

Enrichment of Intrinsically Disordered Residues in Ohnologs Facilitate Abiotic Stress Resilience in *Brassica Rapa*

Shayani Das Laha

Raiganj University

Deeya Saha

University of Calcutta - Rajabazar Science College Campus

Tapash Ghosh

Bose Institute

Soumita Podder (✉ piu2007.soumita@gmail.com)

Raiganj University <https://orcid.org/0000-0002-8140-508X>

Research article

Keywords: Abiotic stress, Whole genome duplication, Functional divergence, Intrinsically disordered regions, Domain

Posted Date: September 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-64001/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: *Arabidopsis thaliana* and *Brassica rapa* shared a common evolutionary clade but Brassica species experienced an extra whole genome triplication (WGT) event compared with the model plant *A. thaliana*. This extra round of WGT confers *B. rapa* more abiotic stress resistant. The study aims to unravel how the consequences of whole genome duplication steer the variation in stress adaptation competency between the two species.

Result: Comparing the duplication status between abiotic stress resistant (ASR) genes in the two species, significant increase in the number of paralogs in ASR genes of *B. rapa* than *A. thaliana* was found. Investigation on the proteomic features suggests that the ohnologs pairs in both species are more enriched with intrinsically disordered residues (IDRs) than other duplicated pairs but IDRs only in *B. rapa* have showed a significant positive correlation with functional divergence between the duplicated pairs. The functional divergence helps to mediate more stress adaptation functions in the ohnologs of *B. rapa* than that of *A. thaliana*. Moreover, domain ontology analysis has revealed that the new domains with stress functions are significantly more enriched in the ohnologs of *B. rapa*. Interestingly, majority of these stress tolerant domains are found to be present in the intrinsically disordered regions of the proteins. Statistical analysis along with these observations make it reasonable to speculate that IDRs expedite stress adaption potentiality in *B. rapa* by enriching more stress related functions and accommodating stress specific domains in ohnologs of this plant species.

Conclusion: With the occurrence of WGT in the Brassica species the stress resilience features of *B. rapa* as compared to *A. thaliana* is less studied. This study unveils the role of intrinsically disordered residues (IDRs) of ohnologs in optimizing more stress resilience in *B. rapa* than *A. thaliana*. Thus, it will open new avenues in understanding the mechanism of succeeding stress adaptation potentiality in *B. rapa* over evolutionary time.

Background

Brassica rapa, member of family Cruciferae (Brassicaceae) is one of the most important economic crops throughout the world since it is a major source of edible mustard oilseed. The Brassicaceae family is one of the largest plant groups comprising about 49 tribes, 321 genera, and 3660 species [1] and each species contains rich morphotypes showing extreme traits. Paleopoloidization events are recurrent among flowering plants, and the duplicate genes originated via such events contribute significantly to plant evolution. There are different mechanisms responsible for gene duplication such as (a) Whole Genome Duplication (WGD) where duplication takes place via increase in ploidy level, (b) Tandem duplication, which occurs via duplication of a gene through unequal crossing over between similar alleles, (c) Transposed duplication which is mediated through transposons mediated replicative transposition. (d) Dispersed duplication (DSD) which occurs through unpredictable and random patterns by mechanisms that remain unclear, generating two gene copies that are neither colinear nor they lie by side [2]. Among these all mechanisms, polyploidy or whole-genome duplication have imparted much in the

morphological and physiological diversity in the course of evolution [3, 4]. While most of the plant's species underwent whole genome duplication there are many species with triplication such as Brassicaceae family. The Brassicaceae genomes have experienced two rounds of ancient whole genome duplication (WGDs) followed by an additional genome triplication in the *Brassica* species approximately 14-24 million years ago [5, 6]. Thus, *Brassica* species endured an extra whole genome triplication (WGT) event compared with the model plant *Arabidopsis thaliana*. Whole genome sequencing of the *Brassica* species demonstrated that WGT plays an imperative role in the speciation and morphotype diversification of *Brassica* plants [7]. Prior studies have revealed that paleopoloidization not only surges gene expression and cell proliferation but also build up different cellular potentiality like development in cell signalling, homeostasis, and majorly stress resistance capability [8, 9]. It was evident in several reports that many crop species increase tolerance to water deficit, temperature, or salt stresses after polyploidization [10-12]. Polyploid plants like *Brassica rapa* L. and black locust (*Robinia pseudoacacia* L.) showed improved acclimatization to salt stress than their diploids counterparts [13-16]. Also, tetraploid *Dioscorea zingiberensis* exhibited higher heat resistance than their diploids [17].

Now, the molecular mechanism behind the stress tolerance by duplicated genes is one of the thrust areas of research. It is inevitable that adaptation to stressful conditions requires modified expression of certain genes which is obtained by genome sequence changes. The duplicated genes undergo many modification or alterations in themselves such as (A) changes in the protein coding region: alternative splicing [18], gain or loss of exon-intron [19], single nucleotide change [20]; (B) alterations in the regulatory regions like changes in cis elements [21], small RNA binding sites [22], and methylation patterns [23]. The other major dynamic aspect was reported by [24] that impose a new dimension in the consequence of gene duplication relating to protein structural change. They have hypothesized that gain or losses of intrinsically disordered regions in the proteins of duplicated genes are responsible for functional divergence in duplicated genes [24]. Lack of stable 3D structure of intrinsically disordered regions (IDRs) entail them extreme flexibility which often enable more interactions with proteins thereby could play multiple biological functions [25, 26]. In different plant species like *Zea mays*, *Glycine max*, *P. trichocarpa* it was noticed that IDRs increased in paralogs after WGD or WGT event contribute to the stability of functions such as the catalytic activity of proteins, metabolic and transport processes, and molecular binding [27]. Henceforth, we keen to investigate the interplay between WGD and IDRs in manifesting more abiotic stress tolerance in *Brassica rapa* compared to *A. thaliana*.

Results

1. Gene duplication profiles in abiotic stress responsive genes in *thaliana* and *B. rapa*.

Gene duplication or Paleopoloidization is thought to contribute in the evolution of ultmorphological and ecological diversity. It is a phenomenon which allows the genes to gain novel features by neofunctionalization and sub functionalization [3,4]. It has been reported earlier that the abiotic and biotic stress responsive genes were increased after whole genome duplication (WGD) [28]. In order to reinvestigate the role of gene duplication in increased stress (abiotic) tolerance in *Brassica rapa* than *A.*

thaliana, we have compared duplication status between abiotic stress resistance (ASR) genes of *A. thaliana* with their corresponding orthologs present in *B. rapa*. We have noticed that proportion duplicates in stress resistant genes in *B. rapa* is significantly (P value = 0.00001, Z proportionality test) higher (73.47%) than the stress resistance genes in *A. thaliana* (52.09%). Moreover, the average number of paralogs of ASR genes are also found to be significantly (P value=0.0001, Mann-Whitney-U test) higher in *B. rapa* (3.3) than *A. thaliana* (1.8). These observations clearly reflect the fact that Brassica species underwent five rounds of WGD whereas *A. thaliana* experienced two rounds of WGD [6, 29]. Thus, this result suggests that duplication may play possible roles in maintaining the stress resilience attribute.

Formerly, it has been reported that different modes of duplication influence the functional role in a biased way [30]. Therefore, in our study mode of duplication of the stress genes was derived and were classified into one of the four categories being derived from WGD (Whole genome duplication), Tandem duplication (TD), Transposed duplication (TRD) and Dispersed Duplication (DSD). The number of duplicate gene pairs for each category in each taxon was determined. Percentage of genes underwent different modes of duplication in *A. thaliana* and *B. rapa* is delineated in Figure 1. From the above figure it could be clearly depicted that the percentage of stress genes experiencing WGD is significantly (P value=0.0001, Mann-Whitney-U test) higher in *B. rapa* (23.5%) than that of *A. thaliana* (12.9%). Thus, WGD could be considered as a driving force for the evolution of stress adaptive genes from *A. thaliana* to *B. rapa* as also described by Rizzon [28].

2. Enrichment of intrinsically disordered regions in proteins encoded by duplicated pairs of ASR genes in *thaliana* and *B. rapa*

Ohnologs, the duplicates derived by whole genome duplication were reported to contain more intrinsically disordered residues than small scale duplicates [31]. It was also reported in plant system that the proteins of dehydrin family being almost completely disordered could involve in the response to drought and other environmental stresses [32]. Thus, we have analysed intrinsically disordered protein profile of ASR coded genes in the two species. Consistent with the previous study [31], we have found that the intrinsically disordered residues (IDRs) are significantly more enriched in WGD derived duplicated proteins than tandem, dispersed and transposed duplicates in both species (Figure 2). Consequently, it is also evident from our result that in both plants, duplicated ASR proteins are more frequently contain IDRs than singleton ASR genes (Figure 3). Because of this fact, average disordered residues are also significantly higher in duplicated ASR encoding proteins compared to singleton ASRs in *A. thaliana* (Average = 0.20) and *B. rapa* (Average= 0.27). Thus, it could be manifested from the above facts that WGD is associated with the enhancement of IDRs in proteins. If so, it could be assumed that ohnologs of *B. rapa* encode more intrinsically disordered proteins than *A. thaliana* since former is underwent three extra round of WGD. Consistent with our conjecture, we too found that 40.5 % of ohnolog pairs in *B. rapa* contain at least one intrinsically disordered region but in *A. thaliana* it is only 29.9 %. The difference is significant at P value<0.0001 in Z-proportionality test. Moreover, the average percentages of intrinsically disordered residues (IDRs) are significantly ($P = 0.001$, Mann-Whitney Utest) high in the ohnologs of *B. rapa* compared to *A. thaliana* (Figure 4). Moreover, we also retrieved a significant positive correlation between

WGD derived paralogs number and content of intrinsically disorder residues in *B. rapa* (Spearman's $\rho = 0.1547$, $P = 2e-06$) but not in *A. thaliana*. Hence it would be interesting to investigate how the IDRs in proteins acquired through WGD during evolution stimulate the stress adaptation potentiality in *B. rapa* than *A. thaliana*.

3. Functional divergence in duplicated pairs of ASR genes originated through whole genome duplication in *thaliana* and *B. rapa*

Genome duplication (polyploidy) is a conventional phenomenon of plant evolution. It was proposed earlier that functional diversification of the surviving duplicated genes is one of the prime attributes for the long-term evolution of polyploids [33]. It is widely accepted that enrichment of intrinsically disordered residues in proteins could able to impose them diverse functional potentiality because of their high flexibility in nature and IDRs are also reported to create functional divergence between ohnologs after WGD [24]. We have measured the functional divergence based on the Gene Ontology terms between the WGD derived duplicates in the two selected species. We have found that paralogs of 67.30% genes in *B. rapa* turned into functionally diverged after whole genome duplication event whereas it is only 50% for *A. thaliana*. The difference in the proportion of gene in these two species is also significant at 0.99 confidence level. The average functional divergence between the paralogous pairs in *B. rapa* is also found to be significantly (Mann-Whitney U test, $P = 0.009$) higher than the paralogous pairs in *A. thaliana* (*B. rapa* = 0.90; *A. thaliana* = 0.88). Next, we checked the correlation between IDRs and functional divergence between duplicated pairs in our datasets and received a strong positive correlation between them in *B. rapa* but not in *A. thaliana* (Spearman's $\rho = 0.1458$, $P = 6.539e-05$). These results suggest that the increased disordered residues in duplicated pairs of *B. rapa* may play an important role in creating new functions in ohnologs which in turn impose more stress resistance potentiality in *B. rapa* compared to *A. thaliana*.

To give a detailed insight into the functions which are gained after WGD, we have performed functional enrichment analysis of ASR genes and their corresponding paralogs in *B. rapa* and *A. thaliana*. We have noticed that three additional stress related functions (response to heat, response to abiotic stimulus, response to temperature stimulus) are more successfully enriched in the paralogs of *B. rapa* than their ancestral genes (Figure 5) but in case of paralogs in *A. thaliana* no such significant stress related functional enrichment is observed (Figure 5). These results indicate that IDRs in duplicated proteins because of their flexible nature can create functional divergence by changing protein interaction pattern which in turn could excite more stress adaptation potentiality in *B. rapa*.

4. Enrichment of protein domains encoded by duplicated pairs of ASR genes in *thaliana* and *B. rapa*

Protein domains entail functional modularity, and it is likely that functional divergence between paralogous pairs exploits this modularity. Experimental evidences reported few domains are assigned for stress tolerance in plants [34] we have assessed protein domains in duplicated pairs by scanning their amino acid sequences. We noticed that duplicated pairs of ASR genes in *B. rapa* encode significantly (Mann-Whitney U test, $P = 0.024$) higher number of domains than the duplicated pairs in *A. thaliana* (*B.*

rapa- average no. of domain = 3.9, *A. thaliana*, average no. of domain = 1.20). Interestingly, when we have performed domain enrichment analysis by Fisher Exact test, we found during divergence from *A. thaliana*, *B. rapa* lost some domains as well as several new domains are gained in the ohnolog pairs of *B. rapa* (Figure 6). Next, we have analysed domain ontology by dcGO to compare functional enrichment of unique domains present in *A. thaliana* and *B. rapa* duplicated pairs. Figure 7 has clearly depicted that stress related functions are more enriched in the domain which are solely present in *B. rapa* than the domains in *A. thaliana*. The domains in *B. rapa* which are annotated with the stress related functions are enlisted in Table 1. In previous research, the distribution of disordered regions in GATA- type binding Domain is responsible for manifesting functional specificity [35]. Therefore, we also analysed the domain content in disordered regions of the protein. The results we found was consistent with the previous results where the percentage of stress specific domain in disordered regions ($9/14=64.3\%$) was significantly ($P=0.001$, Z proportionality test) higher than the domain present in the ordered domain ($5/14=35.7\%$) (Table1). The enrichment of stress related domains in the disordered regions of duplicated pairs provides an important cue that genome triplication in *B. rapa* helps to encode disordered residues which might help in their stress resistance.

Table 1. Table enlisting domains in *B. rapa* annotated with the stress related functions. Annotated domain with disorder are Bold marked

GO term	Annotated domain
Cellular response to environmental stimulus	PF00005, PF00010, PF00013, PF00027
Cellular response to light stimulus	PF00010, PF00013, PF00027
Defense response	PF00031, PF00240, PF00314
Positive regulation of response to stimulus	PF00004, PF00005, PF00010, PF00013, PF00023, PF00026, PF00240
Regulation of DNA-templated transcription in response to stress	PF00004, PF00012, PF00023, PF00240
Regulation of response to external stimulus	PF00010, PF00013, PF00031
Regulation of response to stimulus	PF00004, PF00005, PF00010, PF00012, PF00013, PF00023, PF00026, PF00027, PF00031, PF00240
Regulation of response to stress	PF00004, PF00010, PF00012, PF00013, PF00026, PF00031, PF00240
Regulation of transcription from RNA polymerase II promoter in response to stress	PF00004, PF00012, PF00023, PF00240
Response to abiotic stimulus	PF00004, PF00005, PF00006, PF00010, PF00011, PF00012, PF00013, PF00023, PF00027, PF00155
Response to biotic stimulus	PF00005, PF00012, PF00026, PF00240, PF00314
Response to external stimulus	PF00005, PF00010, PF00012, PF00026, PF00027, PF00240, PF00314
Response to heat	PF00004, PF00011, PF00012, PF00023
Response to hypoxia	PF00004, PF00010, PF00155, PF00240
response to stress	PF00004, PF00005, PF00010, PF00011, PF00012, PF00023, PF00031, PF00155 PF00240, PF00314
response to temperature stimulus	PF00004, PF00010, PF00011, PF00012, PF00013, PF00023

5. Statistical analysis of IDRs and stress resistance potentiality of *rapa*

Here, we have noticed that whole genome gene duplication triggers the enrichment of intrinsically disordered residues in proteins. These IDRs mediate functional divergence in the ohnologs as well as confer the sites for enriching new domains. The functional divergence in the duplicated pairs and the new domains originated during evolution of *A. thaliana* to *B. rapa* contribute a lot for escalating stress tolerance potentiality in *B. rapa*. Thus, we intend to explore whether these four factors (no. of paralogs, functional divergence, IDR content, domain content) in ohnologs acts mutually inclusive way or they are

independent in *B. rapa*. For this, we have categorized stress related genes in two classes- mono stress (associated with one stress condition) and poly stress (associated with more than one stress condition). Then, we have performed linear regression analysis with stress and stress regulatory four factors. We found that only IDRs could independently control stress resistance potentiality in *Brassica rapa* (Spearman's $\rho = 0.185$, $P = 5.943e-03$).

Discussion

Comparative genome analysis have shown that *A. thaliana* shared a common evolutionary clade and underwent whole genome duplication around 20 MYA whereas the *Brassica* lineage went through whole genome triplication (WGT) [6, 36-40]. Gene duplication is the key source of genetic novelty originated during genomic evolution. This genetic novelty drives structural changes within the DNA which may lead to changes in the protein structure as well as its function. Recent study proposed that with the occurrence of this WGT, successive genomic rearrangements, divergence and speciation in *Brassica* species confer *B. rapa* more abiotic stress resilient than *A. thaliana* [41]. It will therefore be very interesting to explore the attributes imposed by gene duplication to mediate stress tolerance in plants. Comparing ASR genes of *A. thaliana* and their corresponding orthologs present in *B. rapa*, we have noticed that most of the ASR genes in *B. rapa* (82.3%) take part in gene duplication events. Along with that numbers of paralogs are significantly higher in *B. rapa* than their ancestral species. Subsequently, it is also evident from our result that the number of paralogs originated via WGD in *A. thaliana* (12.9%) is nearly half to the paralogs in *B. rapa* (23.5%) which reflects the effects of extra round of genome duplication in *B. rapa*. This result is also echoed in a previous study on *A. thaliana* hypothesized that many of the identified duplicated gene pairs evolving by WGD mechanism might function in cell wall integrity pathways, growth and stress tolerance [42]. Unlike small-scale gene duplications, whole-genome duplications copy entire pathways or networks, thus play an important role in the evolution of novel traits and increased biological complexity [43-45]. Previous corroborations illustrate that genomic recombination and chromosomal rearrangements during gene duplication would create flexible or ductile regions in proteins which are termed as intrinsically disordered proteins [27, 46] There is profuse evidence that disordered residues confer flexibility to proteomes [47-49, 35] most importantly they contribute to organismic plasticity by facilitating complex gene regulatory network dynamics and protein multifunctionalities [50-58]. To further explore the role of WGD events in stress tolerance of plant we have analysed disordered content of ASR gene paralogs originated through different modes of duplication i.e. WGD, tandem, transposed and dispersed. We noticed that IDRs content is explicitly enriched in ohnologs of *A. thaliana* and *B. rapa* than any other paralogs (Figure 2). By comparing average IDRs in the ohnologs of *B. rapa* and *A. thaliana*, we have confirmed that WGT events in *B. rapa* increased IDRs in their proteomes thereby may contribute to more successive adaptation to stress. To affirm this hypothesis, we have measured functional divergence in the ohnologs of both species, where we found that *B. rapa* ohnologs undergo more functional divergence than *A. thaliana*. Interestingly, we have observed that 76.03% of ohnologs having at least one disordered region (residues >30) are become functionally diverged and only 0.03 % remain functionally redundant after WGD in *B. rapa* and we also retrieved significant positive correlation between IDRs and functional

divergence (Spearman's $\rho = 0.1458$, $P= 6.539e-05$). However, in case of *A. thaliana* paucity in IDRs in their proteome leads to the deviation of the above fact in them. Moreover, we have illustrated that the functional divergence in the duplicated pairs of *B. rapa* indeed help in the enrichment of stress responsive functions in their paralogs (Figure 5). Our result supports previous evidence on a distinguished NAC family protein in *A. thaliana* that stated the differences in IDR region contributed to functional divergences of NAC92 and NAC59 which are involved in senescence, salt stress responses, and lateral root development [35, 59]. Now, protein functions achieved via protein-protein interaction which is invariably mediated by protein domain. The structure-function continuum concept i.e. unique biological functions of proteins require unique 3D-structures dominated scientific minds for more than a century. However, the growing understanding of the significant functional role of IDRs has challenged the year-old concept, proving indisputably that structure-less IDPs/IDPRs are functional, being able to engage in biological activities and perform impossible tricks that are highly unlikely for ordered proteins. Presence of IDRs in the binding domain of singlish interface hub proteins for pursuing diverse functionality had also been reported in few years back [60-62]. We have studied the domain content in ordered and disordered region of proteins in the ohnologs of *B. rapa* and *A. thaliana*. We found that the number of total domain content (present in ordered and disordered regions) per gene is significantly high in *B. rapa* than *A. thaliana*. Computing domain enrichment analysis, we evidenced that 181 new domains are gained in *B. rapa* which were entirely absent in their orthologs in *A. thaliana* (Figure 6). Thus, we could infer that these domains are acquired after WGT events in *B. rapa*. Interestingly, analysing the ontology of domains present in the ordered and disordered region, we noticed that stress related functional domains are significantly more enriched in disordered regions. These results provide significant evidence on the origin of IDRs by evolution and their implications for stress tolerance. Now, to prove the stress responsive roles of IDRs statistically, we have performed partial and regression analysis in *Brassica rapa* by categorizing the stress genes according to number of stress factors into poly stress (involve in more than one stress response) and mono stress (restricted in one stress response). From this analysis we have reached to the inference that IDRs in proteins could independently guide stress tolerance potentiality in the ohnologs of *B. rapa*.

Conclusion

After the phenomenon of whole genome triplication (WGT) Brassica species experienced an additional round of whole genome duplication compared with the model plant *A. thaliana*. the occurrence of this WGT, confer *B. rapa* more abiotic stress resistant than *A. thaliana* by successive genomic rearrangements, divergence and speciation. Based on our study, it is not unreasonable to speculate that IDRs mediate stress adaption potentiality in *B. rapa*. From, the result it is evidenced that IDRs could independently govern the stress resilience of *B. rapa*. We have illustrated the fact that the structural flexibility of IDRs expedites the stress tolerant attributes in paralogs by enriching more stress related functions and accommodating stress specific domains in *B. rapa*. This study will open new avenues in understanding that over evolutionary time, how duplication confers stress adaptation potentiality in *B. rapa*.

Methods

1. Identification of abiotic stress genes *thaliana* and *B. rapa*

We obtained a list of Arabidopsis stress responsive genes from TAIR (<https://www.arabidopsis.org/>), PHYTOZOME (<https://phytozome.jgi.doe.gov/pz/portal.html>), PLAZA (<https://bioinformatics.psb.ugent.be/plaza/>) using the keyword search salinity, drought, heat, cold stress and abiotic stress as they are the major abiotic stress conditions. The redundant gene identifiers were removed and a total set 1028 genes from *A. thaliana* were obtained. Orthologous genes of *A. thaliana* stress response genes in *B. rapa* were downloaded from Ensembl Plants (<https://plants.ensembl.org/biomart/martview>, Release 43). Coding sequences of *B. rapa* having at least 60% identity and 60% of their length with *A. thaliana* were identified as orthologs. We have retrieved 844 orthologs in *B. rapa*. After scanning the stress related GO functions of these orthologous genes we have finally obtained 773 stress response genes from *B. rapa*. For duplicated genes, paralogs for both *A. thaliana* and *B. rapa* were retrieved using Biomart from ensemble plants. Paralogous genes having at least 60% sequence identity are selected for the study. The coding and amino acid sequences of the stress genes and their paralogs are extracted from Ensembl. The abiotic stress genes were categorized into singleton genes without any or with only one paralog and duplicated genes with more than one paralogs.

2. Identification of Duplication Mode

Plant Duplicate Gene Database (PlantDGD, <http://pdgd.njau.edu.cn:8080>) was used to categorise genes into different modes of duplications as WGD, Tandem, Transposed and Dispersed. Duplicated gene pairs list were downloaded from the database and were matched with the abiotic stress genes and then categorised.

3. Functional Divergence and Functional Enrichment Analysis in duplicated genes

The GO annotation for *B. rapa* and *A. thaliana* ohnologs were obtained from the ENSEMBL Biomart([http://plants.ensembl.org/biomart/martview](https://plants.ensembl.org/biomart/martview)). Among the three top GO categories: molecular function (MF), biological process (BP), and cellular component (CP), only the Biological process were considered for functional divergence analysis. Functional divergence was determined using Czekanowski- Dice distance formula.

Functional distance $(i, j) = [\text{Number of(Terms (i) Terms(j))} / \text{Number of(Terms(i) Terms(j))}]$ Number of (Terms(i) Terms(j))]

Here, i represents GO terms of individual plant genes, j represents GO terms of the paralogous genes, Δ corresponds to the symmetrical between the GO term sets of two genes, \cup and \cap represents the non-redundant and common GO terms, respectively.

Functional enrichment analysis was done by Gene Ontology Annotation Database (ref). We have used corrected P value < 0.05 as the threshold to determine significant overrepresentation of certain GO terms.

4. Identification of Intrinsically Disordered regions

The prediction of intrinsic protein disorder was carried out by the IUPred (<https://iupred.elte.hu/>) algorithm using the option “long disordered regions”. We choose IUPred algorithm for disorder prediction as IUPred didn’t train on any specific dataset and hence it offers an unbiased estimate of disorder score [63]. IUPred assigns disorder score based on the pairwise interaction energy score for each amino acid [64]. Residues having a disorder score above threshold value 0.5 were considered as disordered residues.

5. Prediction of stress related functional domain enrichment in the duplicated genes

Pfam domains analysis was performed using Pfam 32.0 database(<https://pfam.xfam.org/>)which uses the basics of HMMER (v 3.0). Then obtained the Pfam ids were search with dcGOdatabase (<http://supfam.org/SUPERFAMILY/cgi-bin/dcenrichment.cgi>) for the enrichment analysis of domain centric gene ontology. To evaluate significant enrichment of certain GO terms in the identified domains, P values were calculated based on Poisson function. Finally, we used P value < 0.05 as the threshold to determine significant overrepresentation of certain function specific domain.

6. Statistical tests

Online Z-test was carried out for proportionality test between two groups. All other statistical analyses and Heatmap were performed using R.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Additional file 1: Dataset of abiotic stress resistance genes of *A. thaliana* and their corresponding orthologs playing similar functions in *B. rapa*.

Competing interests

All authors did not have any conflict of interest.

Funding

The research discussed in this paper is funded by Swami Vivekanand Merit-Cum-Means Scholarship conducted by Government of West Bengal.

Authors' contributions

SP designed the work, SDL executed the whole work, DS helped in several analyses, SP and SDL wrote the manuscript, TCG helped in critical reading of manuscript.

Acknowledgements

We also sincerely thanks Raiganj University, West Bengal and Bose Institute, Kolkata for substantially providing the workstation to carry out the research work.

References

1. Al-Shehbaz IA. A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon*. 2012; 61: 931–954. [1002/tax.615002](https://doi.org/10.1002/tax.615002).
2. 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: 38. [10.1186/s13059-019-1650-2](https://doi.org/10.1186/s13059-019-1650-2).
3. Paterson AH, Freeling M, Tang H, Wang X. Insights from the comparison of plant genome sequences. *Annu. Rev Plant Biol.* 2010; 61: 349–372. [1146/annurev-arplant-042809-112235](https://doi.org/10.1146/annurev-arplant-042809-112235).
4. Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. Polyploidy and angiosperm diversification. *Am J Bot.* 2009; 96: 336-48. [10.3732/ajb.0800079](https://doi.org/10.3732/ajb.0800079).
5. Koch MA, Haubold B, and Mitchell-Olds T. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol.* 2000; 17: 1483–1498. [10.1093/oxfordjournals.molbev.a026248](https://doi.org/10.1093/oxfordjournals.molbev.a026248).
6. Lysak MA, Koch MA, Pecinka A, Schubert I. Chromosome triplication found across the tribe Brassiceae. *Genome Res.* 2005; 15: 516-525. [1101/gr.3531105](https://doi.org/10.1101/gr.3531105).
7. Cheng F, Wu J, Wang X. Genome triplication drove the diversification of *Brassica* plants. *Horticulture Research.* 2014; 1: 14024. [1038/hortres.2014.24](https://doi.org/10.1038/hortres.2014.24)
8. Yu Z, Haberer G, Matthes M, Rattei T, Mayer KF, Gierl A, Torres-Ruiz RA. Impact of natural genetic variation on the transcriptome of autotetraploid *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences USA.* 2010; 107: 17809–17814. [1073/pnas.1000852107](https://doi.org/10.1073/pnas.1000852107)

9. Li X, Yu E, Fan C, Zhang C, Fu T, Zhou Y. Developmental, cytological and transcriptional analysis of autotetraploid *Arabidopsis*. *Planta*. 2012; 236: 579–596. 10.1007/s00425-012-1629-7.
10. Allario T, Brumos J, Colmenero-Flores JM, Iglesias DJ, Pina JA, Navarro L, Talon M, Ollitrault P, Morillon R. Tetraploid Rangpur lime rootstock increases drought tolerance via enhanced constitutive root abscisic acid production. *Plant Cell and Environment*. 2013; 36: 856–868. [1111/pce.12021](https://doi.org/10.1111/pce.12021)
11. Deng B, Du W, Liu C, Sun WS, Tian S, Dong H. Antioxidant response to drought, cold and nutrient stress in two ploidy levels of tobacco plants: low resource requirement confers polytolerance in polyploids?. *Journal of Plant Growth Regulation*. 2012; 66: 1–37. 10.1007/s10725-011-9626-6
12. Kasajima I, Ide Y, Yokota HM, Fujiwara T. WRKY6 is involved in the response to boron deficiency in *Arabidopsis thaliana*. *Physiologia Plantarum*. 2010; 139: 80–92. 10.1111/j.1399-3054.2010.01349.x.
13. Saleh B, Allario T, Dambier D, Ollitrault P, Morillon R. Tetraploid citrus rootstocks are more tolerant to salt stress than diploid. *Comptes Rendus Biologies*. 2008; 331: 703–710. 10.1016/j.crvi.2008.06.007.
14. Allario T, Brumos J, Colmenero-Flores JM, Tadeo F, Froelicher Y, Talon M, Navarro L, Ollitrault P, Morillon R. Large changes in anatomy and physiology between diploid Rangpur lime (*Citrus limonia*) and its autotetraploid are not associated with large changes in leaf gene expression. *J Exp Bot*. 2011; 62: 2507–2519. 10.1093/jxb/erq467.
15. Meng H, Jiang SS, Hua SJ, et al. Comparison between a tetraploid turnip and its diploid progenitor (*Brassica rapa* L.): the adaptation to salinity stress. *Agricultural Sciences in China*. 2011; 10: 363–375. 10.1016/S1671-2927(11)60015-1
16. Wang Z, Wang M, Liu L, Meng F. Physiological and proteomic responses of diploid and tetraploid black locust (*Robinia pseudoacacia* L.) subjected to salt stress. *International Journal of Molecular Sciences*. 2013; 14: 20299–20325. 10.3390/ijms141020299
17. Zhang XY, Hu CG, Yao JL. Tetraploidization of diploid *Dioscorea* results in activation of the antioxidant defense system and increased heat tolerance. *Journal Plant Physiology*. 2010; 167: 88–94. 10.1016/j.jplph.2009.07.006.
18. Zhang PG, Huang SZ, Pin AL, Adams KL. Extensive divergence in alternative splicing patterns after gene and genome duplication during the evolutionary history of *Arabidopsis*. *Mol Biol Evol*. 2008; 27: 1686–1697. 10.1073/pnas.1109551108
1. Xu G, Guo C, Shan H, Kong H, Divergence of duplicate genes in exon-intron structure. *Proc Natl Acad Sci USA*. 2012; 109: 1187–1192. 10.1073/pnas.1109047109
20. Ali Z, Park HC, Ali A, Oh DH, Aman R, Kropornicka A, Hong H, Choi W, Chung WS, Kim WY, Bressan RA, Bohnert HJ, Lee SY, Yun DJ. *TsHKT1;2*, a *HKT1* homolog from the extremophile *Arabidopsis* relative

TheLLungiella salsuginea, shows K specificity in the presence of NaCl. Plant Physiol. 2012; 158: 1463-1474. 10.1104/pp.111.193110

21. Arsovski AA, Pradinuk J, Guo XQ, Wang S, Adams KL. Evolution of Cis-Regulatory Elements and Regulatory Networks in Duplicated Genes of *Arabidopsis*. Plant Physiol. 2015; 169: 2982-2991. 10.1104/pp.15.00717
22. Wang S and Adams KL. Duplicate gene divergence by changes in microRNA binding sites in *Arabidopsis* and *Brassica*. Genome Biol Evol. 2015; 7: 646-655. 10.1093/gbe/evv023
23. Wang J, Marowsky NC, Fan C. Divergence of gene body DNA methylation and evolution of plant duplicate genes. PLoS One. 2014; 9: e110357. 10.1371/journal.pone.0110357.
24. Montanari F, Shields DC, Khaldi N. Differences in the Number of Intrinsically Disordered Regions between Yeast Duplicated Proteins, and Their Relationship with Functional Divergence. PLoS ONE. 2011; 6(9). 10.1371/journal.pone.0024989
25. Brown CJ, Johnson AK, Daughdrill GW. Comparing models of evolution for ordered and disordered proteins. Mol Biol Evol. 2010; 27: 609–21. 10.1093/molbev/msp277.
26. Nilsson J, Grahn M, and Wright AP. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. Genome Biol. 2011; 12: R65. 10.1186/gb-2011-12-7-r65.
27. Yruela I, Contreras-Moreira B, Dunker AK and Niklas KJ. Evolution of Protein Ductility in Duplicated Genes of Plants. Front. Plant Sci. 2018; 9:1216. 10.3389/fpls.2018.01216/full.
28. Rizzon C, Ponger L, Gaut BS. Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. PLOS Comput Biol. 2006; 2: e115. 10.1371/journal.pcbi.0020115
29. Navabi ZK, Huebert T, Sharpe AG, O'Neill CM, Bancroft I, Parkin IA. Conserved microstructure of the *Brassica* B genome of *Brassica nigra* in relation to homologous regions of *Arabidopsis thaliana*, *rapa* and *B. oleracea*. BMC Genomics. 2009; 14: 250. 10.1186/1471-2164-14-250.
30. Qiao X, Yin H, Li L, Wang R, Wu J, Wu J, Zhang S. Different Modes of Gene Duplication Show Divergent Evolutionary Patterns and Contribute Differently to the Expansion of Gene Families Involved in Important Fruit Traits in Pear (*Pyrus bretschneideri*). Front Plant Sci. 2018; 9: 161. 10.3389/fpls.2018.00161
31. Banerjee S, Feyertag F and Alvarez-Ponce D. Intrinsic protein disorder reduces small-scale gene duplicability. DNA Research. 2017; 24(4): 435–444. 10.1093/dnares/dsx015.
32. Mouillon JM, Gustafsson P, Harryson P. Structural investigation of disordered stress proteins: comparison of full-length dehydrins with isolated peptides of their conserved segments. Plant

- Physiol. 2006; 141: 638–650. 10.1104/pp.106.079848.
33. Blanc G, Wolfe KH. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis*. Plant Cell. 2004; 16: 1679-1691. 10.1105/tpc.021410
34. Chi YH, Koo SS, Oh HT, Lee ES, Park JH, Phan K, Wi SD, Bae SB, Paeng SK, Chae, HB, Kang CH, Kim MG, Kim WY, Yun DJ, Lee SY. The Physiological Functions of Universal Stress Proteins and Their Molecular Mechanism to Protect Plants From Environmental Stresses. Frontiers plant science. 2019; 10: 750. [3389/fpls.2019.00750](#)
35. Yruela I, Oldfield CJ, Niklas KJ, Dunker AK. Evidence for a strong correlation between transcription factor protein disorder and organismic complexity. Genome Biol Evol. 2017; 9: 1248–1265. doi: 10.1093/gbe/evx073. 10.1093/gbe/evx073.
36. Lysak, Martin & Cheung, Kwok & Kitschke, Michaela & Bures, Petr. Ancestral Chromosomal Blocks Are Triplicated in Brassiceae Species with Varying Chromosome Number and Genome Size. Plant physiology. 2007. 145. 402-10. 10.1104/pp.107.104380.
37. Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ, Vigouroux M, Trick M, Bancroft I. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell. 2006; 18: 1348-59. 10.1105/tpc.106.041665
38. Yang TJ, Kim JS, Kwon SJ et al. Sequence-level analysis of the diploidization process in the triplicated FLOWERING LOCUS C region of *Brassica rapa*. Plant Cell. 2006; 18(6): 1339-1347. 10.1105/tpc.105.040535
39. Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, Yu HJ, Kim DS, Kim N, Lim KB, Lee SI, Hahn JH, Lim YP, Bancroft I, Park BS. Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. GenomeBiol. 2009; 10: R111. 10.1186/gb-2010-11-9-r94
40. ZhaoQ, Zou J, Meng J, Mei S, Wang Tracing the Transcriptomic Changes in Synthetic Trigenomic Allohexaploids of *Brassica* Using an RNA-Seq Approach. Plos ONE. 2013; 11: 8(7): e68883. 10.1371/journal.pone.0068883
41. Kagale S, Robinson SJ, Nixon J, Xiao R, Huebert T, Condie J, Kessler D, Clarke WE, Edger PP, Links MG. Polyploid evolution of the Brassicaceae during the Cenozoic era. Plant Cell. 2014; 26: 2777-2791.
- 1105/tpc.114.126391
42. De Smet R, Sabaghian E, Li Z, Saeys Y, Van de Peer Y. Coordinated Functional Divergence of Genes after Genome Duplication in *Arabidopsis thaliana*. Plant Cell. 2017;29(11):2786-2800.

43. Van de Peer Y, Maere S, and Meyer A. The evolutionary significance of ancient genome duplications. *Nat Rev Genet.* 2009; 10: 725–732. 10.1038/nrg2600.
1. Vanneste K, Maere S, and Van de Peer Y. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philos T R Soc B.* 2014; 306: 20130353. 10.1098/rstb.2013.0353.
45. Soltis PS, and Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin Plant Biol.* 2016; 30: 159–165. 10.1016/j.pbi.2016.03.015.
46. Yruela I, and Contreras-Moreira B. Genetic recombination is associated with intrinsic disorder in plant proteomes. *BMC Genomics.* 2013; 14. 10.1186/1471-2164-14-772
47. Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci.* 2002; 27: 527–533. 10.1016/S0968-0004(02)02169-2.
48. Schad E, Tompa P, and Hegyi H. The relationship between proteome size, structural disorder and organism complexity. *Genome Biol.* 2011; 12: R120. 10.1186/gb-2011-12-12-r120
49. Yruela I and Contreras-Moreira B. Protein disorder in plants: a view from the chloroplast. *BMC Plant Biol.* 2010; 12:165. 10.1186/1471-2229-12-165
50. Dyson HJ and Wright PE. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol.* 2002; 12: 54–60. 10.1016/S0959-440X(02)00289-0
51. Dyson HJ and Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol.* 2005; 6: 197–208. 10.1038/nrm1589.
52. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* 2007; 6: 1882–1898. 10.1021/pr060392u
53. Habchi J, Tompa P, Longhi, S and Uversky VN. Introducing protein intrinsic disorder. *Chem Rev.* 2014; 114: 6561–6588. 10.1021/cr400514h
54. Van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. *Chem Rev.* 2014; 114: 6589–6631. 10.1021/cr400525m
55. O’Shea C, Jensen MK, Stender EGP, Kragelund BB, Willemoes M, Skriver K. Protein intrinsic disorder in *Arabidopsis* NAC transcription factors: transcriptional activation by ANAC013 and ANAC046 and their interactions with RCD1. *Biochem J.* 2015; 465: 281–294. 10.1042/BJ20141045.

56. Wright PE, and Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol.* 2015; 16: 18–29. [10.1038/nrm3920](https://doi.org/10.1038/nrm3920).
57. Yruela I. Plant development regulation: overview and perspectives. *J Plant Physiol.* 2015; 182: 62–78. [10.1016/j.jplph.2015.05.006](https://doi.org/10.1016/j.jplph.2015.05.006)
58. Covarrubias AA, Cuevas-Velazquez CL, Romero-Pérez PS, Rendón-Luna DF and Chater CCC. Structural disorder in plant proteins: where plasticity meets sessility. *Cell Mol Life Sci.* 2017; 74: 3119–3147. doi: [10.1007/s00018-017-2557-2](https://doi.org/10.1007/s00018-017-2557-2)
59. Balazadeh S, Siddiqui H, Allu AD, Matallana-Ramirez LP, Caldana C, Mehrnia M et al. A gene regulatory network controlled by the NAC transcription factor ANAC092/AtNAC2/ORE1 during salt-promoted senescence. *Plant J.* 2010; 62: 250–264. [10.1111/j.1365-313X.2010.04151.x](https://doi.org/10.1111/j.1365-313X.2010.04151.x).
60. Singh GP, Ganapathi M, Dash D. Role of intrinsic disorder in transient interactions of hub proteins. *Proteins.* 2007; 66: 761–765. [10.1002/prot.21281](https://doi.org/10.1002/prot.21281)
61. Kim PM, Sboner A, Xia Y, Gerstein M. The role of disorder in interaction networks: a structural analysis. *Mol Syst Biol.* 2008; 4: 179. [10.1038/msb.2008.16](https://doi.org/10.1038/msb.2008.16)
62. Podder S, Ghosh TC. Exploring the Differences in Evolutionary Rates between Monogenic and Polygenic Disease Genes in Human. *Mol Biol Evol.* 2010; 27(4): 934–941. [10.1093/molbev/msp297](https://doi.org/10.1093/molbev/msp297)
63. Dosztanyi Z, Csizmok V, Tompa P and Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 2005a; 21: 3433–4. [10.1093/bioinformatics/bti541](https://doi.org/10.1093/bioinformatics/bti541).
64. Dosztanyi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol.* 2005b; 347: 827–839. [10.1016/j.jmb.2005.01](https://doi.org/10.1016/j.jmb.2005.01)

Figures

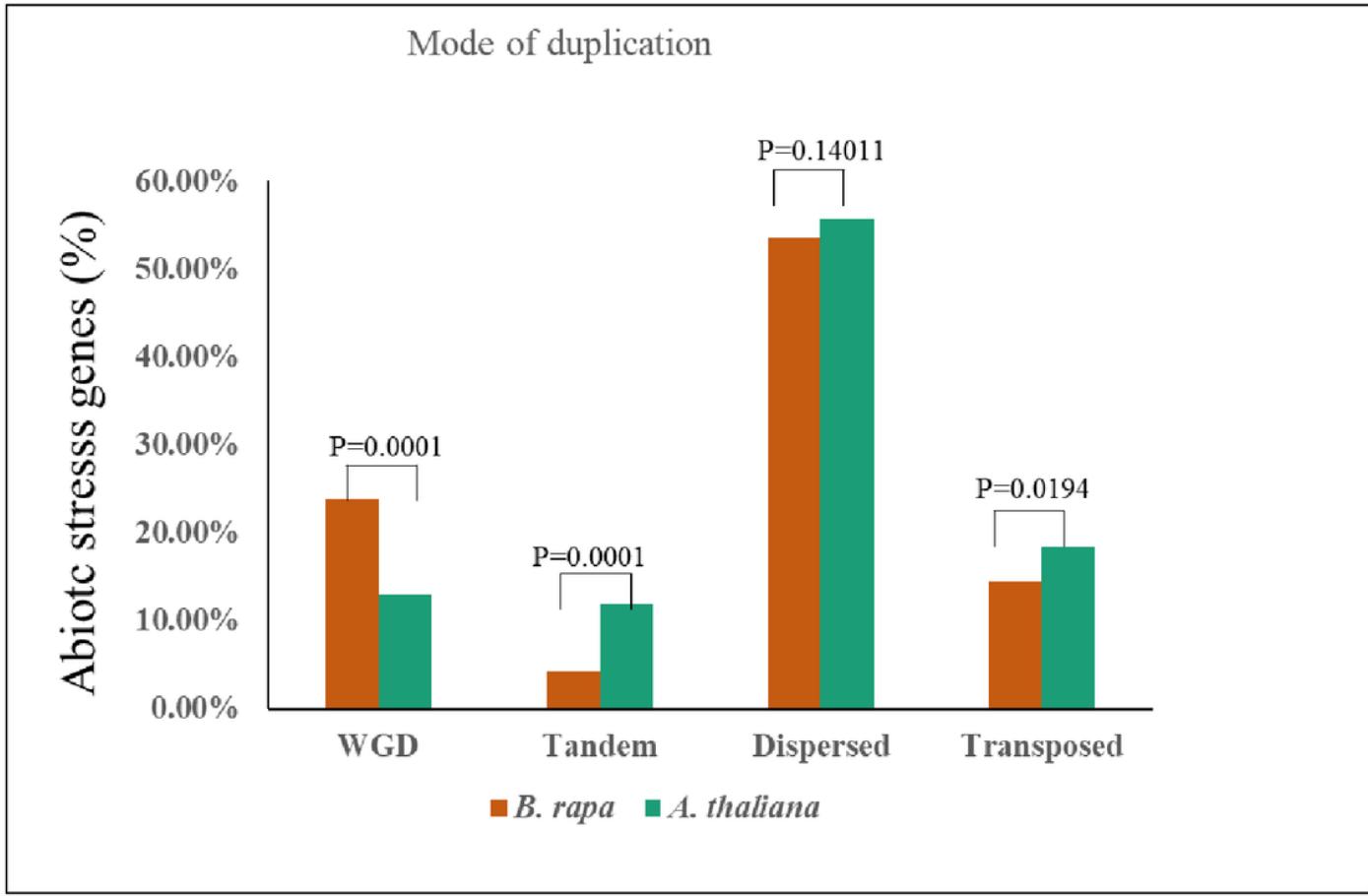


Fig.1

Figure 1

Percentage of abiotic stress genes in different modes of duplication in *A. thaliana* and *B. rapa*

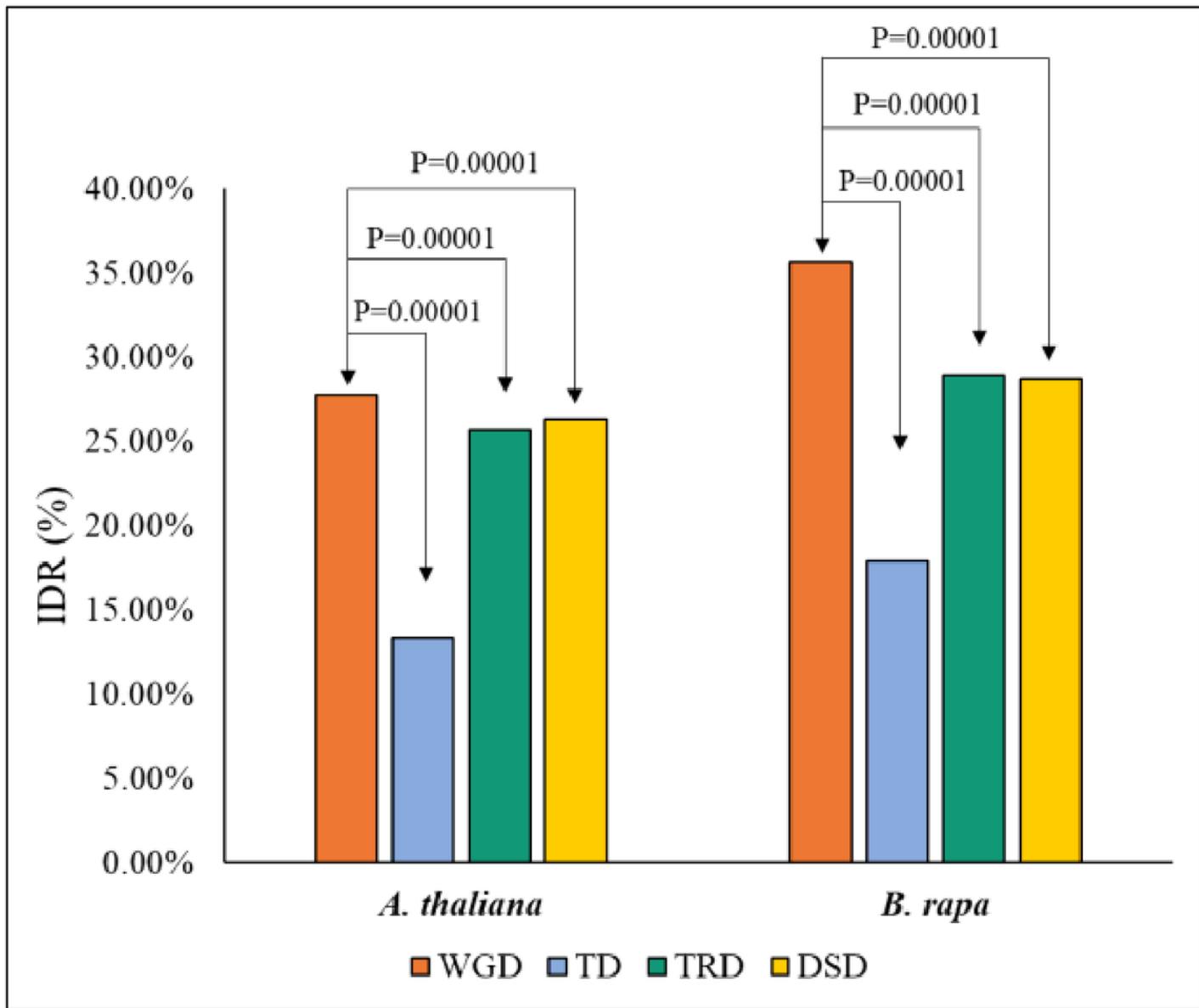


Fig.2

Figure 2

Percentage of intrinsic disorder in different modes of duplicated genes in *A. thaliana* and *B. rapa*

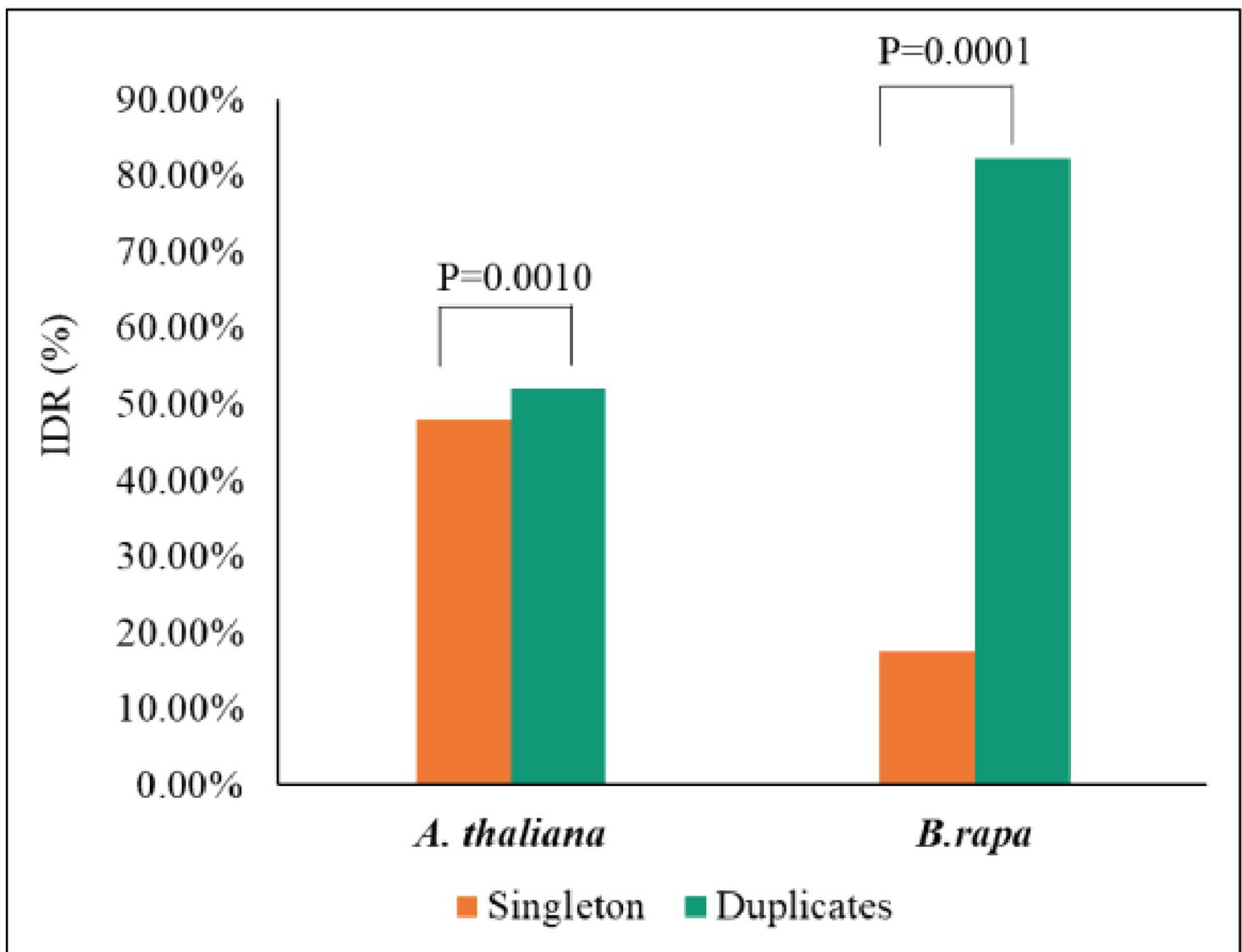


Fig. 3

Figure 3

Comparison of percentage of intrinsic disorder in singleton vs duplicated abiotic stress genes in *A. thaliana* and *B. rapa*.

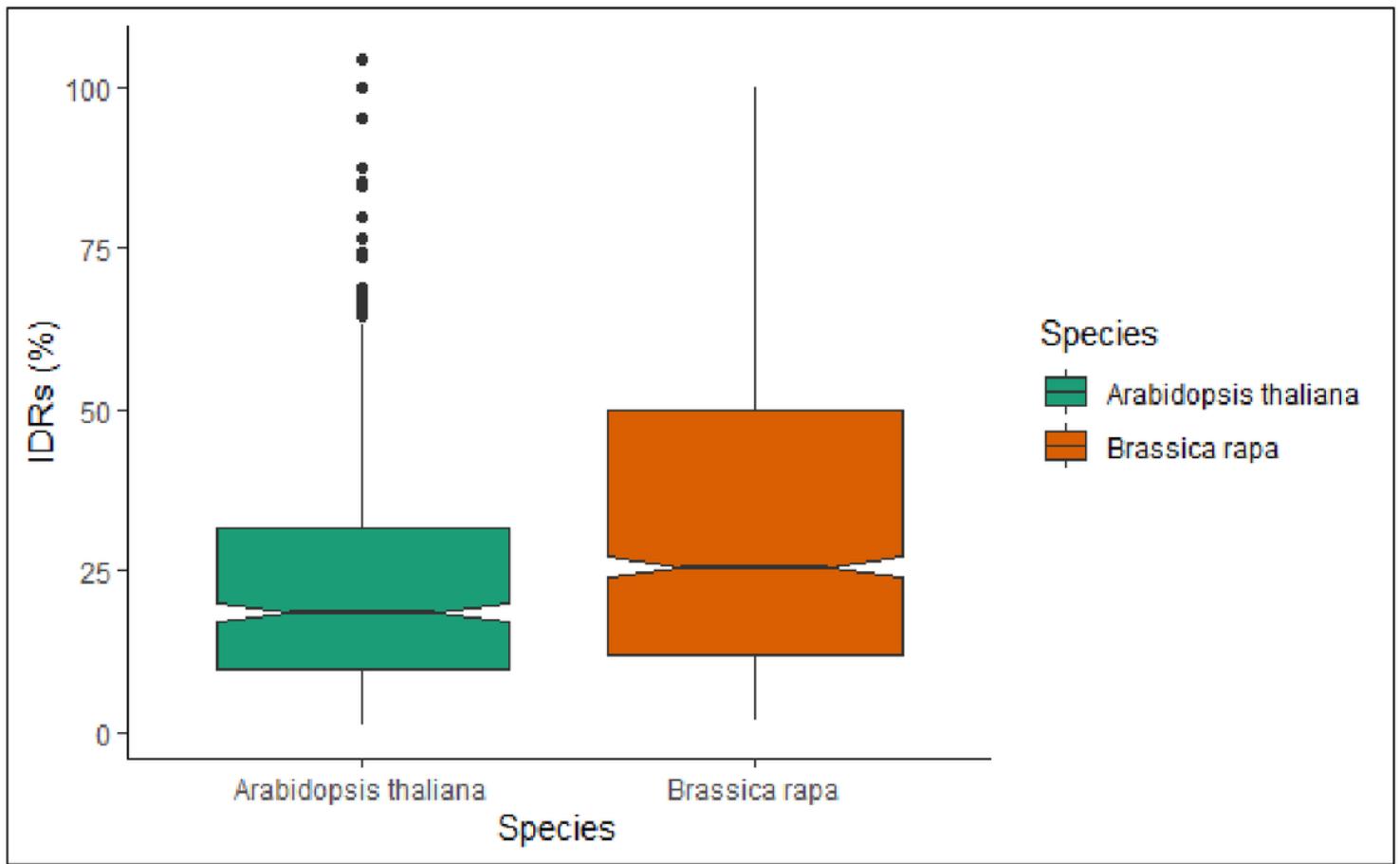


Fig. 4

Figure 4

Percentage of intrinsic disorder in abiotic stress genes in *A. thaliana* and *B. rapa*

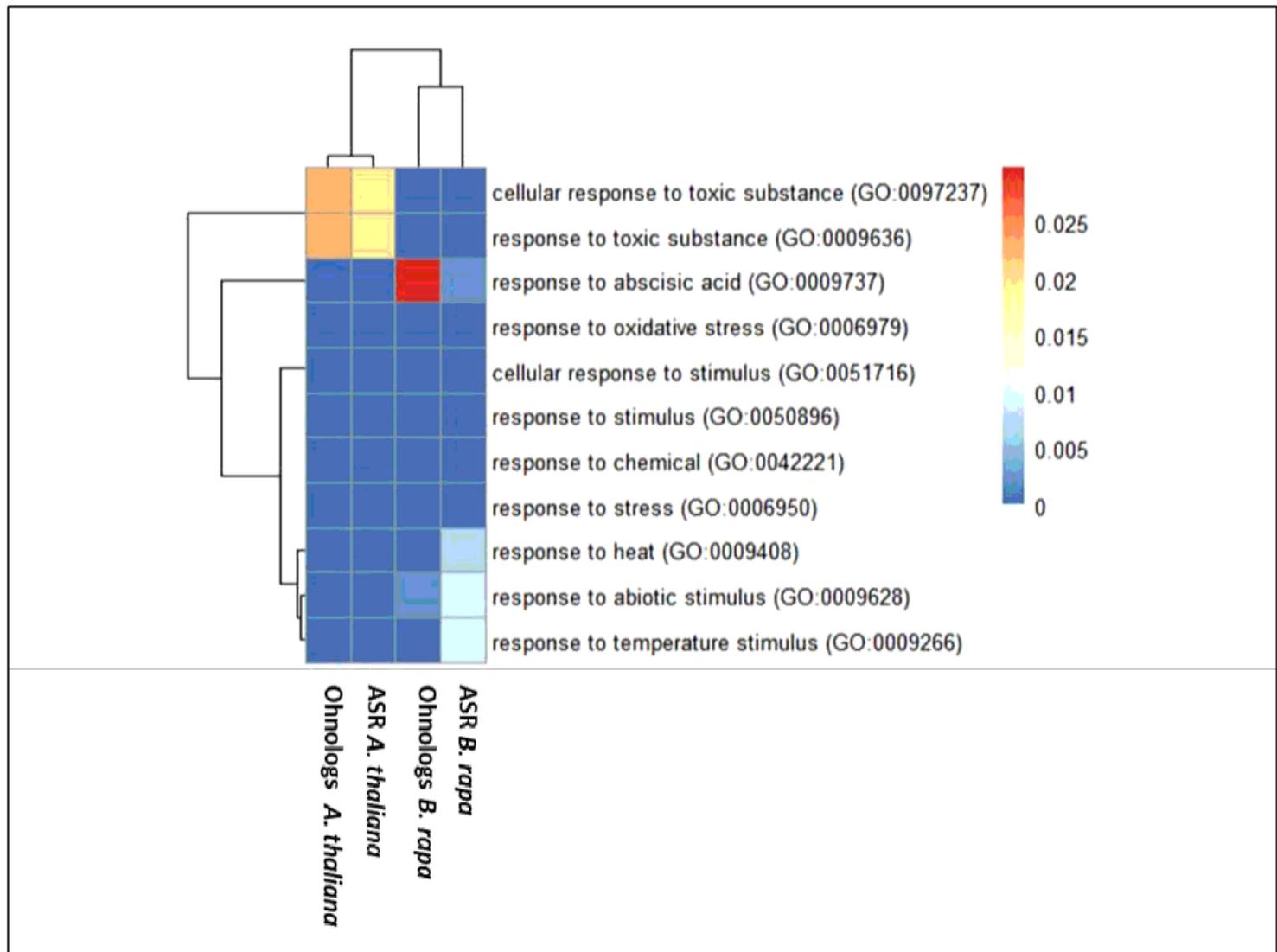


Fig. 5

Figure 5

Heatmap of functional enrichment analysis of ASR genes and their corresponding paralogs in *B. rapa* and *A. thaliana*.

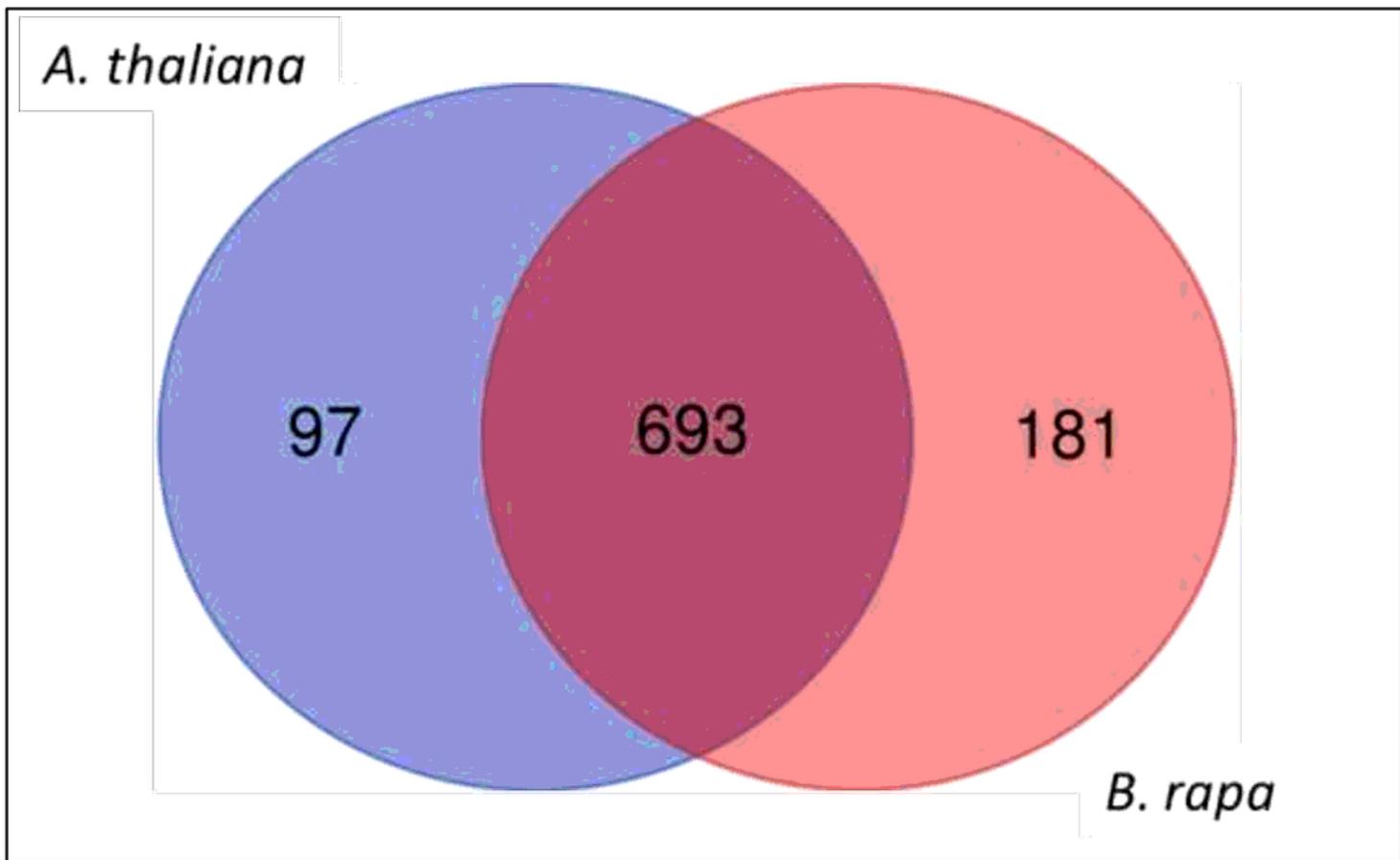


Fig. 6

Figure 6

Venn diagram of domain enrichment analysis WGD derived ASR genes in *B. rapa* an *A. thaliana*.

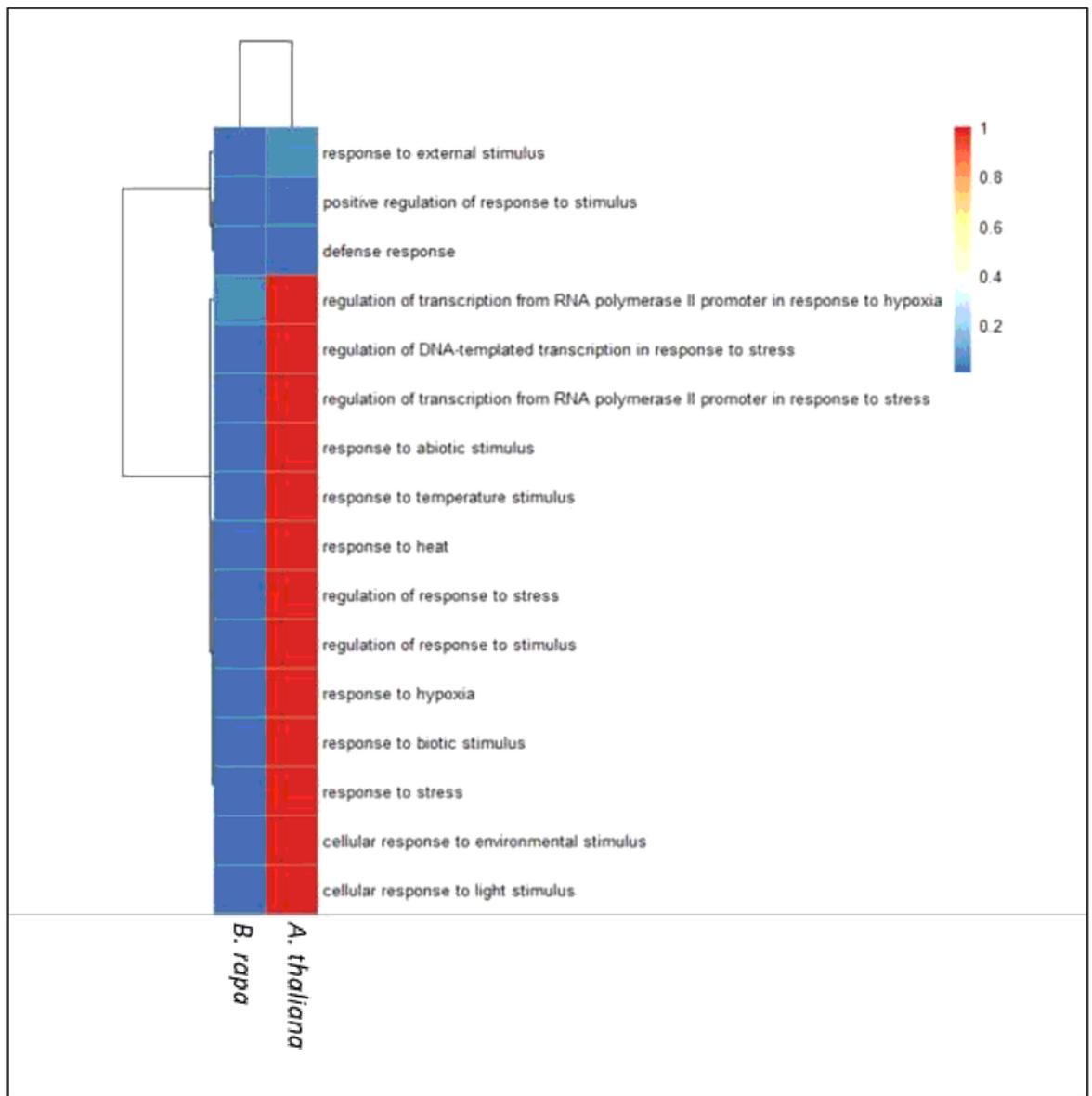


Fig.7

Figure 7

Heatmap of Functional enrichment analysis of unique domain specific to stress in *B. rapa* and *A. thaliana*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryfile1.xlsx](#)