Title	Molecular Genetic Study on Tocopherol Synthesis in Soybean Seeds
Author(s)	朴, 哲旴
Degree Grantor	北海道大学
Degree Name	博士(農学)
Dissertation Number	甲第13930号
Issue Date	2020-03-25
DOI	https://doi.org/10.14943/doctoral.k13930
Doc URL	https://hdl.handle.net/2115/87730
Type	doctoral thesis
File Information	Park_cheol_woo.pdf

Molecular Genetic Study on Tocopherol Synthesis in Soybean Seeds
（ダイズ種子におけるトコフェロール生合成に関する分子遺伝学的研究）

Hokkaido University Graduate School of Agriculture
Division of Agrobiology Doctoral Course

Cheolwoo Park

Contents

Chapter 1 Introduction

Soybean

The soybean [*Glycine max* (L.) Merr.] is an important crop as a major source of oil, protein, starch, dietary fiber, minerals, and vitamins. In East and Southeast Asia, it has been used in various traditional foods, such as soymilk, soy sauce, soy paste, tempeh, miso, tofu, and natto (Wilson, 1995). Nowadays, it is cultivated worldwide and mainly processed into soybean oil and meal (residue after oil extraction) which are used as materials in the production of edible oil, biodiesel, cosmetics and feed (Bandillo et al., 2015). About 55 % of produced soybean oil is used as cooking and salad oil, 24 % as baking and frying fats and oils, 4 % as an ingredient in margarines, 7 % for other food and industrial uses, and 11 % as a substrate for biodiesel production (Wisner, 2012; Medic et al., 2014). The largest soybean producer in 2018 is the USA, with 34 % share in the world's production, followed by Brazil (32 %), Argentina (15 %), China (4 %), Paraguay (3 %), India (3 %), Canada (2 %), and others (6 %). Compared to the early 1960s, the soybean cultivation area has increased about 4 times and the yield per unit area has more than double (Masuda et al. 2009). Consequently, the world's gross output of soybean was increased from 28.6 million metric tons in 1961-65 to 337.48 million metric tons at 2019-20. It reflects the demand to accommodate the needs for food, feedstuff and fuel of the growing world population.

Tocopherol

Tocopherols (the vitamin E family) are lipophilic antioxidants that prevent the oxidation of unsaturated fatty acids. There are four isoforms of tocopherols (Toc), α-, β-, γ-, and δ-Toc, of which α-Toc has the highest vitamin E activity in humans because of its highest affinity with the hepatic tocopherol transfer protein (Grusak and DellaPenna, 1999; Bramley et al., 2000). As well as vitamin E activity, α-Toc also plays a role in the prevention of aging-related diseases such as cardiovascular diseases and cancer (Grusak and DellaPenna, 1999; Bramley et al., 2000) .

Tocopherol biosynthesis in plants

Tocopherol biosynthesis in plants is well characterized (Fritsche et al., 2017; Mène-Saffrané and Pellaud, 2017) (Fig. 1). Fusion between the aromatic head of homogentisic acid (HGA) and the polyprenyl side chain of homogentisate phytyltransferase creates 2-methyl-6-

phytyl-1,4-benzoquinol (MPBQ) which is further methylated by MPBQ methyltransferase (MPBQ-MT) to 2,3-dimethyl-6-phytyl-1,4-benzoquinone (DMPBQ). MPBQ and DMPBQ are converted to γ-Toc and δ-Toc, respectively, through cyclization of the HGA head by tocopherol cyclase. The final step in the tocopherol biosynthesis pathway is the conversion of γ-Toc and δ-Toc to α-Toc and β-Toc, respectively, by γ-tocopherol methyltransferase (γ-TMT). MPBQ-MT and γ-TMT are crucial in determining the seed tocopherol composition. The γ-TMT activity is reflected in α/γ ratio, a ratio of α-Toc content to γ-Toc content, whereas the MPBQ-MT activity is reflected in (α + γ)/total ratio, a ratio of the sum of α-Toc and γ-Toc contents to the total Toc content (Fig. 1). In *Arabidopsis*, MPBQ-MT and γ-TMT are encoded by *VTE3* and *VTE4*, respectively (Van Eenennaam et al., 2003; Bergmüller et al., 2003). The *VTE4* overexpression in soybean seeds was reported to increase the α-Toc ratio by up to 70 % (Van Eenennaam et al., 2003). *VTE3* and *VTE4* co-expression further increased the α-Toc ratio by up to 90 % and decreased both the δ-Toc and γ-Toc ratios in soybean seeds (Van Eenennaam et al., 2003).

Tocopherol contents in soybean oil

The soybean oil has a relatively high total tocopherol content (1884 mg/L), compared with other oilseed crops, maize (1511 mg/L), *Brassica napus* (1240 mg/L) and Sunflower (1408 mg/L) (Li et al., 2019). However, the most predominant form is γ-Toc. The α-Toc content is typically less than 10 % that of the total Toc content (Grusak and DellaPenna, 1999; Bramley et al., 2000). Considering that soybean is a major oil source providing 30 % of the total worldwide oil consumption, increasing the seed α-Toc content may open up opportunities for new food and industrial uses of soybean.

Tocopherol research in soybean

Soybean cultivars have variable seed Toc contents and compositions (Ujiie et al., 2005; Carra˜o-Panizzi et al., 2007; Dwiyanti et al., 2007; Seguin et al., 2009). The genetic and molecular bases underlying this natural variation have been extensively studied (Li et al., 2010; Liu et al., 2010; Dwiyanti et al., 2011; Shaw et al., 2016; Shaw et al., 2017). Dwiyanti et al. (2011) detected a major quantitative trait locus (QTL) for α-Toc ratio in chromosome (Chr) 9 which accounted for 55 % of the phenotypic variation in a recombinant inbred line (RIL) population of a cross between a Japanese standard soybean cultivar Ichihime (α-Toc ratio < 10%) and a high α-Toc cultivar from Eastern Europe, Keszthelyi Aproszemu Sarga (KAS; α-Toc ratio > 20%). The QTL region contained a γ-tocopherol methyltransferase gene designated as *γ-TMT3*, which showed higher expression in developing seeds from RILs having the KAS allele at the QTL than those having the Ichihime allele. A β-glucuronidase reporter-aided analysis of *γ-TMT3* further confirmed that the promoter of *γ-TMT3* from KAS had higher activity than that from Ichihime. The higher promoter activity is likely caused by single nucleotide polymorphisms (SNPs) in *cis*-regulatory elements, such as MYBCORE and CAAT box motifs, in the promoter (Dwiyanti et al., 2011). Based on these results, Dwiyanti et al. (2011) suggested that the use of *γ-TMT3* with high promoter activity from KAS could be a means of improving the α-Toc content in soybean seeds. The function of γ-TMT3 as a methyltransferase was confirmed by the catalytic assay of purified enzyme heterologously expressed in *Escherichia coli* (Tewari et al., 2017). Soybean has additional two tightly linked *γ-TMT* genes, *γ-TMT1* and *γ-TMT2*, on Chr12 (Li et al., 2010). The three TMT proteins (γ-TMT1, γ-TMT2, and γ-TMT3) exhibit high amino acid similarities of 90.5 %–94.4 % with each other. The overexpression of *γ-TMT2* increased the seed α-Toc content 3–4.5 fold and 4–6 fold relative to the wild type in maize (*Zea mays*) and *Arabidopsis* (Zhang et al., 2013), respectively, supporting that it possessed the function of γ-TMT. Based on plastid transit peptide prediction (Emanuelsson et al., 1999), only γ-TMT2 possesses a plastid transit peptide signal, suggesting that γ-TMT2 functions in the plastid (Dwiyanti et al., 2011). However, the intracellular localization and function for the other two TMT proteins remain undetermined.

Purposes of this study

Increasing α-Toc in soybean seed is an important breeding target of soybean seed quality. But molecular mechanisms of tocopherol biosynthesis are not fully identified. The purposes of the present study were first to determine the molecular genetic basis for an elevated α-Toc ratio detected in a wild soybean accession, and second to characterize the thermal responses of tocopherol biosynthesis and determine the association with DNA polymorphisms in the three *γ-TMT* genes.

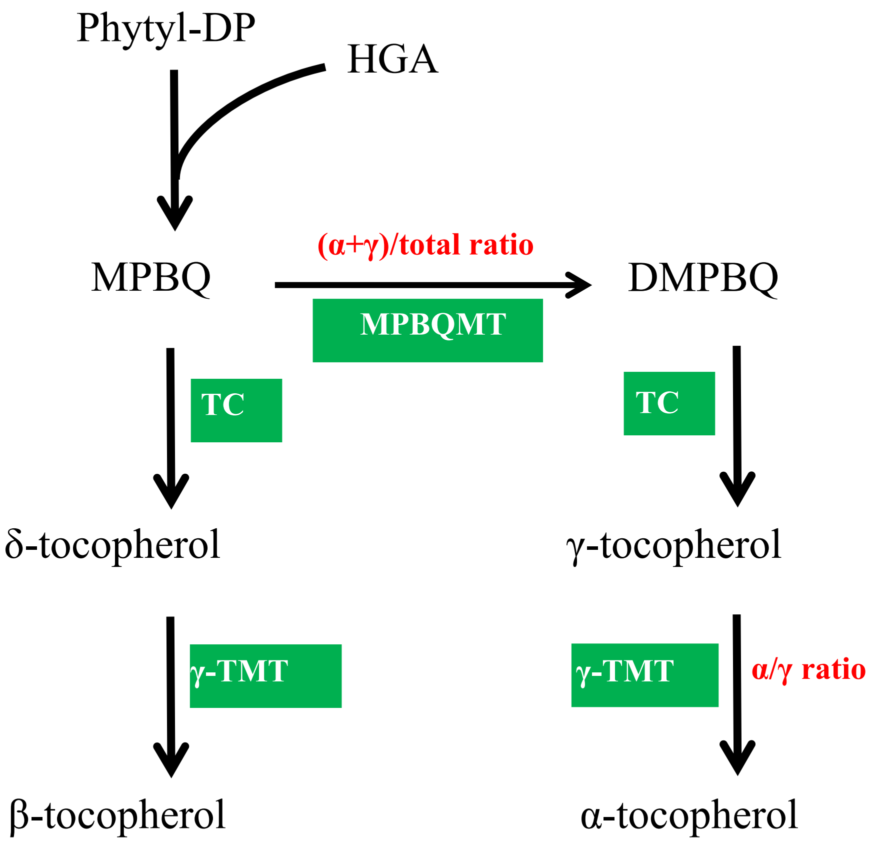


Fig. 1 Tocopherol biosynthesis pathway.

Enzymes are shown in green boxes. Phytyl-DP, phytyl-diphosphate; HGA, homogentisic acid; MPBQ, 2-methyl-6-phytyl-1,4-benzoquinol; DMPBQ, 2,3-dimethyl-6-phytyl-1,4-benzoquinol; MPBQMT, 2-methyl-6-phytyl-1,4-benzoquinol methyltransferase; TC, tocopherol cyclase; γ-TMT; γ-tocopherol methyltransferase.

Chapter 2 Identification of quantitative trait loci for increased α-tocopherol biosynthesis in wild soybean using a high-density genetic map

2.1 Background and Purposes

The wild soybean (*Glycine soja*) is a huge reservoir of potentially useful variants for the improvement of soybean cultivars. To date, it has been used to improve yield, stress tolerances, disease resistances, and nutritional components of seeds in soybean breeding (Hajika et al., 1996; Kikuchi et al., 1999; Concibido et al., 2003; Fukuda et al., 2005; Kanamaru et al., 2008; Kim et al., 2010; Wang et al., 2007; Xu et al., 2012; Dwiyanti et al., 2016). Based on a survey on 528 wild soybean accessions collected from various regions of Japan and South Korea, Dwiyanti et al. (2016) discovered 11 accessions with high α-Toc ratios. Sequencing analyses of the promoter and 5′-untranslated region (UTR) of *γ-TMT3* classified the 11 accessions into four haplotypes, of which one was identical to the *γ-TMT3* sequence of KAS. A molecular genetic study of the high α-Toc ratio of wild accessions, particularly those with novel *γ-TMT3* promoter haplotypes, may therefore be useful in broadening the genetic diversity of α-Toc biosynthesis in soybean.

In this chapter, I identified QTLs for a high α-Toc trait detected in a wild accession B04009, and discussed the candidates for the QTLs, based on the results obtained from sequencing analyses.

2.2 Material and Methods

2.2.1 Plant materials

The soybean breeding line TK780 and a wild accession (B04009) were used in this study. TK780 is an early-flowering line with a low seed α-Toc content and ratio. B04009, a wild soybean accession collected in Yamanashi Prefecture, Japan, has a high seed α-Toc content and ratio (Dwiyanti et al., 2016). A RIL population of 94 F_8 (2016) and F_9 (2017) lines was developed by a single seed descendent method from an F_2 population of the cross between TK780 and B04009. RILs and parents were grown in pots under short-day greenhouse conditions where air temperatures were set to 25°C with a fluctuation from 22 °C to 28 °C from February to May in 2016 and 2018. Lights were supplied with high intensity discharge lamps (HONDA-T; Panasonic Co., Osaka, Japan) in the daytime. Seeds were harvested as a bulk from three to four plants in each line and dried in a desiccator until required for the assay. Parents were also grown at 25 °C in the greenhouse, and after flowering were transferred to growth chambers set to 20 ºC and 30 ºC; seeds were harvested individually from two to five plants.

2.2.2 Tocopherol quantification

Tocopherol contents of seeds were quantified according to Dwiyanti et al. (2011). Ten fully dried seeds for each line were ground to a fine powder by Multi-Beads Shocker (MB755U(S); Yasui Kikai Co, Osaka, Japan). Twenty mg of seed powder was thoroughly mixed with 500 μL of cold 80 % ethanol (4 °C) containing 5 μL dl-Tocol solution (10 μg/ml; Tama Biochemical Co. Ltd., Tokyo, Japan) as an internal standard. After sonication for 10 min, the mixture was thoroughly mixed with 1,000 μL of hexane with added pyrogallol as an antioxidant. The mixture was sonicated again for 10 min, and then centrifuged for 5 min at 18,900 × *g* at 4 °C. The supernatants were analyzed by high-performance liquid chromatography (LaChrom Elite, Hitachi High-Technologies Corp., Tokyo, Japan) using a reverse-phase column (Inertsil ODS-3, 3.0 mm × 250 mm; GL Sciences, Tokyo, Japan) with methanol : acetonitrile (10 : 90 v/v) as a mobile phase at flow rate 0.5 mL/min at a constant temperature of 40 °C. Tocopherol isoforms were detected at 295 nm with one peak for dl-Tocol and three peaks for δ-Toc, the sum of γ-Toc and β-Toc, and α-Toc, in order of retention time. The content of each isoform was calculated with the ratio of the peak area against that of dl-Tocol. In this study, the content calculated from the sum of γ-Toc and β-Toc was considered the γ-Toc content because the actual β-Toc content is very low in soybean seeds (Ujiie et al., 2005). The assay was carried out in triplicate.

2.2.3 Genotype data generation

Total DNA of RILs was extracted from young leaves using the modified CTAB method (Doyle and Doyle, 1987). DNA was digested using the restriction enzymes BglII and EcoRI to create a DNA library for double digest restriction site-associated DNA sequencing (Baird et al., 2008; Peterson et al., 2012). Sequencing was performed with 51 bp single-end reads in one lane of a HiSeq2000 Sequencer (Illumina, San Diego, CA, USA) by Macrogen (Seoul, South Korea). The resulting reads were trimmed with Trimmomatic ver 0.33 (Bolger et al., 2014) using the following parameters: LEADING:19, TRAILING:19, SLIDINGWINDOW:30:20, AVGQUAL:20, and MINLEN:51. These RAD-Seq procedures were carried out by Clockmics, Inc. (Izumi, Osaka, Japan). The trimmed reads were mapped to the soybean reference genome Williams82.v2 using Bowtie2 (Langmead et al., 2012) with a default parameter setting. SNP calling was performed using GATK-Unified Genotyper (McKenna et al., 2010). Imputation of missing genotypes in RILs using parental SNP data was performed using Beagle 4.0 (Browning and Browning, 2007). Filtering for monomorphic SNPs and SNPs having many missing calls was performed using TASSEL.5.2.31 with following parameters: minimum call rate per SNP 90 % and minimum allele frequency 0.05 (to remove monomorphic SNPs) TASSEL (v 5.2.31) (Bradbury et al., 2007). Using custom script in R (RC Team, 2018), the nucleotide information was converted to AB genotype with parentA=TK780 and parentB=B04009. All heterozygous genotypes were converted as missing alleles. Further filtering for duplicated markers or markers having switch alleles was performed in R/QTL (Broman et al., 2003) resulting in a final set of 7,710 SNPs.

2.2.4 Linkage group construction and QTL mapping

QTL IciMapping ver 4.1 (Wang et al., 2016) was used to construct a linkage map with 7,710 SNP markers. The Input algorithm, which re-estimates recombination frequency and genetic distance without changing the marker order in the input file, was used to determine the order of markers on the genetic map. The sum of adjacent recombinant frequencies with a window size of 5 was used as a rippling criterion for fine tuning of the markers. Recombination frequencies between linked loci were transformed into centimorgan (cM) distances using Kosambi's mapping function (Kosambi, 1943). Linkage map was drawn using MapChart (Voorrips, 2002).

QTL analyses were performed with the inclusive composite interval mapping of additive QTLs implemented in QTL IciMapping (Wang et al., 2016). The permutation test was performed to determine the threshold of significant QTLs. Based on the result, LOD scores more than 3.576 were used as criteria to delineate the significance of QTLs ($p < 0.05$). The supporting intervals of QTLs passing threshold were defined by ICIMapping as leftCI = left border of confidence interval and rightCI = right border of confidence interval.

2.2.5 Construction of *γ-TMT* genes based on whole genome resequencing data

Raw reads of TK780 and B04009 from next-generation sequencing Illumina Hiseq XTen were aligned to the soybean reference genome Williams82.a2 (Schmutz et al., 2010). The alignment was performed using Bowtie2-2.2.9 (Langmead and Salzberg, 2012). The resulting alignment was further processed to remove duplicate reads and to fix mate information using

Picard tools (http://broadinstitute.github.io/picard). The Genome Analysis Toolkit (GATK ver 3.8; McKenna et al., 2010) was used to realign small indels. Subsequently, variants (SNP and indels) were called using the GATK Unified Genotyper function which filtered out reads having mapped base quality Phred scores below 20. Using the reference genome Williams82.a2 and SNP dataset of each variety, sequences of *γ-TMT* genes were reconstructed by the FastaAlternateReferenceMaker function available in GATK.

2.2.6 *Cis*-element prediction

New PLACE, a Plant *Cis*-acting Regulatory DNA Elements database (http://www.dna.affrc.go.jp/PLACE/) (Higo et al., 1999) was used to predict *cis*-acting regulatory elements present in the promoter region (2,000-bp upstream of the start codons) of *γ-TMT1*, *γ-TMT2*, *γ-TMT3*, and genes for S-adenosyl-L-methionine-dependent methyltransferases superfamily protein, RING-H2 finger C2A and GATA type zinc finger transcription factor family protein in B04009 and TK780.

2.3 Results

2.3.1 Tocopherol contents and ratios of parental lines under different thermal conditions

Toc contents and compositions were compared between seeds matured at 20 °C and 30 °C. TK780 and B04009 had different characteristics for seed Toc biosynthesis. TK780 produced seeds with tocopherol contents approximately two-fold higher than B04009, irrespective of the temperature during seed development (Table 1). B04009 produced more α-Toc than TK780 both in 20 °C ($t = 8.36$, $p = 1.6 \times 10^{-4}$) and in 30 °C ($t = 5.71$, $p = 1.3 \times 10^{-3}$) despite its lower total Toc content. The α-Toc content elevation in B04009 was associated with increments of both the (α + γ)/total ratio (ratio of the sum of α-Toc and γ-Toc contents to the total Toc content) and the α/γ ratio (ratio of the α-Toc content to the γ-Toc content), reflecting the extent of conversions from MPBQ to DMPBQ and from γ-Toc to α-Toc, respectively. α-Toc contents increased as temperatures increased in both B04009 ($t = 7.57$, $p = 1.3 \times 10^{-4}$) and TK780 ($t = 3.80$, $p = 0.032$), as reported in previous studies (Almonora et al., 1998; Britz et al., 2002; Britz et al., 2008; Seguin et al., 2010; Carrera et al., 2011; Chennupati et al., 2011; Dwiyanti et al., 2016). The (α + γ)/total and α/γ ratios also increased as temperatures rose, but this was particularly marked in the α/γ ratio of B04009 which increased 5.5-fold more at 30 ºC than at 20 ºC (Table 1).

2.3.2 Variation of α-tocopherol contents and ratios in RIL populations

Ninety-four RILs were developed by a single-seed descent method from the F_2 population of the cross between TK780 and B04009. The two parents differed in flowering habits; under natural daylength (ND) conditions of Sapporo (43°07′N, 141°35′E). TK780 flowered in the middle of July whereas B04009 flowered in late September. The flowering time of RILs also varied widely within the range of parents' flowering time under ND conditions (data not shown). To reduce the variation induced by different thermal conditions associated with flowering and maturing times, the RILs were grown under short-day conditions in a greenhouse where the air temperature was controlled at 25 ºC.

The seed Toc contents in the RILs varied continuously in a range of parental values in δ-Toc, γ-Toc, and total Toc contents, but the α-Toc content was slightly over the parental values in the two years tested (2016 and 2017) (Fig. 2). The α-Toc contents ranged from 7–115 μg/g in 2016 (24 μg/g seed in TK780, 62 μg/g seed in B04009) and from 9–91 μg/g in

2017 (32 μg/g seed in TK780, 46 μg/g seed in B04009). The correlation coefficients between years were significant in all contents ($p < 0.01$); the highest was seen in the α-Toc content ($r = 0.772$), suggesting that the biosynthesis of α-Toc was relatively stable compared with that of other tocopherol contents (Table 2). The α-Toc contents did not correlate with δ-Toc and γ-Toc contents in either year (Additional file 1B), but weakly correlated with total Toc contents; the correlation coefficient ($r = 0.292$) was significant only in 2016 ($p < 0.01$; Fig. 3A and 3B). In contrast, there were strong positive correlations ($r = 0.747$ to 0.931) among δ-Toc, γ-Toc, and total Toc contents (Table 3).

Next, the variation of (α + γ)/total and α/γ ratios in the RIL population was examined. The RILs varied continuously almost in a range of parental values with respect to the (α + γ)/total ratio, whereas some lines showed higher α/γ ratios than B04009 (Fig. 4). Both ratios showed significant ($p < 0.01$) positive correlations between years, and the correlation coefficient was higher in the α/γ ratio ($r = 0.768$) than the (α + γ)/total ratio ($r = 0.474$) (Table 2). Both ratios further positively correlated with the α-Toc content in both years; the correlation with the α-Toc content was stronger in the α/γ ratio than the (α + γ)/total ratio (Fig. 3C to 3F). The two ratios also exhibited significantly positive ($p < 0.01$) correlations with each other ($r = 0.541$ in 2016, $r = 0.472$ in 2017). Collectively, these results suggest that the seed α-Toc contents in the RIL population were associated primarily with the α/γ ratio, indicating the extent of conversion from γ-Toc to α-Toc, then the (α + γ)/total ratio indicating conversion from MPBQ to DMPBQ, but only weakly with the total tocopherol production (Fig. 4A and 4B).

2.3.3 Construction of a high-density linkage map

Prior to QTL analysis, a genome-wide SNP-based genetic map was constructed with a total of 7,710 SNPs obtained from restriction site-associated DNA sequencing. The linkage map was constructed using IciMapping software (Wang et al., 2016). The entire length of the linkage map was 3,211.6 cM, and the length of each chromosome ranged from 121.3 cM for the smallest linkage group of Chr9 to 221.1 cM for the largest one of Chr11. The average genetic distance between neighboring SNP markers was 1.4 cM; the largest gap between SNPs in each chromosome ranged from 5 cM in Chr3 and Chr4 to 34.1 cM in Chr11. The gaps were mostly caused by a lack of SNPs available to map between the two lines.

2.3.4 QTLs for Toc contents

QTL analyses were performed with the inclusive composite interval mapping of additive QTLs implemented in QTL IciMapping (Wang et al., 2016). Based on the results of permutation test, a QTL could be considered as significant ($p < 0.05$) if the logarithm of odds (LOD) score exceeded 3.576 for both years. Five significant QTLs were detected in the α-Toc content. They were located in Chr5 (*qαTC-5*), Chr9 (*qαTC-9*), Chr11 (*qαTC-11*), and Chr12 (*qαTC-12*) in 2016, and in Chr9 (*qαTC-9*) and Chr12 (*qαTC-12*) in 2017 (Table 4 and Fig. 5). Additionally, a QTL with a LOD score of 3.2 was detected in Chr11 (*qαTC-11*) in 2017, although the LOD value was below the threshold of significance. The map positions for the QTLs in Chr9, Chr11, and Chr12 were identical or almost the same between the two years, suggesting that the effects on the α-Toc content were caused by respective identical QTLs. The B04009 allele increased the α-Toc contents at *qαTC-9* and *qαTC-12* but decreased the contents at *qαTC-5* and *qαTC-11*. Of these, *qαTC-9* had the highest LOD scores (14.3 in 2016, 13.1 in 2017) with the largest additive effect on the α-Toc content. Collectively, the QTLs detected accounted for 56.4 % (2016) and 54.2 % (2017) of the phenotypic variation observed in the RIL population.

Three QTLs (*qδTC-4*, *qδTC-6*, and *qδTC-19*) were detected in the δ-Toc content only in 2017. The wild allele increased the δ-Toc content at *qδTC-4* and *qδTC-6* and decreased it at *qδTC-19*. Of these, *qδTC-19* exhibited the highest LOD score (14.1), solely accounting for 32 % of the phenotypic variation in the RIL population.

Three QTLs (*qγTC-6*, *qγTC-13*, and *qγTC-14*) were detected in the γ-Toc content in 2017. The wild allele increased the γ-Toc content at both QTLs, which collectively accounted for 20.2 % of the whole variation. One QTL for γ-Toc content in 2016 was detected in Chr7.

Only one QTL (*qTTC-1*) was detected in the total-Toc content in 2016.

Taken together, the four QTLs detected in the α-Toc content did not overlap with those for δ-Toc, γ-Toc, and total-Toc contents, consistent with the absence of correlation between the α-Toc content and the other contents (Fig. 3A and 3B, Table 3).

2.3.5 QTLs for Toc ratios

QTL analysis was also performed for (α + γ)/total and α/γ ratios. Significant QTLs were detected in only the α/γ ratio: four QTLs in Chr9 (*qαγR-9*), Chr11 (*qαγR-11*), and Chr12 (*qαγR-12.1* and *qαγR-12.2*) in 2016 and four QTLs in Chr9 (*qαγR-9*), Chr11 (*qαγR-11*), Chr12 (*qαγR-12.1*), and Chr17 (*qαγR-17*) (Table 4, Fig. 5). Because they had identical or almost the same map positions in the two years, the three QTLs, *qαγR-9*, *qαγR-11*, and *qαγR-*

12.1, consistently controlled the α/γ ratio. The B04009 allele positively controlled the α/γ ratios at *qαγR-9* and *qαγR-12.1* but negatively at *qαγR-11*, *qαγR-12.2*, and *qαγR-17*. *qαγR-9* had the highest LOD scores in both years (18.2 in 2016, 21.1 in 2017). Collectively, the four QTLs accounted for 63.3 % (2016) and 68.2 % (2017) of the whole variation detected in the RIL population. The SNPs flanking *qαγR-9*, *qαγR-11*, and *qαγR-12.1* were identical to or nearby those of QTLs for α-Toc contents (*qαTC-9*, *qαTC-11*, and *qαTC-12*), suggesting that they controlled both the α/γ-Toc ratio and α-Toc contents.

2.3.6 Additive effects of three QTLs on α/γ ratio

Next, the additive effects were evaluated for the major three QTLs (*qαγR-9*, *qαγR-11*, and *qαγR-12.1*). RILs were classified into eight genotypic classes based on the flanking SNPs at the QTLs, and mean α/γ ratios were compared among genotypes. As shown in Fig. 6, alleles from B04009 at *qαγR-9* and *qαγR-12.1* and those from TK780 at *qαγR-11* each additively increased the α/γ ratios in both years.

2.3.7 Candidate genes for QTLs for α-Toc biosynthesis

The Williams 82 genome sequence (Phytozome v12.1/Glycine max Wm82.a2.v1) was surveyed for the genes annotated in the genomic regions of the QTLs in Chr9, Chr11, and Chr12 for α-Toc contents and α/γ ratios. The *qαTC/αγR-9* region contained *γ-TMT3* (Glyma.09G22280, physical position: 44,341,974–44,346,311); this was previously identified as a candidate of the QTL for a high α-Toc ratio in the cross between high and low α-Toc soybean cultivars (Dwiyanti et al., 2011). There was no gene directly involved in tocopherol biosynthesis between the nearest flanking markers of *qαTC/αγR-12.1*, but *γ-TMT2* (Glyma.12G014300.1, physical position: 1,033,151–1,037,054) and *γ-TMT1* (Glyma.12G014200.1, physical position: 1,025,584–1,029,095) were located 703 kbp and 711 kbp, respectively, apart from the flanking marker S12_1740699.

There were 29 genes located within the *qαTC/αγR-11* region between markers S11_31748669 and S11_32039088 (Table 2, Table 5). Of these, 18 genes were expressed in developing seeds and pods (SoyBase; https://soybase.org/). However, no genes were known to be involved in Toc biosynthesis. The candidate gene may be located in the vicinity of QTL. Therefore, I looked into a total of 71 genes in the flanking regions, 35 genes for the upstream region and 36 genes for the downstream region. This survey totally encompassed the region of 400 kbp. Of the selected 100 genes (Glyma.11G219000 to Glyma.11G228900), there were four zinc finger transcription factors, two of which were expressed in seeds and pods; RING-H2 FINGER C2A (Glyma.11G220400) and Znf_GATA (Glyma.11G226400) (Table 5, SoyBase; https://soybase.org/). Additionally, S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (Glyma.11G222800) also existed in the *qαTC/αγR-11* region. This gene was also expressed in seed and pods (SoyBase; https://soybase.org/).

2.3.8 Sequence polymorphism of candidate genes

The sequences of three *γ-TMT* genes were compared between TK780 and B04009. The 1,350 bp sequence in the promoter region of *γ-TMT3* has already been determined in B04009 and shown to differ with respect to 21 SNPs and four indels from KAS (Dwiyanti et al., 2016). TK780 and B04009 possessed the same coding sequence as the soybean reference genome Williams 82 (Glyma.09G222800.1), but there were 13 SNPs in the promoter region, of which 10 were located within known *cis*-elements (Fig.7). TK780 possessed 2SSEEDPROTBANAPA, a *cis*-element conserved in many storage protein gene promoters (Stålberg et al., 1996), and a seed-specific *cis*-element CANBNNAP (Ellerström et al., 1996), whereas B04009 contained MYB1AT and CAATBOX1, which were previously detected as *cis*-elements specific to cultivars with high α-Toc ratios (Dwiyanti et al., 2011) .

A non-synonymous substitution was detected in exon 5 of *γ-TMT2*; the amino acid residue was serine in TK780 in place of threonine in Williams 82 (Glyma.12G014300.1) and B04009 (Fig.8). A total of 46 DNA polymorphisms, 38 SNPs and eight indels were detected in the promoter and introns (Fig.8). Of these, eight polymorphisms were located within known *cis*-elements: B04009 possessed DRE2COREZMRAB17, a *cis*-element for genes expressed during late embryogenesis and induced by abscisic acid (Busk et al., 1997), and SEF4MOTIFGM7S, a *cis*-element bound by soybean embryo factor 4 (Allen et al., 1989).

The coding sequences of *γ-TMT1* in both TK780 and B04009 were identical to that of Williams 82 (Glyma.12G014200.1; Fig.9). Additionally, a total of 17 SNPs and 12 indels were detected in the promoter and introns, of which seven polymorphisms were located within known *cis*-elements; there were 13 B04009-specific and four TK780-specific *cis*-elements, of which PYRIMIDINEBOXHVEPB1 and RYREPEATBNNAPA in B04009 are known to be involved in seed development and function (Cercós et al., 1999).

The sequences of three candidate genes of the QTL in chromosome 11 were also compared between TK780 and B04009.

The coding sequences of a gene for *RING-H2 finger C2A* in both TK780 and B04009 were identical to that of Williams 82 (Glyma.11G220400). Additionally, a total of 9 SNPs were detected in the promoter, 3′UTR and 5′UTR, of which four polymorphisms were located within known *cis*-elements; there were seven B04009-specific and five TK780-specific *cis*-elements. However, there was no *cis*-element known to be involved in seed development and function (Fig.10).

The coding sequences of a gene for *GATA type zinc finger transcription factor family protein* in both TK780 and B04009 were identical to that of Williams 82 (Glyma.11G226400). Additionally, a total of 8 SNPs were detected in the promoter and 3′UTR, of which one polymorphism was located within known *cis*-elements; there were three TK780-specific *cis*-elements, however there is no *cis*-elements known to be involved in seed development and function (Fig.11).

In a gene for S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (Glyma.11G222800.1), there were two non-synonymous substitutions in exon 3 and exon 9. In exon 3, the amino acid residue was serine in TK780 and Williams 82, but cysteine in B04009. In exon 9, the amino acid residue was serine in TK780 and Williams 82 (11G222800.1), but proline in B04009. The amino acid residues in TK780 were conserved in the homoeologous copy (Glyma18G034800.1) and homologs in *Arabidopsis* and other plant species. Therefore, the protein of B04009 might be a missense variant. Additionally, a total of 16 SNPs were detected in the promoter and introns (Fig.12). Of these, 6 polymorphisms were located within known *cis*-elements. However, there were no *cis*-elements known to be involved in seed development and function (Fig.12).

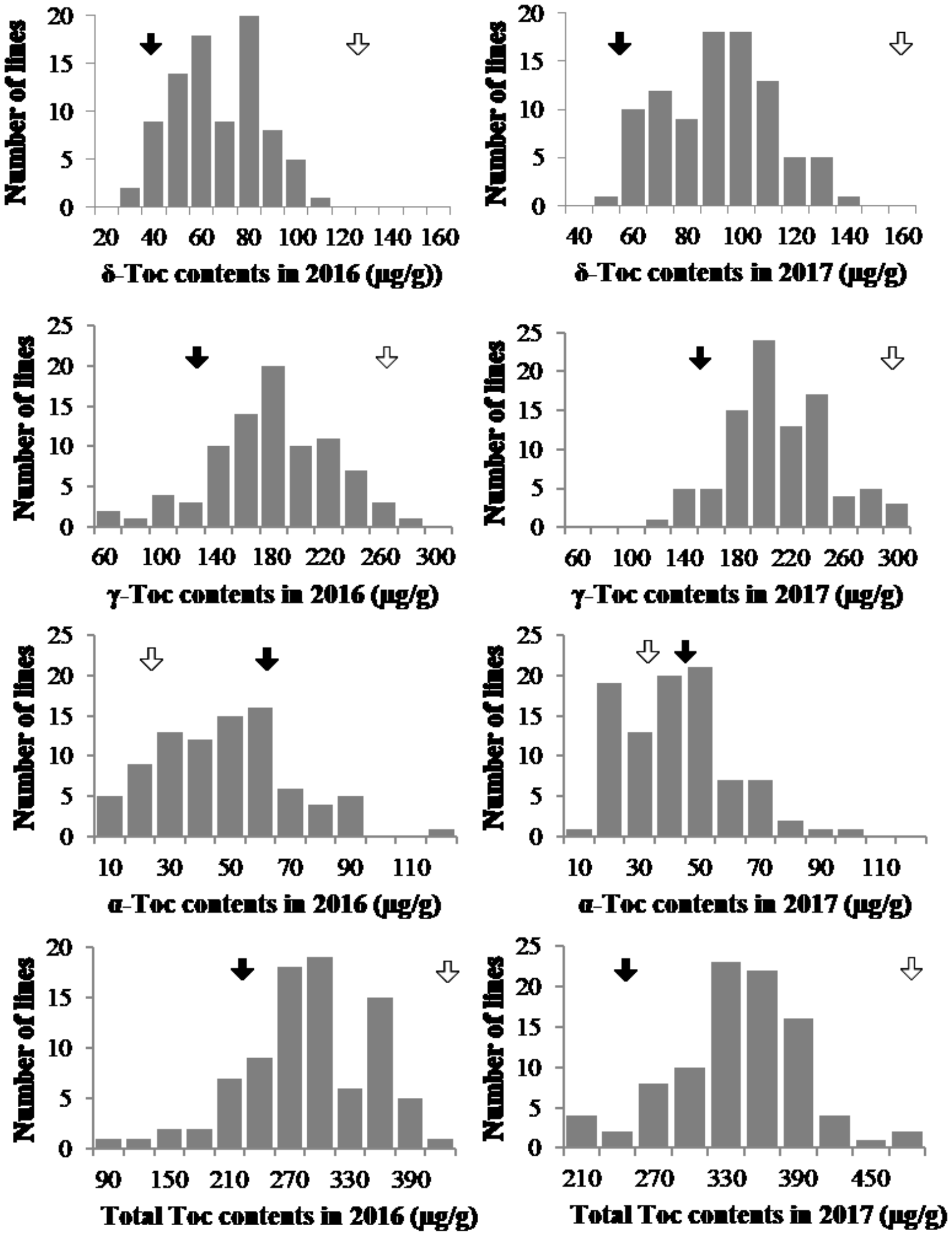


Fig. 2 Tocopherol content variation in seeds in RILs of the cross between TK780 and B04009. Closed arrow; B04009, Open arrow; TK780.

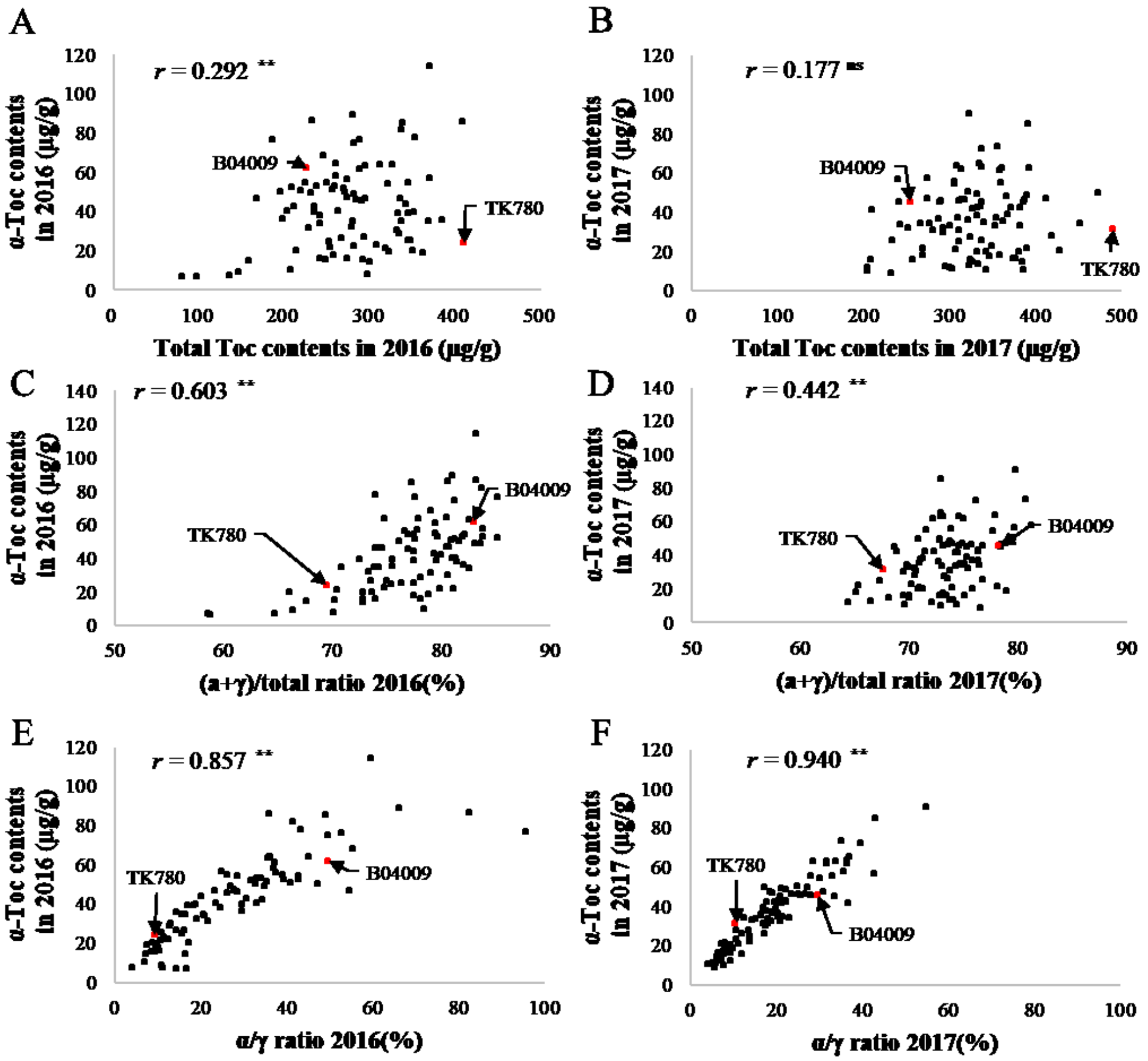


Fig. 3 Scatter diagrams showing relationship between seed α-Toc contents with total tocopherol (Toc) contents and Toc ratios in RILs. Relationship between seed α-Toc contents with total Toc contents in 2016 (A) and with total Toc contents in 2017 (B). Relationship between seed α-Toc contents with the (α + γ)/δ-Toc ratio in 2016 (C) and 2017 (D). (α + γ)/δ-Toc ratio is the ratio of the sum of α- and γ-Toc contents to the δ-Toc content. Relationship between seed α-Toc contents with the α/γ-Toc ratio in 2016 (E) and 2017 (F). α/γ-Toc ratio is the ratio of the α-Toc content to the γ-Toc content.

**, $p < 0.01$; ns, non-significant.

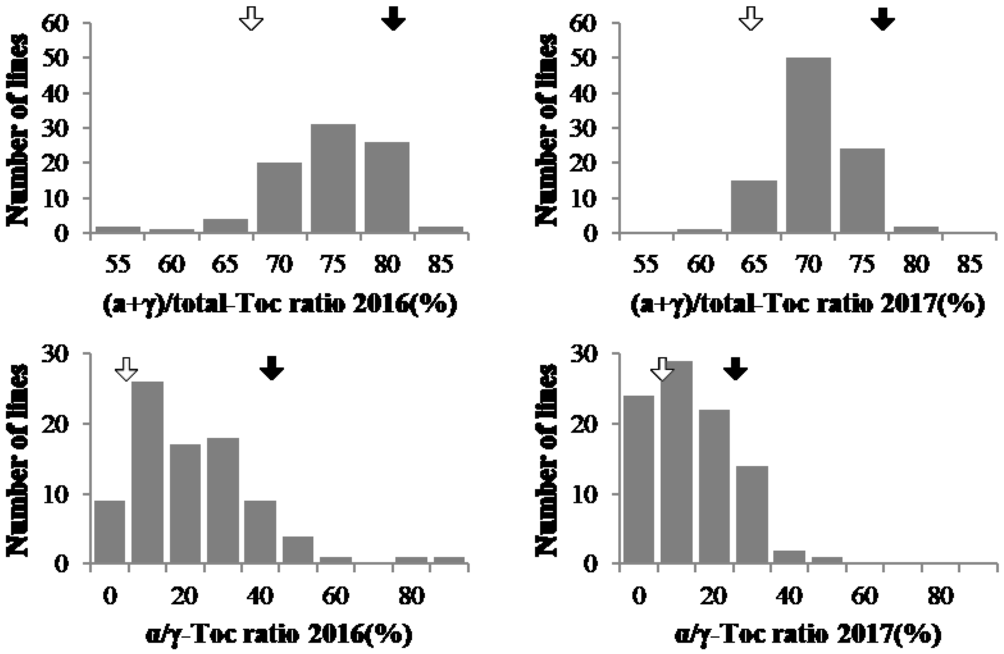


Fig. 4 Variation in tocopherol ratios in seeds in RILs of the cross between TK780 and B04009.

(α + γ)/total ratio; the ratio of the sum of α- and γ-Toc contents to the total Toc content. α/γ ratio; the ratio of the α-Toc content to the γ-Toc content. Closed arrow; B04009, Open arrow; TK780.

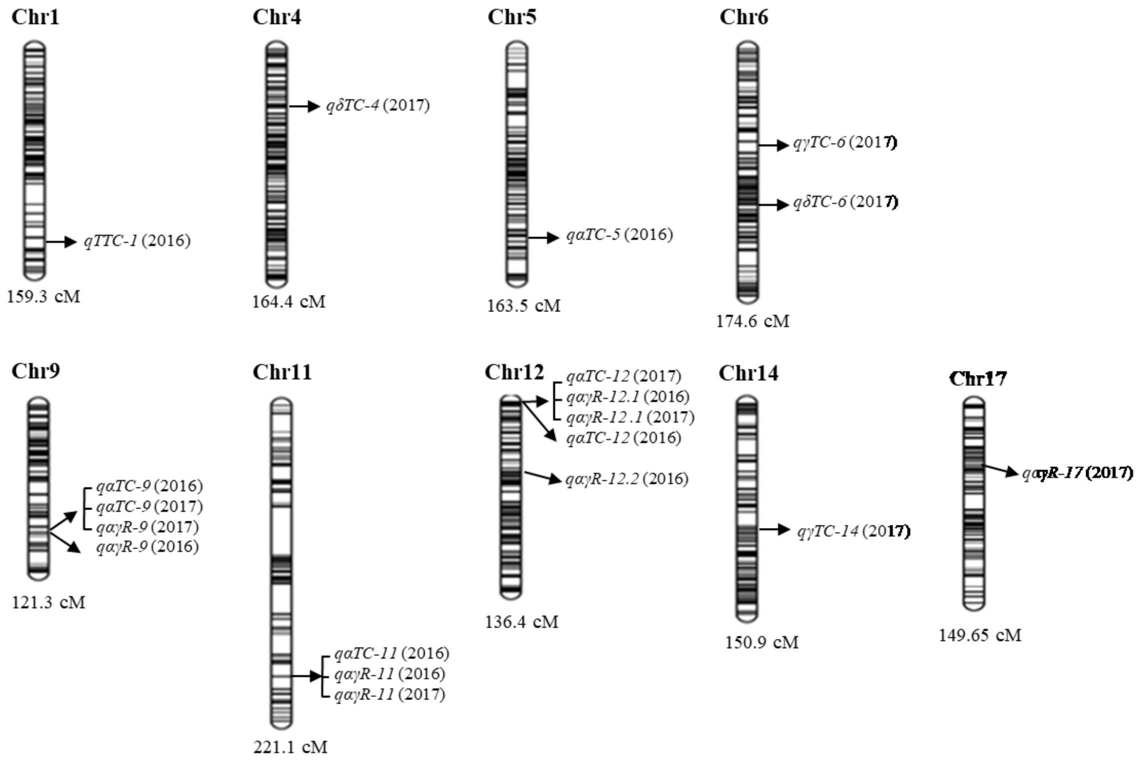


Fig. 5 Location of QTLs for tocopherol contents and ratios in RILs of TK780 and B04009 cross.

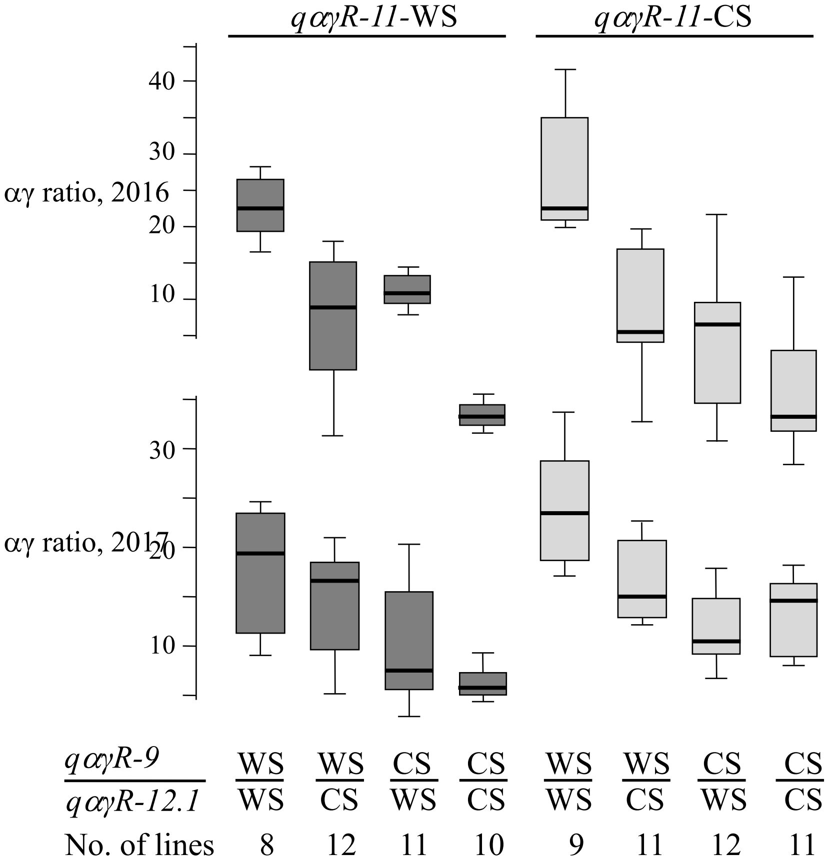
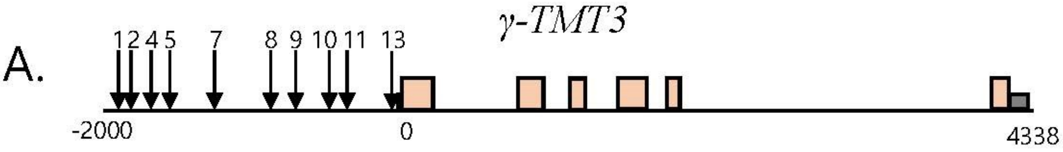
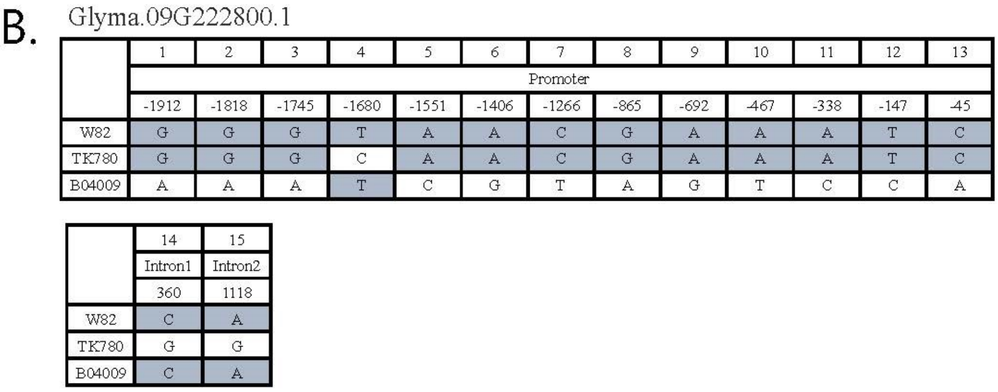


Fig. 6 Additive effects of three QTLs for the seed α/γ ratio in an RIL population of the cross between breeding line TK780 and wild accession B04009.

B04009 and TK780 are designated as WS and CS, respectively.

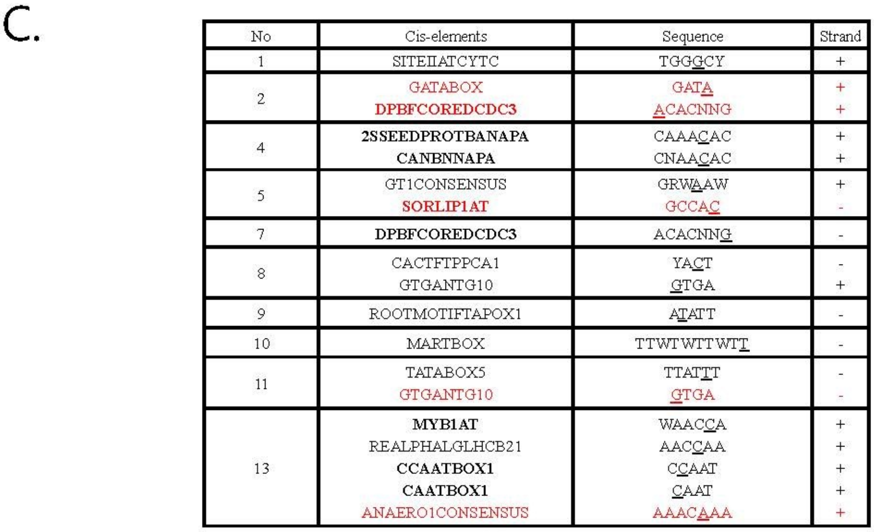




Glyma.09G222800.1

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Promoter												
	-1912	-1818	-1745	-1680	-1551	-1406	-1266	-865	-692	-467	-338	-147	-45
W82	G	G	G	T	A	A	C	G	A	A	A	T	C
TK780	G	G	G	C	A	A	C	G	A	A	A	T	C
B04009	A	A	A	T	C	G	T	A	G	T	C	C	A

	14	15
	Intron1	Intron2
	360	1118
W82	C	A
TK780	G	G
B04009	C	A



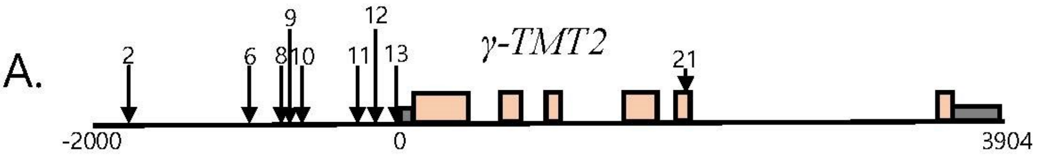
No	Cis-elements	Sequence	Strand
1	SITEIIATCYTC	TGGGCY	+
2	GATABOX **DPBFCOREDCDC3**	GATA ACACNNG	+ +
4	**2SSEEDPROTBANAPA** **CANBNNAPA**	CAAACAC CNAACAC	+ +
5	GT1CONSENSUS **SORLIP1AT**	GRWAAW GCCAC	+ -
7	**DPBFCOREDCDC3**	ACACNNG	-
8	CACTFTPPCA1 GTGANTG10	YACT GTGA	- +
9	ROOTMOTIFTAPOX1	ATATT	-
10	MARTBOX	TTWTWTTWTT	-
11	TATABOX5 GTGANTG10	TTATTT GTGA	- -
13	**MYB1AT** REALPHALGLHCB21 **CCAATBOX1** **CAATBOX1** ANAERO1CONSENSUS	WAACCA AACCAA CCAAT CAAT AAACAAA	+ + + + +

Fig. 7 Sequence polymorphisms in *γ-TMT3* gene and promoter region.

A) *γ-TMT3* gene structure is depicted as follows. Exons are shown as red boxes, and areas between two exons are introns. The 5'-UTR and 3'-UTR are shown as grey boxes. Arrows shows all SNPs and indels located in *cis*-elements (C). Numbers above arrows correspond to SNPs and indels number in (B).

B) All SNPs and indels found in gene and promoter region between TK780 and B04009. Third rows show the position of nucleotide polymorphisms relative to translational start site (ATG).

C) List of SNPs and indels located within *cis*-elements. *Cis*-elements present in TK780 but not in B04009 are shown in black, whereas *cis*-elements present in B04009 but not in TK780 are shown in red. SNPs or indels position are underlined in Sequence column. *Cis*-elements specific to seed based on PLACE prediction are written in bold letters.



B. Glyma.12G014300.1

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Promoter												
	-1797	-1790	-1281	-1200	-1019	-1017	-862	-791	-728	-691	-296	-179	-30
W82	C	A	-	-	-	TATATTTTATACCGACTGG	C	T	G	T	A	T	A
TK780	A	C	T	AA	-	-	A	A	G	C	T	T	A
B04009	C	A	-	-	GG	TATATTTTATACCGACTGG	C	T	T	T	A	A	T

	14	15	16	17	18	19	20	21
	Intron1		Intron3		Exon4	Intron4		Exon5
	432	515	1153	1267	1480	1694	1698	1870
W82	G	A	G	C	C	-	ATTT	A
TK780	G	A	G	T	T	AT	ATTT	T
B04009	T	G	A	C	C	-	-	A

	Exon5
TK780	GLSSLLSS
B04009	GLTSLLSS

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
	Intron5																						3'UTR		
	2723	2724	2731	2795	2817	3063	3081	3086	3087	3103	3138	3142	3168	3181	3192	3230	3247	3267	3271	3274	3277	3322	3641	3814	3856
W82	T	A	T	T	A	A	T	T	C	G	C	C	G	C	A	T	A	T	T	C	-	A	G	A	G
TK780	A	A	T	C	A	G	G	G	G	A	C	T	T	T	G	T	G	T	T	C	-	-	C	C	T
B04009	T	T	C	T	T	A	T	T	C	G	G	C	G	C	A	C	A	G	A	A	AGAAC	A	G	A	G

C.

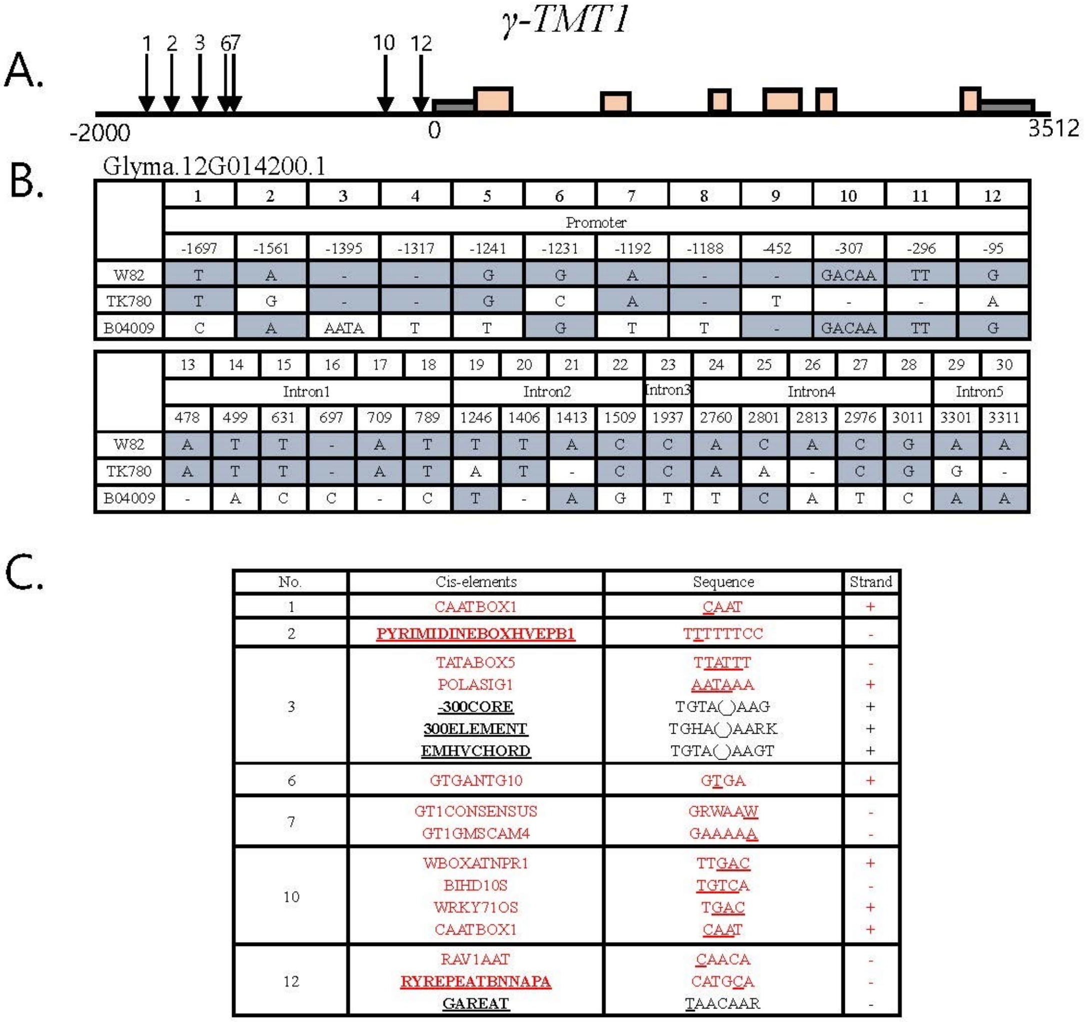
No.	Cis-elements	Sequence	Strand
2	TATABOX5	AAATAA(TTATTT)	-
6	LECPLEACS2 ROOTMOTIFTAPOX1 DRE2COREZMRAB17 DRECRTCOREAT CBFHV LTRECOREATCOR15(T)	TAAAATAT ATATT ACCGAC RCCGAC RYCGAC CCGAC	- + + + + +
8	INRNTPSADB	YTCANTYY	-
9	GATABOX IBOXCORE OSE2ROOTNODULE POLLEN1LELAT52	GATA GATAA CTCTT AGAAA	+ + - +
10	BIHD1OS	TGTCA	-
11	ARR1AT	NGATT	+
12	S1FBOXSORPS1L21	ATGGTA	+
13	ROOTMOTIFTAPOX1 -10PEHVPSBD SEF4MOTIFGM7S	ATATT TATTCT RTTTTTR	+ + -

Fig. 8 Sequence polymorphisms in *γ-TMT2* gene and promoter region.

A) *γ-TMT2* gene structure is depicted as follows. Exons are shown as red boxes, and areas between two exons are introns. The 5'-UTR and 3'-UTR are shown as grey boxes. Arrows shows all SNPs and indels located in *cis*-elements (C). Numbers above arrows correspond to SNPs and indels number in (B).

B) All SNPs and indels found in gene and promoter region between TK780 and B04009. Third rows show the position of nucleotide polymorphisms relative to translational start site (ATG).

C) List of SNPs and indels located within *cis*-elements. *Cis*-elements present in TK780 but not in B04009 are shown in black, whereas *cis*-elements present in B04009 but not in TK780 are shown in red. SNPs or indels position are underlined in Sequence column. *Cis*-elements specific to seed based on PLACE prediction are written in bold letters.



	1	2	3	4	5	6	7	8	9	10	11	12
	Promoter											
	-1697	-1561	-1395	-1317	-1241	-1231	-1192	-1188	-452	-307	-296	-95
W82	T	A	-	-	G	G	A	-	-	GACAA	TT	G
TK780	T	G	-	-	G	C	A	-	T	-	-	A
B04009	C	A	AATA	T	T	G	T	T	-	GACAA	TT	G

	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	Intron1						Intron2				Intron3	Intron4					Intron5	
	478	499	631	697	709	789	1246	1406	1413	1509	1937	2760	2801	2813	2976	3011	3301	3311
W82	A	T	T	-	A	T	T	T	A	C	C	A	C	A	C	G	A	A
TK780	A	T	T	-	A	T	A	T	-	C	C	A	A	-	C	G	G	-
B04009	-	A	C	C	-	C	T	-	A	G	T	T	C	A	T	C	A	A

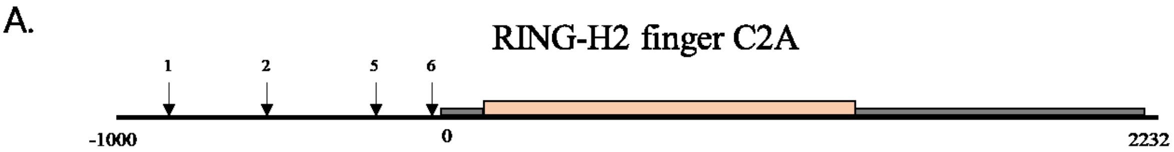
No.	Cis-elements	Sequence	Strand
1	CAATBOX1	CAAT	+
2	**PYRIMIDINEBOXHVEPB1**	TTTTTTCC	-
3	TATABOX5 POLASIG1 **-300CORE** **300ELEMENT** **EMHVCHORD**	TTATTT AATAAA TGTA( )AAG TGHA( )AARK TGTA( )AAGT	- + + + +
6	GTGANTG10	GTGA	+
7	GT1CONSENSUS GT1GMSCAM4	GRWAAW GAAAAA	- -
10	WBOXATNPR1 BIHD1OS WRKY71OS CAATBOX1	TTGAC TGTCA TGAC CAAT	+ - + +
12	RAV1AAT **RYREPEATBNNAPA** **GAREAT**	CAACA CATGCA TAACAAR	- - -

Fig. 9 Sequence polymorphisms in *γ-TMT1* gene and promoter region.

A) *γ-TMT1* gene structure is depicted as follows. Exons are shown as red boxes, and areas between two exons are introns. The 5'-UTR and 3'-UTR are shown as grey boxes. Arrows shows all SNPs and indels located in *cis*-elements (C). Numbers above arrows correspond to SNPs and indels number in (B).

B) All SNPs and indels found in gene and promoter region between TK780 and B04009. Third rows show the position of nucleotide polymorphisms relative to translational start site (ATG).

C) List of SNPs and indels located within *cis*-elements. *Cis*-elements present in TK780 but not in B04009 are shown in black, whereas *cis*-elements present in B04009 but not in TK780 are shown in red. SNPs or indels position are underlined in Sequence column. *Cis*-elements specific to seed based on PLACE prediction are written in bold letters.



B.

Glyma.11G220400

	1	2	3	4	5	6	7	8	9
	Promoter						3'UTR	5'UTR	
	-880	-614	-412	-399	-218	-50	49	1997	2147
W82	A	A	T	G	A	G	A	G	A
TK780	A	A	T	G	A	G	A	G	A
B04009	T	T	C	T	C	T	T	A	C

C.

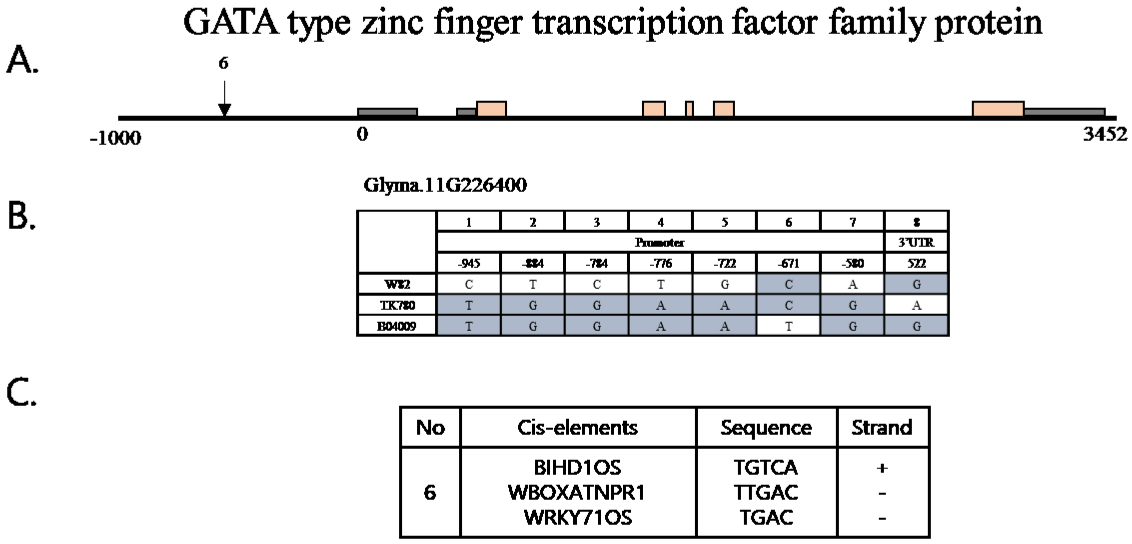
No	Cis-elements	Sequence	Strand
1	RBCSCONSENSUS ARR1AT SEF4MOTIFGM7S REALPHALGLHCB21 MYB1AT	AATCCAA NGATT RTTTTTR AACCAA WAACCA	- + + - -
2	S1FSORPL21 S1FBOXSORPS1L21 MYB1AT	ATGGTATT ATGGTA WAACCA	+ + -
5	ERELEE4 SEF4MOTIFGM7S	AWTTCAAA RTTTTTR	+ -
6	CRTDREHVCBF2 CBFHV	GTCGAC RYCGAC	+ +

Fig. 10 Sequence polymorphisms in the promoter and genic regions of Glyma.11G220400 among TK780, B04009 and Williams 82 (W82).

A) Gene structure of Glyma.11G220400(RING-H2 finger C2A*),* Red boxes, grey boxes and lines show exons, UTRs and introns, respectively. Arrows shows the SNPs and indels located in cis-elements (C). Numbers above arrows correspond to SNPs and indels number in (B).

B) The SNPs and indels found in the gene and promoter region between TK780 and B04009. Third rows show the position of nucleotide polymorphisms relative to the Adenine of translational start site (ATG).

C) List of SNPs and indels located within *cis*-elements. *Cis*-elements specific to TK780 and B04009 are shown in black and red, respectively.



	1	2	3	4	5	6	7	8
	Promoter							3'UTR
	-945	-884	-784	-776	-722	-671	-580	522
W82	C	T	C	T	G	C	A	G
TK780	T	G	G	A	A	C	G	A
B04009	T	G	G	A	A	T	G	G

No	Cis-elements	Sequence	Strand
6	BIHD1OS WBOXATNPR1 WRKY71OS	TGTCA TTGAC TGAC	+ - -

Fig. 11. Sequence polymorphisms in the promoter and genic regions of Glyma.11G226400 among TK780, B04009 and Williams 82 (W82).

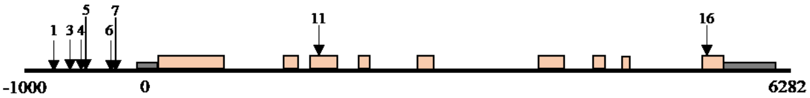
A) Gene structure of Glyma.11G226400(GATA type zinc finger transcription factor family protein *),* Red boxes, grey boxes and lines show exons, UTRs and introns, respectively. Arrows shows the SNPs and indels located in *cis*-elements (C). Numbers above arrows correspond to SNPs and indels number in (B).

B) The SNPs and indels found in the gene and promoter region between TK780 and B04009. Third rows show the position of nucleotide polymorphisms relative to the Adenine of translational start site (ATG).

C) List of SNPs and indels located within *cis*-elements. *Cis*-elements specific to TK780 and are shown in black.

S-adenosyl-L-methionine-dependent methyltransferases superfamily protein

A.



B.

Glyma.11G222800.1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Promoter							Exon1	Intrion1	Intrion2	Exon3	Exon4	Intron4	Intron5	Intron8	Exon9	3'UTR			
	-780	-778	-634	-578	-555	-252	-244	416	915	1600	1759	2181	2304	3048	4847	5626	5836	6076	6094	6104
W82	G	A	G	T	T	G	A	T	G	T	A	C	T	G	G	T	G	G	T	TTA
TK780	G	A	G	T	T	G	A	T	G	T	A	C	T	G	G	T	G	G	T	TTA
B04009	A	T	A	-	A	A	T	A	T	G	T	T	G	C	A	C	T	T	C	C

C.

No	Cis-elements	Sequence	Strand
1	INRNTPSADB	YTCANTYY	-
3	ARR1AT	NGATT	-
4	POLASIG2 ROOTMOTIFTAPOX1	AATTAAA ATATT	- -
5	SEF4MOTIFGM7S	RTTTTR	+
6	ROOTMOTIFTAPOX1	ATATT	+
7	SEF4MOTIFGM7S	RTTTTR	+

	Exon3
W82	GCGVASFGAY
TK780	GCGVASFGAY
B04009	GCGVACFGAY

	Exon9
W82	RCNVSSIMLE
TK780	RCNVSSIMLE
B04009	RCNVPSIMLE

Fig. 12 Sequence polymorphisms in the promoter and genic regions of Glyma.11G222800.1 among TK780, B04009 and Williams 82 (W82).

A) Gene structure of Glyma.11G222800.1(S-adenosyl-L-methionine-dependent methyltransferases superfamily protein*),* Red boxes, grey boxes and lines show exons, UTRs and introns, respectively. Arrows shows the SNPs and indels located in *cis*-elements (C). Numbers above arrows correspond to SNPs and indels number in (B).

B) The SNPs and indels found in the gene and promoter region between TK780 and B04009. Third rows show the position of nucleotide polymorphisms relative to the Adenine of translational start site (ATG).

C) List of SNPs and indels located within *cis*-elements. *Cis*-elements specific to TK780 and B04009 are shown in black and red, respectively.

Table 1. Tocopherol contents and ratios in seeds of TK780 and B04009 produced in different thermal conditions

	Thermal condition	Content (µg/g)				Ratio (%)	
		δ-Toc	γ-Toc	α-Toc	total-Toc	(α+γ)/total ratio	α/γ ratio
TK780	20ºC	178.0 ± 29	204 ± 36	16 ± 8	397 ± 73	55.1 ± 0.9	7.5 ± 2.5
	30ºC	115 ± 4	313 ± 26	43 ± 11	471 ± 20	75.5 ± 0.4	13.8 ± 4.8
B04009	20ºC	63 ± 9	146 ± 38	26 ± 7	236 ± 55	72.8 ± 2.3	17.0 ± 0.9
	30ºC	29 ± 8	82 ± 21	77 ± 18	187 ± 25	84.5 ± 4.1	114.8 ± 15.0

(α + γ)/total ratio= ratio of sum of α-Toc and γ-Toc contents to total-Toc content (%)

α/γ ratio= ratio of α-Toc content to γ-Toc content (%)

Numerals within parenthesis are standard deviations

Table 2. Correlation coefficients (*r*) between years for tocopherol contents and ratios

Traits		*r*
Contents		
	δ-tocopherol	0.696 **
	γ-tocopherol	0.459 **
	α-tocopherol	0.772 **
	Total tocopherol	0.462 **
Ratios		
	(α + γ)/total ratio	0.474 **
	α/γ ratio	0.768 **

(α + γ)/total ratio; ratio of sum of γ-Toc and α-Toc contents to total-Toc content

α/γ ratio; ratio of α-Toc content to γ-Toc content

**, significant at 1% level

Table 3. Correlation coefficients (*r*) between contents of different tocopherol isoforms and total tocopherol.

		2016	2017
α-tocopherol content	γ-tocopherol content	-0.023 ns	-0.172 ns
	δ-tocopherol content	-0.145 ns	-0.112 ns
	Total tocopherol content	0.292 **	0.177 ns
γ-tocopherol content	δ-tocopherol content	0.754 **	0.747 **
	Total tocopherol content	0.931 **	0.909 **
δ-tocopherol content	Total tocopherol content	0.781 **	0.847 **

**, significant at 1% level, ns, non-significant

Table 4. QTLs for seed tocopherol contents in RILs of TK780 and B04009 cross.

Trait	Year tested	QTL-Chr	Position (cM)	Support interval (cM)	LOD score	PVE (%)	AE	Flanking marker[1)]
α-Toc content	2016	*qαTC-5*	142	141.5-144.5	6.1	9.2	-7.6	S05_37473691/37803884
		qαTC-9	92	91.5-92.5	14.3	28.7	12.9	S09_43927286/44366371
		qαTC-11	198	193.5-198.5	4.0	5.8	-5.8	S11_31748669/32054530
		qαTC-12	3	2.5-3.5	7.8	12.7	8.6	S12_2207593/2310293
	2017	*qαTC-9*	92	91.5-92.5	13.1	39.4	11.2	S09_43927286/44366371
		qαTC-11[1)]	196	191.5-198.5	3.2[2)]	8.4	-5.2	S11_31748669/32054530
		qαTC-12	1	0-1.5	6.0	14.7	6.9	S12_1507927/1790872
δ-Toc content	2017	*qδTC-4*	43	42.5-45.5	5.5	9.6	6.9	S04_6780105/7188146
		qδTC-6	117	116.5-118.5	4.7	7.9	6.1	S06_42091687/43647797
		qδTC-19	79	78.5-79.5	14.1	32.4	-12.4	S19_37454169/38007384
γ-Toc content	2016	*qγTC-7*	97	96.5-99.5	4.0	16.4	-19.8	S07_19171390/19946051
	2017	*qγTC-6*	75	74.5-76.5	5.6	7.3	13.9	S06_14027025/14268479
		qγTC-13	20	16.5-20.5	6.8	8.6	-14.8	S13_14124674/14460816
		qγTC-14	91	90.5-91.5	7.7	9.9	16.1	S14_25722294/30081672
Total-Toc content	2016	*qTTC-1*	141	139.5-141.5	5.5	22.5	-27.1	S01_52482586/53582878

1) Physical position of the QTL support intervals can be determined based on the flanking marker name. Two digits after S define chromosome, whereas the number after underline define its physical position on the reference genome Williams82.a2.

2) p= 0.066

PVE= percent of variation explained (%), AE; Additive effect of B04009 allele (µg/g)

Table 5. QTLs for seed tocopherol ratios in RILs of TK780 and B04009 cross.

Trait	Year tested	QTL-Chr	Position (cM)	Support interval (cM)	LOD score	PVE (%)	AE	Flanking marker[1)]
α/γ ratio	2016	*qαγR-9*	96	95.5-97.5	18.2	42.8	6.9	S09_45093824/ 45831868
		qαγR-11	198	194.5-198.5	5.5	8.5	-3.1	S11_31748669 / 32054530
		qαγR-12.1	1	0-1.5	5.7	9.8	3.3	S12_1507927 / 1790872
		qαγR-12.2	49	47.5-49.5	3.7	5.6	-2.5	S12_9268661 / 9725498
	2017	*qαγR-9*	92	91.5-92.5	21.1	42.6	4.8	S09_43927286 / 44366371
		qαγR-11	197	192.5-198.5	4.6	6.3	-1.9	S11_31748669 / 32054530
		qαγR-12.1	1	0-1.5	10.5	16.0	3.0	S12_1507927 / 1790872
		qαγR-17	44	43.5-44.5	4.1	5.1	-1.7	S17_8786113 / 9025866

1) Physical position of the QTL support intervals can be determined based on the flanking marker name. Two digits after S define chromosome, whereas the number after underline define its physical position on the reference genome Williams82.a2.

α/γ ratio (%)= ratio of α-Toc content to γ-Toc content

PVE= percent of variation explained (%), AE; Additive effect of B04009 allele (µg/g)

2.4 Discussion

QTL mapping and genome-wide association studies have revealed molecular and genetic bases underlying the natural variation in seed tocopherol contents and compositions in *Arabidopsis* (Gilliland et al., 2006) and major crops such as maize (Diepenbrock et al., 2017; Wang et al., 2018), barley (Graebner et al., 2015), rice (Wang et al., 2015), tomato (Almeida et al., 2011; Quadrana et al., 2014), soybean (Li et al., 2010; Dwiyanti et al., 2011; Shaw et al., 2016; Liu et al., 2017; Shaw et al., 2017) , and *Brassica napus* (Marwede et al., 2005; Fritsche et al., 2012; Wang et al., 2012). The α-Toc contents and concentrations are strongly associated with *γ-TMT* (*VTE4*) in maize (Diepenbrock et al., 2017) and rice (Wang et al., 2015). In tomato, epialleles with different DNA methylation levels in the promoter region proximal to the start codon of the gene encoding MPBQ-MT (*VTE3*) affect *VTE3* expression levels and are inversely correlated with γ-Toc contents (Quadrana et al., 2014). As well as those in tocopherol biosynthesis pathways, novel genes also exhibit an association with contents of γ-Toc, δ-Toc, and total Toc, including two genes for protochlorophyllide reductase in chlorophyll biosynthesis and a gene for long-chain acyl-coenzyme A synthase in the fatty acid pathway of maize (Diepenbrock., 2017; Wang et al., 2018).

A number of QTLs for seed α-Toc contents have been reported in soybean (Marwede et al., 2005; Li et al., 2010; Dwiyanti et al., 2011; Shaw et al., 2016; Liu et al., 2017; Shaw et al., 2017). Li et al. (2010) identified four QTLs for α-Toc contents by single marker analyses in a cross between the high α-Toc Canadian cultivar OAC Bayfield and the low α-Toc Chinese cultivar Hefeng 25. Shaw et al. (2017) found nine and five QTLs, by single marker analyses and interval mapping, respectively, for α-Toc contents in a cross with OAC Bayfield and a low α-Toc OAC Shire across three locations over two years, of which the QTL tagged by Satt117 (Chr15) had the largest effect, accounting for up to 32% of the phenotypic variation. Liu et al. (2017) reported a total of 18 QTLs for α-Toc contents in an RIL population of a cross with the Chinese high α-Toc local variety Beifeng 9, of which four QTLs in Chr15 had stable and significant additive effects across six environments. These studies similarly detected QTLs in Chr15, although the candidate genes remained undetermined. However, no genes encoding the enzymes directly involved in α-Toc biosynthesis are colocalized in these QTL regions. Only a QTL detected in a cross with the high α-Toc cultivar KAS harbored *γ-TMT3*; the causal factor in the elevated α-Toc in KAS seeds was considered the higher promoter activity of *γ-TMT3* (Dwiyanti et al., 2011).

In the present study, I identified three QTLs that were consistently detected in a 2-year experiment in the RIL population of a cross between the high α-Toc wild accession B04009 and low α-Toc breeding line. In the RIL population, the α-Toc contents varied in close association with the α/γ ratio, which is the index for conversion efficiency from γ-Toc to α-Toc mediated by γ-TMT. However, the α-Toc content was not associated with the contents of total Toc or other isoforms. The three QTLs detected were involved in both the α-Toc content and the α/γ ratio. These QTLs may therefore bring about the accumulation of more α-Toc in seeds by enhancing the conversion from γ-Toc to α-Toc. The QTL with the largest effect was located close to the QTL in Chr9 as previously detected in the cross between Ichihime and KAS (Dwiyanti et al., 2011) (Tables 4 and 5, Fig. 5), suggesting that the high α-Toc trait in B04009 may be controlled by the same QTL as detected in the Ichihime × KAS cross.

The other two QTLs, *qαγR-11*/*qαTC-11* and *qαγR-12.1*/*qαTC-12*, were the novel QTLs which were not detected in the previous studies. The genomic region of *qαγR-12.1*/*qαTC-12* was close to linked pair of *γ-TMT* genes, *γ-TMT1* and *γ-TMT2*. Either of or both genes may be a probable candidate(s) of *qαγR-12.1*/*qαTC-12*.

The QTL in Chr11 had an opposite effect compared with the other two QTLs, and likely contributed to the transgressive segregation in the α-Toc contents and α/γ ratios in the RIL population. The genomic region of the QTL contained no known genes involved in Toc biosynthesis but included two zinc finger transcription factors (Glyma.11G220400 and Glyma.11G226400) and S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (Glyma.11G222800), all of which are expressed in developing seeds (https://soybase.org/). Of these, GATA-type zinc fingers are known GATA-binding transcription factors that control embryo development in *Arabidopsis* (Dean Rider et al., 2003). Furthermore, the synthetic zinc finger transcription factor, fused to a nuclear localization signal and the maize C1 activation domain, successfully upregulated *γ-TMT* by binding to *cis*-elements in the promoter to elevate the seed α-Toc ratio in *Arabidopsis* (Van Eenennaama et al., 2004). Resequencing analyses revealed that the three candidate genes for the QTL in Chr11 had a lot of DNA polymorphisms in the promoter, and additionally, a gene for S-adenosyl-L-methionine-dependent methyltransferases superfamily protein had two non-synonymous substitutions between TK780 and B04009. Because the amino acid residues in TK780 were conserved in the homoeolog (Glyma18G034800.1) and homologs in *Arabidopsis* and the other plant species, the protein from B04009 might be a missense variant. I carried out the expression analysis for the three genes in the next chapter.

Chapter 3

Molecular mechanism of thermal response of α-tocopherol biosynthesis

3.1 Backgroud and Purposes

Temperature during seed development is one of the environmental factors that influence the α-Toc contents and ratios in seeds (Carrera et al., 2016). The α-Toc contents and ratios increase as temperatures rise during seed maturation (Almonora et al., 1998; Britz et al., 2002; Britz et al., 2008; Seguin et al., 2010; Carrera et al., 2011; Chennupati et al., 2011; Dwiyanti et al., 2016). γ-TMT is the key enzyme in the last step of tocopherol biosynthesis, which converts γ-Toc to α-Toc or δ-Toc to β-Toc (DellaPenna, 2005). Soybean has three *γ-TMT* genes, but it remains unknown how these *γ-TMT* genes function in the Toc biosynthesis in response to temperatures. In this chapter, I first evaluated the variation in thermal responses of Toc biosynthesis among soybean cultivars and wild accessions, and second compared expression profiles of *γ-TMT* genes in leaves and immature cotyledons of plants grown in different thermal conditions and investigated the methylation of the promoter regions as one of factors for different expression profiles. Finally, I examined the associations between α-Toc ratios and DNA polymorphisms for the three *γ-TMT* genes.

3.2 Material and Method

3.2.1 Plant materials

Twenty-one soybean cultivars and 11 wild soybean accessions were used in this study. The origins/collection sites are given in Table 5. These materials were grown at 25 °C in the greenhouse, and after flowering, transferred to growth chambers set to 20 ºC and 30 ºC; seeds were harvested individually from one to ten plants.

3.2.2 RNA extraction and expression analysis

Three to four pods of full seed stage (Fehr et al., 1971) and three to four leaves were sampled individually from two to five plants for TK780, B04009 and KAS grown at 20 °C and 30 °C 12 h after lighting. The sampled leaves and immature seeds were bulked separately, immediately frozen in liquid nitrogen, and stored at –80 °C. Frozen tissue was ground to a fine powder under liquid nitrogen by Multi-Beads Shocker. Total RNA was isolated from 200 mg of frozen seed powder using TRIzol reagent (Thermo Fisher Scientific, Yokohama, Japan). DNase I (Takara Bio, Shiga, Japan) was used to remove genomic DNA. cDNAs were synthesized from 1 μg of total RNA by using the M-MLV reverse transcriptase system (Invitrogen, Tokyo, Japan) with an oligo (dT) 20 primer. Transcript levels of three *γ-TMT* genes, *γ-TMT1*, *γ-TMT2* and *γ-TMT3*, and three candidate genes for *qαγR-11*/*qαTC-11,* Glyma.11G220400, Glyma.11G222800 and Glyma.11G226400, were determined by quantitative real-time PCR. The PCR mixture (20 μL) contained 0.1 μL of the cDNA synthesis reaction, 5 μL of 1.2 μM primer premix, and 10 μL SYBR Premix Ex Taq II (Takara). A CFX96 Real-Time System (Bio-Rad, Osaka, Japan) was used. PCR cycling conditions were 95 °C for 3 min followed by 39 cycles of 95 °C for 10 s, 57 °C for 20 s, 72 °C for 20 s, and 78 °C for 2 s. Fluorescence was quantified before and after incubation at 78°C to monitor the formation of primer dimers. Actin mRNA (Glyma.18G222800.1) was used for normalization. A reaction mixture without reverse transcriptase was used as a control to confirm the absence of genomic DNA contamination. Amplification of a single DNA fragment was confirmed by melting curve analysis and gel electrophoresis of PCR products. Averages and standard errors of relative expression levels were calculated from PCR results for four independently synthesized cDNAs. Primer sequences used in expression analyses are listed in Table 6.

3.2.3 Methylation analysis

Genomic DNA was extracted from immature cotyledons of full seed size sampled from TK780 and KAS grown in two different thermal conditions (20 °C and 30 °C) after flowering, using the modified CTAB method. DNA samples were digested with McrBC (Takara Bio, Otsu, Japan) that cleaves DNA sequence containing two methylated cytosine residues ((G/A)^{m}C) separated by 30-2000 bp on one or both strands. Digested and undigested samples were used in PCR amplification targeting different regions of *γ-TMT2* and *γ-TMT3* promoters. Primer sequences used in methylation analyses are listed in Table 7

3.2.4 Tocopherol quantification

Tocopherol contents of seeds were quantified according to the method described in the previous chapter.

3.2.5 Construction of *γ-TMT* genes based on whole genome resequencing data

Raw reads of 15 soybean cultivars and 9 wild soybean accessions from next-generation sequencing Illumina Hiseq XTen were aligned to the soybean reference genome Williams82.a2 (Schmutz et al., 2010). The alignment was performed using Bowtie2-2.2.9 (Langmead et al., 2012). The resulting alignment was further processed to remove duplicate reads and to fix mate information using Picard tools (http://broadinstitute.github.io/picard). The Genome Analysis Toolkit (GATK ver 3.8; (McKenna et al., 2010) was used to realign small indels. Subsequently, variants (SNP and indels) were called using the GATK Unified Genotyper function which filtered out reads having mapped base quality Phred scores below 20. Using the reference genome Williams82.a2 and SNP dataset of each variety, sequences of *γ-TMT* genes were reconstructed by the FastaAlternateReferenceMaker function available in GATK.

3.2.6 Association test of α-tocopherol ratios with SNPs in *γ-TMT1*, *γ-TMT 2* and *γ-TMT3*

The 24 soybean accessions and William 82 were classified based on the nucleotides at SNPs. The significance of association between α-tocopherol ratio and SNPs was analyzed by one-way analysis of variance. The SNPs tested included those in the promoter and introns and the non-synonymous SNPs in exons.

3.2.7 *Cis*-element prediction

New PLACE, a Plant *Cis*-acting Regulatory DNA Elements database (http://www.dna.affrc.go.jp/PLACE/) (Higo et al., 1999) was used to predict the position of *cis*-acting regulatory elements in the promoter region (2,000 bp upstream from the start codons) of *γ-TMT1*, *γ-TMT2*, and *γ-TMT3* of 15 soybean cultivars and 9 wild soybean accessions.

3.3 Results

3.3.1 α-Toc ratios in seeds matured under different thermal conditions

The variation in the α-Toc ratios in seeds matured at 20 °C and 30 °C are presented in Fig. 13-1. The seed α-Toc ratios of 21 cultivars varied widely from 3.9 % to 15.5 % at 20 °C and from 9.1 % to 49.5 % at 30 °C. Similarly, the α-Toc ratios of 11 wild accessions varied widely from 5.8 % to 23.0 % at 20 °C and from 10.1 % to 57.2 % at 30 °C. Of the 32 accessions tested, 14 cultivars and 10 wild accessions were tested in both temperature conditions. Fig. 13-2 presents combined data of α-Toc ratios in the two temperature conditions for these 24 accessions. The α-Toc ratios increased as temperatures increased in whole lines, as reported in Dwiyanti et al. (2016). However, there were large differences in increasing rates of α-Toc ratios among cultivars and wild accessions. Some lines such as Tanishi Daizu, Norin 2 go, KAS, Tachi Nagaha, B00092, and B08040 showed highly elevated α-Toc ratios with the rise of temperatures. On the contrary, Feng Dihuang, Aomori and B08045 showed only small changes in α-Toc ratios.

3.3.2 Expression profiles for *γ-TMT* genes in different tissues under different thermal conditions

The expression levels of *γ-TMT1*, *γ-TMT2*, and *γ-TMT3* were analyzed in leaves and immature cotyledons of full seed size sampled from TK780, B04009 and KAS plants grown in two different thermal conditions after flowering. The expression levels of *γ-TMT1* and *γ-TMT2* in leaves were higher approximately 80 fold than immature cotyledons in both temperatures in all lines. *γ-TMT3* also showed higher expression levels in leaves than cotyledons in the two temperature conditions in TK780 and B04009; however, the extent was low compared with those of *γ-TMT1* and *γ-TMT2.* Interestingly, the *γ-TMT3* expression levels were higher in immature cotyledons than leaves in KAS in both temperature conditions (Fig. 14).

In immature cotyledons, the expression levels of *γ-TMT1* and *γ-TMT3* were higher in KAS than TK780 and B04009 in both temperature conditions. In contrast, the *γ-TMT2* expression was almost the same among three lines in 20 °C, whereas in 30 °C the expression was significantly upregulated in KAS and B04009 (Fig. 14). TK780 and B04009, parents of RIL population used in the QTL analysis, showed similar expression levels in *γ-TMT1*, but the expressions of *γ-TMT3* were higher in the latter than the former at both temperatures.

3.3.3 Expression profiles of candidate genes for *qαγR-11*/*qαTC-11*

The expression levels of three candidate genes of QTL in Chr11, Glyma.11G220400 (RING-H2 FINGER C2A) Glyma.11G222800 (S-adenosyl-L-methionine-dependent methyltransferases superfamily protein) and Glyma.11G226400 (GATA type zinc finger transcription factor family protein), were analyzed in immature cotyledons of full seed size sampled from TK780 and B04009 plants grown in two different thermal conditions after flowering (Fig. 15). The expression levels of Glyma.11G220400 and Glyma.11G226400 did not produce any consistent and large differences between TK780 and B04009; the expression levels were upregulated in 20 °C relative to 30 °C, except for Glyma.11G226400 of B04009. The expression levels of Glyma.11G222800 were higher in TK780 than B04009 at both temperatures (Fig. 15), in accordance with the finding that the TK780 allele at *qαγR-11*/*qαTC-11* promoted the α-Toc ratios, differently from the other two QTLs.

3.3.4 Methylation analysis

As indicated in the expression analyses, the *γ-TMT* genes exhibited different expression profiles between tissues and between temperature conditions. The three lines exhibited different expression profiles in response to different temperatures. To determine whether the methylation is involved in different expressions, I analyzed the cytosine methylation in the promoter regions of *γ-TMT2* and *γ-TMT3*, using methylation-dependent restriction enzyme McrBC. Primers used in PCR were designed for each of the targeted regions to be singly amplified. PCR was performed for the McrBC-digested and non-digested DNA samples. If the targeted genome region contains either or both of the cytosine residues methylated at two McrBC-target sites ((G/A)^{m}C) that are located within a distance of 30–2,000 bp, the region is cleaved by McrBC, so that the PCR targeting the region cannot produce any amplicon.

The results of electrophoresis of PCR products are presented in Fig. 16. The PCR amplification targeting the promoter region of *γ-TMT2* produced the amplicons of expected lengths in both TK780 and KAS, irrespective of thermal conditions during maturation (Fig. 16A and 16C). In contrast, the PCR amplification targeting the promoter region of *γ-TMT3* from the start codon to 1,342-bp upstream produced no amplicon when the McrBC-digested DNAs from TK780 grown at 20 °C were used as a template. The PCR amplification produced the expected PCR product when the DNAs from TK780 grown at 30 °C were used as a template. Therefore, this methylation was dependent on the temperatures. To delineate the methylated region, the region was further divided into three regions of 653 bp (region B), 466bp (region C) and 224 bp (region D). Among them, the region B of 1,342 bp to 689 bp upstream from the start codon did not show any amplicons, whereas the other two regions produced the amplicons with the expected lengths (Fig. 16B and 16D). However, there was no methylation in the regions tested in KAS. Accordingly, the temperature-dependent methylation in the promoter of *γ-TMT3* was specific to TK780.

Then, the sequences of the methylated region were compared among TK780, KAS and Williams 82. There were a total of seven SNPs among the accessions compared, of which two were located in the McrBC-targeted cytosine ((G/A)^{m}C) sites of 1336 bp and 1266 bp upstream of the start codon (Fig. 17). TK780 and William82 possessed the cytosine at the two sites, whereas KAS possessed thymine (Fig. 17). These polymorphisms may be responsible for the difference in methylation between TK780 and KAS.

3.3.5 Association test between α-tocopherol ratios and SNPs

To determine the association between α-Toc ratios and DNA polymorphisms, DNA sequences of 2,000 bp covering the promoter and gene region were analyzed for each of three *γ-TMT* genes.

In the *γ-TMT1* promoter region, a total of 23 SNPs and two indels were detected, but there was no significant association between α-Toc ratios and DNA polymorphisms (Fig. 18). In the *γ-TMT1* gene region, a total of 26 SNPs and two indels were detected; there was no non-synonymous substitution (Fig. 19). However, the accessions having the nucleotide T at a SNP in intron 1 (Polymorphism number 28) produced higher α-tocopherol ratios in 30 °C than those having the C nucleotide ($p = 0.004$) (Fig. 20).

In the *γ-TMT2* promoter region, a total of 24 SNPs and one indel were detected (Fig. 21). The accessions having the nucleotide T at a SNP in the promoter region (Polymorphism number 6; SNP-6) produced higher α-tocopherol ratios in 20 °C than those having the C nucleotide ($p = 0.027$). Similarly, the accessions having the nucleotide A at a SNP in the promoter region (Polymorphism number 25; SNP-25) produced higher α-tocopherol ratios in 30 °C than those having the T nucleotide ($p = 0.003$). Furthermore, these SNPs generated the changes of *cis*-elements; the accessions having the nucleotide T at SNP-6 possessed a cis-element, WRKY71OS, a binding site of transcriptional repressor of the gibberellin signaling pathway (Zhang et al., 1996). The accessions having the nucleotide T at SNP-25 possessed a *cis*-element -10PEHVPSBD which was bound by a chlorophyll-binding protein (Thum et al., 2001)(Fig. 22). In addition, a total of 51 SNPs and seven indels were detected in the *γ-TMT2* gene region. There were four non-synonymous substitutions. However, there was no significant association between α-Toc ratio and these DNA polymorphisms (Fig. 23).

In the *γ-TMT3* promoter region, a total of 72 SNPs and one indels were detected (Fig. 24). Three SNPs (Polymorphism numbers 36, 72 and 73) showed highly significant associations in both temperature conditions. The accessions having the nucleotides G, C and A at SNP-36, SNP-72 and SNP-73, respectively, produced higher α-tocopherol ratios than those having the G, C and A nucleotides (20 °C, $p = 0.058$; 30 °C, $p = 0.00001$). In addition, three SNPs (Polymorphism numbers 3,26 and 57) produced significant associations only in 30 °C; the accessions having the nucleotides A, C and G at SNP-3, SNP-26 and SNP-57, respectively, produced higher α-tocopherol ratios than those having the G, A and A nucleotides (p = 0.005). Furthermore, some of these SNPs generated or abolished *cis*-elements (Fig. 25). Particularly, the nucleotide A at SNP-26 generated a *cis*-element SORLIP1AT, which is conserved in the genes with light-regulated expression patterns during seedling development (Jiao et al., 2005), and the nucleotide A at SNP-73 did CACTFTPPCA1, a key component of expression module found in the cis-regulatory element (Gowik et al., 2004). Total of 28 SNPs and 10 indels were detected in the *γ-TMT3* gene region (Fig. 26). There were three non-synonymous substitutions, any of which were not associated with α-Toc ratios. Only a SNP in intron 1 (Polymorphism number 74) exhibited a highly significant association (p = 0.0005) with α-Toc ratios in 30 °C; the accessions having the nucleotide G produced higher α-tocopherol ratios than those having the C nucleotide (Fig. 27).

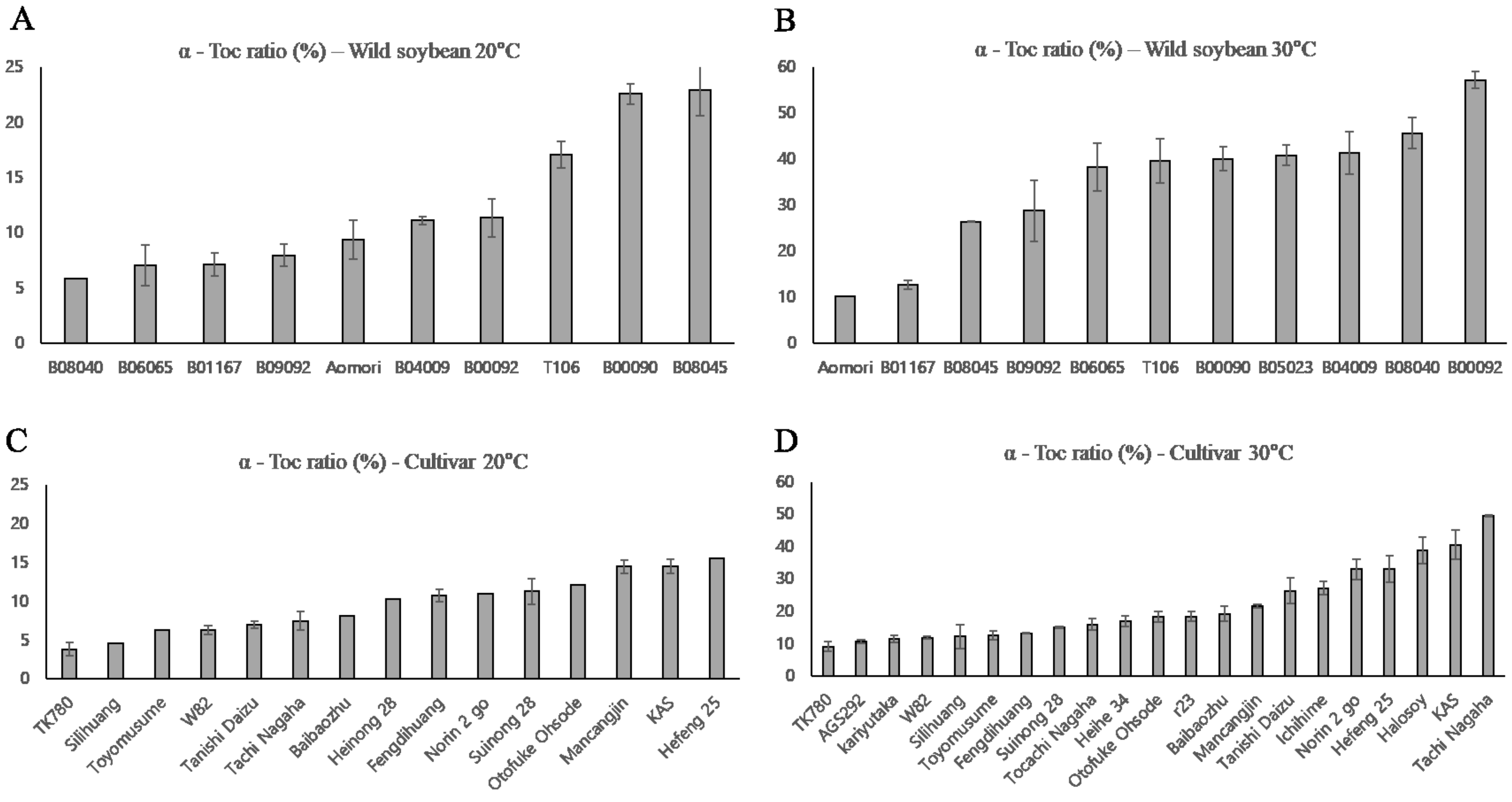


Fig. 13-1 Variation in α -tocopherol ratios of seeds matured at different temperature conditions in various soybean cultivars and wild accessions.

A, B) α -tocopherol ratios in wild accessions in 20 °C (A) and 30 °C (B). C, D) α -tocopherol ratios in soybean cultivars in 20 °C (C) and 30 °C (D).

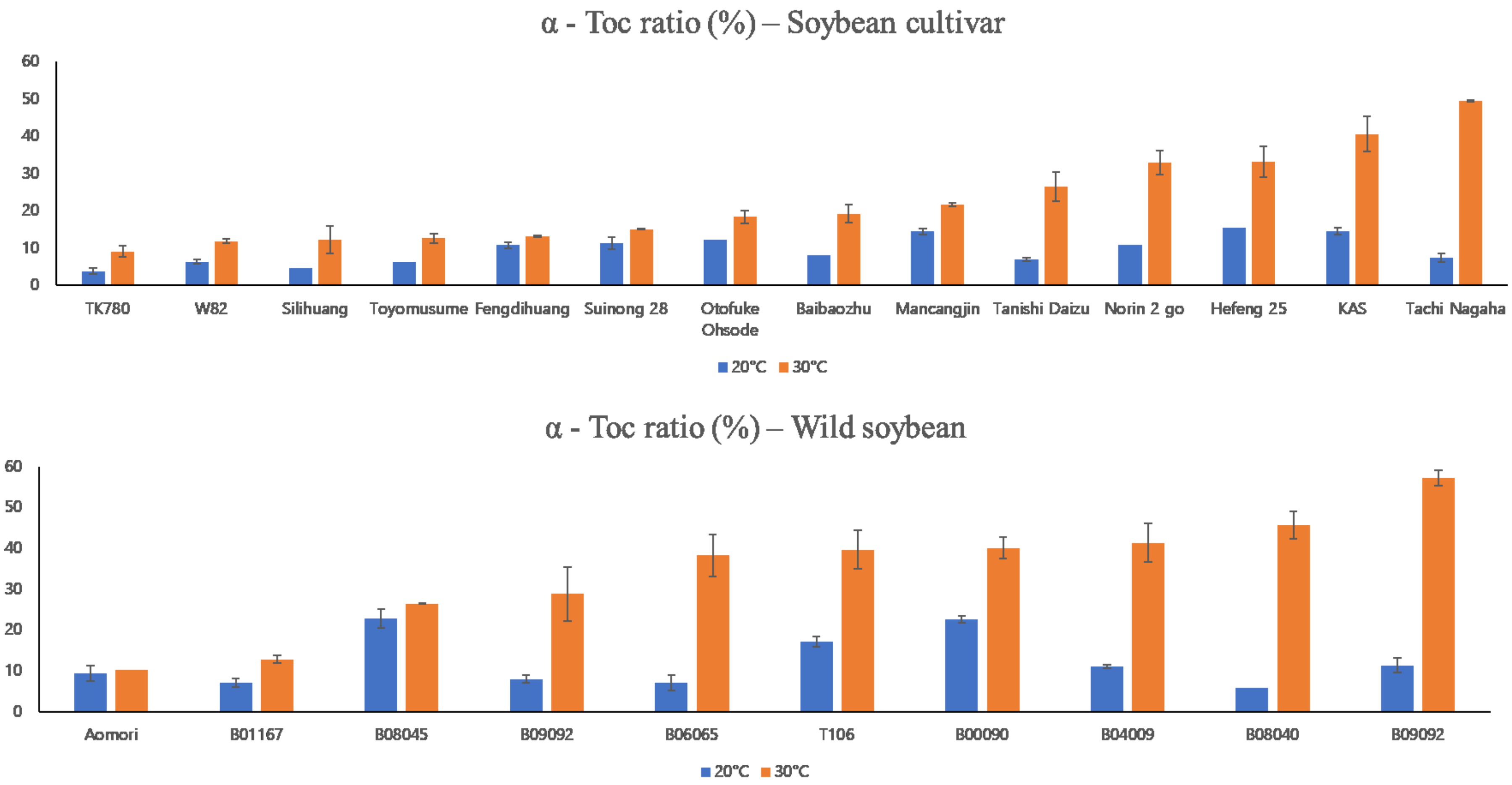


Fig. 13-2 Different responses of α-tocopherol biosynthesis in seeds of various soybean cultivars and wild soybeans to different temperatures during maturation. Blue bar: 20 °C, Orange bar: 30 °C

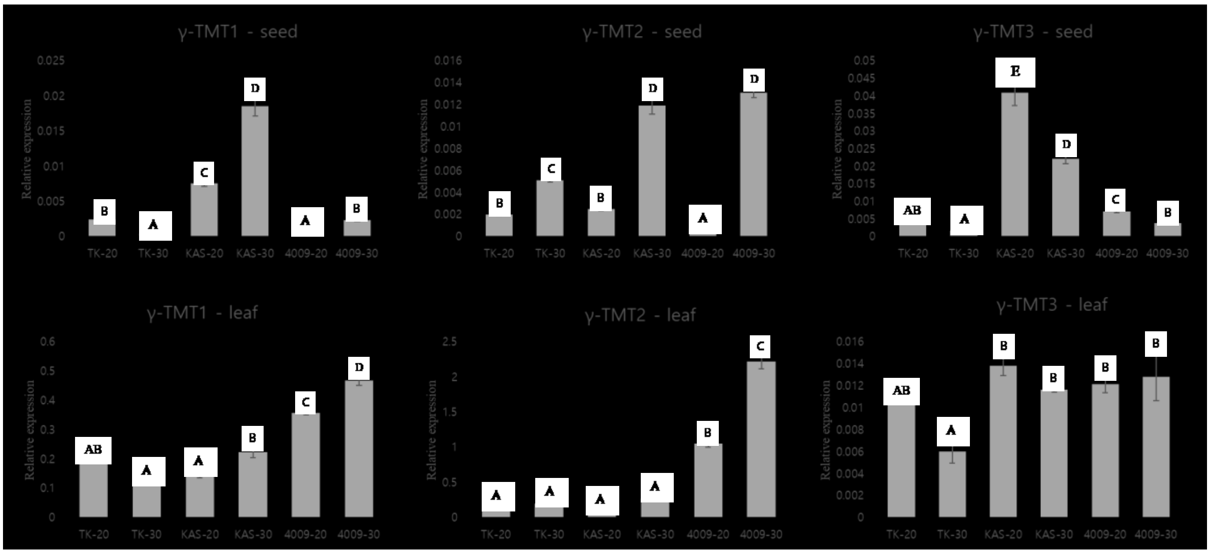


Fig. 14 Expression profiles of three *γ-TMT* genes in leaf and immature cotyledons of plants grown in 20°C and 30°C.

The relative expression is presented using *actin* (Glyma.18G222800.1) as an internal control. The experiments were repeated four times using independently synthesized cDNAs. Error bars; standard errors, Different alphabets indicate statistically significant (at 5 % level) differences between mean values tested by Tukey's HSD.

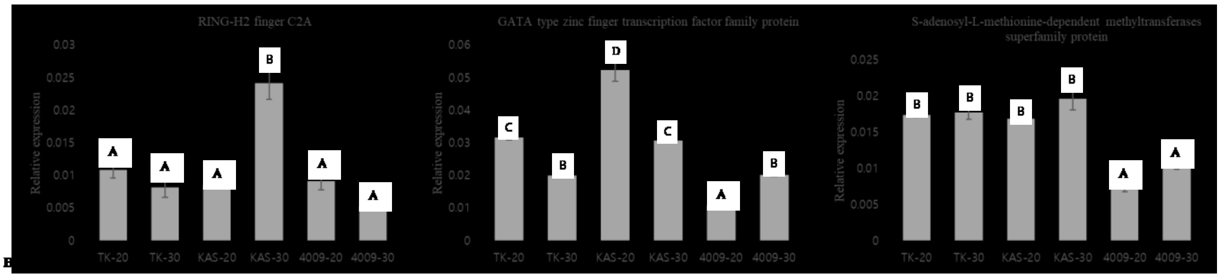


Fig. 15 Expression profiles of three candidate genes for the QTL in chromosome 11 in leaf and immature cotyledons of plants grown in 20°C and 30°C.

The relative expression is presented using *actin* (Glyma.18G222800.1) as an internal control. The experiments were repeated four times using independently synthesized cDNAs. Error bars; standard errors, Different alphabets indicate statistically significant (at 5 % level) differences between mean values tested by Tukey's HSD.

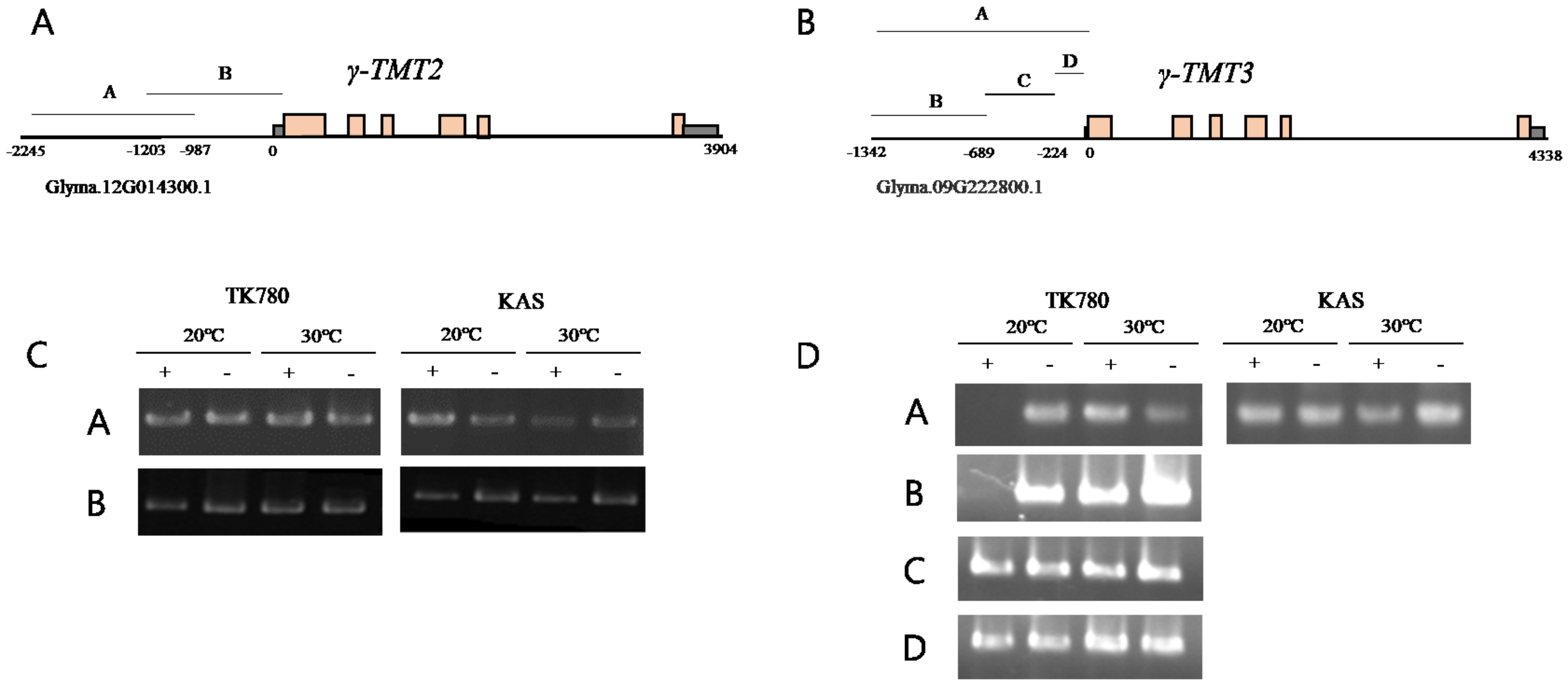
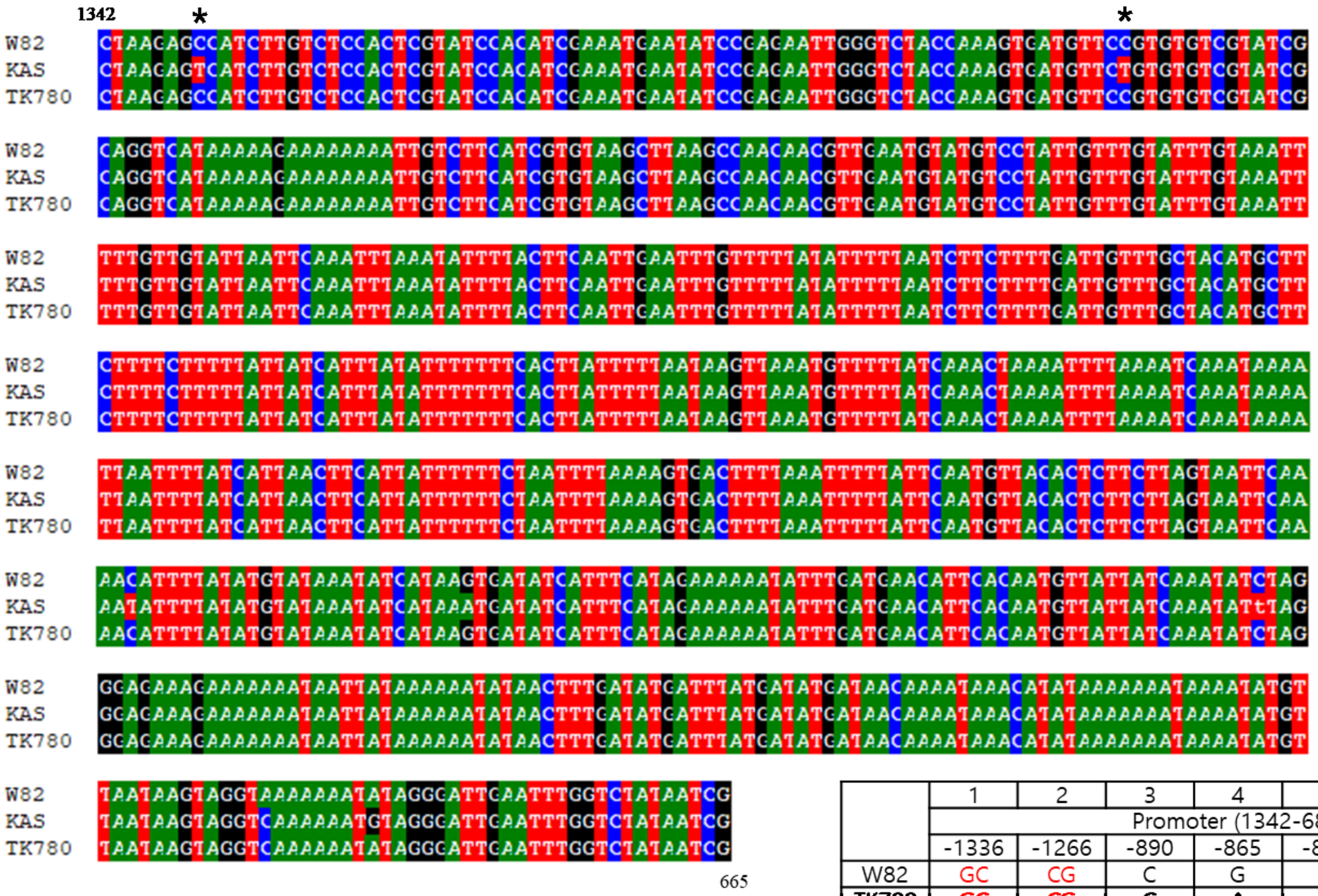


Fig. 16 Analysis of DNA methylation in the promoter regions of *γ-TMT2* and *γ-TMT3* in TK780 and KAS using a methylation-sensitive restriction enzyme McrBC .

A, B) the genomic regions tested in *γ-TMT2* (A) and *γ-TMT3* (B). Two regions (A and B) and four regions (A to D) were tested in *γ-TMT2* and *γ-TMT3*, respectively.

C, D) Amplified products in the genomic regions of *γ-TMT2* (C) and *γ-TMT3* (D). PCR was carried out for the genomic DNAs treated by McrBC (+) and untreated (-). The DNAs were extracted from immature cotyledons of plants grown in 20 ℃ and 30 ℃ conditions. Amplicons were visualized in 1.0 % agarose gels.

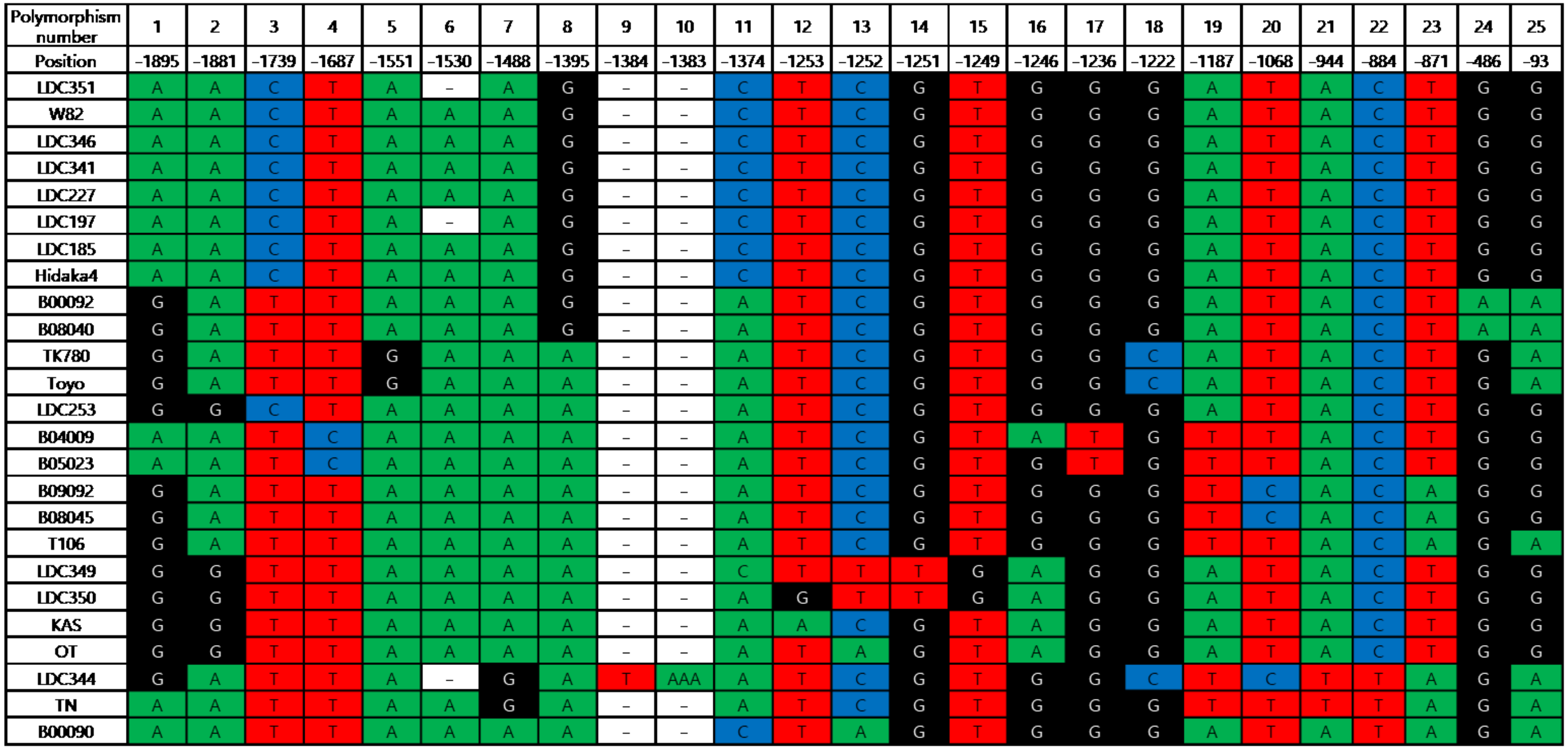


```
1342     *                                                                      *
W82    CTAAGAGCCATCTTGTCTCCACTCGTATCCACATCGAAATGAATATCCGAGAATTGGGTCTACCAAAGTGATGTTCCGTGTGTCGTATCG
KAS    CTAAGAGTCATCTTGTCTCCACTCGTATCCACATCGAAATGAATATCCGAGAATTGGGTCTACCAAAGTGATGTTCTGTGTGTCGTATCG
TK780  CTAAGAGCCATCTTGTCTCCACTCGTATCCACATCGAAATGAATATCCGAGAATTGGGTCTACCAAAGTGATGTTCCGTGTGTCGTATCG

W82    CAGGTCATAAAAACAAAAAAAAATTGTCTTCATCGTGTAAGCTTAAGCCAACAACGTTGAATGTATGTCCTATTGTTTGTATTTGTAAATT
KAS    CAGGTCATAAAAACAAAAAAAAATTGTCTTCATCGTGTAAGCTTAAGCCAACAACGTTGAATGTATGTCCTATTGTTTGTATTTGTAAATT
TK780  CAGGTCATAAAAACAAAAAAAAATTGTCTTCATCGTGTAAGCTTAAGCCAACAACGTTGAATGTATGTCCTATTGTTTGTATTTGTAAATT

W82    TTTGTTGTATTAATTCAAATTTAAATATTTTACTTCAATTGAATTTGTTTTTATATTTTTAATCTTCTTTTGATTGTTTGCTACATGCTT
KAS    TTTGTTGTATTAATTCAAATTTAAATATTTTACTTCAATTGAATTTGTTTTTATATTTTTAATCTTCTTTTGATTGTTTGCTACATGCTT
TK780  TTTGTTGTATTAATTCAAATTTAAATATTTTACTTCAATTGAATTTGTTTTTATATTTTTAATCTTCTTTTGATTGTTTGCTACATGCTT

W82    CTTTTCTTTTTATTATCATTTATATTTTTTTCACTTATTTTTAATAAGTTAAATGTTTTTATCAAACTAAAATTTTAAAATCAAATAAAA
KAS    CTTTTCTTTTTATTATCATTTATATTTTTTTCACTTATTTTTAATAAGTTAAATGTTTTTATCAAACTAAAATTTTAAAATCAAATAAAA
TK780  CTTTTCTTTTTATTATCATTTATATTTTTTTCACTTATTTTTAATAAGTTAAATGTTTTTATCAAACTAAAATTTTAAAATCAAATAAAA

W82    TTAATTTTATCATTAACTTCATTATTTTTTCTAATTTTAAAAGTGACTTTTAAATTTTTATTCAATGTTACACTCTTCTTAGTAATTCAA
KAS    TTAATTTTATCATTAACTTCATTATTTTTTCTAATTTTAAAAGTGACTTTTAAATTTTTATTCAATGTTACACTCTTCTTAGTAATTCAA
TK780  TTAATTTTATCATTAACTTCATTATTTTTTCTAATTTTAAAAGTGACTTTTAAATTTTTATTCAATGTTACACTCTTCTTAGTAATTCAA

W82    AACATTTTATATGTATAAATATCATAAGTGATATCATTTCATAGAAAAAATATTTGATGAACATTCACAATGTTATTATCAAATATCTAG
KAS    AATATTTTATATGTATAAATATCATAAATGATATCATTTCATAGAAAAAATATTTGATGAACATTCACAATGTTATTATCAAATATtTAG
TK780  AACATTTTATATGTATAAATATCATAAGTGATATCATTTCATAGAAAAAATATTTGATGAACATTCACAATGTTATTATCAAATATCTAG

W82    GGAGAAAGAAAAAAAATAATTATAAAAAAATATAACTTTGATATGATTTATGATATGATAACAAAATAAACATATAAAAAAATAAAATATGT
KAS    GGAGAAAGAAAAAAAATAATTATAAAAAAATATAACTTTGATATGATTTATGATATGATAACAAAATAAACATATAAAAAAATAAAATATGT
TK780  GGAGAAAGAAAAAAAATAATTATAAAAAAATATAACTTTGATATGATTTATGATATGATAACAAAATAAACATATAAAAAAATAAAATATGT

W82    TAATAAGTAGGTAAAAAAAATATAGGCATTGAATTTGGTCTATAATCG
KAS    TAATAAGTAGGTCAAAAAAATGTAGGCATTGAATTTGGTCTATAATCG
TK780  TAATAAGTAGGTCAAAAAAATATAGGCATTGAATTTGGTCTATAATCG
                                                    665
```

	1	2	3	4	5	**6**	**7**
	Promoter (1342-689)						
	-1336	-1266	-890	-865	-806	**-700**	**-690**
W82	GC	CG	C	G	C	**A**	**A**
TK780	**GC**	**CG**	**C**	**A**	**C**	**G**	**A**
KAS	**GT**	**TG**	**T**	**A**	**T**	**G**	**G**

Fig. 17 Predicted positions of cytosine methylated in the genomic region A in the promoter of *γ-TMT3* gene.

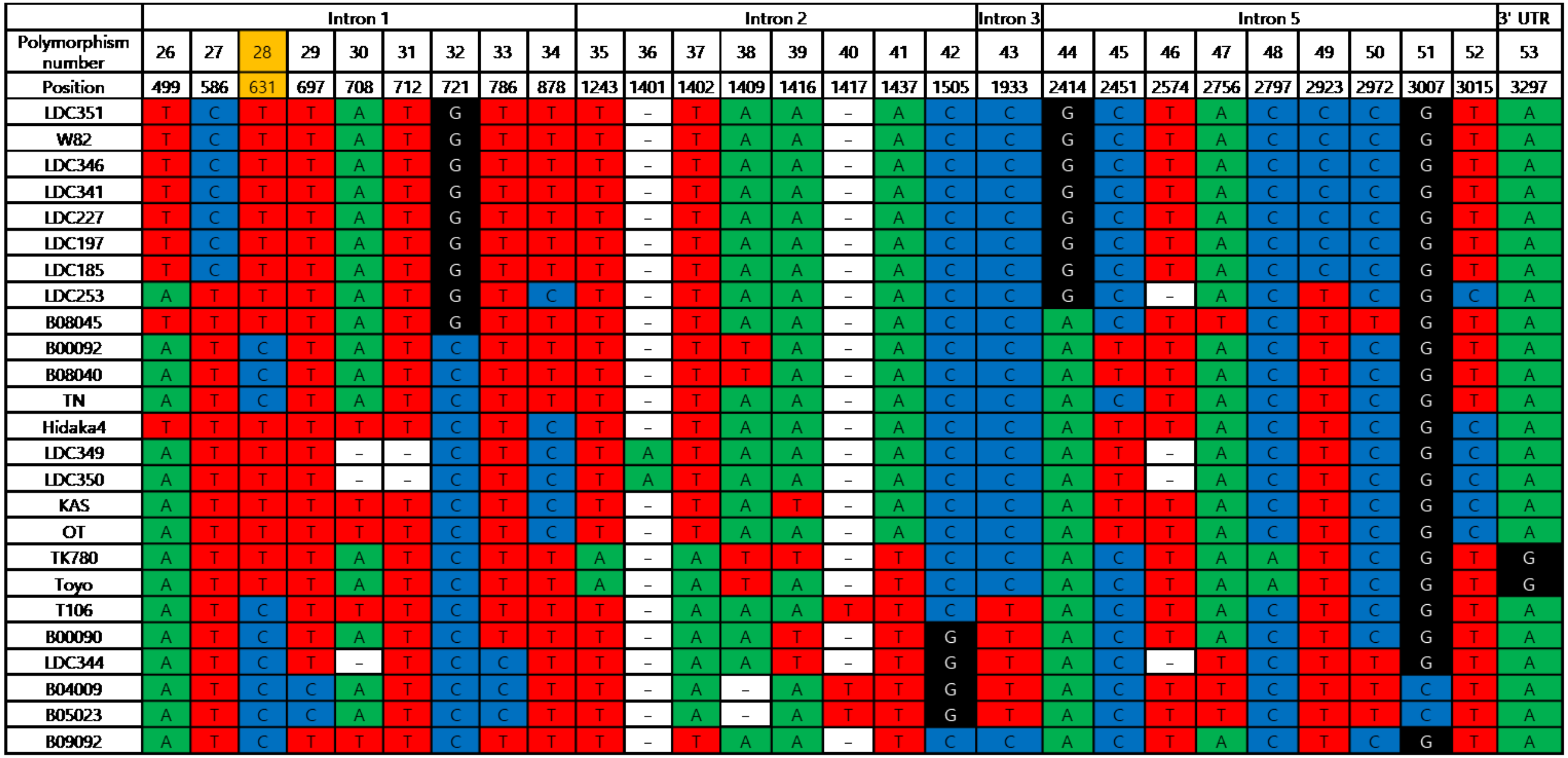
Seven nucleotide substitutions were presented in the target region between TK780 and KAS. Asterisks, cytosine in the TK780 sequence predicted to be methylated



Polymorphism number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Position	–1895	–1881	–1739	–1687	–1551	–1530	–1488	–1395	–1384	–1383	–1374	–1253	–1252	–1251	–1249	–1246	–1236	–1222	–1187	–1068	-944	-884	-871	-486	-93
LDC351	A	A	C	T	A	–	A	G	–	–	C	T	C	G	T	G	G	G	A	T	A	C	T	G	G
W82	A	A	C	T	A	A	A	G	–	–	C	T	C	G	T	G	G	G	A	T	A	C	T	G	G
LDC346	A	A	C	T	A	A	A	G	–	–	C	T	C	G	T	G	G	G	A	T	A	C	T	G	G
LDC341	A	A	C	T	A	A	A	G	–	–	C	T	C	G	T	G	G	G	A	T	A	C	T	G	G
LDC227	A	A	C	T	A	A	A	G	–	–	C	T	C	G	T	G	G	G	A	T	A	C	T	G	G
LDC197	A	A	C	T	A	–	A	G	–	–	C	T	C	G	T	G	G	G	A	T	A	C	T	G	G
LDC185	A	A	C	T	A	A	A	G	–	–	C	T	C	G	T	G	G	G	A	T	A	C	T	G	G
Hidaka4	A	A	C	T	A	A	A	G	–	–	C	T	C	G	T	G	G	G	A	T	A	C	T	G	G
B00092	G	A	T	T	A	A	A	G	–	–	A	T	C	G	T	G	G	G	A	T	A	C	T	A	A
B08040	G	A	T	T	A	A	A	G	–	–	A	T	C	G	T	G	G	G	A	T	A	C	T	A	A
TK780	G	A	T	T	G	A	A	A	–	–	A	T	C	G	T	G	G	C	A	T	A	C	T	G	A
Toyo	G	A	T	T	G	A	A	A	–	–	A	T	C	G	T	G	G	C	A	T	A	C	T	G	A
LDC253	G	G	C	T	A	A	A	A	–	–	A	T	C	G	T	G	G	G	A	T	A	C	T	G	G
B04009	A	A	T	C	A	A	A	A	–	–	A	T	C	G	T	A	T	G	T	T	A	C	T	G	G
B05023	A	A	T	C	A	A	A	A	–	–	A	T	C	G	T	G	T	G	T	T	A	C	T	G	G
B09092	G	A	T	T	A	A	A	A	–	–	A	T	C	G	T	G	G	G	T	C	A	C	A	G	G
B08045	G	A	T	T	A	A	A	A	–	–	A	T	C	G	T	G	G	G	T	C	A	C	A	G	G
T106	G	A	T	T	A	A	A	A	–	–	A	T	C	G	T	G	G	G	T	T	A	C	A	G	A
LDC349	G	G	T	T	A	A	A	A	–	–	C	T	T	T	G	A	G	G	A	T	A	C	T	G	G
LDC350	G	G	T	T	A	A	A	A	–	–	A	G	T	T	G	A	G	G	A	T	A	C	T	G	G
KAS	G	G	T	T	A	A	A	A	–	–	A	A	C	G	T	A	G	G	A	T	A	C	T	G	G
OT	G	G	T	T	A	A	A	A	–	–	A	T	A	G	T	A	G	G	A	T	A	C	T	G	G
LDC344	G	A	T	T	A	–	G	A	T	AAA	A	T	C	G	T	G	G	C	T	C	T	T	A	G	A
TN	A	A	T	T	A	A	G	A	–	–	A	T	C	G	T	G	G	G	T	T	T	T	A	G	A
B00090	A	A	T	T	A	A	A	A	–	–	C	T	A	G	T	G	G	G	A	T	A	T	A	G	A

*** W82, Williams 82, Toyo, Toyomusume, OT, Otofuke Ohsode, TN, Tokachi Nagaha**

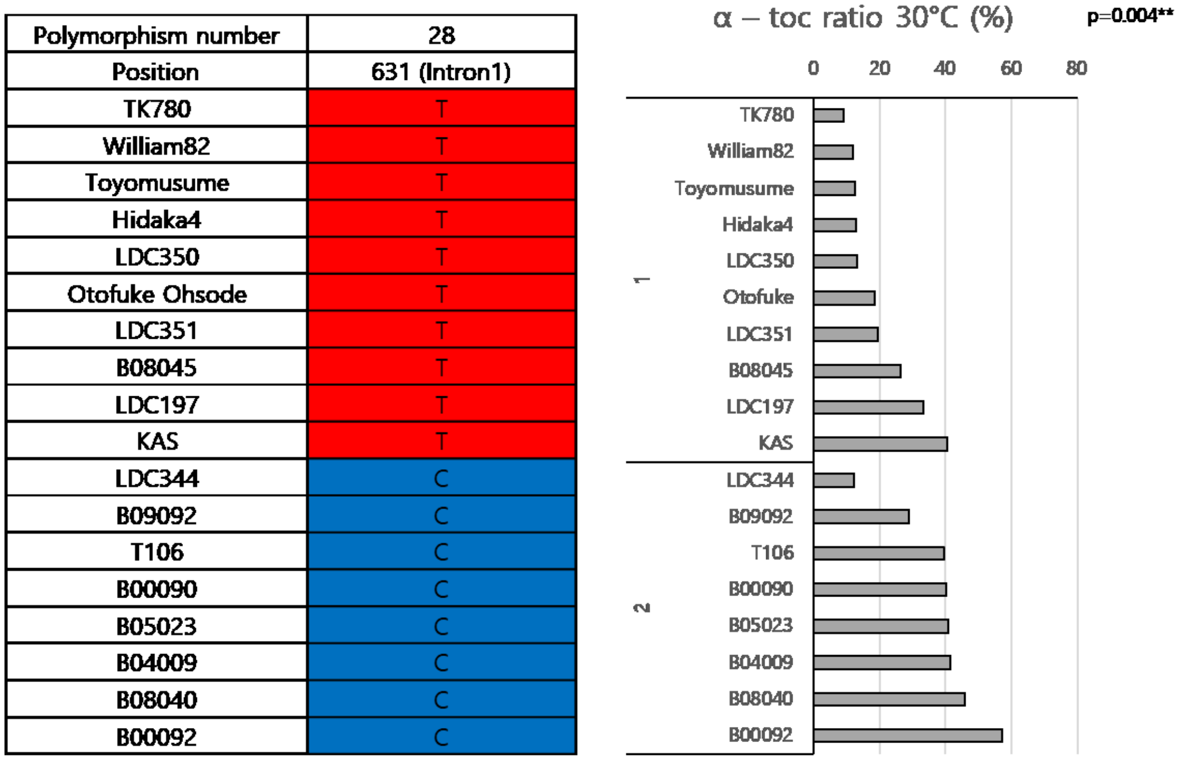
Fig. 18 DNA polymorphism in the promoter region of *γ-TMT1* among 25 wild soybean accessions or soybean cultivars. The positions of nucleotide polymorphisms relative to the adenine of translational start site (ATG) are presented at the second row.



	Intron 1									Intron 2								Intron 3	Intron 5									3' UTR
Polymorphism number	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
Position	499	586	631	697	708	712	721	786	878	1243	1401	1402	1409	1416	1417	1437	1505	1933	2414	2451	2574	2756	2797	2923	2972	3007	3015	3297
LDC351	T	C	T	T	A	T	G	T	T	T	–	T	A	A	–	A	C	C	G	C	T	A	C	C	C	G	T	A
W82	T	C	T	T	A	T	G	T	T	T	–	T	A	A	–	A	C	C	G	C	T	A	C	C	C	G	T	A
LDC346	T	C	T	T	A	T	G	T	T	T	–	T	A	A	–	A	C	C	G	C	T	A	C	C	C	G	T	A
LDC341	T	C	T	T	A	T	G	T	T	T	–	T	A	A	–	A	C	C	G	C	T	A	C	C	C	G	T	A
LDC227	T	C	T	T	A	T	G	T	T	T	–	T	A	A	–	A	C	C	G	C	T	A	C	C	C	G	T	A
LDC197	T	C	T	T	A	T	G	T	T	T	–	T	A	A	–	A	C	C	G	C	T	A	C	C	C	G	T	A
LDC185	T	C	T	T	A	T	G	T	T	T	–	T	A	A	–	A	C	C	G	C	T	A	C	C	C	G	T	A
LDC253	A	T	T	T	A	T	G	T	C	T	–	T	A	A	–	A	C	C	G	C	–	A	C	T	C	G	C	A
B08045	T	T	T	T	A	T	G	T	T	T	–	T	A	A	–	A	C	C	A	C	T	T	C	T	T	G	T	A
B00092	A	T	C	T	A	T	C	T	T	T	–	T	T	A	–	A	C	C	A	T	T	A	C	T	C	G	T	A
B08040	A	T	C	T	A	T	C	T	T	T	–	T	T	A	–	A	C	C	A	T	T	A	C	T	C	G	T	A
TN	A	T	C	T	A	T	C	T	T	T	–	T	A	A	–	A	C	C	A	C	T	A	C	T	C	G	T	A
Hidaka4	T	T	T	T	T	T	C	T	C	T	–	T	A	A	–	A	C	C	A	T	T	A	C	T	C	G	C	A
LDC349	A	T	T	T	–	–	C	T	C	T	A	T	A	A	–	A	C	C	A	T	–	A	C	T	C	G	C	A
LDC350	A	T	T	T	–	–	C	T	C	T	A	T	A	A	–	A	C	C	A	T	–	A	C	T	C	G	C	A
KAS	A	T	T	T	T	T	C	T	C	T	–	T	A	T	–	A	C	C	A	T	T	A	C	T	C	G	C	A
OT	A	T	T	T	T	T	C	T	C	T	–	T	A	A	–	A	C	C	A	T	T	A	C	T	C	G	C	A
TK780	A	T	T	T	A	T	C	T	T	A	–	A	T	T	–	T	C	C	A	C	T	A	A	T	C	G	T	G
Toyo	A	T	T	T	A	T	C	T	T	A	–	A	T	A	–	T	C	C	A	C	T	A	A	T	C	G	T	G
T106	A	T	C	T	T	T	C	T	T	T	–	A	A	A	T	T	C	T	A	C	T	A	C	T	C	G	T	A
B00090	A	T	C	T	A	T	C	T	T	T	–	A	A	T	–	T	G	T	A	C	T	A	C	T	C	G	T	A
LDC344	A	T	C	T	–	T	C	C	T	T	–	A	A	T	–	T	G	T	A	C	–	T	C	T	T	G	T	A
B04009	A	T	C	C	A	T	C	C	T	T	–	A	–	A	T	T	G	T	A	C	T	T	C	T	T	C	T	A
B05023	A	T	C	C	A	T	C	C	T	T	–	A	–	A	T	T	G	T	A	C	T	T	C	T	T	C	T	A
B09092	A	T	C	T	T	T	C	T	T	T	–	T	A	A	–	T	C	C	A	C	T	A	C	T	C	G	T	A

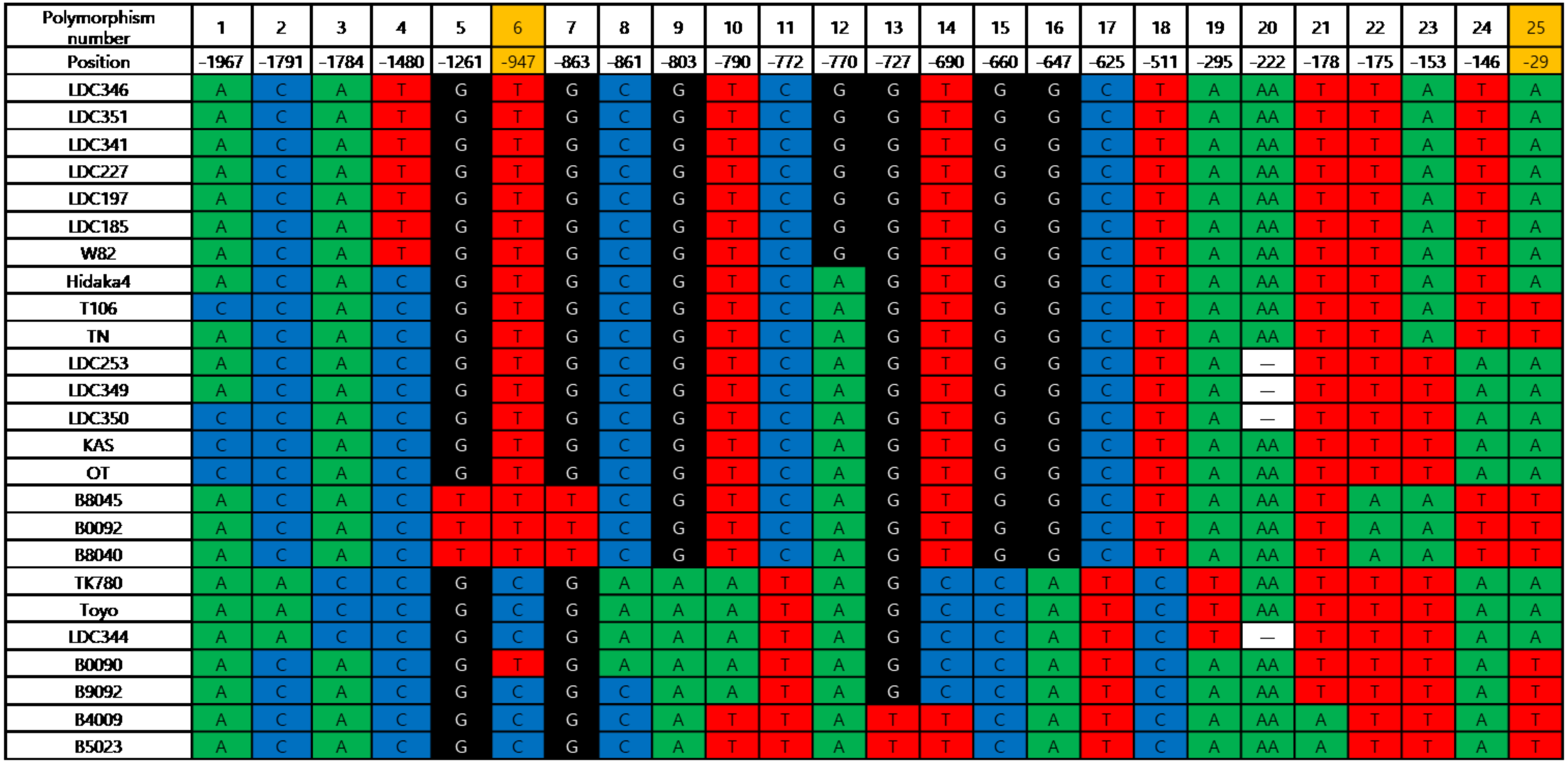
*** W82, Williams 82, Toyo, Toyomusume, OT, Otofuke Ohsode, TN, Tokachi Nagaha**

Fig. 19 DNA polymorphisms in the genic region of 5′ UTR to 3′ UTR in *γ-TMT1* among 25 wild soybean accessions and soybean cultivars. The positions of nucleotide polymorphisms relative to the adenine of translational start site (ATG) are presented at the second row.



Polymorphism number	28
Position	631 (Intron1)
TK780	T
William82	T
Toyomusume	T
Hidaka4	T
LDC350	T
Otofuke Ohsode	T
LDC351	T
B08045	T
LDC197	T
KAS	T
LDC344	C
B09092	C
T106	C
B00090	C
B05023	C
B04009	C
B08040	C
B00092	C

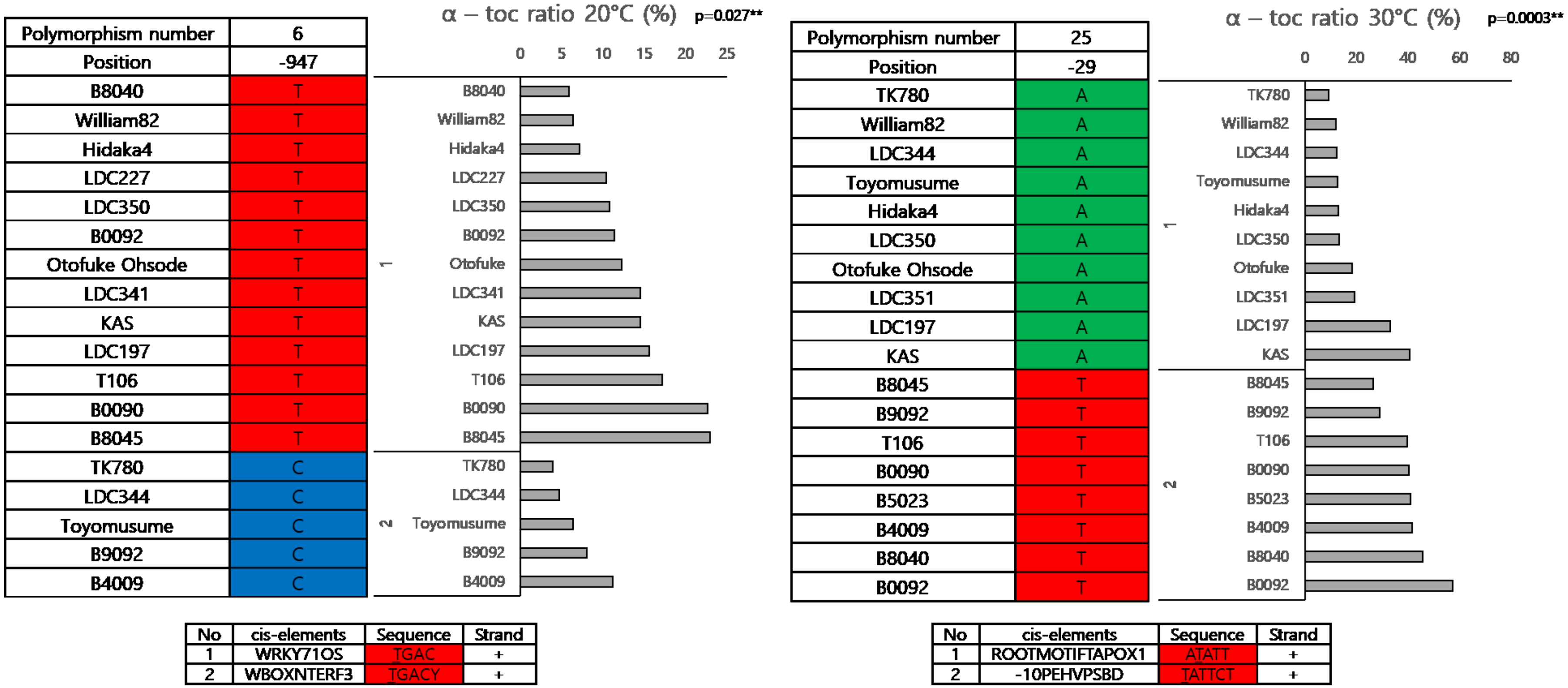
Fig. 20 Significant association between α-tocopherol ratios and a SNP in the intron of *γ-TMT1* gene. The accessions tested were classified based on the nucleotides at SNPs. The association between the SNP and the α-tocopherol ratios were analyzed by one-way analysis of variance.



Polymorphism number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Position	-1967	-1791	-1784	-1480	-1261	-947	-863	-861	-803	-790	-772	-770	-727	-690	-660	-647	-625	-511	-295	-222	-178	-175	-153	-146	-29
LDC346	A	C	A	T	G	T	G	C	G	T	C	G	G	T	G	G	C	T	A	AA	T	T	A	T	A
LDC351	A	C	A	T	G	T	G	C	G	T	C	G	G	T	G	G	C	T	A	AA	T	T	A	T	A
LDC341	A	C	A	T	G	T	G	C	G	T	C	G	G	T	G	G	C	T	A	AA	T	T	A	T	A
LDC227	A	C	A	T	G	T	G	C	G	T	C	G	G	T	G	G	C	T	A	AA	T	T	A	T	A
LDC197	A	C	A	T	G	T	G	C	G	T	C	G	G	T	G	G	C	T	A	AA	T	T	A	T	A
LDC185	A	C	A	T	G	T	G	C	G	T	C	G	G	T	G	G	C	T	A	AA	T	T	A	T	A
W82	A	C	A	T	G	T	G	C	G	T	C	G	G	T	G	G	C	T	A	AA	T	T	A	T	A
Hidaka4	A	C	A	C	G	T	G	C	G	T	C	A	G	T	G	G	C	T	A	AA	T	T	A	T	A
T106	C	C	A	C	G	T	G	C	G	T	C	A	G	T	G	G	C	T	A	AA	T	T	A	T	T
TN	A	C	A	C	G	T	G	C	G	T	C	A	G	T	G	G	C	T	A	AA	T	T	A	T	T
LDC253	A	C	A	C	G	T	G	C	G	T	C	A	G	T	G	G	C	T	A	–	T	T	T	A	A
LDC349	A	C	A	C	G	T	G	C	G	T	C	A	G	T	G	G	C	T	A	–	T	T	T	A	A
LDC350	C	C	A	C	G	T	G	C	G	T	C	A	G	T	G	G	C	T	A	–	T	T	T	A	A
KAS	C	C	A	C	G	T	G	C	G	T	C	A	G	T	G	G	C	T	A	AA	T	T	T	A	A
OT	C	C	A	C	G	T	G	C	G	T	C	A	G	T	G	G	C	T	A	AA	T	T	T	A	A
B8045	A	C	A	C	T	T	T	C	G	T	C	A	G	T	G	G	C	T	A	AA	T	A	A	T	T
B0092	A	C	A	C	T	T	T	C	G	T	C	A	G	T	G	G	C	T	A	AA	T	A	A	T	T
B8040	A	C	A	C	T	T	T	C	G	T	C	A	G	T	G	G	C	T	A	AA	T	A	A	T	T
TK780	A	A	C	C	G	C	G	A	A	A	T	A	G	C	C	A	T	C	T	AA	T	T	T	A	A
Toyo	A	A	C	C	G	C	G	A	A	A	T	A	G	C	C	A	T	C	T	AA	T	T	T	A	A
LDC344	A	A	C	C	G	C	G	A	A	A	T	A	G	C	C	A	T	C	T	–	T	T	T	A	A
B0090	A	C	A	C	G	T	G	A	A	A	T	A	G	C	C	A	T	C	A	AA	T	T	T	A	T
B9092	A	C	A	C	G	C	G	C	A	A	T	A	G	C	C	A	T	C	A	AA	T	T	T	A	T
B4009	A	C	A	C	G	C	G	C	A	T	T	A	T	T	C	A	T	C	A	AA	A	T	T	A	T
B5023	A	C	A	C	G	C	G	C	A	T	T	A	T	T	C	A	T	C	A	AA	A	T	T	A	T

* W82, Williams 82, Toyo, Toyomusume, OT, Otofuke Ohsode, TN, Tokachi Nagaha

Fig. 21 DNA polymorphism in the promoter region of *γ-TMT2* among 25 wild soybean accessions or soybean cultivars. The positions of nucleotide polymorphisms relative to the adenine of translational start site (ATG) are presented at the second row.



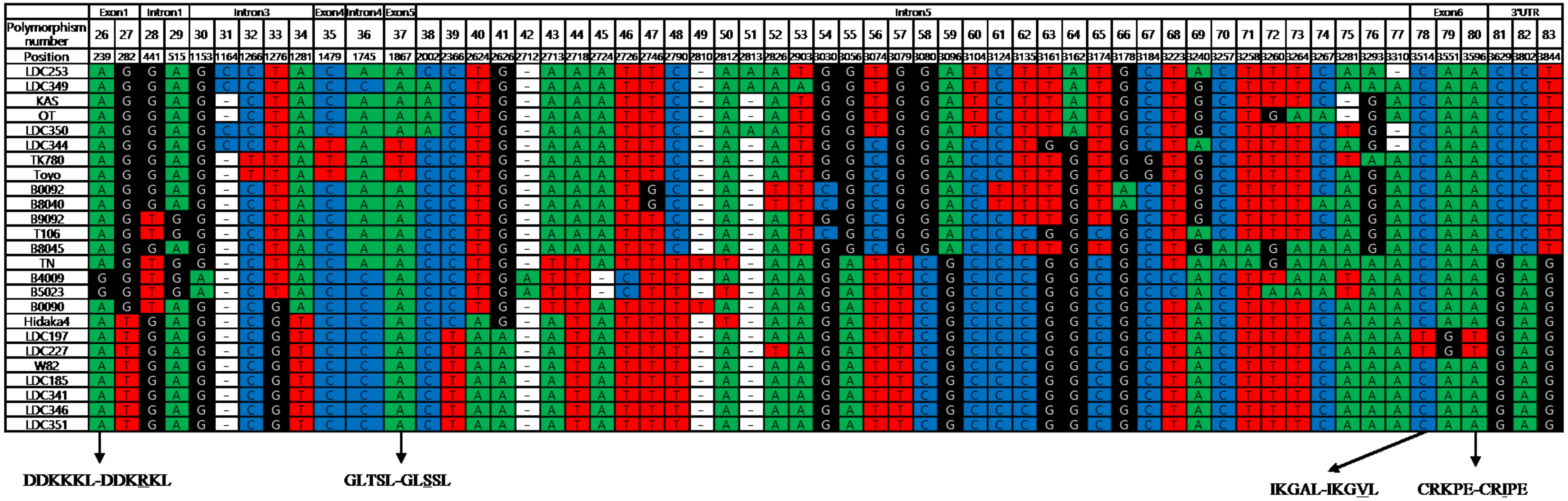
Polymorphism number	6
Position	-947
B8040	T
William82	T
Hidaka4	T
LDC227	T
LDC350	T
B0092	T
Otofuke Ohsode	T
LDC341	T
KAS	T
LDC197	T
T106	T
B0090	T
B8045	T
TK780	C
LDC344	C
Toyomusume	C
B9092	C
B4009	C

No	cis-elements	Sequence	Strand
1	WRKY71OS	TGAC	+
2	WBOXNTERF3	TGACY	+

Polymorphism number	25
Position	-29
TK780	A
William82	A
LDC344	A
Toyomusume	A
Hidaka4	A
LDC350	A
Otofuke Ohsode	A
LDC351	A
LDC197	A
KAS	A
B8045	T
B9092	T
T106	T
B0090	T
B5023	T
B4009	T
B8040	T
B0092	T

No	cis-elements	Sequence	Strand
1	ROOTMOTIFTAPOX1	ATATT	+
2	-10PEHVPSBD	TATTCT	+

Fig. 22 Significant association between α-tocopherol ratios and a SNP in promoter of *γ-TMT2* gene. The accessions tested were classified based on the nucleotides at SNPs. The association between the SNP and the α-tocopherol ratios were analyzed by one-way analysis of variance.

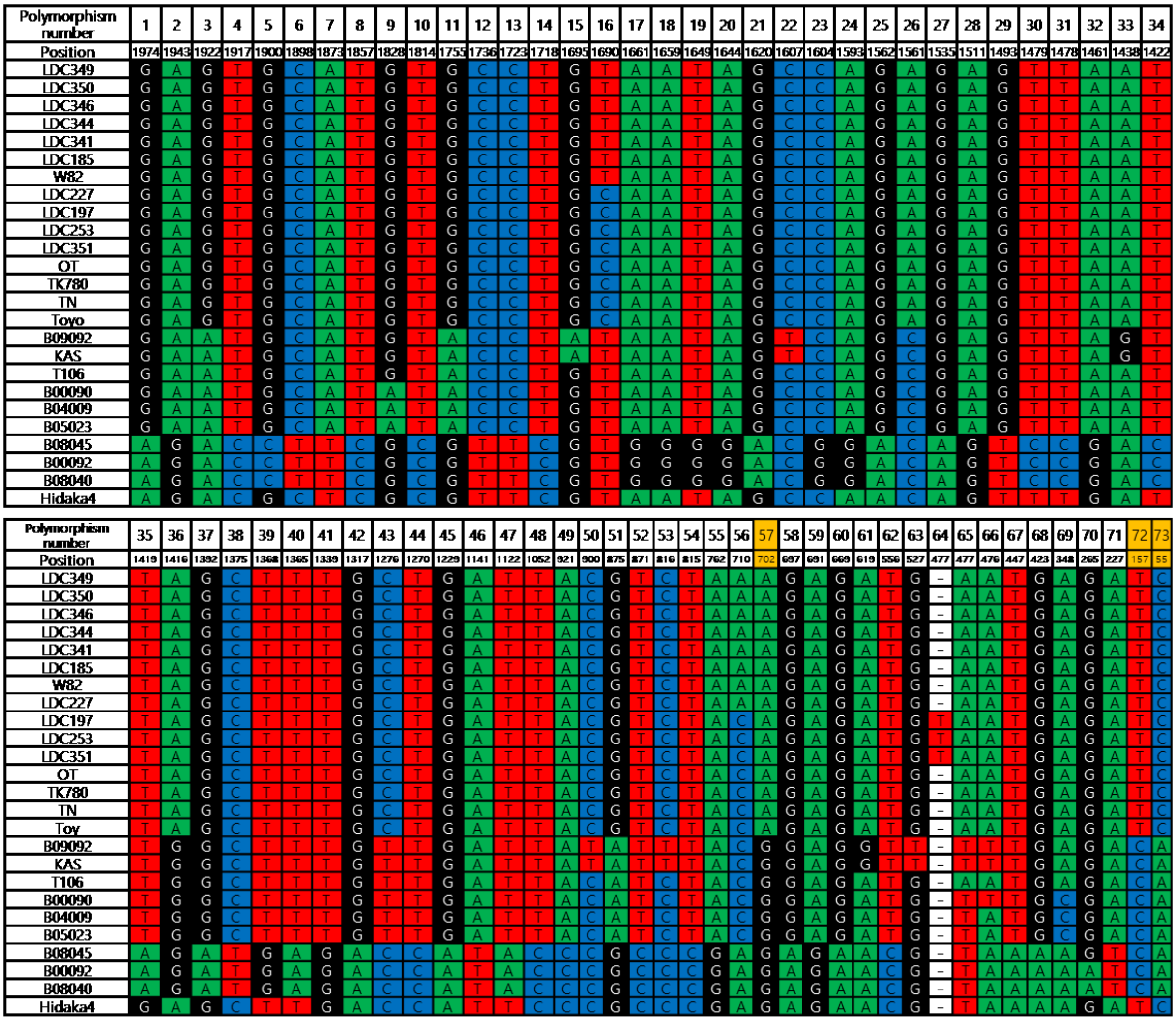


	Exon1		Intron1		Intron3					Exon4	Intron4	Exon5	Intron5																
Polymorphism number	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Position	239	282	441	515	1153	1164	1266	1276	1281	1479	1745	1867	2002	2366	2624	2626	2712	2713	2718	2724	2726	2746	2790	2810	2812	2813	2826	2903	3030
LDC253	A	G	G	A	G	C	C	T	A	C	A	A	C	C	T	G	–	A	A	A	T	T	C	–	A	A	A	T	G
LDC349	A	G	G	A	G	C	C	T	A	C	C	A	A	C	T	G	–	A	A	A	T	T	C	–	A	A	A	A	G
KAS	A	G	G	A	G	–	C	T	A	C	A	A	A	C	T	G	–	A	A	A	T	T	C	–	A	–	A	T	G
OT	A	G	G	A	G	–	C	T	A	C	A	A	A	C	T	G	–	A	A	A	T	T	C	–	A	–	A	T	G
LDC350	A	G	G	A	G	C	C	T	A	C	A	A	A	C	T	G	–	A	A	A	T	T	C	–	A	A	A	T	G
LDC344	A	G	G	A	G	C	C	T	A	T	A	T	C	C	T	G	–	A	A	A	T	T	C	–	A	–	A	T	G
TK780	A	G	G	A	G	–	T	T	A	T	A	T	C	C	T	G	–	A	A	A	T	T	C	–	A	–	A	T	G
Toyo	A	G	G	A	G	–	T	T	A	T	A	T	C	C	T	G	–	A	A	A	T	T	C	–	A	–	A	T	G
B0092	A	G	G	A	G	–	C	T	A	C	A	A	C	C	T	G	–	A	A	A	T	G	C	–	A	–	T	T	C
B8040	A	G	G	A	G	–	C	T	A	C	A	A	C	C	T	G	–	A	A	A	T	G	C	–	A	–	T	T	C
B9092	A	G	T	G	G	–	C	T	A	C	A	A	C	C	T	G	–	A	A	A	T	T	C	–	A	–	A	T	G
T106	A	G	T	G	G	–	C	T	A	C	A	A	C	C	T	G	–	A	A	A	T	T	C	–	A	–	A	T	C
B8045	A	G	G	A	G	–	C	T	A	C	A	A	C	C	T	G	–	A	A	A	T	T	C	–	A	–	A	T	G
TN	A	G	T	G	G	–	C	T	A	C	A	A	C	C	T	G	–	T	T	A	T	T	T	T	T	–	A	A	G
B4009	G	G	T	G	A	–	C	T	A	C	C	A	C	C	T	G	A	T	T	–	C	T	T	–	A	–	A	A	G
B5023	G	G	T	G	A	–	C	T	A	C	C	A	C	C	T	G	A	T	T	–	C	T	T	–	T	–	A	A	G
B0090	A	G	T	A	G	–	C	G	A	C	C	A	C	C	T	G	–	T	T	A	T	T	T	T	A	–	A	A	G
Hidaka4	A	T	G	A	G	–	C	G	T	C	C	A	C	C	A	G	–	A	T	A	T	T	T	–	T	–	A	A	G
LDC197	A	T	G	A	G	–	C	G	T	C	C	A	C	T	A	A	–	A	T	A	T	T	T	–	A	–	A	A	G
LDC227	A	T	G	A	G	–	C	G	T	C	C	A	C	T	A	A	–	A	T	A	T	T	T	–	A	–	T	A	G
W82	A	T	G	A	G	–	C	G	T	C	C	A	C	T	A	A	–	A	T	A	T	T	T	–	A	–	A	A	G
LDC185	A	T	G	A	G	–	C	G	T	C	C	A	C	T	A	A	–	A	T	A	T	T	T	–	A	–	A	A	G
LDC341	A	T	G	A	G	–	C	G	T	C	C	A	C	T	A	A	–	A	T	A	T	T	T	–	A	–	A	A	G
LDC346	A	T	G	A	G	–	C	G	T	C	C	A	C	T	A	A	–	A	T	A	T	T	T	–	A	–	A	A	G
LDC351	A	T	G	A	G	–	C	G	T	C	C	A	C	T	A	A	–	A	T	A	T	T	T	–	A	–	A	A	G

	Intron5																							Exon6			3'UTR		
Polymorphism number	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83
Position	3056	3074	3079	3080	3096	3104	3124	3135	3161	3162	3174	3178	3184	3223	3240	3257	3258	3260	3264	3267	3281	3293	3310	3514	3551	3596	3629	3802	3844
LDC253	G	T	G	G	A	T	C	T	T	A	T	G	C	T	A	C	T	T	T	C	A	A	–	C	A	A	C	C	T
LDC349	G	T	G	G	A	T	C	T	T	A	T	G	C	T	G	C	T	T	T	C	A	A	A	C	A	A	C	C	T
KAS	G	T	G	G	A	T	C	T	T	A	T	G	C	T	G	C	T	T	T	C	–	G	A	C	A	A	C	C	T
OT	G	T	G	G	A	T	C	T	T	A	T	G	C	T	G	C	T	G	A	A	–	G	A	C	A	A	C	C	T
LDC350	G	T	G	G	A	T	C	T	T	A	T	G	C	T	G	C	T	T	T	C	T	G	–	C	A	A	C	C	T
LDC344	G	C	G	G	A	C	C	T	G	G	T	G	C	T	A	C	T	T	T	C	A	G	–	C	A	A	C	C	T
TK780	G	C	G	G	A	C	C	T	T	G	T	G	G	T	G	C	T	T	T	C	T	A	A	C	A	A	C	C	T
Toyo	G	C	G	G	A	C	C	T	T	G	T	G	G	T	G	C	T	T	T	C	A	G	A	C	A	A	C	C	T
B0092	G	C	G	G	A	C	T	T	T	G	T	A	C	T	G	C	T	T	T	C	A	G	A	C	A	A	C	C	T
B8040	G	C	G	G	A	C	T	T	T	G	T	A	C	T	G	C	T	T	T	A	A	G	A	C	A	A	C	C	T
B9092	G	C	G	G	A	C	C	T	T	G	T	G	C	T	G	C	T	T	T	C	A	G	A	C	A	A	C	C	T
T106	G	C	G	G	A	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	G	A	C	A	A	C	C	T
B8045	G	C	G	G	A	C	C	T	T	G	T	G	C	T	G	A	A	G	A	A	A	G	A	C	A	A	C	C	T
TN	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	A	A	G	A	A	A	A	A	C	A	A	G	A	G
B4009	A	T	T	C	G	C	C	C	G	G	C	G	C	C	A	C	T	T	A	A	T	A	A	C	A	A	G	A	G
B5023	A	T	T	C	G	C	C	C	G	G	C	G	C	C	A	C	T	A	A	A	T	A	A	C	A	A	G	A	G
B0090	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	A	A	C	A	A	G	A	G
Hidaka4	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	A	A	C	A	A	G	A	G
LDC197	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	A	A	T	G	T	G	A	G
LDC227	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	A	A	T	G	T	G	A	G
W82	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	A	A	C	A	A	G	A	G
LDC185	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	A	A	C	A	A	G	A	G
LDC341	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	A	A	C	A	A	G	A	G
LDC346	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	A	A	C	A	A	G	A	G
LDC351	A	T	T	C	G	C	C	C	G	G	C	G	C	T	A	C	T	T	T	C	A	A	A	C	A	A	G	A	G

*** W82, Williams 82, Toyo, Toyomusume, OT, Otofuke Ohsode, TN, Tokachi Nagaha**

Fig. 23 DNA polymorphisms in the genic region of 5′ UTR to 3′ UTR in *γ-TMT2* among 25 wild soybean accessions and soybean cultivars. The positions of nucleotide polymorphisms relative to the adenine of translational start site (ATG) are presented at the second row.



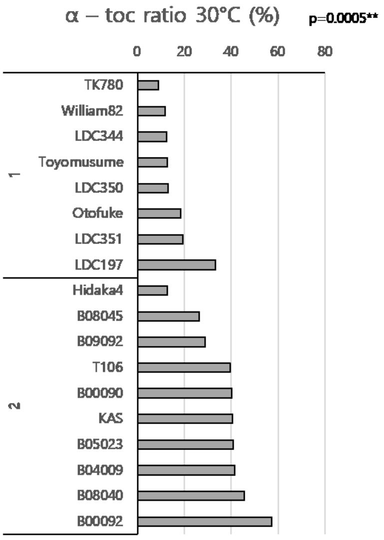
Polymorphism number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Position	1974	1943	1922	1917	1900	1898	1873	1857	1828	1814	1755	1736	1723	1718	1695	1690	1661	1659	1649	1644	1620	1607	1604	1593	1562	1561	1535	1511	1493	1479	1478	1461	1438	1422
LDC349	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	T	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
LDC350	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	T	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
LDC346	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	T	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
LDC344	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	T	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
LDC341	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	T	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
LDC185	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	T	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
W82	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	T	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
LDC227	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	C	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
LDC197	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	C	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
LDC253	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	C	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
LDC351	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	C	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
OT	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	C	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
TK780	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	C	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
TN	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	C	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
Toyo	G	A	G	T	G	C	A	T	G	T	G	C	C	T	G	C	A	A	T	A	G	C	C	A	G	A	G	A	G	T	T	A	A	T
B09092	G	A	A	T	G	C	A	T	G	T	A	C	C	T	A	T	A	A	T	A	G	T	C	A	G	C	G	A	G	T	T	A	G	T
KAS	G	A	A	T	G	C	A	T	G	T	A	C	C	T	A	T	A	A	T	A	G	T	C	A	G	C	G	A	G	T	T	A	G	T
T106	G	A	A	T	G	C	A	T	G	T	A	C	C	T	G	T	A	A	T	A	G	C	C	A	G	C	G	A	G	T	T	A	A	T
B00090	G	A	A	T	G	C	A	T	A	T	A	C	C	T	G	T	A	A	T	A	G	C	C	A	G	C	G	A	G	T	T	A	A	T
B04009	G	A	A	T	G	C	A	T	A	T	A	C	C	T	G	T	A	A	T	A	G	C	C	A	G	C	G	A	G	T	T	A	A	T
B05023	G	A	A	T	G	C	A	T	A	T	A	C	C	T	G	T	A	A	T	A	G	C	C	A	G	C	G	A	G	T	T	A	A	T
B08045	A	G	A	C	C	T	T	C	G	C	G	T	T	C	G	T	G	G	G	G	A	C	G	G	A	C	A	G	T	C	C	G	A	C
B00092	A	G	A	C	C	T	T	C	G	C	G	T	T	C	G	T	G	G	G	G	A	C	G	G	A	C	A	G	T	C	C	G	A	C
B08040	A	G	A	C	C	T	T	C	G	C	G	T	T	C	G	T	G	G	G	G	A	C	G	G	A	C	A	G	T	C	C	G	A	C
Hidaka4	A	G	A	C	G	C	T	C	G	C	G	T	T	C	G	T	A	A	T	A	G	C	C	A	A	C	A	G	T	T	T	G	A	T

Polymorphism number	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Position	1419	1416	1392	1375	1368	1365	1339	1317	1276	1270	1229	1141	1122	1052	921	900	875	871	816	815	762	710	702	697	691	669	619	556	527	477	477	476	447	423	348	265	227	157	55
LDC349	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	A	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
LDC350	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	A	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
LDC346	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	A	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
LDC344	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	A	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
LDC341	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	A	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
LDC185	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	A	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
W82	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	A	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
LDC227	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	A	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
LDC197	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	C	A	G	A	G	A	T	G	T	A	A	T	G	A	G	A	T	C
LDC253	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	C	A	G	A	G	A	T	G	T	A	A	T	G	A	G	A	T	C
LDC351	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	C	A	G	A	G	A	T	G	T	A	A	T	G	A	G	A	T	C
OT	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	C	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
TK780	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	C	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
TN	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	C	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
Toy	T	A	G	C	T	T	T	G	C	T	G	A	T	T	A	C	G	T	C	T	A	C	A	G	A	G	A	T	G	-	A	A	T	G	A	G	A	T	C
B09092	T	G	G	C	T	T	T	G	T	T	G	A	T	T	A	T	A	T	T	T	A	C	G	G	A	G	G	T	T	-	T	T	T	G	A	G	A	C	A
KAS	T	G	G	C	T	T	T	G	T	T	G	A	T	T	A	T	A	T	T	T	A	C	G	G	A	G	G	T	T	-	T	T	T	G	A	G	A	C	A
T106	T	G	G	C	T	T	T	G	T	T	G	A	T	T	A	C	A	T	C	T	A	C	G	G	A	G	A	T	G	-	A	A	T	G	A	G	A	C	A
B00090	T	G	G	C	T	T	T	G	T	T	G	A	T	T	A	C	A	T	C	T	A	C	G	G	A	G	A	T	G	-	T	T	T	G	C	G	A	C	A
B04009	T	G	G	C	T	T	T	G	T	T	G	A	T	T	A	C	A	T	C	T	A	C	G	G	A	G	A	T	G	-	T	A	T	G	C	G	A	C	A
B05023	T	G	G	C	T	T	T	G	T	T	G	A	T	T	A	C	A	T	C	T	A	C	G	G	A	G	A	T	G	-	T	A	T	G	C	G	A	C	A
B08045	A	G	A	T	G	A	G	A	C	C	A	T	A	C	C	C	G	C	C	C	G	A	G	A	G	A	A	C	G	-	T	A	A	A	A	G	T	C	A
B00092	A	G	A	T	G	A	G	A	C	C	A	T	A	C	C	C	G	C	C	C	G	A	G	A	G	A	A	C	G	-	T	A	A	A	A	A	T	C	A
B08040	A	G	A	T	G	A	G	A	C	C	A	T	A	C	C	C	G	C	C	C	G	A	G	A	G	A	A	C	G	-	T	A	A	A	A	A	T	C	A
Hidaka4	G	A	G	C	T	T	G	A	C	C	A	T	T	C	C	C	G	C	C	C	G	A	G	A	G	A	A	C	G	-	T	A	A	A	A	G	A	T	C

*** W82, Williams 82, Toyo, Toyomusume, OT, Otofuke Ohsode, TN, Tokachi Nagaha**

Fig. 24 DNA polymorphism in the promoter region of γ-TMT3 among 25 wild soybean accessions or soybean cultivars. The positions of nucleotide polymorphisms relative to the adenine of translational start site (ATG) are presented at the second row.

Polymorphism number	3	26	57
Position	1922	1561	702
TK780	G	A	A
William82	G	A	A
LDC344	G	A	A
Toyomusume	G	A	A
LDC350	G	A	A
Otofuke Ohsode	G	A	A
LDC351	G	A	A
LDC197	G	A	A
Hidaka4	A	C	G
B08045	A	C	G
B09092	A	C	G
T106	A	C	G
B00090	A	C	G
KAS	A	C	G
B05023	A	C	G
B04009	A	C	G
B08040	A	C	G
B00092	A	C	G

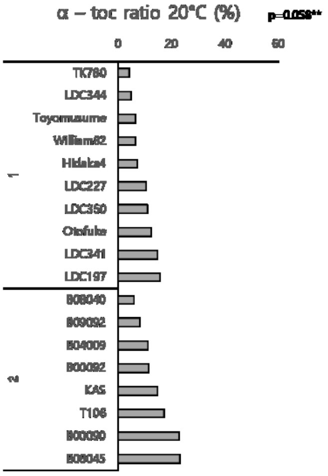


Polymorphism	No	cis-elements	Sequence	Strand
3	1	SITEIIATCYTC	TGGGCY	+

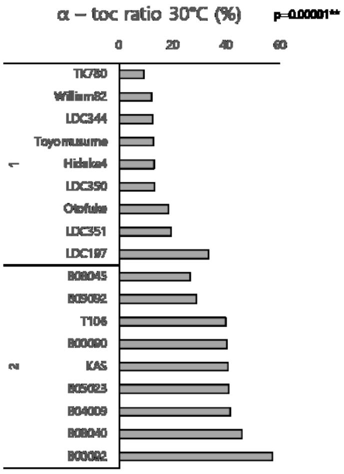
Polymorphism	No	cis-elements	Sequence	Strand
26	1	GT1CONSENSUS	GRWAAW	+
	2	SORLIP1AT	GCCAC	-

Polymorphism	No	cis-elements	Sequence	Strand
57	1	ROOTMOTIFTAPOX1	ATATT	-

Polymorphism number	36	72	73
Position	1416	157	55
TK780	A	T	C
LDC344	A	T	C
Toyomusume	A	T	C
William82	A	T	C
Hidaka4	A	T	C
LDC227	A	T	C
LDC350	A	T	C
Otofuke Ohsode	A	T	C
LDC341	A	T	C
LDC197	A	T	C
B08040	G	C	A
B09092	G	C	A
B04009	G	C	A
B00092	G	C	A
KAS	G	C	A
T106	G	C	A
B00090	G	C	A
B08045	G	C	A



Polymorphism number	36	72	73
Position	1416	157	55
TK780	A	T	C
William82	A	T	C
LDC344	A	T	C
Toyomusume	A	T	C
Hidaka4	A	T	C
LDC350)	A	T	C
Otofuke Ohsode	A	T	C
LDC351	A	T	C
LDC197	A	T	C
B08045	G	C	A
B09092	G	C	A
T106	G	C	A
B00090	G	C	A
KAS	G	C	A
B05023	G	C	A
B04009	G	C	A
B08040	G	C	A
B00092	G	C	A

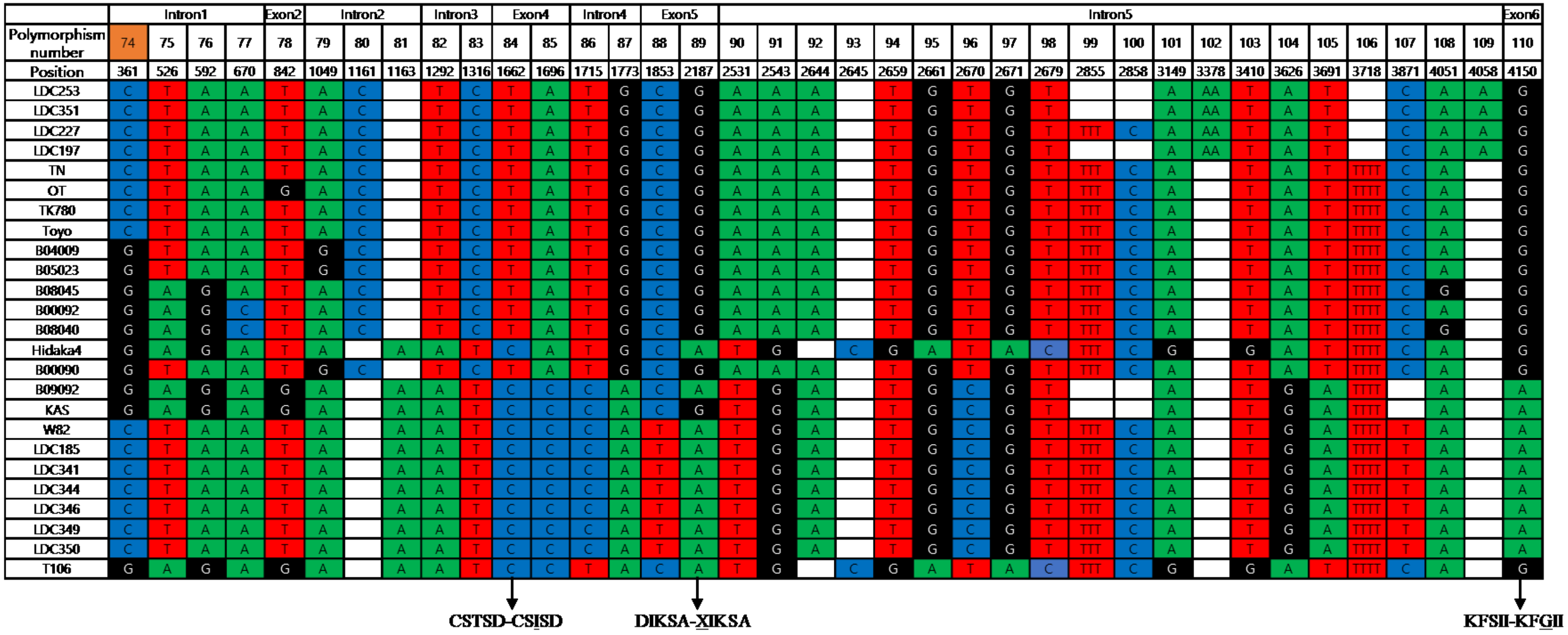


Polymorphism	No	cis-elements	Sequence	Strand
36	1	CGACGOSAMY3	CGACG	+

Polymorphism	No	cis-elements	Sequence	Strand
72	1	ROOTMOTIFTAPOX1	ATATT	+

Polymorphism	No	cis-elements	Sequence	Strand
73	1	MYB1AT	WAACCA	+
	2	REALPHALGLHCB21	AACCAA	+
	3	CCAATBOX1	CCAAT	+
	4	CAATBOX1	CAAT	+
	5	CACTFTPPCA1	YACT	-
	6	ANAERO1CONSENSUS	AAACAAA	+

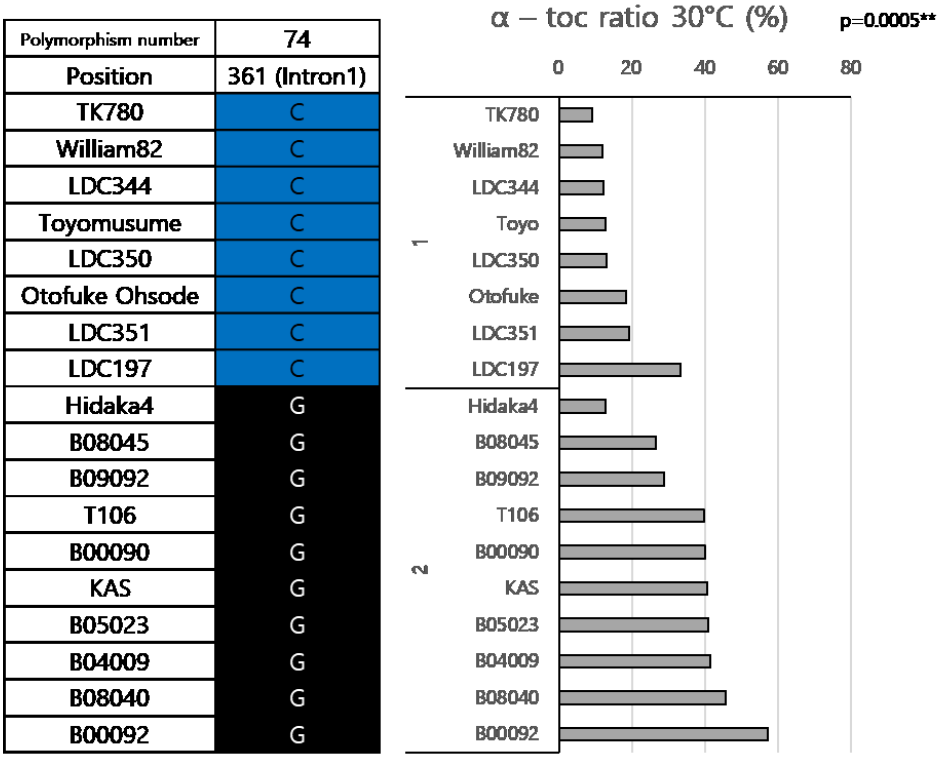
Fig. 25 Significant association between α-tocopherol ratios and a SNP in promoter of *γ-TMT3* gene. The accessions tested were classified based on the nucleotides at SNPs. The association between the SNP and the α-tocopherol ratios were analyzed by one-way analysis of variance.



	Intron1				Exon2	Intron2			Intron3		Exon4		Intron4		Exon5		Intron5																				Exon6
Polymorphism number	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110
Position	361	526	592	670	842	1049	1161	1163	1292	1316	1662	1696	1715	1773	1853	2187	2531	2543	2644	2645	2659	2661	2670	2671	2679	2855	2858	3149	3378	3410	3626	3691	3718	3871	4051	4058	4150
LDC253	C	T	A	A	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T			A	AA	T	A	T		C	A	A	G
LDC351	C	T	A	A	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T			A	AA	T	A	T		C	A	A	G
LDC227	C	T	A	A	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A	AA	T	A	T		C	A	A	G
LDC197	C	T	A	A	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T			A	AA	T	A	T		C	A	A	G
TN	C	T	A	A	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	A		G
OT	C	T	A	A	G	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	A		G
TK780	C	T	A	A	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	A		G
Toyo	C	T	A	A	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	A		G
B04009	G	T	A	A	T	G	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	A		G
B05023	G	T	A	A	T	G	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	A		G
B08045	G	A	G	A	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	G		G
B00092	G	A	G	C	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	A		G
B08040	G	A	G	C	T	A	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	G		G
Hidaka4	G	A	G	A	T	A		A	A	T	C	A	T	G	C	A	T	G		C	G	A	T	A	C	TTT	C	G		G	A	T	TTTT	C	A		G
B00090	G	T	A	A	T	G	C		T	C	T	A	T	G	C	G	A	A	A		T	G	T	G	T	TTT	C	A		T	A	T	TTTT	C	A		G
B09092	G	A	G	A	G	A		A	A	T	C	C	C	A	C	A	T	G	A		T	G	C	G	T			A		T	G	A	TTTT		A		A
KAS	G	A	G	A	G	A		A	A	T	C	C	C	A	C	G	T	G	A		T	G	C	G	T			A		T	G	A	TTTT		A		A
W82	C	T	A	A	T	A		A	A	T	C	C	C	A	T	A	T	G	A		T	G	C	G	T	TTT	C	A		T	G	A	TTTT	T	A		A
LDC185	C	T	A	A	T	A		A	A	T	C	C	C	A	T	A	T	G	A		T	G	C	G	T	TTT	C	A		T	G	A	TTTT	T	A		A
LDC341	C	T	A	A	T	A		A	A	T	C	C	C	A	T	A	T	G	A		T	G	C	G	T	TTT	C	A		T	G	A	TTTT	T	A		A
LDC344	C	T	A	A	T	A		A	A	T	C	C	C	A	T	A	T	G	A		T	G	C	G	T	TTT	C	A		T	G	A	TTTT	T	A		A
LDC346	C	T	A	A	T	A		A	A	T	C	C	C	A	T	A	T	G	A		T	G	C	G	T	TTT	C	A		T	G	A	TTTT	T	A		A
LDC349	C	T	A	A	T	A		A	A	T	C	C	C	A	T	A	T	G	A		T	G	C	G	T	TTT	C	A		T	G	A	TTTT	T	A		A
LDC350	C	T	A	A	T	A		A	A	T	C	C	C	A	T	A	T	G	A		T	G	C	G	T	TTT	C	A		T	G	A	TTTT	T	A		A
T106	G	A	G	A	G	A		A	A	T	C	C	T	A	C	A	T	G		C	G	A	T	A	C	TTT	C	G		G	A	T	TTTT	C	A		G

*** W82, Williams 82, Toyo, Toyomusume, OT, Otofuke Ohsode, TN, Tokachi Nagaha**

Fig. 26 DNA polymorphisms in the genic region of 5′ UTR to 3′ UTR in *γ-TMT3* among 25 wild soybean accessions and soybean cultivars. The positions of nucleotide polymorphisms relative to the adenine of translational start site (ATG) are presented at the second row.



Polymorphism number	74
Position	361 (Intron1)
TK780	C
William82	C
LDC344	C
Toyomusume	C
LDC350	C
Otofuke Ohsode	C
LDC351	C
LDC197	C
Hidaka4	G
B08045	G
B09092	G
T106	G
B00090	G
KAS	G
B05023	G
B04009	G
B08040	G
B00092	G

Fig. 27 Significant association between α-tocopherol ratios and a SNP in the intron of *γ-TMT3* gene. The accessions tested were classified based on the nucleotides at SNPs. The association between the SNP and the α-tocopherol ratios were analyzed by one-way analysis of variance.

Table 6. Primer sequences for *γ-TMT* gene expression analysis.

Primer Name	Primer Sequence
GTMT-1_F1	5'-CTGGAGGCAGAGTATAGCG-3'
GTMT-1_R1	5'-AAACTCCCAGGTCCCACCCAAT-3'
GTMT-2_F1	5'-GAAGCAAGTTTCCAACAGGTCG-3'
GTMT-2_R1	5'-CGCCAATCATAGGAGATATTGCATATG-3'
GTMT-3_F1	5'-CAGTGGACTTAAAACCATAAAGGGAGC-3'
GTMT-3_R1	5'-CCACATACTCTATATCATTCACACGAG-3'
qGmActin-F	5'-CGGTGGTTCCTATCTTGGCATC-3'
qGmActin-R	5'-GTCTTTCGCTTCAATAACCCTA-3'

Table 7. Primer sequences for methylation analysis.

Primer Name	Primer Sequence
G2-PRO-2245F	5'-CTCATTCTCGGCCACGAGATAACGAG-3'
G2-PRO-987R	5'-CTGCAACACACGGTCATAGGATTGG-3'
G2-PRO-1203F	5'-GGGACTAGAAATGGAAGGAGCCAAC-3'
G2-PRO--93R	5'-CGACCCGGATTCTGGCGAAAGTGCG-3'
G3-PRO-6F	5'-CTAAGAGCCATCTTGTCTCCACTC-3'
G3-PRO-6R	5'-GTAGTCCATTTCATGTGGTGCTGG-3'
G3-PRO-689F	5'-GGGATTGAATTTGGTCTATAATCG-3'
G3-PRO-689R	5'-CGATTATAGACCAAATTCAATCCC-3'
G3-PRO-224F	5'-CCGTTAGAAACCGATCTCACAAAC-3'
G3-PRO-224R	5'-GTTTGTGAGATCGGTTTCTAACGG-3'

3.4 Discussion

3.4.1 Variation in thermal response of Toc biosynthesis and *γ-TMT* expression

The α-Toc ratios in seeds in the cultivars and wild accessions tested increased as the temperature rose from 20 °C to 30 °C. In both thermal conditions, α-Toc ratios were highly correlated with α/γ ratios (r = 0.99 in 20 °C and r = 0.98 in 30 °C). The α-Toc ratios were also positively correlated with (α + γ)/total ratios (r = 0.54 in 20 °C and r = 0.55 in 30 °C), but the correlation coefficients were lower than those between α-Toc ratios and α/γ ratios. Furthermore, the ratio of 20 °C : 30 °C was highly correlated between α-Toc ratios and α/γ ratios (r = 0.98). From these results, the variations in α-Toc ratios and their thermal responses among the soybean cultivars and wild accessions were mainly associated with the variations in α/γ ratio, but not in those in (α + γ)/total ratio. The different accumulations of α-Toc may therefore be mainly attributed to the variation in the conversion efficiency from γ-Toc to α-Toc mediated by γ-TMTs.

The expression profiles of three *γ-TMT* genes in immature cotyledons were analyzed in three accessions, TK780, KAS and B04009. The expression levels varied with both temperatures and genotypes. Particularly, the higher temperature upregulated the expression of *γ-TMT*2 in all of the three lines tested, but the extents were marked in KAS and B04009, both of which accumulated α-Toc at higher ratios (Fig. 14). Therefore, *γ-TMT*2 may have an important role in the accumulation of α-Toc as promoted by high temperature. In addition, *γ-TMT*3 showed higher expression levels in KAS than TK780 and B04009 at both temperatures; the expression levels were also slightly but significantly higher in B04009 than TK780. As indicated in the previous chapter, *γ-TMT*3 is likely the candidate gene for the *qαγR-9*/*qαTC-9*. Although the expression levels of *γ-TMT*3 in B04009 were considerably low compared with those of KAS, the different expression levels might be a molecular basis for the QTL, as in the Ichihime × KAS cross. .

3.4.2 Expression profiles of candidate genes for the *qαγR-11*/*qαTC-11*

Van Eenennaam et al. (2003) demonstrated that the *γ-TMT* genes are main factor of high α-Toc contents and ratio in soybean. However, the present study indicated that the QTL located in Chr11 also had an effect on α-Toc content and ratio, despite that there was no gene involved in known Toc biosynthesis. I picked up three candidate genes for the QTL, based on the annotations in Williams 82 genome sequence data base, and analyzed their expressions in the immature cotyledons. The expression levels of genes for RING-H2 finger C2A (Glyma.11G220400), GATA type zinc finger transcription factor family protein (Glyma.11G226400) and S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (Glyma.11G222800) in immature cotyledons varied with both temperatures and genotypes. In Glyma.11G220400 and Glyma.11G226400, there is no consistent association between expression levels and α-Toc ratios. The expression levels of Glyma.11G222800 did not show any distinct responses to temperatures, but the expression levels in TK780 and KAS were significantly higher than those of B04009. A further study is needed to determine whether these genes were involved in the α-Toc biosynthesis or not.

3.4.3 Methylation analysis

The methylation on the promoter or coding region is known to inhibit gene expression in plant (Kilby et al., 1992; Hohn et al., 1996). As a causal factor of different expression levels of *γ-TMT* genes, I analyzed the methylation of the promoter regions of *γ-TMT2* and *γ-TMT3*. As a result, I found that the *γ-TMT3* promoter region of TK780 was methylated under 20 °C condition, but not in 30 °C. The comparison of DNA sequences for the methylated region further detected two SNPs located in the McrBC-targeted sites between TK780 and KAS, in which TK780 possessed the cytosine residues. Because the region was methylated only in TK780, these two cytosine residues might be subject to the methylation (Fig. 16). Furthermore, the methylation was dependent on the temperatures; only low temperature of 20 °C produced the methylation. It would be intriguing to determine how these cytosine residues were methylated in a further study. However, in TK780, the expression levels of *γ-TMT3* increased at 20 °C relative to 30 °C (Fig. 14), in discord with the prediction that the expression level at 20 °C was repressed by methylation. Therefore, the methylation of the *γ-TMT3* promoter may be not a main factor of lowered α-Toc biosynthesis in TK780.

3.4.4 Association of α-Toc ratios with SNPs in *γ-TMT* genes

In this study, I construct the *γ-TMT1*, *γ-TMT2* and *γ-TMT3* sequences based on whole genome resequencing data for 15 soybean cultivars and 9 wild soybean accessions. These materials were classified based on the nucleotides at SNPs. The association analysis revealed that α-Toc ratios were significantly associated with the SNPs in the intron of *γ-TMT1* gene

(SNP-28), *γ-TMT2* promoter (SNP-6 and 25), *γ-TMT3* promoter (Group of SNP-3, 26, 57 and Group of SNP-36, 72, 73) and gene regions (SNP-74). Of these, the SNPs (SNP-36, SNP-72 and SNP-73) in the *γ-TMT3* promoter showed the highest association between the SNP and α-Toc ratios, particularly at 30 °C; All of the accessions having the nucleotides G, C and A at these SNPs produced higher α-Toc ratios than those having different nucleotides (A, T and C), except for only LDC197, a Chinese accession, which produced high α-Toc ratios at both 20 °C and 30 °C. The other associations were not necessarily so close; there were some overlaps between α-Toc ratios and the nucleotides at SNPs. A further study is needed to determine the roles of the SNPs in the control of α-Toc biosynthesis

Wild soybean accessions with high seed α-Toc ratios including B04009 exhibited diverse DNA polymorphisms in the promoter sequence (1,350 bp) of *γ-TMT3*; B04009 differed with respect to 14 SNPs and four indels from KAS (Dwiyanti et al., 2016). However, all of the wild accessions with high α-Toc ratios shared three SNPs with KAS, which were differentiated from the low α-Toc cultivars tested (Dwiyanti et al., 2016). These SNPs corresponded to three SNPs identified in this study, SNP-57, SNP-72 and SNP-73. The group of accessions with SNP-57 included a wild accession with low α-Toc ratios (Hidaka 4). Therefore, SNP-72 and SNP-73 may be more important in the regulation of *γ-TMT3* gene expression. A further investigation using diverse germplasm collections would be useful to confirm the validity of the associations of these SNPs with seed α-Toc contents/ratios, and the identification of critical *cis*-element(s) would aid an understanding of seed α-Toc content control by *γ-TMT3*.

Chapter 4 General Discussion

Tocopherols are lipophilic antioxidants. The soybean oil has a relatively high total tocopherol content compared with other oilseed crops, maize, *Brassica napus* and sunflower (Li et al., 2019). However, the most predominant form is γ-Toc. The α-Toc content is typically less than 10 % that of the total Toc content (Grusak and DellaPenna, 1999; Bramley et al., 2000). In order to increase the α-Toc content and open up opportunities for new food and industrial uses of soybean, the genetic and molecular bases underlying the natural variation have been extensively studied (Li et al., 2010; Liu et al., 2010; Dwiyanti et al., 2011; Shaw et al., 2016; Shaw et al., 2017). Up to now, a number of QTLs for tocopherol biosynthesis have been identified. Dwiyanti et al. (2011) detected a major QTL for α-Toc ratio in chromosome 9, which accounted for 55 % of the phenotypic variation in a RIL population derived from the Ichihime × KAS cross. The molecular basis of this QTL is suggested to be a different promoter activity of *γ-TMT3*.

In this study, I identified two major QTLs and one minor QTL conferring higher α-Toc contents in a cross between a low α-Toc soybean breeding line TK780 and a high α-Toc wild accession B04009. Of these, the *qαγR/αTC-9* was located in the QTL detected in the Ichihime × KAS cross, and harbored *γ-TMT3*. The results suggest that the elevated α-Toc ratios detected in both cultivated and wild soybeans are attributed to the same QTL. The other two QTLs were novel QTLs, which had not been reported so far. The genomic region flanking *qαγR/αTC-12* contained tightly-linked *γ-TMT1* and *γ-TMT*2, whereas there were no genes involved in tocopherol biosynthesis in the *qαγR/αTC-11* region. The alleles from B04009 increased α-Toc ratios at *qαγR/αTC-9* and *qαγR/αTC-12*, but decreased the ratios at *qαγR/αTC-11*. The presence of the QTLs with opposite effects resulted in the transgressive segregation in the RIL population. In the *qαγR/αTC-11* region, I found two zinc finger transcription factors (Glyma.11G220400 and Glyma.11G226400) and S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (Glyma.11G222800.1) as possible candidates for the QTL.

The α-Toc ratios in seeds matured in different thermal conditions were analyzed in the cultivars and wild accessions. The α-Toc ratios were increased in response to the rise of the temperatures from 20 °C to 30 °C. In both thermal conditions, the α-Toc ratios in cultivars/accessions were highly correlated with the α/γ ratios, indicating that the different accumulations of α-Toc is mainly attributed to the variation in the conversion efficiency from γ-Toc to α-Toc. Two parents, TK780 and B04009, and KAS were analyzed for the expression profiles of *γ-TMT* genes. Of the three genes, *γ-TMT*2 exhibited distinct upregulation of expressions by high temperatures in all of the accessions tested, but the extents were marked in KAS and B04009, which accumulated α-Toc at higher ratios in 30 °C condition. Therefore, *γ-TMT*2 may have an important role in the accumulation of α-Toc as promoted by high temperature. The expression levels of *γ-TMT3* were also high in KAS, compared with TK780 and B04009. However, the expression levels were slightly but significantly higher in B04009 than TK780 at both temperatures. Accordingly, the different expression levels of *γ-TMT3* and *γ-TMT2* may be the molecular bases underlying the two QTLs for α-Toc contents and ratios in Chr9 and Chr12 detected in the TK780 × B04009 cross.

The sequences of *γ-TMT* genes and three candidate genes for *qαγR/αTC-11* were compared between the parents of RIL population, and among cultivated and wild soybean accessions. There were a lot of SNPs and indels in the promoter and genic regions for the three *γ-TMT* genes. The associations of SNPs with α-Toc ratios were then analyzed. The α-Toc ratios were significantly associated with a number of SNPs in the promoters and introns of *γ-TMT* genes. Particularly, three SNPs in the promoter of *γ-TMT3* exhibited high association with the α-Toc ratios. These SNPs were located within known *cis*-elements, such as CGACGOSAMY3, ROOTMOTIFAPOX1 and MYB1AT/CAATBOX1/ANAERO1CONSENSIS. A further study is needed to confirm the validity of the associations of these SNPs with the α-Toc ratios and to determine the roles of *cis*-elements in the regulation of expression of *γ-TMT3*.

Besides the contribution to human health as vitamin E activity, tocopherols function in the prevention of non-enzymatic lipid oxidation during seed storage, germination, and early seedling development (Sattler et al., 2004). Different isoforms have their own functions against oxidative stresses. A lack of all isoforms of tocopherols abolished seed longevity, germination, and seedling growths (Sattler et al., 2004). The reduction of α-Toc contents in leaves showed an elevated susceptibility toward salt stress, but a diminished susceptibility toward osmotic and oxidative stresses from methyl viologen-induced reactive oxygen species, suggesting that γ-Toc is more potent than α-Toc in conferring desiccation tolerance (Abbasi et al., 2007). Tavva et al. (2007) found that the increment of α-Toc in seeds of transgenic plants was associated with the reduction of lipid peroxidation products in seeds and germinating seeds, although no specific differences were observed in seed germination or seedling growth between wild-type and transgenic plants. The β-Toc content negatively correlates with seed longevity in rice cultivars (Lee et al., 2017).

In this study, the roles of tocopherol isoforms on stress tolerances in plant development were not studied. Three *γ-TMT* genes in soybean have high amino acid similarities to each other, except for the N-terminal region in which only *γ-TMT2* possesses a plastid transit peptide signal (Dwiyanti et al., 2016). The expression analyses revealed that the expression of *γ-TMT2* was upregulated in response to the rise of temperatures during maturation. This upregulation appeared specific to *γ-TMT2*; there were no distinct responses of expressions in the other two genes. The different protein structures, different expression profiles, and different thermal responses of expressions may therefore indicate that the three *γ-TMT* genes have been subfunctionalized to each other, and possess different functions in the adaptation to oxidative stresses that occur during various phases of development. A further study should be directed to their roles in the adaptation to the oxidative stresses.

Summary

The soybean oil has a relatively high total tocopherol (Toc) content compared with other oilseed crops. However, the content of α-Toc which has the highest vitamin E activity in human body is typically less than 10 % of the total Toc content in soybean. Increasing the seed α-Toc content may therefore open up opportunities for new food and industrial uses of soybean. The objectives of this study were to determine the molecular genetic base for an elevated α-Toc ratio detected in a wild soybean accession, and to characterize the thermal responses of tocopherol biosynthesis and determine the association with DNA polymorphisms in the three *γ-TMT* genes. The results obtained from this study were summarized as follows.

Identification of quantitative trait loci for increased α-tocopherol biosynthesis in wild soybean using a high-density genetic map

A recombinant inbred line population was developed from a cross between the low α-Toc breeding line TK780 and the high α-Toc wild accession B04009. The tocopherol contents and compositions were evaluated in the greenhouse of ambient temperatures of 25 °C to 30 °C in short day condition. The α-Toc content in seeds correlated closely with the ratio of α-Toc to γ-Toc contents (α/γ ratio). QTL analysis using a high-density map constructed with 7,710 single nucleotide polymorphisms (SNPs) generated by restriction site-associated DNA sequencing detected six QTLs involved in Toc biosynthesis. Of these, three in chromosomes (Chr) 9, 11, and 12 produced consistent large effects on α-Toc contents and ratios during a 2-year trial. The B04009 allele at QTLs in Chr9 and Chr12 and the TK780 allele at the QTL in Chr11 each promoted the conversion of γ-Toc to α-Toc, which elevated the seed α-Toc contents and ratios. The opposite effects by the three QTLs thus resulted in the transgressive segregation in the seed α-Toc contents and α/γ ratios in the RIL population. A survey on the Williams 82 genome database revealed that three genes (*γ-TMT1*, *γ-TMT2* and *γ-TMT3*) encoding γ-tocopherol methyltransferase (TMT), which convert γ-Toc to α-Toc, were co-located in the QTLs in Chr9 (*γ-TMT3*) and Chr12 (*γ-TMT1* and *γ-TMT2*), whereas the QTL in Chr11 had no known genes involved in Toc biosynthesis; there were two zinc finger transcription factors (Glyma.11G220400 and Glyma.11G226400) and S-adenosyl-L-methionine-dependent methyltransferases superfamily protein (Glyma.11G222800).

Molecular mechanism on thermal response of α-tocopherol biosynthesis

The α-Toc ratios in seeds were analyzed in 15 cultivars and 10 wild accessions in 20 °C and 21 cultivars and 11 wild accessions in 30 °C. In both thermal conditions, α-Toc ratios and α/γ ratios were highly correlated to each other. The increment of α-Toc ratios was mainly associated with the conversion efficiencies from γ-Toc to α-Toc. Expression analyses of immature cotyledons developed in different thermal conditions were performed in TK780, KAS (a high α-Toc cultivar) and B04009. Of the three *γ-TMT* genes, *γ-TMT2* was upregulated by high temperature, and this upregulation was marked particularly in KAS and B04009. The expression levels of *γ-TMT3* were markedly high in KAS, compared with TK780 and B04009; the expression levels were further slightly but significantly higher in B04009 than TK780 at both temperatures. These findings suggest that different expression levels of *γ-TMT3* and *γ-TMT2* may be the molecular bases underlying the two QTLs for α-Toc contents and ratios in Chr9 and Chr12.

Of the three candidate genes for the QTL in Chr11, Glyma.11G222800 encoding S-adenosyl-L-methionine-dependent methyltransferase superfamily protein exhibited higher expressions in TK780 than B04009 in both 20 °C and 30 °C conditions. There were two non-synonymous substitutions between TK780 and B04009, in which TK780 had the amino acid residues conserved in the homoeolog and homologs in other plant species. Therefore, it was presumed that B04009 had a missense variant of the protein. A further study is needed to determine whether this methyltransferase protein has any function to convert γ-Toc to α-Toc.

I analyzed the methylation of the promoter regions of *γ-TMT2* and *γ-TMT3* as a causal factor for different expression profiles between tissues and between thermal conditions. As a result, I found that a part of the *γ-TMT3* promoter of TK780 was methylated in 20 °C condition, but not in 30 °C. However, the presence of methylation was not directly related to the expression levels of *γ-TMT3* in different temperatures. Therefore, the observed methylation in the *γ-TMT3* promoter was not a main factor, if any, to control the Toc biosynthesis in response to different thermal conditions in TK780.

Finally, I performed the association analysis of SNPs in the *γ-TMT1*, *γ-TMT2* and *γ-TMT3* sequences detected based on whole genome resequencing data with α-Toc ratios for 16 soybean cultivars and 9 wild soybean accessions. I detected a number of SNPs significantly associated with the variation in α-Toc ratios. The association analysis revealed that the α-Toc ratios were significantly associated with one SNP in the intron of *γ-TMT1*, two in the promoter of *γ-TMT2*, and seven in the promoter and genic regions of *γ-TMT3*. Particularly, three SNPs in the promoter region of *γ-TMT3* showed high associations with α-Toc ratios, suggesting an involvement of SNPs in *cis*-elements of *γ-TMT* genes in the different controls of α-Toc biosynthesis among soybean cultivars and wild accessions.

Based on the results from the present study, I discussed 1) the molecular genetic basis of the elevated α-Toc ratio of wild soybean accession, and 2) molecular genetic basis on the variation in thermal responses of α-Toc biosynthesis in cultivated and wild soybeans. The findings obtained in this study would be helpful in understanding of the molecular mechanisms underlying α-Toc biosynthesis in soybean seeds and developing the cultivars with high α-Toc contents in seeds.

Abbreviations

RAD-seq: Restriction-site associated DNA sequencing, SNP: single nucleotide polymorphisms, RIL: recombinant inbred line, QTL: quantitative trait locus, PCR: polymerase chain reaction. qRT-PCR: quantitative real-time PCR, Toc: tocopherol, MPBQ: 2-methyl-6-phytyl-1,4-benzoquinol, MT: methyltransferase, DMPBQ: 2,3-dimethyl-6-phytyl-1,4-benzoquinone, γ-TMT: γ-tocopherol methyltransferase.

References

1. Abbasi AR, Hajirezaei M, Hofius D, Sonnewald U, Voll LM. Specific roles of α- and γ-tocopherol in abiotic stress responses of transgenic tobacco. Plant Physiology, 2007; 143.4: 1720-1738.

2. Allen RD, Bernier F, Lessard PA, Beachy RN. Nuclear factors interact with a soybean beta-conglycinin enhancer. The Plant Cell, 1989; 1.6: 623-631.

3. Almeida J, Quadrana L, Asís R, Setta N, Godoy F, Bermúdez L, Otaiza SN, Corrêa da Silva JV, Fernie AR, Carrari F, Rossi M. Genetic dissection of vitamin E biosynthesis in tomato. Journal of Experimental Botany, 2011; 62.11: 3781-3798.

4. Almonora GO, Fennera GP, Wilson RF. Temperature effects on tocopherol composition in soybeans with genetically improved oil quality. Journal of the American Oil Chemists' Society, 1998; 75.5: 591-596.

5. Arun M, Subramanyam K, Theboral J, Sivanandhan G, Rajesh M, Dev GK, Jaganath B, Manickavasagam M, Girija S, Ganapathi A. Transfer and targeted overexpression of γ-tocopherol methyltransferase (γ-TMT) gene using seed-specific promoter improves tocopherol composition in Indian soybean cultivars. Applied Biochemistry and Biotechnology, 2014; 172.4: 1763-1776.

6. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLOS One, 2008; 3.10: e3376.

7.Bandillo N, Jarquin D, Song Q, Nelson R, Cregan P, Specht J, Lorenz A. A population structure and genome-wide association analysis on the USDA soybean germplasm collection. The Plant Genome, 2015; 8.3.

8. Bergmüller E, Porfirova S, Dörmann P. Characterization of an *Arabidopsis* mutant deficient in gamma-tocopherol methyltransferase. Plant Molecular Biology, 2003; 52.6:1181-1190.

9. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 2014; 30.15: 2114-2120.

10. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 2007; 23: 2633-2635.

11. Bramley PM, Elmadfa I, Kafatos A, Kelly FJ, Manios Y, Roxborough HE, Schuch W, Sheehy PJA, Wagner KH. Vitamin E – review. Journal of the Science of Food and Agriculture, 2000; 80: 913-938.

12. Britz SJ, Kremer DF. Warm temperatures or drought during seed maturation increase free α-tocopherol in seeds of soybean (*Glycine max* [L.] Merr.). Journal of Agricultural and Food Chemistry, 2002; 50.21: 6058-6063.

13. Britz SJ, Kremer DF, Kenworthy WJ. Tocopherols in soybean seeds: genetic variation and environmental effects in field-grown crops. Journal of the American Oil Chemists' Society, 2008; 85.10: 931-936.

14. Broman KW, Wu H, Sen Ś, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics, 2003; 19.7: 889-890.

15. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. American Journal of Human Genetics, 2007; 81: 1084-1097.

16. Busk PK, Jensen AB, Pagès M. Regulatory elements in vivo in the promoter of the abscisic acid responsive gene rab17 from maize. The Plant Journal, 1997; 11.6: 1285-1295.

17. Carra˜o-Panizzi MC, Erhan SZ. Environmental and genetic variation of soybean tocopherol content under Brazilian growing conditions. Journal of the American Oil Chemists' Society, 2007; 84.10: 921-928.

18. Carrera C, Martínezb MJ, Dardanellib J, Balzarinic M. Environmental variation and correlation of seed components in nontransgenic soybeans: protein, oil, unsaturated fatty acids, tocopherols, and isoflavones. Crop Science, 2011; 51.2: 800-809.

19. Carrera CS, Seguin P. Factors affecting tocopherol concentrations in soybean seeds. Journal of Agricultural and Food Chemistry, 2016; 64.50: 9465-9474.

20. Cercós M, Gómez-Cadenas A, Tuan-Hua David H. Hormonal regulation of a cysteine proteinase gene, EPB-1, in barley aleurone layers: cis- and trans-acting elements involved in the co-ordinated gene expression regulated by gibberellins and abscisic acid. The Plant Journal, 1999; 19.2: 107-118.

21. Chen DF, Zhang M, Wang YQ, Chen XW. Expression of γ-tocopherol methyltransferase gene from *Brassica napus* increased α-tocopherol content in soybean seed. Biologia Plantarum, 2012; 56.1: 131-134.

22. Chennupati P, Seguin P, Liu W. Effects of high temperature stress at different development stages on soybean isoflavone and tocopherol concentrations. Journal of Agricultural and Food chemistry, 2011; 59.24: 13081-13088.

23. Concibido V, Vallee BL, Mclaird P, Pineda N, Meyer J, Hummel L, Yang J, Wu K, Delannay X. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theoretical and Applied Genetics, 2003; 106.4: 575-582.

24. Dean Rider S. Jr, Henderson JT, Jerome RE, Edenberg HJ, Romero-Severson J, Ogas J. Coordinate repression of regulators of embryonic identity by PICKLE during germination in *Arabidopsis*. The Plant Journal, 2003; 35.1: 33-43.

25. Dellapenna D. A decade of progress in understanding vitamin E synthesis in plants. Journal of Plant Physiology, 2005; 162.7: 729-737.

26. Diepenbrock CH, Kandianis CB, Lipka AE, Magallanes-Lundback M, Vaillancourt B, Góngora-Castillo E, Wallace JG, Cepela J, Mesberg A, Bradbury PJ, Ilut DC, Mateos-Hernandez M, Hamilton J, Owens BF, Tiede T, Buckler ES, Rocheford T, Buell CR, Gore MA, DellaPenna D. Novel loci underlie natural variation in vitamin E levels in maize grain. The Plant Cell, 2017; 29.10: 2374-2392.

27. Doyle J, Doyle JL. Genomic plant DNA preparation from fresh tissue-CTAB method. Phytochemical Bulletin, 1987; 19.11: 11-15.

28. Dwiyanti MS, Ujiie A, Thuy LTB, Yamada T, Kitamura K. Genetic analysis of high alpha-tocopherol content in soybean (*Glycine max*) seeds. Breeding Science, 2007; 57: 23-28.

29. Dwiyanti MS, Yamada T, Sato M, Abe J, Kitamura K. Genetic variation of γ-tocopherol methyltransferase gene contributes to elevated α-tocopherol content in soybean seeds. BMC Plant Biology, 2011; 11.1: 152.

30. Dwiyanti MS, Maruyama S, Hirono M, Sato M, Park E, Yoon SH, Yamada T, Abe J. Natural diversity of seed α-tocopherol ratio in wild soybean (*Glycine soja*) germplasm collection. Breeding Science, 2016; 66: 653-657.

31. Ellerström M, Stålberg K, Ezcurra I, Rask L. Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm-specific transcription. Plant Molecular Biology, 1996; 32.6: 1019-1027.

32. Emanuelsson O, Nielsen H, von Heijne G. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science, 1999; 8.5: 978-984.

33. Fehr WR, Caviness CE, Burmood DT, Pennington JS. Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Science, 1971; 11: 929-931.

34. Fritsche S, Wang X, Li J, Stich B, Kopisch-Obuch FJ, Endrigkeit J, Leckband G, Dreyer F, Friedt W, Meng J, Jung C. A candidate gene-based association study of tocopherol content and composition in rapeseed (*Brassica napus*). Frontiers in Plant Science, 2012; 3: 129.

35. Fritsche S, Wang X, Jung C. Recent advances in our understanding of tocopherol biosynthesis in plants: An overview of key genes, functions, and breeding of vitamin E improved crops. Antioxidants, 2017; 6.4: 99.

36. Fukuda T, Maruyama N, Kanazawa A, Abe J, Shimamoto Y, Hiemori M, Tsuji H, Tanisaka T, Utsumi S. Molecular analysis and physicochemical properties of electrophoretic variants of wild soybean *Glycine soja* storage proteins. Journal of Agricultural and Food Chemistry, 2005; 53: 3658-3665

37. Gilliland LU, Magallanes-Lundback M, Hemming C, Supplee A, Koornneef M, Bentsink L, DellaPenna D. Genetic basis for natural variation in seed vitamin E levels in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America, 2006; 103.49: 18834-18841.

38. Grusak MA, DellaPenna D. Improving the nutrient composition of plants to enhance human nutrition and health. Annual Review of Plant Biology, 1999; 50.1: 133-161.

39. Graebner RC, Wise M, Cuesta-Marcos A, Geniza M, Blake T, Blake VC, Butler J, Chao S, Hole DJ, Horsley R, Jaiswal P, Obert D, Smith KP, Ullrich S, Hayes PM. Quantitative trait loci associated with the tocochromanol (vitamin E) pathway in barley. PlOS One, 2015; 10.7: e0133767.

40. Hajika M, Takahashi M, Sakai S, Igita M. A new genotype of 7S globulin (β-conglycinin) detected in wild soybean (*Glycine soja* Sieb. et Zucc.). Breeding Science, 1996; 46: 385-386.

41. Higo K, Ugawa Y, Iwamoto M, Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Research, 1999; 27.1: 297-300.

42. Hohn T, Corsten S, Rieke S, Müller M, Rothnie H. Methylation of coding region alone inhibits gene expression in plant protoplasts. Proceedings of the National Academy of Sciences of the United States of America, 1996; 93.16: 8334-8339.

43. Kanamaru K, Wang S. Yamada T. Abe J. Kitamura K. Genetic analysis and biochemical characterization of the high lutein trait of wild soybean (*Glycine soja* Sieb. et Zucc.). Breeding Science, 2008; 58: 393-400.

44. Kikuchi A, Tsukamoto C, Tabuchi K, Adachi T, Okubo K. Inheritance and characterization of a null allele for group A acetylsaponins found in a mutant soybean. Breeding Science, 1999; 49: 167-171.

45. Kilby Nigel J, Leyser HM Ottoline, Furner Ian J. Promoter methylation and progressive transgene inactivation in *Arabidopsis*. Plant Molecular Biology, 1992; 20.1: 103-112.

46. Kim MY, Lee SH, Van K, Kim TH, Jeong SC, Choi IY, Kim DS, Lee YS, Park D, Ma J, Kim WY, Kim BC, Park S, Lee KA, Kim DH, Kim KH, Shin JH, Jang YE, Kim KD, Liu WX, Chaisan T, Kang YJ, Lee YH, Kim KH, Moon JK, Schmutz J, Jackson SA, Bhak J, Lee SH. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proceedings of the National Academy of Sciences of the United States of America, 2010; 107.51: 22032-22037.

47. Kosambi DD. Statistics in function space, Journal of the Indian Mathematical Society, 1943; 76-88.

48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods, 2012; 9.4: 357.

49. Lee JS, Kwak J, Yoon MR, Lee JS, Hay FR. Contrasting tocol ratios associated with seed longevity in rice variety groups. Seed Science Research, 2017; 27.4: 273-280.

50. Li H, Liu H, Han Y, Wu X, Teng W, Liu G, Li W. Identification of QTL underlying vitamin E contents in soybean seed among multiple environments. Theoretical and Applied Genetics, 2010; 120.7: 1405-1413.

51. Liu H, Cao G, Wu D, Jiang Z, Han Y, Li W. Quantitative trait loci underlying soybean seed tocopherol content with main additive, epistatic and QTL × environment effects. Plant Breeding, 2017; 136.6: 924-938.

52. Li Q, Tang X, Lu S, Wu J. Composition and tocopherol, fatty acid, and phytosterol contents in micro-endosperm ultra-high oil corn. Grasas y Aceites, 2019; 70.3: 311.

53. Marwede V, Gul MK, Becker HC, Ecke W. Mapping of QTL controlling tocopherol content in winter oilseed rape. Plant Breeding, 2005; 124.1: 20-26.

54. Masuda, T, Goldsmith PD. World soybean production: area harvested, yield, and long-term projections. International Food and Agribusiness Management Review, 2009; 12.1030-2016-82753: 1-20.

55. Mène-Saffrané L. Pellaud S. Current strategies for vitamin E biofortification of crops. Current Opinion in Biotechnology, 2017; 44: 189-197.

56. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 2010; 20.9: 1297-1303.

57. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLOS One, 2012; 7.5: e37135.

58. Quadrana L, Almeida J, Asís R, Duffy T, Dominguez PG, Bermúdez L, Conti G, Corrêa da Silva JV, Peralta IE, Colot V, Asurmendi S, Fernie AR, Rossi M, Carrari F. Natural

occurring epialleles determine vitamin E accumulation in tomato fruits. Nature Communications, 2014 5: 4027.

59. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2018; URL https://www.R-project.org/.

60. Sattler SE, Gilliland LU, Magallanes-Lundback M, Pollard M, DellaPenna D. Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. The Plant Cell, 2004; 16.6: 1419-1432.

61. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. Genome sequence of the palaeopolyploid soybean. Nature, 2010; 463.7278: 178.

62. Seguin P, Turcotte P, Tremblay G, Pageau D, Liua W. Tocopherols concentration and stability in early maturing soybean genotypes. Agronomy Journal, 2009; 101.5: 1153-1159.

63. Seguin P, Tremblay G, Pageau D, Liu. W. Soybean tocopherol concentrations are affected by crop management. Journal of Agricultural and Food Chemistry, 2010; 58.9: 5495-5501.

64. Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA, Cannon SV, May GD, Vance CP, Shoemaker RC. RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biology, 2010; 10.1: 160.

65. Shaw EJ, Kakuda Y, Rajcan I. Effect of genotype, environment, and genotype× environment interaction on tocopherol accumulation in soybean seed. Crop Science, 2016; 56.1: 40-50.

66. Shaw EJ, Rajcan I. Molecular mapping of soybean seed tocopherols in the cross ‘OAC Bayfield’×‘OAC Shire’. Plant Breeding, 2017; 136.1: 83-93.

67. Stålberg K, Ellerstöm M, Ezcurra I, Ablov S, Rask L. Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds. Planta, 1996; 199.4: 515-519.

68. Tavva VS, Kim YH, Kagan IA, Dinkins RD, Kim KH, Collins GB. Increased α-tocopherol content in soybean seed overexpressing the Perilla frutescens γ-tocopherol methyltransferase gene. Plant Cell Reports, 2007; 26.1: 61-70.

69. Tewari K, Dahuja A, Sachdev A, Kumar V, Ali K, Kumar A, Kumari S. Molecular cloning, heterologous expression and functional characterization of gamma tocopherol methyl transferase (γ-TMT) from *Glycine max*. Protein Expression and Purification, 2017; 140: 81-89.

70. Thum KE, Kim M, Morishige DT, Eibl C, Koop HU, Mullet JE. Analysis of barley chloroplast psbD light-responsive promoter elements in transplastomic tobacco. Plant Molecular Biology, 2001; 47.3: 353-366.

71. Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, Westhoff P. cis-Regulatory elements for mesophyll-specific gene expression in the C4 plant *Flaveria trinervia*, the promoter of the C4 phosphoenolpyruvate carboxylase gene. The Plant Cell, 2004; 16.5: 1077-1090.

72. Ujiie A, Yamada T, Fujimoto K, Endo Y, Kitamura K. Identification of soybean varieties with high α-tocopherol content. Breeding Science, 2005; 55.2: 123-125.

73. Van Eenennaam AL, Lincoln K, Durrett TP, Valentin HE, Shewmaker CK, Thorne GM, Jiang J, Baszis SR, Levering CK, Aasen ED, Hao M, Stein JC, Norris SR, Last RL. Engineering vitamin E content: from *Arabidopsis* mutant to soy oil. The Plant Cell, 2003; 15.12: 3007-3019.

74. Van Eenennaama AL, Li G, Venkatramesh M, Leverin C, Gong X, Jamieson AC, Rebar EJ, Shewmaker CK, Case CC. Elevation of seed α-tocopherol levels using plant-based transcription factors targeted to an endogenous locus. Metabolic Engineering, 2004; 6.2: 101-108.

75. Voorrips RE, MapChart: Software for the graphical presentation of linkage maps and QTLs. The Journal of Heredity, 2002; 93 (1): 77-78.

76. Wang H, Xu S, Fan Y, Liu N, Zhan W, Liu H, Xiao Y, Li K, Pan Q, Li W, Deng M, Liu J, Jin M, Yang X, Li J, Li Q, Yan J. Beyond pathways: genetic dissection of tocopherol content in maize kernels by combining linkage and association analyses. Plant Biotechnology Journal, 2018; 16.8: 1464-1475.

77. Wang J, Li H, Zhang L, Meng L. User's manual of QTL IciMapping ver. 4.1.'. Quantitative Genetics group, Institute of Crop Science, Chinese Academy of Agricultural Sciences (CAAS): Beijing/Genetic Resources Program, International Maize and Wheat Improvement Center (CIMMYT): Mexico City; 2016.

78. Wang S, Kanamaru K, Li W, Abe J, Yamada T, Kitamura K. Simultaneous accumulation of high contents of α-tocopherol and lutein is possible in seeds of soybean (*Glycine max* (L.) Merr.). Breeding Science, 2007; 57.4: 297-304.

79. Wang X, Zhang C, Li L, Fritsche S, Endrigkeit J, Zhang W, Long Y, Jung C, Meng J. Unraveling the genetic basis of seed tocopherol content and composition in rapeseed (*Brassica napus* L.). PlOS One, 2012; 7.11: e50038.

80. Wang XQ, Yoon MY, He Q, Kim TS, Tong W, Choi BW, Lee YS, Park YS. Natural variations in OsγTMT contribute to diversity of the α-tocopherol content in rice. Molecular Genetics and Genomics, 2015; 290.6: 2121-2135.

81. Xu D, Tuyen DD. Genetic studies on saline and sodic tolerances in soybean. Breeding Science, 2012; 61.5: 559-565.

82. Jiao Y, Ma L, Strickland E, Deng XW. Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and *Arabidopsis*. The Plant Cell, 2005; 17.12: 3239-3256.

83. Zhang L, Luo Y, Zhu Y, Zhang L, Zhang W, Chen R, Xu M, Fan Y, Wang L GmTMT2a from soybean elevates the α-tocopherol content in corn and *Arabidopsis*. Transgenic Research, 2013; 22.5: 1021-1028.

84. Zhang ZL, Xie Z, Zou X, Casaretto J, Ho TD, Shen QJ. A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. Plant Physiology, 2004; 134.4: 1500-1513.

NOTE:
Part of the results presented in this thesis has been published in
BMC PLANT BIOLOGY.

Cheolwoo Park, Maria Stefanie Dwiyanti, Atsushi J. Nagano, Baohui Liu, Tetsuya Yamada, Jun Abe. Identification of quantitative trait loci for increased α-tocopherol biosynthesis in wild soybean using a high-density genetic map. *BMC Plant Biology*, 2019, 19.1: 510.

Acknowledgements

北海道大学に戻って来た一昨日ようなのに、あっという間に3年が経って卒業をするようになりました。

まず，担当教授である、 阿部 純 先生に心から感謝を申し上げます。留学期間の全過程にわたる先生の献身的なサポートがなかったら、これまでの成果はなかったと思います。もう一度、本当にありがとうございます。

3 年間私を誠心誠意支援いただいた、マリア先生。本当にありがとうございます。マリア先生のおかげで、大豆研究家として一段階発展することができました。 私の博士学位と論文の成果に欠かせない重要な意見や提案をいただいた金澤先生、山田先生ありがとうございます。 留学生活のアドバイスや様々な実験を教えてくれた徐美蘭さんありがとうございます。

植物遺伝資源科学実験室の学生の皆さんに感謝します。皆さんのおかげで楽しく大学院生活を楽しむことができました。

このような 皆さんの助けがあったから、不足な私ですが、卒業後もずっとやりたかった研究を行うことができました。今に甘んじることなく発展する研究者になれるよう努力します。 後日、笑いながら会えることを祈ります。

最後に、留学中精神的、財政的に支援を惜しまなかった、家族に感謝します。