Hyden et al. *Horticulture Research* (2021)8:170
https://doi.org/10.1038/s41438-021-00606-y

Horticulture Research
www.nature.com/hortres

ARTICLE Open Access

Integrative genomics reveals paths to sex dimorphism in *Salix purpurea* L

Brennan Hyden[1], Craig H. Carlson[1], Fred E. Gouker[1,2], Jeremy Schmutz[3,4], Kerrie Barry[3], Anna Lipzen[3], Aditi Sharma[3], Laura Sandor[3], Gerald A. Tuskan[5], Guanqiao Feng[6], Matthew S. Olson[6], Stephen P. DiFazio[7] and Lawrence B. Smart[1]✉

Abstract

Sex dimorphism and gene expression were studied in developing catkins in 159 F_2 individuals from the bioenergy crop *Salix purpurea*, and potential mechanisms and pathways for regulating sex development were explored. Differential expression, eQTL, bisulfite sequencing, and network analysis were used to characterize sex dimorphism, detect candidate master regulator genes, and identify pathways through which the sex determination region (SDR) may mediate sex dimorphism. Eleven genes are presented as candidates for master regulators of sex, supported by gene expression and network analyses. These include genes putatively involved in hormone signaling, epigenetic modification, and regulation of transcription. eQTL analysis revealed a suite of transcription factors and genes involved in secondary metabolism and floral development that were predicted to be under direct control of the sex determination region. Furthermore, data from bisulfite sequencing and small RNA sequencing revealed strong differences in expression between males and females that would implicate both of these processes in sex dimorphism pathways. These data indicate that the mechanism of sex determination in *Salix purpurea* is likely different from that observed in the related genus *Populus*. This further demonstrates the dynamic nature of SDRs in plants, which involves a multitude of mechanisms of sex determination and a high rate of turnover.

Introduction

Major progress has been made in recent years in identifying the master regulators of sex determination in plants, but less is known about the transcriptional networks that account for sex dimorphism. Transcription factors[1] (*Asparagus*), small RNAs[2] (*Diospyros*), and cytokinin-response regulators (*Actinidia* and *Populus*)[3,4] have all been identified as mechanisms of sex determination in angiosperms. Development of separate sexes, however, requires that the genes controlling sex determination regulate the transcription of many genes and metabolic pathways in order to coordinate development of sex-specific characteristics, such as gametes and floral morphology. With the notable exception of *Diospyros*[5], little is known about the metabolic pathways that are regulated by these sex-determination genes and the resulting transcriptome-wide expression differences.

Evidence from multiple angiosperm taxa, as well as leading sex determination models, suggests that in plants, sex is controlled by one or two regulatory factors, termed "sex determination" or "master regulator" genes[1,2,4,6,7]. These factors in turn lead to primary sex dimorphisms (andrecium and gynecium development) as well as secondary sex dimorphisms, such as floral volatile profiles, pigmentation, floral phenology, and organ morphology, and often involve both sex-linked and autosomal genes[8]. While primary sex dimorphisms are under direct control of the sex determination gene(s), secondary sex dimorphisms may be either under the control of sex-linked genes, or under direct control of the sex-determining genes themselves[7,8].

Correspondence: Lawrence B. Smart (lbs33@cornell.edu)
[1]Horticulture Section, School of Integrative Plant Science, Cornell University, Cornell AgriTech, Geneva, NY, USA
[2]Floral and Nursery Plants Research Unit, US National Arboretum, United States Department of Agriculture, Agricultural Research Service, Beltsville, MD, USA
Full list of author information is available at the end of the article
These authors contributed equally: Brennan Hyden, Craig H. Carlson

Because their expression in the opposite sex may be deleterious, linkage between sex determination genes and genes controlling secondary sexual dimorphisms may be favored by natural selection. Such genes are termed "sexually antagonistic"[9]. As a result, it can be challenging to discriminate among sex determination genes and other sex-linked genes that influence sex dimorphisms.

In angiosperms, the maintenance of separate sexes typically involves factors controlling sex determination residing alongside linked sexually antagonistic genes controlling sex dimorphisms on heterogametic sex chromosomes, where typically one sex is heterogametic and the other homogametic[6]. Two heterogametic systems, XY and ZW, have been observed in angiosperms. The XY system, where the male is the heterogametic sex, tends to be more prevalent, and is found in 84.7% of diecious angiosperm species, including *Asparagus officinalis*, *Carica papaya*, *Diospyros*, and *Phoenix dactylifera*. In the less common ZW system, females are heterogametic, as in *Fragaria* and *Silene otitis*[10,11]. The Salicaceae family is of particular interest for use as a model in understanding the evolution of sex chromosomes and sex determination in plants, because the family is almost exclusively diecious, yet the sex determination region (SDR) has been remarkably dynamic[4,12,13]. The Salicaceae family exhibits both ZW and XY heterogametic systems and SDRs that are localized in different chromosomes across species. The SDR in *S. purpurea* has been localized to a 6.73-Mb pericentromeric region on chr15W, which includes approximately 2 Mb of sequence that is not present in the corresponding region of chr15Z[14]. Other *Salix* spp. in the section *Vetrix* also have a chr15 ZW SDR, including *S. viminalis* and *S. suchowensis*[15,16], whereas *S. nigra* in the section *Salix* has a chr07 XY system[12]. In contrast, most *Populus* species have chr19 SDRs, including XY in *P. trichocarpa*[17] and ZW in *P. alba*[18] but, exceptionally, *P. euphratica* exhibits a chr14 XY SDR[19]. This indicates that sex determination has a complex evolutionary history in the Salicaceae, and that the SDR has shifted chromosomes as well as heterogametic systems, possibly through translocation, the rise of an entirely new locus becoming sexually antagonistic, or independent origins of diecy. Differing sex determination systems (as a result of sex chromosome turnover) in closely related taxa have been observed across both plants and animals. For instance, both XY and ZW systems are represented in *Silene*[10,11]. In cichilid fish, multiple sex determination loci and systems have been identified in specific populations[20]. Further, examples of a sex determination region evolving independently in related taxa include: sticklebacks, medaka fishes, and true frogs (Ranidae)[21]. Thus, even closely related species with similar genomic regions involved in sex determination, as in the Salicaceae, may have evolved independently and can involve unique mechanisms.

The Salicaceae family contains many species of economic importance in the genera *Populus* and *Salix*. Shrub willows (*Salix* section *Vetrix*) in particular, are grown throughout North America and Eurasia for bioenergy and bioproducts[22]. Despite its commonality across the Salicaceae family, diecy presents a challenge for breeding efforts and the cultivation of shrub willow, with sex showing linkage to biomass-related traits, such as leaf area[23] and catkins showing distinct phenology and secondary metabolite profiles between sexes, affecting pollinator and pest attraction[24,25]. There is a strong interest in understanding the genetic mechanisms controlling sex determination in *Salix*, along with the gene pathways involved in sex dimorphism, in order to advance current breeding efforts and genetic studies to improve *Salix* as a bioenergy crop. Nevertheless, there is still relatively little data characterizing sex-determination genes and sex dimorphism pathways in *Salix*, despite substantial research and identification of putative master regulators of sex in the related genus *Populus*[4,18,26,27]. Moreover, willows are typically insect pollinated, whereas members of the genus *Populus* are wind pollinated and show little evidence of sex dimorphism in vegetative traits[28] that may point to unique pathways and genes involved in sex dimorphism and underscore the need for sex-determination research that is specific to *Salix*.

A previous study of transcriptomic data in *Salix* identified differentially expressed genes associated with sex in shoot tips containing floral primordia[29]. It is hypothesized that sexually dimorphic genes are regulated through complex pathways that are ultimately controlled by one or more elements in the SDR, most likely via transcription factors, plant hormones, and/or small RNAs. This would be consistent with characterized sex determination systems in plants[1–4]. Elements in the SDR controlling sex determination are termed "master regulator genes" and likely control several top-level regulatory genes that may or may not be located in the SDR, which in turn regulate both primary and secondary sex characteristics through intermediate metabolic pathways, as described by Feng et al., 2020[11]. Unfortunately, identification of master regulator genes in *Salix purpurea* is complicated by an SDR that comprises nearly 40% of chr15W and contains 488 linked genes, including repetitive regions, and tandem duplications[14], requiring the use of transcriptomic data and coexpression analyses to characterize candidate genes.

This study captures the transcriptome-wide primary and secondary sex dimorphism profile in emerging inflorescences, which contain hundreds of achlamydeous flowers (lacking both sepals and petals) across a range of developmental time points, in addition to exhibiting distinct terpenoid and phenylpropanoid profiles leading to pigmentation and volatile organic compound emission

dimorpism[24]. Using eQTL analysis, we associated the expression levels of differentially expressed genes in catkins with genomic loci. We identified multiple genes associated in *trans* with the SDR that could serve as top-level regulators of primary and secondary floral sex dimorphisms under direct control by master regulator genes, as well as enriched pathways predicted to serve as intermediate pathways involved in sex dimorphism. Furthermore, based on these multi-omics results, six gene families are presented as candidates for the master regulators of sex: homologs of Arabidopsis *GATA15*, *ARR17*, *AGO4*, and *DRB1*, three genes coding hypothetical proteins, and a CCHC zinc finger. Characterizing the master regulator genes and the mechanisms of sex determination in willow provides insight into the complex evolution of diecy in the Salicaceae. These data represent one of the most comprehensive studies of sex dimorphism expression in a diecious crop plant, incorporating RNA expression, genotyping-by-sequencing (GBS), differential methylation, and small RNA expression, and provide a valuable addition to the nascent body of knowledge on sex-determination mechanisms in crop plants.

Results

Differential expression analysis

Principal component analysis showed a clear separation of male and female genotypes based on transcriptome-wide expression (Fig. S1). A total of 36,518 gene models, including alternative transcripts, accounting for 63.6% of all genome-wide protein-coding transcripts, were expressed in floral tissue across the 159 samples, with 24,074 (66%) showing significant differential expression between males (M) and females (F) (FDR ≤ 0.05) (Fig. 1, Fig. S2). There were 4676 genes with $\log_2$ M:F ≥ 1 (male-upregulated), while only 3247 genes with $\log_2$ M:F ≤ − 1 (female-upregulated).

Gene Ontology enrichment was performed on the male and female differentially expressed (DE) genes showing at least twofold expression differences between sexes (Tables S1–S4). Gene Ontology enrichment for the male-upregulated genes showed a significant overrepresentation of 76 terms, including pollen and anther development, male gamete development, the terpenoid biosynthesis pathway, and cytokinin metabolism (Table S1). The male-upregulated gene set showed an underrepresentation for 71 terms, including terms relating to transcription and RNA regulation, splicing, and modification (Table S2). Among the female-upregulated genes, there was an overrepresentation of RNA transcription and metabolism and shoot and organ development (Table S3), and an underrepresentation of 47 terms, including cell metabolism and biosynthesis (Table S4).

Small RNA identification and analysis

A total of 266,272 small RNA (smRNA) loci were identified, 146,807 of which contained smRNAs in the range of 20–25 nucleotides, the canonical length of siRNAs and miRNAs. Forty-five miRNA precursor loci representing 34 unique final miRNAs were predicted based on putative stem-loop structures consistent with miRNA biosynthesis (Table 1, Fig. 1, Table S5, Fig. S3). Forty of the 45 loci had identity with known and annotated plant smRNAs in the PmiREN database (www.pmiren.com), including 27 with matches to miRNAs in *Populus trichocarpa*. psRNATarget (plantgrn.noble.org/psRNATarget) was used to identify putative target genes for all miRNAs. Differential expression analysis on the miRNA loci revealed that 25 were significantly differentially expressed (FDR ≤ 0.05), including 18 with greater expression in males and seven with greater expression in females. Among the miRNA loci were two male-upregulated copies of miR172, which targets the MADS-box gene *APETALA2* in Arabidopsis[30] and four male-upregulated copies of miR156, one of which is located in the SDR.

Bisulfite sequencing and analysis

A total of 604,150 methylated regions were identified using DMRfinder after combining nearby methylated sites. Of these sites, 2018 show differential methylation, including 1465 with greater methylation in males and 553 with greater methylation in females (Fig. 1, Table S6). In total, 170 genes were identified with differentially methylated sites in their putative promoter regions (between 1 and 1000 bp ahead of transcription start site) (Table S7).

WGCNA network analysis and GO enrichment

WGCNA network analysis was performed to explore pathways that may be involved in floral sex dimorphism and development. Twenty-four modules were generated based on gene expression similarities (Fig. 2, Table 2). In total, 17,953 genes were assigned to modules, accounting for 49.8% of all expressed genes and alternative transcripts. The remaining unassigned genes were placed in the "gray" module. Three modules accounted for the majority of assigned genes: the "purple", "cyan", and "brown" modules. The "purple" module was strongly correlated to the female sex ($r^2 = 0.97$) and captured 30.2% of all female-expressed genes (Table 2), this module likely represents the genes involved in primary and secondary female sex dimorphism. In the "purple" module, 17 GO terms were enriched, including RNA and nucleic acid metabolism and regulation, photosynthesis, and phenylpropanoid metabolism (Table S8). This is consistent with the female differential expression of the majority of genes in this module (Table S3) and supports the contention that this module may be responsible for female sex-dimorphic traits. The cyan module was the most male-correlated ($r^2 = 0.92$) and included 45.5% of all

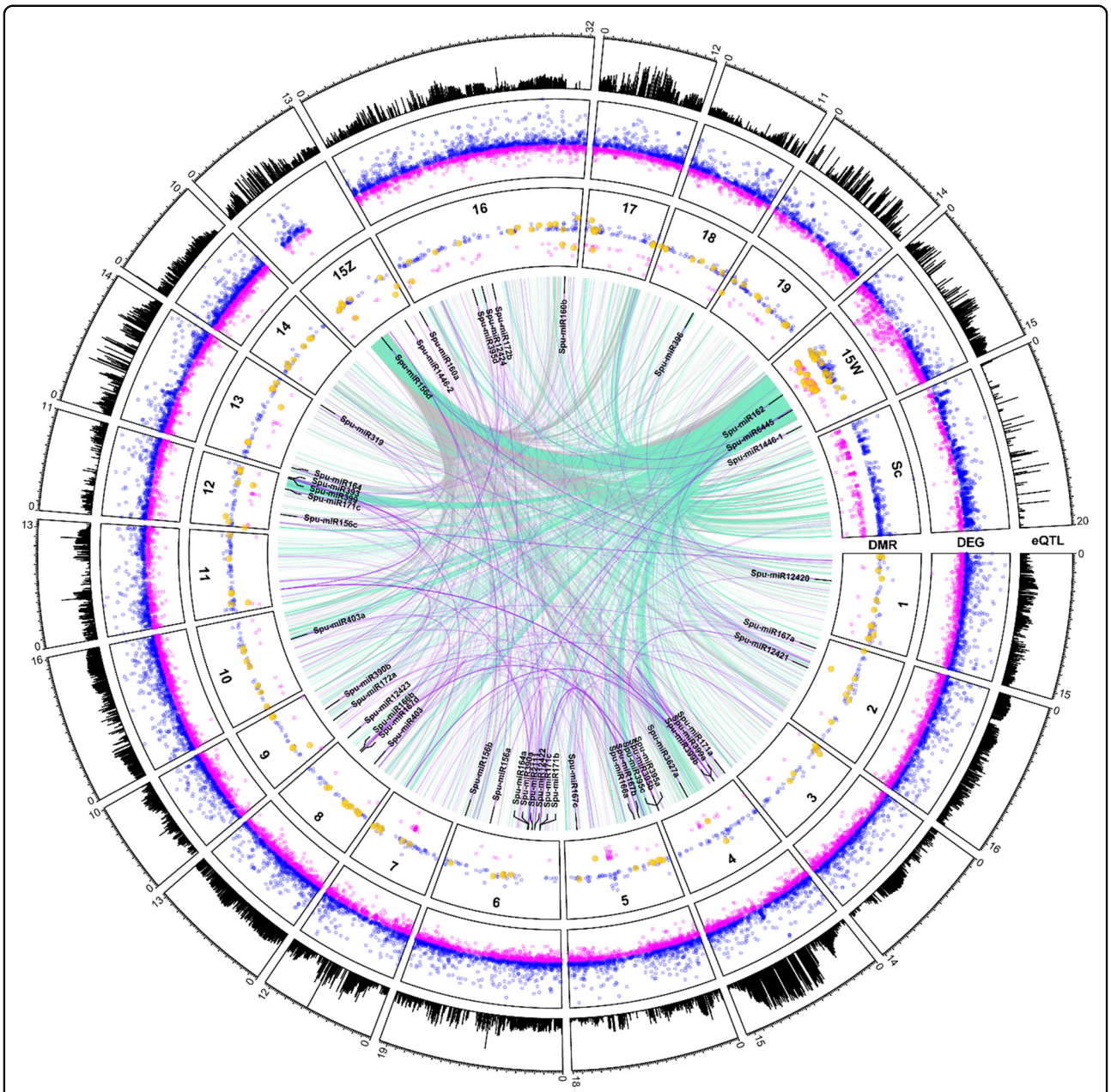

Fig. 1 Circos plot of eQTL, differentially expressed genes (DEG), and differentially methylated regions (DMR). Chromosome length (Mb) and total eQTL per locus are on the outer track (range 0–550), mapping sites of all differentially methylated regions are in the middle track (range −0.44 to 0.72 male:female methylation; male upregulated in blue, female upregulated in magenta, and DMRs in putative gene promoters in gold), and mapping sites of all differentially expressed genes are in the inner track (range $\log_2$ −10.3 to 14.4). Associations between gene expression and polymorphisms in the Z-SDR are shown in gray, while associations with the W-SDR are shown in aquamarine. The top ten predicted target sites for each miRNA are shown in purple. Scaffolds not assigned to a chromosome have been concatenated and labeled "Sc"

male-upregulated genes likely representing the genes in primary and secondary male sex dimorphism pathways. The "cyan" module contained 230 significantly enriched GO terms, notably multiple terms related to pollen development (Table S9). The "brown" module consisted mostly of genes not showing differential expression, and probably represents gene pathways involved in basic cell and biological processes.

eQTL analysis

A total of 1,381,813 *cis*-eQTL and 811,499 *trans*-eQTL (FDR ≤ 0.05) were identified after accounting for

Table 1 Forty-five miRNA loci identified and their respective PmiREN database matches with known plant small RNAs as well as the top predicted hit in the *S. purpurea* genome using psRNATarget

ID	Chr	Start	End	Match	log_2 (M:F)	FDR	Top Predicted Target	Top Target Description
Spu-miR12420	1	5390167	5390252	N/A	−0.731	N/A	Sapur.003G010300.3	helix loop helix DNA-binding domain protein
Spu-miR167a	2	2546206	2546454	Ptr-miR167a	2.265	**6.754E-24**	Sapur.016G022200.1	short-chain dehydrogenase-reductase B
Spu-miR12421	2	6755707	6755829	N/A	−0.062	8.278E-01	Sapur.008G043000.2	DNA/RNA-binding protein Kin17 motif protein
Spu-miR171a	4	476403	476580	Ptr-miR171i	0.751	**4.385E-05**	Sapur.010G159800.2	triosephosphate isomerase
Spu-miR399a	4	3155960	3156093	Ath-miR399a	0.534	6.197E-01	Sapur.011G036300.2	phosphate 2
Spu-miR399b	4	3158638	3158815	Ath-miR399a	0.403	4.865E-01	Sapur.011G036300.2	phosphate 2
Spu-miR3627a	4	9319512	9319746	Mes-miR3627	0.211	5.732E-01	Sapur.005G170700.1	autoinhibited calcium ATPase
Spu-miR395a	4	15610953	15611114	Ptr-miR395a	0.601	2.029E-01	Sapur.002G074200.2	sulfate/bicarbonate/oxalate exchanger and transporter sat-1
Spu-miR395b	4	15612449	15612536	Ptr-miR395a	0.551	3.850E-01	Sapur.002G074200.2	sulfate/bicarbonate/oxalate exchanger and transporter sat-1
Spu-miR395c	4	15628880	15629146	Ptr-miR395a	0.747	9.340E-02	Sapur.002G074200.2	sulfate/bicarbonate/oxalate exchanger and transporter sat-1
Spu-miR167b	5	2940103	2940284	Ptr-miR167a	2.069	**2.510E-23**	Sapur.016G022200.1	short-chain dehydrogenase-reductase B
Spu-miR166a	5	3924142	3924378	Ptr-miR166a	−0.728	**4.897E-04**	Sapur.003G027100.1	Homeobox luecine zipper HB-15
Spu-miR167c	5	15569344	15569484	Ptr-miR167a	1.791	**3.386E-22**	Sapur.016G022200.1	short-chain dehydrogenase-reductase B
Spu-miR171b	6	3831004	3831090	Gma-miR171l	1.445	**3.383E-03**	Sapur.010G081700.2	O-Glycosyl hydrolases family 17 protein
Spu-miR171c	6	3835891	3836056	Gma-miR171l	1.578	**5.840E-03**	Sapur.010G081700.2	O-Glycosyl hydrolases family 17 protein
Spu-miR12422	6	4882875	4883149	N/A	−0.169	5.722E-01	Sapur.016G204200.2	enoyl-CoA hydratase/isomerase D
Spu-miR2111	6	5427587	5427760	Ptr-MIR2111a	−0.672	**1.418E-02**	Sapur.003G163400.1	putative casein kinase
Spu-miR390a	6	6180906	6181056	Ptr-miR390a	0.020	1.000E + 00	Sapur.006G039400.1	LRR receptor-like tyrosine-kinase
Spu-miR164a	6	6317292	6317530	Ptr-miR164a	0.896	**3.360E-07**	Sapur.15 WG015900.1	NAC domain containing protein 100
Spu-miR156a	6	13846849	13846964	Mac-miR156f	2.084	**2.512E-15**	Sapur.15 WG078700.4	Squamosa promoter-like 15
Spu-miR156b	6	19006933	19007076	Ptr-miR156a	2.773	**1.038E-17**	Sapur.15WG078700.4	Squamosa promoter-like 15
Spu-miR403	8	5423300	5423499	Ptr-miR403a	−0.299	5.623E-02	Sapur.012G088900.1	Argonaute 2
Spu-miR167d	8	8894197	8894487	Ptr-miR167a	1.104	**9.631E-12**	Sapur.016G022200.1	short-chain dehydrogenase-reductase B
Spu-miR166b	8	9245708	9245886	Ptr-miR166o	−0.314	1.128E-01	Sapur.006G067500.2	Basic-leucine zipper (bZIP) transcription factor
Spu-miR12423	8	12743624	12743755	N/A	0.300	6.822E-01	Sapur.011G106700.4	snapin/pallidin protein
Spu-miR172a	9	2892935	2893114	Ptr-miR172a	1.193	**6.288E-05**	Sapur.002G186500.2	GTP-binding protein, putative
Spu-miR390b	9	5612879	5613152	Ptr-miR390a	−2.031	**1.640E-39**	Sapur.006G039400.1	LRR receptor-like tyrosine-kinase
Spu-miR403a	10	10076881	10077088	Ptr-miR403a	−0.335	**3.498E-02**	Sapur.012G088900.1	Argonaute 2

Table 1 continued

ID	Chr	Start	End	Match	log_2 (M:F)	FDR	Top Predicted Target	Top Target Description
Spu-miR156c	12	3197390	3197490	Aco-miR156m	3.648	**5.405E-17**	Sapur.017G006900.1	aminophospholipid ATPase 2
Spu-miR171c	12	8351984	8352111	Ath-miR171b	0.411	1.412E-01	Sapur.005G100800.1	GRAS family transcription factor
Spu-miR399	12	10545966	10546254	Ath-miR399a	0.850	**3.106E-02**	Sapur.011G036300.2	phosphate 2
Spu-miR393	12	10816595	10816773	Aof-miR393a	3.748	**2.366E-47**	Sapur.016G197800.1	transport inhibitor response protein
Spu-miR164	13	842171	842291	Ptr-miR164a	0.805	**2.922E-03**	Sapur.15WG015900.1	NAC domain containing protein 100
Spu-miR319	13	14611610	14611886	Ath-miR1319a	0.320	8.193E-02	Sapur.016G120400.1	transcription factor Myb3, putative
Spu-miR162	15W	6070157	6070284	Ptr-miR162a	0.397	**9.255E-03**	Sapur.002G137200.1	endoribonuclease dicer-like protein
Spu-miR6445	15W	9647264	9647437	Ptr-miR6445a	0.417	**4.988E-03**	Sapur.009G127600.1	ribosomal protein L32
Spu-miR1446-1	15W	13324567	13324662	Ptr-miR1446a	0.367	6.057E-01	Sapur.011G018900.2	alpha-farnesene synthase
Spu-miR156d	15Z	5200310	5200523	Aco-miR156m	3.574	**5.520E-17**	Sapur.017G006900.1	aminophospholipid ATPase 2
Spu-miR1446-2	15Z	10970968	10971061	Ptr-miR1446a	0.060	9.441E-01	Sapur.011G018900.2	alpha-farnesene synthase
Spu-miR160a	16	80655	80741	Ptr-miR160a	1.069	2.886E-01	Sapur.008G028800.1	Auxin Response Factor 16
Spu-miR395d	16	11568408	11568564	Ptr-miR395a	0.706	9.748E-02	Sapur.002G074200.1	sulfate/bicarbonate/oxalate exchanger and transporter sat-1
Spu-miR12424	16	13458436	13458691	N/A	0.536	5.116E-01	Sapur.008G128400.1	Serine/Threonine kinase domain protein
Spu-miR172b	16	15596694	15596955	Ptr-miR172a	2.318	**7.535E-04**	Sapur.012G062700.1	armadillo repeat only 4 protein
Spu-miR160b	16	30160705	30160920	Ptr-miR160a	0.730	9.118E-02	Sapur.008G028800.1	Auxin Response Factor 16
Spu-miR396	18	10287527	10287770	Mac-miR396t	0.860	**2.360E-06**	Sapur.007G075900.1	hypothetical protein

Negative log_2-fold change values are female upregulated and positive values are male upregulated. Those passing the significance threshold (FDR ≤ 0.05) are in bold

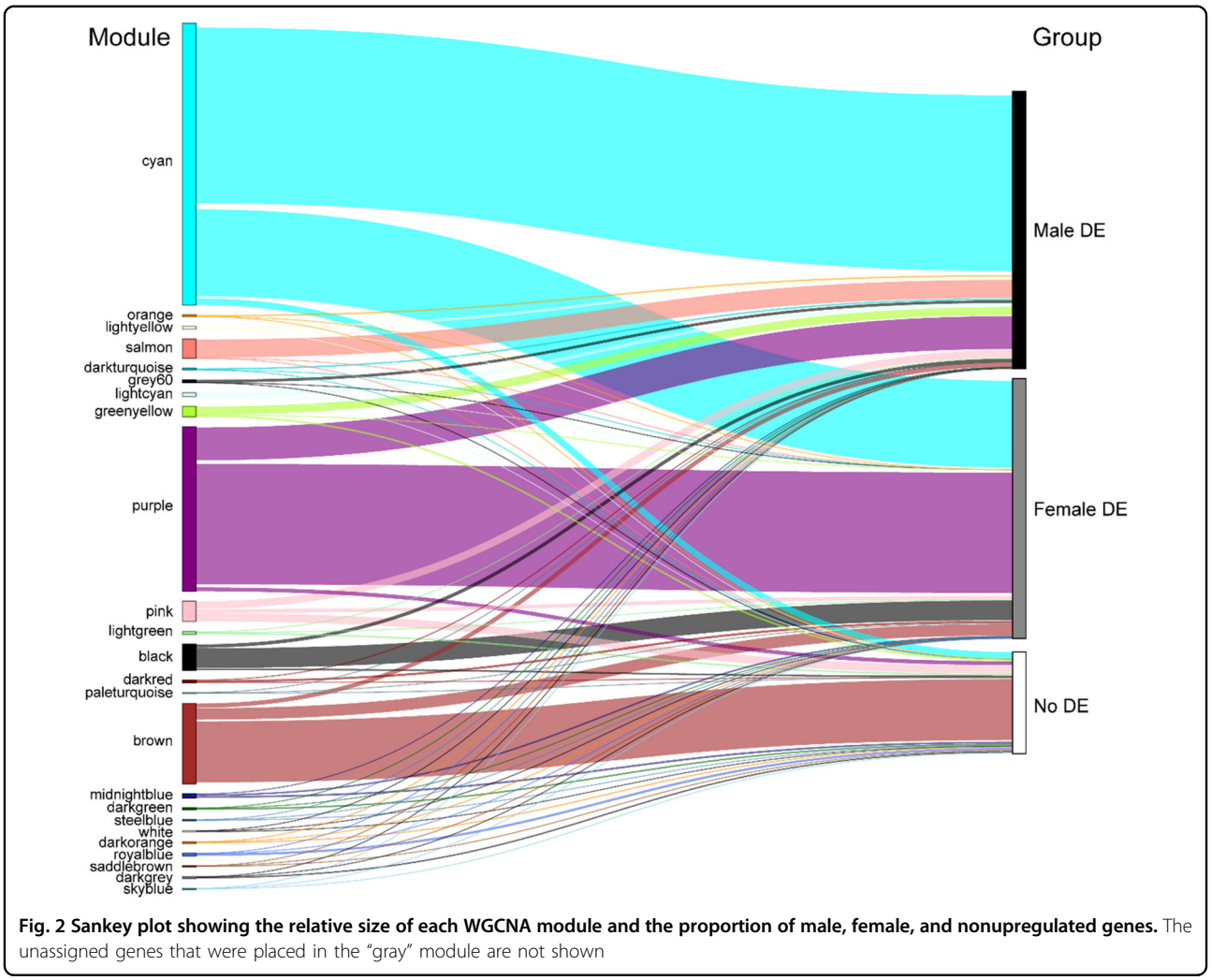

Fig. 2 Sankey plot showing the relative size of each WGCNA module and the proportion of male, female, and nonupregulated genes. The unassigned genes that were placed in the "gray" module are not shown

covariates. Notably, there appears to be an eQTL "hotspot" on chr07, to which the expression levels of 550 genes are associated (Fig. 1) The exceptional number of genome-wide eQTL at this locus suggests that it may have a major role in regulating sex dimorphic gene expression in catkin tissue. Expression levels of 2127 genes were found to be associated with polymorphisms in the SDR, of which 1686 were *trans* (Fig. 1). To identify top-level regulatory and intermediate pathway genes, 2127 genes with eQTL in the SDR were a subset for genes with involvement in secondary metabolism (22), hormone signaling (16), RNA splicing and regulation (4), or transcription (96) (Table S10). Of the 96 genome-wide transcription factors found to have SDR eQTL, 70 were differentially expressed in either males or females. Because these transcription factors include genes relating to floral development, phenylpropanoid production, and cytokinin signaling, they are candidates for top-level regulatory genes that may regulate further downstream expression. However, confirmation of such roles will require further investigation using methods such as ChIP-Seq and DAP-Seq. Fourteen MADS-box and floral development genes, 15 phenylpropanoid pathway genes, and five terpenoid pathway genes were also found to have eQTL in the SDR (Table S10), representing candidates for intermediate pathway genes directly responsible for dimorphisms in floral morphology, pigmentation, and volatile and secondary metabolite profiles.

Candidate master regulator gene identification

We identified eleven genes that are strong candidates as master regulators of sex using the following criteria: (1) presence on chr15W and absence from chr15Z, (2) a significant $\log_2$ M:F < -1, (3) presence in the female-correlated "purple" WGCNA module, and (4) gene annotation either consistent with a possible floral sex dimorphism pathway or of unknown function. Genes meeting all four of these criteria are expected to be present only in females and have expression levels and module membership that would implicate them in sex

Table 2 All WGCNA modules with the total number of genes belonging to each, along with the number of differentially expressed genes; the gray module contains all genes that were unable to be assigned to a specific module

Module Name	Correlation (M:F)	*p*-value	Total	Male DE	Female DE
purple	−0.97	3.0E-94	4624	968	3542
cyan	0.92	9.0E-65	7912	5165	2554
salmon	0.8	3.0E-36	538	533	1
lightyellow	0.68	8.0E-23	78	77	0
orange	0.67	4.0E-22	45	45	0
black	−0.65	2.0E-20	728	100	581
darkturquoise	0.65	4.0E-20	50	0	1
paleturquoise	0.65	4.0E-20	30	0	24
lightcyan	0.55	5.0E-14	107	104	0
grey60	0.46	2.0E-09	86	0	3
darkred	−0.39	6.0E-07	70	0	64
greenyellow	0.38	1.0E-06	294	241	1
white	−0.21	9.0E-03	37	0	21
darkgreen	0.17	3.0E-02	57	23	0
steelblue	0.16	5.0E-02	30	6	0
brown	−0.12	1.0E-01	2259	127	336
darkgrey	−0.12	1.0E-01	15	3	3
pink	0.12	1.0E-01	570	217	113
midnightblue	−0.1	2.0E-01	122	18	47
royalblue	0.11	2.0E-01	76	10	0
saddlebrown	0.085	3.0E-01	33	3	1
darkorange	0.073	4.0E-01	42	5	2
skyblue	0.068	4.0E-01	33	2	1
lightgreen	0.058	5.0E-01	86	29	6
gray	0.19	8.0E-03	18565	4552	4427

dimorphism. Four copies of *ARR17*, a truncated *AGO4* gene, *DRB1*, *GATA15*, a CCHC zinc finger nuclease, and three genes coding hypothetical proteins met these criteria and were identified as candidate master regulator genes (Table 3).

Discussion

We have used a combination of differential expression, coexpression, and eQTL analyses to identify genes that are candidate master and top-level regulators of sex expression in *S. purpurea* floral tissue.

RNA-Seq and differential expression

Total RNA-Seq and small RNA-Seq captured the unique sex-specific transcriptomic profiles during catkin development, after floral meristem differentiation and prior to maturation of any stamens or pistils. Within a single maturing catkin, there are hundreds of individual flowers across a range of developmental stages, resulting in pooled expression data from across floral development time points as well as tissue types (i.e., anthers/pistils, floral bracts, and peduncles). In addition to the primary sex dimorphism genes responsible for anther and carpel development, this enables the identification of secondary sex dimorphisms, such as genes involved in pigmentation, volatile emission, and differences in catkin phenology, which can also inform differences in vegetative emergence and secondary metabolites. By using network analysis and incorporating genomic data through eQTL, we can hypothesize how the SDR may regulate differential gene expression in catkins. Nearly two-thirds of all expressed genes in the floral tissue exhibited differential expression

Table 3 Eleven candidate master regulator genes and their log_2 male:female (M:F) differential expression values and *Arabidopsis thaliana* homologies

Gene ID	log_2 (M:F)	FDR	Arabidopsis homolog	Arabidopsis gene name	Description	Start position (Mb)
Sapur.15WG062800.1	−7.491	2.23E-250	AT3G06740.1	*GATA15*	GATA transcription factor 15	7088851
Sapur.15WG068800.1	−6.282	5.82E-167	AT2G15180.1	–	Zinc knuckle (CCHC-type) family protein	8120121
Sapur.15WG073500.1	−3.973	1.67E-60	AT3G56380.1	*ARR17*	response regulator 17	8795968
Sapur.15WG073900.1	−2.802	1.93E-25	AT3G56380.1	*ARR17*	response regulator 17	8828508
Sapur.15WG074000.1	−4.026	3.82E-62	AT3G56380.1	*ARR17*	response regulator 17	8842708
Sapur.15WG074300.1	−1.845	5.68E-13	AT1G09700.1	*DRB1*	dsRNA-binding domain-like superfamily protein	8862381
Sapur.15WG074400.1	−7.842	3.99E-263	AT2G27040.1	*AGO4*	Argonaute family protein	8875048
Sapur.15WG074900.1	−1.305	1.06E-25	–	–	–	8957099
Sapur.15WG075200.1	−4.004	3.94E-61	AT3G56380.1	*ARR17*	response regulator 17	9007096
Sapur.15WG075300.2	−3.534	4.02985E-52	AT1G77270.1	–	hypothetical protein	9016588
Sapur.15WG075700.1	−3.840	2.71E-53	–	–	–	9065664

between males and females. This number is due in part to the large sample size of 159 individuals, whereby there is enough statistical power to detect even slight differences in expression. Nevertheless, over 21% of the expressed genes showed at least twofold expression differences between sexes, providing evidence of global expression differences, which would require robust transcriptional regulation, ultimately leading back to the sex-determinant genes in the SDR. These genes provide important clues about the regulation of sex determination in this species and the molecular mechanism responsible for diecy and floral sex dimorphism, as described in more detail below.

Type A response regulator (ARR17)

Four copies of *ARR17*, a type A cytokinin-response regulator, in the SDR, show high levels of expression in female *S. purpurea*: Sapur.15WG073500, Sapur.15WG073900, Sapur.15WG074000, and Sapur.15WG075200. Two additional copies of *ARR17* are present on chr19 but are not differentially expressed. The cytokinin signaling pathway has been proposed as a common pathway for sex determination in angiosperms[10]. Cytokinin-response regulators serve as feminizing factors in *Actinidia*, where they are master regulators[31], and *Diospyros*, where they act as top-level regulators downstream of the SDR[2,5]. There is recent evidence implicating *ARR17* as the master regulator of sex in the closely related genus *Populus*, where it may function as a feminizing factor whose expression is suppressed in males by small RNAs[4].

The presence of two complete copies of *ARR17* on *S. purpurea* chr19, expressed in both males and females, suggests that the dosage of *ARR17* (eight copies in females vs. four in males) may play a role in sex determination in willow. Interestingly, these findings suggest a different mechanism for *ARR17* than the leading model for sex determination in *Populus* proposed by Müller et al. (2020)[4]. They proposed that functional *ARR17* in *P. alba* (ZW) is a feminizing factor, and in XY species, *ARR17* is silenced by inverted repeats on the Y chromosome through the RNA-directed DNA methylation (RDDM) pathway, leading to a male phenotype. To confirm this, they silenced the *ARR17* gene in an early-flowering female line and observed male flowers in tissue culture[4]. We found no evidence of an *ARR17* RNA interference mechanism in *S. purpurea* catkins. *Salix purpurea* has a similar truncated inverted repeat of *ARR17* on chr15Z, but we did not observe small RNAs mapping to the *ARR17* genes and their proximal regions, nor to the *ARR17* homologs on *Salix* chr19. There were also no differential methylated regions in the putative promoter regions of any of the *ARR17* genes, and *S. purpurea* males show expression of the *ARR17* copies on chr19, whereas in *P. trichocarpa* males, there is no *ARR17* expression. Furthermore, Carlson et al. (2017) did not find that ARR17 was differentially expressed in shoot tips containing floral primordia, indicating that this mechanism is not present at an earlier floral development stage either[29]. Taken together, these results suggest that the RNA-

interference mechanism of *ARR17* may be absent in *S. purpurea*.

The observation of *ARR17* expression in both male (chr19) and female (chr19 and chr15W) *S. purpurea*, combined with a lack of small RNA loci in these same regions, demonstrates that the *Salix* sex-determination mechanism is likely different from the model proposed by Müller et al. (2020)[4]. Instead, our data suggest that if *ARR17* is a master regulator in *S. purpurea*, it likely involves a unique mechanism, possibly through gene dosage, such that a threshold of *ARR17* expression must be reached to activate a switch from male-to-female development. Alternatively, there may be another feature in the *S. purpurea* SDR that is suppressing this silencing mechanism, one such candidate is the adjacent *AGO4* homolog described below.

ARGONAUTE 4 (AGO4)

A single copy of an Arabidopsis *AGO4* homolog, Sapur.15WG074400, is present within the *ARR17*-inverted repeat region of the chr15W SDR[14] that exhibits a $\log_2$ M:F expression of −7.94, and has a cis-eQTL in the SDR. *AGO4* is a component of the RNA-induced silencing complex (RISC) in the RNA-dependent DNA methylation (RDDM) pathway, where it binds small RNAs and silences mRNA[32]. In the bisulfite sequencing data, nearly three times as many regions showed increased methylation in males (1465) compared with females (553), supporting that methylation activity is downregulated in females and may have a role in mediating sex dimorphisms (Fig. 1). The SDR *AGO4* gene appears to be truncated to only 79 amino acids in length compared with five other catkin-expressed *AGO4* homologs in *S. purpurea*, which are 893–922 amino acids, and has multiple indels and substitutions when aligned (Fig. S4A). The most similar AGO4 paralog to Sapur.15WG074400 by MUSCLE multiple-sequence alignment is Sapur.008G00580 which has a nearly sevenfold greater expression in males (Fig. S4B, Table S11). We speculate it is possible that the truncated version of *AGO4* is interfering with expression of the full-length Sapur.008G00580 in males by a long noncoding RNA. This could have wide-ranging effects on sexually dimorphic gene expression and could explain the decreased genome-wide methylation observed in females. The findings from the bisulfite-sequencing data indicate that methylation is globally reduced in females. We hypothesize that the Sapur.15WG074400 could be competing for binding of siRNAs with a full-length *AGO4* and sequestering male-specific RDDM in females. These global methylation differences could be responsible for sex determination, such as in *Melandrium album* where demethylation of male plants results in monoecy, with no effect on female plants[33]. Such a mechanism could also explain *ARR17* expression levels, and why no small RNAs were observed mapping to *ARR17* in *Salix*, despite evidence for this mechanism in *Populus*.

Double-stranded RNA-binding protein 1 (DRB1)

A copy of a *DRB1* homolog, Sapur.15WG074300, is also located in the *ARR17* inverted repeat region of the chr15W SDR and is adjacent to *AGO4*. In Arabidopsis, *DRB1* is involved in RNA-mediated post-transcriptional silencing, working directly with *DCL1* in miRNA processing[34]. RNA modification, RNA metabolic process, and regulation of transcription were all among the enriched GO terms in the female differentially expressed genes and could be regulated directly or indirectly by *DRB1*.

Transcription factor GATA15

A female-expressed homolog of *GATA15*, Sapur.15WG062800, is located in the W-specific region of the SDR and shows a cis-eQTL association with polymorphisms on Chr15. *GATA15* is a transcriptional regulator that binds GAT or GATA motifs in gene promoters and is involved in cell differentiation, morphogenesis, and development[35]. This is consistent with the GO enrichment analysis of female-expressed genes, which contains many significant terms related to morphogenesis and development. Furthermore, chr15 *GATA15* was found by Carlson et al. (2017) to be differentially expressed in F_1 *S. purpurea* shoot tips containing floral primordia[29], suggesting that this may be the earliest cue for floral sex differentiation, which would implicate it as a master regulator gene. While functional genomics data are required to elucidate its precise function, its expression in females both during floral differentiation and catkin emergence suggests that it could be directly involved in gynecium development.

Genes of unknown function

Four genes were identified that fit the criteria for candidate master regulator genes, but whose functions are not known or whose annotations are insufficient for further analysis. These included Sapur.15WG068800, a CCHC-type zinc finger, and three hypothetical proteins: Sapur.15WG075300, Sapur.15WG074900, and Sapur.15WG075700.

While there is mounting evidence pointing toward *ARR17* as the master regulator in *Populus* spp[4,26,27], the evidence for different expression profiles of *ARR17* in *Salix*, as well as the presence of additional candidate genes, suggests that the mechanism may be more complicated or altogether different in *Salix*. Nevertheless, expression data from the *ARR17* homologs in *S. purpurea* do support a possible role in sex determination, either as a single gene master regulator or part of a two-gene system in conjunction with another master regulator, and would provide further evidence to support cytokinin response as a common mechanism for diecy in angiosperms, as

suggested by Montalvão et al[10]. Further functional genomics studies will be necessary to elucidate the precise functions of candidate master regulators and their role in sex determination.

SDR regulation of floral sex dimorphism

Among the floral development genes with eQTL in the SDR were homologs of *AGL11* and *AGL32* (ovule development and seed formation)[36], *AGL29* and *AGL30* (pollen development)[37], and *AGL6* (floral meristem differentiation and gamete development)[38], as well as *TOC1, WOX1, RGA*, and *CONSTANS*[39] (Table S10). While differential expression of MADS-box genes is expected in floral tissues, their association with the SDR through eQTL, even after accounting for sex as a covariate, suggests that the SDR may have a direct role in controlling expression of these genes and subsequent primary sex dimorphisms. *TOC1* and *CO* regulate circadian rhythm and day-length responses, while *RGA* represses floral growth in the absence of gibberellic acid[39]. *WUSCHEL* (*WOX1*) genes have a well-characterized function in meristem organization and have been shown to interact with *AGAMOUS* in floral meristem development[39]. *AGL11* is important in ovule and seed development, and has been shown to interact with cytokinin to control fruit size[36,39]. These floral development genes may have specific roles as "entry points" for the SDR into the floral development pathway to regulate the development of a particular sex (Fig. 3). Cronk and Müller (2020) proposed that *ARR17* may act as a feminizing master regulator in *Populus* through the suppression of *PISTILLATA* (*PI*) or *APETALLA3* (*AP3*) MADS-box genes[27]. Importantly, there was no association of either *PI* or *AP3* expression with genomic variation in the SDR, despite the fact that *PI* shows very high levels of expression in males ($\log_2$ M:F 13.21) (Table S11), further suggesting that the mechanism of sex determination in *Salix* may be different from *Populus*.

eQTL analysis revealed several loci with an exceptionally high number of *trans*-eQTL (Fig. 1). Intriguingly, the hotspot with the greatest number of *trans*-eQTL is located in the region on chr07 homologous to the *S. nigra* SDR[12], which could implicate its role as an ancestral SDR in *Salix*, or explain a fitness advantage of a chr07 SDR in *S. nigra* when linked with sex-dimorphism genes in this region. Approximately 250 kb from this locus on chr07 is Sapur.007G068100, a homolog of *AGL32*, a MADS-box gene involved in ovule development[36]. The expression of this gene is associated in *trans* with the chromosome-15 SDR region, further supporting its role as a top-level regulatory gene under direct regulation by the SDR.

Compounds produced from the phenylpropanoid and terpenoid pathways are well-characterized in Salicaceae, and there is evidence that floral volatile, terpenoid, and phenolic glycoside profiles differ substantially between males and females, which affects both pollinator attraction and herbivory, traits that are likely to be evolutionary drivers of diecy and affect cultivar yield[24,25]. In support of this, we identified five terpenoid pathway genes and 15 phenylpropanoid pathway genes with eQTL in the SDR. These include both genes involved in the core phenylpropanoid pathway, and biosynthesis of specific compounds, including naringenin, flavenol glucosides, and sesquiterpenoids (Table S10). This provides evidence supporting a direct link between the SDR and synthesis of these compounds, by an as-yet unknown mechanism.

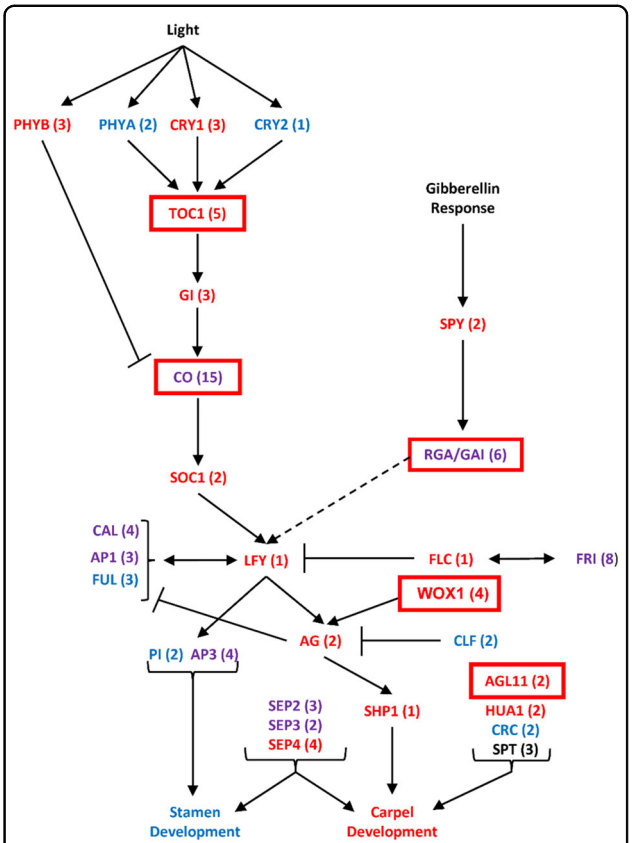

Fig. 3 Flowering gene expression network. The diagram displays known roles of flowering pathways and genes in Arabidopsis whose homologs are expressed in *S. purpurea* catkins. The number of copies of each gene in *S. purpurea* is shown in parentheses. Female differentially expressed genes are shown in red, male differentially expressed in blue, and genes with both male and female differentially expressed copies in purple. Genes with eQTL associations to the sex-determination region are outlined in red boxes and represent possible entry points for SDR regulation of the flowering pathway. Figure adapted from ref. [54]

Small RNA regulation of sex dimorphism

Of the 45 identified miRNA loci, 18 were male differentially expressed and seven were female differentially expressed. Among the putative targets of these miRNAs were dicer-like genes, squamosa promoter-like genes, and

auxin-response factors, and transcription factors (Table 1) providing evidence that miRNAs are likely to be a component of floral sex-dimorphism regulation. Notably, five miRNAs were identified that had no match with any small RNAs in the PmiREN database and could represent genus- or species-specific micro-RNAs. Furthermore, 13 miRNA loci with matches in the pmiREN database were not matched to a known *P. trichocarpa* miRNA, which is the closest species for which extensive small RNA data are available (Table 1). This suggests that S. purpurea may utilize different sets of small RNAs in floral development relative to Poplar, which likely has implications on sex dimorphism and determination. Expression results showed that four miR156 and two miR172 homologs have greater expression in male floral tissue. In Arabidopsis, miR156 and miR172 interact to form a gradient that regulates vegetative-to-floral meristem transition through the targeting of squamosa promoter-like genes (*SPL*)[30]. While all copies of both miR172 and miR156 show male-biased expression in catkins, the overall expression of miR156 is greater than miR172 in males. The majority of *SPL* genes that are targeted are female-upregulated in *S. purpurea*, including an *SPL4* homolog on chr07 (Sapur.007G123900) that also shows increased methylation in males in its promoter region. These data support that the miR156/172 pathway is upregulated in male catkins and may be responsible for sex dimorphisms (Table S11). This pathway may play a role in male floral tissue development or differentiation.

Importantly, one copy of miR156 is located in the SDR region unique to chr15Z. Alignment of the chr15Z miR156 precursor sequence to chr15W reveals that a single indel is responsible for this chr15Z-specific mapping (Fig. S5), which may prevent transcription or processing of this small RNA from chr15W. This could indicate a dosage dependent response in males, which have four copies of this mature miR156 homolog, compared with only three in females. Female-upregulated miRNAs included miR403, which targets *AGO2*, and miR162, which targets a dicer-like gene. All of these are involved in small RNA signaling and DNA methylation, which is consistent with the enrichment of transcription regulation terms in the female-upregulated genes. This may point toward a role of genome-wide DNA methylation or RNA silencing in regulating sex dimorphism, possibly mediated by *AGO4* or *DRB1*.

Conclusion

Taken together, RNA-Seq, small RNA-Seq, bisulfite sequencing, and eQTL mapping results suggest that the sex-determination mechanism in willow is different from the small RNA-mediated, single-gene-sex determination pathway proposed in *Populus*, and reveals that many miRNAs, transcription factors, floral development genes, and secondary metabolism genes are involved in primary and secondary sex-dimorphism pathways in *Salix purpurea* catkins. A major factor inhibiting functional genomics research in *Salix* is the lack of a functional transformation system, which precludes direct assessment of gene function in transgenic plants. We show here that using a combination of genome-wide differential expression, coexpression, eQTL, small RNA, and genomic variation can provide strong evidence supporting the involvement of these proposed candidate master regulator genes in sex determination.

Materials and methods

Plant material and growing conditions

The F_1 family (Family 82) was generated from a cross between female *S. purpurea* clone 94006 and male clone 94001, both collected from different naturalized *S. purpurea* populations in Upstate NY (Fig. S6)[40,41]. A female F_1 individual, *S. purpurea* 'Wolcott' (clone 9882-41) and male F_1 individual, *S. purpurea* 'Fish Creek' (clone 9882-34), were crossed to generate the F_2 *S. purpurea* family (Family 317). All progeny individuals and their parents were planted in nursery beds at Cornell AgriTech, Geneva, NY. For additional information on the F_1 and F_2 families and their parents, see Carlson et al. (2019)[42]. Details on phenological stage of catkin collection can be found in the Supplementary Methods.

Library preparation and sequencing

Catkins of 90 males and 90 females from the 317 F_2 family were collected for RNA-Seq, which, after removing individuals with poor mapping quality, was reduced to 77 females and 82 males. Reads were mapped using STAR[43] (Table S12), and counts assigned using featureCounts[44] in R. DESeq2 were used to identify differentially expressed genes[45]. A total of 22 male and 21 female F_2 progeny, along with the male (94001) and female (94006) grandparents, were collected for small RNA sequencing. Mapping and identification of putative miRNAs was done using ShortStacks[46] and differential expression determined using DESeq2[45]. Six male and six female F_2 progeny, along with twelve male and twelve female unrelated *S. purpurea* from a diversity panel, were collected for bisulfite sequencing. Reads were mapped using Bismark[47] and differentially methylated regions (DMR) determined using DMRfinder[48]. GBS data were collected on all 317 family F_2 progeny and used to call variants in TASSEL 5.0[49]. The RNA-Seq mapping and downstream analysis pipelines are summarized in Fig. S7. All read mapping for RNA-Seq, small RNA-Seq, GBS, and bisulfite reads was conducted using the version 5.0 Salix purpurea reference genome available from Phytozome 13. Full details on library prep, sequencing, and alignment to the reference genome can be found in the Supplementary Methods.

eQTL analysis

The R package, MatrixEQTL[50], was used to map eQTL. Covariates were included for sex, along with the four largest principal components in order to control for the underlying data structure and minimize false-positive eQTL mapping to the SDR. Each unique SNP/gene association is considered separately as an eQTL, independent of any additional genes that may be associated with that SNP. *Cis*-associations were classified as those SNPs within 1 Mb of their associated gene based on the recommended settings of MatrixEQTL, while *trans*-associations were classified as greater than 1 Mb or on different chromosomes. SNPs within the SDR, spanning from 2.34 Mb to 9.07 Mb on chr15W and 2.34 Mb to 6.70 Mb on chr15Z[14], were also considered to be *cis* to all other genes within the SDR, regardless of distance, due to low recombination in this region.

WGCNA network analysis

Library-normalized FPKM expression values were calculated from the raw count data using the edgeR package[51]. Weighted gene coexpression network analysis (WGCNA)[52] was used to construct modules of genes with similar expression values. A clustering of FPKM expression values was generated using hclust, which showed that two male samples (10X-317-161 and 10X-317-020) displayed extreme differences in overall expression patterns from all of the other samples (Fig. S8), and so were removed from downstream analyses. A topological overview matrix was created to perform module identification. In order to correlate sex with module expression, sex was coded numerically with values of −1 for females and 1 for males; thus, a module r^2 near one indicates strong correlation with males, and a module r^2 near-negative one indicates strong correlation with females. The minimum module-size threshold was set at 30 genes per module. Identified modules were merged based on similar expression levels. A total of 32 unmerged modules were merged into 24 modules based on similar expression patterns.

Gene ontology enrichment and pathway analysis

Gene ontology (GO) enrichment at the lowest levels of the GO hierarchy was calculated in the BiNGO app[53] in Cytoscape using the Arabidopsis homolog gene models. Arabidopsis gene models were used as there are limited data available on gene function in *Salix*, and gene functional annotations of nonmodel species are often ultimately based on experimental evidence from Arabidopsis. Furthermore, using Arabidopsis gene models provided compatibility with existing databases, such as TAIR. The background reference was created from the list of Arabidopsis homologs for every transcript model in the *S. purpurea* 94006 v5.1 genome found in the gene annotation information file on Phytozome (https://phytozome-next.jgi.doe.gov/info/Spurpurea_v5_1). The Arabidopsis homologs of the genes present in each module or male: female differential expression cutoff (inclusive of all genes above or below the cutoff) were used as the query to determine overrepresented and underrepresented parent and child terms for each gene set. Descriptions of each over- and underrepresented Gene Ontology ID were generated by BinGO based on TAIR descriptions[39].

Acknowledgements

Support for this research was provided by grants (DEB-1542486, DEB-1542599) from the National Science Foundation and from the USDA National Institute for Food and Agriculture (2015-67009-23957). The work conducted by the DOE JGI is supported by the Office of Science of the US Department of Energy under Contract no. DE-AC02-05CH11231. Partial support for BLH was provided by a Cornell AgriTech Extension and Outreach Assistantship. The authors would like to thank Matt Christiansen, Curt Carter, Rebecca Wilk, Lauren Carlson, Jane Petzoldt, Dawn Fishback, and Sam Knopka for their technical assistance with sample collection and field-trial maintenance and Alex Harkess and Xiohan Yang for critical feedback on drafts of the paper.

Author details

[1]Horticulture Section, School of Integrative Plant Science, Cornell University, Cornell AgriTech, Geneva, NY, USA. [2]Floral and Nursery Plants Research Unit, US National Arboretum, United States Department of Agriculture, Agricultural Research Service, Beltsville, MD, USA. [3]United States Department of Energy, Joint Genome Institute, Berkeley, CA, USA. [4]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA. [5]The Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [6]Department of Biology, Texas Tech University, Lubbock, TX, USA. [7]Department of Biology, West Virginia University, Morgantown, WV, USA

Author contributions

B. L. H. and C. H. C. contributed equally to the paper. B. L. H. designed the data analysis workflow, performed data analysis and interpretation and wrote the paper. C. H. C. contributed to experimental design, performed catkin collections, total RNA and small RNA isolations, data collection, and bioinformatics. F. E. G. contributed to experimental design and data collection. G. F. contributed to development of custom mapping genome for RNA-Seq data. L. B. S., S. P. D., M. S. O., and G. A. T. obtained funding and contributed to the experimental design, execution of various stages of the research, and preparation of the paper. S. P. D. and M. S. O. made substantial revisions and comments on late versions of the paper. A. L., A. S., K. B., and L. S. carried out transcriptome sequencing.

Data availability

Raw RNA-Seq, small RNA-Seq, and bisulfite sequencing data are available at the JGI genome portal under proposal 1690, as well as on the NCBI sequence read archive (SRX5027565-SRX5027793, SRX3886725-SRX3886746, SRX3987232-SRX3987253). Code used in the mapping and analysis is available at https://github.com/Willowpedia/Hyden-et-al.-Hort.-Res.

Competing interests

The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41438-021-00606-y.

Received: 21 March 2021 Revised: 23 May 2021 Accepted: 1 June 2021
Published online: 01 August 2021

References

1. Harkess, A. et al. Sex determination by two Y-Linked genes in garden asparagus. *Plant Cell* **32**, 1790–1796 (2020).

2. Akagi, T., Henry, I. M., Tao, R. & Comai, L. A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. *Science* **346**, 646–650 (2014).
3. Akagi, T. et al. Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nat. Plants* **5**, 801–809 (2019).
4. Muller, N. A. et al. A single gene underlies the dynamic evolution of poplar sex determination. *Nat. Plants* **6**, 630–637 (2020).
5. Yang, H. W., Akagi, T., Kawakatsu, T. & Tao, R. Gene networks orchestrated by MeGI: a single-factor mechanism underlying sex determination in persimmon. *Plant J.* **98**, 97–111 (2019).
6. Charlesworth, D., Charlesworth, B. & Marais, G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**, 118–128 (2005).
7. Akagi, T. & Charlesworth, D. Pleiotropic effects of sex-determining genes in the evolution of dioecy in two plant species. *Proc. R. Soc. B Biol. Sci.* **286**, 20191805 (2019).
8. Barrett, S. C. & Hough, J. Sexual dimorphism in flowering plants. *J. Exp. Bot.* **64**, 67–82 (2013).
9. Albert, A. Y. K. & Otto, S. P. Sexual selection can resolve sex-linked sexual antagonism. *Science* **310**, 119–121 (2005).
10. Montalvão, A. P. L., Kersten, B., Fladung, M. & Müller, N. A. The diversity and dynamics of sex determination in dioecious plants. *Front. Plant Sci.* **11**, 580488 (2020).
11. Feng, G. et al. Pathways to sex determination in plants: how many roads lead to Rome? *Curr. Opin. Plant Biol.* **54**, 61–68 (2020).
12. Sanderson, B. J. et al. Sex determination through X-Y heterogamety in *Salix nigra*. *Heredity* **126**, 630–639 (2021).
13. Yang, G. *et al*. Sex-related differences in growth, herbivory, and defense of two *Salix* species. *Forests*, https://doi.org/10.3390/f11040450 (2020).
14. Zhou, R. et al. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biol.* **21**, 38 (2020).
15. Pucholt, P., Ronnberg-Wastljung, A. C. & Berlin, S. Single locus sex determination and female heterogamety in the basket willow (*Salix viminalis* L.). *Heredity* **114**, 575–583 (2015).
16. Liu, J. et al. Transcriptome analysis of the differentially expressed genes in the male and female shrub willows (*Salix suchowensis*). *PLoS ONE* **8**, e60181 (2013).
17. Tuskan, G. A. et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604 (2006).
18. Paolucci, I. et al. Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across Populus species. *Tree Genet. Genomes* **6**, 863–875 (2010).
19. Yang, W. et al. A general model to explain repeated turnovers of sex determination in the Salicaceae. *Molec. Biol. Evol.* **38**, 968–980 (2020).
20. Gammerdinger, W. J. & Kocher, T. D. Unusual diversity of sex chromosomes in african cichlid fishes. *Genes* **9**, 480 (2018).
21. Jeffries, D. L. et al. A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nat. Commun.* **9**, 4088 (2018).
22. Dickmann, D. I. & Kuzovkina, J. Poplars and willows of the world, with emphasis on silviculturally important species. In *Poplars and willows: Trees for society and the environment* **Vol 22** CABI (2014).
23. Gouker, F. E. *et al*. Sexual dimorphism in the dioecious willow *Salix purpurea*. *Amer. J. Bot.* https://doi.org/10.1002/ajb2.1704 (2021).
24. Fussel, U., Dotterl, S., Jurgens, A. & Aas, G. Inter- and intraspecific variation in floral scent in the genus *Salix* and its implication for pollination. *J. Chem. Ecol.* **33**, 749–765 (2007).
25. Ashman, T. L. Sniffing out patterns of sexual dimorphism in floral scent. *Funct. Ecol.* **23**, 852–862 (2009).
26. Melnikova, N. V. et al. Sex-specific polymorphism of MET1 and ARR17 genes in *Populus x sibirica*. *Biochimie* **162**, 26–32 (2019).
27. Cronk, Q. & Müller, N. A. Default sex and single gene sex determination in dioecious plants. *Front. Plant Sci.*, https://doi.org/10.3389/fpls.2020.01162 (2020).
28. Robinson, K. M. et al. *Populus tremula* (European aspen) shows no evidence of sexual dimorphism. *BMC Plant Biol.* **14**, 276 (2014).
29. Carlson, C. H. et al. Dominance and sexual dimorphism pervade the *Salix purpurea* L. transcriptome. *Genome Biol. Evol.* **9**, 2377–2394 (2017).
30. Wu, G. et al. The Sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* **138**, 750–759 (2009).
31. Akagi, T. et al. A Y-Encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. *Plant Cell* **30**, 780–795 (2018).
32. Fedoroff, N. V. Transposable elements, epigenetics, and genome evolution. *Science* **338**, 758–767 (2012).
33. Janoušek, B., Široký, J. & Vyskot, B. Epigenetic control of sexual phenotype in a dioecious plant *Melandrium album*. *Mol. Gen. Genet. MGG* **250**, 483–490 (1996).
34. Eamens, A. L., Smith, N. A., Curtin, S. J., Wang, M.-B. & Waterhouse, P. M. The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA* **15**, 2219–2235 (2009).
35. Ranftl, Q. L., Bastakis, E., Klermund, C. & Schwechheimer, C. LLM-domain containing B-GATA factors control different aspects of cytokinin-regulated development in *Arabidopsis thaliana*. *Plant Physiol.* **170**, 2295–2311 (2016).
36. Ocarez, N. & Mejía, N. Suppression of the D-class MADS-box AGL11 gene triggers seedlessness in fleshy fruits. *Plant Cell Rep.* **35**, 239–254 (2016).
37. Verelst, W., Saedler, H. & Münster, T. MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific *Arabidopsis* Promoters. *Plant Physiol.* **143**, 447–460 (2007).
38. Dreni, L. & Zhang, D. Flower development: the evolutionary history and functions of the AGL6 subfamily MADS-box genes. *J. Exp. Bot.* **67**, 1625–1638 (2016).
39. The Arabidopsis Information Resource (TAIR), https://www.arabidopsis.org/tools/bulk/go/index.jsp, on www.arabidopsis.org, last accessed on 23 May 2021.
40. Lin, J., Gibbs, J. P. & Smart, L. B. Population genetic structure of native versus naturalized sympatric shrub willows (*Salix*; Salicaceae). *Am. J. Bot.* **96**, 771–785 (2009).
41. Gouker, F. E., DiFazio, S. P., Bubner, B., Zander, M. & Smart, L. B. Genetic diversity and population structure of native, naturalized, and cultivated *Salix purpurea*. *Tree Genetics Genomes*, https://doi.org/10.1007/s11295-019-1359-0 (2019).
42. Carlson, C. H. et al. Joint linkage and association mapping of complex traits in shrub willow (*Salix purpurea* L.). *Ann. Bot.* **124**, 701–716 (2019).
43. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
44. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
45. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
46. Axtell, M. J. Classification and comparison of small RNAs from plants. *Annu Rev. Plant Biol.* **64**, 137–159 (2013).
47. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
48. Gaspar, J. M. & Hart, R. P. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinforma.* **18**, 1–8 (2017).
49. Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
50. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
51. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
52. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* **9**, 1–13 (2008).
53. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).
54. Blázquez, M. A. Flower development pathways. *J. Cell Sci.* **113**, 3547–3548, https://doi.org/10.1242/jcs.113.20.3547 (2000).