QTL analysis on data mapped and genotyped against 'Santhica' haplotypes.**
832
833 **Figure S10. Genotypes and phenotypes at QTL region of interest.**
834
835 **Figure S11. Gene loss analysis using alternative reference genome.**
836
837 **Figure S12. Y-specific gene content distribution.**
838
839 **Figure S13. Phylogenetic reconstruction of KAN4/ATS homologs.**
840
841 **Figure S14. CsKAN4 qPCR using alternative housekeeping gene**
842
843 **Figure S15. Sequence similarity between *lncREM16* and *CsREM16***
844
845 **Figure S16. Gene expression of the candidate sex-determining genes**
846

847 **Figure S17. Transposable element presence in monoecious cultivars 'Santhica 27' and**
848 **'Felina 32'.**
849
850 **Figure S18. Protein sequence conservation between *CsREM16* (LOC115699937) /**
851 ***HlREM16* (LOC133803719) and between *CsKAN4* / *HlKAN4* (LOC133804030)**
852
853 **Figure S19. Sequence conservation of the ethylene biosynthesis gene *CsACS3***
854 **(LOC115696400) between monoecious and dioecious cultivars.**

855 **Figure S20 Expression of the sex determination candidate gene in male, female, and**
856 **silver-induced male flowers.**
857

858 **Supplementary Data**
859
860 **Supplementary Data S1** Single tree phylogenies of the 313 genes where single-copy orthologs
861 between *Cannabis* X and Y and *Humulus* X and Y chromosomes were identified. Lists of the
862 trees with the two possible topologies, distinguishing whether recombination, were identified
863 are included.
864
865 **Supplementary Data S2** Nucleotide alignment of *lncREM16* showing sequence similarity to
866 *CsREM16.*
867
868 **Supplementary Data S3** Nucleotide sequence of the transposable element found upstream of
869 the *CsKAN4* gene in the monoecious 'Felina 32' parent but not the dioecious 'FINOLA' parent.
870
871 **Supplementary Text**
872
873 **Supplementary Text 1** Analysis of unexpected X chromosome genotypes in the F_2 mapping
874 population.
875
876 **Supplementary Text 2** Characterization of the male-specific long non-coding RNA
877 *lncREM16* and its sequence relationship to *CsREM16*.
878
879

880 **References**

881

882 **Adal AM, Doshi K, Holbrook L, Mahmoud SS**. **2021**. Comparative RNA-Seq analysis
883 reveals genes associated with masculinization in female Cannabis sativa. *Planta* **253**: 1–17.

884 **Adhikary D, Kulkarni M, El-Mezawy A, Mobini S, Elhiti M, Gjuric R, Ray A, Polowick**
885 **P, Slaski JJ, Jones MP, *et al.* 2021**. Medical Cannabis and Industrial Hemp Tissue Culture:
886 Present Status and Future Potential. *Frontiers in Plant Science* **12**: 627240.

887 **Akagi T, Segawa T, Uchida R, Tanaka H, Shirasawa K, Yamagishi N, Yaegashi H,**
888 **Natsume S, Takagi H, Abe A, *et al.* 2025**. Evolution and functioning of an X–A balance sex-
889 determining system in hops. *Nature Plants*: 1–14.

890 **Andre CM, Hausman J-F, Guerriero G**. **2016**. Cannabis sativa: The Plant of the Thousand
891 and One Molecules. *Frontiers in Plant Science* **7**.

892 **Bachtrog D**. **2013**. Y chromosome evolution: emerging insights into processes of Y
893 chromosome degeneration. *Nature reviews. Genetics* **14**: 113.

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

896 **Broman KW, Wu H, Sen Ś, Churchill GA**. **2003**. R/qtl: QTL mapping in experimental
897 crosses. *Bioinformatics* **19**: 889–890.

898 **Carey SB, Bentz PC, Lovell JT, Akozbek LM, Myers ZA, Korani W, Havill JS, Padgitt-**
899 **Cobb L, Lynch RC, Allsing N, *et al.* 2025**. An X-linked sex determination mechanism in
900 cannabis and hop. *bioRxiv*: 2024.12.09.627636.

901 **Caselli F, Beretta VM, Mantegazza O, Petrella R, Leo G, Guazzotti A, Herrera-Ubaldo**
902 **H, de Folter S, Mendes MA, Kater MM, *et al.* 2019**. REM34 and REM35 Control Female
903 and Male Gametophyte Development in Arabidopsis thaliana. *Frontiers in Plant Science* **10**.

904 **Charlesworth D**. **2013**. Plant sex chromosome evolution. *Journal of Experimental Botany* **64**:
905 405–420.

906 **Charlesworth D, Harkess A**. **2024**. Why should we study plant sex chromosomes? *The Plant*
907 *Cell* **36**: 1242–1256.

908 **Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE,**
909 **Lunter G, Marth GT, Sherry ST, *et al.* 2011**. The variant call format and VCFtools.
910 *Bioinformatics* **27**: 2156–2158.

911 **DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA,**
912 **del Angel G, Rivas MA, Hanna M, *et al.* 2011**. A framework for variation discovery and
913 genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**: 491–498.

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

Dowling CA, Shi J, Toth JA, Quade MA, Smart LB, McCabe PF, Schilling S, Melzer R. **2024**. A FLOWERING LOCUS T ortholog is associated with photoperiod-insensitive flowering in hemp (Cannabis sativa L.). *The Plant Journal* **119**: 383–403.

Emms DM, Kelly S. **2019**. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.

Ewels P, Magnusson M, Lundin S, Käller M. **2016**. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047–3048.

Faux A-M, Berhin A, Dauguet N, Bertin P. **2014**. Sex chromosomes and quantitative sex expression in monoecious hemp (Cannabis sativa L.). *Euphytica* **196**: 183–197.

Garcia-de Heer L, Guo Q, Mieog J, Nolan M, Liu L, Dimopoulos N, Melzer R, Kretzschmar T. **2025**. Transcriptomic analysis of ethephon-induced sex reversion of male Cannabis sativa reveals changes in expression of floral homeotic genes and a distinct trichome morphology. *Journal of Experimental Botany* **76**: 5401–5418.

Garcia-de Heer L, Mieog J, Burn A, Kretzschmar T. **2024**. Why not XY? Male monoecious sexual phenotypes challenge the female monoecious paradigm in Cannabis sativa L. *Frontiers in Plant Science* **15**: 1412079.

Gomez MD, Ventimilla D, Sacristan R, Perez-Amador MA. **2016**. Gibberellins Regulate Ovule Integument Development by Interfering with the Transcription Factor ATS. *Plant Physiology* **172**: 2403.

Guo R, Guo H, Zhang Q, Guo M, Xu Y, Zeng M, Lv P, Chen X, Yang M. **2018**. Evaluation of reference genes for RT-qPCR analysis in wild and cultivated Cannabis. *Bioscience, Biotechnology, and Biochemistry* **82**: 1902–1910.

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.

Haunold A. **1991**. 29 - Cytology and Cytogenetics of Hops. *Developments in Plant Genetics and Breeding* **Volume 2, Part B**: 551–563.

Hyden B, Zou J, Wilkerson DG, Carlson CH, Robles AR, DiFazio SP, Smart LB. **2023**. Structural variation of a sex-linked region confers monoecy and implicates GATA15 as a master regulator of sex in Salix purpurea. *New Phytologist* **238**: 2512–2523.

Izhaki A, Bowman JL. **2007**. KANADI and Class III HD-Zip Gene Families Regulate Embryo Patterning and Modulate Auxin Flow during Embryogenesis in Arabidopsis. *The Plant Cell* **19**: 495–508.

Ju Y, Guo L, Cai Q, Ma F, Zhu Q-Y, Zhang Q, Sodmergen. **2016**. Arabidopsis JINGUBANG Is a Negative Regulator of Pollen Germination That Prevents Pollination in Moist Environments. *The Plant Cell* **28**: 2131–2146.

Leite Montalvão AP, Kersten B, Fladung M, Müller NA. **2021**. The Diversity and Dynamics of Sex Determination in Dioecious Plants. *Frontiers in Plant Science* **11**.

Lindeløv J. **2020**. mcp: An R Package for Regression With Multiple Change Points.

Love MI, Huber W, Anders S. **2014**. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.

Lubell JD, Brand MH. **2018**. Foliar Sprays of Silver Thiosulfate Produce Male Flowers on Female Hemp Plants. *HortTechnology* **28**: 743–747.

Lynch RC, Padgitt-Cobb LK, Garfinkel AR, Knaus BJ, Hartwick NT, Allsing N, Aylward A, Bentz PC, Carey SB, Mamerto A, *et al.* 2025. Domesticated cannabinoid synthases amid a wild mosaic cannabis pangenome. *Nature* **643**: 1001–1010.

Mantegazza O, Gregis V, Mendes MA, Morandini P, Alves-Ferreira M, Patreze CM, Nardeli SM, Kater MM, Colombo L. **2014**. Analysis of the arabidopsis REM gene family predicts functions during flower development. *Annals of Botany* **114**: 1507–1515.

Marais GAB, Branco C, Rocheta M, Dufay M, Tonnabel J. **2025**. Plant sex-determining genes and the genetics of the evolution towards dioecy. *Journal of Experimental Botany* **76**: 3896–3911.

Margarido GRA, Souza AP, Garcia A a. F. **2007**. OneMap: software for genetic mapping in outcrossing species. *Hereditas* **144**: 78–79.

Masuda K, Akagi T. **2023**. Evolution of sex in crops: recurrent scrap and rebuild. *Breeding Science* **73**: 95–107.

Matias-Hernandez L, Battaglia R, Galbiati F, Rubes M, Eichenberger C, Grossniklaus U, Kater MM, Colombo L. **2010**. VERDANDI is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in Arabidopsis. *The Plant Cell* **22**: 1702–1715.

Mendes MA, Guerra RF, Castelnovo B, Silva-Velazquez Y, Morandini P, Manrique S, Baumann N, Groß-Hardt R, Dickinson H, Colombo L. **2016**. Live and let die: a REM complex promotes fertilization through synergid cell death in Arabidopsis. *Development* **143**: 2780–2790.

Mohan Ram HY, Sett R. **1982**. Induction of fertile male flowers in genetically female Cannabis sativa plants by silver nitrate and silver thiosulphate anionic complex. *Theoretical and Applied Genetics* **62**: 369–375.

Monthony AS, Roy J, Ronne M de, Carlson O, Murch SJ, Torkamaneh D. **2026**. Sex-specific ethylene responses drive floral sexual plasticity in Cannabis sativa. *The Plant Journal* **125**: e70721.

Moraga C, Branco C, Rougemont Q, Jedlička P, Mendoza-Galindo E, Veltsos P, Hanique M, Vega RCR de la, Tannier E, Liu X, *et al.* 2025. The Silene latifolia genome and its giant Y chromosome. *Science (New York, N.Y.)* **387**: 630.

Müller NA, Kersten B, Leite Montalvão AP, Mähler N, Bernhardsson C, Bräutigam K, Carracedo Lorenzo Z, Hoenicka H, Kumar V, Mader M, *et al.* 2020. A single gene underlies the dynamic evolution of poplar sex determination. *Nature Plants* **6**: 630–637.

Pancaldi F, Salentijn EMJ, Trindade LM. **2025**. From fibers to flowering to metabolites: unlocking hemp (Cannabis sativa) potential with the guidance of novel discoveries and tools. *Journal of Experimental Botany* **76**: 109–123.

Petit J, Salentijn EMJ, Paulo M-J, Denneboom C, Trindade LM. **2020**. Genetic Architecture of Flowering Time and Sex Determination in Hemp (Cannabis sativa L.): A Genome-Wide Association Study. *Frontiers in Plant Science* **11**: 569958.

Prentout D, Stajner N, Cerenak A, Tricou T, Brochier-Armanet C, Jakse J, Käfer J, Marais GAB. **2021**. Plant genera Cannabis and Humulus share the same pair of well-differentiated sex chromosomes. *New Phytologist* **231**: 1599–1611.

Ram HYM, Jaiswal VS. **1972**. Induction of male flowers on female plants of Cannabis sativa by gibberellins and its inhibition by abscisic acid. *Planta* **105**: 263–266.

Riera-Begue A, Toscani M, Malik A, Dowling C, Schilling S, Melzer R. **2025**. A simple and reliable PCR-based method to differentiate between XX and XY sex genotypes in Cannabis sativa. *bioRxiv*: 2025.05.02.651914.

Rougemont Q, Lucotte E, Boyer L, Jalaber de Dinechin A, Snirc A, Giraud T, Rodríguez de la Vega RC. **2025**. EASYstrata: an all-in-one workflow for genome annotation and genomic divergence analysis. *NAR Genomics and Bioinformatics* **7**.

Ryu B-R, Gim G-J, Shin Y-R, Kang M-J, Kim M-J, Kwon T-H, Lim Y-S, Park S-H, Lim J-D. **2024**. Chromosome-level Haploid Assembly of Cannabis sativa L. cv. Pink Pepper. *Scientific Data* **11**: 1442.

Safi A, Medici A, Szponarski W, Ruffel S, Lacombe B, Krouk G. **2017**. The world according to GARP transcription factors. *Current Opinion in Plant Biology* **39**: 159–167.

Salz H, Erickson JW. **2010**. Sex determination in Drosophila: The view from the top. *Fly* **4**: 60–70.

Schilling S, Dowling CA, Shi J, Ryan L, Hunt DJL, O'Reilly E, Perry AS, Kinnane O, McCabe PF, Melzer R. **2021**. The Cream of the Crop: Biology, Breeding, and Applications of Cannabis Sativa. : 471–528.

Shi J, Schilling S, Melzer R. **2024**. Morphological and genetic analysis of inflorescence and flower development in hemp (Cannabis sativa L.). : 2024.01.25.577276.

Shi J, Toscani M, Dowling CA, Schilling S, Melzer R. **2025**. Identification of genes associated with sex expression and sex determination in hemp (Cannabis sativa L.). *Journal of Experimental Botany* **76**: 175–190.

Singh Yadav A. **2025**. Time to bloom: GmREM16a promotes flowering time in soybeans. *Plant Physiology* **198**.

Timoteo Junior AA, Oswald IWH. **2024**. Optimized guidelines for feminized seed production in high-THC Cannabis cultivars. *Frontiers in Plant Science* **15**: 1384286.

1029 **Trubanová N, Isobe S, Shirasawa K, Watanabe A, Kelesidis G, Melzer R, Schilling S**.
1030 **2025**. Genome-specific association study (GSAS) for exploration of variability in hemp
1031 (Cannabis sativa). *Scientific Reports* **15**: 8371.

1032 **Vasimuddin Md, Misra S, Li H, Aluru S**. **2019**. Efficient Architecture-Aware Acceleration
1033 of BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed
1034 Processing Symposium (IPDPS). Rio de Janeiro, Brazil: IEEE, 314–324.

1035 **Wang Z, Du C, Li Q, Li M, Wang Y, Bao G, Yin Y, Yang M, Yang Q, Xu P, *et al.* 2025**.
1036 The transcription factor REM16a promotes flowering time in soybean by activating flowering-
1037 related genes. *Plant Physiology* **198**.

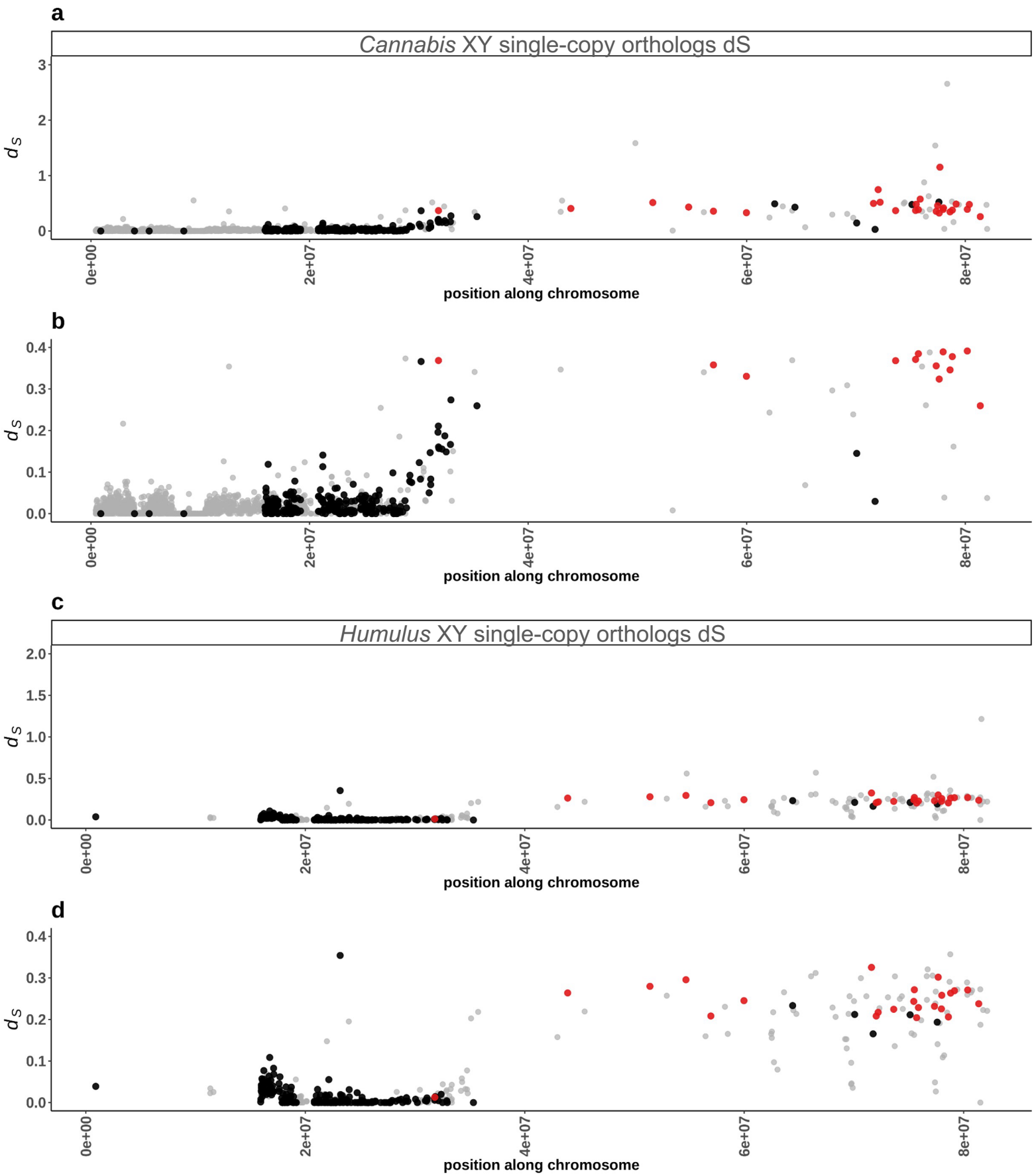
1038 **Wu Z, Xue Y, Wang S, Shih Y-H, Zhong Z, Feng S, Draper J, Lu A, Hoeke CA, Sha J, *et***
1039 ***al.* 2025**. REM transcription factors and GDE1 shape the DNA methylation landscape through
1040 the recruitment of RNA polymerase IV transcription complexes. *Nature Cell Biology* **27**: 1136–
1041 1147.

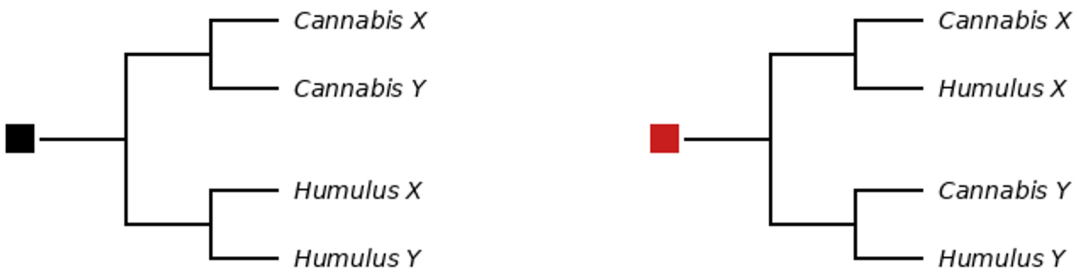
1042 **Xu G, Chen Y, Martins LM, Li E, Wang F, Magana T, Ruan J, Law JA**. **2025**.
1043 Transcription factors instruct DNA methylation patterns in plant reproductive tissues. *Nature*
1044 *Cell Biology* **27**: 2116–2127.

1045 **Yamano K, Haseda A, Iwabuchi K, Osabe T, Sudo Y, Pachakkil B, Tanaka K, Suzuki Y,**
1046 **Toyoda A, Hirakawa H, *et al.* 2024**. QTL analysis of femaleness in monoecious spinach and
1047 fine mapping of a major QTL using an updated version of chromosome-scale pseudomolecules.
1048 *PLOS ONE* **19**: e0296675.

1049 **Yang Z**. **2007**. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology*
1050 *and Evolution* **24**: 1586–1591.

1051 **Zhao Y, Liu Z, She H, Xu Z, Zhang H, Zheng S, Qian W**. **2025**. Comparative Transcriptome
1052 Analysis of Gene Expression Between Female and Monoecious Spinacia oleracea L. *Genes* **16**:
1053 24.





No single-copy ortholog between *Cannabis* and *Humulus* X and Y chromosomes, or tree topology not unambiguous

Figure 1. Synonymous divergence (d_S) and trans-specific polymorphisms analysis of *Cannabis* and *Humulus* X-Y single copy orthologs.

a) Synonymous divergence substitution rates (d_S) between *Cannabis sativa* X-Y single-copy orthologs plotted against their chromosomal position in the X (Mb). The y-axis is truncated at $d_S = 3$ to allow visualization of the majority of gene pairs while excluding extreme outliers. **b)** *Cannabis* X-Y d_S values plotted by gene rank order along the reference chromosome. The y-axis is truncated at $d_S = 0.4$ to improve resolution of functionally constrained genes and exclude possible pseudogenes. **c)** Synonymous divergence substitution rates (d_S) between *Humulus lupulus* X-Y single-copy orthologs plotted against their chromosomal position in *Cannabis sativa* X (Mb), with y-axis truncated at $d_S = 3$. **d)** *Humulus* X-Y d_S values plotted by gene rank order along the reference chromosome, along the x-axis, with y-axis truncated at $d_S = 0.4$.

In all panels, genes with trans-specific polymorphisms, presumed to belong to the genomic regions that stopped recombining before speciation, are highlighted in red. Gene ordering is always based on the position of orthologs on the YMv2a assembly chrX (Lynch et al., 2025), which serves here as the "ancestral" reference to maintain a consistent and comparable ordering across the two species.

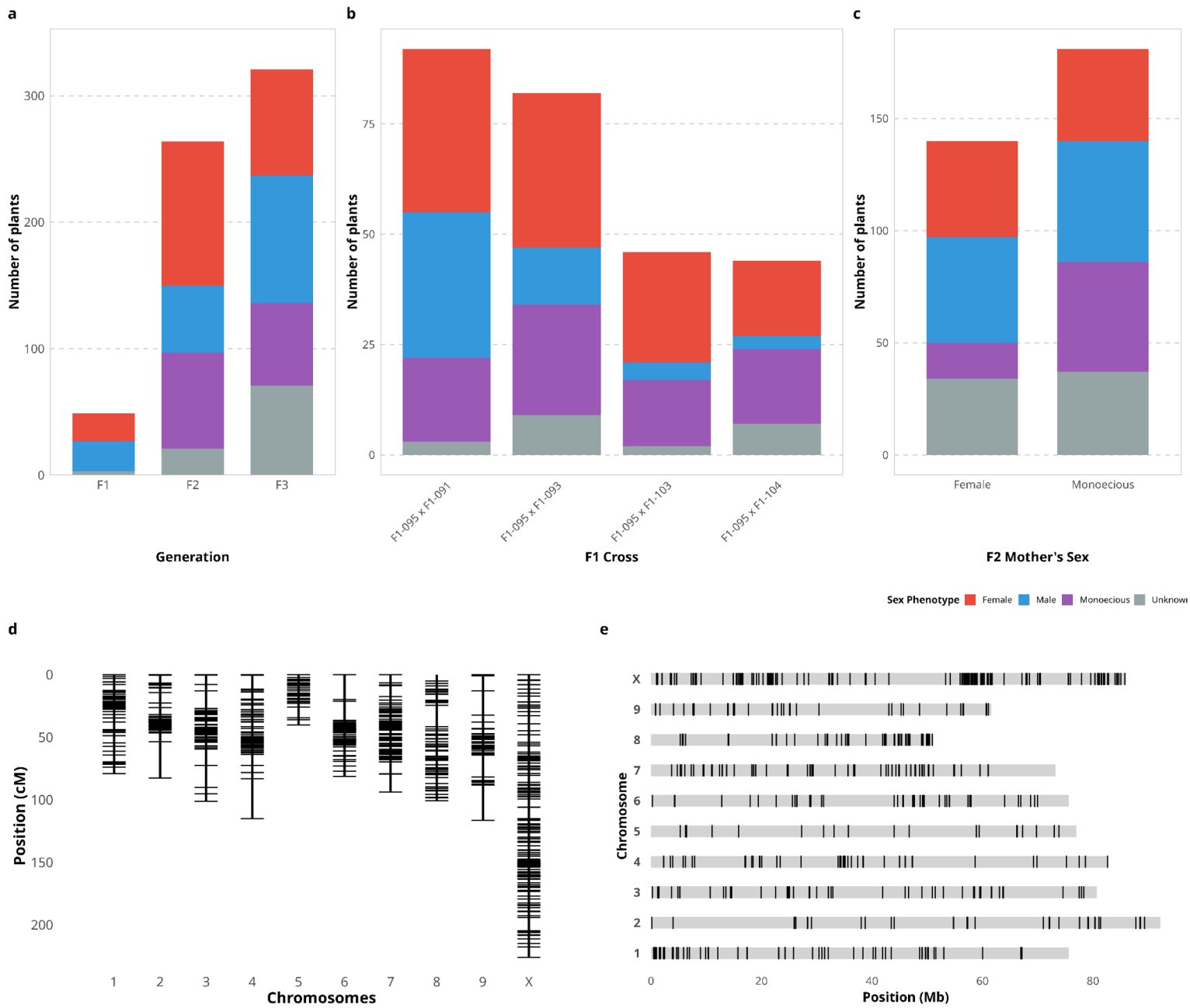


Figure 2. Sex inheritance patterns across generations and genetic mapping in Cannabis sativa. **a)** Sex distribution across three generations of a 'FINOLA' (dioecious) × 'Felina 32' (monoecious) cross. **b)** Sex distribution of F_2 plants grouped by F_1 mother plant. **c)** Sex distribution of F_3 plants grouped by F_2 maternal parent sex phenotype. **d)** Genetic map constructed from 1,400383 SLAF-seq markers spanning 1,055 cM across 10 chromosomes. **e)** Physical map showing marker distribution along chromosomes in megabases (Mb). Figure 2. QTL mapping and sex chromosome evolution in *Cannabis sativa*.

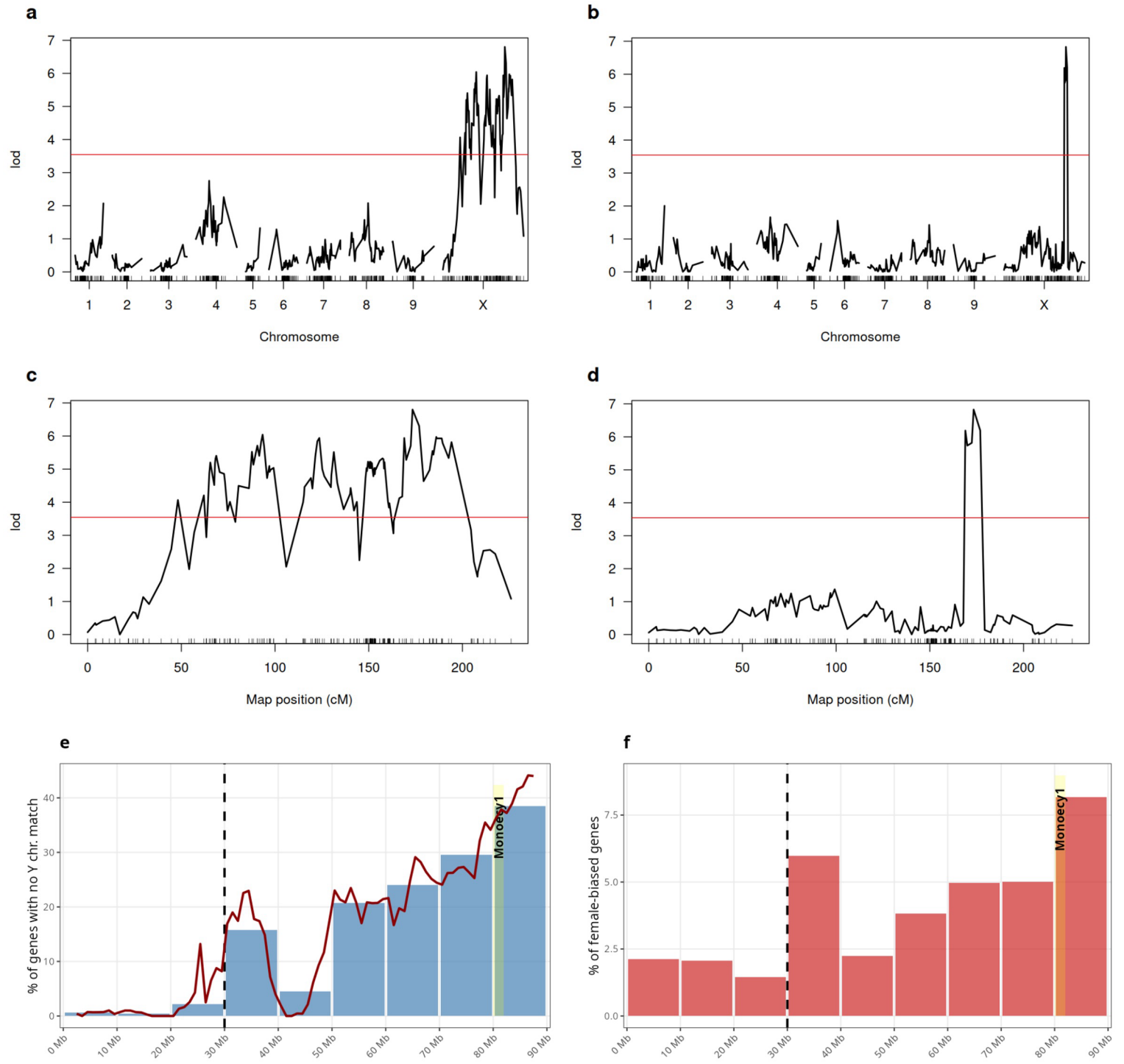


Figure 3. QTL mapping and sex chromosome evolution in *Cannabis*.

a) Genome-wide QTL scan for monoecy trait using binary model (EM algorithm). The red line indicates the 5% significance threshold (10,000 permutations). **b)** Composite Interval Mapping (CIM) confirming X chromosome QTL. **c)** Detailed X chromosome QTL analysis showing peak LOD score at marker X_80375543. **d)** CIM refinement of *Monoecy1* locus to 80-82 Mb interval. **e)** Percentage of X-chromosomal genes lacking Y chromosome homologs in 10 Mb bins. The dashed line indicates PAR boundary; yellow highlights *Monoecy1*. **f)** Distribution of female-biased genes along the X chromosome. Yellow highlights *Monoecy1*. Female-biased genes are defined as those exhibiting significantly higher expression in females compared to males (adjusted p-value < 0.05, log2 fold change > 1) in at least one of two developmental stage comparisons (vegetative or flowering), as identified in Shi et al. (2025).

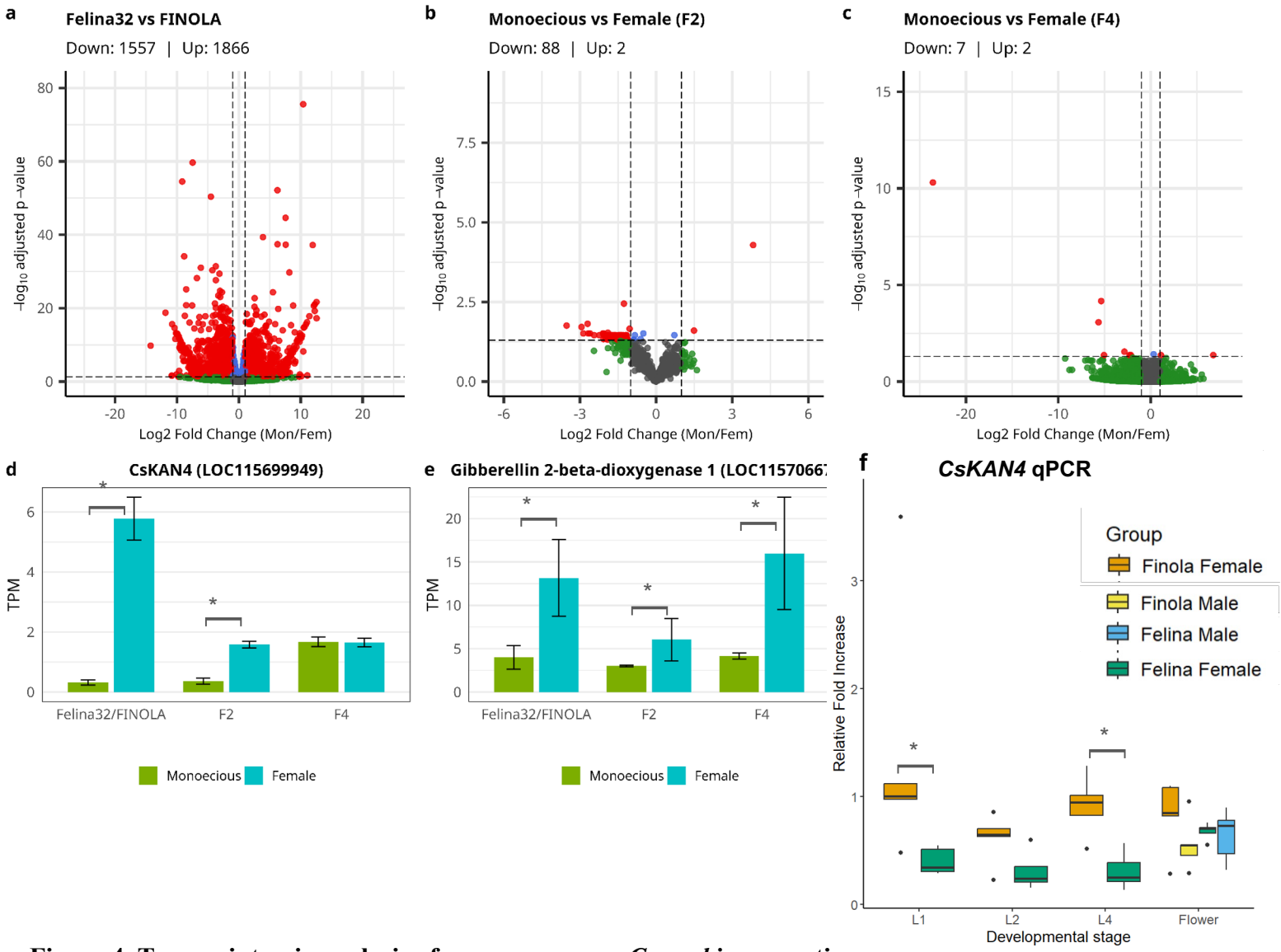


Figure 4. Transcriptomic analysis of monoecy across *Cannabis* generations.

a-c) Volcano plots depicting differential gene expression between monoecious and female *C. sativa* plants across three generations. Log2 fold change (x-axis) represents expression in monoecious relative to female plants; -log10 adjusted p-value (y-axis) indicates statistical significance. Red dots represent significantly differentially expressed genes (DEGs; adjusted $p < 0.05$, $|\log2FC| > 1$). **a)** Comparison between parental lines 'Felina 32' (monoecious) and 'FINOLA' (dioecious females) reveals extensive transcriptional differences. **b)** F_2 segregating population shows a marked reduction in DEGs as genetic recombination has randomised cultivar-specific traits unrelated to sex expression. **c)** F_4 generation demonstrates further reduction of DEGs. **d-f)** ETPM-normalized expression of key monoecy related candidate genes across generations and developmental stages. **d)** *Gibberellin 2-beta-dioxygenase 1* (LOC115706670), thought to be degrading GA, displays consistently downregulation in monoecious plants across generations. Error bars represent standard error of the mean. **e)** *CsKAN4* (LOC115699949) shows downregulation in leaf samples of monoecious parental line 'Felina 32' and in F_2 monoecious plants compared to females, but not in flower meristem samples in F_4 plants. **f)** *CsKAN4* (LOC115699949) qPCR showing is expressioned in the apical meristem in early stages of plant development (two leaf pairs, stage L2) and in later stages (four and nine leaf pairs, L4 and L9) and in flower tissue in cultivars 'Felina 32' (monoecious) and 'FINOLA' (dioecious). .* indicates $p < 0.05$. For panels d and e, p-values were obtained from DESeq2 pairwise group comparisons. For panel f, the Wilcoxon–Mann-Whitney test was used due to non-normality of the data.

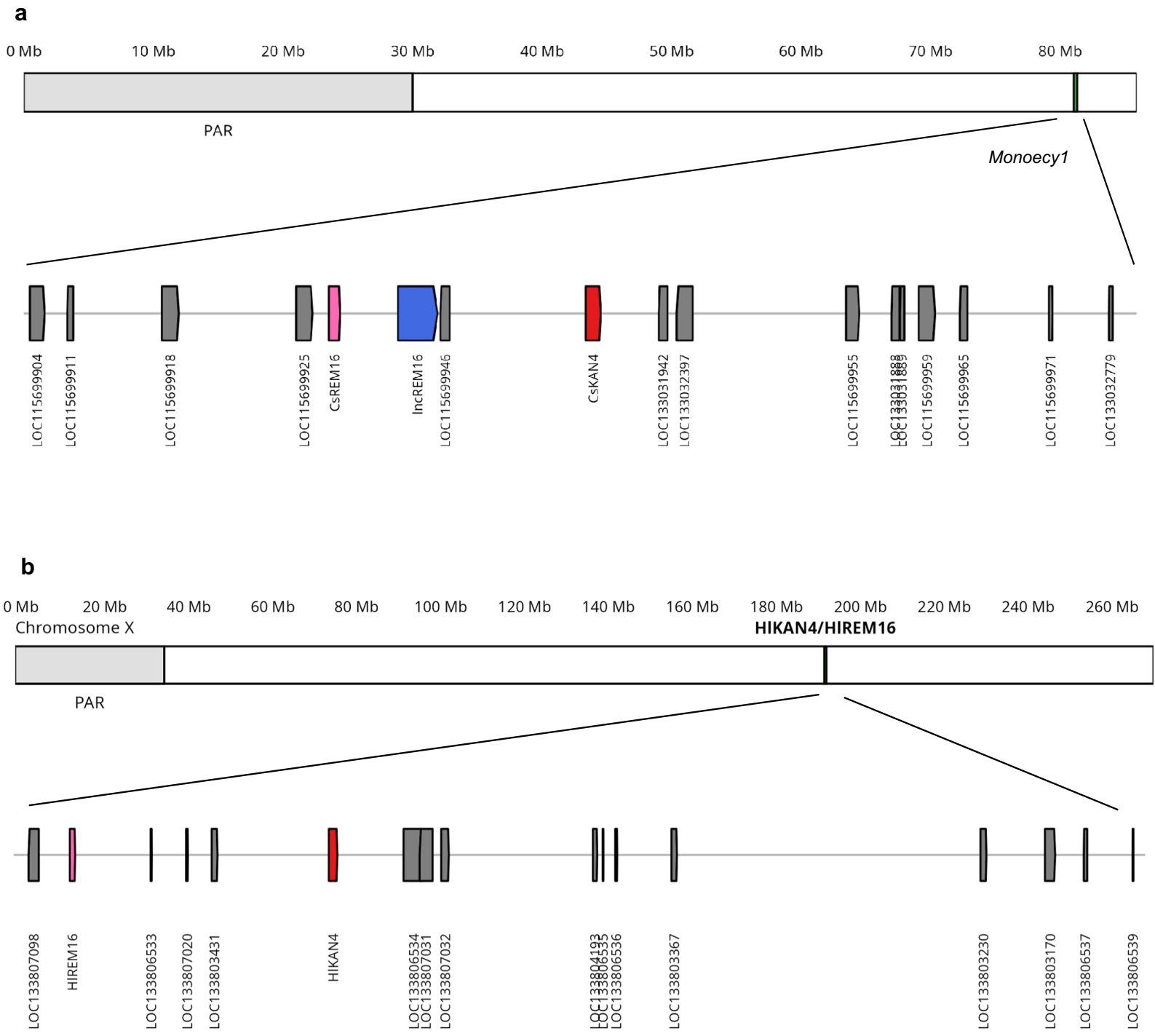


Figure 5. Genomic organization and expression patterns of the potential sex-determining genes in the *Monoecy1* locus.

a) Chromosome X ideogram and detailed gene organization in the *Monoecy1* region surrounding *CsREM16* and *CsKAN4* (81.025-81.300 Mb). The upper panel shows the X chromosome with the pseudoautosomal region (PAR, gray) and the highlighted region containing potential sex-determining genes contained in the *Monoecy1* locus (green). The lower panel displays gene arrangements with directional arrows indicating strand orientation. The putative sex-determining genes are colored according to their sex-bias classification: *CsREM16* (pink, XX-biased), *lncREM16* (blue, male-biased), and *CsKAN4* (red, dioecious-biased). **b)** Genomic organization of candidate sex-determining gene orthologs on the *H. lupulus* X chromosome. *HlREM16* (LOC133803719) and *HlKAN4* (LOC133804030) are located in close proximity within the terminal non-recombining region of the X chromosome, as they are in *Cannabis*, at 189.9-190.0 Mb. Gene positions are indicated with directional arrows showing transcriptional orientation. The lncREM16 present in *Cannabis* was not identified in *H. lupulus* in its entirety.

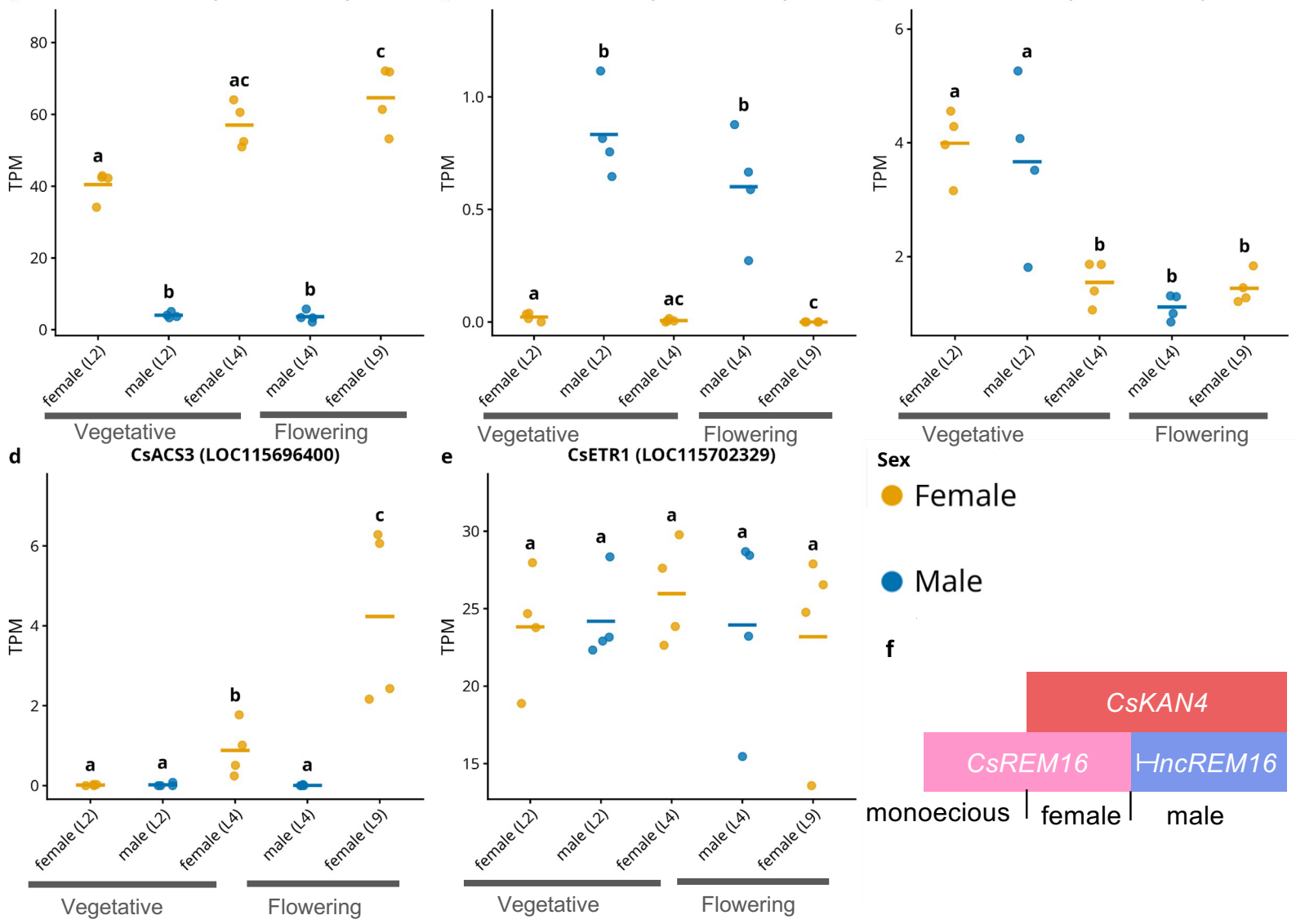


Figure 6. Expression analysis of sex determination candidate genes across developmental stages and cultivars. (a–c) Expression of primary sex-determination candidates in apical meristem samples across developmental stages. **(a)** *CsREM16* (LOC115699937) shows consistent female-biased expression at all stages examined **(b)** lncREM16 (LOC133032448) displays male-specific expression with no detectable expression (TPM = 0) in females. **(c)** CsKAN4 (LOC115699949) shows similar expression between sexes across all developmental stages. **(d–e)** Expression of alternative proposed ethylene signaling pathway sex determinants in apical meristem samples. **(d)** CsACS3 (LOC115696400) shows no or very low expression at early vegetative stages (L2) and in flowering males (L4), with detectable expression emerging in late vegetative females (L4) and flowering females (L9), albeit at low levels (TPM ~1–4). **(e)** CsETR1 (LOC115702329) exhibits constitutive expression across all developmental stages and sex phenotypes, with no significant differences between males and females (adjusted p-value > 0.05). Expression values are shown as transcripts per million (TPM). Statistical significance was assessed using DESeq2 pairwise contrasts with Benjamini-Hochberg adjusted p-values. Groups sharing the same letter are not significantly different (adjusted p < 0.05). Orange: female; blue: male. Horizontal bars indicate group means. Developmental stages are two (L2), four (L4) and 9 (L9) leaf pair stages. Shoot apical meristem or flowering tissues were sampled. Males flower at L4, females at L9. Stages and RNAseq data from Shi *et al.* (2025) **f)** Combinatorial model of sex determination in Cannabis governed by three key regulators: CsREM16, CsKAN4 and lncREM16. XX females express CsREM16 and CsKAN4 but not lncREM16, resulting in female flower development. XY males express lncREM16 and CsKAN4 but CsREM16 is silenced (potentially by lncREM16), leading to male flower development. XX monoecious plants express CsREM16 but not lncREM16 and CsKAN4, enabling development of both male and female flowers on the same individual.