JFLS | 2016 | Vol 1(1) | Pp 8-15

Research Article

Genome-wide analysis of DnaJ protein family in Zebrafish (*Danio rerio*) and comparative phylogenetic analysis with four fish models

Sivaraman Balasubramanian*[1], Renuka Vijayakumar[2], Elavarasan Krishnamoorthi[2] and Shanmughavel Piramanayagam[3]

[1]*Fisheries College and Research Institute, Thoothukudi, India*
[2]*ICAR-Central Institute of Fisheries Technology, Cochin, India*
[3]*Department of Bioinformatics, Bharathiar University, Coimbatore, India*

ISSN: 2456-6268

ARTICLE INFO
Received:18 September 2016
Accepted: 18 December 2016
Available online: 27 December 2016

KEYWORDS
Heat shock proteins
Danio rerio
DnaJ protein
Genome wide analysis
Phylogenetic analysis
Fish models

***CORRESPONDENCE**
sivabala17@gmail.com

ABSTRACT

Polypeptides which synthesized in cell are more prone to interact with the cellular components that may lead to undesirable folding functions. This situation is taken care of a group of molecules called Heat shock proteins (Hsps) come under molecular chaperons. They play -several important roles such as processing of newly synthesized polypeptides into functional native conformation, prevention of degradation of polypeptides, translocation of newly synthesized polypeptides across the membrane and maintaining the homeostasis of protein under stress conditions in an organism. Among Heat shock proteins, Hsp70and DnaJ proteins are well established molecular chaperon machinery for their various cellular functions. Binding of Hsp70s with substrate polypeptides for executing any functions are associated with an ATP hydrolysis cycle and- induced by DnaJ/ Hsp40 family. Knowledge of the correct phylogenetic relationship among vertebrates is crucial for the valid interpretation of evolutionary trends in DnaJ proteins. In the present computational analysis, identification and systematic classification of DnaJ proteins using genome wide analysis of zebrafish (*Danio rerio*) has been outlined. Phylogenetic analysis of DnaJ protein from five fish models, *Danio rerio*, *Takifugu rubripes*, *Oryzias latipes*, *Gasterosteus aculeatus* and *Tetraodon nigroviridis* reveal that a close relationship prevails between fugu and tetraodon, medaka and stickleback. Whereas zebrafish have shown independent evolution. Apart from phylogenetic analysis a comprehensive domain structural organisation, sub cellular localization and gene distribution of zebrafish DnaJ proteins have also been documented.

INTRODUCTION

Polypeptides which are synthesized *in vivo* do not simply fold - into --appropriate structure and take the confirmation. Inside the cell different nature of molecules are present and interact differently with the newly synthesized polypeptides. This interaction adversely affects the native confirmation and function of newly synthesised polypeptides. These effects are overcome by the group of molecules called heat shock proteins (Hsps) / molecular chaperons which regulates many molecular functions like protein degradation, protein folding and translocation of proteins across the membrane (Craig, 2006).

Based on the molecular weight, Hsps have been placed under six categories namely Hsp70, Hsp40, Hsp90, Hsp100, Hsp60 and small Hsp family (Georgopoulos and Welch, 1993). Among them DnaJ proteins have been reported to be scattered in the cell organelle including cytosol, nucleus, endoplasmic reticulum and mitochondrial matrix (Yamamoto *et al.*, 2008). Among the Hsps, Hsp70 proteins play a major role in the regulation of many kinds of cellular process which is essential for the survival under both normal and stress conditions (Bukau *et al.*, 2006).

The number and type of Hsp70 molecules in each cell organelle of Prokaryotes and eukaryotes may vary from one to two. Each type of Hsp70 protein involves in a variety of cellular functions and the specificity of these proteins in a particular process is determined by a group of proteins called DnaJ / Hsp40 proteins and associated chaperons (Craig, 2006). The number of DnaJ / Hsp40 protein family and the number of Hsp70 family are not always equal in a particular organism / organelle. Compared to Hsp70 family, number of members in DnaJ protein family is higher, resultant to multiple DnaJ proteins interaction with same type of Hsp70 protein and forms a different complex network to perform a variety of cellular process (Qiu *et al.*, 2006). The well characterised and most abundant Hsp70 proteins are activated by several co-chaperons, especially DnaJ/Hsp40 protein family which stimulates Hsp70 ATP hydrolysis, there by regulating Hsp70 and client protein

interaction (Hennessy *et al*., 2005). Hsp70 proteins selectively bind with unfolded hydrophobic regions of substrate (polypeptides), and their activity is controlled by the cycle of ATP binding, hydrolysis and nucleotide exchange (Takayama *et al.*, 1999). This ATP hydrolysis turns Hsp70 proteins from open state with high association and dissociation rates for substrates into a closed state with low exchange rates and stabilise the Hsp70s interaction with substrates (Bukau and Horwich, 1998).

Throughout the evolutionary course, molecular chaperons are highly conserved. The DnaJ families of molecular chaperons share this conservation only within a 70 amino acid domain called J-domain, outside this region the family is divergent (Cheetham and Caplan, 1998). This J-domain comprises 4 helices namely I, II, III and IV. A highly conserved Histidine, Proline and Aspartic acid tri-peptide motif (HPD motif) forms a loop region between helices II and III and this motif role is functionally critical (Tsai and Douglas, 1996). Apart from the J-domain, DnaJ proteins contain other conserved domain regions such as Glycine / Phenylalanine rich flexible region (Gly / Phe region), classical repeats of cysteine (CXXCXGXG) motif called Zinc-Finger domain and less conserved C-terminal domain. Gly / Phe region have a greater degree of conformational plasticity thus serves as a flexible linker region and controls the specificity of DnaJ protein (Pellecchia *et al.*, 1996). Zinc-Finger domain plays a critical role in substrate binding and assists Hsp70 for protein folding (Lu and Cyr, 1998). C-domain plays an important role in dimerization activity of Hsp70 proteins and involve in the interaction with a wide variety of client polypeptides (Wu *et al*., 2005). Based on these domain structures DnaJ / Hsp40 proteins are classified into three different types. Type I proteins possess the J-domain, Gly / Phe region, zinc-finger domain and C-domain and Type II proteins lack zinc-finger domain whereas Type III hold only the J-domain (Cheetham and Caplan,1998). Even though Type I and II proteins differ from their conserved region they have the similar functional property and bind to non-native substrates. Type III proteins may not be bind with non-native polypeptide substrates. Apart from this, some DnaJ family members having additional domains may determine the functional diversity of the DnaJ proteins (Jermy *et al.*, 2006).

Although the function of plant Hsps are well understood by extensive researches, a relatively less focus has been given to understand their function in fish models. The primary objective of this paper is to redirect the focus on uncovering molecular details of DnaJ protein and their evolution among fish models. By utilizing comparative and genome wide analysis, DnaJ protein family have been annotated based on sequence homology, molecular structure and protein localization within the five fish models.

MATERIALS AND METHODS

Sequence retrieval and conformation of DnaJ protein

A series of searches were performed to retrieve the zebrafish DnaJ gene and protein from publically available database NCBI using several key words such as DnaJ, J-protein and Danio DnaJ etc. To find a potential DnaJ protein sequence and to avoid redundant sequence, reference sequence option was used (http://www.ncbi.nlm.nih.gov/). These sequences were used as a query in blast programme to identify the corresponding genes with expected value of 0.001 from the Ensembl database (http://www.ensembl.org). A primary sequence similarity search and phylogenetic analysis were carried out to identify duplicates, alleles and partial sequences. These selected sequences were used as references in blast programme for further retrieval of DnaJ protein sequences of four fish models [Fugu (*Takifugu rubripes*), Medaka (*Oryzias latipes*), Stickleback (*Gasterosteus aculeatus*) and Tetraodon (*Tetraodon nigroviridis*)] genome from ensemble database.

Nomenclature

Nomenclature is a complex issue to nominate the DnaJ genes and to avoid contradiction with other DnaJ genes, a complete systematic nomenclature was proposed for zebrafish and other four fish DnaJ genes, based on the classification of mammalian DnaJ proteins with slight modification (Ohtsuka and Hata, 2000). Each protein was named by representing the source organism with one upper (genus name) and one lower (species name) case letters (Dr), DnaJ homology by Dj, and the types I, II, and III by A, B, and C respectively. The chronology of DnaJ-protein has been denoted by Arabic numeral and the variants were represented by lowercase alphabets.

Conserved domain and sub-cellular localization analysis

Protein sequences acquired from the retrieval were subjected to domain prediction analysis using Pfam HMM database to find the J-Protein specific domains with E-value of 0.001 as the cut-off (Bateman *et al.*, 2002). Proteins containing detectable J-domain were regarded as DnaJ protein, otherwise they were excluded from the data set. For the confirmation of conserved motifs in DnaJ protein and genes, the sequences were analysed using Multiple EM for motif elicitation programme in MEME suit (Bailey *et al.*, 2006). Sub cellular localization of the DnaJ proteins were analysed by WoLF PSORT and Euk-mPLoc 2.0 (Horton *et al.*, 2007; Chou and Shen, 2010).

Sequence alignment and Phylogenetic analysis

Sequence alignment was generated in CLUSTAL OMEGA with BLOSUM as the protein weight matrix (McWilliam *et al.*, 2013). Several values for gap opening penalty and gap extension penalty were tried to identify the commonly resolved domain. A combination of gap opening penalty (6.0) and gap extension penalty (0.20) were finally adopted which enable reasonable alignment among conserved domains with few gaps and resulting alignment was manually refined in Bioedit with the removal of poorly aligned positions (Hall, 1999).

Four data sets were taken for phylogenetic analysis. First, all the three types of DnaJ proteins from zebrafish were analyzed and then the relationship of Type I, II and III proteins from zebrafish, fugu, medaka and stickleback were evaluated. The phylogenetic tree was reconstructed using two different approaches, Bayesian and Neighbor-Joining (NJ) methods. Bayesian analysis was conducted in MrBayes 3.1.2 with the gamma invariant sites (Ronquist and Huelsenbeck, 2003). Four independent runs of 1500000 generations, each with four metropolis-coupled Monte Carlo

Markov Chains (MCMCMC), were run in parallel. Markov chains were sampled every 10 generation and the first 25 % of the trees were discarded as burn in and the remaining were used to compute the majority rule consensus tree, the posterior probability of clades and branch lengths. Convergence of four runs was assessed by checking the average standard deviation of split frequency (below 0.01) and the potential scale reduction factor (PSRF- very close to 1.00 for all parameters). Neighbor joining phylogeny was performed in Phylip. Seeqboot was used to produce 2000 replicates for bootstrap analysis to test the significance of nodes. Distance matrix was calculated by using DnaDist program. Neighbor Joining analysis was conducted by neighbor program and finally consensus tree was built using Consense program. Finally the trees were presented by using Figtree and iTOL packages (Letunic and Bork, 2006).

Gene distribution and mapping DnaJ protein on zebra fish chromosomes

For the analysis of physical location of DnaJ genes on zebrafish genome, manual detection was carried out in ensemble database. By using blast program starting and end positions of the DnaJ gene on each chromosome were confirmed. Finally, the chromosome map showing the physical locations of all DnaJ genes was generated using the Genome Pixelizer (Kozik *et al.,* 2002).

RESULTS AND DISCUSSION

Sequence retrieval and conformation of DnaJ protein

Using blast programme in NCBI database, 183 DnaJ protein sequences were obtained, in which repetitive, isoform and fragment proteins were discarded. Careful alignment was carried out among those sequences to identify the redundant sequences. Sequences with more than 95% similarity at the DNA level were considered as likely alleles, unless they were previously reported to be different genes (Zhang *et al.,* 2001). After the removal of redundant, allele and fragment sequences, 72 sequences were retained. These sequences were used as blast queries in ensemble database and 49 genes were confirmed as DnaJ protein genes in zebrafish and their respective protein sequences were also retrieved. A primary phylogenetic analysis was conducted to check the sequence redundancy. After this second confirmation 6 genes were excluded from the dataset and 43 genes used for further analysis. Using the same approach as mentioned above 44, 46, 49, and 50 DnaJ genes were retrieved from fugu, medaka, stickleback and tetraodon respectively.

Conserved Domain and Sub-cellular Localization Analysis

Pfam HMM database was used to predict the presence of respective domains in the confirmed data sets. In zebrafish among the 43 sequences 3 were found to have J-domain, Gly / phe region, Zinc-Finger domain and C-domain and they were placed under type I DnaJ-Protein according to the classification given by Cheetham and Caplan (1998). Another seven protein sequences were noticed to be lacking Zinc-Finger domain and they were included in Type II DnaJ-proteins. In the rest of 33 sequences, 30 sequences were found to have only J-domain and classified as Type III DnaJ protein. Due to absence of J-domain another 3 sequences were excluded and only 40 genes have taken for the further analysis (Table 1). The Domain organization of zebrafish DnaJ proteins has shown in Figure 1.

Other four fish models fugu, medaka, sdtickleback and tetraodon were selected for evaluating the distribution of J-domain and found to have 42, 44, 47 and 43 DnaJ protein sequences respectively. In all the four models namely fugu, medaka, stickleback and tetraodon, maximum of 32, 34, 36 and 33 sequences comes under type III DnaJ protein respectively. 3 of 42 from fugu, 4 of 44 from medaka, 4 of 47 from stickleback and 3 of 43 from tetraodon sequences belong to type I DnaJ protein. 7 sequences of fugu, stickleback and tetraodon and 6 sequences of medaka were classified under type II DnaJ protein. From the results it can be suggested that type III protein sequence distribution is prominent in zebrafish and other fish models. All the five fish models are having a close relationship with respect to type III DnaJ protein sequences. Comparatively type I DnaJ proteins are less distributed in five studied fish species. In ensemble database six sequences of zebrafish were annotated under type II of DnaJ-protein, though they are having only J domain. In the present genome wide analysis we were classified those sequences under type III DnaJ protein based on the classification given by Cheetham and Caplan (1998).

Sequence alignments and Phylogenetic analysis

The-phylogenetic analysis of DnaJ proteins within the zebrafish revealed the 40 DnaJ proteins can be divided into three types. Each type of DnaJ protein formed a relatively separate clade and the relationship among these clades was also generally consistent (Fig. 2). Type III and type II DnaJ proteins are more closely related compared with type I DnaJ protein. Type III members form the largest cluster in the DnaJ-protein family and all of them lacked Gly / phe region, Zinc-Finger and C-domain, which made them more independent in the evolutionary tree. It is interesting to note that 6 sequences of DnaJ proteins were classified under type II in Ensembl database which were classified in the present computation analysis under type III as they are having only J-domain.

To gain an understanding of the evolution of DnaJ protein among fishes, zebrafish DnaJ proteins were compared with other available 4 fish models. Totally 216 putative Dnajprotein sequences, including 40 from zebrafish, 42 from Fugu, 44 from madeka 47 from stickleback and 43 from tetraodon were analysed. Form this phylogenetic analysis DnaJ protein genes were clustered into three groups (I, II and III). The first data set contain only Type I DnaJ genes (17 genes) from all the five fish models likewise the second data set contain 34 type II DnaJ genes and Third datasets contain 165 type III DnaJ genes.

Phylogenetic tree of type I DnaJ gene shows that all the five fish models share DjA2 gene (Fig. 3). DjA1 genes were present only in medaka, stickleback and zebrafish. DjA1 genes are not traceable in Fugu and Tetraodon. Phylogenetic tree of type II DnaJ gene shown that all the five fish models share DnaJ B gene with the exception that DnajB11 genes are not traceable in medaka (Fig. 4). From the analysis it can be suggested that DjB5 genes were independently evaluated. Closely related evolution was found between

tetraodon and fugu and in the same way between stickleback and medaka for all the type II variant genes. This tree suggests that the DjB1a of DnaJ gene in stickleback and medaka were evolve independently. Type III DnaJ genes show the close evolutionary relationship between tetraodon and fugu as well as between stickleback and medaka and isolated evolution in zebrafish.

From the analysis fugu and tetraodon have DnaJ genes which are showing closely related evolution in the same way DnaJ genes in stickleback and medaka are closely evoluated. In comparison with other four fish models DnaJ genes in zebrafish are distantly evoluated.

Gene distribution

The genes responsible for the synthesis of zebrafish DnaJ proteins are distributed in twenty two out of twenty five chromosomes and their locations vary in chromosome to chromosome (Fig. 5). In gene distribution analysis, chromosome 1 - have shown the presence of the respective genes for the production of all the three types of DnaJ proteins namely Type I, II and III. Type I and type II DnaJ protein coding genes are present in both chromosome 2 and 3 but the placement differ from each other whereas chromosome 21 is showing the presence of three genes among which two are coding for type III and one for type II. The maximum number of six DnaJ proteins (5 for Type III, 1 for type II) genes are present in chromosome 9. Chromosome 6, 8, 12 14, 17, 19, 20 and 24 are possessing the single gene coding for the Type III DnaJ protein. Chromosome 13, 23 and 25 are having 2 genes coding for two type III DnaJ protein. Chromosome 7 reveals the presence of two genes coding for Type I and type III DnaJ protein whereas chromosome 18 having the coding sequence of type I DnaJ protein. The second longest Chromosome 5 own only one gene responsible for the Type II DnaJ protein.

From the above results it can be suggested that the genome of Danio have not shown any duplication or tandem duplication of genes coding for DnaJ protein as indicated by the variation in the location of genes in each chromosome. Among the 25 chromosomes 3 chromosomes namely 11, 15 and 16 do not have any gene sequence for the synthesis of DnaJ proteins and 17 chromosomes have shown the coding region for mainly type III protein. This indicates the evolutionary significance of the conservation of type III DnaJ protein genes and in turn implies the functional importance of type III DnaJ protein in zebrafish.

Table 1: Analysis of DnaJ protein genes in zebrafish

Name	Ensembl Gene ID	Associated Gene Name	Chromosome Name	% GC content	Location
Dr_DjA1	ENSDARG00000030972	dnaja1l	1	34.85	Cytosol
Dr_DjA2a	ENSDARG00000023921	dnaja2	7	36.5	Cytosol
Dr_DjA2b	ENSDARG00000010745	dnaja2l	18	37.66	Cytosol
Dr_DjB1a	ENSDARG00000041394	dnajb1a	1	37.48	Cytosol
Dr_DjB1b	ENSDARG00000015831	dnajb1b	3	34.32	Cytosol
Dr_DjB4	ENSDARG00000038978	zgc:91922	2	32.82	Nucleus
Dr_DjB5a	ENSDARG00000052881	NP_001093510.1	5	36.71	Cytosol
Dr_DjB5b	ENSDARG00000004187	zgc:122979	21	39.08	Cytosol
Dr_DjB11	ENSDARG00000015088	dnajb11	9	32.49	Cytosol
Dr_DjB13	ENSDARG00000043157	dnajb13	10	38.32	Cytosol
Dr_DjC1	ENSDARG00000001940	zgc:152779	2	37.09	Cytosol
Dr_DjC2	ENSDARG00000070477	dnajc2	4	34.39	Nucleus
Dr_DjC3	ENSDARG00000041110	dnajc3	9	32.81	Nucleus
Dr_DjC4	ENSDARG00000024090	zgc:77513	14	36.8	Mitochondria
Dr_DjC5a2	ENSDARG00000042948	dnajc5aa	8	36.72	Cytosol
Dr_DjC5b	ENSDARG00000058147	dnajc5b	24	33.78	Cytosol
Dr_DjC5bL	ENSDARG00000041896	dnajc5g	20	35.03	Cytosol
Dr_DjC6	ENSDARG00000079891	dnajc6	6	33.98	Cytosol
Dr_DjC7a	ENSDARG00000058148	dnajc7	3	36.16	Cytosol
Dr_DjC8	ENSDARG00000059373	zgc:136281	19	49.8	Nucleus
Dr_DjC9	ENSDARG00000031293	zgc:92648	12	38.29	Cytosol
Dr_DjC10	ENSDARG00000074727	dnajc10	9	33.84	Cytosol
Dr_DjC11a	ENSDARG00000011196	dnajc11	23	35.22	Nucleus
Dr_DjC11b	ENSDARG00000088495	DNAJC11 (2 of 2)	Zv9_NA732	36.31	Nucleus
Dr_DjC12	ENSDARG00000086691	dnajc12	13	31.4	Cytosol
Dr_DjC15	ENSDARG00000038309	dnajc15	9	37.43	Cytosol
Dr_DjC17	ENSDARG00000026561	dnajc17	Zv9_scaffold3548	37.73	Cytosol
Dr_DjC18	ENSDARG00000056005	dnajc18	21	37.82	Cytosol
Dr_DjC19	ENSDARG00000044420	dnajc19	22	35.01	Mitochondria
Dr_DjC21	ENSDARG00000057080	dnajc21	21	32.37	Nucleus

Dr_DjC22	ENSDARG00000037067	dnajc22	23	36.07	Mitochondria
Dr_DjC24	ENSDARG00000023927	dnajc24	25	35.39	Extracellular membrane
Dr_DjC27	ENSDARG00000070916	dnajc27	17	34.38	Cytosol
Dr_DjC28	ENSDARG00000018181	dnajc28	9	39.6	Mitochondria
Dr_DjCb2	ENSDARG00000058644	zgc:153268	9	38	Nucleus
Dr_DjCb6b	ENSDARG00000020953	dnajb6b	7	34.43	Nucleus
Dr_DjCb6a	ENSDARG00000004680	dnajb6a	2	36.06	Nucleus
Dr_DjCb9	ENSDARG00000052072	zgc:114162	25	33	Cytosol
dnajb12	ENSDARG00000039363	Dr_DjCb12a	13	34.19	Cytosol
Dr_DjCb14	ENSDARG00000069996	zgc:153638	1	34.67	Cytosol

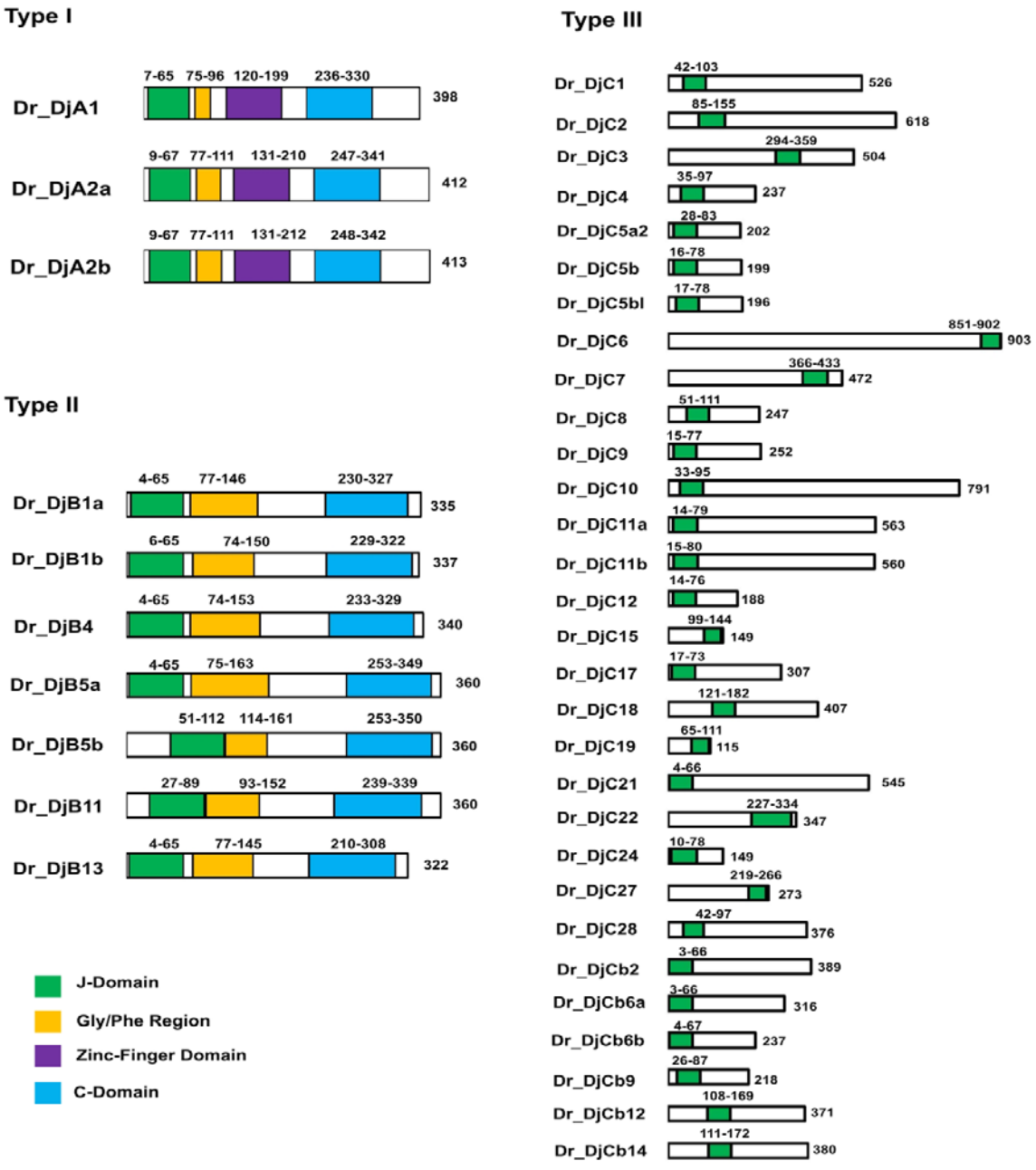

Fig. 1. Structural Classification of DnaJ proteins. DnaJ proteins are categorized according to the domains present in the protein. They are classified into three types. Type I contain 4 different domains namely J-domain, Gly / Phe region, Zinc-Finger Domain and C-Domain. Type II DnaJ protein lacks Zinc-Finger Domain. Type III DnaJ protein Contain only J-Domain.

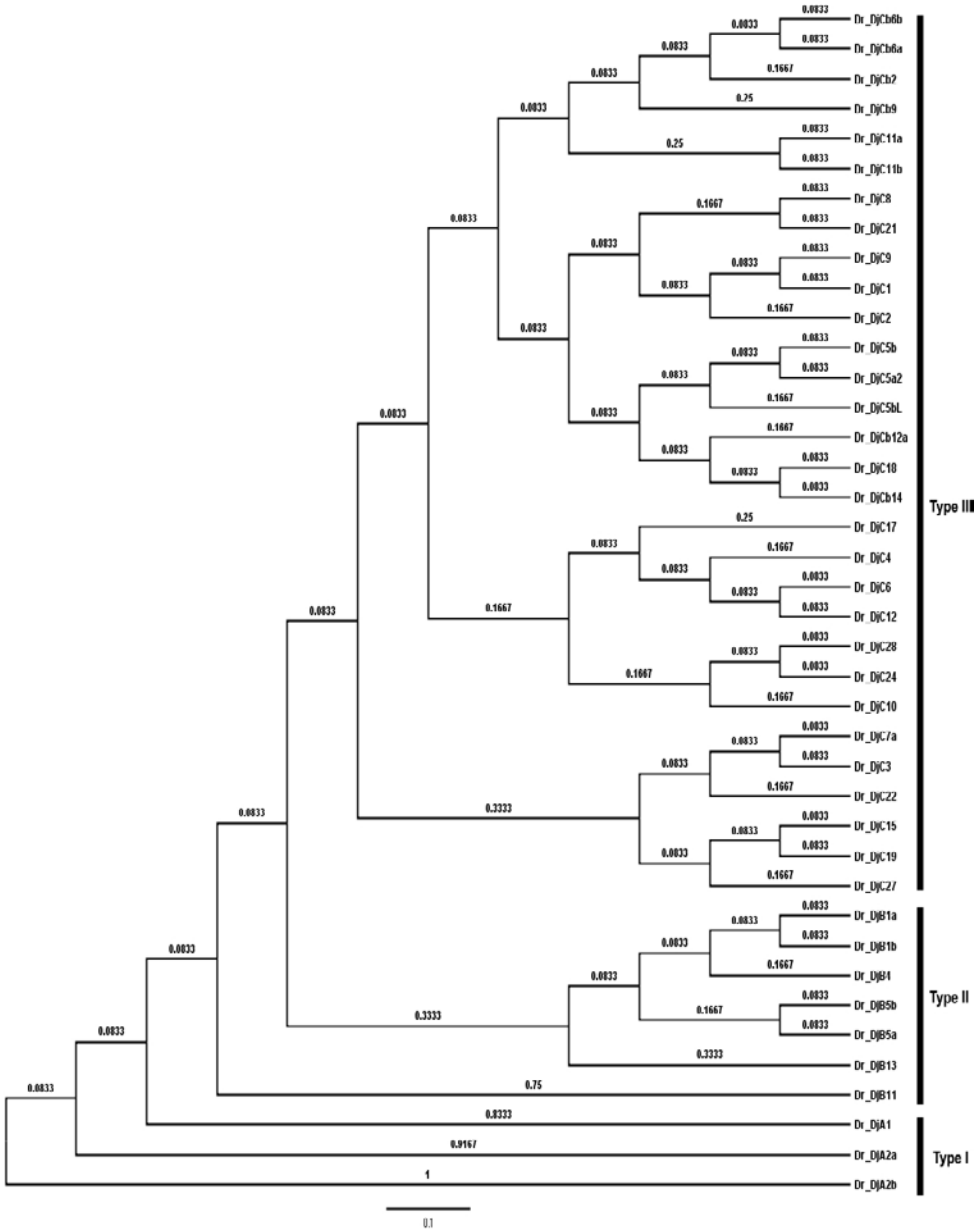

Fig. 2. Phylogenetic relationships of the 40 *Danio rario* DnaJ With 0.001 as the E-value cut-off using neighbor-joining method. The numbers above the branches are branch length based on the time scale. Taxa terminologies (*Danio rario*) are abbreviated using the first letter of genus and species name (Dr). Scale bar indicates the number of substitutions per site.

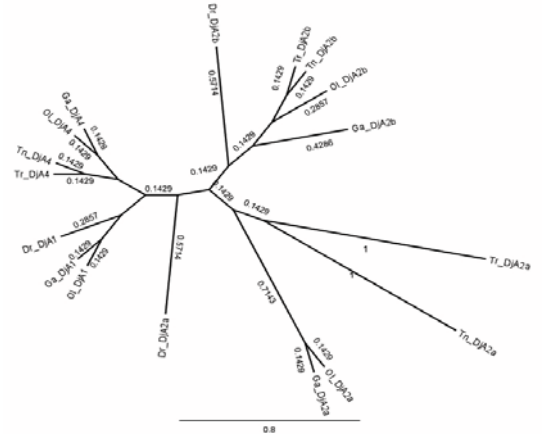

Fig. 3. Unrooted Neighbor-Joining tree of Type I DnaJ protein among the five fish models. The first letter denotes the genus name and the second represent the species name of the fishes. The numbers above the branches are branch length based on the time scale. Scale bar indicates number of substitutions per site.

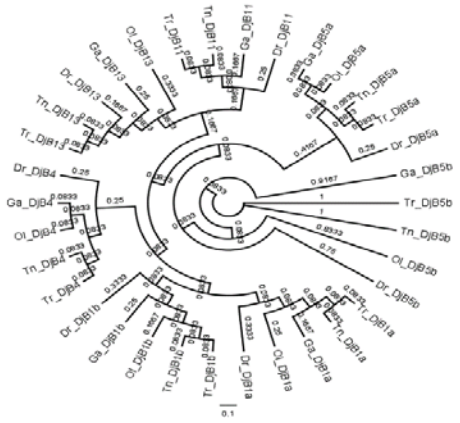

Fig. 4. Phylogenetic Neighbor-Joining tree of Type II DnaJ protein among the five fish models. The numbers above the branches are branch length based on the time scale. Scale bar indicates the number of substitutions per site.

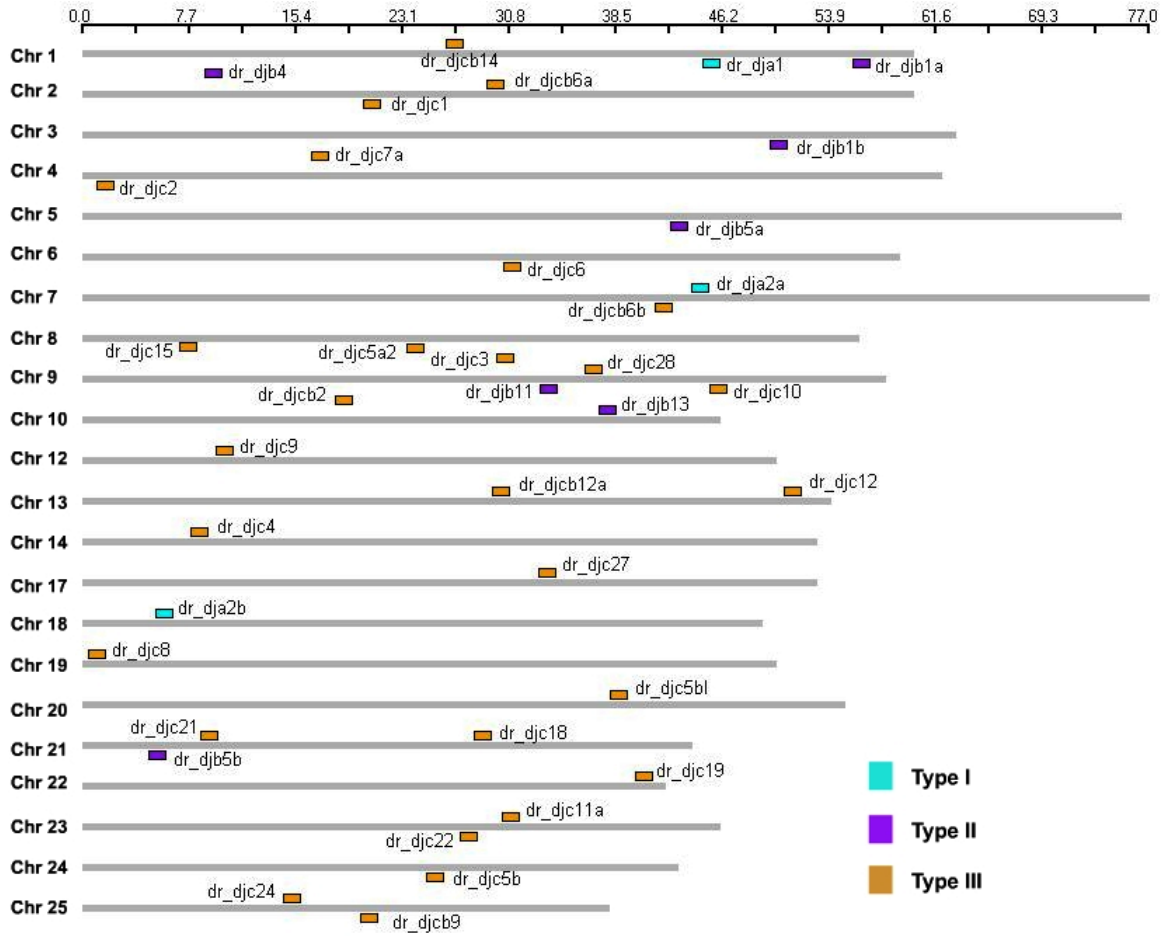

Fig. 5. Physical distributions of DnaJ Gene on *Danio rario* chromosomes. The boxes above and below the chromosomes (Chr represents chromosome) designate the approximate locations of the 40 *Danio rario* DnaJ genes. The color of the boxes indicates the DnaJ protein type.

CONCLUSION

From the computational analysis of five fish models, it can be concluded that Type III of DnaJ- protein are abundantly distributed when compared to type I and type II. The reason for this abundance is yet to be investigated thoroughly with respect to functional aspects. In silico localization analysis reveals that the DnaJ protein is distributed in wide range of cellular sites and cytosol is the major site of DnaJ protein localization. Phylogenetic analysis of DnaJ protein genes shows the close evolutionary relationship between fugu and tetraodon and medaka and stickleback whereas the distinct evolutionary pattern is observed for zebrafish. The genome of all the five fish models have not shown any duplication or tandem duplication of genes coding for DnaJ protein as indicated by the variation in the location of genes in each chromosome.

REFERENCES

Bailey T.L., Williams N., Misleh C., Wli W. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic acids res.*, 34: 369-373.

Bukau B., Horwich A.L. 1998. The Hsp70 and Hsp60 chaperone machines. *Cell*, 92: 351–366.

Bukau B., Weissman J., Horwich A. 2006. Molecular chaperones and protein quality control. *Cell*, 125: 443–451.

Cheetham M.E., Caplan A.J. 1998. Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress Chaperones*, 3:28–36.

Chou K.C, Shen H.B. 2010. A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS One*, 5:e9931.

Craig E.A., Huang P., Aron R., Andrew A. 2006. The diverse roles of J-proteins, the obligate Hsp70 co-chaperone. *Rev. Physiol. Biochem. Pharmacol.*, 156:1–21.

Georgopoulos C., Welch W.J. 1993. Role of the major heat shock proteins as molecular chaperones. *Annu. Rev. Cell Biol.*, 9: 601–634.

Hall T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symp Ser.*, 41:95–98.

Horton P., Park K.J., Obayashi T., Fujita N., Harada H., Adams-Collier C.J., Nakai K. 2007. WoLF PSORT:

protein localization predictor. *Nucleic Acids Res.*, 35: 585-587.

Jermy A.J., Willer M., Davis E., Wilkinson B.M., Stirling C.J. 2006. The Brl domain in Sec63p is required for assembly of functional endoplasmic reticulum translocons. *J. Biol. Chem*., 281: 7899–7906.

Kozik A., Kochetkova E., Michelmore R. 2002. GenomePixelizer- a visualization program for comparative genomics within and between species. *Bioinformatics*, 18:335-336.

Letunic I., Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.*, 44(W1): W242-245

Lu Z., Cyr D.M. 1998. The conserved carboxyl terminus and zinc finger-like domain of the co-chaperone Ydj1 assist Hsp70 in protein folding. *J. Biol. Chem*., 273: 5970–5978.

McWilliam H, Li W., Uludag M., Squizzato S., Park Y.M., Buso N., Cowley A.P., Lopez R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic acids res.*, 41(W): W597-600.

Ohtsuka K., Hata M. 2000. Mammalian HSP40/DNAJ homologs: cloning of novel cDNAs and a proposal for their classification and nomenclature. *Cell Stress Chaperones*, 5: 98–112.

Pellecchia M., Szyperski T., Wall D., Georgopoulos C., Wuthrich K. 1996. NMR structure of the J-domain and the Gly/Phe-rich region of the *Escherichia coli* DnaJ chaperone. *J. Mol. Biol*., 260: 236–250.

Qiu X.B., Shao Y.M., Miao S., Wang L. 2006. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell Mol. Life Sci*., 63: 2560–2570.

Ronquist F., Huelsenbeck J.P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19: 1572-1574.

Takayama S., Xie Z., Reed J.C. 1999. An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. *J. Biol. Chem*., 274:781–786.

Tsai J., Douglas M.G. 1996. A conserved HPD sequence of the J-domain is necessary for YDJ1 stimulation of Hsp70 ATPase activity at a site distinct from substrate binding. *J. Biol. Chem*., 271: 9347–9354.

Wu Y., Li J., Jin Z., Fu Z., Sha B. 2005. The crystal structure of the C terminal fragment of yeast Hsp40 Ydj1 reveals novel dimerization motif for Hsp40. *J. Mol. Biol*., 346: 1005–1011.

Yamamoto M., Maruyama D., Endo T., Nishikawa S. 2008. *Arabidopsis thaliana* has a set of J proteins in the endoplasmic reticulum that are conserved from yeast to animals and plants. *Plant Cell Physiol*., 49: 1547–1562.

Zhang L., Pond S.K., Gaut B.S. 2001. A survey of the molecular evolutionary dynamics of twenty-five multigene families from four grass taxa. *J. Mol. Evol*., 52:144–156.