Gene duplication and stress genomics in *Brassica*s: current understanding and future prospects

Shayani Das Laha[1], Smritikana Dutta[1], Anton R. Schäffner[2], Malay Das[1*]

[1]Department of Life Sciences, Presidency University, Kolkata, India

[2]Institute of Biochemical Plant Pathology, Department of Environmental Sciences, Helmholtz Zentrum München, München, Germany

*Correspondence:

Malay Das

Plant Genomics Laboratory,

Department of Life Sciences,

Presidency University,

86/1 College Street, Kolkata- 700073

E. mail: malay.dbs@presiuniv.ac.in

Keywords: polyploidy, genome evolution, stress response, expression divergence, microRNA, phenotyping, genotyping

Abstract

Polyploidy or whole genome duplication (WGD) is an evolutionary phenomenon that happened in all angiosperms multiple times over millions of years. Extensive studies on the model plant *Arabidopsis thaliana* genome have revealed that it has undergone five rounds of WGDs followed, in the Brassicaceae tribe, by a characteristic whole genome triplication (WGT). In addition, small-scale events such as tandem or segmental duplications and retrotransposition also enable plants to reshape their genomes. Over the decades, extensive research efforts have been undertaken to understand the evolutionary significance of polyploidy. On the other hand, much less attention has been paid to understanding the impact of gene duplication on the diversification of important stress response genes. The main objective of this review is to discuss key aspects of gene and genome duplications with a focus on genes primarily regulated by osmotic stresses. The focal family is the Brassicaceae, since it (i) underwent multiple rounds of WGDs plus WGTs, (ii) hosts many economically important crops and wild relatives that are tolerant to a range of stresses, and (iii) comprises many species that have already been sequenced. Diverse molecular mechanisms that lead to structural and regulatory alterations of duplicated genes are discussed. Examples are drawn from recent literature to elucidate expanded, stress responsive gene families identified from different *Brassica* crops. A combined bioinformatic and transcriptomic method has been proposed and tested on a known stress-responsive gene pair to prove that stress-responsive duplicated allelic variants can be identified by this method. Finally, future prospects for engineering these genes into crops to enhance stress tolerance are discussed, and important resources for *Brassica* genome research are provided.

1. Introduction

Polyploidy or whole genome duplication (WGD) has been identified as a major driving force behind the evolution and diversification of all land plants (Jiao et al., 2011, Panchy et al., 2016, Landis et al., 2018, Oh and Dassanayake, 2019). Extensive phylogenomic analyses have identified the occurrence of at least five rounds of WGD events leading to the origin of the reference plant *Arabidopsis thaliana*. Two of these five WGDs have been shared with other angiosperms. The most ancient polyploidy, known as paleopolyploidy, happened approximately 340 million years ago (MYA) in the common ancestors of all seed plants, and was followed by another at 170-235 MYA in the angiosperm lineage (Fig. 1, Jiao et al., 2011). Subsequently, three additional WGDs known as α/3R, β/2R and γ/1R led to the evolution of *Arabidopsis* (Bowers et al., 2003, Jiao et al., 2011). Initial studies focusing only on the *Arabidopsis* genome estimated that α duplication took place 14.5-20.4 MYA, β 170-235 MYA and γ 300 MYA (Fig. 1, Bowers et al., 2003). However, a recent study involving extensive taxonomic sampling and sequence data estimated a more accurate timing of duplications. According to this study the α duplication happened approximately 47 MYA and the β event took place 124 MYA (Fig. 1, Kagale et al., 2014). The *Brassica* lineage that includes many important crop species separated from *Arabidopsis* approximately 43.2 MYA and from *Thellungiella* 12 MYA (Oh et al., 2014) and finally underwent a genome triplication (WGT) approximately 7.9-14.6 MYA (Lysak et al., 2005). Following this WGT the first diploid *Brassica* species that arose was *B. nigra* (BB) 8 MYA, followed by the split of *B. rapa* (AA) and *B. oleracea* (CC) from their shared ancestor 3.75 MYA (Fig.1, Lysak et al., 2005, Navabi et al., 2013). These diploid *Brassica*s underwent natural hybridization among themselves and many new allopolyploids such as *B. napus* (AACC), *B. juncea* (AABB) and *B. carinata* (BBCC) evolved 0.75 MYA (Fig. 1, Mun et al., 2009). Generation of extensive sequence data from all these species provided opportunities for comparative genomics to characterize the newly sequenced genomic blocks by studying synteny and collinearity to the completely sequenced genome of *Arabidopsis* (Jiao and Schneeberger, 2020).

Although the repeated occurrence of WGDs remains the primary contributor of the majority of the duplicated genes, smaller scale sub-genomic duplications such as tandem duplication (Qiao et al., 2019), transposable element mediated duplication (Vicient and Casacuberta 2017), segmental duplication (Kuo et al., 2019), and retroduplication (Kim et al., 2017) also significantly enriched the duplicated gene pool (Panchy et al., 2016). Regardless of the mechanism of duplication, most of these duplicated genes were lost over time and this loss was correlated with the age of the duplications.

Therefore, the retention rate of α duplicates is much higher (16.3%) than that of γ duplicates (4.4%, Maere et al., 2005). Only a few duplicates were kept in the genome mainly because they conferred adoptive advantages, although the evolutionary disadvantages of polyploidy, such as reproductive isolation and lower fertility rates, have also been pointed out (Ramsey and Ramsey 2014). The existing duplicated genes can lead to functional innovation *via* different routes such as neo-functionalization (acquiring a new gene function), sub-functionalization (splitting ancestral gene functions), genetic redundancy (sharing gene functions), or, ultimately, loss of gene functions by non-functionalization/pseudogenization. Studies have been conducted to test whether the mechanism of gene duplication, i.e. WGD *vs*. sub-genomic duplications, has any impact on the types of functional categories of genes that have been retained. Indeed, genes related to signal transduction, transcriptional regulation and ribosome structures were enriched among WGD categories (Blanc and Wolf 2004, Maere et al., 2005), whereas tandem (Zhang et al., 2016) and segmental (Dossa et al., 2016) duplicates were mostly responsive to environmental stresses. For instance, in *Arabidopsis* 27% of the tandem, 26% of the segmentally duplicated, and 40% of the transposable element-mediated duplicates are responsive to abiotic stresses (Wu et al., 2012).

Over the last few decades extensive research emphasis has been given to understanding the ecological and evolutionary consequences of gene duplication (see above). New findings indicate a possible link between polyploidy and enhanced stress tolerance as well as reproductive advantage under stress conditions (Chao et al., 2013, Ramsey and Ramsey, 2014). For example, the autotetraploid *Arabidopsis thaliana* accessions, e.g., Bla-5 (from Blanes, Spain) and Ciste-2 (from Cisterna di Latina, Italy) had elevated levels of potassium (K) and its analogue rubidium (Rb) in the leaves, and also demonstrated enhanced salinity tolerance compared to the diploids (Chao et al., 2013). These findings highlight the potential of duplicated genes as a 'repository' for new stress-regulated functions. The main objective of this review article is to discuss the mechanisms, operating at the gene structural and regulatory levels, by which duplicated genes can acquire functional divergence. The family Brassicaceae provides a perfect platform for this discussion, since it bears extensive signatures of large scale WGD and WGT as well as of local genome duplications and rearrangements. Furthermore, it hosts many economically important crop species and many wild species/germplasms possessing diverse stress tolerance properties, and enormous amounts of sequence data as well as gene mutant lines are available. Therefore, by using examples of known stress-responsive genes in the family Brassicaceae we will explore how genomic, transcriptomic and genetic methods can be combined to identify new allelic variants that are responsive to stresses. Finally, important research questions and possibilities that are

emerging in this area are discussed. Together, this article summarizes our current understanding of abiotic stress-regulated duplicated genes in the Brassicaceae and renders concepts that can be implemented for potential crop improvement.

2. Current status of abiotic stress research on *Brassica*

The genus *Brassica* comprises approximately 100 species, including major food crops such as rapeseed (*Brassica napus* L.), mustard (*Brassica juncea* L.), turnip rape (*Brassica rapa* L.), cabbage (*B. oleracea* L. var. *capitata*), cauliflower (*B. oleracea* var. *botrytis*), kale (*B. oleracea* var. *acephala*), broccoli (*B. oleracea* var. *italica*), brussels sprouts (*B. oleracea* var. *gemmifera*) and kohlrabi (*B. oleracea* var. *gongylodes* (Zhang et al., 2014, Beacham et al., 2017). Abiotic stress severely affects agricultural crop productivity world-wide, and *Brassica* is no exception (Abou-Hussein, 2012). For instance, drought and high salinity stresses have been shown to result in reduced biomass, seed germination, and shoot and root growth in *B. napus* and *B. oleracea* (Hadi et al., 2014, Rodríguez et al., 2015). Phenotypic variations with respect to salt stress sensitivity have been assessed by growing plants under controlled growth conditions (Siddiqui et al., 2009) with special emphasis on early growth phenotypes such as seed germination and seedling growth rates (Sharma et al., 2013). However, few studies have been conducted under field conditions (Zou et al., 2014, Das Laha et al., 2020). Most of these studies revealed that the germplasms which were predicted to be tolerant in laboratory and greenhouse assays may not be similarly tolerant in field settings (Zou et al., 2014, Das Laha et al., 2020).

Putatively tolerant species/varieties were assessed at the biochemical (Srivastava et al., 2015[a]), genome (Chikkaputtaiah et al., 2017), transcriptome (Bhardwaj, 2015, Srivastava et al., 2015[b], Guo et al., 2017) and proteome levels (Yousuf et al., 2016; Bandehagh et al., 2011), to identify the key determinants underlying salinity tolerance. In addition, genes encoding well-characterized salt stress pathways, such as *SALT OVERLY SENSITIVE 1* (*SOS1)*, *SOS2* and *SOS3*, have been sequenced and characterized by their level of induction as a consequence of increasing salt stress treatments, to infer their possible role in salt tolerance (Chakraborty et al., 2012). Such stress-regulated genes have been identified and functionally characterized in *Brassica rapa* (Kim et al., 2015, Krishnamurthy et al., 2019), *B. rapa* ssp. *oleifera* (Mooney et al., 2019), *B. rapa* ssp. *chinensis* (Wang et al 2018), *B. napus* (Zhao et al., 2016, Lu et al., 2019), *B. campestris* (Zhu et al., 2018, Zhang et al., 2018) and *B. juncea* (Nutan et al., 2018), and subsequently overexpressed in reference plants *Arabidopsis*, *Oryza sativa* or *Nicotiana tabacum*, (Table 1).

3. Protein coding sequence alterations: exon-intron structure, single nucleotide change, alternative splicing

Post duplication gene function divergence can be accompanied by structural changes either in protein coding regions or in regulatory regions such as promoters and small RNA binding sites. A comparative study of 612 pairs of paralogs and 300 pairs of one-to-one orthologs revealed that changes such as exon/intron gain/loss, exonization/pseudo-exonization and insertion/deletion indeed occur more frequently in the duplicated genes than in single copy orthologs (Xu et al., 2012). Large-scale sequence alterations may affect domain organization or similar sequence features and thereby influence the ultimate protein functionality (Xu et al., 2012, Wang et al., 2013). There are instances where single nucleotide substitutions have resulted in enhanced stress responsiveness with one homolog, but not with the others (Ali et al., 2012). For instance, the *HIGH-AFFINITY* K^+ *TRANSPORTER 1* (*HKT1*) gene homolog *Ts.HKT1;2* (*Ts6g08730*) of the salt tolerant species *T. salsuginea* underwent two amino acid changes compared to the susceptible species *A. thaliana* (D207N, D238N). These changes enabled *Ts.HKT1;2* to perform enhanced ion selectivity and transmembrane transport.

Differential retention of spliced transcript variants are observed for many stress-responsive WGD and tandem duplicates (Zhang et al., 2010, Ding et al., 2014, Kannan et al., 2018). For instance, *RADIATION SENSITIVE23 RAD23* (*AT1G16190*, *AT1G79650*) is a pair of WGD-derived DNA repair genes where *AT1G16190* produced only the intron-retained variant when the plants were exposed to heat and drought stresses, whereas the non-intron retaining form was obtained under regular growth conditions. Considering *AT1G79650,* however, the non-intron retaining form was obtained under heat and drought stress conditions (Zhang et al., 2010). Such variation in spliced copies was also noticed for the tandemly duplicated, stress-responsive gene pair *AT1G78380* and *AT1G78370*. There are instances of tandem duplicates where alternative splicing caused premature insertion of stop codons resulting in complete or partial loss of important domains. For instance, *SULPHATE TRANSPORTER 1 SULTR1;2* (*AT1G78000*) and *SULTR2;2* (*AT1G77990*) are a pair of tandem genes encoding sulfate transporter proteins. *SULTR1;2* retained the last intron resulting in premature insertion of a stop codon and thereby causing non-functionalization of the gene. On the other hand, *SULTR2;2* maintains regular protein functionality (Zhang et al., 2010). Such differences in spliced isoforms generated by duplicated genes can be due to differential gain and loss of splicing related elements such as exonic splicing enhancers and suppressors in different paralogous gene copies (Zhang et al., 2009). It has been proven that alternative splicing can aid to combat low- and high-temperature stresses by increasing the splicing rates of several stress responsive genes such as *PASTICCINO2* in *B. napus* (Zhou et al. 2011) and

*HEAT STRESS-TRANSCRIPTION FACTOR*s (*HSF*s) in *Boechera depauperata* (Kannan et al., 2018). Stress exposure may regulate alternative splicing by degrading the splicing machinery (Ding et al. 2014, Ling et al. 2017).

4. Regulatory sequence alterations: promoters and small RNA binding sites

There is growing evidence that duplicated genes can diverge in expression patterns when differential regulatory changes between a pair of duplicated genes are produced by alterations in *cis* elements (Oh et al., 2014, Arsovski et al., 2015), small RNA binding sites (Wang and Adams 2015), and methylation patterns (Wang et al., 2014a). Regulatory neofunctionalization occurs if one duplicated copy maintains the ancestral expression pattern, while the other acquires a new pattern. Gain of diverse *cis* regulatory elements in the promoter regions of *NUCLEAR FACTOR Y* (*NF-Y*, Xu et al., 2014), *LEA* (Liang et al., 2016), *SOD* (Verma et al., 2019), *PYRABACTIN RESISTANCE 1 LIKE* (*PYL*, Cheng et al., 2019), *HOMOLOGOUS TO E6-AP C TERMINUS* (*HECT*, Alam et al., 2019), *CYTOKININ RESPONSE FACTORS* (CRFs, Powell et al., 2019), *AP2/ERF* (Wang et al., 2019), L-myo-inositol 1-phosphate synthase or D-myo-inositol-3-phosphate synthase (MIPS, Hazra et al., 2019), and *LATERAL ORGAN BOUNDARIES DOMAIN* (LBD, Xie et al., 2020) genes resulted in abiotic stress tolerance in different *Brassica* species. In yeast a very high number of TATA boxes were gained after duplication and the TATA-gain duplicates underwent higher expression divergence, particularly under environmental stress conditions, than the TATA-less (ancestral) genes (Zou et al., 2011[a]). In the event of regulatory subfunctionalization, the expression patterns of the two duplicated gene copies are split, e.g. occur in different organs. Alteration in *cis* elements and translocation in the upstream region causes sub functionalization of *NHX8, BOR5* and *AVP1, SOS1* duplicates, and thereby provide elevated lithium, boron and salt tolerance in *T. salsuginea* and *T. parvula* (Oh et al., 2014).

Apart from promoter divergence, gene expression can be also affected in trans due to the action of small and non-coding RNAs at the transcriptional or post-transcriptional level. Indeed, duplicated genes can diverge in their expression and function due to differences in their microRNA binding sites (Wang and Adams, 2015). Out of a total of 6858 pairs of paralogous *A. thaliana* genes originating from WGD, tandem and other duplication mechanisms, 92% were targeted by miRNAs, whereas only 8% of the singleton genes harbored microRNA target sites. Moreover, the majority of the paralogous genes undergo divergence in their microRNA binding sites and also demonstrate divergence in their expression patterns. A similar trend was observed for the WGT-derived genes studied in *B. rapa*. This clearly indicates that microRNAs play a very important role in the functional diversification of

duplicated genes in polyploid genomes. The identification of new microRNAs in *Brassica* (Sun et al., 2015) and *Thellungiella* (Zhang et al., 2013), along with the availability of transcriptome arrays (Lee et al., 2013, Huang et al., 2014), will facilitate future investigations to characterize stress-regulated duplicated genes and their possible regulation by miRNAs and siRNAs.

5. Preferential retention of duplicated genes contributes considerably to the extremophilic nature of wild Brassicaceae

A wide range of tolerance against various stresses exists in the family Brassicaceae. Two such important plants are the halophytic species *Thellungiella salsuginea* (= *Eutrema salsugineum;* Wu et al. 2012) and *T. parvula* (= *Schrenkiella parvula;* Dassanayake et al., 2011) as well as the heavy metal tolerant plant *A. halleri* (Briskine et al., 2016). *T. salsuginea* and *T. parvula* have emerged as model plants for studying salt tolerance mechanisms in plants. These two species split from *A. thaliana* approximately 7-12 MYA (Wu et al., 2012). Although *A. thaliana* and *T. salsuginea* share 95% genomic similarity, they are very different with respect to their sensitivity towards salt stress (Pang et al., 2010, Mishra and Tanna, 2017). Both *Thellungiella* species are capable of sustaining very high salt concentrations, of up to 500 mM (Uzilday et al., 2015). The total number of duplicated genes in *T. salsuginea* (38%) is marginally higher than that of *A. thaliana*, i.e. 36% (Yang et al., 2013). The frequency of diverse duplication categories in *T. salsuginea* is quite comparable to that of *A. thaliana.* For example, tandem, segmental and transposon mediated duplications in *A. thaliana* and *T. salsuginea* are 9.8% *vs.* 9.5%, 30% vs. 28% and 3% vs. 4.8%, respectively (Wu et al., 2012). A total of 813 out of 2723 tandemly duplicated genes (30%) and a total of 2551 out of 8713 segmentally duplicated genes plus transposable elements (30%) that are present in *T. salsuginea* are responsive to abiotic stresses (Wu et al., 2012). In *A. thaliana* the numbers are marginally lower, *i.e.*, 737 out of 2708 tandemly duplicated genes (27%) and 2301 out of 8782 segmentally duplicated genes plus transposable elements (26%) are abiotic stress responsive (Wu et al., 2012). Strikingly, the genome of *T. salsuginea* shows expansion of important stress-inducible genes and families such as K^+ *EFFLUX ANITIPORTER1* (*KEA1*), K^+ *UPTAKE PERMEASE9* (*KUP9*), *ARABIDOPSIS VACUOLAR* H^+*PYROPHOSPHATASE1* (*AVP1,* Oh et al., 2014) *CYTOCHROME P450, FAMILY 707, SUBFAMILY A* (*CYP707A*), *HIGH-AFFINITY* K^+ *TRANSPORTER 1* (*HKT1*), *CYCLIC NUCLEOTIDE GATED CHANNEL* (*CNGC*), *ALPHA CARBONIC ANHYDRASE* (*ACA*), , *BETA GLUCOSIDASE (BGL*), *CBL- INTERACTING PROTEIN KINASE* (*CIPK*), *CALCIUM DEPENDENT PROTEIN KINASE* (*CDPK*), *NUCLEAR TRANSCRIPTION FACTOR X-BOX BINDING 1 (NF-X1), GIBBERELLIC ACID INSENSITIVE*

REPRESSOR OF GA1-3 – RGA LIKE 1-SCARECROW (*GRAS*), *HSF, RELATED TO ABI3/ VP1* (*RAV*) and *SALT TOLERANCE 32* (*SAT32*). Similarly, many important stress genes such as *HKT1*, *CALCINEURIN B-LIKE 10* (*CBL10*), *Na^+/H^+ EXCHANGER 8* (*NHX8)* and *MYB DOMAIN PROTEIN 47* (*MYB47*) have been multiplicated in *T. parvula* (Dassanayake et al., 2011), but not in *A. thaliana*.

Arabidopsis halleri is a heavy metal hyperaccumulator for cadmium and zinc (Briskine et al., 2016). *A. halleri* and *A. lyrata* evolved from a common ancestor approximately 8 MYA (Schlaeppi et al., 2014). Molecular breeding studies in these plants have highlighted two proteins HEAVY METAL ATPASE 4 (HMA4) and METAL TOLERANCE PROTEIN 1 (MTP1) that control metal tolerance (Courbot et al., 2007, Willems et al., 2007). The genes encoding these proteins are multiplicated in the *A. halleri* genome: four copies of *HMA4* (Hanikenne et al., 2008) and five copies of *MTP1* gene have been identified (Shahzad et al., 2010). Such copy number expansion combined with *cis* regulatory modifications resulted in elevated transcriptional expression of *AhHMA4*, a possible mechanism for *A. halleri* metal hyper accumulation.

6. Genome duplication and triplication contributes to the diversification of stress genes in *Brassica* crops

The enrichment of stress-responsive, tandemly duplicated genes is much lower in *Brassica* than in *Thellungiella*. Approximately 3% and 4% of tandemly duplicated genes are responsive to abiotic stresses in *B. rapa* and *B. oleracea*, respectively (Liu et al., 2014). However, the occurrence of WGT in *Brassica* species added extra leverage for generating new allelic variations. Indeed, genes that respond to environmental stresses were preferentially retained after WGT in *B. rapa* (Fang et al., 2012). The Gene Ontology annotation classes of over-represented genes in the sequenced *B. rapa* genome show strong bias towards various abiotic and biotic stresses that include salt, cold, and pathogen stresses (Wang et al., 2011). In addition, recent studies demonstrated that the amphidiploid species *B. juncea, B. napus,* and *B. carinata* are more tolerant to stresses than diploids such as *B. oleracea, B. nigra*, and *B. rapa* (Purty et al., 2008). As a result of WGD, WGT, and tandem and segmental duplications, the number of members of many multigene families increased in different *Brassica* crops and were responsive to diverse osmotic stresses (Table 2). Some, such as *DEHYDRATION RESPONSE ELEMENT BINDING PROTEIN* (*DREB1*, Lee et al., 2012), *TRANSCRIPTION FACTORS OF CONSERVED MOTIF (TIF[F/Y]XG*, (*TIFY*, Saha et al., 2016) *LATE EMBRYOGENESIS ABUNDANT* (*LEA*, Liang et al., 2016), *SUPEROXIDE DISMUTASE* (*SOD*, Verma et al., 2019), *ANNEXIN* (*ANN*, Yadav et al., 2015), *SALT OVERLY SENSITIVE* (SOS, Cheng et al., 2019), and *PHOSPHOLIPASE D*

(*PLD*, Lu et al., 2019), were already known to be stress-regulated, whereas new candidates such as *HOMOLOGOUS TO E6-AP C TERMINUS* (*HECT*, Alam et al., 2019), *LIN-11, ISL-1 AND MEC-3 DOMAIN GENE* (*LIM*, Park et al., 2014), *CCCH ZINC FINGER FAMILY* (*CCCH*, Pi et al., 2018), *SET DOMAIN GROUP* (*SDG*, Dong et al., 2015), and *SUGARS WILL EVENTUALLY BE EXPORTED TRANSPORTER* (*SWEET*, Wei et al., 2019) have also emerged in recent years. Eight of these studies were done on *B. rapa*, three on *B. juncea* and two on *B. napus* (Table 2). Most were published within the last two years, which indicates that the release of additional *Brassica* genome sequences will foster more such findings in the near future.

7. Future strategies to identify potential stress-regulated allelic variants

New allelic variants relating to diverse stresses are important for the engineering of stress-tolerant plants. One potential avenue is to characterize stress-responsive genes that are retained as multiple copies in tolerant species such as *Thellungiella* or *A. halleri*, while present only as a single copy in the susceptible *A. thaliana*. Bioinformatics tools such as OrthoMCL (Li et al., 2003) can be implemented to compare susceptible and tolerant genomes to identify such cases, in which one *Arabidopsis* gene relates to multiple *Thellungiella* genes. Even though OrthoMCL employs a very conservative reciprocal best BLAST hit (RBH) search, each query gene from *Arabidopsis* will return multiple *Thellungiella* genes, resulting in hundreds of gene matches in total (Wu et al., 2012). In a comparative study between *Arabidopsis thaliana* and its close relative *A. lyrata*, 2850 gene clusters were found where either one-to-many or many-to-many situations exist (Das et al., 2016). To identify the best possible *A. lyrata* homolog to the queried *A. thaliana* gene, overall promoter divergence scores (dSM) were calculated between the *A. thaliana* and all *A. lyrata* homologs present in one Ortho MCL cluster. In 78% of the cases such an analysis could correctly predict one *A. thaliana* - *A. lyrata* gene pair that maintained a conserved expression pattern, whereas the other member/s of the same gene cluster did not (Das et al., 2016). Thus, the other member(s) might have lost their function and/or adopted novel one(s), and could be of interest in terms of different stress resistance.

The strategy used by Das et al. (2016) could be further developed and refined. Instead of taking into account the sequence similarity over the entire promoter length, we focussed on specific regions of promoters, i.e. stress-linked *cis* elements that primarily contribute to a gene's response towards stress. We hypothesized that the enrichment of stress-responsive *cis* elements could be used as a criterion to predict *Thellungiella* homologs that are highly stress-responsive as compared to the homologous *Arabidopsis* gene (Fig. 2). To test this hypothesis, we selected examples of *A. thaliana* (*At4g10310*)

and *T. salsuginea* (*TsHKT1;1- Ts6g08650, TsHKT1;2- Ts6g08730, TsHKT1;3- Ts6g08740*) genes. We analyzed 1000 bp upstream regions of these four genes for the detection of putative stress-responsive *cis* elements by using the New PLACE database (https://www.dna.affrc.go.jp/PLACE/?action=newplace). ~~Plant care database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html)~~. The *TsHKT1;2* allele contains forty ~~twelve~~ osmotic (salt/drought/heat/cold) stress-responsive *cis* elements, which are members of seven different *cis* elements families. ~~which are ABA-responsive elements (*ABRE*), heat shock elements (*HSE*), a MYB binding sites (*MBS*), and *TC* rich repeats.~~ This enrichment is much higher than that of *A. thaliana*, which has thirty ~~four~~ elements, and *TsHKT1;1* and *TsHKT1;3*, which have twenty and seventeen, respectively ~~two each~~ (Fig. 2, Supplementary Fig. 1). In order to obtain further significance of these elements in osmotic stress, all available literature were mined to identify candidates, which could be functionally associated with stress responses by mutational analyses (Supplementary Table 1). Again, this analysis supports the hypothesis of enhanced stress sensitivity of *TsHKT1;2* in comparison to the other two homologs. Therefore, we predict that *TsHKT1;2* could be a potential allelic variant possessing an enhanced stress-responsive property. ~~It has been previously found that the position and combination of *cis* elements in the promoter region plays a crucial role in controlling the gene expression level (Zou et al. 2011[b]). In our study 8 out of 12 predicted *cis* elements in *TsHKT1;2* are located in the core promoter region, i.e., within 500 bp upstream, suggesting their active involvement in stress responses.~~

8. *In planta* assay to assess the utility of the identified allelic variants in crop improvement programs

To complement predictions based on *cis* elements, transcriptomic studies need to be performed to identify homologs that are upregulated or downregulated by stress applications in the tolerant species, but not in the susceptible species. The tissue-specific and stress-induced gene expression similarities between *A. thaliana* and *A. lyrata* homologs were successfully used to predict their possible functional fate (Das et al., 2016). The homologs that shared similar expression patterns were called ‘expressologs’ and those that differed in their expression were called ‘non-expressologs’ (Patel et al., 2012). The complementation of *A. thaliana* loss-of-function mutants with *A. lyrata* homologs could experimentally prove that expressologs are indeed functional orthologs, whereas non-expressologs are not (Das et al., 2016). In order to identify potential stress tolerance-related genes one could particularly focus on those stress-induced non-expressologs, which are enriched with stress-responsive *cis* elements in their

promoters and also show upregulation after stress application (Fig. 3). Thus *TsHKT1;2,* for example, emerged as a tentative candidate from the *cis* element prediction (see above). Indeed, the transcriptional expression of *TsHKT1;2* as a result of exposure to salt stress is much higher than those of the *A. thaliana* or two other *T. salsuginea* homologs (Wu et al., 2012). Even its basal level of expression in shoots and roots are much higher than the other three homologs studied. Although this is a clear indication, it needs further verification by the expression of the gene to be tested under the control of its native promoter in the susceptible model plant *A. thaliana* (Fig., 3). Such a strategy should refrain from constitutive over-expression, since this would eliminate the regulatory aspect. If the tested gene variant is indeed responsible for a monogenic trait, a gain-of-function should be identified; otherwise further experiments have to be conducted to decipher whether this particular gene is not conferring the tested stress resistance or whether a polygenic effect is involved. Clearly, the method proposed here will identify primary candidates for genetic engineering of crops such as *Brassica*s to enhance their stress tolerance.

9. Functional genomic resources available for future research in this area

While this article primarily discusses the potential advantages of having multiple copies of a stress response gene in the genome, multiple copies may also create problems with obtaining visible phenotypes due to genetic redundancy. Therefore, in spite of the development of TILLING resources for *B. rapa* (Stephenson et al., 2010, Table 3), *B. napus* (Wang et al., 2008), and *B. juncea* (Himelblau et al., 2009), more targeted gene-centric methods are warranted to save time and to increase the success rate. Recent improvements in the CRISPR-Cas9 mutagenesis method extended its successful application not only to the model plant *Arabidopsis* (Fauser et al., 2014; Feng et al., 2014; Table 3), but also to polyploid crops like wheat (Wang et al., 2014b), rape (Braatz et al., 2017), and *B. oleracea* (Lawrenson et al., 2015). Targeted mutagenesis of four alleles of a pair of *BnALC* homologs was recently demonstrated in rapeseed without any off-target effects (Braatz et al., 2017). Availability of such precise, genome editing technology will be beneficial for functional analysis of genes in the complex, polyploidy genomes (Lu et al., 2019). However, identification of neo-functionalized, stress-related genes and their mobilization via transgenesis to susceptible crops having high yield potential (as discussed above) can be implemented in crop improvement programs in future. Suitable plant transformation methods in *Brassica*s, which limited progress even a few years ago, is not a limitation today with standardized transformation protocols established for different *Brassica* species (Bhalla and Singh, 2008, Baskar et al., 2016; Table 3).

~~Therefore, in addition to a gain-of-function approach (discussed above), the CRISPR-Cas-mediated loss-of-function approach of individual or multiple genes could be tried in the future, to attain better tolerance against stresses (Lu et al., 2019).~~

10. Conclusions

Considerable variability with respect to stress sensitivity exists in the family Brassicaceae at the species and also at the variety level. Although *Thellungiella* has already emerged as a model halophytic plant, there are additional Brassica species (*Lepidium virginicum, L. densiflorum, Cakile maritima*, *Lobularia maritima, Caulanthus amplexicaulis, Thlaspi arvense*) that show promise, since they cover different stress resistance phenotypes (Orsini et al., 2010). Extensive use of high-throughput phenotyping platforms will enable researchers to efficiently screen more germplasms in order to unveil additional useful genes for crop plant engineering (Singh et al., 2016, IPPN: http://www.plant-phenotyping.org/, EPPN: www.plant-phenotyping-network.eu/, DPPN: www.dppn.de/). The rapid reduction of sequencing costs will accelerate genome-sequencing efforts of many of these germplasms by community-driven initiatives such as the Brassicales Map Alignment Project (http://www.brassica.info/resource/sequencing/bmap.php). The release of new genome sequences enables experimental plant biologists to address individual stress gene families, to study gene family expansion, and to identify novel allelic variants suitable for crop improvements. The proposed method of combining bioinformatic analyses along with transcriptomic methods to identify stress-regulated genes and to further confirm this prediction by *in planta* transgenic technology render a framework for multi-genome analyses in the future. Thus, studying stress-regulated duplicated genes can offer more than just basic understanding about gene function and evolution, as it enhances prospects for genetic engineering of crops resilient to stresses.

Acknowledgements

The research discussed in this paper is funded by UGC Major Research Grant (MRP-MAJOR-BIOT-2013-18380) and Faculty Research and Professional Development (FRPDF) grant, Presidency University. MD lab also acknowledge funding from Council of Scientific and Industrial Research, India [38(1386)/14/EMR-II], [38(1493)/19/EMR-II], Department of Biotechnology, Govt. of India (BT/PR10778/PBD/16/1070/2014, BT/PR28859/FCB/125/3/2018) and Alexander von Humboldt

Foundation, Germany. SD acknowledges a SRF fellowship from CSIR, India [08/l55(0055)/ 2019 EMR-I]. The authors thank Sitakant Pattanaik, University of Kentucky, USA for useful discussion on promoter analyses and the review editor for his critical comments to improve the quality of the manuscript.

References

Abou-Hussein, S.D., 2012. Climate change and its impact on the productivity and quality of vegetable crops. J Appl. Sci. Res. 8, 4359-4383.

Alam, I, Cui, D.L., Batool, K., Yang, Y.Q., and Lu, Y.H., 2019. Comprehensive Genomic Survey, Characterization and Expression Analysis of the HECT Gene Family in *Brassica rapa* L. and *Brassica oleracea* L. Genes 10, 400, doi: 10.3390/genes10050400.

Ali, Z., Park, H.C., Ali, A., Oh, D.H., Aman, R., Kropornicka, A., Hong, H., Choi, W., Chung, W.S., Kim, W.Y., Bressan, R.A., Bohnert, H.J., Lee, S.Y., Yun, D.J., 2012. *TsHKT1;2*, a *HKT1* homolog from the extremophile *Arabidopsis* relative *Thellungiella salsuginea*, shows K specificity in the presence of NaCl. Plant Physiol. 158, 1463-1474.

Arsovski, A.A., Pradinuk, J., Guo, X.Q., Wang, S., Adams, K.L., 2015. Evolution of *Cis*-Regulatory Elements and Regulatory Networks in Duplicated Genes of *Arabidopsis*. Plant Physiol. 169, 2982-2991.

Bai, J., Wu, F., Mao, Y., He, Y., 2013. In planta transformation of *Brassica rapa* and *B. napus* via vernalization-infiltration methods. Protocol Exchange doi:10.1038/protex.2013.067.

Bandehagh, A., Salekdeh, G.H., Toorchi, M., Mohammadi, A., Komatsu, S., 2011. Comparative proteomic analysis of canola leaves under salinity stress. Proteomics. 11(10), 1965-75.

Baskar, V., Gangadhar, B. H., Park, S.W., Nile, S.H., 2016. A simple and efficient *Agrobacterium tumefaciens*-mediated plant transformation of *Brassica rapa* ssp. *pekinensis*. 3 Biotech. 6(1), 88.

Beacham, A.M., Hand, P., Pink, D.A., Monaghan, J.M., 2017. Analysis of *Brassica oleracea* early stage abiotic stress responses reveals tolerance in multiple crop types and for multiple sources of stress. J. Sci. Food Agric. (15), 5271-5277.

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., Mathews. S., 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc Natl Acad Sci USA. 107, 18724-8.

Bhalla, P.L., Singh, M.B., 2008. *Agrobacterium*-mediated transformation of *Brassica napus* and *Brassica oleracea.* Nature Protocols. 3, 181-189.

Bhardwaj, A.R., Joshi, G., Kukreja, B., Malik, V., Arora, P., Pandey, R., Shukla, R.N., Bankar, K.G., Katiyar-Agarwal, S., Goel, S., Jagannath, A., Kumar, A., Agarwal, M., 2015. Global insights into high temperature and drought stress regulated genes by RNA-Seq in economically important oilseed crop *Brassica juncea*. BMC Plant Biol. 15, 9.

Blanc, G., Wolfe, K.H., 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell. 16, 1679-1691.

Bowers, J.E., Chapman, B.A., Rong, J., Paterson, A.H., 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature. 422, 433-438.

Braatz, J., Harloff, H.J., Mascher, M., Stein, N., Himmelbach, A., Jung, C., 2017. CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). Plant Physiol. 174(2), 935-942.

Briskine, R.V., Paape, T., Shimizu-Inatsugi, R., Nishiyama, T., Akama, S., Sese, J., Shimizu, K.K., 2016. Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. Mol Ecol Resour. doi: 10.1111/1755-0998.12604.

Chakraborty, K., Sairam, R.K., Bhattacharya, R.C., 2012. Differential expression of salt overly sensitive pathway genes determines salinity stress tolerance in *Brassica* genotypes. Plant Physiol Biochem. 51, 90-101.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.H., Thi, V.H., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H., Wang, X., Canaguier, A., Chauveau, A., Bérard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., Wincker, P., 2014. Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science .345, 950-953.

Chao, D.Y., Dilkes, B., Luo, H., Douglas, A., Yakubova, E., Lahner, B., Salt, D.E., 2013. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. Science. 341(6146), 658-659.

Cheng, C., Zhong, Y., Cai, Z., Su, R., Li, C., 2019. Genome-Wide Identification and Gene Expression Analysis of ABA Receptor Family Genes in *Brassica juncea* var. *tumida*. Genes. 10, 470; doi: 10.3390/genes10060470.

Chikkaputtaiah, C., Debbarma, J., Baruah, I., Havlickova, L., Boruah, H.P.D., Curn, V., 2017. Molecular genetics and functional genomics of abiotic stress responsive genes in oilseed rape (*Brassica napus* L.): a review of recent advances and future. Plant Biotech. Rep. 11, 365-384.

Comai, L., 2005. The advantages and disadvantages of being polyploid. Nat Rev Genet. 6, 836-846.

Courbot, M., Willems, G., Motte, P., Arvidsson, S., Roosens, N., Saumitou-Laprade, P., Verbruggen, N., 2007. A major quantitative trait locus for cadmium tolerance in *Arabidopsis halleri* colocalizes with *HMA4*, a gene encoding a heavy metal ATPase. Plant Physiol. 144, 1052-1065.

Dalton-Morgan, J., Hayward, A., Alamery, S., Tollenaere, R., Mason, A.S., Campbell, E., Patel, D., Lorenc, M.T., Yi, B., Long, Y., Meng, J., Raman, R., Raman, H., Lawley, C., Edwards, D., Batley, J., 2014. A high-throughput SNP array in the amphidiploid species *Brassica napus* shows diversity in resistance genes. Funct Integr Genomics. 14(4), 643-55.

Das, M., Haberer, G., Panda, A., Das Laha, S., Ghosh, T.C., Schäffner, A.R., 2016. Expression pattern similarities support the prediction of orthologs retaining common functions after gene duplication events. Plant Physiol. 171, 2343-2357.

Das Laha, S., Naskar, A.J., Sarkar, T., Guha, S., Mondal, H.A., Das, M., 2020. Field phenotyping for salt tolerance and imaging techniques for crop stress biology. "Intelligent Image Analysis for Plant Phenotyping" edited by A. Samal, and S. Das Choudhury, CRC Press/Taylor & Francis Group. DOI: https://doi.org/10.1201/9781315177304

Dassanayake, M., Oh, D.H., Haas, J.S., Hernandez, A., Hong, H., Ali, S., Yun, D.J., Bressan, R.A., Zhu, J.K., Bohnert, H.J., Cheeseman, J.M., 2011. The genome of the extremophile crucifer *Thellungiella parvula*. Nat Genet 43, 913-918.

Ding, F., Cui, P., Wang, Z., Zhang, S., Ali, S., Xiong, L., 2014. Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in *Arabidopsis*. BMC Genom. 15, 431.

Dong, H., Liu, D., Han, T., Zhao, Y., Sun, J., Lin, S., Cao, J., Chen, Z.H., Huang, L., 2015. Diversification and evolution of the SDG gene family in *Brassica rapa* after the whole genome triplication. Scientific Reports. 5, 16851.

Dossa, K., Diouf, D., Cissé, N., 2016. Genome-wide investigation of *Hsf* genes in sesame reveals their segmental duplication expansion and their active role in drought stress response. Front. Plant Sci. 7, 1522.

Fang, L., Cheng, F., Wu, J., Wang, X, 2012. The impact of genome triplication on tandem gene evolution in *Brassica rapa*. Front Plant Sci. 3, 261.

Fauser, F., Schiml, S., Puchta, H., 2014. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. Plant J. 79 (2), 348-59.

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., Zeng, L., Liu, X., Zhu, J.K., 2014. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. Proc Natl Acad Sci U S A. 111(12), 4632-7.

Guo, Y.M., Samans, B., Chen, S., Kibret, K.B., Hatzig, S., Turner, N.C., Nelson, M.N., Cowling, W.A., Snowdon, R.J., 2017. Drought-tolerant *Brassica rapa* shows rapid expression of gene networks for general stress responses and programmed cell death under simulated drought stress. Plant Mol. Biol. Rep. 35(4), 416-30.

Hadi, F., Ayaz, M., Ali, S., Shafiq, M., Ullah, R., Jan, A.U., 2014. Comparative effect of polyethylene glycol and mannitol induced drought on growth (*in vitro*) of canola (*Brassica napus*), cauliflower (*Brassica oleracea*) and tomato (*Lycopersicum esculentum*) seedlings. Int. J. Biosci. 4(9), 34-41.

Hanikenne, M.W., Talke, I.N., Haydon, M.J., Lanz, C., Nolte, A., Motte, P., Kroymann, J., Weigel, D., Krämer, U., 2008., Evolution of metal hyperaccumulation required *cis*-regulatory changes and triplication of *HMA4*. Nature. 453, 391-395.

Hazra, A., Dasgupta, N., Sengupta, C., Das, S., 2019. MIPS: Functional dynamics in evolutionary pathways of plant kingdom. Genomics. 111(6), 1929-1945.

Himelblau, E., Gilchrist, E.J., Buono, K., Bizzell, C., Mentzer, L., Vogelzang, R., Osborn, T., Amasino, R.M., Parkin, I.A., Haughn, G.W., 2009. Forward and reverse genetics of rapid-cycling *Brassica oleracea*. Theor Appl Genet. 118(5), 953-61.

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J.D., Ossowski, S., Ottilar, R.P., Salamov, A.A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M.E., Bergelson, J., Carrington, J.C., Gaut BS., Schmutz, J., Mayer, K.F., Van de Peer, Y., Grigoriev, I.V., Nordborg, M., Weigel, D., Guo, Y.L., 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet. 43, 476-81.

Huang, Y., Shi, J., Tao, Z., Zhang, L., Liu, Q., Wang, X., Yang, Q., Liu, G., Wang, H., 2014. Microarray expression analysis of the main inflorescence in *Brassica napus*. PLOS One 9, e102024.

Jiao, W., Schneeberger, K., 2020. Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics. Nat. Commun. 11, 989.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., dePamphilis, C.W., 2011. Ancestral polyploidy in seed plants and angiosperms. Nature. 473. 97-100.

Jin, L., Ouyang, N., Huang, Y., Liu, C., Ruan, Y., 2019. Genome-wide analysis of sulfotransferase genes and their responses to abiotic stresses in Chinese cabbage (*Brassica rapa* L.). PLOS One https://doi.org/10.1371/journal.pone.0221422.

Johnston, J.S., Pepper, A.E., Hall, A.E., Chen, Z.J., Hodnett, G., Drabek, J., Lopez, R., Price, H.J., 2005. Evolution of genome size in Brassicaceae. Ann Bot, 95, 229-235.

Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., Links, M.G., Sharpe, A.G., Parkin, I.A., 2014. Polyploid evolution of the Brassicaceae during the Cenozoic era. Plant Cell. 26. 2777-91.

Kannan, S., Halter, G., Renner, T., Waters, E.R., 2018. Patterns of alternative splicing vary between species during heat stress. AoB Plants. 10(2), ply013.

Kim, A.R., Lim, H., Cho, J.I., Kim, C.K., Ji, S.U., Park, S.C., Lee, G.S., 2015. Overexpression of BrTSR53 gene improves tolerance of rice plant to salt stress. Plant Breed. Biotech. 3(4), 376-383.

Kim, S., Park, J., Yeom, S.I., Kim ,Y.M., Seo, E., Kim, K.T., Kim, M.S., Lee, J.M., Cheong, K., Shin, H.S., Kim, S.B., Han, K., Lee, J., Park, M., Lee, H.A., Lee, H.Y., Lee, Y., Oh, S., Lee, J.H., Choi, E., Choi, E., Lee, S.E., Jeon, J., Kim, H., Choi, G., Song, H., Lee, J., Lee, S.C., Kwon, J.K., Lee, H.Y., Koo, N., Hong, Y., Kim, R.W., Kang, W.H., Huh, J.H., Kang, B.C., Yang, T.J., Lee, Y.H., Bennetzen, J.L., Choi, D., 2017. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biol. 18, 210.

Kirchner, T.W., Niehaus, M., Debener, T., Schenk, M.K., Herde, M., 2017. Efficient generation of mutations mediated by CRISPR/Cas9 in the hairy root transformation system of *Brassica carinata*. PLoS One. 12(9), e0185429.

Krishnamurthy, P., Muthusamy, M., Kim, J.A., Jeong, M.J., Lee, S.I. (2019) *Brassica rapa* expansin-like B1 gene (BrEXLB1) regulate growth and development in transgenic Arabidopsis and elicits response to abiotic stresses. J. Plant Biochem. and Biotech. 28(4), 437-446.

Kuo, Y.T., Chao, Y.T., Chen, W.C., Shih, M.C., Chang, S.B., 2019. Segmental and tandem chromosome duplications led to divergent evolution of the chalcone synthase gene family in *Phalaenopsis* orchids. Ann. Bot. 123(1), 69-77.

Landis, J.B., Soltis, D.E., Li, Z., Marx, H.E., Barker, M.S., Tank, D.C., Soltis, P.S., 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. Am. J. Bot. 105(3), 348-363.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., Uauy, C., Harwood, W., 2015. Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. Genome Biol. 16, 258.

Lee, S.C, Lim, M.H., Yu, J.G., Park, B.S., Yang, T.J., 2012. Genome-wide characterization of the *CBF*/*DREB1* gene family in *Brassica rapa*. Plant Physiol Biochem. 61, 142-152.

Lee, Y.P., Giorgi, F.M., Lohse, M., Kvederaviciute, K., Klages, S., Usadel, B., Meskiene, I., Reinhardt, R., Hincha, D.K., 2013. Transcriptome sequencing and microarray design for functional genomics in the extremophile *Arabidopsis* relative *Thellungiella salsuginea* (*Eutrema salsugineum*). BMC Genomics. 14, 793.

Li, L., Stoeckert, C.J, Jr., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178-2189.

Liang, Y., Xiong, Z., Zheng, J., Xu, D., Zhu, Z., Xiang, J., Gan, J., Raboanatahiry, N., Yin, Y., Li, M., 2016. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (*LEA*) gene family formation pattern in *Brassica napus*. Sci Rep. 6, 24265.

Ling, Y., Alshareef, S., Butt, H., Lozano-Juste, J., Li, L., Galal, A.A., Moustafa, A., Momin, A.A., Tashkandi, M., Richardson, D.N., Fujii, H., Arold, S., Rodriguez, P.L., Duque, P., Mahfouz, M.M., 2017. Pre-mRNA splicing repression triggers abiotic stress signaling in plants. Plant J. 89, 291-309.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A., Zhao, M., et al. 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun. 5, 3930.

Lu, S., Fadlalla, T., Tang, S., Li, L., Ali, U., Li, Q., Guo L., 2019. Genome-Wide Analysis of Phospholipase D Gene Family and Profiling of Phospholipids under Abiotic Stresses in *Brassica napus*. Plant Cell Physiol. 60(7), 1556-1566.

Lysak, M.A., Koch, M.A., Pecinka, A., Schubert, I., 2005. Chromosome triplication found across the tribe Brassiceae. Genome Res, 15, 516-525.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., Vande Peer, Y., 2005. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA. 102, 5454–5459.

Mishra, A., Tanna, B., 2017. Halophytes: potential resources for salt stress tolerance genes and promoters. Front. Plant Sci. 8, 829.

Mooney, S, Al-Saharin, R., Choi, C.M., Tucker, K., Beathard, C., Hellmann, H.A., 2019. Characterization of *Brassica rapa* RAP2.4-Related Proteins in Stress Response and as CUL3-Dependent E3 Ligase Substrates. Cells. 8, 336.

Mun, J.H., Kwon, S.J., Yang, T.J., Seol, Y.J., Jin, M., Kim, J.A., Lim, M.H., Kim, J.S., Baek, S., Choi, B.S., Yu, H.J., Kim, D.S., Kim, N., Lim, K.B., Lee, S.I., Hahn, J.H., Lim, Y.P., Bancroft, I., Park, B.S., 2009 Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. Genome Biol. 10, R111.

Navabi, Z.K., Huebert, T., Sharpe, A.G., O'Neill, C.M., Bancroft, I., Parkin, I.A., 2013. Conserved microstructure of the Brassica B genome of *Brassica nigra* in relation to homologous regions of *Arabidopsis thaliana, B. rapa* and *B. oleracea*. BMC Genomics. 14, 250.

Nutan, K.K., Kumar, G., Singla-Pareek, S.L., Pareek, A., 2018. A Salt Overly Sensitive pathway member from *Brassica juncea* BjSOS3 can functionally complement ΔAtsos3 in *Arabidopsis*. Current Genomics 19, 60-69.

Oh, D.H., Dassanayake, M., 2019. Landscape of gene transposition-duplication within the Brassicaceae family, DNA Res. 26 (1), 21-36.

Oh, D.H., Hong, H., Lee, S.Y., Yun, D.J., Bohnert, H.J., Dassanayake, M., 2014. Genome structures and transcriptomes signify niche adaptation for the multiple-ion-tolerant extremophyte *Schrenkiella parvula*. Plant Physiol. 164(4), 2123-38.

Orsini, F., D'Urzo, M.P., Inan, G., Serra, S., Oh, D.H., Mickelbart, M.V., 2010. A comparative study of salt tolerance parameters in 11 wild relatives of *Arabidopsis thaliana*. J Exp Bot. 61. 3787-3798.

Panchy, N., Lehti-Shiu, M., Shiu, S.H., 2016. Evolution of Gene Duplication in Plants. Plant Physiol. 171(4), 2294-2316.

Pang, Q., Chen, S., Dai, S., Chen, Y., Wang, Y., Yan, X., 2010. Comparative proteomics of salt tolerance in *Arabidopsis thaliana* and *Thellungiella halophila*. J. Proteome Res. 9(5), 2584-2599.

Park, J.I., Ahmed, N.U., Jung, H.J., Arasan, S.K.T., Chung, M.Y., Cho, Y.G., Watanabe, M., Nou, I.S., 2014. Identification and characterization of LIM gene family in *Brassica rapa*. BMC Genomics, 15, 641

Patel, R.V., Nahal, H.K., Breit, R., Chen, Y., Provart, N.J., 2012 BAR Expressolog identification: expression profile ranking of "orthologous" genes in plant species. Plant J. 71, 1038-1050.

Pi, B., He, X., Ruan, Y., Jang, J.C., Huang, Y., 2018. Genome-wide analysis and stress responsive expression of CCCH zinc finger family genes in *Brassica rapa*. BMC Plant Biol. 18, 373.

Powell, R.V., Willett, C.R., Goertzen, L.R., Rashotte, A.M., 2019. Lineage specific conservation of cis-regulatory elements in Cytokinin Response Factors. Sci Rep. 9, 13387

Purty, R.S., Kumar, G., Singla-Pareek, S.L., Pareek, a. 2008. Towards salinity tolerance in Brassica: an overview. Physiol. Mol. Biol. Plants 14, 39–49.

Qiao, X., Li, Q., Yin, H, Qi, K., Li, L., Wang, R., Zhang, S., Paterson, A.H., 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol. 20, 38.

Ramsey, J., Ramsey, T.S., 2014. Ecological studies of polyploidy in the 100 years following its discovery. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369(1648), 20130352.

Rodríguez, V.M., Soengas, P., Alonso-Villaverde, V., Sotelo, T., Cartea, M.E., Velasco, P., 2015. Effect of temperature stress on the early vegetative development of *Brassica oleracea* L. BMC Plant Biol. 15, 145.

Saha, G., Park, J.I., Kayum, M.A., Nou, I.S., 2016 A genome-wide analysis reveals stress and hormone responsive patterns of TIFY family genes in *Brassica rapa*. Front Plant Sci, 7, 936.

Schlaeppi, K., Dombrowski, N., Oter, R.G., Themaat, E.V.L.V., Schulze-Lefert, P., 2014. Quantitative divergence of the bacterial root microbiota in *Arabidopsis thaliana* relatives. Proc Natl Acad Sci USA, 111, 585-592.

Shahzad, Z., Gosti, F., Frérot. H., Lacombe, E., Roosens, N., Saumitou-Laprade, P., Berthomieu, P., 2010. The five AhMTP1 zinc transporters undergo different evolutionary fates towards adaptive evolution to zinc tolerance in *Arabidopsis halleri*. PLOS Genet. 6, e1000911.

Sharma, P., Sardana, V., Banga, S.S., 2013 Salt tolerance of Indian mustard (*Brassica juncea*) at germination and early seedling growth. Environmental and Experimental Biology. 11, 39-46.

Siddiqui, M. H., Mohammad F., Khan. M. N. 2009. Morphological and physio-biochemical characterization of *Brassica juncea* L. Czern. & Coss. genotypes under salt stress. J. Plant Interactions 4, 67-80.

Singh, A., Ganapathysubramanian, B., Singh, A.K., Sarkar, S., 2016. Machine learning for high-throughput stress phenotyping in plants. Trends Plant Sci. 21, 110-124.

Srivastava, A.K., Srivastava, S., Lokhande, V.H., D'Souza, S.F., Suprasanna, P., 2015[a]. Salt stress reveals differential antioxidant and energetics responses in glycophyte (*Brassica juncea* L.) and halophyte (*Sesuvium portulacastrum* L.). Front. Environ. Sci. 3(19), 1-9.

Srivastava, S., Srivastava, A.K., Sablok, G., Deshpande, T.U., Suprasanna, P., 2015[b]. Transcriptomics profiling of Indian mustard (*Brassica juncea*) under arsenate stress identifies key candidate genes and regulatory pathways. Front Plant Sci. 6, 646.

Stephenson, P., Baker, D., Girin, T., Perez, A., Amoah, S., King, G.J., Østergaard, L., 2010. A rich TILLING resource for studying gene function in *Brassica rapa*. BMC Plant Biology, 10: 62.

Sun, C., Wu, J., Liang, J., Schnable, J.C., Yang, W., Cheng, F., Wang, X., 2015. Impacts of whole-genome triplication on MIRNA evolution in *Brassica rapa*. Genome Biol Evol. 7, 3085-3096.

Uzilday, B., Ozgur, R., Sekmen, A.H., Yildiztugay, E., Turkan, I., 2015. Changes in the alternative electron sinks and antioxidant defence in chloroplasts of the extreme halophyte *Eutrema parvulum* (*Thellungiella parvula*) under salinity. Ann Bot. 115, 449-463.

Verma, D., Lakhanpal, N., Singh, K., 2019. Genome-wide identification and characterization of abiotic-stress responsive SOD (superoxide dismutase) gene family in *Brassica juncea* and *B. rapa*. BMC Genomics, 20: 227.

Vicient, C.M., Casacuberta, J.M., 2017. Impact of transposable elements on polyploid plant genomes. Ann Bot. 120(2), 195-207.

Waminal, N.E., Perumal, S., Lee, J., Kim, H.H., Yang, T.J., 2016. Repeat Evolution in *Brassica rapa* (AA), B. oleracea (CC), and B. napus (AACC) Genomes. Plant Breed Biotech. 4, 107-122.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., 2011. The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet. 43, 1035-1039.

Wang, Y., Tan, X., Paterson, A.H., 2013. Different patterns of gene structure divergence following gene duplication in *Arabidopsis*. BMC Genomics. 14, 652.

Wang, J., Marowsky, N.C., Fan, C., 2014a. Divergence of gene body DNA methylation and evolution of plant duplicate genes. PLOS One. 9, e110357.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., Qiu, J.L., 2014b. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat. Biotech. 32(9), 947-51.

Wang, L., Ma, H., Lin, J., 2019. Angiosperm-wide and family-level analyses of AP2/ERF genes reveal differential retention and sequence divergence after Whole-Genome Duplication. Front. Plant Sci. 10, 196.

Wang, J., Huang, F., You, X., Hou, X., 2018. Molecular cloning, characterization and expression analysis of BcHHP3 under abiotic stress in Pak-choi (*Brassica rapa* ssp. Chinensis). J. Plant Interactions 14(1), 1-9.

Wang, N., Wang, Y., Tian, F., King, G.J., Zhang, C., Long, Y., Shi, L., Meng, J., 2008. A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of *FAE1* point mutations by TILLING. New Phytol. 180(4), 751-65.

Wang, S., Adams, K.L., 2015. Duplicate gene divergence by changes in microRNA binding sites in *Arabidopsis* and *Brassica*. Genome Biol. Evol. 7, 646-655.

Wei, Y., Xiao, D., Zhang, C., Ho, X., 2019. The Expanded SWEET Gene Family Following Whole Genome Triplication in *Brassica rapa*. Genes, 10, 722.

Willems, G., Dräger, D.B., Courbot, M., Godé, C., Verbruggen, N., Saumitou-Laprade, P., 2007. The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): an analysis of quantitative trait loci. Genetics. 176, 659-674.

Wood, I.P., Pearson, B.M., Garcia-Gutierrez, E., Havlickova. He, Z., Harper, A.L., Bancroft, I., Waldron, K.W., 2017. Carbohydrate microarrays and their use for the identification of molecular markers for plant cell wall composition. Proc Natl Acad Sci U S A. 114(26), 6860-6865.

Wu, H.J., Zhang, Z., Wang, J.Y., Oh, D.H., Dassanayake, M., Liu, B., Huang, Q., Sun, H.X., Xia, R., Wu, Y., Wang, Y.N., Yang, Z., Liu, Y., Zhang, W., Zhang, H., Chu, J., Yan, C., Fang, S., Zhang, J., Wang, Y., Zhang, F., Wang, G., Lee, S.Y., Cheeseman, J.M., Yang, B., Li, B., Min, J., Yang, L., Wang, J., Chu, C., Chen, S.Y., Bohnert, H.J., Zhu, J.K., Wang, X.J., Xie, Q., 2012. Insights into salt tolerance from the genome of *Thellungiella salsuginea*. Proc Nat Acad Sci USA. 109, 12219-12224.

Xie, T., Zeng, L., Chen, X., Rong, H., Wu, J., Batley, J., Jiang, J., Wang, Y., 2020. Genome-wide analysis of *the Lateral Organ Boundaries Domain* gene family in *Brassica napus*. Genes (Basel). 11(3), E280.

Xu, G., Guo, C., Shan, H., Kong, H., 2012. Divergence of duplicate genes in exon-intron structure. Proc Natl Acad Sci USA, 109, 1187-1192.

Xu, L., Lin, Z., Tao, Q., Liang, M., Zhao, G., Yin, X., Fu, R., 2014. Multiple *NUCLEAR FACTOR Y* transcription factors respond to abiotic stress in *Brassica napus* L. PLoS One. 9(10), e111354.

Yadav, D., Ahmed, I., Kirti, P.B., 2015. Genome-wide identification and expression profiling of annexins in *Brassica rapa* and their phylogenetic sequence comparison with *B. juncea* and *A. thaliana* annexins. Plant Gene 4, 109-124

Yang, R., Jarvis, D.E., Chen, H., Beilstein, M.A., Grimwood, J., Jenkins, J., Shu, S., Prochnik, S., Xin, M., Ma, C., Schmutz, J., Wing, R.A., Mitchell-Olds, T., Schumaker, K.S., Wang, X., 2013. The reference genome of the halophytic plant *Eutrema salsugineum*. Front Plant Sci. 4, 1-14.

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., Hu, Z., Chen, S., Pental, D., Ju, Y., Yao, P., Li, X., Xie, K., Zhang, J., Wang, J., Liu, F., Ma, W., Shopan, J., Zheng, H., Mackenzie, S.A., Zhang, M., 2016. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. Nat Genet. 48, 1225-1232.

Yang, H., Wu, J.J., Tang, T., Liu, K.D., Dai, C., 2017. CRISPR/Cas9-mediated genome editing efficiently creates specific mutations at multiple loci using one sgRNA in *Brassica napus*. Scientific Reports. 7, 7489.

Yousuf, P.Y., Ahmad, A., Ganie, A.H., Iqbal, M., 2016. Salt stress-induced modulations in the shoot proteome of *Brassica juncea* genotypes. Environ Sci Pollut Res Int. 23(3), 2391-401.

Yu, J., Ke, T., Tehrim, S., Sun, F., Liao, B., Hua, W., 2015. PTGBase: an integrated database to study tandem duplicated genes in plants. Database (Oxford) 22, 2015.

Zhang, Z., Zhou, L., Wang, P., Liu, Y., Chen, X., Hu, L., Kong, X., 2009. Divergence of exonic splicing elements after gene duplication and the impact on gene structures. Genome Biol. 10, R120.

Zhang, P.G., Huang, S.Z., Pin, A.L., Adams, K.L., 2010. Extensive divergence in alternative splicing patterns after gene and genome duplication during the evolutionary history of *Arabidopsis*. Mol Biol Evol. 27, 1686-1697.

Zhang, Q., Zhao, C., Li, M., Sun, W., Liu, Y., Xia, H., Sun, M., Li, A., Li, C., Zhao, S., Hou, L., Picimbon, J.F., Wang, X., Zhao, Y., 2013. Genome-wide identification of *Thellungiella salsuginea* microRNAs with putative roles in the salt stress response. BMC Plant Biol. 13, 180.

Zhang, T., Moa, J., Zhoua, K., Changa, Y., Liub, Z. 2018. Overexpression of *Brassica campestris* BcICE1 gene increases abiotic stress tolerance in tobacco. Plant Physiol. and Biochem. 132, 515-523.

Zhang, X., Lu, G., Long,W., Zou, X., Li, F., Nishio, T., 2014. Recent progress in drought and salt tolerance studies in *Brassica* crops. Breed Sci. 64(1), 60-73.

Zhang, Y., Liu, Z., Khan, A.A., Lin, Q., Han, Y., Mu, P., Liu, Y., Zhang, H., Li, L., Meng, X., Ni, Z., Xin, M., 2016. Expression partitioning of homeologs and tandem duplications contribute to salt tolerance in wheat (*Triticum aestivum* L.). Sci Rep. 6, 21476.

Zhao, B.Y., Hu, Y.F., Li, J.J., Yao, X., Liu, K.D., 2016. BnaABF2, a bZIP transcription factor from rapeseed (*Brassica napus* L.), enhances drought and salt tolerance in transgenic *Arabidopsis*. Bot. Stud. 57, 12.

Zhou, R., Moshgabadi, N., Adams, K.L., 2011. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. Proc.Nat. Aca. Sci. USA. 108, 16122-16127.

Zhu, X., Wang, Y., Liu, Y., Zhou, W., Yan, B., Yang, J., Shen, Y., 2018. Overexpression of BcHsfA1 transcription factor from Brassica campestris improved heat tolerance of transgenic tobacco. PLOS One https://doi.org/10.1371/journal.pone.0207277.

Zou, Y., Huang, W., Gu, Z., Gu, X., 2011[a]. Predominant gain of promoter TATA box after gene duplication associated with stress responses. Mol Biol Evol. 28, 2893-2904.

Zou, C., Sun, K., Mackaluso, J.D., Seddon, A.E., Jin, R., Thomashow, M.F., Shiu, S.H., 2011[b]. *Cis*-regulatory code of stress-responsive transcription in *Arabidopsis thaliana*. Proc Natl Acad Sci USA. 108, 14992-14997.

Zou, X., Hu, C., Zeng, L., Cheng, Y., Xu, M., Zhang, X., 2014. A Comparison of Screening Methods to Identify Waterlogging Tolerance in the Field in *Brassica napus* L. during Plant Ontogeny. PLoS ONE 9, e89731.

Legend of Figures

Fig. 1 Origin of different Brassicaceae species, time of occurrences of 3 WGDs, 1 WGT and frequency of genes derived from these as well as tandem and segmental duplication events. Abbreviations used- WGD: whole-genome duplications, WGT: whole-genome triplication, SD: segmental duplication, TD: Tandem duplication, MYA: million years ago, na: not available, NA: not applicable.

Fig. 2 Comparison of *A. thaliana* (*At4g10310*) and *T. salsuginea HKT 1* (*TsHKT1;1- Ts6g08650*, *TsHKT1;2- Ts6g08730*, *TsHKT1;3- Ts6g08740*) genes to identify stress specific *cis* elements in the 1000 bp upstream regions of the respective gene copies. Different stress responsive *cis* elements were predicted by new PLACE database (https://www.dna.affrc.go.jp/PLACE/?action=newplace, August, 2020). They are marked with separate colours and those which have been functionally characterized are marked by asterisks. Each number indicated below each *cis* element represent total number of that element present in a particular position. The actual positions of these *cis* elements on the DNA strands have been described in Supplementary Fig. 1.

Predicted stress-responsive *cis* elements available in the Plant care database until June, 2017 (http://bioinformatics.psb.ugent.be/webtools/plantcare/html) were considered for this analyses. Abbreviations used- *ABRE*: ABA-responsive elements, *HSE*: heat shock element, *MBS*: MYB binding site.

Fig. 3 Transcriptome analysis for predicting novel allelic variants with respect to stress responses and analysis of transgenic expression of the gene in the susceptible model plant *A. thaliana* to confirm their usefulness in crop improvement programs. The gene expression data was obtained from Wu et al (2012).

Supplementary Fig. 1. Detailed comparison of 1000 bp upstream regions of A. thaliana (At4g10310) and T. salsuginea HKT 1 (TsHKT1;1- Ts6g08650, TsHKT1;2- Ts6g08730, TsHKT1;3-Ts6g08740) genes to identify stress specific cis elements. Different stress responsive cis elements have been predicted by new PLACE database (https://www.dna.affrc.go.jp/PLACE/?action=newplace, August, 2020). They are marked with separate patterns and those which have been functionally characterized are marked by asterisks. Numbers below indicate number of stress elements that have been identified within 500 bp and 1000 bp upstream of the respective homologs. Positions of the elements were cited above each element. The (+) and (-) signs indicate their locations in the positive and negative DNA strands respectively.

Supplementary Fig. 1

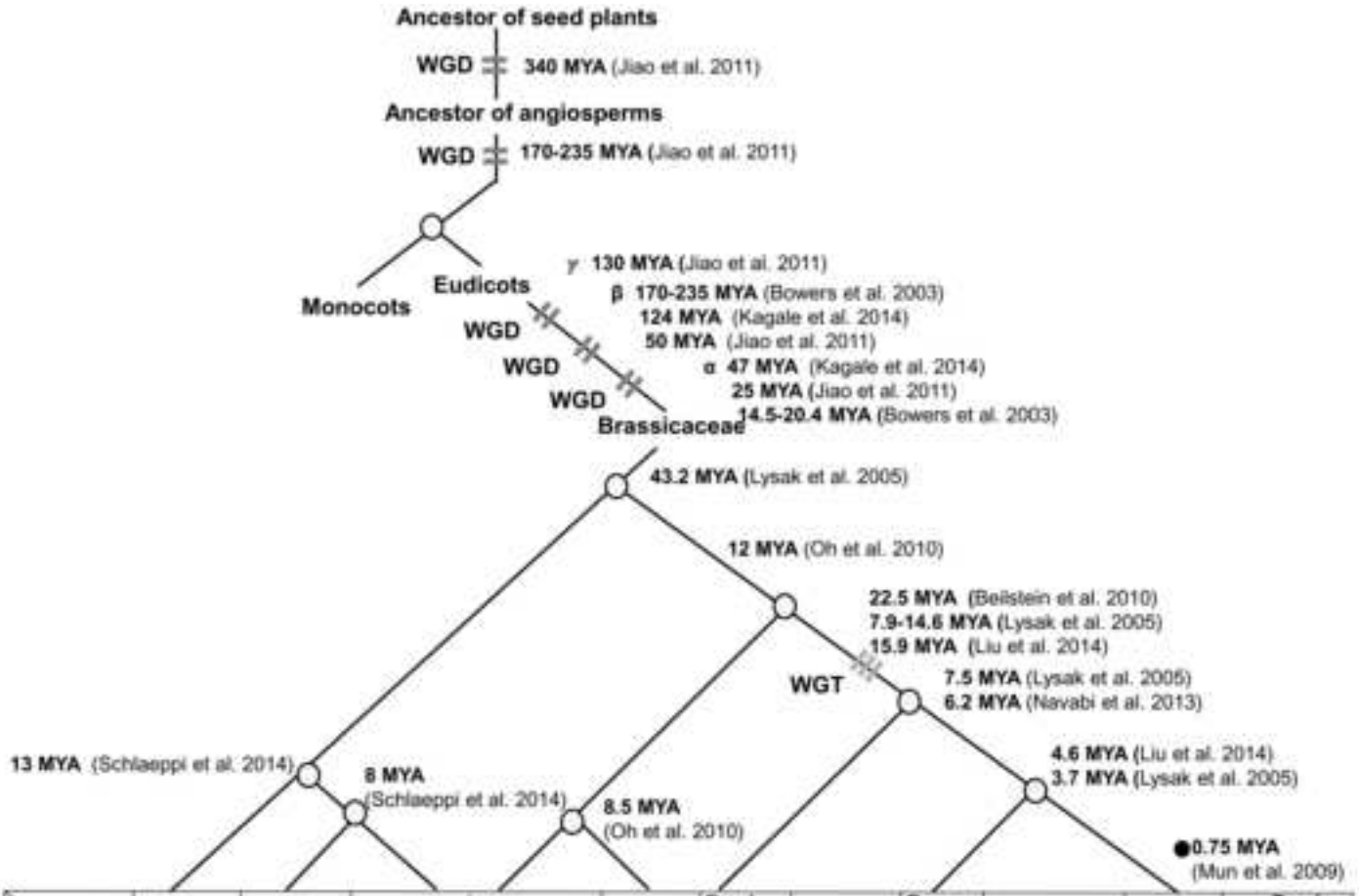
Click here to download e-component: Das Laha et al_Suppl. Fig. 1.tif

Supplementary Table 1

Click here to download e-component: Das Laha et al._Suppl. Table 1.docx

Figure 1

Click here to download high resolution image



Species / Character	*A. thaliana*	*A. halleri*	*A. lyrata*	*T. salsuginea*	*T. parvula*	*B. nigra*	*B. juncea*	*B. rapa*	*B. napus*	*B. oleracea*	*B. carinata*
Genome size (Mb)	**125** (Yu et al. 2015)	**250** (Briskine et al. 2016)	**207** (Hu et al. 2011)	**260** (Wu et al. 2012)	**137.09** (Dassanayake et al. 2011)	**632** (Johnston et al. 2005)	**922** (Yang et al. 2016)	**283.8** (Liu et al. 2014)	**1130** (Chalhoub et al. 2014)	**539.9** (Liu et al. 2014)	**1284** (Waminal et al. 2016)
Total no. of protein coding genes	**27,416** (Wu et al. 2012)	**32,553** (Briskine et al. 2016)	**32,670** (Hu et al. 2011)	**28,457** (Wu et al. 2012)	**28,901** (Dassanayake et al. 2011)	**49,826** (Yang et al. 2016)	**80,050** (Yang et al. 2016)	**41,174** (Liu et al. 2014)	**1,01,040** (Chalhoub et al. 2014)	**45,758** (Liu et al. 2014)	na
Duplicated genes from WGD	**10,483-17,406** (Rizzon et al. 2006)	na	**1,258** (Qiao et al., 2019)	**3,155** (Qiao et al., 2019)	**4,442** (Qiao et al., 2019)	na	na	**5,778** (Liu et al. 2014)	**14,241-15,681** (Qiao et al., 2019)	**5,369** (Liu et al. 2014)	na
Duplicated genes from WGT	NA	NA	NA	NA	NA	na	na	**11,448** (Liu et al. 2014)	na	**11,493** (Liu et al. 2014)	na
Duplicated genes from SD	**8,429** (Wu et al. 2012)	na	na	**8,178** (Wu et al. 2012)	na	na	na	na	na	na	na
Duplicated genes from TD	**2,708** (Wu et al. 2012)	na	**503** (Qiao et al., 2019)	**2,723** (Wu et al. 2012)	**2,940** (Dassanayake et al. 2011)	na	na	**4,841** (Liu et al. 2014)	**3,149** (Xie et al., 2020)	**3, 786** (Liu et al. 2014)	na

Fig. 1 Origin of different Brassicaceae species, time of occurrences of 3 WGDs, 1 WGT and frequency of genes derived from these as well as tandem and segmental duplication events. Abbreviations used-WGD: whole-genome duplications, WGT: whole-genome triplication, SD: segmental duplication, TD: Tandem duplication, MYA: million years ago, na: not available, NA: not applicable.

Figure 2

Click here to download high resolution image

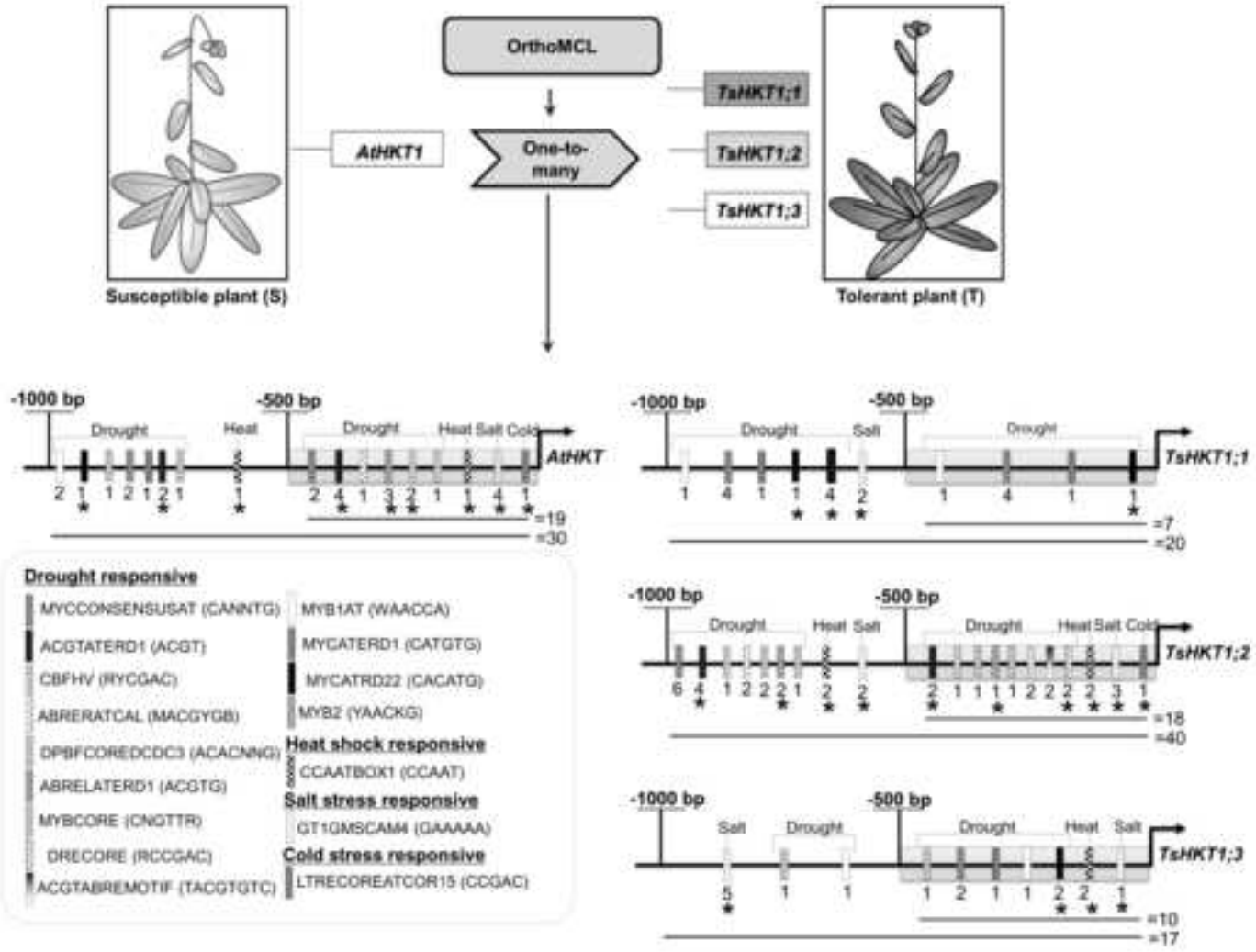


Fig. 2 Comparison of *A. thaliana* (*At4g10310*) and *T. salsuginea HKT 1* (*TsHKT1;1- Ts6g08650, TsHKT1;2- Ts6g08730, TsHKT1;3-Ts6g08740*) genes to identify stress specific cis elements in the 1000 bp upstream regions of the respective gene copies. Different stress responsive cis elements have been predicted by new PLACE database (https://www.dna.affrc.go.jp/PLACE/?action=newplace, August, 2020). They are marked with separate patterns and those which have been functionally characterized are marked by asterisks. Different numbers indicate number of respective stress elements that have been identified.

Figure 3

Click here to download high resolution image

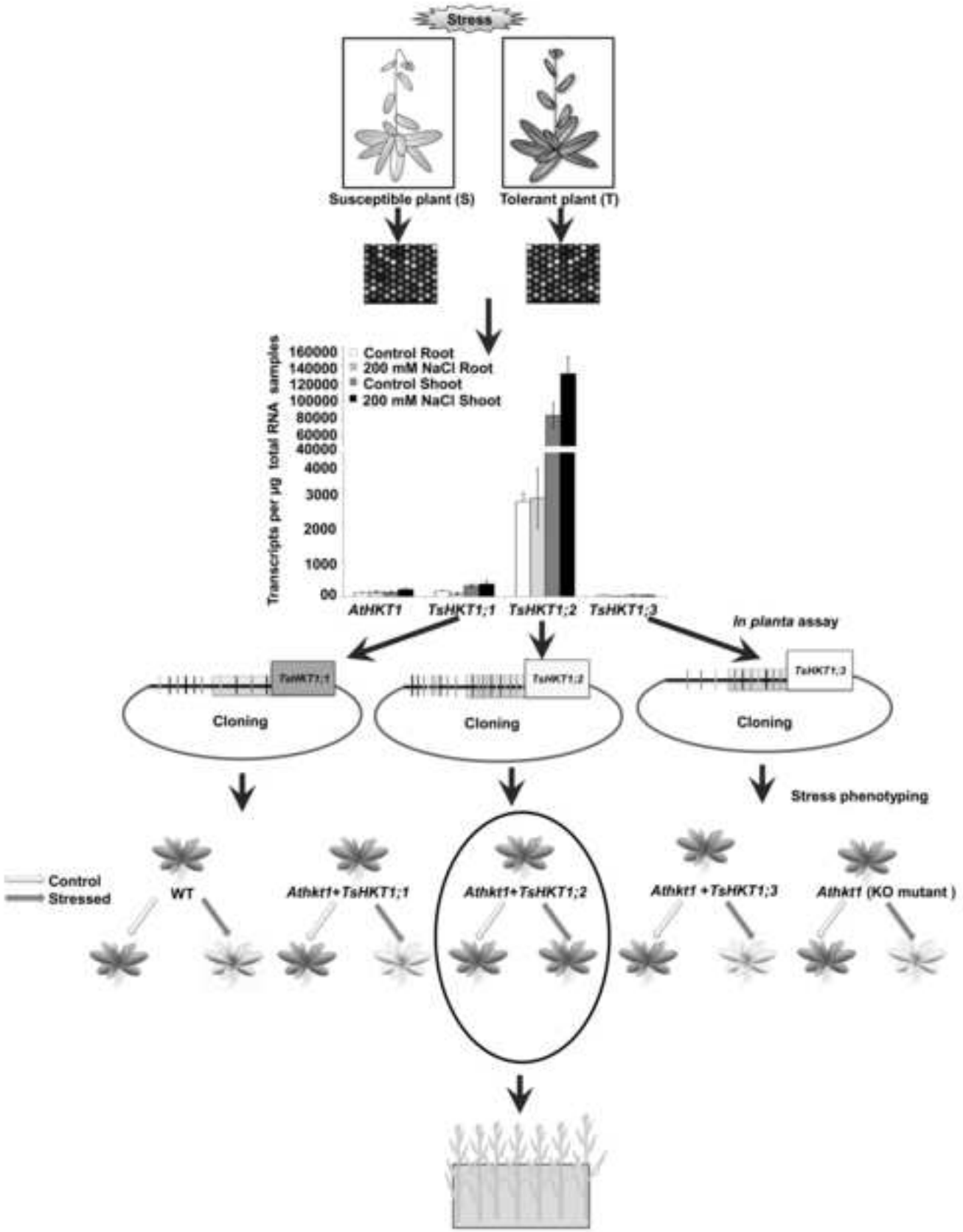


Fig. 3 Transcriptome analysis for predicting novel allelic variant with respect to stress responses and analysis of ectopic expression of the gene in the susceptible model plant *A. thaliana* to confirm their usefulness in crop improvement programs. The gene expression data was obtained from (Wu et al. 2012).

Table 1

Table 1. Stress responsive genes, which were identified in crop species of *Brassica*s and were functionally characterized by transgenic manipulation.
Abbreviation- NA: not applicable, -: not reported.

***Brassica* species**	**Genes**	**Accession numbers**	**Evolutionary origin**	**Heterologous expression**	**Mode of transgene manipulation**	**Promoter**	**Tolerance to NaCl**	**Tolerance to Drought**	**Tolerance to Low/High Temperature**	**References**
B. rapa (Chiifu-401-42)	*Expansin like B1* (*BrEXLB1*)	*Bra040107*	Tandem duplication	*A. thaliana* (Col-0)	Over expression	CaMV35S	-	PEG6000 (4%)	-	Krishnamurthy et al., 2019
B. rapa ssp. *oleifera* var. *R-o-18*	*Apetala2.4/ Wound Induced Dedifferentiation 1* (*BrRAP2.4-1*)	*Bra003659*	Segmental duplication	*A. thaliana* (Col-0)	Over expression	CaMV35S	150 mM	-	-	Mooney et al., 2019
Brassica rapa	*Triple Stress Resistance53* (*BrTSR53*)	*BRAS0001 S00024153*	-	*Oryza sativa* L. cv. *Dongjin*	Over expression	CaMV35S	100 mM	-	-	Kim et al., 2015
B. napus cultivar *Zhongshuang 11*	*Phospholipase D* (*BnPLDα1C5*)	*BnaC05g37540D*	Allotetra ploidization	*B. napus* cultivar *Westar*		*Bna PLDα1-CRISPR /Cas9*	200 mM	-	-	Lu et al., 2019
B. napus L.	*ABRE Binding Factors* (*BnABF2*)	*HE616526*	-	*A. thaliana* (Col-0)	Over expression	AtUBQ10	300 mM	Water deprivation (2 weeks)	-	Zhao et al., 2016
B. campestris (Suzhouqing)	*Heat Shock FactorA1* (*BcHsfA1*)	-	-	*N. tabacum*	Over expression	CaMV35S			42º C	Zhu et al., 2018
B. campestris (Longyou 6)	*Inducer of CBF expression 1* (*BcICE1*)	-	-	*Nicotiana tabacum* (NC89)	Over expression	CaMV35S	200 mM	PEG6000 (20%)	4º C	Zhang et al., 2018
B. juncea var. *CS52*	*Salt Overly Sensitive 3* (*BjSOS3*)	*DQ248965*	-	*A. thaliana* (*ΔAtsos3*-CS3869)	Over expression	CaMV35S	100mM and 200mM	ABA (25 µM)	-	Nutan et al., 2018

Table 2. Important gene families arisen through WGD or WGT in *Brassica*s and their responses to diverse osmotic stresses.
Abbreviations WGD: whole genome duplication, WGT: whole genome triplication, TD: tandem duplication, SD: segmental duplication.

***Brassica* species**	**Gene family**	**Number of homologs in *A. thaliana***	**Number of homologs in *Brassica***	**Modes of multiplication in *Brassica***	**Responsive to Stresses**	**Number of stress induced *Brassica* genes**	**Accession numbers**	**References**
Brassica rapa	*Homologous to E6-AP C Terminus* (*HECT*)	7	10	WGT, TD, SD	**Salt:** NaCl (200 mM)	8	*Bra000779*, *Bra005748*, *Bra021231*, *Bra022201*, *Bra027850*, *Bra028860*, *Bra029461*, *Bra038022*	Alam et al., 2019
					Drought: PEG6000 (20%)	5	*Bra000779*, *Bra005748*, *Bra021231*, *Bra028860*, *Bra029461*	
B. rapa	*Sulfotransferases* (*SOTS*)	21	56	WGD, WGT, TD	**Salt:** NaCl (200 mM)	7	G-IV (*Bra034065*), G-IV (*Bra036913*), G-V (*Bra017368*), G-V (*Bra017370*), G-VI (*Bra009300*), G-VII (*Bra027880*), G-VII (*Bra015936*)	Jin et al., 2019
					Drought: ABA (50μM), PEG6000 (25%)	8	G-I (*Bra005876*), G-IV (*Bra034065*), G-V (*Bra017370*), G-VI (*Bra009300*), G-VI (*Bra028711*), G-VII (*Bra027880*), G-VII (*Bra015936*), G-VII (*Bra015935*)	
B. rapa	*Dehydration Responsive Element Binding Transcription Factor 1* (*DREB1*)	6	10	WGT, TD	**Salt:** NaCl (250 mM)	3	*DREB1B2* (*Bra022770*), *DREB1D* (*Bra028290*), *DREB1F1* (*Bra026963*)	Lee et al., 2012
					Drought: Air drying, ABA (100 μM)	8	*DREB1A* (*Bra010461*), *DREB1B1* (*Bra010460*), *DREB1B2* (*Bra022770*), *DREB1C1* (*Bra010463*), *DREB1C2* (*Bra019162*), *DREB1D* (*Bra028290*), *DREB1F1* (*Bra026963*), *DREB1F3* (*Bra016763*)	
					Low temperature: 4˚C	6	*DREB1A* (*Bra010461*), *DREB1B1* (*Bra010460*), *DREB1B2* (*Bra022770*), *DREB1C1* (*Bra010463*), *DREB1D* (*Bra028290*), *DREB1F1* (*Bra026963*)	
B. rapa	*LIN-11, ISL-1 and MEC-3 Domain Gene* (*LIM*)	13	22	WGT, TD	**pH:** 5, 7	4	*WLIM1a* (*Bra019956*), *WLIM1b* (*Bra018454*), *WLIM2b* (*Bra007586*), *WLIM2c* (*Bra033233*)	Park et al., 2014

					Drought: ABA (100 μM)	7	*WLIM1a* (*Bra019956*), *WLIM1b* (*Bra018454*), WLIM2a (*Bra014447*), *WLIM2b* (*Bra007586*), *WLIM2c* (*Bra033233*)**,** *WLIM2d* (*Bra032650*), *LIM14* (*Bra016514*)	
					Low temperature: 4˚C	6	*WLIM1a* (*Bra019956*), *WLIM1b* (*Bra018454*), *WLIM2c* (*Bra033233*)**,** *WLIM2d* (*Bra032650*), *WLIM2e* (*KJ686594b*), *LIM14* (*Bra016514*)	
B. rapa	*CCCH Zinc Finger Family* (*CCCH*)	68	103	WGD, WGT, TD, tetraploidization	**Salt:** NaCl (250 mM)	4	*BraA10g029760*, *BraA05g005940*, *BraA08g035210*, *BraA05g005940*	Pi et al., 2018
					Drought: ABA (100 μM), 300mM mannitol	5	*BraA07g001000*, *BraA01g008400*, *BraA03g054960*, *BraA09 g046730*, *BraA09g020370*	
B. rapa	*Superoxide Dismutase* (*SOD*)	7	18	WGD	**Drought:** PEG8000 (20%)	5	*FSD3b* (*Scaffold000004.555*), *CSD1a* (*BraA06001242*), *CSD1b* (*BraA09006623*), *CSD2a* (*BraA04002032*), *CSD3* (*BraA10001953*)	Verma et al., 2019
B. rapa cultivar *Chiifu-401-42*	*Sugars Will Eventually Be Exported Transporter* (*SWEET*)	17	34	WGD, WGT, TD	**Salt:** NaCl (200 mM)	3	*SWEET1-LF* (*Bra017916*), *SWEET1-MF1* (*Bra016421*), *SWEET11-MF1* (*Bra029914*)	Wei et al., 2019
					Low temperature: 4˚C	3	*SWEET1-LF* (*Bra017916*), *SWEET1-MF1* (*Bra016421*), *SWEET11-MF1* (*Bra029914*)	
					High temperature: 38 ˚C	2	*SWEET11-MF1* (*Bra029914*), *SWEET11-MF2* (*Bra019560*)	
B. rapa cultivar *YID-1*	*Annexin* (*ANN*)	8	13	WGT, TD, SD	**Salt:** NaCl (100 mM)	3	*Bra034402*, *Bra024346*, *Bra000090*, *Bra008892*	Yadav et al., 2015
					Drought: ABA (100 μM)	3	*Bra034402*, *Bra017102*, *Bra009048*	
					Oxidative: H_2O_2 (10 mM)	4	*Bra034402*, *Bra024346*, *Bra017102*, *Bra000090*	

B. rapa	*TIFY*	18	36	WGD, SD	**Salt:** NaCl (100 mM)	16	*TIFY3a* (*Bra002338*), *TIFY6a* (*Bra021281*), *TIFY6b* (*Bra022254*), *TIFY7b* (*Bra007937*), *TIFY7c* (*Bra016193*), *TIFY9a* (*Bra006190*), *TIFY9b* (*Bra008846*), *TIFY9c* (*Bra023399*), *TIFY10c* (*Bra015880*), *TIFY10d* (*Bra016520*), *TIFY10e* (*Bra031065*), *TIFY10f* (*Bra025713*), *TIFY11a* (*Bra008033*), *TIFY11b* (*Bra016056*), *TIFY11c* (*Bra016604*), *TIFY11e* (*Bra030986*)	Saha et al., 2016
					Drought: ABA (100 μM)	11	*TIFY3a* (*Bra002338*), *TIFY3b* (*Bra020135*), *TIFY6b* (*Bra022254*), *TIFY9a* (*Bra006190*),*TIFY9b* (*Bra008846*), *TIFY9c* (*Bra023399*), *TIFY10b* (*Bra008172*), *TIFY10c* (*Bra015880*), *TIFY10f* (*Bra025713*), *TIFY11e* (*Bra030986*)	
					Low temperature: 4˚C	18	*TIFY1a* (*Bra013805*), *TIFY1b* (*Bra019227*), *TIFY1c* (*Bra031379*), *TIFY4a* (*Bra036885*), *TIFY4b* (*Bra039720*), *TIFY6a* (*Bra021281*), *TIFY6b* (*Bra022254*), *TIFY7a* (*Bra003947*), *TIFY8a* (*Bra011357*), *TIFY8b* (*Bra040086*), *TIFY9c* (*Bra023399*), *TIFY10b* (*Bra008172*), *TIFY10c* (*Bra015880*), *TIFY10f* (*Bra025713*), *TIFY11a* (*Bra008033*), *TIFY11b* (*Bra016056*), *TIFY11d* (*Bra025977*), *TIFY11e* (*Bra030986*)	
B. juncea var. *tumida*	*Salt Overly Sensitive* (*SOS*)	6	12	WGD, allotetra ploidization	**Salt:** NaCl (200 mM)	6	*SOS1-1* (*A002024*), *SOS1-2* (*B027801*), *SOS2* (*B024452*), *SOS3-1* (*B038085I*), *SOS3-2* (*A046788*), *SOS3-3* (*A033133*)	Cheng et al., 2019
					Drought: ABA (50μM)	2	*SOS3-1* (*B038085*), *SOS4-1* (*A015607*)	
					Low temperature: 4˚C	7	*SOS3-1* (*B038085*), *SOS4-1* (*A015607*), *SOS4-2* (*B048173*), *SOS5-1* (*B019581*), *SOS5-2* (*A022634*), *SOS6-1* (*A037683*), *SOS6-1* (*B043857*)	

B. juncea var. *tumida*	*Pyrabactin Resistance 1 Like* (*PYL*)	14	25	WGD, allotetra ploidization	**Salt:** NaCl (200 mM)	5	*PYL5-4* (*BjuB040841*), *PYL5-5* (*BjuA040927*), *PYL6-2* (*BjuB042092*), *PYL8-3* (*BjuA006960*), *PYL8-4* (*BjuA015299*)	Cheng et al., 2019
					Drought: ABA (50μM)	8	*PYL4-1* (*BjuB020198*), *PYL5-4* (*BjuB040841*), *PYL5-5* (*BjuA040927*), *PYL6-1* (*BjuA010539*), *PYL6-2* (*BjuB042092*), *PYL6-3* (*BjuB042125*), *PYL7-3* (*BjuB017238*), *PYL8-3* (*BjuA006960*)	
					Low temperature: 4˚C	7	*PYL3* (*BjuB025977*), *PYL4-1* (*BjuB020198*), *PYL5-5* (*BjuA040927*), *PYL6-1* (*BjuA010539*), *PYL7-3* (*BjuB017238*), *PYL8-3* (*BjuA006960*), *PYL8-4* (*BjuA015299*)	
B. juncea	*Superoxide Dismutase* (*SOD*)	7	29	WGD, allotetra ploidization	**Drought:** PEG8000 (20%)	5	*ACSD4* (*BjuA031027*), *AFSD3b* (*BjuA032504*), *AFSD3d* (*BjuA023518*), *AFSD2a* (*BjuA010212*), *BCSD8* (*BjuB033224*)	Verma et al., 2019
					High temperature: 40˚C	3	*AFSD3d* (*BjuA023518*), *BCSD8* (*BjuB033224*), *ACSD4* (*BjuA031027*)	
B. napus	*Phospholipase D* (*PLD*)	12	32	TD, allotetra ploidization	**Salt:** NaCl (200 mM)	5	*PLDα1A1* (*BnaA01g28530D*), *PLDα1A5* (*BnaA05g23740D*), *PLDα1C1* (*BnaC01g35830D*), *PLDα1C5* (*BnaC05g37540D*), *PLDδC7* (*BnaC07g45790D*)	Lu et al., 2019
					Drought: ABA (25 μM)	6	*PLDα1C1* (*BnaC01g35830D*), *PLDδAnn* (*BnaAnng08250D*), *PLDβ1A10* (*BnaA10g00160D*), *PLDβ1C5* (*BnaC05g00230D), PLDζ1A5* (*BnaA05g23110D*), *PLDζ1C5* (*BnaC05g36550D*)	
					Low temperature: 4˚C	6	*PLDδC7* (*BnaC07g45790D*), *PLDδAnn* (*BnaAnng08250D*), *PLDβ1A10* (*BnaA10g00160D*), *PLDβ1C5* (*BnaC05g00230D), PLDζ1A5* (*BnaA05g23110D*), *PLDζ1C5* (*BnaC05g36550D*)	

B. napus	*Nuclear Factor Y (NF-Y)*	10	33	Not known	**Salt:** NaCl (150 mM)	17	*NF-YA1, NF-YA2, NF-YA4, NF-YA5, NF-YA7, NF-YA8, NF-YA9, NF-YA10, NF-YA11, NF-YB10, NF-YB13, NF-YB3, NF-YB7, NF-YB8, NF-YB9, NF-YB10, NF-YB14*	Xu et al., 2014
					Drought: PEG6000 (15%), ABA (100 μM)	13	*NF-YA8, NFYA9, NF-YA10, NF-YA11, NF-Y12, NF-YA13, NF-YB7, NF-YB10, NF-YB14, NF-YC1, NF-YC2, NF-YC4, NF-YC5*	

Table 3. Important resources useful for *Brassica* genome research.

Purpose	Brassicaceae species	Resource
Genome database	*Arabidopsis thaliana, A. lyrata , Aethionema arabicum , Brassica rapa, B. oleracea , B. napus , B. juncea Camelina sativa, Capsella rubella , Leavenworthia alabamica , Sisymbrium irio, Schrenkiella parvula , Thellungiella halophile, T. salsuginea*	*Brassica* Database (http://brassicadb.org/brad/)
Genome database	*B. rapa*	*Brassica rapa* genome database (http://www.plantgdb.org/BrGDB/)
Genome database	*B. napus*	*Brassica napus* genome browser (http://www.genoscope.cns.fr/brassicanapus/)
Genome database	*B. oleracea*	Bol base (http://ocri-genomics.org/bolbase/)
Agrobacterium mediated transformation	*B. napus*, B. *oleracea* *B.rapa* *B. juncea*	Bhalla and Singh 2008 Baskar et al., 2016 http://agris.fao.org/agris-search/search.do?recordID=CN2014001141
Vernalization-infiltration method of *in planta* transformation	*B. rapa*, *B. napus*	Bai et al., 2013
EMS TILLING resource	*B. rapa*	Stephenson et al., 2010
EMS TILLING resources	*B. napus*	Wang et al., 2008
EMS TILLING resources	*B. oleracea*	Himelblau et al., 2009
TILLING of mutagenized Brassicaceae populations, targeted	*Brassica napus*, *Brassica oleracea*	RevGen (https://www.jic.ac.uk/technologies/

gene knockouts using RNA-guided Cas9, transformation		genomic-services/revgenuk-tilling-reverse-genetics/)
Site-directed mutagenesis by CRISPR/Cas9 method	*B. carinata*	Kirchner et al., 2017
RNA-guided Cas9 mediated genome editing	*B. oleracea*	Lawrenson et al., 2015
CRISPR/Cas9 mediated genome editing	*B. napus*	Yang et al., 2017
Germplasms	*B. cretica* (Heldr. & Halácsy) Snogerup & al. *B. gravinae* Ten. *B. oleracea* L. *B. oxyrrhina* Cossin *B. souliei* (Batt.) Batt.	https://brassibase.cos.uni-heidelberg.de/
Oligonucleotide array	*B. napus* *B. juncea*	Huang et al., 2014 Srivastava et al., 2015b
SNP array platform	*B. napus*	Dalton-Morgan et al., 2014
Carbohydrate microarray	*B. napus*	Wood et al., 2017