Almost all plants in the genus *Populus* are dioecious (i.e. trees are either male or female), but it is unknown whether dioecy evolved in a common ancestor or independently in different subgenera. Here, we sequence the small peritelomeric X- and Y-linked regions of *P. deltoides* chromosome XIX. Two genes are present only in the Y-linked region. One is a duplication of a non-Y-linked, female-specifically expressed response regulator, which produces siRNAs that block this gene’s expression, repressing femaleness. The other is an LTR/Gypsy transposable element family member, which generates long non-coding RNAs. Overexpression of this gene in *A. thaliana* promotes androecium development. We also find both genes in the sex-determining region of *P. simonii*, a different poplar subgenus, which suggests that they are both stable components of poplar sex-determining systems. By contrast, only the duplicated response regulator gene is present in the sex-linked regions of *P. davidiana* and *P. tremula*. Therefore, findings in our study suggest dioecy may have evolved independently in different poplar subgenera.
Most flowering plants produce bisexual or ‘perfect’ flowers (hermaphroditism), but ~10% of angiosperm species bear unisexual flowers, either with male and female flowers on the same plant (monoeacy) or on separate individuals (dioecy). Dioecy has evolved independently hundreds of times from hermaphroditic ancestors, in multiple plant lineages, and recent advances have allowed details of the genetic mechanisms of sex determination to be understood in several dioecious plants. A theoretical model for the origin of sex chromosomes involves a transition from functional hermaphroditism (including monoecy) to dioecy via mutations in two linked genes, affecting female and male functions. Recent empirical studies in garden asparagus (Asparagus officinalis L.) and kiwifruit (Actinidia rufa × A. chinensis) have revealed fully Y-linked sex-determining genes that support this hypothesis.

A two-gene system is not inevitable. First, both mutations could occur in the same gene. Second, a single-gene sex determination, involving a gene that dominantly suppresses female function and promotes male function, can be experimentally created in monocious plants, including maize and melon (Cucumis melo), by mutations in unlinked but interacting genes, one of which acts in a sex-specific manner. Sex determination in persimmon ( Diospyros lotus, a tree) is a naturally evolved single-gene system, involving a non-coding RNA locus, OG1, suppressing femaleness. However, two mutations were necessary for its evolution, and the target gene MeGI is also inferred to have changed during the evolution of females.

Different genetic factors therefore clearly control sex determination in unrelated flowering plant lineages, consistent with independent origins of dioecy. Unlike the ‘cryptic dioecy’ in asparagus and kiwifruit, whose flowers bear apparently normal organs of the opposite sex, sex organ abortion in poplars and willows (family Salicaceae) occurs early, before the initiation of stamen or carpel primordia, and dioecy may have evolved from a monocious ancestral state, consistent with these plants bearing ‘catkins’ of small unisexual florets. In the genus Populus, chromosome XIX carries the sex determining locus. Genetic mapping studies revealed that the locations of the sex determining gene differ in genera Populus and its sister genus Salix, suggesting independently evolved sex determination within the family Salicaceae. Multiple candidate sex determining genes have emerged from genomic association studies and/or transcriptome expression analyses, including a recent study in a poplar species, P. tremula, that revealed a sex determining role for one of these candidates, which was named ARR17 (Arabidopsis Response Regulator 17).

In the present study, in another poplar species P. deltoides, we characterize the regulation and functions of candidate sex-determining genes. Our results suggest how dioecy may have evolved in P. deltoides, and support the view that dioecy evolved independently in P. deltoides and P. tremula.

**Results**

**Mapping the sex-determining locus and reconstructing X and Y haplotypes.** Linkage analysis using simple sequence repeat (SSR) markers (Supplementary Table 1) located the P. deltoides sex-determining locus near the telomeric end of chromosome XIX (Supplementary Fig. 1), and confirmed male heterogamy (XY sex determining system). To identify the sex-determining gene(s), we sequenced and de novo assembled the genomes of a poplar female and one of its male offspring. The assembly for the female is 431 Mb, with contig N50 of 1.4 Mb, and 414 Mb with contig N50 of 2.8 Mb for the male. BUSCO analyses showed that 93.5% and 96.1% of plant conserved single-copy genes are complete in the female and male assemblies, respectively (Supplementary Table 2). Our SSR markers located the sex-determining locus to a 299-kb sex-linked region (SLR) between the telomeric region and marker N362 (Fig. 1a); we refer to the recombining region proximal to N362 as pseudo-autosomal (the PAR).

Because the sequenced male inherited his X chromosome from his sequenced female parent, we could use single-nucleotide polymorphisms (SNPs) to infer the complete SLR-X and SLR-Y haplotypes in the sex-linked region (Fig. 1b). The X and Y haplotypes’ telomeric repeat regions were assembled into two contigs, of sizes 104 kb and 141 kb, both well supported by mapped raw PacBio reads; the annotated repetitive sequences total 6.2 kb for the X telomeric repeat region, versus much longer (24.5 kb) for the Y haplotype (Fig. 1b). Annotation predicted 41 genes in the SLR-X haplotype and 26 in SLR-Y. At least 16 genes with assigned functions were found in both the SLR-X and -Y, including a cluster of 5 tandem genes encoding leucine-rich repeats (LRR) receptor-like protein kinases (Supplementary Table 3). Supplementary Figure 3 shows the alignment of homologous genes between SLR-X and -Y. Divergence between genic sequences on the SLR-X and -Y is highest for genes neighboring the telomeric repeat regions, and some are unalignable (Fig. 1b and Supplementary Fig. 2). Taken together, the differentiation of X and Y haplotypes across the 299 kb region is strong evidence for a completely sex-linked region at the telomeric end of chromosome XIX.

**Identifying sex determining genes.** To identify the P. deltoides sex determining factors, we performed a genome-wide association study (GWAS) based on SNPs, using 49 female and 46 male trees (Supplementary Table 4). Genome resequencing generated a total of 1.15 Tb of Illumina reads with sequence depths of at least 20× for each of the 95 trees (Supplementary Table 4). With the sequenced female as the reference, we detected 435 SNPs with genotypes matching the individuals’ sexes, assuming male heterogamy (SNPs that are homozygous in all females in our samples, but heterozygous in all the males; Fig. 2a and Supplementary Table 5). We refer to such SNPs as SEMSs (SNPs exactly matching with sexes). 315 SEMSs are within three SLR genes, T-complex protein 1 subunit gamma (TCP), Chloride channel protein CLC-c (CLC), and DNA-methyltransferase 1 (MET1). In total, 120 are elsewhere, of which 78 are in a chromosome XIX PAR gene (Fig. 1a) that belongs to a gene family of response regulators (RR) in the cytokinin signaling pathway, and specifically to the A. thaliana type-A RR genes. We named this gene FERR (female-specifically expressed RESPONSE REGULATOR), based on evidence for its function described below. Phenotypes previously reported for mutations in two members of this A. thaliana gene family, arr16 and arr17, include effects on plant morphogenesis, cell division activity and root hydrotripsm, but not changes in floral organs. However, a recent study demonstrated involvement of such a gene in female functions and sex determination in P. tremula and named the FERR ortholog ARR17 (although, as discussed below, it is not the ortholog of the A. thaliana ARR17 gene).

The other 42 SEM variants are in genes on three other chromosomes and one unplaced contig, Contig01665 (Fig. 2a); 27 are in an autosomal gene, HEMA1, on chromosome IX, and 15 in non-coding sequences or genes with unknown functions (Supplementary Table 5). Previous studies in P. balsamifera and P. trichocarpa, using the assembled genome sequence of a P. trichocarpa female as the reference, also found SNPs associated with sex on multiple chromosomes.

However, using a female genome as the reference will produce false positive SEMSs, because reads from Y-linked regions that are missing from the female genome will erroneously map to homologous sequences elsewhere in the genome. Examination of our P. deltoides non-SLR SEMSs indeed revealed sequence...
similarity with the SLR-Y (Supplementary Table 6). Use of the SLR-Y as our reference for GWAS analysis eliminated all the non-SLR SEMSs (Fig. 2c), leaving only the three SLR genes, TCP, CLC, and MET1 genes as sex-determining gene candidates.

Y-hemizygous genes. As mentioned above, some segments of the X and Y haplotypes are unalignable (Supplementary Fig. 2); these indicate sex-specific regions that could include further candidate sex-determining genes. To examine this possibility, we performed coverage-based GWAS in our 95 P. deltoides samples (see Methods section). With the SLR-Y as the reference, we detected three Y-specific hemizygous sequences (YHSs), two large ones, YHS1 in Fig. 1 (34.8 kb), the 4.3 kb YHS2, and one of only 100 bp (YHS3, in an intergenic region, which is not considered further). These sequences are present in all male trees and absent in all...
females (Fig. 2d). No female-specific sequences were detected using the SLR-X as reference (Fig. 2b). The YHS1 sequence was validated by amplifying and Sanger sequencing 13 overlapping fragments in the sequenced male (Fig. 1c, d). This yielded a 41,481-bp sequence identical to the SLR-Y sequence, which is therefore complete, and includes no gaps. We tested 20 trees of each sex from the GWAS samples for the presence of YHS1 using PCR primers designed to amplify eight separate fragments (Fig. 1c); PCR amplification succeeded in all the males but no females yielded amplification products (Fig. 1d), confirming male specificity.

Three gene models were predicted in YHS1, and one in YHS2 (Fig. 1b and Supplementary Table 3). Sequence analysis indicated that three gene models are transposable elements, but one gene (EVM0039006 in YHS1) is the duplication of FERR described above. We named the FERR duplicate FERR-R, standing for its inferred FERR repressor function, as deduced from experiments described below. A duplication of FERR was also recently found in the P. tremula sex-determining region (which is located in a different chromosome XIX region from that of P. deltoides), and was named ARR17 inverted repeat29. EVM0039005 (also within YHS1) was found to produce long non-coding transcripts (see details below), and was named MSL, for male-specific IncRNA. Before describing our evidence about these functions, we first describe data about expression during flower development that eliminates the protein-coding genes in the fully sex-linked region as candidate sex determining genes.

Expression patterns of fully sex-linked genes. Poplars have small florets, without petals or sepals, and many florets are attached to the rachis of morphologically different male or female catkins (Supplementary Fig. 4). A single male floret consists of a group of stamens inserted on a disk, while a female floret has a single-celled ovary seated in a cup-shaped disk (Supplementary Fig. 4). Poplars bloom in early spring before the flush of leaves, but female and male flower primordia start differentiating in June of the previous year (Fig. 3a). Four stages of sex organ development related to occurrence of doiecy are recognized35, T1 (before the initiation of stamen or carpel primordia), T2 (early stamen or carpel development), pre-meiosis (stage T3), and post-meiosis (stage T4); the Methods section defines the stages. Longitudinal sections of flower buds (Fig. 3b and Supplementary Fig. 5) showed that P. deltoides male and female flower primordia are distinguishable starting from early June (T1) and early July (T3), respectively. The sex-determining genes must therefore act at these early stages.

Our RNA-seq experiments detected no expression of two of the Y-specific transposable elements, eliminating them from consideration as candidate sex determining genes. In contrast, TCP,
Fig. 3 Development of flower buds and gene expression profiles. a Morphology of the male and female flower buds at developmental stages T1–T4 defined in the text. b Longitudinal sections of male and female inflorescences at T1 and T3. The red arrows point to the floret primordia, and the yellow arrow denotes the anther primordium. Three independent samples were observed with similar results. c Gene expression profiles in the 8 flower development stages described in the text. Values shown are the mean ± SD from three biological replicates (*P < 0.05; **P < 0.01, two-sided Student’s t test). The data for c are provided as a Source Data file.

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-19559-2
ARTICLE

NATURE COMMUNICATIONS | (2020) 11:5893 | https://doi.org/10.1038/s41467-020-19559-2 | www.nature.com/naturecommunications
**Fig. 4 Origin of FERR-R and its function in repressing FERR.** a) Sequence homology analysis for FERR-R. e1-e5 represent the five exons of FERR (in black) or HEMA1 (in red). S1-S8 indicate the duplicated segments described in the text. b) Abundance of FERR-R transcripts and the FERR-R generated siRNAs, and the differential methylation of FERR in the two sexes. The gray shadow shows the region methylated only in males, and the red vertical bars indicate the methylation levels in this region. c) Transient expression experiment in poplar protoplasts. siFERR is siRNA generated by FERR-R. FERR' is the siRNA-resistant version of FERR. Three independent experiments were performed with similar observation.

CLC, MET1, FERR-R, and MSL were expressed in all four early flower development stages. qRT-PCR bioassays showed that none of these genes has expression limited to flower tissue, and none shows a consistent sex difference in expression (Fig. 3c). The three protein-coding genes, TCP, CLC, and MET1, are present in both the SLR-X and -Y, making them unlikely candidates for the sex determining genes, given their lack of consistent sex differences in expression. However, the two SLR-Y hemizygous genes, FERR-R and MSL, are present only in male *P. deltoides*. Despite being expressed in all four early flower development stages, and not exclusively in flower tissue, these two non-protein-coding genes are considered as candidate sex determining genes in *P. deltoides*. We next describe expression and function data that support this view.

**FERR-R is a female sex suppressor that generates siRNAs suppressing FERR function.** The SLR-Y hemizygous gene FERR-R shows homology with FERR (a PAR gene) and with the autosomal HEMA1 gene (Fig. 4a shows the alignments). Seven segments (S1, S2, S3, S4, S6, S7, and S8) show homology with FERR; these include all regions of FERR (the promoter region, 5'UTR, exon 1, exon 2, exon 3, the first three introns, and two downstream segments) in a rearranged order compared with the progenitor copy, and with parts, including some of the promoter, and exon 1, appearing more than once in the chimeric FERR-R sequence (Fig. 4a). The FERR-R segment S5, between FERR-derived segments S4 and S5, shows homology to the HEMA1 5'-UTR and exon 1 (Fig. 4a).

Expression data from strand-specific lncRNA-Seq and small RNA-Seq revealed that the male-specific FERR-R copy is transcribed into long transcripts that generate small interfering RNAs (siRNAs; Fig. 4b). In other organisms, siRNAs have been found to guide the methylation of homologous DNA through RNA-directed DNA methylation. We found that, in *P. deltoides*, siRNAs generated by FERR-R could guide methylation at the promoter, 5'UTR, exon 1, and the first intron of the FERR gene (Fig. 4b). Bisulfite sequencing showed that methylation of the corresponding regions in FERR occurred specifically in males (Fig. 4b). In *P. balsamifera*, male-specific DNA methylation was also detected in the promoter and first intron of the homologous gene, *PbRR9*. Besides inducing siRNA-directed DNA methylation, siRNAs produced by FERR-R were also found to target FERR exons 1 to 3, suggesting that FERR-R might also trigger siRNA-guided cleavage of FERR transcripts. Transient expression experiment in poplar mesophyll protoplasts confirmed that cleavage indeed occurred, and involved interaction between FERR-R and FERR. Green fluorescence signals were observed in poplar protoplasts transformed with FERR, FERR' (a siRNA-resistant version of FERR), and after co-transformation with...
siFERR+FERR, but not after co-transformation with siFERR+FERR (Fig. 4c). Supplementary Figure 6 shows a model of the interaction between FERR-R and FERR, based on the observations just described.

qRT-PCR revealed that FERR is expressed only during the initiation of female flower primordia and the early development of female flowers in P. deltoides. In the early stages of flower development, the scales form a large proportion of bud tissue, and the flowers are too small to separate into different tissues before stage T5. FERR is expressed in whole flower buds (including scales) of stages T1 to early stage T5. In stage T5 female flower buds, removal of the scales increased FERR expression more than 50-fold, indicating that high FERR levels in earlier stages were obscured by the presence of scale tissue. After stage T5, FERR expression decreased to a low level by stage T8 (Fig. 3c). The effects of the FERR-R duplication described above act by affecting expression of FERR, and must therefore be specific to early carpel development, even though, as described above, expression of FERR-R is not specific to early bud stages.

FERR-like genes resembling A. thaliana ARR17 were among candidate sex determining genes in several previous studies of poplar and willow species (PbARR9 in P. balsamifera26,37, PtARR9 or PtRR11 in P. trichocarpa38, ARR17 in P. tremula29, RR in Salix purpurea22). Phylogenetic analysis of type-A RR genes (Supplementary Fig. 7) shows, however, that neither the P. deltoides FERR (EVM0009215.1) nor P. trichocarpa FERR (Potri019G133600.3) is orthologous to the A. thaliana ARR17 gene (the closest sequence is another gene in this family, EVM0036439.1). To investigate FERR’s function, we therefore made transgenic A. thaliana overexpressing P. deltoides FERR. These had normal androecium development, but often showed stigma exaggeration, in extreme cases producing flowers with two pistils or carpel-like sepals (Fig. 5a). We, therefore, propose that FERR is a female-specifically expressed P. deltoides response regulator (RR) that promotes female functions during the initiation of female flower primordia and early carpel development. Such a function is consistent with our evidence above that the Y-linked FERR-R gene suppresses FERR functions in P. deltoides males, and corresponds to the hypothesized female suppressor, or SiF, involved in the evolution of dioecy7. In our model, the pre-duplication FERR gene of the ancestor promoted female functions, and suppressing these created males in the dioecious descendant species. Type-A RR genes have a very complex regulatory network, not currently well characterized, and, although transcriptomic data from our transgenic A. thaliana overexpressing FERR does detect genes with altered expression (Supplementary Data 1 and Supplementary Table 7), they do not illuminate FERR’s function.

MSL is a candidate gene that generates long non-coding RNAs promoting maleness. Raw reads matching the MSL sequence in our male genome sequence were extracted and assembled to confirm the male-specificity in our GWAS samples (Supplementary Fig. 8a, b), and PCR amplification succeeded in all the males but no females yielded amplification products, confirming male-specificity (Supplementary Fig. 8c). Based on a de novo repeat library constructed from P. deltoides genome sequences (see Methods section), MSL is annotated as a transposable element in the LTR/Gypsy transposon family. Consistent with being a transposable element, we found large numbers of partial sequences with homology to the 5′ end of the MSL sequence (Supplementary Fig. 9), on all P. deltoides chromosomes (Supplementary Data 2, MSL homology regions in Populus and Salix species). However, the only complete copy was the one in YHS1. MSL sequences are completely absent from the A. thaliana and Oryza sativa genomes. Although most other poplar and willow species have only partial sequences (Supplementary Data 2), a complete MSL is present in the genome of a male P. simonii (in a different subgenus from P. deltoides, see Supplementary Data 2). This sequence is also in a YHS at the peritelomeric end of chromosome XIX, as in P. deltoides, and is absent from females of this species, and could be involved in sex determination in these two species.

A function for a transposable element is not implausible, as such elements generate IncRNAs in many species.39–41 qRT-PCR revealed continuous expression of MSL in male P. deltoides (Fig. 3c). Transcripts are detected in IncRNA-seq, but not general RNA-seq (Supplementary Fig. 10a), indicating that MSL indeed produces IncRNAs transcripts lacking a poly-A tail. To test for a phenotypic effect of MSL, we over-expressed it in A. thaliana. Overexpression of PdeMSL in A. thaliana (to levels 5- to 20-fold higher than levels in P. deltoides) did not affect the pistils, and seed set was unaffected, but did affect the androecium, commonly resulting in flowers with six long stamens, or seven or occasionally 8 stamens, stamens bearing two anthers, or branched stamens (Fig. 5b), versus the four long and two short stamens of wild-type A. thaliana flowers. These androecium-specific effects suggest that the MSL gene may promote maleness in P. deltoides. Supporting this, GO analysis revealed that genes with significantly increased expression in our transgenic lines (Supplementary Data 3, gene expression data from A. thaliana overexpressing MSL, Supplementary Table 7), show significant enrichment for pollen development functions (Supplementary Fig. 11).

Sex determination also involves FERR duplication in P. davidiana. P. deltoides belongs to subgenus Aigeiros in the genus Populus. To test whether the hemizygous YHS1 region is present in other poplars, we sequenced the genome of a male P. davidiana, in the same subgenus, Leuce, as P. tremula and P. tremuloides (in an earlier-branching section of Populus than Aigeiros17,29); sex-determining regions of all three subgenus Leuce species map to the pericentromeric region of chromosome XIX17,29. Coverage analysis of genome sequences of 49 P. davidiana females and 47 males, using the male sequence as the reference, detected a 126 kb YHS on a contig (ctg345, Supplementary Fig. 12) that again includes a FERR-R-like gene (named PdaFERR-R). PdaFERR-R shares twelve duplicated segments with its putative progenitor, PdARR9, but, unlike the P. deltoides duplication, only exon 1 of FERR is duplicated, and no HEMA1 exons are present (Supplementary Fig. 12). Aligning the sequences flanking P. davidiana’s YHS to P. deltoides and P. trichocarpa genomes revealed that this YHS is in the pericentromeric region of chromosome XIX. The different locations in species in the two subgenera support the hypothesis that their dioecy evolved independently20, yet involved a similar mechanism.

Neither P. davidiana, nor P. tremula has a complete MSL sequence, and therefore probably have no active PdeMSL orthologs. Partial MSL sequences were found in contigs in many genome regions, but no IncRNA transcripts were detected from the P. davidiana chromosome XIX peritelomeric and pericentromeric regions copies (Supplementary Fig. 10b).

Discussion
The development of female and male flowers in plants usually involves genes with spatially highly specific expression within the plant body or the flower primordia, and specific developmental timing. Most sex-limited and sex-biased genes in plants are not carried on the sex chromosomes, and their expression levels are probably controlled by an upstream sex-determining gene or genes31,45. Many genes probably function in the development of sexual dimorphisms of poplars, but we propose that FERR-R is
Fig. 5 Floral characterization of wild type and transgenic Arabidopsis plants. Each line number represents an independent transgenic line. a The first and second row show the stigma hypertrophy and exertion commonly observed in FERR-overexpressing plants; the third row shows extreme floral phenotypes observed in two FERR-overexpressing lines. A flower with two bent pistils (left) and one with carpel-like sepals (right, red arrow) were found in line 254, and line 454 (middle) produced a flower with one bent pistil and an unfused carpel. Due to the abnormal floral structures, it is difficult to stage the flowers in the third row from the top. However, by relying on the developmental stage of petals, stamens, and sepals, we classified these flowers as late stages 15–16. b Dissected flowers of wild type (WT) and MSL-overexpressing plants. WT flowers have four long and two short stamens (tetradynamous stamen), whereas the flowers from MSL-overexpressing plants had six long stamens (line 2), extra stamens (lines 51 and 13), or reduplicated Y-shaped stamens (lines 7 and 17).
the upstream female-suppressor gene. In *P. deltoides* XX females, FERR function is active due to the absence of the FERR-R gene, which is male-specific (being present only as a Y-linked copy in the YH51 of the Y haplotype, and absent from the X-linked region). Although FERR-R expression is not confined to the flower development stage when sex determination occurs, temporal specificity is provided by FERR, which is expressed only during the initiation of carpel primordia and early female flower development, and our transformation experiments in *A. thaliana* demonstrate that it is a female promoter. FERR belongs to the type-A RR gene family of transcription repressors in cytokinin signaling30, and altering cytokinin signaling has been reported to affect flower development46. RR genes have been suggested as sex determiners in other plants, including kiwifruit54, *S. purpurea*55 and *Ginkgo*57. Finally, a recent study showed that, in *P. tremula* (in subgenus *Leuce*, like *P. davidiana*), knockout of the ortholog of the *P. deltoides* FERR gene (ARR17 in *P. tremula*), in female trees converted them into males, indicating that *P. tremula* has a single-gene sex determination system59.

However, the hypothesis that FERR-R suppresses femaleness, creating males, does not account for the evolution of females from the cosexual ancestor from which dioecious poplars evolved. This ancestor must have had a functional female-promoting FERR gene, or a similar ARR17 homolog. Its loss of activity through partial duplication could have created male, which would produce a population with cosexuals and males (termed androdioecy). Androdioecy is, however, extremely rare, and requires very unusual conditions to evolve5. Even if such an evolutionary change did occur, another change is still necessary to generate females. This would require a mutation causing loss of male functions, to convert the cosexuals into females. A mutation within the same gene as the femaleness suppressor is formally possible, resulting in a single-gene sex determination system. However, it seems unlikely that the FERR gene could also produce male-sterility mutations, given FERR-R’s mode of action to create males, involving siRNA-guided cleavage of FERR transcripts, and our evidence that (in *A. thaliana*) over-expression affects only the pistils. A second gene therefore seems to be required to explain the evolution of dioecy in poplars.

Two possibilities exist for this second gene. First, a mechanism similar to the one discovered in persimmon is possible: a first mutation in an existing gene creates females, and then the gene involved becomes duplicated, suppressing the first mutation’s male-suppressing effects, and thus creating males. In this scenario, the second mutation is a dominant femaleness suppressor, and hence the first mutation becomes redundant for sex determination, and can become fixed in the species, leaving presence/absence of the duplication in control of sex organ development. A second possibility is that the MSL sequence is a domesticated transposable element that produces IncRNAs that promote male functions. If the cosexual ancestor possessed this gene, its loss would create females, but the gene would remain present in the male-determining region (as observed in the fully Y-linked region of *P. deltoides*). Our finding that MSL over-expression affects the *A. thaliana* androecium, but not the gynoecium, and altering expression of genes with male functions, is consistent with this possibility.

To test this possibility further, we asked whether MSL sequences are shared between different poplar species. Presence of a TE sequence at the same location in multiple species is unlikely, because TE insertions are rarely fixed in all individuals of a species. Our finding of a complete MSL sequence in *P. simonii* (in a different subgenus, *Tacamahaca*, from the one, *Aigeiros*, that includes *P. deltoides*, see Supplementary Data 2), therefore suggests that that this sequence may have a function in these species. The retention of a complete sequence of such an insertion for a long evolutionary time, corresponding to the divergence of two subgenera, is unexpected, and could signify that it has a plant function, such as increasing male functions. Alternatively a mutation that increased male functions could have arisen on chromosome XIX after the chromosome acquired its male-determining locus. If it arose on a Y haplotype that happened to carry the MSL insertion, the resulting selective sweep would have led to this haplotype replacing the ancestral version (though the maintenance of its complete sequence remains surprising). However, function of MSL is evident in heterologously expressed *A. thaliana* in this study. Without a test of its function in poplar, such as a knockout, we cannot currently determine its role in androecia development in poplar. Moreover, findings in this study show that MSL cannot be essential for male functions in all Salicaceae species, given that other poplar and willow species have only partial sequences, and that the knockout of a single gene in *P. tremula*, ARR17, converted female trees into males59.

In *P. davidiana*, in the same subgenus as *P. tremula* (subgenus *Leuce*), only partial MSL homologous sequences are present, and the FERR-R duplication is in the pericentromeric region of chromosome XIX (Supplementary Data 2). These two differences from the findings in subgenera *Aigeiros* (*P. deltoides*) and *Tacamahaca* (*P. simonii*) suggest that dioecy evolved independently in these taxa, and in subgenus *Leuce*. It is unknown how females of *P. davidiana*, or *P. tremula* evolved. If a gene like MSL was involved, it might no longer be essential after FERR-R appeared, and have been lost. It will be important in the future to locate the sex determining loci in more poplar species, to determine whether a peritelomeric location generally correlates with the presence of a complete MSL gene in the region, and a pericentromeric location with its absence. Nevertheless, the single-gene and possible two-gene systems in different sections of *Populus* make this genus particularly interesting for studying the evolution of sex determination in plants.

Many sex chromosomes, including the mammalian, bird, and Drosophila Ys, have evolved suppressed crossing over, sometimes followed by loss of large numbers of genes present on the X (and on the Y ancestor), causing hemizygosity in males, a process known as genetic degeneration58,49. The *P. deltoides* sex-determining genes also appear to be in a non-recombining region, but it is physically much smaller than in these animals, and even than the regions in plants such as papaya50,51, kiwifruit52, and asparagus5. Our analysis revealed that FERR-R originated by segmental duplication of the FERR gene (perhaps following a male-sterility mutation involving deletion of the MSL gene from the genome region into which the duplicate copy inserted, or perhaps followed by insertion of an MSL sequence into the region). The resulting hemizygosity of the *P. deltoides* Y-specific region is thus not due to degeneration, but to these two separate mutations. The hemizygosity can explain why the region carrying the sex-determining gene does not recombine between the SLR-Y haplotype and the X counterpart. Intriguingly, a hemizygous region was also found in asparagus5, and also in the date palm, *Phoenix dactylifera* (whose femaleness suppressor in males may also act via expressing a truncated competitor)53, suggesting that this deletion/insertion model represents a previously unrecognized route for the evolution of plant fully Y- and X-linked regions, similar to the evolution of hemizygous non-recombining regions controlling flower morphs in distylous plants54,55. Epigenetic regulation of reproductive genes is common in sex determination mechanisms in plants and animals56, including in the sex-determining mechanism by which monococious plants develop male or female flowers in differing developmental contexts, rather than under control of different plant genotypes57. Specifically, miRNAs and IncRNAs have been found to be involved58,59, including an RNAi process in the plant...
**Methods**

**Plant materials.** To map the sex locus, an intraspecific population of *P. deltoides* was established in 1998 by crossing two randomly selected siblings from a full-sib family. A total of 1077 offspring (550 females and 527 males) of the *F*₂ population were planted (6 × 6 m spacing) and maintained at Sihong Forest Farm in Jiangsu, China. To obtain the reference genomes for female and male *P. deltoides*, we sequenced the maternal parent and a randomly selected male progeny from the *F*₂ mapping pedigree. For genome-wide association study (GWAS), we sequenced the genomes of 49 unrelated females and 46 unrelated males (referred to as GWAS samples of *P. deltoides*), which were selected from the *P. deltoides* germplasm collections maintained at Sihong Forest Farm. The *P. deltoides* germplasm populations at nine different times: T1 (June 3), T2 (June 18), T3 (July 3), T4 (July 18), T5 (August 3), T6 (August 18), T7 (September 3), T8 (December 1), and T9 (January 15). To quantify the expression levels of genes, flower buds at the corresponding times were separately collected from three families of the female and male *P. deltoides* for biological replications. In the early developmental stages, it is difficult to harvest enough tissue for molecular experiments, and thus the flower buds with scales were used for T1–T5. For T5–T9, the descaled flower buds were used in molecular experiments. Since flower tissue only accounts for a small portion of the flower bud, we measured expression by using both the scaled and descaled flowers at T5 for comparison purposes.

Apart from *P. deltoides*, we also explored sex determination in *P. davidiana*, a more primitive species belonging to the subgenus *Leuce* [Leuce]. Leaf samples of *P. davidiana* were collected from a natural population along Duotong River in Qinghai, China in spring 2019. GWAS samples of *P. davidiana* included 49 females and 47 males.

**Sequencing, assembly, and annotation of *P. deltoides* genomes.** Young leaves of the maternal parent and a randomly selected male progeny were collected for DNA extraction and sequencing libraries construction. The detailed description of sequencing, assembly, and annotation processes are provided in Supplementary Methods 1 and 2.

**Mapping the sex locus of *P. deltoides*.** To map the sex locus, complete genetic maps were built for the female and male parents with 94 randomly selected offspring following a completely random block design in 1998. To characterize the developmental processes of flower buds, we collected the female and male flower buds from a female and a male tree in *P. deltoides* at nine different times: T1 (June 3), T2 (June 18), T3 (July 3), T4 (July 18), T5 (August 3), T6 (August 18), T7 (September 3), T8 (December 1), and T9 (January 15). To quantify the expression levels of genes, flower buds at the corresponding times were separately collected from three families of the female and male *P. deltoides* for biological replications. In the early developmental stages, it is difficult to harvest enough tissue for molecular experiments, and thus the flower buds with scales were used for T1–T5. For T5–T9, the descaled flower buds were used in molecular experiments. Since flower tissue only accounts for a small portion of the flower bud, we measured expression by using both the scaled and descaled flowers at T5 for comparison purposes.

**Validation of the SLR haplotypes.** To validate the reconstructed SLR haplotypes for the Y-specific region as described in Supplementary Method 3, sequence-specific primers were designed according to SLR-Y, and these primers were used to amplify DNA extracts from the sequenced male. All PCR reactions were performed with PrimeSTAR® GXL Premix (Takara, China) according to the manufacturer instructions. The details of the primers and PCR conditions were listed in Supplementary Table 1. The amplification products were purified using the AxyPrep DNA Gel Extraction Kit (Axygen Scientific, USA), cloned into pEASY-Blunt vector (Transgen Biotech, China), and then sequenced on the Sanger sequencing platform ABI 3730xl (Applied Biosystems, USA). The obtained sequences were assembled into an integrated contig and aligned to SLR-Y to evaluate the reliability of the haplotype reconstruction for the Y-specific region. The conservation of the Y-specific region was further confirmed by PCR amplification with Y-specific primer pairs of 651F/R, 655F/R, 657F/R, 661F/R, and 712F/R (Supplementary Table 1) against DNA extracts from 20 male and 20 female *P. deltoides*, which were randomly selected from the GWAS samples. The amplification products were examined by electrophoresis on an agarose gel (Supplementary Table 1). Primer extension was performed using the primer extension kit (Applied Biosystems, USA) and converted into gene dose data. The presence or absence of the genomic fragments is associated with individuals’ sexes, the windows were grouped into three read-coverage categories: 0 for no read, 1 for coverage 1–2, and 2 for coverage ≥3 (indicating a reliable number of reads, a widely used indicator to screen reliable SNP sets). Only the primary (best) locations were selected for read counting. The windows associated with sex phenotypes were identified using GEMMA following the description as in snp-GWAS.

**Characterizing flower development and quantifying genes expression.** Flower buds at different times (T1–T4) were collected and fixed using formalin-acetic acid. The samples were further dehydrated in a graded concentration of ethanol and embedded in paraffin. Serial sections were prepared by employing a Leica microtome. The sections were then mounted on microscope slides for staining with 1% safranin and examined using Carl Zeiss Imager M2 microscope (Zeiss, Germany).

RNA prep Pure Plant Kit (Tiangen, China) was used to extract RNAs from flower buds from a female and a male tree in *P. deltoides* to conduct qRT-PCR on a 7500 Fast Real-Time PCR System (Applied Biosystems, USA). In each 20 μL reaction volume, 100 ng of cDNA was used as templates. The PCR parameters used were as follows: 95 °C for 3 min, 40 cycles of 95 °C for 15 s, 60 °C for 1% safranin and examined using Carl Zeiss Imager M2 microscope (Zeiss, Germany).

RNA prep Pure Plant Kit (Tiangen, China) was used to extract RNAs from flower buds from a female and a male tree in *P. deltoides* to conduct qRT-PCR on a 7500 Fast Real-Time PCR System (Applied Biosystems, USA). In each 20 μL reaction volume, 100 ng of cDNA was used as templates. The PCR parameters used were as follows: 95 °C for 3 min, 40 cycles of 95 °C for 15 s, 60 °C for 1% safranin and examined using Carl Zeiss Imager M2 microscope (Zeiss, Germany).

**Quantifying the digital expression of genes and analyzing the interaction between FERR-R and FERR.** To quantify the digital expression of genes and to explore gene regulation patterns, Illumina sequencing experiments (Illumina NovaSeq 6000, USA) were performed at levels of mRNA, lncRNA, small RNA and DNA methylation. RNA-Seq was performed to quantify the expression of the four protein-coding genes (*MET1, CLC, TCP, and FERR*); whereas strand-specific lncRNA-Seq worked for transcripts with polyA tails. The samples applied for sequencing were listed in Supplementary Table 1. The quantification analysis of Illumina reads is described in Supplementary Method 4. The assembly of MSL sequences for *Populus deltoides* trees using Illumina sequencing reads was described in Supplementary Method 5.

**Experimental verification of interaction between FERR-R and FERR.** Isolation of *Populus* mesophyll protoplasts was performed using the PEG-Mediated plant protoplast transformation kit (Shanghai Maokang Biotechnology, China) with 3% (w/v) cellulose R10 (Yakult Pharmaceutical, Japan) and 0.8% (w/v) maceromize R10 (Yakult Pharmaceutical, Japan). The middle section of expanded *Populus* leaves (micropropagated for the clone Nanlin 895) were cut into 0.5–1 mm fine strips, digested for 30 min in the dark using a desiccator for Vacuum infiltrate, and then planted on the agar plate for 24 h. After 24 h, the leaf was immersed in 10 mL of B5 liquid medium supplemented with 3 mg/mL 2,4-D, 0.5% sucrose, and 0.5% gelrite. These leaves were cultured in the dark at 25 °C for 5 h without shaking. The protoplasts were used for transfection.
were harvested by filtering through a 70-μm pore nylon cloth and then suspended in the transfection buffer.

The sequence of one siRNA generated from FERR-R locus was designed onto the backbone of AMIR172a, driven by CaMV 35S promoter. The artificial miRNA generated in the construction mimics the siRNA of 21 nt from FERR-R. The artificial miRNA precursor sequence was synthesized on a Dr. Oligo384 (Biotylit, USA) and constructed into the vector pGWP7 using Gateway technology. The full length of the PCR product was amplified by using PrimeSTAR Max (TaKaRa, China) from cDNA of female bud. Site-directed synonymous mutagenesis of FERR was performed using the single-tube ‘megaprimer’ PCR method to generate the resistant version of FERR (referred to as FERR'). FERR' transcript was synthesized with substitutions in the complementary sequences of cDNA of female

Conif"
28. Geraudel, A. et al. Recent Y chromosome divergence despite ancient origin of dioecy in poplars (Populus). Mol. Ecol. 24, 3243–3256 (2015).

29. Müller, N. A. et al. A single gene underlies the dynamic evolution of poplar sex determination. Nat. Plants 6, 630–637 (2020).

30. Hellmann, E., Grunh, N. & Heyl, A. The more, the merrier: cytokinin signaling beyond Arabidopsis. Plant Signal. Behav. 5, 1384–1390 (2010).

31. Várn, A., Soryas, C. L., Tarr, P. T., Nimchuk, Z. L. & Bergmann, D. C. Modulation of an asymmetric division pathway by cytokinin and SPEECHLESS regulatory interactions in the Arabidopsis stomatal lineage. Dev. Cell 47, 53–66 (2018).

32. Chang, J. et al. Asymmetric distribution of cytokinins determines root hydrotropism in Arabidopsis thaliana. Cell Res. 29, 984–993 (2019).

33. Srivastava, A. K., Dutta, S. & Chattopadhyay, S. MYC2 regulates ARR16, a component of cytokinin signaling pathways, in Arabidopsis seedling development. Plant Cell 3, e0177 (2019).

34. Tuskan, G. A. et al. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313, 1596–1604 (2006).

35. Diggle, P. K. et al. Multiple developmental processes underlie sex differentiation in angiosperms. Trends Genet. 27, 368–376 (2011).

36. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. 15, 394–408 (2014).

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

38. Zhou, R. et al. Sequencing and analysis of the sex determination region of Populus trichocarpa. Genes 11, 843 (2020).

39. Kapusta, A. et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. 9, e1003470 (2013).

40. Johnson, R. & Guigo, R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. RNA 20, 959–976 (2014).

41. Wang, J. et al. Primate-specific endogenous retrovirus-driven transcription defines naïve-like stem cells. Nature 516, 405–409 (2014).

42. Zhao, T. et al. LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neo-functionalization. Genome Biol. 19, 195 (2018).

43. Carlevaro-Fita, J. et al. Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. Genome Res. 29, 208–222 (2019).

44. Wang, M. et al. Phylogenomics of the genus Populus reveals extensive interspecific gene flow and balancing selection. New Phytol. 225, 1370–1382 (2020).

45. Jones-Rhoades, M. W., Bartel, D. P. & Bartel, B. MicroRNAs and their regulatory roles in plants. Annu. Rev. Plant Biol. 57, 19–53 (2006).

46. Li, X. G. et al. Cytokinon overproduction-caused alteration of flower development is partially mediated by CUC2 and CUC3 in Arabidopsis. Gene 450, 109–120 (2010).

47. Zhang, H. et al. Recent origin of an XX/XY sex-determination system in the ancient plant lineage Ginkgo biloba. BiolRxiv https://doi.org/10.1101/517946 (2019).

48. Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B 355, 1563–1572 (2000).

49. Charlesworth, D., Charlesworth, B. & Marais, G. Steps in the evolution of heteromorphic sex chromosomes. Heredity 95, 118–128 (2005).

50. Wang, J. et al. Sequencing papaya X and Y chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc. Natl Acad. Sci. USA 109, 13710–13715 (2012).

51. VanBuren, R. et al. Origin and domestication of papaya Y chromosome. Genome Res. 25, 524–533 (2015).

52. Pilkington, S. M. et al. Genetic and cytological analyses reveal the recombination landscape of a partially differentiated plant sex chromosome in kiwifruit. BMC Plant Biol. 19, 172 (2019).

53. Torres, F. M. et al. Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix. Nat. Commun. 9, 3969 (2018).

54. Hsu, C. N. et al. Presence versus absence of CYP73A450 underlies the single-length dimorphism in primroses. Elife 5, e17956 (2016).

55. Shore, J. S. et al. The long and short of the S-locus in Turnera (Passifloraceae). New Phytol. 224, 1316–1329 (2019).

56. Pfaffler, F. Epigenetics of sex determination and gonadogenesis. Dev. Dyn. 242, 360–370 (2013).

57. Parkinson, S. E., Gross, S. M. & Hollick, J. B. Male sex determination and abaxial leaf fates are canalized by a factor that maintains repressed epigenetic states. Dev. Biol. 308, 462–473 (2007).

58. Chack, G., Meeley, R., Irish, E., Sakai, H. & Hake, S. The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1. Nat. Genet. 39, 1517–1521 (2007).

59. Golicz, A. A., Bhalla, P. L. & Singh, M. B. IncRNAs in plant and animal sexual reproduction. Trends Plant Sci. 23, 195–205 (2018).

60. Yin, T.-M., Wang, X.-R., Andersson, R. & Lerectceau-Kohler, E. Nearly complete genetic maps of Poins sylvestris L. (Scots pine) constructed by AFLP marker analysis in a full-sib family. Theor. Appl. Genet. 106, 1075–1083 (2003).

61. Garrison, E. & Marth, G. T. Haplotype-based variant detection from short-read sequencing. arXiv arXiv:1207.3907 (2012).

62. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).

63. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821–824 (2012).

64. Brunner, A. M., Yakovlev, I. A. & Strauss, S. H. Validating internal controls for quantitative plant gene expression studies. BMC Plant Biol. 4, 14 (2004).

65. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2 -ΔΔCT method. Methods 25, 402–408 (2001).

66. Yoo, S.-D., Cho, Y.-H. & Sheen, J. Arabidopsis mesophyll proplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. 2, 1565 (2007).

67. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16, 735–743 (1998).

68. Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early flower development in Arabidopsis. Plant Cell 2, 755–767 (1990).

69. Xue, L. et al. Evidence for roles of two Y sex genes in sex determination in Populus deltoide. BioRxiv https://doi.org/10.5281/zenodo.4076814 (2020).