PNAS RESEARCH ARTICLE | PLANT BIOLOGY

Genome-wide association identifies a missing hydrolase for tocopherol synthesis in plants

Elise Albert[a], Sungsoo Kim[a], Maria Magallanes-Lundback[a], Yan Bao[a], Nicholas Deason[a], Benoit Danilo[a], Di Wu[b], Xiaowei Li[b], Joshua C. Wood[c], Nolan Bornowski[c], Michael A. Gore[b], C. Robin Buell[c], and Dean DellaPenna[a,1]

Contributed by Dean DellaPenna; received July 24, 2021; accepted April 25, 2022; reviewed by Bernhard Grimm and Harry Klee

The tocopherol biosynthetic pathway, encoded by *VTE* genes 1 through 6, is highly conserved in plants but most large effect quantitative trait loci for seed total tocopherols (totalT) lack *VTE* genes, indicating other activities are involved. A genome-wide association study of *Arabidopsis* seed tocopherols showed five of seven significant intervals lacked *VTE* genes, including the most significant, which mapped to an uncharacterized, seed-specific, envelope-localized, alpha/beta hydrolase with esterase activity, designated *AtVTE7*. *Atvte7* null mutants decreased seed totalT 55% while a leaky allele of the maize ortholog, *ZmVTE7*, decreased kernel and leaf totalT 38% and 49%, respectively. Overexpressing *AtVTE7* or *ZmVTE7* partially or fully complemented the *Atvte7* seed phenotype and increased leaf totalT by 3.6- and 6.9-fold, respectively. VTE7 has the characteristics of an esterase postulated to provide phytol from chlorophyll degradation for tocopherol synthesis, but bulk chlorophyll levels were unaffected in *vte7* mutants and overexpressing lines. Instead, levels of specific chlorophyll biosynthetic intermediates containing partially reduced side chains were impacted and strongly correlated with totalT. These intermediates are generated by a membrane-associated biosynthetic complex containing protochlorophyllide reductase, chlorophyll synthase, geranylgeranyl reductase (GGR) and light harvesting-like 3 protein, all of which are required for both chlorophyll and tocopherol biosynthesis. We propose a model where VTE7 releases prenyl alcohols from chlorophyll biosynthetic intermediates, which are then converted to the corresponding diphosphates for tocopherol biosynthesis.

seed | vitamin E | chlorophyll | GWAS | hydrolase

Significance

Tocopherols (vitamin E) are plant-synthesized, lipid-soluble antioxidants whose dietary intake, primarily from seed oils, is essential for human health. Tocopherols contain a phytol-derived hydrophobic tail whose in vivo source has been elusive. The most significant genome-wide association signal for *Arabidopsis* seed tocopherols identified an uncharacterized, seed-specific esterase (VTE7) localized to the chloroplast envelope, where tocopherol synthesis occurs. *VTE7* disruption and overexpression had large impacts on tissue tocopherol contents with metabolic phenotypes consistent with release of prenyl alcohols, including phytol, during chlorophyll synthesis, rather than from the bulk degradation of thylakoid chlorophylls as has long been assumed. Understanding the source of phytol for tocopherols will enable breeding and engineering plants for vitamin E biofortification and enhanced stress resilience.

Tocopherols are a class of lipid-soluble, plastid-synthesized antioxidants present in all plant tissues, but most abundant in seeds, where they are essential for protecting membrane lipids, especially during seed desiccation, storage, and germination (1–3). Tocopherols consist of a polar chromanol head group, derived from homogentisic acid (HGA), and a lipophilic side chain derived from phytyl-diphosphate (PDP). The four tocopherols produced by plants (δ-, β-, γ- and α-) differ only in the numbers and positions of methyl groups on their chromanol head groups. In humans, tocopherols are an important dietary antioxidant and essential nutrient (vitamin E) with positive impacts on immune function, cardiovascular disease, and age-related macular degeneration in at-risk populations (4).

Tocopherol biosynthesis has been most extensively investigated in *Arabidopsis thaliana* where genes for synthesis of HGA, PDP, and the core tocochromanol pathway itself (*VTE1* through *VTE6*) were initially identified (Fig. 1*A*) (2, 3). The committed step in tocopherol biosynthesis is condensation of HGA and PDP by homogentisate phytyltransferase, VTE2. Two methyltransferases and the tocopherol cyclase (VTE3, VTE4 and VTE1, respectively) then act in different orders and combinations to generate δ-, β-, γ- and α-tocopherols. *VTE* orthologs are readily identified in crop genomes, which has enabled efforts to engineer the pathway (2, 3). Altering methyltransferase expression changes tocopherol composition in leaves and seeds without affecting total tocopherol (totalT) levels, whose engineering has been more challenging. Of the six *VTE* genes, only *VTE2* overexpression had a substantial impact on totalT content of leaves and seeds, while overexpression of other *VTE* genes, singly or in combination with *VTE2*, had minimal impact. These results are consistent with VTE2 activity and the provision of its substrates, HGA and PDP, being key control points for totalT synthesis.

HGA is an intermediate in aromatic amino acid metabolism, a plant pathway tightly regulated by feedback inhibition, and overexpressing pathway enzymes from plants had limited impact on tocopherol content (2, 3). It was only by introducing feedback-insensitive bacterial or fungal enzymes that substantial increases in tocochromanol content were achieved, although often accompanied by massive accumulation of aromatic

Author contributions: E.A., J.W., and D.D.P. designed research; E.A., S.K., M.M.-L., Y.B., N.D., B.D., D.W., and X.L. performed research; E.A., S.K., and J.C.W. contributed new reagents/analytic tools; E.A., S.K., M.M.-L., Y.B., N.D., B.D., D.W., X.L., N.B., M.A.G., C.R.B., and D.D.P. analyzed data; E.A. and D.D.P. wrote the paper; C.R.B. oversaw informatics work; and D.D.P. conceptualized the project.

Reviewers: B.G., Humboldt-Universitat zu Berlin; and H.J.K., University of Florida.

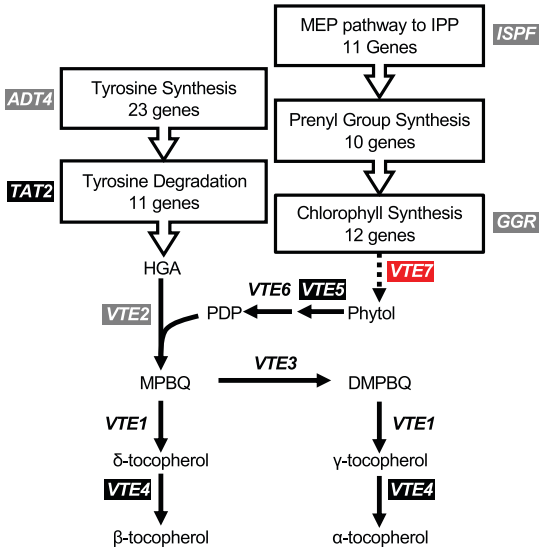
The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: dellapen@msu.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2113488119/-/DCSupplemental.

Published May 31, 2022.

A



B

Chr	# of significant SNPs	Position peak SNP (bp)	Peak SNP *P-value* after FDR correction	Traits	Gene at peak SNP	*A priori* candidate genes +/- 40 kb from peak SNP	*A Priori* Candidate Pathway
Chr1	*2**	*23,708,007*	*2.89E-02**	*γT, totalT*	*AT1G63870*	*ISPF* (AT1G63970)	*IPP Synthesis*
Chr1	3 (40*)	24,131,921	2.44E-05	αT	AT1G64960	*VTE4* (AT1G64970)	Tocochromanol
Chr1	*2**	*27,962,976*	*4.70E-02**	*αT*	*AT1G74390*	*GGR* (AT1G74470)	*Chlorophyll Synthesis*
Chr2	*1**	*8,170,738*	*4.82E-02**	*αT*	*AT2G18876*	*VTE2* (AT2G18950)	*Tocochromanol*
Chr3	1	7,657,301	4.04E-02	totalT	intergenic		
Chr3	*6**	*16,259,409*	*3.62E-02**	*αT, totalT*	*intergenic*	*ADT4* (AT3G44720)	*Tyrosine Synthesis*
Chr4	1	414,024	2.48E-02	γT, totalT	intergenic		
Chr4	1	7,460,123	4.36E-02	γT	AT4G12620		
Chr5	3 (5*)	1,281,561	5.86E-03	γT, totalT	AT5G04490	*VTE5* (AT5G04490)	Tocochromanol
Chr5	4 (5*)	14,265,216	3.37E-03	γT, totalT	intergenic	*TAT2* (AT5G36160)	Tyrosine Degradation
Chr5	104	15,706,147	4.68E-06	γT, totalT	AT5G39220	*VTE7* (AT5G39220)	

* Pathway-level significance only.

Fig. 1. Summary of genome-wide and pathway-level association results for seed tocopherol traits. (*A*) Summarized pathways leading to tocopherol synthesis in *Arabidopsis thaliana* with key genes mentioned in text indicated and (*B*) table summarizing association results. White letters with black and red backgrounds indicate genome-wide significance, while gray background indicates significance only in pathway-level analysis. *VTE7* is indicated by red highlighting. An FDR threshold of 5% was applied to declare significant SNPs in genome-wide and pathway-level analyses. Genomic information for significant SNPs is presented in *SI Appendix*, Fig. S2 and Datasets S2 and S4. Compound abbreviations: IPP, isopentenyl diphosphate; PDP, phytyl-diphosphate; HGA, homogentisic acid; MPBQ, 2-methyl-6-phytyl-1,4-benzoquinol; DMPBQ, 2,3-dimethyl-6-phytyl-1,4-benzoquinol; MEP, 2-C-methyl-D-erythritol 4-phosphate. Gene abbreviations: *VTE1*, tocopherol cyclase; *VTE2*, homogentisate phytyltransferase; *VTE3*, MPBQ methyltransferase; *VTE4*, γ-tocopherol methyltransferase; *VTE5*, phytol kinase; *VTE6*, phytylphosphate kinase; *TAT2*, tyrosine aminotransferase 2; *GGR*, geranylgeranyl reductase; *ISPF*, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase.

pathway intermediates that can have deleterious effects (2, 3). Geranylgeranyl reductase (GGR) can produce the other VTE2 substrate, PDP, by two possible routes: by direct reduction of geranylgeranyl-diphosphate (GGDP) to PDP (5), or indirectly by reducing geranylgeranyl-chlorophyll to (phytyl)-chlorophyll followed by phytol hydrolysis and its sequential phosphorylation to PDP by the VTE5 and VTE6 kinases (6). *vte5* null mutants reduced seed and leaf tocopherols by 80% and 65%, respectively, and in a *vte5 vte6* double mutant leaf tocopherols are below detection. This indicates the VTE5/VTE6-dependent route provides the bulk of PDP for tocopherols and that any PDP made directly from GGDP is not accessible by VTE2 (6). The enzyme(s) that release phytol from chlorophyll are therefore central to tocopherol synthesis but attempts to identify them have been unsuccessful (2, 3, 7, 8).

Linkage and association mapping studies of seed tocopherols has provided important insights into the genetic control of tocopherol synthesis that are generally consistent with transgenic studies. Major effect quantitative trait loci (QTL) for seed tocopherol composition often map to intervals or loci encoding the two pathway methyltransferases, *VTE3* and *VTE4* (2, 9–12). While some QTL intervals for seed totalT contain *VTE2* or *VTE5*, the largest effect QTL intervals typically lack *VTE* genes. In most studies, constraints in panel size, marker diversity and mapping resolution precluded identification of the underlying, novel causal genes. A notable exception was (12), who used the highly powered maize nested association mapping (NAM) panel to identify 18 QTL for kernel totalT with 16 being previously unidentified loci that lacked 81 a priori candidate genes in their intervals. The two unidentified QTL with the largest effects mapped to protochlorophyllide reductase (POR) homologs (12), a key regulated step in chlorophyll synthesis, and when one locus (*POR1*) was overexpressed in kernels it increased totalT levels 19% (13). That PORs are the major determinants of kernel tocopherol content, despite maize kernels being nonphotosynthetic with chlorophyll levels 1,000 to 10,000 times lower than leaves, confirms the dependence of tocopherol synthesis on chlorophyll metabolism (12).

In this study, we investigated natural variation in tocopherol composition and content in *Arabidopsis* seeds which, unlike maize kernels, are green and photosynthetic during development with chlorophyll levels similar to leaves. Using an 814-member association panel (14) we identified 7 significant genome-wide associations, 5 of which lacked *VTE* genes. The strongest association signal was for totalT and resolved to AT5G39220 (hereafter named *AtVTE7*), an envelope-localized, alpha/beta hydrolase (ABH) that is highly expressed across seed development and present in all plants. The maize ortholog

(*ZmVTE7*) localizes within the support interval of an unresolved totalT QTL with the third largest totalT effect size in the NAM panel (12). Knockout and overexpression data are consistent with *VTE7* encoding a hydrolase that provides most of the phytol from chlorophyll for tocopherol biosynthesis in seeds.

Results

Genetic Determinants of Tocopherol Content in *Arabidopsis thaliana* Seeds. We assessed natural variation for dry seed tocopherol content and composition in 814 diverse *A. thaliana* accessions whose genomes have been resequenced as part of the 1001 Genomes Project (14). Samples from two outgrowths were quantified for alpha-, gamma-, delta-, and total tocopherols (αT, γT, δT and totalT, respectively) and these data were used to calculate best linear unbiased predictors (BLUPs) for each trait (Dataset S1). Large natural variation for tocopherol content was observed (CV from 11 to 21%) with γT representing on average 95% of total seed tocopherols. Broad sense heritabilities ranged from 0.72 to 0.92, indicative of a high amount of variance explained by genetic effects (*SI Appendix*, Fig. S1). We ran genome-wide association (GWA) analysis for the four tocopherol BLUPs using 1,802,374 single-nucleotide polymorphisms (SNPs) and a mixed linear model. After applying the Benjamini-Hochberg correction to control for a genome-wide false discovery rate (FDR) of 5%, 122 SNPs were significantly associated with at least one trait (Dataset S2 and *SI Appendix*, Fig. S2).

The 122 SNPs identified seven genomic regions associated with αT, γT or totalT (Fig. 1*B*). Because linkage disequilibrium (LD) decays to $r^2 < 0.2$ within 10 kb in the panel, but with large variation depending on genomic region (*SI Appendix*, Fig. S3) (15), we limited our candidate gene search space to ±40 kb from significant SNPs. Within these constraints, three of the seven regions contained a tocopherol a priori candidate gene: *VTE4* (γ-tocopherol methyltransferase), associated with αT, and *VTE5* (phytol kinase) and *TAT2* (tyrosine aminotransferase 2), both associated with γT and totalT (Fig. 1*B* and Dataset S2). TAT2 produces *p*-hydroxyphenylpyruvate, which is converted to HGA, the aromatic head group for all tocopherols. The remaining four genomic regions represent previously unidentified loci that impact seed tocopherols, with that at 15.71 Mb on chromosome 5 for γT and totalT being the most significant in this study (FDR $P = 4.68E-06$) (Fig. 1*B*). To examine potential impacts of associated SNPs those passing FDR were assessed with the functional effect prediction tool SnpEff (16). In two cases, the most significant SNP association conditioned an amino acid change: Asp596Val for VTE4 and Ala297Thr for VTE5 (Dataset S2).

Tocopherol synthesis uses substrates derived from the methylerythritolphosphate (MEP) and aromatic amino acid pathways (Fig. 1*A*) which, along with *VTE* genes and components of chlorophyll and prenyldiphosphate metabolism totals 73 a priori candidate genes (Dataset S3) that could impact tocopherol synthesis (11, 12). We assessed these a priori genes for more modest associations that did not reach genome-wide significance. A total of 74,399 SNPs located ±40 kb of these genes was used to compute FDR adjustment for each trait, which identified 61 unique significant SNPs (Fig. 1*B*, Dataset S4 and *SI Appendix*, Fig. S2). Not surprisingly, most were within ±40 kb of *VTE4*, *VTE5* and *TAT2*, but 11 SNPs were in the vicinity of four additional a priori genes: *VTE2* and *GGR* for αT, the MEP pathway enzyme 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*ISPF*) for γT and totalT and the aromatic amino acid enzyme arogenate dehydratase 4 (*ADT4*) for αT and totalT (Fig. 1*B*).

***VTE7* Is a Major Regulator of Tocopherol Content in Seeds.** The largest GWA signal for γT and totalT was a 59.1 kb chromosome 5 interval without a priori candidate genes for which the most significant of 104 SNPs mapped to the first intron of AT5G39220 (Figs. 1*B* and 2*A*). To better refine the association signal, LD was computed across the 40-kb region centered on the peak SNP (Fig. 2*A*), delineating a 7,327 bp region encompassing two genes: AT5G39220 and AT5G39210. AT5G39210 encodes a component of the chloroplast NAD(P)H dehydrogenase complex (AtCCR7) that is weakly expressed in developing seeds and for which two previously characterized null alleles were available (*Atcrr7-1* and *Atcrr7-3*) (17) (*SI Appendix*, Fig. S4). AT5G39220 (hereafter named *AtVTE7*), encodes an uncharacterized ABH that is strongly expressed across seed development but not in leaves (*SI Appendix*, Fig. S4) and predicted by iPSORT to be plastid-targeted (18). Four *AtVTE7* T-DNA insertions alleles were isolated and RT-qPCR of 1 cm long seed containing green siliques demonstrated *Atvte7-1* and *Atvte7-5* were null alleles while *Atvte7-3* *Atvte7-4* were leaky (*SI Appendix*, Fig. S5). Growth of *Atvte7* and *Atcrr7* mutants was indistinguishable from WT, as was the seed totalT content of both *Atcrr7* alleles (Fig. 2*B*). Seed totalT in *Atvte7* null alleles was reduced 54% relative to WT (Fig. 2*B*), while leaf totalT was not affected (*SI Appendix*, Fig. S5), indicating *AtVTE7* encodes a seed specific ABH required for tocopherol synthesis.

The *Arabidopsis* ABH gene family contains more than 600 members including AT1G13820, which has 76% protein similarity to *AtVTE7* (*SI Appendix*, Fig. S6) and is also predicted by iPSORT to be plastid-targeted (18). AT1G13820 is strongly expressed in vegetative tissues and at low levels in seeds (*SI Appendix*, Fig. S4) raising the possibility that it provides VTE7 activity in leaves and/or is partially functionally redundant with *AtVTE7* in seeds. An AT1G13820-1 knockout allele was isolated and assessed singly and in combination with *Atvte7* null alleles. AT1G13820 mRNA is undetectable in AT1G13820-1 but seed and leaf tocopherol contents were not significantly different from WT (*SI Appendix*, Fig. S7). Similarly, seed tocopherols in AT1G13820-1 *Atvte7* double mutants were not significantly different from *Atvte7* null alleles, indicating the two proteins have distinct activities and are not functionally redundant (*SI Appendix*, Fig. S7).

VTE7 orthologs are present at low copy number in all sequenced angiosperm genomes (*SI Appendix*, Fig. S6). The maize homolog GRMZM5G898684 (hereafter named *ZmVTE7*) is particularly relevant as it is in the interval of the third largest effect kernel totalT QTL in the maize NAM panel, which could not be resolved to a causal gene (QTL12 in (12)). *ZmVTE7* is strongly expressed across seed development and highest in develop\ing embryos, the predominant site of tocopherol accumulation in maize kernels (*SI Appendix*, Fig. S4). *ZmVTE7* is also expressed in vegetative tissues suggesting that it may also contribute to leaf tocopherol synthesis. A homozygous, leaky Mu insertion 56 bp upstream of the *ZmVTE7* start codon (hereafter named *Zmvte7-1*) (19) reduced *ZmVTE7* RNA levels 80% relative to WT, in developing kernels and leaves (*SI Appendix*, Fig. S8). The totalT content of mature *Zmvte7-1* kernels was decreased by 30% relative to WT, while tocotrienols were unchanged, indicating the role of *VTE7* in seed tocopherol synthesis is functionally conserved between *Arabidopsis thaliana* and *Zea mays* (Fig. 2 *C* and *D*). Leaf totalT in *Zmvte7-1* was reduced

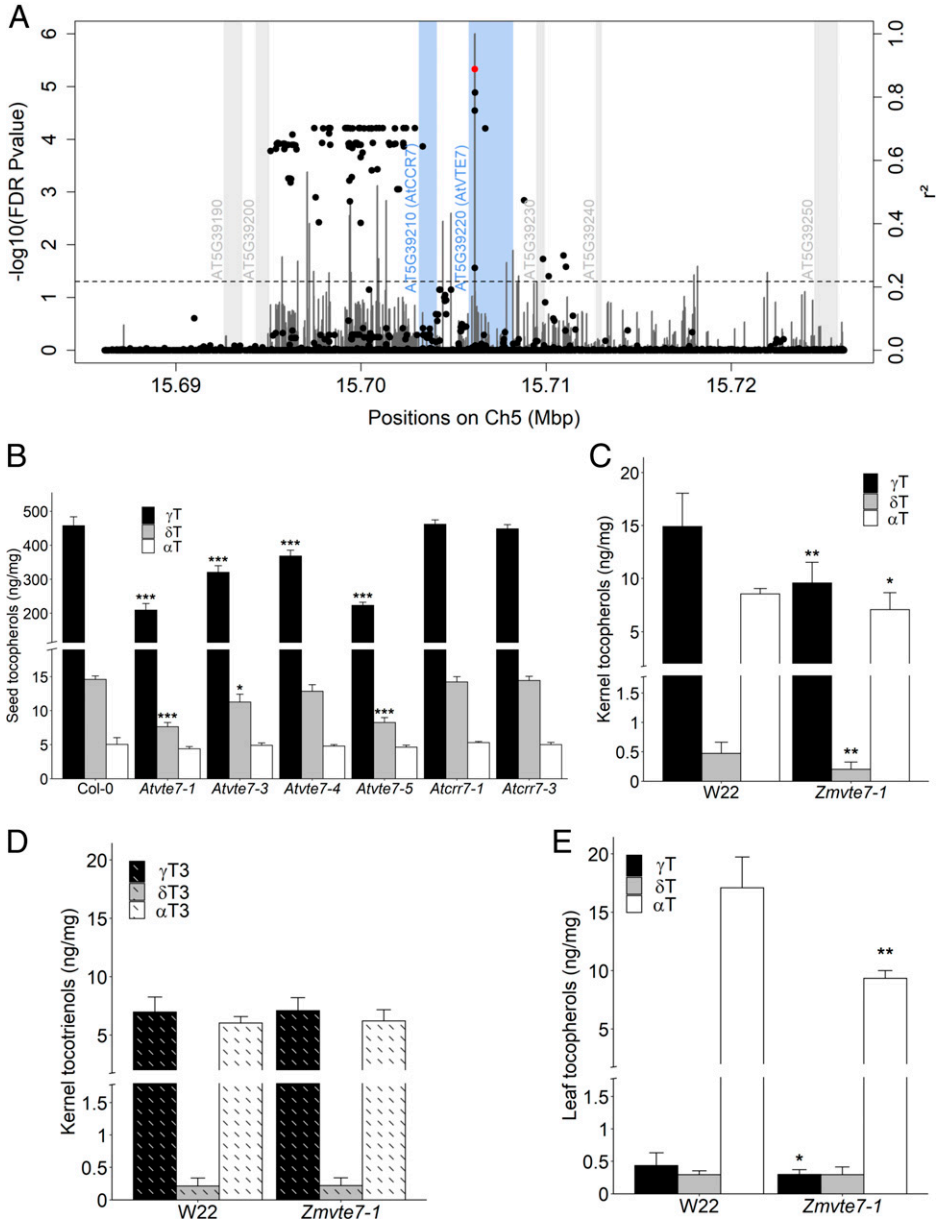


Fig. 2. Association mapping identified *AtVTE7* (AT5G39220) as the major contributor to seed total tocopherol content. (*A*) Manhattan plot for totalT across a 40 kb interval centered on the most significant SNP identified in this study (red filled circle). Circles indicate -log10 FDR corrected *P values* of SNPs plotted against their chromosome 5 positions. The black dashed line indicates the α = 0.05 genome-wide FDR significance threshold. Vertical lines indicate linkage disequilibrium relative to the peak SNP (r^2, right *y*-axis) and shaded vertical bars genes. (*B*) Seed tocopherol content in homozygous alleles of *Atvte7* (AT5G39220) and *Atcrr7* (AT5G39210) mutants and Col-0 wild type. Means and SDs are from four biological replicates. (*C*) Tocopherol and (*D*) tocotrienol content in dry maize kernels and (*E*) tocopherol content in fully expanded fifth leaves of wild-type, W22 and the *Zmvte7-1* mutant. Means and SDs are from eight biological replicates. *p*-values of mutant means significantly different from wild type in Dunnett *t* tests are indicated with asterisks: *< 0.05; **< 0.01; ***< 0.001. Note that *y*-axes in *B-E* are discontinuous.

45% relative to WT, indicating *ZmVTE7* also contributes to tocopherol synthesis in maize leaves (Fig. 2*E*).

We examined variation in gene sequence, gene expression, and structural variation at the *ZmVTE7* locus in the 26 diverse NAM founders. As expected, some structural variation was identified between the NAM founders but overall, none was associated with tocopherol effect estimates in the population (*SI Appendix*, Fig. S9). However, 26 of the NAM founders had totalT QTL at the *ZmVTE7* locus and in all cases the recurring B73 parent was the superior allele (12). In B73 RefGen_v2 and _v3, *ZmVTE7* was incorrectly annotated as a single 26.5 kb gene that was corrected in B73 RefGen_v4 to two, tandem repeats of 4.6kb *ZmVTE7* loci with only 3 bp difference (Zm00001d006778 and Zm00001d006779). B73 is unique in having two tandem copies of *VTE7* and there was a modest positive correlation between *ZmVTE7* expression and tocopherol content across kernel development in the NAM founders (*SI Appendix*, Fig. S10), making it likely that *VTE7* copy number variation in B73 contributes to the superior nature of its allele.

***AtVTE7* Localizes to the Plastid Envelope.** Both *AtVTE7* and *ZmVTE7* were predicted to be chloroplast targeted and to directly assess their localization, C-terminal YFP fusions driven by the 35S promoter were introduced into WT and *Atvte7-1* backgrounds (to assess functional complementation of the mutant). In cotyledons of high expressing, stably transformed plants, the YFP signal of a 35S:YFP control was present in both

the cytoplasm and nucleus while 35S:AtVTE7:YFP consistently colocalized with chlorophyll fluorescence, indicating chloroplast localization (*SI Appendix*, Fig. S11). For 35S:ZmVTE7:YFP, the YFP signal in lines with low expression was restricted to chloroplasts but at high expression levels, strong signal was also present in the cytosol. Because chloroplast targeted proteins from monocots can have reduced targeting efficiency in dicots, especially when expressed at high levels, we deleted the predicted 16 amino acid (monocot) *ZmVTE7* plastid transit peptide and replaced it with the well-characterized *Arabidopsis* (dicot) RecA chloroplast transit peptide (cTP) (20). Lines stably overexpressing the 35S:cTP-ZmVTE7:YFP fusion were faithfully targeted to chloroplasts, even at high expression levels (*SI Appendix*, Fig. S11). Because resolution is inherently limited when imaging whole plant tissues, we also assessed 35S:AtVTE7:YFP localization by transient expression in isolated *Arabidopsis* leaf protoplasts. While 35S:AtVTE7:YFP was still associated with chloroplasts it clearly did not colocalize with (thylakoid) chlorophyll fluorescence but was present in halos surrounding the chloroplast, consistent with chloroplast envelope localization (*SI Appendix*, Fig. S11).

Characterization of *AtVTE7* and *ZmVTE7* Overexpression Lines. A total of 71 independent transgenic *Atvte7-1* lines were isolated that constitutively overexpressed native or YFP-tagged versions of *AtVTE7* or *cTP-ZmVTE7* cDNAs. Pooled T_2 seeds of each line were analyzed for tocopherol content as, although still segregating for the transgene, this would provide a rank order for relative complementation of the *Atvte7-1* seed phenotype. All lines complemented to varying degrees and importantly, there were no significant differences between native and YFP-tagged versions, allowing subsequent analyses to focus on the YFP-tagged versions used to demonstrate localization (*SI Appendix*, Fig. S12 and Dataset S5). The whole plant phenotypes of native and YFP-tagged *AtVTE7* complementation lines were indistinguishable from WT and *Atvte7-1*, while for both native and YFP-tagged *cTP-ZmVTE7* lines approximately half were indistinguishable, and half displayed varying degrees of anomalous leaf expansion (*SI Appendix*, Fig. S13).

Three independent, homozygous, single transgene insertion lines for *Atvte7-1(35S:AtVTE7:YFP)* and *Atvte7-1(35S:cTP-ZmVTE7:YFP)* were selected and used for subsequent analyses. *AtVTE7:YFP* overexpression resulted in a range of complementation with the highest restoring a WT tocopherol seed phenotype, while *cTP-ZmVTE7:YFP* overexpression partially complemented to a maximum of 78% of WT (Fig. 3*A*). *AtVTE7* is not expressed in leaves and its ectopic overexpression increased leaf totalT up to 3.6-fold relative to WT and though all tocopherols increased, the major contributor, increasing more than 20-fold, was γT (Fig. 3*B*). Ectopic overexpression of *ZmVTE7:YFP* in leaves had an even stronger impact, increasing totalT as much as 6.9-fold in line 35, with γT again being the major contributor (Fig. 3*B*). Because line 35 displayed an anomalous leaf phenotype, we assessed two additional homozygous *ZmVTE7:YFP* lines with similar leaf anomalies, lines 34 and 27, both of which also had leaf totalT levels similar to line 35 (*SI Appendix*, Fig. S13 and Dataset S6).

***VTE7* and *VTE2* Overexpression Are Additive in Leaves but Not Seeds.** *VTE2* encodes homogentisate phytyl transferase (HPT, Fig. 1*A*), the committed step for tocopherol synthesis whose constitutive overexpression increased totalT up to 4.4-fold and 40% in *Arabidopsis* leaves and seeds, respectively (21). To test whether co-overexpression of *AtVTE2* and *AtVTE7* might be additive, homozygous *35S:AtVTE7:YFP* lines 74 and 6 were crossed with *35S:AtVTE2_line 11* (21) and tocopherols analyzed in seeds and leaves of the resulting F_1 hybrids. The totalT content of *35S:AtVTE7* lines 6 and 74 leaves were 3.2- and 3.4-fold higher than WT, respectively, while *35S:AtVTE2_line 11* was 4.1-fold higher (Fig. 3*B*). F_1 leaves overexpressing both transgenes were 6.9 and 7.4-fold higher than WT, due predominantly to a > 40-fold increase in γT, indicating the two transgenes are additive in leaves (Fig. 3*B*). In contrast, F_1 seed totalT was not significantly different from the *35S:AtVTE2_line 11* parent, indicating overexpressing the two transgenes is not additive in seeds (Fig. 3*A*).

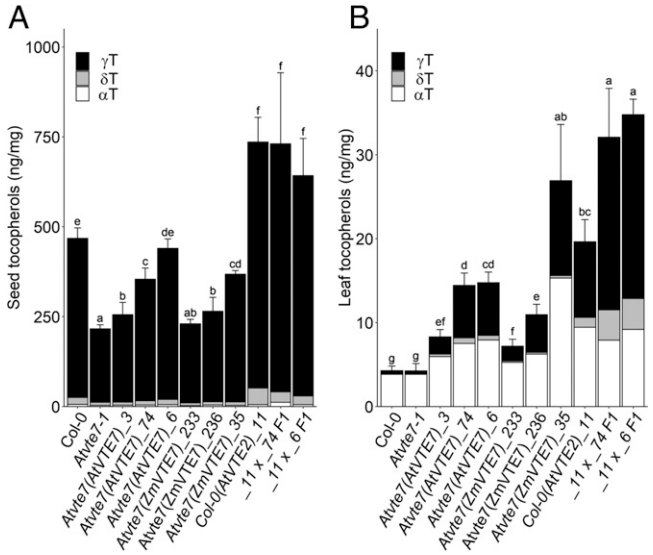


Fig. 3. Seed (*A*) and leaf (*B*) tocopherols in Col-0 wild-type *AtVTE7*, *ZmVTE7*, *AtVTE2* single, and *AtVTE7 AtVTE2* double overexpressing lines. Tissues for single overexpression lines were from homozygous lines and for double overexpression lines were from F_1 seeds or plants from crosses of the indicated single overexpression lines. Each line is represented as an average from four biological replicates which for leaves is five, pooled 4-wk-old rosettes and for seeds is from individual plants. TotalT means significantly different from each other in Tukey's Honest Significant Difference test are indicated with letters ($P < 0.05$).

In Vitro and In Vivo Analysis of *VTE7* Activity and Function. Based on the large effects that overexpressing and mutating *VTE7* has on tocopherol content, we hypothesized VTE7 hydrolyzes esterified phytol from bulk chlorophyll for tocopherol synthesis. To test this in vitro, *AtVTE7* and *ZmVTE7* were expressed as His-tagged protein fusions in *Escherichia coli*, purified and assayed for their ability to hydrolyze phytol from chlorophylls, pheophytins and fatty acid phytyl esters, as well as synthetic 4-nitrophenol (NP) substrates with esterified acyl chains ranging from C_4 to C_{18}. Conditions for optimal activity were determined against the NP-C_4 synthetic substrate and both AtVTE7 and ZmVTE7 showed strong activity that sharply decreased, but was still detectable, as acyl chain length increased (Dataset S7). However, neither AtVTE7 or ZmVTE7 showed in vitro activity against known phytyl ester compounds present in plants: chlorophylls a or b, pheophytins a or b, fatty acid phytyl esters and tocopherol fatty acid esters. Because we could not demonstrate in vitro activity of VTE7 against these substrates, we performed comparative analyses of metabolites in WT, transgenic and mutant tissues to gain insight into the activity of VTE7 in vivo.

We first analyzed developing seeds and leaves of WT, *Atvte7-1* and *VTE7* overexpression lines for tocopherols and the major

chlorophyll metabolites: chlorophylls a and b, chlorophyllides a and b (chlorophylls lacking phytol tails) and pheophytins a and b (chlorophylls lacking Mg^{2+}). Seeds were assessed at four developmental stages: linear embryo (9 d after pollination [DAP]), mature green embryo (14 DAP), yellowing (18 DAP) and brown seeds (24 DAP). In WT, totalT increased 34.2-fold between 9 and 24 DAP, with the largest increase occurring between 14 and 18 DAP. TotalT levels in *Atvte7-1* was significantly lower and approximately half of WT at all stages (Fig. 4*A*). TotalT levels in both *AtVTE7* and *ZmVTE7* transgenic seeds were significantly different from *Atvte7-1* at 14 DAP, indicating both enzymes complemented the *Atvte7-1* tocopherol phenotype during early seedling development, however, thereafter they diverged. *35S:AtVTE7-YFP_6* fully complemented throughout development while *35S:AtVTE7-YFP_74* remained significantly different from *Atvte7-1,* it trended lower than WT at 18 DAP and was significantly so at 24 DAP. In contrast, at 18 and 24 DAP, totalT levels in *ZmVTE7* overexpression lines were no longer significantly different from *Atvte7-1* (Fig. 4*A*).

Developing *Arabidopsis* embryos are green until the mature embryo stage after which time chlorophyll degradation leads to extremely low levels in dry seeds. Total chlorophyll in WT seeds was high at 9 DAP, increased further at 14 DAP and rapidly declined by 24 DAP (*SI Appendix*, Fig. S14*A*). Pheophytin a and chlorophyllide a followed a similar pattern but at levels < 1% that of total WT chlorophyll at all timepoints (*SI Appendix*, Fig. S14 *B* and *C*). With few exceptions, chlorophylls a and b, pheophytin a and chlorophyllide a in developing *Atvte7-1* and transgenic seeds were also not significantly different from WT. Similarly, although leaf tocopherol levels were strongly elevated by *AtVTE7* or *ZmVTE7* overexpression (Fig. 3*B*), chlorophyll metabolites in transgenics were also not significantly different from WT (Datasets S6 and S8). These combined data were unexpected as we had anticipated that if bulk chlorophyll were the source of phytol for tocopherols that one or more major seed chlorophyll metabolites would be significantly impacted by the absence or overexpression of *VTE7*.

To gain insight into the in vivo consequences of altering VTE7 levels, we performed RNAseq analysis and untargeted LC-MS analysis of metabolites isolated from WT and *Atvte7-1* seed at 9, 14 and 18 DAP. Expression levels for the 73 a priori genes listed in Dataset S3, *VTE7,* and 18 additional genes involved in chlorophyll synthesis, degradation and phytol metabolism were compared in Col-0 and *Atvte7-1* seeds at each timepoint. Except for *VTE7*, which was 100-fold lower ($P <$ 6.3E-13) in *Atvte7-1* at 9 and 14 DAP, only one gene in aromatic amino acid metabolism, shikimate kinase 2, had moderately lower expression ($P <$ 2.5E-2) in *Atvte7-1* at 18 DAP (Dataset S9). These data are consistent with *Atvte7-1* only affecting *VTE7* expression and the decreased tocopherols in *Atvte7-1* not resulting from altered expression of other *VTE*, aromatic amino acid biosynthesis or chlorophyll metabolism genes.

For untargeted LC-MS analysis, data were first filtered for relative mass defects that enrich for lipophilic molecules (22)

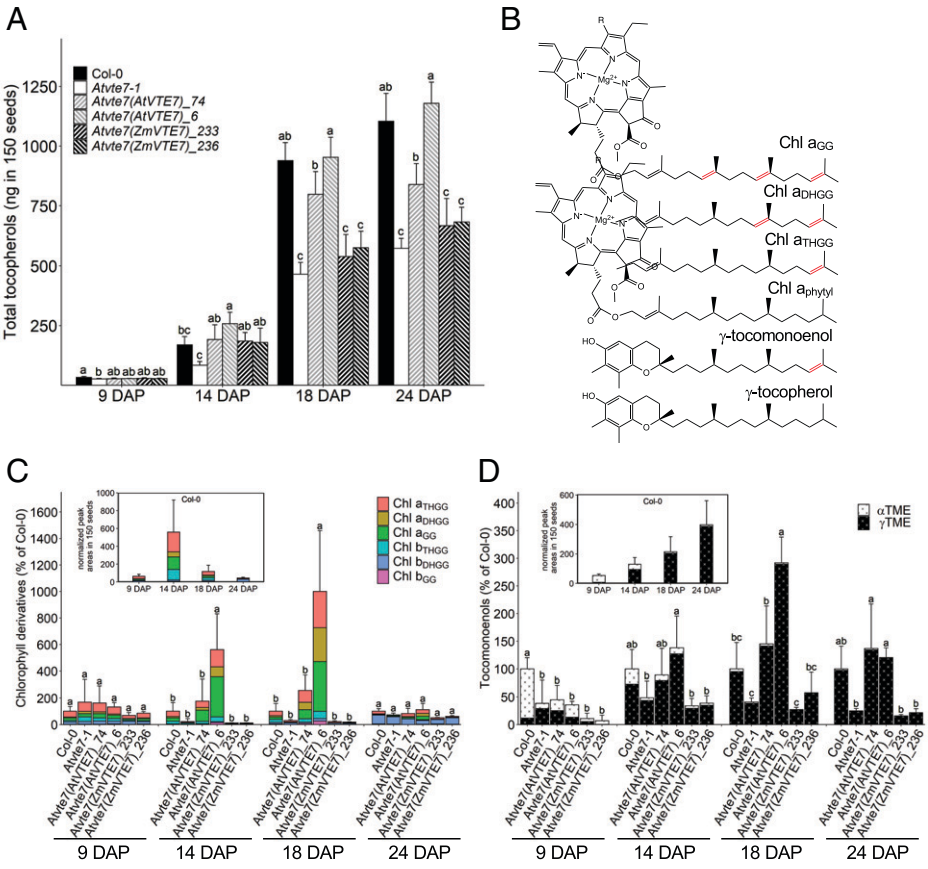


Fig. 4. Total tocopherols (*A*), compound structures (*B*), chlorophyll derivatives (*C*), and tocomonoenols (*D*) in developing seeds of Col-0, *Atvte7-1* and lines overexpressing *AtVTE7* or *ZmVTE7* in *Atvte7-1* mutant. Seeds were analyzed at four stages: linear (9 d after pollination, DAP), mature green (14 DAP), yellowing (18 DAP) and brown (24 DAP). Means and SDs are from four biological replicates of 150 seeds taken from individual plants. Tocopherols are expressed in ng per 150 seeds. Tocomonoenols and chlorophyll derivatives were determined by LC-MS and normalized to recovery of an internal standard (telmesartin). To accommodate the large dynamic range across development (see Col-0 insert plots in *C* and *D*), genotypes in *C* and *D* main graphs are expressed as percentage Col-0 levels at each timepoint. Means significantly different from each other in Tukey's Honest Significant Difference test are indicated with letters at each stage ($P < 0.05$).

resulting in 3,566 mass/time peaks identified as being present in at least one sample. Only eleven of these mass/time peaks passed a statistical cutoff of > 2-fold change and $P < 0.05$ between genotypes (Dataset S10), of which, 10 were significantly lower in *Atvte7-1*. As anticipated, γT was one significantly lower compound but surprisingly, four others had fragmentation patterns with exact masses corresponding to chlorophyll rings, three to chlorophyll a and one to chlorophyll b. However, their intact molecular masses were 2, 4, or 6 atomic mass units (AMUs) less than (phytylated) chlorophyll a or b, indicating the presence of up to three additional double bonds in their hydrophobic tails. These compounds correspond to known chlorophyll biosynthetic intermediates, chlorophyll rings esterified to geranylgeraniol (e.g., Chl a_{GG}; -6 AMU), dihdyrogeranylgeraniol (e.g., Chl a_{DHGG}; -4 AMU) or tetrahydrogeranylgeraniol (e.g., Chl a_{THGG}; -2 AMU) (Fig. 4*B*). These data demonstrate that although bulk chlorophylls and chlorophyllide a and pheophytin a were not affected by loss of AtVTE7 activity in *Atvte7-1* seeds, chlorophyll a and b biosynthetic intermediates were.

Based on these findings, we performed targeted liquid chromatography–mass spectrometry (LC-MS) analysis of developing seeds from WT, *Atvte7-1* and four overexpression lines to assess levels of chlorophyll biosynthetic intermediates and α- and γ-tocomonoenols, minor tocopherol species produced by the VTE2-dependent incorporation of THGG-DP instead of phytyl-DP during synthesis (23). In developing WT seeds the sum of partially reduced chlorophyll intermediates was highly dynamic (Fig. 4*C*, *Inset*), increasing 10-fold between 9 and 14 DAP then rapidly decreasing by 18 DAP, the timeframe of maximal tocopherol increase (Fig. 4*A*). Chl a species, particularly Chl a_{THGG}, were most abundant at all timepoints except 24 DAP and tocomonoenols also increased nearly 10-fold during WT seed development (inset, Fig. 4*D*), paralleling the increase in tocopherols. In *Atvte7-1*, chlorophyll intermediates were similar to WT at 9 and 24 DAP, but substantially lower at 14 and 18 DAP (Fig. 4*C*), as were tocomonoenols at all timepoints (Fig. 4*D*). Chlorophyll intermediates in the four transgenics were also similar to WT at 9 and 24 DAP but 2- to 10-fold higher than WT at 14 and 18 DAP in *AtVTE7-YFP* overexpressers, due almost exclusively to large increases in partially reduced Chl a species (Fig. 4*C*). Tocomonoenol levels in *AtVTE7-YFP* overexpressers were also equivalent to or higher than WT at 14 DAP and beyond. In contrast, the levels of chlorophyll intermediates, tocomonoenols and tocopherols in the two *ZmVTE7-YFP* overexpressers were nearly identical to *Atvte7-1* at 14 and 18 DAP (Fig. 4*D*), mirroring the limited seed complementation of these two lines (Fig. 3*A*).

Arabidopsis leaves provide the opportunity to assess the in vivo impact of ectopic VTE7 activity in a tissue that normally lacks it. Consistent with the lack of *AtVTE7* expression in leaves, WT and *Atvte7-1* were indistinguishable in terms of tocopherols, tocomonoenols and chlorophyll intermediates, and as in seeds, Chl a_{THGG}, was most abundant (Fig. 3*B* and 5*A*). Overexpression of either *AtVTE7* or *ZmVTE7* elevated total chlorophyll intermediates to similar levels, significantly so relative to WT and *Atvte7-1* in transgenic lines with the highest leaf tocopherol levels (Fig. 5*A*), due predominantly to large increases in Chl a_{THGG}, Chl b_{THGG} and Chl a_{DHGG}. Tocomonoenols were differentially impacted, increasing less than twofold and up to eightfold in *AtVTE7* and *ZmVTE7* overexpressers, respectively (Fig. 5*B*). To better understand their biosynthetic relationships, correlation analysis of the four metabolite classes was performed with WT, *Atvte7-1* and transgenic lines. *AtVTE7* overexpressers showed strong, significant, positive correlations of total chlorophyll intermediates with both total tocomonoenols and totalT (r^2 = +0.84 and +0.85, respectively) while total chlorophylls lacked significant correlations (Fig. 5*C*). *ZmVTE7* correlations were more complex with total chlorophyll intermediates also being strongly and positively correlated with total tocomonoenols (r^2 = +0.84) but less so with totalT (r^2 = +0.48). Unlike *AtVTE7* overexpressers, total chlorophyll levels in *ZmVTE7* overexpressers were negatively and significantly correlated with the other three metabolites (Fig. 5*D*).

Discussion

Tocopherols are most abundant in seeds but attempts to engineer increases in seed totalT by overexpressing *VTE* genes have, except for *VTE2*, had limited impact (2). Studies of natural variation for seed tocopherols in various organisms has shown that activities beyond the six core *VTE* pathway are involved (2, 9–12). Our GWA results agree with these prior linkage and association studies in that only two of seven significant GWAs mapped to *VTE* genes. As in other species, the major association for α-tocopherol was *VTE4*, the final enzyme in α-tocopherol synthesis (2, 9–12). The six remaining GWAs were for totalT and/or γT (which accounts for 95% of seed totalT), but only one mapped to a *VTE* gene, *VTE5*, the first of two kinases needed to produce PDP from phytol. That *vte5* null mutants reduce seed and leaf totalT by 80% and 65%, respectively (24), while leaf tocopherols are eliminated in a *vte5 vte6* double mutant (6) indicates the majority of PDP for tocopherol synthesis in both tissues is dependent on VTE5/VTE6 phytol phosphorylation. The association of *VTE5* with totalT is direct evidence that the phytol phosphorylation pathway may limit totalT accumulation in seeds. A third GWA mapped to an a priori enzyme of aromatic amino acid metabolism, tyrosine aminotransferase 2 (TAT2), one of two paralogs that produce the substrate for HGA synthesis. *TAT1* is strongly expressed in *Arabidopsis* leaves where *tat1* knockouts reduced leaf totalT 50% (25), *TAT2* is strongly expressed in developing seeds where it likely plays a similar role. To assess for associations that could not pass the stringent genome-wide significance test, we performed a pathway level analysis using 73 a priori genes, which identified four additional weaker associations: VTE2 (which utilizes HGA and PDP), GGR (which carries out the three-step reduction of Chl_{GG} to Chl_{phytyl}), the aromatic amino acid enzyme arogenate dehydratase 4 (ADT4) and the MEP pathway enzyme 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ISPF). Taken together these genome-wide and pathway level results are consistent with the provision of HGA and PDP, the cosubstrates for VTE2, being major determinants of totalT variation in *Arabidopsis* seeds and, with the exception of *VTE5*, other *VTE* genes making limited contributions.

The most significant totalT genome-wide association signal mapped to AT5G39220 (*AtVTE7*), an uncharacterized, seed-specific, chloroplast envelope-targeted protein containing an ABH fold (Interpro domain IPR029058) and one of more than 600 ABH gene family members in *Arabidopsis* (26). Other relevant ABH family members include the three esterases known to hydrolyze phytol from chlorophylls and/or pheophytin a in vitro: two chlorophyllases (*CLHs*) that are defense-related and not chloroplast-localized, and two thylakoid-associated activities, chlorophyll dephytylase1 (*CLD1*), which is involved in chlorophyll turnover but highly active toward pheophytin a in vitro, and

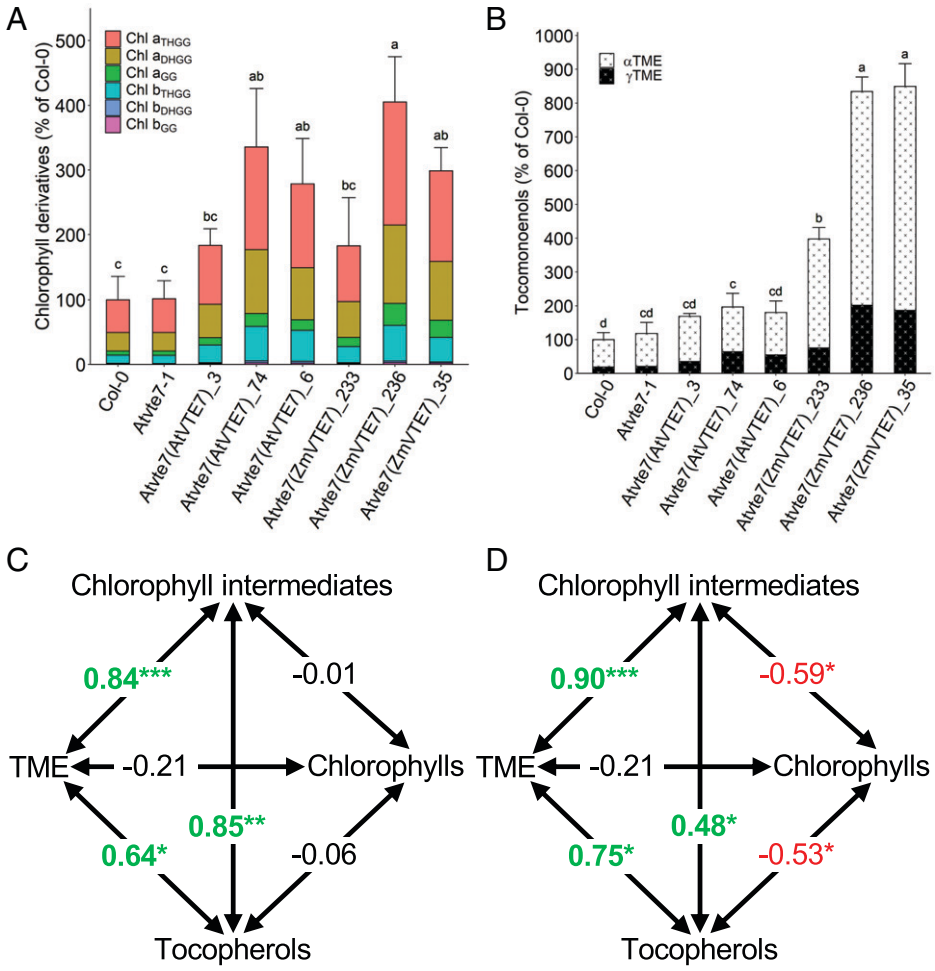


Fig. 5. Chlorophyll derivatives (*A*) and tocomonoenol (*B*) levels in leaves of *Atvte7-1* and lines overexpressing *AtVTE7* or *ZmVTE7* in *Atvte7-1* mutant and correlations between compounds (*C* and *D*). Compounds were assessed by LC-MS and expressed as percentage of Col-0. One biological replicate corresponds to five pooled 4-wk-old rosettes. Means and SDs are from three biological replicates. Means significantly different from each other in Tukey's Honest Significant Difference test are indicated with letters at each stage ($P < 0.05$). Significant positive and negative correlations are indicated by green and red, respectively, in (*C*) and (*D*) with mutant means significantly different from Col-0 in Dunnett *t* tests are indicated by asterisks: *< 0.05; **< 0.01; ***< 0.001.

pheophytinase (*PPH*), which is specific for pheophytin a and only expressed in senescing tissues (8, 27–29). All three activities had been postulated to provide phytol for tocopherol synthesis but knockdowns, knockouts and overexpression, individually and in some combinations, had little effect on seed or leaf tocopherol content, consistent with other activities being involved (7, 27, 30, 31).

In contrast to *CLH*, *CLD1*, and *PPH*, *AtVTE7* has characteristics expected for an esterase that releases phytol from chlorophyll for use in tocopherol synthesis. First, *Atvte7* null mutants decreased seed totalT levels by 55% (Fig. 2), two-thirds of the 80% of seed totalT whose synthesis is *VTE5*-dependent (6, 32). Second, ectopic overexpression of *AtVTE7* in leaves increased totalT 3.6-fold (Fig. 3), second only to *VTE2* in terms of overexpression impact (21), and combined overexpression of *AtVTE7* and *AtVTE2* in leaves was additive, consistent with *VTE7* providing a *VTE2* cosubstrate, phytyl-DP. Finally, all sequenced plant genomes contain at least one *VTE7* ortholog with the maize ortholog, *ZmVTE7*, being present in an unresolved, kernel totalT NAM QTL that lacked a priori genes and, after the two *POR* loci, had the largest effect of 18 totalT QTL identified in the panel (12). A leaky *Zmvte7-1* allele decreased kernel and leaf totalT by 38% and 45%, respectively, which likely underestimates its contribution to totalT synthesis given both tissues still contained *ZmVTE7* mRNA at 20% of WT levels. Overexpressing *ZmVTE7* in *Atvte7-1* partially complemented its seed phenotype while ectopic overexpression in leaves increased totalT 6.9-fold, nearly double that of *AtVTE7* overexpression. These combined results indicate that the in vivo activity and function of VTE7 in tocopherol synthesis is conserved between *A. thaliana* and *Z. mays*, that VTE7 activity makes a major contribution to tocopherol synthesis in tissues where it is endogenously or exogenously expressed and that *VTE7* is a major determinant of seed totalT natural variation in both dicots and monocots.

Degradation/turnover of bulk thylakoid chlorophyll has been assumed to be the source of phytol for tocopherol synthesis and given the impact modifying VTE7 activity has on tocopherols, we anticipated chlorophyll levels would also be impacted and were surprised to find no differences between WT, *Atvte7* mutants and *AtVTE7* transgenics (*SI Appendix*, Fig. S14 and Datasets S6 and S8). Untargeted metabolomics of developing WT and *Atvte7-1* seed showed the impact of *Atvte7-1* on metabolism was limited and focused with only 11 of the more than 3,500 mass/time peaks identified being significantly different between genotypes. Ten of these metabolites were lower in *Atvte7-1* including the most abundant tocopherol in seed, γT, and large fold-changes in four chlorophyll biosynthetic intermediates with partially reduced side chains (GG, DHGG, and THGG; Dataset S10). Ectopic overexpression of *VTE7* in *Arabidopsis* leaves (which normally lack *VTE7* expression) also

conditioned significant increases in both tocopherols and these same chlorophyll biosynthetic intermediates which, unlike chlorophylls, were positively correlated with leaf tocopherol content (Fig. 5*C*). These intermediates are generated by GGR during the final sequence of chlorophyll synthesis, a process that involves GGR, POR, and two integral membrane proteins, chlorophyll synthase (CHS) and light harvesting protein-like 3 (LIL3). LIL3 is nonenzymatic and binds POR and GGR, which posttranslationally stabilizes and localizes them to the membrane (33–36). LIL3 also binds a variety of chlorophyll metabolites (e.g., protochlorophyllide, chlorophyllide a, Chl a_{GG} and Chl a_{phytyl}) and is thought to channel such potentially phototoxic substrates, intermediates, and products between enzymes (34, 37) (Fig. 6). All four proteins are required for tocopherol and chlorophyll synthesis which are both severely impacted by the loss of any single protein (34, 36, 38–41). This dependency of tocopherols on chlorophyll synthesis may explain why VTE7 is active in vitro against synthetic ester substrates but not chlorophylls or pheophytin a: VTE7 may require active chlorophyll biosynthesis for its activity.

While underlying the mechanism(s) are still unclear, alterations in tocopherols and partially reduced chlorophyll intermediates in *VTE7* modified genotypes are consistent with VTE7 affecting chlorophyll biosynthesis during hydrolysis of partially or fully reduced prenyl alcohols for tocopherol synthesis. The prenyl alcohols could be converted to their corresponding diphosphates by VTE5 and VTE6 while the chlorophyllide a produced, which LIL3 binds (42), could enter another cycle of geranylgeranylation and reduction (Fig. 6). Evidence supporting aspects of this model include that GGR carries out the three-step reduction of Chl a_{GG} and GGDP to Chl a_{phytyl} and PDP in vitro, respectively, via their corresponding DHGG and THGG intermediates (Fig. 6) (5). Isolated chloroplasts contain kinase activities that can generate PDP from either phytol or geranylgeraniol (GG) and GGDP from GG (43, 44). Although VTE5 and VTE6 have only been assayed with phytol and phytyl-P as substrates, respectively (6, 24), they are the most likely activities for phosphorylating GG (and DHGG and THGG) in vivo. Consistent with this, the *vte6-1* mutant had a 4-fold increase in its known substrate, phytyl-P, and also a 6-fold increase in geranylgeranyl-P (6, 32). Likewise, VTE2 preferentially uses PDP but can also utilize GGDP both in vitro and in vivo and

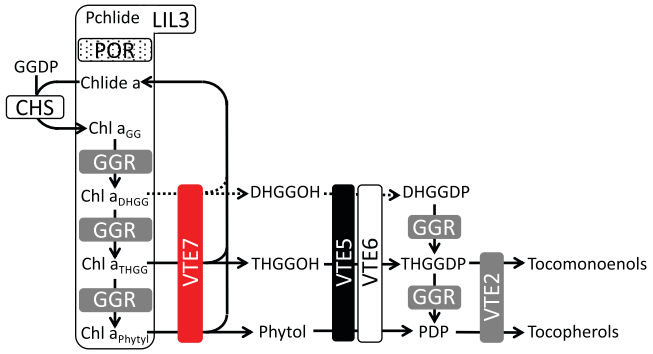


Fig. 6. Model for the role of VTE7 in connecting chlorophyll biosynthesis with tocopherol synthesis. Proteins are indicated by lettered boxes with black highlight indicating associations with genome-wide significance, gray highlight pathway level associations and white highlight no association in this study. VTE7 is highlighted in red and had the highest genome-wide significance in this study and the third largest kernel totalT QTL effect in the maize NAM panel (12). POR is stippled black as though not associated with natural variation in this study, two maize POR paralogs had the largest kernel totalT QTL effects in the maize NAM panel (12). The large white box indicates, LIL3, which physically interacts with POR and GGR associating them with the membrane. Black lines are supported directly by evidence while dashed lines are postulated. VTE5, phytol kinase; VTE6, phytylphosphate kinase; VTE2, homogentisate phytyl transferase.

though not tested (45), it seems likely it could also utilize the intermediates leading from GGDP to PDP, DHGG-DP and THGG-DP. Indeed, the presence of tocomonoenols, whose VTE2-dependent synthesis requires THGG-DP (23, 45) (Figs. 4*D* and 5*C*), is consistent with this proposed activity. Finally, compared to WT, the levels of tocopherols, tocomonoenols and Chl a_{THGG} were substantially reduced in *Atvte7-1* seeds (Fig. 4), while in *VTE7* overexpressing leaves all three compounds were significantly elevated and strongly and positively correlated (Figs. 3 *B* and 5 *A*–*C*). Future experiments to test aspects of the model in Fig. 6 include assaying VTE2, VTE5, and VTE6 against these newly proposed substrates and assessing whether VTE7 interacts with GGR, LIL3, POR, or CHS in vitro or in vivo.

The identification of *VTE7* as a large effect locus underlying natural variation for seed totalT and the accompanying mutant and overexpression phenotypes are consistent with VTE7 playing a key role in providing phytol for tocopherol synthesis from chlorophyll biosynthesis. A key difference between maize and *Arabidopsis* is that in maize VTE7 was recruited for tocopherol synthesis in both leaves and seeds, while in *Arabidopsis* VTE7 is seed specific (Fig. 2 *B*–*E* and *SI Appendix*, Fig. S5). That AtVTE7 plays no role in leaf tocopherol synthesis in *Arabidopsis* indicates additional hydrolases must provide phytol for leaf tocopherol synthesis and for the remaining tocopherol produced in *Atvte7-1* seeds. These activities are most likely to reside in other ABH gene family members but mutation of one likely candidate, the second member of the VTE7 clade, AT1G13820, singly and in combination with *Atvte7-1* had no effect on leaf or seed tocopherol levels (*SI Appendix*, Fig. S7). The limited identity of the known phytol-releasing enzymes in the *Arabidopsis* ABH gene family (*VTE7*, *PPH*, CLD,1 and *CHLs*) indicate phytol hydrolysis has arisen independently multiple times in the ABH gene family. A systematic assessment of other chloroplast targeted ABH family members will be required to identify the additional activities responsible for the remaining VTE5/VTE6 dependent tocopherol synthesis in seeds and leaves.

Materials and Methods

Details of the methods used in this work, including plant material and growth conditions, HPLC quantification, statistical analysis, genome-wide and pathway level association analysis, phylogenetic analysis, plant transformation, gene expression and RNAseq analysis are described in SI Appendix, Materials and Methods. Information about the LC-MS/MS analysis, confocal microscopy and recombinant protein expression and purification are also provided in *SI Appendix, Materials and Methods*. National Center for Biotechnology Information (NCBI) Sequence Read Archive accessions for RNAseq data are under Bioproject ID PRJNA767331. All other study data are included in the article and supporting information.

Data Availability. RNAseq reads data have been deposited in NCBI Sequence Read Archive ID PRJNA767331. All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank the Center for Advanced Microscopy at Michigan State University for assistance with confocal imaging; the Mass Spectrometry and Metabolomics Core at Michigan State University and Dan Jones for assistance with mass spectrometry measurements; undergraduate students Laura Vendal, Joe Dziedziula, and Violet Raterman for excellent assistance with plant cultivation and seed harvesting and Christine Diepenbrock for help with *ZmVTE7* correlated expression and effect QTL (ceeQTL) analysis. This work was supported by the NSF, Plant Genome Research Project (Award No. 1546657).

Author affiliations: [a]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824; [b]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853; and [c]Department of Plant Biology, Michigan State University, East Lansing, MI 48824

1. L. Mène-Saffrané, A. D. Jones, D. DellaPenna, Plastochromanol-8 and tocopherols are essential lipid-soluble antioxidants during seed desiccation and quiescence in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17815–17820 (2010).
2. D. DellaPenna, L. Mène-Saffrané, "Vitamin E" in *Biosynthesis of Vitamins in Plants Part B*, F. Rébeillé, R. Douce, Eds. (Academic Press, Elsevier Ltd., 2011), pp. 179–227.
3. S. Fritsche, X. Wang, C. Jung, Recent advances in our understanding of tocopherol biosynthesis in plants: An overview of key genes, functions, and breeding of vitamin E improved crops. *Antioxidants* **6**, 99 (2017).
4. F. Galli *et al.*, Vitamin E: Emerging aspects and new directions. *Free Radic. Biol. Med.* **102**, 16–36 (2017).
5. Y. Keller, F. Bouvier, A. d'Harlingue, B. Camara, Metabolic compartmentation of plastid prenyllipid biosynthesis–Evidence for the involvement of a multifunctional geranylgeranyl reductase. *Eur. J. Biochem.* **251**, 413–417 (1998).
6. K. Vom Dorp *et al.*, Remobilization of phytol from chlorophyll degradation is essential for tocopherol synthesis and growth of *Arabidopsis*. *Plant Cell* **27**, 2846–2859 (2015).
7. W. Zhang *et al.*, Chlorophyll degradation: The tocopherol biosynthesis-related phytol hydrolase in *Arabidopsis* seeds is still missing. *Plant Physiol.* **166**, 70–79 (2014).
8. Y.-P. Lin, M.-C. Wu, Y. Y. Charng, Identification of a chlorophyll dephytylase involved in chlorophyll turnover in *Arabidopsis*. *Plant Cell* **28**, 2974–2990 (2016).
9. X. Shutu *et al.*, Dissecting tocopherols content in maize (*Zea mays L.*), using two segregating populations and high-density single nucleotide polymorphism markers. *BMC Plant Biol.* **12**, 201 (2012).
10. C. Park *et al.*, Identification of quantitative trait loci for increased α-tocopherol biosynthesis in wild soybean using a high-density genetic map. *BMC Plant Biol.* **19**, 510 (2019).
11. A. E. Lipka *et al.*, Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. *G3 (Bethesda)* **3**, 1287–1299 (2013).
12. C. H. Diepenbrock *et al.*, Novel loci underlie natural variation in vitamin E levels in maize grain. *Plant Cell* **29**, 2374–2392 (2017).
13. W. Zhan *et al.*, An allele of *ZmPORB2* encoding a protochlorophyllide oxidoreductase promotes tocopherol accumulation in both leaves and kernels of maize. *Plant J.* **100**, 114–127 (2019).
14. 1001 Genomes Consortium, 1,135 Genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016).
15. S. Kim *et al.*, Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* **39**, 1151–1155 (2007).
16. P. Cingolani *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
17. M. Kamruzzaman Munshi, Y. Kobayashi, T. Shikanai, Identification of a novel protein, CRR7, required for the stabilization of the chloroplast NAD(P)H dehydrogenase complex in Arabidopsis. *Plant J.* **44**, 1036–1044 (2005).
18. H. Bannai, Y. Tamada, O. Maruyama, K. Nakai, S. Miyano, Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **18**, 298–305 (2002).
19. D. R. McCarty *et al.*, Steady-state transposon mutagenesis in inbred maize. *Plant J.* **44**, 52–61 (2005).
20. R. H. Köhler, J. Cao, W. R. Zipfel, W. W. Webb, M. R. Hanson, Exchange of protein molecules through connections between higher plant plastids. *Science* **276**, 2039–2042 (1997).
21. E. Collakova, D. DellaPenna, Homogentisate phytyltransferase activity is limiting for tocopherol biosynthesis in *Arabidopsis*. *Plant Physiol.* **131**, 632–642 (2003).
22. E. A. P. Ekanayaka, M. D. Celiz, A. D. Jones, Relative mass defect filtering of mass spectra: A path to discovery of plant specialized metabolites. *Plant Physiol.* **167**, 1221–1232 (2015).
23. S. Pellaud *et al.*, WRINKLED1 and ACYL-COA:DIACYLGLYCEROL ACYLTRANSFERASE1 regulate tocochromanol metabolism in *Arabidopsis*. *New Phytol.* **217**, 245–260 (2018).
24. H. E. Valentin *et al.*, The *Arabidopsis vitamin E pathway gene5-1* mutant reveals a critical role for phytol kinase in seed tocopherol biosynthesis. *Plant Cell* **18**, 212–224 (2006).
25. M. Wang, K. Toda, A. Block, H. A. Maeda, TAT1 and TAT2 tyrosine aminotransferases have both distinct and shared functions in tyrosine metabolism and degradation in *Arabidopsis thaliana*. *J. Biol. Chem.* **294**, 3563–3576 (2019).
26. J. T. Mindrebo, C. M. Nartey, Y. Seto, M. D. Burkart, J. P. Noel, Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Curr. Opin. Struct. Biol.* **41**, 233–246 (2016).
27. S. Hörtensteiner, B. Kräutler, Chlorophyll breakdown in higher plants. *Biochimica et Biophysica Acta (BBA). Bioenergetics* **1807**, 977–988 (2011).
28. X. Hu, T. Jia, S. Hörtensteiner, A. Tanaka, R. Tanaka, Subcellular localization of chlorophyllase2 reveals it is not involved in chlorophyll degradation during senescence in Arabidopsis thaliana. *Plant Sci.* **290**, 110314 (2020).
29. X. Hu *et al.*, Reexamination of chlorophyllase function implies its involvement in defense against chewing herbivores. *Plant Physiol.* **167**, 660–670 (2015).
30. B. S. Lira *et al.*, Pheophytinase knockdown impacts carbon metabolism and nutraceutical content under normal growth conditions in tomato. *Plant Cell Physiol.* **57**, 642–653 (2016).
31. Y. P. Lin, Y. Y. Charng, Supraoptimal activity of CHLOROPHYLL DEPHYTYLASE1 results in an increase in tocopherol level in mature *Arabidopsis* seeds. *Plant Signal. Behav.* **12**, e1382797 (2017).
32. K. vom Dorp, "Phytol and tocopherol metabolism in *Arabidopsis thaliana*," PhD thesis, Rheinische Friedrich-Wilhelms University of Bonn, Bonn, Germany (2015).
33. J. Herbst, D. Hey, B. Grimm, "Posttranslational control of tetrapyrrole biosynthesis: Interacting proteins, chaperones, auxiliary factors" in *Metabolism, Structure and Function of Plant Tetrapyrroles: Control Mechanisms of Chlorophyll Biosynthesis and Analysis of Chlorophyll-Binding Proteins*, B. Grimm, Ed. (Academic Press, Elsevier Ltd., 2019), pp. 163–194.
34. D. Hey *et al.*, LIL3, a light-harvesting complex protein, links terpenoid and tetrapyrrole biosynthesis in *Arabidopsis thaliana*. *Plant Physiol.* **174**, 1037–1050 (2017).
35. R. Tanaka, U. Oster, E. Kruse, W. Rüdiger, B. Grimm, Reduced activity of geranylgeranyl reductase leads to loss of chlorophyll and tocopherol and to partially geranylgeranylated chlorophyll in transgenic tobacco plants expressing antisense RNA for geranylgeranyl reductase. *Plant Physiol.* **120**, 695–704 (1999).
36. R. Tanaka *et al.*, LIL3, a light-harvesting-like protein, plays an essential role in chlorophyll and tocopherol biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16721–16725 (2010).
37. A. E. Mork-Jansson, L. A. Eichacker, Characterization of chlorophyll binding to LIL3. *PLoS One* **13**, e0192228 (2018).
38. C. Zhang *et al.*, Chlorophyll synthase under epigenetic surveillance is critical for vitamin E synthesis, and altered expression affects tocopherol levels in *Arabidopsis*. *Plant Physiol.* **168**, 1503–1511 (2015).
39. Y. Zhou *et al.*, Mutation of the *light-induced yellow leaf 1* gene, which encodes a geranylgeranyl reductase, affects chlorophyll biosynthesis and light sensitivity in rice. *PLoS One* **8**, e75299 (2013).
40. C. Li *et al.*, A *lil3 chlp* double mutant with exclusive accumulation of geranylgeranyl chlorophyll displays a lethal phenotype in rice. *BMC Plant Biol.* **19**, 456 (2019).
41. A. Mork-Jansson *et al.*, Lil3 assembles with proteins regulating chlorophyll synthesis in Barley. *PLoS One* **10**, e0133145 (2015).
42. L. Thai *et al.*, Farnesol is utilized for isoprenoid biosynthesis in plant cells via farnesyl pyrophosphate formed by successive monophosphorylation reactions. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13080–13085 (1999).
43. J. Soll, G. Schultz, W. Rüdiger, J. Benz, Hydrogenation of geranylgeraniol: Two pathways exist in spinach chloroplasts. *Plant Physiol.* **71**, 849–854 (1983).
44. C. Zhang *et al.*, Genetic and biochemical basis for alternative routes of tocotrienol biosynthesis for enhanced vitamin E antioxidant production. *Plant J.* **73**, 628–639 (2013).
45. S. Pellaud, L. Mène-Saffrané, Metabolic origins and transport of vitamin E biosynthetic precursors. *Front. Plant Sci.* **8**, 1959 (2017).