Journal of
Proteomics & Bioinformatics

Sheng et al., J Proteomics Bioinform 2014, 7:1
DOI: 10.4172/0974-276X.1000299

Research Article Open Access

A Genome-Wide Analysis of the AAAP Gene Family in Maize

Lei Sheng#, Lin Deng#, Hanwei Yan, Yang Zhao, Qing Dong, Qian Li, Xiaoyu Li, Beijiu Cheng* and Haiyang Jiang*

School of Life Sciences, Anhui Agricultural University, Hefei, 230036, China
#These authors contributed equally to this work

Abstract

The AAAP (amino acid/auxin permease) genes encode a large family of transporters. A few members of the AAAP family have been characterized in *Arabidopsis thaliana* and rice (*Oryza sativa*). However, little is known about AAAP genes in maize (*Zea mays*). Therefore, we performed a systematic bioinformatics analysis to characterize all AAAP genes in maize, which included analysis of the genome sequence, conserved protein domains, chromosomal locations, phylogenetic relationships, gene duplications, and gene expression profiles. In this study, seventy-one *ZmAAAP* genes were identified and named *ZmAAAP01* to *ZmAAAP71*. The number of AAAP genes in maize is more than the number in *Arabidopsis thaliana* (43) and in rice (58). We found a higher percentage of AAAP gene duplications in the maize genome, which contributed to the expansion of the AAAP gene family. Furthermore, segmental duplications played a major role in the AAAP gene expansion in maize. The AAAP genes are unevenly distributed on the 10 maize chromosomes, and 31 of them are distributed within 12 clusters on segmental chromosomes. Both the EST analysis and transcriptome data indicated that most *ZmAAAP* genes exhibited abundant expression patterns, suggesting diverse and novel functions of AAAP gene family in maize. The results presented here provide useful information for further functional analysis of the AAAP gene family in maize.

Keywords: Maize; AAAP gene family; Phylogenetic analysis; Gene duplication; Chromosomal location; EST analysis

Introduction

Amino acids serve as primary sources of organic nitrogen for the growth of many eukaryotic cells [1]. Additionally, they also serve as neurotransmitters and hormones, allowing communication between cells and tissues within multicellular organisms [2,3]. Some amino acids even allow adaptation to environmental change, especially to those causing organismal stress [4]. These functions require the presence of transport systems that catalyze uptake into, and release from specialized cell types [5].

Amino acid transporters are the principal mediators of organic N distribution, and are important regulators of resource allocation in plants. In recent years, a large number of gene families of amino acid transporters have been identified, such as the amino acid/polyamine/choline (APC) family [6,7], the amino acid transporter gene family (AAT) in rice [8], the lysine-histidine-like transporter family (LHT) in *Arabidopsis* [9] etc. These genes are widely distributed in eukaryotic organisms (ranging from yeast and plants to insects and mammals), and many of these genes have important biological functions in eukaryotes.

The amino acid/auxin permease (AAAP) family is one of the largest families of amino acid transporters identified so far, with members found in virtually all eukaryotic organisms [1,2,10]. The AAAP family belongs to the "Electrochemical Potential Driven Transporters" class, which includes hundreds of proteins from plants, animals, yeast, and other fungi. AUX1/LAX proteins [3,11], belong to the Amino Acid/Auxin Permease (AAAP) family are proton-gradient-driven secondary transporters, where auxin influx is mediated as cotransport with proton(s). Individual proteins in AAAP family are amino acid permeases that transport auxin (indole-3-acetic acid), a single amino acid, or multiple amino acids. Some of these permeases exhibit very broad specificities transporting all twenty amino acids naturally found in proteins. Some also transport D-amino acids. Apparent differences in substrate specificity influenced by structure and net charge are observed within these carriers and may be expected also among AUX1/LAX proteins [12]. Genes in the AAAP family encode transporters with a variety of functions involved in the developmental and physiological processes in plants. For example, *AtAAP1* was highly expressed in the cotyledons and the endosperm and regulated the import of amino acid into root cells or developing embryo [13-15]. *OsAAP8* and *OsAAP15* might participate in the uptake and long-distance transport of amino acid [8].

Maize (*Zea mays*) is an important cereal crop that has also become a model species for the study of genetics, evolution, and other basic biological processes. The availability of maize genome sequence has provided an excellent opportunity for whole-genome annotation, classification, and comparative genomics research [16]. Although the roles of many AAAP genes were revealed in extensive studies of *Arabidopsis thaliana* and rice, none of the *ZmAAAP* genes has been functionally characterized. Therefore, there is an urgent need for a thorough bioinformatics analysis of the *ZmAAAP* gene family. In this work, an attempt was made to find out all AAAP genes existed in maize through a systematic EST and genomic DNA sequence data mining, resulting in the identification of 71 AAAP genes. Then the complete survey of genomic organization, chromosomal location and

***Corresponding authors:** Haiyang Jiang, School of Life Sciences, Anhui Agricultural University, Hefei, 230036, China, Tel: +86-551-65786423; Fax: +86-551-65786021; E-mail: hyjiang@ahau.edu.cn

Beijiu Cheng, School of Life Sciences, Anhui Agricultural University, Hefei, 230036, China, E-mail: beijiucheng@ahau.edu.cn

Received November 15, 2013; **Accepted** January 27, 2014; **Published** January 29, 2014

Citation: Sheng L, Deng L, Yan H, Zhao Y, Dong Q, et al. (2014) A Genome-Wide Analysis of the AAAP Gene Family in Maize. J Proteomics Bioinform 7: 023-033. doi:10.4172/0974-276X.1000299

sequence homology of all *ZmAAAP* genes were performed, as well as the classification of all *ZmAAAP* genes and the phylogenetic tree based on 71 *ZmAAAP* protein sequences, 58 *OsAAAP* protein sequences and 43 *AtAAP* protein sequences. In addition, the complete duplication analysis and predicted expression patterns of *ZmAAAP* genes will be helpful for future investigations into the biological functions of individual *ZmAAAP* genes as well as the evolution of AAAP genes in different plant species.

Materials and Methods

Identification of the maize AAAP gene family and analysis of conserved domains

Maize genome sequences were downloaded from http://www.maizesequence.org/. The sequences of AAAP genes from *Arabidopsis thaliana* were downloaded from The *Arabidopsis* Information Resource (TAIR) website (http://www.arabidopsis.org/). Initially, members of the AAAP family from *Arabidopsis thaliana* were used as query sequences and submitted to the PFAM database (http://pfam.janelia.org/search/sequence) to determine conserved domain motifs. The Hidden Markov Model (HMM) profile of AAAP domains from the Pfam database was then used to identify all possible AAAP genes in the maize genome using the BlastP program (*P* value = 0.001). Second, the Pfam database was used to determine if each of the predicted *ZmAAAP* gene was a member of the AAAP family. To avoid identification of overlapping genes, all of the identified *ZmAAAP* genes were aligned using Clustal W [17] as implemented in MEGA v 4.0 [18]. On the basis of these searches, we were able to identify all members of the AAAP family in maize by using currently available genome databases.

Phylogenetic analysis of maize AAAP genes

The complete protein sequences of maize AAAP genes were merged, and multiple-sequence alignments were performed with Clustal X (version 1.83) [17]. In order to better understand the phylogenetic relationships and group the *ZmAAAP* genes, a phylogenetic tree for all complete *ZmAAAP* protein sequences was constructed with the neighbor-joining (NJ) method using MEGA v4.0 [18]. Default parameters were used in the neighbor-joining (NJ) method. The same methods can be applied to analyze the evolutionary relationships between the AAAP proteins in maize, rice and *Arabidopsis thaliana*. The protein sequences of AAAP genes from rice and *Arabidopsis thaliana* were downloaded from the Phytozome website (http://www.phytozome.net/rice.php; http://www.phytozome.net/arabidopsis.php), and a phylogenetic tree based on 71 *ZmAAAP* protein sequences, 58 *OsAAAP* protein sequences and 43 *AtAAP* protein sequences was then constructed from the protein sequence alignment.

Sequence analysis and duplication patterns of *ZmAAAP* genes

On the basis of the phylogenetic analysis, the conserved motifs in the *ZmAAAP* proteins were characterized using MEME (http://meme.sdsc.edu/meme/meme.html) [19] to identify structural divergence between different subgroups of the AAAP gene family. The number of amino acids (length), molecular weight (MW), isoelectric point (pI) and length of the open reading frame (ORF) for each *ZmAAAP* gene were obtained from Expasy (http://www.expasy.ch/tools). Gene duplication events for the AAAP genes in maize B73 were also investigated. All of the confirmed AAAP genes from the maize genome were aligned using ClustalW and analyzed using MEGA v4.0 on the basis of the phylogenetic tree.

Mapping *ZmAAAPs* on the maize chromosomes

Each non-overlapping *ZmAAAP* gene sequence was used as a query against the maize genome sequence (http://maizesequence.org/blast), analyzed with the TBlastN program, and positioned on the 10 maize chromosomes. *ZmAAAP* gene names are assigned according to their position from the top to the bottom of maize chromosomes 1–10. The chromosome map showing the physical location of all *ZmAAAP* genes was generated with Genome Pixelizer software (http://www.niblrrs.ucdavis.edu/GenomePixelizer/GenomePixelizer_Welcome.html).

EST expression profile analysis of the *ZmAAAP* gene family

The analysis of *ZmAAAP* gene expression profiles was accomplished by searching the maize dbEST database (http://www.ncbi.nlm.nih.gov/dbEST/) and finding expression information provided at the Web sites. Maize gene expression data were first obtained through blast searches against the maize dbEST database downloaded from NCBI conducted with the DNATOOLS Blast program.

Search parameters were as follows: maximum identity >95%, length >200 bp, and E value $<10^{-10}$. In addition to the maize EST database, maize gene expression data were also extracted from the Maize Assembled Genomic Island (MAGI) (http://magi.plantgenomics.iastate.edu/) and the Plant Genomic Database (Plant GBD) (http://www.plantgdb.org/) and included ESTs, cDNAs, and PUTs (PlantGDB unique transcripts).

Constructing three-dimensional models of *ZmAAAP* homologous genes

In order to make primary analysis on structures of *ZmAAAP* homologous genes, we selected *ZmAAAP40* and its three homologous genes as an example to construct three-dimensional models. First, the protein sequences of four homologous genes were submitted to the SWISS-MODEL (http://swissmodel.expasy.org/), respectively.

And the homologous protein sequences were selected for constructing three-dimensional models. Second, the selected protein sequences were submitted to the SWISS-MODEL (http://swissmodel.expasy.org/) again for homology modeling analysis. Then the pdb files of four *ZmAAAP* proteins were downloaded. At last, accurate structures were built using Swiss-Pdb Viewer [20].

Results

Identification of AAAP genes in the maize genome

We used BLASTP searches based on the conserved AAAP domain HMM profile to identify the AAAP genes present in the maize genome. Seventy-one potential AAAP protein sequences were predicted with an E value threshold of 0.001 and identified as AAAP genes using the BLAST program from the Pfam database. The 71 AAAP genes were then designated *ZmAAAP01* to *ZmAAAP71*. Descriptive information for the maize AAAP genes included the sequence ID, amino acid length, predicted MW, predicted PI, number of *ZmAAAP* motifs, and the physical locations of the genes on the chromosomes. The predicted PIs for seventy-one *ZmAAAP* gene products are below 11.0, and most of them are about 9.0. Only two gene products have PIs below 5.0. The predicted lengths of the *ZmAAAP* ORFs ranged from 396 bp (*ZmAAAP19*) to 3807 bp (*ZmAAAP67*) and the predicted molecular weights ranged from 13.96 kDa (*ZmAAAP19*) to 140.21 kDa (*ZmAAAP68*). All of this information is listed in Table 1.

Gene name	Sequence ID	Chr	ORF length (bp)	Length (aa)	MW (kDa)	PI
ZmAAAP01	GRMZM2G149481_P01	1	1563	520	57.98	9.14
ZmAAAP02	GRMZM2G129413_P01	1	1713	570	62.74	9.01
ZmAAAP03	GRMZM2G017170_P01	1	1794	597	64.6	5.89
ZmAAAP04	GRMZM2G052461_P01	1	1335	444	49.49	8.59
ZmAAAP05	GRMZM5G894432_P01	1	891	296	31.29	5.92
ZmAAAP06	GRMZM2G042933_P01	2	1668	555	60.28	9.17
ZmAAAP07	GRMZM2G175321_P01	2	1317	438	48.42	9.16
ZmAAAP08	GRMZM2G109865_P01	2	1584	527	56.75	9.56
ZmAAAP09	GRMZM2G155491_P01	2	1377	458	50.37	8.38
ZmAAAP10	GRMZM2G114523_P01	2	1212	403	43.48	9.15
ZmAAAP11	GRMZM2G137161_P01	2	1212	403	42.4	9.34
ZmAAAP12	GRMZM2G097802_P03	2	1722	573	62.02	5.87
ZmAAAP13	GRMZM2G108597_P01	2	1395	464	49.54	8.92
ZmAAAP14	GRMZM2G082434_P02	2	1440	479	51.93	8.8
ZmAAAP15	GRMZM2G125832_P01	3	1527	508	54.88	8.97
ZmAAAP16	GRMZM2G092223_P01	3	1716	571	61.28	5.84
ZmAAAP17	GRMZM2G032304_P01	3	1008	335	36.55	9.03
ZmAAAP18	GRMZM2G031167_P01	3	1500	499	53.52	8.84
ZmAAAP19	GRMZM2G136288_P01	3	396	131	13.96	4.42
ZmAAAP20	GRMZM2G078024_P01	3	1341	446	48.83	9.27
ZmAAAP21	GRMZM2G110195_P01	3	1485	494	53.61	8.58
ZmAAAP22	GRMZM5G830545_P03	3	1407	468	50.07	8.86
ZmAAAP23	GRMZM2G154958_P01	3	1380	459	48.86	9.08
ZmAAAP24	GRMZM2G127949_P01	3	1473	490	54.46	8.63
ZmAAAP25	GRMZM2G057733_P01	3	1404	467	49.59	9.22
ZmAAAP26	GRMZM2G476954_P01	3	759	252	26.81	6.4
ZmAAAP27	GRMZM2G066428_P01	3	1497	498	52.89	9.07
ZmAAAP28	GRMZM2G046743_P01	4	1368	455	50.59	9.06
ZmAAAP29	GRMZM2G180547_P01	4	1440	479	51.12	8.79
ZmAAAP30	GRMZM2G427319_P01	4	1599	532	57.34	9.1
ZmAAAP31	GRMZM2G101125_P01	4	1623	540	58.72	5.42
ZmAAAP32	GRMZM2G045057_P01	4	1458	485	53.87	8.88
ZmAAAP33	GRMZM2G332505_P02	4	1419	472	50.77	8.6
ZmAAAP34	GRMZM2G332562_P01	5	1356	451	49.08	6.19
ZmAAAP35	GRMZM2G074053_P01	5	1299	432	45.48	6.99
ZmAAAP36	GRMZM2G161641_P01	5	1422	473	50.58	7.53
ZmAAAP37	GRMZM2G083788_P01	5	1632	543	59.02	5.08
ZmAAAP38	GRMZM2G080843_P01	5	1587	528	57.25	4.88
ZmAAAP39	GRMZM2G180659_P01	6	1359	452	50.2	9.14
ZmAAAP40	GRMZM2G127342_P01	6	1314	437	47.83	8.66
ZmAAAP41	GRMZM2G127328_P01	6	1419	472	51.98	8.96
ZmAAAP42	GRMZM2G127294_P01	6	1365	454	50.7	9.36
ZmAAAP43	GRMZM2G127338_P01	6	1407	468	52.31	9.04
ZmAAAP44	GRMZM2G429322_P01	6	1482	493	54.61	9.1
ZmAAAP45	AC205362.4_FGP002	6	1455	484	52.12	8.48
ZmAAAP46	GRMZM2G331283_P02	6	1410	469	50.69	5.97
ZmAAAP47	GRMZM2G134888_P01	6	1668	555	60.22	5.94
ZmAAAP48	GRMZM2G010433_P01	6	1200	399	43.24	9.2
ZmAAAP49	GRMZM2G476886_P01	6	1341	446	49.1	9.29
ZmAAAP50	GRMZM2G067022_P02	6	1956	651	72.48	10.07
ZmAAAP51	GRMZM5G894233_P02	7	1299	432	46.93	8.9
ZmAAAP52	GRMZM2G108023_P01	7	1452	483	52.12	8.8
ZmAAAP53	GRMZM2G173967_P01	7	1287	428	45.44	6.54
ZmAAAP54	GRMZM2G164814_P01	8	1437	478	51.98	8.51
ZmAAAP55	GRMZM2G092945_P01	8	1386	461	49.5	8.79
ZmAAAP56	GRMZM2G096407_P01	8	1455	484	50.84	9.04
ZmAAAP57	GRMZM2G036448_P01	8	1413	470	49.23	9.54
ZmAAAP58	GRMZM2G150406_P01	9	1434	477	51.42	5.98
ZmAAAP59	GRMZM2G145989_P01	9	1416	471	51.28	8.77
ZmAAAP60	GRMZM2G136300_P01	9	1461	486	52.95	8.9

ZmAAAP61	GRMZM2G087635_P01	9	1464	487	51.73	6.13
ZmAAAP62	GRMZM2G177659_P01	9	1380	459	49.89	6.06
ZmAAAP63	GRMZM2G105192_P05	9	1476	491	53.74	8.69
ZmAAAP64	GRMZM2G076593_P01	10	1443	480	51.54	8.27
ZmAAAP65	GRMZM2G433162_P01	10	1434	477	51.24	8.61
ZmAAAP66	GRMZM2G157168_P01	10	1449	482	51.67	8.46
ZmAAAP67	GRMZM2G413943_P01	10	3807	1268	140.21	5.35
ZmAAAP68	GRMZM2G149216_P01	10	1341	446	49.65	9.09
ZmAAAP69	GRMZM2G173597_P02	10	1407	468	50.77	8.72
ZmAAAP70	GRMZM2G455128_P01	10	1488	495	52.4	8.5
ZmAAAP71	GRMZM2G360519_P01	10	1554	517	55.73	9.33

ORF=open reading frame; MW=molecular weight; PI=isoelectric point

Table 1: Characteristics of the predicted AAAP proteins in maize.

Phylogenetic analysis of *ZmAAAP* genes

To examine the phylogenetic relationships of the AAAP genes and analyze the evolution of each subfamily, a phylogenetic tree of all full-length maize AAAP protein sequences was generated by the neighbor-joining method using the MEGA program (Figure 1). The phylogenetic tree showed that the 71 *ZmAAAP* genes are distributed into 30 sister groups in eight major clades, with a common ancestor and no additional descendants.

In order to better investigate the phylogenetic relationships between the same gene families in different species, we compared maize with rice and *Arabidopsis thaliana*. The phylogenetic tree derived from 71 *ZmAAAP*, 58 *OsAAAP* and 43 *AtAAP* sequences were aligned and constructed (Figure 2). The phylogenetic tree showed that members of AAAP gene family in three species share a high level of protein sequence homology. Moreover, the major clades of maize and rice are highly unified.

Analysis of conserved motifs and classification of the maize AAAP gene family

The results of the Pfam analysis showed that 71 putative *ZmAAAP* protein sequences contained a typical Aa trans domain. We then aligned the amino acid sequences of *ZmAAAP* proteins using Clustal X, and the *ZmAAAP* protein conserved motifs were defined by submitting their full-length amino acid sequences to MEME [7]. Subsequently, 20 motif sequences were identified and shown in Table 2. Besides, the distributions of these motifs in *ZmAAAP* genes were shown in Figure 3. Analysis of the conserved motifs indicated that 69 of 71 *ZmAAAP* proteins contain motif 5, 68 of 71 *ZmAAAP* proteins contain motif 1, and 68 contain motif 4. Conserved motifs can provide evidence for classification of the maize AAAP gene family. It is possible that proteins within a subgroup that share identical motifs are likely to possess similar functions, and this could be helpful in predicting the function of these proteins.

In order to better analyze the phylogenetic relationships between the same gene families in different species, we submitted the maize and *Arabidopsis* AAAP proteins to MEME(http://meme.sdsc.edu/meme/meme.html). Twenty putative conserved motifs were identified in the maize and *Arabidopsis* AAAP proteins (Table 3). The distributions of these motifs in the maize and *Arabidopsis* AAAP proteins were shown in Figure 4. This figure has shown that the maize and *Arabidopsis* AAAP proteins in the same subgroups have similar distributions of motifs. The comparison indicated that some maize and *Arabidopsis* AAAP genes in the same subgroups are likely to have similar functions.

The AAAP family proteins in maize were classified into eight distinct subfamilies based upon their sequence composition and phylogenetic relationship, including amino acid permeases (AAPs), lysine and histidine transporters (LHTs), proline transporters (ProTs), GABA transporters (GATs), auxin transporters (AUXs), aromatic and neutral amino acid transporters (ANTs) and amino acid transporter-like (ATL) subfamilies. ATL subfamilies consist of two phylogenetic clades that are named as ATLa and ATLb, respectively.

Chromosomal locations of the *ZmAAAP* genes

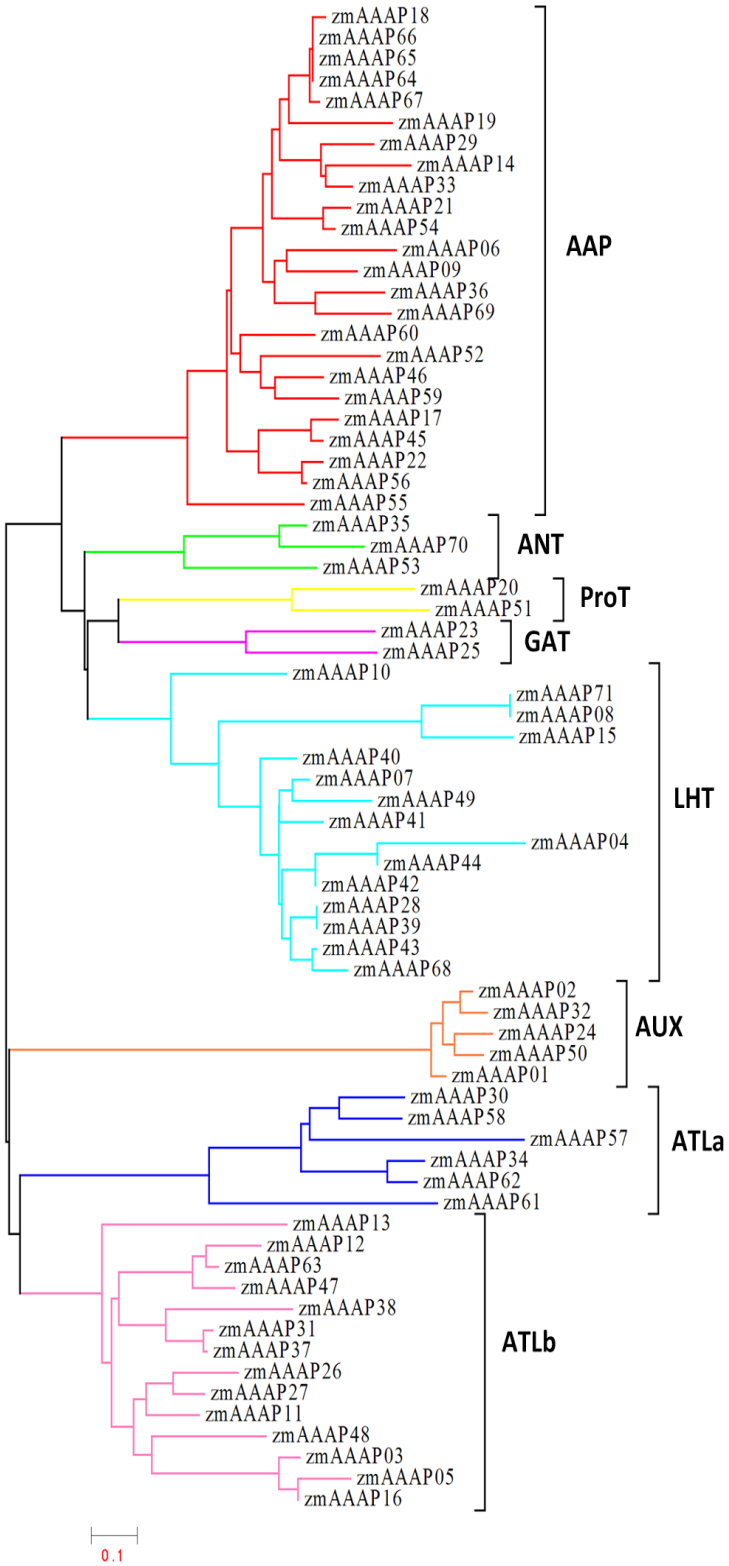
Based on available information (http://www.maizegenome.org/data_portal.html), the identified *ZmAAAP* genes were positioned on the maize chromosomes using Genome Pixelizer software. All 71 *ZmAAAP* genes mapped to the 10 maize chromosomes. There are thirteen AAAP genes located on chromosome 3, 12 genes on chromosome 6, nine genes on chromosome 2, and eight genes on chromosome 10. Six genes each mapped to chromosomes 4 and 9, and there are five genes each on chromosomes 1 and 5. Four AAAP genes mapped to chromosome 8, and three genes to chromosome 7 (Figure 5).

In addition, several *ZmAAAP* genes are clustered together on segmental chromosomes. A gene cluster is a chromosome region containing two or more genes that encode the same or similar products, and gene clusters are useful for tracing recent evolutionary history [21,22]. We found that 31 of the *ZmAAAP* genes are located within 12 clusters, with chromosomes 2, 3, 5, 6, 7, 8, 9 and 10 containing 1, 3, 1, 1, 1, 1, 2, and 2 gene clusters, respectively; the largest gene cluster contained six *ZmAAAP* genes.

Based on the distribution of AAAP genes on the maize chromosomes, we found that the duplication of AAAP genes included tandem and segmental duplication events (Figure 5). Tandem duplication events accounted for less than 10% of the total number, while segmental duplication accounted for the majority of the duplication events. The distribution of AAAP genes on the rice chromosomes has been reported. Tandem duplication and segmental duplication took the similar proportion in rice AAAP family. The result suggested that large-scale segmental and tandem duplication events played a significant role in the expansion of the rice AAAP family [8]. Therefore, the duplications of *ZmAAAP* genes may play an important role in a succession of maize genomic rearrangements and expansions.

Expression patterns of the *ZmAAAP* gene family

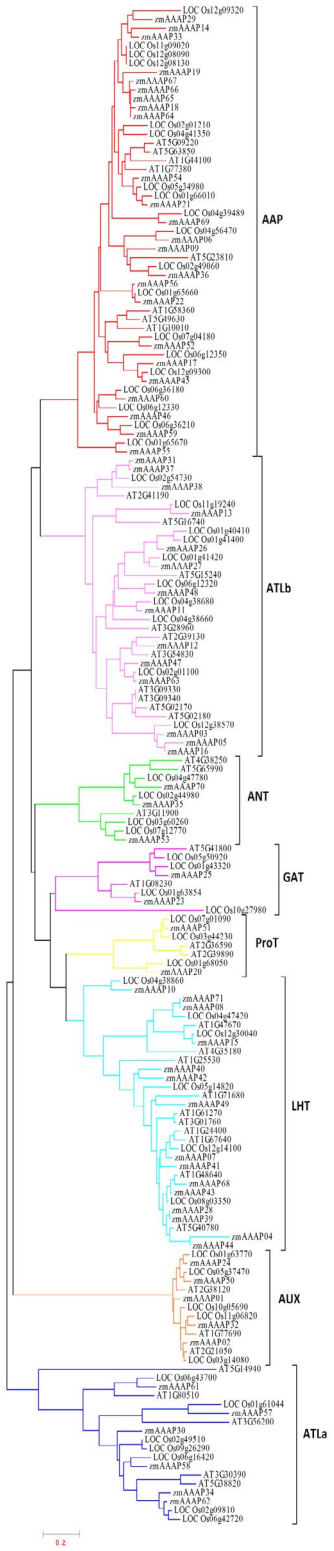
Since EST data can provide valuable information for gene expression research, we examined the expression patterns of maize AAAP genes in various tissues and organs using the NCBI EST database. We divided



Abbreviations: AAP: Amino Acid Permeases; LHT: Lysine and Histidine Transporters; ProT: Proline Transporters; GAT; GABA Transporters; AUX: Auxin Transporters; ATL: Amino Acid Transporter-Like; ANT: Aromatic and Neutral Amino Acid Transporters

Figure 1: Phylogenetic tree showing the evolutionary relationships of the maize AAAP genes. The phylogenetic tree was generated using the neighbor-joining method as implemented in the MEGA v4.0 program.

Citation: Sheng L, Deng L, Yan H, Zhao Y, Dong Q, et al. (2014) A Genome-Wide Analysis of the AAAP Gene Family in Maize. J Proteomics Bioinform 7: 023-033. doi:10.4172/0974-276X.1000299



Abbreviations: AAP: Amino Acid Permeases; LHT: Lysine and Histidine Transporters; ProT: Proline Transporters; GAT: GABA Transporters; AUX: Auxin Transporters; ATL: Amino Acid Transporter-Like; ANT: Aromatic and Neutral Amino Acid Transporters

Figure 2: Phylogenetic tree of AAAP genes in maize, rice and *Arabidopsis thaliana*. The unrooted tree was generated using the MEGA v4.0 program with the neighbor-joining method.

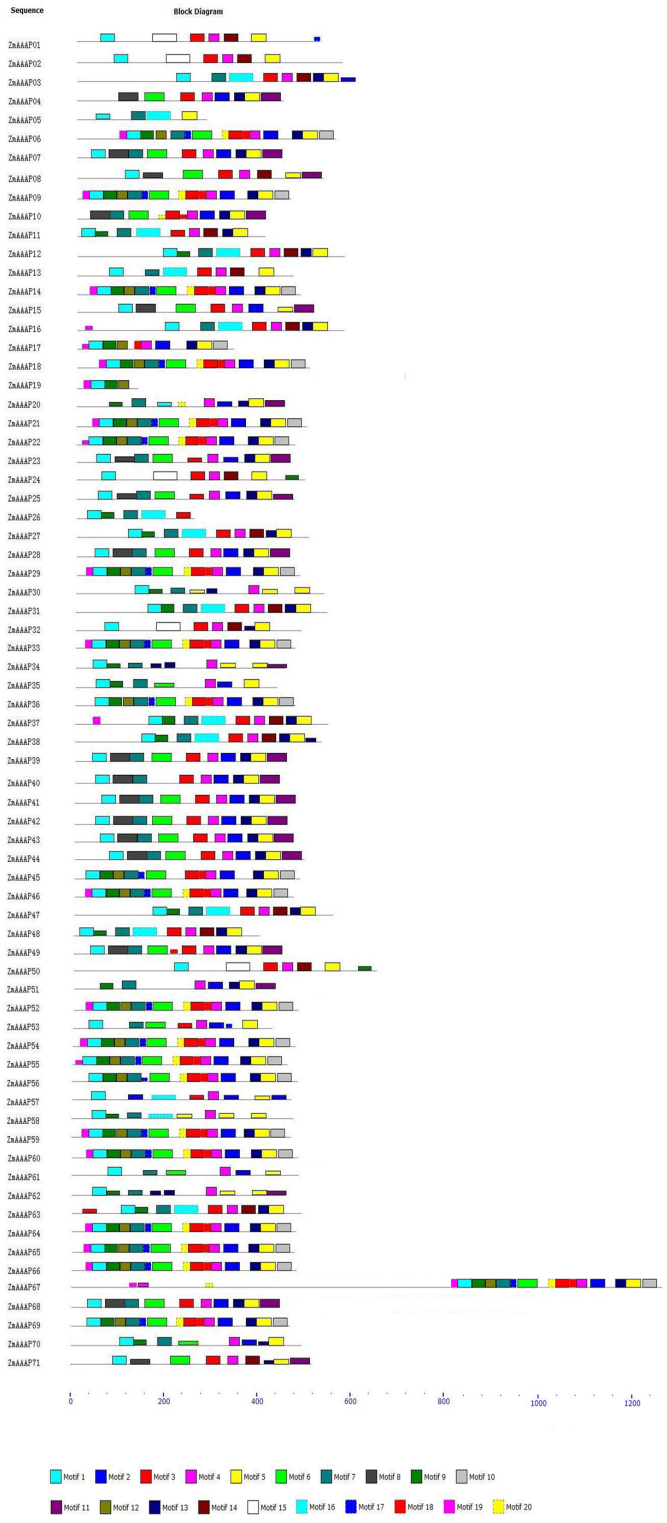


Figure 3: Twenty putative conserved motifs were identified in the maize AAAP proteins using the MEME search tool. Different motifs are indicated by different colors and the gene names are indicated on the left side of the figure. The length of motif in each protein represents the actual length and motif sizes are indicated at the bottom of the figure.

Figure 4: Twenty putative conserved motifs were identified in the maize and *Arabidopsis* AAAP proteins using the MEME search tool. Different motifs are indicated by different colors and the names of all members are shown on the left side of the figure. The markers next to the gene names represent the subgroups to which each AAAP gene belongs. The length of motif in each protein represents the actual length and motif sizes are indicated at the bottom of the figure.

the maize EST data into 16 tissue groups: mixed, silks, apex, husks, shoot tips, leaf, suspension culture, pedicel, root, endosperm, seedling, tassel, embryo, pollen, ear, and multiple tissues (Table 4 data included as supplementary). According to the NCBI EST expression database, 10 *ZmAAAP* genes are not expressed, *ZmAAAP40*-specific mRNA was found only in the embryo, and mRNA from the remaining *ZmAAAP* genes were identified in two or more tissues and organs.

Although the digital EST expression analyses provided a first glimpse of the patterns of AAAP gene expression in maize, we could not draw conclusive inferences regarding AAAP gene expression. Therefore, to gain more insight into the expression patterns of the AAAP genes, a comprehensive expression analysis was further performed by using the publicly available transcriptome data for maize. Firstly, 18 tissues representing five organs have been analyzed using RNA sequencing (RNA-Seq) [21]. Distinct expression profiles were identified for all 71 AAAP genes and the transcriptome data were shown in Table 5 (Data included as supplementary). Then the transcriptome data for the AAAP genes were imported into R and Bioconductor (http://www.bioconductor.org/) for expression analysis. We used the heat map package to make the heat map (Figure 6). Based on the results of heat map, the majority of AAAP genes showed distinct multiple expression patterns across the 18 tissues examined. The *ZmAAAP40* gene was distinctly expressed only in the embryo. The transcriptome analysis of the expression patterns for all AAAP genes was consistent with the above EST analysis (Table 4 data included as supplementary). So far, there are many literatures that have reported the expression profiles of some AAAP genes in other species, such as the microarray data of gene expression for AAAP genes in rice under various abiotic stresses [8], AAAP transporters were significantly upregulated by FA treatment [23], Combining these results to our transcriptome analysis, we could find that the AAAP genes showed multiple expression patterns in many species.

We selected *ZmAAAP40* and its three homologous genes as an example to construct three-dimensional models (Figure 7). By comparing their structures, we have found that homologous genes in the same subfamily have different features. The protein structure of *ZmAAAP40* only contains helix and coil, as well as the protein structure of *ZmAAAP04*. However, the protein structures of *ZmAAAP42* and *ZmAAAP44* not only contain helix and coil, but also contain strands. Although *ZmAAAP04* has the similar protein structure with *ZmAAAP40,* their expression profiles are of great differences. In addition, *ZmAAAP04* and *ZmAAAP44* were duplicated genes, but their protein structures were of great differences.

Duplication Contributed to Functional Divergence of *ZmAAAP* gene Family

When gene duplication occurs, expression patterns and original functions of these genes may be retained [24]. Consistent with that, the comparison analysis for expression pattern of paralogous *ZmAAAP* genes involved in duplication indicates that *ZmAAAP08* and *ZmAAAP71*, and *ZmAAAP46* and *ZmAAAP59*, and *ZmAAAP43*and *ZmAAAP68*, and *ZmAAAP02* and *ZmAAAP32*, and *ZmAAAP05* and *ZmAAAP16*, localized on segmental duplication, exhibited similar expression patterns in development stages. However, it was well known that a great degree of expression and functional divergence might be present between two duplicated genes due to the intense selection pressure and the need for the diversification [25]. Most duplicated *ZmAAAP* genes exhibit distinct expression patterns, such as *ZmAAAP04*

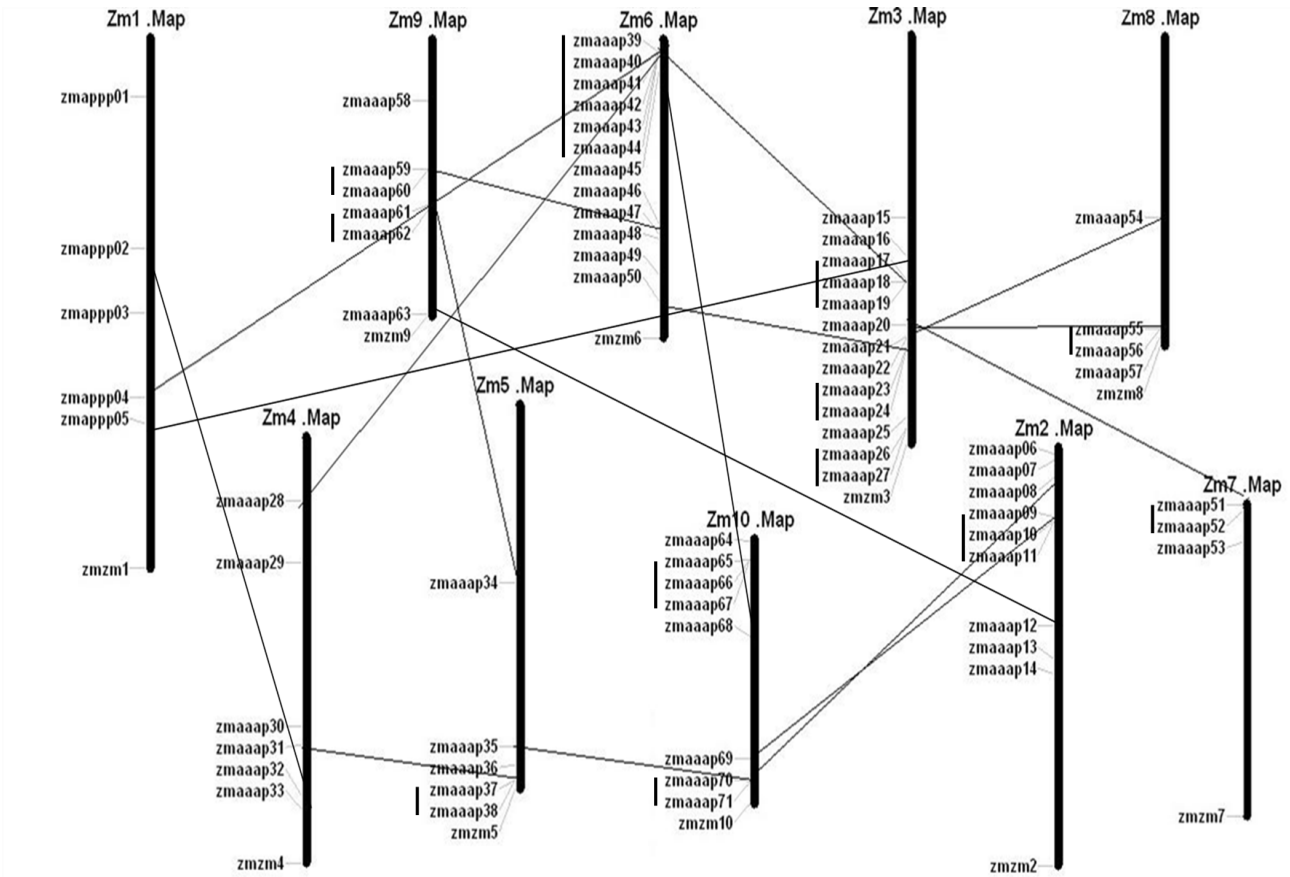


Figure 5: Genomic distribution and duplications of the AAAP genes on the 10 chromosomes of maize. The chromosome numbers are shown at the top of each chromosome (vertical black bars). The names on the left side of each chromosome correspond to the approximate location of each AAAP gene. The thick black lines on the left side of the names of the AAAP genes indicate the clusters of genes on each chromosome. The thin black lines connect the AAAP paralogs lying on duplicated chromosomal segments.

MOTIF	WIDTH	BEST POSSIBLE MATCH
1	29	WTAWFHNITAMIGSGVLSLPWAMKQLGWV
2	30	WLIDVANMCIVVHLVGAYQVYCQPIFDMIE
3	29	GDIAFAYSYHNVLIEIQDTMKSPPKEPSK
4	21	YMPCGCMGYWAFGDDTPDNIL
5	31	PFFNDVMGFFGSFSFWPTTYYFPCEMYLKQY
6	41	TYYMMIFGSVQIVFSQIPNFHQIWWLSIVAAVMSFTYSTIG
7	29	CGPQQYVNLYGCCIEYMITEGDSMKKIHR
8	41	WVITLYTLWQMVEMHEMVPGKRFDRYHELGQHAFGERLGLW
9	27	GPAAMLLFACVTYYTSTLLADCYRTGD
10	29	STRWICLQTLSVVCLLVSIAAAAGSIEDV
11	41	FSLSWFTNWICIIIGVLLMVLSPIGGLRQIILDAKTYKFYQ
12	21	TGKRNYTYMDAVRSNLGGKKV
13	21	NPFRLCWRTAYVCFTTFVAMT
14	29	KIAIWTTVINPFTKYALTCTPLYFSWEEL
15	50	NIYYINDRLDKRTWTYIFGACCATTVFIPSFHNYRIWSFLGLVMTTYTAW
16	50	HLNGHQLFAILTALVVLPTTWLRDLSCLSYLSAGGVIASILVIVCLFWVG
17	11	CFHEHGHDDPC
18	15	MKKATMYSVATTTIF
19	15	EWLDDDGRPRRTGTM
20	15	DVTPTQKVWHTLQAF

Table 2: Twenty putative conserved motifs were identified in the maize AAAP family using MEME search tool.

Motif	Width	Best possible match
1	29	FYEPFWLIDFANMCIVVHLIGAYQVYCQP
2	21	ITAMIGSGVLSLPWAMKQLGW
3	21	YMPCACMGYWAFGDDTPDNIL
4	50	MSWVITLYTLWQMVEMHEMVPGKRFDRYHELGQHAFGEKLGLWIVVPQQL
5	50	FNALGDVAFAYAGHNVVLEIQATIPSTPEKPSKKPMWKGVVVAYIVVAIC
6	21	FPFFNDVMGLFGSFSFWPTTY
7	29	QIVLSQIPNFHQIWWLSIVAAVMSFTYST
8	29	AIGNYAFCYSYHNVFPEIQDTMKKPPKEN
9	21	VNLYGCCIEYMITEGDNMKKI
10	21	CMLLFACVTYYTSTLLKDCYR
11	15	YFPCEMWIKQKKPKR
12	39	DSHQLFAILTALIVLPTVWLRDLSVLSYLSAGGVIASIL
13	28	LNLPQHKFWSKIAVWTTVINPFTKYALT
14	15	DPITGKRNYTYMDAV
15	29	FSLSWCTNWICIILGVLLMILSPIGGLRQ
16	41	STRWICLQMLSVVCLLVSVAAAAGSIQDVIDDLKVYKPFKT
17	15	RLIWRTAYVCFTTFV
18	50	CCATTVFIPSFHNYRIWSFLGLGMTTYTAWYLTIASIIHGQVEGVKHSGP
19	15	RPKRTGTVWTASAHI
20	21	KPKIKTYPDIGQAAFGRTGRI

Table 3: Twenty putative conserved motifs were identified in the maize and *Arabidopsis* AAAP proteins using the MEME search tool.

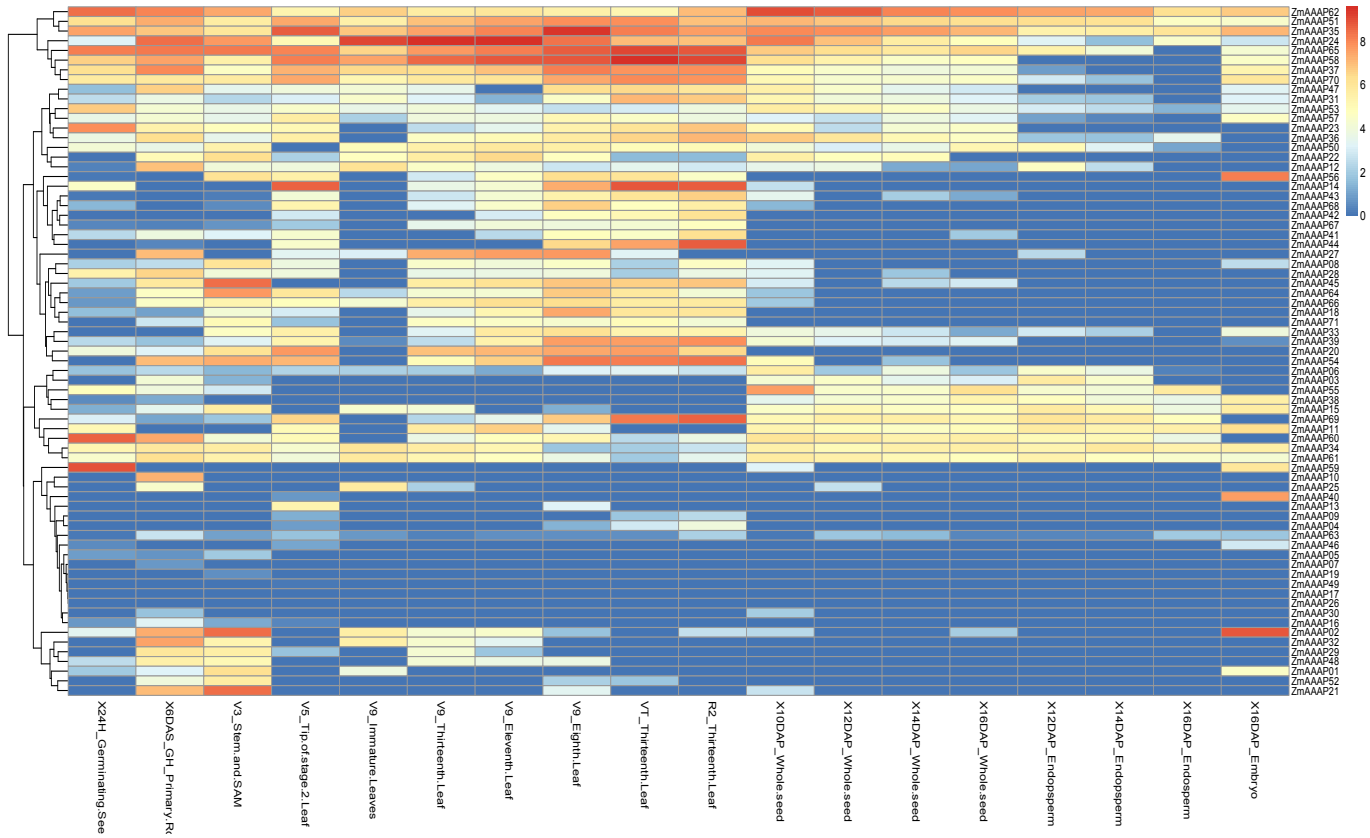


Figure 6: Expression profiles of maize AAAP genes. The heat map was generated by hierarchical clustering using heat map package. The expression data were gene-wise normalized and hierarchically clustered with average linkage. The color scale in the top right corner represents the relative gene expression levels: red, yellow and blue indicate high, medium and low levels of gene expression, respectively.

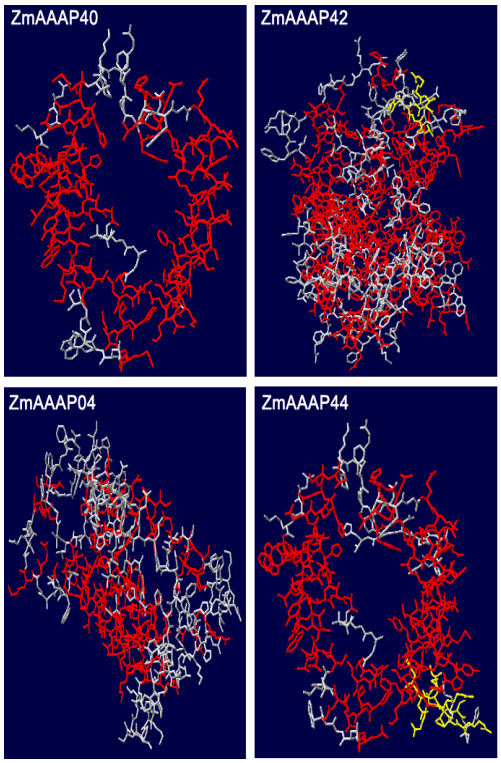


Figure 7: Three-dimensional models of four *ZmAAAP* homologous genes. Red, yellow and gray indicate helix, strands and coil of protein structure, respectively.

and *ZmAAAP44*, *ZmAAAP28* and *ZmAAAP39*, *ZmAAAP24* and *ZmAAAP50*, *ZmAAAP21* and *ZmAAAP54*, etc. These results suggested that chromosomal duplication events not only facilitated the expansion of the *ZmAAAP* gene family, and also leaded to expression divergences between duplicate genes, further contributing to the establishment of gene functional diversity during their evolution.

Discussion

Numerous trans-membrane amino acid transporters have been characterized physiologically and genetically [10,26]. The amino acid/auxin permease (AAAP) proteins constitute a major eukaryotic-specific superfamily of amino acid transport proteins. These are secondary carrier proteins and their activity requires an independently established electrochemical gradient [10,26]. As an important gene family, the eukaryotic-specific amino acid/auxin permease (AAAP) family plays an important role in various aspects of plant growth and development. Significant progress has been made toward the identification and characterization of AAAP gene family in model plants, but little attention has been paid to AAAP gene family in maize.

In previous reports, several AAT genes in *Arabidopsis* were classified and functionally characterized in detail, such as genes in the *AtAAP* subfamily [15,27,28] and the *AtAUX* subfamily [3,29]. However, members of AAAP family are still unknown so far in maize, and none of them is functionally characterized. In this study, we identified 71 *ZmAAAP* genes which were divided into eight subfamilies based on amino acid sequence similarity, and there was obvious difference in numbers among subfamilies. The largest *ZmAAAP* subfamily number was 24, and the smallest *ZmProT* and *ZmGAT* subfamilies were only two (Figure 1). In addition, the number of AAAP genes in maize is more than the number in *Arabidopsis* (43) and in rice (58). The increase in the number of *ZmAAAP* members contributed to the expansion of the AAAP gene family.

The phylogenetic tree describing the evolutionary relationships between AAAP genes in maize, rice and *Arabidopsis thaliana* showed that maize shared more homology with rice, which maybe result from that both maize and rice are monocotyledons. The result will pave the way for further bioinformatics analysis of the evolutionary relationships between the AAAP genes in maize and other plant species. Chromosomal mapping of *ZmAAAP* genes showed their variable distribution on 10 maize chromosomes. The most members are localized on chromosome 3 and the least members are localized on chromosome 7. Meanwhile, we found that gene duplication was common in *ZmAAAP* genes. Gene duplication is thought to be an important means of gene family expansion and functional diversity during evolution, which may occur through three major pathways: chromosomal segmental duplication, tandem duplication and retroposition [25,30,31]. Most gene families in maize contain gene duplications which include tandem and segmental duplications. These duplications play a major role in the expansion of gene families, such as MADS-box genes [32,33], ARF genes [34]. In this study, our analysis on gene duplication reveals that 37 of 71(55.30%) *ZmAAAP* genes are duplicated genes, 32 genes (86.49%) are involved in the segmental duplication and 5 genes (13.51%) in tandem duplication. The analysis indicated that segmental duplication contributes most to the expansion of the *ZmAAAP* gene family, which was different with rice. In rice, the segmental and tandem duplications contribute almost equally to the expansion of the *OsAAAP* gene family [8].

Most *ZmAAAP* genes were found to be expressed in multiple tissues, *ZmAAAP* genes may have different expression patterns in different species. For example, in *Arabidopsis*, *AtAUX1* was primarily expressed in root and promoted lateral root formation [35,36], however in rice, *OsAUX1*, an ortholog of *AtAUX1*, had high expression levels in all the organs, especially in young root and panicle [8]. Our data showed that *ZmAAAP02*, an ortholog of *AtAUX1* was primarily expressed in embryo and shoot apical meristem, and *ZmAAAP24,* an ortholog of *OsAUX1 was* expressed in multiple tissues, especially in leaves. By comparing expression profiles of AAAP genes in maize, rice and *Arabidopsis*, we found that the majority of AAAP genes in each species have multiple expression profiles. Therefore, the AAAP gene family may play an important role in growth and development of different species. In this study, we have also shown that *ZmAAAP40* may be an embryo-specific gene by combining EST with transcriptome data, which provided a useful reference for cloning and further functional analysis. Meanwhile, we analyzed the three-dimensional models of *ZmAAAP40* and its three homologous genes. Our analyses suggested that homologous genes may have great difference of protein structures, which might be result in different expression profiles of homologous genes. We have found that the phenomenon that homologous genes have different expression profiles also exists in other species. For example, the five pairs of paralogous genes (*OsAAP11* and *OsAAP16*, *OsGAT1* and *OsGAT2*, *OsANT1* and *OsANT2*, *OsANT3* and *OsANT4*, *OsATL5* and *OsATL6*) have divergent expression patterns in rice [8].

Our study suggests that most maize AAAP genes play functional roles in more than one type of tissue. Therefore, the AAAP gene family likely plays an important role in maize developmental processes. The comparative and phylogenetic analyses of the AAAP genes in maize are first step towards a comprehensive functional characterization of this important gene family.

Conclusion

In conclusion, the results of this study display the analysis of the genome sequence, classification, chromosomal locations and conserved motifs of the 71 *ZmAAAP* members, along with their expression profiles. All 71 *ZmAAAP* members were divided into eight clades. The AAAP genes were unevenly distributed on 10 chromosomes, and their diverse sequence features provided potential evidence for diversifying functions. In addition, our survey showed that *ZmAAAP* genes exhibit various expression patterns. Our study will provide an insight into further understanding of functions of AAAP genes and their roles in maize growth and development. Overall, our findings presented here can serve as useful information for guiding future experimental work and understanding the structure-function relationship of the members in maize AAAP gene family.

References

1. Fischer WN, Kwart M, Hummel S, Frommer WB (1995) Substrate specificity and expression profile of amino acid transporters (AAPs) in Arabidopsis. J BiolChem 270: 16315-16320.
2. Clark JA, Amara SG (1993) Amino acid neurotransmitter transporters: structure, function, and molecular diversity. Bioessays 15: 323-332.
3. Bennett MJ, Marchant A, Green HG, May ST, Ward SP, et al. (1996) Arabidopsis AUX1 gene: a permease-like regulator of root gravitropism. Science 273: 948-950.
4. Rentsch D, Hirner B, Schmelzer E, Frommer WB (1996) Salt stress-induced proline transporters and salt stress-repressed broad specificity amino acid permeases identified by suppression of a yeast amino acid permease-targeting mutant. Plant Cell 8: 1437-1446.
5. Williams LE, Bick JA, Neelam A, Weston KN, Hall JL (1996) Biochemical and molecular characterization of sucrose and amino acid carriers in *Ricinuscommunis*. J Exp Bot 47 Spec No: 1211-1216.
6. Reizer J, Finley K, Kakuda D, MacLeod CL, Reizer A, et al. (1993) Mammalian integral membrane receptors are homologous to facilitators and antiporters of yeast, fungi, and eubacteria. Protein Sci 2: 20-30.
7. Reizer J, Reizer A, Saier MH Jr (1994) A functional superfamily of sodium/solute symporters. BiochimBiophysActa 1197: 133-166.
8. Zhao H, Ma H, Yu L, Wang X, Zhao J (2012) Genome-wide survey and expression analysis of amino acid transporter gene family in rice (*Oryza sativa* L.). PLoS One 7: e49210.
9. Foster J, Lee YH, Tegeder M (2008) Distinct expression of members of the LHT amino acid transporter family in flowers indicates specific roles in plant reproduction. Sex Plant Reprod 21: 143-152.
10. Young GB, Jack DL, Smith DW, Saier MH Jr (1999) The amino acid/auxin:protonsymportpermease family. BiochimBiophysActa 1415: 306-322.
11. Swarup K, Benková E, Swarup R, Casimiro I, Péret B, et al. (2008) Theauxin influx carrier LAX3 promotes lateral root emergence. Nat Cell Biol 10: 946-954.
12. Fischer WN, Loo DD, Koch W, Ludewig U, Boorer KJ, et al. (2002) Lowand high affinity amino acid H+-cotransporters for cellular import of neutral and charged amino acids. Plant J 29: 717-731.
13. Hirner B, Fischer WN, Rentsch D, Kwart M, Frommer WB (1998) Developmental control of H+/amino acid permease gene expression during seed development of Arabidopsis. Plant J 14: 535-544.
14. Lee YH, Foster J, Chen J, Voll LM, Weber AP, et al. (2007) AAP1 transports uncharged amino acids into roots of Arabidopsis. Plant J 50: 305-319.
15. Sanders A, Collier R, Trethewy A, Gould G, Sieker R, et al. (2009) AAP1 regulates import of amino acids into developing Arabidopsis embryos. Plant J 59: 540-552.
16. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.
17. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876-4882.
18. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. MolBiolEvol 24: 1596-1599.
19. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34: W369-373.
20. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195-201.
21. Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, et al. (2011) Genome-wide atlas of transcription during maize development. Plant J 66: 553-563.
22. Engel W, Hof JO, Wolf U (1970) [Gene duplication by polyploid evolution: the isoenzyme of the sorbitol dehydrogenase in herring- and salmon-like fishes (Isospondyli)]. Humangenetik 9: 157-163.
23. Chi WC, Chen YA, Hsiung YC, Fu SF, Chou CH, et al. (2013) Autotoxicity mechanism of Oryza sativa: transcriptome response in rice roots exposed to ferulic acid. BMC Genomics 14: 351.
24. Kong H, Landherr LL, Frohlich MW, Leebens-Mack J, Ma H, et al. (2007) Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. Plant J 50: 873-885.
25. Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3: 827-837.
26. Wipf D, Ludewig U, Tegeder M, Rentsch D, Koch W, et al. (2002) Conservation of amino acid transporters in fungi, plants and animals. Trends BiochemSci 27: 139-147.
27. Okumoto S, Schmidt R, Tegeder M, Fischer WN, Rentsch D, et al. (2002) High affinity amino acid transporters specifically expressed in xylem parenchyma and developing seeds of Arabidopsis. J BiolChem 277: 45338-45346.
28. Couturier J, de Faÿ E, Fitz M, Wipf D, Blaudez D, et al. (2010) PtAAP11, a high affinity amino acid transporter specifically expressed in differentiating xylem cells of poplar. J Exp Bot 61: 1671-1682.
29. Marchant A, Kargul J, May ST, Muller P, Delbarre A, et al. (1999) AUX1 regulates root gravitropism in Arabidopsis by facilitating auxin uptake within root apical tissues. EMBO J 18: 2066-2073.
30. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290: 1151-1155.
31. He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169: 1157-1164.
32. Irish VF (2003) The evolution of floral homeotic gene function. Bioessays 25: 637-646.
33. Becker A, Theissen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. MolPhylogenetEvol 29: 464-489.
34. Liu Y, Jiang HY, Chen WJ, Qian YX, Ma Q, et al. (2011) Genome-wide analysis of the auxin response factor (ARF) gene family in maize (*Zea mays*). Plant Growth Regul 63: 225-234.
35. Yang Y, Hammes UZ, Taylor CG, Schachtman DP, Nielsen E (2006) High-affinity auxin transport by the AUX1 influx carrier protein. CurrBiol 16: 1123-1127.
36. Ugartechea-Chirino Y, Swarup R, Swarup K, Péret B, Whitworth M, et al. (2010) The AUX1 LAX family of auxin influx carriers is required for the establishment of embryonic root cell organization in *Arabidopsis thaliana*. Ann Bot 105: 277-289.