

## Evolutionary analysis of *PHLPP1* gene in humans and non-human primates

Padmanabhan Anbazhagan<sup>1,2</sup>, Meera Purushottam<sup>1</sup>, H. B. Kiran Kumar<sup>1</sup>, Shobana Kubendran<sup>1</sup>, Odity Mukherjee<sup>2</sup>, Samir Kumar Brahmachari<sup>3</sup>, Sanjeev Jain<sup>1</sup> and Ramanathan Sowdhamini<sup>2,\*</sup>

<sup>1</sup>Molecular Genetics Laboratory, Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bangalore, India; <sup>2</sup>National Centre for Biological Sciences, Bangalore, India; <sup>3</sup>Institute of Genomics and Integrative Biology, Delhi, India; Ramanathan Sowdhamini\* - E-mail: mini@ncbs.res.in; Phone: 91-80-23666250; Fax: 91-80-23666662; \* Corresponding Author

received June 11, 2008; revised July 06, 2008; accepted July 26, 2008; published August 01, 2008

### Abstract:

The chromosome 18q22-23 region has been shown to be implicated in bipolar disorder (BPAD) by several studies. *PHLPP1* gene, in the locus (chromosome 18q22-23), is involved in circadian pathways and bears modules like 'PH domain and leucine rich repeat protein phosphatase'. This gene also contains a polyglutamine (CAG or PolyQ) repeat motif at the carboxyl terminal end. A comparative analysis of the PolyQ repeats of the *PHLPP1* gene in humans, non-human primates and other species has been attempted in order to investigate the possible significance of repeat length as seen in other triplet-repeat associated diseases. Sequencing of the CAG repeat in humans and in non-human primates revealed that the CAG repeat is not polymorphic in humans; whereas, in other species it shows an area of high variability, both in length and sequence composition. Despite the conservation of circadian clock components in different species, there is remarkable diversity in the protein structure, regulation and biochemical functions of the circadian orthologs. These can be due to specific adaptations in accordance with the physiology of the particular species providing a species-specific biological advantage.

**Keywords:** *PHLPP1*; bipolar disorder; CAG repeat; PolyQ; phylogenetic analysis; evolution

### Background:

Disruption in circadian rhythm such as, the sleep-wake cycle, body temperature and cortisol secretion, frequently characterize mood disorders, like bipolar disorder (BPAD). Changes in circadian functioning are a prominent aspect of BPAD [1]. The *PHLPP1* Pleckstrin homology (PH) domain and leucine rich repeat protein phosphatase, a gene involved in the biology of circadian rhythms [2], is expressed ubiquitously. This protein is one of the several brain-expressed genes located on chromosome 18 with a higher expression in the brain and spinal cord than in other tissues (www.genecards.org). The *PHLPP1* protein (NP\_919431.1), with the total length of 1205 residues consists of a pleckstrin homology (PH) domain, a leucine rich repeat (LRR) domain, a protein phosphatase 2C (PP2C)-like domain followed by a polyglutamine repeat sequence (PolyQ) at the carboxy terminus.

Structural analysis of the protein suggests that LRR domain and PH domain are involved in the mechanism of circadian oscillation through membrane targeting and Ras (low molecular weight globular (G) protein) mediates signal transduction [3]. Recent studies implicate the *PHLPP1* protein in the negative regulation of MAP (Mitogen-activated protein) kinases in memory formation in the hypothalamus [4]. This protein causes termination of Akt (protein kinase B) signaling by dephosphorylation of particular amino acid residues in the specific Akt isoforms [5].

CAG stretches beyond six repeats have been found to be polymorphic, leading to disease status in many neurological disorders [6]. Variations in CAG repeats on chromosome 18 have been previously investigated in psychiatric disorders [7]. The correlation between invariant CAG stretches in coding parts of the genome to disease phenotype is not well understood. Considering the importance of CAG repeat length in several neurological and psychiatric disorders, an analysis from an evolutionary standpoint may elucidate the role of these CAG repeats and the gene function. Here, we present the cross-species analysis of the CAG repeat length of the *PHLPP1* gene to look for clues to the molecular evolution and diversity. We examined CAG repeat length by two complementary approaches: We first sequenced the CAG repeat region of *PHLPP1* gene in 26 human subjects (see Methodology), and 5 primates, and subsequently performed a cross-species and comparative analysis of CAG repeat length with the sequenced subjects and other species. We observed that polyglutamine stretch in this gene is not polymorphic in humans. However, a considerable variation is seen in the repeat lengths in non-human primates and in other species that were analyzed.

### Methodology:

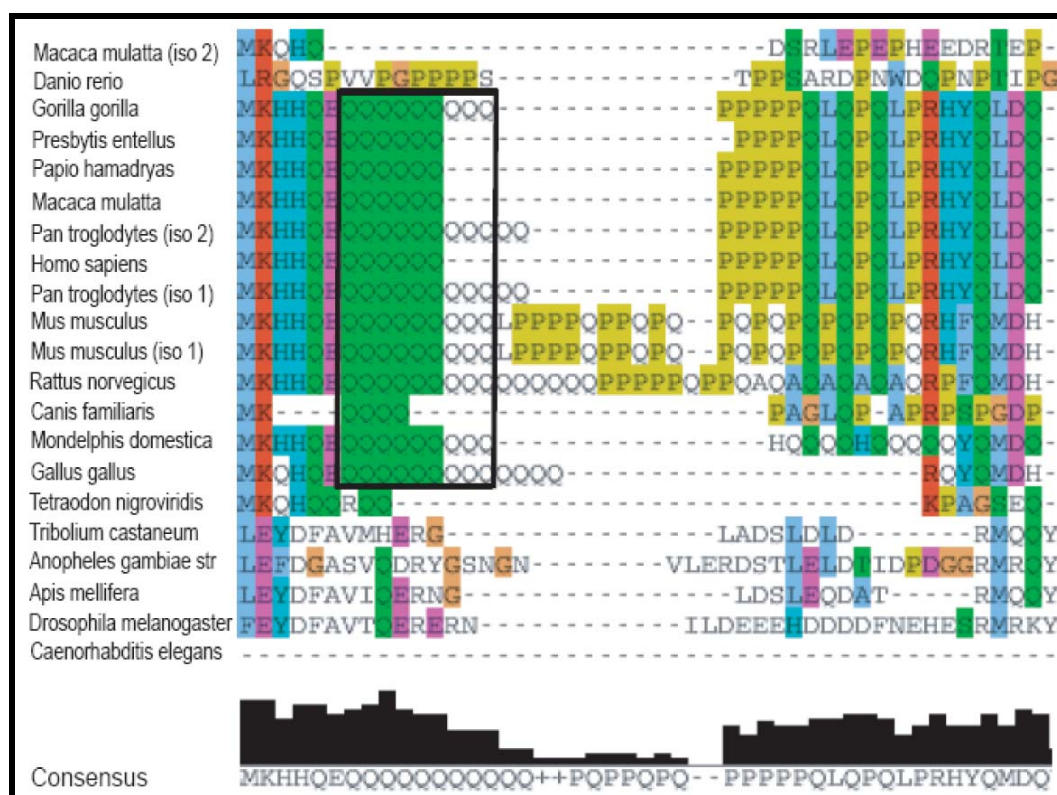
#### Subjects

We sequenced DNA from 26 human subjects (BPAD probands N=16; control subjects (N=10), after due consent, as part of an IRB approved study of the genetics of bipolar disorder. Genomic DNA of the non-human primates, Chimpanzee (*Pan*

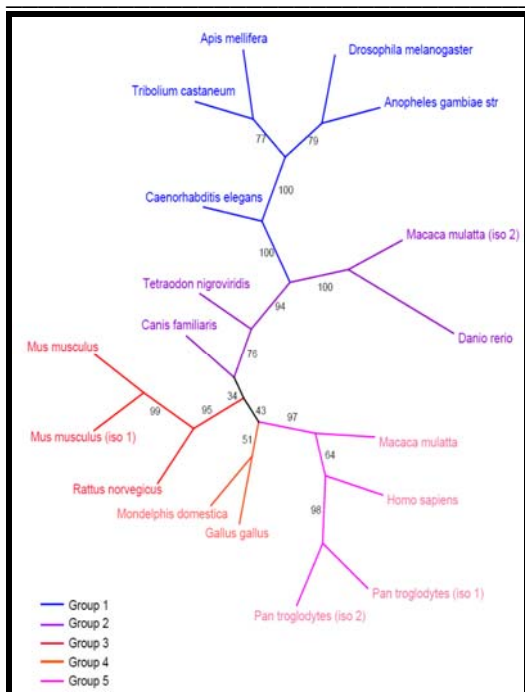
*troglydotes*), Gorilla (*Gorilla gorilla*), Baboon (*Papio hamadryas*), Rhesus monkey (*Macaca mulatta*) and Langur (*Presbytis entellus*) was obtained from the Institute of Genomics and Integrative Biology (IGIB), Mall road, New Delhi. The human *PHLPP1* gene sequence (Genbank accession no-NC\_000018) was retrieved from the Entrez nucleotide query at the National Centre for Biotechnology Information, Bethesda, Md (<http://www.ncbi.nlm.nih.gov>). Genomic DNA was amplified by PCR reactions to amplify the CAG repeat region of the gene using standard protocol. Sequencing was performed and products analyzed on an ABI-377 automated sequencer using appropriate firmware softwares.

Orthologs of the gene (NP\_919431) containing the polyglutamine stretch were obtained from non-redundant (nr) database through PSI-BLAST search [8]. Multiple sequence alignment (MSA) of the orthologs and the genomes under study was aligned using Clustal W 1.83 [9]. The alignment was manually adjusted using Jalview 2.08.1 [10] (Figure 1). A phylogenetic analysis using Maximum likelihood (ML) method tree construction was carried using PHYLIP package V 3.65 [11]. The input alignment file was bootstrapped 100 times using SEQBOOT [11] with no randomization of sequence order, the tree topology was obtained using maximum likelihood method using PROML [11]. Consensus tree was obtained from 100 maximum likelihood trees using CONSENSE [11]. TreeView [12] was used for generating the tree (Figure 2).

### Sequence alignment and phylogenetic analysis



**Figure 1:** ClustalW alignment of the *PHLPP1* sequences showing variations of PolyQ repeats across species, with a conserved core of 6 glutamine repeats. (Sequences shown in the alignment: *Macaca mulatta* (iso 2) (XP\_001105985.1), *Danio rerio* (XP\_700869.2), *Gorilla gorilla*<sup>#</sup>, *Presbytis entellus*<sup>#</sup>, *Papio hamadryas*<sup>#</sup>, *Macaca mulatta* (XP\_001094216.1), *Pan troglodytes* (iso2) (XP\_523952.2), *Homo sapiens* (NP\_919431.1), *Pan troglodytes* (iso 1) (XP\_001145310.1), *Mus musculus* (XP\_990044.1), *Mus musculus* (iso 1) (XP\_129968.3), *Rattus norvegicus* (NP\_067689.1), *Canis familiaris* (XP\_533384.2), *Mondelphis domestica* (XP\_001375913.1), *Gallus gallus* (XP\_418990.2), *Tetraodon nigroviridis* (CAF91595.1), *Tribolium castaneum* (XP\_973398.1), *Anopheles gambiae* str (XP\_319835.3), *Apis mellifera* (XP\_394402.3), *Drosophila melanogaster* (NP\_609938.1), *Caenorhabditis elegans* (NP\_497845.2). <sup>#</sup> - From the DNA sequence of the organism).



**Figure 2:** Unrooted Phylogenetic tree of the *PHLPP1* sequences. Phylogenetic analysis shows distribution of the species into five groups. Bootstrap values are shown.

### Results:

A high degree of conservation at various residues, flanking the consensus polyglutamine sequences, was evident across the 18 species that were analyzed. Sequencing of the region in humans showed an uninterrupted motif of 6 CAG repeats. In the other non-human primates, there was an uninterrupted CAG repeat in langur, baboon and rhesus monkey of the same length as the humans, whereas it was longer in gorilla and chimpanzee (Figure 1).

The homologous protein sequences, obtained from NCBI, revealed that polyglutamine motif was also found in chicken and rodents like mouse and opossum (Figure 1). In these species, the repeat length was variable, and had non-glutamine residue interruptions when compared to the human sequence. The shortest stretch of polyglutamine with 4 repeats was found in the dog, while chicken had 13 repeats. This phenomenon is similar to the intraspecies conservation, but inter-species variability in the coding region at the CAG locus, observed in the circadian genes *Clock* [13] and the Variable Number Tandem Repeat (VNTR) in the *hper3* gene [14].

For many CAG repeat associated diseases such as SCA1, spinocerebellar ataxia 1 protein, or ataxin1; SCA2, spinocerebellar ataxia 2 protein; SCA7, spinocerebellar ataxia 7 protein; SCA17, spinocerebellar ataxia 17 protein; HD, Huntington's disease protein, or huntingtin, the corresponding repeat is much smaller in rodents [15]. A recent study of CAG repeat stretch to characterize the nature of polyglutamine repeat length variation across the human genome has been

conducted to establish the background against which pathogenic repeat expansions can be detected, and to prioritize candidate genes for repeat expansion disorders [16]. The simultaneous occurrence of considerable variation in repeat lengths suggests that this variation, or the lack of it, could be important for the structure of the functional protein, and thus to the physiological function and pathology.

Several motif-based sensitive sequence searches using the PSI-BLAST [8] revealed that there are no direct homologues of the human *PHLPP1* protein in insects and fish. The polyglutamine stretch as well as the PH domain is thus completely absent in some species including worm, fish and insects. The repeat motifs are sometimes lost in paralogous proteins in other species, which might indicate that these repeat motifs are functionally less important for these species [17].

Comparison of the sequence lengths at this locus in different species shows considerable variation and the phylogenetic analysis using the complete protein sequence showed five different groups (Figure 2). The *PHLPP1* protein does not harbor the PAS (Per-Arnt-Sim) domain found in the other circadian genes such as the *per*, *bmal* and *clock* which is known to aid in heterodimer formation. Unlike other aspects of the honeybee circadian rhythm, which appears to have evolved away from the *Drosophila* (*Drosophila melanogaster*) and towards the mouse system [18], this gene does not show any increased similarity in its carboxy terminal end to the mouse sequence.

**Discussion:**

The role of trinucleotide repeats (TNR), especially those in brain-expressed genes, are important from a medical and evolutionary perspective. Our findings suggest that the polyglutamine stretch in *PHLPP1* gene is not polymorphic in humans. However, it exhibits varying repeat lengths in non-human primates and in other species.

Circadian behaviors in the animal kingdom are regulated by a set of conserved genes. Over the course of evolution, there has been a substantial reshuffling of specific functions between structurally homologous components of circadian clocks [19]. Comparative analysis of various circadian genes suggests that despite the highly conserved nature of their individual components, the hierarchical model of mammalian circadian-clock system does not exist in insects like fruit fly, beetle, mosquito and honey bee [20, 21], and vertebrates like fish (zebrafish and puffer fish) [22]. It is likely that variations in certain components of the circadian pathways would be contingent upon the physiological systems and environment of the species. Thus, in species such as insects, reptiles and fish, which lack thermoregulatory feedback correction, an integral part of circadian regulation would be expected to differ from mammalian vertebrates.

Extrapolations in circadian biology need to be made with caution, after understanding of the underlying physiological substrates. It is possible that the aspect of circadian rhythms operated by these parts of the *PHLPP1* protein either have not evolved in these organisms or they could be executed by a different set of proteins distinct from the human pathway.

**Conclusion:**

In summary, *PHLPP1* gene was studied for variation in human and non-human primates, and compared with other species. The CAG repeat length was non-polymorphic in humans. Both shorter and longer CAG lengths were found to be present in non-human primates and in the other species studied. A recent compilation of all proteins encoding four or more amino acid repeats show that they are conserved across vertebrates; however, the nucleotide repeat motif itself is not always completely conserved. Our study indicates a similar species-specific variation in repeat length is observed for the glutamine repeat in the *PHLPP1* protein which might be governed by the role of the protein in the particular species. Our study underlines the importance of this gene in circadian biology based on its presence across a wide range of species. Though this preliminary study showed an apparent monomorphic nature for the CAG repeat in this gene, future studies might reveal clues for a role of *PHLPP1* gene in mood disorders which are associated with disturbances in circadian rhythms.

**Acknowledgement:**

We thank the subjects and controls for participation in the study. We thank the Director, Institute of Genomics and Integrative biology, New Delhi, India for providing the non-Human primate samples. This work was supported by research grants from the Department of Biotechnology (DBT), Government of India, New Delhi. We acknowledge Mr. Kumar Gaurav and Mr. K. Kanagarajadurai from Dr. R. Sowdhamini's lab, NCBS for their assistance.

**References:**

- [01] S. N. Ghaemi, *Schizophr Bull.*, 33: 122 (2007) [PMID: 17122410]
- [02] K. Shimizu, *et al.*, *FEBS Lett.*, 458: 363 (1999) [PMID: 10570941]
- [03] K. Shimizu, *et al.*, *J Biol Chem.*, 278: 14920 (2003) [PMID: 12594205]
- [04] K. Shimizu, *et al.*, *Cell*, 128: 1219 (2007) [PMID: 17382888]
- [05] J. Brognard, *et al.*, *Mol Cell.*, 25: 917 (2007) [PMID: 17386267]
- [06] C. J. Cummings and H. Y. Zoghbi, *Annu Rev Genomics Hum Genet.*, 1: 281 (2000) [PMID: 11701632]
- [07] T. Swift *et al.*, *Psychiatr Genet.*, 15: 91(2005) [PMID: 15900223]
- [08] S. F. Altschul, *et al.*, *Nucleic Acids Res.*, 25: 3389 (1997) [PMID: 9254694]
- [09] J. D. Thompson, *et al.*, *Nucleic Acids Res.*, 22: 4673 (1994) [PMID: 7984417]
- [10] M. Clamp, *et al.*, *Bioinformatics*, 20: 426 (2004) [PMID: 14960472]
- [11] J. Felsenstein, *Cladistics*, 5: 164 (1989)
- [12] R. D. Page, *Comput Appl Biosci.*, 12: 357 (1996) [PMID: 8902363]
- [13] Q. Saleem, *et al.*, *Hum Genet.*, 109: 136 (2001) [PMID: 11511917]
- [14] A. Jenkins, *et al.*, *J Biol Rhythms.*, 20: 470 (2005) [PMID: 16267386]
- [15] H. Huang, *et al.*, *Genome Biol.*, 5: R47 (2004) [PMID: 15239832]
- [16] S. L. Butland, *et al.*, *BMC Genomics*, 8: 126 (2007) [PMID: 17519034]
- [17] P. Siwach, *et al.*, *Mol Biol Evol.*, 23:1357 (2006) [PMID: 16618963]
- [18] E. B. Rubin, *et al.*, *Genome Res.*, 16: 1352 (2006) [PMID: 17065608]
- [19] M. D. Field, *et al.*, *Neuron*, 25: 437 (2000) [PMID: 10719897]
- [20] E. Rosato and C. P. Kyriacou, *Philos Trans R Soc Lond B Biol Sci.*, 356: 1769 (2001) [PMID: 11710984]
- [21] Y. Xu, *et al.*, *Nature*, 434: 640 (2005) [PMID: 15800623]
- [22] D. Bell-Pedersen, *et al.*, *Nat Rev Genet.*, 6: 544 (2005) [PMID: 15951747]

Edited by P. Kanguane

Citation: Anbazhagan *et al.*, *Bioinformatics* 2(10): 471-474 (2008)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.