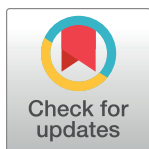


## RESEARCH ARTICLE

Lineage-specific gene duplication and expansion of *DUF1216* gene family in *Brassicaceae*Zai-Bao Zhang<sup>1\*</sup>, Tao Xiong<sup>2</sup>, Xiao-Jia Wang<sup>3</sup>, Yu-Rui Chen<sup>3</sup>, Jing-Lei Wang<sup>3</sup>, Cong-Li Guo<sup>3</sup>, Zi-Yi Ye<sup>1</sup><sup>1</sup> School of Life and Health Science, Huzhou College, Huzhou, Zhejiang, China, <sup>2</sup> College of Life Science, Xinyang Normal University, Xinyang, Henan, China, <sup>3</sup> College of International Education, Xinyang Normal University, Xinyang, Henan, China\* [zaibaozhang79@163.com](mailto:zaibaozhang79@163.com)

## OPEN ACCESS

**Citation:** Zhang Z-B, Xiong T, Wang X-J, Chen Y-R, Wang J-L, Guo C-L, et al. (2024) Lineage-specific gene duplication and expansion of *DUF1216* gene family in *Brassicaceae*. PLoS ONE 19(4): e0302292. <https://doi.org/10.1371/journal.pone.0302292>**Editor:** Achraf El Allali, Mohammed VI Polytechnic University: Universite Mohammed VI Polytechnique, MOROCCO**Received:** January 19, 2024**Accepted:** April 1, 2024**Published:** April 16, 2024**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0302292>**Copyright:** © 2024 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting Information](#) files.

## Abstract

Proteins containing domain of unknown function (DUF) are prevalent in eukaryotic genome. The DUF1216 proteins possess a conserved DUF1216 domain resembling to the mediator protein of Arabidopsis RNA polymerase II transcriptional subunit-like protein. The DUF1216 family are specifically existed in *Brassicaceae*, however, no comprehensive evolutionary analysis of *DUF1216* genes have been performed. We performed a first comprehensive genome-wide analysis of DUF1216 proteins in *Brassicaceae*. Totally 284 *DUF1216* genes were identified in 27 *Brassicaceae* species and classified into four subfamilies on the basis of phylogenetic analysis. The analysis of gene structure and conserved motifs revealed that *DUF1216* genes within the same subfamily exhibited similar intron/exon patterns and motif composition. The majority members of *DUF1216* genes contain a signal peptide in the N-terminal, and the ninth position of the signal peptide in most DUF1216 is cysteine. Synteny analysis revealed that segmental duplication is a major mechanism for expanding of *DUF1216* genes in *Brassica oleracea*, *Brassica juncea*, *Brassica napus*, *Lepidium meyenii*, and *Brassica carinata*, while in *Arabidopsis thaliana* and *Capsella rubella*, tandem duplication plays a major role in the expansion of the *DUF1216* gene family. The analysis of Ka/Ks (non-synonymous substitution rate/synonymous substitution rate) ratios for *DUF1216* paralogous indicated that most of gene pairs underwent purifying selection. *DUF1216* genes displayed a specifically high expression in reproductive tissues in most *Brassicaceae* species, while its expression in *Brassica juncea* was specifically high in root. Our studies offered new insights into the phylogenetic relationships, gene structures and expressional patterns of *DUF1216* members in *Brassicaceae*, which provides a foundation for future functional analysis.

## Introduction

Gene duplication provides raw materials for the production of new genes, which can be divided into tandem duplication, segmental duplication and whole genome duplication [1].

**Funding:** This work was supported by the grants from National Natural Science Foundation of China (32170351), the General Research Projects of Zhejiang Provincial Department of Education (Y202351039), Huzhou Science and Technology Plan Project (2023GZ44), and Research Program of Huzhou College (2023HXKM09). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

There are a wide range of different types of gene duplication in the genome [2, 3]. The eukaryotic genome contains numerous multi-copy gene families that facilitate the generation of novel protein functions [4]. Lineage-specific gene duplication and expansion refers to a notable increase in the copy numbers of certain genes in one species compared to other species, which may have played a crucial role during evolution or under environmental pressures and ultimately became fixed in that particular lineage [5–7]. Papain-like cysteine proteases in *Carica papaya* display lineage-specific gene duplication and expansion, which attributed to the ongoing reciprocal selection pressures of herbivores and plant defense mechanisms [7].

Flowers are among the most complex structures in angiosperms, playing significant roles in the evolution and advancement of sexual reproduction in plants [8]. Stamens, crucial components of flowers, are divided into anthers and filaments. Anther development in *Arabidopsis thaliana* comprises 14 stages, including anther morphogenesis, microspore development, gametophyte genesis, and anther dehiscence [9]. The tapetum, a critical component of the anther, undergoes three stages of cellular fate: tapetum cell differentiation, tapetum secretory cell formation, and tapetum cell apoptosis [9, 10]. Numerous peptides, transcription factors, kinases are expressed in the tapetum, which participate in intercellular communication and play crucial roles in plant reproductive development. *Arabidopsis TPD1* (*TAPETUM DETERMINANT 1*) encodes a cysteine-rich peptide expressed in the tapetum. Mutations in *TPD1* disrupt tapetum development, resulting in a proliferation of microsporocytes and a non-functional tapetum phenotype [11–13]. Transcription factors BES1 (*BRI1 EMS SUPPRESSOR 1*) and BZR1 (*BRASSINAZOLE RESISTANT 1*) serve as downstream targets of the TPD1-EMS1 (*EXCESS MICROSPOROCYTES 1*)-SERK1/2 (*SOMATIC EMBRYO RECEPTOR KINASE 1/2*) pathway [11] and are also expressed in the tapetum. The quintuple mutant of the *BES1* family, which consists of *bes1*, *bzr1*, *beh1*, *beh3*, and *beh4* genes, exhibits a tapetum-absent phenotype similar to that of *ems1*, *tpd1*, and *serk1serk2* mutants [11].

The *Brassicaceae* family, a part of the Brassicales order, is a large and diverse plant family with over 3709 species distributed across approximately 338 genera [14]. This family, also known as the mustard family, is predominantly found in the north temperate zone, particularly in the Mediterranean region. In China alone, there are about 95 genera and 411 species in *Brassicaceae*. *Brassicaceae* plants are characterized by unique morphological features, such as four cross-like arranged petals, which give the family its name. The flowers typically have six stamens, with two being shorter than the others, and a single pistil [15]. The fruits are often elongated capsules called siliques or shorter, rounded structures called silicles [16, 17]. Species within the *Brassicaceae* family hold significant economic and ecological importance, encompassing a wide range of crops, ornamental plants, and wild species. Notable members of this family include Brassica species, such as the model plant *Arabidopsis thaliana*, which is extensively studied for its small genome size and rapid life cycle. Other economically important species, such as *Brassica oleracea* (cabbage, broccoli, cauliflower, kale), *Brassica rapa* (turnip, Chinese cabbage), and *Brassica napus* (rapeseed, canola), are cultivated worldwide for their edible parts and oil production.

In the evolution of *Brassicaceae*, gene duplication is the main way to produce new genes [18–20], such as L-type LecRKs (Lectin Receptor Kinase) and LLPs (L-type lectin domain proteins), which makes an important contribution to their adaptation to the environment [18, 21]. Domains of unknown function (DUFs), deposited in the protein family database (Pfam), represent conserved amino acid sequences within protein domains whose functions remain uncharacterized [22, 23]. Proteins sharing the same DUF are classified into distinct DUF families [22]. The DUF protein family plays a pivotal role in the regulation of plant growth and development, mediating responses to both biotic and abiotic stresses, as well as fulfilling various regulatory functions throughout the life cycle of plants [24]. The following protein

domains have undergone functional characterization in *Arabidopsis thaliana*: DUF6 [25], DUF26 [26], DUF246 [27], DUF538 [28], DUF579 [29], DUF617 [30], DUF642 [31], DUF647 [32], DUF724 [33], DUF784 [34], DUF1117 [35], DUF1218 [36], DUF4005 [37] and DUF4228 [38]. Several of these proteins, including DUF6, DUF246, and DUF579, have been implicated in cell wall development; the involvement of the DUF538 protein has been observed in trichome development; and plant stress responses have been associated with DUF26, DUF1117, and DUF4228. In summary, these investigations have unveiled the participation of proteins containing the domain of unknown function (DUF) in a wide range of biological processes. In this study, we have identified a specific expansion of DUF1216 within the *Brassicaceae* lineage. Through comprehensive phylogenetic analysis, a total of 284 *DUF1216* genes were successfully characterized across 27 species belonging to the *Brassicaceae*, subsequently classified into four distinct subfamilies. The examination of gene structure and conserved motifs unveiled a consistent pattern in intron/exon arrangement and motif composition among *DUF1216* genes belonging to the same subfamily. Synteny analysis revealed that segmental duplication is a predominant mechanism driving the expansion of *DUF1216* genes in *B. oleracea*, *B. juncea*, *B. napus*, *L. meyenii*, and *B. carinata*. Conversely, tandem duplication plays a prominent role in the expansion of the *DUF1216* gene family in *A. thaliana* and *C. rubella*. The *DUF1216* genes exhibited a notably high expression specifically in reproductive tissues across most *Brassicaceae* species, whereas their expression was particularly elevated in the roots of *B. juncea*. Our studies have provided novel insights into the phylogenetic relationships, gene structures, and expression patterns of *DUF1216* members in *Brassicaceae*, thereby establishing a solid foundation for future functional analysis.

## Results

### Whole genome identification of *DUF1216* genes in *Brassicaceae*

Hidden Markov Model (HMM) and BLASTP search were used to identify *DUF1216* proteins in 27 genome sequenced species in *Brassicaceae*. In total, 5, 7, 11, 11, 7, 6, 9, 9, 20, 18, 23, 11, 12, 11, 30, 7, 9, 5, 6, 7, 6, 14, 6, 12, 9, 7, and 6 *DUF1216* proteins were retrieved from *Aethionema arabicum*, *Arabidopsis halleri*, *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Arabidopsis alpina*, *Barbarea vulgaris*, *Boechera retrofracta*, *Boechera stricta*, *Brassica carinata*, *Brassica juncea*, *Brassica napus*, *Brassica nigra*, *Brassica oleracea*, *Brassica rapa*, *Camelina sativa*, *Capsella grandiflora*, *Capsella rubella*, *Cardamine hirsuta*, *Eutrema salsugineum*, *Isatis indigotica*, *Leavenworthia alabamica*, *Lepidium meyenii*, *Microthlaspi erraticum*, *Raphanus sativus*, *Schrenkiella parvula*, *Sisymbrium irio*, and *Thlaspi arvense*, respectively (Table 1 and S1 Table). No *DUF1216* homologues have been found in other species, such as algae, mosses, ferns, monocotyledonous plants, dicotyledonous plants except *Brassicaceae*. Protein domain duplication is the main reason for the functional diversity of protein families [39]. About one-third (93 in 284) of the identified *DUF1216* proteins contain two *DUF1216* domains, indicating that these proteins are involved in complex biological processes (S1 Fig).

The physical and chemical properties of the identified *DUF1216* protein were studied, including the number of amino acids, molecular weight (WM) and isoelectric point (pI) (S1 Table). The *DUF1216* protein sequence varies in length from 68 to 2708 amino acids, with an average length of 602 amino acids. Its MW ranges from 7.49 to 310.10 KDa, while its pI ranges from 3.96 to 10.48.

### Phylogenetic relationship of *DUF1216* genes in *Brassicaceae*

A rootless phylogenetic tree was constructed based on the whole protein sequence of *DUF1216*. According to the structure of the phylogenetic tree, 284 *DUF1216* genes were

**Table 1. Distribution of *DUF1216* gene in *Brassicaceae*.**

| Species                        | DUF1216 Subfamily |     |     |    | Total |
|--------------------------------|-------------------|-----|-----|----|-------|
|                                | I                 | II  | III | IV |       |
| <i>Aethionema arabicum</i>     | 1                 | 1   | 1   | 2  | 5     |
| <i>Arabidopsis halleri</i>     | 0                 | 2   | 3   | 2  | 7     |
| <i>Arabidopsis lyrata</i>      | 1                 | 4   | 4   | 2  | 11    |
| <i>Arabidopsis thaliana</i>    | 1                 | 5   | 3   | 2  | 11    |
| <i>Arabidopsis alpina</i>      | 1                 | 2   | 2   | 2  | 7     |
| <i>Barbarea vulgaris</i>       | 0                 | 3   | 1   | 2  | 6     |
| <i>Boechera retrofracta</i>    | 1                 | 4   | 2   | 2  | 9     |
| <i>Boechera stricta</i>        | 2                 | 3   | 2   | 2  | 9     |
| <i>Brassica carinata</i>       | 2                 | 7   | 5   | 6  | 20    |
| <i>Brassica juncea</i>         | 2                 | 8   | 3   | 5  | 18    |
| <i>Brassica napus</i>          | 3                 | 9   | 5   | 6  | 23    |
| <i>Brassica nigra</i>          | 2                 | 5   | 2   | 2  | 11    |
| <i>Brassica oleracea</i>       | 1                 | 5   | 3   | 3  | 12    |
| <i>Brassica rapa</i>           | 1                 | 4   | 3   | 3  | 11    |
| <i>Camelina sativa</i>         | 3                 | 12  | 8   | 7  | 30    |
| <i>Capsella grandiflora</i>    | 1                 | 2   | 2   | 2  | 7     |
| <i>Capsella rubella</i>        | 1                 | 4   | 2   | 2  | 9     |
| <i>Cardamine hirsuta</i>       | 0                 | 2   | 1   | 2  | 5     |
| <i>Eutrema salsugineum</i>     | 1                 | 2   | 1   | 2  | 6     |
| <i>Isatis indigotica</i>       | 1                 | 3   | 1   | 2  | 7     |
| <i>Leavenworthia alabamica</i> | 0                 | 1   | 1   | 4  | 6     |
| <i>Lepidium meyenii</i>        | 3                 | 3   | 2   | 6  | 14    |
| <i>Microthlaspi erraticum</i>  | 1                 | 1   | 2   | 2  | 6     |
| <i>Raphanus sativus</i>        | 1                 | 6   | 2   | 3  | 12    |
| <i>Schrenkiella parvula</i>    | 2                 | 3   | 2   | 2  | 9     |
| <i>Sisymbrium irio</i>         | 0                 | 3   | 2   | 2  | 7     |
| <i>Thlaspi arvense</i>         | 1                 | 1   | 2   | 2  | 6     |
| Total                          | 33                | 105 | 67  | 79 | 284   |

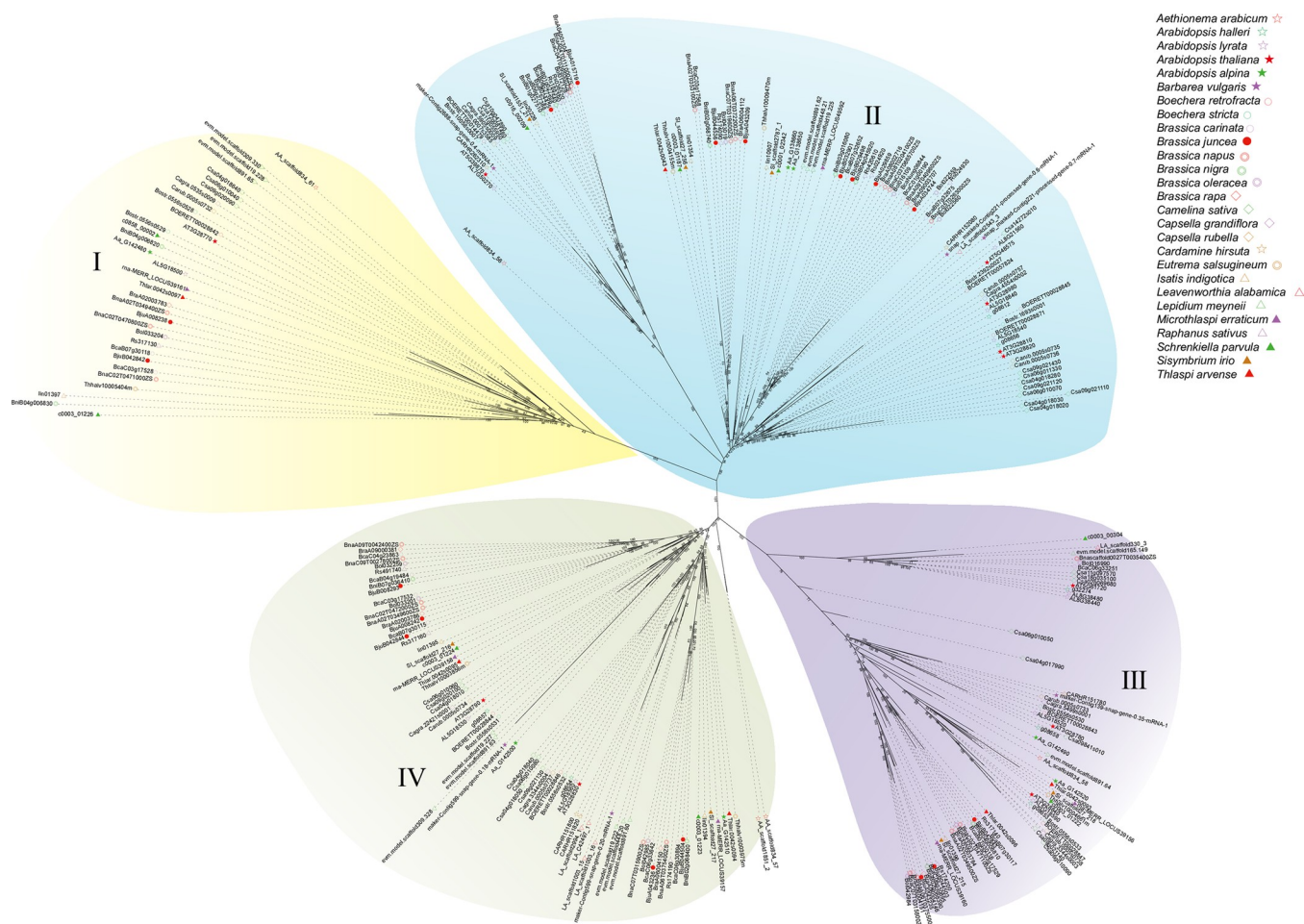
<https://doi.org/10.1371/journal.pone.0302292.t001>

divided into four subfamilies, represented by I, II, III and IV, respectively (Fig 1). The *DUF1216* genes exhibited an uneven distribution among the four subfamilies, with 33, 105, 67 and 79 *DUF1216* genes were identified in subfamilies I, II, III and IV, respectively (Table 1). Subfamilies II, III and IV contain *DUF1216* genes from all these 27 species, whereas in subfamily I, *DUF1216* genes were lost in *A.halleri*, *B.vulgaris*, *C.hirsuta*, *L.alabamica* and *S.irio*. The *DUF1216* genes in most species are evenly distributed across four distinct subfamilies, while *C. sativa* *DUF1216* displayed a significant high number in subfamily II (about 40%, 12 in 30).

### Gene structure and motif conservation of *DUF1216*

Conservation of gene structure and motif offers crucial insights into the evolution and functional conservation of the *DUF1216* gene family. Most *DUF1216* genes (227 in 284, 80%) contain only one or two introns. Totally 51 *DUF1216* genes (51 in 284, 18%) contain more than three introns, and most of them are concentrated in subfamily I (21 in 51, 41%) (Fig 2, S1 Table). In subfamilies II, III, and IV, 84% (88/105), 67% (45/67), 76% (60/79) *DUF1216* genes have 1–2 introns, one intron, and two introns, respectively. However, in subfamily I, 64% (21/33) of *DUF1216* genes have over three introns.





**Fig 1. Phylogenetic tree of DUF1216 in 27 species of *Brassicaceae* based on maximum likelihood method.** There are 284 DUF1216 proteins in the phylogenetic tree. These DUF1216 proteins were divided into four groups, with each branch name displayed next to the corresponding group and marked with a different color.

<https://doi.org/10.1371/journal.pone.0302292.g001>

The conserved motif of DUF1216 protein was predicted by online software MEME, and six conserved motifs were identified (Fig 2). Motif 1 and motif 2 are widely distributed in DUF1216 protein. Motifs 4 and 5 were not identified in subfamily I with the exception of Car-ub.0005s0732 and Csa04g018640; motif4 was not detected in subfamily II with the exception of BjuB024425, Rs036520, Bol012128, BraA04001304, BjuA015719, and Thlar.0042s0043; motif 3 was not identified in subfamily III with the exception of AT5G61720 (Fig 2). The motif composition and position among different subfamilies are different. Motif 4 was specially identified in subfamilies III and IV. Overall, the members of DUF1216 in the same subfamily have similar motif composition, but significant differences were identified among different subfamilies, indicating the functional similarity within subfamilies and the diversity of functions between subfamilies.

### Signal peptide and subcellular localization of DUF1216 protein

Based on the predictions, most DUF1216 proteins (228 in 384, 80%) contain signal peptide, with the length of most signal peptides is approximately 25 amino acids (Fig 3, S2 Fig and S2 Table). Approximately 75.4% of the signal peptides have C (cysteine) and 14.0% of the



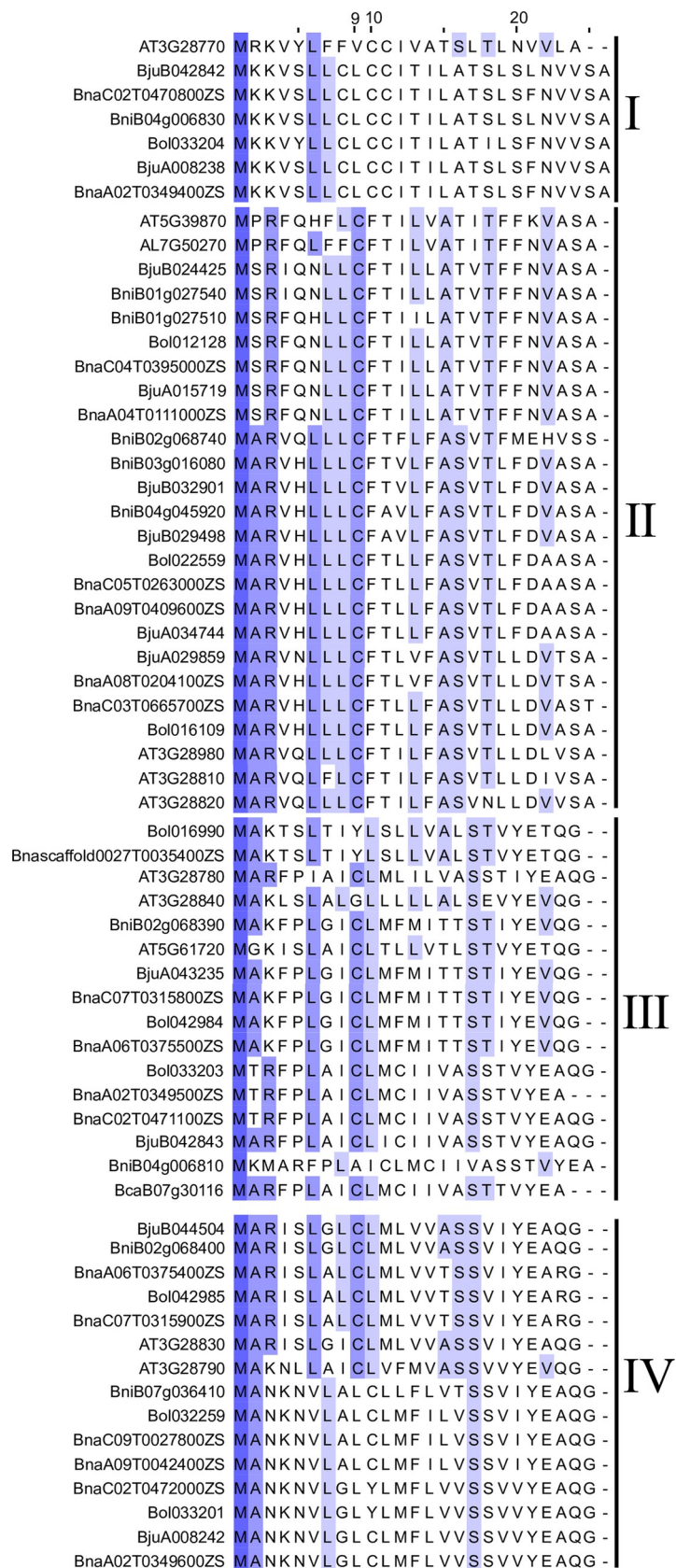
**Fig 2. Gene structure and conserved motif analysis of DUF1216 among *Brassicaceae* species.** Green boxes represent untranslated regions (UTRs), pink boxes correspond to coding sequences (CDSs), and thin black lines indicate introns. Six distinct motifs (numbered 1–6) were identified in DUF1216 proteins using the MEME program and are depicted in different colors.

<https://doi.org/10.1371/journal.pone.0302292.g002>

signal peptides have L (leucine) at position 9, respectively. Additionally, the majority of DUF1216 members are predicted to localize to the nucleus, with only a few proteins are predicted to be located in the cell membrane, cell wall, or cytoplasm (S2 Table). These prediction results can be useful for further functional studies and as references for target screening.

### Promoter *cis*-elements analysis of *DUF1216*s

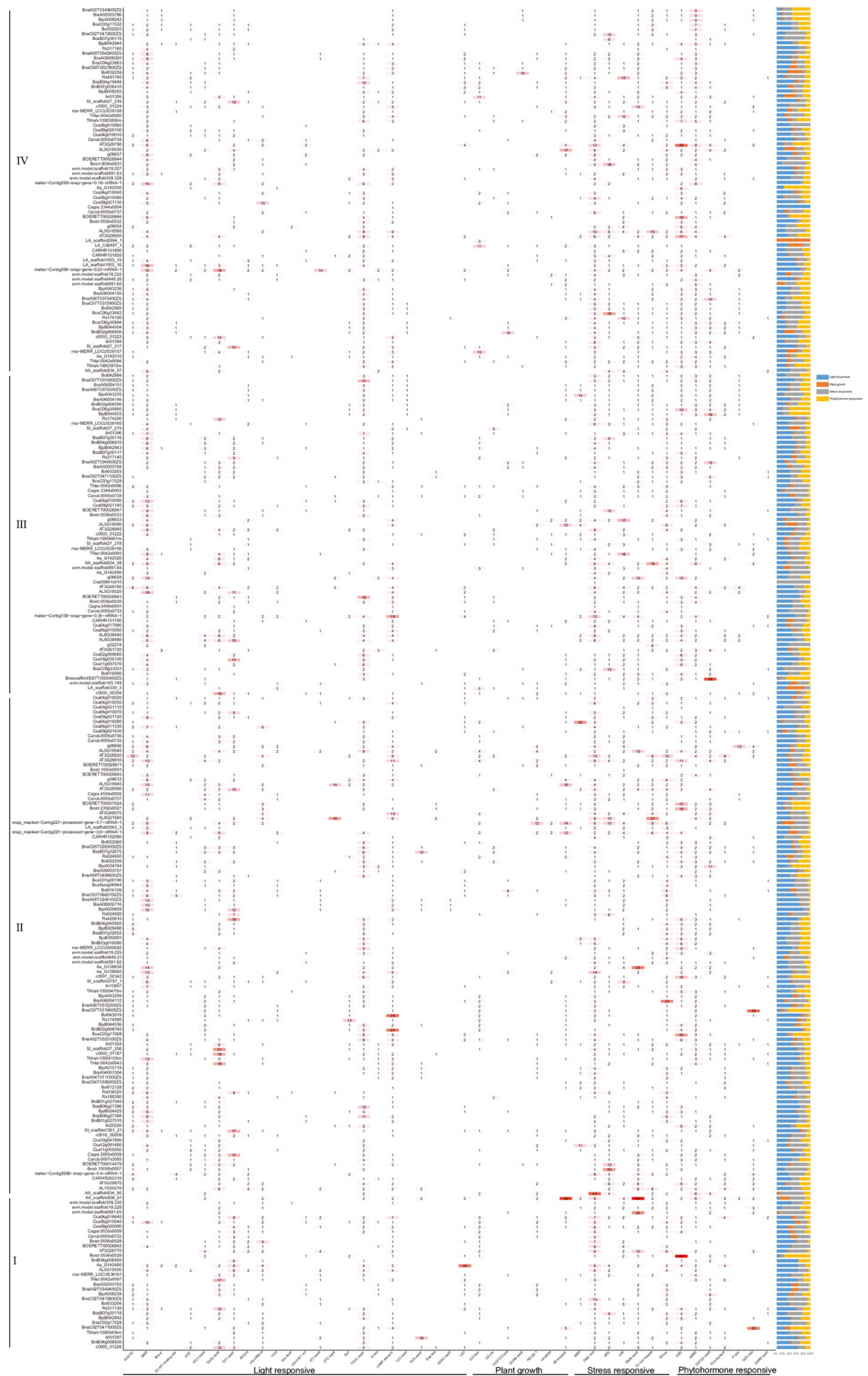
Promoter *cis*-elements play crucial roles in regulating gene expression [40]. Forty-five *cis*-acting elements were identified in the promoter region of *DUF1216* genes, which could be classified into four categories elements, including light-responsive, plant growth, stress, and phytohormone responsive elements (Fig 4, S3 Table). A total of 3846, 626, 2629, and 1775 light-responsive elements, plant growth elements, stress elements, and phytohormone responsive elements, were identified, respectively. In subfamily I, it is worth noting that several *cis*-acting elements, including the light-responsive elements (I-box, ATC-motif, TCT-motif, AAAC-motif) and plant growth elements (O2-site, GCN4-motif, and circadian), were absent. In subfamily II, light-responsive components (TCT-motif elements) were absent. GA-motif and AAAC-motif elements were absent in subfamily III and ACA-motif and gap-



**Fig 3. DUF1216 signal peptide sequence of representative species.**

<https://doi.org/10.1371/journal.pone.0302292.g003>







box components were absent in subfamily IV. The *DUF1216* promoter are characterized by a high abundance of MRE (830) and DRE core (803) elements. These results indicate that *DUF1216*s are expressed under light induction, respond to abiotic stress, and play important roles in hormone response and regulate the plant growth.

### Gene duplication and synteny analysis of *DUF1216* genes

To explore the expansion pattern of *DUF1216* genes, the duplication pattern was analyzed (Fig 5 and S3 Fig). Previous research has indicated that segmental and tandem duplications are two primary factors contributing to the expansion of plant gene families [41, 42]. In detail, 2, 1, 3, 4, 6, 1, 1, 1, 22, 30, 12, 10 and 12 pairs of segmental duplication genes were identified in *A.thaliana*, *A.lyrata*, *S.parvula*, *B.nigra*, *B.oleracea*, *C.rubella*, *R.sativus*, *I.indigotica*, *B.juncea*, *B.napus*, *C.sativa*, *L.meyneii*, and *B.carinata*, respectively (Fig 5, S3 Fig and S4 Table). Furthermore, tandem duplication resulted in 6, 3, 3, 5, 2, 6, 2, 3, 2, 2, 16, 7, and 5 additional *DUF1216* genes in *A.thaliana*, *A.lyrata*, *S.parvula*, *B.nigra*, *B.oleracea*, *C.rubella*, *R.sativus*, *I.indigotica*, *B.juncea*, *B.napus*, *C.sativa*, *L.meyneii*, and *B.carinata*, respectively (Fig 5 and S3 Fig). Various mechanisms play significant roles in the expansion of *DUF1216* genes among different *Brassicaceae* species. Segmental duplication is the primary mechanism responsible for the expansion of *DUF1216* genes in *B.oleracea*, *B.carinata*, *B.napus*, *B.juncea*, and *L.meyneii*, while tandem duplication is the main expansion mechanism for the expansion of *DUF1216* genes in *A.thaliana* and *C.rubella*. This emphasizes the diverse evolutionary strategies utilized by distinct *Brassicaceae* species to expand *DUF1216* genes.

The analysis of Ka/Ks (non-synonymous substitution rate/synonymous substitution rate) ratios was performed to examine the selection pressure on gene evolution (S5 Table). Usually, a Ka/Ks ratio above 1 indicates positive selection, while a Ka/Ks ratio below 1 suggests purifying or negative selection [43]. These results showed that the Ka/Ks ratios of all the *DUF1216* paralogs were all less than 1, indicating that they undergone strong purifying selection pressure contributing to the maintenance of their function and reflecting that they had not diverged much during evolution. Among them, one Ka/Ks ratio was lower than 0.1 (paralog gene pair *Csa04g018280/Csa06g011330*) suggesting a strong purifying selection pressure and making the functions of the paralogous gene toward relative similarity.

To gain a deeper understanding of the evolutionary relationships among *DUF1216* genes, inter-species synteny was analyzed among 14 genomes, including *Arabidopsis thaliana* and *Brassica napus* (Fig 6 and S4 Fig). The synteny analysis showed that syntenic gene pairs were extensively present among these 14 genomes, such as 3 *DUF1216* syntenic gene pairs between *A.thaliana* and *A.alpina*; twelve *DUF1216* syntenic gene pairs between *A.alpina* and *B.napus*; forty-seven *DUF1216* syntenic gene pairs between *B.napus* and *C.sativa*; thirteen *DUF1216* syntenic gene pairs between *C.sativa* and *I.indigotica*; two *DUF1216* syntenic gene pairs between *I.indigotica* and *S.irio* (Fig 6). Individual homologous genes exhibited one-to-many or many-to-one homology, which was more evident in the synteny between *Arabidopsis alpina* and *Brassica napus*, *Brassica napus* and *Camelina sativa*, and *Camelina sativa* and *Isatis indigotica* (Fig 6). Furthermore, one-to-one collinear relationships were also identified between *Isatis indigotica* and *Sisymbrium irio* (Fig 6). The widespread presence of syntenic connections among these 14 genomes suggests that whole genome duplication significantly contributed to the expansion of *DUF1216* families.

### Expression profile of *DUF1216* genes

The expression of *DUF1216* genes in different tissues was detected by using public RNA-seq data (Fig 7). In *Arabidopsis thaliana*, 23 tissues and developmental stages were investigated

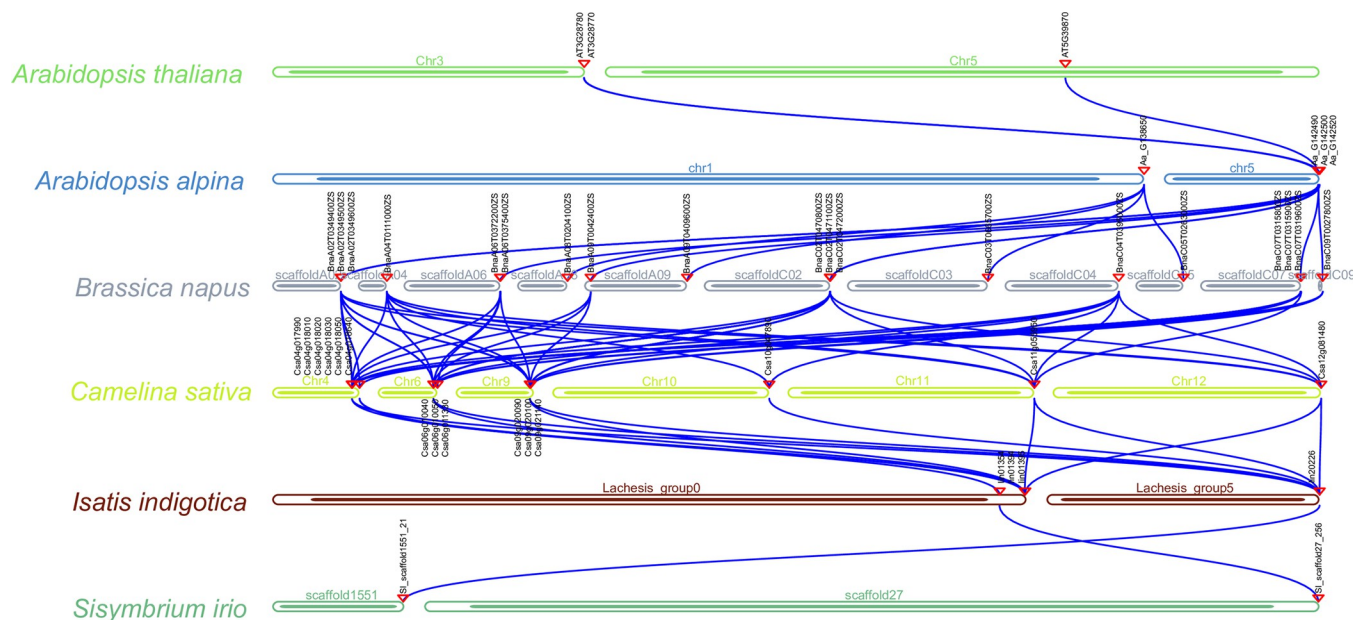


**Fig 5. Intraspecific syntenic relationships of *DUF1216* genes in various plant species.** A, *Arabidopsis thaliana*; B, *Arabidopsis lyrata*; C, *Schrenkiella parvula*; D, *Brassica nigra*; E, *Brassica oleracea*; F, *Capsella rubella*; G, *Raphanus sativus*; H, *Isatis indigotica*. Yellow highlights represent *DUF1216* genes; tandem duplications denoted by yellow arrows with green dots; syntenic genes linked by black lines.

<https://doi.org/10.1371/journal.pone.0302292.g005>

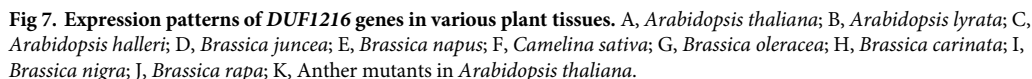
(Fig 7A). Ten *AtDUF1216* genes were significantly high expression in mature pollen and stamens of stage 15 flowers, while its transcripts in vegetable organs were barely detected. These results indicate that *Arabidopsis DUF1216* genes function in stamen development. In *Arabidopsis lyrata*, 7 tissues were examined, i.e., leaf, mature stigma, mature pollen, bolting flower, mature stem, mature root, and young silique (Fig 7B). The *A.lyrata DUF1216* genes also exhibit high expression levels in mature pollen and bolting flower, similar to that in *Arabidopsis thaliana*. In *Arabidopsis halleri*, five tissues were investigated including old leaves, new leaves, roots, self-pollinated pistils, and unpollinated pistils (Fig 7C). All *A.halleri DUF1216* genes exhibited high expression in both self-pollinated and unpollinated pistils, mirroring the expression patterns observed in *Arabidopsis* and *Arabidopsis lyrata*. Most *DUF1216* genes in *Brassica napus*, *Camelina sativa*, *Brassica oleracea*, *Brassica carinata*, *Brassica nigra*, *Brassica carinata*, *Brassica oleracea*, and *Brassica rapa*, displayed high expression in reproductive tissues, especially in the male organs (Fig 7E–7J). These expression patterns were consistent with the *Arabidopsis DUF1216* genes. In *Brassica juncea*, the *DUF1216* genes showed different expression patterns (Fig 7D), with most *DUF1216* genes were highly expressed in roots, indicating the special functions of *B.juncea DUF1216* in root development.

The majority of *AtDUF1216* genes exhibit downregulation in anther mutants, such as *DYT1* (DYSFUNCTIONAL TAPETUM1), *TDF1* (TAPETAL DEVELOPMENT and FUNCTION1) (Fig 7K), *AMS* (ABORTED MICROSPORES), and *MS188* (MALE STERILITY188). This suggests that the expression of *AtDUF1216* genes is under the regulation of anther-specific transcription factors.



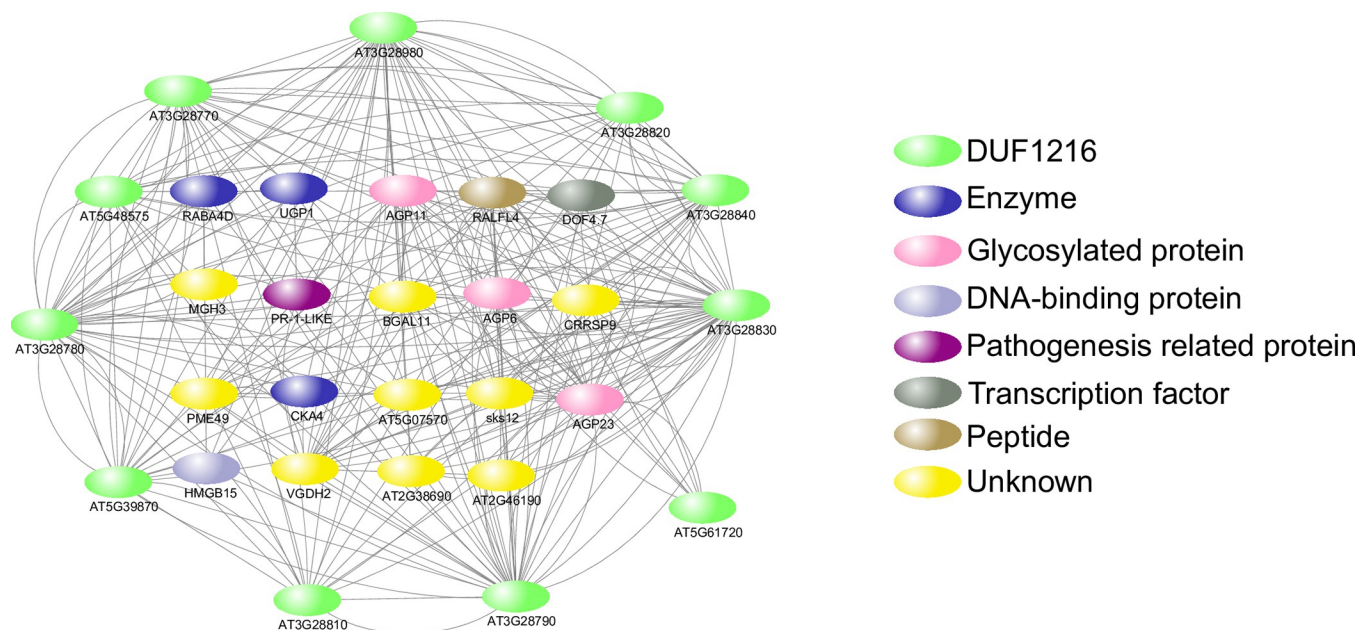
**Fig 6. Interspecies syntenic relationship of *DUF1216* genes in *Brassicaceae*.** Orthologous gene synteny represented by black lines.

<https://doi.org/10.1371/journal.pone.0302292.g006>



12 / 24





**Fig 8. Co-expression network of *DUF1216* genes in *Arabidopsis thaliana*.**

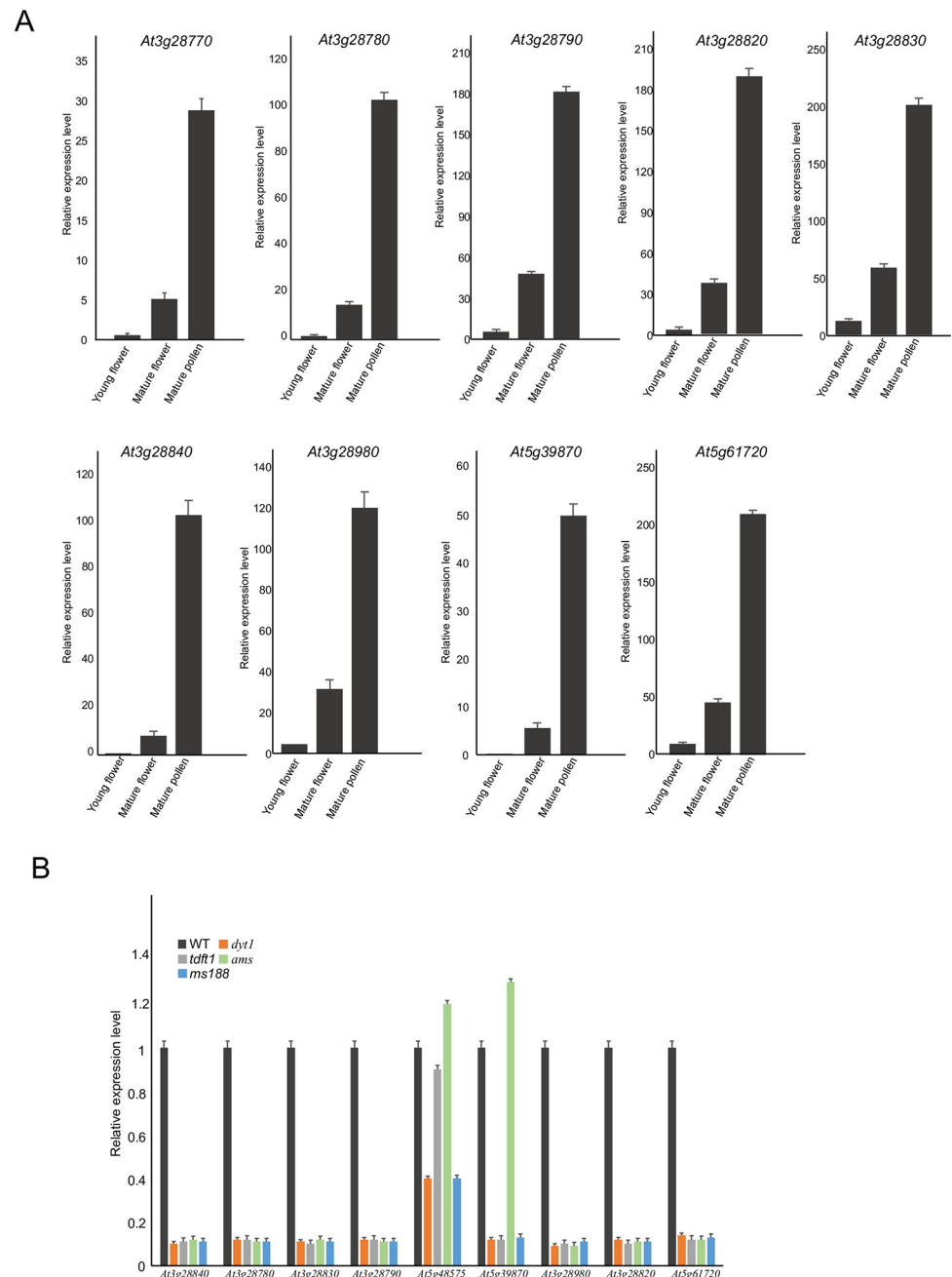
<https://doi.org/10.1371/journal.pone.0302292.g008>

### Analysis of *DUF1216* genes co-expression network

To analyze the interaction of *DUF1216* genes with other genes, co-expression analysis was performed for *Arabidopsis DUF1216* genes (Fig 8). The 11 *Arabidopsis DUF1216* genes are co-expressed with each other and are interacting with a total of 19 transcripts. Many of these co-expressed transcripts were found to function in pollen development. For example, *AtUGP1* (UDP-glucose pyrophosphorylase 1) and *AtUGP2* (UDP-glucose pyrophosphorylase 2) encode UGPase, express in all organs, and the *ugp1 ugp2* double mutant cause male sterility [44]. *AGP6* (Arabinogalactan protein 6) and *AGP11* (Arabinogalactan protein 11), encode cell wall proteoglycans, expressed in pollen, and mutation in *AGP6* and *AGP11* impaired pollen development and hindered pollen tube growth [45–48]. *AGP23* is expressed exclusively in pollen and may play a significant role in microspore development and/or pollen tube growth [49]. *RALF4* (RAPID ALKALIZATION FACTOR) is expressed in pollen tube and inhibits pollen germination. [50]. *AtHMGB15* (ARID-HMG DNA-binding protein 15) is predominantly expressed in pollen grains and pollen tubes, and mutation in *AtHMGB15* result in impaired pollen tube growth and reduced seed set rates [51]. *RABA4D*, a Rab GTPase gene, is specifically expressed in pollen tube and regulates both PRK6 distribution and pollen tube growth [52–54]. In summary, *AtDUF1216* genes are co-expressed with various genes crucial for pollen development, suggesting their potential importance in these processes.

### Gene expression analysis qRT-PCR validation

To verify the expression profiles of *DUF1216* genes, we selected and analyzed the expression of *Arabidopsis DUF1216* genes with qRT-PCR (Fig 9). The expression levels of the tested *Arabidopsis DUF1216* genes were specifically high expressed in mature flowers and pollen grains, but not in vegetable tissues (leaf, root, stem) (Fig 9A). These expression patterns are similar to transcriptomic results (Fig 7), indicating that *Arabidopsis DUF1216* genes involved in reproductive development.



**Fig 9. Expression profiles of Arabidopsis *DUF1216* genes.** A, Expression profiles of Arabidopsis *DUF1216* genes in young flower, mature flower and mature pollen. The error bars were obtained from three measurements. B, qRT-PCR analysis of *DUF1216* genes in the inflorescence of WT, *dtf1*, *tdf1*, *ams*, and *ms188* mutants. Data are presented as mean and SD (n = 3).

<https://doi.org/10.1371/journal.pone.0302292.g009>

Dyt1, TDF1, AMS and MS188 are essential transcription factors that regulate anther development [55–59]. Based on the previous microarray data, most *DUF1216* genes act downstream of these regulators (Fig 7) [59]. We collected the inflorescences of *dtf1*, *tdf1*, *ams*, and *ms188*, for quantitative real-time PCR (qPCR) analysis. The qPCR results showed that the expressions of *DUF1216* genes were downregulated in these mutants (Fig 9B). These results demonstrated

that *Arabidopsis DUF1216* genes were regulated by another important transcription factors during anther development.

## Discussion

The *Brassicaceae* family holds significant scientific and economic importance, particularly in the fields of agriculture and human health [60–62]. The *DUF1216* are existed exclusively in *Brassicaceae*. However, it is unclear how *DUF1216* family has evolved in *Brassicaceae* plants. Here, we conducted a comprehensive evolutionary analysis of *DUF1216* genes in *Brassicaceae*. Our studies provided new insights into the phylogenetic relationships, gene structures, and expression patterns of *DUF1216* genes in *Brassicaceae*. These findings lay the foundation for future functional analysis.

This study represents the first comparative evolutionary analysis of the *DUF1216* genes across *Brassicaceae* species. The number of *DUF1216* genes exhibited significant differences among various species of *Brassicaceae*. Most *Brassicaceae* plants have approximately 10 *DUF1216* genes, while *Brassica carinata*, *Brassica juncea*, *Brassica napus*, and *Camelina sativa* exhibit a higher number of *DUF1216* genes at 20, 18, 23, and 30, respectively (Table 1). Thus, even though the genome sizes of different *Brassicaceae* species are similar, the copy number of the *DUF1216* gene varies among them.

Phylogenetic analysis indicates that the *DUF1216* genes in *Brassicaceae* can be classified into four subfamilies (Fig 1). All *DUF1216* proteins, except for AT5G61720, contain the conserved *DUF1216* domain. Additionally, around one-third of *DUF1216* proteins contain two *DUF1216* domains. These results suggest that the *DUF1216* domain was lost in AT5G61720 during evolution, and that the protein with two *DUF1216* domains may have some additional functions. The majority of *DUF1216* genes in subfamily I contained more than three introns, whereas most genes in the other three subfamilies contained 1–2 introns (Fig 2). These results indicating that the subfamily I *DUF1216* genes may have distinct functions in comparison to those in the other subfamilies. The arrangement of motif within the subfamily was generally consistent, but there were significant differences among different subfamilies (Fig 2). Motif4 is exclusively existed in subfamilies III and IV, and Motif3 is exclusively present in subfamilies I, II, and IV. The existence of different motifs among different *DUF1216* proteins may endow different functions, resulting in functional differentiation.

Gene duplications play crucial roles in the emergence of new gene functions and the expansion of gene families [63]. Our study emphasizes the diverse evolutionary strategies utilized by distinct *Brassicaceae* species to expand the *DUF1216* genes (Fig 5, S3 Fig and S4 Table). The exclusive expansion of *DUF1216* genes in *Brassicaceae* plants suggests that these genes are biologically significant within this specific plant family. In addition, the Ka/Ks values of the duplicated *DUF1216* genes are less than 1 (S5 Table), indicating that they have undergone a more intensive purification selection. Extensive interspecific synteny among *DUF1216* genes was identified, indicating that *DUF1216* are conserved during evolution in *Brassicaceae* (Fig 6 and S4 Fig).

Most *DUF1216* genes displayed high expression in reproductive organs, indicating that they involved in reproductive development (Fig 7). However, in *Brassica juncea*, the *DUF1216* genes are highly expressed in roots, indicating a different role of *B.juncea DUF1216* genes in root growth. In *Arabidopsis*, most of the *DUF1216* genes were located downstream of anther transcription factors, and *AtDUF1216* genes were co-expressing with pollen-related genes. These results suggest that *AtDUF1216* genes may play important regulatory roles in pollen development.

## Conclusions

In this study, we conducted a comprehensive and genome-wide investigation of the *DUF1216* gene family in representative *Brassicaceae* species for the first time. We analyzed gene structure, protein motif, syntenic relationship, expression, and phylogenomic relationship of *DUF1216* genes, thereby providing novel insights into the evolution of this gene family. Our research offers valuable information for future functional analysis and utilization of *DUF1216* genes in *Brassicaceae*.

## Materials and methods

### Data sources and sequence acquisition

*Arabidopsis DUF1216* genes were retrieved from the Arabidopsis Information Resource (TAIR, <https://www.arabidopsis.org/>). Genomic sequences of *Brassicaceae* plants, *Aethionema arabicum*, *Arabidopsis halleri*, *Arabidopsis lyrata*, *Arabidopsis alpina*, *Barbarea vulgaris*, *Boechera retrofracta*, *Boechera stricta*, *Brassica carinata*, *Brassica juncea*, *Brassica napus*, *Brassica nigra*, *Brassica oleracea*, *Brassica rapa*, *Camelina sativa*, *Capsella grandiflora*, *Capsella rubella*, *Cardamine hirsute*, *Eutrema salsugineum*, *Isatis indigotica*, *Leavenworthia alabamica*, *Lepidium meyenii*, *Microthlaspi erraticum*, *Raphanus sativus*, *Schrenkiella parvula*, *Sisymbrium irio* and *Thlaspi arvense*, were downloaded from TBGR (<http://www.tbgr.org.cn/download/download.html>) [64–87].

Two methods were employed to identify *DUF1216* genes in *Brassicaceae*. First, HMMER search (E-value =  $1e-10$ ) with the Hidden Markov Model profile of the *DUF1216* domain (PF06746) was conducted to search local databases [88, 89]. Second, the amino acid sequences of *Arabidopsis thaliana DUF1216* members were used to perform a BLASTP search against the protein database with an E-value less than  $10^{-6}$  [90]. The putative *DUF1216* genes were further validated using online tools such as CDD (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/>) [91], HMM (<https://www.ebi.ac.uk/interpro/>) [92] and SMART (<https://smart.embl-heidelberg.de/>) [93].

### Multiple sequence alignment, protein structure predictions and phylogenetic analysis

*DUF1216* multiple sequence alignments were performed using MAFFT software [94]. Phylogenetic trees were generated based on full protein sequences. The maximum likelihood (ML) phylogenetic tree was constructed using IQ-TREE with the parameter '-m MFP -bb/alrt 1000' and 1000 ultra-bootstrap replicates [95]. Branch support was tested using 1000 bootstrap replicates. The consensus trees topology was visualized with iTOL [96].

### Gene structure, conserved motif analysis, signal peptide prediction and subcellular localization prediction

The protein properties were analyzed using ProtParam (<http://web.expasy.org/protparam/>), while the gene structures were visualized using the Gene Structure Display Server (GSDS) (<http://gsds.cbi.pku.edu.cn/index.php>). Conserved motifs were predicted utilizing MEME (<http://meme.nbcr.net/meme3/mme.html>) [97]. The signal peptides were identified using SignalP (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>) [98], while subcellular localization was predicted utilizing Plant-mPLoc (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>) [99].



### Promoter *cis*-acting element and syntenic analysis

The analysis of promoter *cis*-elements was performed using the Plant Care website (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) [100], while R software was utilized for data visualization. The chromosomal positions of *DUF1216* genes were determined based on comprehensive genome annotation. The gene duplication patterns were identified using MCScanX with default settings [101], while the syntenic map was generated using CIRCOS to visually connect putative duplicated genes.

### Gene expression and co-expression network analysis

The transcriptional levels of *DUF1216* genes assessed by analyzing transcriptome datasets obtained from the National Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/>). The reads were subjected to reanalysis, and the quantification of gene expression was performed as fragments per kilobase per million reads (FPKM). The heat-map with k-means clustering was generated using the R software. The mRNA expression data for four mutants (*dyt1*, *tdf1*, *ams*, and *ms188*) were obtained from a previous study [102]. The co-expression network was evaluated using the Arabidopsis Information Resource (TAIR, <https://www.arabidopsis.org/>) database, while Cytoscape software was employed for visualization.

### RNA extraction and quantitative reverse transcription-PCR

The total RNA was isolated using TRIzol reagent (Invitrogen, Thermo Fisher Scientific, United States) according to the instructions of the manufacturer. The first-strand cDNAs were synthesized from DNase I-treated total RNA using the cDNA Synthesis Kit (Takara, Japan). The gene expression was normalized using tubulin beta8 (TUB8) (At5g23860) as the reference gene, and the relative gene expression was calculated as the mean of three biological replicates and three technical replicates. Relative expression was calculated by the  $2^{-\Delta C_t}$  and  $2^{-\Delta\Delta C_t}$  methods. Gene specific primers used for quantitative reverse transcription-PCR (qRT-PCR) are listed in S6 Table.

### Supporting information

**S1 Fig. *DUF1216* protein sequence.** *DUF1216* domains are framed with red lines.  
(PDF)

**S2 Fig. *DUF1216* signal peptide sequence.**  
(PDF)

**S3 Fig. The intraspecific syntenic relationship of *DUF1216* genes in *Brassica juncea*, *Brassica napus*, *Camelina sativa*, *Lepidium meyenii*, *Brassica juncea* and *Brassica carinata*.** Syntenic genes are linked by black lines.  
(PDF)

**S4 Fig. Syntenic analysis of *DUF1216* among 14 species of *Brassicaceae*.**  
(PDF)

**S1 Table. *DUF1216* family genes identified in 27 *Brassicaceae* plants.**  
(XLSX)

**S2 Table. *DUF1216* signal peptide and subcellular localization prediction.**  
(XLSX)

**S3 Table. *DUF1216* promoter cis-acting element data.**  
(XLSX)

**S4 Table. Segmental duplication gene pairs.**  
(XLSX)

**S5 Table. The Ka/Ks analysis of *DUF1216* genes.**  
(XLSX)

**S6 Table. Primers used in qPCR.**  
(XLSX)

**S7 Table. *DUF1216* protein sequences used to build phylogenetic tree.**  
(DOCX)

## Author Contributions

**Funding acquisition:** Zai-Bao Zhang.

**Visualization:** Xiao-Jia Wang, Yu-Rui Chen, Jing-Lei Wang, Cong-Li Guo, Zi-Yi Ye.

**Writing – original draft:** Zai-Bao Zhang, Tao Xiong.

## References

1. Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon-intron structure. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(4):1187–92. Epub 2012/01/11. <https://doi.org/10.1073/pnas.1109047109> PMID: 22232673; PubMed Central PMCID: PMC3268293.
2. De Bodt S, Maere S, Van de Peer Y. Genome duplication and the origin of angiosperms. *Trends in ecology & evolution*. 2005; 20(11):591–7. Epub 2006/05/17. <https://doi.org/10.1016/j.tree.2005.07.008> PMID: 16701441.
3. Yu J, Wang J, Lin W, Li S, Li H, Zhou J, et al. The Genomes of *Oryza sativa*: a history of duplications. *PLoS biology*. 2005; 3(2):e38. Epub 2005/02/03. <https://doi.org/10.1371/journal.pbio.0030038> PMID: 15685292; PubMed Central PMCID: PMC546038.
4. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of Gene Duplication in Plants. *Plant physiology*. 2016; 171(4):2294–316. Epub 2016/06/12. <https://doi.org/10.1104/pp.16.00523> PMID: 27288366; PubMed Central PMCID: PMC4972278.
5. Shukla V, Habib F, Kulkarni A, Ratnaparkhi GS. Gene Duplication, Lineage-Specific Expansion, and Subfunctionalization in the MADF-BESS Family Patterns the *Drosophila* Wing Hinge. *Genetics*. 2014; 196(2):481–96. <https://doi.org/10.1534/genetics.113.160531> %J Genetics. PMID: 24336749
6. Tran LT, Taylor JS, Constabel CP. The polyphenol oxidase gene family in land plants: Lineage-specific duplication and expansion. *BMC Genomics*. 2012; 13(1):395. <https://doi.org/10.1186/1471-2164-13-395> PMID: 22897796
7. Liu J, Sharma A, Niewiara MJ, Singh R, Ming R, Yu Q. Papain-like cysteine proteases in *Carica papaya*: lineage-specific gene duplication and expansion. *BMC Genomics*. 2018; 19(1):26. <https://doi.org/10.1186/s12864-017-4394-y> PMID: 29306330
8. Alvarez-Buylla ER, Benítez M, Corvera-Poiré A, Chaos Cador A, de Folter S, Gamboa de Buen A, et al. Flower development. *The arabidopsis book*. 2010; 8:e0127. Epub 2010/01/01. <https://doi.org/10.1199/tab.0127> PMID: 22303253; PubMed Central PMCID: PMC3244948.
9. Sanders PM, Bui AQ, Weterings K, McIntire KN, Hsu Y-C, Lee PY, et al. Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. *Sexual Plant Reproduction*. 1999; 11(6):297–322. <https://doi.org/10.1007/s004970050158>
10. Scott RJ, Spielman M, Dickinson HG. Stamen structure and function. *The Plant cell*. 2004; 16 Suppl (Suppl):S46–S60. Epub 2004/05/08. <https://doi.org/10.1105/tpc.017012> PMID: 15131249; PubMed Central PMCID: PMC2643399.
11. Chen W, Lv M, Wang Y, Wang PA, Cui Y, Li M, et al. BES1 is activated by EMS1-TPD1-SERK1/2-mediated signaling to control tapetum development in *Arabidopsis thaliana*. *Nature communications*.

- 2019; 10(1):4164. Epub 2019/09/15. <https://doi.org/10.1038/s41467-019-12118-4> PMID: 31519953; PubMed Central PMCID: PMC6744560.
12. Yang SL, Xie LF, Mao HZ, Puah CS, Yang WC, Jiang L, et al. Tapetum determinant1 is required for cell specialization in the Arabidopsis anther. *The Plant cell*. 2003; 15(12):2792–804. Epub 2003/11/15. <https://doi.org/10.1105/tpc.016618> PMID: 14615601; PubMed Central PMCID: PMC282801.
  13. Jia G, Liu X, Owen HA, Zhao D. Signaling of cell fate determination by the TPD1 small protein and EMS1 receptor kinase. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(6):2220–5. Epub 2008/02/06. <https://doi.org/10.1073/pnas.0708795105> PMID: 18250314; PubMed Central PMCID: PMC2538901.
  14. Warwick SI, Francis A, Al-Shehbaz IA. Brassicaceae: Species checklist and database on CD-Rom. *Plant Systematics and Evolution*. 2006; 259(2):249–58. <https://doi.org/10.1007/s00606-006-0422-0>
  15. Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in plant science*. 2011; 16(2):108–16. Epub 2010/12/24. <https://doi.org/10.1016/j.tplants.2010.11.005> PMID: 21177137.
  16. Dinneny JR, Yadegari R, Fischer RL, Yanofsky MF, Weigel D. The role of JAGGED in shaping lateral organs. *Development (Cambridge, England)*. 2004; 131(5):1101–10. Epub 2004/02/20. <https://doi.org/10.1242/dev.00949> PMID: 14973282.
  17. Robles P, Pelaz S. Flower and fruit development in Arabidopsis thaliana. *The International journal of developmental biology*. 2005; 49(5–6):633–43. Epub 2005/08/13. <https://doi.org/10.1387/ijdb.052020pr> PMID: 16096970.
  18. Das Laha S, Dutta S, Schäffner AR, Das M. Gene duplication and stress genomics in Brassicas: Current understanding and future prospects. *J Plant Physiol*. 2020; 255:153293. <https://doi.org/10.1016/j.jplph.2020.153293> PMID: 33181457
  19. Lv X, Wei F, Lian B, Yin G, Sun M, Chen P, et al. A Comprehensive Analysis of the DUF4228 Gene Family in Gossypium Reveals the Role of GhDUF4228-67 in Salt Tolerance. *International journal of molecular sciences*. 2022; 23(21):13542. <https://doi.org/10.3390/ijms232113542> PMID: 36362330
  20. Hofberger JA, Nsibo DL, Govers F, Bouwmeester K, Schranz ME. A Complex Interplay of Tandem- and Whole-Genome Duplication Drives Expansion of the L-Type Lectin Receptor Kinase Gene Family in the Brassicaceae. *Genome Biol Evol*. 2015; 7(3):720–34. <https://doi.org/10.1093/gbe/evv020> %J Genome Biology and Evolution. PMID: 25635042
  21. Moore RC, Purugganan MD. The early stages of duplicate gene evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100(26):15682–7. Epub 2003/12/13. <https://doi.org/10.1073/pnas.2535513100> PMID: 14671323; PubMed Central PMCID: PMC307628.
  22. Luo C, Akhtar M, Min W, Bai X, Ma T, Liu C. Domain of unknown function (DUF) proteins in plants: function and perspective. *Protoplasma*. 2023. Epub 2024/01/02. <https://doi.org/10.1007/s00709-023-01917-8> PMID: 38158398.
  23. Bateman A, Coggill P, Finn RD. DUFs: families in search of function. *Acta crystallographica Section F, Structural biology and crystallization communications*. 2010; 66(Pt 10):1148–52. Epub 2010/10/15. <https://doi.org/10.1107/S1744309110001685> PMID: 20944204; PubMed Central PMCID: PMC2954198.
  24. Lv P, Wan J, Zhang C, Hina A, Al Amin GM, Begum N, et al. Unraveling the Diverse Roles of Neglected Genes Containing Domains of Unknown Function (DUFs): Progress and Perspective. *International journal of molecular sciences*. 2023; 24(4):4187. Epub 2023/02/26. <https://doi.org/10.3390/ijms24044187> PMID: 36835600; PubMed Central PMCID: PMC9966272.
  25. Ranocha P, Denancé N, Vanholme R, Freydier A, Martinez Y, Hoffmann L, et al. Walls are thin 1 (WAT1), an Arabidopsis homolog of Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *The Plant journal: for cell and molecular biology*. 2010; 63(3):469–83. Epub 2010/05/26. <https://doi.org/10.1111/j.1365-3113X.2010.04256.x> PMID: 20497379.
  26. Vaattovaara A, Brandt B, Rajaraman S, Safronov O, Veidenberg A, Luklová M, et al. Mechanistic insights into the evolution of DUF26-containing proteins in land plants. *Communications biology*. 2019; 2:56. Epub 2019/02/19. <https://doi.org/10.1038/s42003-019-0306-9> PMID: 30775457; PubMed Central PMCID: PMC6368629.
  27. Stonebloom S, Ebert B, Xiong G, Pattathil S, Birdseye D, Lao J, et al. A DUF-246 family glycosyltransferase-like gene affects male fertility and the biosynthesis of pectic arabinogalactans. *BMC Plant Biol*. 2016; 16:90. Epub 2016/04/20. <https://doi.org/10.1186/s12870-016-0780-x> PMID: 27091363; PubMed Central PMCID: PMC4836069.
  28. Yu CY, Sharma O, Nguyen PHT, Hartono CD, Kanehara K. A pair of DUF538 domain-containing proteins modulates plant growth and trichome development through the transcriptional regulation of GLA-BRA1 in Arabidopsis thaliana. *The Plant journal: for cell and molecular biology*. 2021; 108(4):992–1004. Epub 2021/09/09. <https://doi.org/10.1111/tpj.15487> PMID: 34496091.

29. Urbanowicz BR, Peña MJ, Ratnaparkhe S, Avci U, Backe J, Steet HF, et al. 4-O-methylation of glucuronic acid in Arabidopsis glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(35):14253–8. Epub 2012/08/16. <https://doi.org/10.1073/pnas.1208097109> PMID: 22893684; PubMed Central PMCID: PMC3435161.
30. Moriwaki T, Miyazawa Y, Kobayashi A, Uchida M, Watanabe C, Fujii N, et al. Hormonal regulation of lateral root development in Arabidopsis modulated by MIZ1 and requirement of GNOM activity for MIZ1 function. *Plant physiology*. 2011; 157(3):1209–20. Epub 2011/09/24. <https://doi.org/10.1104/pp.111.186270> PMID: 21940997; PubMed Central PMCID: PMC3252132.
31. Zúñiga-Sánchez E, Soriano D, Martínez-Barajas E, Orozco-Segovia A, Gamboa-deBuen A. BIIDX1, the At4g32460 DUF642 gene, is involved in pectin methyl esterase regulation during Arabidopsis thaliana seed germination and plant development. *BMC Plant Biol*. 2014; 14:338. Epub 2014/12/03. <https://doi.org/10.1186/s12870-014-0338-8> PMID: 25442819; PubMed Central PMCID: PMC4264326.
32. Tong H, Leasure CD, Hou X, Yuen G, Briggs W, He ZH. Role of root UV-B sensing in Arabidopsis early seedling development. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(52):21039–44. Epub 2008/12/17. <https://doi.org/10.1073/pnas.0809942106> PMID: 19075229; PubMed Central PMCID: PMC2634920.
33. Cao X, Yang KZ, Xia C, Zhang XQ, Chen LQ, Ye D. Characterization of DUF724 gene family in Arabidopsis thaliana. *Plant molecular biology*. 2010; 72(1–2):61–73. Epub 2009/10/02. <https://doi.org/10.1007/s11103-009-9551-5> PMID: 19795213.
34. Jones-Rhoades MW, Borevitz JO, Preuss D. Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. *PLoS genetics*. 2007; 3(10):1848–61. Epub 2007/10/17. <https://doi.org/10.1371/journal.pgen.0030171> PMID: 17937500; PubMed Central PMCID: PMC2014789.
35. Kim SJ, Ryu MY, Kim WT. Suppression of Arabidopsis RING-DUF1117 E3 ubiquitin ligases, AtRDUF1 and AtRDUF2, reduces tolerance to ABA-mediated drought stress. *Biochemical and biophysical research communications*. 2012; 420(1):141–7. Epub 2012/03/13. <https://doi.org/10.1016/j.bbrc.2012.02.131> PMID: 22405823.
36. Mewalal R, Mizrachi E, Coetzee B, Mansfield SD, Myburg AA. The Arabidopsis Domain of Unknown Function 1218 (DUF1218) Containing Proteins, MODIFYING WALL LIGNIN-1 and 2 (At1g31720/MWL-1 and At4g19370/MWL-2) Function Redundantly to Alter Secondary Cell Wall Lignin Content. *PLoS one*. 2016; 11(3):e0150254. Epub 2016/03/02. <https://doi.org/10.1371/journal.pone.0150254> PMID: 26930070; PubMed Central PMCID: PMC4773003 commercial sponsor of the work. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials (as detailed online in the guide for authors (<http://www.PLOSone.org/static/editorial.action#competing>)).
37. Li Y, Huang Y, Wen Y, Wang D, Liu H, Li Y, et al. The domain of unknown function 4005 (DUF4005) in an Arabidopsis IQD protein functions in microtubule binding. *The Journal of biological chemistry*. 2021; 297(1):100849. Epub 2021/06/01. <https://doi.org/10.1016/j.jbc.2021.100849> PMID: 34058197; PubMed Central PMCID: PMC8246641.
38. Yang Q, Niu X, Tian X, Zhang X, Cong J, Wang R, et al. Comprehensive genomic analysis of the DUF4228 gene family in land plants and expression profiling of ATDUF4228 under abiotic stresses. *BMC Genomics*. 2020; 21(1):12. Epub 2020/01/05. <https://doi.org/10.1186/s12864-019-6389-3> PMID: 31900112; PubMed Central PMCID: PMC6942412.
39. Aluru C, Singh M. Improved inference of tandem domain duplications. *Bioinformatics (Oxford, England)*. 2021; 37(Suppl\_1):i133–i41. Epub 2021/07/13. <https://doi.org/10.1093/bioinformatics/btab329> PMID: 34252920; PubMed Central PMCID: PMC8275333.
40. Yang J, Zhang B, Gu G, Yuan J, Shen S, Jin L, et al. Genome-wide identification and expression analysis of the R2R3-MYB gene family in tobacco (*Nicotiana tabacum* L.). *BMC genomics*. 2022; 23(1):432. Epub 2022/06/11. <https://doi.org/10.1186/s12864-022-08658-7> PMID: 35681121; PubMed Central PMCID: PMC9178890.
41. Zhu Y, Wu N, Song W, Yin G, Qin Y, Yan Y, et al. Soybean (*Glycine max*) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. *BMC plant biology*. 2014; 14:93. Epub 2014/04/12. <https://doi.org/10.1186/1471-2229-14-93> PMID: 24720629; PubMed Central PMCID: PMC4021193.
42. Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC plant biology*. 2004; 4:10. Epub 2004/06/03. <https://doi.org/10.1186/1471-2229-4-10> PMID: 15171794; PubMed Central PMCID: PMC446195.
43. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in genetics: TIG*. 2002; 18(9):486. Epub 2002/08/15. [https://doi.org/10.1016/s0168-9525\(02\)02722-1](https://doi.org/10.1016/s0168-9525(02)02722-1) PMID: 12175810.



44. Park JI, Ishimizu T, Suwabe K, Sudo K, Masuko H, Hakozaiki H, et al. UDP-glucose pyrophosphorylase is rate limiting in vegetative and reproductive phases in *Arabidopsis thaliana*. *Plant & cell physiology*. 2010; 51(6):981–96. Epub 2010/05/04. <https://doi.org/10.1093/pcp/pcq057> PMID: 20435647.
45. Costa M, Nobre MS, Becker JD, Masiero S, Amorim MI, Pereira LG, et al. Expression-based and co-localization detection of arabinogalactan protein 6 and arabinogalactan protein 11 interactors in *Arabidopsis* pollen and pollen tubes. *BMC plant biology*. 2013; 13:7. Epub 2013/01/10. <https://doi.org/10.1186/1471-2229-13-7> PMID: 23297674; PubMed Central PMCID: PMC3546934.
46. Coimbra S, Costa M, Jones B, Mendes MA, Pereira LG. Pollen grain development is compromised in *Arabidopsis* *agp6 agp11* null mutants. *Journal of experimental botany*. 2009; 60(11):3133–42. Epub 2009/05/13. <https://doi.org/10.1093/jxb/erp148> PMID: 19433479; PubMed Central PMCID: PMC2718217.
47. Kaur D, Moreira D, Coimbra S, Showalter AM. Hydroxyproline-O-Galactosyltransferases Synthesizing Type II Arabinogalactans Are Essential for Male Gametophytic Development in *Arabidopsis*. *Frontiers in plant science*. 2022; 13:935413. Epub 2022/07/02. <https://doi.org/10.3389/fpls.2022.935413> PMID: 35774810; PubMed Central PMCID: PMC9237623.
48. Levitin B, Richter D, Markovich I, Zik M. Arabinogalactan proteins 6 and 11 are required for stamen and pollen function in *Arabidopsis*. *The Plant journal: for cell and molecular biology*. 2008; 56(3):351–63. Epub 2008/07/23. <https://doi.org/10.1111/j.1365-3113X.2008.03607.x> PMID: 18644001.
49. Pereira AM, Masiero S, Nobre MS, Costa ML, Solís MT, Testillano PS, et al. Differential expression patterns of arabinogalactan proteins in *Arabidopsis thaliana* reproductive tissues. *Journal of experimental botany*. 2014; 65(18):5459–71. Epub 2014/07/24. <https://doi.org/10.1093/jxb/eru300> PMID: 25053647; PubMed Central PMCID: PMC4400541.
50. Morato do Canto A, Ceciliato PH, Ribeiro B, Ortiz Morea FA, Franco Garcia AA, Silva-Filho MC, et al. Biological activity of nine recombinant AtRALF peptides: implications for their perception and function in *Arabidopsis*. *Plant physiology and biochemistry: PPB*. 2014; 75:45–54. Epub 2013/12/26. <https://doi.org/10.1016/j.plaphy.2013.12.005> PMID: 24368323.
51. Xia C, Wang YJ, Liang Y, Niu QK, Tan XY, Chu LC, et al. The ARID-HMG DNA-binding protein AtHMGB15 is required for pollen tube growth in *Arabidopsis thaliana*. *The Plant journal: for cell and molecular biology*. 2014; 79(5):741–56. Epub 2014/06/14. <https://doi.org/10.1111/tpj.12582> PMID: 24923357.
52. Yang Y, Niu Y, Chen T, Zhang H, Zhang J, Qian D, et al. The phospholipid flippase ALA3 regulates pollen tube growth and guidance in *Arabidopsis*. *The Plant cell*. 2022; 34(10):3718–36. Epub 2022/07/22. <https://doi.org/10.1093/plcell/koac208> PMID: 35861414; PubMed Central PMCID: PMC9516151.
53. Zhou Y, Yang Y, Niu Y, Fan T, Qian D, Luo C, et al. The Tip-Localized Phosphatidylserine Established by *Arabidopsis* ALA3 Is Crucial for Rab GTPase-Mediated Vesicle Trafficking and Pollen Tube Growth. *The Plant cell*. 2020; 32(10):3170–87. Epub 2020/08/21. <https://doi.org/10.1105/tpc.19.00844> PMID: 32817253; PubMed Central PMCID: PMC7534478.
54. Szumlanski AL, Nielsen E. The Rab GTPase RabA4d regulates pollen tube tip growth in *Arabidopsis thaliana*. *The Plant cell*. 2009; 21(2):526–44. Epub 2009/02/12. <https://doi.org/10.1105/tpc.108.060277> PMID: 19208902; PubMed Central PMCID: PMC2660625.
55. Sorensen AM, Kröber S, Unte US, Huijser P, Dekker K, Saedler H. The *Arabidopsis* ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. *The Plant journal: for cell and molecular biology*. 2003; 33(2):413–23. Epub 2003/01/22. <https://doi.org/10.1046/j.1365-3113x.2003.01644.x> PMID: 12535353.
56. Zhang W, Sun Y, Timofejeva L, Chen C, Grossniklaus U, Ma H. Regulation of *Arabidopsis* tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development (Cambridge, England)*. 2006; 133(16):3085–95. Epub 2006/07/13. <https://doi.org/10.1242/dev.02463> PMID: 16831835.
57. Zhang ZB, Zhu J, Gao JF, Wang C, Li H, Li H, et al. Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. *The Plant journal: for cell and molecular biology*. 2007; 52(3):528–38. Epub 2007/08/31. <https://doi.org/10.1111/j.1365-3113X.2007.03254.x> PMID: 17727613.
58. Zhu J, Chen H, Li H, Gao JF, Jiang H, Wang C, et al. Defective in Tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. *The Plant journal: for cell and molecular biology*. 2008; 55(2):266–77. Epub 2008/04/10. <https://doi.org/10.1111/j.1365-3113X.2008.03500.x> PMID: 18397379.
59. Zhu J, Lou Y, Xu X, Yang ZN. A genetic pathway for tapetum development and function in *Arabidopsis*. *Journal of integrative plant biology*. 2011; 53(11):892–900. Epub 2011/10/01. <https://doi.org/10.1111/j.1744-7909.2011.01078.x> PMID: 21957980.

60. Farooq O, Ali M, Sarwar N, Mazhar Iqbal M, Naz T, Asghar M, et al. Foliar applied brassica water extract improves the seedling development of wheat and chickpea. *Asian Journal of Agriculture and Biology*. 2021;8. <https://doi.org/10.35495/ajab.2020.04.219>
61. Ahmad N, Fazli Ahad R, Iqbal T, Khan N, Nauman M, Hameed F, editors. Genetic analysis of biochemical traits in F3 populations of rapeseed (*Brassica napus* L.)2020: *Asian Journal of Agriculture and Biology*.
62. Raiola A, Errico A, Petruk G, Monti DM, Barone A, Rigano MM. Bioactive Compounds in Brassicaceae Vegetables with a Role in the Prevention of Chronic Diseases. *Molecules* (Basel, Switzerland). 2017; 23(1). Epub 2018/01/04. <https://doi.org/10.3390/molecules23010015> PMID: 29295478; PubMed Central PMCID: PMC5943923.
63. Xie T, Chen C, Li C, Liu J, Liu C, He Y. Genome-wide investigation of WRKY gene family in pineapple: evolution and expression profiles during development and stress. *BMC genomics*. 2018; 19(1):490. Epub 2018/06/27. <https://doi.org/10.1186/s12864-018-4880-x> PMID: 29940851; PubMed Central PMCID: PMC6019807.
64. The Arabidopsis Genome I. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000; 408(6814):796–815. <https://doi.org/10.1038/35048692> PMID: 11130711
65. Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, et al. The genome of the extremophile crucifer *Thellungiella parvula*. *Nature genetics*. 2011; 43(9):913–8. Epub 2011/08/09. <https://doi.org/10.1038/ng.889> PMID: 21822265; PubMed Central PMCID: PMC3586812.
66. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature genetics*. 2011; 43(5):476–81. Epub 2011/04/12. <https://doi.org/10.1038/ng.807> PMID: 21478890; PubMed Central PMCID: PMC3083492.
67. Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, et al. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nature genetics*. 2013; 45(8):891–8. Epub 2013/07/03. <https://doi.org/10.1038/ng.2684> PMID: 23817568.
68. Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo YL, et al. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature genetics*. 2013; 45(7):831–5. Epub 2013/06/12. <https://doi.org/10.1038/ng.2669> PMID: 23749190.
69. Yang R, Jarvis DE, Chen H, Beilstein MA, Grimwood J, Jenkins J, et al. The Reference Genome of the Halophytic Plant *Eutrema salsugineum*. *Frontiers in plant science*. 2013; 4:46. Epub 2013/03/23. <https://doi.org/10.3389/fpls.2013.00046> PMID: 23518688; PubMed Central PMCID: PMC3604812.
70. Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, et al. The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nature communications*. 2014; 5:3706. Epub 2014/04/25. <https://doi.org/10.1038/ncomms4706> PMID: 24759634; PubMed Central PMCID: PMC4015329.
71. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, et al. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature communications*. 2014; 5:3930. Epub 2014/05/24. <https://doi.org/10.1038/ncomms4930> PMID: 24852848; PubMed Central PMCID: PMC4279128.
72. Dorn KM, Fankhauser JD, Wyse DL, Marks MD. A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA research: an international journal for rapid publication of reports on genes and genomes*. 2015; 22(2):121–31. Epub 2015/01/30. <https://doi.org/10.1093/dnares/dsu045> PMID: 25632110; PubMed Central PMCID: PMC4401323.
73. Gan X, Hay A, Kwantes M, Haberer G, Hallab A, Ioio RD, et al. The *Cardamine hirsuta* genome offers insight into the evolution of morphological diversity. *Nature plants*. 2016; 2(11):16167. Epub 2016/11/04. <https://doi.org/10.1038/nplants.2016.167> PMID: 27797353; PubMed Central PMCID: PMC8826541.
74. Jeong YM, Kim N, Ahn BO, Oh M, Chung WH, Chung H, et al. Elucidating the triplicated ancestral genome structure of radish based on chromosome-level comparison with the *Brassica* genomes. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik*. 2016; 129(7):1357–72. Epub 2016/04/04. <https://doi.org/10.1007/s00122-016-2708-0> PMID: 27038817.
75. Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, et al. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nature genetics*. 2016; 48(10):1225–32. Epub 2016/09/07. <https://doi.org/10.1038/ng.3657> PMID: 27595476.
76. Zhang J, Tian Y, Yan L, Zhang G, Wang X, Zeng Y, et al. Genome of Plant Maca (*Lepidium meyenii*) Illuminates Genomic Basis for High-Altitude Adaptation in the Central Andes. *Molecular plant*. 2016; 9(7):1066–77. Epub 2016/05/14. <https://doi.org/10.1016/j.molp.2016.04.016> PMID: 27174404.
77. Briskine RV, Paape T, Shimizu-Inatsugi R, Nishiyama T, Akama S, Sese J, et al. Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary

- ecology. Molecular ecology resources. 2017; 17(5):1025–36. Epub 2016/10/27. <https://doi.org/10.1111/1755-0998.12604> PMID: 27671113.
78. Byrne SL, Erthmann P, Agerbirk N, Bak S, Hauser TP, Nagy I, et al. The genome sequence of *Barbarea vulgaris* facilitates the study of ecological biochemistry. Scientific reports. 2017; 7:40728. Epub 2017/01/18. <https://doi.org/10.1038/srep40728> PMID: 28094805; PubMed Central PMCID: PMC5240624.
  79. Cai C, Wang X, Liu B, Wu J, Liang J, Cui Y, et al. Brassica rapa Genome 2.0: A Reference Upgrade through Sequence Re-assembly and Gene Re-annotation. Molecular plant. 2017; 10(4):649–51. Epub 2016/11/29. <https://doi.org/10.1016/j.molp.2016.11.008> PMID: 27890636.
  80. Jiao WB, Accinelli GG, Hartwig B, Kiefer C, Baker D, Severing E, et al. Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. Genome research. 2017; 27(5):778–86. Epub 2017/02/06. <https://doi.org/10.1101/gr.213652.116> PMID: 28159771; PubMed Central PMCID: PMC5411772.
  81. Lee CR, Wang B, Mojica JP, Mandáková T, Prasad K, Goicoechea JL, et al. Young inversion with multiple linked QTLs under selection in a hybrid zone. Nature ecology & evolution. 2017; 1(5):119. Epub 2017/08/16. <https://doi.org/10.1038/s41559-017-0119> PMID: 28812690; PubMed Central PMCID: PMC5607633.
  82. Kliver S, Rayko M, Komissarov A, Bakin E, Zhernakova D, Prasad K, et al. Assembly of the *Boechera retrofracta* Genome and Evolutionary Analysis of Apomixis-Associated Genes. Genes. 2018; 9(4):1–16. Epub 2018/03/31. <https://doi.org/10.3390/genes9040185> PMID: 29597328; PubMed Central PMCID: PMC5924527.
  83. Kang M, Wu H, Yang Q, Huang L, Hu Q, Ma T, et al. A chromosome-scale genome assembly of *Isatis indigotica*, an important medicinal plant used in traditional Chinese medicine: An *Isatis* genome. Horticulture research. 2020; 7:18. Epub 2020/02/07. <https://doi.org/10.1038/s41438-020-0240-5> PMID: 32025321; PubMed Central PMCID: PMC6994597.
  84. Mishra B, Ploch S, Runge F, Schmuker A, Xia X, Gupta DK, et al. The Genome of *Microthlaspi erraticum* (Brassicaceae) Provides Insights Into the Adaptation to Highly Calcareous Soils. Frontiers in plant science. 2020; 11:943. Epub 2020/07/29. <https://doi.org/10.3389/fpls.2020.00943> PMID: 32719698; PubMed Central PMCID: PMC7350527.
  85. Perumal S, Koh CS, Jin L, Buchwaldt M, Higgins EE, Zheng C, et al. A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral Brassica genome. Nature plants. 2020; 6(8):929–41. Epub 2020/08/13. <https://doi.org/10.1038/s41477-020-0735-y> PMID: 32782408; PubMed Central PMCID: PMC7419231.
  86. Song JM, Guan Z, Hu J, Guo C, Yang Z, Wang S, et al. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. Nature plants. 2020; 6(1):34–45. Epub 2020/01/15. <https://doi.org/10.1038/s41477-019-0577-7> PMID: 31932676; PubMed Central PMCID: PMC6965005.
  87. Song X, Wei Y, Xiao D, Gong K, Sun P, Ren Y, et al. Brassica carinata genome characterization clarifies U's triangle model of evolution and polyploidy in Brassica. Plant physiology. 2021; 186(1):388–406. Epub 2021/02/19. <https://doi.org/10.1093/plphys/kiab048> PMID: 33599732; PubMed Central PMCID: PMC8154070.
  88. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic acids research. 2018; 46(W1):W200–w4. Epub 2018/06/16. <https://doi.org/10.1093/nar/gky448> PMID: 29905871; PubMed Central PMCID: PMC6030962.
  89. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic acids research. 2021; 49(D1):D412–d9. Epub 2020/10/31. <https://doi.org/10.1093/nar/gkaa913> PMID: 33125078; PubMed Central PMCID: PMC7779014.
  90. Christiam C, Grzegorz MB, Victor J, Roberto Vera A, Thomas LM. ElasticBLAST: Accelerating Sequence Search via Cloud Computing. BMC bioinformatics. 2023; 24(1):2023.01.04.522777. <https://doi.org/10.1186/s12859-023-05245-9> PMID: 36967390
  91. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic acids research. 2015; 43(Database issue):D222–D6. Epub 2014/11/22. <https://doi.org/10.1093/nar/gku1221> PMID: 25414356; PubMed Central PMCID: PMC4383992.
  92. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic acids research. 2023; 51(D1):D418–D27. Epub 2022/11/10. <https://doi.org/10.1093/nar/gkac993> PMID: 36350672; PubMed Central PMCID: PMC9825450.
  93. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic acids research. 2021; 49(D1):D458–D60. Epub 2020/10/27. <https://doi.org/10.1093/nar/gkaa937> PMID: 33104802; PubMed Central PMCID: PMC7778883.

94. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013; 30(4):772–80. Epub 2013/01/19. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690; PubMed Central PMCID: PMC3603318.
95. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*. 2015; 32(1):268–74. Epub 2014/11/06. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430; PubMed Central PMCID: PMC4271533.
96. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic acids research*. 2021; 49(W1):W293–w6. <https://doi.org/10.1093/nar/gkab301> PMID: 33885785
97. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*. 2009; 37(Web Server issue):W202–W8. Epub 2009/05/22. <https://doi.org/10.1093/nar/gkp335> PMID: 19458158; PubMed Central PMCID: PMC2703892.
98. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology*. 2019; 37(4):420–3. Epub 2019/02/20. <https://doi.org/10.1038/s41587-019-0036-z> PMID: 30778233.
99. Chou KC, Shen HB. Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PloS one*. 2010; 5(6):e11335. Epub 2010/07/03. <https://doi.org/10.1371/journal.pone.0011335> PMID: 20596258; PubMed Central PMCID: PMC2893129.
100. Rombauts S, Déhais P, Van Montagu M, Rouzé P. PlantCARE, a plant cis-acting regulatory element database. *Nucleic acids research*. 1999; 27(1):295–6. Epub 1998/12/10. <https://doi.org/10.1093/nar/27.1.295> PMID: 9847207; PubMed Central PMCID: PMC148162.
101. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic acids research*. 2012; 40(7):e49. Epub 2012/01/06. <https://doi.org/10.1093/nar/gkr1293> PMID: 22217600; PubMed Central PMCID: PMC3326336.
102. Li DD, Xue JS, Zhu J, Yang ZN. Gene Regulatory Network for Tapetum Development in *Arabidopsis thaliana*. *Frontiers in plant science*. 2017; 8:1559. Epub 2017/09/29. <https://doi.org/10.3389/fpls.2017.01559> PMID: 28955355; PubMed Central PMCID: PMC5601042.