Diversity of ABC transporter genes across the plant kingdom and their potential utility in biotechnology

Thomas S. Lane1,3, Caroline S. Rempe1,3, Jack Davitt2, Margaret E. Staton2, Yanhui Peng3, Douglas Edward Soltis4,5,6,7, Michael Melkonian7, Michael Deyholos8, James H. Leebens-Mack9, Mark Chase10,11, Carl J. Rothfels12, Dennis Stevenson13, Sean W. Graham14, Jun Yu15, Tao Liu16, J. Chris Pires17, Patrick P. Edger18, Yong Zhang19, Yinlong Xie18,19,20, Ying Zhu18, Eric Carpenter21, Gane Ka-Shu Wong18,21,22 and C. Neal Stewart Jr.1,3*

Abstract

Background: The ATP-binding cassette (ABC) transporter gene superfamily is ubiquitous among extant organisms and prominently represented in plants. ABC transporters act to transport compounds across cellular membranes and are involved in a diverse range of biological processes. Thus, the applicability to biotechnology is vast, including cancer resistance in humans, drug resistance among vertebrates, and herbicide and other xenobiotic resistance in plants. In addition, plants appear to harbor the highest diversity of ABC transporter genes compared with any other group of organisms. This study applied transcriptome analysis to survey the kingdom-wide ABC transporter diversity in plants and suggest biotechnology applications of this diversity.

Results: We utilized sequence similarity-based informatics techniques to infer the identity of ABC transporter gene candidates from 1295 phylogenetically-diverse plant transcriptomes. A total of 97,149 putative (approximately 25% were full-length) ABC transporter gene members were identified; each RNA-Seq library (plant sample) had 88 ± 30 gene members. As expected, simpler organisms, such as algae, had fewer unique members than vascular land plants. Differences were also noted in the richness of certain ABC transporter subfamilies. Land plants had more unique ABCB, ABCC, and ABCG transporter gene members on average (p < 0.005), and green algae, red algae, and bryophytes had significantly more ABCF transporter gene members (p < 0.005). Ferns had significantly fewer ABCA transporter gene members than all other plant groups (p < 0.005).

Conclusions: We present a transcriptomic overview of ABC transporter gene members across all major plant groups. An increase in the number of gene family members present in the ABCB, ABCC, and ABCD transporter subfamilies may indicate an expansion of the ABC transporter superfamily among green land plants, which include all crop species. The striking difference between the number of ABCA subfamily transporter gene members between ferns and other plant taxa is surprising and merits further investigation. Discussed is the potential exploitation of ABC transporters in plant biotechnology, with an emphasis on crops.

Keywords: ABC transporter, Transcriptomics, Computational biology, Taxonomic diversity

* Correspondence: nealstewart@utk.edu
1 The Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN 37996, USA
2 Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996, USA
3 Full list of author information is available at the end of the article

© 2016 The Author(s). Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background
The ATP-binding cassette (ABC) transporter family is one of the largest known protein superfamilies in biology, which is represented in all living organisms [1–7]. 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 a wide variety of substrates across various types of membranes in cells [3].

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, which dimerize in pairs to create virtual full transporters as homodimers or heterodimers. A third type of transporter has no TMDs but two NBDs [5]. 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 (LSGGQ). The D-loop primarily functions in holding dimers together, the switch H-loop interacts with the transmembrane domain, the Walker A and B motifs form the P-loop, which binds to ATP, and the Q-loop and H-loop contain residues important for interacting with the γ-phosphate of the ATP [3]. Finally, the signature motif (LSGGQ) is exclusively found in ABC proteins, which enables ABC proteins to be distinguished from other ATPases [7].

In plants, a full inventory of ABC transporters was first catalogued for *Arabidopsis thaliana* [8], partially characterized [9, 10]. 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 [5]. In our current study we used the system described in [5], which is 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. Soluble proteins lacking TMDs such as the ABCF and ABCE members of the ABC transporter protein superfamily, despite being called such, lack any transport phenomena [5]. Despite the obvious importance of ABC transporters in plants, relative to other organisms there is a dearth of functional studies on individual members in plant as well as –omics studies; typically studies are limited to a few dicot species [8–14].

There are several known functions of ABC transporters of agricultural importance. One of the most important is detoxification. Detoxification mechanisms are of particular interest in herbicide resistance studies, particularly those involving putative glyphosate sequestration into vacuoles [15–17]. The *Arabidopsis* pleiotropic drug resistance 9 (PDR9) is also known to confer auxinic herbicide resistance [18]. Export proteins are also involved in environmental response and plant development such as auxin exporters ABCC19 and ABCB1 [19]. 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 [20].

The main objective of our study was to survey the diversity of ABC transporters across plants using a new public resource for plant transcriptomes, the One Thousand Plants Transcriptome Project (1KP). 1KP includes raw transcript data as well as assembled reference transcriptomes generated with standardized processing procedures. These data included transcriptomes from two species with extant reference genomes: *Ricinus communis* (castor bean) and *Linum usitatissimum* (flax), which acted as internal controls for matching the 1KP transcriptomes with the published reference genomes. These resources offer a source of transcriptome sequencing data spanning the breadth of diversity of plants [13, 21–24].

Methods
Data acquisition and quality control
All of the data in this project, including raw sequence reads, transcriptome assemblies, and assembly statistics, were gathered from the 1KP collaborative project [21]. The 1KP consortium performed all sample collection, sequencing, quality control, and assembly [13, 21]. A total of 1462 assemblies, assembled with SOAP de novo, were downloaded from the 1KP repository [11, 25], because some crucial taxonomic information was missing from accompanying data files, 161 samples were excluded from subsequent analysis. Another six samples were removed because of contamination that was discovered during the quality control step. The remaining 1295 1KP samples were translated into six frames using the Transeq program from EMBOSS (The European Molecular Biology Open Software Suite) [26]. A six-frame translation coupled with searches amongst confirmed ABC transporter protein sequences enabled us to assign each transcript into the correct reading frame for analysis. Since well-annotated genomes exist for only a handful of species analyzed and most species lacked sufficiently large and diverse tissue sequencing results, alternative splice variants were not considered.

HMMER domain e-value selection
The full sets of *Arabidopsis* and rice protein sequences were obtained from Phytozome version 10, including the *Arabidopsis* ABC transporter protein list [5] and rice list
| HUGO subfamily | Sánchez-Fernández subfamily | ABCISSE family$^{bc}$ | ABCISSE subfamily$^{bc}$ | TC subfamily$^d$ | Domain organization | Taxa |
|----------------|-----------------------------|-----------------------|--------------------------|----------------|-------------------|------|
| ABCA           | ABC1 homologue (AOH)        | Drug and antibiotic resistance (DRA) | ABCA                     | Cholesterol/phospholipid/retinal flipase (CPR) | (TMD-NBD)$^2$ | eukaryotes (not yeast) eukaryotes |
|                | 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 | - | (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: $^a$[8], $^b$[6], $^c$[4], $^d$[52], Figure reprinted from [5] with permission
The hmmscan program from the HMMER package version 3.1 was used to search for the pfam record PF00005 across the full Arabidopsis and rice protein data sets [5, 28, 29]. A pfam record, such as the PF00005 domain, is a hidden Markov model based on a multiple sequence alignment of protein domains; in the case of PF00005 this multiple sequence alignment is of the NBD of many different ABC transporter proteins. A cutoff e-value of 3.10E-18 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 [30, 31]. 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 [11, 25]. 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 [28]. Search results containing the PF00005 pfam domain with an e-value of 3.10E-18 were subjected to further classification using best BLAST searches [32]. The de novo assembled transcriptomes from the 1KP data set were fragmented, transporters containing multiple nucleotide binding domains may only have had one of their binding domains accounted for in the transcriptomes sequencing results. Determination of ABC transport protein family was made strictly using BLAST searches against known ABC transporters in Arabidopsis and rice. 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 (Additional file 2: Table S1). All known plant ABC transporter subfamilies were represented in our BLAST database.

Table 2 A summary of HMMER search results

| E-value cutoff | True positives | False positives | True negatives | False negatives | Specificity | Sensitivity |
|----------------|----------------|-----------------|----------------|-----------------|-------------|-------------|
| 3.10E-21       | 222            | 48              | 84,882         | 11              | 0.9994      | 0.9528      |
| 3.10E-18       | 230            | 52              | 84,870         | 3               | 0.9994      | 0.9871      |
| 2.10E-15       | 231            | 52              | 84,869         | 2               | 0.9994      | 0.9914      |
| 2.10E-13       | 232            | 55              | 84,865         | 1               | 0.9994      | 0.9957      |
| 2.10E-10       | 233            | 64              | 84,855         | 0               | 0.9992      | 1.0000      |

Summary of sensitivity and specificity of 85,152 HMMER queries against pfam as a method for selecting an appropriate cutoff e-value for the identification of PF00005 (ABC transporter) domains

Each of the 1295 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 of ABC diversity
JMP version 11 [33] was used to perform a one-way analysis of counts by species and performed pairwise non-parametric comparisons using the Wilcoxon method. All subfamily and total ABC transporter gene member classifications were subjected to pairwise analysis between each of the groups: angiosperms, conifers, ferns, lycophytes, mosses, hornworts, liverworts, chlorophytes, charophytes, glaucophytes, rhodophytes, and Chromista.

Results and discussion
ABC protein discovery
The pfam record PF00005 represents the highly conserved ATP-binding domain of ABC transporters. We used this pfam record to identify ABC transporter gene members across an entire transcriptome database as well as full gene sets of Arabidopsis and rice (Table 2). ABC transporter protein sequences had much lower expectation values for the PF00005 pfam domain from the hmmscan than did known non-ABC transporter protein sequences (85 % of the top 200 lowest e-value hits were known ABC transporter proteins) (Additional file 3: Table S2). These results were obtained using 233 known ABC transporters and 85,152 known non-ABC transporter protein sequences. While the pfam search was effective in 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 five highest e-values we identified the largest gap between e-values of 3.10E-18 and 2.10E-15. This gap represents three 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 the 282 sequences identified as ABC transporters, 230 had been previously classified as ABC transporters, but 52 were not (putative false positives). Therefore, the cutoff 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 (true positives divided by true positive plus false negatives) of 0.9871 and a specificity of 0.9994 (true negatives divided by true negatives plus false positives). Consequently, an e-value of 3.10E-18 was selected as the optimum, which allowed a total of three false negatives for ABC transporter detection for all further HMMER analysis.

Comparison of genome and transcriptome data sources
The flax and castor bean transcriptomes allowed us to examine how the number of expressed ABC transporters identified in 1KP corresponds to the number of known ABC transporters in their entire genomes, which allowed us to make inferences throughout the plant kingdom, with caveats. ABC transporter genes are known to have variable expression across tissues [11], and thus their identification from RNA-Seq data is unlikely to yield a complete complement of gene family members for an organism.

The 1KP-generated assemblies for flax 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. These figures represented 42 – 47 % of the number of ABC transporters identified from the reference genome protein sequences. A total of 59 transcripts were putatively identified as ABC transporters from the single available castor bean transcriptome from 1KP, 42 % of the 142 putative genes 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 (Figs. 1a and b).

In general, the 1KP data per accession contained nearly half the number of unique ABC transporter gene members, by comparison, to their reference genome counterparts as identified by hmmscan searches. In addition, hmmscan found nearly the same number of unique ABC transporters for each of the three 1KP data sets for flax. The relative consistency between corresponding 1KP data samples is encouraging for extension of inferences across data sets in order to estimate the total number of gene family members for all plant taxa.

ABC transporter subfamily classification
The majority of our knowledge about the ABC transporter protein superfamily in plants is from studies limited to a few angiosperm species. In an effort to expand our knowledge of the varying representation of ABC transporter protein subfamilies across the plant kingdom, we classified nearly 97,149 transcripts across 1295 samples into one of the seven ABC transporter HUGO subfamilies (A, B, C, D, E, F, G) and the plant-only subfamily, I [5]. The 1295 1KP samples span a diversity of plant taxa: 816 from angiosperms, 76 from conifers, 71 ferns, 22 lycophytes, 25 liverworts, 41 mosses, 6 hornworts, 123 chlorophytes, 46 charophytes, 5 glaucophytes, and 28 rhodophytes. In general, the results comparing hornworts/glaucophytes and other plants were not significant, likely owing to the small sample sizes in these samples.
groups. The breadth of samples across various plant groups provides robust statistical support for comparisons amongst groups in most cases. The gymnosperms were mainly represented by conifers, with few other gymnosperms being represented: three Gnetales, four Cycadales, and one Ginkgoales.

The hmmscan analysis with a cutoff of 3.10E-18 found at least 50 unique ABC genes (transcripts) containing PF00005 pfam domains per plant sample (Fig. 2a). In many cases, we recorded outliers, the largest of which, Blasia sp. (526 ABC transporters), more than doubled the next largest sample. For ease of interpretation, the

![Fig. 2 Number of unique ABC transporter gene members distributed over plant taxa inferred from RNA-Seq transcriptome data.](image)

- **a** Box and whisker plot of subfamily ABC transporter unique genes among each set of samples in indicated plant groups: Chromista, rhodophytes, glaucophytes, charophytes, chlorophytes, hornworts, liverworts, mosses, lycophytes, ferns, conifers, and angiosperms. 
- **b** ABCC transporter gene members. 
- **c** ABCD transporter gene members. 
- **d** ABCD transporter gene members. 
- **e** ABC transporter gene members. 
- **f** ABC transporter gene members. 
- **g** ABC transporter gene members. 
- **h** ABC transporter gene members. 
- **i** ABC transporter gene members. 

Classifications are based on hmmscan searches for the PF00005 pfam domain. Significant differences between unique gene counts of the twelve groups were determined by a comparison of means using the Wilcoxon method. Filled circles indicate outliers. The total samples per plant group are indicated in parentheses beside each label along the x-axis. Statistical differences are indicated by letter groupings (p < 0.005).
Blasia sp. sample was omitted from figures in most cases. Bryophytes, lycophytes, conifers, ferns, and angiosperms had 16 to 43 more ABC transporter transcripts than rhodophytes, chlorophytes, charophytes, glaucophytes, and Chromista. Many observed differences in average number of transporters between plant groups were found to be significant (p < 0.005). Although green algae and land plants have been shown to harbor similar numbers of ABC transporter subfamily gene members and total ABC transporter genes [34], the simpler transport systems of green algae, rhodophytes, glaucophytes, and Chromista may contribute to the reduced number of unique ABC transporter genes we observed on our study. Angiosperms, conifers, ferns, chlorophytes, charophytes, glaucophytes, rhodophytes, and Chromista had 22 to 43 fewer unique ABC transporter gene members than bryophytes. The differences in mean number of transporters between these groups is also significant, with exception to hornworts (p < 0.005).

Results from BLAST subfamily classifications were used to identify full-length transcripts based on query coverage. Because 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 of the same subfamily. Transcripts with greater than 90 % of the query aligned to the target sequence were classified as full length. All 1295 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, 66 ABC transporter gene members discovered per RNA-Seq library with a standard deviation of 26 members on average, 66 ABC transporter gene members detected across all 1295 tissues sampled. This finding would appear to support the reduced expression of ABC transport proteins in an aquatic environment (Additional file 4: Table S3). Algae also do not produce a waxy cuticle - so reduced ABCG transcripts may be related to a lack of need by algae for stomata regulation and/or detoxification. While liverworts also lack stomata we do not see significantly fewer ABCC transporter proteins among liverworts. Algae do not require the specialized transport systems of land plants because most cells have direct access to nutrients in their water environment. Such direct access to an aquatic environment may contribute to fewer unique gene members of subfamilies ABCC and ABCC. Only 26 ABC transport proteins were found in Ceratophyllum demersum, relative to average 66 ABC transport proteins seen across all 1295 tissues sampled. This finding is conferred by the expression of ABCB27 in root tips [38]. ABCC transporter proteins have been identified among plants for their role in detoxification and the regulation of stomatal guard cells [39]. The gap in types of ABCB transporter transcript number seen between algae and other plant taxa might be explained by a lack of need by algae for stomata regulation and/or detoxification. While liverworts also lack stomata we do not see significantly fewer ABCB transporter proteins among liverworts. Algae do not require the specialized transport systems of land plants because most cells have direct access to nutrients in their water environment. Such direct access to an aquatic environment may contribute to fewer unique gene members of subfamilies ABCB and ABCC.

The number of unique ABC transporter subfamily gene members varied between distantly related plants. Among the functionally diverse subfamilies ABCD, ABCE, and ABCI, we saw significantly fewer unique ABC transporter gene members on average in algae (5 to 15 gene members) than in other groups: angiosperms, conifers, ferns, lycophytes, and bryophytes (p < 0.005) (Figs. 2b and c). 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 groups: angiosperms, conifers, and bryophytes (p < 0.005) (Fig. 2d).

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 [5, 35]. Our results revealed additional variability among plants regarding the size of these subfamilies, specifically in that algae contain relatively fewer unique ABCB, ABCC, and ABCG transporter members. The larger number of unique ABCB transporter gene members seen in vascular plant species may have adaptive significance for complex and heterogeneous environments of “higher plants.” Land plants’ best defense against xenobiotics, such as heavy metals, might come from ABCB transporters’ ability to sequester and/or transport foreign chemicals. For example, AtABCB25, an Arabidopsis ABCB export protein, allows some tolerance to cadmium and lead when overexpressed [36, 37]. The overexpression data are useful to infer functional significance, but also guide potential utility in biotechnology. Aluminum tolerance is conferred by the expression of ABCB27 in root tips [38]. ABCC transporter proteins have been identified among plants for their role in detoxification and the regulation of stomatal guard cells [39]. The gap in types of ABCB transporter transcript number seen between algae and other plant taxa might be explained by a lack of need by algae for stomata regulation and/or detoxification. While liverworts also lack stomata we do not see significantly fewer ABCB transporter proteins among liverworts. Algae do not require the specialized transport systems of land plants because most cells have direct access to nutrients in their water environment. Such direct access to an aquatic environment may contribute to fewer unique gene members of subfamilies ABCB and ABCC.
have an evolutionary advantage over plants that do not, given the importance of the glyoxylate cycle to survival across all plant taxa. Rhodophytes have nearly double, on average, the number of ABCD subfamily members compared with other plants. We can speculate that since these ABCD transporters are peroxisome-localized and important for fatty acid metabolism, especially long chain fatty acids, and that Rhodophyta produce a high number of long-chain fatty acids [42], that there may be functional significance. Rhodophytes might be good targets to mine ABCD transporters for use to produce novel oils in crops. ABCE proteins are found in archaea, bacteria, and eukaryotes, suggesting that this class of ABC transporter proteins is essential to life and are thus conserved [5]. ABCE1 is a translation termination factor [43]. Once a ribosome reaches a stop codon, ABCE1 helps to detach the ribosome from the mRNA [44], and also plays an important role in ribosome recycling [45]. The maintenance of similar numbers of unique ABCD and ABCE subfamily gene members may be because of the conservation of essential functions provided by those subfamilies. Our understanding of the ABCI subfamily, is expected to be significantly improved as more information regarding the structure and function of the subfamily members becomes available. However, current data suggest the origins of ABCI subfamily transporter genes may be from prokaryotic genomes and so they may result from a movement of genes from mitochondria and plastids to the nuclear genome. As such, the ABCI transporter proteins have been classified as a heterogeneous group composed of multicomponent transporters (reviewed in [5]) so the similarity seen in average number of unique gene members across plant taxa was unexpected.

In the case of ABCF subfamily proteins, we observed significantly higher transcript count averages in Chromista, glaucophytes, rhodophytes, chlorophytes, charophytes, and bryophytes than in all other groups, four to seven more transcripts on average than angiosperms, conifers, ferns, and lycophytes (p < 0.005) (Fig. 2h). When directly comparing algae and bryophytes, we found that bryophytes had more unique ABCF subfamily gene members on average (p < 0.005). Understanding the increased variation, specifically regarding the ABCF subfamily size among algae and bryophytes is difficult with a lack of functional characterization of ABCF proteins among plants. Studies in yeast and humans have indicated ABCF protein function in the activation of eIF-2α kinase [5]. Activation of eIF-2α kinase is important in the regulation of stress response factors [46]. Indeed one member of the ABCF, ABC50 appears to play an especially important role in mammalian translation processes in interactions with eIF2 and the ribosome [47, 48]. There is a published patent application in which ABC50 is claimed to be useful to increase protein in plants, among other organisms [49]. We noted an increase in the total number of ABCF protein gene members among algae and bryophytes; we cannot speculate about the potential adaptive significance.

The average number of unique ABCA subfamily transporter gene members was significantly lower in ferns, three to four less gene members on average in most cases (p < 0.005) (Fig. 2i). There was a significant difference between the average number of unique ABCA transporter gene members in ferns, conifers, and angiosperms (p < 0.005). While little is known about the function of ABCA transporter proteins among plants, they are suspected to be involved in lipid metabolism based on their known function in humans (reviewed in [50]). Upon comparison with Arabidopsis, rice was found to be missing an orthologous full length ABCA transporter, which is consistent with previously published results [51]. Our results indicate an occurrence of decreased ABCA transporter genes among ferns relative to other plant groups, possibly from gene loss. While the single gene loss observed in rice and the results observed in fern are likely unrelated, they may point to similar phenomena regarding the importance of ABCA transport proteins in sterol metabolism [52].

Conclusions

ABC transport proteins have diverse functions and are ubiquitous in plants. We catalogued ABC transporters across the plant kingdom from hundreds of transcriptomes. While our current study was constrained by imperfect transcriptome assemblies taxonomic patterns should be of interest to the community to build hypotheses. Specifically, we found significant differences in the number of unique ABC transporter gene members between algae (Chromista, rhodophytes, glaucophytes, chlorophytes, and charophytes), bryophytes (mosses, liverworts, hornworts), lycophytes, ferns, conifers (other gymnosperms were not well represented among the 1KP data), and angiosperms. Algae and bryophytes have fewer unique ABC transporter gene members than lycophytes, ferns, conifers, and angiosperms. The number of unique ABCD, ABCE, and ABCI subfamily protein gene members show few significant differences amongst plants. With regards to plant biotechnology applications, it is noted that these three subfamilies might not provide much novelty. In contrast, the diversity of ABCB and ABCF proteins, known for detoxification and stress resistance, respectively, are attractive genes to screen for improving crops via biotechnology. Especially intriguing are the ABCF members in algae and bryophytes for potential use in crop biotechnology. Given the diversity and function of ABCF members, algae and bryophytes have been largely ignored as sources of potentially useful genes in agriculture.
Additional files

Additional file 1: file S1. Fasta file of ABC transporter proteins from rice and Arabidopsis. (FASTA 139 kb)

Additional file 2: Table S1. Detailed table of HMMER testing results. (XLSX 82 kb)

Additional file 3: Table S2. Detailed table of classification results and associated metadata. (XLSX 163 kb)

Additional file 4: Table S3. Filtered BLAST results of transcripts with 90 % or greater query coverage of ABC transporter protein subject. (XLSX 2317 kb)

Acknowledgments

The authors would like to thank Dylan Storey for his suggestions and advice regarding the handling, storage, and analysis of the 1KP data.

Funding

This work was supported by the 1KP consortium.

Availability of data and material

All sequence data used in this paper are made available in the supplements. See 1KP.org for updates on all data on the larger 1KP project.

Authors’ contributions

TSL designed and performed the data analyses and drafted the manuscript. CSR co-conceived the strategy for separating putative ABC transporter transcripts into respective ABC transporter subfamilies and was involved in the collection of known ABC transporter proteins. JD performed statistical analysis based on data. YP provided insights regarding the form and function of ABC transporters among plant species and provided the initial list of known plant ABC transporters. MES contributed strategies for verifying analyses performed. DES, MM, MD, JHL-M, MC, CJR, DS, SWG, JY, TL, JCP, PPE, YZ, YX, YZ, EC, GK-SW, and CNS are members of the 1KP consortium and were involved in 1KP organization and structure, sample collection and sequencing. CNS was partially responsible for the conception of the study and writing. All authors agree with the contents of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

No animals or humans were used in the research.

Author details

1The Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN 37996, USA. 2Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN 37996, USA. 3Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996, USA. 4Department of Biology, University of Florida, Gainesville, FL 32611, USA. 5Florida Museum of Natural History, Gainesville, FL 32611, USA. 6Genetics Institute, University of Florida, Gainesville, FL 32611, USA. 7Department of Biology, Cologne Biocenter, University of Cologne, Cologne 50674, Germany. 8Department of Biology, University of British Columbia, Kelowna, BC V1V 1V7, Canada. 9Department of Biological Sciences, University of Georgia, Athens, GA 30602, USA. 10Jodrell Laboratory, Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AE, UK. 11Department of Biology, Duke University, Durham, NC 27708, USA. 12Department of Integrative Biology, University of California, University Herbarium, Berkeley, CA 94720-2465, USA. 13New York Botanical Garden, Bronx, New York, NY 10458, USA. 14Department of Botany, University of British Columbia, Vancouver, BC, Canada. 15CAS Key Laboratory of Genome Sciences and Precision Medicine Technologies, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, People’s Republic of China. 16College of Marine Life Sciences, Ocean University of China, Qingdao 266003, People’s Republic of China. 17Bond Life Sciences Center, Division of Biological Sciences, University of Missouri, Columbia, MO, USA. 18BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China. 19Department of Computer Science, HKU-BGI Bioinformatics Algorithms and Core Technology Research Laboratory, University of Hong Kong, Pokfulam, Hong Kong. 20School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China. 21Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada. 22Department of Medicine, University of Alberta, Edmonton, AB T6G 2E1, Canada.

Received: 1 March 2016 Accepted: 23 May 2016

Published online: 31 May 2016

References

1. Higgins CF. ABC transporters: from microorganisms to man. Annu Rev Cell Biol. 1992;6:73–113.
2. Jones PM, George AM. The ABC transporter structure and mechanism: perspectives on recent research. Cell Mol Life Sci. 2004;61:882–99.
3. Davidson AL, Dassa E, Oreille C, Chen J. Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol Mol Biol Rev. 2008;72:317–64.
4. Bourie P, Laurent D, Pilyon I, Dassa E. Phylogenetic and functional classification of ATP-binding cassette (ABC) systems. Curr Protein Pept Sci. 2002;3:541–59.
5. Perrier PJ, Bird D, Burla B, Dassa E, Forester C, Geiser M, Klein M, Kolukisaoglu U, Lee Y, Martinoa E et al. Plant ABC proteins - a unified nomenclature and updated inventory. Trends Plant Sci. 2008;13:151–9.
6. Dassa E, Bourie P. The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. Res Microbiol. 2001;152:211–9.
7. Davies TGE, Coleman JD. The Arabidopsis thaliana ABC-binding cassette proteins: an emerging superfamily. Plant Cell Env. 2002;23:431–43.
8. Sanchez-Fernandez R, Davies TG, Coleman JD, Rea PA. The Arabidopsis thaliana ABC protein superfamily, a complete inventory. J Biol Chem. 2001;276:30231–44.
9. Schneider E, Hunke S. ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. FEMS Microbiol Rev. 1998;22:1–20.
10. van den Brule S, Smart CC. The plant PDR family of ABC transporters. Planta. 2002;216:95–106.
11. 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. J Plant Physiol. 2014;171:1276–88.
12. Cakir B, Kilickaya O. Whole-genome survey of the putative ATP-binding cassette transporter family genes in Vitis vinifera. PLoS One. 2013;8(11):e78680.
13. Matasci N, Hung L-H, Yan Z, Carpenter EL, Wickert NJ, Miralbar S, Nguyen N, Warnow T, Ayyampalayam S, Barker ME et al. Data access for the 1,000 Plants (1KP) project. Gigascience. 2014;3:17.
14. Adolfo G, Ruocco M, Di Donato A, Frusciante L, Lorito M, Scala F, Ercolano MR. Genetic variability and evolutionary diversification of membrane ABC transporters in plants. BMC Plant Biol. 2015;15:51.
15. Peng Y, Abercombie LL, Yuan JS, Riggins CW, Sammons RD, Tranp PJ, Stewart JR CN. 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 Manag Sci. 2010;66:1053–62.
16. Ge X, d’Avignon DA, Ackerman JJ, Sammons RD. Rapid vascular sequestration: the horseweed glyphosate resistance mechanism. Pest Manag Sci. 2010;66:345–8.
17. Hanson BD, Shrestha A, Shaner DL. Distribution of glyphosate-resistant horseweed (Conyza canadensis) and its relationship to cropping systems in the Central Valley of California. Weed Sci. 2009;57:486–500.
18. Itos H, Gray WA. A gain-of-function mutation in the Arabidopsis pleiotropic drug-resistance transporter PDR9 confers resistance to auxinic herbicides. Plant Physiol. 2006;142:63–74.
19. Yang H, Murphy AS. Functional expression and characterization of Arabidopsis ABCB, AUX1 and PIN auxin transporters in Schizosaccharomyces pombe. Plant J. 2009;59:179–91.
20. Theodoulou FL, Holdsworth M, Baker A. Peroxisomal ABC transporters. FEBS Lett. 2006;580:1139–55.
21. The One Thousand Transcriptomes Project [https://sites.google.com/a/ulbbera.ca/onekp/]. Accessed 28 May 2016.
22. Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Pudd D, Melake-Berhan A, Jones KM, Redman J, Chen G. Draft genome sequence of the oiledseed species Ricinus communis. Nat Biotechnol. 2010;28:591–6.
et al. BMC Biotechnology

23. 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. Genet Res Crop Evol. 2007;55:365–78.

24. Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datta R et al. The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. Plant J. 2012;72:461–75.

25. 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.

26. Rice P, Longden I, Bleasby A. EMBOSS: The European molecular biology open software suite. Trends Genet. 2000;16:276–7.

27. Crouzet J, Trombik T, Fraysse AS, Boutry M. Organization and function of the plant pleiotropic drug resistance ABC transporter family. FEBS Lett. 2006;580:1123–30.

28. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39(Web Server issue):W29–37.

29. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J. The Pfam protein families database. Nucleic Acids Res. 2012;40(Database issue):D290–301.

30. Phytozome [http://phytozome.jgi.doe.gov/pz/portal.html]. Accessed 28 May 2016.

31. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazio J, Mitros T, Dirks W, Hellsten U, Putnam N, Rohs AR. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40(D1):D1178–86.

32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

33. JMP®, Version 11. SAS Institute Inc., Cary, NC, 1989–2007.

34. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin K, Marechal-Drouard L. The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science. 2007;318:245–50.

35. Geisler M, Murphy AS. The ABC of auxin transport: the role of p-glycoproteins in plant development. FEBS Lett. 2006;580:1094–102.

36. Kim DY, Bovet L, Kushnir S, Neh EW, Martinoa E, Lee Y. AtATM3 is involved in heavy metal resistance in Arabidopsis. Plant Physiol. 2006;140:22–32.

37. 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. J Biol Chem. 2007;282:15611–71.

38. 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;225:1447–58.

39. Klein M, Buria B, Martincia E. The multidrug resistance-associated protein (MRP/ABCC) subfamily of ATP-binding cassette transporters in plants. FEBS Lett. 2006;580:1112–22.

40. 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;306:702–4.

41. Panikashvili D, Savaldi-Goldstein S, Mandel T, Yifhar T, Franke RB, Hofer R, Schreiber L, Chory J, Aharoni A. The Arabidopsis MRP1/AABP1 transporters are involved in cutin and wax secretion. Plant Physiol. 2007;145:1345–56.

42. Lang I, Hodak L, Friedl T, Feussner I. Fatty acid profiles and their distribution in Arabidopsis. Plant J. 2006;46:54–65.

43. Khoshnevis S, Gross T, Rolpe C, Baierlein C, Ficner R, Krebber H. The iron-sulphur protein RNase L inhibitor functions in translation termination. Biochim Biophys Acta. 2006;1762:510–24.

44. Paytubi S, Wang X, Lam YW, Isquierdo L, Hunter MJ, Jan E, Hundal HS, Pighin JA, Zheng H, Balakshin LJ, Goodman IP, Western TL, Jetter R, Kunst L. ABC50 promotes translation initiation in mammalian cells. J Biol Chem. 2009;284:24061–73.

45. Berger SA. Increasing protein production by increasing ABC50 expression or activity. 2012. US Patent Publication US20120094329 A1.

50. Kaminski WE, Pfeiler A, Wenzel JJ. ABC A-subfamily transporters: structure, function and disease. Biochim Biophys Acta. 2006;1762:510–24.

51. Garcia O, Bouige P, Forestier C, Dassa E. Inventory and comparative analysis of rice and Arabidopsis ATP-binding cassette (ABC) systems. J Mol Biol. 2004;343:249–65.

52. Saier MH. A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol Mol Biol Rev. 2000;64:354.