Journal of Biotechnology & Biomaterials

Ashfaq H, J Biotechnol Biomater 2021, 11:5

Research Article Open Access

Genome Wide Identification of FH2 Protein (Formin) Gene Family in *Homo Sapiens*

Hira Ashfaq*

Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan

***Corresponding author:** Hira Ashfaq, Department of Bioinformatics and Biotechnology, Government College University, Faisalabad, Pakistan, Tel: 3181723649; E-mail: hiraashfaq37015@gmail.com

Received date: September 06, 2021; **Accepted date:** September 20, 2021; **Published date:** September 27, 2021

Citation: Ashfaq H (2021) Genome Wide Identification of FH2 Protein (Formin) Gene Family in *Homo Sapiens*. J Biotechnol Biomater 11: 892.

Abstract

Formin (FH2) proteins are involved in the actin filament nucleation, elongation, microtubule organization, cytokinesis, regulation of cytoskeleton during cell division, and cell polarization. These genes consist of highly conserved FH2 domain which are involved in the assembly of actin filaments. In current study, a genome wide analysis of formin genes was performed in *Homo sapiens* by using *Arabidopsis thaliana* FH2 protein as a reference genome. In total, 89 members of FH2 containing formin genes were identified in *H. sapiens* and designated as HSF, respectively. Phylogenetic analysis clustered formin genes into seven subfamilies (INF, DRF, DIAPH, FHOD, FMN, FMNL, DAAM). Conserved domain analysis of the formin genes revealed that all the formin genes share a common FH2 domain. Phylogenetic analysis revealed that all the formin genes undergo duplication event during evolution, so show close relationship with each other. Serial cloner analysis reveals the graphical construction and restriction map of the INF gene. Current study offers enough understanding of formin genes in *H. sapiens*, that will be useful for additional functional analysis of these genes.

Keywords: Formin genes; FH2 domain; Phylogenetic analysis; Conserved domains; Restriction map; Actin assembly

Introduction

Formin protein play a role in actin filament nucleation and elongation and microtubule organization that helps to maintain cytoskeleton during cell division and migration. They are present in almost all the eukaryotes such as metazoan, fungi, slime molds, plants and humans. The globular actin proteins converted into linear actin filaments via formin proteins. These proteins are involved in different processes in different organisms such as budding of yeast cell, cytokinesis in fruit fly, and formation of fruiting bodies in *Dictyostelium*. They are also involved in cell polarization, adhesion, movement, and division in mammals.

Formin proteins are large proteins consist of approximately 1000 residues and different conserved domains such as Formin Homology 1 (FH1) and Formin Homology 2 (FH2) domains. The major domain of the formin proteins is FH2 domain which is conserved and consist of 350 amino acids approximately that are involved in actin filament assembly. The FH2 domain of the formin protein is present in different eukaryotes such as yeast consist of 2-3 formin proteins, *Drosophilla* contain 6 formin proteins, and mammals contain 15 formin proteins. These formin proteins are divided into nine subtypes but the formin proteins in mammals are divided into seven subtypes. The seven subfamilies include Diaphanous Related Formins (DRF), Formin Like Protein (FMNLs), Disheveled Associated Activator Of Morphogenesis (DAAM), Inverted Formin (INF), Formin Homology Domain Containing Protein (FHOD), Formin (FMN) [1].

The Diaphanous Related Formin (DRF) is a member of formin proteins that are involved in processes such as cell polarization and migration. The FMNLs encode for formin related protein and involved in the processes such as morphogenesis, cell polarity and cytokinesis. The protein encoded by DAAM gene are involved in the regulation of the cytoskeleton structure, and GTPase Rho activates via Wnt/Fz signaling for the control of cell polarization and migration. INF belongs to the formin protein family which consist of diaphanous inhibitory domain at its N-terminus and involve in polymerization and depolymerization of actin. FHOD also belongs to the diaphanous protein family and found in spleen. FMN gene encoded protein perform function in the production of adherent junctions and actin polymerization.

In the current study, a genome wide analysis was performed on FH2 protein (formin) gene family of the *Homo sapiens*. This study includes genome wide identification of formin protein in *H. sapiens*, phylogenetic relationship of these proteins, gene structure analysis, conserved domain analysis and restriction site analysis. This study based on the potential knowledge about evolutionary history and important features of the formin genes.

Material and Methods

Identification of formin genes in *Homo sapiens*

The sequences of formin genes were retrieved from NCBI. Protein sequences of formin gene from Arabidopsis thaliana were used as a query. BLASTP program was applied to identify the formin genes in *Homo sapiens*. All the selected protein sequences of formin genes were examined for conserved domain using the SMART Tool, pfam domain, and prosite. This process was done to separate out the

sequences that do not have conserved domains, and are not required for the analysis [2].

Gene structure analysis

The genomic and mRNA sequences of formin genes were retrieved from NCBI in the FASTA format for gene structure analysis. The Gene Structure Display Server (GSDS) program was used to display the organization of intron/exon, coding sequences, Untranslated Regions (UTRs) in the gene structure of Homo sapiens FH2 protein.

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignment was done to verify the conserved domains of the FH2 proteins (formin) genes in humans. The full length FH2 proteins from *Homo sapiens* were aligned in MEGA 7 and the evolutionary tree was formed which depends on the multiple sequence alignment. Maximum Likelihood (ML) method was used in building of phylogenetic tree.

Serial cloner analysis

For precise mapping of the genes encoding for FH2 proteins, the sequences were retrieved from NCBI. The Serial cloner tool used to analyze the patterns of the sequences and their graphical construction, finding restriction sites in particular sequences.

Results and Discussion

Identification of formin genes in *Homo sapiens*

The FH2 protein sequence of *A. thaliana* (query sequence) was obtained from NCBI. The query sequence was used in BLASTp search in NCBI against Homo sapiens, and a total of 89 genes of FH2 protein (formin) gene family from *Homo sapiens* were found. An online tool SMART and pfam domain and prosite was used to identify the conserved domains of these sequences. In the present study, 89 FH2 protein (formin) genes were identified from *Homo sapiens* by using *A. thaliana* FH2 protein sequence as a query sequence. In previous studies, Rivero et al. identified 10 formin genes in the genome of *Dictyostelium discoideum*. Identified 12 formin related proteins present in the *Leishmania spp* at NCBI, Zigmond identified 9 formin genes in mammalian genomes, Higgs and Peterson identified 101 formin genes in different eukaryotes and 15 genes specifically in mice, Pruyne identified 9 formin subtypes in metazoan. This data showed the relevance of formin genes with different types of organisms [3].

Conserved domain analysis

The INF2 consist of WH2 domain which consist of 18 residues actin binding motifs. It is present in all eukaryotes, some bacteria and viruses and perform major function in the mammalian cytoskeleton regulation. The WH2 forms tandem repeat and act as a modular part of the protein and work in association with other domains such as WH1 domain and CRIB domain. The FMN2 consist of conserved domain DEP. The globular DEP domain consists of 80 amino acids which is involved in G signaling pathway. The major role of DEP domain is to target DEP domain-containing proteins to specific G protein signaling pathway and to specific sites of the membranes at subcellular level. Formin proteins involved in cytoskeletal processes and work as a key regulator in the nucleation and elongation of actin filament. The INF2, FMN2, FMNL3, DAAM1, FMNL1, FMNL2, DIAPH2, GRID2IP, consist of FH2 domain. FH2 domain is a highly conserved region of 400 amino acids which forms a dimer in head to tail ring shape and bind to the actin filament at its sarcastic end.

There are five subdomains of FH2 domain such as N-terminal “lasso”, a “linker” part, a globular “knob” part, a “coiled-coil” region and a C-terminal “post” subdomain.

The interaction between “lasso” and “post” of the other FH2 domain forms a closed ringed structure which results in the formation of dimer. The INF2, DAAM1, FMNL2, DIAPH2, FMNL3 consist of two conserved domains GBD-FH3.

The Rho GTPases Binding Domain (GBD) consist of 380 amino acids in length. The GBD N-terminal region involved in localization of the intracellular processes by encountering different targets.

The GBD also work as a regulator in the activation of intramolecular interaction between Diaphanous Autoregulatory Domain (DAD) and the N-terminal. It is a α-helical domain and consist of three subdomains such as GBD N-terminal segment, an armadillo repeats region (ARR) and dimerization region.

The DAAM1, FMNL1, FMNL2, DIAPH2, FMNL3 consist of the DAD domain. The DAD domain forms α-helical structure and consists of 32 amino acids approximately and known as autoinhibitory domain which involves in intramolecular binding.

The GRID2IP formin protein consist of PDZ domain which is globular conserved domain consist of approximately 80-100 residues. In eukaryotes, PDZ domain present as tandem repeat units in the form of single or multiple copies in a diverse range of proteins.

Some copies of PDZ also found in bacteria proved by horizontal gene transfer. These PDZ domains are involved in the organization of protein network on membrane and play a significant role in the clustering of molecules that are involved in signaling pathways [4].

Phylogenetic analysis

For multiple sequence alignment and phylogenetic analysis, protein sequences of FH2 protein (formin) were taken from *Arabidopsis thaliana* and *Homo sapiens*.

The phylogenetic analysis was done by using Mega 7 to generate a rooted tree based on the maximum likelihood (ML) method (Figure 1).

The evolutionary analysis clustered 14 formin proteins in humans that are divided into different isoforms with their orthologues from *A. thaliana*. All members of the formin proteins are based on homology of the specific region and conserved domains.

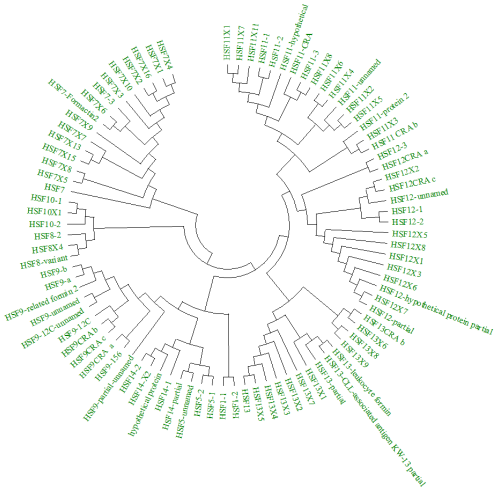

Figure 1: Ancestral states were inferred using the Maximum Likelihood method.

During evolution, it was thought that FH2 domain evolve only for one time which means that the formin proteins and their subfamilies undergoes repeated duplications of ancestral formin proteins that lead to the sequence divergence over time.

Moreover, a comparison between important features of all the formin genes of *H. sapiens* were discussed in Table 1. The retrieved data described that all formin genes of *H. sapiens* contain conserved FH2 domain from *A. thaliana*.

Similar proteins were observed in previous reports in the formin protein gene family in *Arabidopsis thaliana*, *Dictyostelium discoideum*, *Leishmania spp*, metazoan, mice, and mammals.

Proposed names	Gene locus	Exons	Chr #	ORF length	Amino acid length	Start of genomic location
HSF1	INF2	23	14q32.33	3720	1240	5001
HSF2	DIAPH1	30	5q31.3	3789	1263	5056
HSF3	FHOD1	26	16q22.1	3570	1190	5042
HSF4	DIAPH3	38	13q21.2	3579	1193	5058
HSF5	DAAM1	26	14q23.1	3204	1068	5057
HSF6	FMN1	29	15q13.3	3588	1196	5020
HSF7	FHOD3	33	18q12.2	4266	1422	5018

Table 1: Proposed nomenclature and important features of formin genes in Homo sapiens.

In mice, formin proteins was first identified and studied. Formins play a significant role in the nucleation and elongation of the actin filament and in the regulation of cytoskeletal structures and subcellular processes. Formin proteins consist of highly conserved domains such as Formin Homology 1 (FH1) and Formin Homology 2 (FH2) as well as a less conserved domain such as Formin Homology 3 (FH3). Formins have the ability of f dimerization and form homodimer by using dimerization domain of both FH3 domain and FH2 domain. FH2 domain nucleates actin by making head to tail dimer. For filament elongation, FH1 domain activates profilin associated G-actin filaments and gets closer to the sarcastic end of the domain. There is total 15 members of the formin protein gene family that are present in mammals. These members can be further divided into different isoforms based on their protein domain structure. All these formin proteins have their own mechanism of localization, binding, and regulation and have variable ability and strength in the nucleation and elongation of the actin filament [5]. Formins also help in the regulation of microtubules to maintain the cytoskeleton. There are some non-FDD type formin homology proteins such as FHOD1, FHOD3, FMN1, FHDC1 and GRID2IP and some belongs to the FDD type homology proteins such as DIAPH1, DIAPH2, FMNL1, FMNL2, FMNL3, DAAM1 and DAAM2. INF2 gene belongs to the formin family of proteins. At the N-terminal of this protein consist of diaphanous inhibitory domain so also known as diaphanous formin. In mouse, this protein plays a major role in in polymerization and depolymerization of actin filaments. Defects in this gene results in the production of focal segmental glomerulosclerosis. FMN2 is a formin homology protein that perform function in the regulation of the actin cytoskeleton and maintain polarity of the cell. The set of states at each node is ordered from most likely to least likely, excluding states with probabilities below 5%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value.

The rates among sites were treated as being uniform among sites (Uniform rates option). This analysis involved 89 amino acid sequences. There were a total of 2360 positions in the final dataset. Evolutionary analyses were conducted in MEGA 7. DIAPH1 is a member of the diaphanous subfamily of the FH protein family. This

gene performs major function in the growth and regulation of ovaries. Mutation in this gene product results in the production of ovarian failure. FHOD3 protein belongs to the Diaphanous-Related Formin (DRF). It consists of domains, such as GTPase-binding domain (GBD), Diaphanous Inhibitory Domain (DID), Formin Homology 1 (FH1), Formin Homology 2 (FH2), and Diaphanous Auto-Regulatory Domain (DAD). In cardiomyocytes, FHOD3 protein work for polymerization of actin filament.

Serial cloner analysis

The Serial cloner tool is used to analyze the patterns of the sequences and their graphical construction, finding restriction sites in particular sequence. The graphical construction of the IFN2 gene results in the production of restriction map which shows all the unique sites that are present on the IFN2 genomic sequence for cloning (Figure 2).

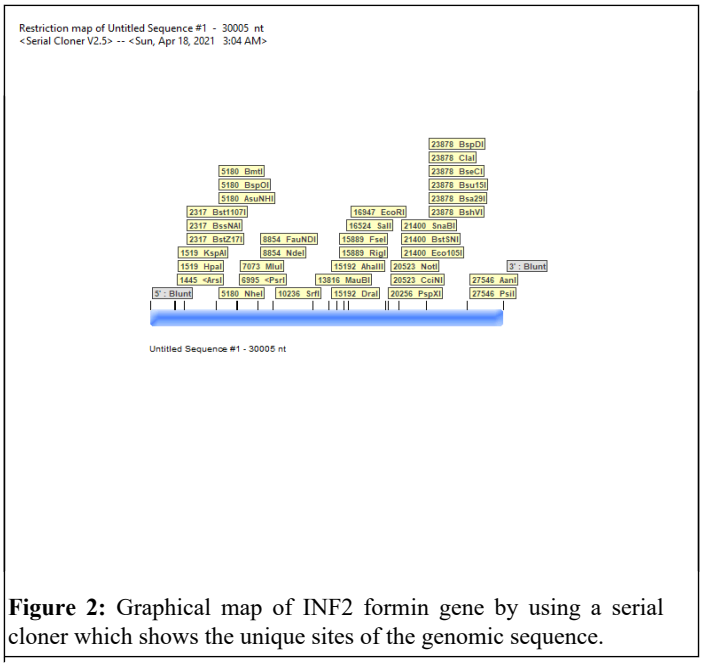

Figure 2: Graphical map of INF2 formin gene by using a serial cloner which shows the unique sites of the genomic sequence.

These unique sites specific for restriction enzymes that are required during cloning. The restriction map of the IFN2 protein also constructed by using serial cloner which shows all the unique sites on the protein sequence. After the identification of unique sites, virtual cutter analysis of the protein sequence of INF2 formin gene of Homo sapiens by using Serial cloner is done which is incubated with different enzymes such as Acc1I+AccI+AccIII+Acc361+AciI and resulted into 3 fragments. This bioinformatic analysis is easy to develop restriction map and PCR analysis to know which restriction endonuclease site is present in the targeted sequence.

Conclusion

This study reveals the identification of FH2 protein (formin) in *H. sapiens*. This study reveals closed phylogenetic relationship of formin proteins in *H. sapiens* and *A. thaliana*. This study reveals highly conserved domains of formin genes of *H. sapiens*. This study also reveals the genomic and proteomic restriction map of the IFN gene which plays a beneficial role during cloning and PCR to analyze restriction endonuclease for the specific sites. Current study offers sufficient information about formin genes in *H. sapiens* that will be useful for more detailed functional analysis of these genes.

References

1. Zigmond SH (2004) Formin-induced nucleation of actin filaments. Curr Opin Cell Biol 16: 99-105.
2. Huckaba TM, Pon LA (2002) Cytokinesis: Rho and formins are the ringleaders. Curr Biol 12: R813-R814.
3. Chesarone MA, DuPage AG, Goode BL (2010) Unleashing formins to remodel the actin and microtubule cytoskeletons. Nat Rev Mol Cell Biol 11: 62-74.
4. Pruyne D (2016) Revisiting the phylogeny of the animal formins: Two new subtypes, relationships with multiple wing hairs proteins, and a lost human formin. PLoS One 11: e0164067.
5. DeWard AD, Eisenmann KM, Matheson SF, Alberts AS (2010) The role of formins in human disease. Biochim Biophys Acta 1803: 226-233.