Ilyas, M., Mir, A., Rauf, S., Nazir, S., & Javed, H. (2016). Huntington gene ontological study and orthological analysis. *World Journal of Biology and Biotechnology*, 1(2), 65-69.

(Online)
ISSN2522-6754
2016
WORLD JOURNAL OF BIOLOGY AND BIOTECHNOLOGY
(Print)
ISSN 2518-2463
Volume: 1
www.sciplatform.com
Issue: 2

HUNTINGTON GENE ONTOLOGICAL STUDY AND ORTHOLOGICAL ANALYSIS

[a] Muhammad Ilyas, [a] Asif Mir, [a] Sobiah Rauf, [a] Sidrah Nazir and [b] Humera Javed
[a] Department of Bioinformatics & Biotechnology, International Islamic University (IIU), Islamabad, Pakistan,
[b] Faulty of life Sciences and informatics, Balochistan University of Information Technology, Engineering and Management Sciences Quetta
*Corresponding email address: ilyas_baloch@ymail.com

ABSTRACT

Huntington gene is located on chromosome 4p16.3 IT15 locus considered a major causative gene of Huntington disorder. HTT is a neurodegenerative disorder mutation in gene cause abnormal movements and psychiatric disturbances. HTT is inherited in an autosomal dominant manner with almost complete penetrance and till now, no research studies provide insight into HTT gene. Bioinformatics analysis includes transcription factors binding sites, phylogenetic studies with reference to various selected orthologs and syntenic relationship of HTT gene. Our study showed that in HTT gene majority of the portion is conserved among two orthologs chimpanzee and mouse in significance to human. These studies also revealed information about conservation of genes among different ortholog species and their evolutionary relationship.

Key word: Huntington disorder, orthologs, CAG repeats, genome synteny analysis

INTRODUCTION

Huntington's disease is caused by the expansion of cytosine, adenine and guanine (CAG) repeats in the Huntington gene and the expansion of CAG repeats leads to psychiatric disturbances, abnormal movements and cognitive decline in affected person (Evans *et al.*, 2013). HTT is one of nine poly glutamine diseases with the only common feature being the expansion of a poly glutamine domain in the disease-specific protein (Orr and Zoghbi, 2007; Shao and Diamond, 2007). Huntington disease is inherited in an autosomal dominant fashion, so each child of an affected parent, regardless of gender, has a 50% chance of suffering from the disease (Telenius *et al.*, 1993). The symptom of Huntington's (HTT) disease can become appear at any time between the ages of 1 and 80 years but they usually appear between the ages of 35 and 45, occasionally earlier and sometimes later (Walker, 2007). Huntington's disease affects the mind, body and emotions. The common symptoms of Huntington disease in early stages can include poor memory, difficulty making decisions, mood changes such as increased depression, anger or irritability, a growing lack of coordination, twitching or other uncontrolled movements, difficulty walking, speaking and swallowing. The way in which symptoms develop will vary from person to person. As the disease continues, the symptoms become progressively worse(Tyagi *et al.*, 2010). The other most frequently observe symptoms are chorea movements and psychiatric disturbance. The onset of chorea movements generally leads to the diagnosis of HTT although psychiatric disturbance may be note as early as a decade or more before this movements begin (Di Maio *et al.*, 1993). In 1983 researchers of Massachusetts general hospital mapped the HTT gene located on the (p)short arm of chromosome (GUSELLA *et al.*, 2004). Huntington disorder was first mapped on to chromosome 4. HTT gene is also known as IT15 gene and this gene encodes the Huntington protein. HTT gene expression occurs in brain but functioning and expression pattern of the Huntington protein in the brain relics unknown. The anomalous HTT gene contains prolonged and unsteady DNA segment, unstable segment of DNA is composed with trinucleotide, CAG, these CAG are repetitive many time in row. These CAG repeating pieces are longer on the HTT chromosome than on the normal chromosome and are unstable and repeatedly changed size when it is conceded to next generation. These CAG repeats codes for the amino acid glutamine. The CAG expansion therefore causes the HTT protein to contain more glutamine than it normally contains specific no of CAG repeats (Telenius *et al.*, 1993). *HTT* locus contains normally up to 35 CAG repeats, while affected alleles have 36 and up to100 CAG repeats (McNeil *et al.*, 1997). The size of the CAG fragments can be uneven, especially when transmit by the male germ line (Telenius *et al.*, 1993). Different factors may be involved to add CAG instability, including the size of the CAG fragment, CAG tract interruptions, sex and age of the transmitting parent, environmental stress are also involved in CAG flux (Nguyen *et al.*, 2003; Nørremølle *et al.*, 2004). The identification of HTT gene in 1993, lead to the foundation of genetic mouse models of the disease to further study the disease mechanism. In particular, to understand the early change and progression of the disease could be followed and examined thoroughly (Cepeda *et al.*, 2010). No of genetic mouse models have been studied and generated and the first mouse model of HTT disease. These models include transgenic model, knock-in and conditional models with different controlled conditions (Cepeda *et al.*, 2010). Each model used to understand the mechanism of HTT but knock-in models are more realistic in terms of genetic context and in recapitulate the late onset, slow natural progression and neuropathology of HTT disease

(Cepeda *et al.*, 2010). Knock-in models is helpful in terms of expression of full-length Huntington's disease gene in its native genomic perspective. Several models have been generated, that differ mainly in the number of CAG repeats from 48 to 200 (Heng *et al.*, 2007). Behavioral changes are observed in knock-in mice, that is sensitive and careful testing shows abnormalities as early as 1–2 months of age (Menalled *et al.*, 2002; Menalled *et al.*, 2003). Other two knock-in models, first one isHTThQ150 and second one is HTThQ200, shows a behavioral change and that change appears phenotypically that is CAG- length dependent, with motor abnormalities appearing at 50 or 100 weeks of age and reduced DA receptors in heterozygotes (Heng *et al.*, 2007). The HTThQ200 mice also show regional-selective pathology in the striatum and cortex, and display age-dependent weight loss beginning at approx. 70 weeks (Heng *et al.*, 2010). HTT Gene annotation and gene ontology was performed, HTT transcription factors binding sites would be predicted computationally and HTT gene evolutionary pathway will be traced using probabilistic modeling and ortholog testing.

MATERIALS AND METHODS

The HTT gene and protein sequences were obtained from online Bioinformatics centers such as National Centre for Biotechnology Information (NCBI) and European Molecular Biology Laboratory (EMBL). The data set includes HTT (Huntington gene) sequences using the following gene identifiers from different sources. Postscript file is parsed by using NCBI gene bank browser and the gene identifiers are extracted and sequences are retrieved in FASTA format. For finding the binding sites for transcription factor the HTT gene identifiers were used as query identifiers on the "ChiP Transcription Factor Search Portal. MEGA5 was used for phylogenetic tree reconstruction to estimate evolutionary relationship (Tamura *et al.*, 2011).

Sequence retrieval: Eleven ortholog species with reference to human have been considered in the current study. These species include: Chimpanzee (*Pan troglodytes*), Mouse (*Musmusculus*), Macaque (*Macacamulatta*), Guinea Pig (*Caviaporcellus*), Megabat (*Pteropusvampyrus*), Dog (*Canisfamiliaris*), Anole Lizard (*Anoliscarolinensis*), Zebrafish (*Daniorerio*),Fugu (*Takifugurubripes*), Cionaintestinalis, Opossum (*Monodelphisdomestica*). Sequences of query gene and of all orthologs were collected through ensemble database. Sequence similarity of ortholog species with human gene sequence was analyzed through alignment using BLASTP against the protein database in order to choose closest putative orthologous protein sequences. This analysis was carried out in order to select the sequences of ortholog species (Altschul *et al.*, 1990; Johnson *et al.*, 2008).

Phylogenetic tree reconstruction: After sequence acquisition, pairwise and multiple alignments were carried out using ClustalW algorithm. The algorithm calculates similarity percentage between sequences and generates an alignment file which is further used as input file during tree reconstruction. Neighbor joining (NJ) method was used to construct tree. Tree was generated based on this alignment file which was used as input for tree reconstruction. NJ shows a high performance as compared to rest methods resulting the correct tree, but more perceptive as compared to other methods and do not build tree when evolutionary time varies amongst the genes with high rate (Saitou and Imanishi, 1989). Computer-based method (Bootstrap analysis) that assign corrective measures to sample estimate was also used (Efron and Tibshirani, 1994). Bootstrap values show the confidence and consistency of clusters. A tree topology test based on the bootstrap value which further validates the branching pattern of the tree. It is a precise way to control and check the stability of results obtained from Bootstrap analysis. In the current study bootstrap test uses 1000 replicates and assigns each branch a value ranging from 0 to 100 which gives an idea that how much a sequence is evolutionary closer to each other and also validates each branch. Synteny analysis was performed using Ensemble synteny view in Ensemble database (Hubbard *et al.*, 2002) and the visual analysis of conserved regions was carried out using web-based genome synteny viewer GSV. For this analysis only four ortholog species of human have been considered (Revanna *et al.*, 2011).

RESULTS

Data set: The data set includes HTT (Huntington gene) sequences using the following gene identifiers from different sources:

HGNC: 4851
Entrez Gene: 3064
Ensembl: ENSG00000197386
UniProtKB: P42858

Extraction of identifiers: HTT gene identifiers obtained after parsing the post script files were GC04P003046 and GC04P003014

Search result for transcription factors regulating HTT gene: CTFSP provides the most relevant transcription factors binding sites for the HTT gene

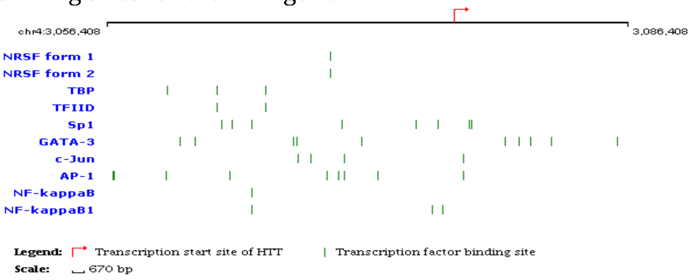

Fig.1. Displays the most relevant transcription factors binding sites in the HTT gene promoter as predicted by the SA Biosciences Text Mining Application and the UCSC.

Phylogenetic tree of orthologos: Neighbor joining tree for HTT is shown in Figure 1. According to tree, human and chimpanzee are in one cluster with a bootstrap value of 100 showing 100% reliability of this cluster. This also indicates that human is closely related to the chimpanzee as compare to other species. Macaque is close to the cluster of human/chimpanzee. Bootstrap values only greater than 70 are mentioned in the tree. Megabat and dog are in one cluster. Similarly mouse/guinea pig and fugu/zebrafish are making clusters with 77 and 99 as bootstrap values, respectively. High bootstrap values of the clusters show their reliability. The phylogenetic tree is reconciling species divergence time. Vertebrate cionaintestinal is as out group in this tree. Evolutionary time for the tree is 0.05.

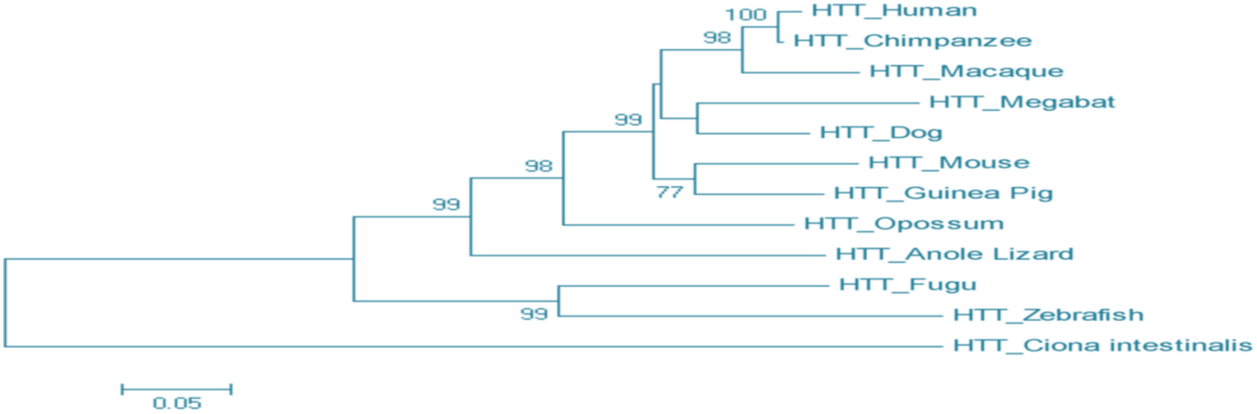

Fig. 2. Neighbor Joining (NJ) Tree for human HTT using MEGA5, numbers on branches represent bootstrap values (based on 1000 replications).

Orthologs testing for HTT gene: Genome synteny analysis: To find out the genomic element that are functionally conserved, we find out set genes or loci that show conserved regions, in the same locus on a set of homologous chromosomes (of human and its four orthologs). Study shows conservation of human 8 genes (both upstream and downstream of HTT) with four orthologs genes of HTT. Data collected from ensembl synteny view in ensembl database and its summary is given in Table 1.

S/N	Chimpanzee [Gene (Location)]	Mouse [Gene (Location)]	Human [Gene (Location)]	Rabbit [Gene (Location)]	Platypus [Gene (Location)]
Upstream Genes	**C4orf48** (4:2087420-2089804)	**Gm1673** (5:33983437-33985013)	**C4orf48** (4:2043689-2045697)	No homologues	**C4orf48** (36-1698)
	No homologues	**NAT8L** (5:33995984-34005916)	**NAT8L** (4:2061239-2070816)	**NAT8L** (:104228-106710)	**NAT8L** (6258-6662)
	POLN (4:2115292-2274953)	POLN (5:34007198-34169448)	**POLN** (4:2073645-2243848)	**POLN** (:111903-303031)	No homologues
	HAUS3 (4:2277567-2287854)	**HAUS3** (5:34153921-34169445)	**HAUS3** (4:2229191-2243891)	**HAUS3** (:305104-320164)	**HAUS3** (1607768-1617825)
Downstream Genes	HTT (4:3144882-3291490)	HTT (5:34761740-34912534)	HTT (4:3076408-3245676)	**HTT** (:966937-1141172)	**HTT** (3260-9792)
	MSANTD1 (4:3322257-3334641)	**MSANTD1** (5:34915915-34923839)	**MSANTD1** (4:3246096-3273465)	**MSANTD1** (:1147609-1153393)	No homologues
	RGS12 (4:3392745-3518475)	**RGS12** (5:34949445-35039644)	**RGS12** (4:3294755-3441640)	No homologues	**ox_plat1_124442** (4568-42527)
	HGFAC (4:3520338-3527954)	**HGFAC** (5:35041553-35048436)	**HGFAC** (4:3443614-3451211)	No homologues	**HGFAC** (288-9126)

Table 1. HTT with 8 genes (Upstream and Downstream) in human and its four Orthologs.

Different orthologs used to consider for this study and these orthologs are chimpanzee (*pan troglodytes*), mouse (*mus musculus*), rabbit (*oryctolagus cuniculus*) and platypus (*ornithorhynchus* Anatinus). These four orthologs were selected due the reason that are closely related as well as those diverged with respect to human. Through this way we were able to clearly reveal the presence and absence of conserved synteny between human and its orthologs species. Conserved region were also generated by using genome synteny viewer GSV web server which produced graphical representations and facilitated the quick visualization of conserved areas in the form of colored blocks with the ruler indicating location of these conserved regions (Figure 2). Colored blocks in relevance to human showing conserved portion between human and other orthologs. Our analysis showed that in HTT majority of the portion is conserved among two orthologs (chimpanzee and mouse) in relevance to human with only 3 deletions and less conservation found with rabbit and platypus with 11 and 10 deletions, respectively. Changes which lead towards the evolution of these organisms are given in Table 2. Common deletion in four orthologs in relevance to human in case of HTT is AL590235.1 gene (Table 2).

Organism	Observed No. of Deletions	Deleted Genes
Organism	3	NAT8L, **AL590235.1** , STX18
Chimpanzee	3	RP11-503N18.3, **AL590235.1**, LINC00955
Mouse	11	C4orf48, RP11-503N18.3, MFSD10, RGS12,HGFAC, DOK7, LRPAP1, **AL590235.1**, LINC00955, ADRA2C, MSX1
Rabbit	10	POLN, MXD4, ZFYVE28, RP11-503N18.3, MSANTD1, DOK7, LRPAP1, **AL590235.1**, LINC00955, ADRA2C
Platypus	3	NAT8L, **AL590235.1** , STX18

*Bold indicates common deletions in four orthologs.

HTT is present in four ortholog species (chimpanzee, mouse, dog and Platypus) in relevance to human which shows its importance in these species.

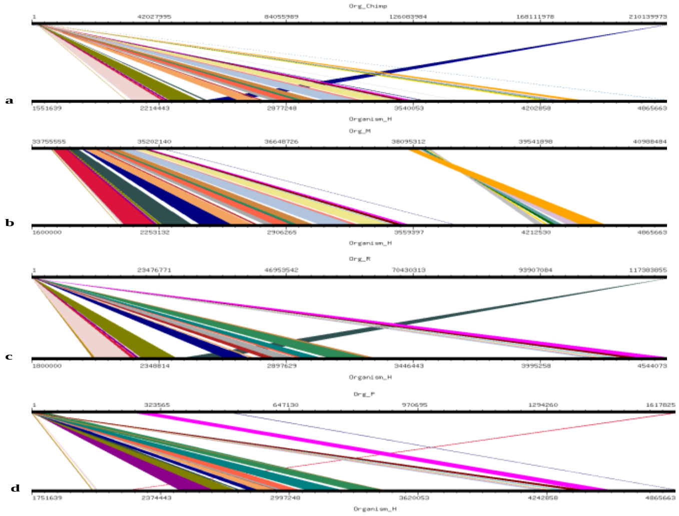

Fig 3. Results of GSV for Human HTT shows highly Conserve Region; a) Org (Chimpanzee) vs Organism (Human); b) Organism (Mouse) vs Organism (Human); c) Organism (dog) vs Organism (Human); d) Org (Chicken) vs Organism (Human)

DISCUSSION

Huntington disease (HTT) is a progressive neurodegenerative disorder with an established genetic origin. It belongs to the polyQ accumulation disease with mutant Huntington protein. The mutant Huntington has expanded CAG triplet repeats, which make it easy to accumulate (Finkbeiner, 2011). HTT has number of transcription binding sites in the promoter region. Transcription factor regulate the gene expression and these are helpful gaining insight into mechanism of transcription regulation (Nguyen *et al.*, 2014). More than ten different transcription factor bind to the HTT gene. Out of those ten transcription factor NRSF has only two different binding sites to HTT gene while GATA-3 and AP-1 shows ten different binding sites at HTT gene. Neighbor joining tree shows that human and chimpanzee is in common cluster with 100% similarity. Evolutionary relationship shows that human is closely related with chimpanzee. The ortholog testing shows conserved region similar to HTT gene. HTT gene most commonly in chimpanzee, mouse, rabbit (*oryctolagus cuniculus*) and platypus (*Ornithorhynchus anatinus*) followed by several others and the phylogenetic linkage deduced the evolutionary pathway and conserved regions. To find out functionally conserved elements in genome, genome Synteny analysis conform that 8 different genes in human show conserved region to four orthologs genes of HTT. A common deletion is found out in four orthologs species and the deletion is AL590235.1 in reference to human (Tamura *et al.*, 2011). Bioinformatics provides an efficient and comfortable approach towards gene study and protein analysis. The methodology utilized for this research work provides effective and efficient HTT gene based study including ontology, binding site prediction, ortholog detection. The syntenic relationship of the HTT gene has determined conservation of genomic elements among four human orthologs i.e. chimpanzee, mouse, dog and chicken (with respect to 15 upstream and downstream genes of human HTT.

Acknowledgements

We are obliged to all those individuals who have provided us moral and functional support during this research work. Our deep gratitude and appreciation are extended to Baluchistan University of Information Technology, Engineering and Management Sciences and Faculty of Life Sciences & Informatics, for using their bioinformatics lab in addition to their continuous help throughout the study. We also thank departments of bioinformatics and biotechnology international Islamic university Islamabad.

REFERENCE

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990. Basic local alignment search tool. Journal of molecular biology, 215(3): 403-410.

Cepeda, C., D. M. Cummings, V. M. André, S. M. Holley and M. S. Levine, 2010. Genetic mouse models of huntington's disease: Focus on electrophysiological mechanisms. ASN neuro, 2(2): AN20090058.

Di Maio, L., F. Squitieri, G. Napolitano, G. Campanella, J. A. Trofatter and P. M. Conneally, 1993. Onset symptoms in 510 patients with huntington's disease. Journal of medical genetics, 30(4): 289-292.

Efron, B. and R. J. Tibshirani, 1994. An introduction to the bootstrap. CRC press.

Evans, S. J., I. Douglas, M. D. Rawlins, N. S. Wexler, S. J. Tabrizi and L. Smeeth, 2013. Prevalence of adult huntington's disease in the uk based on diagnoses recorded in general practice records. Journal of Neurology, Neurosurgery & Psychiatry: jnnp-2012-304636.

Finkbeiner, S., 2011. Huntington's disease. Cold Spring Harbor perspectives in biology, 3(6): a007476.

Gusella, J. F., N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, k. Ottina, M. R. Wallace and A. Y. SakaguchI, 2004. A polymorphic DNA marker genetically linked to huntington’s disease. Landmarks in Medical Genetics: Classic Papers with Commentaries, 306(51): 153.

Heng, M. Y., D. K. Duong, R. L. Albin, S. J. Tallaksen-Greene, J. M. Hunter, M. J. Lesort, A. Osmand, H. L. Paulson and P. J. Detloff, 2010. Early autophagic response in a novel knock-in model of huntington disease. Human molecular genetics, 19(19): 3702-3720.

Heng, M. Y., S. J. Tallaksen-Greene, P. J. Detloff and R. L. Albin, 2007. Longitudinal evaluation of the hdh (cag) 150 knock-in murine model of huntington's disease. The Journal of neuroscience, 27(34): 8989-8998.

Hubbard, T., D. Barker, E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen and T. Down, 2002. The ensembl genome database project. Nucleic acids research, 30(1): 38-41.

Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis and T. L. Madden, 2008. Ncbi blast: A better web interface. Nucleic acids research, 36(suppl 2): W5-W9.

McNeil, S. M., A. Novelletto, J. Srinidhi, G. Barnes, I. Kornbluth, M. R. Altherr, J. J. Wasmuth, J. F. Gusella, M. E. MacDonald and R. H. Myers, 1997. Reduced penetrance of the huntington's disease mutation. Human molecular genetics, 6(5): 775-779.

Menalled, L. B., J. D. Sison, I. Dragatsis, S. Zeitlin and M. F. Chesselet, 2003. Time course of early motor and neuropathological anomalies in a knock-in mouse model of huntington's disease with 140 cag repeats. Journal of Comparative Neurology, 465(1): 11-26.

Menalled, L. B., J. D. Sison, Y. Wu, M. Olivieri, X.-J. Li, H. Li, S. Zeitlin and M.-F. Chesselet, 2002. Early motor dysfunction and striosomal distribution of huntingtin microaggregates in huntington's disease knock-in mice. The Journal of neuroscience, 22(18): 8266-8276.

Nguyen, G. H., J. Bouchard, M. G. Boselli, L. G. Tolstoi, L. Keith, C. Baldwin, N. C. Nguyen, M. Schultz, V. L. Herrera and C. L. Smith, 2003. DNA stability and schizophrenia in twins. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 120(1): 1-10.

Nguyen, T. T., J. S. Mattick, Q. Yang, M. A. Orman, M. G. Ierapetritou, F. Berthiaume and I. P. Androulakis, 2014. Bioinformatics analysis of transcriptional regulation of circadian genes in rat liver. BMC bioinformatics, 15(1): 1.

Nørremølle, A., L. Hasholt, C. B. Petersen, H. Eiberg, S. G. Hasselbalch, P. Gideon, J. E. Nielsen and S. A. Sørensen, 2004. Mosaicism of the cag repeat sequence in the huntington disease gene in a pair of monozygotic twins. American journal of medical genetics Part A, 130(2): 154-159.

Orr, H. T. and H. Y. Zoghbi, 2007. Trinucleotide repeat disorders. Annu. Rev. Neurosci., 30: 575-621.

Revanna, K. V., C.-C. Chiu, E. Bierschank and Q. Dong, 2011. Gsv: A web-based genome synteny viewer for customized data. BMC bioinformatics, 12(1): 316.

Saitou, N. and T. Imanishi, 1989. Relative efficiencies of the fitch margoliash, maximum parsimony, maximum likelihood, minimum-evolution, and neighbor joining methods of phylogenetic tree construction in obtaining the correct tree. Mol. Biol. Evol, 6(5): 514-525.

Shao, J. and M. I. Diamond, 2007. Polyglutamine diseases: Emerging concepts in pathogenesis and therapy. Human molecular genetics, 16(R2): R115-R123.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar, 2011. Mega5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution, 28(10): 2731-2739.

Telenius, H., H. Kremer, J. Thellmann, S. Andrew, E. Almqvist, M. Anvret, C. Greenberg, J. Greenberg, G. Lucotte and F. Squltierl, 1993. Molecular analysis of juvenile huntington disease: The major influence on (cag) n repeat length is the sex of the affected parent. Human molecular genetics, 2(10): 1535-1540.

Tyagi, S., L. K. Tyagi, R. Shekhar, M. Singh and M. Kori, 2010. Symptomatic treatment and management of huntington's disease: An overview. Global Journal of Pharmacology, 4(1): 06-12.

Walker, F. O., 2007. Huntington's disease. The Lancet, 369(9557): 218-228.

Date Published (M-D-Y): 15-8-2016