Evidence for dosage compensation in *Coccinia grandis*, a plant with a highly heteromorphic XY system

Cécile Fruchard[1], Hélène Badouin[1], David Latrasse[2], Ravi S. Devani[3], Aline Muyle[1], Bénédicte Rhoné[1], Susanne S. Renner[4], Anjan K. Banerjee[3], Abdelhafid Bendahmane[2], Gabriel A.B. Marais[1,*]

1. *Laboratoire de Biométrie et Biologie Evolutive (LBBE), UMR5558, Université Lyon 1, 69622 Villeurbanne, France.*
2. *Institute of Plant Sciences Paris Saclay (IPS2), University of Paris Saclay, 91405 Orsay, France.*
3. *Biology Division, Indian Institute of Science Education and Research (IISER), Pune, Maharashtra, India.*
4. *Systematic Botany and Mycology, University of Munich (LMU), Menzinger Str. 67, 80638 Munich, Germany.*

* corresponding author: gabriel.marais@univ-lyon1.fr

Abstract

Some ~15.000 angiosperms are dioecious, but mechanisms of sex determination in plants remain poorly understood. In particular, how Y chromosomes evolve and degenerate, and whether dosage compensation evolves as a response, are matters of debate. Here we focus on *Coccinia grandis*, a dioecious cucurbit with the highest level of X/Y heteromorphy recorded so far. We identified sex-linked genes using RNA sequences from a cross and a model-based method termed SEX-DETector. Parents and F1 individuals were genotyped and the transmission patterns of SNPs then analysed. In the >1300 sex-linked genes studied, X-Y divergence was 0.13 - 0.17, and substantial Y degeneration is implied by an average Y/X expression ratio of 0.63 and an inferred gene loss on the Y of ~40%. We also found reduced Y gene expression being compensated by elevated expression of corresponding genes on the X and an excess of sex-biased genes on the sex chromosomes. Molecular evolution of sex-linked genes in *C. grandis* is thus comparable to that in *Silene latifolia*, another dioecious plant with a strongly heteromorphic XY system, and cucurbits are the fourth plant family in which dosage compensation is described, suggesting it might be common in plants.

Key words

dioecy; sex chromosomes; heteromorphy; Y degeneration; dosage compensation; cucurbits

Introduction

Some 5 or 6% of the angiosperms, depending on the assumed total species number, have male and female sporophytes, a sexual system termed dioecy, (Renner, 2014). Transitions from other sexual systems towards dioecy are estimated to have occurred between 871 and 5,000 times independently (Renner, 2014). Chromosomes and sex determination have been studied in few dioecious plants, however, and microscopically distinguishable (heteromorphic) sex chromosomes have been reported in about 50 species only (Ming et al., 2011). An important question is whether sex chromosomes evolve similarly in plants and animals. For example, the evolution towards heteromorphy might be common between both lineages, with an autosomal origin of the sex chromosomes, gradual recombination suppression between X and Y chromosomes, and genetic degeneration of the Y chromosome (Charlesworth et al., 2005; Bergero & Charlesworth, 2009; Ming et al., 2011; Muyle et al. 2017). Animal Y chromosomes tend to shrink over time and can become tiny as in the old heteromorphic systems of mammals and *Drosophila* (Charlesworth et al., 2005; Bergero & Charlesworth, 2009; Bachtrog, 2013 but see Mahajan et al., 2018). In plants, the size of the Y chromosomes might evolve in a non-linear way, with Y chromosomes of intermediate age being larger than their X counterparts, but very old Y chromosomes being smaller (Ming et al., 2011). Systems with large Y chromosomes have not been found in animals and might be plant-specific; the reason for this is unknown. *Silene latifolia* has one such system and has been intensively studied (e.g. Filatov et al., 2000; Nicolas et al., 2005; Bergero et al., 2008; Cermak et al., 2008; Marais et al., 2008; Chibalina & Filatov, 2011; Bergero & Charlesworth, 2011; Muyle et al., 2012; Papadopoulos et al., 2015; Muyle et al., 2018; Rodríguez Lorenzo et al., 2018). Studying more plant systems, in particular those with larger-than-X Y chromosomes, is necessary to get a more precise picture of the evolution of heteromorphic sex chromosomes in plants.

An important aspect of the evolution of animal sex chromosomes is dosage compensation, which has been reported in old heteromorphic systems (Gu & Walters, 2017). The genetic degeneration of the Y chromosome causes a global decrease of Y gene expression through gene loss and gene silencing (Bachtrog, 2013). Without compensation, this phenomenon would result in a dosage imbalance in males because many Y genes have reduced expression or are not expressed at all, compared to females where both X copies are fully expressed (Gu & Walters, 2017). Three such compensation mechanisms have been described: the mammalian system, the *Drosophila* system, and the *C. elegans* system (Ercan, 2015). The mechanism in fruit fly appears to be straightforward with a chromosome-wide doubling of X expression in males, re-establishing the proper dosage (Ercan, 2015). The mammal and *C. elegans* mechanisms are less straightforward, with an apparent doubling of expression on the X chromosome in both sexes and then a mechanism to correct expression in females/hermaphrodites (X-inactivation in mammals, downregulation of both Xs in *C. elegans*; Ercan, 2015). In both lineages, dosage compensation seems to affect certain genes only (Pessia et al., 2012, 2014; Mank, 2013; Albritton et al., 2014; Ercan, 2015; Veitia et al., 2015; Gu & Walters, 2017). In birds, dosage compensation also is local, affecting only a few dosage-sensitive genes (Arnold et al., 2008; Zimmer et al., 2016).

In plants, dosage compensation was first documented in *S. latifolia* (Muyle et al., 2012) although its existence was initially disputed as different patterns were observed for X/Y gene pairs and X-hemizygous genes (lacking Y copies) in RNA-seq-based studies (Chibalina &Filatov, 2011; Muyle et al., 2012; Bergero et al. 2015). However, a study comparing sex-linked genes found in BAC sequences and in RNA-seq data revealed that X-hemizygous genes were probably under-represented in RNA-seq-based studies of *S. latifolia* (Blavet et al.,

2015). Partial genome sequencing of *S. latifolia* confirmed this finding and provided a large set of X-hemizygous genes (Papadopoulos et al., 2015). This last study also confirmed that partial dosage compensation does exist in *S. latifolia* and that a fraction of both the X/Y gene pairs and X-hemizygous genes are compensated (Papadopoulos et al., 2015). Dosage compensation in *S. latifolia* seems to be achieved through an upregulation of the maternal X chromosome in both sexes, which is reminiscent of the scenario envisioned by Ohno (1967) for the evolution of dosage compensation in mammals (Muyle et al., 2018). The precise molecular mechanism remains unknown (Krasovec et al., 2019), but epigenetic studies suggest that the two female X chromosomes bear different epigenetic marks, implying different expression regulation (Siroky et al., 1998; Bacovsky et al., 2019). Additionally, evidence for dosage compensation has been found in *Silene otites*, which has a ZW system younger than the *S. latifolia* XY one (Martin et al., 2019), in *Rumex rothschildianus*, which has an XY_1Y_2 system that is 8-10 My old (Crowson et al., 2017), and in *Cannabis sativa*, which has an XY system that is 20-30 My old (Prentout et al., 2020). An important question is whether dosage compensation is a general feature of plant sex chromosomes.

Here, we focus on *Coccinia grandis*, a species in the Cucurbitaceae with a large Y chromosome (Kumar & Vishveshwaraiah, 1952; Sousa et al., 2012, 2016, 2017). *C. grandis* is a perennial, tropical liana that can produce fruits in the first year but can reach stem diameters of >8 cm and ages of at least 20 years. The genus comprises 25 species, all of them dioecious. It belongs to the tribe Benincaseae, where its sister genus (*Diplocyclos*) consists of four monoecious species (Holstein & Renner, 2011). The *C. grandis* Y chromosome comprises about 200 Mb, making it four times larger than the X chromosome, mainly due to the accumulation of transposable elements (TEs) and satellite repeats, resulting in a 10% difference in the size of male and female genomes (Sousa et al., 2012, 2016, 2017). Plastid

and mitochondrial-like sequences also have accumulated on the *C. grandis* Y chromosome. The repetitive fraction of the male and female genomes of *C. grandis* is mainly composed of Ty1 copia and Ty3 gypsy LTR elements. Of these elements, five are found in much greater abundance on the Y than on other chromosomes (Sousa et al., 2016). Despite this heteromorphy, the species has been estimated to be only 3.1 My old based on a molecular-clock model applied to a phylogeny with all 25 species (Holstein & Renner, 2011). *C. grandis* males and females show no morphological differences except in their flowers, but between 2 and 8% of the genes appear to be differentially expressed between males and females (Devani et al., 2017; Mohanty et al., 2017). The species is a promising system in which to study sex chromosome evolution because of its relatively small genome size (~ 1 Gb) and its phylogenetic proximity to *Cucumis* (Schaefer et al., 2009), which includes the fully sequenced *Cucumis sativus* and *Cucumis melo* genomes (Huang et al., 2009; Garcia-Mas et al., 2012; Ruggieri et al., 2018; Li et al., 2019). However, no sex-linked genes have been identified so far and no reference genome (with identified sex chromosomes) is currently available in *Coccinia*. Information about the extent of gene loss, the degradation of Y gene expression, the existence of dosage compensation, the genomic distribution of sex-biased genes is thus currently missing.

Here we identify sex-linked genes in *C. grandis* using a model-based method that we have developed and termed SEX-DETector (Muyle et al., 2016) and which uses RNA-seq data to genotype the parents and F1 individuals from a cross. For each SNP, the transmission from parents to offspring of each allele is analysed. Sex-linkage or autosomal segregation types have typical patterns that SEX-DETector is able to differentiate even when there are genotyping errors. For example, a bi-allelic SNP in which one allele is transmitted exclusively from father to sons, while the other is transmitted from both parents to all progeny will be

identified as an X/Y SNP (with the male-specific allele being the Y allele). The information of all SNPs in a gene is then combined into a probability for the gene to be sex-linked. RNA-seq-based segregation analysis is both relatively cheap and efficient, and has been applied in several plant systems in which sex-linked genes have been identified successfully, initially using empirical methods (Chibalina & Filatov, 2011; Bergero & Charledsworth, 2011; Muyle et al., 2012; Hough et al., 2014; Michalovova et al., 2015). More recently, SEX-DETector has been applied in *Cannabis sativa*, *Mercurialis annua*, *Silene*, and *Vitis vinifera* (Muyle et al., 2016, 2018; Zemp et al., 2017; Martin et al., 2019; Veltsos et al. 2019; Prentout et al. 2020; Badouin et al. 2020). Based on the detected sex-linked genes, we aimed to estimate the age of the sex chromosomes and test for Y degeneration, dosage compensation, and sex-biased genes in *Coccinia grandis*.

Material and Methods

Plant material

RNA sequencing data were obtained from a cross between a male and a female individual of the dioecious plant *C. grandis*, both grown in the experimental fields of IISER Pune, India. Seeds from the cross were collected as soon as the fruits matured. The 24 seedlings raised from these seeds took four to seven months to begin flowering, which allowed sexing of the individuals. Flower buds at early developmental stages 3-4 (defined in Ghadge et al., 2014) were sampled from plants being grown in the experimental fields. RNAs were isolated from 5 males (sons) and 5 females (daughters) from the F1 generation as well as from their parents.

RNA sequencing

The flower buds were sent to IPS2 Paris, France using RNA later ICE kits by Thermo Fisher. Total RNA was extracted from 12 flower bud samples using Agilent's spin column purification method, mRNA was isolated with Oligo-dT Beads from NEB and RNAseq libraries were constructed with the Directional Kit from NEB. Sequencing was performed at IPS2 Paris, France, with Illumina NextSeq500 following a paired-end protocol of library preparation (fragment lengths 100-150 bp, 75 bp sequenced from each end). RNA samples were checked for quality, individually tagged and sequenced (see Supplementary Table 1 for library sizes).

De novo transcriptome assembly

A reference transcriptome was built for *C. grandis* using Trinity (Haas et al., 2013) on the combined libraries of the 12 individuals (the parents and their ten offspring). For each individual, 100% identical reads, assumed to be PCR duplicates, were filtered out using the ConDeTri v2.3 trimming software (Smeds & Kunstner, 2011). Reads were then filtered out for sequencing adapters and low quality using ea-utils FASTQ processing utilities v1.04.636 (Aronesty, 2011). Cleaned reads from all male and female individuals were combined and assembled with Trinity version 2.4.0 with default settings (Haas et al., 2013). 276,225 contigs were obtained. Poly-A tails were removed from contigs using PRINSEQ v0.20.4 (Schmieder & Edwards, 2011) with parameters -trim_tail_left 5 -trim_tail_right 5. rRNA-like sequences were removed using riboPicker version 0.4.3 (Schmieder et al., 2012) with parameters -i 90 -c 50 -l 50 and the following databases: SILVA Large subunit reference database, SILVA Small subunit reference database, the GreenGenes database and the Rfam database. To ensure that X and Y gametologs are assembled in consensus contigs (required for the SEX-DETector analysis, see Muyle et al., 2016), Trinity components were merged using Cap3 (Huang & Madan, 1999), with parameter -p 90 and custom perl scripts. Coding sequences were

predicted using Trinity TransDecoder version 3.0.1 (Haas et al., 2013) and including Pfam domain searches as ORF retention criteria. This assembly included 128,904 ORFs. To avoid mapping X and Y reads on separate contigs of the same gene, we chose to work on the longest ORF predicted per Trinity isoforms, which resulted in a final set of 82,699 contigs (see Table 1). BUSCO v3.0.2 (Benchmarking Universal Single-Copy Orthologs) was used to assess the completeness of our transcriptome according to conserved gene content from the Plant Dataset (Simao et al., 2015). Results are shown for full assembly with all ORFs and for the longest ORF per Trinity isoforms, hereafter referred to as our reference transcriptome (128,904 and 82,699 contigs respectively) in Supplementary Table 2.

Functional annotation and Gene Ontology enrichment analysis

De novo annotation of our transcriptome was performed using Trinnotate v3.1.0 (Haas et al., 2013) and resulted in 59,319 annotations for 82,699 contigs (71.73%). Gene Ontology (GO) was assessed using GOSeq (Young et al., 2010) version 1.30.0 on R version 3.4.3 (2017-11-30) to identify over or under-represented GO terms (p-value cutoff = 0.05).

Inferring sex-linked contigs

The raw Illumina reads were mapped on the reference transcriptome using BWA mapping (Li et al., 2009) version 0.7.15 with following commands: bwa aln -n 5 and bwa sampe. Mapping statistics are shown in Supplementary Table 3. Mapped reads were kept with Samtools version 1.3.1 and individual genotypes were predicted with reads2snps version 2.0.64 with paralog detection (Gayral et al., 2013; Tsagkogeorga et al., 2012), option -aeb that allows alleles to have different expression levels, and -par 0 to avoid removal of paralogous positions by the paraclean program that tends to filter out X/Y SNPs.

SEX-DETector (Muyle et al., 2016) version 1.0 (3rd September 2017) was used to infer contig segregation types using a stochastic expectation maximization (SEM) algorithm. The detected SNPs were filtered using Perl scripts to retrieve the autosomal or sex-linked SNPs, when their posterior probability to be either autosomal or sex-linked was higher than 0.8. A contig was then inferred as sex-linked if its global probability of being sex-linked was higher than the probability of it being autosomal and if it at least had one sex-linked SNP without genotyping error. Amongst sex-linked genes, X-linked contigs without a detectable homologous Y-linked copy are called X-hemizygous. Sex-linked contigs with no Y expression were considered as X-hemizygous, the rest as X/Y.

Correcting mapping bias

To avoid biases towards the reference allele in expression level estimates, a second mapping was done with GSNAP (Wu & Nacu, 2010), a SNP-tolerant mapping software (see Supplementary Table 3). A SNP file of X/Y SNPs identified in the first run of SEX-DETector was produced with home-made perl scripts as described in Muyle et al. (2018). Raw Illumina reads were mapped with GSNAP version 2017-11-15 and parameters -m 10 and -N 1. Only uniquely mapped and concordant paired reads were kept for expression analysis. SEX-DETector was run a second time on this new mapping, and the new inferences were used afterwards for all analyses. No significant difference in the number of sex-linked inferences was observed: 1,196 X/Y and 168 X-hemizygous contigs were found as shown in Table 2.

Estimating the age of sex chromosomes

The X and Y ORF sequences were produced by SEX-DETector using only X/Y segregating SNPs, and pairwise dS was estimated by the codeml program implemented in the PAML suite (Yang, 2007) version 4.8 (see Supplementary Figure 1). The age of the sex chromosomes can

be obtained from the X/Y gene pairs with the highest synonymous divergence (the first to stop recombining, see Charlesworth et al., 2005; Bergero & Charlesworth 2009). To get age estimates in millions years, we used three Brassicaceae molecular clocks: 1.5×10^{-8} substitution/synonymous site/year, derived from an assumed divergence time of *Barbarea* and *Cardamine* of 6.0 MYA (Koch et al., 2000), 7.1×10^{-9} substitutions/site/generation based on spontaneous mutations in *Arabidopsis thaliana* (Ossowski et al., 2010), and 4×10^{-9} substitutions/synonymous site/year, derived from a phylogeny calibrated with six Brassicales fossils (Beilstein et al., 2010). We obtained the age estimates as follows: age (in years) = dSmax / rate, using the molecular clock of Koch et al. (2000) and Beilstein et al. (2010), and age (in number of generations) = dSmax / 2μ, using the molecular clock of Ossowski et al. (2010) and assuming a *C. grandis* generation time of 1 to 5.5 years.

Estimating gene loss

X-hemizygous genes (X-linked genes without detectable Y copies) have been used to infer the extent of gene loss on Y chromosomes. This only gives a rough idea of gene loss as X-hemizygous genes inferred by SEX-DETector comprise both genes with deleted or silenced Y copies (true lost Y genes) and genes with Y copies that are expressed in some tissues but not in the one used for RNA-seq (false lost Y genes). Also, X-hemizygous contigs are inferred by SEX-DETector from X polymorphism, as explained in Muyle et al., 2016, whereas X/Y contig inference relies on fixed mutations. X-hemizygous contigs can therefore only be detected in contigs with X polymorphism, resulting in their underestimation (Bergero & Charlesworth, 2011). We corrected for this by using the number of X-hemizygous contigs (168) relative to X/Y contigs with X polymorphism (424) that were listed in the output of SEX-DETector. Premature stop codons were detected using a custom script on X and Y alleles.

Analysis of expression level differences between X and Y alleles

Allelic expression measurement. Normalized allelic expression of sex-linked contigs was computed as in Muyle et al. (2018) from SEX-DETector outputs. Expression of X and Y alleles was computed using reads spanning diagnostic X/Y SNPs only and were normalized using the library size and the number of studied SNPs in the contig. Normalized expression levels were lower in males compared to females in autosomal contigs, as seen in Supplementary Figure 2. This may be because a small subset of genes are very highly expressed in developing male organs, as observed in other plants (e.g. Badouin et al., 2017), resulting in an apparent lower expression of housekeeping genes after normalization for total library size. We applied a correction coefficient to male expression using the ratio of median male autosomal expression over median female autosomal expression, to have a comparable median expression in males and females for autosomal contigs. The same ratio was then applied to sex-linked contigs. The results of this correction are shown in Supplementary Figures 2, 3 and 4. We used corrected expression levels to prepare Figure 1 and to perform all downstream analyses.

Analysis of dosage compensation in X-hemizygous contigs. Contigs with Xmale/2Xfemale ratios above 8 or under 0.125 were excluded as in Muyle et al. (2012) because we do not expect dosage compensation to occur in these sex-biased genes. To test for dosage compensation, we filtered genes based on their $\log_2$ male-to-female expression ratio following Mullon et al. (2015). X-hemizygous contigs that showed the same expression level in males and in females, i.e. with of $\log_2$(male/female) of 0 +/- 0.2 were considered as compensated and X-hemizygous contigs with twice as much expression in females compared to males, i.e. with of $\log_2$(male/female) of -1 +/- 0.2 were considered as non-compensated (see Figure 2). A wider range of $\log_2$(male/female) of +/- 0.5, as used in Zimmer et al. (2016) for dosage-

compensated contigs, had the same GO enrichment categories. A smaller range of +/- 0.1 had not enough contigs per category to allow for a significant GO enrichment analysis.

Identifying contigs with sex-biased expression.

We used the R packages DESeq2 (Love et al., 2014) version 1.18.1, edgeR (McCarthy et al., 2012; Robinson et al., 2010) version 3.20.8, and Limma-Voom (Law et al., 2014) version 3.34.8 in R version 3.4.3 (2017-11-30) to perform biased gene expression analysis between males and females. DGE analysis was performed on the raw read counts (untransformed, not normalized for sequencing depth). Contigs with a CPM (Count Per million) lower than 0.5 (corresponding to a count of 10) were filtered out. Regularized log-transformation of the DESeq2 package was used to reduce variance of low read counts. Normalization with edgeR was made with a tagwise dispersion and GLM normalization method (calcNormFactors, estimateTagwiseDisp and glmLRT functions). With LimmaVoom, counts were fitted to a linear model and differential expression was computed by empirical Bayes (lmFit and eBayes functions).

Contigs identified with at least two methods with an FDR cutoff of 0.0001 were retained as differentially expressed (see Supplementary Figure 5 and Supplementary Table 4). The three methods do not have the same characteristics, and keeping contigs identified with at least two methods as differentially expressed is more robust. DESeq2 and edgeR are based on the same method, but edgeR has fewer false positives. LimmaVoom is very specific (see Supplementary Figure 5), and therefore also has few false positives. The union of intersections is a way to both remove false negatives of DESeq2 and EdgeR and retain true positives that LimmaVoom tends to discard.

310 *Statistics*

311 Unless stated otherwise in the relevant sections, all statistical analyses and graphs were done
312 with R version 3.4.3 (R Core Team, 2014). Fisher's exact tests were two-tailed and p-values
313 were adjusted with the FDR method (Benjamini & Hochberg, 1995). Exact adjusted p-values
314 are provided for each test.

315

316

317 # Results

318

319 *Sex-linked genes identified by SEX-DETector*

320 We assembled a *de novo* transcriptome for *C. grandis* with male and female reads (82,699
321 contigs, see Table 1). We used RNA-seq data from a *C. grandis* F1 cross mapped to our
322 reference transcriptome to identify genes located on the sex chromosomes (Supplementary
323 Table 1). The raw reads were mapped on open reading frames (ORF). A total of 45.76% reads
324 were mapped with standard mapping and 49.24% with SNP-tolerant mapping (see
325 Supplementary Table 3). We divided the contigs expressed in buds into autosomal, sex-linked
326 X/Y (defined as contigs having both X- and Y-linked alleles), and X-hemizygous contigs
327 (sex-linked, but with no Y-copy expression). These categories were inferred from single
328 nucleotide polymorphisms (SNPs) segregating in a family, using a probabilistic model (Muyle
329 et al., 2016). Out of the 82,699 contigs, 5,070 had enough informative SNPs to be assigned to
330 a segregation type, 3,706 were inferred as autosomal (73,10% of contigs with enough
331 informative SNPs), 1,196 as X/Y (23,59%), and 168 as X-hemizygous (3,31%) (see Table 2).

332

333 *Age of the C. grandis XY system*

Age estimates are based on the divergence between X and Y copies and three Brassicaceae molecular clocks (see Table 3). We obtained a maximum dS of 0.17 in all contigs and of 0.13 in contigs longer than 1kb. The molecular clocks obtained from fossil-calibrated Brassicaceae phylogenies (Koch et al., 2000; Beilstein et al., 2010) returned age estimates between 8.7 to 34.7 My old. The molecular clock obtained from *Arabidopsis thaliana* substitution rates (Ossowski et al., 2010) returned estimates ranging from 9.3 to 12.1 My, when assuming a generation time of 1 year. The true generation time of *C. grandis*, however, is unknown. One year is the onset of sexual maturity in this plant and corresponds to a lower bond estimate of the generation time. *Coccinia grandis* can reach 20 years in the wild, and individuals of 10 years are known from botanical gardens (S.S. Renner, pers. com). Assuming an average generation time of 5.5 years (from the minimum value of 1 year and the conservative maximum value of 10 years), yields much higher age estimates for XY divergence (see Table 3).

Patterns of Y degeneration

We looked for patterns of degeneration in our data. Males showed lower gene expression than females for most of the genes, because a small subset of genes are very highly expressed in developing male flower buds resulting in apparent lower mean expression after normalization for total library size (Supplementary Figure 2). After correcting for this expression bias between males and females (see Methods), we found that sex-linked genes are less expressed in males than in females (see Supplementary Figures 2 and 4), Wilcoxon ranked test p-value = 2.57×10^{-8}). To refine our analysis of Y chromosome degeneration in *C. grandis*, we analysed the allelic expression of genes inferred as sex-linked, which showed that Y-linked alleles were significantly less expressed than X-linked alleles in males (see Figure 1, Wilcoxon ranked test p-value = 3.06×10^{-13}). Lost Y genes can be detected by SEX-DETector when the Y copy is

359 absent or unexpressed and are assigned as X-hemizygous. But given that X-hemizygous
360 contigs can be inferred from segregation patterns only if there is polymorphism on the X
361 chromosome (Bergero & Charlesworth, 2011; Muyle et al., 2016; Materials and Methods),
362 their number may be underestimated when segregation patterns are analyzed. We corrected
363 for this as described in Materials and Methods, which resulted in a corrected rate of gene loss
364 of 39.62%. Another hallmark of degeneration is the presence of premature stop codons. We
365 detected 17 X (1.4%) and 56 Y (4.7%) alleles with a premature stop, implying that premature
366 stop codons are more abundant in Y alleles (Fisher's exact test p-value = 6.9×10^{-6}). These
367 observations clearly point to a significant level of genetic degeneration on the Y chromosome
368 in *C. grandis*.
369
370 *Patterns of dosage compensation*
371 To determine whether some genes are dosage-compensated, we first studied the $\log_2$ fold
372 change between male and female expressions. In the absence of dosage compensation, the
373 Xmale/2Xfemale expression ratio is expected to be 0.5, so the $\log_2$ of the ratio is expected to
374 be -1, because males (XY) have one X-linked copy and females (XX) have two. This is what
375 we observed for contigs that do not show reduced expression of the Y-linked allele relative to
376 the X-linked allele, i.e., that have a Y/X expression ratio close to 1 (median of $\log_2$
377 Xmale/2Xfemale ratio is -1.29 for contigs with Y/X > 1; see Figure 2). For contigs with
378 reduced Y expression (low Y/X ratios), we observed a higher Xmale/2Xfemale expression
379 ratio, which suggests that dosage compensation occurs for some genes (median of contigs
380 with Y/X ≤ 0.5 is -0.85; see Figure 2). Finally, in X-hemizygous contigs, the distribution of
381 Xmale/2Xfemale expression ratio was bimodal, with a set of 31 contigs centered on a $\log_2$
382 Xmale/2Xfemale ratio of -1 (no compensation), and 23 contigs centered on 0 (total
383 compensation). This suggests that these two sets of genes exhibited respectively no

compensation and total compensation resulting in equal expression in males and females. This trend is also present in a less visible pattern for X/Y contigs with low Y expression (see Figure 2). To investigate dosage compensation further, we compared expression of X-linked and Y-linked alleles in males and females for different Y/X expression ratio categories (Figure 3), using female expression as a reference. We excluded 1% of the sex-linked contigs that showed either an elevated Y expression (high Y/X ratios) or sex-biased X expression (very high or very low Xmale/2Xfemale ratios, see Methods). The Y/X ratio was computed in *C. grandis* males and averaged between individuals, and used as a proxy for Y degeneration. In the absence of dosage compensation, X male/2X female expression ratio is expected to be 0.5. Instead, we found that X expression in males increases with decreasing Y expression, which results in similar expression levels of sex-linked contigs in both sexes and provides further evidence for dosage compensation in *C. grandis*.

X/Y genes appeared to be depleted in hormone-related functions, such as response to ethylene and negative regulation of ethylene biosynthetic process (GO:0009723 and GO:0010366). X-hemizygous genes were depleted in dosage-sensitive functions such as macromolecular complex, intracellular ribonucleoprotein complex, ribonucleoprotein complex, transferase complex, and membrane protein complex (GO:0032991, GO:0030529, GO:1990904, GO:1990234 and GO:0098796). However, these same dosage-sensitive functions were enriched in X-hemizygous genes that show dosage compensation when compared to all sex-linked genes, which suggests compensation targets dosage-sensitive genes.

Genomic distribution of sex-biased genes

We detected 3,453 sex-biased genes with edgeR, 4,881 with DESeq, and 538 with LimmaVoom (Supplementary Figure 1). To establish a robust set of sex-biased genes we

409 retained genes that were identified as sex-biased with at least two methods (3,273 genes).
410 Among these, 2,682 (81.94%) were male-biased and 591 (18.06%) female-biased. Genes with
411 sex-biased expression were significantly over-represented among sex-linked genes (see
412 Supplementary Table 4, Fisher's exact test p-value $< 2.2 \times 10^{-16}$), with 241 out of 1,364 sex-
413 linked genes being sex-biased (17.67% of sex-linked genes, with respectively 13.42% and
414 4.25% having male and female-biased expression), and 206 out of 3,706 autosomal genes
415 being sex-biased (5.56% of them, 4.45% with male and 1.11% with female-biased
416 expression). Out of the sex-biased genes that were localized on the sex chromosomes, 228
417 (94.61%) had a X/Y segregation type (181 male-biased and 47 female-biased), and only 11
418 male-biased genes and 2 female-biased were X-hemizygous. X-hemizygous contigs were not
419 enriched for differentially expressed genes when compared to autosomal contigs (see
420 Supplementary Table 4, Fisher's exact test p-value = 0.2302), which might be due to the small
421 sample of X-hemizygous genes. Gene ontology analysis revealed several biological processes
422 that are significantly over-represented among female-biased genes and under-represented
423 among male-biased genes, or vice versa. GO categories related to pollen production were
424 enriched in male-biased genes, such as pectin catabolic process, pollen wall assembly,
425 sporopollenin biosynthetic process, pollen exine formation, pectin metabolic process,
426 pollination, anther wall tapetum development, pollen sperm cell differentiation, anther
427 development, rejection of self pollen, and regulation of pollen tube growth (GO:0045490,
428 GO:0010208, GO:0080110, GO:0010584, GO:0045488, GO:0009856, GO:0048658,
429 GO:0048235, GO:0048653, GO:0060320 and GO:0080092). Functions related to hormone
430 signaling were enriched in female-biased genes, such as response to auxin, auxin-activated
431 signaling pathway, regulation of ethylene-activated signaling pathway, brassinosteroid
432 mediated signaling pathway, auxin polar transport, auxin mediated signaling pathway
433 involved in phyllotactic patterning, ethylene receptor activity, ethylene binding, auxin

434 transport, jasmonic acid and ethylene-dependent systemic resistance, and ethylene mediated
435 signaling pathway (GO:0009733, GO:0009734, GO:0010104, GO:0009742, GO:0009926,
436 GO:0060774, GO:0038199, GO:0051740, GO:0038199, GO:0051740, GO:0060918 and
437 GO:0009871). These functional enrichments suggest that sex-biased expression may have
438 evolved to support contrasting biological functions in *C. grandis* females and males.
439
440

441 ## Discussion

442

443 *C. grandis XY are of intermediate age, similarly to other highly heteromorphic plant systems*
444 The divergence between X and Y copies in *C. grandis* reaches 0.13 to 0.17, and Brassicaceae-
445 derived molecular clocks returned age estimates from 8.7 to 34.7 My for their divergence (and
446 up to >50 My old, see Table 3). By contrast, a phylogeny that included 24 species of *Coccinia*
447 and six outgroup taxa, calibrated with the divergence time of *Coccinia* and *Diplocyclos* of 15
448 ± 2.6 My, yielded a divergence time of *Coccinia grandis* from its sister species (whose sex
449 chromosomes have not been studied) of 3.1 My (Holstein & Renner, 2011). The other plant
450 species with a huge Y chromosome, *S. latifolia*, has been dated to 11.5 My old based on
451 mutation rate measurements (Krasovec et al., 2018) or instead <1 to 10 My using molecular
452 clocks (Nicolas et al., 2005: Brassicaceae and *Ipomoea* rates yield 5-10 My for the sex
453 chromosomes of *S. latifolia*; Rautenberg et al., 2012: fossil-calibrated Caryophyllaceae
454 phylogeny yields a divergence time of <1 My for *S. latifolia* and its sister clade). *Silene*
455 *latifolia* has a generation time of 1 year, and it is thus tempting to think that *C. grandis* sex
456 chromosomes should be as old or older than the ones of *S. latifolia*, unless mutation rate is a
457 lot higher in *C. grandis*, which would be unexpected (Krasovec et al., 2018).
458

C. grandis Y chromosome degeneration is moderate, with an unusually reduced Y expression

Our results suggest that not only is the Y chromosome of *C. grandis* accumulating repeats as shown previously, but it is also losing genes and becoming silent. Out of the 1,364 sex-linked genes, 168 did not have a Y-linked homolog, and we estimated a total gene loss of about 40%, which is similar to what has been found in *S. latifolia* using a methodology similar to this study and also others (Blavet et al, 2015; Papadopoulos et al., 2015; Muyle et al., 2016, 2018). The Y/X expression ratio, however, was lower than what has been found in *S. latifolia* (0.63 vs. 0.77 respectively, see Muyle et al., 2012). The difference in size between the sex chromosomes is larger in *C. grandis* compared to *S. latifolia* (Sousa et al., 2012), which might indicate that the TE load is larger in the former. TEs trigger host defence mechanisms, such as DNA methylation to silence them. This can affect genes close to TE insertions and reduce their expression level (Hollister et al., 2011; Root, 2003). The possibly higher TE abundance on the *C. grandis* Y chromosome might trigger a higher level of gene silencing and thus explain its unusually low Y/X expression ratio. To test this idea further, the full sequences of the sex chromosomes in *C. grandis* and a DNA methylation study of TEs and neighboring genes would be needed.

C. grandis exhibit sex chromosome dosage compensation, a phenomemon observed in several plant systems

When analysing the expression of X and Y copies in both sexes, we found that the reduction of the Y copy was compensated by increasing expression of the X copy to maintain similar expression of the pair in both sexes (Figure 3). This suggests that dosage compensation has evolved in *C. grandis*. Again, a very similar pattern has been found in *S. latifolia* (Muyle et al., 2012). Compensation is probably partial, and not all the sex-linked genes are compensated (Figure 2), something also observed in *S. latifolia* (Papadopoulos et al., 2015). We mentioned

above that X-hemizygous genes are underrepresented in RNA-seq-based studies such as this one, and conclusions about dosage compensation are therefore more difficult to draw for this category of genes (discussed in Blavet et al., 2015). Here we found that the signal of dosage compensation was weak in X-hemizygous genes when taken together (Figure 3) as observed in *S. latifolia* with the same approach (Muyle et al., 2012). However, when looking at genes individually as in Papadopoulos et al., (2015), we found evidence that some X-hemizygous genes are fully compensated (see bimodal distribution in Figure 2D). Strickingly, those genes are enriched in dosage-sensitive functions in agreement with findings in animals (Pessia et al., 2012, 2014; Mank, 2013; Albritton et al., 2014; Ercan, 2015; Veitia et al., 2015; Zimmer et al., 2016; Gu & Walters, 2017). Evidence for dosage compensation has been found so far in Caryophyllaceae, Cannabaceae, and Polygonaceae (Muyle et al., 2012, 2018; Papadopolous et al., 2015; Crowson et al., 2017; Martin et al., 2019; Prentout et al., 2020). Cucurbitaceae are thus the fourth plant family in which dosage compensation is documented. Future work is needed to find out whether dosage compensation relies on genomic imprinting as in *S. latifolia* (Muyle et al., 2018) or on a different mechanism.

Coccinia grandis sex chromosomes are enriched in sex-biased genes

We found that 4% of the genes expressed in *C. grandis* floral buds are sex-biased (Supplementary Table 4), with a total of 3,273 sex-biased genes identified (2,682 male-biased and 591 female-biased), in agreement with prior studies on sex-biased genes in *C. grandis* (Devani et al., 2017; Mohanty et al., 2017). Our results support enrichment for pollen production-related functions in male-biased genes as previously found in *C. grandis* (Devani et al., 2017; Mohanty et al., 2017) and in other plant systems (Muyle et al. 2017). Female-biased genes, on the other hand, were significantly enriched for hormone-signaling functions. Male-biased genes are also significantly more numerous than female-biased genes, a pattern

509 that is common in dioecious plants (Harkess et al. 2015; Zemp et al., 2016; Muyle et al.,
510 2017; Cossard et al., 2019, but see Darolti et al., 2018; Sanderson et al., 2019; Muyle, 2019).
511 Lastly, sex-biased genes were found both on autosomes and on sex chromosomes but the
512 latter were significantly enriched in such genes, again a common pattern in dioecious plants
513 (for review, see Muyle et al., 2017; see also Darolti et al., 2018; Sanderson et al., 2019). The
514 most probable hypothesis is that sexually antagonistic selection favors sex linkage of sex-
515 biased genes involved in sexual dimorphism, but further work to identify the footprints of
516 sexually antagonistic selection (such as in Zemp et al., 2016) will be needed to test this idea in
517 *C. grandis*.
518

519 ## Acknowledgments

520 This project was supported through the Agence Nationale de la Recherche (ANR) grant ANR-
521 14-CE19-0021-01 to GABM and AB.
522

523 ## Author contributions

524 The study was conceived and coordinated by GABM with input from AB and AKB. Funding
525 was secured by GABM, AB and SSR. The cross was done by RSD and AKB. Samples were
526 collected by RSD and AKB. Sequencing was done by DL and AB. CF and AM built the
527 reference transcriptome. HB performed the GO annotation. BR provided scripts and helped
528 interpreting the differential expression analysis. CF made all subsequent analyses and
529 prepared the figures. Data analysis and interpretation were done by CF, HB and GABM. CF,
530 GABM and SSR wrote the manuscript with input from all authors.
531

532 ## References

533 Albritton, S. E., Kranz, A. L., Rao, P., Kramer, M., Dieterich, C., & Ercan, S. (2014). Sex-biased gene
534 expression and evolution of the x chromosome in nematodes. *Genetics, 197*(3), 865-883.
535 Arnold, A. P., Itoh, Y., & Melamed, E. (2008). A bird's-eye view of sex chromosome dosage
536 compensation. *Annu. Rev. Genomics Hum. Genet., 9*, 109-127.
537 Aronesty, E. (2011). Ea-utils: "Command-line tools for processing biological sequencing data".
538 *Available Online at:* https://expressionanalysis.github.io/ea-utils/
539 Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome
540 degeneration. *Nature Reviews Genetics, 14*(2), 113-124.
541 Bačovský, V., Houben, A., Kumke, K., & Hobza, R. (2019). The distribution of epigenetic histone
542 marks differs between the X and Y chromosomes in Silene latifolia. *Planta, 250*(2), 487-494.
543 Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., … Langlade, N. B. (2017).
544 The sunflower genome provides insights into oil metabolism, flowering and asterid evolution.
545 *Nature*, *546*(7656), 148–152. doi:10.1038/nature22380
546 Badouin, H., Velt, A., Gindraud, F., Flutre, T., Dumas, V., Vautrin, S., ... & Santoni, S. et al. (2020).
547 The wild grape genome sequence provides insights into the transition from dioecy to
548 hermaphroditism during grape domestication. *bioRxiv*.
549 Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R., & Mathews, S. (2010).
550 Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. *Proc. Natl.*
551 *Acad. Sci. U.S.A.*, *107*(43), 18724–18728.
552 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful
553 approach to multiple testing. *Journal of the Royal statistical society: series B*
554 *(Methodological), 57*(1), 289-300.
555 Bergero, R., & Charlesworth, D. (2011). Preservation of the Y transcriptome in a 10-million-year-old
556 plant sex chromosome system. *Curr. Biol.*, *21*(17), 1470–1474.
557 Bergero, R, Charlesworth, D., Filatov, D. a, & Moore, R. C. (2008). Defining regions and
558 rearrangements of the Silene latifolia Y chromosome. *Genetics*, *178*(4), 2045–53.
559 Bergero, R., Qiu, S., & Charlesworth, D. (2015). Gene loss from a plant sex chromosome system.
560 *Current Biology*, *25*(9), 1234–1240.

561 Bergero, Roberta, & Charlesworth, D. (2009). The evolution of restricted recombination in sex
562 chromosomes. *Trends in Ecology & Evolution*, *24*(2), 94–102. doi:10.1016/j.tree.2008.09.010
563 Blavet, N., Blavet, H., Muyle, A., Käfer, J., Cegan, R., Deschamps, C., … Marais, G. (2015).
564 Identifying new sex-linked genes through BAC sequencing in the dioecious plant *Silene*
565 *latifolia*. *BMC Genomics*, *16*(1), 546.
566 Cermak, T., Kubat, Z., Hobza, R., Koblizkova, A., Widmer, A., Macas, J., ... & Kejnovsky, E. (2008).
567 Survey of repetitive sequences in Silene latifolia with respect to their distribution on sex
568 chromosomes. *Chromosome Research, 16*(7), 961-976.
569 Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex
570 chromosomes. *Heredity*, 95(2), 118-128.
571 Chibalina, M. V., & Filatov, D. A. (2011). Plant Y chromosome degeneration is retarded by haploid
572 purifying selection. *Curr. Biol.*, *21*(17), 1475–1479.
573 Cossard, G. G., Toups, M. A., & Pannell, J. R. (2019). Sexual dimorphism and rapid turnover in gene
574 expression in pre-reproductive seedlings of a dioecious herb. *Annals of botany, 123*(7), 1119-
575 1131.
576 Crowson, D., Barrett, S. C. H., & Wright, S. I. (2017). Purifying and Positive Selection Influence
577 Patterns of Gene Loss and Gene Expression in the Evolution of a Plant Sex Chromosome
578 System. *Molecular Biology and Evolution*, *34*(5), 1140–1154.
579 Darolti, I., Wright, A. E., Pucholt, P., Berlin, S., & Mank, J. E. (2018). Slow evolution of sex-biased
580 genes in the reproductive tissue of the dioecious plant Salix viminalis. *Molecular ecology,*
581 *27*(3), 694-708.
582 Devani, R. S., Sinha, S., Banerjee, J., Sinha, R. K., Bendahmane, A., & Banerjee, A. K. (2017). De
583 novo transcriptome assembly from flower buds of dioecious, gynomonoecious and chemically
584 masculinized female Coccinia grandis reveals genes associated with sex expression and
585 modification. *BMC Plant Biol.*, *17*(1), 241.
586 Ercan, S. (2015). Mechanisms of x chromosome dosage compensation. *J Genomics*, *3*, 1–19.
587 Filatov, D. A., Monéger, F., Negrutiu, I., & Charlesworth, D. (2000). Low variability in a Y-linked
588 plant gene and its implications for Y-chromosome evolution. *Nature*, *404*(6776), 388-390.

589 Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V. M., ... & Alioto, T.
590 (2012). The genome of melon (Cucumis melo L.). *Proceedings of the National Academy of*
591 *Sciences, 109*(29), 11872-11877.
592 Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., … Galtier, N. (2013).
593 Reference-free population genomics from next-generation transcriptome data and the
594 vertebrate-invertebrate gap. *PLoS Genetics*, *9*(4), e1003457.
595 Ghadge, A. G., Karmakar, K., Devani, R. S., Banerjee, J., Mohanasundaram, B., Sinha, R. K., …
596 Banerjee, A. K. (2014). Flower development, pollen fertility and sex expression analyses of
597 three sexual phenotypes of Coccinia grandis. *BMC Plant Biology*, *14*(1), 325.
598 Gu, L., & Walters, J. R. (2017). Evolution of sex chromosome dosage compensation in animals: a
599 beautiful theory, undermined by facts and bedeviled by details. *Genome biology and*
600 *evolution, 9*(9), 2461-2476.
601 Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... & MacManes,
602 M. D. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity
603 platform for reference generation and analysis. *Nature protocols, 8*(8), 1494.
604 Harkess, A., Mercati, F., Shan, H. Y., Sunseri, F., Falavigna, A., & Leebens-Mack, J. (2015).
605 Sex-biased gene expression in dioecious garden asparagus (Asparagus officinalis). *New*
606 *Phytologist, 207*(3), 883-892.
607 Hollister, J. D., Smith, L. M., Guo, Y. L., Ott, F., Weigel, D., & Gaut, B. S. (2011). Transposable
608 elements and small RNAs contribute to gene expression divergence between Arabidopsis
609 thaliana and Arabidopsis lyrata. *Proceedings of the National Academy of Sciences, 108*(6),
610 2322-2327.
611 Holstein, N., & Renner, S. S. (2011). A dated phylogeny and collection records reveal repeated biome
612 shifts in the African genus Coccinia (Cucurbitaceae). *BMC Evolutionary Biology*, *11*(1), 28.
613 Hough, J., Hollister, J. D., Wang, W., Barrett, S. C., & Wright, S. I. (2014). Genetic degeneration of
614 old and young Y chromosomes in the flowering plant Rumex hastatulus. *Proc. Natl. Acad.*
615 *Sci. U.S.A.*, *111*(21), 7713–7718.

616 Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., ... & Ren, Y. (2009). The genome of the
617 cucumber, Cucumis sativus L. *Nature genetics, 41*(12), 1275-1281.
618 Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, *9*(9),
619 868–877.
620 Krasovec, M., Chester, M., Ridout, K., & Filatov, D. A. (2018). The mutation rate and the age of the
621 sex chromosomes in Silene latifolia. *Current Biology, 28*(11), 1832-1838.
622 Krasovec, M., Kazama, Y., Ishii, K., Abe, T., & Filatov, D. A. (2019). Immediate Dosage
623 Compensation Is Triggered by the Deletion of Y-Linked Genes in Silene latifolia. *Current*
624 *Biology, 29*(13), 2214-2221.
625 Koch, M. A., Haubold, B., & Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone
626 synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera
627 (Brassicaceae). *Mol. Biol. Evol.*, *17*(10), 1483–1498.
628 Kumar, L. S., & Vishveshwaraiah, S. (1952). Sex mechanism in Coccinia indica Wight and Arn.
629 *Nature* 170: 330–331.
630 Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model
631 analysis tools for RNA-seq read counts. *Genome Biol.*, *15*(2), R29.
632 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform.
633 *bioinformatics, 25*(14), 1754-1760.
634 Li, Q., Li, H., Huang, W., Xu, Y., Zhou, Q., Wang, S., ... & Zhang, Z. (2019). A chromosome-scale
635 genome assembly of cucumber (Cucumis sativus L.). *GigaScience, 8*(6), giz072.
636 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for
637 RNA-seq data with DESeq2. *Genome Biol.*, *15*(12), 550.
638 Mahajan, S., Wei, K. H. C., Nalley, M. J., Gibilisco, L., & Bachtrog, D. (2018). De novo assembly of
639 a young Drosophila Y chromosome using single-molecule sequencing and chromatin
640 conformation capture. *PLoS biology, 16*(7), e2006348.
641 Marais, G. A., Nicolas, M., Bergero, R., Chambrier, P., Kejnovsky, E., Monéger, F., ... &
642 Charlesworth, D. (2008). Evidence for degeneration of the Y chromosome in the dioecious
643 plant Silene latifolia. *Current Biology*, *18*(7), 545-549.

644 Martin, H., Carpentier, F., Gallina, S., Godé, C., Schmitt, E., Muyle, A., ... & Touzet, P. (2019).
645 Evolution of young sex chromosomes in two dioecious sister plant species with distinct sex
646 determination systems. *Genome biology and evolution, 11*(2), 350-361.
647 Mank, J. E. (2013). Sex chromosome dosage compensation: definitely not for everyone. *Trends*
648 *Genet.*, *29*(12), 677–683.
649 McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor
650 RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.*, *40*(10), 4288–
651 4297.
652 Michalovova, M., Kubat, Z., Hobza, R., Vyskot, B., & Kejnovsky, E. (2015). Fully automated pipeline
653 for detection of sex linked genes using RNA-Seq data. *BMC bioinformatics, 16*(1), 78.
654 Ming, R., Bendahmane, A., & Renner, S. S. (2011). Sex chromosomes in land plants. *Annual Review*
655 *of Plant Biology*, *62*(1), 485–514.
656 Mohanty, J. N., Nayak, S., Jha, S., & Joshi, R. K. (2017). Transcriptome profiling of the floral buds
657 and discovery of genes related to sex-differentiation in the dioecious cucurbit Coccinia grandis
658 (L.) Voigt. *Gene*, *626*, 395–406.
659 Mullon, C., Wright, A. E., Reuter, M., Pomiankowski, A., & Mank, J. E. (2015). Evolution of dosage
660 compensation under sexual selection differs between X and Z chromosomes. *Nat Commun*, *6*,
661 7720.
662 Muyle, A. (2019). How different is the evolution of sex-biased gene expression between plants and
663 animals? A commentary on:'Sexual dimorphism and rapid turnover in gene expression in pre-
664 reproductive seedlings of a dioecious herb'. *Annals of botany, 123*(7), iv-v.
665 Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., & Marais, G. A. (2012). Rapid de
666 novo evolution of X chromosome dosage compensation in Silene latifolia, a plant with young
667 sex chromosomes. *PLoS Biol.*, *10*(4), e1001308.
668 Muyle, A., Zemp, N., Fruchard, C., Cegan, R., Vrana, J., Deschamps, C., … Marais, G. A. B. (2018).
669 Genomic imprinting mediates dosage compensation in a young plant XY system. *Nature*
670 *Plants*, *4*(9), 677–680. doi:10.1038/s41477-018-0221-y

671 Muyle, A., Käfer, J., Zemp, N., Mousset, S., Picard, F., & Marais, G. A. (2016). SEX-DETector: A
672 probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biology*
673 *and Evolution*, *8*(8), 2530–2543. doi:10.1093/gbe/evw172
674 Muyle, Aline, Shearn, R., & Marais, G. A. (2017). The Evolution of Sex Chromosomes and Dosage
675 Compensation in Plants. *Genome Biology and Evolution*, *9*(3), 627–645.
676 doi:10.1093/gbe/evw282
677 Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., … Moneger, F. (2005).
678 A gradual process of recombination restriction in the evolutionary history of the sex
679 chromosomes in dioecious plants. *PLoS Biol.*, *3*(1), e4.
680 Ohno, S. (1967) *Sex Chromosomes and Sex-linked Genes*. Springer, Berlin.
681
682 Ossowski, S., Schneeberger, K., Lucas-Lledo, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., …
683 Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in Arabidopsis
684 thaliana. *Science*, *327*(5961), 92–94.
685 Papadopulos, A. S. T., Chester, M., Ridout, K., & Filatov, D. A. (2015). Rapid Y degeneration and
686 dosage compensation in plant sex chromosomes. *Proceedings of the National Academy of*
687 *Sciences*, *112*(42), 13021–13026.
688 Pessia, E., Engelstadter, J., & Marais, G. A. (2014). The evolution of X chromosome inactivation in
689 mammals: the demise of Ohno's hypothesis? *Cell. Mol. Life Sci.*, *71*(8), 1383–1394.
690 Pessia, E., Makino, T., Bailly-Bechet, M., McLysaght, A., & Marais, G. A. (2012). Mammalian X
691 chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive
692 genes on the X chromosome. *Proceedings of the National Academy of Sciences, 109*(14),
693 5346-5351.
694 Prentout, D., Razumova, O., Rhoné, B., Badouin, H., Henri, H., Feng, C., ... & Marais, G. A. (2020).
695 An efficient RNA-seq-based segregation analysis identifies the sex chromosomes of *Cannabis*
696 *sativa*. *Genome Research, 30*(2), 164-172.

Rautenberg, A., Sloan, D. B., Aldén, V., & Oxelman, B. (2012). Phylogenetic relationships of Silene multinervia and Silene section Conoimorpha (Caryophyllaceae). *Systematic Botany, 37*(1), 226-237.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Renner, S. S. (2014). The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany*, *101*(10), 1588–1596.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. doi:10.1093/bioinformatics/btp616

Rodriguez Lorenzo, J. L. R., Hobza, R., & Vyskot, B. (2018). DNA methylation and genetic degeneration of the Y chromosome in the dioecious plant Silene latifolia. *BMC genomics, 19*(1), 540.

Root, G. (2003). Evolution of dioecy and sex chromosomes via methylation driving Muller's ratchet. *Biological Journal of the Linnean Society*, *80*(2), 353–368.

Ruggieri, V., Alexiou, K. G., Morata, J., Argyris, J., Pujol, M., Yano, R., ... & Benhamed, M. (2018). An improved assembly and annotation of the melon (Cucumis melo L.) reference genome. *Scientific reports, 8*(1), 1-9.

Sanderson, B. J., Wang, L., Tiffin, P., Wu, Z., & Olson, M. S. (2019). Sex-biased gene expression in flowers, but not leaves, reveals secondary sexual dimorphism in Populus balsamifera. *New Phytologist, 221*(1), 527-539.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, *27*(6), 863–864.

Schmieder, R., Lim, Y. W., & Edwards, R. (2012). Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*, *28*(3), 433–435.

Schaefer, H., Heibl, C., & Renner, S. S. (2009). Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proceedings of the Royal Society B: Biological Sciences, 276*(1658), 843-851.

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. doi:10.1093/bioinformatics/btv351

Siroky, J., Ruffini Castiglione, M., & Vyskot, B. (1998). DNA methylation patterns of Melandrium album chromosomes. *Chromosome Research*, 6(6), 441-446.

Smeds, L., & Kunstner, A. (2011). ConDeTri–a content dependent read trimmer for Illumina data. *PLoS ONE*, *6*(10), e26314.

Sousa, A., Bellot, S., Fuchs, J., Houben, A., & Renner, S. S. (2016). Analysis of transposable elements and organellar DNA in male and female genomes of a species with a huge Y chromosome reveals distinct Y centromeres. *Plant Journal*, *88*(3), 387–396.

Sousa, A., Fuchs, J., & Renner, S. S. (2012). 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*.

Sousa, A., Fuchs, J., & Renner, S. S. (2017). Cytogenetic comparison of heteromorphic and homomorphic sex chromosomes in Coccinia (Cucurbitaceae) points to sex chromosome turnover. *Chromosome Research*, *25*(2), 191–200.

Tsagkogeorga, G., Cahais, V., & Galtier, N. (2012). The population genomics of a fast evolver: High levels of diversity, functional constraint, andmolecular adaptation in the tunicate Ciona intestinalis. *Genome Biology and Evolution*, *4*(8), 740–749.

Veitia, R. A., Veyrunes, F., Bottani, S., & Birchler, J. A. (2015). X chromosome inactivation and active X upregulation in therian mammals: facts, questions, and hypotheses. *J Mol Cell Biol*, *7*(1), 2–11.

Veltsos, P., Ridout, K. E., Toups, M. A., González-Martínez, S. C., Muyle, A., Emery, O., ... & Marais, G. A. (2019). Early sex-chromosome evolution in the diploid dioecious plant *Mercurialis annua*. *Genetics, 212*(3), 815-835.

751 Wu, T. D., & Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in
752 short reads. *Bioinformatics, 26*(7), 873-881.
753 Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and*
754 *Evolution*, *24*(8), 1586–1591. doi:10.1093/molbev/msm088
755 Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for
756 RNA-seq: accounting for selection bias. *Genome Biol.*, *11*(2), R14.
757 Zemp, N., Tavares, R., Muyle, A., Charlesworth, D., Marais, G. A. B., & Widmer, A. (2016).
758 Evolution of sex-biased gene expression in a dioecious plant. *Nature Plants*, *2*(11), 16168.
759 Zimmer, F., Harrison, P. W., Dessimoz, C., & Mank, J. E. (2016). Compensation of Dosage-Sensitive
760 Genes on the Chicken Z Chromosome. *Genome Biology and Evolution*, *8*(4), 1233–1242.
761

Tables

Table 1: Transcriptome assembly statistics of *C. grandis* flower buds. Statistics in the final Trinity transcriptome and the working transcriptome containing the longest ORF predicted per Trinity isoform.

	Full Transcriptome	Longest ORF per isoform
Total contigs	128,904	82,699
Total assembled bases (bp)	103,275,123	27,290,670
Median contig length	552	836
Average contig length	801.18	836.83
Maximum contig length	16,296	16,296
Minimum contig length	297	297
N50	1,029	1,086
Total contigs longer than 1kb	30,795	21,587
GC content (%)	42.96	42.96

Table 2: Results of the SEX-DETector pipeline on the *C. grandis* dataset. Number of contigs assigned by SEX-DETector to an autosomal, X/Y or X-hemizygous segregation type (see Methods) before and after SNP-tolerant mapping (see Methods).

	SEX-DETector with BWA mapping	SEX-DETector with GSNAP SNP-tolerant mapping
Contigs in final assembly	82,699	82,699
Contigs with enough coverage to be studied	82,689	70,298
Contigs with enough informative SNPs to compute a segregation probability	4,320	3,801
Contigs assigned to an autosomal segregation type	2,889	3,706
Contigs assigned to a X-Y segregation type	1,239	1,196
Contigs assigned to a X-hemizygous segregation type	192	168

770

771 **Table 3: age estimates of the *C. grandis* XY system.** These estimates were obtained using
772 the maximum synonymous divergence between X and Y chromosomes and several molecular
773 clocks from Brassicaceae (see Methods). Estimates are shown in increasing order.

Molecular clocks	Age estimates of the sex chromosomes, with dS max = 0.17	Age estimates of the sex chromosomes, with dS max = 0.13
Koch et al. (2000), calibrated with an assumed divergence time of *Barbarea* and *Cardamine* of 6.0 My	11.3	8.7
Ossowski et al. (2010), generation time = 1 year, assumed for *Arabidopsis thaliana*	12.1	9.3
Ossowski et al. (2010), generation time = 1.5 year	18.2	13.9
Beilstein et al. (2010), calibrated with six Brassicales fossils	34.7	26.5
Ossowski et al. (2010), generation time = 5.5 year	66.8	51.1

774

775

775 Figure legends

776

777 **Figure 1: Y/X expression ratio in *C. grandis***. Distribution of normalized expression ratio
778 between X and Y alleles. Total Y and X read numbers were summed at sex-linked SNP
779 locations for each contig and normalized for each male separately (see Methods), then
780 averaged across males to obtain the Y/X ratio. The median is shown in red.

781

782 **Figure 2: X expression in males versus females in *C. grandis*.** Distribution of the ratio
783 between the expression of the single X in males and the two X copies in females (log_2
784 Xmale/2Xfemale) for all sex-linked contigs. Distributions are shown for Y/X expression ratio
785 categories in males: (A) Y/X>1; (B) X/Y in]0.5-1]; (C) X/Y in]0-0.5]; (D) X-hemizygous
786 genes (no Y copy expression, see Methods). Total X read numbers were summed at sex-
787 linked SNP locations in each contig and normalized for each individual separately, then
788 averaged among males and females to get the Xmale/2Xfemale ratio (see Methods).
789 Distribution is shown in log_2 scale with its density curve. Contigs with Xmale/2Xfemale
790 ratios above 8 or under 0.125 were excluded, which reduced the dataset to 1,351 sex-linked
791 contigs. Sample sizes are: 0, 168; 0–0.5, 460 (8 outliers); 0.5–1, 389 (2 outliers); >1, 334 (3
792 outliers). Medians are indicated for each Y/X ratio category.

793

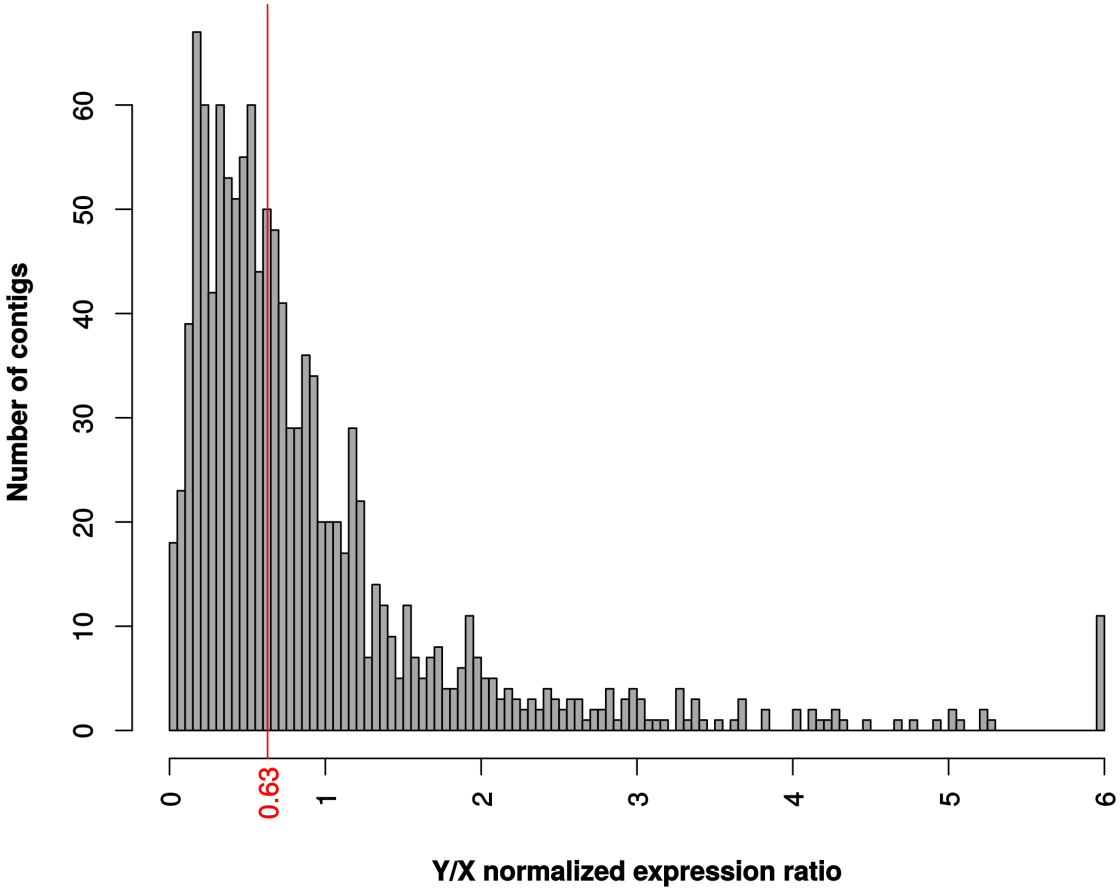
794

Figure 3: Allele-specific expression of sex-linked genes in both sexes in *C. grandis*. Expression levels of sex-linked contigs in both sexes are shown for different Y/X expression ratio categories. Total read numbers were summed at sex-linked SNP locations in each contig and normalized for each individual separately; medians for all contigs and individuals of the same sex were then obtained. XX females, median expression level of both X-linked alleles in females; X males, median expression level of the single X-linked allele in males; Y males, median expression level of the Y-linked allele in males; XY males, median expression level of the X-linked plus Y-linked alleles in males. To compare different Y/X expression ratio categories, medians were normalized using the XX expression levels in females. Sample sizes are: 0, 168; 0–0.25, 207; 0.25–0.5, 261; 0.5–0.75, 243; 0.75–1, 148; 1–1.25, 108. Error bars indicate 95% confidence intervals of the median.

806 Figures

807

808 Fig. 1



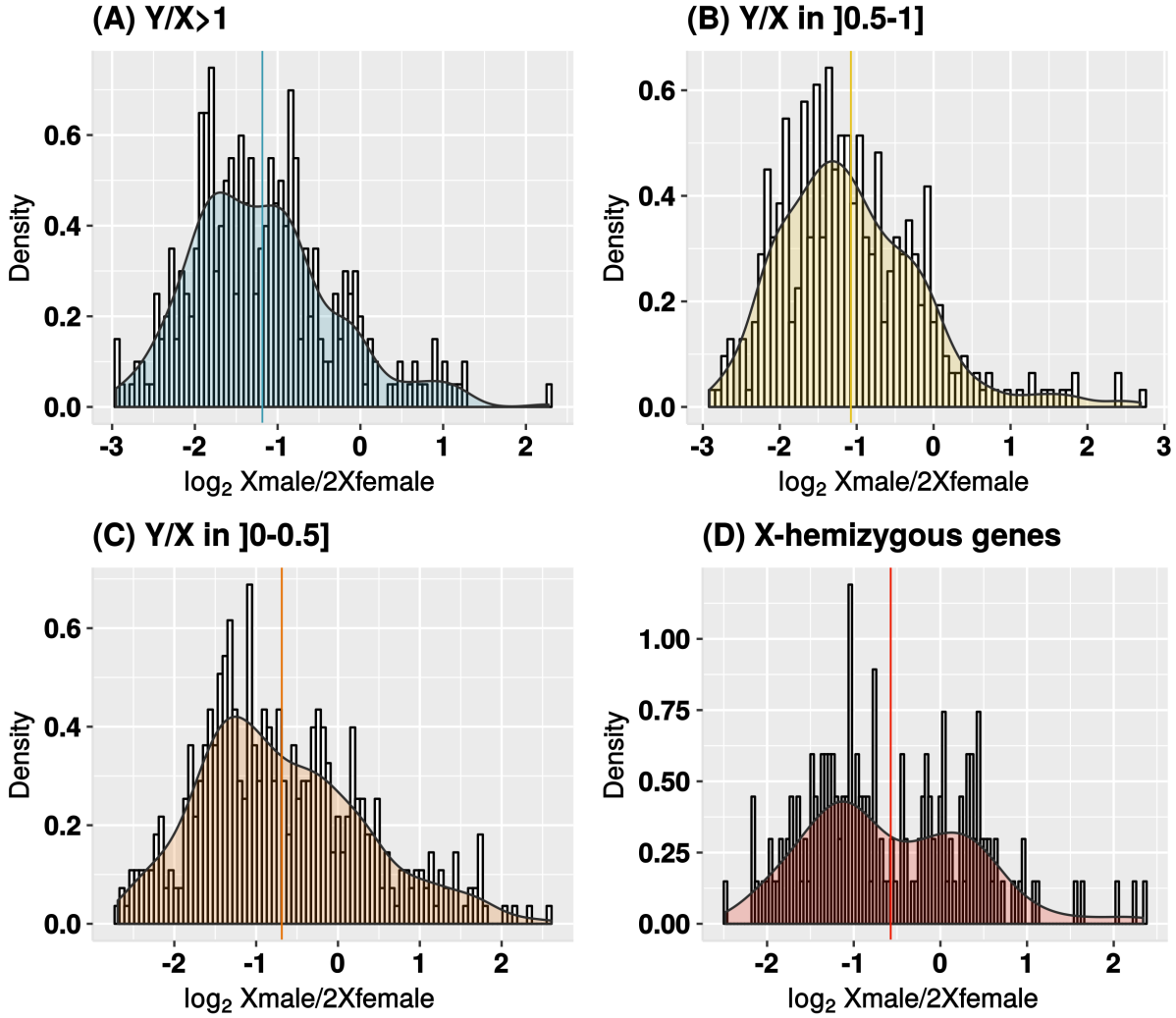
809

810

811

811

812 Fig. 2

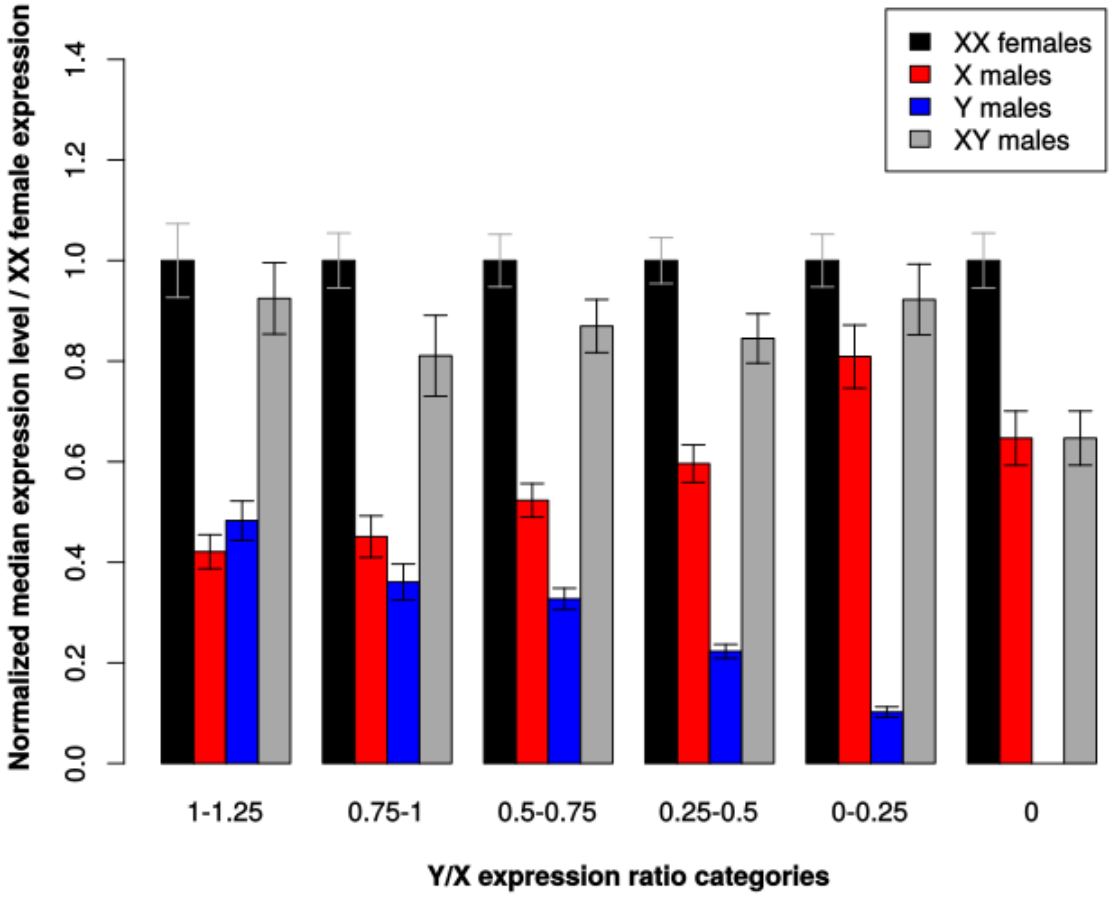


813

814

814

815 Fig. 3



816

817

818

Supplementary Material

Supplementary Table 1: Library sizes and statistics for RNA sequencing from *C. grandis* male and female flower buds. Data characteristics before and after filtering out for sequencing adapters and low quality.

	Total number reads	GC (%)	Total bases	Duplicates (%)	Total number reads	GC (%)	Total bases	Duplicates (%)
C. grandis 1 – father	71,230,290	45.55	5,413,502,040	24.85	62,168,386	45.36	4,724,687,045	19.93
C. grandis 2 – mother	81,477,288	43.86	6,192,273,888	25.94	67,354,070	43.55	5,118,780,884	19.20
C. grandis 3 – son	74,106,838	42.40	5,632,119,688	39.80	53,535,608	42.02	4,068,592,348	30.21
C. grandis 4 – son	85,096,602	42.67	6,467,341,752	40.49	58,127,652	42.34	4,417,567,083	30.05
C. grandis 5 – son	76,653,508	43.63	5,825,666,608	37.12	56,078,580	43.14	4,261,849,202	27.72
C. grandis 6 – son	77,798,734	44.55	5,912,703,784	35.52	58,797,412	44.19	4,468,480,916	26.74
C. grandis 7 – son	72,948,152	43.55	5,544,059,552	36.73	55,608,456	43.27	4,226,126,310	27.83
C. grandis 8 – daughter	95,001,666	45.20	7,220,126,616	25.83	79,158,198	44.69	6,015,873,957	18.45
C. grandis 9 – daughter	76,074,360	44.79	5,781,651,360	20.00	67,478,518	44.44	5,128,249,651	14.98
C. grandis 10 – daughter	83,484,940	44.85	6,344,855,440	20.44	73,348,162	44.50	5,574,328,675	15.09
C. grandis 11 – daughter	77,883,860	44.74	5,919,173,360	20.94	73,348,162	44.36	5,194,162,562	15.44
C. grandis 12 – daughter	85,273,198	44.66	6,480,763,048	18.62	75,439,724	44.42	5,733,285,393	14.08
Total	957,029,436	44.20	72,734,237,136	28.86	780,442,928	43.86	58,931,984,026	21.64

Supplementary Table 2: BUSCO results for *C. grandis* flower bud transcriptomes. Assessment of transcriptome assembly completeness based on gene content from near-universal single copy orthologs. Contents of the final Trinity transcriptome and the working transcriptome containing the longest ORF predicted per Trinity isoform.

	Full Transcriptome	%	Longest ORF per isoform	%
Complete	1,115	77.43	1,033	71.74
Complete and single-copy	530	36.81	557	38.68
Complete and duplicated	585	40.63	476	33.06
Fragmented	176	12.22	193	13.40
Missing	149	10.35	214	14.86
Total groups searched	1,440		1,440	

Supplementary Table 3: Mapping statistics of all the samples. Number of reads from the *C. grandis* individuals mapping to the reduced trancriptome (the longest ORF predicted per Trinity isoform, see Methods) with BWA and with GSNAP, a SNP-tolerant mapper. The male offspring libraries had higher PCR duplication rates than the female libraries (35.52-0.49% and 18.62-25.83% of raw reads respectively, see Supplementary Table 1) resulting in lower mapping rates (36.49 compared to 49.42% with SNP-tolerant mapping).

		BWA		GSNAP	
	#reads	#mapped reads	% mapping	#mapped reads	% mapping
C. grandis 1 – father	71,230,290	38,454,717	53.99	41,111,554	57.72
C. grandis 2 – mother	81,477,288	34,954,495	42.90	37,572,482	46.11
C. grandis 3 – son	74,106,838	24,609,806	33.21	27,041,875	36.49
C. grandis 4 – son	85,096,602	31,121,997	36.57	33,923,437	39.86
C. grandis 5 – son	76,653,508	29,660,482	38.69	32,261,535	42.09
C. grandis 6 – son	77,798,734	35,416,940	45.52	38,445,259	49.42
C. grandis 7 – son	72,948,152	30,938,500	42.41	33,763,165	46.28
C. grandis 8 – daughter	95,001,666	45,792,447	48.20	49,016,666	51.60
C. grandis 9 – daughter	76,074,360	38,958,015	51.21	41,562,663	54.63
C. grandis 10 – daughter	83,484,940	43,255,348	51.81	46,118,479	55.24
C. grandis 11 – daughter	77,883,860	39,070,638	50.17	41,756,525	53.61
C. grandis 12 – daughter	85,273,198	45,684,988	53.57	48,644,666	57.05
Total	957,029,436	437,918,373	45.76	471,218,306	49.24

835

836 **Supplementary Table 4: Number of Differentially Expressed Genes (DEG) per gene**

837 **categories**.

	All		Autosomal		Sex-linked		X/Y		X-hemizygous	
	# contigs	%	# contigs	%	# contigs	%	# contigs	%	# contigs	%
All	82,699		3,706		1,364		1,196		168	
Differentially expressed	3,273	3.96	206	5.56	241	17.67	228	19.06	13	7.74
Male-biased	2,682	3.24	165	4.45	183	13.42	181	15.13	2	1.19
Female-biased	591	0.71	41	1.11	58	4.25	47	3.93	11	6.55

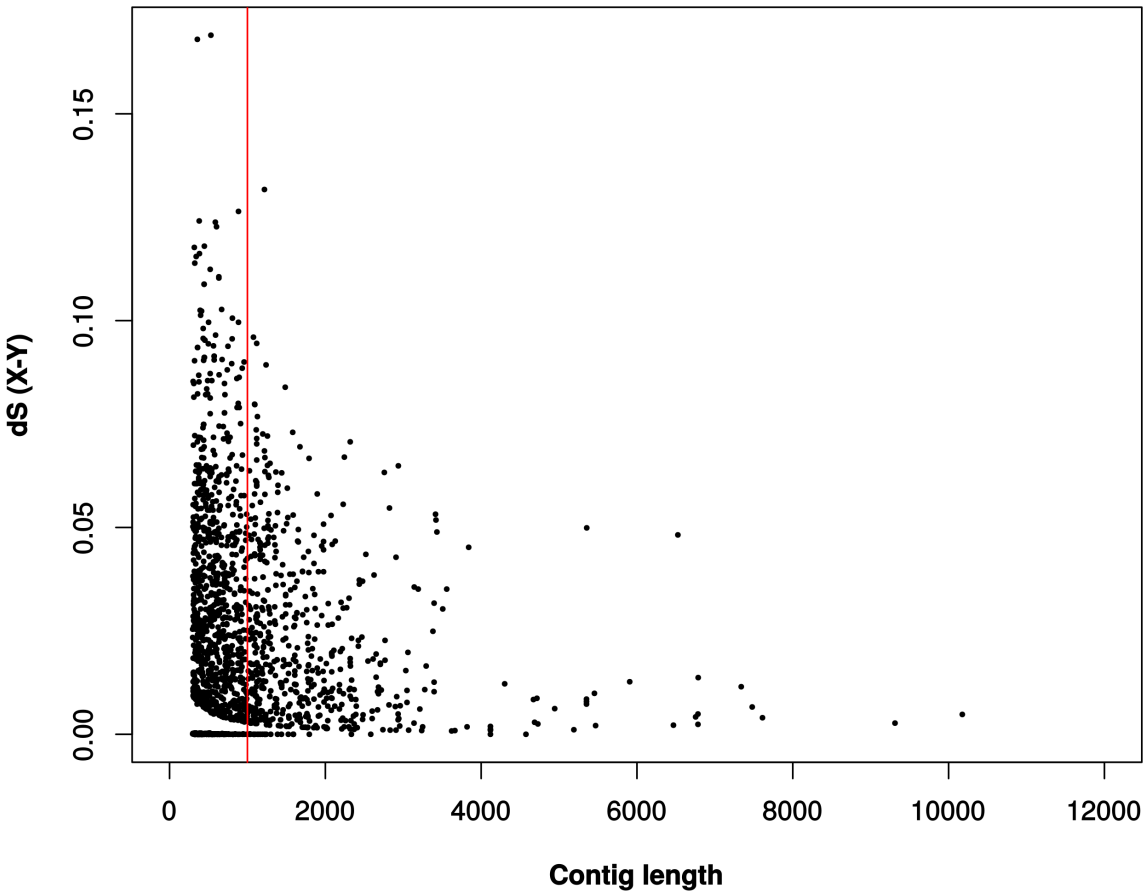
838

838

839 **Supplementary Figure 1: Distribution of pairwise synonymous divergence (dS) between**

840 **X and Y alleles in X/Y contigs.** The vertical line indicates contigs longer than 1kb.

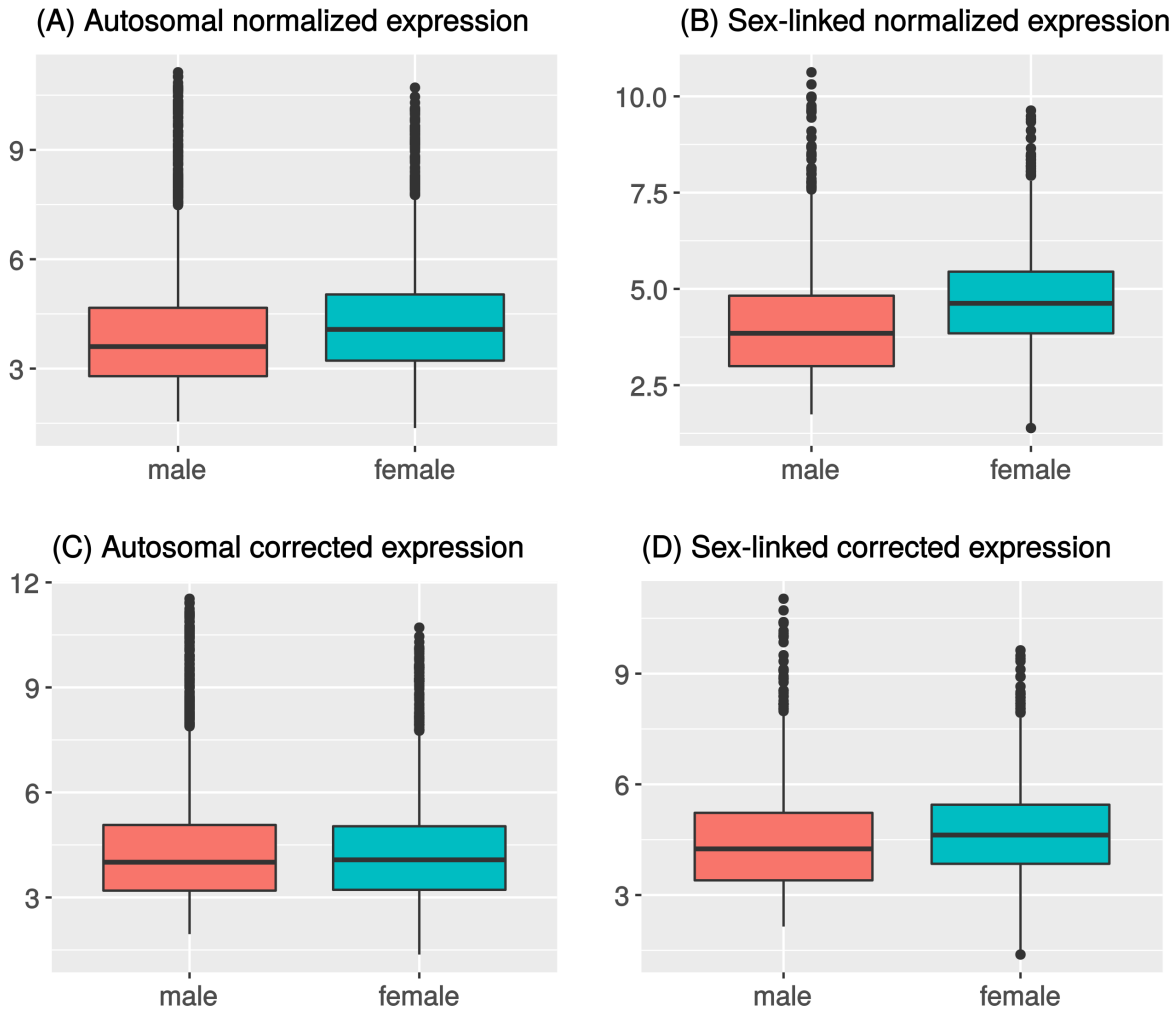
841



842

843

Supplementary Figure 2: Expression levels of autosomal and sex-linked contigs in both sexes before and after correction. Total read numbers were summed at SNP locations in each autosomal (A and C) and sex-linked (B and D) contig and normalized for each individual separately; medians for all individuals of the same sex were then obtained. Expression was corrected (C and D) using the autosomal median male/female ratio (see Methods).



852

853 **Supplementary Figure 3: Male vs. female autosomal gene expression ratio after**

854 **correction.** Distribution of normalized expression difference between males and females in

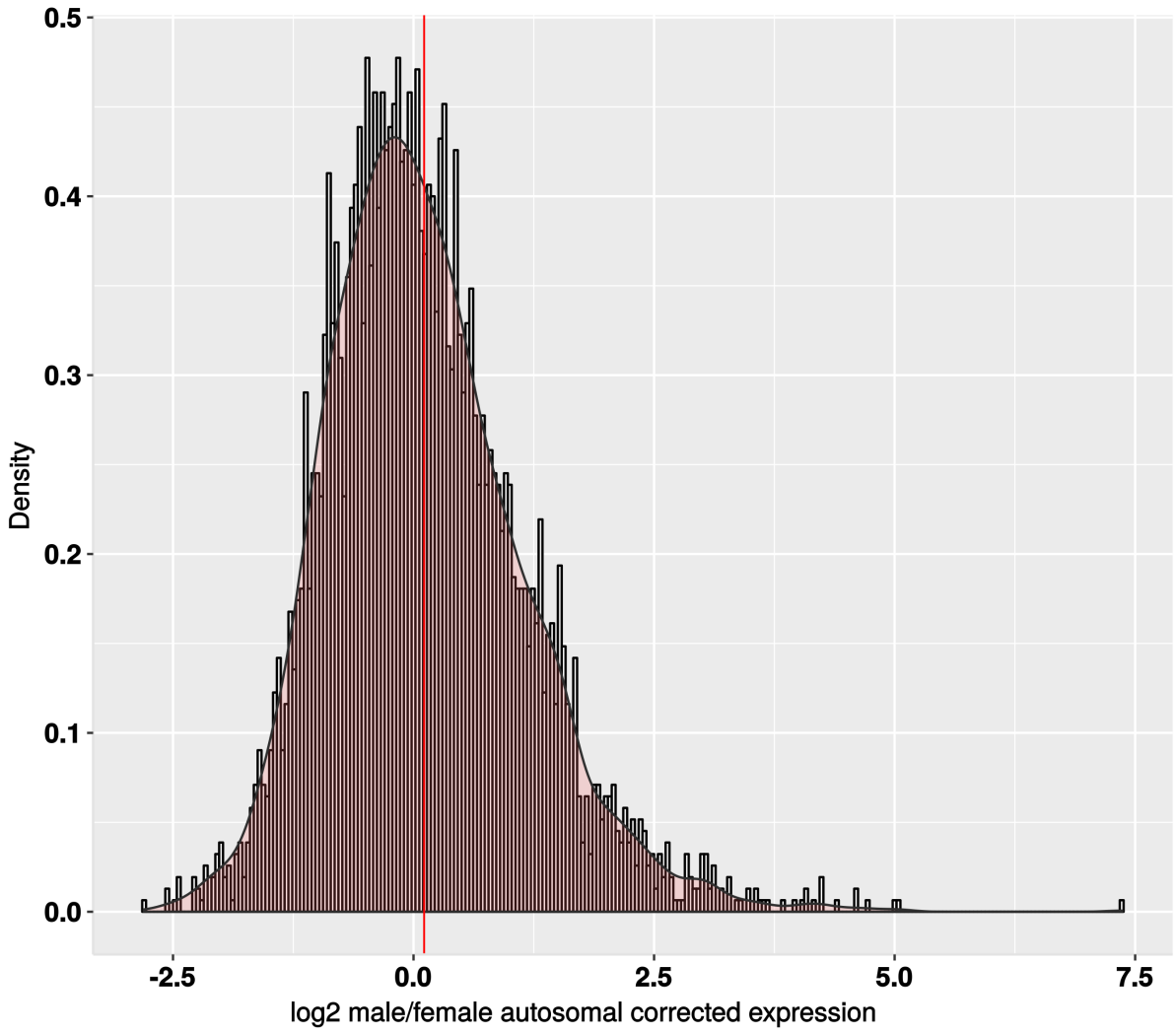
855 autosomal contigs. Total read numbers were summed at autosomal SNP locations for each

856 contig and normalized for each individual separately, then averaged across sexes to obtain the

857 male/female ratio. Expression was corrected using the autosomal median male/female ratio.

858 Distribution is shown in $\log_2$ scale with its density curve. The median is shown in red.

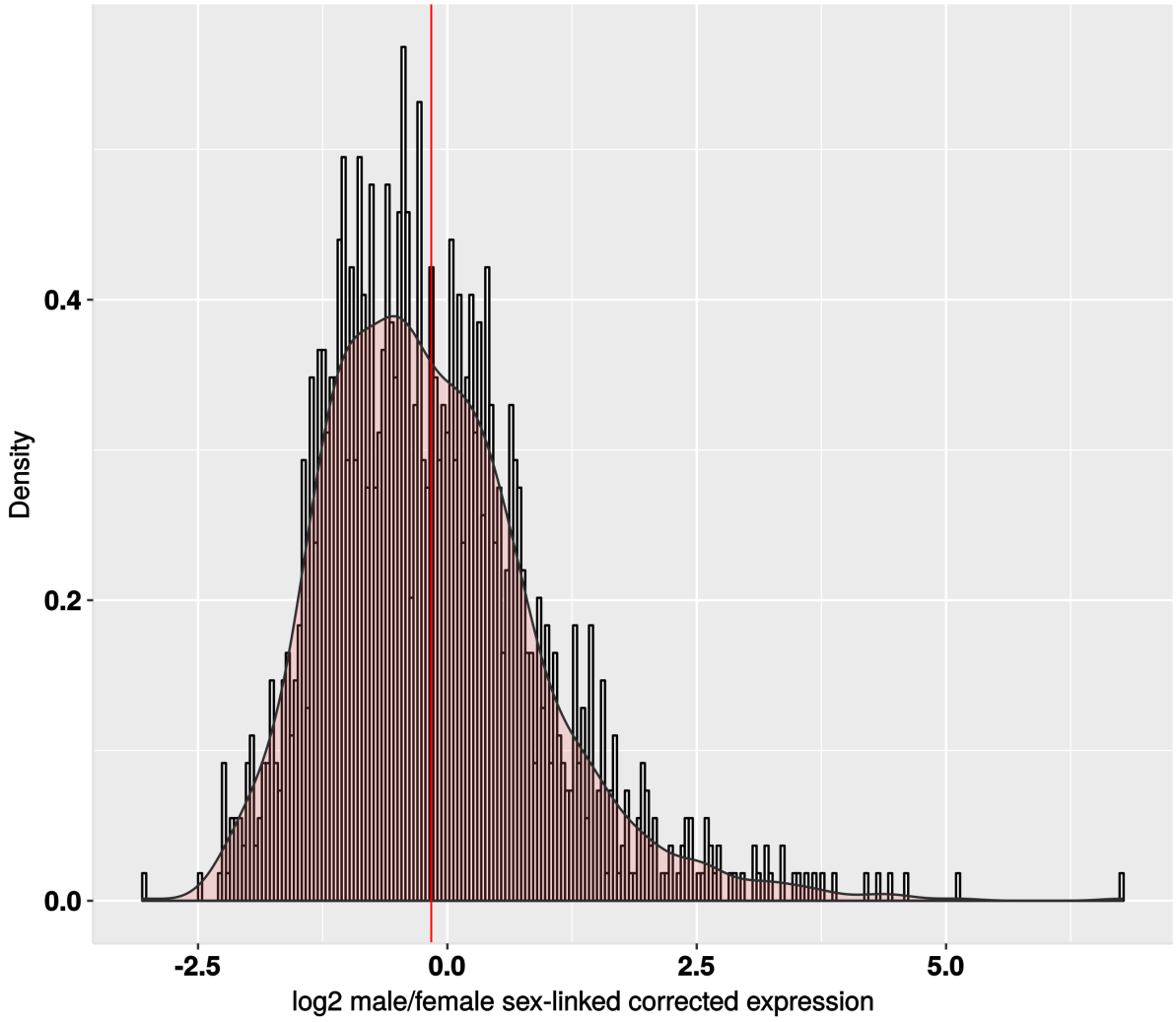
859



860

861

Supplementary Figure 4: Male vs. female sex-linked genes expression ratio after correction. Distribution of normalized expression difference between males and females in sex-linked contigs. Total read numbers were summed at sex-linked SNP locations for each contig and normalized for each individual separately, then averaged across sexes to obtain the male/female ratio. Expression was corrected using the autosomal median male/female ratio. Distribution is shown in $\log_2$ scale with its density curve. The median is shown in red.



Supplementary Figure 5: Venn diagram of Differentially Expressed Genes identified by DESeq2, edgeR, and LimmaVoom. Number of genes with FDR <0.0001 for each method. 3,273 differentially expressed genes were identified by at least 2 methods.