Gap-free X and Y chromosome assemblies of *Salix arbutifolia* reveal an evolutionary change from male to female heterogamety in willows, without a change in the position of the sex-determining locus

Yi Wang[1,2], Guang-Nan Gong[1], Yuan Wang[1], Ren-Gang Zhang[3], Elvira Hörandl[4], Zhi-Xiang Zhang[2], Deborah Charlesworth[5] and Li He[1]

[1]Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai, 201602, China; [2]Laboratory of Systematic Evolution and Biogeography of Woody Plants, School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, 100091, China; [3]Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650201, Yunnan, China; [4]Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), University of Goettingen, 37073, Göttingen, Germany; [5]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3FL, UK

Authors for correspondence:
Deborah Charlesworth
Email: deborah.charlesworth@ed.ac.uk

Li He
Email: lhe@cemps.ac.cn

Received: *16 November 2023*
Accepted: *21 March 2024*

New Phytologist (2024) **242:** 2872–2887
doi: 10.1111/nph.19744

Key words: evolutionary strata, heteromorphism, inversion, pericentromeric regions, *Salix arbutifolia*, sex chromosome turnovers.

Summary

- In the *Vetrix* clade of *Salix*, a genus of woody flowering plants, sex determination involves chromosome 15, but an XY system has changed to a ZW system. We studied the detailed genetic changes involved.
- We used genome sequencing, with chromosome conformation capture (Hi-C) and PacBio HiFi reads to assemble chromosome level gap-free X and Y of *Salix arbutifolia*, and distinguished the haplotypes in the 15X- and 15Y-linked regions, to study the evolutionary history of the sex-linked regions (SLRs).
- Our sequencing revealed heteromorphism of the X and Y haplotypes of the SLR, with the X-linked region being considerably larger than the corresponding Y region, mainly due to accumulated repetitive sequences and gene duplications.
- The phylogenies of single-copy orthogroups within the SLRs indicate that *S. arbutifolia* and *Salix purpurea* share an ancestral SLR within a repeat-rich region near the chromosome 15 centromere. During the change in heterogamety, the X-linked region changed to a W-linked one, while the Z was derived from the Y.

Introduction

Genetic sex determination systems have evolved independently in different angiosperm lineages, including male heterogametic (XX/XY) and female heterogametic (ZZ/ZW) systems (Ming *et al.*, 2011; Renner, 2014; Käfer *et al.*, 2017). In plants, in which separate sexes have evolved from cosexual ancestors (hermaphroditic or monoecious populations), this initial step may involve mutation in two genes, one with an essential male function, creating females, and one creating males by a femaleness-suppressing mutation, defining a Y-linked region (Charlesworth & Charlesworth, 1978). An incompletely Y-linked region carrying two such factors creates selection favoring closer linkage between them, potentially leading to complete recombination suppression between the Y- and X-linked regions, allowing the Y to become differentiated. The single gene systems that have recently been discovered (Akagi *et al.*, 2014; Muller *et al.*, 2020), including new ones generated by sex chromosome turnovers (Vicoso, 2019), will not be selected for changed recombination (reviewed in Charlesworth & Harkess, 2024). Moreover, Y- and X-linked regions can also arise if sex-determining mutations arise within a recombinationally inactive region (Charlesworth, 2019) and examples of sex-linked regions (SLRs), within pericentromeric regions are now known in several plants (Zhou *et al.*, 2018; Rifkin *et al.*, 2021; She *et al.*, 2023; Yue *et al.*, 2023).

Y-linked regions often accumulate repetitive elements (Charlesworth *et al.*, 1994), directly reducing gene density compared with homologous X-linked regions, and with autosomal or pseudo-autosomal regions (PARs), for example, in *Silene latifolia* (Bergero *et al.*, 2008). After enough evolutionary time, genetic degeneration may also occur, sometimes eventually leading to major gene losses, creating hemizygosity of genes and regions in

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

males. Low recombination rates also allow the accumulation of Y-specific gene duplications, chromosome rearrangements, and eventual changes in size and heterochromatinization of the sex chromosomes (Charlesworth *et al.*, 2005; Bergero & Charlesworth, 2009). W chromosomes in species with female heterogamety undergo similar changes (Fraisse *et al.*, 2017; Picard *et al.*, 2018; L. Xu *et al.*, 2019).

By contrast, the X and Z chromosomes, which recombine in one sex, are expected to largely retain the characteristics of their ancestors before sex linkage evolved, though, in XY species, the X recombines only in females, reducing its recombination rate compared with the autosomes, so it may also undergo some changes (e.g. accumulation of repeats; Mrnjavac *et al.*, 2023). Although empirical data are limited in plants, assemblies of the *Silene latifolia* X and *Salix dunnii* X found a lower gene density, compared with most autosomes (He *et al.*, 2021; Yue *et al.*, 2023).

If recombination suppression has occurred, study of the mechanism(s) is important. It has sometimes spread along the sex chromosome, forming 'evolutionary strata' with different levels of Y-X sequence divergence, reflecting different times when recombination stopped (Lahn & Page, 1999). Although data are very scarce, strata have been detected in a few flowering plants, namely *Silene latifolia*, *Cannabis sativa*, and *Humulus lupulus* (Lahn & Page, 1999; Papadopulos *et al.*, 2015; Prentout *et al.*, 2021) with heteromorphic, highly degenerated, sex chromosomes. Lahn & Page (1999) suggested that inversions initiate new sex chromosome strata. Flowering plants are well-suited to testing for inversions, because dioecy has often evolved recently (Charlesworth & Harkess, 2024) so that later rearrangements have not had time to obscure such events. Inversions were identified in the papaya SLR (Wang *et al.*, 2012).

In the family Salicaceae, turnovers have occurred, creating young sex-determining regions (Yang *et al.*, 2021; Wang *et al.*, 2023), and strata have been suggested, possibly involving an inversion, in the SLR of *Populus euphratica* (S. Zhang *et al.*, 2022). In *Salix viminalis*, two strata were suggested, but the assemblies were low quality, especially in the SLR (Almeida *et al.*, 2020). Further studies in this family are therefore needed. The Salicaceae also display sex chromosome turnovers, including 'heterogametic' transitions between XY and ZW systems transitions, as well as turnovers retaining the same heterogametic sex. Like homogametic transitions, in which the heterogametic sex is unchanged, heterogametic ones can involve the same or different chromosomes (Bull, 1983; Saunders *et al.*, 2018). However, it has been suggested that W-linked regions are likely to evolve from ancestral X-linked regions (or vice versa), because these carry feminizing genes, while the Z and Y may often carry masculinizing genes (Miura, 2007; Ma *et al.*, 2018; Ogata *et al.*, 2018; Furman *et al.*, 2020).

In *Salix*, recent analyses of genome-wide sequences from many species show that the traditional subgenera *Chamaetia*, *Vetrix*, *Longifoliae*, *Protitea*, and *Salix*, for example in Argus *et al.* (2010), are inconsistent with the phylogenetic trees, which indicate two main clades, *Salix* and *Vetrix* (Gulyaev *et al.*, 2022; Sanderson *et al.*, 2023). Under this phylogenetic framework, an XY to ZW transition has been found in *Vetrix* species, whose SLRs are on chromosome 15. Basal *Vetrix* species (Fig. 1a) have male heterogamety (abbreviated to 15XY), but other species, including *Salix purpurea*, have female heterogamety (15ZW; Wang *et al.*, 2023; Xue *et al.*, 2024). Reference genome sequences of several *Salix* species, including *S. purpurea*, and the *Salix* clade 7XY outgroup species, *S. dunnii*, include SLRs occupying between *c.* 18% and 44% of their sex chromosomes, with larger X- or W-linked regions than Y or Z ones (Zhou *et al.*, 2020; He *et al.*, 2023).

The mechanism of the XY to ZW change remains to be understood. In the Salicaceae, an *ARR17*-like gene (in a cytokinin response regulator family) has been shown to be essential for female development. CRISPR-Cas9-induced mutation in a *Populus* species with male heterogamety (Muller *et al.*, 2020) showed that when *ARR17*-like gene expression is silenced or absent, individuals develop into males, and partial duplicates in the SLRs can act as male-determining factors, by producing small RNAs that silence intact *ARR17*-like copies. The SLRs of both *Salix arbutifolia*

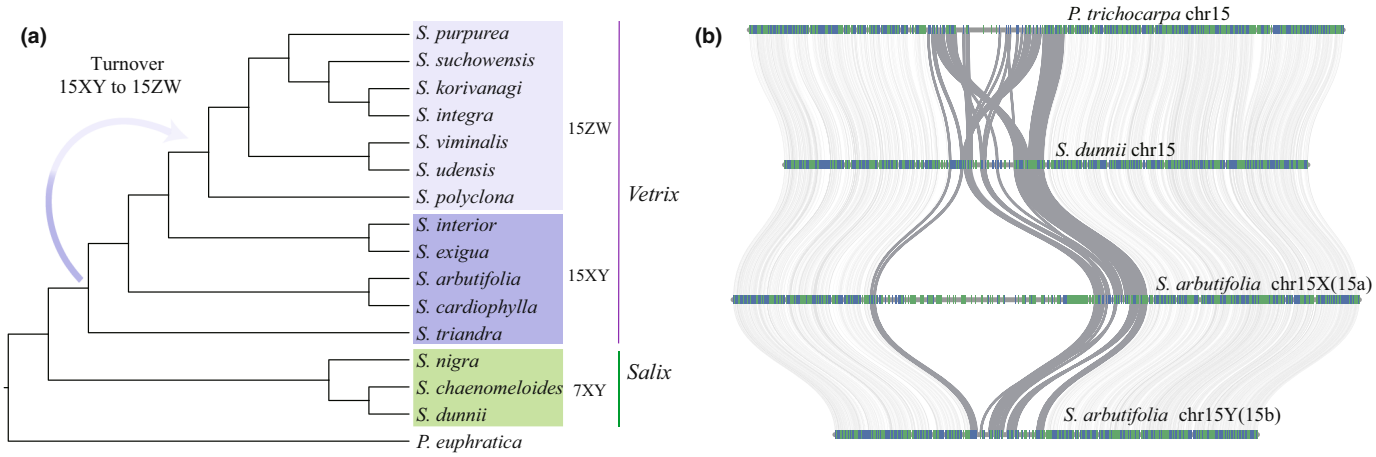

Fig. 1 (a) Phylogenetic relationships of species with known sex determination systems in *Salix* and turnover events in the *Vetrix* clade, adapted from Xue *et al*. (2024). (b) Microsynteny between the *Salix dunnii*, *Populus trichocarpa* autosome 15, and the *Salix arbutifolia* sex chromosomes 15X and 15Y, showing the heteromorphism due to a large region present only in the 15X sex-linked region. The dark gray lines indicate genes inferred to be within the *S. arbutifolia* sex-linked regions and the corresponding regions of *S. dunnii* and *Populus trichocarpa*.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

(15XY) and *Salix chaenomeloides* (7XY) also show these properties and produce small RNAs that can silence intact *ARR17*-like copies on chromosome 19 (Wang *et al.*, 2022). *ARR17*-like genes might therefore be involved in sex determination across the Salicaceae family (Yang *et al.*, 2021; Wang *et al.*, 2022). If the partial duplicates are the male-determining genes in these species, they should be located in the oldest-established SLR(s), which can potentially be identified by comparing closely related species whose chromosomes have not undergone too many rearrangements, though rearrangements are documented in the family Salicaceae (He *et al.*, 2021). A recent paper used the 15XY species, *Salix exigua*, to study the XY to ZW transition (Hu *et al.*, 2023), and proposed that both the W and the Z arose from the Y chromosome. However, their data were based on partially unphased (hybrid) XY sequences, which cannot provide definitive conclusions about the origins of the different sex chromosomes. To better understand the change in heterogamety, we developed new gap-free assemblies of *S. arbutifolia* X and Y chromosomes, and compared their organization with previous genome assemblies of *S. purpurea*, (Zhou *et al.*, 2020). We combined the results with new *S. arbutifolia* RNA-Seq reads from female and male catkins, and population resequencing data (Wang *et al.*, 2023), and with the partially unphased sequences of the 15XY regions of *S. exigua* (Hu *et al.*, 2023), genome of *S. dunnii* (He *et al.*, 2021), genome (with phased 15Z- and 15W-linked regions), and expression data from *S. purpurea* (Zhou *et al.*, 2020; Hyden *et al.*, 2021), as well as genome of *Populus trichocarpa* v.4.1 (https://genome.jgi.doe.gov).

Materials and Methods

Plant material

We collected young leaves from a male *S. arbutifolia* Pall., B649AF, for genome sequencing, plus young leaf, catkin and stem samples, and catkins from a female plant B649V, for transcriptome sequencing and gene annotation. Supporting Information Table S1 lists these materials. To study sex linkage, we also analyzed the resequencing data from 41 wild individuals of Wang *et al.* (2023). Three other species with published genome assemblies, *S. dunnii* C.K. Schneid. (He *et al.*, 2021), *S. purpurea* L. (Zhou *et al.*, 2020), and *P. trichocarpa* Torr. & A. Gray were included in relevant analyses.

Genome sequencing

For Illumina PCR-free sequencing, total genomic DNA was extracted from B649AF using a Qiagen DNeasy Plant Mini Kit. Sequencing libraries were generated using the Illumina TruSeq DNA PCR-Free Library Preparation Kit. After quality assessment on an Agilent Bioanalyzer 2100 system, the libraries were sequenced on an Illumina NovaSeq 6000 platform by Beijing Novogene Bioinformatics Technology (hereafter Novogene). For Hi-C and HiFi sequencing, total genomic DNA was extracted by the CTAB method (Chang *et al.*, 1993). For Hi-C sequencing, fresh leaves from B649AF were fixed in 4% formaldehyde solution in MS buffer. Subsequently, crosslinked DNA was isolated from nuclei. The DpnII restriction enzyme was used to digest the DNA, and the digested fragments were labelled with biotin, purified, and ligated before sequencing. Hi-C libraries were sequenced on an Illumina Hiseq X Ten platform by Novogene. The HiFi library for single-molecule real-time (SMRT) sequencing was constructed with an insert size of 15-kb using the SMRTbell Express Template Prep Kit 2.0. The DNA was sheared using the Diagenode Megaruptor system and concentrated with AMPure PB Beads (Pacific Biosciences, Menlo Park, CA, USA). The libraries were size selected on a BluePippin System and sequenced by Novogene using the PacBio Sequel II platform.

RNA extraction and library preparation

Total RNA was extracted from young leaves, catkins, and stems using the Plant RNA Purification Reagent (Invitrogen). RNA-Seq transcriptome libraries were prepared using the TruSeq RNA sample preparation kit, and sequencing was performed on an Illumina Novaseq 6000 by Novogene.

Genome assembly

The CCS software (https://github.com/PacificBiosciences/ccs) was used to produce high-precision PacBio HiFi reads. The reads were used to assemble initial contigs using the Hifiasm pipeline v.0.16-r375 (Cheng *et al.*, 2021) with default parameters and assemble haplotypes of each chromosome for subsequent analyses. We filtered the Hi-C reads using Fastp v.0.23.2 (Chen *et al.*, 2018). Juicer v.1.5.6 (Durand *et al.*, 2016) and 3d-DNA pipeline v180922 (Dudchenko *et al.*, 2017) were then used to achieve complete chromosome assemblies, based on the clean Hi-C reads and the assembled genome sequences. We numbered the chromosome according to the 19 chromosomes of *Salix brachista* (Chen *et al.*, 2019). LR_Gapcloser v.1.1 (G. C. Xu *et al.*, 2019) and Minimap2 (Li, 2018) were used to improve the contiguity of the genome assemblies based on HiFi reads. We also employed NextPolish v.1.41 (Hu *et al.*, 2020) to improve the assembly's base accuracy using Illumina short reads. Redundans v.0.13c (Pryszcz & Gabaldon, 2016) was used to remove redundancy and contamination from external sources. GetOrganelle v.1.7.5 (Jin *et al.*, 2020) was used to assemble mitochondrial and chloroplast genomes based on clean Illumina short reads. The assembly was evaluated by mapping the Illumina paired-end reads, HiFi reads, and RNA-Seq reads to the genome assembly using the Bwa-Mem v.0.7.8 algorithm (Li & Durbin, 2009), Minimap2, and Hisat2 v.2.1.0 (Kim *et al.*, 2015). We also used Busco v.4.1.2 (http://busco.ezlab.org/) to check the completeness of the genome assembly.

Annotation of repetitive sequences

Repeat elements were identified and classified using RepeatModeler v.2.0.5 (Flynn *et al.*, 2020) to produce a repeat library for *S. arbutifolia*. Then, RepeatMasker (http://www.repeatmasker.org/RepeatMasker) was used to identify repetitive regions in the genome.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Transcriptome assembly and gene annotation

The genome was annotated by combining evidence from transcriptome data, protein homology-based, and *ab initio* prediction. We used three different methods: (1) RNA-Seq reads *de novo* assembled with Trinity v.2.0.6 (Grabherr *et al.*, 2011); (2, 3) RNA-Seq reads mapped to the reference genome using Hisat2 v.2.1.0 (Kim *et al.*, 2015), and then assembled using either Trinity v.2.0.6 with genome-guided mode, or using StringTie v.1.3.5 (Pertea *et al.*, 2015). CD-Hit v.4.6 (Fu *et al.*, 2012) was used to control the transcript quality, and Pasa v.2.4.1 (Haas *et al.*, 2003) was used to obtain high-quality loci from the transcriptome data, to train the Augustus v.3.4.0 (Stanke *et al.*, 2008) gene modeler. A total of 278 011 protein sequences were obtained from published data from 17 species of Salicaceae and *Arabidopsis thaliana* (Table S2) and used as reference proteins for homology-based gene annotation. The Maker2 pipeline v.2.31.9 (Cantarel *et al.*, 2008) was used to annotate genes.

We annotated tRNA, rRNA sequences, and other noncoding RNAs (ncRNAs)as described in He *et al.* (2021). The functions of protein-coding genes were annotated based on three strategies: (1) the annotated genes were aligned to proteins in the eggNOG databases v.4.0 (Powell *et al.*, 2014) using eggNOG-Mapper v.2.0.0 (Huerta-Cepas *et al.*, 2017), and assigned to GO (http://geneontology.org/) and KEGG (https://www.genome.jp/kegg/pathway.html) metabolic pathways; (2) the protein sequences were aligned with the Uniprot database (including Swiss-Prot and TrEMBL, https://www.uniprot.org/), NR (https://www.ncbi.nlm.nih.gov/), and *A. thaliana* reference genome sequence using Diamond v.0.9.24 (Buchfink *et al.*, 2015; with E value $< 10^{-5}$ and identity >30%); (3) motifs and functional domains were identified by searching various domain libraries (ProDom, PRINTS, Pfam, SMART, PANTHER and PROSITE) using Interproscan v.5.27-66.0 (Jones *et al.*, 2014). Pseudopipe (Zhang & Yu, 2006) was used with default parameter settings to detect pseudogenes in the whole genome. We also used Blastn v.2.5.0 (Altschul *et al.*, 1990) to identify pseudogenization within the 15X- and 15Y-SLRs of *S. arbutifolia*, and to compare the exons of the male B649AF's 15X-SLR specific alleles with the 15Y-SLR sequences. Sequences in the 15Y-SLR with hits to only parts of the corresponding 15X-SLR exons were classified as pseudogenes. In a second approach, we extracted possible pseudogenes identified by PseudoPipe in the 15Y-SLR with the 15X gene IDs. We used Blastp against all *S. arbutifolia* peptide sequences excluding 15X, to exclude any of these candidate pseudogenes that proved to have complete 15Y-SLR exons. Finally, when both approaches identified the same pseudogene, we kept only one. We identified pseudogenes in the 15X-SLR similarly.

Identifying SLRs of *S. arbutifolia*

Short resequencing reads from 21 individual *S. arbutifolia* females and 20 males were used to distinguish the 15X and 15Y haplotypes. We first applied the chromosome quotient (CQ) method (Hall *et al.*, 2013) to our haplotype sequences. The CQ is the normalized ratio of female to male alignments to a given reference sequence, using the stringent criterion that the entire read must align with zero mismatches. CQ-calculate.pl (Hall *et al.*, 2013) was used to map all reads from the samples to sliding 50-kb nonoverlapping windows of haplotype *a* or *b*. In an XY system, provided that X- and Y-linked sequences differ sufficiently, or many sequences are absent from the Y due to genetic degeneration (or X insertion), ratios of female to male alignments to a given reference sequence (CQ values) may be around two using the X reference haplotype, vs close to 1 and zero, respectively, for autosomal and Y-linked sequences.

Bwa v.0.7.12 (Li & Durbin, 2009) was used to align the clean reads to two haplotype sequences. Samtools v.0.1.19 (Li *et al.*, 2009) was used to process the mapped sequence data. Variants were called for all individuals using Gatk v.4.1.8.1 (McKenna *et al.*, 2010). Gatk and VCFtools v.0.1.16 (Danecek *et al.*, 2011) were used to select high-quality SNPs after hard filtering (with the parameter values QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum <−12.5, ReadPosRankSum <−8.0, SOR > 3.0). We: excluded sites with coverage above twice the mean depth at variant sites across all samples; retained only bi-allelic SNPs; and removed SNPs with > 10% missing information and/or minor allele frequency < 5%. Weighted F_{ST} values between the sexes were then calculated in overlapping 100-kb windows with 10-kb steps, using Weir & Cockerham's (1984) estimator. A changepoint package (Killick & Eckley, 2014) was used to detect regions with significantly different F_{ST} values, in order to detect candidate SLRs (He *et al.*, 2021). We used LDBlockShow v.1.36 (Dong *et al.*, 2021) to calculate and visualize the LD (linkage disequilibrium) pattern of chromosome 15a (15X)- and 15b (15Y)-based SNP datasets that were separated by at least > 5 kb. High densities of TEs, and correspondingly low gene contents, are expected in the centromeric regions of chromosomes, and regions with such prominent patterns were therefore identified as putative centromeric regions.

X- and Y- SLRs genes in *S. arbutifolia*

We compared the SLRs of the two sex chromosomes and their presumed ancestral autosome, chromosome 15 of *S. dunnii* and *P. trichocarpa*, by synteny analysis using MCScan v.1.2.7 (Tang *et al.*, 2008) with '--cscore = 0.99'. To identify SLR genes present in both the B649AF male's 15X- and 15Y-sequences, we performed a reciprocal Blast (Götz *et al.*, 2008) of all potential peptide sequences in the 15X and 15Y regions. 15X-SLR genes with no Blast hits in the 15Y-SLR were classified as possibly 15X-SLR specific, and vice versa for the 15Y-SLR. Autosome 15 of *S. dunnii* (He *et al.*, 2021) was used as an outgroup representing the likely state of the ancestral chromosome. Gene gains and losses affecting the 15X- and 15Y-SLR were classified as follows: (1) genes found only in the 15X (or 15Y) sequence with hits to outgroup autosome 15 were respectively classified as 'Y (or X) deletion'; and (2) genes without hits to autosome 15 were categorized as 'X (or Y) insertion'. We also employed Blastp to identify duplicated genes within the X- and Y-SLRs.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Relationships of the SLRs during the change in the *Vetrix* clade

To trace the evolutionary history of 15XY and 15ZW in the *Vetrix* clade, we used the phased 15-SLRs of *S. arbutifolia* and *S. purpurea*. First, to test whether the SLRs inherited similar sequences from their ancestors, we aligned all *S. arbutifolia* 15X- and 15Y-SLR specific genes with the 15Z- and 15W-SLRs of *S. purpurea*. Second, we used OrthoFinder v.2.5.5 (Emms & Kelly, 2019) to identify single-copy orthologs (SCOs) in the SLRs and in chromosome 15 of *S. dunnii*. MAGUS software (Smirnov & Warnow, 2021) was used to align the sequences of each SCO set. We also used Blastn to search chromosome 15 of *S. exigua* to find likely orthologs of *S. arbutifolia* single-copy sex-linked genes. We used ModelFinder (Kalyaanamoorthy *et al.*, 2017) to select the best model, and IQ-Tree v.2.2.0.3 (Nguyen *et al.*, 2015) to estimate gene trees, and perform 1000 bootstrap replicates. Astral v.5.7.8 (Zhang *et al.*, 2018) was used to obtain species trees based on the gene trees. The local posterior probabilities were used to calculate clade support (Sayyari & Mirarab, 2016).

To identify any *ARR17*-like gene duplicates in the SLRs, we used Blastn to search the whole *S. arbutifolia* and *S. purpurea* genomes for sequences similar to the exons of Potri.019G 133600, the likely female-determining factor (Muller *et al.*, 2020). This *ARR17*-like gene includes five protein-coding exons. Blastn results that lacked any of the five exons were treated as partial duplicates, and the others as intact, following (Wang *et al.*, 2022). *ARR17*-like gene phylogenies were estimated with IQ-Tree, using exon sequences.

Tests for evolutionary strata in the X and Z chromosomes

We examined the topologies of SCO sets of 15X-, 15Y-, 15Z-, and 15W-SLR genes from *S. arbutifolia* and *S. purpurea*, since genes in a sex chromosome region in which recombination was suppressed at different times, relative to the split between the two species, will exhibit different topologies and thus can be used to infer evolutionary strata (Handley *et al.*, 2004; H. Zhang *et al.*, 2022). For species with a shared SLR, their X and Y or/and W and Z alleles are expected to cluster by gametologs rather than species (Cortez *et al.*, 2014). We also calculated synonymous site divergence (Ks) values between XY or ZW gametolog pairs identified by reciprocal Blast searches, excluding sequences with amino acid identity < 70% (Sievers *et al.*, 2011); divergence was estimated using PAML's yn00 function (Yang, 2007).

Gene expression

We used Fastp v.0.23.2 to filter the raw reads from male and female catkins (Table S1), mapped the clean RNA-Seq reads to 15X and 15Y in the assembled genome using Hisat2 v.2.1.0 (Kim *et al.*, 2015) and assigned counts using featureCounts v.2.0.1 (Liao *et al.*, 2014). After filtering out unexpressed genes (counts = 0 in all samples, excluding non-mRNA), we converted the read counts to TPM (transcripts per million reads; Vera Alvarez *et al.*, 2019). As shown in Table S1, we used three biological replicates of catkins of each sex. Differential expression analyses between male and female groups were performed using the DESeq2 R package v.1.16.1 (Love *et al.*, 2014). We classified genes as differentially expressed genes (DEGs) using two criteria (1) a corrected *P*-value <0.05 (Padj) and (2) absolute $\log_2$FoldChange > 1.

Table 1 Statistics of the *Salix arbutifolia* genome assembly.

Category	Summary statistics
Total assembly size (Mb)	612
Total number of contigs	44
Maximum contig length (Mb)	28.382
Minimum contig length (kb)	15.522
Contig N50 length (Mb)	16.045
Contig L50 count	17
Contig N90 length (Mb)	10.093
Contig L90 count	35
Total number of scaffolds	40
Maximum scaffold length (Mb)	28.382
Minimum scaffold length (kb)	15.522
Scaffold N50 length (Mb)	16.446
Scaffold L50 count	17
Scaffold N90 length (Mb)	13.288
Scaffold L90 count	33
Gap number	4
BUSCO (%)	98.1
GC content (%)	34.42
Gene number	62 709
Repeat content (%)	40.62

Results

Genome assembly

Our haplotype-resolved assembly of the *S. arbutifolia* genome was derived by integrating a total of 45.2 Gb Illumina reads, 44.9 Gb of HiFi reads with an average length of 18 kb, and 44.7 Gb of Hi-C reads (Table S3). The HiFi and Hi-C reads yielded a chromosome-scale assembly of 40 scaffolds, N50 = 16.05 Mb (Fig. S1; Table 1). Thirty-eight contigs were assigned to the 19 *S. arbutifolia* chromosomes. Most assembled chromosomes were gap free (except for chr04a, chr07a, chr07b, and chr19a). The mitochondrial (Mt) and chloroplast (Cp) genomes formed two scaffolds, of 659 371 and 155 219 bp, respectively. Around 99.6% of the Illumina short reads mapped to the assembly, of which *c.* 99.4% was covered by at least 20× reads. Around 99.6% of the HiFi reads were mapped to the assembly and 99.4% had coverage ≥ 20 (Table S4). The primary assembly covered 95.9% of complete BUSCO genes. Thus, our genome assembly has high continuity, coverage, and accuracy.

Genome annotation

We identified a total of 31 263 gene models in the *S. arbutifolia* haplotype *a* set of sequences, and 31 156 in *b*, plus 91 genes in the plastid genome and 199 in the mitochondrial genome

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

(Table S5). The average length of a predicted protein-coding gene is 3457.5 bp, with 6.2 exons (Table S6). 98.32% matched a predicted protein in the public database(s) (Table S7). We also detected 248.46 Mb of repetitive sequences (40.62% of the assembled genome, Table S8).

Identifying SLRs in *S. arbutifolia*

The two haplotypes of most chromosomes have similar sizes (Table S5). However, although no sex chromosome heteromorphism has previously been detected, the *S. arbutifolia* chromosome 15 haplotypes differ; *a* is 16.17 Mb, while *b* is only *c.* 10.98 Mb (Figs 1b, S2, S3). Part of the 15*a* haplotype yielded a CQ close to 2, identifying haplotype 15*a* as the X. Supporting this, the corresponding region has a CQ close to 0 using haplotype 15*b*, indicating that females consistently lack these sequences, and identifying it as the 15Y (Figs S2, S3).

The mapping ratios ranged from 86.00% to 93.80% using haplotype *a* as the reference (Table S9), and 87.40% to 94.00% with the smaller haplotype *b* (Table S10). We obtained 3308 616 and 3324 319 high-quality SNPs, respectively, in the same short-read sequences, and used these to estimate F_{ST} values between the sexes. Changepoint analysis detected higher F_{ST} values in a region corresponding to 7.06 Mb of the 15X (*a*) haplotype, compared with the flanking regions, locating an 15X-SLR occupying megabases 3.64–10.7, nearly half of the chromosome. F_{ST} analysis using the 15Y haplotype *b* to map variants identified a much smaller 15Y-SLR (1.81 Mb, between 3.62 and 5.43 Mb, only 16% of the chromosome). The two ends of the chromosome assembly are undifferentiated and are PARs. X-PAR1 and X-PAR2, total 9.11 Mb (Figs 2a, S4), and the Y-PARs are similar (9.17 Mb), as expected (Figs 2, S5).

Details of heteromorphism of the X- and Y- SLRs of *S. arbutifolia*

The PARs have higher gene density and lower repeat density than the SLRs (Table 2), consistent with the SLRs being nonrecombining. 238 and 85 protein-coding genes were annotated in 15X- and 15Y-SLR, respectively (Tables S5, S11). The heteromorphism revealed by the sequencing could be caused by deletions from the Y. However, degeneration is probably not the reason for the larger 15X-SLR size. First, the 15 and 15Y-SLRs include only one and three pseudogenes, respectively (Table S12). Further evidence for this conclusion is that synteny analysis detected a single extra *c.* 4.36 Mb 15X-specific region occupying 4.98–9.34 Mb (Fig. 1b, 'extended region' in Fig. 2). The extended region is highly repetitive (86.70%; Table S13). This creates an overall extremely high repeat sequence density in the 15X-SLR (80.73%), vs 64.09% in the 15Y-SLR and 71.11% in the rest of the 15X-SLR (which is syntenic with the 15Y-SLR, both of which have more LTR-Gypsy and LTR-Copia retrotransposons than the PARs; Table 2; Fig. 2). Repeat-rich regions, sometimes extensive, are detected in the middle of many *S. arbutifolia* chromosomes. These probably indicate centromeric and pericentromeric rarely recombining regions. The chromosome 15 results suggest that the SLRs are within such regions, accounting for their low gene densities and LD patterns (Fig. S6).

Consistent with such a location, we detected many duplicated genes in the 15Y- and 15X-SLRs, especially the latter (174% or 60.21% of the 15X-SLR genes, vs only 19.2% of those in the 15Y-SLR, a total of only 19 genes). Of the duplicated 15X-SLR genes, 134 (including a tandem array of 34 duplicates of the homologous gene Saarb15bG0044200 in the 15Y-SLR, with average length 8585 bp, and a mean of 11.6 exons) are in the 'extended region' (totaling 0.51 Mb, or 11.67% of the region; Figs 2, S7; Table S14). This array is discussed later. Together, repeats and duplicated genes constitute 98.17% of this X region, which completely lacks SCOs.

While 81% of the 15Y-SLR genes have 15X-SLR homologs, 120 complete genes were detected only in the 15X-SLR (we therefore term them 'X-specific'), of which 78 are 'X insertion' genes defined using homologous genome regions in outgroup species, as described in the Methods section (Fig. 2; Table S15). These genes include 64 duplicates in the extended region, vs only 42 in the 'Y deletion' category (including duplicates, 27 of which are in the extended region). Sixteen SLR genes were detected only in the 15Y-SLR, including seven 'X deletion' and nine 'Y insertion' genes (Fig. 2; Table S16). We calculated the proportions of genes lost from the X or Y, respectively, excluding duplicated genes, as the numbers in the 'X deletion' or 'Y deletion' categories, divided by the total number of genes ancestrally present (either shared by the 15X- and 15Y-SLRs, or classified as 'X deletion' or 'Y deletion'; as expected, most of these genes are also on chromosome 15 of the outgroup, *S. dunnii*). Ten genes were estimated to have been lost from the 15Y-SLR (11.49%), vs 6.90% from the 15X-SLR (Table S17). Overall, the results support the view that the extended 15X-SLR does not reflect Y degeneration, but an extra 15X-specific region (Fig. 1b).

A similar calculation for the *S. purpurea* ZW system estimated 8.44% loss for the 15Z-SLR, vs only 2.50% for the 15W-SLR (Table S17). Interestingly, the 15Y- and 15Z-SLRs share three deleted genes, and the 15X- and 15W-SLRs share one deleted gene, supporting the idea that the Z chromosome evolved from an ancestral Y chromosome with these deletions, and the W from the X.

Details of the sex chromosomes heterogamety change

We compared the 120 *S. arbutifolia* X-SLR and 16 Y-SLR specific genes with genes in the *S. purpurea* 15Z- and 15W-SLRs. Including duplicates, we found homologs of 25 15X-SLR specific genes (Tables 3, S15) in the *S. purpurea* 15W-SLR, but not its 15Z-SLR, suggesting that these genes were present on a chromosome ancestral to both the X and W, but not the Y or Z (Table 3). We also detected homologs of two 15Y-SLR specific genes in the 15Z- but not the 15W-SLR (Tables 3, S16; Fig. S8). These results support the prediction above that the W originated from an ancestral X chromosome, and the Z from a Y. However, 44 15X-SLR specific duplicated genes have no homologs in either the 15Z- or 15W-SLR, suggesting that they duplicated within the *S. arbutifolia* lineage (Tables 3, S15).

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

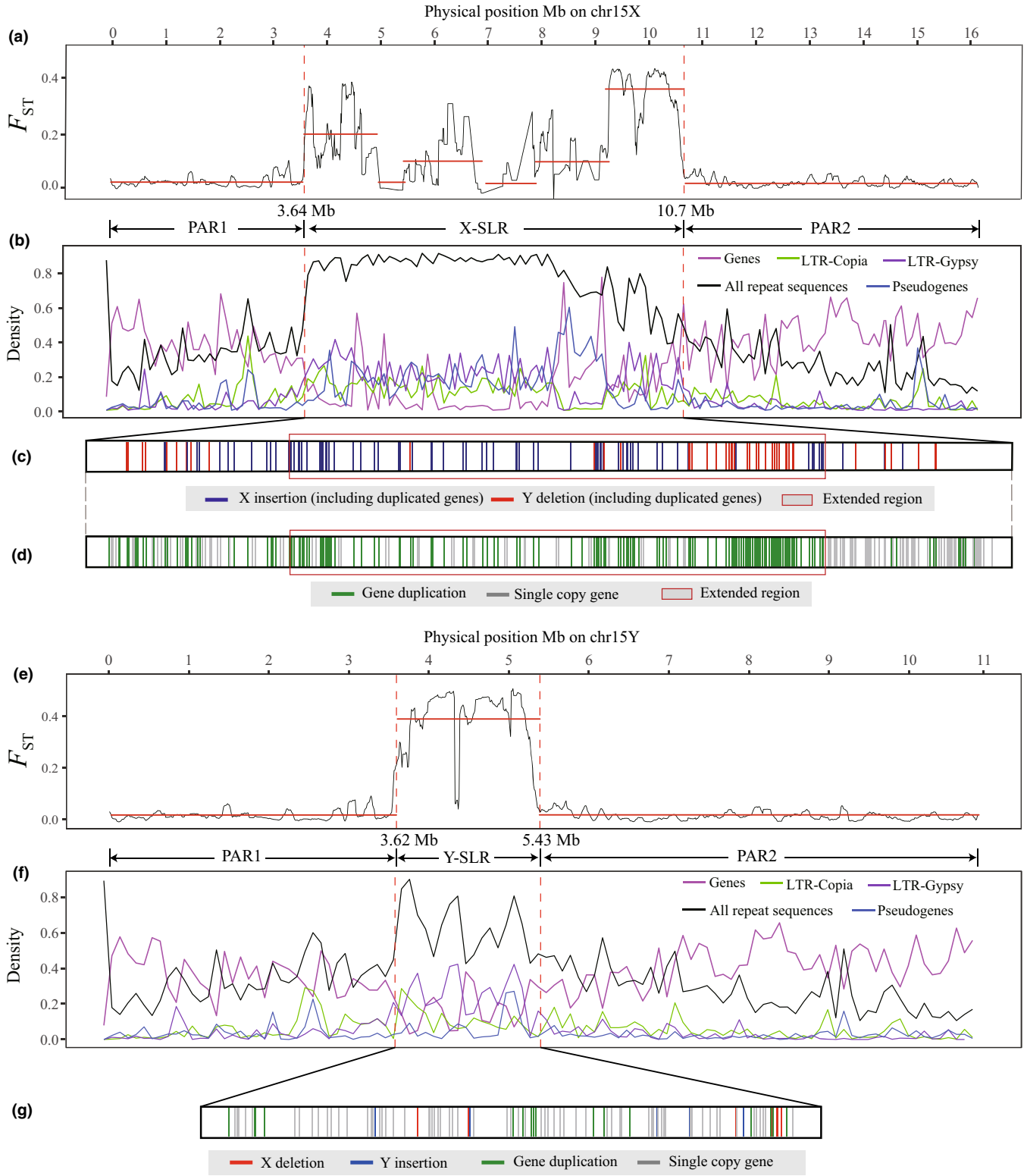

Fig. 2 Identification of the *Salix arbutifolia* 15X- and 15Y-SLRs. (a–d) Analysis of the 15X region. (a) F_{ST} values between the sexes in 100-kb overlapping windows, with 10-kb steps. Red horizontal lines indicate three regions that changepoint analysis suggests have different F_{ST} values. (b) Densities of LTR-Gypsy (purple line), LTR-Copia (green line), all repeat sequences (black line), pseudogenes (blue line), and genes (magenta line) in the entire 15X. (c) 15X-SLR specific genes (including duplicates). The blue vertical lines represent 'X insertion' genes, the red ones are 'Y deletion' genes, and the thin red box (also in d) indicates the X 'extended region'. (d) Gene duplication events in 15X-SLR. The green vertical lines indicate gene duplicates, and gray ones indicate single-copy genes, for example analysis of the 15Y region. The explanations above apply to these plots in (e) and (f) (for the other haplotype), except that there is no red boxed region (no part corresponding to (c) above), and, in (g) we show 15Y-SLR specific genes, with red vertical lines indicating 'X deletion', and blue ones 'Y insertion' genes. SLR, sex-linked region.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Table 2 Genes and repeats in different regions of the *Salix arbutifolia*, *Salix purpurea*, and *Salix dunnii* genomes.

Species and regions analyzed	Length of sequence occupied by genes, in Mb (percentage of the length[1])	Estimated amounts of repetitive sequences in Mb (percentage of the length[1])		
		Total repeats	LTR-Copia	LTR-Gypsy
S. arbutifolia				
15X-SLR	1.01 (14.37)	5.70 (80.73)	0.84 (11.91)	1.44 (20.34)
15X-SLR extended region	0.52 (11.93)	3.78 (86.70)	0.49 (11.24)	0.78 (17.89)
15X-SLR excluding the extended region	0.49 (18.15)	1.92 (71.11)	0.35 (12.96)	0.66 (24.44)
15Y-SLR	0.31 (17.23)	1.16 (64.09)	0.20 (10.81)	0.46 (25.20)
15X-PARs	3.82 (41.96)	2.77 (30.41)	0.52 (5.71)	0.28 (3.04)
15Y-PARs	3.64 (39.66)	2.83 (30.86)	0.50 (5.47)	0.28 (3.10)
S. dunnii chromosome 15 region syntenic with the *S. arbutifolia* SLR and PARs				
SLR	0.73 (30.79)	1.41 (59.49)	0.10 (4.18)	0.84 (35.65)
PARs	4.61 (40.30)	4.20 (36.71)	0.55 (4.81)	1.39 (12.14)
S. purpurea				
15W-SLR	1.56 (23.80)	3.16 (48.10)	0.72 (10.90)	0.86 (13.20)
15W-PARs	3.72 (41.90)	2.58 (29.00)	0.37 (4.10)	0.38 (4.30)
15Z-SLR	1.14 (26.80)	1.81 (42.40)	0.25 (5.90)	0.55 (12.80)
15Z-PARs	3.72 (41.90)	2.58 (29.00)	0.37 (4.10)	0.38 (4.30)
S. dunnii chromosome 15 region syntenic with the *S. purpurea* SLRs and PARs				
SLR	1.47 (30.50)	2.51 (52.16)	0.21 (4.36)	1.3 (26.97)
PARs	3.87 (43.05)	3.10 (34.37)	0.44 (4.89)	0.93 (10.39)

[1]The percentages do not add up to 100%, because there are other sequence types (e.g. unannotated DNA sequences) in the relevant regions.

Table 3 Genes in the 15X- and 15Y-SLRs of *Salix arbutifolia* and/or in the 15Z- and 15W-SLRs of *Salix purpurea*.

S. arbutifolia 15X-specific genes	X insertion (including duplicated genes)	Y deletion (including duplicated genes)	Total number (%)
Present in the 15W-SLRs	12	13	25 (20.83)
Present in the 15W- and 15Z-SLRs	29	22	51 (42.50)
Present only in the 15X-SLR	37	7	44 (36.67)
S. arbutifolia 15Y-specific genes	Y insertion	X deletion	Total number
Present in the 15Z-SLRs	1	1	2 (12.50)
Present in the 15Z- and 15W-SLRs	2	5	7 (43.75)
Present only in the 15Y-SLR	6	1	7 (43.75)

The *S. arbutifolia* 15Y- and *S. purpurea* 15Z-SLRs, both of which should carry factors for male functions, have seven and eight partial *ARR17*-like gene duplicates, respectively, which cluster together in the Fig. S9, supporting their previously suggested common ancestry (Wang *et al.*, 2022). Chromosome 15W also carries *ARR17*-like genes, four of which are intact and very similar to copies on chromosome 19 of *S. purpurea* (Fig. S9; Zhou *et al.*, 2020), suggesting that they support female function and were translocated to 15W after the split from *S. arbutifolia*, whose 15X-SLR does not carry such genes.

Among the protein-coding genes in the SLRs of *S. arbutifolia* and *S. purpurea* and also on chromosome 15 of *S. dunnii*, 44 were SCOs. These display three major phylogenetic tree topologies: (I) for two SCOs, the X- and W-linked sequences of *S. arbutifolia* and *S. purpurea* cluster together, while their Y- and Z-linked genes form a well-supported separate clade, consistent with a lack of recombination before these species split (Fig. 3a), and with Z-linked sequences evolving from Y-linked ones, as concluded above; (II) for 31 SCOs, the *S. arbutifolia* 15X-SLR and 15Y-SLR form one cluster, and *S. purpurea* 15Z-SLR and 15W-SLR another, consistent with the species tree; some or all of these genes may still recombine with the sex-determining locus, or recombination may have become suppressed, evolving new evolutionary strata (in either or both species) after the species split (Fig. 3b); (III) finally, for seven SCOs, the 15W-, 15X-, and 15Y-SLRs formed one clade, with the 15Z-SLR sequences sister to it (Fig. 3c). We identified both *S. exigua* alleles of the likely orthologs of the two Topology I SCOs, using the published data (Hu *et al.*, 2023). One sequence clusters with the 15X-15W sequences, and the other with the 15Y-15Z ones. This further supports the conclusion that the 15W and 15Z in willows arose from 15X and 15Y sequences, respectively (Fig. 3a,d). We also identified both allele sequences of orthologs of 24 other single-copy sex-linked genes, using the unphased *S. exigua* chromosome 15 sequences (for 14 SCOs, we could identify only one sequence, Table S18). These results suggest that phased SLRs are needed to test the sex chromosome evolution of *S. exigua*.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

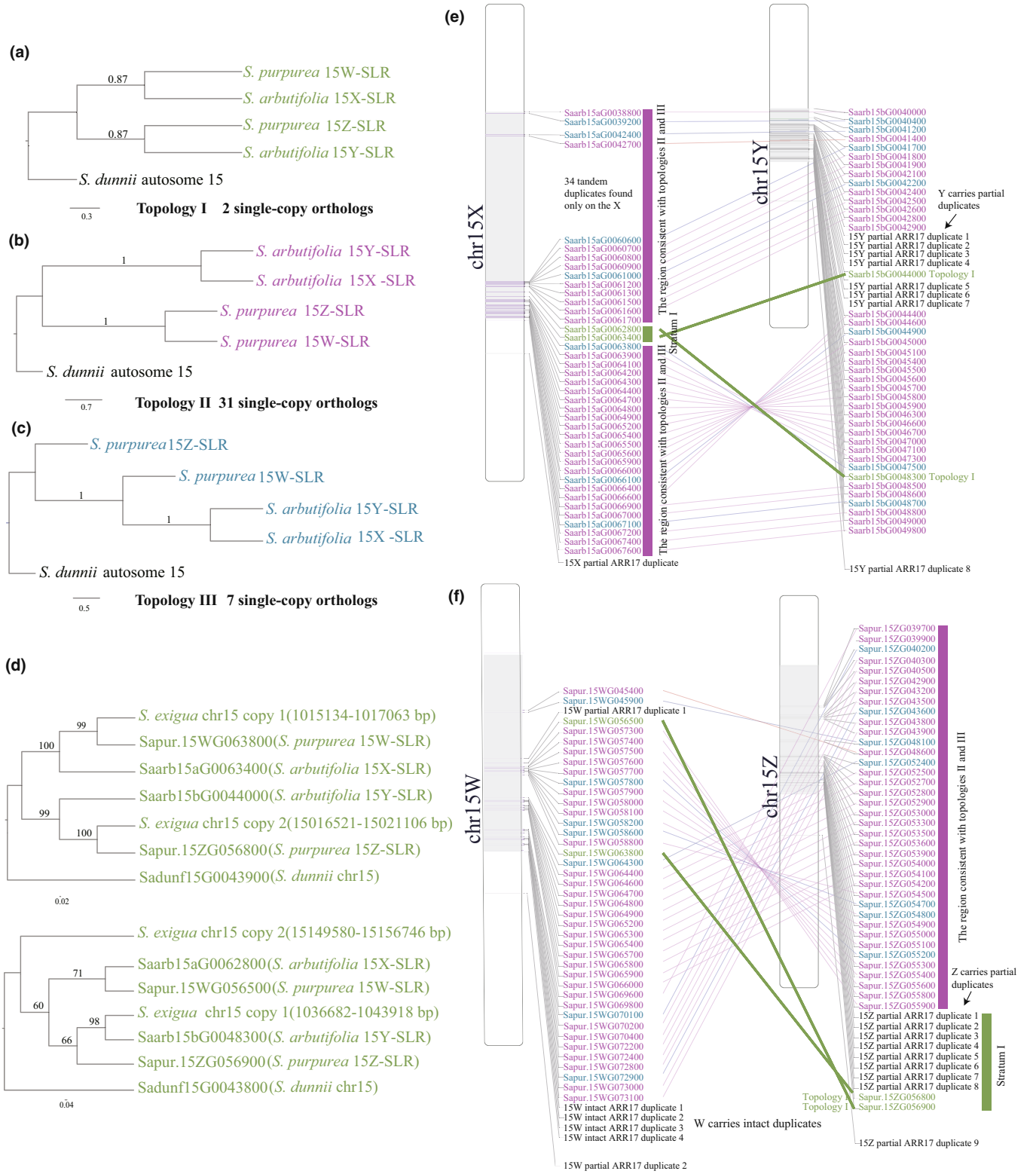

Fig. 3 Heterogametic sex transition history of 15X, 15Y, 15Z, and 15W of willows. (a) Topology I (ASTRAL tree), proto-X-Y recombination stopped before the split of the two species. (b) Topology II (ASTRAL tree), X-Y, and Z-W recombination might have stopped after the split. (c) Topology III (ASTRAL tree) showed that 15W-, 15X- and 15Y-SLRs formed a clade and sister to 15Z-SLR. (d) Gene trees of Topology I genes and their homologs on chromosome 15 of *Salix exigua*. (e, f) showed the positions of these single-copy orthologs with the three tree topologies on 15X, 15Y, 15W, and 15Z. The numbers above branches of ASTRAL trees refer to local posterior probabilities, and the gene trees refer to bootstrap support values. SLR, sex-linked region.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Topology I suggest that at least two genes (in green in Fig. 3d, e) were fully sex-linked before the species split, within a possible old evolutionary stratum 1 of an ancestral SLR. These genes are both within a region spanning 180 kb of the *S. arbutifolia* 15X-SLR (9.72–9.90 Mb), and 460 kb in the *S. purpurea* 15Z-SLR (6.25–6.70 Mb, including partial *ARR17*-like gene duplicates). Partial *ARR17*-like gene duplicates are also found near the Y copy of one *S. arbutifolia* gene, while the second gene is further away in this species' Y.

To estimate the age of this evolutionary stratum, we estimated synonymous site divergence values (Ks) between the gametolog pairs of all genes in the SLRs. High Y-X and W-Z divergence estimates are detected, but they are scattered across the SLRs (Fig. S10; Tables S19, S20), suggesting that these regions have been rearranged, as our assemblies suggest (Figs 1b, 3e, S8). They are not clearly organized into older and younger strata.

Rearrangements have occurred during the considerable evolutionary time since *S. arbutifolia* and *S. purpurea* diverged, as their 15Y and 15Z regions differ in size (10.98 vs 13.25 Mb, respectively), and both differ from the *S. dunnii* autosome 15 (13.81 Mb, Fig. S11). The Z- and W-SLRs in the *S. purpurea* also differ by several inversions (Fig. S12), but again, these may not be fixed Z-W differences, as the SLR contigs relied on a linkage map to scaffold the contigs in the reference genome, not on HiFi sequencing (Zhou *et al.*, 2020; Hyden *et al.*, 2023), and could be mis-assembled.

Differential expression of genes in *S. arbutifolia* and *S. purpurea*

More than 90% of our RNA-Seq reads mapped uniquely to the *S. arbutifolia* genome assembly (Table S21). Among genes in the identified SLRs, many 15X-SLR genes show female-biased expression. If a Y-linked region has undergone genetic degeneration, X-SLR genes will be present in two copies in females, vs only one in males, and, unless dosage compensation has evolved, will show female-biased expression in the transcriptome. This is not the explanation in *S. arbutifolia*, because most 15Y-SLR genes also have X homologs, and we did not separately estimate expression specifically for the X and Y alleles. These results therefore probably reflect expression differences related to sex functions. We also detected overall male-biased expression of 15Y-SLR genes that have X homologs (Table S22). Among the 33 SCOs with topologies I or II, 12 15X-sex-linked genes showed significant female-biased expression, six were male-biased, and 15 either had unbiased or undetectable expression (Fig. 4a). All 34 genes within the 15X-SLR tandem array show female-biased expression (Fig. 4b; Table S23). In the *S. purpurea* 15Z-SLR, 12 of the 33 SCOs were male-biased, while only three showed female-biased expression (10 did not yield expression data; Fig. 4a; Table S24).

We also estimated expression levels in catkins for 120 and 16 genes that appear to be specific to the 15X- or 15Y-SLRs, respectively. The former genes exhibited a significant overall female expression bias, but two X insertion genes showed male-biased expression (Table S25). We detected expression of only one 15Y-SLR specific gene. The genes shared only by the 15W- and 15X-SLRs showed very diverse gene expression patterns. Two genes are shared only by the 15Z- and 15Y-SLRs; we did not obtain expression data from their 15Y-SLR alleles, but the 15Z-SLR genes showed male biases (Table S26). The 15W-SLR includes 175 genes that are specific to this region (of which 155 are not present in the *S. arbutifolia* 15X-SLR), and are expressed in female *S. purpurea* catkins (Hyden *et al.*, 2021; Table S27).

Discussion

Expansion of the X-linked regions in willows

In *S. arbutifolia*, as in other Salicaceae species, the 15X-SLR is considerably longer than the Y-linked region. This is due to X expansion, not Y degeneration, as the *S. arbutifolia* 15Y-SLR shows only slight degeneration, and (including the three detected pseudogenes) has lost functional copies of only 11.49% of the ancestral genes. Furthermore, the 15X-SLR includes an 'extended region' that is not present in the *S. dunnii* chromosome 15. This region has an extremely high repetitive sequence density (Table 2) and all genes are duplicates, not single copy. Although X and Y chromosomes may change, and some animal X chromosomes may have expanded (Bellott *et al.*, 2010), changes in the X are expected to be much smaller than those in Y. In SLRs, the accumulation of repeats should mainly affect Y-linked regions and is in females largely prevented by recombination of the corresponding X-linked regions; preferential Y-linked accumulation is indeed observed (Bachtrog, 2013; Charlesworth, 2017). Initial expansion of the Y-linked region is therefore expected, though a larger X may later evolve, through Y degeneration, as observed in many mammals (Bergero & Charlesworth, 2009). The SLR size differences observed here are the opposite of those just predicted (Fig. 5a). The results described here support the view above that *Salix* species' chromosomes have extensive pericentromeric regions that recombine rarely, and hence are repetitive.

Salix arbutifolia's large 15X-SLR, relative to its Y-linked counterpart, does not reflect a reduced size of the Y, or loss of genes by genetic degeneration, which is the most common reason why the X is larger than the Y, as in mammals and *Drosophila* (Beukeboom & Perrin, 2014). This is consistent with the generally low Y-X Ks values (Fig. S10), indicating that, even in the old stratum, the Y has not been diverging from the X for a long evolutionary time. We therefore suggest that pericentromeric locations may explain the SLR recombination suppression, and hence their repeat-richness (Fig. S6). The X 'extended region' might thus reflect repetitive sequence accumulation in an ancestrally nonrecombining region, as is also likely in papaya (Ming *et al.*, 2011; Bachtrog, 2013; Charlesworth, 2017). *Salix dunnii* also has a larger 7X- than 7Y-SLR (5.87 vs 2.95 Mb, respectively), and the 7Y-SLR lacks only one gene present on autosome 7 of outgroups, supporting an absence of degeneration (He *et al.*, 2023).

Change in the heterogametic sex in willows

The Topology I genes shared by *S. arbutifolia* and *S. purpurea* support the conclusion that recombination in the SLR was

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

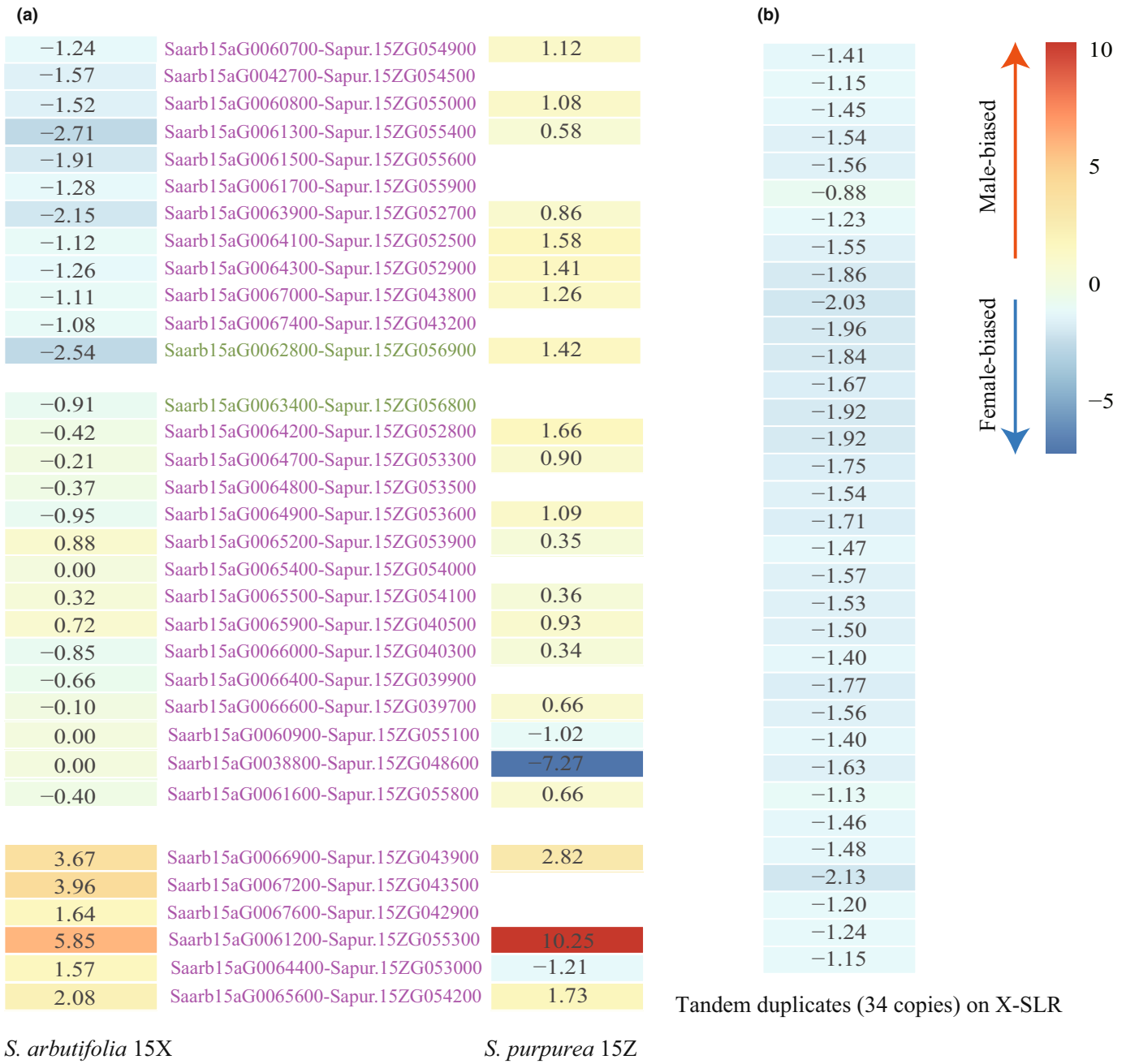

Fig. 4 Gene expression estimates ($\log_2$ M/F ratios). (a) *Salix arbutifolia* (left-hand panel) and their *Salix purpurea* homologs (right-hand panel) for all 33 single-copy orthologs with topologies I and II. (b) Tandemly duplicated genes (34 copies in total) in the *S. arbutifolia* 15X-SLR. Blank areas indicate the absence of expression data. SLR, sex-linked region

already rare in the common ancestor of these 15XY and 15ZW species. The physically small size of the region in both species (Fig. 3e,f) suggests that the same gene or genes might be involved in sex determination in both systems.

We argued above that the *Salix* species so far studied appear to have nondegenerated systems (Wang *et al.*, 2022; He *et al.*, 2023), which make turnovers plausible, compared with highly degenerated systems (Bull, 1983). A study in a frog (Ogata *et al.*, 2021) documented that, as predicted in Bull (1983), a slight advantage to the new system, such as selection to maintain a 1 : 1 sex ratio can favor a turnover. Our gene trees support an X to W, and Y to Z, transition (Figs 3d, 5b). This could have involved a new and more highly expressed female-determining factor on an X chromosome (perhaps the four intact *ARR17*-like gene copies in *S. purpurea*) creating a 'strong W' that makes the W/Y heterozygotes (where the W was formerly the X) female, instead of male. In other words, the W factor behaves as dominant over the Y factor (though these factors may not be alleles of

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

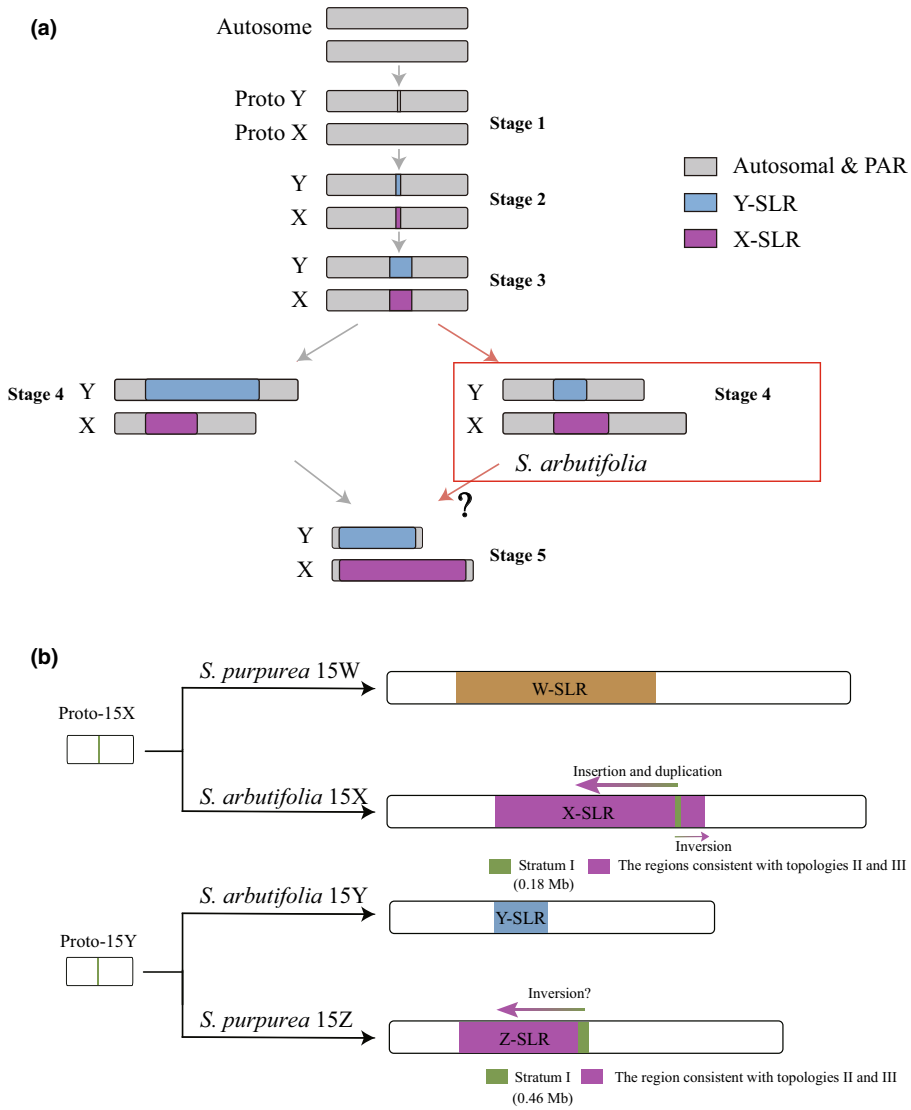

Fig. 5 (a) Stages of sex chromosome evolution. The gray arrows indicate the different stages, adapted from Bergero & Charlesworth (2009). The red arrows (in the red box) indicate an alternative route for sex chromosome evolution that is found in the willow *Salix arbutifolia* (a species with a 15 XY system). (b) Hypothetical evolutionary relationship of 15X, 15Y, 15Z, and 15W, and the changes that may have occurred during their evolution. SLR, sex-linked region.

the same gene), allowing a transition to a ZW system with the Z derived from the ancestral Y (which is viable when homozygous, and still carries masculinizing factors; Ming *et al.*, 2011).

In studying heterogamety changes in the platyfish, *Xiphophorus maculatus* (Volff & Schartl, 2001) suggested that a changed copy number of a sex-determining gene might be involved, and might allow the same gene to act as a dominant masculinizing factor in an XY system and a recessive one in a ZW system. It is therefore intriguing that the systems in *Salix* involves copy number differences, and that the sex-determining regions, as in the platyfish, are highly repetitive genome regions (Tomaszkiewicz *et al.*, 2014), where copy number changes can readily occur. The angiosperm system studied here may be suitable for testing Volff & Schartl's model. A similar mechanism may apply to other changes in the heterogametic sex.

Interestingly, the *S. arbutifolia* 15X-SLR (but not the 15Y-SLR) includes a tandem array of 34 duplicate gene copies showing female-biased expression (Fig. 4b). Since the *S. purpurea* 15ZW-SLRs do not carry such a tandem array, the duplications probably occurred in *S. arbutifolia* after the two species split, possibly because they have some female-specific function. Intriguingly, the repeated gene's *A. thaliana* homolog is BRR2C, which encodes the DExH-box ATP-dependent RNA helicase, a highly conserved spliceosome component required for efficient splicing of this species' Flowering locus C (FLC) introns, and for regulation of two genes, Flowering locus T (FT) and Suppressor of Overexpression of Constans 1 (SOC1; Mahrez *et al.*, 2016). However, additional studies are needed to test the functions of the tandem array.

15X-SLR specific genes in *S. arbutifolia* showed predominantly female-biased expression, and a gene that is inserted into the 15Y-SLR is male expressed specifically (Table S25), suggesting that the X- and Y- linked regions may respectively have accumulated female-biased and male-biased factors. The SCOs with topologies I and II in Fig. 3 generally show female-biased expression of the 15X alleles, and male-biased expression of 15Z ones (Fig. 4a; Table S24). Some single-copy sex-linked genes may therefore have evolved expression differences after the change in heterogamety.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Acknowledgements

This study was financially supported by the National Natural Science Foundation of China (grant no.: 32171813) and the Special Fund for Scientific Research of Shanghai Landscaping & City Appearance Administrative Bureau (grant nos.: G232403 & G242417). We are also grateful to De-Yan Wang, Tao Ma, Zhi-Qing Xue, and Xin-Jie Cai for their suggestions for our analysis, and Xiao-Bo An for sampling plants for the study, and Nan Hu and Matthew S. Olson for their kind help for analysis of *S. exigua*.

Competing interests

None declared.

Author contributions

Yi Wang and LH planned and designed the research; Yi Wang, LH, RGZ, Yuan Wang and GNG analyzed the data; Yi Wang, LH, EH, ZXZ and DC wrote the paper.

ORCID

Deborah Charlesworth https://orcid.org/0000-0002-3939-9122
Elvira Hörandl https://orcid.org/0000-0002-7600-1128
Li He https://orcid.org/0000-0002-4591-8056
Yi Wang https://orcid.org/0000-0001-6611-5319
Zhi-Xiang Zhang https://orcid.org/0000-0002-1161-5065

Data availability

Sequencing data for genome assembly and annotation can be downloaded from the National Center for Biotechnology Information (NCBI) under the BioProject accession no.: PRJNA882493. The genome assembly and annotation can be downloaded from the National Genomics Data Center (NGDC) under the BioProject accession no.: PRJCA016000. Transcriptome data from male and female catkins from two *S. arbutifolia* individuals B649AF and B649V can be downloaded from the NCBI Sequence Read Archive (SRA) under the BioProject accession no.: PRJNA990983.

References

Akagi T, Henry IM, Tao R, Comai L. 2014. Plant genetics. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **346**: 646–650.

Almeida P, Proux-Wera E, Churcher A, Soler L, Dainat J, Pucholt P, Nordlund J, Martin T, Ronnberg-Wastljung AC, Nystedt B *et al.* 2020. Genome assembly of the basket willow, *Salix viminalis*, reveals earliest stages of sex chromosome expansion. *BMC Biology* **18**: 78.

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

Argus GW, Eckenwalder JE, Kiger RW. 2010. Salicaceae. In: Flora of North America Editorial Committee, ed. *Flora of north America, vol. 7*. New York, NY, USA: Oxford University Press, 3–164.

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews. Genetics* **14**: 113–124.

Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, Kremitzki C, Brown LG, Rozen S, Warren WC, Wilson RK *et al.* 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466**: 612–616.

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

Bergero R, Forrest A, Charlesworth D. 2008. Active miniature transposons from a plant genome and its nonrecombining Y chromosome. *Genetics* **178**: 1085–1092.

Beukeboom LW, Perrin N. 2014. *The evolution of sex determination.* Oxford, UK: Oxford University Press.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12**: 59–60.

Bull JJ. 1983. *Evolution of sex determining mechanisms.* Menlo Park, CA, USA: The Benjamin/Cummings Publishing Co.

Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188–196.

Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11**: 113–116.

Charlesworth B, Charlesworth D. 1978. A model for the evolution of dioecy and gynodioecy. *The American Naturalist* **112**: 975–997.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215–220.

Charlesworth D. 2017. Evolution of recombination rates between sex chromosomes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **372**: 20160456.

Charlesworth D. 2019. Young sex chromosomes in plants and animals. *New Phytologist* **224**: 1095–1107.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**: 118–128.

Charlesworth D, Harkess A. 2024. Why should we study plant sex chromosomes? *Plant Cell*: koad278.

Chen JH, Huang Y, Brachi B, Yun QZ, Zhang W, Lu W, Li HN, Li WQ, Sun XD, Wang GY *et al.* 2019. Genome-wide analysis of Cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. *Nature Communications* **10**: 5230.

Chen S, Zhou Y, Chen Y, Gu J. 2018. FASTP: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**: 170–175.

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

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

Dong SS, He WM, Ji JJ, Zhang C, Guo Y, Yang TL. 2021. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics* **22**: bbaa227.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP *et al.* 2017. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**: 92–95.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems* **3**: 95–98.

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

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. REPEATMODELER2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences, USA* **117**: 9451–9457.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Fraisse C, Picard MAL, Vicoso B. 2017. The deep conservation of the Lepidoptera Z chromosome suggests a non-canonical origin of the W. *Nature Communications* **8**: 1486.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

Furman BLS, Cauret CMS, Knytl M, Song XY, Premachandra T, Ofori-Boateng C, Jordan DC, Horb ME, Evans BJ. 2020. A frog with three sex chromosomes that co-mingle together in nature: *Xenopus tropicalis* has a degenerate W and a Y that evolved from a Z chromosome. *PLoS Genetics* **16**: e1009121.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**: 3420–3435.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**: 644–652.

Gulyaev S, Cai XJ, Guo FY, Kikuchi S, Applequist WL, Zhang ZX, Horandl E, He L. 2022. The phylogeny of *Salix* revealed by whole genome re-sequencing suggests different sex-determination systems in major groups of the genus. *Annals of Botany* **129**: 485–498.

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD *et al.* 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**: 5654–5666.

Hall AB, Qi Y, Timoshevskiy V, Sharakhova MV, Sharakhov IV, Tu Z. 2013. Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females. *BMC Genomics* **14**: 273.

Handley LJ, Ceplitis H, Ellegren H. 2004. Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* **167**: 367–376.

He L, Jia KH, Zhang RG, Wang Y, Shi TL, Li ZC, Zeng SW, Cai XJ, Wagner ND, Hörandl E *et al.* 2021. Chromosome-scale assembly of the genome of *Salix dunnii* reveals a male-heterogametic sex determination system on chromosome 7. *Molecular Ecology Resources* **21**: 1966–1982.

He L, Wang Y, Wang Y, Zhang R, Wang Y, Hörandl E, Mank JE, Ming R. 2023. Allopolyploidization from two dioecious ancestors leads to recurrent evolution of sex chromosomes and reversion to autosomes. *bioRxiv*. doi: 10.1101/2023.10.18.562857.

Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**: 2253–2255.

Hu N, Sanderson BJ, Guo M, Feng G, Gambhir D, Hale H, Wang D, Hyden B, Liu J, Smart LB *et al.* 2023. Evolution of a ZW sex chromosome system in willows. *Nature Communications* **14**: 7144.

Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Molecular Biology and Evolution* **34**: 2115–2122.

Hyden B, Carlson CH, Gouker FE, Schmutz J, Barry K, Lipzen A, Sharma A, Sandor L, Tuskan GA, Feng G *et al.* 2021. Integrative genomics reveals paths to sex dimorphism in *Salix purpurea* L. *Horticulture Research* **8**: 170.

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.

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biology* **21**: 241.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al.* 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.

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

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

Killick R, Eckley IA. 2014. changepoint: an R package for changepoint analysis. *Journal of Statistical Software* **58**: 1–19.

Kim D, Langmead B, Salzberg SL. 2015. Hisat: a fast spliced aligner with low memory requirements. *Nature Methods* **12**: 357–360.

Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. *Science* **286**: 964–967.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

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

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

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

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

Ma WJ, Veltsos P, Sermier R, Parker DJ, Perrin N. 2018. Evolutionary and developmental dynamics of sex-biased gene expression in common frogs with proto-Y chromosomes. *Genome Biology* **19**: 156.

Mahrez W, Shin J, Muñoz-Viana R, Figueiredo DD, Trejo-Arellano MS, Exner V, Siretskiy A, Gruissem W, Köhler C, Hennig L. 2016. BRR2a affects flowering time via FLC splicing. *PLoS Genetics* **12**: e1005924.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al.* 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**: 1297–1303.

Ming R, Bendahmane A, Renner SS. 2011. Sex chromosomes in land plants. *Annual Review of Plant Biology* **62**: 485–514.

Miura I. 2007. An evolutionary witness: the frog rana rugosa underwent change of heterogametic sex from XY male to ZW female. *Sexual Development* **1**: 323–331.

Mrnjavac A, Khudiakova KA, Barton NH, Vicoso B. 2023. Slower-X: reduced efficiency of selection in the early stages of X chromosome evolution. *Evolution Letters* **7**: 4–12.

Muller NA, Kersten B, Leite Montalvao AP, Mahler N, Bernhardsson C, Brautigam 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.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-Tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Ogata M, Lambert M, Ezaz T, Miura I. 2018. Reconstruction of female heterogamety from admixture of XX-XY and ZZ-ZW sex-chromosome systems within a frog species. *Molecular Ecology* **27**: 4078–4089.

Ogata M, Suzuki K, Yuasa Y, Miura I. 2021. Sex chromosome evolution from a heteromorphic to a homomorphic system by inter-population hybridization in a frog. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **376**: 20200105.

Papadopulos AS, Chester M, Ridout K, Filatov DA. 2015. Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proceedings of the National Academy of Sciences, USA* **112**: 13021–13026.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**: 290–295.

Picard MAL, Cosseau C, Ferre S, Quack T, Grevelding CG, Coute Y, Vicoso B. 2018. Evolution of gene dosage on the Z-chromosome of schistosome parasites. *eLife* **7**: e35684.

Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei T, Creevey C, Kuhn M *et al.* 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Research* **42**: D231–D239.

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

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Pryszcz LP, Gabaldon T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research* **44**: e113.

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

Rifkin JL, Beaudry FEG, Humphries Z, Choudhury BI, Barrett SCH, Wright SI. 2021. Widespread recombination suppression facilitates plant sex chromosome evolution. *Molecular Biology and Evolution* **38**: 1018–1030.

Sanderson BJ, Gambhir D, Feng G, Hu N, Cronk QC, Percy DM, Freaner FM, Johnson MG, Smart LB, Keefover-Ring K *et al.* 2023. Phylogenomics reveals patterns of ancient hybridization and differential diversification that contribute to phylogenetic conflict in willows, poplars, and close relatives. *Systematic Biology* **72**: 1220–1232.

Saunders PA, Neuenschwander S, Perrin N. 2018. Sex chromosome turnovers and genetic drift: a simulation study. *Journal of Evolutionary Biology* **31**: 1413–1419.

Sayyari E, Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* **33**: 1654–1668.

She H, Liu Z, Li S, Xu Z, Zhang H, Cheng F, Wu J, Wang X, Deng C, Charlesworth D *et al.* 2023. Evolution of the spinach sex-linked region within a rarely recombining pericentromeric region. *Plant Physiology* **193**: 1263–1280.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J *et al.* 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**: 539.

Smirnov V, Warnow T. 2021. MAGUS: multiple sequence alignment using graph clUStering. *Bioinformatics* **37**: 1666–1672.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**: 637–644.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. *Science* **320**: 486–488.

Tomaszkiewicz M, Chalopin D, Schartl M, Galiana D, Volff JN. 2014. A multicopy Y-chromosomal SGNH hydrolase gene expressed in the testis of the platyfish has been captured and mobilized by a Helitron transposon. *BMC Genetics* **15**: 44.

Vera Alvarez R, Pongor LS, Mariño-Ramírez L, Landsman D. 2019. TPMCalculator: one-step software to quantify mRNA abundance of genomic features. *Bioinformatics* **35**: 1960–1962.

Vicoso B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nature Ecology & Evolution* **3**: 1632–1641.

Volff JN, Schartl M. 2001. Variability of genetic sex determination in poeciliid fishes. *Genetica* **111**: 101–110.

Wang D, Li Y, Li M, Yang W, Ma X, Zhang L, Wang Y, Feng Y, Zhang Y, Zhou R *et al.* 2022. Repeated turnovers keep sex chromosomes young in willows. *Genome Biology* **23**: 200.

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

Wang Y, Cai X, Zhang Y, Hörandl E, Zhang Z, He L. 2023. The male-heterogametic sex determination system on chromosome 15 of *Salix triandra* and *Salix arbutifolia* reveals ancestral male heterogamety and subsequent turnover events in the genus *Salix*. *Heredity* **130**: 122–134.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Xu GC, Xu TJ, Zhu R, Zhang Y, Li SQ, Wang HW, Li JT. 2019. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience* **8**: giy157.

Xu L, Wa Sin SY, Grayson P, Edwards SV, Sackton TB. 2019. Evolutionary dynamics of sex chromosomes of paleognathous birds. *Genome Biology and Evolution* **11**: 2376–2390.

Xue Z, Appleguist WL, Hörandl E, He L. 2024. Sex chromosome turnover plays an important role in the maintenance of barriers to post-speciation introgression in willows. *Evolution Letters*: qrae013.

Yang W, Wang D, Li Y, Zhang Z, Tong S, Li M, Zhang X, Zhang L, Ren L, Ma X *et al.* 2021. A general model to explain repeated turnovers of sex determination in the Salicaceae. *Molecular Biology and Evolution* **38**: 968–980.

Yang Z. 2007. Paml 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Yue J, Krasovec M, Kazama Y, Zhang X, Xie W, Zhang S, Xu X, Kan B, Ming R, Filatov DA. 2023. The origin and evolution of sex chromosomes, revealed by sequencing of the Silene latifolia female genome. *Current Biology* **33**: 2504–2514 e2503.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 153.

Zhang H, Sigeman H, Hansson B. 2022. Assessment of phylogenetic approaches to study the timing of recombination cessation on sex chromosomes. *Journal of Evolutionary Biology* **35**: 1721–1733.

Zhang S, Wu Z, Ma ZJ, Han X, Jiang Z, Liu S, Xu J, Jiao P, Li Z. 2022. Chromosome-scale assemblies of the male and female *Populus euphratica* genomes reveal the molecular basis of sex determination and sexual dimorphism. *Communications Biology* **5**: 1186.

Zhang Z, Yu J. 2006. Evaluation of six methods for estimating synonymous and nonsynonymous substitution rates. *Genomics, Proteomics & Bioinformatics* **4**: 173–181.

Zhou R, Macaya-Sanz D, Carlson CH, Schmutz J, Jenkins JW, Kudrna D, Sharma A, Sandor L, Shu S, Barry K *et al.* 2020. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biology* **21**: 38.

Zhou R, Macaya-Sanz D, Rodgers-Melnick E, Carlson CH, Gouker FE, Evans LM, Schmutz J, Jenkins JW, Yan J, Tuskan GA *et al.* 2018. Characterization of a large sex determination region in *Salix purpurea* L. (Salicaceae). *Molecular Genetics and Genomics* **293**: 1437–1452.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Genome-wide analysis of chromatin interactions in the *Salix arbutifolia* genome based on Hi-C data.

Fig. S2 All the chromosome quotient results of haplotype *a* of *Salix arbutifolia*.

Fig. S3 All the chromosome quotient results of haplotype *b* of *Salix arbutifolia*.

Fig. S4 All the F_{ST} results of haplotype *a* of *Salix arbutifolia*.

Fig. S5 All the F_{ST} results of haplotype *b* of *Salix arbutifolia*.

Fig. S6 Gene density, transposable element density, LTR-Copia density and LTR-Gypsy density of genome *Salix arbutifolia*.

Fig. S7 The 34 tandemly repeated genes.

Fig. S8 Circular plot showing *Salix arbutifolia* 15X- and 15Y-specific genes that are present in the *Salix purpurea* 15W- and/or 15Z-SLRs.

Fig. S9 The gene tree of all intact and partial *ARR17*-like genes of *Salix dunnii*, *Salix purpurea*, and *Salix arbutifolia*.

Fig. S10 Distribution of Ks values of *Salix arbutifolia* 15X-15Y gene pairs and *Salix purpurea* 15Z-SLR-15W-SLR gene pairs.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation

Fig. S11 Collinearity analysis of different sex chromosomes.

Fig. S12 Collinearity analysis of *Salix purpurea* 15ZW and *Salix dunnii* autosome 15.

Table S1 Details of plant materials used in this study.

Table S2 Details of 17 species of Salicaceae and *Arabidopsis thaliana* used as references for homology-based gene annotation.

Table S3 Sequencing statistics of WGS-HiFi, WGS-Illumina paired-end sequences, Hi-C, and RNA-Seq datasets.

Table S4 Statistics of mapping results of different sequencing data sets.

Table S5 Estimated numbers of sequences in the *a* and *b* haplotypes of *Salix arbutifolia* chromosomes.

Table S6 Statistics of RNAs of the genome of *Salix arbutifolia*.

Table S7 Functional annotation of the predicted genes of *Salix arbutifolia*.

Table S8 Summary of repeat contents of the genome of *Salix arbutifolia*.

Table S9 Summary of mapping results of 41 samples of *Salix arbutifolia* using haplotype *a* as reference.

Table S10 Summary of mapping results of 41 samples of *Salix arbutifolia* using haplotype *b* as reference.

Table S11 Genes in the 15X- and 15Y-linked regions of *Salix arbutifolia*.

Table S12 Pseudogenes of sex chromosome 15X and 15Y of *Salix arbutifolia*.

Table S13 The type and size of the repeat sequences in the 15X-SLR extended region.

Table S14 List of all 34 tandem genes in the 15X-SLR extended region.

Table S15 *Salix arbutifolia* 15X-SLR specific genes that have homologs on other chromosomes.

Table S16 *Salix arbutifolia* 15Y-SLR specific genes that have homologs on other chromosomes.

Table S17 Losses of sex-linked genes in *Salix arbutifolia* and *Salix purpurea*.

Table S18 Homologous copies of Topology I-III genes on chromosome 15 of *Salix exigua*.

Table S19 Synonymous site divergence estimates between gametologous gene pairs in *Salix arbutifolia*.

Table S20 Synonymous site divergence estimates between gametologous gene pairs in *Salix purpurea*.

Table S21 Statistics of transcriptome data sets of female and male catkins.

Table S22 Statistics on the number of sex-biased genes on chromosomes of *Salix arbutifolia*.

Table S23 Expression of 34 genes in the tandem array within the 15X-SLR.

Table S24 The gene expression patterns for 33 single-copy orthologs representing topologies I and II.

Table S25 The gene expression (TPM values) of *Salix arbutifolia* 15X- and 15Y-specific genes.

Table S26 The expression of specific genes shared by the 15W- and 15X-SLRs, or by the 15Z- and 15Y-SLRs.

Table S27 Summary of gene expression in the sex-linked regions of *Salix arbutifolia* and *Salix purpurea*.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

© 2024 The Authors
New Phytologist © 2024 New Phytologist Foundation