

Tyamagondlu V. Venkatesh · Balasulojini Karunanandaa  
Daniel L. Free · Jeannie M. Rottnek · Susan R. Baszis  
Henry E. Valentin

## Identification and characterization of an *Arabidopsis* homogentisate phytyltransferase paralog

Received: 14 July 2005 / Accepted: 7 November 2005 / Published online: 12 January 2006  
© Springer-Verlag 2006

**Abstract** Tocochromanols (tocopherols and tocotrienols) are micronutrients with antioxidant properties synthesized by photosynthetic bacteria and plants that play important roles in animal and human nutrition. There is considerable interest in identifying the genes involved in tocochromanol biosynthesis to allow transgenic modification of both tocochromanol levels and tocochromanol composition in agricultural crops. The first committed reaction in tocopherol biosynthesis is the condensation of homogentisic acid (HGA) with phytyldiphosphate or geranylgeranyldiphosphate, catalyzed by the homogentisate phytyltransferase (VTE2) or by the homogentisate geranylgeranyl transferase (HGGT). In this study, we describe the identification of conserved amino acid sequences within VTE2 and HGGT and the application of these conserved sequences for a motif analysis resulting in the discovery of a *VTE2*-paralog in the *Arabidopsis* genome. We designated this new gene *VTE2-2* and renamed the old *VTE2* to *VTE2-1*. Seed-specific expression of *VTE2-2* in *Arabidopsis* resulted in increased seed-tocopherol levels, similar to the transgenic expression of *VTE2-1*. Bioinformatics analysis revealed that *VTE2-2* is conserved in both monocotyle-

donous and dicotyledonous plants and is distinct from *VTE2-1* and HGGT.

**Keywords** *Arabidopsis* · Homogentisate phytyltransferase · Homogentisate prenyltransferase · Motif analysis · Tocopherol · Vitamin E

**Abbreviations** CTP: Chloroplast target peptide · GGDP: Geranylgeranyldiphosphate · HGA: Homogentisic acid · HGGT: Homogentisate geranylgeranyltransferase · HMM: Hidden Markov Model · MPSS: Massively Parallel Signature Sequencing · TNF: Tumor necrosis factor · PDP: Phytyldiphosphate · PUFA: Polyunsaturated fatty acids · VTE2: Homogentisate phytyltransferase · VTE3: 2-methyl-6-phytylbenzoquinol methyltransferase

**Electronic Supplementary Material** Supplementary material is available for this article at <http://www.dx.doi.org/10.1007/s00425-005-0180-1> and is accessible for authorized users.

Tyamagondlu V. Venkatesh, and Balasulojini Karunanandaa have equally contributed.

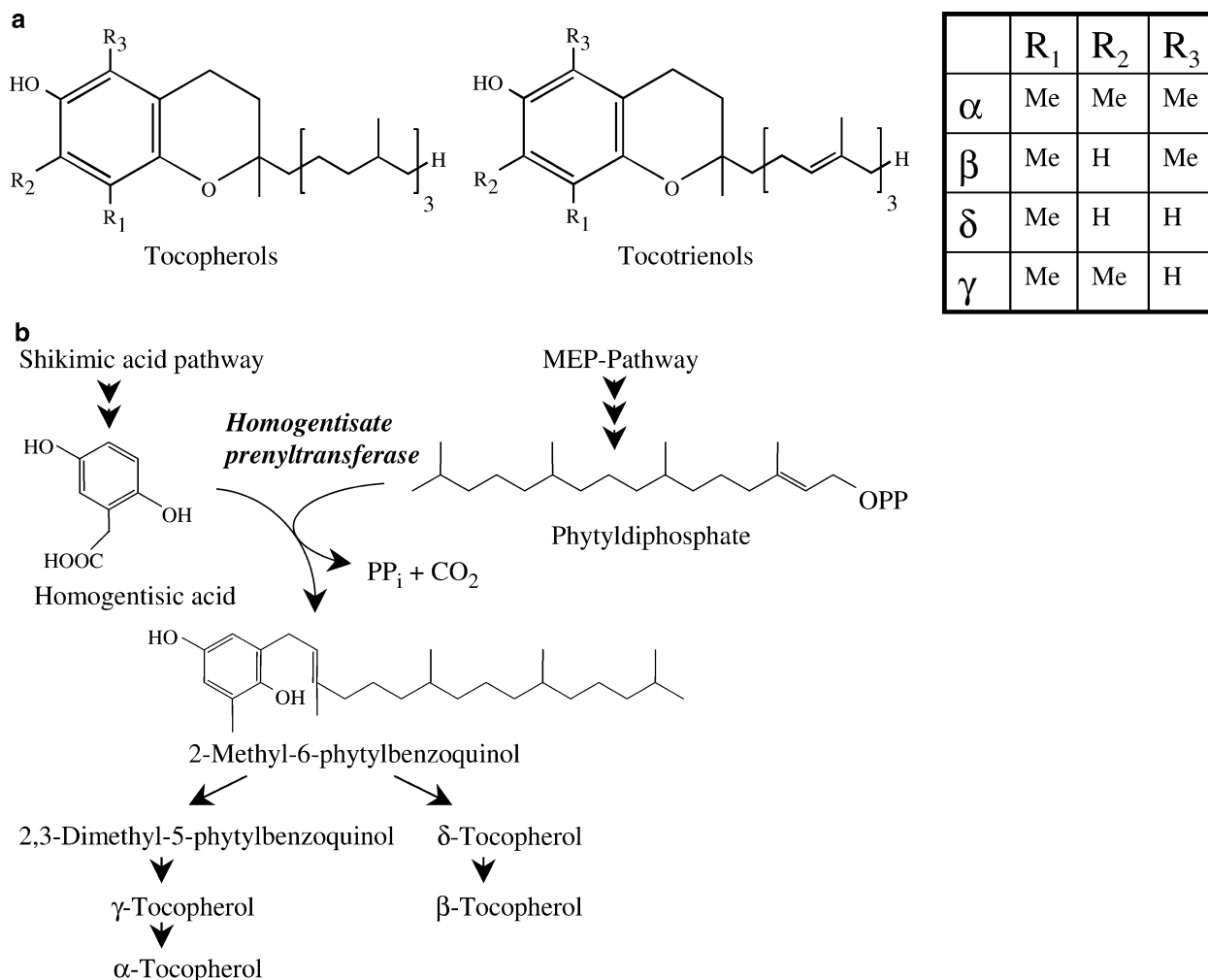
T. V. Venkatesh · B. Karunanandaa  
Monsanto Company, 800 N. Lindbergh Boulevard,  
St. Louis, MO 63167, USA

D. L. Free · J. M. Rottnek · S. R. Baszis  
Monsanto Company, 700 Chesterfield Parkway West,  
Chesterfield, MO 63017, USA

H. E. Valentin (✉)  
Monsanto Company, Calgene Campus, 1920 Fifth Street,  
Davis, CA 95616, USA  
E-mail: [henry.e.valentin@monsanto.com](mailto:henry.e.valentin@monsanto.com)  
Tel.: +1-530-7922136  
Fax: +1-530-7922454

### Introduction

Tocopherols and tocotrienols function as the principle lipid-soluble oxidation chain-breaking antioxidants in biological membranes and lipoproteins (Liebler 1998; Brigelius-Flohé et al. 2002). In contrast to cytoplasmic systems, such as the glutathione redox cycle, or the superoxide dismutase, which depend on enzymatic inactivation of oxygen radicals, the antioxidant reactions of tocochromanols do not require involvement of enzymes. The physiological role of tocopherols and tocotrienols is thought to be the protection of polyunsaturated fatty acids (PUFA) from lipid oxidation by quenching free radicals in cell membranes and other lipophilic environments (Kamal-Eldin and Appelqvist 1996). Tocopherols and tocotrienols occur in four major isoforms each:  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -tocopherol, and  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -tocotrienol (Pongracz et al. 1995; Fig. 1). Tocotrienols are distinguished from tocopherols by the presence of three double bonds in the isoprenoid side chain, and greek letters refer to the grade of methylation on the aromatic head group, with  $\alpha$ -tocopherol or  $\alpha$ -tocotrienol being the highest methylated isoforms



**Fig. 1** Homogentisate prenyltransferase reaction in context of the tocopherol biosynthetic pathway. Tocopherols and tocotrienols with major isoforms (**a**) and schematic drawing of the homogentisate prenyltransferase reaction (**b**). *MEP* methylerythritol phosphate

(Fig. 1). Of these,  $\alpha$ -tocopherol has the highest vitamin E activity in animals and humans (Sheppard et al. 1993; Bramley et al. 2000), presumably due to its preferred retention and distribution throughout the mammalian body (Traber and Sies 1996).

The tocopherol biosynthesis begins with the prenylation of homogentisic acid (HGA) with phytyldiphosphate (PDP) catalyzed by the homogentisate prenyltransferase (*VTE2*; Collakova and DellaPenna 2001; Schledz et al. 2001; Savidge et al. 2002), or by prenylation of HGA with geranylgeranyldiphosphate (GGDP) catalyzed by the homogentisate geranylgeranyl transferase (*HGGT*; Cahoon et al. 2003). HGA and PDP plus GGDP are derived from the shikimate pathway (Norris et al. 1998) and the methylerythritol-phosphate pathway (Rohmer 2003), respectively (Fig. 1). The reaction products of this first committed step, 2-methyl-6-phytylbenzoquinol, or 2-methyl-6-geranylgeranylbenzoquinol are subsequently cyclized and methylated resulting in the formation of various tocopherol isoforms (Fig. 1).

Previous evidence suggests that the first committed reaction in tocopherol biosynthesis regulates tocopherol flux and to some extent tocopherol composition (Savidge et al. 2002; Cahoon et al. 2003; Collakova and DellaPenna 2003a, b; Karunanandaa et al. 2005). *VTE2* has for these reasons been identified to be critical for tocopherol biosynthesis pathway engineering. Interestingly, *VTE2* was discovered via bioinformatics approaches in parallel by two different groups (Collakova and DellaPenna 2001; Savidge et al. 2002). In this study, we report on further bioinformatics characterization of homogentisate prenyltransferases leading to the identification of four conserved sequence motifs and on the application of two of these motifs for database searches to specifically search homogentisate prenyltransferase sequences. Application of these motifs for additional database searches identified a *VTE2* paralog in the Arabidopsis genome. This new *VTE2* paralog was named *VTE2-2* and the previously known *VTE2* was renamed *VTE2-1*. Transgenic overexpression of this new gene supports a function of this new gene in tocopherol biosynthesis.

## Material and methods

### Bioinformatics analysis

BLAST 2.2 algorithms (Altschul et al. 1997) and the non-redundant amino acid database were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov>). Profile Hidden Markov Models (HMMs) for the motifs described were built and profile HMM searches using these motifs were performed using HMMER 2.0 software as described in the users manual (<http://hmmerr.wustl.edu/>). Gene prediction software FGENESH was procured from Softberry Inc., NY, USA. Dicot model provided with this software was used to predict coding regions from Arabidopsis genomic DNA sequences. Multiple sequence alignments using ClustalX (Thompson et al. 1997) were performed as described in the user's manual (<http://www.igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html>). Multiple sequence alignments were visualized and edited using GeneDoc (Nicholas et al. 1997). The residues were shaded using conserved residue shading mode set to level 4 using default settings. Similar amino acid residues conserved in all columns were shaded dark. Phylogenetic analysis was conducted using MEGA version 2.1 (Kumar et al. 2001). Separate neighbor-joining trees were generated using both p-distance and gamma distance amino acid substitution models. Both pair wise and complete deletion approaches for handling gaps or missing data were applied separately for each model. The trees were tested by bootstrapping, performing 500 replications. Subsequently, the trees were rooted for the *Chloroflexus aurantiacus VTE2*.

For further functional characterization of *VTE2-2* the expression patterns of *VTE2-1* and *VTE2-2* genes were analyzed in the Arabidopsis Massively Parallel Signature Sequencing (MPSS) data set available at <http://mpss.udel.edu/at/> (Meyers et al. 2004). MPSS produces short sequence signatures chosen from a defined position within an mRNA, and the relative abundance of these signatures in a given library represents a quantitative estimate of expression of that gene. The MPSS signatures are 17 bp in length, and can uniquely identify >95% of all genes in Arabidopsis.

### Arabidopsis transformation

Using standard cloning techniques the *VTE2-2* full-length cDNA was excised from an in house EST clone (pMON69960) to generate binary vectors for sense and antisense expression under the control of the seed-specific Napin promoter (Kridl et al. 1991) and the constitutive e35S promoter (McPherson and Kay 1994) as described (Valentin et al. 2003). The plant binary vectors, pMON69963, and pMON69965 were used to transform *Arabidopsis thaliana* Col-0 for seed-specific sense and antisense expression of *At-VTE2-2*, respectively. Binary vectors, pMON69964 and pMON69966

were used for constitutive sense and antisense expression of *VTE2-2*, respectively. The binary vectors were transformed into *Agrobacterium tumefaciens* strain ABI by electroporation (Bio-Rad Electroporation Manual, Dower et al. 1988). Transgenic Arabidopsis plants were obtained by *Agrobacterium*-mediated transformation as described (Valvekens et al. 1988; Bechtold et al. 1993; Bent et al. 1994). Transgenic plants were selected by sprinkling the transformed seed onto selection plates containing MS basal salts (4.3 g/l), Gamborg's B-5, 500× (2.0 g/l), sucrose (10 g/l), Mes (0.5 g/l), phytagar (8 g/l), carbenicillin (250 mg/l), cefotaxime (100 mg/l), plant preservation medium (2 ml/l), and kanamycin (60 mg/l) and then vernalizing them at 4°C in the absence of light for 2–4 days. Subsequently, the seed were transferred to 23°C, and 16/8 h light/dark cycle for 5–10 days until seedlings emerged. Once, one set of true leaves were formed on the kanamycin resistant seedlings, plantlets were transferred to soil and grown to maturity. Transgenic lines generated through kanamycin selection were grown under 16 h light and 8 h dark.

### Tocochromanol analysis

The tocochromanol content was analyzed on R2 Arabidopsis seed as described by Savidge et al. (2002).

### Northern analysis

RNA was isolated from 100 mg fully matured but not dried, Arabidopsis siliques. Tissue samples were ground with 0.25 g polyvinylpyrrolidone in liquid nitrogen. While the ground tissues were frozen, 10 ml REC.8+ buffer containing 50 mM Tris-HCl (pH 9.0), 800 mM NaCl, 10 mM EDTA, 0.5% CTAB and 0.5%  $\beta$ -mercaptoethanol was added to each sample, mixed and centrifuged for 5 min at 18,000g in a Sorvall centrifuge, model Super T21. The supernatant was filtered through miracloth (Calbiochem) onto 3 ml chloroform, mixed and centrifuged again at 18,000g for 5 min. The supernatant was separated and extracted twice with an equal volume of phenol:chloroform mix (1:1, v/v) and then ethanol precipitated. The ethanol pellet was re-suspended in 50 mM EDTA. Equal amounts of total RNA (20  $\mu$ g) were fractionated on 1.2% agarose gels containing 1 M formaldehyde. Gels were blotted onto nylon membranes (Ambion) according to the manufacturer's instructions and hybridized in Sigma PerfectHyb<sup>TM</sup> buffer, containing 10 mg/ml of salmon sperm DNA at 65°C. A cDNA fragment (*Xho*I/*Bam*HI) of the Arabidopsis *VTE2-2* gene was labeled and used as a gene-specific probe. Probe labeling with <sup>32</sup>P-dCTP (1.85 MBq) was performed using Amersham Rediprime<sup>TM</sup> labeling kit (Amersham Biosciences Corp.). Blots were washed in 2× SSC, 0.1% SDS at 65°C and exposed to X-ray film (BioMax MS film, Sigma). After detection of *VTE2-2* mRNA the blot was stripped with

0.1× SSC containing 0.1% SDS to remove the radiolabelled *VTE2-2* probe. Subsequently the blot was again pre-hybridized and hybridized with a rDNA radiolabelled probe using the same methodology described above. After 2 h hybridization with this probe, the blot was washed and exposed to X-ray film.

## Results

### Bioinformatic analysis

Motif analysis I: *Arabidopsis thaliana* and *Synechocystis* sp. PCC 6803 *VTE2* genes (*At-VTE2*, and *Sy-VTE2*, respectively) have been described previously (Collakova and DellaPenna 2001; Schledz et al. 2001; Savidge et al. 2002). Using the deduced amino acid sequence of *Sy-VTE2*, the *Nostoc* and *Anabaena* *VTE*-genes were identified from GenBank (gi|17230940, and gi|46135490, respectively) based on sequence homology. In addition, sequences homolog to the *At-VTE2* and *Sy-VTE2* deduced amino acid sequences were obtained from soybean (*Glycine max*), leek (*Allium porrum*) and *Cuphea pulcherrima* cDNA libraries (GenBank accession no. DQ231059, DQ231057, and DQ231058, respectively). The deduced *VTE2* amino acid sequences from *Arabidopsis*, leek, soybean, *Cuphea pulcherrima*, *Synechocystis*, *Nostoc*, and *Anabaena* were aligned and four motifs conserved in all *VTE2* sequences were identified (Fig. S1 in supplemental material). These motifs have been numbered 1.1, 2.1, 3.1 and 4.1, respectively. They correspond to amino acids 99–112, 167–174, 285–298, and 371–390 of *At-VTE2-1* respectively.

We used Hidden Markov Model (HMM) profile techniques (Eddy 1998) to test the specificity and sensitivity of these motifs. Profile HMMs are statistical models of multiple sequence alignments. They capture position-specific information about how conserved each column of the alignment is, and which residues are likely to be sensitive in identifying proteins sharing homology to a conserved domain of interest. Profile HMM models built from multiple alignments of these motifs were used to search all sequences of the NCBI non-redundant amino acid database available on March 2003. Motifs 1.1 through 4.1 identified all previously known deduced *VTE2* amino acid sequences present in the database at high specificity. Surprisingly, this search identified two new genomic *Arabidopsis VTE2* variants representing contradicting predicted protein sequences from the same area of genomic *Arabidopsis* DNA. When both of these sequences (gi15229898 and gi10998133) were used to search the non-redundant amino acid database, the BLAST search results indicated that these sequences were most related to *VTE2* sequences from cyanobacteria and *Arabidopsis VTE2-1* (from here on referred to as *VTE2-1*). However, gi15229898 appeared to encode a much larger protein of 970 amino acids, which was homologous to *VTE2-1* over its carboxy terminal half. A blast search using the 561 amino-terminal deduced

amino acids of gi15229898 that were not homologues to *VTE2-1*, identified homology to deduced amino acid sequences from genes of unknown function from *Musa acuminata* (gi|40850572, 53% identity over 310 amino acids), *Oryza sativa* (gi|50930768, 44% identity over 335 amino acids), *Zea mays* (gi|54652075, 41% identity over 224 amino acids), and TNF receptor-associated factor 5 from *Gallus gallus* (gi5|4020685, 27% identity over 179 amino acids). PFAM (<http://www.sanger.ac.uk/Software/Pfam/index.shtml>) analysis of this region identified a conserved TRAF-type zinc finger (Zf-TRAF) domain. Mammalian signal transducers associated with the cytoplasmic domain of the 75 kDa tumor necrosis factor (TNF) receptor are known to contain this domain. The presence of a putative Zf-TRAF domain may therefore indicate a regulatory function of the gi|15229898 amino terminal sequence. Analysis of plant EST databases indicated that the deduced amino terminal sequence of gi|15229898 that was not homologous to *VTE2-1* may code for a gene of functional significance which is conserved in plants. EST sequences of this gene were identified in multiple plant species including *Arabidopsis thaliana*, *Brassica napus*, *Zea mays* and *Triticum aestivum* (data not shown). However, all cDNA sequences with homology to gi|15229898 lacked the carboxy-terminal region that contained sequence similarities to *VTE2*, suggesting that the N-terminal region of gi|15229898 is transcribed independent from the sequence that was homolog to *VTE2-1*.

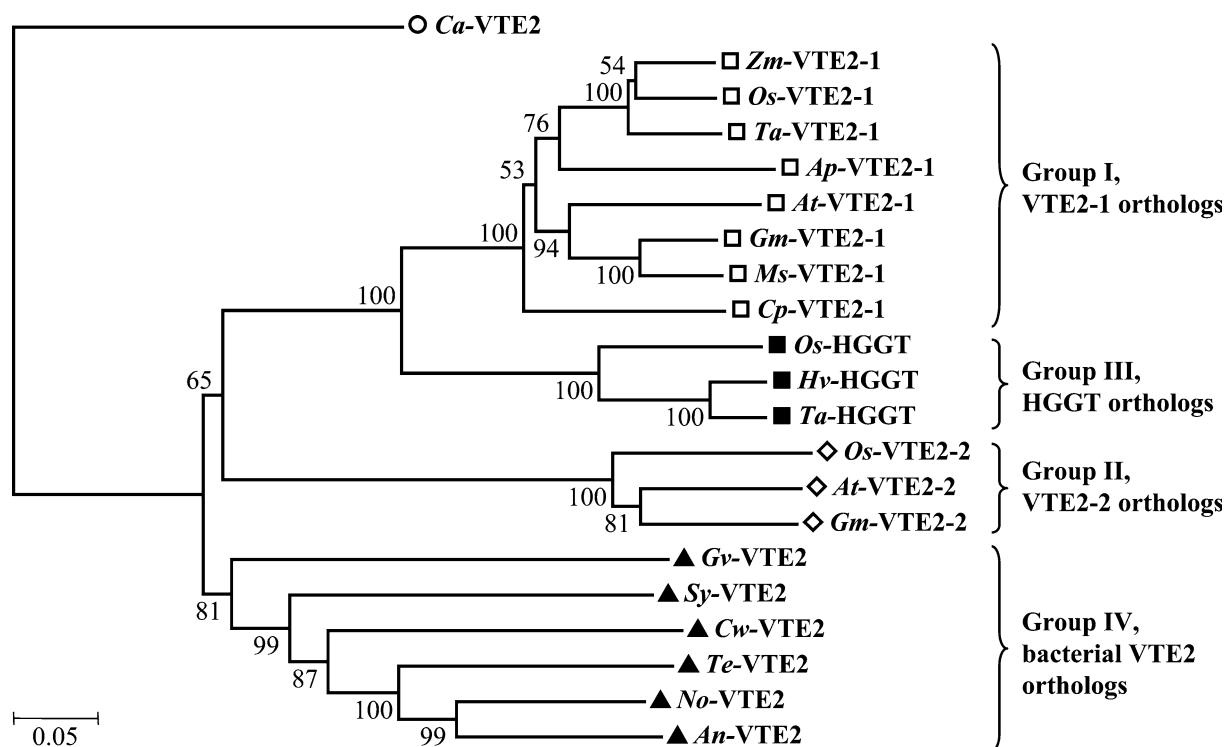
Gi|10998133 encoded a 441 amino acid protein that was highly similar to *VTE2-1* but it appeared to miss approximately 40–50 amino acids from its carboxy terminus. Gene prediction algorithms are often mistaken in predicting terminal exons (Zhang 2002). Such predicted proteins are often found to have misannotated N-terminal and/or C-terminal sequences and require further verification. In order to verify the GenBank sequence prediction, the nucleotide sequence of a BAC clone that corresponded to the *Arabidopsis* genomic sequence that contained *VTE2*-motifs (gi|12408742|gb|A-C016795.6|ATAC016795, 10,0835 bp) was analyzed, and its coding sequences were predicted using the FGENESH gene prediction program (Solovyev 2001). FGENESH identified 28 coding regions on this BAC clone (data not shown). To identify new homogenisate phytyltransferase proteins among these 28 sequences, all 28 predicted amino acid sequences were searched against the non-redundant amino acid database using BLAST tool (Altschul et al. 1997). Predicted protein no. 25 (base 80,571 to 83,410 on the BAC clone), with a length of 402 amino acids was most similar to gi10998133 (441 amino acids), the carboxyl terminal half of gi15229898 (970 aa) and other deduced *VTE2* amino acid sequences. The deduced amino acid sequence of protein no. 25 aligned well with other *VTE2*s over the entire length of the protein (data not shown). To provide functional and transcriptional evidence and to confirm the coding sequence for this gene, plant EST sequence databases were searched. Several ESTs from *Arabidopsis* matched the



amino terminal and carboxy terminal portions of this gene (data not shown). ESTs with significant similarity were also identified from other plant species (data not shown). The new gene was named *VTE2-2* (GenBank accession no. DQ231060) and previously known *VTE2s* from plant sources were renamed *VTE2-1*. The deduced *At-VTE2-2* amino acid sequence is quite distinct from *At-VTE2-1* and is about 32% identical (Figs. 2 and 3). Similar as all plant derived *VTE2-1* sequences represented in Fig. 2, the deduced amino acid sequences of all *VTE2-2* represented in Fig. 2 contain a predicted chloroplast target peptide (CTP) according to ChloroP (Emanuelsson et al. 1998).

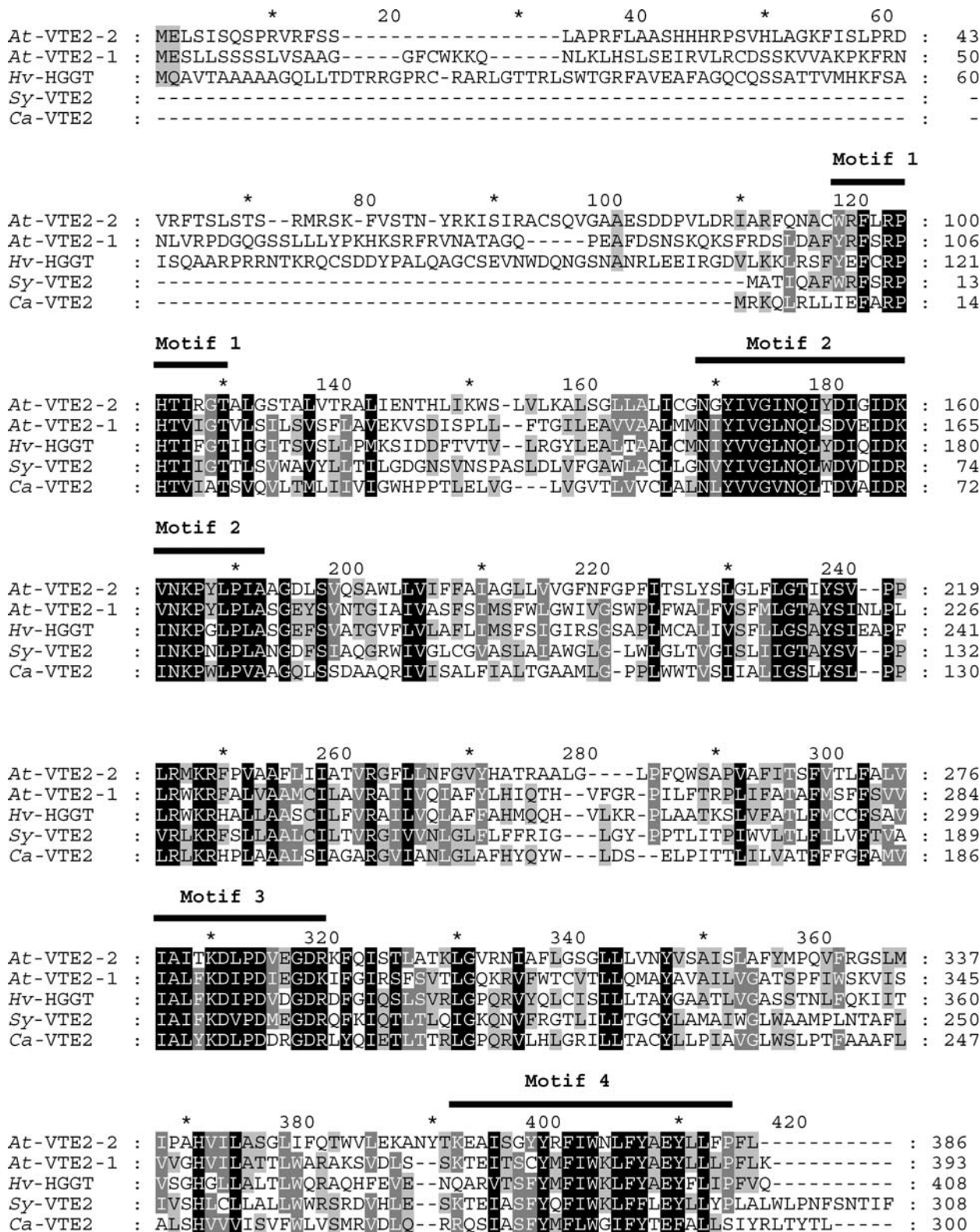
**Motif analysis II:** Additional EST database searches using the previously identified motifs revealed that *At-VTE2-2* orthologs were present in soybean and rice as well. These newly identified *VTE2-2* sequences were cloned and their sequences included in the multiple sequence alignment used in the previous motif analysis, for further motif optimization. The resulting new motifs were designated 1.2, 2.2, 3.2, and 4.2 (Fig. 4) corre-

sponding to motifs 1.1, 1.2, 3.1 and 4.1 from previous motif analysis (Fig. S1). Profile HMM models were built and HMM searches were performed using the optimized motifs on the NCBI non-redundant amino acid database, containing more than 2.1 million sequences (Nov. 2004). The results of this search are shown in Table 1. All four motifs identified all previously known homogentisate phytyltransferase genes from plants and cyanobacteria found in the database. In addition to *VTE2-1* and *VTE2-2* sequences, these motifs also identified the recently identified *HGGT* sequences from monocotyledon plants (Cahoon et al. 2003). Motifs 3.2 and 4.2 were found to be highly specific to VTE2 type amino acid sequences as evident by lower *E* values. Motif 4.2 also identified a VTE2 related sequences from the photosynthetic bacterium *Chloroflexus aurantiacus*. Motif 1.2 is also specific to *VTE2* type genes; however, its sensitivity for HGGT sequences is low as indicated by higher *E* values. Motif 2.2 identified five bacterial *ubiA* prenyltransferase sequences in addition to *VTE2s* and *HGGTs* as it spans the conserved catalytic domain of



**Fig. 2** Neighbour-joining tree of homogentisate prenyltransferases. The tree has been rooted for *Chloroflexus aurantiacus* (*Ca*) VTE2. Branch length calculated based on amino acid substitutions per site is shown below (0.05=5% difference between sequences). *Blank squares* VTE2-1 sequences from plants, *blank diamonds*, VTE2-2 sequences from plants, *filled squares* HGGT sequences from plants, *filled triangles* VTE2 sequences from cyanobacteria, *blank circle* VTE2 from photobacterium. Abbreviations: *An* *Anabaena*, *Ap* *Allium porrum*, *At* *Arabidopsis thaliana*, *Cp* *Cuphea pulcherrima*, *Cw*, *Crocospaera watsonii*, *Gm* *Glycine max*, *Gv* *Gloeobacter violaceus*, *Ms* *Medicago sativa*, *No* *Nostoc*, *Os* *Oryza sativa*, *Sy* *Synechocystis*, *Ta* *Triticum*

*aestivum*, *Te* *Trichodesmium erythraeum*, *Zm* *Zea mays*. Reference numbers: *An-VTE2*, gi|46135490; *Ap-VTE2-1*, accession no. DQ231057; *At-VTE2-1*, gi|21281072; *At-VTE2-2*, accession no. DQ231060; *Ca-VTE2*, gi|53795310; *Cp-VTE2-1*, accession no. DQ231058; *Cw-VTE2*, gi|45525087; *Gm-VTE2-1*, accession no. DQ231059; *Gm-VTE2-2*, accession no. DQ231061; *Gv-VTE2*, gi|37519852; *Hv-HGGT*, gi|33391138; *Ms-VTE2-1*, gi|51949754; *No-VTE2*, gi|17230940; *Os-HGGT*, gi|33391144; *Os-VTE2-1*, gi|51536170; *Os-VTE2-2*, gi|50938601; *Sy-VTE2*, gi|16330366; *Ta-HGGT*, gi|33391142; *Ta-VTE2-1*, accession no. DQ231056; *Te-VTE2*, gi|48893591; *Zm-VTE2-1*, accession no. DQ231055



**Fig. 3** Multiple alignment of selected homogentisate prenyltransferases. Motif 1, 2, 3, and 4 are labeled with a bold line. Species abbreviations *At* *Arabidopsis thaliana*, *Hv* *Hordeum vulgare*, *Sy* *Synechocystis*, *Ca* *Chloroflexus aurantiacus*

prenyltransferases (Collakova and DellaPenna 2001). However, the *E*-values for *ubiA* sequences were higher by several orders of magnitude, indicating a lower se-

quence homology and greater sequence divergence of tocochromanol related prenyltransferases from *ubiA* type prenyltransferases.

**Motif 1.2 (Motif 1.1 in Supplemental Fig. 1)**

At-VTE2-1 : YRFSRPHTVIGT : 12  
 No-VTE2 : WKFSRPHTIIGT : 12  
 An-VTE2 : WKFSRPHTIIGT : 12  
 Sy-VTE2 : WRFSRPHTIIGT : 12  
 Zm-VTE2-1 : YRFSRPHTVIGT : 12  
 Ta-VTE2-1 : YRFSRPHTIIGT : 12  
 Ap-VTE2-1 : YKFSRPHTIIGT : 12  
 Gm-VTE2-1 : YRFSRPHTVIGT : 12  
 Cp-VTE2-1 : YRFSRPHTIIGT : 12  
 At-VTE2-2 : WRFLRPHTIRGT : 12  
 Gm-VTE2-2 : WRFLRPHTIRGT : 12  
 Os-VTE2-2 : WRFLRPHTIRGT : 12

Consensus : (YW) (RK) F (SL) RPHT (IV) (IR) GT

**Motif 2.2 (Motif 2.1 in Supplemental Fig. 1)**

At-VTE2-1 : NIYIVGLNQLSDVEIDKVNKPYLPLA : 26  
 No-VTE2 : NVYIVGLNQLEDVDIDKINKPHLPLA : 26  
 An-VTE2 : NVYIVGLNQLEDIEIDKVNKPHLPLA : 26  
 Sy-VTE2 : NVYIVGLNQLWDVDIDRINKPNLPLA : 26  
 Zm-VTE2-1 : NIYIVGLNQLFDIEIDKVNKPTLPLA : 26  
 Ta-VTE2-1 : NIYIVGLNQLFDIEIDKVNKPTLPLA : 26  
 Ap-VTE2-1 : NVYIVGLNQLEDIEIDKVNKPDPLA : 26  
 Gm-VTE2-1 : NIYIVGLNQLSDVEIDKINKPYLPLA : 26  
 Cp-VTE2-1 : NIYIVGLNQLSDIDIDKVNKPYLPLA : 26  
 At-VTE2-2 : NGYIVGINQIYDIGIDKVNKPYLPIA : 26  
 Gm-VTE2-2 : NGYIVGINQIYDISIDKVNKPYLPIA : 26  
 Os-VTE2-2 : NGYIVGINQIYDIRIDKVNKPYLPIA : 26

Consensus : N (GVI) YIVG (LI) NQ (LI) (SEFWY) D (IV) (EDSRG) IDK (IV) NKP (YDTNH) LP (LI) A

**Motif 3.2 (Motif 3.1 in Supplemental Fig. 1)**

At-VTE2-1 : IALFKDIPDIEGDK : 14  
 No-VTE2 : IAIKDIKDIEGDR : 14  
 An-VTE2 : IAIKDIKDIEGDR : 14  
 Sy-VTE2 : IAIKDIKDIEGDR : 14  
 Zm-VTE2-1 : IALFKDIPDIEGDR : 14  
 Ta-VTE2-1 : IALFKDIPDIEGDR : 14  
 Ap-VTE2-1 : IALFKDIPDIEGDK : 14  
 Gm-VTE2-1 : IALFKDIPDIEGDK : 14  
 Cp-VTE2-1 : IALFKDIPDIEGDK : 14  
 At-VTE2-2 : IAITKDLDPVEGDR : 14  
 Gm-VTE2-2 : IAITKDLDPVEGDR : 14  
 Os-VTE2-2 : IAITKDLDPVEGDR : 14

Consensus : IA (LI) (FT) KD (IVL) PD (VIM) (ED) GD (KR)

**Motif 4.2 (Motif 4.1 in Supplemental Fig. 1)**

At-VTE2-1 : KTEITSYMFYIWKLFYAEYLLLP : 23  
 No-VTE2 : KSAIAQFYQFIWKLFYAEYLLFP : 23  
 An-VTE2 : KTAIAQFYQFIWKLFYAEYLLMFP : 23  
 Sy-VTE2 : KTEIASFYQFIWKLFYAEYLLYP : 23  
 Zm-VTE2-1 : KAAITSYMFYIWKLFYAEYLLIP : 23  
 Ta-VTE2-1 : KAAITSFYMLIWRWFYAEYLLIP : 23  
 Ap-VTE2-1 : KTTITSFYMFVWVWKLFYAEYLLIP : 23  
 Gm-VTE2-1 : KASITSFYMFYIWKLFYAEYLLIP : 23  
 Cp-VTE2-1 : KAAITSFYMFYIWKLFYAEYLLIP : 23  
 At-VTE2-2 : KEAISGYRFIWNLFYAEYLLFP : 23  
 Gm-VTE2-2 : KDAISGYRFIWNLFYAEYLLFP : 23  
 Os-VTE2-2 : KDAISQYRFIWNLFYAEYLLFP : 23

Consensus : K (TASED) (EATS) I (TAS) (SQG) (FC) Y (QMR) (FL) (IV) W (KRN) LF (FY) (ALI) EY (LAI) (LIFM) (IFYL) P

**Fig. 4** Optimized VTE-2 sequence motifs identified based on multiple alignment of VTE2 sequences. Numbers on the right border represent the number of amino acid residues in each motif. Consensus sequences are shown for each motif. Abbreviations: An

*Anabaena*, *Ap Allium porrum*, *At Arabidopsis thaliana*, *Cp Cuphea pulcherrima*, *Gm Glycine max*, *No Nostoc*, *Sy Synechocystis*, *Ta Triticum aestivum*, *Zm Zea mays*

**Table 1** HMM search of non-redundant NCBI amino acid databases

Gi No.	Species	Gene	Length	Motif 1.2	Motif 2.2	Motif 3.2	Motif 4.2	Comments
30681973	<i>Arabidopsis thaliana</i>	<i>VTE2-2</i>	954	0.072	2.9E-11	0.011	9.2E-09	N-terminal sequence misannotated
50938601	<i>Oryza sativa</i> (japonica cultivar-group)	<i>VTE2-2</i>	379	0.072	1.4E-11	0.011	3E-08	
51963498	<i>Oryza sativa</i> (japonica cultivar-group)	<i>VTE2-2</i>	104	0.072				Incomplete sequence, not detected by blast search
10998133	<i>Arabidopsis thaliana</i>	<i>VTE2-2</i>	441	0.072	2.9E-11	0.011		C-terminal sequence misannotated
21281072	<i>Arabidopsis thaliana</i>	<i>VTE2-1</i>	393	0.031	1.1E-12	0.0054	7.9E-09	
3004556	<i>Arabidopsis thaliana</i>	<i>VTE2-1</i>	210		1.1E-12			Incomplete sequence
51536164	<i>Oryza sativa</i> (japonica cultivar-group)	<i>VTE2-1</i>	188		6.20E-07			Incomplete sequence
51536170	<i>Oryza sativa</i> (japonica cultivar-group)	<i>VTE2-1</i>	402		1.3E-12	0.0026	3.6E-09	N-terminal sequence misannotated
51949754	<i>Medicago sativa</i>	<i>VTE2-1</i>	411	0.031	1.6E-12	0.0054	1.1E-07	
16330366	<i>Synechocystis</i> sp. PCC 6803	<i>VTE2</i>	308	0.011	4.4E-12	0.0026	8E-08	
17230940	<i>Nostoc</i> sp. PCC 7120	<i>VTE2</i>	318	0.013	7.2E-13	0.0017	2.5E-08	
23125851	<i>Nostoc punctiforme</i> PCC 73102	<i>VTE2</i>	322	0.013	1.9E-12	0.0021	5.4E-08	
37519852	<i>Gloeobacter violaceus</i> PCC 7421	<i>VTE2</i>	298	0.56	1.3E-11	0.0017	8.80E-06	
45525087	<i>Crocospaera watsonii</i> WH 8501	<i>VTE2</i>	315	0.17	2.5E-11	0.029	9.2E-08	
46135490	<i>Anabaena variabilis</i> ATCC 29413	<i>VTE2</i>	318	0.013	7.2E-13	0.0017	2.5E-08	
48893591	<i>Trichodesmium erythraeum</i> IMS101	<i>VTE2</i>	349	0.046	3.5E-12	0.0021	1.5E-07	
53795310	<i>Chloroflexus aurantiacus</i>	<i>VTE2</i>	300		5.8E-09	7.8	0.024	
33391138	<i>Hordeum vulgare</i>	<i>HGGT</i>	408	3.5	5.2E-11	0.0093	1.70E-06	
33391142	<i>Triticum aestivum</i>	<i>HGGT</i>	408	2	5.2E-11	0.0093	1.70E-06	
33391144	<i>Oryza sativa</i> (japonica cultivar-group)	<i>HGGT</i>	404	0.62	1.3E-11	0.0095	1.50E-06	
51535642	<i>Oryza sativa</i> (japonica cultivar-group)	<i>HGGT</i>	270		1.3E-11	0.0095	1.50E-06	Short Sequence
15921742	<i>Sulfolobus tokodaii</i> str. 7	<i>ubiA</i>			0.06			
15897501	<i>Sulfolobus solfataricus</i> P2	<i>ubiA</i>			0.12			
41719117	<i>Methanococcoides burtonii</i> DSM 6242	<i>ubiA</i>			0.17			
15668454	<i>Methanocaldococcus jannaschii</i> DSM 2661	<i>ubiA</i>			0.17			
45359240	<i>Methanococcus maripaludis</i> S2	<i>ubiA</i>			2.1			

Numbers in motif columns indicate *E* values

## Phylogenetic analysis

Four neighbor-joining trees were generated using two different amino acid substitution models (see Materials and methods) and two different approaches for handling gaps or missing data. All four neighbor-joining trees clustered the homogentisate phytyltransferases in four clades. One representing plant *VTE2-1* sequences, a separate clade representing plant *VTE2-2* sequences, a third representing *HGGT* sequences, and a fourth clade representing bacterial homogentisate phytyltransferases. A tree generated using p distance and complete deletion model is shown in Fig. 2. The topology of the tree suggests that the bacteria and cyanobacteria harbor only one homogentisate phytyl transferase gene. Dicotyledonous and monocotyledonous plants harbor two and

three homogentisate prenyltransferase genes, respectively.

## Expression pattern of *VTE2-1* and *VTE2-2* in Arabidopsis

For characterization of the functional significance of *VTE2-2* the Arabidopsis expression patterns of the two *VTE2*-paralogs were characterized in silico. While the frequency of *VTE2-1* and *VTE2-2* hits in public and proprietary cDNA-libraries from flower buds, seed, silique, roots and leaves suggested a similar expression pattern of these two genes (data not shown), a search in the more quantitative Arabidopsis MPSS data set (<http://www.mpss.udel.edu/at/>) revealed a preferred



expression of *VTE2-1* in inflorescence tissue, and a preferred expression of *VTE2-2* in leaf (Fig. 5). Expression levels in callus, root and silique tissues were found to be similar.

#### Overexpression and antisense of *VTE2-2* in Arabidopsis

Expression of *At-VTE2-2* under napin promoter control resulted in 38% increased seed tocopherol levels in the R2 seed population of all events (Table 2). This increase is slightly less than the increase obtained with *VTE2-1* using the same promoter and 3'-UTR (Savidge et al. 2002), but it represents a significant change in total seed tocopherol content. Constitutive expression of *VTE2-2* under e35S promoter control resulted in a significant tocopherol increase of 13% compared to the vector control (Table 2).  $\beta$ -Tocopherol was not detected in any of the *VTE2-2* over expressing lines, and  $\gamma$ -tocotrienol was found only in trace amounts ( $\leq 7$  ppm) of some samples harboring the napin driven expression construct for *VTE2-2* (Table 2). Mean tocopherol levels of all events in Arabidopsis seed harboring seed-specific and constitutive *VTE2-2* antisense constructs were not significantly different from tocopherol levels found in control seed populations. However, these constructs represented simple antisense constructs, which are frequently found to provide only a low percentage of events that exhibit the antisense effect.

#### Expression analysis of transgenic *VTE2-2* expressers

Siliques from seven selected R3, kanamycin resistant, transgenic lines originating from transformation with pMON69963 and an empty vector control line were used for Northern analysis to detect the expression level of the *VTE2-2* transgene. All transgenic lines exhibited a signal migrating at a molecular mass of 1.2 kb

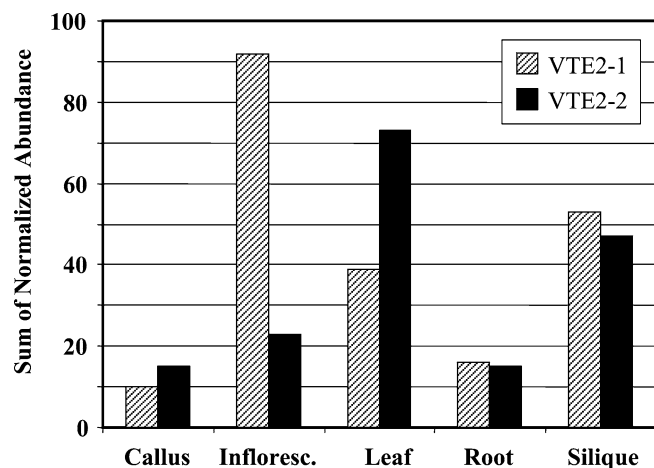


Fig. 5 Expression pattern of *VTE2* paralogs in Arabidopsis. The expression pattern are based on analysis of the Arabidopsis MPSS data set (<http://mpss.udel.edu/at/>)

corresponding with the expected molecular mass of the *VTE2-2* transgene mRNA that was not present in wild-type control preparations. Transgenic lines with low, medium and high tocopherol increase were used for this experiment. The transgene expression level in the lines analyzed correlated with the level of tocopherol increase in R2 seed (Fig. 6).

## Discussion

In this study, we describe a motif analysis approach which identified a set of conserved motifs in *VTE2-1* and *HGGT* amino acid sequences. Application of these motifs in sequence database searches resulted in the discovery of a *VTE2*-paralog in the Arabidopsis genome. A function of this new gene in tocopherol biosynthesis was confirmed through over expression in Arabidopsis resulting in increased seed tocopherol levels. Interestingly, this new gene maps to the same location as a so far not characterized mutation on the top of chromosome 3 (*pds2*) that caused a tocopherol deficient phenotype (Norris et al. 1995).

The conserved motifs were identified through alignment of homologous sequences. Profile HMMs constructed from multiple alignments of these motifs were used in sequence database searches. In comparison to PFAM domains (<http://pfam.wustl.edu>) the motifs used here are small, vary in size from 12 to 30 amino acids, and are specific to a subset of proteins of a large protein family. In contrast, the corresponding PFAM motif is approximately 260 amino acids long and identifies a large group of prenyltransferases including *ubiA* and *VTE2* orthologs found in multiple taxonomic groups. Motifs 3.2 and 4.2 described in this paper are specific to *VTE2* and *HGGT* sequences found in bacteria and higher plants. While motif 2 covers the catalytic domain found in other *ubiA* family prenyltransferases, the functional significance of the other conserved domains identified in this study is not known. Additional studies using *VTE2* or *HGGT* genes mutagenized in the conserved regions identified here may help to decipher the structural and biochemical functions of these amino acid sequences. We found this approach sensitive in identifying short sequences that can be easily missed by sequence similarity search methods like BLAST.

It is interesting to note that all other tocopherol pathway genes appear to occur as single copy genes in Arabidopsis. The 2-methyl-6-phytylbenzoquinolmethyltransferase (*VTE3*) even appears to carry out reactions in two different metabolic pathways, the methylation of 2-methyl-6-phytylbenzoquinol for tocopherol biosynthesis, and the methylation of the plastoquinol precursor 2-methyl-6-solanylbenzoquinol (Cheng et al. 2003; Motohashi et al. 2003). In this context it may appear surprising to find two *VTE2* paralogs in the Arabidopsis genome. However, *VTE2* catalyzes the first committed reaction in tocopherol biosynthesis. These reactions frequently coincide with enzymes regulating the pathway

**Table 2** Sense and antisense expression of *VTE2-2* in *Arabidopsis thaliana*

Plasmid	Relevant genotype	Events analyzed	Mean tocopherol composition <sup>a</sup> (ppm)			Mean total tocopherol <sup>a</sup> (ppm)	Significance of mean
			$\alpha$ -	$\gamma$ -	$\delta$ -		
pMON10098	Empty vector control	10	12.7 ± 0.6	410 ± 9.9	16.9 ± 0.5	439 ± 10.7	C
pMON69963 <sup>b</sup>	pNapin:: <i>VTE2-2</i> <sub>At</sub> sense::napin 3'	27	13.2 ± 0.4	569 ± 8.4	25.8 ± 0.6	608 ± 8.8	A
pMON69964	pe35S:: <i>VTE2-2</i> <sub>At</sub> sense::E9 3'	32	12.4 ± 0.3	464 ± 5.9	20.3 ± 0.5	497 ± 6.1	B
pMON69965	pNapin:: <i>VTE2-2</i> <sub>At</sub> antisense::napin 3'	31	12.7 ± 0.3	419 ± 3.2	17.7 ± 0.4	449 ± 3.5	C
pMON69966	pe35S:: <i>VTE2-2</i> <sub>At</sub> antisense::E9 3'	31	14.4 ± 1.7	412 ± 5.7	16.9 ± 0.4	443 ± 5.7	C

Statistical data evaluation was performed according to Tukey–Kramer. Populations with different letters represent populations with statistically significant different total tocopherol levels at  $\alpha=0.05$

<sup>a</sup>Mean values with standard error

<sup>b</sup>Traces of  $\gamma$ -tocotrienol ( $\leq 7$  ppm) were detected in some of the samples

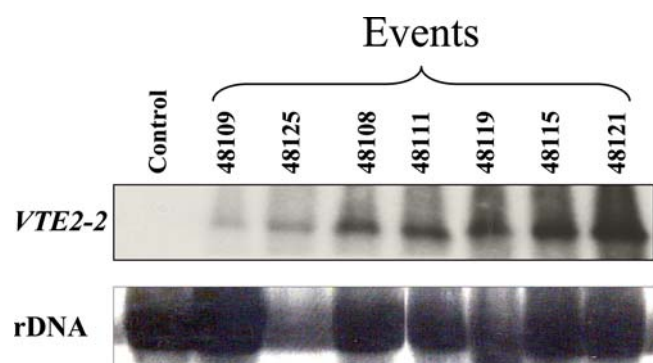
flux. Indeed several recent publications suggest a flux limitation for tocopherol biosynthesis by *VTE2* under certain growth conditions (Savidge et al. 2002; Cahoon et al. 2003; Collakova and DellaPenna 2003a, b; Karunanandaa et al. 2005). The occurrence of two *VTE2* paralogs may provide a broader flexibility in tocopherol flux regulation, or it may help to separate tocopherol flux regulation from plastoquinol biosynthesis. This would however require differences in substrate specificity for the two *VTE2* paralogs. A slight difference in the substrate spectrum of *VTE2-1* and *VTE2-2* may be suggested by the observation that some *Arabidopsis* seed samples expressing *VTE2-2* under napin promoter control contained traces of  $\gamma$ -tocotrienol (Table 2), while a similar construct expressing *VTE2-1* did not result in tocotrienols accumulation (Karunanandaa et al. 2005). *VTE2-2* might therefore also help to explain the accumulation of tocotrienols in transgenic *Arabidopsis* in the presence of high HGA-levels.

Previous studies (Rippert et al. 2004; Karunanandaa et al. 2005) have demonstrated the accumulation of

substantial tocotrienol levels in plants co-expressing a prephenate dehydrogenase and a *p*-hydroxyphenylpyruvate dioxygenase. The accumulation of tocotrienols could not be explained on the basis of currently published data, as previous studies did not detect prenyl-transferase activity of the *Arabidopsis VTE2-1* with the tocotrienol precursor GGDP (Collakova and DellaPenna 2001). The identification of *VTE2-2* and the data presented in Table 2 might suggest that this enzyme can utilize GGDP as substrate, and thereby solve the previous dilemma on how tocotrienols were formed. Whether *VTE2-2* has a function in plastoquinol biosynthesis or fulfills other as yet to be identified physiological functions will require additional experimentation, such as enzyme assays using a variety of different substrates, or parallel complementation experiments to elucidate if the *pds2* mutant as described by Norris et al. (1995) can be complemented by *VTE2-1* or *VTE2-2* or both.

Analysis of the *Arabidopsis* MPSS data set suggested that compared to *VTE2-2* the *VTE2-1* gene is expressed at substantially higher levels in inflorescence and at lower expression levels in leaf. Further enzymatic characterization of both *VTE2* expression products with an emphasis on substrate preferences will help to decipher the physiological function of these two genes.

**Acknowledgements** Tyamagondlu V. Venkatesh and Balasulojini Karunanandaa have equally contributed to this work. We would like to thank Jon Schumeke and team with Taqman analysis and Jing Dong Liu for critical review of the manuscript.



**Fig. 6** Northern-blot analysis for detection of *At-VTE2-2* transcripts in R3 transgenic *Arabidopsis* siliques harboring the *pNapin::VTE2-2* expression cassette. The top panel indicates detection of the *VTE2-2* mRNA from total RNA isolated from transgenic siliques. The bottom panel shows the same northern blot that was used for detection of *VTE2-2*, re-probed with rDNA. Among the lines compared, events 48109 and 48125 represent low (1.2-fold); 48108, 48111, and 48119 represent medium (1.4-fold), and 48115 and 48121 represent high (1.5-fold) seed tocopherol lines for seed of the R2 generation

## References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Bechtold N, Ellis J, Pelletier G (1993) In planta *Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C R Acad Sci III* 316:1194–1199
- Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ (1994) RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science* 265:1856–1860

- Bramley PM, Elmadfa I, Kafatos A, Kelly FJ, Manios Y, Roxborough HE, Schuch W, Sheehy PJA, Wagner KH (2000) Vitamin E. *J Sci Food Agric* 80:913–938
- Brigelius-Flohé R, Kelly FJ, Salonen JT, Neuzil J, Zingg JM, Azzi A (2002) The European perspective on vitamin E: current knowledge and future research. *Am J Clin Nutr* 76:703–716
- Cahoon EB, Hall SH, Ripp KG, Ganzke TS, Hitz WD, Coughlan SJ (2003) Metabolic redesign of vitamin E biosynthesis in plants for tocotrienol production and increased antioxidant content. *Nat Biotechnol* 21:1082–1087
- Cheng Z, Sattler S, Maeda H, Sakuragi Y, Bryant DA, DellaPenna D (2003) Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes. *Plant Cell* 15:2343–2356
- Collakova E, DellaPenna D (2001) Isolation and functional analysis of homogentisate phytyltransferase from *Synechocystis* sp. PCC 6803 and *Arabidopsis*. *Plant Physiol* 127:1113–1124
- Collakova E, DellaPenna D (2003a) Homogentisate phytyltransferase activity is limiting for tocopherol biosynthesis in *Arabidopsis*. *Plant Physiol* 131:632–642
- Collakova E, DellaPenna D (2003b) The role of homogentisate phytyltransferase and other tocopherol pathway enzymes in regulation of tocopherol biosynthesis during abiotic stress. *Plant Physiol* 133:930–940
- Dower WJ, Miller JE, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* 16:6127–6145
- Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763
- Emanuelsson O, Nielsen H, von Heijne G (1999) ChloP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* 8:978–984
- Kamal-Eldin A, Appelqvist L (1996) The chemistry and antioxidant properties of tocopherols and tocotrienols. *Lipids* 31:671–701
- Karunanandaa B, Qi Q, Hao M, Baszsis S, Jensen P, Wong YH, Jiang J, Venkatramesh M, Gruys KJ, Moshiri F, Post-Beittenmiller D, Weiss JD, Valentin HE (2005) Metabolically engineered oilseed crops with enhanced seed tocopherol. *Metab Eng* 7:384–400
- Kridl JC, McCarter DW, Rose RE, Scherer DE, Knutzon DS, Radke SE, Knauf VC (1991) Isolation and characterization of an expressed napin gene from *Brassica rapa*. *Seed Sci Res* 1:209–219
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Liebler DC (1998) Antioxidant chemistry of  $\alpha$ -tocopherol in biological systems; roles of redox cycles and metabolism. *Subcell Biochem* 30:301–317
- McPherson JC, Kay R (1994) DNA sequence for enhancing the efficiency of transcription. US Patent 5,322,938
- Meyers BC, Tej SS, Vu TH, Haudenschild CD, Agrawal V, Edberg SB, Ghazal H, Decola S (2004) The use of MPSS for whole-genome transcriptional analysis in *Arabidopsis*. *Genome Res* 14:1641–1653
- Motohashi R, Ito T, Kobayashi M, Taji T, Nagata N, Asami T, Yoshida S, Yamaguchi-Shinozaki K, Shinozaki K (2003) Functional analysis of the 37 kDa inner envelope membrane polypeptide in chloroplast biogenesis using a *Ds*-tagged *Arabidopsis* pale-green mutant. *Plant J* 34:719–731
- Nicholas KB, Nicholas HB Jr, Deerfield DW II (1997) GeneDoc: analysis and visualization of genetic variation. *Embnet News* 4:1–4
- Norris SR, Barrette TR, DellaPenna D (1995) Genetic dissection of carotenoid synthesis in *Arabidopsis* defines plastoquinone as an essential component of phytoene desaturation. *Plant Cell* 7:2139–2149
- Norris SR, Shen X, DellaPenna D (1998) Complementation of the *Arabidopsis pds1* mutation with the gene encoding *p*-hydroxyphenylpyruvate dioxygenase. *Plant Physiol* 117:1317–1323
- Pongracz G, Weiser H, Matzinger D (1995) Tocopherols—natural antioxidants. *Fett Wiss Technol* 97:90–104
- Rippert P, Scimemi C, Dubald M, Matringe M (2004) Engineering plant shikimate pathway for production of tocotrienol and improving herbicide resistance. *Plant Physiol* 134:92–100
- Rohmer M (2003) Mevalonate-independent methylerythritol phosphate pathway for isoprenoid biosynthesis. Elucidation and distribution. *Pure and Appl Chem* 75:375–387
- Savidge B, Weiss JD, Wong Y-HH, Lassner MW, Mitsky TA, Shewmaker CK, Post-Beittenmiller D, Valentin HE (2002) Isolation and characterization of tocopherol phytyltransferase genes from *Synechocystis* sp. PCC 6803 and *Arabidopsis*. *Plant Physiol* 129:321–332
- Schledz M, Seidler A, Beyer P, Neuhaus G (2001) A novel phytyltransferase from *Synechocystis* sp. PCC 6803 involved in tocopherol biosynthesis. *FEBS Lett* 499:15–20
- Sheppard AJ, Pennington JA, Weihrauch JL (1993) Analysis and distribution of Vitamin E in vegetable oils and foods. In: Packer L, Fuchs J (eds) *Vitamin E in health and disease*. Marcel Dekker, New York, pp 9–31
- Solovyyev VV (2001) Statistical approaches in eukaryotic gene prediction. In: Balding D et al (eds) *Handbook of statistical genetics*. Wiley, New York, pp 83–127
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL  $\times$  windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Traber MG, Sies H (1996) Vitamin E and humans: Demand and delivery. *Annu Rev Nutr* 16:321–347
- Valentin HE, Venkatesh TV, Karunanandaa B (2003) Homogentisate prenyl transferase (“HPT”) nucleic acids and polypeptides and uses thereof. International patent application WO 03080647
- Valvekens D, Van Montagu M, Van Lijsebettens M (1988) *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. *Proc Natl Acad Sci USA* 85:5536–5540
- Zhang MQ (2002) Computational prediction of eukaryotic protein-coding genes. *Nat Rev Genet* 3:698–709