Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

Ram Kumar Choudhary
Dr. Rajendra Prasad
Agricultural University, Pusa,
Samastipur, Bihar, India

Tanmaya Kumar Sahu
ICAR-Indian Agricultural
Statistics Research Institute,
New Delhi, India

Himansu Kumar
ICAR-Indian Agricultural
Statistics Research Institute,
New Delhi, India

Atmakuri Ramakrishna Rao
ICAR-Indian Agricultural
Statistics Research Institute,
New Delhi, India

Shiv Kumar Choudhary
Bihar Agriculture University,
Sabaur, Bhagalpur, Bihar, India

Tusar Kanti Behera
ICAR- Indian Agricultural
Research Institute, New Delhi,
India

**International Web-Conference
On
New Trends in Agriculture, Environmental & Biological Sciences for
Inclusive Development
(21-22 June, 2020)**

Computational identification of putative genes and vital amino acids involved in biennial rhythm in mango (*Mangifera indica* L.)

Ram Kumar Choudhary, Tanmaya Kumar Sahu, Himansu Kumar, Atmakuri Ramakrishna Rao, Shiv Kumar Choudhary and Tusar Kanti Behera

Abstract
Mango is one of the important perennial fruit crops that exhibits bienniality and results in high economic loss. The molecular mechanisms involved in the bienniality behavior of mango are yet to be fully explored. In this study, an attempt was made to identify putative genes and amino acid residues playing an important role in biennial rhythm in mango through comparative genomics approach. Here, the reported genes responsible for bienniality in *Arabidopsis* and apple were used to identify the corresponding orthologs in mango. The phylogenetic analysis was also carried out to understand the evolutionary relationships among the bienniality related proteins of mango, *Arabidopsis* and apple. Further, the amino acid residues conserved in the protein products of the candidate genes in mango were identified by both sequence and structure alignment. From both the alignments, arginine and glycine were found conserved in the candidate bienniality proteins of mango that indicates the possible involvement of these residues in the bienniality behavior. A total of twenty nine putative genes of mango related to bienniality rhythm were identified and are populated in the Mango Bienniality Gene Database (MBGDB), accessible at http://webapp.cabgrid.res.in/mangodb. The results of this study and the developed database are expected to help enable the scientists, researchers and mango growers to minimize the economical loss due to bienniality.

Keywords: Bienniality, AGAMOUS-like protein, terminal flowering, MADS1, Mango Bienniality Gene Database.

Introduction
Most of the perennial crops exhibit biennial rhythm in fruiting and growth. Mango is an important perennial fruit crop that shows biennial behavior resulting high economic loss [1]. Bienniality is a complex phenomenon, where heavy fruiting occurs in one year (*on year*) and little or no fruiting occurs in the next year (*off year*). The concentration of carbohydrate in apical shoot and maturity level of terminal shoots are important causes of the biennial behaviour in mango [2]. Especially, absence or reduction of carbohydrate concentration in the apical shoot results into *off year* in mango cultivation. Carbohydrate reserves have a direct relationship with flower formation in mango [3]. Further, it has been found that highest polysaccharide level in the wood of shoots with enlarged apical buds and lowest level in the absence of apical buds in the *Haden* cultivar of mango [4]. Periodic cessation of the fruit is required for the proper development of physiological maturity and flowering [5]. However, flowering in mango depends upon *on* and *off* stage of the crop irrespective of cessation of apical shoot growth. In addition to carbohydrates, Gibberellic Acid (GA)-like substances also have an important role in flowering [6-7]. Greater amount of GA retards flowering whereas high amount of auxin promotes the flowering. KNO_3 is favourable for the induction of flowering in the mango [8].
Despite of the availability and importance of genomic resources, mango genomic resources lag behind other commercial crops. However, there are few but highly redundant expressed

Correspondence
Ram Kumar Choudhary
Dr. Rajendra Prasad
Agricultural University, Pusa,
Samastipur, Bihar, India

sequence tag (EST) sequences available in GenBank for mango [9]. In order to supplement this lag several researchers are working on sequencing and characterization of mango genome. Such information is useful for understanding the fundamental molecular biology of mango. A draft assembly of 412,728 (about 450 Mbps) contigs has been generated covering more than 95 % of the genome at ICAR-National Research Centre on Plant Biotechnology, India [10].

Characterized the mango leaf and chloroplast (cp) using next generation sequencing technology [9]. Apple orthologues of flowering and hormone-related genes and found that flowering related genes are less likely to be responsible for biennial bearing than hormones-related genes [11]. An exhaustive *in silico* inventory of floral and hormone-related genes in apple was performed in order to establish a list of candidate genes that have possible role in the process of biennial bearing. The genes co-locating with Quantitative Trait Loci (QTLs) were considered as potential candidates and specific markers were developed in order to position them on the genetic map to test their relationship with QTLs. In citrus bud morphology, expression of flowering control genes, and global gene expression in *on* and *off* crop buds were studied [12]. They concluded that some of the metabolic and regulatory processes are altered in *on* and *off* buds. A comprehensive study on olive miRNA related to alternate bearing was performed and found potential alternate bearing-associated processes such as development, hormone-mediated signaling and organ morphogenesis [13]. Influence of carbohydrate, minerals and hormones in alternate bearing in black pepper were studied [14]. Isolated and analyzed the expression of cDNAs that are associated with alternate bearing in olive and annotated 37 putative transcripts corresponding to the known genes [15].

Several databases exist on mango involving the information on genomic resources, morphology, related IPR issues *etc*. in public domain [16]. Have developed a mango resource information system (MRIS) to manage the phenotypic as well as molecular information of mango. The MRIS enables searching DNA sequences, protein sequences, (expressed sequence tag) ESTs *etc*. However, MRIS lack information on the genes responsible for bienniality in mango. So, there is a need to identify and conserve the information on these genes and related protein products, which in turn will be useful for research on bienniality behavior of perennial crops. Thus, here an attempt has been made in this study to (i) identify the genes responsible for biennial rhythm in mango through computational approaches and (ii) develop a database based on this information for providing easy access to the users working in the area of mango genomics and breeding.

Materials and methods
Identification of candidate genes responsible for bienniality
The sequences of protein products of candidate genes that could be possibly related with biennial rhythm in mango were collected from National Centre for Biotechnology Information (NCBI; *http://www.ncbi.nlm.nih.gov*) based on the literature. The collected proteins were mostly the products of floral development, carbohydrate depletion and hormone-related genes in mango. The protein sequences of *Arabidopsis thaliana* [17] and apple [11]. Corresponding to the genes related to biennial rhythm that are involved in regulating floral development, carbohydrate depletion and hormonal activity were also retrieved from NCBI. Further, offline Basic Local Alignment Search Tool (BLAST) [18] was executed by considering the mango protein sequences as the database and

protein sequences of *Arabidopsis* and apple as queries. Top hits were filtered out corresponding to e-values lying in the first quartile of all e-values from the BLAST result. Further, distinct sequences of *Arabidopsis* and apple matched with corresponding mango proteins were collected. The filtered genes of mango shared a homology with the genes related to biennial rhythm in *Arabidopsis* and apple, hence, these genes were considered as putative genes involved in the exhibition of biennial rhythm in mango.

Evolutionary relationship among biennial rhythm genes of Mango, Apple and Arabidopsis
Evolutionary relationship among the putative genes responsible for biennial behavior of *Arabidopsis*, Apple and Mango species has been established by phylogenetic analysis using the MEGA version 6.06 software [19]. Initially, multiple sequence alignment (MSA) was created for all the above mentioned proteins and the MSA was further used as a reference to create a phylogenetic tree. The phylogenetic tree was constructed using maximum parsimony method by fixing 1000 bootstrap samples for testing.

Identification of key residues responsible for bienniality
Functional domains, namely, AGAMOUS-like protein, APETALA1-like protein, CONSTANS, flowering locus T-like protein, LEAFY-like protein, MADS1, MAP1, SEPALLATA1-like protein, SOC1 protein and terminal flowering proteins are to be associated with the biennial behavior in the fruit crops reported [20]. Hence, the proteins having these domains were considered to analyze the structural similarity and to find out the conserved amino acid residues at tertiary structure level. Discovery Studio 3.5 was used for both MSA and structural alignment of putative proteins of mango related to bienniality. The sequence alignment was treated as a reference to the structural alignment in order to identify the amino acid residues conserved at both primary and tertiary protein structure levels. Before the structural alignment, protein structures were initially modeled by using SWISS-MODEL server 3.07 (*http://swissmodel.expasy.org*) and then validated by using SAVES server.

The stereo-chemical stability of the protein structures were examined using Ramachandran plot that showed the empirical distribution of amino acids in a plot, where the Φ and Ψ angles of the amino acids are represented in X and Y axes respectively. The Rampage server (*http://www.mordred.bioc.cam.ac.uk/ rapper/rampage.php*) was used to generate Ramachandran plots for the predicted protein structures. The incorrect loops of the protein structures were refined using the ModLoop server (*http://www.modbase.compbio.ucsf.edu/modloop/*), as loops in a protein framework frequently determine the functional specificity in terms of recognition of its interacting partners, enzymatic reaction with substrate and conformational changes [21]. Further, the structures were revalidated using SAVES server and Rampage server. This process was repeated till obtaining a stereo-chemically correct structure. The energy minimization of validated structures was done using YASARA (*http://www.yasara.org/minimizationserver.htm*) minimization server [22].

Finally, the energy minimized structures of the proteins were superposed in Discover Studio 3.5 using the sequence alignment as reference. From the resultant superposition, the common geometrical cores along with the structurally conserved amino acid residues in all the proteins were

identified. As domains are the functional part of the proteins, the residues found conserved in domain co-ordinates are likely to have meaningful biological significance. Thus, the residues found conserved in both structural and sequence alignments of protein domains are designated as the putative key residues responsible for bienniality in mango.

Mango Bienniality Gene Database (MBGDB)
A database on the putative genes related to biennial rhythm in mango was developed to share the information on the genes related to biennial rhythm in *Arabidopsis,* apple and mango. The database was developed based on relational database modeling using MySQL database system. The web interface of the database was designed using Hyper Text Markup Language (HTML) and Preprocessor Hypertext (PHP 5.2.1). The *MBGDB* is accessible at www.webapp.cabgrid.res.in/mangodb.

Results and discussion
Putative genes for biennial rhythm in mango
The biennial rhythm is a complex phenomenon being exhibited by perennial crops including mango. The most probable causes are due to the presence of carbohydrate depletion, floral and hormones-related genes [2]. In the present investigation, a total of twenty nine putative genes related to bienniality in mango were identified. The identified genes are based on the identification of biennial rhythm in apple [11] and *Arabidopsis* [17]. Also found Quantitative Treat Loci (QTL) and mapped the putative genes of floral and hormones-related genes on the whole genome of apple [11]. These two species were considered because apple is a fruit crop as mango and *Arabidopsis* is a model organism for plants. Hence, the reliability on results can be expected as mango shares similarity with these plants. The candidate genes of mango identified in this study are listed in Table 1.

Phylogenetic analysis (Fig. 1) of all the proteins sequences considered under this study revealed the similarity of the mango proteins with the apple proteins, as expected. However, an exception was also noticed, *i.e.,* the terminal flowering protein of mango showed more closeness to the centroradialis protein (CEN gene) of *Arabidopsis*. The CEN gene is responsible for terminal flowering and a conditional mutation in this gene results in delayed terminal flowering in *Arabidopsis* [23]. This implies that the mechanism of terminal flowering might be similar in mango and *Arabidopsis*. Interestingly, it has been reported that the floral initiation of *Arabidopsis* and mango is dependent on temperature and other factors whereas in apple it is mostly dependent on the aspects of vegetative development which corroborates with the findings of phylogenetic analysis [20].

Identification of conserved residues related to biennial rhythm in mango
Similarity based on only at sequence level does not fully explain the functional relationships between the proteins; hence, structural alignment was carried out. The sequence analysis and conserved domain database search revealed the existence of nine different domains (AMAGOUS, MAP1,

SEPALLATA1-like protein, LEAFY, MADS1, SOC, CONSTANS, FLOWERING LOCUS T-like protein and APETAL1-like protein) in the considered protein sequences. The domains, for which tertiary structure were not available in PDB, were modeled *in silico* using homology modeling or fold recognition or *abinitio* methods based on the availability of template structures. Modeled structures were validated through RMAPAGE server using Ramachandran plot. Ramchandran plot confirmed the stability of the 3D structures, as 95-99% of residues were found in the favored region (Table 2). Energy minimization and loop refinement was performed on some of the modeled structures which were less stable and revalidated. After revalidation, the structures were found stereo chemically stable with allowed number of residues in favored area of Ramchandran plot. The atomic coordinate files of the stable structures were used for the superposition to identify the structurally conserved residues. The super-positioning of the structures were done by taking the sequence alignment as reference, with the expectation of finding the amino acid residues conserved both at sequence and structure level. Fig. 2 presents both structure and sequence alignments of the considered domains. The conserved residues in the sequence alignment are highlighted in deep color. Arginine was found conserved in the aligned domains at the alignment positions 33 and 64 (Fig. 2) and fortunately the role of arginine in halting the growth of biennial and perennial crops, has been confirmed [24]. Besides, Glycine was also found conserved in some of the protein domains.

Mango Bienniality Gene Database (MBGDB)
A database, Mango Bienniality Gene Database (MBGDB) was developed to manage information on genes related to biennial rhythm in mango. The home page of MBGDB is presented in Fig. 3. Home page of the database enables users to get information about conserved domains, homology of mango bienniality genes with that of Apple and *Arabidopsis*. Detailed information on conserved domains including gene ID, CDD accession, name of the conserved domain, description, co-ordinates, e-value, and bit score etc. are also provided as a link in the database. Further, the information on putative genes of mango that share homology with apple and *Arabidopsis* are provided as separate links.

The study reveals twenty nine putative genes responsible for mango bienniality identified by *in silico* methods and presents a database consisting of biomolecular information related to bienniality in mango. The conserved residues having a possible role in biennial rhythm in mango were also identified in this study by sequence and structure alignment. The phylogenetic analysis also revealed homology of mango genes with apple and *Arabidopsis* bienniality genes. The developed database is expected to be useful to scientists, breeders and other researchers working in this field. The MBGDB may serve as a model for other tropical and subtropical fruit crops related to biennial rhythm. Further, MBGDB is believed to supplement the existing mango databases.

Table 1: List of putative genes identified for mango bienniality in mango

Sl. No.	Gene id	Accession	Description
1	219964711	ACL68407.1	MAP1
2	219964713	ACL68408.1	MAP1
3	224760941	ACN62415.1	CONSTANS-like protein
4	225638983	ACN97631.1	AGAMOUS-like protein, partial
5	225638985	ACN97632.1	LEAFY-like protein, partial
6	240130270	ACS45102.1	APETALA1-like protein
7	240130272	ACS45103.1	APETALA1-like protein
8	28300409	AAO37817.1	Leafy, partial
9	323650469	ADX97315.1	LEAFY-like protein
10	323650471	ADX97316.1	LEAFY-like protein
11	323650473	ADX97317.1	LEAFY-like protein
12	323650475	ADX97318.1	LEAFY-like protein
13	323650477	ADX97319.1	LEAFY-like protein
14	323650479	ADX97320.1	LEAFY-like protein
15	323650483	ADX97322.1	CONSTANS
16	323650485	ADX97323.1	CONSTANS
17	323650487	ADX97324.1	SOC1
18	323650495	ADX97328.1	SEPALLATA1-like protein
19	346683573	AEO45959.1	SEPALLATA3-like protein
20	354683682	AER34988.1	AGAMOUS-like protein
21	354683684	AER34989.1	AGAMOUS-like protein
22	384080992	BAM10941.1	Flowering locus T-like protein
23	401606224	AFP95335.1	Flowering locus T-like protein
24	448047728	AFR54358.2	Flowering locus T-like protein
25	544388846	AGW23412.1	Terminal flower
26	544452523	AGW23594.1	MADS1
27	545793525	AGW80455.1	Constans-like protein
28	827520732	AKJ85721.1	SOC1 protein
29	857280783	AKO60174.1	SEPALLATA1-like protein

Table 2: Ramchandaran plot results showing percentage of amino acid in Favored, Allowed and Outlier regions.

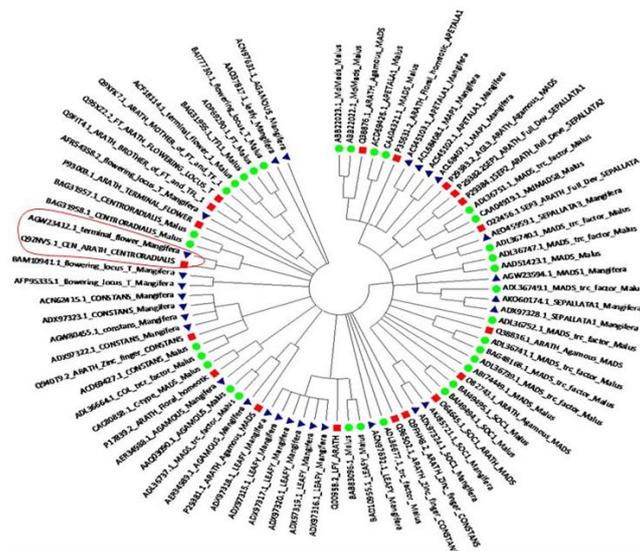| Sl. No.	Domains	Favored	Allowed	Outlier
1.	AMAGOUS	95.70%	4.30%	0.00%
2.	BAM01	96.30%	3.70%	0.00%
3.	SEPA	98.50%	1.50%	0.00%
4.	LEAFY	99.20%	0.80%	0.00%
5.	MADS1	96.70%	1.10%	2.20%
6.	SOC	96.70%	1.10%	2.20%
7.	AptaLike	87.30%	11.30%	1.40%
8.	Terminal Flower	94.20%	4.10%	1.70%
9.	Constant	85.70%	9.10%	5.20%
10.	MAP1	85.90%	9.90%	4.20%

Fig. 1. Phylogenetic tree showing evolutionary relationship proteins of Mango, Apple and Arabidopsis related biennial rhythm. The marked region shows the terminal flowering genes of mango clustered with the CEN gene of *Arabidopsis*.
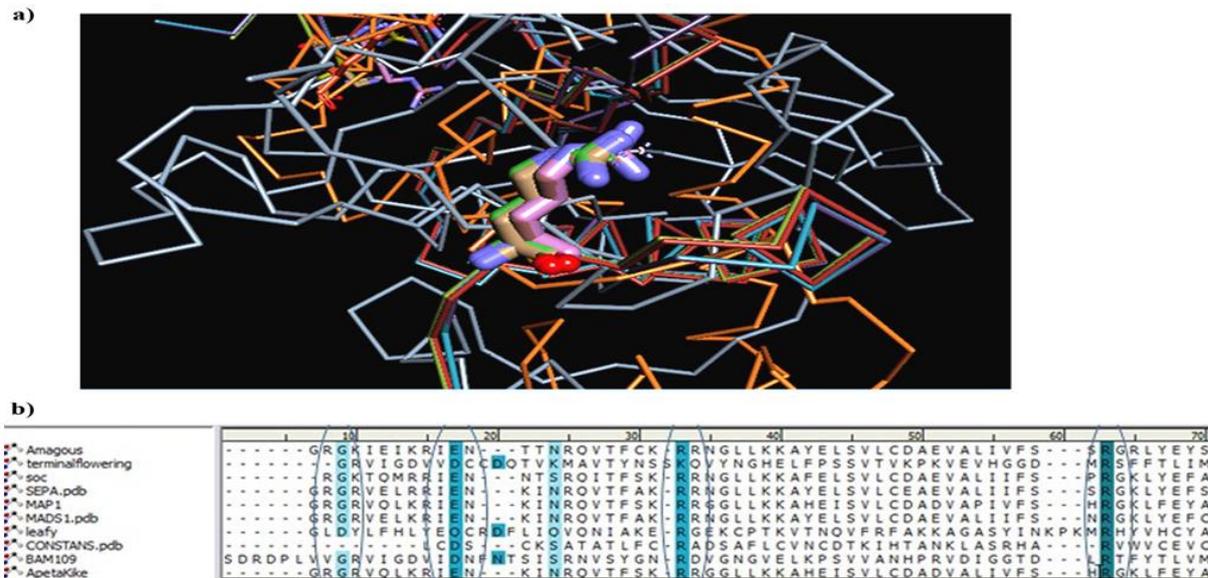
Fig 2: (a) Superimposed protein structures showing conserved arginine at 64th position.
(b) Multiple sequence alignment of protein domains showing conserved residues.

Fig 3: Home page of Mango Bienniality Gene Database (MBGDB)

References
1. Singh VK, Sharma K. Physiological and Biochemical Changes during Flowering of Mango (*Mangifera indica* L.). Int. J Dev. Biol. 2008; 2(2):100-105.
2. Pandey RM. Physiology of flowering in mango. Acta Hortic. (Netherlands), 1989.
3. Reece PC, Furr JR, Cooper WC. Further studies of floral induction in the Haden mango (*Mangifera indica* L.). Am. J Bot. 1949, 734-740.
4. Gazit S. Initiation and development of flower buds in various mango varieties, 1960.
5. Nakasone HY, Bowers FAI, Beaumont JH. Terminal growth and flowering behavior of the Pirie mango (*Mangifera indica* L.) in Hawaii, 1955.
6. Kulkarni P. Studies on regulation of vegetative growth, flowering and fruit drop in mango (Mangifera indica L.) cv. Alphonso. Thesis Abstracts, Haryana Agric. Univ. 1983; 9:344-345.
7. Reddy YTN, Kohli RR, Singh G, Bhargava BS. Effects of rootstocks on growth, yield and leaf nutrient composition of mango. Fruits. 1989; 44:409-413.
8. Chadha KL, Pal RN. Biennial bearing in mango. In: Chadha, K.L., Pareek, O.P. (Eds.), Advances in Horticulture. 4. Malhotra Publishing House, New Delhi. 1993, 2032-2033.
9. Azim MK, Khan IA, Zhang Y. Characterization of mango (*Mangifera indica* L.) transcriptome and chloroplast genome. Plant Mol. Biol. 2014; 85(1-2):193-208.
10. Singh NK. A draft genome of the king of fruit, mango (*Mangifera indica* L.). In *Plant and Animal Genome XXII Conference*. Plant and Animal Genome, 2014.
11. Guitton B, Kelner JJ, Velasco R, Gardiner SE, Chagne D, Costes E. Genetic control of biennial bearing in apple. J. Exp. Bot. 2011, 261.
12. Shalom L, Samuels S, Zur N, Shlizerman L, Zemach H, Weissberg M *et al*. Alternate bearing in citrus: changes in

the expression of flowering control genes and in global gene expression in on- versus off-Crop trees. PLoS One. 2012; 7(10):e46930.

13. Yanik H, Turktas M, Dundar E, Hernandez P, Dorado G, Unver T. Genome-wide identification of alternate bearing-associated microRNAs (miRNAs) in olive (Olea europaea L.). BMC Plant Biol. 2013; 13(1):10.

14. Krishnamurthy KS, Ankegowda SJ, Srinivasan V, Hamza S. Influence of carbohydrates, mineral nutrients and plant hormones in alternate bearing of black pepper (*Piper nigrum* L.). Am. J Plant Sci. 2013; 4(10):1960.

15. Dundar E, Suakar O, Unver T, Dagdelen A. Isolation and expression analysis of cDNAs that are associated with alternate bearing in *Olea europaea* L. cv. Ayvalik. BMC Genomics. 2013; 14(1):219.

16. Rajan S, Sahu TK, Yadava LP. Mango resources information system: an open access portrayal of phenotypical, genetic and chemical information on mango. In IX International Mango Symposium. 2010; 992:99-104.

17. Bernier G, Perilleux C. A physiological overview of the genetics of flowering time control. Plant Biotechnol. J. 2005; 3(1):3-16.

18. Altschul SF, Gish W, Miller W, Myers EW. and Lipman D. J. Basic local alignment search tool. J Mol. Biol. 1990; 215(3):403-410.

19. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 2013; 30(12):2725-2729.

20. Wilkie JD, Sedgley M, Olesen T. Regulation of floral initiation in horticultural trees. J Exp. Bot. 2008; 59(12):3215-3228.

21. Rossi DJ, Bryder D, Seita J, Nussenzweig A, Hoeijmakers J, Weissman IL. Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. Nature. 2007; 447(7145):725-729.

22. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson K *et al*. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. Protiens Struct. Funct. Bioinf. 2009; 77(S9):114-122.

23. Cremer F, Lonnig WE, Saedler H, Huijser P. The delayed terminal flower phenotype is caused by a conditional mutation in the centroradialis gene of snapdragon. Plant Physiol. 2001; 126(3):1031-1041.

24. Sagisaka S. Amino acid metabolism in nongrowing environments in higher plants. Amino Acids. 1993; 4(1):141-155.