

Genome-wide Investigation of MADS-box Members and Validation of a New Member MADS8 in *Fragaria Vesca*

Min ZHOU^{1,#}, Yan WANG^{2,#}, Hao-ru TANG¹ and Qing CHEN^{1, a,*}

¹College of Horticulture, Sichuan Agricultural University, Chengdu 611130, Sichuan, China

²Institute of Pomology and Olericulture, Sichuan Agricultural University, Chengdu 611130, Sichuan, China

[#]These authors contributed equally to this work

^asupnovel@gmail.com

^{*}Corresponding author

Keywords: Strawberry, MADS, Genome wide identification, Clone.

Abstract. Based on the whole genome sequence of *Fragaria vesca*, the members of the MADS gene family in strawberry genome were identified and analyzed. The MADS-box containing proteins were searched using both hidden markov model fitting with *Arabidopsis thaliana* proteins, and de novo prediction. 85 MADS-box protein coding genes were identified, with 74 with confidence. Five new proteins that were not annotated in the genome were found, one of which, MADS8, was experimentally validated. These results provide useful information for functional studies of this important protein family.

Introduction

The growth and development regulation mechanism of plants is the key and research hot spot for a long time. Among various factor, the MADS-box transcription factor mediated regulation of flower development in higher plants has become increasingly clear in recent years. MADS-box proteins comprised of a highly conserved DNA-binding domain. The first reported three factors were the yeast Mini Chromosome Maintenance 1 (MCM1), the *Arabidopsis* AGAMOUS (AG), the snapdragon DEFICIENS (DEF) and the human serum response factor Serum Response Factor (SRF) [1]. MADS-box protein is the acronym of these four factors, which have similar structure characteristics.

To date, most of the genes involved in the development of plant flowers belong to the MADS-box gene family. Based on the genetic studies of dicotyledonous plant mutants, the ABCDE model of flower development has been proposed. A, B, C, D and E represent the five different function types of floral genes. They function alone or collaboration to control the morphology of flower development [2, 3]. With the study of plants such as *Arabidopsis* and *Petunia*, many MADS-box genes were cloned, with some being functionally characterized. There are more than 100 in *Arabidopsis* genome [4]. So far, hundreds of MADS-box genes have been found in other plants. Most of these factors play significant roles in the development of plant reproductive organs and some were also involved in the development of plant vegetative organs [5].

Strawberry is an important berry fruit in the horticultural industry. Its cultivation area ranks only second to that of grape in the world. The fruit is formed by the receptacle and the one-seed-achenes on it, which is a good material for studying the development of

flower/fruit organs. In contrast to the woody fruit trees, strawberries are perennial herbs. As a major crop in horticulture industry, the genome of strawberry has been released for years [6]. Primary investigations of MADS family members has been reported by Zhou et al. [7], but detailed and comprehensive analysis and validation still need to be conducted for these gene family members. What's more, expression analysis can bring us valuable information to find those MADS-box members that might get involved in strawberry flower development.

In this study, *Fragaria vesca* genome information was used to characterized the MADS thoroughly. Manually curated annotation of each member was carried out here. A new MADS-box gene named MADS8, located on LG2, which was not annotated in genome, was experimentally validated. These results pave a way for the further exploration of the family's genetic function.

Materials and Methods

Genome-wide MADS-box investigation

The whole genome sequence data for *Fragaria vesca* from Genome Database for Rosaceae (<https://www.rosaceae.org/>) v2.1 were used to isolate the MADS-box domain-containing sequences. Two strategies were used: 1) Genomic sequences for each of the seven linkage groups were all translated based on all possible open reading frames according to the '+' and '-' chain. The deduced amino acid information was searched using hidden markov model implemented in HMMER (v3.1b2) software against MADS-box pfam document 'PF00319' (e-value 0.00001). The obtained target segments were the candidate MADS-box gene coding regions. 2) Referring to the pipeline of Gramzow et al. [8], MADS-box encoding genes were semi-automatically predicted in the *Fragaria vesca* genome. The two parts of results were combined, redundant sequences were removed. To manually curate the encoding factors, functional domain searches were conducted through domain scanning (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Those candidates without MADS domain structure were excluded. The protein coding genes were manually checked and the intron and exon information were extracted from the genome database.

Phylogenetic Analysis

The protein sequences of all MADS-box genes were aligned with clustalW implemented in bioedit software (v7.2.6). Maximum likelihood tree of each gene was inferred using IQTREE (v1.5.5) with automated substitution model and 1000 regular bootstrap validation.

Cloning of Strawberry MADS8 Gene

On the basis of the existing strawberry genome annotation, MADS8 was selected as a new finding. The existence of this new factor was validated by RT-PCR. According to the corresponding amino acid sequence of MADS in the genome, we designed two primers MADS8For: ATG GGG AAG AAA TCA CAA AAC ACC G, MADS8Rev: TTA ATT TCT TCC AAT ATC TCC TAT TGA TTT GTT GGA ACC. The CTAB method was used to extract the RNA of the cultivated strawberry 'Toyonaka' with flower as materials. PCR amplicons were gel-purified and cloned into pEASY-blunt vector. Positive clones were subjected to sequencing confirmation.

Results

Members of the MADS gene family

Eighty-five MADS genes of *Fragaria vesca* were identified through the two methods (Table 1). They were distributed evenly on the seven pairs of chromosomes, with seven on the first chromosome, ten on the second chromosome, eight on the third. The sixth chromosome have the largest number, 24 MADS encoding genes. Five new MADS gene sites were identified compared with the results of genomic annotation. Eight of the 85 sites, although containing the MADS domain, were not identifiable due to the inclusion of gaps. The average intron length of the 74 MADS encoding genes that were identified was 2.6, which can be as high as 10 (mRNA31290). Among the proteins with confidence, the longest CDS is 1921 bp (mRNA13277).

Table 1. MADS-box members identified in *Fragaria vesca* genome

Gene name	Gene ID ^z	Chromosome	Intron No.	Extron No.	Amino Acids No.	e-value	DNA amplification	mRNA length ^y	expected other domain
FvMADS1	mrna26429	LG1	1	2	66	8.13E-38	289	201	
FvMADS2	mrna16491	LG1	0	1	165	7.04E-21	498	498	
FvMADS3	mrna14896	LG1	4	5	232	1.26E-35	1761	699	k-box
FvMADS4	mrna05564	LG1	0	1	203	1.36E-02	612	612	k-box
FvMADS5	mrna31068	LG1	0	1	211	4.91E-25	636	636	k-box
FvMADS6	mrna31594	LG1	2	3	167	2.61E-28	856	504	
FvMADS7	mrna16489	LG1	0	1	186	3.98E-19	561	561	
FvMADS8	NEW	LG2	6	7	232	2.09E-38	2256	699	k-box
FvMADS9	mrna11233	LG2	0	1	314	1.31E-23	945	945	
FvMADS10	mrna01788	LG2	0	1	231	1.74E-33	696	696	
FvMADS11	mrna11267	LG2	6	7	201	4.09E-35	2417	606	K-box
FvMADS12	mrna15090	LG2	0	1	224	1.55E-38	675	675	
FvMADS13	mrna01789	LG2	0	1	220	2.50E-34	663	663	
FvMADS14	mrna24978	LG2	3	4	468	1.73E-16	3524	1407	bZIP
FvMADS15	mrna11244	LG2	2	3	507	1.81E-17	3263	1524	
FvMADS16	mrna11268	LG2	3	4	142	2.16E-36	2090	429	
FvMADS17	mrna27834*	LG2	0	1	359	2.23E-08	1080	1080	
FvMADS18	mrna14058	LG3	0	1	184	2.97E-35	555	555	
FvMADS19	mrna25070	LG3	6	7	230	8.82E-41	4461	693	K-box
FvMADS20	mrna24852	LG3	6	7	250	1.56E-44	4799	753	K-box
FvMADS21	mrna20134	LG3	8	9	276	4.70E-31	6161	839(CT-AC)	K-box
FvMADS22	mrna20015	LG3	5	6	254	1.67E-39	1816	770(CT-AC)	K-box
FvMADS23	mrna28641	LG3	0	1	606	4.18E-07	1821	1821	
FvMADS24	mrna30741	LG3	0	1	222	1.06E-31	669	669	K-box
FvMADS25	mrna30567	LG3	0	1	182	5.20E-35	549	549	
FvMADS26	mrna00615	LG4	0	1	227	2.35E-35	684	684	
FvMADS27	mrna03748	LG4	5	6	224	1.80E-30	5260	675	K-box
FvMADS28	mrna04563	LG4	6	7	213	4.15E-43	2935	642	K-box
FvMADS29	mrna03714	LG4	7	8	224	5.33E-30	5318	675	K-box
FvMADS30	mrna04562	LG4	7	8	245	3.92E-40	2931	738	K-box
FvMADS31	mrna05180	LG4	2	3	98	9.22E-25	1571	297	
FvMADS32	mrna16933	LG4	3	4	442	5.42E-06	4009	1329	RuBisCO_small
FvMADS33	mrna17037	LG4	0	1	174	6.69E-07	525	525	
FvMADS34	mrna27105	LG5	5	6	196	1.65E-40	2262	591	K-box
FvMADS35	mrna31290	LG5	10	11	350	6.32E-36	3366	1053	K-box
FvMADS36	NEW	LG5	3	7	225	5.64E-38	3661	678	
FvMADS37	mrna32155	LG5	0	1	224	1.72E-18	675	675	
FvMADS38	mrna26118	LG5	6	7	209	1.29E-26	6371	636(CT-AC)	K-box
FvMADS39	mrna11073	LG5	1	2	66	2.55E-15	319	201	
FvMADS40	mrna12119	LG5	7	8	235	1.11E-33	10797	708 (N included)	K-box
FvMADS41	mrna12189	LG5	2	3	227	8.70E-09	963	684	
FvMADS42	mrna25669	LG5	0	1	225	5.21E-29	678	678	K-box
FvMADS43	mrna29294	LG5	5	6	499	3.53E-08	6982	1500 (N included)	
FvMADS44	mrna06301	LG5	4	5	157	2.74E-32	1185	474	K-box
FvMADS45	mrna24650	LG5	0	1	150	3.01E-13	453	453	
FvMADS46	mrna26028	LG5	14	15	700	1.24E-26	7302	2103 (N included)	LpIA
FvMADS47	mrna22696	LG5	5	6	264	4.05E-33	4545	795	K-box

Table 1, cont. MADS-box members identified in *Fragaria vesca* genome

FvMADS48	mrna26119	LG5	7	8	284	1.44E-42	3546	862(CT-AC)	K-box
FvMADS49	mrna12120	LG5	7	8	235	6.26E-32	7012	708	K-box
FvMADS50	mrna10739	LG5	0	1	244	1.96E-20	735	735	
FvMADS51	mrna25871	LG6	7	8	431	5.59E-16	4775	1296	EnY2
FvMADS52	mrna23924	LG6	1	2	451	7.44E-31	2692	1356	
FvMADS53	mrna04228	LG6	7	8	257	2.66E-42	4698	774	K-box
FvMADS54	mrna05411	LG6	6	7	528	5.84E-22	4168	1587	CIA30
FvMADS55	mrna15899	LG6	4	5	296	1.25E-27	6411	891	
FvMADS56	mrna18094	LG6	0	1	165	1.61E-27	498	498	
FvMADS57	mrna02453	LG6	3	4	176	2.73E-14	3025	531 (N included)	
FvMADS58	mrna34182	LG6	1	2	54	3.57E-07	404	165	
FvMADS59	mrna14413	LG6	0	1	225	1.46E-08	678	678	
FvMADS60	mrna15394	LG6	4	5	377	1.94E-04	2249	1134	
FvMADS61	NEW	LG6	0	1	223	6.91E-31	672	672	Glyco_tranf_GTA
FvMADS62	NEW	LG6	0	1	219	6.00E-33	660	660	
FvMADS63	mrna28409	LG6	0	1	216	2.18E-12	651	651	
FvMADS64	mrna24494	LG6	7	8	262	3.28E-44	8874	789	K-box
FvMADS65	mrna22916	LG6	0	1	278	2.26E-19	837	837	
FvMADS66	mrna22967	LG6	0	1	291	4.54E-22	876	876	
FvMADS67	mrna14402	LG6	0	1	204	5.17E-10	615	615	
FvMADS68	mrna05451	LG6	0	1	233	2.23E-26	702	702	
FvMADS69	mrna01336	LG6	1	2	125	6.74E-36	2170	378	
FvMADS70	mrna04229	LG6	7	8	249	7.70E-45	3550	750	K-box
FvMADS71	mrna15563	LG6	0	1	198	1.44E-15	597	597	
FvMADS72	mrna18029	LG6	0	1	186	1.39E-20	561	561	
FvMADS73	mrna17829	LG6	2	3	325	3.74E-25	2056	978 (N included)	
FvMADS74	mrna04269	LG6	0	1	426	2.31E-10	1281	1281	
FvMADS75	mrna29503	LG7	8	9	309	3.56E-35	3918	930	K-box
FvMADS76	mrna13354	LG7	6	7	214	6.18E-39	3672	645	K-box
FvMADS77	NEW	LG7	6	7	215	1.88E-41	5115	648 (N included)	K-box
FvMADS78	mrna21477	LG7	0	1	153	1.15E-08	462	462	
FvMADS79	mrna04949	LG7	0	1	169	6.99E-28	510	510	
FvMADS80	mrna13356	LG7	23	24	929	5.67E-37	14596	2790 (NC)	K-box
FvMADS81	mrna19428	LG7	9	10	465	1.15E-43	5838	1398 (NC)	K-box
FvMADS82	mrna13358	LG7	1	2	155	6.77E-29	1836	468 (N included)	
FvMADS83	mrna13277	LG7	6	7	646	6.04E-19	4680	1941	NT-C2, SYCE1
FvMADS84	mrna18483	LG7	24	25	863	3.84E-23	9245	2592 (NC)	
FvMADS85	mrna05027	LG4	4	5	274	1.94E-40	3832	825 (N included)	

^z Inferred as confirmed by RNA-seq;

^y none 'GT-AG' intron boundary was marked out in parentheses; NC, not confirmed due to annotation errors in the genome. Sequences with gaps in the genome were marked as N including.

Phylogeny of MADS-box Proteins in *Fragaria Vesca*

The amino acids of MADS-box proteins identified were aligned using clustalW. The ML tree (Fig. 1) was reconstructed with JTT as the most suitable amino acid substitution model in IQTREE. As can be seen from the figure, the 74 members with confidence (no gap and no annotation error) of the identified sequence can be divided into two broad categories (TypeI and TypeII). The former contains 20 members and the latter 54 members. Among them, Type II can be further divided into five subgroups, two of which contain more than 20 members (arrow). It can be recognized from the figure that members with other domain structures other than MADS, are mainly concentrated in the second type, such as k-box containing factors. It is noteworthy that only one member (MADS14) has its own special b-zip domain, indicating its special function.

Verification of the Newly Found MADS8 Gene in the Genome

To verify the accuracy of the above analysis, MADS8, a member that was not annotated in the strawberry genome was selected as one representative to be verified. Specific primers targeting LG2:22698117-22700372 were used to amplify the expected coding sequences. The amplification product was subjected to gel analysis (Fig. 2) and

subsequent sequencing confirmation. The abtained results were identical with the sequence of analysis.

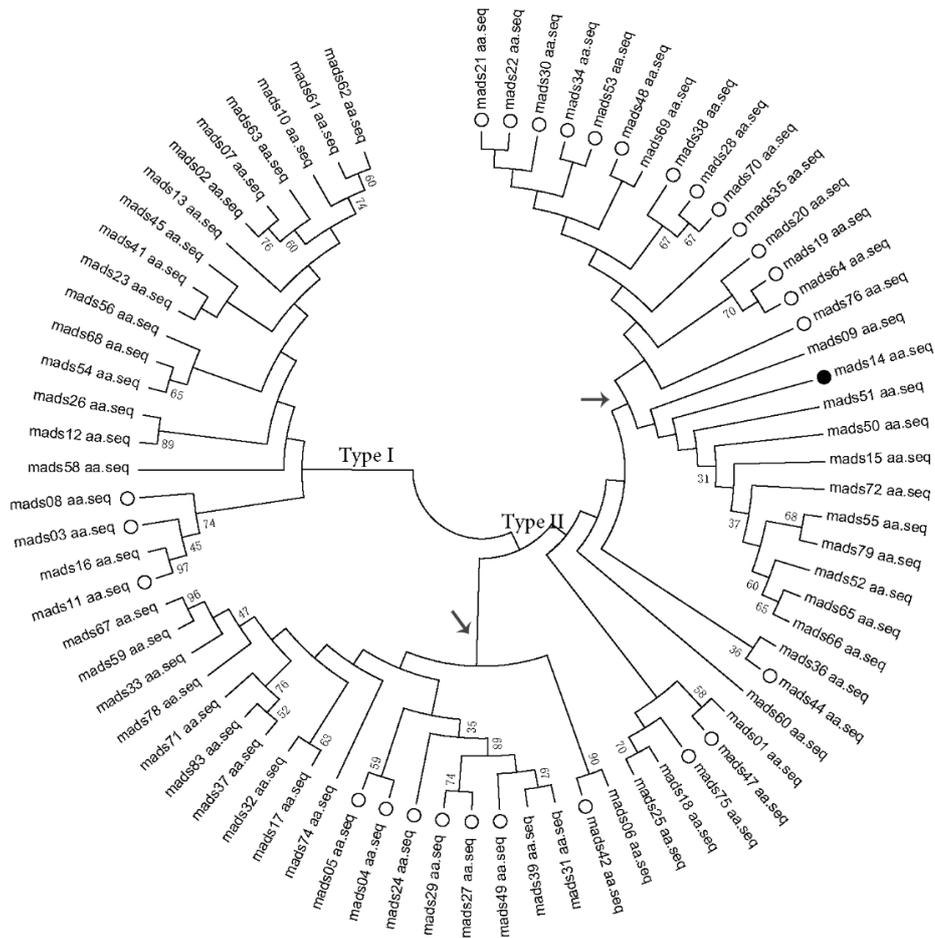


Figure 1. Phylogenetic tree based on maximum likelihood method for *Fragaria vesca* MADS members. Arrows showed two large subgroups in type II, open circles represent members that contain the K-box domain, and solid circles are members that contain bZIP.

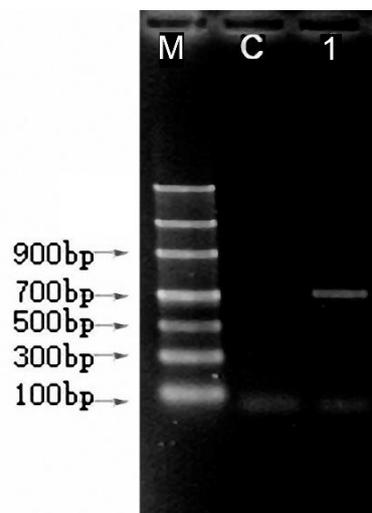


Figure 2. PCR amplification of FvMADS8. M, molecular marker; C, Negative control; 1, FvMADS8

Discussion

Quantities of MADS-box encoding genes have been identified in model plants [4]. In the *Fragaria vesca* genome, Zhou et al. [7] found 70 members. In this study, we identified 85 members in the genome. The main reason might be due to the strategies used. Moreover, we excluded some uncertain candidates. New members were reported here too, one of which has been validated by cloning. Overall, our results provide basic information for MADS-box genes in the model *Fragaria vesca* genome sequence.

References

- [1] P. Shore, A. D. Sharrocks, The MADS-Box family of transcription factors, *Europ J. Biochem.* 1995, 229(1): 1-13.
- [2] G. Theissen, A. Becker, A.D. Rosa, A. Kanno, J.T. Kim, T. Munster, K.U. Winter, H. Saedler, A short history of MADS-box genes in plants, *Plant Mol. Biol.* 2000, 42(1), 115.
- [3] S. Ferrario, R.G. Immink, G.C. Angenent, Conservation and diversity in flower land, *Curr. Opin. Plant Biol.* 2004, 7(1), 84-91.
- [4] S. de Folter, R.G. Immink, M. Kieffer, L. Parenicová, S.R. Henz, D. Weigel, M. Busscher, M. Kooiker, L. Colombo, M.M. Kater, B. Davies, G.C. Angenent, Comprehensive interaction map of the Arabidopsis MADS box transcription factors, *Plant Cell* 2005, 17(5), 1424-1433.
- [5] M. Ng, M.F. Yanofsky, Function and evolution of the plant MADS-box gene family, *Nat. Rev. Genet.* 2001, 2(3), 186-195.
- [6] V. Shulaev, D.J. Sargent, R.N. Crowhurst, T.C. Mockler, O. Folkerts, A.L. Delcher, P. Jaiswal, K. Mockaitis, A. Liston, S.P. Mane, P. Burns, T.M. Davis, J.P. Slovin, N. Bassil, R.P. Hellens, C. Evans, T. Harkins, C. Kodira, B. Desany, O.R. Crasta, R.V. Jensen, A.C. Allan, T.P. Michael, J.C. Setubal, J.M. Celton, D.J. Rees, K.P. Williams, S.H. Holt, J.J. Ruiz Rojas, M. Chatterjee, B. Liu, H. Silva, L. Meisel, A. Adato, S.A. Filichkin, M. Troggio, R. Viola, T.L. Ashman, H. Wang, P. Dharmawardhana, J. Elser, R. Raja, H.D. Priest, D.W. Jr Bryant, S.E. Fox, S.A. Givan, L.J. Wilhelm, S. Naithani, A. Christoffels, D.Y. Salama, J. Carter, E. Lopez Girona, A. Zdepski, W. Wang, R.A. Kerstetter, W. Schwab, S.S. Korban, J. Davik, A. Monfort, B. Denoyes-Rothan, P. Arus, R. Mittler, B. Flinn, A. Aharoni, J.L. Bennetzen, S.L. Salzberg, A.W. Dickerman, R. Velasco, M. Borodovsky, R.E. Veilleux, K.M. Folta, The genome of woodland strawberry (*Fragaria vesca*), *Nat. Genet.* 2011, 43(2), 109-116.
- [7] H.C. Zhou, G. Li, X. Zhao, Z.R. Kang, A.G. Guo, Genome-wide sequence identification of MADS-box transcription factor gene family from *Fragaria vesca* and cloning of FaMADS1 gene from *F. ananassa* fruit. *Plant Physiol. J.* 2014, 50(11), 1630-1638
- [8] L. Gramzow, G. Theißen, Phylogenomics of MADS-Box genes in plants - two opposing life styles in one gene family, *Biology* 2013, 2(3), 1150-1164.