

Research Article

Repertoire of Protein Kinases Encoded in the Genome of *Takifugu rubripes*

**R. Rakshambikai,¹ S. Yamunadevi,^{1,2,3} K. Anamika,^{1,4}
N. Tyagi,^{1,5} and N. Srinivasan¹**

¹Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

²Max Planck Institute for Intelligent Systems (Formerly Max Planck Institute for Metals Research), BioQuant BQ0038, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

³German Cancer Research Center, BioQuant BQ0038, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

⁴Department of Functional Genomics and Cancer and Department of Structural Biology and Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 964, Université de Strasbourg, 1 rue Laurent Fries, 67404 Illkirch Cedex, France

⁵European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD Cambridge, UK

Correspondence should be addressed to N. Srinivasan, ns@mbu.iisc.ernet.in

Received 27 November 2011; Revised 14 February 2012; Accepted 28 February 2012

Academic Editor: G. Pesole

Copyright © 2012 R. Rakshambikai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Takifugu rubripes is teleost fish widely used in comparative genomics to understand the human system better due to its similarities both in number of genes and structure of genes. In this work we survey the fugu genome, and, using sensitive computational approaches, we identify the repertoire of putative protein kinases and classify them into groups and subfamilies. The fugu genome encodes 519 protein kinase-like sequences and this number of putative protein kinases is comparable closely to that of human. However, in spite of its similarities to human kinases at the group level, there are differences at the subfamily level as noted in the case of KIS and DYRK subfamilies which contribute to differences which are specific to the adaptation of the organism. Also, certain unique domain combination of galectin domain and YkA domain suggests alternate mechanisms for immune response and binding to lipoproteins. Lastly, an overall similarity with the MAPK pathway of humans suggests its importance to understand signaling mechanisms in humans. Overall the fugu serves as a good model organism to understand roles of human kinases as far as kinases such as LRRK and IRAK and their associated pathways are concerned.

1. Introduction

Takifugu rubripes is a teleost fish native to northwest pacific seas. It belongs to the family Tetraodontidae and order Tetraodontiformes. The fugu genome is rather compact with a size of ~400 Mb although the number of genes is comparable to that of higher eukaryotes indicating a considerable reduction in the intergenic regions [1]. Detailed analysis reveals that the intron-exon boundaries [2] and in certain cases alternative splicing [3], synteny [4, 5] have been conserved with respect to that of humans suggesting the possibility of conserved elements from a common ancestor. Thus, due to such features indicating close relationship, the fugu is suggested to be

a good model organism and an effective way to study evolution of structure of complex vertebrate genomes [4, 6].

In 2002, the first draft sequence of the fugu genome was reported by the International fugu genome Consortium using the “whole-genome shotgun” strategy. Subsequently many versions of the genomic data have been made available at <http://www.fugu-sg.org/>. The latest version (v-5) was released in 2010 which covers about 392 Mb and 72% of genome being organized into chromosomes.

Response to environmental stimulus via complex signaling systems is a central feature of all living cells. Phosphorylation is one such posttranslational modification employed in signaling circuits which usually results in a functional

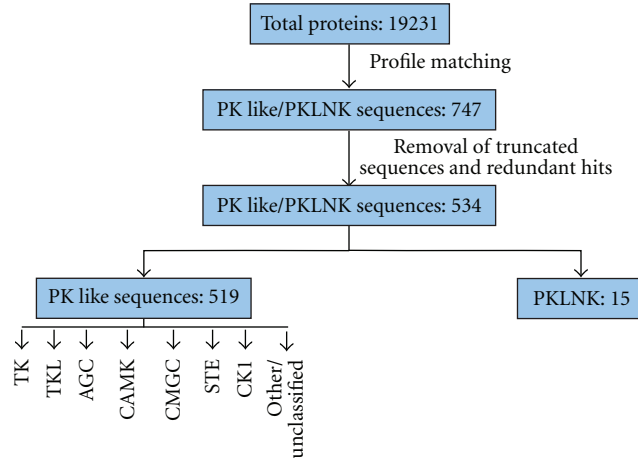


FIGURE 1: Distribution of fugu kinases into various Hanks and Hunter groups. PKLNK: protein-kinase-like nonkinases; CAMK: calcium/calmodulin-dependent protein kinase; CMGC: the group of cyclin-dependent protein kinase, mitogen-activated protein kinase, glycogen synthase kinase, casein kinase-2; AGC: the group of protein kinase A, protein kinase G, protein kinase C; STE: sterile (homologs of yeast STE); TK: tyrosine kinase; TKL: tyrosine kinase like; CK1: casein kinase 1.

change in the substrate by changing enzyme activity, cellular location, or association with other proteins. Thus, protein kinases have implication in regulation of various cellular processes encompassing metabolism, stress responses, cell cycle control, organ development, and intercellular communication [28, 29]. Abnormalities in the functioning of these kinases usually have implications in developmental disorders and malignancies [30, 31]. The eukaryotic kinases mainly constitute the Ser/Thr and Tyr kinases which share a common three-dimensional fold and the catalytic core spanning to about 300 residues [32].

The fugu genome has been used in a myriad of comparative genomic studies to elucidate the function of proteins involved in neurodegenerative diseases [33, 34], signaling systems [35, 36], and so forth and has been suggested as a method to elucidate cognate pathways in humans. In this paper, using sensitive sequence analysis [37–54] we recognize the repertoire of Ser/Thr and Tyr kinases encoded in the fugu genome. This is not trivial as homologous kinases are known to be characterized by weak sequence similarity. In addition, we classify these kinases on the basis of their catalytic domain sequence and the domains covalently tethered to the catalytic kinase domain [38]. Finally, we provide comparative analysis with distribution of the kinases in other model organisms and other proteins of the MAPK pathway especially in relation to the higher eukaryotes like human.

2. Materials and Methods

2.1. Identification of Protein Kinases. The complete set of predicted proteins from the ORF's of the *Takifugu rubripes* fifth assembly genome has been obtained from <http://www.fugu-sg.org/>. We have adopted sensitive sequence profile matching algorithms to identify and examine Ser/Thr and Tyr kinases encoded in the genome. The protocol used is identical to that adopted for analysis of kinases of other organisms earlier in this laboratory [43–49]. Briefly, we have employed

multiple sensitive sequence search and analysis methods PSI-BLAST [37], MulPSSM [39–41] involving extensive use of RPS-BLAST [41] and HMMer [42] which match Hidden Markov Models (HMMs) to identify protein kinase catalytic domain and their co-occurring domains. The criteria used to associate a given protein kinase to a given subfamily on the basis of its primary structure include the degree of sequence identity greater than 30% with members of known subfamily of kinases and the presence of signature amino acids that are characteristics of protein kinase subfamilies [43] which include the glycine rich loop and catalytic aspartate of consensus sequence HRDLKXXN. In addition, search procedures such as PSI-BLAST have been used to detect sequences homologous to the kinase catalytic domain using an *E*-value cutoff of 0.0001 which is decided on the basis of previous prototypic study [50]. Truncated sequences which are less than 200 amino acids long were eliminated to arrive at a set of 534 PK- (protein-kinase-) like sequences. The data set of putative PK-like sequences has been obtained from the compilation of hits obtained during various search procedures. Out of these, 15 sequences lack aspartate in the catalytic loop and, therefore, are unlikely to function as kinases. These are referred as protein-kinase-like non-kinases (PKLNKs) [55]. These sequences were subjected to fold recognition approach PHYRE (<http://www.sbg.bio.ic.ac.uk/phyre/>) [51, 52] to ensure that they fold like kinases. The final number of 519 sequences is likely to function as protein kinases. The entire operation works stepwise with filtering of sequences at every stage in order to recognize kinases. The number of sequences involved at various stages is depicted in Figure 1.

2.2. Classification of Kinases into Hanks and Hunter Groups. Hanks and Hunter have proposed classification of kinases based on sequence analysis [38]. In order to classify the fugu kinases into these groups and subfamilies reverse PSI-BLAST (RPS-BLAST) was used to search each of the 519 PK-like sequences as a query against the database containing 2810

position-specific scoring matrices (PSSMs) created for the various subgroups of protein kinases corresponding to sub-families of kinases. A query kinase sequence was associated to its subfamily based on the extent of sequence similarity. Sequences with greater than 30% identity and 70% profile coverage with at least one of the members of a kinase groups have been considered as members of the group or subfamily concerned. CLUSTALW [53] was used to generate multiple sequence alignment for the 519 kinases that were associated to specific groups. MEGA version 4 [54] was used to generate the dendrogram showing various groups of protein kinases. The sequences belonging to the group “Others” have been clustered using another dendrogram. Also MEGA 4 [54] has been used to cluster sequences of the kinase domain regions of certain families from other organisms including human, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae*.

2.3. Assignment of Domains to Multidomain Kinases. Domain assignments have been made for protein kinase catalytic domain containing gene products using the HMMer method by querying each of the kinase domain containing proteins against the protein family HMMs available in the Pfam database [56] and MULPSSM profiles [40, 41] of families in Pfam database. Transmembrane segments were detected using TMHMM [57].

2.4. Identification of MAPK Pathway Proteins. Sequences of MAPK proteins involved in the MAPK pathways in human were obtained from KEGG pathway database <http://www.genome.jp/kegg/pathway.html> [58]. The sequences were then used as a query to search using BLAST against the predicted protein sequences of fugu genome with an E-value cut off of 0.0001. The hits were obtained after pruning on the basis of coverage and percentage identity. The proteins which did not identify any homolog were then queried with an integrated dataset of fugu genome and SWISSPROT using PSI-BLAST [37] with an E-value cutoff of 0.001. The pathway was then generated as a network using CYTOSCAPE2.6.3 [59].

3. Results

The genome of fugu encodes 534 PK-like sequences. Of the 534 PK-like sequences 519 them possess the critical aspartate residue at the location characteristic of catalytic base and have at least one glycine conserved in the “glycine rich” loop GXGXXG present in the subdomain I of the kinase catalytic domain. A list of all these 519 sequences identified in this work along with classification and domain combinations are deposited in the KinG database [43] which was developed in this laboratory and the information is publicly available at <http://king.mbu.iisc.ernet.in/>. In addition, these 519 kinases are listed in the supplementary data file 1 in Supplementary Material available online at doi. 10.1155/2012/258284. The 15 sequences lacking the critical aspartate are unlikely to be functional as kinases though they are likely to adopt the kinase fold. Though the exact roles of these PKLNK's

(protein kinase like non kinases) is not entirely clear, such sequences have been reported to have implications in various signaling pathways [60]. Out of the 519 putative functional kinases, 407 of them have an arginine residue preceding the critical aspartate and these are called as the “RD” kinases [48, 61]. This arginine has been indicated to be involved in an interaction with phosphate group of a phosphorylated sidechain in the activation loop of kinases. This interaction is known to be a critical step in switching a kinase from inactive to active state through conformational changes. Hence the switching mechanism of these RD kinases is likely to be mediated by phosphorylation at the activation segment in the kinase catalytic domain.

Among the putative PKs, 135 are likely to be tyrosine kinases and 296 are likely to be Ser/Thr kinases. 17 of the 296 Ser/Thr kinases are predicted to have membrane spanning regions. Similarly, 50 of the tyrosine kinases are predicted to be receptor tyrosine kinases. List of these kinases including the information on predicted transmembrane region is included in the supplementary data file 1.

3.1. Groupwise Distribution of Protein Kinases. The 519 kinases have been classified on the basis of the Hanks and Hunter [38] scheme of classifying protein kinases into 7 groups with clearly defined functional roles, namely, AGC (regulated by binding of second messengers), STE (proteins featuring in the MAPK signaling cascades in yeast), CMGC (include MAPK, CDK proteins), CAMK (Calcium/Calmodulin regulated kinases), CK1 (Casein kinase 1), TK (tyrosine kinase), and TKL (tyrosine kinase like). Apart from the 7 groups enlisted, there is yet another group which comprises of sequences that cannot be classified into any of the standard groups and are termed as “Other/Unclassified.” The group-wise distribution of kinases in fugu has been shown in Figure 2 with the TK group being most prevalent and CK1 the least prevalent. Clustering of the sequences of the catalytic kinase domain using BLAST-CLUST was performed and 27 sequences are identified as outliers to the group/subfamily concerned. A dendrogram was constructed without these 27 sequences or the TK group and it resulted in clear distinct grouping of kinases (Figure 3). Though the members belonging to the “Other/Unclassified” group are significantly different from the classical groups, clusters are observed indicating high similarity among the sequences within a node (Figure 4). This points to the possible emergence of newer subfamilies of protein kinases that do not conform to the known groups [38] in the classification of kinases. However it is also possible that some of these “new subfamilies” represent outliers of currently known subfamilies. The group-wise distribution (percentage number of sequences in each group) has been compared with that of other model eukaryotic organisms such as human, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae* by performing similar analysis (Figure 5 and Table 1). The distribution of kinases in fugu is very similar to that of other eukaryotes considered and in most cases very similar to that of humans at the group level.

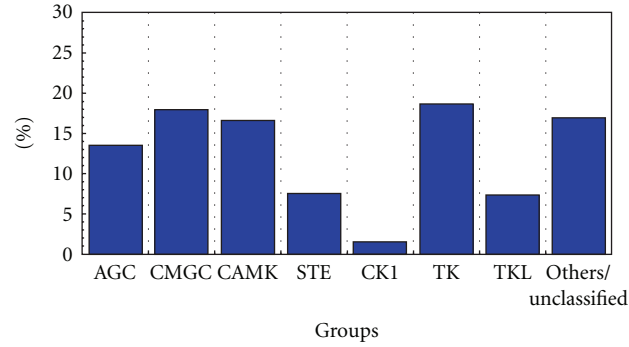


FIGURE 2: Percentage-wise distribution of fugu kinase groups. CAMK: calcium/calmodulin dependent protein kinase; CMGC: a group of cyclin dependent protein kinase, mitogen activated protein kinase, glycogen synthase kinase and casein kinase-2; AGC: a group of protein kinase A, protein kinase G, and protein kinase C; STE: sterile (homologs of yeast STE); TK: tyrosine kinase; TKL: tyrosine kinase like; CK1: casein kinase 1.

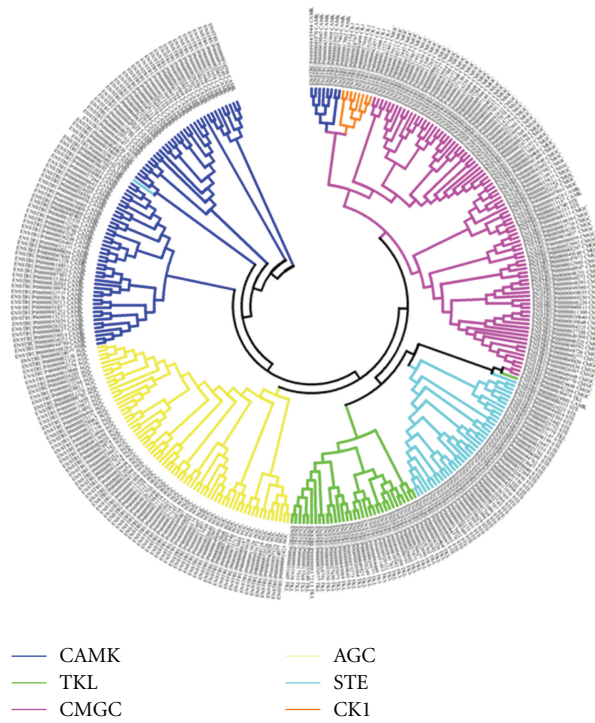


FIGURE 3: Dendrogram depicting clustering of fugu kinases. Abbreviations followed in the diagram are same as listed in the legend of Figure 2.

TABLE 1: Distribution of kinases of fugu, yeast, human, Drosophila, C. elegans into various Hanks and Hunter groups. Numbers indicated as percentage.

Groups	Organism				
	Fugu	Yeast	Human	<i>D. melanogaster</i>	<i>C. elegans</i>
AGC	13.48	15.23	11.72	17.35	11.04
CMGC	17.91	20.95	19.82	16.43	15.82
CAMK	16.57	22.85	16.72	15.51	12.53
STE	7.51	13.33	8.9	8.21	6.56
CK1	1.54	3.8	1.8	5.45	19.4
TK	18.68	0	20.34	15.52	20.59
TKL	7.32	0	7.2	6.39	3.28
Others/Unclassified	16.95	26.66	13.27	15.06	10.74

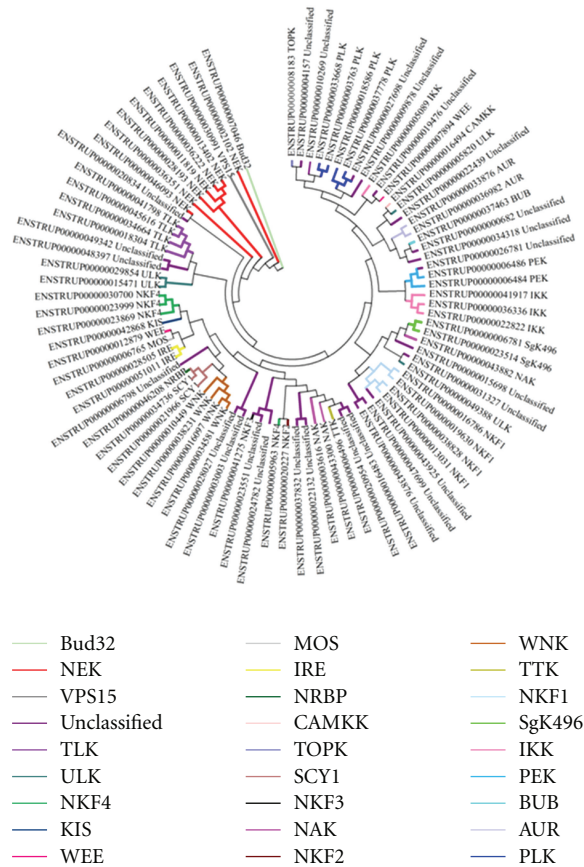


FIGURE 4: Dendrogram depicting clustering of kinases belonging to the “Other/Unclassified” group.

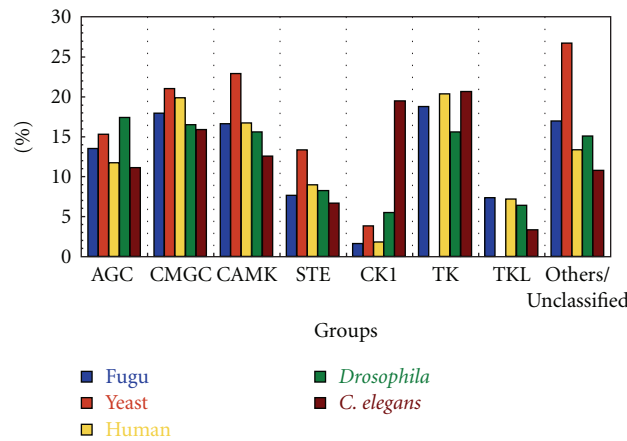


FIGURE 5: Comparison of group wise distribution of kinases from fugu, human, yeast, *Drosophila* and *C. elegans*. Abbreviations followed in the figure are same as given in the legend to Figure 2.

Though the group-wise distribution is quite comparable we wanted to explore at the level of sub-families to see if the same trend is observed. Interestingly, there are certain subfamilies in which the representation by fugu kinases is noticeably higher or lower than those of other model organisms considered. The percentage-wise distribution of these kinase subfamilies has been indicated in Figures 6 and 7.

Figure 6 indicates distribution of kinases for 15 subfamilies in which fugu kinases occur in higher frequency with DYRK being highest and those of tyrosine kinase group generally being highly represented. Likewise Figure 7 shows distribution of kinases for 7 subfamilies in which fugu kinases occur at a lower frequency with CAMKK and KIS subfamilies being least. The functions of each of the

TABLE 2: Kinase sub-families with abnormal distribution in fugu. Overall function of each of the family is also indicated.

Family	Function
<i>Highly represented families</i>	
JakA	Receptor tyrosine kinase. Activates STAT involved in interferon signaling [7]
Lmr	Receptor tyrosine kinase involved in apoptosis [8]
PDGFR	Receptor tyrosine kinase activating factors for growth, differentiation, development [7]
Eph	Receptor tyrosine kinase component of developmental pathways [9]
InsR	Receptor that binds insulin and has a tyrosine-protein kinase activity [10]
PIM	Phosphorylating chromatin proteins and controlling transcription [11]
CDKL	Mediates phosphorylation of MECP2 [12]
Trio	Guanine nucleotide exchange factors that mediate cell invasiveness [13]
DYRK	Directs cellular response to stress conditions and also implicated in neuropathological characteristics of Down's syndrome [14, 15]
Trk	Receptor for neurotrophin-3 (NT-3) [16]
Src	Nonreceptor protein tyrosine kinase that plays pivotal roles in numerous cellular processes such as proliferation, migration, and transformation [17]
VEGFR	The VEGF-kinase ligand/receptor signaling system plays a key role in vascular development and regulation of vascular permeability [18]
Sgk496	Induces both caspase-dependent apoptosis and caspase-independent cell death [19]
Trbl	Interacts with MAPK kinases and regulates activation of MAP kinases [20]
Wnk	Controls sodium and chloride ion transport [21]
<i>Underrepresented families</i>	
KIS	Function unknown
MLCK	Calcium/calmodulin-dependent enzyme implicated in smooth muscle contraction via phosphorylation of myosin light chains [22]
MLK	Involved in the JNK pathway [23]
MAPKAPK	Integrative element of signaling in both mitogen and stress responses [24]
CDK	Involved in regulation of cell cycle by binding to cyclins [25]
CAMKK	Calcium/calmodulin-dependent protein kinase that belongs to a proposed calcium-triggered signaling cascade involved in a number of cellular processes [26]
Ste7	MAP2K homologous to yeast Ste7 [27]

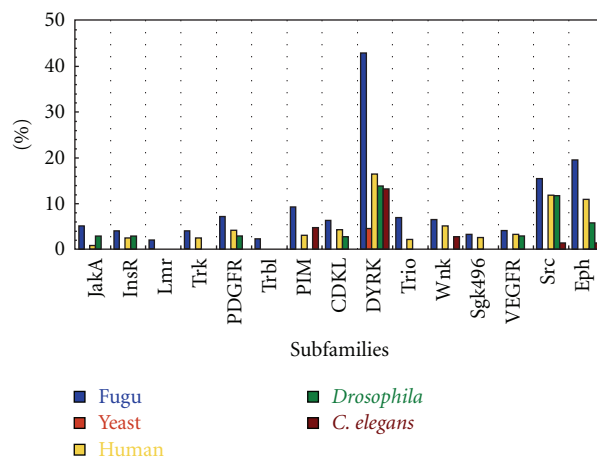


FIGURE 6: Percentage distribution of subfamilies that are overrepresented in fugu in comparison to various model organisms. JakA: janus kinase A; InsR: insulin receptor, Lmr: lemur kinase; Trk: neurotrophic tyrosine kinase receptor type 1; PDGFR: platelet derived growth factor receptor; Eph: ephrin receptor; VEGFR: vascular endothelial growth factor receptor; Trbl: tribbles; CDKL: cyclin-dependent kinase like; DYRK: dual specificity tyrosine regulated kinase, Wnk: with no lysine (K) kinases.

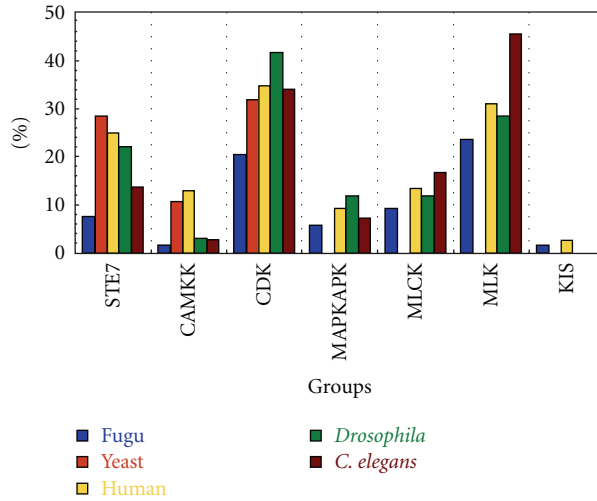


FIGURE 7: Percentage distribution of subfamilies that are under-represented in various model organisms. MLK: mixed lineage kinase; MAPKAPK: mitogen- activated protein kinase- activated protein kinase; MLCK: myosin light chain kinase; CDK: cyclin dependent kinase; CAMKK: calmodulin dependent protein kinase kinase; KIS: kinase interacting with stathmin; Ste7: sterile7.

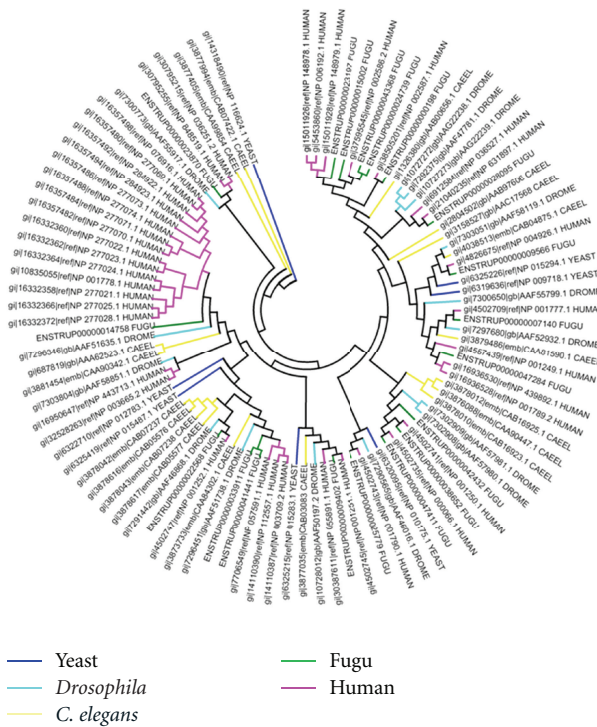


FIGURE 8: Dendrogram depicting clustering of kinases of CDK sub-family from fugu, human, *Drosophila*, and *C. elegans*.

proteins have been indicated in Table 2. Mutations or change in expression levels of these proteins has been implicated in various cancers (see the publications cited in Table 2). Also, profound differences in the number of paralogous proteins in two organisms could result in diverse outputs resulting in distinct features in the molecular processes in the two organisms [62]. The percentage distribution of each of the subfamilies for all the organisms considered can be obtained from Supplementary data file 2.

We have analyzed the nature of clustering of these fugu kinases with respect to those in other organisms. For most sub-families, fugu kinases group quite closely to those of human and in a few cases with other organisms. A similar trend is observed in the underrepresented sub-families with different clusters being observed that are not organism specific. The example of CDK subfamily is depicted in Figure 8. In each of the clusters human kinases dominate, with few representatives from fugu. Since paralogous proteins

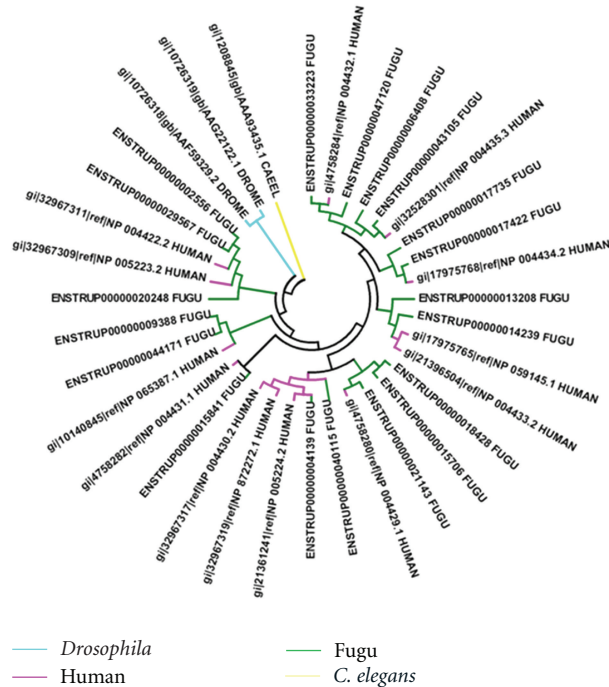


FIGURE 9: Dendrogram depicting clustering of kinases of ephrin receptor subfamily from fugu, human, *Drosophila*, and *C. elegans*.

perform diverse functions, their presence in large numbers is indicative of high functional diversity [62]. Therefore, it appears that paralogous CDKs in humans show larger functional diversity than fugu CDKs which are fewer in number. However, in the highly represented subfamilies of fugu kinases, apart from the above trend few kinases cluster separately suggesting higher divergence in a subset of DYRK kinases. Figure 9 shows the example of ephrin receptor family. In case of the DYRK family, 30 out of 40 kinases cluster separately (Figure 10). We then explored if this is a fish-specific trait. Fortunately another fish (zebrafish) genome has been sequenced and the list of zebrafish kinases are available in the KinG database [43]. However, the zebrafish has a large kinome (~900 kinases). Analysis on these shows that in most cases the fugu kinases do not cluster closely with those of zebrafish (supplementary data file 3). We performed a BLAST [37] search against the SWISSPROT database to identify the nearest homologs for these kinases in other organisms irrespective of the genome being sequenced fully or not. It was observed that, in the sequences which cluster close to human kinases, the percentage identity of those sequences with that of human counterpart is greater than 60%. Nevertheless, the kinases belonging to the fugu-specific cluster indicate low sequence identity with that of human and relatively higher identity of about 40% with that of certain lower eukaryotes like *Xenopus tropicalis*, *Dictyostelium discoideum*, and fungal species (supplementary data file 4). These may be fugu-specific sequences which are functionally divergent compared to those paralogs which show significant similarity to the human sequences. Similar expansions have been seen in other organisms as well [63] that help the organism to adapt to its environment.

3.2. Domain Combinations. The domain combinations for all the predicted kinases of fugu are provided in supplementary data file 1. Most of the domain combinations observed have been observed in various higher eukaryotes according to the Pfam database [56]. However there are 2 cases with an unusual domain combination as depicted in Figure 11.

In the first case, a galectin binding domain has been associated with protein kinase domain. The galectin domain is a carbohydrate binding domain and its function has been attributed to regulation of immunity and inflammatory responses, progression of cancer, and in specific developmental cascades [64]. These domains may function within or outside the cell. In this particular context, since there is no transmembrane component or domains which localize it to the membrane, the protein is likely to be cytosolic. This is further corroborated by the lack of any signal peptide motifs in this sequence which was analyzed using SignalP server [65]. The protein kinase domain may play a regulatory role wherein the activation state of the kinase might dictate the binding abilities of the galectin binding domain.

The second case involves a YkyA domain tethered to the protein kinase, CNH, PBD, PH, DMPK_Coil, C1.1, and the M protein repeats. Such a combination has been predicted also in another closely related fresh water species of Puffer fish, *Tetraodon nigroviridis*, however, without the YkyA domain. The YkyA domain has been reported only in bacterial species and is a putative lipoprotein binding domain occurring as a single-domain protein which aids in virulence [66]. It is likely that this protein is localized to the membrane due to the presence of the PH domain which is a reasonable indicator for membrane localization [67]. However, the role of YkyA in fugu is unclear.

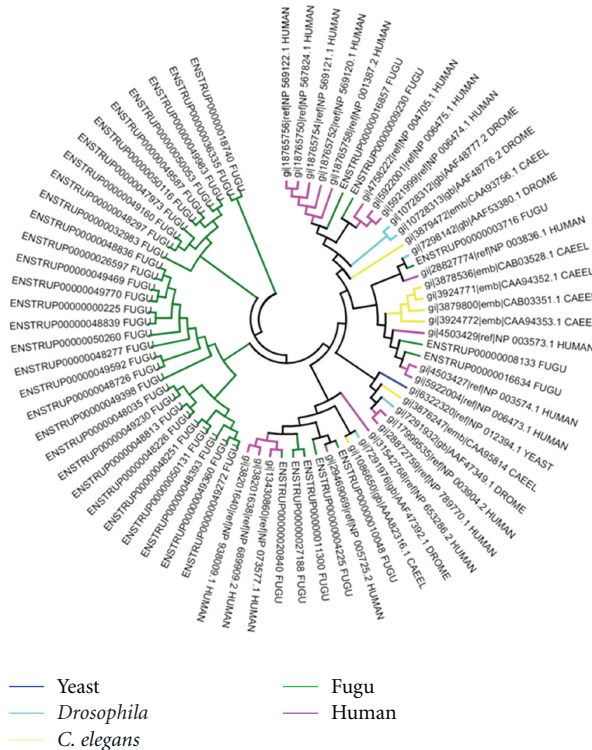


FIGURE 10: Dendrogram depicting clustering of kinases of dual specificity tyrosine regulated kinases (DYRKs) subfamily from fugu, human, yeast, *Drosophila*, and *C. elegans*.

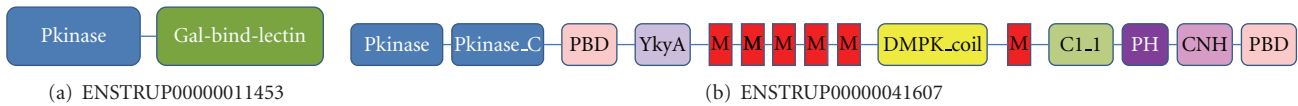


FIGURE 11: Unusual domain combinations. (a) Protein kinase tethered to a galectin binding domain. (b) Yka_A domain associated with protein kinase domain long with PBD, YkyA, M repeats, DMPK_coil, C1.1, PH, CNH. PBD: P21-Rho-binding domain; C1.1: phorbol esters/diacylglycerol binding domain; PH: pleckstrin homology.

Eight fugu kinase sequences are predicted to have two protein kinase domains each containing all the prerequisites for a functional kinase. This feature is distinct from that of Janus kinase which comprises a nonfunctional kinase domain (lacking critical aspartate) apart from a functional kinase domain. Five of these 8 cases correspond to the AGC group while the rest belong to the CAMK group. Such twin kinases have been reported in other higher eukaryotes in proteins like MAPKAP-K1 [68] wherein the C-terminal kinase domain is regulatory in nature and is involved in the activation of the N-terminal kinase domain. In all these 8 cases, both the kinase domains within a protein belong to the same group and sub-family which may be indicative of a duplication event and it may act by fine-tuning the activity levels of the protein. Interestingly, in most cases the C-terminal protein kinase domain has lower sequence identity to the respective groups than the N-terminal domain.

3.3. MAPK Signaling Pathway in Fugu. Given the slight unusual distribution of fugu kinases in terms of subfamilies we wanted to investigate the overall effect on a signaling

pathway by considering distribution of all the proteins involved (including nonkinases) in the pathway. Although we did not observe any significant skewing in the clustering of the MAPK subfamily we chose to work on the MAPK pathway as the proteins involved in this pathway are extremely well characterized in other eukaryotes, especially for humans. Both kinases and non-kinases (including upstream factors and downstream effectors) were considered for the analyses. The extent of sequence similarity is an approximate indicator of the similarities of the functions of human and fugu proteins involved in MAPK pathway. Results are depicted pictorially in Figure 12. The cases with high sequence identity (>30%) and coverage (>70%) (compared to human proteins) have been represented as green boxes. These indicate the presence of functional counterparts, which are present in approximately 90% of the cases. Interestingly in a few cases (represented as yellow boxes) the sequence similarity levels reveal distant homologs suggesting differences in molecular events. These include proteins which act as ligands (FAS, TNF) and receptors (those of IL1, TNF, and certain lipopolysaccharides) which are mainly

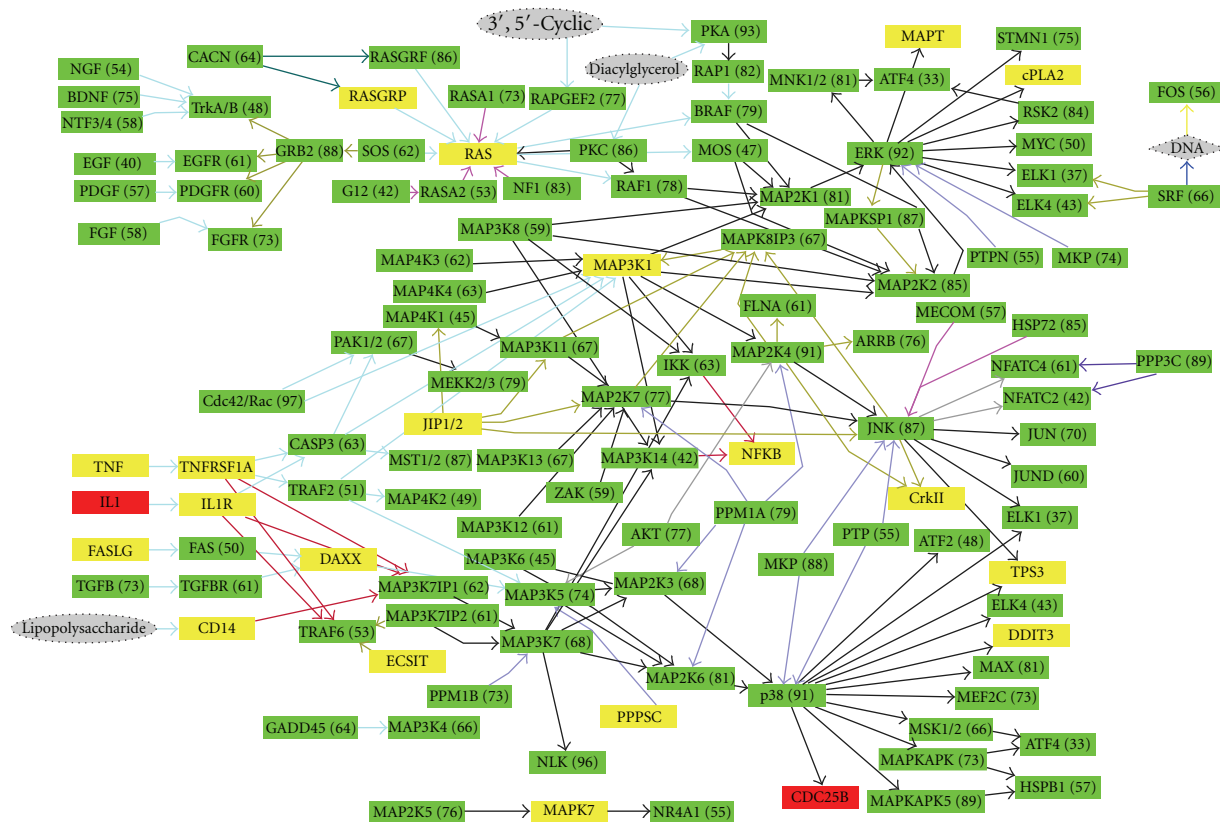


FIGURE 12: Human-fugu homologs MAPK pathway. Boxes in green represent closely related homologs of human MAPK pathway proteins present in fugu; yellow boxes indicate remote homologs and red boxes indicate potential absence of homologs in fugu.

implicated in signaling pathways of immune system along with few proteins of the three tier MAPK (MAPKKK, MAPKK, MAPK) cascade. It also includes Ras and its activating protein RasGrp. This is especially interesting because Ras is a key component involved in GTP exchange which is a crucial step in activating the MAPK cascades. Even more glaringly, two proteins (IL1, CDC25b) do not have any identified homologs in fugu. Potential absence of IL1 is especially interesting because its absence indicates alternate ligands which are able to bind to its receptor. The other protein without any detectable homolog is CDC25b, which is a phosphatase involved in activating CDK. Absence of CDC25b may be viewed in the light of the fact that CDK subfamily is underrepresented in the fugu genome.

4. Conclusions

The current analysis on the fugu genome indicates a kinase repertoire of ~3% of the total genome which is slightly higher than an average of around 2% in the other model organisms. All the groups of eukaryotic protein kinases are found to be present in this genome with comparable numbers to that of humans. The observed distribution of few kinase subfamilies is fugu specific. The presence of unique domain combinations gives an insight into possibility of new regulatory functions of kinases. Finally, the similarity to signaling pathways in human may not only provide

a platform for studying signaling systems but can also be used in kinase drug screening.

Acknowledgments

The authors are grateful to the anonymous referee for excellent suggestions in improving the presentation of this paper. This research is supported by Department of Biotechnology, Government of India.

References

- [1] S. Aparicio, J. Chapman, E. Stupka et al., "Whole-genome shotgun assembly and analysis of the genome of fugu rubripes," *Science*, vol. 297, no. 5585, pp. 1301–1310, 2002.
- [2] P. Gilligan, S. Brenner, and B. Venkatesh, "Fugu and human sequence comparison identifies novel human genes and conserved non-coding sequences," *Gene*, vol. 294, no. 1-2, pp. 35–44, 2002.
- [3] O. Coutelle, G. Nyakatura, S. Taudien et al., "The neural cell adhesion molecule II: genomic organisation and differential splicing is conserved between man and the pufferfish fugu," *Gene*, vol. 208, no. 1, pp. 7–15, 1998.
- [4] G. Elgar, R. Sandford, S. Aparicio, A. Macrae, B. Venkatesh, and S. Brenner, "Small is beautiful: comparative genomics with the pufferfish (fugu rubripes)," *Trends in Genetics*, vol. 12, no. 4, pp. 145–150, 1996.

- [5] B. Brunner, T. Todt, S. Lenzner et al., "Genomic structure and comparative analysis of nine fugu genes: conservation of synteny with human chromosome xp22.2-p22.1," *Genome Research*, vol. 9, no. 5, pp. 437–448, 1999.
- [6] S. Brenner, G. Elgar, R. Sandford, A. Macrae, B. Venkatesh, and S. Aparicio, "Characterization of the pufferfish (fugu) genome as a compact model vertebrate genome," *Nature*, vol. 366, no. 6452, pp. 265–268, 1993.
- [7] G. Manning, "Genomic overview of protein kinases," *Wormbook : the Online Review of C. Elegans Biology*, pp. 1–19, 2005.
- [8] S. Kawa, J. Fujimoto, T. Tezuka, T. Nakazawa, and T. Yamamoto, "Involvement of brek, a serine/threonine kinase enriched in brain, in ngf signalling," *Genes to Cells*, vol. 9, no. 3, pp. 219–232, 2004.
- [9] R. A. Lindberg and T. Hunter, "Cdna cloning and characterization of eck, an epithelial cell receptor protein-tyrosine kinase in the eph/elk family of protein kinases," *Molecular and Cellular Biology*, vol. 10, no. 12, pp. 6316–6324, 1990.
- [10] A. Ullrich, J. R. Bell, and E. Y. Chen, "Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes," *Nature*, vol. 313, no. 6005, pp. 756–761, 1985.
- [11] N. Koike, H. Maita, T. Taira, H. Ariga, and S. M. M. Iguchi-Ariga, "Identification of heterochromatin protein 1 (hp1) as a phosphorylation target by pim-1 kinase and the effect of phosphorylation on the transcriptional repression function of hp1," *FEBS Letters*, vol. 467, no. 1, pp. 17–21, 2000.
- [12] F. Mari, S. Azimonti, I. Bertani et al., "Cdkl5 belongs to the same molecular pathway of mecp2 and it is responsible for the early-onset seizure variant of rett syndrome," *Human Molecular Genetics*, vol. 14, no. 14, pp. 1935–1946, 2005.
- [13] B. Salhia, N. L. Tran, A. Chan et al., "The guanine nucleotide exchange factors trio, ect2, and vav3 mediate the invasive behavior of glioblastoma," *American Journal of Pathology*, vol. 173, no. 6, pp. 1828–1838, 2008.
- [14] K. Yoshida, "Nuclear trafficking of pro-apoptotic kinases in response to dna damage," *Trends in Molecular Medicine*, vol. 14, no. 7, pp. 305–313, 2008.
- [15] J. Park, Y. Oh, and K. C. Chung, "Two key genes closely implicated with the neuropathological characteristics in down syndrome: dyrk1a and rcan1," *Bmb Reports*, vol. 42, no. 1, pp. 6–15, 2009.
- [16] F. Lamballe, R. Klein, and M. Barbacid, "Trkc, a new member of the trk family of tyrosine protein kinases, is a receptor for neurotrophin-3," *Cell*, vol. 66, no. 5, pp. 967–979, 1991.
- [17] J. T. Parsons, V. Wilkerson, and S. J. Parsons, "Structural and functional motifs of the rous sarcoma virus src protein," *Gene Amplification and Analysis*, vol. 4, pp. 1–19, 1986.
- [18] M. A. McTigue, J. A. Wickersham, C. Pinko et al., "Crystal structure of the kinase domain of human vascular endothelial growth factor receptor 2: a key enzyme in angiogenesis," *Structure*, vol. 7, no. 3, pp. 319–330, 1999.
- [19] J. Zha, Q. Zhou, L. G. Xu et al., "Rip5 is a rip-homologous inducer of cell death," *Biochemical and Biophysical Research Communications*, vol. 319, no. 2, pp. 298–303, 2004.
- [20] E. Kiss-Toth, S. M. Bagstaff, H. Y. Sung et al., "Human tribbles, a protein family controlling mitogen-activated protein kinase cascades," *Journal of Biological Chemistry*, vol. 279, no. 41, pp. 42703–42708, 2004.
- [21] F. Verissimo and P. Jordan, "Wnk kinases, a novel protein kinase subfamily in multi-cellular organisms," *Oncogene*, vol. 20, no. 39, pp. 5562–5569, 2001.
- [22] V. Lazar and J. G. N. Garcia, "A single human myosin light chain kinase gene (MLCK; MYLK) transcribes multiple nonmuscle isoforms," *Genomics*, vol. 57, no. 2, pp. 256–267, 1999.
- [23] M. Masaki, A. Ikeda, E. Shiraki, S. Oka, and T. Kawasaki, "Mixed lineage kinase lzk and antioxidant protein-1 activate nf- κ b synergistically," *European Journal of Biochemistry*, vol. 270, no. 1, pp. 76–83, 2003.
- [24] M. M. McLaughlin, S. Kumar, P. C. McDonnell et al., "Identification of mitogen-activated protein (MAP) kinase-activated protein kinase-3, a novel substrate of csbp p38 map kinase," *Journal of Biological Chemistry*, vol. 271, no. 14, pp. 8488–8492, 1996.
- [25] J. Pines, "The cell cycle kinases," *Seminars in Cancer Biology*, vol. 5, no. 4, pp. 305–313, 1994.
- [26] S. Yano, H. Tokumitsu, and T. R. Soderling, "Calcium promotes cell survival through cam-k kinase activation of the protein-kinase-b pathway," *Nature*, vol. 396, no. 6711, pp. 584–587, 1998.
- [27] B. R. Cairns, S. W. Ramer, and R. D. Kornberg, "Order of action of components in the yeast pheromone response pathway revealed with a dominant allele of the ste11 kinase and the multiple phosphorylation of the ste7 kinase," *Genes and Development*, vol. 6, no. 7, pp. 1305–1318, 1992.
- [28] T. Pawson, "Introduction: protein kinases," *Faseb Journal*, vol. 8, no. 14, pp. 1112–1113, 1994.
- [29] D. O. Frost, "Bdnf/trkb signaling in the developmental sculpting of visual connections," *Progress in Brain Research*, vol. 134, pp. 35–49, 2001.
- [30] M. H. Lee and H. Y. Yang, "Negative regulators of cyclin-dependent kinases and their roles in cancers," *Cellular and Molecular Life Sciences*, vol. 58, no. 12–13, pp. 1907–1922, 2001.
- [31] R. B. Irby, W. Mao, D. Coppola et al., "Activating src mutation in a subset of advanced human colon cancers," *Nature Genetics*, vol. 21, no. 2, pp. 187–190, 1999.
- [32] J. Zheng, D. R. Knighton, L. F. ten Eyck et al., "Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor," *Biochemistry*, vol. 32, no. 9, pp. 2154–2161, 1993.
- [33] W. P. Yu, J. M. M. Tan, K. C. M. Chew et al., "The 350-fold compacted fugu parkin gene is structurally and functionally similar to human parkin," *Gene*, vol. 346, pp. 97–104, 2005.
- [34] S. Baxendale, S. Abdulla, G. Elgar et al., "Comparative sequence analysis of the human and pufferfish huntington's disease genes," *Nature Genetics*, vol. 10, no. 1, pp. 67–76, 1995.
- [35] J. N. Volff and M. Schartl, "Evolution of signal transduction by gene and genome duplication in fish," *Journal of Structural and Functional Genomics*, vol. 3, no. 1–4, pp. 139–150, 2003.
- [36] Y. Niimura, "On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species," *Genome Biology and Evolution*, vol. 1, pp. 34–44, 2009.
- [37] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped blast and psi-blast: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [38] S. K. Hanks and T. Hunter, "The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification," *Faseb Journal*, vol. 9, no. 8, pp. 576–596, 1995.
- [39] B. Anand, V. S. Gowri, and N. Srinivasan, "Use of multiple profiles corresponding to a sequence alignment enables effective detection of remote homologues," *Bioinformatics*, vol. 21, no. 12, pp. 2821–2826, 2005.
- [40] V. S. Gowri, O. Krishnadev, C. S. Swamy, and N. Srinivasan, "Mulpssm: a database of multiple position-specific scoring

- matrices of protein domain families,” *Nucleic Acids Research*, vol. 34, pp. D243–246, 2006.
- [41] V. S. Gowri, K. G. Tina, O. Krishnadev, and N. Srinivasan, “Strategies for the effective identification of remotely related sequences in multiple pssm search approach,” *Proteins: Structure, Function and Genetics*, vol. 67, no. 4, pp. 789–794, 2007.
- [42] S. R. Eddy, “Profile hidden markov models,” *Bioinformatics*, vol. 14, no. 9, pp. 755–763, 1998.
- [43] A. Krupa, K. R. Abhinandan, and N. Srinivasan, “King: a database of protein kinases in genomes,” *Nucleic Acids Research*, vol. 32, pp. D153–D155, 2004.
- [44] A. Krupa and N. Srinivasan, “The repertoire of protein kinases encoded in the draft version of the human genome: atypical variations and uncommon domain combinations,” *Genome Biology*, vol. 3, no. 12, p. RESEARCH0066, 2002.
- [45] A. Krupa and N. Srinivasan, “Diversity in domain architectures of ser/thr kinases and their homologues in prokaryotes,” *BMC Genomics*, vol. 6, 2005.
- [46] Anamika, N. Srinivasan, and A. Krupa, “A genomic perspective of protein kinases in plasmadium falciparum,” *Proteins: Structure, Function and Genetics*, vol. 58, no. 1, pp. 180–189, 2005.
- [47] A. Krupa, G. Preethi, and N. Srinivasan, “Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in ser/thr and tyr kinases,” *Journal of Molecular Biology*, vol. 339, no. 5, pp. 1025–1039, 2004.
- [48] A. Krupa, Anamika, and N. Srinivasan, “Genome-wide comparative analyses of domain organisation of repertoires of protein kinases of arabidopsis thaliana and oryza sativa,” *Gene*, vol. 380, no. 1, pp. 1–13, 2006.
- [49] K. Anamika and N. Srinivasan, “Comparative kinomics of plasmodium organisms: unity in diversity,” *Protein and Peptide Letters*, vol. 14, no. 6, pp. 509–517, 2007.
- [50] A. Müller, R. M. MacCallum, and M. J. E. Sternberg, “Benchmarking psi-blast in genome annotation,” *Journal of Molecular Biology*, vol. 293, no. 5, pp. 1257–1271, 1999.
- [51] L. A. Kelley, R. M. MacCallum, and M. J. E. Sternberg, “Enhanced genome annotation using structural profiles in the program 3d-pssm,” *Journal of Molecular Biology*, vol. 299, no. 2, pp. 499–520, 2000.
- [52] K. Bryson, L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi, and D. T. Jones, “Protein structure prediction servers at university college london,” *Nucleic Acids Research*, vol. 33, no. 2, pp. W36–W38, 2005.
- [53] R. Chenna, H. Sugawara, T. Koike et al., “Multiple sequence alignment with the clustal series of programs,” *Nucleic Acids Research*, vol. 31, no. 13, pp. 3497–3500, 2003.
- [54] S. Kumar, K. Tamura, and M. Nei, “Mega: molecular evolutionary genetics analysis software for microcomputers,” *Computer Applications in the Biosciences*, vol. 10, no. 2, pp. 189–191, 1994.
- [55] K. Anamika, K.R. Abhinandan, K. Deshmukh et al., “Classification of nonenzymatic homologues of protein kinases,” *Comparative and Functional Genomics*, vol. 2009, Article ID 365637, 17 pages, 2009.
- [56] A. Bateman, E. Birney, L. Cerruti et al., “The pfam protein families database,” *Nucleic Acids Research*, vol. 30, no. 1, pp. 276–280, 2002.
- [57] A. Krogh, B. Larsson, G. Von Heijne, and E. L. L. Sonnhammer, “Predicting transmembrane protein topology with a hidden markov model: application to complete genomes,” *Journal of Molecular Biology*, vol. 305, no. 3, pp. 567–580, 2001.
- [58] M. Kanehisa and S. Goto, “Kegg: kyoto encyclopedia of genes and genomes,” *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [59] P. Shannon, A. Markiel, O. Ozier et al., “Cytoscape: a software environment for integrated models of biomolecular interaction networks,” *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [60] L. R. Potter and T. Hunter, “Phosphorylation of the kinase homology domain is essential for activation of the a-type natriuretic peptide receptor,” *Molecular and Cellular Biology*, vol. 18, no. 4, pp. 2164–2172, 1998.
- [61] L. N. Johnson, M. E. M. Noble, and D. J. Owen, “Active and inactive protein kinases: structural basis for regulation,” *Cell*, vol. 85, no. 2, pp. 149–158, 1996.
- [62] E. V. Koonin, “Orthologs, paralogs, and evolutionary genomics,” *Annual Review of Genetics*, vol. 39, pp. 309–338, 2005.
- [63] G. Manning, D. S. Reiner, T. Lauwaet et al., “The minimal kinome of giardia lamblia illuminates early kinase evolution and unique parasite biology,” *Genome Biology*, p. R66, 2011.
- [64] H. Leffler, S. Carlsson, M. Hedlund, Y. Qian, and F. Poirier, “Introduction to galectins,” *Glycoconjugate Journal*, vol. 19, no. 7–9, pp. 433–440, 2002.
- [65] T. N. Petersen, S. Brunak, G. vonHeijne et al., “SignalP 4.0: discriminating signal peptides from transmembrane regions,” *Nature Methods*, vol. 8, no. 10, pp. 785–786, 2011.
- [66] W. Wei, Z. W. Cao, Y. L. Zhu et al., “Conserved genes in a path from commensalism to pathogenicity: comparative phylogenetic profiles of staphylococcus epidermidis rp62a and atcc12228,” *Bmc Genomics*, vol. 7, article no. 112, 2006.
- [67] M. Saraste and M. Hyvonen, “Pleckstrin homology domains: a fact file,” *Current Opinion in Structural Biology*, vol. 5, no. 3, pp. 403–408, 1995.
- [68] M. Deak, A. D. Clifton, J. M. Lucocq, and D. R. Alessi, “Mitogen- and stress-activated protein kinase-1 (msk1) is directly activated by mapk and sapk2/p38, and may mediate activation of creb,” *Embo Journal*, vol. 17, no. 15, pp. 4426–4441, 1998.