

Allopolyploidization from two dioecious ancestors leads to recurrent evolution of sex chromosomes

Received: 13 March 2024

Accepted: 31 July 2024

Published online: 12 August 2024

 Check for updates


Li He^{1,10} , Yuàn Wang^{1,10} , Yi Wang^{1,2} , Ren-Gang Zhang^{3,4} , Yuán Wang¹,
Elvira Hörandl⁵ , Tao Ma⁶ , Yan-Fei Mao⁷, Judith E. Mank⁸ & Ray Ming⁹ 

Polyploidization presents an unusual challenge for species with sex chromosomes, as it can lead to complex combinations of sex chromosomes that disrupt reproductive development. This is particularly true for allopolyploidization between species with different sex chromosome systems. Here, we assemble haplotype-resolved chromosome-level genomes of a female allotetraploid weeping willow (*Salix babylonica*) and a male diploid *S. dunnii*. We show that weeping willow arose from crosses between a female ancestor from the *Salix*-clade, which has XY sex chromosomes on chromosome 7, and a male ancestor from the *Vetrix*-clade, which has ancestral XY sex chromosomes on chromosome 15. We find that weeping willow has one pair of sex chromosomes, ZW on chromosome 15, that derived from the ancestral XY sex chromosomes in the male ancestor of the *Vetrix*-clade. Moreover, the ancestral 7X chromosomes from the female ancestor of the *Salix*-clade have reverted to autosomal inheritance. Duplicated intact *ARR17*-like genes on the four homologous chromosomes 19 likely have contributed to the maintenance of dioecy during polyploidization and sex chromosome turnover. Taken together, our results suggest the rapid evolution and reversion of sex chromosomes following allopolyploidization in weeping willow.

Polyploidy is an important feature of genome evolution^{1,2}, and although highly degenerate sex chromosomes might act as a barrier to polyploidy³, polyploidy in lineages with less differentiated sex chromosomes has been observed (e.g., dioecious plants, amphibians, and fishes, etc.)^{1,4,5}. Polyploidization presents several interesting questions

in the context of sex chromosomes^{3,4,6}. For example, how does the polyploid lineage overcome unbalanced sex chromosome dose effects? In autopolyploid lineages, where both genome copies arise from the same species⁷, polyploid progeny could inherit the sex chromosomes from the diploid ancestor⁸. However, allopolyploidy,

¹Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China. ²Laboratory of Systematic Evolution and Biogeography of Woody Plants, School of Ecology and Nature Conservation, Beijing Forestry University, Beijing 100083, China. ³Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201 Yunnan, China. ⁴University of Chinese Academy of Sciences, Beijing 101408, China. ⁵Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), University of Göttingen, Göttingen, Germany. ⁶Key Laboratory for Bio-Resource and Eco-Environment of Ministry of Education & Sichuan Zoige Alpine Wetland Ecosystem National Observation and Research Station, College of Life Science, Sichuan University, Chengdu, China. ⁷CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China. ⁸Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada. ⁹Centre for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou 350002, China. ¹⁰These authors contributed equally: Li He, Yuàn Wang.

 e-mail: lhe@cemps.ac.cn

where hybridization between species combines one duplicated copy of each ancestral genome, could lead to diverse sex chromosome combinations when the diploid ancestors carry different sex chromosome types (XX/XY, male heterogametic and ZW/ZZ, female heterogametic), or even the same sex chromosome type but on different chromosomes^{9,10}. The increased sex chromosome complement in polyploids could disrupt the sex-determination process, leading to the loss of dioecy, such as the reversal to monoecy observed in polyploid persimmon^{11–14}. Alternatively, the allopolyploidization from lineages with competing sex chromosome systems could lead to transitions between sex-determining systems, but how this occurs, and how the allele dosage effects of competing sex chromosome systems are overcome, remains an open question.

Allopolyploidization also raises interesting questions about how one of the duplicated sex chromosome pairs reverts to autosomal inheritance. Most cases of polyploidization are followed by rediploidization, sometimes instantaneously from the point of polyploidization, but more often rediploidization is a gradual process^{15,16}. Rediploidization presents peculiar issues for duplicated sex chromosomes, as only one pair is retained as the sex-determining chromosomes^{4,6}, as has been observed in *Rumex acetosella*¹⁷, although see the ZZZW/ZZZZ system in autotetraploid *Salix polyclona*⁸. Presumably, the superfluous pair of sex chromosomes transition into autosomes, but the XY or ZW chromosomes of a pair may carry different gene content, and some genes might be lost in the rediploidization process⁴. For example, doublesex and mab-3 related transcription factor (*DMRT1*) gene copies, male-related autosomal genes, were pseudogenized multiple times independently in *Xenopus* following polyploidization in order to balance sex determination¹⁸. However, the retention and loss of sex chromosomes during allopolyploidization, and how this affects gene content and sex determination, remains largely unexplored due to the previous lack of haplotype-resolved genome assemblies of polyploid dioecious plants.

The weeping willow (*Salix babylonica*), native to eastern Asia, is a frequently cultivated ornamental tree in the northern hemisphere^{19,20}. The species is tetraploid ($2n = 76$) and a member of a tetraploid group of *Salix*²¹, that likely arose from crosses between species from the *Salix*- and *Vetrix*-clade within the genus *Salix* according to phylogenetic incongruence between the chloroplast and nuclear trees of *Salix*^{10,20,22}. Diploid species within the *Salix*-clade have XY sex chromosomes on chromosome 7, whereas species in the *Vetrix*-clade have XY or ZW sex chromosomes on chromosome 15²³, with 15XY inferred as the ancestral sex chromosomes of the clade^{23–26}. The weeping willow is dioecious and produces viable seeds^{22,27}, suggesting that it also has sex chromosomes and has overcome the allele dosage effect of the two different sex chromosome systems of its ancestors. This system provides an opportunity to understand how allopolyploidization resulting from

the hybridization of two species with different sex chromosome systems resolves into stable dioecy.

The genetic control of the sexes in the Salicaceae is well known from poplars (*Populus*), the sister genus of *Salix*. In the male-specific region of the Y chromosome, duplicates of the partial *ARABIDOPSIS RESPONSE REGULATOR 17* (*ARR17*) orthologue produce small RNAs (sRNA) that silence the intact *ARR17*-like gene of poplar sex chromosome 19, thereby indirectly activating *PISTILLATA* (*PI*, required for stamen development) and suppressing female development^{28–30}. Partial *ARR17*-like duplicates have also been found in the Y-linked regions (7Y in the *Salix*-clade and 15Y in the *Vetrix*-clade) of willows, which can produce sRNAs to suppress the expression of intact *ARR17*-like genes on chromosome 19²⁴. The diploid 15XY and 15ZW species of the *Vetrix*-clade inherited the sex-linked region (SLR) from their 15XY ancestor, with the W- and X-SLR coming from the ancestral X-SLR, and Z- and Y-SLR from the ancestral Y-SLR. The ancestral Y-linked genes are close to partial *ARR17*-like duplicates^{23–25}, and two ancestral X-Y homologous gene pairs diverged in the progenitor of the *Vetrix*-clade²⁵.

To understand how different parental sex chromosome systems affect allopolyploidization, we assemble a haplotype-resolved chromosome-level genome of a female allotetraploid weeping willow and a haplotype-resolved genome of a male diploid *S. dunnii* of the *Salix*-clade with an XY system on chromosome 7. We previously assembled a female genome, which includes 7X, of *S. dunnii* based on Oxford Nanopore Technologies (ONT) long reads³¹, and the work reported here includes the 7Y chromosome. We show that the weeping willow inherited two alternative sex chromosome systems, 7XX from a female ancestor of the *Salix*-clade, and most likely 15XY from a male ancestor of the *Vetrix*-clade. Following allopolyploidization, the *Salix*-clade 7X sex chromosomes reverted to autosomal inheritance, while the *Vetrix*-clade 15XY sex chromosome system gave rise to a 15ZW sex chromosome system to form stable dioecy and female heterogamety observed in the weeping willow today.

Results

Genome assembly and annotation

From a single female *S. babylonica*, we generated 46 Gb (64X) of HiFi reads, 79 Gb (94X) of Illumina reads, and 164 Gb (254X) of Hi-C reads (Supplementary Table 1). From a single male *S. dunnii* individual (with XY sex chromosomes on chromosome 7³¹), we generated 39 Gb (112X) of HiFi reads, 40 Gb (108X) of Illumina reads, and 37 Gb (107X) of Hi-C reads (Supplementary Table 1). After assembling each species separately with the PacBio HiFi and Hi-C reads, we used Illumina short reads to correct errors for each genome. The *S. babylonica* assembly was 1286 Mb, comprised of 102 contigs (contig N50 = 16 Mb) (Table 1), with a final chromosome-scale assembly of 76 pseudochromosomes and gap-free sex chromosomes (Fig. 1a, Supplementary Fig. 1, and Supplementary Table 2). We obtained an *S. dunnii* assembly of 696 Mb, comprised of 40 contigs (contig N50 = 19 Mb) and a final gap-free chromosome-scale *S. dunnii* genome with 38 pseudochromosomes (Fig. 1b, Table 1, and Supplementary Fig. 2).

For the *S. babylonica* assembly, about 98.91% of Illumina short reads and 99.43% HiFi reads could be aligned back to the genome assembly, and 99.41 and 95.79% of the assembly was covered by at least 20X reads, respectively. Similarly, about 99.95% of Illumina short reads and 99.86% of HiFi reads were mapped back to the *S. dunnii* genome assembly, and around 99.18 and 99.86% of the assembly was covered by at least 20X reads, respectively. BUSCO analysis suggested that 1422 (98.8%) highly conserved core proteins were identified in the *S. babylonica* genome, and 1416 (98.4%) highly conserved core proteins in the *S. dunnii* genome (Supplementary Table 3).

We recovered 525.5 Mb (40.86%) and 333.7 Mb (47.92%) of repetitive sequences in the *S. babylonica* and *S. dunnii* genomes (Supplementary Data 1). Among them, long terminal repeats (LTRs) represented the most common component, accounting for 19.37% of

Table 1 | Statistics of the *S. babylonica* and *S. dunnii* genome assemblies

| | <i>S. babylonica</i> | <i>S. dunnii</i> |
|----------------------------|----------------------|------------------|
| Total assembly size (Mb) | 1286 | 696 |
| Total number of contigs | 102 | 40 |
| Maximum contig length (Mb) | 32 | 37 |
| Minimum contig length (Kb) | 64 | 156 |
| Contig N50 length (Mb) | 16 | 19 |
| Contig N50 count | 33 | 16 |
| Contig N90 length (Mb) | 10 | 13 |
| Contig N90 count | 70 | 33 |
| Gap number | 24 | 0 |
| GC content (%) | 34 | 33 |
| Gene number | 123,231 | 65,485 |
| Repeat content (%) | 41 | 48 |

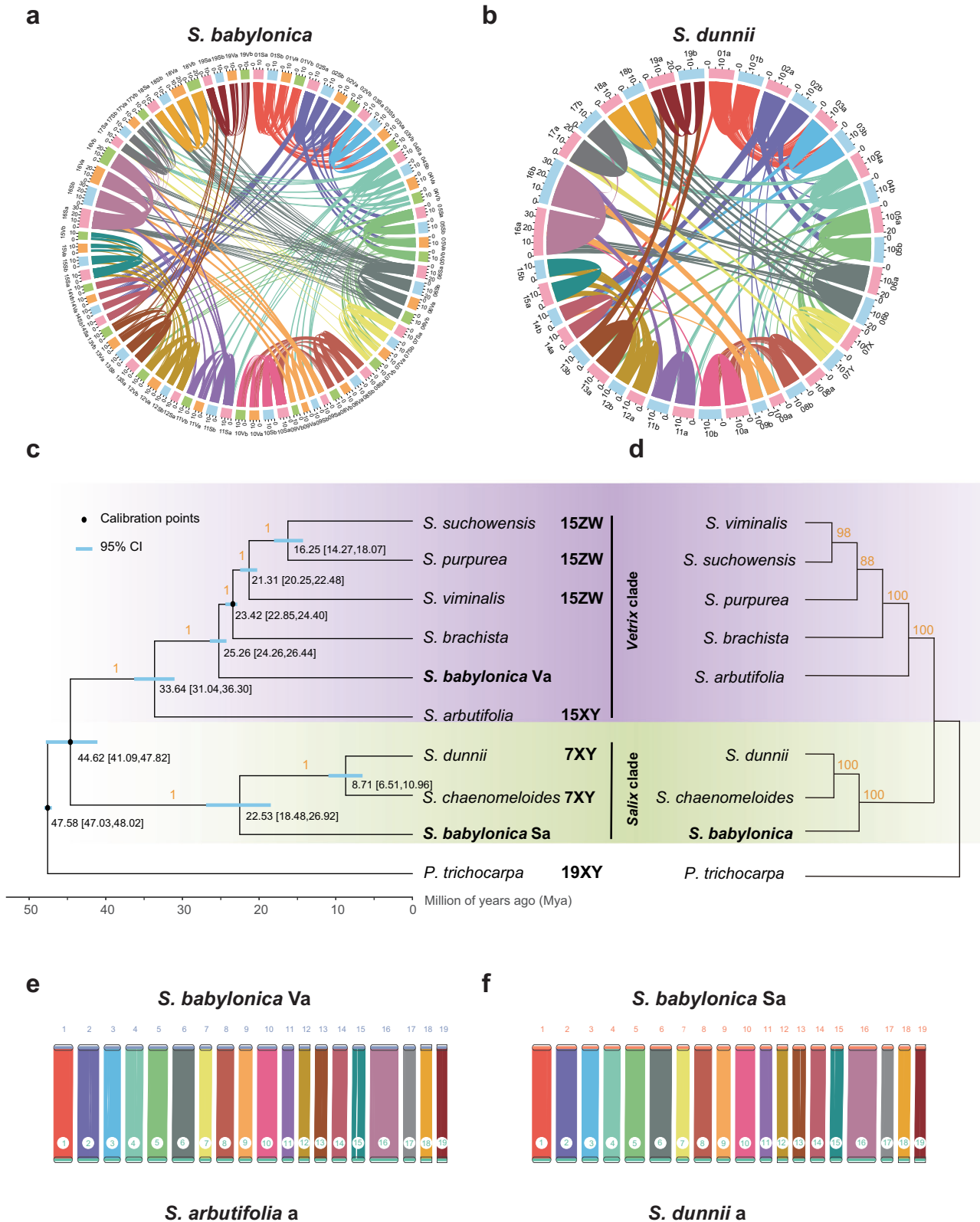


Fig. 1 | Origin and divergence time estimation of *S. babylonica* subgenomes. **a** Synteny of female *S. babylonica* genome. **b** Synteny of male *S. dunzii* genome. **c** Inferred phylogenetic tree and divergence times based on nuclear genome sequences of eight willows and the outgroup *P. trichocarpa*. Blue node bars are 95% confidence intervals, and the black nodes are three fossil calibration points³². Black numbers marked around nodes represent divergence time. Orange numbers marked above branches represent support values. 19XY, 7XY, 15XY and 15ZW on the

right represent different sex-determination systems. **d** Inferred phylogenetic tree based on chloroplast genome sequences of the nine species. Numbers marked on the tree represent bootstrap values (Supplementary Fig. 5). **e** Synteny between *Va* haplotype of *S. babylonica* and *a* haplotype of *S. arbutifolia*. **f** Synteny between *Sa* haplotype of *S. babylonica* and *a* haplotype of *S. dunzii*. Source data are provided as a Source Data file.

the *S. babylonica* genome and 24.78% of the *S. dunnii* genome (Supplementary Data 1). A total of 123,231 genes (116,745 mRNAs excluding alternative splicing, 2428 transfer RNAs (tRNAs), 1528 ribosomal RNAs (rRNAs), and 2530 unclassifiable noncoding RNAs (ncRNAs)) were identified in *S. babylonica* (Supplementary Data 2). We obtained 65,485 genes, (61,872 mRNAs, 1139 tRNAs, 1102 rRNAs, and 1372 ncRNAs) in *S. dunnii* (Supplementary Data 3). The average *S. babylonica* protein-coding gene is 3421.6 bp long, with an average of 5.9 exons, and the average *S. dunnii* protein-coding gene is 3683.4 bp long, with an average of 6.4 exons (Supplementary Table 4). By combining several strategies, we were able to match the vast majority of our predicted genes to predicted proteins in public databases such as GO, KEGG, and Swiss Prot (Supplementary Data 4), and only 1.77% of *S. babylonica* protein-coding genes and 2.24% *S. dunnii* protein-coding genes were not annotated.

Phylogenetic analysis and divergence time estimation

We confirmed that *S. babylonica* is a tetraploid based on flow cytometry and an allotetraploid according to our *k*-mer results (Supplementary Figs. 3, 4). To determine the ancestry of the two subgenomes (S from *Salix*-clade and V from *Vetrix*-clade) within the genus *Salix*, we obtained 5146 single-copy orthologs from our *S. babylonica* *Salix* (*Sa*) and *Vetrix* (*Va*) haplotypes, and a haplotype genome assembly from *S. dunnii*, as well as from six other willows with available genomes, and from the outgroup *Populus trichocarpa* (Supplementary Table 5). The resulting species tree is broadly consistent with previous studies and resolves the *Salix*- and *Vetrix*-clade^{24,25,31,32}.

Importantly, our results show that the V subgenome of *S. babylonica* originated from an ancestor in the *Vetrix*-clade and diverged from the *S. brachista*-*S. viminalis*-*S. purpurea*-*S. suchowensis* clade about 25.26 Mya (million years ago), while the S subgenome originated from an ancestor in the *Salix*-clade, diverging from the *S. dunnii*-*S. chaenomeloides* about 22.53 Mya (Fig. 1c, e). We also estimated the phylogenetic tree from chloroplast genomes, which suggests that the maternal ancestor of *S. babylonica* originated from the *Salix*-clade³³ (Fig. 1d, f and Supplementary Fig. 5).

According to the inferred time of transposable elements divergence (see methods) between the subgenomes of *S. babylonica*, the estimated allotetraploidization time is around 6.2–2.91 Mya (Supplementary Fig. 6), suggesting that *S. babylonica* may have emerged during this period. Since the genus *Salix* contains >400 species³⁴ and *S. babylonica* is nested in a tetraploid group in a plastome tree²¹, we cannot conclude whether the parental species of weeping willow are extant or extinct based on currently available genomes.

Identification of the sex-determination system

We used 4417.1 million clean Illumina reads from 20 females and 20 males of allotetraploid weeping willow (97–135.8 million reads per individual, mean 110.4, Supplementary Data 5), to identify sex-specific *k*-mers, and found a significant enrichment of female-specific *k*-mers on chromosome 15Va, between 5.32–9.71 Mb (Fig. 2a and Supplementary Fig. 7a), consistent with a W chromosome. We did not observe a significant enrichment of male-specific *k*-mers on any chromosome of *S. babylonica*, as expected in a female-heterogametic system (Supplementary Fig. 7b). Our results from the chromosome quotient (CQ) method³⁵ are concordant with the *k*-mer analysis, and we detected the sex-linked region (SLR) between 5.28–9.69 Mb on chromosome 15Va, with mean M:F (male: female) CQ (chromosome quotient) = 0.27, consistent with a W chromosome. The M:F CQ of 2.83–7.88 Mb of chromosome 15Vb had a mean of 2.1, consistent with a Z chromosome. We did not find significant CQ signals or partial *ARR17*-like duplicates on chromosomes 15Sa and 15Sb (Supplementary Fig. 8 and Supplementary Data 6). These results suggest that weeping willow has a female heterogametic system on chromosome 15, 15Va and 15Vb are W (15V_W) and Z (15V_Z) chromosomes, respectively.

Our synteny analysis revealed four inversion events between the 15V_W and 15V_Z chromosomes of *S. babylonica*, and further confirmed a 6.3 Mb W sex-linked region (SLR) between 3.41–9.71 Mb on 15Va, with a corresponding 5.25 Mb Z-SLR between 2.66–7.91 Mb on 15Vb (Fig. 2a and Supplementary Fig. 9). The other regions of the two chromosomes form pseudoautosomal regions (PARs) with 8.74 Mb in 15Va and 8.13 Mb in 15Vb, respectively.

S. dunnii has a male heterogametic system on chromosome 7³¹. The F:M (female: male) CQ analysis revealed a 5.35 Mb region, spanning 6.15–11.5 Mb on 7X, with mean CQ = 2.03. We also observe a 2.35 Mb region spanning 6.25–8.6 Mb, on 7Y with mean CQ = 0.22 (Fig. 2b). We obtained 6,687,746 and 6,705,953 SNPs for haplotype *a* (including 7X) and *b* (including 7Y) of *S. dunnii*, and used these to calculate the F_{ST} values between the 18 male and 20 female genomes. Change-point analyses detected significantly higher F_{ST} values between 5.77–11.64 Mb on chromosome 7X and 5.76–8.71 Mb on chromosome 7Y than in other regions (PARs), which covered the regions identified by CQ. We identified three inversion events between the X and Y in *S. dunnii* using synteny analysis (Fig. 2b and Supplementary Fig. 10), which are all within the region identified as sex-linked by F_{ST} analysis. SLRs of both *S. babylonica* and *S. dunnii* show low gene density and high repetitive sequence density (Fig. 2c, d and Supplementary Data 7).

Evolutionary origin of *Salix babylonica*

The nuclear and chloroplast trees suggest that the female ancestor of *S. babylonica* arose from the *Salix*-clade, while the male ancestor most likely came from the *Vetrix*-clade. The *Vetrix*-clade contains both 15X15X/15X15Y and 15Z15W/15Z15Z species, however phylogenetic analysis of ancestral sex-linked genes within the *Vetrix*-clade reveals that alleles cluster by gametologs, with one clade comprising of *S. arbutifolia* 15X, *S. babylonica* 15W and *S. purpurea* 15W, and another clade comprising of *S. arbutifolia* 15Y, *S. babylonica* 15Z and *S. purpurea* 15Z (Fig. 3a). This result suggests that *S. babylonica* inherited 15X and 15Y, which correspond to 15W and 15Z respectively, or 15W and 15Z from its *Vetrix*-clade ancestor. A diploid male (15Z15Z) in a female heterogametic species can only produce 15Z gametes, and cannot produce 15W, which must come from a female, while a diploid male (15X15Y) in male heterogametic species can produce unreduced gametes with 15X and 15Y. This suggests that the 15XY is the most likely sex chromosome system for the male *S. babylonica* *Vetrix*-clade ancestor. Therefore, this suggests that the ancestral *Vetrix*-clade 15Y (hereafter, 15V_Y) transitioned to 15Z (hereafter, 15V_Z) in *S. babylonica*, and 15X (hereafter, 15V_X) transitioned to 15W (hereafter, 15V_W) (Figs. 1c, d, 2a, 3a).

To better understand the transition from 15V_XV_Y in the *S. babylonica* male *Vetrix*-clade ancestor to the current *S. babylonica* 15V_ZV_W sex chromosome system, we examined the distribution of *ARR17*-like genes in *S. arbutifolia* (15X15X/15X15Y) and *S. babylonica*, which are involved in sex determination in *Salix* and *Populus*^{24,28}. We found *ARR17*-like partial duplicates on 15Y-SLR of *S. arbutifolia* and on 15V_Z-SLR of *S. babylonica* (Supplementary Fig. 11 and Supplementary Data 6), while they are absent in 15X-SLR and 15V_W-SLR of these two species. This further confirmed the transitions (15V_X→15V_W and 15V_Y→15V_Z) in *S. babylonica*. The 15W and 15Z of diploid *S. purpurea* also arose from ancestral *Vetrix*-clade 15X and 15Y, respectively, but its 15W has accumulated intact *ARR17*-like genes²⁵.

We also assessed collinearity between the sex chromosomes (7X and 7Y) of *S. dunnii* and the homologous autosomes (7Sa, 7Sb, 7Va, and 7Vb) of *S. babylonica*. Our results suggest that 7X of the female *Salix*-clade ancestor has reverted to autosomal inheritance in *S. babylonica* (Figs. 1c, d, 3b). There are inversions between the sex chromosomes of *S. dunnii*, but the gene order on autosome 7Sa and 7Sb of *S. babylonica* is most similar to 7X of *S. dunnii*, again consistent with a female *Salix*-

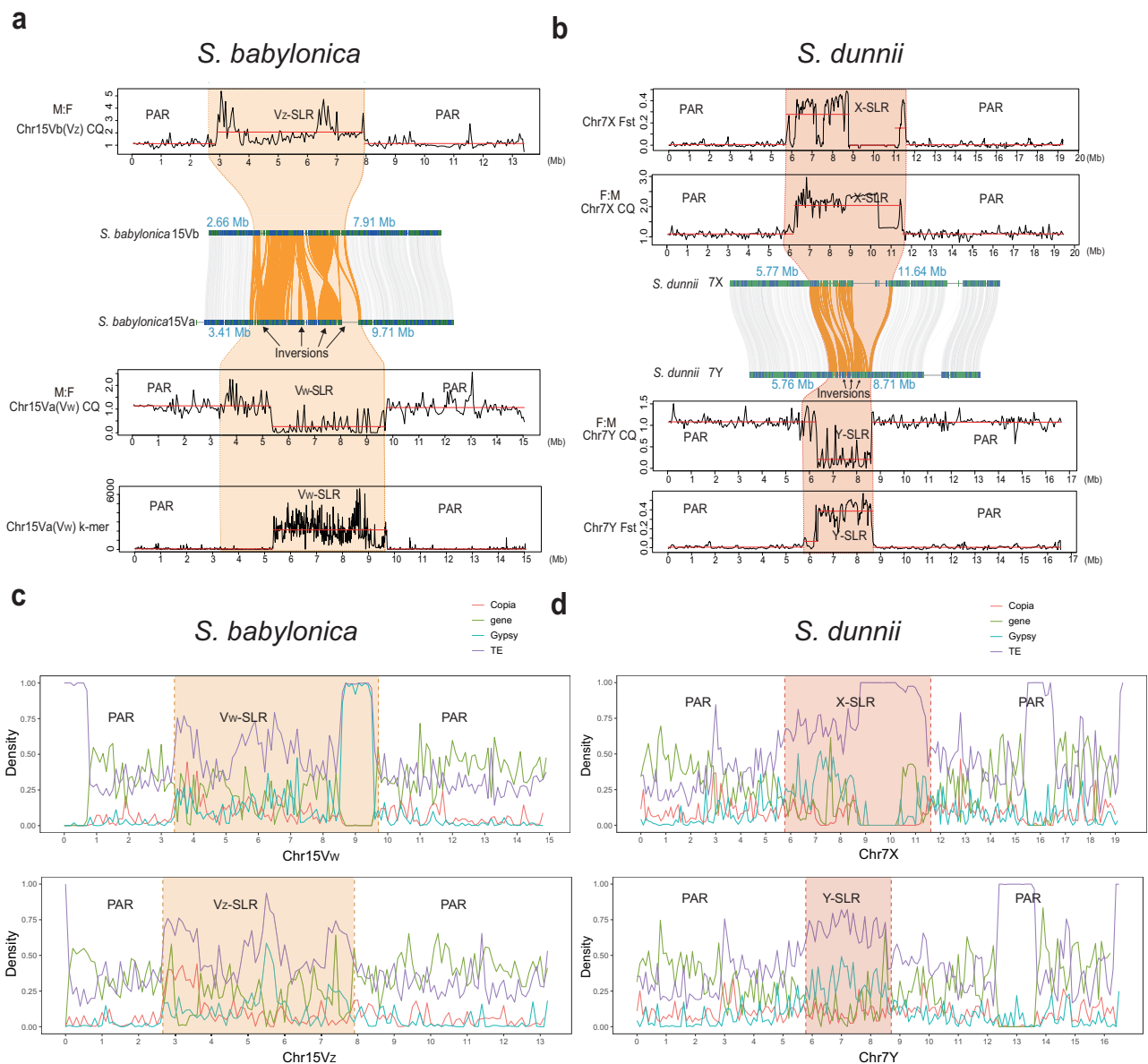


Fig. 2 | Sex-linked regions in *S. babylonica* and *S. dunnii*. **a** CQ (male vs. female alignments), collinearity and female-specific *k*-mer results for *S. babylonica* sex chromosomes 15Vb(Vz) and 15Va(Vw). Supplementary Fig. 7a shows *k*-mer results of the whole genome. The collinearity results between two sex chromosomes show four inversions. **b** Male-female *Fst*, CQ (female vs. male), and collinearity results for

S. dunnii sex chromosomes 7X and 7Y. The collinearity results between two sex chromosomes show three inversions. **c** Gene and transposable elements (TEs) landscape of *S. babylonica* sex chromosomes (100-kb windows). **d** Gene and TE landscape of *S. dunnii* sex chromosomes (100-kb windows). SLR: sex-linked region. PAR: pseudoautosomal region. Source data are provided as a Source Data file.

clade ancestor (Fig. 3b). In addition, the number of genes in the 7Sa&Sb regions syntenic with 7X-SLR is between those in 7X-SLR and 7Va&Vb region (autosome) syntenic with 7X-SLR, supporting that 7Sa&Sb are derived from 7S_X (Supplementary Data 7).

Allopolyploidization from these ancestors could have involved diploid unreduced gametes (Supplementary Fig. 12). Alternatively, crosses between two tetraploid ancestors is also possible, though arguably less likely^{7,36}. The first scenario suggests that the unreduced gametes from the male *Vetrix*-clade ancestor gametes were likely 15X15Y (hereafter, 15V_X15V_Y) (skipping meiosis I), or 15X15X (hereafter, 15V_X15V_X) or 15Y15Y (hereafter, 15V_Y15V_Y) (skipping meiosis II). The unreduced gametes from the female *Salix*-clade ancestor were likely 7X7X (hereafter, 7S_X7S_X) (Figs. 1d, 3b). These female and male gametes can produce 7S_X7S_X15V_X15V_Y (now 7Sa7Sb15V_Z15V_W) females and 7S_X7S_X15V_Y15V_Y (now 7Sa7Sb15V_Z15V_Z) males.

Evolution of the sex-linked region

In *S. babylonica*, gene counts are similar between the V_W-SLR (323 protein-coding genes) and V_Z-SLR (306 protein-coding genes), though their total lengths (V_W-SLR 6.31 Mb vs. V_Z-SLR 5.25 Mb) and repeat lengths (V_W-SLR 3.86 Mb vs. V_Z-SLR 2.96 Mb) differ (Supplementary Data 7). This suggests that sex chromosome turnover or/and polyploidization inhibited substantial heteromorphy of V_Z and V_W⁴. We compared the collinearity and genomic composition in *S. babylonica* (15V_ZV_W), *S. purpurea* (15ZW), and *S. arbutifolia* (15XY) (Fig. 3c, d), and our results indicate that these regions evolved independently in different sex chromosomes despite having the same origin. The *S. babylonica* 15V_Z-SLR, *S. purpurea* 15Z-SLR, and *S. arbutifolia* 15Y-SLR, constitute 39.24, 32.77, and 16.48% of their Z or Y chromosome. The *S. babylonica* 15V_W-SLR, *S. purpurea* 15W-SLR, and *S. arbutifolia* 15X-SLR, constitute 41.95, 42.95, and 43.66% of their W or X chromosome

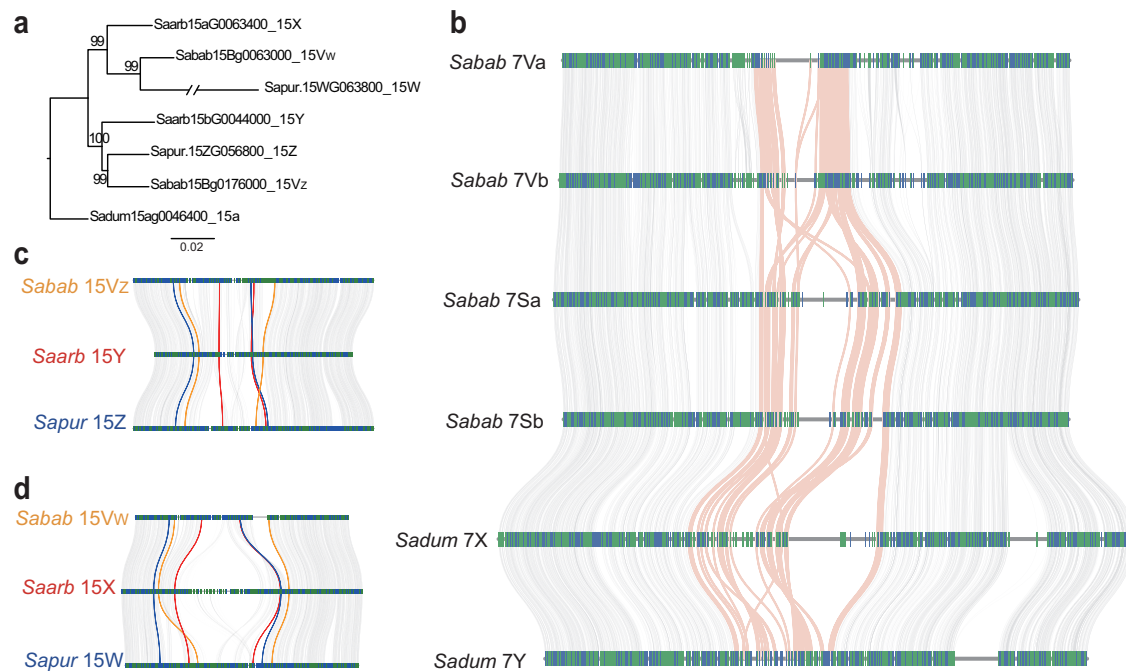


Fig. 3 | Evolutionary relationships, genomic composition, and collinearity of sex chromosomes in different sex-determination systems. **a** The phylogenetic tree is reconstructed from ancestral sex-linked genes proposed by ref. 25. **b** Collinearity and genomic composition of *S. babylonica* 7Va, 7Vb, 7Sa, 7Sb and *S. dunnii* 7X, 7Y. Pink indicates the collinear regions of 7X-SLR, 7Y-SLR, and 7Sa, 7Sb, 7Va, and 7Vb. **c** Collinearity and genomic composition of the *S. babylonica* 15V_Z-SLR, the *S. arbutifolia* 15Y-SLR and the *S. purpurea* 15Z-SLR. Orange, red, and blue

represent the sex-linked regions in *Sabab*, *Saarb*, and *Sapur*, respectively. **d** Collinearity and genomic composition of the *S. babylonica* 15V_W-SLR, the *S. arbutifolia* 15X-SLR, and *S. purpurea* 15W-SLR. Orange, red, and blue represent the sex-linked regions in *Sabab*, *Saarb*, and *Sapur*, respectively. *Sabab*: *S. babylonica*; *Sadum*: *S. dunnii*; *Saarb*: *S. arbutifolia*; *Sapur*: *S. purpurea*. Source data are provided as a Source Data file.

(Supplementary Data 7). Furthermore, we detected different inversions in the sex chromosomes of the three species (Fig. 4 and Supplementary Fig. 9).

In *S. dunnii*, the X-SLR (5.87 Mb) is larger than the Y-SLR (2.95 Mb). Similarly, the *S. arbutifolia* X-SLR is larger (7.06 Mb) than the Y-SLR (1.81 Mb)²⁵. There are 128 and 96 protein-coding genes within X-SLR and Y-SLR in *S. dunnii* (Supplementary Data 3), and the X-SLR contains a total of 2.72 Mb more repeat sequences than the Y-SLR (Supplementary Data 7). We also identified 31 X-SLR-specific genes, of which 13 are tandem duplicates, and 16 Y-SLR-specific genes, of which three are tandem duplicates (Supplementary Data 8). Therefore, the accumulation of repeat sequences and specific tandem duplications may contribute to the longer X-SLR than Y-SLR in *S. dunnii*, which is consistent with *S. arbutifolia*²⁵. This result suggests that the XY sex chromosomes in willows may be somewhat different from other XY plant sex chromosome systems^{37,38}, with a longer Y-SLR compared to the corresponding X-SLR counterpart.

Our analysis reveals *ARR17*-like partial duplicates near or within the inversions between 7X and 7Y of *S. dunnii*, the inversions between 15V_Z and 15V_W of *S. babylonica*, and the inversion between 15X and 15Y of *S. arbutifolia* (Fig. 4 and Supplementary Data 9). It is difficult to determine whether the detected inversions are a catalyst or consequence of recombination suppression³⁹, but these results suggest that *ARR17*-like partial duplicates are associated with sex-linked inversions. This is either because recombination has been selectively suppressed in these regions to maintain sex-specific segregation patterns of the partial *ARR17*-like duplicates, or alternatively, the inversions may have resulted from the fact that these loci are in areas of the pericentromeric region with low recombination (Supplementary Figs. 13, 14).

Sex-determination mechanism in *S. babylonica*

The *ARR17*-like sequences play an important role in the sex determination of Salicaceae. The partial *ARR17*-like duplicates on Y-SLRs of diploid *S. chaenomeloides* (7XY) and *S. arbutifolia* (15XY) can produce sRNA to silence four intact *ARR17*-like genes on chromosome 19 to determine maleness^{24,25}. *S. dunnii* likely shows a similar strategy. Therefore, we used 1181.68 million clean RNA-Seq Illumina reads (56.42–77.76 million reads per individual, mean 65.65) to estimate the expression of intact *ARR17*-like and *PI*-like genes, and 43.24 million sRNA reads (1.70–3.52 million reads per individual, mean 2.40) surrounding the partial *ARR17*-like duplicates of weeping willow flower buds from three stages. We identified more sRNAs in the partial *ARR17*-like duplicate (exon1) region and its upstream region in male flower buds than in female flower buds of weeping willow (Fig. 5a, stage 2 and Supplementary Fig. 15, stage 1 and 3). Consistent with this, we found that all of exon1 of intact *ARR17*-like genes were expressed in female flower buds, with significantly lower or no expression in male flower buds of weeping willow (Fig. 5b–e). We also detected the accumulation of sRNAs in intact *ARR17*-like gene regions on chromosomes 19Va, 19Vb, 19Sa, and 19Sb in males (Supplementary Fig. 16). Furthermore, the *PI*-like genes were only expressed in male flower buds (Supplementary Fig. 17). This result suggests that sRNAs produced by partial *ARR17*-like duplicates of two 15V_{ZS} alleles may inhibit eight intact *ARR17* genes on the four homologs of chromosome 19, activating the *PI*-like genes and leading to the production of male individuals (Fig. 6). Our results suggest a dose effect related to sex determination, with one 15V_Z carrying partial *ARR17*-like duplicates in females insufficient to inhibit intact *ARR17*-like genes, similar to the sex-determining mechanism of chicken⁴⁰. *S. babylonica* 15V_W could also acquire a female factor to inhibit sRNA production of partial *ARR17*-like duplicates, but more assemblies of allotetraploid willows are needed to detect whether such factor(s) exist.

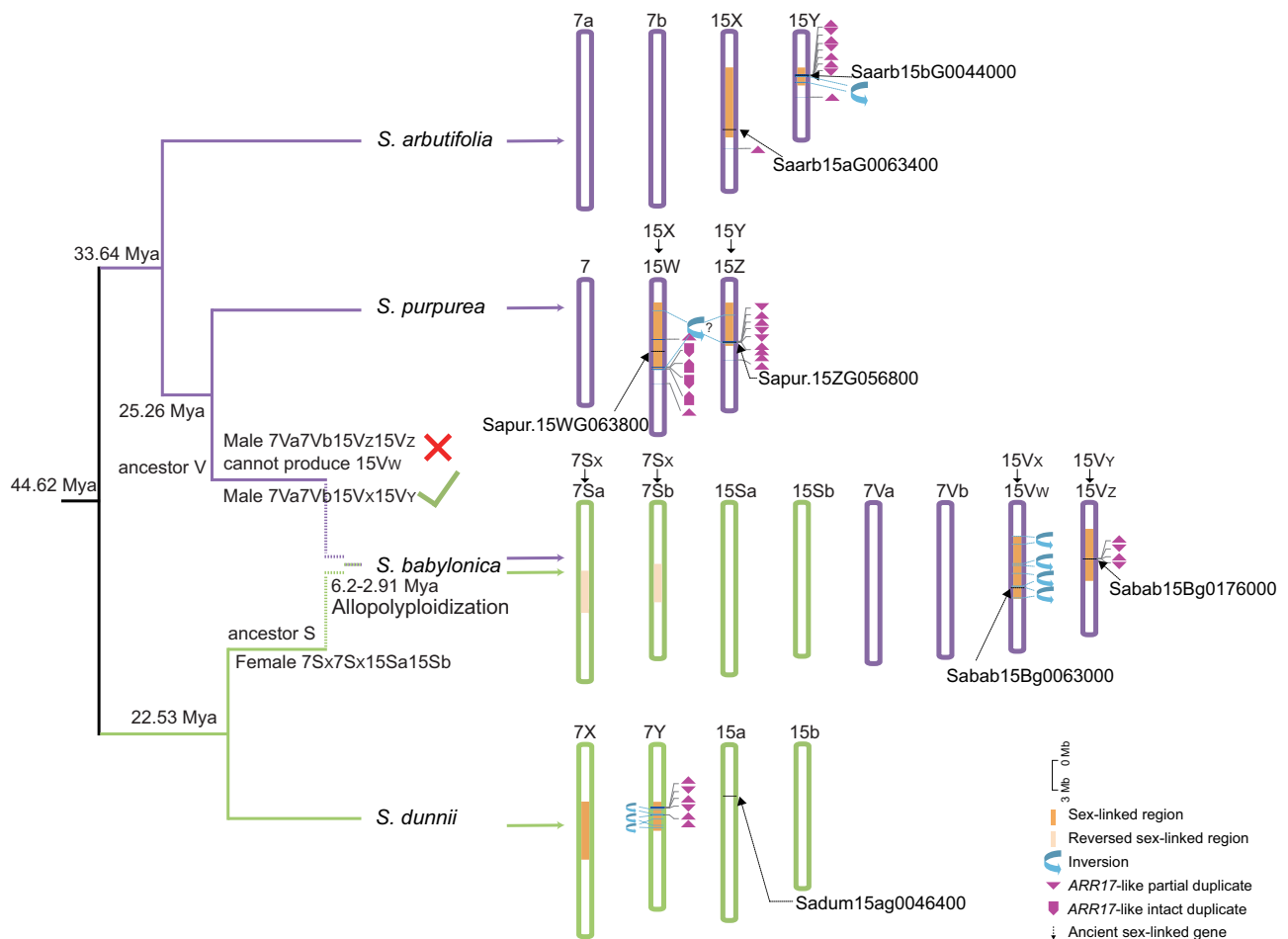


Fig. 4 | Hypothetical origination of *S. babylonica* and evolution of sex-linked regions (SLRs) in relevant diploid willows. Purple represents genomic components of *S. babylonica* the *Vetrix*-clade ancestor, that likely has 15XY sex chromosomes, while green represents components from the *Salix*-clade ancestor with 7XY sex chromosomes. The chromosomes on the right use the scaled real physical

length, and dark yellow or light yellow represents sex-linked or reversed sex-linked regions. Purple triangles indicate partial *ARR17*-like duplicates (the tip is the end of the duplicate), and purple arrows indicate intact *ARR17*-like duplicates (the tip is the end of the gene). Blue arrows indicate inversions on SLRs. Genes marked by the black arrows are ancestral sex-linked genes that were used in Fig. 3a.

Dosage compensation in different sex chromosomes

We used 854.73 million clean Illumina reads of weeping willow catkins (69.86–73.08 million reads per individual, mean 71.23), 410.10 million clean Illumina reads of *S. arbutifolia* catkins (33.86–34.52 million reads per individual, mean 34.17), and 380.73 million clean Illumina reads of *S. dunnii* catkins (30.44–32.89 million reads per individual, mean 31.73) to observe the expression level between SLRs and autosomes. We observed degeneration signals in the SLRs of the 15V_Z and 15V_W of *S. babylonica*, the 15X and 15Y of *S. arbutifolia*, and the 7X and 7Y of *S. dunnii*. Gene density was lower in all these SLRs compared to the corresponding PARs, while repetitive elements density was higher (Fig. 2c, d and Supplementary Data 7). The W-SLR and Y-SLRs have lost genes compared with their orthologous autosomes, i.e., the *S. babylonica* 15V_W-SLR lost 2.84% genes, the *S. arbutifolia* 15Y-SLR lost 11.49% genes, and the *S. dunnii* 7Y-SLR lost 1.22% genes (Supplementary Table 6). We also found evidence of gene loss from Z-SLRs and X-SLRs. The *S. babylonica* 15V_Z-SLR lost 3.9% genes, the *S. arbutifolia* 15X-SLR lost 9.20% genes, and the *S. dunnii* 7X-SLR lost 1.22% genes. The V_Z-SLR lost more genes than the V_W-SLR in *S. babylonica*, which may be due to the turnover (Y→Z) mentioned above.

Overall, the expression values between sexes are similar but significantly lower (Wilcoxon test *: $p \leq 0.05$) than in autosomes among 15ZW (male 15Z-SLR15Z-SLR vs female 15Z-SLR), 15XY (female 15X-SLR15X-SLR vs male 15X-SLR), and 7XY (female 7X-SLR7X-SLR vs male

7X-SLR) (Fig. 7). Based on these results, we revealed similar incomplete dosage compensation patterns in *S. babylonica*, *S. arbutifolia*, and *S. dunnii*.

Discussion

Although sex chromosomes were once thought to be an obstacle to polyploidization³, it is increasingly clear that polyploidy can arise from diploid ancestors with sex chromosomes in both plants and animals^{4,9,41}. How these polyploid lineages overcome the complex and unbalanced allelic combinations of sex determination alleles to establish stable dioecy is not yet known. Previous studies did not provide direct evidence that dieocious plants overcome the polyploidy limit^{14,42,43}. *Diospyros kaki* and *Mercurialis annua*, both male heterogametic systems, reverted to non-dieocious sex-determination systems after genome doubling^{14,42}, while the diploid and polyploid ancestors of the octoploid dieocious *Fragaria chiloensis* were likely hermaphroditic⁴³. For other polyploids with sex chromosomes, such as *Rumex acetosella*, a male heterogametic tetraploid, genomic resources are not yet available to investigate sex chromosome evolution and its ancestral species¹⁷.

We assembled haplotype-resolved genomes of female allotetraploid *S. babylonica* and male diploid *S. dunnii*. Our nuclear and chloroplast genome sequences trees indicate that *S. babylonica* has a female ancestor from the *Salix*-clade and a male ancestor from the

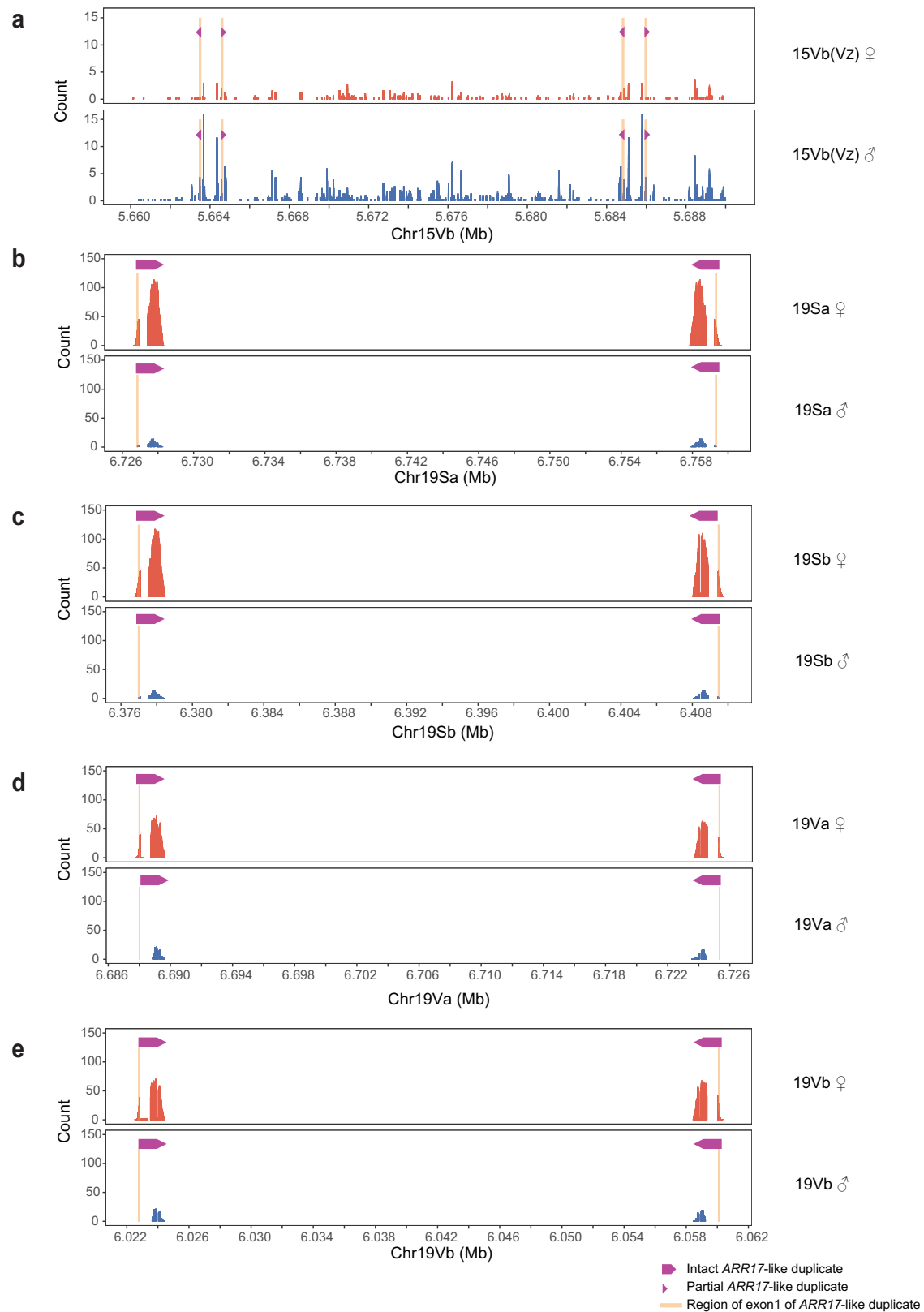


Fig. 5 | Expression patterns of *ARR17*-like duplicates in female and male flower buds of weeping willow (stage 2). **a** Transcript level of partial *ARR17*-like duplicate sRNA and surrounding regions on sex chromosome 15V_Z. The orange area indicates exon1 of *ARR17*-like copies. **b–e** Transcript level of intact *ARR17*-like genes on

Chr19Sa, Chr19Sb, Chr19Va and Chr19Vb. Purple triangles indicate partial *ARR17*-like duplicates (the tip is the end of the duplicate). The purple arrows show the intact *ARR17*-like duplicates (the tip is the end of the gene). Source data are provided as a Source Data file.

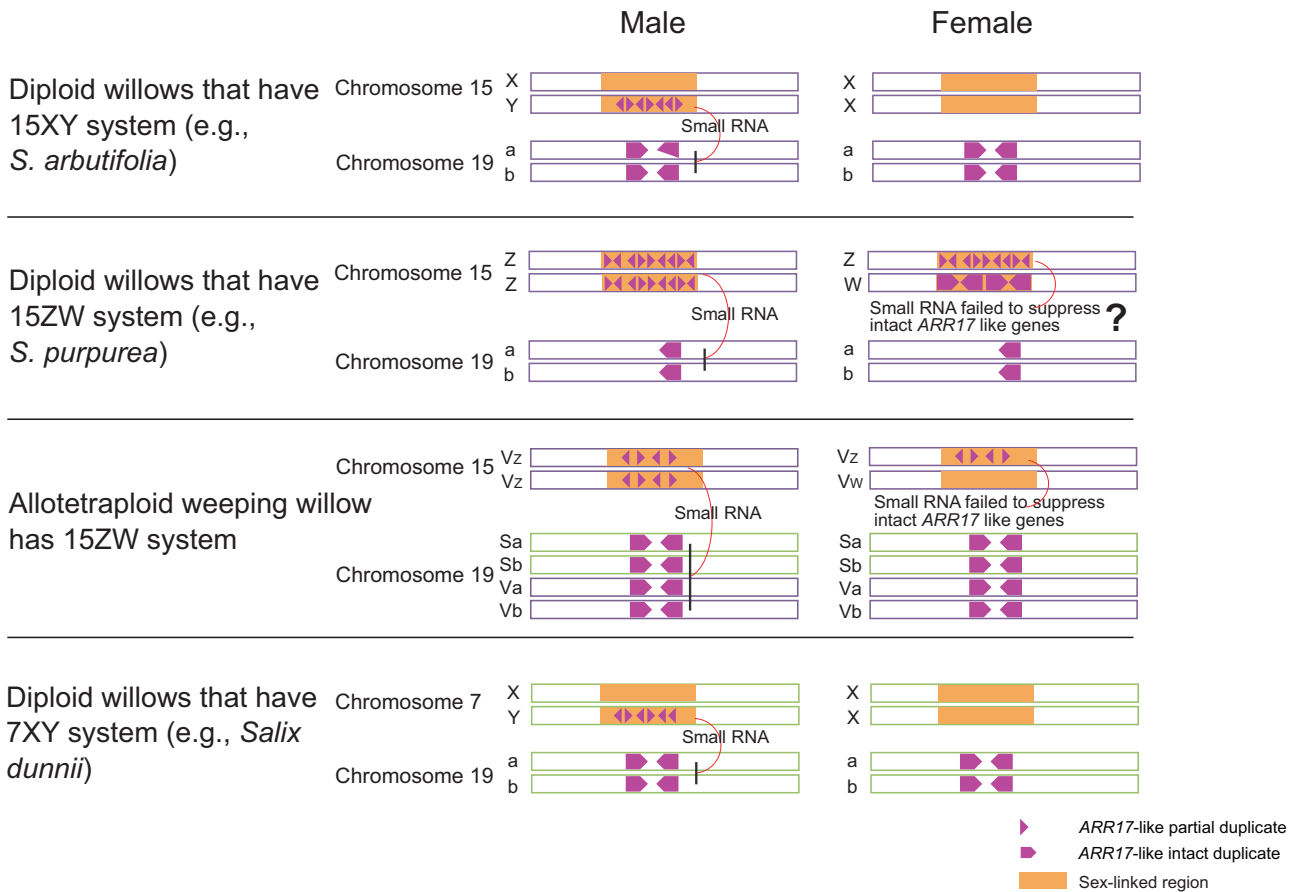


Fig. 6 | Hypothetical model for *ARR17*-like duplicates and sex determination in willows. One Z/Y sex chromosome carrying partial *ARR17*-like duplicates can produce sRNA to inhibit the expression of four intact *ARR17*-like genes, but six or eight intact *ARR17*-like genes can overcome the influence of one Z. Purple triangles within

sex-linked regions indicate partial *ARR17*-like duplicates (the tip is the end of the duplicate). Purple arrows show intact *ARR17*-like duplicates (the tip is the end of the gene). The relevant data can be found in the Supplementary Data 6.

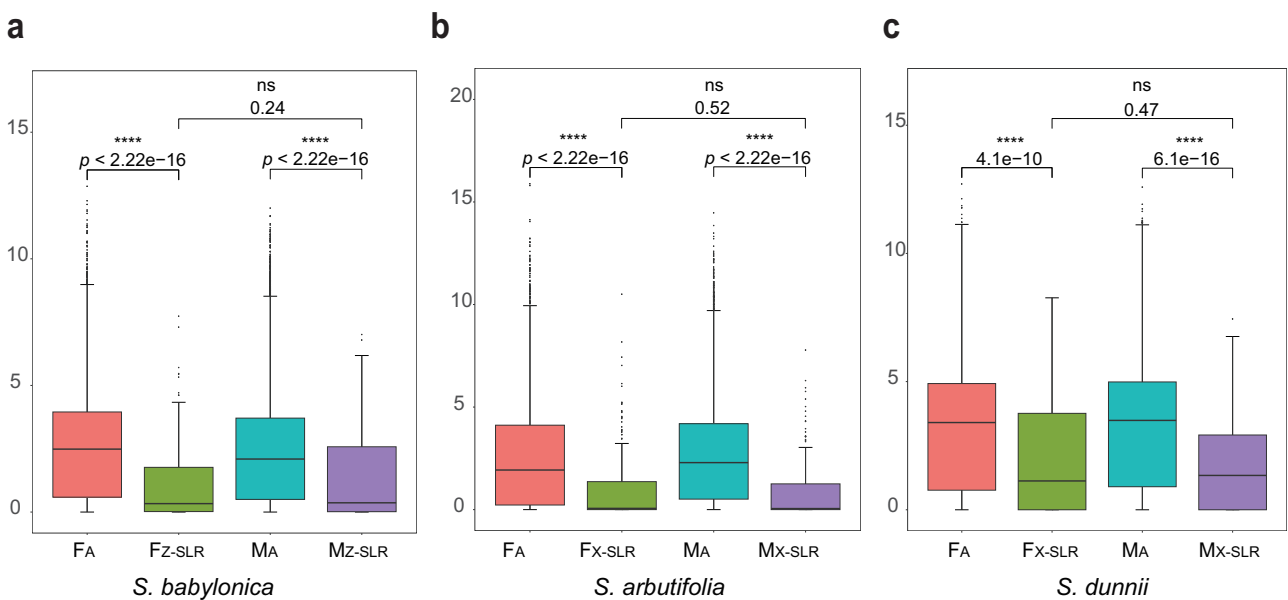


Fig. 7 | Dosage compensation pattern in different sex-linked regions of willows. The expression of SLR and autosomal genes in **a** *S. babylonica* (gene number: F_A (56063), F_{Z-SLR} (306), M_A (56063), M_{Z-SLR} (306)), **b** *S. arbutifolia* (gene number: F_A (28097), F_{X-SLR} (238), M_A (28097), M_{X-SLR} (238)), and **c** *S. dunnii* (gene number: F_A (29623), F_{X-SLR} (128), M_A (29623), M_{X-SLR} (128)). F and M represent females and males. Each sex contained three independent biological replicates. A and -SLR

represents autosome and sex-linked region. Significance based on two-sided Wilcoxon test (ns: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$). Box edges indicate upper and lower quartiles, centerlines indicate median values, and whiskers extend to 1.5 times the interquartile range. Source data are provided as a Source Data file.

Vetrix-clade within the *Salix* genus and that allopolyploidization occurred between 6.2–2.91 Mya, long before plant domestication⁴⁴. Our results also suggested that sex chromosome turnover (15XY to 15ZW) happened during allopolyploidization.

Similar to its male ancestor, the *S. babylonica* SLRs are located on chromosome 15Va(15V_w) and 15Vb(15V_z) according to *k*-mer and CQ analysis (Fig. 2a and Supplementary Figs. 7, 8), indicating that only one pair of sex chromosomes was retained⁶. According to the ancestral sex-linked gene tree, the X-Y homologous gene pair diverged in the ancestor of the *Vetrix*-clade²⁵, and the genomic distribution and phylogenetic analysis of partial *ARR17*-like duplicates (Supplementary Fig. 11), the 15V_w(15Va) and 15V_z(15Vb) derived from the ancestral 15V_x and 15V_y chromosomes in the male ancestor, respectively (Figs. 1c, 3a, 4 and Supplementary Figs. 11, 12). The 15W and 15Z chromosomes of diploid *S. purpurea* were also derived from ancestral 15X and 15Y chromosomes of the *Vetrix*-clade²⁵, respectively. These results suggest that XY to ZW transitions occurred independently in different willow species.

Early generations of allotetraploid *S. babylonica* were likely comprised of 7S_x7S_x15V_x15V_x and 7S_x7S_x15V_x15V_y females (see Sex-determination mechanism in *S. babylonica*), and 7S_x7S_x 15V_y15V_y males. Initially, the two 15V_ys in males could potentially undergo homologous recombination, but would only be present together when the maternal gamete contained 15V_y, which can only be produced by a female-fertile 7S_x7S_x15V_x15V_y genotype (Supplementary Fig. 12). Furthermore, the 7S_x7S_x15V_x15V_x genotype can only produce 7S_x15V_x gametes, which when combined with a male gamete carrying 7S_x15V_y, would produce 7S_x7S_x15V_x15V_y progeny. In this process, recombination was reduced for the 15V_x chromosome in females, and ultimately 15V_x transitioned to 15V_w. Our evidence suggests that the 7S_x chromosomes from the female ancestor likely reverted to autosomal inheritance, as evidenced by the similar gene order shared between the resulting autosome and the 7X chromosome of *S. dunnii*, genomic composition of 7Sa, 7Sb, 7Va, 7Vb, 7X, and 7Y, and the phylogenetic trees (Figs. 1c, 3b and Supplementary Data 7).

The *ARR17*-like gene (partial and intact) acts as a sex-determination factor in some members of the Salicaceae^{24,28}, and is well known from poplars (*Populus*), the sister genus of *Salix*. In the male-specific region of the Y chromosome, partial *ARR17*-like duplicates ortholog produce sRNAs and silence the intact *ARR17*-like gene on poplar sex chromosome 19, thereby determining maleness^{28,29}.

Partial *ARR17*-like duplicates have been found in Y-linked regions in the genus *Salix* (7Y in the *Salix*-clade and 15Y in the *Vetrix*-clade), which can produce sRNAs to suppress the expression of intact *ARR17*-like genes on chromosome 19^{23,25} (Fig. 6 and Supplementary Data 6), determining maleness. We observe partial *ARR17*-like duplicates near or within inversion boundaries in the *S. babylonica* and *S. dunnii* sex chromosomes, suggesting that these loci act as sex-determining factors in these species as well. The partial *ARR17*-like duplicates on the two 15V_z-SLRs of male *S. babylonica* can produce more sRNA than that on one 15V_z-SLR in females due to dose effects, and likely silenced all the eight intact *ARR17*-like genes on the four homologs of chromosome 19, thereby determining maleness (Fig. 5 and Supplementary Fig. 15). The eight intact *ARR17*-like genes on chromosome 19 in *S. babylonica* likely overcame the influence of partial *ARR17*-like duplicates of one 15V_z-SLR and indirectly suppressed the expression of *PI*-like genes in females to maintain dioecy (Fig. 5 and Supplementary Fig. 17). In diploid *S. purpurea* (15ZW), the four intact *ARR17*-like genes on 15W-SLR and two intact copies on chromosome 19 can determine femaleness like in the weeping willow²⁴ (Fig. 6 and Supplementary Data 6). Although it is unclear whether the inversions are the cause of recombination suppression between sex chromosomes or are a consequence of it³⁹, the association with *ARR17*-like partial duplicates suggests selection to suppress recombination in these regions to maintain sex-specific segregation patterns.

Gene loss from the SLRs of sex chromosomes is common in plants⁴⁵. Although dioecy is ancestral in willows, and originated at least 47 Mya, sex chromosomes in species within the *Vetrix*-clade, including *S. viminalis* (15ZW)⁴⁶ and *S. purpurea* (15ZW)⁴⁷ show low levels of divergence. The results indicate that the 15Y-SLR of *S. arbutifolia* lost more genes compared with the 15X-SLR (Supplementary Table 6). This signature was also found in other plant sex chromosome systems, including papaya⁴⁸, *Silene latifolia*⁴⁵, and *Rumex hastatulus*⁴⁹, and is also a general feature of animal sex chromosomes⁵⁰. Our results suggest that the 15V_w in *S. babylonica* originated from 15V_x, and 15V_z originated from 15V_y. It is likely that the gene loss from the 15V_z-SLR in *S. babylonica* represents the ancestral loss of genes from *Vetrix* 15V_y-SLR.

The loss of genes from W-SLRs and Y-SLRs occurs in many systems following recombination suppression (reviewed by refs. 51–53), and this results in gene dose differences between males and females. Dosage compensation mechanisms have evolved in some species that increase expression levels in the heterogametic sex, potentially restoring expression to levels to the ancestral chromosome pair, and making them equal in both sexes^{54–57}. However, polyploidization complicates dose differences associated with Y and W chromosome degeneration, and it is difficult to predict how polyploidization might affect dosage compensation dynamics. We observe similar patterns of incomplete dosage compensation in *S. babylonica* (15V_zV_w), *S. dunnii* (7XY), and *S. arbutifolia* (15XY) (Fig. 7), similar to many other plant sex chromosome systems⁵⁸.

Methods

Plant material

We collected young leaves from a male *Salix dunnii* (FNU-M-1) and a female *S. babylonica* (saba01F) plant for genome sequencing. Young leaf, catkin, stem, and root samples for transcriptome sequencing were collected from FNU-M-1 and saba01F, as well as catkins from two other female and three male *S. babylonica* plants, and one other female *S. dunnii* plant. In October 2023 (stage 1), December 2023 (stage 2), and February 2024 (stage 3), we collected flower buds from three male and three female individuals of *S. babylonica* for sRNA sequencing and mRNA sequencing to detect possible sex-determination mechanisms. Samples were frozen in liquid nitrogen and stored at –80 °C until total genomic DNA or RNA extraction. We sampled 40 flowering individuals of *S. babylonica* and dried their leaves in silica-gel for resequencing. Voucher specimens collected for this study are deposited in the herbarium of Shanghai Chenshan Botanical Garden (CSH). Supplementary Data 10 gives detailed information on all the samples. We also downloaded genome sequence data of 38 individuals of *S. dunnii* published by ref. 31, the genome of *S. arbutifolia* (with gap-free 15X and 15Y sex chromosomes)²⁵, the genome of *S. purpurea* (with phased 15Z-SLR and 15W-SLR)⁴⁷, and other available willows genomes and *Populus trichocarpa* for relevant analyses (Supplementary Table 5).

Ploidy determination

The ploidy of saba01F was measured by flow cytometry, using diploid *Salix dunnii* (2x = 2n = 38)³¹ as an external standard. The assay followed the protocol of ref. 59. The leaf tissue was incubated for 80 min in 1 mL LB01 buffer and chopped with a razor blade. The homogenate was then filtered through a 38-μm nylon mesh and treated with 80 μg/mL propidium iodide (PI) and 80 μg/mL RNase followed by 30 min incubation on ice to stain the nuclei. DNA content measurements were done in a MoFlo-XDP flow cytometer and evaluated using Summit v.5.2 (Beckman Coulter Inc.). The ploidy level was calculated as:

$$\text{sample ploidy} = \text{reference ploidy} \times \frac{\text{mean position of the sample peak}}{\text{mean position of reference peak}} \quad (1)$$

Determination of allo- or autotetraploid origin in weeping willow

We used jellyfish version 2.3.0⁶⁰ to construct *k*-mer frequency distributions of saba01F based on PCR-free Illumina short reads, with *k*-mer length set to 21. Genomescope 2.0⁶¹ was run with 21-mer to distinguish between autotetraploid and allotetraploid origin based on the patterns of nucleotide heterozygosity. Allotetraploids are expected to have a higher proportion of aabb than aaab, while autotetraploids have a higher proportion of aaab^{8,61}.

Genome sequencing

For Illumina PCR-free sequencing of FNU-M-1 and saba01F, and whole-genome sequencing (WGS) of all 40 weeping willow individuals, their total genomic DNAs were extracted using the Qiagen DNeasy Plant Mini kit following the manufacturer's instructions (Qiagen). PCR-free sequencing libraries were generated using the Illumina TruSeq DNA PCR-Free Library Preparation Kit (Illumina) following the manufacturer's recommendations. Paired-end libraries were constructed for all 40 samples for WGS. These libraries were sequenced on an Illumina platform (NovaSeq 6000) by Beijing Novogene Bioinformatics Technology (hereafter Novogene).

The Hi-C library was prepared following standard procedures⁶². In brief, the leaves from FNU-M-1 and saba01F were fixed with a 4% formaldehyde solution. Subsequently, cross-linked DNA was isolated from nuclei. The restriction enzyme MboI was then used to digest the DNA, and the digested fragments were labeled with biotin, purified, and ligated before sequencing. Hi-C libraries were controlled for quality and sequenced on NovaSeq 6000 by Novogene.

For PacBio HiFi libraries and sequencing, total genomic DNA was extracted using the CTAB method. PacBio large insert libraries were prepared with SMRTbell Express Template Prep Kit 2.0. These libraries were sequenced by Novogene using the PacBio Sequel II platform.

RNA extraction and library preparation

Total mRNA was extracted from young leaves, female and male catkins, stems and roots of *S. dunnii* and weeping willow, and female and male buds (three stages, see Plant material) of weeping willow using the CTAB method. RNA integrity was assessed using the Fragment Analyzer 5400 (Agilent Technologies, CA, USA). Libraries were generated using NEBNext[®] UltraTM RNA Library Prep Kit for Illumina[®] (NEB, USA) following the manufacturer's recommendations, and sequencing was performed on an Illumina Novaseq 6000 by Novogene.

For sRNA sequencing, total RNA from flower buds (three stages) of *S. babylonica* was extracted using RNAPrep Pure Plant Plus Kit (Tiangen, China). Adapters were ligated to the ends of the sRNA, and then the first strand of cDNA was synthesized after hybridization with reverse transcription primers. PCR enrichment was used to generate the double-stranded cDNA library. Libraries with insertions of 18–40 bp were ready for sequencing after purification and size selection. sRNA sequencing was performed on an Illumina Novaseq 6000 by Novogene.

Genome assembly

Hifiasm⁶³ was used to assemble initial genomes based on PacBio HiFi reads, with the resulting haplotype assemblies used for subsequent analysis. For chromosome-level genome assembly, the Hi-C reads were first aligned to the haplotype contig genome using Juice⁶⁴. Subsequently, a preliminary Hi-C-assisted chromosome assembly was carried out using 3d-dna⁶⁵. This was followed by manual inspection and adjustment using Juicebox⁶⁶, primarily focused on refining chromosome boundaries, removing incorrect insertions, adjusting orientations, and rectifying assembly errors. To optimize the assembly further, gap filling based on HiFi reads was performed using the LR_Gapcloser⁶⁷. Additionally, a two-round polishing approach using

Nextpolish⁶⁸ was employed for genome base correction utilizing our short-read data.

We obtained chromosome-scale, haplotype-resolved genome assemblies of *S. babylonica* and *S. dunnii*. Although the majority of chromosomes exhibited well-assembled telomeric ends containing the characteristic telomere sequence (TTTAGGG)_n, there were a few cases where this sequence was either short or missing. Assuming incomplete assembly or insufficient extension, the HiFi reads were remapped to the genome. Reads aligning near the telomeres were selected, and contig assembly was performed using hifiasm⁶³. The resulting contigs were then aligned back to the chromosomes, allowing for the extension of the chromosomes towards the outer ends to enhance the completeness of the telomere sequence assembly. For the *S. babylonica* genome assembly, the chromosomes were compiled as chromosome 01 to chromosome 19 (Sa/Aa, Sb/Ab, Va/Ba, and Vb/Bb) according to the homologous relationship with *S. dunnii* (*S. Salix*-clade) and *S. brachista* (*V. Vetrrix*-clade). The chromosome number of our male *S. dunnii* is consistent with that of the female *S. dunnii* genome reported previously³¹.

The chloroplast and mitochondrial genomes were assembled using GetOrganelle⁶⁹. Afterward, fragmented contigs were mapped to the chromosome-level genome and organelle genome sequences by Redundans⁷⁰, enabling the identification of redundant segments. Furthermore, low-coverage fragments or haplotigs within the scattered sequences and rDNA fragments were discarded.

Genome annotation

Homology-based prediction, transcript prediction, and de novo prediction approaches were used for genome annotation. For homology-based prediction, we used the publicly available Salicaceae protein sequences (Supplementary Table 5) as homologous protein evidence for gene annotation. For transcript prediction, we used three strategies to assemble transcripts. In addition to using Trinity⁷¹ for de novo assembly directly, the reads were also aligned to genomes with HISAT2⁷² before assembly with the Trinity genome-guided model and StingTie⁷³. Then we combined all the transcript sequences and removed redundancy using CD-HIT⁷⁴ (identity >95%, coverage >95%). Based on transcript evidence, we used the PASA pipeline⁷⁵ to annotate gene structure, and aligned to the reference protein (Supplementary Table 5) to identify full-length genes. These genes were used for AUGUSTUS⁷⁶ training, and we performed five rounds of optimization.

MAKER2⁷⁷ was used for genome annotation based on ab initio prediction, transcript and homolog protein evidence. After masking repeat sequences with RepeatMasker, AUGUSTUS⁷⁸ was used for ab initio predict coding genes. We then aligned transcript and protein sequences to the genomes using BLASTN and TBLASTX. After optimizing the comparison results using Exonerate⁷⁹, we integrated and predicted gene models with AUGUSTUS⁷⁸, and expression sequence tag (EST). EvidenceModeler (EVM) gene structure annotation tool⁸⁰ was further used to integrate MAKER and PASA annotation results. In addition, we performed TEsorter⁸¹ and EVM to identify and mask TE protein domains. Then, PASA⁷⁵ was used to upgrade the EVM annotation, adding untranslated regions (UTRs) and alternative splicing. Finally, missense (internal stop codon or ambiguous base, no start codon or stop codon) annotations, and genes <50 amino acids were removed.

For noncoding RNA (ncRNA) annotation, tRNAs were annotated using tRNAScan-SE⁸², and RfamScan was used to align and annotate various ncRNAs. We also performed Barrnap (<https://github.com/tseemann/barrnap>) to remove partial results. Finally, we removed redundancies and integrated all the annotation results.

The functions of protein-coding genes were annotated based on three strategies: (1) Gene functions were identified using eggno-mapper⁸³, aligning with homologous gene databases; (2) We used DIAMOND to perform sequence similarity searches, and compared

protein sequences with protein databases, including Swiss_Prot, TrEMBL, and NR (identity >30%, E-value <1e-5); (3) InterProScan⁸⁴ was used to search domain similarity. We compared sequences with the PRINTS, Pfam, SMART, PANTHER, and CDD databases and obtained amino acid-conserved sequences, motifs, and domains.

We used EDTA⁸⁵ to identify transposable elements (–sensitive 1, –anno 1), producing a TE library. Then, RepeatMasker (<http://www.repeatmasker.org/RepeatMasker/>) was used to determine repetitive regions within our genome assemblies.

Phylogenetic analysis and divergence time estimation

We performed a phylogenetic analysis of the eight willow genomes (*S. suchowensis*, *S. purpurea*, *S. viminalis*, *S. brachista*, *S. babylonica* (haplotypes *Sa* and *Va*), *S. arbutifolia* (haplotype *a*), *S. dunnii* (haplotype *a*), *S. chaenomeloides*), and used *P. trichocarpa* as the outgroup (Supplementary Table 5). After obtaining the longest mRNA among them, OrthoFinder⁸⁶ was used to identify single-copy orthologous genes. To avoid the effect of sex chromosomes, we removed genes from chromosomes 7, 15, and 19. Then, these protein sequences were aligned using MAFFT⁸⁷. We constructed gene trees with IQ-TREE (–m MFP –bb 1000 –bnni –redo)⁸⁸. Then, ASTRAL was used to infer a species tree based on gene trees' results⁸⁹. Finally, MCMCTREE in PAML⁹⁰ was used to estimate the divergent time (burnin = 400,000, sampfreq = 10, nsample = 100,000) based on two fossils and one inferred time at three nodes: (1) the root node of Salicaceae (48 Mya), (2) the divergence time between *Salix*- and *Vetrix*-clade (37.15 Mya to 48.42 Mya), and (3) the ingroup of *Vetrix* (*Chamaetia*-*Vetrix*)-clade (23 Mya)³². The chloroplast genomes of the nine species were obtained from available assemblies or assembled using GetOrganelle⁶⁹ based on whole-genome sequencing reads (Supplementary Table 5). These chloroplast genome sequences were aligned by HomBlocks.pl with default parameters⁹¹, and then IQ-TREE was used to construct a chloroplast phylogenetic tree.

In order to determine the allotetraploidization time of *S. babylonica*, we calculated TE (transposable element) divergence values from the two subgenomes because TE substitution rates of the subgenomes of allotetraploids differ before and after polyploid formation⁹². Among the four haplotypes of *S. babylonica*, we used *Sa* vs. *Va* and *Sa* vs. *Vb* to perform this analysis, removing chromosomes 7 and 15. First, we used Nucmer⁹³ to identify the matching region between two subgenomes of *S. babylonica* (alignment length >1000, alignment identity >90), then RepeatModeler (<http://www.repeatmasker.org/>) was used to create repeat sequences database based on above matching sequences (RMblast), and RepeatMasker was used to identify TEs in subgenomes. Subsequently, the substitutions of these identified TEs were calculated with calcDivergenceFromAlign.pl (a script in RepeatMasker). Divergence was obtained by comparing TEs between *S. babylonica* and the repeat sequences database obtained above. Finally, the ggplot2⁹⁴ R package was used to print divergence values using GAM (Generalized Additive Model) to fit the curves. Two time points of divergence and merger between two subgenomes corresponded to TE divergence rates of 18% and 2.5% in *Sa* vs. *Va*, as well as 23% and 1.5% in *Sa* vs. *Vb* (Supplementary Fig. 6). The divergence time of *Salix*- and *Vetrix*-clade was 44.62 Mya according to the result of phylogenetic analysis above. Therefore, the allotetraploidization time was estimated: 44.62 Mya × 2.5/18 ≈ 6.2 Mya and 44.62 Mya × 1.5/23 ≈ 2.91 Mya.

Identification of the sex-linked regions of weeping willow and *Salix dunnii*

Sequence reads of weeping willow and *S. dunnii* were filtered and trimmed by fastp⁹⁵ with parameters “–cut_by_quality3 –cut_by_quality5 –n_base_limit 0 –length_required 60 –correction”. The KMC 3.1.1⁹⁶ was used to count the 31-mer of each clean dataset of the weeping willow. The *k*-mers with <30% missing and an average count >1 in each sex group were retained. The *k*-mers in all 40 individuals with

minor allele (*k*-mer absent or present) frequency <0.1 were discarded. The filtered *k*-mers were mapped to the genome (76 chromosomes) of weeping willow using Bowtie⁹⁷. *k*-mers with a coverage of 0 in all 20 female individuals were defined as male-specific *k*-mers, while *k*-mers with a coverage of 0 in 20 male individuals were defined as female-specific *k*-mers. We expect male (male heterogamety, male carrying XY, and female carrying XX) or female (female heterogamety, female carrying ZW, and male carrying ZZ) specific *k*-mers on a sex-linked region of sex chromosome(s).

We employed the CQ method³⁵ to detect the sex chromosomes of the two willows. We made combined female and male clean read datasets of the two willows, respectively. The cq-calculate.pl software³⁵ was used to calculate the CQ for each 50-kb nonoverlapping window of the genomes. For male heterogamety, the CQ is the normalized ratio of female to male alignments to a given reference sequence, and the CQ value is close to 2 (for XX/XY, 1.33 for XXXX/XXXXY) in windows in the X-linked region and zero in windows in the Y-linked region. For female heterogamety, the CQ is the normalized ratio of male to female alignments, and the CQ value is close to 2 (for ZW/ZZ, 1.33 for ZZZW/ZZZZ) in windows in Z-linked region and to zero in windows in the W-linked region.

We aligned clean reads of *S. dunnii* to each genome (both haplotypes) using the BWA-MEM algorithm from bwa 0.7.12^{98,99} with default parameters. Samtools 0.1.19¹⁰⁰ was used to extract primary alignments, sort, and merge the mapped data. PCR replicates were filtered using sambamba 0.7.1¹⁰¹. The variants were called and filtered using Genome Analysis Toolkit v. 4.1.8.1 and VCFtools 0.1.16¹⁰². Hard filtering of the SNP calls was carried out with “QD <2.0, FS >60.0, MQ <40.0, MQRankSum <–12.5, ReadPosRankSum <–8.0, SOR >3.0”. Only biallelic sites were kept for subsequent filtering. The sites with coverage greater than twice the mean depth at all variant sites across all samples were discarded. Genotypes with depth <4 were treated as missing, and sites with >10% missing data or minor allele frequency <0.05 were removed.

We used VCFtools to calculate weighted F_{ST} values between 18 male and 20 female genomes of *S. dunnii*¹⁰³ with 100-kb windows and 10-kb steps. The Changepoint package¹⁰⁴ was used to detect the boundaries of the SLRs based on F_{ST} and CQ values between the sexes of *S. dunnii* and CQ values and sex-specific *k*-mer counts of weeping willow, respectively. Furthermore, we used the Python version of MCScan¹⁰⁵ to analyze chromosome collinearity between the protein-coding sequences detected in 7 and 15 chromosomes of *S. dunnii* and weeping willow and their homologous autosomes, to detect possible inversions in sex chromosomes. The “–score = 0.99” was used to obtain reciprocal best hit (RBH) orthologs for the collinearity analysis. We then combined all the previous analyses to detect the SLRs.

ARR17 and *PI* identification and phylogeny of *ARR17*

In order to obtain *ARR17*-like sequences in target species, we used BLASTN to blast *ARR17*-like gene (Potri.019G133600²⁸) against the *S. purpurea*⁴⁷, *S. arbutifolia*²⁵, *S. dunnii*, and *S. babylonica* genomes with parameters “–evalue 1e-5 –word_size 8”. *ARR17*-like gene includes five exons, so only the sequences including all the exons were classified as intact *ARR17*-like genes, while sequences with <5 exons were regarded as partial *ARR17*-like duplicates²⁴. We also used BLASTN to identify the *PI*-like genes in the *S. babylonica* genome using Potri.002G079000 as the query³⁰. We used the exon regions of the identified *ARR17*-like sequences for phylogenetic reconstruction. These exons were aligned with MAFFT⁸⁷, then we constructed a phylogenetic tree with IQ-TREE.

Ancestral sex-linked gene identification

We used the OrthoFinder⁸⁶ to identify single-copy genes in SLRs of W and Z of *S. babylonica* and *S. purpurea*, 15X and 15Y of *S. arbutifolia*, and chromosome 15a of *S. dunnii*. We then extracted homologous genes, that diverged between ancestral X and Y in ancestors of *Vetrix*-clade

and close to partial *ARR17*-like gene duplicates, as proposed by ref. 25. We obtained one ancestral single-copy homologous gene, then used MAFFT and IQ-TREE to align and reconstruct the phylogenetic tree using *S. dunnii* as outgroup.

Sex-linked region features and gene loss

We calculated the content of gene and repeat sequences (total repeat, TE, LTR-Gypsy, LTR-Copia) in the PARs and SLRs for *S. babylonica* (15V_w, 15V_z) and *S. dunnii* (7X, 7Y). We calculated the difference between the SLRs and the PARs based on 100 kb windows.

We used chromosome 15a of *S. dunnii* as the reference to identify protein-coding gene loss in 15V_w, 15V_z, 15X, and 15Y, and used 7Va in *S. babylonica* as the reference to identify gene loss in 7X and 7Y. Firstly, we identified shared genes in W-Z and X-Y, and specific genes in W, Z, X, and Y. For specific genes in X-SLR/W-SLR, there is no homologous gene in Y-SLR/Z-SLR. Similarly, there is no homologous gene in X-SLR/W-SLR for specific genes in Y-SLR/Z-SLR. Then we used these specific genes to determine gene loss among them. The degradation rate of W-SLR = W-SLR loss/(W-SLR loss + Z-SLR loss + their shared genes), and the degradation rate of Z-SLR = Z-SLR loss/(W-SLR loss + Z-SLR loss + their shared genes). Similarly, we obtained the degradation rate of X-SLR and Y-SLR. We also obtained the relevant data on PARs.

ARR17-like duplicates and *PI* expression analyses

Analysis and identification of sRNAs from female and male buds of *S. babylonica* was performed using sRNAmimer v1.1.2¹⁰⁶. sRNAs were identified with sRNAanno database¹⁰⁷. Adapters, noncoding RNA (rRNA, tRNA, snoRNA, snRNA), and plasmid contamination were removed from the sRNA-Seq datasets. The clean reads were aligned to the reference genome of *S. babylonica* using sRNAmimer and read (per site) coverage estimated using IGV-sRNA (<https://gitee.com/CJchen/IGV-sRNA>). We calculated the average read (per site) coverage of biological replicates of each partial *ARR17*-like duplicate and around regions and each intact *ARR17*-like duplicate.

We calculated the gene expression (excluding non-mRNA) among female and male catkins of *S. babylonica*, *S. arbutifolia*, and *S. dunnii*, respectively. The RNA datasets of *S. arbutifolia* are from Wang et al.²⁵. Each sex and individual contained three independent biological replicates (Supplementary Data 10). After filtering, clean transcript reads from each sample were mapped to their own genome with HISAT2 v2.1.0¹⁰⁸. We used a haplotype genome assembly and sex chromosomes (15X and 15Y or 7X and 7Y) as reference genomes in *S. arbutifolia* and *S. dunnii*, and *Sa* and *Va* haplotypes genome assembly and sex chromosomes (15V_w and 15V_z) as reference genomes in *S. babylonica*. The number of reads mapping to each gene was calculated using featureCounts¹⁰⁹. Then we converted these read counts to TPM (transcripts per million reads). After filtering out unexpressed genes (counts = 0 in all samples), then we used the expression levels of the SLRs in males and females and their autosomes to identify the dosage compensation pattern¹¹⁰. The expression level of flower buds from *S. babylonica* was calculated using HISAT2 and IGV v2.17.1¹¹¹. We obtained the average read coverage of biological replicates of each *ARR17*-like and *PI*-like gene.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The genome assembly sequences generated in this study have been deposited in the National Genomics Data Center (NGDC) database under BioProject [PRJCA016000](https://www.ncbi.nlm.nih.gov/bioproject/PRJCA016000) and in the NCBI database under BioProject [PRJNA1130083](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1130083), [PRJNA1130084](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1130084), [PRJNA1130085](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1130085), and [PRJNA1130087](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1130087). The sequencing datasets generated in this study have been deposited in the NCBI database under BioProject accession codes

[PRJNA882493](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA882493), [PRJNA1020583](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1020583), and [PRJNA1020619](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1020619). Source data are provided with this paper.

References

- Evans, B. J., Alexander Pyron, R. & Wiens, J. J. Polyploidization and sex chromosome evolution in amphibians. *in Polyploidy and genome evolution* (eds Pamela S. & Soltis, D. E. S.) 385–410 (Springer, 2012).
- Van de Peer, Yves, Mizrachi, Eshchar & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
- Muller, H. J. Why polyploidy is rarer in animals than in plants. *Am. Nat.* **59**, 346–353 (1925).
- He, L. & Hörandl, E. Does polyploidy inhibit sex chromosome evolution in angiosperms? *Front. Plant Sci.* **13**, 976765 (2022).
- Mable, B. K., Alexandrou, M. A. & Taylor, M. I. Genome duplication in amphibians and fish: an extended synthesis. *J. Zool.* **284**, 151–182 (2011).
- Gates, R. R. Polyploidy and sex chromosomes. *Nature* **117**, 234 (1926).
- Ramsey, J. & Schemske, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* **29**, 467–501 (1998).
- He, L. et al. Evolutionary origin and establishment of a dioecious diploid-tetraploid complex. *Mol. Ecol.* **32**, 2732–2749 (2023).
- Stöck, M. et al. Sex chromosomes in meiotic, hemiclinal, clonal and polyploid hybrid vertebrates: along the ‘extended speciation continuum’. *Philos. Trans. R. Soc. B* **376**, 20200103 (2021).
- Gulyaev, S. et al. The phylogeny of *Salix* revealed by whole genome re-sequencing suggests different sex-determination systems in major groups of the genus. *Ann. Bot.* **129**, 485–498 (2022).
- Otto, S. P. & Whitton, J. Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437 (2000).
- Scott, M. F., Osmond, M. M. & Otto, S. P. Haploid selection, sex ratio bias, and transitions between sex-determining systems. *PLoS Biol.* **16**, e2005609 (2018).
- Mawaribuchi, S. et al. Sex chromosome differentiation and the W- and Z-specific loci in *Xenopus laevis*. *Dev. Biol.* **426**, 393–400 (2017).
- Akagi, T., Henry, I. M., Kawai, T., Comai, L. & Tao, R. Epigenetic regulation of the sex determination gene *MeGI* in polyploid persimmon. *Plant Cell* **28**, 2905–2915 (2016).
- Robertson, F. M. et al. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol.* **18**, 1–14 (2017).
- Mason, A. S. & Wendel, J. F. Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. *Front. Genet.* **11**, 564174 (2020).
- Cunado, N. et al. The evolution of sex chromosomes in the genus *Rumex* (Polygonaceae): identification of a new species with heteromorphic sex chromosomes. *Chromosom. Res.* **15**, 825–833 (2007).
- Bewick, A. J., Anderson, D. W. & Evans, B. J. Evolution of the closely related, sex-related genes *DM-W* and *DMRT1* in African clawed frogs (*Xenopus*). *Evolution* **65**, 698–712 (2011).
- Newsholme, C. *Willows: The Genus Salix* (Timber Press, 1992).
- Isebrands, J. G. & Richardson, J. *Poplars and Willows: Trees for Society and the Environment* (CABI, 2014).
- Wagner, N. D., Volf, M. & Hörandl, E. Highly diverse shrub willows (*Salix* L.) share highly similar plastomes. *Front. Plant Sci.* **12**, 662715 (2021).
- Fang, Z., Zhao, S. D. & Skvortsov, A. K. Salicaceae. *Flora China* **4**, 139–274 (1999).
- Wang, Y. et al. 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**, 177 (2023).
24. Wang, D. et al. Repeated turnovers keep sex chromosomes young in willows. *Genome Biol.* **23**, 1–23 (2022).
 25. Wang, Y. et al. Gap-free X and Y chromosomes of *Salix arbutifolia* reveal an evolutionary change from male to female heterogamety in willows, without a change in the sex-determining region. *N. Phytol.* **242**, 2872–2887 (2024).
 26. Xue, Z. Q., Applequist, W. L. & Elvira Hörandl, L. H. Sex chromosome turnover plays an important role in the maintenance of barriers to post-speciation introgression in willows. *Evol. Lett.* **8**, 467–477 (2024).
 27. Teng, S. C. & Yu, H. Propagation of weeping willow from seed. *Bot. Bull. Acad. Sin.* **2**, 131–132 (1948).
 28. Müller, N. A. et al. A single gene underlies the dynamic evolution of poplar sex determination. *Nat. Plants* **6**, 630–637 (2020).
 29. Xue, L. et al. Evidences for a role of two Y-specific genes in sex determination in *Populus deltooides*. *Nat. Commun.* **11**, 5893 (2020).
 30. Leite Montalvão, A. P., Kersten, B., Kim, G., Fladung, M. & Müller, N. A. *ARR17* controls dioecy in *Populus* by repressing B-class MADS-box gene expression. *Philos. Trans. R. Soc. B* **377**, 20210217 (2022).
 31. He, L. et al. Chromosome-scale assembly of the genome of *Salix dunnii* reveals a male-heterogametic sex determination system on chromosome 7. *Mol. Ecol. Resour.* **21**, 1966–1982 (2021).
 32. Wu, J. et al. Phylogeny of *Salix* subgenus *Salix* s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. *BMC Evol. Biol.* **15**, 1–13 (2015).
 33. Zhang, Q., Liu, Y. & Sodmergen Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant Cell Physiol.* **44**, 941–951 (2003).
 34. He, L., Wagner, N. D. & Hörandl, E. Restriction-site associated DNA sequencing data reveal a radiation of willow species (*Salix* L., Salicaceae) in the Hengduan mountains and adjacent areas. *J. Syst. Evol.* **59**, 44–57 (2021).
 35. Hall, A. B. et al. Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics* **14**, 1–13 (2013).
 36. Harlan, J. R., deWet, J. M. J. & On, Ö. Winge and a prayer: the origins of polyploidy. *Bot. Rev.* **41**, 361–390 (1975).
 37. Ming, R., Bendahmane, A. & Renner, S. S. Sex chromosomes in land plants. *Annu. Rev. Plant Biol.* **62**, 485–514 (2011).
 38. Prentout, D. et al. An efficient RNA-seq-based segregation analysis identifies the sex chromosomes of *Cannabis sativa*. *Genome Res.* **30**, 164–172 (2020).
 39. Wright, A. E., Dean, R., Zimmer, F. & Mank, J. E. How to make a sex chromosome. *Nat. Commun.* **7**, 12087 (2016).
 40. Beukeboom, L. W. & Perrin, N. *The Evolution of Sex Determination* (Oxford Univ. Press, 2014).
 41. Troups, M. A., Vicoso, B. & Pannell, J. R. Dioecy and chromosomal sex determination are maintained through allopolyploid speciation in the plant genus *Mercurialis*. *PLoS Genet.* **18**, e1010226 (2022).
 42. Gerchen, J. F., Veltos, P. & Pannell, J. R. Recurrent allopolyploidization, Y-chromosome introgression and the evolution of sexual systems in the plant genus *Mercurialis*. *Philos. Trans. R. Soc. B* **377**, 20210224 (2022).
 43. Cauret, C. M. S., Mortimer, S. M. E., Roberti, M. C., Ashman, T.-L. & Liston, A. Chromosome-scale assembly with a phased sex-determining region resolves features of early Z and W chromosome differentiation in a wild octoploid strawberry. *G3* **12**, jkac139 (2022).
 44. Hillman, G., Hedges, R., Moore, A., Colledge, S. & Pettitt, P. New evidence of Lateglacial cereal cultivation at Abu Hureyra on the Euphrates. *Holocene* **11**, 383–393 (2001).
 45. Papadopulos, A. S. T., Chester, M., Ridout, K. & Filatov, D. A. Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proc. Natl Acad. Sci. USA* **112**, 13021–13026 (2015).
 46. Almeida, P. et al. Genome assembly of the basket willow, *Salix viminalis*, reveals earliest stages of sex chromosome expansion. *BMC Biol.* **18**, 1–18 (2020).
 47. Zhou, R. et al. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biol.* **21**, 1–19 (2020).
 48. Wang, J. et al. Sequencing papaya X and Y^h chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl Acad. Sci. USA* **109**, 13710–13715 (2012).
 49. Hough, J., Hollister, J. D., Wang, W., Barrett, S. C. H. & Wright, S. I. Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. *Proc. Natl Acad. Sci. USA* **111**, 7713–7718 (2014).
 50. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
 51. Charlesworth, B., Sniegowski, P. & Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220 (1994).
 52. Ellegren, H. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat. Rev. Genet.* **12**, 157–166 (2011).
 53. Muyle, A. et al. Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* **10**, e1001308 (2012).
 54. Charlesworth, B. Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl Acad. Sci. USA* **75**, 5618–5622 (1978).
 55. Mank, J. E. Sex chromosome dosage compensation: definitely not for everyone. *Trends Genet.* **29**, 677–683 (2013).
 56. Muyle, A., Shearn, R. & Marais, G. A. B. The evolution of sex chromosomes and dosage compensation in plants. *Genome Biol. Evol.* **9**, 627–645 (2017).
 57. Ohno, S. *Sex Chromosomes and Sex Linked Genes* (Springer, 1967).
 58. Muyle, A., Marais, G. A. B., Bačovský, V., Hobza, R. & Lenormand, T. Dosage compensation evolution in plants: theories, controversies and mechanisms. *Philos. Trans. R. Soc. B* **377**, 20210222 (2022).
 59. Doležel, J., Greilhuber, J. & Suda, J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2**, 2233–2244 (2007).
 60. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
 61. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1–10 (2020).
 62. Belton, J.-M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
 63. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
 64. Durand, N. C. et al. JuiceR provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
 65. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
 66. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
 67. Xu, G.-C. et al. LR_GapCloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* **8**, giy157 (2019).

68. Hu, J., Fan, J., Sun, Z. & Liu, S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
69. Jin, J.-J. et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 1–31 (2020).
70. Prysycz, L. P. & Gabaldón, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**, e113–e113 (2016).
71. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
72. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
73. Perteu, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
74. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
75. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
76. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
77. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
78. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
79. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 1–11 (2005).
80. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, 1–22 (2008).
81. Zhang, R.-G. et al. TEsorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. *Hortic. Res.* **9**, uhac017 (2022).
82. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
83. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
84. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
85. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 1–18 (2019).
86. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
87. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
88. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
89. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 15–30 (2018).
90. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
91. Bi, G., Mao, Y., Xing, Q. & Cao, M. HomBlocks: a multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block searching. *Genomics* **110**, 18–22 (2018).
92. Xu, P. et al. The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. *Nat. Commun.* **10**, 1–11 (2019).
93. Delcher, A. L., Salzberg, S. L. & Phillippy, A. M. Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinformatics* **Chapter 10**, Unit 10.3 (2003).
94. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
95. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
96. Kokot, M., Dlugosz, M. & Deorowicz, S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* **33**, 2759–2761 (2017).
97. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, 1–10 (2009).
98. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv1303.3997 (2013).
99. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
100. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
101. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
102. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
103. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution*. **38**, 1358–1370 (1984).
104. Killick, R. & Eckley, I. changepoint: an R package for changepoint analysis. *J. Stat. Softw.* **58**, 1–19 (2014).
105. Tang, H. et al. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **18**, 1944–1954 (2008).
106. Li, G., Chen, C., Chen, P., Meyers, B. C. & Xia, R. sRNAmminer: a multifunctional toolkit for next-generation sequencing small RNA data mining in plants. *Sci. Bull.* **69**, 784–791(2023).
107. Chen, C. et al. sRNAanno—a database repository of uniformly annotated small RNAs in plants. *Hortic. Res.* **8**, 45 (2021).
108. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
109. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
110. Darolti, I. et al. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proc. Natl Acad. Sci. USA* **116**, 19031–19036 (2019).
111. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

Acknowledgements

This study was financially supported by the National Natural Science Foundation of China (grant no. 32171813 to L.H.) and the Special Fund for Scientific Research of Shanghai Landscaping & City Appearance Administrative Bureau (grant nos. G232403 and G242417 to L.H.). We are grateful to Si-Wen Zeng for sampling. We are indebted to James H. Leebens-Mack, Suhua Yang, Zhenyang Liao, Zhiqing Xue, Guangnan Gong, and Zhiying Zhu for their kind help during the preparation of our paper. We are grateful to the staff at Urban Horticulture Research and Extension Center of Shanghai Chenshan Botanical Garden for providing their facilities and help.

Author contributions

L.H. conceived the study. L.H., Yuàn W. and Yi W. designed and conceptualized the study. L.H., Yuàn W., Yi W., R.Z. and Yuàn W. performed the data analysis. L.H. and Yi W. collected materials. L.H. and Yuàn W. wrote the manuscript. E.H., T.M., Y.M., J.E.M. and R.M. revised the manuscript. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-51158-3>.

Correspondence and requests for materials should be addressed to Li He.

Peer review information *Nature Communications* thanks Kanae Masuda and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024