

## SEARCHING ORPHAN GENES IN HYBRID POPLAR

Chen Hsieh<sup>1</sup>, Ran Zhou<sup>2</sup>, and Chung-Jui Tsai<sup>3</sup>

<sup>1</sup> PhD student, Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA  
(Chen.Hsieh@uga.edu);

<sup>2</sup> Post-Doctoral Associate, Warnell School of Forestry and Natural Resources, University of Georgia,  
Athens, GA 30602, USA;

<sup>3</sup> Professor, School of Forestry and Natural Resources, Department of Plant Biology, Department of  
Genetics and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA

Orphan genes by definition have restricted taxonomic occurrence. They don't share sequence similarity to other known genes in the bioinformatic databases, thus making it difficult to predict their biological function using traditional approaches. Since orphan genes are unique to specific taxa, they may confer adaptive significance during evolution. We focus on the *Populus tremula* × *P. alba* INRA 717-1B4 hybrid (hereafter '717'), an important study system in poplar functional genomics because of its high transformation efficiency. To identify protein-coding orphan genes, a total of 15 genomes were selected, including well annotated model species and those closely related to our focal species for analysis. Orthogroups were constructed according to the amino acid sequence similarity among the selected proteomes. Genes belonging to orthogroups that only contain *Populus* sequences as well as those which cannot be assigned to any orthogroups were categorized as putative orphan candidates. Collinearity of certain 717-specific orphan genes can be observed between the two subgenomes in the current 717 draft genome assembly. The candidate orphans showed several properties, including shorter coding sequences and higher isoelectric points. Among these orphan genes, seven showed interesting expression patterns in different tissues and under drought stress treatment. Three genes showed higher expression in xylem. Two genes were only expressed in callus and the predicted polypeptides are shorter than 100 amino acids. The other two genes were induced under drought stress and were only expressed in roots. In addition, they were all predicted to contain intrinsically disordered regions. We hypothesize that these candidate orphans may act as chaperone proteins and contribute to the fitness of '717'.