

GWAS analysis combined with QTL mapping identify *CPT3* and *ABH* as genes underlying dolichol accumulation in *Arabidopsis*

Katarzyna Gawarecka¹, Joanna Siwinska², Jaroslaw Poznanski¹, Agnieszka Onysk¹, Przemyslaw Surowiecki¹, Karolina Sztompka¹, Liliana Surmacz¹, Ji Hoon Ahn³, Arthur Korte⁴, Ewa Swiezewska¹, and Anna Ihnatowicz²

¹Polska Akademia Nauk Instytut Biochemii i Biofizyki

²Intercollegiate Faculty of Biotechnology of the University of Gdańsk and the Medical University of Gdańsk

³Korea University

⁴Center for Computational and Theoretical Biology, University of Wurzburg

September 25, 2021

Abstract

Dolichols (Dols), ubiquitous components of living organisms, are indispensable for cell survival. In plants, as well as other eukaryotes, Dols are crucial for posttranslational protein glycosylation, aberration of which leads to fatal metabolic disorders in humans and male sterility in plants. Until now, the mechanisms underlying Dol accumulation remain elusive. In this report, we have analyzed the natural variation of the accumulation of Dols and six other isoprenoids between more than 120 *Arabidopsis thaliana* accessions. Subsequently, by combining QTL and GWAS approaches, we have identified several candidate genes involved in the accumulation of Dols, polyprenols, plastoquinone, and phytosterols. The role of two genes implicated in the accumulation of major Dols in *Arabidopsis* – the AT2G17570 gene encoding a long searched for cis-prenyltransferase (*CPT3*) and the AT1G52460 gene encoding an alpha-beta hydrolase (*ABH*) – is experimentally confirmed. These data will help to generate Dol-enriched plants which might serve as a remedy for Dol-deficiency in humans.

GWAS analysis combined with QTL mapping identify *CPT3* and *ABH* as genes underlying dolichol accumulation in *Arabidopsis*

Katarzyna Gawarecka^{1,3}, Joanna Siwinska², Jaroslaw Poznanski¹, Agnieszka Onysk¹, Przemyslaw Surowiecki¹, Karolina Sztompka¹, Liliana Surmacz¹, Ji Hoon Ahn³, Arthur Korte⁴, Ewa Swiezewska^{1*}, Anna Ihnatowicz^{2*}

¹Institute of Biochemistry and Biophysics Polish Academy of Sciences, 02-106 Warszawa, Poland

²Intercollegiate Faculty of Biotechnology of University of Gdansk and Medical University of Gdansk, 80-307 Gdansk, Poland

³Department of Life Sciences, Korea University, Seoul, 02841, Korea

⁴Center for Computational and Theoretical Biology, University of Wurzburg, 97074 Wurzburg, Germany

*Correspondence

Anna Ihnatowicz, Department of Biotechnology, Intercollegiate Faculty of Biotechnology of University of Gdansk and Medical University of Gdansk, 80-307 Gdansk, Poland

E-mail: anna.ihnatowicz@biotech.ug.edu.pl

Ewa Swiezewska, Laboratory of Lipid Biochemistry, Institute of Biochemistry and Biophysics Polish Academy of Sciences, 02-106 Warszawa, Poland

E-mail: ewas@ibb.waw.pl

Running title: New determinants of dolichol accumulation

Abstract

Dolichols (Dols), ubiquitous components of living organisms, are indispensable for cell survival. In plants, as well as other eukaryotes, Dols are crucial for posttranslational protein glycosylation, aberration of which leads to fatal metabolic disorders in humans and male sterility in plants. Until now, the mechanisms underlying Dol accumulation remain elusive. In this report, we have analyzed the natural variation of the accumulation of Dols and six other isoprenoids between more than 120 *Arabidopsis thaliana* accessions. Subsequently, by combining QTL and GWAS approaches, we have identified several candidate genes involved in the accumulation of Dols, polyprenols, plastoquinone, and phytosterols. The role of two genes implicated in the accumulation of major Dols in *Arabidopsis* – the AT2G17570 gene encoding a long searched for cis-prenyltransferase (CPT3) and the AT1G52460 gene encoding an alpha-beta hydrolase (ABH) – is experimentally confirmed. These data will help to generate Dol-enriched plants which might serve as a remedy for Dol-deficiency in humans.

Keywords: isoprenoid, polyprenol, dolichol, natural variation, plant-environment interactions, secondary metabolism, QTL mapping, GWAS

INTRODUCTION

Isoprenoids (also known as terpenes) are a large and diverse group of compounds comprised of more than 40,000 chemical structures (Bohlmann & Keeling, 2008). Linear polymers containing from 5 to more than 100 isoprene units are called polyisoprenoids (Swiezewska & Danikiewicz, 2005). Due to the hydrogenation status of their OH-terminal, (α -) isoprene unit, polyisoprenoids are subdivided into α -unsaturated polyprenols (hereafter named Prens) and α -saturated dolichols (hereafter named Dols) (Figure 1). Prens are common for bacteria, green parts of plants, wood, seeds, and flowers, while Dols are constituents of plant roots as well as animal and fungal cells (Rezanka & Votruba, 2001). In eukaryotic cells, the dominating polyisoprenoid components are accompanied by traces of their counterparts, e.g., Prens are accompanied by Dols in photosynthetic tissues (Skorupinska-Tudek et al., 2003).

All isoprenoids are synthesized from isopentenyl and dimethylallyl diphosphate (IPP and DMAPP) molecules, which in plants are derived from the cytoplasmic mevalonate (MVA) and plastidial methylerythritol phosphate (MEP) pathways (Hemmerlin, Harwood, & Bach, 2012; Lipko & Swiezewska, 2016). Formation of the polyisoprenoid chains of both Pren and Dol from IPP is executed by enzymes called *cis*-prenyltransferases (CPTs), which are responsible for elongation of an all-*trans* initiator molecule, most commonly farnesyl or geranylgeranyl diphosphate. This reaction generates a mixture of polyprenyl diphosphates (PolyprenylIPP) of similar, CPT-specific, lengths. In *Arabidopsis thaliana* (hereafter named *Arabidopsis*), only three (Oh, Han, Ryu, & Kang, 2000; Cunillera, Arrò, Fors, Manzano, & Ferrer, 2000; Surowiecki, Onysk, Manko, Swiezewska, & Surmacz, 2019; Kera, Takahashi, Sutoh, Koyama, & Nakayama, 2012; Surmacz, Plochocka, Kania, Danikiewicz, & Swiezewska, 2014; Akhtar et al., 2017) out of nine putative CPTs (Surmacz & Swiezewska, 2011) have been characterized at the molecular level. Interestingly, none of these well-characterized CPTs (CPT1, -6 or -7) is responsible for the synthesis of the major ‘family’ of Dols (Dol-16 dominating) accumulated in *Arabidopsis* tissues. The polyprenyl diphosphates resulting from CPT activity undergo then either dephosphorylation to Prens and/or reduction to Dols. The reduction reaction is catalyzed by polyprenol reductases, two of which have been recently described in *Arabidopsis* (Jozwiak et al., 2015). Although this biosynthetic scheme is generally accepted some steps of Pren and Dol biosynthesis pathways remain unknown.

Isoprenoids are implicated in vital processes in plants, e.g. in photosynthesis and stress response (chlorophylls, carotenoids, plastoquinone, and tocopherols), or in the synthesis of plant hormones (carotenoids,

sterols), or they function as structural components of membranes (sterols) (Tholl, 2015). Polyisoprenoids are modulators of the physico-chemical properties of membranes, but they are also involved in other specific processes. Dolichyl phosphate (DoLP) serves as an obligate cofactor for protein glycosylation and for the formation of glycosylphosphatidylinositol (GPI) anchors, while Prenyl, in turn, have been shown to play a role in plant photosynthetic performance (Akhtar et al., 2017). Importantly, an increased content of Prenyl improves the environmental fitness of plants (Hallahan & Keiper-Hrynko, 2006). Additionally, it has also been suggested that in plants Prenyl and Dols might participate in cell response to stress since their content is modulated by the availability of nutrients (Jozwiak et al., 2013) and by other environmental factors (xenobiotics, pathogens, and light intensity) (summarized in Surmacz & Swiezewska, 2011). Moreover, the cellular concentration of Prenyl and Dols is also considerably increased upon senescence (summarized in Swiezewska & Danikiewicz, 2005). These observations suggest that eukaryotes might possess, so far elusive, regulatory mechanisms allowing them to control polyisoprenoid synthesis and/or degradation.

Most traits important in agriculture, medicine, ecology, and evolution, including variation in chemical compound production, are of a quantitative nature and are usually due to multiple segregating loci (Mackay 2001). Arabidopsis is an excellent model for studying natural variation due to its genetic adaptation to different natural habitats and its extensive variation in morphology, metabolism, and growth (Alonso-Blanco et al., 2009; Fusari et al., 2017). Natural variation for many traits has been reported in Arabidopsis, including primary and secondary metabolism (Mitchell-Olds & Pedersen, 1998; Kliebenstein, Gershenzon, & Mitchell-Olds, 2001; Sergeeva et al., 2004; Tholl, Chen, Petri, Gershenzon, & Pichersky, 2005; Keurentjes et al., 2006; Meyer et al., 2007; Lisek et al., 2008; Rowe, Hansen, Halkier, & Kliebenstein, 2008; Siwinska et al., 2015). Until now, no systematic analysis of the natural variation of polyisoprenoids has been performed for any plant species.

Therefore, in this study, we decided to use the model plant Arabidopsis to explore the natural variation of Prenyl and Dols. Importantly, Arabidopsis provides the largest and best-described body of data on the natural variation of genomic features of any plant species (Kawakatsu et al., 2016; The 1001 Genomes Consortium, 2016). Over 6,000 different Arabidopsis accessions that can acclimate to enormously different environments (Kramer, 2015) have been described so far (Weigel & Mott, 2009).

In order to identify genes that are responsible for modulation of polyisoprenoid content, we used both a quantitative trait loci (QTL) mapping approach and genome-wide association studies (GWAS). So far, neither QTL nor GWAS has been used for the analysis of Prenyl and Dols. Traditional linkage mapping usually results in detection of several QTLs with a high statistical power, making it a powerful method in the identification of genomic regions that co-segregate with a given trait in mapping populations (Koornneef, Alonso-Blanco, & Vreugdenhil, 2004; Korte & Farlow, 2013). But the whole procedure including the identification of underlying genes is usually time-consuming and laborious. Moreover, the mapped QTL regions can be quite large, making it sometimes impossible to identify the causative genes. Another issue is that the full range of natural variation is not analyzed in QTL studies using bi-parental populations, because they are highly dependent on the genetic diversity of the two parents and may reflect rare alleles. GWAS studies profit from a wide allelic diversity, high resolution and may lead to the identification of more evolutionarily relevant variation (Kooke et al., 2016). Therefore, it is possible to overcome some limitations of QTL analyses by using the GWAS approach, which can be used to narrow down the candidate regions (Korte & Farlow, 2013; Han et al., 2018). But it should be kept in mind that GWAS also has its limitations, such as dependence on the population structure, the reliance on SNPs rather than gene structural variants or the potential for false-positive as well as false-negative errors (Zhu, Gore, Buckler, & Yu, 2008; Korte & Farlow, 2013). We have applied here both QTL mapping and GWAS analyses because it has been shown that the combination of these two methods can alleviate their respective limitations (Zhao et al., 2007; Brachi et al., 2010).

The described here application of QTL and GWAS led to identification of several candidate genes underlying the accumulation of polyisoprenoids. Additionally, to get insight into the biosynthetic pathways of Dols and Prenyl in a broader cellular context, a set of seven isoprenoid compounds was analyzed and subsequently candidate genes were selected. The most interesting of the identified genes were *cis-prenyltransferase*

3(*CPT3* , AT2G17570, identified through QTL) and *alpha-beta hydrolase* (*ABH* , AT1G52460, identified through GWAS). *CPT3*, although biochemically not characterized, has been demonstrated to efficiently incorporate *in vitro* IPP *intocis* -polyisoprenoid of an undefined chain-length thus to possess a CPT-like activity; moreover, its expression complemented the γ CTP deficiency (Kwon, Kwon, & Ro, 2016), whereas alpha-beta hydrolase has not been previously connected with polyisoprenoid biosynthesis. In this work, their involvement in Dol biosynthesis/accumulation is experimentally confirmed using mutant approach, metabolite profiling, yeast transformation, transient expression in *Nicotiana benthamiana* leaves, bimolecular fluorescence complementation (BiFC) and yeast two-hybrid (Y2H) assays. Although obtained results clearly suggest the role of *CPT3* and *ABH* in Dol accumulation one should remember that in this report analysis of the level of terpene was limited to the seedling stage and might differ for mature plants. Moreover, it should be kept in mind, however, that although *CPT3* and *ABH*, together with other genes depicted in this report, are strong candidates for being causal for the observed natural variation more studies are required to prove such role. Importantly, identification of *CPT3* fills the gap in the Dol biosynthetic route in *Arabidopsis* and, together with newly depicted *ABH*, makes the manipulation of Dol content in plants feasible. Consequently, an option for the generation of plant tissues with increased Dol content as dietary supplements for individuals suffering from Dol-deficiency is emerging. Moreover, presuming conserved role of *ABH* in Dol pathway in eukaryotes a design of a new therapeutic strategy ameliorating Dol deficiency via manipulation of the activity of respective human *ABH* seems plausible.

MATERIALS AND METHODS

Plant materials

Arabidopsis thaliana accessions used in this study are listed in the Supporting Information (Table S9). All accessions were obtained from the stock center NASC (<http://arabidopsis.info/>).

A population of advanced intercross recombinant inbred lines (AI-RIL, EstC) was obtained after crossing of the Est-1 (Estland) and Col-0 (Columbia) accessions (Balasubramanian et al., 2009). All lines were kindly provided by Maarten Koornneef from Max Planck Institute for Plant Breeding Research in Cologne, Germany. The EstC mapping population with all sequence variant database is available at the NASC under the stock number CS39389.

For miRNA-mediated knockdown of the *CPT3* gene, two pairs of primers specific to amiRNA and amiRNA* targeting the gene were designed using the Web MicroRNA Designer WMD3. The vector pRS300 was used as a template for subsequent PCR amplification and replacement of the endogenous miR319a and miR319a* sequences with appropriate amiRNA and amiRNA* of *CPT3* as described in the website protocol wmd3.weigelworld.org (Ossowski Stephan, Fitz Joffrey, Schwab Rebecca, Rieger Markus and Weigel Detlef, personal communication). The obtained stem-loop was used as a template for PCR to generate the 454 bp fragment with a CACC overhang at the 5' end, which was used for directional cloning into the pENTR/D-TOPO vector system (Invitrogen). The recombination reaction from pENTR/D-TOPO to the pGWB602 binary vector was carried out with the Gateway LR clonase II system (Invitrogen). All primers used in the construction of the *AtCPT3* silencing vector are listed in Table S10. The obtained plasmid was introduced into *Agrobacterium tumefaciens* strain GV3101, which was then used to transform *Arabidopsis* (Col-0) by the floral dip method (Weigel and Glazebrook, 2002). T1 seeds were germinated on soil and transgenic plants were selected by spraying with 0.1% BASTA in the greenhouse. Spraying was performed one week after germination and was repeated two times at two-day intervals. Additionally, the plants that survived were verified by PCR.

AtCPT3-over-expressing lines (*CPT3-OE*) were generated using a 35S::AtCPT3 construct introduced into the *A. tumefaciens* GV3101 strain. Transformation of *Arabidopsis* (Col-0) plants was performed by the floral dip method (Weigel & Glazebrook, 2002). Transformant selection was performed as described previously (Surowiecki, Onysk, Manko, Swiezewska, & Surmacz, 2019).

The T-DNA insertion mutant lines for AT1G52460, SALK_066806 and GK_823G12, were obtained from the Nottingham *Arabidopsis* Stock Center, their progeny was genotyped, and heterozygous lines were isolated.

Growth conditions

Plants were grown in a growth chamber in a long day (16 h light) photoperiod at 22 °C/18 °C at day/night. The seeds were surface-sterilized by treatment with an aqueous solution of 5% calcium hypochlorite for 8 min, subsequently rinsed four times with sterile water and planted on plates. Before location in the growth chamber, plates with seeds were kept for 4 days at 4 °C in darkness for stratification. The Arabidopsis accessions, the AI-RIL mapping population and the T-DNA insertion mutant lines were grown on Petri dishes on solid $\frac{1}{2}$ Murashige-Skoog medium with vitamins (1L of medium contained 0.5 µg nicotinic acid, 0.5 µg pyridoxine, 0.1 µg thiamine, 2 µg glycine) and 0.8% agar. For each genotype analyzed (accessions, mapping population, T-DNA mutants), plants were cultivated in at least three biological replicates, full set of plates with accessions or RIL lines was grown at the same time.

Isolation of isoprenoids

Unless indicated otherwise, entire 3-week-old seedlings were used for the isolation of all isoprenoid compounds. To elucidate the correlation between polyisoprenoid content vs. *CPT3* transcript level, the Arabidopsis seedlings, leaves, and flowers were used. For qualitative and quantitative analysis of isoprenoids, either internal (Prens, Dols, and phytosterols) or external (plastoquinone and tocopherol) standards were employed. For quantitative analysis of Prens, Dols, phytosterols, plastoquinone, and tocopherols signals corresponding to compounds of well-characterized structure were taken into consideration, exclusively.

Prens, Dols, phytosterols, plastoquinone, tocopherols, carotenoids and chlorophylls were isolated and quantified using standard methods – for details see Supplementary Methods.

δημπλεμεντατιον οφ της ψεαστ ρερ2Δ μυταντ

To express *CPT3* and *LEW1* in *Saccharomyces cerevisiae* mutant cells (*rer2* Δ mutant: *rer2* ::kanMX4*ade2-101 ura3-52 his3-200 lys2-801*), coding sequences of *CPT3* and *LEW1* (AT1G11755) were subcloned into the pESC-URA yeast dual expression vector (Agilent, Santa Clara, CA, USA) according to the manufacturer's protocol. Transformant selection and growth, as well as analyses of polyisoprenoid profile and CPY glycosylation status, were performed as described previously (Surowiecki, Onysk, Manko, Swiezewska, & Surmacz, 2019).

Subcellular localization and BiFC assays

For subcellular localization analysis of *35S::CPT3*, *A. tumefaciens* cells carrying the vectors CPT3-GFP and cd3-954 (ER-CFP, used as an organelle marker) were introduced into the abaxial side of *Nicotiana benthamiana* leaves. A BiFC assay was performed based on split EYFP. EYFP was fused to the C-terminus of CPT3 and the N-terminus of Lew1, resulting in the expression of CPT3:EYFPC and Lew1:EYFPN. CPT3:EYFPC was co-infiltrated with Lew1:EYFPN into the abaxial side of *N. benthamiana* leaves. A positive fluorescence signal (EYFP) is indicative of the restoration of EYFP due to the heterodimerization of CPT3 with Lew1.

The transient expression of CPT3, ER-CFP, and CPT3/Lew1-YFP fusion proteins was observed under a Nikon C1 confocal system built on TE2000E with 408, 488 and 543 nm laser excitations for CFP (450/35 nm emission filter) and GFP (515/30 nm emission filter), respectively.

Yeast two-hybrid (Y2H) assay

To test protein-protein interactions coding sequences of CPT3 and LEW1 were subcloned into the pENTR/D-TOPO vector and next recombined into Y2H vectors (pGADT7-GW and pGBKT7-GW) using LR Clonase II. Selected constructs were transformed into *S. cerevisiae* AH109 strain [MATa, trp 1- 901, leu2 - 3, 112, ura 3 - 52, his3 - 200, gal4:gal80:LYS2 : GAL1UAS – GAL1TATA – HIS3, GAL2UAS – GAL2TATA – ADE2,URA3 : MEL1UAS – MEL1TATA – LacZ] using the lithium acetate method. Double transformant colonies selected by colony PCR were grown in media lacking leucine (*Leu* / *Trp*). Serial dilutions of the selected double transformants were grown in plates lacking leucine, tryptophan and histidine (–*Leu* / *Trp* – *His*) supplemented with 1mM 3-AT (3-Amino-1, 2, 4-triazole). The experiments were performed in at least three replicates.

Y2H vectors pGADT7-GW (Addgene plasmid #61702; <http://n2t.net/addgene:61702>; RRID:Addgene.-61702) and pGBKT7-GW (Addgene plasmid #61703; <http://n2t.net/addgene:61703> ; RRID:Addgene.-61703) were a gift from Yuhai Cui (Lu et al., 2010).

Statistical analyses

Quantitative genetic analyses

Mean values of at least three replicates were calculated for each isoprenoid compound measured, for each AI-RIL and each natural accession. These values were used in QTL mapping and GWAS. The broad sense heritability (H^2) for isoprenoid accumulation for the AI-RIL population was estimated according to the formula: $H^2 = V_G / (V_G + V_E)$, where V_G is the among-genotype variance component and V_E is the residual (error) variance. For GWAS heritability, estimates have been extracted from the mixed model accordingly.

QTL analyses in the AI-RIL population

All obtained phenotypical data were used in QTL mapping that was performed using R software (R Core Team, 2012, <https://www.R-project.org/>) with R/qtl package (Arends, Prins, Jansen, & Broman, 2010; Broman, Wu, Sen, & Churchill, 2003; <http://www.rqtl.org/>). Stepwise qtl function was used to detect multiple-QTL models (Broman, 2008, http://www.rqtl.org/tutorials/new_multiqtl.pdf). This function requires single-QTL genome scan to locate QTLs with the highest LOD scores, then the initial model is tested using arguments for additional QTLs and interactions between QTLs search, model refinement and backward elimination of each QTL detected back to the null model. Obtained QTL models were refined with the refineqtl function; any possible interactions between QTLs were verified by the addint function. See Table S2 for detailed description of the procedure of selection of candidate genes from chosen QTL intervals.

GWAS

Genome-wide association mapping was performed on measurements for 115 – 119 different natural accessions per phenotype. The phenotypic data are available at the AraPheno database (Seren et al., 2016). The genotypic data were based on whole-genome sequencing data (The 1001 Genomes Consortium, 2016) and covered 4,314,718 SNPs for the 119 accessions. GWAS was performed with a mixed model correcting for population structure in a two-step procedure, where first all polymorphisms were analyzed with a fast approximation (emmaX, Kang et al., 2010) and afterwards the top 1000 polymorphisms were reanalyzed with the correct full model. Only polymorphisms with a minor allele count greater than 5 are reported. The kinship structure has been calculated under the assumption of the infinitesimal model using all sequence variants with a minor allele frequency of more than 5% in the whole population. The analysis was performed in R (R Core Team, 2016). The R scripts used are available at <https://github.com/arthurkorte/GWAS>. The genotype data used for GWAS are available at the 1001 Genomes Project (www.1001genomes.org).

Correlation analyzes of isoprenoid accumulation - a statistical meta-analysis

All correlation analyses were performed with the aid of R version 3.3.0 (R Core Team, 2016, <https://www.R-project.org/>) using the outliers (Komsta, 2011, R package version 0.14, <https://CRAN.R-project.org/package=outliers>) and the gplots (Warnes et al., 2016, R package version 3.0.1, <https://CRAN.R-project.org/package=gplots>). The significance level α of 0.001 was assumed in all statistical tests.

Although for each accession the level of each metabolite was measured in triplicate, the values thus obtained were analyzed separately, as indicated by the number of experimental points in the respective figures (which equals three times the number of accessions). Means were not calculated, and this approach was employed deliberately to avoid the problem of adjusting and weighing mean values and to allow testing for outliers among single replicates instead of among mean values.

The Shapiro-Wilk test (Shapiro and Wilk 1965) was used to assess the agreement of isoprenoid content in the populations with the Gaussian distribution. Since, even after filtering out of extreme values with the Grubbs' test for outliers (Grubbs 1950), a vast majority of the distributions were found non-Gaussian,

further analyses were performed using non-parametric methods. Consequently, a correlation matrix for the seven investigated isoprenoids was calculated accordingly to the Spearman's rank correlation coefficients (Spearman 1904).

A hierarchical cluster analysis of the correlation matrix was performed according to the Ward criterion (Ward 1963).

Quantitative real-time PCR analysis

Total RNA from Col-0, Stw-0, and Or-0 seedlings (1-, 2-, and 3-week-old) and leaves (4-, 5-, and 6-week-old plants) was isolated and purified using RNeasy Plant Mini Kit (Qiagen) following the manufacturer's instructions. RNA concentration and purity were verified using a NanoDropTM 1000 Spectrophotometer (Thermo Scientific, Waltham, MA). RNA was treated with RNase-free DNase I (Thermo Scientific) according to the manufacturer's instructions. 160 ng RNA per each sample was used for first-strand synthesis using SuperScriptTM II First-Strand Synthesis System for RT-PCR (Thermo Scientific) and oligo-dT primers according to the manufacturer's procedure. 2 μ l of cDNA was used for real-time PCR analysis, using 0.6 μ l each of gene-specific primers listed in Table S10 in a total volume of 20 μ l of Luminaris HiGreen High ROX qPCR Master Mix (Thermo Scientific) in a real-time thermal cycler STEPOnePlus (A&B Applied Biosystems, Waltham, MA) as instructed. Statistical analysis was performed using Anova with a post-hoc Turkey test.

RESULTS

Phenotypic variation in isoprenoid content among Arabidopsis accessions

A set of 116 natural Arabidopsis accessions, originating from various geographical locations, was carefully selected for a detailed analysis of seven isoprenoid compounds (carotenoids, chlorophylls, Dols, phytosterols, plastoquinone, Prens, and tocopherols). Levels of seven selected isoprenoids were quantified in 3-week-old seedlings grown on solid Murashige-Skoog medium. For all analyzed accessions, the same profiles of isoprenoids were observed, however, their level differed remarkably. Thus, for all accessions, one 'family' of Prens composed of 9 to 12 isoprene units (Pren-9 to -12, Pren-10 dominating) and one 'family' of Dols (Dol-15 to -18, Dol-16 dominating) were detected (HPLC/UV, Figure S1A); however, the content of Prens and Dols revealed remarkable variation between accessions (Figure 1A). The highest difference in Pren content was observed for the accessions Est-1 and Uod-7 (20-fold), while in Dol content - for LL-0 and Bur-0 (4-fold). Similar observations were noted for the remaining isoprenoids - although the profile was the same for all accessions (Figure S1B-D), their content revealed substantial differences (Figure S2A-E). For phytosterols - 5-fold (Sav-0 vs. Est-1), for plastoquinone - 25-fold (Mr-0 vs. Er-0), for tocopherols - 8-fold (Lip -0 vs. Edi-0), for carotenoids - 4-fold (Est-1 vs. CS22491) and for chlorophylls - 5-fold (Br-0 vs. CS22491). Detailed analyses revealed considerable differences in the content of 5 out of 7 analyzed compounds (i.e., Prens, Dols, phytosterols, carotenoids, and plastoquinone) between Est-1 and Col-0 (Figure 1A, Figure S2A-E).

Moreover, Est-1 and Col-0 are the parents of the advanced intercross recombinant inbred lines (AI-RILs) mapping population (EstC), which is an excellent resource for QTL analyses due to a large number of fixed recombination events and the density of polymorphisms (Balasubramanian et al., 2009). For these reasons, the EstC population was selected for further analyses in addition to the analysis of the natural accessions.

Phenotypic variation in isoprenoid content in the AI-RIL mapping population

Next, the seven isoprenoid compounds described above (carotenoids, chlorophylls, Dols, phytosterols, plastoquinone, Prens, and tocopherols) were quantified in 146 lines of the EstC mapping population and its parental lines (Col-0 and Est-1). The profiles of analyzed isoprenoids were similar to those described above for different accessions, while the level of particular compounds varied among lines of the mapping population (shown in details in Figure 1B, Figure S3, and Table 1). The range of the content of Prens (Figure 1B), Dols (Figure 1B) and other compounds (Figure S3) was broader than that observed for both parental lines, which might suggest that several loci within the EstC population contribute to this phenomenon and it may be explained by the presence of transgressive segregation.

Estimation of the heritability of isoprenoid levels

To identify the fraction of the observed variation that is genetically determined, we estimated the broad sense heritability (H^2) for each isoprenoid (Table 1) as described in the Material and Methods section. In the AI-RIL population, the broad sense heritability ranged from 0.33 (for Phytosterols) to 0.55 (for Pren and Dol) and 0.57 (for Tocopherols) (Table 1).

Identification of QTLs for the accumulation of Dols, Prens, chlorophylls, and carotenoids

The collected biochemical data for the EstC mapping population were subsequently used to map QTL regions underlying the observed phenotypic variation in isoprenoid accumulation. We were able to map QTLs for four types of compounds (Prens, Dols, chlorophylls, and carotenoids). We detected three QTLs on chromosome 5 for Pren accumulation (Figure S4A) (127.3-133.4 cM, 166.5-170.8 cM, and 171.1-173.3 cM), explaining approximately 33% of the phenotypic variance explained (PVE) by these QTLs containing 948 loci (Table S1). For Dol, we detected a QTL region on chromosome 2 (Figure S4B) (64.8-74.4 cM) containing 308 loci (Table S1), which explains approximately 16.8% of the PVE.

Two QTLs were detected for chlorophyll accumulation on chromosome 2 (160.8-191.6 cM) and 3 (111.6-188.1 cM) (Figure S4C), which together explain 16% of the PVE (Table S1). On chromosome 2, 3, and 5 (159.3-196.5 cM, 131.3-145.6 cM, and 151.3-187.2 cM, respectively) (Figure S4D) we identified three QTLs underlying the variation in carotenoid accumulation, as the whole model explains together almost 24% of the PVE (Table S1). It should be underlined that the QTL on chromosome 3 (for chlorophylls) and the QTL on chromosome 5 (for carotenoids) were included in this analysis even though their LOD scores were below the threshold (below 3) (Figure S4C and Figure S4D, respectively). Interestingly, two of the QTLs identified for chlorophylls and carotenoids, localized on chromosomes 2 and 3, were overlapping.

Our search also revealed two small QTL regions for phytosterols (data not shown); however, they were not analyzed further due to their statistical insignificance ($LOD < 3.0$). Despite the large set of numerical data, no QTLs were identified for plastoquinone or tocopherols. This might indicate that the mapping population used in this study was not appropriate for investigating these metabolites.

Selection of candidate genes from QTL mapping

In order to select and prioritize positional candidate genes from the QTL confidence intervals, we conducted a literature screen and an *in silico* analysis (explained in more detail in the Materials and Methods section) that were based on functional annotations, gene expression data and tissue distribution of the selected genes. We analyzed loci from the Dol-associated QTL (DOL1) and from the three Pren-associated QTLs (PRE1, PRE2, PRE3). We selected the intervals that were characterized by the highest percentage of phenotypic variance related to each QTL and the highest LOD score values linked with the lowest number of loci (Table S1). As a result of the above-described procedure of selection and prioritization, we generated four sets of genes - three for Prens (Table S2) and one for Dol (Table S3).

Within a set of potential candidate genes for Pren (Table S2), there was the AT5G45940 gene encoding the Nudix hydrolase 11 (Kupke, Caparrós-Martín, Malquichagua Salazar, & Culiáñez-Macià, 2009) with putative IPP isomerase activity. For Dol biosynthesis, we identified three loci that might be directly implicated in the process: AT2G17570, encoding a *cis*-prenyltransferase 3 (CPT3), AT2G17370, encoding HMGR2 (hydroxymethylglutaryl Coenzyme-A reductase 2, also called HMG2, a highly regulated enzyme that constitutes a rate-limiting step in the MVA pathway), and AT2G18620, encoding a putative GGPPS2 (geranylgeranyl diphosphate synthase 2). A brief comment on the putative role of the two latter genes in the Dol pathway is presented in Table S3, while an in-depth characteristic of AT2G17570 (CPT3) is presented below.

The role of CPT3 in Dol synthesis in Arabidopsis – genetic and biochemical studies

Remarkably, the CPT responsible for the formation of the hydrocarbon backbone of the major Dols (Dol-15 to Dol-17) accumulated in Arabidopsis has not been identified yet. The AT2G17570 gene encoding CPT3 (sometimes named CPT1 (Kera, Takahashi, Sutoh, Koyama, & Nakayama, 2012)) is ubiquitously expressed

in Arabidopsis organs and, among all nine AtCPTs, it is by sequence homology the closest counterpart of the yeast CPTs that synthesize Dols (Surmacz & Swiezewska, 2011). Preliminary studies revealed that *CPT3*, when co-expressed with *LEW1*, was capable of rescuing the growth defect of a yeast strain devoid of both yeast CPTs: $\rho\epsilon\rho2\Delta \sigma\rho\tau1\Delta$, and a microsomal fraction of thus obtained yeast transformant was able to incorporate *in vitro* a radioactive precursor into polyisoprenoids, although their profile had not been presented (Kwon, Kwon, & Ro, 2016).

At the time, no T-DNA insertion mutant in the *CPT3* gene was available from the NASC collection. For this reason, to analyze *in planta* the involvement of CPT3 in Dol formation, four independent RNAi lines targeting CPT3 for mRNA knockdown (RNAi-1, -12, -14 and -23) and a transgenic line overexpressing CPT3 (OE-7) were generated. The expression level of *CPT3* and the polyisoprenoid content were examined in 4-week-old leaves of these mutants. qRT-PCR analyses revealed that the *CPT3* transcript is significantly reduced (by 40-50%) in the four RNAi lines, and it is nearly 5-fold elevated in the OE line, in comparison to wild-type plants (Figure 2A). No visible phenotypic changes were observed between wild type plants and the studied mutant lines under standard growth conditions (data not shown). In contrast, HPLC/UV analysis of total polyisoprenoids revealed a significant decrease in dolichol (Dol-15 – Dol-17, dominating Dol-16) accumulation in *CPT3* RNAi lines – to approx. 50% of the WT for three lines (RNAi-1, -12, and -23) and to approx. 80% for RNAi-14. Not surprisingly, *CPT3-OE* plants accumulated significantly higher amounts of dolichols, reaching 300% of the WT levels (Figure 2B). These results clearly suggest that CPT3 is involved in the biosynthesis of the major family of Dols in Arabidopsis. In line with this, we observed a positive correlation between the level of *CPT3* transcript and the content of Dol during plant development for three of the selected accessions (Figure 2C). This further supports the role of CPT3 in Dol formation; interestingly, no such correlation was noted for Prens (Figure 2C).

CPT3, similarly to numerous other eukaryotic CPTs engaged in Dol biosynthesis (Grabińska, Park, & Sessa, 2016), is located in the endoplasmic reticulum (ER), as documented by confocal laser microscopy – in transiently transformed *N. benthamiana* leaves the fluorescence signal of CPT3-GFP fully overlapped with that of the ER marker ER-CFP (Figure 2D). Moreover, the physical interaction of CPT3 with Lew1 (Arabidopsis homologue of mammalian NgBR and yeast Nus1, an accessory protein required for activity of some eukaryotic CPTs, Grabińska, Park, & Sessa, 2016) was confirmed *in planta* using a BiFC assay (nEYFP-C1/CPT3 was transiently co-expressed with cEYFP-N1/Lew1 in *N. benthamiana* leaves, Figure 2E) and Y2H system (Figure 2F).

Finally, functional complementation of the yeast mutant *rer2* Δ by Arabidopsis *CPT3* followed by an analysis of the polyisoprenoid profile of transformants (Figure 2G) revealed that solely co-expression of *CPT3* and *LEW1* resulted in the synthesis of the major family of Dols (Dol-14 to Dol-16, Dol-15 dominating, Figure 2G). Moreover, in line with the cellular function of Dol as an obligate cofactor of protein *N*-glycosylation, only simultaneous expression of *CPT3* and *Lew1* fully rescued the defective glycosylation of the marker protein CPY in *rer2* Δ mutant cells (Figure 2H).

Taken together, the genetic and biochemical data presented here clearly show that Arabidopsis CPT3 is a functional ortholog of yRer2 and this verifies the QTL mapping by demonstrating that CPT3 is responsible for Dol synthesis in Arabidopsis. It should be kept in mind however, that further experiments are needed to document that *CPT3* is causal of the natural variation between Col-0 and Est.

Genetic analyses of the variations in metabolite levels in natural accessions - GWAS

As a following step, we used a multi-trait mixed model (Korte et al., 2012) to calculate the genetic correlations between the different traits studied (see Table S4). Here, we found a strong correlation for the four traits – Prens, phytosterols, plastoquinone, and Dols, which argues for a common genetic correlation of these four traits, and at the same time it shows that they have a negative genetic correlation with the remaining three traits, namely tocopherols, chlorophylls, and carotenoids.

Next, we used the mean phenotypic values of the 116 natural Arabidopsis accessions per trait to perform GWAS. Eighty-six of these lines have been recently sequenced as part of the 1,001 genomes project and full

sequence information is readily available (1001 Genomes Consortium, 2016). For the remaining accessions, high-density SNP data have been published earlier (Horton et al., 2012). We used an imputed SNP dataset that combined both sets and has been published earlier (Togninalli et al., 2008). This data set contains ~ 4 million polymorphisms that segregate in the analyzed accessions. Two million polymorphisms, which had a minor allele count of at least 5, were included in the analysis. At a 5% Bonferroni corrected significance threshold of 2.4×10^{-8} , significant associations have been found only for three of the seven different compounds analyzed (Dols, plastoquinone, and phytosterols), while no significant associations have been found for the other four compounds (chlorophylls, carotenoids, Prens, and tocopherols). In summary, 2, 7, and 5 distinct genetic regions were significantly associated with Dols, plastoquinone, and phytosterols, respectively. One region on chromosome 1 is found for all three traits. The respective Manhattan plots are shown in Figure 3 and Figure S5.

Summarizing, we found 4 SNPs, representing two different regions, that were associated with Dol content. The first of the associated polymorphisms is at position 19,545,459 on chromosome 1 and it codes for a non-synonymous AA-exchange (Q270K) in the first exon of AT1G52450, a gene involved in ubiquitin-dependent catabolic processes. The second polymorphism is located at position 19,540,865: it is upstream of AT1G52450 and in the 3' UTR of the neighboring gene AT1G52440, which encodes a putative alpha-beta hydrolase (ABH). A second putative ABH (AT1G52460) is also within 10 kb of these associations. The remaining two significant associations are on chromosome 3 (positions 18,558,714 and 18,558,716, respectively) and they code for one non-synonymous (V113G) and one synonymous substitution in an exon of the gene AT3G50050, which encodes the auxin-related transcription factor Myb77 (Shin et al., 2007).

Moreover, several SNPs were detected for plastoquinone (26) and phytosterols (10) – for details see Figure S5.

The identification of AT1G52450 and two neighboring genes as putative effectors of the accumulation of Dols, plastoquinone, and phytosterols prompted us to analyze the phenotypes of the respective Arabidopsis T-DNA insertion mutants (Figure 4). Interestingly, a significant increase in the content of Dols (approx. 2-fold, comparing to control WT plants) was noted for two analyzed heterozygous AT1G52460-deficient lines: SALK_066806 and GK_823G12. Moreover, in the SALK_066806 line, phytosterol content was also increased (167.8 ± 20.3 vs. 117.4 ± 23.2 $\mu\text{g/g}$ of fresh weight) and plastoquinone content was considerably decreased (27.3 ± 2.0 vs 56.7 ± 5.2 $\mu\text{g/g}$ of fresh weight). It is worth noting that mutations in the AT1G52460 gene did not affect the content of Prens – this gene has not come up as that putatively affecting Pren accumulation (Figure 4). Additionally, these mutant plants developed deformed, curled leaves (Figure S6). Expression analysis of genes of interest in the genetic backgrounds of heterozygous AT1G52460-deficient lines (both SALK_066806 and GK_823G12) revealed that in comparison to WT (Columbia-0) plants, the level of AT1G52460 mRNA was considerably decreased while that of AT1G52440 and AT1G52450 remained unchanged (Figure 4B).

To establish the reason for the inability to obtain homozygous AT1G52460-deficient mutant plants, we analyzed and genotyped the progenies of heterozygous plants originating from the SALK_066806 (n=61) and GK_823G12 (n=151) lines. The lack of AT1G52460-deficient homozygotes among analyzed plants of each mutant line suggested that disruption of this gene was lethal (Table S5). Since the fraction of aborted seeds per silique was higher for both mutants (approx. 17.9% and 25.5% for GK_823G12 and SALK_066806, respectively) than for WT line (2.6%), the seeds produced by mutants showed a reduced germination rate comparing to WT plants (Table S5). It suggests that homozygous mutation in AT1G52460 most probably results in embryo lethality. Other analyzed homozygotic mutants (carrying insertions in the genes AT1G52440 and AT1G52450) did not show significant differences neither in isoprenoid content nor in macroscopical appearance (data not shown).

Taken together, identification of the involvement of putative ABH, encoded by AT1G52460, in Dol biosynthesis sheds new light on metabolic pathway in eukaryotes, although the cellular mechanism underlying this process as well as the causative role of ABH variants in the natural variation of Dol accumulation awaits clarifications.

Correlation analyses of isoprenoid accumulation in the various accessions and in the mapping population- a statistical meta-analysis

As a final step, we conducted a detailed statistical meta-analysis of the studied traits in the different Arabidopsis accessions and in the lines of the EstC mapping population. Numerous correlations were found for the content of seven isoprenoid compounds estimated in the seedlings of natural accessions and the mapping population (Figure 5A and 5B, respectively). Moreover, we clearly identified some outliers (Grubbs test at significance level $\alpha=0.001$) (Grubbs 1950). For plastoquinone, seven values corresponding to three accessions (Er-0, Est-1, and Fei-0) were unequivocally assigned as outliers, for carotenoids – three values corresponding to a single accession (Ren-1), for phytosterols a single outlier was identified in the natural accessions and for Dols in the mapping population (Figure S7). All these outliers, denoted by red triangles in Figure 5, were filtered out in the statistical analysis of metabolite distribution and the correlation analyses (Figure 5A and 5B). For both datasets, the analysis of metabolite correlations revealed the highest correlation for chlorophylls vs. carotenoids ($R>0.97$), while four other metabolites – phytosterols, Prens, plastoquinone, and Dols – also correlated with each other significantly ($p<0.0001$) Table S6. Tocopherol accumulation correlated only occasionally with the other metabolites (Table S6). Based on the structural similarity between Prens and Dols, some level of similarity between the mechanisms of their accumulation might be expected. However, the obtained values for the correlation between Prens and Dols among the tested accessions (0.325, $p=0.0001$) and among the AI-RILs (0.608, $p=0.0001$) suggest differences between these two subgroups of polyisoprenoids. Relationships between levels of metabolites analyzed in this report were also confirmed using hierarchical clustering Figure S8.

Importantly, all the strongest genetic correlations detected for particular metabolites (Table S4) were also identified as the most significant ($p < 0.0001$) for metabolic data-based analysis and this is valid both for the natural accessions and for the EstC mapping population lines (Table S6). Moreover, a consistent trend of correlations (either positive or negative) between individual metabolites in the natural accessions was observed for both genetic- and metabolic-based analysis (Table S4 and Table S6). Taken together, results of the meta-analysis indicate genetic co-regulation of the biosynthesis of specific isoprenoids.

DISCUSSION

Dolichol is a vital component of eukaryotic cell synthesis and accumulation of which is tightly regulated in response to physiological requirements and environmental stimuli. The identification of a long searched for CPT3 makes the biosynthetic route of Dol in plants complete and implicates possible integration of this pathway into Dol-producing biotechnological platforms. Association of a putative alpha-beta hydrolase, encoded by AT1G52460, with Dol accumulation in Arabidopsis provides novel insight into possible determinants of Dol level in all eukaryotes (Figure 6). Understanding of the cellular mechanisms underlying this connection appeals for clarification still it would be difficult to get unrevealed without genetic association approach applied in this study.

It is intriguing that we could detect QTLs for four different compounds: Prens, Dols, chlorophylls, and carotenoids, while we found significant GWAS associations for three: phytosterols, plastoquinone, and Dols. Consequently, Dols are the only compounds where both approaches detected associations. Still, the reported QTL on chromosome 2 does not overlap with the GWAS results, which are located on chromosomes 1 and 3, respectively (summarized in Table S7). While, at a first glimpse, this lack of accordance might be disturbing, there could be many good reasons for it. It is well known that both methods have different power to detect associations (see Figure 4 in Weigel and Nordborg, 2015). For example, on chromosome 1, we identified a significant GWAS association for three different compounds, but we detected no corresponding QTL in the mapping population even though the associated polymorphism segregates in the AI-RIL population. The three traits for which this association is detected (the content of phytosterols, plastoquinone, and Dols) show a strong genetic correlation, so one would expect to find shared genetic factors that regulate all three traits, despite a slightly lesser phenotypic correlation of the traits. The associated sequence variant is located in the gene AT1G52450, which is thus an excellent candidate to modulate all three traits and would not have been found using QTL mapping alone. AT1G52450 is annotated to encode a ubiquitin carboxyl-terminal

hydrolase (UCH)-related protein, while the neighboring gene AT1G52460 encodes an alpha-beta hydrolase, ABH (PubMed Gene database). Neither of these proteins has been characterized yet. Eukaryotic cells usually possess a family of UCHs (e.g., three in Arabidopsis) (Isono and Nagel, 2014) responsible for releasing ubiquitin (Ub) from ubiquitinated proteins. A tight balance between ubiquitination and deubiquitination is required for cellular survival since ubiquitin controls numerous bioactivities, such as protein degradation by the 26S proteasome, cell cycle regulation, signal transduction, or membrane trafficking. In turn, the ABH superfamily proteins are found across all domains of life. They are implicated in primary and secondary metabolism by serving highly diverse enzymatic activities, e.g., as esterases, thioesterases, lipases, proteases. Additionally, proteins with the α/β hydrolase fold function as receptors in the strigolactone, gibberellin and karrikin-smoke response pathways (Mindrebo, Nartey, Seto, Burkart, & Noel, 2016 and references therein). In Arabidopsis, more than 600 proteins with ABH folds have been predicted by the InterPro database (Mitchell et al., 2019) with the majority remaining uncharacterized.

Taken together, hydrolytic enzymes, as ABH, encoded by AT1G52460, and/or UCH, encoded by AT1G52450, might control isoprenoid biosynthesis in eukaryotic cells. Interestingly, both *ABH* and *UCH* show a high dN/dS ratio (ratio of nonsynonymous to synonymous divergence) in the Arabidopsis population, arguing for strong selection on these genes (see Table S8). Further studies are needed to identify the cellular target(s) of AT1G52460 and the mechanisms underlying its involvement in the metabolism of Dol, phytosterol, and plastoquinone.

It is worth noting that in previous reports, the AT1G52460 gene was identified as one of the maternally expressed imprinted genes (MEGs) that was shown to be predominantly expressed from maternal alleles in reciprocal crosses (Wolff et al., 2011). Notably, the AT1G52460 was among the MEGs (30% of all the MEGs tested in that study) for which authors reported a dN/dS value greater than one (Wolff et al., 2011). The dN/dS value can be used to measure the rate of molecular evolution of genes (Warren et al., 2010); therefore, the results of Wolff et al. (2011) provide particularly strong evidence for the fast evolution of AT1G52460. Taking into account that we detected only heterozygotic lines for the AT1G52460 gene, we consider that a loss-of-function allele may lead to a lethal phenotype. A 2:1 ratio (the frequency of heterozygous:WT plants in F₂) fitted the data ($\chi^2=2.6$ and $\chi^2=0.2$ for GK_823G12 and SALK_066806 lines, respectively, at the value of $p>0.05$). This finding could be particularly important, and it deserves further investigation since very few imprinted genes have been confirmed in plants and even fewer of them have been functionally investigated (He et al., 2017).

The confidence intervals of the detected QTLs include hundreds of different genes. This is within the typical mapping resolution of QTL studies but leads to the problem of prioritizing candidate genes. The most promising gene identified in the QTL analysis, AT2G17570 (*CPT3*), is a long-searched enzyme responsible for backbone synthesis for the major family of dolichols in Arabidopsis, with Dol-15 and Dol-16 dominating. Interestingly, the different product specificity of the Arabidopsis enzymes CPT3, CPT6 (which produces *in planta* a single Dol-7 (Surmacz, Plochocka, Kania, Danikiewicz, & Swiezewska, 2014)) and the recently characterized CPT1 (producing a family of Dols with Dol-21 dominating (Surowiecki, Onysk, Manko, Swiezewska, & Surmacz, 2019)) suggests that the particular AtCPTs play dedicated, non-redundant roles in isoprenoid synthesis in Arabidopsis tissues. For further comments regarding CPT3 see also Table S8.

Even though no overlapping associations have been found for the GWAS and QTL results, one can try, using the GWAS results, to prioritize candidate genes in the QTL interval. In the confidence interval of the detected QTL for Dol on chromosome 2, we could analyze 6,668 independent segregating polymorphisms with a minor allele frequency greater than 5%. None of these reached the genome-wide significance threshold; the most significant polymorphism had a p-value of 4.88×10^{-6} and was located in the proximity of AT2G17570, which encodes CPT3. Although this score is marginal, it is locally significant, if we restrict our analysis to sequence variants within the QTL region. So, the combined results of GWAS and QTL strongly indicate that *CPT3* is the gene underlying the detected QTL for Dol, despite the plethora of other tempting candidate genes. Detailed SNP analyses of *CPT3* revealed that this gene shows a high amount of variation with a total number of 30 non-synonymous substitutions and 5 alternative starts and 1 premature stop codon in

the *Arabidopsis* population (Table S8).

It is worth underlying that both genetic- and metabolic-based analysis revealed correlations of the analyzed traits indicating genetic co-regulation of the biosynthesis of specific isoprenoids.

In summary, several candidate genes for potential new factors that might affect polyisoprenoid accumulation have been identified in this study. The regulation of isoprenoid pathways is complex but using the combination of both GWAS and QTL it is possible to prioritize the underlying genes. Genetic and biochemical evidences described in this report document the role of *CPT3* and *ABH* in Dol pathway (Figure 6) still further studies are needed to prove their causal role in the natural variation of this trait.

Last but not least, it should be kept in mind that this study is based on terpene levels at the seedling stage and might not be representative for later growth stages. Anyhow, obtained results clearly suggest the role of *CPT3* and *ABH* in Dol accumulation.

Understanding of the mechanisms of Dol synthesis/accumulation in eukaryotes is important since the shortage of dolichol/dolichyl phosphate results in serious defects in all studied organisms, most probably caused by defective protein glycosylation. In plants, it is lethal due to male sterility (Jozwiak et al., 2015; Lindner et al., 2015) while in humans mutations in genes encoding enzymes involved in Dol/DolP synthesis lead to rare genetic disorders collectively called Congenital Disorders of Glycosylation (CDG type I); supplementation of the diet with plant tissues that can be utilized as a source of dolichol/dolichyl phosphate has been suggested (summarized in Buczkowska, Swiezewska, & Lefeber, 2015). The identification of genes involved in the synthesis/accumulation of Dols – such as the here detected *CPT3* and *ABH* – opens a perspective for the manipulation of Dol content in plants and consequently makes it feasible to think of constructing plants with increased Dol content. Moreover, involvement of *ABH* in Dol synthesis in *Arabidopsis* might also suggest analogous role of *ABH* in mammalian cells indicating new potential therapeutic strategy for CDG patients.

ACKNOWLEDGEMENTS

This research was supported by grants from the National Science Centre of Poland [UMO-2014/15/N/NZ3/04316] (KG), [UMO-2018/29/B/NZ3/01033] (ES), and [UMO-2014/15/B/NZ2/01073] (AI), and the National Research Foundation (NRF) of Korea [NRF- 2017R1A2B3009624] (JHA). We would like to express our gratitude to Professor Maarten Koornneef for providing the AI-RILs seeds used in this study. We also would like to thank Dr Agata Lipko for initial characterization of mutant lines. Dr Marta Hoffinan-Sommer is kindly acknowledged for help with preparation of the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

All data obtained and/or analyzed in this study are available from the corresponding authors upon reasonable request. The phenotypic data used for GWAS are available at the AraPheno, <https://arapheno.1001genomes.org/>. GWAS script: the R scripts used are available at <https://github.com/arthurkorte/GWAS>. The genotype data used for GWAS are available via the AraGWAS Catalog, <https://aragwas.1001genomes.org/#/download-center>

Supporting data for this article is available online.

ORCID

Anna Ihnatowicz <https://orcid.org/0000-0001-7883-5954>

REFERENCES

1001 Genomes Consortium. (2016) 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*166: 481–491

- Akhtar TA, Surowiecki P, Siekierska H, Kania M, Van Gelder K, Rea K, Virta L, Vatta M, Gawarecka K, Wojcik J, Danikiewicz W, Buszewicz D, Swiezewska E, Surmacz L (2017) Polyprenols are synthesized by a plastidial *cis*-prenyltransferase and impact photosynthetic performance in *Arabidopsis thaliana*. *Plant Cell* 29: 1709–1725
- Alonso-Blanco C, Aarts L, Bentsink MGM, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M (2009) What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* 21: 1877–1896
- Arends D, Prins P, Jansen RC, Broman KW (2010) R/qtl: high-throughput multiple QTL mapping. *Bioinformatics* 26: 2990–2992
- Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO, Chory J, Weigel D (2009) QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. *PLoS One* 4: e4318 1–8
- Beck G, Coman D, Herren E, Ruiz-Sola MA, Rodríguez-Concepción M, Gruissem W, Vranová E (2013) Characterization of the GGPP synthase gene family in *Arabidopsis thaliana*. *Plant Molecular Biology* 82: 393–416
- Bohlmann J, Keeling CI (2008) Terpenoid biomaterials. *Plant J* 54: 656–669
- Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics* 6: e1000940
- Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889–890
- Broman KW (2008) New functions for exploring multiple-QTL models.
<http://www.rqtl.org/tutorials/new-multiqtl.pdf>
- Buczowska A, Swiezewska E, Lefeber DJ (2015) Genetic defects in dolichol metabolism. *J. Inher. Metabolic Disease* 38: 157–169
- Cunillera N, Arrò M, Fors O, Manzano D, Ferrer A (2000) Characterization of dehydrodolichyl diphosphate synthase of *Arabidopsis thaliana*, a key enzyme in dolichol biosynthesis. *FEBS Lett* . 477: 170–174
- Fusari CM, Kooke R, Lauxmann MA, Annunziata MG, Enke B, Hoehne M, Krohn N, Becker FFM, Schlereth A, Sulpice R, Stitt M, Keurentjes JJB (2017) Genome-Wide Association mapping reveals that specific and pleiotropic regulatory mechanisms fine-tune central metabolism and growth in *Arabidopsis*. *Plant Cell* 29: 2349–2373
- Gawarecka K, Swiezewska E (2014) Analysis of plant polyisoprenoids. *Meth Molec Biol* 1153: 135 – 147
- Grafińska KA, Park EJ, Sessa WC (2016) *cis*-Prenyltransferase: new insights into protein glycosylation, rubber synthesis, and human diseases. *J Biol Chem* 291: 18582–18590
- Gregory R, Warnes BB, Bonebakker L, Gentleman R, Huber A, Liaw W, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B (2016) gplots: Various R programming tools for plotting data. <https://cran.r-project.org/web/packages/gplots/gplots.pdf>
- Grubbs FE (1950) Sample criteria for testing outlying observations. *Annals Math Statistics* 21: 27–58
- Hallahan D, Keiper-Hrynko NM (2006) USA United States Patent No 7205456
- Han K, Lee HY, Ro NY, Hur OS, Lee JH, Kwon JK, Kang BC (2018) QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. *Plant Biotech J* 16: 1546–1558

- He S, Sun Y, Yang Q, Zhang X, Huang Q, Zhao P, Sun M, Liu J, Qian W, Qin G, Gu H, Qu LJ (2017) A novel imprinted gene NUWA controls mitochondrial function in early seed development in Arabidopsis. *PLoS Genet* 13: e1006553
- Hemmerlin A, Harwood JL, Bach TJ (2012) A raison d'etere for two distinct pathways in the early steps of plant isoprenoids biosynthesis? *Prog Lipid Res* 51: 95–148
- Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muiyati NW, Platt A, Sperone FG, Vilhjálmsson B., Nordborg M, Borevitz JO, Bergelson J (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet* 44: 212–216
- Isono E, Nagel MK (2014) Deubiquitylating enzymes and their emerging role in plant biology. *Front Plant Sci* 5: 56
- Jozwiak A, Gutkowska M, Gawarecka K, Surmacz L, Buczkowska A, Lichocka M, Nowakowska J, Swiezewska E (2015) POLYPRENOL REDUCTASE2 deficiency is lethal in Arabidopsis due to male sterility. *Plant Cell* 27: 3336–3353
- Jozwiak A, Lipko A, Kania M, Danikiewicz W, Surmacz L, Witek A, Wojcik J, Zdanowski K, Paczkowski C, Chojnacki T, Poznanski J, Swiezewska E (2017) Modelling of dolichol mass spectra isotopic envelopes as a tool to monitor isoprenoid biosynthesis. *Plant Physiol* 174: 857–874
- Jozwiak A, Ples M, Skorupinska-Tudek K, Kania M, Dydak M, Danikiewicz W, Swiezewska E (2013) Sugar availability modulates polyisoprenoid and phytosterol profiles in *Arabidopsis thaliana* hairy root culture. *Biochim Biophys Acta* 1831: 438–447
- Kadioglu Y, Demirkaya F, Demirkaya AK (2009) Quantitative determination of underivatized α -tocopherol in cow milk, vitamin and multivitamin drugs by GC-FID. *Chromatographia* 70: 665–670
- Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42: 348–354
- Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, Chen H, Dubin M, Lee C-R, Wang C, Bemm F, Becker C, O'Neil R, O'Malley RC, Quarless DX (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* 166: 492–505
- Kera K, Takahashi S, Sutoh T, Koyama T, Nakayama T (2012) Identification and characterization of a *cis*, *trans*-mixed heptaprenyl diphosphate synthase from *Arabidopsis thaliana*. *FEBS J* 279: 3813–3827
- Keurentjes JJB, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, van der Plas LH, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. *Nat Genet* 38: 842–849
- Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics* 159: 359–370
- Komsta L (2011) Package 'outliers', Tests for outliers, R package version 0.14. <https://cran.r-project.org/web/packages/outliers/outliers.pdf>
- Kooke R, Kruijer W, Bours R, Becker F, Kuhn A, van de Geest H, Buntjer J, Doeswijk T, Guerra J, Bouwmeester H, Vreugdenhil D, Keurentjes JJB (2016) Genome-Wide Association mapping and genomic prediction elucidate the genetic architecture of morphological traits in Arabidopsis. *Plant Physiol* 170: 2187–2203
- Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu Rev Plant Biol* 55: 141–172

- Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet* 44: 1066–1071
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9: 29
- Kramer U (2015) Planting molecular functions in an ecological context with *Arabidopsis thaliana*. The natural history of model organisms. *eLIFE* 4: 1–13
- Kupke T, Caparrós-Martín JA, Malquichagua Salazar KJ, Culiñez-Macià FA (2009) Biochemical and physiological characterization of *Arabidopsis thaliana* AtCoAse: a Nudix CoA hydrolyzing protein that improves plant development. *Physiol Plant* 135: 365–378
- Kwon M, Kwon EJ, Ro DK (2016) *cis*-Prenyltransferase and polymer analysis from a natural rubber perspective. *Meth Enzymol* 576: 121–145
- Lichtenthaler HK, Buschmann C (2001) Chlorophylls and carotenoids: measurement and characterization by UV/VIS spectroscopy. *Curr Prot Food Anal Chem* F4.3.1–F4.3.8
- Lindner H, Kessler SA, Müller LM, Shimosato-Asano H, Boisson-Dernier A, Grossniklaus U (2015) TURAN and EVAN mediate pollen tube reception in *Arabidopsis* synergids through protein glycosylation. *PLoS Biol* 13: e1002139
- Lipko A, Swiezewska E (2016) Isoprenoid generating systems in plants - A handy toolbox how to assess contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthetic process. *Prog Lipid Res* 63: 70–92
- Lisek J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Torjek O, Selbig J, Altmann T, Willmitzer L (2008) Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J* 53: 960–972
- Lovell JT, Mullen JL, Lowry DB, Awole K, Richards JH, Sen S, Verslues PE, Juenger TE, McKay JK (2015) Exploiting differential gene expression and epistasis to discover candidate genes for drought-associated QTLs in *Arabidopsis thaliana*. *Plant Cell* 27: 969–983
- Lu Q, Tang X, Tian G, Wang F, Liu K, Nguyen V, Kohalmi SE, Keller WA, Tsang EW, Harada JJ, Rothstein SJ, Cui Y (2010) *Arabidopsis* homolog of the yeast TREX-2 mRNA export complex: components and anchoring nucleoporin. *Plant J* 61: 259–270
- Mackay TFC (2001) The genetic architecture of quantitative traits. *Annu Rev Genet* 35: 303–339
- Meyer RC, Steinfath M, Lisek J, Becher M, Witucka-Wall H, Törjék O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, Altmann T (2007) The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 104: 4759–4764
- Mindrebo JT, Nartey CM, Seto Y, Burkart MD, Noel JP (2016) Unveiling the functional diversity of the Alpha-Beta hydrolase fold in plants. *Curr Opin Struct Biol* 41: 233–246
- Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI, M.I., Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong S-Y, Finn RD (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 47(D1): D351–D360
- Mitchell-Olds T, Pedersen D (1998) The molecular basis of quantitative genetic variation in central and secondary metabolism in *Arabidopsis*. *Genetics* 149: 739–747

- Oh SK, Han KH, Ryu SB, Kang H (2000) Molecular cloning, expression, and functional analysis of a *cis*-prenyltransferase from *Arabidopsis thaliana*. *J Biol Chem* 275: 18482–18488
- Rezanka T, Votruba J (2001) Chromatography of long chain alcohols (polyprenols) from animal and plant sources. *J Chromatogr A* 936: 95–110
- Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant Cell* 20: 1199–1216
- Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M (2012) An efficient multi-locus mixed model approach for genome-wide association studies in structured populations. *Nat Genet* 44: 825–830
- Seren U, Grimm D, Fitz J, Weigel D, Nordborg M, Borgwardt K, Korte A (2016) AraPheno: a public database for *Arabidopsis thaliana* phenotypes. *Nucleic Acids Res* 45: D1054–D1059
- Sergeeva LI, Vonk J, Keurentjes JJB, van der Plas LH, Koornneef M, Vreugdenhil D (2004) Histochemical analysis reveals organ-specific quantitative trait loci for enzyme activities in *Arabidopsis*. *Plant Physiol* 134: 237–245
- Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52: 591–611
- Shin R, Burch AY, Huppert KA, Tiwari SB, Murphy AS, Guilfoyle TJ, Schachtman DP (2007) The *Arabidopsis* transcription factor MYB77 modulates auxin signal transduction. *Plant Cell* 19: 2440–2453
- Siwinska J, Kadzinski L, Banasiuk R, Gwizdek-Wisniewska A, Olry A, Banecki B, Lojkowska E, Ihnatowicz A (2014) Identification of QTLs affecting scopolin and scopoletin biosynthesis in *Arabidopsis thaliana*. *BMC Plant Biol* 14: 280–294
- Skorupinska-Tudek K, Bienkowski T, Olszowska O, Furmanowa M, Chojnacki T, Danikiewicz W, Swiezewska E (2003) Divergent pattern of polyisoprenoid alcohols in the tissues of *Coluria geoides*: A new electrospray ionization MS approach. *Lipids* 38: 981–990
- Skorupinska-Tudek K, Poznanski J, Wojcik J, Bienkowski T, Szostkiewicz I, Zelman-Femiak M, Bajda A, Chojnacki T, Olszowska O, Grunler J, Meyer O, Rohmer M, Danikiewicz W, Swiezewska E (2008) Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of dolichols in plants. *J Biol Chem* 283: 21024–21035
- Spearman C (1904) The proof and measurement of association between two things. *Am J Psychol* 15: 72–101
- Surmacz L, Swiezewska E (2011) Polyisoprenoids - Secondary metabolites or physiologically important superlipids? *Biochem Biophys Res Commun* 407: 627–632
- Surmacz L, Plochocka D, Kania M, Danikiewicz W, Swiezewska E (2014) *cis*-Prenyltransferase AtCPT6 produces a family of very short-chain polyisoprenoids in planta. *Biochim Biophys Acta* 1841: 240–250
- Surowiecki P, Onysk A, Manko K, Swiezewska E, Surmacz L (2019) Long-chain polyisoprenoids are synthesized by AtCPT1 in *Arabidopsis thaliana*. *Molecules* 24: 2789
- Suzuki M, Nakagawa S, Kamide Y, Kobayashi K, Ohyama K, Hashinokuchi H, Kiuchi R, Saito K, Muranaka T, Nagata N (2009) Complete blockage of the mevalonate pathway results in male gametophyte lethality. *J Exp Bot* 60: 2055–2064
- Swiezewska E, Danikiewicz W (2005) Polyisoprenoids: structure, biosynthesis and function. *Prog Lipid Res* 44: 235–258
- Tholl D (2015) Biosynthesis and biological functions of terpenoids in plants. *Adv Biochem Eng Biotechnol* 148: 63–106

Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E (2005) Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers. *Plant J* 42: 757–771

Togninalli M, Seren Ü, Meng D, Fitz J, Nordborg M, Weigel D, Borgwardt K, Korte A, Grimm DG (2018) The AraGWAS Catalog: a curated and standardized *Arabidopsis thaliana* GWAS catalog. *Nucleic Acids Res* 46: D1150–D1156

Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Amer Statist Assoc* 58: 236–244

Weigel D, Glazebrook J *Arabidopsis: a laboratory manual*. Cold Spring Harbor: New York, NY, USA, 2002; pp. 119–141

Weigel D, Mott R (2009) The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biol* 10: 107

Weigel D, Nordborg M (2015) Population genomics for understanding adaptation in wild plant species. *Annu Rev Genet* 49: 315–338

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MTA, Spillane C, Nordborg M, Rehmsmeier M, Köhler C (2011) High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet* 7 : e1002126

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *The Plant Genome* 1: 5–20

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet* 3: e4

Table 1. Isoprenoid content: parental values, ranges, and heritabilities in the AI-RIL mapping population (see Materials and Methods).

Isoprenoid compound

values [$\mu\text{g/gFW}$]

Parental lines

Parental lines

AI-RILs

AI-RILs

AI-RILs

Col-0

Est-1

Range

Median (quartiles)

Heritability^a

Prenols

116 \pm 10

179 \pm 5

60 – 209

129 (104; 153)

0.55

Dolichols

0.9 ± 0.1

1.6 ± 0.5

$0.7 - 2.0$

1.1 (0.9; 1.2)

0.55

Chlorophylls

503 ± 29

250 ± 8

$222 - 604$

392 (349; 441)

0.42

Carotenoids

125 ± 18

75 ± 8

$57 - 140$

94 (84; 104)

0.43

Phytosterols

98 ± 11

125 ± 3

$74 - 154$

107 (97; 117)

0.33

Plastoquinone

99 ± 12

148 ± 12

$50 - 176$

111 (97; 127)

0.47

Tocopherols

138 ± 34

226 ± 36

76 – 288

142 (121; 163)

0.57

^aMeasure of total phenotypic variance attributable to genetic differences among genotypes (broad sense heritability) calculated as $V_G/(V_G+V_E)$.

^aMeasure of total phenotypic variance attributable to genetic differences among genotypes (broad sense heritability) calculated as $V_G/(V_G+V_E)$.

^aMeasure of total phenotypic variance attributable to genetic differences among genotypes (broad sense heritability) calculated as $V_G/(V_G+V_E)$.

^aMeasure of total phenotypic variance attributable to genetic differences among genotypes (broad sense heritability) calculated as $V_G/(V_G+V_E)$.

^aMeasure of total phenotypic variance attributable to genetic differences among genotypes (broad sense heritability) calculated as $V_G/(V_G+V_E)$.

^aMeasure of total phenotypic variance attributable to genetic differences among genotypes (broad sense heritability) calculated as $V_G/(V_G+V_E)$.

FIGURE LEGENDS

FIGURE 1. Polyisoprenoid lipids of *Arabidopsis thaliana*.

A Content of polyprenols (Pren) and dolichols (Dol) in Arabidopsis accessions . Bars representing Col-0 and Est-1 are marked in red. Shown are means \pm SD (n=3). Content of other isoprenoids (chlorophylls, carotenoids, phytosterols, plastoquinone and tocopherols) in the seedlings of Arabidopsis accessions are given in Figure S2.

B Frequency distribution of the content of polyprenols and dolichols in the seedlings of AIRILs and their parental lines, Col-0 and Est-1 . Each bar covers the indicated range of a particular isoprenoid compound. Frequency distribution of the content of other isoprenoids (chlorophylls, carotenoids, phytosterols, plastoquinone and tocopherols) are given in Figure S3. Structures of polyprenol (Pren) and dolichol (Dol) are shown in the inset: t and c stand for the number of internal isoprene units in *trans* and *cis* configuration, respectively. The α - and ω -terminal isoprene units are depicted.

FIGURE 2. Role of CPT3 in Dol biosynthesis – studies *in planta* and functional assay in yeast.

A, B Relative expression of *CPT3* and content of dolichols (Dol-15 to Dol-17) in the leaves of 4-week-old Arabidopsis plants, measured for wild-type Col-0, four independent RNAi lines targeting *CPT3* (RNAi-1, -12, -14 and -23) and a *CPT3* -overexpressing line (OE-7). The results are means (\pm SD) of three independent experiments. Asterisks indicate statistically significant differences between WT and mutant plants (* 0.01<P<0.05, ** 0.001<P<0.01, *** P<0.001, Student's t-test).

C Changes in the levels of *CPT3* mRNA (black curves) and in the content of Dols and Prens (respectively, orange and green bars) in the tissues of three Arabidopsis accessions: Col-0, Or-0, and Stw-0, during the plant life-span. Transcript levels and lipid content were estimated in Arabidopsis seedlings (1-3 weeks, bright colors) and leaves (4-6 weeks, deep colors), shown are means \pm SD (n=3). Please note that the content of Prens is rescaled (0.01 multiples are presented) due to their high cellular level.

D Co-localization of fluorescence signals of CPT3-GFP (green) and ER-CFP (compartmental marker, blue) upon transient expression in *Nicotiana benthamiana* leaves. Bar: 10 μ m.

E Analysis of CPT3 and Lew1 protein–protein interaction using split yellow fluorescent protein (YFP) BiFC assay in tobacco leaf cells. Shown is the co-localization of fluorescence signals from the CPT1/Lew1 complex (green) and from the compartmental marker ER-CFP (blue). Bar: 10 μ m. See Materials and methods.

F Analysis of CPT3 and Lew1 interaction using yeast two-hybrid system.

G, H Polyisoprenoid profiles and glycosylation status of CPY analyzed for *rer2* Δ transformed with empty vector, *CPT3*, *LEW1*, and *CPT3/LEW1*. Representative HPLC/UV chromatograms are shown. The positions of mature CPY (mCPY) and its hypoglycosylated glycoforms (lacking between one and four *N*-linked glycans, -1 to -4) are indicated. See Materials and Methods.

FIGURE 3. Manhattan plot of genome-wide associations for phytosterols, plastoquinone, and dolichols. The dotted horizontal lines indicate a significance level of 0.05 after Bonferroni correction for multiple testing (see Materials and Methods). Manhattan plot of genome-wide association results for polyprenols, chlorophylls and tocopherols are shown in Figure S5.

FIGURE 4. Effect of AT1G52460 deficiency on the level of transcripts of neighboring genes and the content of polyisoprenoids.

A AT1G52460 gene structure. Exons and introns are indicated by thick and thin lines, respectively. The T-DNA insertion sites in two independent mutant lines: SALK_066806 and GK_823G12 are depicted.

B Levels of AT1G52440, AT1G52450, and AT1G52460 transcripts (qPCR analysis) in the leaves of 3-week-old plants – WT (Col) and AT1G52460-deficient plants were compared. P[?] \leq 0.0001 (one-way ANOVA followed by Turkey post-tests); ns, non-significant.

C The content of Dols and Prens estimated in leaves of 3-week-old plants using HPLC/UV, shown are means \pm SD of five independent biological replicates. The phenotypic appearance of 4-week-old detached leaves of AT1G52460-deficient line (SALK_066806, *abh* heterozygous mutant) and wild-type (Col-0) plants grown in soil is shown in Figure S6. Seed germination and segregation analysis of F1 progeny of self-pollinated of heterozygous SALK_066806 and GK_823G12 lines is shown in Table S5.

FIGURE 5. Correlations between the content of seven metabolites estimated in the seedlings of Arabidopsis accessions (A) and the EstC mapping population (B). The original distributions (green bars), together with the approximation of the normal distribution of the data (blue curve) with outliers removed, are presented on the diagonal. Correlation patterns for each metabolite pair are presented at the appropriate intersection; please note that outliers (red dots) were not taken into consideration for the analysis. Above each diagonal panel, the Shapiro-Wilk statistics (W, p) for normal distribution is presented, while for out-of-diagonal panels Pearson (P) and Spearman (S) correlation coefficients together with the associated significance levels are shown (please note that ‘0’ means $p < 1e-7$). Bearing in mind the statistically significant deviations from normal distribution shown in the diagonal panels, the significance of the observed correlations should be interpreted in terms of the Spearman rather than Pearson coefficient (see Materials and methods). Cumulative distribution functions (CDF) of the content of seven studied metabolites analyzed in the seedlings of Arabidopsis accessions and AI-RILs are shown in Figure S7.

FIGURE 6. Biosynthetic routes leading to isoprenoids in a plant cell; the involvement of the genes *cis*-prenyltransferase 3 (*CPT3*) and putative role of alpha-beta hydrolase (*ABH*) is indicated. Depicted are seven metabolites analyzed in this study. Two pathways, the mevalonate (MVA) and methylerythritol phosphate (MEP) pathways, are generating IPP in parallel, both contributing to particular isoprenoids (Hemmerlin, Harwood, & Bach, 2012; Akhtar et al., 2017; Jozwiak et al., 2017). Blue arrows illustrate the exchange of intermediates between the MVA and MEP pathways. Abbreviations: DMAPP-dimethylallyl diphosphate, FPP-farnesyl diphosphate, GPP-geranyl diphosphate, GGPP-geranylgeranyl diphosphate, IPP-isopentenyl diphosphate.

Hosted file

Figure 1_Gawarecka et al. 2021_PC&E_RESUBMISSION.docx available at <https://authorea.com/users/436897/articles/538910-gwas-analysis-combined-with-qt1-mapping-identify-cpt3-and-abh-as-genes-underlying-dolichol-accumulation-in-arabidopsis>

Hosted file

Figure 2_Gawarecka et al. 2021_PC&E_RESUBMISSION.docx available at <https://authorea.com/users/436897/articles/538910-gwas-analysis-combined-with-qt1-mapping-identify-cpt3-and-abh-as-genes-underlying-dolichol-accumulation-in-arabidopsis>

Hosted file

Figure 3_Gawarecka et al. 2021_PC&E_RESUBMISSION.docx available at <https://authorea.com/users/436897/articles/538910-gwas-analysis-combined-with-qt1-mapping-identify-cpt3-and-abh-as-genes-underlying-dolichol-accumulation-in-arabidopsis>

Hosted file

Figure 4_Gawarecka et al. 2021_PC&E_RESUBMISSION.docx available at <https://authorea.com/users/436897/articles/538910-gwas-analysis-combined-with-qt1-mapping-identify-cpt3-and-abh-as-genes-underlying-dolichol-accumulation-in-arabidopsis>

Hosted file

Figure 5_Gawarecka et al. 2021_PC&E_RESUBMISSION.docx available at <https://authorea.com/users/436897/articles/538910-gwas-analysis-combined-with-qt1-mapping-identify-cpt3-and-abh-as-genes-underlying-dolichol-accumulation-in-arabidopsis>

Hosted file

Figure 6_Gawarecka et al. 2021_PC&E_RESUBMISSION.docx available at <https://authorea.com/users/436897/articles/538910-gwas-analysis-combined-with-qt1-mapping-identify-cpt3-and-abh-as-genes-underlying-dolichol-accumulation-in-arabidopsis>