Phylogenetic and Chronological Analysis of Proteins Causing Alzheimer's, Parkinson's and Huntington's Diseases

Bilal Hussain[1*], Hina Khalid[1], Shahid Nadeem[1], Tayyaba Sultana[2], Saima Aslam[3]

[1]*Department of Bioinformatics and Biotechnology*
GC University, Faisalabad-38000, Pakistan
E-mails: profbilal@yahoo.com, hina_huny24@yahoo.com, snadeem63@yahoo.com

[2]*Department of Zoology*
GC University, Faisalabad-38000, Pakistan
E-mail: arif143@yahoo.com

[3]*Department of Computer Science*
University of Agriculture, Faisalabad-38000, Pakistan

[*]*Corresponding author*

Received: June 4, 2012 **Accepted: September 12, 2012**

Published: September 15, 2012

Abstract: *It is evident that Neurodegenerative diseases (Alzheimer's, Parkinson's and Huntington's) have many similarities at cellular and molecular level as they carry parallel mechanisms including protein aggregation and inclusion body formation caused by protein mis-folding. The main objective of this study was to have detailed insight on variation and resemblance among these proteins. One hundred and four protein sequences, both directly and indirectly involved in disease mechanism to perform phylogenetic analysis revealing insight on evolutionary relationship among these proteins, were selected. The percentage of replicate trees, in which the associated taxa clustered together in the bootstrap test, was 1000 replicates. Various statistical tests were performed for the confirmation of results e.g., Tajma's Neutrality Test showed* $D > 6$*, nucleotide diversity* $\pi > 0.6$ *and ps value as greater than 1. Phylogenetic analysis showed that the protein sequences of neurodegenerative diseases had high sequence similarity and identity to each other as depicted by the evolutionary tree. It showed the similar mechanism of evolving from each other and had similar mechanism of generating mis-folding leading towards symptoms of disease.*

Keywords: *Phylogenetics, Alzheimer's, Parkinson's, Huntington's diseases.*

Introduction

Several neurodegenerative diseases are believed to emerge from the accumulation of mis-fold (incorrectly folded) proteins. Aggregated proteins are associated with such as amyloid-related illnesses such as Alzheimer's disease and aggregation diseases Huntington's and Parkinson's disease. Parkinson's disease (PD) is a chronic progressive neurodegenerative movement disorder characterized by a profound and selective loss of nigrostriatal dopaminergic neurons. Generally, mutations in the LRRK2, PARK2, PARK7, PINK1, and SNCA genes cause PD. Alzheimer's disease is a degenerative disease of the brain and causes dementia, which is a gradual loss of memory, judgment, and ability to function and characterized by cognitive impairment, progressive neurodegeneration and formation of amyloid-β (Aβ)-containing plaques and neurofibrillary tangles composed of hyperphosphorylated taut. Huntington's disease is a progressive brain disorder caused by mutations in the HTT gene. The HTT mutation involves a DNA segment known as a CAG trinucleotide repeat. It results into

uncontrolled movement and cognitive function begins to deteriorate into dementia. Mutations in the PTEN-induced putative-kinase-1 (PINK1 a mitochondrial serine–threonine kinase), and Parkin, (an E3 ubiquitin ligase), are associated with autosomal-recessive forms of PD [1].

Mutations in DJ-1 cause recessively transmitted early-onset PD, and oxidative damage to DJ-1 has been associated with the pathogenesis of late-onset sporadic PD. Moreover, it will be tried to find out that the C-terminally cleaved form of DJ-1 with activated protease function exhibits enhanced cytoprotective action against oxidative stress-induced apoptosis [2, 10]. The aim of this study was to asses the evolutionary relationship among the proteins directly or indirectly, involved in these three neurodegenerative diseases (Parkinson's, Huntington's and Alzheimer's disease) by using 32, 24 and 48 proteins, respectively.

Materials and methods

Sequence retrieval and analysis

To find similarity or diversity, we retrieved sequences from UniProt which are directly or indirectly involved in Parkinson's, Huntington's and Alzheimer's disease causing mechanism. Total of 104 sequences were used for analysis which were reviewed and whose domain units are with known functions. Among the selected proteins are E3 ubiquitin-protein ligase parkin (O60260), Protein DJ-1 (Q99497), Parkinson disease 7 domain-containing protein 1 (Q8NB37), Glucosylceramidase (P04062), Serine/threonine-protein kinase PINK1, mitochondrial (Q9BXM7), Serine protease HTRA2, mitochondrial (O43464), Synphilin-1 (Q9Y6H5), PERQ amino acid-rich with GYF domain-containing protein 2 (Q6Y7W6), 5'-AMP-activated protein kinase subunit gamma-2 (Q9UGJ0), Ceruloplasmin (P00450), Transitional endoplasmic reticulum ATPase (P55072), Charged multivesicular body protein 2b (Q9UQN3), DNA damage-inducible transcript 4 protein (Q9NX09), RING finger protein 11 (Q9Y3C5), F-box only protein 7 (Q9Y3I1), Brain-derived neurotrophic factor (P23560), Probable G-protein coupled receptor 37 (O15354), Tyrosine 3-monooxygenase (P07101), Thrombomodulin (P07204), Caveolin-1 (Q03135), Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 1 (Q96FE5), Parkin coregulated gene protein (Q96M98), E3 ubiquitin-protein ligase RNF19A (Q9NV58), Whirlin (Q9P202), Forkhead box protein L2 (P58012), NADPH-cytochrome P450 reductase (P16435), Septin-5 (Q99719), Regulator of nonsense transcripts 3B (Q9BZI7), NADH-ubiquinone oxidoreductase chain 6 (P03923), Complexin-1 (O14810), Synaptotagmin-11 (Q9BT88), Nuclear receptor subfamily 4 group A member 2 (P43354) for Parkinson's disease. We also retrieved proteins of Huntington's disease, namely Huntingtin (P42858), Major prion protein (P04156), Junctophilin-3 (Q8WXH2), Histone-lysine N-methyltransferase SETDB1 (Q15047), Cdc42-interacting protein (Q15642), Antithrombin-III (P01008), Huntingtin-interacting protein 1 (O00291), Histone-lysine N-methyltransferase SETD2 (Q9BYW2), Transcription elongation regulator 1 (O14776), Fibroblast growth factor receptor 3 (P22607), Prothrombin (P00734), Alpha-1-antitrypsin (P01009), Transcription initiation factor TFIID (P216750, Huntingtin-associated protein 1 (P54257), SH3 domain-binding protein (P78314), Palmitoyltransferase ZDHHC17 (Q8IUH5), Caspase recruitment domain-containing protein (Q5EG05), Pre-mRNA-processing factor 40 homolog B (Q6NWY9), Heparin cofactor 2 (P05546), Nucleolar protein 14 (P78316), Complexin-2 (Q6PUV4), Alpha-adducin (P35611), RING finger protein 4 (P78317), Capucin (A6NDD5).

Moreover, proteins causing Alzhemeir's disease were also retrieved, namely Amyloid beta A4 protein (P05067), Alpha-1-antichymotrypsin (P01011), Protein AATF (Q9NY61), Abl interactor 1 (Q8IZP0), Acyl-coenzyme A thioesterase 2, mitochondrial (P49753), Amyloid-like protein 1 (P51693), Apolipoprotein E (P02649), A disintegrin and

metalloproteinase with thrombospondin motifs (O75173), Beta-secretase 1 (P56817), Beta-secretase 2 (Q9Y5Z0), Nucleosome-remodeling factor subunit BPTF (Q12830), Cathepsin D (P07339), Cannabinoid receptor 2 (P34972), Collectin-12 (Q5KU26), Collagen alpha-1(XXV) chain (Q9BXS0), Calsenilin (Q9Y2W7), Calsyntenin-1 (O94985), Complement decay-accelerating factor (P08174), 24-dehydrocholesterol reductase (Q15392), Dickkopf-related protein 1 (O94907), Dickkopf-related protein 2 (Q9UBU2), Dickkopf-related protein 3 (Q9UBP4), Dickkopf-related protein 4 (Q9UBT3), Dedicator of cytokinesis protein 3 (Q8IZD9), Endothelin-converting enzyme 2 (O60344), Translation initiation factor eIF-2B subunit beta (P49770), Filamin-B (O75369), G-protein coupled receptor 3 (P46089), Histamine H2 receptor (P25021), C-Jun-amino-terminal kinase-interacting protein 1 (Q9UQF2), Kallikrein-6 (Q92876), Kallikrein-8 (Q5S007), Leucine-rich repeat serine/threonine-protein kinase 2 (O60259), Myc-associated zinc finger protein (P56270), Metallothionein-2 (P02795), Metallothionein-3 (P25713), Napsin-A (O96009), Collectin-12 (Q5KU26), Ubiquitin carboxyl-terminal hydrolase isozyme (P15374), Ubiquitin carboxyl-terminal hydrolase isozyme (P09936), Visinin-like protein 1 (P62760), Alpha-synuclein (P37840), Beta-synuclein (Q16143), Homocysteine-responsive endoplasmic reticulum-resident ubiquitin-like domain member 1 protein (Q15011), RNA-binding protein 25 (P49756), Protein numb homolog (P49757), RAB GTPase-activating protein 1-like (Q5R372).

Multiple sequence alignment

Multiple sequence alignment was carried out using graphical Custal X user interface version of Clustal. Multiple sequence alignment was performed using protein matrix Blosum with gap penalty and gap extension penalty set to 10 and 0.2 respectively.

Evolutionary relationships analysis

Phylogenetic trees were constructed through MEGA version 5 by the Neighbor-Joining method [9]. The bootstrap consensus trees were inferred from 1000 replicates, the evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated.

Neutrality analysis

The Tajima test statistics was estimated through MEGA version 5. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). The abbreviations used are as follows: m – number of sites, a_i – quadrate components, S – number of segregating sites, $p_s = S/m$, $\Theta = p_s/a_1$, and π – nucleotide diversity. D is the Tajima test statistic [9].

Results

Phylogenetic inference

The Neighbor-Joining method used for phylogenetic inference was applied to the topologies. The reliability of branch length estimates was tested by Bootstrap method [4]. Felsenstein's Bootstrap test is one of the most commonly used tests to check the reliability of an inferred tree. In order to statistically evaluate the accuracy and reliability of our phylogenetic tree, Tajima's Neutrality Test (TNT) was applied. Figs. 1-8 illustrate the phylogenetic relationships and TNT among 104 selected proteins [9].

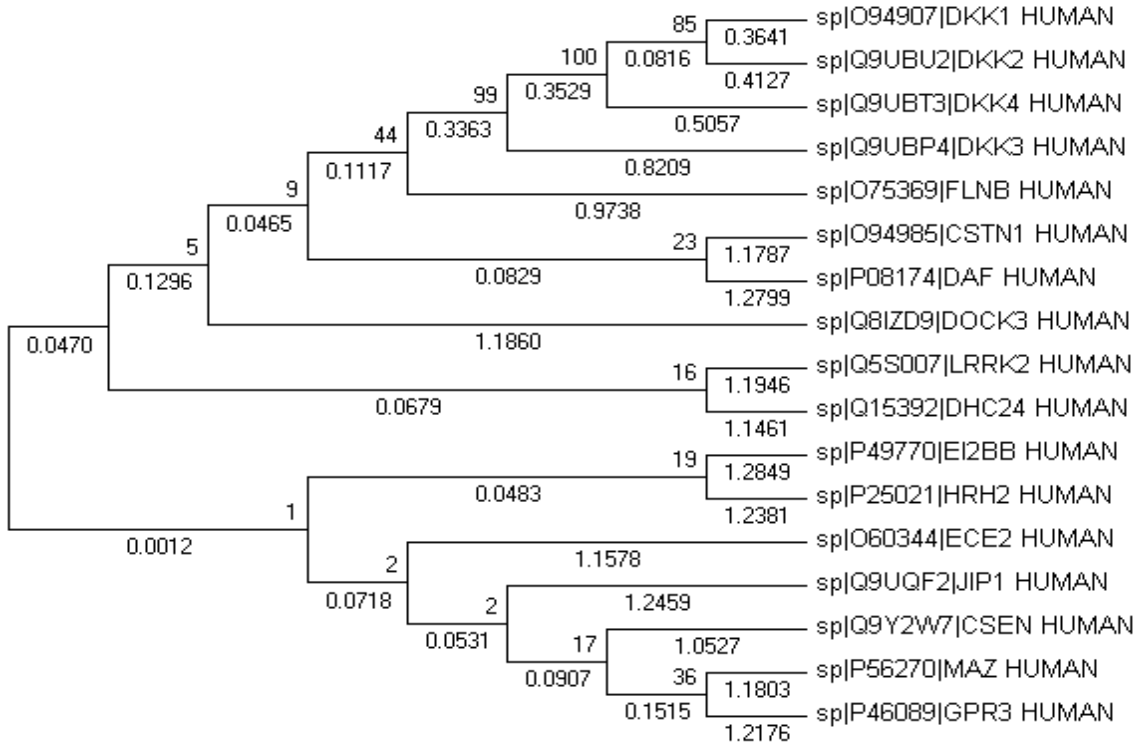


Fig. 1 Evolutionary relationships among 17 taxa of Alzheimer's disease-causing proteins;
a phylogram of 17 selected Alzheimer's disease-causing proteins indicating the relationship
between the proteins as well as conveying a sense of time or rate of evolution;
the optimal tree with the sum of branch length = 19.11279952;
all positions containing gaps and missing data were eliminated; there were 187 positions.

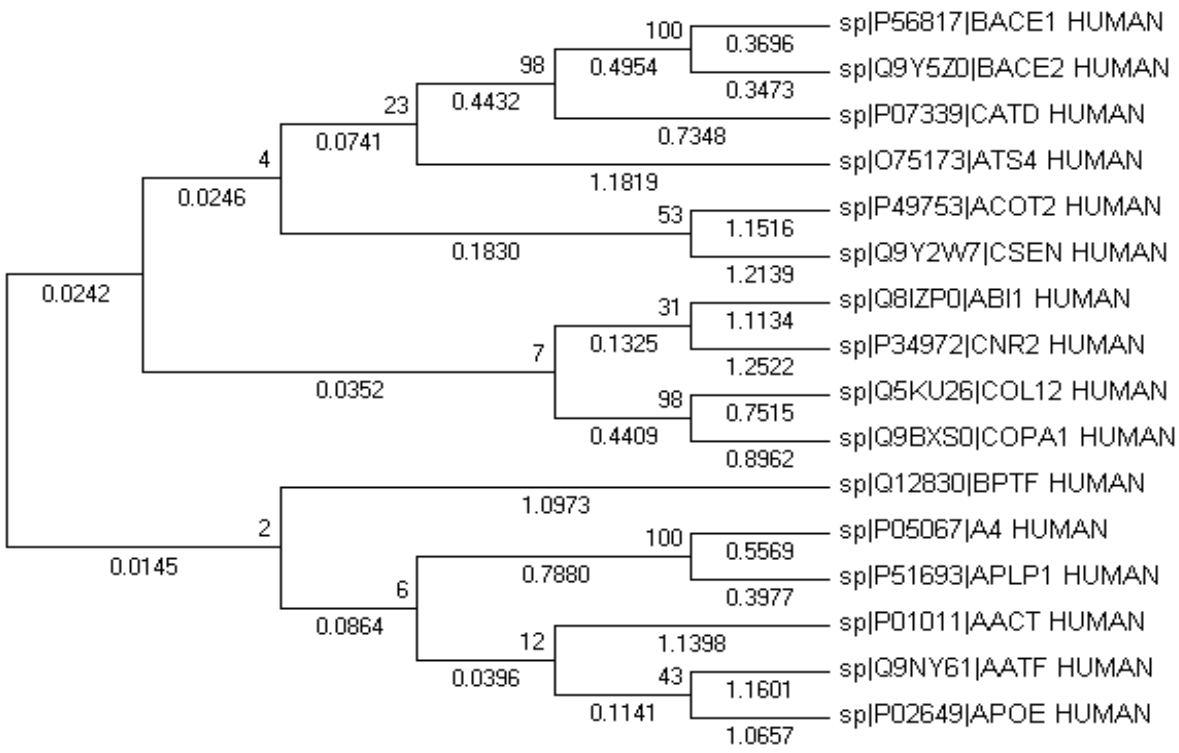


Fig. 2 Evolutionary relationships among 16 taxa of Alzheimer's disease-causing proteins;
a phylogram of 16 selected of Alzheimer's disease-causing proteins indicating the relationship between the proteins as well as conveying a sense of time or rate of evolution;
the optimal tree with the sum of branch length = 17.32545732 is shown;
all positions containing gaps and missing data were eliminated;
there were 213 positions in the final dataset.

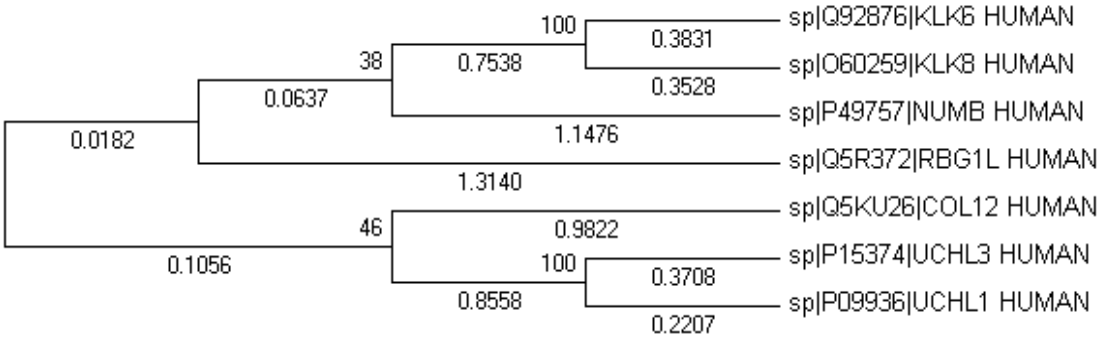


Fig. 3 Evolutionary relationships among 7 taxa of Alzheimer's disease-causing proteins; a phylogram of 7 selected of Alzheimer's disease-causing proteins indicating the relationship between the proteins as well as conveys a sense of time or rate of evolution; the optimal tree with the sum of branch length = 6.56816110 is shown; all positions containing gaps and missing data were eliminated; there were a total of 215 positions in the final dataset.

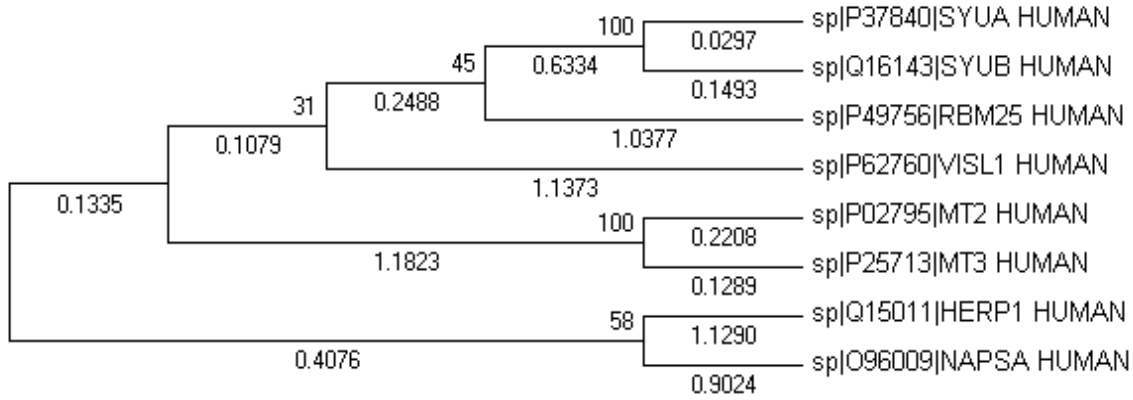


Fig. 4 Evolutionary relationships among 8 taxa of Alzheimer's disease-causing proteins; a phylogram of 8 selected of Alzheimer's disease-causing proteins indicating the relationship between the proteins and also conveying the sense of time or rate of evolution; the optimal tree with the sum of branch length = 7.44873208 is shown with 61 positions in the final dataset.

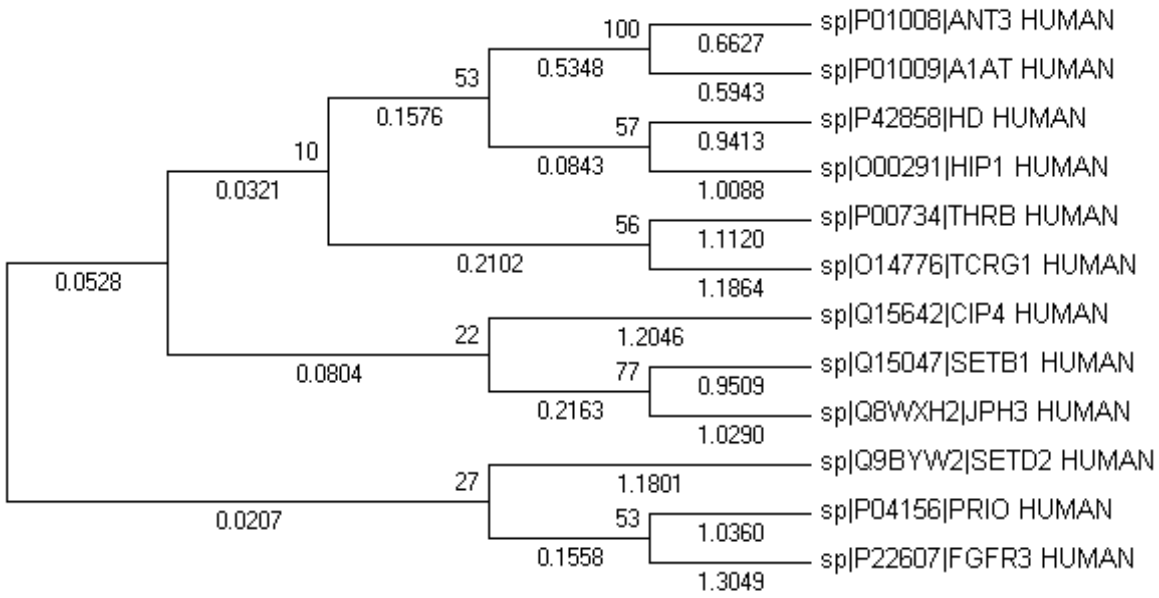


Fig. 5 Evolutionary relationships among 12 taxa of Huntington's disease-causing proteins; a phylogram of 12 selected of Huntington's disease-causing proteins those indicates the relationship between the proteins and also conveying a sense of time or rate of evolution; the optimal tree with the sum of branch length = 13.75600116 is shown; all positions containing gaps and missing data were eliminated; there were 239 positions in the final dataset.

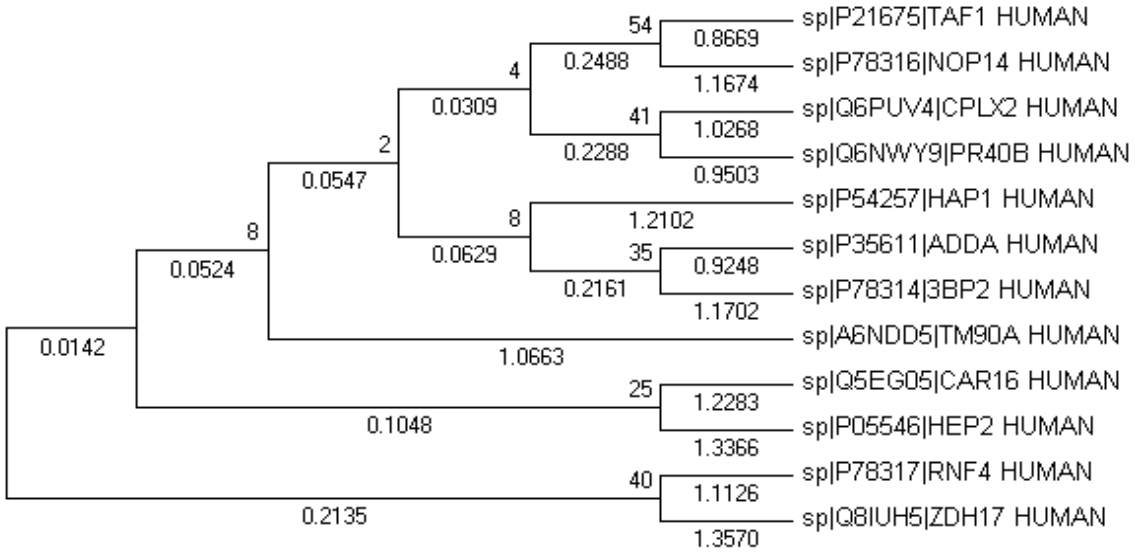


Fig. 6 Evolutionary relationships among 12 taxa of Huntington's disease causing proteins;
a phylogram of 12 selected of Huntington's disease-causing proteins indicating
the relationship between the proteins and provide rate of evolution;
the optimal tree with the sum of branch length is 14.64463276;
all positions containing gaps and missing data were eliminated;
there were 130 positions in the final dataset.

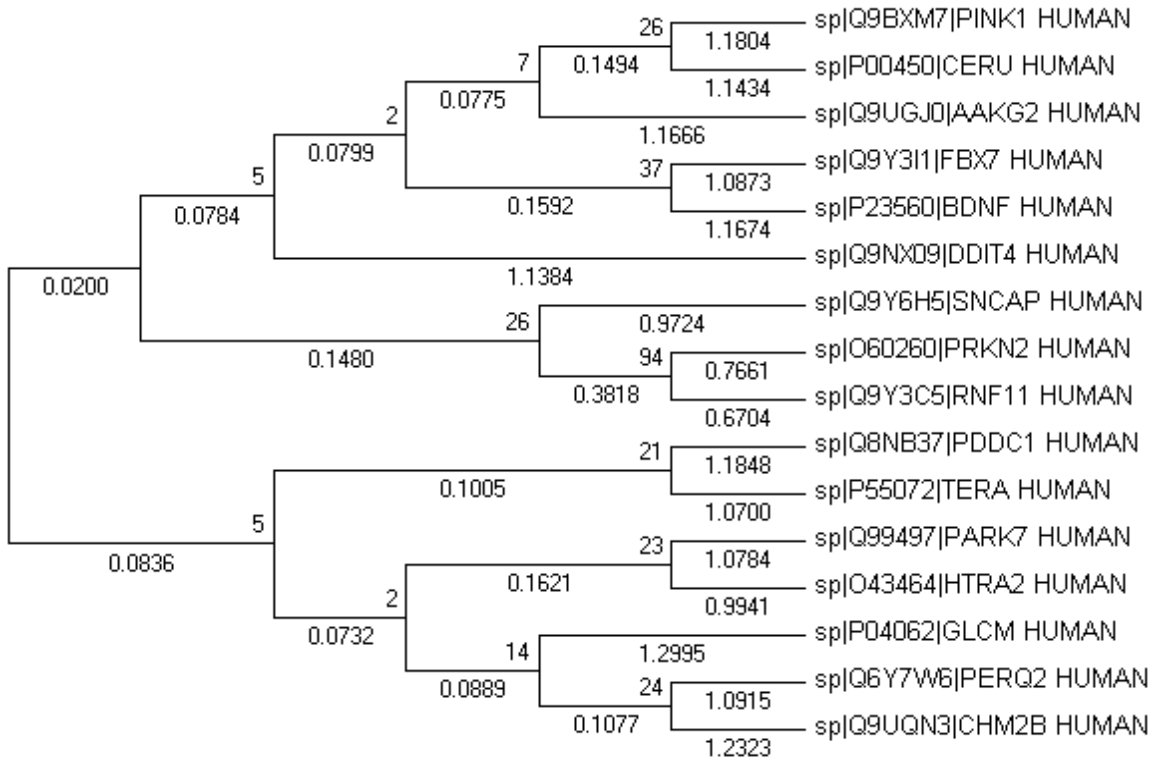


Fig. 7 Evolutionary relationships among 16 taxa of Parkinson's disease-causing proteins; a phylogram of 16 selected of Parkinson's disease-causing proteins indicating the relationship between the proteins and also conveys a sense of time or rate of evolution; the optimal tree with the sum of branch length = 18.95340502 is shown; there were 143 positions in the final dataset.

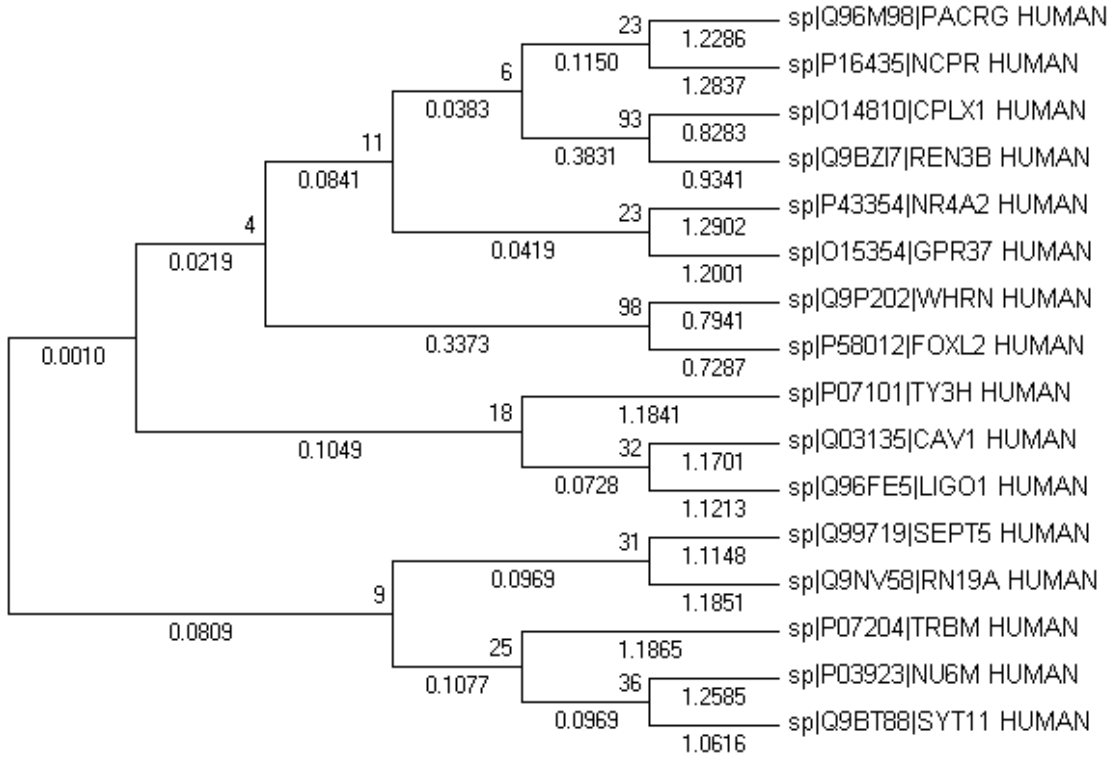


Fig. 8 Evolutionary relationships among 16 taxa of Parkinson's disease-causing proteins; a phylogram of 16 selected of Parkinson's disease-causing proteins indicating the relationship between the proteins and provide a sense of time or rate of evolution; branch length of tree is 19.15230121; all ambiguous positions were removed for each sequence pair with 915 positions in final dataset.

Discussion

Neurodegenerative diseases have common mechanism of protein mis-folding, cellular dysfunction and neurodegeneration. Olzscha *et al.* stated that "Not all the proteins are affected by aggregation", [6], "especially those proteins are susceptible, which possess specific structural characteristics and are involved in important biological processes". This phylogenetic analysis done for various proteins revealed genetic divergence into distinct lineages which indicated that majority of the proteins involved, were evolving in a similar manner regarding time scale. The misfolding and progressive aggregations of specific proteins are seminal occurrence in many neurodegenerative diseases. Results corroborated Jellinger findings for providing accumulative evidence of overlap between synucleinopathies, tauopathies and other protein-misfolding diseases [4]. This suggested that prion-like induction and spreading, involving secreted protein were the major pathogenic mechanism involved in neurodegenerative diseases.

Proteins with specific structural characteristics and significant biological functions are more liable to be influenced by aggregation [6]. Here computational estimation revealed significant prevalence of intrinsically unstructured protein in these three diseases, which might contribute to disease pathogenecity. Therefore, the interaction network of these proteins is large and complex. The majority of these proteins are associated with biological processes like neuronal activities, signal transduction, apoptosis, intracellular trafficking and cell differentiation [8]. From the present study, information, added to what was already known, brought the total number of Alzheimer's-related interactions up to 6000, involving 1700 proteins, resulting into the largest network of interactions between proteins related to Alzheimer's disease [7].

Recently, studies revealed [5] that a variant in the NEDD9 gene might have been the common genetic factors. We attempted to confirm this initial observation by conducting a similar analysis in terms of pathologies and sample size. We genotyped the NEDD9 rs760678 SNP in three independent AD case-control studies (n = 3176) and two independent PD case-control studies (n = 1855). However, we could not find any association of this SNP with the risk of developing AD or PD, in any of these populations. Concluding, these data indicate that the rs760678 SNP of the NEDD9 gene is a weakest genetic determinant of AD or PD and these findings are quite inline with those of Julien *et al.* studies [5]. According to phylogenetic analysis, the evolutionary tree indicates the closely related proteins, as well as distant and distinct sequences. The inference of phylogeny is an essential factor for understanding biological variation and similarities among various proteins involved in neurodegenerative diseases. Furthermore, this hierarchical relationship would contribute to further studies on their pathogenesis and therapeutic aspects.

Conclusion

Higher bootstrap values convey evolutionary sense among proteins causing Huntinton's, Parkinson's and Alzheimer's diseases. The present study showed that proteins involved in development and pathway of Huntinton's, Parkinson's and Alzheimer's disease are closely related to each other. The study will be useful for knowing the taxonomy and evolution of proteins participating in neurodegenerative diseases.

References

1. Aleksandar R., P. Seibler, A. Ramirez, C. Klein (2010). Effect of Endogenous Mutant and Wild-type PINK1 on Parkin in Fibroblasts from Parkinson Disease Patients, Hum. Mol. Genet., 19(16), 3124-3137.
2. Chen J., L. Lian, L. S. Chin (2010). Parkinson Disease Protein DJ-1 Converts from a Zymogen to a Protease by Carboxyl-Terminal Cleavage, Hum. Mol. Genet., 19(12), 2395-2408.
3. Jain E., A. Bairoch, S. Duvaud, I. Phan, N. Redaschi, B. E. Suzek, M. J. Martin, P. McGarvey, E. Gasteiger (2009). Infrastructure for the Life Sciences: Design and Implementation of the UniProt Website, BMC Bioinformatics, 10:136, doi:10.1186/1471-2105-10-136.
4. Jellinger K. A. (2012). Interaction between Pathogenic Proteins in Neurodegenerative Disorders, J. Cell. Mol. Med., 16(6), 1166-1183.
5. Julien C., F. Moisan, G. Mellick, A. Elbaz, P. Silburn, C. Lendon (2008). Association Study of the NEDD9 Gene with the Risk of Developing Alzheimer's and Parkinson's Disease, Hum. Mol. Genet., 17(18), 2863-2867.
6. Olzscha H., S. M. Schermann, A. C. Woerner, S. Pinkert, M. H. Hecht, G. G. Tartaglia, M. Vendruscolo, M. Hayer-Hartl, F. U. Hartl, R. M. Vabulas (2011). Amyloid-like

Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions, Cell, 144(1), 67-78.
7. Panjkovich A., P. Aloy (2010). Predicting Protein-protein Interaction Specificity through the Integration of Three-dimensional Structural Information and the Evolutionary Record of Protein Domains, Mol. BioSyst., 6, 741-749.
8. Paychaudhuri S., S. Dey, N. P. Bhattacharyya, D. Mukhopadhyay (2009). The Role of Intrinsically Unstructured Proteins in Neurodegenerative Diseases, PLoS One, 4(5), e5566.
9. Shelbourne P. F., C. Keller-McGandy, W. L. Bi, S. R. Yoon, L. Dubeau, N. J. Veitch, J. P. Vonsattel, N. S. Wexler (2007). Triplet Repeat Mutation Length Gains Correlate with Cell-type Specific Vulnerability in Huntington Disease Brain, Hum. Mol. Genet., 16(10), 1133-1142.
10. Tamura K., D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods, Mol. Biol. Evol., 28(10), 2731-2739.
11. Thompson J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, D. G. Higgins (1997). The ClustalX Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools, Nucl. Acids. Res., 25, 4876-4882.

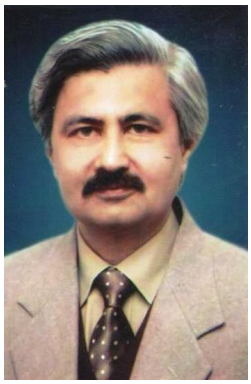
Bilal Hussain, M.Phil., Ph.D. Scholar
E-mail: profbilal@yahoo.com

Bilal Hussain is a Ph.D. Scholar from the Government College University – Faisalabad with the distinction of Bronze Medal in the session. Since 2003 he is working as a Lecturer in the Department of Bioinformatics and Biotechnology. He is among the founders of Department of Bioinformatics and Biotechnology in GC University Faisalabad. He has specialization in the field of Molecular Biology and fisheries. He is performing many administrative duties in the Department of Bioinformatics and Biotechnology and is an active member of research group of the department with many publications. He is working on molecular study of genetic disorders.

Hina Khalid, B.Sc.
E-mail: hina_huny24@yahoo.com

Hina Khalid has completed her B.Sc. Bioinformatics and now she is student of M.Sc. Bioinformatics and Biotechnology in GC University Faisalabad. She has specialization in using Bioinformatics tools, phylogenetic analysis and structure prediction.

Dr. Shahid Nadeem, Ph.D.
E-mail: Snadeem63@yahoo.com



Dr. Shahid Nadeem got his Ph.D. from the Punjab University, Lahore in 2002. He is working as a senior scientist in NIAB, Faisalabad. In 2009 he joined the Department of Bioinformatics and Biotechnology as a Chairman. He is among the founders of Department of Bioinformatics and Biotechnology in GC University Faisalabad. He is also the founder of Database and Software Engineering Research Group in the Department of Bioinformatics and Biotechnology. He is also an active member of teaching Faculty of the department, teaching some of the interesting subjects like Biotechnology and its applications, Social, Ethical aspects of Biotechnology also publishing textbooks. Scientific interests: Biotechnology, Plant breeding and Genetics, Microbiology.

Assist. Prof. Tayyaba Sultana, Ph.D. Genetics
E-mail: arif143@yahoo.com

Dr. Tayyaba Sultana, Assistant Professor, Department of Zoology, GC University, Faisalabad, Pakistan. She is a HEC approved Ph.D. Supervisor, too. During her academic career she got the distinction's scholarships like Quaid-i-Azam Merit Scholarship, INFAQ Foundation Merit Scholarship and The University Grants Commission (UGC) Merit Scholarship. She has a partial Research Project execution on my credit working as Co-link coordinator (2006-2009) under the British council-Higher Education Commission (JHELP-II) on 'Capacity building for molecular biology studies in Fish Production'. She has more than ten research publications in the national and international impact factor research journals and one book by the VDM Publishers (2010).

Saima Aslam, M.Sc. Computer Science

Mrs. Saima Aslam has completed her master degree in Computer Sciences. She has specialization in the field of Software development. Now she is working on the project Patient management system of civil hospital Faisalabad Pakistan.