9-1-2012

Using the Neandertal and Denisova Genetic Data to Understand the Common *MAPT* 17q21 Inversion in Modern Humans

Núria Setó-Salvia
Àrea de Prehistoria, Universitat Rovira i Virgili (URV), Tarragona, Spain

Federico Sánchez-Quinto
Institut de Biologia Evolutiva (UPF-CSIC), CEXS Universitat Pompeu Fabra, Barcelona, Spain

Eudald Carbonell
Àrea de Prehistoria, Universitat Rovira i Virgili (URV), Tarragona, Spain

Carlos Lorenzo
Àrea de Prehistoria, Universitat Rovira i Virgili (URV), Tarragona, Spain

David Comas
Institut de Biologia Evolutiva (UPF-CSIC), CEXS Universitat Pompeu Fabra, Barcelona, Spain

See next page for additional authors

Authors
Núria Setó-Salvia, Federico Sánchez-Quinto, Eudald Carbonell, Carlos Lorenzo, David Comas, and Jordi Clarimón

Using the Neandertal and Denisova genetic data to understand the common *MAPT* 17q21 inversion in modern humans

Núria Setó-Salvia,[1,2,3] Federico Sánchez-Quinto,[4] Eudald Carbonell,[2,3,5] Carlos Lorenzo,[2,3] David Comas,[4] and Jordi Clarimón[1,6]

[1] IIB Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain.

[2] Àrea de Prehistoria, Universitat Rovira i Virgili (URV), Tarragona, Spain.

[3] Institut Català de Paleoecologia Humana i Evolució Social (IPHES), Tarragona, Spain.

[4] Institut de Biologia Evolutiva (UPF-CSIC), CEXS Universitat Pompeu Fabra, Barcelona, Spain.

[5] Institute of Vertebrate Paleontology and Paleoanthropology of Beijing (IVPP), China.

[6] Center for Networker Biomedical Research in Neurodegenerative Diseases (CIBERNED), Madrid, Spain.

Corresponding author: Dr. Jordi Clarimón. Hospital de la Santa Creu i Sant Pau. Genetics of Neurodegenerative Diseases Unit. Alzheimer's Laboratory. Building 19. Sant Antoni M. Claret 167, 08025 Barcelona, Spain. Phone: +34 93 2919050; Fax: +34 93 5565602E-mail: jclarimon@santpau.cat

Keywords: Denisova; Microtubule associated protein Tau; MAPT; Haplotypes; Neandertal; Genetics.

Abstract

The polymorphic inversion on 17q21, that includes the *MAPT* gene, represents a unique locus in the human genome characterized by a large region with strong linkage disequilibrium. Two distinct haplotypes, H1 and H2, exist in modern humans, and H1 has been unequivocally related to several neurodegenerative disorders. Recent data indicates that recurrent inversions of this genomic region have occurred through primate evolution, with the H2 haplotype being the ancestral state. Neandertals harbored the H1 haplotype, however until now no data was available for the Denisova hominin. Neandertals and Denisovans are sister groups that share a common ancestor with modern humans. We analyzed the *MAPT* sequence and assessed the differences between modern humans, Neandertals, Denisovans, and great apes. Our analysis indicated that the Denisova hominin carried the H1 haplotype and the Neandertal and Denisova common ancestor probably shared the same subhaplotype (H1j). We also found 68 intronic variants within the *MAPT* gene, 23 exclusive to Denisova hominin, 6 limited to Neandertals and 24 exclusive to present-day humans. Our results reinforce previous data suggesting that the 17q21 inversion arose within the modern human lineage. The data also indicates that archaic hominins that coexisted in Eurasia probably shared the same *MAPT* subhaplotype, that can be found in almost 2% of chromosomes from European ancestry.

The polymorphic ~970 Kb inversion on human chromosome 17q21 constitutes an excellent example of the evolutionary pattern and genomic architecture of chromosomal rearrangements. This inversion encompasses several genes, including the microtubule associated protein Tau (*MAPT*) (OMIM ID: 157140), and is characterized by an unusual and strong linkage disequilibrium region that extends ~1.6 Mb (Baker et al. 1999), which in turn differentiates two major haplotypes: H1 (direct orientation) and H2 (inverted orientation), with no recombination between them. The H1 haplotype is the most frequent variant and is present in all present-day human populations, whereas the H2 is mainly found in Southwest Asian and Southern European populations, with frequencies ranging from 21% to 32% (Evans et al. 2004; Donnelly et al. 2010). Interestingly, different H1 variants have been associated with several neurodegenerative disorders, such as Alzheimer's disease, Parkinson's disease, progressive supranuclear palsy, corticobasal degeneration, and frontotemporal dementia (Simón-Sánchez et al. 2009; Dickson et al. 2011; Höglinger et al. 2011; Setó-Salvia et al. 2011). On the other hand, the H2 haplotype has been associated with increased fecundity and higher recombination rates over the genome, which might be related to its positive selection in the Icelandic population (Stefansson et al. 2005). Due to the peculiarities of this unique genomic region, several studies have focused on the reconstruction of its evolutionary history. Genetic analyses in non-human primates such as macaques and orangutans suggest that H2 is the ancestral state, and recurrent inversion events have occurred independently in chimpanzees and humans (Zody et al. 2008; Donnelly et al. 2010). However, whether the H1 haplotype is the primitive orientation in the *Homo* genus and H2 arose either by a recent inversion in modern humans from Africa, or by introduction to humans from other hominins, is still a matter of controversy. High throughput sequencing of nuclear DNA from Neandertal specimens has recently leaded to the

conclusion that Neandertals carried the more common H1 haplotype (Green et al. 2010). Also recently, DNA from a bone excavated in Denisova Cave in the Altai Mountains (Southern Siberia) has been sequenced. This Denisova individual belongs to a hominin group that shares common ancestor with Neandertals (640,000 years ago) and also lived during the Late Pleistocene (Reich et al. 2010). We have used DNA sequence data from these two archaic hominins to follow the evolution of this singular polymorphic inversion. We have also employed our data from a previous analysis of the *MAPT* region in present-day European population (Setó-Salvia et al. 2011) to determine the present-day frequency of the haplotype variant carried by both archaic hominins.

Materials and Methods

In order to analyze the Neandertal and Denisova low coverage genomes, bam read files from all Neandertal samples from the UCSC ftp site (ftp://hgdownload.cse.ucsc.edu/gbdb/hg18/neandertal/seqAlis) were downloaded and merged and Denisova data was obtained from ftp://hgdownload.cse.ucsc.edu/gbdb/hg18/denisova/. The same filters for mapping quality and damage treatment were largely used as in previous studies (Reich et al. 2010; Skoglund et al. 2011; Sánchez-Quinto et al. 2012). Given that both ancient hominin genomes are at low coverage, only reads with a base quality of at least 30 were used for comparison with the corresponding human *MAPT* sequences.

The first Denisova high-coverage genome has been available online since February 2012. Since low coverage genomes could be prone to sequence errors, and in the case of ancient DNA, additionally influenced by postmortem modifications, we performed the same analysis using the high coverage Denisova genome.

The entire human *MAPT* sequence was retrieved through the Unified Genotyper program from the GATK package (http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit) version 1.0.4641M, using default parameters.

This region comprised the 134,002 bp genomic segment in chromosome 17 (from position 41,327,544 to position 41,461,546 of the human genome hg18 assembly). The biallelic polymorphisms rs1467967, rs242557, rs3785883, rs2471738, and rs7521, where used to define the different subhaplotypes, as described earlier (Myers et al. 2005; Pittman et al. 2005; Myers et al. 2007). To analyze the 238 bp insertion/deletion polymorphism in intron 9 (*del-In9*), we used the genetic variant rs17652121, which is in complete linkage disequilibrium and can be used as a surrogate marker for the inversion status (Baker et al. 1999). Data regarding the frequency of *MAPT* subhaplotypes in European individuals was retrieved from our previous data (Setó-Salvia et al. 2011) resulted from the genotyping of the six polymorphisms in a series of 374 unrelated Spanish individuals from European ancestry.

Results

Up to 50.8% and 87% of the 134,002 bp region including the *MAPT* gene has been sequenced in Neandertals and Denisova hominin, respectively. These sequences include five biallelic polymorphisms (rs1467967, rs242557, rs3785883, rs2471738 and rs7521) and a 238 bp deletion between exons 9 and 10 of *MAPT* that distinguishes H1 from H2 haplotypes (Myers et al. 2005; Pittman et al. 2005; Myers et al. 2007). Specific alleles of this genomic region in different human linages and non-human primates are shown in Table 1. Our analysis indicates that the haplotype in Denisova hominin represents the H1 chromosome. Detailed analysis of its composition disclosed that both hominins

harbored the H1j variant, although the heterozygous genotype at the rs3785883 polymorphism in Denisova could also designate the H1l variant in this hominin. The H1j variant represents ~ 1.7% of all possible subhaplotypes in present-day humans of European ancestry (Figure 1). Subsequently, the 134,002 bp genetic region comprising the *MAPT* gene was compared between modern humans, Neandertals, Denisova and other non-human primates (Table 2). A total of 68 intronic variants were found, 23 were exclusive to the Denisova hominin, 6 were exclusive to Neandertals and 24 were exclusive to modern humans.

Discussion

The Neandertal and modern human lineages diverged about 400,000-800,000 years ago (Reich et al. 2010; Langergraber et al., 2012). The remains from Denisova, a phalanx found in 2008 and a molar found in 2010 (Reich et al. 2010), are dated between 50,000 and 30,000 years ago. There is no doubt that these three species co-existed during the same time period, but how the *MAPT* gene sequence differed between them is still unknown (Figure 2) (Holzer et al. 2004; Cruts et al. 2005; Hardy et al. 2005; Stefansson et al. 2005; Zody et al. 2008; Donnelly et al. 2010) and so far there is no evidence of *MAPT* variants exchange between them. To understand when the H2 inversion appeared in evolution, several studies have analyzed *MAPT* SNPs that are in linkage disequilibrium with the H1 and H2 haplotypes in human populations (Steinberg et al. 2012) and other primates (Steffanson et al. 2005; Donnelly et al. 2010). Recent analysis of nuclear DNA from Neandertals showed the presence of an H1 haplotype (Green et al. 2010) suggesting that the inversion appeared recently in the human lineage, however there is still uncertainty over when the inversion occurred.

The present study analyses the currently available data and indicates that Denisova hominin carried an H1 haplotype as well. This observation could be explained by the fact that Denisovans and Neandertals were closely related, though there is still debate about the classification of Denisovans as a distinct species and about their temporal and geographical distribution. Our analysis shows that both archaic hominins most probably carried the H1j subhaplotype as well. However, the presence of a possible heterozygous state of one biallelic polymorphisms in Denisova could also indicate that this hominin carried the H1l subhaplotype. Interestingly, none of these variants (H1j or H1l) have been related to any of the neurodegenerative diseases common to nowadays humans, such as Alzheimer's dementia or progressive supranuclear palsy, where H1c is a risk haplotype, or Parkinson's disease, where the H1p variant has been overrepresented in patients compared to healthy controls (Myers et al. 2005; Pittman et al. 2005; Setó-Salvia et al. 2011).

Krause et al. (2010) suggested that Denisova derives from a hominin migration out of Africa [ca. 1.0 My] distinct from that of the ancestors of Neandertals and of modern humans. But an alternative interpretation has been proposed by Martinón-Torres et al. (2011) who stated that an Asian origin of the Denisovans cannot be excluded. Both hypotheses can neither be confirmed nor refuted on the basis of the available *MAPT* sequence data. These results also argue against the introduction of the H2 haplotype into European populations through a gene flow from any of these two archaic hominins. Since data from both Neandertals and Denisova are scarce, we cannot discard the existence of H2 chromosome carriers (either heterozygous or homozygous) co-inhabiting during the same period. Haplotype data of the H2 lineage in modern humans indicates that it is extremely homogeneous relative to the diverse H1 haplotype (Stefansson et al., 2005). Because of the ancient (2.3 million years) coalescence time of

these two haplotypes, and given the high frequencies of the H2 linage in Europeans, a potential positive selection or extraordinary genetic drift has been suggested for this particular genomic inversion. "The comparison of the *MAPT* gene sequence between these three hominids and great apes has provided a greater knowledge of their genetic variation at the *MAPT* locus which will in turn provide a better understanding of the causes human evolution, variation and disease.

Acknowledgements

This work was supported by a research grant from 'Ciberned', and 'Fondo de Investigaciones Sanitarias' PI09/00098. Also this work has been developed in the framework of the Spanish MICINN (CGL2009-12703-C03-02). We thank Dr. Belbin for editorial help.

REFERENCES

Baker, M., I. Litvan, H. Houlden et al. 1999. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Hum. Mol. Genet. 8:711–715.

Cruts, M., R. Rademakers, I Gijselinck et al. 2005. Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. Hum. Mol. Genet. 14:1753-1762.

Dickson, D.W., N. Kouri, M.E. Murray et al. 2011. Neuropathology of Frontotemporal Lobar Degeneration-Tau (FTLD-Tau). J. Mol. Neurosci. 45:384-389.

Donnelly, M.P., P. Paschou, E. Grigorenko et al. 2010. The Distribution and Most Recent Common Ancestor of the 17q21 Inversion in Humans. Am. J. Hum. Genet. 86:161-171.

Evans, W., H.C. Fung, J. Steele et al. 2004. The tau H2 haplotype is almost exclusively Caucasian in origin. Neurosci. Lett. 369:183–185.

Green, R.E., J. Krause, A.W. Briggs et al. 2010. A draft sequence of the Neandertal genome. Science. 328:710–722.

Hardy, J., A.M. Pittman, A. Myers et al. 2005. Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens. Biochem. Soc. Trans. 3:582-585.

Höglinger, G.U., N.M. Melhem, D. Dickson et al. 2011. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nat. Genet. 43:699-705.

Holzer, M., M. Craxton, R. Jakes et al 2004. Tau gene (MAPT) sequence variation among primates. Gene. 341:313-322.

Krause, J., Q. Fu, J.M. Good, B. Viola, M.V. Shunkov, A.P. Derevianko, S. Paabo 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature 464: 894-897.

Langergraber, K.E., K. Prüfer, C. Rowney et al. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc. Natl. Acad. Sci. U S A.

Martinón-Torres, M., R. Dennell, J.M. Bermúdez de Castro 2011. The Denisova hominin need not be an out of Africa story. J. Hum. Evol. 60:251-255.

Myers, A.J., M. Kaleem, L. Marlowe et al. 2005. The H1c Haplotype at the MAPT Locus is associated with Alzheimer's disease. Hum. Mol. Genet. 14:2399–2404.

Myers, A.J., A.M. Pittman, A.S. Zhao et al. 2007. The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. Neurobiol. Dis. 25:561-570.

Pittman, A.M., A.J. Myers, P. Abou-Sleiman et al. 2005. Linkage disequilibrium fine-mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. J. Med. Genet. 42:837–846.

Reich, D., R.E. Green, M. Kircher et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 468:1053-1060.

Sánchez-Quinto, F., H. Schroeder, O. Ramirez et al. 2012. Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gathereres. Curr. Biol. [*in press*]

Setó-Salvia, N., J. Clarimón, J. Pagonabarraga et al. 2011. Dementia risk in Parkinson disease:disentangling the role of MAPT haplotypes. Arch. Neurol. 68:359-364.

Simón-Sánchez, J., C. Schulte, J.M. Bras et al. 2009. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat. Genet. 41:1308-1312.

Skoglund, P., H. Malmström, M. Raghavan et al. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science. 336:466–469.

Stefansson, H., A. Helgason, G. Thorleifsson et al. 2005. A common inversion under selection in Europeans. Nat. Genet. 37:129–137.

Zody, M.C., Z. Jiang, H.C. Fung et al. 2008. Evolutionary toggling of the MAPT 17q21.31 inversion region. Nat. Genet. 40:1076–1083.

Table 1. *MAPT* allele and haplotype diversity in Neanderthal, Denisova (high coverage 30X) and non-human primates.

	rs1467967	**rs242557**	**rs3785883**	**rs2471738**	***(del/In9)***	**rs7521**	**Haplotype**
Rhesus	A	G	C	C	*del*	G	H2
Orangutan	A	G	C	C	*del*	G	H2
Gorilla	A	G	G	C	*del*	G	H2
Chimpanzee	A	G	G	C	*del*	G	H2
Denisova	A	G	G/A	C	*ins*	G	H1j/H1l
Neanderthal	A	G	G	C	*ins*	G	H1j

Table 2. *MAPT* sequence comparison between great apes and human linages. High coverage 30X for Denisova results.

Location	dbSNP	Human	Neanderthal	Denisova	Chimp	Orang	Macaque
41328711	rs3744457	C/t	-	T/T	T	T	T
41329294	rs4792891	T/g	-	G/G	G	G	G
41330390	rs7224541	A/g	-	A/A	G	G	G
41332031	rs11079726	A/g	G	G/G	G	G	G
41332480	rs9303523	T/c	C	C/C	C	C	C
41335605		G	-	T/T	G	G	G
41335768	rs930119	A/g	G	G/G	G	G	G
41336236		A	-	T/T	A	-	A
41342162		T	-	C/C	T	T	T
41342491		C	-	T/T	C	C	C
41346738		C	-	T/T	C	C	T
41346752		G	-	C/C	G	G	G
41347275		G	-	T/T	G	G	G
41347343	rs62056842	G/t	-	G/T	T	T	T
41347381	rs9915721	A/g	-	G/G	G	G	G

41348771	rs9899833	G/a	-	A/A	A	-	C
41349951		T	C	T/T	T	T	T
41351747		A	-	T/A	A	A	A
41351968	rs8080903	T/c	-	C/C	C	C	C
41352389	rs1560313	A/g	-	G/G	G	G	A
41355112		A	-	C/C	A	A	A
41358099	rs9904290	A/g	G	G/G	G	G	G
41358460	rs2055797	T/a	-	T/T	A	A	A
41361649	rs1001945	G/c	G	G/G	C	C	C
41363929	rs8078967	T/c	-	C/C	C	C	C
41365392		G	-	A/A	G	G	G
41369240		C	-	G/C	C	C	C
41374593	rs2435205	G/a	-	A/A	A	A	A
41375573	rs242557	A/g	-	G/G	G	G	G
41375997		G	-	C/C	G	G	G
41380709		G	-	A/A	G	G	G
41381748	rs242559	A/c	-	C/C	C	-	C
41382191	rs242560	A/g	-	G/G	G	G	G

41385632		G	-	A/G	G	G	G
41387681		C	-	T/C	C	C	C
41391156		A	-	G/G	A	A	A
41391300		T	C	T/T	T	T	-
41391869		G	-	T/T	G	G	-
41392246	rs62063276	T/g	G	G/G	G	A	-
41392426		C	-	T/C	C	C	-
41395201	rs242554	T/c	-	C/C	C	C	T
41405165		C	-	A/A	C	C	C
41406664	rs9896485	C/g	G	G/G	G	G	G
41410532	rs754593	A/g	-	G/G	G	G	G
41410869		A	G	A/A	A	A	A
41417319		G	-	A/A	G	G	G
41418440	rs6503453	G/a	-	A/A	A	A	A
41421578	rs919463	A/g	-	G/G	G	G	G
41422301	rs2435212	A/g	G	G/G	G	G	G
41423219	rs2258689	C/t	-	T/T	T	T	T
41423684	rs2435213	T/g	G	G/G	G	G	G

41427734		C	-	T/T	C	C	C
41430007		C	-	T/T	C	C	C
41432926		C	-	T/T	C	C	C
41435218		C	T	C/C	C	C	C
41436191	rs2471739	C/t	-	T/T	T	T	T
41436673	rs2435202	A/g	G	G/G	G	G	G
41439134	rs2435203	C/t	-	T/T	T	T	T
41442104	rs4792897	G/a	-	A/A	A	A	A
41445598		G	-	A/A	G	G	A
41445599		T	-	A/A	T	T	T
41447700		C	-	T/T	C	C	C
41449536		G	-	C/C	G	G	-
41450397		T	-	C/C	T	T	-
41454707		A	G	A/A	A	A	A
41456634		T	-	A/T	T	T	T
41459277		C	T	C/C	C	C	C
41459318		A	G	A/A	A	A	A

Lower cases indicate the ancestral state of human polymorphisms.

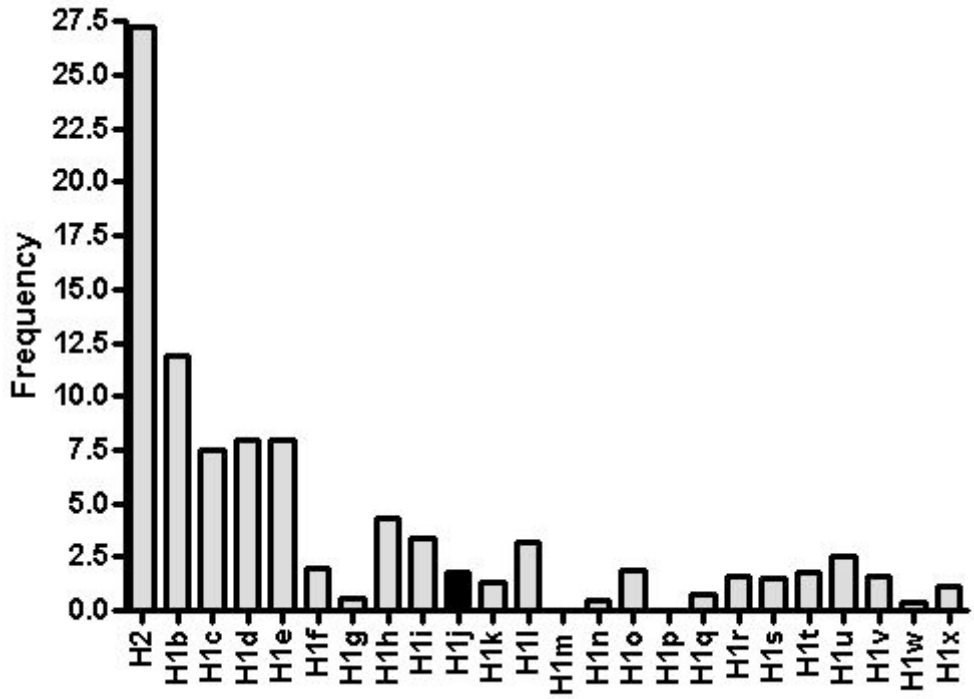

Figure 1. Subhaplotypes frequencies of present-day humans of European ancestry (data from Setó-Salvia et al. 2010). In black is the H1 subhaplotype encountered in Neandertal and Denisova.

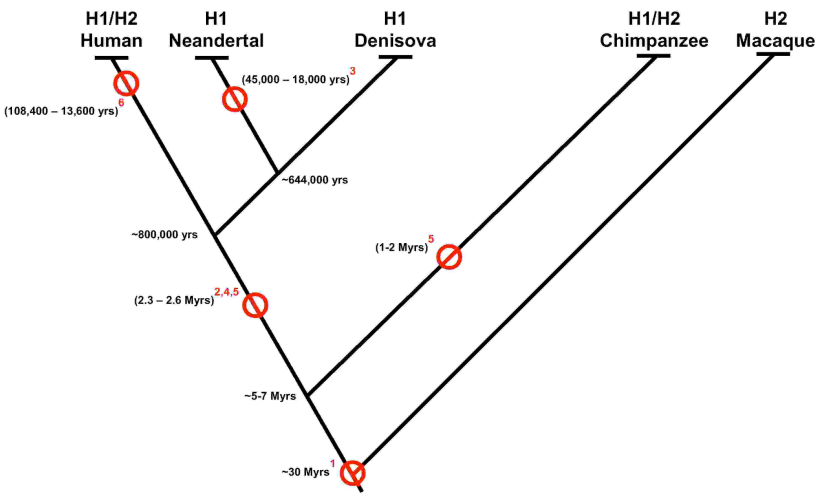

Figure 2. Evolutionary model of *MAPT* inversions between great apes and humans. Red circles indicate the chronological period suggested for the presence of H2 haplotype. References for Figure 2 are as follows:

1- Holzer, M., M. Craxton, R. Jakes et al. 2004. Tau gene (MAPT) sequence variation among primates. Gene. 341:313-322.

2- Stefansson, H., A. Helgason, G. Thorleifsson et al. 2005. A common inversion under selection in Europeans. Nat. Genet. 37:129–137.

3- Hardy, J., A.M. Pittman, A. Myers et al. 2005. Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens. Biochem. Soc. Trans. 3:582-585.

4- Cruts, M., R. Rademakers, I. Gijselinck et al. 2005. Genomic architecture of human 17q21 linked to frontotemporal dementia uncovers a highly homologous family of low-copy repeats in the tau region. Hum. Mol. Genet. 14:1753-1762.

5- Zody, M.C., Z. Jiang, H.C. Fung et al. 2008. Evolutionary toggling of the MAPT 17q21.31 inversion region. Nat. Genet. 40:1076–1083.

6- Donnelly, M.P., P. Paschou, E. Grigorenko et al. 2010. The Distribution and Most Recent Common Ancestor of the 17q21 Inversion in Humans. Am. J. Hum. Genet. 86:161-171.