www.ThePharmaJournal.com

The Pharma Innovation

**ISSN (E): 2277- 7695**
**ISSN (P): 2349-8242**
**NAAS Rating: 5.03**
TPI 2019; 8(1): 422-428

www.thepharmajournal.com
Received: 13-11-2018
Accepted: 16-12-2018

**Suman Dutta**
ICAR-Indian Agricultural Research Institute, New Delhi, India

**Vignesh Muthusamy**
ICAR-Indian Agricultural Research Institute, New Delhi, India

**Rajkumar U Zunjare**
ICAR-Indian Agricultural Research Institute, New Delhi, India

**Rakesh Bhowmick**
ICAR-National Research Centre for Plant Biotechnology, New Delhi, India

**Firoz Hossain**
ICAR-Indian Agricultural Research Institute, New Delhi, India

**Correspondence**
**Vignesh Muthusamy**
ICAR-Indian Agricultural Research Institute, New Delhi, India

# Genome wide study of *fatty acid hydroxylase* (*FAH*) superfamily containing β-*carotene hydroxylase* (*crtRB1*) in maize (*Zea mays* L.)

**Suman Dutta, Vignesh Muthusamy, Rajkumar U Zunjare, Rakesh Bhowmick and Firoz Hossain**

**Abstract**
The *Fatty acid hydroxylase* (*FAH*) superfamily in plant is a broad group of proteins comprising carotene and fatty acid hydroxylases, and sterol desaturases. Favourable allele of *FAH* superfamily gene, β-*carotene hydroxylase* (*crtRB1*) was introgressed through marker assisted selection for enrichment of provitamin-A in maize. In this study, putative *FAH* genes were identified in maize and rice genome, and physicochemical properties of ZmFAH proteins were predicted. A total of twenty nine and twenty two *FAH* genes were found in maize and rice, respectively. Phylogenetic analysis revealed a total of four clades in *ZmFAHs* and *OsFAHs* and predominantly contain transmembrane domain. Isoelectric point of ZmFAH proteins varied from 6.23 to 10.48. Phylogenetic tree and intron exon organization revealed that member within a group had similar gene structure and almost well distributed across ten chromosomes in maize. *In-silico* expression of *ZmFAH* genes indicates their differential expression across different tissues. Therefore it may be speculated that *FAH* family members in a single group may have similar function.

**Keywords:** *FAH*, β-carotene, sterol, hydroxylases, desaturases, gene evolution

## 1. Introduction
The *Fatty acid hydroxylase* (*FAH*) superfamily includes a broad sprectrum of proteins which comprises of carotene and fatty acid hydroxylases; and sterol desaturases in plants. This protein family is involved in synthesis of zeaxanthin due to hydroxylation of the β-carotene, the major provitamin-A (proA) carotenoid. Carotenoids are large group (comprising more than 700 compounds) of isoprenoids (5 carbon) synthesized by around 20 biosynthetic enzymes (Britton 1995) [1]. However, only α-carotene, β-carotene and β-cryptoxanthin act as a precursor molecule for vitamin A biosynthesis in human beings whereas lutein, zeaxanthin play an important role for scavenging free radicals (Olson 1989; Gupta *et al*. 2015) [2, 3]. *β-carotene hydroxylase* (*crtRB1*) gene which is a member of *FAH* superfamily, is involved in carotenoids biosynthesis pathway. It hydroxylates β-carotene into compounds having lesser proA activity (Yan *et al*., 2010; Vignesh *et al.* 2012) [4, 5]. Using marker-assisted selection (MAS), favourable allele of *crtRB1* was introgressed into promising maize inbreds for enrichment of kernel proA (Muthusamy *et al*., 2014; Zunjare *et al*., 2018) [6, 7]. Research efforts conducted at CIMMYT and other institutions worldwide has also led to the development of proA rich maize genotypes through exploitation of either or both of the two key genes viz. *β-carotene hydroxylase* (*crtRB1*) and *lycopene-ε-cyclase* (*lcyE*) that causes higher accumulation of proA in maize kernel (Manjeru *et al*., 2017) [8].

This superfamily also includes C-5 sterol desaturase and C-4 sterol methyl oxidase. Sterols are membrane bound lipophilic components of most eukaryotic cells which are required for structural integrity (Gulati *et al*., 2010) [9] and sterol along with phospholipids maintains the fluidity of the biological membrane (Lagace *et al*., 2013) [10] and sterol also acts as a precursors molecule of bile salts and steroid hormones in mammals, brassinosteroids in fungi and plants (Tomazic *et al*., 2011) [11]. They also act as precursor for ecdysteroids in arthropods (Gilbert *et al*., 2002) [12]. Sterol are de-novo synthesized in different forms viz., cholesterol in vertebrates, stigmasterol, sitosterol and campesterol in plants and ergosterol in fungi which are categorized into four protein families (Tomazic *et al*., 2011) [11]. Apart from these, proteins of this *FAH* superfamily are involved in cholesterol biosynthesis and plant cuticular wax (Kunst *et al*. 2003) [13].

*FAH* superfamilygene, *C-5 sterol desaturase* gene (*ERG3*) is putative gene which encodes C-5 sterol desaturase protein required for ergosterol biosynthesis in yeast (Arthington *et al.*, 1991) [14]. Targeted disruption of the *erg3A* and *erg3B* genes in *Aspergillus fumigates* revails that C-5 sterol desaturase enzymes (Erg3A and Erg3B) were not essential for *A. fumigatus* viability (Alcazar *et al.*, 2006) [15]. The *dwarf7/sterol1* (*dwf7*/*ste1*) mutant discovered in Arabidopsis reduces Brassinosteroid biosynthesis due to shortage of substrate sterols defection in the $\Delta^7$ Sterol C-5 desaturation step leads to dwarf phenotype (Gachotte *et al.*, 1995; Choe *et al.*, 1999) [16, 17]. C-4 sterol methyl oxidase (ERG25) in *Saccharomyces cerevisiae* performs the first of three enzymatic steps prerequisite for production of cholesterol (animal), ergosterol (fungal), and stigmasterol (plant) after the removal of two C-4 methyl groups (Bard *et al.*, 1996) [18]. Therefore *FAH* superfamily has very important role for the plant physiological and morphological activities in crop plant. In this study, candidate *FAH* genes were identified in maize genome and phylogenetic relationships among candidate *FAH* genes were established to infer molecular evolution. Maize *FAH* genes were further analyzed to predict their subcellular localization, biochemical properties, intron-exon organization and their expression. This study will enhance our understanding of the evolution and function of the *FAH* superfamily in maize. The evolutionary and structural divergence analysis of *FAH* superfamily in maize presented here provides useful information for further probing the molecular mechanism by which *FAH* superfamily contributes to the sterol biosynthesis and carotenoids pathway.

## 2. Materials and methods

### 2.1 Identification of the Fatty acid hydroxylase superfamily proteins

Hidden Markov model (HMM) profile of *FAH* superfamily (PF04116) was retrieved from Pfam database (https://pfam.xfam.org/, Finn *et al.*, 2015) [19]. Predicted protein sequences of maize and rice were downloaded from Ensemble plant (plants.ensembl.org/index.html, Zerbino *et al.*, 2018) [20]. To identify *FAH* genes both HMM profile search and BLAST search were used. HMM profile search against maize and rice protein database was performed using HMMER 3.0. (hmmer.org/). Putative FAH proteins identified from HMM search were again used as a query and BLASTP was employed against maize protein database to identify *FAH* genes with more than 50% sequence identity and with an E-value of less than $10^{-5}$. All the predicted ZmFAH proteins were further searched against Pfam database and CD databse and Inter Pro Scan (https://www.ebi.ac.uk/interpro/interproscan.html, Quevillon *et al.*, 2005) to ensure presence of FAH domain. Multiple isoforms were predicted for multiple genes, but only one isoforms was taken for each gene.

### 2.2 Phylogenetic Analysis

Multiple sequence alignment of 29 *ZmFAH* genes and 22 *OsFAH* genes were performed using CLUSTALWin MEGA6 software (Tamura *et al.*, 2013) [22]. This alignment file was used to establish phylogenetic relationship among rice and maize *FAH* genes using the Neighbor-Joining method (Saitou *et al*., 1987) [23] and unrooted bootstrap consensus tree was inferred from 1000 replicates (Felsenstein *et al.*, 1985) [24]. Less than 50% bootstrap replicates were collapsed to reproduce branches corresponding to partitions. Poisson correction method was used to compute evolutionary distances.

### 2.3 Physicochemical property of the protein

CELLO v2.5 (cello.life.nctu.edu.tw; Yu *et al*., 2006) [25] was used to determine sub-cellular localization of the FAH proteins.Iso-electric point of the proteins was determined by Protein isoelectric point calculator (isoelectric.org/; Kozlowski *et al.*, 2016) [26]. Trans-membrane helices in FAH proteins were predicted using TMHMM server version 2.0 (http://www.cbs.dtu.dk/services/TMHMM; Krogh *et al.*, 2001) [27] with the standard parameter.

### 2.4 Gene structure analysis and chromosomal location

The exon/intron structure of each *FAH*gene was generated using Gene Structure Display Server (GSDS; http://gsds.cbi.pku.edu.cn/, Hu *et al.*, 2014) [28] by alignment of the CDS with their corresponding genomic DNA sequences. Chromosomal location of maize *FAH* genes was retrieved from Biomart data mining tool of Ensemble plant (plants.ensembl.org/index.html, Zerbino *et al.*, 2018) [20]. Chromosomal distribution of 29 *ZmFAHs* were presented using Map Chart software (https://www.wur.nl/en/show/Mapchart.htm, Voorrips *et al.*, 2002) [29] based on their genomic co-ordinate.

### 2.5 Expression analysis *FAH* genes family

Expression of maize *FAH* genes in twelve different tissues was analyzed using the GENEVESTIGATOR software (https://genevestigator.com/, Hruz *et al*., 2008) [30] which utilizes publicly available transcriptome data. Expressions of *FAH* genes were calculated as "percent of expression potential" (PEP = average expression of a gene across all the samples in a particular tissue / maximum expression value across samples) and the results were presented as heat map usinglogarithm (log2) scale.

## 3. Results and discussion

### 3.1 *Fatty acid hydroxylase* superfamily genes and their phylogenetic analysis

A total of 29 genes were found in maize genome having *FAH* protein like activity by using Hidden Markov model (HMM) search. Single isoform of each gene were taken for downstream analysis. All the predicted FAH proteins were found to contain FAH domain by Pfam database and CD databse and InterProScan. Apart from the FAH domain, four proteins (ZmFAH14, ZmFAH18, ZmFAH23, ZmFAH28) also have wax2-c-terminal domain (PF12076). To study phylogenetic relationship between 22 *OsFAH* and 29 *ZmFAH* from *O. sativa* and *Z. mays*, protein sequences were clustered into four sub groups with bootstrap support 45in an un-rooted Neighbor-Joining tree. The largest group (group-I) contain 16 *FAH* genes and smallest group (group-IV) contain 7 *FAH* genes. *FAH* genes clustered into same group may have similar function (Fig 1).

### 3.2 Physicochemical property of FAH proteins and their gene structure analysis

The FAH proteins were mainly plasma membrane bound protein as confirmed by CELLO v2.5 and TMHMM software. However *ZmFAH1*, *ZmFAH2*, *ZmFAH6*, *ZmFAH11* and *ZmFAH29* proteins were localized in mitochondria; *ZmFAH5*, *ZmFAH29* were localized in chloroplast; *ZmFAH10* was

localized in endoplasmic reticulum (ER) and *ZmFAH14* was extra-cellularly localized. Isoelectric point of ZmFAH proteins varied from 6.23 to 10.48. Maximum number of trans-membrane helices was 6 as determined by TMHMM server version 2.0. Three ZmFAH proteins (*ZmFAH11*, *ZmFAH14*and *ZmFAH17*) do not contain any transmenbrane domain. The number of the amino- acids was obtained from TMHMM server version 2.0 and it varied 180 to 635 (Table 1). Details of the Gene ID, isoelectric point (pI) of the proteins, their sub-cellular localization, number of the trans-membrane domain and total number of amino acid were given in Table 1. To further understand *FAH* gene structure, *ZmFAH* gene transcripts were analyzed using GSDS server for Intron-Exon organization. Structural feature of the transcript were presented based on evolutionary relationship. Genes having similar intron-exon organization were clustered together (Fig 2). Number of introns in *ZmFAH* genes varied from one to eleven. Eleven genes contain only single intron whereas one gene has (*ZmFAH27*) 11 introns (Figure 2). Size of *ZmFAH* genes varied from 0.87 kb (*ZmFAH8*) to 9.59 kb (*ZmFAH7*) and average size of *ZmFAH* genes was 3.7 kb. Phylogenetic tree and intron exon organization revealed that member within a group had similar gene structure. Therefore it may be speculated that *FAH* family members in a single group may have similar function. *ZmFAH* genes were almost well distributed across ten chromosomes. However, Chromosome 4 has five *ZmFAH* genes and Chromosome10 has one *ZmFAH* gene. Each of the chromosomes namely Chromosome 1, Chromosome 3 and Chromosome 5 carry four *ZmFAH* genes (Fig 3).

### 3.3 Expression profile *ZmFAH* genes

Expression of twenty eight *ZmFAH* genes was analyzed in twelve different tissues using GENEVESTIGATOR software. Expression of one gene (*ZmFAH15*) was not measured as GENEVESTIGATOR database has no expression data of that gene. It was found that expression potential of *GRMZM2G166853*, *GRMZM2G007289*, *GRMZM2G423229* and *GRMZM2G030118* was very less across all the tissues (Fig 4). Interestingly *GRMZM2G163008* gene was over expressed only in root as compared to other tissues. It could be due to tissue dependent expression of the *ZmFAH* genes. Highest expression potential was found in *GRMZM2G038964*in pollen. *In-silico* expression of *ZmFAH* genes indicates their differential expression across different tissues.

**Table 1:** Features of *Zea mays fatty acid hydroxylase* (*FAH*) genes

| Gene Name | Gene ID | Chromosome location | pI | Sub-cellular localization | No. of TMH | No. of AA. | Domain |
|---|---|---|---|---|---|---|---|
| *ZmFAH1* | *Zm00001d027443 GRMZM2G090051* | 1 | 9.66 | PM, MIT | 4 | 299 | FAH(136-269) |
| *ZmFAH2* | *Zm00001d029810 GRMZM2G127067* | 1 | 9.24 | PM, MIT | 1 | 410 | FAH(239-374) |
| *ZmFAH3* | *Zm00001d031149 GRMZM6G247892* | 1 | 7.44 | PM | 2 | 267 | FAH(87-256) |
| *ZmFAH4* | *Zm00001d034319 GRMZM2G038964* | 1 | 7.63 | PM | 3 | 237 | FAH(87-226) |
| *ZmFAH5* | *Zm00001d026056 GRMZM2G152135* | 2 | 8.07 | PM, CHL | 4 | 319 | FAH(134-267) |
| *ZmFAH6* | *Zm00001d002589 GRMZM2G164318* | 2 | 9.90 | PM, MIT | 4 | 309 | FAH(153-286) |
| *ZmFAH7* | *Zm00001d006920 GRMZM2G044175* | 3 | 6.44 | PM | 5 | 258 | FAH(101-237) |
| *ZmFAH8* | *Zm00001d039965 GRMZM5G850019* | 3 | 6.32 | PM | 2 | 276 | FAH(135-265) |
| *ZmFAH9* | *Zm00001d040979 GRMZM2G062531* | 3 | 6.23 | PM | 2 | 218 | FAH(130-218) |
| *ZmFAH10* | *Zm00001d041458 GRMZM2G066321* | 3 | 7.20 | PM, ER | 3 | 293 | FAH(132-267) |
| *ZmFAH11* | *Zm00001d043994 GRMZM2G003526* | 4 | 8.83 | PM, MIT | 0 | 275 | FAH(114-250) |
| *ZmFAH12* | *Zm00001d048585 GRMZM2G000531* | 4 | 7.86 | PM | 3 | 228 | FAH(105-210) |
| *ZmFAH13* | *Zm00001d049213 GRMZM2G007289* | 4 | 7.34 | PM | 1 | 250 | FAH(93-230) |
| *ZmFAH14* | *Zm00001d051932 GRMZM2G099097* | 4 | 7.68 | EX | 0 | 204 | FAH(1-92); WAX2(88-202) |
| *ZmFAH15* | *Zm00001d052258* | 4 | 7.47 | PM | 3 | 258 | FAH(100-236) |
| *ZmFAH16* | *Zm00001d053106 GRMZM2G166853* | 5 | 7.54 | PM | 1 | 265 | FAH(107-243) |
| *ZmFAH17* | *Zm00001d014055 GRMZM2G066578* | 5 | 8.24 | PM | 0 | 414 | FAH(131-272) |
| *ZmFAH18* | *Zm00001d015477 GRMZM2G083526* | 5 | 8.97 | PM | 4 | 627 | FAH(131-272); WAX2(456-626) |
| *ZmFAH19* | *Zm00001d017251 GRMZM2G075255* | 5 | 8.87 | PM | 2 | 318 | FAH(1-105) |
| *ZmFAH20* | *Zm00001d018041 GRMZM5G836378* | 6 | 7.68 | PM | 3 | 258 | FAH(100-236) |
| *ZmFAH21* | *Zm00001d037225* | 6 | 7.14 | PM | 5 | 263 | FAH(97-233) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | *GRMZM2G423229* | | | | | | |
| *ZmFAH22* | *Zm00001d038630 GRMZM2G163008* | 7 | 8.00 | PM | 1 | 266 | FAH(102-238) |
| *ZmFAH23* | *Zm00001d020557 GRMZM2G114642* | 8 | 9.11 | PM | 4 | 621 | FAH(128-271); WAX2(450-620) |
| *ZmFAH24* | *Zm00001d008569 GRMZM2G167085* | 8 | 6.97 | PM | 2 | 316 | FAH(175-305) |
| *ZmFAH25* | *Zm00001d011503 GRMZM2G030118* | 8 | 8.61 | PM | 1 | 277 | FAH(116-252) |
| *ZmFAH26* | *Zm00001d011765 GRMZM2G176301* | 9 | 9.53 | PM | 2 | 180 | NONE |
| *ZmFAH27* | *Zm00001d044833 GRMZM2G124276* | 9 | 7.13 | PM | 3 | 264 | FAH(98-234) |
| *ZmFAH28* | *Zm00001d046865 GRMZM2G029912* | 9 | 8.65 | PM | 6 | 635 | FAH(143-284); WAX2(464-634) |
| *ZmFAH29* | *Zm00001d048469 GRMZM2G382534* | 10 | 10.48 | PM, CHL, MIT | 3 | 293 | FAH(129-263) |

pI: Iso electric point; TMH: Trans Membrane Helix; AA: Amino Acid

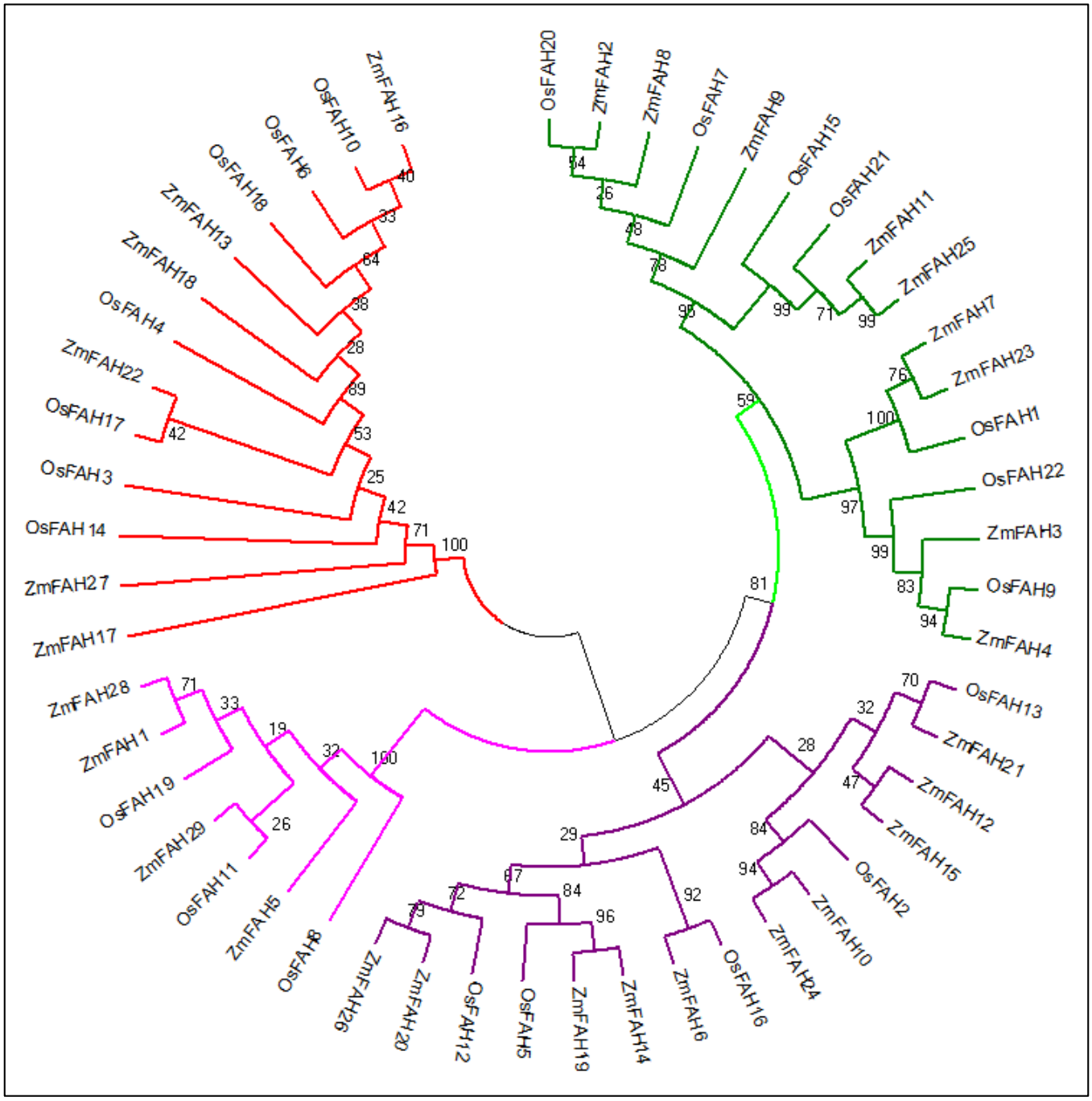


**Fig 1:** Neighbor-Joining phylogeny of *FAH* genes – Number on branches are bootstrap percentage

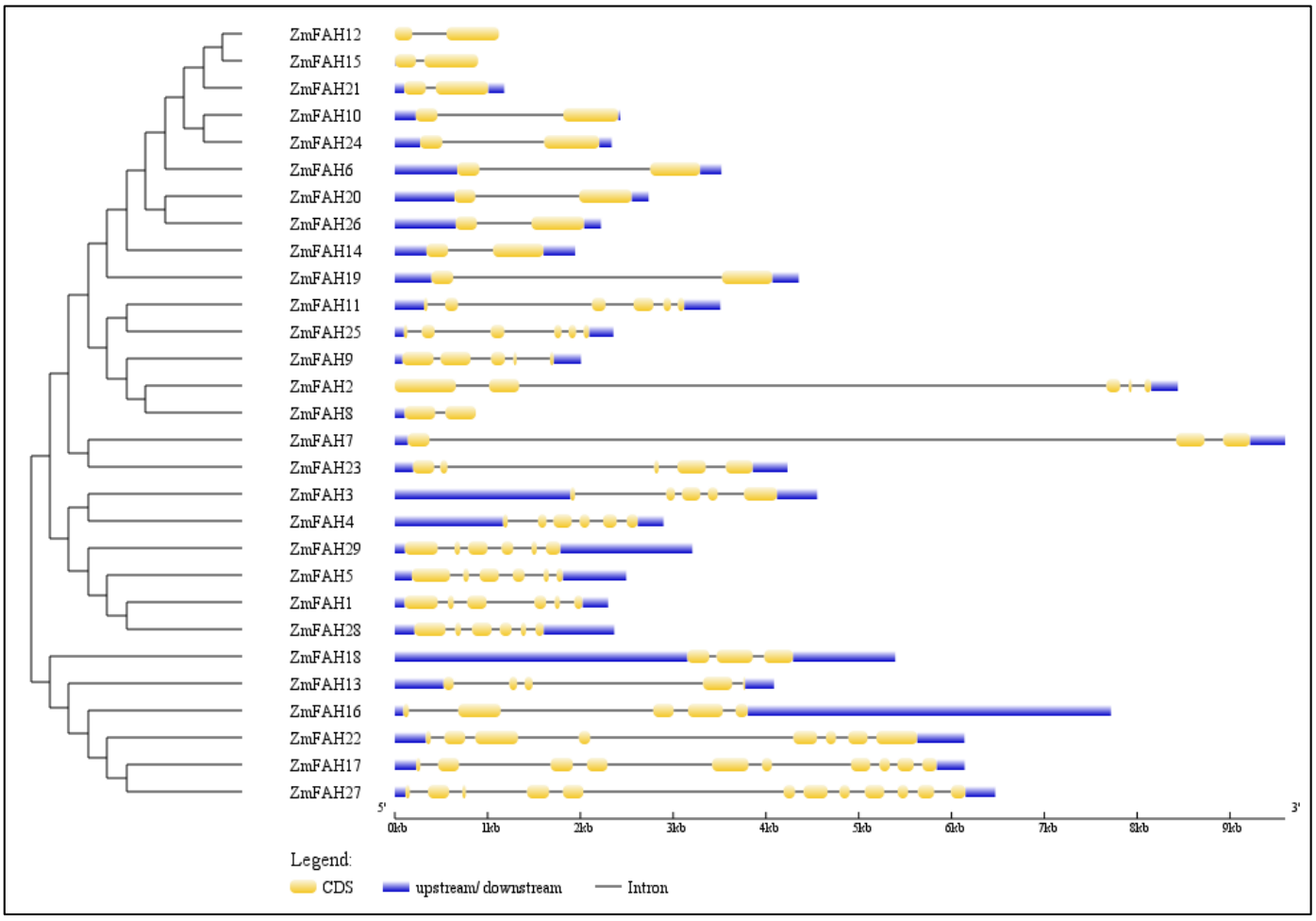


**Fig 2:** *ZmFAH* genes structure with intron-exon boundaries

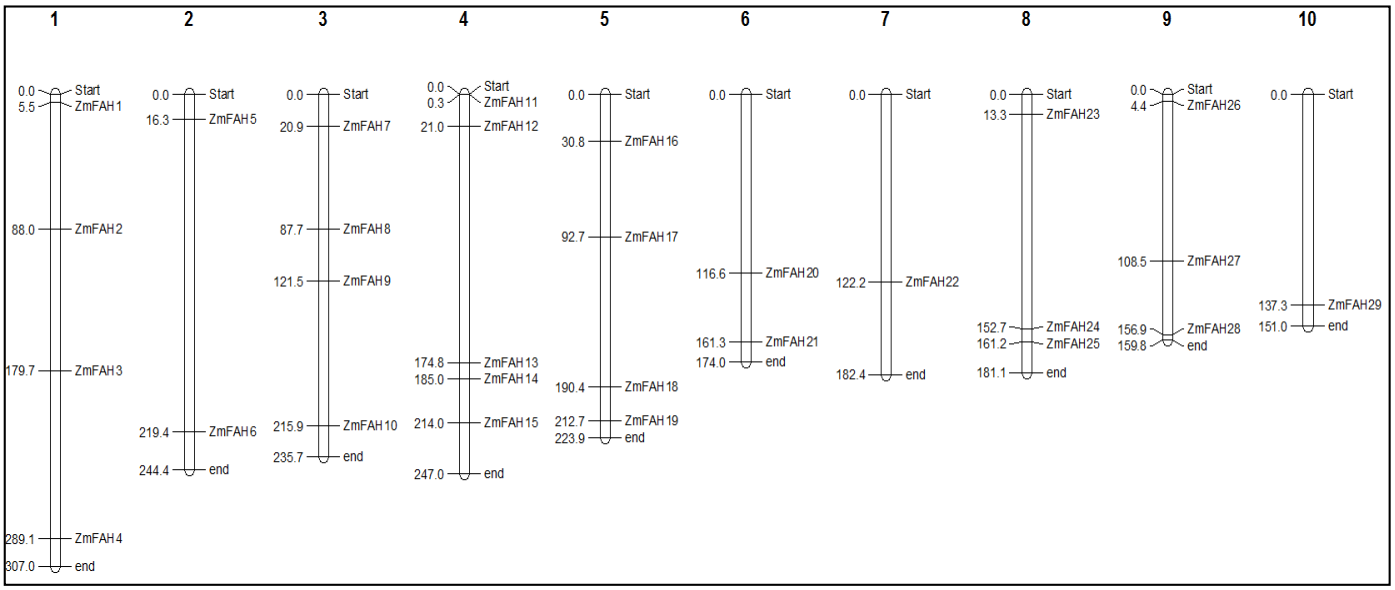


**Fig 3:** Chromosomal location of *FAH* genes in the maize genome

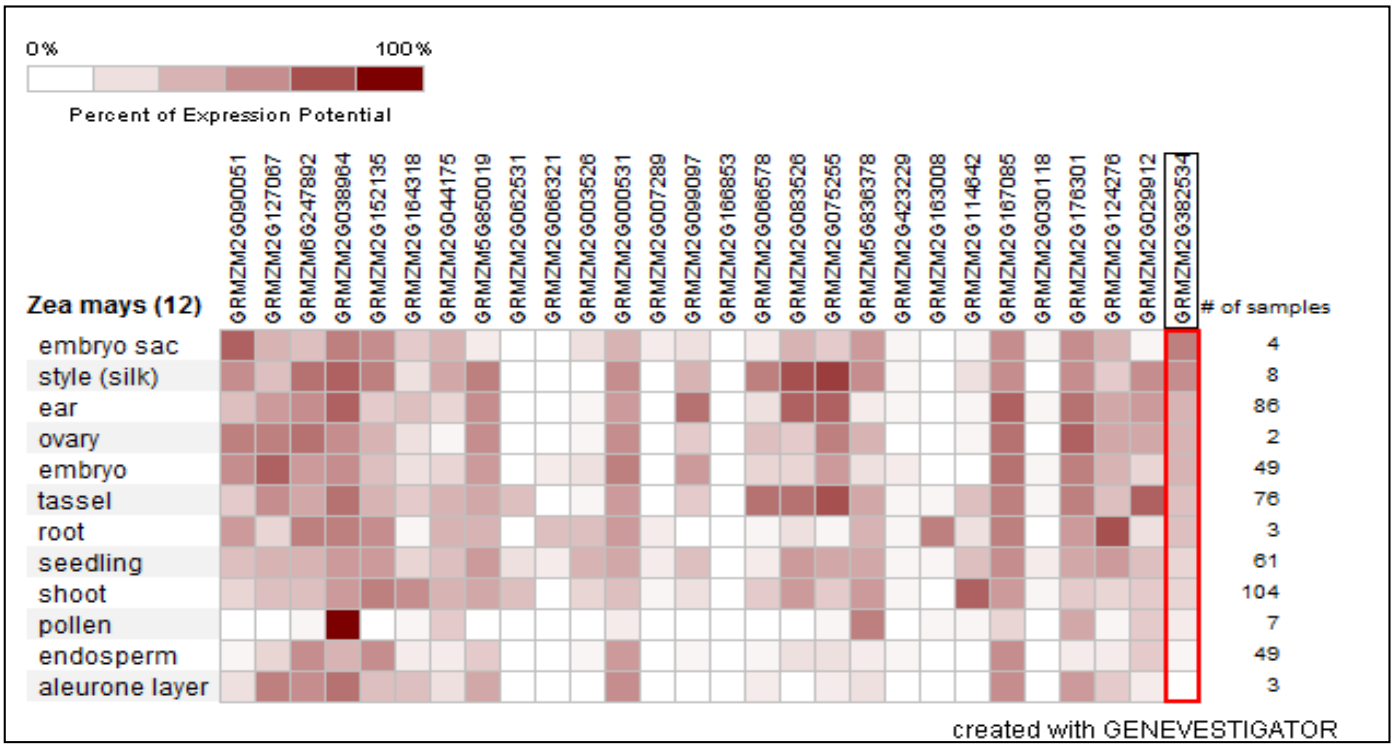


**Fig 4:** Heatmap of expression profiles of *ZmFAH* genes in different tissues

## 4. Conclusion

In this study maize genome has been analyzed to identify and characterize *FAH* gene family by using different bioinformatics tool. HMM profile search and standalone BLAST search was employed to identify candidate *FAH* genes. Iso-electric point was obtained from Protein isoelectric point calculator. Chromosomal location, locus ID was retrieved from Ensemble plant database. Multiple sequence alignment and tree contraction was carried out in MEGA 6 software. Exon-intron distribution in *ZmFAH* genes were analyzed using GSDS server. Chromosomal location of *FAH* genes were visualized using MAPCHART software. *In-silico* gene expression was performed using GENEVESGATOR software. This comprehensive analysis of *ZmFAH* genes paves the way for deciphering the functional role of all *FAH* genes in maize. *crtRB1* belonging to the *ZmFAH* gene family has a detrimental role in carotenoids biosynthesis pathway. This analysis presented here may be useful in selection of candidates genes for further functional characterization related to maize carotenoids biosynthesis pathway.

## 5. Acknowledgements

The first author is thankful to Indian Council of Agricultural Research for the Junior Research Fellowship to pursue the master's degree programme. Financial support from the ICAR funded CRP on Biofortification is thankfully acknowledged. We also thank Dr. Anil Dahuja, Principal Scientist (Biochemistry), IARI for the help.

## 6. References

1. Britton G. Structure and properties of carotenoids in relation to function. The FASEB Journal. 1995; 9(15):1551-1558.
2. Olson JA. Biological actions of carotenoids. Journal of Nutrition. 1989; 119(1):94-95.
3. Gupta HS, Hossain F, Muthusamy V. Biofortification of maize: An Indian perspective. Indian J Genet. 2015; 75(1):1-22.
4. Yan J, Kandianis BC, Harjes EC, Bai L, Kim HE, Yang X *et al*. Rare genetic variation at *Zea mays* crtRB1 increases beta carotene in maize grain. Nature Genetics. 2010; 42(4):322-329.
5. Vignesh M, Hossain F, Nepolean T, Saha S, Agarwal PK, Guleria SK *et al*. Genetic variability for kernel β-carotene and utilization of *crtRB1* 3'TE gene for biofortification in maize (*Zea mays* L.). Indian J Genet. 2012; 72(2):189-194.
6. Muthusamy V, Hossain F, Thirunavukkarasu N, Choudhary M, Saha S, Bhat JS *et al*. Development of β-carotene rich maize hybrids through marker-assisted introgression of β-carotene hydroxylase allele. PLoS ONE. 2014; 9(12):e113583.
7. Zunjare RU, Hossain F, Muthusamy V, Baveja A, Chauhan HS, Bhat JS *et al*. Development of Biofortified Maize Hybrids through Marker-Assisted Stacking of β-carotene hydroxylase, lycopene-ε-cyclase and opaque2 Genes. Frontiers in Plant Science. 2018; 9:178.
8. Manjeru P, Biljon AV, Labuschagne M. The development and release of maize fortified with Provitamin A carotenoids in developing countries. Critical Reviews in Food Science and Nutrition https://doi.org/10.1080/10408398.2017.1402751.
9. Gulati S, Liu Y, Munkacsi AB, Wilcox L, Sturley SL. Sterols and sphingolipids: dynamic duo or partners in crime?. Progress in Lipid Research. 2010; 49(4):353-365.
10. Lagace TA, Ridgway ND. The role of phospholipids in the biological activity and structure of the endoplasmic reticulum. Biochimicaet Biophysica Acta (BBA)-Molecular Cell Research. 2013; 1833(11):2499-2510.
11. Tomazic ML, Najle SR, Nusblat AD, Uttaro AD, Nudel CB. A novel sterol desaturase like gene promoting dealkylation of phytosterols in *Tetrahymena thermophila*. Eukaryotic Cell, 2011.
12. Gilbert LI, Rybczynski R, Warren JT. Control and biochemical nature of the ecdysteroidogenic pathway. Annual Review of Entomology. 2002; 47(1):883-916.

13. Kunst L, Samuels AL. Biosynthesis and secretion of plant cuticular wax. Progress in lipid research. 2003; 42(1):51-80.
14. Arthington BA, Bennett LG, Skatrud PL, Guynn CJ, Barbuch RJ, UIbright CE *et al*. Cloning, disruption and sequence of the gene encoding yeast C-5 sterol desaturase. Gene. 1991; 102(1):39-44.
15. Alcazar-Fuoli L, Mellado E, Garcia-Effron G, Buitrago MJ, Lopez JF, Grimalt JO *et al*. *Aspergillus fumigatus* C-5 sterol desaturases Erg3A and Erg3B: role in sterol biosynthesis and antifungal drug susceptibility. Antimicrobial Agents and Chemotherapy. 2006; 50(2):453-460.
16. Gachotte D, Meens R, Benveniste P. An Arabidopsis mutant deficient in sterol biosynthesis: heterologous complementation by ERG 3 encoding a Δ7-sterol-C-5-desaturase from yeast. The Plant Journal. 1995; 8(3):407-416.
17. Choe S, Noguchi T, Fujioka S, Takatsuto S, Tissier CP, Gregory BD *et al*. The Arabidopsis dwf7/ste1 mutant is defective in the Δ7 sterol C-5 desaturation step leading to brassinosteroid biosynthesis. The Plant Cell. 1999; 11(2):207-221.
18. Bard M, Bruner DA, Pierson CA, Lees ND, Biermann B, Frye L, Barbuch R. Cloning and characterization of ERG25, the *Saccharomyces cerevisiae* gene encoding C-4 sterol methyl oxidase. Proceedings of the National Academy of Sciences. 1996; 93(1):186-190.
19. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL *et al*. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research. 2015; 44(D1):D279-D285.
20. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J *et al*. Ensembl 2018. Nucleic Acids Research. 2017; 46(D1):D754-D761.
21. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R *et al*. Inter Pro Scan: protein domains identifier. Nucleic acids research. 2005; 33(2):W116-W120.
22. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution. 2013; 30(12):2725-2729.
23. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987; 4(4):406-425.
24. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 1985; 39(4):783-791.
25. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins: Structure, Function and Bioinformatics. 2006; 64(3):643-651.
26. Kozlowski LP. IPC–isoelectric point calculator. Biology direct. 2016; 11(1):55.
27. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology. 2001; 305(3):567-580.
28. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2014; 31(8):1296-1297.
29. Voorrips RE. Map Chart: software for the graphical presentation of linkage maps and QTLs. Journal of heredity. 2002; 93(1):77-78.
30. Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L *et al*. Genevestigator v3: A reference expression database for the meta-analysis of transcriptomes. Advances in bioinformatics, 2008, 2008. http://dx.doi.org/10.1155/2008/420747