**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 graveoleens*. 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 monoxygenase [9]. An independent lineage of FNS I, a soluble Fe²⁺/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 flavone 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 moellendorfii 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 oxioreductases [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](image-url)

**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 neofunctionalization [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 Apiaceae 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 Additional file 1. 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 (Additional file 1). 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 (Additional file 2) and aligned to the D. carota genome sequence using STAR v2.7.3a [32] with previously described parameters [33].

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 phylogenetic tree constructed with FastTree v2.1.10 (-wag -nosupport) [35]. A final tree was constructed with RAxML-NG v1.0.1 [36] using LG+G8+F and 10300 rounds of bootstrapping. The resulting tree topology was compared against trees constructed by FastTree (-wag) [35] and IQ-Tree v1.6.12 (-alrt 1000 -bb 1000) [37, 38].

Synteny analysis
JCVI/MCscan [39] 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 (Additional file 3). TBLASTN v2.8.1 (-evalue 0.00001) [40] was run with the P. crispum FNS I against the genome sequence of C. asiatica and A. graveolens to identify gene copies that might be missing in the annotation. The results of this search were compared against the annotation to find BLAST hits outside of gene models [34]. The best hits were assessed in a phylogenetic tree with previously characterized F3H and FNS I sequences.

Gene expression analysis
Paired-end RNA-seq data sets were retrieved from the Sequence Read Archive via fastq-dump v2.8.1 [41] (Additional file 2). kallisto v0.44 [42] was applied with default parameters for the quantification of gene expression (counts and TPMs). A Python script was developed for the generation of violin plots to
illustrate the variation of gene expression (TPMs) across various samples [34]. Outliers, defined as data points which are more than three interquartile ranges away from the median, were excluded from this visualization. Co-expression was analyzed by calculating pairwise Spearman correlation coefficients of gene expression values across all samples. Lowly expressed genes were excluded and only pairs with a correlation coefficient $>0.65$ and an adjusted p-value $< 0.05$ were reported. Functional annotation of the *Daucus carota* genes (Dcarota_388_v2.0) was inferred from *Arabidopsis thaliana* based on reciprocal best BLAST hits of the representative peptide sequences as previously described [29, 43]. This co-expression analysis was implemented in a Python script (coexp3.py) which is available via github [34] and as an online service [44].

**Results**

Apiaceae FNS I sequences show high similarity to F3H that suggest a close phylogenetic relationship of both lineages. For example, *D. carota* FNS I and F3H have 78% identical amino acids, while the proportion of identical amino acids between different FNS I candidates is 80% (Additional file 4). A phylogenetic tree was constructed to visualize the relationship of all these sequences in a larger context. FNS I sequences of the non-APIaceae species *Arabidopsis thaliana, Oryza sativa, Zea mays,* and *Parmotrema appendiculatum* clustered outside the F3H clade in this tree. The FNS I sequences of seven Apiaceae species formed a distinct clade (Fig. 2). This FNS I clade is embedded within a large clade of F3H sequences that included a wide range of phylogenetically distant plants. The position of the FNS I sequences within the F3H clade suggests that Apiaceae FNS I originated from F3H. The pattern also supports a single FNS I origin within the Apiaceae. The FNS I sequences of non-APIaceae species seem to have an independent origin.
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 (Additional file 1). 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
lineage seems to predate a duplication in the FNS I-like lineage that produced the two copies discovered in *D. carota* and *A. graveolens*.

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

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Name</th>
<th>106</th>
<th>115</th>
<th>116</th>
<th>131</th>
<th>195</th>
<th>200</th>
<th>215</th>
<th>216</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCAR_009489</td>
<td>DcFNS I</td>
<td>T</td>
<td>T</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>I</td>
<td>V</td>
<td>R</td>
</tr>
<tr>
<td>DCAR_009487</td>
<td>DcFNS I-like</td>
<td>P</td>
<td>I</td>
<td>V</td>
<td>F</td>
<td>E</td>
<td>I</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>DCAR_009488</td>
<td>DcFNS I-like</td>
<td>T</td>
<td>T</td>
<td>V</td>
<td>F</td>
<td>E</td>
<td>I</td>
<td>V</td>
<td>R</td>
</tr>
<tr>
<td>DCAR_009483</td>
<td>DcF3H</td>
<td>M</td>
<td>I</td>
<td>V</td>
<td>I</td>
<td>D</td>
<td>V</td>
<td>L</td>
<td>K</td>
</tr>
<tr>
<td>CM020904_g37715</td>
<td>AgFNS I-like</td>
<td>T</td>
<td>T</td>
<td>I</td>
<td>F</td>
<td>K</td>
<td>I</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td>CM020901_g36861</td>
<td>AgFNS I</td>
<td>T</td>
<td>T</td>
<td>I</td>
<td>F</td>
<td>E</td>
<td>I</td>
<td>V</td>
<td>R</td>
</tr>
<tr>
<td>CM020901_g36676</td>
<td>AgF3H</td>
<td>M</td>
<td>I</td>
<td>V</td>
<td>I</td>
<td>D</td>
<td>V</td>
<td>L</td>
<td>K</td>
</tr>
</tbody>
</table>

To narrow down the origin of the Apiaceae FNS I, we compared highly contiguous genome sequences of Apiaceae and outgroup species. The Apiaceae members *Daucus carota* and *Apium graveolens* show microsynteny in a region that harbors both, F3H and FNS I genes (Fig. 3). Both species differ from the *Centella asiatica* (basal Apiaceae species) and *Panax ginseng* (outgroup species) which do not show a FNS I gene in this region or elsewhere in the genome sequence. However, the presence of F3H and synteny of many flanking genes indicates that the correct region was analyzed.
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 (Additional file 2). 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 (Additional file 5). In contrast, *FNS I* shows a strong association with the flavonol biosynthesis most notable is the co-expression with *FLS1* (Additional file 6).
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 (Additional file 2). 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 weakly 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 confirm a hypothesis about the evolution of FNS I from F3H [20, 45] and narrow down the initial duplication event. We show that the F3H duplication took place in a shared ancestor of D. carota and A. graveolens, and is not shared with all members of the Apiaceae 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.

The inspection of conserved amino acid residues in D. carota and A. graveolens candidate sequences confirmed the presence of one F3H and FNS I in each species. Additionally, both species have at least one sequence that lacks some of the functionally important amino acid residues of a bona fide FNS I without having all residues of a F3H (Table 1). This might indicate a different enzymatic activity or promiscuity of these enzymes. It is striking to see that the F3H and FNS I cannot be distinguished based...
on a P/Y substitution at position 240 (based on *Plagiochasma appendiculatum* PaFNS/F2H). This difference was previously reported between ancestral FNS I sequences and the sequences of monocot/dicot F3H sequences [8]. All Apiacea FNS I and F3H sequences show a conserved P at this position suggesting that two independent shifts between F3H and FNS I activity are possible. According to substitution experiments [10], the presence of T106, F131, and E195 in DCAR_009488 indicates that this enzyme has at least some basal FNS activity. It is possible that FNS I-like enzymes have multiple activities [8]. Only based on the diagnostic amino acid residues, 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). Based on their phylogenetic relationship, we hypothesize that two FNS I copies were present in the common ancestor of *D. carota* and *A. graveolens*. One of these copies was again duplicated in *D. carota* after separation of the lineages leading to *D. carota* and *A. graveolens* hence explaining the presence of three copies in *D. carota*. The preservation of these sequences since the separation of both species indicates a relevance of these FNS I-like sequences. The expression analysis suggests that these genes are active in specific tissues like petiole and root where FNS I is also active.

The physical clustering of FNS I and F3H in the genome could be due to the recent tandem duplication. However, it could be interesting to investigate whether this clustering does also provide an evolutionary benefit. Biosynthetic gene clusters (BGCs) were previously described in numerous plant species [46, 47]. These BGCs are often associated with an evolutionary young trait that provides a particular advantage e.g. in the defense against a pathogen [46]. Given the relevance of flavones in the defense against pathogens [7, 48], it seems possible that the flavone biosynthesis could be a similar trait that evolved in the Apiaceae.

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

**Conclusions**

In conclusion here we uncovered the duplication mechanism that gave rise to *FNS I* within the Apiaceae family. The gene evolved from a tandem duplication of F3H followed by neofunctionalization. The origin of Apiaceae *FNS I* appears to be independent from *FNS I* genes described in *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays*.
Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

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

Competing interests
The authors declare that they have no competing interests.

Funding
MI was supported by the United States Department of Agriculture National Institute of Food and Agriculture, Hatch project 1008691. We acknowledge support by the Open Access Publication Funds of Technische Universität Braunschweig.

Authors’ contribution
BP and MI performed the analyses and wrote the manuscript.

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

Additional files

Additional file 1. Collection of F3H, FNS I, and FNS I-like sequences that were used for the analyses of this study.

Additional file 2. Gene expression values (TPMs) of *Daucus carota* F3H, FNS I, and FNS I-like genes.


Additional file 4. Pairwise comparison of F3H and FNS I sequences. Percentage of identical amino acid residues is displayed in this matrix.

Additional file 5. Results of a co-expression analysis of F3H (DCAR_009483) in *Daucus carota* based on 146 RNA-seq data sets.
Additional file 6. Results of a co-expression analysis of FNS I (DCAR_009489) in *D. carota* based on 146 RNA-seq data sets.

References


