

# The monosaccharide transporter gene family in land plants shows differential subfamily expression and expansion across lineages

Deborah A. Johnson, Jeffrey P. Hill and Michael A. Thomas, Dept. of Biological Sciences, Idaho State University, Pocatello, ID

## Introduction

Gene families evolve through a combination of tandem, segmental and whole genome duplication (polyploidy) events. The classical model of gene duplication and divergence proposed that, after duplication, one duplicate maintains function, allowing the other to diverge [1]. Recent theoretical and empirical studies suggest that duplicate genes diverge predominantly through the partitioning of expression [2, 3] and that breadth of gene expression is related to the rate of protein sequence evolution [4]. Here, we utilize expressed sequence tag (EST) data to study gene duplication and expansion patterns in the monosaccharide transporter (MST) gene family across the land plants.

In plants, MSTs are integral membrane proteins that transport various sugars across the membranes that separate cells (plasma membrane) and cell compartments (tonoplast, plastid) [5, 6]. In *Arabidopsis*, there are 53 MST genes that form seven distinct subfamilies (Figure 1).

We created profile hidden Markov models (HMMs) of each subfamily and searched EST databases representing diverse land plant lineages to address the following questions:

- Are homologs of each *Arabidopsis* subfamily present in the earliest land plants?
- Do expression patterns among subfamilies and individual genes within subfamilies differ across lineages?
- Has gene duplication resulted in lineage-specific subfamily expansion patterns?
- Is there a correlation between highly expressed *Arabidopsis* genes and similarity to orthologs in early lineages?

## Methods

One hundred twenty MST protein sequences from across the green plant lineage was assembled, aligned and subject to phylogenetic analysis. Profile HMMs were then produced from subsets of the data corresponding to each subfamily. Nine EST databases (*Marchantia polymorpha*, *Physcomitrella patens*, *Physcomitrella patens* subsp. *patens*, *Selaginella lepidophylla*, *Ceratopteris richardii*, *Pinus taeda*, *Zea mays*, *Arabidopsis thaliana*, and *Lycopersicon esculentum*) were searched using the Wise2 package and the profile HMM-based consensus sequences for each MST subfamily. ESTs with a significant match to the profile HMMs ( $\leq 1e-10$ ) were then matched against the *Arabidopsis* protein reseq database (BLASTN) to determine best-match MST genes. Cap3 was then used to create unique EST contigs. The contigs and remaining singlets which had significant overlap (>50 amino acids) and significant sequence differences were inferred to represent unique loci in each EST database.

## Results

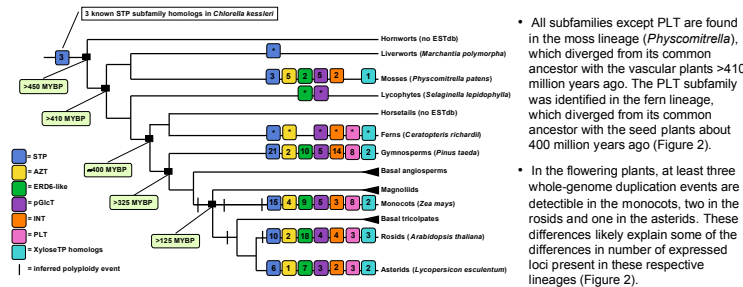
### Patterns of subfamily expression and expansion across lineages

- Relative subfamily EST database representation varies across lineages, with greater expression of the STP and ERD6-like subfamilies in the gymnosperm and flowering plant lineages and relatively high expression of the AZT and pGlcT subfamilies in the moss lineage.
- Lineage-specific subfamily expansion patterns are present that often appear to be due to multiple gene duplications (probably tandem) arising from a single ancestral gene.
  - In *Pinus*, the STP (21), ERD6-like (10), INT (14), and PLT (8) subfamilies are expanded.
  - In *Zea*, the STP (15), ERD6-like (9), and PLT (8) subfamilies are expanded.
  - In *Lycopersicon*, the STP (6) and ERD6-like (7) subfamilies appear to be slightly expanded.

Table 1. Summary of EST database search results.

Taxon	ESTs in database	Tissues	ESTs showing significant homology to the MST subfamily							Total MST ESTs	
			STP	AZT	ERD6-like	pGlcT	INT	PLT	XyloseTP		% MST ESTs
<i>Marchantia</i> (liverwort)	1,415	Sex organs	2							2	0.14%
<i>Physcomitrella</i> (moss)	140,617	Multiple tissues/stages	14	54	5	54	11		6	144	0.10%
			9.3%	37.5%	3.2%	37.5%	7.6%		4.2%	15	
<i>Selaginella</i> (lycophyte)	1,046	Desiccated frond	3	5	2	5	2		1	18	0.19%
<i>Ceratopteris</i> (fern)	5,085	Germinating spore	1	1		1	1		1	6	0.12%
<i>Pinus</i> (gymnosperm)	291,588	Multiple tissues/stages	83	8	36	64	58	41	13	303	0.10%
			27.4%	2.6%	11.9%	21.1%	19.1%	13.5%	4.3%	62	
<i>Zea</i> (monocot angiosperm)	417,803	Multiple tissues/stages	19,616	16,326	19,959	13,116	12,225	15,176	3,926	112	0.07%
			4.7%	3.9%	4.8%	3.1%	2.9%	3.6%	0.9%	46	
<i>Lycopersicon</i> (dicot angiosperm - asterid)	189,735	Multiple tissues/stages	27	3	33	8	4	10	8	83	0.05%
			1.4%	0.2%	1.7%	0.4%	0.2%	0.5%	0.4%	24	
<i>Arabidopsis</i> (dicot angiosperm - rosid)	415,250	Multiple tissues/stages	507	41	352	71	32	111	41	1155	0.28%
			12.2%	1.0%	8.5%	1.7%	0.8%	2.7%	1.0%	44	

Figure 2. Land plant phylogeny of major clades showing approximate dates of divergence, presence of inferred polyploidy events and the number of expressed MST subfamily loci in the moss, gymnosperm, monocot, rosid and asterid lineages inferred from EST data.



- All subfamilies except PLT are found in the moss lineage (*Physcomitrella*), which diverged from its common ancestor with the vascular plants >410 million years ago. The PLT subfamily was identified in the fern lineage, which diverged from its common ancestor with the seed plants about 400 million years ago (Figure 2).
- In the flowering plants, at least three whole-genome duplication events are detectable in the monocots, two in the rosids and one in the asterids. These differences likely explain some of the differences in number of expressed loci present in these respective lineages (Figure 2).

## Arabidopsis MST gene function and EST database representation

Figure 3. EST database representation of MST genes in *Arabidopsis thaliana*. Branches with yellow highlighting indicate the presence of ESTs in the database of 415,250 ESTs. Small callouts show the total number of ESTs with a best match to each indicated gene and the percentage of total subfamily ESTdb representation. Large callouts display known function and expression of characterized genes. Red \*s indicate genes with best match homologs in at least one early lineage (liverwort, moss, lycophyte, fern and/or gymnosperm).

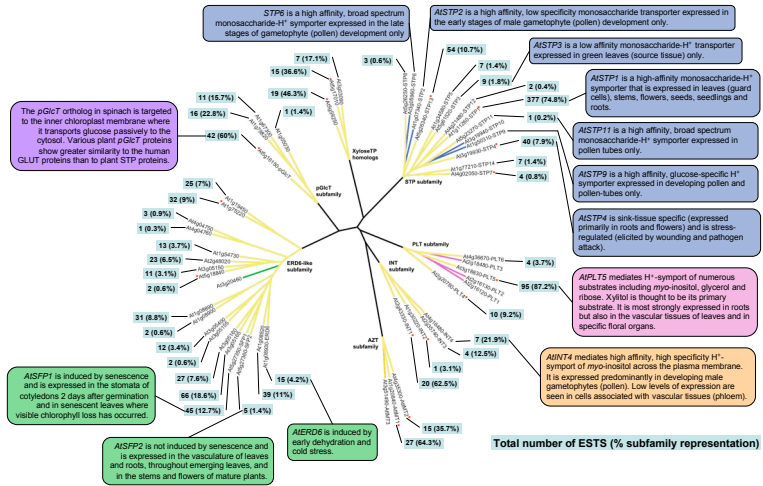
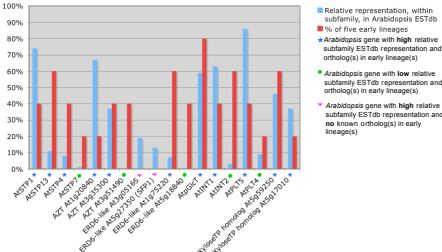


Figure 4. Relationship between *Arabidopsis* MST genes with high EST database representation and the presence of best-match orthologs in five early land plant lineages (liverwort, moss, lycophyte, fern, and/or gymnosperm).



## Correlation of expression with orthologs in early lineages

- Eleven of 13 (85%) *Arabidopsis* MST genes with high ESTdb representation have one or more best-match orthologs in the five early divergent lineages in this study (liverwort, moss, lycophyte, fern and gymnosperm) (Figure 4).
- Four *Arabidopsis* genes with low ESTdb representation are the best match to orthologous genes in early divergent lineages (Figure 4).
- Two *Arabidopsis* genes with high relative subfamily ESTdb representation do not have best-match orthologs in early divergent lineages. These two genes are members of a subclade of tandem duplicates in the ERD6-subfamily in which representation is more evenly spread among members than in other subfamilies. (Figure 4)

## Conclusions

- All seven subfamilies of the *Arabidopsis* MST gene family were present in the land plants at least 400 million years ago [7].
- MST subfamily ESTdb representation patterns reveal that the AZT and pGlcT subfamilies have higher relative representation in the early divergent moss and the STP and ERD6-like subfamilies have higher relative representation in the flowering plants [7].
- Lineage-specific subfamily expansions have resulted in large expansion of the STP and ERD6-like subfamilies in *Arabidopsis*, of the STP, ERD6-like, INT and PLT subfamilies in *Pinus taeda* (loblolly pine), of the STP, ERD6-like, and PLT subfamilies in *Zea mays* (corn), and in small expansion of the STP and ERD6-like subfamilies in *Lycopersicon esculentum* (tomato) [7].
- 85% of *Arabidopsis* MST genes with high representation in the EST database have best-match homologs in early lineages, suggesting that genes with high expression are usually highly conserved [7].

## References

- Ohno S. *Evolution by Gene Duplication*. Springer-Verlag; 1970.
- Force A, Lynch M, Pickett FB, Amores A, Yan Y-L, Postlethwait J. *Preservation of Duplicate Genes by Complementary, Degenerative Mutations*. *Genetics* 1999, 151(4):1531-1545.
- Gu Z, Nicolae D, Lu H-H, Li W-H. *Rapid divergence in expression between duplicate genes inferred from microarray data*. *Trends Genet* 2002, 18(12):608-613.
- Duret L, Mouchiroud D. *Determinants of Substitution Rates in Mammalian Genes: Expression Pattern Affects Selection Intensity but Not Mutation Rate*. *Mol Biol Evol* 2000, 17(11):68-70.
- Buttner M, Sauer N. *Monosaccharide transporters in plants: structure, function and physiology*. *Biochim Biophys Acta* 2000, 1465:263-274.
- Williams LE, Lemoine R, Sauer N. *Sugar transporters in higher plants - a diversity of roles and complex regulation*. *Trends Plant Sci* 2000, 5(7):283-290.
- Johnson DA, Hill JP, Thomas MA. *The monosaccharide transporter gene family in land plants is ancient and shows differential subfamily expansion and expansion across lineages*. *BMC Evol Biol* 2006 (in press).

## Acknowledgements

This study was made possible by NIH Grant #P20 RR061454 from the INBRE Program of the National Center for Research Resources and funding from the PhRMA Foundation. All bioinformatic workflows were performed in the Evolutionary, Ecological and Environmental Genomics Group Laboratory on the Apple Xserve cluster. We thank Luobin Yang for writing custom Perl scripts, troubleshooting and technical support.