Functional analysis of *Salix purpurea* genes support roles for *ARR17* and *GATA15* as master regulators of sex determination

Authors: Brennan Hyden[1,2], Dana L. Carper[2], Paul E. Abraham[2], Guoliang Yuan[2], Tao Yao[2], Leo Baumgart[3], Yu Zhang[3], Cindy Chen[3], Ronan O'Malley[3], Jin-Gui Chen[2], Xiaohan Yang[2], Robert L. Hettich[2], Gerald A. Tuskan[2], Lawrence B. Smart[1]*

*Corresponding author email: lbs33@cornell.edu, mailing address: Cornell University, 630 West North Street, Geneva, NY 14456

1. Horticulture Section, School of Integrative Plant Science, Cornell University, Cornell AgriTech, Geneva, NY 14456
2. Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831
3. US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Running head: Functional analysis of willow sex determination genes

Abstract

The Salicaceae family is of growing interest in the study of dioecy in plants because the sex determination region (SDR) has been shown to be highly dynamic, with differing locations and heterogametic systems between species. Without the ability to transform and regenerate *Salix* in tissue culture, previous studies investigating the mechanisms regulating sex in the genus *Salix* have been limited to genome resequencing and differential gene expression, which are mostly descriptive in nature, and functional validation of candidate sex determination genes has not yet been conducted. Here we used Arabidopsis to functionally characterize a suite of previously identified candidate genes involved in sex determination and sex dimorphism in the bioenergy shrub willow *Salix purpurea*. Six candidate master regulator genes for sex determination were heterologously expressed in Arabidopsis, followed by floral proteome analysis. In addition, 11 transcription factors with predicted roles in mediating sex dimorphism downstream of the SDR were tested using DAP-Seq in both male and female *S. purpurea* DNA. The results of this study provide further evidence to support models for the roles of *ARR17* and *GATA15* as master regulator genes of sex determination in *S. purpurea*, contributing to a regulatory system that is notably different from that of its sister genus *Populus*. Evidence was also obtained for the roles of two transcription factors, an *AP2*/*ERF* family gene and a homeodomain-like transcription factor, in downstream regulation of sex dimorphism.

Keywords: Arabidopsis, dioecy, proteomics, Salix, sex determination, transgenics, willow

Introduction

Understanding the genetic regulation of sex determination in dioecious plants is of interest in the plant biology community because, while dioecy is only observed in about six percent of angiosperm species, it is present in numerous taxa and families and is thought to have evolved independently on as many as 5,000 occasions (Käfer et al., 2017). As such, understanding the genetic mechanisms that lead to separation of sexes in different taxa can provide insight into the repeated evolution and maintenance of dioecy. The Salicaceae family is of particular interest in this effort, as nearly all species in the family are dioecious, and it contains two genera of economic importance: poplars (*Populus*) and willows (*Salix*). In particular, *Salix* contains over 300 species, which are native to every continent except Antarctica and Australia, and grow in a diverse range of biomes, including subarctic tundra, deserts and temperate and tropical forests (Argus, 1997; Kuzovkina et al., 2007). *Salix* species also exhibit a variety of growth habits, ranging from prostrate dwarfs to shrubs and trees (Argus, 1997). Despite this remarkable diversity in species range and form, dioecy has been maintained throughout the evolution of most of the family, including all *Populus* and *Salix*. Moreover, the sex determination region (SDR) in *Salix* appears to have shifted among chromosomes and varied heterogametic systems on multiple occasions, with tree willows such as *S. nigra* and *S. chaenomeloides* containing SDR on Chr07 with an XY system (Sanderson et al., 2021; Wang et al., 2022) and alternatively on Chr15 with an XY system in *S. arbutifolia* (Wang et al., 2022). Shrub willows, including *S. purpurea* and *S. viminalis*, contain the SDR on Chr15 under a ZW system (Pucholt et al., 2015; Zhou et al., 2018; Wilkerson et al., 2022). Because of the dynamic nature of sex determination in Salicaceae, there is an opportunity to characterize the precise mechanisms of sex determination in diverse species across this family and to add to our understanding of the evolution, conservation, and transition of the SDR.

Much work has already been done to identify candidate master regulator genes in both *Populus* and *Salix*. Müller et al (2020) demonstrated that a homolog of *ARR17*, an Arabidopsis type-C response regulator, is likely the sole master regulator gene in *Populus*, with females expressing *ARR17* while in males it is either absent or silenced by smRNA produced from exon 1 repeats located on the Y chromosome (Müller et al., 2020). A similar *ARR17* mediated sex determination system is thought to be present in some willows with Chr07 and Chr15 XY systems, but has not been confirmed through expression or functional analysis (Wang et al., 2022). *ARR17* was first proposed as a candidate master regulator gene of sex determination in *S. purpurea* (Chr15 ZW SDR) by Zhou et al. (2020), where the authors identified four inverted repeats of the gene on Chr15W in a female *S. purpurea* (Zhou et al.,

2020). Another study conducted RNA-Seq and smRNA-Seq of an F_2 family of *S. purpurea* and proposed several candidate master regulator genes in addition to *ARR17*, but notably did not find evidence for the smRNA silencing mechanism in males that exists in XY *Populus* (Hyden et al., 2021). This lack of smRNA expression, along with confirmed expression of *ARR17* in both male and female *S. purpurea*, suggests that if *ARR17* is a master regulator gene in *S. purpurea*, it may operate via a mechanism that differs from *Populus* and the tree willows (Hyden et al., 2021). Based on sequencing and expression analysis, several other genes have been proposed as candidate genes for sex determination in addition to *ARR17*, including homologs of *GATA15*, *AGO4*, *DRB1*, and several hypothetical proteins (Hyden et al., 2021). However, no functional validation has been conducted to confirm the role of these proposed sex-determination genes. Recently, sequencing and expression data from a monoecious *S. purpurea* revealed a structural variant on Chr15W that includes deletions of *ARR17*, *AGO4*, and *DRB1*, but not *GATA15* (Hyden et al., 2023). Based on these data, we hypothesize that *GATA15* is a master regulator gene of sex determination that functions to promote female development, while *ARR17* acts as a suppressor of male development.

Previous research on willow sex determination and dimorphism has relied primarily on the comparison of DNA and RNA sequencing data between males and females (Carlson et al., 2017; Zhou et al., 2020; Hyden et al., 2021); however, functional validation is ultimately needed to support these hypotheses. Yet, unlike poplar, there is not a facile protocol for *Salix* transformation and regeneration from tissue culture, so it is not feasible to study gene function in willow using transformation-based gain/loss-of function approaches. Our best alternative is to use a model plant species, such as Arabidopsis, for transgenic manipulation. In addition, transcription factors can also be studied *in vitro* using DAP-Seq as it does not require plant transformation. In DAP-Seq, transcription factors are transiently expressed and incubated with native genomic DNA. Transcription factor bound DNA fragments are then sequenced and aligned to the reference genome (O'Malley et al., 2016). This system enables transcription factor binding analysis in any species from which candidate genes can be cloned and high-quality DNA can be sequenced.

In this study, we sought to elucidate the mechanisms of sex determination and dimorphism in *S. purpurea* by investigating both master regulator genes of sex and transcription factors with predicted roles in sex dimorphism. Six candidate master regulator genes for sex determination in *S. purpurea* were heterologously expressed in Arabidopsis using a constitutive promoter. Bottom-up proteomics was performed on floral tissue of the transgenic Arabidopsis plants to identify proteins regulated by the

overexpressed candidate master regulators of primary sex dimorphism (anther or stamen development). To characterize regulation of sex dimorphism downstream of the SDR, which includes floral development and floral secondary metabolite production, DAP-Seq was performed on 11 transcription factors. Nine of the TFs tested are located on autosomes and have eQTL in the *S. purpurea* SDR and two are located in the SDR and are candidate master regulator genes of sex, for which expression in Arabidopsis and proteomic analysis was also performed (Hyden et al., 2021). Using data from these combined methods, we constructed a conceptual model for the functional role of several genes involved in sex determination and dimorphism and provide evidence to support of *ARR17* and *GATA15* as master regulator genes of sex determination in *S. purpurea*.

Results

Proteomic Analysis

Transgenic Arabidopsis floral proteomics data were used to assess the broad impact of each candidate *S. purpurea* gene on the proteome as well as changes among proteins relevant to floral development and sex determination biological processes. Eight transgenic constructs were generated (six candidate master regulator gene and two methodological control genes), each with 4-6 independent insertion events (Table 1) and proteome data were obtained from three full-sibling T_2 plants from each event in addition to five empty vector control plants, for a total of 122 plants evaluated. Across all 122 samples, a total of 17,191 Arabidopsis protein accessions were identified and the total proteins identified for each transgenic line ranged from 11,395 to 16,885 (Table 1). The relative impact of each heterologous expression transgenic line on proteome expression was compared against the empty vector control proteome data. Across the eight expression lines, the number of proteins of differential abundance (PDA) ranged from 103 to 5,970 (Supplemental Dataset S20). Enriched MapMan functional categories (Schwacke et al., 2019) were identified for all eight heterologously expressed genes to identify biological processes affected by expression of each candidate master regulator (Fig. S1-S2). As expected, the impact of each expressed gene on the floral proteome was not uniform, as shown by the variation in total peptides detected, PDAs, and MapMan enrichment categories. Plants expressing *GATA15* showed the greatest number of PDAs when compared against the empty vector control (Fig. 1, Table 1), followed by hypothetical protein Sapur.15WG074900 and *ARR17* with 5,318 and 4,305 PDAs respectively (Fig. 1, Fig. S3, Table 1). The positive control genes *LEAFY* and *FT* affected only 444 and 179 PDAs, respectively (Fig. S3, Table 1). The CCHC Zinc finger Sapur.15WG068800 produced the fewest number of PDAs, at only 103.

In general, most transgenic proteomes contained a substantial number of downregulated proteins when compared to the empty vector control and this observation was particularly pronounced when expressing *ARR17*, *GATA15*, and Sapur.15WG074900 (Fig. 2). Resulting PDAs across each heterologous expression line were compared to assess similarity in the resulting proteome changes. Among all the PDAs, there was considerable proteome changes observed between *ARR17*, *GATA15*, and Sapur.15WG074900 hypothetical protein lines (Fig. 2). Notably, in addition to the large number of resulting PDAs, GATA15 and ARR17 expression lines exhibited unique protein abundance profiles of multiple floral development genes, indicating a likely role in floral and reproductive development and sex determination.

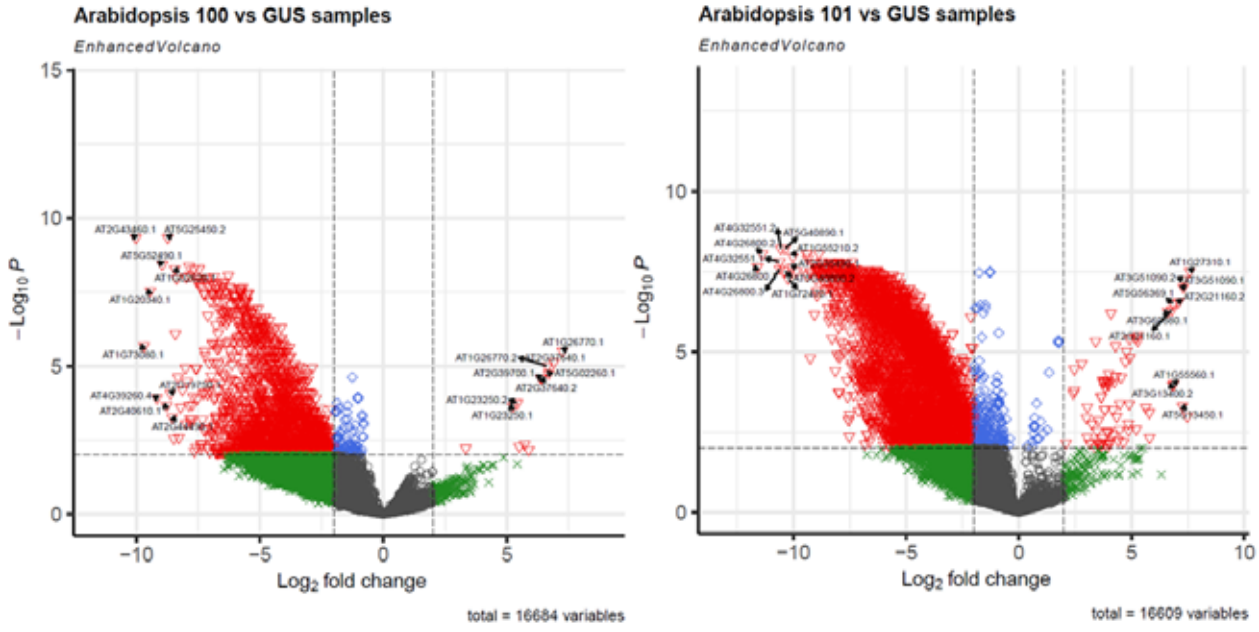


Figure 1. Volcano plots displaying the total proteins of differential abundance (PDA) results for the *ARR17* (100, left) and *GATA15* (101, right) expression lines, relative to the empty vector control. Grey indicates non-significant proteins, green indicates those that meet the log2 FC cutoff but not p-value, blue indicates those that meet the p-value threshold but not log2 FC, and red indicates proteins that have met the log2 FC cutoff and p-value threshold. Identities of the most extreme PDA are indicated.

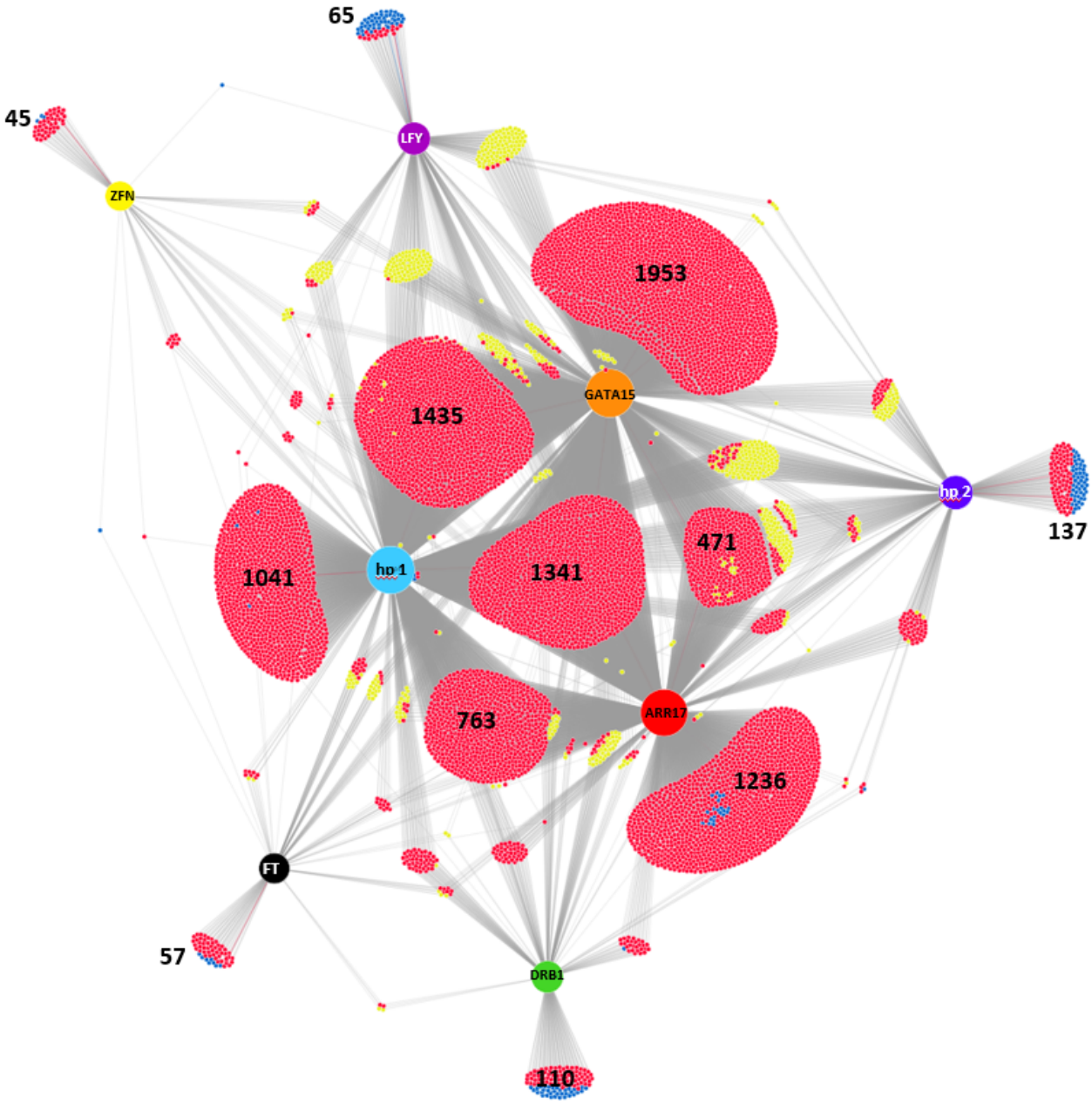


Figure 2. DiVenn diagram comparing proteomic expression patterns in transgenic plants heterologously expressing each of the eight *S. purpurea* genes relative to the empty vector control. Colored nodes reperesent expressed genes (ZFN: Sapur.15WG068800; hp 1: Sapur.15WG074900; hp 2: Sapur.15WG075700). Proteins are grouped according to unique differential abundance or shared differential abundance among the Arabidopsis lines. Relative to the empty vector control, red points represent downregulated PDAs, blue points are upregulated PDAs, and yellow points are PDAs which show differing abundance patterns between expression lines. Groupings containing unique PDAs and the largest groupings of shared PDA are labeled with the respective number of proteins in each group.

158 *Sapur.15WG072500 ARR17 expression results*

159 The *ARR17*-OX lines (pBH100) showed 4,305 PDAs, 1,236 of which had unique abundance patterns when
160 compared to other heterologously expressed candidate genes (Table 1, Fig. 1-2). Enriched MapMan
161 categories included chromatin organization, coenzyme metabolism, transferase and hydrolase enzyme
162 activity, multi-process regulation, protein homeostasis and modification, RNA biosynthesis and
163 processing, and vesicle trafficking (Fig. S1). Among the proteins showing differential abundance unique
164 to the *ARR17*-OX lines were several with annotations related to floral development, including a homolog
165 of *PISTILLATA* and seven genes involved in tapetum and pollen development (Table 2). Among the
166 proteins most regulated in the *ARR17*-OX lines were multiple expansin family proteins, which have been
167 shown to be involved in pollen tube development and cell expansion (Table S1, Fig. 1) (Liu et al., 2021).

168 *Sapur.15WG062800 GATA15 expression results*

169 The *GATA15*-OX lines (pBH101) showed the greatest number of floral PDAs by a substantial margin at
170 5,970, of which 1,953 had unique abundance patterns (Table 1, Fig. 1-2), as well as the greatest number
171 of enriched MapMan functional categories, at 22. Among the enriched MapMan categories were cell
172 cycle organization, transcriptional regulation, RNA modification, and vesicle transport (Fig. S1). *GATA15*
173 was one of two genes tested, the other being *ARR17,* whose expression resulted in a unique differential
174 abundance pattern of proteins with floral development annotations, including the upregulation of
175 *FLOR1*, a gene involved in floral meristem development and transition (Acevedo et al., 2004), and
176 downregulation of two isoforms each of *SHATTERPROOF1* homologs and *AGL72* homologs, involved in
177 fruit dehiscence and floral transition, respectively (Liljegren et al., 2000; Dorca-Fornell et al., 2011)
178 (Table 2). Among the proteins with the greatest abundance in the *GATA15*-OX lines were *SKU5*-Similar
179 13 and *SKU5*-Similar 14, the former of which has been shown to be essential for pollen tube growth
180 through regulation of jasmonic acid biosynthesis (Zhang et al., 2022). However, these proteins do not
181 appear to be uniquely upregulated by *GATA15*, as they were also upregulated in the hypothetical
182 protein Sapur.15WG074900 transgenic lines. Among the most downregulated proteins in *GATA15*-OX
183 lines were two isoforms of *LEUNIG*, which is involved in regulating gynoecium development (Table 2,
184 Table S1, Fig. 1).

185

186

187

Sapur.15WG068800 CCHC Zinc Finger Transcription Factor expression results

The *Sapur.15WG068800-OX* lines (pBH102) had the fewest PDAs relative to the control at only 103, of which 45 had unique protein abundance patterns (Table 1, Fig. S3). Enriched functions included phosphorylation, carrier-mediated transport, and solute channel transport (Fig. S2).

Sapur.15WG074300 DRB1 expression results

The lines expressing *DRB1* (pBH103) had 343 PDAs relative to the empty vector control, of which 110 showed unique abundance patterns (Table 1, Fig. 2, Fig. S3). Enriched functional categories included sucrose metabolism, chromatin structure, hydrolase enzyme activity, and MAP kinase cascade signaling (Fig. S2).

Sapur.15WG74900 hypothetical protein expression results

In addition to *ARR17* and *GATA15*, plants expressing the hypothetical protein Sapur.15WG074900 (pBH104) also showed an exceptionally high number of PDAs at 5,318, of which 1,041 had unique abundance patterns (Table 1, Fig. 2, Fig. S3). Despite this large number of floral PDAs, there were not any protein annotations among these that have been previously related to floral development (two *SKU5*-like proteins were upregulated, but their upregulation was also observed in the *GATA15*-OX lines). Sapur.15WG074900 lines also showed the second greatest number of enriched functional categories, with 20 that were significant (Fig. S2), including vesicle trafficking, RNA splicing, homeostasis, and modification, phosphorylation, ubiquitin-proteasome system, calcium-dependent signaling, fatty acid metabolism, and sucrose metabolism.

Sapur.15WG075700 hypothetical protein expression results

Plants expressing the hypothetical protein Sapur.15WG075700 (pBH106) had 611 PDAs, of which 137 were unique (Table 1, Fig. S3). Microfilament network, fatty acid metabolism, protein quality control, phosphorylation, transcriptional regulation, RNA export, and primary active transport of solutes were all enriched functions in these plants (Fig. S2).

Sapur.15WG122200 LEAFY control expression results

A *S. purpurea* homolog of *LEAFY*, located on Chr15, was expressed as a positive experimental control (pBH107). The *LEAFY*-OX lines had 444 PDAs, of which only 65 were unique (Table 1, Fig. S3). Among the

significantly enriched functional terms were pectin, microtubular network, MAP kinase cascade signaling, and transcriptional regulation (Fig. S2).

Sapur.008G061900 Flowering Locus T control expression results

A *S. purpurea Flowering Locus T* (*FT*) homolog was also included as a positive experimental control (pBH108), and showed 179 PDAs, of which 57 had a unique abundance pattern (Table 1, Fig. S3). The *FT*-OX lines showed enrichment for the fewest functional categories, with only two that were significant: oxidoreductase enzymes and transcriptional regulation (Fig. S2). Arabidopsis T_2 progeny from all five *S. purpurea Flowering Locus T* (*FT*) expression events showed early flowering phenotypes when compared to the empty vector control, producing inflorescences in just 23 days after germination (Fig. S4). These observations are consistent with the Arabidopsis *FT* overexpression phenotype (Kardailsky et al., 1999) and indicated the *S. purpurea FT* homolog was both expressed and functional in Arabidopsis, which is the first such example of functional expression of an *S. purpurea* gene in Arabidopsis. Among the most upregulated proteins in the FT-OX lines were three isoforms of *FASCIATA5*, which is involved in floral initiation and is consistent with the role of *FT* in inducing early flowering (Fig. 1, Table S1) (Albert et al., 2015).

Genome-wide identification of transcription factor binding sites

In DAP-Seq, peak calling is performed by mapping transcription factor bound DNA fragments to the reference genome and comparing the relative read abundance to background levels of mapping, with a three-fold mapping rate relative to background being a standard minimum cutoff for identifying putative transcription factor binding sites, also termed “summits”. Candidate target genes are identified as those nearest to a summit with expression in the direction away from the summit (i.e. antisense if upstream of the summit, positive sense if downstream of the summit). Of the 11 transcription factors tested by DAP-Seq, three (Sapur.001G003600.1 (*AP2/ERF*, GCCGGC binding sequence), Sapur.003G027300.1 (Homeodomain-like, TGGATAA binding sequence), and Sapur.15WG062800.1 (*GATA15*, GATCA binding sequence)) produced large numbers of significant peaks with a threshold of three-fold or greater mapping in the 94006 female and ‘Fish Creek’ male libraries along with similar binding motif predictions in both libraries (Table 3, Fig. S5). In particular, Sapur.001G003600.1 had 12 peaks in 94006 and five peaks in ‘Fish Creek’ with a mapping rate at least 10-fold over background, while Sapur.003G027300.1 had 36 in 94006 and 56 in ‘Fish Creek’. The remaining eight transcription factors that were tested produced inconsistent motif predictions between the two libraries and fewer than 20 significant peaks

in any library (Table 3, Fig. S5, Supplemental Dataset S21). Many of the peaks were shared among these eight transcription factors, suggesting that the results from these latter eight genes are likely the result of background mapping and not true transcription factor binding sites. All three genes that produced prominent binding motifs and more than 20 significant binding sites also targeted multiple floral development genes (Table S2), confirming a likely role in the regulation of primary sex dimorphism and development.

Discussion

Transgenic Arabidopsis and proteomic analysis

In this study we were able to measure and identify over 17,000 total proteins, including 11,000 protein models for each overexpressed gene and a multitude of PDAs (Table 1, Fig. S1-S3). These protein numbers exceed that of recent studies on Arabidopsis floral tissue, which identified between 8,000 and 12,000 proteins (Jing et al., 2020; Lu et al., 2020). This is the first study reporting on *S. purpurea* gene heterologous expression in Arabidopsis, confirming the validity of this system for functional genomics studies in *Salix*, which is especially useful since stable transformants in *S. purpurea* cannot be generated.

Three of the candidate genes used in this study have vague annotations and have not been previously well characterized: Sapur.15WG068800 (pBH102, CCHC Zinc Finger nuclease), Sapur.15WG074900 (pBH104, hypothetical protein), and Sapur.15WG075700 (pBH106, hypothetical protein). The MapMan functional enrichment analysis of PDAs from these genes' expression lines can provide some insight into their potential role. For Sapur.15WG068800, enriched terms included phosphorylation, carrier-mediated transport, and solute transport channels, which suggest a role in regulating transmembrane transport. In the lines with expression of Sapur.15WG074900, the exceptional number of PDAs observed (5,318) is particularly interesting. Sapur.15WG074900 is a hypothetical protein that unique to Chr15W and female *S. purpurea* and shows high levels of RNA expression in female catkins, but lacks a homolog in Arabidopsis, despite lacking an Arabidopsis homolog. The closest homologs of this gene in *P. trichocarpa* and *P. deltoides* and are also uncharacterized (Tuskan et al., 2006; Goodstein et al., 2011; Hyden et al., 2021). Nevertheless, the large number of floral PDAs suggest that this gene likely has conserved patterns of transcriptional activation in Arabidopsis. The MapMan enrichment categories from the proteomic data suggest a potential role in either directly or indirectly regulating RNA or protein modification and stability. Sapur.15WG075700 is another gene annotated as a hypothetical protein and appears to be unique to *Salix*, as there are no homologs in either *Arabidopsis* or *Populus* (Tuskan et al., 2006; Lamesch

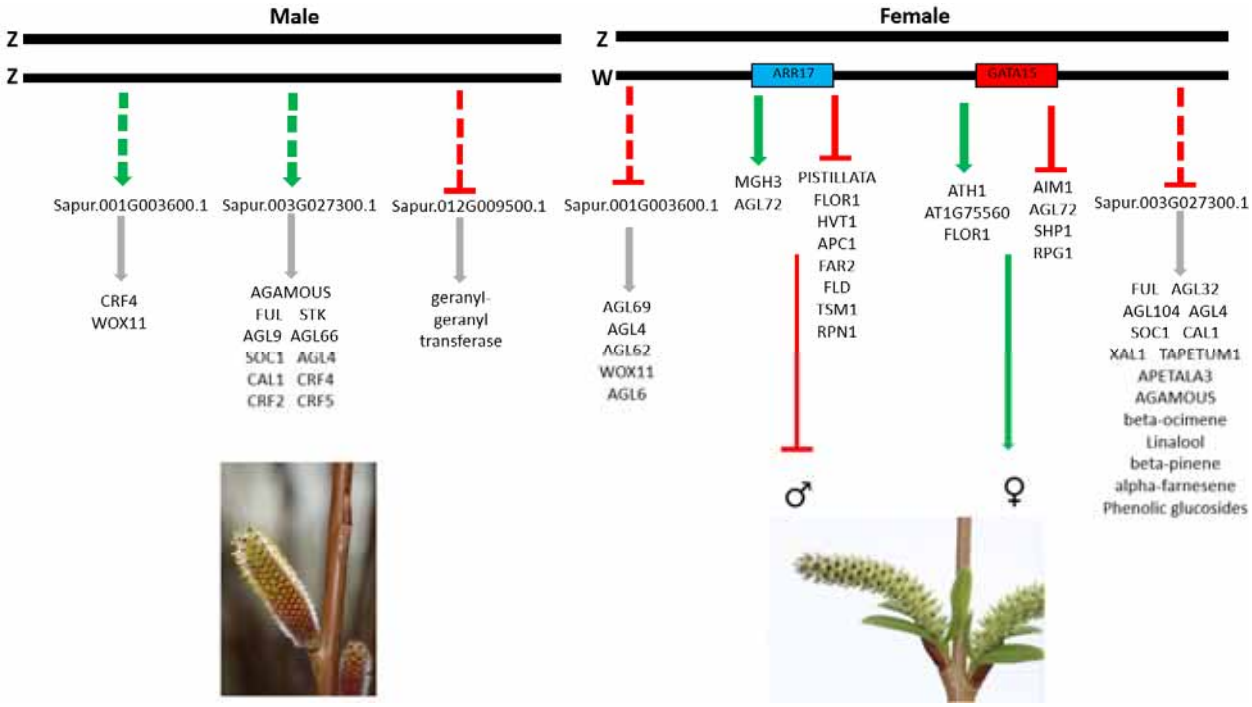
et al., 2012). Among the MapMan enriched terms for Sapur.15WG075700 PDAs were microfilament network, primary active transport, and fatty acid metabolism, which together could suggest a role in intracellular transport.

The *ARR17*-OX and *GATA15*-OX lines stood out in this study as having an exceptional number of PDAs when compared with most of the other lines, as well as unique differential expression of multiple proteins with floral development annotations (Tables 1-2, Fig. 2). The downregulation of *PISTILLATA* in the *ARR17* expression lines is particularly interesting. *PISTILLATA* is a well-characterized B-class MADS box gene that is necessary for stamen development (Krizek and Meyerowitz, 1996), which in *S. purpurea* has also been confirmed to have exceptionally high expression in males (Hyden et al., 2021). *ARR17* is hypothesized to act as a switch from male to female development in *Populus* species through the downregulation of *PISTILLATA* expression (Cronk and Müller, 2020). The results from this study support a similar mechanism in *S. purpurea*. Downregulation of *PISTILLATA*, and multiple tapetum and pollen development genes identified exclusively in the *ARR17*-OX lines, along with the differential expression of over 4,300 floral proteins, provides the first functional evidence supporting a role of *ARR17* as a master regulator of sex determination in *S. purpurea*, functioning in the suppression of male floral development (Fig. 3).

GATA15 is a proposed master regulator gene of sex determination in *S. purpurea* that shows female-specific expression in mature catkins as well as differential expression in females in early floral shoot development (Carlson et al., 2017; Hyden et al., 2021). It is also the only candidate sex determination gene that is still present on Chr15W in monoecious willows, as other putative master regulator genes were deleted. These monoecious genotypes contain a Chr15W with structural variation and produce both male and female flowers. The presence of *GATA15* and *ARR17* has led us to hypothesize that the role of *GATA15* is to promote female development (Hyden et al., 2023). *GATA15* homologs have been shown to have a role in floral development in Arabidopsis and *Lagerstroemia speciosa* (Ranftl et al., 2016; Hu et al., 2019). In the present study, expression of *S. purpurea GATA15* produced the greatest number of floral PDAs, including several with annotations related to floral development and transition, consistent with previous data that show *GATA15* is expressed early in the transition from vegetative to floral meristem identity when catkin development is determined (Zhang and Fernando, 2005; Carlson et al., 2017). Moreover, among the most downregulated genes in the *GATA15* lines were two isoforms of *LEUNIG*, which is involved in gynoecium development and whose knockout in Arabidopsis has been shown to convert sepals to carpels, reduce stamen number, and alter expression on *PISTILATTA,*

306 *AGAMOUS*, *AP3*, and *AP1* MADS-box genes (Liu and Meyerowitz, 1995; Lamesch et al., 2012). These
307 proteomic results further support the hypothesized role of *GATA15* as a master regulator of sex
308 determination in *S. purpurea* with an involvement in female floral development.

309 Taken together, the floral proteomic data from this study indicate that of the genes tested *GATA15* and
310 *ARR17* are the most likely to be master regulators of sex determination in *S. purpurea*, with *ARR17* likely
311 suppressing male flower development and *GATA15* promoting female flower development (Fig. 3). Such
312 a system is consistent with the two-gene model of sex determination, which is common among
313 angiosperms (Charlesworth, 2002) and has been identified in other species, including garden asparagus
314 (Harkess et al., 2020) and kiwifruit (Akagi et al., 2019).



315

316 **Figure 3.** Model for regulation of sex determination in *S. purpurea.* Males represent the default sex
317 when the W chromosome is absent. Genes predicted to be up (green arrow) and down (red line)
318 regulated by the putative master regulators *ARR17* and *GATA15* are shown, which result in simultaneous
319 suppression of male floral development and promotion of female floral development, respectively.
320 Predicted targets of each DAP-Seq transcription factor in males and females that have likely involvement
321 in sex dimorphism are listed, although the exact regulation of expression of each DAP-Seq gene is
322 unknown.

323

324

325

DAP-Seq

Of the three transcription factors (TFs) that produced consistent binding motifs and large number of significant peaks, two (Sapur.001G003600.1 and Sapur.003G027300.1) have eQTL that map to the SDR and are predicted to be genes in the sex dimorphism pathway that are regulated either directly or indirectly by the master regulator genes of sex (Hyden et al., 2021), while Sapur.15WG062800.1 (*GATA15*) is a candidate master regulator gene. Among the genes targeted by the *AP2/ERF* TF Sapur.001G003600.1 were a *Wuschel*-like *WOX11* in both libraries, a cytokinin response factor 4 in the male 'Fish Creek', and four MADS-box genes: *AGL4*, *AGL6*, *AGL62*, and *AGL69*, in the female 94006. The *AGL4* target is particularly interesting, as it is involved in ovule development (Rounsley et al., 1995), and showed a 4.96-fold increase in TF binding in the 94006 library. This preponderance of MADS-box gene binding exclusively in the female library suggests a potential role of this TF in promoting female floral development.

Sapur.003G027300.1 exhibited binding near both floral development genes and secondary metabolism genes. Among the floral development genes with adjacent TF binding in both libraries were homologs of two *AGAMOUS*-like genes (involved stamen and ovule identity) (Mizukami and Ma, 1992), *AGL32* (ovule endothelial identity) (De Folter et al., 2006), and *AGL4* (ovule development) (Rounsley et al., 1995), pointing towards involvement of this gene in floral development in both sexes. In the 94006 library, Sapur.003G027300.1 binding sites were identified observed near *TAPETUM1*, *TPD1*, and *AP3* (Krizek and Meyerowitz, 1996; Yang et al., 2003; Lamesch et al., 2012), all of which are directly involved in stamen development and may suggest a role of this gene in downregulating male floral development in females. Furthermore, Sapur.003G027300.1 displayed binding near multiple cytokinin response factor genes exclusively in the male 'Fish Creek' library, indicating a role in regulating cytokinin in males. This is particularly interesting considering the proposed role of *ARR17*, a cytokinin response regulator, in sex determination in *S. purpurea* (Zhou et al., 2020; Hyden et al., 2021). Indeed, given that eQTL for this gene map to the sex determination region (Hyden et al., 2021), it is possible that *ARR17* may directly or indirectly regulate expression of Sapur.003G027300.1, although the precise mechanism for this remains unclear. Sapur.003G027300.1 also binds near a multitude of genes involved in terpenoid, phenolic glucoside, and flavonoid production in the 94006 library, including genes specifically annotated as being involved in production of beta-ocimene, beta-pinene, limonene, and alpha-farnesene. These aforementioned compounds are terpenoids that are differentially produced in male and female *S. purpurea* catkins and are associated with pollinator and pest attraction (Keefover-Ring et al., 2022). TF

binding activity near genes responsible for production of these metabolites in 94006 but not 'Fish Creek' suggests that, under the presence of the Chr15W and the sex determination genes, the CpG methylation profile is altered, which in turn affects TF binding and regulation of these metabolites in females resulting in differential expression and sex dimorphism.

The DAP-Seq assay of Sapur.15WG062800 *GATA15*, a candidate master regulator gene, showed peaks associated with *ATH1* and a CCHC Zinc finger in 94006. *ATH1* is activated by the C class MADS box gene *AGAMOUS*, and in turn regulates GA synthesis (Gómez-Mena et al., 2005). The CCHC Zinc finger targeted by *GATA15* is also described as having a likely role in reproductive development in Arabidopsis (Lamesch et al., 2012). This binding activity of Sapur.15WG062800 was only observed in the female 94006 library. These data are consistent with previous research hypothesizing *GATA15* as a female-specific master regulator gene of sex in *S. purpurea* with a role in promoting female floral development (Hyden et al., 2023).

The targeting of different gene families involved in sex dimorphism between male and female libraries was consistently observed across three transcription factors tested with DAP-Seq and further supports a role for CpG methylation, which is nearly three-fold higher in male *S. purpurea* catkins compared to females, in sex dimorphism (Hyden et al., 2021).

Comparison of Arabidopsis heterologous expression and DAP-Seq results

Among the *Salix* genes tested in this study, the CCHC zinc finger Sapur.15WG098800 gene and *GATA15* Sapur.15WG062800 gene were both analyzed through heterologous expression in Arabidopsis and in DAP-Seq assays, due to their annotation as transcription factors and hypothesized role as master regulators of sex. In both cases, the DAP-Seq and proteomics results provided complementary data supporting or rejecting a role in sex determination. For the CCHC zinc finger Sapur.15WG068800 gene, the fewest PDAs were observed (103) out of any of the transgenic Arabidopsis lines, and the DAP-Seq results indicated only seven and one significant peaks in the female and male libraries, respectively. Moreover, none of these PDAs or genes adjacent to DAP-Seq peaks had annotations that would suggest a role in regulating sex determination. From these data, the precise function of Sapur.15WG06800 remains inconclusive, but a role in regulating sex determination seems unlikely. For *GATA15* on the other hand, both the DAP-Seq and the Arabidopsis expression data supported a potential role in sex determination as a promoter of female development, with the greatest number of PDAs in the

Arabidopsis heterologous expression lines, multiple floral development proteins with unique abundance patterns, and DAP-Seq peaks near two well-characterized floral development genes.

In summary, results from this study support the role of *ARR17* and *GATA15* as master regulator genes of sex determination in *S. purpurea*. *ARR17* appears to suppress expression of *PISTILLATA* and tapetum development genes, implicating a role as a male suppressor gene, while *GATA15* appears to promote female floral development by regulating expression of floral transition and ovule development genes, including *LEUNIG*. This system is clearly distinct from the single-gene system in *Populus* and underscores the dynamic nature of the sex determination system in the Salicaceae family.

Materials and Methods

Generation and evaluation of Arabidopsis heterologous expression lines

Salix purpurea genes used for heterologous expression in Arabidopsis were obtained from the list of eight candidate sex-determination genes described in a previous study (Hyden et al., 2021) including Arabidopsis Response Regulator 17 (*ARR17*) (Sapur.15WG073500), Double-stranded RNA-Binding 1 (*DRB1*) (Sapur.15WG074300), a CCHC znc finger nuclease (Sapur.15WG068800), *GATA15* (Sapur.15WG062800), and two genes annotated as hypothetical proteins (Sapur.15WG074900, Sapur.15WG075700). One gene, Sapur.15WG074400, a homolog of *AGO4*, did not contain either a start codon or a canonical stop codon, and was therefore dropped from further consideration. Attempts to clone Sapur15WG075300, annotated as a hypothetical protein, did not produce any colonies containing the transgene in *E. coli*, suggesting that it may result in a toxic product, and therefore this gene was also dropped from further consideration. *Salix purpurea* homologs of *Flowering Locus T* (*FT*, Sapur.008G061900) and *LEAFY* (Sapur.15WG122200) were included as positive controls to test the effectiveness of the transformation methods and validity of the results, since their role and function are well characterized in Arabidopsis. All coding sequences (CDS) were obtained from the *S. purpurea* female 94006 v5.1 reference (Zhou et al., 2020) available on Phytozome (Goodstein et al., 2011). CDSs were synthesized as gblocks by Integrated DNA Technologies (Coralville, IA, USA) and contained 30 bp and 26 bp overlap sequences homologous to the pGFPGUSPlus vector (Vickers et al., 2007) on the 3' and 5' ends, respectively, along with a six His tail immediately prior to the stop codon. NEBuilder HiFi DNA assembly (New England Biolabs, Ipswich, MA, USA) was used to assemble each gblock into the pGFPGUSPlus vector, replacing the GFP CDS adjacent to a 35S promoter (Supplemental Datasets S1-S8). Each vector contained plant selectable markers for GUS and hygromycin, also driven by 35S promoters,

as well as a bacterial selectable marker for kanamycin resistance. Assembled constructs were used to transform chemically competent TOP10 *E. coli* cells obtained from ThermoFisher (Waltham, MA, USA) following the manufacturer's protocol. Successful insertion of each gblock in the correct orientation was confirmed by restriction digest and PCR amplification using Q5 polymerase (New England Biolabs, Ipswich, MA, USA) followed by Sanger sequencing of plasmid DNA in the Cornell University Institute for Biotechnology (Ithaca, NY, USA). Plasmid DNA was extracted from *E. coli* cells using a miniprep kit from Qiagen (Germantown, MD, USA) following the recommended protocol. *Agrobacterium tumefaciens* GV3101 cells were transformed using a standard electroporation protocol, and insertion of each plasmid and gene of interest were confirmed with PCR using Q5 polymerase followed by Sanger sequencing. Columbia-0 ecotype Arabidopsis was transformed by floral dip following the protocol described by (Zhang et al., 2006). T_1 generation Arabidopsis seeds were grown on standard MS media containing hygromycin for selection of transformants at 30 µg L^{-1} and surviving seedlings were transferred to potting mix. The presence of each transgene was confirmed using PCR from genomic DNA followed by Sanger sequencing of PCR products. Four to six T_1 plants, each representing a unique transgene insertion event, were self-pollinated to generate T_2 seeds, which were grown on MS media containing 30 µg L^{-1} hygromycin before being transferred to potting mix. All Arabidopsis were grown at 21°C under fluorescent lighting with an 8/16 hour photoperiod prior to bolting. Upon initiation of floral shoots, day length was switched to 16 hours. Plants were sub-irrigated regularly whenever the potting mix became dry. Expression of the plasmid in floral tissue of T_2 plants was confirmed by GUS staining assay following the manufacturer's protocol (Millipore Sigma, Burlington, MA, USA) (Figure S6). Floral buds from three T_2 plants (biological replicates) with confirmed gene insertion events for each event were harvested prior to anthesis, flash frozen in liquid nitrogen, and stored at -80°C.

Protein Extraction and Proteome Analysis

Three floral buds collected prior to anthesis from each from three biological replicates (full sibling T_2 plants) for each transgenic insertion event were selected for proteomics. Floral buds were ground in liquid nitrogen using 2.3mm zirconia/silica beads with a Geno/Grinder 2010 (SPEX) at a rate of 1200rpm for 1 minute. Ground tissue was resuspended in 200µL of lysis buffer (4% sodium dodecyl sulfate, 10mM dithiothreitol) and incubated for 10 minutes at 90°C with constant shaking. Proteins were alkylated with 30mM iodoacetamide and incubated in the dark for 15 minutes to prevent the reformation of disulfide bonds. For each sample, all of the crude protein extract was transferred to a fresh tube, Sera-Mag beads were added (100µg), and proteins were extracted by protein aggregation capture (Batth et al., 2019).

Precipitated protein was resuspended in 100mM ammonium bicarbonate (ABC) and then digested with two separate and sequential aliquots of sequencing grade trypsin (Promega) in the ratio of 1:75 trypsin to sample protein ratio overnight followed by a 3-hr digestion. Peptide mixtures were adjusted to 0.5% formic acid (FA) and physically separated from the Sera-Mag beads with an AcroPrep Advance 96-well 10KDa omega filter plate (Pall Corporation) by centrifuging at 1500xg for 30 minutes. Peptides were freeze dried (Labconco FreeZone 72040) and then resuspended in an aqueous solvent (0.1% formic acid, 5% ACN). Peptide concentrations were estimated using a Nanodrop One spectrophotometer. For each sample, 2µg aliquots were measured by one-dimensional liquid chromatography tandem mass spectrometry (1D-LC-MS/MS) using a RSLCnano UHPLC system (Thermo Scientific) coupled to a Q Exactive Plus mass spectrometer (Thermo Scientific). Peptide mixtures were first injected across an in-house built strong cation exchange (SCX) Luna trap column (5 µm, 150 µm X 50mm; Phenomenex, USA) followed by a nanoEase symmetry reverse phase (RP) C18 trap column (5 µm, 300 µm X 50mm; Waters, USA) and then washed with the aqueous solvent. A 1M ammonium acetate inject was used to elute peptides to the C18 trap column, which was then switched to be in-line with an in-house pulled nanospray emitter analytical column (75 µm X 350 mm) packed with Kinetex RP C18 resin (1.7 µm; Phenomenex, USA). Peptides were separated over a 160-minute linear gradient from 2 to 25% of mobile phase (0.1% FA, 80% ACN) at a flow rate of 250nL/min and analyzed using a Top10 data dependent acquisition strategy (Villalobos Solis et al., 2019). All MS data were acquired with Thermo Xcalibur (version 4.2.47) and analyzed using the Proteome Discoverer software (Thermo-Fisher Scientific, version 2.5) (Orsburn, 2021). Each MS raw data file was processed by the SEQUEST HT database search algorithm (Eng et al., 1994) and confidence in peptide-to-spectrum (PSM) matching was evaluated by Percolator (Käll et al., 2007). The TAIR 11 reference genome (Lamesch et al., 2012) was used for mapping proteins. Peptide and PSMs were considered identified at $q < 0.01$ and proteins were required to have at least one unique peptide sequence. Proteins with at least one unique peptide were exported from Proteome Discoverer. Log2-transformation of protein abundances was performed followed by local regression (LOESS) normalization and mean-centering across the entire dataset in R using scripts from the InfernoRDN software (v1.1.7995) (Larsson, 2014). The abundance values for proteins with missing values were imputed with random values drawn from the normal distribution (width 0.3, downshift 2.2) using R. Proteins of differential abundance (PDAs) were calculated using the “limma” package in R (Ritchie et al., 2015) by comparing all events and biological replicates for each overexpressed gene against the pGFPGUSPlus empty vector controls. Volcano plots were generated using EnhancedVolcano (v.1.13.2) (Blighe et al., 2018) with a fold change cutoff of 2 and p-value

threshold of 0.01. Mapman functional categories (Schwacke et al., 2019) were assigned to each FASTA sequence using the Mercator4 online submission tool (https://www.plabipd.de/portal/web/guest/mercator4). Functional enrichment for MapMan categories was performed for each gene expression relative to the empty vector control using clusterProfiler (v.4.2.2) (Wu et al., 2021). Briefly, each protein was assigned to a MapMan category, significantly differentially enriched proteins (|logFC|>2 and p.val<0.05) were used as the differentially enriched genes, while all genes in each comparison were used as background. Each differentially abundant protein set was searched for genes with annotations that are well characterized in floral development.

Transcription factor cloning for DAP-Seq

Genes for DAP-Seq analysis were selected from among 97 transcription factors (TFs) identified as having a potential role as top level regulator genes of sex determination from an eQTL study (Hyden et al., 2021). These genes were prioritized based on transcription factor family, likelihood of success in a DAP-Seq assay, functional annotation, and floral differential gene expression (O'Malley et al., 2016; Hyden et al., 2021). The 11 genes with the highest prioritization score were advanced for DAP-Seq analysis (Table 2). TF CDS regions were successfully cloned via RT-PCR from catkin RNA obtained from the *S. purpurea* 317 F_2 family (Hyden et al., 2021), or were generated using gblocks from IDT based on the female 94006 v5.1 reference genome (Zhou et al., 2020). Each of the 11 genes were cloned into a pENTR-DTOPO vector and finally into a pIX-HALO expression vector using a Gateway cloning kit obtained from NEB and following the manufacturer's protocol (Supplemental Datasets S9-S19). PCR followed by Sanger sequencing of plasmid DNA was used to confirm the presence and correct sequence and orientation of each CDS sequence in the entry and destination vectors. Genomic DNA for the DAP-Seq assay was extracted from catkins of female (clone 94006) and male (clone 'Fish Creek') *S. purpurea* using a modified Qiagen plant mini kit protocol.

DAP-Seq experiments

DAP-Seq experiments were conducted as described previously in (O'Malley et al., 2016), with minor modifications, described in (Baumgart et al., 2021). DNA libraries were prepared by fragmenting genomic DNA of either *S. purpurea* 'Fish Creek' or *S. purpurea* 94006 using a Covaris LE220-Plus focused-ultrasonicator (Covaris), followed by library preparation with the KAPA HyperPrep kit (Roche) following the manufacturer's recommendations. Insert sizes were targeted to an average of 150 bp. Before use in the DAP-seq assay, libraries were PCR amplified for 10 cycles.

For *in vitro* protein expression, linear fragments were first PCR amplified from each pIX-HALO plasmid using primers targeting the upstream T7 promoter (5' GTGAATTGTAATACGACTCACTATAGGG 3') and downstream of the poly-A stretch (5' CAAGGGGTTATGCTAGTTATTGCTC 3'). The correct size of each PCR product was verified using a Tapestation (Agilent Technologies), and PCR products were purified using SPRI beads. Transcription factors were expressed using at least 2000 ng PCR product per sample with the TnT T7 Quick for PCR DNA in vitro protein expression kit (Promega). All reaction volumes were doubled to yield a total of 100 µL protein product per transcription factor. Each DAP-Seq reaction was run with 100 µL expressed protein, 150 ng of the previously prepared fragment library, and 15 µg salmon sperm DNA to reduce non-specific binding. The final DAP-seq libraries were pooled for sequencing on a NovaSeq using the S4 flowcell (Illumina), targeting 30 million 2x150 reads per sample. Primary data analyses included quality filtering, alignment to the reference genome, peak-calling, and gene assignment as described in (Baumgart et al., 2021). Binding motifs for transcription factors were predicted using MEME version 5.3.0 (Bailey and Elkan, 1994).

Tables

Table 1. Total, differentially abundant proteins (PDA), and unique PDA determined for each transgenic heterologous expression line in Arabidopsis.

Construct	Expressed Gene	Gene Annotation	Events	Total Proteins	PDA (FDR < 0.05)	Unique PDA
pBH100	Sapur.15WG073500	*ARR17* response regulator	6	16,684	4305	1236
pBH101	Sapur.15WG062800	*GATA15* transcription factor	4	16,609	5970	1953
pBH102	Sapur.15WG068800	CCHC Zinc Finger	4	16,023	103	45
pBH103	Sapur.15WG074300	*DRB1* dsRNA binding	6	12,306	343	110
pBH104	Sapur.15WG074900	Hypothetical Protein	5	16,885	5318	1041
pBH106	Sapur.15WG075700	Hypothetical Protein	4	11,355	611	137
pBH107	Sapur.15WG122200	*LEAFY*	5	12,271	444	65
pBH108	Sapur.008G061900	*Flowering Locus T*	5	11,928	179	57

528 **Table 2.** Floral and reproductive development genes showing differential expression unique to either
529 *ARR17* or *GATA15* expression lines. log_2FC, log_2 fold-change; FDR, false discovery rate

OX Gene	Arabidopsis Gene	log2FC	FDR	Annotation
	AT1G19890.1	5.42	2.03E-02	male-gamete-specific histone H3, *MALE-GAMETE-SPECIFIC HISTONE H3*
	AT5G51860.2	1.80	4.54E-02	*AGAMOUS*-like 72
	AT3G12145.1	-1.02	4.12E-02	*FLOR1, FLORAL TRANSITION AT THE MERISTEM*
	AT2G30800.2	-1.36	4.24E-02	helicase in vascular tissue and tapetum
	AT5G05560.1	-1.87	4.52E-02	pollen calcium-binding protein 1, *EMBRYO DEFECTIVE 2771*, anaphase promoting complex 1
ARR17	AT5G05560.3	-2.01	2.57E-02	pollen calcium-binding protein 1, *EMBRYO DEFECTIVE 2771*, anaphase promoting complex 1
	AT3G11980.1	-2.75	2.98E-02	*FATTY ACID REDUCTASE 2, MALE STERILITY 2*
	AT5G20240.2	-3.02	7.83E-03	*PISTILLATA*
	AT3G10390.3	-3.03	4.59E-03	*FLOWERING LOCUS D*, Reduced Systemic immunity 1
	AT1G67990.1	-3.28	3.33E-02	*TAPETUM-SPECIFIC METHYLTRANSFERASE 1*
	AT3G10390.2	-3.37	2.55E-04	*FLOWERING LOCUS D*, Reduced Systemic immunity 1
	AT3G10390.1	-3.54	1.56E-04	*FLOWERING LOCUS D*, Reduced Systemic immunity 1
	AT3G10390.4	-3.54	3.72E-04	*FLOWERING LOCUS D*, Reduced Systemic immunity 1
	AT1G25260.1	-6.50	1.43E-08	*REDUCED POLLEN NUMBER 1, REDUCED POLLEN NUMBER*
	AT3G12145.1	0.79	4.07E-04	*FLOR1, FLORAL TRANSITION AT THE MERISTEM*
	AT4G29010.1	-1.31	4.93E-02	*ABNORMAL INFLORESCENCE MERISTEM*
	AT3G58780.3	-1.37	1.76E-02	*SHATTERPROOF 1, AGAMOUS*-like 1
GATA15	AT3G58780.2	-1.42	2.64E-02	*SHATTERPROOF 1, AGAMOUS*-like 1
	AT5G51860.1	-2.43	4.76E-03	*AGAMOUS*-like 72
	AT5G51860.2	-2.45	1.37E-03	*AGAMOUS*-like 72
	AT5G40260.1	-6.42	1.27E-07	*RUPTURED POLLEN GRAIN1*
	AT4G32551.1	-10.60	1.55E-08	*ROTUNDA2, LEUNIG*
	AT4G32551.2	-10-55	6.10E-09	*ROTUNDA2, LEUNIG*

530

531

532 **Table 3.** Analyzed transcription factors and libraries with total number of significant summits and target
533 genes from DAP-Seq analysis.

Library	Salix Gene ID	Description	TF Source	Summit fold-change > 10	Summit fold-change > 5	Summit fold-change > 3	Total Target Genes
94006	Sapur.001G003600.1	*AP2/ERF* transcription factor	RT-PCR 10X-317-124 (M)	12	351	1082	1293
	Sapur.003G027300.1	Homeodomain-like protein	RT-PCR 11X-317-194 (F)	36	1264	5919	9377
	Sapur.003G155500.1	*scarecrow*-like 3	gblock	0	0	1	7
	Sapur.004G110200.3	transcription factor *VRN1*	gblock	0	0	7	21
	Sapur.005G077400.1	NAC transcription factor 030	gblock	0	0	17	81
	Sapur.006G140600.1	*CONSTANS*-like 5	RT-PCR 11X-317-194 (F)	0	0	1	10
	Sapur.007G074000.1	WRKY transcription factor, putative	gblock	0	0	7	18
	Sapur.012G009500.1	*scarecrow*-like 18	RT-PCR 11X-317-118 (F)	0	2	12	48
	Sapur.017G014200.1	GRAS family transcription factor	gblock	0	1	7	21
	Sapur.15WG062800.1	*GATA* transcription factor 15	gblock	0	3	85	150
	Sapur.15WG068800.1	CCHC Zinc Finger	RT-PCR 11X-317-194 (F)	0	0	7	28
'Fish Creek'	Sapur.001G003600.1	*AP2/ERF* transcription factor	RT-PCR 10X-317-124 (M)	5	233	541	357
	Sapur.003G027300.1	Homeodomain-like protein	RT-PCR 11X-317-194 (F)	56	1952	9148	8336
	Sapur.003G155500.1	*scarecrow*-like 3	gblock	0	0	3	4
	Sapur.004G110200.3	transcription factor *VRN1*	gblock	0	0	4	3
	Sapur.005G077400.1	NAC transcription factor 030	gblock	0	0	5	7
	Sapur.006G140600.1	*CONSTANS*-like 5	RT-PCR 11X-317-194 (F)	0	0	0	0
	Sapur.007G074000.1	WRKY transcription factor, putative	gblock	0	0	7	4
	Sapur.012G009500.1	*scarecrow*-like 18	RT-PCR 11X-317-118 (F)	0	1	23	9
	Sapur.017G014200.1	GRAS family transcription factor	gblock	0	0	5	5
	Sapur.15WG062800.1	*GATA* transcription factor 15	gblock	0	8	62	28
	Sapur.15WG068800.1	CCHC Zinc Finger	RT-PCR 11X-317-194 (F)	0	0	1	1

534

535

Acknowledgements

The authors are grateful for excellent technical support provided by Michael Quade, McKenzie Schessl, and Alexander Wares. We appreciate insightful comments on the manuscript from Steven DiFazio, Jeffrey Doyle, and Jocelyn Rose.

Author Contributions

B.L.H., J.G.C., X.Y., R.L.H., G.A.T., R.O., and L.B.S. designed the research; B.L.H., D.L.C., P.E.A., G.Y., T.Y., L.B., Y.Z., C.C., and R.O. performed the experiments, B.L.H. wrote the paper with contributions from all authors.

Funding

This manuscript is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-SC0014664. This study was also supported by The Center for Bioenergy Innovation, a US Department of Energy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The work was also supported by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract Number DE- AC05-00OR22725. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. This work was partially funded by a graduate fellowship from USDA NIFA AFRI (Award #2021-67034-35116) and a grant from the National Science Foundation (DEB-1542486).

Conflict of interest statement: The authors declare no competing interests.

Data Availability

All proteomics spectral data in this study were deposited at the ProteomeXchange Consortium via the MASSIVE repository (https://massive.ucsd.edu/). The data can be reviewed under the username "reviewer_MSV000091180" and password "BHArabidopsis"

Supplemental Figures

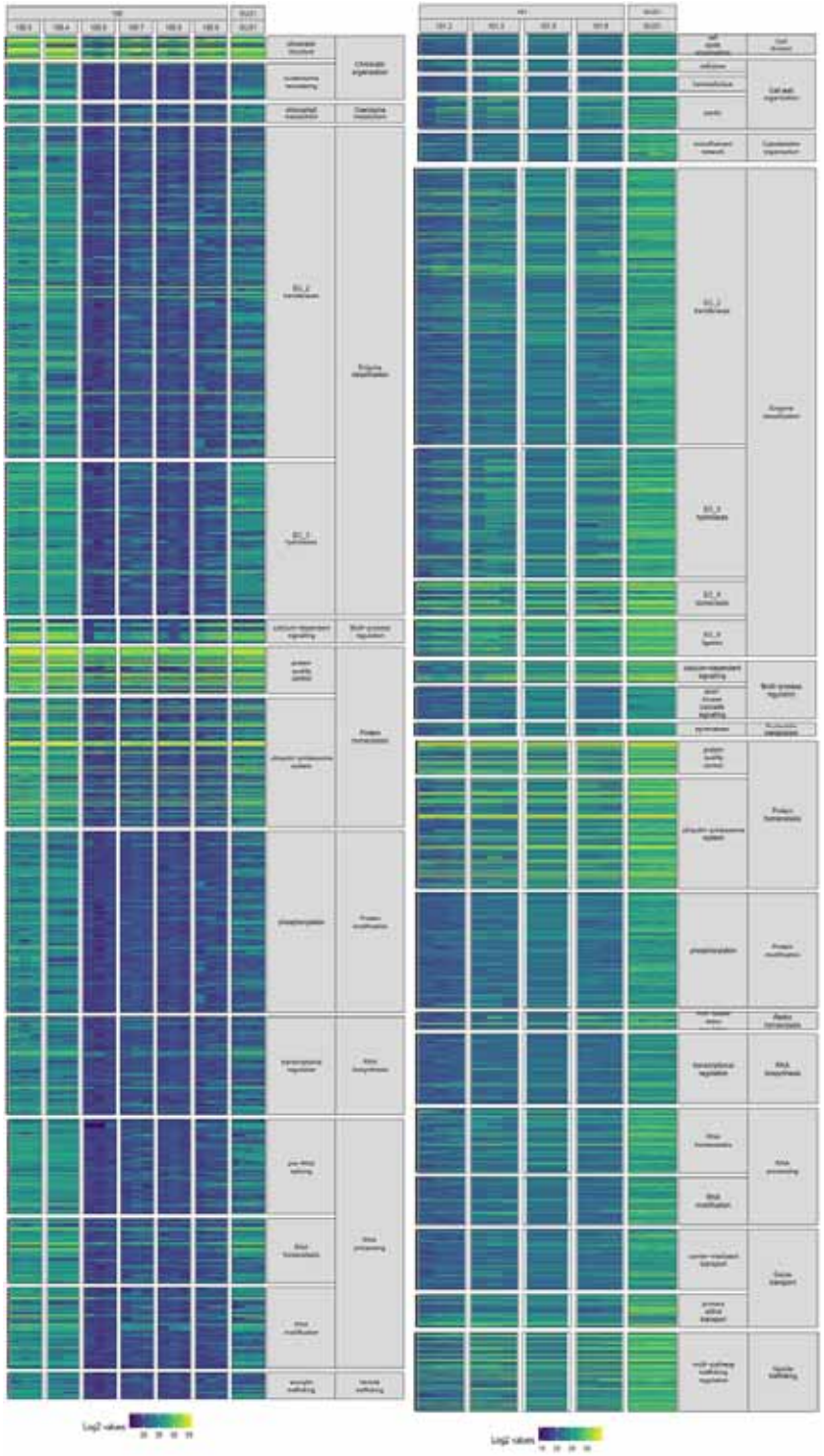


Figure S1. Heatmap displaying the total expression of significantly differentially abundant proteins in the MapMan enriched functional categories for the *ARR17* (100) and *GATA15* (101) expression lines, compared to the pGFPGUSPlus empty vector control (GUS1). Each column represents data from a unique transgene event.

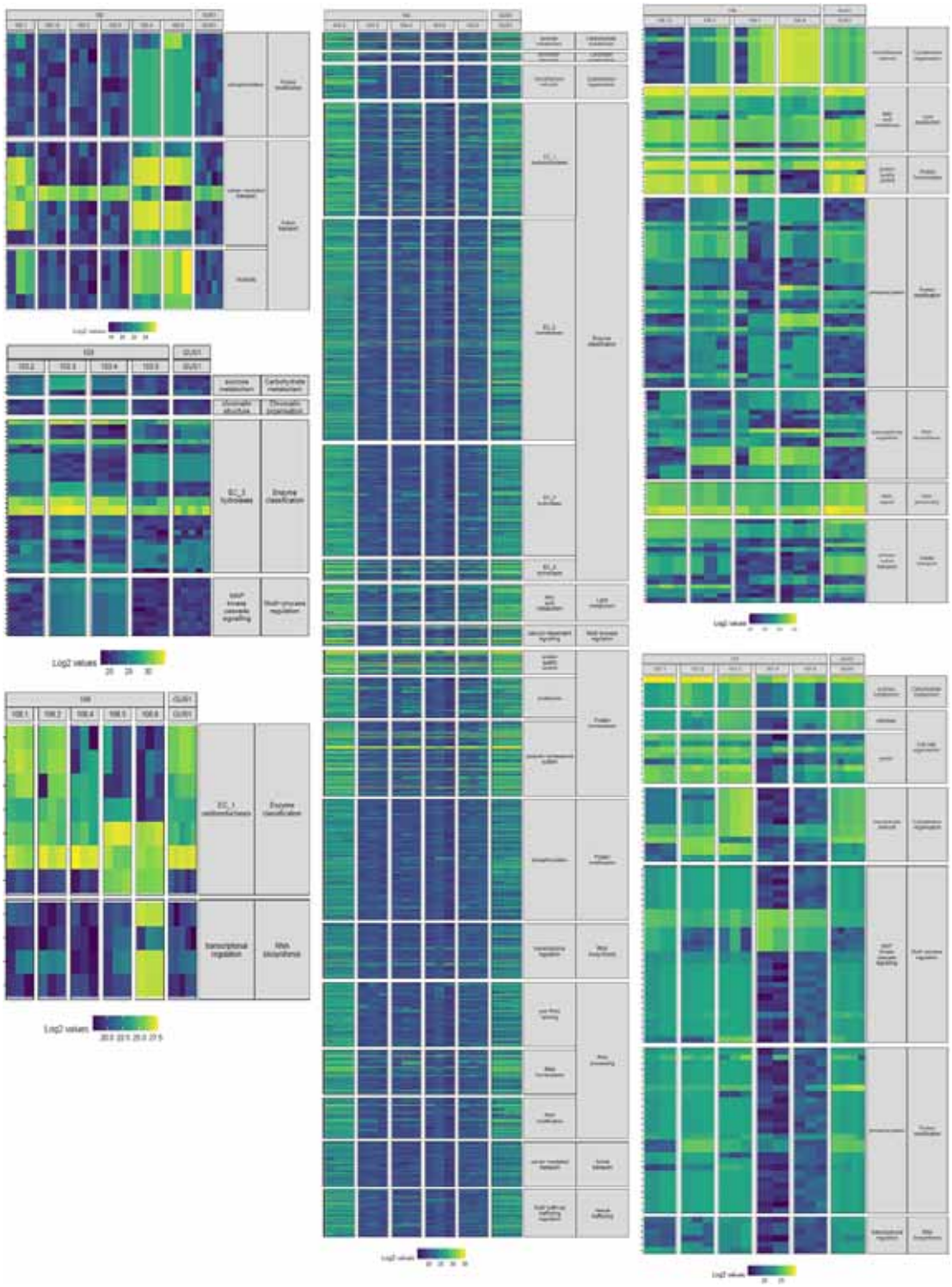


Figure S2. Heatmap displaying the total expression of each gene in the MapMan enriched functional categories for the CCHC Zinc Finger (102) *DRB1* (103) Sapur.15WG074900 hypothetical protein (104), Sapur.15WG075700 hypothetical protein (106), *LEAFY* (107), and *FT* (108) expression lines, compared to the pGFPGUSPlus empty vector control (GUS1). Each column represents data from a unique transgene event.

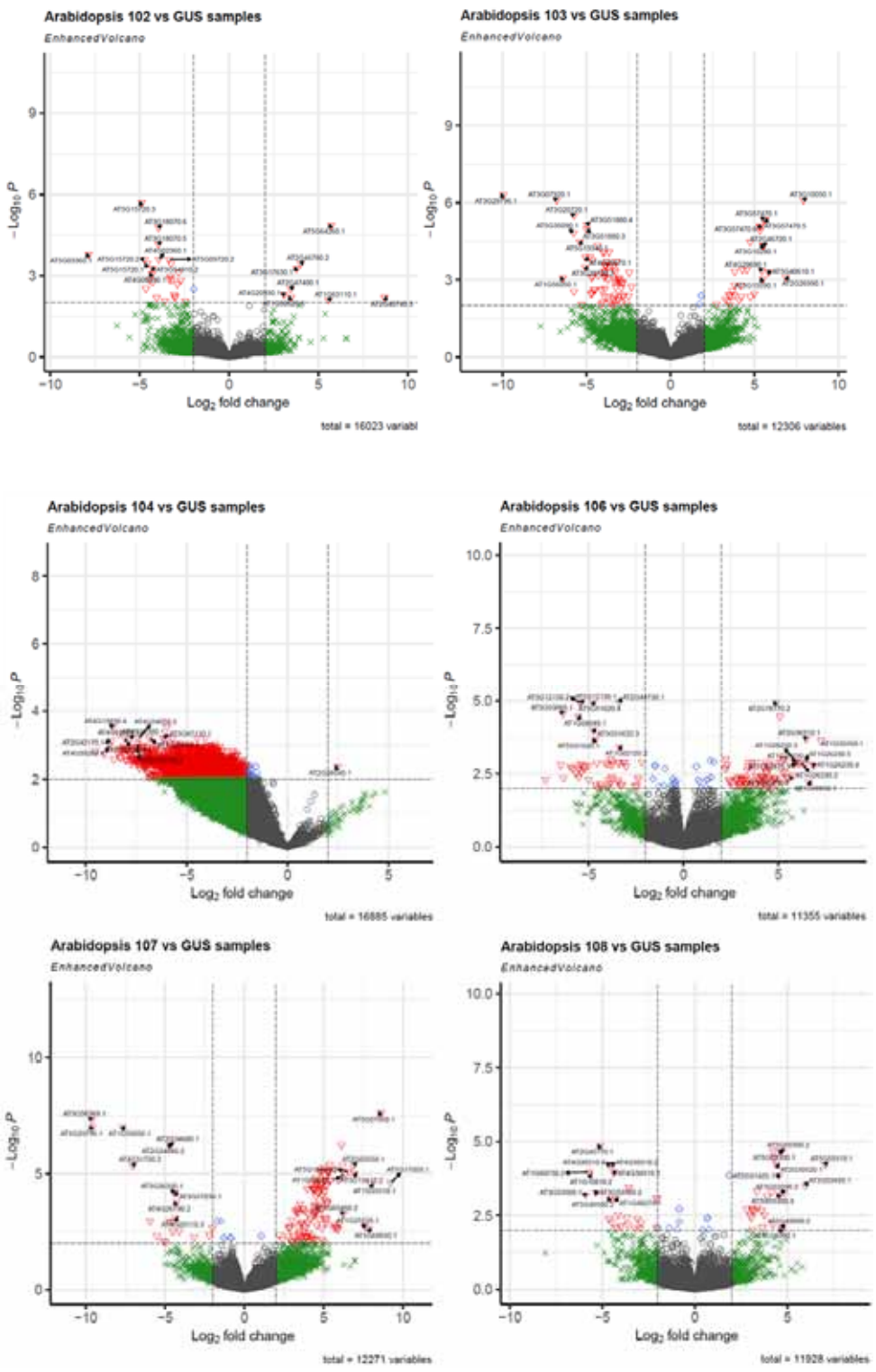


Figure S3. Volcano plots displaying the total differential abundant protein results and top ten significant up and down regulated proteins for the Sapur.15WG068800 CCHC Zinc Finger nuclease (102), Sapur.15WG074300 *DRB1* (103), Sapur.15WG074900 hypothetical protein (104), Sapur.15WG075700 hypothetical protein (106), *LEAFY* (107), and *FT* (108) expression lines, relative to the empty vector control.

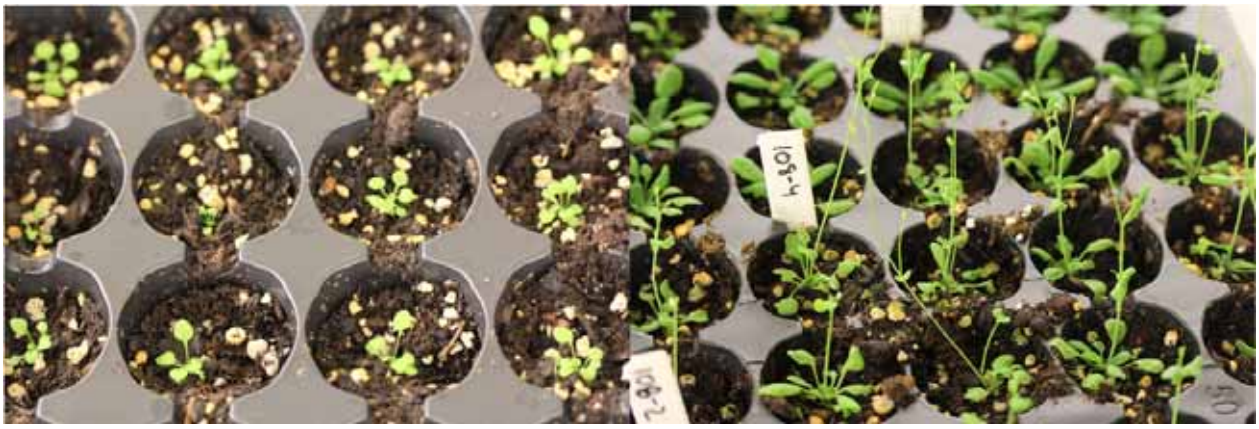


Figure S4. Comparison of Arabidopsis transformed with the empty vector control (left) and overexpressing *S. purpurea FT* (Sapur.008G061900, right)

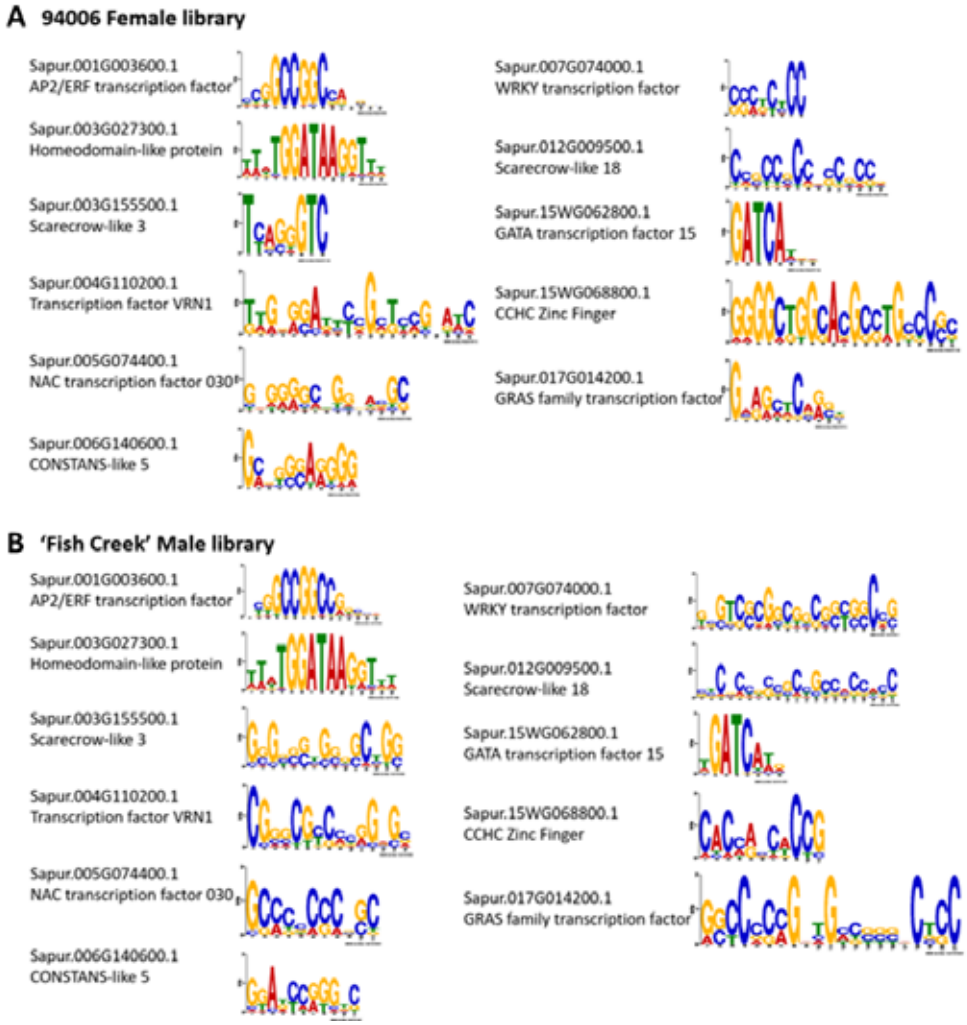


Figure S5. Predicted binding motifs for each transcription factor tested in DAP-Seq. A. Predictions from the 94006 library (female); B. Predictions from the 'Fish Creek' library (male).

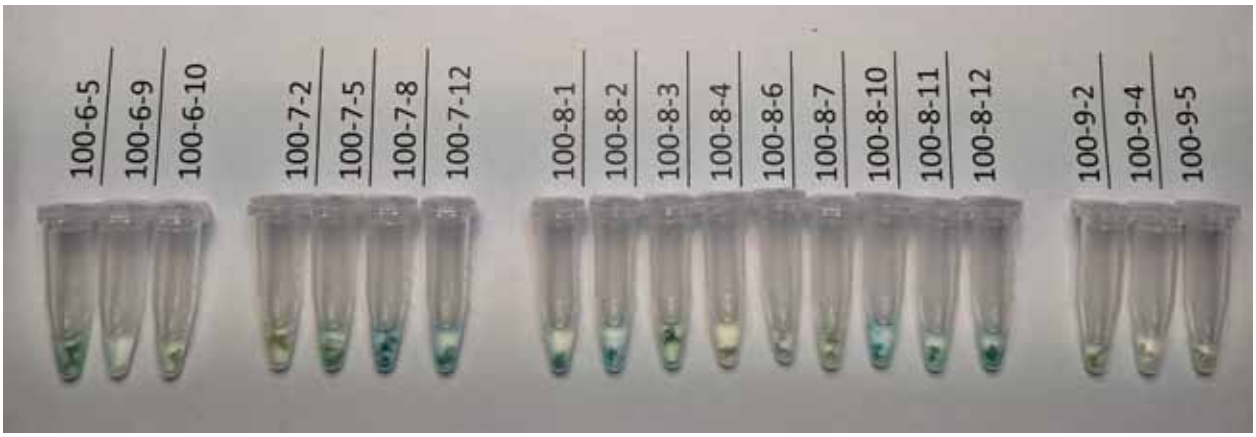


Figure S6. GUS staining assay results for representative samples of T_2 flowers after ethanol staining

Supplemental Tables

Table S1. Top 10 greatest up- and down-regulated proteins for each Arabidopsis expression line.

OX Gene	Regulation	Arabidopsis Gene	log2FC	FDR	Description
ARR17	Upregulated	AT1G26770.1	7.25	3.11E-06	EXPANSIN A10
ARR17	Upregulated	AT2G37640.1	6.90	7.36E-06	EXPANSIN 3
ARR17	Upregulated	AT5G02260.1	6.69	1.77E-05	EXPANSIN A9
ARR17	Upregulated	AT1G26770.2	6.64	9.90E-06	EXPANSIN A10
ARR17	Upregulated	AT2G37640.2	6.41	2.50E-05	EXPANSIN 3
ARR17	Upregulated	AT2G39700.1	6.36	2.50E-05	EXPANSIN A4
ARR17	Upregulated	AT1G09560.1	5.89	6.77E-03	PLASMODESMAL GERMIN-LIKE PROTEIN 1
ARR17	Upregulated	AT1G13370.1	5.72	4.35E-03	Histone superfamily protein
ARR17	Upregulated	AT5G48940.1	5.50	5.20E-03	RGF1 INSENSITIVE 2
ARR17	Upregulated	AT1G23250.2	5.42	1.77E-04	Caleosin-related family protein
ARR17	Downregulated	AT2G43460.1	-10.02	4.62E-10	Ribosomal protein EL38Z
ARR17	Downregulated	AT1G73080.1	-9.69	2.05E-06	PEP1 receptor 1
ARR17	Downregulated	AT1G20340.1	-9.43	2.87E-08	PLASTOCYANIN 2
ARR17	Downregulated	AT4G39260.4	-9.16	1.18E-04	cold, circadian rhythm, and RNA binding 1
ARR17	Downregulated	AT5G52490.1	-8.95	3.66E-09	Fibrillarin family protein
ARR17	Downregulated	AT2G40610.1	-8.78	2.14E-04	Expansin A8
ARR17	Downregulated	AT5G25450.2	-8.73	4.62E-10	Cytochrome bd ubiquinol oxidase
ARR17	Downregulated	AT2G19750.1	-8.62	8.35E-05	Ribosomal protein ES30Z
ARR17	Downregulated	AT2G44490.1	-8.55	5.42E-04	PENETRATION 2
ARR17	Downregulated	AT4G39200.1	-8.48	2.79E-03	Ribosomal protein ES25W
GATA15	Upregulated	AT1G27310.1	7.58	3.07E-08	nuclear transport factor 2A
GATA15	Upregulated	AT5G13450.2	7.47	1.03E-03	delta subunit of Mt ATP synthase
GATA15	Upregulated	AT3G51090.2	7.31	5.76E-08	coiled-coil 90B-like protein
GATA15	Upregulated	AT3G51090.1	7.30	9.81E-08	coiled-coil 90B-like protein
GATA15	Upregulated	AT5G13450.1	7.26	4.93E-04	delta subunit of Mt ATP synthase
GATA15	Upregulated	AT2G21160.2	7.09	2.74E-07	Translocon-associated protein
GATA15	Upregulated	AT1G55560.1	6.87	9.82E-05	SKU5-Similar 14
GATA15	Upregulated	AT5G56369.1	6.85	3.25E-07	no full name available
GATA15	Upregulated	AT3G13400.2	6.83	1.08E-04	SKU5-Similar 13

GATA15	Upregulated	AT2G21160.1	6.79	4.93E-07	Translocon-associated protein
GATA15	Downregulated	AT4G26800.1	-11.59	2.27E-08	Pentatricopeptide repeat protein
GATA15	Downregulated	AT4G26800.2	-11.37	8.80E-09	Pentatricopeptide repeat protein
GATA15	Downregulated	AT4G26800.3	-10.60	2.60E-08	Pentatricopeptide repeat protein
GATA15	Downregulated	AT4G32551.1	-10.60	1.55E-08	ROTUNDA 2, LEUNIG
GATA15	Downregulated	AT4G32551.2	-10.55	6.10E-09	ROTUNDA 2, LEUNIG
GATA15	Downregulated	AT5G40890.2	-10.40	3.07E-08	CHLORIDE CHANNEL-A
GATA15	Downregulated	AT5G40890.1	-10.36	6.10E-09	CHLORIDE CHANNEL-A
GATA15	Downregulated	AT1G72470.1	-10.24	5.23E-08	exocyst subunit exo70 family protein D1
GATA15	Downregulated	AT2G35490.1	-10.06	2.40E-08	FIBRILLIN2
GATA15	Downregulated	AT1G55210.2	-10.02	1.08E-08	Disease resistance-responsive protein
Sapur.15WG068800	Upregulated	AT2G40740.3	0.01	3.32E+00	WRKY DNA-binding protein 55
Sapur.15WG068800	Upregulated	AT5G64260.1	0.00	1.11E+01	EXORDIUM like 2
Sapur.15WG068800	Upregulated	AT1G63110.1	0.01	3.07E+00	no full name available
Sapur.15WG068800	Upregulated	AT4G29260.1	0.03	1.40E+00	Vegetative Storage Protein 3
Sapur.15WG068800	Upregulated	AT2G46780.2	0.00	6.94E+00	RNA-binding family protein
Sapur.15WG068800	Upregulated	AT5G06260.3	0.01	2.41E+00	oxidation resistance 3
Sapur.15WG068800	Upregulated	AT3G17630.1	0.00	6.30E+00	cation/H+ exchanger 19
Sapur.15WG068800	Upregulated	AT1G12800.1	0.03	9.93E-01	S1 domain-containing RBP
Sapur.15WG068800	Upregulated	AT1G30620.6	0.01	3.39E+00	HIGH SUGAR RESPONSE8
Sapur.15WG068800	Upregulated	AT2G47400.1	0.00	4.24E+00	CP12 domain-containing protein 1
Sapur.15WG068800	Downregulated	AT5G03360.1	-7.85	1.69E-04	no full name available
Sapur.15WG068800	Downregulated	AT4G19850.1	-5.25	2.75E-02	phloem protein 2-A2
Sapur.15WG068800	Downregulated	AT5G15720.3	-4.88	2.05E-06	GDSL-motif lipase 7
Sapur.15WG068800	Downregulated	AT1G59520.4	-4.84	2.24E-02	CW7
Sapur.15WG068800	Downregulated	AT4G26180.1	-4.82	4.81E-02	CoA Carrier 2
Sapur.15WG068800	Downregulated	AT5G15720.2	-4.73	2.51E-04	GDSL-motif lipase 7
Sapur.15WG068800	Downregulated	AT1G14560.2	-4.71	4.54E-02	CoA Carrier 1
Sapur.15WG068800	Downregulated	AT3G11650.1	-4.68	3.02E-02	NDR1/HIN1-like 2
Sapur.15WG068800	Downregulated	AT2G43420.1	-4.65	2.98E-03	reticulon 20
Sapur.15WG068800	Downregulated	AT4G26180.2	-4.64	4.39E-02	CoA Carrier 2
DRB1	Upregulated	AT3G10050.1	7.92	7.91E-07	L-O-methylthreonine resistant 1
DRB1	Upregulated	AT2G26990.1	6.87	8.02E-04	CONSTITUTIVE PHOTOMORPHOGENIC 12

DRB1	Upregulated	AT5G40610.1	5.83	5.13E-04	Glycerol-3-phosphate dehydrogenase plastidic
DRB1	Upregulated	AT3G57470.5	5.65	4.57E-06	Insulinase family protein
DRB1	Upregulated	AT3G57470.1	5.56	4.44E-06	Insulinase family protein
DRB1	Upregulated	AT2G46720.1	5.54	4.52E-05	3-ketoacyl-CoA synthase 13
DRB1	Upregulated	AT3G10280.1	5.50	4.65E-05	3-ketoacyl-CoA synthase 14
DRB1	Upregulated	AT3G15590.1	5.50	9.39E-04	DWEORG1
DRB1	Upregulated	AT4G29690.1	5.44	4.41E-04	no full name available
DRB1	Upregulated	AT3G57470.6	5.35	8.23E-06	Alkaline-phosphatase-like family protein
DRB1	Downregulated	AT3G29796.1	-9.95	4.93E-07	hypothetical protein
DRB1	Downregulated	AT3G07920.1	-6.78	7.91E-07	Translation initiation factor IF2/IF5
DRB1	Downregulated	AT1G56050.1	-6.40	8.57E-04	GTP-binding like protein
DRB1	Downregulated	AT1G79490.1	-5.93	1.51E-02	embryo defective 2217
DRB1	Downregulated	AT3G09220.1	-5.87	4.37E-02	laccase 7
DRB1	Downregulated	AT1G35310.1	-5.83	1.47E-02	MLP-like protein 168
DRB1	Downregulated	AT4G35560.2	-5.82	4.67E-02	DUO1-activated WD40 1, Tomosyn-like
DRB1	Downregulated	AT5G35090.1	-5.77	1.49E-05	hypothetical protein
DRB1	Downregulated	AT3G28840.1	-5.75	2.81E-03	hypothetical protein DUF1216
DRB1	Downregulated	AT3G20720.1	-5.75	3.13E-06	amino terminal region of chorein
Sapur.15WG074900	Upregulated	AT5G56369.1	4.83	2.34E-02	defensin-like family protein
Sapur.15WG074900	Upregulated	AT3G13400.2	4.34	2.60E-02	SKU5-Similar 13
Sapur.15WG074900	Upregulated	AT1G55560.1	4.18	3.14E-02	SKU5-Similar 14
Sapur.15WG074900	Upregulated	AT2G01818.2	3.85	2.99E-02	PLATZ transcription factor
Sapur.15WG074900	Upregulated	AT2G01818.1	3.80	2.68E-02	PLATZ transcription factor
Sapur.15WG074900	Upregulated	AT3G60880.1	3.78	3.97E-02	dihydrodipicolinate synthase 1
Sapur.15WG074900	Upregulated	AT4G00550.2	3.72	3.68E-02	digalactosyl diacylglycerol deficient 2
Sapur.15WG074900	Upregulated	AT3G14530.1	3.68	4.25E-02	geranylgeranyl pyrophosphate synthase 6
Sapur.15WG074900	Upregulated	AT3G45230.1	3.61	4.74E-02	ARABINOXYLAN PECTIN ARABINOGALACTAN PROTEIN 1
Sapur.15WG074900	Upregulated	AT2G35780.1	3.19	4.83E-02	serine carboxypeptidase-like 26
Sapur.15WG074900	Downregulated	AT4G05180.1	-9.53	1.52E-03	photosystem II subunit Q-2
Sapur.15WG074900	Downregulated	AT1G72470.1	-9.07	1.81E-03	exocyst subunit exo70 family protein D1
Sapur.15WG074900	Downregulated	AT4G39260.4	-8.81	9.99E-04	cold, circadian rhythm, and RNA binding 1
Sapur.15WG074900	Downregulated	AT2G42170.1	-8.73	7.21E-04	Actin family protein
Sapur.15WG074900	Downregulated	AT4G13850.4	-8.66	2.94E-04	glycine-rich RNA-binding protein 2

Sapur.15WG07490	Downregulate	AT4G21280.			
0	d	2	-8.60	2.56E-03	photosystem II subunit QA
Sapur.15WG07490	Downregulate	AT1G20440.			
0	d	1	-8.38	1.83E-03	cold-regulated 47
Sapur.15WG07490	Downregulate	AT4G12600.			
0	d	1	-8.37	4.71E-04	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein
Sapur.15WG07490	Downregulate	AT4G35560.			
0	d	1	-8.36	1.40E-03	DUO1-activated WD40 1, Tomosyn-like
Sapur.15WG07490	Downregulate	AT1G19870.			
0	d	2	-8.30	2.32E-03	IQ-domain 32
Sapur.15WG07570		AT1G35550.			
0	Upregulated	1	7.24	2.32E-04	elongation factor Tu
Sapur.15WG07570		AT1G26230.			
0	Upregulated	4	6.82	1.60E-03	chaperonin-60beta4
Sapur.15WG07570		AT1G68650.			
0	Upregulated	1	6.60	6.20E-03	photosynthesis-affected mutant 71 like 5
Sapur.15WG07570		AT2G36910.			
0	Upregulated	1	6.48	2.05E-04	ATP-binding cassette B1
Sapur.15WG07570		AT1G26230.			
0	Upregulated	5	6.45	1.35E-03	chaperonin-60beta4
Sapur.15WG07570		AT1G26230.			
0	Upregulated	2	6.18	1.35E-03	chaperonin-60beta4
Sapur.15WG07570		AT1G25520.			
0	Upregulated	1	6.12	1.07E-02	photosynthesis-affected mutant 71 like 4
Sapur.15WG07570		AT1G62570.			
0	Upregulated	1	6.04	1.35E-03	flavin-monooxygenase glucosinolate S-oxygenase 4
Sapur.15WG07570		AT1G26230.			
0	Upregulated	3	5.98	1.24E-03	chaperonin-60beta4
Sapur.15WG07570		AT1G26230.			
0	Upregulated	1	5.89	1.24E-03	chaperonin-60beta4
Sapur.15WG07570	Downregulate	AT2G47960.			
0	d	1	-7.28	4.95E-03	TRAPPC13
Sapur.15WG07570	Downregulate	AT5G54730.			
0	d	2	-6.47	1.59E-03	homolog of yeast autophagy 18 (ATG18) F
Sapur.15WG07570	Downregulate	AT5G53060.			
0	d	1	-6.34	2.78E-05	Enhanced Stress Response 1
Sapur.15WG07570	Downregulate	ATCG00650.			
0	d	1	-6.32	4.50E-03	ribosomal protein S18
Sapur.15WG07570	Downregulate	AT1G65470.			
0	d	1	-5.95	1.81E-03	FASCIATA 1
Sapur.15WG07570	Downregulate	AT1G65470.			
0	d	2	-5.84	2.10E-03	FASCIATA 1
Sapur.15WG07570	Downregulate	AT1G27045.			
0	d	5	-5.67	3.47E-02	homeobox protein 54
Sapur.15WG07570	Downregulate	AT1G27045.			
0	d	4	-5.64	3.29E-02	homeobox protein 54
Sapur.15WG07570	Downregulate	AT1G74330.			
0	d	2	-5.63	2.88E-03	Protein kinase superfamily protein
Sapur.15WG07570	Downregulate	AT1G74330.			
0	d	3	-5.54	4.25E-03	Protein kinase superfamily protein
		AT5G17050.			
LEAFY	Upregulated	1	9.23	2.51E-05	UDP-glucosyl transferase 78D2
		AT5G07600.			
LEAFY	Upregulated	1	8.61	2.45E-08	Oleosin family protein
		AT1G35310.			
LEAFY	Upregulated	1	7.96	2.87E-05	MLP-like protein 168
		AT1G68650.			
LEAFY	Upregulated	1	7.83	2.47E-03	photosynthesis-affected mutant 71 like 5
		AT1G25520.			
LEAFY	Upregulated	1	7.61	1.93E-03	photosynthesis-affected mutant 71 like 4
		AT5G13410.			
LEAFY	Upregulated	2	6.94	1.03E-05	FKBP-like peptidyl-prolyl cis-trans isomerase family protein
		AT2G03550.			
LEAFY	Upregulated	1	6.87	4.45E-06	alpha/beta-Hydrolases superfamily protein
		AT5G13410.			
LEAFY	Upregulated	3	6.57	8.27E-06	FKBP-like peptidyl-prolyl cis-trans isomerase family protein

LEAFY	Upregulated	AT1G78670.1	6.28	1.03E-05	gamma-glutamyl hydrolase 3
LEAFY	Upregulated	AT5G65460.2	6.25	5.41E-04	kinesin like protein for actin based chloroplast movement 2
LEAFY	Downregulated	AT5G56369.1	-9.62	4.78E-08	defensin-like family protein
LEAFY	Downregulated	AT3G29796.1	-9.60	1.01E-07	hypothetical protein
LEAFY	Downregulated	AT1G56050.1	-7.69	1.01E-07	ENGD-2
LEAFY	Downregulated	AT4G31720.2	-7.03	4.45E-06	SALT TOLERANCE DURING GERMINATION 1
LEAFY	Downregulated	AT4G22150.1	-5.95	1.14E-03	plant UBX domain-containing protein 3
LEAFY	Downregulated	AT4G10090.1	-5.50	4.76E-03	elongator protein 6
LEAFY	Downregulated	AT4G04210.1	-5.09	7.80E-03	plant UBX domain containing protein 4
LEAFY	Downregulated	AT1G65650.1	-4.96	8.92E-03	ENGD-2
LEAFY	Downregulated	AT5G53060.1	-4.86	2.62E-02	Enhanced Stress Response 1
LEAFY	Downregulated	AT4G03110.2	-4.73	1.14E-03	Bruno-like 1, RNA-binding protein-defense related 1
FT	Upregulated	AT5G55310.1	7.01	6.00E-05	TOPOISOMERASE 1, DNA topoisomerase 1 beta
FT	Upregulated	AT3G03430.1	5.96	3.13E-04	EF-hand family protein
FT	Upregulated	AT5G17000.1	4.93	2.17E-02	Zinc-binding dehydrogenase family protein
FT	Upregulated	AT3G52090.2	4.82	5.31E-04	NRPB11
FT	Upregulated	AT5G49680.2	4.70	7.78E-03	similar to SABRE
FT	Upregulated	AT5G16990.1	4.69	9.36E-03	hypothetical protein
FT	Upregulated	AT5G55300.1	4.67	2.10E-05	FASCIATA5
FT	Upregulated	AT5G55300.2	4.67	2.10E-05	FASCIATA5
FT	Upregulated	AT2G35740.1	4.64	2.53E-02	nositol transporter 3
FT	Upregulated	AT5G55300.3	4.57	6.31E-04	FASCIATA5
FT	Downregulated	AT3G53500.1	-5.90	6.40E-04	arginine/serine-rich zinc knuckle-containing protein 32
FT	Downregulated	AT1G60730.2	-5.73	1.10E-04	NAD(P)-linked oxidoreductase superfamily protein
FT	Downregulated	AT1G10810.2	-5.70	1.25E-04	NAD(P)-linked oxidoreductase superfamily protein
FT	Downregulated	AT3G53500.2	-5.36	5.31E-04	arginine/serine-rich zinc knuckle-containing protein 32
FT	Downregulated	AT3G09040.1	-5.29	3.22E-02	mitochondrial RNA editing factor 12
FT	Downregulated	AT2G45770.1	-5.11	1.36E-05	FERRIC CHELATE REDUCTASE DEFECTIVE 4
FT	Downregulated	AT4G30310.4	-4.64	6.00E-05	FGGY family of carbohydrate kinase
FT	Downregulated	AT1G01780.1	-4.63	1.06E-02	PLIM2b
FT	Downregulated	AT1G74740.1	-4.62	4.21E-03	calcium-dependent protein kinase 30
FT	Downregulated	AT5G49500.2	-4.57	8.34E-04	Signal recognition particle SRP54 subunit protein

608

609

610

611

612

613 **Table S2.** Sex dimorphism related genes involved in floral development and secondary metabolism
614 adjacent to significant peaks in DAP-Seq analysis.

Library	Description	Target Gene	Description	Fold-change
94006	*AP2/ERF* Transcription Factor	Sapur.003G127200	AGL69,FCL4,MAF4, MADS-box transcription factor, flowering time	5.05
	AP2/ERF Transcription Factor	Sapur.008G077600	AGL4,SEP2, MADS-box transcription factor, ovule development	4.96
	AP2/ERF Transcription Factor	Sapur.009G065900	AGL62, MADS-box transcription factor, endosperm development	4.61
	AP2/ERF Transcription Factor	Sapur.013G069000	WUSCHEL related homeobox 11	5.88
	AP2/ERF Transcription Factor	Sapur.014G055900	AGL6, MADS-box transcription factor, floral organ development	3.34
'Fish Creek'	*AP2/ERF* Transcription Factor	SpFC.15G019600	cytokinin response factor 4	6.29
	AP2/ERF Transcription Factor	SpFC.13G066100	WUSCHEL related homeobox 11	4.35
94006	Homeodomain-like protein	Sapur.007G105000	beta-ocimene synthase	6.65
	Homeodomain-like protein	Sapur.019G014800	terpene synthase 21	5.94
	Homeodomain-like protein	Sapur.019G014700	terpene synthase 21	5.52
	Homeodomain-like protein	Sapur.004G024600	geranyl linalool synthase	4.91
	Homeodomain-like protein	Sapur.019G015500	beta-pinene synthase	3.99
	Homeodomain-like protein	Sapur.016G185200	pinene synthase	4.15
	Homeodomain-like protein	Sapur.011G018900	alpha-farnesene synthase	3.52
	Homeodomain-like protein	Sapur.016G013600	UDP-glucose flavonoid 3-O-glucosyltransferase	3.69
	Homeodomain-like protein	Sapur.011G018900	alpha-farnesene synthase	3.30
	Homeodomain-like protein	Sapur.019G014700	terpene synthase 21	3.29
	Homeodomain-like protein	Sapur.012G102300	AGL42, MADS-box transcription factor	6.25
	Homeodomain-like protein	Sapur.004G074100	AGL8,FUL, MADS-box transcription factor	5.31
	Homeodomain-like protein	Sapur.016G063100	AGL48, MADS-box transcription factor	5.45
	Homeodomain-like protein	Sapur.007G068100	ABS,AGL32,TT16, MADS-box transcription factor, necessary for ovule endothelial identity	5.63
	Homeodomain-like protein	Sapur.013G018400	AGL80,FEM111, MADS-box transcription factor	5.50
	Homeodomain-like protein	Sapur.012G022600	UDP-sugar flavonoid 7-O-glycosyltransferase	5.57
	Homeodomain-like protein	Sapur.007G047700	AGL104, MADS-box transcription factor, pollen tube development and growth	4.83
	Homeodomain-like protein	Sapur.012G045200	AGL8,FUL, MADS-box transcription factor	4.49
	Homeodomain-like protein	Sapur.004G074100	AGL8,FUL, MADS-box transcription factor	4.18
	Homeodomain-like protein	Sapur.004G074000	AGL4,SEP2, MADS-box transcription factor, ovule development	3.86
	Homeodomain-like protein	Sapur.013G001400	AGL80,FEM111, MADS-box transcription factor	4.40
	Homeodomain-like protein	Sapur.013G127700	UDP-glucose:flavonoid 3-O-glucosyltransferase	4.29
	Homeodomain-like protein	Sapur.006G043300	UDP-glucose:flavonoid 7-O-glucosyltransferase	3.30

	Homeodomain-like protein	Sapur.002G123800	AGL20,ATSOC1,SOC1, MADS-box transcription factor, flowering time	4.09
	Homeodomain-like protein	Sapur.016G014300	UDP-glucose flavonoid 3-O-glucosyltransferase	4.41
	Homeodomain-like protein	Sapur.003G126600	AGL10,CAL,CAL1, MADS-box transcription factor	3.16
	Homeodomain-like protein	Sapur.019G074900	AGL11,STK, MADS-box transcription factor, carpel expressed, seed development	4.07
	Homeodomain-like protein	Sapur.019G074500	AGL12,XAL1, MADS-box transcription factor, flowering time	3.79
	Homeodomain-like protein	Sapur.008G114500	TAPETUM 1	3.29
	Homeodomain-like protein	Sapur.016G203100	AGL8,FUL, MADS-box transcription factor	3.15
	Homeodomain-like protein	Sapur.008G083600	TPD1, tapetum determinant 1, anther and pollen development	3.18
	Homeodomain-like protein	Sapur.007G014400	AP3,ATAP3, MADS-box transcription factor, specifies stamen identity	4.73
	Homeodomain-like protein	Sapur.011G052400	AG, MADS-box transcription factor, specifies carpel and stamen identity	4.86
	Homeodomain-like protein	Sapur.004G044800	AG, MADS-box transcription factor, specifies carpel and stamen identity	4.04
	Homeodomain-like protein	Sapur.005G094500	AP3,ATAP3, MADS-box transcription factor, specifies stamen identity	3.03
	Homeodomain-like protein	Sapur.003G126500	AGL10,CAL,CAL1, MADS-box transcription factor	3.10
	Homeodomain-like protein	SpFC.13G018400	AGL80,FEM111, MADS-box transcription factor	7.96
	Homeodomain-like protein	SpFC.07G049600	AGL104, MADS-box transcription factor, pollen tube development and growth	6.48
	Homeodomain-like protein	SpFC.03G127900	AGL9,SEP3, MADS-box transcription factor, stamen development	6.47
	Homeodomain-like protein	SpFC.07G110800	AGL66, MADS-box transcription factor, pollen tube development and growth	6.20
	Homeodomain-like protein	SpFC.12G109600	AGL42, MADS-box transcription factor	6.10
	Homeodomain-like protein	SpFC.11G052600	AG, MADS-box transcription factor, specifies carpel and stamen identity	5.52
	Homeodomain-like protein	SpFC.07G069400	ABS,AGL32,TT16, MADS-box transcription factor, necessary for ovule endothelial identity	5.36
	Homeodomain-like protein	SpFC.19G078600	AGL11,STK, MADS-box transcription factor, carpel expressed, seed development	4.99
	Homeodomain-like protein	SpFC.12G046800	AGL8,FUL, MADS-box transcription factor	4.84
	Homeodomain-like protein	SpFC.04G045900	AG, MADS-box transcription factor, specifies carpel and stamen identity	4.83
	Homeodomain-like protein	SpFC.03G128100	AGL10,CAL,CAL1, MADS-box transcription factor	4.69
	Homeodomain-like protein	SpFC.13G001700	AGL80,FEM111, MADS-box transcription factor	4.68
	Homeodomain-like protein	SpFC.13G001800	AGL80,FEM111, MADS-box transcription factor	4.53
'Fish Creek'	Homeodomain-like protein	SpFC.02G131600	AGL20,ATSOC1,SOC1, MADS-box transcription factor, flowering time	4.44
	Homeodomain-like protein	SpFC.16G214600	AGL8,FUL, MADS-box transcription factor	4.41
	Homeodomain-like protein	SpFC.08G086200	AGL4,SEP2, MADS-box transcription factor, ovule development	4.34
	Homeodomain-like protein	SpFC.04G076000	AGL8,FUL, MADS-box transcription factor	4.06
	Homeodomain-like protein	SpFC.05G076600	ABS,AGL32,TT16, MADS-box transcription factor, necessary for ovule endothelial identity	3.78
	Homeodomain-like protein	SpFC.04G075900	AGL4,SEP2, MADS-box transcription factor, ovule development	3.71
	Homeodomain-like protein	SpFC.12G083700	AGL10,CAL,CAL1, MADS-box transcription factor	3.58
	Homeodomain-like protein	SpFC.16G214600	AGL8,FUL, MADS-box transcription factor	3.38
	Homeodomain-like protein	SpFC.04G099300	AGL29, MADS-box transcription factor	3.19
	Homeodomain-like protein	SpFC.19G078200	AGL12,XAL1, MADS-box transcription factor, flowering time	3.13
	Homeodomain-like protein	SpFC.15G019600	cytokinin response factor 4	6.25
	Homeodomain-like protein	SpFC.13G147000	cytokinin response factor 4	5.70
	Homeodomain-like protein	SpFC.01G079600	cytokinin response factor 2	4.59
	Homeodomain-like protein	SpFC.02G156100	cytokinin response factor 5	3.23

	Homeodomain-like protein	SpFC.16G119100	AGL82, MADS-box transcription factor	4.74
'Fish Creek'	*scarecrow*-like 18	SpFC.10G054800	RAB geranylgeranyl transferase alpha subunit 1	3.40
94006	*GATA* transcription factor 15	Sapur.012G027300	AT1G75560.1 homolog, zinc knuckle (CCHC-type), regulation of reproductive development	5.36
	GATA transcription factor 15	Sapur.006G190400	ATH1 transcription factor activated by AGAMOUS, regulates GA biosynthesis, flowering delay	5.10

Supplemental Datasets

Supplemental Dataset S1. Plasmid map (.dna format) for *ARR17* Sapur.15WG073500 expression plasmid (pBH100)

Supplemental Dataset S2. Plasmid map (.dna format) for *GATA15* Sapur.15WG062800 expression plasmid (pBH101)

Supplemental Dataset S3. Plasmid map (.dna format) for CCHC Zinc Finger Sapur.15WG068800 expression plasmid (pBH102)

Supplemental Dataset S4. Plasmid map (.dna format) for *DRB1* Sapur.15WG074300 expression plasmid (pBH103)

Supplemental Dataset S5. Plasmid map (.dna format) for hypothetical protein Sapur.15WG074900 expression plasmid (pBH104)

Supplemental Dataset S6. Plasmid map (.dna format) for hypothetical protein Sapur.15WG075700 expression plasmid (pBH106)

Supplemental Dataset S7. Plasmid map (.dna format) for *LEAFY* Sapur.15WG122200 expression plasmid (pBH107)

Supplemental Dataset S8. Plasmid map (.dna format) for *FT* Sapur.008G061900 expression plasmid (pBH108)

Supplemental Dataset S9. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.15WG062800 (pBH217)

Supplemental Dataset S10. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.15WG068800 (pBH218)

Supplemental Dataset S11. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.012G009500 (pBH219)

Supplemental Dataset S12. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.006G140600 (pBH220)

Supplemental Dataset S13. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.007G074000 (pBH225)

Supplemental Dataset S14. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.005G077400 (pBH226)

Supplemental Dataset S15. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.003G155500 (pBH227)

Supplemental Dataset S16. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.004G110200 (pBH228)

Supplemental Dataset S17. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.017G014200 (pBH229)

Supplemental Dataset S18. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.001G003600 (pBH232)

Supplemental Dataset S19. Plasmid map (.dna format) for pIX-HALO expression vector with Sapur.003G027300 (pBH242)

Supplemental Dataset S20. Expression data on all significant differentially abundant proteins for Arabidopsis expression lines 100 to 108.

Supplemental Dataset S21. Listing of all significant target genes for each DAP-Seq assay.

Bibliography

Acevedo FG, Gamboa A, Paéz-Valencia J, Jiménez-García LF, Izaguirre-Sierra M, Alvarez-Buylla ER (2004) FLOR1, a putative interaction partner of the floral homeotic protein AGAMOUS, is a plant-specific intracellular LRR. Plant Sci. **167:** 225-231

Akagi T, Pilkington SM, Varkonyi-Gasic E, Henry IM, Sugano SS, Sonoda M, Firl A, McNeilage MA, Douglas MJ, Wang T, Rebstock R, Voogd C, Datson P, Allan AC, Beppu K, Kataoka I, Tao R (2019) Two Y-chromosome-encoded genes determine sex in kiwifruit. Nature Plants **5:** 801-809

Albert E, Kavai-Ool U, Ezhova T (2015) Studying the role of *FASCIATA5* gene in the regulation of flower development in *Arabidopsis thaliana*. Russ. J. Dev. B. **46:** 19-26

Argus GW (1997) Infrageneric classification of *Salix* (Salicaceae) in the new world. Syst. Bot. Monog. **52:** 1-121

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in bipolymers. UCSD Technical Report: CS94-351.

Batth TS, Tollenaere MX, Rüther P, Gonzalez-Franquesa A, Prabhakar BS, Bekker-Jensen S, Deshmukh AS, Olsen JV (2019) Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. Mol. Cell Proteomics **18:** 1027-1035

Baumgart LA, Lee JE, Salamov A, Dilworth DJ, Na H, Mingay M, Blow MJ, Zhang Y, Yoshinaga Y, Daum CG (2021) Persistence and plasticity in bacterial gene regulation. Nat. Methods **18:** 1499-1505

Blighe K, Rana S, Lewis M (2018) EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling, Github.

Carlson CH, Choi Y, Chan AP, Serapiglia MJ, Town CD, Smart LB (2017) Dominance and sexual dimorphism pervade the *Salix purpurea* L. transcriptome. Genome Biol. Evol. **9:** 2377-2394

Charlesworth D (2002) Plant sex determination and sex chromosomes. Heredity **88:** 94-101

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

De Folter S, Shchennikova AV, Franken J, Busscher M, Baskar R, Grossniklaus U, Angenent GC, Immink RG (2006) A Bsister MADS-box gene involved in ovule and seed development in petunia and Arabidopsis. Plant J. **47:** 934-946

Dorca-Fornell C, Gregis V, Grandi V, Coupland G, Colombo L, Kater MM (2011) The Arabidopsis *SOC1*-like genes *AGL42*, *AGL71* and *AGL72* promote flowering in the shoot apical and axillary meristems. Plant J. **67:** 1006-1017

Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. **5:** 976-989

Gómez-Mena C, de Folter S, Costa MMR, Angenent GC, Sablowski R (2005) Transcriptional program controlled by the floral homeotic gene *AGAMOUS* during early organogenesis. Development **132**: 429-438

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2011) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. **40:** D1178-D1186

Harkess A, Huang K, van der Hulst R, Tissen B, Caplan JL, Koppula A, Batish M, Meyers BC, Leebens-Mack J (2020) Sex determination by two Y-Linked genes in garden asparagus. Plant Cell **32:** 1790-1796

Hu L, Zheng T, Cai M, Pan H, Wang J, Zhang Q (2019) Transcriptome analysis during floral organ development provides insights into stamen petaloidy in *Lagerstroemia speciosa*. Plant Physiol. Biochem. **142:** 510-518

Hyden B, Carlson CH, Gouker FE, Schmutz J, Barry K, Lipzen A, Sharma A, Sandor L, Tuskan GA, Feng G, Olson MS, DiFazio SP, Smart LB (2021) Integrative genomics reveals paths to sex dimorphism in *Salix purpurea* L. Hort. Res. **8:** 170

Hyden B, Zou J, Wilkerson DG, Carlson CH, Rivera Robles A, DiFazio SP, Smart LB (2023) Structural variation of a sex-linked region confers monoecy and implicates *GATA15* as a master regulator of sex in *Salix purpurea*. New Phytol. DOI 10.1111/nph.18853

Jing D, Chen W, Hu R, Zhang Y, Xia Y, Wang S, He Q, Guo Q, Liang G (2020) An integrative analysis of transcriptome, proteome and hormones reveals key differentially expressed genes and metabolic pathways involved in flower development in loquat. Int. J. of Mol. Sci. **21:** 5107

Käfer J, Marais GAB, Pannell JR (2017) On the rarity of dioecy in flowering plants. Mol. Ecol. **26:** 1225-1241

Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat. Methods **4:** 923-925

Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D (1999) Activation tagging of the floral inducer FT. Science **286:** 1962-1965

Keefover-Ring K, Carlson CH, Hyden B, Azeem M, Smart LB (2022) Genetic mapping of sexually dimorphic volatile and non-volatile floral secondary chemistry of a dioecious willow. J. Exp. Bot. **73**: 6352-6366

Krizek BA, Meyerowitz EM (1996) The Arabidopsis homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. Development **122:** 11-22

Kuzovkina YA, Weih M, Romero MA, Charles J, Hust S, McIvor I, Karp A, Trybush S, Labrecque M, Teodorescu TI (2007) *Salix*: botany and global horticulture. Hortic. Rev. **34:** 447-489

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. **40:** D1202-D1210

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics **30:** 3276-3278

Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in Arabidopsis. Nature **404:** 766-770

Liu W, Xu L, Lin H, Cao J (2021) Two expansin genes, *AtEXPA4* and *AtEXPB5*, are redundantly required for pollen tube growth and *AtEXPA4* is involved in primary root elongation in *Arabidopsis thaliana.* Genes **12:** 249

Liu Z, Meyerowitz EM (1995) *LEUNIG* regulates *AGAMOUS* expression in Arabidopsis flowers. Development **121:** 975-991

Lu J, Fu Y, Li M, Wang S, Wang J, Yang Q, Ye J, Zhang X, Ma H, Chang F (2020) Global quantitative proteomics studies revealed tissue-preferential expression and phosphorylation of regulatory proteins in Arabidopsis. Int. J. Mol. Sci. **21:** 6116

Mizukami Y, Ma H (1992) Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic Arabidopsis plants alters floral organ identity. Cell **71:** 119-131

Muller NA, Kersten B, Leite Montalvao AP, Mahler N, Bernhardsson C, Brautigam K, Carracedo Lorenzo Z, Hoenicka H, Kumar V, Mader M, Pakull B, Robinson KM, Sabatti M, Vettori C, Ingvarsson PK, Cronk Q, Street NR, Fladung M (2020) A single gene underlies the dynamic evolution of poplar sex determination. Nat. Plants **6:** 630-637

O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. Cell **165:** 1280-1292

Orsburn BC (2021) Proteome Discoverer—A community enhanced data processing duite for protein informatics. Proteomes **9:** 15

Pucholt P, Ronnberg-Wastljung AC, Berlin S (2015) Single locus sex determination and female heterogamety in the basket willow (*Salix viminalis* L.). Heredity **114:** 575-583

Ranftl QL, Bastakis E, Klermund C, Schwechheimer C (2016) LLM-domain containing B-GATA factors control different aspects of cytokinin-regulated development in *Arabidopsis thaliana*. Plant Physiol. **170:** 2295-2311

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. **43:** e47-e47

Rounsley SD, Ditta GS, Yanofsky MF (1995) Diverse roles for MADS box genes in Arabidopsis development. The Plant Cell **7:** 1259-1269

Sanderson BJ, Feng G, Hu N, Carlson CH, Smart LB, Keefover-Ring K, Yin T, Ma T, Liu J, DiFazio SP (2021) Sex determination through X–Y heterogamety in *Salix nigra*. Heredity **126:** 630-639

Schwacke R, Ponce-Soto GY, Krause K, Bolger AM, Arsova B, Hallab A, Gruden K, Stitt M, Bolger ME, Usadel B (2019) MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. Mol. Plant **12:** 879-892

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596-1604

Vickers CE, Schenk PM, Li D, Mullineaux PM, Gresshoff PM (2007) pGFPGUSPlus, a new binary vector for gene expression studies and optimising transformation systems in plants. Biotechnol. Lett. **29:** 1793-1796

Villalobos Solis MI, Giannone RJ, Hettich RL, Abraham PE (2019) Exploiting the dynamic relationship between peptide separation quality and peptide coisolation in a multiple-peptide matches-per-spectrum approach offers a strategy to optimize bottom-up proteomics throughput and depth. Anal. Chem. **91:** 7273-7279

Wang D, Li Y, Li M, Yang W, Ma X, Zhang L, Wang Y, Feng Y, Zhang Y, Zhou R, Sanderson BJ, Keefover-Ring K, Yin T, Smart LB, DiFazio SP, Liu J, Olson M, Ma T (2022) Repeated turnovers keep sex chromosomes young in willows. Genome Biol. **23:** 200

Wilkerson DG, Taskiran B, Carlson CH, Smart LB (2022) Mapping the sex determination region in the *Salix* F_1 hybrid common parent population confirms a ZW system in six diverse species. G3-Genes Genom. Genet. **12:** jkac071

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation **2:** 100141

Yang S-L, Xie L-F, Mao H-Z, Puah CS, Yang W-C, Jiang L, Sundaresan V, Ye D (2003) *Tapetum determinant1* is required for cell specialization in the Arabidopsis anther. The Plant Cell **15:** 2792-2804

Zhang MJ, Zhao TY, Ouyang XK, Zhao X-Y, Dai X, Gao X-Q (2022) Pollen-specific gene *SKU5-SIMILAR 13* enhances growth of pollen tubes in the transmitting tract in Arabidopsis. J. Exp. Bot. **73:** 696-710

Zhang S, Fernando DD (2005) Structural, histochemical, and protein analysis of male reproductive development in willow. Sex. Plant Reprod. **18:** 37-46

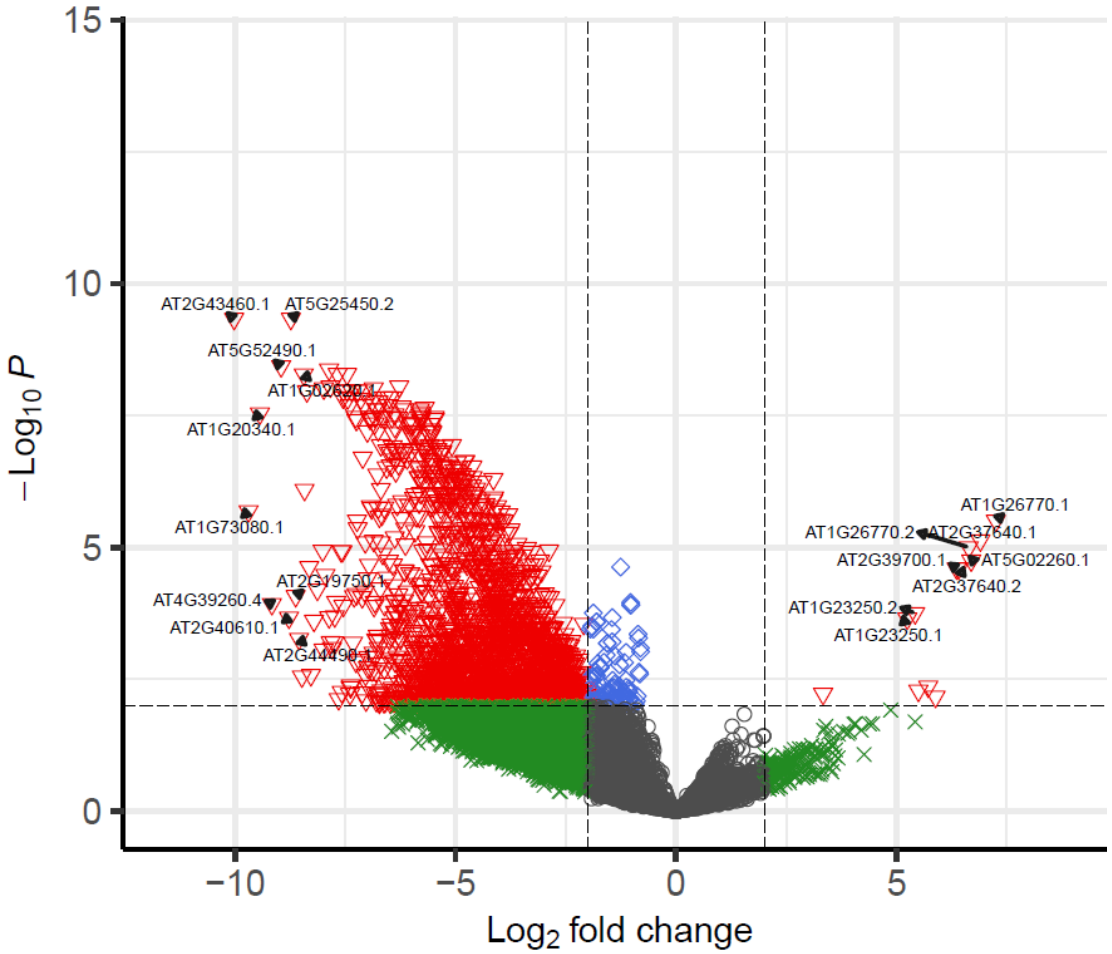
Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H (2006) *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. Nat. Protoc. **1:** 641-646

Zhou R, Macaya-Sanz D, Carlson CH, Schmutz J, Jenkins JW, Kudrna D, Sharma A, Sandor L, Shu S, Barry K, Tuskan GA, Ma T, Liu J, Olson M, Smart LB, DiFazio SP (2020) A willow sex chromosome reveals convergent evolution of complex palindromic repeats. Genome Biol. **21:** 38

Zhou R, Macaya-Sanz D, Rodgers-Melnick E, Carlson CH, Gouker FE, Evans LM, Schmutz J, Jenkins JW, Yan J, Tuskan GA, Smart LB, DiFazio SP (2018) Characterization of a large sex determination region in *Salix purpurea* L. (Salicaceae). Mol. Genet. Genom. **293:** 1437-1452

Arabidopsis 100 vs GUS samples

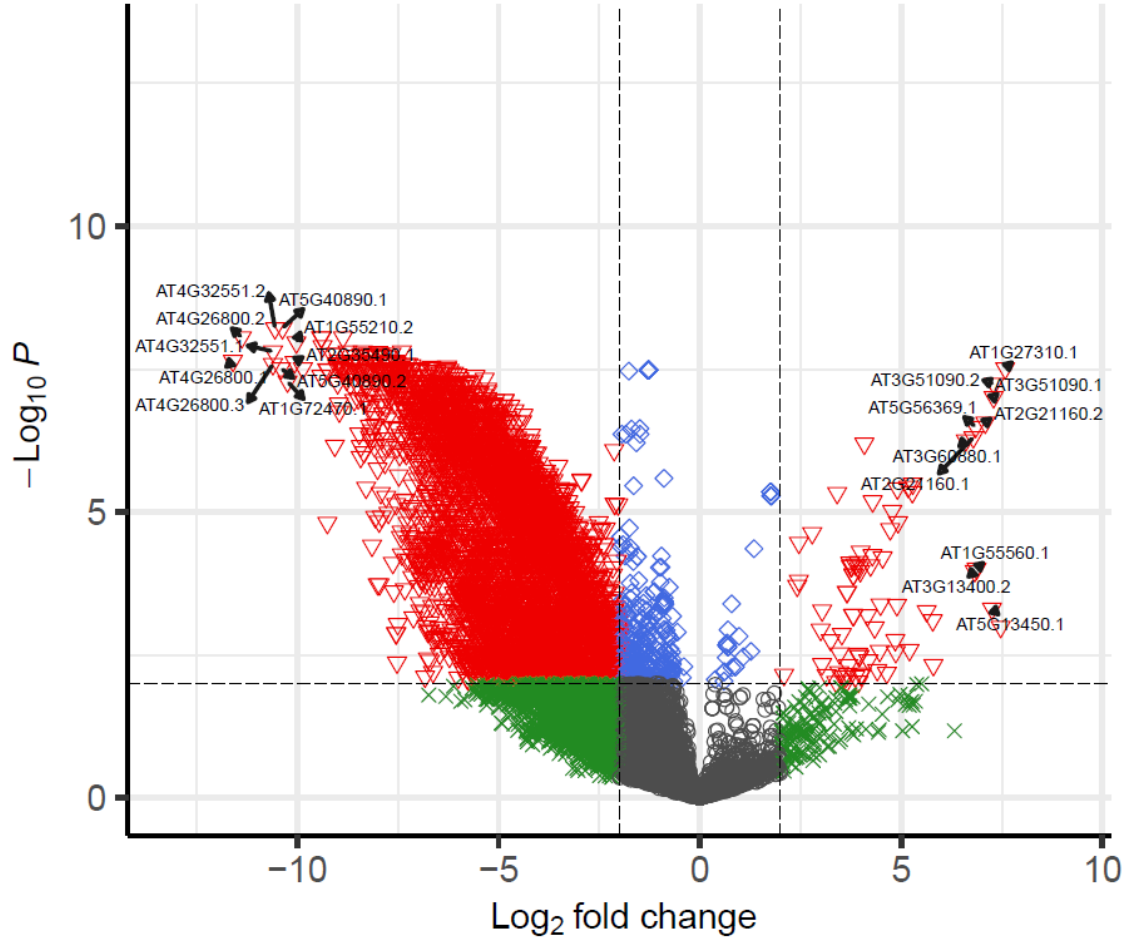
EnhancedVolcano



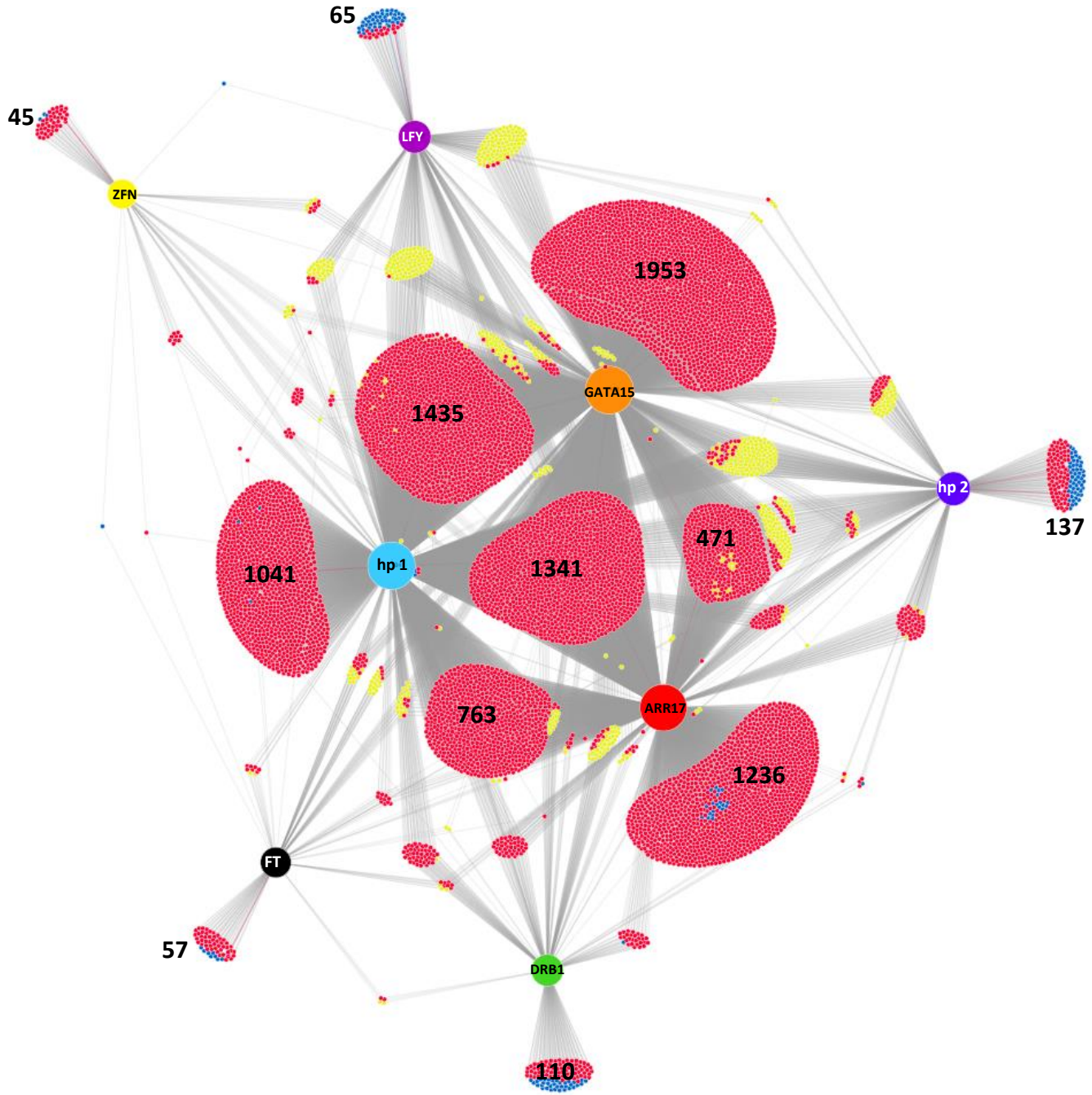
total = 16684 variables

Arabidopsis 101 vs GUS samples

EnhancedVolcano



total = 16609 variables


65
45
LFY
ZFN
1953
GATA15
1435
hp 2
137
471
1041
hp 1
1341
ARR17
763
1236
FT
57
DRB1
110

Male

Z

Z

Sapur.001G003600.1

CRF4
WOX11

Sapur.003G027300.1

AGAMOUS
FUL STK
AGL9 AGL66
SOC1 AGL4
CAL1 CRF4
CRF2 CRF5

Sapur.012G009500.1

geranyl-geranyl transferase

Female

Z

W

ARR17

GATA15

Sapur.001G003600.1

AGL69
AGL4
AGL62
WOX11
AGL6

MGH3
AGL72

PISTILLATA
FLOR1
HVT1
APC1
FAR2
FLD
TSM1
RPN1

ATH1
AT1G75560
FLOR1

AIM1
AGL72
SHP1
RPG1

Sapur.003G027300.1

FUL AGL32
AGL104 AGL4
SOC1 CAL1
XAL1 TAPETUM1
APETALA3
AGAMOUS
beta-ocimene
Linalool
beta-pinene
alpha-farnesene
Phenolic glucosides

♂ ♀

Parsed Citations

Acevedo FG, Gamboa A, Paéz-Valencia J, Jiménez-García LF, Izaguirre-Sierra M, Alvarez-Buylla ER (2004) FLOR1, a putative interaction partner of the floral homeotic protein AGAMOUS, is a plant-specific intracellular LRR. Plant Sci. 167: 225-231
Google Scholar: Author Only Title Only Author and Title

Akagi T, Pilkington SM, Varkonyi-Gasic E, Henry IM, Sugano SS, Sonoda M, Firl A, McNeilage MA, Douglas MJ, Wang T, Rebstock R, Voogd C, Datson P, Allan AC, Beppu K, Kataoka I, Tao R (2019) Two Y-chromosome-encoded genes determine sex in kiwifruit. Nature Plants 5: 801-809
Google Scholar: Author Only Title Only Author and Title

Albert E, Kavai-Ool U, Ezhova T (2015) Studying the role of FASCIATA5 gene in the regulation of flower development in Arabidopsis thaliana. Russ. J. Dev. B. 46: 19-26
Google Scholar: Author Only Title Only Author and Title

Argus GW (1997) Infrageneric classification of Salix (Salicaceae) in the new world. Syst. Bot. Monog. 52: 1-121
Google Scholar: Author Only Title Only Author and Title

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in bipolymers. UCSD Technical Report: CS94-351.
Google Scholar: Author Only Title Only Author and Title

Batth TS, Tollenaere MX, Rüther P, Gonzalez-Franquesa A, Prabhakar BS, Bekker-Jensen S, Deshmukh AS, Olsen JV (2019) Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. Mol. Cell Proteomics 18: 1027-1035
Google Scholar: Author Only Title Only Author and Title

Baumgart LA, Lee JE, Salamov A, Dilworth DJ, Na H, Mingay M, Blow MJ, Zhang Y, Yoshinaga Y, Daum CG (2021) Persistence and plasticity in bacterial gene regulation. Nat. Methods 18: 1499-1505
Google Scholar: Author Only Title Only Author and Title

Blighe K, Rana S, Lewis M (2018) EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling, Github.

Carlson CH, Choi Y, Chan AP, Serapiglia MJ, Town CD, Smart LB (2017) Dominance and sexual dimorphism pervade the Salix purpurea L. transcriptome. Genome Biol. Evol. 9: 2377-2394
Google Scholar: Author Only Title Only Author and Title

Charlesworth D (2002) Plant sex determination and sex chromosomes. Heredity 88: 94-101
Google Scholar: Author Only Title Only Author and Title

Cronk Q, Müller NA (2020) Default sex and single gene sex determination in dioecious plants. Front. Plant Sci. 11: 1162
Google Scholar: Author Only Title Only Author and Title

De Folter S, Shchennikova AV, Franken J, Busscher M, Baskar R, Grossniklaus U, Angenent GC, Immink RG (2006) A Bsister MADS-box gene involved in ovule and seed development in petunia and Arabidopsis. Plant J. 47: 934-946
Google Scholar: Author Only Title Only Author and Title

Dorca-Fornell C, Gregis V, Grandi V, Coupland G, Colombo L, Kater MM (2011) The Arabidopsis SOC1-like genes AGL42, AGL71 and AGL72 promote flowering in the shoot apical and axillary meristems. Plant J. 67: 1006-1017
Google Scholar: Author Only Title Only Author and Title

Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5: 976-989
Google Scholar: Author Only Title Only Author and Title

Gómez-Mena C, de Folter S, Costa MMR, Angenent GC, Sablowski R (2005) Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis. Development 132: 429-438
Google Scholar: Author Only Title Only Author and Title

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2011) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40: D1178-D1186
Google Scholar: Author Only Title Only Author and Title

Harkess A, Huang K, van der Hulst R, Tissen B, Caplan JL, Koppula A, Batish M, Meyers BC, Leebens-Mack J (2020) Sex determination by two Y-Linked genes in garden asparagus. Plant Cell 32: 1790-1796
Google Scholar: Author Only Title Only Author and Title

Hu L, Zheng T, Cai M, Pan H, Wang J, Zhang Q (2019) Transcriptome analysis during floral organ development provides insights into stamen petaloidy in Lagerstroemia speciosa. Plant Physiol. Biochem. 142: 510-518

Google Scholar: Author Only Title Only Author and Title

Hyden B, Carlson CH, Gouker FE, Schmutz J, Barry K, Lipzen A, Sharma A, Sandor L, Tuskan GA, Feng G, Olson MS, DiFazio SP, Smart LB (2021) Integrative genomics reveals paths to sex dimorphism in Salix purpurea L. Hort. Res. 8: 170
Google Scholar: Author Only Title Only Author and Title

Hyden B, Zou J, Wilkerson DG, Carlson CH, Rivera Robles A, DiFazio SP, Smart LB (2023) Structural variation of a sex-linked region confers monoecy and implicates GATA15 as a master regulator of sex in Salix purpurea. New Phytol. DOI 10.1111/nph.18853
Google Scholar: Author Only Title Only Author and Title

Jing D, Chen W, Hu R, Zhang Y, Xia Y, Wang S, He Q, Guo Q, Liang G (2020) An integrative analysis of transcriptome, proteome and hormones reveals key differentially expressed genes and metabolic pathways involved in flower development in loquat. Int. J. of Mol. Sci. 21: 5107
Google Scholar: Author Only Title Only Author and Title

Käfer J, Marais GAB, Pannell JR (2017) On the rarity of dioecy in flowering plants. Mol. Ecol. 26: 1225-1241
Google Scholar: Author Only Title Only Author and Title

Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat. Methods 4: 923-925
Google Scholar: Author Only Title Only Author and Title

Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D (1999) Activation tagging of the floral inducer FT. Science 286: 1962-1965
Google Scholar: Author Only Title Only Author and Title

Keefover-Ring K, Carlson CH, Hyden B, Azeem M, Smart LB (2022) Genetic mapping of sexually dimorphic volatile and non-volatile floral secondary chemistry of a dioecious willow. J. Exp. Bot. 73: 6352-6366
Google Scholar: Author Only Title Only Author and Title

Krizek BA, Meyerowitz EM (1996) The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. Development 122: 11-22
Google Scholar: Author Only Title Only Author and Title

Kuzovkina YA, Weih M, Romero MA, Charles J, Hust S, McIvor I, Karp A, Trybush S, Labrecque M, Teodorescu TI (2007) Salix: botany and global horticulture. Hortic. Rev. 34: 447-489
Google Scholar: Author Only Title Only Author and Title

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40: D1202-D1210
Google Scholar: Author Only Title Only Author and Title

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30: 3276-3278
Google Scholar: Author Only Title Only Author and Title

Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF (2000) SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. Nature 404: 766-770
Google Scholar: Author Only Title Only Author and Title

Liu W, Xu L, Lin H, Cao J (2021) Two expansin genes, AtEXPA4 and AtEXPB5, are redundantly required for pollen tube growth and AtEXPA4 is involved in primary root elongation in Arabidopsis thaliana. Genes 12: 249
Google Scholar: Author Only Title Only Author and Title

Liu Z, Meyerowitz EM (1995) LEUNIG regulates AGAMOUS expression in Arabidopsis flowers. Development 121: 975-991
Google Scholar: Author Only Title Only Author and Title

Lu J, Fu Y, Li M, Wang S, Wang J, Yang Q, Ye J, Zhang X, Ma H, Chang F (2020) Global quantitative proteomics studies revealed tissue-preferential expression and phosphorylation of regulatory proteins in Arabidopsis. Int. J. Mol. Sci. 21: 6116
Google Scholar: Author Only Title Only Author and Title

Mizukami Y, Ma H (1992) Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. Cell 71: 119-131
Google Scholar: Author Only Title Only Author and Title

Muller NA, Kersten B, Leite Montalvao AP, Mahler N, Bernhardsson C, Brautigam K, Carracedo Lorenzo Z, Hoenicka H, Kumar V, Mader M, Pakull B, Robinson KM, Sabatti M, Vettori C, Ingvarsson PK, Cronk Q, Street NR, Fladung M (2020) A single gene underlies the dynamic evolution of poplar sex determination. Nat. Plants 6: 630-637
Google Scholar: Author Only Title Only Author and Title

O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. Cell 165: 1280-1292
Google Scholar: Author Only Title Only Author and Title

Orsburn BC (2021) Proteome Discoverer-A community enhanced data processing duite for protein informatics. Proteomes 9: 15
Google Scholar: Author Only Title Only Author and Title

Pucholt P, Ronnberg-Wastljung AC, Berlin S (2015) Single locus sex determination and female heterogamety in the basket willow (Salix viminalis L.). Heredity 114: 575-583
Google Scholar: Author Only Title Only Author and Title

Ranftl QL, Bastakis E, Klermund C, Schwechheimer C (2016) LLM-domain containing B-GATA factors control different aspects of cytokinin-regulated development in Arabidopsis thaliana. Plant Physiol. 170: 2295-2311
Google Scholar: Author Only Title Only Author and Title

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43: e47-e47
Google Scholar: Author Only Title Only Author and Title

Rounsley SD, Ditta GS, Yanofsky MF (1995) Diverse roles for MADS box genes in Arabidopsis development. The Plant Cell 7: 1259-1269
Google Scholar: Author Only Title Only Author and Title

Sanderson BJ, Feng G, Hu N, Carlson CH, Smart LB, Keefover-Ring K, Yin T, Ma T, Liu J, DiFazio SP (2021) Sex determination through X–Y heterogamety in Salix nigra. Heredity 126: 630-639
Google Scholar: Author Only Title Only Author and Title

Schwacke R, Ponce-Soto GY, Krause K, Bolger AM, Arsova B, Hallab A, Gruden K, Stitt M, Bolger ME, Usadel B (2019) MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. Mol. Plant 12: 879-892
Google Scholar: Author Only Title Only Author and Title

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313: 1596-1604
Google Scholar: Author Only Title Only Author and Title

Vickers CE, Schenk PM, Li D, Mullineaux PM, Gresshoff PM (2007) pGFPGUSPlus, a new binary vector for gene expression studies and optimising transformation systems in plants. Biotechnol. Lett. 29: 1793-1796
Google Scholar: Author Only Title Only Author and Title

Villalobos Solis MI, Giannone RJ, Hettich RL, Abraham PE (2019) Exploiting the dynamic relationship between peptide separation quality and peptide coisolation in a multiple-peptide matches-per-spectrum approach offers a strategy to optimize bottom-up proteomics throughput and depth. Anal. Chem. 91: 7273-7279
Google Scholar: Author Only Title Only Author and Title

Wang D, Li Y, Li M, Yang W, Ma X, Zhang L, Wang Y, Feng Y, Zhang Y, Zhou R, Sanderson BJ, Keefover-Ring K, Yin T, Smart LB, DiFazio SP, Liu J, Olson M, Ma T (2022) Repeated turnovers keep sex chromosomes young in willows. Genome Biol. 23: 200
Google Scholar: Author Only Title Only Author and Title

Wilkerson DG, Taskiran B, Carlson CH, Smart LB (2022) Mapping the sex determination region in the Salix F1 hybrid common parent population confirms a ZW system in six diverse species. G3-Genes Genom. Genet. 12: jkac071
Google Scholar: Author Only Title Only Author and Title

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation 2: 100141
Google Scholar: Author Only Title Only Author and Title

Yang S-L, Xie L-F, Mao H-Z, Puah CS, Yang W-C, Jiang L, Sundaresan V, Ye D (2003) Tapetum determinant1 is required for cell specialization in the Arabidopsis anther. The Plant Cell 15: 2792-2804
Google Scholar: Author Only Title Only Author and Title

Zhang MJ, Zhao TY, Ouyang XK, Zhao X-Y, Dai X, Gao X-Q (2022) Pollen-specific gene SKU5-SIMILAR 13 enhances growth of

pollen tubes in the transmitting tract in Arabidopsis. J. Exp. Bot. 73: 696-710

Google Scholar: Author Only Title Only Author and Title

Zhang S, Fernando DD (2005) Structural, histochemical, and protein analysis of male reproductive development in willow. Sex. Plant Reprod. 18: 37-46

Google Scholar: Author Only Title Only Author and Title

Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat. Protoc. 1: 641-646

Google Scholar: Author Only Title Only Author and Title

Zhou R, Macaya-Sanz D, Carlson CH, Schmutz J, Jenkins JW, Kudrna D, Sharma A, Sandor L, Shu S, Barry K, Tuskan GA, Ma T, Liu J, Olson M, Smart LB, DiFazio SP (2020) A willow sex chromosome reveals convergent evolution of complex palindromic repeats. Genome Biol. 21: 38

Google Scholar: Author Only Title Only Author and Title

Zhou R, Macaya-Sanz D, Rodgers-Melnick E, Carlson CH, Gouker FE, Evans LM, Schmutz J, Jenkins JW, Yan J, Tuskan GA, Smart LB, DiFazio SP (2018) Characterization of a large sex determination region in Salix purpurea L. (Salicaceae). Mol. Genet. Genom. 293: 1437-1452

Google Scholar: Author Only Title Only Author and Title