Molecular cloning of *Brassica napus TRANSPARENT TESTA 2* (*BnTT2*) gene family encoding potential MYB regulatory proteins of proanthocyanidin biosynthesis

WEI Yunliang, LI Jiana, LU Jun, TANG Zhanglin, PU Dongchun, CHAI Yourong

College of Agronomy and Biotechnology, Southwest University, 216 Tiansheng Road, Beibei, Chongqing, 400716, P. R. China. Email: chaiyourong@163.com

Abstract

Three members of *Brassica napus TRANSPARENT TESTA 2* (*BnTT2*) gene family encoding potential R2R3-MYB regulatory proteins of proanthocyanidin biosynthesis were isolated. *BnTT2-1*, *BnTT2-2* and *BnTT2-3* all are 1102 bp with 2 introns, and have a 938-bp full-length cDNA with a 260-amino-acid open reading frame. They share 98.2%-99.3% nucleotide and 96.5%-98.5% amino acid identities to each other, and are orthologous to *Arabidopsis thaliana TT2* (*AtTT2*) with 74.1%-74.8% nucleotide and 71.1%-71.8% amino acid identities. An mRNA type of *BnTT2-2* was found to contain unspliced intron 2 and encode a premature protein. They all have an alternative polyadenylation site. *BnTT2-1* and *BnTT2-3* also have an alternative transcription initiation site. Aligned with *AtTT2*, their 5′ UTRs are astonishingly conserved, and 2 conserved regions were also found in their 3′ UTRs. Oligonucleotide deletion leads to double start codons of them. Resembling AtTT2, BnTT2 proteins are nuclear-located R2R3-MYB proteins containing predicted DNA binding sites, bHLH interaction residues and transcription activation domains. Southern blot indicated that there might be three *BnTT2* members in *B. napus*, lower than triplication-based prediction. Semi-quantitative RT-PCR revealed that the expression of *BnTT2-2* is most like *AtTT2* with intensive expression in young seeds, but it is also expressed in root in which *AtTT2* has no expression. *BnTT2-1* shows lower tissue specificity and transcription levels, while *BnTT2-3* is the lowest. Comparative cloning and RT-PCR indicated that seed-color near-isogenic lines L1 and L2 have equivalent *BnTT2* genes, and the yellow seed color in L2 might be caused by locus/loci other than *BnTT2*. Our results lay the basis for further investigating the regulatory mechanism of *BnTT2* genes in flavonoid pathway and for transgenic creating of novel yellow-seeded *B. napus* stocks.

Key words: Rapeseed (*Brassica napus*), Flavonoid, MYB, proanthocyanidin, *TRANSPARENT TESTA 2* (*TT2*)

Introduction

Yellow seed is widely accepted as a good-quality trait of rapeseed (*Brassica napus* L.), because it contains much less dark pigments in the crude oil and indirectly enhances seed oil content and meal protein content. Yellow-seeded *B. napus* germplasms used today are exclusively generated via interspecific hybridization or artificial mutation. But in *B. napus* yellow seed color shows drastic variation in phenotype stability, which has been the major obstacle in yellow-seeded rapeseed breeding (Heneen & Brismar, 2001). Elucidation of molecular mechanism of seed color formation is crucial to molecular breeding of applicable yellow-seeded cultivars of *B. napus*.

Seed coat pigment is polymers of proanthocyanidin (PA), the so-called condensed tannin (CT). PA is synthesized through phenylpropanoid-flavonoid pathway. In *Arabidopsis thaliana*, mutants of many flavonoid pathway genes have been identified as *transparent testa* (*tt,* i.e. yellow to pale brown seed) phenotypes. Quite a few regulatory genes of flavonoid pathway have been isolated from *A. thaliana*, such as *TT1*, *TT2*, *TT8*, *TT16*, *TTG1* (*TRANSPARENT TESTA GLABRA 1*) and *TTG2* (Debeaujon et al., 2003). MYB transcription factors are involved in a wide array of cellular processes, and some are important regulators of phenylpropanoid pathway (Mehrtens et al., 2005). *A. thaliana TT2* (*AtTT2*) gene encodes a R2R3-MYB transcription factor playing an important role in regulating PA synthesis and transport (Nesi et al., 2001). *TT2* is expressed in seed coat during early stages of embryogenesis. In combination with *TT8* and *TTG1*, it regulates the expression of flavonoid late biosynthetic structural genes *TT3*, *TT18*, *TT12* and *BAN*. In an attempt to elucidate the molecular mechanism of seed color trait of rapeseed, the *B. napus TT2* (*BnTT2*) gene family was cloned and characterized in this study.

Materials and methods

Plant materials and nucleic acid isolation: B. napus black-seeded line 5B and seed color near-isogenic lines L1 (black-seeded) and L2 (yellow-seeded) were traditionally grown. For 5B, root, hypocotyl, cotyledon, stem, leaf, bud, flower, pericarp, and seed of 10, 20 and 30 d after flowering (DAF) were sampled. For L1 and L2, bud, flower, pericarp, and seed of the 3 stages were sampled. Total RNA was isolated using a CTAB method and treated with RNase-free DNase I. Total genomic DNA was extracted from leaves using a CTAB method.

Rapid amplification of 3′ *and* 5′ *cDNA ends of* BnTT2 *members:* Five μg of total RNA from 5B reproductive organs was used to generate first-strand total cDNA using GeneRacer kit (Invitrogen, USA). Sense primers FTT2-31 (5′-CAGATGGTCGTTGATAGCTGG-3′) and FTT2-32 (5′-TTGATAGCTGGGAGGCTTCCAGG-3′) were paired with kit

primers 3′ Primer and 3′ Nested Primer for primary and nested amplifications of 3′ cDNA ends respectively. While kit primers 5′ Primer and 5′ Nested Primer were paired with RTT2-51 (5′-ATTCTTTATTTCATTGTCTGTTCG-3′) and RTT2-52 (5′-CCTGGAAGCCTCCCAGCTATCAA-3′) for primary and nested amplifications of the 5′ cDNA ends respectively. 50-μL standard *Taq* PCR system was adopted. In primary amplification, 2 μL of first-strand cDNA was used as template. Then 0.1 μL of primary amplification product was used for nested amplification. Annealing temperature of 52°C was used. PCR products were subcloned using pMD18-T vector (TaKaRa) and sequenced using primers M13F and M13R.

Amplification of full-length cDNAs and genomic sequences of BnTT2 *members:* Sense primer FBnTT2 (5′-ACCCGGGAATCTATTCTCAACACAACGCTAAAG-3′) and antisense primer RBnTT2 (5′-AGAGCTCAGATATTACTTGTAATTAGAAGCGTTC-3′) were synthesized to simultaneously amplify the 3 gene members in one tube. Two μL of total cDNA and 0.5 μg of total genomic DNA of line 5B were used as templates for cDNA and genomic sequence amplifications respectively. After subcloning, PCR-positive colonies were further screened. 3 member-specific antisense primers were designed: RBNTT2-1S (5′-GTATCAGTCGAAACATAATCTCCT-3′), RBNTT2-2S (5′-CGATGAAGGAGAACTAGCTTGT-3′) and RBNTT2-3S (5′-CCAAACCATCAAAGCCCATTAA-3′). Annealed at 59°C, primer pairs FTT2-32/RBNTT2-1S, FTT2-32/RBNTT2-2S and FTT2-32/RBNTT2-3S were used to screen for *BnTT2-1*, *BnTT2-2* and *BnTT2-3* respectively. Double-positive colonies were sequenced.

Southern blot detection: 35 μg of 5B genomic DNA for each enzyme was fully digested with *Eco*RI, *Eco*RV, *Hin*dIII and *Xba*I respectively for electrophoresis and transfer onto nylon membrane. Using primer pair FTT2-3/RBnTT2, a 601-bp fragment of *BnTT2-1* cDNA was amplified and labeled with Digoxigenin-11-dUTP to carry out Southern blot hybridization at 39°C for 16 h and immunological detection (DIG Kits, Roche).

Semi-quantitative RT-PCR detection: 5 μg of each RNA sample were reverse-transcribed using primer Oligo$(dT)_{20}$ (SuperScript III First-Strand Synthesis SuperMix, Invitrogen). 0.5-μL first strand cDNA of each sample was used in a 50-μL *Taq* PCR. Amplification for *ACT2* was used as internal control. Annealed at 55°C in a 30-cycle amplification, primers FBnTT2A (5′-AGAGCTCAACAGAGGAGCTTGGACCGATCA-3′) and RBnTT2A (ACCCGGGCCAGCACCTCGTCAGATGAAACA-3′) were adopted to detect the overall expression of *BnTT2* gene family in various organs of *B. napus*. In detection of individual members, primer pairs FTT2-32/RBNTT2-1S, FTT2-32/RBNTT2-2S and FTT2-32/RBNTT2-3S were used for *BnTT2-1*, *BnTT2-2* and *BnTT2-3* with annealing temperatures of 62.0°C, 64.0°C and 60.2°C, respectively. *BnTT2-1* and *BnTT2-2* were amplified for 35 cycles, while 40 cycles for *BnTT2-3*.

Comparative identification of L1 and L2: Amplifications of genomic sequences of *BnTT2* members from total genomic DNA of L1 and L2 respectively were performed. RNA samples from bud, flower, 10 DAF seed, 20 DAF seed and 30 DAF seed of L1 and L2 were reverse-transcribed and amplified for 30 and 35 cycles respectively using primer pair FBnTT2A/RBnTT2A.

Results

Isolation of full-length sequences of the 3 BnTT2 *members*: Gel detection of nested PCRs of the 3′ and 5′ RACE showed a wide band around 650 bp and a bright band of about 380 bp respectively. Subcloned *E. coli* colonies showed polymorphism in insert length. Representative colonies of each insert length were sequenced and they most probably represented 3 different genes with alternative polyadenylation sites and alternative initiation sites in some members. Full-length cDNA amplification yielded a band of about 950 bp. Nested PCR checking successfully screened out colonies corresponding to full-length cDNAs of *BnTT2-1, BnTT2-2* and *BnTT2-3* which were all 938 bp proved by sequencing. A 1013-bp premature version of *BnTT2-2* full-length cDNA was also sequenced. Amplification of the genomic sequences yielded a bright band of about 1100 bp. Colonies corresponding to the 3 genes all had a sequenced length of 1102 bp.

Nucleotide characterization of BnTT2 *gene family: BnTT2-1*, *BnTT2-2* and *BnTT2-3* are all 1102 bp. They all contain 2 introns, one at 201-289 bp and another at 420-494 bp. G+C contents of non-coding regions especially the intron 1 and the 3′ UTRs are obviously lower than coding regions. They all have a 783-bp open reading frame (A_{59}-G_{841}) encoding a 260-aa protein. *BnTT2-1* and *BnTT2-3* have an alternative transcription start site A_{12}. Alternative polyadenylation sites were also found, C_{1035} of *BnTT2-1* and G_{1080} of *BnTT2-2* and *BnTT2-3* respectively. *BnTT2-2* has another type of mature mRNA (*BnTT2-2PM*) with the 75-bp second intron not excised. This leads to pre-stop of the ORF by intron-derived TAA.

These 3 genes share surprisingly high similarities to each other. On whole genomic scale, identities of *BnTT2-1* to *BnTT2-2*, *BnTT2-1* to *BnTT2-3* and *BnTT2-2* to *BnTT2-3* are 98.2%, 99.3% and 98.2% respectively, while on mRNA level the values are 98.4%, 99.1% and 98.4% respectively. *BnTT2-1*, *BnTT2-2* and *BnTT2-3* have 74.8%, 74.1% and 74.1% of nucleotide identities to *AtTT2* on full-length genomic scale, but on ORF scale the identities rise to 79.8%, 79.2% and 79.2% respectively. The 5′ coding region of the 3 genes has significantly higher similarities to *AtTT2* and other *MYB* genes than the 3′ coding region does. The first 59 bp (mainly the 5′ UTR) of the 3 genes show higher identities to *AtTT2* than most of the coding regions do. However, the sequence AGAAAAGTGAGAA just prior to the start codon ATG in *AtTT2* has been degenerated to a novel start codon ATG in the 3 *BnTT2* genes, which leads to a style of double start codons. The sequences have been submitted to GenBank under accession numbers from DQ778643 to DQ778649.

Conservation and structural features of the deduced BnTT2 proteins: BnTT2-1, BnTT2-2 and BnTT2-3 all are 260 aa. BnTT2-1 possesses a Mw of 29.72 kDa and a pI of 9.16, while 29.64 kDa and 9.25 for BnTT2-2, and 29.57 kDa and 9.25 for BnTT2-3, respectively. The *BnTT2-2PM* mRNA encodes a polypeptide of only 113 aa with a Mw of 13.28 kDa and a pI of

10.21. The 3 BnTT2 proteins have no signal peptide and transmembrane domain. WoLFPSORT predicted a nuclear localization signal RKRLPKSQTNQQKSRKH in them, which is quite similar to that of AtTT2. NetPhos 2.0 predicted 14-15 potential phosphorylation sites in each member. Resembling AtTT2 (Nesi et al., 2001), BnTT2 proteins are typical R2R3-MYB proteins. NCBI Conserved Domain search detected two SANT/Myb_DNA-Binding conserved domains tandemly located at N-terminal regions, SANT at N_{16}-P_{66} and K_{69}-K_{117}, and Myb at R_{17}-L_{64} and R_{70}-L_{115}.

BnTT2-1 shows 98.5%/98.9% identities/positives to BnTT2-3, while 96.5%/96.9% between BnTT2-1 and BnTT2-2, and 96.9%/97.3% between BnTT2-2 and BnTT2-3, respectively. BnTT2-1, BnTT2-2 and BnTT2-3 show 71.4%/77.1%, 71.8%/77.8% and 71.1%/77.1% of overall identities/positives to AtTT2 and moderate similarities to other 7 non-cruciferous MYB proteins such as GHMYB36. These 7 proteins are extremely conserved in the two MYB_DNA-binding domains. A conserved sequence between AtTT2 and OsMYB3 in the highly variable C-terminal region has been revealed (Nesi et al., 2001). Our multi-alignment undoubtedly proved the conservation of this region. It has a 9-residue consensus V(I/V)R(T/P)(K/R)A(I/L/V)(R/K)C, with the sequence VIRTKAIRC from AtTT2 represents the most frequently occurred consensus. The A residue in the central is the only invariable residue, and most proteins contain at least 2-3 basic (R/K) residues but no acidic residues (D/E) in this motif and its proximal regions, so we tentatively call this motif as bA-motif. In the 3 α-helix regions of R2-MYB and R3-MYB and in the linker region of the 2 MYB repeats, the 3 BnTT2 proteins are identical to AtTT2 except for divergence at 4 non-conserved residues in the 1st α-helix of R2-MYB. Of all the conserved W residues in MYB repeats (Stracke et al., 2001), the linker sequence between the R2 and R3 repeats (Nesi et al., 2001), the 2 cysteines involved in REDOX-dependent DNA binding (Heine et al., 2004), and residues of conserved signature sequence [DE]Lx2[RK]x3Lx6Lx3R proposed for interaction with R/B-like bHLH proteins (Zimmermann et al., 2004), BnTT2 proteins are identical to AtTT2 and other MYB proteins aligned. Their R2R3-MYB domains also share very wide homologies to various plant R2R3-MYB proteins.

Based on SOPMA prediction, the secondary structures of the 3 BnTT2 proteins are similar to each other. α-helices (34.62%-37.69%) and random coils (50.00%-55.38%) constitute the most part, while extended strands (5.38%-6.92%) and beta turns (3.85%-6.15%) also contribute a little. α-helices mainly distribute in the R2R3-MYB region and the C-terminus. In each MYB repeat there are 3 helices corresponding to those reported in AtTT2 and other MYB proteins (Nesi et al., 2001). Like AtTT2, the 3 BnTT2 proteins also contain 2 strong helices in the C-terminus (Nesi et al., 2001). An N-terminal helix was found preceding the R2-MYB domain, and the bA-motif in the variable C-terminal region is invariably associated with some extended strands. SWISS-MODEL prediction yielded tertiary structures of only the MYB domain regions (K_{12}-K_{117}) of the 3 BnTT2 proteins. In this region the 3 BnTT2 proteins are identical to each other in primary sequences. Their predicted tertiary structures conform to those of typical R2R3-MYB domains reported.

Number of paralogous members of BnTT2 *gene family in* B. napus *genome:* In Southern blot, *Eco*RI and *Eco*RV both yielded 3 distinct hybridization bands, while 2 bands were detected by *Hin*dIII or *Xba*I digestion. All the 3 *BnTT2* members have no cutting sites of these 4 enzymes. It is postulated that there are most probably 3 genes orthologous to *AtTT2* in the genome of *B. napus*.

Transcription levels of the 3 BnTT2 *members in various organs of* B. napus: In 30-cycle detection using primer pair FBnTT2A/RBnTT2A, the highest overall expression of *BnTT2* family was observed in 20 DAF seed, followed by 30 DAF seed and 10 DAF seed, then by flower and root, and very weak in bud. 35 cycles of amplification revealed that *BnTT2-1* has the highest expression in 20 DAF seed, followed by flower, seed of other stages, pericarp, bud and leaf. Very weak expression was also observed in root, hypocotyl and cotyledon. Only stem did not show obvious expression of *BnTT2-1*. Among the 3 members, *BnTT2-2* is the most tissue-specific one with expression pattern most like *AtTT2*. Revealed by 35-cycle amplification, it is dominantly and intensively expressed in 20 DAF seed, 10 DAF seed and root. Expression of *BnTT2-2* was also observed in bud, 30 DAF seed, hypocotyl, cotyledon and flower, but the levels are extremely low. *BnTT2-3* could only be detected with weak bands by 40 cycles of amplification. It shows little tissue-specificity, with very weak expression in all the tested 11 organ tissues.

Comparison of BnTT2 *gene sequences and expression levels between L1 and L2:* Genomic sequences of *BnTT2-1*, *BnTT2-2* and *BnTT2-3* were cloned from L1 and L2 and no substantial difference was found between L1, L2 and 5B. 30 cycles of RT-PCR showed distinct *BnTT2* bands only in 10 DAF seed and 20 DAF seed, and little difference was found between the L1 and L2. In 35-cycle RT-PCR, all the 5 reproductive organ tissues showed amplification bands, and no drastic difference was observed between the 2 lines. It is speculated that the yellow seed colour of L2 may be caused by other locus/loci.

Discussion

The TT2 *gene was monogenic in* Brassicaceae *ancestor and probably not triplicated in* Brassica *after* Brassica-Arabidopsis *split:* "Diploid" *Brassica* species are likely derived from hexaploid ancestry, and it is suggested that in *B. napus* there might exist about 6 ortholog genes corresponding to each gene from *A. thaliana* (Lysak et al., 2005). Here 3 highly homologous *BnTT2* genes were isolated, and Southern blot also demonstrated about 3 *BnTT2* genes/alleles. Since *BnTT2-3* can be assumed of an in–*napus* origination, it is possible that both *B. rapa* and *B. oleracea* contain only one *TT2* gene and the *Brassicaceae* ancestor was monogenic at *TT2* locus.

BnTT2 *genes are orthologs of* AtTT2*:* Alignments, BLASTs and phylogenetic analysis all showed that *BnTT2* genes share highest homologies to *AtTT2*. They share nearly identical R2R3-MYB domains with AtTT2, with identical residues presumably involved in MYB definition, DNA binding and bHLH interaction, implying that they may have identical target

DNA binding ability and interact with the same bHLH protein. The C-terminal regions of MYB proteins are typically featured with very low conservation. The C-terminal regions of BnTT2 proteins and AtTT2 share identical positions and very similar sequences of NLS, two α-helices presumably involved in transcriptional activation (Nesi et al., 2001), and a conserved bA-motif. This implies that they have the same nuclear-location site, transcriptional activation ability, and unknown functional features defined by the bA-motif etc. Orthology analysis by OrthoMCL showed that BnTT2 proteins are most similar to AtTT2.

Notable possible cis regulation mechanisms of BnTT2 *expression:* BnTT2-2PM is surely of no transcription activation ability since it lacks the entire C-terminal region. *BnTT2-2* may play a leading role among the 3 members. But overexpression of a transcription factor is unnecessary and may be harmful, so *BnTT2-2PM* may represent a kind of post-transcriptional fine regulation, i.e. limiting total "efficient" mRNAs to an adequate level, or even translating premature proteins to competitively bind target DNA for transiently negative regulation. *BnTT2-1* and *BnTT2-3* have an alternative transcription initiation site A_{12}, and the 3 members also have an alternative poly(A) tailing site. Whether selective length of UTRs is a kind of active cis regulation in *BnTT2* expression, or just a result of non-stringent action, deserves further study. The 5′ UTRs of *AtTT2* and the 3 *BnTT2* members share even higher identities than MYB-domain-coding regions. Furthermore, 2 conserved sites were also found in their 3′ UTRs. These strongly suggest roles of them in cis regulation. The double-start-codon of *BnTT2* genes is as if elaborately evolved, and may serve to promote translation efficiency.

Fast divergence of tissue specificity of TT2 *orthologs and paralogs:* In 5B, all the 3 members of *BnTT2* gene family are expressed mainly in reproductive organ tissues, with the highest in seed. This pattern is quite similar to that of *AtTT2*. But *BnTT2* is also expressed in root, which is the largest difference between *BnTT2* and *AtTT2*. Whether the root expression of *BnTT2* is engaged in synthesis of PAs or other compounds, and what kind of functions they play in *B. napus* root, need to be elucidated. Furthermore, *BnTT2-1* and *BnTT2-3* also have low expression in organs far beyond those for *AtTT2*. Since BnTT2 proteins are predicted as counterparts of AtTT2, the broader expression of *BnTT2* can be better explained as fast divergence of tissue specificity of orthologous *TT2* genes across relative species. Though *BnTT2* members show very high sequence similarities to each other, they show distinct differences in transcription patterns. Concerning tissue specificity and transcription levels, *BnTT2-2* is most similar to *AtTT2* and may play a leading role, *BnTT2-1* shows lowered levels, while *BnTT2-3* is the lowest. This can be explained as fast functional division of different members or fast functional degradation of certain redundant members.

References

Heneen W.K. & Brismar K. (2001) Maternal and embryonal control of seed colour by different *Brassica alboglabra* chromosomes. Plant Breed. 120: 325-329.

Debeaujon I., Nesi N., Perez P., Devic M., Grandjean O., Caboche M. & Lepiniec L. (2003) Proanthocyanidin-accumulating cells in *Arabidopsis* testa: regulation of differentiation and role in seed development. Plant Cell 15: 2514-2531.

Nesi N., Jond C., Debeaujon I., Caboche M. & Lepiniec L. (2001) The *Arabidopsis TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell 13: 2099-2114.

Mehrtens F., Kranz H., Bednarek P. & Weisshaar B. (2005) The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. Plant Physiol. 138: 1083-1096.

Stracke R., Werber M. & Weisshaar B. (2001) The *R2R3-MYB* gene family in *Arabidopsis thaliana*. Curr. Opin. Plant Biol. 4: 447-456.

Heine G.F., Hernandez M.J. & Grotewold E. (2004) Two cysteines in plant R2R3 MYB domains participate in REDOX-dependent DNA binding. J. Biol. Chem. 279: 37878-37885.

Zimmermann I.M., Heim M.A., Weisshaar B. & Uhrig J.F. (2004) Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. Plant J. 40: 22-34.

Lysak M.A., Koch M.A., Pecinka A. & Schubert I. (2005) Chromosome triplication found across the tribe *Brassiceae*. Genome Res. 15: 516-525.