Bulletin of Biological and Allied Sciences Research
***ISSN:* 2521-0092**
www.bbasr.org
DOI: https://doi.org/10.54112/bbasr.v2023i1.38
*Bull. Biol. All. Sci. Res., Volume, **8**: 38*

Original Research Article

GENOME-WIDE BIOINFORMATICS ANALYSIS OF 1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE (*ACS*), 1-AMINOCYCLOPROPANE-1-CARBOXYLATE OXIDASE (*ACO*) AND ETHYLENE OVERPRODUCER 1 (*ETO1*) GENE FAMILY OF *FRAGARIA VESCA* (WOODLAND STRAWBERRY)

FATIMA S[1], CHEEMA KUHT[1], SHAFIQ M[1*], MANZOOR MT[2], ALI Q[3], HAIDER MS[3], SHAHID MA[4*]

[1]*Department of Horticulture, Faculty of Agricultural Sciences, University of the Punjab Lahore, Pakistan*
[2]*Department of Plant Pathology, Faculty of Agricultural Sciences, University of the Punjab Lahore, Pakistan*
[3]*Department of Plant Breeding and Genetics, Faculty of Agricultural Sciences, University of the Punjab Lahore, Pakistan*
[4]*Horticultural Science Department, North Florida Research and Education Center, University of Florida/IFAS, Quincy, FL, United States*
**Correspondence author email address: shafiq.iags@pu.edu.pk, mshahid@ufl.edu*

(Received, 23rd January 2023, Revised 25th August 2023, Published 26th August 2023)

Abstract *This study was made on three genes ACS, ACO, and ETO, which are involved in ethylene biosynthesis pathway in Fragaria vesa plant, to know about evolution, conserved motifs and domains, gene expressions, and phylogeny of these genes. After carefully screening using Phytozome plant gene database, NCBI gene database, Motif finder, and MegaX phylogenetic tree 10 gene sequences of ACS, 5 gene sequences of ACO and 3 gene sequences of ETO were identified. Four ETO gene sequences of Arabidopsis thaliana were also used to authenticate this research because only 3 ETO gene sequences of Fragaria vesca analyses cannot be done. MegaX evolutionary analysis, TB tools domain analysis, Meme motif analysis, Cis-regulatory analysis, Wolf analysis were made on these sequences to acquire detailed knowledge. The presence of light, anaerobic induction, abscisic acid, MeJA, gibberellin, low temperature, drought, cell cycle, and endosperm expression responsive elements were identified in FeACS, FeACO, and FeETO genes by cis-regulatory analysis. This study will help for further practical experimentation on ethylene regulators. The bioinformatics-based genome-wide assessment of the family of Fragaria vesca attempted in the present study could be a significant step for further practical investigation on ethylene regulators based on genome-wide expression profiling.*

*[Citation: Fatima, S., Cheema, K., Shafiq, M., Manzoor, M., Ali, Q., Haider, M., Shahid, M. (2023). Genome-wide bioinformatics analysis of 1-aminocyclopropane-1-carboxylate synthase (ACS), 1-aminocyclopropane-1-carboxylate oxidase (ACO) and ethylene overproducer 1 (ETO1) gene family of fragaria vesca (woodland strawberry). Bull. Biol. All. Sci. Res. **8**: 38. doi: https://doi.org/10.54112/bbasr.v2023i1.38]*
Keywords: *Fragaria vesca, ACS, ACO, ETO, Genome-wide, Bioinformatics*

Introduction

The wild strawberry, *Fragaria vesca*, an important plant for experimental purposes, is diploid (2n = 2x = 14). *Fragaria vesca* belongs to the rosacea family, a herbaceous but perennial plant. Its genome is small, consisting of only 240 Mb, prone to genetic changes, identical to *Fragaria ananassa* (cultivated strawberry) and other rosaceous plants. About 34,809 genes have been identified by gene prediction using transcriptome mapping. *Fragaria ananassa* is one of the youngest crops and originated only about 250 years ago. Its fruit could not be categorized as berry or true fruit because its surface has abundant seeds. More broadly, *F. vesca* offers many benefits as a model genomic system for Rosaceae family, it is perennial but short-term generation time, and its vegetative propagation and small herbaceous stature also add ease to the experiments done with Fragaria *vesca* as compared with tree species of rosaceae family such as peach or apple. *F. vesca*, encouraging the creation of forward and reverse genetic devices just as primary and utilitarian investigations (Shulaev et al., 2011). Two main enzymes are involved in ethylene biosynthesis 1- aminocyclopropane-1-carboxylate (ACC) oxidase and 1-aminocyclopropane-1-carboxylate (ACC) synthase. ACC synthase act on the S-adenosylmethionine (SAM) and converts it to ACC (1-aminocyclopropane-1-carboxylate); and ACC oxidase take ACC as substrate and oxidizes it to ethylene (Hans Kende, 1989). ACC synthase and ACC oxidase are not encoded from a single gene instead, they belong to multigene families in different plants. ACC synthase belongs to a multigene family called PLP-dependent enzyme family and cannot perform its role without a cofactor called pyridoxal-

5'- phosphate (PLP or vitamin B6) (Huai et al., 2001; Yip, Dong, Kenny, Thompson, & Yang, 1990). ACC synthase was first found in 1979 in tomato (Boller, Herner, & Kende, 1979). Then, the genomic study of many plants such as Arabidopsis, wheat, apple, pear, carnation, rose, banana, potato, winter squash, and rice have shown that they also have ACC synthase genes (Ge et al., 2000; Phoebe R. Johnson & Joseph R. Ecker, 1998) etc.

ACC oxidase (ACO) is engaged with the last advance of ethylene creation in plant tissues. A few ACO isoforms have been disengaged in different plants, which a multigenic family encodes. There is solid proof that ACO quality articulation is related to the ethylene creation rates, and its various isoforms are a work in progress and natural control. Subsequently, the regulation of ACO quality action may act either as an extra or in a few cases additionally as a fundamental level for controlling ethylene biosynthesis in higher plants. (Ruduś, Sasiak, & Kępczyński, 2012). Arabidopsis thaliana ETO1 (ethylene overproducer) and ETO3 mutants produce raised degrees of ethylene as etiolated seedlings. (Woeste, Ye, & Kieber, 1999). In Arabidopsis, ETO1 (ETHYLENE-OVERPRODUCER1) is a negative controller of ethylene development by communicating with *AtACS*5, an isoform of the rate-restricting catalyst, 1-aminocyclopropane-1-carboxylate synthases (ACC synthase or ACS), in ethylene biosynthetic pathway. ETO1 directly inhibits the enzymatic activity of *AtACS*5 (Yoshida, Nagata, Saito, Wang, & Ecker, 2005)

This research paper found 10, 5, and 3 sequences of ACS, ACO, and ETO genes in *Fragaria vesca*, respectively. Using MegaX, various phylogenetic trees were constructed to study the evolutionary relationship among gene sequences. Our statistical and systematic analysis of ACS, ACO, and ETO genes of *Fragaria vesca* plays a key role and acts as a base for an extensive and comprehensive study of these genes. Certainly, this study will open doorways for further experimentation on ethylene.

Materials and methods

Database search and retrieval of sequence

Fragaria vesca ACS, ACO and ETO genes were identified in the *Fragaria vesca* Genome Database Phytozome (https://Phytozome.jgi.doe.gov/pz/portal.html) (Finn et al., 2014). At first, they were 16, 146, and 7 for ACS, ACO, and ETO, respectively, from which correct sequences were identified later. Subsequently, protein, gene, and virtual cDNA sequences were all retrieved from the Plant Genome Database (https://Phytozome.jgi.doe.gov/pz/portal.html).

Determination of physio-chemical properties of woodland strawberry proteins:

Gene accession number, transcript ID, linkage group number or chromosome number, gene location, gene orientation, orthologues in other plants, number of introns, miRNA length in base pairs were taken from Phytozome gene database. For amino acid length, pI, and molecular weight, the online ProtParam tool (http://web.expasy.org/protparam/) was used(Gasteiger et al., 2005). For this purpose, peptide sequences were taken of all genes individually and pasted in box of ProtParam tool-ExPASY, then the results.

Conserved domain analysis:

The conserved domains in the protein sequences of ACS, ACO, and ETO genes of Fragaria vesca were identified by a web-search tool "NCBI Conserved Domain Search – NIH" (https://ncbi.nlm.nih.gov/Structure/cdd/wrpsb/cgi). ACS, ACO, and ETO gene protein sequences were searched against "Pfam v32.0 – 17919 PSSMs". MEGA X software was used to make a phylogenetic tree among peptide sequences of a gene using "neighborhood joining method" at 1000 bootstraps. One tree was made for each gene, and the files were saved in "newick" format. These files were added to "TB tools" software, and the results were saved as a graphical diagram after the necessary edition.

Multiple sequence alignment and phylogenetic analysis

To study the evolutionary relationship among the sequences of ACS, ACO, and ETO genes, peptide or protein sequences of these genes were used. 10, 5, and 3 sequences of ACS, ACO and ETO, respectively were taken. The MEGA X (molecular evolutionary genetics analysis) program was used for making phylogenetic trees. The tree was constructed using "Construct/Test UPGMA Tree" option, and the number of bootstrap replications was set to "1000". (Kumar et al., 2018).

Cis-regulatory elements and conserved motifs recognition

To study the conserved motifs in the sequences two online search databases are used "Multiple EM for Motif Elicitation (MEME)" program (http://meme-suite.org/tools/meme) and "MOTIF: Searching Protein Sequence Motifs – GenomeNet" (https://www.genome.jp/tools/motif/) search database (Bailey et al., 2015). The peptide sequences of genes were pasted individually in MEME search data. For making the "Cis-regulatory element analysis" the promoter regions of the sequences for genes ACS, ACO, and ETO were collected from the "Phytozome gene database". For making FASTA file of the promoter regions 1000-bp sequences before the start codon "ATG" of the genomic sequences were collected for each sequence. Then these promoter sequences were further analyzed using "PlantCARE, a database of plant promoters and their cis-acting analysis". This statistical data was then used to make a "HeatMap" using "TB tools" program (Rombauts et al., 1999)

Prediction of the subcellular localization

ACO and ETO proteins' peptide sequences were used to predict the subcellular localizations of ACS. Fasta file of these peptide sequences was subjected to a

web-based tool "WoLF PSORT" (https://wolfpsort.hgc.jp/) to predict protein subcellular localization. Results of WoLF PSORT were collected in an Excel sheet for convenience (Horton et al., 2006).

Identification of orthologues

Orthologues of ACS, ACO and ETO genes in plants other than *Fragaria vesca* were isolated from an online gene search database "Phytozome" (https://Phytozome.jgi.doe.gov/phytomine/portal.do?externalid=PAC:27262948&class=gene) with default settings.

Nuclear localization signals calculation

Nuclear localization signals were predicted using a web-based tool "NLSdb – the Rostlab" (Nuclear localization signals database) (https://rostlab.org/services/nlsdb/). The ACS, ACO, and ETO gene peptide sequences were subjected to NLSdb for prediction. Results were downloaded in an Excel sheet (Cokol et al., 2000).

Expressed sequence tag (EST) analysis

The expression status of identified sequences of ACS, ACO, and ETO genes of *Fragaria vesca* in various organs and tissues of the plant was checked from an online databank "PlantGDB-Resources for Plant Comparative Genomics" (http://plantgdb.org/cgi-bin/blast/PlantGDB). ACO and ETO genes were used to dig out EST data from PlantGDB DNA sequences of ACS. E-value was set to "1e−04" for blastn search, but all other parameters were default. TB tools were used to illustrate EST data about the expression of genes in different parts of a plant in a heatmap for investigation (Kantety et al., 2002).

Results

Identification of ACS, ACO and ETO genes in *Fragaria vesca*

After carefully surveying, 10 members of ACS, 5 of ACO, and 3 of ETO gene in Fragaria vesca were identified and used for further analyses. The proteins encoded by the same gene isoforms and proteins containing the truncated ACS, ACO and ETO DNA-binding domains were excluded from the analysis. The CDS and protein sequences of ACS genes family were all downloaded from Phytozome plant gene database (https://Phytozome.jgi.doe.gov/pz/portal.html). The *FeACS* genes encode proteins ranges from 409–546 amino acids in length and with a molecular weight that ranges from 105199.27 to 99026.83 kD, with *FeACS*10 being the smallest and *FeACS*1 being the longest protein (Table 1). The isoelectric points of identified proteins ranged from 7.35 to 9.75. The *FeACO* genes encode proteins ranges from 299–363 amino acids in length and with a molecular weight that ranges from 103099.27 to 98026.83 kD, with *FeACO*1 being the smallest and *FeACO*5 being the longest protein (Table 1). The isoelectric points of identified proteins ranged from 7.15 to 9.35. The *FeETO* genes encode proteins ranging from 898–1251 amino acids in length and with a molecular weight that ranges from 113099.27 to 99726.83 kD, with *FeETO*1 being the smallest and *FeETO*2 being the longest protein (Table 1). The isoelectric points of identified proteins ranged from 7.55 to 9.65.

Table 1: Table of different characteristics of gene sequences. It contains information about gene name, ID in Phytozome, chromosome number on which gene located, location by number of start and end nucleotides, orientation, and orthologues in other plants. Table also has data about different characteristics of protein sequences of these genes such as the number of introns, mRNA base pairs, amino acids, pI, and molecular weight

Gene	Gene Accession No Phytozome	Transcript ID	Chromosome no/Contig NO	Location	Orientation	Orthologe in Arbidopsis/ other plants	Protein No of Intron	mRNA (bp)	AA length	pI	M_w (D)
FveACS1	gene02640-v1.0-hybrid	mrna02640.1-v1.0-hybrid (primary)	LG2	18063529..18065375	forward	orange1.1g037587m.g	3	1380	459	8.74	51467.49
FveACS2	gene05580-v1.0-hybrid	mrna05580.1-v1.0-hybrid (primary)	LG1	15450203..15452461	forward	orange1.1g043923m.g	4	1641	546	8.39	61291.90
FveACS3	gene11391-v1.0-hybrid	mrna11391.1-v1.0-hybrid (primary)	LG1	4632483..4637290	reverse	evm.TU.supercontig_922.3	4	1368	455	5.63	50624.47
FveACS4	gene11392-v1.0-hybrid	mrna11392.1-v1.0-hybrid (primary)	LG1	4642656..4644989	reverse	orange1.1g046900m.g	3	1479	492	7.59	55555.34

FveACS5	gene13406-v1.0-hybrid	mrna13406.1-v1.0-hybrid (primary)	LG7	22420488..22422725	forward	orange1.1g019899m.g	3	1230	409	5.20	45649.48
FveACS6	gene19023-v1.0-hybrid	mrna19023.1-v1.0-hybrid (primary)	LG7	2593550..2595998	forward	orange1.1g011801m.g	3	1482	493	9.04	55357.65
FveACS7	gene30630-v1.0-hybrid	mrna30630.1-v1.0-hybrid (primary)	LG3	3118601..3120385	reverse	orange1.1g037587m.g	3	1410	469	7.14	52744.10
FveACS8	gene30682-v1.0-hybrid	mrna30682.1-v1.0-hybrid (primary)	LG3	2699983..2701731	forward	orange1.1g037587m.g	3	1386	461	8.85	51850.21
FveACS9	gene31839-v1.0-hybrid	mrna31839.1-v1.0-hybrid (primary)	LG5	1983769..1985918	reverse	orange1.1g011570m.g	3	1464	487	6.06	54940.75
FveACS10	gene32323-v1.0-hybrid	mrna32323.1-v1.0-hybrid (primary)	LG5	702099..704680	reverse	evm.TU.supercontig_29.56	3	1608	535	8.78	58771.43
FveACO1	gene01202-v1.0-hybrid	mrna01202.1-v1.0-hybrid (primary)	LG6	31374856..31376752	reverse	orange1.1g020953m.g	3	1092	363	5.77	41614.59
FveACO2	gene11261-v1.0-hybrid	mrna11261.1-v1.0-hybrid (primary)	LG2	17115740..17116826	forward	evm.TU.supercontig_64.148	2	912	303	4.93	34153.67
FveACO3	gene11421-v1.0-hybrid	mrna11421.1-v1.0-hybrid (primary)	LG1	4770859..4771937	forward	orange1.1g021636m.g	2	918	305	5.29	34561.97
FveACO4	gene11424-v1.0-hybrid	mrna11424.1-v1.0-hybrid (primary)	LG1	4775876..4776980	forward	orange1.1g021636m.g	2	924	307	5.61	35042.92
FveACO5	gene19733-v1.0-hybrid	mrna19733.1-v1.0-hybrid (primary)	LG3	888704..894071	forward	EuGene.0900010377	3	900	299	5.75	34475.49
FveETO1	gene13489-v1.0-hybrid	mrna13489.1-v1.0-hybrid (primary)	LG6	7023230..7029121	forward	orange1.1g002100m.g	4	3756	1251	5.25	141630.84
FveETO2	gene20345-v1.0-hybrid	mrna20345.1-v1.0-hybrid (primary)	LG2	14457222..14460715	forward	orange1.1g002716m.g	4	2697	898	5.78	101898.45
FveETO3	gene28226-v1.0-hybrid	mrna28226.1-v1.0-hybrid (primary)	LG3	20826281..20829787	reverse	orange1.1g002379m.g	3	2784	927	5.93	105240.50

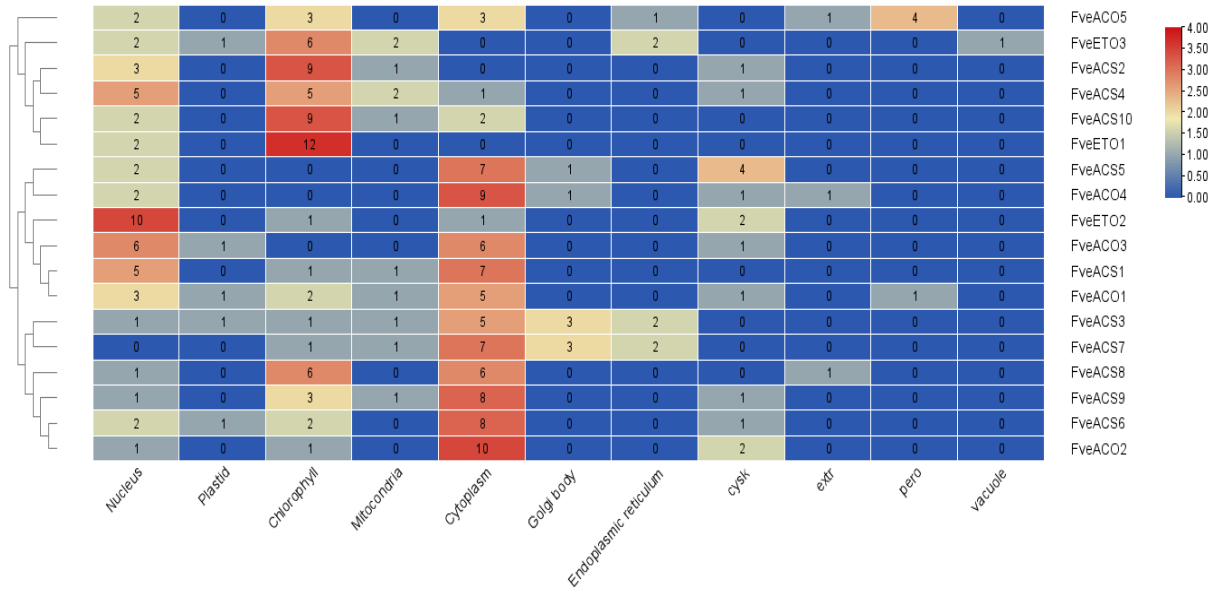

Figure 1: Diagram of Wolf analysis for ease in data analysis. Different values has been represented with different colors.

Comparative phylogenetic relatedness of *F. vesca ACS, ACO* and *ETO1* gene family with Arabidopsis

To investigate the molecular evolution and phylogenetic relationships among gene sequences phylogenetic trees were constructed using the neighbor-joining (NJ) method. For statistical reliability, we conducted bootstrap analysis with 1000 replicates. The NJ phylogenetic tree showed 10 Fve ACS sequences can be divided into three groups "Group A", "Group B" and "Group C" (Fig 2a). Group A has two sequences *FeACS*5 and *FeACS*10. The two sequences are in one clad, which shows they are parallel. Group B consists of a clad with four sequences subdivided into two subgroups "B1" and "B2". Group B1 consists of a clad with two sequences FeACS3 and FeACS4, which are paralogs. Group B2 consists of another pair of paralog sequences *FeACS*6 and *FeACS*9. Quite similar to Group B, Group C consists of a clad with four sequences which is further subdivided into three subgroups "C1", "C2" and "C3". Subgroup C1 consists of a clad with two sequences *FeACS*7 and *FeACS*8, which are paralogs to each other. Subgroups C2 and C3 do not have clads. *FeACS*1 and *FeACS*2 lie in subgroups C2 and C3, respectively. Sequences *FeACS*1 and *FeACS*2 are not paralogs to each other (Fig 2a). Phylogenetic tree of 5 *FveACO* sequences can be sub divided into three groups "Group A", "Group B" and "Group C". Group C consists of a clad with two paralog sequences *FeACO*3 and *FeACO*4. Group B has only one sequence *FeACO*2, and does not have a clad. The two paralog sequences *FeACO*1 and *FeACO*5 lies in Group A. Group C consists of the most primitive and ancestors of other *FeACO* sequences. *FeACO*2 which lies in Group B evolved from Group C. Group A sequences are evolved from group from Group B and are the least primitive (Fig 2b). Although only three sequences of *FeETO* were found but phylogenetic tree was made with seven gene sequences. Four ETO gene sequences of Arabidopsis thaliana were added to make the results more authentic. These Arabidopsis thaliana sequences were named the same as in the database. Through care analysis phylogenetic tree can be divided into two groups Group A and Group B. Group A consists of only two sequences *FeETO*2 and atEOL1_888, while Group B has five members "*atETO*1_951, *atETO*1_956, atEOL2_925, *FeETO*1, *FeETO*3" respectively (Fig 2c). Two *Fragaria vesca* FeETO1 and FeETO3 sequences showed more similarity with Arabidopsis thaliana gene sequences. Instead of making a clad with *FeETO*2 and *FeETO*3, *FeETO*1 were in same clad with three Arabidopsis thaliana sequences "*atETO*1_951, *atETO*1_956 and atEOL2_925". *FeETO*2 was in the same clad as fourth sequence of Arabidopsis thaliana "atEOL1_888. According to the tree *FeETO*2 is the most primitive ancestor and *FeETO*3 evolved from it. Later *FeETO*1 evolved from *FeETO*2 (Fig 2c). So *FeETO*1 is the most advanced one according to the evolutionary history. Moreover, our subclades classification was consistent with the report from Jakubowicz (Jakubowicz, 2002). Segmental duplication, tandem duplication and transposition events are three main reasons for gene family expansion (Kong et al., 2007).

Figure 2a: Phylogenetic tree of ACS gene sequences to understand the phylogenetic relationship among 10 FveACS gene sequences and determine different groups formed based on gene similarity.

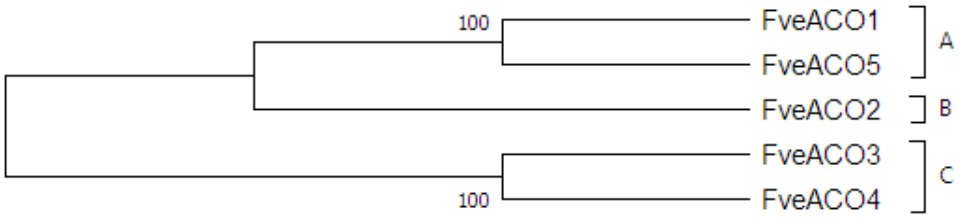

Figure 2b: Phylogenetic tree of ACO gene sequences to understand the phylogenetic relationship among 5 FveACO gene sequences and determine different groups formed based on similarity among genes.

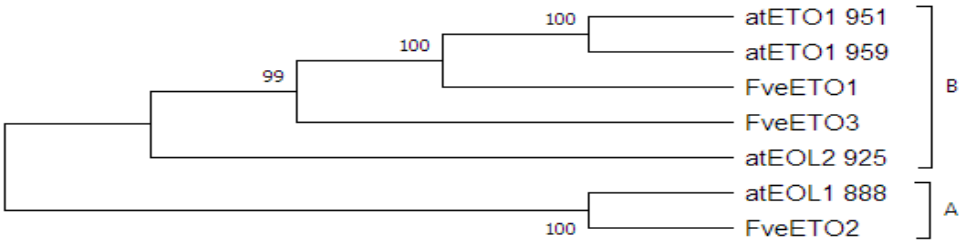

Figure 2c: Phylogenetic tree of ETO gene sequences to understand the phylogenetic relationship among 3 FveETO and 4 AthETO gene sequences and determine different groups formed based on similarity

Gene Structures and Recognition of Conserved Motifs and Domain

The domain graph of *Fagaria vesca* for ACS gene made by Tb tools has shown surprising results. Four domains were found in 10 sequences of Fve ACS, of which two domains "Aminotransferase 1_2" and "AAT_1 superfamily" were found in all the sequences. These two domains completely overlap each other in every peptide sequence. Two sequences "*FeACS*3" and "*FeACS*9" have another other domain "Beta_elim_lyase" along with the former two domains. "*FeACS*4" showed different results from other sequences and have four domains. Another domain "Cys_Met_Meta_PP" was found in this sequence along with the former three domains (Fig 3a). The domain analysis of 5 ACO sequences of *Fragaria vesca* has shown same pattern in every sequence. Total of four domains "2OG-Fell_Oxy", "2OG-Fell_Oxy superfamily", "DIOX_N", "DIOX_N superfamily" were found, and these four domains were present in every sequence. But the diagram of TB tools displays only two domains in each sequence. This is because 2OG-Fell_Oxy superfamily domain completely overlap 2OG-Fell_Oxy domain and DIOX_N superfamily domain completely overlap DIOX_N domain. When the diagram was distorted, it became clear that four domains are in each sequence (Fig 3b). As mentioned already, only 3 sequences of Fve ETO were identified from Phytozome gene database. But for making MEGA X phylogenetic tree minimum four sequences were required. So, 4 sequences of Arabidopsis thaliana ETO were added to them. These 3 sequences of Fve ETO showed quite different results from each other. Overall, 6 domains were found in seven gene sequences. Only one domain "PEP_TPR_lipo superfamily" was conserved and found in all gene sequences. "BTB_POZ_ETO1-like" domain was found in all four Arabidopsis thaliana sequences but was absent in all the Fragaria vesca ETO gene sequences. Instead of "BTB_POZ_ETO1-like" domain all the three ETO sequence of Fragaria vesca have another quite similar domain at the same position, "BTB_POZ superfamily". In each ETO gene sequence of Arabidopsis thaliana, two similar domains "PEP_TPR_lipo superfamily" and "BTB_POZ_ETO1-like" were identified. Three domains were identified in each Fragaria vesca ETO sequence. Out of three two domains "PEP_TPR_lipo superfamily" and "BTB_POZ superfamily" were same in all *Fragaria vesca* ETO gene sequences. Every *Fragaria vesca* ETO sequence was identified with a unique domain. *FeETO*1, *FeETO*2 and *FeETO*3 were identified with three different domains "DUF688 superfamily", "GAF superfamily," and "3a0801s09 superfamily," respectively (Fig 3c).

.

.

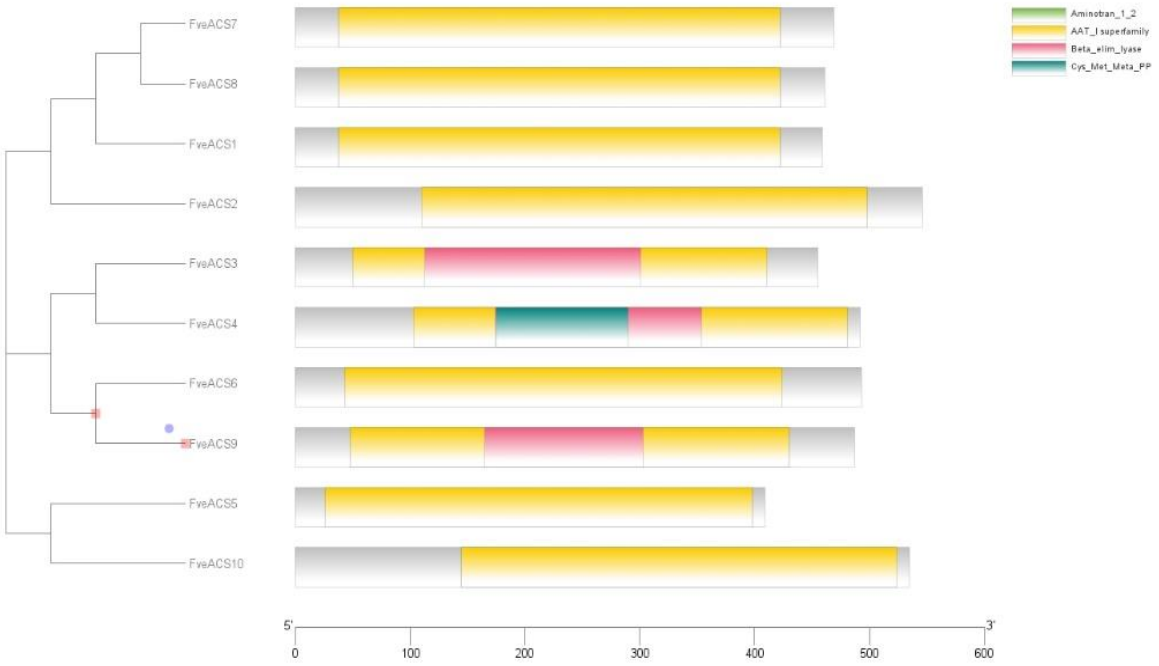

Figure 3a: Diagram of FveACS domain analysis for 10 FveACS gene sequences to clearly depict the type and position of different domains using different colors found in gene sequences and the phylogenetic relationship.

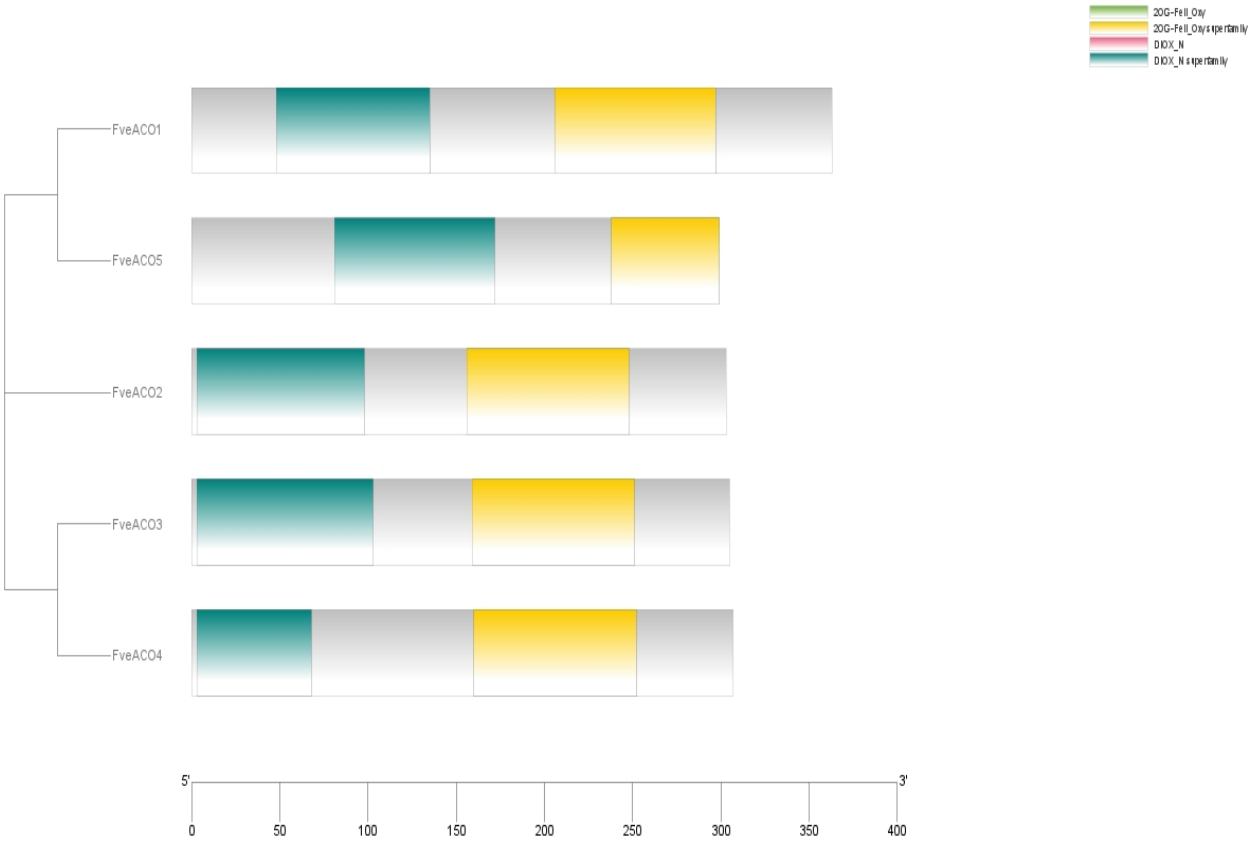

Figure 3b: Diagram of FveACO domain analysis for 5 FveACO gene sequences to clearly depict the type and position of different domains using different colors found in gene sequences and the phylogenetic relationship.

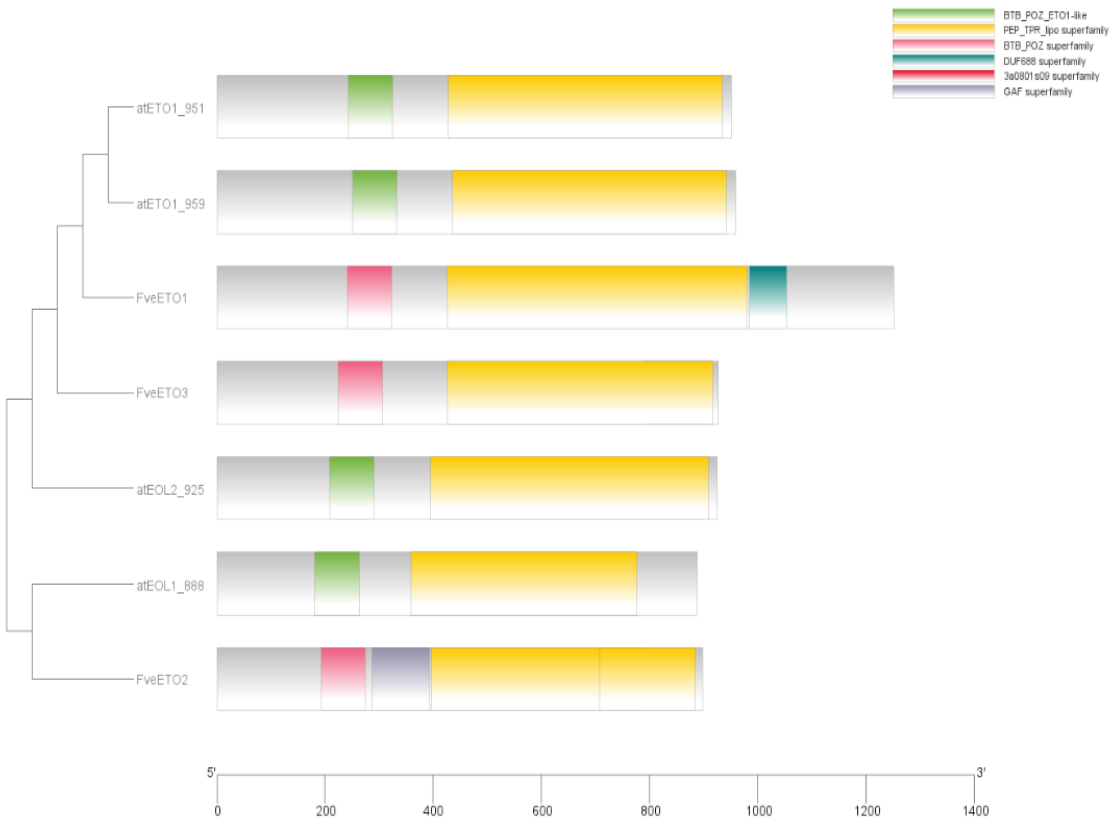

Figure 3c: Diagram of FveETO and AthETO domain analysis for 3 FveETO and 4 AthETO gene sequences to clearly depict the type and position of different domains using different colors found in gene sequences and the phylogenetic relationship.

Meme motif analysis

Most motifs were consistent in all *FeACS* sequences, showing that the gene evolved during the gene expansion process. Motifs 1, 2, 3, 4, 6, 7, 8 and 10 were found in all sequences, and the same pattern predicted that all *FeACS* sequences might have a common ancestor. Motif 5 is found in all except in *FeACS*5. Ten *FeACS* sequences were characterized into three groups A, B and C. Group C has four members *FeACS*7, *FeACS*8, *FeACS*1 and *FeACS*2. Group B has four members *FeACS*3, 4, 6, and 9. Group A consists of only two members *FeACS*5 and *FeACS*10. Motif 17, 12 and 14 were identified in all members of Group A. Group B was divided into subgroups B1 and B2 containing *FeACS*3, *FeACS*4 and *FeACS*6, *FeACS*9, respectively. Motif 16 was identified only in members of subgroup B1. Motifs 13, 18 and 20 were recognized only in members of subgroup B2. Motif 19 was found only in two Group A. Motif 15 members showed astonishing behavior at 5' terminal in *FeACS*10 and at 3' terminal in *FeACS*2 (Fig 4a)(Fig 4.1a). *According to the phylogenetic tree, FeACO gene sequences were classified into three groups A, B and C*. Group C, B, and A consists of (*FeACO*3, *FeACO*4), (*FeACO*2) and (*FeACO*1, *FeACO*5) respectively. Motifs 1, 2, 4, and 7 were found in all groups without exception. Motifs 11, 10 and 9 were identified only in Group C. Motifs 12, 15, 17 and 18 were recognized only in Group A but their position is not same there might be some evolution in their position. Motif 16 was found in *FeACO*1 and *FeACO*5 members of group A, but their position is completely reciprocal. It was identified at 3' terminal of *FeACO*1 while in *FeACO*5 it lies at 5' terminal. Motifs 13 and 20 showed quite surprising results. They were found at similar positions in *FeACO*1 and *FeACO*2, belonging to two groups. *FeACO*1 and *FeACO*2 might have some close or similar ancestor. They also depict some similarity in their function (Fig 4b)(Fig 4.1b). Seven peptide sequences were classified into two groups A and B. Most of the motifs were conserved and consistent in both groups. Motifs 1… 14, 16, 17, and 19 were identified in all members at same positions. Motif 15 differentiate groups A and B. It was found in five members of Group B (*atETO*1_951, *atETO*1_956, atEOL2_925, *FeETO*1, *FeETO*3) and was absent in Group A members (*FeETO*2 and atEOL1_888). Two members of Group B "*atETO*1_951" and "*atETO*1_956" have same conserved motifs. All motifs 1… 20 were identified in them (Fig 4c)(Fig 4.1c).

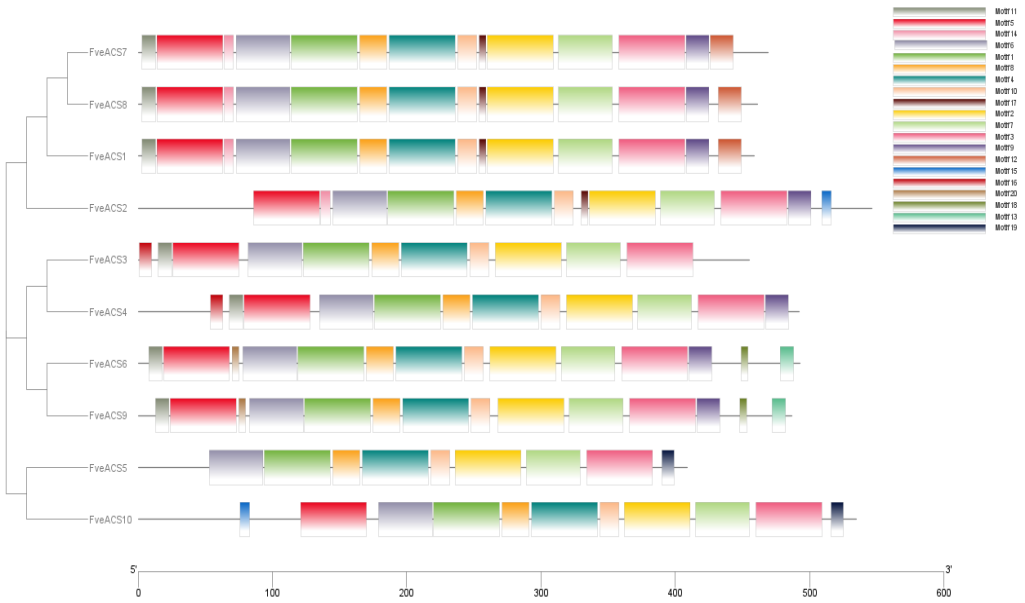

Figure 4a: Diagram of FveACS motif analysis for 10 FveACS gene sequences to clearly depict the type and position of different motifs using different colors found in gene sequences and the phylogenetic relationship to understand the similarities among gene sequences.

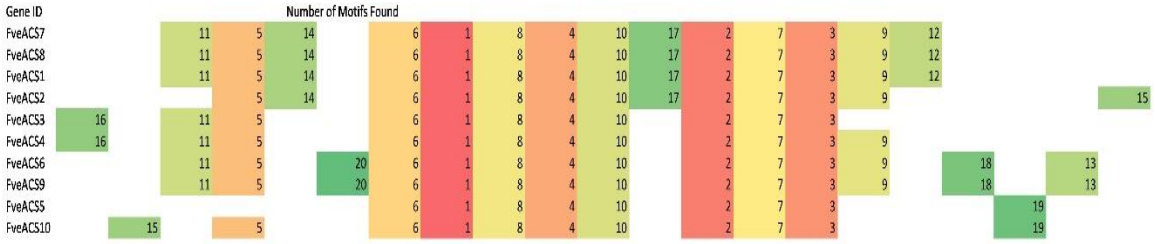

Figure 4.1a: FveACS Motif table for 10 FveACS genes. Every motif from 1 to 20 is colored distinctly to easily understand the presence and absence of different motifs in gene sequences and to predict the similarities and differences in functions of gene sequence based on the presence and absence of similar motifs.

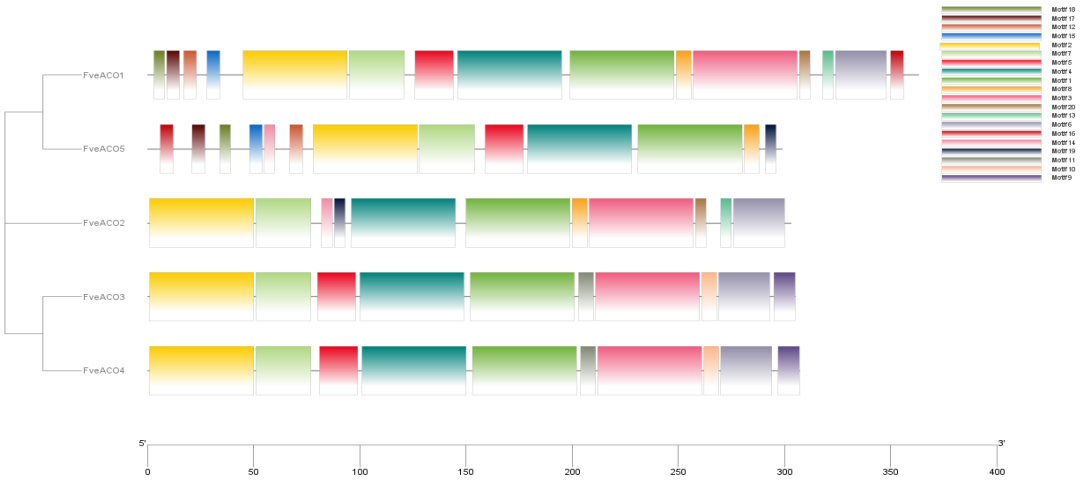

Figure 4b: Diagram of FveACO motif analysis for 5 FveACO gene sequences to clearly depict the type and position of different motifs using different colors found in gene sequences and the phylogenetic relationship to understand the similarities among gene sequences.

Gene ID	Motifs Found																				
FveACO1	18	17	12	15			2	7	5		4	1	8	3	20		13	6	16		
FveACO5	16	17	18	15	14	12	2	7	5		4	1	8	19							
FveACO2							2	7	14	19	4	1	8	3	20		13	6			
FveACO3							2	7	5		4	1	11	3		10		6	9		
FveACO4							2	7	5		4	1	11	3		10		6	9		

Figure 4.1b: FveACO Motif table for 5 FveACO genes. Every motif from 1 to 20 is colored distinctly to easily understand the presence and absence of different motifs in gene sequences and to predict the similarities and differences in functions of gene sequences based on the presence and absence of similar motifs.

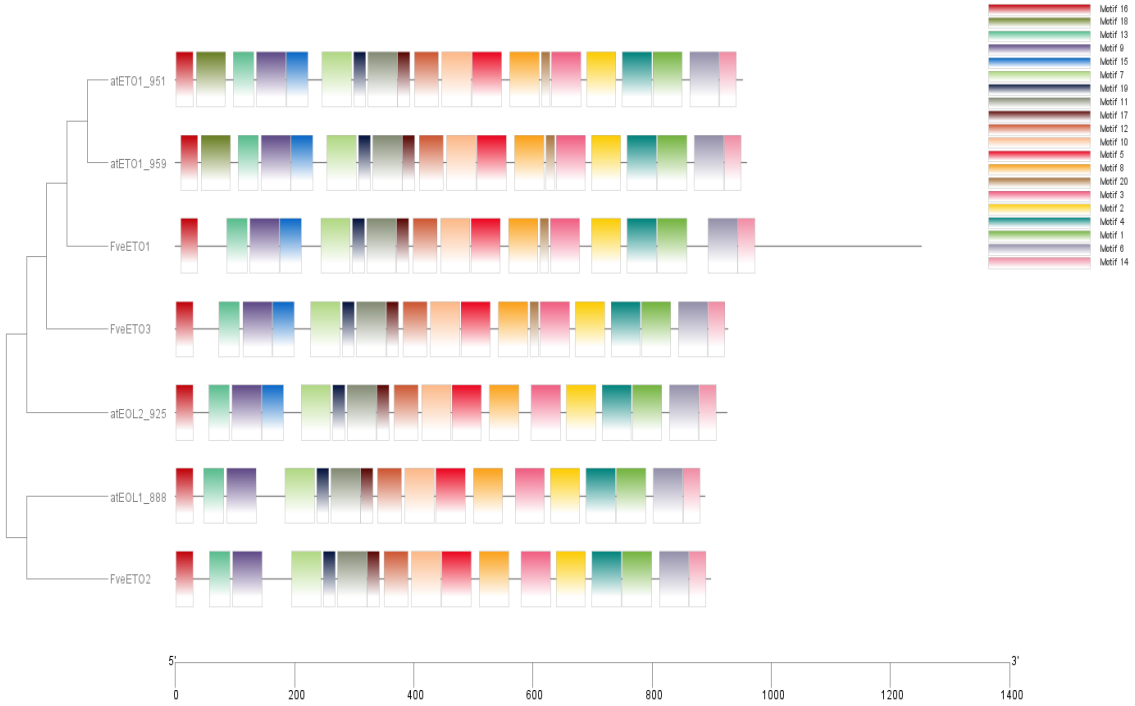

Figure 4c: Diagram of FveETO motif analysis for 3 FveETO and 4 AthETO/EOL gene sequences to depict the type and position of different motifs using different colors found in gene sequences along with the phylogenetic relationship to understand the similarities among gene sequences.

Gene ID	Motifs Found																				
atETO1_951	16	18	13	9	15	7	19	11	17	12	10	5	8	20	3	2	4	1	6	14	
atETO1_959	16	18	13	9	15	7	19	11	17	12	10	5	8	20	3	2	4	1	6	14	
FveETO1	16		13	9	15	7	19	11	17	12	10	5	8	20	3	2	4	1	6	14	
FveETO3	16		13	9	15	7	19	11	17	12	10	5	8	20	3	2	4	1	6	14	
atEOL2_925	16		13	9	15	7	19	11	17	12	10	5	8		3	2	4	1	6	14	
atEOL1_888	16		13	9		7	19	11	17	12	10	5	8		3	2	4	1	6	14	
FveETO2	16		13	9		7	19	11	17	12	10	5	8		3	2	4	1	6	14	

Figure 4.1c: FveETO Motif table for 3 FveETO and 4 AthETO/EOL genes. Every motif from 1 to 20 is colored distinctly to easily understand the presence and absence of different motifs in gene sequences and to predict the similarities and differences in functions of gene sequences based on the presence and absence of similar motifs.

Analysis of Cis-regulatory Elements

Cis-regulatory analysis was made to probe the cis-regulatory elements, which are responsible for the expression of the genes. The expression of genes is affected by the presence and organization of various cis-regulatory elements at the binding site of transcription factors on promoter region. An in-silico analysis of the various cis-regulatory elements was made to find the putative functions of genes. Various cis-regulatory elements "AE-box, ATCT-motif, Box 4, G-Box, G-box, GA-motif, GATA-motif, GT1-motif, GTGGC-motif, Gap-box, I-box, LAMP-element, MRE, SP 1, TCCC-motif, TCT-motif, chs-CMA1a, and chs-CMA2a" responsible for the light responses were found in all *FeACS* gene sequences. Cis-regulatory element "ARE" responsible for the anaerobic induction were identified in all the *FeACS* sequences except *FeACS*10. "MeJA" responsive cis-regulatory elements "CGTCA-motif" and "TGACG-motif" were not found only in two sequences *FeACS*6 and *FeACS*9. "Abscisic acid" responsive cis-regulatory element "ABRE" was absent in four genes *FeACS*3, *FeACS*6, *FeACS*8 and *FeACS*9. The conserved DNA module array (CMA3) fragments were found in *FeACS*2 and *FeACS*10, which are responsible for light responses. *FeACS*1, *FeACS*4 and *FeACS*6 genes contain gibberellins responsive cis-regulatory elements "P-box and TATC-box". Cis-regulatory element "CAT-box" related to Meristem expression was found only in *FeACS*7 gene. *FeACS*2, *FeACS*4, *FeACS*6 and *FeACS*8 genes possess "LTR" cis-regulatory element responsible for low temperature responsiveness. Only two genes *FeACS*6 and *FeACS*10 contain "GCN4_motif" cis-regulatory element responsible for the endosperm expression. "Drought inducibility" cis-regulatory element "MBS" was found in *FeACS*2, *FeACS*4, *FeACS*9 and *FeACS*10 genes. Cis-regulatory elements "MBSI" and "TCA-element" responsible for "flavonoid biosynthetic genes regulation" and "salicylic acid responsiveness" respectively were possessed by only two genes *FeACS*7 and *FeACS*10. "O2-site", a cis-regulatory element which is responsible for "zein metabolism regulation" was identified in *FeACS*4, *FeACS*5, and *FeACS*8 genes. "Seed-specific regulation" responsive cis-regulatory element "RY-element" was found in *FeACS*4 and *FeACS*7 genes. Only one gene, FeACS9, possesses "defense and stress" responsive cis-regulatory element "TC-rich repeats". "Auxin responsive" cis-regulatory element "TGA-element" was probed in *FeACS*1, *FeACS*8 and *FeACS*10 genes (Fig 5a). "Light" and "Abscisic acid" responsive cis-regulatory elements were found in all five *FeACO* genes. Cis-regulatory element "ARE" responsible for anaerobic induction was found in four genes but absent in *FeACO*5. "MeJA-responsive" cis-regulatory element "CGTCA-motif" was identified in three genes *FeACO*1, *FeACO*2 and *FeACO*5, while "gibberellin-responsive" cis-regulatory elements "GARE-motif" and "P-box" were found in two genes *FeACO*2 and *FeACO*5 respectively. "Cell cycle regulation" and "zein metabolism regulation" responsive cis-regulatory elements "MSA-like" and "O2-site" respectively were possessed by only one gene *FeACO*2. "Auxin responsive" and "low temperature responsive" cis-regulatory elements "TGA-element" and "LTR" respectively were only found in *FeACO*5 gene. *FeACO*5 also contains an enhancer like the cis-regulatory element "GC-motif" involved in "anoxic specific inducibility". "Drought inducibility" responsive cis-regulatory element "MBS" was found in *FeACO*3 and *FeACO*4, while "salicylic acid" responsive cis-regulatory element "TCA-element" was probed in *FeACO*4 and *FeACO*5 genes (Fig 5b). Cis-regulatory elements involved in "Light responsiveness" and "Anaerobic induction" were found in all the three genes of *FeETO*. *FeETO*2 also have "3-AF3 binding site" which part of a conserved DNA module array (CMA3) involved in light responsiveness. Cis-regulatory elements "CAT-box" and "LTR" involved in "meristem expression" and "low-temperature responsiveness" respectively were found only in *FeETO*1 gene. Only *FeETO*3 possess "GARE-motif", "MBS" and "TCA-element" cis-regulatory elements which are involved in "gibberellin responsiveness", "drought inducibility" and "salicylic acid responsiveness" respectively. "MeJA responsive" cis-regulatory element "CGTCA-motif" was found only in *FeETO*2 gene. "Defense and stress responsive" cis-regulatory element "TC-rich repeats" was identified in *FeETO*1 and *FeETO*2 genes, while "Auxin responsive" cis-regulatory element "TGA-element" was probed in *FeETO*2 and *FeETO*3 genes (Fig 5c).

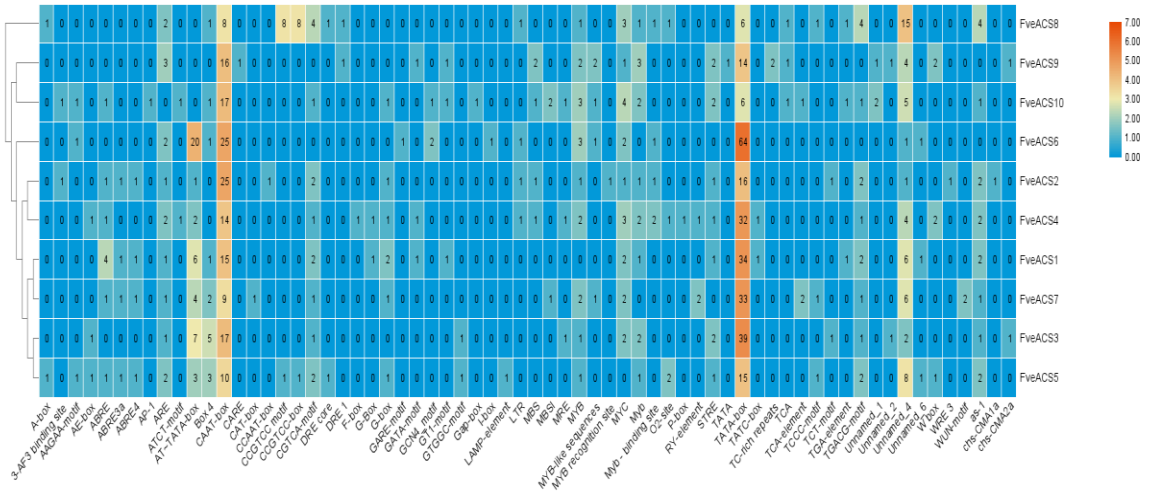

Figure 5a: FveACS Cis-regulatory elements Heat map. The numbers in this heat map how much repetition of that specific element was found in a gene promoter sequence. Different numerical values have been represented with different colors for easy data analysis.

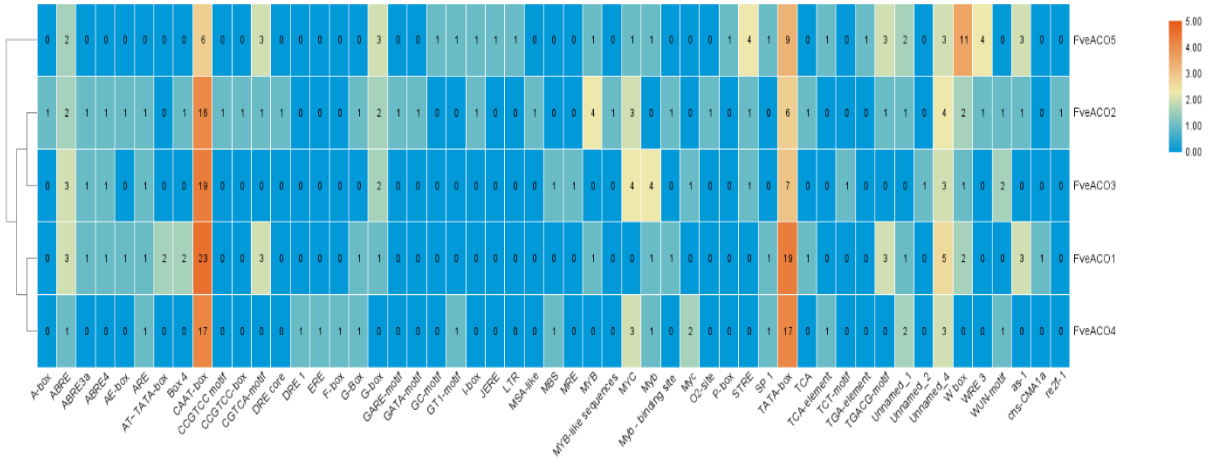

Figure 5b: FveACO Cis-regulatory elements Heat map. The numbers in this heat map how much repetition of that specific element was found in a gene promoter sequence. Different numerical values have been represented with different colors for easy data analysis.

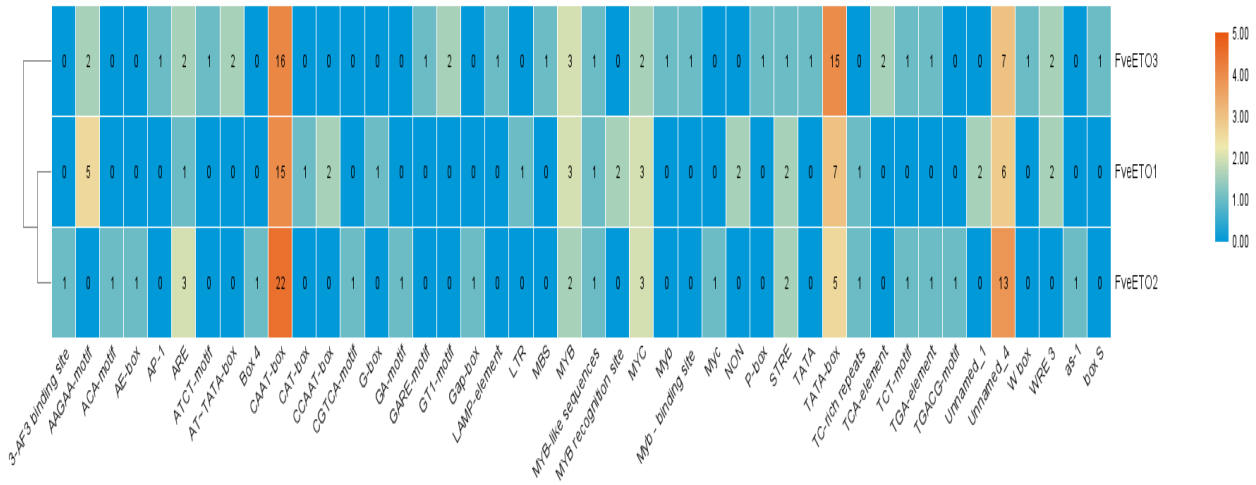

Figure 5c: FveETO Cis-regulatory elements Heat map. The numbers in this heat map how much repetition of that specific element was found in a gene promoter sequence. Different numerical values have been represented with different colors for easy data analysis.

EST analysis

Data about the gene expression was collected from an online database search "Plantdb" using CDS sequences of the genes. This analysis can identify the expression of five genes FeACS3, FeACS4, FeACS5, FeACS9 and FeACS10. Expression of these genes was studied in different plant organ at different life stages. The expression of *FeACS*3, *FeACS*4 and *FeACS*9 genes was identified in "seedlings" under "cold stress". *FeACS*5 gene also expresses at seedling stage but under "drought stress". *FeACS*10 has expression under drought, heat and salt stresses but the stage of life couldn't probe (Fig 6a). This analysis probed the expression of three genes FeACO1, FeACO2 and FeACO5. *FeACO*3 and *FeACO*4 gene expression was not identified. *FeACO*2 gene expresses only at seedling stage under cold stress, *FeACO*5 expresses in cold, drought and heat stressed seedlings and *FeACO*1 have its expression in cold, drought, salt and heat stressed seedlings. *FeACO*1 also expressed at seedlings level when salt and heat stresses were applied simultaneously (Fig 6b). *FeETO*1 gene expression was not found, *FeETO*3 expresses only in cold stress seedlings while *FeETO*2 expresses in cold, salt, heat, salt and heat combine and drought stressed seedlings. Fveerto2 also have its expression at flowering stage of the Fragaria vesca life (Fig 6c).

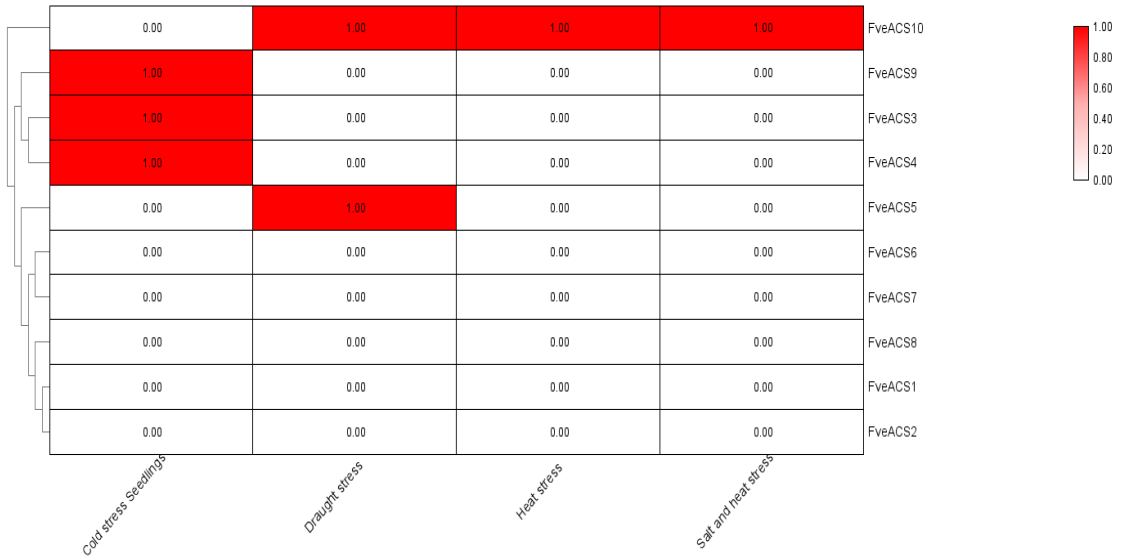

Figure 6a: FveACS EST analysis diagram give us a clear view about the expression of various ACS genes under different stress conditions and enable us to analyze results accurately.

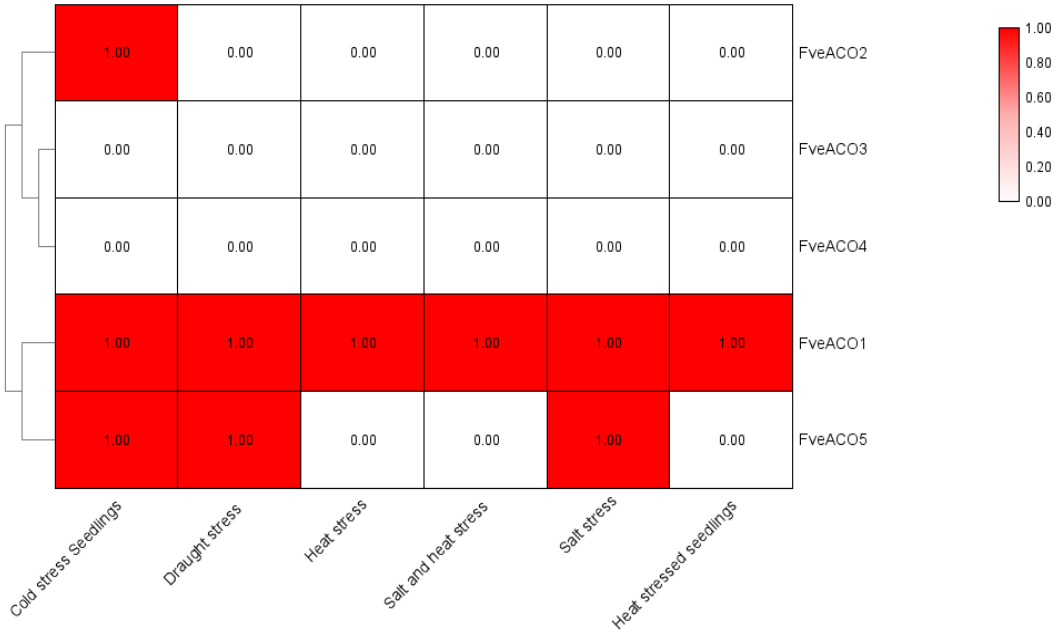

Figure 6b: FveACO EST analysis diagram give us a clear view about the expression of various ACO genes under different stress conditions and enable us to analyze results accurately.

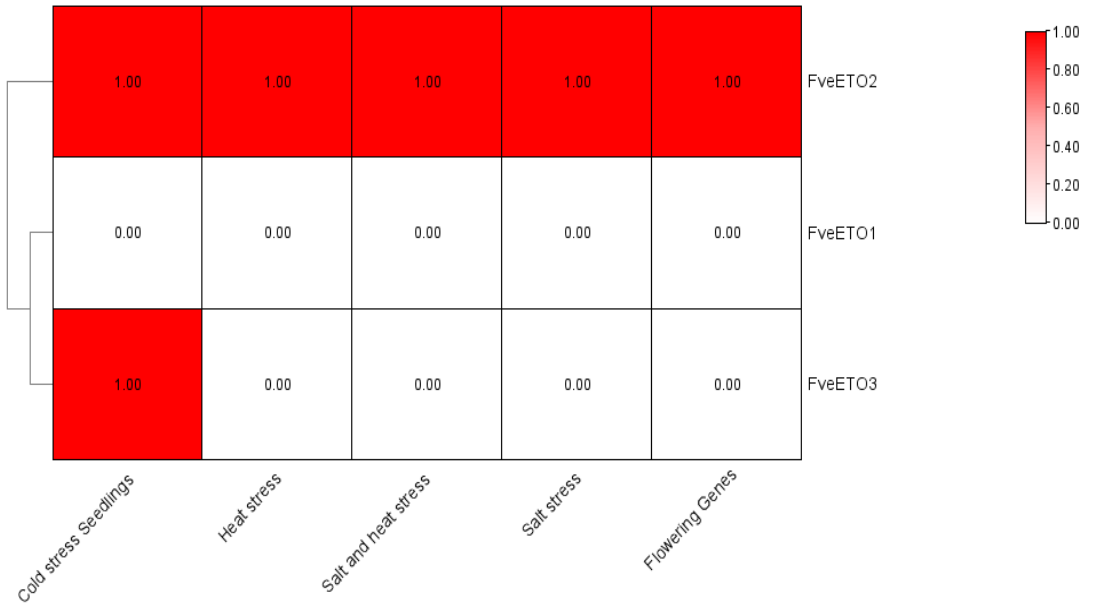

Figure 6c: FveETO EST analysis diagram give us a clear view about the expression of various ETO genes under different stress conditions and enable us to analyze results accurately.

Discussion

Transcription factors (TFs) are important regulatory molecules and are mainly responsible for regulating gene transcription and networking. Characterization and identification of TFs provide a better understanding of the plant growth and development under environmental stimuli (Jones and Vandepoele, 2020; Wen et al., 2016; Yanagisawa and Schmidt, 1999). According to the phylogenetic and domain analysis of *A. thaliana*, citrus (Wu et al., 2016) and eggplant (Wei et al., 2018). In this study, we used the recently released Fragaria vesca genome database (https://Phytozome.jgi.doe.gov/pz/portal.html) to identify 10 ACS, 5 ACO and 3 ETO *genes* at the genome wide level. Using the phylogenetic analysis, 10 FeACS genes were classified into six subfamilies (Group A, B1, B2, C1, C2, C3) (Zou et al., 2013). Motif 1, 2, 3, 4, 6, 7, 8 and 10 were found in all *FeACS* gene sequences, and the same pattern predicted that all *FeACS* sequences might have a common ancestor. In *FeACS* Motif 17, 12, and 14 were identified in all members of Group C showed that these sequences are closer to each other. They might have evolved from each other and have similar functions. Motif 16 was identified only in members of subgroup B1. Motifs 13, 18, and 20 were recognized only in members of subgroup B2. Motif 19 was found only in two members of Group C. Presence of Motif 15 in *FeACS*10 and *FeACS*5 is and evidence of similarity in the structure of these two genes. They might have same ancestor. In *FeACO* Motifs 13 and 20 showed quite surprising results. They were found at similar positions in *FeACO*1 and *FeACO*2, belonging to two groups. *FeACO*1 and *FeACO*2 might have some close or similar ancestor. They depict some similarity in their function also. In *FeETO* and *atETO* genes two members of Group B "*atETO*1_951" and "*atETO*1_956" have same conserved motifs showing that they are more similar than other members. The distribution of motifs among *Fve ACS, ACO and ETO* proteins indicates evolutionary relationship as deduced by phylogenetic tree (Gupta et al., 2015; Malviya et al., 2015).

The motif data analysis by MEME and domain analysis using NCBI CDD and the alignment of *Fve ACS*, protein sequences revealed highly conserved "Aminotransferase_1_2" and "Aminotransferase_1_2 superfamily" domains in all *FeACS* genes in the central region. Any of these two domains is compulsory for a gene sequence to be characterized as ACS. ACS is a PLP (active vitamin B6) dependent enzyme, which act a cofactor for ACS function. These two domains provide a binding site for the cofactor essential for ACS role. In *FeACO* genes sequences total four domains "2OG-FeII_Oxy", "2OG-FeII_Oxy superfamily", "DIOX_N", "DIOX_N superfamily" were found and these four domains were present in all five sequences. Presence of these four domains is compulsory for a gene sequence to be called ACO gene because ACO is a dioxygenase enzyme, and these four domains are compulsory for a dioxygenase enzyme to perform its function. ETO gene play its role to inhibit ethylene production by breaking down ACS protein through ubiquitination. It plays a role of a bridge to bind ubiquitin protein with ACS protein. Two domains "PEP_TPR_lipo superfamily" and "BTB_POZ superfamily" were same in all Fragaria vesca ETO gene sequences. Every ETO gene must have these two domains as these domains bind ubiquitin with other proteins. If anyone of these domains is absent ETO will be unable to perform its function. Although an evolution has been observed

among AthETO and *FeETO* genes. AthETO genes have "PEP_TPR_lipo" and "BTB_POZ" domains not superfamilies like *FeETO* genes. In addition, *Fve ACS ACO and ETO* genes showed structural conservation in subfamilies and was consistent with other plants such as *arabidopsis* (Dong et al., 2016; Lijavetzky et al., 2003; Nasim et al., 2016; Yang and Tuskan, 2006). In addition, as predicted by *in silico* analyses, only "*FeACS*8" showed nuclear localization signals (NLS) predicted using online NLS database (https://rostlab.org/services/nlsdb). The NLSdb signal for *FeACS*8 was found to be "KKHIH. Subcellular localization was predicted using online database WoLF PROST (https://wolfpsort.hgc.jp). WoLF PSORT converts protein amino acid sequences into numerical localization features; based on sorting signals, amino acid composition and functional motifs such as DNA-binding motifs. It gives the data about localization of genes at different sites within a cell.

EST analyses provide important clue about the function of Fve ACS, ACO and ETO genes. Fve ACS, ACO and ETO showed specific and temporal expression different organ and developmental stages. As mentioned earlier, 5 out of 10 FeACS genes was found in the data downloaded from the NCBI GEO datasets experiment (Ramirez-Tejero et al., 2020). From the expression comparison graph, it can be observed that different Fve ACS genes expressed under different stress conditions. *FeACS*2 and *FeACS*3 were expressed in cold stress seedlings although in different groups. *FeACS*5 express under drought stress and drought inducibility regulatory elements found from cis-regulatory analysis. *FeACS*10 has expression under drought, heat and salt stresses but the stage of life couldn't probe (Peng and Weselake, 2011; Ramirez-Parra et al., 2017; Rymen et al., 2017; Xu and Cai, 2019; Xu et al., 2016). This analysis could probe the expression of three FeACO genes FeACO1, FeACO2 and FeACO5. *FeACO*3 and *FeACO*4 gene expression could not identified. *FeACO*2 gene expresses only at seedling stage under cold stress, *FeACO*5 expresses in cold, drought, and heat-stressed seedlings and *FeACO*1 have its expression in cold, drought, salt and heat stressed seedlings. *FeACO*1 was also expressed at seedlings level when salt and heat stresses were applied simultaneously. These genes are generally expressed all over the plant, more specifically in root, stem, leaf, flower, seed, guard cell, plant embryo and pollen (Peng and Weselake, 2011). *FeETO*1 gene expression was not found, *FeETO*3 expresses only in cold stress seedlings while *FeETO*2 expresses in cold, salt, heat, salt and heat combined and drought-stressed seedlings. FvETO2 also has its expression at the flowering stage of the Fragaria vesca life. At the flowering stage, ethylene is not produced in plants, and FeETO2 expression at the flowering stage is evidence that ETO inhibits ethylene production during flowering. It revealed that this gene might have been associated with the Fragaria vesca's reproductive functioning. (Moreno-Risueno et al., 2007; Yanagisawa, 2002a). These results are consistent with their orthologue partner in Group B2, which also showed expression during early flower development in arabidopsis (Wellmer et al., 2006), which is a key process in the life cycle of a plant, during which floral patterning and the specification of floral organs is established (Wellmer et al., 2006). Cis-regulatory analysis also predicts *FeACS2 and FeACS3* also have a role relating to light responsiveness and anaerobic induction. *FeACS*4 also showed good expression under drought inducibility, and cis analysis also found elements related to drought inducibility. *FeACS*4 is closely related to *FeACS*3, which is also expressed under drought stress (Guo, 2009; Miyashima, 2019; Yanagisawa, 2002b). So, it can be inferred that these *FeACS* proteins may also have similar roles and functions in the *Fragaria vesca* plant as of its orthologs in Arabidopsis. In the end, *Fve ACS7* appeared to have slight to moderate expression in various organs and parts of the plant. The Arabidopsis are *ATACS*2, *ATACS*5, and *ATACS*7 regulates a photoperiodic flowering response (Fornara, 2009; Fornara et al., 2009; Imaizumi, 2005).

Conclusion

This study comprehensively analyzed Fve ACS ACO and ETO genes in the *Fragaria vesca* genome. The 10 *FeACS*, 5 *FeACO*, and 3 *FeETO* genes were categorized into subgroups, and each member's structural and functional properties were characterized. Different conserved domains and motifs have been identified to understand gene function and role of these domains in gene function. Same motifs were observed in genes of same subgroups. Cis-regulatory analysis has revealed the binding sites of different transcription factors, thus helped in predicting the expression and suppression of genes under various conditions. Light responsive, anaerobic induction, drought inducibility, gibberellin-responsive, endosperm expression, low-temperature responsive, meristem expression, zein metabolism regulation, seed-specific regulation, defense and stress responsiveness, salicylic acid responsiveness, auxin-responsive element, flavonoid biosynthetic genes regulation, anoxic specific inducibility, cell cycle regulation responsive transcription factors binding sites or cis-regulatory elements have been found in Fve ACS, ACO, and ETO genes. Most of the Fve ACS, ACO, and ETO genes are involved in stress-related responsiveness such as heat, drought, cold, and salinity stress. The detailed computational inspection of Fve ACS, ACO, and ETO proteins revealed in the current study might be selected for cloning purposes at the molecular level, portraying gene expression and studying their interaction with different transcription factors.

Declarations

Data Availability statement

All data generated or analyzed during the study are included in the manuscript.

Ethics approval and consent to participate
Not applicable
Consent for publication
Not applicable
Funding
Not applicable
Conflict of Interest
The authors affirm that the research was conducted without any involvement of commercial or financial relationships that could be perceived as a possible conflict of interest.
Acknowledgments
The University of the Punjab, Lahore, Pakistan supported this work.
References
Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME Suite. *Nucleic acids research* **43**, W39-W49. DOI: https://doi.org/10.1093/nar/gkv416

Cokol, M., Nair, R., and Rost, B. (2000). Finding nuclear localization signals. *EMBO reports* **1**, 411-415. DOI: https://doi.org/10.1093/embo-reports/kvd092

Dong, C., Hu, H., and Xie, J. (2016). Genome-wide analysis of the DNA-binding with one zinc finger (Dof) transcription factor family in bananas. *Genome* **59**, 1085-1100. DOI: https://doi.org/10.1139/gen-2016-0081

Fornara, F. (2009). Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Europe PMC*.

Fornara, F., Panigrahi, K. C., Gissot, L., Sauerbrunn, N., Rühl, M., Jarillo, J. A., and Coupland, G. (2009). Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Developmental cell* **17**, 75-86. DOI: https://doi.org/10.1016/j.devcel.2009.06.015

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *In* "The Proteomics Protocols Handbook" (J. M. Walker, ed.), pp. 571-607. Humana Press, Totowa, NJ. DOI: https://doi.org/10.1385/1-59259-890-0:571

Guo, Y. (2009). Dof5.6/HCA2, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in Arabidopsis. *Europe PMC*. DOI: https://doi.org/10.1105/tpc.108.064139

Gupta, S., Malviya, N., Kushwaha, H., Nasim, J., Bisht, N. C., Singh, V., and Yadav, D. (2015). Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. *Planta* **241**, 549-562. DOI: https://doi.org/10.1007/s00425-014-2239-3

Horton, P., Park, K.-J., Obayashi, T., and Nakai, K. (2006). Protein subcellular localization prediction with WoLF PSORT. *In* "Proceedings of the 4th Asia-Pacific bioinformatics conference", pp. 39-48. World Scientific. DOI: https://doi.org/10.1142/9781860947292_0007

Imaizumi, T. (2005). FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. *Europe PMC*. DOI: https://doi.org/10.1126/science.1110586

Jakubowicz, M. (2002). Structure, catalytic activity and evolutionary relationships of 1-aminocyclopropane-1-carboxylate synthase, the key enzyme of ethylene synthesis in higher plants. *Acta Biochim Pol* **49**, 757-74. DOI: https://doi.org/10.18388/abp.2002_3784

Jones, D. M., and Vandepoele, K. (2020). Identification and evolution of gene regulatory networks: insights from comparative studies in plants. *Curr Opin Plant Biol* **54**, 42-48. DOI: https://doi.org/10.1016/j.pbi.2019.12.008

Kantety, R. V., La Rota, M., Matthews, D. E., and Sorrells, M. E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology* **48**, 501-510. DOI: https://doi.org/10.1023/A:1014875206165

Kong, H., Landherr, L. L., Frohlich, M. W., Leebens-Mack, J., Ma, H., and dePamphilis, C. W. (2007). Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *Plant J* **50**, 873-85. DOI: https://doi.org/10.1111/j.1365-313X.2007.03097.x

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* **35**, 1547-1549. DOI: https://doi.org/10.1093/molbev/msy096

Lijavetzky, D., Carbonero, P., and Vicente-Carbajosa, J. (2003). Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC evolutionary biology* **3**, 17. DOI: https://doi.org/10.1186/1471-2148-3-17

Malviya, N., Gupta, S., Singh, V., Yadav, M., Bisht, N., Sarangi, B., and Yadav, D. (2015). Genome wide in silico characterization of Dof gene families of pigeonpea (Cajanus cajan (L) Millsp.). *Molecular biology reports* **42**, 535-552. DOI: https://doi.org/10.1007/s11033-014-3797-y

Miyashima, S. (2019). Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Europe PMC*. DOI: https://doi.org/10.1038/s41586-018-0839-y

Moreno-Risueno, M. Á., Martínez, M., Vicente-Carbajosa, J., and Carbonero, P. (2007). The family of DOF transcription factors: from green unicellular algae to vascular plants. *Molecular Genetics and Genomics* **277**, 379-390. DOI: https://doi.org/10.1007/s00438-006-0186-9

Nasim, J., Malviya, N., Kumar, R., and Yadav, D. (2016). Genome-wide bioinformatics analysis of Dof transcription factor gene family of chickpea and its comparative phylogenetic assessment with Arabidopsis and rice. *Plant Systematics and Evolution* **302**, 1009-1026. DOI: https://doi.org/10.1007/s00606-016-1314-6

Peng, F. Y., and Weselake, R. J. (2011). Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. *BMC Genomics* **12**, 286. DOI: https://doi.org/10.1186/1471-2164-12-286

Ramirez-Parra, E., Perianez-Rodriguez, J., Navarro-Neila, S., Gude, I., Moreno-Risueno, M. A., and Del Pozo, J. C. (2017). The transcription factor OBP4 controls root growth and promotes callus formation. *New Phytol* **213**, 1787-1801. DOI: https://doi.org/10.1111/nph.14315

Ramirez-Tejero, J. A., Jimenez-Ruiz, J., Leyva-Perez, M. O., Barroso, J. B., and Luque, F. (2020). Gene Expression Pattern in Olive Tree Organs (Olea europaea L.). *Genes (Basel)* **11**. DOI: https://doi.org/10.3390/genes11050544

Rombauts, S., Déhais, P., Van Montagu, M., and Rouzé, P. (1999). PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Research* **27**, 295-296. DOI: https://doi.org/10.1093/nar/27.1.295

Rymen, B., Kawamura, A., Schafer, S., Breuer, C., Iwase, A., Shibata, M., Ikeda, M., Mitsuda, N., Koncz, C., Ohme-Takagi, M., Matsui, M., and Sugimoto, K. (2017). ABA Suppresses Root Hair Growth via the OBP4 Transcriptional Regulator. *Plant Physiol* **173**, 1750-1762. DOI: https://doi.org/10.1104/pp.16.01945

Wei, Q., Wang, W., Hu, T., Hu, H., Mao, W., Zhu, Q., and Bao, C. (2018). Genome-wide identification and characterization of Dof transcription factors in eggplant (Solanum melongena L.). *PeerJ* **6**, e4481. DOI: https://doi.org/10.7717/peerj.4481

Wellmer, F., Alves-Ferreira, M., Dubois, A., Riechmann, J. L., and Meyerowitz, E. M. (2006). Genome-wide analysis of gene expression during early Arabidopsis flower development. *PLoS Genet* **2**, e117. DOI: https://doi.org/10.1371/journal.pgen.0020117

Wen, C. L., Cheng, Q., Zhao, L., Mao, A., Yang, J., Yu, S., Weng, Y., and Xu, Y. (2016). Identification and characterisation of Dof transcription factors in the cucumber genome. *Sci Rep* **6**, 23072. DOI: https://doi.org/10.1038/srep23072

Wu, J., Fu, L., and Yi, H. (2016). Genome-Wide Identification of the Transcription Factors Involved in Citrus Fruit Ripening from the Transcriptomes of a Late-Ripening Sweet Orange Mutant and Its Wild Type. *PLoS One* **11**, e0154330. DOI: https://doi.org/10.1371/journal.pone.0154330

Xu, P., and Cai, W. (2019). Nitrate-responsive OBP4-XTH9 regulatory module controls lateral root development in Arabidopsis thaliana. *PLoS Genet* **15**, e1008465. DOI: https://doi.org/10.1038/srep27705

Xu, P., Chen, H., Ying, L., and Cai, W. (2016). AtDOF5.4/OBP4, a DOF Transcription Factor Gene that Negatively Regulates Cell Cycle Progression and Cell Expansion in Arabidopsis thaliana. *Sci Rep* **6**, 27705. DOI: https://doi.org/10.1016/S1360-1385(02)02362-2

Yanagisawa, S. (2002a). The Dof family of plant transcription factors. *Trends in plant science* **7**, 555-560.

Yanagisawa, S. (2002b). The Dof family of plant transcription factors. *Europe PMC*.

Yanagisawa, S., and Schmidt, R. J. (1999). Diversity and similarity among recognition sequences of Dof transcription factors. *The Plant Journal* **17**, 209-214. DOI: https://doi.org/10.1046/j.1365-313X.1999.00363.x

Yang, X., and Tuskan, G. A. (2006). Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. *Plant physiology* **142**, 820-830. DOI: https://doi.org/10.1104/pp.106.083642

Zou, H. F., Zhang, Y. Q., Wei, W., Chen, H. W., Song, Q. X., Liu, Y. F., Zhao, M. Y., Wang, F., Zhang, B. C., Lin, Q., Zhang, W. K., Ma, B., Zhou, Y. H., Zhang, J. S., and Chen, S. Y. (2013). The transcription factor AtDOF4.2 regulates shoot branching and seed coat formation in Arabidopsis. *Biochem J* **449**, 373-88. DOI: https://doi.org/10.1042/BJ20110060