Sex-biased gene expression in flowers, but not leaves, reveals secondary sexual dimorphism in *Populus balsamifera*

Brian J. Sanderson\(^1\), Li Wang\(^{1,2}\), Peter Tiffin\(^3\), Zhiqiang Wu\(^{1,2}\) and Matthew S. Olson\(^1\)

\(^1\)Department of Biological Sciences, Texas Tech University, Box 43131, Lubbock, TX 79409, USA; \(^2\)Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA; \(^3\)Department of Plant and Microbial Biology, University of Minnesota, St Paul, MN 55108, USA

*These authors contributed equally to this work.*

### Summary

- Because sexual dimorphism in plants is often less morphologically conspicuous than in animals, studies of sex-biased gene expression may provide a quantitative metric to better address their commonality, molecular pathways, consistency across tissues and taxa, and evolution. The presence of sex-biased gene expression in tissues other than the androecium or gynoecium, termed secondary sexual characters, suggests that these traits arose after the initial evolution of dioecy. Patterns of sequence evolution may provide evidence of positive selection that drove sexual specialization.
- We compared gene expression in male and female flowers and leaves of *Populus balsamifera* to assess the extent of sex-biased expression, and tested whether sex-biased genes exhibit elevated rates of protein evolution.
- Sex-biased expression was pervasive in floral tissue, but nearly absent in leaf tissue. Female-biased genes in flowers were associated with photosynthesis, whereas male-biased genes were associated with mitochondrial function. Sex-biased genes did not exhibit elevated rates of protein evolution, contrary to results from other studies in animals and plants.
- Our results suggest that the ecological and physiological constraints associated with the energetics of flowering, rather than sexual conflict, have probably shaped the differences in male and female gene expression in *P. balsamifera*.

### Introduction

Sexual dimorphism in dioecious plants (species with separate male and female individuals) tends to be less common and less conspicuous than in animals (Barrett & Hough, 2012). The lack of vegetative sexual dimorphism was particularly striking in recent studies of dioecious *Populus tremula, P. trichocarpa* and *P. balsamifera*, which failed to uncover any sexual dimorphism in the morphology or physiology of vegetative tissues, despite large sample sizes and extensive phenotyping (Robinson et al., 2014; McKown et al., 2017). Nonetheless, dimorphism in vegetative traits is present in some dioecious plants, such as the dramatic differences in leaf and branch architecture between males and females of *Leucodendron* spp., which allow easy identification of sexes, even when the plants are not flowering (Midgley, 2010; Welsford et al., 2016). *Silene latifolia* also exhibits sex-specific differences in floral and vegetative morphology and physiology that probably evolved in conjunction with sex-specific tradeoffs in resource allocation (Delph et al., 2002; Steven et al., 2007; Delph & Bell, 2008). Sexual dimorphism in other plants may be manifest not as differences in morphology, but as differences in habitat preference or life history traits, such as growth rate, phenology and longevity (Lloyd & Webb, 1977; Sakai & Burris, 1985; Dawson & Bliss, 1989; Barrett & Hough, 2012). Sexual dimorphism in plant reproductive traits tends to be more common than in vegetative traits (Barrett & Hough, 2012), and includes differences in flower number, flower size, floral defense, inflorescence size, flowering phenology, nectar and floral longevity (see reviews in Delph, 1999; Eckhart, 1999; Barrett & Hough, 2012).

Sexual dimorphism evolves when reproductive output is maximized by different traits in males and females, either through increasing mating success or because maximum reproductive output relies on differing resource allocation in the two sexes (Geber, 1999). Thus, conflicts arise when an allelic variant is expressed in the wrong sex, and can be resolved either by the evolution of linkage between the sexually antagonistic locus and sex determination regions (SDRs) (Charlesworth & Charlesworth, 1980; Rice, 1987) or by the evolution of sex-biased gene regulation (Ellegren & Parsch, 2007). If sexual antagonism has been a major force driving the evolution of sex-limited or sex-biased genes, we might expect that those genes would evolve in response to positive selection (Ellegren & Parsch, 2007). Signatures of elevated adaptive evolution of sex-biased genes are found in both animals and plants (Grath & Parsch, 2016). The rate of protein evolution (the ratio of non-synonymous to synonymous substitutions, dN/dS)
is higher for genes with male-biased expression in *Drosophila* (Pröschel et al., 2006; Baines et al., 2008; Grath & Parsch, 2012) and for genes in mouse sperm membrane cells, which are associated with sperm competition and mate recognition (Dorus et al., 2010). In birds (Galloanserae clade), genes with male-biased expression also exhibit elevated rates of protein evolution, and the degree of sexual ornamentation is correlated with the proportion of genes with male-biased expression, suggesting that sexual selection targets expression levels in males (Harrison et al., 2015). In hermaphroditic plants, genes with pollen-limited expression in both the self-fertilizing *Arabidopsis thaliana* and the outcrossing *Capsella grandiflora* exhibit increased rates of amino acid substitutions and stronger effects of purifying selection compared with genes expressed only in seedlings (*Capsella*) and sperm (*Arabidopsis*; Arunkumar et al., 2013; Gossmann et al., 2014). By contrast, in the dioecious *Salix viminalis*, male-biased genes show slower rates of protein evolution than female-biased or unbiased genes (Darolti et al., 2018).

In plants, traits with a direct relation to gamete development, including the production of the androecium and gynoecium, are characterized as primary sexual dimorphisms, whereas secondary sexual dimorphisms are not directly involved in gamete production. This classification allows a distinction between traits that arose as a direct consequence of the evolution of dioecy (primary), and those that may also have evolved as specializations during or after the evolution of dioecy (secondary; reviewed by Charlesworth, 2018). With regard to gene expression, genes that are expressed in both sexes, but at different levels, are termed sex-biased genes. An extreme form of sex bias is sex limitation, such as those genes that occur only on the W or Y chromosome, and thus are found only in one sex. Genes on the X, Z or autosomes may also express sex-limited expression, resulting from either the lack of production of corresponding tissues in both sexes or because selection on secondary traits has driven their expression to be completely lost in one sex.

In dioecious animal and plant species, even when there are highly diverged heteromorphic sex chromosomes, males and females only differ in the small number of genes that lie inside the non-recombining SDRs (Rice, 1987; Bergero & Charlesworth, 2011; Wang et al., 2012; Charlesworth et al., 2014; Geraldes et al., 2015). Thus, most phenotypic sexual dimorphism is determined by differences in gene dosage and gene regulation, rather than the expression of genes found only in one sex (Parsch & Charlesworth, 2011; Wang et al., 2012; Charlesworth et al., 2014; Geraldes et al., 2015). Therefore, sex-biased gene expression will also be uncommon.

In this study, we sought to determine the extent of sex-biased and sex-limited gene expression in the wind-pollinated dioecious tree *P. balsamifera*. Our objectives were two-fold. First, we characterized gene expression differences between the sexes in leaves and flowers, and the genetic pathways that are enriched for sex-biased and sex-limited gene expression. Second, we asked whether sex-biased and sex-limited genes show evidence of response to positive selection, which is predicted if a history of sexual conflict has shaped the differences between the sexes.
Materials and Methods

Sample collection, RNA extraction, sequencing and assembly

Dormant stems of five female and five male balsam poplar (*Populus balsamifera* L.) trees bearing both flower and leaf buds were sampled in early spring 2013 from one natural population in Fairbanks, Alaska (64°48’19”N, 147°50’33”W). Sampled individuals were growing at distances separated by >100 m interspersed with trees of different species to minimize the possibility of collecting multiple samples from the same individual. The cuttings were kept frozen during transport to Texas Tech University, and then flushed by placing the unrooted stems in water at room temperature. Whole male and female catkins were collected 3 d after catkin bud burst, when the catkin was fully expanded, but before the flowers opened (Fig. 1). Flowers were cut from the catkin rachis and immediately frozen in liquid nitrogen for RNA extraction. Leaves were sampled 10 d after leaf out, when they were fully expanded, and immediately frozen in liquid nitrogen for RNA extraction.

RNA of the two tissues (flower and leaf) from five females and five males was extracted using a Qiagen RNeasy Plant kit (Qiagen, Hilden, North Rhine-Westphalia, Germany) with the addition of 4% w/v PVP-40 (polyvinylpyrrolidone) during tissue maceration, just before adding the RLT extraction buffer. The quality and quantity of total RNA were determined using an RNA 6000 Pico Chip on an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA, USA). RNA was applied to MultiMACS columns (Miltenyi Biotec, Bergisch Gladbach, North Rhine-Westphalia, Germany) to enrich for mRNA before library preparation. Libraries were barcoded and pooled, eight per lane, before sequencing on an Illumina Hi-Seq 2000 (Illumina, San Diego, CA, USA) to produce 24,345,776 ± 1992,470 (mean ± SD; Supporting Information Table S1) 50-bp paired-end reads per library. Library preparation, labeling and sequencing were conducted at the University of Minnesota Genomics Center (http://genomics.umn.edu/). The raw reads are available on the National Center for Biotechnology Information (NCBI) sequence read archive (https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP154132).

Adapter sequences were trimmed from the reads, and the read sequences were filtered for quality (Q = 15) and minimum length (l = 30) using FASTQ-MCF (Table S1; Aronesty, 2011, 2013). To identify which genes were differentially expressed, we aligned the trimmed and filtered reads to the *Populus trichocarpa* reference genome version 3 (Tuskan et al., 2006) using the program HISAT2 v.2.0.4 with the parameters --no-mixed, --no-discordant, --score-min L,0,-0.1 and --dta (Kim et al., 2015). An average of 63.5% of read pairs mapped once concordantly to the genome with these parameters. The resulting alignments were sorted by name and compressed into BAM files using SAMTOOLS v.1.3.1 (Li et al., 2009; Li, 2011). Gene expression was quantified using the program HTSEQ-COUNT v.0.6.0 (Anders et al., 2015).

Identification of sex-limited and sex-biased genes

We classified genes as sex and/or tissue limited if they were expressed at values > 0.1 counts per million (CPM) in the target libraries (for example, all five male flowers for male flower-limited genes, or all five male flowers and all five female flowers...
for flower-limited genes), and < 0.1 CPM in all of the non-target libraries, using the cpm function in the R package edgeR (Table S2; Robinson et al., 2010; R Core Team, 2016). Although these strict cutoffs resulted in some non-intuitive results (e.g., a gene would not be classified as male flower specific when all five male flowers exhibited high expression and it was also expressed in only one male leaf sample and no female tissues), we present this analysis to provide a heuristic assessment of the patterns of sex-limited expression. Qualitatively similar patterns of sex-limited gene expression were identified regardless of the strictness of cutoffs (e.g., if only four of five replicates of tissues were required for inclusion/exclusion). The script used to identify the patterns of sex limitation is available at https://github.com/BrianSanderson/gene-expression/. We plotted a Venn diagram of the sex- and tissue-limited genes using the function venn.diagram in R (Hanbo, 2018).

We classified genes as being sex biased (expressed at different levels in each sex) using metrics of differential gene expression (DGE). For this analysis, flower and leaf tissues were analyzed separately. To limit false positives resulting from lowly expressed genes, genes with read counts below 0.1 CPM for five or more individuals were removed before downstream analyses using the function cpm in edgeR (Robinson et al., 2010). It should be noted that these cutoffs resulted in some genes being classified as both sex limited and differentially expressed. Because different methods can yield dramatically different estimates of DGE (Zyprych-Walczak et al., 2015; Conesa et al., 2016), we estimated DGE using three methods implemented in edgeR v.3.14.0 (Robinson et al., 2010), DESeq2 1.12.4 (Love et al., 2014) and limma 3.28.21 (Ritchie et al., 2015). A multidimensional scaling (MDS) plot of the expression differences among libraries showed that samples clustered by sex and tissue type rather than by sequencing lane, indicating that batch effects were minimal (Fig. S1). Normalization of the read count data was performed using the trimmed mean of ‘M’ values (TMM) method in edgeR and limma, and the analogous median method implemented in DESeq2. We specified a false discovery rate (FDR) of 0.01 as a threshold for identifying the significance of DGE, using the Benjamini–Hochberg method implemented in edgeR, DESeq2 and limma. We used the consensus of all three estimates of DGE that passed the FDR threshold of 0.01 to identify the set of genes that differed significantly in expression between males and females in both flower and leaf tissues, following Colicchio et al. (2015) (see Figs S2 and S3 for a comparison of the results from individual analyses). The R function used to perform this analysis is available at https://github.com/BrianSanderson/gene-expression/. We visualized DGE for the log2 fold change (log2 FC) values for reads expressed in floral tissues using the heatmap function in the R package PLOTS (Warnez et al., 2016).

Gene ontology (GO) and pathway enrichment analysis

The R package TOPGO (Alexa & Rahnenführer, 2010) was used for GO enrichment analysis. GO terms with < 5 annotated genes were pruned from the hierarchy. Fisher exact tests were used to estimate the enrichment P values.

Annotated functions of differentially expressed (DE) genes were further characterized to identify enriched pathways with MapMan (Thimm et al., 2004; Usadel et al., 2005). The protein sequences of DE genes were extracted from the P. trichocarpa v3 proteins sequences database via cdbfasta (https://sourceforge.net/projects/cdbfasta/). Mercator (Lohse et al., 2014) was employed to annotate the DE gene sequences with default settings in order to generate the mapping file, required as input in the pathway enrichment analysis in MapMan. A Wilcoxon rank sum test was used to identify the enriched pathways.

Signatures of adaptive evolution of sex-biased and sex-limited genes

To quantify the rates of molecular evolution for sex-biased and sex-limited genes, we calculated dN/dS for the P. trichocarpa genes since their split with P. tremula and using P. euphratica as an outgroup. We used the annotated P. trichocarpa gene sequences for this comparison, because it is sister to P. balsamifera (Ismaiel et al., 2012; Levens et al., 2012) and an annotated genome does not exist for P. balsamifera. We identified homologs from P. trichocarpa v3 (Tuskan et al., 2006) in the P. tremula v.1.1 (Sundell et al., 2015) and P. euphratica v.2 (Ma et al., 2013) genome annotations using BLASTN v2.5.0 (Camacho et al., 2009), based on the best hit that optimized the length of the local alignment and the alignment score (R script available at https://github.com/BrianSanderson/gene-expression/). For each gene, nucleotide sequences were translated to protein sequences using the transeq function in EMBOSs v.6.6.0 (Rice et al., 2000) before protein sequence alignment using MAFFT v.7.123b (Katoh & Standley, 2013). Alignments were translated back into DNA codon alignments using PAL2NAL v.1.4 (Suyama et al., 2006). This resulted in a set of 24 537 three-species alignments of annotated genes.

The values of dN/dS for both the P. trichocarpa and P. tremula lineages were estimated using codeml (PAML v.4.7), employing a free-ratio branch model (Yang, 2007). We excluded from downstream analyses 4908 alignments with unrealistically high divergence along the P. balsamifera lineage (S × dS > N × dN > 15% of the alignment length). Also, because dN/dS is a ratio, genes with very low dS may provide unreasonable estimates of protein evolution, and so 7052 alignments with very low numbers of synonymous substitutions (S × dS < 3) were excluded. The mean value of S × dS before applying these filtering criteria was 13.60 ± 146.37, and the mean value of S × dS after filtering was 8.68 ± 11.10. The final set of 13 047 alignments was used to compare values of dN/dS between sex-biased, sex-limited and randomly selected genes. These filtering criteria did not change the outcomes of the following statistical tests, compared with an analysis using the unfiltered dataset.

To determine whether the sex-biased and sex-limited genes exhibited elevated rates of protein evolution, we performed two comparisons of dN/dS values for the lineage from the P. trichocarpa–P. tremula common ancestor to P. trichocarpa (hereafter P. trichocarpa lineage), and for the lineage from P. trichocarpa to P. tremula. These two species harbor different
SDRs (Kersten et al., 2014; Geraldes et al., 2015). First, we compared the median values of dN/dS for sex-biased and sex-limited genes with a random sample of unbiased genes, using a non-parametric Kruskal–Wallis test (function kruskal.test in R, post hoc tests were conducted using dunn.test). Also, to determine whether there were differences in the rate of neutral evolution, we compared the median values of dS of sex-biased and sex-limited genes with randomly selected unbiased genes. Second, we tested whether sex-biased and sex-limited genes were more likely to exhibit extreme values of dN/dS by comparing the 95% quantiles of these categories with a distribution of the 95% quantile values of 5000 bootstrap samples of unbiased genes using the package boot (Davison & Hinkley, 1997; Canty & Ripley, 2017). Because the results of the analyses for the P. trichocarpa lineage and the P. trichocarpa–P. tremula lineage were qualitatively the same, we only report the results for the P. trichocarpa lineage.

Results

Sex-biased and sex-limited gene expression

We found 30 503 genes expressed in flowers and 29 482 genes expressed in leaves, after filtering for low expression (<0.1 CPM; Table S3). Of these, 11 068 genes showed sex-biased expression in flowers (FDR < 0.01; Fig. 2). Twice as many genes were more highly expressed in female flowers than male flowers (7767 up-regulated in female flowers and 3301 in male flowers; Table S4). Only 325 of the 11 068 genes differed in transcript abundance by less than two-fold (log₂ FC < 1; Figs S4, S5), and so we concluded that these results reflect biologically relevant levels of gene expression dimorphism. In contrast with the large number of genes differentially expressed in flowers, in leaf tissue, only one gene was differentially expressed between males and females (Potri.018G0758000, function unknown; log₂ FC > 2.51; FDR < 8.9 × 10⁻⁴).

Far fewer genes showed sex-limited than sex-biased expression. We found that 428 genes (426 + 2; Fig. 3) were expressed in all male flowers and no other tissues, whereas 640 genes (76 + 564; Fig. 3) were expressed in all female flowers and no other tissues. With respect to sex-limited gene expression, male flowers were more unique than female flowers: 99.5% (426/428) of the genes expressed in all male, but no female, flowers were also not expressed in leaves, whereas only 11.9% (76/640) of genes expressed in all female, but no male, flowers were unique to female flowers. We found no genes with expression limited to only male leaves or only female leaves, and only 52 genes were expressed in all leaves, but no flowers (Fig. 3).

Genes in the SDR were no more likely to exhibit sex-biased expression than genes outside of the SDR (likelihood ratio χ² = 0.698, P > 0.40). Three of the 12 genes located in the genomic SDR in P. balsamifera (Geraldes et al., 2015) exhibited sex-biased expression in flowers: PbRR9 (Potri.019G133600), which was up-regulated in males, and PbCLC-C (Potri.018G138100) and PbTCP-1 (Potri.018G138200), which were up-regulated in females. However, transcript abundance of PbRR9 in male flowers was low (median raw numbers of transcripts sequenced per RNA library = 7, range 3–43; Table S3), and transcripts were detected in only one female flower and not in any male and female leaves. None of the male or female sex-limited genes (genes expressed in only one sex) were in the genomic SDR of P. balsamifera (as identified by Geraldes et al., 2015).

![Fig. 2 Heatmap of differential gene expression in male and female flowers of Populus balsamifera. The color gradient represents variation in natural log-transformed counts per million (CPM) reads of gene expression. Cooler colors (purple-blue) represent low gene expression, and warmer colors (red-orange-yellow) represent high gene expression. Each row in the heatmap is a single gene, and each column is an individual male (M1–M5) or female (F1–F5) tree.](image-url)
Biological processes associated with sex-biased genes

MapMan and GO analyses identified similar patterns of enrichment among sex-biased genes in flowers (Fig. 4; Tables 1, S5). As expected from the greener color of female than male flowers, the photosynthesis pathway \( (P < 10^{-16}) \), including light reactions, photorespiration and the Calvin cycle, was dominated by female-biased genes (161 of 164, 98.2%; Table S6). In addition, the tetrapyrrole synthesis pathway \( (P = 1.06 \times 10^{-9}) \), which is associated with chlorophyll production, was composed mainly of female-biased genes (30 of 31, 96.8%; Table S6). Elevated expression in both chlorophyll \( a \) and \( b \) pathways in females is consistent with the observation that female flowers are green and probably remain photosynthetically active during fruit development (Fig. 1). By contrast, male-biased genes were overrepresented in the tricarboxylic acid cycle (TCA) pathway \( (P = 2.82 \times 10^{-4}; 22 of 27, 81.5\%); \) Table S6) and the mitochondrial electron transport/ATP synthesis pathway \( (P = 6.2 \times 10^{-4}; 25 of 35, 71.4\%); \) Table S6). The enrichment of these pathways in males is consistent with increased numbers of mitochondria, which is associated with viable pollen production (Hanson, 1991).

Additional metabolic processes in flowers also exhibited sexual dimorphism in gene expression. Floral transcripts in the glucosinolate and flavonoid pathways, which produce important defense compounds (as identified by Tsai et al., 2006), exhibited increased abundance in female flowers relative to male flowers (26 female-biased genes among a total of 29; Fig. S6; Table S7). The lysine synthesis pathway \( (P = 0.05) \) was also dominated by female-biased genes (eight of eight). By contrast, transcript abundance was higher in male relative to female flowers in the sucrose degradation \( (P = 0.0066; 15 of 19) \) and sucrose synthesis \( (P = 0.0038; six of eight) \) pathways, the cell wall precursor synthesis pathway \( (P = 4.02 \times 10^{-6}; 24 of 27) \) and the methionine synthesis pathway \( (P = 0.005; 11 of 13; \) Fig. 4; Table S6).

Biological processes associated with sex-limited genes

Sex-limited genes combined for both male and female flowers were overrepresented in the RNA \( (P < 0.0001; 33 male-\) and 20 female-limited genes), hormone metabolism \( (P < 0.005; six male-\) and five female-limited genes) and transport \( (P < 0.05; 30 male-limited genes) \) pathways (Table S8). When transcripts from male flower-limited genes (426) were compared with the genes expressed in all tissues except male flowers (564; Fig. 3), some of the same pathways as discovered for the DE genes were identified as enriched: cell wall \( (P = 0.001; 19 male flower-limited genes of a total of 23) \), photosynthesis \( (P = 0.003; only one male flower-limited gene of a total of 16) \), lipid metabolism \( (P = 0.024; 10 male flower-limited genes of a total of 13) \) and mitochondrial electron transport/ATP synthesis \( (P = 0.046; Table S8) \).

Evolution of sex-limited and sex-biased genes

The median dN/dS did not differ among genes expressed in only male or female flowers (sex-limited) and 3000 randomly chosen unbiased genes (Fig. 5; Kruskal–Wallis test, \( P = 0.960, df=2 \)). We also identified 82 genes that were present in only the male flowers of our study and were also expressed during pollen germination and pollen growth in hybrid poplars (Zhao et al., 2016). The median dN/dS for these 82 genes was not significantly different from that of genes selected randomly from the genome (Fig. 5; Kruskal–Wallis test, \( P = 0.525, df=1 \)).

For sex-biased genes, the median values of dN/dS were significantly lower for both male and female sex-biased genes than
unbiased genes (Kruskal–Wallis test, $P = 5.06 \times 10^{-9}$, df = 2), although the median values of male- and female-biased genes did not differ from each other (Dunn’s test, $P = 0.214$). The distribution of values of dS showed the same patterns as those observed for values of dN/dS. Sex-limited genes and randomly chosen unbiased genes did not differ (Kruskal–Wallis test, $P = 0.115$, df = 2), dS was significantly lower for both male- and female-biased genes compared with randomly selected unbiased genes (Kruskal–Wallis test, $P = 1.17 \times 10^{-14}$, df = 2), and dS in male- and female-biased genes did not differ from each other (Dunn’s test, $P = 0.321$).

Because the large number of sex-biased genes in this comparison may be masking a signal of elevated rates of molecular evolution, we conducted two additional analyses. First, we compared the distribution of dN/dS values of the 100 genes with the strongest male bias (based on the magnitude of the log$_2$FC values) and the 100 genes with the strongest female bias with randomly selected unbiased genes. Similar to the results from the full set of sex-biased genes, these highly sex-biased genes showed lower rates of dN/dS than unbiased genes, although this relationship was not statistically significant (Kruskal–Wallis test, $P > 0.08$, df = 2). Second, we compared the upper 95% quantile of sex-biased and sex-limited genes with the distribution of upper 95% quantiles from 5000 random bootstrap samples of dN/dS values from unbiased genes, and found that the quantiles for the male- and female-biased genes (0.889 and 0.906, respectively) fell below all of the upper 95% quantiles from the bootstrapped sample (unbiased 95% quantile range: 1.050–1.202). We conclude that these categories of genes are not enriched for extremely high values of dN/dS, compared with unbiased genes.
Table 1 Enriched functional pathways identified by MapMan for sex-biased genes in Populus balsamifera.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>No. of genes</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photosynthesis</td>
<td>164</td>
<td>$&lt; 10^{-16}$</td>
</tr>
<tr>
<td>Transport</td>
<td>512</td>
<td>$3.63 \times 10^{-16}$</td>
</tr>
<tr>
<td>RNA</td>
<td>1330</td>
<td>$3.11 \times 10^{-14}$</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>211</td>
<td>$1.25 \times 10^{-7}$</td>
</tr>
<tr>
<td>Protein</td>
<td>1516</td>
<td>$5.17 \times 10^{-6}$</td>
</tr>
<tr>
<td>Cell wall</td>
<td>205</td>
<td>$1.06 \times 10^{-5}$</td>
</tr>
<tr>
<td>Tetrapyrole synthesis</td>
<td>31</td>
<td>$1.06 \times 10^{-5}$</td>
</tr>
<tr>
<td>Secondary metabolism</td>
<td>221</td>
<td>$3.19 \times 10^{-5}$</td>
</tr>
<tr>
<td>Cell</td>
<td>384</td>
<td>$1.21 \times 10^{-4}$</td>
</tr>
<tr>
<td>Misc.</td>
<td>609</td>
<td>$2.61 \times 10^{-4}$</td>
</tr>
</tbody>
</table>
| Tricarboxylic acid cycle (TCA)/org.
  synthesis                               | 24           | $2.82 \times 10^{-4}$ |
| Mitochondrial electron transport/ATP
  synthesis                               | 35           | $6.20 \times 10^{-4}$ |
| Minor carbohydrate (CHO) metabolism     | 66           | $7.37 \times 10^{-3}$ |
| Hormone metabolism                       | 295          | $2.15 \times 10^{-2}$ |
| Signaling                                | 738          | $3.97 \times 10^{-2}$ |
| Major CHO metabolism                     | 62           | $4.49 \times 10^{-2}$ |

**Fig. 5** Distribution of values of dN/dS for genes limited in expression to female (red) or male (blue) flowers in Populus balsamifera, compared with genes randomly chosen from the genome (white) among sex-biased and randomly selected genes from the genome. The ‘pollen’ category represents the distribution of dN/dS values for genes identified by another study in Populus (Zhao et al., 2016), which were also limited in expression to male flowers in our data. The pollen genes are a subset of the male-limited genes, and so the statistical tests reported in the Results section are from separate analyses. The center line represents the median value of dN/dS for that distribution.

**Discussion**

Males and females in *Populus balsamifera* differ in gene expression almost solely in their flowers. In tissues sampled from the same stem, flowers showed sex-biased expression in nearly one-third of annotated genes, whereas leaves showed sex bias in only one gene. This low level of sex-biased gene expression in leaves is consistent with findings from two recent studies that both provided strong evidence for the absence of sexual dimorphism in non-floral tissues in the genus *Populus*, including *P. balsamifera* (Robinson et al., 2014; McKown et al., 2017). Because many vegetative characteristics are sexually homomorphic in *P. balsamifera* (McKown et al., 2017), sexually dimorphic adaptations in vegetative characters may either be absent or isolated to specific environments or tissues not sampled in our study. Robinson et al. (2014) also found no sexual dimorphism in vegetative traits and no sex-biased expression in leaf tissues for *P. tremula*, except for one gene located in the SDR. Nonetheless, the broad patterns of sex-biased gene expression that we found in floral tissues, including sex-biased expression in photosynthetic and defense compounds, indicate the presence of secondary sexual dimorphism in *P. balsamifera*. Although these traits are conspicuous and ephemeral on large poplar trees, they probably reflect the impacts of sexual specialization.

Sex-biased gene expression is often greater in reproductive than in non-reproductive tissues, although the magnitude of the differences between tissues can vary substantially (Mank, 2017). Several examples in animals have shown a more than ten-fold greater proportion of genes differentially expressed in gonad than in somatic tissues (*Drosophila*: Assis et al., 2012; Catalán, 2012; guppies and zebrafish: Sharma et al., 2014; Yang et al., 2016; Galloanserae birds: Pointer et al., 2013; Harrison et al., 2015). There are fewer comparisons of sex-biased gene expression across reproductive and vegetative tissues in dioecious plants. In *Silene latifolia*, 17% of genes are differentially expressed in flowers, whereas <2% are differentially expressed in leaves (Zemp et al., 2016), and, in *Salix viminalis*, 43% of genes in catkins are differentially expressed, whereas only 0.09% of genes exhibit sex bias in leaves (Darolti et al., 2018). In the current study of *P. balsamifera*, we found that 36% of genes expressed in catkins showed sex bias, whereas 0.003% showed sex bias in leaves. The 100–1000-fold reduction in genes that exhibit sex-biased expression between flowers and leaves suggests that sexual dimorphism in the Salicaceae may be much more restricted to reproductive tissues than in other dioecious organisms.

The gene expression differences in floral tissues of *P. balsamifera* probably underlie both primary and secondary sexual floral dimorphisms. Because both male and female flowers consist almost solely of the androecium and gynoecium, with no petals, sepals or nectaries (Fisher, 1928), we expected much of the differential expression in flowers to represent differences in primary sexual characteristics (Sakai & Weller, 1999). However, we observed phenotypic differences in catkin color (Fig. 1) and expression dimorphism in photosynthesis-related and chemical defense (phenolic glycoside and condensed tannin) pathways, suggesting that female flowers are photosynthetically more active and better defended against herbivores than are male flowers (Hwang & Lindroth, 1997). Because neither of these traits is directly related to structures or functions specific to the androecium or gynoecium, and the transcripts are expressed in both...
sexes, but at different levels, they are classified as secondary sexual characters, which we hypothesize arose to increase sex specialization after the evolution of dioecy.

The SDR of *P. balsamifera* is small, with only 12 genes (Gerald et al., 2015; McKown et al., 2017), yet more than 11,000 genes exhibited sex-biased expression in flowers. This pattern is consistent with a small number of regulatory genes causing the downstream cascade of gene expression, resulting in both primary and secondary sexual dimorphism (Beukeboom & Perrin, 2014). The genes responsible for sex determination in *P. balsamifera* have not been definitively identified, but three of the 12 genes in the SDR (*PbrR9*, *PbCCl*C, and *PbTCP-I*) exhibited sex-biased gene expression in floral tissues in our study. *PbrR9* (Potri.019G133600) has been proposed to be a potential master regulator of sex determination in *P. balsamifera*, because it is the only gene that displays sex-biased methylation in xylem tissues (Bräutigam et al., 2017). *PbrR9* is a homolog of *AAR16* (At2G40670.1) in *Arabidopsis* (At3G56380.1), where it functions in the phosphorylation signal transduction system (Ramireddy et al., 2013) and exhibits elevated expression in stamens relative to carpels (Schmid et al., 2005; Winter et al., 2007). Using reverse transcription-polymerase chain reaction (RT-PCR), Ramirez-Carvaljal et al. (2008) detected elevated *PbrR9* expression in female catkins relative to xylem, and in xylem they detected truncated transcripts. The second candidate, *PbCCl*C (Potri.018G138100), is a homolog of *Arabidopsis* *ArCCl*C (At5G49890.1), which is a chloride channel gene important for nitrate accumulation and pH adjustment in intracellular organelles (Jossier et al., 2010), and exhibits strong expression in guard cells and pollen (Jossier et al., 2010). Moreover, chloride transport is known to be important for pollen germination and growth (Tavares et al., 2011). Finally, *PbTCP-I* (Potri.018G138200) is a homolog of a TCP-1.cpn60 chaperonin family protein in *Arabidopsis* (At5g26360.1), where it has the lowest expression in stamens and pollen (Schmid et al., 2005; Winter et al., 2007), but otherwise little is known about its function.

We found that sex-biased genes were more commonly female biased than male biased. This pattern of sex-biased expression stands in contrast with that of two insect-pollinated dioecious species, *Silene latifolia* and *Asparagus officinalis*, which exhibited more male-biased genes in flowers (Harkess et al., 2015; Zemp et al., 2016). In *Salix viminalis*, which is both wind and insect pollinated, the numbers of male- and female-biased genes were roughly even (Darolti et al., 2018). The differences in the direction of sex bias among these studies may partially result from the greater number of secondary sexual traits associated with insect attraction, which, in turn, may be more strongly influenced by sexual selection (Moore & Pannell, 2011; Delph & Herlhy, 2012) than are wind-pollinated poplar flowers. Alternatively, the greater degree of female-biased expression in *P. balsamifera* flowers may reflect the lower energetic demands and reduced benefit of chemical protection against insect herbivores in the ephemeral male flowers.

Male- and female-biased genes in *P. balsamifera* exhibited molecular evolutionary signals consistent with greater purifying selection (lower dN/dS) when compared with unbiased genes. In the dioecious *Salix viminalis*, male-biased genes had slower rates of molecular evolution than either female-biased or unbiased genes (Darolti et al., 2018). For *P. balsamifera*, this result partially reflects the high frequency of genes with fundamental functions, such as photosynthesis and mitochondrial electron transport, for which there is probably strong selection against deleterious alleles. Although Darolti et al. (2018) suggested that haploid selection of genes expressed in gametes may be responsible for stronger purifying selection and lower dN/dS, we found that the values of dN/dS for genes with pollen-limited expression did not differ from those of unbiased genes. Several studies have revealed evidence for positive selection driving the evolution of sex-biased genes in animals (Pröschel et al., 2006; Baines et al., 2008; Dorus et al., 2010; Grath & Parsch, 2012; Harrison et al., 2015), and two studies found evidence for positive selection in sex-biased genes of plants. In the hermaphroditic plant *C. grandiflora*, genes with pollen-limited expression were shown to have higher dN/dS than genes with seedling-limited expression (Arunkumar et al., 2013), and, in hermaphroditic *A. thaliana*, genes with female-limited expression had elevated rates of dN/dS (Gossmann et al., 2014) and genes expressed in pollen and pollen tubes exhibited elevated rates of amino acid substitutions (ε; Eyre-Walker & Keightley, 2009; Gossmann et al., 2014). These observations in these hermaphroditic species point to sexual selection resulting from pollen competition and pollen–pistil interactions, and it is curious that similar patterns were not detected in dioecious species.

Several possibilities exist for why no evidence for the elevated evolution of sex-biased or sex-limited genes was found in *P. balsamifera*. First, sexual and natural selection may be relatively weak forces driving the evolution of sexually dimorphic traits in poplars (McKown et al., 2017). Indeed, male- and female-biased genes had significantly lower values of dN/dS and dS than randomly chosen genes. Second, adaptation of the sex-biased expression of genes in *P. balsamifera* may have occurred largely before the divergence of *P. balsamifera* and *P. tremula*, which are both dioecious. Our comparison of dN/dS values for sex-biased and sex-limited genes in the branch shared by *P. balsamifera* and *P. tremula* yielded the same results as the comparison of only the *P. balsamifera* branch, which suggests that any signal of adaptation may be most prominent in the lineage that includes the evolution of dioecy, which occurred before the divergence of the Salicaceae. Finally, selection may be driving the evolution of sex-biased and sex-limited genes, but the target of selection may be primarily at unexpressed regulatory regions, rather than the coding region of these genes. Our studies were unable to detect patterns of evolution at regulatory regions outside of the exons.

**Conclusions**

Although the non-recombining SDR on *Populus balsamifera* chromosome 19 is small (McKown et al., 2017), we found that there were a remarkably large number of genes exhibiting sex-biased expression in flowers, indicating that the few genes in the SDR have an outsized effect on differences in gene expression between male and female flowers. Major categories of sex-biased genes in flowers were associated with secondary sexual characters, such as energetic demands and defense against herbivores. This result indicates that sexual specialization in resource allocation
has probably influenced the evolution and maintenance of dioecy in poplars. We were unable to identify a subset of genes targeted by recent positive selection that would be consistent with sexual selection. Instead, sex-biased genes exhibited an overall pattern of greater purifying selection than did unbiased genes. A similar pattern was found in S. viminalis, which probably shares a dioecious common ancestor with P. balsamifera. Finally, we confirmed a result in other species in the Salicaceae, i.e. that the pattern of sex-biased expression in flowers seems to be largely uncorrelated with the expression patterns in leaf tissues, potentially because of differences in the regulation of common gene networks, or sex- and tissue-specific selection on energetic demands. The Salicaceae are a large family with primarily dioecious members (Cronk et al., 2015), and the two most speciose genera, Populus and Salix, diverged over 45 Ma in the early Eocene (Wu et al., 2015); thus, sexual specialization has had a long time to evolve and mature. As patterns of sexual dimorphism and sex-biased expression are explored in a more diverse array of plants, we will be able to discern the patterns of sexual dimorphism in gene expression found in P. balsamifera are common among groups with ancient dioecy, or are related to other aspects of their biology or ecology, such as the relative investment by males and females in reproduction or their long lifespan.

Acknowledgements

We thank Jianquan Liu and Steve Difazio for help in accessing poplar genomes, Diana Wolf, Naoki Takebayashi, Gery Allen and Guanqiao Feng for help with photographs of poplar flowers, and Lynda Delph and four anonymous reviewers for helpful comments on the manuscript. This research was supported by Texas Tech University and National Science Foundation grants IOS-1137001 to MSO and PT, and DEB-1542599 to MSO.

Author contributions

LW, ZW, PT and MSO designed the experiment. LW, ZW and PT collected tissues and sequenced the mRNA. BJS, LW and MSO analyzed the data, interpreted the results and wrote the manuscript. PT edited later drafts. BJS and LW contributed equally to this work.

ORCID

Matthew S. Olson http://orcid.org/0000-0002-0798-145X
Brian J. Sanderson http://orcid.org/0000-0002-0310-0440
Peter Tiffin http://orcid.org/0000-0003-1975-610X
Li Wang http://orcid.org/0000-0003-2068-7535
Zhiqiang Wu http://orcid.org/0000-0002-4238-7317

References


Research 11

New Phytologist


Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article:

**Fig. S1** Multidimensional scaling (MDS) plot of expression differences among libraries.

**Fig. S2** Venn diagram comparing the genes identified in flower tissue as differentially expressed by the individual analyses in DESeq2, limma and edgeR.

**Fig. S3** Venn diagram comparing the genes identified in leaf tissue as differentially expressed by the individual analyses in DESeq2, limma and edgeR.

**Fig. S4** MA plot comparing the relationship in flower tissue between expression level (log2 counts per million (log2 CPM)) and differential expression (log2 fold change (log2 FC)).

**Fig. S5** MA plot comparing the relationship in leaf tissue between expression level (log2 counts per million (log2 CPM)) and differential expression (log2 fold change (log2 FC)).

**Fig. S6** Secondary metabolism pathways with different expression profiles in female and male flower tissues.

**Table S1** Summary of read pairs and mapping efficiency for RNA-sequencing (RNA-seq) data from leaf and flower tissues of male and female *Populus balsamifera*. 
**Table S2** Gene names and gene ontology (GO) terms for genes that exhibit patterns of sex- and tissue-limited expression.

**Table S3** Raw counts of gene expression quantified by mapping each paired-end library to the *Populus trichocarpa* v3 genome.

**Table S4** The log₂ fold change (log₂ FC) and false discovery rate (FDR) values from EDGER, LIMMA and DESEQ2 analyses for genes showing significant differential gene expression between male and female *Populus balsamifera* flowers.

**Table S5** Gene ontology (GO) terms for genes identified as differentially expressed between male and female flower tissue.

**Table S6** Enriched functional pathways identified by MapMan for sex-biased genes.

**Table S7** Enriched functional pathways identified by MapMan for genes in secondary metabolism pathways identified by Tsai *et al.* (2006).

**Table S8** Enriched functional pathways identified by MapMan for sex-limited genes.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.