Int J Pharm Bio Sci 2015 July; 6(3): (B) 992 - 1005

Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

COMPARATIVE GENOMICS OF MYOCILIN PROTEIN

R. RAMANATHAN*[1], R. RAMASAMY[2] AND S. KRISHNAN[3]

[1] Department of Physics, Government Arts College, Kulithalai, India
[2] Research and P.G. Department of Physics, National College (Autonomous), Tiruchi, India.
[3] Department of spine surgery, Sunshine Hospitals, Secunderabad, India

ABSTRACT

Glaucoma is one of the leading diseases which causes irreversible visual impairment and blindness throughout the world. It constitutes clinically and genetically heterogenous group of optic neuropathies. Specific mutations in the myocilin gene cause primary open angle glaucoma with varying age-of-onset and degree of severity. Though the structure of the gene and protein have been elucidated still the role of mutant protein is not clearly understood. To study myocilin and glaucoma-associated myocilin mutants in higher animals and to understand, how these molecules may alter intraocularpressure and outflow resistance, it would be more helpful if we have animal models which could be used in the study. For finding out suitable animal models, sequence anlysis approach has been the first step in many diseases. Here we took a sequence analysis approach to know more about the related orthologes of myocilin proteins from different species.

KEYWORDS: Myocilin, Glaucoma, Phylogenic tree,Genomics.

R. RAMANATHAN
Department of Physics, Government Arts College, Kulithalai, India

***Corresponding author**

INTRODUCTION

Glaucoma is a term used to describe a group of disorders that have in common a characterestic degeneration of optic nerve associated with typical visual field defects and usually elevated introcular pressure, if left untreated it leads to absolute irreversible blindness. Despite many decades of research little is known about the molecular biology of the defect. Enzymes, structural proteins and proteins involved in embryogenesis and development of the eye may be important to normal physiology of trabecular meshwork and defects in the genes coding for these proteins may play a role in the genetic predisposition to glaucoma. Glaucoma is often associated with elevated levels of intraocular pressure (IOP). Elevated IOP is neither necessary nor sufficient for the development of the disease. Individuals who are ocular hypertensives, have no signs of optic damage[1], however on the other hand people with normal ocular tension show damage due to glaucoma[2], thus increase in the IOP is considered as a risk factor for the disease. There are various forms of the disease but the two most common are primary open angle glaucoma (POAG) and juvenile open angle glaucoma (JOAG). POAG is autosomal recessive and it occurs more frequently. Discovery of genetic factors that influence the development and progression of glaucoma will hopefully lead to a better diagnosis and treatment. The disease results in a characteristic degeneration of the optic nerve that is usually associated with an elevation of intraocular pressure. Pressure within the eye is dependent on the rate of production of a fluid (aqueous humor) by the ciliary body and on the rate of removal of the fluid by the trabecular meshwork[3]. There is a complex genetic mechanism when IOP is generally elevated. A total of about 73 disease causing mutations are known to be present in myocilin (MYOC) gene[4]. Using micro satellite markers the region of linkage of the disease is to be associated with region 1q21-1q31[5]. All people with Gln368STOP mutation in MYOC mutation had elevated IOP. Other factors in addition to this mutation also plays a role in the development of elevated IOP and glaucoma[6]. Further, normal IOP may also lead to eye damage due to glaucoma. Oxidative and nitrative stress appears to play a role in glaucomatous optic nerve damage[7]. Oxidative stress may play a significant role in the pathogenesis of POAG[8]. Myocilin is expressed in many ocular and non-ocular tissues, is found in both intracellular and extracellular spaces[9] and much more is to be learnt about the mechanism of action of myocilin. In addition to the mutation of MYOC which is one of the reasons for glaucoma, oxidative stress may be a cause for glaucoma. MYOC protein is found from Pufferfish to Humans. This protein has sustained the evolution and there is a need to understand the mechanism of evolution and also its function, which is till date not understood properly. To understand the function of myocilin several animal models are considered such as rabbit[10], rat[11],cat[12] and various other animals. In the present study various available gene and protein sequences are compared and a phylogenic tree is constructed. From this study, it will be possible to select a suitable orthologous animal myocilin protein or gene sequence which matches with the mutated human sample for further analysis. If the study requires a particular mutation at a particular place, from the available data source, it will be possible to select a sequence of our preference from the animal source. This particular protein is expressed right from the zebra fish to humans and study of this protein and its gene sequence is important for understanding developmental genetics in various species. Further, the function of this protein is not understood completely and its presence and function in tissues other than the eye is yet to be established in detail. Even the presence of MYOC protein in the eye needs further investigation. Oxidative stress is known to cause glaucoma and there are several antioxidant nanoparticles such as Zinc Oxide[13], which can be used to evaluate the effect of antioxidant property on MYOC protein. Further, there is a need for integrating and presenting the available data in the area of research[14], here, the tools in bioinformatics are used as for as MYOC is concerned.

MATERIALS AND METHODS

Human MYOC Protein sequence was used to fetch protein otrhologues from Chimp, Rhesus monkey, Rabbit, Cattle, Horse, Dog, Cat, Pig, Mouse, Rat, Chicken, Puffer fish and Zebra fish. Sequences of Chimp, Monkey, Horse and Chicken were predicted orthologues. Different orthologous protein sequences were used in BLAST, separately with reference mRNA sequences to fetch orthologus mRNA sequences. The orthologus protein sequences were subjected to multiple alignment using the the Application "CLC Workbench software"[15], to generate the protein alignment along with sequence LOGO and consensus and the pH/Charge information for different orthologues and also the phylogenetic tree with bootstrap. The first step in a bootstrap analysis is to resample the alignment columns with replacement. That is, in the resampled alignment, a given column in the original alignment may occur two or more times, while some columns may not be represented in the new alignment at all. The re-sampled alignment represents an estimate of how a different set of sequences from the same genes and the same species may have evolved on the same tree. If a new tree reconstruction on the re-sampled alignment results in a tree similar to the original one, this increases the confidence in the original tree. By resampling a number of times it is possibly to put reliability weights on each internal branch of the inferred tree. If the data was bootstrapped a 100 times, a bootstrap score of 100 means that the corresponding branch occurs in all 100 trees made from re-sampled alignments. Thus, a high bootstrap score is a sign of greater reliability. In this case, the data was bootstrapped a 1000 times and is highly reliable. Here the tree is constructed using "The neighbour joining method"[16] with a bootstrap value of 100 and 1000. Using MEME[17] all the 13 orthologues were compared to get information on the blocks of conservation based on MAST output[18]. The human mRNA was used in BLAST against available reference database sequence database to get the Myocilin gene sequence which was then compared with other available genomic sequences using the program VISTA[19]. Using blast P different othologous protein sequences were compared and the percentage of similarity between Myocilin protein orthologs[20,21]. Evolutionary analyses were conducted in MEGA6[22]. The repeat table of the genomic region of the Human Myocilin gene was tabulated using the repeat masker[23].

Accession Numbers Used

AAS68633, XP_422235, AAN59763, CAF97613, AAT42260, BAA24532, AAD46401, AAI04572, NP_034995.2, AAZ78213.1, XP_513995.2, XP_001492349.1, AAO38666.1, XP_001099905.1

RESULTS

1. Estimates of Evolutionary Divergence between Sequences

Using blastP different othologous protein sequences were compared, for evaluating the percentage of similarities between the Myocilin protein orthologs. Using MEGA6, pairwise identity were determined and these results are tabulated in table 1.

Table 1

Pairwise identity and Percentage of identity of Myocilin Protein between different orthologs

		HUMAN	CHIMP	MONKEY	HORSE	CATTLE	PIG	CAT	DOG	RABBIT	RAT	MOUSE	CHICK	ZEBRA FISH	PUFFER FISH
HUMAN	Pairwise	**0**	0.01	0.03	0.13	0.17	0.16	0.14	0.15	0.13	0.18	0.18	0.56	0.72	0.70
	% identity	100	98	96	84	82	81	86	83	86	81	82	65	43	45
CHIMP	Pairwise		0	0.04	0.14	0.18	0.17	0.14	0.16	0.14	0.19	0.19	0.57	0.74	0.71
	% identity		100	95	84	81	81	86	82	85	81	81	65	43	45
MONKEY	Pairwise			0	0.13	0.17	0.17	0.14	0.15	0.13	0.17	0.17	0.56	0.72	0.71
	% identity			100	84	81	81	86	84	86	81	82	66	43	45
HORSE	Pairwise				0	0.15	0.19	0.15	0.15	0.17	0.21	0.21	0.53	0.70	0.70
	% identity				100	85	81	75	82	84	80	80	68	45	46
CATTLE	Pairwise					0	0.16	0.17	0.18	0.19	0.21	0.22	0.56	0.69	0.69
	% identity					100	84	83	82	82	79	79	67	44	48
PIG	Pairwise						0	0.15	0.17	0.17	0.20	0.20	0.55	0.72	0.71
	% identity						100	84	82	83	79	79	67	43	46
CAT	Pairwise							0	0.12	0.17	0.20	0.21	0.55	0.72	0.69
	% identity							100	87	84	79	81	52	43	47
DOG	Pairwise								0	0.17	0.23	0.22	0.51	0.71	0.72
	% identity								100	82	78	78	69	44	47
RABBIT	Pairwise									0	0.18	0.18	0.58	0.72	0.72
	% identity									100	82	82	66	42	45
RAT	Pairwise										0	0.07	0.61	0.76	0.73
	% identity										100	92	64	41	46
MOUSE	Pairwise											0	0.59	0.74	0.75
	% identity											100	66	42	45
CHICK	Pairwise												0	0.71	0.75
	% identity												100	57	60
ZEBRA	Pairwise													0	0.47
	% identity													100	56
PUFFER	Pairwise														0
	% identity														100

The number of amino acid substitutions per site from between sequences is shown. Analyses were conducted using the Poisson correction model[24]. The analysis involved 14 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 418 positions in the final dataset.

2. Conservation between genomic sequences

Vista plot gives the overall graphic representation of conservation between the genomic sequences of different species with a window size of 100Bp with the conservation of a minimum of 40% and in each panel the line in the middle represents 60% conservation level.

vista plot

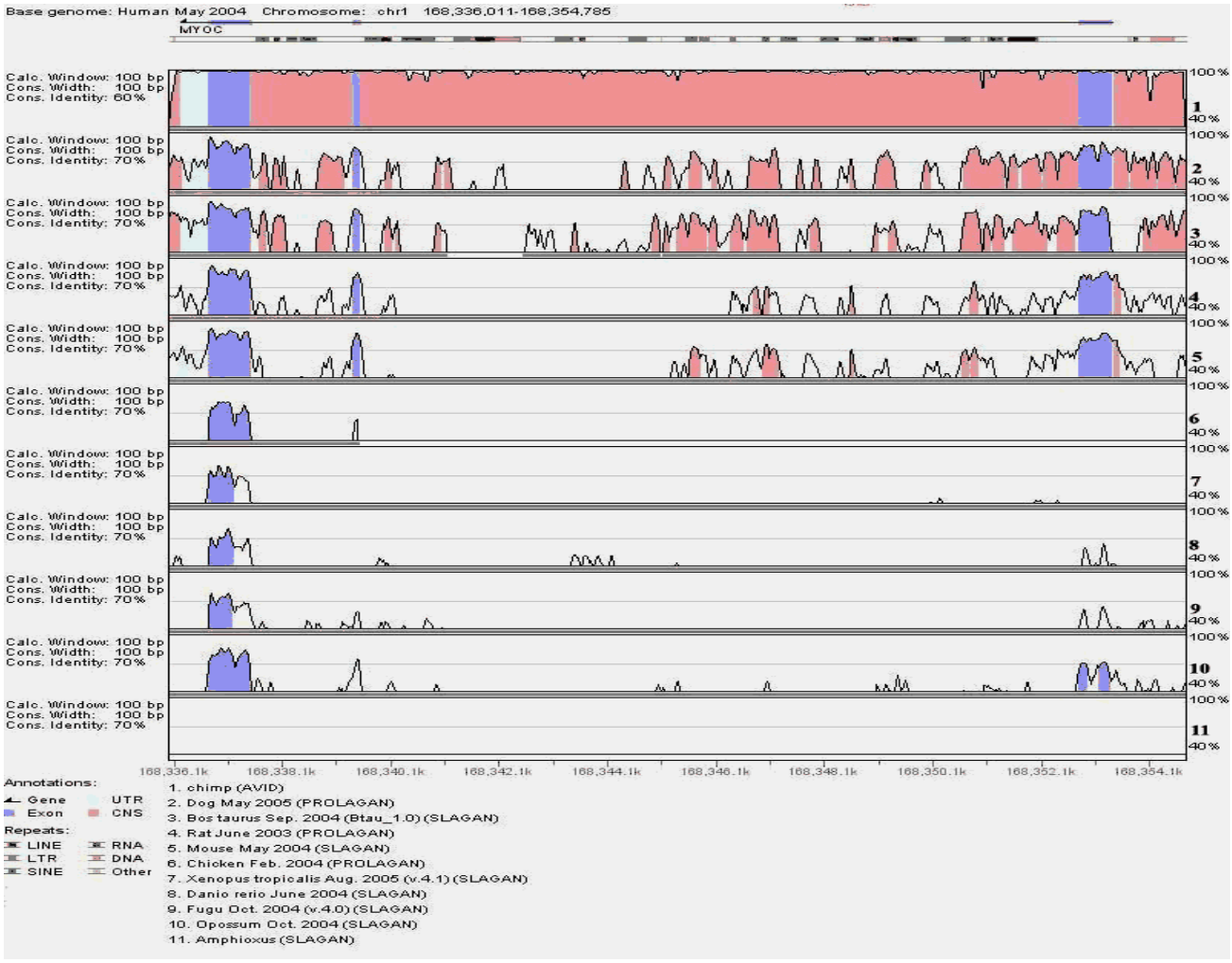


Figure 1

Vista plot showing the result of different Myocilin genomic sequences compared with the Human myocilin genomic sequence. Annotations and representations given in the figure itself. UTR (Untranslated Region), NS (Non coding sequences)

3. Construction of Phylogenic tree

Phylogenetic Tree of orthologs, compared shows that the proteins have evolved from the common ancestor, Protein from the species namely Human, Chimp, Rhesus monkey, Rabbit, Rat evolved from a common ancestor, another group consisting of cat, dog, horse, cattle and pig evolved from another common ancestor which diverged from the root ancestor, which gave rise to different groups long ago during the process of evolution. The common ancestor which gave rise to a group consisting of Puffer fish and Zebra fish diverged from the other groups very early during evolutionary process. Among these different groups Human and Chimp, Dog and Cat, Cattle and Horse, Puffer fish and Zebra fish shows a greater degree of similarities within their respective groups. Evolutionary divergence is also marked to scale and is represented in the phylogenic tree.

Phylogenic tree of the Myocilin proteins

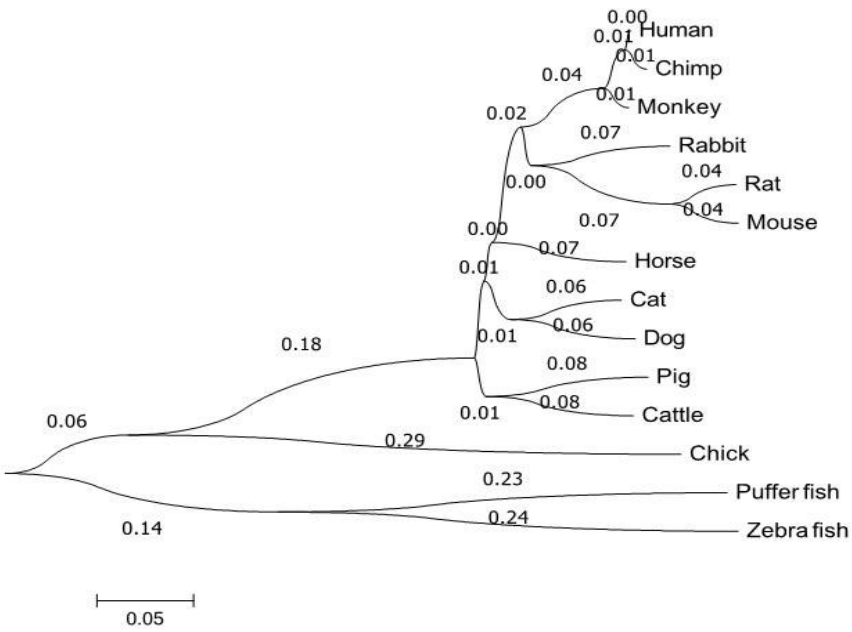
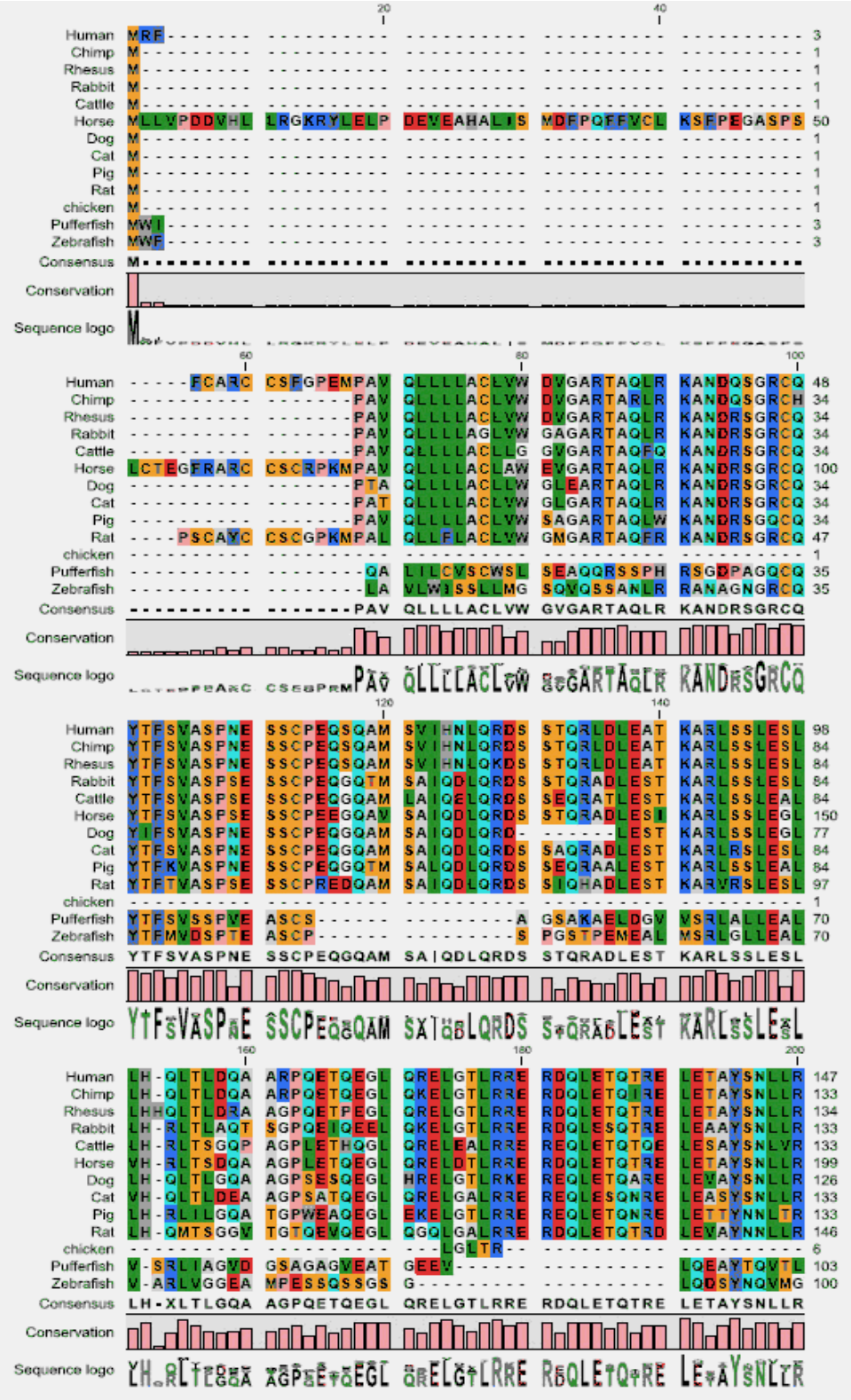
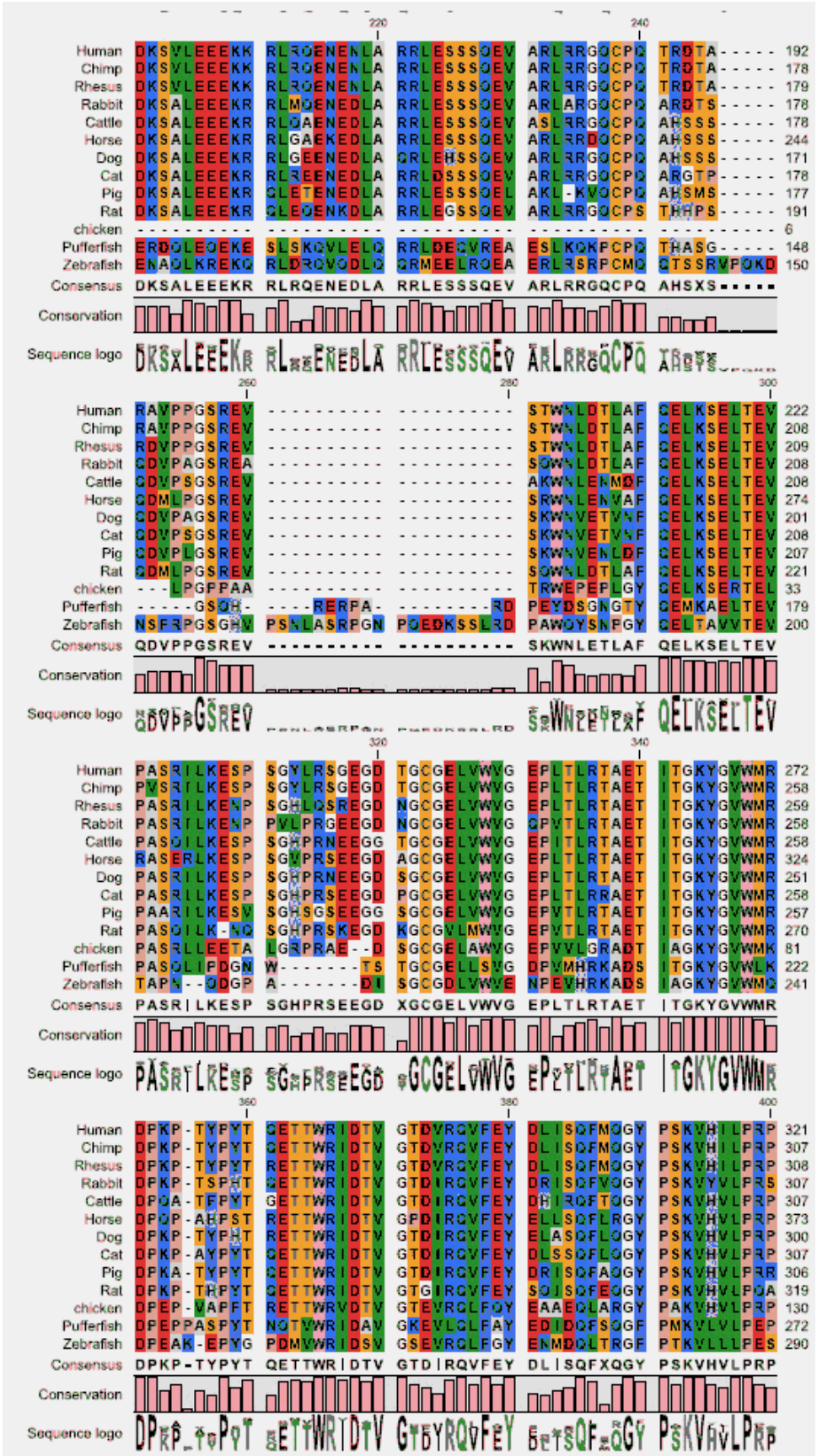


Figure 2
Phylogenetic tree of the Myocilin proteins from different species

4. Comparitive alignment between different MYOC orthologues and conservation pattern with charge plot

Alignment between different orthologues of MYOC




Human
Chimp
Rhesus
Rabbit
Cattle
Horse
Dog
Cat
Pig
Rat
chicken
Pufferfish
Zebrafish
Consensus
Conservation
Sequence logo

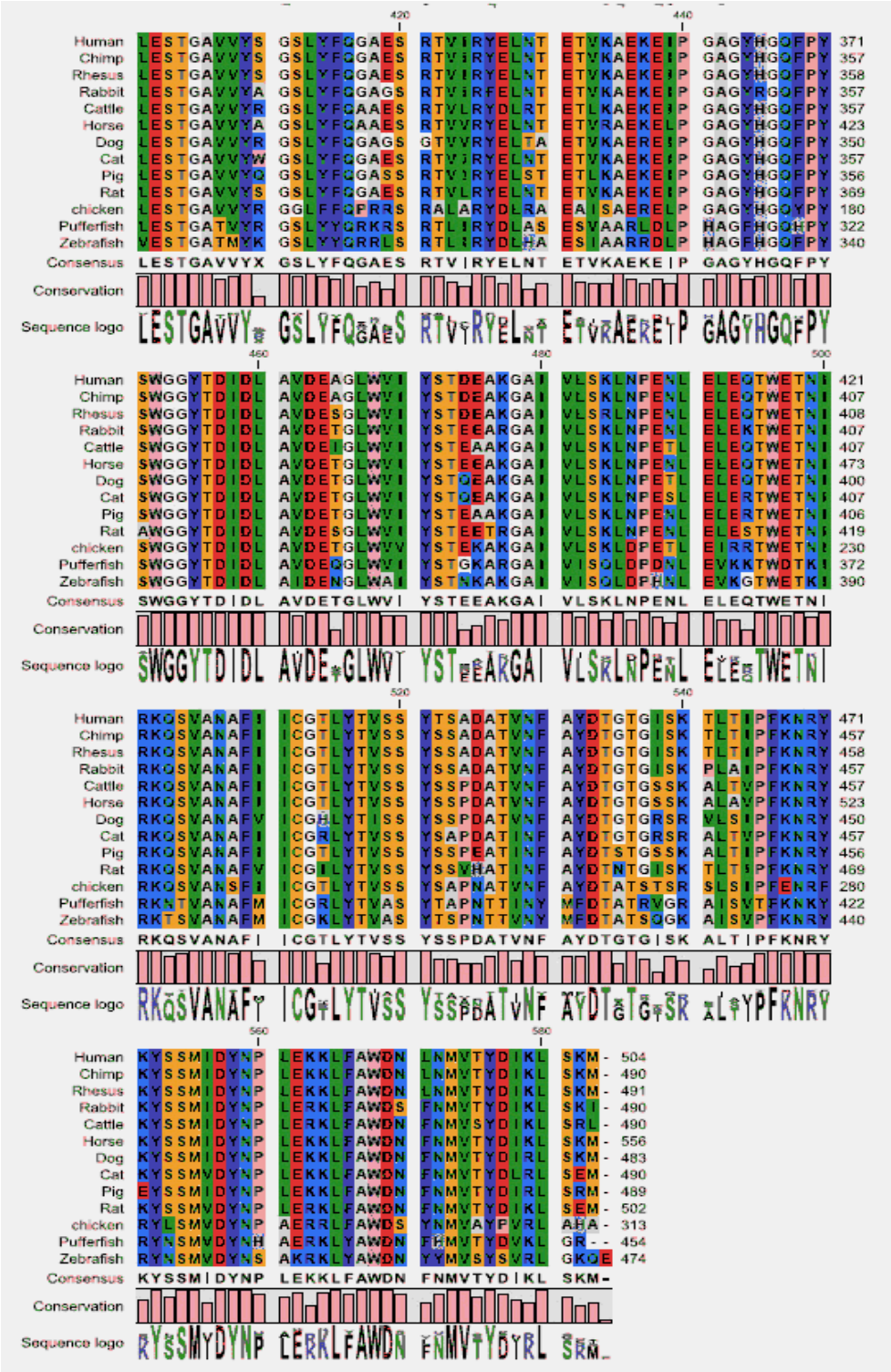


Figure 3

The tabular format of a multiple alignment of 13 MYOC protein sequences. Sequence names appear at the beginning of each row and the residue position is indicated by the numbers at the top of the alignment columns as well as the extreme left included in each panel are the protein alignment along with consensus in the particular stretch, Bar diagram showing the level of conservation at a particular position, and sequence LOGO of the given stretch.

The input to multiple alignment algorithms is a number of homologous sequences i.e. sequences that share a common ancestor and most often also share molecular function. The generated alignment is a table, where each row corresponds to an input sequence and each column corresponds to a position in the alignment. An individual column in this table represents residues that have all diverged from a common ancestral residue. Gaps in the table (commonly represented by a '-') represent positions where residues have been inserted or deleted and thus do not have ancestral

counterparts in all sequences. The alignment algorithm has three parameters concerning gap costs: Gap open cost, Gap extension cost and End gap cost. The degree of conservation is high among closely related species and it increases among higher groups in the phyla with respect to Phylogenetic tree. Some stretches of protein are highly conserved which may be according to the functional domain and the higher degree of conservation.

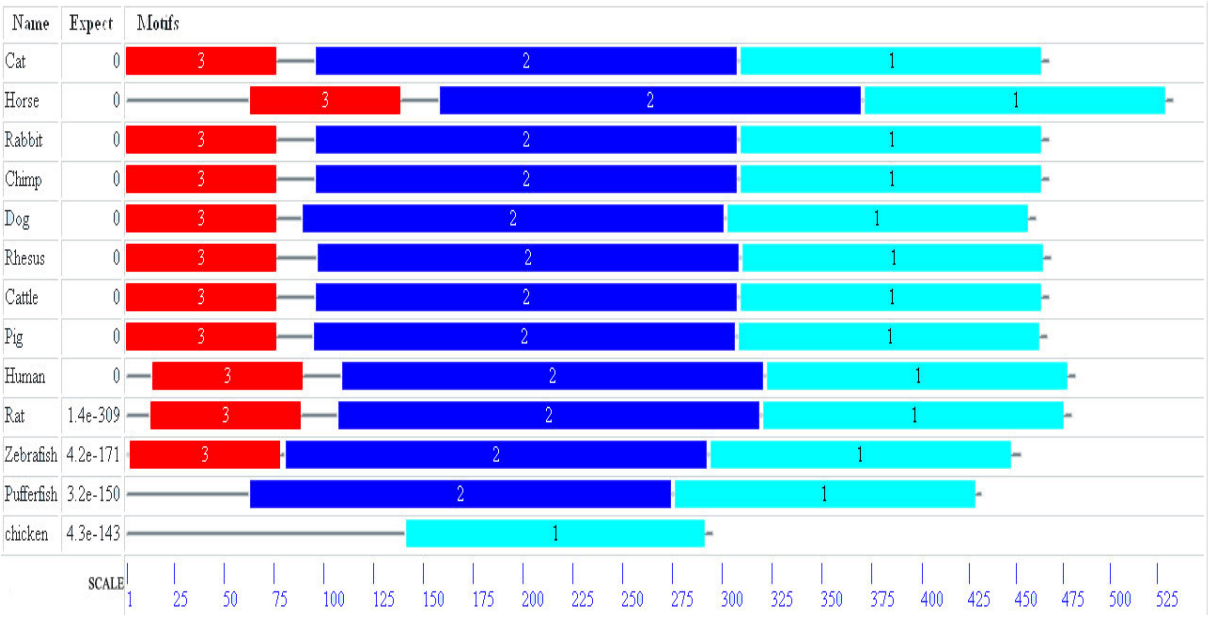
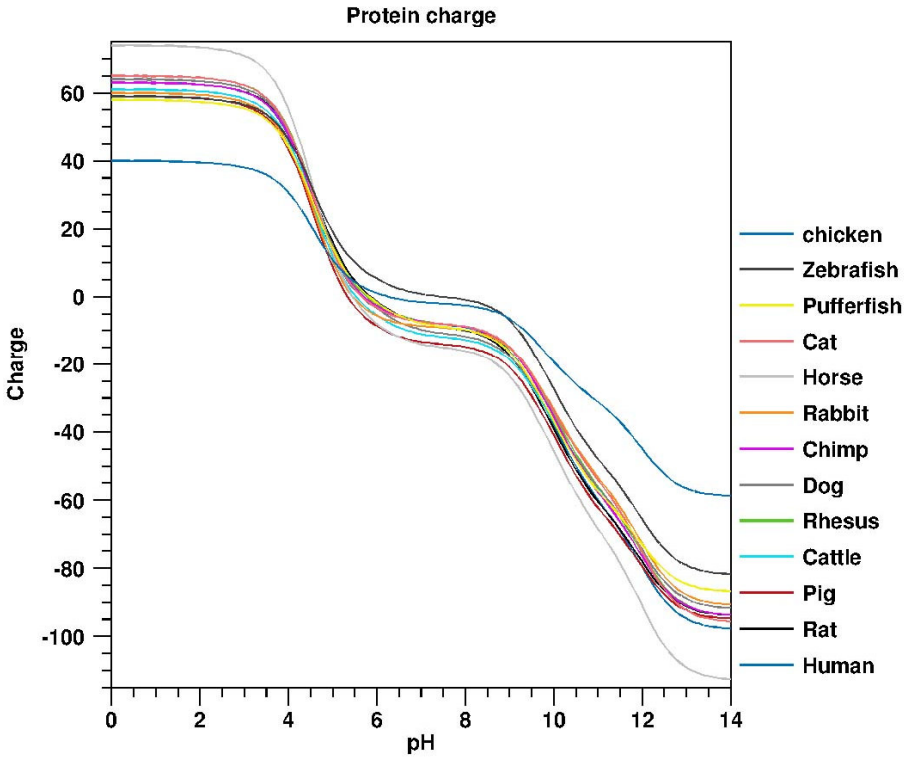


Figure 4
The conservation pattern observed as three different blocks In the different orthologues of the protein.

Using MEME, all the 13 orthologues are compared (Fig 4) which shows that the orthologues show a very high degree of conservation and the proteins show three different blocks of conservation. There is no significant degree of conservation in the intronic regions among species, there by suggesting the importance for different exons of the gene which try to remain without accumulating changes, while the introns have accumulated lots of variations among species there by suggesting that these regions are less likely to be involved in pathogenesis. Till date almost all of the mutations in Myocilin are in the coded protein and none of the mutations has been reported in the non coding regions of the gene.



Graph 1
Graph of the electric charge as a function of pH for all the MYOC orthologues is shown

Graph 1 is particularly useful for finding the net charge of the protein at a given pH. This knowledge can be used e.g. in relation to isoelectric focusing on the first dimension of 2D-gel electrophoresis. The isoelectric point (pI) is found where the net charge of the protein is zero. And here all the orthologues seem to have pI between pH of 5-7.

Table 2
Repeat table of the human Myocilin gene

	Number of elements*	Length occupied	Percentage of sequence
SINEs:	23	5298 bp	25.84 %
ALUs	16	4166 bp	20.32 %
MIRs	7	1132 bp	5.52 %
LINEs:	6	1345 bp	6.56 %
LINE 1	5	1286 bp	6.27 %
LINE 2	1	59 bp	0.29 %
L3/CR1	0	0 bp	0.00 %
LTR elements:	4	1594 bp	7.78 %
MaLRs	0	0 bp	0.00 %
ERVL	3	1484 bp	7.24 %
ERV_classI	0	0 bp	0.00 %
ERV_classII	0	0 bp	0.00 %
DNA elements:	2	750 bp	3.66 %
MER1_type	2	750 bp	3.66 %
MER2_type	0	0 bp	0.00 %
Unclassified:	0	0 bp	0.00 %
Total interspersed repeats:		8987 bp	43.84 %
Small RNA:	0	0 bp	0.00 %
Satellites:	0	0 bp	0.00 %
Simple repeats:	5	228 bp	1.11 %
Low complexity:	1	21 bp	0.10 %

**** most repeats fragmented by insertions or deletions have been counted as one element.***

total length: 20501 bp (20501 bp excl N/X-runs)
C level: 45.09 %
bases masked: 9236 bp (45.05 %)

Myocilin gene is infested with repeat elements like SINES and LINE although not to very extreme, for the repeat constituency of the genomic sequence of the myocilin gene as seen in the repeat table 2. Though the conservation among different is more pronounced in the UTR elements which seems to represent the common mode of regulation of this gene among different species.

DISCUSSION

Pairwise identity and Percentage of identity of Myocilin protein between different orthologs shows that and Zebra fish and Puffer fish myocilin seems to have diverged very early in evolution form the ancestors of other groups thus showing 40% to 50% similarity with other orthologes. Monkey and Chimp Myocilin protein shows 95% and 98% similarity with Human Myocilin protein, which are later in evolution than Zebra fish and Puffer fish. Further, this method gives more accurate method of estimation of the relative closeness between different Myocilin orthologes. Vista Plot of genomic sequences (Fig 1) shows that the human and chimp to be closely related species with about 98% similarities of human myocilin genomic sequences with other species shows a greater degree of conservation in the exons and the regulatory sequences. The phylogenic tree which is obtained with thier evolutionary divergence is of great importance because it is bootstrapped a 1000 times and is more reliable than other methods, hence, this phylogenic tree can be considered as most accurate. From the alignment of the protein sequence, and through the conservation pattern of the three different blocks, in different MYOC orthologes, from this comparative approach, valuable evolutionary information can be obtained about which amino acid substitutions are functionally tolerant to the organism and which are not. This information can be used to identify substitutions that affect the protein

function stability, and is of high importance to the study of protein.

CONCLUSION

Long-term studies will be important in determining the role of Myocilin's physiological role in affecting intraocular pressure and the aqueous outflow pathway. As in other diseases caused by conformational changes, mutations alone need not be responsible for the disease as changes in the molecular environment and other factors in addition to the presence of the key mutations seem to play an important role in POAG. To elucidate the role of Myocilin in POAG it is important to generate animal models, in which mutated myocilin is to be inserted in the genomic locus of wild-type myocilin. The regulatory region of the gene has to be studied in detail so as to understand the molecular mechanisms involving the Myocilin in detail. The phylogenic tree which is constructed is more accurate, than already published data, which has the evolutionary weights attached to it. Further, the LOGO construction gives the change in amino acid at a particular position for different orthologes of MYOC. This gives an insight to the evolutionary basis of MYOC and will be helpful to analze the mutations which are caused in POAG, which has clinical significance.

REFERENCES

1. Maier PC., Funk J., Schwarzer G., Antes G., Falck-Ytter YT. Treatment of ocular hypertension and open angle glaucoma: meta-analysis of randomised controlled trials. BMJ 331, 134, (2005)
2. Henry E., Newby DE., Webb DJ., O'Brien C. Peripheral endothelial dysfunction in normal pressure glaucoma. Investig. Ophthalmol. Vis. Sci. 40, 1710–1714, (1999)
3. Wiggs JL, et al., Prevalence of mutations in TIGR/Myocilin in patients with adult and juvenile primary open-angle glaucoma. Am. J. Hum. Genet. 63, 1549–1552, (1998)
4. Gong G., Kosoko-Lasaki O., Haynatzki GR., Wilson MR. Genetic dissection of myocilin glaucoma. Hum. Mol. Genet. 13 Spec No, R91–R102, (2004)
5. Sheffield VC., et al., Genetic linkage of familial open angle glaucoma to chromosome 1q21-q31. Nat. Genet. 4, 47–50, (1993)
6. Allingham RR., et al., Gln368STOP myocilin mutation in families with late-onset primary open- angle glaucoma. Investig. Ophthalmol. Vis. Sci. 39, 2288–2295, (1998)
7. Aslan M., Cort A., Yucel I. Oxidative and nitrative stress markers in glaucoma. Free Radical Biology and Medicine 45, 367–376, (2008)
8. Saccà SC., Izzotti A., Rossi P., Traverso C. Glaucomatous outflow pathway and oxidative stress. Exp. Eye Res. 84, 389–399, (2007)
9. Resch ZT., Fautsch MP. Glaucoma-associated myocilin: A better understanding but much more to learn. Experimental Eye Research 88, 704–712, (2009)
10. Shepard AR., et al., Characterization of rabbit myocilin: Implications for human myocilin glycosylation and signal peptide usage. BMC Genet. 4, 5, (2003)
11. Taguchi M., et al., Molecular cloning and expression profile of rat myocilin. Mol. Genet. Metab. 70, 75–80, (2000)
12. Fautsch MP., Vrabel AM., Johnson DH. Characterization of the Felix domesticus (cat) glaucoma-associated protein myocilin. Exp. Eye Res. 82, 1037–1045, (2006)
13. Das D., Nath BC., Phukon P., Kalita A., Dolui SK. Synthesis of ZnO nanoparticles and evaluation of antioxidant and cytotoxic activity, Colloids Surfaces B Biointerfaces, 1411,556-560, (2013)
14. Joseph J., Hassan S., Vasanthi R., Vanaja Kumar. Microbial Genome Databases: A user's perspective, IJPBS V1(2), (2010)
15. CLC Bio-Qiagen, Aarhus, Denmark. www.clcbio.com
16. Saitou N., Nei M. The neighbor-joining method: a new method for reconstructing

phylogenetic trees, Mol.Biol.Evol, 4, 406-425, (1987)

17. Bailey TL., Bodén M., Buske FA., Frith M., Grant CE., Clementi L., Ren J., Li WW.,Noble WS. "MEME SUITE: tools for motif discovery and searching". Nucleic Acids Research, 37:W202-W208, (2009) http://meme.nbcr.net/meme/meme.html
18. Gribskov TL. Combining evidence using p-values: application to sequence homology searches, Bioinformatics 14, 48-54, (1998)
19. Frazer KA., Pachter L., Poliakov A., Rubin EM., Dubchak I. "VISTA: computational tools for comparative genomics". Nucleic Acids Research 32(Web Server issue):W273-W279. doi:10.1093/nar/gkh458, (2004) http://pipeline.lbl.gov/cgi-bin/GenomeVista
20. Altschul SF., Gish W., Miller W., Myers EW., Lipman DJ. "Basic local alignment search tool." J. Mol. Biol. 215:403-410, (1990) http://www.ncbi.nlm.nih.gov/blast
21. Altschul SF., Madden TL., Schäffer AA., Zhang J., Zhang Z., Miller W., Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402, (1997).
22. Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution30: 2725-2729, (2013)
23. Smit AFA., Hubley R., Green P. unpublished data. Current Version: open-4.0.5 (RMLib: 20140131 & Dfam: 1.3), http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker
24. E. Zuckerkandl, and L. Pauling. Evolutionary divergence and convergence in proteins. In: V.Bryson and H.J. Vogel (eds), Evolving Genes and Proteins, Academic Press, New York, 1965, pp. 97-166.