Dynamic evolution of a sex-linked region

Xiaomeng Mao
Uppsala University
Nima Rafati
Uppsala University
Christian Tellgren-Roth
Uppsala University
Pär Ingvarsson
Swedish University of Agricultural Sciences
Sophie Karrenberg
sophie.karrenberg@ebc.uu.se
Uppsala University

Research Article

Keywords: Sex chromosome evolution, Inversion, Sex determination, Salicaceae

Posted Date: February 10th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-5980036/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Genome Biology on March 19th, 2026. See the published version at https://doi.org/10.1186/s13059-026-04038-6.

Abstract

Background

Sex chromosomes often evolve exceptionally fast and degenerate after recombination arrest. However, the underlying evolutionary processes are under persistent debate, particularly whether or not recombination arrest evolves in a stepwise manner and how switches in sex determination genes contribute to sex chromosome evolution. Here, we study sex determination in the dioecious plant genus *Salix* with a high turnover of sex chromosomes.

Results

We identified Z and W sex-linked regions (~ 8 Mb) on chromosome 15 of the dwarf willow *Salix herbacea* using a new haplotype-resolved assembly. The W sex-linked region harboured a large (5 Mb) embedded inversion. Analyses of synteny with other *Salix* species, sequence divergence between sex chromosomes and degeneration suggest that the inversion recently incorporated pseudoautosomal sequences into the sex-linked region, extending its length nearly three-fold. W-hemizygous regions exclusively contained seven pairs of inverted partial repeats of the male essential floral identity gene *PISTILLATA*, suggesting a possible *PISTILLATA* suppression mechanism by interfering RNA in females. Such *PISTILLATA* pseudogenes were also found in other *Salix* species with ZW sex determination but not in those with XY sex determination.

Conclusions

Our study provides rare and compelling direct support for the long-standing theory of stepwise recombination reduction mediated by inversions and suggests that the turnover of sex chromosomes in the Salicaceae family is associated with a switch of the sex determination gene.

Background

The evolution of sex determination and sex chromosomes are fundamentally important for the evolution of most animals and some plants. Sex chromosomes act most commonly in an XY system (XX females and XY males) or in a ZW system (ZW females and ZZ males). Y and W chromosomes often experience recombination arrest and ensuing degeneration, including accumulation of deleterious variants and repetitive elements, massive gene loss and breakdown of synteny between sex chromosomes [1, 2]. This process is, however, not inevitable. On the one hand, sex-linked regions can escape degeneration due to frequent turn-over leading to a change in the sex chromosome, the sex determining gene or even the sex determining system, as seen in frogs [3], cichlids [4] and in willows and poplars (Salicaceae) [5–7]. On the other hand, small sex-linked regions can also remain stable and without substantial degeneration over extended evolutionary time, for example, in ratite birds and sturgeons [8, 9]. These findings

accentuate the need to understand the mechanisms and evolutionary processes leading to sex chromosome degeneration.

Theoretical studies suggest at least four non-exclusive evolutionary drivers for the establishment of recombination arrest on sex chromosomes, but empirical evidence is currently scarce or controversial for all of them [10, 11]. These drivers are (1) sexually antagonistic mutations, first described in the classic model by Rice [12], (2) regulatory evolutions [13], (3) sheltering of deleterious mutations on the Y or W (recent update in [14]) and (4) neutral sequence divergence [15]. Theories on the first three drivers often involve (1) or require (2 and 3) inversions on the Y or W chromosome that capture mutations under selection and therefore predict stepwise reduction in recombination between sex chromosomes. The fourth theory, neutral sequence divergence, in contrast, predicts a gradual decrease in recombination reduction. Many empirical studies report discontinuous patterns of sequence divergence in X-Y or Z-W homologous genes that are assigned to evolutionary strata in accordance with theories predicting stepwise recombination reduction [11]. Evolutionary strata have been convincingly associated with fusions between established sex chromosomes and autosomes (neo-sex chromosomes) [16–18]; however, there is only very little direct evidence of inversions associated with evolutionary strata [11], for example in humans [19] and sticklebacks [20]. Assessing the role of inversions for sex chromosome evolution remains a challenging endeavour, not only due to the need for a haplotype-resolved assembly but also because sex chromosomes can degenerate to an extent that makes the detection of inversions impossible.

Sex determination commonly involves an upstream sex-determining gene located on a sex chromosome that acts on downstream cascades of genes to result in either male or female development [1, 21]. Interestingly, upstream sex determining genes in animals appear to have higher evolutionary rates than downstream genes that often are highly conserved [21]. It is unclear if a similar model applies to plants, where sex determining genes have thus far been identified in only a handful of species [22]. Two linked, sex-determining genes, a male promoter and a female suppressor, have been reported in asparagus [23] and in kiwifruit [24]. Sex determination by a single gene, in contrast, has been found in persimmon [25] and in poplar [26]. In both of these cases, Y-encoded small RNAs were shown to repress sex determining genes (*MeGI* and *ARABIDOPSIS RESPONSE REGULATOR 17*) via methylation, with downstream effects on the expression of conserved genes involved in floral identity [25–27]. Interestingly, genes related to the cytokinin signaling pathway, such as *ARR17*, and downstream MADS-box genes are generally overrepresented in candidate genes for plant sex determination [28].

In the Salicaceae family, both ZW and XY sex determination has been reported within *Populus* (chromosome 14 or 19) and within *Salix* (chromosomes 7 or 15), suggesting numerous turnover events [6, 7]. In *Populus tremula* with XY sex determination (chromosome 19), CRISPR/Cas9-induced loss of function mutation of *ARABIDOPSIS RESPONSE REGULATOR 17* (*ARR17*) switches females to males [26]. Genetic network analysis revealed that *ARR17* is a dominant suppressor of maleness, downregulating the genes involved in floral organ identity such as *USUAL FLORAL ORGANS* (*UFO*) and *PISTILLATA* (*PI*) [27]. In *P. alba* and *S. purpurea* with ZW sex determination systems, *ARR17* is present on the W

chromosome but not on the Z chromosome and thus suggested as a candidate sex determining gene, although functional validation is currently not available [7, 29]. In *Salix* and *Populus* species with XY sex determination, inverted repeats of *ARR17* were detected on the sex-linked region of the Y chromosome, suggesting a possible silencing mechanism of *ARR17* by small interfering RNA. Whereas sex-linked regions (SLR) in described in *Populus* are small (0.2–1.7 Mb) and only one common homologous gene, *ARR17* [7], the SLR in *Salix* varies greatly in size, from 1.8 to 6.7 Mb, and typically contains many candidate sex determination genes [5, 6, 30, 31]. It thus remains unsolved whether there is a common upstream sex determining gene in the *Populus* and *Salix*.

In this study, we analyze the evolution of sex chromosomes using a new haplotype-resolved assembly of the arctic-alpine dwarf shrub *Salix herbacea* and publicly available data from the Salicaceae family. We identified chromosome 15 as the sex chromosome with female-heterogametic (ZW) sex determination in *S. herbacea*. Our analyses suggest that a large (5Mb) species-specific inversion on the W-chromosome recently incorporated pseudoautosomal sequence (PAR) into the sex-linked region, supporting a model of stepwise recombination suppression. We further propose a new candidate gene for sex determination in ZW *Salix* species: W-specific inverted partial repeats of the male essential gene *PISTILLATA* (*PI*) may act to suppress *PI* in females, superseding the role of the upstream *ARR17* in sex determination.

Results

Genome assembly and annotation

Leveraging the long-read capabilities of PacBio HiFi sequencing and the high-resolution chromosome conformation capture provided by Hi-C, we assembled the genome of a female individual of *Salix herbacea* (Table 1, Additional file 3: Table S1). We generated the initial draft assembly using PacBio HiFi long reads. Both haplotypes were anchored by Hi-C physical mapping, which showed a well-organized pattern of contact interactions along the diagonals within pseudo-chromosomes (Additional file 1: Fig. S1-3). After manual curation, haplotype 1 and 2 comprised 328 Mb and 331 Mb, with more than 98% anchored into 19 pseudo-chromosomes (Table 1). BUSCO assessment revealed a high completeness of over 97% (Table 1, Additional file 3: Table S2). The accuracy and continuity of the genome assembly were further verified by nearly complete mapping coverage of genomic short reads (Table 1). We annotated 31,082 protein-coding genes on *S. herbacea* haplotype 1 and 30,945 on haplotype 2 (Table 1). 85% of the genes were functionally annotated in public databases [32]. Repetitive sequences comprised over 49% of both haplotype assemblies, including 21% retrotransposons (Additional file 3: Table S3).

ZW sex-determination system on chromosome 15

Analysis of read depth using short-read data of 10 females and 10 males as well as Pool-Seq data from 50 females and 50 males revealed regions with consistent and strong sex-biased coverage in a large region on chromosome 15 spanning over 8 Mb (Fig. 1b, Additional file 1: Fig. S4-5). From 3.2 Mb to 11.6

Mb on haplotype 1, coverage was female-biased, whereas, in the corresponding region on haplotype 2 (from 3 Mb to 11 Mb), coverage was predominantly male-biased, as expected for a sex-linked region (Additional file 1: Fig. S4-5). We also detected regions with sex-biased coverage on chromosomes 4, 6, 7, 12, 13, 18, 19 both in pooled and individual data (Additional file 1: Fig. S4-5). However, these regions exhibited either only weak bias (chromosomes 4 and 7), a mix of male and female coverage bias on both haplotypes (chromosomes 6, 12, 18 and 19), or female coverage bias on both haplotypes (chromosome 13).

The chromosome 15 region with biased coverage had high between-sex genetic differentiation (F_{ST}) as compared to autosomes, particularly at both ends of the region (3.2-6 Mb and 10.9-11.6 Mb on haplotype 1), where we detected exceptionally high levels of between-sex F_{ST} with values exceeding the genome-wide 99^{th} percentile (Fig. 1c, Additional file 1: Fig. S6-7)). We obtained similar F_{ST} results for chromosome 15 using individual re-sequencing and Pool-Seq data (Additional file 1: Fig. S6-7). However, for the individual re-sequencing dataset, some regions with sex-biased coverage on chromosomes 7 and 12 also exhibited weaker signals of high between-sex F_{ST}. Those signals were not detected when comparing female and male pools with more individuals (Additional file 1: Fig. S6-7).

The chromosome 15 region with biased coverage and high between-sex F_{ST} was further associated with a striking excess of heterozygous sites in females compared to males (Fig. 1e, Additional file 1: Fig. S8-9). Integrating these findings, we thus find support for ZW sex-determination in *S. herbacea* where haplotype 1 corresponds to the female-specific W-haplotype and haplotype 2 to the Z haplotype.

Structural variation on the sex chromosomes: a large inversion and two female-specific regions

W- and Z-haplotypes had similar lengths and gene content (Table 2). In an alignment of the two haplotypes, all chromosomes aligned along the diagonal, except for a fragmented gap (2.7-6.1 Mb on the W haplotype) and an adjacent large inversion within the sex-linked region of chromosome 15 (Fig. 1a, Additional file 1: Fig. S3), revealing sex-linked structural variation. The inversion on the W-SLR (5.1 Mb, 6.1-11.2 Mb) is considerably larger than the corresponding region on the Z haplotype (3Mb, 7.5-10.5 Mb). We can exclude that the inversion is an artifact because the breakpoints of the inversion were covered by continuous PacBio HiFi long reads and Illumina short reads (Additional file 2: Fig. S10).

Through a detailed whole-genome scan of coverage in 10 females and 10 males, we identified 700 W-hemizygous regions (101,094 bp in length) on chromosome 15, present in all females and absent from all males. No Z-hemizygous sites were detected. W-hemizygous sequences were detected almost exclusively in two regions at both ends of the proposed W-SLR (3.3-5.9 Mb and 10.9-11.3 Mb, Fig. 1d) and correspond to the two high F_{ST} regions described above (Fig. 1c). We refer to these two regions with female-specific sequences as FS1 and FS2. While FS1 mapped outside of the inversion, FS2 occurred

within the inversion. We thus describe five regions of the W-chromosome: a first pseudoautosomal region (PAR1), FS1, the inversion (INV), FS2 and a second pseudoautosomal region (PAR2, Fig. 1).

Degeneration of sex-linked regions on chromosome 15

Recombination arrest on sex chromosomes allows for sequence degeneration over time, such as gene loss, the accumulation of repetitive sequences and disruptive mutations [2]. Compared to the PARs, the W-SLR was characterized by a strong enrichment in LTR sequences up to 50% with a concomitant decrease in gene content, particularly in the INV region (Fig. 2a, Table 2, Additional file 3: Table S5). LTR insertion times along the W-chromosome ranged from over 14.6 million generations to less than 0.5 million generations ago (Fig. 2c). A clear clustering of LTR insertions, 1-2 million generations ago, occurred at the 5' end of the inversion (~11.4 Mb) suggesting that LTRs might have driven the inversion event recently (Fig. 2c). The proportion of the sequence constituting highly degenerated genes (pseudogenes) was much higher in FS1 and FS2 (0.011, 0.007) than in the inversion (INV, 0.003) and the two PAR regions (0.004 and 0.003) (Fig. 2a, Additional file 3: Table S5). We further annotated and classified mutations on chromosome 15 within our study population (20 individuals) according to their potential impact on gene function (Fig. 2b). In the PARs, mutations with disruptive effects were 20 times less common than mutations with moderate effects. In the INV and FS1 regions, the ratio between moderate and disruptive mutations was reduced to 17 and 11, respectively. In contrast, the FS2 region had a strongly increased incidence of disruptive mutations, leading to a ratio of moderate to disruptive mutations of only 2. Together, these results suggest a stronger degeneration in the female-specific regions (FS1 and FS2) than in the INV region which differs from the PARs mainly by higher LTR and lower gene content (Table 2). The FS2 region was particularly degenerated in terms of disruptive mutations, suggesting that FS2 might be the oldest region of the W-SLR.

We further estimated the loss of ancestral genes from the SLR, defined as genes present in *S. herbacea* and also found on the same chromosome in at least one closely related species (*S. purpurea*, *S. suchowensis*, *S. arbutifolia* and *P. trichocarpa*). 441 ancestral genes were present on either the Z-SLR, W-SLR or both. Of these, 102 were absent from the W-SLR (23%) and 65 were absent from the Z-SLR (15%) (Table 2).

Divergence of sex-linked regions on chromosome 15

Sequence degeneration and recombination arrest promote progressive divergence of the two sex-linked haplotypes from each other [2]. We estimated sequence divergence between 1456 ZW-homologous genes based on synonymous substitution rates (K_s). Genes on FS1 and FS2 were highly divergent between haplotypes, while divergence of genes located in the INV region within the W-SLR and in the PARs at both ends of the sex chromosomes was similarly low (Fig. 2d). Among the 13 ZW-homologous genes in FS2, one gene (*DUF4219*) with a K_s value greater than 1.5, encodes an LTR protein, which may be related to an LTR-driven inversion event. Using phylogenetic trees of 409 ZW-homologous genes with

single-copy orthologs in *S. purpurea* and using *P. trichocarpa* as an outgroup, wefound that the great majority of the gene phylogenies followed the species phylogeny (Fig. 2e, Fig. 5b). In the phylogenies of16 genes, W and Z haplotypes clustered across both *Salix* species; 15 of these genes were in the W-SLR region (Fig. 2d, e). This indicates that the divergence of the W and Z haplotypes of *S. herbacea* occurred mainly after the speciation event separating *S. herbacea* and *S. purpurea,* dated to 14.5 million years before present in our analysis (Fig. 5b).

Structural variation of sex chromosomes

A synteny plot of the two sex chromosomes of *S. herbacea* supports the inversion within the W-SLR as described above, as well as the overlap of FS2 with the inversion (Fig. 3a, b). The two other haplotype-resolved *Salix* genomes, *S. pupurea* from the same clade as *S. herbacea* (*Vetrix*), also with ZW sex determination, and *S. arbutifolia* from the *Salix* clade with XY sex determination, generally show high synteny with *S. herbacea* in the PAR regions but much-reduced synteny in the SLRs, indicating massive structural divergence between species in the SLRs (Fig. 3a, b). *Salix pupurea* and *S. arbutifolia* do not share the large inversion within the SLR of *S. herbacea*. About half of the *S. herbacea* inversion on the W-SLR mapped to the PAR of *S. purpurea* and to a small region at the upstream end of the *S. purpurea* W-SLR, as well as to a small region within the PAR of *S. arbutifolia* (Fig. 3a, b). These findings suggest that the W-specific interversion in *S. herbacea* recently incorporated part of the PAR region into the middle of the W-SLR, splitting the ancestral W-SLR into two parts (FS1 and FS2, Fig. 3c).

Intriguingly, the FS1 of the W-SLR of *S. herbacea* exhibited crossed lines of synteny with the corresponding region of the *S. herbacea* Z-SLR, the W-haplotype of *S. purpurea*, and the X haplotype of *S. arbutifolia* (Fig. 3a, b) suggesting an ancient, species-specific inversion event involving the FS1 of *S. herbacea*. The FS2 region of the *S. herbacea* W-SLR exhibited several translocated alignments (10.7-11.2 Mb) with the middle of the *S. purpurea* W-SLR (6.2-6.6 Mb) and the *S. arbutifolia* X-SLR (9.4-9.7 Mb) indicating that re-arrangements formed this region.

Evolutionary history of sex-linked genes

Using an evaluation of sex-linked variation, we detected 46 sex-linked genes on FS1 and FS2 of the W-SLR, corresponding to 36 homologous functional genes annotated in *Arabidopsis* (Fig. 4, Table 3, Additional file 3: Table S6). Of these, 13 genes contained W-hemizygous sites that were present in all ten female individuals but not in any of the ten male individuals; 26 genes had W-specific polymorphisms that were heterozygous in all females and homozygous in all males (ZW-ZZ pattern); 7 additional genes harboured both types of polymorphisms. Whereas sex-linked genes in the FS1 region were somewhat scattered, all 19 genes in the FS2 region were located consecutively on the upstream side (10.85-11.2 Mb), representing half of all genes in that region (Fig. 4). Three W-specific alleles in FS2 corresponded to disruptive mutations (frameshift mutations or the loss of stop codon) in homologs of the genes MEE29, GSTT3 and CRF9 in *Arabidopsis* [33].

To investigate the evolutionary trajectories of sex-linked genes on the W-SLR of *S. herbacea*, we analyzed gene synteny with six other species of the Salicaceae family (Fig. 4). 18 sex-linked genes of *S. herbacea* were homologous to genes on chromosome 15 in other *Salix* and *Populus* species with different sex-determining systems, providing strong evidence for an ancestral origin (Fig. 4a-c, Additional file 3: Table S6). Interestingly, these ancestral homologs contained not only genes related to floral development, such as a B-box type zinc finger protein (*BBX15*) [34] and *CRF9* [35], but were also enriched for male-function genes, such as *HIK*, *PAPS1*, *ENOLDL6*, *ARID*, *ARP4* that are involved in pollen and anther development [36–40] (Fig. 4d, Additional file 3: Table S6). 15 genes with various functions had homologs in one or two *Salix* or *Populus* species (Additional file 3: Table S6). Another 13 genes were identified as *S. herbacea*-specific sex-linked genes, including two DEAD/H box RNA helicase-like genes (*MEE29* and *NOF1*) that are involved in female gametogenesis [41, 42], *SCD1* and *SWEETIE* that are associated with inflorescence formation [43, 44]*, MiP1* that functions as a flowering activity switch [45, 46](Fig. 4d, Table 3, Additional file 3: Table S6).

We further examined the W-hemizygous regions and W-specific alleles in *S. herbacea* in three other ZW species, *S. viminalis, S. purpurea,* and *S. suchowensis*, by comparing female and male individuals. 215 of 101,094 hemizygous sites and 3 of 573 W-specific alleles were present in multiple ZW species and were located in one gene, *ARP4*, and in the intergenic regions of the gene block in the FS2 region (Fig. 4d, Additional file 4: Dataset S1, Additional file 5: Dataset S2). In *S. viminalis*, the species most closely related to *S. herbacea* (Fig. 5b), we detected more shared W-specific alleles (71bp, 6 genes) and hemizygous sites (12779bp, 5 genes), resulting in ten common sex-linked genes (Fig. 4d).

In addition to sex-linked genes on the W-chromosome, we also scanned the Z chromosome for candidate male-function genes degenerated on the W chromosome. We identified 94 pseudogenes on the W chromosome, with 55 complete genes specific to the Z chromosome, corresponding to 26 different functional genes (Additional file 3: Table S7). One of these is the well-known gene *PISTILLATA* (*PI*) which is involved in male organ identity in many species, including *Populus tremula* [25–28]. For most of the remaining genes, the connection to floral development is unclear (Additional file 3: Table S7) .

Previously proposed sex determination gene: *ARR17*

In *Populus* and some *Salix* species, the *ARABIDOPSIS RESPONSE REGULATOR 17* (*ARR17*) has been proposed as the single sex-determining gene [6, 26]. We did not detect any *ARR17* genes or partial duplicates on chromosome 15 in *S. herbacea*, but two hemizygous *ARR16*-like genes were found in FS1 of the W chromosome (Fig. 5a). *ARR16* and *ARR17* are a homologous type-A cytokinin gene pair in the *RR* gene family [47]. However, sequence identity between *ARR17* and *ARR16* or other *RR* genes was low, 68-71% between *ARR17* and *ARR16* (Additional file 6: Dataset S3). We searched for complete or partial homologs of *ARR16/17* in 12 other *Salix* and *Populus* species, using all exons or only the first exon. Two *ARR17*-like genes and one *ARR16*-like gene on chromosome 19 were ancestrally shared among *Salix* and *Populus*. Of the *Salix* species, only *S. purpurea* and *S. brachista* had *ARR17* genes on chromosome 15,

while the other six *Salix* species lacked *ARR17* homologs on chromosome 15 (Fig. 5b). Partial duplicates of *ARR17* were found on chromosome 15 in 8 *Salix* species (W, Z or Y haplotypes) but were completely lacking in *S. herbacea* and *S. dunnii* (Fig. 5b). An additional *ARR16* gene on chromosome 15W was only detected in *S. herbacea* and in the female genome assembly of *S. viminalis* (Additional file 3: Table S8).

PISTILLATA (*PI*) is a new candidate sex determination gene

We found one homolog of an intact *PISTILLATA* (*PI*) gene on the Z-SLR of chromosome 15, and 16 partial duplicates of the first exon on the W-SLR, including 7 pairs of inverted repeats in the FS1 and FS2 regions and two unpaired partial duplicates (Fig. 5a). This suggests a possible mechanism of suppressing the male essential *PI* by small interfering RNA mediated gene silencing [48]. We searched for complete or partial homologs of *PI* genes in 9 *Salix* and 2 *Populus* species by processing all exons or only the first exon. All these species, including *S. herbacea,* shared conserved copies of *PI* on chromosomes 2 and 5 (Additional file 1: Fig. S11). We observed different numbers of partial duplicates on chromosome 15 across all five ZW genome assemblies except for *S. viminalis*, possibly due to a poor assembly from short reads. Partial duplicates of *PI* were not found in any male assemblies, such as the ZZ assembly of *S. udensis* or in species with XY sex determination. The inverted partial repeats of *PI* were also detected on chromosome 15W of *S. purpurea* and *S. brachista* (Fig. 5b,Additional file 3: Table S8). In a phylogeny of *PI* genes, the full-length *PI* on the Z-SLR of *S. herbacea*, 6 pairs of inverted repeats and the unpaired partial duplicates clustered with the *PI* homologs on chromosome 5, while one pair of inverted repeats clustered with the *PI* homologs on chromosome 2 (Additional file 2: Fig. S11), indicating that inverted repeats originated from both chromosomes.

Discussion

A large inversion causes a three-fold extension of the sex-linked region

Using a haplotype-resolved *de novo* genome assembly and population whole-genome re-sequencing data, we have identified female heterogamety (ZW) on chromosome 15 as the sex determination mechanism of dwarf willow, *Salix herbacea*. The sex-linked region (SLR) comprised 8.4 Mb, 53% of chromosome 15, making it the largest reported SLR in the Salicaceae family (*Salix* and *Populus*). Y- or W-linked regions in other *Salix* species cover between 17% and 43% of the sex chromosome [5, 30], but only 2–10% in *Populus* [49–52]. The *S. herbacea* W-SLR contained a large (5 Mb) inversion not found in other *Salix* species. Two regions with a high incidence of female-specific sequences were detected, one at the 3’ end of the W-SLR (FS1, 2.6 Mb), adjacent to the inversion, and the other at the 5’ end of the W-SLR, as part of the inversion (FS2, 0.4 Mb). The inversion region between FS1 and FS2 (INV) had high synteny with the pseudoautosomal regions (PAR) of *S. purpurea* and *S. suchowensis*, but not with the sex-linked regions of these species. This suggests that the inversion incorporated a PAR region into the SLR, extending its length nearly threefold. Two lines of evidence indicate that the inversion on the W-SLR is

recent: First, synonymous sequence divergence (K_s), between ZW homologous genes as well as the incidence of disruptive mutations were low in the INV region and the peudo-autosomal regions, and higher in the female-specific regions FS1 and FS2. Second, we found support for a recent burst of TE insertions close to the breakpoint of the inversion. Our data thus support the model of stepwise evolution of recombination suppression via inversions as a proximate mechanism [2, 10, 11], in contrast to studies in other *Salix* species [6, 30, 31]. While inversions are commonly found within sex-linked regions [17, 18, 53, 54], we report one of the few cases showing a clear link between a recent inversion and the extension of the recombination-suppressed region on sex chromosomes [11].

Degeneration of the sex-linked region

Plant sex chromosomes are often described as young, with limited degeneration, small SLRs and frequent turnover [55]. This has also been supported by findings in several other *Salix* and *Populus* species [6, 26]. Dwarf willow, *S. herbacea*, does not fit this pattern. The INV region on W-SLR is highly collinear with the Z haplotype and exhibits a prominent accumulation of repetitive elements, likely representing the earliest stages of degeneration. FS1 and FS2, in contrast, show a breakdown of synteny between W and Z haplotypes. Synteny breakdown between sex chromosomes also emerges from other studies with haplotype-resolved assemblies and likely reflects advanced sequence degeneration rather than difficulties with assembling repeat-rich regions [17, 53, 54]. The W-SLR of *S. herbacea* has lost 23% of the ancestral genes, and the female-specific regions within the W-SLR have accumulated disruptive mutations and pseudogenes, providing clear evidence of sequence degeneration. Sequence divergence in the remaining ZW homologous genes reached K_s values of 0.05 in FS2, higher than in other *Salix* species [6, 30]. *Salix herbacea* had a similar degree of gene loss and differentiation (K_s) as reported for the younger SLR regions in *Humulus* and *Cannabis* [56], *Rumex hastatulus [17]* and *Silene latifolia [53]*, as well as in primates [54], sticklebacks [20] and birds [16]. Phylogenetic analyses of single-copy orthologs among the ZW homologous genes suggest that divergence between the sex chromosomes of *S. herbacea* occurred mostly after the split from *S. purpurea*, which has been dated to nearly 14–20 million years before the present, using analyses with fossil calibration [57]. This is a comparatively long divergence time, considering that the heteromorphic sex chromosomes of the white campion (*Silene latifolia*) have evolved within the last 11 million years [53]. However, divergence time estimates are difficult to compare as they depend on mutation rates and generation times. This is an issue particularly for S. *herbacea* since it relies on both clonal and sexual reproduction and is very long-lived with age estimates exceeding 2000 years [58]. Nonetheless, our data suggests that sex chromosome evolution in the Salicaceae family involves considerable degeneration over extended periods, in accordance with the classic predictions for sex chromosome evolution [1, 2].

Candidate sex determination mechanism: ARR17 replaced by regulation of downstream PISTILLATA by inverted repeats

We detected a conspicuous pattern for the gene *PISTILLATA* (*PI*) in *S. herbacea. PI* is a conserved gene essential for the production of male flowers in many species, including *Populus tremula [25–28]*. Full-

length sequences of *PI* were located on two autosomes (chromosomes 2 and 5) and on the Z-SLR. *PI* pseudogenes were present in the FS1 and FS2 regions of the W-SLR in the form of seven pairs of partial *PI* copies arranged as inverted tandem repeats. This peculiar arrangement suggests the formation of double-stranded RNA hairpin structures that can produce siRNA acting to silence the *PI* coding RNAs by sequence complementarity [59, 60]. This mechanism is prevalent in angiosperms [61] and has been suggested as a sex-determination mechanism in persimmon [25] and *Populus* [26]. Suppression of stamen development by interfering RNA targeting *PI* has further been demonstrated experimentally in legumes [48]. This raises the possibility that hairpin-like duplicates of *PI* on the W chromosome silence *PI* on autosomes and Z chromosomes in females, suppressing the development of male organs.

Interestingly, the upstream sex determination gene *ARR17* in *Populus*, also suppresses *PI*, leading to the development of females [26–28]. In XY males of *P. tremula, ARR17* is itself suppressed by small RNAs generated from partial duplicates of *ARR17* on the Y chromosome. This double suppression mechanism has also been suggested for *Salix* species with XY sex determination [5, 6], but *ARR17* involvement in sex determination remains unclear for *Salix* species with ZW sex determination [62, 63]. In *S. herbacea*, we detected *ARR17* only on chromosome 19 and partial duplicates were not found. We did find another *RR* gene, *ARR16*, on the W-SLR; however, this gene was also present on autosomes, and there is no evidence that *ARR16* has the same function as *ARR17* in *Populus*, given the strong sequence divergence between *ARR16* and *ARR17*. If the autosomal *ARR17* in *S. herbacea* had the same function as in *P. tremula [26]*, all individuals would experience *PI* suppression and develop as females; this mechanism is therefore unlikely. Instead, our findings suggest that the role of *ARR17* in regulating the downstream *PI* is superseded by *PI* silencing mediated by *PI* pseudogenes on the W-SLR, which leads to a suppression of male development in ZW females only.

Autosomal *PI* copies on chromosomes 2 and 5 are found in all *Salix* and *Populus* species. Inverted partial duplicates of *PI*, however, are only found in *Salix* species with ZW sex determination but not in those with XY sex determination. In *S. herbacea*, *PI* sex determination could thus be ancestral. An increase in the number of partial duplicate pairs to seven could, however, be species-specific. Conversely, autosomal *ARR17* genes together with potentially male-determining partial *ARR17* copies on the Y chromosomes were common, whereas only some of the ZW *Salix* species had partial *ARR17* copies. This leads us to the tentative suggestion that a shift from *ARR17*-controlled sex determination to *PI*-controlled sex determination could be associated with the transition between XY sex determination and ZW sex determination [57].

Sex-differentiation found in both ancestral and translocated genes

A search for sex-linked genes with W-hemizygosity or fixed allelic differences revealed 46 genes within the W-SLR. Forty percent of these genes had ancestral homologs on chromosome 15 in *Salix* and *Populus* species with different sex determination systems. Interestingly, we detected a block of such ancestral, sex-linked genes in FS2, the most differentiated and presumably oldest part of the W-SLR.

Many of these genes were enriched for male-specific functions, such as anther and pollen development, for example, *CRF9*, *ENOLDL6*, *ARP4* [35, 36, 64, 65]. Some of these genes shared sex-specific polymorphisms with other ZW *Salix* species, especially the male fertility-related gene *ARP4*. These ancestral genes with male-specific functions may have originated even before sex determination on chromosome 15 was established. The sex-specific variants of these genes in *S. herbacea* might result from sequence degeneration on the W-SLR in females where there is no selection on male-specific functions.

Additional W-hemizygous genes in the FS2 gene block likely were recruited from autosomes after the split of *S. herbacea* from *S. purpurea*, as they had homologs only on autosomes in *S. purpurea* and other *Salix* species. These genes often had functions associated with female gametogenesis or flowering, such as *MEE29*, *NOF1*, *SWEETIE*, and *MIP1* [45, 46]. This finding could thus represent the capture of female beneficial genes onto the W-SLR, one of the processes that can lead to selection for recombination reduction [10, 12]. We further detected sex-specific differences in *DEAD* genes that have functions related to female meiosis. This finding is fascinating as *Salix* species, including *S. herbacea*, frequently have female-biased sex ratios that may arise through genetic rather than ecological mechanisms [66, 67]. In theoretical studies, sex ratio selection and meiotic drive have been shown to lead to sex chromosome turnover [68], offering another possible explanation for the processes behind the evolution of sex chromosomes in *Salix*.

Conclusions

With this study, we contribute to the understanding of sex chromosome evolution. We show that the dwarf willow *Salix herbacea* had an unexpectedly large sex-linked region (ZW) comprising more than half of chromosome 15. Part of this region has undergone considerable sequence degeneration over a long period of 14-20 million years, comparable to other plant species with large sex chromosomes [17, 53, 56]. Our study provides one of the few examples of the proximate mechanisms producing large sex-linked regions: a large inversion (5Mb) on the W-chromosome recently incorporated pseudoautosomal sequence into the sex-linked region, resulting in stepwise recombination suppression.We further identified 46 sex-linked genes on the W chromosome that are possible targets of sexually antagonistic selection, one of the theoretically inferred evolutionary processes allowing for the fixation of inversions. Our results suggest the involvement of conserved genetic pathways in sex determination across the closely related genera *Salix* and *Populus* but with differences in the regulation mechanisms. Given our data, it is unlikely that the sex determination gene proposed for *Populus, ARR17 [26]*, a regulator of male essential gene *PISTILLATA (PI),* is involved in the sex determination of *Salix* species with ZW sex determination. Instead, we propose the down-regulation of *PI* in females by W-specific pseudogenes of *PI* as a new candidate mechanism for sex determination in these species. Our findings align well with results in animals where changes in upstream regulators of sex determining pathways were common, whereas downstream genes were more conserved [21]. Overall, we show that dynamic sex chromosome evolution in *Salix* conforms to classic predictions for sequence degeneration [1, 2] and involves a

combination of ancestrally retained and newly transposed genes, as well as stepwise recombination suppression.

Methods

Plant material

In a natural population of *Salix herbacea* near Jakobshorn, Switzerland (46.7720N, 9.8554E, 2535 m above sea level), we collected one female plant (*F005*) to be used for whole genome sequencing and sampled leaves from 60 female and 60 male individuals. We cultivated *F005* under controlled conditions and collected fresh tender leaves continuously to gain enough material for whole genome sequencing. Leaves were immediately frozen in liquid nitrogen, and stored at -80℃. The leaf material of the population collection was dried with silica gel (VWR / Avantor, Switzerland). For transcriptome sequencing, we collected five fresh tissues from two further individuals cultivated from seeds of the same populations: leaf, stem, root, female catkin and male catkin.

DNA and RNA extraction, library construction and sequencing

Long read sequencing

High molecular weight DNA for PacBio HiFi sequencing was extracted from individual F005 using a modified protocol by Workman et al. [69]. The sequencing library for PacBio HiFi sequencing was constructed using the SMRTbell Express Template Prep Kit 2.0 according to the manufacturer's instructions (PacBio, USA) and sequenced on a PacBio Sequel II system at the Uppsala Genome Center, National Genomics Infrastructure Sweden.

Hi-C

The nuclei of the same plant individual used for long-read sequencing were extracted from the lysed cells using gradient density centrifugation, and the nuclei were then digested using restriction endonucleases. Hi-C library preparation was performed after DNA cross-linking, restriction endonuclease digestion (*HindIII*), end-repairing, end-ligation, DNA capturing and purification. The Hi-C library was constructed and sequenced on the Illumina NovaSeq 6000 platform by Biomarker Technologies (BMK) GmbH (China).

Transcriptome sequencing

The total RNA of five tissues was isolated using the RNeasy Kit (Qiagen, Hilden, Germany) and subsequently pooled based on RNA quantity for library construction. The concentration and RNA integrity were accessed using the Bioanalyzer Agilent 2100 (Agilent Technologies, Böblingen, Germany). One mRNA library was prepared with polyA enrichment and sequenced on NovaSeq PE 150 by Novogene (UK).

Short-read sequencing of pools and individuals

The genomic DNA of 60 female and 60 male individuals of one population was isolated using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). DNA concentration was measured using Quant-iT PicoGreen dsDNA kit (Thermo Fisher Scientific, Waltham, MA, USA) and a Fluorescence Microplate Reader (ThermoFisher, Waltham, MA, USA); the quality of DNA was assessed using a Nanodrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). A female and a male pool with 50 individuals each was prepared based on equimolar DNA concentrations to create two pooled DNA libraries. Additionally, 20 individual libraries were prepared separately for ten females and ten males. In total, 22 Illumina TruSeq PCR-free libraries were constructed using unique dual indexes, targeting an insert size of 350 bp. The sequencing was performed on NovaSeq 6000 by SNP&SEQ Technology Platform in Uppsala and National Genomics Infrastructure in Stockholm (Sweden).

Chromosome-scale assembly with HiFi and Hi-C data

For the female individual F005, high-quality HiFi Circular Consensus Sequencing (CCS) reads were used for an initial contig assembly using HiFiasm v0.15.4 [70] with default settings. We then purged the redundant haplotypic duplications using purge_dups v1.2.5 [71].

We filtered HiC reads using Fastp v0.23.1 [72] to remove the adapters and low quality reads (--qualified_quality_phred 20, --length_required 100), trim reads in front and tail (--cut_front --cut_tail --cut_window_size 4 --cut_mean_quality 20) and correct mismatched base pairs. The filtered HiC reads were mapped to the initial assembly by Juicer v1.6 [73] to generate HiC map bins. We followed 3D-DNA pipelines [74] to scaffold the genome and eliminate misjoins with three iterative steps. The output assembly of each step was visualized and evaluated through Juicebox [75]. The assembly from the early step was retained as the candidate 'mega-scaffold' with the least fragments and errors in breaking contigs. We manually curated it to correct mis-joins, order, and orientation. The revised assembly was polished, split, sealed and merged in the post-process by 3D-DNA [74].

We applied the same processes to both haplotypes. In the end, we manually modified the order and orientation of the chromosomes by aligning both haplotypes to keep them parallel. The chromosome assignment was done according to the 19 chromosomes of *Salix purpurea*.

The final assembly was evaluated by mapping the HiFi reads and Illumina paired-end reads (see below), using Minimap v2 [76], BWA v0.7.17 [77], respectively. BUSCO v5.0.0 [78] was used to assess the completeness of the genome assembly.

Genome annotation

We identified repetitive sequences in the genome assembly using RepeatMasker v4.1.5 [79] based on repeat sequences registered in Repbase [80], species-specific repeat datasets from *Populus tricocarpa*

[81], *de novo* repeat libraries built with RepeatModler v2.0.3 [82] and LTR_retriever [83]. To further annotate unknown repeats, we adopted four to six rounds of self-classification and the database of *Arabidopsis* [84] by Repclassifier v1.1[82]. Tandem repeats were predicted using trf v4.10.0 under the recommended parameters [85].

Based on the repeat-masked genome, gene prediction was performed following automated pipelines using BRAKER [86] with GeneMark-ETP and AUGUSTUS for three separate rounds: homology-based only [87], RNA-seq only [88, 89], and combining homology-based and RNA-seq [90–94]. Transcriptome sequences from five tissues of *S. herbacea* were mapped to the genome assembly using HISAT2 [95] and SAMtools [96]. For the homology-based prediction, protein sequences of six species (*Arabidopsis thaliana*, *Salix purpurea*, *S. brachista*, *S. dunnii*, *Populus deltoides*, *P. trichocarpa*, Additional file 3: Table S4) were aligned to the assembly using BLAST [97]. High-quality alignments were selected and integrated using TSEBRA [98]. The function of the predicted genes was annotated using three strategies: 1) based on the homologous genes using EGGNOG-MAPPER [32]; 2) based on domain similarities using InterProScan [99]; 3) based on sequence identity with the gene database of *Arabidopsis* using BLAST [97].

Identification of sex-linked regions

Genotype calling of females and males

Whole-genome short reads for both pools and individuals were preprocessed with Fastp [72] using the following parameter settings: -q 20 -u 20 -l 100 -y 30 -g 10 -x 10 -p 20, and quality control was performed using FastQC [100]. We mapped clean paired-end reads to our genome assembly using BWA v0.7.17 [77]. The mapping output was filtered and sorted according to coordinates using SAMtools v1.17 [101]. Duplicate reads from sample preparation, amplification and sequencing were detected and marked using PicardTools [102]. Variants of individual samples were called and genotyped using GATK v4.1.4.1 [103] and split into SNPs and INDELs. We applied hard filtering pipelines to SNPs (QD<2.0, QUAL<30.0, SOR>3, FS> 60, MQ<40, MQRankSum<-12.5, ReadPosRankSum<-8, DP<30) and INDELs (FS>200, QD<2, QUAL< 30, ReadPosRankSum<-20, DP<30) separately. High-quality SNPs were filtered using VCFtools [104] with the criteria: MAF>0.1, minDP=10, maxDP=100, max-missing=1. For the pooled samples, the variants were called using SAMtools v1.17 [101] and PoPoolation2 [105] were then used to call SNPs with counts of at least 10 for the alternate allele and a depth between 50 and 1,000× at the position concerned. All SNPs with coverage lower than 10 or higher than 500 were discarded.

Comparing females and males

To identify sex-linked regions in *S. herbacea*, we utilized depth-based, genetic differentiation-based and heterozygosity-based approaches. Here, a higher mapping depth in one sex indicates sex-specific sequences on one of the haplotypes, especially if combined with high heterozygosity in the same sex.

High genetic differentiation between sexes is expected in or near sex-linked, recombination-suppressed regions [55].

We calculated mapping depth in 10kb windows with Mosdepth [106] using all bases with mapping quality over 20. After normalization with the overall mean depth in each pool, we calculated the proportion of females reads calculated as F/(F+M). We applied the same approach to compare depths between the individually resequenced individuals (10 females and 10 males). We calculated genetic differentiation as weighted Weir and Cockerham's F_{ST} between sexes in 10kb windows with 10kb steps for female and male pools, as well as for individuals using VCFtools [104, 107]. The proportion of heterozygous sites in 10kb windows of each re-sequenced individual was calculated and averaged within the sexes.

In addition to the window-based analyses above, we also conducted a genome-wide search for sex-limited regions (present in one sex but not in the other) at single base pair resolution. For this analysis, we used BEDtools [94] and the individual re-sequencing data from 10 individuals of each sex. We identified these regions as absent from all 10 individuals of one sex and present in all individuals of the other sex. To account for mis-mapping in the short read data, we scored sites with up to 5 reads, corresponding to the lowest 20 percent of the depth distribution as absent. W-hemizygous regions were strictly defined as those properly mapped in all ten females and which showed no proper mapping in any of the ten males. Conversely, Z-hemizygous regions were identified based on the opposite criteria.

Structural variation between haplotypes

To detect structural variation between sex chromosomes, we aligned the two haplotypes from the genome assembly using the MUMmer v3.23 pipeline [108]. The maximum gap between two adjacent matches in a cluster was set to 500bp, and the minimum length of a cluster was 100bp. We subsequently applied hard filtering and kept only unique alignments with identity >89% and alignment length >1000bp.

Degeneration of sex chromosomes

Degeneration of genes and proportion of repetitive elements

Using the annotation described above, we estimated the number and sequence proportion of genes along the two PARs, the SLR of haplotype 1 of chromosome 15, as well as the proportion of LTRs and pseudogenes.

We annotated pseudogenes using a modified version of the pipelines of Xie et al. [109], including 1) masking of repetitive sequences and coding regions of the reference genome; 2) identification of intergenic regions with sequence identity to the annotated proteins of the same haplotype using tBLASTn [97]; 3) filtering of matches with identity <40%, match length <30 amino acids; 4) linking adjacent regions within 50bp distance on the chromosome when matched to the same protein; 5)

stacking homologous genes on the same location of the chromosome with at least 80% sequence overlap. In this process, a unique chromosome region can match with multiple genes.

To estimate the proportion of ancestral genes lost, we first identified ZW homologous genes under a strict threshold (e-value >1e^{-10}, identity > 80 and alignment/query length >70%). Z-specific and W-specific genes without homologs on the other haplotype were determined by excluding homologous genes under a soft threshold (e-value >1e^{-5}, identity > 60, alignment/query length >50%). Ancestral genes were defined a Z-W homologs with orthologs in at least one of the following four species: *S. purpurea*, *S. suchowensis*, *S. arbutifolia* and *P. trichocarpa*, Additional file 3: Table S4) as detected by OrthoFinder v2.5.2 [110]. The proportion of ancestral genes lost on each haplotype was calculated as (1 - the number of ancestral genes on a haplotype divided by the joint number of ancestral genes across both haplotypes).

Mutation accumulation

We concatenated SNPs and INDELs for the ten individuals of each sex, using conservatively filtered SNPs with the following parameters in VCFtools [104]: minDP=10, maxDP=100, MAF=0.1, max-missing=0.8, minQ=30. We categorized the effect of these variants within coding regions into three groups using SnpEff v5.2 [111]: 1) disruptive; the variant is expected to have a substantial impact on protein function, alternatively, nonsense-mediated decay, such as stop gained or frameshift variant; 2) moderate; the variant may change protein effectiveness, e.g., missense variant or inframe deletion; 3) no impact; the variant is not expected to change protein functionality or does not affect the protein as for synonymous variants or downstream gene variants.

Estimation of LTR insertion times

LTR_harvest and LTR_finder were used to predict full-length LTR retrotransposons in the genome assembly [112, 113]. The insertion time of LTRs on the W-chromosome was estimated based on the substitutions between 5’ and 3’ flanking LTR repeats, assuming a mutation rate of 2.5×10^{-9} per generation.

Divergence between Z and W haplotypes

We estimated the substitution rates (K_s) of homologous genes between the Z and W haplotypes based on the method of YN [114] using KaKs_caculator 2.0 [115]. Protein-coding sequences of ZW homologous genes were extracted using amino acid alignment with ParaAT [116]. To ensure accurate alignments, ambiguous sites and gaps were removed, and only aligned nucleotide regions longer than 300 bp were retained.

To reconstruct phylogenetic relationships among the Z-W homologous genes, we identified single-copy orthologs in *S. herbacea*, *S. purpurea* and *P. trichocarpa* using OrthoFinder v2.5.2 [110]. A maximum likelihood tree for each gene was built using IQ-TREE [117] with the best model and partition scheme

selected by ModelFinder [118]. A topological distance matrix between phylogenetic trees was generated in R with the APE package [119] in R version v4.4.0 [120] using *P. trichocarpa* as an outgroup. We performed hierarchical cluster analysis with the UPGMA method and generated consensus trees for gene trees bootstrap support for nodes > 80%.

To infer the phylogenetic position of *S. herbacea*, we downloaded eight published *Salix* genomes (*S. herbacea*, *S. purpurea*, *S. udensis* (ZZ), *S. suchowensis*, *S. koriyanagi*, *S. brachista*, *S. viminalis*, *S. dunnii*, *S. arbutifolia*) and two *Populus* genomes (*P. trichocarpa*, *P. qiongdaoensis*, Additional file 3: Table S4). Gene trees were constructed using the pipeline described above. Genes from chromosome 7, 15 and 19 were excluded. A species tree was inferred from these gene trees using ASTRAL [121]. We estimated divergence time among different species using MCMCtree in the PAML package [122] under the parameters “clock = 2, model = 4, BDparas = 1 1 0.1, kappa_gamma = 6 2, alpha_gamma = 1 1, rgene_gamma = 10 25 1, sigma2_gamma = 1 10” . We used fossil calibration information as 48-52 million years for the separation of *Salix* and *Populus* clades [123], and at least 23 million years ago for the formation of the subgenus *Vetrix* in the late Oligocene [124].

Structural variation in chr15 across *Salix* species

To gain insight into the evolution of structural variation, we performed pairwise genome alignments between the two haplotypes of *Salix herbacea* and two published *Salix* genomes (*S. purpurea* and *S. arbutifolia*, Additional file 3: Table S4). We used Minimap2 v2.24 [76] allowing for sequence divergence of 0.5% (asm5). For structural variation calling, the alignment results were processed to SyRI v1.6.4 [125] with parameters set to “--tdmaxolp 0.5 --hdrseq --cigar --maxsize 100000”.

We focused on inversion blocks, filtering out the ambiguous alignments. We excluded blocks with inverted regions covering less than 10% of their length while retaining syntenic regions or structural rearrangements within these blocks. Syntenic blocks between multiple genome assemblies were identified and visualized using Plotsr [126].

Evolutionary history of sex-linked genes

W-hemizygous regions and fixed W-specific variants

We annotated the hemizygous regions identified above using our gene annotation of the reference genome. The regions without hits were scanned for pseudogenization using the pseudogene pipeline described above. Using both haplotypes, we first searched for fixed W-specific variants that were consistently heterozygous in all ten females but homozygous in all ten males. To reduce the effects of duplication and multiple alignments in the sex-linked region (SLR), individuals were scored conservatively as heterozygous when the proportion of reads of the minor allele was greater than 0.3 (the ratio of minor allele and major allele > 0.43). We further restricted the identification of W-specific variants (ZW in females and ZZ in males) to biallelic loci, where males were homozygous for one of the

two alleles found in females. We annotated these loci using the repeat and gene annotations described above. We further evaluated the impact of the fixed W-specific variants on gene function using SnpEff v5.2 [111].

Z-specific genes

To detect potential male function genes present on the Z chromosome but degenerated on the W chromosome, we searched for gene residuals retained on the W haplotype using Z-specific genes as the query after masking out coding and repetitive sequences. The partial copies were determined as sequences with sequence identities with the conserved exon > 80%.

Gene synteny in other *Salix and Populus* species

For candidate genes identified in *S. herbacea* here, we searched for the homologous genes in 10 *Salix* genomes using BLASTp [97]. The gene multiple synteny plots were generated using MCScan (python version) under JCVI [127, 128].

To identify whether the W-hemizygous regions and fixed W-specific variants are also present in other *Salix* species, we mapped the Illumina reads of both sexes of *S. pupurea*, *S. suchowensis*, *S. viminalis* (Additional file 3: Table S4) to the *S. herbacea* reference genome using BWA v0.7.17 [77]. The mapping output and genotype calling were filtered following the same process as described above. The candidate regions were extracted and verified using custom scripts.

Candidate sex determination genes

To investigate the previously identified candidate sex determination genes in *Populus* and *Salix*, *ARR17 [26],* we used BLASTp [97] searches for intact homologous genes in *S. herbacea* and the other 9 *Salix* and 2 *Populus* genome assemblies (data from Additional file 3: Table S4) using *ARR17* (Potri.019G133600) as the query. To detect partial duplicates, *ARR17* segmental sequences and five complete exon nucleotide sequences from Hu et al. [63] were searched in all these genome assemblies using BLASTN [97]with “-evalue 1e-5 -word_size 8”. Genes with sequence identity over 80% in all exons were considered complete genes and sequences matching only on the first exon were considered partial duplicates. The same workflow was applied to the candidate gene (*PI*) proposed in this study. The first exon of all complete and partial duplicate copies was aligned by MUSCLE v 5.1[129]. Phylogenies were constructed using maximum likelihood analyses in IQ-TREE [117] with the best model and partition scheme selected by ModelFinder [118].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Data supporting the findings of this work are available within the paper and associated Supplementary Information files. The whole genome assembly project for *Salix herbacea* is deposited in GenBank under the accessions PRJNA1199913. All raw reads of *Salix herbacea* generated in this study are deposited in the NCBI database under BioProject accession PRJNA1196844 (Dynamic evolution of a sex-linked region).

The code used in this study is available in the GitHub repositories at https://github.com/miaoxm2/Salix-SexChr.Genome.git.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by a project grant no. 2019-03844 from the Swedish Research Council (VR) to SK. The computations and data handling were enabled by resources in project NAISS 2024/6-88, NAISS 2023/6-63 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX, funded by the Swedish Research Council through grant agreement no. 2022-06725.

Authors' contributions

S.K. and X.M. conceived and designed the study. X.M. performed genome assembly, annotation, and analysis. C.T. contributed to the initial genome assembly. S.K., P.I. and N.R. provided valuable input to the analysis. S.K. and X.M. wrote the manuscript. All authors read, revised, and approved the final manuscript.

Acknowledgements

We thank Christian Rixen and Francesca Jaroszynska for performing field work and Yvonne Meyer-Lucht for help with the extraction of high molecular weight DNA. We would like to acknowledge support of the National Genomics Infrastructure (NGI) /Uppsala Genome Center and UPPMAX for providing assistance in massive parallel sequencing and computational infrastructure.

References

1. Beukeboom L, Perrin N. The Evolution of Sex Determination. OUP Oxford; 2014.
2. Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet. 2013;14:113–24.
3. Jeffries DL, Lavanchy G, Sermier R, Sredl MJ, Miura I, Borzée A, et al. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. Nat Commun. 2018;9:4088.
4. 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.
5. Wang Y, Gong G-N, Wang Y, Zhang R-G, Hörandl E, Zhang Z-X, et al. Gap-free X and Y chromosome assemblies of *Salix arbutifolia* reveal an evolutionary change from male to female heterogamety in willows, without a change in the position of the sex-determining locus. New Phytol. 2024.
6. 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;23:200.
7. 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.
8. Yazdi HP, Ellegren H. Old but not (so) degenerated--slow evolution of largely homomorphic sex chromosomes in ratites. Mol Biol Evol. 2014;31:1444–53.
9. Kuhl H, Guiguen Y, Höhne C, Kreuz E, Du K, Klopp C, et al. A 180 Myr-old female-specific genome region in sturgeon reveals the oldest known vertebrate sex determining system with undifferentiated sex chromosomes. Philos Trans R Soc Lond B Biol Sci. 2021;376:20200089.
10. Saunders PA, Muyle A. Sex chromosome evolution: hallmarks and question marks. Mol Biol Evol. 2024;:msae218.
11. Jay P, Jeffries D, Hartmann FE, Véber A, Giraud T. Why do sex chromosomes progressively lose recombination? Trends Genet. 2024.
12. Rice WR. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution. 1987;41:911–4.
13. Lenormand T, Roze D. Y recombination arrest and degeneration in the absence of sexual dimorphism. Science. 2022;375:663–6.
14. Olito C, Ponnikas S, Hansson B, Abbott JK. Consequences of partially recessive deleterious genetic variation for the evolution of inversions suppressing recombination between sex chromosomes1. Evolution. 2024;78:1499–510.
15. Jeffries DL, Gerchen JF, Scharmann M, Pannell JR. A neutral model for the loss of recombination on sex chromosomes. Philos Trans R Soc Lond B Biol Sci. 2021;376:20200096.
16. Sigeman H, Downing PA, Zhang H, Hansson B. The rate of W chromosome degeneration across multiple avian neo-sex chromosomes. Sci Rep. 2024;14:16548.

17. Sacchi B, Humphries Z, Kružlicová J, Bodláková M, Pyne C, Choudhury BI, et al. Phased assembly of Neo-sex chromosomes reveals extensive Y degeneration and rapid genome evolution in *Rumex hastatulus*. Mol Biol Evol. 2024;41:msae074.
18. Xu L, Ren Y, Wu J, Cui T, Dong R, Huang C, et al. Evolution and expression patterns of the neo-sex chromosomes of the crested ibis. Nat Commun. 2024;15:1670.
19. Lemaitre C, Braga MDV, Gautier C, Sagot M-F, Tannier E, Marais GAB. Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes. Genome Biol Evol. 2009;1:56–66.
20. Peichel CL, McCann SR, Ross JA, Naftaly AFS, Urton JR, Cech JN, et al. Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol. 2020;21:177.
21. Zhu Z, Younas L, Zhou Q. Evolution and regulation of animal sex chromosomes. Nat Rev Genet. 2024;:1–16.
22. Feng G, Sanderson BJ, Keefover-Ring K, Liu J, Ma T, Yin T, et al. Pathways to sex determination in plants: how many roads lead to Rome? Curr Opin Plant Biol. 2020;54:61–8.
23. Harkess A, Huang K, van der Hulst R, Tissen B, Caplan JL, Koppula A, et al. Sex determination by two Y-linked genes in garden asparagus. Plant Cell. 2020;32:1790–6.
24. Akagi T, Pilkington SM, Varkonyi-Gasic E, Henry IM, Sugano SS, Sonoda M, et al. Two Y-chromosome-encoded genes determine sex in kiwifruit. Nat Plants. 2019;5:801–9.
25. Akagi T, Henry IM, Tao R, Comai L. Plant genetics. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. Science. 2014;346:646–50.
26. 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.
27. 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 Lond B Biol Sci. 2022;377:20210217.
28. Leite Montalvão AP, Kersten B, Fladung M, Müller NA. The diversity and dynamics of sex determination in dioecious plants. Front Plant Sci. 2020;11:580488.
29. 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:e3546.
30. 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;21:38.
31. 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;18:78.

32. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. EggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. bioRxiv. 2021.
33. Cheng C, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. The Plant Journal. 2017;89:789–804.
34. Yang Y, Ma C, Xu Y, Wei Q, Imtiaz M, Lan H, et al. A zinc finger protein regulates flowering time and abiotic stress tolerance in chrysanthemum by modulating gibberellin biosynthesis. Plant Cell. 2014;26:2038–54.
35. Swinka C, Hellmann E, Zwack P, Banda R, Rashotte AM, Heyl A. Cytokinin response factor 9 represses cytokinin responses in flower development. Int J Mol Sci. 2023;24:4380.
36. Kandasamy MK, Deal RB, McKinney EC, Meagher RB. Silencing the nuclear actin-related protein AtARP4 in *Arabidopsis* has multiple effects on plant development, including early flowering and delayed floral senescence: Role of AtARP4 in plant development. Plant J. 2005;41:845–58.
37. Zhou D, Chen C, Jin Z, Chen J, Lin S, Lyu T, et al. Transcript profiling analysis and ncRNAs' identification of male-sterile systems of Brassica campestris reveal new insights into the mechanism underlying anther and pollen development. Front Plant Sci. 2022;13:806865.
38. Ramming A, Kappel C, Kanaoka MM, Higashiyama T, Lenhard M. Poly(A) polymerase 1 contributes to competence acquisition of pollen tubes growing through the style in Arabidopsis thaliana. Plant J. 2023;114:651–67.
39. Mashiguchi K, Asami T, Suzuki Y. Genome-wide identification, structure and expression studies, and mutant collection of 22 early nodulin-like protein genes in *Arabidopsis*. Biosci Biotechnol Biochem. 2009;73:2452–9.
40. Xia C, Wang Y-J, Liang Y, Niu Q-K, Tan X-Y, Chu L-C, et al. The ARID-HMG DNA-binding protein AtHMGB15 is required for pollen tube growth in *Arabidopsis thaliana*. Plant J. 2014;79:741–56.
41. Pagnussat GC, Yu H-J, Ngo QA, Rajani S, Mayalagu S, Johnson CS, et al. Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. Development. 2005;132:603–14.
42. Harscoët E, Dubreucq B, Palauqui J-C, Lepiniec L. NOF1 encodes an *Arabidopsis* protein involved in the control of rRNA expression. PLoS One. 2010;5:e12829.
43. Falbel TG, Koch LM, Nadeau JA, Segui-Simarro JM, Sack FD, Bednarek SY. SCD1 is required for cytokinesis and polarized cell expansion in *Arabidopsis thaliana*. Development. 2003;130:4011–24.
44. Veyres N, Danon A, Aono M, Galliot S, Karibasappa YB, Diet A, et al. The *Arabidopsis* sweetie mutant is affected in carbohydrate metabolism and defective in the control of growth, development and senescence. Plant J. 2008;55:665–86.
45. Graeff M, Straub D, Eguen T, Dolde U, Rodrigues V, Brandt R, et al. MicroProtein-mediated recruitment of CONSTANS into a TOPLESS trimeric complex represses flowering in Arabidopsis. PLoS Genet. 2016;12:e1005959.

46. Atanasov V, Schumacher J, Muiño JM, Larasati C, Wang L, Kaufmann K, et al. *Arabidopsis* BBX14 is involved in high light acclimation and seedling development. Plant J. 2024;118:141–58.
47. D'Agostino IB, Deruère J, Kieber JJ. Characterization of the response of the *Arabidopsis* response regulator gene family to cytokinin. Plant Physiol. 2000;124:1706–17.
48. Benlloch R, Roque E, Ferrándiz C, Cosson V, Caballero T, Penmetsa RV, et al. Analysis of B function in legumes: PISTILLATA proteins do not require the PI motif for floral organ development in *Medicago truncatula*. Plant J. 2009;60:102–11.
49. 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. Molecular Ecology. 2023;32:1366–80.
50. Kersten B, Pakull B, Groppe K, Lueneburg J, Fladung M. The sex-linked region in *Populus tremuloides* turesson 141 corresponds to a pericentromeric region of about two million base pairs on P. trichocarpa chromosome 19. Plant Biol (Stuttg). 2014;16:411–8.
51. Zhang S, Wu Z, Ma D, Zhai J, Han X, Jiang Z, et al. Chromosome-scale assemblies of the male and female *Populus euphratica* genomes reveal the molecular basis of sex determination and sexual dimorphism. Commun Biol. 2022;5:1186.
52. Geraldes A, Hefer CA, Capron A, Kolosova N, Martinez-Nuñez F, Soolanayakanahally RY, et al. Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Molecular Ecology. 2015;24:3243–56.
53. Moraga C, Branco C, Rougemont Q, Veltsos P, Jedlička P, Muyle A, et al. The Silene latifolia genome and its giant Y chromosome. bioRxiv. 2023;:2023.09.21.558754.
54. Zhou Y, Zhan X, Jin J, Zhou L, Bergman J, Li X, et al. Eighty million years of rapid evolution of the primate Y chromosome. Nat Ecol Evol. 2023;7:1114–30.
55. Renner SS, Müller NA. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. Nat Plants. 2021;7:392–402.
56. Prentout D, Stajner N, Cerenak A, Tricou T, Brochier-Armanet C, Jakse J, et al. Plant genera *Cannabis* and *Humulus* share the same pair of well-differentiated sex chromosomes. New Phytol. 2021;231:1599–611.
57. Ogutcen E, de Lima Ferreira P, Wagner ND, Marinček P, Vir Leong J, Aubona G, et al. Phylogenetic insights into the Salicaceae: The evolution of willows and beyond. Mol Phylogenet Evol. 2024;199:108161.
58. Centenaro G, Petraglia A, Carbognani M, Piotti A, Hudek C, Büntgen U, et al. The oldest known clones of Salix herbacea growing in the Northern Apennines, Italy are at least 2000 years old. Am J Bot. 2023;:e16243.
59. Borges F, Martienssen RA. The expanding world of small RNAs in plants. Nat Rev Mol Cell Biol. 2015;16:727–41.
60. Slotkin RK, Freeling M, Lisch D. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nat Genet. 2005;37:641–4.

61. Erdmann RM, Picard CL. RNA-directed DNA methylation. PLoS Genet. 2020;16:e1009034.
62. 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;8:170.
63. 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.
64. Licausi F, Ohme-Takagi M, Perata P. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. New Phytol. 2013;199:639–49.
65. Lalanne E, Honys D, Johnson A, Borner GHH, Lilley KS, Dupree P, et al. SETH1 and SETH2, two components of the glycosylphosphatidylinositol anchor biosynthetic pathway, are required for pollen germination and tube growth in *Arabidopsis*. Plant Cell. 2004;16:229–40.
66. Mao X, Cortés AJ, Rixen C, Karrenberg S. Female-biased population sex ratios caused by genetic rather than ecological mechanisms in dwarf willow (*Salix herbacea* L.). J Ecol. 2024;112:1731–42.
67. Myers-Smith IH, Hik DS. Uniform female-biased sex ratios in alpine willows. Am J Bot. 2012;99:1243–8.
68. Scott MF, Osmond MM, Otto SP. Haploid selection, sex ratio bias, and transitions between sex-determining systems. PLoS Biol. 2018;16:e2005609.
69. Workman R, Fedak R, Kilburn D, Hao S, Liu K, Timp W. High molecular weight DNA extraction from recalcitrant plant species for third generation sequencing. Protocols.io. 2019.
70. 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.
71. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36:2896–8.
72. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. Imeta. 2023;2:e107.
73. 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.
74. 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.
75. Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. Juicebox.js provides a cloud-based visualization system for Hi-C data. Cell Syst. 2018;6:256–8.e1.
76. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37:4572–4.
77. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
78. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38:4647–54.

79. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009;Chapter 4:4.10.1–4.10.14.
80. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:11.
81. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 2006;313:1596–604.
82. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117:9451–7.
83. Ou S, Jiang N. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant Physiol. 2018;176:1410–22.
84. Kapitonov VV, Jurka J. Distribution of transposable and repetitive elements in the A. thaliana chromosomes. 2002.
85. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.
86. 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.
87. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 2021;3:lqaa108.
88. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and Augustus. Bioinformatics. 2016;32:767–9.
89. Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 2014;42:e119.
90. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. bioRxiv. 2024.
91. Bruna T, Lomsadze A, Borodovsky M. A new gene finding tool GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. bioRxiv. 2024.
92. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019;20:278.
93. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Res. 2020;9.
94. Quinlan AR. BEDTools: The Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics. 2014;47:11.12.1–34.
95. 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.
96. 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.

97. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
98. Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. BMC Bioinformatics. 2021;22:566.
99. 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.
100. Andrews S. FastQC: A quality control tool for high throughput sequence data. 2023. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 13 May 2024.
101. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27:2987–93.
102. Broad Institute. Picard toolkit. Broad Institute; 2019.
103. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2018;:201178.
104. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.
105. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011;27:3435–6.
106. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 2018;34:867–8.
107. Weir BS, Cockerham CC. Estimating f-statistics for the analysis of population structure. Evolution. 1984;38:1358–70.
108. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.
109. Xie J, Li Y, Liu X, Zhao Y, Li B, Ingvarsson PK, et al. Evolutionary origins of pseudogenes and their association with regulatory sequences in plants. Plant Cell. 2019;31:563–78.
110. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:157.
111. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. Apr-Jun 2012;6:80–92.
112. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 2008;9.
113. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35 Web Server issue:W265–8.
114. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 2000;17:32–43.

115. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics. 2010;8:77–80.
116. Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, et al. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. Biochem Biophys Res Commun. 2012;419:779–81.
117. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von HA, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution. 2020;37:1530–4.
118. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
119. Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20:289–90.
120. R Core Team. R: A Language and Environment for Statistical Computing. 2024.
121. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 2018;19 Suppl 6:153.
122. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.
123. Manchester SR, Dilcher DL, Tidwell WD. Interconnected reproductive and vegetative remains of Populus (Salicaceae) from the middle Eocene Green River Formation, northeastern Utah. Am J Bot. 1986;73:156–60.
124. Boucher LD, Manchester SR, Judd WS. An extinct genus of Salicaceae based on twigs with attached flowers, fruits, and foliage from the Eocene Green River Formation of Utah and Colorado, USA. Am J Bot. 2003;90:1389–99.
125. Goel M, Sun H, Jiao W-B, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 2019;20:277.
126. Goel M, Schneeberger K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics. 2022;38:2922–6.
127. Tang H, Krishnakumar V, Zeng X, Xu Z, Taranto A, Lomas JS, et al. JCVI: A versatile toolkit for comparative genomics analysis. Imeta. 2024;3:e211.
128. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49.
129. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.

Tables

Tables 1 to 3 are available in the Supplementary Files section

Figures

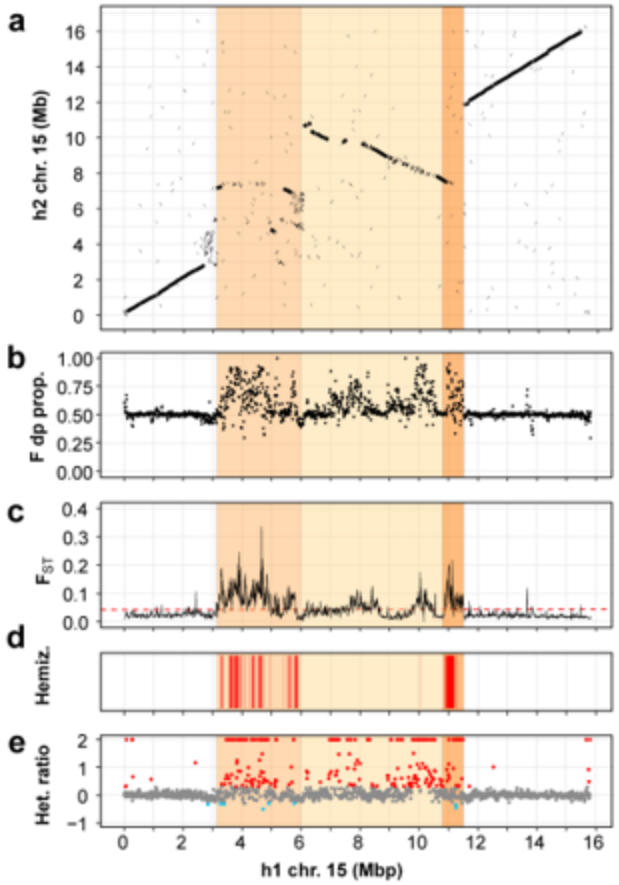

Figure 1

Sex-linked regions on chromosome 15 in the dwarf willow *Salix herbacea*. a Dot plot of an alignment of chromosome 15 between haplotype 1 and haplotype 2. **b** Read depth proportion (F dp prop.) of females in pooled sequences of 50 females (F) and 50 males (M), calculated as F/(F+M) in 10 Kb windows. **c** Genetic differentiation (F_{ST}) between pools of 50 females and 50 males in 10 Kb windows with the genome-wide 99th percentile indicated as a red dashed line. **d** Location of W-hemizygous sites that were present in all 10 female individuals but absent from all 10 male individuals. **e** Ratio of the number of heterozygous sites (Het. ratio) between females (F) and males (M) in 10 Kb windows (F/ M); a twofold excess of female or male heterozygous sites is indicated in red and blue color, respectively. The haplotype alignment, as well as read depth, F_{ST}, and heterozygous sites for the both haplotypes and all chromosomes are shown in Additional file 1: Fig. S3-9.

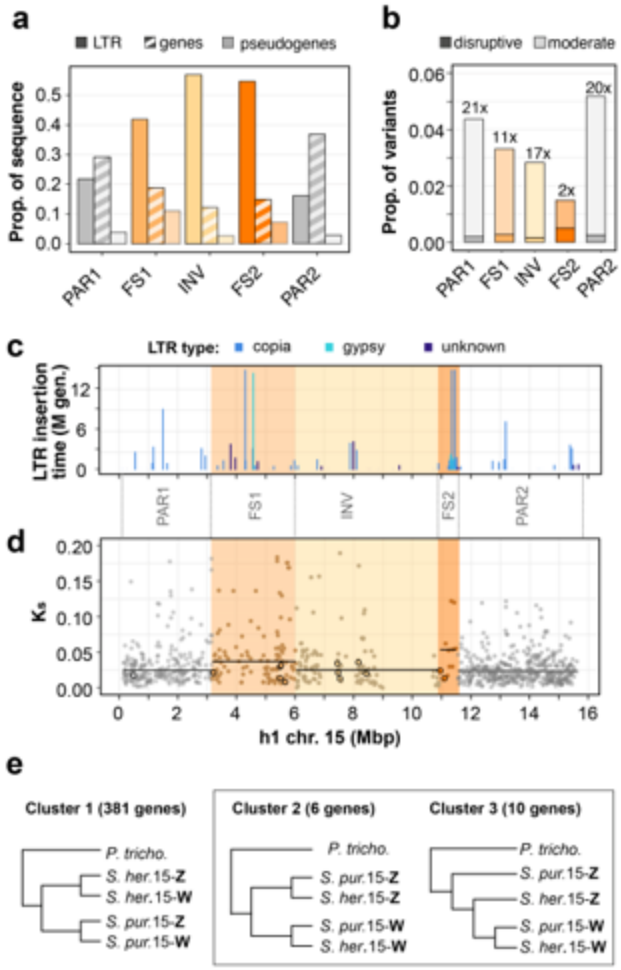

Figure 2

Degeneration and divergence along the W-chromosome of the dwarf willow *Salix herbacea*. **a** Sequence proportions of long terminal repeats (LTRs), genes and pseudogenes (x10) in five regions, two distal pseudoautosomal regions (PAR1, PAR2), two-female specific regions (FS1, FS2) and a central inversion region (INV). **b** Proportion of variants across the study population with predicted moderate and disruptive effects on gene function with the ratio between the numbers of moderate and disruptive variants. **c** Insertion time estimates for LTRs in millions of generations (M gen.), assuming a mutation rate of 2.5×10^{-9} per generation, with colors distinguishing between Copia (green), Gypsy (red), and unknown LTR (purple) types. **d** Synonymous substitution rates (K_S) of Z-W homologous genes with median K_S indicated in each region. Ks values are shown within the range of 0–0.2 for clarity. Additional 28 genes with $K_S > 0.2$ are not shown. **e** Three primary consensus trees of single-copy orthologs (SCOs) for *S. herbacea* (Saher) and *S. purpurea* (Sapur) with *P. trichocarpa* as an outgroup. Genes supporting sex chromosome divergence before species divergence (clusters 2 and 3) are marked as black circles in d.

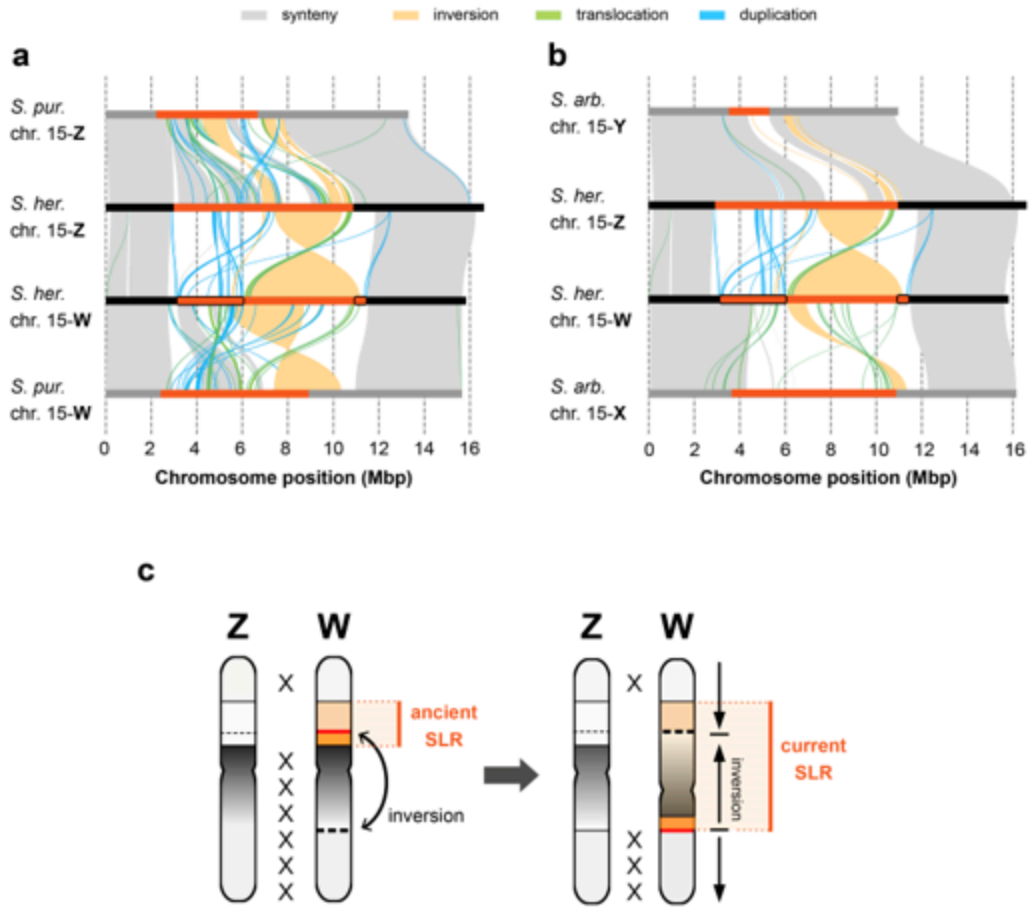

Figure 3

Structural variation across *Salix* species. Collinearity between Z and W chromosomes of *S. herbacea* (S. her.) with *S. purpurea* (S. pur, **a**) and *S. arbutifolia* (S. arb., **b**). Sex-linked regions of each haplotype are colored in orange and the two female specific regions on the W-chromosome of *S. herbacea* with boxes. Inversions, translocations and duplications are shown in light orange, green and blue, respectively. Grey lines show synteny alignments. **c.** Illustration of inferences on the structural evolution of sex chromosomes in *S. herbacea*: step-wise recombination reduction is mediated by a large inversion that incorporates pseudoautosomal sequence into the sex-linked region.

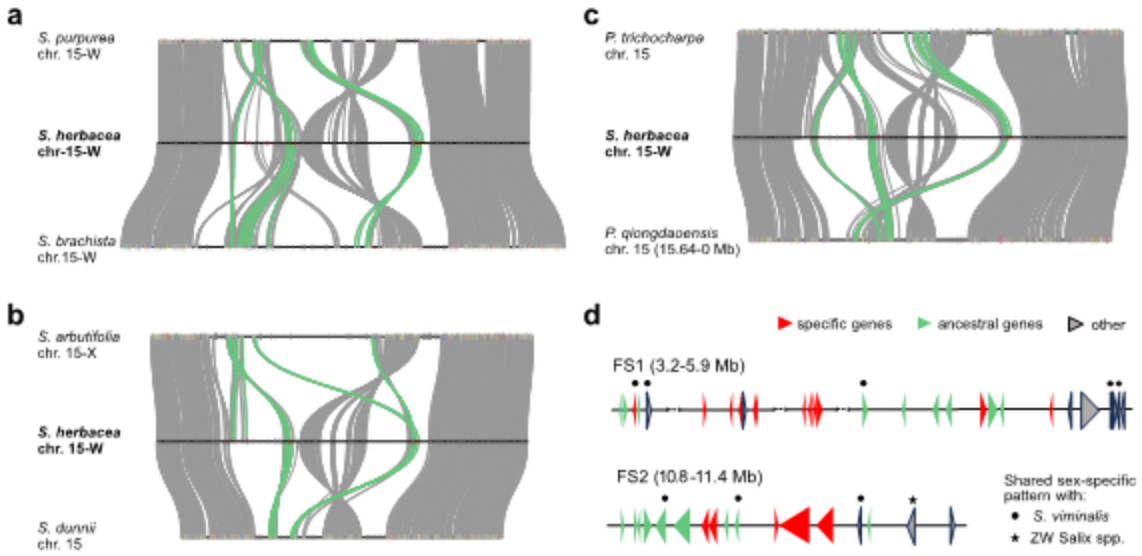

Figure 4

Gene synteny on chromosome 15 within the Salicaceae family. a. Synteny of *Salix herbacea* with other *Salix*species with 15ZW sex determination (*S. purpurea*, *S. brachista*), **b.** with *Salix* species with 15XY and 7XY sex-determining systems (*S. arbutifolia*, *S. dunnii*), **c.** with *Populus*species with 19XY and 19ZW sex-determining systems (*P. trichocarpa*, *P. qiongdaoensis*). Red ticks indicate sex-linked genes in *S. herbacea*, with homologous genes highlighted by green lines. **d.** Schematic representation of sex-linked genes in FS1 and FS2 of chromosomes 15 of *S. herbacea*; blue triangles represent genes ancestrally homologous across all three clades (a, b, c), orange triangles denote genes specific to the *S. herbacea*chromosome 15, and grey triangles indicate genes with homologs in at least one but not all clades. Genes with W-hemizygosity or sex-specific polymorphisms shared with *S. viminalis* or with multiple species with a ZW sex-determination system are marked by blue circles or a green star, respectively.

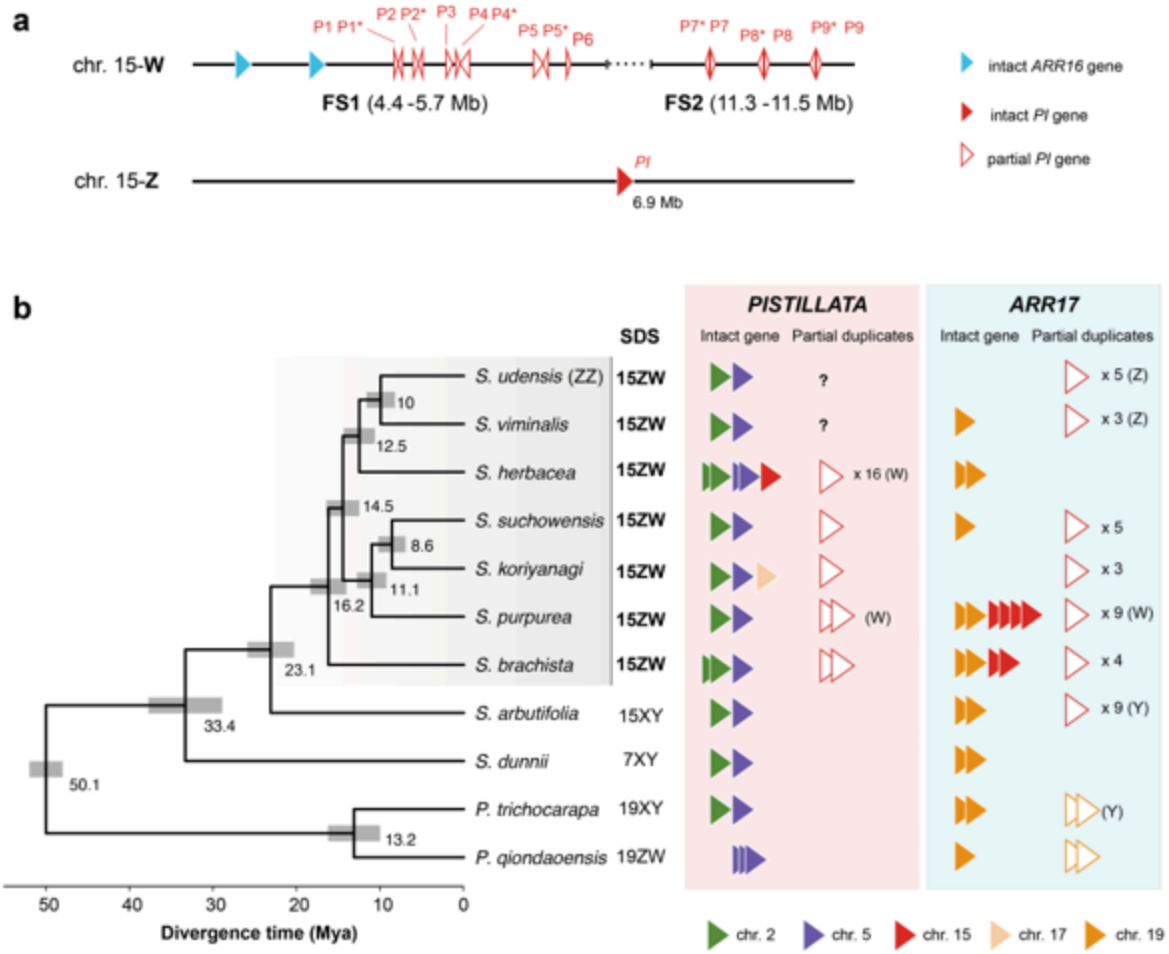

Figure 5

Candidate sex determination genes in *Salix* and *Populus* species. a.Distribution of *PISTILLATA* (*PI*) and *RR16* on both haplotypes of chromosome 15 in *Salix herbacea*. **b.**Species phylogenetic tree and copy numbers of candidate sex-determination genes (*PI* and *ARR17*) in the genomes of nine *Salix* species and two *Populus* species. The chromosomal locations are depicted by distinct colors and complete and partial genes by full and empty triangles, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- Table2.docx
- Table3.docx
- Additionaldata1.supplementaryfiguresS1S9.pdf
- Additionaldata2.supplementaryfiguresS10S11.pdf

- Additionaldata3.supplementarytables.pdf
- Additionaldata4.Dataset1.Wpolymorphisms.xlsx
- Additionaldata5.Dataset2.Whemizygous.xlsx
- Additionaldata6.Dataset3.blastofARR17.xlsx