Apiaceae *FNS I* originated from *F3H* through tandem gene duplication

Boas Pucker[1,2] and Massimo Iorizzo[3,4]

[1] Institute of Plant Biology, TU Braunschweig, Braunschweig, Germany

[2] BRICS, TU Braunschweig, Braunschweig, Germany

[3] Plants for Human Health Institute, NC State University, Kannapolis, North Carolina, USA

[4] Department of Horticultural Science, NC State University, Raleigh, North Carolina, USA

Corresponding authors: b.pucker@tu-braunschweig.de; miorizz@ncsu.edu

Abstract

Background: Flavonoids are specialized metabolites with numerous biological functions in stress response and reproduction of plants. Flavones are one subgroup that is produced by the flavone synthase (FNS). Two distinct enzyme families evolved that can catalyze the biosynthesis of flavones. While the membrane-bound FNS II is widely distributed in seed plants, one lineage of soluble FNS I appeared to be unique to Apiaceae species.

Results: We show through phylogenetic and comparative genomic analyses that Apiaceae *FNS I* evolved through tandem gene duplication of flavanone 3-hydroxylase (*F3H)* followed by neofunctionalization. Currently available datasets suggest that this event happened within the Apiaceae in a common ancestor of *Daucus carota* and *Apium graveolens*. The results also support previous findings that *FNS I* in the Apiaceae evolved independent of *FNS I* in other plant species.

Conclusion: We validated a long standing hypothesis about the evolution of Apiaceae FNS I and predicted the phylogenetic position of this event. Our results explain how an Apiaceae-specific *FNS I* lineage evolved and confirm independence from other *FNS I* lineages reported in non-Apiaceae species.

Keywords

Flavone synthase, flavanone 3-hydroxylase, flavonoid biosynthesis, carrot, comparative genomics, neofunctionalization, evolution, tandem gene duplication

Introduction

A plethora of specialized metabolites including flavonoids are produced by plants. These compounds provide an evolutionary advantage under certain environmental conditions. Flavonoids are produced in response to stresses like ultra violet (UV) radiation, cold, or drought [1, 2]. Especially visible is the pigmentation of flowers and fruits by anthocyanins which are one subclass of the flavonoids [3, 4]. Other subclasses include the proanthocyanidins which contribute to the pigmentation of seed coats [5] or flavonols which are produced in response to UV stress [6]. These branches of the flavonoid biosynthesis are well studied and conserved in many plant species and represent a model system for the investigation of the specialized metabolism in plants. A less conserved branch of the flavonoid biosynthesis leads to flavones (Fig 1), which are important in signaling and defense against pathogens [7]. Flavones are derivatives of phenylalanine which is channeled through the general phenylpropanoid pathway to the chalcone synthase (CHS). This enzyme is the first committed step of the flavonoid biosynthesis. Chalcone isomerase (CHI) and flavone synthase (FNS) represent the following steps involved in the formation of the flavone apigenin. F3'H can convert naringenin into eriodictyol which serves as substrate for the formation of the flavone luteolin (Fig 1). FNS activity evolved independently in different phylogenetic lineages [8], and to date two types of FNS genes have been described, named FNSI and FNSII. Distributed over a wide phylogenetic range is FNS II, a membrane-bound cytochrome P450-dependent monooxygenase [9]. An independent lineage of FNS I, a soluble Fe^{2+}/2-oxoglutarate-dependent dioxygenase (2-ODD), was identified in the Apiaceae and appeared to be restricted to that family [10]. However, other studies report FNS I functionality in other plant species like OsFNSI in *Oryza sativa* [11], EaFNSI in *Equisetum arvense* [12], PaFNSI in *Plagiochasma appendiculatum* [13], AtDMR6 in *Arabidopsis thaliana* [14], and ZmFNSI-1 in *Zea mays* [14]. These lineages were presented as independent evolutionary events and are not orthologs of Apiaceae *FNS I* [8, 15]. Recently, a study revealed that FNS I is widely distributed in liverworts and is the most likely origin of seed plant flavanone 3-hydroxylase (F3H) [8]. Reports of enzymes with multiple functions like F3H/FNS I [8, 10] or F3H/FLS [16, 17] indicate that the 2-ODD family has a high potential for the acquisition of new functionalities and that independent evolution of these new functions might happen frequently.

It is assumed that flavone biosynthesis is an evolutionarily old trait that predates flavonol and anthocyanin biosynthesis, because the ancestor of the F3H was probably a FNS I [8]. Minor changes in sequence and protein structure can determine the change in enzyme function. One particularly important residue is Y240 in the liverwort *Plagiochasma appendiculatum* PaFNSI/F2H [13]. Bifunctional *Physcomitrella patens* and *Selaginella moellendorffii* enzymes show M or F residues at this site. Most angiosperms and gymnosperms show a P at the corresponding position of their F3Hs [8]. Substitution of this P by M or F resulted in reduced F3H activity and increased FNS I activity, while a replacement with Y resulted in dominant FNS I activity [8]. This indicates that this site played a crucial role in the transition from FNS I to F3H activity.

Apiaceae FNS I show high sequence similarity to F3H thus both were previously classified as DOXC28 in a systematic investigation of the 2-ODD family [18]. Another study called this group of 2-ODD sequences 'POR', because they are NADPH-cytochrome P450 oxidoreductases [19]. It was also hypothesized that Apiaceae *FNS I* evolved from *F3H* of seed plants by duplication and subsequent divergence [10, 19, 20].

F3H and FNS I accept the same substrate (Fig 1) which suggests that competition takes place if both enzymes are present at the same intracellular location. The specific activity of both enzymes in the Apiaceae is defined by a small number of diagnostic amino acid residues [8, 10]. It is important to note that the P/Y substitution [8] described above does not play a role in the Apiaceae, because FNS I and F3H sequences show a conserved P at this position. Substitution of several other amino acids in F3H results in FNS I activity though [10]. For instance, I131F, M106T, and D195E are sufficient to confer partial FNS I function to F3H [10]. Also, I131F with L215V and K216R can be sufficient to confer FNS I functionality [10]. A substitution of these seven amino acid residues substantially modifies the pocket of the active site hence changing the orientation of the substrate [10]. This is expected to cause a syn-elimination of hydrogen from carbon-2 (FNS activity) instead of hydroxylation of carbon-3 (F3H activity) [10].

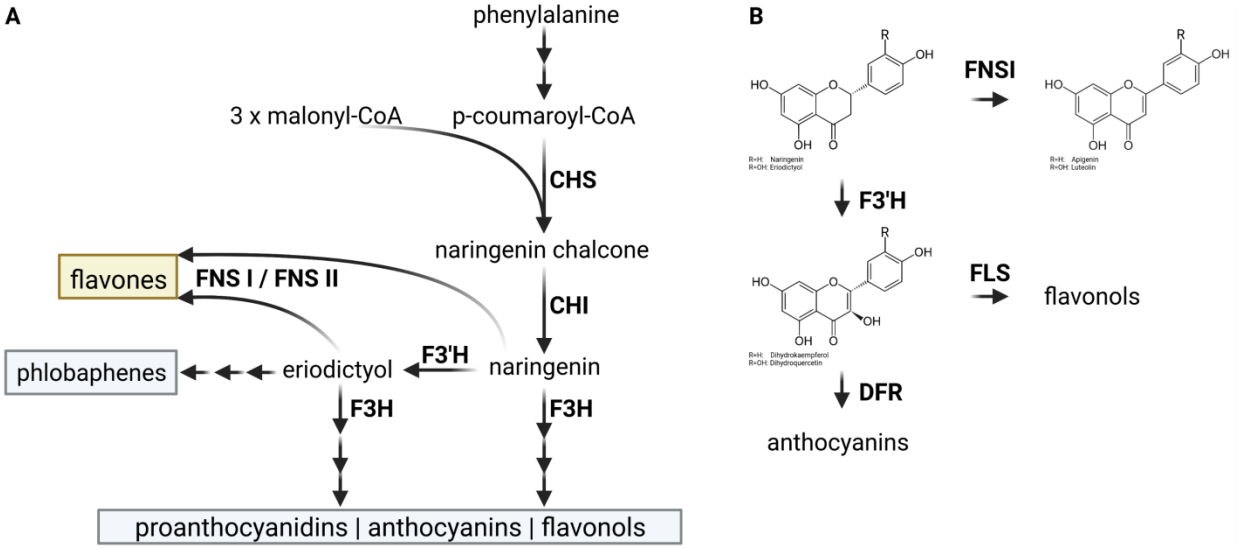

Fig 1. (A) Simplified illustration of the flavonoid biosynthesis with focus on the flavone biosynthesis. CHS (naringenin-chalcone synthase), CHI (chalcone isomerase), F3H (flavanone 3-hydroxylase), F3'H (flavonoid 3'-hydroxylase), and FNS (flavone synthase). (B) Reactions catalyzed by FNS I and F3H in the context of the flavonoid biosynthesis.

Although previous work hypothesized that Apiaceae FNS I originated from F3H through duplication and neofunctionlization [10, 19, 20], this hypothesis has not yet been validated. The recent release of high quality genome sequences representing most angiosperm lineages including members of the Apiacea family [21–24] opens the opportunity to address this hypothesis. Here, we investigated the evolution of FNS I in the Apiaceae through phylogenetic analysis and comparative genomics. The results indicate that *FNS I* originated from a tandem duplication of *F3H* that was followed by a neofunctionalization event.

Methods

Datasets

The genome sequences and the corresponding annotation of *Daucus carota* 388_v2.0 [21] and *Panax ginseng* GCA_020205605.1 [22] were retrieved from Phytozome [25]. The genome sequences of *Apium graveolens* GCA_009905375.1 [23] and *Centella asiatica* GCA_014636745.1 [24] were downloaded from NCBI. Sequences of F3H and FNS I were retrieved from the KIPEs v1 data set [26] and are included in S1 File. The phylogenetic relationships of Apiaceae species were inferred from a previously constructed species tree [27]. This tree was used to arrange genome sequences in the synteny analysis.

Gene prediction

Since no complete annotation of the coding sequences was publicly available for *Apium graveolens* and *Centella asiatica*, we applied AUGUSTUS v3.3 [28] for an *ab initio* gene prediction with previously described settings [29]. The *Daucus carota* annotation of *F3H* and *FNS I* was manually checked in the Integrated Genomics Viewer [30] and revised (S1 File). Polishing of the gene models was based on a TBLASTN v2.8.1 [31] alignment of the *Petroselinum crispum* FNS I sequence against the *D. carota* genome sequence. Additionally, RNA-seq reads were retrieved from the Sequence Read Archive (S2 File) and aligned to the *D. carota* genome sequence using STAR v2.7.3a [32] with previously described parameters [33].

Alignment and phylogenetic tree construction

F3H and FNS I sequence collections [26] were used to search for additional candidates in *C. asiatica*, *D. carota*, and *A. graveolens* using a BLAST-based Python script [34]. Initial candidates were validated through a phylogeny constructed with FastTree v2.1.10 (-wag -nosupport) [35] based on a MAFFT v7.475 [36] alignment of the polypeptide sequences. Additional phylogenies were constructed with different approaches to validate important relationships. MAFFT v7.475 [36] and MUSCLE v5.1 [37] were applied for the alignment construction with default parameters. A customized Python script (algntrim.py, [34]) was applied to remove alignment columns with less than 10% occupancy. The maximum likelihood tree displayed in Fig 2 is based on the MAFFT alignment running RAxML-NG v1.0.1 [38] with LG+G8+F and 10300 rounds of bootstrapping. RAxML-NG was also run with the same model based on a MUSCLE5 alignment of the same sequences. Phylogenies for both alignments were also generated with FastTree2 [35] (-wag), IQ-TREE v1.6.12 (-alrt 1000 -bb 1000)[39, 40], and MEGA v11.0.13 [41] (neighbor-joining, 1000 bootstrap replicates, poisson model, uniform rates, pairwise deletion). The resulting tree topologies were manually compared to validate (1) important nodes, (2) monophyly of FNS I, and (3) position of the FNS I clade within the F3H clade of the Apiaceae.

Synteny analysis

JCVI/MCscan [42] was applied to compare the genome sequences of *P. ginseng*, *C. asiatica*, *D. carota*, and *A. graveolens*. The region around *F3H* and *FNS I* was manually selected. Connections of genes between the species were manually validated and revised based on phylogenetic trees (S3 File).

132 TBLASTN v2.8.1 (-evalue 0.00001) [43] was run with the *P. crispum* FNS I against the genome sequence
133 of *C. asiatica* and *A. graveolens* to identify gene copies that might be missing in the annotation. The
134 results of this search were compared against the annotation to find BLAST hits outside of gene models
135 [34]. The best hits were assessed in a phylogenetic tree with previously characterized F3H and FNS I
136 sequences.

137 ### Gene expression analysis

138 Paired-end RNA-seq data sets were retrieved from the Sequence Read Archive via fastq-dump v2.8.1
139 [44] (S2 File). kallisto v0.44 [45] was applied with default parameters for the quantification of gene
140 expression (counts and TPMs). A Python script was developed for the generation of violin plots to
141 illustrate the variation of gene expression (TPMs) across various samples [34]. Outliers, defined as data
142 points which are more than three interquartile ranges away from the median, were excluded from this
143 visualization. Co-expression was analyzed by calculating pairwise Spearman correlation coefficients of
144 gene expression values across all samples. Lowly expressed genes were excluded and only pairs with a
145 correlation coefficient >0.65 and an adjusted p-value < 0.05 were reported. Functional annotation of the
146 *Daucus carota* genes (Dcarota_388_v2.0) was inferred from *Arabidopsis thaliana* based on reciprocal
147 best BLAST hits of the representative peptide sequences as previously described [29, 46]. This co-
148 expression analysis was implemented in a Python script (coexp3.py) which is available via github [34]
149 and as an online service [47].

150

151 ## Results

152 Apiaceae FNS I sequences show high similarity to F3H that suggest a close phylogenetic relationship of
153 both lineages. For example, *D. carota* FNS I and F3H have 78% identical amino acids, while the
154 proportion of identical amino acids between different FNS I candidates is 80% (S4 File). A phylogenetic
155 tree was constructed to visualize the relationship of all these sequences in a larger context. FNS I
156 sequences of the non-Apiaceae species *Arabidopsis thaliana*, *Oryza sativa*, *Zea mays*, and *Parmotrema*
157 *appendiculatum* clustered outside the F3H clade in this tree. The FNS I sequences of seven Apiaceae
158 species formed a distinct clade (Fig 2). This FNS I clade is embedded within a large clade of F3H
159 sequences that included a wide range of phylogenetically distant plants. The position of the FNS I
160 sequences within the F3H clade suggests that Apiaceae *FNS I* originated from *F3H*. The pattern also
161 supports a single *FNS I* origin within the Apiaceae. The critical node separating Apiaceae FNS I from
162 Apiaceae F3H is well supported in the results of all applied tools (S3 File). The monophyly of the
163 Apiaceae FNS I clade is also well supported in all analyses. The FNS I sequences of non-Apiaceae species
164 seem to have an independent origin.

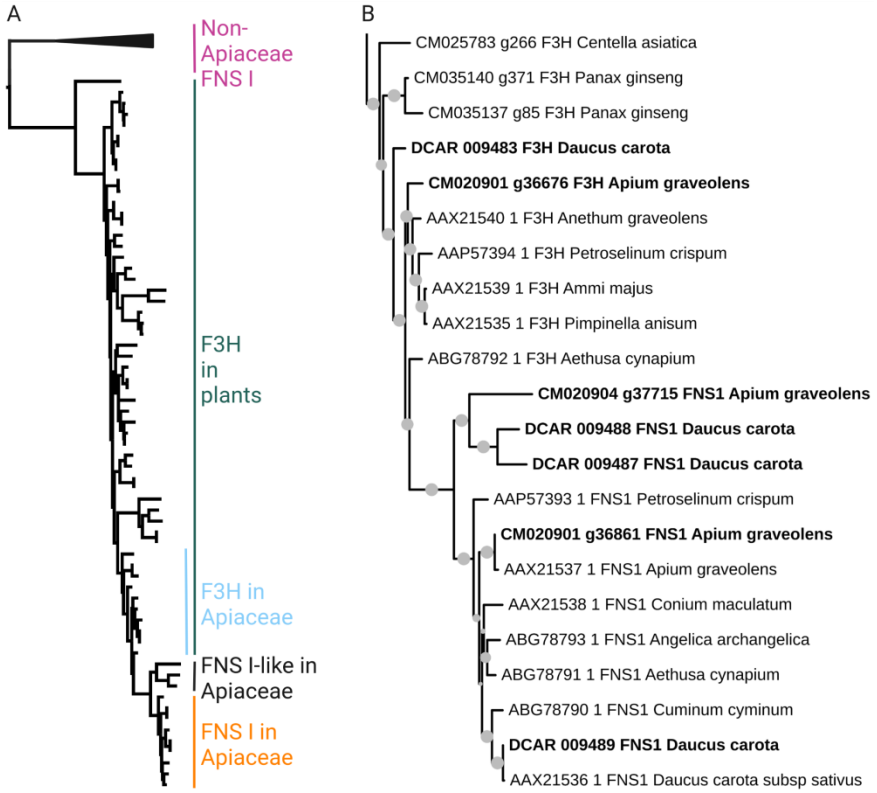

Fig 2. (A) Maximum likelihood tree of F3H and FNS I sequences. Apiaceae FNS I sequences form a nested cluster within the F3H context. The FNS I sequences of non-Apiaceae species are placed outside the F3H clade. Polypeptide sequences used for the construction of this tree are included in S1 File. A full phylogenetic tree with sequence names is included as S3 File. (B) Apiaceae clade in the maximum likelihood tree of F3H and FNS I sequences. Sequences analyzed in Table 1 are written in bold. Classification of sequences in the Apiaceae F3H, Apiaceae FNS I, and Apiaceae FNS I-clades were supported by RAxML, FastTree, MEGA, and IQ-TREE analyses. Support for the individual branches is indicated by the size of grey circles.

A previous study identified diagnostic amino acid residues that determine the FNS or F3H activity, respectively [10]. It was demonstrated that a substitution of selected amino acid residues can convert one enzyme into the other. We inspected these characteristic features of the FNS I and F3H sequences of *Daucus carota* and *Apium graveolens* (Table 1). The results suggest that there is one *bona fide* F3H in *D. carota* (DCAR_009483) and *A. graveolens* (CM020901_g36676), respectively (S1 File). We also identified one FNS I in each of these species: DCAR_009489 and CM020901_g36861, respectively. In addition, there are FNS I-like copies which lack some of the functionally important amino acid residues of a *bona fide* FNS I (Table 1). Since the residues do not always match F3H, we cannot tell whether (1) these sequences have lost their FNS function in secondary events or (2) represent intermediates in the evolution from F3H towards FNS I. However, the incorporation of additional residues places these sequences in an intermediate clade (Fig 2). The separation of the *FNS I*-like lineage from the *FNS I*

186 lineage seems to predate a duplication in the *FNS I*-like lineage that produced the two copies discovered
187 in *D. carota* and *A. graveolens*.

188

189

190 **Table 1.** Inspection of diagnostic amino acid residues in FNS I and F3H candidates of *Daucus carota* and
191 *Apium graveolens*. FNS I residues are highlighted in orange, F3H residues are highlighted in skyblue.
192 Positions are based on the FNS I of *Petroselium crispum* (AAP57393.1).

Sequence	**Name**	**106**	**115**	**116**	**131**	**195**	**200**	**215**	**216**
DCAR_009489	*DcFNS I*	T	T	I	F	E	I	V	R
DCAR_009487	*DcFNS I- like*	P	I	V	F	E	I	C	R
DCAR_009488	*DcFNS I-like*	T	T	V	F	E	I	V	R
DCAR_009483	*DcF3H*	M	I	V	I	D	V	L	K
CM020904_g37715	*AgFNS I-like*	T	T	I	F	K	I	C	R
CM020901_g36861	*AgFNS I*	T	T	I	F	E	I	V	R
CM020901_g36676	*AgF3H*	M	I	V	I	D	V	L	K

193

194

195 To narrow down the origin of the Apiaceae *FNS I*, we compared highly contiguous genome sequences of
196 Apiaceae and outgroup species. The Apiaceae members *Daucus carota* and *Apium graveolens* show
197 microsynteny in a region that harbors both, *F3H* and *FNS I* genes (Fig 3). Both species differ from the
198 *Centella asiatica* (basal Apiaceae species) and *Panax ginseng* (outgroup species) which do not show a
199 *FNS I* gene in this region or elsewhere in the genome sequence. However, the presence of *F3H* and
200 synteny of many flanking genes indicates that the correct region was analyzed.

201

202

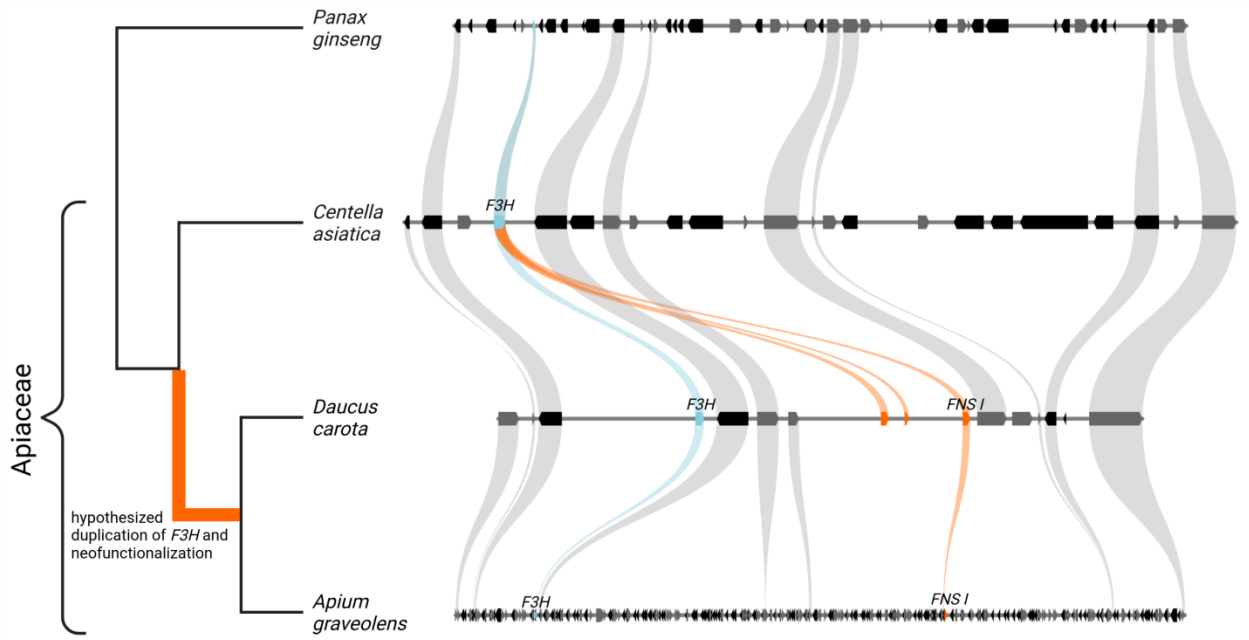

Fig 3. Syntenic region between the Apiaceae species *Daucus carota* and *Apium graveolens* shows *F3H* (skyblue) and *FNS I* (orange) in close proximity, while *FNS I* was not observed in the basal Apiaceae species *Centella asiatica* or in the outgroup *Panax ginseng*.

Although multiple gene copies were identified based on the available genome sequences, expression of these genes determines their relevance. Expression of the *F3H*, *FNS I*, and *FNS I*-like genes in carrots was analyzed across 146 RNA-seq samples (Fig 4). The results show that *F3H* and *FNS I* show substantially higher expression than any of the *FNS I*-like genes (DCAR_009487) while the other *FNS I*-like gene (DCAR_009488) is almost not expressed. Expression analysis in specific plant parts and tissues revealed that *F3H* (DCAR_009483) is strongly expressed in phloem and flowers, while *FNS I* (DCAR_009489) was dominant in leaf, petiole, and root (Fig 4). Expression patterns of both *FNS I*-like genes are more similar to *FNS I* than to *F3H* expression. Strongest expression of DCAR_009487 was observed in the phloem and xylem of the root and in the petiole. DCAR_009488 showed the highest expression in whole flowers and stressed leaves of orange cultivars (S2 File). Co-expression analyses results are only available for *F3H* and *FNS I*, because the expression of the *FNS I*-like genes is too low to be analyzed. F3H shows a strong co-expression with genes of the anthocyanin biosynthesis including *CHS*, *CHI*, *F3'H*, *DFR*, *LDOX*, several glycosyl transferase encoding genes, and a potential anthocyanin transporter encoding gene (S5 File). In contrast, *FNS I* shows a strong association with the flavonol biosynthesis most notable is the co-expression with *FLS1* (S6 File).

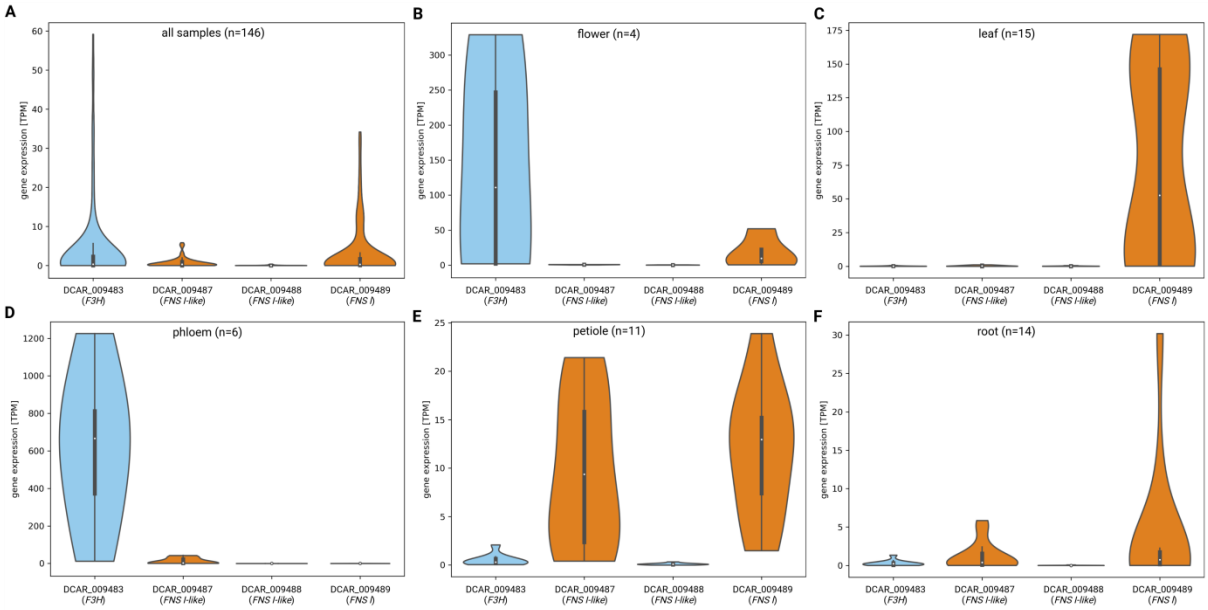

Fig 4. Expression of *F3H*, *FNS I*, and *FNS I*-like genes in carrots. (A) This plot shows the distribution of transcript per million (TPM) values across 146 RNA-seq data sets derived from different tissues/organs and conditions (S2 File). These aggregated expression data reveal that *F3H* (DCAR_009483) and *FNS I* (DCAR_009489) are expressed in several samples, while *FNS I*-like (DCAR_009487) is only weekly expressed in a few selected samples and *FNS I*-like (DCAR_009488) is almost not expressed at all. (B-F) Plots show the expression of the four genes in specific tissues covering flower, leaf, phloem, petiole, and root. This tissue-specific gene expression analysis was restricted to samples with available metadata about the respective sample.

Discussion

We provide genomic and phylogenetic evidence for the evolution of the Apiaceae *FNS I* from *F3H* through tandem duplication followed by neofunctionalization. These results support a hypothesis about the evolution of *FNS I* from *F3H* [20, 48] and narrow down the initial duplication event. The phylogenetic analysis provides strong support for a single event of Apiaceae FNS I evolution. The nested position of the Apiaceae FNS I clade within the Apiaceae F3H clade is also well supported, while some relationships between F3H sequences of the angiosperms have only low or moderate support.

We show that the F3H duplication most likely took place in a shared ancestor of *D. carota* and *A. graveolens,* and is probably not shared with all members of the Apiacea family as previously hypothesized. Since there is no evidence for this gene duplication in *C. asiatica* which branches early in the Apiaceae, we hypothesize that the F3H duplication took place after the separation of *C. asiatica* from the Daucus/Apium lineage (Fig 3). Additional genome sequences will help to support this hypothesis and to narrow down the precise duplication event within the Apiaceae lineage.

246 The inspection of conserved amino acid residues in *D. carota* and *A. graveolens* candidate sequences
247 confirmed the presence of one F3H and FNS I in each species. Additionally, both species have at least
248 one sequence that lacks some of the functionally important amino acid residues of a *bona fide* FNS I
249 without having all residues of a F3H (Table 1). This might indicate a different enzymatic activity or
250 promiscuity of these enzymes. It is striking to see that the F3H and FNS I cannot be distinguished based
251 on a P/Y substitution at position 240 (based on *Plagiochasma appendiculatum* PaFNSI/F2H). This
252 difference was previously reported between ancestral FNS I sequences and the sequences of
253 monocot/dicot F3H sequences [8]. All Apiaceae FNS I and F3H sequences show a conserved P at this
254 position suggesting that two independent shifts between F3H and FNS I activity are possible. According
255 to substitution experiments [10], the presence of T106, F131, and E195 in DCAR_009488 indicates that
256 this enzyme has at least some basal FNS activity. It is possible that FNS I-like enzymes have multiple
257 activities [8]. Only based on the diagnostic amino acid residues, we cannot tell whether (1) these
258 sequences have lost their FNS function in secondary events or (2) represent intermediates in the
259 evolution from F3H towards FNS I. However, the incorporation of additional residues places these
260 sequences in an intermediate clade (Fig 2). Based on their phylogenetic relationship, we hypothesize
261 that two *FNS I* copies were present in the common ancestor of *D. carota* and *A. graveolens*. One of these
262 copies was again duplicated in *D. carota* after separation of the lineages leading to *D. carota* and *A.*
263 *graveolens* hence explaining the presence of three copies in *D. carota*. The preservation of these
264 sequences since the separation of both species indicates a relevance of these FNS I-like sequences. The
265 expression analysis suggests that these genes are active in specific tissues like petiole and root where
266 *FNS I* is also active.

267 The physical clustering of *FNS I* and *F3H* in the genome could be due to the recent tandem duplication.
268 However, it could be interesting to investigate whether this clustering does also provide an evolutionary
269 benefit. Biosynthetic gene clusters (BGCs) were previously described in numerous plant species [49, 50].
270 These BGCs are often associated with an evolutionary young trait that provides a particular advantage
271 e.g. in the defense against a pathogen [49]. Given the relevance of flavones in the defense against
272 pathogens [7, 51], it seems possible that the flavone biosynthesis could be a similar trait that evolved in
273 the Apiaceae.

274 *FNS I* genes were also discovered in a small number of non-Apiaceae species [11, 13, 14]. However,
275 these genes belong to an independent *FNS I* lineage [8]. As more high quality genome sequences of seed
276 plants are released, a systematic search for additional non-Apiaceae *FNS I* sequences could become
277 feasible in the near future. The number of independent FNS I origins remains unknown. Exploration and
278 comparison of additional *FNS I* lineages across plants has the potential to advance our understanding of
279 enzyme evolution.

280

281 Conclusions

282 In conclusion here we uncovered the duplication mechanism that gave rise to *FNS I* within the Apiaceae
283 family. The gene probably evolved from a tandem duplication of F3H followed by neofunctionalization.

284 The origin of Apiacea *FNS I* appears to be independent from *FNS I* genes described in *Arabidopsis*
285 *thaliana*, *Oryza sativa*, and *Zea mays*.

286

287

288 # Declarations

289 ## Ethics approval and consent to participate
290 Not applicable.

291 ## Consent for publication
292 Not applicable.

293 ## Availability of data and materials
294 All datasets underlying this study are publicly available or included within the additional files. Scripts
295 developed for this work are freely available on github: https://github.com/bpucker/ApiaceaeFNS1.

296 ## Competing interests
297 The authors have declared that no competing interests exist.

298 ## Funding
299 MI was supported by the United States Department of Agriculture National Institute of Food and
300 Agriculture, Hatch project 1008691 and award number 2022-67013-36389. We acknowledge support by
301 the Open Access Publication Funds of Technische Universität Braunschweig.

302 ## Authors' contribution
303 BP and MI performed the analyses and wrote the manuscript.

304 ## Acknowledgements
305 Many thanks to the German network for bioinformatics infrastructure (de.NBI, grant 031A533A) and the
306 Bioinformatics Resource Facility (BRF) at the Center for Biotechnology (CeBiTec) at Bielefeld University
307 for providing an environment to perform the computational analyses. We used bioRender.com for the
308 construction of some figures. We thank Hanna Marie Schilbert for discussion and comments on the
309 manuscript.

310

311 # Supporting information
312 **S1 File**. Collection of F3H, FNS I, and FNS I-like sequences that were used for the analyses of this study.

313 **S2 File**. Gene expression values (TPMs) of *Daucus carota F3H*, *FNS I*, and *FNS I-like* genes.

314 **S3 File**. Phylogenetic trees of F3H, FNS I, and FNS I-like sequences. (A) Constructed by RAxML based on a
315 MAFFT alignment, (B) constructed by RAxML based on a MUSCLE5 alignment, (C) Constructed by
316 FastTree2 based on a MAFFT alignment, (D) Constructed by FastTree2 based on a MUSCLE5 alignment,

317 (E) constructed by IQ-TREE based on a MAFFT alignment, (F) constructed by IQ-TREE based on a
318 MUSCLE5 alignment, (G) constructed by MEGA based on a MAFFT alignment, and (H) constructed by
319 MEGA based on a MUSCLE5 alignment.

320 **S4 File**. Pairwise comparison of F3H and FNS I sequences. Percentage of identical amino acid residues is
321 displayed in this matrix.

322 **S5 File**. Results of a co-expression analysis of *F3H* (DCAR_009483) in *Daucus carota* based on 146 RNA-
323 seq data sets.

324 **S6 File**. Results of a co-expression analysis of *FNS I* (DCAR_009489) in *D. carota* based on 146 RNA-seq
325 data sets.

326

327 ## References

328

329 1. Winkel-Shirley B. Biosynthesis of flavonoids and effects of stress. Current Opinion in Plant Biology.
330 2002;5:218–23.

331 2. Nakabayashi R, Mori T, Saito K. Alternation of flavonoid accumulation under drought stress in
332 Arabidopsis thaliana. Plant Signal Behav. 2014;9:e29518.

333 3. Winkel-Shirley B. Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology,
334 and Biotechnology. Plant Physiology. 2001;126:485–93.

335 4. Grotewold E. The genetics and biochemistry of floral pigments. Annu Rev Plant Biol. 2006;57:761–80.

336 5. Todd JJ, Vodkin LO. Pigmented Soybean (Glycine max) Seed Coats Accumulate Proanthocyanidins
337 during Development. Plant Physiology. 1993;102:663–70.

338 6. Emiliani J, Grotewold E, Falcone Ferreyra ML, Casati P. Flavonols protect Arabidopsis plants against
339 UV-B deleterious effects. Mol Plant. 2013;6:1376–9.

340 7. Jiang N, Doseff AI, Grotewold E. Flavones: From Biosynthesis to Health Benefits. Plants. 2016;5:27.

341 8. Li D-D, Ni R, Wang P-P, Zhang X-S, Wang P-Y, Zhu T-T, et al. Molecular Basis for Chemical Evolution of
342 Flavones to Flavonols and Anthocyanins in Land Plants. Plant Physiol. 2020;184:1731–43.

343 9. Martens S, Forkmann G. Cloning and expression of flavone synthase II from Gerbera hybrids. The
344 Plant Journal. 1999;20:611–8.

345 10. Gebhardt YH, Witte S, Steuber H, Matern U, Martens S. Evolution of Flavone Synthase I from Parsley
346 Flavanone 3β-Hydroxylase by Site-Directed Mutagenesis. Plant Physiol. 2007;144:1442–54.

347 11. Lee YJ, Kim JH, Kim BG, Lim Y, Ahn J-H. Characterization of flavone synthase I from rice. BMB Rep.
348 2008;41:68–71.

349 12. Bredebach M, Matern U, Martens S. Three 2-oxoglutarate-dependent dioxygenase activities of
350 Equisetum arvense L. forming flavone and flavonol from (2S)-naringenin. Phytochemistry. 2011;72:557–
351 63.

352 13. Han X-J, Wu Y-F, Gao S, Yu H-N, Xu R-X, Lou H-X, et al. Functional characterization of a Plagiochasma
353 appendiculatum flavone synthase I showing flavanone 2-hydroxylase activity. FEBS Lett. 2014;588:2307–
354 14.

355 14. Falcone Ferreyra ML, Emiliani J, Rodriguez EJ, Campos-Bermudez VA, Grotewold E, Casati P. The
356 Identification of Maize and Arabidopsis Type I FLAVONE SYNTHASEs Links Flavones with Hormones and
357 Biotic Interactions. Plant Physiol. 2015;169:1090–107.

358 15. Wang Q-Z, Downie SR, Chen Z-X. Genome-wide searches and molecular analyses highlight the
359 unique evolutionary path of flavone synthase I (FNSI) in Apiaceae. Genome. 2018;61:103–9.

360 16. Prescott AG, Stamford NPJ, Wheeler G, Firmin JL. In vitro properties of a recombinant flavonol
361 synthase from Arabidopsis thaliana. Phytochemistry. 2002;60:589–93.

362 17. Schilbert HM, Schöne M, Baier T, Busche M, Viehöver P, Weisshaar B, et al. Characterization of the
363 Brassica napus Flavonol Synthase Gene Family Reveals Bifunctional Flavonol Synthases. Frontiers in
364 Plant Science. 2021;12.

365 18. Kawai Y, Ono E, Mizutani M. Evolution and diversity of the 2–oxoglutarate-dependent dioxygenase
366 superfamily in plants. The Plant Journal. 2014;78:328–43.

367 19. Andersen TB, Hansen NB, Laursen T, Weitzel C, Simonsen HT. Evolution of NADPH-cytochrome P450
368 oxidoreductases (POR) in Apiales – POR 1 is missing. Molecular Phylogenetics and Evolution.
369 2016;98:21–8.

370 20. Martens S, Forkmann G, Britsch L, Wellmann F, Matern U, Lukačin R. Divergent evolution of
371 flavonoid 2-oxoglutarate-dependent dioxygenases in parsley 1. FEBS Letters. 2003;544:93–8.

372 21. Iorizzo M, Ellison S, Senalik D, Zeng P, Satapoomin P, Huang J, et al. A high-quality carrot genome
373 assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat Genet.
374 2016;48:657–66.

375 22. Jiang Z, Tu L, Yang W, Zhang Y, Hu T, Ma B, et al. The chromosome-level reference genome assembly
376 for Panax notoginseng and insights into ginsenoside biosynthesis. Plant Communications.
377 2021;2:100113.

378 23. Li M-Y, Feng K, Hou X-L, Jiang Q, Xu Z-S, Wang G-L, et al. The genome sequence of celery (Apium
379 graveolens L.), an important leaf vegetable crop rich in apigenin in the Apiaceae family. Hortic Res.
380 2020;7:1–10.

381 24. Pootakham W, Naktang C, Kongkachana W, Sonthirod C, Yoocha T, Sangsrakru D, et al. De novo
382 chromosome-level assembly of the Centella asiatica genome. Genomics. 2021;113:2221–8.

383 25. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative
384 platform for green plant genomics. Nucleic Acids Res. 2012;40 Database issue:D1178–86.

385 26. Pucker B, Reiher F, Schilbert HM. Automatic Identification of Players in the Flavonoid Biosynthesis
386 with Application on the Biomedicinal Plant Croton tiglium. Plants. 2020;9:1103.

387 27. Downie SR, Katz-Downie DS, Watson MF. A phylogeny of the flowering plant family Apiaceae based
388 on chloroplast DNA rpl16 and rpoC1 intron sequences: towards a suprageneric classification of subfamily
389 Apioideae. Am J Bot. 2000;87:273–92.

390 28. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of
391 alternative transcripts. Nucleic Acids Res. 2006;34 suppl_2:W435–9.

392 29. Pucker B, Holtgräwe D, Weisshaar B. Consideration of non-canonical splice sites improves gene
393 prediction on the Arabidopsis thaliana Niederzenz-1 genome sequence. BMC Research Notes.
394 2017;10:667.

395 30. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative
396 Genomics Viewer. Nat Biotechnol. 2011;29:24–6.

397 31. Gertz EM, Yu Y-K, Agarwala R, Schäffer AA, Altschul SF. Composition-based statistics and translated
398 nucleotide searches: Improving the TBLASTN module of BLAST. BMC Biology. 2006;4:41.

399 32. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq
400 aligner. Bioinformatics. 2013;29:15–21.

401 33. Haak M, Vinke S, Keller W, Droste J, Rückert C, Kalinowski J, et al. High Quality de Novo
402 Transcriptome Assembly of Croton tiglium. Front Mol Biosci. 2018;5.

403 34. Pucker B. Apiaceae FNS I. https://github.com/bpucker/ApiaceaeFNS1. 2022.

404 35. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large
405 Alignments. PLOS ONE. 2010;5:e9490.

406 36. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in
407 Performance and Usability. Mol Biol Evol. 2013;30:772–80.

408 37. Edgar RC. High-accuracy alignment ensembles enable unbiased assessments of sequence homology
409 and phylogeny. 2022;:2021.06.20.449169.

410 38. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly
411 tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35:4453–5.

412 39. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model
413 selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.

414 40. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2:
415 New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology
416 and Evolution. 2020;37:1530–4.

417 41. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol
418 Biol Evol. 2021;38:3022–7.

419 42. Tang H, Krishnakumar V, Li J. jcvi: JCVI utility libraries (v0.5.7). 2015.
420 https://doi.org/10.5281/zenodo.31631.

421 43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol.
422 1990;215:403–10.

423 44. Leinonen R, Sugawara H, Shumway M, on behalf of the International Nucleotide Sequence Database
424 Collaboration. The Sequence Read Archive. Nucleic Acids Research. 2011;39 suppl_1:D19–21.

425 45. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat
426 Biotechnol. 2016;34:525–7.

427 46. Pucker B, Holtgräwe D, Sörensen TR, Stracke R, Viehöver P, Weisshaar B. A De Novo Genome
428 Sequence Assembly of the Arabidopsis thaliana Accession Niederzenz-1 Displays Presence/Absence
429 Variation and Strong Synteny. PLOS ONE. 2016;11:e0164321.

430 47. Rempel A, Pucker B. BioInfToolServer. BioInfToolServer. 2022. http://pbb.bot.nat.tu-bs.de/KIPEs/.
431 Accessed 31 May 2022.

432 48. Gebhardt Y, Witte S, Forkmann G, Lukacin R, Matern U, Martens S. Molecular evolution of flavonoid
433 dioxygenases in the family Apiaceae. Phytochemistry. 2005;66:1273–84.

434 49. Polturak G, Osbourn A. The emerging role of biosynthetic gene clusters in plant defense and plant
435 interactions. PLOS Pathogens. 2021;17:e1009698.

436 50. Polturak G, Liu Z, Osbourn A. New and emerging concepts in the evolution and function of plant
437 biosynthetic gene clusters. Current Opinion in Green and Sustainable Chemistry. 2022;33:100568.

438 51. Du Y, Chu H, Wang M, Chu IK, Lo C. Identification of flavone phytoalexins and a pathogen-inducible
439 flavone synthase II gene (SbFNSII) in sorghum. J Exp Bot. 2010;61:983–94.

440

441