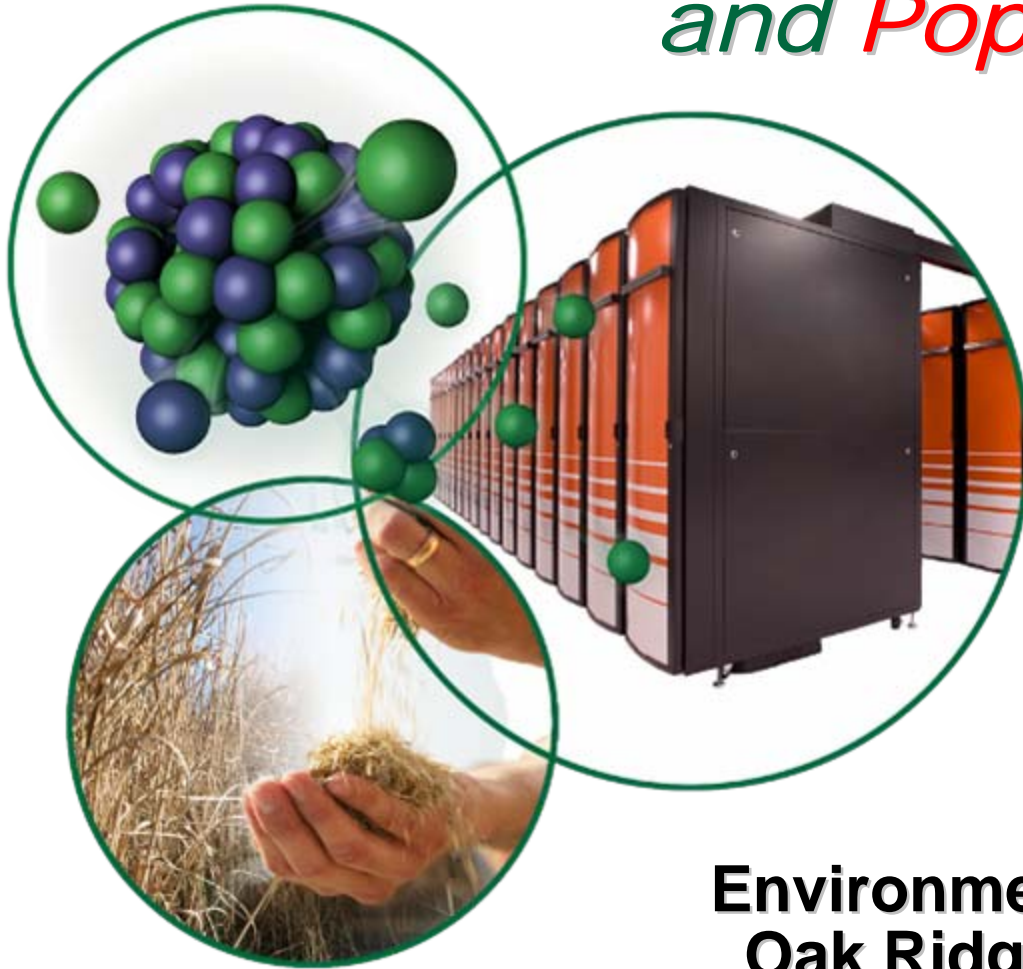


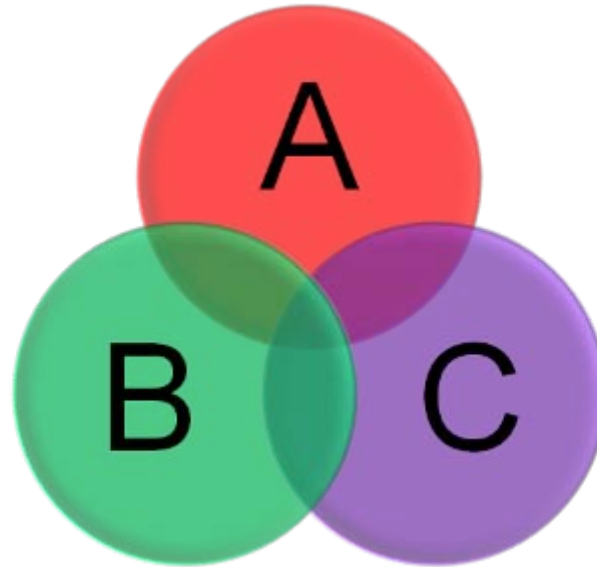
# *Genome-wide Identification of Lineage Specific Genes in Arabidopsis, Oryza and **Populus***



**Xiaohan Yang  
Sara Jawdy  
Timothy Tschaplinski  
Gerald Tuskan**

**Environmental Sciences Division  
Oak Ridge National Laboratory  
USA**

*Identification of taxon-specific genes has both scientific and practical value*



**Conserved function (Common) ←-----→ New function (Unique)**

# Plant genome sequences

## Eudicot

## Monocot

### Eurosid II

### Eurosid I



*Arabidopsis*

*Carica*

*Medicago*

*Glycine*

*Populus*

*Vitis*

*Zea*

*Sorghum*

*Oryza*

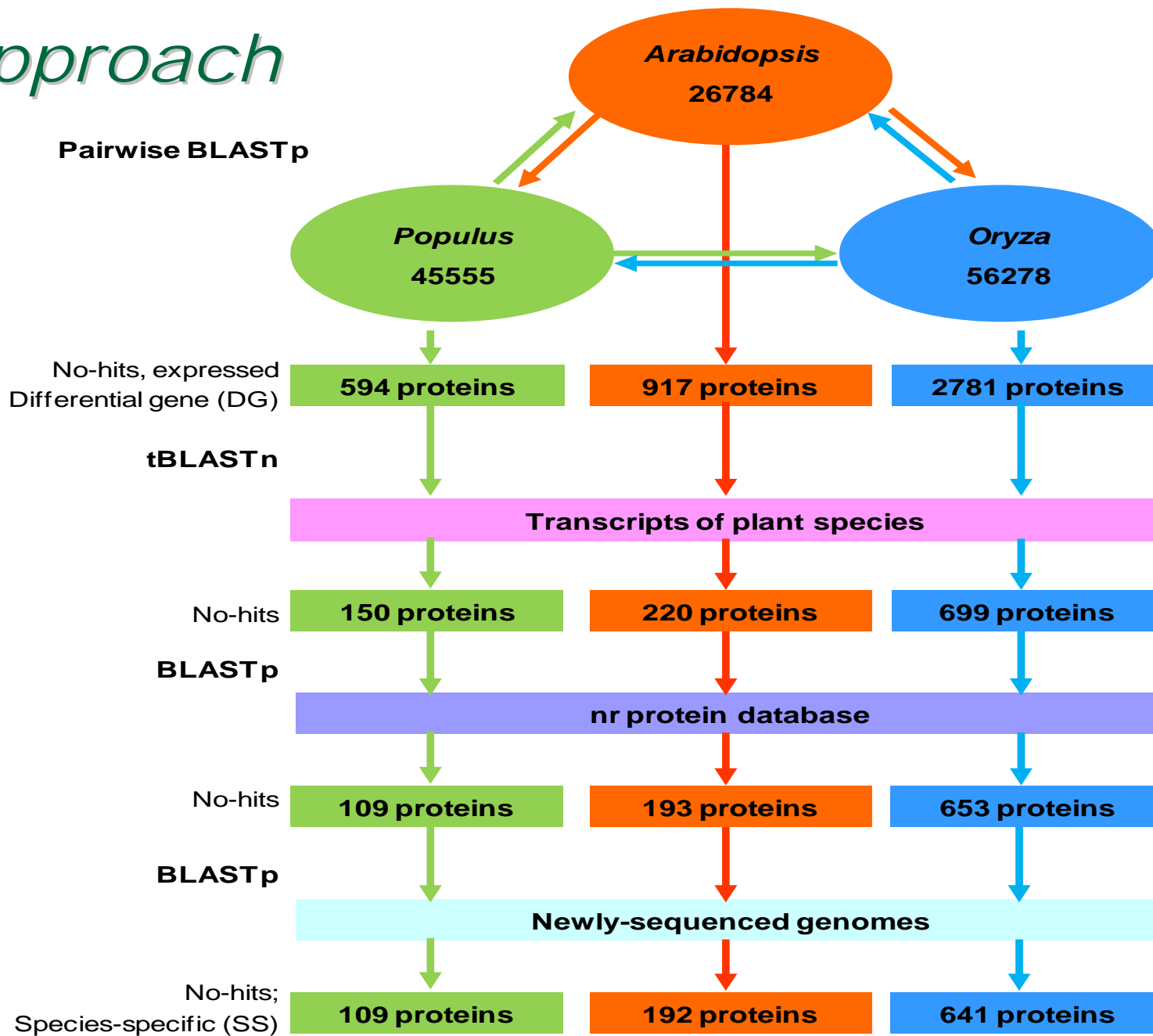
# Objectives

- *To identify lineage-specific genes in reference plant species*
- *To explore functions of the specific genes via expression analysis*
- *To study the evolution of the lineage-specific genes*

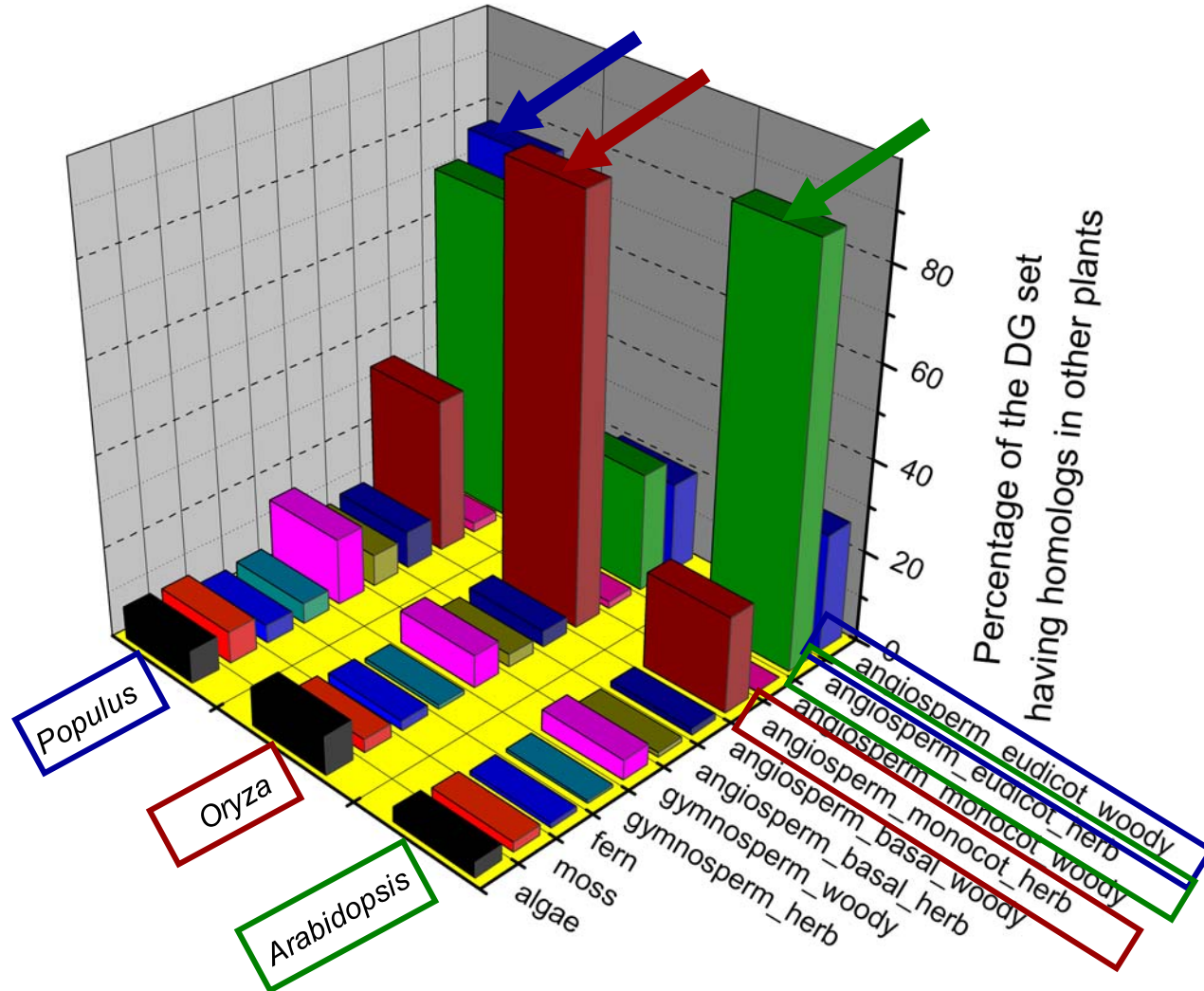


*Application of the knowledge gained to bioenergy and carbon sequestration research  
(Candidate genes)*

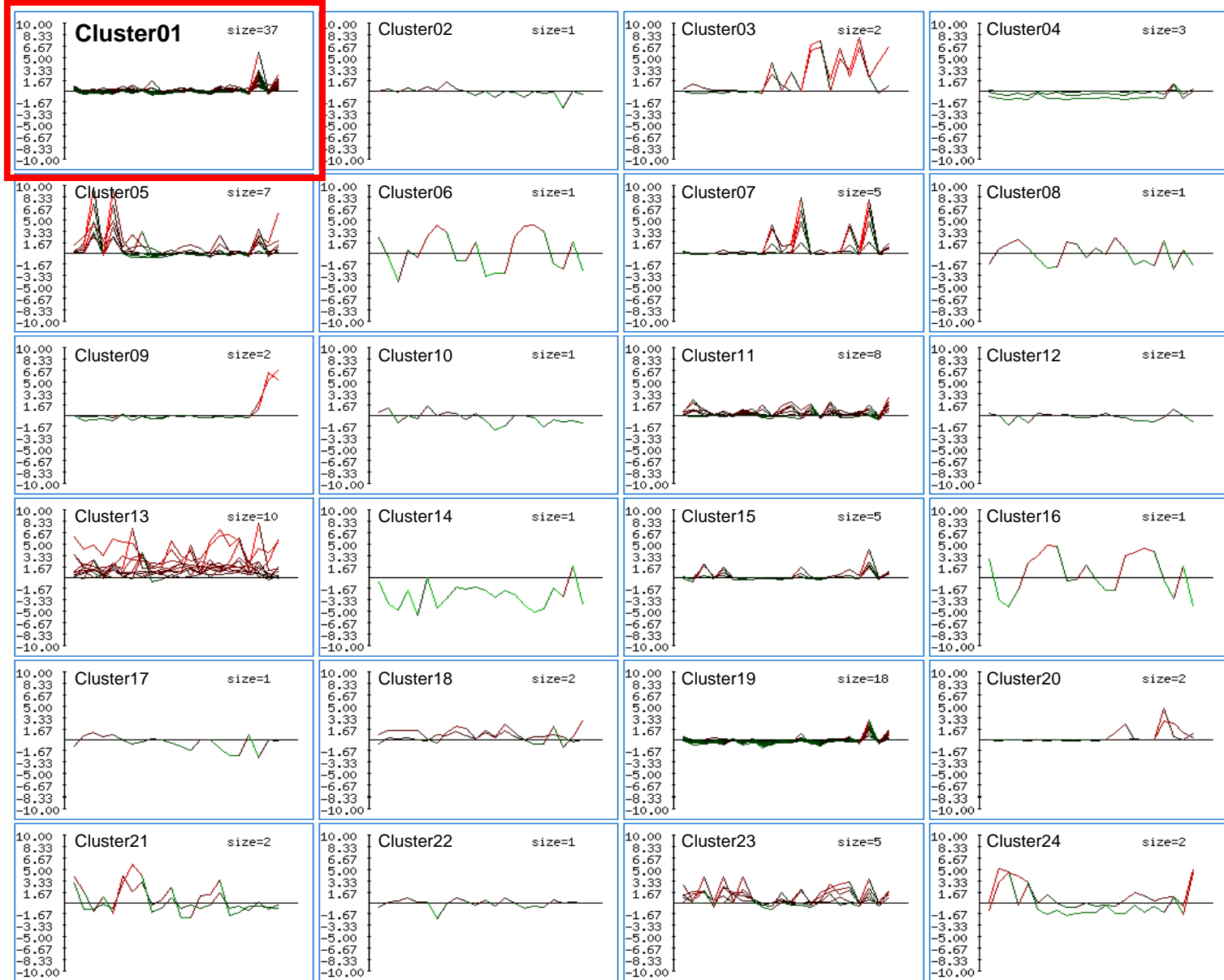
# Approach



# The percentage of differential genes showing homology to other plants

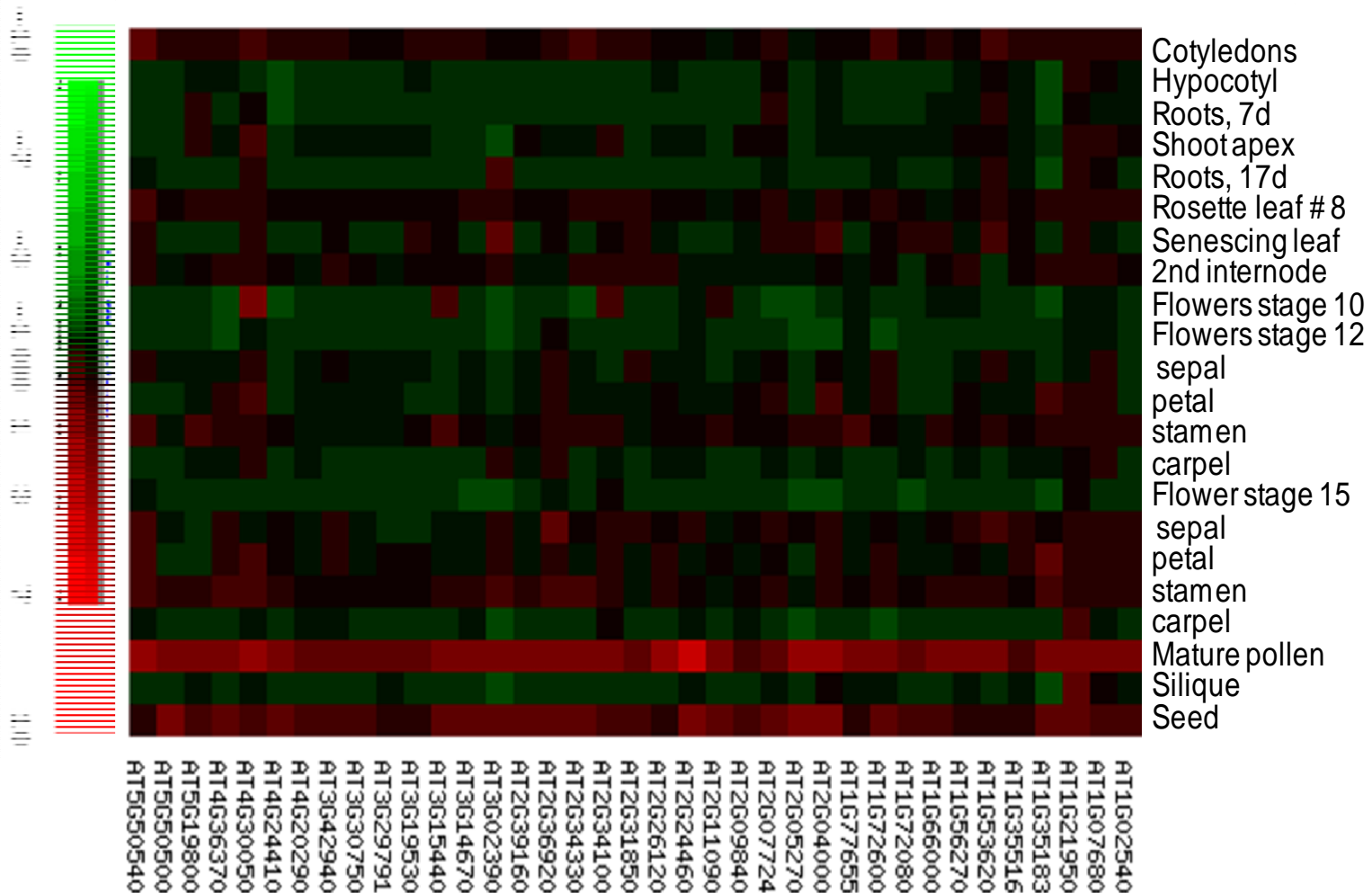


# Expression of the Arabidopsis SS

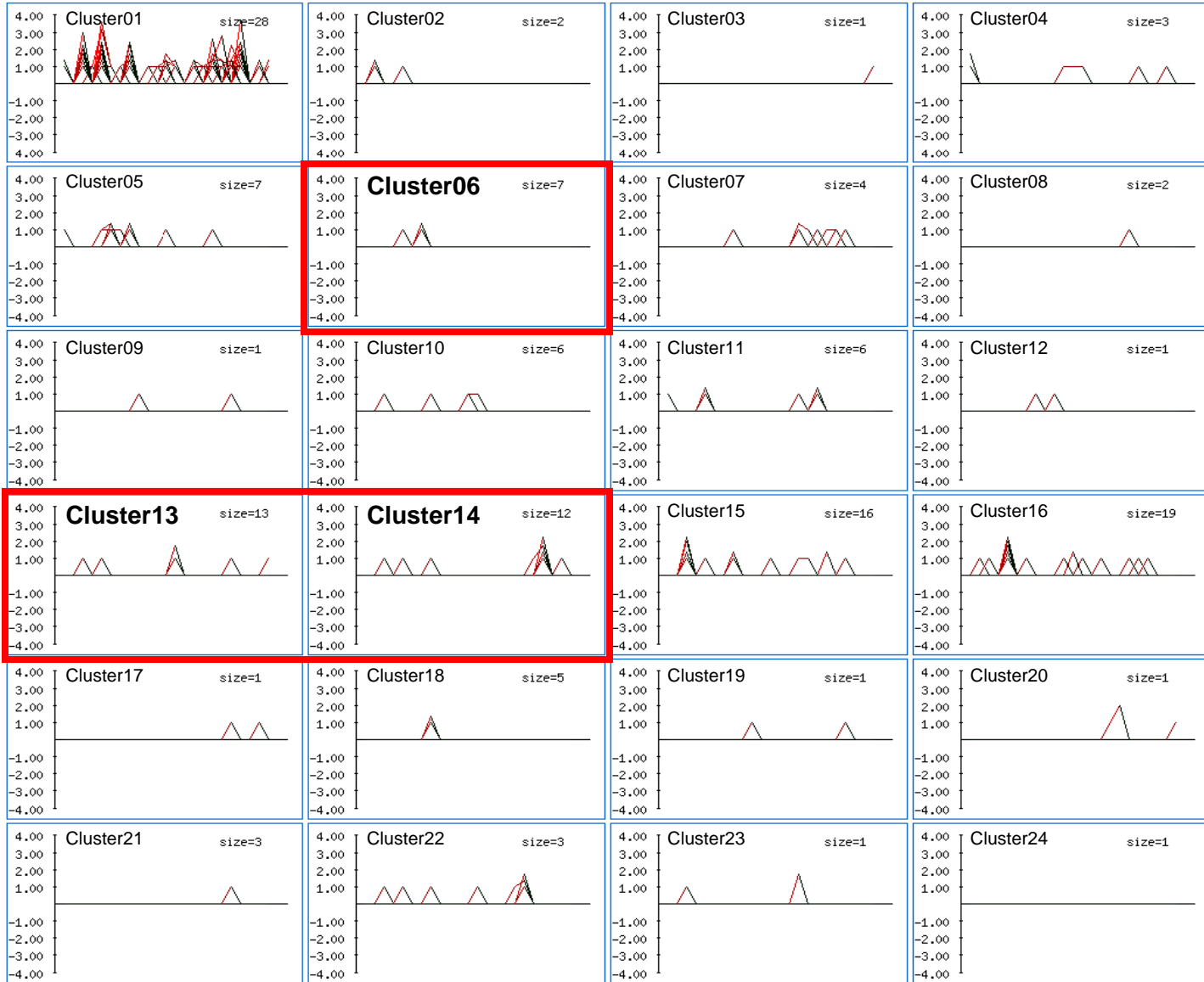


Cotyledons  
 Hypocotyl  
 Roots, 7d  
 Shoot apex  
 Roots, 17d  
 Rosette leaf # 8  
 Senescing leaf  
 2nd internode  
 Flowers stage 10  
 Flowers stage 12  
 sepal  
 petal  
 stamen  
 carpel  
 Flower stage 15  
 sepal  
 petal  
 stamen  
 carpel  
 Mature pollen  
 Silique  
 Seed

# Expression of the Arabidopsis SS (Cluster01)



# Expression of the Populus SS genes

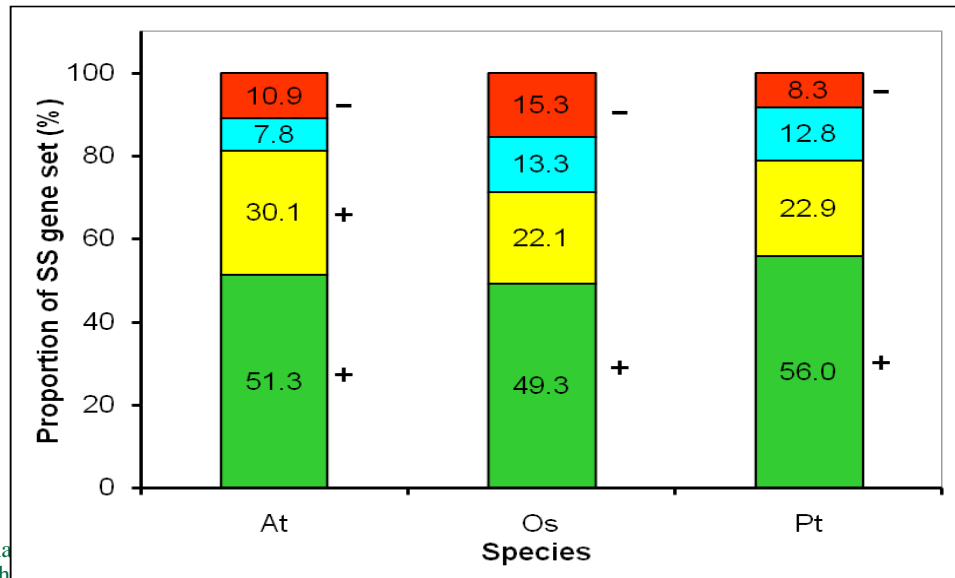
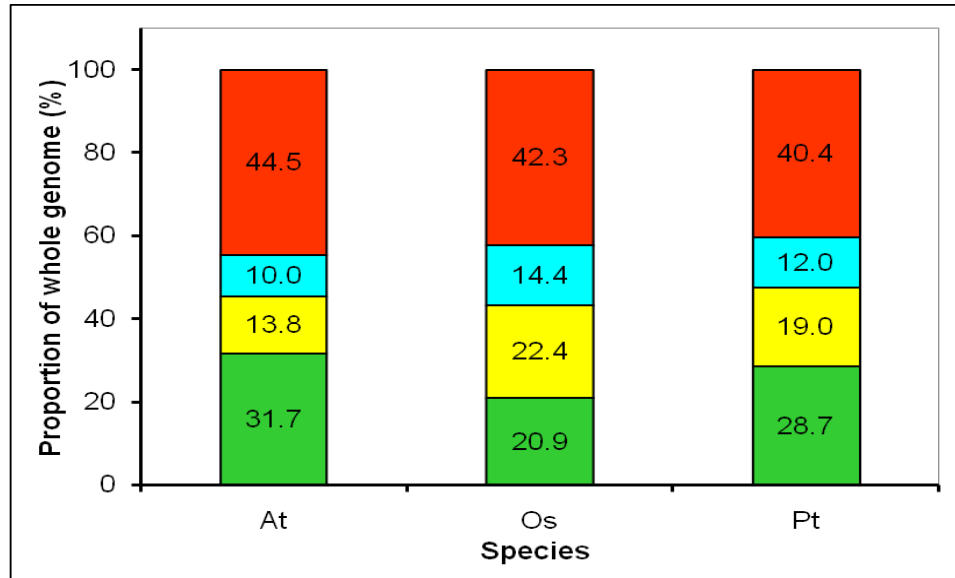


Bark  
 Buds\_dormant  
 Cambium  
 Cambium\_dormant  
 Flower\_buds  
 Flower\_catkins\_female  
 Flower\_catkins\_male  
 Leaf  
 Leaf\_cold\_stressed  
 Leaf\_senescing  
 Leaf\_virus\_infected  
 Petioles  
 Root  
 Root\_cap  
 Seed\_imbibed  
 Shoot  
 Shoot\_tip  
 Stem  
 Suspension\_cell  
 Xylem  
 Xylem\_Phloem  
 Xylem\_Phloem\_primary  
 Xylem\_Phloem\_secondary

# Expression of the *Populus* SS genes



# Number of introns per gene in Arabidopsis, Oryza & Populus



- Intron number per gene:
  - 3-or-more introns*
  - 2 introns*
  - 1 intron*
  - 0 intron (Intronless)*
- “+” indicates that genes are over-represented and “-” under-represented at  $P < 1 \times 10^{-5}$ , as compared with all genes in the whole genome.

# *qPCR analysis of expression of Populus-specific & differential gene sets*

Gene ID	bark	SEM	leaf	SEM	root	SEM	shoot	SEM	stem	SEM
SS01	130.6	39.9	14.2	3.8	140.0	38.6	1.0	0.6	91.8	29.8
SS02	1.1	0.0	2.4	0.4	1.5	0.2	1.1	0.1	1.0	0.2
SS03	1.0	0.3	12.2	2.7	1.5	0.2	1.6	0.4	7.6	1.6
SS04	1.0	0.1	4.3	0.4	1.6	0.2	1.2	0.1	2.5	0.3
SS05	1.4	0.2	3.5	0.3	1.2	0.1	2.5	0.4	1.0	0.1
SS06	1.7	0.4	78.7	15.4	6.8	0.6	1.0	0.2	31.8	5.7
SS07	1.0	0.4	27.5	4.5	11.7	1.8	7.6	0.0	10.8	1.9
SS08	2.2	0.1	9.1	1.2	9.4	0.9	1.0	0.1	14.6	2.1
SS09	7.4	0.9	2.8	0.4	19.8	4.9	12.5	1.8	1.0	0.2
SS10	1.6	0.4	11.2	2.7	1.0	0.2	5.1	1.1	0.0	0.0
SS11	18.2	3.3	4.8	0.3	2.1	0.5	1.0	0.2	0.0	0.0
SS12	1.2	0.1	5.3	0.5	1.0	0.1	2.4	0.2	1.8	0.2
SS13	1.4	0.1	5.5	0.5	1.9	0.2	1.0	0.1	2.0	0.2
SS14	3.1	0.8	11.2	1.7	1.0	0.2	2.0	0.4	2.8	0.6
SS15	1.4	0.2	2.8	0.2	1.6	0.2	1.3	0.1	1.0	0.1
DG01	6.0	1.1	95.7	31.9	1.0	0.3	6.2	0.7	16.8	2.6
DG38	1.0	0.2	18.3	2.3	11.3	1.3	6.6	0.0	10.0	1.5

# Conserved motifs in Arabidopsis- and Oryza-specific proteins

Motif ID	Motif consensus	Genes
<b>Motif 1</b>	D[E]-E-E-E-E-E-E-E[D]-E[R]-D[E]-D[E]-D[ELR]-E[R]	9 genes from <i>Arabidopsis</i> specific group 1
<b>Motif 2</b>	E-K[R]-x-E[R]-E[R]-G[R]-E-E[K]-E-K[E]-E-E-E-x-D[E]	22 genes from <i>Oryza</i> specific group 1
<b>Motif 3</b>	L[V]-R-F-N-D-E[K]-P[L]-R-G-D[NL]-L[PQS]-L[S]-L[W]-S[K]-P[QS]-V[G]-M[ET]-A[FHLP]-T[HP]-P[QT]-K[IN]	5 genes from <i>Oryza</i> specific group 1
<b>Motif 4</b>	H[D]-H-H-H-R[CG]-H-H-H	8 genes from <i>Oryza</i> specific group 1
<b>Motif 5</b>	K[ERS]-G[KN]-E[GIKL]-D[NRI]-S[KNQ]-D[KMT]-D[KE]-D[LS]-D[KLY]-D[EGQ]-S[TV]-D[FVY]-G[HSTW]-S[DK]-S[DNW]-N[GTV]-D[EKNV]-D[NG]-L[NSY]-D[EFG]-S[GT]-D[Q]-D[HY]-D[FL]-D[EKMS]-D[T]-A[DKSV]-M[DNS]-K[HLM]-D[CHK]-K[DH]-I[LD]-S[FGT]-D[ES]-L[SQ]-F[DHY]-K[HI]-D[EHLV]-K[CGS]-D[I]	5 genes from <i>Oryza</i> specific group 1
<b>Motif 6</b>	E[K]-E[R]-R[EG]-E[GL]-E[DR]-D[E]-G[S]-R[EK]-K-K[E]-K[L]-E[DK]-E[R]-E-K-E[EKRT]-K	4 genes from <i>Oryza</i> specific gene expression cluster 18

# Other species sharing motifs with Arabidopsis- and Oryza-specific genes

Motif	Organism		Protein	
<b>Motif 01</b>	Eukaryota; Metazoa	6	hypothetical protein	5
	Eukaryota; Alveolata	3	unknown protein	1
	Eukaryota; Fungi	1	Hmgb1-prov protein	1
			Casein kinase II $\beta$ subunit	1
			alpha 2B adrenergic receptor	1
			MGC85274 protein	1
<b>Motif 02</b>	Eukaryota; Metazoa	9	hypothetical protein	9
	Archaea; Euryarchaeota	1	melanoma antigen family A	1
<b>Motif 03</b>	Bacteria; Proteobacteria	1	hypothetical protein	1
	Viruses; dsDNA viruses	9	lef-8	9
<b>Motif 04</b>	Bacteria; Firmicutes	1	hypothetical protein	5
	Eukaryota; Alveolata	2	histidine-rich protein	2
	Eukaryota; Metazoa	4	unknown protein	1
	Bacteria; Proteobacteria	2	SJCHGC08151 protein	1
	Eukaryota; Mycetozoa	1	proline-rich region	1
<b>Motif 05</b>	Eukaryota; Alveolata	5	hypothetical protein	8
	Eukaryota; Mycetozoa	4	MIF4G domain protein	1
	Eukaryota; Fungi	1	Adenylate cyclase catalytic	1
<b>Motif 06</b>	Eukaryota; Metazoa	6	hypothetical protein	6
	Bacteria; Firmicutes	1	LOC566027 protein	1
	Archaea; Crenarchaeota	1	LRRGT00096	1
	Eukaryota; Viridiplantae;	2	unknown protein (1), lti45 (1)	2

# Summary

- *192, 641 and 109 species-specific (SS) genes were discovered in Arabidopsis, Oryza and Populus, respectively*
- *Some SS genes were preferentially expressed in flowers, roots, xylem (and cambium or up-regulated by stress, data not shown)*
- *Six conserved motifs in Arabidopsis and Oryza SS proteins were found in other distant lineages*
- *The SS gene sets were enriched with intronless genes*
- *The Populus-specific genes are likely candidates for carbon sequestration & biofuel*



# Acknowledgements

*Gerald Tuskan, Sara Jawdy, Timothy Tschaplinski  
Stan Wullschleger, Udaya Kalluri, Lee Gunter*

*<http://www.esd.ornl.gov/PGG/index.htm>*

