



Article

Transcriptome Sequencing Analyses between the Cytoplasmic Male Sterile Line and Its Maintainer Line in Welsh Onion (*Allium fistulosum* L.)

Qianchun Liu ^{1,†}, Yanping Lan ^{2,†}, Changlong Wen ¹, Hong Zhao ¹, Jian Wang ¹ and Yongqin Wang ^{1,*}

¹ Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences/Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (North China), Ministry of Agriculture, Beijing 100097, China; lqc072150@126.com (Q.L.); wenchanglong@nercv.org (C.W.); zhaohong@nercv.org (H.Z.); wyqty@sohu.com (J.W.)

² Institute of Agricultural Integrated Development, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; lanyanping2000@126.com

* Correspondence: wangyongqin@nercv.org; Tel.: +86-10-5150-3125; Fax: +86-10-5150-5002

† These authors contributed equally to this work.

Academic Editor: Jianhua Zhu

Received: 1 April 2016; Accepted: 27 June 2016; Published: 1 July 2016

Abstract: Cytoplasmic male sterility (CMS) is important for exploiting heterosis in crop plants and also serves as a model for investigating nuclear–cytoplasmic interaction. The molecular mechanism of cytoplasmic male sterility and fertility restoration was investigated in several important economic crops but remains poorly understood in the Welsh onion. Therefore, we compared the differences between the CMS line 64-2 and its maintainer line 64-1 using transcriptome sequencing with the aim of determining critical genes and pathways associated with male sterility. This study combined two years of RNA-seq data; there were 1504 unigenes (in May 2013) and 2928 unigenes (in May 2014) that were differentially expressed between the CMS and cytoplasmic male maintainer Welsh onion varieties. Known CMS-related genes were found in the set of differentially expressed genes and checked by qPCR. These genes included F-type ATPase, NADH dehydrogenase, cytochrome c oxidase, etc. Overall, this study demonstrated that the CMS regulatory genes and pathways may be associated with the mitochondria and nucleus in the Welsh onion. We believe that this transcriptome dataset will accelerate the research on CMS gene clones and other functional genomics research on *A. fistulosum* L.

Keywords: Welsh onion; transcriptome sequencing; cytoplasmic male sterility (CMS)

1. Introduction

As an important vegetable crop, The Welsh onion (*Allium fistulosum* L.) is cultivated worldwide from tropical Asia to Northeast Asia. Because it is rich in propylene sulfide having bactericidal and anti-inflammatory effects, the Welsh onion has also been used as an herbal medicine for diverse diseases, such as febrile disease, headache, diarrhea, abdominal pain, eye-related disorders, and habitual abortion [1]. The Welsh onion is a cross-pollinated crop, and heterosis has been extensively applied in breeding [2]. Artificial emasculation is difficult in this species because of its small flower. Therefore, cytoplasmic male sterility (CMS) plays a crucial role in heterosis breeding.

CMS is a phenotypic trait that is widespread among plants and results in the inability of a plant to produce viable pollen [3], using CMS lines to obtain F₁ hybrids, which is the preferred method in heterosis breeding. The CMS plants are usually obtained by natural populations [4,5], crossed by different genera and species plants [6,7], physical and chemical mutagenesis [8], and plant protoplast

fusion [9,10]. As early as 1972, Nishimura and Shibano discovered spontaneous CMS of onion in Japan [11]. Thereafter, Chinese researchers discovered the CMS line in the Welsh onion (*Allium fistulosum* L.) [12]. It has been proven to be an efficient way to obtain a heterosis hybrid, harboring important commercial benefit in the Welsh onion seed market [13–15].

Recently, the chromosomal location of the pollen fertility-restoring gene in *A. fistulosum* has been reported [15]. Furthermore, the CMS is closely linked to molecular markers. Restriction fragment length polymorphism (RFLP), rapid amplification of cDNA ends (RACE) and PCR-markers were used to identify cytoplasm types in the onion [16–18]. Subsequently, isozymes, random amplified polymorphic DNA (RAPD) and sequence characterized amplified region (SCAR) markers linked to the fertility restoring gene (*Rf*) locus were successfully developed [13]. However, the genetic gene or linked molecular marker of CMS in the Welsh onion has not been performed.

Some studies have shown that male-sterility-inducing factors are present in mitochondrial genomes [19], while others have reported that nuclear–mitochondrial interaction results in CMS [20,21]. This study, focusing on transcriptome sequencing comparative analysis of the CMS line and its maintainer line, may aid in screening the CMS-related genes and pathways in the Welsh onion and contribute to a better understanding of the CMS interaction of the nuclear and mitochondrial genomes in *A. fistulosum*.

2. Results

2.1. Sequencing and Transcriptome Assembly

To maximize the range of transcript diversity, a mixed RNA sample from three inflorescences in the flowering stage of the Welsh onion was prepared for RNA-seq using the Illumina HiSeqTM 2000. After stringent quality assessment and data filtering, we generated 57.19 million read pairs that corresponded to 11.55 Gb of sequence data in May 2013, and 56.82 million read pairs that corresponded to 11.48 Gb of sequence data in May 2014 (Table 1). The Q30 percentages exceeded 85% in both years. In May 2013, using the Trinity de novo assembly program, the next-generation short-read sequences were assembled into 97,230 transcripts which had a mean length of 917.56 bp. A total of 52,126 unigenes with an average length of 755.69 bp were obtained. The N50 values of the transcripts and unigenes were 1419 and 1247 bp, respectively. In May 2014, there were 308,353 transcripts with a mean length of 1122.65 bp, and there were 70,360 unigenes with an average length of 879.72 bp. The N50 values of the transcripts and unigenes were 1813 and 1267 bp, respectively (Table S1). The longest transcript was taken as the sample unigene for data from each year.

Table 1. Output statistics of sequencing analysis.

Sample	Total Reads	Total Nucleotides (bp)	GC Percentage	Q30 Percentage	Total Mapping Reads Percentage
CMS-2013	29,169,887	5,891,890,308	49.30%	87.75%	89.99%
CMM-2013	28,020,145	5,659,402,069	45.46%	89.18%	83.95%
CMS-2014	20,978,451	4,237,366,829	41.80%	95.07%	79.22%
CMM-2014	23,309,156	4,708,122,991	41.92%	95.11%	79.55%

CMS-2013: Cytoplasmic male sterility in May 2013; CMM-2013: Cytoplasmic male maintainer in May 2013; CMS-2014: Cytoplasmic male sterility in May 2014; CMM-2014: Cytoplasmic male maintainer in May 2014. Total mapping reads percentage: Total number of mapping reads percentage, including uniquely mapping reads and multiply mapping reads.

The distribution curve for each sample was relatively flat in the randomly fragmented transcriptome. The number of sample sequences reached saturation under current conditions (Figure S1). The length distributions of the ORFs are shown in Figure S2. These results indicate that the sequencing quality and throughput were sufficient for the subsequent analyses.

2.2. Functional Annotation and Classification

Unigenes that had at least one significant match to an existing gene model in BLAST were determined using searches against the NCBI NR, Swiss-Prot, KEGG, GO, and COG. After screening all differentially expressed genes, we extracted identified information from the unigenes (Figure 1), and presented the overview of the results in Table S2.

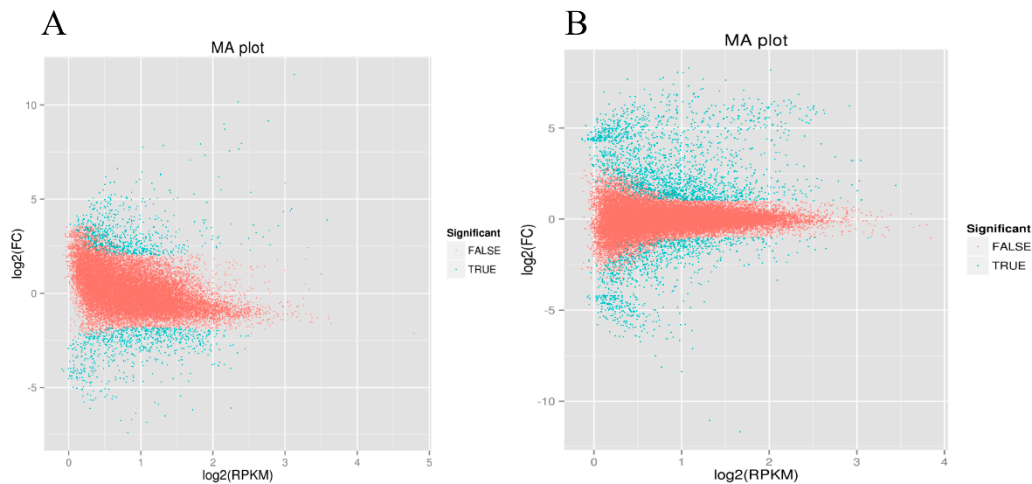


Figure 1. Differentially expressed genes in CMS vs. cytoplasmic male maintainer: (A) Whole-study overview of log-fold changes in gene expression in CMS vs. cytoplasmic male maintainer in May 2013; (B) Whole-study overview of log-fold changes in gene expression in CMS vs. cytoplasmic male maintainer in May 2014. The x-axis indicates the absolute expression levels (\log_2 RPKM (Reads Per Kilobase per Million mapped reads)). The y-axis indicates the \log_2 -FC (fold changes) between the two samples. Genes for which differential expression is significant are shown as blue dots ($\text{Log}_2\text{FC} > 1$ or < -1 ; FDR (False Discovery Rate) < 0.01).

Using the COG functional classification, only 8326 of 52,126 (15.97%, May 2013) and 8201 of 70,360 (11.66%, May 2014) unigenes were annotated with COG in each assembly (Table S2). The unigenes were annotated in 22 (May 2013) and 23 (May 2014) COG categories (Figure 2). Among these, “the general function prediction” was the largest group, followed by “the replication, recombination and repair”, “transcription”, and “signal transduction mechanisms”, whereas the “Nuclear structure”, “Extracellular structures”, and “Cell motility” groups were the smallest.

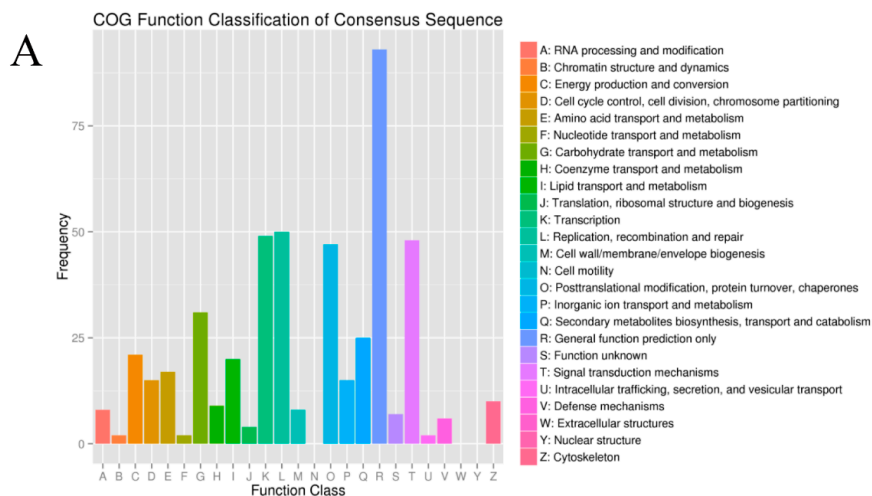


Figure 2. Cont.

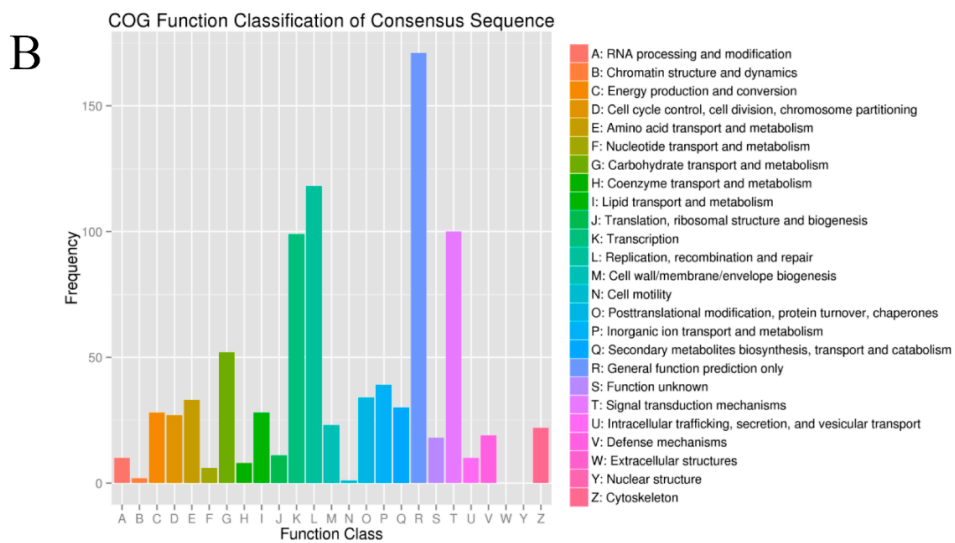


Figure 2. Cluster of Orthologous Groups (COG) categories of the unigenes: (A) sequencing data from May 2013; and (B) sequencing data from May 2014.

2.3. Gene Ontology (GO) Classification

We applied GO terms to classify the functions of the assembled transcripts and describe gene products. A total of 20,599 unigenes (819 different unigenes in May 2013) and 17,541 unigenes (1201 different unigenes in May 2014) were assigned to 56 functional groups using GO assignment. The major different sub-categories within each of the three main categories (i.e., cellular component, molecular function and biological process) of the GO classification are shown in Figure 3: there were five sub-categories in the cellular component cluster, i.e., “nucleus” (GO:0005634), “plasma membrane” (GO:0005886), “mitochondrion” (GO:0005739), “cytoplasmic membrane-bounded vesicle” (GO:0016023), and “membrane” (GO:0016020); there were four sub-categories in the molecular function cluster, i.e., “ATP binding” (GO:0005524), “binding” (GO:0005488), “metal ion binding” (GO:0046872) and “mitochondrial respiratory chain” (GO:0005746); and there were four sub-categories in the biological process cluster, i.e., “pollen tube growth” (GO:0009860), “phosphorylation” (GO:0016310), “oxidation-reduction process” (GO:0055114), and “response to abscisic acid” (GO:0009737).

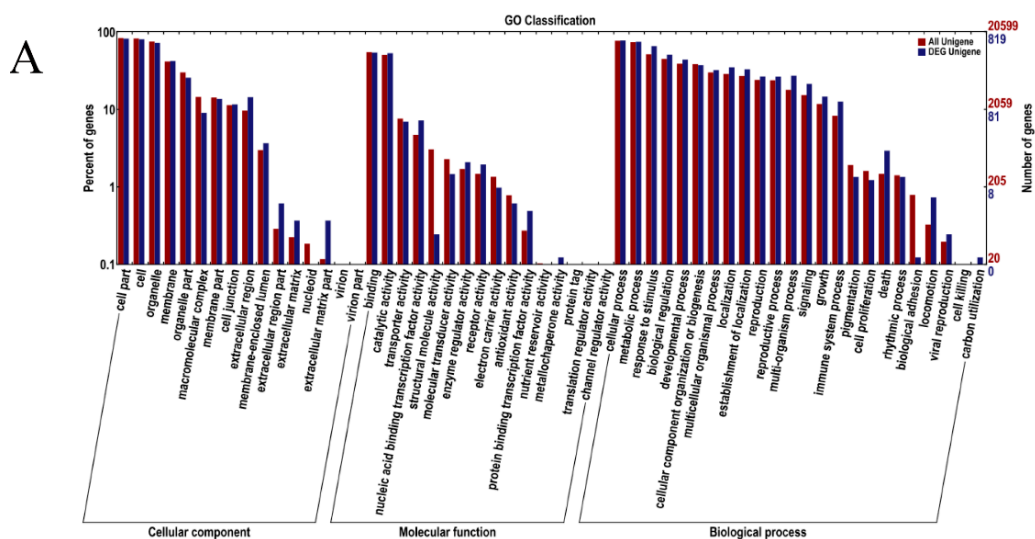


Figure 3. Cont.

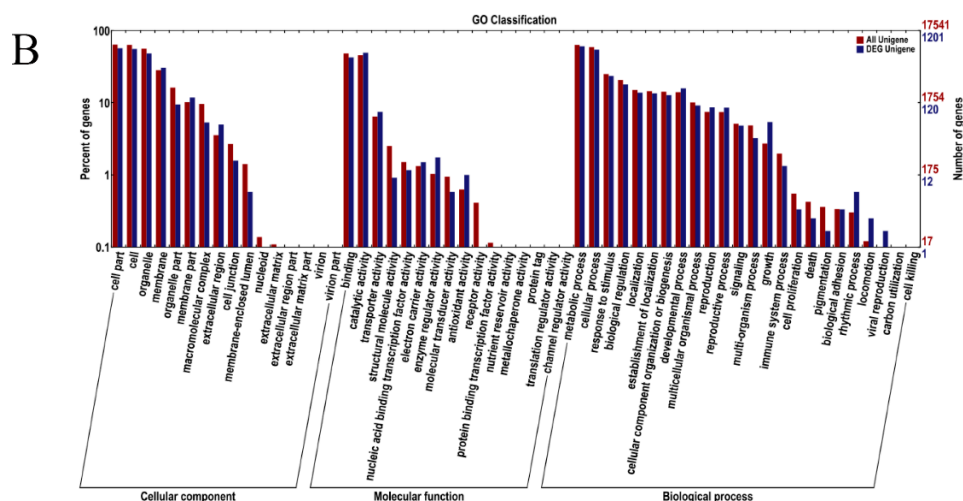


Figure 3. Gene Ontology Classification: (A) sequencing data from May 2013; and (B) sequencing data from May 2014. The unigenes are summarized into three main categories: cellular component, molecular function and biological process. The right side of the *y*-axis is the number of the genes; the left side of the *y*-axis is the percent of genes.

2.4. Pathway Enrichment Analysis of DEGs (Differentially Expressed Unigenes)

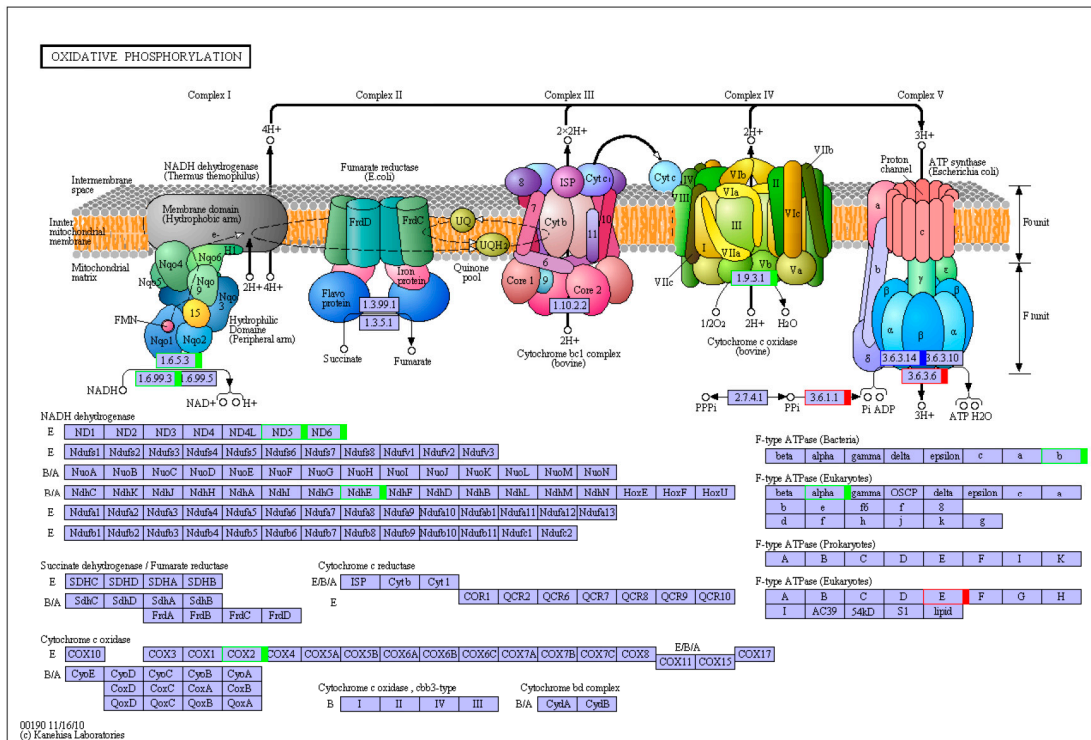
To categorize the biological functions of the DEGs, we performed a KEGG pathway enrichment analysis. The DEGs that significantly enriched pathways in May 2013 and May 2014 were the “plant-pathogen interaction” pathway, followed by the “oxidative phosphorylation” and “plant hormone signal transduction” pathways. In addition, pathways such as “starch and sucrose metabolism”, “glycerophospholipid metabolism”, and “protein processing in endoplasmic reticulum” composed a large component of the DEGs with pathway annotation (Table 2). In the above-mentioned DEGs with the best-represented KEGG pathways, there was consistency in enzyme regulation in the “oxidative phosphorylation” pathway between May 2013 and May 2014 (Figure 4). The enzymes “NADH: ubiquinone reductase (H⁺-translocating)”, “ND5” (NADH dehydrogenase 5), “COX2” (cytochrome-*c* oxidase 2), and “F-type ATPase α” (Eukaryotes) were related to down-regulated genes, whereas “inorganic diphosphatase”, “ATP phosphohydrolase” (H⁺-exporting), and “F-type ATPase E” (Eukaryotes) were related to up-regulated genes. “ATP phosphohydrolase” (H⁺-transporting) was related to both up- and down-regulated genes.

Table 2. Differentially expressed unigenes with significantly enriched pathways.

Pathway	DEGs with Pathway Annotation		All genes with Pathway Annotation		<i>p</i> -Value		Pathway ID
	May 2013	May 2014	May 2013	May 2014	May 2013	May 2014	
Plant-pathogen interaction	13 (9.35%)	17 (8.85%)	111 (2.73%)	113 (3.04%)	8.55×10^{-5}	5.04×10^{-5}	ko04626
Protein processing in endoplasmic reticulum	16 (11.51%)	4 (2.08%)	231 (5.69%)	186 (5%)	5.03×10^{-3}	9.89×10^{-1}	ko04141
Oxidative phosphorylation	13 (9.35%)	15 (7.81%)	171 (4.21%)	161 (4.33%)	5.24×10^{-3}	1.76×10^{-2}	ko00190
Starch and sucrose metabolism	13 (9.35%)	15 (7.81%)	173 (4.26%)	152 (4.08%)	5.77×10^{-3}	1.08×10^{-2}	ko00500
Plant hormone signal transduction	8 (5.76%)	15 (7.81%)	173 (4.26%)	152 (4.08%)	2.40×10^{-1}	1.08×10^{-2}	ko04075
Glycerophospholipid metabolism	5 (3.6%)	10 (5.21%)	84 (2.07%)	90 (2.42%)	1.59×10^{-1}	1.64×10^{-2}	ko00564

DEGs with pathway annotation: Figures in brackets represent the ratio of the number of DEGs that are annotated to this pathway and the number of DEGs that are annotated to all pathways. All genes with pathway annotation: Figures in brackets represent the ratio of all genes that are annotated to this pathway and all genes that are annotated to all pathways.

A



B

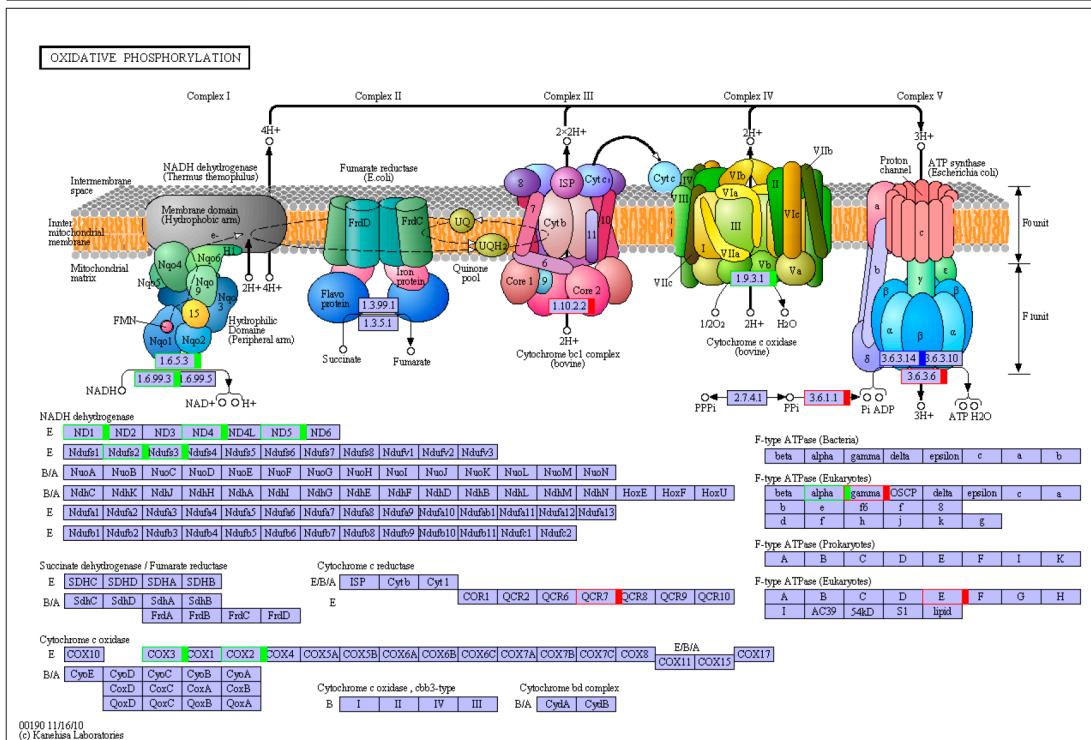


Figure 4. Welsh onion unigenes involved in oxidative phosphorylation pathways: (A) sequencing data from May 2013; and (B) sequencing data from May 2014. Blue boxes: enzymes specific to this species; Green boxes: down-regulated genes; Red boxes: up-regulated genes.

2.5. Analysis of CMS Related Genes in the Welsh Onion

This study utilized RNA-seq technology to explore CMS related genes between a CMS line and its maintainer line in the Welsh onion. All of the DEGs obtained from two years of RNA-seq data were investigated by BLAST in the NCBI database, and 559 unigenes were predicted to associate with CMS. Due to the DEGs that significantly enriched oxidative phosphorylation pathways in May 2013 and May 2014, the mitochondrial respiratory chain enzymes and enzyme complexes in oxidative phosphorylation pathways were important to CMS lines in other plants [22–24]. This study obtained six unigenes (IDs: c116086, c175619, c159049, c160965, c113452 and c50467) related to mitochondrial respiratory chain enzymes and enzyme complexes (Table 3), and they were significantly differentially expressed between the CMS line and the CMM line (Figure 5). These six unigenes were validated using RT-qPCR; the transcription expression of five unigenes was down-regulated in the cytoplasmic male maintainer line. Although most RT-qPCR results indicated smaller differences compared with the RNA-seq analysis, there was consistent expression. The unigenes functioned in “Energy production and conversion” (c175619, c159049) and “Posttranslational modification, protein turnover, and chaperones” (c173435) in the COG class annotation; they annotated to “Plasma membrane ATPase” (c116086), “Cytochrome c oxidase subunit 2” (c175619), “Cytochrome c oxidase subunit 3” (c159049), “Probable F-box protein” (c113452) and “Polygalacturonase inhibitor 1” (c50467) in the Swiss-Prot annotation; and, for the nt annotation, they were involved in the “*Cucumis sativus* plasma membrane ATPase 4-like” (LOC101221564) (c116086), “*Hibiscus cannabinus* cultivar P3B cytochrome c oxidase subunit II (*cox2*) gene” (c175619), “*Allium cepa* cultivar saski cytochrome oxidase subunit 3 (*cox3*) gene, complete cds; mitochondrial” (c159049), and “*Allium cepa* cultivar saski NADH dehydrogenase subunit 1 (*nad1*) gene, partial cds; and ATPase α subunit (*atpA*) gene, complete cds; mitochondrial” (c160965).

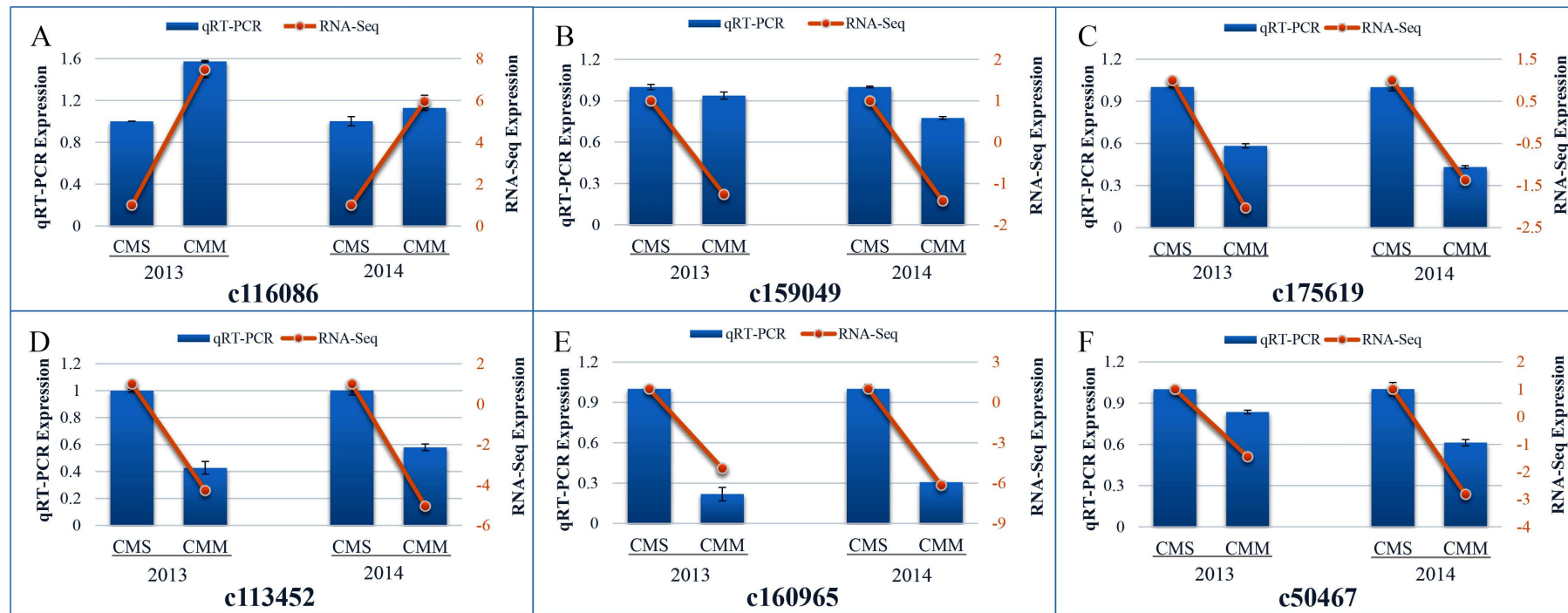


Figure 5. RT-qPCR and RNA-seq verification of differentially expressed candidate unigenes. CMS: Cytoplasmic male sterility; CMM: Cytoplasmic male maintainer. Data represent the average of three independent experiments. Error bars indicate the standard deviation (STDEV) for three biological replicates and three technical replicates in each experiment. (A–F) Different unigene IDs: (A) c116086; (B) c159049; (C) c175619; (D) c113452; (E) c160965; and (F) c50467.

Table 3. The unigenes associated with Cytoplasmic male sterility (CMS).

Unigene ID	COG Class Annotation	KEGG Annotation	Swissprot Annotation	nr Annotation	nt Annotation
c116086	–	–	Plasma membrane ATPase (Wheat)	plasma membrane H ⁺ -ATPase (<i>Cucumis sativus</i>)	PREDICTED: <i>Cucumis sativus</i> plasma membrane ATPase 4-like (LOC101221564), mRNA >gi 449510556 ref XM_004163650.1
c175619	Energy production and conversion	K02261 1e-149 pop:POPTR_936218 hypothetical protein	Cytochrome c oxidase subunit 2 (Mouse-ear cress)	cytochrome oxidase subunit 2 (<i>Boehmerianivea</i>)	<i>Hibiscus cannabinus</i> cultivar P3B cytochrome c oxidase subunit II (<i>cox2</i>) gene, complete cds; mitochondrial
c159049	Energy production and conversion	K02262 0.0 ath:AT2G07687 cytochrome c oxidase subunit 3 (EC:1.9.3.1)	Cytochrome c oxidase subunit 3 (Wheat)	cytochrome oxidase subunit 3 (<i>Cliviaminiata</i>)	<i>Allium cepa</i> cultivar saski cytochrome oxidase subunit 2 (<i>cox2</i>) gene, exon 2 and partial cds; and cytochrome oxidase subunit 3 (<i>cox3</i>) gene, complete cds; mitochondrial
c160965	–	–	–	–	<i>Allium cepa</i> cultivar saski NADH dehydrogenase subunit 1 (<i>nad1</i>) gene, partial cds; and ATPase alpha subunit (<i>atpA</i>) gene, complete cds; mitochondrial
c113452	–	–	Probable F-box protein (Mouse-ear cress)	hypothetical protein MTR_139s0011 (<i>Medicago truncatula</i>)	–
c50467	–	–	Polygalacturonase inhibitor 1 (Mouse-ear cress)	PREDICTED: polygalacturonase inhibitor-like (<i>Setaria italica</i>)	–

–: No annotation information.

3. Discussion

3.1. The Welsh Onion CMS Related Genes Were Enriched in the Oxidative Phosphorylation Pathway of the Mitochondrial Respiratory Chain

In this study, we profiled the Welsh onion transcriptome using the Illumina HiSeq 2000 platform, and classified the functions of the unigenes using the COG and GO classification and pathway enrichment analysis. A total of 1504 unigenes (in May 2013) and 2928 unigenes (in May 2014) were differentially expressed between the CMS line and its maintainer line in the Welsh onion.

In this study, we found many DEGs enriched in oxidative phosphorylation of the mitochondrial respiratory chain. The GO terms that identified the DEGs were related to CMS, i.e., “mitochondrion”, “mitochondrial respiratory chain”, “phosphorylation”, and “oxidation-reduction process”, and the cluster frequency of these DEGs was >5% (cluster frequency is the ratio of the number of DEGs that are annotated to this GO term to the number of DEGs that are annotated to all GO terms). The DEGs from the pathway analysis also showed many enzymes associated with the respiratory chain, and the regulation of numerous enzymes enriched in the “Oxidative phosphorylation” pathway, which is consistent in the RNA-seq data from both years.

In electron transfer system of mitochondrial respiratory chain, the respiratory chain enzymes and enzyme complexes (NADH dehydrogenase, cytochrome reductase, and cytochrome oxidase) play important roles in plants, their major function is to link electron transfer with proton translocation out of the mitochondrion, generating a transmembrane proton motive force that subsequently drives ATP synthesis by H⁺-ATPase [25]. In addition, cytochrome oxidase is the marker enzyme of the mitochondrial inner membrane with strong activity. Numerous studies have shown that cytochrome oxidase is relevant to CMS in plants [22,23]. In the pepper CMS line, a novel chimeric gene *orf 456*, was found to be inserted into the 3'-end of the *coxII* gene, and it may change the sterile characterization in the CMS line [22]. Luo et al. demonstrated that a new mitochondrial gene *WA352* encodes a protein that interacts with the nuclear-encoded mitochondrial protein COXII, thus causing premature tapetal programmed cell death and consequent male sterility [23]. In some studies, the CMS related candidate genes had close relationships with ATPase genes [2,11]. The F-type ATPase was proved to be embedded in the inner membranes of mitochondria, and this enzyme is a coupling factor in oxidative phosphorylation or photophosphorylation and plays an important role in energy conversion [24]. In this study, several key genes in the oxidative phosphorylation pathway were observed to be down-regulated by the RNA-seq approach; these gene-related enzymes were “NADH: ubiquinone reductase (H⁺-translocating)”, “ND5” (NADH dehydrogenase 5), “COXII” (cytochrome-c oxidase II), and “F-type ATPase α ” (Eukaryotes). Two unigenes, c175619 and c159049, were annotated in “Energy production and conversion (COG)”; the unigene c116086 was identified as a plasma membrane ATPase (Swiss-Prot annotation), and the unigene c160965 was an ATPase α subunit (*atpA*) gene (nr annotation). In our previous study, the analysis of Two-Dimensional Electrophoresis results showed the ATPase α subunit was specifically expressed in the CMS line compared with the cytoplasmic male maintainer line (unpublished data). Based on the above research, we conjectured that CMS was closely related to the respiratory chain enzymes and enzyme complexes.

3.2. The Nuclear Cellular Component Function Genes Are Important to CMS in the Welsh Onion

In addition, more DEGs genes were obtained in the nucleus than in the mitochondrion: their cellular component functions (GO) were clustered in “nucleus” (Cluster frequency 25.42% in May 2013 and 13.43% in May 2014), “cytoplasmic membrane-bounded vesicle” (Cluster frequency 13.84% in May 2013 and 16.24% in May 2014), and “membrane” (Cluster frequency 15.68% in May 2013 and 12.70% in May 2014). In the COG functional classification, the “replication, recombination and repair” (Cluster frequency 10.22% in May 2013 and 13.27% in May 2014) and “transcription” (Cluster frequency 10.02% in May 2013 and 11.14% in May 2014) were the largest groups, and the processes of gene transcription mainly occur in the nucleus. Some research has reported that the sterility is

caused by chimeric mitochondrial genes regulated by nuclear genes [26–29]. Thus, the nucleus may play an important role in the mechanisms of cytoplasmic male sterility. The results from Liu et al. suggest the CMS is the result of the nuclear–mitochondrial interaction [30]. In addition, Hong et al. suggested that nuclear–mitochondrial interaction may regulate pollen development and that the failure of sporogenous cell differentiation leads to sterility [31]. Therefore, we hypothesized that the Welsh onion male sterility regulatory genes and pathways are not only implicated in the mitochondria, but are also implicated in nuclear–mitochondria interaction. In addition, the identification of genes associated with male sterility in the nucleus remains to be further validated.

4. Methods and Methods

4.1. Sample Preparation

Two Welsh onion varieties (termed 64) were used in this study: 64-1, a CMS mutant in natural populations, and 64-2, a cytoplasmic male maintainer line, which when bred with 64-1 can produce fertile offspring (Figure 6). Both lines were bred at the Beijing Vegetable Research Center and the Beijing Academy of Agriculture and Forestry Sciences. For each variety, RNA was isolated from three inflorescences in the flowering stage of the Welsh onion in May 2013. We then collected experimental materials the same way in May 2014 as a repeat experiment. All samples were ground immediately after harvest in a mortar and pestle using liquid nitrogen. Total RNA was extracted using RNAiso for polysaccharide-rich Plant Tissue Kit (TaKaRa Biotechnology, Dalian, Liaoning Province, China), and RNA integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA-seq was performed at Beijing BioMarker Technologies (Beijing, China). The paired-end library preparation, cDNA library and sequencing were performed following standard Illumina methods on the Illumina sequencing platform (HiSeq™ 2000). The read length of sequencing in both years was 2×103 bp.

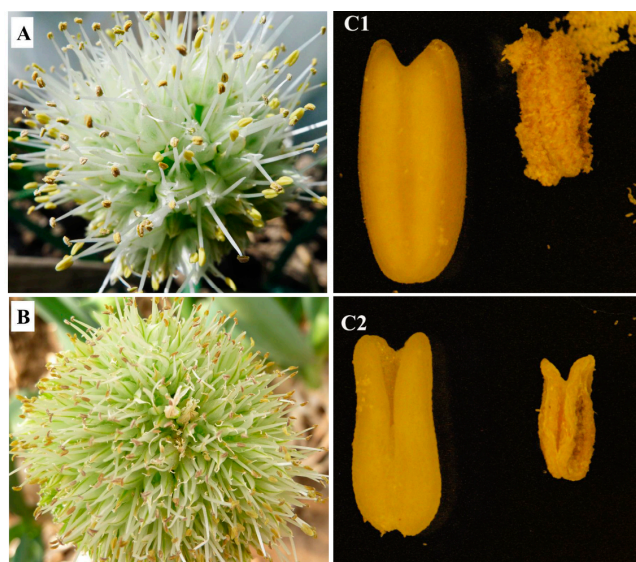


Figure 6. Comparison of cytoplasmic male maintainer and sterile varieties: (A) 64-1, cytoplasmic male maintainer; and (B) 64-2, cytoplasmic male sterility; (C1) 64-1 cytoplasmic male maintainer anther; and (C2) 64-2 cytoplasmic male sterility anther.

All of the datasets from the Illumina sequencing platform can be found in the Short Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) under accession number SRP071555.

4.2. Sequence Data Analysis and Assembly

The Bcl2fastq software was used for pre-processing. The raw reads were cleaned by removing the adaptor sequences, duplication sequences, low-quality reads (reads with ambiguous bases “N”), and reads with more than 50% of the bases with a Q -value ≤ 5 [32] to obtain high-quality clean read data. The Trinity method [33] was used for de novo assembly. First, contigs were obtained by extension based on the overlap between sequences. Next, the resultant contigs were joined into transcripts with the paired-end information, and the following parameters were used: seqTypefq, min kmer cov 2, min contig length 200, group pairs distance 500 and other default parameters. The longest transcript was taken as the sample unigene, and the unigenes were combined to produce the final assembly used for annotation. Getorf [34] was used to predict the open reading frame (ORF) of the unigenes, and the longest ORF was extracted for each unigene. The Bowtie [35] was used to compare between unigene database and each sequencing read sample, and default parameters of the software were used.

4.3. Functional Annotation and Differential Gene Expression Analysis

The unigenes were aligned with the NCBI’s nucleotide (nt) databases and non-redundant (nr) protein, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database, the Swiss-Prot protein database, and the Cluster of Orthologous Groups (COG) database. The Blast2GO program was used to assign GO terms [36]. The BLAST [37] program was used to performed COG and KEGG pathway annotation. The above searches were performed with a cut-off e -value of 10^{-5} . The unigenes were aligned to the COG and KEGG databases to predict and classify the possible functions and metabolic pathways of the unigenes. For differential gene expression analysis, Differentially expressed genes were identified using DESeq software [38,39] in each comparison, and the results were analyzed as described previously [1].

4.4. Quantitative RT-PCR (RT-qPCR) Analysis

The cDNA was synthesized from the flowering stage of the Welsh onion with total RNA in a 10- μ L reaction system using PrimeScript II RTase (TaKaRa Biotechnology, Dalian, Liaoning Province, China). The reverse transcription reaction mixture contained 5 μ L of total RNA (0.8 μ g), 3 μ L of diethylpyrocarbonate (DEPC) water, 1 μ L of oligo dT (50 μ M) and 1 μ L of dNTP (10 mM), and the reaction was performed following the kit protocol. Three biological replicates and three technical replicates for each experiment were performed. The qPCR reactions were performed in 96-well plates on the LightCycler 480 instrument (Roche Diagnostic, Mannheim, Germany) using SYBR Green I (TaKaRa Biotechnology, Dalian, Liaoning Province, China). The qPCR primers, which were designed using Primer3 (<http://frodo.wi.mit.edu/primer3/>), are shown in Table S3. Each gene was analyzed using the method described previously by Qianchun Liu et al. [1].

5. Conclusions

This study investigated the important CMS-related genes and pathways in the Welsh onion by transcriptome analysis. Several key genes such as F-type ATPase, NADH dehydrogenase, and cytochrome *c* oxidase were observed by RNA-seq and validated by RT-qPCR. These genes were not only closely related to mitochondria but also to the nuclear–mitochondria interaction. Further functional characterization is in progress to determine the specific functions of these genes in relation to CMS in *Allium fistulosum* L. This study will accelerate the research on CMS gene clones and other functional genomics research on *A. fistulosum*.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/17/7/1058/s1>.

Acknowledgments: This work was supported by grants from the National Natural Science Foundation of China (No. 31372066), The Natural Science Foundation of Beijing (No. 6132016), the Ministry of agriculture “948” project

(No. 2011-Z61), and the Science and Technology Innovation Capability Construction of the Beijing Academy of Agriculture and Forestry Sciences (No. KJCX20140111).

Author Contributions: Conceived and designed the experiments: Yongqin Wang and Changlong Wen; Performed the experiments: Qianchun Liu; Analyzed the data: Qianchun Liu and Yanping Lan; Contributed reagents/materials/analysis tools: Hong Zhao and Jian Wang; Wrote the paper: Qianchun Liu.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, Q.C.; Wen, C.L.; Zhao, H.; Zhang, L.Y.; Wang, J.; Wang, Y.Q. RNA-Seq reveals leaf cuticular wax-related genes in welsh onion. *PLoS ONE* **2014**, *9*, e113290. [[CrossRef](#)] [[PubMed](#)]
2. Wang, X.J. *Vegetable Breeding*; China Agriculture Press: Beijing, China, 1995; pp. 214–216.
3. Laser, K.D.; Lersten, N.R. Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. *Bot. Rev.* **1972**, *38*, 425–454. [[CrossRef](#)]
4. Ogura, H. Studies on the new male-sterility in Japanese radish, with special reference to the utilization of this sterility towards the practical raising of hybrid seeds. *Mem. Fac. Agric. Kagoshima Univ.* **1968**, *6*, 39–78.
5. Fu, T.D.; Yang, X.N.; Yang, G.S. Development and studies on polima cytoplasmic male sterile “three lines” in *Brassica napus* L. *J. Huazhong Agric. Univ.* **1989**, *8*, 201–207.
6. Hinata, K.; Konno, N. Studies on a male sterile strain having the *Brassica campestris* nucleus and the *Diplotaxismuralis* cytoplasm. *Jpn. J. Breed.* **1979**, *29*, 305–311. [[CrossRef](#)]
7. Wang, D.J.; Guo, A.G.; Li, D.R.; Tian, J.H.; Huang, F.; Sun, G.L. Cytological and molecular characterization of a novel monogenic dominant GMS in *Brassica napus* L. *Plant Cell Rep.* **2007**, *26*, 571–579. [[CrossRef](#)] [[PubMed](#)]
8. Sanders, P.M.; Bui, A.Q.; Weterings, K.; McIntire, K.N.; Hsu, Y.C.; Lee, P.Y.; Truong, M.T.; Beals, T.P.; Goldberg, R.B. Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. *Sex. Plant. Reprod.* **1999**, *11*, 297–322. [[CrossRef](#)]
9. Hu, Q.; Hansen, L.; Laursen, J.; Dixelius, C.; Andersen, S. Intergeneric hybrids between *Brassica napus* and *Orychophragmus violaceus* containing traits of agronomic importance for oilseed rape breeding. *Theory Appl. Genet.* **2002**, *105*, 834–840.
10. Zubko, M.K.; Zubko, E.I.; Adler, K.; Grimm, B.; Gleba, Y.Y. New CMS-associated phenotypes in cybrids *Nicotiana tabacum* L. (+*Hyoscyamus niger* L.). *Ann. Bot.* **2003**, *92*, 281–288. [[CrossRef](#)] [[PubMed](#)]
11. Nishimura, J.; Shino, M. Studies on the male sterility in Welsh onion. *Jpn. Soc. Hortic. Sci.* **1972**, *9*, 180–181.
12. Zhang, Q. Welsh Onion male sterility in natural populations. *J. Shandong Agric. Univ.* **1987**, *18*, 1–10.
13. Yamashita, K.; Tashiro, Y. Seed productivity test of CMS lines of Japanese bunching onion (*Allium fistulosum* L.) possessing the cytoplasm of a wild species, *A. galanthum* Kar. et Kir. *Euphytica* **2004**, *136*, 327–331. [[CrossRef](#)]
14. Peffley, E.B.; Hou, A. Bulb-type onion introgressants possessing *Allium fistulosum* L. genes recovered from interspecific hybrid backcrosses between *A. cepa* L. and *A. fistulosum* L. *Theory Appl. Genet.* **2000**, *100*, 528–534.
15. Yamashita, K.; Takatori, Y.; Tashiro, Y. Chromosomal location of a pollen fertility-restoring gene, *Rf*, for CMS in Japanese bunching onion (*Allium fistulosum* L.) possessing the cytoplasm of *A. galanthum* Kar. et Kir. revealed by genomic in situ hybridization. *Theory Appl. Genet.* **2005**, *111*, 15–22.
16. Li, H.Y.; Wang, C.; Li, L.L.; Wang, Y.Q.; Zhao, R. Cloning and Expression Analysis of A Putative B Class MADS-box Gene of AcPI in Onion. *Sci. Agric. Sin.* **2012**, *45*, 4759–4769.
17. Engelke, T.; Terefe, D.; Tatlioglu, T. A PCR-based marker system monitoring CMS-(S), CMS-(T) and (N)-cytoplasm in the onion (*Allium cepa* L.). *Theory Appl. Genet.* **2003**, *107*, 162–167.
18. Havey, M.J. A putative donor of S-cytoplasm and its distribution among open-pollinated populations of onion. *Theory Appl. Genet.* **1993**, *86*, 128–134. [[CrossRef](#)] [[PubMed](#)]
19. Hanson, M.R. Plant mitochondrial mutations and male sterility. *Annu. Rev. Genet.* **1991**, *25*, 461–486. [[CrossRef](#)] [[PubMed](#)]
20. Chase, C.D. Cytoplasmic male sterility: A window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.* **2007**, *23*, 81–90. [[CrossRef](#)] [[PubMed](#)]
21. Jing, B.; Heng, S.P.; Tong, D.; Wan, Z.J.; Fu, T.D.; Tu, J.X.; Ma, C.Z.; Yi, B.; Wen, J.; Shen, J.X. A male sterility-associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. *J. Exp. Bot.* **2012**, *63*, 1285–1295. [[CrossRef](#)] [[PubMed](#)]

22. Kim, D.H.; Kang, J.G.; Kim, B.D. Isolation and characterization of the cytoplasmic male sterility-associated *orf456* gene of chili pepper (*Capsicum annuum* L.). *Plant Mol. Biol.* **2007**, *63*, 519–532. [[CrossRef](#)] [[PubMed](#)]
23. Luo, D.P.; Xu, H.; Liu, Z.L.; Guo, J.X.; Li, H.Y.; Chen, L.T.; Fang, C.; Zhang, Q.Y.; Bai, M.; Yao, N.; et al. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat. Genet.* **2013**, *45*, 573–577. [[CrossRef](#)] [[PubMed](#)]
24. Bowler, M.W.; Montgomery, M.G.; Leslie, A.G.W.; Walker, J.E. Howazide inhibits ATP hydrolysis by the F-ATPases. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8646–8649. [[CrossRef](#)] [[PubMed](#)]
25. Weiss, H.; Friedrich, T.; Hofhaus, G.; Preis, D. The respiratory-chain NADH dehydrogenase (complex I) of mitochondria. *Eur. J. Biochem.* **1991**, *197*, 563–576. [[CrossRef](#)] [[PubMed](#)]
26. L'Homme, Y.; Stahl, R.J.; Li, X.Q.; Hameed, A.; Brown, G.G. Brassica nap cytoplasmic male sterility is associated with expression of a mtDNA region containing a chimeric gene similar to the pol CMS-associated *orf224* gene. *Curr. Genet.* **1997**, *31*, 325–335. [[CrossRef](#)] [[PubMed](#)]
27. Yuan, M.; Yang, G.S.; Fu, T.D.; Li, Y. Transcriptional control of *orf224/atp6* by the pol CMS restorer *Rfp* gene in *Brassica napus* L. *J. Genet. Genom.* **2003**, *30*, 469–473.
28. Singh, M.; Hamel, N.; Menassa, R.; Li, X.Q.; Young, B.; Jean, M.; Landry, B.S.; Brown, G.G. Nuclear genes associated with a single Brassica CMS restorer locus influence transcripts of three different mitochondrial gene regions. *Genetics* **1996**, *143*, 505–516. [[PubMed](#)]
29. Singh, M.; Brown, G.G. Suppression of cytoplasmic male sterility by nuclear genes alters expression of a novel mitochondrial gene region. *Plant Cell* **1991**, *3*, 1349–1362. [[CrossRef](#)] [[PubMed](#)]
30. Liu, C.; Ma, N.; Wang, P.Y.; Fu, N.; Shen, H.L. Transcriptome sequencing and de novo analysis of a cytoplasmic male sterile line and its near-isogenic restorer line in chili pepper (*Capsicum annuum* L.). *PLoS ONE* **2013**, *8*, e65209. [[CrossRef](#)] [[PubMed](#)]
31. An, H.; Yang, Z.; Yi, B.; Wen, J.; Shen, J.; Tu, J.; Ma, C.; Fu, T. Comparative transcript profiling of the fertile and sterile flower buds of *pol* CMS in *B. napus*. *BMC Genom.* **2014**, *15*, 162–182. [[CrossRef](#)] [[PubMed](#)]
32. Shi, C.Y.; Yang, H.; Wei, C.L.; Yu, O.; Zhang, Z.Z.; Jiang, C.J.; Sun, J.; Li, Y.Y.; Chen, Q.; Xia, T.; et al. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genom.* **2011**, *12*. [[CrossRef](#)] [[PubMed](#)]
33. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.D.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
34. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **2000**, *16*, 276–277. [[CrossRef](#)]
35. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*. [[CrossRef](#)] [[PubMed](#)]
36. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [[CrossRef](#)] [[PubMed](#)]
37. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.H.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)] [[PubMed](#)]
38. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, 1–12. [[CrossRef](#)] [[PubMed](#)]
39. Wang, L.; Feng, Z.; Wang, X.; Wang, X.; Zhang, X. DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **2010**, *26*, 136–138. [[CrossRef](#)] [[PubMed](#)]

