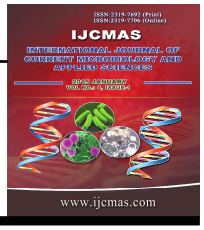
Int.J.Curr.Microbiol.App.Sci (2015) 4(3): 721-732

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 4 Number 3 (2015) pp. 721-732**
http://www.ijcmas.com



Original Research Article

Bioinformatics study of Tocopherol biosynthesis pathway genes in *Brassica rapa*

Essa Ali[1†], Kaleem Ullah Kakar[2†], Jawad Munawar Shah[1†], Mumtaz Ali Saand[3†], Nazim Hussain[1†], Du Xue[1†] and Lixi Jiang[1†]*

[1]Zhejiang Key Laboratory of Crop Gene Resources, College of Agriculture and Biotechnology, Zhejiang University, 866 Yu-Hang-Tang Road, Hangzhou 310058, PR China
[2]State Key Laboratory of Rice Biology, Zhejiang University, PR China
[3]Department of plant pathology, Zhejiang University, PR China
**Corresponding author*
† All authors equally contributed

Keywords

Tocopherols, *Brassica rapa*, Phylogenetic analysis, antioxidant activity, gene structure

A B S T R A C T

Vitamin E comprises four tocopherols and four tocotrienols, collectively termed tocochromanols that play an essential role as antioxidants in humans, animals and photosynthetic organisms and are also believed to play a role in modulation of signal transduction and gene expression pathways. In *Brassica rapa* we identified 9 tocopherols biosynthesis genes. In this study we find the subcellular localization and chromosomal distribution of these genes. In addition domain structure and phylogentic analysis were also carried out. This information is expected to be helpful for further functional characterization of tocopherols biosynthesis genes in different plant tissues under diverse growth conditions.

Introduction

Plants produce numerous organic compounds that not only perform vital functions in plant cells but also are essential or beneficial in human nutrition. One such class of compounds consists of different derivatives, which are collectively known as vitamin E. Vitamin E is a lipid-soluble molecule that covers a family of eight structurally related derivative with different biological activity. The biological activity of each vitamin E derivatives is the measure of the potency or functional use in the body (Traber and Packer, 1995; Lichtenthaler *et al*., 1997).

Many of the proposed tocopherol functions in animals and plants are related to their antioxidant properties, the most prominent of which is protection of polyunsaturated fatty acids from lipid peroxidation by quenching and scavenging various reactive oxygen species (ROS) including singlet oxygen, superoxide radicals, and alkyl peroxy radicals (Fukuzawa and Gebicky, 1983; Munne-Bosch and Alegre, 2002a).

In plants, tocopherol levels and composition vary in different tissues and fluctuate during development and in response to abiotic stresses. Dry and germinating seeds of many

plants accumulate predominantly gama-tocopherol, whereas alpha-tocopherol is the major tocopherol in leaves, which may reflect distinct roles of individual tocopherols in these tissues (Shintani and DellaPenna, 1998). Significant increases in leaf alpha-tocopherol levels are observed during aging and senescing of plants (Tramontano *et al*., 1992), possibly to protect cellular components from increased oxidative stress (Munne-Bosch and Alegre, 2002b). Enhanced tocopherol accumulation also occurs in response to a variety of abiotic stresses including high light, drought, salt, and cold and may provide an additional line of protection from oxidative damage (Munne-Bosch and Alegre, 2002a).

Historically, the purification and analysis of biosynthesis pathway enzymes from plant sources has been extremely difficult due to many factors like low specific activity, poor stability and membrane association of these enzymes. Different molecular approaches have contributed towards much of the understanding of genes and enzymes of these pathways. In the model dicot plant, Arabidopsis thaliana, and the model cyanobacterium, Synechocystis sp. PCC6803, molecular genetics and biochemical genomics-based approaches have revealed all the genes coding for tocochromanol pathway enzymes (Dellapenna and Pogson, 2006). The regulation, activities, integration, and evolution of individual enzymes have been studied using mutant and transgenic approaches. There have been many reports where important crop plants having different combinations and proportions of these compounds are generated, to study their role in plant growth, development and stress response (Hunter and Cahoon, 2007).

Here we report the identification and analysis of tocopherols biosynthesis genes in *Brassica rapa*. To study the mechanisms responsible for differential abundance of vitamin E compounds, the genes responsible for tocopherols biosynthesis were identified and studied in detail. Genomic organization and phylogenetic relationships of these genes with other known tocopherols biosynthesis pathway enzymes was undertaken to study the potential mechanisms responsible for the evolution of these genes within and among species.

Materials and Methods

Identification of tocopherols biosynthesis genes in *Brassica rapa*

In Arabidopsis, 6 genes coding for tocopherol biosynthesis enzymes have been characterized (Dellapenna and Pogson, 2006). To identify tocopherol biosynthesis genes in *Brassica rapa*, BLAST search of all the annotated proteins in the brassica genome was performed at NCBI (annotated software version 6.1). In order to identify proteins similar to 4-hydroxyphenylpyruvate dioxygenase (HPPDase), homogentisic acid prenyltransferase (HPT), homogentisic acid geranylgeranyl transferase (HGGT), MPBQ methyltransferase (MPBQ MT), tocopherol cyclase (TC) and gamma tocopherol methyltransferase (gTMT) enzymes in *Brassica rapa*, these sequences from arabidopsis were used as query. Only those genes which have the required sequence similarity and conserved domain architecture were further considered as the candidate genes.

Domain analysis:

All retrieved non-redundant sequences were collected from TAIR and NCBI were subjected to domain analysis by using six different domain analysis programs: the Pfam 27.0 (http://pfam.sanger.ac.uk/), CDD

(http://www.ncbi.nlm.nih.gov/Structure/ cdd/ wrpsb.cgi), SMART (http://smart.embl-heidelberg.de/.), PROSITE profiles (http://prosite.expasy.org/), supfam (http://supfam.org/SUPERFAMILY/hmm.html.) and Gene3D (http://gene3d.biochem.ucl.ac.uk/.), with the default cut off parameters. Genes without specific domains were rejected. The protein sequences of A. thalian tocochromanol biosynthesis genes were used as a query to search against the brassica genome in NCBI (annotated software version 6.1). The same sequence database was searched repeatedly using the newly identified genes. In BLAST result, all the significant hits were taken and analyzed by alignment with known proteins using ClustalX.

Chromosomal localization and gene duplication

Positional information on the *Brassica rapa* genes was provided by the Brassica genome databases.

Prediction of gene structure

The gene structure (Exon-intron distribution) analyses of the tocopherol biosynthetic genes of *Brassica rapa* and Arabidopsis were carried out using the Gene Structure Display Server (GSDS) with default settings (http://gsds.cbi.pku.edu.cn/).

Subcellular localization

The presence of putative chloroplast transit peptides was analyzed by ChloroP1.1 program (http:// www.cbs.dtu.dk/services/ChloroP/)

Phylogenetic analysis of CNGC genes in Arabidopsis and rice

The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [Jones *et al*, 1992]. The tree with the highest log likelihood (-4647.1413) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 9 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 249 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [Tamura *et al*, 2013].

Results and Discussions

Tocopherol biosynthesis genes in *Brassica rapa* and Arabidopsis thaliana

Arabidopsis thaliana, the model dicot plant has a completely sequenced genomeand the information thus obtained is expected to be helpful for studies with other plants also. Tocochromanols are a group of amphiphilic lipids in plastids, the tocopherols and tocotrienols, which are synthesized exclusively by photosynthetic organisms.To study the absence and abundance of these compounds in different tissues, tocopherol biosynthesis genes in B. rapa were identified and studied in detail. In the current study, BLAST searches of the A. thaliana genome of tocopherols biosynthesis enzymes as query was use in NCBI to search for *Brassica rapa* tocopherol genes. Complete sequence search of the genome and domain analysis of the putative genes, revealed that

the tocopherol biosynthesis genes comprises of 9 potential members in *Brassica rapa*. All the relevant information about A.thalina and B. rapatocopherol biosynthesis genes arelisted inTable 1 and 2, respectively.

Structure of tocopherols biosynthesis genes

Comparisons between genomic DNA sequences withcorresponding cDNA sequences showed that coding sequences ofB. rapa tocopherol genes are interrupted by introns. Gene structural analysis showed that the *Brassica rapa* VTE (Fig. 1) had different gene structures as compared to Arabidopsis VTE (Fig. 2), with a variable number of introns and lengths. The numbers of introns in Arabidopsis VTE ranged from 1 to 12 introns, while *Brassica rapa* VTE rangedfrom 0 to12 introns. This suggests that all tocopherol genes with introns did not ascend from a retrotrans position event, and the different numbers may be caused by the insertion and loss of introns during their evolution [Cheng *et al*., 2003]

Ptotein Features

The nine *Brassica rapa* tocopherols proteins range in length from 3154 (Bra008507) to 1555 (Bra015511) amino acids (aa). ExPASy analysis showed that B. rapa tocopherols proteins vary greatly in isoelectric point (pI) values (ranging from 5.60to 9.59) and molecular weights (ranging from 29287.84to 54920.63kDa). According to their instability index values, not a single protein can be stable with an instability index of >40 (Table 2) [Guruprasad, 1990]. Protein features of A. thaliana tocopherol genes are given in table 3.

Subcellular localization

The subcellular localization of each Brassica tocopherol and AtVTE were predicted by PSORT analysis. Our results evidenced that except AtPDS1, Bra015511, and Bra032415 which localized in mitochondria all other were localized in chloroplast (Table 1 and 2).

The main function of tocopherol is to protect the lipid peroxidation of polyunsaturated fatty acids (PUFA) and chloroplastic membranes mostly consist of PUFA, therefor localization of most of tocopherol enzymes in chloroplast are reasonable. In the model species A. thaliana, genes encoding the key enzymes of the tocopherol biosynthesis pathway have been identified and functionally characterized and present in chloroplast (DellaPenna and Pogson 2006, Falk and Munne-Bosch 2010, Mene-Saffrane and DellaPenna 2010).

Post-translational modification

Translated proteins are often subjected to post–translational modifications (PTMs) to become functionally active. PTMs are the chemical modification of a protein after its translation, and have wide effects on broadening its range of functionality [Khoury *et al*., 2011]. Post-translational modifications of AtVTE are presented in Table 4.

Domain analysis

The proteins encoded by B. rapa tocopherol biosynthesis genes were thoroughly studied with respect to their domain structure (Fig. 2). Tocopherols are biosynthesized from two precursors, homogentisic acid (HGA) and phytyl diphosphate. The two precursors are condensed by HGA phytyl transferase,generating MPBQ. MPBQ is methylated to become2,3-dimethyl-6-phytyl-1,4-benzoquinol (DMPBQ).MPBQ and DMPBQ are converted by tocopherolcyclase to δ-tocopherol and g-tocopherol, respectively. The last step of the

tocopherol biosynthesis pathway ismethylation of δ-tocopherol and g-tocopherol by g-tocopherolmethyltransferase (g-TMT), yielding b-tocopheroland a-tocopherol, respectively [Bramley et al., 2000]. All the enzymes have a highly conserved catalytic domain according to their role in the biosynthesis pathway. After analyzing all the biosynthesis proteins from A. thaliana and B. rapa plant species, we noted that all homologous proteins have similar sequence and highly conserved catalytic domains, thus likely to have a common function (Table 3, Fig. 2).

Chromosomal distribution

To determine the chromosomal distribution of all the 9 tocopherol biosynthesis genes, the position for each gene were determined on the B.rapa chromosome pseudomolecules, as shown inFig. 3. The nine tocopherol biosynthesis genes were found to be distributed on only 5 of the 10B. rapa chromosomes (n=10, 2n=20). Bra022503 and Bra008507 are present on chromosome 2, Bra005809 and Bra037042 on chromosome 3, Bra038834 and Bra003524 on chromosome 7, Bra032415 on chromosome 9, and Bra015511 and Bra009449 on chromosome 10.

A maximum of 2 genes were present on a chromosome, hence the distribution of 9tocopherol biosynthesis genes did not reveal any evident clusters. In Arabidopsis maximum of two and manimum of one gene present on each chromosome. Chromosome 1 has two gene (AtVTE4 and AtPDS1), AtVTE2, AtVTE3, and AtVTE5 are localized on Chromosome 2, 3, and 5, respectively (Table 1)

Phylogenetic relationship between rice and Arabidopsis CNGC family genes

To determine the phylogenic relationship of tocopherol biosynthetic pathway genes between B. rapa and Arabidopsis, a maximum likelihood (ML) phylogenetic tree was constructed using full-lengthamino acid sequences. Threegroups were identified containing representative gene of both B. rapa andArabidopsis (Fig. 4). Each group contain 2 genes from Arabidopsis and 3 from B. rapa.

Group I comprises three members from B. rapa (Bra037.42, Bra008507, Bra022503) and 2 from Arabidopsis (AtVTE1, AtVTE4). Similarly, Group II contains three Brassica(Bra003524, Bra015511, and Bra032415) and 2 AtVTE (AtVTE3, AtPDS1). In the same way, Group III comprised of 3 Brassica and 2 arabidopsis tocopherol genes (Fig. 4). Phylogenetic data provide a framework for making more appropriate intergenomic comparisons, by determining whether chromosomal duplications within taxa pre-date or post-date divergence among taxa [Bowers, 2003].

In conclusion *Brassica rapa* tocopherol biosynthesis pathway comprise of nine genes, distributed on different chromosomes of this species. Domain analysis reveal all the catalytic domain responsible for catalyzing biosynthesis of tocopherols. Phylogenetic analysis show its relationship with the ancestors. This study will help in the experimental analysis of tocopherol genes in future.

Table.1 Features of Tocopherole biosynthetic pathway genes in A. thaliana

s.no	Gene name	Gene locus	Gene length	CDS length	ORF length	No. of exons	No. of introns	Position on genome	Chr	SCL
1	ATVTE1	AT4G32770	3069	1467		10	9	15804981-15807790	4	Plastid
2	ATVTE2	AT2G18950	3122	1182		13	12	8207491-8210047	2	Plastid
3	ATVTE3	AT3G63410	1531	1017		3	2	23415816-23417002	3	Plastid
4	ATVTE4	AT1G64970	2025	1047		4	3	24134337-24135993	1	Plastid
5	ATVTE5	AT5G04490	1910	915		6	5	1279867-1281587	5	Plastid
6	ATPDS1	AT1G06570	1661	1422		2	1	2012015-2013543	1	Mitoch-ondria

Table.2 Features of Tocopherol biosynthesis pathway genes in B. rapa

	Name	ID	PI	MW	Gene length	Protein Length	CDS	instability index (II)	Aliphatic index	Grand average of hydropathicity (GRAVY)	chr	strand	SCL
1	Bra037042	LOC103862197	6.00	54920.63	2866	490	1473	48.50	59.14	-0.482	A03	+	Chl
2	Bra038834	LOC103828352	9.59	43738.71	2994	394	1185	40.38	115.05	0.595	A07	-	Chl
3	Bra003524	LOC103830388	9.11	37757.55	1400	337	1014	44.27	82.76	-0.244	A07	-	Chl
4	Bra008507	LOC103853294	7.08	38262.86	3154	347	1044	60.03	88.85	-0.146	A02	+	Chl
5	Bra022503	LOC103852208	6.60	29287.84	1800	264	795	52.82	91.25	-0.077	A02	-	Chl
6	Bra005809	LOC103855573	9.39	33685.36	2114	308	927	39.03	108.57	0.473	A03	+	Chl
7	Bra009449	LOC103847284	9.56	32842.57	2158	299	900	37.28	108.23	0.475	A10	-	Chl
8	Bra015511	LOC103843547	5.81	48386.46	1555	443	1332	42.69	76.59	-0.244	A10	+	mito
9	Bra032415	LOC103844028	5.60	48787.97	1754	443	1332	40.92	78.98	-0.247	A09	-	mito

Table.3 Protein features of A. thaliana tocopherol genes

Locus ID	Gene name	Protein length	M. wt	pfam	CDD	PROSITE	SUP FAMILY	GENE 3D
AT4G32770	VTE1	488	54720.3	Tocopherol_ cycl	Tocopherol_cycl	-------	-------	Tocopherol_cycl
AT2G18950	VTE2	393	43908.4	UbiA	PT_UbiA	-------	-------	UbiA
AT3G63410	VTE3	338	37926	Methyltransf_11	AdoMet_MTases	-------	S-adenosyl-L-methionine-dependent methyltransferases 2.02e-03	MPBQ/MSBQ MT Pfam-B_2050
AT1G64970	VTE4	348	38075.1	Methyltransf_11	AdoMet_MTases PLN02244	SAM_MT95 SER_RICH	S-adenosyl-L-methionine-dependent methyltransferases 2.79e-56	Gamma-tocopherolmethyltransferase
AT5G04490	VTE5	304	33089.4	CTP_transf_1	CTP_transf_1 SEC59	-------	-------	CTP_transf_1
AT1G06570	PDS1	473	51920.2	Glyoxalase	HPPD_C Glo_EDI_BRP_like HPPD_N_like PLN02875	-------	Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyldioxygenase 3.24e-96	4-hydroxyphenylpyruvate dioxygenase domains(2)

Table.4 Post transcriptional modification of A. thaliana tocopherol

Locus ID	aa	gDNA	CDS	N-Glyc	N-Myr	PKC	Casein	Tyr	cAMP-cGMP	Amidation
AT4G32770	VTE1	3069	1467	4	16	14			2	2
AT2G18950	VTE2	3122	1182	1	6	8	1	1		1
AT3G63410	VTE3	1531	1017	2	12	6	4			
AT1G64970	VTE4	2025	1047	3	4	5	5			
AT5G04490	VTE5	1910	915	2	4	4	1			
AT1G06570	PDS1	1661	1422	1	6	7	8	1	4	

Fig.1 Gene structure of B. rapa Tocopherol genes

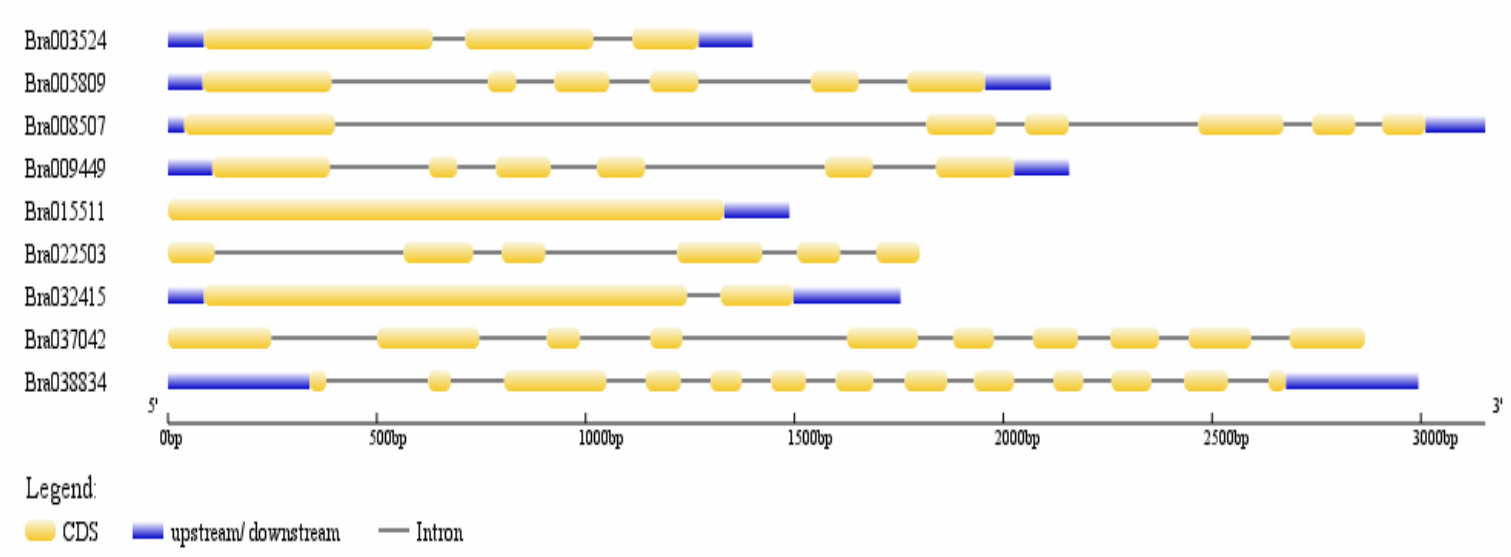


Fig.2 Domain analysis of B. rapa tocopherol genes

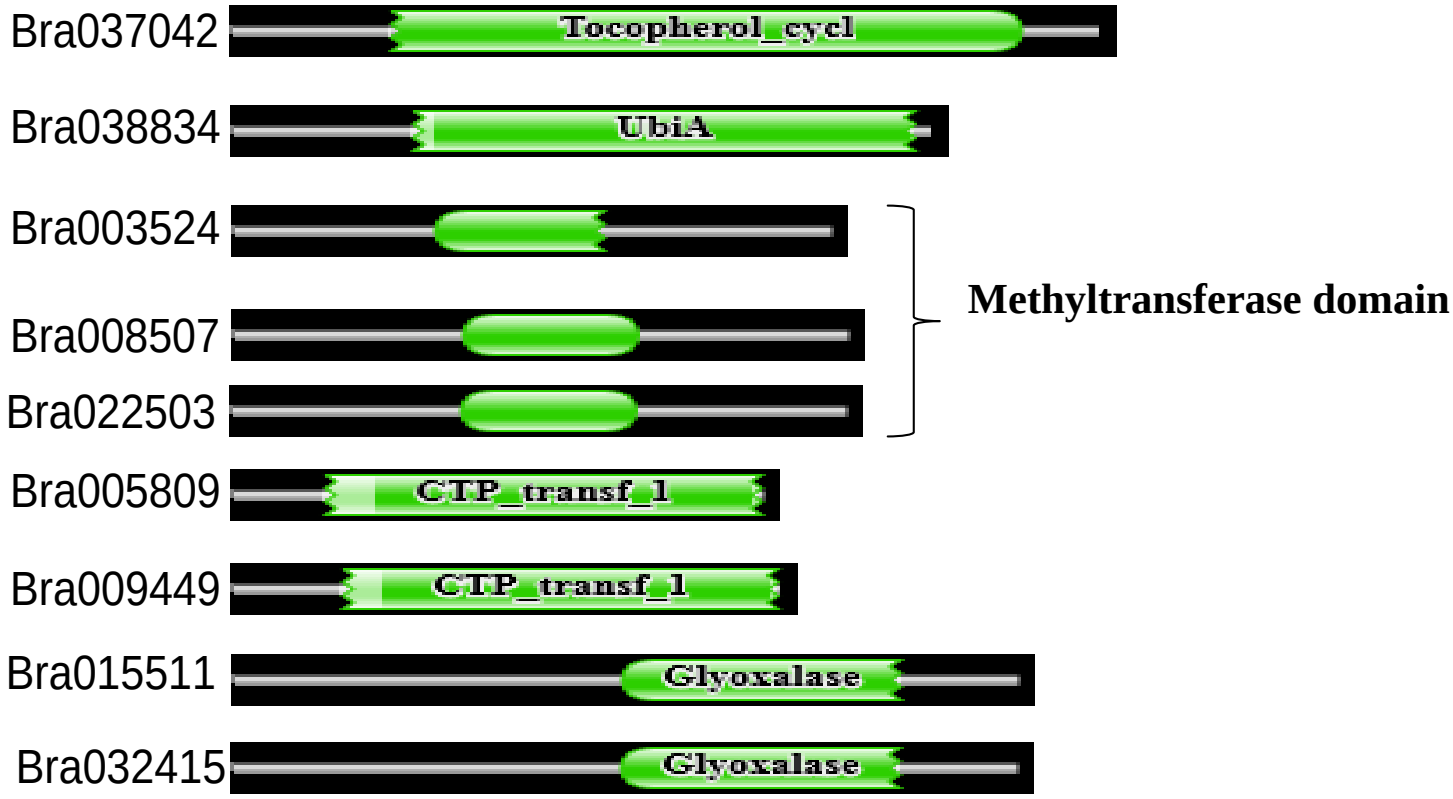


Fig.3 Chromosomal distribution of B. rapa tocopherol genes

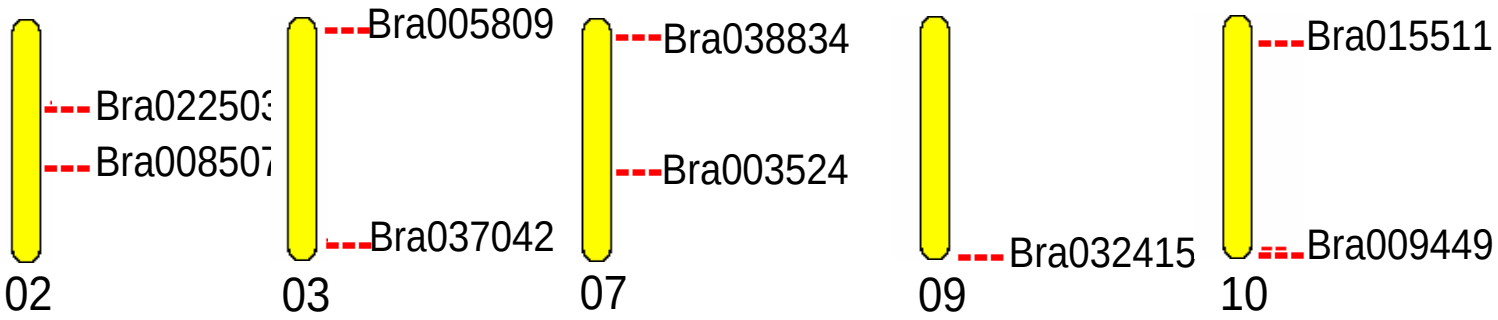
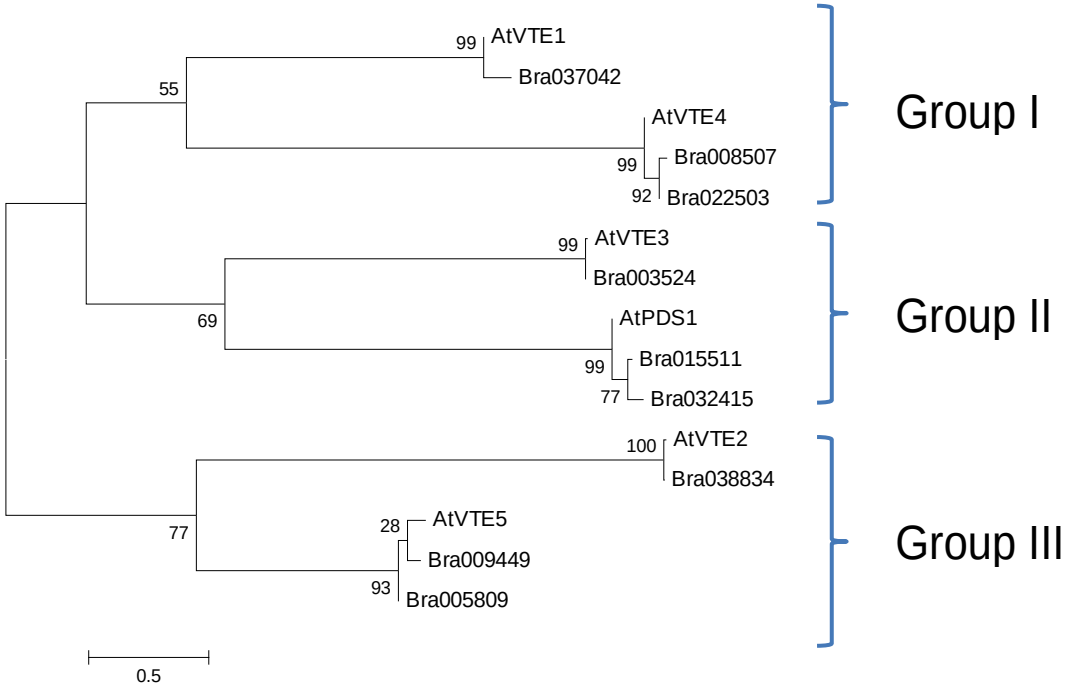


Fig.4 Phylogenetic analysis of B.rapa tocopherol genes



Acknowledgements

The work of our lab was sponsored by the National Key Basic Research Project (abbreviated as 973 project, Code No. 2015CB150205) and Natural Science Foundation of China (Grant nos. 31371542).

References

Bowers, J.E., Chapman, B.A., Rong, J., Paterson, A.H. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature.,* 422:433–438

CDD: a Conserved Domain Database. http://www.ncbi.nlm.nih.gov/ Structure/cdd/wrpsb.cgi.

Cheng, Z., Sattler, S., Maeda, H., Sakuragi, Y., Bryant, D.A., Dellapenna, D. 2003. Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. Plant Cell., 15 : 2343–2356.

Dellapenna, D., Pogson, B.J. 2006. Vitamin synthesis in plants: tocopherols and carotenoids. *Annu. Rev. Plant Biol*., 57:711–738.

Gene3D.http://gene3d.biochem.ucl.ac.uk/

GSDS: Gene Structure Display Server.http://gsds.cbi.pku.edu.cn/.

Guruprasad, K., Reddy B.B., Pandit, M.W. 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng*., 4:155–161

Hunter, S.C., Cahoon, E.B., 2007. Enhancing tocochromanol in oilseeds: unraveling tocopherol and tocotrienol biosynthesis. *Lipids.,* 42: 97–108

Jones D.T., Taylor W.R., and Thornton J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences.*, 8: 275-282.

Khoury, G.A., Baliban, R.C., Floudas, C.A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep 2011,1:1–5.

Prosite profiles.http://prosite.expasy.org/.

Smart: a Simple Modular Architecture Research Tool. http://smart.embl-heidelberg.de/.SUPER FAMILY. http://supfam.org /SUPER FAMILY/ hmm.html.

Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution.,* 30: 2725-2729.

The Pfam database. http://pfam. sanger. ac.uk/.