

*Full Length Research Paper*

# Characterization and comparison of key genes involved with flowering time regulation from *Arabidopsis thaliana*, *Oryza sativa* and *Zea mays*

Bi Ma<sup>1</sup>, Lijuan Wei<sup>2</sup>, Annaliese S. Mason<sup>3</sup>, Zeshan An<sup>2</sup>, Meili Xiao<sup>2</sup>, S. S. Banga<sup>4</sup>, Junjun Zhang<sup>5</sup>, Qinghong Zhou<sup>5</sup> and Donghui Fu<sup>5\*</sup>

<sup>1</sup>State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, China.

<sup>2</sup>Engineering Research Center of South Upland Agriculture of Ministry of Education,

College of Agronomy and Biotechnology, Southwest University, Chongqing, P. R. 400715, China.

<sup>3</sup>Centre for Integrative Legume Research and School of Agriculture and Food Sciences, The University of Queensland, Brisbane 4072, Australia.

<sup>4</sup>Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana 141001, India.

<sup>5</sup>Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Jiangxi Agricultural University, Nanchang 330045, China.

Accepted 29 August, 2012

Flowering is an important phase in the angiosperms life cycle, as it signifies the transition from vegetative to reproductive phase. Recently, increasing number of flowering genes are being cloned and analyzed. However, few studies have focused on the comparative analysis of key candidate genes involved in the regulation of flowering time across species. We explore the common characteristics of key flowering genes across species and compare their diversity and evolutionary divergence. We analyze key flowering candidate genes in three species, *Arabidopsis thaliana*, *Oryza sativa* and *Zea mays*. The homology of flowering genes across the three species was more than 60%. Nucleotide polymorphism analysis for these genes revealed high sequence conservation during evolution, although some small insertion/deletions (InDels) were also detected. Furthermore, these candidate genes harbored simple repeat sequence repeats that perhaps drove genetic variation and divergence of these genes. The flowering genes were expressed preferentially in the organelle and cellular components and participated mainly in metabolic processes with binding and catalytic functionality. These results suggested that key flowering genes have more common characteristics that may allow the identification and analysis of other flowering genes, especially in species where genome sequencing has not yet been performed.

**Key words:** Evolution, flowering gene, simple repeat sequences, homologous genes.

## INTRODUCTION

The onset of flowering is one of the most important, complex stages in the life cycle of angiosperms and it marks the transition from vegetative to reproductive phase. Flowering time affects sexual reproduction of plants, seed and fruit development and consequently plant productivity. For example, the flowering locus CA (FCA) gene not only accelerates the flowering but also

affects root development (Macknight et al., 2002). Similarly, days to heading on chromosome 8 (DTH8) regulates both flowering time in the photoperiod pathway as well as plant height and yield potential (Wei et al., 2010). Appropriate flowering time not only facilitates niche farming but also helps to realize the genetic potential of crops. Increasing numbers of flowering genes

have now been cloned, especially in *Arabidopsis thaliana* and rice. Furthermore, sketch maps of the regulation network of key flowering genes are now available in *A. thaliana* (Wellmer and Riechmann, 2010; Jung and Muller, 2009), rice (Komiya et al., 2009). However, the common characteristics, divergence and evolutionary differences between key flowering candidate genes remain unknown.

In *A. thaliana*, four pathways simultaneously control flowering. These include, the photoperiod pathway, the vernalization pathway, the autonomous pathway, and the gibberellin pathway (Jung and Muller, 2009; He and Amasino, 2005; Boss et al., 2004). In fact, these four pathways mutually interact to regulate flowering with some genes acting as bridges. For example, the autonomous and vernalization pathways are integrated by *FLC* gene (Kim et al., 2009; Amasino, 2010). The *SOC1* (Lee et al., 2000), and *FT* (Halliday et al., 2003) genes are also regarded as integrators between the temperature- and light-signaling pathways, with only scattered studies on other pathways (Komiya et al., 2009). There is now significant interest in the core genes or integrator genes in these pathways because the loss or alteration of these key genes can greatly influence flowering time.

The few comparisons of homology between flowering genes have shown that the flowering genes are highly conserved across plant species. For example, the *Hd3a* gene in rice is highly homologous to the *FT* gene in *A. thaliana* (Kojima et al., 2002). *Hd1* is also closely related to the *A. thaliana* flowering time gene *CO* (Yano et al., 2000). These results suggested that the GI-CO-FT pathway in *A. thaliana* is conserved in rice (*OsGI-Hd1-Hd3a*). Taylor et al. (2010) also found that the flowering genes, *AcGI* and *AcFKF1* in onion are homologous to genes involved in photoperiod regulation of *A. thaliana*. These results suggest that extensive homology of flowering genes exists across species, but degree of key genes across diverse species is still unknown.

In the present study, a comprehensive comparative analysis of key flowering genes was performed in the three typical plant species *A. thaliana*, *Oryza sativa*, and *Zea mays*. The objective of this comparison was to reveal the extent of sequence conservation, parallel nucleotide

divergence, and functionality of key flowering genes between these species. The results are expected to aid in molecular investigations into trait variation, evolutionary patterns and regulatory mechanisms of flowering genes in other less-investigated plant species.

## MATERIALS AND METHODS

### Mining sequences of key flowering genes

Two approaches were followed to collect information on the flowering genes from *A. thaliana*, *O. sativa*, and *Z. mays*. Initially, the first batch of flowering genes of each species was retrieved and downloaded from National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>) by the key word "flowering" and the organism name. Three basic screening criteria were used: (1) Selection of full-length cDNAs of flowering genes rather than truncated genes; (2) removal of duplicated genes derived from different databases in the NCBI website to retain only one set of genes, and (3) involvement of selected genes in flowering time regulation rather than in floral organogenesis. Finally to mine more putative flowering genes, the first batch of downloaded flowering time genes (full-length cDNA sequences) were queried in each of the three species against the remaining two species using BLASTN with  $e < 10^{-30}$  and  $> 75\%$  identity (alignment length \*100 / query length). Both the hits and the first batch of original flowering genes were used for further investigation. BioEdit (Hall, 1999) was used to analyze the length of flowering gene sequences.

### Comparative homologies of flowering genes

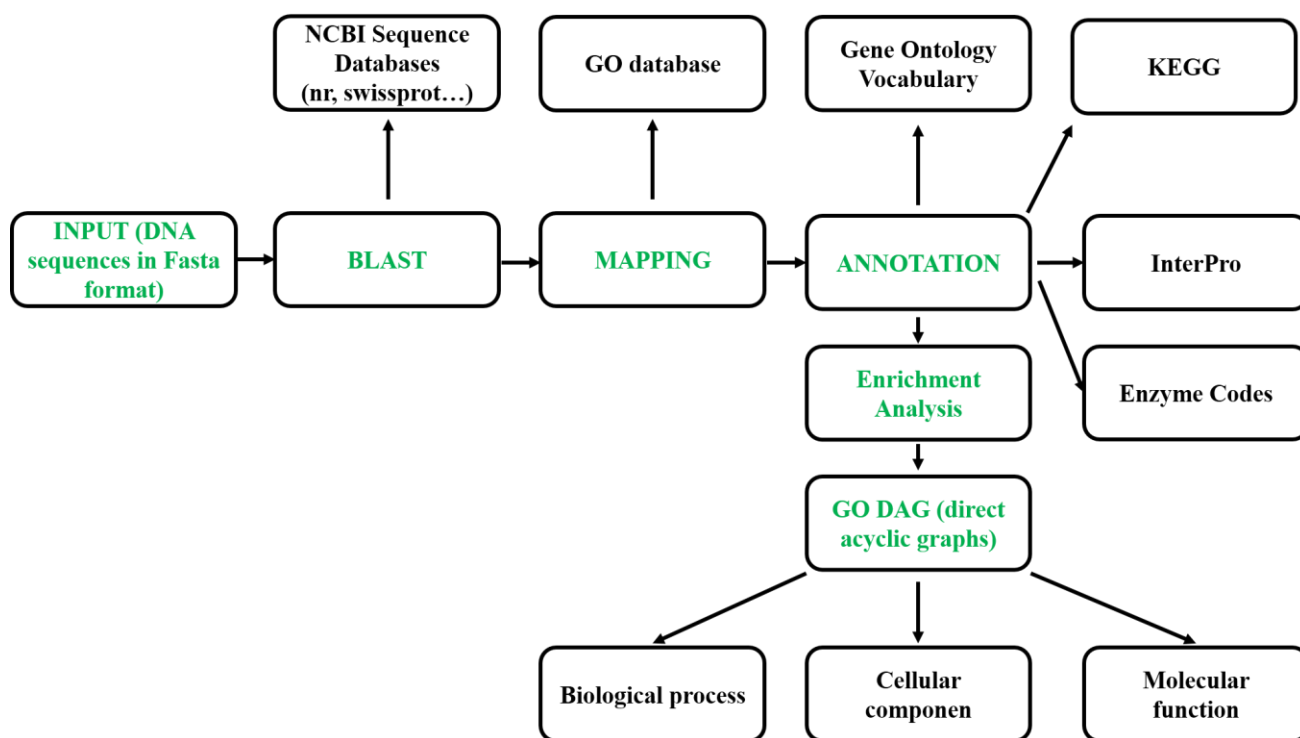
To compare the homology of flowering genes from *A. thaliana*, *O. sativa*, and *Z. mays*, flowering genes from each species were used as queries to perform BLASTP (<http://www.ncbi.nlm.nih.gov/>) queries against flowering genes of the remaining two species. If two genes from different species showed more than 75% (alignment length \*100/query length) identity with  $e < 10^{-20}$ , they were considered homologous genes. An open source platform for complex network analysis and visualization, Cytoscape (<http://www.cytoscape.org/>) (Angelovici et al., 2009) was used to produce an intuitive graphical representation of the homology of the flowering genes.

### Nucleotide polymorphism analysis of key flowering genes

Key flowering genes are generally located in the central node positions of regulatory networks. TBLASTN (<http://www.ncbi.nlm.nih.gov/>) analysis was conducted against gene databases of other species except for *A. thaliana*, *O. sativa* and *Z. mays* to acquire homologies of other species at 75% (alignment length\*100/query length) identity and  $e < 10^{-20}$ . Each flowering gene and flowering gene homologue was used to perform DNA sequence polymorphism analysis (DnaSP, <http://www.ub.edu/dnasp/>, version 5.10.01) (Rozas and Rozas, 1999). The characteristics of homologue divergence were calculated by nucleotide variation patterns, including  $\pi$  value (a parameter which is used to illustrate insertion/deletion (InDel) diversity per site and the nucleotide diversity) (Nei, 1987; Tajima, 1983), the total number of mutations ( $\theta$ ) (Watterson, 1975), Tajima's D value, which is used to determine whether sequence divergence is consistent with the neutral evolution model (Tajima, 1989) and InDel information (Jander et al., 2002). All parameters were set to default commands in accordance

\*Corresponding author. E-mail: fudhui@163.com. Tel: +086-0791-83813142. Fax: +086-0791-83813185.

**Abbreviations:** **FCA**, Flowering locus CA; **DTH8**, days to heading on chromosome 8; **CO**, constans; **FLC**, flowering locus C; **SOC1**, suppressor of overexpression of CO 1; **FT**, flowering locus T; **Ehd1**, early heading date 1; **GI**, gigantea; **Hd3a**, heading date 3a; **FKF1**, flavin-binding kelch repeat; **DnaSP**, DNA sequence polymorphism analysis; **InDel**, insertion/deletion; **GO DAG**, go directed acyclic graph.



**Figure 1.** Schematic representation of Blast2GO annotation. GO annotations are generated through a five-step process: Blast, mapping, annotation, enrichment analysis (optional) and visualization.

with the standard procedures for DnaSP.

#### Detection of simple repeat sequences in flowering genes

To judge whether flowering genes harbored abundant simple repeat sequences and to analyze the possible relationship between variations of flowering genes and simple sequence repeats (SSRs) Locator (Da Maia et al., 2008) was used to detect SSR loci in flowering genes of the three species with the parameters set at default values. The minimum numbers of repeats used to identify presence of SSR loci were 20, 10, 7, 4, 5, 5, 3, 3, 3, and 3 for monomers, dimers, trimers, tetramers, pentamers, hexamers, heptamers, octamers, nonamers, and decamers, respectively. A space of less than 100 and more than 5 between imperfect SSRs was permitted. The number and frequency of SSRs in all known genes of the three species were also analyzed for comparison.

#### Functional annotations of flowering genes

To determine whether flowering genes from each species showed similar spatiotemporal expression (Conesa and Gotz, 2008; Conesa et al., 2005), functional annotations were performed by Blast2GO (<http://www.blast2go.org/>) to categorize flowering time genes by cellular component, biological process, and molecular function. Blast2GO is a popular tool to annotate, visualize and analyze DNA sequences. Blast2GO analysis comprises five steps: (1) Blasting: a set of DNA sequences are blasted against the NCBI database, (2) mapping: the blast results are matched with GO terms (cellular component, biological process, and molecular function) using annotation files provided by the GO Consortium, (3) annotation:

certain annotation rules are used to annotate sequences, (4) enrichment analysis (optional): statistical analysis of different annotation results and (5) visualization: display of annotation and statistics results by the GO DAG (GO directed acyclic graph). Figure 1 shows the schematic diagram of Blast2GO annotation for the flowering time genes.

## RESULTS

### Mining flowering genes of *A. thaliana*, *O. sativa*, and *Z. mays*

The entire set of flowering genes from each species was obtained using two approaches. Initially, 40 genes in *A. thaliana*, 101 genes in *O. sativa*, and 84 genes in *Z. mays* were collected by searching with the keyword “flowering” and the species name in the NCBI website. Finally, 1 gene in *A. thaliana*, 9 genes in *O. sativa*, and 15 genes in *Z. mays* were added to the first batch of flowering genes of the three species by BLASTN analysis. Forty-one flowering genes were thus collected in *A. thaliana*, 110 in *O. sativa*, and 99 in *Z. mays*. Average lengths of cDNA sequences were 2206.0, 1896.3, and 1454.2 bp in *A. thaliana*, *O. sativa*, and *Z. mays*, respectively. The average protein lengths were 586.5, 448.5, 337.3 amino acids (aa) in *A. thaliana*, *O. sativa*, and *Z. mays*, respectively (Table 1).

**Table 1.** The basic statistics of flowering-related genes of *A. thaliana*, *O. sativa*, and *Z. mays*.

Species	<i>A. thaliana</i>	<i>O. sativa</i>	<i>Z. mays</i>
NCBI <sup>a</sup>	40	101	84
BLASTN <sup>b</sup>	1	9	15
The total number of mRNA	41	110	99
The length of mRNA length (bp)	2206.0 ± 916.4	1896.3 ± 1112.8	1454.2 ± 1194.2
The number of protein	40	98	22
The length of protein length (aa)	586.5 ± 307.9	448.5 ± 301.2	337.3 ± 196.9

<sup>a</sup>The number of flowering genes from NCBI found by the species name and "flowering". <sup>b</sup>The number of flowering genes obtained by BLASTN analysis.

### Comparative homologies of flowering time regulatory genes

Homologous relationships between the three species are presented in Figure 2. These comprised a total of 130 nodes and 627 edges, comparing 27 (62.8%), 87 (81.3%), and 16 (16%) nodes in *A. thaliana*, *O. sativa*, and *Z. mays*, respectively. Twenty-two of these (16.9%) node genes (marked in yellow) were inferred to be critical flowering genes due to the high degree of homology in these genes between the three species. These genes are likely to become principal targets of research in species for which whole genome sequence is not available.

The proportion of homologous flowering genes between any two of the three species is depicted in Figure 3. *A. thaliana* flowering genes had 77.5% homologous identity with *O. sativa*, and 80.0% homologous identity with *Z. mays* (average 78.8%). *O. sativa* flowering genes shared 50% identity with *A. thaliana* and 60% with *Z. mays* (average 55%). *Z. mays* flowering genes shared 35.7% identity with *A. thaliana* and 64% identity with *O. sativa* (average 49.9%). Thus, the average homology between the flowering networks of the three species was 61.2%, suggesting that the components of the regulatory networks of flowering genes were highly conserved during species evolution.

### Nucleotide polymorphism analysis of the key genes

In order to find out whether the key genes had homologous sequences in species other than the three target species, the BLASTP analysis with default parameters was performed in NCBI. Each key gene shared sequence homology with about 15 to 67 different species, with an average of 45 species (Figure 4). For example, homologies for *SOC1* (*A. thaliana*) were found for at least 51 species. We also analyzed the number of homologues that each key gene had within each species. On average, only one homologue of a key gene was found in 59.3% of species, two homologues in 22.6% of species and three or more homologues in 18.1% of species.

Each of the 22 key genes and their corresponding homologues were analyzed for divergence and evolution by DNAsp. The InDel information for nucleotide diversity and the Tajima's D information for the neutral evolution model are presented in Tables 2 and 3, respectively. The diversity index of these key flowering genes ranged from 0.15675 to 0.52519 with an average of 0.27983, exhibiting a high level of variation. The average InDel length event was 5.594 (range = 5.594 ± 5.450). The average InDel length was 7.323 (range = 7.323 ± 10.421). The statistics of the Tajima's D test (Table 3) showed that the key flowering genes were under negative selection pressure, although this effect was not significant for all key genes.

### Analysis of simple sequence repeats (SSRs) in flowering genes

To explore the basic principles underlying variation of flowering genes in nature, the number and frequency of the SSRs in flowering genes was analyzed. Sixty-one SSR loci were detected in 64 (25.6%) cDNAs of 250 flowering genes, including 4 (9.3%), 36 (33.6%), and 24 (24.0%) cDNAs in *A. thaliana*, *O. sativa*, and *Z. mays*, respectively (Figure 5). Thirty-three trinucleotide-motif SSR occupied the highest proportion (47.8%), followed by 25 (36.2%) mononucleotide-motif SSR, 7 (10.1%) pentanucleotide-motif SSRs, and 5.8% SSRs with other motifs. Trimernucleotide-motif SSRs were the main SSR type in *A. thaliana* and *O. sativa*. The predominant-motif SSR motifs in *Arabidopsis* were TCT (50%), CAA (25%), and CTT (25%) whereas the main motifs of SSRs in *O. sativa* were CGC/GCG (22.5%), CCG/CGG (17.5%), GGA (10.0%), GTG (10.0%), AAG (5.0%), and TCT (5.0%).

### Functional annotations

To assess whether the flowering genes could have similar expression across different species, Blast2Go analysis was performed to annotate the flowering genes.

**Table 2.** DNA polymorphism in 22 groups of key flowering related genes.

Protein accession	Average InDel length event	Average InDel length	Number of InDel Haplotypes	InDel Haplotype Diversity	InDel Diversity, k(i)	InDel Diversity per site, $\pi(i)$	Theta (per sequence) from I, Theta(i)-W	Statistical significance
AT1G65480.1	18.2	6.824	6	0.183	0.189	0.00037	1.029	*
AT1G69120.1	3.425	3.516	29	0.96	5.075	0.00775	9.013	Not significant
AT2G45660.1	2.455	3.053	13	0.807	1.28	0.00254	2.529	Not significant
AT4G18960.1	4.708	6.198	23	0.925	3.039	0.00451	5.059	Not significant
AT5G10140.1	10.4	4.318	6	0.778	1.156	0.00299	1.767	Not significant
AT5G61380.1	4.25	4.008	14	0.99	20.743	0.01702	25.834	Not significant
BAH30234.1	2.333	2.952	4	0.444	0.824	0.00189	0.693	Not significant
BAH30235.1	2.333	2.952	4	0.219	0.31	0.00063	0.693	Not significant
BAH30236.1	2	1.977	6	0.337	0.676	0.00155	1.387	Not significant
BAH30243.1	1.8	1.989	5	0.175	0.227	0.00055	1.149	*
BAH30244.1	2	1.235	5	0.216	0.396	0.00094	1.379	*
NP_001105151.1	3.389	2.408	17	0.883	3.225	0.00632	4.312	Not significant
NP_001105959.1	21.857	43.532	13	0.877	2.129	0.00279	3.45	Not significant
NP_001106241.1	6.304	10.679	14	0.892	3.152	0.0052	5.667	Not significant
NP_001106242.1	3	3.042	9	0.613	1.312	0.00266	1.712	Not significant
NP_001106243.1	8.938	15.411	8	0.589	1.462	0.00246	3.942	*
NP_001106244.1	11.545	31.243	10	0.849	2.276	0.0037	2.671	Not significant
NP_001106245.1	2.143	2.085	8	0.811	1.64	0.00321	1.725	Not significant
NP_001106247.1	2.625	2.435	7	0.602	1.173	0.00224	1.87	Not significant
NP_001106256.1	3	4.82	5	0.336	0.481	0.00095	1.358	Not significant,
NP_001120722.1	3.861	3.444	7	0.944	14.111	0.01338	13.246	Not significant
Q8VWH2.1	2.5	2.977	5	0.179	0.186	0.00042	0.924	*

<sup>a</sup> The key flowering genes were used to perform TBLASTN analysis to acquire homologous genes from other species. The hit and key flowering genes were used together to perform DNA polymorphism analysis. \*  $p < 0.05$ . Not significant,  $p > 0.10$ .

The annotation of cellular components of regulatory network of flowering genes allowed categorization into six subclasses (Figure 6a). The cellular-component subclass occupied 24.4, 43.5 and 30.0% of *A. thaliana*, *O. sativa*, and *Z. mays* flowering genes, respectively. Next most common was organelle localization, found for 26.7, 19.3 and 50.0% of *A. thaliana*, *O. sativa*, and *Z. mays* flowering genes, respectively. The intracellular localization subclass also constituted 22.8, 17.3 and 20.0%, of *A. thaliana*, *O. sativa*, and *Z. mays*

flowering genes, respectively. Apart from these three larger subclasses, other subclasses, such as the macromolecular complex, membrane, and cytoskeleton showed clear differences among the three species. For example, 23.2 and 9.8% of the flowering genes of *A. thaliana* and *O. sativa*, respectively, were expressed in the macromolecular complex, whereas 2.9 and 8.3% of *A. thaliana* and *O. sativa* functioned in membrane subclass.

For the annotation of biological processes,

flowering genes were divided into nine subclasses (Figure 6b). The largest subclass was metabolic process, containing 33.2, 57.9 and 80.0% of the flowering genes of *A. thaliana*, *O. sativa*, and *Z. mays*, respectively. The next largest subclass was biological processes, occupying 17.2, 15.0 and 16.0% of *A. thaliana*, *O. sativa*, and *Z. mays* genes, respectively. In *A. thaliana* and *O. sativa*, 24.9 and 6.5% of genes respectively fell into the subclass of cell division, whereas 13.0 and 6.5% of the flowering genes of *A. thaliana* and *O. sativa*

**Table 3.** Tajima's D information for neutral versus selective bias in DNA sequence evolution.

Protein accession <sup>a</sup>	Number of polymorphic (segregating) sites, S	Total number of mutations, $\theta$	Average number of nucleotide differences, k	Nucleotide diversity, $\pi$	Theta (per sequence) from $\theta$	Theta (per site) from $\theta$	Tajima's D	Statistical significance
AT1G65480.1	301	588	111.468	0.2629	120.9675	0.2853	-0.27487	Not significant,
AT1G69120.1	418	789	146.6436	0.2647	177.7842	0.32091	-0.64291	Not significant,
AT2G45660.1	382	765	157.111	0.32596	175.8621	0.36486	-0.39547	Not significant,
AT4G18960.1	409	770	148.7476	0.24958	162.314	0.27234	-0.29650	Not significant,
AT5G10140.1	158	193	52.51111	0.15675	68.22275	0.20365	-1.14875	Not significant,
AT5G61380.1	865	1924	511.5333	0.52519	591.7156	0.60751	-0.60435	Not significant,
BAH30234.1	281	558	112.5814	0.26182	128.9654	0.29992	-0.47185	Not significant,
BAH30235.1	316	619	128.7708	0.26551	143.0637	0.29498	-0.37127	Not significant,
BAH30236.1	288	551	109.7265	0.25818	127.3475	0.29964	-0.51389	Not significant,
BAH30243.1	282	542	108.1723	0.26842	124.5977	0.30918	-0.48814	Not significant,
BAH30244.1	284	549	106.445	0.26026	126.2069	0.30857	-0.57985	Not significant,
NP_001105151.1	344	657	131.9144	0.2783	157.3819	0.33203	-0.61339	Not significant,
NP_001105959.1	315	589	125.6421	0.26676	145.1277	0.30813	-0.51683	Not significant,
NP_001106241.1	396	728	135.1193	0.28933	179.3768	0.3841	-0.95059	Not significant,
NP_001106242.1	332	620	129.5116	0.27094	151.6338	0.31723	-0.55935	Not significant,
NP_001106243.1	394	714	136.2936	0.29185	175.9273	0.37672	-0.86791	Not significant,
NP_001106244.1	328	601	134.4941	0.27336	145.9372	0.29662	-0.29937	Not significant,
NP_001106245.1	326	610	136.0947	0.27438	150.302	0.30303	-0.36392	Not significant,
NP_001106247.1	357	692	136.0829	0.26683	161.7373	0.31713	-0.59333	Not significant,
NP_001106256.1	353	671	132.5301	0.27102	151.9238	0.31068	-0.46940	Not significant,
NP_001120722.1	599	919	312.5556	0.33681	338.134	0.36437	-0.39508	Not significant,
Q8VWH2.1	297	567	112.8062	0.26295	131.0455	0.30547	-0.51699	Not significant,

<sup>a</sup> The key flowering genes were used to perform TBLASTN analysis to acquire homologous genes from other species. The hit and key flowering genes were used together to perform Tajima's D test. Not significant,  $p > 0.10$ .

were involved in the defense subclass of defense. The remaining subclasses were inconsistent. Overall, most flowering genes participated in the metabolic process subclass.

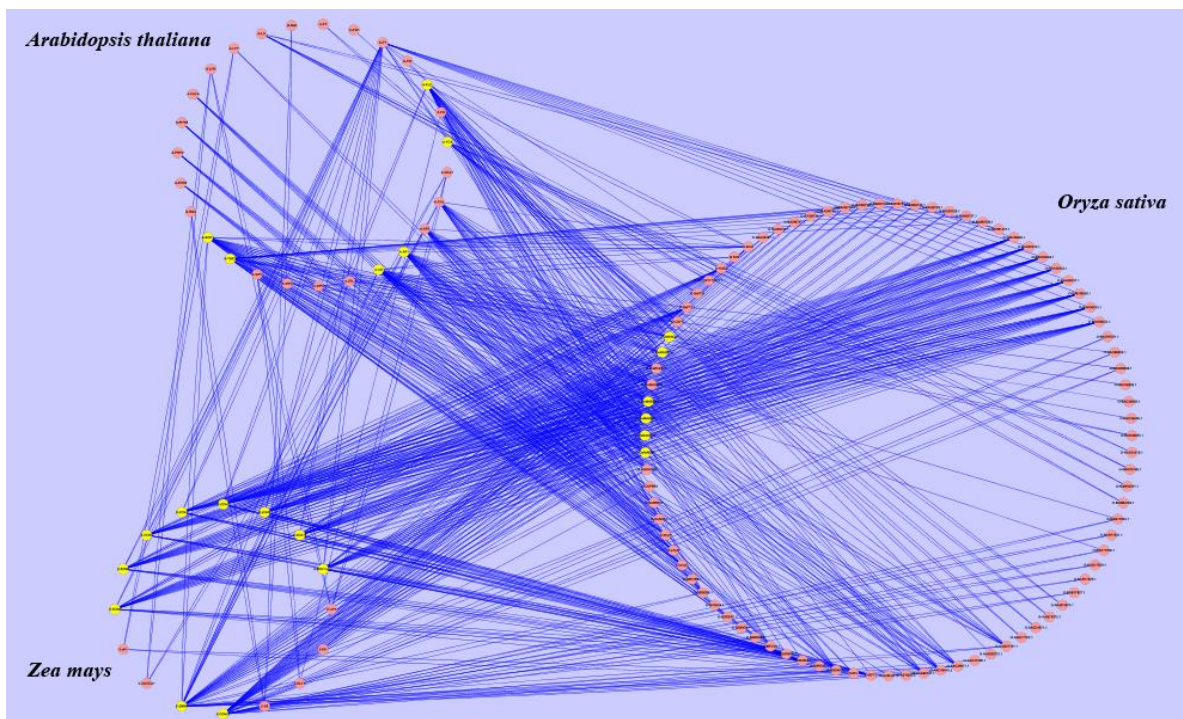
In the annotation of the molecular function, the flowering genes were assigned to six subclasses (Figure 6c). The subclass with the highest proportion of genes was binding, which covered 35.0, 54.9 and 69.7% of the flowering genes of *A.*

*thaliana*, *O. sativa*, and *Z. mays*. Second, catalytic activity occupied 40.8 and 13.7% of *A. thaliana* and *O. sativa* genes respectively. The third largest subclass was molecular transducer activity, which constituted 11.0% of *A. thaliana* flowering genes and 8.4% of *O. sativa* genes. The remaining minor subclasses mainly included transcription regulation and electron carrier activities. The results indicated that the main molecular function

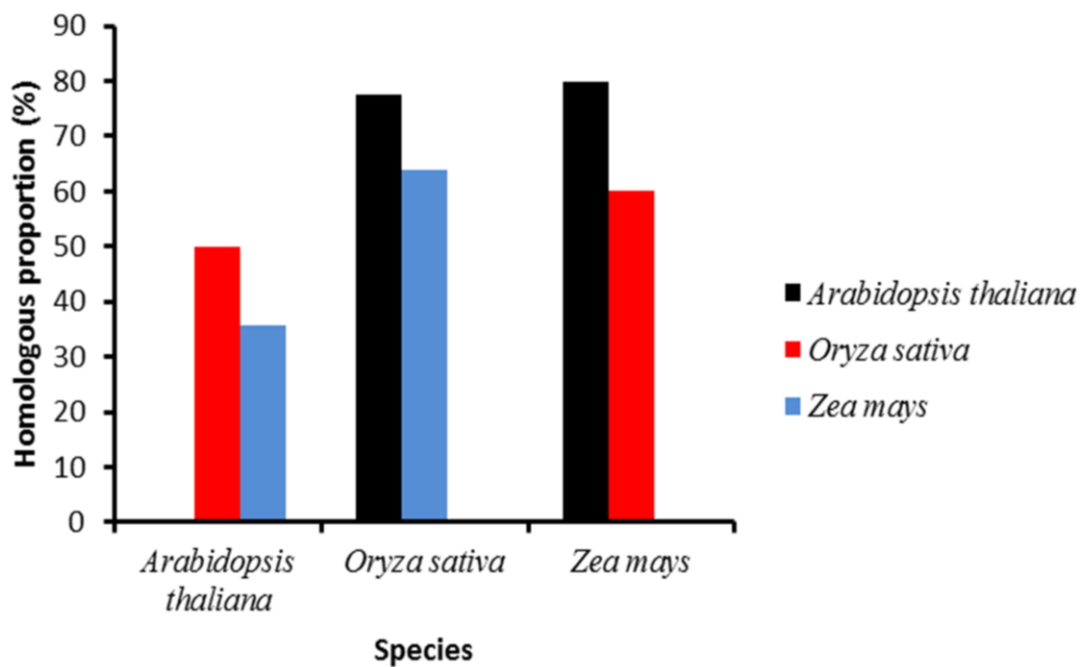
of the flowering genes is binding.

## DISCUSSION

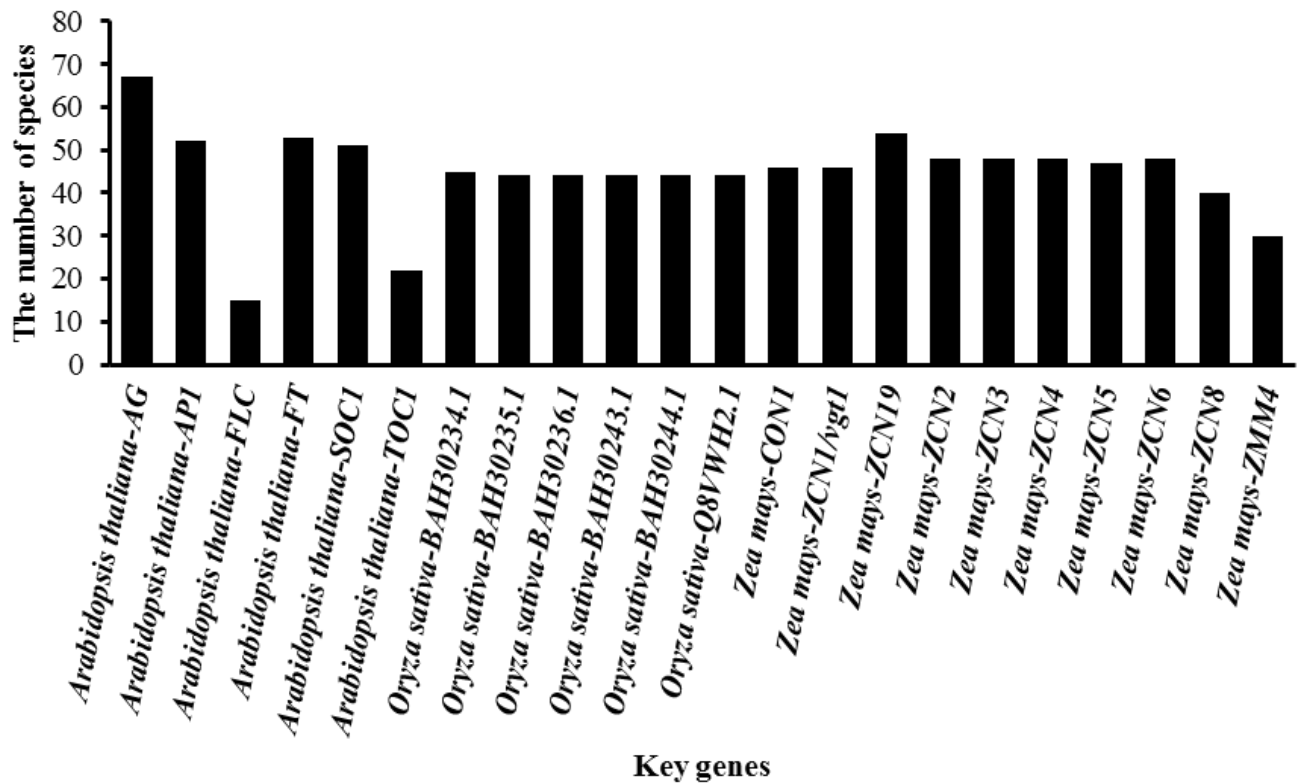
The dissection and comparison of the flowering genes of different plants is important for understanding molecular mechanisms behind the regulatory network of flowering. In the present



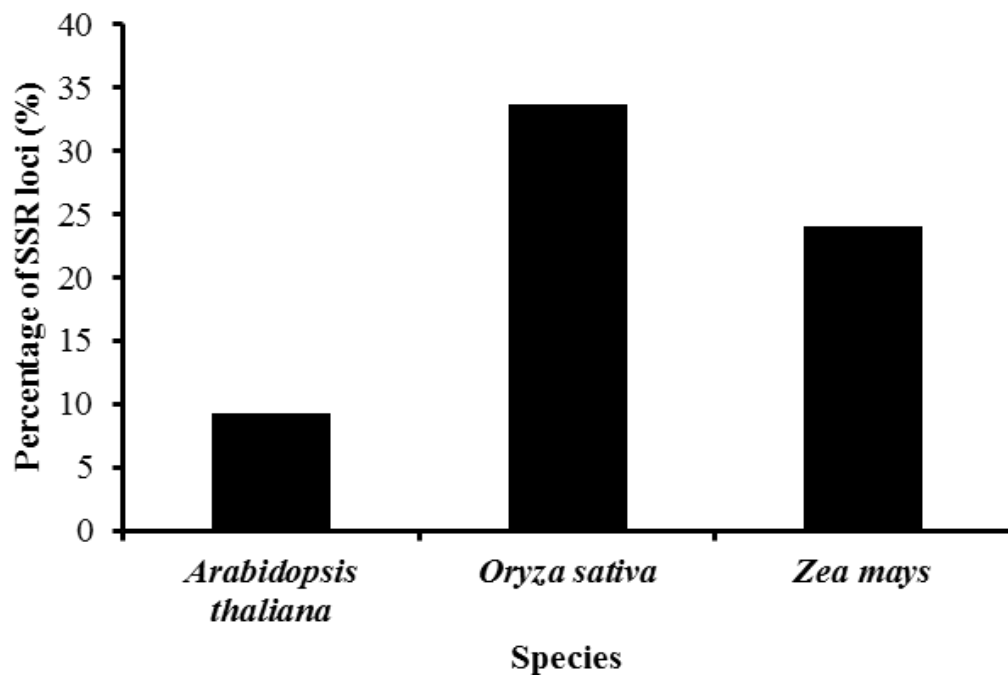
**Figure 2.** The similarity relationship analysis results for the regulatory networks of *A. thaliana*, *O. sativa*, and *Z. mays*. The edges (blue lines) suggest the relationship between two homologous node genes. A total of 22 (16.9%) node genes are marked by yellow and were inferred to be critical flowering genes (Edges  $\geq$  14). The pink nodes indicate the common flowering genes between these three species (Edges  $<$  14).



**Figure 3.** The pair-wise proportions of homologous flowering-related genes between *A. thaliana*, *O. sativa*, and *Z. mays*. The x-axis represents the three different species, and the y-axis denotes the proportion of homologous flowering genes between each two species.

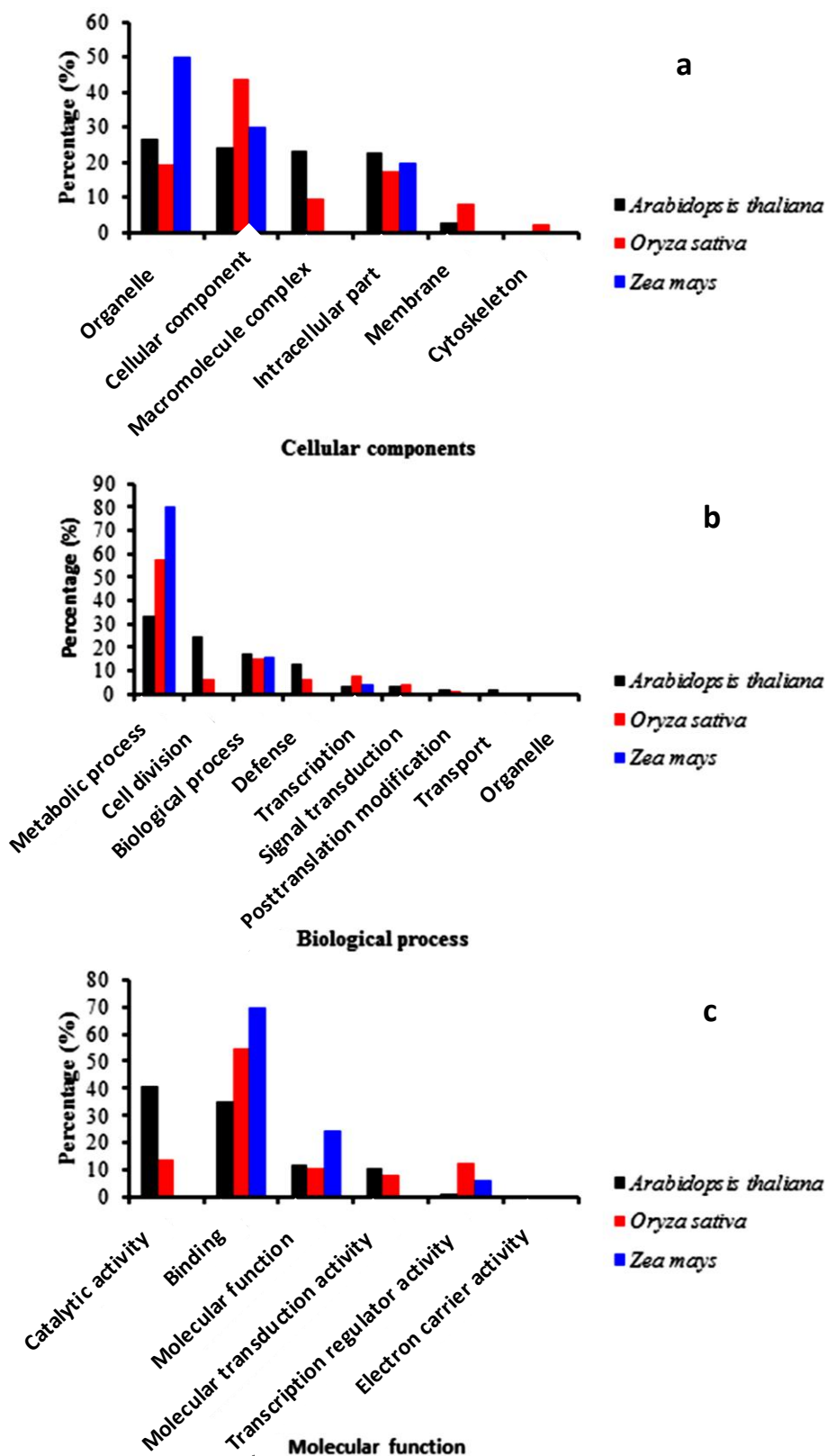


**Figure 4.** The number of species, which have identified homology, sequences with the key flowering genes other than *A. thaliana*, *O. sativa*, and *Z. mays*. The x-axis represents the proteins related to flowering genes in the three species, and the y-axis denotes the number of other species which have homologues to key flowering genes of these three species.



**Figure 5.** The percentage of flowering related genes with SSR loci. The x-axis represents the three different species, and the y-axis denotes the percentage of genes with SSR loci.





**Figure 6.** Gene ontology classification, showing assignment of flowering genes into three broad categories, namely, (a) cellular components, (b) biological process and (c) molecular function.

study, we characterized and compared the flowering genes of *A. thaliana*, *O. sativa*, and *Z. mays*. Nucleotide polymorphism analysis of key genes revealed universal sequence conservation during evolution, though some small InDels were also detected. Furthermore, many of these flowering genes harbored simple repeat sequences and these repeats perhaps drove the variation and divergence of flowering genes. These results will identify key flowering genes, facilitating their efficient cloning and an in-depth understanding of the molecular mechanism of flowering. Approximately 61.2% homology was detected between key flowering genes in the three species. It has been previously observed that 80.6% of the genes of *A. thaliana* were homologous to those of *O. sativa* (the number of homologous genes/ the number of genes of *A. thaliana*), whereas only 49.4% genes of *O. sativa* are homologous to those of *A. thaliana* (the number of homologous genes/ the number of genes of *O. sativa*) (Yu et al., 2002). A large proportion of rice genes have no recognizable homologs, possibly accrued from a gradient in the GC content of rice coding sequences (Yu et al. 2002). In our studies, we observed that *A. thaliana* genes had 77.5% identity with *O. sativa*, whereas only 50% of the flowering genes of *O. sativa* were homologous to those of *A. thaliana*. The overall homology observed in our study (77.5%) is close to the previously produced genome analysis figure (80.6%), supporting our approach for the comparative analysis of flowering genes between the three species.

Twenty-two (16.9%) of genes were considered to be highly conserved key genes in the regulatory network of flowering genes (marked by yellow in Figure 2). These key genes also had counterparts in many other species with an average of 45 homologues per gene. This result indicates that these conserved genes have some relative conservation and that the rapid spread of the ancestral genes across more species may be related to gene functionality and species adaptation to environments (Blanc and Wolfe, 2004). Previous studies in *O. sativa* (Tsuji et al., 2011) and *B. vulgaris* (Abou-Elwafa et al., 2011) have indicated that a certain degree of gene differentiation helped facilitate the evolution of unique genes that were more suitable for specific ecological niches. Tajima's D values of some genes indicated that these genes were subjected to negative selection and were not significant (Baudry et al., 2004; Tajima, 1989; Carlson et al., 2005).

Gene annotation studies suggested that most of flowering genes across different species displayed high functional similarity. There were obvious biases of flowering genes in cellular components, biological process and molecular function classifications: the organelle was the main cellular component in which the flowering genes of *A. thaliana* and *Z. mays* operated, whereas the largest proportion of *O. sativa* genes operated in the cellular component. Metabolic processes were the main biological process of the three species. In

addition, binding was the main molecular function for most of the flowering genes of *O. sativa* and *Z. mays*. For *A. thaliana*, the main molecular function appeared to be catalytic activity. These results were consistent with the previous research outcomes on the floral regulation pathway, which suggest that photoreceptors (Simpson and Dean, 2002) and specifically phytochromes, cryptochromes, phototropin, and circadian clock regulators (Dunlap, 1999), function in the organelles for many steps in the photoperiod pathway. The autonomous flowering pathway controls flowering time by transcriptional modification of chromatin and RNA (Simpson, 2004; He et al., 2003). DNA and histone methylation are involved in the vernalization pathway (Jung and Muller, 2009), and flower development is also regulated by the activity of gibberellin in the gibberellin pathway (Davidson et al., 2005; He and Amasino, 2005). All of these mechanisms are associated with metabolic processes and binding functions.

Gene polymorphism includes both length polymorphism and nucleotide-composition polymorphisms. Gene divergence directly results in the diversification of biological process to adapt to the surrounding biological environment. Sixty-nine SSR loci were detected in 250 flowering genes with an average of 0.276 SSRs per gene, which suggested that flowering genes are SSR enriched. SSR's are among the most mutable tracts in plant genomes (Tautz, 1989) and are not only found frequently within genes, including within protein-coding regions, untranslated regions, and introns, but are also common in transcribed spacer regions (Morgante et al., 2002). Some SSR mutations can result in the alteration of the gene function (Li et al., 2004; Kashi and King, 2006). A high ratio of trinucleotide-motif SSRs in flowering time genes in *A. thaliana* and *O. sativa* suggested that mutations in these genes increase and decrease the number of repeats to change a few amino acids might be tolerated, as these mutations do not thoroughly change the composition of proteins through frameshift mutations. However, in *Z. mays*, the mononucleotide-motif SSR A was the predominant SSR motif type (68.0%). The results suggest SSR enrichment in flowering genes has distinct characteristics and may be related to flowering genes mutation. Flowering is a complex trait controlled by multiple genes and is generally influenced by environment. The variability of SSRs in flowering genes may create allelic variation in species, allowing swift response to changing external environments (Kashi et al., 1997). Thus, SSRs of flowering genes may be an internal force to drive the evolution and variation of these genes. Such abundant SSRs of flowering genes can also be used to design SSR primers as functional markers for use in gene mapping and marker-assisted selection in plant breeding.

In summary, we collected the flowering genes of *A. thaliana*, *O. sativa*, and *Z. mays* and compared the

commonalities and divergence points between them with respect to structure, nucleotide variation pattern and trait differentiation. We also posit the possible reason for the differences between the flowering genes, and describe the functional annotation of the flowering genes in detail. Overall, these flowering genes or regulatory network showed higher conservation and mutation comprised the variation of mostly small fragments. Therefore, for a species without available genome sequence, it is possible to select and clone some key flowering genes by way of homology with established genome sequences, to achieve rapid understanding of the mechanism of flowering genes.

## REFERENCES

- Abou-Elwafa SF, Buttner B, Chia T, Schulze-Buxloh G, Hohmann U, Mutasa-Gottgens E, Jung C, Muller AE (2011). Conservation and divergence of autonomous pathway genes in the flowering regulatory network of *Beta vulgaris*. *J. Exp. Bot.* 62(10):3359-3374.
- Amasino R (2010). Seasonal and developmental timing of flowering. *Plant J.* 61(6):1001-1013.
- Angelovici R, Fait A, Zhu X, Szymanski J, Feldmesser E, Fernie AR, Galili G (2009). Deciphering transcriptional and metabolic networks associated with lysine metabolism during *Arabidopsis* seed development. *Plant Physiol.* 151(4):2058-2072.
- Baudry E, Viginier B, Veuille M (2004). Non-African populations of *Drosophila melanogaster* have a unique origin. *Mol. Biol. Evol.* 21(8):1482-1491.
- Blanc G, Wolfe KH (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant cell* 16(7):1679-1691.
- Boss PK, Bastow RM, Mylne JS, Dean C (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell.* 16 Suppl:S18-31.
- Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, Nickerson DA (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* 15(11):1553-1565.
- Conesa A, Gotz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genom.* 2008:619832.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674-3676.
- Da Maia LC, Palmieri DA, de Souza VQ, Kopp MM, de Carvalho FI, Costa de Oliveira A (2008). SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. *Int. J. Plant. Genom.* 2008:412696.
- Davidson SE, Swain SM, Reid JB (2005). Regulation of the early GA biosynthesis pathway in pea. *Planta* 222(6):1010-1019.
- Dunlap JC (1999) Molecular bases for circadian clocks. *Cell* 96(2):271-290.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41:95-98.
- Halliday KJ, Salter MG, Thingnaes E, Whitelam GC (2003). Phytochrome control of flowering is temperature sensitive and correlates with expression of the floral integrator *FT*. *Plant J.* 33(5):875-885.
- He Y, Amasino RM (2005). Role of chromatin modification in flowering-time control. *Trends Plant Sci.* 10(1):30-35.
- He Y, Michaels SD, Amasino RM (2003). Regulation of flowering time by histone acetylation in *Arabidopsis*. *Science* 302(5651):1751-1754.
- Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002). *Arabidopsis* map-based cloning in the post-genome era. *Plant Physiol.* 129(2):440-450.
- Jung C, Muller AE (2009). Flowering time control and applications in plant breeding. *Trends Plant Sci.* 14(10):563-573.
- Kashi Y, King D, Soller M (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* 13(2):74-78
- Kashi Y, King DG (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends Genet* 22(5):253-259
- Kim DH, Doyle MR, Sung S, Amasino RM (2009). Vernalization: winter and the timing of flowering in plants. *Annu. Rev. Cell Dev. Biol.* 25:277-299.
- Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002). *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant Cell Physiol.* 43(10):1096-1105.
- Komiya R, Yokoi S, Shimamoto K (2009). A gene network for long-day flowering activates *RFT1* encoding a mobile flowering signal in rice. *Development* 136(20):3443-3450.
- Lee H, Suh SS, Park E, Cho E, Ahn JH, Kim SG, Lee JS, Kwon YM, Lee I (2000). The *AGAMOUS-LIKE 20* MADS domain protein integrates floral inductive pathways in *Arabidopsis*. *Genes Dev.* 14(18):2366-2376.
- Li YC, Korol AB, Fahima T, Nevo E (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21(6):991-1007.
- Macknight R, Duroux M, Laurie R, Dijkwel P, Simpson G, Dean C (2002). Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter *FCA*. *Plant Cell* 14(4):877-888.
- Morgante M, Hanafey M, Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* 30(2):194-200.
- Nei M (1987). *Molecular evolutionary genetics*. Columbia University Press, New York.
- Rozas J, Rozas R (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15(2):174-175.
- Simpson GG (2004). The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Curr. Opin. Plant Biol.* 7(5):570-574.
- Simpson GG, Dean C (2002). *Arabidopsis*, the Rosetta stone of flowering time? *Science* 296(5566):285-289.
- Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2):437-460.
- Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585-595.
- Tautz D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17(16):6463-6471.
- Taylor A, Massiah AJ, Thomas B (2010). Conservation of *Arabidopsis thaliana* photoperiodic flowering time genes in onion (*Allium cepa* L.). *Plant Cell Physiol.* 51(10):1638-1647.
- Tsuji H, Taoka K, Shimamoto K (2011). Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. *Curr. Opin. Plant Biol.* 14(1):45-52.
- Watterson GA (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7(2):256-276.
- Wei X, Xu J, Guo H, Jiang L, Chen S, Yu C, Zhou Z, Hu P, Zhai H, Wan J (2010). *DTH8* suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant Physiol.* 153(4):1747-1758.
- Wellmer F, Riechmann JL (2010). Gene networks controlling the initiation of flower development. *Trends. Genet.* 26(12):519-527.
- Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000). *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* 12(12):2473-2484.
- Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296(5565):79.