

Functional Divergence and Evolutionary Dynamics of the Putative *AAAP* Gene Family in *Brassica rapa*

Li-Zong Hu · Wei-Bo Yin · Yu-Hong Chen · Zan-Min Hu

Published online: 15 October 2013
© Springer Science+Business Media New York 2013

Abstract The amino acid/auxin permease (AAAP) protein family is ubiquitously present in almost all eukaryotic species and functions in various aspects of growth and development. To investigate the evolution of AAAP proteins, here 83 *AAAP* genes in *Brassica rapa* were identified, and their sequence features, and evolutionary relationships were analyzed using in silico methods. According to the phylogenetic analysis, the *AAAP* genes of *B. rapa* are divided into six clades, and these clades share relatively similar sequence features, including gene structures, conserved motifs, and domain organizations. Synteny mapping strongly suggested that segmental duplications could be responsible for the expansion of this family. Adaptive evolution analysis demonstrated that most of AAAP proteins were subject to purifying selection. However, the site Tyr²⁵⁷ on eight AAAP proteins from clade 2b underwent significant positive selection. Functional divergent analysis showed that type I functional divergence coefficients (θ_I) were significantly greater than zero in six pair-wise comparisons. However, functional divergence sites ($Q_k > 0.95$) found only in the AAAP I/II and AAAP I/III comparisons were localized mainly to the trans-membrane (TM) regions, suggesting highly divergent TM structures between these groups might be associated with group-specific functions. Our results could provide a valuable clue for further investigations of the evolutionary history and biological functions of the *AAAP* genes in *B. rapa*.

Keywords *Brassica rapa* · AAAP proteins · Selection · Functional divergence

Introduction

Amino acids and auxins function as nitrogen sources, neurotransmitters, and hormones and are essential for the growth and development of multicellular organisms (Bennett et al. 1996; Clark and Amara 2005; Tegeder and Ward 2012). The amino acid/auxin permeases (AAAP) are enzymes that mediate the transport of a broad range of amino acids or auxins into and out of cells (Saier et al. 2009). Therefore, studying the *AAAP* gene family helps us to understand the molecular mechanisms for transferring and distributing amino acids and auxins in organisms. Previous studies have shown that there is a low sequence similarity among the AAAP transporters, but they all recognize certain amino acids or closely related compounds and have 10~12 predicted trans-membrane (TM) α -helical spanners (Young et al. 1999; Chang and Bush 1997). Additionally, AAAP proteins with highly divergent sequences have various substrate specificities, catalyzing the symport of auxin (indole-3-acetic acid), single amino acids, or multiple amino acids across lipid membranes (Young et al. 1999; Saier et al. 2009; Tegeder et al. 2007).

Since the first mammalian AAAP transporter, named the vesicular GABA transporter (VGAT), was identified, three different subfamilies of AAAP transporters have been identified in mammals, including the VGAT subfamily (McIntire et al. 1997; Sagné et al. 1997), the system A/N subfamily (Chaudhry et al. 1999; Sugawara et al. 2000), and the proton/amino acid transporter (PAT) subfamily (Sagné et al. 2001; Boll et al. 2002). In plants, some *Arabidopsis* amino acid transporters were identified by complementation of yeast mutants defective in the uptake of amino acids (Chen and Bush 1997; Fischer et al. 1995). Amino acid transporters from other

Electronic supplementary material The online version of this article (doi:10.1007/s11105-013-0671-3) contains supplementary material, which is available to authorized users.

L.-Z. Hu · W.-B. Yin · Y.-H. Chen · Z.-M. Hu (✉)
Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Datun Road, Chaoyang District, Beijing 100101, China
e-mail: zmhu@genetics.ac.cn

L.-Z. Hu
University of Chinese Academy of Sciences, Beijing 100049, China

yeast and plant species (Mai and Lipp 1994; Tegeder and Rentsch 2010; Lu et al. 2012) were also isolated and functionally characterized. In the plant, members of the *AAAP* gene family were divided into different subfamilies, including the amino acid permease (AAP) (Frommer et al. 1993; Fischer et al. 1995, 2002; Okumoto et al. 2002; Lee et al. 2007), the lysine–histidine transporter (LHT) (Chen and Bush 1997; Himer et al. 2006), auxin transporter-like protein (AUX1) (Bennett et al. 1996; Young et al. 1999), and the proline transporter (ProT) (Breitkreuz et al. 1999; Grallath et al. 2005). A series of studies strongly suggested that the *AAAP* gene family participates in the regulation of vital biological processes exclusive to plants (Frommer et al. 1993, 1995; Hsu et al. 1993; Rentsch et al. 1996; Grallath et al. 2005; Lee et al. 2007; Tegeder and Ward 2012). Therefore, a phylogenetic analysis of this gene family in representative species of different taxonomic groups is pertinent. Recent analysis of the *Arabidopsis* genome revealed that at least 43 putative amino acid carriers were present at the plasma membrane and tonoplast (Ward 2001; Wipf et al. 2002).

Brassica rapa ($2n=2x=20$, AA genome) is a diploid species that belongs to the *Cruciferae* family and is one of the major oilseed crops cultivated for raw materials for vegetable oil and animal feed protein in many areas. *B. rapa* has a variety of the advantages over oilseed rape (*Brassica napus* L., $2n=4x=38$), including high yield potential, early maturity, and stable male sterile lines (Niemelä et al. 2006). In addition, *B. rapa* is a good genetic model that allows us to investigate duplicated gene fates, gene origin and expansion, gene dosage effects, and gene rearrangement after paleopolyploidizations (Cheng et al. 2011, 2012; Tang et al. 2012). Ideally, one locus in the *Arabidopsis* genome is orthologous to three counterparts in the triplicated *B. rapa* genome due to a whole genome triplication event approximately 11–12 MYA (Lysak et al. 2007; Mun et al. 2009). Recently, the draft genome of *B. rapa* offered an opportunity to investigate the *AAAP* gene family in this species (Wang et al. 2011). Here, we first computationally identified the putative *AAAP* genes in *B. rapa*. These genes were further classified according to a comparative phylogenetic relationship with *Arabidopsis* orthologs. The chromosomal locations and duplication patterns of these genes were also studied. Finally, we examined selective signatures and functional divergence of these genes. The results provide new insights into the evolutionary history of this family, and useful clues for further functional analysis of each group of AAAP proteins in *B. rapa*.

Materials and Methods

Retrieval and Identification of Putative *AAAP* Genes

First, the known *AAAP* genes in *Arabidopsis thaliana* were downloaded from the TAIR database (<http://www.arabidopsis.org/browse/genefamily/AAAP.jsp>).

We used these genes as queries, and retrieved the BRAD database (*B. rapa* ssp. *pekinensis* cv. Chiifu genome V1.0, <http://Brassicadb.org>) via the Basic Local Alignment Search Tool (BLAST) tool to identify *AAAP* candidate genes in *B. rapa*. Second, the HMM profile of the Aa_trans domain (PF01490) from the Pfam database (<http://pfam.sanger.ac.uk/search>) was reconstructed using HMMER 2.3.2 software (Eddy 1998), which was also used to search the BRAD database for *AAAP* candidate genes. Third, we took the *AAAP* candidate genes identified by the above methods as queries to further search the BRAD and National Center for Biotechnology Information databases. A collection of putative *AAAP* genes was obtained, and the redundant sequences were removed. Finally, the Pfam tool (Punta et al. 2012) was used to screen the genes for Aa_trans domains (PF01490) to confirm the identity of the genes as *AAAP* family members. If there was more than one allele at the same locus, only one allele was chosen as representative.

Chromosomal Mapping and Gene Expansion Patterns

The predicted *AAAP* genes were selected as queries, and BLASTN was used to search the BRAD database to identify their corresponding genomic position. Subsequently, the *AAAP* genes were manually mapped on chromosomes. Gene amplification patterns include two major molecular mechanisms, including segmental and tandem duplication. To elucidate the expansion patterns of this gene family, duplicate genes, generated by segmental duplication, were retrieved from the PGDD database for *B. rapa* and *Arabidopsis*. Furthermore, if two or more members of the *AAAP* gene family were located on the same region with a sequence length of less than 50 kb, they were considered tandem duplication genes (Riechmann et al. 2000).

Phylogenetic Analysis and Sequence Features

We aligned the full-length sequences of the AAAP proteins from *Arabidopsis* and *B. rapa* using Clustal X with default parameters (Thompson et al. 1997). The phylogenetic tree of *AAAP* genes from *B. rapa* and *Arabidopsis* was constructed using the neighbor-joining method (Saitou and Nei 1987) with *p* distance and pairwise deletion options. The distinct layouts of the phylogenetic trees were viewed using MEGA (Tamura et al. 2007) and iTOL (Letunic and Bork 2011). Gene structures were illustrated using the Gene Structure Display Server (GSDS; Guo et al. 2007) by a comparison of the corresponding coding and genomic sequences of the *B. rapa* *AAAP* genes. MEME version 2.2 (Bailey et al. 2006) was used to identify the conserved motif patterns with a set of parameters as follows: number of repetitions—any, maximum number of motifs—20, and optimum motif width set to 6–200.

Tests for Selective Pressure and Functional Divergence

To improve the accuracy and power of selective pressure analysis, raw sequences were filtered according to the following criteria: (1) 83 BrAAAP protein sequences were clustered using BLASTCLUST (<http://toolkit.tuebingen.mpg.de/blastclust/>) with 90 % coverage and 30 % identity. Only clusters containing four or more sequences were included in further analysis. (2) Multiple codon alignments of 42 filtered *BrAAAP* genes were generated using PAL2NAL (Suyama et al. 2006). (3) Gaps and highly divergent regions were automatically removed.

To detect the selective pressures acting on each amino acid site of AAAP proteins, two codon-based likelihood methods, including site and branch site models (Yang and Nielsen 2002; Zhang et al. 2005; Yang 2007), were carried out using the CODEML package of PAML version 4.0. For site models, M0 (one ratio), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), and M8 (beta and ω) were used to calculate the related parameters and likelihood scores, and then likelihood ratio test (LRT) was performed by a comparison of M0 with M3, M1a with M2a, and M7 with M8. Finally, Bayes empirical Bayes (BEB) was used to infer posterior probabilities for all of amino acid sites under the M3, M2a, and M8 models. For the branch site model, foreground clades were labeled, and positive selection sites at these clades were identified by test 2. The LRT was performed in test 2 by the comparison between the null model MA1 (model=2, NSSites=2, $\omega=1$) and the alternative model MA (model=2, NSSites=2, estimated ω ; Yang 2007). The result of LRT was applied to measure the statistical significance and the BEB method was used to infer posterior probabilities.

Functional divergence between different groups of the AAAP protein family in *B. rapa* was identified by DIVERGE2 (Gu 1999). The coefficient of type I functional divergence (θ_I) was estimated (Gu 1999; Gu and Vandervelden 2002). The fact that the values of θ_I were significantly greater than 0 revealed that a site-specific evolutionary rate shift existed between groups (Gu 2003).

Prediction of Trans-Membrane Domains and Protein Structures

TMPred (http://www.ch.embnet.org/software/TMPRED_form.html) was used to predict trans-membrane domains of BrAAAP1 proteins that were illustrated by TOPO2 (<http://www.sacs.ucsf.edu/TOPO2/>). The homology models for BrAAAP2 and BrAAAP11 proteins were automatically generated by the SWISS-MODEL web server (Arnold et al. 2006), and their three-dimensional structures were visualized by PDBpaint (Fournier and Andrade-Navarro 2011). The critical amino acid sites with functional divergence between

groups and under positive selection were mapped on the representative TM and protein structures.

Results

Identification and Phylogenetic Relationships of *AAAP* Genes in *Arabidopsis* and *B. rapa*

The highly conserved *Aa_trans* domain allows us to easily identify all members of the AAAP family of proteins. Using the combined methods of BLAST, HMMER, and Synteny mapping, a total of 83 *AAAP* genes were identified from the sequenced *B. rapa* genome. All AAAP proteins in *B. rapa* except for six (BrAAAP8/43/47/61/69/78) contained at least one *Aa_trans* domain. The proteins without an *Aa_trans* domain were also considered members of the AAAP family based mainly on a high degree of collinearity between them and well-known *AAAP* genes. Detailed information of all genes is listed in Electronic Supplementary Material (ESM) Table S1.

To investigate the phylogenetic relationships of *AAAP* genes between *B. rapa* and *Arabidopsis*, a phylogenetic tree of the *AAAP* genes was constructed using the neighbor-joining method with the *p* distance and pairwise deletion options (Fig. 1). Based on the phylogenetic analysis, the typical relationships of *AAAP* genes between *Arabidopsis* and *B. rapa* can be classified into three evolutionary relationships, 1:3, 1:2, and 1:1 orthologous relationships. Of the 83 *AAAP* genes identified in *B. rapa*, 30 (36.1 %), 28 (33.7 %), and 11 (13.3 %) genes are 1:3, 1:2, and 1:1 orthologs, respectively. Additionally, the orthologous counterparts of *BrAAAP22/81*, *BrAAAP5/59*, *BrAAAP30/72*, and *BrAAAP65* were absent in *Arabidopsis* genome, suggesting that their orthologous relationships to *Arabidopsis* counterparts have been broken by rapid evolution. *BrAAAP26* is orthologous to two *Arabidopsis* loci (AT3G09330 and AT3G09340) that are tandem duplicated within a 5.7 kb region of chromosome 3.

Gene Structures and Conserved Motifs of *AAAP* Genes in *B. rapa*

To clarify the sequence–phylogeny relationships between the *AAAP* genes in *B. rapa*, a phylogenetic tree was constructed using the neighbor-joining method (Saitou and Nei 1987) implemented in MEGA software (Tamura et al. 2007). Additionally, the gene structures and conserved motifs were identified using the GSDS (Guo et al. 2007) and MEME (Bailey et al. 2006) tools and were illustrated on the right of the phylogenetic tree (Fig. 2).

Based on our phylogenetic analysis, the *AAAP* genes in *B. rapa*, could be divided into six major clades: A, B, C, D, E, and F (Fig. 2). Analyses of gene structures showed that the

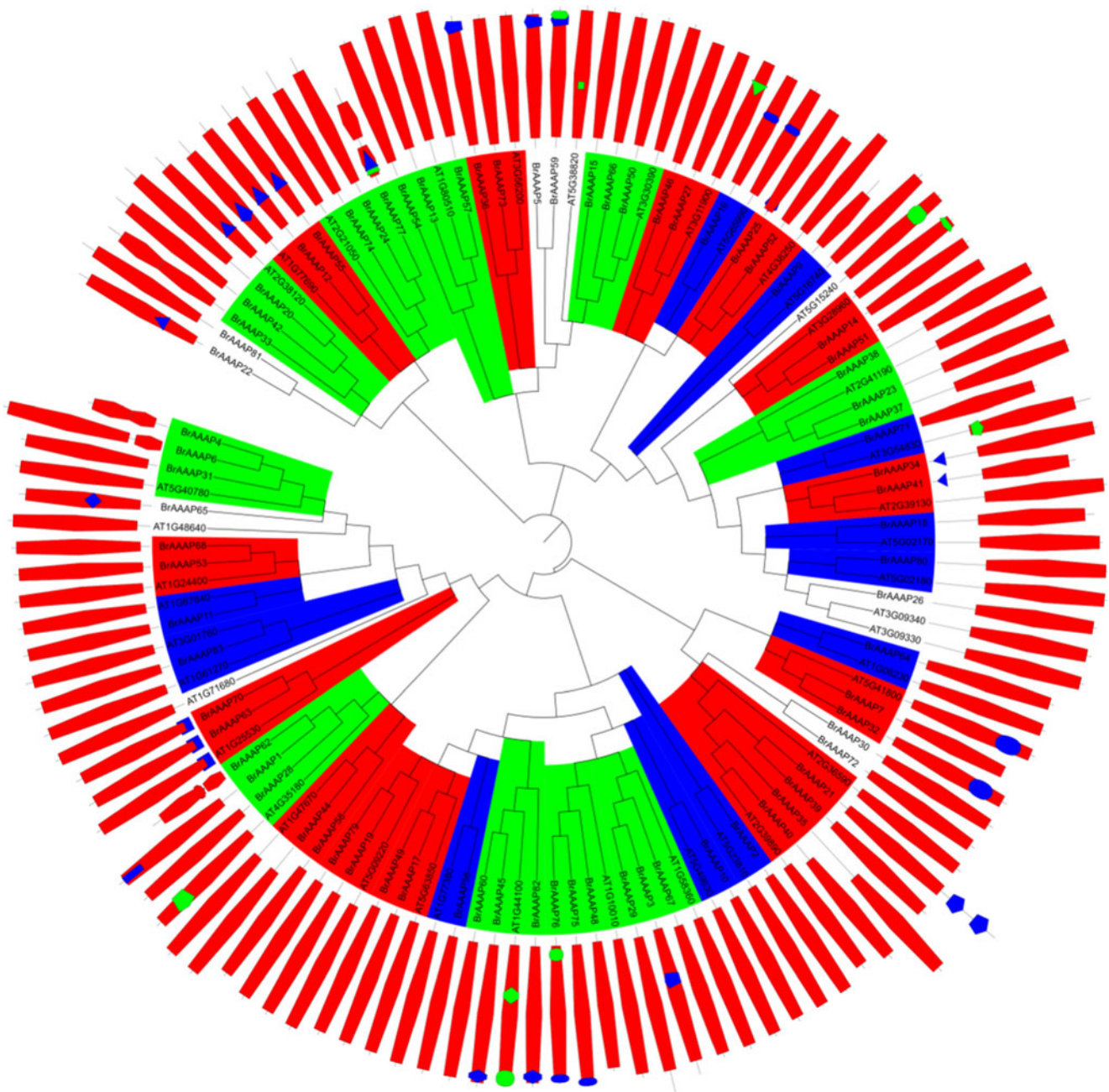


Fig. 1 Phylogenetic relationships of AAAP family proteins in *Arabidopsis* and *B. rapa*. The circular tree based on the full-length sequences of proteins was viewed using iTOL. Blue, red, and green clades indicate 1:3, 1:2, and 1:1 orthologous relationships between

Arabidopsis and AAAP proteins in *B. rapa*. Each of the shape–color combinations represent one independent domain. The red vertical hexagon is representative of the *Aa_trans* (PF01490) domain. The remnant shape–color combinations are representative of 27 distinct domains

coding sequences of all *BrAAAP* genes were disrupted by introns except for *BrAAAP13/54/57* in clade C and *BrAAAP25/52* in clade D. According to the number of introns, 83 *BrAAAP* genes were divided into three groups with a range of 1–3, 4–7, and 8–11 introns. We found that 52 (62.7 %) genes had between four and seven introns, indicating that the most common structures of AAAP genes in *B. rapa* could be found in each of the clades in the phylogenetic tree. Of

BrAAAP genes with between one and three introns, *BrAAAP16* in clade D and *BrAAAP61* in clade E had one intron; *BrAAAP62* in clade A, *BrAAAP9/14/27/46/51* in clade D and *BrAAAP8/47/69/78* in clade E had two introns; *BrAAAP5/73* had a three-intron structure that was specific to clade C. For *BrAAAP* genes with between 8 and 11 introns, *BrAAAP6* in clade A and *BrAAAP35* in clade F had 8 introns; *BrAAAP18/26/37/71* in clade D and *BrAAAP20* in clade E

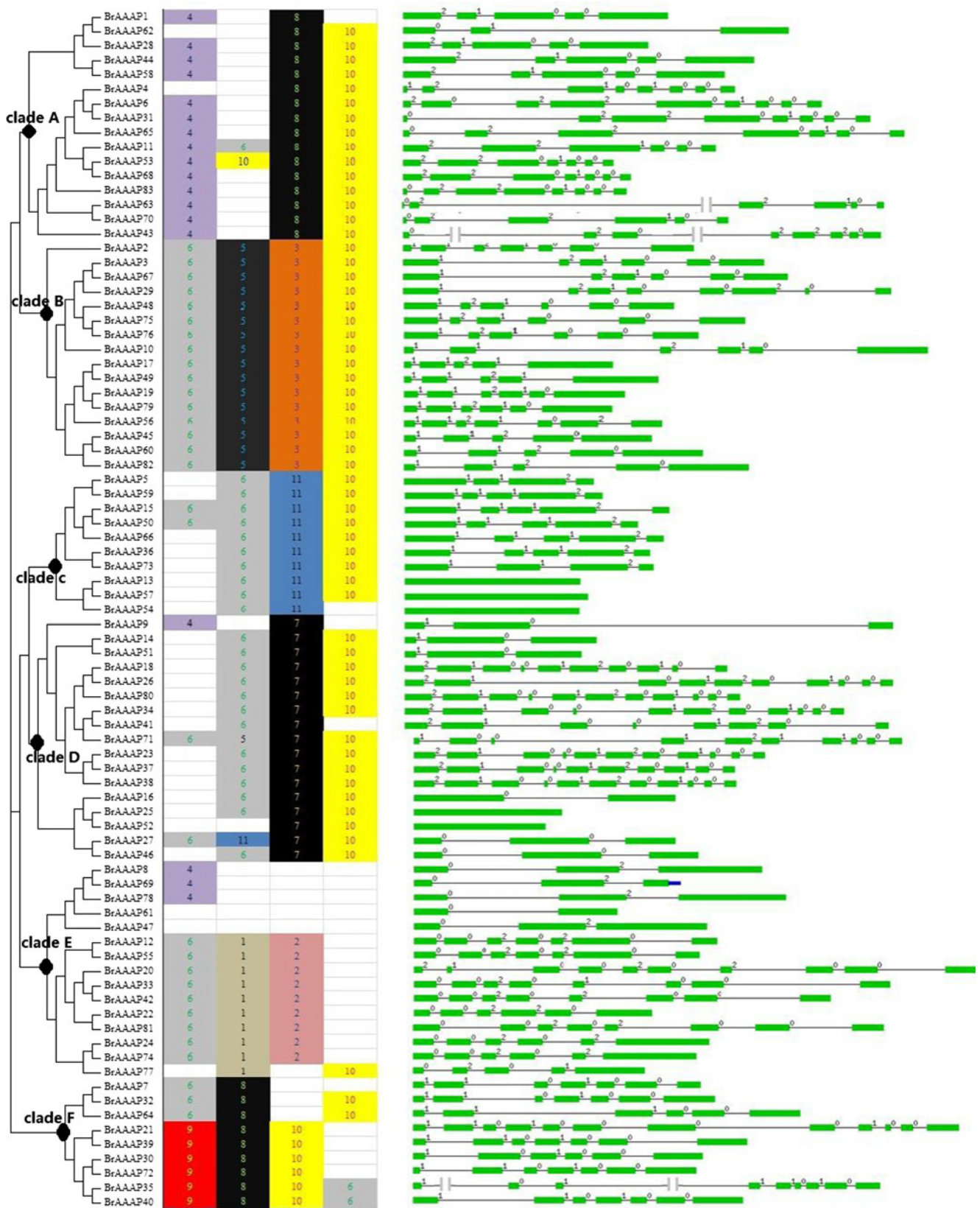


Fig. 2 Phylogenetic relationships and sequence features of AAAP genes in *B. rapa*. The phylogenetic tree was constructed based on the full-length sequences of 83 AAAP proteins. *Black diamond* highlight different clades on the left. The *rectangles* with different numbers and colors

demonstrate distinct conserved motifs in the middle. On the right, exons are represented by *green boxes*; introns are represented by *black lines*; UTR regions are represented by *blue thick lines*; 0, 1, and 2 indicate distinct intron phases

had 9 introns; *BrAAAP23/34/38/80* had a 10-intron structure that was specific to clade D; and *BrAAAP21* had an 11-intron structure that was specific to clade F.

The number, order, and position of motifs within BrAAAP proteins were graphically presented (Fig. 2). Motif 4/8/10, motif 6/5/3/10, motif 6/11/10, motif 6/7/10, motif 6/1/2, and motif 9/8/10 constituted the most common patterns of motif architectures in clade A, B, C, D, E, and F, respectively. In clade A, motif 4 was absent in BrAAAP4/62; motif 10 was also absent in BrAAAP1. However, motif 6 was inserted into BrAAAP11, and another motif 10 was observed in BrAAAP53. All proteins in clade B had the same motif organization pattern. In clade C, motif 6 was added in BrAAAP15/50 and motif 10 was deleted in BrAAAP54. A gain of motif 5 in BrAAAP71 and motif 11 in BrAAAP27 was found, but a loss of motif 10 in BrAAAP41 and motif 6 in BrAAAP52 was also observed in clade D. BrAAAP9 had a distinct motif pattern with motif 4 and 7. Interestingly, BrAAAP47/61 in clade E did not contain any motifs, and BrAAAP8/69/78 contained only motif 4, but BrAAAP77 had a distinct motif pattern with motif 1 and 10. In clade F, BrAAAP7 contained motif 6/8, BrAAAP32/64 contained motif 6/8/10, BrAAAP21/30/39/72 contained motif 9/8/10, and BrAAAP35/40 contained motif 9/8/10/6.

Physical Mapping and Duplication Patterns of *AAAP* Genes in *B. rapa*

Physical mapping of *AAAP* genes showed that ~81 *BrAAAP1* genes were unevenly distributed on 10 chromosomes in *B. rapa*, while the chromosomal positions of the *BrAAAP82* and *BrAAAP83* genes were still unknown because they were mapped to scaffolds (Scaffold000123 and Scaffold000217) that were not associated with a specific *B. rapa* chromosome. The number of *AAAP* genes per chromosome ranged from 4 to 13. Gene mapping revealed that 4, 13, 10, 8, 10, 5, 5, 7, 12, and 5 *AAAP* genes mapped to chromosome A01, A02, A03, A04, A05, A06, A07, A08, A09, and A10, respectively.

Based on the phylogenetic results (Fig. 1), 27 pairs of duplicated co-paralogs were identified in *BrAAAP* genes. Twenty protein-coding genes flanking each pair of duplicated co-paralogs were retrieved from the *B. rapa* sequencing database using the PGDD tool. Novel co-paralogs were found in protein-coding genes near each pair of 23 duplicated co-paralogs (Fig. 3), suggesting that these 23 duplicated co-paralogs could be segmental duplicates. In addition, four pairs of *B. rapa AAAP* genes (*BrAAAP21/22*, *BrAAAP24/25*, *BrAAAP39/40* and *BrAAAP75/76*) occurred within a distance of 50 kb, and thus they were considered to be the results of tandem duplications according to the criterion reported by Riechmann et al. (2000). Other duplicated co-paralogs might be explained by different duplication events, such as random duplication and/or insertion events.

Testing for Selection in AAAP Proteins

Site-specific models assumed variable selective pressures among sites. M0 (one ratio, ω), M1a (nearly neutral, $\omega_0 < 1$ and $\omega_1 = 1$), and M7 (beta, p, q) were the null models that did not allow for any codons with $\omega > 1$. However, when compared with three null models above, M2a (positive selection; $\omega_0 < 1$, $\omega_1 = 1$, and $\omega_2 > 1$), M3 (discrete; $\omega_0, \omega_1, \omega_2$), and M8 (beta and ω ; $p, q, \omega_s > 1$) were the alternative models that were added to an extra class of sites with $\omega > 1$. An LRT was performed by the comparison of M0 and M3, M1a and M2a, and M7 and M8. The integrative analysis of LRT and BEB were used for an investigation of positively selected sites. The results of the selective pressure analysis of AAAP proteins are listed in Table 1. For the M1a/M2a and M7/M8 comparison, LRT analysis revealed no sites under positive selection. According to the results of the LRT for M0 and M3 ($2\Delta l = 771.155$, $df = 4$, $P < 0.01$), we found amino acid sites under heterogeneous selection pressures, but the LRT of the M0/M3 comparison was not adopted in most cases. This indicated that *AAAP* genes were under strong purifying selection across amino acid sites. Additionally, no site under positive selection was identified via Bayes empirical Bayes inference using a probability criterion of 95 %.

Some of amino acid sites on the AAAP proteins from the specific clades in the phylogenetic tree (ESM Fig. S1) might be driven by positive selection. Therefore, we further detected this type of amino acid site for AAAP proteins using the branch site model. Model A (MA, model=2, NSSites=2, estimated ω) was the alternative model, and its corresponding null model (MA1, model=2, NSSites=2, $\omega = 1$) fixed $\omega_2 = 1$ for the foreground branch (Yang and Nielsen 2002; Zhang et al. 2005). The LRT of MA and MA1 was used to identify some positively selected sites of specific clades. The LRT statistic of comparison between MA and MA1 in clade 1c ($2\Delta l = 4.26$, $df = 1$, $P < 0.05$), clade 3a ($2\Delta l = 7.84$, $df = 1$, $P < 0.01$), and clade 2b ($2\Delta l = 10.24$, $df = 1$, $P < 0.01$) suggested that MA was significantly better than MA1 for these specific clades. Furthermore, regardless of the probability, the BEB analysis showed that there were three positively selected sites in clade 1c (46Y, 201L), one positively selected site in clade 2b (133 W) and one positively selected site in clade 3a (39 V). However, only one positively selected site (133 W) was identified from eight AAAP proteins in clade 2b at significant level (Table 2).

Analysis of Functional Divergence

To identify key amino acids involved in functional divergences between protein subfamilies, AAAP proteins in *B. rapa* were divided into five groups (ESM Fig. S2). Pairwise comparisons between AAAP groups, including AAAP I, AAAP II, AAAP III, AAAP IV, and AAAP V, were carried

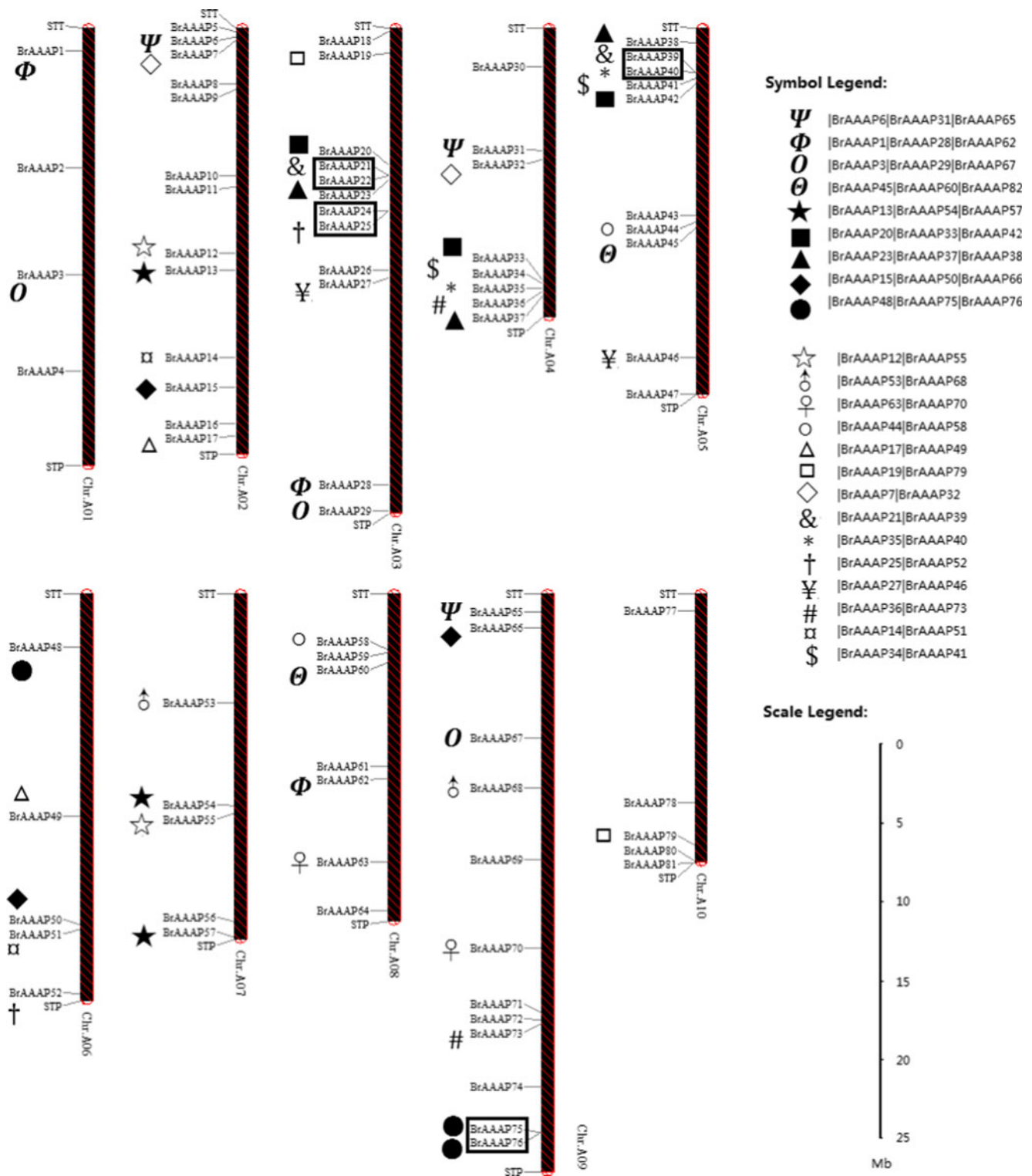


Fig. 3 Chromosomal localization and expansion pattern of *B. rapa* AAAP genes. Eighty-one AAAP genes were mapped to 10 chromosomes in *B. rapa*. Four tandem-duplicated gene pairs are indicated by

rectangles. Twenty-three segmental duplicated gene pairs are marked by a broad range of symbols. The start and stop positions of the chromosomes are abbreviated as STT and STP, respectively

out using the DIVERGE program (Gu and Vander-Velden 2002). The coefficients of type I functional divergence with standard errors, significance levels, and critical amino acids

are shown in Table 3. Type I coefficients of functional divergence in comparisons between different groups, apart from the AAAP II/IV, AAAP II/V, AAAP III/IV, and AAAP IV/V

Table 1 Tests for selection among codons of *B. rapa* AAAP proteins using site models included in Codeml program implemented in PAML 4

Model	NFP ^a	Parameter estimate	lnL ^b	2Δl ^c	P ^d	PSS ^e
M0 (one-ratio)	83	$\omega_0=0.10368$	-16,179.645			Not allowed
M1a (neutral)	84	$P_0=0.984$ $\omega_0=0.102$	$p_1=0.0162$ $\omega_1=1.000$	-16,171.396		Not allowed
M2a (positive)	86	$p_0=0.984$ $p_2=0.0130$ $\omega_1=1.000$	$p_1=0.00312$ $\omega_0=0.102$ $\omega_2=1.000$	-16,171.396	(M1a vs. M2a) 0	>0.05 None
M3 (discrete)	87	$p_0=0.166$ $p_2=0.267$ $\omega_1=0.0814$	$p_1=0.568$ $\omega_0=0.0215$ $\omega_2=0.220$	-16,009.173	(M0 vs. M3) 771.155**	<0.01 None
M7 (beta)	84	$p=1.784$	$q=14.373$	-16,013.371		Not allowed
M8 (beta & ω)	86	$p=1.784$ $p_0=0.99999$ $\omega=2.602$	$q=14.373$ $(p_1=0.00001)$	-16,013.373	(M7 vs. M8) -0.002	>0.05 None

^a The number of free parameters^b Likelihood of the model^c 2(11–10)^d Probability value^e Positive selective sites

comparisons were greater than zero at a significant level ($P < 0.05$). Furthermore, the site-specific profile generated by posterior analysis was used to predict the critical amino acid residues with $Q_k > 0.85$ as a cutoff. Although θ_I of functional divergence in the AAAP I/IV, AAAP I/V, and AAAP II/III

comparisons were significantly greater than zero, no amino acid sites with $Q_k > 0.85$ were identified. However, in the AAAP I/II, AAAP I/III, and AAAP III/IV comparisons, we detected 37, 200 and 1 amino acid positions using $Q_k > 0.85$ as a cutoff, respectively (Fig. 4). Take BrAAAP1 as an example.

Table 2 Parameters estimation and likelihood ratio tests for the branch-site models

Model	df ^a	Parameter estimate	lnL ^b	2Δl ^c	P ^d	PSS ^e
MA1 Clade 1c	3	$\omega_0=0.10, P_0=0.45, \omega_1=1.00, P_1=0.007,$ $\omega_{2a \text{ fore}}=0.40, \omega_{2a \text{ back}}=0.10, P_{2a}=0.54,$ $\omega_{2b \text{ fore}}=0.40, \omega_{2b \text{ back}}=1.00, P_{2b}=0.009$	-16,167.37			Not allowed
MA Clade 1c	4	$\omega_0=0.10, P_0=0.65, \omega_1=1.00, P_1=0.01,$ $\omega_{2a \text{ fore}}=1.78, \omega_{2a \text{ back}}=0.10, P_{2a}=0.33,$ $\omega_{2b \text{ fore}}=1.78, \omega_{2b \text{ back}}=1.00, P_{2b}=0.005$	-16,165.24	4.26*	<0.05	46 Y (0.941) 201 L (0.904)
MA1 Clade 2b	3	$\omega_0=0.10, P_0=0.93, \omega_1=1.00, P_1=0.01,$ $\omega_{2a \text{ fore}}=0.40, \omega_{2a \text{ back}}=0.10, P_{2a}=0.06,$ $\omega_{2b \text{ fore}}=0.40, \omega_{2b \text{ back}}=1.00, P_{2b}=0.0009$	-16,166.27			Not allowed
MA Clade 2b	4	$\omega_0=0.10, P_0=0.96, \omega_1=1.00, P_1=0.01,$ $\omega_{2a \text{ fore}}=1.47, \omega_{2a \text{ back}}=0.10, P_{2a}=0.03,$ $\omega_{2b \text{ fore}}=1.47, \omega_{2b \text{ back}}=1.00, P_{2b}=0.0004$	-16,161.15	10.24**	<0.01	133 W (0.998**)
MA1 Clade 3a	3	$\omega_0=0.10, P_0=0.72, \omega_1=1.00, P_1=0.01,$ $\omega_{2a \text{ fore}}=0.40, \omega_{2a \text{ back}}=0.10, P_{2a}=0.26,$ $\omega_{2b \text{ fore}}=0.40, \omega_{2b \text{ back}}=1.00, P_{2b}=0.004$	-16,167.07			Not allowed
MA Clade 3a	4	$\omega_0=0.10, P_0=0.82, \omega_1=1.00, P_1=0.01,$ $\omega_{2a \text{ fore}}=2.44, \omega_{2a \text{ back}}=0.10, P_{2a}=0.17,$ $\omega_{2b \text{ fore}}=2.44, \omega_{2b \text{ back}}=1.00, P_{2b}=0.003$	-16,163.15	7.84**	<0.01	39V (0.737)

^a The number of free parameters^b Likelihood of the model^c 2(11–10)^d Probability value^e Positive selective sites (take BrAAAP1 as an example)* $P < 0.05$, significant level; ** $P < 0.01$, significant level

Table 3 Estimation of the coefficient of type I functional divergence (θ_I) from pairwise comparisons between AAAP groups in *B. rapa*

Comparison	$\theta_I \pm SE^a$	P^b	Critical amino acid sites (Q_k) ^c
AAAP I vs. AAAP II	0.797±0.135**	<0.01	76, 212, 264, 345, 357, 417 ($Q_k > 0.95$)
AAAP I vs. AAAP III	0.934±0.132**	<0.01	76, 103, 110, 159, 212, 226, 302, 357, 392, 409 ($Q_k > 0.98$)
AAAP I vs. AAAP IV	0.295±0.192*	<0.05	Not found ($Q_k > 0.85$)
AAAP I vs. AAAP V	0.404±0.262*	<0.05	Not found ($Q_k > 0.85$)
AAAP II vs. AAAP III	0.339±0.133**	<0.01	Not found ($Q_k > 0.85$)
AAAP II vs. AAAP IV	0.075±0.118	>0.05	Not found ($Q_k > 0.85$)
AAAP II vs. AAAP V	0.096±0.219	>0.05	Not found ($Q_k > 0.85$)
AAAP III vs. AAAP IV	0.366±0.158**	<0.01	195 ($Q_k > 0.85$)
AAAP III vs. AAAP V	0.194±0.248	>0.05	Not found ($Q_k > 0.85$)
AAAP IV vs. AAAP V	0.0546±0.147	>0.05	Not found ($Q_k > 0.85$)

^a Standard error

^b Significant level

^c Threshold values of posterior probabilities for all individual sites (take BrAAAP1 as an example)

* $P < 0.05$; ** $P < 0.01$

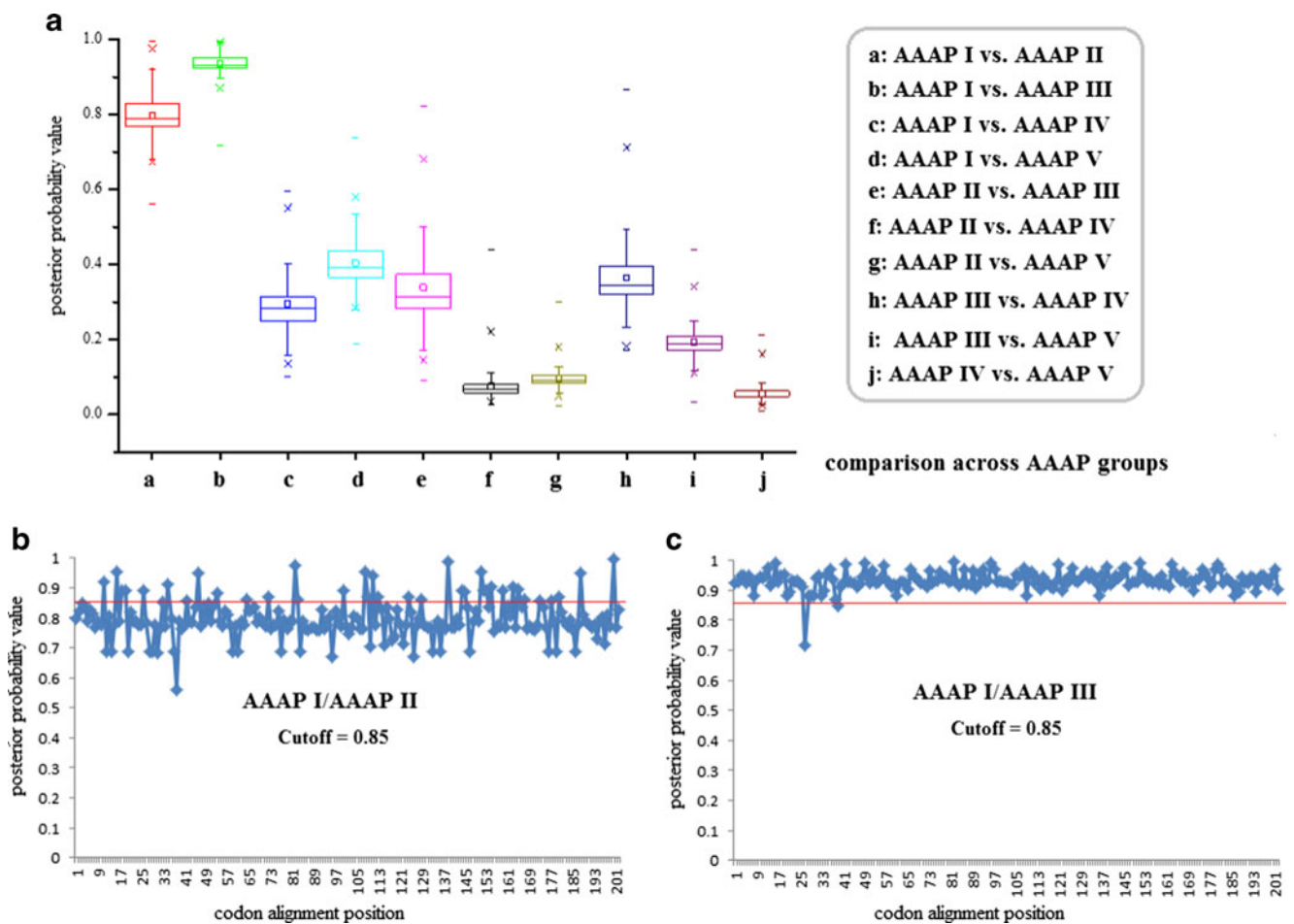


Fig. 4 The estimation of type I functional divergence between subgroups of AAAP proteins in *B. rapa*. The positions with gaps involved in each pairwise comparison were retained. The *line* indicates a posterior probability cutoff of 0.85. **a** The distribution of the posterior probability values

of each AAAP comparison is plotted side by side, showing the minimum, the 1 % percentile, the 25 % percentile, the median, the mean, the 75 % percentile, the 99 % percentile and the maximum values. **b** and **c** Posterior probability values in AAAP I/II and AAAP I/III comparisons, respectively

We identified critical amino acid positions of the AAAP I/II comparison using $Q_k > 0.95$ as a cutoff, which included six amino acid sites: Gly⁷⁶, Leu²¹², Ser²⁶⁴, Phe³⁴⁵, Ser³⁵⁷, and Val⁴¹⁷. Moreover, we also identified 10 amino acid sites, including Gly⁷⁶, Ala¹⁰³, Trp¹¹⁰, Leu¹⁵⁹, Leu²¹², Trp²²⁶, Ala³⁰², Ser³⁵⁷, Ser³⁹², and Tyr⁴⁰⁹, which were critical sites determining functional divergence (Table 3).

Mapping of Critical Amino Acids on the Representative TM and Three-Dimensional Protein Structures

To gain insight into the potential effect of positive selection and type I functional divergence sites on AAAP protein structures and/or functions, we located the sites of selection on the representative TM and three-dimensional structures. The results showed that most of the type I functional divergence sites with posterior probability values $>95\%$ were found in the TM helix. For example, 83.7 % of type I functional divergence sites between AAAP I and AAAP III existed in the TM helix, and this type of site between AAAP I and AAAP II was found in the TM region (Fig. 5a). BrAAAP2 and BrAAAP11 were representative of the rapidly evolutionary clade 1c and 2b of the phylogenetic tree (ESM Fig. S1), and their protein structures were constructed using SWISS-MODEL based the template protein (PDB ID: 3ob6). Four positively selected sites were found via a branch site model and mapped on the corresponding protein structures. The positively selected site Tyr²⁵⁷ in clade 2b was mapped to the coil structure inside the cytoplasm of the BrAAAP11 protein (Fig. 5b), but the positively selected sites Val⁵¹ and Leu¹⁶² in clade 1c were mapped to the α -helix structure of the BrAAAP2 protein (Fig. 5c). In addition, Cys²⁶³ in clade 3a ($P=0.737$) was found in the region inside the cytoplasm of the BrAAAP2 protein (Fig. 5c).

Discussion

B. rapa, one of the major vegetable and oil crops in Asia and Europe, has a simple diploid genome that has been fully sequenced and is a potential model species for the Brassicaceae family. Extensive studies have shown that the *Brassica* and *Arabidopsis* species shared three whole-genome duplication events, named *At- α* (3R), *At- β* (2R), and *At- γ* (1R) (Lagercrantz and Lydiat 1996; Lan et al. 2000; Lukens et al. 2003; Babula et al. 2003; Lysak et al. 2005; Yang et al. 2006; Cheung et al. 2009; Beilstein et al. 2010; Cheng et al. 2011). Moreover, after the split of the *Brassica* and *Arabidopsis* species approximately 10–20 Ma ago (Yang et al. 1999), the *Brassica* species experienced an additional whole-genome triplication named 4R (Lan et al. 2000; Babula et al. 2003; Lukens et al. 2003; Mun et al. 2009; Tang et al. 2012; Cheng et al. 2012). Therefore, it was clear that lineage-specific expansion occurred in the *B. rapa* genome (Johnston

et al. 2005), and theoretically, three paralogous genes in *B. rapa* should be co-orthologous to one *Arabidopsis* counterpart. In our study, we identified 83 members of the AAAP gene family in the *B. rapa* genome. They were unevenly distributed on 10 chromosomes. The nearly twofold increase of AAAP gene members in *B. rapa* might result from the lineage-specific expansion described above. Based on the circular phylogenetic tree, it was clear that there were three types of phylogenetic relationships between *Arabidopsis* and *B. rapa* AAAP orthologous genes, including 1:3, 1:2, and 1:1 orthologous relationships (Fig. 1). These complex relationships might be related to a preference for gene retention and loss after triplication caused by chromosomal rearrangements (Yang et al. 2005, 2006; Cheung et al. 2009; Beilstein et al. 2010; Cheng et al. 2011). It is well-known that three duplication patterns, including polyploidy (whole genome duplication), tandem duplication, and transposition events were responsible for the lineage-specific expansion (Kong et al. 2007). To investigate the evolutionary relationships of the AAAP genes in *B. rapa*, we further analyzed the syntenic relationships of paralog pairs at the terminal nodes of the phylogenetic tree using the PGDD tool. The results revealed that 23 and 4 paralog pairs resulted from segmental and tandem duplications, respectively, suggesting that segmental and tandem duplication might have contributed greatly to the expansion of this family in *B. rapa*.

Correct classification was important for an in-depth understanding of the structural, functional, and evolutionary information about the AAAP genes from *B. rapa*. Unlike the *Arabidopsis* AAAP gene family that was divided into four major clades, the *B. rapa* AAAP genes were divided into six different clades. Comparative analysis showed that AAAP genes in clade A, B, E, and F from the phylogenetic tree (Fig. 2) corresponded to the annotated *LHT*, *AAP*, *AUX1*, and *ProT* genes in *Arabidopsis*, respectively (Rentsch et al. 2007; Tegeder and Rentsch 2010; Tegeder and Ward 2012). However, AAAP genes in clades C and D corresponded to unannotated AAAP genes in *Arabidopsis*, suggesting that little was known about their functions. Homology analysis demonstrated that all members of the AAAP family shared approximately 8 % sequence similarity, and members of the AAAP family in clade A, B, C, D, E, and F sequence shared approximately 31, 48, 40, 22, 26, and 28 % sequence similarities at protein level, respectively. These results indicated that the AAAP gene family might have multiple cellular functions and be involved in several biological processes (Frommer et al. 1993, 1995; Rentsch et al. 1996). According to the orthologous relationship between BrAAAP proteins in clade B and AtAAP1~8, we predicted that the motif 6/5/3/10 was closely related to the function of generally transporting neutral and acidic amino acids with moderate affinity (Okumoto et al. 2002). BrAAAP proteins in clade E corresponding to AUX1

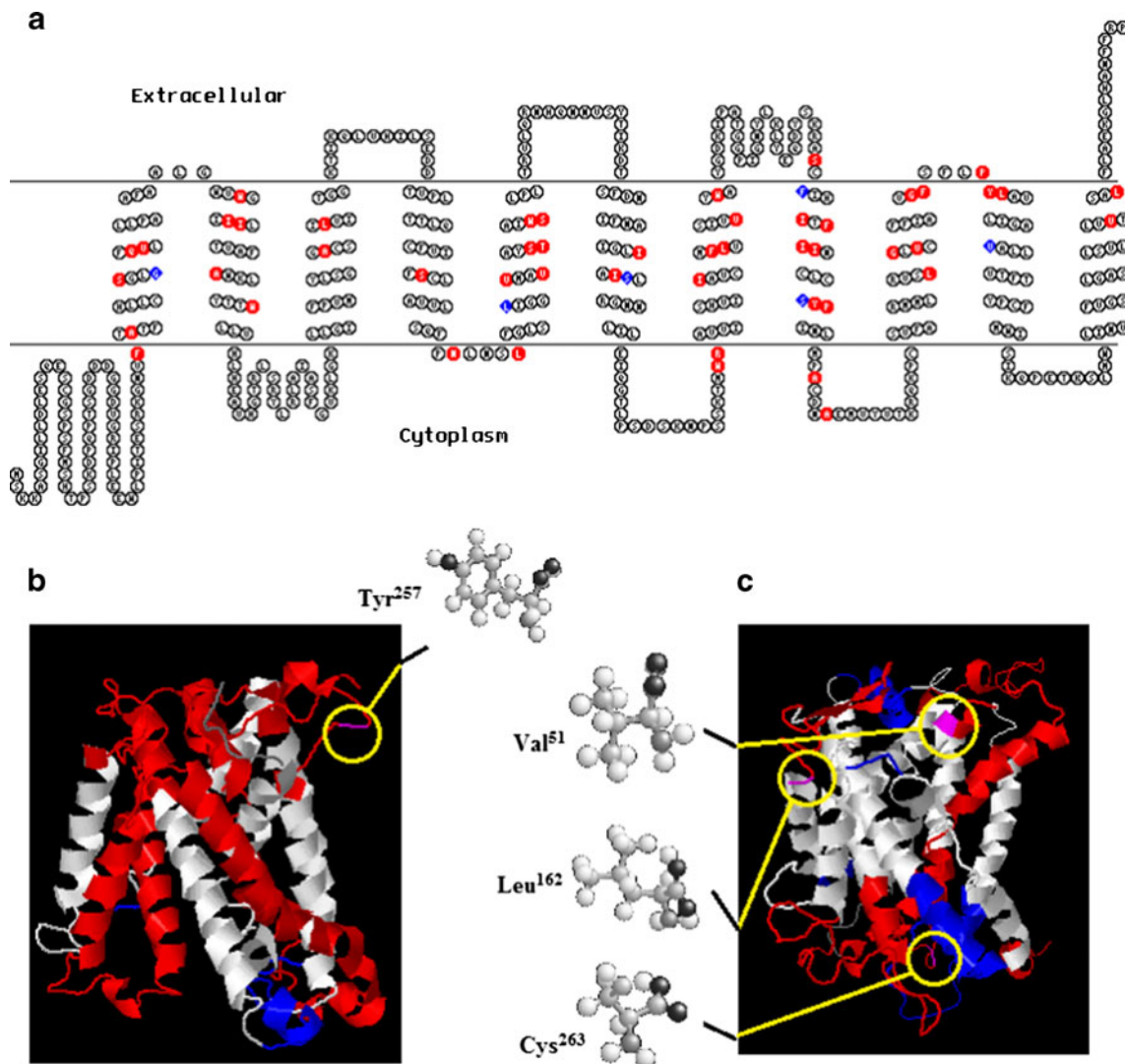


Fig. 5 Localization of critical amino acid sites on the representative TM and 3D structures of AAAP proteins in *B. rapa*. Type I functional divergence sites and positive selection sites mapped on BrAAAP1 TM structures (a) are labeled by red circles and blue diamonds, respectively.

did not contain motif 10 shared by most AAAP proteins in *B. rapa*. This finding implied that the presence of motif 10 could depress the activity of auxin transporters (Rentsch et al. 1996). Six *BrAAAP* genes (*BrAAAP8/43/47/61/69/78*) shared high collinearity with well-known AAAP genes, but their proteins had no Aa_trans domain. This implied that there was domain loss during the evolution of the *BrAAAP* gene family.

To detect selective pressure on the *AAAP* family, the site-specific model in PAML was used to evaluate positive evolution in the phylogenetic tree. Positively selected sites were found using the BEB method. This suggested that AAAP proteins evolved under purifying selection. The branch site model has advantages over the site model because it can detect positively selected sites that occurred in certain positions in specific clades (Yang and Nielsen 2002; Zhang et al. 2005).

For BrAAAP11 (b) and BrAAAP2 (c) proteins, white, red, and blue regions show TM-helix structures, inside residues, and outside residues, respectively, and purple amino acid sites included in the yellow circle were under positive selection

Here, we identified the site (Trp²⁴¹), shared by eight proteins in clade 2b, as being under positive selection. These proteins were orthologous to LHT based on phylogenetic relationships. Therefore, we predicted that LHT had undergone rapid evolution, which was associated with an increase in the number of LHT gene and was closely related to additional physiological functions, except for the function of a lysine and histidine selective transporter (Foster et al. 2008; Tegeder and Rentsch 2010; Tegeder and Ward 2012). Moreover, some sites of AAAP proteins (e.g., *BrAAAP2*, *BrAAAP32*, and *BrAAAP64*) had a value of $\omega > 1$ (ESM Fig. S1). These AAAP proteins were orthologous to *Arabidopsis* ProT proteins, and thus we predicted that ProT proteins in *B. rapa* were under relaxed functional constraints (Rentsch et al. 1996).

LRT analysis of functional divergence showed that type I functional divergence occurred in six group comparisons

(Fig. 4 and Table 3). Interestingly, the members of the AAAP I group, including *BrAAAP30*, *BrAAAP32*, *BrAAAP39*, *BrAAAP40*, *BrAAAP64*, and *BrAAAP72*, were highly divergent from the other groups (ESM Fig. S2). These genes in the AAAP I group were orthologous to *ProT* genes in *Arabidopsis* based on phylogenetic relationships. Previous studies showed that the ProT proteins functioned mainly as low-affinity transporters for glycine, betaine, proline, and quaternary ammonium compounds (Rentsch et al. 1996, 2007; Young et al. 1999; Breitkreuz et al. 1999; Grallath et al. 2005; Tegeder and Rentsch 2010). Therefore, we predicted that *ProT* genes had undergone rapid evolution and developed distinct functions. This observation was consistent with the fact that many amino acid sites under positive selection or relaxed purifying selection were identified in clade 3. Collectively, these data provided new insights into the evolutionary history and diverse functions of the AAAP gene family.

Conclusion

The AAAP gene family members were divided into six clades. The AAAP genes were unevenly distributed on 10 chromosomes, and their diverse sequence features provided potential evidence for diversifying functions. Syntenic analysis suggested that segmental and tandem duplications contributed to the expansion of the AAAP gene family in the *B. rapa* genome. Most of BrAAAP proteins evolved under purifying selection, but some sites of the AAAP proteins from clades 1c, 2b, and 3a were subject to positive or relaxed purifying selection. Interestingly, a total of 238 type I functional divergence sites ($Q_k > 0.85$) were identified in the AAAP I/II, AAAP I/III, and AAAP III/IV comparisons, implying that these sites were responsible for functional divergence of AAAP proteins in *B. rapa*.

Acknowledgments This work was supported by National Natural Science Foundation of China (31271755) and the Ministry of Science and Technology of China (2006CB101600).

References

- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics* 22:195–201
- Babula D, Kaczmarek M, Barakat A, Delseny M, Quiros CF, Sadowski J (2003) Chromosomal mapping of *Brassica oleracea* based on ESTs from *Arabidopsis thaliana*: complexity of the comparative map. *Mol Genet Genomics* 268:656–665
- Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34:369–373
- Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 107:18724–18728
- Bennett MJ, Marchant A, Green HG, May ST, Ward SP, Millner PA, Walker AR, Schulz B, Feldmann KA (1996) *Arabidopsis* AUX1 gene: a permease-like regulator of root gravitropism. *Science* 273:948–950
- Boll M, Foltz M, Rubio-Aliaga I, Kottra G, Daniel H (2002) Functional characterization of two novel mammalian electrogenic proton-dependent amino acid co-transporters. *J Biol Chem* 277:22966–22973
- Breitkreuz KE, Shelp BJ, Fischer WN, Schwacke R, Rentsch D (1999) Identification and characterization of GABA, proline and quaternary ammonium compound transporters from *Arabidopsis thaliana*. *FEBS Lett* 450:280–284
- Chang HC, Bush DR (1997) Topology of NAT2, a prototypical example of a new family of amino acid transporters. *J Biol Chem* 272:30552–30557
- Chaudhry FA, Reimer RJ, Krizaj D, Barber D, Storm-Mathisen J, Copenhagen DR, Edwards RH (1999) Molecular analysis of system N suggests novel physiological roles in nitrogen metabolism and synaptic transmission. *Cell* 99:769–780
- Chen L, Bush DR (1997) LHT1, a lysine- and histidine-specific amino acid transporter in *Arabidopsis*. *Plant Physiol* 115:1127–1134
- Cheng F, Liu S, Jian Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang XW (2011) BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biol* 11:136
- Cheng F, Jian Wu J, Fang L, Sun S, Liu B, Lin K, Bonnema G, Wang XW (2012) Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* 7(5):e36442
- Cheung F, Trick M, Drou N, Lim YP, Park JY, Kwon SJ, Kim JA, Scott R, Pires JC, Paterson AH, Town C, Bancroft I (2009) Comparative analysis between homoeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. *Plant Cell* 21:1912–1928
- Clark JA, Amara SG (2005) Amino acid neurotransmitter transporters: structure, function, and molecular diversity. *BioEssays* 15:323–332
- Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763
- Fischer WN, Kwart M, Hummel S, Frommer WB (1995) Substrate specificity and expression profile of amino acid transporters (AAPs) in *Arabidopsis*. *J Biol Chem* 270:16315–16320
- Fischer WN, Loo DD, Koch W, Ludewig U, Boorer KJ, Tegeder M, Rentsch D, Wright EM, Frommer WB (2002) Low and high affinity amino acid H⁺-cotransporters for cellular import of neutral and charged amino acids. *Plant J* 29:717–731
- 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
- Fournier D, Andrade-Navarro MA (2011) PDBpaint, a visualization webservice to tag protein structures with sequence annotations. *Bioinformatics* 27:2605–2606
- Frommer WB, Hummel S, Riesmeier JW (1993) Expression cloning in yeast of a cDNA encoding a broad specificity amino acid permease from *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 90:5944–5948
- Frommer WB, Hummel S, Unseld M, Ninnemann O (1995) Seed and vascular expression of a high-affinity transporter for cationic amino acids in *Arabidopsis*. *Proc Natl Acad Sci U S A* 92:12036–12040
- Grallath S, Weimar T, Meyer A, Gumy C, Suter-Grotemeyer M, Neuhaus JM, Rentsch D (2005) The AtProT family. Compatible solute transporters with similar substrate specificity but differential expression patterns. *Plant Physiol* 137:117–126
- Gu X (1999) Statistical methods for testing functional divergence after gene duplication. *Mol Biol Evol* 16:1664–1674

- Gu X (2003) Functional divergence in protein (family) sequence evolution. *Genetica* 118:133–141
- Gu X, Vander-Velden K (2002) DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics* 18:500–501
- Guo AY, Zhu QH, Chen X, Luo JC (2007) GSDB: a gene structure display server. *Yi Chuan (Chinese)* 29:1023–1026
- Hirner A, Ladwig F, Stransky H, Okumoto S, Keinath M, Harms A, Frommer WB, Koch W (2006) *Arabidopsis* LHT1 is a high-affinity transporter for cellular amino acid uptake in both root epidermis and leaf mesophyll. *Plant Cell* 18:1931–1946
- Hsu LC, Chiou TJ, Chen L, Bush DR (1993) Cloning a plant amino acid transporter by functional complementation of a yeast amino acid transport mutant. *Proc Natl Acad Sci U S A* 90:7441–7445
- Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, Drabek J, Lopez R, Price HJ (2005) Evolution of genome size in *Brassicaceae*. *Ann Bot* 95:229–235
- Kong H, Landherr LL, Frohlich MW, Leebens-Mack J, Ma H, dePamphilis CW (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
- Lagercrantz U, Lydiat DJ (1996) Comparative genome mapping in *Brassica*. *Genetics* 144:1903–1910
- Lan TH, DelMonte TA, Reischmann KP, Hyman J, Kowalski SP, McFerson J, Kresovich S, Paterson AH (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. *Genome Res* 10:776–788
- Lee YH, Foster J, Chen J, Voll LM, Weber APM, Tegeder M (2007) AAP1 transports uncharged amino acids into roots of *Arabidopsis*. *Plant J* 50:305–319
- Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* 39:W475–W478
- Lu Y, Song Z, Lü K, Lian X, Cai H (2012) Molecular characterization, expression and functional analysis of the amino acid transporter gene family (*OsAATs*) in rice. *Acta Physiol Plant* 34:1943–1962
- Lukens L, Zou F, Lydiat D, Parkin I, Osborn T (2003) Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. *Genetics* 164:359–372
- Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe *Brassicaceae*. *Genome Res* 15:516–525
- Lysak MA, Cheung K, Kitzchke M, Bureš P (2007) Ancestral chromosomal blocks are triplicated in *Brassicaceae* species with varying chromosome number and genome size. *Plant Physiol* 145:402–410
- Mai B, Lipp M (1994) Cloning and chromosomal organization of a gene encoding a putative amino-acid permease from *Saccharomyces cerevisiae*. *Gene* 143(1):129–133
- McIntire SH, Reimer RJ, Schuske K, Edwards RH, Jorgensen EM (1997) Identification and characterization of the vesicular GABA transporter. *Nature* 389:870–876
- Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, Yu HJ, Kim DS, Kim N, Lim KB, Lee SI, Hahn JH, Lim YP, Bancroft I, Park BS (2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol* 10:111
- Niemelä T, Seppänen M, Jauhianen L, Tulisalo U (2006) Yield potential of spring turnip rape synthetics and composite hybrids compared with open-pollinated commercial cultivars. *Can J Plant Sci* 86:693–700
- Okumoto S, Schmidt R, Tegeder M, Fischer WN, Rentsch D, Frommer WB, Koch W (2002) High affinity amino acid transporters specifically expressed in xylem parenchyma and developing seeds of *Arabidopsis*. *J Biol Chem* 277:45338–45346
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40:290–301
- 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
- Rentsch D, Schmidt S, Tegeder M (2007) Transporters for uptake and allocation of organic nitrogen compounds in plants. *FEBS Lett* 581:2281–2289
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim BP, Zhang JZ, Ghandehari D, Sherman BK, Yu GL (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290:2105–2110
- Sagné C, Mestikawy SEI, Isambert M, Hamon M, Henry J, Giros B, Gasnier B (1997) Cloning of a functional vesicular GABA and glycine transporter by screening of genome databases. *FEBS Lett* 417:177–183
- Sagné C, Agulhon C, Ravassard P, Darmon M, Hamon M, Mestikawy SEI, Gasnier B, Giros B (2001) Identification and characterization of a lysosomal transporter for small neutral amino acids. *Proc Natl Acad Sci U S A* 98:7206–7211
- Saier M, Yen M, Noto K, Tamang D, Elkan C (2009) The Transporter Classification Database: recent advances. *Nucleic Acids Res* 37:274–278
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sugawara M, Nakanishi T, Fei YJ, Huang W, Ganapathy ME, Leibach FH, Ganapathy V (2000) Cloning of an amino acid transporter with functional characteristics and tissue expression pattern identical to that of system A. *J Biol Chem* 275:16473–16477
- Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:609–612
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC (2012) Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. *Genetics* 190:1563–1574
- Tegeder M, Rentsch D (2010) Uptake and partitioning of amino acids and peptides. *Mol Plant* 3:997–1011
- Tegeder M, Ward JM (2012) Molecular evolution of plant AAP and LHT amino acid transporters. *Front Plant Sci* 3(21):1–8
- Tegeder M, Tan Q, Grennan AK, Patrick JW (2007) Amino acid transporter expression and localisation studies in pea (*Pisum sativum*). *Funct Plant Biol* 34:1019–1028
- 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
- Wang X, Wang H, Wang J, Sun R, Wu J et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 43:1035–1039
- Ward J (2001) Identification of novel families of membrane proteins from the model plant *Arabidopsis thaliana*. *Bioinformatics* 17:560–563
- Wipf D, Ludewig U, Tegeder M, Rentsch D, Koch W, Frommer WB (2002) Conservation of amino acid transporters in fungi, plants and animals. *Trends Biochem Sci* 27:139–147
- Yang Z (2007) PAML4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917

- Yang YW, Lai KN, Tai PY, Li WH (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *J Mol Evol* 48:597–604
- Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22: 1107–1118
- Yang TJ, Kim JS, Kwon SJ, Lim KB, Choi BS, Kim JA, Jin M, Park JY, Lim MH, Kim HI, Lim YP, Kang JJ, Hong JH, Kim CB, Bhak J, Bancroft I, Park BS (2006) Sequence-level analysis of the diploidization process in the triplicated FLOWERING LOCUS C region of *Brassica rapa*. *Plant Cell* 18:1339–1347
- Young GB, Jack DL, Smith DW, Saier MH (1999) The amino acid/auxin: proton symport permease family. *Biochim Biophys Acta* 1415:306–322
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479