Genome Biology

https://doi.org/10.1186/s13059-026-04026-w

Article in Press

Recurrent sex chromosome turnover mediated by distinct *ARR17* and *PISTILLATA* duplications in willows

Received: 17 Jun 2025

Accepted: 04 Mar 2026

Published online: 12 March 2026

Cite this article as: Wang, Y., Xue, Z., Zhang, R. *et al*. Recurrent sex chromosome turnover mediated by distinct *ARR17* and *PISTILLATA* duplications in willows. *Genome Biol* (2026). https://doi.org/10.1186/s13059-026-04026-w

Yuàn Wang, Zhi-Qing Xue, Ren-Gang Zhang, Zhi-Ying Zhu, Elvira Hörandl, Xiao-Ru Wang, Yan-Fei Mao, Deborah Charlesworth & Li He

Recurrent sex chromosome turnover mediated by distinct *ARR17* and *PISTILLATA* duplications in willows

Yuàn Wang[1,§], Zhi-Qing Xue[1,§], Ren-Gang Zhang[2], Zhi-Ying Zhu[1], Elvira Hörandl[3], Xiao-Ru Wang[4], Yan-Fei Mao[5], Deborah Charlesworth[6,*], Li He[1,*]

[1] Key Laboratory of East China Plant Conservation and Utilization, National Forestry and Grassland Administration, Shanghai Chenshan Botanical Garden, Shanghai, 201602, China

[2] 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

[3] Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), University of Goettingen, Göttingen, Germany

[4] Department of Ecology and Environmental Science, Umeå Plant Science Centre, Umeå University, Umeå, Sweden

[5] CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, 200032, China

[6] Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, UK

[§]Contributed equally to this study.

*Corresponding Author: Deborah Charlesworth, deborah.charlesworth@ed.ac.uk; Li He, lhe@cemps.ac.cn

Abstract

Background

Sex chromosome turnovers evolve via translocation or duplication of established sex-determining genes, or their replacement by newly evolved ones. Few cases of replacements by new factors have been documented in dioecious plants, but are suspected in *Salix*, in which both XY and ZW systems occur, with sex-linked regions (SLRs) of different species on various chromosomes. The male-determining genes in XY species' SLRs are partial duplicates of autosomal *ARR17*-like genes and regulate the expression of downstream genes involved in stamen development by producing small RNAs that suppress the expression of intact copies.

Results

Here we describe phased chromosomal assemblies of three *Salix* species with a ZW system derived from an XY system, including four lineages of the *Salix polyclona* complex (six assemblies in total). Their SLRs are within the same repeat-rich pericentromeric region of chromosome 15 as in the XY system. Although these Z- and W- SLRs carry intact and/or partial *ARR17* duplicates, few RNA products are detectable in our sampled tissues. However, the W-SLRs include partial duplicates of *PISTILLATA* (*PI*), a stamen development gene. These are arranged in inverted repeats and express small interfering RNAs targeting the autosomal intact *Salix PI* gene, suggesting that they reduce its expression, and therefore act as maleness-suppressing factors.

Conclusions

The turnover events involving intact *ARR17* and partial *PI* duplications in the 15ZW clade I species involve pericentromeric regions that recombine rarely, making changes possible in these *Salix* species that would be unlikely in other genome regions.

Keywords

Salix, sex determination, sex chromosome turnovers, pericentromeric regions, recombination landscape, translocation

Background

Sex chromosomes were first recognised due to their heteromorphism in organisms such as beetles, mammals, and birds. These sex chromosomes are ancient [1]. However, many vertebrates have younger sex chromosomes, including many fish [2], lizards [3] and amphibians [4], in which the evolutionary process restarted after "turnover events" in which new sex-determining factors arose, or existing factors moved to new chromosomal locations with various possible recombination rates [5]. Such new sex-determining regions may not become genetically degenerated because they are physically small regions in recombining genome regions, but, even if the new sex-determining locus is within an extensive region that recombines rarely, occasional crossovers can prevent degeneration [6]. As the sex-determining systems in many flowering plants evolved recently, and many have non-degenerated sex chromosomes [7], they are well suited for detailed studies of such changes. Plants are also of interest for examining the suggestion that rarely recombining sex-linked regions may often evolve when turnovers lead to new sex-determining loci evolving within pericentromeric regions that already recombine rarely [8]. The pericentromeric regions usually contain repetitive sequences and transposable elements, with low gene density and low recombination rate surrounding the centromere, but covering the centromere in our study [9]. Such regions can be extensive, and are generally enriched in repetitive elements, compared with other genome regions (as reviewed in [10,11]). These properties make such regions permissive for genomic rearrangements, such as inversions and duplications [12], and their low gene densities may allow deletions to occur. Several examples of such pericentromeric sex-linked regions have now been discovered, for example *Rumex* [13]. Here, we examine a genus in which such regions appear to be involved in turnover events.

In different flowering plants, different sex-determining genes have been discovered [14], reflecting multiple independent origins of dioecy [15,16], though, in some plants, an existing sex-determining gene has moved its location, as has been documented in strawberries [17,18] and kiwifruit [19]. In

fish, many cases are known of a new gene replacing an ancestral sex determining gene [20] (sometimes termed “rewiring” the sex determining pathway [21]). Such turnover events can cause changes between male and female heterogamety (for brevity, termed XY and ZW systems) [10,22]. Theoretical modelling has shown that sexually-antagonistic (SA) polymorphisms can favour movements of sex-determining loci to genomic locations near such polymorphic loci [23,24], and some animal turnovers have been suggested to reflect this process, including in the hessian fly, *Mayetiola destructor* [25], cichlid fish [26], and a halibut species, *Hippoglossus hippoglossus* [27], but definitive evidence for involvement of SA polymorphisms is lacking, and other possibilities exist, as we discuss below.

Although, as noted above, sex chromosome turnovers have been documented in plants, closely related species have rarely been compared in a reliable phylogeny to infer the events involved [28]. Poplar and willow species (the Salicaceae genera *Populus* and *Salix*, respectively) are suitable for such studies, as the sex-determining genes are on different, non-homologous, chromosomes in different species in the two genera, and changes from XY to derived ZW systems are documented [29–32], as described further below. Before describing turnover events in the Salicaceae, we first outline the sex development genes that have been identified in the two sister genera. In the genus *Populus*, *P. tremula* has an XY system. Its sex-determination gene was the first to be identified in the Salicaceae. An *ARABIDOPSIS RESPONSE REGULATOR 17* (*ARR17*)-like gene was shown to have Y-linked partial duplicates arranged as inverted repeats that can generate double-stranded RNAs (dsRNAs), producing small interfering RNAs (siRNAs) [33]. These siRNAs can silence expression of the autosomal *P. tremula* intact *ARR17*-like gene, causing maleness [10,29–36]; in the absence of this silencing, female development occurs, as the intact *ARR17* copy, which is essential for female development, can be expressed [37]. A similar system, with different partial duplications, appears to act in *P. deltoides* [38]. In *Populus alba*, however, an ancestral male heterogametic sex determination system with Y-chromosomal partial *ARR17* duplicates (present in *P. trichocarpa* and related species) has changed to female heterogamety, and there is a W-specific intact *ARR17* copy [30,33]. In all cases, expression of intact *ARR17*-like genes ‘triggers’ the expression of several downstream genes,

indirectly suppressing expression of the homolog of the *A. thaliana* B-class gene *PISTILLATA* (*PI*), which is essential for stamen development. This regulation was confirmed in *Populus* by *ARR17* knockout CRISPR mutants [36]; similar *PI*-like genes participate in the downstream regulation of male sex determination in the persimmon [39].

In *Salix*, the sister genus to *Populus*, partial *ARR17*-like duplicates are again present in the sex-linked regions of clades with both male and female heterogamety (below, we term their respective Y- and Z-linked regions SLRs). The *Salix* clade, including *Salix mesnyi*, *S. chaenomeloides* and *S. dunnii*, has SLRs on chromosome 7, and these are termed 7XY species. The *Vetrix* clade includes both XY and ZW systems, both with SLRs on chromosome 15 [29,31,32,40]. In the 15XY *Vetrix* species, *S. triandra*, *S. arbutifolia* and *S. exigua* [41], intact *ARR17*-like orthologs are located on an autosome (chromosome 19). When *ARR17* is expressed (in the absence of partial *ARR17*-like sequences, which are found only in the 15Y-SLR), the *PI*-like gene is likely indirectly silenced, resulting in female development [29,31,42]. In the *Vetrix* 15ZW species, *S. purpurea*, intact *ARR17*-like sequences were found on chromosome 19 and in the chromosome 15W-SLR, while partial *ARR17*-like sequences were found only in the 15Z-SLR [43]. Only *ARR17* on 15W had female-specific expression, but the expression levels were low, suggesting that the sex-determination mechanism might differ from the system in *Populus* and XY willows [44].

The *S. purpurea* 15W-SLR also includes a sequence similar to *GATA* transcription factors (named *GATA15* [45]), which might promote female development, as some members of this group of DNA-binding proteins (with a class IV zinc finger motif, involving light- and nitrate-dependent control of transcription), act downstream of *PI* in *A. thaliana* [46,47]. The sex-determining genes in other 15ZW *Salix* species are still unknown, and one goal of the present study was to investigate whether the sex chromosome turnovers described above also involved changes in candidates, thereby identifying the sex-determining genes of further species, and potentially helping to understand the evolution of the turnovers.

The 15ZW sex determining system evolved from an ancestral 15XY system, and studies of the diploid species *S. exigua* (15XY) and *S. purpurea* (15ZW) suggested that the Y changed into a Z chromosome [41]. This change was inferred by comparing phased *S. purpurea* SLR sequences with the 15XY-SLRs of *S. arbutifolia* [10] (Fig. 1a), and has been confirmed in further willow species [42]; specifically, for a gene whose sequences cluster by gametologs, the 15Ys of *S. triandra* and *S. arbutifolia* group with the 15Zs of *S. purpurea* and *S. polyclona*, while their 15Xs group with the *S. purpurea* and *S. polyclona* 15Ws, supporting 15X → 15W and 15Y → 15Z transitions in this clade [42]. Wang et al. [29] suggested that duplicative translocation of intact *ARR17*-like genes from the autosomal chromosome 19 to chromosome 15 might be associated with this turnover event. To distinguish among different lineages, we use the term "15ZW clade I" to refer to all *Vetrix* clade species with 15ZW sex determination systems that evolved from diploid ancestors, to distinguish them from the hybrid allotetraploid *Vetrix* clade species (termed 15ZW clade II), which includes one species studied here, *S. babylonica* [31]. *S. babylonica* arose from crosses between an unknown ancestral 7XY *Salix* clade species and a 15XY species from the *Vetrix* clade (Fig. 1, in which *S* and *V* denote these sources, respectively); each of its homoeologous chromosome 19 pairs carries two copies of the intact *ARR17* gene, resulting in a total of eight intact copies in its genome. This duplication probably occurred in a 7XX 15XY ancestor, and may have affected the original sex determination, resulting in XY females and an independent transition to a ZW system in *S. babylonica* [31] (Fig. 1a, 15XY to 15ZW in 4X ancestors).

To study sex chromosome evolution in *Vetrix* clade 15ZW clade I species, and compare their genomes with that of *S. purpurea*, we sequenced and assembled phased (haplotype-resolved) chromosomes of two additional diploid species in this subclade (*Salix baileyi* and *S. gordejevii*) and four lineages representing the core phylogenetic groups of the *S. polyclona* complex [48] (Table 1). He et al. [48] inferred female-heterogamy with a W-linked region on chromosome 15 in both diploid and tetraploid *Salix polyclona*, and Xue et al. [32] confirmed that the sex-determination system differs between major subgroups of *Salix*, while heterogamety and SLR locations are conserved within the subgroups. As the *Vetrix* clade I includes species with confirmed 15ZW systems, *S.*

purpurea and *S. viminalis*, we expected the other members of this clade studied here, *S. baileyi* and *S. gordejevii*, to also share this system. To include both the W- and Z-linked regions, we therefore sequenced and assembled female genomes of these species. Because different species or lineages can have different SLRs [29,49], which might be partially or fully sex-linked, we identified the sex chromosomes and SLRs independently in each of the taxa studied, using the re-sequencing datasets described in the Methods section, and to explore changes we compared these with previous genome assemblies of *Salix* clade 7XY species and *Vetrix* clade 15XY and 15ZW species for which phased sex-linked regions have been published [10,31,42,43]. We identified candidate sex-determining factors, and validated them with new expression data and dual fluorescence experiments. The analyses reveal multiple changes in genus *Salix* lineages, and support the hypothesis that the XY → ZW change in the 15ZW clade I species involved evolution of female heterogamety involving suppression of *PISTILLATA*, without changing the genomic location of the SLR, which is within the homologous repeat-rich pericentromeric regions in all these species. We discuss how this location suggests an evolutionary basis for the changes.

Results

Ploidy analysis

As reviewed in [50], diploid species in the Salicaceae are palaeotetraploids resulting from a whole genome duplication (WGD) event in an ancestor of *Populus* and *Salix* about 58 million years ago, before the split of the different *Salix* lineages [31,51], from which about 30% of genes remain as paralogs; more recent polyploids also exist. Among the newly sequenced 15ZW clade I species, *S. purpurea* [43], *S. baileyi* [52] and *S. polyclona*-E-TS [48] are diploids. Flow cytometry analysis (Additional file 1: Fig S1) further confirmed diploidy of *S. gordejevii* and the *S. polyclona*-E and -W1 lineages, while the W2 lineage of *S. polyclona* is tetraploid. Our *k-mer* results (Additional file 1: Fig S2) confirm a previous report that *S. polyclona*-W2 is an autotetraploid [48].

Chromosome-level genome assemblies

From initial primary assemblies, we generated phased chromosomal assemblies for 15ZW clade I taxa, using PacBio HiFi reads, Illumina short reads and Hi-C reads, followed by error correction, polishing, and scaffolding (see the Methods section and sequencing statistics in Additional file 2: Table S1). For the tetraploid *S. polyclona*-W2, 54.28 Gb of Oxford Nanopore Technology (ONT) long reads were also used for assembly.

In our six assemblies (Table 1), Hi-C analyses successfully phased the homologous chromosomes (Additional file 1: Fig S3). The 19 chromosomes were numbered according to their homologies with *S. arbutifolia* chromosomes [10]. To test the accuracy of phasing of the sex-linked region, we aligned unitigs (unique contiguous sequences) produced after the unambiguous path extension step and before contig assembly (see Methods) to the assembled genomes. Chimeric sequences were rare with hifiasm (Hi-C) phasing [53], as divergence between Z- and W-linked alleles so far studied is high enough to distinguish the two haplotypes clearly [31,43], and hifiasm assembled them into different unitigs [54]. The alignments of unique contiguous sequences to the assembled genomes indicated that all six assemblies were fully phased in the sex-linked regions (Additional file 1: Fig S4). In our assemblies of *S. baileyi*, *S. polyclona*-E, E-TS, and W1, the sex chromosomes were gap-free. *S. gordejevii* had only two gaps, and *S. polyclona*-W2 only one (Additional file 2: Table S2); as described in the section below entitled “Identification of the SLRs”, these gaps were not in the regions inferred to be sex-linked by the criteria described below (defined as including the sex-determining factors, as discussed shortly, and termed SLRs, in what follows).

The assembled genome sizes of the five diploid individuals sequenced (see Table 1) were between 652 and 831 Mb, with 38 pseudochromosomes including the two haplotypes (*a* and *b*) of each chromosome, corresponding to the expected 19 pairs of chromosomes (Table 1, Fig. 1, Additional file 1: Fig S3). The 1,507 Mb tetraploid *S. polyclona*-W2 assembly (Table 1) assembled into 77 pseudochromosomes, each represented by four haplotypes (a, b, c and d), and a potential B chromosome, termed chromosome 20 (Fig. 1, Additional file 1: Fig S3). Consistent with the Hi-C

results described above, all chromosome pairs in both diploids and tetraploids exhibit excellent alignment between their haplotype assemblies, whereas chromosome 20 shows no blocks with homology to any other chromosomes (Additional file 1: Fig S5). The contig N50 lengths range from 15 to 23 Mb (Table 1), and both the HiFi reads and Illumina short reads map to more than 99% of the assembled regions. Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis suggested completeness exceeding 97% for all six assemblies (Additional file 2: Table S3).

The predicted numbers of genes and protein-coding genes for each assembly are shown in Table 1 and Additional file 2: Table S4. The average protein-coding gene length in these species (including exons, introns, and both 5′ and 3′ UTRs)**,** ranges from 3,435.3 bp to 3,753.2 bp, and the mean coding sequence lengths range from 1,328.1 bp to 1,362.3 bp (Additional file 2: Table S4). After combining three annotation methods, eggNOG-mapper, DIAMOND, and InterProScan, less than 1.5% of genes remained unannotated in each assembled genome (Additional file 2: Table S5). All the genomes have high overall repetitive sequence contents, from 43.95% to 57.28% (Table 1, Additional file 2: Table S6).

Phylogenetic relationships among the *Salix* species studied

To understand the phylogenetic relationships of the three newly sequenced species (six individuals, including the four lineages from the *S. polyclona* complex), we estimated a species tree including other species from the *Vetrix* and *Salix* clades (Fig. 1a), using *Populus trichocarpa* as an outgroup (Additional file 1: Fig S6, Additional file 2: Table S7). In the *Vetrix* clade 15ZW clade I, *S. baileyi* clusters with *S. purpurea* (also in the clade I), and then with *S. gordejevii*. These species together form sister clades to the *S. polyclona* complex, with the *S. babylonica* V sequences forming a basal lineage, and two additional *Vetrix* clade 15XY species as further basal lineages. The *Salix* clade species are more distant outgroups of the *Vetrix* clade, and all branches of the species tree are strongly supported, with posterior probabilities of 1.

Identification of sex-linked regions in the new genome assemblies and evidence that they are pericentromeric regions

Analyses of male-female sequence differences (chromosome quotient, or CQ, F_{ST} and *k-mer* analyses) and synteny were used to determine the heterogamety and locate the SLRs in the sequenced individuals, following methods previously used for other *Salix* species [10,31,42], including changepoint analysis to determine their boundaries. Due to the limited size of the *S. polyclona*-E-TS population, F_{ST} analysis was omitted for this lineage [55]. When a sex-determining factor (which is completely sex-linked, by definition) is within a genome region that recombines very rarely, such as the pericentromeric regions inferred from our findings described below, occasional recombination may prevent strong differentiation of the two haplotypes in regions other than the sex-determining gene itself. There may therefore be no definitive wider fully sex-linked region. We therefore divided the analyses into two steps, first determining sex-linked regions, defined as contiguous genomic regions that include the sex-determining locus. An SLR is detectable by analyses of F_{ST}, CQ and *k-mer* values, as a region with a peak of consistent signals of sequence divergence between the two haplotypes, reflecting either complete or partial sex-linkage with linkage disequilibrium between sequence variants and the sex-determining locus (we excluded short regions on various chromosomes with inconsistent signals in these different analyses). After describing the results of these analyses, we describe evidence for the pericentromeric locations of the SLRs identified, and then describe phylogenetic analyses that can detect recombination in the region, which revealed that indeed few genes show complete sex linkage. In all the newly sequenced taxa, signals of sex linkage are detected on chromosome 15, similar to the previously studied *Vetrix* clade 15ZW clade I species *S. purpurea* [43] (Additional file 3, Additional file 1: Figs. S7 to S18). In the summaries in Table 2 and Fig. 2, we included regions inverted between the two haplotypes at the borders of these regions in our inferred SLRs.

The sizes of these inferred SLRs differ considerably between the species, ranging from 3.61 to 10.93 Mb. The Z-linked regions occupy between 17.84 and 44.81% of chromosome 15 (with between 266 and 529 genes), versus 240 to 661 genes in the W-linked regions (16.94 to 46.56% of the total

chromosome sizes). The large size differences between the Z- and W-SLR regions are due to different amounts of repetitive sequences and gene duplications (Additional file 2: Table S8). In all species, both haplotypes have higher TE densities in these regions (exceptionally high for the terminal repeat (LTR) elements *Copia* and *Gypsy*), than in the PARs or the rest of these species' genomes (Fig. 3a, b, Additional file 1: Fig S19). The ends of all *Salix* chromosomes are heavily enriched in tandem repeats (Additional file 1: Fig S20 shows the sex chromosomes of the sequenced species) and are completely devoid of protein-coding genes, like the telomeres and sub-telomeric regions [56] of many organisms [57–59], including other members of the Salicaceae family [60]. Excluding these regions from our analysis of TEs (see Methods), TE densities decrease steadily on each arm of chromosome 15 in all species, and gene densities correspondingly increase, with distance from a region of elevated linkage disequilibrium that we infer is closest to the centromeric non-recombining region (Additional file 1: Fig S21).

In plants, CENH3 (centromere-specific histone H3) is an epigenetic marker defining centromere identity and location [61,62]. Wang et al. [60] found that CRMs (Centromeric Retrotransposon in Maize) showed high CENH3 enrichment in three Salicaceae species, *Populus alba* var. pyramidalis, *P. euphratica*, and *Salix chaenomeloides*. We also found CRM-enriched regions overlapped with the SLRs defined by other methods in our species (Additional file 1: Fig S22). We therefore conclude that the sex-linked regions of these 15ZW species (which are derived from 15XY ancestors) all coincide with the pericentromeric regions of these chromosomes [60], which are metacentrics with various arm ratios [63], consistent with findings in species with 15XY systems, including *S. arbutifolia* and *S. triandra* [10,42,60]. As gene and TE density change continuously, precise pericentromeric region boundaries cannot be defined, but the finding that regions identified above are also highly rearranged between the two haplotypes within each species supports our inferences (Additional file 1: Fig S23). The autosomes homologous to the sex chromosomes studied here (chromosome 15 of *S. dunnii* and *S. mesnyi*), also have regions containing CRMs, within regions with similar pericentromeric characteristics supporting rarity of recombination, including extensive LD blocks, high densities of repetitive sequences, and low gene density (Additional file 1: Fig S24-

S25). However, these autosomes show many fewer rearrangements than the sex chromosomes (Additional file 1: Fig S13-S18, S24-S25).

Evolution of sex-linked genes in the newly sequenced *Salix* taxa

Although the genus *Salix* is at least 44 million years old (Fig S6 of [52]), chromosome 15 is largely collinear in 15ZW and 15XY species (Additional file 1: Fig S23). Of the several hundred genes found in most 15ZW species' SLRs, an analysis using the autosomal 15a haplotype of *S. dunnii* (7XY) as an outgroup shows that the vast majority are shared by the Z and W, with rarely more than 10% loss of genes from the W-linked regions, and quite similar losses from the Z-linked regions, suggesting at most slight genetic degeneration of the W (Fig. 3c, Additional file 2: Table S9).

The highest proportions of gene losses are found in the 15XY-clade species, consistent with having evolved early, and ZW being the derived state (see the Introduction). Degeneration is much lower than in the oldest-established *Silene latifolia* sex-linked region (more than 70%), in which Y-X synonymous site divergence (Ks values) are between 10 and 20% [11]. The evolution of these pericentromeric regions will be analysed in the future.

The *Salix* ZW-linked regions might therefore be much younger than that in *S. latifolia*, and we estimated W-Z Ks values to test this. Genes are absent from both the Z- and W-linked regions of one or more of these species and SLRs' boundaries are different, precluding their use in Ks analysis of the whole newly sequenced species set. To estimate relative times since W-Z recombination ceased in these species, we therefore used a sub-set of 21 single- or low-copy orthogroups of protein-coding genes shared by SLRs of all three newly sequenced species, and in the previously published genomes of *S. purpurea*, *S. babylonica*, *S. triandra*, *S. arbutifolia*, and on chromosome 15a of *S. dunnii* (Additional file 2: Table S10). Additional file 2: Table S11 lists genes found in the SLRs of all *Salix* species studied, including 21 single- or low-copy orthogroups shared by all species. The SLR boundaries of 15XY and 15ZW species changed independently, partly explaining the small number of orthogroups found within the SLRs [42], and gains and losses of genes have also occurred. Note

that these do not include sex-determining factor candidates, as they are frequently not in both SLRs (X and Y or Z and W) in willows, as discussed in the following section.

These 21 genes are a sample sufficient for reliable divergence estimates. As expected, despite the many changes documented in Additional file 2: Table S11, the Ks values for gametologous pairs are consistently low, with the highest value, in *S. polyclona*-E-TS, around 3% (Additional file 1: Fig S26a), with lower values in *S. gordejevii*, *S. polyclona*-E and W1 suggesting that recombination may have continued in these lineages, and might still occur. To further examine ZW divergence, we also estimated Ks values for 77 pairs of orthologous genes found in only the four 15ZW clade I species. Based on an estimated neutral mutation rate of 1×10^{-9} per year in *Salix* nuclear genes (Berlin et al., 2011), their age estimates also range widely, from 7.88 Mya for *S. polyclona*-W1 to 14.31 Mya for *S. polyclona*-TS, and several differences between the Ks values of different species are statistically significant (Additional file 1: Fig S26b, Additional file 2: Table S12). These results suggest that recombination stopped at different times in different lineages. Importantly, all the Z-W divergence time estimates within these species were shorter than the previously estimated divergence time (23 million years) for 15ZW clade I species [31,52]. It should be noted that as the latter's time was inferred from nuclear gene data using fossil-calibrated ancestral nodes and Bayesian posterior probability analysis, a direct comparison could be potentially misleading. In *S. polyclona*-W2, the divergence times between chromosome 15a (W-sex linked genes) and its 15b (Zb-sex linked genes), 15c (Zc-sex linked genes), and 15d (Zd-sex linked genes) homeologs were 17.58, 16.63 and 14.62 Mya respectively.

To test more rigorously for complete sex linkage, we estimated phylogenetic trees for these 21 orthogroups. Consistent with the conclusion that recombination occurred until recently, their topologies mostly show no consistent patterns (Additional file 1: Fig S27). Nevertheless, our sequencing of further species confirm the previously published conclusion [10,42] that three gametolog pairs have become completely sex linked, at times that depend on the 15ZW species analysed. For the *S. arbutifolia* Saarb15bG0048300 sequence and its orthologs, the topology

suggests that full sex-linkage was established before the divergence of the two 15XY species *S. triandra* and *S. arbutifolia*, and that the gene is also full sex-linked in the ZW taxa *S. purpurea*, *S. baileyi*, *S. polyclona*-W2, and *S. polyclona*-E-TS (Fig. 4a, set1). The *S. arbutifolia* Saarb15bG0044000 sequence and its orthologs are sex linked in this species and the 15ZW species (Fig. 4c). The third gene (Sapur.15WG058100 and its orthologs) is different; we found evidence for partial sex-linkage until after the two XY species diverged, but complete sex-linkage in the 15ZW species (Fig. 4e).

Including further species (set 2: *S. gordejevii*, *S. polyclona*-W1 and E, and the *S. babylonica* V homeolog) in the analysis revealed further details of the evolution of complete sex-linkage. For the first two genes, the *S. gordejevii* Z gametologs cluster with its W sequences and other species' W and X alleles, indicating partial sex-linkage before the 15XY clade speciated (Fig. 4b and Fig. 4d, Additional file 1: Fig S28a, b). Similarly, the Z gametolog of *S. gordejevii* clusters with its W sequences and other species' W alleles for the third gene, with recombination continuing after the speciation of the 15XY clade (Fig. 4f, Additional file 1: Fig S28c). The results from the *S. polyclona*-E and W1 lineages suggest that a 15W gametolog (labelled "new" in Fig. 4b, 4d and 4f) arose from a 15Z ancestor, not from a 15W gametolog like those in the set 1 species (the E and W1 individual's W sequence clusters with the Z ones from other species).

Although the results from the allotetraploid *S. babylonica* are difficult to interpret, as exchanges between homeologous sequences may have occurred, Saarb15bG0044000 and its orthologs tree supports its Z and W derived from the ancestral Y and X, respectively (Fig. 4d) [31]. Our interpretations are also supported by the finding that sequences of the ZW species' Z-linked alleles and the Y-linked alleles of 15XY species (including partial *ARR17*-like duplicates and the *MSF* identified in *S. arbutifolia* and *S. triandra* [42]), are detected in both the 15Z- and 15W-SLRs of the *S. polyclona*-E and W1 individuals, whereas, in *S. gordejevii*, they are only in the 15W-SLR (Fig. 4, Table 3, Additional file 2: Table S13). The results for the third gene, Sapur.15WG058100, also indicate incomplete sex-linkage in the 15XY and both sets of 15ZW species (Fig. 4f, Additional file

1: Fig S28c).

Overall, we conclude that only two genes have been fully sex-linked since before the split of the ancestor of the 15ZW taxa from the 15XY species (Saarb15bG0048300 and its orthologs became fully sex-linked before the divergence of *S. triandra*, *S. arbutifolia* and the 15ZW species; Saarb15bG0044000 and its orthologs became sex-linked before the divergence of *S. arbutifolia* and the other 15ZW species). Sapur.15WG058100 became fully sex-linked in an ancestor of the set 1 15ZW taxa, but remained partially sex-linked in the set 2 15ZW taxa and in the 15XY species so far studied, until a recent change led to complete sex linkage in the set 2 15ZW taxa. When we narrow our detection of fully sex-linked genes to fewer species, we are able to identify a greater number of such genes that cluster according to gamete type, further suggesting progressive/lineage-specific expansion of recombination suppression regions (Additional file 1: Fig S29).

Sex-determining factor candidates in 15ZW clade I willows

Given that the 15ZW *Salix* species appear likely to have arisen in a turnover event that created a single-gene female-determining factor in a clade I species ancestor, we hypothesised that this factor was duplicated into the ancestor's chromosome 15 pericentromeric region. If so, it can potentially be identified by searching for a gene present only in the W region assemblies, and not in their Z counterparts. We therefore identified the genes in both assemblies of each of the three new *Salix* species. In a given species, some genes have BLAST hits in only one haplotype or the other (Fig. 5a, Additional file 1: Fig S30, Additional file 2: Tables S14 and S15), and this reflects similar levels of gains and losses by both haplotypes, based on using the autosomal 15a haplotype of *S. dunnii* (7XY) as an outgroup (Additional file 2: Table S9), and not genetic degeneration of the W-linked region. Comparing all three species, no genes were overall Z-SLR-specific (shared by the Z-SLRs of all lineages investigated; Additional file 1: Fig S30). However, one sequence is shared by all the W-linked region assemblies (Fig. 5a). Strikingly, this shows homology to the *A. thaliana* B-class flower development gene *PISTILLATA* (*PI*); this gene was also found in the *S. purpurea* W-SLR (Fig. 5b, Additional file 2: Table S13), and is therefore W-specific. *PISTILLATA* is essential for male organ

(stamen) development in *A. thaliana*, and it acts downstream of the *ARR17*-like gene in *Populus*, which acts primarily by inhibiting downstream genes such as *LEAFY* (*LFY*) and *UNUSUAL FLORAL ORGANS* (*UFO*), which in turn suppress the maleness-promoting *PI* gene [36]. Therefore, the shared *PI* homolog sequences in the W-linked region may be a conserved W-specific female-determining factor in the 15ZW *Salix* species, and it was investigated further, as described below.

The intact *ARR17*-like genes involved in determining femaleness in *Salix* and *Populus* [29,33] are located on an autosome, chromosome 19, in all studied *Salix* species with male heterogamety, including 7XY and 15XY species, as well as in the 15ZW allotetraploid *S. babylonica* [10,31,42]. In the 15ZW clade I species *S. purpurea*, intact *ARR17*-like genes are present on chromosome 15W as well as chromosome 19 [43]. However, blastn searches (Additional file 2: Table S16) detected no intact *ARR17* genes in the genome of another 15ZW clade I species, *S. viminalis*; contrary to a reported intact *ARR17* gene on the 15W chromosome by Almeida et al. [64], we detected only a few incomplete *ARR17* exon fragments (exon1, exon3, or exon5) on the whole genome, suggesting that the numbers of these genes varies among these species; this confirms the recent *S. viminalis* genome assembly, which detected only a single intact *ARR17*-like gene, on chromosome 19, and none on chromosome 15 [65]. Consistent with these previous results, we found intact *ARR17*-like genes on both chromosomes 19 and 15 in our sequenced females of two 15ZW species, *S. baileyi* and the *S. polyclona* complex (in both ZW diploids and the 15ZZZW autotetraploid), and the chromosome 15 copies were within their SLRs (Fig. 5b, Table 3, Additional file 2: Table S13), in variable numbers. These results suggest that the number of intact *ARR17*-like genes varies among these species. The number ranges from a prematurely terminated copy in *S. gordejevii* (whose entire genome contains only one full-length *ARR17*-like gene copy on chromosome 19, bearing a premature termination codon in the fourth exon), to one copy on chromosome 19 in its close relative *S. viminalis* [65,66], to a small number in *S. baileyi*, and up to 15 copies in the newly assembled 15W of *S. polyclona*-W2 individual (Fig. 5b, Table 3 and Additional file 1: Fig S31). The high copy number of *ARR17* within 15W of *S. polyclona*-TS and W2 may simply reflect random gene amplification (Additional file 2: Table S8). The sequence presence/absence results are summarised in Table 3 and an interpretation of

these findings is discussed below. Phylogenetic analysis of the intact *ARR17*-like sequences (from 411 to 468 bp, spanning five exons), suggests that chromosome 19 was the ancestral location of the *ARR17*-like sequences now found on chromosome 15W (Additional file 1: Fig S32; the origin of the *ARR17*-like sequences on chromosome 15Z in the four *S. polyclona* taxa remains unresolved); consistent with a chromosome 19 origin, all willow species have one or two copies at this site (Table 3), but not on its homeolog, chromosome 13 [42].

Turning to the *PI*-like sequences on the 15W-SLRs, homology-based blastn analysis revealed that these sequences are incomplete compared with their intact autosomal homologs (Additional file 1: Fig S33, Additional file 2: Table S17), which, in all the species studied here, including *S. babylonica* [31], were found on the homeologous autosome pair 2 and 5 (Table 3; this pair originated in the WGD event mentioned above [42,51]). The intact *PI*-like copies on chromosomes 2 and 5 are collinear across all the 15ZW clade I species (Additional file 1: Fig S34). We detected partial *PI*-like duplicates in the 15W-SLRs of all the 15ZW clade I species studied here, but not in 15Y-SLRs of XY willows or the allotetraploid *S. babylonica* (Table 3, Additional file 2: Table S13). Searches of the entire genome found *PI*-like duplicates (~189 bp) only on chromosomes 15W, 2, and 5 (Additional file 2: Table S17). Phylogenetic analysis of the partial sequences (Fig. 5c) revealed that some duplicates in the 15W-SLRs of all lineages (except *S. purpurea*) cluster with intact sequences on autosome 2, and others with ones on autosome 5 (except *S. baileyi*), suggesting duplicative transpositions from both these autosomes. It is, of course, unlikely that *PI*-like duplicates from autosomes 2 and 5 could simultaneously duplicate to the SLRs and form inverted repeat elements. However, the sequences that form inverted repeats usually originate from a single autosome, either 2 or 5, suggesting that such repeats can readily form (Fig. 5c, Additional file 2: Table S17). The intact *PI* gene sequence includes both the MADS domain and the K-box domain, while the annotated partial copies have only the MADS domain (Additional file 1: Fig S33, Additional file 2: Table S18).

These conclusions are supported by ancestral character inference (see the Methods section). In the diploid ancestors of 15ZW clade I species, before the 15ZW species diverged, chromosome 15

apparently gained duplicated copies of the chromosome 19 intact *ARR17*-like sequences, and partial *PI*-like sequences (from chromosomes 2 and/or 5); these conclusions are supported by probabilities of 54% and 96%, respectively (Additional file 1: Fig S35).

Sex determination in members of the 15ZW clade I

Any of the genes known to act upstream or downstream of the intact *ARR17*-like genes might have the potential to evolve into new sex-determining factors. We can exclude the previously proposed *UFO* and *LFY* genes [36] as candidates for such changes in the 15ZW clade I species, because neither is located in the SLRs of any of the taxa studied here (Additional file 2: Table S19). The involvement of other genes was tested using studies of gene expression in buds in early developmental stages, at a stage that we confirmed is in the appropriate developmental stage (see Methods, Additional file 1: Fig S36). Furthermore, in flower bud and catkin tissues of these six taxa, *UFO* shows negligible expression levels, while *LFY* expression is similar in both sexes (Additional file 2: Table S19). The other two potential sex-determining factors of *S. purpurea*, dsRNA-binding domain-like superfamily protein (*DRB1*) and Argonaute family protein (*AGO4*) were also detected on the sex chromosomes of our sequenced genomes (Additional file 2: Table S19) [44]. However, they are found either in the PARs, or in both the Z- and W-SLRs, and no consistent expression differences were found between the sexes in flower buds or catkins (Additional file 2: Table S19), so they are also excluded from our candidates. We can also probably exclude a role for a member of the *GATA* gene family in sex-determination of most willows (except *S. purpurea*). Such a role was proposed because, in *S. purpurea*, a member (*GATA15*) is within the 15W-SLR, and (as outlined in the Introduction) might reinforce the *ARR17*-like gene's effect, suppressing male flower development and inducing female flowers [44]. However, no expression or experimental evidence directly suggests a function in sex-determination in this species [45], and, although we confirmed the presence of *GATA15* in *S. purpurea*, we did not detect a chromosome 15 copy in the other 15ZW species (Table 3).

The intact *ARR17*-like sequences in the 15W-SLRs of the species studied here are expressed at very

low levels in early developmental stages of flower buds and catkins (Fig. 5d), and we confirmed that when expression is detected, the sequences within their exons are not or low expressed [45] (see Additional file 1: Fig S37, Additional file 2: Table S20). Some of these sequences include premature stop codons or insertions changing the coding sequence (Additional file 1: Fig S38), which may account for their extremely low expression, or affect their function. Together with the copy number variation described above, and the complete absence of intact *ARR17*-like gene copies from the genomes of *S. gordejevii*, we suggest that intact *ARR17*-like genes may not be essential for female functions in the 15ZW clade I species studied here. In *S. purpurea*, expression of the sex-linked (15ZW) *ARR17* sequence is female-specific, but it is very low [44]. However, expression differences between the sexes are quite striking in catkins [43]. Studies of expression in early flower bud development are therefore still needed to test whether the intact copies in this species function in the sex determination pathway, as would be expected. In *S. baileyi*, the intact *ARR17* gene is specifically expressed at low levels in female flower buds (Fig. 5d, Additional file 1: Fig S37). This expression pattern may resemble that observed in *S. purpurea*. Hence, experimental validation is required to determine the functional role of 15W-linked intact *ARR17*-like copies, given that *ARR17* expression may in fact be more challenging to quantify experimentally than currently appreciated.

However, the previous result that no small RNAs map to the *ARR17*-like genes or their nearby regions [44] is consistent with our observations that the partial *ARR17*-like duplicates produced almost no detectable small RNAs in either female or male flower buds of the studied ZW willows (Additional file 2: Table S21). This suggests that *ARR17* RNA-interference no longer occurs in the 15ZW clade I species, and that their partial *ARR17*-like duplicates, although often still found on the Z chromosomes in this clade, may have lost any sex-determining function. Using one of the intact *ARR17* on chromosome 19 of *S. triandra* as reference, we calculated divergence of all intact *ARR17* copies in 15ZW clade I species and *S. arbutifolia*, indicating accumulation of more non-synonymous substitutions in 15ZW clade I species (Additional file 2: Table S22). This may suggest that the intact *ARR17* copies in the 15ZW clade I species have undergone functional changes, consistent with the previous transcriptome data.

Instead, the partial *PI*-like duplicates in the *Salix* 15ZW clade I might function as femaleness-promoters. In several of these species, expression of the autosomal intact *PI* homolog is high in male flower buds and catkins, whereas hardly any expression was detected in female buds and catkins (Fig. 5e, Additional file 2: Table S18). We therefore propose that low *PI* expression in female buds and catkins in *Salix* is likely to be due to the presence of the multiple W-linked partial copies.

In our genome assemblies, the partial *PI*-like duplicates form inverted repeat elements within the W-SLRs of all 15ZW clade I species sequenced, suggesting that they might form hairpin structures through complementary pairing of the arms (Additional file 1: Fig S39-S45), yielding double-stranded RNA (dsRNA) intermediates that (like the partial *ARR17*-like genes described above) can be processed into small interfering RNAs (siRNAs) that reduce expression of the autosomal intact *PI* genes. Supporting this hypothesis, we found a large accumulation of 22-24 bp sRNAs with sequences homologous to regions near and within these tandem duplicate arrays of partial *PI*-like sequence (Fig. 6a; we also detected sRNAs with sequences homologous to those in and around the intact *PI*-like duplicates on chromosomes 2 and 5, as shown in Additional file 1: Fig S46, probably reflecting degraded transcripts from intact copies). In contrast, the partial *PI* sequences showed very low coverage in mRNA sequences (Additional file 1: Fig S47, Additional file 2: Table S18), consistent with functioning through siRNA-mediated silencing.

To further test our hypothesis, we performed dual-luciferase reporter assays in *Nicotiana benthamiana*. As the intact *PI*-like duplicates and siRNA sequences are similar in all 15ZW clade I species, sequences from a single species, *S. baileyi*, were used for these experiments. When the luciferase-fused intact *PI*-like sequence was co-expressed with siRNAs derived from partial duplicate sequences (in leaf regions containing both elements), luminescence ratios decreased, compared with controls (*t-test*, $p < 0.0001$), indicating that these siRNAs can indeed silence intact *PI*-like gene expression (Fig. 6b, Additional file 1: Fig S48), as hypothesised. We conclude that the siRNAs produced from the W haplotypes in ZW or ZZZW females probably inhibit expression of

intact autosomal *PI* copies that promote male functions (Fig. 6c), while these can function normally in the ZZ or ZZZZ males. The *PI* gene's crucial role in stamen development is well-documented, but there is currently no reported evidence demonstrating that its expression inhibits pistil development. Further studies are needed to (1) directly validate the function of intact *PI*-like genes in willows, and (2) investigate whether other W-specific factors, similar to *GATA15* in *S. purpurea*, can promote pistil development in other 15ZW species, or whether a factor that inhibits stamen development will pleiotropically increase female flower part development, so that a trade-off occurs, such that the genotype that is not male automatically develops as a female [67].

Discussion

Our results and conclusions are summarised in Fig. 7. As reviewed in the introduction, the ancestral system in both poplars and willows with male heterogamety, in both the 7XY and 15XY, groups involves partial *ARR17*-like duplicates in Y-SLRs which generate small RNAs silencing intact autosomal *ARR17*-like genes whose expression determines female development; their silencing should lead to activation of intact *PI*-like genes, determining male development [31,36,42].

In 2020, Wang et al. described evidence that the appearance of intact female-promoting *ARR17*-like genes on chromosome 15, by duplicative transposition, led to the 15XY→15ZW sex chromosome turnover in *S. purpurea* [29], with an increase in the copy number of these sequences in the pericentromeric repetitive 15W-SLR promoting female development, changing this genome region to a W-linked region, similar to a change documented in sex-determination in *Populus alba* [30,33]. Willows may possess a take-over mechanism similar to that in *P. alba*, in which the intact *ARR17* gene becomes the primary sex-determining factor, replacing the ancestral partial *ARR17*-like duplicates. Such a change would initially have created a system with a W-linked region as well as the ancestral X- and Y-linked ones, similar to the situation observed in the platyfish, *Xiphophorus maculatus* [68].

The allotetraploid species supports this interpretation. In the species tree, the *S. babylonica* V sub-genome is paraphyletic to other ZW species in the *Vetrix* clade (Fig. 1, 7), and its 15W-SLR is similar to its ancestral X-SLR, having no intact *ARR17*-like or partial *PI*-like duplicates. The eight intact *ARR17*-like duplicates on its autosome 19 copies can evidently outweigh the maleness-promoting action of the 15Z-SLR partial *ARR17*-like duplicates, to produce XY females.

A new female-determining function then appears to have evolved in the W haplotype of the chromosome 15 sex-linked region, involving partial duplication of *PI*-like sequences (from chromosomes 2 or 5), probably creating a stronger female-determiner. This was followed by loss of the male-determining function of the ancestral 15Y-linked *ARR17*-like system (Fig. 7). This order of appearance of sequences in the 15W-SLR is supported by synonymous divergence (*K*s estimates); pairwise divergence from intact chromosome 19 *ARR17* sequences (based on 36 sequence pairs) is about four times higher than that between partial *PI*-like duplicates on chromosome 15W and chromosome 2 or 5 (26 pairs). As Additional file 1: Fig S49 shows, the difference is highly statistically significant, despite the short sequence lengths, especially those of the partial *PI*-like sequences, which include only exon 1.

The new system involving partial *PI*-like sequences resembles the ancestral Y-linked *ARR17*-like system, as both involve small RNAs controlling an autosomal gene, *PI*, though the default sex (the sex that develops in the absence of the sex-determining "trigger" factor) differs in the two cases. The involvement of small RNAs is increasingly recognised in both animal and flowering plant sex determination [33,38,69,70], including a W-linked PIWI-interacting RNA (piRNA) that functions as feminising factor in the silkworm [70]. In the plant genus *Diospyros*, the male-specific (Y-linked) *OGI* sequence is a duplicated gene whose inverted repeat forms an RNA hairpin structure, which produces a small RNA that silences its autosomal progenitor gene, *MeGI* [69], which is probably involved in controlling allocation to male versus female flower structures [8]. The *ARR17*-like and *PI*-like genes may participate in such an allocation system.

Under the interpretation above, new sex determining factors have evolved at least twice in *Salix* from the ancestral Salicaceae XY system. In *S. purpurea*, the *GATA15* gene may promote female development in females, alongside the female-promoting function of intact *ARR17*, which also causes suppression of male development. In the 15ZW clade I *Salix* species studied here, the maleness suppressing function, creating females, appears to have been replaced by suppressive effects of partial *PI* sequences. These observations carry the interesting implication that sex-determination in both cases involves trade-offs between female and male functions, such that the development of one sex phenotype leads to loss of structures with the other sex functions. It is also interesting that, in both cases, the new factor appears to have evolved from a gene that acts downstream in the sex-determination pathway, not by adding a new sex determining gene upstream of the ancestral sex-determination pathway, as Wilkins suggested [71]. There seems no reason why this should not occur, and another example was outlined in the introduction. In the 15ZW clade I *S. polyclona* complex individuals, with sex-linked intact *ARR17*-like gene copies, these sequences have likely transferred from their initial 15W-SLR to the 15Z counterpart. The XY systems also underwent changes in *Salix*. Among the ancestral 7XY species, 7Y chromosomes were lost in two species, and in the *Salix* clade, new 7Y chromosomes probably evolved from the ancestral 7X [42]. An XY→ZW change in *Populus* is also documented [30]. These hypotheses can be further tested in the future in basal *Vetrix* clade species, e.g., *S. reticulata* and *S. setchelliana* [72].

The 15XY→15ZW sex chromosome turnover studied here involves transposition, followed by duplications of intact *ARR17*-like genes (as in *S. purpurea* [29]), and a subsequent turnover in the *Vetrix* 15ZW clade I, involving partial duplications onto the 15W chromosome of a *PI*-like gene (Fig. 6), taking over the female-determining function. It remains unclear whether the *ARR17*-like copies retain functionality, as gene expression was assessed at only a single stage of floral development in each sampled species or lineages. A very recent report (preprint) found a similar mechanism in 15ZW clade I *S. herbacea* [73]. These changes suggest that effects of dosage of genes involved in

sex determination are important in the poplars and willows, probably in the development of stamens and pistils during flower development. This suggestion is consistent with the idea of trade-offs between allocation to male and female functions, because trade-offs are likely to involve dosage-dependence, with greater expression of some genes promoting maleness, and leading automatically to lower expression of female functions.

Recent changes, such as those within the *Vetrix* sub-clades of genus *Salix*, create new sex chromosomes, and some of the SLRs appear to still be capable of recombining, as the W-Z Ks values are low. In *S. gordejevii*, there is evidence for Z-W recombination (see Fig. 4). These SLRs may be too young to have degenerated or differentiated, or occasional recombination may have prevented degeneration and maintained homomorphic, as in some frogs [74].

A common feature of all these systems is that sex-determining loci are located within pericentromeric regions of the chromosomes involved. These regions probably recombine rarely (based on genetic evidence in *S. arbutifolia*, *S. triandra*, *S. babylonica*, *S. mesnyi*, and the newly studied species [10,31,42] (Fig. 3a), and on their high repetitive content in all species where genome assemblies are available to test this. The lack of major degeneration of the SLRs, but evidence of small proportions of genes being missing from both the W- and Z-linked regions, is consistent with genes in repetitive pericentromeric genome regions being expressed at low levels, and being prone to gene losses and gains [75,76]. The *S. gordejevii* W-SLR is smaller than the Z-SLR (Fig. 2), consistent with occasional recombination preventing accumulation of repeats. However, the S. *arbutifolia* and *S. triandra* X-linked regions are larger than the corresponding older-established Y regions, due to the accumulation of repetitive sequences, as expected if recombination is rare [42], and most of the 15ZW species derived from XY ones, have SLRs that exhibit micro-heteromorphism, with larger W- than Z-SLRs (Fig. 2).

The lack of degeneration (Additional file 2: Table S9) suggests that the changes documented here are unlikely to have been favoured by loss of genes from Y-linked regions in the ancestral species with

male heterogamety [77,78]. In the Introduction, we mentioned the possibility that movements create linkage to a sexually antagonistic polymorphism, but other possibilities have also been modelled [79]. One is that high TE density may increase the rate of movements of genome regions into and out of these regions, including movements of genes. If a gene involved in the sex-determining pathway moves into such a region, its copy number may increase through tandem duplication (another process that is facilitated in rarely recombining genome regions). If the gene acts in a dosage-dependent manner during development, this might cause a change in the sex that develops. Such changes might affect the sex ratio of a population, resulting in a sex chromosome turnover, and/or in compensatory evolution to restore a 1:1 sex ratio [80]. Such sex ratio selection is one of the main causes proposed for sex chromosome turnovers [80]. Unequal sex ratios can arise for many different reasons, but we speculate that the large numbers of turnovers in willows might be related to their chromosomes' large pericentromeric regions. Turnovers involving different chromosomes carrying the sex-determining locus are also documented in the genera *Silene* and *Rumex*, both of which share the property of having large pericentromeric regions [81–84]. However, testing this possibility requires analyses of more taxa.

Conclusions

In *Salix*, the sex-linked regions in the 15ZW taxa studied here evolved within the rarely recombining pericentromeric region and show little or no degeneration. Our interpretation of their evolution involves the duplication and translocation of sex-determining factors, including *ARR17* and *PI* sequences. Partial *PI* duplicates were found exclusively in the 15W-SLRs of 15ZW clade I species, and are absent from the SLRs of previously studied *Salix* species with 7XY and 15XY systems, and from the 15ZW clade II species. The siRNA sequences transcribed from the newly identified partial *PI* duplicates can directly inhibit the expression of intact autosomal copies of a *PI*-like gene, acting as a dominant femaleness-determiner. This change followed transposition of intact *ARR17* sequences to an ancestral 15X-linked region, that likely triggered transitions of XY to ZW. All these events did

not alter the location of the sex-determining regions in pericentromeric regions of the genomes, consistent with the known tendency for genes in this genomic compartment to be gained and lost.

Methods

Plant materials

We collected leaves from females of *S. gordejevii* (HL00135) and *S. baileyi* (HL00025) for whole genome sequencing. To represent the diversity in the *S. polyclona* complex, our material includes one autotetraploid individual, West 2 (W2), and three diploid female individuals from populations named West 1 (W1), East (E), and East Taishan (E-TS) [48].

Catkin, flower bud, young leaf, stem, and root samples from female plants, and male catkin and bud samples, were collected for DNA or RNA sequencing; flower buds of *S. gordejevii* for RNA sequencing were collected in October 2024, for *S. baileyi* in September 2024, for *S. polyclona*-E in April 2024, for *S. polyclona*-E-TS in March 2024, for *S. polyclona*-W1 in April 2024, and for *S. polyclona*-W2 in March 2024. The flower buds were identified based on their morphological characteristics and positions on branches [85]. Samples were frozen in liquid nitrogen and stored at −80°C until total DNA or RNA extraction. Pistillate and staminate flower bud development were studied in the same *Salix baileyi* individuals as used for transcriptome samples, using OCT-embedding techniques. At the developmental stage sampled for transcriptome studies (see below), the pistillate and staminate buds or inflorescences were fixed in OCT compound, followed by vacuum treatment, and stored at -80°C. Sectioning was performed using a cryostat microtome (Leica CM30508). The sections were then observed and photographed under a LEICA M205 FA optical microscope and LAS v4.13 software.

For analyses including variant frequency comparisons between the sexes based on Illumina sequencing, we collected leaves of 19 male and 19 female of *S. gordejevii*; for the *S. polyclona*

complex, we collected leaves from ten males of E-TS, three females and 14 males of E, 14 males of W1, and nine males of W2. These samples were dried in silica-gel for whole genome resequencing and ploidy estimation. All the plant materials we used were identified morphologically based on published references (*S. gordejevii* [86], *S. baileyi* [87], *S. polyclona* [48,88]). Voucher specimens of the newly sequenced samples are deposited in the herbarium of Shanghai Chenshan Botanical Garden (CSH). We also downloaded genome sequence data of four E-TS, 22 W1, 19 E, nine W2 individuals of the complex published by [48], and retrieved *S. baileyi* resequencing data from NCBI [52]. Detailed information of these datasets is in Additional file 2: Table S23.

Ploidy determination

The ploidy levels of five females, five males of *S. gordejevii* and 31 *S. polyclona* complex individuals were estimated by flow cytometry, followed the protocol in ref. [89], using diploid *Salix dunnii* (2x = 2n =38 [51]) and *S. baileyi* (2x = 2n = 38 [90]) as external standards, respectively. Chopped leaf tissue was incubated for 80 min in 1 mL LB01 buffer and then filtered through a 38-μm nylon mesh. The filtered homogenate was treated with 80 μg/mL propidium iodide (PI) and 80 μg/mL RNase to stain the nuclei. Estimates used a MoFlo-XDP flow cytometer and Summit v.5.2 (Beckman Coulter Inc.). The ploidy level was calculated as sample ploidy = reference ploidy × mean position of the sample peak/mean position of reference peak.

We used Jellyfish [91] to construct *k-mer* frequency distributions of tetraploid *S. polyclona*-W2 (HL00183) (*k-mer* = 21). Genomescope 2.0 [92] was used to distinguish autotetraploidy from allotetraploidy based on progeny genotypes. This approach is based on the fact that, in allotetraploids, preferential pairing of homologous chromosomes from the same subgenome during meiosis produces high proportions of progeny with two alleles from each subgenome (aabb), whereas in autotetraploids whose chromosomes pair randomly in quadrivalents, progeny can inherit one allele from one given subgenome and three from the other (aaab).

Genome sequencing

PacBio HiFi libraries for long-read sequencing and genome assembly were prepared from genomic DNA extracted from *S. gordejevii*, *S. baileyi*, and the four *S. polyclona* complex individuals using a CTAB method modified and optimised by Novogene to obtain high quality DNA from woody plants for such sequencing. PacBio large insert libraries were then created by SMRTbell Express Template Prep Kit 2.0 and sequenced by Novogene using the PacBio Sequel II platform with the Circular Consensus Sequencing (CCS) model. To improve accuracy of the assemblies, we also performed ONT sequencing for autotetraploid *S. polyclona*-W2, using DNA extracted using the phenol–chloroform method. ONT libraries were prepared following the Nanopore 1D Genomic DNA by ligation protocol, and sequenced by Novogene on the PromethION platform.

Hi-C libraries were constructed using standard procedures [93]. Tender leaves from *S. gordejevii*, *S. baileyi*, and four lineages of the *S. polyclona* complex were used for library preparation. The leaves were fixed in a 4% formaldehyde solution, and then the cross-linked DNA was isolated from nuclei. Subsequently, DNA was digested with the restriction enzyme MboI, and restriction fragment ends were biotinylated, purified and ligated. Hi-C libraries were controlled for quality and sequenced on Illumina HiSeq X Ten platform and NovaSeq 6000 by Novogene.

We obtained short-read sequences from 38 *S. gordejevii* individuals, and 50 individuals (41 diploids and 9 tetraploids) of the *S. polyclona* complex, using genomic DNA extracted from leaves with the Magnetic Plant Genomic DNA Kit (Tiangen, China). Paired-end libraries were constructed for all samples. Whole-genome sequencing with expected depths of 20× for the diploids and 80× for the tetraploids was performed on Illumina NovaSeq 6000 and NovaSeq X Plus by Beijing Novogene Bioinformatics Technology.

RNA sequencing

Total mRNA was extracted from samples from *S. baileyi*, *S. gordejevii* and four lineages of the *S.*

polyclona complex using a Polysaccharide Polyphenol Plant Total RNA Extraction Kit (Tiangen, China). The RNA integrity was assessed using the Fragment Analyzer 5400 (Agilent Technologies, CA, USA), and libraries were generated using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA).

We also performed small RNA sequencing for female and male flower buds from all species studied, using total RNA extracted with RNAprep Pure Plant Plus Kit (Tiangen, China). After adaptor ligation to the ends of the small RNA molecules, reverse transcription primer hybridization was used to create the first strand cDNA. Double-stranded cDNA libraries were created by PCR enrichment, and libraries with insertions ranging from 18 to 40 bp were purified and sequenced. Both mRNA-seq and small RNA-seq sequencing were performed on an Illumina NovaSeq 6000 by Novogene.

Genome assembly

We used a similar strategy to assemble and annotate the genome assemblies of all the species. Initial contigs were assembled based on PacBio HiFi reads using hifiasm [54]. For the autopolyploid *S. polyclona*-W2, ONT reads were also used to assemble initial contigs. Hi-C reads were then aligned to the haplotype contig genomes using Juicer [94], for chromosome-level genome assembly. Preliminary Hi-C-assisted chromosomal assemblies were completed using 3d-dna [95]. Subsequent manual inspection and correction using Juicebox [96] refined the chromosome boundaries, eliminated improper insertions, modified orientations, and corrected assembly errors. Finally, LR_Gapcloser [97] gap filling based on HiFi readings was implemented to improve the assembly. GetOrganelle [98] was used to assemble chloroplast and mitochondrial genomes.

The fragmented contigs were aligned to the chromosome-level genome and organelle genome sequences with Redundans [99]. rDNA fragments and low-coverage fragments or haplotigs among the dispersed sequences were eliminated. Furthermore, our short-read data was polished with Nextpolish [100] to correct bases.

To improve our haplotype-resolved chromosome assemblies, reads mapped near the telomeres were selected and assembled into contigs using hifiasm, and then mapped back to the chromosomes. The chromosomes were numbered from 01 to 19 following the homologous *S. arbutifolia* chromosomes [10]. We obtained two haplotypes (a, b) for the diploid species, and four (a, b, c, d) for the tetraploid *S. polyclona*-W2.

To assess the accuracy of the sex chromosome phasing and detect any chimeric regions in the assemblies, unitig sequences (unique contiguous sequences inferred as high-confidence, branch-free paths in the assembly graph) were mapped to the assembled genomes using minimap2 (-x asm5) [101]. The SLRs and PARs described in the Results section were identified and highlighted in unitig graphs visualised using Bandage [102].

Genome annotation

Coding gene prediction combined evidence from homology-based prediction, transcript prediction, and *de novo* prediction strategies. For homology-based prediction, we used the publicly available Salicaceae protein sequences as homologous protein evidence for gene annotation. For transcript prediction, we assembled transcripts from our mRNA data using Trinity [103] and StringTie [104], and removed redundant sequences with CD-HIT [105] (with settings identity > 95%, and coverage > 95%). The PASA pipeline [106] was used to annotate gene structures based on our transcriptome data, and full-length genes were identified by aligning with their homologous proteins. These full-length genes were used for AUGUSTUS [107] training with five optimisation replicates. Gene structures were then predicted by the MAKER [108] pipeline based on the repeat-masked genome with *de novo* prediction, transcript and homolog protein evidence. Next, the MAKER and PASA annotation results were integrated to produce consensus gene sets using the EVidenceModeler (EVM) gene structure annotation tool [109]. Finally, untranslated regions (UTRs) and alternative splicing were annotated using PASA [106], and genes encoding fewer than 50 amino acids were removed, plus sequences with internal stop codons or ambiguous bases, or no start or stop codon in the assembled annotation. Non-coding RNAs (ncRNAs) were annotated using tRNAScan-SE [110], RfamScan, and Barrnap

pipelines（https://github.com/tseemann/barrnap).

Gene functions were annotated based on homology and similarity searches. In addition to the eggNOG-mapper [111] annotation, DIAMOND [112] was used for sequence similarity searches (identity > 30%, E-value < 1e-5), and protein databases including Swiss_Prot, TrEMBL and NR. The InterProScan software [113] was used to search for domain similarity. BUSCO was used to assess genome completeness, using the embryophyta_odb9 database.

Repeat elements were identified by EDTA [114] (--sensitive 1 --anno 1), and RepeatMasker (http://www.repeatmasker.org/RepeatMasker/) was then used to determine repetitive regions within our assemblies. We used Tandem Repeats Finder (v4.09) to identify tandem repeats [115], and detect the telomeric regions of chromosomes [56].

Phylogenetic analysis

To understand the relationships of the species whose sex-linked regions are investigated here, we performed a phylogenetic analysis of autosomal sequences from 12 willow genome sequences, including the *a* haplotypes of *S. dunnii*, *S. mesnyi*, *S. triandra*, *S. arbutifolia*, *S. purpurea*, the four *S. polyclona* samples, W2, W1, E-TS and E, *S. baileyi* and *S. gordejevii*, and the *S. babylonica* S and V genomes, with *Populus trichocarpa* as an outgroup (Additional file 2: Table S7). Single-copy homologous protein sequences were identified using OrthoFinder [116]. The inferred protein sequences were aligned using MAFFT [117], before using the nucleotide sequences to estimate gene trees with IQ-TREE (-m MFP -bb 1000 -bnni -redo) [118], and then the species tree, using ASTRAL [119] according gene trees.

Read mapping and variant calling

Fastp [120] was used to filter all sequence reads, and clean reads were used for subsequent analysis. The BWA-MEM algorithm from bwa 0.7.12 [121,122] was used to align clean reads to each genome (both haplotypes for diploid *S. gordejevii*, *S. polyclona*-W1 and E, with default parameters; for the

W2 individual, we used the four haplotypes as references for calling SNPs, for consistency with the methods for the diploids, as described previously [48]. Samtools 0.1.19 [123] was used to filter and convert the mapped data to BAM. PCR replicates were filtered using sambamba 0.7.1 [124]. GATK 4.2.2.0 version was used to call variants in our species, with "--sample-ploidy 2". Hard filtering was carried out with parameters "QD<2.0, FS>60.0, MQ<40.0, MQRankSum<–12.5, ReadPosRankSum< –8.0, and SOR>3.0". Only biallelic sites were used in subsequently filtering steps. Due to the potential for false positives and artifacts, sites with coverage above twice the mean depth at variant sites across all samples were discarded (DP > 2 * average DP) [125,126]. To control quality, sites where the samples have average depth < 4× were treated as missing, and sites missing in> 10% of the samples (using GATK) or with minor allele frequency <0.05 (using Samtools) were filtered out, as described previously [10,31,42,51].

Identification of the SLRs

We identified the SLRs of the *S. gordejevii*, *S. baileyi*, and the four *S. polyclona* complex lineages using a combination of CQ, *k-mer*, F_{ST}, the location of inversions, and synteny between the regions identified. The CQ, F_{ST} and *k-mer* analyses are based on 64-82 million clean Illumina reads per individual (mean 70) of *S. gordejevii* (average read depths from 25.9× to 33.9×), and 65-80 million clean Illumina reads per individual (mean 70) of *S. baileyi* (average read depths from 21.4× to 26.9×). For *S. polyclona*, we obtained 61-83 million clean Illumina reads per individual (mean 66) of the 41 newly sequenced diploids (average read depths from 19.3× to 28.7×), and 253-263 million reads per individual (mean 256) of the nine tetraploids (average read depths from 80.8× to 87.2×) using the relevant lineages as the reference genome.

Previous studies [32,48] suggest that the 15ZW clade I species have a similar female heterogametic system. The CQ method [127] was used to detect Z and/or W sex chromosomes using cq-calculate.pl software based on clean reads. We calculated the CQ for each 50-kb non-overlapping window across the chromosomes based on the combined female and male clean read data sets. Under female heterogamety, the CQ is the normalised ratio of male to female alignments, and the value should be

close to 2 in windows in Z-linked region and zero in windows in the W-linked region. The Changepoint package [128] was used to detect the boundaries of the SLRs based on CQ values between the sexes.

Weighted F_{ST} values between the sexes were calculated in 100 kb windows and 10 kb steps across chromosome 15, using Weir and Cockerham's [129] estimator implemented in VCF tools, for 5 of the 6 taxa sampled (the exception was the E-TS lineage, which is a very small population from the top of Tai Mountain, where only 34 individuals have been recorded [55] from which we obtained re-sequencing datasets from only four female and 10 male individuals [48]). We used the changepoint package [128] to detect significant differences in F_{ST} values.

The SLR regions identified using the above approach were further validated through *k-mer* analysis, following the methodology described by Carey et al. (2024). We used Jellyfish (v2.3.0) to extract 21-mers from each individual [91], and the Unix 'comm' command to identify sex-specific *k-mers*. These kmer-mers were then mapped to the haplotyped assemblies using BWA-MEM with the parameters -k 21, -T 21, -a, and -c 10 [121]. Mapping results were visualised by calculating coverage in 50,000-bp sliding windows using bedtools [130], and the coverage profiles were plotted in R using ggplot2 [131].

The SLRs in each species' 15Z- and 15W-linked regions were analysed for synteny, using the Python version of MCScan [132] with parameter "-- cscore=.99". Inversions surrounding the SLRs were incorporated into the final boundaries of the SLRs.

Analysis of evolution of the sex-linked region

We calculated the density of gene and TE in 50 kb windows across the autosomes and sex chromosomes in the *Salix* species studied here. The SLRs exhibited high repeat content, particularly of gypsy and copia elements, and low gene density. In plants, the CENH3 serves as an epigenetic hallmark of centromere position and function [61,62]. Consistent with this, Wang et al. [60] reported

that likely centromere-associated repeats, CRMs, based on presence in the known (rough) chromosomal locations of the centromeres are highly enriched for CENH3 in three Salicaceae species, consistent with their being components of active centromeric chromatin. We therefore characterised CRM elements in our genome assemblies to assess whether they can be used in other Salicaceae species. As CRM elements belong to a gypsy-type LTR retrotransposon lineage, we further used TEsorter software [133] to identify CRM repetitive sequences and calculated their genomic density in 100 kb windows across the autosomes and sex chromosomes in the *Salix* species studied here. In addition, higher LD is expected in pericentromeric regions, we hence used LDBlockShow v.1.40 [134] to calculate and visualise LD patterns between SNPs for relevant sex chromosomes and autosome 15.

We then used CRM-rich regions, together with patterns of repeat and gene densities, extended linkage disequilibrium (LD) blocks, to delineate putative pericentromeric regions in each genome. As expected for pericentromeric regions, those including the inferred centromeres showed high repeat density, low gene density, and LD blocks signal. In the Results section "Identification of sex-linked regions in the new genome assemblies and evidence that they are pericentromeric regions", we refer to these as pericentromeric regions and show that the sex-linked regions are largely overlapped with them, even though their boundaries are not sharp. Based on these combined signals of high repeat, low gene density, LD blocks signal, and high CRM density [60], we systematically identified likely pericentromeric regions in the six sequenced genomes and in the homologous autosome 15 of *S. mesnyi* and *S. dunnii*, and characterised the different types of repetitive sequences in them.

For each of the six new *Salix* genome assemblies, we separately identified presence or absence of genes in the W- and Z-SLRs using BLASTP with e-value 1e-5, in order to infer gains, losses and sharing for each species' SLR or PAR, as described previously [17,26], using the autosomal 15a haplotype of *S. dunnii* (7XY) as an outgroup. These genes were annotated in TAIR (https://www.arabidopsis.org/) to predict possible biological functions. We visualised the shared and unique gene sets using upset diagrams generated in TBtools [135]. For each species, we conducted

BLASTP searches within regions—Z-SLR vs. Z-SLR and W-SLR vs. W-SLR (identity > 30%, e-value < 1e−5). Based on these alignment results, we counted duplicate genes separately for the Z-SLRs and the W-SLRs.

The degeneration rate of sex chromosome was estimated according to the number of shared gene and specific gene between Z-SLR vs. W-SLR or X-SLR vs. Y-SLR. For the specific genes, we calculated the gain and loss of genes as follows: specific genes with hits to outgroup autosome 15a haplotype of *S. dunnii* (7XY) in 15Z-SLR (or 15W-SLR) were categorised as 'W (or Z) deletion', respectively; specific genes without hits to autosome 15 were classified as 'Z (or W) insertion'. Finally, we calculated degeneration rates based on gene loss situation. Degeneration rate for W-SLR = W-SLR loss / (W-SLR loss + Z-SLR loss + their shared genes), and degeneration rate for Z-SLR = Z-SLR loss / (W-SLR loss + Z-SLR loss + their shared genes). We performed a similar analysis for genes in PARs. The same method was used to calculate the degeneration rate in 15XY SDS species.

Sex chromosomes may stop recombining at several different times, initiating accumulation of variants in sequences specific to the sex-limited chromosome (the Y or W). The resulting divergence from the X (or Z) creates distinctive evolutionary topologies [136], and species that share a common ancestral SLR will share separate clusters of X and Y (or W and Z) alleles, termed gametologs, allowing a lack of recombination before the species diverged to be detected, whereas the sequences of each species will cluster together if recombination occurs [10,137]. In *Salix,* previous studies found two such genes shared by XY and ZW species [10], and seven genes shared only by ZW species, with 15Z and 15W sex-linked sequences clustered by gametologs, while their orthologous 15X- and 15Y-linked sequences cluster by species [42]. We used OrthoFinder [116] to identify sex-linked orthogroups in different species sets (all 15ZW clade I species and 15XY species *S. arbutifolia* and *S. triandra*; 15ZW clade I species only; *S. polyclona* TS and W2) using the *Salix* clade species *S. dunnii* as outgroup. We used MAFFT [117] to align the genes, and IQ-TREE [118] to reconstruct phylogenetic trees using the *S. dunnii* haplotype a of chromosome 15 as an outgroup.

To detect inversions and structural variants in SLRs, we also conducted synteny analysis between the sex chromosomes of the newly sequenced samples and the homologous autosomes of *S. dunnii* and *S. mesnyi*, using the Python version of MCScan [132] with parameter "-- cscore=.99".

Genes proposed as sex-determining factors and phylogenetic analysis of their sequences, and reconstruction of ancestral states

Intact *ARR17*-like genes (with all 5 exons intact) and partial *ARR17*-like duplicates (including fewer than 5 exons) were identified in the assemblies of the newly sequenced samples using the same strategy as in previous studies of other willows [10,29,31]. We used BLASTN ("-evalue 1e-5 -word_size 8") to identify orthologs of the *Populus trichocarpa ARR17*-like gene (Potri.019G133600). To ensure that all regions of the *ARR17*-like duplicates were identified, we extracted the aligned regions plus 200 bp of upstream and downstream sequence, and aligned them from the start codon ATG, using Geneious Prime 2023.2.1 (https://www.geneious.com/), which was also used to visualise the alignments. The coding sequences (CDS) of intact *ARR17* genes were used to construct a phylogenetic tree with IQ-TREE [118], using *P. trichocarpa* as an outgroup.

Using the BLASTN approach, we also identified sequences of the other genes mentioned in the Introduction: the male specific factor (*MSF*) [42], which is shared by the Y-SLRs of both 7XY and 15XY species, and the genes *UFO*, and *LFY* that act downstream of intact *ARR17*-like genes in the sex development pathway (*UFO* and *LFY* query sequences were the *P. trichocarpa* Potri.001G160900 and Potri.015G106900 genes), and the *DRB1* and *AGO4* genes, which have been proposed sex-determining candidates in *S. purpurea* (the query sequences from that species were Sapur.15WG074300 and Sapur.15WG074400 [44]).

Finally, we identified intact and partially duplicated sequences homologous to the *PI* gene, whose sequence is shared by all ZW species studied (see Fig. 5)**.** We used the *P. trichocarpa* sequence (Potri.002G079000, with 7 exons) as the query. To reconstruct a phylogenetic tree including both partial and intact *PI*-like duplicates, we analysed only the first exons (as explained above for the

ARR17-like sequences); this exon is around 189bp [36].

To trace the evolutionary history of the potential sex-determining genes that were not excluded from consideration (see the Results section), we used the willow species tree (Fig. 1a) with gene presence or absence as character states. Reconstruction of ancestral character states was performed in Mesquite v.3.81 [138], employing both character tracing based on the phylogenetic tree and the likelihood method for ancestral state reconstruction.

Gene expression analyses, including small RNAs

We estimated gene expression (excluding non-mRNA transcripts) in both female and male catkins and flower buds of the six newly sequenced taxa, with three biological replicates for each sex (Additional file 2: Table S23). After filtering, clean transcript reads were aligned to the genome assemblies with HISAT2 [139], respectively, and then featureCounts [140] was used to calculate the number of reads mapping to each gene. These counts were converted to TPM (transcripts per million reads).

To analyse and identify small RNAs in the six newly sequenced taxa, we used sRNAminer v1.1.2 [141], with three biological replicates of bud tissues from females and males in each taxon (Additional file 2: Table S23). SRNAanno database [142] was used to identify sRNAs, after removing adaptors, non-coding RNA including rRNA, tRNA, snoRNA, snRNA, and plasmid contamination from the sRNA-Seq datasets. The clean reads were aligned to their respective reference genomes with sRNAminer, and we applied IGV-sRNA (https://gitee.com/CJchen/IGV-sRNA) to estimate read coverage per site, and the means over the biological replicates. Estimates were made for each partial *PI*-like duplicate and for their surrounding regions.

Inverted repeat and hairpin structure identification

We annotated the repeat sequence of partial *PI*-like duplicates using the Repeat Finder Plugin (https://www.geneious.com/plugins/repeat-finder). RNA fold prediction was performed by the

Vienna package RNAfold tool [143], and hairpin structures were visualised with Geneious Prime 2023.2.1 (https://www.geneious.com/).

Luciferase reporter assay for *PI* gene and siRNAs

The *Salix baileyi PI*-like gene sequence (Sabai02aG0062500) with firefly luciferase reporter gene at its C-terminus was cloned into *pGreenII 0800-LUC* vector, to drive expression by the CaMV 35S promoter. Full-length siRNA sequence produced from partial *PI*-like sequences was artificially synthesised and cloned into *pGreenII-SK-62* vector for functional validation [144].

For the experiment (see Results), four distinct treatment groups were assigned to the same *N. benthamiana* leaves, to ensure consistency of results by infiltration into tissue in the same growth background. The first three groups served as the following controls: mixtures of the empty SK-62 vector and the empty 0800-LUC-35S vector (area 1), SK-62-siRNA and the empty 0800-LUC-35S vector (area 2), the empty SK-62 vector and the 0800-LUC-35S-intact *PI* construct (area 3). The fourth, area 4, was the experimental group, with mixtures of SK-62-siRNA and the 0800-LUC-35S-intact *PI* construct. All of the recombined plasmids were transformed into *Agrobacterium tumefaciens* (GV3101(pSoup-p19)). These *Agrobacterium* strains were cultured and resuspended with infiltration medium (10 mM 4-morpholineethane sulfonic acid, pH5.6; 10 mM $MgCl_2$; 150 mM acetosyringone), OD_{600} to 0.75. For all treatment groups, *A. tumefaciens* mixtures containing SK and the 3´UTR sensor were prepared in a ratio of 10:1, respectively. 72 hours after infection, luciferase activities were estimated in five biological replicates by the firefly and Renilla methods using a Promega GloMax96 instrument and the Dual-Luciferase Reporter Assay System (Promega) [145].

Estimation of the relative ages of gene duplications from Ks values, Ka/Ks estimates

As described in the Results section, transpositions have duplicated an intact *ARR17* homolog from chromosome 19 to 15W, and partial *PI* homologs from chromosome 2 and/or 5 to 15W, during the evolution of female heterogamety. To estimate the relative timings of these events, we used the intact *S. babylonica Vetrix* clade chromosome 19 *ARR17* sequence, and the chromosome 2 or 5 *PI* sequence

with the lower Ks value, as the likely progenitor of the chromosome 15 copies, consistent with the estimated phylogenetic relationships (Fig. 5c); for both genes this used only exon1, as explained above. The estimates were made using the yn00 model of PAML [146], we estimated the divergence between the sequences of the duplicate copies on chromosome 15W and those of their progenitors in the six newly sequenced taxa. The same software was used to obtain Ka/Ks estimates between the intact *ARR17* of 15ZW clade I species and one of the ancestral intact *ARR17* copy of *S. triandra*, to test whether the 15ZW clade I species accumulated more mutations.

We identified orthologous genes between the Z and W chromosomes within SLR across four species in the 15ZW clade I (*S. gordejevii*, *S. purpurea*, *S. baileyi*, *S. polyclona*), and used the same method mentioned before to calculate the Ks values of these Z-W homologous gene pairs within each species. Based on the mutation rate of *Salix* ($\mu = 1 \times 10^{-9}$ per year) [147], we estimated the divergence time between Z and W chromosomes for each species ($T = Ks/2\mu$).

Declarations

Peer review information

Wenjing She was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The genome assembly sequences have been deposited in the Figshare [148]. The sequencing datasets have been deposited in NCBI under the BioProject accession numbers PRJNA1104312 (*S. polyclona*) [149], PRJNA1110087 [150] and PRJNA1108470 (*S. gordejevii*) [151], PRJNA1110157 (*S. baileyi*) [152]. Publicly available resequencing data for *S. baileyi* and *S. polyclona* complex can be found in NCBI under the BioProject number PRJNA1108974 (*S. baileyi*) [153] and PRJNA765215 (*S. polyclona* complex) [154] respectively.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was financially supported by the National Natural Science Foundation of China (grant no. 32500303 to Z.-Q. X., 32171813 to L.H.) and Special Fund for Scientific Research of Shanghai Landscaping & City Appearance Administrative Bureau (grant nos. G232403, G242417, G252403 to L.H. & G252411 to Z.-Q. X.).

Author Contributions

L.H., Y.W. and Z.-Q. X. conceived, designed, and conceptualised the study. Y.W., Z.-Q. X., L.H., and R.-G. Z. performed the data analysis. Y.-F. M., Z.-Y. Z. and Z.-Q. X. performed the experiment. L.H. collected materials. Y.W., Z.-Q. X., D. C. and L.H. wrote the manuscript. D. C., E.H., X.-R. W. revised the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

We are grateful to Yi Wang, Jie Ding, Yi-Qun Liang, Qi-Chao Wu, Xin-Xin Zhu, Guan-Zhi Liu, Lu-Yao Li and Guang-Nan Gong for their kind help during preparation of our study. We are indebted to Judith E. Mank for her help with a previous version of the paper.

Supplementary Information

Additional file 1: Supplementary Figs. Figs S1-S49. A single file containing all supplementary figures referenced in the manuscript. Captions and legends for each supplementary figure are included within the file.

Additional file 2: Supplementary Tables. Tables S1–S23. A compiled multi-tab Excel file containing all supplementary tables referenced in the manuscript.

Additional file 3: Supplementary text. Detailed methods for identifying sex-linked regions.

References

1. Abbott JK, Nordén AK, Hansson B. Sex chromosome evolution: Historical insights and future perspectives. Proc R Soc B Biol Sci. 2017;284.

2. Kikuchi K, Hamaguchi S. Novel sex-determining genes in fish and sex chromosome evolution. Dev Dyn. 2013;242:339–53.

3. Ezaz T, Sarre SD, O'Meally D, Marshall Graves JA, Georges A. Sex chromosome evolution in lizards: independent origins and rapid transitions. Cytogenet Genome Res. 2009;127:249–60.

4. Miura I. Sex determination and sex chromosomes in Amphibia. Sex Dev. 2018;11:298–306.

5. Vicoso B. Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat Ecol Evol. 2019;3:1632–41.

6. Perrin N. Sex reversal: a fountain of youth for sex chromosomes? Evolution. 2009;63:3043–9.

7. Renner SS, Müller NA. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. Nat plants. 2021;7:392–402.

8. Charlesworth D. Young sex chromosomes in plants and animals. New Phytol. 2019;224:1095–107.

9. Lambing C, Tock AJ, Topp SD, Choi K, Kuo PC, Zhao X, et al. Interacting genomic landscapes of REC8-cohesin, chromatin, and meiotic recombination in *Arabidopsis*. Plant Cell. 2020;32:1218–39.

10. Wang Y, He L, Gong G, Zhang R, Hoerandl E, Charlesworth D. 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. New Phytol. 2024;242:2872–87.

11. Akagi T, Fujita N, Shirasawa K, Tanaka H, Nagaki K, Masuda K, et al. Rapid and dynamic evolution of a giant Y chromosome in *Silene latifolia*. Science. 2025;387:637–43.

12. Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature. 1994;371:215–20.

13. Rifkin JL, Beaudry FEG, Humphries Z, Choudhury BI, Barrett SCH, Wright SI. Widespread recombination suppression facilitates plant sex chromosome evolution. Mol Biol Evol. 2021;38:1018–30.

14. Charlesworth D, Harkess A. Why should we study plant sex chromosomes? Plant Cell. 2024;36:1242–56.

15. Ming R, Bendahmane A, Renner SS. Sex chromosomes in land plants. Annu Rev Plant Biol. 2011;62:485–514.

16. Charlesworth D. Distribution of dioecy and selfincompatibility in angiosperms. Evol honour John Maynard Smith. 1985;237–67.

17. Tennessen JA, Wei N, Straub SCK, Govindarajulu R, Liston A, Ashman T-L. Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. PLoS Biol. 2018;16:e2006062.

18. Qiao Q, Cao Q, Zhang R, Wu M, Zheng Y, Xue L, et al. Genomic analyses provide insights into sex differentiation of tetraploid strawberry (*Fragaria moupinensis*). Plant Biotechnol J. 2024; 22:1552–1565.

19. Akagi T, Varkonyi-Gasic E, Shirasawa K, Catanach A, Henry IM, Mertten D, et al. Recurrent neo-sex chromosome evolution in kiwifruit. Nat Plants. 2023;9:393–402.

20. Herpin A, Schartl M. Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. EMBO Rep. 2015;16:1260–74.

21. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, et al. Sex determination: why so many ways of doing it? PLoS Biol. 2014;12:e1001899.

22. Bull JJ. Evolution of sex determining mechanisms. 1983.

23. van Doorn GS, Kirkpatrick M. Transitions between male and female heterogamety caused by sex-antagonistic selection. Genetics. 2010;186:629–45.

24. Van Doorn GS, Kirkpatrick M. Turnover of sex chromosomes induced by sexual conflict. Nature. 2007;449:909–12.

25. Benatti TR, Valicente FH, Aggarwal R, Zhao C, Walling JG, Chen M-S, et al. A neo-sex chromosome that drives postzygotic sex determination in the Hessian fly (*Mayetiola destructor*). Genetics. 2010;184:769–77.

26. Roberts RB, Ser JR, Kocher TD. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science. 2009;326:998–1001.

27. Einfeldt AL, Kess T, Messmer A, Duffy S, Wringe BF, Fisher J, et al. Chromosome level reference of Atlantic halibut *Hippoglossus hippoglossus* provides insight into the evolution of sexual determination systems. Mol Ecol Resour. 2021;21:1686–96.

28. El Taher A, Ronco F, Matschiner M, Salzburger W, Böhne A. Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. Sci Adv. 2021;7:eabe8215.

29. Wang D, Li Y, Li M, Yang W, Ma X, Zhang L, et al. Repeated turnovers keep sex chromosomes young in willows. Genome Biol. 2022;23:1–23.

30. Yang W, Wang D, Li Y, Zhang Z, Tong S, Li M, et al. A general model to explain repeated turnovers of sex determination in the Salicaceae. Mol Biol Evol. 2021;38:968–80.

31. He L, Wang Y, Wang Y, Zhang R-G, Wang Y, Hörandl E, et al. Allopolyploidization from two dioecious ancestors leads to recurrent evolution of sex chromosomes. Nat Commun. 2024;15:6893. https://doi.org/10.1038/s41467-024-51158-3

32. Xue Z-Q, Applequist W-L, Hörandl E, He L. Sex chromosome turnover plays an important role in the maintenance of barriers to post-speciation introgression in willows. Evol Lett. 2024;8: 467–477.

33. Müller NA, Kersten B, Leite Montalvão AP, Mähler N, Bernhardsson C, Bräutigam K, et al. A single gene underlies the dynamic evolution of poplar sex determination. Nat plants. 2020;6:630–7.

34. Li Y, Wang D, Wang W, Yang W, Gao J, Zhang W, et al. A chromosome-level *Populus qiongdaoensis* genome assembly provides insights into tropical adaptation and a cryptic turnover of sex determination. Mol Ecol. 2023;32:1366–80.

35. Gladysh NS, Kovalev MA, Lantsova MS, Popchenko MI, Bolsheva NL, Starkova AM, et al. The Molecular and Genetic Mechanisms of Sex Determination in Poplar. Mol Biol. 2024;58:178–91.

36. Leite Montalvão AP, Kersten B, Kim G, Fladung M, Müller NA. *ARR17* controls dioecy in *Populus* by repressing B-class MADS-box gene expression. Philos Trans R Soc B. 2022;377:20210217.

37. Cronk Q, Müller NA. Default sex and single gene sex determination in dioecious plants. Front Plant Sci. 2020;11:1162.

38. Xue L, Wu H, Chen Y, Li X, Hou J, Lu J, et al. Evidences for a role of two Y-specific genes in

sex determination in *Populus deltoides*. Nat Commun. 2020;11:5893.

39. Yang H-W, Akagi T, Kawakatsu T, Tao R. Gene networks orchestrated by *MeGI*: a single-factor mechanism underlying sex determination in persimmon. Plant J. 2019;98:97–111.

40. Wang Y, Cai X, Zhang Y, Hörandl E, Zhang Z, He L. 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 (Edinb). 2023;130:177.

41. Hu N, Sanderson BJ, Guo M, Feng G, Gambhir D, Hale H, et al. Evolution of a ZW sex chromosome system in willows. Nat Commun. 2023;14:7144.

42. Wang Y, Zhang R-G, Hörandl E, Zhang Z-X, Charlesworth D, He L. Evolution of sex-linked genes and the role of pericentromeric regions in sex chromosomes: insights from diploid willows. Mol Biol Evol. 2024;msae235.

43. Zhou R, Macaya-Sanz D, Carlson CH, Schmutz J, Jenkins JW, Kudrna D, et al. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. Genome Biol. 2020;21:1–19.

44. Hyden B, Carlson CH, Gouker FE, Schmutz J, Barry K, Lipzen A, et al. Integrative genomics reveals paths to sex dimorphism in *Salix purpurea* L. Hortic Res. 2021;8.

45. Hyden B, Zou J, Wilkerson DG, Carlson CH, Robles AR, DiFazio SP, et al. Structural variation of a sex-linked region confers monoecy and implicates *GATA15* as a master regulator of sex in *Salix purpurea*. New Phytol. 2023;238:2512–23.

46. Mara CD, Irish VF. Two *GATA* transcription factors are downstream effectors of floral homeotic gene action in *Arabidopsis*. Plant Physiol. 2008;147:707–18.

47. Reyes JC, Muro-Pastor MI, Florencio FJ. The *GATA* family of transcription factors in *Arabidopsis* and rice. Plant Physiol. 2004;134:1718–32.

48. He L, Guo F-Y, Cai X-J, Chen H-P, Lian C-L, Wang Y, et al. Evolutionary origin and establishment of a dioecious diploid-tetraploid complex. Mol Ecol. 2023;32:2732–49.

49. Hallast P, Ebert P, Loftus M, Yilmaz F, Audano PA, Logsdon GA, et al. Assembly of 43 human Y chromosomes reveals extensive complexity and variation. Nature. 2023;1–10.

50. Wang J, Zhang L, Wang J, Hao Y, Xiao Q, Teng J, et al. Conversion between duplicated genes

generated by polyploidization contributes to the divergence of poplar and willow. BMC Plant Biol. 2022;22:298.

51. He L, Jia KH, Zhang RG, Wang Y, Shi T Le, Li ZC, et al. Chromosome-scale assembly of the genome of *Salix dunnii* reveals a male-heterogametic sex determination system on chromosome 7. Mol Ecol Resour. 2021;21:1966–82.

52. Gong G-N, Wang Y, Zhu Z-Y, Wang Y, Hörandl E, Wang X-R, et al. Evolutionary population dynamics and conservation strategies for *Salix baileyi* - a species with extremely small populations. Glob Ecol Conserv. 2025;58:e03504.
https://www.sciencedirect.com/science/article/pii/S2351989425001052

53. Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, et al. Haplotype-resolved assembly of diploid genomes without parental data. Nat Biotechnol. 2022;40:1332–1335.

54. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–5.

55. Liu H, Zang F, Wu Q, Ma Y, Zheng Y, Zang D. Genetic diversity and population structure of the endangered plant *Salix taishanensis* based on CDDP markers. Glob Ecol Conserv. 2020;24:e01242.

56. Podlevsky JD, Chen JJ-L. Evolutionary perspectives of telomerase RNA structure and function. RNA Biol. 2016;13:720–32.

57. Mefford HC, Trask BJ. The complex structure and dynamic evolution of human subtelomeres. Nat Rev Genet. 2002;3:91–102.

58. Bredeson J V, Mudd AB, Medina-Ruiz S, Mitros T, Smith OK, Miller KE, et al. Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs. Nat Commun. 2024;15:579.

59. Rhie A, Nurk S, Cechova M, Hoyt SJ, Taylor DJ, Altemose N, et al. The complete sequence of a human Y chromosome. Nature. 2023;621:344–54.

60. Wang Y, Zhao L, Wang D, Chen K, Luo T, Luo J, et al. Four near-complete genome assemblies reveal the landscape and evolution of centromeres in Salicaceae. Genome Biol. 2025;26:111.

61. Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, et al. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell. 2002;14:2825–36.

62. Talbert PB, Henikoff S. The genetics and epigenetics of satellite centromeres. Genome Res. 2022;32:608–15.

63. Ansai S, Toyoda A, Yoshida K, Kitano J. Repositioning of centromere-associated repeats during karyotype evolution in *Oryzias* fishes. Mol Ecol. 2024;33:e17222.

64. Almeida P, Proux-Wera E, Churcher A, Soler L, Dainat J, Pucholt P, et al. Genome assembly of the basket willow, *Salix viminalis*, reveals earliest stages of sex chromosome expansion. BMC Biol. 2020;18:1–18.

65. Hyden B, Feng K, Yates TB, Jawdy S, Cereghino C, Smart LB, et al. De Novo Assembly and Annotation of 11 Diverse Shrub Willow (*Salix*) Genomes Reveals Novel Gene Organization in Sex-Linked Regions. 2023;24:2904.

66. Wan J, Tang H, Yu Z, Wang J, Yao X, Li X. Characterization of the complete *Salix viminalis* var. *gmelinii* Turcz 1854 chloroplast genome from the northeast of China. Mitochondrial DNA Part B. 2022;7:1768–70.

67. Hyden B, Carper DL, Abraham PE, Yuan G, Yao T, Baumgart L, et al. Functional analysis of *Salix purpurea* genes support roles for *ARR17* and *GATA15* as master regulators of sex determination. Plant Direct. 2023;7:e546.

68. Volff J-N, Schartl M. Variability of genetic sex determination in Poeciliid fishes. Genetica. 2001;111:101–10.

69. Akagi T, Henry IM, Tao R, Comai L. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. Science. 2014;346:646–50.

70. Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, et al. A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature. 2014;509:633–6.

71. Wilkins AS. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. Bioessays. 1995;17:71–7.

72. Marinček P, Léveillé-Bourret É, Heiduk F, Leong J, Bailleul SM, Volf M, et al. Challenge accepted: Evolutionary lineages versus taxonomic classification of North American shrub willows (*Salix*). Am J Bot. 2024;111:e16361.

73. Mao XM, Rafati N, Tellgren-Roth C, Ingvarsson P, Karrenberg S.. Dynamic evolution of a sex-

linked region. preprint. 2025.

74. Dufresnes C, Borzée A, Horn A, Stöck M, Ostini M, Sermier R, et al. Sex-chromosome homomorphy in Palearctic tree frogs results from both turnovers and X-Y recombination. Mol Biol Evol. 2015;32:2328–37.

75. Méndez-Lago M, Bergman CM, De Pablos B, Tracey A, Whitehead SL, Villasante A. A large palindrome with interchromosomal gene duplications in the pericentromeric region of the *D. melanogaster* Y chromosome. Mol Biol Evol. 2011;28:1967–71.

76. Du J, Tian Z, Sui Y, Zhao M, Song Q, Cannon SB, et al. Pericentromeric effects shape the patterns of divergence, retention, and expression of duplicated genes in the paleopolyploid soybean. Plant Cell. 2012;24:21–32.

77. Blaser O, Neuenschwander S, Perrin N. Sex-chromosome turnovers: the hot-potato model. Am Nat. 2014;183:140–6.

78. Blaser O, Grossen C, Neuenschwander S, Perrin N. Sex-chromosome turnovers induced by deleterious mutation load. Evolution. Blackwell Publishing Inc Malden, USA; 2013;67:635–45.

79. Schenkel MA, Billeter J-C, Beukeboom LW, Pen I. Divergent evolution of genetic sex determination mechanisms along environmental gradients. Evol Lett. 2023;7:132–47.

80. Kozielska M, Weissing FJ, Beukeboom LW, Pen I. Segregation distortion and the evolution of sex-determining mechanisms. Heredity (Edinb). 2010;104:100–12.

81. Rifkin JL, Hnatovska S, Yuan M, Sacchi BM, Choudhury BI, Gong Y, et al. Recombination landscape dimorphism and sex chromosome evolution in the dioecious plant *Rumex hastatulus*. Philos Trans R Soc B. 2022;377:20210226.

82. Hibbins MS, Rifkin JL, Choudhury BI, Voznesenska O, Sacchi B, Yuan M, et al. Phylogenomics resolves key relationships in *Rumex* and uncovers a dynamic history of independently evolving sex chromosomes. Evol Lett. 2024;qrae060.

83. Martin H, Carpentier F, Gallina S, Godé C, Schmitt E, Muyle A, et al. Evolution of young sex chromosomes in two dioecious sister plant species with distinct sex determination systems. Genome Biol Evol. 2019;11:350–61.

84. Balounova V, Gogela R, Cegan R, Cangren P, Zluvova J, Safar J, et al. Evolution of sex

determination and heterogamety changes in section *Otites* of the genus *Silene*. Sci Rep. Nature Publishing Group UK London; 2019;9:1045.

85. Skvortsov AK. Willows of Russia and adjacent countries. Faculty of Mathematics and Natural Sciences Report Series, no. 39. University of Joensusu Press, Joensusu; 1999.

86. Fang Z, Zhao SD, Skvortsov AK. Salicaceae. In: Wu ZY, Raven PH, eds. Flora of China. Beijing and St. Louis: Science Press and Missouri Botanical Garden Press. 1999;4:139–274.

87. He L. Taxonomy and nomenclature of *Salix baileyi*, *S. rehderiana*, and *S. disperma*. Phytotaxa. 2018;349:54–60.

88. Liu L, He L, Applequist WL. Untangling two Chinese *Salix* species (Salicaceae) published by CK Schneider, with lectotypification of four names. Willdenowia. 2020;159–63.

89. Doležel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry. Nat Protoc. 2007;2:2233–44.

90. Gulyaev S, Cai X-J, Guo F-Y, Kikuchi S, Applequist WL, Zhang Z-X, 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. 2022;129:485–98.

91. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70.

92. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11:1–10.

93. Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi--C: a comprehensive technique to capture the conformation of genomes. Methods. 2012;58:268–76.

94. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3:95–8.

95. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356:92–5.

96. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016;3:99–

101.

97. Xu G-C, Xu T-J, Zhu R, Zhang Y, Li S-Q, Wang H-W, et al. LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. Gigascience. 2019;8:giy157.

98. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:1–31.

99. Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. 2016;44:e113-e113.

100. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 2020;36:2253–5.

101. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.

102. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31:3350–2.

103. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.

104. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5.

105. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.

106. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31:5654–66.

107. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24:637–44.

108. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18:188–96.

109. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene

structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:1–22.

110. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.

111. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017;34:2115–22.

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

113. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.

114. Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20:1–18.

115. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.

116. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:1–14.

117. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

118. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015

119. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 2018;19:15–30.

120. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884--i890.

121. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv

Prepr arXiv13033997. 2013;

122. Li H, Durbin R. Fast and accurate short read alignment with Burrows--Wheeler transform. bioinformatics. 2009;25:1754–60.

123. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. bioinformatics. 2009;25:2078–9.

124. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 2015;31:2032–4.

125. Kelly BJ, Fitch JR, Hu Y, Corsmeier DJ, Zhong H, Wetzel AN, et al. Churchill: an ultra-fast, deterministic, highly scalable and balanced parallelization strategy for the discovery of human genetic variation in clinical and population-scale genomics. Genome Biol. 2015;16:6.

126. Yu L, Stachowicz JJ, DuBois K, Reusch TBH. Detecting clonemate pairs in multicellular diploid clonal species based on a shared heterozygosity index. Mol Ecol Resour. 2023;23:592–600.

127. Hall AB, Qi Y, Timoshevskiy V, Sharakhova M V, Sharakhov I V, Tu Z. Six novel Y chromosome genes in *Anopheles mosquitoes* discovered by independently sequencing males and females. BMC Genomics. 2013;14:1–13.

128. Killick R, Eckley I. changepoint: An R package for changepoint analysis. J Stat Softw. 2014;58:1–19.

129. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;1358–70.

130. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

131. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. https://ggplot2.tidyverse.org

132. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res. 2008;18:1944–54.

133. Zhang R-G, Li G-Y, Wang X-L, Dainat J, Wang Z-X, Ou S, et al. TEsorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. Hortic Res. 2022;9:uhac017.

134. Dong S-S, He W-M, Ji J-J, Zhang C, Guo Y, Yang T-L. LDBlockShow: a fast and convenient

tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. Brief Bioinform. 2021;22:bbaa227.

135. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, et al. TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. Mol Plant. 2023;16:1733–42.

136. Handley L-JL, Ceplitis H, Ellegren H. Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. Genetics. 2004;167:367–76.

137. Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al. Origins and functional evolution of Y chromosomes across mammals. Nature. 2014;508:488–93.

138. Maddison W, Maddison D. Mesquite 2. A Modul Syst Evol Anal. 2007;3.

139. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.

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

141. Li G, Chen C, Chen P, Meyers BC, Xia R. sRNAminer: A multifunctional toolkit for next-generation sequencing small RNA data mining in plants. Sci Bull. 2023;

142. Chen C, Li J, Feng J, Liu B, Feng L, Yu X, et al. sRNAanno—a database repository of uniformly annotated small RNAs in plants. Hortic Res. 2021;8.

143. Lorenz R, Bernhart SH, zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:1–14.

144. Wang W, Wang J, Wu Y, Li D, Allan AC, Yin X. Genome-wide analysis of coding and non-coding RNA reveals a conserved miR164-NAC regulatory pathway for fruit ripening. New Phytol. 2020;225:1618–34.

145. Liu Q, Wang F, Axtell MJ. Analysis of complementarity requirements for plant microRNA targeting using a *Nicotiana benthamiana* quantitative transient assay. Plant Cell. 2014;26:741–53.

146. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.

147. Berlin S, Fogelqvist J, Lascoux M, Lagercrantz U, Rönnberg-Wästljung AC. Polymorphism and divergence in two willow species, *Salix viminalis* L. and *Salix schwerinii* E. Wolf. G3 Genes|

Genomes| Genet. 2011;1:387–400.

148. Wang Y, Xue Z-Q, Zhang R-G, Zhu Z-Y, Hörandl E, Wang X-R, Mao Y-F, Charlesworth D, He L. Recurrent sex chromosome turnover mediated by distinct *ARR17* and *PISTILLATA* duplications in willows. Figshare. https://doi.org/10.6084/m9.figshare.31408812. (2026).

149. Wang Y, Xue Z-Q, Zhang R-G, Zhu Z-Y, Hörandl E, Wang X-R, Mao Y-F, Charlesworth D, He L. Recurrent sex chromosome turnover mediated by distinct *ARR17* and *PISTILLATA*duplications in willows. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1104312. (2026).

150. Wang Y, Xue Z-Q, Zhang R-G, Zhu Z-Y, Hörandl E, Wang X-R, Mao Y-F, Charlesworth D, He L. Recurrent sex chromosome turnover mediated by distinct *ARR17* and *PISTILLATA* duplications in willows. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1110087. (2026).

151. Wang Y, Xue Z-Q, Zhang R-G, Zhu Z-Y, Hörandl E, Wang X-R, Mao Y-F, Charlesworth D, He L. Recurrent sex chromosome turnover mediated by distinct *ARR17* and *PISTILLATA* duplications in willows. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1108470. (2026).

152. Wang Y, Xue Z-Q, Zhang R-G, Zhu Z-Y, Hörandl E, Wang X-R, Mao Y-F, Charlesworth D, He L. Recurrent sex chromosome turnover mediated by distinct *ARR17* and *PISTILLAT*A duplications in willows. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1110157. (2026).

153. Gong GN, Wang Y, Zhu ZY, Wang Y, Hörandl E, Wang XR, Xue ZQ, He L. Resequencing data for *S. baileyi*. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1108974. (2025).

154. He L, Guo FY, Cai XJ, Chen HP, Lian CL, Wang Y, Shang C, Zhang Y, Wagner ND, Zhang ZX, Hörandl E, Wang XR. Resequencing data for *S. polyclona* complex. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/ PRJNA765215. (2023).

Figure legends

Figure 1 Phylogenetic relationship and genome synteny of *Salix*. a Inferred phylogenetic tree based 6,467 single-copy autosomal genes present in genome sequences of 12 willows and the outgroup *P. trichocarpa*. Latin names in bold indicate the six newly sequenced genomes. Two independent 15XY to 15ZW transitions were mapped on relevant positions following refs. [10,31]. The deep black arrows indicate the turnover events in the 15ZW clade I lineage, which is the focus of our study. Numbers marked above branches represent posterior probabilities. **b** Genome synteny between chromosomes, and sex chromosomes were marked in figure.

Figure 2 Sex-linked regions in *S. gordejevii*, *S. baileyi* and the *S. polyclona* complex (W1, E, E-TS, and W2 samples). For each of these species with female heterogamety (indicated by their names), the F_{ST} values together with the synteny results (shown with chromosome diagrams in between the F_{ST} and CQ values based on the assembled Z and W haplotypes, the upper and lower plots, respectively) indicate sex-linked regions on their chromosome 15s, and the coverage differences of short reads from male and female samples to the two haplotypes, in the CQ analyses, confirm their locations; *k-mer* analysis results are shown only for the W-linked regions, as this analysis tests for female-specific sequences. The detailed results are provided in Additional file 1: Figs. S7-S18. The inferred SLR boundaries are indicated in the chromosome diagrams by black lines, and the SLRs are shaded in pink; blue and green lines on these diagrams indicate forward-and reverse-oriented genes, respectively. Inversions near the SLR boundaries are indicated in green.

Figure 3 Genomic features of the sex chromosomes of *S. gordejevii*, *S. baileyi*, and *S. polyclona* complex. a Densities of genes and four kinds of repeats (TEs of all types, Copia, Gypsy, and CR (centromeric retrotransposon in maize), see Methods). The gray shaded areas indicate the SLRs. **b** The percentage of LTR-gypsy and -copia sequences in SLRs, PARs, and the whole genomes. **c** The proportion of genes inferred to have been lost (labelled degeneration rate) from the sex-linked regions of the chromosomes in *Salix* species with male and female heterogamety sex determination

systems involving sex-linked regions on chromosome 15. The inferences were based on gene loss situation compared with the autosomal 15*a* haplotype of *S. dunnii* (7XY). Degeneration rate for W-SLR = W-SLR loss / (W-SLR loss + Z-SLR loss + their shared genes) and degeneration rate for Z-SLR = Z-SLR loss / (W-SLR loss + Z-SLR loss + their shared genes).

Figure 4 Phylogenies of three sex-linked genes in two 15 XY species (*S. arbutifolia* and *S. triandra*) and several 15ZW species. Species set1 includes *S. purpurea*, *S. baileyi*, *S. polyclona*-W2, and *S. polyclona*-E-TS, and species set2 includes *S. gordejevii*, *S. polyclona*-W1, *S. polyclona*-E, and the *S. babylonica* V. In the left-hand column only the four ZW taxa in set 1 (a, c, e), while the right-hand column results include species of set 1 and set 2 (b, d, f). **a-d** Two genes whose complete sex-linkage evolved before the split between the 15ZW clade I and the *S. babylonica* V genome. **e** and **f** show results for a gene that remained partially sex-linked after the initial radiation of the 15ZW clade I, as discussed in the text. Numbers above branches are bootstrap values.

Figure 5 Genes on the sex chromosomes of the newly sequenced species with female heterogamety, *S. gordejevii*, *S. baileyi*, and the *S. polyclona* complex, including possible sex-determining factors. a Upset plot of shared and unique gene numbers across of W-SLR specific gene (W-SLR genes with no BLAST hits in the counterpart Z-SLR) of the six genome assemblies, showing that only the *PI*-like sequence is shared by all six taxa. **b** The locations of intact *ARR17*-like genes and partial *PI*-like sequences in the W-linked regions. **c** Phylogenetic tree of intact and partial *PI*-like duplicates. Numbers marked on the tree represent bootstrap values. E, E-TS, W1, W2 indicate the species, as follows: *S. polyclona*-E, *S. polyclona*-E-TS, *S. polyclona*-W1, *S. polyclona*-W2; Sagor: *S. gordejevii*; Sapur: *S. purpurea*; Sabai: *S. baileyi*. **d and e** Estimated expression (TPM values) in the tissues indicated of intact *ARR17*-like and *PI*-like duplicates, respectively, in the same species as those in part c (the same abbreviated names are used).

Figure 6 The expression and functional validation of partial *PI*-like duplicates. **a** The numbers of small RNAs (y axes) corresponding to sequences in the region carrying the partial *PI*-like

duplicates in the six newly sequenced taxa (with the same abbreviated names as in Fig. 5); vertical yellow lines indicate the duplicates' positions (x axes). The yellow triangles represent partial *PI*-like duplicates, with triangles pointing right indicating forward orientation and triangles pointing left indicating reverse orientation. **b** Design of the dual-luciferase reporter assay to test the effects of siRNAs on expression of the intact willow *PI* homologs. Colours indicate luciferase the fluorescence intensity levels, expressed as p/s/cm²/sr (photons per second per square centimeter per steradian; see the key). **c** Diagram of the proposed sex determination mechanisms in *Salix* as described in the Results section and in the papers of[33,36].

Figure 7 Hypothesis for changes in sex determination in *Salix*. The left-hand side shows the ancestral reconstruction results (see Additional file 1: Fig S35) based on duplicates of intact *ARR17*-like genes and partial *PI*-like genes in the 15W-linked regions (and the *S. purpurea*-specific *GATA15* in its W-linked region), and the parts on the right show diagrams of the genes involved in each set of species, and their chromosomal locations. In 7XY and 15XY species (at the bottom), Y-linked partial *ARR17*-like duplicates are present, and can inhibit the intact *ARR17*-like copies on chromosome 19. The XY ancestor of the 15ZW allotetraploid *S. babylonica* had eight intact *ARR17*-like genes on chromosome 19, due to a duplication, but the 15Y-SLR had only a single set of partial *ARR17*-like duplicates, which may have led to the 15XY→15ZW change, as previously proposed [31]; the partial *ARR17*-like duplicates on the two 15Z-SLRs present in extant ZZ *S. babylonica* males can suppress the intact chromosome 19 *ARR17*-like copies. In the 15ZW clade I species (at the top), partial *PI*-like duplicates are present on chromosome 15, determining femaleness, and the system based on *ARR17*-like genes has been lost.

Table

Table 1 Statistics of six willows genome assemblies.

	Salix gordejevii	***Salix baileyi***	***S. polyclona*-W1**	***S. polyclona*-E**	***S. polyclona*-TS**	***S. polyclona*-W2**
Ploidy	2X	2X	2X	2X	2X	4X
Total assembly size (two or four haplotypes, Mb)	652	670	785	823	831	1,507
Total number of contigs	51	41	44	44	48	93
Maximum contig length (Mb)	30	33	35	35	35	44
Minimum contig length (Kb)	156	155	156	156	156	156
Contig N50 length (Mb)	15	18	21	22	23	19
Contig N50 count	17	16	16	17	16	32
Contig N90 length (Mb)	9	12	13	14	13	14
Contig N90 count	37	34	34	34	34	68
Gap number	11	1	4	4	8	12
GC content (%)	34.83	34.77	34.97	35.19	35.27	34.89
Gene number	61,481	61,894	62,004	62,396	63,175	132,186
Transcript number	84,714	80,128	87,775	82,136	87,039	165,793
Repeat content (%)	43.95	44.66	51.21	53.78	57.28	46.73
BUSCO assessment	97.90%	98.50%	98.20%	98.60%	98%	98.90%

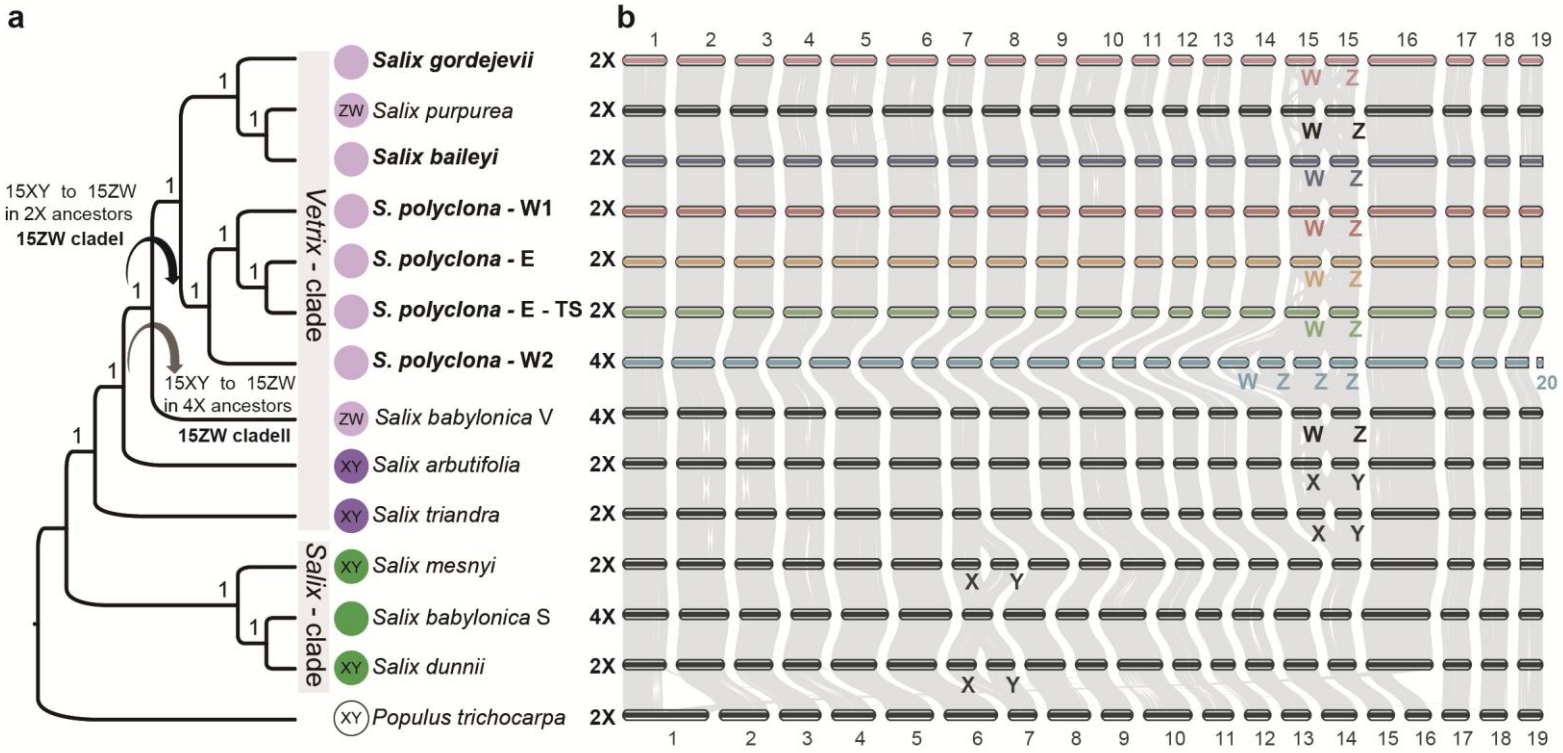
Table 2 Identification of SLRs in the newly sequenced *Salix* species. The sex-linked region locations are shown in megabases (Mb) in the relevant assemblies. "/" indicates that the specific signal region cannot be determined by this method.

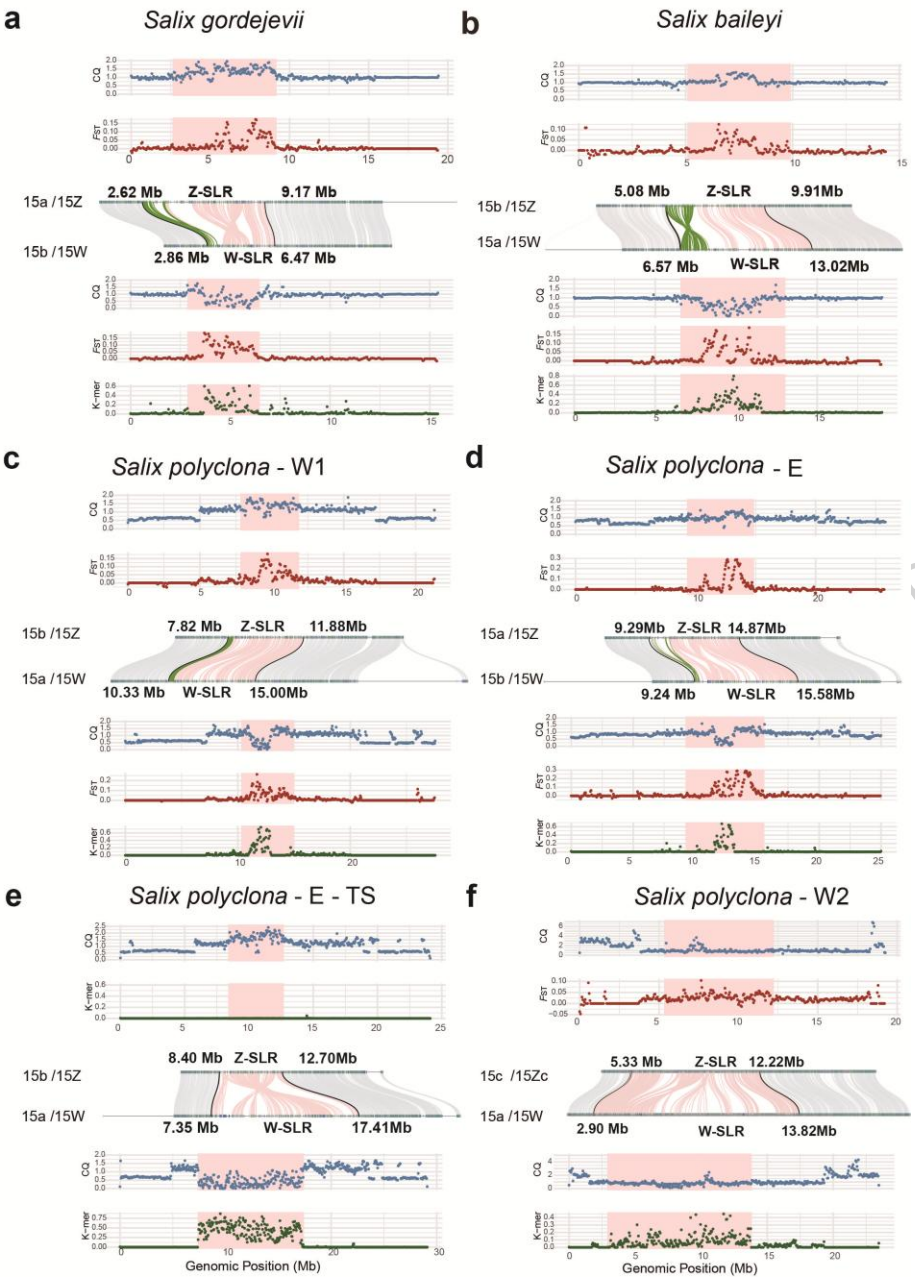
Species	Chromosome	Based on CQ results	Mean CQ in the estimated SLRs	Based on F_{ST} results	Based on *k-mer* results	Positions of inversions	Based on all criteria
S. gordejevii	15a(Z)	3.40 – 9.15	1.41	5.34 – 9.17	/	2.74 – 3.71	2.62 – 9.17
	15b(W)	3.65 – 6.40	0.57	3.62 – 6.47	3.65 – 6.05	2.86 – 2.97	2.86 – 6.47
S. baileyi	15a(W)	7.85 – 11.55	0.54	7.88 – 12.63	7.60 – 11.60	6.57 – 8.24	6.57 – 13.02
	15b(Z)	6.90 – 8.30	1.4	6.37 – 9.91	/	5.08 – 6.43	5.08 – 9.91
S. polyclona-E	15a (Z)	12.45 - 14.10	1.26	10.66 – 14.87	/	9.29 - 10.34, 10.78 - 12.26	9.29 - 14.87
	15b (W)	11.55 - 13.05	0.33	11.30 - 15.58	11.5 - 13.20	9.24 - 9.52, 10.30 - 11.74	9.24 - 15.58
S. polyclona-E-TS	15a (W)	7.40 - 17.25	0.4	No data	7.35 - 17.40	8.76 - 16.36	7.35 - 17.41
	15b (Z)	8.40 - 12.70	1.67	No data	/	9.09 - 12.22	8.40 - 12.70
S. polyclona-W1	15a (W)	11.25 - 12.90	0.39	11.06 – 15.00	11.05 – 13.00	10.33 - 10.60	10.33 - 15.00
	15b (Z)	8.10 - 12.00	1.38	9.10 – 11.49	/	7.82 - 8.10	7.82 - 11.88
S. polyclona-W2 (tetraploid)	15a (W)	6.60 - 10.30	0.59	8.43 - 11.92	2.90 - 13.82	2.90 - 13.82	2.90 - 13.82
	15b (Z)	/	/	5.35 - 8.53	/	4.30 - 10.82	4.30 - 10.82
	15c (Z)	6.90 - 7.85	1.88	6.52 - 12.22	/	5.33 - 12.22	5.33 - 12.22
	15d (Z)	/	/	3.72 - 6.46	/	2.62 -9.28	2.62 - 9.28

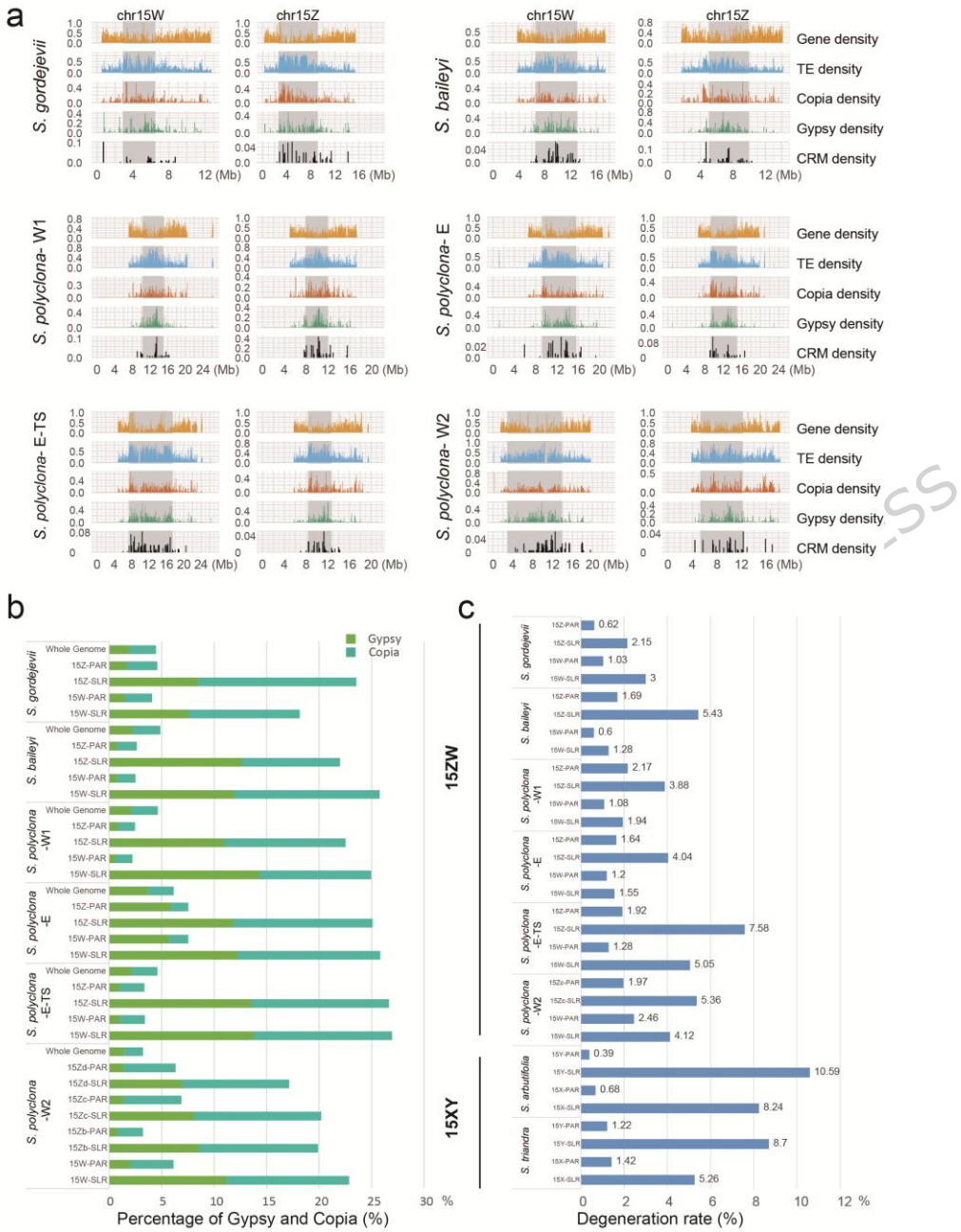
Table 3 The distribution of sex-determining factors in *Salix*.

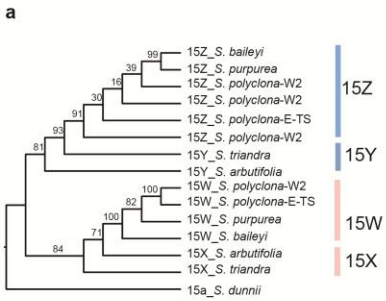
Clade and sex determination system	**Taxon**	***MSF*: Male specific factor**		**Intact *ARR17*-like duplicates**		**Partial *ARR17*-like duplicates**		**Intact *PI*-like duplicates**		**Partial *PI*-like duplicates**	***GATA15***
		Chr15	Chr7	Chr15	Chr19	Chr7	Chr15	Chr2	Chr5	Chr15	Chr15
Vetrix clade 15ZW clade I	*S. gordejevii*	1(W)	0	0	1	0	2(W)	2	2	5(W)	0
Vetrix clade 15ZW clade I	*S. purpurea*	1(Z)	0	4(W)	2	0	7(Z)	1	1	2(W)	1(W)
Vetrix clade 15ZW clade I	*S. baileyi*	1(Z)	0	1(W)	2	0	3(Z)	2	2	2(W)	0
Vetrix clade 15ZW clade I	*S. polyclona*-W1	2(Z) & 2(W)	0	2(Z) & 2(W)	4	0	7(Z) & 7(W)	2	2	4(W)	0
Vetrix clade 15ZW clade I	*S. polyclona*-E	1(Z) & 2(W)	0	2(Z) & 2(W)	4	0	4(Z) & 6(W)	3	2	9(W)	0
Vetrix clade 15ZW clade I	*S. polyclona*-E-TS	1(Z)	0	2(Z) & 14(W)	4	0	4(Z)	2	2	9(W)	0
Vetrix clade 15ZW clade I	*S. polyclona*-W2	2(Zb) & 2(Zc) & 2(Zd)	0	1(Zb) & 2(Zc) & 2(Zd) & 15(W)	6	0	3(Zb) & 5(Zc) & 8(Zd)	5	4	3(W)	0

Vetrix clade 15ZW clade II	*S. babylonica*	2(Z)	0	0	8	0	4(Z)	4	4	0	0
Vetrix clade, 15XY	*S. arbutifolia*	2(Y)	0	0	4	0	7(Y)	2	2	0	0
Vetrix clade, 15XY	*S. triandra*	4(Y)	0	0	4	0	7(Y)	2	2	0	0
Salix clade, 7XY	*S. mesnyi*	0	2(Y)	0	4	4(Y)	0	2	2	0	0
Salix clade, 7XY	*S. dunnii*	0	2(Y)	0	4	6(Y)	0	2	2	0	0

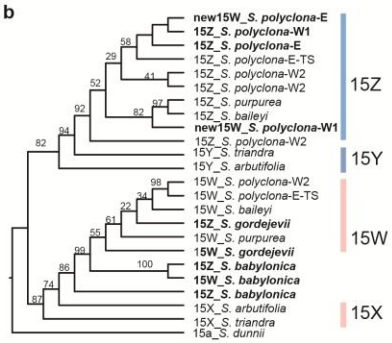

a
b
15XY to 15ZW
in 2X ancestors
15ZW cladeI
15XY to 15ZW
in 4X ancestors
15ZW cladeII
Vetrix - clade
Salix - clade
Salix gordejevii
Salix purpurea
Salix baileyi
S. polyclona - W1
S. polyclona - E
S. polyclona - E - TS
S. polyclona - W2
Salix babylonica V
Salix arbutifolia
Salix triandra
Salix mesnyi
Salix babylonica S
Salix dunnii
Populus trichocarpa
2X
4X
W Z
X Y
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19


a
Salix gordejevii
CQ
FST
2.62 Mb Z-SLR 9.17 Mb
15a /15Z
15b /15W
2.86 Mb W-SLR 6.47 Mb
K-mer
b
Salix baileyi
5.08 Mb Z-SLR 9.91Mb
15b /15Z
15a /15W
6.57 Mb W-SLR 13.02Mb
c
Salix polyclona - W1
7.82 Mb Z-SLR 11.88Mb
15b /15Z
15a /15W
10.33 Mb W-SLR 15.00Mb
d
Salix polyclona - E
9.29Mb Z-SLR 14.87Mb
15a /15Z
15b /15W
9.24 Mb W-SLR 15.58Mb
e
Salix polyclona - E - TS
8.40 Mb Z-SLR 12.70Mb
15b /15Z
15a /15W
7.35 Mb W-SLR 17.41Mb
Genomic Position (Mb)
f
Salix polyclona - W2
5.33 Mb Z-SLR 12.22Mb
15c /15Zc
15a /15W
2.90 Mb W-SLR 13.82Mb
Genomic Position (Mb)

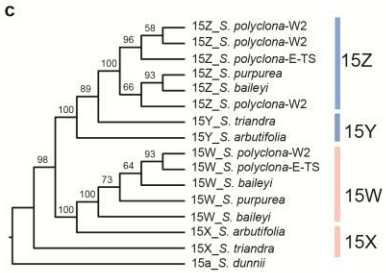

a
chr15W
chr15Z
S. gordejevii
S. baileyi
S. polyclona- W1
S. polyclona- E
S. polyclona- E-TS
S. polyclona- W2
Gene density
TE density
Copia density
Gypsy density
CRM density
b
Gypsy
Copia
Whole Genome
15Z-PAR
15Z-SLR
15W-PAR
15W-SLR
15Zd-PAR
15Zd-SLR
15Zc-PAR
15Zc-SLR
15Zb-PAR
15Zb-SLR
Percentage of Gypsy and Copia (%)
c
15ZW
15XY
15Z-PAR
15Z-SLR
15W-PAR
15W-SLR
15Zc-PAR
15Zc-SLR
15Y-PAR
15Y-SLR
15X-PAR
15X-SLR
S. arbutifolia
S. triandra
Degeneration rate (%)



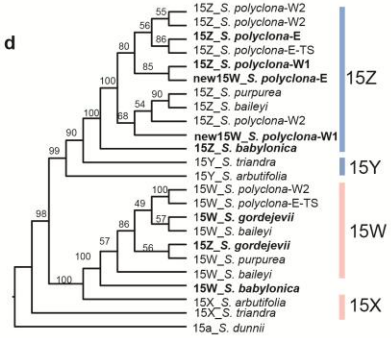
Saarb15bG0048300 and its ortholog shared across the *Vetrix*-clade species



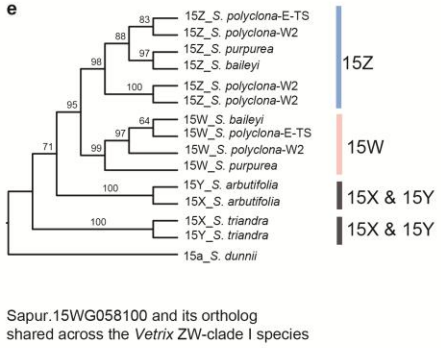
Tree **a** adding *S. gordejevii*, *S. polyclona*-W1, *S. polyclona*-E, *S. babylonica*



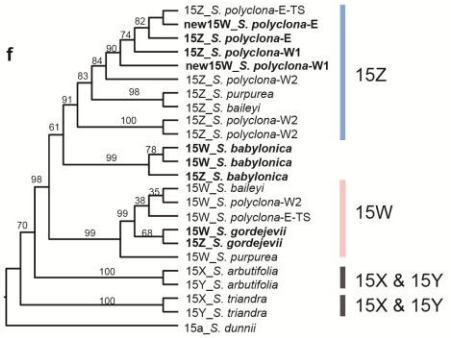
Saarb15bG0044000 and its ortholog shared across the *Vetrix*-clade species



Tree **c** adding *S. gordejevii*, *S. polyclona*-W1, *S. polyclona*-E, *S. babylonica*



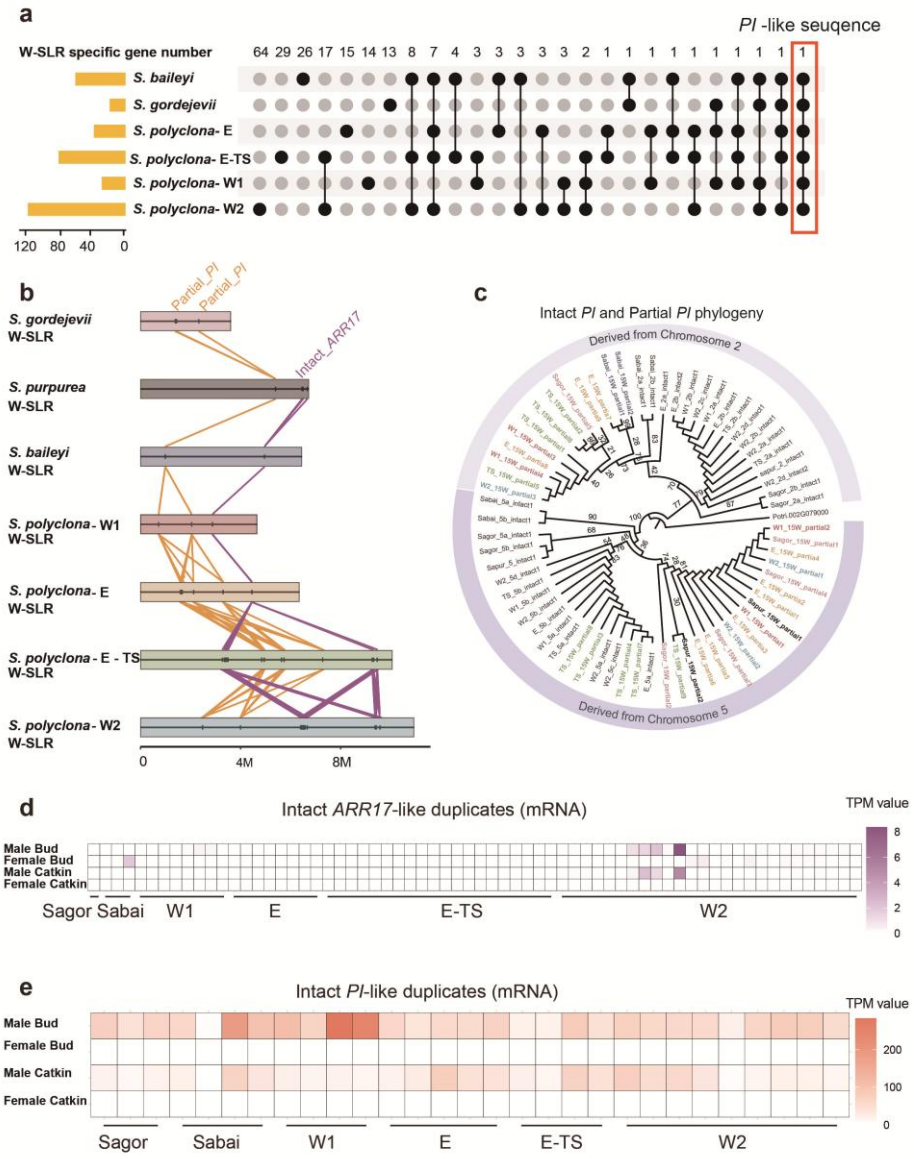
Sapur.15WG058100 and its ortholog shared across the *Vetrix* ZW-clade I species

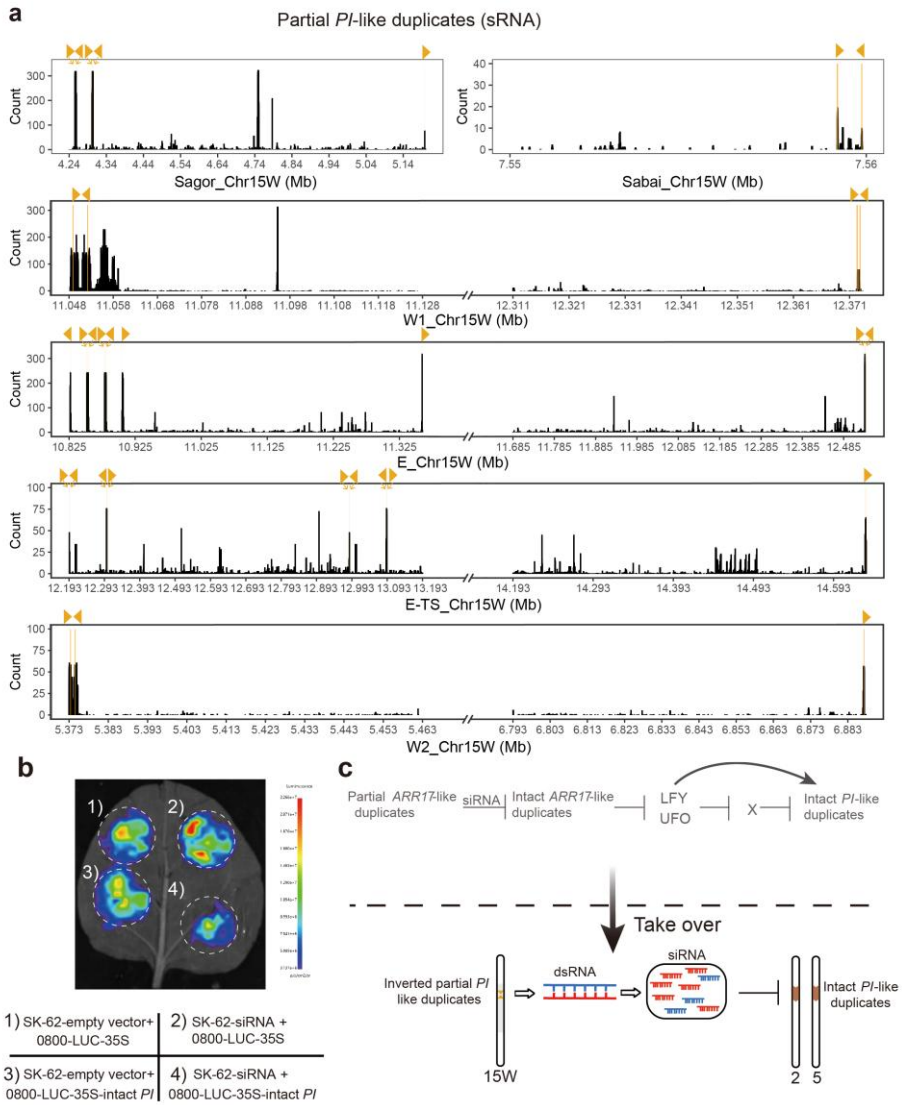


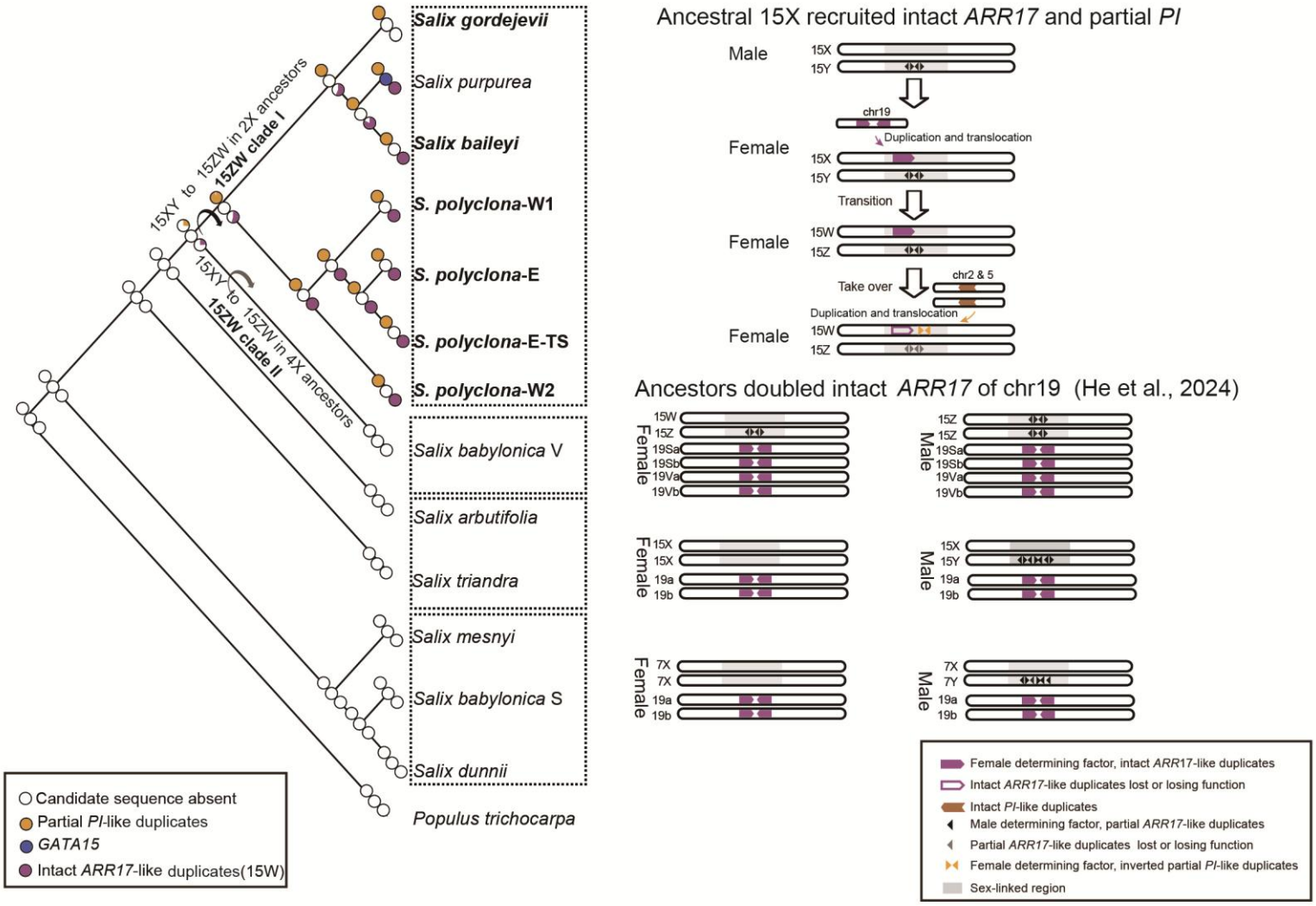
Tree **e** adding *S. gordejevii*, *S. polyclona*-W1, *S. polyclona*-E, *S. babylonica*

Species set 1 (a, c, e)

Species set1 + set2 (b, d, f)


a
PI -like seuqence
W-SLR specific gene number
64 29 26 17 15 14 13 8 7 4 3 3 3 3 2 1 1 1 1 1 1 1 1 1
S. baileyi
S. gordejevii
S. polyclona- E
S. polyclona- E-TS
S. polyclona- W1
S. polyclona- W2
120 80 40 0
b
Partial_PI
Partial_PI
Intact_ARR17
S. gordejevii W-SLR
S. purpurea W-SLR
S. baileyi W-SLR
S. polyclona - W1 W-SLR
S. polyclona - E W-SLR
S. polyclona - E - TS W-SLR
S. polyclona - W2 W-SLR
0
4M
8M
c
Intact PI and Partial PI phylogeny
Derived from Chromosome 2
Derived from Chromosome 5
d
Intact ARR17-like duplicates (mRNA)
TPM value
8
6
4
2
0
Male Bud
Female Bud
Male Catkin
Female Catkin
Sagor Sabai W1 E E-TS W2
e
Intact PI-like duplicates (mRNA)
TPM value
200
100
0
Male Bud
Female Bud
Male Catkin
Female Catkin
Sagor Sabai W1 E E-TS W2


a
Partial PI-like duplicates (sRNA)
Count
Sagor_Chr15W (Mb)
Sabai_Chr15W (Mb)
W1_Chr15W (Mb)
E_Chr15W (Mb)
E-TS_Chr15W (Mb)
W2_Chr15W (Mb)
b
1) SK-62-empty vector+ 0800-LUC-35S
2) SK-62-siRNA + 0800-LUC-35S
3) SK-62-empty vector+ 0800-LUC-35S-intact PI
4) SK-62-siRNA + 0800-LUC-35S-intact PI
c
Partial ARR17-like duplicates
siRNA
Intact ARR17-like duplicates
LFY
UFO
X
Intact PI-like duplicates
Take over
Inverted partial PI like duplicates
dsRNA
siRNA
Intact PI-like duplicates
15W
2 5


Salix gordejevii
Salix purpurea
Salix baileyi
S. polyclona-W1
S. polyclona-E
S. polyclona-E-TS
S. polyclona-W2
Salix babylonica V
Salix arbutifolia
Salix triandra
Salix mesnyi
Salix babylonica S
Salix dunnii
Populus trichocarpa
15XY to 15ZW in 2X ancestors
15ZW clade I
15XY to 15ZW in 4X ancestors
15ZW clade II
Candidate sequence absent
Partial PI-like duplicates
GATA15
Intact ARR17-like duplicates(15W)
Ancestral 15X recruited intact ARR17 and partial PI
Male
15X
15Y
chr19
Duplication and translocation
Female
15X
15Y
Transition
Female
15W
15Z
Take over
chr2 & 5
Duplication and translocation
Female
15W
15Z
Ancestors doubled intact ARR17 of chr19 (He et al., 2024)
Female
15W
15Z
19Sa
19Sb
19Va
19Vb
Male
15Z
15Z
19Sa
19Sb
19Va
19Vb
Female
15X
15X
19a
19b
Male
15X
15Y
19a
19b
Female
7X
7X
19a
19b
Male
7X
7Y
19a
19b
Female determining factor, intact ARR17-like duplicates
Intact ARR17-like duplicates lost or losing function
Intact PI-like duplicates
Male determining factor, partial ARR17-like duplicates
Partial ARR17-like duplicates lost or losing function
Female determining factor, inverted partial PI-like duplicates
Sex-linked region