The Journal of Animal & Plant Sciences, 29(6): 2019, Page: 1630-1642
ISSN: 1018-7081

ASSESSMENT OF GENOME-WIDE EVOLUTIONARY RELATIONSHIP AND CHARACTERIZATION OF *GhFPF1* FAMILY IN *G. HIRSUTUM*

A. H. Jarwar[1,3], X. Wang[2], W. Long[1], A. Ahmad[1], N. Mangi[1], Q. Ma[1*], and F. Shuli[1*],

[1]State Key Laboratory of Cotton Biology, Institute of Cotton Research of CAAS, Anyang Henan, 455000, China; [2]Anyang Institute of Technology, College of Biology and Food Engineering, Anyang, 455000, China; [3]Cotton Research Institute, Agriculture Research Institute, Tandojam, Sindh, Pakistan,
[*]Corresponding author: Email: Fsl427@126.com,

ABSTRACT

A flowering promoting factor 1 was first cloned from Arabidopsis, and it is a model type of species which have revealed a network of regulatory association between proteins that integrate and transduce environmental and developmental signals to advanced or inhibit the transition to flowering. *FPF1* homologus sequences were examined from diploid *G. arboreum L.* genome (A-genome, n=13), and *Gossypium raimondii L.* (D-genome, n=13) data base. There are two orthologus genes are propose that distinctions at amino acid and nucleic acid levels were not significance because of degeneracy of codon. Atleast 6 genes of *FPF1* were obtained from cultivated allotetraploid species *Gossypium hirsutum* L (genome AD, n=26). Analysis of six genes from various tissues discloses that this type of gene family shows that expression level of tissue specific is very strong. In previous studies in Arabidopsis it shows significantly suppressed Flowering Locus C and activated in APETALA 1 expression were stimulate by over expression of *GhFPF1* in the Arabidopsis.

Keywords: *GhFPF1*, wild type, transgenic, G-Protein.

INTRODUCTION

The cotton is most important renewable and natural crop in the world, approximately 33 million ha (5% of the total worlds arable land) is used for cotton cultivation (Wang *et al*., 2012). In addition the genus *Gossypium* had a wide phenotypic diversity, it contains more than 50 species (Wendel, 2009; Wendel and Grover, 2015; Gallagher, 2017). The *Gossypium* are cytogenetically distinguished into 8 sub groups (A-G, and K), between them most of the species is diploid which is (n=13), but among them 5 are allopolyploids (n=26), Global cotton products are demonstrated from the 2 allotetraploid species *Gossypium hirsutum* and *Gossypium barbadense* (Wendel, 2013; Wendel, 1992). The study of homoeologous gene development in allopolyploid upland cotton, we induced a phylogenetic perspective by including orthologus duplicated from representative of each of the progenitor diploid genes (Cronn RC, 1999).

Because of the environmental pressures, such as climatic change and land use, the early maturity of cotton has become a critical theme for plant breeders. Various traits collaborate to influence the early maturity of cotton; along flowering time is especially significant (Jung and Muller, *et al*., 2009). The Flowering Promoting Factor 1 was first cloned from Arabidopsis by (Kania, *et al*., 1997). The apical meristem being determined or in determined contributes to the fate of shoot architecture. The growth of plants initiates from a very small number of various cells which is called as meristems. Determinate apical meristems undergo terminal differentiation, whereas indeterminate apical meristems hold vegetative stem cells of population indefinitely which execute tissue and other organ differentiation below and on the flask of main stem (Melzer *et al*., 2008). The independent and vernalization pathways unite on the regulation of the expression of the floral repressor FLC (Levy, and Dean, *et al*., 1998). *FLC* encodes a MADS-box transcription factor that subdue flowering. The overexpression of flowering locus C from a broadly expressed viral promoter decreases the abundance of SOC1 and flowering time mRNAs in young plants (Searle, *et al*., 2006). We found that the vernalization response late flowering mutants had a high-level of Flowering Locus C transcript than the wild types species and commonly that the non-vernalization-response late flowering mutants did not grownup Flowering Locus C transcript level (Sheldon, *et al*., 2000).

This promotion is known as vernalization, which does not initiate flowering directly but provides the meristem capable to respond to other developmental and environmental flowering signals (Michaels, and Amasino, *et al*., 2001). The timing of floral induction is vital for reproductive success, and high number of plant species has germinated various pathways to modulate flowering time (Michaels, and Amasino, *et al*., 1999). Many genes are recognized which are intricate in controlling the flowering time, and they play a vital role in genetic as well as in molecular pathways were also characterized (Bernier *et al*., 2002; Bernier *et al*., 2005). In previous studies there are four important floral pathways, such as gibberellin, photoperiod, vernalization, and independent pathways these are modulate floral integrator genes such as Flowering locus T (FT), APETALA 1 (AT1), LEAFY (LFY), and SUPPRESSOR OF OVEREXPRESSION OF CONSTANT 1 (SOC1) (Corbesier, *et al*., 2007; Li J, *et al*.. 2013).

These are the activated genes for necessarily reproductive development (Simpson, and Dean, *et al*., 2002). Flowering promoting factor 1 gene is explicit immediately in apical meristems after the photoperiodic initiation of flowering in long day plants that could flowering in response to long day conditions (Kania, *et al*., 1997). The Flowering Promoter factor 1 gene was originally understood on account of its role in flowering. The high *FPF1* gene expression (Y11988) led to early flowering in Arabidopsis (Kania, *et al*., 1997). The previous study indicates that Flowering Promoter Factor might play a vital role in regulating the competence of apical meristems to immediately react to the floral meristems identity gene *LFY* and *AP1*. *GhFPF1* gene constantly active at a same time as *LFY*, and earlier than AP1 (Melzer, *et al*., 1999). Still yet corresponding genes of Flowering Promoting Factor 1 have also been described in tobacco (*Nicotiana tabacum*) and rice (*Oryza sativa*) (Ge, *et al*., 2004; Smykal, *et al*., 2004).

MATERIALS AND METHODS

The Plant material and RNA and DNA Extraction: The experiment was carried out in the field of Cotton Research Institute of CAAS, Anyang, china, the cotton genotypes were obtained from various mainland's, such as, Wake Atoll, a Hawaiian island, Galapagos island, Dominican Republic, San Cristobal, and Santa Cruz.

DNA Extraction and Plant Material: We extracted a sum of 36 genotypes which is belongs to various species, such as, 18 genes from *G. hirsutum*, 8 genes from *G. arboreum*, and 9 genes from *G. raimondii* from various mainland's and wild type germplasm in china. Among them 36 genotypes which were collectively collected from a Hawaiian Island, Dominican Republic, Wake Atoll, Santa Cruz, and San Cristobal in respect manners. When the seedlings were grown at cotyledon stage the young leaves, stem and buds were collected to extract DNA for DNA extraction the DNA was extracted from frozen stem leaves and buds using CTAB method cetyltrimethylammonium bromide, as previously reported by (Zhang and Stewart, *et al*., 2000). For tissue specific expression profile, true leaves, young leaves, cotyledons, roots, and stems, were collected during various plants development stages (Ondati, *et al*., 2016). When the young plants are at leaf development stages, the leaves were Reap from cotton seedlings at 5 development stages. For RNA all the collected samples were put in liquid nitrogen for frozen and then stored at -80 °C. For RNA extraction, the RNA was extracted using RNA prep pure Tiangen kit (Polysaccharides & Polyphenolics). QRT-PCR analysis was executed using CWBIO (Low ROX) Ultra SYBR Mixture (Supplementary figure 1).

QRT-PCR analysis and expression of *GhFPF1* gene and wild type Plants: For QRT-PCR the expression of GhFPF1 gene and wild type under different plant growth hormones. The young and fresh leaves are collected for RNA extraction. Using CWBIO Ultra SYBR Mixture (Low ROX) 5ml obtained from Cotton Research Institute Anyang Henan, was used to obtained RNA from fresh Leaf, Stem, and Root (Livak KJ, Schmittgen TD, 2001). The quality and purification was detected using spectrophotometer Nanodrop 2000 (Waltahm, MA, Thermal Fisher, USA) and using Gel electrophoresis.

RESULTS

Analysis of phylogenetic tree of *GhFPF1*: To know the relationship and evolutionary history of *GhFPF1* gene family, a tree was constructed through an alignment sequence of the full length *GhFPF1* protein sequence with Neibour-joining method on the basis of similarities of *GhFPF1* gene in *G.hirsutum*, furthermore, the phylogenetic tree of *GhFPF1* gene members with gene combination and motif compositions.

Gene duplication events and chromosomal locations: The information about the physiological positions of *GhFPF1* genes on chromosomes was abstract through the soft wear BlastN searches against *G. hirsutum* all the *GhFPF1* gene were then measured on the chromosomes using the map chart soft wear. To detect the *GhFPF1* genomic duplication happens was done which is based on alignment results. In the previous studies it was described that the smaller sequence which cover over 87% of the larger sequence after alignment and the small identity of aligned regions is greater than 80% (Zhang, and Stewart, *et al*., 2000). Moreover, for further proceeds, genomic similarities events, and synonymous substitution rate and the nonsynonymous substitution rate (Ka) were estimated using the DNA SP soft wear the time represents with T was estimated according to the equation $T=Ks/2\lambda \times 10^{-6}$ Mya (Librado, *et al*., 2009).

Analysis of sequence comparative and gene screening: Candidate gene sequencing was abstracted through performing blast searches of genomic data bases of *G. arboreum* L. (unpublished), and *G. raimondii L.* from website (http;//cgp.genomics.org.cn) the Flowering promoter factor gene 1 as an inquiry sequence. Various alignment sequence of the infer amino acids was executed using the Clustal X soft wear (http;//www.ebi.ac.uk), and the tree was performed by

MEGA-X soft wear (Wang, *et al*., 2012) (http;//www.megasoftwear.net), and the promoter sequence was analyzed in the data base of plant cis-acting regulatory element, DNA elements (http;//bioinformatics.psb.ugent.be/webtools/plantcare/html/) Plant CARE.

Promoter analysis: There are many functions of Cis-regulatory elements they are depending on their different styles, localization, types, and orientation on the promoter (Kumar, and Tamura, *et al*., 2016). Our study reports, a high number of *GhFPF1* gene promoters includes five elements groups among 3 of which (hormones responsiveness, stress responsiveness, and light responsiveness elements) are most abundant. While as the other 2 groups (binding site and cellular development) are minor abundant. Evidently, it was confirmed and sure by exogenous application of hormones to express their role in abiotic and biotic stress regulation (Biłas, *et al*., 2016; Roy Choudhury, *et al*., 2009).

Table 1. Specific gene primers used in real time qRT-PCR analysis for the expression of different plant hormones in the transgenic cotton *FPF1* gene and Wild type cotton.

Genes	Forward sequence	Reverse sequence
GhFPF1	ATGAGCGGTCCTTGGTGTTT	TCATTTATCCATAACCATGAACAT
GhFPF1	TATGGACTGTATCACCACCAC	TTGGACACACAAAGAAACACC

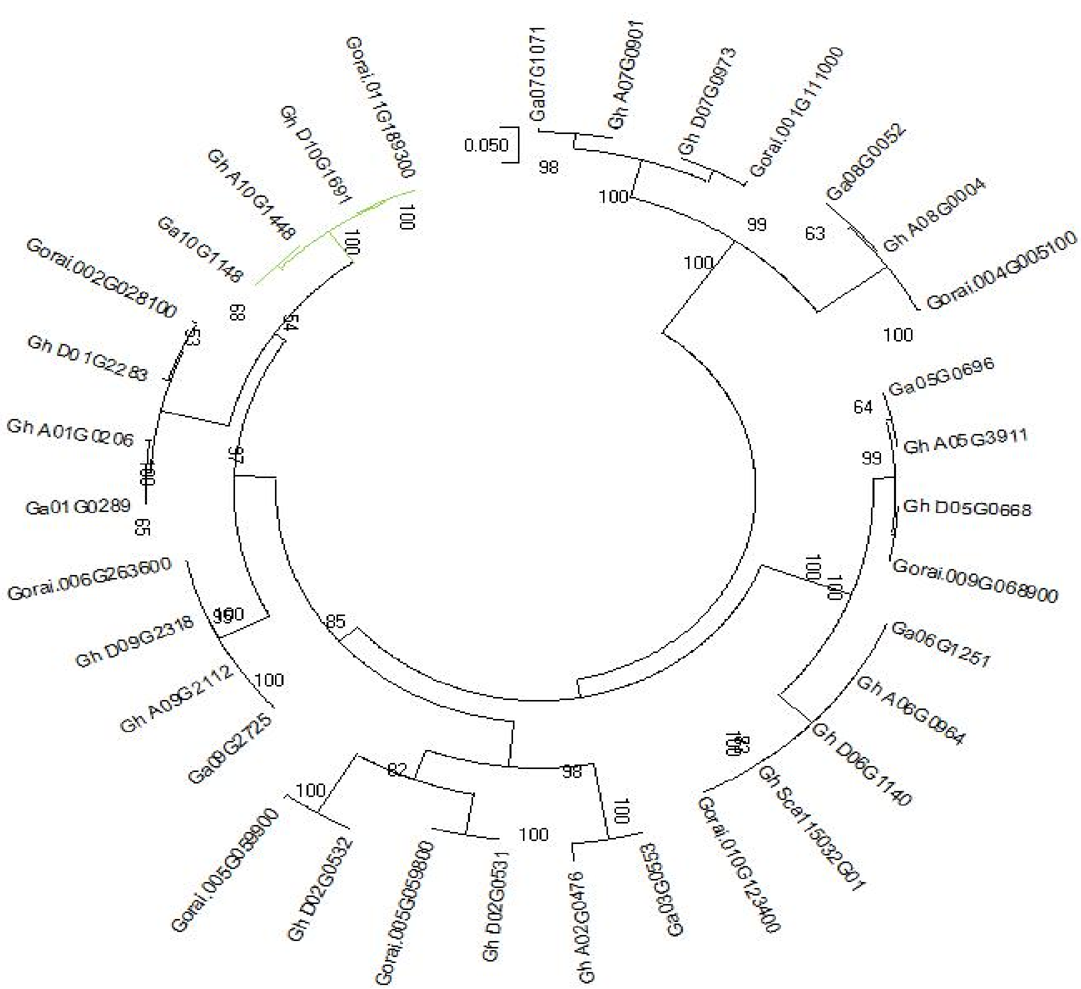

Figure 1. Phylogenetic tree, of the *GhFPF1* gene protein, in *G. hirsutum, G. arboreum, G. raimondii,* were aligned with the Neighbor joining (NJ) method in MEGA-X soft wear. The phylogenetic tree is divided into four main groups, SF7, SF3 and *FPF1* gene.

Gene Ontology: Actually, the gene Ontology is a very useful method to depict the gene and their functions (Escobar-sepu, *et al*., 2017). In our study research, gene ontology analysis disclosed that *GhFPF1* gene involved in the catalytic activity. The apical meristem being determinate or indeterminate contributes to the fate of shoot architecture. The determinate apical meristems undergo terminal differentiation, usually in inflorescence or flower. Meanwhile indeterminate apical meristems hold a population of vegetative stems cells indefinitely which shows tissue and other organ initiation below and on the flanks of the stem. In annually plants, the floral meristem initiation process happens when vegetative shoot meristem grows into inflorescence meristems, and flowers are grown up (Melzer, *et al*., 2008).

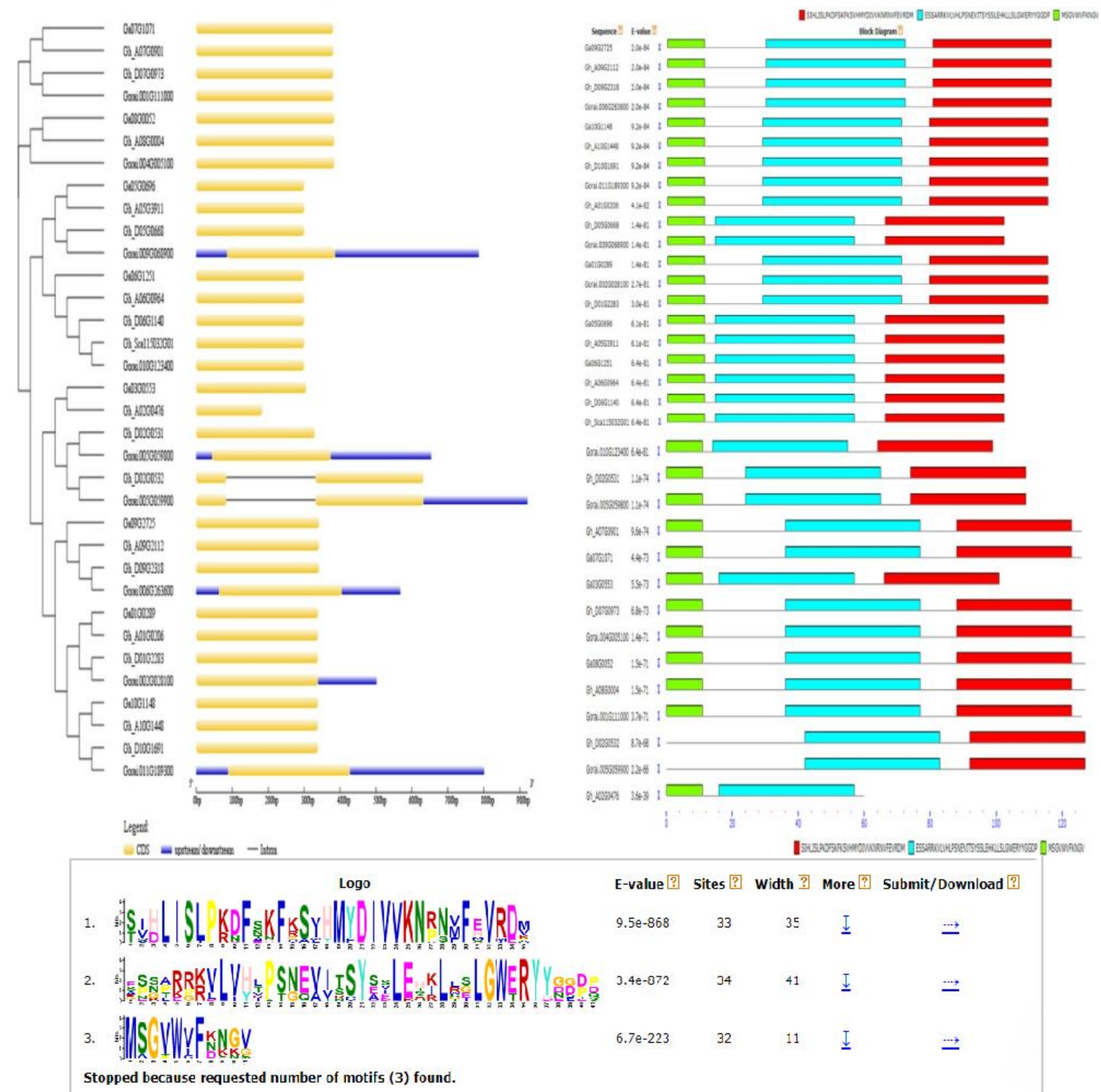

Figure 2. (A) Phylogenetic tree, gene structure, and motif of all the cotton genes. Yellow color indicates CDS, blue color indicates up/downstream, and the lines which are indicating intron. The phylogenetic tree was constructed with the aligned *GhFPF1* proteins sequence using the MEGA-X soft wear and the neighbor- joining method with bootstrap 1000 replicates. Gene structure was determined using gene structure display server (http;//gsds.cbi.pku.edu.cn). (B) Motif analysis was constructed using MEME soft wear (http;//meme.sdsc.edu/meme4_6_1/cgi-bin/meme.cgi). Common motifs which define the cotton and *GhFPF1* genes.

Extraction of RNA and CDNA synthesis: The extraction of RNA was isolated from different samples with using pure plant kit RNA prep (TIANGEN). The RNA purity and concentration were measured by (Nano, Spectrophometer Drop 2000) and with electrophoresis (Zhu, *et al*., 2008). RNA different profile pictured captured during doing RNA see in (Fig S1). The qRT-PCR was carried out CWBIO ultra SYBR Mixture (low ROX) 5ml. As written in the manufacturer's instructions form.

Primers: Three Genes were indiscriminately selected randomly. And the primers were intentional using Oligo7 and synthesized by Gene WIZ. Primers are listed in additional file.

Analysis for qRT-PCR attempt: The Quantitative RT-PCR was carried out using CWBIO Ultra SYBR Mixture (Low ROX) 5ml and ABI 7500 sequence sleuthing system. The cotton action was used as a house keeping gene to appraise the quantity of model cDNA in each amplification reaction (Livak and Schmittgen, 2001). The reaction was performed according to the protocol manufacturers.

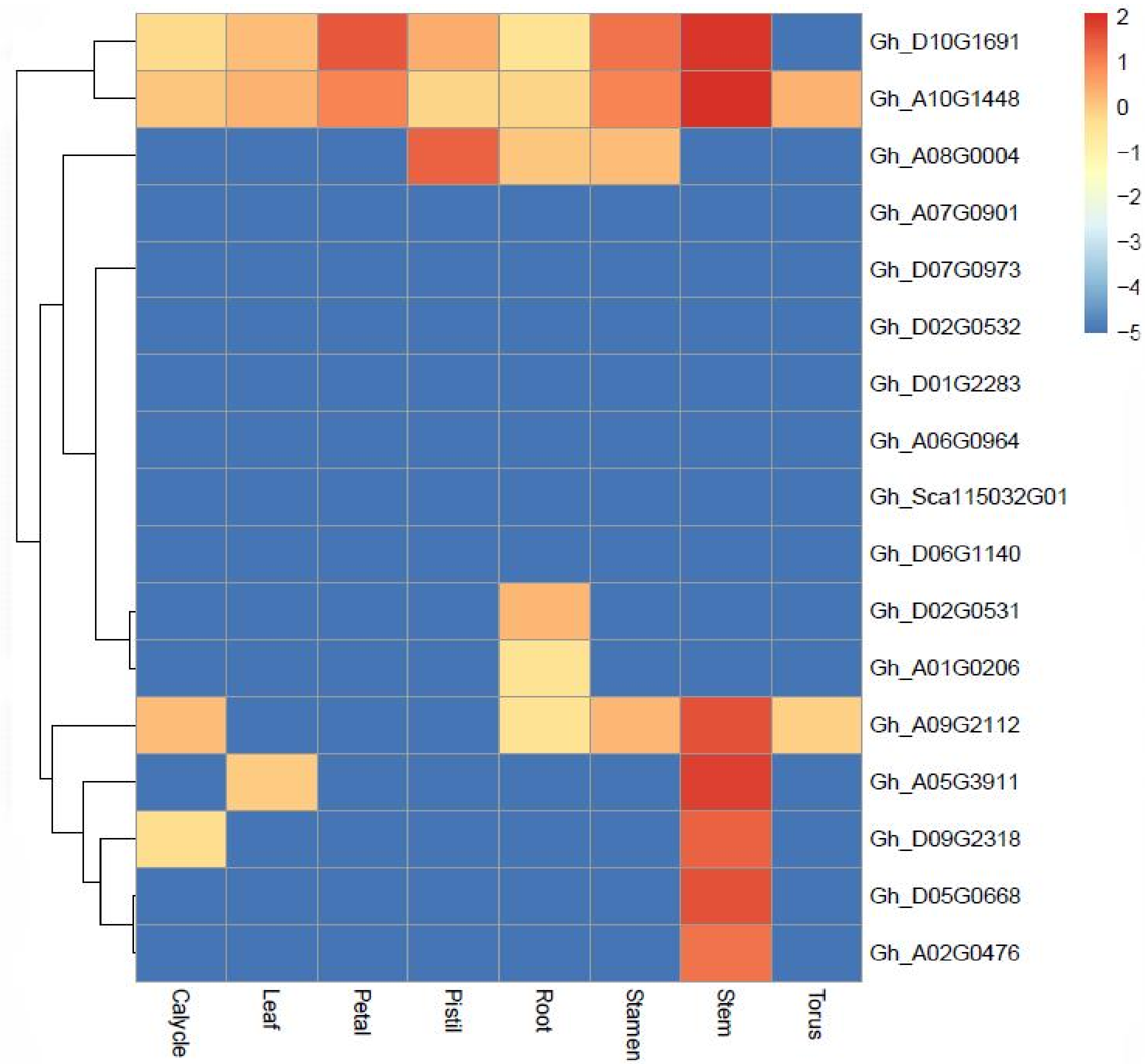

Figure 3. Expression Profile of 17 expressed genes with *GhFPF1* gene in different tissues. The different tissues of plants parts are shown on the bottom; and, the names of genes are listed to the right of the figure. Scale bars at the right hand whcich represents log2- transforemed FPFM values. The heat map was constructed from the log10 (FPKM) of the expression valiues by using R softwear.

Gene Ontology: Actually, gene Ontology (GO) is a method through which bioinformatics initiative to merge the presentation of gene and genomic product impute across all species. More importantly it intent to hold and evolve its controlled vocabulary of gene and genomic product attributes. Annotate genes and genomic products.

And disseminate and assimilate annotation data. The functions of GhFPF1 genes were find out using Blast 2GO (Meng, *et al*., 2011). Moreover, WEGO (Conesa and Stefan, 2009) was also used to envision these functional items. Amino acid sequences and protein sub cellular localization of *GhFPF1* genes were subjected to the WOLF PSORT programme website (https;//www.genscript.com/wolf-psort.html/) for protein and localization forecasting.

Biotic treatment and phytochromes: To observe the transcriptional levels of these genes under various a biotic stress and exogenous phytochromes treatments, *FPF1* gene and wild type WT seeds were grown in small pots at 28°C in a green house at sixteen-hour light/ eight hours dark cycle. When the plants are grown at cotyledons stage they were subjected to various treatments. At the stage of seedlings, the leaves were sprayed with (ABA), (IAA), (GA), and (Ethephon) with amount of 200 µM ABA, ethephon 0.5 $_{M}$M, GA 3460 µL, (Ye *et al*., 2006; Jin, *et al*., 2008). And then after treatments of these above growth regulators the samples were collected at 0, 6,12,24,36, and 48 hours respectively, samples were collected and put in liquid nitrogen quickly for frozen and stored at -80 °C for further RNA extraction. The research experiments were carried out in using complete randomized design.

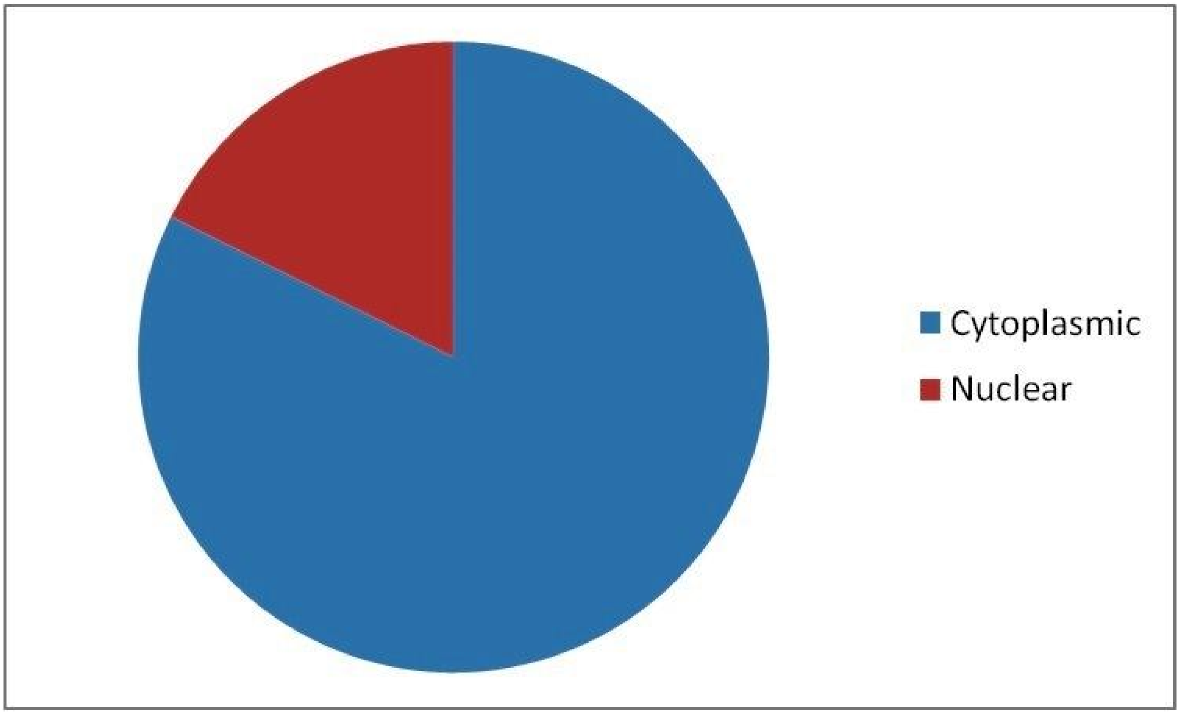

Figure 4. Protein localization for *GhFPF1* family gene forebode by using the programme Wolf PSORT website (http;//www.genscript.com/wolfpsort.html), the two colors are showing that the localization like as Cytoplasmic (blue) color, and Nuclear shows in (Red) color.

This studies first report on genomic-wide characterization and investigation of *GhFPF1* gene family in cotton genotypes. We are able to find out that there are total of 18 genes of *G. hirsutum*, eight genes in *G. arboreum*, and about nine genes in *G. raimondii*. *GhFPF1* genes were classified into three subgroups (*GhFPF1*, *GhFPF1* like protein 1, *GhFPF1* like protein 2). This study result agreed with previous studies reports in other species such as Arabidopsis, rice, and maize, (Wang, *et al*., 2014). Moreover, usually the *GhFPF1* gene in cotton belonged to subfamilies *FPF1*, and *FPF1* like protein 1, similar as described in previous studies in maize, rice, and in radish (Kania *et al*., 1997). The intron and exon structure pattern confirm subfamily characterization and subgroups appellative of phylogenetic tree investigations (He, and Wang, 2016). The same intron level was previously used to characterize plants associated with common origins (Meyer *et al*., 2008). In this our studies, *GhFPF1* gene closely related in the similar subgroups had the same function and structure in cotton. The tree analysis clustered *GhFPF1* into five subgroups. All the *GhFPF1* gene family members for different species (*G. hirsutum, G. arboreum,* and *G. raimondii*,), were distributed through the five groups with GhFPF1 subfamilies.

Recognition and analysis of *GhFPF1* gene in upland cotton: There are 3 cotton genome sequence which were downloaded from 3 various cotton genomic websites;

G.arborium is originally downloaded from Beijing institute of Genome (http//www.bgi.com/, Beijing china), *G.raimondi* protein sequence was obtained was downloaded from Institute of Joint Genome, www.phytozome.net, and *G.hirsutum* protein sequence was obtained from cotton Nanjing research cotton Institute of (http;//mascotton.njau.edu.cn, China), along with E-value < 0.01. The protected domain of GhFPF1 gene protein (Gh_D02G0531.1) which is mainly downloaded from protein Pfams families (http;//pfam.xfarm.net.cn, /).

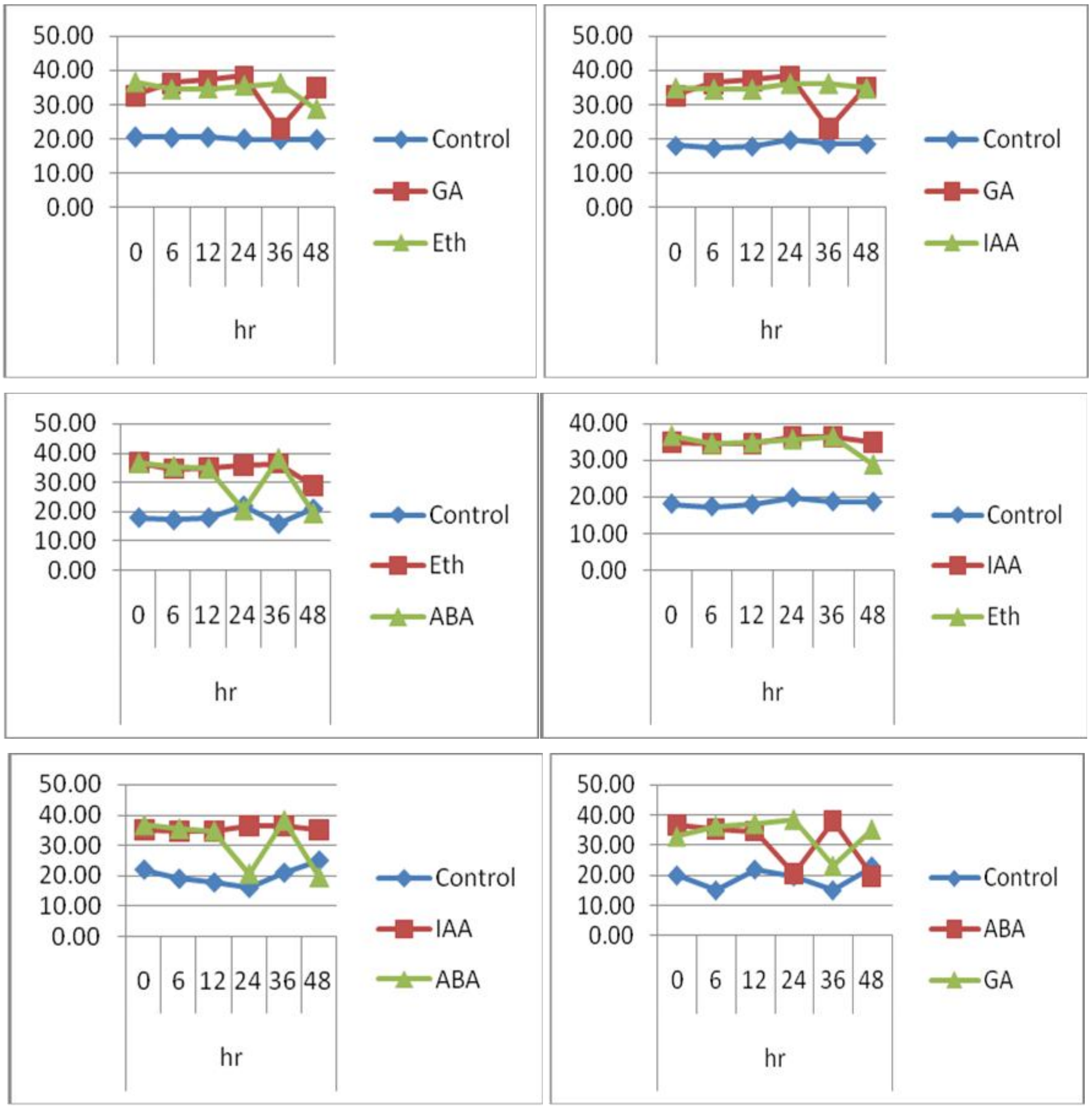

Figure 5. QRT-PCR different expression levels of treatments such as IAA, ABA, GA, and ETH growth hormones between wild type and *GhFPF1* gene, the samples were collected at 0, 6, 12, 24, 36, 48 hours.

Chlorophyll content: Weighing of total 200 mg leaf sections were collected from both WT & transgenic line, quickly put in liquid nitrogen after that put in ethanol of 5 ml (99.9%), then shocked in hot water bath at 80 c for 18 to 20 minute, the chlorophyll was measure in the extracts alcohol reading, using the precision extinction co-efficient chlorophyll content was measure (mg/g fresh leaves weight) was estimated at 100 x A654 (39.8 x sample fresh leaves weight) as suggested by (Tetley, and Thiemann, *et al*., 1974; Yang, 2002).

***GhFPF1* gene role under different plant hormones regulators:** Actually, the plant growth regulators are very important regulators of plant growth to abiotic factors. There are four most and specific plant hormones are Gibberellic Acid (GA), abscisic acid (ABA), Indolic

Acetic Acid (IAA), and Ethylene (ETH) (Cramer *et al*., 2011). Transgenic line FPF1 gene and wild type is grown in green house at optimum temperature 25C. The both seeds from transgenic line and wild species is grown in small pots soil when the seedlings are grown at cotyledons stage were highly induced by four hormones, for plant growth hormones prepare stock solution 200ml than we use the concentration amount of plant growth hormones.

Construct of Phylogenetic tree: To determine the evolutionary relationship a phylogenetic tree was constructed to know the relationship between the *G. hirsutum, G. raimondii,* and *G. arboreum* proteins sequence, a tree was constructed using method Neibour joining in MEGA-X soft wear (Kumar, *et al*., 2018). A total of 35 *GhFPF1* gene sequences from *G.hirsutum* (19), *G.arborium* (08), and *G.raimondi* (09), proteins sequences were downloaded from Tair website (http;//www.arabidopsis.org) and from NCBI website http;//www.ncbi.com/ and the full length of amino acid sequences of the *GhFPF1* alignment sequence were used to construct a phylogenetic tree. The *GhFPF1* full length of amino acid sequences were aligned by the clustalx1.81 analysis soft wear and for phylogenetic tree was constructed using MEGA-X programme (Figure 1). Finally, we construct the heat map using RNA sequence expression to formalize our results. The 17 *G. hirsutum* genes with different expression in *G. hirsutum* were selected to draw a heat map. The genes were both down regulated and up regulated in various developmental stages of different tissue organs, such as torus, stem, stamen, root, pistil, petal, leaf, and cotyledons (Figure 9). Total of 17 genes are divided into eight groups, Gh_D10G1691 and Gh_A10G1448 these both genes are expressed in all developmental stages of plants organs, while as from Gh_A07G0901 to Gh_D06G1140 were did not expressed in any part of the plant tissue, Gh_D02G0531, Gh_A01G0206, and Gh_A09G2112 were all expressed in the root, Gh_A09G2112, Gh_A05G3911, Gh_D09G2318, Gh_D05G0668, Gh_A02G0476 these all were expressed in stems.

Protein Subcellular Localization Prediction: Protein localization we used website (http;www.genescript.com/wolf-psort.html), to distinguished the *GhFPF1* family gene in cotton in the above figure-5 the protein localization for GhFPF1 family gene forebode by using the programme Wolf PSORT website the two colors are showing that the localization like as Cytoplasmic (blue) color, and Nuclear shows in (Red) color.

Analysis of phylogenetic tree and database searches: Among the cotton genomic database, 35 protein amino acid sequences of *GhFPF1* were aligned by the Clustalx 1.81 programme, for phylogenetic tree was constructed using MEGA-X soft wear (Kumar, *et al*., 2018) along an open gap penalty of ten and extension penalty gap of 0.2. Hence, a neighbor joining tree was constructed with support for each node, (Amura, *et al*., 2013). The protein alignment sequences were analyzed for motif recongnization and the motif were determined with the use of multiple Motif Elicitation programme (http; // meme.sdu.edu/meme4_6_1/meme.cgi), by (Bailey, 2009), and envision with an aligned sequence logo website generator (http; Weblogo.berkeley.edu/cn) Web logo. Multiple Em for motif Elicitation was run from the web server with the maximum to minimum parameter 6 to fifty amino acids, for every motif. The highest motif number was 15. A morphological analysis of *GhFPF1* gene was characterized using Gene morphological display server web (http; // gsds.cbi.pk.cn) (Guo, *et al*., 2007) through compare the predicted sequence with genomic sequence from cotton genome (Baloglu, *et al*., 2014).

Subcellular localization and chromosome mapping of the *FPF1* coupled G-protein receptors in cotton: The genes were not calculated in all chromosomes, in some diploid as well as tetraploid cotton species; this will be explicated by numerous genomics to be calculated in all the cotton chromosomes. A total of 17 chromosomes against *G.hirsutum*, 8 genes in *G.arborium,* and 9 genes in *G.raimondi*, respectively, out of 17 chromosomes of the diploid cotton species the *FPF1* genes were calculated in 6 chromosomes out of 17 in *G.hirsutum*, while as in *G.arborium* total 8 genes four genes calculated *FPF1* gene, whereas in *G.raimondi* out of 9 chromosomes 5 were mapped in *FPF1* gene respectively. This shows that these genes play a vital role as they possess high coverage of entire cotton genomes in spite of their few numbers. In association to the Subcellular localization all the bioinformatics tools used give same results (Table 2).

Table 2. Subcellular localization of the cotton *GhFPF1* candidate genes of G-protein coupled receptors in cotton.

Gene ID	Principle Transcript ID	Species	Assembly	Transcript Length (bp)	CDS Length (bp)	CDS GC Content (%)	Exon Number	Mean Exon Length (bp)	Mean Intron Length (bp)
Gh_D05G0668	Gh_D05G0668.1	Gossypium hirsutum	NAU	300	300	46.3	1	300	No intron
Gh_A05G3911	Gh_A05G3911.1	Gossypium hirsutum	NAU	300	300	46	1	300	No intron
Gh_Sca115032G01	Gh_Sca115032G01.1	Gossypium hirsutum	NAU	300	300	45.7	1	300	No intron
Gh_D06G1140	Gh_D06G1140.1	Gossypium hirsutum	NAU	300	300	45.7	1	300	No intron
Gh_A06G0964	Gh_A06G0964.1	Gossypium hirsutum	NAU	300	300	45.7	1	300	No intron
Gh_A09G2112	Gh_A09G2112.1	Gossypium hirsutum	NAU	342	342	45.3	1	342	No intron
Gh_D09G2318	Gh_D09G2318.1	Gossypium hirsutum	NAU	342	342	46.8	1	342	No intron
Gh_D10G1691	Gh_D10G1691.1	Gossypium hirsutum	NAU	339	339	45.4	1	339	No intron
Gh_A10G1448	Gh_A10G1448.1	Gossypium hirsutum	NAU	339	339	44.2	1	339	No intron
Gh_D02G0531	Gh_D02G0531.1	Gossypium hirsutum	NAU	330	330	42.7	1	330	No intron
Gh_A01G0206	Gh_A01G0206.1	Gossypium hirsutum	NAU	339	339	46.3	1	339	No intron
Gh_D01G2283	Gh_D01G2283.1	Gossypium hirsutum	NAU	339	339	47.8	1	339	No intron
Gh_D02G0532	Gh_D02G0532.1	Gossypium hirsutum	NAU	384	384	39.1	2	192	248
Gh_D07G0973	Gh_D07G0973.1	Gossypium hirsutum	NAU	381	381	50.7	1	381	No intron
Gh_A07G0901	Gh_A07G0901.1	Gossypium hirsutum	NAU	381	381	49.6	1	381	No intron
Gh_A08G0004	Gh_A08G0004.1	Gossypium hirsutum	NAU	384	384	53.1	1	384	No intron
Gh_A02G0476	Gh_A02G0476.1	Gossypium hirsutum	NAU	183	183	41.5	1	183	No intron
Ga05G0696	Ga05G0696.1	Gossypium arboreum	CRI	300	300	46	1	300	No intron
Ga06G1251	Ga06G1251.1	Gossypium arboreum	CRI	300	300	45.3	1	300	No intron
Ga09G2725	Ga09G2725.1	Gossypium arboreum	CRI	342	342	45	1	342	No intron
Ga10G1148	Ga10G1148.1	Gossypium arboreum	CRI	339	339	44.5	1	339	No intron
Ga01G0289	Ga01G0289.1	Gossypium arboreum	CRI	339	339	45.7	1	339	No intron
Ga03G0553	Ga03G0553.1	Gossypium arboreum	CRI	306	306	39.2	1	306	No intron
Ga07G1071	Ga07G1071.1	Gossypium arboreum	CRI	381	381	49.3	1	381	No intron
Ga08G0052	Ga08G0052.1	Gossypium arboreum	CRI	384	384	52.6	1	384	No intron
Gorai.009G068900	Gorai.009G068900.1	Gossypium raimondii	JGI	786	300	46.3	1	786	No intron
Gorai.010G123400	Gorai.010G123400.1	Gossypium raimondii	JGI	300	300	45.3	1	300	No intron
Gorai.006G263600	Gorai.006G263600.1	Gossypium raimondii	JGI	568	342	46.5	1	568	No intron
Gorai.011G189300	Gorai.011G189300.1	Gossypium raimondii	JGI	801	339	44.8	1	801	No intron
Gorai.005G059800	Gorai.005G059800.1	Gossypium raimondii	JGI	654	330	42.7	1	654	No intron
Gorai.002G028100	Gorai.002G028100.1	Gossypium raimondii	JGI	502	339	48.7	1	502	No intron
Gorai.005G059900	Gorai.005G059900.1	Gossypium raimondii	JGI	674	384	39.3	2	337	248
Gorai.004G005100	Gorai.004G005100.1	Gossypium raimondii	JGI	384	384	52.1	1	384	No intron
Gorai.001G111000	Gorai.001G111000.1	Gossypium raimondii	JGI	381	381	50.4	1	381	No intron

DISCUSSION

The plant creates mechanisms to control the reproductive and vegetative growth by integration of developmental and environmental prompt (Zhang *et al*., 2016). Sum of 35 similar homologous proteins were discovered in GhFPF1 gene, corresponding to 19 which is present *G. hirsutum*, 8 is present in *G. arboreum*, and 9 in *G. raimondii* as presented in the (Figure) 1 respectively. Using MEGA-X soft wear to construct a tree with Neighbor Joining (NJ) method; the tree is classified into four main groups, SF7, SF3, and *FPF1* gene. In our research study work it provides the first report on genomic-wide characterization and investigation of *GhFPF1* family genes in cotton genomes. We are able to find out that there are total of 18 genes which are mainly present in *G. hirsutum*, eight genes which is mainly present in *G. arboreum*, and Atleast nine genes in *G. raimondii*. As mentioned in the domain number, *GhFPF1* genes were classified into three subgroups (*GhFPF1*, *GhFPF1* like protein 1, *GhFPF1* like protein 2). Our result is agreed with previous studies in other species such as Arabidopsis, rice, and maize, (Wang, *et al*., 2014).

To estimate the biological possible functions, A heat map was constructed to see the expression level of different genes in different tissue, which includes Torus, stem, stamen, root, pistil, petal, leaf, and cotyledons at various developmental stages. Using the *G. hirsutum* (TM-1) transcriptome datasets (Zhang, *et al*., 2015). Gene expression profile FPKM value is the expression of genes in different plant parts so we constructed a heat map to investigate the expression value (Figure 3) Cluster analysis of 17 genes which is expressed in 7 different parts of the cotton tissues, genes names are listed on the right side of the figure. To evoluate the relationship between these genes a heat map was constructed from the log 10 (FPKM) of the expression valiues by using R softwear. In the heat map the Gh_D10G1691 the expresion value of this gene is expressed in all parts of the plant except in torus. Whereas Gh_A10G1448 this gene express in every plant tissue, and from Gh_A07G0901, Gh_D07G0973, Gh_D02G0532, Gh_D01G2283, Gh_A06G0964, Gh_Sca115032G01, and Gh_D06G1140 above these genes have no expression in even one plant parts or tissues as mentioned in the heat map.

To know the evolutionary relationship *GhFPF1* gene with *G. raimondii*, *G.hirsutum, G. Arboreum,* a phylogenetic trees was constructed *GhFPF1* gene proteins are classified into subfamilies according to subfamilies in *Gossypium*, we performed a phylogenetic tree analysis of total of 19 genes are present in the *G.hirsutum*, 9 genes in *G. arboreum*, and 8 genes are in *G. raimondii* to constructed unrooted phylogenetic tree. The *GhFPF1* genes were classified into four main subfamilies (I, II, III, and IV) as in (Figure 1, and additional File-1). For furthermore to know the diversification and conservation of the *GhFPF1*, gene their intron and exon structures and conserved motifs were conducted and were shown in (Figure 2). The subfamily members of 1 and members of 2 every one contains 3 exons and 4 introns respectively. Most of the members of subfamily presence 3 exons. While the other members which possess 2 to 3 exons. These results in (Figure 2) which indicates that functions of *GhFPF1* might be diverse. Meanwhile the closely related *GhFPF1* genes they showed similar intron and exons structure, so these genes they might play similar role in plant development and growth. (Figure 2).

Plant growth hormones play important roles in fiber strength, and development, including ethylene, auxin, jasmonic acid (JA) and gibberellic acid (GA), (Qin, *et al*., 2007; Hao, and Zhang, *et al*., 2012; Xiao, and Yin *et al*., 2010; Zhang, and Song, *et al.* 2011; Tan, and Zhang, *et al*., 2012). Our studies have shown that the ABA and Auxin can regulate the activity in flowering promoting factors (Laskowski, and Prusty, *et al*., 2006; Palusa, and Reddy, *et al*., 2007; Payasi, and Sanwal, *et al*., 2004), the expression features of 16 *GhFPF1* were explored after IAA treatment (Figure 5). Three genes were notably up-regulated, and two genes were down regulated, at the 24 h and 36 h (Figure 5a, and 5c). As mentioned in the figure 5 the different genes are expressed at various time periods. So, it indicates that some GhFPF1 genes might take part in the biological pathways regulated by many growth regulators. The aligned sequences were estimated through MUSCLE (R.C. Edgar, *et al*., 2004) both are used at the distance and Bayesian phylogenetic analysis. A tree was generated using Neighbor-joining, using excluding and including gapped characters, by using through Clustalx version 2.1 as previously used by (Larkin, and William, *et al*., 2007). The Bayesian analysis was carried out with Bayes version 3.1.2 (Ronquist, *et al*., 2003) using the parameters are followed as; ♈ portioning of sequence evolution, burn-in fraction of 25%, and 3 runs with 4 chains for 100,000 generations.

Conclusion: In our study research, we find out the tree structure, introns and exons structure and motif analysis of all *GHFPF1* gene family members in *G. hirsutum*, also, *GhFPF1* gene characterization and functions were studies by using through gene Ontology, promoter analysis and protein localization. The profiling, phytochromes and expression of abiotic stress trend of the *GhFPF1* genes were discovered. Our study which shows clearly that *GhFPF1* gene is potential regulators of various biological processes in the development and growth of cotton. Hence, the results which offer a good base on *GhFPF1* gene family especially their physiological role during abiotic and biotic stress.

Author Contributions: FS conceived the idea managed and supervised all the experiment, AHJ, MQ & FS designed the experiment, and grew the cotton, AHJ, XW, NM, and WL executed field experiment and collected data, AHJ & MAA processed data, performed stastical analysis. AHJ & MAA interpreted results and made tables and figures ready to be presented in manuscript. AHJ & MQ drafted manuscript. All authors read and approved the final manuscript for publication.

Funding; This research project was financially supported by the Henan Province key R & D and extension Project (182102110048) and Key Scientific Research projects in Henan Colleges and Universities (17B180001).

Acknowledgement: I am really very thankful to my entire lab colleague they help me during this research project. I also heartly thank full to Prof. Dr. Fan Shuli and teacher ma Qi Feng they guide me and support me during in this research experiment work.

Conflict of interests: The authors declare that there is no conflict of interests.

REFERENCES

Amura, K., G. Stecher, D. Peterson, A. Filip ski, and S. Kumar (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30(12):2725–2729.

Baloglu, MC (2014). Genome-wide in silico identification and comparison of growth regulating factor (GRF) genes in Cucurbitaceae family. Plant. OMICS. 7(4):260–270.

Bailey, T.L., M. Bodén, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, Li. W.W, and W.S. Noble (2009). MEME SUITE: Tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208.

Bernier, G., L. Corbesier, and C. Périlleux (2002). The flowering process: on the track of controlling factors in Sinapis Alba. Russ. J. Plant Physiol. 49: 445–450.

Bernier, G. and C. Périlleux (2005). A physiological overview of the genetics of flowering time control. Plant Biotechnol. J. 3: 3–16.

Biłas, R., K. Szafran, and K. Hnatuszko-konka (2016). Cis-regulatory elements used to control gene expression in plants. Plant Cell Tissue Organ Cult. 127(2):269–287.

Corbesier, L., C. Vincent, S. Jang, F. Fornara, and Q. Fan (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science 316: 1030–1033.

Conesa, A. and G. Stefan (2009). Blast2GO tutorial. Bioinformatics and Genomics Department Prince Felipe Research Center, Valencia SPAIN.

Cronn, RC., RL. Small and JF. Wendel (1999). Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. 96: 14406–14411.

Cramer, GR., K. Urano, S. Delrot, M. Pezzotti, and K. Shinozaki (2011). Effects of abiotic stress on plants: a systems biology perspective. BMC Plant. Biol. 11(1):163.

Escobar-sepu, HF., FC. Go, and M. Leo (2017). Expression patterns and promoter analyses of aluminum-responsive NAC genes suggest a possible growth regulation of rice mediated by aluminum, hormones and NAC transcription factors. PloS. One. 12(10): e0186084.

Gallagher, J.P., C.E. Grover, K. Rex, M. Moran, and J.F. Wendel (2017). A New Species of Cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). Syst. Bot. 42, 115–123.

Ge, L., H. Chen, JF. Jiang, Y. Zhao, and ML. Xu (2004). Over expression of OsRAA1 causes pleiotropic phenotypes in transgenic rice plants, including altered leaf, flower, and root development and root response to gravity. Plant. Physiol. 135: 1502–1513.

Guo, AY., QH. Zhu, X. Chen, and JC. Luo (2007). GSDS: a gene structure display server. Heredities. 29:1023–1026.

Hao, J., L. Tu, H. Hu, J. Tan, F. Deng, W. Tang, Y. Nie, and X. Zhang (2012). GbTCP, a cotton TCP transcription factor, confers fiber elongation and root hair development by a complex regulating system. *J. Exp. Bot.* 63, 6267–6281.

He, Q., DC. Jones, W. Li, F. Xie, J. Ma, R. Sun, and Q, Wang (2016). Genome-wide identification of R2R3-MYB genes and expression analyses during abiotic stress in *Gossypium raimondii*. Nature. Publishing Group, (august 2015), 1–14.

Jin, LG., and JY. Liu (2008). Molecular cloning, expression profile and promoter analysis of a novel ethylene responsive transcription factor gene GhERF4 from cotton (*Gossypium hirsutum*). Plant. Physiol. Biochem. 46(1):46–53.

Jung, C., and AE. Muller (2009). Flowering time control and applications in plant breeding. Trends. Plant. Sci. 14: 563–573.

Kania, T., D. Russenberger, S. Peng, K. Apel, and S. Melzer (1997). FPF1 promotes flowering in Arabidopsis. Plant. Cell. 9: 1327–1338.

Kumar, S., G. Stecher, and K. Tamura (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.

Molecular. Biology and Evolution, *35*(6), 1547-1549.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P .A. Mc. Gettigan, and H. Mc. William (2007). Clustal W and Clustalx version 2.0. Bioinformatics. 23:2947–2948.

Laskowski, M., S. Biller, K. Stanley, T. Kajstura, and R. Prusty (2006). Expression profiling of auxin-treated Arabidopsis roots: toward a molecular analysis of lateral root emergence. Plant. Cell. Physiol. 47(6):788–92.

Levy, Y.Y. and C. Dean (1998). The transition to flowering. Plant. Cell. 10: 1973-1990.

Li, J., SL. Fan, MZ. Song, CY. Pang, and HL. Wei (2013). Cloning and characterization of a FLO/LFY ortholog in *Gossypium hirsutum* L. Plant. Cell. Rep. 32: 1675–1686.

Librado, P., and J. Rozas (2009). DNA SP v5: software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25(11): 1451–1452.

Livak, KJ., and TD. Schmittgen (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods. 25: 402–408.

Melzer, S., F. Lens, J. Gennen, S. Vanneste, and A. Rohde (2008). Flowering-time genes modulate meristem determinacy and growth form in Arabidopsis thaliana, Nat Genet. 40: 1489–1492.

Melzer, S., G. Kampmann, J. Chandler, and K. Apel (1999). FPF1 modulates the competence to flowering in Arabidopsis. Plant. J. 18: 395–405.

Meng, Y., F. Liu, C. Pang, S. Fan, M. Song, D. Wang, and S. Yu (2011). Label-free quantitative proteomics analysis of cotton leaf response to nitric oxide. J. Proteome. Res. 10(12):5416–5432.

Meyer, Y., Y. Meyer, W. Siala, T. Bashandy, C. Riondet, F. Vignols, and JP. Reichheld (2008). Glutaredoxin and thioredoxin in plants. Biochim. Biophys. Acta. 1783:589–600.

Michaels, SD. and RM. Amasino (2001). Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. Plant. Cell. 13: 935–941.

Michaels, SD., and RM. Amasino (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. Plant. Cell. 11: 949–956.

Ondati, E., D. Lingling, G. Yaning, P. Chaoyou, W. Hengling, S. Meizhen, and Y. Shuxun (2016). GhNAC18, novel cotton (*Gossypium hirsutum L.*) NAC gene is involved in leaf senescence and diverse stress responses. Afri. J. Biotechnol. 15(24):1233–1245.

Palusa, SG., M. Golovkin, SB. Shin, DN. Richardson, and AS. Reddy (2007). Organ-specific, developmental, hormonal and stress regulation of expression of putative pectate lyase genes in Arabidopsis. New Phytol. 174 (3):537–50.

Payasi, A., PC. Misra, and GG. Sanwal (2004). Effect of phytohormones on pectate lyase activity in ripening Musa acuminata. Plant. Physiol. Biochem. 42(11):861–5.

Qin, Y.M., C.Y. Hu, Y. Pang, A.J. Kastaniotis, J.K. Hiltunen, and Y.X. Zhu (2007). Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis. Plant. Cell. 19, 3692–3704.

Roy Choudhury, A., and DN. Sengupta (2009). The promoter-elements of some abiotic stress-inducible genes from cereals interact with a nuclear protein from tobacco. Biol. Plant. 53(3):583–587.

Ronquist, F., and J.P. Huelsenbeck (2003). Mr Bayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

R.C., Edgar (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic. Acids. Res. 32:1792–1797.

Searle, I., Y. He, F. Turck, C. Vincent, and F. Fornara (2006). The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes. Dev. 20: 898–912.

Sheldon, CC., DT. Rouse, EJ. Finnegan, WJ. Peacock, and ES. Dennis (2000). The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC). Proc. Natl. Acad. Sci. 97: 3753–3758.

Simpson, GG., and C. Dean (2002). Arabidopsis, the Rosetta stone of flowering time? Science. 296: 285–289.

Smykal, P., R. Gleissner, L. Corbesier, K. Apel, and S. Melzer (2004). Modulation of flowering responses in different Nicotiana varieties. Plant. Mol. Biol. 55: 253–262.

Tan, J., L. Tu, F. Deng, R. Wu, and X. Zhang (2012). Exogenous jasmonic acid inhibits cotton fiber elongation. J. Plant Growth Regul. 31, 599–605.

Tetley, R.M., and K.V. Thiemann (1974). The metabolism of oat leaves during senescence: I. respiration, carbohydrate metabolism, and the action of cytokinin's. Plant. Physiol. 54, 294–303.

Wang, K., Z. Wang, F. Li, W. Ye, J. Wang, G. Song, and S. Yu (2012). The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet. 44*(10), 1098-1103.

Wang, X., S. Fan, M. Song, C. Pang, H. Wei, J. Yu, and S. Yu (2014). Upland Cotton Gene GhFPF1 Confers Promotion of Flowering Time and Shade-Avoidance Responses in Arabidopsis thaliana. PLOS ONE. *9*(3), e91869.

Wendel, J.F., C.L. Brubaker, and A.E. Percival (1992). Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. Am. J. Bot. 79, 1291–1310.

Wendel, JF., and RC. Cronn (2013). Polyploidy and the evolutionary history of cotton. Adv. Agron. 78: 139–186.

Wendel, J.F., C. Brubaker, I. Alverez, R. Cronn, and J.M. Stewart (2009). Evolution and natural history of cotton genus. Genet. Genome. 3, 3–22.

Wendel, J.F., and C.E. Grover (2015). Taxonomy and Evolution of Cotton Genus, Gossypium; Issue Agronmonogr; Cotton, American Society of Agronomy, Inc.: Madison, WI, USA; Crop Science Society of America, Inc.: Madison, WI, USA; Soil Science Society of America, Inc.: Madison, WI, USA, Volume 57.

Xiao, Y.H., D.M. Li, and M.H. Yin (2010). Gibberellin 20-oxidase promotes initiation and elongation of cotton fibers by regulating gibberellin synthesis. J. Plant. Physiol. 167, 829–837.

Yang, MW (2002). Study on Rapid Determination of Chlorophyll Content of Leaves. Chinese J. Spectroscopy Laboratory. 19: 479–481.

Ye, J., L. Fang, H. Zheng, Y. Zhang, J. Chen, Z. Zhang, and J. Wang (2006). WEGO: a web tool for plotting GO annotations. Nucleic. Acids. Res. 34:293–297.

Zhang, T., Y. Hu, W. Jiang, L. Fang, X. Guan, J. Chen, J. Zhang, CA. Saski, BE. Scheffler, and DM. Stelly (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat. Biotechnol. 33 (5):531–7.

Zhang, J., and J. Stewart (2000). Economical and rapid method for extracting cotton genomic DNA. J. Cotton. Sci. 4, 193–201.

Zhang, X., C. Wang, C. Pang, H. Wei, H. Wang, M., Song, and S, Yu (2016). Characterization and Functional Analysis of PEBP Family Genes in Upland Cotton (*Gossypium hirsutum* L.). *PLOS ONE. 11*(8), e0161080.

Zhang, M., X. Zheng, and S. Song (2011). Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. *Nat. Biotech.* 29, 453–458.

Zhu, Y., Y. Liu, Y. Li, Y. Shu, F. Meng, Y. Lu, and B. E. I. Liu (2008). Cis regulatory element-based gene finding: An application Arabidopsis thaliana. Genome. Informatics. 21: 177–187.