# Differential Expression, Regulatory Divergence, and Sex Dimorphism Pervade the Shrub Willow (*Salix* spp.) Transcriptome

Craig H. Carlson[1*], Yongwook Choi[2], Agnes Chan[2], Michelle J. Serapiglia[1], Fred E. Gouker[1], Stephen P. DiFazio[3], Christopher D. Town[2], and **Lawrence B. Smart**[1]

[1] Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY
[2] Plant Genomics Group, J. Craig Venter Institute, Rockville, MD
[3] Department of Biology, West Virginia University, Morgantown, WV

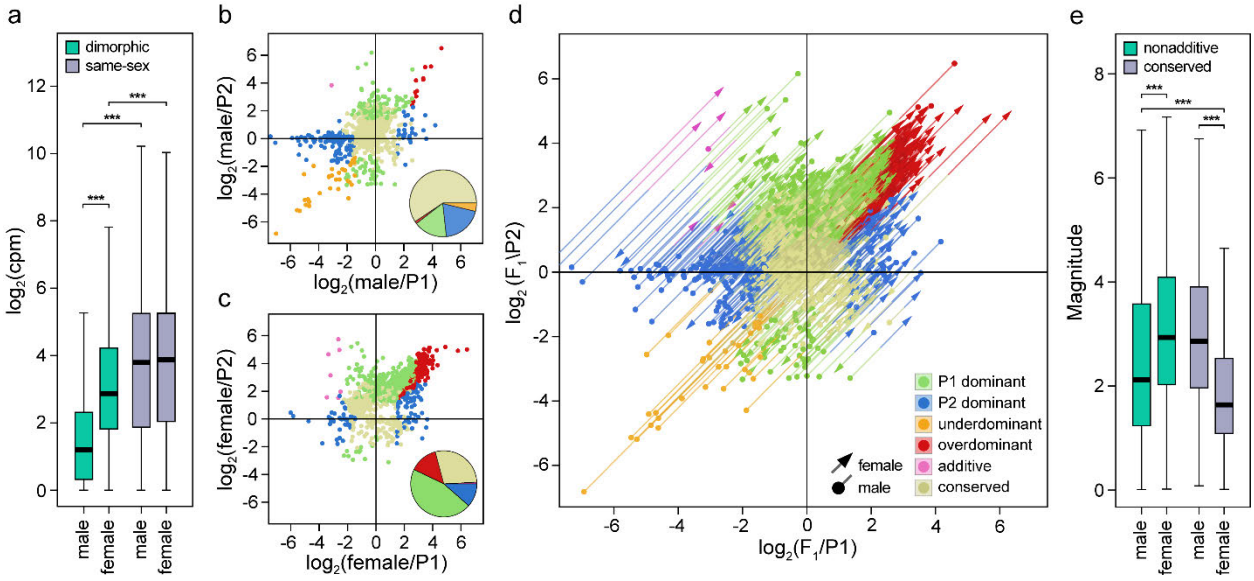
http://willow.cals.cornell.edu

**Project Goals:** Many studies have highlighted the complex, multigenic basis for heterosis (hybrid vigor) in inbred crops. Despite the lack of a consensus model, it is vital that we turn our attention to understanding heterosis in undomesticated, outcrossing and polyploid species, including shrub willow (*Salix*). A consistent trend in willow breeding is the success of triploid progeny produced from crosses between diploid and tetraploid species. We will quantify heterosis for yield and biomass traits across 8 families representing intraspecific and interspecific diploid and interspecific triploid progeny. We will quantify allele-specific gene expression and inheritance patterns in hybrid progeny using RNA-Seq. We will correlate these expression data with phenotypic characterization of heterosis for yield and biomass composition determined in replicated greenhouse and field trials. We will look for networks of coordinated gene regulation controlling yield and lignocellulosic deposition.

Recent genomic advances have provided the biomass feedstock community with new tools to improve traits related to biomass yield and wood chemical composition. In *Salix*, hybrid vigor is apparent in interspecific crosses and tends to be more pronounced in triploid progeny derived from the hybridization of tetraploid and diploid parents. Progeny and parents of full-sib intra- and interspecific $F_1$ and $F_2$ shrub willow families have been examined in order to define the basic architectures and inheritance patterns of transcriptome-wide gene expression. Our main objectives of this study were (1) to test for differential expression among the transcriptomes of segregating diploid and triploid families with regards to contrasting sex, tissue type, and midparent expression, (2) to categorize transgressive gene expression into modes of inheritance, (3) to assess the magnitude of regulatory divergent expression, and (4) to correlate modules of co-expressed genes in parents and progeny with traits important for biomass yield. We show allele-specific expression is largely conserved in intraspecific families and highly divergent in interspecific families. Akin to the heterosis observed in the field, as the complexity of the cross increases, regulatory divergent gene expression increases. In addition, we utilize a full-sib $F_2$ *S. purpurea* mapping population, planted in a replicated field trial, to supplement expression data. Candidate genes that have been or have yet to be identified will be confirmed via allele-specific assays. These data will be used to develop predictive models of heterosis and complement the growing genomic resources available for the improvement of shrub willow bioenergy crops.

**Figure 1. Magnitude of sexually dimorphic inheritance in *Salix purpurea*.**

Boxplots **(a)** summarize the $\log_2$ normalized expression differences for genes with sexually dimorphic inheritance patterns (teal) and those with same-sex inheritance (grey), by sex. Scatterplots compare $\log_2$ normalized expression of $F_1$ **(b)** males and **(c)** females to the maternal (P1, x-axis) and paternal (P2, y-axis) expression. Points represent only genes with dimorphic inheritance patterns. Pie charts within the scatterplots summarize patterns of gene expression inheritance for genes with dimorphic gene expression for each sex. The scatterplot **(d)** illustrates overlain coordinates of gene expression inheritance for males and females, where each gene is represented by male ($m_{xy}$, points) and female ($f_{xy}$, arrows) vectors, connected by a single line segment. Each segment is equally divided by two colors which correspond to the male and female inheritance class for each gene. The absolute magnitude of dimorphic gene expression inheritance was calculated for each gene as the absolute Euclidean distance ($L^2$) between $m_{xy}$ and $f_{xy}$ on the same Cartesian plane. For those genes with dimorphic inheritance, boxplot distributions **(e)** of conserved (grey) and nonadditive (teal) inheritance patterns for males and females depict differences in their absolute magnitude.




*This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0008375.*