

Distribution of ABC Transporter Genes across the Plant  
Kingdom

A Thesis Presented for the  
Master of Science  
Degree  
The University of Tennessee, Knoxville

Thomas Scott Lane  
May 2015

Copyright © 2015 by Thomas Lane  
All rights reserved.

# **DEDICATION**

To my Brother Joseph Lane

## **ACKNOWLEDGEMENTS**

Thanks to the Neal Stewart Lab. Thanks to Dr. Harry Richards, Dylan Storey, and Dr. Meg Staton for your support. Thanks to Dr. Amy Lawton-Rauh for pushing me to seek a graduate level education.

## ABSTRACT

The ATP-binding cassette (ABC) transporter gene superfamily is ubiquitous among extant organisms. ABC transporters act to transport compounds across cellular membranes and are involved in a diverse range of biological processes and functions including cancer resistance in humans, drug resistance among vertebrates, and herbicide resistance in weeds. This superfamily of genes appears to be larger and more diverse in the plant kingdom—yet, we know relatively less about ABC transporter function in plants compared with mammals and bacteria. Therefore, we undertook a plant kingdom-wide transcriptomic survey of ABC transporters to better understand their diversity.

We utilized sequence similarity-based informatics techniques to deduce likely ABC transporter genes from 1,301 phylogenetically-diverse plant transcriptomes. A total of 97,149 putative ABC transporter gene members were identified with an average of 88 gene members per RNASeq library and a standard deviation of 30 gene members. Of 97,149 ABC transporter sequences identified, 22,343 were full length. Fewer unique ABC transporter gene members appeared among algae compared with vascular and non-vascular land plants. Differences were also noted in abundance of certain ABC transporter subfamilies. Non-algae taxa had relatively more unique ABCB, ABCC, and ABCG transporter gene members on average ( $p < 0.005$ ), and both algae and early non-vascular plants averaged significantly more ABCF transporter gene members ( $p < 0.005$ ). Ferns had significantly fewer ABCA transporter gene members than most plant taxa ( $p < 0.005$ ).

Our study provides an overview of ABC transporter protein gene members among many plant taxa including: algae, non-vascular land plants, lycophytes, ferns, conifers, and angiosperms. An increase in the number of gene family members present in the ABCB, ABCC, and ABCD transporter subfamilies may indicate greater variability of the ABC transporter protein superfamily among plant taxa since the divergence of non-vascular plants from algae. We conclude that the number of ABCD, ABCE, and ABCI subfamily proteins gene members has remained relatively unchanged throughout the plant kingdom. The striking difference between the number of ABCA subfamily transporter protein gene members between ferns and other plant taxa is surprising and merits further investigation.

# TABLE OF CONTENTS

CHAPTER I Introduction .....	1
ABC Transporter Protein Function .....	1
ABC Transporter Protein Structure .....	2
Plant ABC Transporter Nomenclature.....	2
One Thousand Transcriptomes Project (1KP) .....	5
CHAPTER II Methods.....	8
Data acquisition and quality control .....	8
HMMER domain e-value selection .....	8
Comparison of genome and transcriptome data sources.....	9
ABC transporter classification.....	9
Nonparametric Comparison Using Wilcoxon Method .....	9
CHAPTER III .....	11
Results and discussion .....	11
Pfam HMMER domain e-value .....	11
Comparison of genome and transcriptome data sources.....	12
ABC transporter subfamily classification.....	17
CHAPTER IV Conclusions .....	25
LIST OF REFERENCES .....	27
VITA .....	32

## LIST OF FIGURES

Figure 1. Flax ABC transporter unique gene distribution deduced by RNA-Seq transcriptome data.....	14
Figure 2. Castor bean ABC transporter unique gene distribution deduced by RNA-Seq transcriptome data.....	15
Figure 3. Number of unique ABC transporter gene members distributed over plant taxa deduced by RNA-Seq transcriptome data. ....	16
Figure 4. Number of unique ABC transporter gene members of subfamilies B, C, and G distributed over plant taxa deduced by RNA-Seq transcriptome data.....	19
Figure 5. Number of unique ABC transporter gene members of subfamilies D, E, and I distributed over plant taxa deduced by RNA-Seq transcriptome data.....	21
Figure 6. Number of unique ABC transporter gene members of subfamilies F and A distributed over plant taxa deduced by RNA-Seq transcriptome data. ....	23

# CHAPTER I

## INTRODUCTION

### **ABC Transporter Protein Function**

The ATP-binding cassette (ABC) transporter family is one of the largest known and most diverse protein superfamilies (Higgins 1992, Jones *et. al.* 2003), and all living organisms are thought to contain ABC proteins (Jones *et. al.* 2003). Shared among members within the ABC transporter protein superfamily is the ability to hydrolyze adenosine triphosphate (ATP), which is used in a wide array of functions, including: DNA repair, RNA translocation, and most commonly, active transport of substrates across membranes (Davidson *et. al.* 2008). Most ABC transporters can be classified as either exporters or importers. Both importers and exporters can move a large variety of substrates across membranes. Export serves a number of roles including detoxification. The detoxification mechanism is of interest in weedy plant herbicide resistance studies, particularly those involving putative glyphosate sequestration into vacuoles (Peng *et. al.* 2010, Ge, X. *et. al.* 2010, Hanson *et. al.* 2009). Export proteins are also involved in environmental response and plant development, such as auxin exporters ABCB19 and ABCB1 (Yang *et. al.* 2009). Importers primarily function in the acquisition of substrates important for cellular activity and ultimately the plant's survival. Such is the case of AtPMP2, which imports critical substrates needed in the glyoxylate cycle during germination (Theodoulou *et. al.* 2006).

## **ABC Transporter Protein Structure**

ABC transporters have three structural types. Full transporters are composed of two transmembrane domains (TMD) and two nucleotide-binding domains (NBD). Half transporters are composed of one TMD and one NBD, and dimerize in pairs to create virtual full transporters as homodimers or heterodimers. A third type of transporter has no TMDs but two NBDs (Verrier *et. al.* 2008). The NBD is present in all three structural types and contains many key conserved motifs: Walker A, Q-loop, Walker B, D-loop, switch H-loop, and a signature motif. Functionally, all of these components interact with ATP with the exception of the D-loop, which primarily functions in holding dimers together, and the switch H-loop, which interacts with the transmembrane domain. The Walker A and B motifs form the P-loop, which binds to ATP. The Q-loop and H-loop contain residues that are important for interacting with the  $\gamma$ -phosphate of the ATP (Davidson *et. al.* 2008). The signature motif LSGGQ is found only in ABC proteins, which makes it valuable for distinguishing ABC proteins from other ATPases.

## **Plant ABC Transporter Nomenclature**

The current and most widely used classification system for ABC transporter subfamilies in plants is based on protein solubility, presence of TMDs, function, and amino acid sequence (Verrier *et. al.* 2008). In our current study we used the system described by Verrier *et. al.* (2008) consistent with the Human Genome Organization (HUGO) designation but including another subfamily, subfamily I. Plant ABC transporters fall into eight subfamilies using this nomenclature: A, B, C, D, E, F, G, and I (Table 1). ABCH subfamily members, a ninth subfamily, have not been identified in plants. ABCE and ABCF subfamily members do not have TMDs, and as a result, are soluble proteins. Soluble proteins such as the ABCF and ABCE

members of the ABC transporter protein superfamily, despite being called such, lack any transport phenomena (Verrier *et. al.* 2008).

#### **ABCA:**

ABCA transporters also referred to as AOH (ABC1 homologue), have full and half size variants among plant genomes (Verrier *et. al.* 2008). These transporters have a forward orientation TMD-NBD as defined by Sanchez-Fernandez *et. al.* (2001). ABCA subfamily members are likely involved in the transportation of sterols and, in general, lipids based on evidence seen in human movement of iron-sulfur clusters out of mitochondria (Chen, S. *et. al.* 2007). Full-length members, P-glycoprotein multidrug resistance proteins, are involved in the transport of auxins (Geisler *et. al.* 2006) and secondary metabolites (Yazaki, K. 2006).

#### **ABCC:**

ABCC transporters are full size and forward oriented TMD-NBD. Members of this family generally play some role in detoxification (Verrier *et. al.* 2008). Multidrug resistance-associated proteins are thought to be involved in the transport of organic anions and/or xenobiotic anions into the vacuole (Klein, M. *et. al.* 2005).

#### **ABCD:**

D subfamily members are involved in importing substrates relevant to glyoxylate cycle in particular  $\beta$ -oxidation (Theodoulou, F. L. *et. al.* 2006). ABCD transporters have a forward orientation TMD-NBD and are always half size. They are capable of forming homodimers or heterodimers. These subfamily members are in involved in the transport of fatty acids into the peroxisome.

**ABCE:**

ABCE transporters contain two NBD domains and no TMD domains. As a result these transporters are soluble. It is not clear if these subfamily members are involved in transport (Verrier *et. al.* 2008). A ribosome detaches from mRNA with the help of ABCE1, a translation termination factor, when the ribosome reaches a stop codon (Sarmiento *et. al.* 2006).

**ABCF:**

Like ABCE transporters, ABCF transporters contain two NBD domains and no TMD domains and are soluble (Verrier *et. al.* 2008). Little is known about these transporters in plants.

**ABCG:**

Members of the ABCG subfamily have a reverse orientation NBD-TMD (Verrier *et. al.* 2008). Full-length members of this subfamily also referred to as pleiotropic drug resistant transporters, have been implicated in resistance to lead and auxinic herbicides (Crouzet, J. *et. al.* 2006, Peng *et. al.* 2010, Ito, H. *et. al.* 2006). The half-length members of this subfamily also called white-brown complex transporters, which are involved in the export of alkanes, cuticular lipids, and cutins (Panikashvili, D. *et. al.* 2007, Pighin, J. A. *et. al.* 2004).

**ABCI:**

ABCI transporters are difficult to identify and categorize. Members of the ABCI subfamily were originally a group of heterogeneous proteins that could not be grouped with other ABC transporters (Sanchez-Fernandez, R. *et. al.* 2001). ABCI subfamily transporters likely derived from organelle genomes based on similarities in domain organization to multi-subunit ABC transporters found in prokaryotes (Verrier *et. al.* 2008). A few examples of well-categorized systems in plants have been deduced: iron-sulphur centre biogenesis complex (Lill,

R. *et. al.* 2005), cytochrome c maturation complex (Rayapuram, N. *et. al.* 2007), and trigalactosyldiacyl glycerol complex (Lu, B. *et. al.* 2007).

## **One Thousand Transcriptomes Project (1KP)**

ABC transporter studies in plants have focused on identifying gene family members and their function using published reference genome sequences from *Arabidopsis thaliana* (*Arabidopsis*), *Oryza sativa* (rice) and *Vitis vinifera* (grape) (Verrier *et. al.* 2008, Nguyen *et. al.* 2014, Cakir *et. al.* 2013). Reference genomes enable the discovery of the complete set of gene family members for a species, however, there are only 48 reference genomes for plant species in Phytozome ([www.phytozome.net](http://www.phytozome.net)), leaving many unexplored phylogenetic branches. Despite the incomplete nature of *de novo* transcriptome sequencing, the abundance of publicly-available RNASeq data sets across species facilitate the examination of gene family evolution across a wider set of species and lineages (Matasci N *et. al.* 2014).

The objective of our study is to survey the diversity of ABC transporters across the plant kingdom using a new public resource for plant transcriptomes, the One Thousand Transcriptomes Project (1KP; [www.onekp.com](http://www.onekp.com)), which offers a source of transcriptome sequencing data spanning the breadth of diversity of the plant kingdom. The 1KP project provides raw data as well as assembled reference transcriptomes generated with standardized processing procedures.

The 1KP project includes sequenced transcriptomes from two species that also have available reference genome sequences: *Ricinus communis* (castor bean) and *Linum usitatissimum* (flax), which are both eudicots. The castor bean is originally from Africa, but now grown as an oilseed crop (Chan *et. al.* 2010) throughout the world in tropical and subtropical regions (Allan

**Table 1. A comparison of different nomenclature systems for ABC proteins**

HUGO subfamily	Sánchez-Fernández subfamily <sup>a</sup>	ABCISSE family <sup>b,c</sup>	ABCISSE subfamily <sup>b,c</sup>	TC subfamily <sup>d</sup>	Domain organisation	Taxa
ABCA	ABC1 homologue (AOH)	Drug and antibiotic resistance (DRA)	ABCA	Cholesterol/phospholipid/retinal flippase (CPR)	(TMD-NBD)2	eukaryotes (not yeast)
	ABC2 homologue (ATH)				TMD-NBD	eukaryotes
ABCB	Multidrug resistance (MDR)	Drug, peptides and lipid export (DPL)	p-glycoprotein (p-GP)	MDR	(TMD-NBD)2	prokaryotes and eukaryotes
	Transporter associated with antigen processing (TAP)		TAP and multidrug resistance-like protein (MDL)	TAP and mitochondrial peptide exporter (MPE)	TMD-NBD	eukaryotes
	ABC transporter of the mitochondria (ATM)		Heavy metal tolerance (HMT)	HMT	TMD-NBD	eukaryotes
	-		Lipid A-like exporter, putative (LLP)	-	TMD-NBD	prokaryotes and plants
ABCC	Multidrug resistance associated protein (MRP)	Organic anion conjugates and drug export (OAD)	MRP	Conjugate transporter (CT)	(TMD-NBD)2	eukaryotes
ABCD	Peroxisomal membrane protein (PMP)	Fatty acid export (FAE)	-	Peroxisomal fatty acyl-CoA transporter (P-FAT)	TMD-NBD; (TMD-NBD)2	bacteria and eukaryotes
ABCE	RNase L inhibitor (RLI)	RNase L inhibitor (RLI)	-	-	NBD-NBD	archaea and eukaryotes
ABCF	General control non-repressible (GCN)	Antibiotic resistance and translation regulation (ART)	Gene expression regulation (REG)	-	NBD-NBD	bacteria and eukaryotes
ABCG	White-brown complex homologue (WBC)	Eye pigment precursors and drugs (EPD)	WHITE	Eye pigment precursor transporter (EPP)	NBD-TMD	bacteria and eukaryotes
	Pleiotropic drug resistance (PDR)		PDR	PDR	(NBD-TMD)2	plants and fungi
ABCH	-	Drug resistance, bacteriocin and lantibiotic immunity (DRI)	YHIH		NBD-TMD	prokaryotes, slime moulds, echinoderms, insects and fish

References: a (Sanchez-Fernandez *et. al.* 2001), b (Dassa *et. al.* 2001), c (Bouige *et. al.* 2002), d (Saier *et. al.* 2000). Printed with permissions from Verrier *et. al.* 2008.

*et. al.* 2007). Flax, also known as linseed, is in the Linaceae family, and is a cool-temperate crop grown for both food and fiber (Wang *et. al.* 2012). These two species provide an opportunity to examine the effectiveness of the 1KP transcriptome data set analyses with respect to available reference genomes. Specifically, these data will serve as a touchstone for validating the usage of RNASeq transcriptomes for detection and categorization of ABC transporter gene family members.

## CHAPTER II

### METHODS

#### **Data acquisition and quality control**

Total data, including raw sequence reads, transcriptome assemblies, and assembly statistics, were gathered from the 1KP collaborative project ([www.onekp.com](http://www.onekp.com)). The 1KP consortium performed all sample collections, sequencing, quality control, and assembly (Matasci *N et. al.* 2014). A total of 1,462 assemblies were downloaded from the 1KP repository. Since some crucial taxonomic information was missing from accompanying data files, 161 samples were excluded from further analysis. All remaining 1,301 1KP samples were translated into six frames using the transeq program from EMBOSS (The European Molecular Biology Open Software Suite)(Rice P. *et. al.* 2000).

#### **HMMER domain e-value selection**

The full sets of Arabidopsis and rice protein sequences were obtained from Phytozome version 10. The list of Arabidopsis ABC transporter proteins was obtained from Verrier *et. al.* (2008). The list of rice ABC transporter proteins was obtained from Crouzet *et. al.* (2006). The hmmscan program from the HMMER package version 3.1 was used to search for the PF00005 pfam domain across the full Arabidopsis and rice protein data sets (Finn *et. al.* 2011, Punta M. *et. al.* 2012, Verrier *et. al.* 2008). An e-value cutoff was selected in the two data sets that maximizes true positives (reported actual ABC transporters) and minimizes false negatives (reported as non-ABC transporters).

## **Comparison of genome and transcriptome data sources**

The full set of flax and castor bean proteins were obtained from Phytozome v10 (www.phytozome.net). Three sets of transcriptomic data for flax (shoot sample 1, shoot sample 2, shoot sample 3) and one set of castor bean (mixed tissue sample) were obtained from the 1KP data set. The hmmscan program was used to search for PF00005 pfam domains among all six sets of data using an e-value of  $3.10E-18$ , yielding six sets of putative ABC transporter genes.

## **ABC transporter classification**

Each sample's translated transcriptome was searched using hmmscan (Finn *et. al.* 2011). Search results containing the PF00005 pfam domain with an e-value of  $3.10E-18$  were subjected to further classification using best BLAST (Basic Local Alignment Search Tool) searches (Altschul *et. al.* 1990). A specialized database of known plant ABC transporters from Arabidopsis and rice was constructed for subfamily classification purposes using legacy BLAST formatdb version 2.2.26. All known plant ABC transporter subfamilies were represented in our BLAST database.

Each of the 1,301 1KP subsamples along with flax and castor bean genome-derived proteomes were used to query the ABC transporter database using legacy BLAST version 2.2.26, in each case the best hit to the ABC transporter database was used to classify the query among one of the eight ABC transporter subfamilies.

## **Nonparametric Comparison Using Wilcoxon Method**

JMP version 11 (Cary, NC) was used to perform oneway analysis of counts by species and tested using nonparametric comparisons for each pair using Wilcoxon method. All subfamily

and total ABC transporter protein gene member classifications were tested pairwise between each of the taxa: angiosperms, conifers, ferns, lycophytes, non-vascular land plants, and algae. A single exception was made among non-vascular land plants, a sample of *Blasia sp.* was not included in the pairwise tests.

## CHAPTER III

### RESULTS AND DISCUSSION

#### **Pfam HMMER domain e-value**

The pfam database is a set of protein family records, each with multiple sequence alignments and HMMs (Hidden Markov Models) that may be used to identify protein family members and/or domains from sequence sets (Punta *et. al.* 2012). The pfam record PF00005 represents the highly conserved ATP-binding domain of ABC transporters. In order to test the ability of a pfam search to accurately identify ABC transporter gene members across an entire transcriptome, we tested the pfam discovery approach against the full gene sets of Arabidopsis and rice. Of the 233 known ABC transporters and 85,152 known non-ABC transporters protein sequences, the hmmscan searches for the PF00005 pfam domain revealed that ABC transport protein sequences often had much lower expectation values for containing the PF00005 pfam domain than other known non-ABC transport protein sequences (85% of the top 200 lowest e-value hits were known ABC transport proteins). While the pfam search proved effective at ABC transporter gene member discovery, there was not a clear delineation in e-value scores between known gene family members and other genes. The highest e-value among known ABC transporters was  $2.10E-10$ . However, some non-ABC transporters showed e-values lower than the  $2.10E-10$  cutoff. Searching among the top 5 highest e-values we identified the largest gap between e-values of  $3.10E-18$  and  $2.10E-15$ , a separation of 3 orders of magnitude between non-ABC transporters and ABC transporters. Using the  $3.10E-18$  threshold of significance, we identified 282 sequences as ABC transporters. Of 282 sequences identified as ABC transporters 230 have been previously classified as ABC transporters, but 52 have not (putative false

positives). The  $3.10E-18$  e-value cut-off excluded three known ABC transporters (putative false negatives). With the entire set of 85,152 starting sequences, the usage of the pfam PF00005 domain yielded a sensitivity of 0.9871 and a specificity of 0.9994. An e-value of  $3.10E-18$  was selected as the best level of significance for ABC transporter detection for all further HMMER analysis.

## **Comparison of genome and transcriptome data sources**

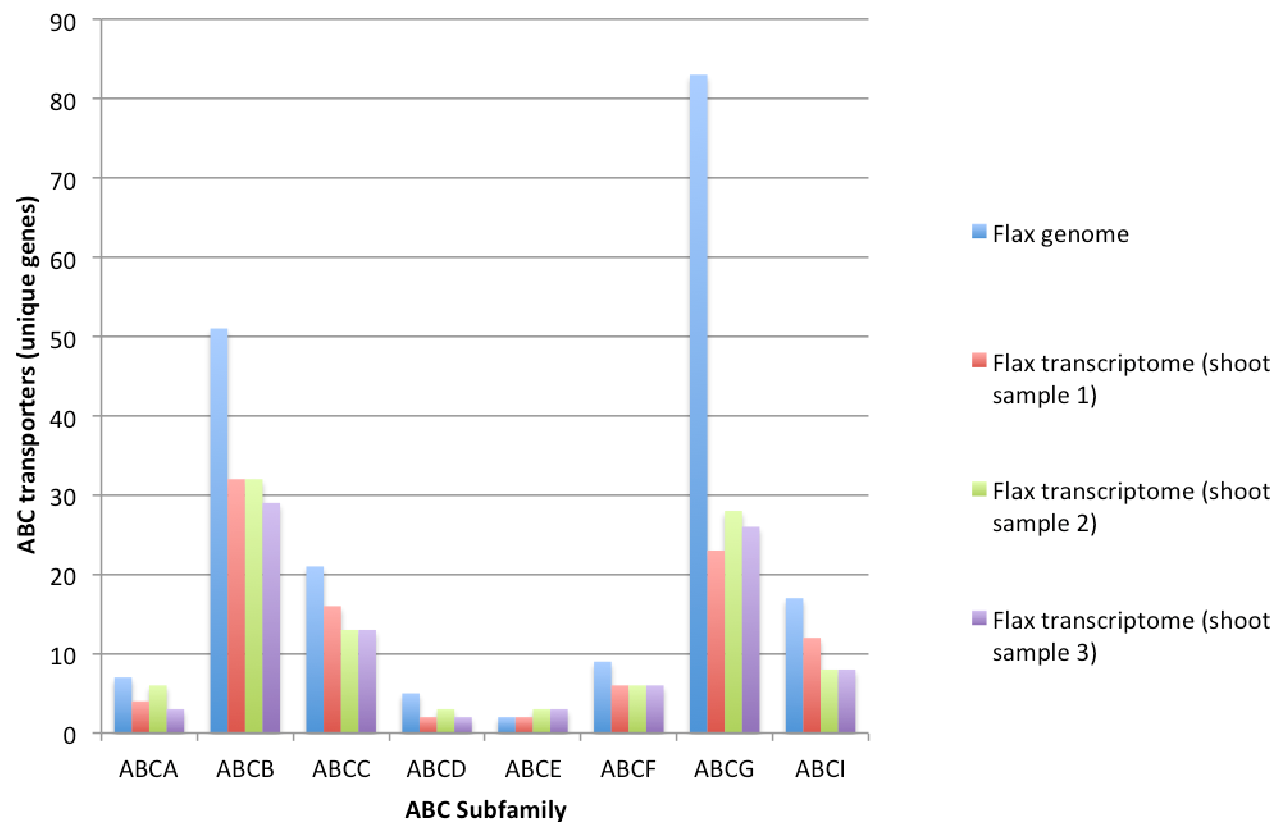
Transcriptome sequencing yields sequences only from the genes expressed in the particular sample, usually a subset of the total genes in an organism's genome. The ABC transporter genes are known to have variable expression across tissues (Nguyen *et. al.* 2014), and thus their identification from RNASeq is unlikely to yield a complete complement of gene family members for the organism. The 1KP set includes transcriptomes from flax and castor bean, both of which have published reference genome sequences (Wang *et. al.* 2012, Chan *et. al.* 2010). These provide an opportunity to examine how the number of ABC transporters identified in a transcriptome sequencing project correlates to the number of ABC transporters in the entire genome.

Three flax libraries were produced during the 1KP project, each from a different plant, and all three from shoot tissue samples. The 1KP-generated assemblies had 99,855, 95,813, and 101,110 total putative transcripts for each library. The number of transcripts identified by hmmscan as ABC protein transporter gene members was similar between the three flax samples: 82, 91 and 92 putative ABC transporters. This represents 42-47% of the number of ABC transporters identified from the 195 flax reference genome protein sequences. In the case of castor bean, only one library was produced within the 1KP and it originated from a mixed tissue

sample. A total of 59 transcripts were putatively identified as ABC transporters, out of 142 (42%) putatively identified in the genome-sequence derived gene set.

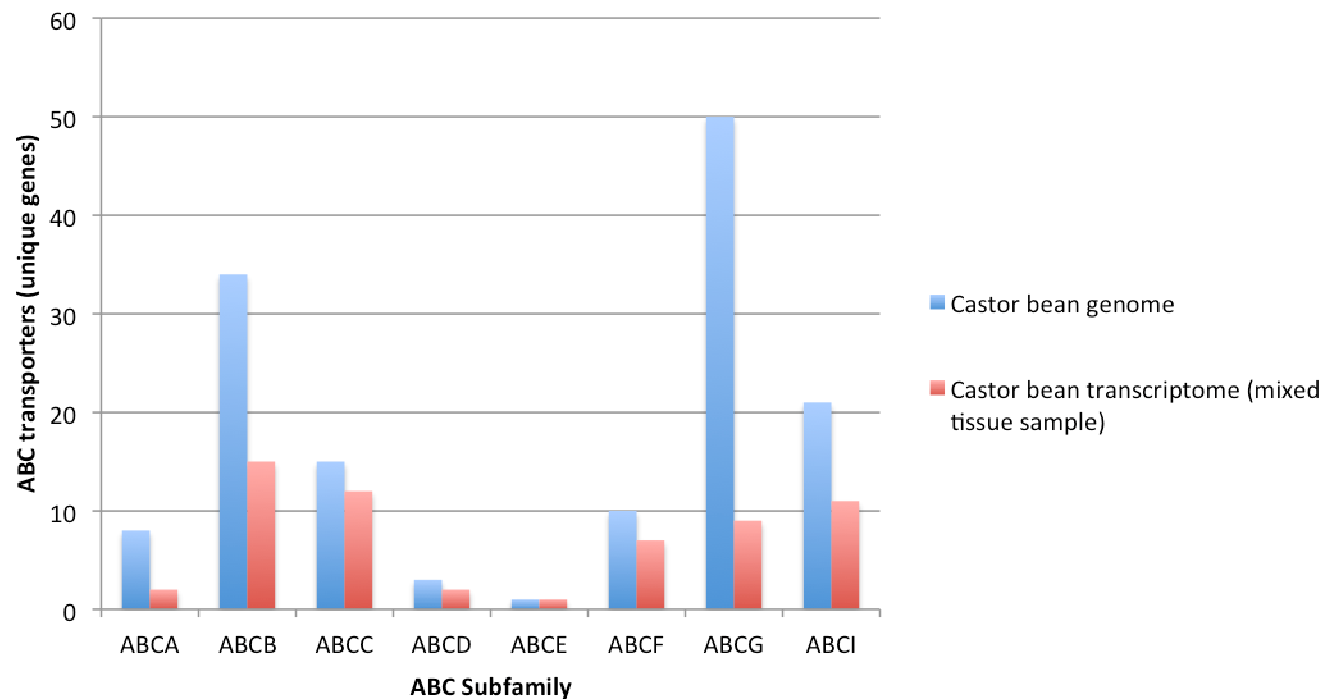
Subfamilies for putative ABC transporters were assigned on the basis of the most similar amino acid sequence from *Arabidopsis* or rice. The distribution of the ABC transporters among the eight subfamilies shared a similar pattern between the 1KP transcriptome samples and the reference genome-derived gene sets (Figures 1 and 2). Hmmscan searches of the 1KP data set with a cutoff of  $3.10E-18$  assigned 1KP samples 50 or more PF00005 pfam domains (Figure 3). In many cases we recorded outliers, the largest of which, *Blasia sp.* (526 ABC transporters), more than doubled the next largest sample, *Phaeoceros caronlinianus* (181 ABC transporters). For ease of interpretation the *Blasia sp.* sample was left out of figures as its inclusion largely skewed the y-axis, total ABC transporter proteins, in most cases. Among the total ABC transporters classified: non-vascular land plants, lycophytes, conifers, and angiosperms had 16 to 43 more ABC transporter protein transcripts than algae. The average number of transporters between algae and these four other lineages were found to be significant ( $p < 0.005$ ). Although algae and higher-level plants in at least one case have been shown to exhibit similar numbers of ABC transporter subfamily gene members and total ABC transporter genes (Merchant *et al.* 2007) the less complex transport systems of algae may contribute to the reduced number of unique ABC transporter protein genes observed. Angiosperms, conifers, ferns, and algae had 22 to 43 fewer unique ABC transporter protein gene members than non-vascular land plants. The differences in mean number of transporters between these groups is also significant ( $p < 0.005$ ).

In general, the 1KP data contained nearly half the number of unique ABC transport gene members by comparison to their reference genome counterparts as identified by hmmscan searches. Relative consistency between 1KP data samples is encouraging for extrapolation of



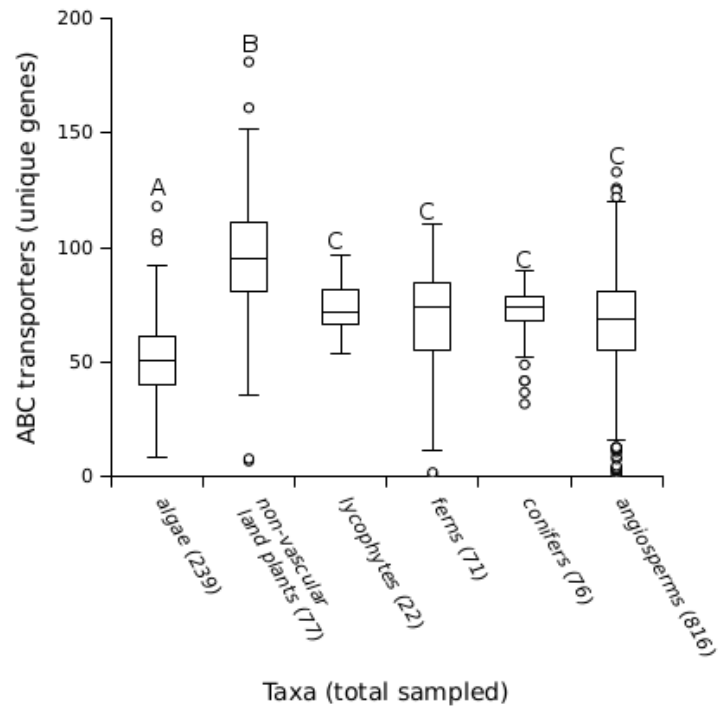
**Figure 1. Flax ABC transporter unique gene distribution deduced by RNA-Seq transcriptome data**

Plot comparing total ABC transporter unique genes observed in the flax genome and the three 1KP transcriptomes for flax (MJAV, OGSY, OZJZ) based on hmmscan searches for the PF00005 pfam domain and subsequent BLAST queries against a custom BLAST database of ABC transporters from Arabidopsis and rice.



**Figure 2. Castor bean ABC transporter unique gene distribution deduced by RNA-Seq transcriptome data**

Plot comparing total ABC transporter unique genes observed in the castor bean genome and the 1KP transcriptomes for castor bean (PAZJ) based on hmmscan searches for the PF00005 pfam domain and subsequent BLAST queries against a custom BLAST database of ABC transporters from Arabidopsis and rice.



**Figure 3. Number of unique ABC transporter gene members distributed over plant taxa deduced by RNA-Seq transcriptome data.**

Box and whisker plot of subfamily ABC transporter unique genes among each set of samples in indicated taxa: algae, non-vascular land plants, lycophytes, ferns, conifers, angiosperms. Classifications are based on hmmscan searches for the PF00005 pfam domain. Significant differences between unique gene counts of the six taxa were determined by a comparison of means using the Wilcoxon method. Unfilled circles indicate outliers. The total samples per taxa are indicated in parentheses beside each taxon along the x-axis. Statistical differences are indicated by letter groupings ( $p < 0.005$ ). *Blasia sp.* (not pictured), a non-vascular land plant sample, had 526 ABC transporter proteins.

total number of gene family members across transcriptome data sets. In addition, hmmscan found nearly the same number of unique ABC transporters for each of the three 1KP data sets for flax.

## **ABC transporter subfamily classification**

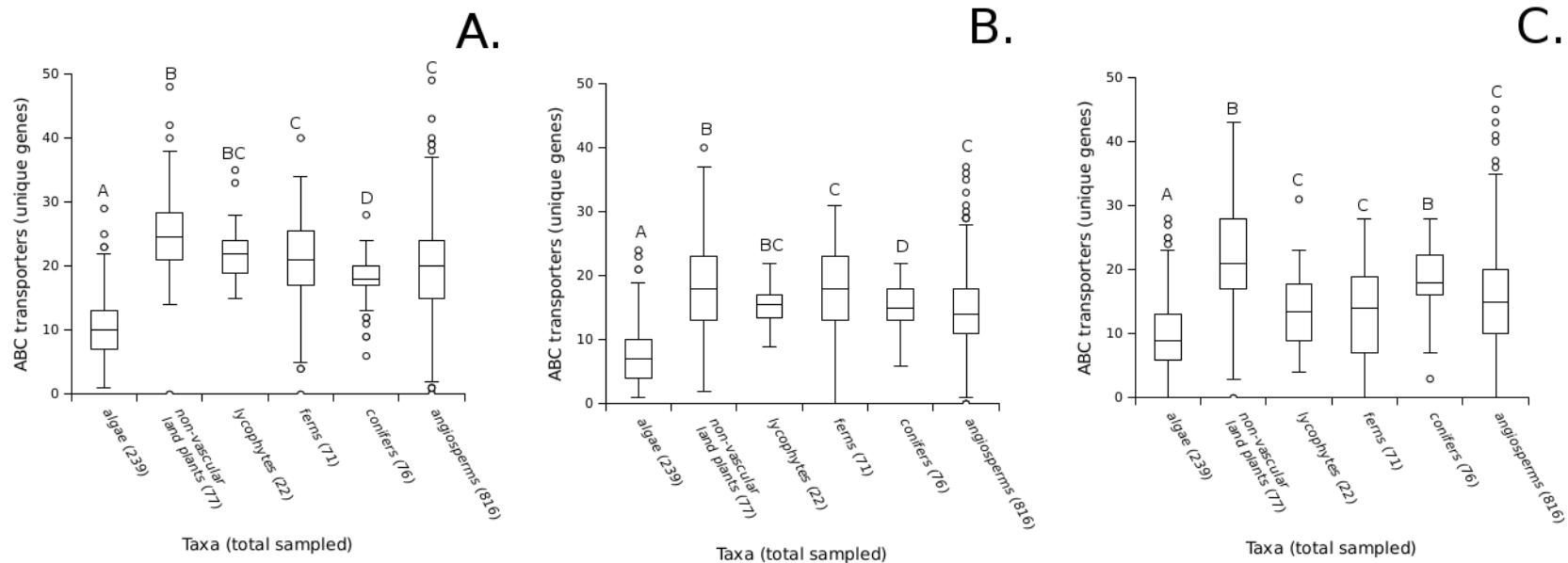
Studies in *Arabidopsis*, rice, and grape have largely shaped our understanding of the ABC transporter protein superfamily among plant taxa. Thus, information of ABC transporter diversity throughout the plant kingdom has been primarily limited to angiosperms. In an effort to expand our knowledge of the varying representation of ABC transporter protein subfamilies across plant taxa, we have classified nearly 97,149 transcripts across 1,301 samples into one of the seven ABC transporter HUGO subfamilies (A, B, C, D, E, F, G) and the plant-only subfamily, I (Verrier *et. al.* 2008). The 1,301 1KP samples span a diversity of plant taxa: 816 angiosperm samples, 76 conifer samples, 71 fern samples, 22 lycophytes samples, 77 non-vascular land plant samples, and 239 algae samples. The breadth of samples across different groups provides robust statistical support for comparisons between the groups.

Results from BLAST subfamily classifications were used to identify full-length transcripts based on query coverage. Since the ABC transport protein subfamilies split prior to the most recent common ancestor of all plants it is likely that ABC transport protein subfamily members will closely match an ortholog within the same subfamily. Transcripts with greater than 90% of the query aligned to the target sequence were classified as full length. All 1,301 transcriptome samples contained at least one putative unique ABC transporter gene member with 97,149 ABC transporters identified across all samples. A total of 22,343 ABC transporter sequences were identified as full length. There were, on average, 88 ABC transporter gene

members discovered per RNASeq library with a standard deviation of 30 gene members. Most samples (1052 of 1301) contained at least one of each of the ABC transporter protein subfamily members.

The number of unique ABC transport protein subfamily gene members varied between distantly related plant taxa. Among the functionally diverse subfamilies ABCB and ABCC, we saw significantly fewer unique ABC transport protein gene members on average in algae (5 to 15 gene members) than in other groups: angiosperms, conifers, ferns, lycophytes, and non-vascular land plants ( $p < 0.0001$ )(Figure 4A and 4B). Among ABCG subfamily transporters the differences in the average number of ABC transporter gene members seen between algae was significantly lower (5 to 11 gene members) than the following taxa: angiosperms, conifers, and non-vascular land plants ( $p < 0.0001$ )(Figure 4C).

Several studies have indicated that plant genomes have more ABCB, ABCC and/or ABCG subfamily gene members in comparison to other eukaryotes such as human and yeast (Verrier *et. al.* 2008, Geisler *et. al.* 2006, Crouzet, J. *et. al.* 2006). Our results reveal additional variability within plant taxa regarding the size of these subfamilies, specifically that algae contain fewer unique ABCB, ABCC and ABCG transport protein gene members. The larger amount of unique ABCB transport protein gene members seen in vascular plant species may result from an evolutionary need for the export of heavy metals. As in the case of AtABCB25, an Arabidopsis ABCB export protein, which allows some tolerance to cadmium and lead when overexpressed (Kim, D. Y. *et. al.* 2006). ABCC transport proteins have been identified among plants for their role in detoxification, and the regulation of stomata guard cells (Klein, M. *et. al.* 2006). The gap in ABCC transport protein transcripts seen between algae and other plant taxa might be explained by a lack of need by algae for stomata regulation and/or detoxification. Algae

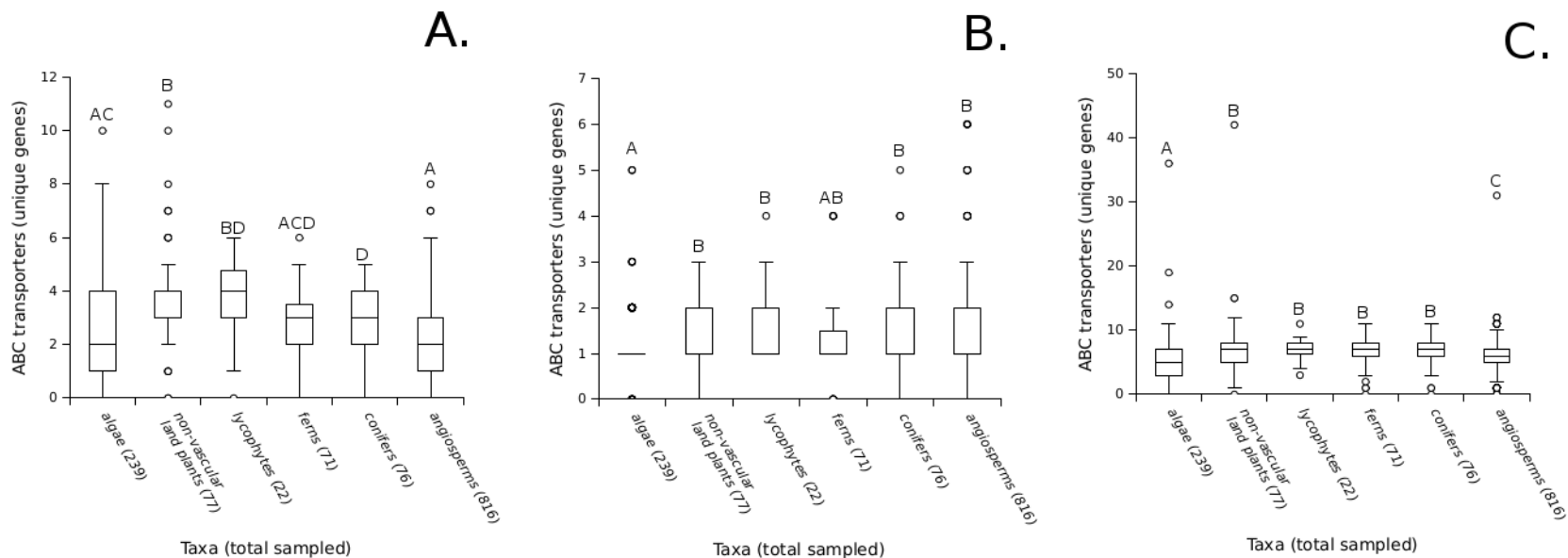


**Figure 4. Number of unique ABC transporter gene members of subfamilies B, C, and G distributed over plant taxa deduced by RNA-Seq transcriptome data.**

(A) ABCB transporter protein unique genes. (B) ABCC transporter protein unique genes. (C) ABCG transporter protein unique genes. Box and whisker plot of subfamily ABC transporter unique genes among each set of samples in indicated taxa: algae, non-vascular land plants, lycophytes, ferns, conifers, angiosperms. Classifications are based on hmmscan searches for the PF00005 pfam domain and subsequent BLAST queries against a custom blast database of ABC transporters from *Arabidopsis* and rice. Significant differences between unique gene counts of the six taxa were determined by a comparison of means using the Wilcoxon method. Unfilled circles indicate outliers. The total samples per taxa are indicated in parentheses beside each taxon along the x-axis. Statistical differences are indicated by letter groupings ( $p < 0.005$ ). *Blasia sp.* (not pictured), a non-vascular land plant sample, had (A) 137 ABCB transporters, (B) 11 ABCC transporters, and (C) 30 ABCG transporters.

do not require the specialized transport systems of land plants since most cells have direct access to nutrients in their water environment. Such direct access to a water environment may contribute to fewer unique gene members of subfamilies ABCB and ABCC. Algae also do not produce a waxy cuticle—so reduced ABCG transcripts may be related to a lack of the ABCG transport proteins that are necessary for the transport of cuticular lipids (Pighin *et. al.* 2004). Whether the variation in size of these families is a result of gene expansion is unclear and requires further investigation for validation.

Among subfamilies ABCD, ABCE, and ABCI, we did not see many significant differences in the average number of gene members across taxa (Figure 5). The few significant differences in the average number of ABC transport protein gene members detected among ABCD, ABCE, and ABCI were small, ranging from zero to two ( $p < 0.0001$ ). The presence of ABCD, ABCE, and ABCI transporter protein transcripts with very similar gene member averages might suggest that the proteins which result from the translation of these transcripts are essential to life among all plant taxa. Plant ABCD transporters are well known for their import of substrates necessary in the glyoxylate cycle (Theodoulou *et. al.* 2006). It seems likely that plants retaining ABCD transporters would have an evolutionary advantage over plants that do not given the importance of the glyoxylate cycle to survival across all plant taxa. ABCE transporters are found in archaeobacteria, bacteria, and eukaryotes suggesting that this class of ABC transport proteins is essential to life and are thus conserved (Verrier *et. al.* 2008). ABCE1 is a translation termination factor. Once a ribosome reaches a stop codon ABCE1 helps detach the ribosome from the mRNA (Sarmiento *et. al.* 2006). The maintenance of similar numbers of unique ABCD and ABCE subfamily transport gene members may be due to the conservation of essential



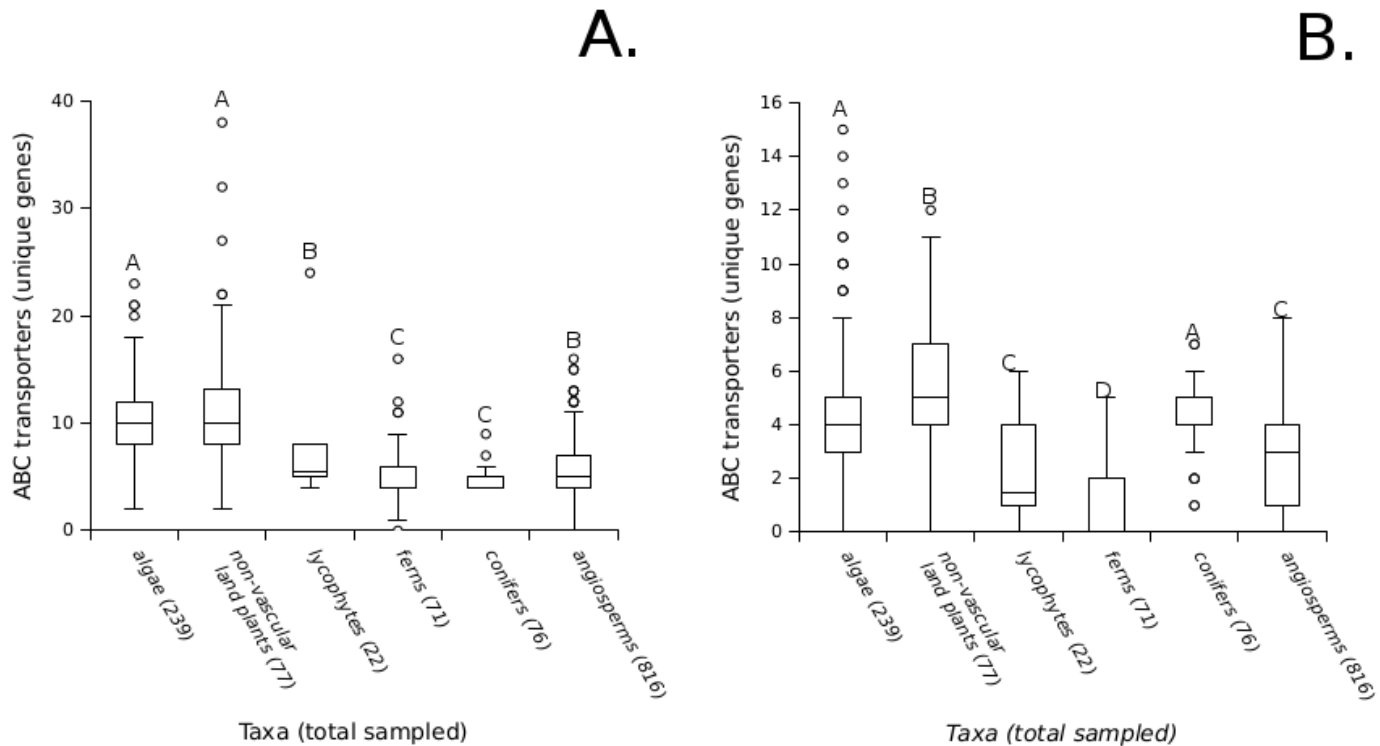
**Figure 5. Number of unique ABC transporter gene members of subfamilies D, E, and I distributed over plant taxa deduced by RNA-Seq transcriptome data.**

(A) ABCD transporter protein unique genes. (B) ABCE protein unique genes. (C) ABCI transporter protein unique genes. Box and whisker plot of subfamily ABC transporter unique genes among each set of samples in indicated taxa: algae, non-vascular land plants, lycophytes, ferns, conifers, anigosperms. Classifications are based on hmmscan searches for the PF00005 pfam domain and subsequent BLAST queries against a custom blast database of ABC transporters from Arabidopsis and rice. Significant differences between unique gene counts of the six taxa were determined by a comparison of means using the Wilcoxon method. The total samples per taxa are indicated in parentheses beside each taxon along the x-axis. Statistical differences are indicated by letter groupings ( $p < 0.005$ ). *Blasia sp.* (not pictured), a non-vascular land plant sample, had (A) 5 ABCD transporters, (B) 0 ABCE transporters, and (C) 229 ABCI transporters.

functions provided by those families. The ABCI subfamily is less understood and is expected to be altered significantly as more information regarding their structure and function becomes available. However, current data suggests the origins of ABCI subfamily transport genes from prokaryotic genomes and result from a movement of endosymbiont genes from mitochondria and chloroplasts to the nuclear genome. As such the ABCI transport proteins had been classified as a heterogeneous group composed of multicomponent transporters (reviewed in Verrier *et. al.* 2008), so the similarity seen in average number of unique gene members across plant taxa was unexpected.

In the case of ABCF subfamily transport proteins, we observed significantly higher transcript count averages in algae and non-vascular land plants than in all other groups, four to seven more transcripts on average than angiosperms, conifers, ferns, and lycophytes ( $p < 0.0001$ )(Figure 6A). When directly comparing algae and non-vascular land plants, we found that non-vascular land plants had more unique ABCF subfamily transporter gene members on average ( $p < 0.001$ ). Understanding the increased variation, specifically regarding the ABCF subfamily size among algae and non-vascular land plants is difficult with a lack of functional characterization of ABCF transport proteins among plant taxa. Studies in yeast and humans have indicated ABCF transport protein function in the activation of eIF-2 $\alpha$  kinase (Verrier *et. al.* 2008). Activation of eIF-2 $\alpha$  kinase is important in the regulation of stress response factors (Kimball SR 1999). An increase in the total number of ABCF transport protein gene members among algae and non-vascular land plants could indicate the need for more stress responses among algae and non-vascular land plants relative to other plant taxa.

The average number of unique ABCA subfamily transport protein gene members was significantly lower in ferns, three to four less gene members on average in most cases ( $p <$



**Figure 6. Number of unique ABC transporter gene members of subfamilies F and A distributed over plant taxa deduced by RNA-Seq transcriptome data.**

(A) ABCF subfamily transporter protein unique genes. (B) ABCA subfamily transporter protein unique genes. Box and whisker plot of subfamily ABC transporter unique genes among each set of samples in indicated taxa: algae, non-vascular land plants, lycophytes, ferns, conifers, angiosperms. Classifications are based on hmmscan searches for the PF00005 pfam domain and subsequent BLAST queries against a custom blast database of ABC transporters from Arabidopsis and rice. Significant differences between unique gene counts of the six taxa were determined by a comparison of means using the Wilcoxon method. Unfilled circles indicate outliers. The total samples per taxa are indicated in parentheses beside each taxon along the x-axis. Statistical differences are indicated by letter groupings ( $p < 0.005$ ). *Blasia sp.* (not pictured), a non-vascular land plant sample, had (A) 51 ABCF transporters and (B) 63 ABCC transporters.

0.0001)(Figure 6B). While there was no significant difference between ferns and lycophytes in number of unique ABCA transporter gene members, there was a significant difference between the average number of unique ABCA transporter gene members in ferns and angiosperms ( $p < 0.0001$ ). While little is known about the function of ABCA transport proteins among plant taxa, they are suspected to be involved in lipid metabolism based on their known function in humans (reviewed in Kaminski *et. al.* 2006). Upon comparison with *Arabidopsis*, rice was found to be missing a full length ABCA transporter. This is thought to be due to gene loss based on observations made in early land plants containing orthologues of this particular full length ABCA transporter protein gene (Garcia *et. al.* 2004). Our results indicate an occurrence of decreased ABCA transporter protein genes among ferns relative to plant taxa, possibly from gene loss. While the single gene loss observed in rice and the results observed in fern are likely unrelated, however they may point to similar phenomena regarding the importance of ABCA transport proteins.

## CHAPTER IV

### CONCLUSIONS

The results of this study provide insights into the variation of the average number of unique ABC transport protein genes across the plant kingdom. Specifically, we have seen significant differences in number of unique ABC transport protein gene members between algae, non-vascular land plants, lycophytes, ferns, conifers, and angiosperms. We show, that in general, algae and non-vascular land plants have fewer unique ABC transport protein gene members than lycophytes, ferns, conifers, and angiosperms. This suggests that fewer ABC transport protein genes may exist among algae and non-vascular land plants than other plant taxa or that these taxa express fewer kinds of ABC transport protein transcripts. When looking at specific ABC transporter subfamilies we see more interesting patterns of variation across plant species. Algae have the fewest ABCB, ABCC, and ABCD subfamily transport protein gene members as compared to all other plant taxa represented in this study—this fits with the trend we see in general among plant ABC transport protein gene members. The number of unique ABCD, ABCE, ABCI subfamily transport protein gene members show few significant differences throughout plant taxa. We suspect that ABCD, ABCE, and ABCI have been tightly conserved given the small variation we see among their unique gene members, due in part to their importance in survival among all plant taxa. Perhaps most surprising, the average number of unique ABCA transporter gene members is very small among fern samples compared to the rest of the plant kingdom. Further investigation is required to understand the circumstances that have led to such a low number of unique ABCA transport protein gene members among ferns. ABCF transport protein transcripts show an inverse to what is seen in general among plant ABC transport protein gene members: algae and non-vascular land plants have more unique ABCF

transporter gene members than lycophytes, ferns, conifers, and angiosperms.

## **LIST OF REFERENCES**

1. Allan G, Williams A, Rabinowicz PD, Chan AP, Ravel J, Keim P: **Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs.** *Genetic Resources and Crop Evolution* 2007, **55**(3):365-378.
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**(3):403-410.
3. Bouige P, Laurent D, Piloyan L, Dassa E: **Phylogenetic and functional classification of ATP-binding cassette (ABC) systems.** *Curr Protein Pept Sc* 2002, **3**(5):541-559.
4. Brocchieri L: **Phylogenetic inferences from molecular sequences: review and critique.** *Theoretical population biology* 2001, **59**(1):27-40.
5. Burleigh JG, Bansal MS, Eulenstein O, Hartmann S, Wehe A, Vision TJ: **Genome-scale phylogenetics: inferring the plant tree of life from 18,896 gene trees.** *Syst Biol* 2011, **60**(2):117-125.
6. Cakir B, Kilickaya O: **Whole-genome survey of the putative ATP-binding cassette transporter family genes in *Vitis vinifera*.** *PloS one* 2013, **8**(11):e78860.
7. Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G *et al*: **Draft genome sequence of the oilseed species *Ricinus communis*.** *Nature biotechnology* 2010, **28**(9):951-956.
8. Chen S, Sanchez-Fernandez R, Lyver ER, Dancis A, Rea PA: **Functional characterization of AtATM1, AtATM2, and AtATM3, a subfamily of *Arabidopsis* half-molecule ATP-binding cassette transporters implicated in iron homeostasis.** *The Journal of biological chemistry* 2007, **282**(29):21561-21571.
9. Crouzet J, Trombik T, Fraysse AS, Boutry M: **Organization and function of the plant pleiotropic drug resistance ABC transporter family.** *FEBS letters* 2006, **580**(4):1123-1130.
10. Dassa E, Bouige P: **The ABC of ABCS: a phylogenetic and functional classification of ABC systems in living organisms.** *Res Microbiol* 2001, **152**(3-4):211-229.
11. Davidson AL, Dassa E, Orelle C, Chen J: **Structure, function, and evolution of bacterial ATP-binding cassette systems.** *Microbiology and molecular biology reviews : MMBR* 2008, **72**(2):317-364, table of contents.
12. Davies TGE, Coleman JOD: **The *Arabidopsis thaliana* ATP-binding cassette proteins: an emerging superfamily.** *Plant Cell and Environment* 2000, **23**(5):431-443.
13. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *Nucleic acids research* 2011, **39**(Web Server issue):W29-37.
14. Garcia O, Bouige P, Forestier C, Dassa E: **Inventory and comparative analysis of rice and *Arabidopsis* ATP-binding cassette (ABC) systems.** *Journal of molecular biology* 2004, **343**(1):249-265.
15. Ge X, d'Avignon DA, Ackerman JJ, Sammons RD: **Rapid vacuolar sequestration: the horseweed glyphosate resistance mechanism.** *Pest management science* 2010, **66**(4):345-348.
16. Geisler M, Murphy AS: **The ABC of auxin transport: the role of p-glycoproteins in plant development.** *FEBS letters* 2006, **580**(4):1094-1102.
17. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N *et al*: **Phytozome: a comparative platform for green plant genomics.** *Nucleic acids research* 2012, **40**(Database issue):D1178-1186.

18. Hanson BD, Shrestha A, Shaner DL: **Distribution of Glyphosate-Resistant Horseweed (*Conyza canadensis*) and Relationship to Cropping Systems in the Central Valley of California.** *Weed Science* 2009, **57**(1):48-53.
19. Higgins CF: **ABC transporters: from microorganisms to man.** *Annu Rev Cell Biol* 1992, **8**:67-113.
20. Ito H, Gray WM: **A gain-of-function mutation in the *Arabidopsis* pleiotropic drug resistance transporter PDR9 confers resistance to auxinic herbicides.** *Plant physiology* 2006, **142**(1):63-74.
21. Jasinski M, Ducos E, Martinoia E, Boutry M: **The ATP-binding cassette transporters: structure, function, and gene family comparison between rice and *Arabidopsis*.** *Plant physiology* 2003, **131**(3):1169-1177.
22. Jones PM, George AM: **The ABC transporter structure and mechanism: perspectives on recent research.** *Cellular and molecular life sciences : CMLS* 2004, **61**(6):682-699.
23. Kaminski WE, Piehler A, Wenzel JJ: **ABC A-subfamily transporters: structure, function and disease.** *Biochimica et biophysica acta* 2006, **1762**(5):510-524.
24. Kim DY, Bovet L, Kushnir S, Noh EW, Martinoia E, Lee Y: **AtATM3 is involved in heavy metal resistance in *Arabidopsis*.** *Plant physiology* 2006, **140**(3):922-932.
25. Kimball SR: **Eukaryotic initiation factor eIF2.** *The International Journal of Biochemistry & Cell Biology* 1999, **31**(1):25-29.
26. Klein M, Burla B, Martinoia E: **The multidrug resistance-associated protein (MRP/ABCC) subfamily of ATP-binding cassette transporters in plants.** *FEBS letters* 2006, **580**(4):1112-1122.
27. Kumar A, Shukla S, Mandal A, Shukla S, Ambudkar SV, Prasad R: **Divergent signature motifs of nucleotide binding domains of ABC multidrug transporter, CaCdr1p of pathogenic *Candida albicans*, are functionally asymmetric and noninterchangeable.** *Biochimica et biophysica acta* 2010, **1798**(9):1757-1766.
28. Larsen PB, Cancel J, Rounds M, Ochoa V: ***Arabidopsis ALS1* encodes a root tip and stele localized half type ABC transporter required for root growth in an aluminum toxic environment.** *Planta* 2007, **225**(6):1447-1458.
29. Lill R, Muhlenhoff U: **Iron-sulfur-protein biogenesis in eukaryotes.** *Trends in biochemical sciences* 2005, **30**(3):133-141.
30. Lu B, Xu C, Awai K, Jones AD, Benning C: **A small ATPase protein of *Arabidopsis*, TGD3, involved in chloroplast lipid import.** *The Journal of biological chemistry* 2007, **282**(49):35945-35953.
31. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y *et al*: **SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler.** *Gigascience* 2012, **1**(1):18.
32. Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M *et al*: **Data access for the 1,000 Plants (1KP) project.** *GigaScience* 2014, **3**(17).
33. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L *et al*: **The *Chlamydomonas* genome reveals the evolution of key animal and plant functions.** *Science* 2007, **318**(5848):245-250.

34. Nguyen VN, Moon S, Jung KH: **Genome-wide expression analysis of rice ABC transporter family across spatio-temporal samples and in response to abiotic stresses.** *Journal of plant physiology* 2014, **171**(14):1276-1288.
35. Panikashvili D, Savaldi-Goldstein S, Mandel T, Yifhar T, Franke RB, Hofer R, Schreiber L, Chory J, Aharoni A: **The *Arabidopsis* DESPERADO/AtWBC11 transporter is required for cutin and wax secretion.** *Plant physiology* 2007, **145**(4):1345-1360.
36. Peng Y, Abercrombie LL, Yuan JS, Riggins CW, Sammons RD, Tranel PJ, Stewart CN, Jr.: **Characterization of the horseweed (*Conyza canadensis*) transcriptome using GS-FLX 454 pyrosequencing and its application for expression analysis of candidate non-target herbicide resistance genes.** *Pest management science* 2010, **66**(10):1053-1062.
37. Pighin JA, Zheng H, Balakshin LJ, Goodman IP, Western TL, Jetter R, Kunst L, Samuels AL: **Plant cuticular lipid export requires an ABC transporter.** *Science* 2004, **306**(5696):702-704.
38. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al*: **The Pfam protein families database.** *Nucleic acids research* 2012, **40**(Database issue):D290-301.
39. Rayapuram N, Hagenmuller J, Grienemberger JM, Giege P, Bonnard G: **AtCCMA interacts with AtCcmB to form a novel mitochondrial ABC transporter involved in cytochrome c maturation in *Arabidopsis*.** *The Journal of biological chemistry* 2007, **282**(29):21015-21023.
40. Rice P, Longden I, Bleasby A: **EMBOSS: The European molecular biology open software suite.** *Trends in Genetics* 2000, **16**(6):276-277.
41. Saier MH: **A functional-phylogenetic classification system for transmembrane solute transporters.** *Microbiology and Molecular Biology Reviews* 2000, **64**(2):354-+.
42. Sanchez-Fernandez R, Davies TG, Coleman JO, Rea PA: **The *Arabidopsis thaliana* ABC protein superfamily, a complete inventory.** *The Journal of biological chemistry* 2001, **276**(32):30231-30244.
43. Sarmiento C, Nigul L, Kazantseva J, Buschmann M, Truve E: **AtRLI2 is an endogenous suppressor of RNA silencing.** *Plant molecular biology* 2006, **61**(1-2):153-163.
44. Schneider E, Hunke S: **ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains.** *FEMS Microbiol Rev* 1998, **22**(1):1-20.
45. Theodoulou FL, Holdsworth M, Baker A: **Peroxisomal ABC transporters.** *FEBS letters* 2006, **580**(4):1139-1155.
46. van den Brule S, Smart CC: **The plant PDR family of ABC transporters.** *Planta* 2002, **216**(1):95-106.
47. Verrier PJ, Bird D, Burla B, Dassa E, Forestier C, Geisler M, Klein M, Kolukisaoglu U, Lee Y, Martinoia E *et al*: **Plant ABC proteins--a unified nomenclature and updated inventory.** *Trends in plant science* 2008, **13**(4):151-159.
48. Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R *et al*: **The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads.** *The Plant journal : for cell and molecular biology* 2012, **72**(3):461-473.

49. Ward JM: **Identification of novel families of membrane proteins from the model plant *Arabidopsis thaliana***. *Bioinformatics* 2001, **17**(6):560-563.
50. Yang H, Murphy AS: **Functional expression and characterization of *Arabidopsis* ABCB, AUX 1 and PIN auxin transporters in *Schizosaccharomyces pombe***. *The Plant journal : for cell and molecular biology* 2009, **59**(1):179-191.
51. Yazaki K: **ABC transporters involved in the transport of plant secondary metabolites**. *FEBS letters* 2006, **580**(4):1183-1191.
52. JMP<sup>®</sup>, Version 11. SAS Institute Inc., Cary, NC, 1989-2007.

## VITA

Thomas Lane was born October 28th 1988 in Salt Lake City, Utah. He was the younger of three brothers: Charles and Joseph. The youngest and fourth brother is Kodi. His parents were Scott and Jennifer. Thomas was educated at many different elementary and junior high schools before graduating from Pendleton High School in upstate South Carolina. Upon graduating, Thomas received a multitude of scholarships and had the opportunity to pursue a degree in science at Clemson University. He completed a Bachelors of Science in Genetics in 2011. Through the guidance of his mentor Dr. Amy Lawton-Rauh, he sought a graduate level degree to further his training as a scientist. Thomas accepted a NSF training fellowship, SCALE-IT, at the University of Tennessee, Knoxville. He joined the Genome Science and Technology program where he studied bioinformatics and found work in Dr. Neal Stewart's lab. Thomas completed a Masters of Science degree in Life Sciences in December 2014. He currently works at the University of Tennessee, Knoxville under the direction of Dr. Meg Staton.