*Biotech Studies 30(2), 79-85*
https://doi.org/10.38042/biotechstudies.977788

BIOTECH STUDIES

RESEARCH PAPER

# Genome-wide analysis of *Fragaria vesca* Three-Amino-Acid-Loop-Extension (TALE) genes

**Gizem Kabak[1], Seray Şahsuvar[1], Sıla Turgut[1], Şeyma Gökdemir[1,*]**

[1]Department of Molecular Biology and Genetics, Baskent University, Bağlıca, 06790 Ankara, Turkey.

**How to cite**

Kabak, G., Şahsuvar, S., Turgut, S., & Gökdemir, Ş. (2021). Genome-wide wnalysis of *Fragaria vesca* Three-Amino-Acid-Loop-Extension (TALE) genes. *Biotech Studies. 30*(2), 79-85, https://doi.org/10.38042/biotechstudies.977788

**Article History**
Received 24 January 2021
Accepted 15 July 2021
First Online 29 July 2021

**Corresponding Author**
Tel.: +90 554 778 78 83
E-mail:
fsgokdemir@baskent.edu.tr

**Keywords**
TALE
*Fragaria vesca*
Genome-wide identification

## Abstract

The present study is aimed to identify and characterize the three-amino-acid-loop-extension (TALE) genes in *Fragaria vesca* as bioinformatics. TALE superclass homeoproteins have important roles in regulating certain signal pathways in the plant system. However, there is no knowledge about the role of *TALE* genes in *Fragaria vesca*. According to this study, a total of 18 candidate *FvescaTale* genes were identified. Identification of motifs, exon and intron analysis, genome mapping, localization in the cell, three-dimensional modeling, and ontology analysis were made according to these genes. This bioinformatic analysis revealed that *FvescaTale* genes might play an important role in stress response for *Fragaria vesca* cultivars and suggests that these genes could be used as functional markers for *in silico* analysis for future studies.

## Introduction

*Fragaria vesca*, a diploid (2n=14) type of strawberry, is a plant that belongs to the Rosaceae family (Li et al., 2019; Shulaev et al., 2011). *Fragaria vesca* is a model used in this family due to its ease of proliferation, small genome length, and short generation cycle compared to other plants. Despite all these qualities of *Fragaria vesca*, its genome was sequenced and revealed in 2010 (Shulaev et al., 2011). *Fragaria* has not a real fruit, it consists of seeds that modified the surface of the reservoir formed at the tip of the shoot. Strawberry, which bred from four different diploid ancestors, is important in genomic research both phylogenetically and because it has a small genome. It has an extensive genotypic diversity thanks to its optimum growth range, vegetative reproduction and many positive recessive properties. Therefore, it is an important plant to determine the genomic characteristics of the Rosaceae family (Darrow, 1966; Shulaev et al., 2011).

A homeobox (HB) encodes a homeodomain (HD) region that is 60 amino acids in length and interacts with DNA specifically. This region is conserved in transcription factors (TF) in all eukaryotic organisms (Ariel et al., 2007). The protein class of homeoproteins called the three-amino-acid-loop-extension (TALE) has been shown to direct organ morphogenesis, meristem continuity or formation, various properties of the reproductive phase, and organ orientation (Hamant & Pautot 2010). Homeobox proteins have been shown to have a TALE superclass and are recognized by an extension of three amino acids (Pro-Tyr-Pro) between α-helices 1 and 2 in the homeodomain. Genes encoding these proteins are highly conserved. It has been observed that plants, fungi, and animals also have transcriptional regulatory functions in their common ancestors and they are crucial for signaling and communication network (Chen, 2003; Burglin, 1997).

Knotted 1-like homeobox (KNOX) proteins, a homeodomain transcription factor, control genes that regulate hormone homeostasis in the shoot apical meristems of plants. KNOX genes contain the TALE

homeobox gene family, which acts as a regulator in the diploid development of plants (Hay & Tsiantis 2010). Homeobox genes play an important role in transcriptional regulation in various plants, shoot apical development and flowering, lignin and cellulose accumulation, cell wall biosynthesis, plant growth and development under high temperature and humidity stress (Rutzens et al., 2009; Hirano et al., 2013; Liu et al., 2014). There are some studies on TALE members, but functional genomic studies are lacking in most plants. Therefore, elucidating the function of the TALE family in plants provides an important genomic resource for genome-wide analysis (Ma et al., 2019).

The aim of study is to define and characterize the TALE family, one of the genes to be used at the transcriptional level in future biotic and abiotic stress studies to be carried out with *Fragaria vesca*. Thus, the genomic functions of the relevant genes will be known and the results will be evaluated *in silico*. This study is an important preliminary resource that includes many bioinformatic analysis for future studies. According to this research, genome-wide analyzes were performed for *F. vesca* TALE genes using various bioinformatics tools. As a result, it is suggested that these genes can be used as functional markers *in silico* analysis for future studies.

## Materials and Methods

### Identification of TALE genes in *Fragaria vesca*

*F. vesca* TALE protein sequences were retrieved from Phytozome database v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html) using keywords in the search with Plant Transcription Factor Database (http://planttfdb.cbi.pku.edu.cn/) and Pfam Database (http://pfam.xfam.org/). To identify TALE proteins in the *F. vesca* genome, both blastp at Phytozome database v12.1 (http://www.phytozome.net) and Hidden Markov model (HMM), (http://www.ebi.ac.uk) searches were performed against the *F. vesca* genome.

The solid and chemical traits of TALE proteins in *F. vesca* were identified using the ProtParam tools (http://web.expasy.org/protparam/) such as: molecular weight, atomic composition, extinction coefficient, theoretical pI, estimated half-life, amino acid composition, aliphatic index, instability index and grand average of hydropathicity.

### TALE proteins of identification of motif paterns, locations, and 3D modelling

Motifs present in the TALE protein families were identified by using the Multiple Expectation Maximization for Motif Elicitation (MEME) tool (http://meme-suite.org/). Protein is classified by families, predicting domains, and important sites by using Interpro Tool (https://www.ebi.ac.uk/interpro/).

Three Dimensional (3D) Structures for all the proteins were reported by using Phyre2 Tool. (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) Phyre2 is also used for functions and mutations.

### Phylogenetic analysis, physical, chromosomal location

Phylogenetic analysis was reported by MegaX tool for 18 TALE genes. Gene Structure Display Server program tool (GSDS; http://gsds.cbi.pku.edu.cn/) was used to estimate the exon/intron set up of the TALE genes. The location of the genes on the chromosome was determined by using MapGene2 tool (http://mg2c.iask.in/mg2c_v2.0/).

### Examination of *cis*-regulatory elements in the promoter region and subcellular localization

2 kb sequences in the upstream region of the 18 genes were analyzed in the PlantCARE program (http://sphinx.rug.ac.be:8080/PlantCARE/).

WoLF PSORT tool was used to analyze where TALE genes are located in the model organism (https://wolfpsort.hgc.jp/). WoLF PSORT converts amino acid sequences of a protein into numerical localization properties so the results show where genes are located.

### Plant small RNA target analysis

psRNATarget was used for the relationship between the transcript of the gene and the miRNA targeted for silencing. The psRNATarget analysis program includes the latest innovations in miRNA target recognition in the plant. This program was used to display sRNA targets in the plant by finding the match found between the sRNA sequence and the target mRNA sequence (https://plantgrn.noble.org/psRNATarget/).

### Ontology analysis

Ontology analysis, biological and molecular functions were reported by using AgriGO tool. It is a major bioinformatics initiative to combine gene and gene product characteristics across all species. This tool supports special focus on agricultural species. (http://bioinfo.cau.edu.cn/agriGO/)

## Results

### Analysis of TALE gene family in *Fragaria vesca*

As a result of screening and profiling the TALE gene from the existing protein databases, the presence of eighteen proteins for *Fragaria vesca* were found. The protein sequences of these genes were obtained from Phytozome database v12.1. Firstly, these sequences were analyzed by ProtParam. The results were given in Table 1. Accordingly, the amino acid length of the *FvescaTale.1* gene is found to be 670. This gene is 73882.80 kDA, and it has 5.67 pI. This gene is unstable and the instability index is 48.65. According to our data, only *FvescaTale.14* gene is found stable with 38.27 instability index.

**Table 1.** Information of TALE genes in Fragaria vesca along with their gene codes, number/length of amino acids, mass (kDa), pI, stability and instability index

| Gene Code | Amount of Amino acid | Mass (kDa) | pI | Stability | Instability Index |
|---|---|---|---|---|---|
| *FvescaTale.1* | 670 | 73882.80 | 5,67 | unstable | 48.65 |
| *FvescaTale.2* | 809 | 88184.05 | 6,04 | unstable | 43.11 |
| *FvescaTale.3* | 662 | 74135.92 | 5,41 | unstable | 49.79 |
| *FvescaTale.4* | 598 | 66747.32 | 6,91 | unstable | 56.50 |
| *FvescaTale.5* | 399 | 44532.13 | 6,28 | unstable | 49.14 |
| *FvescaTale.6* | 815 | 87527.38 | 6,73 | unstable | 53.86 |
| *FvescaTale.7* | 216 | 24764.15 | 6,31 | unstable | 48.22 |
| *FvescaTale.8* | 406 | 46114.01 | 8,85 | unstable | 54.04 |
| *FvescaTale.9* | 795 | 88721.63 | 5,58 | unstable | 52.68 |
| *FvescaTale.10* | 318 | 35355.75 | 8,70 | unstable | 59.47 |
| *FvescaTale.11* | 470 | 52130.66 | 6,79 | unstable | 43.53 |
| *FvescaTale.12* | 933 | 103801.21 | 7,86 | unstable | 39.27 |
| *FvescaTale.13* | 477 | 54344.88 | 7,74 | unstable | 48.12 |
| *FvescaTale.14* | 330 | 37277.88 | 5,04 | stable | 38.27 |
| *FvescaTale.15* | 184 | 20981.38 | 5,66 | unstable | 60.73 |
| *FvescaTale.16* | 391 | 44500.16 | 5,62 | unstable | 41.34 |
| *FvescaTale.17* | 323 | 36600.06 | 5,48 | unstable | 48.13 |
| *FvescaTale.18* | 289 | 32623.65 | 6,24 | unstable | 57.46 |

## Identification and domain relationships of motifs in the TALE family in *Fragaria vesca*

Using MEME program, eighteen motifs were identified by predicting the motifs of TALE genes in the *Fragaria vesca* family. The motifs of the seven genes between the *FvescaTale.12* and *FvescaTale.18* are colorless unlike the other motifs. This is due to the lack of connection between homologs and the irregularity of motifs. In addition, motif locations are checked over MEME Suite.

The domain relationships of the motifs were analyzed via InterProScan program. Figure 1 has showed eleven TALE proteins logo patterns. The other proteins have unstable sequences. They have been looked uninspiring as a result of MEME analysis. All logo patterns have diffirent aminoacid configurations (Figure 1). But most of them contain similar consensus sequence.

When Figure 2 is examined, it is observed that FvescaTale.1, FvescaTale.2, FvescaTale.3 and FvescaTale.4 motifs have homology similarity. Likewise, similar motifs were observed in FvescaTale.16, FvescaTale.17 and FvescaTale.18 due to the presence of domains in close proximity (Figure 2).

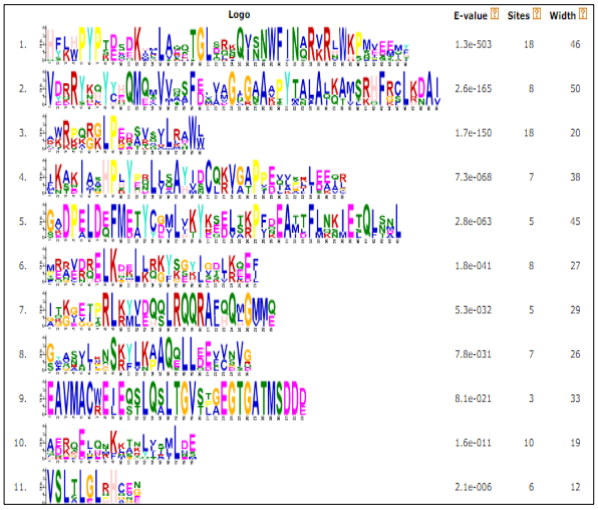


**Figure 1.** Motif sequences analysis in amino acid sequences of TALE from *Fragaria vesca.*

## Exon and intron analysis of TALE gene family in *Fragaria vesca*

In the Supplementary data 1, the yellow areas indicate exons, while the black lines indicate introns. The positions and numbers of the exon and intron regions on the TALE genes of *Fragaria vesca* were shown.

This analysis was carried out by separating exon and intron from GeneStructure Display. The importance of exon and intron number gives the relationship between genes. Four exons and three introns are found in the *FvescaTale.8* gene, eleven exons and ten introns are found in *FvescaTale.12* and finally, five exons and four introns are found in the *FvescaTale.18* gene. The

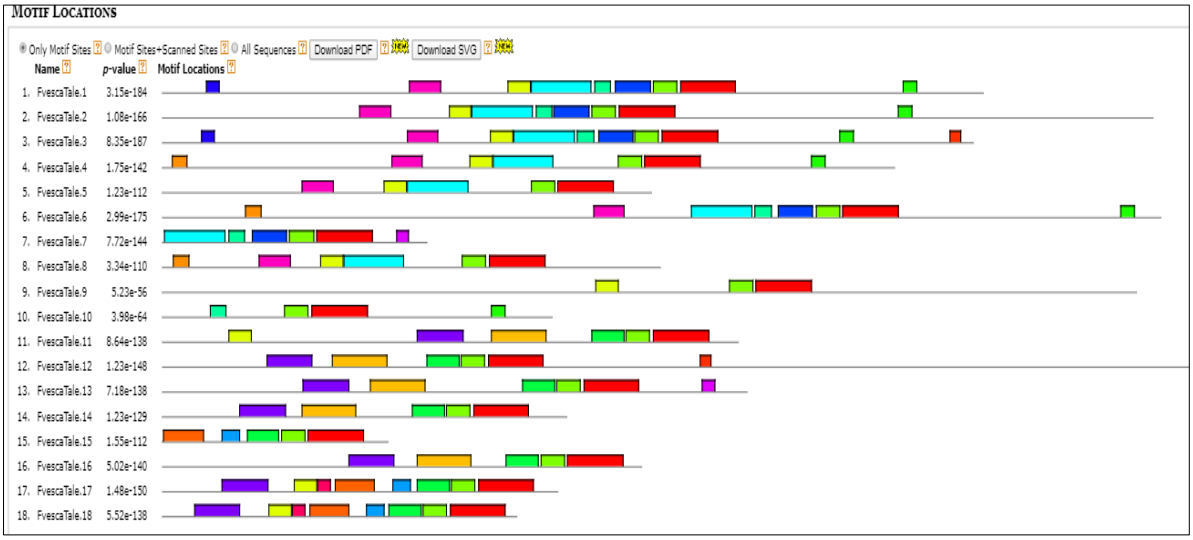


**Figure 2.** Motif locations analysis in amino acid sequences of TALE from *Fragaria vesca.*

importance of intron numbers is still controversial. However, the length of existing introns indicates that the gene region will be stable and conserved. Sequences of preference for gene expression and RNAseq analysis e with long introns. *FvescaTale.2, FvescaTale.12* and *FvescaTale.16* have long intron regions (Supplementary data 1).

**Genome mapping of TALE gene family in *Fragaria vesca***

The studies on MapGene2 were shown in Figure 3. Eighteen TALE transcription factors identified on the 7 chromosomes of *F. vesca* were distributed. According to these distributions, *FvescaTale.17* and *FvescaTale.14* genes were found on chromosome 1. *FvescaTale.1, FvescaTale.15*, *FvescaTale.9, FvescaTale.3* and *FvescaTale.5* genes were localized on chromosome 2. Chromosome 3 contained the *FvescaTale.11* and *FvescaTale.20* genes. *The FvescaTale.16* gene was found on the chromosome 4.

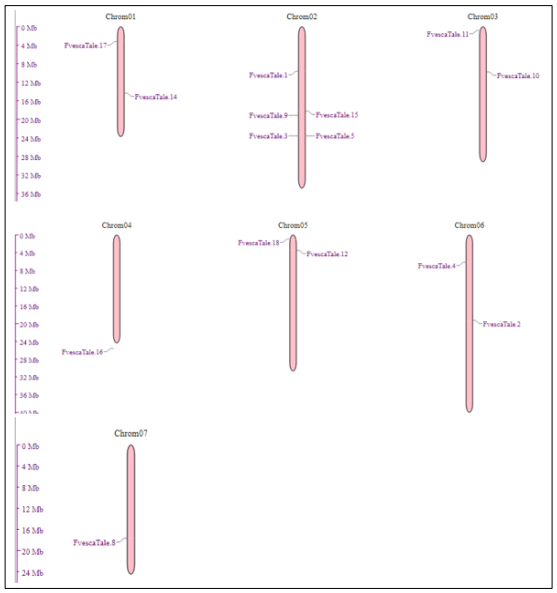


**Figure 3.** Genome mapping by positioning the TALE family on chromosomes.

*The FvescaTale.18* and *FvescaTale.12* genes were observed on the chromosome 5. *FvescaTale.4* and *FvescaTale.2* are located in the chromosome 6. The *FvescaTale.8* gene was localized on the chromosome 7. *Fragaria vesca* has small genome, but it appears that genes are largely distributed on different chromosomes. This is proof that it is important evolutionarily.

**Distribution of TALE gene family in *Fragaria vesca* in the cell**

The distribution of Tale genes in *Fragaria vesca* in the cell is given in Table 2. Cellular localizations of the eighteen genes belonging to *F. vesca* were evaluated by WoLF PSORT. According to Table 2, 7.5 of the *FvescaTale.2* gene were found in the cytoplasm and 13.5 were found in the nucleus. Fourteen of the *FvescaTale.5* gene were localized in the nucleus. While thirteen of the *FvescaTale.10* gene were in the nucleus and one in the chloroplast. Thirteen of the *FvescaTale.14* gene were observed in the nucleus and one in the peroxisome (Table 2).

**Table 2.** Distribution of TALE family in *Fragaria vesca* in the cell compartments

| Gene names | Cytoplasm | Nucleus | Peroxisome | Chloroplast |
|---|---|---|---|---|
| ***FvescaTale1*** | | 14 | | |
| ***FvescaTale2*** | 7.5 | 13.5 | | |
| ***FvescaTale3*** | 7.5 | 13.5 | | |
| ***FvescaTale4*** | | 14 | | |
| ***FvescaTale5*** | | 14 | | |
| ***FvescaTale6*** | 7.5 | 13.5 | | |
| ***FvescaTale7*** | | 13 | 1 | |
| ***FvescaTale8*** | | 14 | | |
| ***FvescaTale9*** | | 13 | 1 | |
| ***FvescaTale10*** | | 13 | | 1 |
| ***FvescaTale11*** | 7.5 | 13.5 | | |
| ***FvescaTale12*** | | 14 | | |
| ***FvescaTale13*** | 7.5 | 13.5 | | |
| ***FvescaTale14*** | | 13 | 1 | |
| ***FvescaTale15*** | | 14 | | |
| ***FvescaTale16*** | | 14 | | |
| ***FvescaTale17*** | | 14 | | |
| ***FvescaTale18*** | | 14 | | |

**3D modeling of TALE proteins in *Fragaria vesca***

The 3D motifs of the eighteen TALE proteins studied were analyzed using the Phyre2 program. These results were shown in Figure 4. According to the data, 38% of the remnants of FvescaTale.5 were modeled with >90% confidence. It was estimated that 54% of the series are irregular (Figure 4).

**Ontology analysis of TALE gene family in *Fragaria vesca***

The results of AgriGO are attached in the form of a table in the Supplementary data 2. AgriGO database is used to determine ontology analysis of genes. According to Figure 5, 80% of our genes participate in cellular processes while 15% have function. Together with these, 5% function as a cellular components (Supplementary data 2).

**Phylogenetic analysis of Tale proteins family in *Fragaria vesca***

In the MegaX program, relationships between 18 TALE proteins were determined. According to the result shown in Figure 6, the relationship between FvescaTale.14 and FvescaTale.5 proteins were remote, while FvescaTale.11 and FvescaTale.1 were closely related.

Eighteen TALE proteins belonging to *F. vesca* organism and *TALE* proteins found in *O. brachyantha* and *F. ananassa* organisms were compared. According to Figure 7, FvescaTale.11 and FananassaTale.8 were closely related. FananassaTale.6 and ObrachyanthaTale.15 have been observed in distant association.

The eighteen TALE genes from the *F. vesca* organism were compared with miRNA sequences from other plant species via psRNATarget. The results are given in the Supplementary data 3.

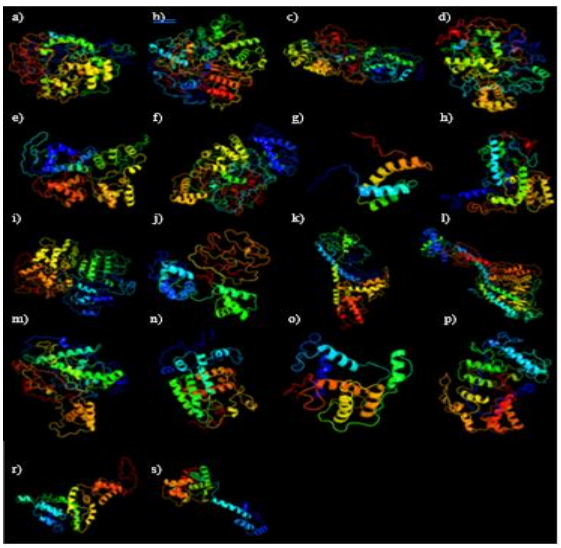

**Figure 4.** 3D modeling of TALE proteins a)FvescaTale1, b)FvescaTale2, c)FvescaTale3, d)FvescaTale4, e)FvescaTale5, f)FvescaTale6, g)FvescaTale7, h)FvescaTale8, i)FvescaTale9, j)FvescaTale10, k)FvescaTale11, l)FvescaTale12, m)FvescaTale13, n)FvescaTale14, o)FvescaTale15, p)FvescaTale16, r)FvescaTale17, s)FvescaTale18. Image coloured by rainbow N → C terminus.

According to the data, 1759 relationships were found (Supplementary data 3). Two of the *FvescaTale.4* genes were associated with the *Arabidopsis thaliana* miRNA. Similarly, 3 of the *FvescaTale.14* genes and one of the *FvescaTale.16* gene were associated with *Oryza sativa*.

## Discussion

In this present study, a total of eighteen candidate *FvescaTale* genes were identified. While the lowest number of *FvescaTale* genes was on chromosomes 4 and 7 (one Tale gene), the highest number of *FvescaTales* was on chromosome 2 (five TALE genes). One gene, *FvescaTale.7,* was found to have no-introns. *FvescaTale.12* had the highest number of introns (10 introns).

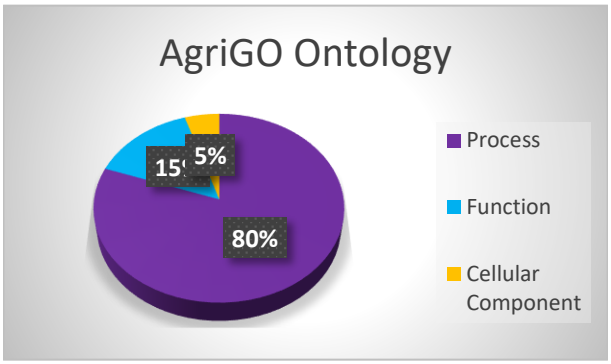


**Figure 5.** AgriGO Ontology graphical datas.

Even if the intron number does not give clear information, the lengths of the introns are important. In a possible gene expression, gene regions with long introns are preferred. Because it is more stable. In addition, all TALE genes were found in the nucleus, while *FvescaTale.2, 3, 6, 11* and *13* were found in the cytoplasm. Moreover, studies have been carried out on the 3-dimensional structures and ontology analysis of these genes. Phylogenetic studies of these 18 genes have been carried out and comparisons with different organisms have also been made. When the motif patterns of the proteins identified according to the results were compared, it was seen that the members with similar motif patterns were located in a close cluster on the phylognetic tree. Although protein domains are not characterized by a known family, it has been observed that the figures obtained have similar motif locations with the beta helix loop helix. The fact that the 18 TALE proteins have different motif patterns indicates that the *cis*-regulatory elements in the upstream region are very different. Phylogenetic tree data also support this hypothesis. Because each tree branch is clustered at a certain distance from each other.

After the whole genome sequence analysis is done, definitions and characterizations should be made on certain parts of the genome in order to determine the

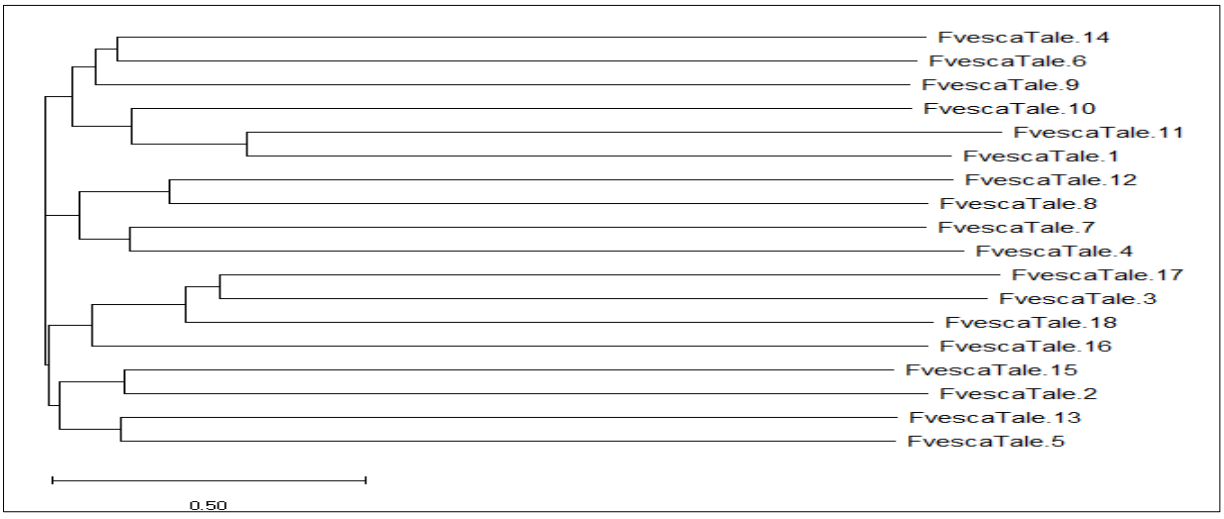


**Figure 6.** Phylogenetic tree of TALE proteins.

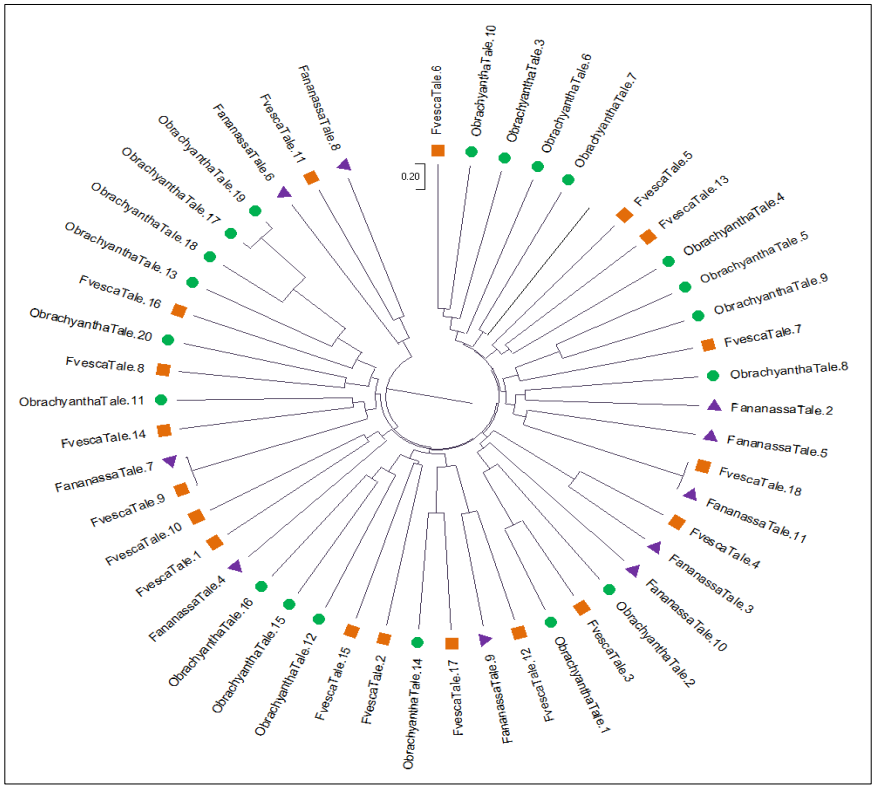


**Figure 7.** Phylogenetic circle tree of TALE proteins in *F. vesca, O.brachyantha* and *F.ananassa.*

genomic features functionally. Compared to existing studies, the data obtained allows us to infer whether the analyzed region of the genome is involved in any stress-related pathway. For example, after the genome sequences of 4 different cotton species were completed, the development of functional genomic analyzes of cotton was proposed and the genome-wide characterization of the genes of the TALE family was made. Following this process, it has been demonstrated that *TALE* genes regulate secondary cell wall synthesis. Unlike *F. vesca*, 46, 47, 94 and 88 *TALE* genes were identified, respectively (Ma et al., 2019).

*TALE* homeobox genes are an important group of genes that play a role in developmental processes in both plants and animals. It is also expressed from time to time in the early development stages of some living things. A study conducted on Spiralians has also been shown to play an important role in the cellular separation mechanism (Morino et al. 2017). Since we know that these genes are expressed along the plant-animal line, *TALE* genes are also very important for the evolutionary process. Especially, thanks to common *cis*-regulatory elements found in both animals and plants, genomic inference can be made. Related the elements have some common functions such as transcriptional regulation, *cis*-regulator regions, stress-related processes, functional processes, circadian rhythm processes and etc. Therefore, it enables inferences in plant-animal interaction and genome-wide association studies.

It is not known whether *TALE* genes, which are expressed in almost every living organisms, arise from a single ancestor or independently of the ancestral *TALE* genes. Therefore, it is necessary to determine the role of *TALE* genes in both plant and especially stress response pathways and plant-animal interactions with common conserved protein domains and *cis*-regulatory elements. Identification and characterization of *TALE* genes at the stages of the evolutionary process at the species and subspecies level will at least provide an elucidation of the mechanism in the pathways that respond to stress.TALE homodomains contain polar residues such as glutamine, lysine, cysteine, histidine or serine. These small polar residues show that the DNA-Protein interactions of *TALE* genes are very different. This proves the existence of a species-specific transcriptional arrangement. Conserved motifs are found between TALE and other homodomain genes in similar pathways. The ELK region in the *Zea mays* plant is an example. In addition, at least 1 intron position is conserved within the KNOX genes. There is a protected GSE-box area within the KNOX and ELK area (Akam 1993).

The determined intron positions do not shed much light on the evolutionary distinction, but indicate that genes belonging to two different subclasses in the TALE family, such as KNOX and BEL, which are expressed in similar pathways and have the same homodoma, may be expressed at different levels in different organisms. This shows that these genes are actually different gene groups specific to each organism (Reiser et al. 1995; Bürglin 1997). This situation shows the importance of genome-wide characterization and identification studies.

*In silico* analysis enables understanding of DNA-protein interactions. Here, we made the genome-wide *in silico* analysis of the *Fragaria* and *TALE* genes, which is a plant species. Although it gives us limited information on the evolutionary process, we were able to obtain important reference data for plant-plant or plant-animal interreactions. For *in-vivo* analyzes to be performed with *TALE* genes in the future, data defined at the transcriptional level will be available. In order to determine the efficiency of *TALE* genes in stress-related pathways, *in silico* analysis of the relevant genes should be done before qrt PCR studies. Therefore, this study is a preliminary study to understand transcriptional regulation at the level of gene expression. Besides, this bioinformatic analysis revealed that *FvescaTale* genes might play an important role in stress response for *Fragaria vesca* cultivars and suggests that these genes could be used as functional markers for *in silico* analysis for future studies.

## Conclusion

In summary, this study provides important clues for further elucidating the functions of *TALE* genes regulating and development in *Fragaria vesca*. In future studies, researchers could use this information to correlate results regarding gene expression. Therefore, this study is part of the genome-wide identification and characterization step *in silico*. The characteristics of the TALE homeodomain family in *F. vesca* have been determined bioinformatically.

## Author Contributions

Conceptualization Ideas; FŞG, Data Curation; GK, SŞ, ST, FŞG, Formal Analysis; GK, SŞ, ST, FŞG Investigastion; GK, SŞ, ST, FŞG, Methodology: GK, SŞ, ST, FŞG, Validation; GK, SŞ, ST, FŞG, Visualization; GK, SŞ, ST, FŞG, Writing-original Draft Preparation; GK, SŞ, ST, FŞG.

## Additional Information

Supplementary data accompanies this paper at https://biotechstudies.org/uploads/BIO-126_Supp1.pdf https://biotechstudies.org/uploads/BIO-126_Supp2.pdf https://biotechstudies.org/uploads/BIO-126_Supp3.pdf

## References

Akam, M. (1993). Evolutionary conservation of developmental mechanisms: edited by Allan Spradling, Wiley-Liss, 1993. $98.00 (xii+ 219 pages) ISBN 0 471 58843 1.

Ariel, F. D., Manavella, P. A., Dezar, C. A., & Chan, R. L. (2007). The true story of the HD-Zip family. Trends in Plant Science, *12*(9), 419–426. https://doi.org/10.1016/j.tplants.2007.08.003

Burglin, T. R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Res*. 25:4173–4180. https://doi.org/10.1093/nar/25.21.4173

Chen, H. (2003). Interacting Transcription Factors from the Three-Amino Acid Loop Extension Superclass Regulate Tuber Formation. *Plant Physiology*, *132*(3), 1391–1404. https://doi.org/10.1104/pp.103.022434

Darrow, G. M. (1966). The strawberry. History, breeding and physiology. *The strawberry. History, breeding and physiology.* https://doi.org/10.1104/pp.103.022434

Hamant, O., & Pautot, V. (2010). Plant development: a TALE story. *Comptes rendus biologies*, *333*(4), 371-381. https://doi.org/10.1016/j.crvi.2010.01.015

Hay, A., & Tsiantis, M. (2010). KNOX genes: versatile regulators of plant development and diversity. *Development*, *137*(19), 3153–3165. https://doi.org/10.1242/dev.030049

Hirano, K., Kondo, M., Aya, K., Miyao, A., Sato, Y., Antonio, B. A., Namiki, N., Nagamura, Y., & Matsuoka, M. (2013). Identification of transcription factors involved in rice secondary cell wall formation. *Plant and Cell Physiology*, *54*(11), 1791-1802. https://doi.org/10.1093/pcp/pct122

Li, Y., Pi, M., Gao, Q., Liu, Z., & Kang, C. (2019). Updated annotation of the wild strawberry *Fragaria vesca* V4 genome. *Horticulture Research*, *6*(1). https://doi.org/10.1038/s41438-019-0142-6

Liu, Y., You, S., Taylor-Teeples, M., Li, W. L., Schuetz, M., Brady, S. M., & Douglas, C. J. (2014). BEL1-LIKE HOMEODOMAIN6 and KNOTTED ARABIDOPSIS THALIANA7 interact and regulate secondary cell wall formation via repression of REVOLUTA. *The Plant Cell*, *26*(12), 4843-4861. https://doi.org/10.1105/tpc.114.128322

Ma, Q., Wang, N., Hao, P., Sun, H., Wang, C., Ma, L., Wang, H., Zhang, X., Wei, H., & Yu, S. (2019). Genome-wide identification and characterization of TALE superfamily genes in cotton reveals their functions in regulating secondary cell wall biosynthesis. *BMC plant biology*, *19*(1), 1-20. https://doi.org/10.1186/s12870-019-2026-1

Morino, Y., Hashimoto, N., & Wada, H. (2017). Expansion of TALE homeobox genes and the evolution of spiralian development. *Nature ecology & evolution, 1*(12), 1942-1949. https://doi.org/10.1038/s41559-017-0351-z

Reiser, L., Modrusan, Z., Margossian, L., Samach, A., Ohad, N., Haughn, G. W., & Fischer, R. L. (1995). The BELL1 gene encodes a homeodomain protein involved in pattern formation in the Arabidopsis ovule primordium. *Cell*, *83*(5), 735-742. https://doi.org/10.1016/0092-8674(95)90186-8

Rutjens, B., Bao, D., Van Eck-Stouten, E., Brand, M., Smeekens, S., & Proveniers, M. (2009). Shoot apical meristem function in Arabidopsis requires the combined activities of three BEL1-like homeodomain proteins. *The Plant Journal*, *58(*4), 641-654. https://doi.org/10.1111/j.1365-313X.2009.03809.x

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S. P., Burns, P., Davis, T. M., Slovin, J. P., Bassil, N., Hellens, R. P., Evans, C., Harkins, T., Kodira, C., Desany, B., Crasta, O. R., … & Folta, K. M. (2011). The Genome of Woodland Strawberry *(Fragaria Vesca)*. *Nature Genetics*, *43*(2), 109-16. DOI.org (Crossref), https://doi.org/10.1038/ng.740