Phylogenetic Analysis of *Brassica rapa* MATH-Domain Proteins

Liming Zhao^{1,2,#}, Yong Huang^{1,2,#}, Yan Hu^{1,2}, Xiaoli He³, Wenhui Shen^{1,2}, Chunlin Liu^{1,4,*} and Ying Ruan^{1,2,*}

¹Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha, 410128 Hunan, China; ²College of Bioscience and Biotechnology, Hunan Agricultural University, Changsha, 410128 Hunan, China; ³Experimental Center of Hunan Agricultural University, Changsha, 410128 Hunan, China; ⁴Agricultural College of Hunan Agricultural University, Changsha, 410128 Hunan, China

Abstract: The MATH (meprin and TRAF-C homology) domain is a fold of seven anti-parallel β -helices involved in protein-protein interaction. Here, we report the identification and characterization of 90 MATH-domain proteins from the *Brassica rapa* genome. By sequence analysis together with MATH-domain proteins from other species, the *B. rapa* MATH-domain proteins can be grouped into 6 classes. Class- I protein has one or several MATH domains without any other recognizable domain; Class- II protein contains a MATH domain together with a conserved BTB (Broad Complex, Tramtrack, and Bric-a-Brac) domain; Class-III protein belongs to the MATH/Filament domain family; Class-IV protein contains a MATH domain frequently combined with some other domains; Class-V protein has a relative long sequence but contains only one MATH domain; Class-VI protein is characterized by the presence of Peptidase and UBQ (Ubiquitinylation) domains together with one MATH domain. As part of our study regarding seed development of *B. rapa*, six genes are screened by SSH (Suppression Subtractive Hybridization) and their expression levels are analyzed in combination with seed developmental stages, and expression patterns suggested that *Bra001786*, *Bra03578* and *Bra036572* may be seed development specific genes, while *Bra001787*, *Bra020541* and *Bra040904* may be involved in seed and flower organ development. This study provides the first characterization of the MATH domain proteins in *B. rapa*.

Received on: November 06, 2012- Revised on: April 12, 2013- Accepted on: April 12, 2013

Keywords: *Brassica rapa*, Phylogenetic analysis, MATH domain protein, Protein domain organization, Gene expression, Seed development.

INTRODUCTION

Meprins are mammalian tissue-specific metalloendopeptidases of the astacin family implicated in developmental, normal and pathological processes by hydrolysing a variety of proteins [1]. TRAF (TNF-receptor associated factors) proteins were first isolated through their ability to interact with TNF receptors [2]. The MATH domain (meprin and TRAF homology domain) is found in cytosolic signaling molecules such as TRAF class proteins, which are characterized by a C-terminal region encompassing about 180 amino acids, forming a 7-8 anti-parallel β -sheets fold (TRAF-C domain / MATH domain) [3-5], and the domain forms a new, light-stranded antiparallel beta sandwich structure [6]. A coiled-coil region adjacent to the MATH domain is important for oligomerisation, essential for establishing appropriate connections to form signalling complexes with TNF receptor-1 [7]. The ligand binding surface of TRAF proteins is located in beta-strands 6 and 7, and consensus motif (P/S/A/T)x (Q/E)E is the major motif of MATH domain [8]. The MATH

domain seems to be important for protein-protein interaction and several studies on human and *C. elegans* MATH proteins indicate that they might have important functions in the regulation of protein processing [9]. In TRAF proteins, the N-terminal of the MATH domain has been shown to be necessary and sufficient for self-association (homodimerization) and receptor interaction [6, 10-14].

Many MATH domain proteins that are found in humans [15-17] and mammals [2, 8, 18] are mainly involved in individual development and cell growth, differentiation, and aging as molecular adaptors. For example, TRAF6 is a critical factor for dendritic cell maturation and development [19-21]. Recently, this type of MATH proteins are also found in plants (*Arabidopsis*, *Medicago*, rice), in lower eukaryotes (Trypanosoma, Plasmodium) and in lower metazoa (*C. elegans*) [22].

MATH domain containing proteins are found usually associated to a discrete set of other protein domains, including peptidases, filamin and RluA domains, broad-complex, tramtrack and brie-a-brae (BTB) domain, tripartite motif (TRIM), astacin domain and RING and Zinc finger domains [9]. The BTB domain is an evolutionarily conserved domain broadly distributed in eukaryotes [23, 24]. At least 76 BTB-domain proteins exist in *Arabidopsis* belonging to 11 major families [25]. Proteins carrying both BTB and MATH motifs are common in plants. For example, there are at least 74

*Address correspondence to these authors at the Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha, 410128 Hunan, China; Tel: 0086-731-84635294; Fax: 0086-731-84673765; E-mails: yingruan@hotmail.com; liucl100@126.com

These authors contributed equally to this work.

likely functional *MATH-BTB* genes in rice, but also at least another 40 *MATH-BTB* pseudogenes [26], whereas, only six members (AtBPM1~6) are annotated in *Arabidopsis* [27-29]. The MATH domain of BPM (BTB/POZ-MATH) proteins is used to assemble with members of the ethylene response factor/Apetala2 (ERF/AP2) transcription factor class [30]. In addition, MATH-BTB proteins may directly interact with and target the homeobox-leucine zipper (HD-ZIP) transcription factor ATHB6 (*Arabidopsis thaliana* homeobox gene 6) for proteasomal degradation [31].

Brassicaceae-specific PSV-embedded proteins (BPEPs) have an N-terminal signal peptide and two tandem MATH domains and are localized in PSVs (protein storage vacuole), and at least one BPEP is tightly associated with the phytate contained within PSV globoids [32]. Additionally, six MATH domain proteins were identified by SSH in *Brassica napus*, among them four proteins are involved in lipin metabolism and two proteins are related to sugar, and all have two MATH domains [33].

Phylogenetic analysis of MATH domain proteins has proven helpful as a guide for genetic and molecular studies of this large family of proteins. Zapata and colleagues reported that MATH domain proteins can be divided into the USP7 family, the MATHd-only family, the MATHd/BTB family and the MATHd/Filament family in *Brassica*, but rice lacks the MATHd/Filament family; and the TRAF, TRIM37 and Meprin families are only found in animals [9]. Phylogenetic and domain organization analysis may infer to putative functions, for example, the USP7 class containing a UBP domain at the C-terminal side of MATH domains has ubiquitin proteases (UBPs) activity [9].

Here, we identified and analyzed MATH-domain proteins from *Brassica rapa* whole genome sequence [34], together with most of *Arabidopsis thaliana*, *Oryza sativa* and *Ostreococcus tauri* [35-37], and also some from animals whose functions are known [9]. Our data provide a platform for future functional characterization of these genes in *Brassica* species.

MATERIALS AND METHODS

MATH Domain Protein Identification

Sequences of MATH domain proteins from *A. thaliana* were retrieved from the TAIR Database (<http://www.arabidopsis.org>, version: 2.2.8) with the key word MATH in species database. Using the *Arabidopsis thaliana* sequences of MATH proteins, we performed a search in the *O. sativa* using the BLASTp algorithm in the NCBI browser (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Then these sequences,

primarily those from *A. thaliana* and *O. sativa*, were used as queries to search Brassica Database (<http://brassicadb.org/brad/searchAll.php>, version: 1.1) and *Ostreococcus tauri* database (<http://genome.jgi-psf.org/Ostta4/Ostta4.home.html>, version: 2.0) by using the BLASTp and tBLASTn tools (<http://blast.ncbi.nlm.nih.gov>). The Expect threshold was set at 1.0 and other parameters were set at default values.

Protein Domain Organization Analysis

The protein sequences were analyzed for domain organization using NCBI-CD searches (<http://ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). The low-complexity filter was turned off, and the Expect value was set at 1.0 to detect short domains or regions of less conservation in this analysis. Domains were also verified and named according to the SMART database (<http://smart.embl-heidelberg.de/>).

Phylogenetic Analysis

Multiple sequence alignments of MATH-domain proteins sequences were performed using the Clustal W program [38]. The full-length proteins were subjected to phylogenetic analysis using the MEGA5.1 program [39]. The trees were constructed with the following settings: Tree Inference as Neighbor-Joining; Include Sites as pair wise deletion option for total sequences analysis; Substitution Model: Poisson correction; and Bootstrap test of 1000 replicates for internal branch reliability [40].

qRT-PCR Analysis

B. rapa plants were grown at 18-22°C under a 12h light (10,000 Lx)/12h dark photoperiod. Leaves were collected from 20, 30, 35 days-old plants after artificial pollination; roots, stems and flower buds were collected from blossoming plants. Total RNA was extracted using Trizol Reagent (Invitrogen, USA) from about 100 mg of collected plant tissue. The RNA preparation was then treated with DNaseI (Promega, USA) for 30 min at 37°C, followed by enzyme inactivation by incubation at 65°C for 5min. First strand cDNA was made using an RT-PCR Kit (RevertAid™ First Strand cDNA Synthesis Kit, Fermentas, CA). The RT-solution with first strand cDNA was stored at -80°C [40]. Fluorescence-based quantitative-PCR was performed used SYBR Premix Ex Tag Reagent (Takara, Japan). Primers used for Q-PCR reactions are listed in (Table 1). Conditions for the Q-PCR reactions were as follows: 95°C for 10 min; then 40 cycles of 95°C for 15 s, 60°C for 30 s, and 72°C for 30s (AB, USA).

Table 1. qRT-PCR Primers for MATH Domain Protein Encoding Genes Involved in Seed Development.

Gene	Forward	Reverse
Bra001786	CAATCCCTCATACGACAATCCA	GTCACCTGGTTTGAAGGTGAGA
Bra001787	GGAGTTTCTCCGCGTTGGT	GCACGGCCACTTGGATTTAC
Bra035787	GTTGGGAAACTGGACGTGCTT	CCACAAGAAACCCCTTTGATG
Bra040904	CCAATGTGTGGCGTTTTAATGT	ATCCGTTTGGCGGGTCTT
Bra036572	GTTGAGAAGCGAGCGACCAT	ACTGAGGATTCTGACGGTTCAA
Bra020541	TCCCTTTGGTTGGAACTCA	GCCCCATCTGTGACAGTCAA

RESULTS

Identification and Annotation of MATH-Domain Proteins from the *Brassica rapa* Genome

Using BLASTp and tBLASTn bioinformatic tools, we identified 90 genes encoding different MATH-domain proteins from the *B. rapa* genome (<http://brassicadb.org/brad>, Table 2) compared to 63 genes in *Arabidopsis* (NCBI), 36 genes in rice (NCBI), only 2 genes in *O. tauri* database (<http://genome.jgi-psf.org>) and 16 genes from different animals (Table 2, Supplement 1). There are ten chromosomes in *Brassica rapa*, just twice to *Arabidopsis*. It suggested that genes encoding MATH domain proteins duplicated or lost during chromosome duplication. Those genes disperse to all ten chromosomes in *B. rapa* (Table 2), except the localization of Bra029646, Bra040903 and Bra040904 are still be uncertain. Ch3 and Chr9 have much more MATH domain encoding genes than the others, and both have 16 genes (Table 2).

Phylogenetic Analysis of MATH-Domain Proteins

To examine the phylogenetic relationships among *B. rapa* MATH-domain proteins and group them within the established classes, we subtracted all MATH domain proteins from several species, including 63 proteins from *A. thaliana*, 36 proteins from *O. sativa*, 2 proteins from *O. tauri*. We also included the TRAF1~6 proteins, TRIM37, meprin A and B, SPOP and HAUSP from Human, MuBM-90, MmSPOP and MmPOZ4 from Mouse, SPOPL from *Gallus gallus*, BL2960 from *Cryptococcus neoformans* and USP7 from *Drosophila melanogaster*. According to the latest report, proteins encompassing MATH domains and their association with other protein domains are grouped in 8 families, and 4 families in Brassica [9]. However, based on our analysis, MATH domain proteins in *B. rapa* were divided into 6 classes, and have not TRAF and Meprin family numbers (Fig. 1, Fig. 2 and Supplement 2). The last 3 classes belong to a branches group in the tree, but had much difference in domain organization.

Conserved Domains in MATH-Domain Proteins

Additionally, we analyzed the architecture of 90 Brassica MATH domain proteins and found that they can be classified into six families based on existence of the other conserved domains combined with sequence similarity (Fig. 2). Class-I, including 30 MATH domain proteins, only has two tandem MATH domains except Bra001787 whose domain organization is similar to Class-III including longer sequence and more MATH domains. But alignment results showed that Bra001787 is much similar to Class-I member (Supplement 3), and 74% amino acid identity with Bra040904, 25% and 27% sequence identities with Class-III member Bra007802 and Bra034251 which also have 4 MATH domains, respectively. Therefore, we classified it into Class-I, suggesting that Bra001787 might originate from Bra040904 by duplication of MATH domain tandem. All the 10 members of Class-II have MATH and BTB domain, and some also have another BACK domain in the C-terminus. Class-III was fallen into a branch which *Arabidopsis* proteins had Filament domain at the C-terminus [9], but this domain couldn't be de-

tected in *B. rapa* and *Arabidopsis* in NCBI and SMART. Class-IV contains 37 members, and is the largest class. Besides the MATH domains, some proteins have another pearl1-4 domain and Bra003329 has another Ribosomal domain in the C-terminus. But in this class, 35 proteins only contained one MATH domain. In rice, MATH only domain class is the largest family and contains about 50% MATH domain proteins [26]. Taken together, MATH only proteins have the largest members. Class-V proteins are larger than the members in Class-I to IV except that Bra023279 only have a MATH domain. Both Class-IV and Class-V classes belonging to same branch (Fig. 1) and having similar MATH domain may have some uncertain relation. Class-VI numbers are also longer than first four classes and all have MATH, UBP and Peptidase domains. Previous researches suggested that the last 2 classes have ubiquitin proteases (UBPs) activity and may be derived from a common ancestor, but some sequences lacked the ubiquitin protease domain during evolutionary process [9].

Expression Analyses of MATH Proteins Encoding Genes in Seed Development

SSH libraries from 20 /30 days seed development of *Brassica napus* were constructed [33], and Bra020541 is homolog to EST T200139, which was screened in the 20-day-old seed SSH library, and was weakly expressed in root, stem, and leaf, but highly in young flower bud. During seed development, the expression level decreased more dramatically in seed at the 20-day than that at 35-day. Bra001786, Bra001787, Bra040904 and Bra035787 are all homolog to EST T350008, and Bra036572 is homolog to EST T350054 in the 30-day-old seed SSH library, respectively [33]. Expression of Bra001786 or Bra036572 was only detected in seed and gradually increased during seed development. Bra001787 had stronger expression in leaf and flower bud than in seed, and no expression in root and stem, and the expression level in seed development reached the peak at 30 days. Bra035787 was highly expressed in root, leaf, flower bud and seed, but less in stem, and the expression level increased with seed development process. Bra040904 was expressed in all detected organs, and much higher in flower bud and stem than in root and leaf, and the expression level increased and reached the top in seed at 30 days during seed development (Fig. 3). Among those genes, Bra036572 expressed much higher in seed at 30 and 35 days than other genes, especially Bra020541.

According to expression of genes in seed development process, there are three expression patterns. Bra020541 expresses in early stage of seed development. Bra001786 and Bra036572 have much similar expression pattern: the expression level increases with seed development process, reaches the top in seed at 35 days, and undetectable expression in root, stem and flower bud. Similarly, Bra035787 expresses much higher in seed than other organs, and also reaches the top in seed at 35days, suggesting that Bra001786, Bra036572 and Bra035787 are late developmental stage seed genes. Bra001787 and Bra040904 both express higher in seed at 30 days, suggesting that they may be the regulation genes at middle stage of seed development.

Table 2. MATH Domain Proteins in *Brassica rapa*.

Name	Arabidopsis Synteny	ORF Length (bp)	Protein Length (aa)	Genomic Length (bp)	No. of Introns	Chromosome
Class-I						
Bra000640	AT4G09770	963	320	1460	4	3
Bra001660	AT3G17380	864	287	1345	4	3
Bra001785	AT3G20360	849	282	1499	4	3
Bra001786	AT3G20360	1200	399	10892	6	3
Bra001787	AT3G20360	3993	1330	10191	14	3
Bra003836	UN	684	227	1481	2	7
Bra003837	UN	1014	337	1660	4	7
Bra003989	UN	873	290	1873	2	7
Bra003990	AT1G69650	903	300	1407	4	7
Bra004405	AT1G69650	1101	366	1666	5	7
Bra005554	AT2G32870	927	308	1565	5	5
Bra007873	AT1G69650	948	315	1308	4	2
Bra009894	AT5G26280	768	255	1216	4	6
Bra009895	AT5G26280	1053	350	1705	6	6
Bra015981	UN	963	382	1783	4	7
Bra020540	AT5G26280	900	299	1921	5	2
Bra020541	AT5G26280	1056	351	1886	5	2
Bra022923	AT2G32870	921	306	1865	5	3
Bra023852	AT3G22080	963	320	1509	4	1
Bra031322	AT3G22080	933	310	1285	4	5
Bra032150	UN	1074	357	1506	4	4
Bra034175	UN	1005	334	3167	1	1
Bra035787	AT3G20370	1215	404	1694	5	5
Bra036572	AT5G26280	1035	344	1753	6	9
Bra036573	AT5G26280	1053	350	1864	5	9
Bra037358	AT4G00780	867	288	1269	4	9
Bra037411	AT4G01390	861	286	1225	4	9
Bra039646	UN	1332	443	2936	5	Scaffold000172
Bra040903	UN	1149	382	1862	5	Scaffold000303
Bra040904	UN	1149	382	1813	5	Scaffold000303
Class-II						
Bra000147	AT2G39760	1149	382	2759	3	3
Bra001186	AT3G06190	1212	403	1815	3	3
Bra002200	AT5G19000	1203	400	2348	3	10
Bra006489	AT5G19000	1212	403	1969	3	3

(Table 2) contd....

Name	Arabidopsis Synteny	ORF Length (bp)	Protein Length (aa)	Genomic Length (bp)	No. of Introns	Chromosome
Class-II						
Bra006591	AT5G21010	1233	410	2144	3	3
Bra017053	AT2G39760	1200	399	1941	3	4
Bra020142	AT5G21010	1248	415	2220	3	2
Bra020764	AT3G06190	1158	385	1819	3	2
Bra036437	AT3G03740	1341	446	2703	3	1
Bra040233	AT3G06190	1170	389	2174	3	7
Class-III						
Bra007802	AT2G25320	4995	1664	5968	9	9
Bra034251	AT2G25320	4944	1647	5945	9	4
Class-IV						
Bra000258	AT2G42460	924	307	1861	4	3
Bra000259	AT2G42460	861	286	1335	3	3
Bra002613	UN	1257	418	1636	3	10
Bra002615	UN	1347	448	1665	3	10
Bra003329	AT3G58230	885	294	1595	4	7
Bra003330	AT3G58230	1440	479	6302	7	7
Bra003331	AT3G58230	981	326	3134	5	7
Bra003332	AT3G58230	2016	671	2659	5	7
Bra003337	AT3G58400	1740	579	4083	5	7
Bra004695	AT2G42460	1215	404	1593	3	5
Bra004696	AT2G42460	1392	463	1778	3	5
Bra004697	AT2G42460	1518	505	2209	4	5
Bra004699	AT2G42460	1371	456	1848	3	5
Bra004700	AT2G42460	1896	631	2301	2	5
Bra007386	AT3G58200	981	326	1319	3	9
Bra007387	AT3G58200	1416	471	2662	6	9
Bra007388	AT3G58200	915	304	1317	3	9
Bra007389	AT3G58200	960	319	1302	3	9
Bra007390	AT3G58200	741	246	1077	3	9
Bra007392	AT3G58200	981	326	1393	3	9
Bra008725	UN	909	302	1209	3	10
Bra012666	UN	819	272	1252	3	3
Bra014088	UN	810	269	1859	4	8
Bra014572	AT3G58400	1347	448	2885	4	4
Bra014574	AT3G58320	855	284	1670	3	4

(Table 2) contd....

Name	Arabidopsis Synteny	ORF Length (bp)	Protein Length (aa)	Genomic Length (bp)	No. of Introns	Chromosome
Class-IV						
Bra014575	AT3G58320	870	289	1538	3	4
Bra015255	UN	909	302	1198	3	10
Bra015948	UN	843	280	1208	3	7
Bra016873	AT2G42460	813	270	1281	3	4
Bra016874	AT2G42460	840	279	1078	2	4
Bra016875	AT2G42460	858	285	1034	2	4
Bra027149	UN	888	295	1209	3	9
Bra031498	UN	1062	353	1459	3	1
Bra031517	UN	1017	338	1287	3	1
Bra032543	UN	831	276	1207	4	9
Bra032595	UN	1623	540	2136	5	9
Bra035414	UN	1071	356	1559	3	1
Class-V						
Bra015302	AT1G04300	3183	1060	4995	12	10
Bra023279	UN	1050	349	1860	6	9
Bra030564	AT1G04300	5052	1683	8328	13	8
Bra032518	AT1G04300	2985	994	4568	10	9
Bra033684	AT5G43560	2784	927	4176	11	6
Class-VI						
Bra001426	AT3G11910	3339	1112	8265	32	3
Bra005900	AT5G06600	3348	1115	8460	32	3
Bra009210	AT5G06600	3303	1100	7823	31	3
Bra028727	AT5G06600	3417	1138	8289	32	2
Bra034802	AT3G11910	3348	1115	8732	32	5
Bra038685	AT3G11910	3309	1102	7657	31	1

Data coming from web <http://brassicadb.org/brad/>.

UN=uncertainly.

DISCUSSION

MATH domains seem to be very important for the regulation of protein processing [5, 6, 9-13]. Except class- I and class-III, all have additional domains, such as BTB, UBQ, BACK, and GAF domains (Fig. 2). The BTB domain also known as the POZ (for Pox virus and Zinc finger) domain [41], is an evolutionarily conserved domain broadly distributed in eukaryotes [23, 24]. Proteins carrying both BTB and MATH motifs are common in plants [23, 24, 41, 42]. Arabidopsis has 6 numbers AtBMP1-AtBMP6, but Brassica has 10 numbers (Fig. 1), and rice has more than 30 numbers. It suggested that those genes divided and evolved after speciation. Ubiquitin domain proteins (UDPs) and ubiquitin-like

(UBL) domain proteins belong to a diverse group of proteins which are characterized by an integral UBQ or UBL domain. The majority of UDPs described so far are components of the ubiquitin system which is crucial for the degradation of most cellular proteins [43, 44]. In Arabidopsis, the EVE1 protein containing a 52 amino acid ubiquitin domain (UBQ) is involved in the control of inflorescence stem formation [45], and so far 7 UBQ domain proteins have been identified (Fig. 1 and Fig. 2).

Previous reports have shown that Brassica MATH proteins fall into four classes [9], but we suppose that six classes may be more reasonable according to phylogenetic analysis and specific domain architectures in the whole genome level.

Fig. (1). Phylogenetic tree of MATH-domain proteins in *Brassica rapa*. The MATH domain sequences of the different 185 proteins (90 proteins in *B. rapa*, 63 proteins in *Arabidopsis*, 36 proteins in rice, 2 proteins in *O. tauris*, 2 proteins in mouse, 8 proteins in human, one in bee, chicken and *Drosophila melanogaster* were aligned using Clustal W, and the phylogenetic tree analysis was performed using MEGA4.0. The trees were constructed with the following settings: Tree Inference as Neighbor-Joining; Include Sites as pairwise deletion option for total sequences analysis; Substitution Model: Poisson correction; and Bootstrap test of 1000 replicates for internal branch reliability. Supplement 2 is the tree based on maximum likelihood, and both trees are much similar.

Class-I MATH-domain-only proteins is a large number of hypothetical proteins containing multiple MATH-domains in tandem except only one in Bra003836. Proteins encompassing a MATH domain and a BTB/POZ domain are broadly represented in eukaryotes (Class-II). Many reports showed MATH-BTB proteins interacted with CUL-3 carried had E3 ubiquitin ligase activity [26, 31, 46, 47], and MAB1 expressed in the germ lineages and the zygote of maize [48]. Here, ten MATH-BTB domain proteins present in *B. rapa*, and conserved domain organization suggests that they might have similar function. Comparisons of sequence similarity and synteny of *B. rapa* and *A. thaliana* MATH-domain proteins revealed occurrence of recent gene duplication events (<http://brassicadb.org/brad/searchSynteny.php>). Bra00200,

Bra006489 and Bra023700 show synteny to AtBPM1, but the putative protein of Bra023700 has 170 aa without MATH domain, suggesting that it may shift its function during the evolution, and also fails to be clustered with others; Brao20764, Bra040233 and Bra001186 show synteny to AtBPM2; Both Bra017053 and Bra000147 show synteny to AtBPM3; Bra036437 show synteny to AtBPM4; Bra006591 and Bra020142 show synteny to AtBPM5, but no sequence show synteny to AtBPM6. It suggests MATH-BTB genes diploidization or triploidization or losing in *B. rapa*. Class-III only has two Brassica proteins and two Arabidopsis proteins, and previous research showed that Arabidopsis proteins had FILAMENT domain, but they can't be detected here. Ubiquitin Proteases (UBPs) are also found

Fig. (2). Domain architecture of the different classes of *Brassica rapa* MATH domain proteins. Domains were identified using the Conserved Domain Search service (<http://ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and verified and named according to the SMART database (<http://smart.embl-heidelberg.de/>). The low-complexity filter was turned off, and the Expect value was set at 1.0 to detect short domains or regions of less conservation in this analysis. Class-I only had two tandem MATH domain except Bra001787; Class-II had MATH and BTB domain; Class-III had Filament domain at the C-terminus in *Arabidopsis* [9], but this domain couldn't be detected in *B.rapa*; Class-IV had one or two MATH domains, in addition to that, some proteins also has a pearl1-4 domain; Class- proteins were longer than others, and Class-members had MATH, UBQ and Peptidase domain.

Fig. (3). The expression of encoding MATH domain proteins genes involved in seed development. Expression pattern suggest that *Bra001786*, *Bra03578* and *Bra036572* may be seed development special genes, but *Bra001787*, *Bra020541* and *Bra040904* may involve in seed and flower organ development. Material: seed, from 20, 25, 30, 35 days young seed after artificial pollination; Roots, from principal root of 15 days seedling; Stems, from young stem of 15 days seedling; Leaf, from young leaf of 15 days seedling; Bud: from young flower bud, length $\approx$ 0.5cm.

in plants (*Arabidopsis*, *Oryza*) and in Metazoa [9]. Except that Bra003330 and Bra003331, Class-IV proteins have only one MATH domain, and about half of them as well as have unknown function PEARLI-4 domain (*Arabidopsis* phospholipase-like protein). AtRTM3 belonging to Class-IV, is the first biological function identified in a resistance mechanism in plant, encodes a MATH and CC domain protein [5], and restricts plant viruses long distance movement [49], and its syntenic gene Bra014574 may have similar function. All MATH domains in Class-V are localized at the N-terminal except Bra030564 in the middle. Interestingly, Class-VI has 7 putative proteins with UBP domain which have ubiquitin proteases (UBPs) activity and Peptidase-C19 domain.

MATH domain proteins are found as cytosolic signaling molecules in animals [3]. Less plant MATH domain proteins are identified, such as, RTM3 (belong to Class-IV) can restricts plant viruses long distance movement [5, 49], and MATH-BTB domain proteins (Class-II) directly interact with and target transcription factor ATHB6 for proteasomal degradation [31]. Our previous research showed that 6 MATH domain proteins are involved in seed development of *Brassica nupas* [33], here 6 *B.rapa* homologous genes Bra001786, Bra001787, Bra020541, Bra036572, Bra035787 and Bra040904 are identified and all have only MATH domain and belong to Class-I. The expression of Bra020541, homolog to EST T200139 in 20-day-old seed SSH library, Bra001786, Bra001787, Bra040904, Bra035787, homolog to EST T350008 and Bra036572, homolog to EST T350054 in 35-day-old seed SSH library [33]. The expression pattern suggest that Bra001786, Bra03578 and Bra036572 may be seed development specific genes, but Bra001787, Bra020541 and Bra040904 may be involved in seed and flower organ development, indicating that MATH domain proteins may have common and/or separate functions during the evolution.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

This work was supported in part by National Basic Research Program of China (973 Program, 2012CB910500), Natural Science Foundation of China (NSFC31071129, NSFC31071455 and NSFC31140010), Hunan Provincial Natural Science Foundation of China (11JJ4024) and Scientific Research Fund of Hunan Provincial Education Department of China (11C0666).

SUPPLEMENTARY MATERIALS

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

- [1] Marchand, P.; Tang, J.; Johnson, G.D.; Bond, J.S. COOH-terminal proteolytic processing of secreted and membrane forms of the alpha subunit of the metalloprotease meprin A. Requirement of the I domain for processing in the endoplasmic reticulum. *J Biol Chem*, **1995**, 270: 5449-5456.
- [2] Rothe, M.; Wong, S.C.; Henzel, W.J.; Goeddel, D.V. A novel family of putative signal transducers associated with the cytoplasmic domain of the 75 kDa tumor necrosis factor receptor. *Cell*, **1994**, 78(4): 681-692.
- [3] Sunnerhagen, M.; Pursglove, S.; Fladvad, M. The new MATH: homology suggests shared binding surfaces in meprin tetramers and TRAF trimers. *FEBS Lett*, **2002**, 530(1-3): 1-3.
- [4] Uren, A.G.; Vaux, D.L. TRAF proteins and meprins share a conserved domain. *Trends Biochem Sci*, **1996**, 21(7): 244-245.
- [5] Cosson, P.; Sofer, L.; Le, Q.H.; Léger, V.; Schurdi-Levraud, V.; Whitham, S.A.; Yamamoto, M.; Gopalan, S.; Gall, O.L.; Candresse, T.; Carrington, J.C.; Revers, F. RTM3, which controls long distance movement of potyviruses, is a member of a new plant gene family encoding a Meprin and TRAF homology domain-containing protein. *Plant Physiol*, **2010**, 154(1): 222-232.
- [6] Park, Y.C.; Burkitt, V.; Villa, A.R.; Tong, L.; Wu, H. Structural basis for self-association and receptor recognition of human TRAF2. *Nature*, **1999**, 398(6727): 533-538.
- [7] Lotocki, G.; Alonso, O.F.; Dietrich, W.D.; Keane, R.W. Tumor necrosis factor receptor 1 and its signaling intermediates are recruited to lipid rafts in the traumatized brain. *J Neurosci*, **2004**, 24(49): 11010-11016.
- [8] Ye, H.; Park, Y.C.; Kreishman, M.; Kieff, E.; Wu, H. The structural basis for the recognition of diverse receptor sequences by TRAF2. *Mol Cell*, **1999**, 4(3): 321-330.
- [9] Zapata, J.M.; Martinez-Garcia, V.; Lefebvre, S. Phylogeny of the TRAF/MATH domain. *Adv Exp Med Biol*, **2007**, 597: 1-24.
- [10] Park, Y.C.; Ye, H.; Hsia, C.; Segal, D.; Rich, R.L.; Liou, H.C.; Myszkowski, D.G.; Wu, H. A novel mechanism of TRAF signaling revealed by structural and functional analyses of the TRADD-TRAF2 interaction. *Cell*, **2000**, 101(7): 777-787.
- [11] Yeh, W.C.; Hakem, R.; Woo, M.; Mak, T.W. Gene targeting in the analysis of mammalian apoptosis and TNF receptor superfamily signaling. *Immunol Rev*, **1999**, 169(1): 283-302.
- [12] Cheng, G.; Cleary, A.M.; Ye, Z.S.; Hong, D.I.; Lederman, S.; Baltimore, D. Involvement of CRAF1, a relative of TRAF, in CD40 signaling. *Science*, **1995**, 267(5203): 1494-1498.
- [13] Chan, F.K.; Chun, H.J.; Zheng, L.; Siegel, R.M.; Bui, K.L.; Leonardo, M.J. A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling. *Science*, **2000**, 288(5475): 2351-2354.
- [14] Kwon, J.E.; La, M.; Oh, K.H.; Oh, Y.M.; Kim, G.R.; Seol, J.H.; Baek, S.H.; Chiba, T.; Tanaka, K.; Bang, O.S.; Joe, C.O.; Chung, C.H. BTB Domain-containing Speckle-type POZ Protein (SPOP) serves as an adaptor of Daxx for ubiquitination by Cul3-based ubiquitin ligase. *J Biol Chem*, **2006**, 281(18): 12664-12672.
- [15] Zapata, J.M.; Pawlowski, K.; Haas, E.; Ware, C.F.; Godzik, A.; Reed, J.C. A diverse family of proteins containing tumor necrosis factor receptor-associated factor domains. *J Biol Chem*, **2001**, 276(26): 24242-24252.
- [16] Ozato, K.; Shin, D.M.; Chang, T.H.; Morse, H.C. TRIM family proteins and their emerging roles in innate immunity. *Nat Rev Immunol*, **2008**, 8(11): 849-860.
- [17] Arthur, J.F.; Shen, Y.; Gardiner, E.E.; Coleman, L.; Kenny, D.; Andrews, R.K.; Berndt, M.C. TNF Receptor-Associated Factor 4 (TRAF4) is a novel binding partner of glycoprotein Ib and glycoprotein VI in human platelets. *J Thromb Haemost*, **2011**, 9(1): 163-172.
- [18] Brink, R.; Lodish, H.F. Tumor necrosis factor receptor (TNFR) associated factor 2A (TRAF2A), a TRAF2 splice variant with an extended RING finger domain that inhibits TNFR2-mediated NF-kB activation. *J Biol Chem*, **1998**, 273(7): 4129-4134.
- [19] Kobayashi, T.; Walsh, P.T.; Walsh, M.C.; Speirs, K.M.; Chiffoleau, E.; King, C.G.; Hancock, W.W.; Caamano, J.H.; Hunter, C.A.; Scott, P.; Turka, L.A.; Choi, Y. TRAF6 is a critical factor for dendritic cell maturation and development. *Immunity*, **2003**, 19(3): 353-363.
- [20] He, L.; Wu, X.; Siegel, R.; Lipsky, P.E. TRAF6 regulates cell fate decisions by inducing caspase 8-dependent apoptosis and the activation of NF-kB. *J Biol Chem*, **2006**, 281(16): 11235-11249.
- [21] Kongchum, P.; Hallerman, E.M.; Hulata, G.; David, L.; Palti, Y. Molecular cloning, characterization and expression analysis of TLR9, MyD88 and TRAF6 genes in common carp (*Cyprinus carpio*). *Fish & Shellfish Immun*, **2011**, 30(1): 361-371.
- [22] Peškan-Berghöfer, T.; Shahollari, B.; Giong, P.H.; Hehl, S.; Markert, C.; Blanke, V.; Kost, G.; Varma, A.; Oelmüller, R. Association of *Piriformospora indica* with *Arabidopsis thaliana* roots represents a novel system to study beneficial plant-microbe

- interactions and involves early plant protein modifications in the endoplasmic reticulum and at the plasma membrane. *Physiol Plantarum*, **2004**, 122(4): 465-477.
- [23] Bardwell, V.J.; Treisman, R. The POZ domain: A conserved protein-protein interaction motif. *Genes Dev*, **1994**, 8(14): 1664-1677.
- [24] Stogios, P.J.; Downs, G.S.; Jauhal, J.J.; Nandra, S.K.; Privé, G.G. Sequence and structural analysis of BTB domain proteins. *Genome Biol*, **2005**, 6(10): R82.
- [25] Dieterle, M.; Thomann, A.; Renou, J.P.; Parmentier, Y.; Cognat, V.; Lemonnier, G.; Müller, R.; Shen, W.H.; Kretsch, T.; Genschik, P. Molecular and functional characterization of Arabidopsis Cullin 3A. *Plant J*, **2005**, 41(3): 386-399.
- [26] Gingerich, D.J.; Hanada, K.; Shiu, S.H.; Vierstra, R.D. Large-scale, lineage-specific expansion of a bric-a-brac / tramtrack / broad complex ubiquitin-ligase gene family in rice. *Plant Cell*, **2007**, 19(8): 2329-2348.
- [27] Weber, H.; Bernhardt, A.; Dieterle, M.; Hano, P.; Mutlu, A.; Estelle, M.; Genschik, P.; Hellmann, H. Arabidopsis AtCUL3a and AtCUL3b form complexes with members of the BTB/POZ-MATH protein family. *Plant Physiol*, **2005**, 137(1): 83-93.
- [28] Weber, H.; Hano, P.; Hellmann, H. The charming complexity of CUL3. *Int J Plant Dev Biol*, **2007**, 1(1): 178-184.
- [29] Nagai, Y.; Kojima, T.; Muro, Y.; Hachiya, T.; Nishizawa, Y.; Wakabayashi, T.; Hagiwara, M. Identification of a novel nuclear speckle-type protein, SPOP. *FEBS Lett*, **1997**, 418(1-2): 23-26.
- [30] Weber, H.; Hellmann, H. *Arabidopsis thaliana* BTB/POZ-MATH proteins interact with members of the ERF/AP2 transcription factor family. *FEBS*, **2009**, 276(22): 6624-6635.
- [31] Lechner, E.; Leonhardt, N.; Eisler, H.; Parmentier, Y.; Alioua, M.; Jacquet, H.; Leung, J.; Genschik, P. MATH/BTB CRL3 peptidases target the Homeodomain-Leucine Zipper ATHB6 to modulate Absciscic acid signaling. *Dev Cell*, **2011**, 21(6): 1116-1128.
- [32] Gillespie, J.; Rogers, S.W.; Deery, M.; Dupree, P.; Roger, J.C. A unique family of proteins associated with internalized membranes in protein storage vacuoles of the Brassicaceae. *Plant J*, **2005**, 41(3): 429-441.
- [33] Peng, Q.; Hu, Y.; Du, P.F.; Xie, Q.X.; Ruan, Y.; Liu, C.L. Construction of SSH library with different stages of seeds development in *Brassica napus* L. *Acta Agron Sin*, **2009**, 35(9): 1576-1583.
- [34] Wang, X.; Wang, H.; Wang, J.; Sun, R.; Wu, J.; Liu, S.; Bai, Y.; Mun, J.H.; Bancroft, I.; Cheng, F.; Huang, S.; Li, X.; Hua, W.; Wang, J.; Wang, X.; Freeling, M.; Pires, J.C.; Paterson, A.H.; Chalhoub, B.; Wang, B.; Hayward, A.; Sharpe, A.G.; Park, B.S.; Weisshaar, B.; Liu, B.; Li, B.; Liu, B.; Tong, C.; Song, C.; Duran, C.; Peng, C.; Geng, C.; Koh, C.; Lin, C.; Edwards, D.; Mu, D.; Shen, D.; Soumpourou, E.; Li, F.; Fraser, F.; Conant, G.; Lassalle, G.; King, G.J.; Bonnema, G.; Tang, H.; Wang, H.; Belcram, H.; Zhou, H.; Hirakawa, H.; Abe, H.; Guo, H.; Wang, H.; Jin, H.; Parkin, I.A.; Batley, J.; Kim, J.S.; Just, J.; Li, J.; Xu, J.; Deng, J.; Kim, J.A.; Li, J.; Yu, J.; Meng, J.; Wang, J.; Min, J.; Poulain, J.; Wang, J.; Hatakeyama, K.; Wu, K.; Wang, L.; Fang, L.; Trick, M.; Links, M.G.; Zhao, M.; Jin, M.; Ramchiary, N.; Drou, N.; Berkman, P.J.; Cai, Q.; Huang, Q.; Li, R.; Tabata, S.; Cheng, S.; Zhang, S.; Zhang, S.; Huang, S.; Sato, S.; Sun, S.; Kwon, S.J.; Choi, S.R.; Lee, T.H.; Fan, W.; Zhao, X.; Tan, X.; Xu, X.; Wang, Y.; Qiu, Y.; Yin, Y.; Li, Y.; Du, Y.; Liao, Y.; Lim, Y.; Narusaka, Y.; Wang, Y.; Wang, Z.; Li, Z.; Wang, Z.; Xiong, Z.; Zhang, Z. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*, **2011**, 43(10): 1035-1039.
- [35] Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **2000**, 408(6814): 796-815.
- [36] Palenik, B.; Grimwood, J.; Aerts, A.; Rouzé, P.; Salamov, A.; Putnam, N.; Dupont, C.; Jorgensen, R.; Derelle, E.; Rombauts, S.; Zhou, K.; Otillar, R.; Mercchant, S.S.; Podell, S.; Gaasterland, T.; Napoli, C.; Gendler, K.; Manuek, A.; Tai, V.; Vallon, O.; Piganeau, G.; Jancek, S.; Heijde, M.; Jabbari, K.; Bowler, C.; Lohr, M.; Robbens, S.; Wernner, G.; Dubchak, I.; Pazour G.J.; Ren, Q.; Paulsen, L.; Delwiche, C.; Schmutz, J.; Rokhsar, D.; de Peer, Y.V.; Moreau, H.; Grigoriev, I.V. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *PNAS*, **2007**, 104(18): 7705-7710.
- [37] Goff, S.A.; Ricke, D.; Lan, T. H.; Presting, G.; Wang, R.; Dunn, M.; Glazebrook, J.; Sessions, A.; Oeller, P.; Varma, H.; Hadley, D.; Martin, C.; Katagiri, F.; Lange, B. M.; Moughamer, T.; Xia, Y.; Budworth, P.; Zhong, J.; Miguel, T.; Paszkowski, U.; Zhang, S.; Colbert, M.; Sun, W.L.; Chen, L.; Cooper, B.; Park, S.; Wood, T.C.; Mao, L.; Quail, P.; Dean, R.; Yu, Y.; Zharkikh, A.; Shen, R.; Saharabudhe, S.; Thomas, A.; Cannings, R.; Gutin, A.; Pruss, D.; Reid, J.; Tavtigian, S.; Mitchell, J.; Eldredge, G.; Scholl, T.; Miller, R.M.; Bhatnagar, S.; Adey, N.; Rubano, T.; Tusneem, N.; Robinson, R.; Feldhaus, J.; Macalma, T.; Oliphant, A.; Briggs, S. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). *Science*, **2002**, 296(5565): 92-100.
- [38] Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **1994**, 22(22): 4673-4680.
- [39] Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*, **2011**, 28: 2731-2739.
- [40] Huang, Y.; Liu, C.; Shen, W.H.; Ruan, Y. Phylogenetic analysis and classification of the *Brassica rapa* SET-domain protein family. *BMC Plant Biol*, **2011**, 11: 175.
- [41] Ahmad, K.F.; Engel, C.K.; Privé, G. G. Crystal structure of the BTB domain from PLZF. *PNAS*, **1995**, 95(21): 12123-12128.
- [42] Albagli, O.; Dhordain, P.; Dewindt, C.; Lecocq, G.; Leprince, D. The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell Growth Differ*, **1995**, 6(9): 1193-1198.
- [43] Madsen, L.; Schulze, A.; Seeger, M.; Hartmann-Petersen, R. Ubiquitin domain proteins in disease. *BMC Biochem*, **2007**, 8 (Suppl 1): s1.
- [44] Zhang, W.; Savelieva, K.V.; Suwanichkul, A.; Small, D.L.; Kirkpatrick, L.L.; Xu, N.; Lanthorn, T.H.; Ye, G.L. Transmembrane and ubiquitin-like domain containing 1 (Tmub1) regulates locomotor activity and wakefulness in mice and interacts with CAMLG. *PLoS One*, **2010**, 5(6): e11261.
- [45] Hwang, H.J.; Kim, H.; Jeong, Y.M.; Choi, M.Y.; Lee, S.Y.; Kim, S.G. Overexpression of EVE1, a novel ubiquitin family protein, arrests inflorescence stem development in *Arabidopsis*. *J Exp Bot*, **2011**, 62(13): 4571-4581.
- [46] Zhuang, M.; Calabrese, M.F.; Liu, J.; Waddell, M.B.; Nourse, A.; Hammel, M.; Miller, D.J.; Walden, H.; Duda, D.M.; Seyedin, S.N.; Hoggard, T.; Harper, J.W.; White, K.P.; Schulman, B.A. Structures of SPOP-Substrate Complexes: Insights into molecular architectures of BTB-Cul3 ubiquitin ligases. *Mol Cell*, **2009**, 36(1): 39-50.
- [47] Mazzucotelli, E.; Belloni, S.; Marone, D.; Leonardis, A.M. De.; Guerra, D.; Fonzo, N. Di.; Cattivelli, L.; Mastrangelo, A.M. The E3 ubiquitin ligase gene family in plants: Regulation by degradation. *Curr Genomics*, **2006**, 7(8): 509-522.
- [48] Juranić, M.; Srilunchang, K.; Krohn, N.G.; Leljak-Levanić, D.; Sprunck, S.; Dresselhaus, T.; Germline-Specific MATH-BTB substrate adaptor MAB1 regulates spindle length and nuclei identity in Maize. *Plant Cell*, **2012**, 24(12): 4974-4991.
- [49] Cosson, P.; Sofer, L.; Schurdi-Levraud, V.; Revers, F. A member of a new plant gene family encoding a meprin and TRAF homology (MATH) domain containing protein is involved in restriction of long distance movement of plant viruses. *Plant Signal Behav*, **2010**, 5(10): 1321-1323.