

Neofunctionalization of Mitochondrial Proteins and Incorporation into Signaling Networks in Plants

Sbatie Lama,¹ Martyna Broda,² Zahra Abbas,² Dries Vanechoutte,^{3,4} Katharina Belt,^{†,2} Torbjörn Säll,¹ Klaas Vandepoele,^{3,4} and Olivier Van Aken^{*,1}

¹Department of Biology, Lund University, Lund, Sweden

²ARC Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley, Australia

³Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

⁴VIB Center for Plant Systems Biology, Ghent, Belgium

[†]Present address: CSIRO, Floreat, WA, Australia

*Corresponding author: E-mail: olivier.van_aken@biol.lu.se.

Associate editor: Michael Purugganan

Abstract

Because of their symbiotic origin, many mitochondrial proteins are well conserved across eukaryotic kingdoms. It is however less obvious how specific lineages have obtained novel nuclear-encoded mitochondrial proteins. Here, we report a case of mitochondrial neofunctionalization in plants. Phylogenetic analysis of genes containing the Domain of Unknown Function 295 (DUF295) revealed that the domain likely originated in Angiosperms. The C-terminal DUF295 domain is usually accompanied by an N-terminal F-box domain, involved in ubiquitin ligation via binding with ASK1/SKP1-type proteins. Due to gene duplication, the gene family has expanded rapidly, with 94 DUF295-related genes in *Arabidopsis thaliana* alone. Two DUF295 family subgroups have uniquely evolved and quickly expanded within Brassicaceae. One of these subgroups has completely lost the F-box, but instead obtained strongly predicted mitochondrial targeting peptides. We show that several representatives of this DUF295 *Organellar* group are effectively targeted to plant mitochondria and chloroplasts. Furthermore, many DUF295 *Organellar* genes are induced by mitochondrial dysfunction, whereas F-Box DUF295 genes are not. In agreement, several Brassicaceae-specific DUF295 *Organellar* genes were incorporated in the evolutionary much older ANAC017-dependent mitochondrial retrograde signaling pathway. Finally, a representative set of DUF295 T-DNA insertion mutants was created. No obvious aberrant phenotypes during normal growth and mitochondrial dysfunction were observed, most likely due to the large extent of gene duplication and redundancy. Overall, this study provides insight into how novel mitochondrial proteins can be created via “intercompartmental” gene duplication events. Moreover, our analysis shows that these newly evolved genes can then be specifically integrated into relevant, pre-existing coexpression networks.

Key words: mitochondria, evolution, retrograde signaling, stress, neofunctionalization.

Introduction

Mitochondria are membrane-bound organelles with crucial roles in eukaryotic cells, including ATP production, Fe–S cluster synthesis, the Krebs cycle, and many other metabolic functions. The endosymbiotic theory proposes that mitochondria are derived from ancestral bacteria that were engulfed and retained by a host cell, probably of archaeal origin (Spang et al. 2015). The exact lineage of bacteria that was the precursor to mitochondria is not known, but they are likely related to (alpha)proteobacteria (Martijn et al. 2018). During evolution, the genecontent of the mitochondrial genome was heavily reduced, and the vast majority of mitochondrial proteins are encoded in the nuclear genome (Ku et al. 2015). This required the mitochondrial import machinery to evolve, allowing proteins translated in the cytosol to be imported into the different compartments of the mitochondria, often with help of specific targeting peptides. Other originally bacterial functions were not redirected to

mitochondria and became operational elsewhere in the cell, for example, the cytosol and peroxisomes (Huynen et al. 2013), or were lost entirely.

Mitochondria usually contain >1,000 different proteins, for example, around 1,800 proteins in mammals (Palmfeldt and Bross 2017), and perhaps >2,000 in plant mitochondria (Rao et al. 2017). As the mitochondria-containing host cell was probably the ancestor for all eukaryotic lineages including plants, animal, and fungi, one would expect the majority of mitochondrial proteins to be of clear bacterial origin. However, of the ~800 human nuclear genes that bear clear resemblance to alpha-proteobacterial genes, only about 200 are present in the human mitochondrial proteome (Szklarczyk and Huynen 2010). The current view is that the mitochondrial proteome is a mixture of alpha-proteobacteria-related proteins, proteins from other (proteo-)bacteria obtained via lateral gene transfer, and viral proteins. Additionally, about 40% of the mitochondrial proteome has no clear viral or bacterial origin. These proteins are thought to be of premitochondrial host cell origin

or are of “lineage-specific” origin (e.g., plant specific), having originated after the Last Eukaryotic Common Ancestor (Roger et al. 2017). The proteome of the plastid (derived from a photosynthetic cyanobacterial endosymbiont) appears to have a similarly complex origin (Leister 2016; Roger et al. 2017).

For the multitude of mitochondrial proteins that are related to bacterial, viral, or archaeal proteins, different mechanisms including endosymbiotic gene transfer (from the endosymbiont to the nucleus) or lateral gene transfer can be envisaged. It is, however, less evident how different eukaryotic lineages have obtained lineage-specific mitochondrial (or plastid) protein types. One possibility is via random creation of new open reading frames that encode completely novel proteins. Another mechanism may be gene duplication, whereby a new copy of a gene is created in the genome. In most cases, the encoded protein retains its original subcellular localization (intracompartamental duplication) (Szklarczyk and Huynen 2009). However, in rare cases, the duplication results in one of the encoded proteins becoming targeted to another subcellular location (intercompartmental duplication).

Another consequence of the endosymbiotic nature of the eukaryotic cell is the need for more complex transcriptional regulation. As most of the genes encoding mitochondrial or plastid proteins are found in the nuclear genome, the individual mitochondrial or plastid genomes can no longer directly control all transcript levels. Instead, the organelles must provide feedback to the nucleus to steer gene expression, a process called retrograde signaling. Such retrograde signaling pathways have been described in fungi, animals, and plants. Especially, when the cellular or metabolic situation in the organelle changes (e.g., availability of substrates or light, inhibition of important enzymes, and reactive oxygen species), adequate adjustments in transcript levels are needed to fine-tune the organellar proteomes. This further raises the question of how lineage-specific organellar proteins become regulated appropriately after their appearance. To be of optimal use to the cell, the new genes may become incorporated into existing transcriptional networks relevant to organellar function. Alternatively, specific new needs may require new transcriptional modules to evolve. Indeed, the best-known retrograde signaling pathways in yeast, animals, or plants appear to be quite different between lineages and employ different (even lineage-specific) transcription factors (Ng et al. 2014; da Cunha et al. 2015).

In this study, we describe the phylogenetic history of lineage-specific Domain of Unknown Function 295 (DUF295) genes in plants. Despite its poorly understood function, the gene family is strongly expanded with 94 representatives in the *Arabidopsis thaliana* genome. Our findings show that relatively recent tandem gene duplications in the Brassicaceae family have led to neofunctionalization in plant mitochondria. Most likely through incomplete gene duplication, an ancestral DUF295 domain gene has lost its N-terminus and has instead obtained a functional mitochondrial targeting peptide. Furthermore, we show that several of these new mitochondrial proteins have been specifically integrated

into pre-existing gene-expression networks containing “old” genes that regulate mitochondrial function.

Results

The DUF295 Proteins Form a Large Gene Family within Angiosperms

Despite many decades of intensive research, many conserved protein domains still have unknown functions. The term Domain of Unknown Function (DUF) was originally coined to describe two bacterial domains (DUF1 and DUF2) found in bacterial signaling proteins (Schultz et al. 1998). Subsequent bioinformatics approaches identified thousands of additional uncharacterized domains that were assigned numbers in the PFAM database. The latest PFAM release 32 contains nearly 4,000 DUF families (up to DUF5654), representing around >20% of the known families (Bateman et al. 2010). The DUF295 domain was identified by PFAM release 7.0 (Bateman et al. 2004) and currently contains 4,353 family members, with an average domain length of 57.80 amino acids. The *A. thaliana* Col-0 genome sequence was searched for proteins containing the DUF295 domain, based on the PFAM motif PF03478 (supplementary fig. 1, Supplementary Material online) and the TAIR10 annotation (www.arabidopsis.org; last accessed February 25, 2019). Using further homology searches, in total 94 unique loci encoding DUF295-related proteins were found (fig. 1A, table 1). To examine the evolutionary conservation and origin of the DUF295 protein family, homology searches were performed to identify representative family members in other lineages. The DUF295 domain was not found in prokaryotes and Animalia. Interestingly, a single DUF295 domain protein was found by EBI (<http://www.ebi.ac.uk/interpro/entry/IPR005174/taxonomy>; last accessed February 25, 2019) in the Basidiomycete *Exigia glandulosa* (KZV93935.1; Fungi). However, this order-specific protein (Auriculariales) did not have any homologs with the DUF295 domain, suggesting it is not a true DUF295 domain protein, or has evolved independently in a single known fungal species. We thus concluded that the DUF295 domain is green lineage specific.

Through the application of a second round of homology searches using the PLAZA 4.0 comparative genomics database containing >70 genomes of species within the Viridiplantae, no DUF295 proteins were found in gymnosperms (fig. 1 and supplementary fig. 2, Supplementary Material online). However, DUF295 proteins were found in monocots (e.g., *Oryza sativa* subsp. *japonica* contains 264 DUF295 proteins) and dicots (e.g., *Populus trichocarpa* contains 23). The presence of DUF295 proteins in the Angiosperm *Amborella trichopoda*, which is thought to be a “sister species” of flowering plants that branched off after the gymnosperms, but before the monocot–dicot divergence, was ambiguous (Amborella Genome Project 2013). Using a search with the DUF295 PFAM HMM profile, no *A. trichopoda* proteins were identified. In conclusion, the DUF295 domain is strongly represented in monocots and dicots and has likely originated around the branching of *A. trichopoda*, after the Gymnosperm/Angiosperm divergence.

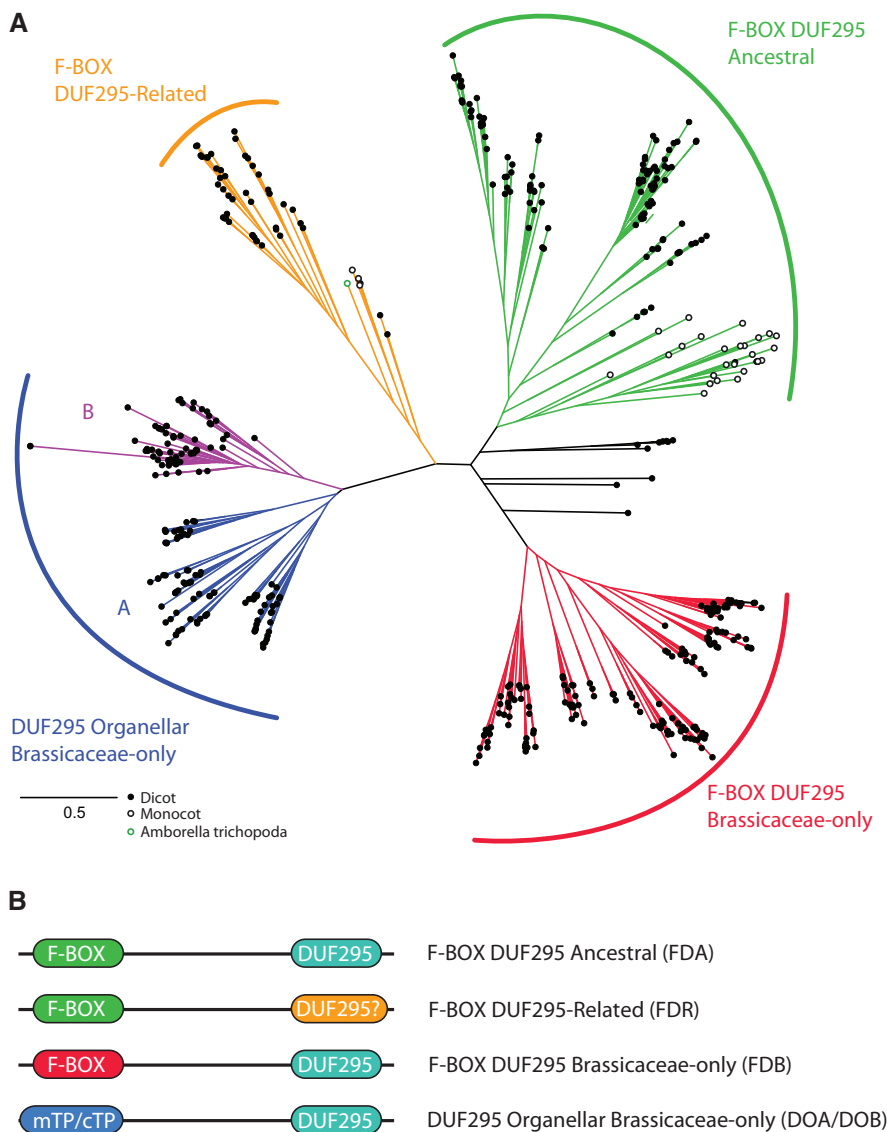


Fig. 1. Phylogenetic analysis of the DUF295 protein family. (A) Unrooted phylogenetic tree of proteins containing DUF295 domains in representative Angiosperm species. Scale bar indicates percentage divergence. For clarity gene and species names have been removed, but information on dicot, monocot, or *Amborella trichopoda* is indicated by different circles (see figure). A fully annotated phylogenetic tree with species/gene names and bootstrap values can be found in [supplementary figure 2, Supplementary Material](#) online. (B) General domain structure of the four groups of DUF295 proteins found in Brassicaceae. mTP, mitochondrial targeting peptide; cTP, chloroplast transit peptide.

Most DUF295 Domain Proteins Also Contain an F-Box Domain

From the phylogenetic analysis, it was apparent that the DUF295 proteins can be divided in four major classes (fig. 1A). One class of DUF295 proteins (indicated in green in fig. 1) was represented in the genomes of all studied plant species, with a clear subgroup of monocot and dicot representatives. This group is thus most likely the ancestral DUF295 protein class. Within the dicot subgroup, a clear expansion of Brassicaceae DUF295 homologs was observed. Analysis of the domain structure of the proteins in the group showed the presence of an N-terminal F-box domain and a C-terminal DUF295 domain (fig. 1B). F-box domains are about 50 amino acids long and involved in protein–protein interactions. They are often found in Skp1-cullin-F-box (SCF)

ubiquitin E3-ligases that mark proteins for degradation, with the F-box imparting specificity of the target proteins. Many key plant hormone receptors have been found to be SCF proteins, including SCF^{TIR} (auxin receptor) and SCF^{CO11} (jasmonic acid receptor) (Kepinski and Leyser 2005; Katsir et al. 2008). The Arabidopsis genome contains 20 ancestral-type F-box/DUF295 proteins. The SKP1-interacting Protein SKIP23 (*At2g17030*) is part of this group and was previously found to interact with ASK1 (Risseuw et al. 2003), a component of, for example, the strigolactone SCF^{MAX2} receptor complex (Yao et al. 2016). In a more recent study, six ancestral-type F-box/DUF295 proteins were found to interact with ASK1 and related proteins by yeast two-hybrid screens (Kuroda et al. 2012). SKIP23 was also found to interact with Arabidopsis 14-3-3 proteins (Hong et al. 2017). Upward Curly

Table 1. Overview of DUF295-Related Genes in *Arabidopsis thaliana*.

AGI	Gene Name	<i>A. trich</i>	Mono/ Dicot	Brassic- Only	F-Box	DUF295	SUBAcon	MS	ASK1 Binding	MRR	Coexpressed Phylostratum
AT1G44080	AtFDA1	—	+	—	+	+	nu	—	—	—	Ang Eud
AT1G57906	AtFDA2	—	+	—	+	+	cyt	—	—	—	
AT1G64840	AtFDA3	—	+	—	+	+	cyt, pm	—	+	—	Vir
At1g65375	AtFDA4	—	+	—	+	+	—	—	—	—	
AT1G65740	AtFDA5/UCL1	—	+	—	+	+	cyt	—	+	—	Eud Brass
AT1G65760	AtFDA6	—	+	—	+	+	pm	—	—	—	
AT1G65770	AtFDA7	—	+	—	+	+	mito	—	+	—	Eud Brass
AT2G16290	AtFDA8	—	+	—	+	+	pm	—	—	—	Ang Eud Brass
AT2G16300	AtFDA9	—	+	—	+	(+)	pm	—	—	—	Ang Eud Brass
AT2G16365	AtFDA10	—	+	—	+	(+)	nu	—	—	—	Vir
AT2G17030	AtFDA11/SKIP23	—	+	—	+	+	golgi	—	+	—	Vir
AT2G17036	AtFDA12	—	+	—	+	+	cyt	—	—	—	
AT2G17690	AtFDA13	—	+	—	+	+	cyt	cyto	+	—	Ang Eud Brass
AT2G24250	AtFDA14	—	+	—	+	+	cyt	—	+	—	Land
AT2G24255	AtFDA15	—	+	—	+	+	pm	—	—	—	Ang Eud
AT2G26160	AtFDA16	—	+	—	+	+	cyt	—	—	—	Ang Eud
AT3G25750	AtFDA17	—	+	—	+	+	nu	—	—	—	Ang Eud
AT4G35733	AtFDA18	—	+	—	+	+	cyt	—	—	—	Ang Eud
AT5G24040	AtFDA19	—	+	—	+	+	cyt	—	—	—	Land
AT5G60060	AtFDA20	—	+	—	+	+	cyt	—	—	—	
AT1G10110	AtFDB1	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT1G27540	AtFDB2	—	—	+	+	+	plastid	—	—	—	Vir
AT1G27550	AtFDB3	—	—	+	+	+	mito	—	—	—	Brass
AT1G27580	AtFDB4	—	—	+	+	+	mito	—	—	—	Ang Eud
AT1G67160	AtFDB5	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT1G69090	AtFDB6	—	—	+	+	+	plastid	—	—	—	Brass
AT2G03560	AtFDB7	—	—	+	+	+	cyt	—	—	—	Ang Eud Brass
AT2G03610	AtFDB8	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT2G04810	AtFDB9	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT2G04830	AtFDB10	—	—	+	+	+	cyt	—	—	—	Ang Brass
AT2G04840	AtFDB11	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT2G05970	AtFDB12	—	—	+	+	+	cyt	—	—	—	Ang Eud Brass
AT2G14290	AtFDB13	—	—	+	+	+	cyt	—	—	—	Land
AT2G14500	AtFDB14	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT2G24080	AtFDB15	—	—	+	+	+	cyt	—	—	—	Eud
AT2G33190	AtFDB16	—	—	+	+	+	cyt	—	—	—	
AT2G33200	AtFDB17	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT3G03726	AtFDB18	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT3G03730	AtFDB19	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT3G22333	AtFDB20	—	—	+	+	+	—	—	—	—	Ang Eud
AT3G22345	AtFDB21	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT4G10820	AtFDB22	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT4G12810	AtFDB23	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT4G12820	AtFDB24	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT4G14165	AtFDB25	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT4G17565	AtFDB26	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT4G22035	AtFDB27	—	—	+	+	+	—	—	—	—	Eud Brass
AT4G22030	AtFDB28	—	—	+	+	+	plastid	—	—	—	Ang
AT4G22060	AtFDB29	—	—	+	+	+	pm, golgi	—	—	—	Ang Eud
AT4G22165	AtFDB30	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT4G22170	AtFDB31	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT4G22180	AtFDB32	—	—	+	+	+	cyt	—	—	—	
AT4G22660	AtFDB33	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT5G14160	AtFDB34	—	—	+	+	+	cyt	—	—	—	Land
AT5G25290	AtFDB35	—	—	+	+	+	pm	—	—	—	Ang Eud
AT5G25300	AtFDB36	—	—	+	+	+	nu	—	—	—	Ang Eud
AT5G38270	AtFDB37	—	—	+	+	+	cyt	—	—	—	Ang Eud
AT5G66830	AtFDB38	—	—	+	+	+	cyt	—	—	—	
AT1G57790	AtFDR1	+	+	—	+	Not clear	cyt	—	+	—	
AT5G55150	AtFDR2	+	+	—	+	Not clear	pm	—	—	—	Land
AT1G05540	AtDOA1	—	—	+	—	+	mito	—	—	—	Land
AT1G05550	AtDOA2	—	—	+	—	+	pm	—	—	—	Land Ang
AT1G30160	AtDOA3	—	—	+	—	+	cyt	—	—	+	Eud

(continued)

Table 1. Continued

AGI	Gene Name	<i>A. trich</i>	Mono/ Dicot	Brassic- Only	F-Box	DUF295	SUBAcon	MS	ASK1 Binding	MRR	Coexpressed Phylostratum
AT1G30170	AtDOA4	–	–	+	–	+	mito	–	–	–	Ang
AT1G68960	AtDOA5	–	–	+	–	+	mito	–	–	–	Ang Eud
AT2G45940	AtDOA6	–	–	+	–	+	cyt	–	–	–	Land Vir
AT4G14260	AtDOA7	–	–	+	–	+	nu	–	–	–	Brass Eud
AT4G16080	AtDOA8	–	–	+	–	+	mito	mito	–	–	
AT4G25920	AtDOA9	–	–	+	–	+	mito	–	–	–	Eud
AT4G25930	AtDOA10	–	–	+	–	+	mito	–	–	+	Ang Eud
AT5G03390	AtDOA11	–	–	+	–	+	mito	plastid	–	–	Land
AT5G46130	AtDOA12	–	–	+	–	+	mito	–	–	–	
AT5G46140	AtDOA13	–	–	+	–	+	cyt	–	–	–	
AT5G53780	AtDOA14	–	–	+	–	+	cyt	–	–	–	Eud Brass
AT5G53790	AtDOA15	–	–	+	–	+	pm	–	–	–	Brass
AT5G55440	AtDOA16	–	–	+	–	+	mito	–	–	–	Eud Brass
AT5G67040	AtDOA17	–	–	+	–	+	cyt	–	–	–	Brass
AT3G25200	AtDOB1	–	–	+	–	+	cyt	–	–	–	Ang Eud
AT3G43170	AtDOB2	–	–	+	–	+	perox	–	–	–	Ang Eud
AT4G13680	AtDOB3	–	–	+	–	+	mito	–	–	–	Brass Eud
AT5G52930	AtDOB4	–	–	+	–	+	mito	–	–	+	
AT5G52940	AtDOB5	–	–	+	–	+	mito	–	–	+	Vir
AT5G53230	AtDOB6	–	–	+	–	+	mito	–	–	+	Ang Eud
AT5G53240	AtDOB7	–	–	+	–	+	mito	–	–	–	Brass
AT5G54320	AtDOB8	–	–	+	–	+	mito	–	–	–	Ang Eud
AT5G54330	AtDOB9	–	–	+	–	+	mito	–	–	–	Ang Eud
AT5G54420	AtDOB10	–	–	+	–	+	pm	–	–	–	Land Brass
AT5G54450	AtDOB11	–	–	+	–	+	mito	–	–	+	Ang Eud
AT5G54550	AtDOB12	–	–	+	–	+	mito	–	–	+	Ang Eud
AT5G54560	AtDOB13	–	–	+	–	+	mito	–	–	+	Land
AT5G55270	AtDOB14	–	–	+	–	+	plastid	–	–	–	Eud
AT5G55870	AtDOB15	–	–	+	–	+	mito	–	–	–	Ang Eud
AT5G55880	AtDOB16	–	–	+	–	+	mito	–	–	–	Ang Eud
AT5G55890	AtDOB17	–	–	+	–	+	mito	–	–	–	Ang Eud

NOTE.—*A. trich*, conserved in *Amborella trichopoda*; Mono/Dicot, conserved in monocots and dicots; Brassic-only, only conserved in Brassicaceae; F-box, protein contains an F-box domain; DUF295, protein contains a DUF295 domain according to the PFAM motif; SUBAcon, subcellular location as suggested by the SUBAcon algorithm; MS, subcellular location as detected by MS (Zybailov et al. 2008; Hummel et al. 2012; Senkler et al. 2017); ASK1 binding, experimentally protein–protein interaction with SKP1/ASK1 proteins; MRR, transcriptionally regulated by mitochondrial retrograde signaling; and coexpressed phylostratum, overrepresented phylostrata in the 300 most strongly coexpressed genes (Ang, angiosperms; Eud, eudicots; Vir, Viridiplantae; Brass, Brassicaceae; and Land, land plants).

Leaf 1 (UCL1, *At1g65740*) was found to be a nuclear protein interacting with Curly Leaf Polycomb proteins and ASK1 (Jeong et al. 2011). This indicates that ASK1 binding is a common feature of this protein group. As most Arabidopsis DUF295 proteins lack a systematic gene name, we named the genes in this group *AtFDA1-20* (F-box/DUF295 Ancestral) (fig. 1 and table 1).

Interestingly, *AtFDA9* (*At2g16300*) is nearly identical to the adjacent gene *AtFDA8* (*At2g16290*), however a frameshift has occurred due to a single base deletion just before the start of the DUF295 domain. This leads to a premature stop codon and truncated *AtFDA9* protein of 322 amino acids, instead of around 415 amino acids as in *AtFDA8* (where the DUF295 domain is at position 319–360). If the *AtFDA9* transcript sequence after the premature stop codon is translated in the +2 frame, the DUF295 domain can be clearly identified, indicating that *AtFDA9* was originally a DUF295-containing gene. The structure of the close-by gene *AtFDA10* (*At2g16365*) appeared even more complex. In the current TAIR10 annotation, *At2g16365* is named *photoperiodic control of hypocotyl 1* (*PCH1*), which is 778 amino acids long. The *PCH1* “domain” (a phytochrome interaction domain) is

located at the N-terminal, and the F-box is at amino acids 459–505. In the other *AtFDA* genes, the F-box is located right at the N-terminal (e.g., residues 3–48 in *AtFDA8*), suggesting it is a compound gene. Indeed, six splice forms have been annotated for *At2g16365*, where *At2g16365.2* does not contain the F-box and downstream sequence. We checked a range of RNA-Seq data sets but could not find evidence for reads spanning the suggested third intron, which would connect the *PCH1* region to the *FDA* region. There is also no proteomic support for the existence of proteins containing both *PCH1* and *AtFDA10* sequence, and only the short splice variant could be cloned (*At2g16365.2*) (Huang et al. 2016). Therefore, we propose that the currently annotated *At2g16365* locus actually encodes two separate genes, *PCH1* and *AtFDA10*. Additionally, *AtFDA10* has a two-base insertion upstream of the DUF295 domain, causing a premature stop codon and loss of the actual DUF295 domain, as observed in *AtFDA9*.

A second group of DUF295 proteins was identified (indicated in red in fig. 1), which also contained an F-box/DUF295 arrangement, but was clearly divergent from the ancestral *FDA*-type proteins. This type of protein was only represented

in Brassicaceae genomes, with *A. thaliana* containing 38 homologs, *Arabidopsis lyrata* 56 homologs, and *Brassica rapa* 32 homologs, respectively. As none of these genes have systematic gene names, we named the *A. thaliana* genes in this group AtFDB1-38 (F-Box/DUF295 Brassicaceae specific). Despite their relatively large number, not much functional information could be found on these proteins. The same study that identified six AtFDA proteins to interact with ASK1 and its homologs in *Arabidopsis*, could not detect ASK1-interaction for the six tested AtFDB proteins (Kuroda et al. 2012). This strongly suggests that this relatively recent, Brassicaceae-specific group has significantly diverged from the ancestral FDA DUF295 proteins.

A third group of DUF295 domain proteins was found (Indicated in blue and purple in fig. 1), but again only in Brassicaceae species (fig. 1). Within this group, two clear subgroups were observed, each containing 17 *A. thaliana* proteins, and many orthologs in the other included Brassicaceae species. Remarkably, none of the proteins in this third group contained the F-box domain, and only the C-terminal DUF295 domain could be found as an annotated domain (fig. 1). Yeast two-hybrid interactions have been reported only for At4g25920 (*Arabidopsis Interactome Mapping Consortium* 2011) (supplementary table 1, Supplementary Material online).

A fourth group (indicated in orange in fig. 1), including At1g57790 and At5g55150, contains the F-box domain, but the DUF295 domain mentioned in the TAIR annotation was not identified using the PFAM profile, indicating it has diverged substantially. A yeast two-hybrid interaction was found with ASK1 for At1g57790 (Kuroda et al. 2012), indicating at least partial functional similarity to the FDA ancestral DUF295 proteins. At1g57790 and At5g55150 were most closely related to *A. trichopoda* ATR0851G001 in the phylogenetic tree. Therefore, we propose that this group represents an older precursor or sister-group to the “proper” DUF295 protein family, and named it AtFDR1-2 (F-Box/DUF295-Related).

The DUF295 Gene Family Has Expanded Rapidly by Tandem Gene Duplication

It was surprising to find that both groups of Brassicaceae-specific DUF295 proteins (fig. 1) contained more members (38 and 34 in *Arabidopsis*) than the ancestral DUF295 group (20 in *Arabidopsis*). This indicates a very rapid expansion of the gene family in a relatively short evolutionary time, as Brassicaceae are thought to have branched off about 32 Ma (Hohmann et al. 2015). When examining the chromosomal locations of genes in the three groups, it was obvious that many homologs were tandem duplications, as evidenced by (nearly) adjacent locations and belonging to the same protein (sub)group. Across the whole DUF295-related family in *A. thaliana* (94 genes), nearly 70% (63 genes) were present as tandem repeats, representing 27 tandem groups (supplementary table 2, Supplementary Material online). Duplication rates were particularly high in the DUF295-only group, with 27 of 34 genes ($\pm 80\%$) spread over 13 tandems. One tandem even contained six genes spanning At5g54320 to At5g54560.

The F-box/DUF295 FDB group contained 26 of 38 ($\pm 68\%$) tandem duplicated genes spread over 10 tandems. Also here tandems of up to six genes were found (At4g22030 to At4g22180). Finally, also the ancestral FDA family contained many tandem duplications, with 10 of 20 genes (50%) spread over 4 tandems. In conclusion, it appears that the DUF295 family has achieved its large size, particularly in *Brassicaceae*, via numerous rounds of tandem duplications.

The Brassicaceae-Specific DUF295 Group Has Replaced the F-Box Domain with a Mitochondrial Targeting Peptide

Since the N-terminal F-box domain was lost or missing in the third group of DUF295 proteins, we analyzed the N-terminal region of these proteins in greater detail. Surprisingly, the large majority of the *A. thaliana* representatives contained predicted mitochondrial targeting peptides, as suggested by several prediction tools including iPSORT, Mitoprot, and Mitopred (supplementary table 3, Supplementary Material online) (Hooper et al. 2017). Of 34 *A. thaliana* homologs, 26 were predicted to be mitochondrial by 6 or more prediction tools, and 14 were predicted to be mitochondrial by 10 or more prediction tools. Twenty-two out of 34 proteins received a consensus localization prediction to the mitochondria based on the SUBA4 consensus algorithm (Hooper et al. 2014). Twenty-four of 26 proteins were also predicted as plastid localized by at least 1 prediction tool (SUBA4), but only 4 were predicted to plastid targeted by 5–7 prediction tools. The SUBA consensus algorithm suggested plastid localization for only 1 protein, as opposed to 22 receiving a mitochondrial consensus prediction. Again, no systematic naming system is present for these proteins so we named the *A. thaliana* homologs AtDOA1-17 (DUF295 Organellar A) and AtDOB1-17 (DUF295 Organellar B), based on the two apparent subgroups (fig. 1).

Despite the strong predictions, experimental evidence for organellar location of the DOA and DOB proteins was very limited. AtDOA8 (At4g16080) was identified in purified mitochondria by mass spectrometry (MS) (Senkler et al. 2017), whereas AtDOA11 (AT5G03390) was identified by MS in purified chloroplasts (Zybailov et al. 2008). As no green fluorescent protein (GFP) localization data were published for any of the predicted mitochondrial isoforms, we cloned three representatives AtDOA10 (At4g25930), AtDOB5 (At5g52940), and AtDOB12 (At5g54550) into C-terminal GFP-fusion vectors (see below for more information on why these were selected). The localization of the fusion proteins was analyzed by transient transformation of *A. thaliana* cell cultures (fig. 2). Both AtDOB5 and AtDOB12 showed clear mitochondrial localization, as evidenced by colocalization with an alternative oxidase (AOX)-red fluorescent protein (RFP) marker. AtDOB5 also showed a weaker signal in plastids, suggesting dual localization (fig. 2 and supplementary fig. 3, Supplementary Material online). For AtDOA10, only diffuse cytosolic localization was found, with no clear colocalization with the AOX-RFP marker. In conclusion, independent sources and experimental approaches support that several of the DOA/DOB proteins

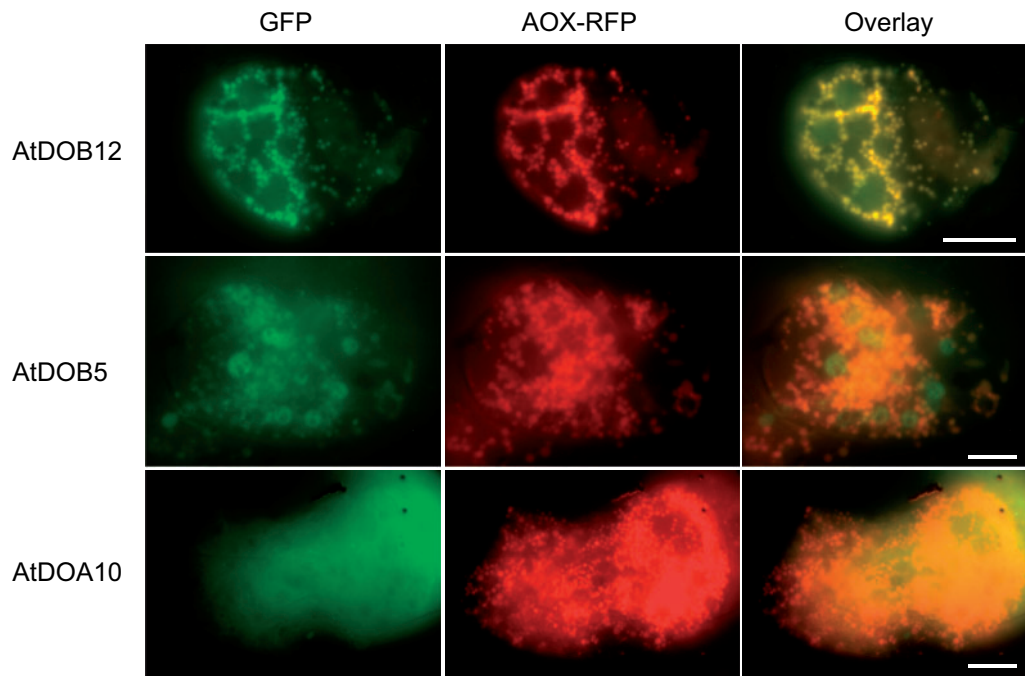


Fig. 2. DUF295 organellar proteins are targeted to the mitochondria. C-terminally GFP-tagged fusion proteins were transiently transformed into *Arabidopsis thaliana* cell culture and cotransformed with mitochondrial marker AOX-RFP. Scale bar indicates 10 μ m.

have obtained functional mitochondrial and/or plastid targeting peptides, in line with their strong organellar prediction.

DUF295 Genes Show Remarkably Specific Expression Patterns

The strong expansion of DUF295 genes does not necessarily indicate that the genes are functional and expressed. Therefore, the transcript levels of the 94 *Arabidopsis* genes were analyzed in a large set of available gene-expression experiments (fig. 3 and supplementary fig. 5, Supplementary Material online). Starting from the transcript counts for 206 public RNA-Seq experiments, a gene-expression matrix was generated by summing transcript counts per locus (Vanechoutte et al. 2017). Out of 94 DUF295-related genes, 73 (78%) appeared to be expressed in one or more conditions (maximum transcripts per million >2). More than 50% (11) of the nonexpressed genes were of the *AtFDB* type, whereas only 1 *AtFDA* (*AtFDA8*) did not seem to be significantly expressed. Four *AtDOA* and five *AtDOB* genes also were not clearly expressed. Remarkably, most of the expressed genes were expressed under relatively specific conditions with only four DUF295 genes showing strong ubiquitous expression (*AtFDA3*, *AtFDA11/SKIP23*, *AtFDA14*, and *AtFDB2*). DUF295-related *AtFDR1* also appears to be expressed in most tissues and conditions. In contrast, most DUF295 genes were expressed under very specific tissues or conditions, often reproductive tissues such as young anthers, pollen, siliques, and young seeds. Others were specifically expressed during abiotic stress, or biotic stress (*Botrytis cinerea*).

As many DUF295-containing genes are present in tandem duplicates, often with up to six related genes in close proximity, we examined whether tandem duplicated genes are coexpressed. In many cases, tandem pairs of two genes

were found to cluster together and showed very similar expression patterns (e.g., *AtDOA1/2*, *AtDOA14/15*, *AtDOB4/5*, and *AtDOB16/17*). Interestingly such paired expression patterns were often observed for *AtDOA* and *AtDOB* genes, whereas F-box containing genes only rarely showed such clear coexpression between tandem repeated genes: *AtFDB30* and *AtFDB33*, though these are interspersed by two non-coexpressed genes in the tandem repeat). We also noted that in the larger tandem repeats like *AtDOB8-13* (six genes), only groups of maximum two genes were similarly expressed (*AtDOB8/9* and *AtDOB10/11*), but these two pairs were very different from each other (fig. 3). Several groups of genes showed remarkably similar expression patterns, such as eight mixed *AtFDA/AtFDB* genes expressed in siliques, or eight genes induced by *Botrytis cinerea* infection (with members of *AtFDA/FDB/DOA/DOB* groups). Clearly, the genes in these groups were not tandem repeats, so the mechanism behind their coexpression is most likely not tandem duplication of promoter regions.

High-Impact Mutation Analysis across 1,135 *A. thaliana* Genomes

To get more insight into which DUF295-related genes may be more active and/or functionally important, we assessed whether they are retained as intact open reading frames in the genome sequences of 1,135 *A. thaliana* accessions published by the 1001 Genomes Consortium (2016). For all 94 *A. thaliana* Col-0 DUF295-related genes, the occurrence of “high-impact mutations” (HIMs; e.g., gain or loss of start/stop codons and loss of splice acceptor sites) was searched in the other ecotypes. This varied widely, with some genes having accumulated no HIMs in other accessions, whereas others have accumulated many hundreds (supplementary

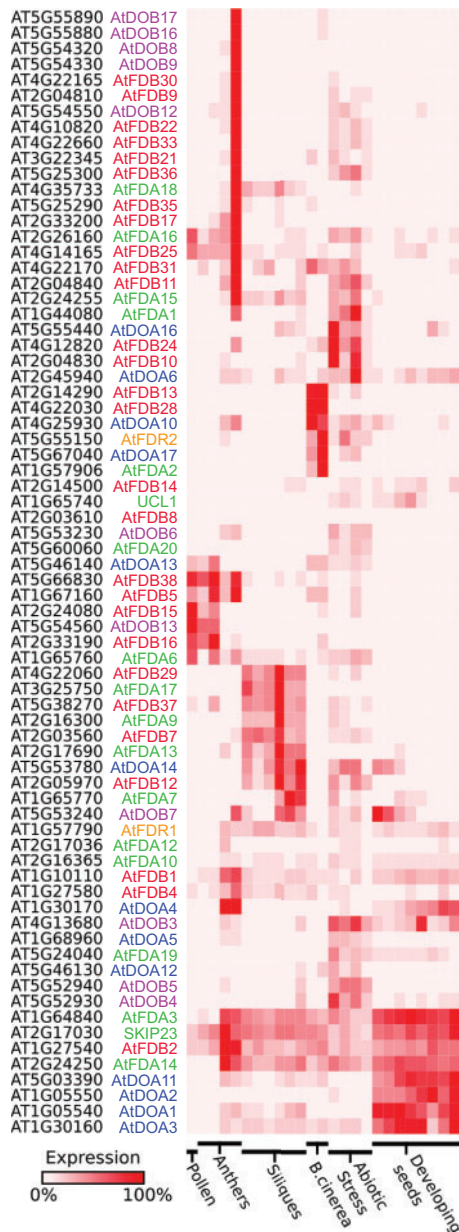


Fig. 3. Expression patterns of DUF295 genes. Expression values, normalized per gene, are shown for 72 expressed DUF295 genes. Gene names are colored to indicate the family subgroups (see fig. 1). Only a subset of 25 samples is shown. Expression data for all 206 samples in the Vanechoutte et al. (2017) data set are available in [supplementary figure 4, Supplementary Material online](#).

table 4, [Supplementary Material online](#)). To clarify, if the same variant compared with Col-0 was found in multiple accessions, it was counted as the number of accessions it occurred in. In other words, if one mutation occurred in 300 accessions, this was counted as 300. Next, we plotted the number of HIMs against the transcript expression strength (maximal transcripts per million (max tpm) in the above 206 RNA-Seq data sets) (fig. 4). A clear trend could be observed that genes with high expression usually had a lower number of HIMs. Conversely, genes with low expression often had many mutations. The only clear exception was *At2g16365* which had both the highest expression and the highest number of

HIMs. As stated above, this locus actually contains two separate genes *PCH1* and *AtFDA10*, so it was excluded from the analysis.

Given the high rate of gene duplication, we postulated that a recently duplicated gene may develop into a functional gene (“consolidated”: max tpm > 5, HIM < 50), gradually turn into a pseudogene and eventually disappear via mutations (“degenerating”: HIM < 50), or temporarily remain in an intermediate stage (“undecided”: max tpm < 5, HIM < 5) (fig. 4B). More than 90% of the genes fell inside the intervals using cut-offs max_tpm of 5 and HIMs of 50, suggesting they are relevant (fig. 4A). When examining ancestral FDA genes, it appears that this selection is nearing completion, as nearly all *AtFDA* genes are either “consolidated” or “degenerating” based on our cut-offs, with only one remaining “undecided” (*AtFDA16*). This further supports the idea that the FDA genes are relatively ancient. Similarly, both *AtFDR* genes show strong expression and very low HIMs (0–1), and thus seem completely “consolidated,” supporting their premonocot/dicot divergence origin. For the probably more recent Brassicaceae-specific genes, the situation looks different. For the *AtFDB* F-box genes an even distribution across the three groups can be seen, suggesting selection is still ongoing and balanced. More than 50% of the *AtDOA* genes seem to be “degenerating,” whereas fewer are being consolidated. Conversely, although most *AtDOB* genes are still in a more “undecided” state, far more are being “consolidated” than are “degenerating.” This suggests that there is higher selective pressure on *AtFDR*, *AtDOB*, and *AtFDA* genes, whereas *AtDOA* genes may be degenerating more often.

DUF295 Organellar Genes Were Incorporated into the ANAC017 Retrograde Signaling Pathway

Previously, we reported that eight *DUF295* genes were constitutively induced in Arabidopsis mutants with mitochondrial defects (Van Aken et al. 2016). Surprisingly, all eight of these are members of the *DUF295 Organellar* group (two *AtDOA* and six *AtDOB*), whereas none of the F-box *DUF295* proteins were represented (table 1). To further examine the specificity of *DUF295 Organellar* proteins in responding to mitochondrial dysfunction, an antimycin A treatment time course was set up. Gene-expression levels were measured for the most highly induced *AtDOA* representative (*AtDOA10*, according to [supplementary table 2, Supplementary Material online](#)), and two highly induced *AtDOB* representatives (*AtDOB12* and *AtDOB5*). Furthermore, *AtFDA11/SKIP23* and *AtFDB2* were selected from the F-box *DUF295* proteins, based on their relatively high expression in Col-0 seedlings of similar age in previous RNA-Seq data sets (Van Aken et al. 2016) ([supplementary table 3, Supplementary Material online](#)). Figure 5 shows that only *AtDOB12*, *AtDOB5*, and *AtDOA10* were strongly induced by antimycin A, whereas *AtFDA11/SKIP23* and *AtFDB2* showed no induction.

As antimycin A is known to induce gene expression via retrograde signaling, the response of the selected *DUF295* genes was also monitored in mutants lacking ANAC017, a key transcription factor in plant mitochondrial and

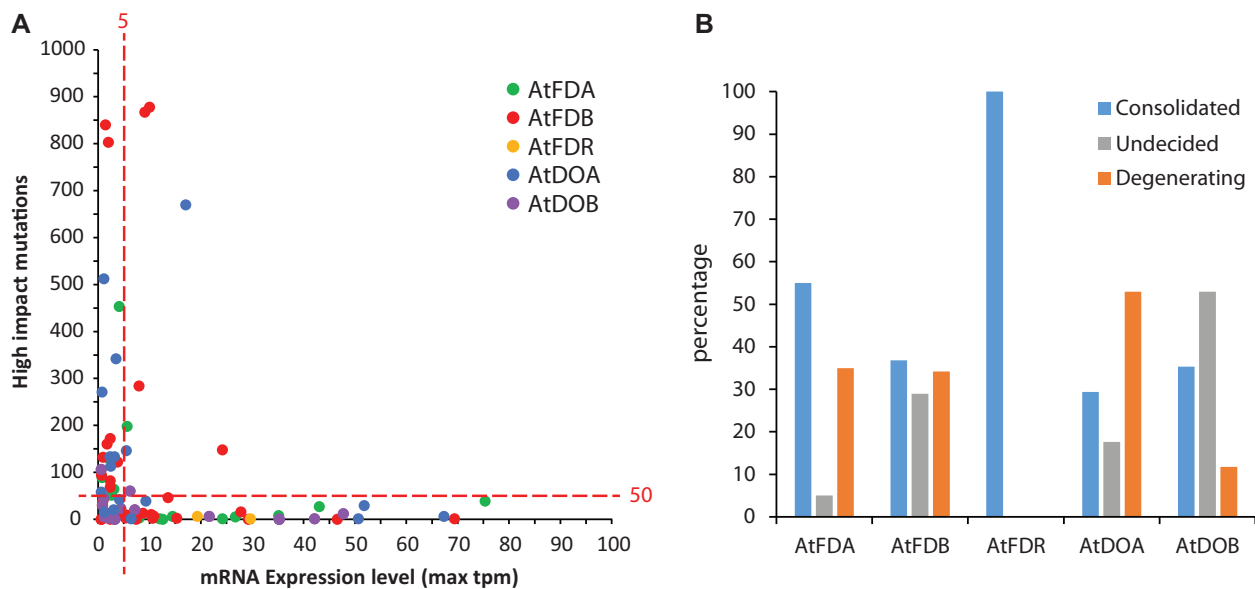


Fig. 4. High-Impact Mutations (HIMs) identified by the 1001 Genomes tool. (A) Accumulation of HIMs in 1,135 *Arabidopsis thaliana* accessions compared with Col-0 were plotted against the gene-expression levels (transcripts per million, max tpm). Genes were color coded by DUF295 subgroup. The red dashed lines indicate the cut-offs used for B. (B) Percentage of genes within each DUF295-related subgroup that are postulated to be “consolidated,” “degenerating,” or “neutral,” based on the cut-offs in A.

chloroplast regulation (De Clercq et al. 2013; Ng et al. 2013; Van Aken et al. 2016). The antimycin A-induced gene expression was almost completely suppressed during the first 6 h, which was where the peak expression occurs in wildtype plants (fig. 5). Some delayed expression was observed toward 9–12 h, which was most likely due to contributions by ANAC017 homologs, such as ANAC013, ANAC053, and ANAC078 (De Clercq et al. 2013; Van Aken et al. 2016). No significant differences in gene expression for AtFDA11/SKIP23 or AtFDB2 were observed between Col-0 and the *anac017* mutants. In summary, the tested *DUF295 Organellar* genes were strongly induced by mitochondrial dysfunction in an ANAC017-dependent way. The F-box *DUF295* genes, however, seem to be largely unresponsive to mitochondrial-stress signaling.

The promoters of the 8 *DUF295 Organellar* genes that were found as responsive to mitochondrial dysfunction based on RNA-Seq data (Van Aken et al. 2016) were searched for binding motifs of ANAC017 and/or its related NAC transcription factors (mitochondrial dysfunction motif) (De Clercq et al. 2013). The TF2Network tool was used and for seven out of eight genes an MDM like motif (CTTGnnnnnCAAG or similar) was found (Kulkarni et al. 2018). Only for AtDOA3 (At1g30160) no MDM could be found, which is in line with its ANAC017-independent gene expression (supplementary table 2, Supplementary Material online). Furthermore, by using DNA affinity purification sequencing (DAP-seq, a variant of chromatin immunoprecipitation ChIP), we found that the promoters of these seven genes bind to ANAC017 and/or its homologs (supplementary table 2, Supplementary Material online) (O’Malley et al. 2016).

It was surprising that only *DUF295* gene variants with (predicted) mitochondrial targeting peptides have become incorporated into a mitochondrial signaling network.

Additionally, this must have occurred relatively recently in evolutionary history, since the *DUF295 Organellar* proteins only evolved in Brassicaceae. As mitochondrial retrograde signaling occurs in all eukaryotic kingdoms (da Cunha et al. 2015), the plant-specific ANAC017-dependent mitochondrial retrograde pathway is most likely much more ancient than Brassicaceae (Kim et al. 2007). This would require that the recent *DUF295 Organellar* genes have been “adopted” by a much older, pre-existing coexpression set. To test this hypothesis, we performed a phylostratic coexpression analysis of all *DUF295* genes. The phylostratic classification grouped *A. thaliana* genes in 13 classes based on their evolutionary conservation (Quint et al. 2012), ranging from genes universally conserved in cellular organisms (phylostratum 1), via Viridiplantae, to genes that are Brassicaceae- (phylostratum 12) or even *A. thaliana*-specific (phylostratum 13). Next, a coexpression analysis was performed using publicly available gene-expression data, to identify the 300 most similarly expressed *A. thaliana* genes for each of the 92 *DUF295* genes. Finally, these 300 coexpressed genes were searched for overrepresentation of genes from the different phylostrata (supplementary table 5, Supplementary Material online) (Ruprecht et al. 2017). Based on this, coexpressed phylostrata were assigned to all *DUF295* genes, giving an indication of the evolutionary age of their coexpression network.

For 82 *DUF295* genes, one or more coexpressed phylostrata were identified (supplementary table 5, Supplementary Material online). When comparing the overall distribution of the *A. thaliana* genome (represented by 32,833 genes used in this analysis), the *DUF295* gene family was particularly enriched in coexpression networks with genes from the Angiosperm and Eudicot phylostrata (fig. 6). This is in line with the presumed age of the *DUF295* domain, which appears to have originated early on in the Angiosperm lineage. The

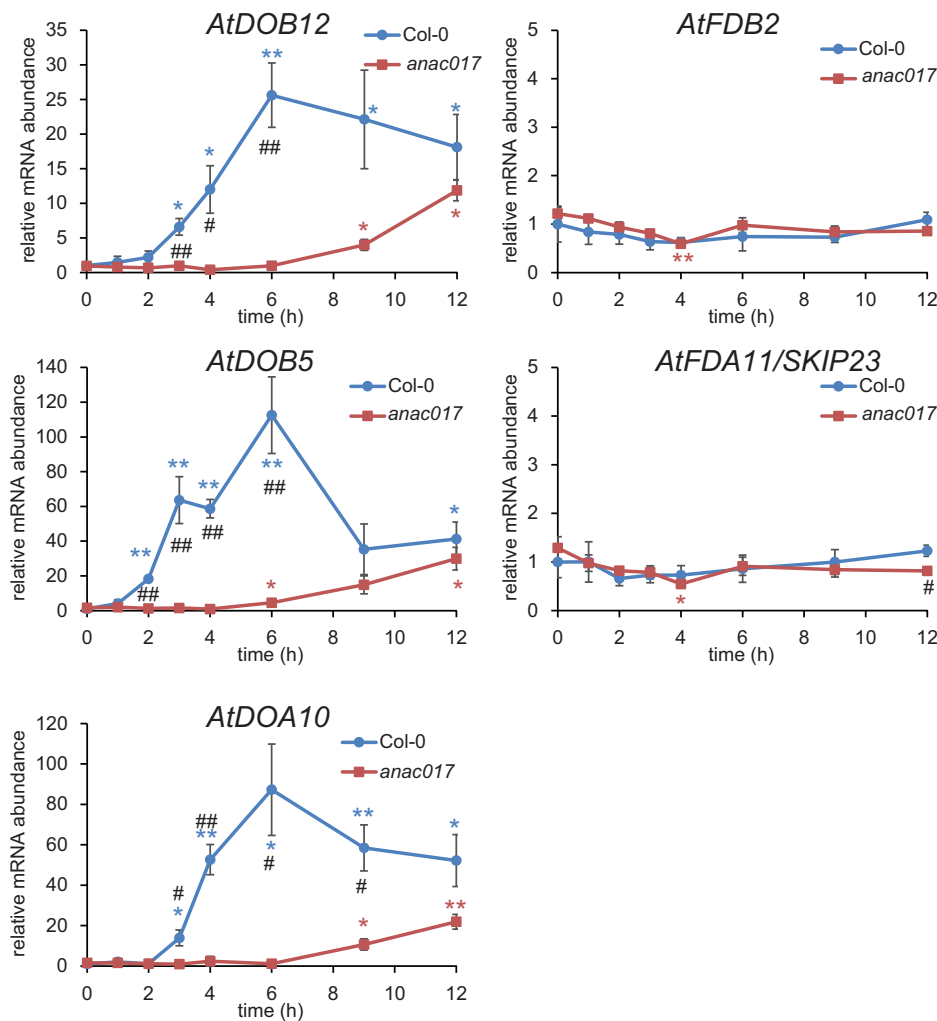


Fig. 5. DUF295 organellar genes are incorporated into mitochondrial retrograde signaling networks. Two-week-old Col-0 and *anac017* mutant plants were treated with antimycin A and samples were collected in triplicate pools of plants. mRNA levels were quantified using qRT-PCR and normalized to Col-0 at time point 0 h. Asterisks indicate statistically significant difference in expression level compared with time point 0 in the same genotype (* $P < 0.05$; ** $P < 0.01$); hash-tags indicate significant difference at the same time point between Col-0 and *anac017* (# $P < 0.05$; ## $P < 0.01$).

specific group of eight Brassicaceae-specific DUF295 organellar genes that are responsive to mitochondrial dysfunction was also enriched in coexpression networks consisting of Angiosperm and Eudicot phylostrata. Interestingly, none of these eight genes are coregulated with Brassicaceae-specific phylostrata, indicating they have been incorporated in coexpression network that is much older than the genes themselves. A similar analysis was performed for 21 “core” mitochondrial retrograde target genes regulated by ANAC017, based on previous data (supplementary table 5, Supplementary Material online). This indicated that the strongest coexpression of core ANAC017-target genes is also found with genes from Angiosperm phylostratum (fig. 6). Stronger coexpression of core ANAC017-target genes was also found in the Landplants phylostratum. In summary, this analysis further supports that the recent mitochondrial-stress responsive DUF295 genes have been adopted by a much older coexpression network, which is of largely similar age to the ANAC017 core regulon.

Characterization of DUF295 T-DNA Insertion Mutants

To investigate the function of DUF295 genes in plants, we isolated T-DNA insertion mutants for representatives of the three main DUF295 groups (fig. 7A). For the ancestral F-box DUF295 genes, *AtFDA11/SKIP23* was chosen, as it has been picked up in several protein–protein interaction screens with ASK1/SKP1-type proteins and 14-3-3 proteins (Risseuw et al. 2003; Kuroda et al. 2012; Hong et al. 2017). It also had the second highest gene-expression level in 2-week-old Col-0 based on RNA-Seq data (supplementary table 3, Supplementary Material online). For Brassicaceae-specific F-Box DUF295 genes, *AtFDB2* was selected as it was by far the most strongly expressed gene in this group. For DUF295 organellar genes, *AtDOA10* was selected as it showed the highest fold-change induction to mitochondrial dysfunction (supplementary table 2, Supplementary Material online). Also *AtDOB5* and *AtDOB12* were chosen because they were the most highly induced representatives of two different *AtDOB*

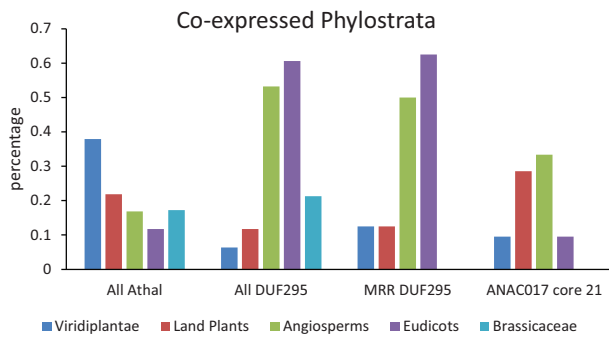


FIG. 6. Phylostratum coexpression analysis. For each gene the 300 most strongly coexpressed *Arabidopsis thaliana* genes were identified. Overrepresented evolutionary conserved phylostrata were searched in these 300 genes. This analysis was done on all *A. thaliana* genes (All Athal), all 94 DUF295-related genes (All DUF295), the 8 DUF295 genes responding to mitochondrial dysfunction or mitochondrial retrograde regulation (MRR DUF295), and on 21 genes that have been consistently found to be regulated by ANAC017 in MRR. The percentage indicates how many genes within this category showed statistically enriched coexpression with member of the indicated phylostrata.

tandem duplications (supplementary table 2, Supplementary Material online). They were also found to be targeted to mitochondria using GFP-fusions (fig. 2). Suitable T-DNA lines were selected from T-DNA express, and homozygous lines were isolated using polymerase chain reaction (PCR)-based genotyping (supplementary table 6, Supplementary Material online).

First, the overall growth rate and phenotype of the mutant lines was compared with Col-0. The rosette surface area was monitored from 14 to 29 days after transfer to the growth room. However, no clear alterations in growth were found compared with Col-0 for any of the lines tested (fig. 7B). Also, no obvious phenotypical differences in plant appearance were observed. As the selected *DUF295 organellar* genes clearly responded to mitochondrial dysfunction, root growth inhibition by antimycin A and methylviologen was tested (fig. 7C–E). Again, no obvious differences in root growth and resistance to inhibitors was observed for any of the lines compared with Col-0. Overall, no clear aberrant phenotypes were observed for any of the T-DNA lines analyzed, which is likely explained by the large extent of gene duplication leading to redundancy, for instance shown by often similar gene-expression patterns of tandem duplications (fig. 3).

Discussion

Through extensive phylogenetic analysis of the DUF295 family, this study found that the F-box/DUF295 domain combination is the most prevalent and conserved configuration in Angiosperms (fig. 1). Most likely these types of proteins derived from F-box precursor proteins, and the DUF295 domain evolved gradually sometime after the Gymnosperm/Angiosperm divergence. A common factor in the limited functional information that is available on the FDA proteins is interaction with SKP1/ASK1-type proteins, which are part of SCF-type ubiquitin E3-ligases. ASK1 seems to mediate the

interaction of the F-box protein with CUL1 (Jeong et al. 2011). The DUF295 domain is likely to be also a protein–protein interaction domain that may be bridging ASK1 and other proteins such as Curly Leaf (CLF), a polycomb SET-domain protein, thereby marking them for degradation. Overexpression of the DUF295 protein UCL1 resulted in similar phenotypes as a loss-of-function mutant in *CLF*, in line with the model that the interaction results in proteasome-mediated degradation of CLF (Jeong et al. 2011). Given the large number of FDA proteins within the same species, it is likely that a large range of proteins may be posttranslationally regulated by such a mechanism. The binding of at least six other AtFDA proteins with ASK1/SKP1 was shown using yeast two-hybrid assays, suggesting this is a common feature (Kuroda et al. 2012). AtFDA11/SKIP23 was also found to interact with 14-3-3 proteins but could not be shown to be directly involved in ubiquitination (Hong et al. 2017). From the limited amount of information available, it thus seems that the DUF295 domain may be a protein–protein interaction domain. For the ancestral FDA proteins, it may help recruit target proteins to SCF E3-ligases for proteasomal degradation.

After several rounds of gene duplication in Brassicaceae, a variant to the F-box DUF295 configuration seems to have arisen (FDB proteins). The yeast two-hybrid screens could not identify an interaction with ASK1 (Kuroda et al. 2012) for any of the six tested FDB proteins, indicating that the F-box domain has diverged significantly. Whether these proteins have obtained a different function is currently unclear. At least a single loss-of-function mutation in *AtFDB2* did not result in obvious phenotypic differences, but this may be due to the extensive redundancy.

The other group of Brassicaceae-specific DUF295 gene variants have led to more radical rearrangement, with the loss of the F-box domain, and the gain of a functional/predicted organellar targeting peptide. Mitochondrial targeting of two AtDOB proteins was confirmed by GFP-fusions in this study (fig. 2), whereas proteomics identified at least one AtDOA protein in isolated mitochondria (Senkler et al. 2017). From an evolutionary standpoint, this represents a clear example of how (partial) gene duplication can result in new organellar proteins. If these new DUF295 Organellar proteins were not useful to plants, one would expect a fast accumulation of point mutations. However, at least for AtDOB proteins, there seems to be some selection pressure, indicating that the genes are being kept in a functional state. Also, most DUF295 genes are expressed at the mRNA level, often in very specific patterns, suggesting they are not pseudogenes. However, their function remains unclear for now. Assuming that the DUF295 domain is a protein–protein interaction domain, they may directly bind other proteins. Due to the loss of the F-box domain, this is unlikely to lead to ubiquitination and protein degradation of the potential binding partners. The DUF295 Organellar are generally only slightly shorter than FDA proteins (most between 350 and 400 amino acids), and the DUF295 domain is close to the C-terminal. Even without the F-box domain and the likely removal of the N-terminal organellar targeting peptide upon import, one would expect

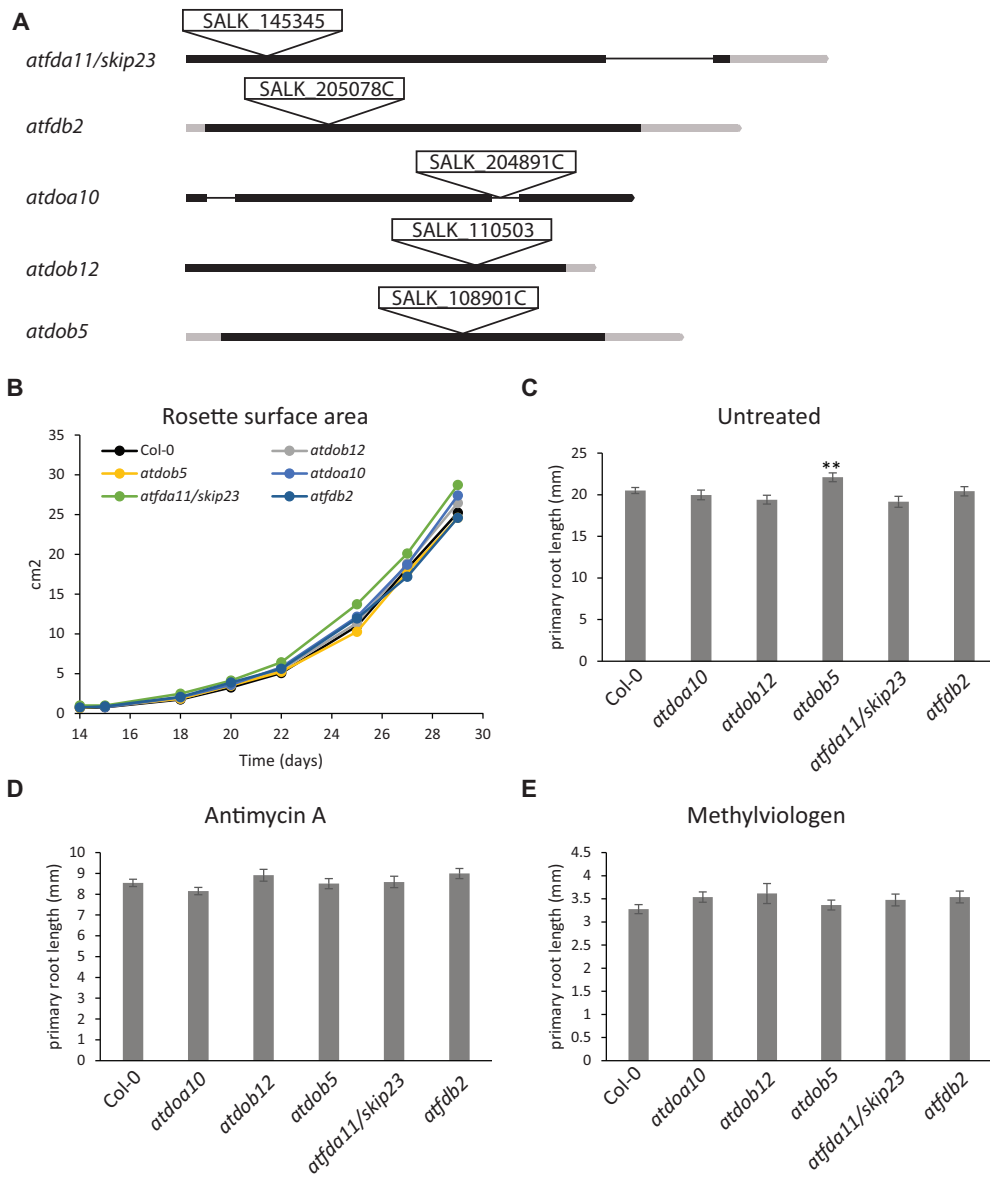


Fig. 7. Phenotypic analysis of DUF295 T-DNA mutants. (A) Overview of T-DNA insertion locations for selected *Arabidopsis thaliana* DUF295 mutant lines. Black bars indicate coding exon, gray bars indicate untranslated regions in exons, and black lines indicate introns. (B) Average rosette surface area of soil-grown plants of the different genotypes monitored over time. (C) Average primary root length after 7 days of growth on vertically oriented MS plates, MS plates supplemented with antimycin A (D) or methylviologen (E). Asterisks indicate statistically significant difference compared with Col-0 (** $P < 0.01$).

at least 200–250 amino acids present in DUF295 proteins outside of the DUF295 domain itself. This would be more than sufficient for other (unknown) functions that are assisted by the DUF295 domain, or perhaps act as a flexible linker between the two domains. To some extent the organellar DUF295 proteins show similarities to microProteins, which are proteins that only contain a protein–protein interaction domain, but no other clear functional domains (Bhati et al. 2018). MicroProteins are thought to have regulatory effects for instance by preventing proteins from forming functional dimers, thus having dominant effects. MicroProteins have also been found in mitochondria, where they can bind mitochondrial elongation factors and stimulate mito-ribosome

translation (Rathore et al. 2018). Further studies with gain/loss-of-function mutants and protein interaction screens may shed further light on the function of these evolving proteins.

A broad gene-expression analysis revealed that only very few DUF295 genes are ubiquitously expressed (fig. 3), for example, *AtFDA11/SKIP23* and *AtFDB2*, which were selected for further study. Most other genes had relatively specific expression patterns. Besides two groups of stress-responsive DUF295 genes, most patterns were strongly biased toward young reproductive tissues, such as siliques, anthers and pollen. Such a bias toward expression of recently evolved genes in the male germ line (“out of testis”) has been reported in animal systems. It has been proposed that male

gametophytes act as an “innovation incubator,” driving species specification and providing new genes to support the arms race against microbial pathogens (Cui et al. 2015). Also other studies support the concept that gene duplication may be a mechanism to adapt organisms to variable environments (Kondrashov 2012). Furthermore, it has been observed in mammalian genomes that the “new” copies of gene after a duplication event show much longer bursts of sequence evolution than the copies in the original location (Pich and Kondrashov 2014). Our data suggest that this is also the case in plants, with both FDB and DOA/DOB groups expanding much more quickly than the ancestral FDA groups. This may be because the (incorrectly) duplicated genes have undergone significant functional changes and are thus evolving toward a new function. From each of these recent subgroups, several individual genes indeed appear to be “consolidating” into conserved state (fig. 4), suggesting their functions are beneficial to the plant.

Considering the recently obtained mitochondrial localization of several DUF295 Organellar proteins, it makes sense that many of them have become integrated in a mitochondrial signaling network, regulated by ANAC017 and its close homologs. This in contrast to the FDA and FDB proteins, which do not appear to be regulated by mitochondrial signaling. Thus, the regulatory information surrounding the coding sequences of the ancestral F-box DUF295 genes (e.g., promoter) may not have been strongly retained during the gene duplication events or at least considerably altered. The question is therefore how the integration into the mitochondrial signaling network occurred. One mechanism could be a gradual shift in regulation, whereby a nonmitochondrially regulated duplicated gene evolved novel regulatory information in the promoter, allowing it to perform its mitochondrial function in a more directed way. An alternative mechanism that could explain both mitochondrial targeting and integration into the ANAC017 regulatory network, could be that the partially duplicated DUF295 gene integrated into a gene encoding an existing mitochondrially target protein. In this way, the partial DUF295 gene may have “hi-jacked” the targeting peptide sequence, as well as the promoter and regulatory information. In such an event, the protein would have instantly become mitochondrially targeted, and regulated by a relevant signaling pathway like ANAC017. The original mitochondrial protein was thereby probably lost.

This second alternative seems the most plausible, as the majority of DOA and DOB genes are predicted to be mitochondrially targeted, and members of both subgroups are ANAC017 regulated (fig. 5). Of the seven DUF295 Organellar genes that are apparently regulated by ANAC017, five are present in tandem gene duplications. *At5g54450*, *At5g54550*, and *At5g54560* form a consecutive group of three (from a total of six in close proximity) (see table 1), whereas *At5g52930* and *At5g52940* form a consecutive group (two out of two at this locus). This suggests their coregulation is possibly caused by coduplication of regulatory information. This coexpression of neighboring genes has been observed previously for unrelated MATE multidrug and toxin efflux carriers *At2g04040*, *At2g04050*, and *At2g04070*, which

are also part of the ANAC017-regulated mitochondrial retrograde pathway (Van Aken et al. 2007, 2016). Possibly, the currently non-ANAC017-regulated organellar DUF295 genes may have lost some of their regulatory information over time.

In conclusion, this study provides compelling evidence for neofunctionalization of proteins via intercompartmental gene duplication in plants, thus adding to the lineage-specific organellar proteome (fig. 8) (Szklarczyk and Huynen 2009). Our study further shows that such duplications can then result in integration into existing and relevant gene regulatory networks, which can be considered as the next stage in the creation of new and useful function. The precise function of the DUF295 proteins is only beginning to be understood, especially of the Brassicaceae-specific subtypes. As they represent 0.3% of the total *A. thaliana* protein-coding gene content, it is likely that future studies will find out more by both targeted and untargeted approaches.

Materials and Methods

Plant Materials and Growth Conditions

Arabidopsis thaliana (L.) Heynh. Col-0 was used in all experiments. Seeds were sown on soil mix or MS media with 2% sucrose and stratified for 2–3 days at 4 °C, then grown under long-day conditions (16 h light/8 h dark) at 22 °C and 100 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Previously published transgenic lines were obtained from Ng et al. (2013) *anac017-anac017-1* SALK_022174. DUF295 mutant lines (fig. 5) were genotyped using PCR on genomic DNA using primers shown in supplementary table 4, Supplementary Material online.

Stress Treatments of Plants

Seeds were sown on petri dishes containing MS medium (Duchefa-Biochimie) + 2% sucrose, stratified for 2–3 days in the cold room and then incubated in long-day growth conditions for 14 days. Pools of plants were then collected before or after treatment and immediately placed in liquid nitrogen for storage and further processing. For transcript analysis, plants were sprayed with 50 μM antimycin A. In vitro stress assays were performed as previously described (De Clercq et al. 2013). For root growth assays, the different plant lines were incubated on vertically positioned plates supplemented with 50- μM antimycin A or 20 μM methylviologen. Plants were stratified in the cold room for 3 days and incubated for 7 days in long-day conditions. Primary root length was measured using ImageJ. Statistical analysis was performed using Student’s *t*-test.

Quantitative Reverse Transcription-PCR and Microarray Analysis

RNA isolation, cDNA generation, and quantitative reverse transcription-PCR (qRT-PCR) were performed as described in Van Aken et al. (2013) using Spectrum RNA Plant extraction kits (Sigma-Aldrich, Sydney, Australia), iScript cDNA synthesis kit (Bio-Rad), and a Roche LC480 Lightcycler using SYBRgreen detection assays. All primers for qRT-PCR are shown in supplementary table 4, Supplementary Material online. Relative expression values were normalized, with

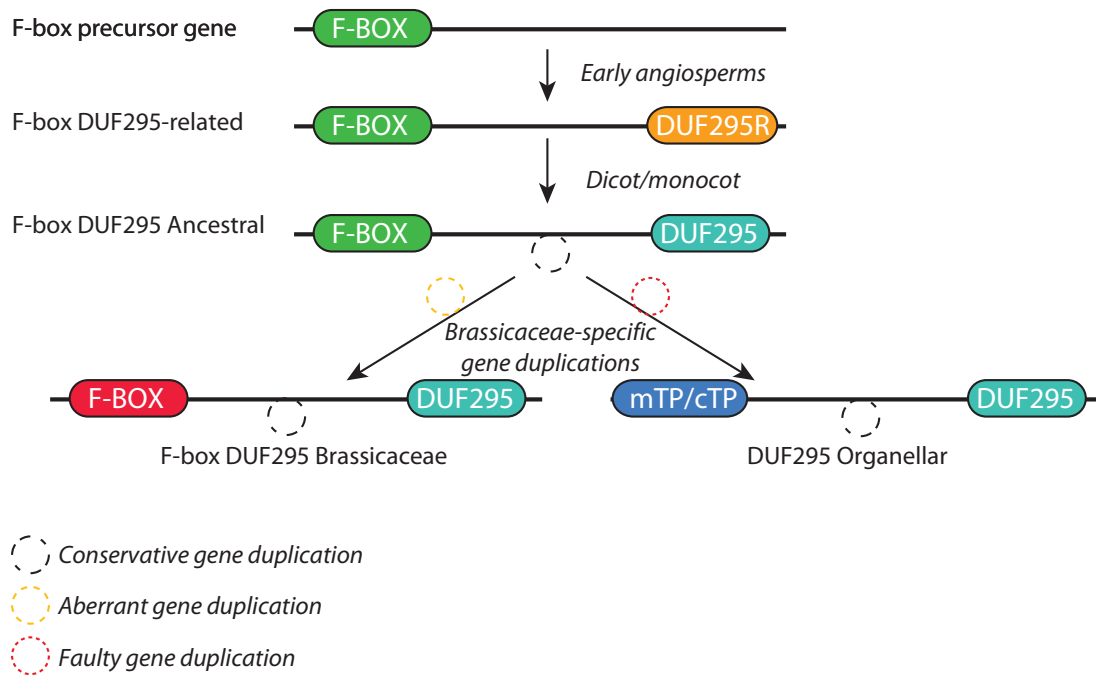


Fig. 8. Model for the evolution of the DUF295 gene family. Our analyses suggest that the DUF295-related domain evolved as an additional C-terminal domain to existing F-box proteins in early angiosperms (presumably 140–180 Ma). The DUF295 domain was then consolidated in monocots and dicots forming the ancestral F-box DUF295 (FDA) protein family, which expanded extensively in the different species via (tandem) gene duplication events. More recently (~32 Ma), aberrant gene duplications specifically in the Brassicaceae resulted in divergent F-box DUF295 (FDB) proteins. The F-box domain was most likely lost t by “faulty” or incomplete gene duplication and replaced with a mitochondrial or chloroplast targeting peptide (mTP/cTP), resulting in the Brassicaceae-specific DUF295 Organellar (DOA and DOB) protein family. All DUF295 protein families appear to have expanded strongly by subsequent (tandem) gene duplication events.

untreated Col-0 samples set as 1. Statistical analyses were performed using Student’s *t*-test throughout the manuscript, except where indicated.

GFP Localization and Microscopy

Coding sequences for full-length DUF295 genes were PCR amplified from *A. thaliana* cDNA and cloned into the pDONR201 Gateway vector (Invitrogen, CA). Cloning into the final GFP pDEST-CGFP vectors was done as described (Carrie et al. 2008). The 42 amino acids targeting signal of AOX was fused to RFP as a mitochondrial marker, and the Rubisco small subunit (SSU-RFP) as a plastid marker (Carrie et al. 2008). Biolistic cotransformation using gold particles of the GFP and RFP fusion vectors was performed on *Arabidopsis* cell culture as previously reported (Carrie et al. 2008). In brief, GFP and RFP plasmids (5 µg each) were coprecipitated onto gold particles and transformed using a PDS-1000/He biolistic transformation system (Bio-Rad). Two to three milliliters of *Arabidopsis* suspension cell culture (4–5 days after 6× dilution of a 1-week-old culture in fresh medium) were placed on osmoticum medium and bombarded. Cells were then incubated for 24–48 h at 22 °C in the dark. GFP and RFP expression and targeting were visualized using a BX61 Olympus microscope (Olympus) using excitation wavelengths of 460/480 nm (GFP) and 535/555 nm (RFP), and emission wavelengths of 495–540 nm (GFP) and 570–625 nm (RFP). Subsequent images were captured using Cell^R imaging software.

Phylogenetic Analysis

Arabidopsis thaliana DUF295 genes were identified using a combination of searches for PFAM motif PF03478, TAIR10 annotation, and homology searches. Representative DUF295 genes from other plant species were obtained using homology searching. Protein sequences were aligned using MAFFT multiple sequence aligner (Katoh and Standley 2013) and edited in BioEdit. Phylogeny was inferred using the IQ-Tree webserver (<http://iqtree.cibiv.univie.ac.at/>; last accessed February 25, 2019) using BLOSUM62 algorithm and 1,000 bootstraps (Katoh and Standley 2013). Phylogenetic trees were visualized using FigTree v1.4.2.

Gene Duplication Analysis

Starting from the set of DUF295 genes reported in table 1, the PLAZA 4.0 Dicots comparative genomics platform was used to retrieve information about gene duplications (Van Bel et al. 2018). Specifically, the PLAZA Workbench was used to define different gene sets and to determine the number of genes involved in a tandem gene duplication event. In the PLAZA database, tandem gene duplicates were identified using i-ADHoRe v3.0.01 (gap_size 30, tandem_gap 30, cluster_gap 35, q_value 0.85, probab_cutoff 0.01, anchor_points 5, and multiple_hypothesis_correction FDR) (Proost et al. 2012).

1001 Genomes SNP Analysis

The 1001 Genomes polymorph tool (<https://tools.1001genomes.org/polymorph/>; last accessed February 25,

2019) was searched for single nucleotide polymorphisms with high impact for all 94 *A. thaliana* DUF295-related genes. The number of ecotypes where specific polymorphisms compared with Col-0 occurred was added.

Phylostratum and Expression Analysis

Starting from the transcript counts reported by Vanechoutte et al. (2017), a gene-expression matrix was generated by summing transcript counts per locus. Subsequently, for each gene, the top 300 coexpressed genes (denoted knn300 cluster) were determined based on the Pearson Correlation Coefficient. Starting from phylostrata information derived from gene families defined in PLAZA 3.0 Dicots (Proost et al. 2015), significantly overrepresented phylostrata per knn300 cluster were identified using the hypergeometric distribution (incl. Benjamini-Hochberg correction for multiple hypothesis testing). All enrichments with corrected P value <0.05 were retained as significant. Expression patterns of DUF295 genes in 206 samples (from the Vanechoutte et al. 2017 data set) were examined in an expression heatmap (fig. 3 and supplementary fig. 2, Supplementary Material online). For this, TPM expression values were first normalized for each gene by dividing them with the maximum TPM observed for that gene. Only genes with a maximum TPM larger than 2 were considered to be expressed and others were excluded from the heatmap. No expression data were available for AtFDA4 and so it was excluded from this analysis as well. Figure 3 shows a manually selected subset of 25 samples to highlight interesting expression behavior of the DUF295 genes.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

O.V.A. was supported by the Swedish Research Council (Vetenskapsrådet 2017-03854), the Australian Research Council (DP160103573), Crafoord Foundation (20170862), Carl Trygger Foundation (CTS 17:487), and Carl Tesdorpf Stiftelse. M.B., K.B., and Z.A. were supported by the ARC Centre of Excellence in Plant Energy Biology (CE140100008). D.V. was supported by the Agency for Innovation by Science and Technology (IWT) in Flanders (predoctoral fellowship).

Author Contributions

The study and experiments were designed by O.V.A., K.V., and T.S. Experiments were performed by S.L., M.B., Z.A., K.B., D.V., K.V., and O.V.A. The manuscript was written by O.V.A. with contributions from the coauthors.

References

1001 Genomes Consortium. 2016. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166(2): 481–491. Amborella Genome Project. 2013. The Amborella genome and the evolution of flowering plants. *Science* 342(6165): 1241089.

Arabidopsis Interactome Mapping Consortium. 2011. Evidence for network evolution in an Arabidopsis interactome map. *Science* 333(6042): 601–607.

Bateman A, Coggill P, Finn RD. 2010. DUFs: families in search of function. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 66(Pt 10): 1148–1152.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer E, et al. 2004. The Pfam protein families database. *Nucleic Acids Res.* 32(Database Issue):D138–D141.

Bhati KK, Blaakmeer A, Paredes EB, Dolde U, Eguen T, Hong SY, Rodrigues V, Straub D, Sun B, Wenkel S. 2018. Approaches to identify and characterize microProteins and their potential uses in biotechnology. *Cell Mol Life Sci.* 75(14): 2529–2536.

Carrie C, Murcha MW, Kuehn K, Duncan O, Barthet M, Smith PM, Eubel H, Meyer E, Day DA, Millar AH, Whelan J. Type II NAD(P)H dehydrogenases are targeted to mitochondria and chloroplasts or peroxisomes in *Arabidopsis thaliana*. *FEBS Lett.* 2008;582(20): 3073–3079.

Cui X, Lv Y, Chen M, Nikoloski Z, Twell D, Zhang D. Young Genes out of the Male: An Insight from Evolutionary Age Analysis of the Pollen Transcriptome. *Mol Plant.* 2015;8(6):935–945.

da Cunha FM, Torelli NQ, Kowaltowski AJ. 2015. Mitochondrial retrograde signaling: triggers, pathways, and outcomes. *Oxid Med Cell Longev.* 2015:1.

De Clercq I, Vermeirssen V, Van Aken O, Vandepoele K, Murcha MW, Law SR, Inze A, Ng S, Ivanova A, Rombaut D, et al. 2013. The membrane-bound NAC transcription factor ANAC013 functions in mitochondrial retrograde regulation of the oxidative stress response in *Arabidopsis*. *Plant Cell* 25(9): 3472–3490.

Hohmann N, Wolf EM, Lysak MA, Koch MA. 2015. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. *Plant Cell* 27(10): 2770–2784.

Hong JP, Adams E, Yanagawa Y, Matsui M, Shin R. 2017. AtSKIP18 and AtSKIP31, F-box subunits of the SCF E3 ubiquitin ligase complex, mediate the degradation of 14-3-3 proteins in *Arabidopsis*. *Biochem Biophys Res Commun.* 485(1): 174–180.

Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH. 2017. SUBA4: the interactive data analysis centre for *Arabidopsis* subcellular protein locations. *Nucleic Acids Res.* 45(D1): D1064–D1074.

Hooper CM, Tanz SK, Castleden IR, Vacher MA, Small ID, Millar AH. 2014. SUBAcon: a consensus algorithm for unifying the subcellular localization data of the *Arabidopsis* proteome. *Bioinformatics* 30(23): 3356–3364.

Huang H, Yoo CY, Bindbeutel R, Goldsworthy J, Tielking A, Alvarez S, Naldrett MJ, Evans BS, Chen M, Nusinow DA. 2016. PCH1 integrates circadian and light-signaling pathways to control photoperiod-responsive growth in *Arabidopsis*. *Elife* 5:e13292.

Hummel M, Cordewener JH, de Groot JC, Smeekens S, America AH, Hanson J. 2012. Dynamic protein composition of *Arabidopsis thaliana* cytosolic ribosomes in response to sucrose feeding as revealed by label free MSE proteomics. *Proteomics* 12(7): 1024–1038.

Huynen MA, Duarte I, Szklarczyk R. 2013. Loss, replacement and gain of proteins at the origin of the mitochondria. *Biochim Biophys Acta* 1827(2): 224–231.

Jeong CW, Roh H, Dang TV, Choi YD, Fischer RL, Lee JS, Choi Y. 2011. An E3 ligase complex regulates SET-domain polycomb group protein activity in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 108(19): 8036–8041.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4): 772–780.

Katsir L, Schillmiller AL, Staswick PE, He SY, Howe GA. 2008. COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. *Proc Natl Acad Sci U S A.* 105(19): 7100–7105.

Kepinski S, Leyser O. 2005. The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature* 435(7041): 446–451.

Kim SY, Kim SG, Kim YS, Seo PJ, Bae M, Yoon HK, Park CM. 2007. Exploring membrane-associated NAC transcription factors in

- Arabidopsis*: implications for membrane biology in genome regulation. *Nucleic Acids Res.* 35(1): 203–213.
- Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci.* 279(1749): 5048–5057.
- Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani-Covo E, McInerney JO, Landan G, Martin WF. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524(7566): 427–432.
- Kulkarni SR, Vanechoutte D, Van de Velde J, Vandepoele K. TF2Network: predicting transcription factor regulators and gene regulatory networks in *Arabidopsis* using publicly available binding site information. *Nucleic Acids Res.* 2018;46(6):e31.
- Kuroda H, Yanagawa Y, Takahashi N, Horii Y, Matsui M. 2012. A comprehensive analysis of interaction and localization of *Arabidopsis* SKP1-like (ASK) and F-box (FBX) proteins. *PLoS One* 7(11): e50009.
- Leister D. 2016. Towards understanding the evolution and functional diversification of DNA-containing plant organelles. *F1000 Res.* 5:330.
- Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature.* 2018;557(7703):101–105.
- Ng S, De Clercq I, Van Aken O, Law SR, Ivanova A, Willems P, Giraud E, Van Breusegem F, Whelan J. 2014. Anterograde and retrograde regulation of nuclear genes encoding mitochondrial proteins during growth, development, and stress. *Mol Plant* 7(7): 1075–1093.
- Ng S, Ivanova A, Duncan O, Law SR, Van Aken O, De Clercq I, Wang Y, Carrie C, Xu L, Kmiec B, et al. 2013. A membrane-bound NAC transcription factor, ANAC017, mediates mitochondrial retrograde signaling in *Arabidopsis*. *Plant Cell* 25(9): 3450–3471.
- O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. 2016. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 166(6): 1598.
- Palmfeldt J, Bross P. 2017. Proteomics of human mitochondria. *Mitochondrion* 33:2–14.
- Pich IRO, Kondrashov FA. 2014. Long-term asymmetrical acceleration of protein evolution after gene duplication. *Genome Biol Evol.* 6(8): 1949–1955.
- Proost S, Fostier J, De Witte D, Dhoedt B, Demeester P, Van de Peer Y, Vandepoele K. 2012. i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* 40(2): e11.
- Proost S, Van Bel M, Vanechoutte D, Van de Peer Y, Inze D, Mueller-Roeber B, Vandepoele K. 2015. PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* 43(Database issue): D974–D981.
- Quint M, Drost HG, Gabel A, Ullrich KK, Bonn M, Grosse I. 2012. A transcriptomic hourglass in plant embryogenesis. *Nature* 490(7418): 98–101.
- Rao RS, Salvato F, Thal B, Eubel H, Thelen JJ, Moller IM. 2017. The proteome of higher plant mitochondria. *Mitochondrion* 33:22–37.
- Rathore A, Chu Q, Tan D, Martinez TF, Donaldson CJ, Diedrich JK, Yates JR, Saghatelian A. 2018. MIEF1 microprotein regulates mitochondrial translation. *Biochemistry* 57(38): 5564–5575.
- Risseuw EP, Daskalchuk TE, Banks TW, Liu E, Cotelesage J, Hellmann H, Estelle M, Somers DE, Crosby WL. 2003. Protein interaction analysis of SCF ubiquitin E3 ligase subunits from *Arabidopsis*. *Plant J.* 34(6): 753–767.
- Roger AJ, Munoz-Gomez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. *Curr Biol.* 27(21): R1177–R1192.
- Ruprecht C, Proost S, Hernandez-Coronado M, Ortiz-Ramirez C, Lang D, Rensing SA, Becker JD, Vandepoele K, Mutwil M. 2017. Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. *Plant J.* 90(3): 447–465.
- Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A.* 95(11): 5857–5864.
- Senkler J, Senkler M, Eubel H, Hildebrandt T, Lengwenus C, Schertl P, Schwarzlander M, Wagner S, Wittig I, Braun HP. 2017. The mitochondrial complexome of *Arabidopsis thaliana*. *Plant J.* 89(6): 1079–1092.
- Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551): 173–179.
- Szklarczyk R, Huynen MA. 2009. Expansion of the human mitochondrial proteome by intra- and inter-compartmental protein duplication. *Genome Biol.* 10(11): R135.
- Szklarczyk R, Huynen MA. 2010. Mosaic origin of the mitochondrial proteome. *Proteomics* 10(22): 4012–4024.
- Van Aken O, De Clercq I, Ivanova A, Law SR, Van Breusegem F, Millar AH, Whelan J. 2016. Mitochondrial and chloroplast stress responses are modulated in distinct touch and chemical inhibition phases. *Plant Physiol.* 171(3): 2150–2165.
- Van Aken O, Ford E, Lister R, Huang S, Millar AH. 2016. Retrograde signalling caused by heritable mitochondrial dysfunction is partially mediated by ANAC017 and improves plant performance. *Plant J.* 88(4): 542–558.
- Van Aken O, Pečenková T, van de Cotte B, De Rycke R, Eeckhout D, Fromm H, De Jaeger G, Witters E, Beemster GTS, Inzé D, et al. 2007. Mitochondrial type-I prohibitins of *Arabidopsis thaliana* are required for supporting proficient meristem development. *Plant J.* 52(5): 850–864.
- Van Aken O, Zhang B, Law S, Narsai R, Whelan J. 2013. AtWRKY40 and AtWRKY63 modulate the expression of stress-responsive nuclear genes encoding mitochondrial and chloroplast proteins. *Plant Physiol.* 162(1): 254–271.
- Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y, Coppens F, Vandepoele K. 2018. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* 46(D1): D1190–D1196.
- Vanechoutte D, Estrada AR, Lin YC, Loraine AE, Vandepoele K. 2017. Genome-wide characterization of differential transcript usage in *Arabidopsis thaliana*. *Plant J.* 92(6): 1218–1231.
- Yao R, Ming Z, Yan L, Li S, Wang F, Ma S, Yu C, Yang M, Chen L, Chen L, et al. 2016. DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature* 536(7617): 469–473.
- Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ. 2008. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* 3(4): e1994.