IMAM KHOMEINI INTERNATIONAL UNIVERSITY
IJGPB
Iranian Journal of Genetics and Plant Breeding
Print ISSN : 2251-9610
Online ISSN : 2676-346X

In-silico identification of SSRs based on RNA-sequencing data from fully open flowers of a frost-tolerant almond under frost stress

Batool Hosseinpour[1*], Mohammad Reza Bakhtiarizadeh[2], Hamed Yousefzadeh[3], Hamid Bina[3]

[1]Department of Agriculture, Iranian Research Organization for Science and Technology (IROST), Institute of Agriculture, Tehran, Iran.
[2]Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Tehran, Iran.
[3]Department of Environment, Faculty of Natural Resources, Tarbiat Modares University, Tehran, Iran.
[*]Corresponding author, 0000-0003-0552-6696. Email: hosseinpour@irost.ir. Tel: +98-21-56276639.

ABSTRACT INFO

Research Paper

Received: 18 Sep 2022

Accepted: 18 Dec 2022

ABSTRACT

Spring frost injury is a major environmental limit for the productivity of stone fruits such as almond. Thus, it is important to screen and breed spring frost tolerant trees. Among molecular markers, SSR derived from EST sequences are more desirable for breeding programs. In our previous study, we employed a frost treatment for flowers of a frost-tolerant almond genotype at full blooming stage as the important phenological stage to study frost stress. The transcriptomes of the frost-treated and non-frost-treated flowers (after post-thaw period) were analyzed and compared by using next generation sequencing technology. A total of 63218 sequences were generated and analyzed using SSR-Locator. In total, 2680 SSRs were identified and 2601 of sequences contained SSRs. In 11 sequences, there was more than one EST-SSRs. Among the potential EST-SSRs, five types of the motifs were identified; di-nucleotide and tri-nucleotide had the highest frequencies followed by penta-nucleotide, hexa-nucleotide and tetra-nucleotide repeats. Our study revealed that di- and tri-nucleotide motifs are more frequent. TA/TA and GA/TC were the most abundant di-nucleotide motifs while GAA, AAC and ACC were common in tri-nucleotide SSRs. GO annotation showed that the top GO terms associated with SSRs containing tri-nucleotide motif were involved in the regulation of transcription, regulation of cellular biosynthetic processes, regulation of gene expression, regulation of cellular processes, response to stress and response to stimulus. Our study is the first report on SSR development for fully open flowers of a frost-tolerant almond genotype.

Key words: Almond, Frost, Fully open flowers, SSR, RNA-seq.

How to cite this article:
Hosseinpour B., Bakhtiarizadeh M. R., Yousefzadeh H., and Bina H. (2021). *In-silico* identification of SSRs based on RNA-sequencing data from fully open flowers of a frost-tolerant almond under frost stress. *Iranian Journal of Genetics and Plant Breeding,* 10(1): 101-106.
DOI: 10.30479/IJGPB.2022.17816.1329

INTRODUCTION

Almond (*Prunus amygdalus* Batsch) is one of the world's most important nut crop. Almond belongs to the genus *Prunus*, subgenus *Amygdalus*. It is an early blooming tree and flower emergence inevitably coincides with spring frost, which damages flowers. Thus, almond breeders aim to create and select late-blooming cultivars (Socias *et al.,* 1997). Nevertheless, this solution is not enough to overcome spring frost damage because the late-blooming trait is usually accompanied by undesirable traits, and late-blooming varieties cannot be cultivated in all areas. Therefore, frost hardiness is considered a selection objective in breeding programs (Kester and Griggs, 1959). Frost-tolerant genotypes can be found among trees adapted to harsh weather conditions. Frost damage is highly dependent on the phenological stage of the bud/flower/fruit, with the early developing fruit being the most susceptible stage (Proebsting and Mills, 1978). In almond, phenological stages of flowers are defined from A to L by Felipe (Felipe, 1977). Winter buds (flower buds at stage A) are the most tolerant ones compared to flowers at later stages. Almond trees are more affected by spring frost at the full bloom stage (or stage F) rather than at the popcorn (or stage D) stage in flowers (Miranda *et al.,* 2005). The level of tolerance to frost varies among the buds of different cultivars and is higher among different species in some cases (Pfammatter and Evequoz, 1975).

To promote crop improvement, it is crucial to be aware of the genetic diversity and relatedness in the breeding material. Therefore, molecular characterization has become the most important tool for evaluating genetic diversity and similarity within and between populations or accessions, mapping of functional genes, assembly of genetic linkage maps, marker-assisted selection (MAS), and phylogenetic experiments in crop species (Arús *et al.,* 2010).

RNA-Seq, which is based on next generation sequencing, is a high throughput technology that has great advantages in examining the fine structure of a transcriptome. When no genome sequence is available, transcriptome sequencing provides an effective way to obtain large amounts of sequence data. RNA-Seq has been widely used in many organisms to obtain mass sequence data for transcriptional analysis (Liu *et al.,* 2022; Singh *et al.,* 2022), gene discovery and molecular marker development (Baytar *et al.,* 2022; Shu *et al.,* 2021). The genetic relationships and diversity among almond germplasm collections have been investigated mostly using simple sequence repeat (SSR) markers (Xu *et al.,* 2004; Xie *et al.,* 2006; Shiran *et al.,* 2007). Compared with other types of molecular markers, SSR markers have many advantages, such as simplicity, effectiveness, abundance, hypervariability, reproducibility, codominant inheritance, and extensive genomic coverage (Powell *et al.,* 1996). But no SSR markers have been reported related to frost tolerance at full blooming stage of flowering in almond. In our previous study, comprehensive expression profiling of genes expressed in fully open flowers was performed after being exposed to frost temperatures (during post-thaw period) by using RNA-seq technology (Hosseinpour *et al.,* 2018). Here, we used the generated RNA-seq data to develop expressed sequenced tag SSRs (EST-SSRs) related to frost tolerance of fully open flowers of almond. These results provide a very useful genomic data for breeding of almond in future.

MATERIALS AND METHODS

EST-SSR mining from RNA-seq data

In the previous research, we provided the molecular signature of pistils of fully open flowers from a frost-tolerant almond genotype, 6-8, and an RNA-seq experiment was conducted using Illumina HiSeq2000 platform (Hosseinpour *et al.,* 2018). In brief, full bloom flowers of 6–8 almond genotypes were exposed to 4 °C/6 h and −3 °C/30 min followed by 21 °C/24 h. Uninjured pistils were used for further physiological and gene expression analysis. After RNA sequencing, the low quality reads were filtered out and adaptor sequences were trimmed using FASTQC software. Clean reads, 27,104,070 and 32,730,772, were obtained for non-frost-treated (NT) and frost-treated (FT) libraries, respectively. Then, clean sequences were assembled using the short read assembly program Trinity. The assembly results from Trinity were passed to CD-HIT (Fu *et al.,* 2012) and CAP3 (Huang and Madan, 1999) for multiple alignments and consensus building in order to reduce redundancy. Thereafter, the final reference transcripts were created, and the longest transcript in a clustering unit was selected as the unigene. A total of 62.24 Mb was assembled, generating 50,896 unigenes and 66,906 transcripts.

Six types of microsatellites were investigated using SSR locator. The parameters were set as follows: the SSRs were considered to contain di-, tri-, tetra-, penta- and hexa-nucleotides with minimum repeat numbers of 10, 6, 5, 5, 5, and 5, respectively.

Statistical analysis and gene ontology annotation of SSRs

SSR motifs were compared using IDEG6. Gene

ontology terms were assigned to SSR containing tri- and hexa-nuleotides and visualized by online software WEGO to understand the distribution of the gene functions.

Primer design

Unigenes with a sequence of more than 150 bp before and after the SSR region were used for primer design by Primer v3.0. The major parameters for primer design were set as follows: primer length of 18-25 bases (optimal 21 bases), PCR product size of 80-200 bp (optimal 100-150 bp), GC content of 40-60% (optimal 50%), and DNA melting temperature of 57-64 °C (optimal annealing temperature 55-59 °C).

RESULTS

The frequency and distribution of SSRs

A total of 63218 sequences were examined. The total size of sequences was 58392338 bp. In total, 2680 SSRs were identified and 2601 of sequences contained SSRs. Among the identified SSRs, 11 were present in compound formation, while others were of perfect one –repeat type. In 11 sequences, there were more than one EST-SSRs. Among the potential EST-SSRs, five types of the motifs were identified; di-nucleotide and tri-nucleotide had the highest frequencies followed by penta-nucleotide, hexa-nucleotide and tetra-nucleotide repeats (Table 1).

The number of repeat unit within SSR loci

The SSR was also analyzed according to the different number of repeat unit or repeat length within SSR locus (Table 2). The repeat length of 1188 of di-nucleotide SSRs was 10. Most tri-nucleotide SSRs had a repeat length of 7. The longer motifs, penta- and hexa-nucleotide showed a dramatic reduction in frequency as the number of repeats was increased.

Motif types

The motifs could be divided into 33 types, which demonstrated nonrandom distribution of SSRs in almond genome (Table 3). The frequency of motifs varied greatly even within the same-unit-sized SSRs, for example, GA and TA were common in di-nucleotide SSRs while GAA, AAC and ACC were common in tri-nucleotide SSRs.

Gene ontology classification

GO annotation showed that the top GO terms associated with SSRs containing tri-nucleotide motif were involved in the regulation of transcription, regulation of cellular biosynthetic processes, regulation of gene expression, regulation of cellular processes, response to stress and response to stimulus (Supplementary 1).

Table 1. Summary of RNA-sequence for SSRs in almond [*Prunus dulcis* (Mill.)].

Parameters used in screening	Generated data by SSR Locator
Total number of sequences examined	63218
Total size of examined sequences (bp)	58392338
Total number of identified SSRs	2680
Number of SSR containing sequences	2601
Number of sequences containing more than 1 SSR	11
Number of SSRs present in compound formation	11
Di-nucleotide	1483
Tri-nucleotide	496
Tetra-nucleotide	172
Penta-nucleotide	275
Hexa-nucleotide	254

Table 2. The SSR number based on the number of repeat unit.

Number of repeat unit	Di-nucleotide	Tri- nucleotide	Tetra- nucleotide	Penta- nucleotide	Hexa- nucleotide
4	0	0	0	254	230
5	0	0	133	16	11
6	0	0	37	3	11
7	0	458	1	2	1
8	0	33	0	0	0
9	0	4	0	0	0
10	1188	1	0	0	0
11	283	0	0	0	0
12	10	0	1	0	0
13	1	0	0	0	0
14	0	0	0	0	0
≥15	1	0	0	0	1

Table 3. The frequency of different motif types.

Motif type	Count	Motif type	Count	Motif type	Count
GA/TC	580	CAG/CTG	2	TAA/TTA	2
AG/CT	119	AAC/GTT	64	TAAA/TTTA	1
AT/AT	41	CAA/TTG	15	GCA/TGC	2
TA/TA	589	CTC/GAG	21	TAAAA/TTTTA	22
GAA/TTC	97	AGG/CCT	6	ACC/GGT	67
AGA/TCT	22	TCA/TGA	33	AAAAC/GTTTT	4
AC/GT	40	GGA/TCC	30	AAGAA/TTCTT	12
CA/TG	15	AAAAT/ATTTT	22	CCA/TGG	19
AAG/CTT	5	AAT/ATT	7	AGAGGA/TCCTCT	1
ATC/GAT	11	AGC/GCT	27	CAACC/GGTTG	15
ATG/CAT	1	GAAAA/TTTTC	10	GGCA/TGCC	20

For SSRs containing hexa-nucleotide motif, the top GO terms were involved in the regulation of primary metabolic process, regulation of metabolic processes, regulation of macromolecule metabolic process, regulation of cellular biosynthetic process and regulation of biological process (Supplementary 1).

DISCUSSION

SSR markers are codominant, easily assayed, highly repeatable with a high polymorphism information content (PIC). Also, most SSR loci are defined by a unique pair of primers (Chen *et al.,* 2006).

The limiting factor of SSRs was the development cost in the past, but they are valuable markers for genome mapping and population studies (Ma *et al.,* 2020; Wang *et al.,* 2020). In recent years, high throughput sequencing technologies have provided a good opportunity to develop thousands of SSR candidates at one time. Great advances have been made by SSRs in breeding of many plants (Roshan *et al.,* 2021; Song *et al.,* 2021) but a few SSR markers were validated in almond and applied in practice. Frost tolerance of flowers at full blooming stage is a major breeding aim in almond. Up to now, there is no report on SSR candidates for this trait. Here, we analyzed RNA-seq output data from a frost-tolerant genotype and developed SSR candidates. Our study revealed that di- and tri-nucleotide motifs are more frequent. TA/TA and GA/TC were the most abundant di-nucleotide motifs. However, GA/TC and AG/CT were the most abundant di-nucleotide motifs in almond flowers at popcorn stage (Alisoltani *et al.,* 2016), suggesting different mechanisms of cold tolerance for this stage. An overall predominance of AG/CT and AT/AT dimer motifs in EST sequences was reported by previous studies in angiosperms (Ranade *et al.,* 2014), almond and peach (Xu *et al.,* 2004). In accordance with our results, AT/AT was the most abundant di-nucleotide motif in the cold-stressed transcriptome of *Notopterygium incisum* (Jia *et al.,* 2019). In the cold-stressed transcriptome of centipedegrass, AG/CT, AC/GT and AT/AT were the most abundant di-nucleotide motif (Wang *et al.,* 2017). In a recent study, AT/AT was identified as the most dominant di-nucleotide in the cold-stresses transcriptome of *Pinus koraiensis* and in accordance with our study, GAA/TTC was also a significantly frequent tri-nucleotide (Li *et al.,* 2020). However, in contrast to our study, AAC/GTT and ACC/GGT were more frequent in *Pinus koraiensis.*

High frequency of GAA/TTC in our study was in accordance with the previous study in almond flowers at popcorn stage (Alisoltani *et al.,* 2016). Also, AAC/GTT and ACC/GGT were previously predicted in non-redundant EST sequences of almond and peach (Xu *et al.,* 2004). A/T, AG/CT and AAG/CTT were the most frequent motifs in EST sequences of lentil (*Lens culinaris*) under cold stress (Sohrabi *et al.,* 2018). In the cold- and drought stressed transcriptome of *Ammopiptanthus mongolicus*, the AG/CT was the most frequent repeat motif and accounted for 10.3% (16/155), followed by AAG/CTT (7.7%, 12/155), AGA/TCT (7.7%, 12/155), GAA/TTC (6.4%, 10/155) and AG/CT (5.8%, 9/155) (Liu *et al.,* 2013).

For the current study, we provided a number of EST-SSR markers for fully open flowers in almond under frost stress condition, based on a *de novo* RNA-Seq assembly. To the best of our knowledge, this is the first attempt to develop SSR markers at this phenological stage. In the light of the above and previous results, generation and validation of EST-SSRs and SNPs from the transcriptome of the flowers at this stage from

various genotypes are efficient for molecular breeding programs.

ACKNOWLEDGMENTS

Not applicable.

REFERENCES

Alisoltani A., Ebrahimi S., Azarian S., Hematyar M., Shiran B., Jahanbazi H., Fallahi H., Mousavi-Fard S., and Rafiei F. (2016). Parallel consideration of SSRs and differentially expressed genes under abiotic stress for targeted development of functional markers in almond and related Prunus species. *Scientia Horticulturae,* 198: 462-472.

Arús P., Yamamoto T., Dirlewanger E., and Abbott A. G. (2010). Synteny in the Rosaceae. *Plant Breeding Reviews,* 27: 175-211.

Baytar A. A., Peynircioğlu C., Sezener V., Frary A., and Doğanlar S. (2022). Association analysis of germination level cold stress tolerance and candidate gene identification in Upland cotton (*Gossypium hirsutum* L.). *Physiology and Molecular Biology of Plants,* 28: 1049-1060. DOI: https://doi.org/10.1007/s12298-022-01184-6.

Chen C., Zhou P., Choi Y. A., Huang S., and Gmitter F. G. (2006). Mining and characterizing microsatellites from citrus ESTs. *Theoretical and Applied Genetics,* 112: 1248-1257.

Felipe A. (1977). Phenological states of almond. Proceedings of the Third GREMPA Colloquium, Bari, Italy, 3-7.

Fu L., Niu B., Zhu Z., Wu S., and Li W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics,* 28: 3150-3152.

Hosseinpour B., Sepahvand S., Kamali Aliabad K., Bakhtiarizadeh M., Imani A., Assareh R., and Salami S. A. (2018). Transcriptome profiling of fully open flowers in a frost-tolerant almond genotype in response to freezing stress. *Molecular Genetics and Genomics,* 293: 151-163.

Huang X., and Madan A. (1999). CAP3: A DNA sequence assembly program. *Genome Research,* 9: 868-877.

Jia Y., Bai J., Liu M., Jiang Z., Wu Y., Fang M., and Li Z. (2019). Transcriptome analysis of the endangered *Notopterygium incisum*: Cold-tolerance gene discovery and identification of EST-SSR and SNP markers. *Plant Diversity,* 41: 1-6.

Kester D., and Griggs W. (1959). Fruit setting in the almond: The effect of cross-pollinating various percentages of flowers and the pattern of flower and fruit drop. *Proceedings of The American Society For Horticultural Science,* 74: 206-213.

Li X., Liu X., Wei J., Li Y., Tigabu M., and Zhao X. (2020). Development and transferability of EST-SSR markers for *Pinus koraiensis* from cold-stressed transcriptome through Illumina sequencing. *Genes,* 11: 500.

Liu M., Shi J., and Lu C. (2013). Identification of stress-responsive genes in *Ammopiptanthus mongolicus* using ESTs generated from cold-and drought-stressed seedlings. *BMC Plant Biology,* 13: 1-14.

Liu N., Ren Z., Ren Q., Chang Z., Li J., Li X., Sun Z., He J., Niu Q., and Xing X. (2022). Full length transcriptomes analysis of cold-resistance of *Apis cerana* in Changbai Mountain during overwintering period. *Gene,* 830: 146-503.

Ma J., Gao B., Wang R., Li X., and Chen S. (2020). Transcriptome analyses of *Ditylenchus destructor* in responses to cold and desiccation stress. *Genetics and Molecular Biology,* 43(1). DOI: https://doi.org/10.1590/1678-4685-GMB-2018-0057.

Miranda C., Santesteban L. G., and Royo J. B. (2005). Variability in the relationship between frost temperature and injury level for some cultivated *Prunus* species. *HortScience,* 40: 357-361.

Pfammatter W., and Evequoz M. (1975). Resistance aux gelees de differentes cultures. *Revue Suisse de Viticulture, Arboriculture, Horticulture,* 19-23.

Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S., and Rafalski A. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding,* 2: 225-238.

Proebsting E., and Mills H. (1978). Low temperature resistance of developing flower buds of six deciduous fruit species. *Journal of the American Society for Horticultural Science,* 103: 192-198.

Ranade S. S., Lin Y.-C., Zuccolo A., Van de Peer Y., and García-Gil M. D. R. (2014). Comparative insilico analysis of EST-SSRs in angiosperm and gymnosperm tree genera. *BMC Plant Biology,* 14: 1-10.

Roshan N. M., Ashouri M., and Sadeghi S. M. (2021). Identification, evolution, expression analysis of phospholipase D (PLD) gene family in tea (Camellia sinensis). *Physiology and Molecular Biology of Plants,* 27: 1219-1232.

Shiran B., Amirbakhtiar N., Kiani S., Mohammadi S., Sayed-Tabatabaei B., and Moradi H. (2007). Molecular characterization and genetic relationship among almond cultivars assessed by RAPD and SSR markers. *Scientia Horticulturae,* 111: 280-292.

Shu H., Li K., Ou Y., Zhan R., and Chang S. (2021). Identification cold tolerance of pineapple germplasms at seedling stage. *American Journal of Plant Sciences,* 12: 1768-1779.

Singh B. K., Sutradhar M., Lahkar C., Singh A. K., Meetei N. T., and Mandal N. (2022). Screening of rice germplasms (*Oryza sativa* L.) for seedling stage cold tolerance utilizing morphological and molecular markers. *Ecological Genetics and Genomics,* 24: 100128. DOI: https://doi.org/10.1016/j.egg.2022.100128.

Socias R., Felipe A. J., Gómez Aparisi J., García J. E., and Dicenta F. (1997). The ideotype concept in almond. In II International Symposium on Pistachios and Almonds, 470: 51-56.

Sohrabi S. S., Ismaili A., Firouz-Abadi F. N., and Fallahi H. (2018). Discovery of EST-SSRs markers in lentil (Lens culinaris) under cold stress. *Crop Biotechnology,* 8: 1-14.

Song X., Li N., Guo Y., Bai Y., Wu T., Yu T., Feng S., Zhang Y., Wang Z., and Liu Z. (2021). Comprehensive identification and characterization of simple sequence repeats based on the whole-genome sequences of 14 forest and fruit trees. *Forestry Research,* 1: 1-10.

Wang P., Yang L., Zhang E., Qin Z., Wang H., Liao Y., Wang X., and Gao L. (2017). Characterization and development of EST-SSR markers from a cold-stressed transcriptome of centipedegrass by illumina paired-end sequencing. *Plant Molecular Biology Reporter,* 35: 215-223.

Wang Z., Wang S., Yu C., Han X., and Zou D. (2020). QTL analysis of rice photosynthesis-related traits under the cold stress across multi-environments. *Euphytica,* 216: 1-21.

Xie H., Sui Y., Chang F.-Q., Xu Y., and Ma R.-C. (2006). SSR allelic variation in almond (*Prunus dulcis* Mill.). *Theoretical and Applied Genetics,* 112: 366-372.

Xu Y., Ma R.-C., Xie H., Liu J.-T., and Cao M.-Q. (2004). Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. *Genome,* 47: 1091-1104.

SUPPLEMENTAL DATA

Supplementary 1. Please click.