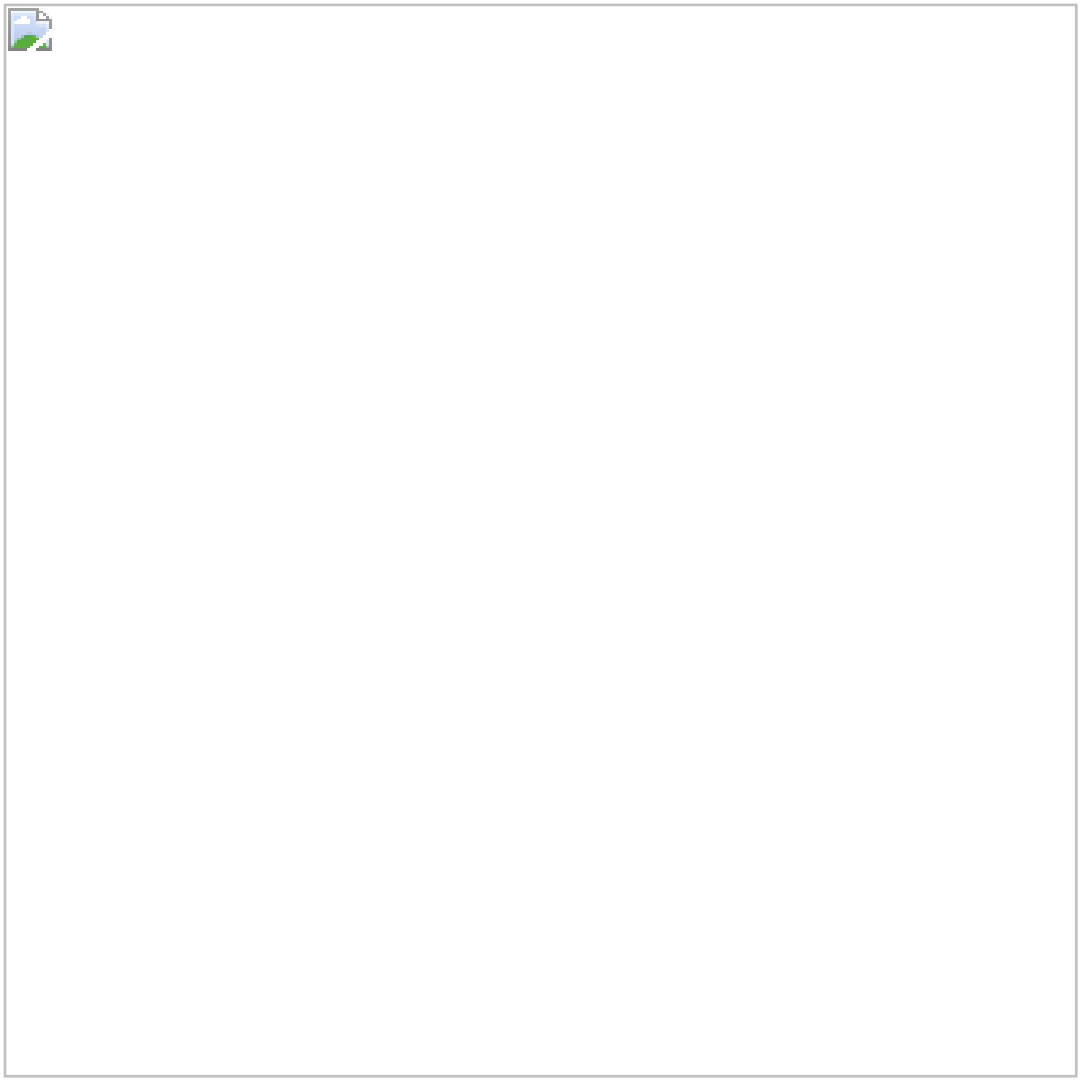

De Novo Transcriptome Reconstruction of a Thermo-Sensitive Male Sterility Mutant in Rapeseed (Brassica napus; Brassicaceae)

Authors: Liu, Xi-Qiong, Yu, Cheng-Yu, Dong, Jun-Gang, Xu, Ai-Xia, and Hu, Sheng-Wu

Source: Applications in Plant Sciences, 5(12)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1700077

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

Applications in Plant Sciences 2017 5(12): 1700077

Genomic Resources Article

De novo transcriptome reconstruction of a thermo-sensitive male sterility mutant in rapeseed (*Brassica napus*; Brassicaceae)[1]

Xi-Qiong Liu[2], Cheng-Yu Yu[2,3], Jun-Gang Dong[2], Ai-Xia Xu[2], and Sheng-Wu Hu[2]

[2]College of Agronomy, Northwest A&F University, 3 Taicheng Road, Yangling 712100, People's Republic of China

- *Premise of the study:* SP2S is a spontaneous thermo-sensitive genic male sterility (TGMS) mutation that facilitates two-line hybrid breeding in *Brassica napus* (Brassicaceae). De novo assembly of the floral bud transcriptome of SP2S can provide a foundation for deciphering the transcriptional regulation of SP2S in response to temperature change.
- *Methods:* mRNAs of the young floral buds of SP2S and its near-isogenic line SP2F grown under cool (16°C)/warm (22°C) conditions were sequenced on an Illumina Solexa platform, producing 239.7 million short reads with a total length of 19.95 Gbp.
- *Results:* The reads were assembled de novo using the Trinity program, resulting in 135,702 transcripts with an average length of 784 bp, an N50 value of 1221 bp, and a total length of 107 Mbp. We identified 24,157 cDNA-derived simple sequence repeats in the assembly. We found 137 and 195 single-nucleotide polymorphisms and 49 and 51 differentially regulated KEGG orthology groups when comparing sample SP2S at 22°C vs. SP2S at 16°C and sample SP2S at 22°C vs. SP2F at 22°C, respectively.
- *Discussion:* The numerous differentially expressed genes and the derived single-nucleotide polymorphisms show abnormal transcriptional regulation in the TGMS system. These results outline an intricate transcriptional regulation that occurred in the rapeseed TGMS SP2S when the temperature changed.

Key words: *Brassica napus*; Brassicaceae; RNA-seq; thermo-sensitive genic male sterility; transcriptome; Trinity.

Pollen development is a biological process involving many events, including anther cell division and differentiation, microsporocyte meiosis, tetrad microspore release, microspore mitosis, and pollen coat development. These events rely on the functions of numerous genes from the microspore, tapetum, and other sporophytic anther tissues (Ma, 2005; Zhu et al., 2011); therefore, dysfunction of these genes may lead to male sterility. Global transcriptome profiling can provide insights into the functional elements of the genome and reveal the molecular constituents of male sterility–related tissues. Thus, thousands of transcripts expressed in flowers and anthers have been identified in brassica species such as *Brassica oleracea* L. (Kim et al., 2014; Ma et al., 2015), *B. rapa* L. (Tong et al., 2013), *B. juncea* (L.) Czern. (Paritosh et al., 2014), and *B. napus* L. (An et al., 2014; Qu et al., 2015). Currently, high-throughput RNA sequencing (RNA-Seq) is a powerful and cost-effective tool for transcription profiling (Bancroft et al., 2011; Tong et al., 2013; An et al., 2014; Zhao et al., 2014; Ma et al., 2015; Qu et al., 2015). RNA-Seq is superior to other technologies in detecting low-abundance transcripts, differentiating biologically critical isoforms, and identifying genetic variants (Zhao et al., 2014), including alternative splicing (Marioni et al., 2008), gene fusion (Ozsolak and Milos, 2011), simple sequence repeats (SSRs), and single-nucleotide polymorphisms (SNPs).

Male sterility in plants is widely studied because it can facilitate the manipulation of hybrid vigor (heterosis). Among various male sterility types, environment-sensitive male sterility has been considered a special model for examining the interactions between genes and temperature or photoperiod (Ding et al., 2012; Pan et al., 2014; Zhou et al., 2014). Male sterility contributes greatly to sustainable increases of rapeseed production to meet the growing demands for both edible and biofuel oils. A new type of rapeseed thermo-sensitive genic male sterility (TGMS) mutation, SP2S, was discovered by us in 2007 (Yu et al., 2015). SP2S is male fertile under cool conditions (temperature <18°C) but sterile under warm conditions (>20°C). TGMS SP2S is controlled by at least two pairs of recessive genes, and all other cultivars can restore SP2S fertility. The combining ability of SP2S on seed yield and seed quality was satisfactory, and the male sterility genes in SP2S showed no adverse effect on the performance of F_1 hybrids (Yu et al., 2015). Thus, SP2S can be used as a promising pollination control system for hybrid production. Some structural abnormalities in SP2S were observed, including the premature breakdown of an extremely vacuolated tapetum and a delayed dissolution of the tetrad wall covering the microspore (Yu et al., 2015). We attempted to find some molecular clues to the TGMS system using a proteomic comparison of the floral buds of SP2S with its near-isogenic line (NIL) SP2F, but we identified only 28 differentially accumulated

[1]Manuscript received 29 July 2017; revision accepted 17 October 2017.

This work was financially supported by the National Transgenic Research Projects of China (2018ZX08020001) and Yangling Sci-Tech Plan Program (2016NY-04). The authors thank Mr. Zhiyuan Huang at LC-Bio, Hangzhou, China, for his assistance with data analysis.

[3]Author for correspondence: yu1009@nwafu.edu.cn

doi:10.3732/apps.1700077

Applications in Plant Sciences 2017 5(12): 1700077; http://www.bioone.org/loi/apps © 2017 Liu et al. Published by the Botanical Society of America. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY-NC-SA 4.0), which permits unrestricted noncommercial use and redistribution provided that the original author and source are credited and the new work is distributed under the same license as the original.

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

TABLE 1. Primers designed for the selected differentially expressed transcripts.

Accession	Gene name	Forward sequence	Reverse sequence	Length (bp)
comp52020_c1_seq1	glucan endo-1,3-beta-glucosidase A6	TAACCCTCTGCCTAAACC	GCTCGTAACATTCTCTGC	166
comp80937_c0_seq1	bHLH25-like	GAGGAGCAAGCCAGTCAGAG	TTCAGCCATAGTGATCTCAACAAC	189
comp17045_c0_seq1	Thioredoxin reductase 2-like	GGAAGCAATCAGAGCGGTTC	TCGGCGAGCACAGTTCTC	112
comp49044_c0_seq1	Small RNA degrading nuclease 2-like	GAAGTAAGGAACTGGATCAAGG	GGAGCACATAATAGCCACAAG	189
comp55352_c1_seq3	Phosphoinositide phospholipase C1	CTCTCCGTTGCCTCCTTCC	GGTGCTGCTTCTACTGTTGAG	124
comp63945_c1_seq7	Leucine-rich repeat receptor-like serine/threonine-protein kinase	GAGTTCCAGTAAGCAGAATCAAG	CGATGACAGAGCAGCCTATC	112
comp49237_c0_seq2	Serine/threonine-protein phosphatase 2A	GTACGCCTCAACATCATAAGC	CGATTATAGCAAGACGAACTCTC	87
comp52644_c0_seq1	Glucan endo-1,3-beta-glucosidase, acidic isoform-like	TTGGCTTGTAACTGCTCATCTG	CCTTCTTGTTCGTCCTATCTCTTG	109
comp52222_c0_seq7	Pollen-specific protein SF3	TGAATCGCCGTTCTCTTCC	GTCCTCCAATCCGCCAAC	105
comp58831_c1_seq3	Probable transcription factor WRKY 70	AGGCACCAACTCGTTCAAG	TGACAAGAGAGGAGGAGGAG	178
EV220887.1	BnActin7	CGCGCCTAGCAGCATGAA	GTTGGAAAGTGCTGAGAGATGCA	101

peptides (Zhang et al., 2016). Transcriptomic analysis of a physiological process may obtain additional information that differs from proteomic analysis due to the very high resolution of transcriptome profiling and a limited correspondence between the mRNA level and protein abundance (Kubala et al., 2015). Therefore, in this study, we performed RNA-Seq and de novo transcriptome assembly to generate a flower bud transcriptome of rapeseed TGMS line SP2S. Numerous SNPs, SSRs, and differentially expressed transcripts (DETs) were identified in the transcriptome. The transcriptome assembly provides a platform for our future study of reproductive responses of rapeseed TGMS to temperature changes.

MATERIALS AND METHODS

***Plant materials*—**The plants of the TGMS line SP2S and its NIL SP2F, originating from same genetic background (Yu et al., 2015), were grown in the experimental field of Northwest A&F University (Yangling, China) in September 2013 and transplanted into a greenhouse in December 2013. When the plants were bolting, both SP2S and SP2F plants were divided into two groups and assigned to cool (16°C) or warm (22°C) treatments in temperature-controlled rooms where the photoperiod was longer than 14 h. The four treatments were designated S16 (SP2S at 16°C), S22 (SP2S at 22°C), F16 (SP2F at 16°C), and F22 (SP2F at 22°C). The plants of S22 would be male sterile when blossoming, but the other three treatments would be fertile. Young floral buds 1–3 mm long (from microsporocyte to uninucleate microspore stage) were excised from a minimum of 10 plants from each treatment. The samples were preserved in RNA*later* (LC-Bio, Hangzhou, China) for later use.

***RNA extraction and RNA-Seq*—**Samples were crushed in liquid nitrogen, and the total RNA was extracted using TRIzol Reagent (Invitrogen, Carlsbad, California, USA). After purification using a TRK1001 Kit (LC Science, Houston, Texas, USA), the quantity and purity of the total RNA were analyzed using a Bioanalyzer 2100 (Agilent, Palo Alto, California, USA). Then, 10 μg RNA of each sample with an RNA integrity number (RIN) ≥7.0 was used to isolate mRNA using poly(T) oligo-attached magnetic beads. The mRNA was fragmented into small pieces using divalent cations, and the cleaved mRNA fragments were reverse-transcribed to produce a normalized cDNA library using the mRNA-Seq Sample Preparation Kit (Illumina, San Diego, California, USA). The average insert size for libraries was 300 ± 50 bp. The cDNAs were sequenced on an Illumina HiSeq 2500 platform as 150-bp paired-end reads, with at least five technical replications. The generated raw data were trimmed to remove nonsense sequences including adapters, low-quality reads, vectors, and very short sequences. The clean reads of four samples were pooled together to assemble a transcriptome in the Trinity software package (Haas et al., 2013) using the default parameters (http://trinityrnaseq.sourceforge.net/). Trinity is a well-known method for the efficient and robust de novo reconstruction of transcriptomes from RNA-Seq data. After assembly, the longest transcript at each locus (designed as comp*_c*_) was considered the unigene for subsequent annotation.

***Analysis of sequence variations*—**To obtain information about gene mutants and RNA editing sites, SNPs from sample S22, compared to S16 and F22, were screened using Isaac Variant Caller (Starling2) in the Bowtie package (Langmead et al., 2009), and highly repeated loci were further designated using SAMtools (Li et al., 2009). We also searched for cDNA-derived SSRs in the assembled transcripts using the MIcroSAtellite Identification Tool (MISA) (Thiel et al., 2003).

***Functional annotation and classification*—**We searched the assembled transcripts in the databases of the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) and the Joint Genome Institute (JGI; U.S. Department of Energy; https://phytozome.jgi.doe.gov/pz/portal.html#) using Bowtie (Langmead et al., 2009) to find the best hits in *Brassica* genomes. In addition, the coding sequences (CDS) of the assembled transcripts were identified using GENSCAN (Burge and Karlin, 1997). Then, the translated amino acid sequences were compared with the proteins in the NCBI nonredundant (nr), Swiss-Prot, Pfam, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Clusters of Orthologous Groups (COG) protein databases using Blast2GO software (Götz et al., 2008) with an *E*-value threshold of 10^{-5}. The expression of each transcript was measured by the FPKM (fragments per kilobase of exon model per million mapped reads) value, which was calculated using Cufflinks (Trapnell et al., 2010), and the statistical analyses were performed using Cuffdiff program in Cufflinks. Functional categorization of Gene Ontology (GO) terms corresponding to the selected genes was performed by searching the best protein match against the GO database (http://www.geneontology.org/) according to their molecular function, biological process, and cellular component ontologies. The KEGG pathway assignments were carried out by sequence searching against the database retrieved from http://www.genome.jp/kegg/. The GOs and KEGGs among the four samples were collected and the significances of changes were tested using a hyper-geometric test (one-tailed Fisher's exact test). A hierarchical clustering analysis of selected DETs across the four samples was performed based on their FPKM values using Cluster 3.0 software (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm).

TABLE 2. Summary of the transcriptomic data of the four samples.

Sample	Treatment	Raw reads	Clean reads	Valid ratio, %	Expected gene mapped by Bowtie
FL	SP2F 16°C	65,401,334	65,130,980	99.59	122,160
FR	SP2F 22°C	57,598,454	56,934,326	98.85	123,431
SL	SP2S 16°C	56,748,432	56,486,936	99.54	121,413
SR	SP2S 22°C	59,978,572	59,738,442	99.60	130,772

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

TABLE 3. Statistics of transcript length of the transcriptome assembly.

All	Minimum length	Median length	Mean length	N50[a]	Maximum length	Total length
Transcript 322134	201	662	916	1353	11,871	295,272,130
Gene 135702	201	495	784	1221	11,871	107,029,193

[a]N50 length is defined as the shortest sequence length at 50% of the transcriptome.

Validation of gene expression level by real-time PCR—To validate the gene expression estimated by FPKM value in RNA-Seq, specific primers (Table 1) were designed for several DETs using Beacon Designer 7.0 (Bio-Rad Laboratories, Hercules, California, USA). Quantitative real-time PCR assays were performed in triplicate using SYBR Green PCR Master Mix (Applied Biosystems, Waltham, Massachusetts, USA) on a QuantStudio 3 thermal cycler (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The *B. napus beta-actin7* gene was used as the internal control, and the variations in the four samples were calculated using a well-known delta-delta threshold cycle relative quantification method $2^{-\triangle CT}$ (Livak and Schmittgen, 2001), where $\triangle CT = (CT_{gene} - CT_{actin})$.

RESULTS

RNA-Seq and de novo transcriptome assembly—The earliest cytological abnormalities in the SP2S anther occurred during the stage from microsporocyte meiosis to tetrad microspore (Yu et al., 2015), and the length of the corresponding young floral buds was ≤3 mm. Thus, the young floral buds of the four treatments, which were designated S16 (SP2S at 16°C), S22 (SP2S at 22°C), F16 (SP2F at 16°C), and F22 (SP2F at 22°C), were collected simultaneously to extract mRNA. The four transcribed cDNA libraries were sequenced using an Illumina HiSeq 2500 system. We assessed the quality of the raw reads and trimmed them to form four sets of data, which had an average length of 95.16 bp. In total, we obtained 239.7 million clean reads, corresponding to 19.95 Gbp of sequences (Table 2). The sequencing depth of the four samples with an average of 4.98 Gbp data was approximately 49× because it was presumed that the total coding nucleotides of the 101,040 predicted genes in *B. napus* was 101.157 million bp (Chalhoub et al., 2014). The raw sequencing data of the above samples were submitted to NCBI (Sequence Read Archive [SRA] accessions SRR2052475, SRR2052499, SRR2052502, and SRR2052505).

The reads from all four samples were pooled together for de novo assembly using Trinity, and 322,134 fragments corresponding to 135,702 unigenes were produced (Table 3). The sequences of all 135,702 transcripts have been deposited in the NCBI Transcriptome Shotgun Assembly (TSA) database (accession GDFQ00000000). The total length was 107,029,193 bp, with an average length of 784 bp (201–11871 bp). There were 37,350 short fragments 200–300 bp in length and 9070 long transcripts (≥2000 bp), as shown in a vertical histogram (Fig. 1) of the size distribution of the transcripts.

Identification of cDNA-derived SSRs—We searched the de novo transcriptome assembly for SSRs that were defined as monomer, dimer, trimer, tetramer, pentamer, and hexamer repeats, with at least five repeats and being at least 18 bp in length. The results revealed a total of 24,157 potential cDNA-derived SSRs that were distributed in 22,927 unigenes (Fig. 2; Appendix S1). Most of these SSRs were dimers (9312), trimers (7792), and monomers (6274), but only a small portion of them were tetramers (191), pentamers (329), and hexamers (259). There were 200 transcripts containing two or more SSRs, and 1141 SSRs represented compound formation. Among all SSRs, the motif AG/CT had the highest frequency (792), followed by motifs GA/TC (715), A/T (426), CTC/GAG (271), AAG/CTT (218), and GAA/TTC (190). The SSRs identified in this study are valuable resources for identifying polymorphic SSR markers for the genetic analysis of rapeseed.

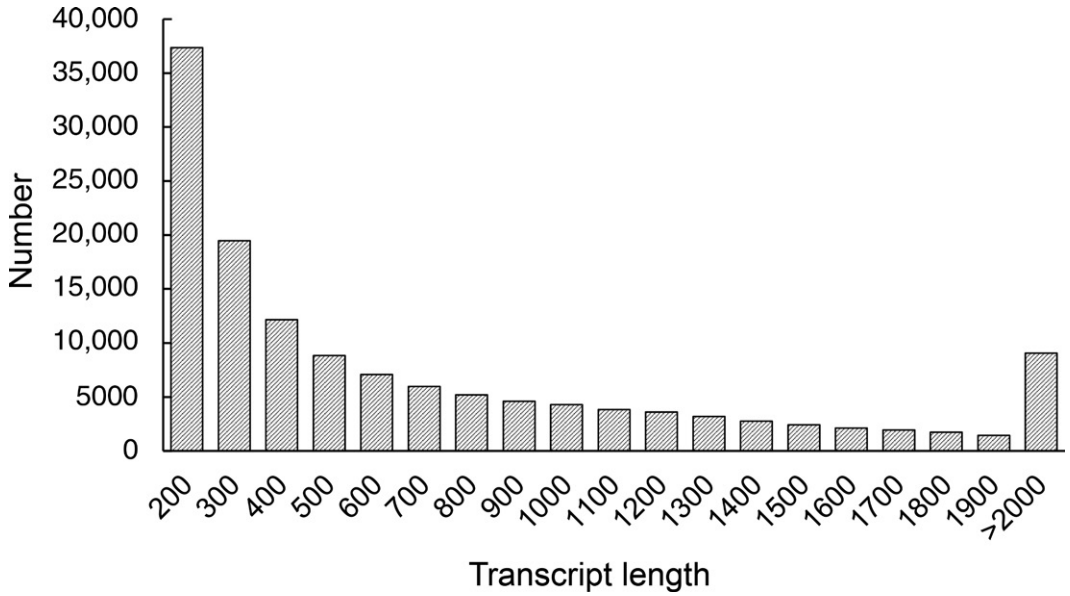

Fig. 1. Size distribution of the unigenes in the transcriptome assembly.

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

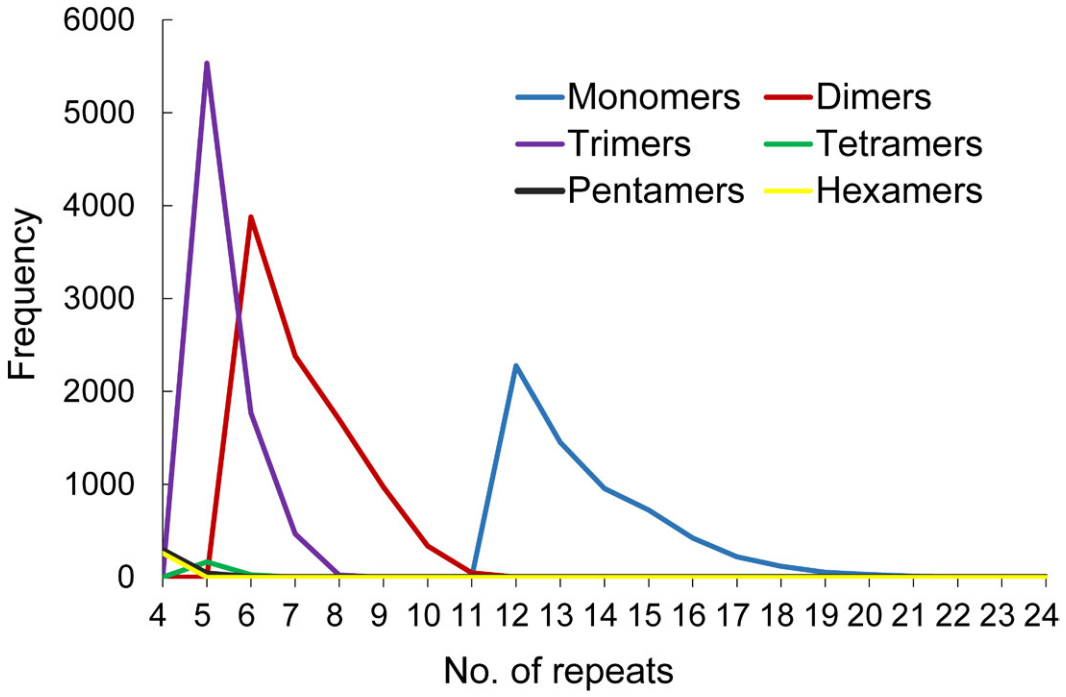

Fig. 2. Distribution of potential SSR motifs in the transcriptome assembly.

***Analysis of SNPs in the transcriptome*—**We searched for SNPs between the alignments of S22 vs. S16 and S22 vs. F22 using Bowtie software (Langmead et al., 2009). In total, 137 and 195 SNPs were identified in the comparisons of S22 vs. S16 and S22 vs. F22 (Appendix S2), respectively. Some SNPs found among the four samples may be produced by DNA mutation, alternative splicing, or RNA editing.

***Functional annotation of the assembled transcripts*—**We searched the 135,702 transcripts in the databases of NCBI and JGI using Bowtie (Langmead et al., 2009) and obtained 117,366 and 98,087 positive hits (Appendix S3), respectively. The remaining 13.51% and 27.72% of the 135,702 contigs had no matches in the NCBI and JGI *Brassica* databases. More than 90% of the identified transcripts had the highest homologies with *A. thaliana* (L.) Heynh. or *A. lyrata* (L.) O'Kane & Al-Shehbaz (Fig. 3), and fewer than 5% of the top matches hit *B. rapa*, *B. napus*, or *B. oleracea* due to the limited number of the *Brassica* protein sequences available in the NCBI database. We also searched the putative proteins transformed from the CDS of the 135,702 transcripts in various databases, and 33.77–55.59% of them had positive hits in the NCBI nr protein, Swiss-Prot,

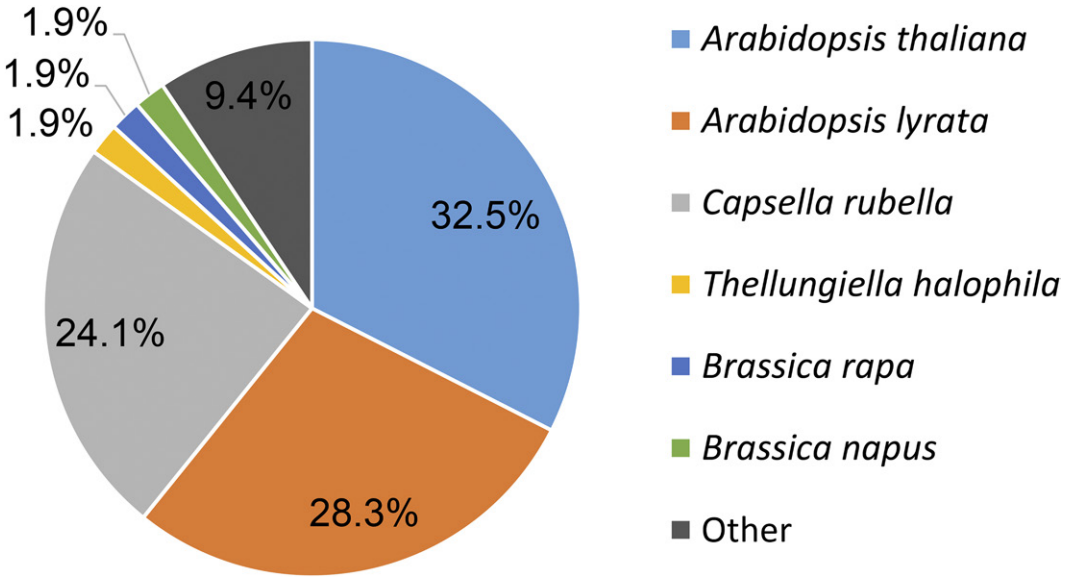

Fig. 3. Distribution of top-hit species by the transcriptome assembly in the NCBI nonredundant protein database.

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

Table 4. Percentage of BLAST hits of the 135,702 transcripts in the databases of NCBI (nucleotide and nonredundant protein), JGI, Swiss-Prot, Pfam, KEGG, and COG.

Database	No. of BLAST hits	% BLAST hits
NCBI	117,366	86.49
JGI	98,087	72.28
Swiss-Prot	45,829	33.77
NR	75,435	55.59
Pfam	52,063	38.35
KEGG	71,986	53.05
COG	43,680	32.19

Note: COG = Clusters of Orthologous Groups; KEGG = Kyoto Encyclopedia of Genes and Genomes; JGI = Joint Genome Institute; NCBI = National Center for Biotechnology Information nucleotide collection; NR = NCBI nonredundant protein.

Pfam, KEGG, and COG databases (Table 4). Moreover, the 43,680 identified proteins in COG were categorized into 25 clusters. Among these categories, the cluster for "general function prediction only" was the largest group, followed by the categories "posttranslational modification, protein turnover, chaperones" and "signal transduction mechanisms." The categories "cell motility" and "nuclear structures" had the fewest corresponding genes (Fig. 4).

***Analysis of differentially expressed transcripts and KEGG pathways*—**The expression level of each transcript in the four samples was estimated by FPKM and the significant DETs; the expression patterns were different among the four samples, with P value ≤0.05 and fold change ≥2 extracted (Appendix S4). When the transcript expressions of the same line treated under warm (22°C) and cool (16°C) conditions were compared, we found 2532 DETs in S22 vs. S16 (286 downregulated and 2235 upregulated in S22), whereas in the NIL SP2F, there were only 1224 DETs in F22 vs. F16 (221 downregulated and 988 upregulated in F22) (Table 5). There were 2320 DETs between SP2S (S22) and SP2F (F22) at 22°C (448 upregulated and 1782 downregulated in S22), whereas at 16°C, there were only 576 DETs between samples S16 and F16 (237 downregulated and 306 upregulated in S16) (Table 5). These DET numbers suggest that the transcriptional profile in TGMS SP2S was more affected by temperature than it was in SP2F and that warm conditions resulted in more DETs than cool temperatures. The comparisons of S22 vs. F22 and S22 vs. S16 indicated that 1013 DETs were shared (Fig. 5). In the sample of SP2S at 22°C, we found an abnormal regulation of some interesting genes that are important for cell wall construction, cell division and growth, lipid metabolism, hormone and temperature response, autophagy, RNA splicing, and nuclease activity (Fig. 6). For example, three genes encoding callose synthase GSL2, GSL10, and GSL12 were upregulated in SP2S at 22°C. Of these, GSL2 is required for the formation of the cell wall surrounding the microsporocyte and GSL10 is involved in pollen development at the mitotic division stage (Shi et al., 2015). Moreover, the genes encoding A6 and other beta-glucanases involved in callose dissolution were also highly upregulated. Some genes encoding the proteins that control meiosis and the cell cycle, including cyclin-dependent kinase inhibitor 7, ribonuclease H2 subunit B, S-phase kinase-associated protein 1, ribosomal protein, sister chromatid cohesion protein PDS, and cohesin complex subunit SA-1/2, were downregulated. In contrast, cyclin A and cell division cycle 20 were upregulated in the male sterile plants. These transcriptional regulations corresponded well to the cytological abnormalities, including asynchronous microsporocyte meiosis (Liu et al., unpublished data) and a delayed degradation of the tetrad wall, which led to aborted microspores with defective exine (Yu et al., 2015).

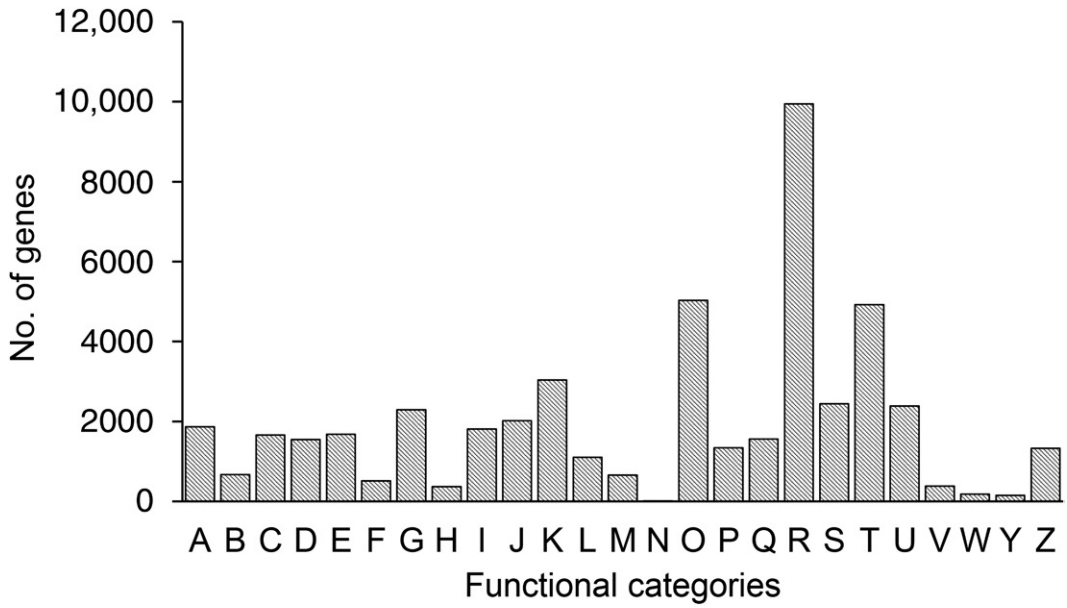

Fig. 4. COG functional categories of the transcriptome assembly. A = RNA processing and modification; B = chromatin structure and dynamics; C = energy production and conversion; D = cell cycle control, cell division, chromosome partitioning; E = amino acid transport and metabolism; F = nucleotide transport and metabolism; G = carbohydrate transport and metabolism; H = coenzyme transport and metabolism; I = lipid transport and metabolism; J = translation, ribosomal structure, and biogenesis; K = transcription; L = replication, recombination, and repair; M = cell wall/membrane/envelope biogenesis; N = cell motility; O = posttranslational modification, protein turnover, chaperones; P = inorganic ion transport and metabolism; Q = secondary metabolites biosynthesis, transport, and catabolism; R = general function prediction only; S = function unknown; T = signal transduction mechanisms; U = intracellular trafficking, secretion, and vesicular transport; V = defense mechanisms; W = extracellular structures; Y = nuclear structure; Z = cytoskeleton.

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

TABLE 5. The number of differentially expressed transcripts (DETs) among the four samples.

Comparison	Total DETs	Downregulated	Upregulated	Former sample specific	Latter sample specific
S22 vs. S16	2532	286	2235	2	9
S22 vs. F22	2320	448	1782	64	26
F22 vs. F16	1224	221	988	3	12
S16 vs. F16	576	237	306	10	23

Because the DETs were identified without biological replication, these results are only preliminary and require additional verification. For example, the expression of 10 selected genes involved in callose degradation, transcription regulation, pollen development, and signal transduction were further assayed by quantitative real-time PCR. The change of FPKM values of each selected gene across the four samples (Fig. 7) had a similar trend to the relative fold change estimated by the △Ct method. Thus, we believe that the RNA-Seq experiment can produce a useful gene expression profile.

The further evaluation of each of the numerous DETs individually would be challenging. Alternatively, we considered a rough comparison of GO and KEGG among those samples because GO and KEGG enrichment can reflect the overall changes of most transcripts. There were a total of 270 GOs (Appendix S5) that were influenced in SP2S under warm vs. cool temperature conditions. A comparison between SP2S and SP2F under the warm temperature condition also revealed 357 GOs being disturbed (Appendix S5). There were 51 and 49 plant KOs (KEGG orthology) showing significant differences between S22 and F22 and between S22 and S16 (Table 6, Appendix S6), respectively, including cell division (meiosis, circadian rhythm, cell cycle, and MAPK signaling pathway), transcription and translation related (ribosome, spliceosome, RNA polymerase, basal transcription factors), and energy supply related (oxidative phosphorylation, CO_2 fixation, citrate cycle, pyruvate metabolism, photosynthesis, and sugar metabolism). Some pathways related to amino acid metabolism, especially proline metabolism, which is important for pollen development (Mattioli et al., 2012), and protein modification and degradation-related KOs (ubiquitin-mediated proteolysis, proteasome, endocytosis, and lysosome), were also influenced. Additionally, lipid metabolism and the biosynthesis of flavonoids and carotenoids were disturbed.

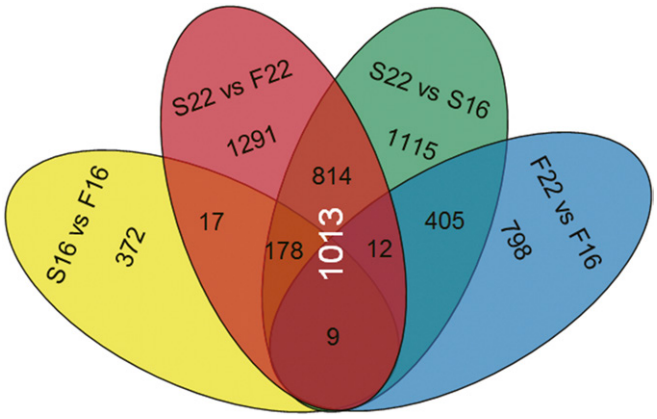

Fig. 5. Venn diagram showing the differentially expressed transcript (DET) number distribution in the pairwise comparisons among four samples.

DISCUSSION

***De novo transcriptome assembly revealed new information for future study*—**De novo assembly of *B. napus* transcriptome data has not been previously reported, with the exception of some related species such as *B. juncea* and *B. oleracea*, which have transcriptome assemblies containing 133,641 and 205,046 transcripts, respectively (Kim et al., 2014; Sinha et al., 2015). Alternatively, alignment of shotgun sequencing RNAs against a reference genome (reference-based strategy) is facilitated by the establishment of reference genome sequences of some *Brassica* species (http://brassicadb.org/brad/blastPage.php?, http://www.genoscope.cns.fr/brassicanapus/, and https://phytozome.jgi.doe.gov/pz/portal.html#). Although the reference-based strategy is powerful, it is unable to detect structural alterations that are not present in the reference sequence data; this is particularly true when the read lengths are short (Birol et al., 2009). Spliced reads that span large introns can be missed because aligners often allow only a fixed length of introns (Martin and Wang, 2011). The success of the reference-based strategy depends on the quality of the reference genome, but many genome assemblies contain misassemblies and large deletions (Martin and Wang, 2011). Thus, even if a reference genome is available for *Brassica* plants, de novo transcriptome assembly should be performed (Xu et al., 2016). Moreover, transcriptome sequencing of different genotypes can provide novel insights into the diverse rearrangements of chromosome fragments, which may be caused by gene translocation, deletion, fusion, and/or recombination, underlying the genetic diversity among various genetic resources. In addition, the sequencing data obtained in this study can be used to develop cDNA-derived molecular markers, including SSRs and SNPs.

De novo transcriptome assembly can recover transcripts that are transcribed from segments of the genome that are missing from the genome assembly (Martin and Wang, 2011). Our assembly, containing 135,702 contigs, was larger than a previous estimation of approximately 101,040 gene models in the *B. napus* genome (Chalhoub et al., 2014), as were the assemblies of *B. juncea* and *B. oleracea* (Kim et al., 2014; Sinha et al., 2015). In addition to redundant information, overlapping of contigs, alternative splicing, and mRNA edition, we believed that the assembly should also contain some novel genes and/or transcripts because it was unlikely that the rest of the unmatched contigs (13.51% in the NCBI nr protein database and 27.72% in JGI) in the 135,702 transcripts were false assemblies. Although some unavoidable mistakes might exist in our assembly, it should be useful as a reference for the further analysis of gene expression and functional genomics studies. For instance, in the transcriptomic profiling comparison between the male sterile plants induced by the herbicide amidosulfuron and the control using a digital gene expression tag method, we identified 516 DETs that aligned well in the de novo transcriptome assembly (Liu et al., 2017). Of those, 486 DETs had perfect match in the NCBI database. Moreover, our transcriptome sequences will enhance the

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

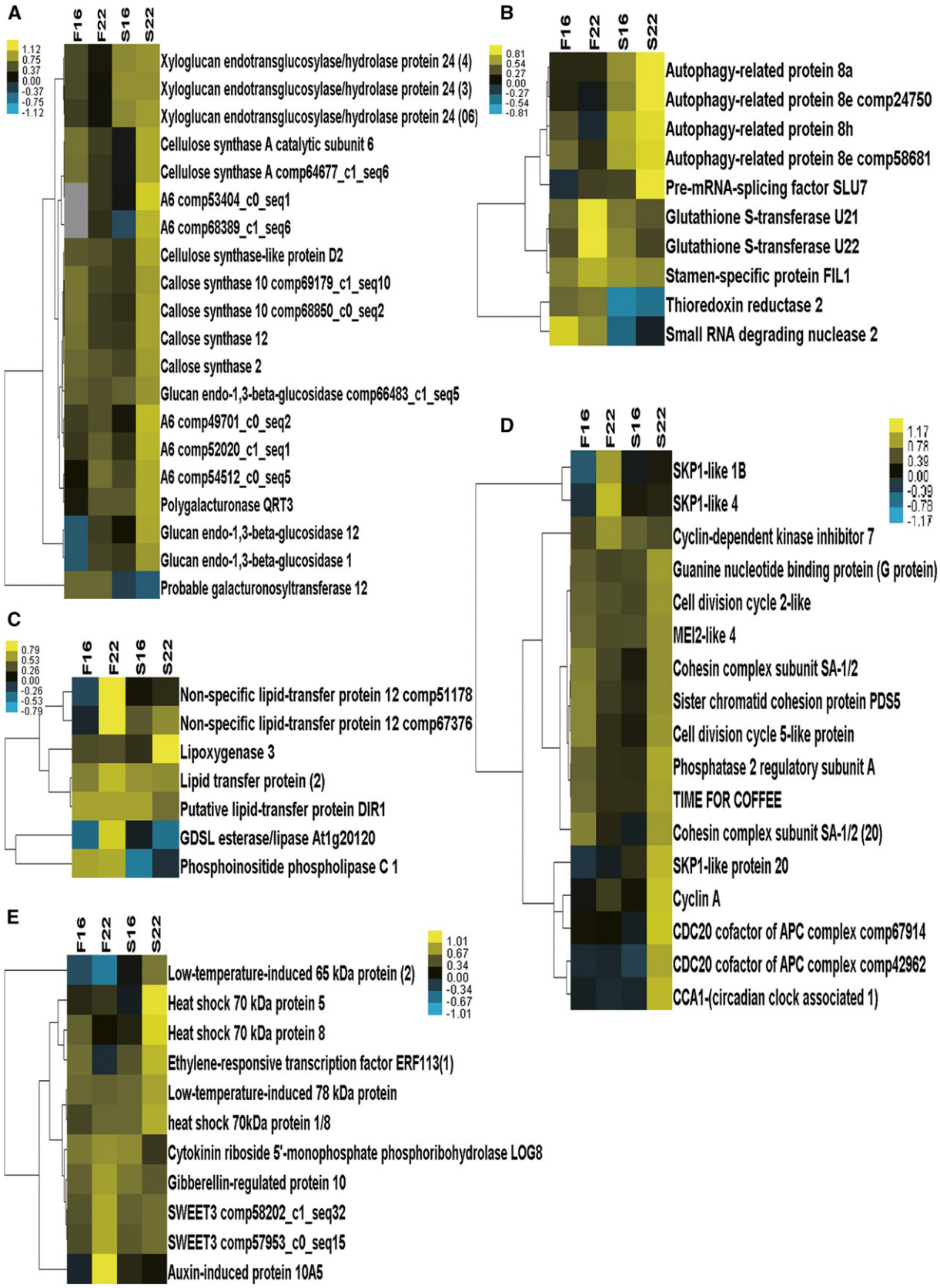

Fig. 6. Heatmap analysis of five subsets of the selected differentially expressed transcripts (DETs) in the four samples: cell wall (A), 10 selected genes (B), lipid metabolism (C), cell division (D), and hormone and stress response (E). Expressed genes are FPKM normalized, and the values are log10 transformed.

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

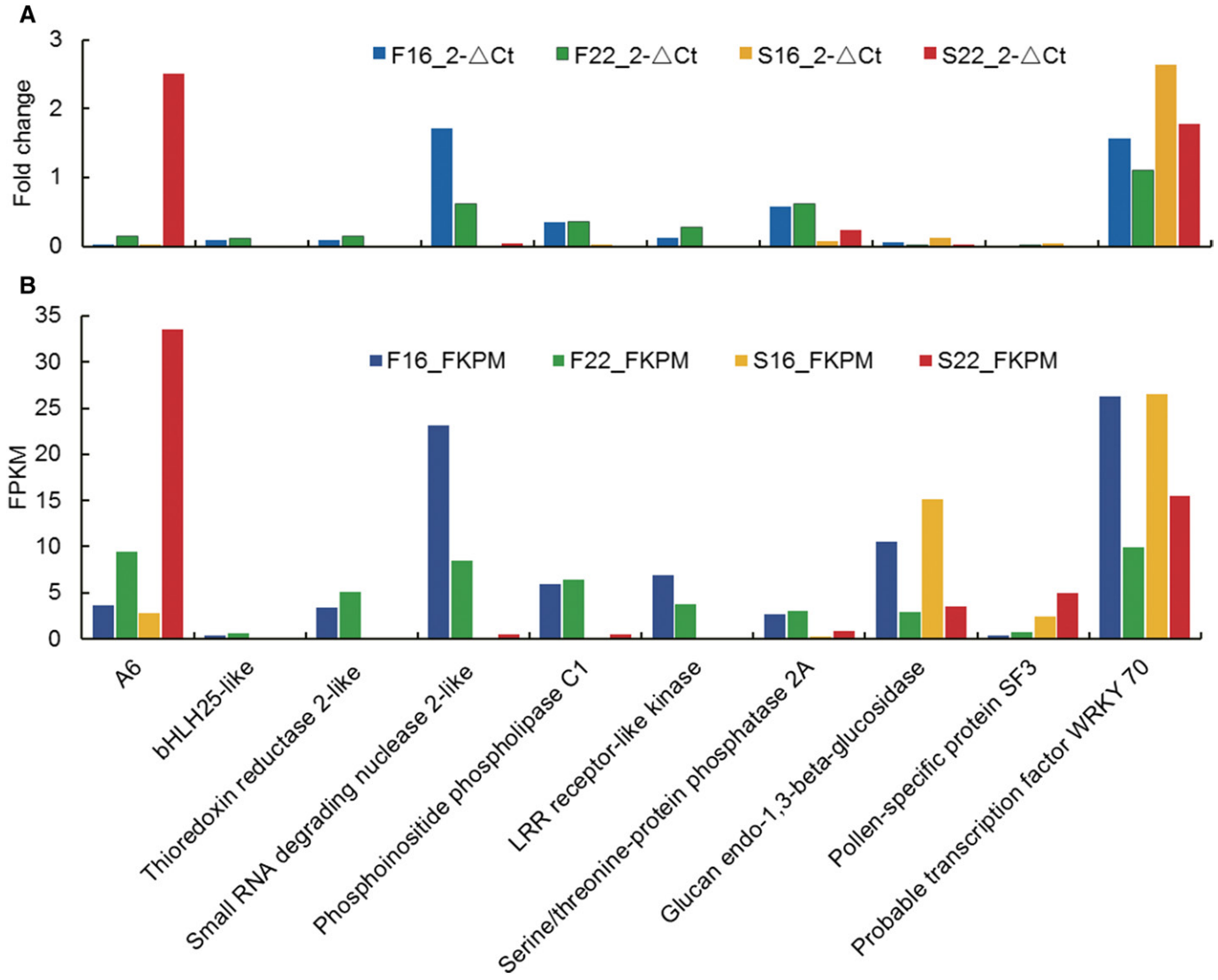

Fig. 7. Comparison of the relative fold change (estimated by the $2^{-\triangle Ct}$ method) (A) and the FPKM values (B) of the 10 selected genes across the four samples.

quality of gene annotation and functional analysis of the rapeseed genome because most genes in the shotgun sequencing assemblies were not yet annotated.

The aberrant transcriptional regulations corresponded to the abnormal phenotypes—Although it is difficult to anchor the male sterility gene itself from analysis of global differential expression between the male sterile mutant and wild type, this approach is useful to unveil some subsequent genes involved in anther abortion (i.e., male sterility–associated genes). The results of this study suggest that although male sterility genes are the trigger, the subsequent pollen abortion often involves many genes and multilevel biological pathways, as most male sterilities do (Ma, 2005; Zhu et al., 2011; An et al., 2014; Kim et al., 2014; Ma et al., 2015; Qu et al., 2015). In other studies, some common pathways were disrupted, such as the ubiquitin-proteasome system in the TGMS system considered in the current study, photoperiod-sensitive male sterility cotton (Ma et al., 2013), and rice *tms5* mutation (Zhou et al., 2014). This finding suggests that the pollen-abortion patterns of different species could share some common steps (Parish et al., 2012; De Storme and Geelen, 2014). Moreover, microsporogenesis of plants undergoing abiotic stress typically shows a reduced level of energy supply in the anthers (Kubala et al., 2015). The differentially expressed transcripts in such pathways as citrate cycle, pyruvate metabolism, oxidative phosphorylation, photosynthesis, CO_2 fixation, and sucrose metabolism will affect pollen development in SP2S because carbohydrates and energy supply are crucial to anther and pollen development. Environmental stress and, more specifically, temperature stress impact the subsequent processes in meiotic cell division (De Storme and Geelen, 2014). The impact of temperature stress on meiosis, circadian rhythm, cell cycle, basal transcription factors, RNA polymerase, and purine and pyrimidine metabolism will arrest the anther and pollen development of TGMS SP2S. Moreover, the biosynthesis of flavonoids and carotenoids, as well as fatty acid metabolism, which were impaired in SP2S, are also thought to be crucial for pollen wall construction (Parish et al., 2012; Ma et al., 2013). These

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

TABLE 6. Plant KEGG Orthologies (KOs) that were significantly influenced in the comparisons between S22 and S16 and between S22 and F22.

KO ID	KO name	No. of DET S22 vs. S16	No. of DET S22 vs. F22	Genes in the pathway
ko00520	Amino sugar and nucleotide sugar metabolism	55	44	254
ko00330	Arginine and proline metabolism	47	41	192
ko04144	Endocytosis	65	52	280
ko00190	Oxidative phosphorylation	82	78	440
ko03010	Ribosome	326	187	616
ko03040	Spliceosome	57	45	99
ko04712	Circadian rhythm (plant)	47	49	171
ko04010	MAPK signaling pathway	36	31	142
ko03060	Protein export	35	31	136
ko00310	Lysine degradation	33	18	116
ko00071	Fatty acid metabolism	36	28	142
ko00565	Ether lipid metabolism	35	21	136
ko00592	α-Linolenic acid metabolism	25	29	94
ko00440	Phosphonate and phosphinate metabolism	13	13	35
ko00640	Propanoate metabolism	18	15	69
ko00620	Pyruvate metabolism	35	30	231
ko00561	Glycerolipid metabolism	15	12	65
ko00564	Glycerophospholipid metabolism	24	25	154
ko00380	Tryptophan metabolism	22	21	135
ko04113	Meiosis	29	22	209
ko00710	Carbon fixation	12	16	233
ko00591	Linoleic acid metabolism	13	7	35
ko00500	Starch and sucrose metabolism	18	16	240
ko00230	Purine metabolism	36	43	340
ko03020	RNA polymerase	14	13	83
ko03050	Proteasome	22	20	171
ko04111	Cell cycle	31	18	281
ko00650	Butanoate metabolism	13	12	103
ko04142	Lysosome	18	16	167
ko04120	Ubiquitin-mediated proteolysis	31	33	345
ko00240	Pyrimidine metabolism	24	32	251
ko00195	Photosynthesis	16	23	155
ko00720	CO_2 fixation	8	10	59
ko04745	Phototransduction (fly)	4	2	21
ko00360	Phenylalanine metabolism	16	24	188
ko00280	Valine, leucine, and isoleucine degradation	33	38	161
ko00290	Valine, leucine, and isoleucine biosynthesis	8	8	97
ko00020	Citrate cycle	12	18	167
ko00910	Nitrogen metabolism	18	10	131
ko00360	Phenylpropanoid biosynthesis	14	20	203
ko00196	Photosynthesis (antenna proteins)	3	6	40
ko04114	Oocyte meiosis	29	18	258
ko00410	beta-Alanine metabolism	15	13	59
ko00340	Histidine metabolism	13	7	54
ko00941	Flavonoid biosynthesis	4	5	43
ko00906	Carotenoid biosynthesis	5	9	60
ko04146	Peroxisome	14	22	102
ko00310	Lysine degradation	23	18	70
ko00970	Aminoacyl-tRNA biosynthesis	15	8	198
ko03022	Basal transcription factors	8	—	105
ko00563	Glycosylphosphatidylinositol	3	—	22

Note: DET = differentially expressed transcript.

aberrant genetic regulations in SP2S corresponded well to abnormal phenotypes such as an asynchronous microsporocyte meiosis (Liu et al., unpublished data), delayed dissolution of the tetrad wall, exine fusion, and premature breakdown of the tapetum (Yu et al., 2015).

CONCLUSIONS

In summary, a total of 135,702 transcripts were revealed in the young flower buds after RNA-Seq and de novo assembly using Trinity software. To our knowledge, this is the first large-scale transcriptome sequencing of a rapeseed TGMS line. These transcriptome data may serve as a reference and provide important insights for future studies of genetic regulation of rapeseed microsporogenesis.

LITERATURE CITED

AN, H., Z. YANG, B. YI, J. WEN, J. SHEN, J. TU, C. MA, ET AL. 2014. Comparative transcript profiling of the fertile and sterile flower buds of pol CMS in *B. napus*. *BMC Genomics* 15: 258.

BANCROFT, I., C. MORGAN, F. FRASER, J. HIGGINS, R. WELLS, L. CLISSOLD, D. BAKER, ET AL. 2011. Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nature Biotechnology* 29: 762–766.

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use

Birol, I., S. D. Jackman, C. B. Nielsen, J. Q. Qian, R. Varhol, G. Stazyk, R. D. Morin, et al. 2009. *De novo* transcriptome assembly with ABySS. *Bioinformatics (Oxford, England)* 25: 2872–2877.

Burge, C., and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* 268: 78–94.

Chalhoub, B., F. Denoeud, S. Liu, I. A. P. Parkin, H. Tang, X. Wang, J. Chiquet, et al. 2014. Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345: 950–953.

De Storme, N., and D. Geelen. 2014. The impact of environmental stress on male reproductive development in plants: Biological processes and molecular mechanisms. *Plant, Cell & Environment* 37: 1–18.

Ding, J. H., Q. Lu, Y. D. Ouyang, H. L. Mao, P. B. Zhang, J. L. Yao, C. G. Xu, et al. 2012. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proceedings of the National Academy of Sciences, USA* 109: 2654–2659.

Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36: 3420–3435.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494–1512.

Kim, H. A., C. J. Lim, S. Kim, J. K. Choe, S. H. Jo, N. Baek, and S. Y. Kwon. 2014. High-throughput sequencing and *de novo* assembly of *Brassica oleracea* var. *capitata* L. for transcriptome analysis. *PLoS One* 9: e92087.

Kubala, S., M. Garnczarska, Ł. Wojtyla, A. Clippe, A. Kosmala, A. Żmieńko, S. Luttse, et al. 2015. Deciphering priming-induced improvement of rapeseed (*Brassica napus* L.) germination through an integrated transcriptomic and proteomic approach. *Plant Science* 231: 94–113.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics (Oxford, England)* 25: 2078–2079.

Liu, X. Q., C. Y. Yu, J. G. Dong, S. W. Hu, and A. X. Xu. 2017. Acetolactate synthase-inhibiting gametocide amidosulfuron causes chloroplast destruction, tissue autophagy, and elevation of ethylene release in rapeseed. *Frontiers in Plant Science* 8: 1625.

Livak, K. J., and T. D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods (San Diego, Calif.)* 25: 402–408.

Ma, H. 2005. Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annual Review of Plant Biology* 56: 393–434.

Ma, J., H. Wei, J. Liu, M. Song, C. Pang, L. Wang, W. Zhang, et al. 2013. Selection and characterization of a novel photoperiod-sensitive male sterile line in upland cotton. *Journal of Integrative Plant Biology* 55: 608–618.

Ma, Y., J. Kang, J. Wu, Y. Zhu, and X. Wang. 2015. Identification of tapetum-specific genes by comparing global gene expression of four different male sterile lines in *Brassica oleracea. Plant Molecular Biology* 87: 541–554.

Marioni, J. C., C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad. 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* 18: 1509–1517.

Martin, J. A., and Z. Wang. 2011. Next-generation transcriptome assembly. *Nature Reviews. Genetics* 12: 671–682.

Mattioli, R., M. Biancucci, C. Lonoce, P. Costantino, and M. Trovato. 2012. Proline is required for male gametophyte development in *Arabidopsis. BMC Plant Biology* 12: 236.

Ozsolak, F., and P. M. Milos. 2011. RNA sequencing: Advances, challenges and opportunities. *Nature Reviews. Genetics* 12: 87–98.

Pan, Y., Q. Li, Z. Wang, Y. Wang, R. Ma, L. Zhu, G. He, et al. 2014. Genes associated with thermosensitive genic male sterility in rice identified by comparative expression profiling. *BMC Genomics* 15: 1114.

Parish, R. W., H. A. Phan, S. Iacuone, and S. F. Li. 2012. Tapetal development and abiotic stress: A centre of vulnerability. *Functional Plant Biology* 39: 553–559.

Paritosh, K., V. Gupta, S. K. Yadava, P. Singh, A. K. Pradhan, and D. Pental. 2014. RNA-seq based SNPs for mapping in *Brassica juncea* (AABB) synteny analysis between the two constituent genomes A (from *B. rapa*) and B (from *B. nigra*) shows highly divergent gene block arrangement and unique block fragmentation patterns. *BMC Genomics* 15: 396.

Qu, C., F. Fu, M. Liu, H. Zhao, C. Liu, J. Li, Z. Tang, et al. 2015. Comparative transcriptome analysis of recessive male sterility (RGMS) in sterile and fertile *Brassica napus* lines. *PLoS One* 10: e0144118.

Shi, X., X. Sun, Z. Zhang, D. Feng, Q. Zhang, L. Han, J. Wu, et al. 2015. *GLUCAN SYNTHASE-LIKE 5* (*GSL5*) plays an essential role in male fertility by regulating callose metabolism during microsporogenesis in rice. *Plant & Cell Physiology* 56: 497–509.

Sinha, S., V. K. Raxwal, B. Joshi, A. Jagannath, S. Katiyar-Agarwal, S. Goel, A. Kumar, et al. 2015. *De novo* transcriptome profiling of cold-stressed siliques during pod filling stages in Indian mustard (*Brassica juncea* L.). *Frontiers in Plant Science* 6: 932.

Thiel, T., W. Michalek, R. K. Varshney, and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106: 411–422.

Tong, C., X. Wang, J. Yu, J. Wu, W. Li, J. Huang, C. Dong, et al. 2013. Comprehensive analysis of RNA-seq data reveals the complexity of the transcriptome in *Brassica rapa. BMC Genomics* 14: 689.

Trapnell, C., B. Williams, G. Pertea, A. Mortazavi, G. Kwan, J. van Baren, S. Salzberg, et al. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28: 511–515.

Xu, H., L. Chen, B. Song, X. Fan, X. Yuan, and J. Chen. 2016. *De novo* transcriptome sequencing of pakchoi (*Brassica rapa* L. *chinensis*) reveals the key genes related to the response of heat stress. *Acta Physiologiae Plantarum* 38: 252.

Yu, C., Y. Guo, J. Ge, Y. Hu, J. Dong, and Z. Dong. 2015. Characterization of a new temperature sensitive male sterile line SP2S in rapeseed (*Brassica napus* L.). *Euphytica* 206: 473–485.

Zhang, J., Z. Liu, X. Liu, J. Dong, H. Pang, and C. Yu. 2016. Proteomic alteration of a thermo-sensitive male sterility SP2S in rapeseed (*Brassica napus*) in response to mild temperature stress. *Plant Breeding* 135: 191–199.

Zhao, S., W. P. Fung-Leung, A. Bittner, K. Ngo, and X. Liu. 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9: e78644.

Zhou, H., M. Zhou, Y. Yang, J. Li, L. Zhu, D. Jiang, J. Dong, et al. 2014. RNase Z^{S1} processes Ub_{L40} mRNAs and controls thermosensitive genic male sterility in rice. *Nature Communications* 5: 4884.

Zhu, J., Y. Lou, X. Xu, and Z. N. Yang. 2011. Genetic pathway for tapetum development and function in *Arabidopsis. Journal of Integrative Plant Biology* 53: 892–900.

Downloaded From: https://staging.bioone.org/journals/Applications-in-Plant-Sciences on 25 Dec 2024
Terms of Use: https://staging.bioone.org/terms-of-use