

RESEARCH

Open Access



Integrating differential expression under drought with gene family expansion unique to drought-tolerant species prioritizes candidate genes for drought adaptation in Brassicaceae species

Carolyn Uebermuth-Feldhaus¹ and Heiko Schoof^{1*}

Abstract

Background In order to prioritize candidate genes for drought adaptation, we analyze genomic data of the more drought-resistant Brassicaceae species *Eutrema salsugineum* and *Arabidopsis lyrata* compared to more drought-sensitive species *Arabidopsis thaliana* and *Brassica napus*. We combine gene family expansion, which is an important driver of evolution in plants, unique to the drought-resistant species with differential expression under drought (DE).

Results We show that combining trait-specific gene family expansion with differential expression identifies a concise set of candidate genes. To demonstrate that these are relevant for drought adaptation in tolerant species, we show enrichment of DE conserved between both tolerant species, DE unique to the tolerant species, and up-regulation. We show that specific functions are enriched, and that the set contains genes with functions such as root development in line with drought adaptation based on evidence from other species, while the background of all differentially expressed genes (DEGs) contains many general stress response genes. Whereas DEGs in general are rarely under diversifying selection, signatures of diversifying selection are slightly enriched in the candidate gene families, highly significantly enriched in DEGs in tolerant species-specific expanded gene families, and, in contrast, not enriched in DEGs in sensitive species-specific expanded gene families.

Conclusions Our approach identifies a concise and functionally relevant set of candidate genes for drought adaptation with promising targets for functional studies and crop improvement for drought tolerance. We propose that our method can also be used to prioritize candidate genes for adaptation to other environmental factors.

Keywords Adaptation, Drought, Comparative genomics, Gene family expansion, Differential expression, Evolution, Brassicaceae, Diversifying selection, Abiotic stress resistance, Bioinformatics

*Correspondence:

Heiko Schoof
schoof@uni-bonn.de

¹Faculty of Agricultural, Nutritional and Engineering Sciences, University of Bonn, INRES Crop Bioinformatics, Katzenburgweg 2, 53115 Bonn, Germany



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

Drought stress experiments often identify thousands of differentially expressed genes and many of them are general stress genes ([1], [2], [3]). Moreover, drought reactions of plants include effects which are unfavorable for crop production, such as the reduction of photosynthesis ([4], [5], [6], [7]), because they reduce biomass accumulation. The goal of this study is to prioritize genes which function in the adaptation to drought from all genes which react to drought. For this purpose, we combine gene family expansion unique to drought-resistant plant species with differential expression under drought (DE).

Gene family expansion through whole genome duplication (WGD) is common in plants ([8], [9]). Gene duplicates can acquire new functions (neofunctionalization, including new regulation patterns) or split their functions (subfunctionalization) (reviewed in e.g., [10]). Examples of experimentally confirmed duplicates which confer abiotic stress tolerance in *Arabidopsis thaliana* are (1) a duplication in an E3 ubiquitin ligase, which is responsible for drought-tolerance [11], and (2) a duplication of a mitochondrial GrpE (Mge) protein which is responsible for heat stress tolerance [12]. In *Eutrema salsugineum*, there are gene family expansions which were experimentally shown to be relevant for its salt tolerance: (1) a duplication in a calcium sensor (*ATCBL10*, AT4G33000) in *Eutrema salsugineum* (EUTSA_v10026019mg and EUTSA_v10028908mg) is important for calcium-signaling and increased salt tolerance in *Arabidopsis thaliana* mutants [13], and (2) HKT1 (AT4G10310) is duplicated and relevant for salt tolerance in *Eutrema salsugineum* [14].

Differential expression upon drought can link a gene duplication to a function under drought stress. The authors of different studies have pointed out expanded gene families among differentially expressed genes, e.g., in a salt tolerant *Populus* species under salt treatment [15], in a desiccation tolerant *Lindernia* species under drought and recovery [16] or in roots of soybeans during nodulation [17]. The relationship between differential expression under abiotic stress and gene family expansion has been systematically described: 30% of duplicated genes in *Eutrema salsugineum* and 26–27% of duplicated genes in *Arabidopsis thaliana* respond to abiotic stresses [18], and tandem duplicates in the cold tolerant grass *Lolium perenne* responded significantly more (17.48%) often to cold treatment than all genes (9.91%) [19].

A duplication of the *NICOTIANAMINE SYNTHASE3* (AT1G09240) in *Arabidopsis lyrata* is proposed as candidate gene for stress resistance, as it has likely neofunctionalized, i.e., acquired expression in a different tissue and under drought and salt stress [20]. Laha and colleagues [21] review the research on stress genomics focused on gene family expansion in Brassicaceae species

and discuss approaches to identify stress-responsive duplicates. They propose a strategy to identify stress-responsive allelic variants for the engineering of stress tolerant plants by characterizing stress-responsive genes retained as multiple copies in tolerant species, while present only as a single copy in the susceptible *Arabidopsis thaliana*. Oh and Dassanayake [22] state that it is important to perform comparative studies on species with natural tolerance to abiotic stress because it is not possible to identify stress tolerance relevant expression and physiological reactions in species which are sensitive to this stress.

In this study, we apply a generalizable approach to identify trait-specific evolutionary adaptations [23] to drought tolerance in Brassicaceae. We identify gene family expansion in the drought-tolerant Brassicaceae species *Arabidopsis lyrata* (*Aly*) and *Eutrema salsugineum* (*Esa*) compared to the more drought-sensitive *Arabidopsis thaliana* (*Ath*) and *Brassica napus* (*Bna*) and combine this with genes which are differentially expressed under drought. *Arabidopsis lyrata* is a biennial or perennial which naturally grows on dry soils. Bouzid and colleagues [24] observed drought avoidance and tolerance mechanisms in *Aly*: The leaf water content decreased slower before wilting and the photosynthetic capacity was kept high at wilting compared to *Ath*. Additionally, *Aly* plants had a higher survival rate and showed less damage from wilting at re-watering ([25], [24]). *Aly* reacts earlier to drought than *Esa* and *Ath* with transcriptional change and growth reduction [25] and wilting compared to *Ath* [24]. Marín-de la Rosa and colleagues [25] showed that the stomata in *Aly* are less open already under normal conditions compared to *Ath*, and they close less upon ABA treatment. Congruent with this, several ABA receptors are constantly higher expressed. They observed this “stress-aware” state in *Aly* and *Esa*. *Eutrema salsugineum* (syn: *Thellungiella salsuginea*) is a halophytic model plant which also shows tolerance towards several other stresses, including drought tolerance (reviewed in [26]). Traits related to drought tolerance in *Esa* include morphological plasticity depending on growth conditions and an increased accumulation of solutes ([27]). *Esa* shows an increased proline content and a constantly high expression of the proline biosynthesis gene P5CS1, even under well-watered conditions [25]. This together with the limited opening of stomata, described above for *Aly*, suggests that *Esa* has several “stress-awareness” mechanisms. *Esa* can maintain the plant water content even under drought conditions [26]. Further, *Esa* responds later with growth reduction, i.e., at a lower soil water content, than *Aly* and *Ath* [25]. Its survival rate after re-watering is higher than that of *Ath* and equal to that of *Aly* [25], probably due to several tolerance mechanisms, some of which are described above. *Arabidopsis thaliana*

is an annual plant which can escape drought by quickly terminating its life cycle. *Ath* also shows some drought avoidance mechanisms, e.g., *Ath* was able to maintain thicker leaves and started wilting later than *Aly* [24]. But the survival rate of *Ath* after wilting is much lower than that of *Aly* and *Esa* ([24], [25]), which is eminently important for crop production, i.e., yield stability. *Brassica napus* is an annual or biennial oil seed crop. It is an allopolyploid species which contains the A and C *Brassica* genome with $2n=4x=38$ [28]. Its yield is highly reduced by drought [29].

Knowledge about the genetic basis of drought adaptation is relevant for breeding of drought-resistant crops. While there are many indicators that neo- or subfunctionalized gene duplicates play an important role in adaptation to abiotic stresses, the knowledge about gene family expansion which confers drought resistance to adapted Brassicaceae species is low. We took advantage of these indicators to prioritize candidate genes for drought resistance from the general drought-responsive genes using our recently published automated workflow [23]. For this we integrated publicly available transcriptomic and genomic data from tolerant and sensitive Brassicaceae species. We systematically described the relationship between gene family expansion and differential expression under drought with focus on the drought-tolerant species. We evaluated the prioritization of candidate genes by identifying enrichment with signatures of diversifying selection.

Results

Protein coding gene families in four Brassicaceae species

We assigned the genes of proteins with a minimum length of 100 amino acids (aa) from the four Brassicaceae species *Eutrema salsugineum* (*Esa*), *Arabidopsis lyrata* (*Aly*), *Arabidopsis thaliana* (*Ath*) and *Brassica napus* (*Bna*) (Fig. 1) to gene families (orthologous groups, HOGs) using OrthoFinder ([30], version 2.5.4) in order to perform comparative genome analyses. There were 17,488 HOGs which were conserved between the four species (Supplementary Fig. 2, Additional File 1). These conserved HOGs contained 19,990 genes from *Esa*, 21,173 genes from *Aly*, 19,919 genes from *Ath* and 38,861 genes from *Bna*. There were considerably less HOGs (less than 1000) in all other combinations, e.g., species-specific HOGs or HOGs with representatives of only some species, except for *Bna* specific HOGs (7,573). All analyses in this study were performed on the set of HOGs which are conserved between all four species (Conserved Set, 17,488 HOGs, *Conserved_dataset_matrix.tsv*, [31]), all other HOGs were not analyzed further.

Upon drought, differential expression as well as regulation are conserved between the four Brassicaceae species

We inferred differentially expressed genes (DEGs) between drought and control from publicly available leaf transcriptomes from *Esa*, *Aly* and *Ath* ([25]) and *Bna* ([32]) ($p\text{-adj.} \leq 0.1$, see Methods). In the Conserved Set, 13,531 HOGs (77.4%) contained at least one DEG. There were 7,790 DEGs from *Esa* in 7,434 HOGs (42.5%). In

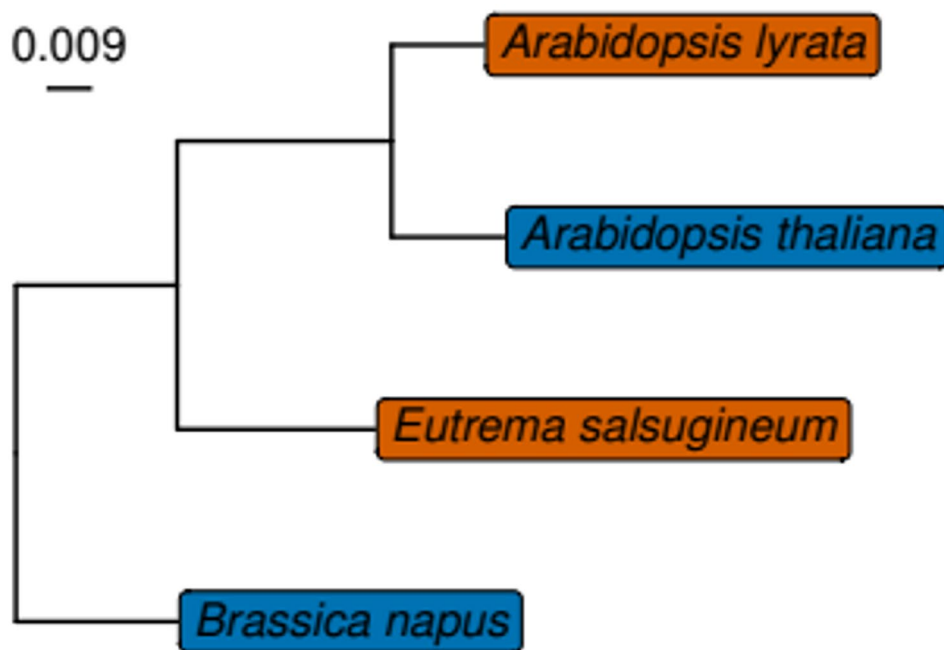


Fig. 1 Phylogenetic tree of the Brassicaceae species in this study as calculated by OrthoFinder [30] on the proteomes of the sequences. Orange marks the drought-tolerant and blue marks the drought-sensitive species. Scale bar represents the number of substitutions per site

Aly, there were 9,478 DEGs in 8,866 HOGs (50.7%). In *Ath*, there were 3,141 DEGs in 2,865 HOGs (16.4%). In *Bna*, there were 10,065 DEGs in 6,956 HOGs (39.8%) (background pie charts in Fig. 2).

In order to find out if the differential expression (DE) was conserved between the four species or rather species specific, we compared the frequency of HOGs with DEG from all four species and HOGs with DEG from none of the species to the frequencies expected from the distribution of HOGs with at least one DEG per species. Using a hypergeometric test for over representation at significance level 0.05, we found significantly more HOGs than expected which contained DEG from all four species (observed: 846/17488 = 4.8%, expected = 245 based on $p_{s4} = 1.4\%$, $p < 2.2e-16$, see Methods) as well as significantly more HOGs with no DEG from any of the four species (observed: 3957/17488 = 22.6%, expected: 2500 based on $p_{f4} = 14.3\%$, $p < 2.2e-16$). We conclude that, even though the treatments were not the same between the drought experiment in *Bna* (3 days) compared to *Ath*, *Aly* and *Esa* (11 days), differential expression upon drought is mostly conserved between the four species.

We then wanted to know the predominant direction of regulation, i.e., if the reaction to drought is manifested rather by up or down-regulation and if the regulation is rather conserved or if it differs between the species. For

this purpose, we categorized the DE gene families with at least 2 DEGs regarding their pattern of regulation, where “up” means that all DEGs are up-regulated, “down” means that all DEGs are down-regulated and “up_and_down” means that there are up-regulated as well as down-regulated genes in the HOG. The regulation within the HOGs with at least two DEGs in the Conserved Set was equally distributed among the regulation categories (2768 only up, 2935 only down and 3097 gene families with up- and down-regulated genes, Supplementary Fig. 4, Additional File 1). Hence, in 2/3 of the HOGs which had more than one DEG, these genes were equally regulated. In only 1/3 of the tested HOGs, the genes were regulated in opposing directions (light blue in Supplementary Fig. 4, Additional File 1 and light blue columns in Supplementary Fig. 5, Additional File 1). We found that HOGs which showed contrasting regulation between the species mostly involved DE in *Bna* (Supplementary Fig. 5, Additional File 4, highlighted with asterisks). There was also more within-species divergent regulation in *Bna* compared to the other species (down_and_up in *Bna*: 230 HOGs, *Aly* 76 HOGs, *Esa* 49 HOGs and *Ath* 6 HOGs, Supplementary Fig. 5, Additional File 1).

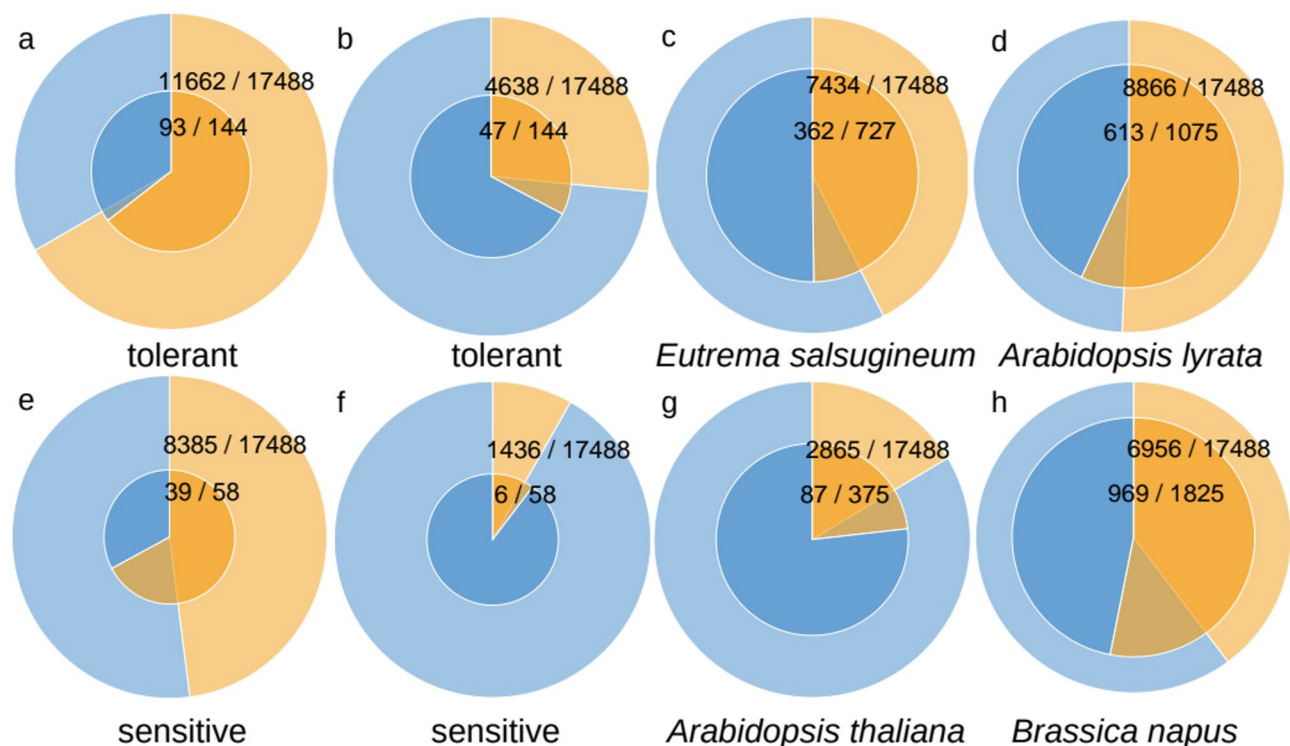


Fig. 2 Pie charts of the number of HOGs with DEG (orange) or without DEG (blue) within the Conserved Set (outer circle) or within the subset of expanded HOGs (inner circle, log10-scale) as defined by the names below the circles. The DEGs are from **a**) any of the tolerant species, **b**) both tolerant species, **c**) *Esa*, **d**) *Aly*, **e**) any of the sensitive species, **f**) both sensitive species, **g**) *Ath* or **h**) *Bna*

Gene family expansion

We defined gene family expansion as HOGs that contained at least twice the number of genes from the species of interest compared to each of the species with the contrasting drought tolerance (see Methods). From the HOGs which were conserved between the four species we identified 727 HOGs which were expanded in the drought-tolerant species *Esa* compared to the drought-sensitive species *Ath* and *Bna*, and 1075 HOGs which were expanded in *Aly* compared to both sensitive species. 144 of these expansions in *Esa* or *Aly* were common to both tolerant species. As for the drought-sensitive species, 375 HOGs were expanded in *Ath* compared to the drought-tolerant species *Esa* and *Aly*, and 1825 HOGs were expanded in *Bna* compared to both tolerant species. 58 of these expansions in *Ath* or *Bna* were common to both sensitive species.

When considering the lineage-specific duplications, we identified 449 HOGs which were commonly expanded in *Esa* and both *Arabidopsis* species compared to *Bna* and 1762 HOGs which were expanded in *Bna* compared to *Esa* and both *Arabidopsis* species. 441 HOGs were expanded in both *Arabidopsis* species compared to *Bna* and *Esa* and 103 HOGs were commonly expanded in *Bna* and *Esa* compared to both *Arabidopsis* species.

Differential expression under drought in expanded gene families in tolerant and sensitive species

Next we analyzed the differential expression (DE) of genes in expanded HOGs. We observed that DE in drought-tolerant species correlated with gene family expansion unique to drought-tolerant species. The HOGs which were expanded in *Esa* but not in any of the sensitive species were significantly more often differentially expressed in *Esa* than expected from the distribution of DEGs in the Conserved Set (363/727 vs. 7434/17488, hypergeometric test for over representation $p = 3.2e-05$, Fig. 2c). The same applied for DE and expansion in *Aly* (613/1075 vs. 8866/17488, $p = 1.0e-05$, Fig. 2d), but also for DE in a drought-sensitive species in HOGs which were expanded in a drought-sensitive species (*Ath*: 87/375 vs. 2865/17488, $p = 0.00035$, *Bna*: 969/1825 vs. 6956/17488, $p = 3.8e-34$, Fig. 2g and h). HOGs which were expanded in both species with the same tolerance compared to the species with contrasting drought tolerance were slightly, but not significantly, enriched for DE of both expanded species (tolerant species 47/144 vs. 4638/17488, $p = 0.06$, sensitive species 6/58 vs. 1436/17488, $p = 0.34$, Fig. 2b and f). We conclude that duplicate genes in general are more often DE.

Thus, we asked if there was a significant enrichment of conserved DE, that is, genes from both species were differentially expressed, in the expanded HOGs. Of the 93 HOGs expanded in both tolerant with DE in at least one

of the tolerant species, 47 HOGs (50%) showed DE in both tolerant species (Supplementary Fig. 4a and b, Additional File 1). This was significantly more than expected from the background, where 4,638 (40%) from 11,662 HOGs which were differentially expressed in any of the two tolerant species were differentially expressed in both tolerant species ($p = 0.022$). In contrast, of the 39 HOGs expanded in both sensitive species with DE in at least one of the sensitive species, only 6 HOGs (15%) showed DE in both sensitive species (Supplementary Fig. 4e and f, Additional File 1). This was slightly less than in the background, where from 8,385 HOGs which were differentially expressed in any sensitive species, 1,436 HOGs (17%) were differentially expressed in both sensitive species. Thus, conserved DE was significantly enriched only in HOGs expanded in both drought-tolerant species, but not in HOGs expanded in both sensitive species.

Then, we analyzed the direction of regulation in the expanded HOGs, i.e., whether up- or down-regulation under drought was more frequent (see above) and if this varied between the expanded species. We found that in the HOGs which were expanded and differentially expressed in the tolerant species, a significantly larger proportion contained only up-regulated genes (51%, $p = 3.9e-3$ in both, 38%, $p = 5e-4$ in *Esa* and 39%, $p = 4e-4$ in *Aly*, Fig. 3a-c) when compared to all HOGs which showed DE in the respective species, where all three regulation categories were equally frequent (Fig. 3a-c background pie charts). In contrast to this, the HOGs expanded in both sensitive species more often showed divergent, both up- and down-regulation (67% HOGs). While this was not significantly more than in the background of all HOGs with DEGs from both sensitive species (Fig. 3d, $p = 0.195$), the sample was only four out of six HOGs expanded in both sensitive species. Looking at the sensitive species individually, the HOGs expanded in *Ath* showed significant enrichment with divergent regulation compared to all HOGs with DE from *Ath* (39%, $p = 0.015$, Fig. 3e), while they were not significantly enriched for only up-regulation (32%, $p = 0.195$, Fig. 3e). The regulation categories of HOGs expanded in *Bna* were distributed similarly to those in all HOGs with DE from *Bna*: nearly half (47%, Fig. 3f) of HOGs showed up- and down-regulation.

To summarize, in tolerant species, but not in sensitive species we observed enrichment of HOGs with (a) conserved differential expression and (b) specifically only up-regulation in expanded HOGs vs. the background.

Interestingly and in contrast to the observation that differential expression was likely to be conserved across species within a HOG, in the HOGs expanded only in tolerant species we found that most genes were differentially expressed only in the tolerant species but not the sensitive species: 28 (59%) out of 47 HOGs (shown in

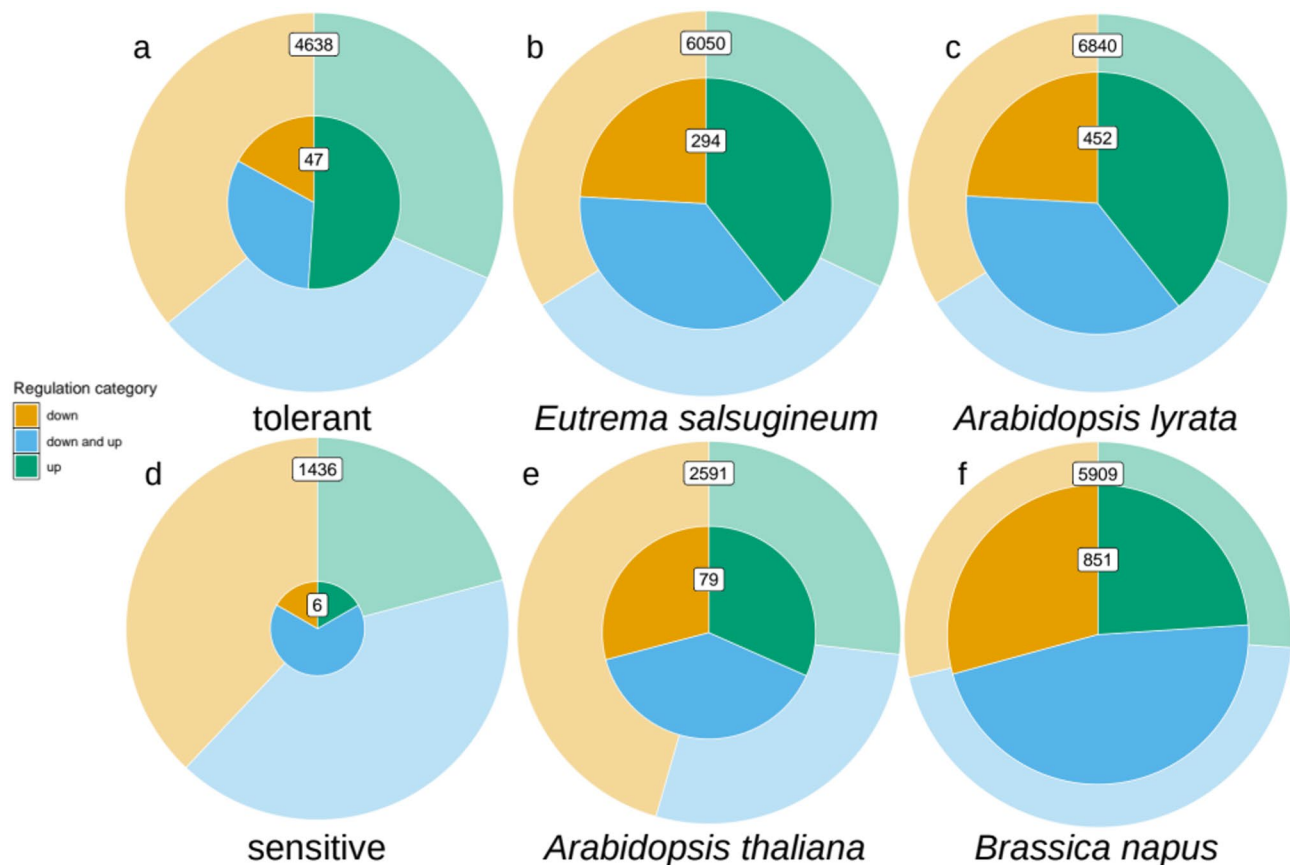


Fig. 3 Pie charts of the proportions of HOGs per regulation category. Green = “up”, Orange = “down”, Blue = “down_and_up”. Each pie chart represents all HOGs with DEGs from the depicted species in the Conserved Set (background) and in its subset (foreground) of HOGs expanded in the depicted species. Only HOGs with at least two DEGs regardless of the species are considered

blue in Fig. 4) which were expanded and showed DE in both tolerant species did not show DE in any of the sensitive species (Fig. 4). This was significantly more ($p = 4.6e-4$) than in the background, where only 1,621 (35%) of 4,638 HOGs which showed DE in both tolerant species were not differentially expressed in the sensitive species (Supplementary Fig. 3, Additional File 1).

Given their strong correlations to the drought-tolerance phenotype we defined the gene families which were commonly expanded in the tolerant species and differentially expressed under drought in at least one of the tolerant species as candidate genes for drought adaptation. We put special focus on two subsets, (a) genes which showed conserved differential expression in both tolerant species and (b) out of these those showing only up-regulation and no DE in sensitive species. Information on these candidate gene families, including regulation under drought of the genes from all four species, are provided in Additional File 2_Candidate_gene_families_and_tolerant_specific_expansions.xlsx, where the corresponding subsets are indicated.

Functions of candidate gene families for drought adaptation

We explored the functions of the prioritized candidate genes for drought adaptation based on function prediction generated by AHRD [33], compared them to the functions of all HOGs in the Conserved Set as well as all DE HOGs and tested for functional enrichment. In addition, we explored the functions of HOGs which were expanded and differentially expressed only in one tolerant species, i.e., in *Esa* only or in *Aly* only (see Supplement). The candidate HOGs as well as the tolerant species-specific expansions are listed in Additional File 2_Candidate_gene_families_and_tolerant_specific_expansions.xlsx, including their predicted functions (human-readable descriptions, HRD).

The set of candidate genes has diverse functions. Only the biological processes “modification-dependent protein catabolic process” and “mRNA splicing, via spliceosome” were significantly enriched in the set of 93 candidate HOGs (Supplementary Table 3, Additional File 1). In its subset of 47 candidate HOGs which were DE in both tolerant species, the biological processes “modification-dependent protein catabolic process” and

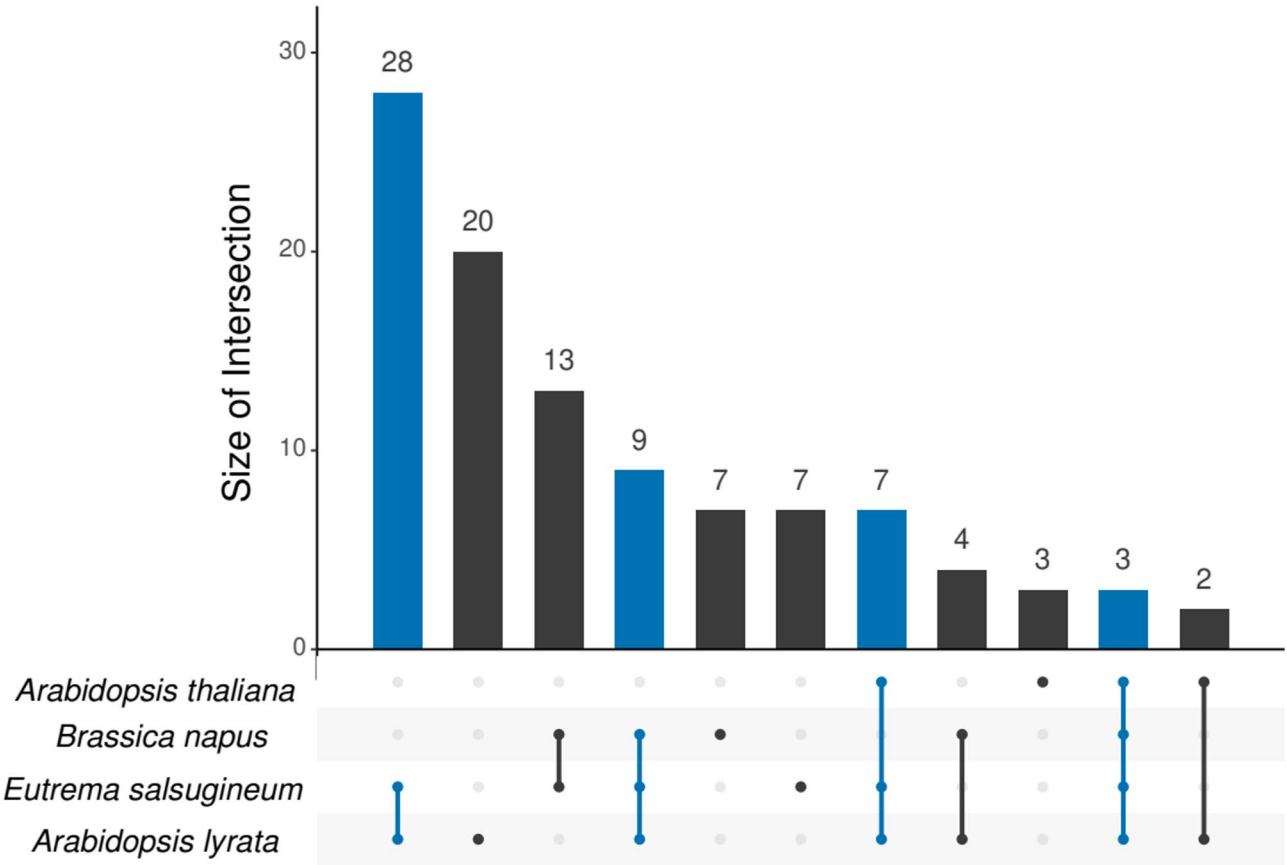


Fig. 4 UpSet plot showing all HOGs expanded in both tolerant species that contain a DEG. The intersections of HOGs with DEG from the given species are shown, where blue marks intersections with both tolerant species. The species are ordered by the number of HOGs, with the largest set on the bottom

“translation” were significantly over-represented (Supplementary Table 4, Additional File 1). Interestingly, neither “modification-dependent protein catabolic process” nor any related post translational regulatory process was enriched in the background of all HOGs which show DE in any or in both tolerant species (Supplementary Figs. 7 and 8, Additional File 1). Candidate genes which function in the enriched processes are described in detail in the Supplementary Results, Additional File 1. In the following, we focus on the candidate genes which are uniquely up-regulated in both tolerant species. Because of the small sample size of 17 HOGs, we have not tested for enriched functions but display all functions in Supplementary Fig. 12, Additional File 1. We inspected all 17 HOGs manually and included these notes in Additional_File_2_Candidate_gene_families_and_tolerant_specific_expansions.xlsx.

The following examples show an interesting relation to drought tolerance. N0.HOG0007350 (Lysine-specific demethylase REF6, Supplementary Fig. 13, Additional File 1), which functions in “positive regulation of lateral root development” (GO:1901333) and “ABA catabolism” (GO:0046345), promotes lateral root formation in *Ath* [34] and is involved in the regulation of flowering [35].

N0.HOG0008546 (AT3G16630) encodes a Kinesin-like protein KIN-13 A (Supplementary Fig. 14, Additional File 1) which functions in trichome morphogenesis [36] and in the formation of secondary cell wall pits [37].

There are several genes which function in regulatory processes: N0.HOG0007005 encodes transcription initiation factor TFIID subunit 9 (AT1G54140, *TAF9*), which functions in “DNA-templated transcription initiation” (GO:0006352). It is duplicated in both tolerant species and highly up-regulated in *Aly* (Supplementary Fig. 15, Additional File 1). N0.HOG0005205 (AT3G49000), DNA-directed RNA polymerase III subunit RPC3, functions in “DNA-templated transcription” (GO:0006351). Moreover, two genes function in epigenetic regulation of gene expression: (1) the highly expanded N0.HOG0005443, which is most likely a family of Histone-lysine N-methyltransferases (AT5G47150). One duplicate from *Aly* is up-regulated and shows signatures of diversifying selection. Both duplicates from *Esa* in this gene family are also up-regulated. (2) The above-mentioned Lysine-specific demethylase REF6, which functions in “positive regulation of lateral root development”. N0.HOG0008832 is a family of E3 UFM1-protein ligase 1 homologs, which regulates protein homeostasis (Supplementary Fig. 9,

Table 1 Number of genes for which diversifying selection was predicted by absrel [38]

Species	DEGs	Background	P_{lower}
<i>Aly</i>	544 (8%)	1423 (10%)	1.6e-15
<i>Ath</i>	133 (6%)	1165 (8%)	1.1e-05
<i>Esa</i>	489 (8%)	1371 (10%)	2.6e-06
<i>Bna</i>	596 (8%)	2732 (10%)	3.5e-14

P_{lower} = p-value from hypergeometric test for under representation in the set of differentially expressed genes, compared to the background of all genes from the given species in the Conserved Set

Table 2 Number of genes for which diversifying selection was predicted by absrel [38] in expanded HOGs

Species	Gene family expansion in	Genes under diversifying selection	Genes under diversifying selection in background	P_{upper}
<i>Aly</i>	<i>Aly</i>	326 (25%)	1423 (10%)	5.9e-65
<i>Aly</i>	<i>AlyandEsa</i>	41 (24%)	1423 (10%)	3.2e-08
<i>Esa</i>	<i>Esa</i>	209 (21%)	1371 (10%)	1.4e-30
<i>Esa</i>	<i>EsaandAly</i>	33 (20%)	1371 (10%)	3.8e-05
<i>Ath</i>	<i>Ath</i>	97 (20%)	1165 (8%)	3.7e-17
<i>Ath</i>	<i>AthandBna</i>	17 (25%)	1165 (8%)	2.1e-05
<i>Bna</i>	<i>Bna</i>	565 (10.4%)	2732 (9.8%)	0.04
<i>Bna</i>	<i>BnaandAth</i>	21 (15%)	2732 (10%)	0.04

For a given species, the number of genes showing diversifying selection is given in HOGs expanded in the given species (upper row, normal letters) or in HOGs expanded in both species with the same tolerance phenotype (lower row, bold letters), but not expanded in species of the opposite tolerance phenotype. P_{upper} = p-value from hypergeometric test for over representation in the set of expanded HOGs, compared to the background of all genes from the given species in the Conserved Set

Table 3 Number of differentially expressed genes for which diversifying selection was predicted by absrel [38] in HOGs which are expanded in the given species (but not in any species with opposite tolerance phenotype)

Species	Gene family expansion in	DEGs under diversifying selection	DEGs under diversifying selection in background	P_{upper}
<i>Aly</i>	<i>Aly</i>	70 (17%)	544 (8%)	1.3e-10
<i>Ath</i>	<i>Ath</i>	6 (10%)	133 (6%)	0.16
<i>Esa</i>	<i>Esa</i>	41 (14%)	489 (8%)	5.0e-4
<i>Bna</i>	<i>Bna</i>	109 (8%)	596 (8%)	0.29
<i>Aly</i> or <i>Esa</i>	<i>Aly</i> and <i>Esa</i>	12 (13%)	1033 (8%)	0.06
<i>Ath</i> or <i>Bna</i>	<i>Ath</i> and <i>Bna</i>	4 (8%)	729 (7%)	0.48

P_{upper} = p-value from hypergeometric test for over representation compared to the background of all DEGs from the given species under diversifying selection in the Conserved Set

Additional File 1). One of the up-regulated duplicates from *Esa* is under diversifying selection. It functions in “response to endoplasmic reticulum stress”, “protein K69-linked ufmylation” and “regulation of proteasomal ubiquitin-dependent protein catabolic process”.

Diversifying selection is enriched in candidate gene families for drought adaptation

We tested the genes within the conserved HOGs for signatures of diversifying selection in their coding sequences by performing an exploratory analysis using the adaptive branch site model absrel [38]. 72% of the conserved genes from *Esa* (14,412 / 19,990), 71% from *Aly* (14,974 / 21,173), 72% from *Ath* (14,408 / 19,919) and 72% from *Bna* (27,962 / 38,861) passed the quality filter criteria (see Methods) and were tested (*OGs_pos_selection_pvalues.tsv*, [31]). From these, 1,371 (9.5%) genes from *Esa*, 1,423 (9.5%) genes from *Aly*, 1,165 (8.1%) genes from *Ath*, and 2,732 (9.8%) genes from *Bna* showed signatures of diversifying selection on significance level 0.05. We used the signatures of diversifying selection as a marker for an increased probability of a gene to be relevant for adaptation. To use this marker to evaluate the prioritization of candidate genes for drought adaptation, we tested if the sets of (1) DEGs (Table 1), (2) expanded HOGs (Table 2) and (3) differentially expressed and expanded HOGs (Table 3), including the candidate genes for drought adaptation, were enriched or depleted for genes with signatures of diversifying selection using a one-sided hypergeometric test at significance level 0.05. We found that DEGs from all four species are significantly less often under diversifying selection than all genes from the respective species in the Conserved Set (Table 1). On the other hand, we expected that gene duplicates are frequently under diversifying selection because neofunctionalization including positive selection is an important model for the fixation of a gene duplicate in a population. We confirmed this hypothesis: duplicated genes of each of the four species, which are not duplicated in the species with contrasting drought tolerance, were significantly more often under diversifying selection compared to all genes of the respective species in the Conserved Set (Table 2). The enrichment was smaller in duplicates from *Bna* compared to duplicates from the other species.

Genes duplicated in both species with the same tolerance phenotype, which are not duplicated in the species with opposite drought tolerance, were also enriched for signatures of diversifying selection, as were expansions common to phylogenetically closely related species (Supplementary Table 7, Additional File 1). However, expansions in *Bna* or expansions common to *Bna* and *Esa* were not enriched for signatures of diversifying selection.

In order to evaluate the quality of the prioritized candidate genes for drought adaptation we used the enrichment of genes with signatures of diversifying selection as a marker for the relevance of the gene for adaptation (Table 3). In the background of all DEGs from any of the tolerant species, 8% were under diversifying selection, whereas 12 of 92 tested candidate gene families (13%) had a DEG from any of the tolerant species under diversifying

selection. However, this enrichment had a low level of significance ($p=0.06$). We did not see enrichment in the contrasting set, which were DEGs from any of the sensitive species in HOGs expanded in both sensitive species, but not in the tolerant species (8% vs. 7%, $p=0.48$). As the number of genes under diversifying selection was very low in the candidate and contrasting gene sets, we also analyzed the enrichment in HOGs expanded in only one species, but not expanded in the species with opposite drought-tolerance phenotype. When looking at DEGs in species-specific expansions, we found that DEGs in HOGs which are expanded in either of the tolerant species, but not in any sensitive species, were highly significantly enriched for DEGs with signatures of diversifying selection (Table 3). This was most prominent for expansions in *Aly*, where 17% of the DEGs were under diversifying selection, whereas only 8% of all DEGs from *Aly* were under diversifying selection ($p=1.26e-10$). We did not find significant enrichment in DEGs in the sensitive species expanded HOGs.

Candidate gene families with differentially expressed genes under diversifying selection

There were 12 candidate HOGs which had a DEG from a tolerant species under diversifying selection (Table 3): 7 of these DEGs were from *Esa* and 5 from *Aly*, while there was no HOG with a DEG from both tolerant species under diversifying selection. Within the focus set of HOGs expanded and specifically up-regulated in both tolerant species, we found N0.HOG0008832 (E3 UFM1-protein ligase 1, Supplementary Fig. 9, Additional File 1), N0.HOG0007350 (Lysine-specific demethylase REF6, Supplementary Fig. 13, Additional File 1) and N0.HOG0005443, which is most likely a family of Histone-lysine N-methyltransferases. More example HOGs have been described above or in Supplementary Results, Additional File 1. Detailed information on the candidate HOGs with a DEG under diversifying selection can be found in Additional File 2_Candidate_gene_families_and_tolerant_specific_expansions.xlsx. We also described the functions and examples of DEGs which show signatures of diversifying selection in HOGs expanded in either *Esa* or *Aly* in the Supplementary Results, Additional File 1.

Discussion

We hypothesized that gene family expansion unique to drought-tolerant species is a criterion to select from differentially expressed genes (DEGs) those which underlie drought adaptation of the tested species. Gene family expansion is an important driver for the evolution of new functions, and retention of duplicates indicates a

selective advantage. Correlation with a tolerance trait can be indicative of successful adaptation.

Using our A2TEA workflow and webapp [23] to combine trait-specific gene family expansion with differential expression (DE) we selected a more concise set of candidate genes. In order to demonstrate that this is a meaningful set we analyzed features that are relevant for drought adaptation and showed that the candidate gene set is enriched for (1) DE under drought conserved in both tolerant species, (2) DE unique to the tolerant species, and (3) up-regulation. Specific functions are enriched, and the set contains genes with functions in line with drought adaptation based on evidence from other species. We showed that signatures of diversifying selection are slightly enriched in the candidate gene families (expanded in both tolerant and DE in any tolerant species) and highly significantly enriched in DEGs in tolerant species-specific expanded gene families. Strikingly, they are not enriched in differentially expressed gene families which are expanded in sensitive species.

The prioritization of candidate genes in this study was based on the classification of drought resistance of the four analyzed *Brassicaceae* species. Drought resistance involves many processes and phenotypic effects, which complicates its definition (reviewed by [39]). For plant production in areas with periods of drought, a high survivability with a minimum of growth reduction is desirable ([40], [41], [24], [42]). Based on the literature, we classified the extremophytes *Eutrema salsugineum* (*Esa*) and *Arabidopsis lyrata* (*Aly*) as more drought resistant than their relatives *Arabidopsis thaliana* (*Ath*) and *Brassica napus* (*Bna*) (see Background). A strength of this set was that gene family expansion, one important pillar on which we based our prioritization of candidate genes for drought adaptation, is detached from the phylogenetic relationship of the analyzed species, as we have used one tolerant and one sensitive species from the *Brassicaceae* lineages I and II, each ([43]). Thereby, our study most likely identified old gene family expansions which are derived from whole genome duplication (WGD) in the common ancestor of the *Brassicaceae* ([44], [45], [46]) but were kept only in the drought-tolerant species. This indicates their selective advantage under drought rather than phylogenetic conservation of lineage-specific expansions not related to the trait, such as the 449 lineage-specific expansions in *Esa*, *Aly* and *Ath*. Similarly, the 441 lineage-specific expansions in both *Arabidopsis* species and the 103 lineage-specific expansions in *Esa* and *Bna* are phylogenetically conserved, but we cannot relate their conservation to drought adaptation.

We argue that retention of duplicates in both drought-resistant species but not sensitive species, even if phylogenetically closer, is indicative of adaptation. Alternatively, two independent duplications in both tolerant

species could have led to adaptation through convergent evolution.

Gene family expansions in DE gene families

Differential expression under drought allows identifying genes which are functionally linked to drought response but not necessarily the result of adaptation. Based on the prediction of orthologous groups (HOGs), we prioritized from the thousands of HOGs which were DE under drought hundreds of HOGs which were expanded in any of the drought-resistant Brassicaceae species. From these we selected 93 candidate HOGs which were expanded in both tolerant species and DE in at least one tolerant species.

This prioritization of candidate genes for drought adaptation relied on the prediction of HOGs by OrthoFinder [30]. Additionally, we used an automated tool, the A2TEA.WebApp [23], to manually evaluate the identified candidate HOGs. This quality control included the visualization of the MSA with the closest related HOGs and the base mean expression per gene in the HOG. We observed several cases where the MSA of gene duplicates was fragmented. This could be due to pseudogenization, or sequencing, assembly, or annotation errors. Nevertheless, we also expect such cases in the background and did not exclude them from enrichment analysis. Furthermore, when manually validating the candidate gene families, we found several HOGs where OrthoFinder predicted an expansion only in the tolerant species, but likely failed to include additional gene copies in *Bna*: when adding to the phylogenetic tree the most closely related HOGs, we observed that genes from a HOG composed of exclusively *Bna* genes are probably members of the HOG in question. This means that the predicted expansion in the tolerant species is not exclusive, but also found in a sensitive species. In the complete dataset, there is an unexpectedly large number of HOGs specific for *Bna*, so this may be a frequent prediction error.

We also found discrepancies between the HOGs predicted in our study to the gene families predicted in other studies. For example, Ali and colleagues [14] found gene family expansion in *HKT1* in *Esa* but not in *Ath*. Laha and colleagues [21] found even three homologs to the single Arabidopsis *HKT1* gene At4G10310 in *Esa*. We did not find this gene duplication in the candidate HOGs, but instead, we identified *HKT1* from *Ath* in N0.HOG0002756 together with a single homolog from *Aly*, which is up-regulated under drought. The three homologs from *Esa* were found separately, each in a HOG with two *Bna* homeologs (N0.HOG0002757: EUTSA_v10028767mg, N0.HOG0002758: EUTSA_v10028594mg and N0.HOG0002759: EUTSA_v10028595mg, refer to A2TEA_finished.Rdata, [31]). So *HKT1* is expanded both in *Esa* and in *Bna*. Thus, as the triplication is also

found in the drought-sensitive species *Bna*, in our study these are not candidate genes for drought adaptation. We speculate that the neofunctionalization of *HKT1* under increased salt concentrations is a special adaptation of *Esa* to salt stress.

In contrast to Das and colleagues [20], who found a duplication in *Aly* in *NICOTINAMINE SYNTHASE3* genes which was not present in *Ath*, we found that the duplication in *NICOTINAMINE SYNTHASE3* genes was present in all four analyzed Brassicaceae species and that the duplicates showed divergent regulation under drought in the species in our study (N0.HOG0003383 and N0.HOG0003384, refer to A2TEA_finished.Rdata, [31]). Hence, we did not identify this HOGs as a candidate HOG for drought adaptation according to our criteria. Nevertheless, the duplicates might be neofunctionalized in all these species.

The definition of orthologous groups is crucial for the reliable prioritization of candidate gene families with our criteria. To achieve the best possible reliability, we have compared tolerant and sensitive species of the same family and have used the most recent version of OrthoFinder. Our manual inspection allowed us to identify some prediction errors, and the A2TEA.WebApp offered a useful visualization for this. However, even though in some cases we came to different conclusions than other studies, we see no indication that the reliability of the identified candidate HOGs is insufficient for statistical evaluation.

We characterized the identified candidate genes for drought adaptation regarding their patterns of regulation and their biological functions.

Differential expression was more conserved in the candidate gene set. Among the set of 93 candidate HOGs we found differential expression under drought in both species more often than expected: 47 (50%) showed differential expression in both species, whereas only 6 (15%) of the contrasting set (expanded in both sensitive species with differential expression in at least one sensitive species) showed differential expression in both. Also with respect to the background set, where 4,638 (40%) were differentially expressed in both tolerant species, in tolerant species conserved differential expression was significantly enriched in the candidate set: 47 (50%) showed DE in both tolerant species ($p=0.02$). We did not find enrichment in the contrasting set. We interpret this conservation to indicate common drought adaptation, which resulted in conserved gene family expansion and regulation under drought in tolerant species, but not sensitive species. This underpins the prioritization as candidate gene families for drought adaptation. Interestingly, in the candidate HOGs where both species showed differential expression, the direction of differential expression was more often only up-regulation and less often

only down-regulation (Fig. 3a). This was different from the background of all HOGs with conserved differential expression in both tolerant species, where up-regulation was equally frequent as down-regulation. We infer from this that up-regulated genes are more likely to have been targets for adaptation to drought. As a result, these duplicates have been retained specifically in the drought-tolerant species, along with their expression pattern.

A different perspective regards differential expression unique to tolerant species, where genes from sensitive species are not differentially expressed. 35% of HOGs from the Conserved Set which were regulated in both tolerant species were not regulated in the sensitive species, i.e., a majority of HOGs shows differential expression also in a sensitive species. This is in line with the observation that differential expression is mostly conserved between species. In contrast, in the candidate HOGs the proportion showing unique differential expression, i.e., only in the tolerant species but not in any sensitive species, was significantly increased to 59% ($p=4.6e-4$, Fig. 4). This could point to emergence of differential expression under drought as a result of adaptation, which would be cases of neofunctionalization [10], or to the loss of differential expression in the sensitive species.

To summarize, the candidate genes for drought adaptation identified by our approach, which additionally show DE in both tolerant species, were more often up- than down-regulated, and mostly uniquely differentially expressed in the tolerant species but not the sensitive species. These observations are consistent with the hypothesis that our approach selects genes that are adaptive for drought.

With respect to gene function, we observed that the selected candidate genes do not show enrichment of functions directly relevant to drought, e.g., general stress response genes or known drought response functions. However, focusing on genes uniquely up-regulated in both tolerant species enriches for genes which function in processes such as “plant-type secondary cell wall biogenesis”, “trichome morphogenesis”, “sugar mediated signaling” and “positive regulation of lateral root development”, functions which could reasonably play a role in drought adaptation. It needs to be noted that this excludes genes differentially expressed upon drought in *Ath*. As many functional annotations are deduced from experiments conducted in *Ath*, the genes annotated with drought response functions are mostly differentially expressed upon drought in *Ath*. It is not surprising that the 17 gene families of interest are not annotated with general drought response functions as they are not regulated upon drought in *Ath*. Instead, we conclude that their regulation upon drought was acquired through adaptation in the drought-resistant species. This hypothesis is supported by genes from this set that function in

processes known to impact drought tolerance such as root development [40] or cell wall organization ([47], [48]) and that were differentially regulated upon drought only in tolerant species.

In contrast to candidate genes uniquely up-regulated in both tolerant species, the background of all DEGs from the tolerant species was enriched for many stress response processes such as “response to water deprivation”, “response to salt stress”, “response to cold”. These results confirmed that drought response genes are often general stress response genes (see Background). Moreover, the high enrichment of genes which function in “photoinhibition”, “photosynthesis” and “photosynthesis, light harvesting in photosystem I” underpins that drought strongly affects photosynthesis, which can have negative effects on crop yield ([5], [6], [7]). Apparently, by selecting from the DEGs candidate gene families which are expanded and up-regulated only in the tolerant species, we differentiate processes which are relevant for drought adaptation from the many general stress responses in all DEGs. We conclude that our approach identified a concise and functionally relevant set of candidate genes for drought adaptation and allowed identifying novel functions not linked to drought response in *Ath*.

We evaluated the prioritization of candidate genes for drought adaptation by identifying enrichment with signatures of diversifying selection to underpin a gene’s relevance for adaptation. While the enrichment of diversifying selection in the candidate gene families was not significant ($p=0.06$), we could still show that selecting gene family expansion unique to drought-tolerant species from DEGs under drought links the selected DEGs to drought adaptation because of the highly significant enrichment of diversifying selection in the DEGs in tolerant species-specific expanded gene families, while there was no enrichment in the DEGs in sensitive species-specific expanded gene families. The proportion of genes which were tested for diversifying selection was 71–72% for each of the four species, which is why we did not expect biases between the species in the sets of tested genes.

We showed that differential expression (DE) under drought was overall not linked to diversifying selection in any of the species. This is in accordance to our observation that DE under drought was conserved between the four species and as such we expected that DEGs are rather under purifying selection. Gene families specifically expanded in both drought-tolerant or drought-sensitive species were enriched for diversifying selection. Thus, these genes, while they are probably a result of adaptation, are not specifically adapted to drought. By considering DE under drought, we could link them to an adaptive function under drought. Even though we found a strong correlation of DE and gene family expansion in

all four species, diversifying selection was only enriched in gene families which combined these criteria in the tolerant, but not in the sensitive species. This indicates that expansions specific to tolerant species are adaptive for the tolerance trait, whereas expansions specific to sensitive species are adaptive for other traits. So while diversifying selection was not generally linked to DE, it was linked to genes DE under drought that were specifically expanded in drought-tolerant plants. We propose that in these genes, the signatures of diversifying selection specifically indicate adaptation to drought. This supports our hypothesis that combining expansion in tolerant species with differential expression under drought selects candidate genes related to adaptation to drought.

There were 14% and 17% of DE gene families expanded in *Esa* or in *Aly*, respectively, in which a DEG from the respective drought-tolerant species showed signatures of diversifying selection. This also means that there were many expanded gene families in which the DE duplicates have not experienced diversifying selection. These gene duplications were most likely subfunctionalized and fixed by genetic drift according to the DDC model ([49], reviewed in [50] and [10]) and are as relevant for adaptation as the ones which have experienced diversifying selection.

Discussion of example candidate genes

Among all 93 candidate genes, we identified several processes which are implicated in drought tolerance. We propose that the candidate genes which function in processes whose relevance has not yet been implicated in drought tolerance are likely to be relevant, too. We describe some interesting cases here. We identified the gene family of *CER9* (N0.HOG0006674) among our candidate genes, where one of two *CER9* homologs is up-regulated in *Aly*. *CER9* encodes a E3 ubiquitin-ligase with a negative regulatory role in the early cuticle wax and lipid biosynthesis [51]. Further, a mutant which is deficient in *cer9* in *Ath* wilts later [51]. Hence, the up-regulation of the gene duplicate in *Aly* might be related to the earlier wilting which Bouzid and colleagues [24] observed in *Aly*. Based on these two findings we speculate that an over-expression in *Ath* might lead to earlier wilting. The gene is also duplicated but not significantly regulated in *Esa*. This could be due to the generally later reaction to drought of *Esa* compared to fast-responding *Aly* [25]. *Aly* has mechanisms to tolerate earlier wilting, for example keep the photosynthetic rate high under wilting, and can better survive wilting than *Ath* ([25], [24]). Possibly, even if the earlier wilting is a negative trait, it was retained because the up-regulated duplicate regulates new targets and influences a positive mechanism for drought tolerance. Actually, the deficiency in *cer9* results in an up-regulation of 21 genes which function in protein

ubiquitination and degradation [51]. These are potential targets for a diverged regulation by the drought-specific up-regulated gene duplicate in *Aly*. The *cer9* mutants also showed an increase in very long-chain fatty acids (VLCFAs), and the effects of *cer9* were mostly additive to mutations in *Long chain acyl-CoA synthetase 1* (*LACS1*, *CER8*) [51], which is a second important component of the early wax and cutin biosynthesis pathway. Mutations in *LACS1* more than doubled the amount of VLCFAs in *Ath* [52]. We found *LACS1* (N0.HOG0008907) among our candidate genes, where one of the homologs from *Esa* and from *Bna*, each, are down-regulated. Even though the increase in VLCFAs has not been shown to be causal for an increased barrier for cuticular water loss, structural differences in cuticular wax composition influence the transpiration ([51], [53]). It is yet unclear, how the duplication of the here described wax biosynthesis genes underlie drought adaptation, but the conservation of the duplication in the tolerant species and their regulation under drought indicate their relevance for the drought tolerance.

Among our candidate gene families which were up-regulated uniquely in the tolerant species was a family coding for a Kinesin-like protein (*KIN-13 A*, N0.HOG0008546, AT3G16630) which functions in the morphogenesis of trichomes [36] and in the formation of secondary cell wall pits [37]. In a hot and dry adapted *Shepherdia* hybrid, trichome density was increased under drought and positively influenced light reflectance and leaf temperature regulation under heat [54]. While *Aly* populations with trichomes were more resistant to herbivory compared to populations without trichomes [55], trichome production did not increase tolerance to drought in *Aly* [56]. But the function of *KIN-13 A* in the development of metaxylem could explain the adaptive role of its duplication in both tolerant species: the surface of secondary cell wall pits, areas where no secondary cell wall is deposited, increased when over expressing *KIN-13 A* in *Ath* [37], which could influence the xylem sap transport. We suggest that the duplication and up-regulation of *KIN-13 A* in *Aly* and *Esa* is adaptive for drought by altering the secondary cell wall patterns in developing xylem tissues in response to drought, which could potentially increase the xylem sap transport. The formation of secondary cell wall pits depends on the depolymerization of the cortical microtubules by the interaction of *KIN-13 A* with *ROP11* and *MIDD1* [37]. Interestingly, both *KIN-13 A*, one of two *ROP11* (N0.HOG0005842) homologs, and the single *MIDD1* (N0.HOG0021372) were up-regulated in *Esa* under drought (refer to *A2TEA_finished.Rdata*, [31]). While *KIN-13 A* was duplicated uniquely in the two tolerant species, *ROP11* was duplicated in all four species and *MIDD1* was not duplicated. It is less clear, how the pathway is regulated in *Aly*, where both

KIN-13 A homologs were, similar to *Esa*, up-regulated, while both *ROP11* (N0.HOG0005842) homologs were down-regulated under drought. In our study (genome assembly GCA_000004255.1), we did not find *MIDD1* in *Aly*, but an OrthoDB search and a NCBI blast search (GCF_000004255.2) for orthologs to the *MIDD1* homolog from *Ath* (AT3G53350) identified a single homolog (110228692) in *Aly* of which we do not have information regarding its expression under drought. These findings suggest that the up-regulation of the *KIN-13 A-ROP11-MIDD1* pathway is adaptive for drought at least in *Esa*.

Another candidate gene regulates flowering in *Ath*: The Lysin-specific demethylase *REF6* (AT3G48430, N0.HOG0007350) was duplicated and up-regulated under drought in both tolerant species but in none of in the sensitive species. Both duplicates from *Esa* showed signatures of diversifying selection. Over-expression of *REF6* in *Ath* lead to early flowering and altered leaf development [57]. It also promoted lateral root formation in *Ath* by demethylation of the repressive H3K27me3 mark on the Chromatin of *PIN1/3/7* [34]. We propose that the over-expression of *REF6*, as we observed it in the up-regulation of two homologs in each of the tolerant species, can improve drought tolerance, also in the sensitive species, likely by increasing the root biomass or altering the regulation of flowering.

Among the candidate genes, transcription factor II D subunit 9 (*TAF9*, N0.HOG0007005) was uniquely up-regulated in both tolerant species. We speculate that the up-regulation of both duplicates in the tolerant *Aly* underlies the generally fast [25] and strong reaction of *Aly* to drought, whereas in contrast in *Esa* only one of the gene duplicates was slightly up-regulated while the other was not regulated under drought. We found that an additional TFIID subunit (*TAF14b*, N0.HOG0011415) was expanded and up-regulated in *Aly*, while the second homolog was not expressed in this experiment and lacked large parts of N- and C-terminal domains. As loss and gain of domains is common in the evolution of diverged TAFs [58], the diverged gene copies in our study likely have specialized functions. *TAF14* is a regulator of *FLC* [59], and we speculate that its duplication can affect the regulation of flowering under drought.

The described cases require further research to prove their relevance in drought adaptation. We report them here to highlight that further investigations about the unknown candidates will likely be rewarding.

Conclusions

We identified a concise set of candidate genes for drought adaptation in two drought-tolerant Brassicaceae species using comparative genomics and transcriptomics. DEGs which were expanded in drought-tolerant, but not in sensitive Brassicaceae species were related to drought

adaptation more than the background of all genes which reacted to drought when regarding conserved DE, DE specific to tolerant species, or up-regulation. Based on functional annotation we found candidate genes involved in processes which are relevant for drought adaptation, while the background set of DEGs was enriched for general stress reactions. As an independent evaluation of the relevance of the identified candidate genes we found that only DEGs which were duplicated uniquely in a tolerant, but not DEGs duplicated uniquely in a sensitive species, were enriched for signatures of diversifying selection. Our method allowed identifying both conserved adaptations and species-specific adaptations in only one tolerant species, see Supplement.

As our approach is exploratory, not based on pre-defined groups of genes, the identification of candidate genes which have not yet been implicated in drought tolerance presents novel but probable hypotheses for further analyses. In vivo experiments using mutations or genetic engineering of candidate genes, e.g., following a strategy outlined in [21], will be necessary to demonstrate a role in drought tolerance and give more insights into which molecular mechanisms underlie the drought tolerance traits of *Esa* and *Aly*.

Methods

Data

Genomic data and proteome files from *Eutrema salsugineum* [60], *Arabidopsis lyrata* [61], *Arabidopsis thaliana* [62] and *Brassica napus* [28] were downloaded from Ensembl genomes [63] as listed in Supplementary Table 1, Additional File 1, on June 07, 2022. RNA-Seq Data from drought stress studies in *Eutrema salsugineum*, *Arabidopsis lyrata* and *Arabidopsis thaliana* [25] and *Brassica napus* [32] were downloaded from the Sequence Read Archive (SRA, [64]) as listed in Supplementary Table 2, Additional File 1, on March 16, 2022. The proteome files were filtered for a minimum sequence length of 100 amino acids using the scripts *filter_min100aa.sh*, Additional File 3 followed by *normalize.sh*, Additional File 3.

Candidate gene prioritization with the A2TEA.Workflow

We calculated the differential expression, the orthologous group prediction and the gene family expansion using our recently published automated snakemake workflow *A2TEA.Workflow* [23] commit e6d4aad58e-0a685e4e8eec6fc0b7a99521730ea4. Briefly, *A2TEA.Workflow* uses OrthoFinder ([30], version 2.5.4) to predict orthologous groups (HOGs) from proteome files of the species of interest, and selects expanded gene families from the predicted HOGs, where the expansion criteria are defined by the user. Expansions in a HOG are always defined by the number of genes in a species of interest

compared to background species, taking into account polyploidy. In the present study, expansions of interest were those, where the species of interest has twice the number of genes compared to the background species in a HOG. In our case, the background species are a group of species with contrasting drought tolerance phenotype as the species of interest. As *Bna* is tetraploid, the gene count in *Bna* is halved. This means that e.g., a HOG which is expanded in *Esa* has at least twice the number of genes from *Esa* than from *Ath* and at least twice the number of genes from *Esa* than half of the number of genes from *Bna*. Additionally, *A2TEA.Workflow* calculates differential gene expression using DESeq2 [65] from RNA-Seq reads of the same species and combines the information of differential gene expression with the gene family expansion (Supplementary Fig. 1a, Additional File 1).

Raw genomic files (cdna and annotation.gff3) as listed in Supplementary Table 1, Additional File 1, and the filtered proteome files were used as input to the *A2TEA.Workflow*. Functional annotation files from a previous *A2TEA.Workflow* run, which was based on the same input data, were used and depicted in the column “function” in *config/species.tsv*. The functional annotation had been produced by AHRD [33] and the file had the first two lines (# and empty lines) deleted, as stated in the *A2TEA.Workflow README.md* file. Parameters in the *config.yaml* were adjusted to use only the longest isoform when there were several isoforms of one protein (*auto-isoform-filtering*=YES). We used kallisto [66] for pseudoalignment of the RNA-Seq reads, and the expression was quantified on gene level (*transcript_level_quantification*=NO). For the differential expression, a false discovery rate (FDR) of 0.1 was defined to identify differentially expressed genes between drought and control conditions (*DEG_FDR*=0.1). During the workflow, phylogenetic trees of expanded gene families were calculated with additionally up to five closest related HOGs (*add_OGs*=5) for subsequent visual inspection using the *A2TEA.WebApp* v1.1.5. We used the *A2TEA.Workflow* to identify gene families which were expanded in both of the tolerant species but not in any of the sensitive species (Hypothesis 1), or in each of the tolerant species but not in any of the sensitive species (Hypotheses 2 and 3), and vice versa (Hypotheses 4–6 in the *config/hypotheses.tsv*). As *Bna* is a young (~7500 y.a.) allopolyploid with the A and C *Brassica* genome (2n=4x=38, [28]), we account for the doubled number of genes in *Bna* when calculating gene family expansions by setting the column “ploidy” in the *config/species.tsv* to “4”.

The results of the *A2TEA.Workflow*, including differential expression, information about gene family expansion on all HOGs, as well as MSA and phylogenetic trees of gene families of each of the tested hypotheses and the

functional annotations predicted by AHRD [33] are in *A2TEA_finished.Rdata*, [31].

Analyses

We characterized the selected candidate genes for drought adaptation regarding their functions, differential expression upon drought, calculated their phylogeny and predicted signatures of diversifying selection (Supplementary Fig. 1b, Additional File 1). The analyses on the differential expression of the gene families and plotting were performed using the R programming language (vers. 4.4, [67]) with the Analysis scripts *Analysis_A2TEA_output.R*, Additional File 3 and *Analysis_expression_pattern_per_OG_only_at_least_2_sig.R*, Additional File 3).

Conservation of DE

To test if the DE was more conserved than expected from the distribution of DEGs from each of the species in the Conserved Set, we calculated the probability of success ($p_s = s/n$) and failure ($p_f = 1 - p_s$) of a HOG to contain at least one DEG per species from the number of HOGs with at least one DEG from species (s) in the background of all conserved HOGs ($n=17488$). We then calculated the probability of a HOG to have a DEG from all 4 species as $p_{s4} = p_{sAt} * p_{sAl} * p_{sEs} * p_{sBn}$ and the probability of a HOG not to contain a DEG from any of the four species as $p_{f4} = p_{fAt} * p_{fAl} * p_{fEs} * p_{fBn}$. We compared the probabilities to the observed number of HOGs with a DEG from all four species or from none of the four species. Subsequently, we computed the chi-square test using the function *chisq.test()* for each of the two hypotheses on a level of significance $\alpha=0.05$ (*Analysis_A2TEA_output.R*, Additional File 3).

GO term over representation analyses

The GO term enrichment analyses were calculated using the integrated TopGO analysis in the *A2TEA.WebApp*, always using the Conserved Set as a background distribution. The GO annotations were the union of the GO annotations of each gene in the orthologous group, hence annotations of all four species were considered. GO terms were only considered over-represented, when the one-sided fisher’s test p-value was below 0.05 and the GO term was annotated at least twice in the set of interest.

Revigo maps of gene functions

Gene functions were summarized and visualized in maps using the Revigo webtool [68] with default parameters.

Phylogenetic trees of candidate gene families

The candidate gene families were evaluated by plotting the phylogeny including up to 5 closest HOGs and the underlying multiple sequence alignment (MSA) of the peptide sequences with the *A2TEA.Webapp* [23].

Diversifying selection

Coding sequences of the four species were downloaded from Ensembl genomes [63] as listed in Supplementary Table 1, Additional File 1. We applied a snakemake ([69], version 7.0.0) workflow to test the gene families predicted by OrthoFinder (HOGs) for signatures of diversifying selection using the exploratory mode in *aBSREL* [38], an adaptive branch-site model. The workflow is publicly available in a github repository [70]. It includes the following steps: (1) We supplied the coding sequences of the proteins from all four species and a matrix which contained the gene families with the corresponding protein names. (2) The data were pre-processed to obtain a codon aware multiple sequence alignment per HOG for the *aBSREL* test for diversifying selection. First, all proteins whose coding sequence contained “N” were removed from the workflow. Then, the amino acid sequences were translated from the coding sequences using the hyphy [71] *pre-msa.bf* script (<https://github.com/veg/hyphy-analyses/tree/master/codon-msa>) using the --E 0.05 option to allow alignment of sequences with low homology. (3) The amino acid sequences of each HOG were then aligned to each other with muscle ([72] version 5.1) and the nucleotide multiple sequence alignment (MSA) was obtained from the peptide alignment with the hyphy *post-msa.bf* script. (4) The phylogenetic trees were calculated with fasttree ([73], version 2.1) only for the HOGs which did not contain “?” in the nucleotide MSA. (5) The test for diversifying selection was then applied to all remaining HOGs. We used a false discovery rate of 0.05 to define signatures of diversifying selection in a gene. Hence, genes with a corrected p-value ≤ 0.05 have most likely experienced diversifying selection. The results of the test for diversifying selection were analyzed with the script *Analysis_A2TEA_output_with_Selection_v04_with_corrected_pvalues.R*, Additional File 3. The over-representation analyses of the signatures of diversifying selection within the subsets of candidate gene families were calculated with the script *Selection_over-representation_v04_with_corrected_pvalues.Rmd*, Additional File 3.

Abbreviations

AA	Amino acid
ABA	Abscisic acid
Aly	Arabidopsis lyrata
Ath	Arabidopsis thaliana
Bna	Brassica napus
BP	Biological process
DE	Differential expression
DEG	Differentially expressed gene
Esa	Eutrema salsugineum
FDR	False discovery rate
GO	Gene ontology
HOG	Hierarchical Orthologous Group
HRD	Human-readable description
MSA	Multiple sequence alignment
VLCFA	Very long-chain fatty acid

WGD Whole genome duplication

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11737-0>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements

We thank Tyll Stöcker for advice and help with the A2TEA.Workflow.

Author contributions

CUF and HS designed the study. CUF performed all analyses. CUF and HS interpreted the results of the analyses. CUF and HS wrote the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. Carolin Uebermuth-Feldhaus was funded by a PhD scholarship from Evangelisches Studienwerk e.V. Villigst.

Data availability

No datasets were generated in this study. The analyses were performed on publicly available genomic (10.3389/fpls.2013.00046, 10.1038/ng.807, 10.1111/tpj.13415, 10.1126/science.1253435) and transcriptomic (10.1111/nph.15841, 10.1093/jxb/erac253) data (details see Methods). The raw results from the A2TEA.Workflow (A2TEA_finished.Rdata), a table of all conserved gene families and the respective genes of the four species (Conserved_dataset_matrix.tsv), and a table with the p-values of the test for diversifying selection for all tested genes (OGs_pos_selection_pvalues.tsv) are available in the bonndata repository 10.60507/FK2/JBUJFW under the CC-BY-4.0 license. All analysis scripts are in Additional_File_3_Scripts.zip. The code for the workflow for predicting signatures of diversifying selection is available in the github repository 10.5281/zenodo.14540010 under the Apache-2.0 license.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 30 January 2025 / Accepted: 21 May 2025

Published online: 19 June 2025

References

- Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J*. 2007;50:347–63. <https://doi.org/10.1111/j.1365-3113x.2007.03052.x>.
- Cohen SP, Leach JE. Abiotic and biotic stresses induce a core transcriptome response in rice. *Sci Rep*. 2019;9:1–11. <https://doi.org/10.1038/s41598-019-42731-8>.
- Dossa K, Mmadi MA, Zhou R, Zhang T, Su R, Zhang Y, Wang L, You J, Zhang X. Depicting the core transcriptome modulating multiple abiotic stresses responses in Sesame (*Sesamum indicum* L). *Int J Mol Sci*. 2019;20:3930. <https://doi.org/10.3390/ijms20163930>.

4. Reddy AR, Chaitanya KV, Vivekanandan M. Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *J Plant Physiol*. 2004;161:1189–202. <https://doi.org/10.1016/j.jplph.2004.01.013>.
5. Lipiec J, Doussan C, Nosalewicz A, Kondracka K. Effect of drought and heat stresses on plant growth and yield: a review. *Int Agrophys*. 2013;27. <https://doi.org/10.2478/intag-2013-0017>.
6. Yuan G-F, Jia C-G, Li Z, Sun B, Zhang L-P, Liu N, Wang Q-M. Effect of brassinosteroids on drought resistance and abscisic acid concentration in tomato under water stress. *Sci Hort*. 2010;126:103–8. <https://doi.org/10.1016/j.scienta.2010.06.014>.
7. Sreenivasulu N, Harshavardhan VT, Govind G, Seiler C, Kohli A. Contrapuntal role of ABA: does it mediate stress tolerance or plant growth retardation under long-term drought stress? *Gene*. 2012;506:265–73. <https://doi.org/10.1016/j.gene.2012.06.076>.
8. Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol Plant*. 2018;11:414–28. <https://doi.org/10.1016/j.molp.2018.01.002>.
9. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol*. 2019;20. <https://doi.org/10.1186/s13059-019-1650-2>.
10. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet*. 2008;9:938–50. <https://doi.org/10.1038/nrg2482>.
11. Jiménez-Morales E, Aguilar-Hernández V, Aguilar-Henonin L, Guzmán P. Molecular basis for neofunctionalization of duplicated E3 ubiquitin ligases underlying adaptation to drought tolerance in *Arabidopsis thaliana*. *Plant J*. 2020;104:474–92. <https://doi.org/10.1111/tpj.14938>.
12. Hu C, Lin S-y, Chi W-t, Chang Y-y. Recent gene duplication and subfunctionalization produced a mitochondrial GrpE, the nucleotide exchange factor of the Hsp70 complex, specialized in thermotolerance to chronic heat stress in *Arabidopsis*. *Plant Physiol*. 2011;158:747–58. <https://doi.org/10.1104/pp.111.1.87674>.
13. Monihan SM, Ryu C-H, Magness CA, Schumaker KS. Linking duplication of a calcium sensor to salt tolerance in *Eutrema salsugineum*. *Plant Physiol*. 2019;179:1176–92.
14. Ali Z, Park HC, Ali A, Oh D-H, Aman R, Kropornicka A, Hong H, Choi W, Chung WS, Kim W-Y and others. TSHKT1; 2, a HKT1 homolog from the extremophile *Arabidopsis* relative *Thellungiella salsuginea*, shows K⁺ specificity in the presence of NaCl. *Plant Physiol*. 2012;158:1463–74.
15. Ma T, Wang J, Zhou G, Yue Z, Hu Q, Chen Y, Liu B, Qiu Q, Wang Z, Zhang J. Others. Genomic insights into salt adaptation in a desert Poplar. *Nat Commun*. 2013;4:2797. <https://doi.org/10.1038/ncomms3797>.
16. VanBuren R, Wai CM, Pardo J, Giarola V, Ambrosini S, Song X, Bartels D. Desiccation tolerance evolved through gene duplication and network rewiring in *Lindernia*. *Plant Cell*. 2018;30:2943–58.
17. Hayashi S, Reid DE, Lorenz MT, Stiller J, Edwards D, Gresshoff PM, Ferguson BJ. Transient Nod factor-dependent gene expression in the nodulation-competent zone of soybean (*Glycine max* [L.] Merr.) roots. *Plant Biotechnol J*. 2012;10:995–1010.
18. Wu H-J, Zhang Z, Wang J-Y, Oh D-H, Dassanayake M, Liu B, Huang Q, Sun H-X, Xia R, Wu Y and others. Insights into salt tolerance from the genome of *Thellungiella salsuginea*. *Proceedings of the National Academy of Sciences*. 2012;109:12219–12224.
19. Wang W, Li X, Fan S, He Y, Wei M, Wang J, Yin Y, Liu Y. Combined genomic and transcriptomic analysis reveals the contribution of tandem duplication genes to low-temperature adaptation in perennial ryegrass. *Front Plant Sci*. 2023;14:1216048. <https://doi.org/10.3389/fpls.2023.1216048>.
20. Das M, Haberger G, Panda A, Das Laha S, Ghosh TC, Schäffner AR. Expression Pattern Similarities Support the Prediction of Orthologs Retaining Common Functions after Gene Duplication Events. *Plant Physiology*. 2016;171:2343–2357. Available from: <https://doi.org/10.1104/pp.15.01207> DOI: 10.1104/pp.15.01207.
21. Laha SD, 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>.
22. Oh D-H, Dassanayake M. Landscape of gene transposition–duplication within the Brassicaceae family. *DNA Res*. 2019;26:21–36.
23. Stöcker T, Uebermuth-Feldhaus C, Boecker F, Schoof H. A2TEA: identifying trait-specific evolutionary adaptations. *F1000Research*. 2022;11:1137. <https://doi.org/10.12688/f1000research.126463.1>.
24. Bouzid M, He F, Schmitz G, Häusler R, Weber A, Mettler-Altmann T, De Meaux J. *Arabidopsis* species deploy distinct strategies to cope with drought stress. *Ann Botany*. 2019;124:27–40. <https://doi.org/10.1093/aob/mcy237>.
25. Marin-de la Rosa N, Lin C-W, Kang YJ, Dhondt S, Gonzalez N, Inzé D, Falter-Braun P. Drought resistance is mediated by divergent strategies in closely related Brassicaceae. *New Phytol*. 2019;223:783–97. <https://doi.org/10.1111/nph.15841>.
26. Amtmann A. Learning from evolution: *Thellungiella* generates new knowledge on essential and critical components of abiotic stress tolerance in plants. *Mol Plant*. 2009;2:3–12. <https://doi.org/10.1093/mp/ssn094>.
27. Guevara DR, Champigny MJ, Tattersall A, Dedrick J, Wong CE, Li Y, Labbe A, Ping C-L, Wang Y, Nuin P. others. Transcriptomic and metabolomic analysis of Yukon *Thellungiella* plants grown in cabinets and their natural habitat show phenotypic plasticity. *BMC Plant Biol*. 2012;12:1–17. <https://doi.org/10.1186/1471-2229-12-175>.
28. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B and others. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*. 2014;345:950–3. <https://doi.org/10.1126/science.1253435>.
29. Wan J, Griffiths R, Ying J, McCourt P, Huang Y. Development of drought-tolerant *Canola* (*Brassica napus* L.) through genetic modulation of ABA-mediated stomatal responses. *Crop Sci*. 2009;49:1539–54.
30. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:1–14. <https://doi.org/10.1186/s13059-019-1832-y>.
31. Uebermuth-Feldhaus C. Supplementary Data to the study on Candidate gene prediction for drought adaptation in Brassicaceae species. 2025. <https://doi.org/10.60507/FK2/JBUJFW>.
32. Zhu X, Liu J, Sun X, Kuang C, Liu H, Zhang L, Zheng Q, Liu J, Li J, Wang H, Hua W. Stress-induced higher vein density in the C3–C4 intermediate *Morandia suffruticosa* under drought and heat stress. *Journal of Experimental Botany*. 2022;73:6334–6351. Available from: <https://doi.org/10.1093/jxb/erac253> DOI: 10.1093/jxb/erac253.
33. Hallab A, Klee K, Boecker F, Srinivas G, Schoof H. AHRD. Available from: <https://github.com/groupschoof/AHRD>.
34. Wang X, Gao J, Gao S, Li Z, Kuai B, Ren G. REF6 promotes lateral root formation through de-repression of PIN1/3/7 genes. *J Integr Plant Biol*. 2018;61:383–7. <https://doi.org/10.1111/jipb.12726>.
35. Luo M, Hung F-Y, Yang S, Liu X, Wu K. Histone lysine demethylases and their functions in plants. *Plant Mol Biology Report*. 2013;32:558–65. <https://doi.org/10.1007/s11105-013-0673-1>.
36. Lu L, Lee Y-RJ, Pan R, Maloof JN, Liu B. An internal motor Kinesin is associated with the golgi apparatus and plays a role in trichome morphogenesis in *Arabidopsis*. *Mol Biol Cell*. 2005;16:811–23.
37. Oda Y, Fukuda H. Rho of Plant GTPase Signaling Regulates the Behavior of *Arabidopsis* Kinesin-13A to Establish Secondary Cell Wall Patterns. *The Plant Cell*. 2013;25:4439–4450. Available from: <https://doi.org/10.1105/tpc.113.117853> DOI: 10.1105/tpc.113.117853.
38. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol*. 2015;32:1342–53. <https://doi.org/10.1093/molbev/msv022>.
39. Cattivelli L, Rizza F, Badeck F-W, Mazzucotelli E, Mastrangelo AM, Francia E, Maré C, Tondelli A, Stanca AM. Drought tolerance improvement in crop plants: an integrated view from breeding to genomics. *Field Crops Res*. 2008;105:1–14. <https://doi.org/10.1016/j.fcr.2007.07.004>.
40. Sharp R. Interaction with ethylene: changing views on the role of abscisic acid in root and shoot growth responses to water stress. *Plant Cell & Environ*. 2002;25:211–22. <https://doi.org/10.1046/j.1365-3040.2002.00798.x>.
41. Blum A. Osmotic adjustment is a prime drought stress adaptive engine in support of plant production. *Plant Cell & Environ*. 2017;40:4–10.
42. Blum A. Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. *Field Crops Res*. 2009;112:119–23. <https://doi.org/10.1016/j.fcr.2009.03.009>.
43. Nikolov LA, Shushkov P, Nevado B, Gan X, Al-Shehbaz IA, Filatov D, Bailey CD, Tsiantis M. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytol*. 2019;222:1638–51. <https://doi.org/10.1111/nph.15732>.
44. Kagale S, Robinson SJ, Nixon J, Xiao R, Huebert T, Condie J, Kessler D, Clarke WE, Edgar PP. Links MG and others. Polyploid evolution of the Brassicaceae during the cenozoic era. *Plant Cell*. 2014;26:2777–91. <https://doi.org/10.1105/tpc.114.126391>.

45. Edger PP, Heide-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafu EK, Tang M and others. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences*. 2015;112:8362–8366. <https://doi.org/10.1073/pnas.1503926112>
46. Hendriks KP, Kiefer C, Al-Shehbaz IA, Bailey CD, van Huysduynen AH, Nikolov LA, Nauheimer L, Zuntini AR, German DA, Franzke A and others. Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. *Curr Biol*. 2023;33:4052–68. <https://doi.org/10.1016/j.cub.2023.08.026>
47. Moore JP, Vitré-Gibouin M, Farrant JM, Driouch A. Adaptations of higher plant cell walls to water loss: drought vs desiccation. *Physiol Plant*. 2008;134:237–45. <https://doi.org/10.1111/j.1399-3054.2008.01134.x>
48. Ganie SA, Ahammed GJ. Dynamics of cell wall structure and related genomic resources for drought tolerance in rice. *Plant Cell Rep*. 2021;40:437–59. <https://doi.org/10.1007/s00299-020-02649-2>
49. Force A, Lynch M, Pickett FB, Amores A, Yan Y-I, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 1999;151:1531–45. <https://doi.org/10.1093/genetics/151.4.1531>
50. Zhang J. Evolution by gene duplication: an update. *Trends Ecol Evol*. 2003;18:292–8. [https://doi.org/10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8)
51. Lü S, Zhao H, Des Marais DL, Parsons EP, Wen X, Xu X, Bangarusamy DK, Wang G, Rowland O, Juenger T and others. Arabidopsis ECERIFERUM9 involvement in cuticle formation and maintenance of plant water status. *Plant Physiol*. 2012;159:930–44.
52. Lü S, Song T, Kosma DK, Parsons EP, Rowland O, Jenks MA. Arabidopsis CER8 encodes LONG-CHAIN ACYL-COA SYNTHETASE 1 (LACS1) that has overlapping functions with LACS2 in plant wax and Cutin synthesis. *Plant J*. 2009;59:553–64.
53. Grünhofer P, Herzog L, Zhang Q, Vitt S, Stöcker T, Malkowsky Y, Brüggemann T, Fladung M, Schreiber L. Changes in wax composition but not amount enhance cuticular transpiration. *Plant Cell Environ*. 2024;47:91–105.
54. Chen J-J, Sun Y, Kopp K, Oki L, Jones SB, Hipps L. Effects of water availability on leaf trichome density and plant growth and development of *Shepherdia xutahensis*. *Front Plant Sci*. 2022;13. <https://doi.org/10.3389/fpls.2022.855858>
55. Kärkkäinen K, Løe G, Ågren J. Population structure in *Arabidopsis lyrata*: evidence for divergent selection on trichome production. *Evolution*. 2004;58:2831–6. <https://doi.org/10.1111/j.0014-3820.2004.tb01634.x>
56. Huttunen P, Kärkkäinen K, Løe G, Rautio P, Ågren J. Leaf trichome production and responses to defoliation and drought in *Arabidopsis lyrata* (Brassicaceae). 2010;47:199–207.
57. Lu F, Cui X, Zhang S, Jenuwein T, Cao X. Arabidopsis REF6 is a histone H3 lysine 27 demethylase. *Nat Genet*. 2011;43:715–9. <https://doi.org/10.1038/ng.854>
58. Antonova SV, Boeren J, Timmers HM and Snel B. Epigenetics and transcription regulation during eukaryotic diversification: the saga of TFIIID. *Genes & Development*. 2019;33:888–902. DOI: 10.1101/gad.300475.117.
59. Choi K, Kim J, Hwang H-J, Kim S, Park C, Kim SY, Lee I. The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in arabidopsis, by recruiting chromatin modification factors. *Plant Cell*. 2011;23:289–303. <https://doi.org/10.1105/tpc.110.075911>
60. Yang R, Jarvis DE, Chen H, Beilstein MA, Grimwood J, Jenkins J, Shu S, Prochnik S, Xin M, Ma C, Schmutz J, Wing RA, Mitchell-Olds T, Schumaker KS, Wang X. The reference genome of the halophytic plant *Eutrema salsugineum*. *Front Plant Sci*. 2013;4. <https://doi.org/10.3389/fpls.2013.00046>
61. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H and others. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet*. 2011;43:476–81. <https://doi.org/10.1038/ng.807>
62. Cheng C, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. Araport1.1: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J*. 2017;89:789–804. <https://doi.org/10.1111/tpj.13415>
63. Yates A, Allen J, Amode RM, Azov AG, Barba M, Becerra A, Bhai J, Campbell L, Carbajo Martinez M, Chakiachvili M, Chougule K, Christensen M, Contreras-Moreira B, Cuzick A, Da Rin Fioretto L, Davis P, De Silva N, Diamantakis S, Dyer S, Elser J, Filippi CV, Gall A, Grigoriadis D, Guizarro-Clarke C, Gupta P, Hammond-Kosack K, Howe KL, Jaiswal P, Kaikala V, Kumar V, Kumari S, Langridge N, Le T, Luypaert M, Maslen GL, Maurel T, Moore B, Muffato M, Mushtaq A, Naamati G, Naithani S, Olson A, Parker A, Paulini M, Pedro H, Perry E, Preece J, Quinton-Tulloch M, Rodgers F, Rosello M, Ruffier M, Seager J, Sitnik V, Szpak M, Tate J, Tello-Ruiz M, Trevanion S, Urban M, Ware D, Wei S, Williams G, Winterbottom A, Zarowiecki M, Finn R and Flicek P. Ensembl genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Res*. 2021;50:D996–1003. <https://doi.org/10.1093/nar/gkab1007>
64. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau D, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick B, Pruitt K, Sherry S. Database resources of the National center for biotechnology information. *Nucleic Acids Res*. 2021;50:D20–6. <https://doi.org/10.1093/nar/gkab112>
65. Love MI, Huber W, Anders S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1–21. <https://doi.org/10.1186/s13059-014-0550-8>
66. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34:525–7. <https://doi.org/10.1038/nbt.3519>
67. R Core Team. R: A Language and Environment for Statistical Computing. 2024. Available from: <https://www.R-project.org/>
68. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. 2011;6:e21800. <https://doi.org/10.1371/journal.pone.0021800>
69. Mölder F, Jablonski K, Letcher B, Hall M, Tomkins-Tinch C, Sochat V, Forster J, Lee S, Twardziok S, Kanitz A, Wilm A, Holtgrewe M, Rahmann S, Nahnsen S, Köster J. Sustainable data analysis with snakemake. *F1000Research*. 2021;10. <https://doi.org/10.12688/f1000research.29032.2>. version 2; peer review: 2 approved.
70. Uebermuth-Feldhaus C. caroue/phd-selection: Brassicaceae Drought Project. 2024. <https://doi.org/10.5281/zenodo.14540010>
71. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 2004;21:676–9. <https://doi.org/10.1093/bioinformatics/bti079>
72. Edgar RC. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun*. 2022;13. <https://doi.org/10.1038/s41467-022-34630-w>
73. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010;5:e9490.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.