



OPEN ACCESS

EDITED BY

Pradeep Sharma,
Indian Institute of Wheat and Barley
Research (ICAR), India

REVIEWED BY

Amit Kumar,
Central Sericultural Research &
Training Institute (CSRTI), India
Mansi Sharma,
Dalhousie University, Canada

*CORRESPONDENCE

Upendra Kumar
baliyan.upendra@gmail.com
Vijai Malik
gathwalajai@gmail.com

SPECIALTY SECTION

This article was submitted to
Plant Abiotic Stress,
a section of the journal
Frontiers in Plant Science

RECEIVED 05 September 2022

ACCEPTED 07 October 2022

PUBLISHED 31 October 2022

CITATION

Kumar V, Sharma H, Saini L, Tyagi A,
Jain P, Singh Y, Balyan P, Kumar S,
Jan S, Mir RR, Djalovic I, Singh KP,
Kumar U and Malik V (2022)
Phylogenomic analysis of 20S
proteasome gene family reveals
stress-responsive patterns in rapeseed
(*Brassica napus* L.).
Front. Plant Sci. 13:1037206.
doi: 10.3389/fpls.2022.1037206

COPYRIGHT

© 2022 Kumar, Sharma, Saini, Tyagi,
Jain, Singh, Balyan, Kumar, Jan, Mir,
Djalovic, Singh, Kumar and Malik. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

Phylogenomic analysis of 20S proteasome gene family reveals stress-responsive patterns in rapeseed (*Brassica napus* L.)

Vivek Kumar¹, Hemant Sharma², Lalita Saini¹, Archasvi Tyagi¹, Pooja Jain¹, Yogita Singh³, Priyanka Balyan⁴, Sachin Kumar², Sofora Jan⁵, Reyazul Rouf Mir⁵, Ivica Djalovic⁶, Krishna Pal Singh^{7,8}, Upendra Kumar^{3*} and Vijai Malik^{1*}

¹Department of Botany, Chaudhary Charan Singh University, Meerut, UP, India, ²Department of Genetics and Plant Breeding, Chaudhary Charan Singh University, Meerut, UP, India, ³Department of Molecular Biology & Biotechnology, College of Biotechnology, Chaudhary Charan Singh (CCS) Haryana Agricultural University, Hisar, India, ⁴Department of Botany, Deva Nagri Post Graduate (PG) College, Chaudhary Charan Singh (CCS) University, Meerut, India, ⁵Division of Genetics and Plant Breeding, Faculty of Agriculture, Sher-e-Kashmir University of Agricultural Sciences & Technology (SKUAST)-Kashmir, Wadura, India, ⁶Institute of Field and Vegetable Crops, National Institute of the Republic of Serbia, Maxim Gorki, Novi Sad, Serbia, ⁷Biophysics Unit, College of Basic Sciences & Humanities, Govind Ballabh (GB) Pant University of Agriculture & Technology, Pantnagar, India, ⁸Vice-Chancellor's Secretariat, Mahatma Jyotiba Phule Rohilkhand University, Bareilly, India

The core particle represents the catalytic portions of the 26S proteasomal complex. The genes encoding α - and β -subunits play a crucial role in protecting plants against various environmental stresses by controlling the quality of newly produced proteins. The 20S proteasome gene family has already been reported in model plants such as Arabidopsis and rice; however, they have not been studied in oilseed crops such as rapeseed (*Brassica napus* L.). In the present study, we identified 20S proteasome genes for α - (PA) and β -subunits (PB) in *B. napus* through systematically performed gene structure analysis, chromosomal location, conserved motif, phylogenetic relationship, and expression patterns. A total of 82 genes, comprising 35 *BnPA* and 47 *BnPB* of the 20S proteasome, were revealed in the *B. napus* genome. These genes were distributed on all 20 chromosomes of *B. napus* and most of these genes were duplicated on homoeologous chromosomes. The *BnPA* (α 1-7) and *BnPB* (β 1-7) genes were phylogenetically placed into seven clades. The pattern of expression of all the *BnPA* and *BnPB* genes was also studied using RNA-seq datasets under biotic and abiotic stress conditions. Out of 82 *BnPA/PB* genes, three exhibited high expression under abiotic stresses, whereas two genes were overexpressed in response to biotic stresses at both the seedling and flowering stages. Moreover, an additional eighteen genes were expressed under normal conditions. Overall, the current findings developed our understanding of the organization of the 20S proteasome genes in *B. napus*,

and provided specific *BnPA/PB* genes for further functional research in response to abiotic and biotic stresses.

KEYWORDS

Brassica napus, 20S proteasome, phylogenomics, digital expression, environmental stress

1 Introduction

Rapeseed (*Brassica napus* L.; $2n=38$; AACC genome) is an allotetraploid species that is derived through interspecific hybridization of turnip rape (*B. rapa*; $2n=20$; AA) with cabbage (*B. oleracea*; $2n=18$; CC) (Peng et al., 1999; Chalhoub et al., 2014). Globally, *B. napus* is one of the major oil crops and ranks third in vegetable oil production. It is commonly used as edible oil (Maršalkienė et al., 2009; Lozano-Baena et al., 2015), condiments (Gilbert et al., 2012; Wiersema and Leon, 2016) and fodder (Cartea et al., 2005) in many parts of the world. In addition, canola oil has medicinal properties including a diuretic (Lust, 1983; Grieve, 1984), analgesic and anticancer activities (Duke, 1983). The crop is adapted to the temperate regions of the world and is therefore, very sensitive to a series of environmental threats such as fungal/bacterial diseases, and cold, heat, and drought stresses. These stresses limit the growth and development at different stages during the crop cycle (Kutcher et al., 2010). The presence of high-quality genome sequences and bioinformatics toolkits may help to identify various genes or gene families to elucidate their functional relevance to environmental stress factors to breed climate-smart cultivars.

The ubiquitin-mediated proteolysis system (UPS) also known as the ubiquitin-proteasome pathway (UPP), regulates the degradation of many proteins in eukaryotic cells (Vierstra, 2009). The UPS consists of two distinct, consecutive steps: ubiquitylation (identification of a substrate) and proteasomal degradation (elimination of the ubiquitinated protein) (Kleiger and Mayor, 2014; Sharma et al., 2016). Briefly, ubiquitination is a complicated process that usually necessitates the use of three enzymes: E1, E2, and E3. Ubiquitin is covalently attached to substrate proteins catalyzed by a cascade of enzymes consisting of ubiquitin activator (E1), conjugase (E2), and ligase (E3). Ubiquitin is commonly conjugated to an internal lysine (Lys) residue but can also be conjugated to the free amino terminus of the substrate *via* its carboxy-terminal glycine. A polyubiquitin chain generated by multiple rounds of ubiquitylation can serve as a signal for degradation by the 26S proteasome machinery, a multiprotein complex consisting of a 20S core particle and 19S regulatory particles (Kleiger and Mayor, 2014). The proteasome has two main origins: the primitive form seen in *Thermoplasma*

acidophilum (Archaea bacteria) and the evolutionarily improved form found in yeast, plants, and mammals. Numerous investigations in eukaryotes have been conducted to understand the dynamics of the proteasomal complex during the last decade (Dantuma and Bott, 2014; Stone, 2019).

The genetics of 20S proteasome core particles in plant species, including Arabidopsis (Yang et al., 2004), rice (Fu et al., 1998) and wheat (Sharma et al., 2022), have been well studied. Morphologically, the 20S core particle is barrel-shaped, and is composed of 28 nonidentical subunits arranged in 4 axially stacked rings. The rings at the two ends are identically formed by 7α subunits, and the rings in the middle are identically formed by 7β subunits. This gives rise to the 20S core particle having a symmetric configuration of $\alpha 1-7/\beta 1-7/\beta 1-7/\alpha 1-7$ (Chen and Hochstrasser, 1996; Wolf and Hilt, 2004). The β -subunits $\beta 1$, $\beta 2$, and $\beta 5$ have proteolytic properties for different substrates. The protein enters the 20S core particle through α -subunits present at the top, and after processing this protein comes out through the α -subunits located at the bottom of the core particle (Chen and Hochstrasser, 1996). In Arabidopsis and many eukaryotes, there are 7α - ($\alpha 1-7$) and 7β -subunits ($\beta 1-7$). Based on nomenclature, these 7α - and 7β -subunits can be represented as proteasome alpha A-G and proteasome beta A-G (i.e., PAA-PAG and PBA-PBG) respectively. There are 23 genes encoding 12 α - and 11 β -subunits in Arabidopsis (Fu et al., 1998). Of these subunits, one gene encodes four subunits (1α and 3β) whereas two genes encode the rest of the subunits. In common wheat (*Triticum aestivum*), 67 members of the 20S proteasome α (*TaPA*) and β (*TaPB*) gene family have recently been discovered (Sharma et al., 2022). These 67 *TaPA* and *TaPB* genes were distributed in all 21 chromosomes of wheat and some of them were found to be involved in heat/drought stress tolerance.

Phylogenomics uses genome-wide data to infer the evolution of genes, genomes, and the tree of life. (Eisen, 1998). The term conventional phylogenetics is based upon the study of a few genes or morphology whereas phylogenomics is an update of the term phylogenetics and concentrates on genome-level analysis (DeSalle et al., 2020). According to the principle of phylogenomics, orthologous sequences conserve more protein function than paralogous sequences. The most commonly used data in phylogenomics are sequence data including whole

genome nucleotide sequences, orthologous genomic blocks, core genomes, core coding genomes, exons, introns, and other conserved biological molecules. The present study involves a phylogenomic analysis of the *B. napus* 20S proteasome gene family and its putative role in biotic and abiotic stresses.

2 Materials and methods

2.1 Analysis of gene sequence

2.1.1 Identification and retrieval of 20S proteasome gene sequences

Since 20S proteasome family genes have already been characterized in *Arabidopsis thaliana* and rice (*Oryza sativa*), the DNA and protein sequences of *AtPA/PB* and *OsPA/PB* were used as a reference to detect the true orthologs in the *B. napus* genome. Criteria for the identification of true orthologs that were developed (Dhaliwal et al., 2014) were strictly followed in the present analysis. The coding DNA sequences of 20S proteasome family genes related to *A. thaliana* and rice were retrieved from the Ensembl Plants database (<https://plants.ensembl.org/index.html>). Consequently, coding sequences of 24 genes of *A. thaliana* and 23 genes of rice were subjected to tBLASTx and BlastP searches against the *B. napus* genome assembly available at the Ensembl Plants database (https://plants.ensembl.org/Brassica_napus/Info/Index?db=core).

2.1.2 Structural features of the 20S proteasome gene family

The tools of Ensembl Plants (<https://plants.ensembl.org/index.html>), Multiple Em for Motif Elicitation Suite Version 5.4.1 (MEME v5.4.1; <https://meme-suite.org/meme/tools/meme>) and TB tools (<https://github.com/CJ-Chen/TBtools/releases>) were employed for gene structure analysis. RepeatMaskerv4.0.9 (<https://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) and BatchPrimer3 v1.0 (<http://probes.pw.usda.gov/batchprimer3>) were used to identify transposable elements (TEs) and microsatellites or simple sequence repeats (SSRs) in gene sequences, respectively. The Plant Care database was used to search for the presence of cis-acting regulatory elements in gene sequences, 1500 base pairs (bp) upstream of the promoter region (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>). The plant small RNA target analysis server psRNATarget (<https://www.zhaolab.org/psRNATarget/>) was used to search for putative microRNAs and their targets in the *BnPA/BnPB* genes of *B. napus*. Here, 0-3 e-value was used (Dai and Zhao, 2011; Dai et al., 2018). All the abovementioned analyses were performed with default parameters.

2.1.3 Gene duplication and synteny analysis

The gene tree pipeline available in the Ensembl Plants database (Vilella et al., 2009; Bolser et al., 2015) was utilized to

infer evolutionary relationships among proteasome genes (*BnPA* and *BnPB*) with the help of a gene identifier. A phylogenetic tree of homologous genes across the genome of *B. napus* and *Arabidopsis* was constructed using the Plant Compara option. This gene tree was further utilized to identify duplication and speciation events. Synteny and collinearity between chromosomes of *B. napus* and *Arabidopsis* were determined covering the stretch of 25 genes. For this purpose, the genome browser biotool GENOMICUSv49.01 (<https://www.genomicus.bio.ens.psl.eu/genomicus-plants-49.01/cgi-bin/search.pl>) was used.

2.2 Analysis of protein sequence

2.2.1 Structure, conserved motifs and physio-chemical properties of protein sequences

The CD-search program of the conserved domain database (CDD) at NCBI was used to identify domain features in *BnPA/BnPB* protein sequences. The physiochemical properties, including amino acid composition, molecular weight, theoretical PI, number of positively/negatively charged residues, instability index, aliphatic index, and grand average of hydropathy (GRAVY) were computed using ExPASy's ProtParam tool (<https://web.expasy.org/protparam/>). The network protein sequence (NPS) analysis was performed using a self-optimized prediction method with an alignment (SOPMA) tool (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). Motifs in protein sequences were searched using the MEME suite (<https://meme-suite.org/meme/tools/meme>). Annotation and visualization of identified motifs were performed using InterPro Scan (<https://www.ebi.ac.uk/interpro/search/sequence/>) and TB tool, respectively.

2.2.2 Computational analysis, structure and validation of predicted proteins

Homology modeling (HM) was applied to deduce the 3D structure of the predicted proteins. For HM, PSI-BLAST was performed against two databases, the Swiss Model template library (<https://swissmodel.expasy.org/>), and the protein data bank (<http://www.rcsb.org/pdb/home/home>) with 100 iterations. A structure analysis and validation server (SAVES; <https://saves.mbi.ucla.edu/>) was used to verify the predicted 3D structures of *BnPA* and *BnPB* proteins. The relative proportion of amino acids that appear in the favored region was found with the help of the PROCHECK option of SAVES v6.0 (Laskowski et al., 1993). VERIFY 3D was utilized to determine the compatibility of the atomic model (3D) with its amino acid sequence (1D) (Eisenberg et al., 1997). The ERRAT program was used to verify the protein structures through patterns of nonbonded interactions among C, N, and O atoms (Colovos and Yeates, 1993).

2.2.3 Functional annotation and superimposition of 3D structures

The Flexible structure Alignment by Chaining Aligned Fragment pairs allowing Twists (FATCAT) server (https://fatcat.godziklab.org/fatcat/fatcat_pair.html) was used to compare the 3D structure of the proteins encoded by various genes belonging to *A. thaliana* with that of predicted *B. napus* proteins by aligning the representative structure. The similarity of 3D structures was measured by the root mean square deviation (RMSD) value of the C α atoms.

2.3 Phylogenetic analysis

The MultAlin program (<http://multalin.toulouse.inra.fr/multalin/>) was used for multiple sequence alignments to determine conserved amino acid residues in *B. napus*, Arabidopsis and rice. A mutual information server to infer coevolution (MISTIC; <http://mistic.leloir.org.ar/results.php?jobid=202112021211022296>) was employed to determine the mutual information (MI) between two amino acid positions in multiple sequence alignment (MSA). The identical amino acid between two positions (homologous protein), called mutual information (MI), was predicted to correlate and compensate for mutations, and was used to identify coevolving residues (Sharma et al., 2022). The amino acid sequences of proteins were utilized for phylogenetic analysis by molecular evolutionary genetics analysis version 6.0 (MEGA v6.0) software using the neighbor-joining method (Tamura et al., 2011). The Newick format tree prepared in MEGA was visualized in iTOL (<https://itol.embl.de/>).

2.4 *In silico* expression profile of 20S proteasome genes

Digital expression of *BnPA* and *BnPB* genes was analyzed using transcriptomic data available at the *Brassica expression database* (<https://brassica.biodb.org/>). Expression was studied in FPKM values using six different tissues (seed, cotyledon, hypocotyls, radical, root, & leaf) belonging to the germination and seedling stages under normal conditions. Under both biotic and abiotic stress conditions, expression was observed at the seedling and flowering stages. Under both normal and stress conditions, the expression pattern of the selected genes was studied.

3 Result

3.1 Analysis of gene sequence

3.1.1 Identification of genes for the 20S proteasome in *B. napus*

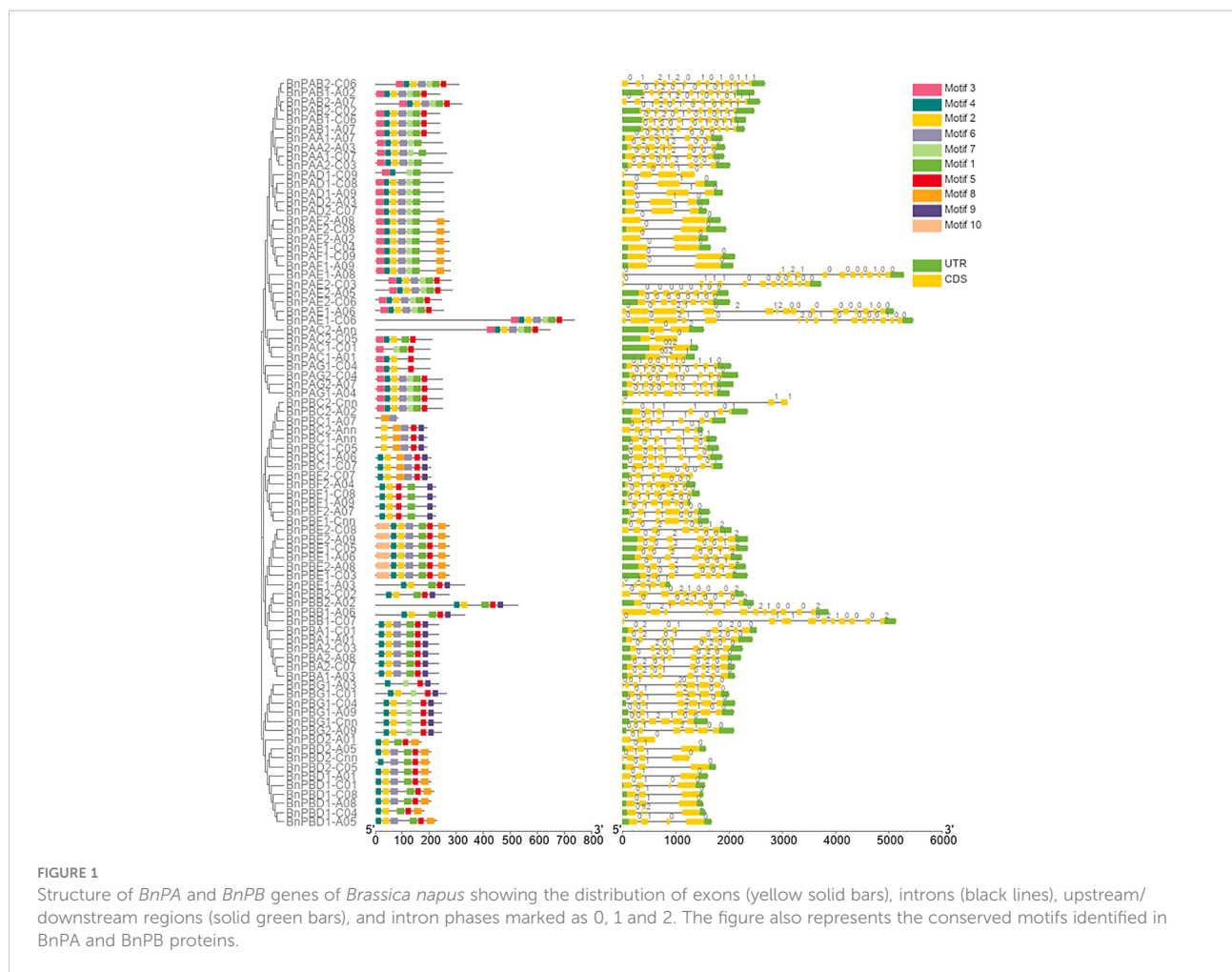
Out of 92 20S proteasome genes initially identified in the *B. napus* genome, 10 genes were incomplete; therefore, these genes

were removed from further analysis. The remaining 82 full-length genes involving 35 *BnPA* and 47 *BnPB* of the 20S proteasome family were further characterized in the present study. Comprehensive information on the full-length gene and coding sequences of *B. napus* α -subunits (*BnPAA-BnPAG*) and β -subunits (*BnPBA-BnPBG*) and their comparison with corresponding *A. thaliana* and rice genes are provided in **Supplementary Table S1**. The name of all 82 *B. napus* genes of the 20S proteasome family was designated according to the corresponding genes reported earlier for Arabidopsis, rice and wheat (Fu et al., 1998; Sassa et al., 2000; Sharma et al., 2022). The sizes of the *BnPA* and *BnPB* genes varied from 1039-5449 bp and 618-5132 bp, respectively. The cDNA sequence of *BnPA* genes ranged from 852 to 2355 bp and that of *BnPB* genes ranged from 249 to 1818 bp. Furthermore, variations in coding DNA sequence (CDS) were also observed individually in *BnPA* (606-2199 bp) and *BnPB* (249-1578 bp) genes (**Supplementary Table S1**). The intron-exon features of *B. napus* 20S proteasome genes were studied to gain an improved understanding of their structural patterns. The number of exons and introns in *BnPA* genes varied from 2-17 and 1-16, respectively. All 35 *BnPA* genes had introns. The number of exons and introns in *BnPB* genes varied from 3-12 and 2-11, respectively (**Supplementary Table S2**). The intron phases were phase 0 (56.75%), phase 1 (30.45%) and phase 2 (12.79%) (**Figure 1**).

3.1.2 Assignment of chromosome and gene duplication

All 82 *BnPA* and *BnPB* genes were physically mapped to 18 individual chromosomes of *B. napus* along with two additional Cnn and Ann chromosomes. The maximum number of six genes were located on each A07, A09, and C07, while the minimum number of two genes each was located on A04, A02, and C09. All the genes were located in the terminal and subterminal positions (**Figure 2**). The tree involving *B. napus* genes along with genes from other taxa was developed using the Ensembl Plants Compara pipeline (**Supplementary Figure S1**). This was done to examine orthology and paralogy among 35 *BnPA* and 47 *BnPB* genes.

Detailed information on the orthologous and paralogous relationships of the *BnPA* and *BnPB* genes with different species is given in **Supplementary Table S1**. Orthologs and paralogs among Brassica, Arabidopsis and rice 20S proteasome genes have been investigated using the order of genes of 14 subunits (7- α and 7- β). Out of the 24 genes in Arabidopsis (13- α and 11- β), the number of duplicated genes was 10 and those that remained single were 4. Out of the 23 rice genes (13- α and 10- β), duplication was observed in seven genes, triplication in 1, and the rest 6 genes remained single (**Supplementary Table S1**). The observed pattern of duplicated genes in *B. napus* was similar to that in Arabidopsis and rice for the corresponding *PA/PB* genes. The orthologs were distinguished from paralogs using Ensembl Plant Compara.



3.1.3 Synteny between *B. napus*, *A. thaliana*, and rice

Synteny analysis of 82 *BnPA* and *BnPB* genes with the corresponding Arabidopsis genes was highly conserved (Figure 3). The collinearity of *BnPA* and *BnPB* genes with Rice was negligible; however, collinearity with *A. thaliana* was found with only one gene, i.e., *BnPBC2-Cnn* with *At1*.

3.1.4 SSRs in *BnPA* and *BnPB* genes

A total of 66 SSRs were identified in 45 (54.87%) of the 82 genes. Of these, 28 SSRs were identified in 16 *BnPA* genes and 38 SSRs were identified in 29 *BnPB* genes. The SSR numbers vary per gene. There were only six SSRs in *BnPAE1-A08*, only four SSRs in *BnPBB1-C07*, three SSRs in two genes (*BnPAE2-C03*, *BnPBD2-C05*), and two SSRs each in nine genes (*BnPAA2-A03*, *BnPAE1-C06*, *BnPAE2-A05*, *BnPAF1-C09*, *BnPAG2-C04*, *BnPBA1-A01*, *BnPBC1-A06*, *BnPBD2-Cnn* and *BnPBF2-A07*), whereas single SSRs were present in the remaining 32 genes (Supplementary Table S3). SSRs with dinucleotide motifs (20) are more frequent followed by tetranucleotide (10) and

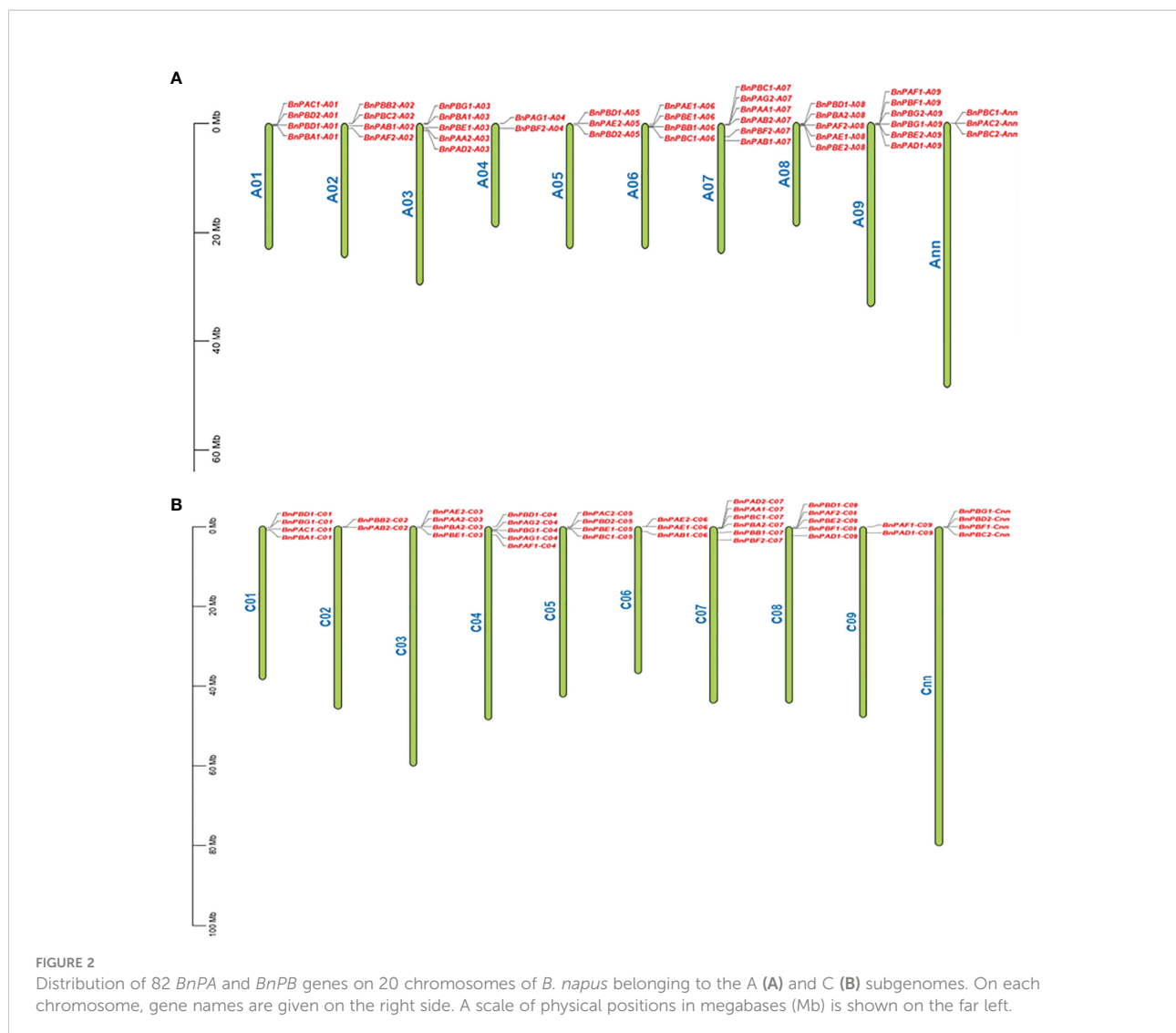
hexanucleotide (10), trinucleotide and mononucleotide (9), pentanucleotide (5) and heptanucleotide (3) motifs.

3.1.5 Analysis of promoter and cis-acting regulatory elements

Cis-acting regulatory elements present 1500 bp upstream of the 5' promoter sequence were studied in each of 82 *BnPA/BnPB* genes. It was observed that the promoter region of all the genes contains many light-responsive elements along with phytochrome and other cis-acting elements. The maximum number of genes was for light, methyl jasmonate (MeJA), and endosperm elements (Supplementary Figure S2).

3.1.6 MicroRNAs and their targets in *BnPA* and *BnPB* genes

A total of 44 microRNAs (miRNAs) were found. These miRNAs had their targets in 21 genes (8 *BnPA* and 13 *BnPB*). The maximum number of target sites i.e., six microRNAs were available for each of two genes (*BnPAC1-A01* and *BnPBD2-A01*). The targets for *BnPBD1-C08*, *BnPBD1-A08* and *BnPBD2-Cnn* were available



for four miRNAs. The target for *BnPAF2-C08* was available for three miRNAs. Similarly, the targets for *BnPBA1-A01* and *BnPBA2-C07* were available for two miRNA each. The target for the remaining 13 miRNAs was available in different *BnPA* and *BnPB* genes distributed on several Brassica chromosomes (Supplementary Table S4).

3.2 *In silico* expression analysis of *BnPA* and *BnPB* genes

The expression pattern of genes was studied under normal and stress conditions. Under normal conditions, two different developmental stages (germination and seedling) for six different organs (seed, cotyledon, hypocotyl, radicle, leaf, and root) were examined. Under both abiotic and biotic stress conditions, two

different developmental stages (seedling both flowering) were examined.

3.2.1 Tissue-specific expression under normal conditions

Under two different developmental stages (germination and seedling), the six different organs (seed, cotyledons, hypocotyl, radical, leaf, and root) showed variation in expression (Figure 4). Out of 82, 18 genes showed higher expression in the two stages and the expression data are provided in the form of FPKM values.

3.2.2 Abiotic stresses

Cold stress. We observed two genes under 24 h cold stress (seedling and flowering stages) whose expression was upregulated i.e., >1.70-fold change (FC). During this same duration, the expression of 33 genes ranged from 0.00 to 0.99 FC i.e., the expression of these genes in response to cold stress was very poor (Figure 5A).

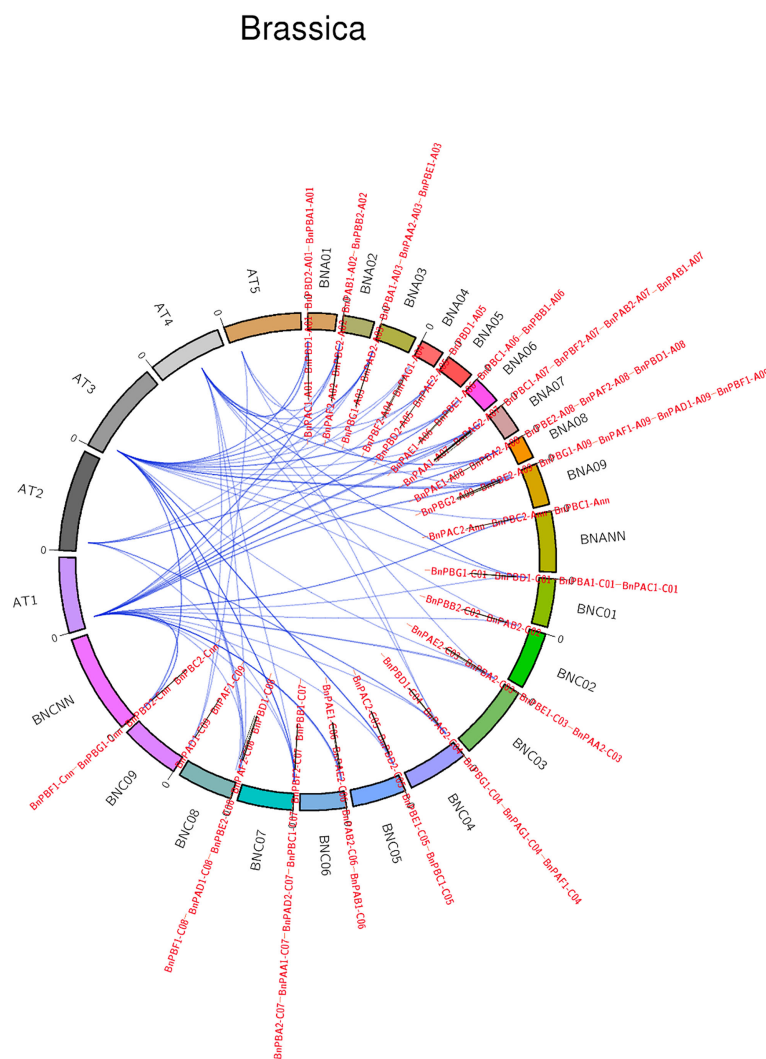


FIGURE 3
A circular map showing synteny among PA and PB genes of *B. napus*, *A. thaliana* and rice (name of chromosomes given outside the circle, name of genes given outside and inside the circle).

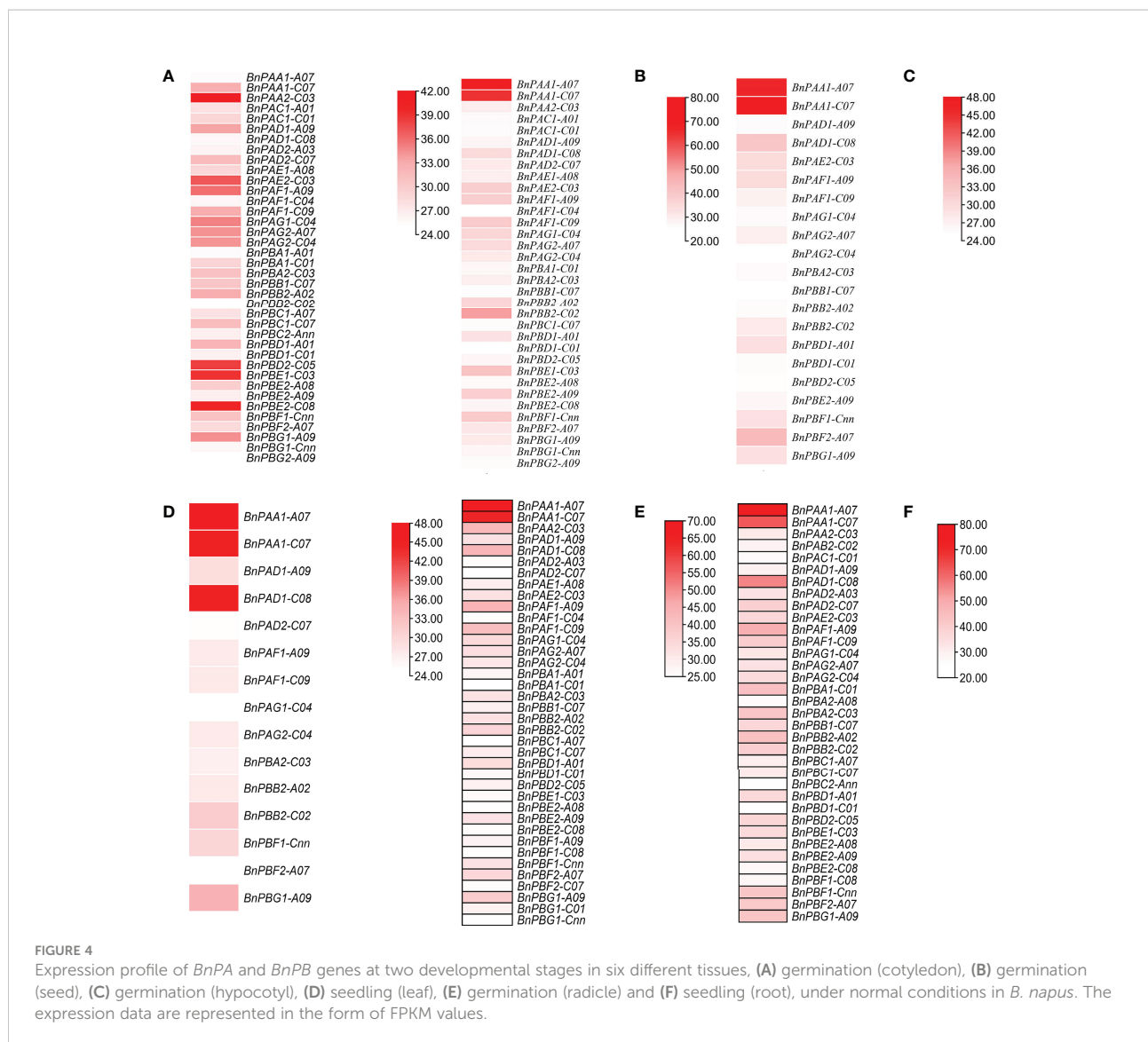
Drought stress. Only one gene showed high expression ($FC > 1.70$) under 24 h drought stress (seedling and flowering stages), whereas 62 genes showed their expression from 0.00 to 0.99 FC (Figure 5B).

Heat Stress. Under 24 h of heat stress, all genes (seedling and flowering stages) showed expression $FC < 1.70$. However, the expression of 57 genes ranged from 0.00 to 0.80.

As a result, we concluded that in *B. napus*, there are only five genes that can tolerate cold stress and there is only one gene that is expressed under drought conditions. Moderate to poor expression of genes was observed under heat stress conditions.

3.2.3 Biotic stresses

Under biotic stress in *B. napus*, 20S proteasome genes were expressed against two fungal pathogens *Leptosphaeria maculans* (blackleg) and *Sclerotinia sclerotiorum* (white mold), at the seedling and flowering stages, respectively. After 264 h of *L. maculans* infection at the seedling stage, one gene with $FC > 0.99$ was expressed against fungal infection (Figure 5C). Relative to the control, 1 gene was stimulated ($FC > 1.70$) after 72 h of infection at the seedling stage. At the flowering stage, BnPAB1-A07 and BnPBA2-A08 genes were expressed with $FC > 0.99$ after 48 h and 96 h of *S. sclerotiorum* inoculation, respectively (Figure 5D).



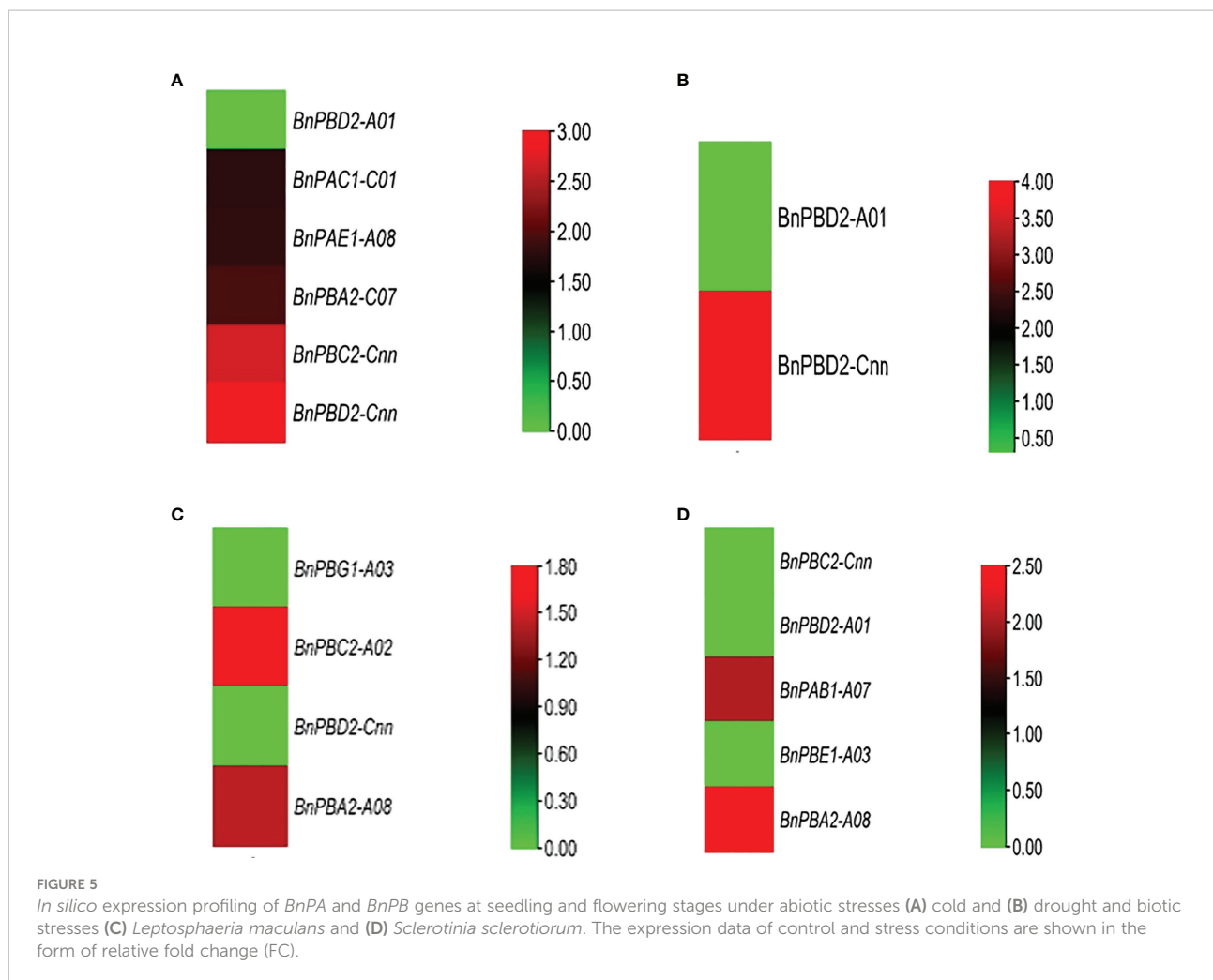
3.3 Analysis of Proteins

3.3.1 Characterization of BnPA and BnPB proteins

The molecular weight of BnPA and BnPB proteins ranged from 9510.19 to 82935.95. the isoelectric point (PI) ranged from 4.65 to 9.93. The total acidic and alkaline proteins were 71 and 11 respectively. The proteins (41) with a lower aliphatic index i.e., 67.30 to 98.52 were unstable, whereas the remaining proteins (41) had a higher aliphatic index of 79.60 to 100.36 and were stable. The grand average of hydropathy (GRAVY) ranged from -0.669 to 0.093 (Supplementary Table S5).

3.3.2 Functional domains and motifs of BnPA and BnPB proteins

The number of amino acids in 82 BnPA/BnPB proteins varied from 82 amino acids (aa) (BnPBC2-Cnn) to 732 aa (BnPAAE1-A06), the mean of which was 251.85 aa (Supplementary Table 2). The length of α -subunits ranged from 201 to 732 aa, and the β -subunits ranged from 82 to 525 aa (Supplementary Table 2). A total of 10 separate motifs are given in Figure 1. The logo of 10 separate motifs and the associated amino acids identified in the BnPA and BnPB proteins are provided in Supplementary Figure S3. The list of identified motifs along with their sequence and e-value are given in Supplementary Table S7.



3.3.3 Sequence alignment and assessment of conserved amino acids

The percent similarity among the α -subunit of *B. napus* was 30 (Supplementary Table S8) whereas in the β -subunit it was 20.72 (Supplementary Table S9). The similarity percentage between the α and β -subunits of *B. napus* was 21.33. The percentsimilarity among *B. napus* and Arabidopsis *BnPA* proteins was 91.06 (Supplementary Table S8) and for *BnPB* protein was 99.14 (Supplementary Table S9). High amino acid similarity was observed among the α and β -subunits of *B. napus*, Rice, and Arabidopsis (Supplementary Figures 4a, b). A high conservation pattern of 31 amino acids (aa) of α -subunits and 15 β -subunits was found among *B. napus*, rice, and Arabidopsis. These 31 (α -subunits) and 15 (β -subunits) residues also had the highest MI (Supplementary Figures S5, S6).

3.3.4 Localization of subcellular proteins and their functions

Gene ontology analysis and functional annotation suggested that *BnPA* and *BnPB* proteins have proteolytic functions and

other molecular functions such as threonine type endopeptidase activity, threonine type peptidase activity, peptidase activity and endopeptidase activity (Figure 6).

3.3.5 Protein structures

The secondary structure of all the proteins was compared. It has been found that the secondary structure is dominated by α -helices followed by random coils (Supplementary Table S10). The latter form irregular structures that permit polypeptide chains to fold uniquely. Proteins such as *BnPA* and *BnPB* tend to form a highly stable structure. *In-silico* 3D structures were determined for 20 (24.39%) *BnPA* and *BnPB* proteins with similarity ranges of 10.85-100% with the corresponding template of Arabidopsis (Supplementary Figure S7). The GMQE for these 20 proteins ranged from 0.59 to 0.82. A high-quality protein model was suggested by it. Q-mean value ranged from 0.73 ± 0.06 to 0.82 ± 0.06 . The similarity between the protein model and reference (Arabidopsis) proteins ranged from 42 to 69%. The quality factor was determined with the help of ERRAT and ranged from 88.2096 to 100%. The 3D-1D score was

determined with the help of VERIFY 3D and ranged from 77.95 to 99.58% (Supplementary Table S11)

3.3.6 Alignment and functional annotation of 3D-structure

The 3D protein structures (with minimum energy) of *B. napus* were superimposed with the reference protein (3D) of Arabidopsis (Figure 7 and Supplementary Table S12). The 3D structure of five (BnPAA1-A07, BnPAA1-C07, BnPAF1-A09, BnPBE1-C03, and BnPBF1-Cnn) proteins showed a similarity of 33 to 98% with the corresponding 3D structure of the AtPAF2 protein with RMSD values in the range of 0.04 to 2.66 Å.

3.4 Phylogenetic analysis

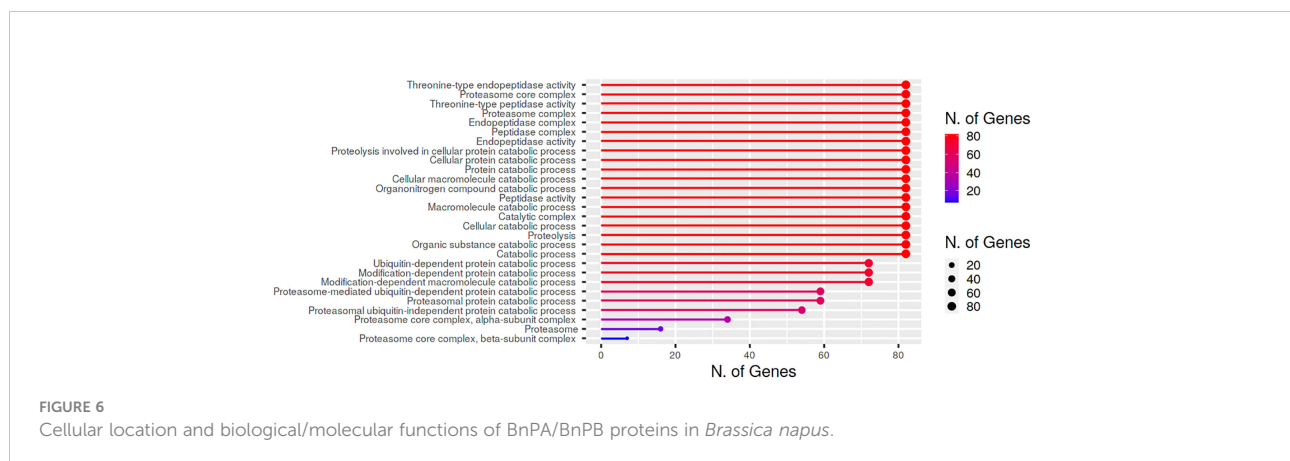
The amino acid sequences of *B. napus*, *A. thaliana* and *O. sativa* for α - (PA) and β -subunits (PB) were utilized separately to construct a phylogenetic tree. Seven clades that belong to 1 to 7 subunits were observed in each phylogenetic tree. These clades showed similarities with the conserved motifs (Figure 1). It was interesting to note that the orthologs of all three taxa belong to each of the α and β -subunits and formed seven clades in both phylogenetic trees. In the α -subunit tree, 7-10 orthologs formed clades (Figure 8A) whereas, in the β -subunit tree, 7-13 orthologs formed clades (Figure 8B).

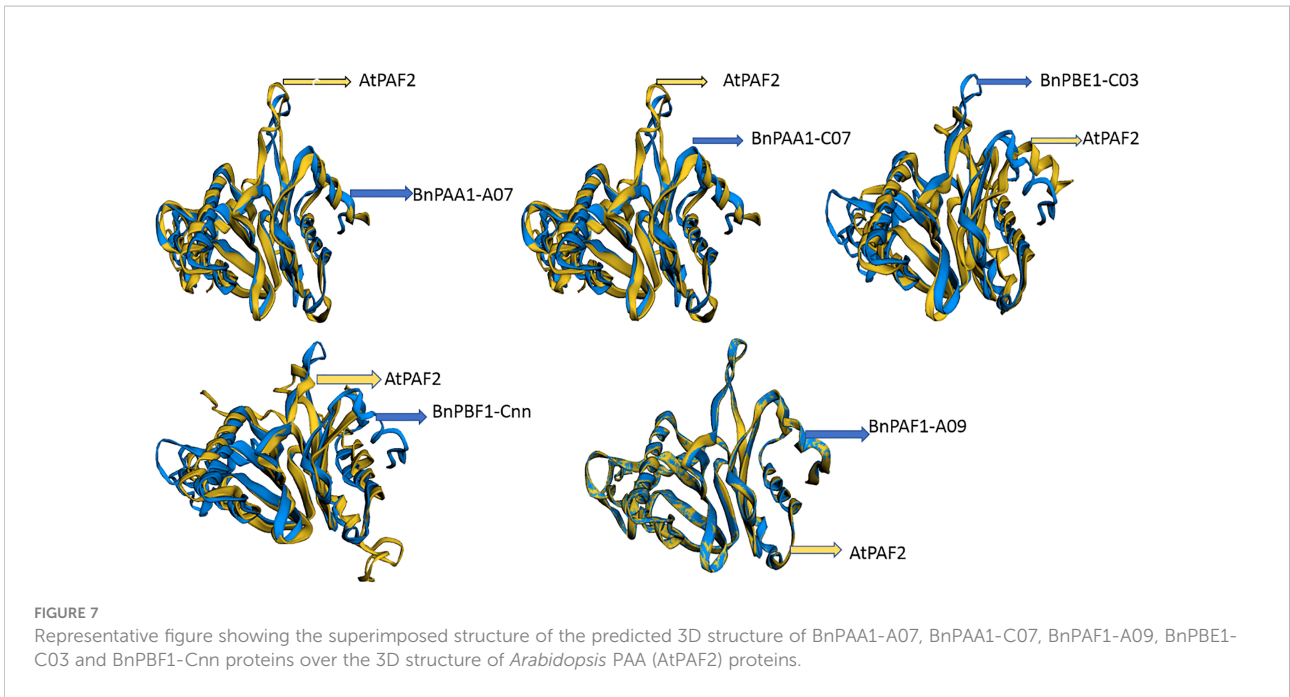
4 Discussion

Plant genome sequencing has been utilized to study genes related to various developmental stages and stress tolerance in many crops (Lu et al., 2019; Mehla et al., 2022). Crops whose genomes have not been sequenced are receiving benefits from those crops whose genomes have been sequenced. *A. thaliana* and *O. sativa* (Rensink and Buell, 2004) have been utilized for such studies.

4.1 Identification of *BnPA* and *BnPB* genes

The present study is the first report in *B. napus* on phylogenomic analysis for identifying and describing genes that encode for various subunits (α 1-7 and β 1-7) of the 20S proteasome. In this study, 82 *BnPA* and *BnPB* genes of *B. napus* were arranged into seven different α and β types of 20S proteasome. The sub-genomes A and C contain 42 and 40 genes, respectively. When compared to previously identified *A. thaliana* (*AtPAA-AtPAG* and *AtPBA-AtPBG*) and rice genes (*OsPAA-OsPAG* and *OsPBA-OsPBG*) (Fu et al., 1998; Sharma et al., 2022), a fairly large number of these genes were detected in *B. napus*. These findings may justify the fact that *B. napus* has a large genome and evolved with a higher ploidy level. The presence of two more genes in genome A of *B. napus* may be due to duplication, and a lower number of genes in genome C than genome A may be due to the loss of the gene. In the majority of the *BnPA* and *BnPB* genes, the structural pattern of exons and introns was found to be similar; however, in some cases, this similar pattern deviates. This deviation in gene structure may be due to the loss/gain of introns during the course of evolution (Rogozin et al., 2003; Yu et al., 2019). In cDNA sequences, this may be due to differences in intron number and size located in *BnPA* and *BnPB* genes. In addition, the variation in the length of UTRs present on the borders of cDNA may be due to differences in the length of cDNA sequences. The phylogenomic analysis of 82 genes revealed seven clades of each α (BnPAA-BnPAG) and β (BnPBA-BnPBG) subunit. The different α and β subunits of *B. napus* varied from each other but they showed similarities with Arabidopsis, a close relative of *B. napus*. This means that the 20S proteasome might have not changed after the divergence of these two taxa due to common ancestry. The distribution pattern of *BnPA* and *BnPB* genes on A-09, C-09, Ann and Cnn differed. In our findings, we found an uneven distribution of genes in genomes A and C. The general pattern is four genes per chromosome but in some cases, there is a deviation from this pattern. This deviation may be due to gene duplication or



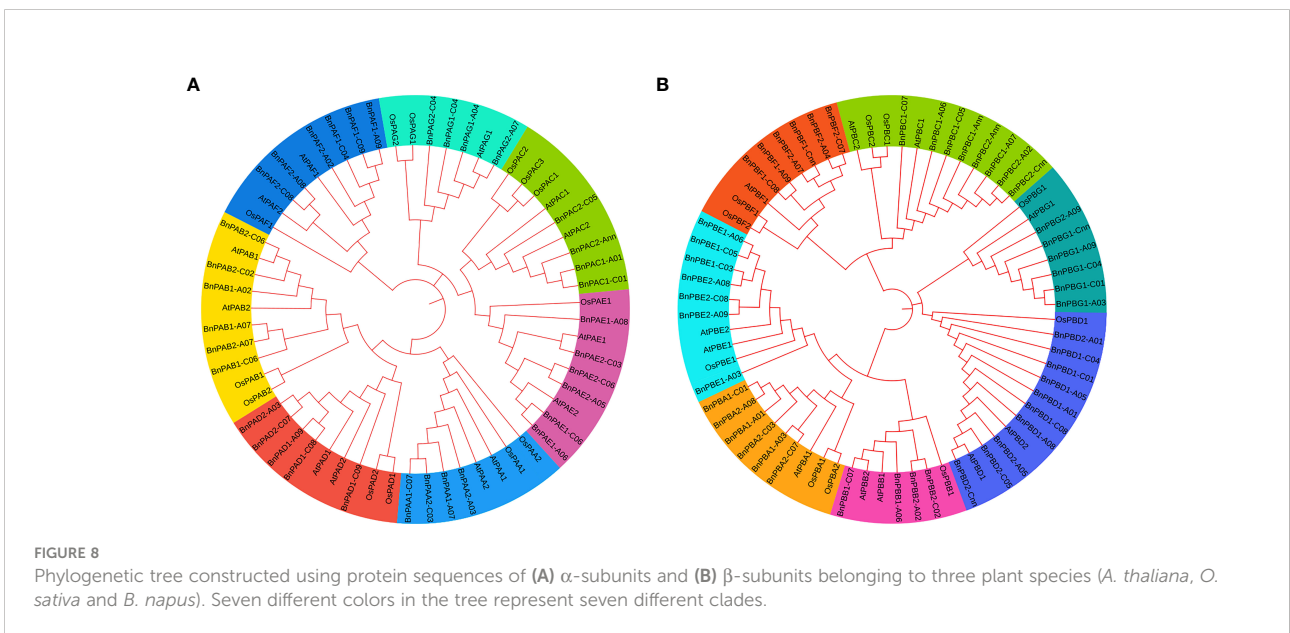


gene loss (Clavijo et al., 2017; Yu et al., 2019). The difference in the distribution of orthologous genes in the A and C genomes may be due to inversion and translocation.

4.2 Number of genes in Arabidopsis, rice and *B. napus*

In *B. napus*, 82 *BnPA* and *BnPB* genes of 20S PA and PB have been reported so far which represent the highest numbers in the

plant kingdom. This number is thrice the number reported in *Arabidopsis* and rice (Figure 9). The number of genes in A and C suggests that different genes have undergone duplication (paralogy) and speciation events (orthology). In *Arabidopsis*, out of 24 genes (Supplementary Table S2), instead of 14 genes as published in earlier studies (Fu et al., 1998), 13 genes with duplications for all α -subunits except *AtPAG* (α -7), whereas 11 genes with duplications in only 4 β -subunits, *AtPBB* (β -2), *AtPBC* (β -3), *AtPBD* (β -4), and *AtPBE* (β -5), have been reported (Parmentier et al., 1997; Fu et al., 1998).



Similarly, a phylogenomic survey of rice determined 23 genes instead of 14, as published in previous studies (Sassa et al., 2000). The duplications and heterogeneity have also been resolved in other organisms by phylogenomic surveys. A phylogenomic survey in *B. napus* revealed that the α -subunit included 35 genes with duplications except for *BnPAC1*, *BnPAA2*, *BnPAD2*, *BnPAG1*, *BnPAG2*, *BnPAA1*, and *BnPAC2*, and the β -subunit include 47 genes with duplication except for *BnPBB2*, *BnPBB1*, and *BnPBG2* (Supplementary Table S2). The α -genes found with maximum duplication are *BnPAB1*, *BnPAP2*, *BnPAP1*, *BnPAB2*, *BnPAP1*, and *BnPAD1*. Similarly, the β -genes found with maximum duplication are *BnPBD1*, *BnPBG1*, and *BnPBC1*.

4.3 Duplication and analysis of synteny in *BnPA* and *BnPB* genes

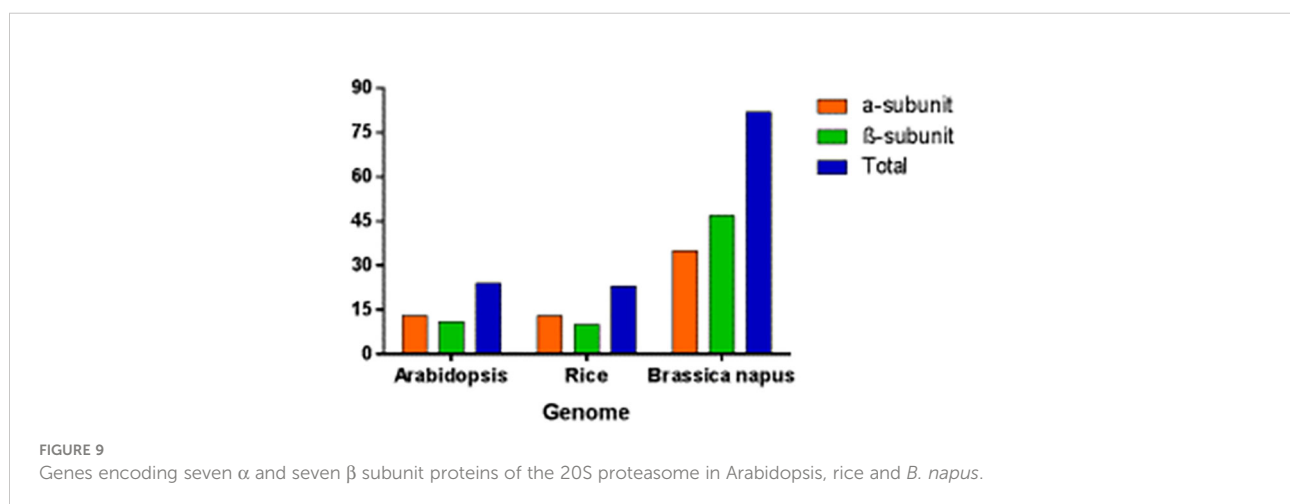
The present results on the 20S proteasome gene family of *B. napus* showed that in most cases duplication events have occurred in both the A and C genomes. The duplication events are from 1 (*BnPBG2-A09*) to 6 (*BnPBD1*). *BnPBD1* was reported on A01, A05, A08, C01, C04, and C08 suggesting maximum duplication (paralogy). These duplications may be due to translocation during the course of evolution. The absence of any duplication in *BnPBG2-A09* may be due to deletion. We have checked the synteny of all five chromosomes of Arabidopsis with all proteasome genes of *B. napus*. It was found that 81 genes of *B. napus* were syntenic with the genes on the chromosomes of *A. thaliana*. One gene of *B. napus* (*BnPAD1-C09*) was found to be nonsyntenic with *A. thaliana*. This could be due to deletion. Chromosome 1 and chromosome 3 of *A. thaliana* showed synteny with 29 genes of *B. napus*, whereas the least synteny was observed with chromosome 5, i.e., only with 4 genes of *B. napus*. All the genes on the chromosome of *O. sativa* were nonsyntenic.

4.4 Analysis of promoter sequences in *BnPA* and *BnPB* genes

The cis regulatory elements, present mostly near the start codon, are noncoding in nature and under different environmental conditions, they modulate gene expression in response to different transcription factors. Gene regulation is determined by cis regulatory elements at the basic level (Mondal and Das, 2020). The identified cis regulatory elements that spanned the promoter of 82 proteasomal genes were found to be associated with development and stress responses (light responsive, endosperm, and MeJA signaling) in Brassica (Supplementary Figure S2). Thus, it is evident that MeJA regulates plant growth and environmental stress (Wang et al., 2020) and defense against pathogens and herbivores (Fürstenberg-Hägg et al., 2013). There is a large number of evidence suggesting that hormonal signals affect the expression of the 20S proteasome gene. This expression leads to biotic and abiotic stresses (Kurepa et al., 2009). The presence of light-responsive elements indicates that they are required for light-dependent transcriptional regulation (Hiratsuka and Chua, 1997; Mondal et al., 2022). Similarly, the presence of endosperm elements is expressed during developmental stages. A large number of GBREs along with other cis-elements respond to phytohormones such as MeJA, and ABA available in *BnPA* and *BnPB* genes may control the expression of genes (*BnPA* and *BnPB*) under heat stress (Cao et al., 2015).

4.5 SSR with SSRs and miRNA with miRNAs

In this study, out of 82 genes, we found 66 SSRs in only 45 genes. This suggests that a fraction of genes in a gene family contain SSRs. The structural and functional aspects of SSR have been reported in a large number of genes (Li et al., 2004; Gupta



and Rustgi, 2004; Varshney et al., 2005). The frequency of dinucleotide was highest (20). It was followed by tetranucleotide and hexanucleotide. Such occurrence is unusual because trinucleotide repeats are generally found most frequently in comparison to other SSRs (Gupta and Varshney, 2000). The SSR reported in the gene that encodes the α and β -subunits of proteasome core particles, shows polymorphism which can be helpful to develop functional molecular markers. Such a marker can be used to improve tolerance against different physiological stresses in the plant. The absence of transposable and retroelements in the genes studied indicates the nonexpression of the 20S proteasome family in *B. napus*.

The mi-RNAs are small noncoding RNAs that have a regulatory function inside the cells at the post-transcriptional and translocation levels. These cause the degradation of targets of the gene (Budak & Zhang, 2017; Liu et al., 2017). In this study, 44 miRNAs involving sequences of *BnPA* and *BnPB* genes were identified. The target sites for these miRNAs were found only in 21 genes (8 *BnPA* and 13 *BnPB*). Among the predicted miRNAs, candidates of bna-miR166 (miR166 a-f; each 21 nucleotides long) were detected in *BnPAC1-A01* and those of miR160 (miR160 a-d) were detected in *BnPBD-C08*, *BnPBD1-A08*, *BnPBD2-A01*, and *BnPBD2-Cnn* (Supplementary Table S4). Micro-RNA166 (miR166), a highly conserved family of miRNAs is involved in several cellular and physiological processes in plants (Li et al., 2017), such as drought stress (Kantar et al., 2011), cold stress (Liu et al., 2008) and heat stress (Khraiwesh et al., 2012). F-box/SCF is encoded by miR2111, which is involved in ubiquitin-mediated proteolysis and then in the adaptive response of Pi-deficiency (Kumar et al., 2017). miR160 has a role in both abiotic and biotic stress responses in Arabidopsis and maize (Zhang et al., 2009; Moldovan et al., 2010). miR160a is involved in regulating auxin response genes, club root development, and disease modulation (Verma et al., 2014). In this way, the function of miR166 and miR160 is understood in plants. The presence of miR166 and miR160 in *B. napus* may be significant for future studies on the biological role of these miRNAs.

4.6 Structural and functional features of *BnPA* and *BnPB* proteins

The predicted proteins of *BnPA* and *BnPB* genes were found to be variable and, differed in the protein length of similar genes from *A. thaliana* and *O. sativa*. The number of amino acids in *BnPA* and *BnPB* varies from 82 aa (*BnPBC2-Cnn*) to 732 aa (*BnPAE1-A06*). This length is more variable than those of rice proteins (237 to 256 aa). Similarly, the length of proteins (199 to 298 aa) of *A. thaliana* is the least variable (Town et al., 2006). The GRAVY ranged from -0.669 to 0.093. This suggests the hydrophobic nature of these proteins. Due to this property, there

will be proper folding of proteins to keep them stable and biologically active.

The *BnPA* and *BnPB* genes encode proteins that contain 10 different motifs. Similarly, 10 motifs are present in the database for Arabidopsis and rice (Parmentier et al., 1997; Fu et al., 1998; Sassa et al., 2000). The length of the individual motif varied from 15 aa (motif 3) to 41 aa (motif 10). Motifs 1, 4, 5, and 6 are concerned with catabolic processes at the cellular level (<https://www.ebi.ac.uk/interpro/result/InterProScan/iprscan5-R20211209-091508-0331-1308065-p2m/>). The remaining 6 motifs were found to be novel and need to be characterized at the molecular level. Earlier studies reported that in eukaryotes, proteins of the 20S proteasome could be in the nucleus and cytoplasm, and the same case may be true for *BnPA* and *BnPB* proteins (Fu et al., 1998). A single specific domain of α -type (α 1-7) or β -type (β 1-7) is present in all 82 *BnPA* and *BnPB* proteins. Of these, 35 *BnPA* proteins contain a single α -type domain and 47 *BnPB* proteins contain a single β -type domain (Supplementary Table 2). Similar findings have been observed in yeast, Arabidopsis and rice (Parmentier et al., 1997; Groll et al., 1997; Fu et al., 1998; Arias and Pires, 2012). A majority of aa residues (>90%) that fall in the most favored region were shown by modeled 3D structures of 20 *BnPA/BnPB* proteins. This suggested that the predicted models are reliable (Laskowski et al., 1993; Kumar et al., 2018; Batra et al., 2019). The predicted 3D structures of *BnPA* and *BnPB* proteins may be helpful for the initial understanding of molecular functions.

4.7 Phylogenetic analysis of *BnPA* and *BnPB* protein sequences

The different orthologs of the α and β -subunits of 20S proteasome were examined, and phylogenetic trees were constructed. The orthologs of Brassica, Rice and Arabidopsis that belong to different α and β -subunits were depicted into seven clades for the α and β subfamilies. Similar findings have been reported for yeast, Arabidopsis and rice proteins (Fu et al., 1998; Sassa et al., 2000). These findings suggest that the α and β -subunits of *B. napus* have a higher level of similarity with those of Arabidopsis and rice.

4.8 Digital expression analysis of *BnPA* and *BnPB* genes

All 20S proteasome genes and core particles are involved in plant responses to different biotic and abiotic stresses (Xu and Xue, 2019). This is due to different signaling molecules that control the development of plants under different stress conditions (Livneh et al., 2016). Earlier a few studies (Fu et al., 1998; Li et al., 2015) examined digital expression analysis of the

20S proteasome gene in a plant system. The present study is the first report in *B. napus* on phylogenomic analysis and digital expression of the 20S proteasome gene under normal tissue-specific and stress conditions in the germination and seedling stages. This was done using a transcriptome database in which approximately 18/82 genes showed higher expression at different germination and seedling stages of *B. napus* in root, seed, leaf, cotyledon, hypocotyls, and radicle (Figure 4). A greater mRNA level expression for six 20S proteasome subunit genes compared to β -tubulin in flower and fruit tissues of Arabidopsis has been observed via RNA blot analysis (Fu et al., 1998). Under 24 h of abiotic stress (cold, drought, and heat), two genes for cold and one gene for drought exhibited high expression at the seedling and flowering stage. No gene was reported for heat. However, in wild rice, the *OgTT1* gene was reported under heat stress conditions (Li et al., 2015). Under 264 h and 72 h of biotic stress (*L. maculans*) as many as one gene exhibited higher expression at the seedling stage. Under 96 h, 48 h and 24 h of biotic stress (*S. sclerotiorum*) as many as 1, 1 and 0 genes respectively, showed high expression in the flowering stage. In this way, the 20S proteasome gene plays a significant role in the development of different plant organs such as roots, seeds, leaves, cotyledons, hypocotyls, and radicles, as well as in abiotic stresses (Xu and Xue, 2019). The functions of the aforementioned genes that showed differential expression under biotic and abiotic stresses may be investigated in future studies.

5 Conclusion

The present study is the first report in *B. napus* on phylogenomic analysis for identifying and describing genes that encode 35 *BnPA* and 47 *BnPB* genes of the 20S proteasome. The orthology and paralogy of *BnPA* and *BnPB* genes were inferred and identified on the basis of the pattern of speciation and duplication of genes in Arabidopsis and rice. A full-length 3D model was predicted for the proteins encoded by *BnPA* and *BnPB* genes. A number of *BnPA* and *BnPB* genes were found to be expressed in many organs of *B. napus* under normal as well as abiotic and biotic stress conditions. In this way, the present study provides much information that can be utilized for the development of climate resilient cultivars of *B. napus*.

References

- Arias, T., and Pires, J. C. (2012). A fully resolved chloroplast phylogeny of the brassica crops and wild relatives (Brassicaceae: Brassicaceae): Novel clades and potential taxonomic implications. *Taxon* 61 (5), 980–988. doi: 10.1002/tax.615005
- Batra, R., Agarwal, P., Tyagi, S., Saini, D. K., Kumar, V., Kumar, A., et al. (2019). A study of CCD8 genes/proteins in seven monocots and eight dicots. *PLoS One* 14 (3), e0213531. doi: 10.1371/journal.pone.0213531

Author contributions

VM, SK and UK planned and guided the design of this study. VK, HS, LS, AT, PJ, PB, SJ and ID retrieved sequence data, conducted computational analysis. VK, HS, LS, AT, PJ and PB performed data visualization, and wrote the first draft of the manuscript. VM, RM, KS, SK, YS, SJ, ID and UK critically revised and edited the manuscript. All authors have read and agree to the final version of the manuscript.

Acknowledgments

The authors thank the Head, Department of Botany, Chaudhary Charan Singh University, Meerut, India for providing the necessary facilities to carry out this study and Head of Department of Molecular Biology & Biotechnology, College of Biotechnology, Chaudhary Charan Singh Haryana Agricultural University, Hisar, India for providing the necessary facilities to carry out this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1037206/full#supplementary-material>

- Bolser, D. M., Kerhornou, A., Walts, B., and Kersey, P. (2015). Triticeae resources in ensembl plants. *Plant Cell Physiol.* 56 (1), e3. doi: 10.1093/pcp/pcu183
- Budak, H., and Zhang, B. (2017). MicroRNAs in model and complex organisms. *Funct. Integr. Genomics* 17 (2), 121–124. doi: 10.1007/s10142-017-0544-1
- Cao, Y., Jiang, Y., Ding, M., He, S., Zhang, H., Lin, L., et al. (2015). Molecular characterization of a transcriptionally active Ty1/copia-like retrotransposon in gossypium. *Plant Cell Rep.* 34 (6), 1037–1047. doi: 10.1007/s00299-015-1763-3

- Cartea, M. E., Soengas, P., Picoaga, A., and Ordas, A. (2005). Relationships among brassica napus (L.) germplasm from Spain and great Britain as determined by RAPD markers. *Genet. Resour. Crop Evol.* 52 (6), 655–662. doi: 10.1007/s10722-003-6014-8
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic brassica napus oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435
- Chen, P., and Hochstrasser, M. (1996). Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* 86 (6), 961–972. doi: 10.1016/S0092-8674(00)80171-3
- Clavijo, B. J., Venturini, L., Schudoma, C., Accinelli, G. G., Kaithakottil, G., Wright, J., et al. (2017). An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res.* 27 (5), 885–896. doi: 10.1101/gr.217117.116
- Colovos, C., and Yeates, T. (1993). ERRAT: An empirical atom-based method for validating protein structures. *Protein Sci.* 2 (9), 1511–1519. doi: 10.1002/pro.5560020916
- Dai, X., and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39, W155–W159. doi: 10.1093/nar/gkr319
- Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: a plant small RNA target analysis server, (2017 Release). *Nucleic Acids Res.* 46, W49–W54. doi: 10.1093/nar/gky316
- Dantuma, N. P., and Bott, L. C. (2014). The ubiquitin-proteasome system in neurodegenerative diseases: precipitating factor, yet part of the solution. *Front. Mol. Neurosci.* 7. doi: 10.3389/fnmol.2014.00070
- DeSalle, R., Tessler, M., and Rosenfeld, J. (2020). *Phylogenomics: a primer* (Boca Raton, London, New York: CRC Press) Boca Raton, London, New York.
- Dhaliwal, A. K., Mohan, A., and Gill, K. S. (2014). Comparative analysis of ABCB1 reveals novel structural and functional conservation between monocots and dicots. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00657
- Duke, J. A. (1983). *Handbook of energy crops* (Sydney, Australia: Center for New Crops & Plants Products).
- Eisen, J. A. (1998). Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* 8 (3), 163–167. doi: 10.1101/gr.8.3.163
- Eisenberg, D., Lüthy, R., and Bowie, J. U. (1997). [20] VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 277, 396–404. doi: 10.1016/S0076-6879(97)77022-8
- Fu, H., Doelling, J. H., Arendt, C. S., Hochstrasser, M., and Vierstra, R. D. (1998). Molecular organization of the 20S proteasome gene family from *Arabidopsis thaliana*. *Genetics* 149 (2), 677–692. doi: 10.1093/genetics/149.2.677
- Fürstenberg-Hägg, J., Zagrobelny, M., and Bak, S. (2013). Plant defense against insect herbivores. *Int. J. Mol. Sci.* 14 (5), 10242–10297. doi: 10.3390/ijms140510242
- Gilbert, A., Cosgrove, L., and Wilkinson, J. (2012). “The royal horticultural society *VegeTable and fruit gardening in australia*,” (Melbourne, Australia: Dorling Kindersly Australia Pty Ltd.).
- Grieve, M. A. (1984). “*Modern herbal*,” (Melbourne, Australia: Penguin Books Australia).
- Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D., et al. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386 (6624), 463–471. doi: 10.1038/386463a0
- Gupta, P. K., and Rustgi, S. (2004). Molecular markers from the transcribed/expressed region of the genome in higher plants. *Funct. Integr. Genomics* 4 (3), 139–162. doi: 10.1007/s10142-004-0107-0
- Gupta, P. K., and Varshney, R. K. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113 (3), 163–185. doi: 10.1023/A:1003910819967
- Hiratsuka, K., and Chua, N. H. (1997). Light-regulated transcription in higher plants. *J. Plant Res.* 110 (1), 131–139. doi: 10.1007/BF02506852
- Kantar, M., Lucas, S. J., and Budak, H. (2011). miRNA expression patterns of *Triticum dicoccoides* in response to shock drought stress. *Planta* 233, 471–484. doi: 10.1007/s00425-010-1309-4
- Khravish, B., Zhu, J. K., and Zhu, J. (2012). Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim. Biophys. Acta (BBA) Gene Regul. Mech.* 1819 (2), 137–148. doi: 10.1016/j.bbagr.2011.05.001
- Kleiger, G., and Mayor, T. (2014). Perilous journey: a tour of the ubiquitin-proteasome system. *Trends Cell Biol.* 24 (6), 352–359. doi: 10.1016/j.tcb.2013.12.003
- Kumar, A., Sharma, P., Gomar-Alba, M., Shcheprova, Z., Daulny, A., Sanmartín, T., et al. (2018). Daughter-cell-specific modulation of nuclear pore complexes controls cell cycle entry during asymmetric division. *Nat. Cell Biol.* 20 (4), 432–442. doi: 10.1038/s41556-018-0056-9
- Kumar, S., Verma, S., and Trivedi, P. K. (2017). Involvement of small RNAs in phosphorus and sulfur sensing, signaling and stress: Current update. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00285
- Kurepa, J., Wang, S., Li, Y., and Smalle, J. (2009). Proteasome regulation, plant growth and stress tolerance. *Plant Signaling Behav.* 4 (10), 924–927. doi: 10.4161/psb.4.10.9469
- Kutcher, H. R., Warland, J. S., and Brandt, S. A. (2010). Temperature and precipitation effects on canola yields in Saskatchewan, Canada. *Agric. For. Meteorol.* 150 (2), 161–165. doi: 10.1016/j.agrformet.2009.09.011
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26 (2), 283–291. doi: 10.1107/S0021889892009944
- Li, X. M., Chao, D. Y., Wu, Y., Huang, X., Chen, K., Cui, L. G., et al. (2015). Natural alleles of a proteasome $\alpha 2$ subunit gene contribute to thermotolerance and adaptation of African rice. *Nat. Genet.* 47 (7), 827–833. doi: 10.1038/ng.3305
- Li, Y. C., Korol, A. B., Fahima, T., and Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21 (6), 991–1007. doi: 10.1093/molbev/msh073
- Liu, W., Ma, R., and Yuan, Y. (2017). Post-transcriptional regulation of genes related to biological behaviors of gastric cancer by long noncoding RNAs and microRNAs. *J. Cancer* 8 (19), 4141–4154. doi: 10.7150/jca.22076
- Liu, H.-H., Tian, X., Li, Y.-J., Wu, C.-A., and Zheng, C.-C. (2008). Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. *RNA* 14, 836–843. doi: 10.1261/rna.895308
- Livneh, I., Cohen-Kaplan, V., Cohen-Rosenzweig, C., Avni, N., and Ciechanover, A. (2016). The life cycle of the 26S proteasome: from birth, through regulation and function, and onto its death. *Cell Res.* 26 (8), 869–885. doi: 10.1038/cr.2016.86
- Li, X., Xie, X., Li, J., Cui, Y., Hou, Y., Zhai, L., et al. (2017). “Conservation and diversification of the miR166 family in soybean and potential roles of newly identified miR166s.” *BMC Plant Biol.* 17 (1), 1–18. doi: 10.1186/s12870-017-0983-9
- Lozano-Baena, M. D., Tasset, I., Obregón-Cano, S., de Haro-Bailon, A., Muñoz-Serrano, A., and Alonso-Moraga, Á. (2015). Antigenotoxicity and tumor growing inhibition by leafy *Brassica carinata* and sinigrin. *Molecules* 20 (9), 15748–15765. doi: 10.3390/molecules200915748
- Lust, J. (1983). *The herb book* (Melbourne, Australia: Penguin Books Australia).
- Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., et al. (2019). Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. *Nat. Commun.* 10 (1), 1–12. doi: 10.1038/s41467-019-09134-9
- Maršalkienė, N., Slesaravičius, A., Karpavičienė, B., and Dastkaitė, A. (2009). Oil content and fatty acid composition of seeds of some Lithuanian wild crucifer species. *Agron. Res.* 7, 654–661.
- Mehla, S., Kumar, U., Kapoor, P., Singh, Y., Sihag, P., Sagwal, V., et al. (2022). Structural and functional insights into the candidate genes associated with different developmental stages of flag leaf in bread wheat (*Triticum aestivum* L.). *Front. Genet.* 13. doi: 10.3389/fgene.2022.933560
- Moldovan, D., Spriggs, A., Yang, J., Pogson, B. J., Dennis, E. S., and Wilson, I. W. (2010). Hypoxia-responsive microRNAs and trans-acting small interfering RNAs in *Arabidopsis*. *J. Exp. Bot.* 61 (1), 165–177. doi: 10.1093/jxb/erp296
- Mondal, R., and Das, P. (2020). Data-mining bioinformatics: suggesting *Arabidopsis thaliana* l-type lectin receptor kinase IX. 2 (LecRK-IX. 2) modulate metabolites and abiotic stress responses. *Plant Signaling Behav.* 15, 1818031. doi: 10.1080/15592324.2020.1818031
- Mondal, R., Madhurya, K., Saha, P., Chattopadhyay, S. K., Antony, S., Kumar, A., et al. (2022). Expression profile, transcriptional and post-transcriptional regulation of genes involved in hydrogen sulphide metabolism connecting the balance between development and stress adaptation in plants: a data-mining bioinformatics approach. *Plant Biol.* 24, 602–617. doi: 10.1111/plb.13378
- Parmentier, Y., Bouchez, D., Fleck, J., and Genschik, P. (1997). The 20S proteasome gene family in *Arabidopsis thaliana*. *FEBS Lett.* 416 (3), 281–285. doi: 10.1016/S0014-5793(97)01228-3
- Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., et al. (1999). ‘Green revolution’ genes encode mutant gibberellin response modulators. *Nature* 400, 256–261. doi: 10.1038/22307
- Rensink, W. A., and Buell, C. R. (2004). *Arabidopsis* to rice: applying knowledge from a weed to enhance our understanding of a crop species. *Plant Physiol.* 135 (2), 622–629. doi: 10.1104/pp.104.040170
- Rogozin, I. B., Wolf, Y. I., Sorokin, A. V., Mirkin, B. G., and Koonin, E. V. (2003). Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* 13 (17), 1512–1517. doi: 10.1016/S0960-9822(03)00558-X
- Sassa, H., Oguchi, S., Inoue, T., and Hirano, H. (2000). Primary structural features of the 20S proteasome subunits of rice (*Oryza sativa*). *Gene* 250 (1–2), 61–66. doi: 10.1016/S0378-1119(00)00190-6

- Sharma, H., Batra, R., Kumar, S., Kumar, M., Kumar, S., Balyan, H. S., et al. (2022). Identification and characterization of 20S proteasome genes and their relevance to heat/drought tolerance in bread wheat. *Gene Rep.* 27, 101552. doi: 10.1016/j.genrep.2022.101552
- Sharma, B., Joshi, D., Yadav, P. K., Gupta, A. K., and Bhatt, T. K. (2016). Role of ubiquitin-mediated degradation system in plant biology. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00806
- Stone, S. L. (2019). Role of the ubiquitin-proteasome system in plant response to abiotic stress. *Int. Rev. Cell Mol. Biol.* 343, 65–110. doi: 10.1016/bs.ircmb.2018.05.012
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28 (10), 2731–2739. doi: 10.1093/molbev/msr121
- Town, C. D., Cheung, F., Maiti, R., Crabtree, J., Haas, B. J., Wortman, J. R., et al. (2006). Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *Plant Cell* 18 (6), 1348–1359. doi: 10.1105/tpc.106.041665
- Varshney, R. K., Graner, A., and Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23 (1), 48–55. doi: 10.1016/j.tibtech.2004.11.005
- Verma, S. S., Rahman, M. H., Deyholos, M. K., Basu, U., and Kav, N. N. (2014). Differential expression of miRNAs in *Brassica napus* root the following infection with *Plasmodiophora brassicae*. *PLoS One* 9 (1), e86648. doi: 10.1371/journal.pone.0086648
- Vierstra, R. D. (2009). The ubiquitin–26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* 10 (6), 385–397. doi: 10.1038/nrm2688
- Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara gene trees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19 (2), 327–335. doi: 10.1101/gr.073585.107
- Wang, J., Song, L., Gong, X., Xu, J., and Li, M. (2020). Functions of jasmonic acid in plant regulation and response to abiotic stress. *Int. J. Mol. Sci.* 21 (4), 1446. doi: 10.3390/ijms21041446
- Wiersema, J. H., and Leon, B. (2016). “World economic plants: A standard reference,” (Melbourne, Australia: CRC press).
- Wolf, D. H., and Hilt, W. (2004). The proteasome: a proteolytic nanomachine of cell regulation and waste disposal. *Biochim. Biophys. Acta (BBA) Mol. Cell Res.* 1695 (1–3), 19–31. doi: 10.1016/j.bbamcr.2004.10.007
- Xu, F. Q., and Xue, H. W. (2019). The ubiquitin-proteasome system in plant responses to environments. *Plant Cell Environ.* 42 (10), 2931–2944. doi: 10.1111/pce.13633
- Yang, P., Fu, H., Walker, J., Papa, C. M., Smalle, J., Ju, Y. M., et al. (2004). Purification of the arabidopsis 26 s proteasome: biochemical and molecular analyses revealed the presence of multiple isoforms. *J. Biol. Chem.* 279 (8), 6401–6413. doi: 10.1074/jbc.M311977200
- Yu, X., Han, J., Wang, E., Xiao, J., Hu, R., Yang, G., et al. (2019). Genome-wide identification and homoeologous expression analysis of PP2C genes in wheat (*Triticum aestivum* L.). *Front. Genet.* 10. doi: 10.3389/fgene.2019.00561
- Zhang, L., Chia, J. M., Kumari, S., Stein, J. C., Liu, Z., Narechania, A., et al. (2009). A genome-wide characterization of microRNA genes in maize. *PLoS Genet.* 5 (11), e1000716. doi: 10.1371/journal.pgen.1000716