Indian J. Genet. Plant Breed., (2024); 84(4): 551-560
https://doi.org/10.31742/ISGPB.84.4.5
ISSN: 0975-6906
www.isgpb.org

RESEARCH ARTICLE

Development of functional markers for high α-tocopherol content in Indian mustard [*Brassica juncea* (L.) Czern. and Coss] employing genetic and molecular analysis of *VTE4* gene

Vijay Kamal Meena[$], Yashpal Taak, Rajat Chaudhary, Subhash Chand[1], Manoj Kumar Patel, Ranjit Saroj, Vignesh Muthusamy, Joghee Nanjundan[2], Navinder Saini, Sujata Vasudev and Devendra Kumar Yadava*

Abstract

Tocopherol (vitamin E) is considered an important vitamin carrying antioxidant properties. It plays a vital role in maintaining the quality and stability of oil in *Brassica* species. Molecular mechanisms of tocopherol content have been studied in Brassicas; however, it is untapped in Indian mustard (*Brassica juncea*). In the experiment, the expression profile and sequence variation of the candidate gene *VTE4* controlling α-tocopherol content (ATC) were studied between two diverse parents (RLC-3 and NPJ-203) of *B. juncea*. The *VTE4* gene expression in different tissues was almost double in NPJ-203 (high ATC genotype) as compared to RLC-3 (low ATC genotype). Moreover, sequence analysis of *VTE4* in NPJ-203 and RLC-3 revealed the presence of two SNPs in the 6th exon, resulting in a shorter coding sequence (CDS) in RLC-3 (996 bp) as compared to NPJ-203 (1044 bp). Using these SNPs, an allele-specific marker was developed and validated in the F_3 population. The single marker analysis revealed that the marker was significantly linked to the tocopherol content, contributing 16.46% to the total phenotypic variance. Thus, the study suggested that *VTE4* is the major gene contributing to the tocopherol content, and the developed marker can be effectively used in marker-assisted breeding to improve tocopherol content in *B. juncea*.

Keywords: *VTE4* gene, Indian mustard, α-tocopherol content, AS-PCR, Marker-assisted selection

Introduction

Indian mustard [*Brassica juncea* (L.) Czern & Coss] is an important oilseed crop that is widely cultivated in many countries of the world, including India, China, and Canada (Chand et al. 2021; Yadava et al. 2022). The most prevalent forms of vitamin E in Indian mustard oil are α- and γ-tocopherol, whereas a trivial amount of δ-tocopherol is also reported (Gupta et al. 2015). Amongst the four isoforms (α, β, γ and δ), a-tocopherol has the utmost activity in humans for various purposes (Rizvi et al. 2014). For instance, it acts as an antioxidant that protects the plant from the damage caused by harmful free radical molecules. In addition, a-tocopherol is considered the most effective antioxidant due to its high free radical scavenging activities (Rizvi et al. 2014). Tocopherols and tocotrienols, a class of amphipathic molecules made up of a polar chromanol head group derived from the shikimate pathway and a polyprenyl lipophilic side chain derived from the methylerythritol phosphate pathway and chlorophyll degradation, are the only forms of vitamin E that can accumulate in photosynthetic organisms (Wang et al. 2012). Four variants of tocopherols and tocotrienols differ from each other in the

Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India.
[1]ICAR-Indian Grassland and Fodder Research Institute, Jhansi 284 003, Uttar Pradesh, India.
[2]ICAR-Indian Agricultural Research Institute, Regional Station, Wellington 643 231, Tamil Nadu, India.
[$]Present address: Agricultural Research Sub-station, Sumerpur, Agriculture University Jodhpur 306 902, Rajasthan, India.
***Corresponding Author:** Devendra Kumar Yadava, Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India, E-Mail: dkygenet@gmail.com

How to cite this article: Meena V.K., Taak Y., Chaudhary R., Chand S., Patel M.K., Saroj R., Muthusamy V., Nanjundan J., Saini N., Vasudev S. and Yadava D.K. 2024. Development of functional markers for high α-tocopherol content in Indian mustard (*Brassica juncea* (L.) Czern. and Coss) employing genetic and molecular analysis of *VTE4* gene. Indian J. Genet. Plant Breed., **84**(4): 551-560.

Source of support: ICAR-Consortium Research Platform on Biofortification

Conflict of interest: None.

Received: June 2024 **Revised:** Oct. 2024 **Accepted:** Nov. 2024

number and location of methyl groups on the chromanol ring (Della Penna 2005; Valentin et al. 2006). Indian mustard has a sufficient quantity of γ-tocopherol (GTC), whereas it contains a low amount of α-tocopherol content (ATC). The ATC plays a vital role in maintaining the quality and stability of the oil by protecting the polyunsaturated fatty acids present in the seeds, as these fatty acids are more susceptible to cause damage from free radicals (Rizvi et al. 2014; Meena et al. 2022; Grygier 2023).

B. juncea genotypes are rich in GTC but poor in ATC (Yusuf and Sarin 2007). ATC is the most active form of vitamin E and is vital for the human daily diet (DellaPenna and Pogson 2006). Among the α-forms, stereoisomers RRR-α-tocopherols have the highest biological activity and can be stored and transported in the body due to specific selection by the hepatic α-tocopherol transfer protein (α-TTP) (Brigelius-Flohé and Traber 1999). Hence to ensure the high availability of ATC, the higher yielder Indian mustard cultivars with high ATC are required, which can be accelerated through a marker-assisted approach. The ATC in *B. juncea* is influenced by genetic factors, and functional markers have not been reported for high ATC (Meena et al. 2022).

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation, and these have gained significant attention as useful genetic markers for identifying functional variation in a wide range of species, including humans, plants and animals (Du et al. 2022; Meena et al. 2023; Najafi et al. 2024). In recent years, SNP allele-specific PCR (AS-PCR) has emerged as a powerful tool for the detection of SNPs and their alleles in various applications (Liu et al. 2012). In the modern era, the cost of genotyping has decreased significantly and high-throughput and cost-effective methods for SNP detection, such as SNP AS-PCR, are available (Liu et al. 2012; Patel et al. 2022). AS-PCR has been used by many researchers in various crop breeding programs for the development of SNP-based molecular markers (Liu et al. 2012).

In *B. juncea*, to the best of our knowledge, no molecular study on tocopherol content has been reported. Although, in different Brassicas, it has been reported that gene *VTE4*, is the major gene governing the tocopherol content (Endrigkeit et al. 2009; Yusuf et al. 2010). In the present study, *VTE4* gene was targeted in *B. juncea* genotypes. Therefore, with this background, in the present investigation, *VTE4* gene was cloned and sequenced on two diverse parents (RLC-3 and NPJ-203) for tocopherol content and using sequence information allele-specific markers were developed and validated on a segregating population of *B. juncea*. In addition, these parents were subjected to expression studies in different tissues and developmental stages.

Materials and methods

Selection of suitable genotypes and development of mapping population

In the present study, a diverse panel of 96 genotypes of *B. juncea* collected from different sources was raised at the experimental farm (latitude 28°380 N; altitude 77°090 E and 228.61 m amsl) of the Division of Genetics, ICAR-Indian Agricultural Research Institute(IARI), New Delhi, India (Supplementary Table S1). In 2017–18, the germplasm panel was evaluated for the tocopherol content in the mature seed and the contrasting genotypes (RLC-3 and NPJ-203) were selected (Table 1) and a mapping population was developed using these parents. Selected parents, *viz*., RLC-3 (low ATC) and NPJ-203 (high ATC) were crossed at the research farm, ICAR-IARI during 2019–20, and F_1s seeds were harvested. Further, F_2 population, derived by selfing of F_1 plants, was developed at ICAR-IARI Regional Station, Wellington, Tamil Nadu (off-season nursery; 11.37°N, 76.8°E; 1855 m amsl) in 2020. Later, $F_{2:3}$population was developed by selfing of F_2 population at the experimental farm, ICAR-IARI, during 2020–2021 and $F_{2:3}$ population was subsequently used for marker validation.

Experimental layout and crop maintenance

The seeds of F_2 population were sown in 9 rows of 5 m length each and the plants were selfed. The row-to-row and plant-to-plant spacing were kept at 45 cm and 15 cm, respectively. All the recommended agricultural practices, from sowing to harvesting, were followed to raise a healthy crop and good plant stand. Further, 211 plants were selected in F_2 population and $F_{2:3}$ seeds harvested from the individual plant were used for the estimation of ATC and GTC.

Estimation of tocopherol contents

Tocopherol traits viz., ATC and GTC contentfrom the harvested seeds were estimated using high-performance liquid chromatography (HPLC) (Fritsche et al. 2012). For tocopherols estimation, 50 mg freshly harvested seeds were ground in 1500 µL n-heptane and extracted tocopherol solution was incubated at -20°C for 2 hours, and 20 µL was used for HPLC analysis. A silica gel column (5 µM LiChrospher® Si 60, Merck, Darmstadt, Germany) was used to extract tocopherols employing a mobile phase consisting of an n-heptane/isopropanol mixture (99 + 1; v + v). Further, the fluorescence detection (excitation at λ = 290 nm and emission at λ = 328 nm) method was used for tocopherol quantification. The retention times were compared with Merck's tocopherol kit (Merck, Darmstadt, Germany) to identify specific tocopherol forms (GTC: 8.9 minutes and ATC: 10.6 minutes), and the concentration of single forms was correlated with signal output to calibrate each tocopherol form. Moreover, a linear calibration range was obtained for

Table 1. Details of the parents used in the estimation of tocopherol content trait

S. N.	Genotypes	Pedigree/ Source	Parent institute	Tocopherol content traits (mg/kg)		
				ATC	GTC	AGR (%)
1	RLC-3	JM06003×JM06020	PAU, Ludhiana, India	0.74	51.17	1
2	NPJ-203	(EJ9913×SEJ8) × Laxmi	IARI, New Delhi, India	3.49	53.14	7

ATC: α- tocopherol content; GTC: γ- tocopherol content; AGR: α-/γ- tocopherol ratio

the concentration of the analyzed samples in this study. The tocopherol composition was expressed as the ratio of ATC and GTC and called AGR.

DNA and RNA isolation

Newly formed young leaves from both the parents (RLC-3 and NPJ-203) and 211 F_2 plants were selected and the standard CTAB method was used for DNA extraction (Doyle and Doyle 1987). Further, nanodrop™ (Nanodrop Technologies, USA) was used for DNA quantification and quality was assessed by electrophoresis. Then, DNA was diluted to a final concentration of 20 ng/µL with nuclease-free water and stored at -20°C. Moreover, different plant parts *viz*., leaf, flower bud and the immature seed of RLC-3 and NPJ-203 were used to isolate total RNA using TRIzol reagent (MRC, USA). Further, RNase-free DNase (Qiagen, USA) was used to eliminate residual DNA contamination in isolated RNA samples. RNA quality and quantity were determined by spectrophotometer (ND-1000 Nanodrop Technologies, USA) and RNA samples having OD value > 2.0 at 260/230 nm wavelength were used for further analysis. RNA integrity was confirmed by resolving the samples on the agarose gel (1.5% in 1X TAE) through electrophoresis. Further, the first strand of cDNA was synthesized from 2 µg of total RNA using a Verso cDNA Synthesis Kit (Thermo Scientific Cat. No. AB1453A).

In-silico analysis and cloning of tocopherol genes

In different Brassicas, previous reports suggest that gene *VTE4* is the major candidate gene governing the tocopherol content (Endrigkeit et al. 2009; Yusuf et al. 2010). Therefore, coding sequence (CDS) of *VTE4* gene was extracted from NCBI (https://www.ncbi.nlm.nih.gov/) and Brassicaceae Database (http://brassicadb.cn/) using orthologues*VTE4* genes as a reference. Further, *VTE4* gene primers were designed using PrimerQuest™ Tool (https://eu.idtdna.com/pages/tools/primerquest) and amplified in RLC-3 and NPJ-203 (Table 2A). The gene amplicon products of parents were gel eluted and ligated into pGEM-T easy vector (Promega Biotech India Pvt. Ltd.) and sequenced by Barcode Biosciences, Bangalore, India. Later, the gene sequence analysis for RLC-3 and NPJ-203 was done using multiple sequence alignment tools by NCBI blast and CLUSTALw program (https://www.ebi.ac.uk/Tools/msa/clustalo/). Eventually, based on the sequencing results between RLC-3 and NPJ-203, allele-specific primers were developed and validated in F_2 population.

Expression analysis of VTE4 gene

The qRT-PCR reactions were performed using SYBR® green detection chemistry in a real-time PCR machine (Applied Biosystems, USA). KAPA SYBR® FAST qPCR Master Mix (2X) Kit (KAPA Biosystems, USA) was used to complete qRT-PCR. In this process, 10 µL reaction mixture was constituted by adding 2 µL diluted cDNA, 5 µL qPCR master mix (2X), 0.4 µL of ROX high reference dye (50X), 0.4 µL of 10 µM forward and reverse primers each and 1.8 µL double distilled water. In the PCR, an initial denaturation was kept for 3 min at 95°C followed by 40 cycles having three steps such as 95°C for 10 seconds, 60°C for 30 seconds, and 72°C for 30 seconds. Then, melt curve analysis of PCR products was carried out by a constant increase of temperature between 60 and 95°C. Glyceraldehyde-3-phosphate-dehydrogenase(GAPDH) gene transcript was used in qRT-PCR normalization. Every time, three biological replicates and three technical replicates were used for qRT-PCR experiments. Primers used for the expression analysis are presented in Table 2B.

Development of PCR based allele specific primers

Allele-specific primers were developed using the method explained by Liu et al. (2012)**.** In this process, two forward primers were designed with single nucleotide differences (T and A) at 3′ end and an additional mismatch (G) was incorporated into the primer sequence at the third nucleotide from 3′ end to increase specificity. Besides, an additional SNP (C/T) was present at 7th base from 3′ end in both primers, which has further enhanced its specificity. Allele-specific markers are dominant in nature; therefore to make them co-dominant markers, two forward primers (TocoASP1 and TocoASP2) and one reverse primer (TocoRev) were designed where one primer set (TocoASP1 & TocoRev) presented positive band of 220 bp only in NPJ203, whereas second set (TocoASP2 &TocoRev) gave a positive band of 220 bp only in RLC3. Therefore, the reaction was operated twice in separation to differentiate between homozygotes and heterozygotes. Primer sequence and specific PCR conditions are presented in Table 2C.

Segregation and single marker analysis in F_2 population

Marker segregation was tested using χ^2–test given by Karl Pearson (Pearson 1916). χ^2 calculated value was compared

Table 2. Details of the primers used in the study

A. Primers used for cloning of *VTE4* gene			
Gene	Primer sequence (5'-3')	F/R	Amplicon size (bp)
VTE4.1	TGTGAGACTGTTCCATCAACAT	F	1624
	CACCCAACATCCACCACTATC	R	
VTE4.2	CGGTTCCAATACTCACACACA	F	1457
	GAATGGTCGAATGTACGAGAAGA	R	
VTE4.3	ATGCGCATTCGGTGAAA	F	1382
	CCACATTTCTCACGTGACTCT	R	
B. Primers used in the study for qRT-PCR analysis			
Gene	Primer sequence (5'-3')	F/R	Amplicon size (bp)
GAPDH	CTAACTGCCTTGCTCCACTT	F	101
	TGTCTTCTGAGTTGCAGTGATAG	R	
VTE4.RT	GCGCTGCGAGAAGGAATAGCG	F	289
	AGTGTGATGCCAATGCATTCGGC	R	

C. Primers used for the development of PCR-based allele-specific markers for a-tocopherol content

Allele-specific Primer	Primer sequence (5'-3')	Annealing T_m and time	Extension T_m and time	Amplicon size (bp)
TocoASP1	TGGAAGGGCCTCGTGGCT	58°C, 30 seconds	72°C, 25s	220
TocoRev	AGATTAAACTTAGAGAGGCTTCTGG			
TocoASP2	TGGAAGGGCCTTGTGGCA	58°C, 30 seconds	72°C, 25s	220
TocoRev	AGATTAAACTTAGAGAGGCTTCTGG			

with χ^2 tabulated value at 5% level of significance with (n-1) degrees of freedom. If χ^2 calculated value is significantly deviating from χ^2 tabulated value, then it can be concluded that markers are not following Mendelian segregation and *vice versa*.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where Oi = Observed frequency of i^{th} individual plant, Ei = Expected frequency of i^{th} individual plant. Single marker analysis (SMA) was carried out using single factor analysis of variance (SF-ANOVA) and simple regression analysis to calculate linkage analysis.

Results

In-silico analysis of VTE4 gene

The CDS sequence of *VTE4* gene from *Arabidopsis thaliana* (AT1G64970) and different *Brassica* species *viz*., *B. rapa* (Bra008507), *B. nigra* (BniB032508), *B. oleracea* (BolC2t09793H), *B. carinata, B. napus* (C02p37750), *B. juncea* (DQ864978) retrieved from NCBI database and compared using multiple sequence alignment tool CLUSTALw program (https://www.ebi.ac.uk/Tools/msa/clustalo/). Our investigation revealed the presence of two paralogs of the *VTE4* gene located on the A02 chromosome of *B. juncea*. The CDS of the first homolog was 1044 base pairs in length, encoding a protein of 347 amino acids. In contrast, the CDS of the second homolog was 996 base pairs long, resulting in a protein consisting of 331 amino acids. While examining *VTE4* gene orthologs across *Brassica* species, it was observed that *B. oleracea* exhibited the highest number of exons, totaling 7, with a CDS sequence measuring 1086 base pairs in length. In contrast, other *Brassica* species typically possessed six exons and five introns in their *VTE4* gene sequences. The results of *in-silico* analysis for *VTE4* gene in *A. thaliana* and different *Brassica* species are presented in Table 3. The *VTE4* gene was found to be located on A–genome (*A. thaliana, B. rapa* and *B. juncea*), B–genome (*B. nigra*), C-genome (*B. oleracea* and *B. carinata*), and A– and C–genome (*B. napus*).

The *VTE4* (γ- tocopherol methyl transferase) gene length varied from 2312 bp (*B. juncea*) to 2903 bp (*B. oleracea*) in *Brassica* species. Likely, the CDS length also varied from 996 bp (*B. juncea*) to 1086 bp (*B. oleracea*). Besides, it had six exons and five introns in most of the *Brassica* species except *B. oleracea* and *A. thaliana* (Table 3). In addition, the shortest and longest amino acid chains were found in *B. juncea* (331 amino acids) and *B. oleracea* (361 amino acids), respectively. The phylogenetic relationship among different *Brassica* species was studied using their CDS sequences (Fig. 1). It revealed that *B. juncea* and *B. rapa* are phylogenetically close

Table 3. *In-silico* analysis of *VTE4* or *γ-TMT* (γ-tocopherol methyltransferase) gene

Gene ID	Species	Gene length (bp)	CDS length (bp)	No. of exons	No. of introns	No. of amino acids	Chromosome
AT1G64970	*A. thaliana*	2412	1047	4	3	348	A1
Bra008507	*B. rapa*	2354	1044	6	5	347	A2
BniB032508	*B. nigra*	2604	1077	6	5	358	B2
BolC2t09793H	*B. oleracea*	2903	1086	7	6	361	C2
Predicted	*B. carinata*	2501	1044	6	5	347	C3
C02p37750	*B. napus*	2415	1047	6	5	348	A2 and C2
DQ864978	*B. juncea*	2312	996	6	5	331	A2

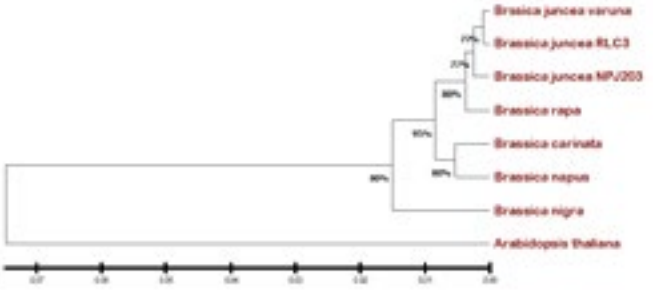


Fig. 1. Neighbor-joining phylogenetic tree of *VTE4* gene [Comparison of the deduced CDS sequences of *VTE4* from related species of *Brassica*. An unrooted tree based on CDS sequence similarity was obtained by using Mega 4.0 software. Bootstrapping was performed with 1,000 replicates]

to each other and share a recent common ancestor with each other than the other related species. Likewise, *B. carinata* and *B. napus* have more similarities, whereas *A. thaliana* and *B. nigra* are more diverse than the other studied groups.

Gene cloning and sequence analysis

The full-length *VTE4* gene was cloned and sequenced from RLC-3 and NPJ-203 genotypes (Fig. 2). The total gene length was 2323 and 2321bp in RLC-3 and NPJ-203, respectively and it had 66 bp long promoter, 6 exons and 5 introns (Fig. 3). After gene sequence alignment of both the genotypes, it was found that there was no variation in the promoter region, introns and exons except exon-6 where 2 SNPs were found. These SNPs led to shorter CDS in RLC-3 (996 bp) and longer CDS in NPJ-203 (1044 bp) genotypes. All sequences related to the *VTE4* gene (Supplementary Table S2), including the amino acid sequences (Supplementary Table S3) and the figure comparing these sequences (Supplementary Fig. 1), have been provided in the supplementary file.

Expression analysis of VTE4 gene

The relative transcript levels of *VTE4* gene in different plant tissues viz., leaf, flower bud and the immature seed of parents (NPJ-203 and RLC-3) were estimated employing gene-specific primers in real-time qRT-PCR. The results revealed that *VTE4* gene expression showed a significant (p <0.05) difference among studied plant tissues of both parents. For instance, the *VTE4* gene expression level was

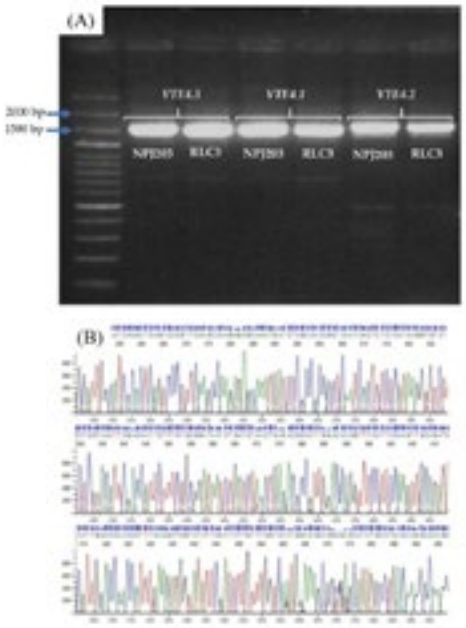


Fig. 2. (A) Gel picture illustration of amplified product of *VTE4* gene from NPJ203 and RLC3. [To amplify the whole gene, three pairs of primers were used, i.e., *VTE4.1, VTE4.2* and *VTE4.3*] (B) Chromatogram illustrating the sequencing quality of complete *VTE4* gene sequence

Fig. 3. Diagrammatic Representation of *VTE4* Gene Variations in NPJ203 and RLC-3. In the *VTE4* gene of RLC-3, a shorter Exon 6 spanning 54 bp is indicated by the Yellow Box, while the Black Line represents a longer intronic region. The Green Arrow highlights the shared Promoter Region, and the common exonic region in both genotypes is denoted by Purple Boxes, with the intronic region indicated by a Red Line

2.4-fold in leaves, 1.9-fold in flower buds and 2.2-fold in immature seeds of NPJ-203 than RLC-3 (Fig. 4). Besides, the *VTE4* gene expression was found more in leaves than the other plant tissues.

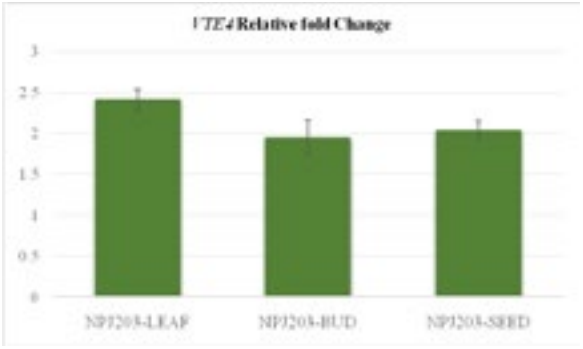


Fig. 4. The expression analysis of VTE4 gene in Leaves, Flower bud and Immature seeds of NPJ203 (high α-tocopherol) and RLC3 (low a-tocopherol). Total RNA was assayed for the VTE4 gene expression by qRT-PCR using GAPDH as a normalizer. Foldchange values represent mean ± SE and comparison of means was carried out by Student's t-test ($p < 0.05$). [In this figure, RLC-3 expression is considered as the baseline (set to 1), and the fold change indicates how much higher the expression of NPJ-203 is relative to this baseline

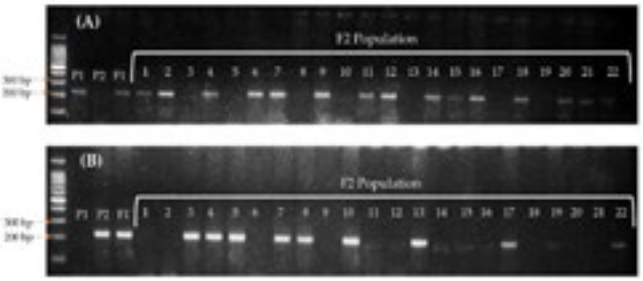


Fig. 5. Gel picture illustration of Primer amplification in Parents, F_1 and F_2 population (A) Amplification pattern of Primers (TocoASP1 + TocRev) in high (P1-NPJ203) ATC, low ATC (P2-RLC3) genotypes, F_1 (RLC3 × NPJ203) and F_2 Population (B) Amplification pattern of Primers (TocoASP2 + TocRev) in high ATC (P1 is NPJ203), low ATC (P2 is RLC3) genotypes, F_1 (RLC3 × NPJ203) and F_2 population

SNP detection and PCR based allele specific markers development

Sequence comparison between NPJ-203 and RLC-3 parents led to the identification of two SNPs at 2125 (C/T) and 2131 bp (T/A) positions. These identified SNPs were used to develop PCR-based allele-specific primers for ATC. Specific combination of forward and reverse primers were used to amplify the target gene, for instance, TocoASP1 + TocRev amplified in high ATC genotype (NPJ203), whereas TocoASP2 + TocRev amplified in low ATC genotype (RLC3) (Fig. 5). Further, F_2 population, derived from cross between RLC-3 and NPJ-203, were used to study the marker segregation pattern using chi-square test. The chi-square calculated value (3.92) was found to be lower than the chi-square tabulated value (5.99), suggesting that the marker was following a monogenic Mendelian segregation pattern (1:2:1) in F_2 population (Fig. 6). Further, the phenotypic variance (R^2) was calculated for these markers to understand the extent of trait contribution and it showed high R^2 value with magnitude of 16.46% (Table 4).

Discussion

Tocopherols, a group of lipid-soluble compounds known for their vitamin E activity, assume a pivotal role in both the stability of *B. juncea* oil and human health (Yusuf and Sarin 2007; Yusuf et al. 2010). *B. juncea*, a plant species recognized for its oil-rich seeds, relies on tocopherols as potent antioxidants to safeguard its oil from oxidative degradation, thereby extending its shelf life. The inherent susceptibility of unsaturated fatty acids in *B. juncea* oil to oxidation necessitates the presence of tocopherols as essential constituents to prevent the development of off-flavors and rancidity (Dua et al. 2014). The biosynthesis of tocopherols in *B. juncea* involves a complex enzymatic pathway situated within the plastids of plant cells (Yusuf et al. 2010). Among the various tocopherol isomers synthesized, α-tocopherol assumes particular importance. Recognized as the biologically most active form of vitamin E, α-tocopherol exhibits the highest antioxidant capacity. Its ability to efficiently scavenge free radicals is a key factor in protecting lipids and cellular components from oxidative harm (Szewczyk et al. 2021). The α-tocopherol significance underscores the importance of optimizing tocopherol biosynthesis in *B. juncea* plants to enhance both the stability of oil and its nutritional value for human health(Yusuf et al. 2010).

In the intricate biosynthesis pathway of tocopherols, *VTE4* gene, which catalyzes γ-tocopherol methyltransferase enzyme, emerges as a central player. This gene is responsible for the conversion of γ-tocopherol to α-tocopherol, a transformation that significantly augments the antioxidant potential of the compound (Liao et al. 2018). The enzymatic activity of γ-tocopherol methyltransferase is paramount as it dictates the ultimate tocopherol composition in *B. juncea* oil (Yusuf et al. 2010). Understanding and manipulating the role of this gene can lead to improved oil quality and enhanced health benefits, establishing it as a key target for future studies in this field.

Table 4. ANOVA for single marker analysis for marker-trait association in the F2 population developed by the cross RLC3 × NPJ203

ANOVA	df	SS	MSS	F	p-value	Multiple R	R^2	Adjusted R^2
Regression	1	16.98	16.98	42.38	5.47E-10	0.41	16.86%	16.46%
Residual	209	83.72	0.40					
Total	210	100.69						

Fig. 6. Graphical representation of Marker segregation in F_2

Gene cloning or DNA cloning refers to amplifying and isolating a specific DNA sequence(typically a particular gene) from the rest of a cell DNA. In our findings, parent RLC–3 had two additional base pair than NPJ–203 in total gene length but NPJ–203 had high CDS (1044 bp) length than RLC3 (996 bp). Earlier, *VTE4* gene was cloned in *B. juncea* var. Varuna (a low ATC variety) (Yusuf et al. 2010) and had a similar CDS length like RLC3 genotype. Likewise, γ-tocopherol methyltransferase gene from *B. oleracea* was cloned and reported with 1265 bp CDS and 1044 bp CDS that encodes a 347 amino acid long protein (Ouyang et al. 2003). *VTE4* is the major gene that regulates high ATC in plants. For instance, ATC was absent in leaves of *VTE4*. *A. thaliana* mutants, but high GTC was accumulated (Bergmüller et al. 2003). In addition, overexpression of *VTE4* gene in leaves alters the quantity and composition of tocopherol and increases ATC with upto 30% rise in total tocopherol content (Li et al. 2010). Likewise, *B. napus VTE4* homolog had increased ATC by 50-fold in transgenic *Arabidopsis* seeds through marker-assisted selection (Endrigkeit et al. 2009).

Genbank sequencing database has >810000EST records for U-triangle species (www.brassica.info). Full-length sequences of *VTE4* gene for *B. oleracea* (AF381248), *B. napus* (DQ508019) and *B. juncea* (DQ864978) are available and exhibited 96% similarity to the *VTE4* sequence among them. The genic construct of *VTE4* gene was conserved in both the genotypes except for 2 SNP differences in the exon-6. The *VTE4* gene affects the accumulation of ATC in various plant tissues and is also affected by temporal stages. In the study, *VTE4* gene expression was more in high ATC parent (NPJ-203) tissues viz., leaves, flower bud and immature seeds than RLC-3 genotype. Besides, *VTE4* gene expression also varied from tissue to tissue and developmental stages. Vinutha et al. (2015) observed comparable outcomes, revealing that elevated expression of the γ-TMT3 gene corresponds with heightened α-tocopherol concentrations in seeds of the Bragg soybean genotype in contrast to the DS-2706 genotype. This unveils the linkage between gene expression and α-tocopherol buildup in soybean seeds. The differential expression of a gene might be due to UTR regions or varied activity of the gamma-tocopherol methyl transferase enzyme (Das and Bansal 2019). However, an enzymatic assay for the gamma-tocopherol methyl transferase enzyme is needed to understand the wide-ranging expression of the gene.

The present findings revealed that the *VTE4* gene is responsible for the conversion of GTC to ATC. Previously, the loss of function of the *VTE4* gene was responsible for replacing ATC with GTC in safflower (García-Moreno et al. 2011) and sunflower (Hass et al. 2006). Also, overexpression of *VTE4* promoter in soybean (Dwiyanti et al. 2011), additional insertion of *VTE4* gene copies in sunflower (García-Moreno et al. 2012), and overexpression of *VTE4* in other plant species viz., *Arabidopsis*, soybean, and Indian mustard (Shintani and Dellapenna 1998; Endrigkeit et al. 2009) has increased ATC and reduced GTC.

Genetic manipulations are vital to improve oil productivity and quality in Indian mustard. It is possible only when modern breeding tools such as marker-assisted breeding, are used to supplement the conventional breeding approaches. Molecular markers were used comparatively less in oilseeds than the cereals. However, many high-density linkage maps and molecular marker systems were developed in *Brassica* species (Yadava et al. 2012). In the past, though a few reports have been published on total tocopherol content (Gupta et al. 2015) but, there is no evidence of molecular marker(s) and/or QTL(s) specifically for ATC in *B. juncea*. In other crops,14 QTL were reported for tocopherol content and composition in *Arabidposis* seeds (Gilliland et al. 2006), and two QTLs for ATC on the A2 chromosome in *B. napus* (Wang et al. 2012). Likewise, 26 and 12 SNPs within two genes (*BnaX.VTE3.a*, *BnaA.PDS1.c*) in *B. napus* were associated with tocopherol traits and explained up to 16.93% of the genetic variance for tocopherol composition and up to 10.48% for total tocopherol content (Fritsche et al. 2012). In this study, allele-specific SNPs were identified and functional molecular markers were developed for high ATC in *B. juncea*. The developed functional AS-PCR markers for *VTE4* loci are highly efficient and would help to reduce time cost and also increase accuracy in the marker-assisted selection for the development of high ATC *Brassica* genotypes. Further, these markers could expedite the identification of homozygous/heterozygous state of alleles in inbred/pure lines and in segregating progenies through marker-assisted selection for high ATC in *B. juncea*.

In-silico analysis or computer-aided analysis uses computational methods and algorithms to study biological data. It has become an important analysis in plant-related research in recent times as it provides valuable insights and in-depth information into various aspects of plant biology. In the present study, *in-silico* analysis revealed the presence of *VTE4* gene on different genomes (A/B/C) of *Brassica* species. For instance, it was located on only A–genome of *A. thaliana, B. rapa* and *B. juncea*, B–the genome of *B. nigra,* C–genome of *B. oleracea* and *B. carinata*, and both A and C–genome of *B. napus*. Whole gene length varied from

species to species in *Brassica* species, but the CDS region was mostly conserved among the species. It might be due to more nucleotide deletion or insertion in the intronic than the exonic regions of the gene. These variations in the *VTE4* gene can be responsible for the varied expression of γ–tocopherol methyl transferase enzyme in the various species (Gilliland et al. 2006; Endrigkeit et al. 2009).

Genomic sequence-based homology is challenging but very crucial for biological studies. Also, phylogenetic studies are dependent on homology between the genomic sequences and provide evidence for evolutionary relationships among the species (Ullah et al. 2022). Homology can be examined at various levels viz., molecular, anatomical, and developmental in the living organisms. In this study, CDS sequences from various *Brassica* species were used to find out the phylogenetic relationship among them. It highlighted that *B. juncea* is very close to *B. rapa* phylogenetically than the other species and also shared common ancestor. It has been reported and proven that *B.juncea* carried A–genome from *B. rapa*, therefore, they showed more similarity in the CDS of the *VTE4* gene than others (Zou et al. 2016; Kang et al. 2021). The similarity between two species at the DNA or amino acid level indicates a common evolutionary history as these sequences are passed down from one generation to the next generation (Suchard et al. 2003; Fonseca et al. 2018). Besides, the codon usage pattern of the *VTE4* gene might be another reason for similarity, as it has been reported that genes with similar functions have alike codon usage modes (Tatarinova et al. 2010; WU et al. 2018; Chaudhary et al. 2022). In the future, the complete characterization and role of *VTE4* gene in *B. juncea* should be addressed. Besides, ATC is affected by environmental and physiological factors, *viz*., temperature, light, water availability, nutrient availability etc. Therefore, understanding the specific mechanisms of factors affecting ATC is very crucial and further research on this aspect can help to enhance the ATC content in Indian mustard.

Supplementary material

Supplementary Tables S1 to S3 are available at www.isgpb.org

Authors' contribution

Conceptualization of research (DKY, NS, SV); Designing of the experiments (DKY, NS, VKM); Contribution of experimental materials (NS, JN, YT); Execution of field/lab experiments and data collection (VKM, JN, RC, MKP); Analysis of data and interpretation (VKY, RS, SC); Preparation of the manuscript (VKY, SC, VM, MKP).

Acknowledgments

The First author is grateful to NCBI for accessing the genomic data, CSIR, New Delhi, for providing a Senior Research Fellowship for the PhD program, ICAR-CRP on Biofortification for financial support and ICAR-IARI, New Delhi, for providing laboratory facility and scientific support during the project.

References

Bergmüller E., Porfirova S. and Dörmann P. 2003. Characterization of an *Arabidopsis* mutant deficient in γ-tocopherol methyltransferase. Plant Mol. Biol., **52**: 1181–1190. https://doi.org/10.1023/B:PLAN.0000004307.62398.91

Brigelius-Flohé R. and Traber M.G. 1999. Vitamin E: function and metabolism. The FASEB Journal, **13:** 1145–1155. https://doi.org/10.1096/fasebj.13.10.1145

Chand S., Patidar O.P., Chaudhary R., Saroj R., Chandra K., Meena V.K., Limbalkar O.M., Patel M.K., Pardeshi P.P. and Vasisth P. 2021. Rapeseed-mustard breeding in India: Scenario, achievements and research needs. Islam A. K. M. A., Hossain M. A. and Islam A. K. M. M. (eds) Brassica breeding and biotechnology, IntechOpen, London, UK, p 174. http://dx.doi.org/10.5772/intechopen.96319.

Chaudhary R., Chand S., Alam B.K., Yadav P., Meena V.K., Patel M.K., Pardeshi P., Rathore S.S., Taak Y., Saini N. and Yadava,D.K. 2022. Codon Usage Bias for Fatty Acid Genes *FAE1* and *FAD2* in Oilseed *Brassica* Species. Sustainability, **14** (17): p.11035. https://doi.org/10.3390/su141711035

Das S. and Bansal M. 2019. Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. PLoS One, **14**: e0212678. https://doi.org/10.1371/journal.pone.0212678

Della Penna D. 2005. A decade of progress in understanding vitamin E synthesis in plants. J. Plant. Physiol., **162**: 729–737. https://doi.org/10.1016/j.jplph.2005.04.004

Della Penna D. and Pogson B.J. 2006. Vitamin synthesis in plants: Tocopherols and carotenoids. Annu. Rev. Plant. Biol., **57**: 711–738. https://doi.org/10.1146/annurev.arplant.56.032604.144301

Doyle J.J. and Doyle J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemi. Bulletin, **19**:11–15

Du H., Raman H., Kawasaki A., Perera G., Diffey S., Snowdon R., Raman R. and Ryan P.R. 2022. A genome-wide association study (GWAS) identifies multiple loci linked with the natural variation for Al. Funct. Plant Biol., **49**: 845–860. https://doi.org/10.1071/FP22073

Dua A., Chander S., Agrawal S. and Mahajan R. 2014. Antioxidants from defatted Indian Mustard (*Brassica juncea*) protect biomolecules against in vitro oxidation. Physiol. Mol. Biol. Plants, **20**: 539–543. https://doi.org/10.1007/s12298-014-0260-4

Dwiyanti M.S., Yamada T., Sato M., Abe J. and Kitamura K. 2011. Genetic variation of γ-tocopherol methyltransferase gene contributes to elevated α-tocopherol content in soybean seeds. BMC Plant Bio., **11**: 1-17. https://doi.org/10.1186/1471-2229-11-152

Endrigkeit J., Wang X., Cai D., Zhang C., Long Y., Meng J. and Jung C. 2009. Genetic mapping, cloning, and functional characterization of the *BnaX. VTE4* gene encoding a γ-tocopherol methyltransferase from oilseed rape. Theor. Appl. Genet., **119**: 567-575. https://doi.org/10.1007/s00122-009-1066-6

Fonseca I.C., Nogueira E., Figueirêdo P.H. and Coutinho S. 2018.

Analyzing similarities in genome sequences. Euro. Phys. Journ. E, **41**: 8. https://doi.org/10.1140/epje/i2018-11609-8

Fritsche S., Wang X., Li J., Stich B., Kopisch-Obuch F.J., Endrigkeit J., Leckband G., Dreyer F., Friedt W., Meng J. and Jung C. 2012. A candidate gene-based association study of tocopherol content and composition in rapeseed (*Brassica napus*). Front. Plant Sci., **3**: p.129. https://doi.org/10.3389/fpls.2012.00129

García-Moreno M.J., Fernández-Martínez J.M., Velasco L. and Pérez-Vich B. 2011. Molecular tagging and candidate gene analysis of the high gamma-tocopherol trait in safflower (*Carthamus tinctorius* L.). Mol. Breed., **28**: 367–379. https://doi.org/10.1007/s11032-010-9489-y

García-Moreno M.J., Fernández-Martínez J.M., Velasco L. and Pérez-Vich B. 2012. Genetic basis of unstable expression of high gamma-tocopherol content in sunflower seeds. BMC Plant Biol., **12**:71. https://doi.org/10.1186/1471-2229-12-71

Gilliland L.U., Magallanes-Lundback M., Hemming C., Supplee A., Koornneef M., Bentsink L. and DellaPenna D. 2006. Genetic basis for natural variation in seed vitamin E levels in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci., **103** (49): 18834-41. https://doi.org/10.1073/pnas.0606221103

Grygier A. 2023. Mustard Seeds as a Bioactive Component of Food. Food Reviews International **39**:4088–4101. https://doi.org/10.1080/87559129.2021.2015774

Gupta S., Sangha M.K., Kaur G., Banga S., Gupta M. Kumar, H. and Banga S.S. 2015. QTL analysis for phytonutrient compounds and the antioxidant molecule in mustard (*Brassica juncea* L.). Euphytica, **201**: 345–356. https://doi.org/10.1007/s10681-014-1204-3

Hass C.G., Tang S., Leonard S., Traber M.G., Miller J.F. and Knapp S.J. 2006. Three non-allelic epistatically interacting methyltransferase mutations produce novel tocopherol (vitamin E) profiles in sunflower. Theor. Appl. Genet., **113**: 767–782. https://doi.org/10.1007/s00122-006-0320-4

Kang L., Qian L., Zheng M., Che, L., Chen H., Yang L., You L., Yang B., Yan M., Gu Y. and Wang T. 2021. Genomic insights into the origin, domestication and diversification of *Brassica juncea*. Nat. Genet., **53**: 1392–1402. https://doi.org/10.1038/s41588-021-00922-y

Li Y., Zhou Y., Wang Z., Sun X. and Tang K. 2010. Engineering tocopherol biosynthetic pathway in *Arabidopsis* leaves and its effect on antioxidant metabolism. Plant Science, **178**: 312–320. https://doi.org/10.1016/j.plantsci.2010.01.004

Liao P., Chen X., Wang M., Bach T.J. and Chye M.L. 2018. Improved fruit α-tocopherol, carotenoid, squalene and phytosterol contents through manipulation of *Brassica juncea* 3-HYDROXY-3-METHYLGLUTARYL-COA SYNTHASE1 in transgenic tomato. Plant Biotechnol. J., **16**: 784–796. https://doi.org/10.1111/pbi.12828

Liu J., Huang S., Sun M., Liu S., Liu Y., Wang W., Zhang X., Wang H. and Hua W. 2012. An improved allele-specific PCR primer design method for SNP marker analysis and its application. Plant Methods., **8**: 34. https://doi.org/10.1186/1746-4811-8-34

Meena V.K., Taak Y., Chaudhary R., Chand S., Patel M.K., Muthusamy V., Yadav S., Saini N., Vasudev S. and Yadava D.K. 2022. Deciphering the Genetic Inheritance of Tocopherols in Indian mustard (*Brassica juncea* L. Czern and Coss). Plants, **11**:1779. https://doi.org/10.3390/plants11131779

Meena V.K., Shekhawat,H.V.S., Chand S., Choudhary K., Sharma J.K. and Lekha L., 2023. Advances in Molecular Marker Technology and Their Significance in Plant Improvement Strategies. In: El-Esawi M.A. (eds) Recent Trends in Plant Breeding and Genetic Improvement, Intech Open, 1-29. http://dx.doi.org/10.5772/intechopen.1002773.

Najafi M., Nasr-Esfahani M., Vatandoost J., Hassanzade-Khankahdani H. and Moeini M.J. 2024. Transcriptome-based analysis of candidate gene markers associated with resistance mechanism to *Phytophthora melonis* that causes root and crown rot in pumpkin. Funct. Plant Bio., **51** (2): 23038. https://doi.org/10.1071/FP23038

Ouyang Q., Fan C., Sun H., Zhang Y., Bai S. and Cai W. 2003. Cloning and analysis of a γ-tocopherol methyltransferase gene from *Brassica oleracea* and the function of its recombinant protein. Progress in Natur. Scien., **13**: 511–516. https://doi.org/10.1080/10020070312331343940

Patel M.K., Chaudhary R., Taak Y., Pardeshi P., Nanjundan J., Vinod K.K., Saini N., Vasudev S. and Yadava D.K. 2022. Seed coat colour of Indian mustard [*Brassica juncea* (L.) Czern. and Coss.] is associated with *Bju.TT8* homologs identifiable by targeted functional markers. Front. Plant Sci., **13**: 1012368. https://doi.org/10.3389/fpls.2022.1012368

Pearson K (1916) On a brief proof of the fundamental formula for testing the goodness of fit of frequency distributions, and on the probable error of "P." The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 31:369–378. https://doi.org/10.1080/14786440408635509

Rizvi S., Raza S.T., Ahmed F., Ahmad A., Abbas S. and Mahdi F. 2014. The role of Vitamin E in human health and some diseases. Sultan Qaboos Univ. Med. J., **14**:157–165

Shintani D. and Dellapenna D. 1998. Elevating the vitamin E content of plants through metabolic engineering. Science, **282**: 2098–2100. https://doi.org/10.1126/science.282.5396.2098

Suchard M.A., Weiss R.E., Sinsheimer J.S., Dorman K.S., Patel M. and McCabe E.R.B. 2003. Evolutionary Similarity Among Genes. J. Am. Stat. Assoc. **98**: 653–662. https://doi.org/10.1198/016214503000000558

Szewczyk K., Chojnacka A. and Górnicka M. 2021. Tocopherols and Tocotrienols—Bioactive Dietary Compounds; What Is Certain, What Is Doubt? Int. J. Mol. Sci., **22**: 6222. https://doi.org/10.3390/ijms22126222

Tatarinova T.V., Alexandrov N.N., Bouck J.B. and Feldmann K.A. 2010. GC3 biology in corn, rice, sorghum and other grasses. BMC Genomics, **11**: 308. https://doi.org/10.1186/1471-2164-11-308

Ullah H., Bashir K., Idrees M., Ullah A., Hassan N., Khan S., Nasir B., Nadeem T., Ahsan H., Khan M.I. and Ali Q. 2022. Phylogenetic analysis and antimicrobial susceptibility profile of uropathogens. PLoS One, **17**: e0262952. https://doi.org/10.1371/journal.pone.0262952

Valentin H.E., Lincoln K., Moshiri F., Jensen P.K., Qi Q., Venkatesh T.V., Karunanandaa B., Baszis S.R., Norris S.R., Savidge B. and Gruys K.J. 2006. The *Arabidopsis* vitamin E pathway gene 5-1 mutant reveals a critical role for phytol kinase in seed tocopherol biosynthesis. Plant Cell, **18**: 212–224. https://doi.org/10.1105/tpc.105.037077

Vinutha T., Maheswari C., Bansal N., Prashat G.R., Krishnan V., Kumari S., Dahuja A., Sachdev A. and Rai R.D. 2015. Expression analysis of γ-tocopherol methyl transferase genes and α-tocopherol content in developing seeds of soybean [*Glycine max* (L.) Merr.]. Indian J. Biochem. Biophys., **52**:

267–273.

Wang X., Zhang C., Li L., Fritsche S., Endrigkeit J., Zhang W., Long Y., Jung C. and Meng, J. 2012. Unraveling the Genetic Basis of Seed Tocopherol Content and Composition in Rapeseed (*Brassica napus* L.). PLoS One, **7**: 8–14. https://doi.org/10.1371/journal.pone.0050038

Wu Y., Li Z., Zhao D. and Tao J. 2018. Comparative analysis of flower-meristem-identity gene *APETALA2* (*AP2*) codon in different plant species. J. Integr. Agric., **17**: 867–877. https://doi.org/10.1016/S2095-3119(17)61732-5

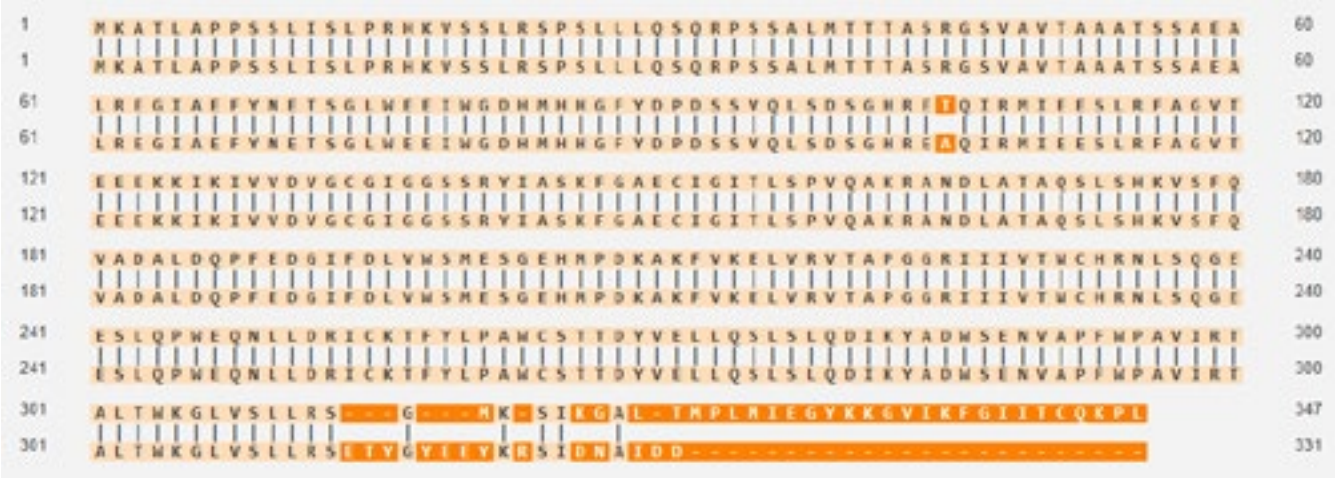
Yadava D.K., Taak Y., Saini N., Nanjundan J. and Vasudev S. 2022. Brassica Breeding. In: Yadava DK, Dikshit HK, Mishra GP, Tripathi S (eds) Fundamentals of Field Crop Breeding. 1st edn. Springer Singapore, pp 779–835. https://doi.org/10.1007/978-981-16-9257-4_15

Yadava S.K., Arumugam N., Mukhopadhyay A., Sodhi Y.S., Gupta V., Pental D. and Pradhan A.K. 2012. QTL mapping of yield-associated traits in *Brassica juncea*: meta-analysis and epistatic interactions using two different crosses between east European and Indian gene pool lines. Theor. Appl. Genet., **125**: 1553–1564. https://doi.org/10.1007/s00122-012-1934-3

Yusuf M.A., Kumar D., Rajwanshi R., Strasser R.J., Tsimilli-Michael M. and Sarin N.B. 2010. Overexpression of γ-tocopherol methyl transferase gene in transgenic *Brassica juncea* plants alleviates abiotic stress: Physiological and chlorophyll a fluorescence measurements. BiochimBiophys Acta Bioenerg., **1797**: 1428–1438. https://doi.org/10.1016/j.bbabio.2010.02.002

Yusuf M.A. and Sarin NB. 2007. Antioxidant value addition in human diets: Genetic transformation of *Brassica juncea* with *γ-TMT* gene for increased α-tocopherol content. Transgenic Res., **16**: 109–113. https://doi.org/10.1007/s11248-006-9028-0

Zou J., Hu D., Liu P., Raman H., Liu Z., Liu X., Parkin I.A., Chalhoub B. and Meng J. 2016. Co-linearity and divergence of the A subgenome of *Brassica juncea* compared with other *Brassica* species carrying different A subgenomes. BMC Genom., **17**: 18. https://doi.org/10.1186/s12864-015-2343-1



Supplementary Fig. 1. ***VTE4*** **amino acid sequence comparison between NPJ203 (Upper- 347 amino acids) and RLC3 (Lower- 331 amino acids)**

Supplementary Table S1. Detailed information of diverse panel of 96 ***B. juncea*** **genotypes used in the present study**

S. N.	*Genotypes*	*Pedigree/Source*	*Origin*
1	AJ-11	China germplasm	China
2	AJ-3	China germplasm	China
3	Ashavati	Advanced Breeding line	-
4	Basanti	Indian Variety (Pedigree)	CSAUA&T Kanpur, India
5	CN-101813	Introduced from Canada gene bank	Canada
6	CN-101845	Introduced from Canada gene bank	Canada
7	CN-105233	Introduced from Canada gene bank	Canada
8	CN-105257	Introduced from Canada gene bank	Canada
9	CN-105305	Introduced from Canada gene bank	Canada
10	CN-105306	Introduced from Canada gene bank	Canada
11	CN-105309	Introduced from Canada gene bank	Canada
12	CN-105312	Introduced from Canada gene bank	Canada
13	CN-105364	Introduced from Canada gene bank	Canada
14	CN-105379	Introduced from Canada gene bank	Canada
15	CN-113780	Introduced from Canada gene bank	Canada
16	CN-34005	Introduced from Canada gene bank	Canada
17	CN-34008	Introduced from Canada gene bank	Canada
18	DJ-12	Resynthesized derived mustard	-
19	DJ-5	Resynthesized derived mustard	-
20	DJ-65	Resynthesized derived mustard	-
21	DRMRIJ 17-38	Indian germplasm	Bharatpur, Indian
22	EC-564949-1	Exotic germplasm	-
23	EJ-22	Advanced Breeding line	-
24	EM-1	Advanced Breeding line	-
25	EC-564949-1	Exotic germplasm	-
26	Giriraj	HB 9908 X HB 9916	Bharatpur, India
27	GR-325	German introduction	Germany
28	I-79 (M)	Introgression line	-
29	IC-597869	Indian germplasm	India
30	IC-597873	Indian germplasm	India
31	IC-597875	Indian germplasm	India
32	IC-597876	Indian germplasm	India
33	IC-597880	Indian germplasm	India
34	IC-597881	Indian germplasm	India
35	IC-597882	Indian germplasm	India
36	IC-597890	Indian germplasm	India
37	IC-597904	Indian germplasm	India
38	IC-597910	Indian germplasm	India
39	IC-597949	Indian germplasm	India

40	IC-766097	Indian germplasm	India
41	I-79 (M)	Introgression line	-
42	I-79 (M)	Introgression line	-
43	I-79 (M)	Introgression line	-
44	I-79 (M)	Introgression line	-
45	I-79 (M)	Introgression line	-
46	I-79 (M)	Introgression line	-
47	JAGANNATH	Multicross between Varuna/inter cross derivatives/synthetic *Brassica juncea*	IARI, New Delhi, India
48	YSRL-9	Indian germplasm	India
49	RH749	Indian germplasm	India
50	JM-06010-1	Australia germplasm	Australia
51	JM-1	Australia germplasm	Australia
52	KDM-1049	Indian germplasm	India
53	KDM-49-1	Indian germplasm	India
54	KRISHNA	Indian Variety	GBPUAT, Pantnagar, India
55	NC-37362	East European germplasm	East European
56	NC-533726	East Europe germplasm	East European
57	NC-660	East Europe germplasm	East European
58	NDRE-4	Indian Variety	India
59	NPJ-113	VEJ Open × Pusa Agrani	IARI, New Delhi, India
60	NPJ-161	Indian germplasm	IARI, New Delhi, India
61	PBR-357	Indian Variety	India
62	PBR-97	Indian Variety	India
63	PCR-7	Advanced Breeding line	Bharatpur, India
64	PCR-9403	Advanced Breeding line	-
65	PM-67	Indian Variety	Gujarat, India
66	PCR-9403	Advanced Breeding line	-
67	PR-2001-42	Advanced Breeding line	-
68	PR-2001-42	Advanced Breeding line	-
69	PR-2001-42	Advanced Breeding line	-
70	PUSA BAHAR	HI1011/HD2348// MENDOS//IWP 72/DL 153-2	IARI, New Delhi, India
71	PUSA BOLD	Varuna × BIC 1780	IARI, New Delhi, India
72	Pusa jai kisan	Somaclone of Varuna	IARI, New Delhi, India
73	PUSA MAHAK	Pusa Bold × Glossy mutant	IARI, New Delhi, India
74	RB-60	Indian germplasm	India
75	RC-1270	Indian germplasm	India
76	RC-1271	Indian germplasm	India
77	RC-132	Indian germplasm	India
78	RC-371	Indian germplasm	India
79	RC-371-1	Indian germplasm	India
80	RC-571	Indian germplasm	India
81	RC-891-1	Indian germplasm	India
82	RCQR-9901	Advanced Breeding line	-
83	RE-11	East European germplasm	East European
84	RE-15	East European germplasm	East European
85	RE-44	East European germplasm	East European
86	RE-7-1	East European germplasm	East European
87	RGN-34	Advanced Breeding line	-
88	RGN-73	Indian Variety	India
89	RH-4193	Indian germplasm	India
90	SITARA SAGAR	Indian Variety	India
91	TN-3	Advanced Breeding line	-
92	VASUNDHRA	RH 839 × RH 30	CCS HAU, Hisar, India
93	RLC-3	JM06003 × JM06020	PAU, Ludhiana, India
94	NPJ203	Pedigree	IARI, New Delhi, India
95	IC-597878	Indian germplasm	India
96z	PM-29	(ZEM-2 × Pusa Barani) × EC-287711	IARI, New Delhi, India

Supplementary Table S2. *VTE4* **gene sequences from Parents (NPJ203 & RLC3) used in the study**

1. VTE4 gene sequence of NPJ203

5′TAATAAAAAAACATAATCCCGTACACAGCCACATTTCTTGTTTCTCCAACCAACCTCTCATTATAAATGAAAGC
GACACTCGCACCACCCTCCTCTCTCATAAGCCTCCCCAGGCACAAAGTATCTTCTCTCCGTTCACCGTCGCTTCT CCTTCAGTCCCAACGGCCATCCTCA
GCCTTAATGACGACGACGGCATCACGTGGAAGCGTGGCTGTGACGGC TGCTGCTACCTCCTCCGC TGAGGCGCTGCGAGAAGGAATA GCGGAATTCTAC
AACGAGACGTCGGGATTATG GGAGGAGATTTGGGGAGATCATATGCATCACGGCTTCTACGATCCTGATTCCTCTGTTCAACTTTCAGATTCCGGTCACC
GGGAAACTCAGATCCGGATGATTGAAGAGTCTCTACGTTTCGCCGGCGTTACTGGTTCGCTTCTCATGCTCTACAC TTGAGTTGATACGTTGT
TTATTAT AAACATTTTTTTGAACTTTTATTATAACAATTCTTACAAACAAATTACTC TTTGAACTCTTTAACATCTATAACAAAGGTTTA
GTTTTACTTTTGATTTGTTGTTGGTAACAGAAATGAGTAGGGA TGTTTGAAGTCAGATATAGCCTTTCTG TTTATCCCTTGGGAAGAAAGGCTTACAGT
AAGCCACGTCCCATCCAGAAGCAGACCCATTCCCTAACTAATCATTTTTATGAACAATTTGTAACACTATTATTCCTAGATATTTTTTTTT
ACGTTTAGTTACCCTAACTCTTTGTATATAAGACACGAGGTGATTTTTCACATTATATATCAAAACATAGACATAGTTTTTTTGAGAA
AATATATCATACATAGTTGTAACTTAGAATTATATATTTTTGAGAAAAAAAACTCAGTAATAATTTTCTTATAATTATTCA
TAGTTTTATATTTATTAATAATAAGATTTTGTAAGCTCTTTTTGAAACTATTATGGATAATGAATAAGTTCCCCATTTCAAGATT
AAGCATACAATTTAAACTGAAATAATAATGCGCATTCGGTGAAAATATCTTCTGCTTGGGATTGTTGTT GTTAATATTAATTTA TATTATAAACACATGAC
TTTTTTAAAGAAGAGGAGAAAAAGATAAAGATAGTGGTGGATGTTGGGTGTGGGATCGGAGGAAGCTCAAGGTATATTGCCTCTAAATTT
GGTGCCGAATGCATTGGCATCACACTCAGTCCCGTTCAAGCCAAGAGAGCAAATGATCTCGCCACCGCTCAATCACTCTCTC
ATAAGGTGTCTTCTCGTACATTCGACCATTCTTTCTGCGGATAATCTGATCTAACTGAGACGCCATTGGACCAGGTTTCC
TTCCAAGTTGCAGATGCATTGGACCAACCATTTGAAGATGGTATATTCGATCTTGTTTGGTCAATGGAAAGCGGCGAGC ATATGCCTGA CAAGGCCAA
GGTATACTAGCTCAGCATAACTTTTATACTAGAT TTACTAGACAATATCTATCTT TTCATGTCAATGATGTCCAATA ATTTTAAAATA AACAAAAGAA
GGATGTGGGGTAAAATTTTGTCAAATTTATATAACAACACGTTTTCTATTTAGTTATGTCATGGTTTCTTTTTGTCTAAAAAATTTTAGGCAGAG
TTTACAAAAAGAAAATTGTAGTATCTGTTCGAAAACAGAATCTTAGTGTGGTATTTCAGAAACTCATTCATGAGTCTTCCTTGTGGAAGC
ATATTTACTGTGTGTGCGAAATGAGTGTAGTTCGTGAAGGAATTGGTACGTGTGACGGCTCCAG GAGGAAGGATAATAATA GTGACATGGTGC
CACAGAAATCTATCTCAAGGGGAAGAATCTTTGCAGCCATGGGAGCAGAACCTCTTGGACAGAATCTGCAAAACATTTTATCTCCCGGCCTG
GTGCTCCACCACTGATTATGTCGATTGCTTCAATCCCTCTCGCTCCAGGTTATTATATTTCTCACGCTCTGATGCTAAAATCA
GTAAGTATTGTCTCAAATATATGTGTGTTTGTAGGATATTAAGTATGCAGATTGGTCAGAGAACGTAGCTCCTTTCTGG
CCGGCGGTTATACGAACCGCATTAACGTGGAAGGGCCTCGTGTCTCTGCTTCGTAGTGGTATGTTTCCGCAATGTTG
TTTACATTCATGATTCCAAATGTTTATAAGATTAGAAACATACAGGTATGAAGAGTATAAAAGGAGCATTGACAATGCCATTGATGATTGAAGGGTACAA
GAAAGGTGTCATTAAGTTTGGCATCATCACTTGCCAGAAGCCTCTCTAAGTTTAATCTAAACA3′

2. VTE4 gene sequence of RLC3

5′TAATAAAAAAACATAATCCCGTACACAGCCACATTTCTTGTTTCTCCAACCAACCTCTCATTATAAATGAAAGCGACACTCGCACCGCCCTCCTCTCTCATAA
GCCTCCCCAGGCACAAAGTATCTTCTCTCCGTTCACCGTCGCTTCTCCTTCAGTCCCAACGGCCATCCTCAGCCTTAATGACGACGACGGCATCACGTGGAA
GCGTGGCTGTGACGGCTGCTGCTACCTCCTCCGCTGAGGCGCTGCGAGAAGGAATAGCGGAATTCTACAACGAGACGTCGGGATTATGGGAGGAGATTTG
GGGAGATCATATGCATCACGGCTTCTACGATCCTGATTCCTCTGTTCAACTTTCAGATTCCGGTCACCGGGAAGCTCAGATCCGGATGATTGAAGAGTCTCTA
CGTTTCGCCGGCGTTACTGGTTCGCTTCTCATGCTGTACACTTGAGTTTGATACGTTGTTTATTATAAACATTTTTTTGAACTTTTATTATAAACAATTCTTACA
AACAAATTACTCTTTGAACTCTTTAAAATCTATAACAAAGGTGTAGTTTTACTTCTGATTTGTTGTTGGTAACAGAAATGAGTAGGGATGTTTGAAGTCAGATA
TAGCCTTTCTGTTTATCCCTTGGGAAGAAAGGCTTACAGTAAGCCACGTCCCATCCAGAAGCAGACCCATTCCCTAACTAATCATTTTTATGAACAATTTGTA
ACACTATTATTCCTAGATATTTTTTTTTTACGTTTAGTTACCCTAACTCTTTGTATATAAGACACGAGGTGATTTTTCACATTATATATCAAAACATAGACATAGTTT
TTTTGAGAAAATATATCATACATAGTTGTAACTTAGAATTATATATTTTTGAGAAAAAAAACTCAGTAATAATTTTCTTATAATTATTCATAGTTTTATATTTATTAAT
AATAAGATTTTGTAAGCTCTTTTTGAAACTATTATGGATAATGAATAAGTTCCCCATTTCAAGATTAAGCATACAATTTAAACTGAAATAATAATGCGCATTCGG
TGAAAATATCTTCTGCTTGGGATTGTTGTTGTTAATATTAATTTATATTATAAACACATGACTTTTTTAAAGAAGAGGAGAAAAAGATAAAGATAGTGGTGGAT
GTTGGGTGTGGGATCGGAGGAAGCTCAAGGTATATTGCCTCTAAATTTGG TGCCGAATGCATTGGCATCACACTCAGTCCCGTTCAAGCCAAGAGAGCAA
ATGATCTCGCCACCGCTCAATCACTCTCTCATAAGGTGTCTTCTCGTACATTCGACCATTCTTTCTGCGGATAATCTGATCTAACTGAGACGCCATTGGACCAGG
TTTCCTTCCAAGTTGCAGATGCATTGGACCAACCATTTGAAGATGGTATATTCGATCTTGTTTGGTCAATGGAAAGCGGTGAGCATATGCCTGACAAGGCCAA
GGTATACTAGCTCAGCATAACTTTTATACTAGATTTACTAGACAATATCTATCTTTTCATGTCAATGATGTCCAATAATTTTAAAATAAACAAAAGAAGGATGTG
AGGGTAAAATTTTGTCAAATTTATATAACAACACGTTTTCTATTTAGTTATGTCATGGTTTCTTTTTGTCTAAAAAATTTTAGGCAGAGTTTACAAAAAGAAAA
TTGTAGTATCTGTTCGAAAACAGAATCTTAGTGTGGTATTTCAGAAACTCATTCATGAGTCTTCCTTGTGGAAGCATATTTACTGTGTGTGCGAAATGAGTGTA
GTTCGTGAAGGAATTGGTACGTGTGACGGCTCCAGGAGGAAGGATAATAATAGTGACATGGTGCCACAGAAATCTATCTCAAGGGGAAGAATCTTTGCAGC
CATGGGAGCAGAACCTCTTGGACAGAATCTGCAAAACATTTTATCTCCCGGCCTGGTGCTCCACCACTGATTATGTCGAGTTGCTTCAATCCCTCTCGCTCCA
GGTTATTATATTCTCACGCTCTGATGCTAAAATCAGTAAGTATTGTCTCAAATATATGTGTGTTTGTAGGATATTAAGTATGCAGATTGGTCAGAGAACGTAGCTC
CTTTCTGGCCGGCGGTTATACGAACCGCATTAACGTGGAAGGGCCTTGTGTCACTGCTTCGTAGTGGTATGTTTCCGCAATGTTGTTTACATTCATGATTCCA
AATGTTTATAAGATTAGAAACATACGGGTATGAAGAGTATAAAAGGAGCATTGACAATGCCATTGATGATTGAAGGGTACAAGAAAGGTGTCATTAAGTTTGG
CATCATCACTTGCCAGAAGCCTCTCTAAGTTTAATCTAAACA3′

Supplementary Table S3. ***VTE4*** **amino acid sequences from Parents (NPJ203 & RLC3)**

1. VTE4 amino acid sequence of NPJ203 (347 amino acids)
MKATLAPPSSLISLPRHKVSSLRSPSLLLQSQRPSSALMTTTASRGSVAVTAAATSSAEALREGIAEFYNETSGLWEEIWGDHMHHGFYDPDSSVQLSDSGHRETQI RMIEESLRFAGVTEEEKKIKIVVDVGCGIGGSSRYIASKFGAECIGITLSPVQAKRANDLATAQSLSHKVSFQVADALDQPFEDGIFDLVWSMESGEHMPDKAKFVKE LVRVTAPGGRIIIVTWCHRNLSQGEESLQPWEQNLLDRICKTFYLPAWCSTTDYVELLQSLSLQDIKYADWSENVAPFWPAVIRTALTWKGLVSLLRSGMKSIKGALT MPLMIEGYKKGVIKFGIITCQKPL

2. VTE4 amino acid sequence of RLC3 (331 amino acids)
MKATLAPPSSLISLPRHKVSSLRSPSLLLQSQRPSSALMTTTASRGSVAVTAAATSSAEALREGIAEFYNETSGLWEEIWGDHMHHGFYDP DSSVQLSDSGHREAQIRMIEESLRFAGVTEEEKKIKIVVDVGCGIGGSSRYIASKFGAECIGITLSPVQAKRANDLATAQSLSHKVSFQVADA LDQPFEDGIFDLVWSMESGEHMPDKAKFVKELVRVTAPGGRIIIVTWCHRNLSQGEESLQPWEQNLLDRICKTFYLPAWCSTTDYVELLQ SLSLQDIKYADWSENVAPFWPAVIRTALTWKGLVSLLRSETYGYEEYKRSIDNAIDD