Universidade de Aveiro Secção Autónoma de Ciências da Saúde
2012

MARIA INÊS MORGADO OLIVEIRA MARTINS

EVOLUÇÃO E RELEVÂNCIA FUNCIONAL DOS PARÁLOGOS DO GENE *ATAXIN-3*

EVOLUTION AND FUNCTIONAL RELEVANCE OF *ATAXIN-3* PARALOGUES

Universidade de Aveiro
2011

Secção Autónoma de Ciências da Saúde

MARIA INÊS MORGADO OLIVEIRA MARTINS

EVOLUÇÃO E RELEVÂNCIA FUNCIONAL DOS PARÁLOGOS DO GENE *ATAXIN-3*

EVOLUTION AND FUNCTIONAL RELEVANCE OF *ATAXIN-3* PARALOGUES

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biomedicina Molecular, realizada sob a orientação científica da Doutora Sandra Cristina da Silva Martins, Investigadora do Instituto de Patologia e Imunologia Molecular da Universidade do Porto, e a co-orientação da Doutora Ana Gabriela da Silva Cavaleiro Henriques, Professora Auxiliar Convidada da Secção Autónoma das Ciências da Saúde da Universidade de Aveiro.

Esta dissertação foi desenvolvida no Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), um laboratório associado do Ministério Português da Educação e Ciência.

o júri

Presidente

Prof. Doutora Odete Abreu Beirão da Cruz e Silva
Professora Auxiliar com Agregação da Secção Autónoma das Ciências da Saúde da Universidade de Aveiro

Doutora Sandra Cristina da Silva Martins
Investigadora no Instituto de Patologia e Imunologia Molecular da Universidade do Porto

Prof. Doutora Ana Gabriela da Silva Cavaleiro Henriques
Professora Auxiliar Convidada da Secção Autónoma das Ciências da Saúde da Universidade de Aveiro

Doutora Ana Goios Borges de Almeida
Investigadora no Instituto de Patologia e Imunologia Molecular da Universidade do Porto

agradecimentos

Ao Professor António Amorim pela oportunidade de integração no seu entusiasmante grupo de investigação.

À Doutora Sandra Martins por me ter acompanhado incansavelmente, pela sua disponibilidade, paciência e compreensão. Agradeço muito também por me ter inspirado e mostrado toda a sua dedicação e motivação.

À Professora Ana Gabriela Henriques, por desde sempre me ter acompanhado e mostrado os caminhos certos a seguir.

Ao IPATIMUP pelo financiamento, pelo acolhimento e apoio científico sem o qual o sucesso do trabalho experimental não teria sido possível.

A todos os elementos do Grupo de Genética Populacional, que tão bem me acolheram.

À Doutora Odete da Cruz e Silva por me ter proporcionado estes excelentes 5 anos de aprendizagem, bem como todas as oportunidades únicas que tive e me fizeram crescer, quer a nível profissional como pessoal. Agradeço ainda a todos os Professores/Tutores que me acompanharam e ajudaram da melhor maneira possível (principalmente às professoras Sandra Vieira, Ana Gabriela Henriques, Margarida Fardilha e Sandra Rebelo).

À Catarina Seabra, pelo constante e incondicional apoio, conselhos e presença nos momentos bons e menos bons.

Ao Semião por, apesar da distância, ser um grande amigo e estar sempre lá, pronto a ouvir-me nos momentos de felicidade e tristeza.

À Joana Tavares, Chita, Ana Sol e Cátia pela amizade, carinho e apoio. Obrigada ainda por todos os ótimos momentos partilhados, os quais nunca esquecerei.

À Catarina Xavier, Filipa, Marisa, Sofia e Alexandre pelos ótimos momentos passados e o apoio constante ao logo de todo o ano.

A todos os meus colegas de Ciências Biomédicas de Aveiro, que me fizeram crescer e aprender a cada passo.

Aos meus amigos de Aveiro, principalmente à Raleira, ao Recha, Tubrinho, Badaró, Big, Renato, Baka, Jorge e ao David, que me acompanharam desde cedo, e me deram a amizade e os melhores anos de Universidade que mais poderia desejar!

Ao João Fernandes por sempre me apoiar e animar nos momentos mais difíceis, e principalmente, por ser tão importante e me dar tantos dos melhores momentos da minha vida.

À minha irmã Mariana, por ter estado sempre ao meu lado, apoiado e ouvido, principalmente nos momentos em que mais precisava que me ouvissem.

Aos meus pais, pelo amor, ensinamentos sábios e pelo incansável apoio e motivação que sempre me deram ao longo da vida.

Às minhas avós por todo o amor, orgulho e interminável apoio.

palavras-chave

Ataxina-3, Retrotransposição, Genes parálogos, Evolução, Doença de Machado-Joseph

resumo

O gene *ataxin-3* (*ATXN3*; 14q32.1) codifica uma proteína expressa ubiquamente, envolvida na via ubiquitina-proteassoma e na repressão da transcrição. Grande relevância tem sido dada ao gene *ATXN3* após a identificação de uma expansão $(CAG)_n$ na sua região codificante, responsável pela ataxia mais comum em todo o mundo, SCA3 ou doença de Machado-Joseph (DMJ). A DMJ é uma doença neurodegenerativa, autossómica dominante, de início tardio. O tamanho do alelo expandido explica apenas uma parte do pleomorfismo da doença, evidenciando a importância do estudo de outros modificadores. Em doenças de poliglutaminas (poliQ), a toxicidade é causada por um ganho de função da proteína expandida; no entanto, a proteína normal parece ser, também, um dos agentes modificadores da patogénese. O gene *ATXN3* possui dois parálogos humanos gerados por retrotransposição: *ataxin-3 like* (*ATXN3L*) no cromossoma X, e *LOC100132280*, ainda não caracterizado, no cromossoma 8. Estudos *in vitro* evidenciaram a capacidade da ATXN3L para clivar cadeias de ubiquitina, sendo o seu domínio proteolítico mais eficiente do que o domínio da ATXN3 parental.

O objetivo deste estudo foi explorar a origem e a evolução das retrocópias *ATXN3L* e *LOC100132280* (aqui denominadas *ATXN3L1* e *ATXN3L2*), assim como testar a relevância funcional de ambas através de abordagens evolutivas e funcionais. Deste modo, para estudar a divergência evolutiva dos páralogos do gene *ATXN3*: 1) analisaram-se as suas filogenias e estimou-se a data de origem dos eventos de retrotransposição; 2) avaliaram-se as pressões seletivas a que têm sido sujeitos os três parálogos, ao longo da evolução dos primatas; e 3) explorou-se a evolução das repetições CAG, localizadas em três contextos genómicos diferentes, provavelmente sujeitos a diferentes pressões seletivas. Finalmente, para o retrogene que conserva uma *open reading frame* (ORF) intacta, *ATXN3L1*, analisou-se, *in silico*, a conservação dos locais e domínios proteicos da putativa proteína. Ademais, para este retrogene, foi estudado o padrão de expressão de mRNA, através da realização de PCR de Transcriptase Reversa, em 16 tecidos humanos.

Os resultados obtidos sugerem que dois eventos independentes de retrotransposição estiveram na origem dos retrogenes *ATXN3L1* e *ATXN3L2*, tendo o primeiro ocorrido há cerca de 63 milhões de anos (Ma) e o segundo após a divisão Platirrínios-Catarrínios, há cerca de 35 Ma. Adicionalmente, outras retrocópias foram encontradas em primatas e outros mamíferos, correspondendo, no entanto, a eventos mais recentes e independentes de retrotransposição. A abordagem evolutiva mostrou a existência de algumas constrições selectivas associadas à evolução do gene *ATXN3L1*, à semelhança do que acontece com *ATXN3*. Por outro lado, *ATXN3L2* adquiriu codões *stop* prematuros que, muito provavelmente, o tornaram num pseudogene processado. Os resultados da análise de expressão mostraram que o gene *ATXN3L1* é transcrito, pelo menos, em testículo humano; no entanto, a optimização final da amplificação específica dos transcriptos ATXN3L1 permitirá confirmar se a expressão se estende a outros tecidos. Relativamente ao mecanismo de mutação inerente à repetição CAG, os dois parálogos mostraram diferentes padrões de evolução: a retrocópia *ATXN3L1* é altamente interrompida e pouco polimórfica, enquanto a *ATXN3L2* apresenta tratos puros de $(CAG)_n$ em algumas espécies e tratos hexanucleotídicos de CGGCAG no homem e no chimpanzé. A recente aquisição da repetição CGGCAG pode ter resultado de uma mutação inicial

de CAG para CGG, seguida de instabilidade que proporcionou a expansão dos hexanucleótidos.Estudos futuros poderão ser realizados no sentido de confirmar o padrão de expressão do gene *ATXN3L1* e de detetar proteína endógena *in vivo*. Adicionalmente, a caracterização da proteina ataxina-3 like 1 e dos seus interatores moleculares poderá povidenciar informação acerca da sua relevância no estado normal e patológico.

keywords

Ataxin-3, Retrotransposition, Paralogue genes, Evolution, Machado-Joseph disease.

abstract

Ataxin-3 gene (*ATXN3*; 14q32.1) encodes a ubiquitously expressed protein involved in the ubiquitin-proteasome pathway and in transcription repression. Much attention has been given to *ATXN3* since the identification of an expanded $(CAG)_n$ tract in its coding region, responsible for the most common dominant ataxia worldwide, SCA3 or Machado-Joseph disease (MJD). MJD is an autosomal dominant, late-onset neurodegenerative disorder. The size of the expanded allele explains only part of the disease pleomorphism, highlighting the importance of studying other modifiers. In polyglutamine (polyQ) diseases, toxicity is caused by gain of function of the expanded protein; however, the normal protein appears to be one of the pathogenesis-modifying agents.

The gene *ataxin-3* has two human paralogues, generated by retrotransposition: *ataxin-3 like* (*ATXN3L*) on the X chromosome of humans and the yet uncharacterized *LOC100132280* on chromosome 8. An *in vitro* study showed the ability of ATXN3L to cleave ubiquitin from their substrates, with its proteolytic domain even more efficient than the domain of the parental ATXN3.

The aim of this study was to explore the origin and evolution of *ATXN3L* and *LOC100132280* retrocopies (here named *ATXN3L1* and *ATXN3L2*) and to test the functional relevance of both human paralogues through evolutionary and functional approaches. Thus, to study the evolutionary divergence of *ATXN3* paralogues, we have 1) analysed their phylogeny and estimated the time of the retrotransposition events; 2) assessed the selective constraints that have been underlying all the three paralogues over primate evolution; and 3) explored the evolution of the $(CAG)_n$ tracts placed on three different chromosomal backgrounds within paralogue genes that have, probably, been under different selective pressures. Finally, for the retrogene that conserved the ORF, *ATXN3L1*, we analyzed the protein domain conservation and the mRNA expression pattern by performing Reverse Transcriptase-PCR in 16 human tissues.

Our results suggested that two independent retrotransposition events have been on the origin of ATXN3L1 and ATXN3L2, the first occurred about 63 million years ago (MYA) and the second after the Platyrrhini-Catarrhini split, about 35 MYA. In addition, several other retrocopies have been found in primates and other mammals, but additional independent and younger retrotransposition events seemed to be on their origin. Our evolutionary studies suggested that *ATXN3L1* has been under some selective constrains over primate evolution, as the parental *ATXN3*. *ATXN3L2* gained premature stop codons that seem to have turned it into a pseudogene. In addition, we confirmed that *ATXN3L1* is a transcriptionally-active retrogene since we observed mRNA expression, at least, in human testis. A refined optimization of the methodology to specifically amplify *ATXN3L1* cDNA will be, however, necessary to assess the complete expression pattern of this retrogene. As for the $(CAG)_n$ tract, the *ATXN3* paralogues presented different evolutionary patterns: *ATXN3L1* showed a poorly polymorphic and highly interrupted tract across the primate lineage, whereas *ATXN3L2* presented a pure $(CAG)_n$ in some species and a polymorphic hexanucleotide repeat in humans and chimpanzees. This recent acquisition of a repetitive CGGCAG resulted, probably, from a CAG to CGG mutation followed by instability that encompassed the six bases instead of the CAG alone.

Future studies may be performed in order to confirm the expression pattern of *ATXN3L1* and to detect the endogenous protein *in vivo*. Further characterization of ataxin-3 like 1 and its molecular interactors will give us insight into the actual relevance of this protein in normal and/or pathological states.

Contents

Abbreviations 3

Chapter 1 - Introduction 5

1.1 - Mechanisms of retroposition 8

1.2- Machado-Joseph disease 10

1.2.1 - Clinical presentation and epidemiology 10

1.2.2 - Molecular genetics and pathogenesis 11

ATXN3 gene 11

Mechanisms of *ATXN3* mutation 12

Ataxin-3 protein and its physiologic role 13

Polyglutamine expansion of ATXN3 and neural cell death 14

1. 3 – Studies on the relevance of *ataxin* paralogues in SCAs 15

Aims 17

Chapter 2 - Material and methods 19

Subjects 21

Part 1 – Evolutionary approach 22

Compilation and alignment of *ATXN3L1* and *ATXN3L2* sequences 22

Phylogenetic trees and genetic distances 23

Synteny of *ATXN3*, *ATXN3L1* and *ATXN3L2* 23

dN/dS ratio (omega) calculations 23

Primer design, DNA amplification and sequencing 24

Nucleotide diversity of human *ATXN3* retrocopies 28

$(CAG)_n$ tract analysis in *ATXN3*, *ATXN3L1* and *ATXN3L2* 28

Part 2 – Functional approach 28

ORF prediction for *ATXN3L1* and *ATXN3L2* 28

Conservation of the ATXN3L1 putative protein 28

mRNA expression of *ATXN3L1* 29

Chapter 3 - Results 32

Part 1 – Evolutionary history of *ATXN3* paralogues 33

1. Onset of *ATXN3* retrocopies 33

1.1. Identification of *ATXN3L1* and *ATXN3L2* orthologues in the primate lineage ... 33

1.2. Phylogeny of *ATXN3* paralogues in primates 39

1.3. Identification of *ATXN3* retrocopies in other mammals 40

2. *ATXN3* transcripts involved in the retrotransposition events 45

3. Selective signatures underlying *ATXN3L1* and *ATXN3L2* 45

4. Nucleotide diversity of *ATXN3L1* and *ATXN3L2* 47

5. Evolution of the $(CAG)_n$ tract in *ATXN3*, *ATXN3L1* and *ATXN3L2* 48

Part 2 – Insights into *ATXN3L1* and *ATXN3L2* functional relevance 50

1. ORF predictions for *ATXN3L1* and *ATXN3L2* 50

2. Analysis of ATXN3L1 putative protein domains in comparison the parental ATXN3 52

3. mRNA expression profile of *ATXN3L1* in humans 53

Chapter 4 - Discussion 55

Chapter 5 - Final remarks and future perspectives 63

References 69

Appendix 73

Abbreviations

AO – Age-at-onset

ATXN1 – Ataxin-1

ATXN1L – Ataxin-1 like

ATXN3 – Ataxin-3

ATXN3L1 – Ataxin-3 like 1

ATXN3L2 – Ataxin-3 like 2

BLAST - Basic local alignment search tool

BLAT - BLAST-like alignment tool

bp – base pairs

CIC – Capicua

DUB – Deubiquitinating enzyme

JD – Josephin domain

JOSD1 – Josephin 1

JOSD2 – Josephin 2

L1 – Long interspersed nuclear element 1

LINE – Long interspersed nuclear element

LTR – Long terminal repeats

MJD – Machado-Joseph disease

MYA – Million years ago

NCBI – National Center for Biotechnology Information

NES – Nuclear export signal

NI – Nuclear inclusions

NLS – Nuclear localization signal

ORF – Open reading frame

PolyQ – poly glutamine

RT-PCR – Reverse transcriptase polymerase chain reaction

SCA1 – Spinocerebellar ataxia type 1

SCA3 – Spinocerebellar ataxia type 3

SNP – Single nucleotide polymorphism

STR – Simple tandem repeat

SVA – (SINE/VNTR/*Alu*)

Ub – Ubiquitin

UCSC –University of California, Santa Cruz

UIM – Ubiquitin interacting motif

UTR – Untranslated region

Chapter 1

Introduction

Chapter 1

Introduction

New genes are thought to contribute to the origin of adaptive evolutionary innovations and, thus, to lineage- or species-specific phenotypic traits. Traditionally, the origin of these new genes is associated with gene duplication, but other mechanisms have recently received increasing attention, such as retrotransposition, which may play an important role in mammal genome evolution.[1, 2, 3] Retrotransposition consists in the re-integration of reverse transcribed mRNA molecules in the genome.[4] At the sequence level, retroposed copies (retrocopies) can be intact or not, depending on the presence of mutations or premature stop codons. Still, if the open reading frame is conserved, retrocopies lack many of the genetic features of their parental genes (such as introns and regulatory elements) and, for this reason, they have been considered as confounding factors for a long time. Most retrocopies turned into pseudogenes in mammals, but some of them have recruited upstream regulatory elements and became functional (these are commonly called retrogenes).[5] Indeed, many recent studies have shown that a larger part of retrotransposons than the previously predicted are functional, and that they can modulate other pre-existing cellular factors.[1, 5]

This thesis project is focused on the study of two retrocopies of *ataxin-3* (*ATXN3*), a ubiquitously expressed gene that codes a protein mainly involved in deubiquitination and in transcription repression. *ATXN3* is also the gene responsible for Machado-Joseph disease (MJD), when its coding repetitive CAG tract is expanded above 61 repeats. MJD or spinocerebellar ataxia type 3 (SCA3) is a late-onset neurodegenerative disorder, characterized by a large pleomorphism. Thus, the search for disease modifiers has been a major point of interest in this disorder.

The origin, features and functional relevance of the poorly studied *ATXN3* retrocopies, *ATXN3L* and *LOC100132280* (here named *ATXN3L1* and *ATXN3L2*, respectively) remain completely unexplored until these days. With this project we aimed at characterizing *ataxin-3* paralogues in the human genome and throughout the evolution to gain insight into the potential roles that they have been playing.

1.1 - Mechanisms of retroposition

The key retrotransposition enzyme stems from different types of retrotransposable elements, depending on the organism. These elements can be subdivided in two groups, distinguished by the presence or absence of long terminal repeats (LTRs). In humans, LTR elements are endogenous retroviruses (HERVs), presently with very limited activity, which accounts for only ~8% of the genome. On the other hand, human non-LTRs include long interspersed nuclear element 1 (LINE 1 or L1), *Alu* and SVA (an element composed of a short interspersed region, a variable number of tandem repeats region and an *Alu*-like region - SINE/VNTR/*Alu*) elements, which collectively account for approximately one-third of the human genome (Figure 1).[6]

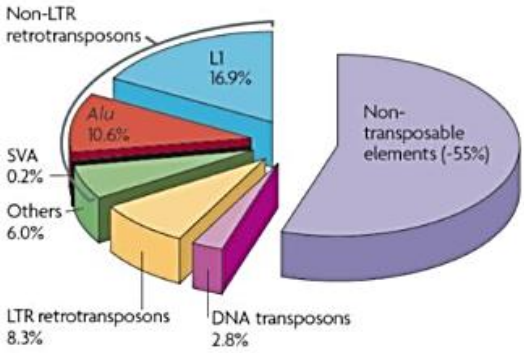

Figure 1 - Transposable and non-transposable element content of the human genome. About 33.7% of the human genome can currently be recognized as being derived from non-LTR retrotransposable elements (Cordaux et al., 2009).[6]

L1, widely present in mammals (about 25% of their genome), are responsible for a burst on the number of retrocopies in mammalian lineages.[2] These retrotransposable elements possess a reverse transcriptase and a endonuclease activity that can recognize any polyadenylated mRNA.[7] After gene transcription, the respective RNA is reverse-transcribed into DNA by the L1 elements, and the resulting cDNA is inserted into the genome at a new location (Figure 2).[4, 7]

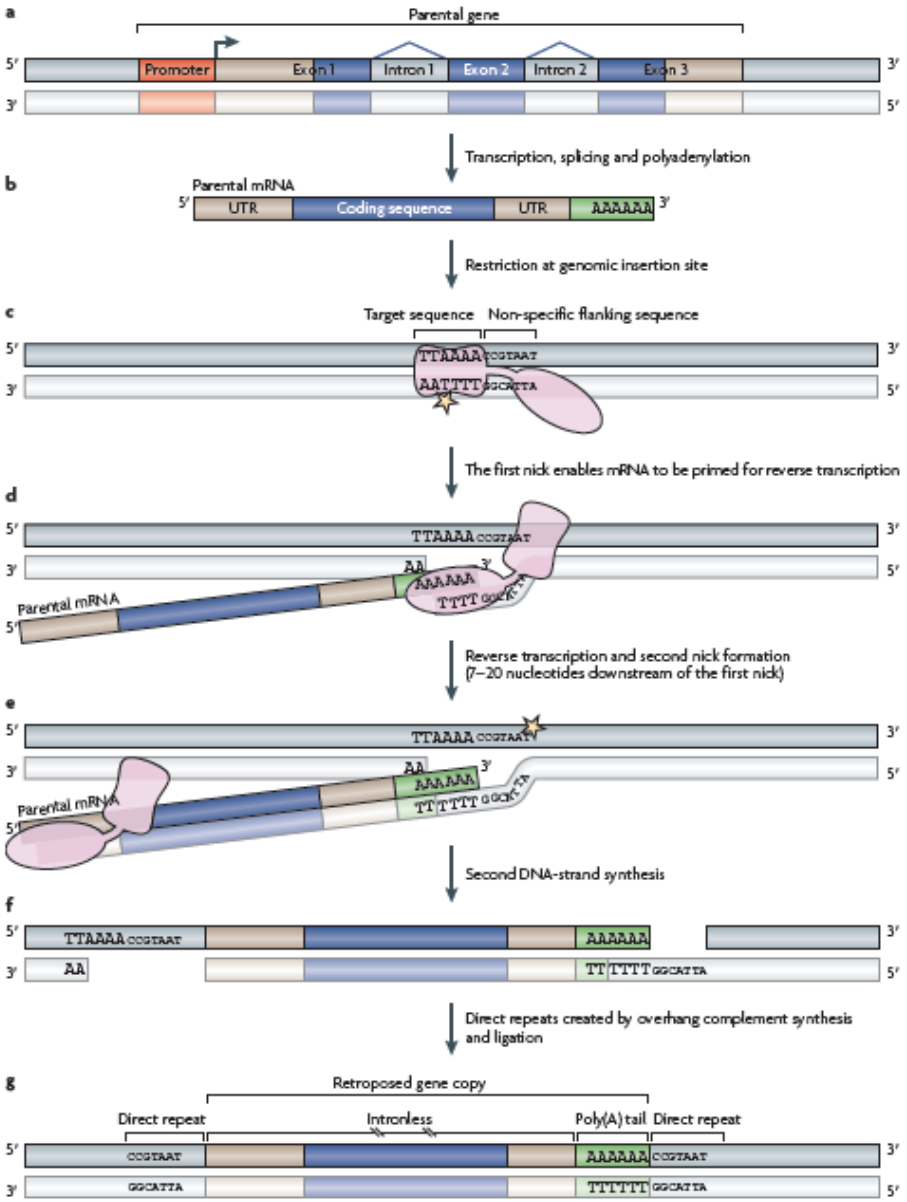

Figure 2 - Retrotransposition mechanism: a) retroposition is initiated with the transcription of a parental gene by RNA polymerase; b) a mature mRNA is produced, with the processing of the resulting RNA (by splicing and polyadenylation); c) L1 endonuclease domain (pink rectangle) mediates retroposition, by creating a first nick (yellow star) at the genomic site of insertion at the TTAAAA target sequence; d) this nick enables the mRNA to be primed for reverse transcription by the L1 reverse transcriptase domain (pink oval), which uses the parental mRNA as a template; e) second-strand nick generation, f) second DNA-strand synthesis; g) cDNA synthesis in the overhang regions created by the two nicks. (adapted from Kaessmann, 2009)[1]

To become expressed at a significant level and in a meaningful way, a new retrogene needs to obtain a core promoter and probably other elements, such as enhancers, that regulate its expression.[1] Generally, the expression of a retrocopy might benefit from: 1) its insertion into intronic sequences of host genes, enabling it to be integrated into new splice variants of the host gene; 2) its insertion into actively transcribed regions with an open chromatin structure, as this increases accessibility for the transcriptional machinery; 3) the recruitment of distant promoters in the genomic neighbourhood via the acquisition of a new untranslated exon–intron structure; 4) the recruitment of proto-promoters from retrotransposons or CpG island; 5) the inheritance of parental promoters through alternative transcriptional start sites used by the parental gene; and 6) *de novo* promoter evolution in the upstream flanking region of the insertion site by single nucleotide substitutions.[1, 5] Since retroposed copies often need to recruit regulatory elements to become transcribed, they are prone to accumulate genetic variants and to evolve new expression patterns relatively to the traditional gene duplication. As a consequence, novel functional roles can emerge from the transcription of new formed retrogenes.[1] Studies have revealed that retrogenes seem to evolve biased functions related to male functions, but others functions have also been described, for example, related to the brain. [8, 9]

In addition, to be heritable and hence of evolutionary relevance, retrotransposition needs to occur in the germ line. The fact that retrotransposition relies on duplication through an mRNA intermediate implies that only genes expressed in the germ line can be the source of new retrocopies.[1]

1.2- Machado-Joseph disease

1.2.1 - Clinical presentation and epidemiology

Machado-Joseph disease (MJD, also called spinocerebellar ataxia type 3, SCA3) is the most common type of spinocerebellar ataxia. SCAs are a large and complex group of late-onset diseases, characterized by progressive cerebellar dysfunction of afferent and efferent pathways, variably associated with other symptoms of the central and peripheral nervous systems. Other nervous system structures affected include the basal ganglia, brainstem nuclei, pyramidal tracts and post superior column and anterior horn of the spinal cord, as well as peripheral nerves.[10, 11] Although there are sporadic forms of ataxia, the term SCA is most often used to refer the

hereditary forms, and in particular the autosomal dominant ataxias.[8] Nearly 30 subtypes of SCAs have been described, differently classified based on the underlying causative mutation. MJD is one of the polyglutamine diseases since the causative mutation is a CAG repeat expansion in ATXN3 gene, which encodes a stretch of glutamine amino acids in the corresponding protein.[12, 13]

In the particular case of MJD, the major signs are the progressive cerebellar ataxia and pyramidal signs. Minor, but more specific, features are external progressive ophtalmoplegia, dystonia, intention fasciculation-like movements of facial and lingual muscles, as well as bulging eyes but these symptoms can vary among patients. The mean age at onset (AO) of the disease is around 40 years, although it can vary greatly, with extremes of 4 and 70 years.[11] The mean survival time is 21 years (ranging from 7 to 29).[14] Thus, MJD is characterized by a high degree of pleomorphism, not only in the variability of the AO, but also in the neurological signs presented and in the resulting degree of incapacity.[10, 15]

SCAs are considered rare disorders, with estimated prevalence described as varying from 0.3 to 2.0 per 100,000 individuals. The relative frequency of MJD among SCAs varies also largely among populations. It is higher in Brazil (69-92%),[16] Portugal (58-74%),[17] Singapore (53%),[18] China (48-49%),[19] the Netherlands (44%),[20] Germany (42%)[21] and Japan (28-63%)[22] but even within each country, the geographic distribution pattern of MJD is not homogeneous.

Haplotype-based studies have suggested that two independent-origin mutations may explain the current worldwide geographic distribution of MJD. The first occurred about 6000 years ago in Asia (TTACAC or Joseph lineage, observed in the 5 continents); and the most recent, with less than 2000 years old, is responsible for the presence of MJD in Portugal (together with the most ancient mutation) and in a few other populations mostly linked to Portugal (GTGGCA or Machado lineage).[23]

1.2.2 - Molecular genetics and pathogenesis

ATXN3 gene

ATXN3 was located in the long arm of chromosome 14 (14q32.1), by Takiyama in 1993.[24] The genomic structure of *ATXN3* spans about 48 kb, with the most frequent transcript containing 11 exons.[25] Two additional exons, 6a and 9a, have recently been described (Figure 3).[26]

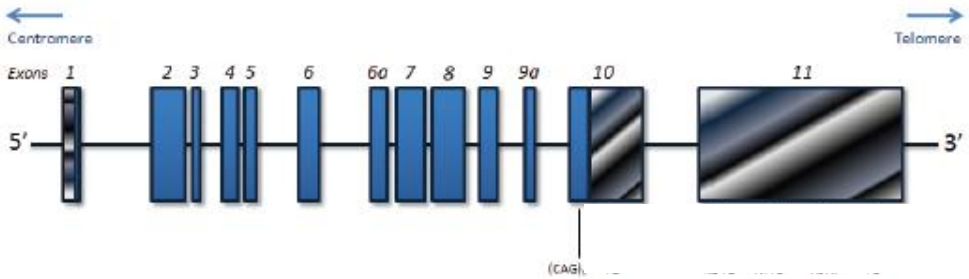

Figure 3 - *ATXN3* gene structure, representing the first described exons and the two new exons: 6a and 9a; the (CAG)n tract is localized in the exon 10. Transcribed regions are represented in blue, and not-transcribed and UTR regions in grey. (adapted from Bettencourt, 2011)[13]

The expansion of a $(CAG)_n$ tract within *ATXN3* exon 10 is the causative mutation of MJD.[27] Usually, wild-type alleles range from 12 to 44 CAG repeats, whereas expanded alleles comprise 61 to 87 CAGs. Intermediate size alleles are rare, and their role in disease presentation is still under debate.[28, 29] The size of the expansion has been found, until date, the parameter that best correlates with the AO of patients; however, it explains only 45-76% of AO variability suggesting a major influence of other factors.[30, 31] Gender was initially proposed to be one of these factors[32], but this was later contradicted, as familial factors were suggested to mask the gender effect.[33] It was demonstrated, instead, that variation in AO accounted for an effect common to a sibship, independently of the CAG repeat length. Less controversial is the dosage effect on disease presentation: homozygous patients with two mutant alleles show, most frequently, a more severe disease phenotype, with earlier onset than those presenting only one mutated allele.[34]

The process underlying the $(CAG)_n$ instability across successive generations is currently thought to involve the generation of abnormal DNA structures in replication slippage, as well as DNA repair and recombination, acting either separately or in combination.[35]

Mechanisms of *ATXN3* mutation

The *ATXN3* gene is present in a large range of living organisms, but the expansion seems to have occurred only in the human lineage, after the divergence from other hominids. A recent study in human *ATXN3* has shown a bimodal distribution of $(CAG)_n$ alleles for all most frequent stable SNP-defined haplotypes. When flanking STR diversity was compared between modal $(CAG)_n$ alleles within each lineage, little differences have been noticed while alleles one or two repeats apart showed much higher genetic distances. Based on these results, a multistep mutation mechanism was suggested for the evolution of this *locus* in humans.[36] Little is known, however,

about the mutation process underlying the evolution and instability over mammalian and even primate evolution that led to the currently observed $(CAG)_n$ tract configuration and level of polymorphism. In *Pan troglodytes*, *Gorilla gorilla*, *Mus musculus* and *Gallus gallus*, the repetitive tract in the *ATXN3* gene is conserved, but shorter than in the human homologue.[37]

Ataxin-3 protein and its physiologic role

The *ATXN3* gene encodes for ataxin-3, a protein with an approximate molecular weight of 42 kDa (in its longest form) that is ubiquitously expressed in neuronal and non-neuronal tissues.[25, 38] In the human brain, ataxin-3 is widely expressed in different regions, with variable expression levels. Granular layer and purkinje cells of cerebellum, hippocampus, striatum and pyramidal cells of motor cortex seems to be the regions of higher expression levels, whereas mesencephalon (substantia nigra pars compacta), occipital cortex, globus pallidus (internal and external) and white matter of cerebellum present lower expression patterns.[25, 38] In the *Drosophila* model, increased levels of the mutant ATXN3 expression have been shown to underlie more severe degeneration and earlier onset of protein accumulation, suggesting that abnormal accumulation of the mutant protein is central to disease and degeneration.[39]

At the cellular level, ATXN3 was found to be present in the cytoplasm (mitochondria included) and nucleus, with varying degrees of predominance depending on the cell type.[40, 41] In human brain cells, ATXN3 is present mainly in the perikarya, but it was also detected on proximal processes, axons and nuclei. This heterogeneity suggests that the regulation of ATXN3 expression levels and localization may have functional relevance.[38]

ATXN3 participates in cellular protein quality control pathways but its biological function has not yet been completely understood. Studies in knockout mice for *ATXN3* have demonstrated a cytoplasmic increase in ubiquitinated proteins, supporting an *in vivo* role of this protein in the ubiquitin/proteasome pathway as a deubiquitinating enzime (DUB).[42, 43] The covalent attachment of ubiquitin to a protein is a reversible signal that can alter a protein's function, control its trafficking, or mark it for degradation. Once attached, ubiquitin can be removed by DUBs, thereby generating a dynamic balance in ubiquitin signaling pathways.[44] Ataxin-3 can also regulate its own cellular turnover in a ubiquitin-dependent manner; abolition of its catalytic activity disrupts this regulation.[45] Hence, it is clear that ataxin-3 DUB activity is physiologically relevant. In addition, MJD mouse models have shown transcriptional deregulation, confirming a possible important role of ATXN3 in transcription.[46]

ATXN3 has a papain-like fold and is essentially composed by a structured globular N-terminal domain, followed by a flexible unfolded C-terminal tail. The N-terminal domain, designed Josephin domain (JD), displays ubiquitin (Ub) protease activity, while the flexible tail presents two or three Ub-interacting motifs (UIMs), depending on the isoform, and a polyQ region of variable length.[47, 48] Notably, the most common isoform found in the human brain has three UIMs in the C-terminal region (Figure 4).[49] The highly conserved catalytic triad on the Josephin domain possesses the predicted catalytic aminoacids found in cysteine proteases: Cys14, His119 and Asn134.[50] Two other serine residues are important in the UIM regions for the interaction with Ub: Ser232 in UIM1 and Ser260 in UIM2.[51]

The Josephin domain is shared by three other human proteins (Ataxin-3 like, ATXN3L; Josephin-1, JOSD1; and Josephin-2, JOSD2), which together with ATXN3 form the Josephin family of DUBs.[52]

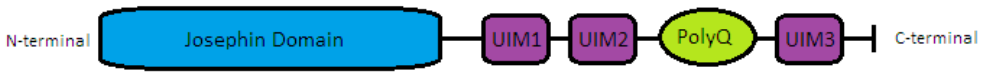

Figure 4 - Ataxin-3 protein structure. UIM – ubiquitin interacting motif; PolyQ – poly-glutamine tract.

A total of 56 human alternative splicing variants of *ATXN3* have been identified with, at least, 21 isoforms expected to be translated. The biological relevance of all these variants remains unknown.[13, 53]

Polyglutamine expansion of ATXN3 and neural cell death

The expanded ATXN3 gains a neurotoxic function through yet unclear mechanisms. In brains of MJD patients, ATXN3 forms nuclear inclusions (NI) present only in neurons, however, more recently axonal inclusions have also been observed in fibers known to degenerate.[41] It is known, however, that this toxicity is linked to abnormal folding and aggregation of the mutant protein to itself as to other members of the cell quality control system, like proteasome constituents, ubiquitin and molecular chaperones.[54, 55] Additionally, intracellular inclusions have been linked to the disease pathogenesis as a major means of gain of toxicity by the expanded protein, through several possible mechanisms: a) hindrance of transcription, by direct decrease of gene expression regulation or through sequestration of other molecules involved in transcription regulation;[56] b) transcription alteration via formation of histone-deacetylating repressor complexes on target

chromatin regions;[57, 58] c) interference in the axonal transport, resulted from motor protein titration and physical blocking;[59] and d) other disturbances caused by the recruitment of Ub-binding proteins (since inclusions are heavily ubiquitinated) or other polyQ-containing proteins.[60] In contrast, NI have been recently suggested to play a protective role on MJD by sequestering the mutant proteins from the toxic interactions.[61]

Although the polyQ expansion appears to be the triggering factor leading to the development of MJD, ATXN3 regions outside the polyQ, as well as other protein properties seem also to define the development of MJD and its particular aspects. In animal and cellular models, the C-terminal fragment of expanded ataxin-3 alone was shown to be more toxic than the full-length protein, suggesting that ataxin-3 cleavage might be a contributor factor for the MJD pathology.[62-64] Indeed, the ataxin-3 expanded polyQ tract was shown to exhibit several potential cleavage sites for caspases.[65] Another study has also demonstrated a possible involvement of the N-terminal on MJD pathogenesis, since a mutant mouse model with an ATXN3 truncated form, containing only its N-terminal, presented typical neurological symptoms.[66]

Therefore, clinical variability of MJD is only partially explained by the size of the $(CAG)_n$ tract and by the ATXN3 protein itself, which leaves a residual variance that should be explained by other unknown factors. In addition to the wild-type protein, the three other JD-like containing proteins may exert similar functions to ATXN3 and compensate for its absence in knockout models.[52, 67] In fact, several studies on animal models have shown the importance of the normal non-mutated protein in the MJD pathogenesis. An interesting feature of non-expanded ataxin-3 is that it is also recruited to NI in several PoliQ diseases.[68] In the *Drosophila* model, for example, normal ATXN3 has a neurodegeneration repression effect *in vivo* (a protective role) by retarding and reducing the accumulation of the pathogenic protein. This suppressor activity requires ubiquitin-associated activities at the normal protein, since, by mutating the two most important serine residues in UIM1 and UIM2 (Ser232 and Ser260, respectively), the ability of ataxin-3 to suppress degeneration become compromised.[69]

1. 3 – Studies on the relevance of *ataxin* paralogues in SCAs

Studies focusing on the modulation of parental genes by the respective expressed paralogues may be important to gain insight into the mechanisms by which *ATXN3* retrocopies may be functional relevant. The gene responsible for spinocerebellar ataxia type 1 (SCA1), *ATXN1*, has an

evolutionary conserved paralogue called *ATXN1-like*. Recent studies in fly and mouse models have shown that the overexpression of *ATXN1L* partially suppresses the neuropathology caused by the polyglutamine-expanded ATXN1, by inducing sequestration of polyglutamine-expanded ATXN1 into nuclear inclusions.[70, 71] Similarly to MJD, molecular and genetic data suggested that SCA1 pathogenesis is caused by a gain of function mechanism, while other studies have shown, however, that the deletion of wild-type protein enhances disease pathogenesis. In SCA1, additionally, it has been suggested that increased levels of the *ATXN1L* retrogene ameliorate the clinical phenotype. These data indicated that both gain and partial loss of function may contribute to the disease progress, with the partial loss of *ATXN1* alone being sufficient to cause some transcriptional changes that are pathogenic in the cerebellum.[70]

The mechanism underlying the functional relevance of ATXN1L may be related to the shared protein interactions by ATXN1 and ATXN1L. The two proteins were shown to share the transcriptional regulator Capicua (CIC) that is known to form complexes (Atxn1-CIC and Atxn1L-CIC) that bind the promoters of target genes, repressing them effectively. Thus, in *ATXN1*$^{-/-}$ mice, the overexpression of *ATXN1L* rescues the level of stable complexes with CIC, maintaining its function as a transcription repressor. This mechanism is an evidence that transcribed paralogues can, in some degree, replace the loss-of-function observed in SCA1.[70] Additionally, it was proposed that, as ATXN1L compete with wild-type and mutant ATXN1 for association with CIC, the increased levels of free mutant ATXN1 lead to an increase of aggregation and nuclear inclusions.[70, 71] This data provide genetic evidence that evolutionary conserved paralogues may have an important role in the mechanisms of pathogenesis of neurodegenerative diseases.

As for ATXN3 and MJD/SCA3, Weeks (2011) performed an in vitro study of the human ATXN3L retrocopy (here named ATXN3L1) by analysing the crystal structure of its predicted protein.[52] The authors found that although ataxin-3 and ATXN3L adopt similar folds, they bind ubiquitin in different, overlapping sites. Additionally, by mutating ataxin-3 at selected positions (and introducing the corresponding ATXN3L residue), only three mutations were sufficient to increase the catalytic activity of ataxin-3. This suggested that ATXN3L Josephin domain could be significantly more efficient than the ataxin-3 domain itself, opening a broad road for further research concerning the study of this ATXN3 paralogue.

Aims

Two human copies of the *ATXN3* gene, formed by retrotransposition events, were recently discovered and annotated in databases. However, they remain unexplored as their description has not been received much attention. Therefore, we wanted to better characterize these retrocopies since, if transcribed, they may play a role in the pathogenesis of MJD. Additionally, by comparing the $(CAG)_n$ tracts among the 3 paralogue sequences, we hope to gain insight into the repeat expansion mechanism that occurred at *ATXN3* in the human lineage.

Therefore, the aims of this project were:

1) to estimate the onset of the retrotransposition events that originated the two human *ATXN3* paralogues;

2) to compare the rates of evolution and selective constrains underlying the three *ATXN3* paralogues throughout the primate lineage;

3) to gain insight into the mechanisms that led to the human-specific $(CAG)_n$ expansion and tract configuration of *ATXN3*; and

4) to investigate if the human *ATXN3* paralogues are transcribed and, if so, their mRNA transcription pattern.

Chapter 2

Material and methods

Chapter 2

Material and methods

To achieve the aims of this project, an evolutionary approach (Part 1) was firstly performed, followed by functional assays (Part 2). We have done evolutionary studies to compare the rates of evolution and selective constrains among *ATXN3*, *ATXN3L1* and *ATXN3L2* genes, and to date the onset of the retrotransposition events. We started by aligning homologous sequences from several primate species (followed by more distant mammals) available in databases. Afterwards, we calculated genetic distances and determined both the origin of *ATXN3* retrocopies and the underlying selective pressures throughout evolution. In Part 2, we started by comparing the Open Reading Frames (ORFs) among *ataxin-3* paralogues. For *ATXN3L1*, which showed a conserved ORF, we assessed the mRNA expression pattern in different human tissues, by Reverse Transcriptase-PCR (RT-PCR). The two approaches, Part 1 and Part 2, are described below in detail. As it was previously referred, *ATXN3* paralogues, described as *ATXN3L* and *LOC100132280*, will be named as *ataxin-3 like 1* (*ATXN3L1*) and *ataxin-3 like 2* (*ATXN3L2*), respectively, for easier reading.

Subjects

We analysed a total of 49 healthy human subjects from European (n=6) and Asian (n=43) origins. DNA samples were available at our lab, previously coded to ensure confidentiality. Informed consent was provided by all individuals.

As for non-human primate species, we analysed DNA samples from *Pan troglodytes* (chimpanzee) (n=2), *Gorilla gorilla* (gorilla) (n=3), *Pongo abelii* (orangutan) (n=2), *Macaca mulatta* (rhesus monkey) (n=5), *Macaca fascicularis* (cynomolgus monkey) (n=3), *Callithrix jacchus* (marmoset) (n=3) and *Papio* (baboon) (n=3). DNA was quantified with NanoDrop spectrophotometer (Thermo Scientific) to further make work aliquots with a DNA concentration of approximately 30 ng/µL.

Part 1 – Evolutionary approach

Compilation and alignment of *ATXN3L1* and *ATXN3L2* sequences

To study the origin of the retrotransposition events, we first needed to identify *ATXN3* paralogues in public databases. We started by searching for annotated *ATXN3L1* and *ATXN3L2* orthologues in primates and other mammalian species. Sequences of *ATXNL1* were available in Ensembl database (http://www.ensembl.org/) for *Homo sapiens*, *Pan troglodytes*, *Pongo abelii*, *Macaca mulatta*, *Callithrix jacchus* and *Nomascus leucogenys*. The sequence from human *ATXN3L2*, termed LOC100132280, was obtained from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/), but no other annotated orthologues were found in either Ensembl or NCBI Gene databases. Therefore, we performed BLAT and BLAST algorithms provided by University of California, Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/) and NCBI databases, respectively. This way, we obtained not yet annotated homologous sequences of *ATXN3L2* in the genome of several primate species. Following the same strategy, we searched for additional *ATXN3L1* sequences of primate species not retrieved from Ensemble. The same approach was also applied to search for *ATXN3L1* and *ATXN3L2* orthologues in other mammals; in this case, however, we used mostly the Trace Archives specialized search of BLAST algorithm (NCBI).

The alignment of the collected sequences (coding regions for *ATXN3*) was performed in Geneious Pro, 5.5.6 software and the homology (percentages of pairwise identity and identical sites) between sequences was obtained among paralogues of primate species. This allowed to discern whether two independent retrotransposition events occurred or, alternatively, one of the copies duplicated from the most ancient retrocopy.

Next, to identify the transcript(s) of the parental gene involved in the retrotransposition events, we aligned *ATXN3L1* and *ATXN3L2* with the 21 protein coding transcripts described for *ATXN3*. Transcripts were retrieved from Ensembl: ATXN3-001, ATXN3-003, ATXN3-004, ATXN3-005, ATXN3-008, ATXN3-015, ATXN3-017, ATXN3-019, ATXN3-020, ATXN3-026, ATXN3-029, ATXN3-032, ATXN3-201, ATXN3-202, ATXN3-203, ATXN3-204, ATXN3-205, ATXN3-206, ATXN3-207, ATXN3-208 and ATXN3-209. In humans, ATXN3-001 has been described as the most common transcript, followed by ATXN3-003, 004 and 005.

Phylogenetic trees and genetic distances

Based on the aligned sequences, we calculated genetic distances among orthologues and paralogues, and constructed phylogenetic trees for all species, by using the same software, Geneious Pro 5.5.6. To build the trees, we applied the Neighbor-Joining method and the Tamura-Nei genetic distance model (chicken, *Gallus gallus,* was included as outgroup). This approach was performed to identify *ATXN3L1* and *ATXN3L2* orthologues. Since the Josephin domain is the most conserved region of *ATXN3* and the $(CAG)_n$ region is usually a source of much variation, we assessed genetic distances and built phylogenetic trees for all three paralogues by using: 1) the entire sequences; 2) the sequences without the $(CAG)_n$ tract; and 3) the Josephin domain alone.

Synteny of *ATXN3*, *ATXN3L1* and *ATXN3L2*

We analysed the genes flanking *ATXN3*, *ATXN3L1* and *ATXN3L2* in a region of, at least, 600000 base pairs (bp) to elucidate if these regions were conserved along the primate lineage, confirming thus gene orthologies. This was performed by using NCBI Gene database, UCGC Genome Brower genomes and Ensembl Synteny analyses. If two flanking regions of a given gene have obvious collinear genomic features, they are syntenous. From this information, therefore, we can conclude that predicted orthologue genes having syntenous flanking regions have higher probability to be real orthologues. In addition, the same approach was applied to analyse sequences of non-primate mammals: their flanking regions were compared to those retrieved from primates in order to identify if they were ancestral retrocopies of *ATXN3L1* and/or *ATXN3L2*, or alternatively, originated from independent retrotransposition events.

dN/dS ratio (omega) calculations

Evolutionary pressures on proteins can be quantified by the ratio of substitution rates at non-synonymous and synonymous sites. We calculated dN/dS values (omega) of *ATXN3* and *ATXN3L1* genes for different primate species, using the DnaSP v5 software. The dN value represents the rate of non-synonymous substitutions (nucleotide changes that result in amino acid alterations) per site, whereas dS is the rate of synonymous substitutions (silent alterations that do not change the amino acid residue). Thus, a dN/dS ratio (omega) is expected to exceed unity if substitutions are equally frequent at all three codon positions, indicating that no constrains have been underlying the gene. On the other hand, omega less than unity is expected if selection suppresses protein changes, suggesting that the gene function has been kept in check. These calculations

allowed us to compare the selective constrains that have been underlying *ATXN3* paralogues in several lineages of primates. As *ATXN3L2* presented a disrupted ORF, with the start codon mutated and several premature stop codons, we did not proceed with this calculation for this retrocopy.

Primer design, DNA amplification and sequencing

Based on the alignment of gene orthologues, primers were designed to specifically amplify *ATXN3* (exonic regions), *ATXN3L1* and *ATXN3L2* in all primate species. Exons of *ATXN3* annotated in NCBI database represent only those included in the most frequent transcript (11 exons). When analysing all *ATXN3* protein coding transcripts represented in Ensembl data base, we retrieved other exons transcribed only in some transcripts (protein coding) and others only predicted to be transcribed. As reference, in this study, we used all *ATXN3* exons annotated in Ensemble (21 in total), although for primer design, we selected only exons known to be transcribed. Figure 5 summarizes this information.

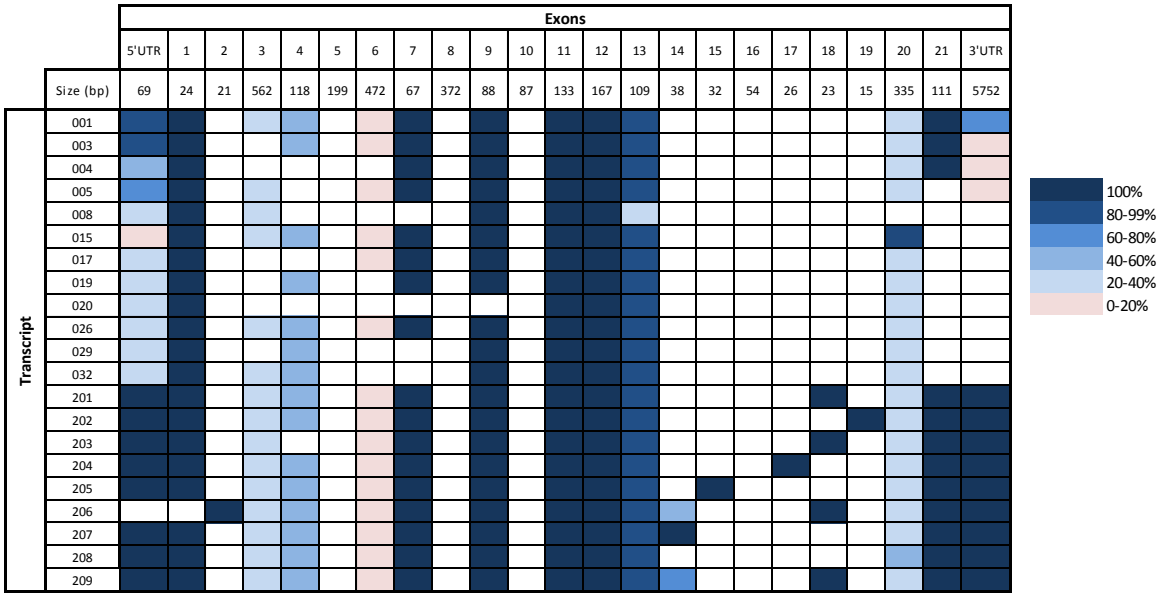

Figure 5 – Exons of the ATXN3 protein coding transcripts, according to Ensemble database. As some transcripts do not contain the entire exon sequence, percentages of coding sequences relatively to the total number of nucleotides of the exon were calculated and represented with different colors. Color legend is on the right side of the figure.

For primate species that lacked the annotated sequence of one or more *ATXN3* paralogues (namely chimp, gorilla, orangutan, gibbon, marmoset, cynomolgus monkey and baboon), we performed sequencing; the same PCR conditions and oligo pairs were used, regarding the sequence conservation of the selected primers for most of the primates.

For primer design, we selected the following criteria: 20 to 24 base pairs; a percentage of GC between 45 and 60%; and a melting temperature comprised between 56 and 64°C. The OligoCalc algorithm (http://www.basic.northwestern.edu/biotools/oligocalc.html) was used to calculate melting temperatures and GC percentages, as well as to predict the formation of hairpins and primer dimers. Primers predicted to form these structures were rejected. In addition, to avoid the formation of primer-dimmers, we used the AutoDimmer software to select primers with no such predicted structures. Finally, to analyze primers' specificity for each species, we performed an UCSC Genome Browser *In-Silico PCR* (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start). The list of selected primers is shown in Tables 1 and 2.

Table 1 – Primers designed to specifically amplify *ATXN3* exons.

ATXN3					
Reaction	**$T_{annealing}$**	**Exons**	**Amplicon**	**Primer sequence**	**Species**
Pentaplex	57°C	Part of 2 and 3 Part of 6 9 12 16	F2-R3 F6-R6 F9-R9 F12-R12 F16-R16	F2 – GTGGTAAGCTGAGATTGCTCC R3 – CCAGCTGATGTGCAATTGAGG F6 – CACAACTAACATAGCTACACTTCC R6 – AAGGCTACAGGGCAGATGCT F9 – CCTGGCCAATGTGGCAAATG R9 – CACTGTCATCTAATGTGCCTG F12 – GGTTGCAGTTATTACCAGTGC R12 – GAAATCTAAAGGGAAAGCCCAC F16 – CTGATCCTAGGTCAGAAAACAGA R16 – GGCCGTGTGCTAGTATTTGTTG	*H. sapiens* *P. troglodytes* *G. gorilla* *P. abelii*
Hexaplex	60°C	4 Part of 6 and 7 11 17 18,19 and part of 20 21	F4-R4 F6'-R7 F11-R11 F17-R17 F18-R20 F21-R21	F4 – GCACGCTAATGACAGTTTGTATCC R4 – GGTGAAACCCCACTATCTCTAC F6'– CAACAGTCCAGAGTATCAGAGGC R7 – GACAGGACCTCCCTTTGTTGCC F11 – TCCAGTGTTCTGTGCTGCCTTTT R11 – AGTCGCCAACAACACAAGGACC F17 – GGAAAGGCATCTCTGGGGAG R17 – GAAGTTTGACACGAGCCTGGAC F18 – CCACTCCTGGCCATGATAGGT R20 – GAATGGTGAGCAGGCCTTACC F21 –CTGGTGGCTATCTGGGATTAGGA R21 – GGACCCTATGCTGTAATCACACAG	*H. sapiens* *P. troglodytes* *G. gorilla* *P. abelii*
Duplex 1	64°C	1 3	F1-R1 F3-R3'	F1 – CGTGTCCCCGGCGTTCACTC R1 – AGATCGGCATGGGGGCGACT F3 – GCAAGAAGGCTCACTTTGTGCTC R3' – CAAGGGTGGGGGTGGGGAAA	*H. sapiens* *P. troglodytes* *G. gorilla* *P. abelii*
Duplex 2	64°C	13 and 14 Part of 20	F13-R14 F20-R20	F13 – ACGCCCAGCCAGAAGAGTAG R14 – CTCCTGACCTCAGGCAATCTG F20 – GGCCAGCCACCAGTTCAGGAG R20' – TCCTCTCCTGCCTTGGTTTCCC	*H. sapiens* *P. troglodytes* *G. gorilla* *P. abelii*

Table 2 - Primers designed to specifically amplify *ATXN3L1* and *ATXN3L2* genes.

ATXN3L1* and *ATXN3L2					
Reaction	**$T_{annealing}$**	**Gene**	**Amplicon**	**Primer sequence**	**Species**
Singleplex 1	59°C	*ATXN3L1*	L1F1-L1R1	L1F1 – CTCTAACTAGGATACCAGCAAAG L1R1 – GGAAAAAGTTCTATGGCAAGAGC	*H. sapiens* *P. troglodytes* *G. gorilla* *P. abelii* *M. mulatta* *N. leucogenys* *C. jacchus*
Singleplex 2	57°C	*ATXN3L2*	L2F1-L2R1	L2F1 – CATTAACCAAAGAAAGTGGGATAC L2R1 – GGAATCCTATGCTGTAATCACAC	*H. sapiens* *P. troglodytes* *G. gorilla* *P. abelii* *M. mulatta* *N. leucogenys*
Singleplex 3	57°C	*ATXN3L2*	cL2F1-cL2R2	L2F1 – CATTAACCAAAGAAAGTGGGATAC L2R1 – GGAATCCTATGCTGTAATCACAC	*C. jacchus**

* As the *ATXN3L2* sequence of *Callithrix jacchus* differed more than 25% from the other sequences, different primers were designed for this species.

We optimized polymerase chain reactions (PCRs) with the designed primers described above. PCRs were done in a final volume of 10 µL, with 1x of Taq polymerase 2x (MyTaq; BIOLINE) and 0.25 µM (in singleplex reactions) or 0.125 µM (in multiplex reactions) of each primer. PCR conditions are described in Figure 6.

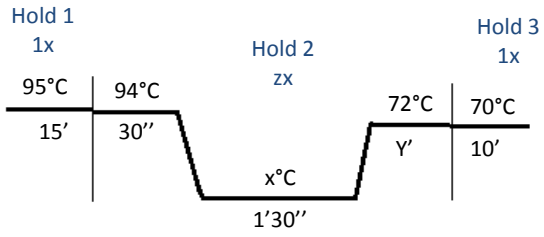

Figure 6 – General PCR protocol with time ('- minutes; '' – seconds) and temperatures (°C) for the amplification of *ATXN3*, *ATXN3L1* and *ATXN3L2* loci. X, Y and Z vary according to the multiplex/singleplex reaction: X (annealing temperature) – described in Tables 1 and 2; Y (extension time) – 2 minutes for singleplex, and 1 minute for multiplexes; Z (number of cycles) - 30 for *ATXN3*; 35 for *ATXN3L1*; and 40 for *ATXN3L2*.

To confirm specificity of DNA amplification, PCR products were submitted to electrophoresis in a polyacrylamide gel prepared with 0.375M Tris/HCl gel buffer (pH=8.8). The gel was obtained by mixing 3 mL T_9C_5 (9% acrylamide, 5% N-N-metileno-bis-acrylamide), 170 µL of 1% ammonium persulfate and 7 µL of TEMED as catalysis agent. Glass supports, with one side covered by a hydrophilic gel-bond film, were used to obtain a 3 mm thick gel. After loading the samples in the

gel, two paper strips soaked in buffer were used at both anode and cathode to allow the horizontal run. To monitor the process of electrophoresis, we added bromophenol blue dye to the anode strip. The electrophoretic system was submitted to refrigeration at 4°C, and to a voltage between 220 and 250 V. The gel was then submitted to silver staining. The coloration method comprised (1) a fixation step of the DNA, with 10% ethanol for 10 minutes followed by 1% nitric acid for 5 minutes; (2) two washes with deionized water, of about 20 seconds each; (3) a coloration step with 0,2% silver nitrate solution, for 20 minutes; (4) two washes again with deionized water, for 20 seconds each; and finally, (5) a revelation step of the DNA fragments with a solution of 0,28 M sodium carbonate and 0,02% formaldehyde. The revelation reaction was stopped with 10% acetic acid for approximately 30 seconds. The resulting gels were washed with water and dried at room temperature.

After optimizing PCR reactions, we proceeded with sequencing. Products were first purified with ExoStar (GE Health Care), for 15 minutes, at 37°C, followed by 15 minutes at 80°C to inactivate the enzyme. ExoStar contains a mix of Alkaline Phosphatase and Exonuclease 1, formulated to remove unincorporated primers and nucleotides from de resulting PCR products.

For sequencing procedures, we used Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) as mix containing normal deoxynucleotides, dye dideoxynucleotides, buffer and AmpliTaq DNA Polymerase. For each reaction of 5 µL, we combined 1 µl of Big Dye mix, sequencing buffer 2.5x, and 0.5 µM of primer with 2.5 µL of purified PCR product. Cycling conditions are described in Figure 7.

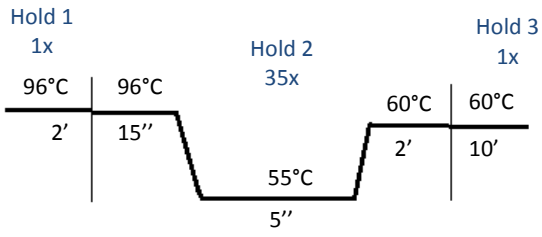

Figure 7 - General protocol with time ('- minutes; '' – seconds) and temperatures (°C) for sequencing A*TXN3*, *ATXN3L1* and *ATXN3L2* PCR products

Finally, sequencing products were purified using Sephadex columns, a cross-linked dextran-gel used to separate the low (unincorporated nucleotides and primers) from high molecular weight molecules of DNA (sequencing products); formamide was then added to the final product to increase the stability of single-stranded DNA for the capillary electrophoresis run in an ABI 3130

Genetic Analyzer. Sequences were analysed with Sequencing Analysis v5.2 software (Applied Biosystems).

Nucleotide diversity of human *ATXN3* retrocopies

To evaluate the nucleotide diversity underlying *ATXN3L1* and *ATXN3L2* human retrocopies, we aligned sequences obtained from the previous step and annotated all variations relative to the reference sequence. We calculated frequencies of variants and predicted the alterations in the respective putative coding sequences.

$(CAG)_n$ tract analysis in *ATXN3*, *ATXN3L1* and *ATXN3L2*

We compared the $(CAG)_n$ tract configuration among the three paralogues for all species based on (1) sequences collected from NCBI, Ensembl and UCSC Genome browser, and (2) results obtained from sequencing (*Pan troglodytes*, *Gorilla gorilla*, *Pongo abelii*, *Macaca mulatta*, *Macaca fascicularis* and *Papio*).

Part 2 – Functional approach

ORF prediction for *ATXN3L1* and *ATXN3L2*

Since little was known regarding the functional relevance of *ATXN3L1* and *ATXN3L2*, we started by predicting the respective ORFs. Predictions were performed by using NCBI Open Reading Frame Finder (ORF Finder - http://www.ncbi.nlm.nih.gov/projects/gorf/) graphical analysis tool.

Conservation of the ATXN3L1 putative protein

The expected protein coding sequence of ATXN3L1 was aligned and compared to the most common coding sequence of the parental gene, the ATXN3-001 transcript. The Josephin domain, UIMs and polyQ tract were annotated in sequences, as well as other important sites/amino acids such as the nuclear export signals NES77 and NES141, the nuclear localization signal NLS273, and the catalytic amino acid triad. With this comparative study, we were able to determine if ATXN3L1 conserved the most important functional components of the parental protein and, thus, a

potential similar functional activity. As ATXNL2 presented premature stop codons and is unlikely to be translated, we did not proceed with further analysis in this retrocopy.

mRNA expression of *ATXN3L1*

To evaluate if *ATXN3L1* is transcribed, primers were designed to specifically amplify ATXN3L1 cDNA. The cDNA of 16 different human tissues was previously obtained by conversion from mRNA, using reverse transcriptase PCR (RT-PCR). Analysed tissues included: ovary, bladder, trachea, esophagus, thymus, thyroid, colon, kidney, skeletal muscle, testis, small intestine, heart, spleen, prostate, liver and brain.

We started by aligning the *ATXN3L1* sequence with the concatenated exons of parental *ATXN3*; this way, we annotated the exons' limits of the parental gene in the retrocopy. Then, primers were designed in regions flanking the exon junctions of the corresponding *ATXN3* exons. Primers are listed in Table 3.

Table 3 - Primers designed to specifically amplify ATXN3L1 in human cDNA.

cDNA *ATXN3L1* primers				
Reactions	**$T_{annealing}$ (°C)**	***ATXN3* exons**	**Amplicon**	**Primer sequence**
ATXN3L1 cDNA	58	UTR, 1 and 2	UTR-E2F - UTR-E2R	UTR-E2F - GCATACAACATCTTCCGGCATACC UTR-E2R - CAGACAGTGCTGAGCACACAGGA
		7 and 8	E7-8F - E7-8R	E7-8F - CAGTGTCGAAGAGATGGATAC E7-8R - CCTCATCTTGGTCTGATGTTCCAGACT
		8 and 9	E8-9F - E8-9R	E8-9F - GAACTAAGCCGCCAAGAAACC E8-9R - GCAGGAGTTACACATGATGTCTTTGGA
		10 and 11	E10-11F - E10-11R	E10-11F - GGGCCACAGTTCATACCTACAC E10-11R - TGTCGACAGCGGCCTGTACTG

PCR conditions were similar to those used for sequencing described above (Figure 1), using 58°C as annealing temperature, 1 minute for extension time, and 35 cycles. To confirm amplification, we have done an electrophoresis run, followed by silver stained, as described above. The expected product sizes are approximately 110 bp for UTR-E2, 140 bp for E7-8, 130 bp for E8-9 and 110 bp for E10-11 amplicon.

Chapter 3

Results

Chapter 3

Results

Part 1 – Evolutionary history of *ATXN3* paralogues

1. Onset of *ATXN3* retrocopies

1.1. Identification of *ATXN3L1* and *ATXN3L2* orthologues in the primate lineage

We retrieved the coding sequence of *ATXN3* for several primates (Table 4). Although the size and location of the collected sequences varied among species, we found homogeneity on both parameters for great apes: humans, chimpanzees, gorillas and orangutans. Notably, when compared to the human sequence, differences were found in *ATXN3* sequences of 1) gorilla, in which exon 2 is not annotated as being translated; 2) orangutang, which does not possess a complete annotated transcription for exon 2; and 3) tarsier, for which databases do not display the complete sequence.

As for *ATXN3* retrocopies, a high homology to the human *ATXN3L1* and *ATXN3L2* sequences was found in *Pan troglodytes*, *Gorilla gorilla, Pongo abelii*, *Macaca mulatta*, *Callithrix jacchus*, *Nomascus leucogenys*, *Tarsius syrichta* and *Otolemur garnettii.* Based on the percentages of homology to human retrocopy sequences, calculated through NCBI-BLAST analysis, we designated the collected sequences as *ATXN3L1* or *ATXN3L2* (Tables 5 and 6). Tarsier and bushbaby sequences (found to be aligned with the human *ATXN3L1*, by using the "comparative genomics" tool of Ensembl) presented a high divergence from both human *ATXN3L1* and *ATXN3L2*. For this reason, the alignment performed by Ensembl did not give us a reliable hint on the origin of these copies in the tarsier and bushbaby species; we, thus, named them as "*ATXN3L?*" at this point of the project. One other sequence for tarsier was retrieved from trace archives specialized BLAST of NCBI (Table 7). No sequence homologies were found in other primate species with genomes included in trace archives, such as Pan paniscus (bonobo), Pongo pygmaeus (bornean orangutan), *Lemur Catta* (ring-tailed lemur), *Aotus nancymaae* (Nancy Ma's night monkey), *Callicebus moloch* (red-bellied Titi), *Chlorocebus aethiops* (grivet or green monkey), *Hylobates concolor* (black

crested gibbon), *Ateles geoffroyi* (Geoffroy's spider monkey), and others. This might be partially explained by the fact that trace archives contain highly fragmented sequences and that these data files are not official entries of the GenBank database, thus, with no associated feature annotations.

Additionally to the *ATXN3L1* and *ATXN3L2* sequences retrieved above, other less homologous sequences (displaying, even though, more than 80% sequence identity to human *ATXN3* coding sequence) were found in Ensembl data base for *Macaca mulatta* and *Callithrix jacchus* (Table 7). These sequences, here classified also as *ATXN3L?*, were aligned with all *ATXN3*, *ATXN3L1* and *ATXN3L2* (Figure A1 of the Appendix section) to next calculate genetic distances among them in topic 1.3 of the Results section.

Table 4 – *ATXN3* genomic and coding sequences compiled for 9 primate species. a) Genomic location of *Macaca mulatta ATXN3* was assessed through the UCSC Genome Browser; b) Incomplete annotated sequence.

ATXN3

Species	Source	Location	Size of genomic/coding sequences (bp)
Homo sapiens	Ensembl	chr14: 92,524,896-92,572,965	48070/1113
Pan troglodytes	Ensembl	chr14: 91,609,895-91,656,494	46600/1086
Gorilla gorilla	Ensembl	chr14: 73834975-73877125:-1	42151/937
Pongo abelii	Ensembl	chr14: 93360787-93404545:-1	43759/996
Nomascus leucogenys	Ensembl	GL397280.1: 29388208-29433598:-1	45391/1080
Macaca mulatta	Ensembl	Scaffold 1099553000000: 10,051-45,764 (Chr7)[a]	35714/1059
Callithrix jacchus	Ensembl	chr10: 117,483,688-117,689,710	206023/1101
Otolemur garnettii	Ensembl	GL873539.1: 18068653-18100321:-1	31669/867[b]
Tarsius syrichta	Ensembl	GeneScaffold_567: 5192-20314:-1	15123/1089[b]

Table 5 – *ATXN3L1* sequences compiled for 7 primate species.

ATXN3L1

Species	Source	Location	Size (bp)
Homo sapiens	Ensembl	chrX: 13336770-13338518:-1	1068
Pan troglodytes	Ensembl	chrX: 13251386-13252912:-1	1062
Gorilla gorilla	UCSC	chrX: 13240248-13241315	1068
Pongo abelii	Ensembl	chrX: 13230000-13231067:-1	1069
Nomascus leucogenys	Ensembl	GL397281.1: 10,640,111-10,641,175	1065
Macaca mulatta	Ensembl	chrX: 11,016,157-11,017,194	1065
Callithrix jacchus	Ensembl	chrX: 11,295,476-13,523,142	1062

Table 6 - *ATXN3L2* sequences compiled for 7 primate species, mainly by using BLAST and BLAT algorithms, from NCBI and UCSC Genome Browser, respectively.

ATXN3L2

Species	Source	Location	Size (bp)
Homo sapiens	NCBI (Gene ID: 100132280)	8q23.2	1109
Pan troglodytes	BLAST/BLAT search	chr8: 109011629-109012743	1115
Gorilla gorilla	BLAST/BLAT search	chr8: 109925669-109926225	1084
Pongo abelii	BLAST/BLAT search	chr8: 117579303-117579866	1075
Nomascus leucogenys	BLAST/BLAT search	GL397267:24598565-24599123	1075
Macaca mulatta	BLAST/BLAT search	chr8:112963801-112964361	1079
Callithrix jacchus	BLAST/BLAT search	chr19:6218140-6218664	1056

Table 7 - Additional ATXN3 paralogues found in primates.

ATXN3L?

Species	Source	Location/Accession
Callithrix jacchus	Ensembl	chr20: 28,630,500-28,631,553
Macaca mulata	Ensembl	chr11: 122178105-122178311
Tarsius syrichta (seq1)	Ensembl - Comp. genomics	Unknown
Tarsius syrichta (seq2)	BLAST - Trace Archives	Unknown
Otoloemur garnettii	Ensembl - Comp. genomics	Unknown

Next, to undoubtedly classify *ATXN3L1* and *ATXN3L2* orthologues in primates and distinguish them from other independent-origin copies, we started by assessing the synteny of all *ATXN3*, *ATXN3L1* and *ATXN3L2* (Table 8 – A, B and C). For *Tarsius syrichta* and *Otolemur garnettii*, this search over flanking regions was not possible in any database. To assess their origin, we compared genetic distances and constructed phylogenetic trees in the next topic (1.2 of the Results section).

By analysing the synteny of *ATXN3* genomic regions, we noticed that up and downstream genes were constant throughout the primate lineage. Upstream genes include, mainly, *CPSF2* (*cleavage and polyadenylation specific factor 2*) and *NDUFB1* (*NADH dehydrogenase 1 beta subcomplex*), whereas downstream genes include *TRIP11* (*thyroid hormone receptor interactor 11*) and *FBLN 5* (*fibulin 5*). The *sodium/potassium/calcium exchanger* gene was only detected upstream to *ATXN3* of some primates (*Homo sapiens*, *Pongo abelii* and *Otolemur garnetii*); however, it may be also present in other species since some genes may have currently not been annotated for all species. The same applies to *PTMAP7* (*prothymosin, alpha pseudogene 7*), only detected downstream to *ATXN3* in humans and chimps.

When we analysed the synteny of *ATXN3L1*, we have found a conserved upstream region across the 7 primates, which included *TCEANC* (*transcription elongation factor A N-terminal and central domain containing*) and *EGFL6* (*epidermal growth factor-like protein 6*) genes. In addition, downstream sequences contained typically two or three of the following genes: *FAM9C* (*family with sequence similarity 9, member C*), *TMSB4X* (*thymosin beta 4, X-linked*) and *TLR8* (*toll receptor 9 precursor, CD289 antigen*).

As for *ATXN3L2*, the majority of primate species presented *KCNV1* (*potassium channel, subfamily V, member 1*) in the 5' flanking region and *CSMD3* (*CUB and Sushi multiple domains 3*) downstream. Human 3' genes were distinct from all other primates: two pseudogenes were present, one of them, with the respective parental gene located upstream of *ATXN3* in humans (*NADH dehydrogenase 1 beta subcomplex pseudogene*). Upstream regions are, however, conserved, supporting this retrocopy is the human *ATXN2* orthologue. Therefore, from this analysis, we confirmed that the majority of collected sequences had been correctly classified based on sequence homology. In the case of the marmoset retrocopy, however, previously classified as *ATXN3L2*, synteny did not favour our first hypothesis of a common origin shared with other *ATXN3L2* orthologues. We, therefore, reclassified this *Callithrix jacchus* sequence as ATXN3L?2 and performed additional analyses to identify its origin (ATXN3L?1 denotes the first detected independent-origin retrocopy for *Callithrix jacchus*).

Table 8 - Synteny of *ATXN3* (A), *ATXN3L1* (B) and *ATXN3L2* (C) through evidence based on collinearity of genes in primates. Two or three genes were retrieved within an interval of 600000 bp, except those marked with an asterisk (*) which were found outside this interval. a, b and c next to the species name represent the source of the information: Ensembl, NCBI and UCSC Genome Browser, respectively. When information for a specific gene or location was retrieved from a source different from the indicated at species name, a, b or c specify the source.

A - ATXN3

Species	Flanking 5'	Gene location	Flanking 3'
Homo sapiens[a]	*NM_153648.3* (*sodium/potassium/calcium exchanger 4 isoform 3*) *CPSF2* (*cleavage and polyadenylation specific factor 2*) *NDUFB1* (*NADH dehydrogenase (ubiquinone) 1 beta subcomplex*)	Chr.14q32.1	*PTMAP7* (*prothymosin, alpha pseudogene 7*)[b] *TRPIP11* (*thyroid hormone receptor interactor 11*) *FBLN5* (*fibulin 5*)
Pan troglodytes[a]	*CPSF2* (*cleavage and polyadenylation specific factor 2*) *NDUFB1* (*NADH dehydrogenase (ubiquinone) 1 beta subcomplex*)	Chr.14: 91,609,895-91,656,494	*LOC100614355* (*prothymosin alpha-like*)[b] *TRPIP11* (*thyroid hormone receptor interactor 11*) *FBLN 5* (*fibulin 5*)
Gorilla gorilla[a]	*CPSF2* (*cleavage and polyadenylation specific factor 2*) *NDUFB1* (*NADH dehydrogenase (ubiquinone) 1 beta subcomplex*)	Chr.14: 73,834,975-73,877,125	*TRIP11* (*thyroid hormone receptor interactor 11*) *FBLN 5* (*fibulin 5*)
Pongo abelii[a]	*LOC100443929* (*cleavage and polyadenylation specific factor 2*)[b] *NDUFB1* (*NADH dehydrogenase (ubiquinone) 1 beta subcomplex*) *LOC100446268* (*Retinal Na+/ Ca2+/K+ Exchanger*)	Chr.14: 93,360,787-93,404,545	*LOC100441607* (*thyroid hormone receptor interactor 11*) *FBLN 5* (*fibulin 5*)
Macaca mulatta	?	Chr.7 [b] (Scaffold 1099553000000: 10,051-45,764)[a]	*LOC100423180* (*uncharacterized*)[b]
Callithrix jacchus[b]	*CPSF2* (*cleavage and polyadenylation specific factor 2*) *LOC100414432* (*uncharacterized*)	Chr.10: 117,483,688-117,689,710[a]	*TRPIP11* (*thyroid hormone receptor interactor 11*) *FBLN 5* (*fibulin 5*)[a]
Nomascus leucogenys[a]	*CPSF2* (*cleavage and polyadenylation specific factor 2*) *NDUFB1* (*NADH dehydrogenase (ubiquinone) 1 beta subcomplex*)	(GL397280.1: 29,388,208-29,433,598)	*TRIP11* (*thyroid hormone receptor interactor 11*) *FBLN 5* (*fibulin 5*)
Otolemur gametii[a]	*SLC24A4* (*Solute carrier family 24 (sodium/potassium/calcium exchanger), member 4*) *ENSOGAG00000032385* (*cleavage and polyadenylation specific factor 2*)	(GL873539.1: 18,068,653-18,100,321)	*TRIP11* (*thyroid hormone receptor interactor 11*) *FBLN 5* (*fibulin 5*)

B - ATXN3L1

Species	Flanking 5'	Sequence location	Flanking 3'
Homo sapiens[a]	*TCEANC* (*transcription elongation factor A (SII) N-terminal and central domain containing*) *GPX1P1* (*glutathione peroxidase pseudogene 1 provided*)[b] *EGFL6* (*epidermal growth factor-like protein 6*)	Xp.22.2	*FAM9C* (*family with sequence similarity 9, member C*) *TMSB4X* (*thymosin beta 4, X-linked*) *TLR8* (*Toll Receptor 9 Precursor, CD289 Antigen*)
Pan troglodytes[a]	*TCEANC* (*transcription elongation factor A (SII) N-terminal and central domain containing*) *EGFL6 (epidermal growth factor-like protein 6*)	ChrX: 13,251,386-13,252,912	*LOC735851* (*family with sequence similarity 9, member C*) *LOC465493* (uncharacterized) *B3Y655_PANTR* (*Toll Receptor 9 Precursor, CD289 Antigen*)
Gorilla gorilla[c]	*TCEANC* (*transcription elongation factor A (SII) N-terminal and central domain containing*) *EGFL6* (*epidermal growth factor-like protein 6*)	ChrX: 13240776-13241315	*FAM9C* (*family with sequence similarity 9, member C*) *TMSB4X* (*thymosin beta 4, X-linked*)
Pongo abelii[a]	*TCEANC* (*transcription elongation factor A (SII) N-terminal and central domain containing* *LOC100461104* (uncharacterized) *LOC100460736 (epidermal growth factor-like protein 6)*	ChrX: 13,230,000-13,231,067	*LOC100452699* (*protein FAM9C-like*) *TYB4_PONAB* (*thymosin beta 4, X-linked*) *TLR8* (*Toll Receptor 9 Precursor, CD289 Antigen*)
Macaca mulatta[c]	*LOC711362* (*transcription elongation factor A (SII) N-terminal and central domain containing*) *EGFL6* (*epidermal growth factor-like protein 6*)	ChrX: 11,016,157-11,017,194	*LOC711039* (*family with sequence similarity 9, member C*) *LOC710959* (Unknown) *ENSMMUG00000032533* (Unknown) *ENSMMUG00000032532* (Unknown) *B6CK00_MACMU* (*Toll Receptor 9 Precursor, CD289 Antigen*)
Callithrix jacchus[a]	*LOC100388682* (*transcription elongation factor A (SII) N-terminal and central domain containing*) *EGFL6* (*epidermal growth factor-like protein 6*)	ChrX: 11,295,476-13,523,142	*LOC100415563* (*thymosin beta 4, X-linked*) *TLR8* (*Toll Receptor 9 Precursor, CD289 Antigen*)
Nomascus leucogenys[a]	*TCEANC* (*transcription elongation factor A (SII) N-terminal and central domain containing*) *EGFL6* (*epidermal growth factor-like protein 6*)	GL397281.1: 10,640,111-10,641,175	*FAM9C* (*family with sequence similarity 9, member C*) *TLR8* (*Toll Receptor 9 Precursor, CD289 Antigen*)

C - ATXN3L2

Species	Flanking 5'	Sequence location	Flanking 3'
Homo sapiens[b]	*KCNV1* (*potassium channel, subfamily V, member 1*)[c] *RPSAP48* (*ribosomal protein SA pseudogene 48*)*	Chr8q23.2 (Chr8: 111567628-111568736)	*LOC100129370* (*NADH dehydrogenase (ubiquinone) 1 beta subcomplex*) - *pseudogene* *EEF1A1P37* (*eukaryotic translation elongation factor 1 alpha 1 pseudogene 37*)
Pan troglodytes	*unknown*		*unknown*
Gorilla gorilla[c]	*KCNV1* (*potassium channel, subfamily V, member 1*) *ENSGGOG00000022520* (*40S Ribosomal*)	Chr8: 109925669-109926225	*ENSGGOG00000023880* (Unknown) *ENSGGOG00000035637* (*U4 spliceosomal RNA*)* *CSMD3* (*CUB and Sushi multiple domains 3*)*
Pongo abelii[c]	*KCNV1_PONAB* (*potassium channel, subfamily V, member 1*)*	Chr8: 117579303-117579866	*CSMD3 (CUB and Sushi multiple domains 3)**
Nomascus leucogenys[c]	*ENSNLEG00000024811* (*U2 spliceosomal RNA*)* *KCNV1* (*potassium channel, subfamily V, member 1*)	GL397267:24598565-24599123	*7SK* (*7SK RNA*) *ENSNLEG00000022484* (*U4 spliceosomal RNA*)* *CSMD3* (*CUB and Sushi multiple domains 3*)*
Macaca mulatta[c]	*ENSMMUG00000034903* (*U2 spliceosomal RNA*) *KCNV1* (*potassium channel, subfamily V, member 1*)	Chr8:112963801-112964361	*LOC700556* (unknown) *LOC699756* (*Nucleophosmin*) *CSMD3* (*CUB and Sushi multiple domains 3*)*
Callithrix jacchus[c]	*5S_rRNA* (*5S ribosomal RNA*) *LOC100389980* (*Lysophospholipase 1*)*	Chr19: 6218140-6218664	*U3* (*Small nucleolar RNA U3*) *C1orf143* (Unknown) *TGFB2* (*Transforming Growth Factor Beta 1 Precursor*)

1.2. Phylogeny of *ATXN3* paralogues in primates

To assess the phylogenetic relationships of *ATXN3* paralogues identified in primates, we calculated the genetic distances among them (Table A1 of the appendix section). Based on these distances, a phylogenetic tree was constructed (Figure 8). Although we could not reproduce the phylogenetic tree of species for each one of the paralogues, we could observe that ATXN3L1 and ATXN3L2 orthologues clustered to each other, confirming their conservation along the primate lineage.

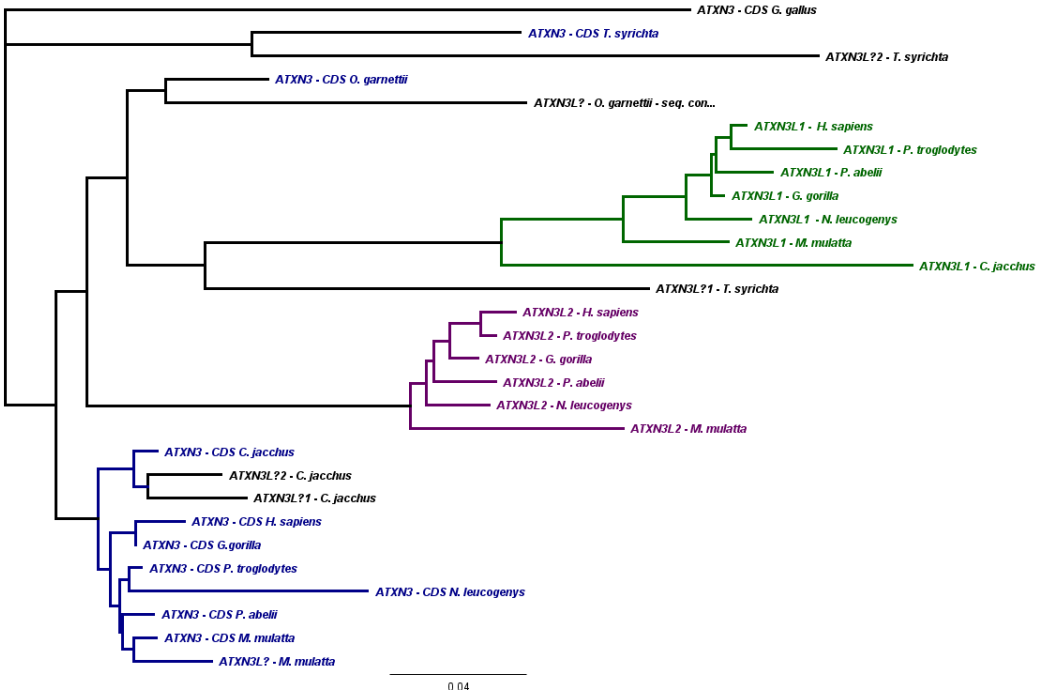

Figure 8 – Phylogenetic tree of *ATXN3*, *ATXN3L1* and *ATXN3L2* for primates, based on the genetic distances presented on Table A1 of the Appendix section. *Gallus gallus* was used as outgroup.

Marmoset *ATXN3L?*2 was shown to be similar to its *ATXN3* coding sequence by sharing a recent node on the phylogenetic tree. This fact confirmed our reclassification of marmoset *ATXN3L2* into *ATXN3L?2* in the previous topic.

ATXN3L?1 sequence of *Tarsius syrichta* was found to cluster next to *ATXN3L1* orthologues and to share more similarity with these orthologues than with the parental or *ATXN3L2* gene; for this reason, we reclassified it as *ATXN3L1*. On the other hand, the other Tarsier sequence, *ATXN3L?2*, shared a more recent ancestor with its parental sequence; therefore, its classification as *ATXN3L?* was maintained. As for *Otolemur garnettii*, *ATXN3L?* sequence did not seem to be orthologue of

ATXN3L1 or *ATXN3L2*; instead, and as it displayed also less genetic distance to its parental gene, *ATXN3L?* classification was maintained.

The obtained phylogenetic tree was compared to others built based on *ATXN3*, *ATXN3L1* and *ATXN3L2* sequences (1) without the $(CAG)_n$ tract, and (2) containing the Josephin domain alone (Figures A2 and A3 of the Appendix section). No differences were found on the main clusters and nodes previously analysed to identify the origin of *ATXN3L1* and *ATXN3L2*.

After identifying *ATXN3L1* and *ATXN3L2* orthologues in primates, we wanted to gain insight into the events underlying their birth, namely the gene source. We, thus, calculated percentages of pairwise identities and identical sites by using aligned sequences of primates in which both retrocopies had previously been identified (Table 9).

Table 9 - Pairwise identities and identical sites between *ATXN3* paralogues of several primate species. The coding sequence of the parental gene and the corresponding aligned sequences of *ATXN3L1* and *ATXN3L2* were used in these calculations. A single value is shown when both percentages are identical.

	Pairwise identity (%)/identical sites (%)					
	H. sapiens	*P. troglodytes*	*G. gorilla*	*P. abelii*	*N. leucogenys*	*M. mulatta*
ATXN3 vs *ATXN3L1*	77.8	78.2	75.6/79.2	77.7/79.6	76.5	80.0/80.5
ATXN3 vs *ATXN3L2*	83.5	83.4	78.1/81.3	80.9/82.5	80.7	80.5/80.9
ATXN3L1 vs *ATXN3L2*	71.1	69.8	74.1	70.7/70.8	73.0	71.2

A common pattern was observed for all species considered in this analysis: *ATXN3L2* presented a higher sequence identity to the *ATXN3* coding sequence than *ATXN3L1*. In addition, both *ATXN3L1* and *ATXN3L2* shared a higher sequence identity to the parental gene than between them, suggesting that two independent events of retrotransposition have occurred, instead of one event followed by duplication.

1.3. Identification of *ATXN3* retrocopies in other mammals

After clarifying the origin of *ATXN3* retrocopies in primates, we performed similar analyses for non-primate mammals to estimate the onset of the two most ancient retrotransposition events. Sequences homologous to *ATXN3L1* and *ATXN3L2* were obtained for *Pteropus vampyrus* (macrobat), *Cavia Porcellus* (guinea pig), *Oryctolagus cuniculus* (rabbit), *Canis familiaris* (dog), *Cloloepus hoffmanni* (sloth), *Bos taurus* (cow) and *Monodelphis domestica* (opossum). As none of these sequences was assigned as *ATXN3L1* or *ATXN3L2*, we named them as *ATXN3L?.* Table 10 resumes the main features of these sequences.

Table 10 – Additional *ATXN3* paralogues found in 7 non-primate mammals.

ATXN3L?

Species	Source	Location/Accession
Pteropus vampyrus (seq1)	BLAST - Trace Archives	gnl | ti:1371399947
Pteropus vampyrus (seq2)	BLAST - Trace Archives	gnl | ti:1328204199
Cavia porcellus	BLAST - Trace Archives	scaffold_33: 11,426,907-11,427,913
Oryctolagus cuniculus	BLAST - Trace Archives	gnl | ti:1979192258
Canis familiaris (seq1)	Ensembl	chr3: 10550724-10551819
Canis familiaris (seq2)	BLAST	NW_003726057.1 (chr5)
Cloloepus hoffmanni (seq1)	BLAST - Trace Archives	gnl | ti:1361780914
Cloloepus hoffmanni (seq2)	BLAST - Trace Archives	gnl | ti:1314261250
Bos taurus (seq1)	BLAST	chr8: 63086126-63086972
Bos taurus (seq2)	BLAST	chr10: 8893215-8893754
Bos taurus (seq3)	BLAST	chr4: 3842581-3843107
Monodelphis domestica	BLAST	chr5: 24707322-24707854

For the majority of the copies, here named *ATXN3L?*, size and location were not homogeneous among them. As for their syntenies, we could not determine chromosomal location in cases where searches were done in the trace archives of NCBI since only scaffolds were available (*Pteropus vampyrus*, *Oryctolagus cuniculus* and *Choloepus hoffmanni*). For species annotated in the UCSC Genome Browser, however, sequences were localized through BLAT algorithm to obtain the chromosomal region and further analysed for the respective neighbouring genes. Therefore, in addition to the previous referred ATXN3L? of rhesus monkey and marmoset, synteny was analysed for dog, guinea pig, cow and opossum retrocopies (Table 11). None of these copies shared any flanking genes with *ATXN3*, *ATXN3L1* or *ATXN3L2*; instead, they differed also from each other. This data suggested that *ATXN3L?* sequences found in these mammalian species are in fact independent-origin rectrocopies of *ATXN3*, and that none of the non-primate mammal sequences is orthologue of *ATXN3L1* or *ATXN3L2*.

Table 11 - Synteny of *ATXN3L?* copies through evidence based on collinearity of genes in non-primate mammals. Two or three genes were retrieved within an interval of 600000 bp, except those marked with an asterisk (*) which were found outside this interval. a, b and c next to the species name represent the source of the information: Ensembl, NCBI and UCSC Genome Browser, respectively. When information for a specific gene or location was retrieved from a source different from the indicated at species name, a, b or c specify the source.

ATXN3L?

Organism	Flanking 5'	Sequence location	Flanking 3'	Human homologues of flanking genes[a]
Callithrix jacchus[a]	*FA2H* (*fatty acid 2-hydroxylase*) *MLKL* (*mixed lineage kinase domain-like*)	Chr20: 28,630,500-28,631,553	*WDR59* (*WD Repeat containing 59*) *ENSCJAG00000014388* (*Class I antigen*)	Chr16
Macaca mulata[a]	*P2RX7* (*P2X Purinoceptor 7*) *ENSMMUG00000031514* (Unknown)	Chr11: 122178105-122178311	*SNORA70* (*Small nucleolar RNA SNORA70*) *OASL* (*2'-5'-oligoadenylate synthetase-like*) *LOC701243* (*Hepatocyte nuclear factor 1 alpha*)	Chr12
Canis familiaris[a]	*NUDT12* (*Peroxisomal NADH pyrophosphatase*) *GIN1* (*Gypsy Retrotransposon Integrase 1*) *ENSCAFG00000024608* (Unknown)	Chr3: 10550724-10551819	*ENSCAFG00000002024* (*Heterogeneous Nuclear Ribonucleoprotein*) *CDH1* (*chromodomain helicase DNA binding protein 1*)	Chr5
Cavia porcellus[c]	*ENSCPOG00000026477* (miRNA) *ENSCPOG00000016622* (5S_rRNA)	scaffold_33: 11,426,907-11,427,913	*ENSCPOG00000021308* (*Scavenger Receptor Cystein Rich Type 1 M130 Precursor*) *ENSCPOG00000009339* (Unknown)*	Unknown

Organism	Flank	Sequence location	Flank	Human homologues of flanking genes[a]
Canis familiaris[c]	*ELAC2* (*ElaC homolog 2 - Zinc phosphodiesterase*)* *5S_rRNA* (*5S ribosomal RNA*) *ENSCAFG00000017903* (*Heparan sulfate glucosamine 3-O-sulfotransferase*)	Chr5: 40677963-40679029	*COX10* (*COX10 homolog, cytochrome c oxidase assembly protein)* *ENSCAFG00000017893* (*Heparan sulfate glucosamine 3-O-sulfotransferase*)	Chr17
Bos taurus[c]	*C9orf174 - Chromosome 9 open reading frame 174* (uncharacterized) hypothetical protein *LOC507550* (uncharacterized)	Chr8: 63086126-63086972	*TDRD7_BOVIN* (*Tudor domain-containing protein 7*) *TMOD1_BOVIN* (*Tropomodulin-1*)	Chr9
Bos taurus[c]	*SNORA31* (*small nucleolar RNA, H/ACA box 31*)	Chr4: 3842581-3843107	*A0JNG1* (*Cordon Bleu Gene*) *Q6U8D5_Bovin* (*Growth Factor Receptor Bound Adapter*)	Chr7
Bos taurus[c]	*OTP* (*Orthopedia Homeobox*) *TBCA* (*Tubulin-specific chaperone A*)	Chr10: 8893215-8893754	*AP3B1* (*AP-3 complex subunit beta-1*) *Q3T0D2_Bovin* (*Secretory Carrier-Associated Membrane Protein 1*)	Chr5
Monodelphis domestica[c]	*C4orf37* (*Chromosome 4 open reading frame 379* *PPM1K* (*protein Phosphatase, Mg2+/Mn2+ dependent, 1K*)*	Chr5:24707322-24707854	*XM_001376187.1* (*Casein Kinase I*)*	Chr4

In order to confirm the origin of *ATXN3L?* copies, sequences were aligned with the coding regions of the respective parental gene, as well as with paralogue sequences previously identified for primates. From this overall alignment (Figure A1 of Appendix section), we calculated genetic distances (Table A4 of the Apendix section) and constructed a phylogenetic tree (Figure 9). Results showed all *ATXN3L?* sequences more similar to the respective parental genes than to primate *ATXN3L1* or *ATXN3L2* sequences. Indeed, they seemed to have originated independently and more recently after species divergence, instead of sharing common ancestral origins. In addition, the extra paralogue sequences found for *M. mulatta* (*ATXN3L?)* and *C. jacchus* (*ATXN3L?1* and *ATXN3L?2*) also showed a most common recent ancestor with their parental sequences than with the other paralogues. This suggests that all sequences classified as *ATXN3L?*resulted, most likely, from additional retrotransposition events originated, more recently, from the respective parental genes.

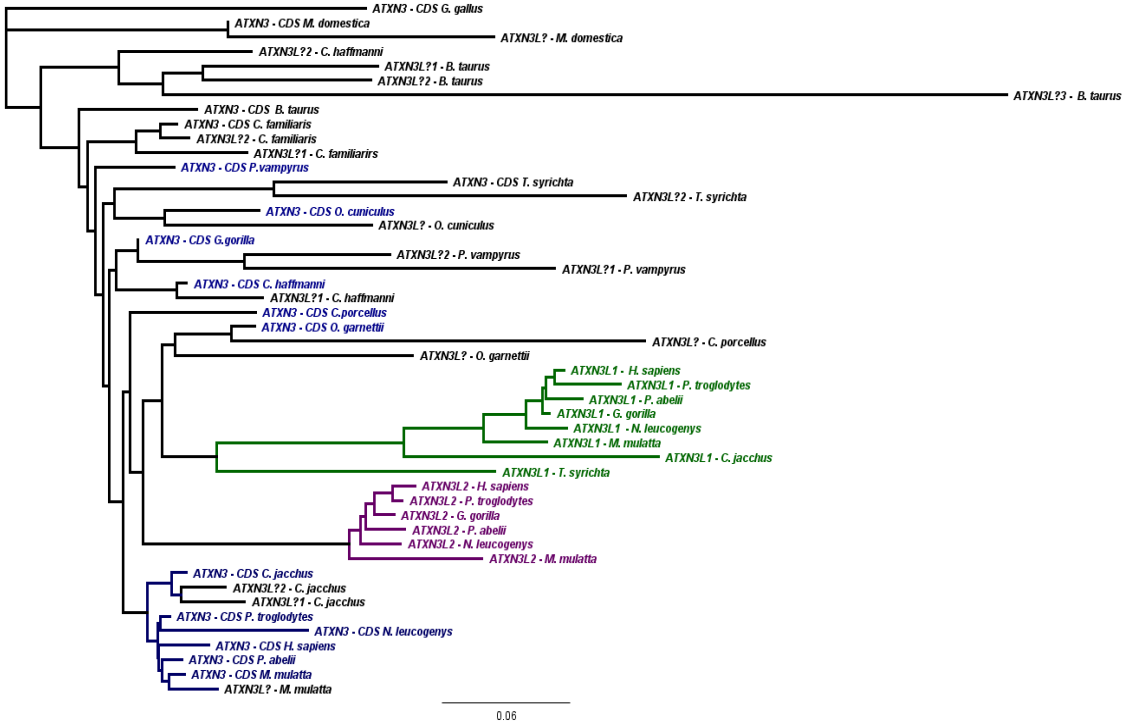

Figure 9 - Phylogenetic tree for *ATXN3*, *ATXN3L1*, *ATXN3L2* and *ATXN3L?* retrocopies of mammals, based on the genetic distances represented on table A4 of the Appendix section. *Gallus gallus* was used as outgroup.

To confirm these results, we constructed phylogenetic trees of these three genes for all mammalian species (1) without the $(CAG)_n$ tract (usually a source of much variation), and (2) for the high conserved Josephin domain alone (Figures A4 and A5). For the first, a similar tree was obtained, whereas the phylogenetic tree based exclusively on the sequence that encodes the Josephin domain (in *ATXN3* and homologous sequences in its paralogues), has shown *ATXN3L1* genetically closer to the parental *ATXN3* than *ATXN3L2*, suggesting that both have very similar Josephin domains, with most of the variation accumulated in the unstructured C-terminal.

Finally, after performing all previous studies, we were able to estimate the time for the origin of *ATXN3* paralogues (Figure 10), based on the tree of species and respective divergence times, in Figure A6 of the appendix section.

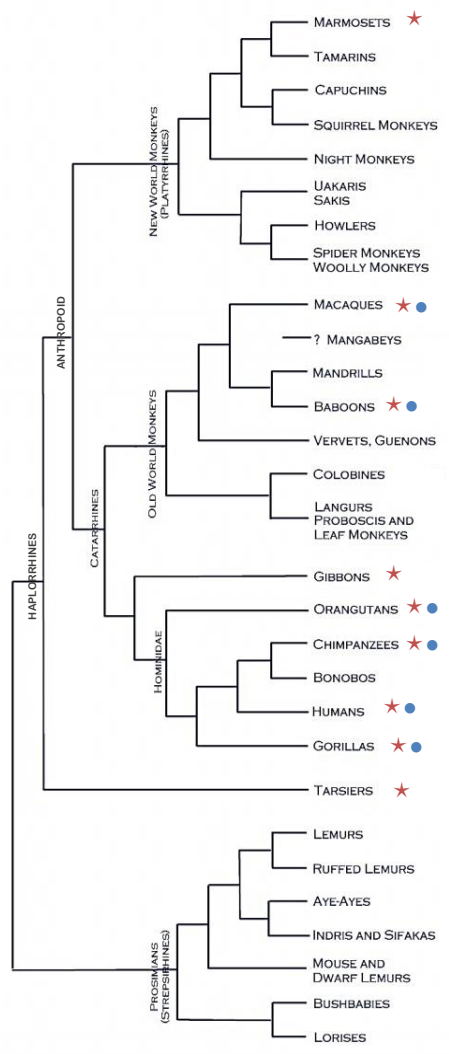

Figure 10 – Phylogenetic tree of primates. The presence of *ATXN3L1* and *ATXN3L2* is marked in each species with a red star and a blue circle, respectively. For *ATXN3L1*, the absence of the red star in Haplorrhines may be explained by incomplete gene annotation in databases. The same applies to blue circles marking *ATXN3L2* in Catarrhines (adapted from http://whozoo.org/mammals/Primates/primatephylogeny.htm).

ATXN3L1 is present in the entire clade of Haplorrhines but absent in Strepsirrhines, indicating that the retrotransposition event on its origin occurred about 63 million years ago (MYA). As for *ATXN3L2*, since orthologues are present in the entire clade of Catarrhines but absent in Platyrrhines, the retrotransposition event is likely to have occurred about 35 MY ago.[79]

2. *ATXN3* transcripts involved in the retrotransposition events

The alignment of *ATXN3L1* and *ATXN3L2* with the 21 human protein coding transcripts annotated in Ensembl has shown both paralogues more similar to ATXN3-001 than to any other transcript. Based on these results, both events of retrotransposition on the origin of *ATXN3L1* and *ATXN3L2* seemed to have involved this transcript, since they all share exons 1, 3 (part), 4 (part), 6 (part), 7, 9, 11, 12, 13 (part), 20 (part) and 21 (based on Ensembl exon classification).

3. Selective signatures underlying *ATXN3L1* and *ATXN3L2*

Evolutionary pressures on proteins were quantified by calculating the omega ratio, or the ratio of substitution rates at non-synonymous and synonymous sites, for *ATXN3* and *ATXN3L1* genes, for different primate species (Table 12).

Table 12 – dN/dS (omega) ratio calculations for *ATXN3* and *ATXN3L1*, to illustrate the selective signatures underlying their evolution.

		ATXN3			*ATXN3L1*		
vs.		**dN**	**dS**	**ω**	**dN**	**dS**	**ω**
Hsap	Ptro	0,0026	0,0094	0,2766	0,0291	0,0549	0,5301
Hsap	Ggor	0,1146	0,1613	0,7105	0,0100	0,0246	0,4065
Hsap	Pabe	0,0026	0,0190	0,1368	0,0163	0,0601	0,2712
Hsap	Nleu	0,0051	0,0287	0,1777	0,0252	0,0979	0,2574
Hsap	Mmul	0,0051	0,0686	0,0743	0,0323	0,2266	0,1425
Hsap	Cjac	0,0129	0,1004	0,1285	0,1431	0,3902	0,3667
Hsap	Tsyr	0,6000	0,8995	0,6670	0,2099	0,5639	0,3722
Ptro	Ggor	0,1175	0,1500	0,7833	0,0304	0,0396	0,7677
Ptro	Pabe	0,0051	0,0094	0,5426	0,0389	0,0730	0,5329
Ptro	Nleu	0,0077	0,0191	0,4031	0,0421	0,1142	0,3687
Ptro	Mmul	0,0077	0,0585	0,1316	0,0507	0,2388	0,2123
Ptro	Cjac	0,0103	0,0900	0,1144	0,1605	0,4522	0,3549
Ptro	Tsyr	0,6270	0,8730	0,7182	0,2335	0,5919	0,3945
Ggor	Pabe	0,1146	0,1613	0,7105	0,0163	0,0345	0,4725
Ggor	Nleu	0,1160	0,1678	0,6913	0,0233	0,6830	0,0341
Ggor	Mmul	0,1190	0,1792	0,6641	0,0278	0,1919	0,1449
Ggor	Cjac	0,1174	0,2500	0,4696	0,1378	0,3792	0,3634
Ggor	Tsyr	0,1414	1,1120	0,1272	0,2022	0,5319	0,3801
Pabe	Nleu	0,0026	0,0287	0,0906	0,0291	0,0867	0,3356
Pabe	Mmul	0,0026	0,0483	0,0538	0,0375	0,1940	0,1933
Pabe	Cjac	0,0103	0,1004	0,1026	0,1501	0,3617	0,4150
Pabe	Tsyr	0,0572	0,8694	0,0658	0,2183	0,5221	0,4181
Nleu	Mmul	0,0051	0,0792	0,0644	0,0355	0,1980	0,1793
Nleu	Cjac	0,0129	0,0900	0,1433	0,1398	0,3840	0,3641
Nleu	Tsyr	0,0572	0,8745	0,0654	0,2126	0,5783	0,3676
Mmul	Cjac	0,0129	0,1561	0,0826	0,1334	0,4354	0,3064
Mmul	Tsyr	0,0614	0,7724	0,0795	0,1979	0,7145	0,2770
Cjac	Tsyr	0,0627	0,9714	0,0645	0,2529	0,7266	0,3481

Omega values were always less than unity, for both *ATXN3* and *ATXN3L1* in all species comparisons. However, the rate of variance of these values were higher in the parental gene (from approximately 0.06 to 0.80) than in the *ATXN3L1* paralogue, which showed less different values (mainly varying from 0.10 to 0.50). The fact that parental genes presented higher variance can be partly explained by the sequence differences found in gorilla, orangutan and tarsier. Still, as omega is less than unity in all calculations, our results suggested that selective constrains are suppressing protein changes, indicating that, as *ATXN3*, also the function of *ATXN3L1* has been kept in check.

4. Nucleotide diversity of *ATXN3L1* and *ATXN3L2*

We have assessed genetic variation of all *ATXN3* paralogues by sequencing a total of 49 human DNA samples (Table 13). For *ATXN3*, we have found a total of seventeen variations, mostly in intronic regions (or in regions that can be transcribed only by less common alternative splicing), according to the Ensemble exon/intron sequence annotations. This is consistent with the fact that transcriptionally active regions of the genome are more prone to be conserved. In exonic regions, we have found six variations: five non-synonymous and one synonymous coding. From these six variations, however, only two (one synonymous and one non-synonymous) were common to all transcripts; the other four were present only in three or fewer transcripts, none of them on the four most common *ATXN3* transcripts.

Table 13 – ATXN3, *ATXN3L1* and *ATXN3L2* nucleotide diversities. Base positions of *ATXN3* were determined based on the sequence annotation of Ensembl; for *ATXN3L1* and *ATXN3L2*, the A nucleotide of the initiation codon (the corresponding first AAG codon in the case of *ATXN3L2*) is base number 1. a – in exonic regions of ATXN3-206, 207 and 209 transcripts; b - in exonic regions of ATXN3-015 transcript; *ancestral alleles determined from primate orthologues alignments; Amino acids: V - Valine; T – Threonine; E - Glutamic acid; R – Arginine; M- Methionine; K – Lysine; Y – Tyrosine; D - Aspartic acid; G – Glycine; NA – Not applicable.

	Variation ID	Base position	Number of chromosomes analysed	Ancestral allele	Base change	Absolute frequencies	Relative frequencies	Annotated frequencies	Amino acid change
ATXN3 **(Chr. 14)**	rs17847278	160 (Intron 4)	20	G	A/G	7/13	0.35/0.65	0.29/0.71	NA
	rs1997920	194 (Exon 6)	48	T	T/C	24/24	0.50/0.50	0.31/0.69	NA
	rs1997919	340 (Exon 6)	42	G	G/A	22/20	0.52/0.48	0.69/0.31	NA
	rs1997918	417 (Exon 6)	42	A	G/A	3/39	0.07/0.93	0.01/0.99	NA
	rs1997917	434 (Exon 6)	44	G	G/A	23/21	0.48/0.52	0.27/0.73	NA
	rs4904834	524 (Intron6)	42	C*	G/C	24/18	0.57/0.53	0.45/0.55	NA
	rs8003520	1081 (Intron 8)	16	G	G/A	3/13	0.19/0.81	0.30/0.70	NA
	rs12590497	129 (Intron 10)	16	T	G/T	10/6	0.63/0.37	0.68/0.32	NA
	rs16999141	17 (Exon 11)	12	T	C/T	6/6	0.50/0.50	0.47/0.53	V/V
	rs1048755	26 (Exon 12)	16	G	A/G	6/10	0.31/0.69	0.26/0.74	M/V
	rs761553	1220 (Intron 12)	16	C	G/C	6/10	0.38/0.62	0.68/0.32	NA
	rs761552	31 (Exon 14)	14	C	T/C	7/7	0.50/0.50	0.32/0.68	M/T[a]
	rs761551	24 (Intron14)	14	C	A/C	7/7	0.50/0.50	0.31/0.69	NA
	rs10467858	127 (Intron 17)	12	G*	A/G	6/6	0.50/0.50	0.27/0.73	NA
	rs7158733	201 (Exon20)	18	C*	C/A	13/5	0.72/0.28	0.68/0.32	Y/Stop[b]
	rs3092822	261 (Exon 20)	22	A*	A/C	14/8	0.64/0.36	0.68/0.32	E/D[b]
	c.2926A>G	319 (Exon20)	22	A*	A/G	8/14	0.36/0.64	undetermined	K/E[b]
ATXN3L1 **(Chr. X)**	c.939G>A	939	38	G*	A/G	2/36	0.05/0.95	undetermined	R/R
	c.995A>G	995	37	A*	A/G	6/31	0.16/0.84	undetermined	D/G
ATXN3L2 **(Chr. 8)**	c.230T>C	230	68	T*	C/T	2/66	0.03/0.97	undetermined	NA
	c.398G>A	398	76	G*	A/G	1/74	0.01/0.99	undetermined	NA
	c.752G>A	752	64	G*	A/G	3/61	0.04/0.96	undetermined	NA

For *ATXN3L1*, only two variations were detected, both near the end of the putative coding sequence: one synonymous and the other non-synonymous, but both relatively rare, varying from 0.05 to 0.16%, respectively. Considering *ATXN3L2*, three variations were observed along the sequence; however, these variants were rare, among our analysed samples, with percentages lower than 0.04%. We have notice, however, that not all SNPs annotated in databases were detected by us, in the resulting sequences, probably due to the low number of chromosomes analysed.

5. Evolution of the $(CAG)_n$ tract in *ATXN3*, *ATXN3L1* and *ATXN3L2*

To gain insight into the processes by which the $(CAG)_n$ tracts have been accumulating variation throughout each paralogue evolution, we analysed sequences from all primates obtained from databases (Table 14). However, this analysis involved mainly a single sequence from each species, which did not allow us to study *loci* diversity. To analyse the level of polymorphism, *ATXN3L1* and *ATXN3L2* from 6 primate species were sequenced (Table 15) and data gathered with the reference ones.

Table 14 – *ATXN3*, *ATXN3L1* and *AtXN3L2* $(CAG)_n$ tracts of several primates collected from NCBI, Ensembl and UCSC Genome Browser databases.

	ATXN3	*ATXN3L1*	*ATXN3L2*
Homo sapiens	$(CAG)_2$ CAA AAG CAG CAA $(CAG)_{17}$ Q_2 Q K Q Q Q_{17}	$(CAG)_2$ GAA CAG AAG $(CAG)_2$ $(CAA)_2$ CAG Q_2 E Q K Q_2 Q_2 Q	$(CAG)_3$ (CGG CAG)n
Pan troglodytes	CAG CAA $(CAG)_{12}$ Q Q Q_{12}	$(CAG)_2$ GAA CAG AAG $(CAG)_2$ CAA CAC CAG Q_2 E Q K Q_2 Q H Q	CAG (CGG $CAG)_9$
Gorilla gorilla	$(CAG)_2$ CAA AAG CAG CAA AAG $(CAG)_n$ Q_2 Q K Q Q K Q_n	$(CAG)_2$ GAA CAG AAG $(CAG)_2$ $(CAA)_2$ CAG Q_2 E Q K Q_2 Q_2 Q	$(CAG)_{10}$
Pongo abelii	$(CAG)_2$ CAA AAG CAG CAA $(CAG)_8$ Q_2 Q K Q Q Q_8	$(CAG)_2$ GAA CAG AAG CTG $(CAG)_4$ Q_2 E Q K L Q_4	$(CAG)_6$
Nomascus leucogenys	$(CAG)_3$ CAA $(CAG)_9$ Q_3 Q Q_9	$(CAG)_2$ GAA CAG AAG $(CAG)_2$ $(CAA)_2$ CAG Q_2 E Q K Q_2 Q_2 Q	$(CAG)_8$
Macaca mulatta	$(CAG)_2$ CAA $(CAG)_2$ AAG $(CAG)_7$ Q_2 Q Q_2 K Q_7	$(CAG)_2$ GAA CAG AAG $(CAG)_2$ $(CAA)_2$ CAG Q_2 E Q K Q_2 Q_2 Q	$(CAG)_6$ CAC CAG
Callithrix jacchus	CAG CAA $(CAG)_5$ CAA $(CAG)_3$ $[GAG\ (CAG)_3]_2$ Q Q Q_5 Q Q_n $(E\ Q_3)_2$	CAA GAA CAA $(CAG)_2$ Q E Q Q_2	Not described
Tarsius syrichta	Not completly sequenced	CAG CAA $(CAG)_2$ Q Q Q_2	Not described
Otolemur garnetii	$(CAG)_5$ CAA $(CAG)_5$ AGG Q_5 Q Q_5 K	Not described	Not described

Analysing the compiled sequences alone, $(CAG)_n$ tract was found to be more conserved in the parental gene along the primate lineage, mainly in human, gorilla and orangutan, with a $(CAG)_2$ CAA AAG CAG CAA tract, followed by AAG and/or $(CAG)_n$. This configuration gives rise to an almost pure polyQ tract, with a single K (lysine) interruption (two in gorilla). Chimp, on the other hand, presents a more simple CAG CAA $(CAG)_n$ configuration, encoding a pure glutamine stretch. The $(CAG)_n$ tract in gibbon, rhesus monkey, marmoset and bushbaby has a $(CAG)_n$ CAA $(CAG)_n$, followed by a variable end. The protein tract is pure (in gibbon) or interrupted only by a lysine in rhesus monkey and bushbaby, or a glutamic acid in marmoset. For *ATXN3L1*, the $(CAG)_n$ is also conserved among primates, having a $(CAG)_2$ GAA CAG AAG $(CAG)_2$ stretch followed by a small variable end with CAA, CAG or CAC triplets. The resulting putative protein sequence is interrupted, generally with the Q_2 E Q K Q_n configuration. Marmoset and tarsier have a different configuration, with a smaller tract: CAA GAA CAA $(CAG)_2$ and CAG CAA $(CAG)_2$, respectively. If translated, these tracts give rise to a smaller pure (for tarsier) or almost pure (for marmoset)

polyglutamine stretches, which may suggest that ATXN3L1 CAG repeat has expanded and gained interruptions along the primate lineage. As for *ATXN3L2*, the $(CAG)_n$ tract is more variable. For gorilla, orangutan and gibbon, the $(CAG)_n$ tract is pure, whereas for human and chimp it is highly interrupted by CGG codons. For the first time, a hexanucleotide repeat is observed associated within ataxin-3 paralogues, instead of a trinucleotide pattern. These results suggested that *ATXN3L1* acquired interruptions in its CAG repeat region that may turn it more stable. On the other hand, in the case of *ATXN3L2*, the almost pure $(CAG)_n$ in species which diverged earlier in primate evolution, and the polymorphic $(CGG\ CAG)_n$ in humans and chimps has shown the highest instability associated to this *locus*.

Table 15 - *ATXN3*, *ATXN3L1* and *ATXN3L2* $(CAG)_n$ tracts obtained by sequencing for several primates.* mark the minimum number of chromosomes analysed taking into account that we ignored whether the respective gene location in these species were in autosomal or sex chromosomes (which, in case of homozygous regions could result in different number of alleles analysed: n vs 2n.

	ATXN3L1			*ATXN3L2*		
	Chr.	**Number of chromossomes**	**Sequence**	**Chr.**	**Number of chromossomes**	**Sequence**
Homo sapiens	X	36	$(CAG)_2$ GAA CAG AAG $(CAG)_2$ $(CAA)_2$ CAG Q_2 E Q K Q_2 Q_2 Q	8	1 4 38 3 2 10	$(CAG)_3$ $(CGG\ CAG)_{11}$ $(CAG)_3$ $(CGG\ CAG)_9$ $(CAG)_3$ $(CGG\ CAG)_8$ $(CAG)_3$ $(CGG\ CAG)_6$ $(CAG)_3$ $(CGG\ CAG)_5$ $(CAG)_3$ $(CGG\ CAG)_4$
Pan troglodytes	X	3	$(CAG)_2$ GAA CAG AAG CTG CAG $(CAA)_2$ CAG Q_2 E Q K L Q Q_2 Q	8	3 1	CAG $(CGG\ CAG)_9$ CAG $(CGG\ CAG)_6$
Gorilla gorilla	X	2	$(CAG)_2$ GAA CAG AAG CTG CAG $(CAA)_2$ CAG Q_2 E Q K L Q Q_2 Q	8	1 1 2 2	$(CAG)_{16}$ $(CAG)_{10}$ $(CAG)_7$ $(CAG)_6$
Pongo abelii	X	2	$(CAG)_2$ GAA CAG AAG CTG CAG $(CAA)_2$ CAG Q_2 E Q K L Q Q_2 Q	8	1 1	$(CAG)_7$ $(CAG)_6$
Papio	?			?	1* 2* 1*	$(CAG)_8$ CAC CAG $(CAG)_7$ CAC CAG CAG CAT $(CAG)_5$ CAC CAG
Macaca mulatta	X	2	$(CAG)_2$ GAA CAG AAG $(CAG)_2$ $(CAA)_2$ CAG Q_2 E Q K Q_2 Q_2 Q	8	4 4	$(CAG)_8$ CAC CAG $(CAG)_6$ CAC CAG
Macaca fascicularis	?	2*	$(CAG)_2$ GAA CAG AAG $(CAG)_2$ $(CAA)_2$ CAG Q_2 E Q K Q_2 Q_2 Q	?	3*	$(CAG)_6$ CAC CAG

Comparing the complied sequences with those obtained by us through sequencing (Table 15), we noticed that *ATXN3L1* sequences of some species may not be correctly described in databases. For chimp, gorilla and orangutan, observed CAG repeat configurations differed from those described in Table 14. This could be simply explained by the level of polymorphism associated to these tracts; nevertheless, all repeat tracts sequenced by us for individuals from these 3 species presented no variation, all sharing the same configuration and length: $(CAG)_2$ GAA CAG AAG CTG

CAG $(CAA)_2$ CAG, possibly encoding a Q_2 E Q K L Q_4 protein stretch. This means that this region may be even more conserved along the primate lineage than it has been described. For *ATXN3L2*, the $(CAG)_n$ tract of all sequenced primate samples showed a similar configuration to reference sequences available in databases, illustrated in Table 14; the repeat size showed, however, to be more variable.

Part 2 – Insights into *ATXN3L1* and *ATXN3L2* functional relevance

1. ORF predictions for *ATXN3L1* and *ATXN3L2*

To explore the possibility of *ATXN3L1* and *ATXN3L2* paralogues being transcribed, the analysis of the respective ORFs was performed; the obtained results are illustrated in Figure 11. ATXN3L1 showed a complete and intact ORF that is maintained in the 7 primate species previously analysed by us. These results confirmed the hypothesis that this paralogue is likely to be subjected to selective constrains that are keeping its sequence on check and almost intact along the primate lineage. On the other hand, for *ATXN3L2*, there are no intact ORFs predicted in any of the primate sequences; instead, only short ORFs of approximately 200 bp or less are expected.

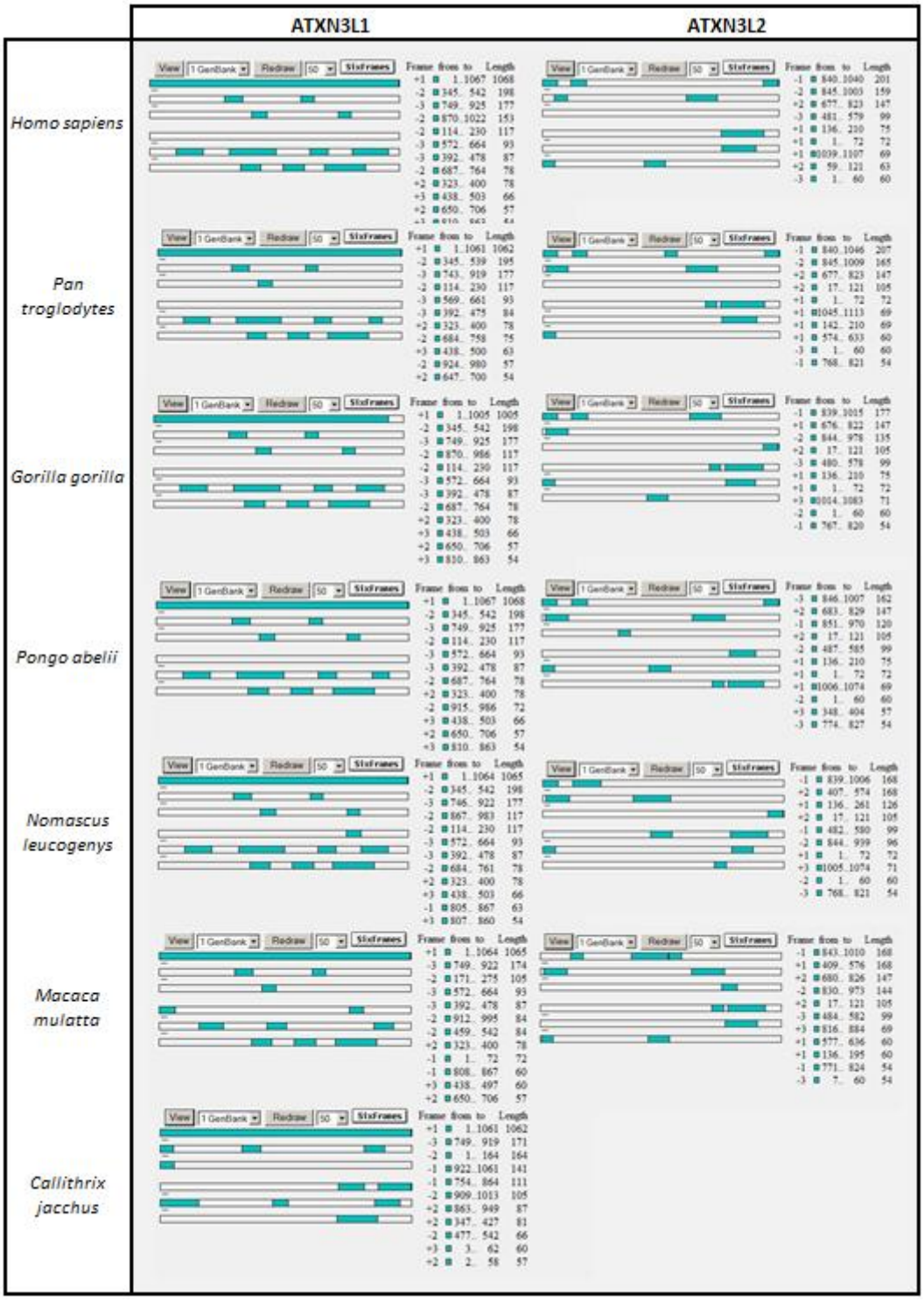

Figure 11 - Predicted ORFs for *ATXN3L1* and *ATXN3L2* primate sequences, calculated by NCBI ORF finder. Obtained ORF predictions are illustrated in blue and the corresponding sizes are described on the right side of each illustration.

2. Analysis of ATXN3L1 putative protein domains in comparison to the parental ATXN3

To understand how differences on the *ATXN3L1* sequence relatively to the parental gene could possibly alter its putative coding protein, the principal domains and regions important to ataxin-3 functions were compared between the two sequences (Figure 12).

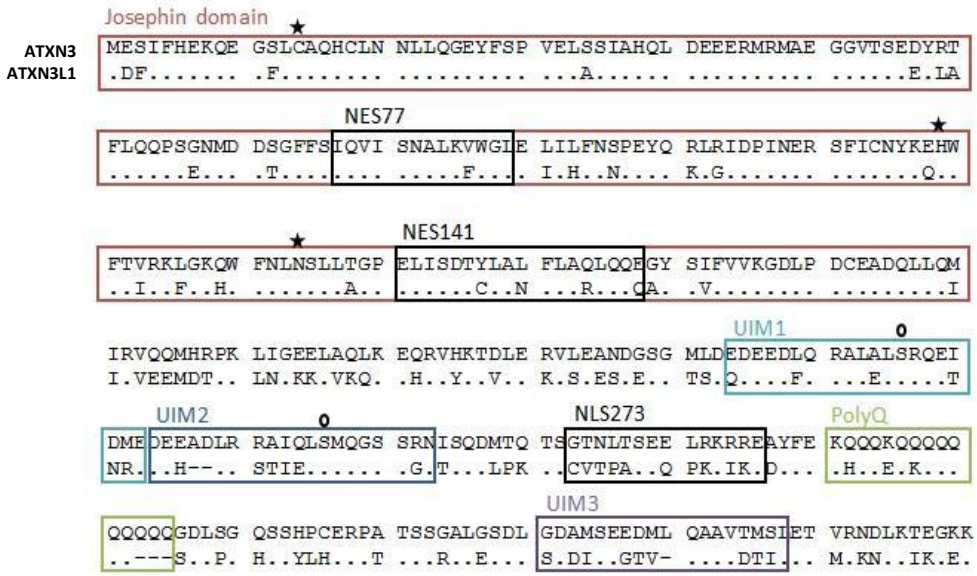

Figure 12 – Comparison between ATXN3 and ATXN3L1 protein coding sequences. Dots show conserved amino acids and letters show the amino acid changes. The main functional Josephin domain, the ubiquitin interacting motifs (UIMs) and polyQ tract are highlighted. Other important sites of the protein are also represented, as the nuclear export signal (NES) and the nuclear localization signal (NLS).★ marks amino acids composing the catalytic triad; 0 marks the conserved serine-232 and serine-260 in UIM1 and UIM2, respectively.

The catalytic Josephin domain was found to be highly conserved in ATXN3L1, with the important catalytic triad, composed of C14, H119 and N134, totally preserved. In addition, NES77 and NES141 are composed mostly by the same amino acids. Additionally, UIM1 of ATXN3L1 is quite similar to the parental protein, although the same cannot be observed for UIM2 and UIM3, which possess more amino acid changes and some gaps. The NLS273 is highly different when compared to the parental protein. Finally, the polyQ tract, as described above, has some alterations that turned it more interrupted in ATXN3L1. It is important to understand that despite these two protein sequences present some amino acid divergences, these small alterations may not correspond to a significant change on the protein conformation and/or function. To better understand this divergence features will be necessary to proceed with more functional studies to characterize the putative ATXN3L1 protein sequence.

3. mRNA expression profile of *ATXN3L1* in humans

We tested for the presence of *ATXN3L1* in cDNA from different human body tissues by using primers designed to specifically amplify this putative transcript. First, to overview the transcriptional pattern among the 16 selected tissues, two of the four planned amplifications were done at a lower temperature (56°C). Primers for the amplification of actin cDNA were used as control (Figure 13).

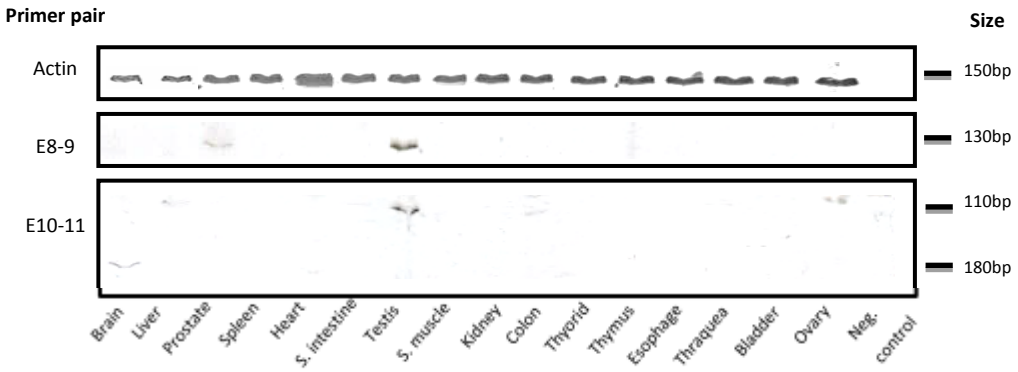

Figure 13 – ATXN3L1 transcriptional pattern of 16 human tissues using E8-9 and E10-11 primers. The human tissues used are listed at the bottom of the image and the sizes of obtained bands are labeled on the right side. Actin cDNA was amplified as a control.

With these conditions, we have found discordant results for the two amplifications tested. By using E8-9 primers, amplification was obtained in only two tissues (prostate and testis), whereas for E10-11, amplified products were observed in brain, liver, testis, colon and ovary. In both cases, testis cDNA was the tissue for which we observed more amplified product, allowing the formation of a clearly visible gel band (a similar amplification pattern for actin cDNA in all tissues discarded the possible heterogeneity among different tissue samples). Sequencing of each band showed that all amplified bands were specific human ATXN3L1 sequences. Additionally, E10-11 primers amplified a longer fragment (180 bp) than the expected 110 bp on brain cDNA; we still need to clarify the reason for this difference on fragment sizes.

To increase the specificity of ATXN3L1 cDNA amplification, the experiment was repeated by applying new different condition of amplification. We increased the annealing temperature to 58°C on the PCR reaction and tested the amplification with the four designed pairs of primers, on the 8 cDNA samples for which we had previously observed amplification (Figure 13).

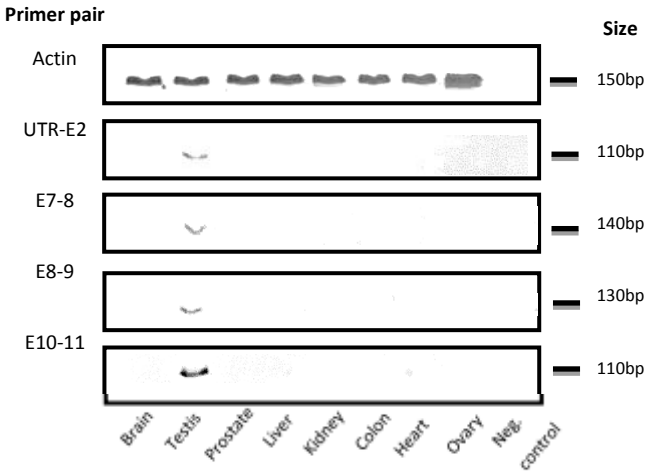

Figure 14 - ATXN3L1 transcriptional pattern for the 4 designed pairs of primers for 8 human tissues. Actin cDNA was amplified as control.

After applying the new conditions, testis was the only tissue where products were constantly and clearly amplified. To clarify the transcription pattern of ATXN3L1, a better optimization of the amplification protocol must be performed.

Chapter 4

Discussion

Chapter 4

Discussion

Retrotransposition and retrogenes are gaining increasing attention as recent studies have shown that they may play an important role in genome evolution and in the formation of new genes.[1, 72]

This project allowed us to investigate the evolution of *ataxin-3* paralogues formed by retrotransposition events, and to shed light on the functional relevance of these copies. Our results showed that *ATXN3* gene has been on the origin of many different retrocopies along the mammalian lineage. Two of these retrocopies, *ATXN3L1* and *ATXN3L2*, were found to have orthologues in several species from the primate lineage. The remaining copies, all diverging from each other, resulted from independent-origin retrotransposition events occurred after the speciation of several mammalian species. These observations are compatible with studies performed by Pan and Zhang (2007 and 2009) where they explain that a burst of young and independent retrogenes has recently arisen in several species branches, mainly in mammalian genomes. [2, 73] The key factor underlying this important active process, that has been shaping the dynamics of mammalian genomes, might have been the recruitment of specific L1 retrotransposons in mammals.[2, 73]

In the case of *ATXN3L1* and *ATXN3L2*, the identification of several orthologues in the primate lineage suggested an ancestral origin for each one of these copies, before the split of several primate clades, with their evolution until these days. Our results suggested that the two retrocopies have independent origins, with *ATXN3L1* as the most ancient copy (63 MYA), followed by the birth of *ATXN3L2*, about 35 MYA. Across their evolution, these copies had the chance to become functional or remain as non-relevant pseudogenes, having consequently more or less influence in the respective genomes of the considered species. Our results have shown some evidences that *ATXN3L1* is being conserved and its ORF kept in check. In addition, our results suggested that *ATXN3L1* is more likely to be transcribed than *ATXN3L2*. This difference may have resulted from the genomic background where copies were inserted. A study by Emerson et al. (2004) claims that retrogenes are not randomly located on chromosomes, but that genes are more likely to be retroposed into and/or out-of the X chromosome in mammals.[74] The authors demonstrated that, during evolution of human and

mouse genomes, the mammalian X chromosome has generated and recruited a disproportionately high number of functional retroposed genes, whereas the autosomes experienced lower gene turnover. [8, 74] Therefore, it is interesting to notice that the most conserved *ATXN3* retrocopy, *ATXN3L1*, is located on the X chromosome, whereas *ATXN3L2* (suggested by our results to be, most likely, a processed pseudogene) resides in an autossome.

As it was previously described, to become fixed in a given species genome, retrocopies must arise in the germ line and the gene source must be expressed at that time.[1] *ATXN3* is ubiquitously present in human tissues and cell types however, the retropositions of *ATXN3L1* and *ATXN3L2* occurred, not in the human species, but in an ancestor of Haplorrhines and Catarrhines, respectively. [25, 38] Thus, to be feasible the occurrence of retrotransposition events from *ATXN3*, this gene must have been transcribed in the germ line of these ancestral species.

The analysis of the $(CAG)_n$ tract of *ATXN3*, among different primate species, is a key factor to understand the process of repeat instability over the primate lineage until reaching the expanded/pathogenic range in humans. In the parental gene, the repeat tract configuration has been maintained stable along the primate lineage, varying mainly on the repeat size, which increased over time. Most of the studies have described a range between 14-40 in humans, 14-20 in chimps, 8-11 in gorillas and 24-25 in orangutans.[75] According to Andrés et al. (2004), humans show, not only a higher mean number of repeats than other species, but also a higher variance and coefficient of variation, with a decreasing trend observed as more anciently diverged primates are analysed. Thus, a relationship has been suggested between variability levels and expansion potential. In addition, the authors hypothesized the existence of some balancing selection favouring the existence of different alleles in human populations that would maintain the high (CAG)n diversity currently observed in our species.[75]

As for *ATXN3L1*, the CAG repeat showed a poorly polymorphic and highly interrupted tract across the primate lineage, whereas *ATXN3L2* presented a pure $(CAG)_n$ in some species, and a polymorphic hexanucleotide repeat in humans and chimpanzees. This recent acquisition of a repetitive CGGCAG resulted, probably, from a CAG to CGG mutation followed by instability that encompassed the six bases instead of the CAG alone.

In addition, studies on the CAG region of both parental and *ATXN3* paralogues can be important to gain insight into the involvement of these repetitive tracts in both pathological and non-pathological states of the protein. In fact, Schaefer et al. (2012) have claimed that $(CAG)_n$ tracts are not randomly distributed in the genome.[76] Also, recent studies have

demonstrated that the polyQ tract may actually have a function in repeat-containing proteins. In fact, we can observe a biased involvement of these proteins towards functions related to transcriptional regulation and nuclear localization.[77, 78] Moreover, recent studies have suggested the involvement of polyQ repeats in protein-protein interaction networks, by stabilizing structural changes to facilitate the interaction between own coiled-coil regions with coiled-coil regions of other proteins (Figure 14).[76] These facts suggest that $(CAG)_n$ tracts can actually be subjected to selective pressures, in order to maintain a possible functional role in the protein context.

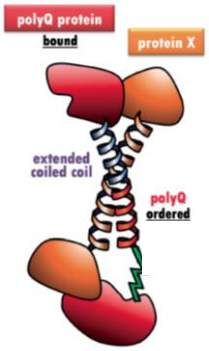

Figure 14 – Possible role for the polyQ stretch in protein-protein interactions: upon interaction with a protein partner, the polyQ region adopts a coiled-coil structure. (adapted from Schaefer, 2012)[76]

Analysing ORF predictions for both *ATXN3L1* and *ATXN3L2*, it was evident that the first copy, although more ancient, maintained its intact putative coding sequence. In opposition, the most recent retrocopy *ATXN3L2* displayed a disrupted ORF, with maximum length fragments of 200 bp predicted as alternative ORFs. These facts come to reinforce that both copies have accumulated genetic differences relatively to the parental gene along their evolution; nevertheless, *ATXN3L1* seems to be under selective constrains that might be keeping its sequence in check whereas *ATXN3L2* has accumulated premature stop codons that interrupted its original sequence. Yet, the short ORF of *ATXN3L2* may be transcribed into mRNA and play a trans-regulatory role in the expression of the parental gene. Indeed, it was recently found that some mammalian retropseudogenes have evolved the capacity to encode small interference RNAs, important for the regulation of their parental source genes.

Concerning the analysis of *ATXN3L1* expression in human tissues, our results have demonstrated that this paralogue is expressed, at least, in human testis. Results for other human tissues, such as brain, prostate, colon, liver and ovary, were less clear and, thus, need to be confirmed with further analyses. By comparing our results with expression data available in NCBI (EST profile) and in the Human Protein Atlas (http://www.proteinatlas.org/, updated at

11/11/2011), data is not consensual. Figure 15 shows a summary of the *ATXN3L1* expression levels, and Figure A7 of the appendix section a more complete version of these data.

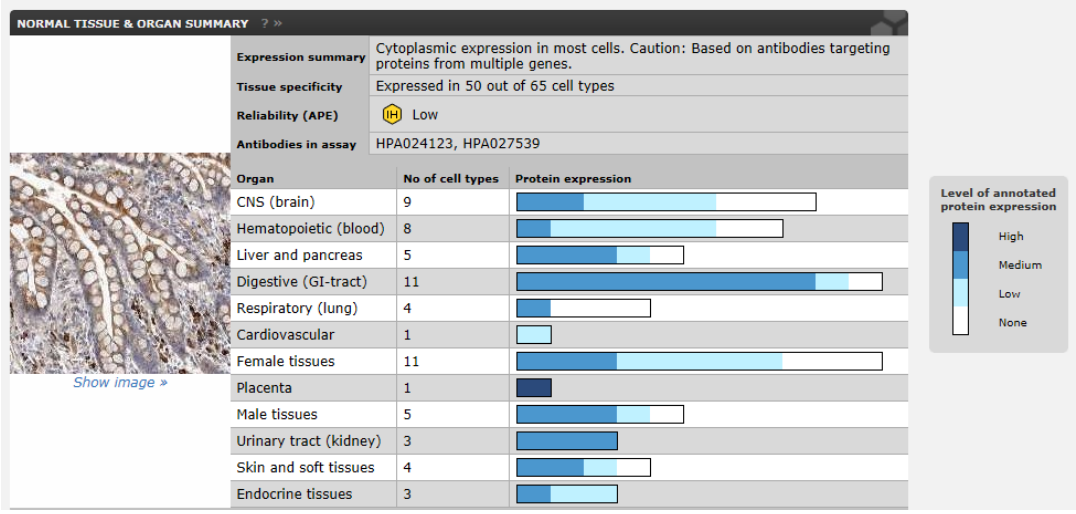

Figure 15 – ATXN3L1 expression levels in different human organs/tissues. (data from the Human Protein Atlas: http://www.proteinatlas.org/ENSG00000123594)

It is important, however, to notice that expression patterns annotated in The Human Protein Atlas are based on antibody assays classified as "Low Reliability" in the case of ATXN3L1. On the other hand, ATXN3L1 expression patterns consulted in the NCBI – UniGene – EST Profile are illustrated in Figure A8 of the appendix section. Here, the expression levels were measured by EST counts in unbiased cDNA libraries. Testis and brain are the only human body tissues where ATXN3L1 is shown to be expressed, with higher levels of expression in testis. These data are, in fact, more similar to our results, but no further conclusions can be taken before confirming the obtained results

By comparing the principal domains of ATXN3 and ATXN3L1 proteins, we observed a conserved Josephin domain (catalytic domain), NES77, NES141 and UIM1, whereas less similarity was found for the remaining regions NLS273, UIM2 and UIM3. As the catalytic domain is the most conserved one (with intact catalytic sites: C14, H119 and N134), it can be hypothesized that ATXN3L1 is likely to exert an ubiquitin protease activity as the parental ATXN3. The two other important residues in UIMs (serine-232 and serine-260 in UIM1 and

UIM2), as well as the UIM1 itself, are also conserved in ATXN3L1 suggesting that ubiquitin interactions can also be carried out by this retrocopy. However, possible functions of ATXN3L1 cannot be speculated based only on its conserved domains. More functional studies must be performed, for example, in order to explore common binding partners to the parental protein.

Chapter 5

Final remarks and future perspectives

Chapter 5

Final remarks and future perspectives

The general aims traced for this project were accomplished, but still, there are techniques here developed that can be optimized, as well as the enlargement of sample sizes in order to enrich and reinforce the obtained results.

From our study, it was possible to conclude that the analysis of paralogue genes involved in neurodegenerative diseases can be of great importance to gain insight into the mechanisms of pathogenesis. Namely, on Machado-Joseph disease, the ATXN3L1 was observed to be transcribed in human tissues; expression patterns will be assessed next through the optimization of specific amplifications of ATXN3L1 cDNA. At the protein level, ATXN3L1 possesses a highly conserved catalytic domain. Next, to detect the endogenous protein in human tissues, an anti-ATXN3L1 specific antibody can be generated to perform Western blot and immunohistochemistry. Additionally, future studies on the protein characterization will allow us to evaluate if this retrocopy has evolved towards functional diversification (neofunctionalization or neosubfunctionalization) relatively to the parental protein. To test if ATXN3L1 shares some of the most important molecular interactors (common binding partners) with ATXN3, one could detect common binding partners by co-immunoprecipitation with the transcriptional coactivators and DNA repair proteins that interact with the parental ataxin-3.

Further studies can be made, also, to confirm if the small ATXN3L2 ORFs predicted in this study are transcribed or not. This can be done by following the same approach we used here for ATXN3L1: the design of specific primers to amplify concatenated regions corresponding to ATXN3 exons, in gene-specific nucleotides with highly conserved sites across species within orthologues. If the transcription of ATXN3L2 occurs, one could explore the possibility of them being encoding small interference RNAs and acting as trans-regulatory factors of ATXN3.

References

References

1. Kaessmann H, Vinckenbosch N, Long M. RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet. Jan 2009;10(1):19-31.
2. Pan D, Zhang L. Burst of young retrogenes and independent retrogene formation in mammals. PLoS One. 2009;4(3):e5040.
3. Ostertag EM, Kazazian HH, Jr. Biology of mammalian L1 retrotransposons. Annu Rev Genet. 2001;35:501-538.
4. Kazazian HH, Jr., Goodier JL. LINE drive. retrotransposition and genome instability. Cell. Aug 9 2002;110(3):277-280.
5. Vinckenbosch N, Dupanloup I, Kaessmann H. Evolutionary fate of retroposed gene copies in the human genome. Proc Natl Acad Sci U S A. Feb 28 2006;103(9):3220-3225.
6. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet. Oct 2009;10(10):691-703.
7. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. Nat Genet. Apr 2000;24(4):363-367.
8. Betran E, Thornton K, Long M. Retroposed new genes out of the X in Drosophila. Genome Res. Dec 2002;12(12):1854-1859.
9. Burki F, Kaessmann H. Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. Nat Genet. Oct 2004;36(10):1061-1063.
10. Carlson KM, Andresen JM, Orr HT. Emerging pathogenic pathways in the spinocerebellar ataxias. Curr Opin Genet Dev. Jun 2009;19(3):247-253.
11. Teive HA. Spinocerebellar ataxias. Arq Neuropsiquiatr. Dec 2009;67(4):1133-1142.
12. Soong BW, Paulson HL. Spinocerebellar ataxias: an update. Curr Opin Neurol. Aug 2007;20(4):438-446.
13. Bettencourt C, Lima M. Machado-Joseph Disease: from first descriptions to new perspectives. Orphanet J Rare Dis. 2011;6:35.
14. Kieling C, Morales Saute JA, Jardim LB. When ataxia is not just ataxia. Nat Clin Pract Neurol. May 2007;3(5):E2.
15. Coutinho P, Andrade C. Autosomal dominant system degeneration in Portuguese families of the Azores Islands. A new genetic disorder involving cerebellar, pyramidal, extrapyramidal and spinal cord motor functions. Neurology. Jul 1978;28(7):703-709.
16. Jardim LB, Silveira I, Pereira ML, et al. A survey of spinocerebellar ataxia in South Brazil - 66 new cases with Machado-Joseph disease, SCA7, SCA8, or unidentified disease-causing mutations. J Neurol. Oct 2001;248(10):870-876.
17. Vale J, Bugalho P, Silveira I, Sequeiros J, Guimaraes J, Coutinho P. Autosomal dominant cerebellar ataxia: frequency analysis and clinical characterization of 45 families from Portugal. Eur J Neurol. Jan 2010;17(1):124-128.
18. Zhao Y, Tan EK, Law HY, Yoon CS, Wong MC, Ng I. Prevalence and ethnic differences of autosomal-dominant cerebellar ataxia in Singapore. Clin Genet. Dec 2002;62(6):478-481.
19. Tang B, Liu C, Shen L, et al. Frequency of SCA1, SCA2, SCA3/MJD, SCA6, SCA7, and DRPLA CAG trinucleotide repeat expansion in patients with hereditary spinocerebellar ataxia from Chinese kindreds. Arch Neurol. Apr 2000;57(4):540-544.
20. van de Warrenburg BP, Sinke RJ, Verschuuren-Bemelmans CC, et al. Spinocerebellar ataxias in the Netherlands: prevalence and age at onset variance analysis. Neurology. Mar 12 2002;58(5):702-708.
21. Schols L, Amoiridis G, Buttner T, Przuntek H, Epplen JT, Riess O. Autosomal dominant cerebellar ataxia: phenotypic differences in genetically defined subtypes? Ann Neurol. Dec 1997;42(6):924-932.
22. Maruyama H, Izumi Y, Morino H, et al. Difference in disease-free survival curve and regional distribution according to subtype of spinocerebellar ataxia: a study of 1,286 Japanese patients. Am J Med Genet. Jul 8 2002;114(5):578-583.
23. Martins S, Calafell F, Gaspar C, et al. Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. Arch Neurol. Oct 2007;64(10):1502-1508.
24. Takiyama Y, Nishizawa M, Tanaka H, et al. The gene for Machado-Joseph disease maps to human chromosome 14q. Nat Genet. Jul 1993;4(3):300-304.

25. Ichikawa Y, Goto J, Hattori M, et al. The genomic structure and expression of MJD, the Machado-Joseph disease gene. J Hum Genet. 2001;46(7):413-422.
26. Bettencourt C, Santos C, Montiel R, et al. Increased transcript diversity: novel splicing variants of Machado-Joseph disease gene (ATXN3). Neurogenetics. May 2010;11(2):193-202.
27. Kawaguchi Y, Okamoto T, Taniwaki M, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet. Nov 1994;8(3):221-228.
28. Wang JL, Xiao B, Cui XX, et al. Analysis of SCA2 and SCA3/MJD repeats in Parkinson's disease in mainland China: genetic, clinical, and positron emission tomography findings. Mov Disord. Oct 15 2009;24(13):2007-2011.
29. Gilman S. The spinocerebellar ataxias. Clin Neuropharmacol. Nov-Dec 2000;23(6):296-303.
30. Maruyama H, Nakamura S, Matsuyama Z, et al. Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. Hum Mol Genet. May 1995;4(5):807-812.
31. van de Warrenburg BP, Hendriks H, Durr A, et al. Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch-French cohort. Ann Neurol. Apr 2005;57(4):505-512.
32. Kawakami H, Maruyama H, Nakamura S, et al. Unique features of the CAG repeats in Machado-Joseph disease. Nat Genet. Apr 1995;9(4):344-345.
33. Maciel P, Gaspar C, DeStefano AL, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. Am J Hum Genet. Jul 1995;57(1):54-61.
34. Carvalho DR, La Rocque-Ferreira A, Rizzo IM, Imamura EU, Speck-Martins CE. Homozygosity enhances severity in spinocerebellar ataxia type 3. Pediatr Neurol. Apr 2008;38(4):296-299.
35. Panigrahi GB, Lau R, Montgomery SE, Leonard MR, Pearson CE. Slipped (CTG)*(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. Nat Struct Mol Biol. Aug 2005;12(8):654-662.
36. Martins S, Calafell F, Wong VC, Sequeiros J, Amorim A. A multistep mutation mechanism drives the evolution of the CAG repeat at MJD/SCA3 locus. Eur J Hum Genet. Aug 2006;14(8):932-940.
37. Djian P, Hancock JM, Chana HS. Codon repeats in genes associated with human diseases: fewer repeats in the genes of nonhuman primates and nucleotide substitutions concentrated at the sites of reiteration. Proc Natl Acad Sci U S A. Jan 9 1996;93(1):417-421.
38. Trottier Y, Cancel G, An-Gourfinkel I, et al. Heterogeneous intracellular localization and expression of ataxin-3. Neurobiol Dis. Nov 1998;5(5):335-347.
39. Bilen J, Bonini NM. Genome-wide screen for modifiers of ataxin-3 neurodegeneration in Drosophila. PLoS Genet. Oct 2007;3(10):1950-1964.
40. Macedo-Ribeiro S, Cortes L, Maciel P, Carvalho AL. Nucleocytoplasmic shuttling activity of ataxin-3. PLoS One. 2009;4(6):e5834.
41. Paulson HL, Das SS, Crino PB, et al. Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. Ann Neurol. Apr 1997;41(4):453-462.
42. Bichelmeier U, Schmidt T, Hubener J, et al. Nuclear localization of ataxin-3 is required for the manifestation of symptoms in SCA3: in vivo evidence. J Neurosci. Jul 11 2007;27(28):7418-7428.
43. Chai Y, Koppenhafer SL, Shoesmith SJ, Perez MK, Paulson HL. Evidence for proteasome involvement in polyglutamine disease: localization to nuclear inclusions in SCA3/MJD and suppression of polyglutamine aggregation in vitro. Hum Mol Genet. Apr 1999;8(4):673-682.
44. Mukhopadhyay D, Riezman H. Proteasome-independent functions of ubiquitin in endocytosis and signaling. Science. Jan 12 2007;315(5809):201-205.
45. Todi SV, Laco MN, Winborn BJ, Travis SM, Wen HM, Paulson HL. Cellular turnover of the polyglutamine disease protein ataxin-3 is regulated by its catalytic activity. J Biol Chem. Oct 5 2007;282(40):29348-29358.
46. Chou AH, Yeh TH, Ouyang P, Chen YL, Chen SY, Wang HL. Polyglutamine-expanded ataxin-3 causes cerebellar dysfunction of SCA3 transgenic mice by inducing transcriptional dysregulation. Neurobiol Dis. Jul 2008;31(1):89-101.

47. Nicastro G, Todi SV, Karaca E, Bonvin AM, Paulson HL, Pastore A. Understanding the role of the Josephin domain in the PolyUb binding and cleavage properties of ataxin-3. PLoS One. 2010;5(8):e12430.
48. Masino L, Musi V, Menon RP, et al. Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. FEBS Lett. Aug 14 2003;549(1-3):21-25.
49. Harris GM, Dodelzon K, Gong L, Gonzalez-Alegre P, Paulson HL. Splice isoforms of the polyglutamine disease protein ataxin-3 exhibit similar enzymatic yet different aggregation properties. PLoS One. 2010;5(10):e13695.
50. Scheel H, Tomiuk S, Hofmann K. Elucidation of ataxin-3 and ataxin-7 function by integrative bioinformatics. Hum Mol Genet. Nov 1 2003;12(21):2845-2852.
51. Mueller T, Breuer P, Schmitt I, Walter J, Evert BO, Wullner U. CK2-dependent phosphorylation determines cellular localization and stability of ataxin-3. Hum Mol Genet. Sep 1 2009;18(17):3334-3343.
52. Weeks SD, Grasty KC, Hernandez-Cuebas L, Loll PJ. Crystal structure of a Josephin-ubiquitin complex: evolutionary restraints on ataxin-3 deubiquitinating activity. J Biol Chem. Feb 11 2011;286(6):4555-4565.
53. Matos CA, de Macedo-Ribeiro S, Carvalho AL. Polyglutamine diseases: the special case of ataxin-3 and Machado-Joseph disease. Prog Neurobiol. Sep 15 2011;95(1):26-48.
54. Ferrigno P, Silver PA. Polyglutamine expansions: proteolysis, chaperones, and the dangers of promiscuity. Neuron. Apr 2000;26(1):9-12.
55. Muchowski PJ, Schaffar G, Sittler A, Wanker EE, Hayer-Hartl MK, Hartl FU. Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. Proc Natl Acad Sci U S A. Jul 5 2000;97(14):7841-7846.
56. Chai Y, Wu L, Griffin JD, Paulson HL. The role of protein composition in specifying nuclear inclusion formation in polyglutamine disease. J Biol Chem. Nov 30 2001;276(48):44889-44897.
57. Li F, Macfarlan T, Pittman RN, Chakravarti D. Ataxin-3 is a histone-binding protein with two independent transcriptional corepressor activities. J Biol Chem. Nov 22 2002;277(47):45004-45012.
58. Evert BO, Araujo J, Vieira-Saecker AM, et al. Ataxin-3 represses transcription via chromatin binding, interaction with histone deacetylase 3, and histone deacetylation. J Neurosci. Nov 1 2006;26(44):11474-11486.
59. Gunawardena S, Her LS, Brusch RG, et al. Disruption of axonal transport by loss of huntingtin or expression of pathogenic polyQ proteins in Drosophila. Neuron. Sep 25 2003;40(1):25-40.
60. Donaldson KM, Li W, Ching KA, Batalov S, Tsai CC, Joazeiro CA. Ubiquitin-mediated sequestration of normal cellular proteins into polyglutamine aggregates. Proc Natl Acad Sci U S A. Jul 22 2003;100(15):8892-8897.
61. Paulson HL, Perez MK, Trottier Y, et al. Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. Neuron. Aug 1997;19(2):333-344.
62. Goti D, Katzen SM, Mez J, et al. A mutant ataxin-3 putative-cleavage fragment in brains of Machado-Joseph disease patients and transgenic mice is cytotoxic above a critical concentration. J Neurosci. Nov 10 2004;24(45):10266-10279.
63. Yoshizawa T, Yamagishi Y, Koseki N, et al. Cell cycle arrest enhances the in vitro cellular toxicity of the truncated Machado-Joseph disease gene product with an expanded polyglutamine stretch. Hum Mol Genet. Jan 1 2000;9(1):69-78.
64. Haacke A, Broadley SA, Boteva R, Tzvetkov N, Hartl FU, Breuer P. Proteolytic cleavage of polyglutamine-expanded ataxin-3 is critical for aggregation and sequestration of non-expanded ataxin-3. Hum Mol Genet. Feb 15 2006;15(4):555-568.
65. Wellington CL, Ellerby LM, Hackam AS, et al. Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. J Biol Chem. Apr 10 1998;273(15):9158-9167.
66. Hubener J, Vauti F, Funke C, et al. N-terminal ataxin-3 causes neurological symptoms with inclusions, endoplasmic reticulum stress and ribosomal dislocation. Brain. Jul 2011;134(Pt 7):1925-1942.
67. Costa Mdo C, Paulson HL. Toward understanding Machado-Joseph disease. Prog Neurobiol. May 2012;97(2):239-257.
68. Fujigasaki H, Uchihara T, Koyano S, et al. Ataxin-3 is translocated into the nucleus for the formation of intranuclear inclusions in normal and Machado-Joseph disease brains. Exp Neurol. Oct 2000;165(2):248-256.
69. Warrick JM, Morabito LM, Bilen J, et al. Ataxin-3 suppresses polyglutamine neurodegeneration in Drosophila by a ubiquitin-associated mechanism. Mol Cell. Apr 1 2005;18(1):37-48.

70. Crespo-Barreto J, Fryer JD, Shaw CA, Orr HT, Zoghbi HY. Partial loss of ataxin-1 function contributes to transcriptional dysregulation in spinocerebellar ataxia type 1 pathogenesis. PLoS Genet. Jul 2010;6(7):e1001021.
71. Bowman AB, Lam YC, Jafar-Nejad P, et al. Duplication of Atxn1l suppresses SCA1 neuropathology by decreasing incorporation of polyglutamine-expanded ataxin-1 into native complexes. Nat Genet. Mar 2007;39(3):373-379.
72. Kaessmann H. Genetics. More than just a copy. Science. Aug 21 2009;325(5943):958-959.
73. Pan D, Zhang L. Quantifying the major mechanisms of recent gene duplications in the human and mouse genomes: a novel strategy to estimate gene duplication rates. Genome Biol. 2007;8(8):R158.
74. Emerson JJ, Kaessmann H, Betran E, Long M. Extensive gene traffic on the mammalian X chromosome. Science. Jan 23 2004;303(5657):537-540.
75. Andres AM, Soldevila M, Lao O, et al. Comparative genetics of functional trinucleotide tandem repeats in humans and apes. J Mol Evol. Sep 2004;59(3):329-339.
76. Schaefer MH, Wanker EE, Andrade-Navarro MA. Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. Nucleic Acids Res. May 1 2012;40(10):4273-4287.
77. Harrison PM. Exhaustive assignment of compositional bias reveals universally prevalent biased regions: analysis of functional associations in human and Drosophila. BMC Bioinformatics. 2006;7:441.
78. Alba MM, Guigo R. Comparative analysis of amino acid repeats in rodents and humans. Genome Res. Apr 2004;14(4):549-554.
79. Jobling MA, Hurles, ME, Tyler-Smith C. Human EvolutionaryGenetics – Origins, Peoples & Disease. Garland Publishing (2004)

Appendix

Appendix

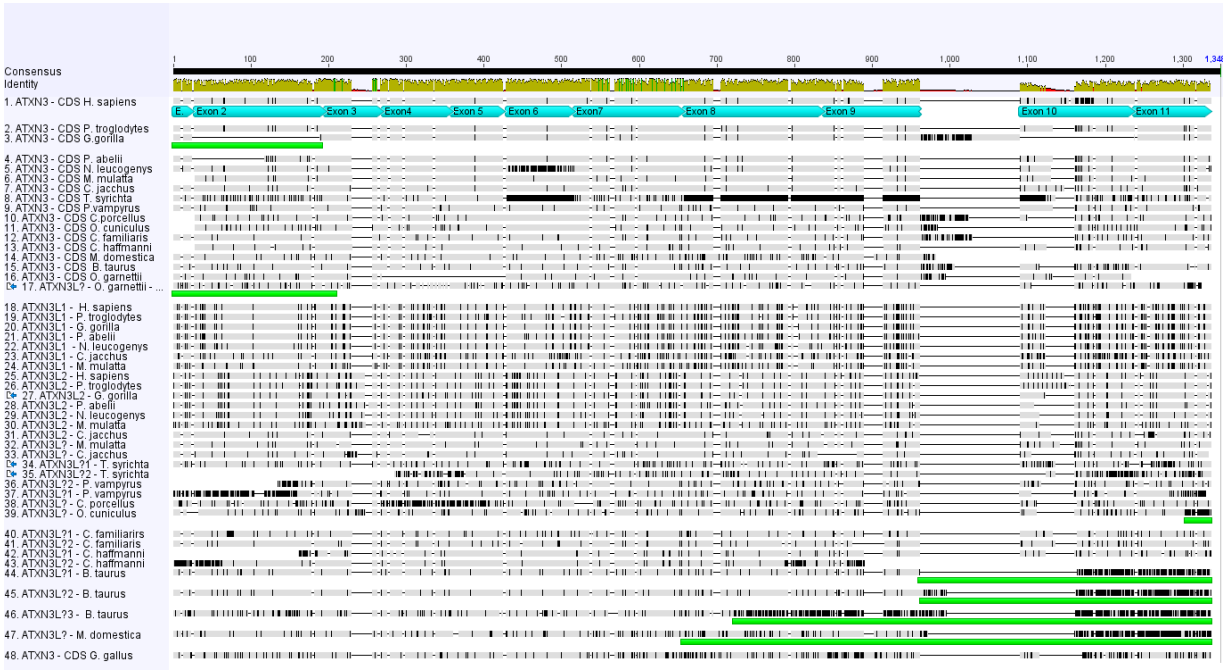

Figure A1 – Alignment of all sequences searched on Ensembl, NCBI and UCSC Genome Browser databases. From top to bottom: *Gallus gallus* as outgroup; ATXN3 coding sequences for primates; *ATXN3* coding sequences for non-primate mammals; *ATXN3L1* for primates; *ATXN3L2* for primates; *ATXN3L?* of primates and non-primate mammals. The human exons present in transcript ATXN3-001 were annotated in the *ATXN3* coding sequence to serve as reference. Green annotations correspond to sequence portions added to those sequences obtained from trace archives.

Table A1– Matrix of genetic distances among *ATXN3*, *ATXN3L1* and *ATXN3L2* of several primates. Cells with darker background represent higher similarity between two sequences, whereas lighter cells show lower similarity. Notably, a lighter gray means further apart while darker means closer together. The longest distance in the tree is white and the shortest distance is black (may not be zero).

	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXNIL...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3 - ...
ATXN3 - CDS H. sapiens		94.9%	64.8%	86.6%	93.7%	94.3%	88.5%	60.4%	72.0%	76.4%	74.3%	76.8%	75.9%	76.0%	72.7%	76.3%	81.5%	82.3%	80.2%	78.4%	79.4%	76.4%	88.9%	93.0%	83.9%	77.4%	69.3%	75.2%	72.3%
ATXN3 - CDS P. troglodytes	94.9%		66.3%	90.1%	97.7%	96.2%	92.5%	62.1%	75.3%	79.8%	77.6%	80.2%	79.2%	79.2%	75.7%	79.6%	82.6%	82.6%	83.6%	81.8%	82.7%	79.5%	92.9%	95.1%	87.5%	80.6%	71.6%	78.6%	74.9%
ATXN3 - CDS G.gorilla	64.8%	66.3%		71.5%	65.3%	64.6%	60.9%	40.0%	48.7%	54.0%	52.2%	54.2%	53.6%	53.6%	51.5%	53.6%	55.7%	55.8%	56.8%	57.1%	57.1%	54.9%	61.8%	63.4%	60.4%	54.4%	60.1%	55.6%	53.8%
ATXN3 - CDS P. abelii	86.6%	90.1%	71.5%		89.4%	87.5%	84.2%	55.4%	67.1%	71.8%	69.7%	72.3%	71.4%	71.4%	68.1%	71.8%	74.9%	74.9%	76.0%	74.1%	75.1%	72.2%	84.5%	87.3%	79.2%	72.4%	71.8%	71.6%	67.9%
ATXN3 - CDS M. mulatta	93.7%	97.7%	65.3%	89.4%		95.0%	92.4%	61.2%	74.4%	79.5%	77.3%	80.0%	78.9%	78.9%	75.4%	79.3%	82.0%	82.0%	83.5%	81.4%	82.6%	79.7%	92.5%	95.3%	87.0%	79.9%	72.0%	78.6%	74.8%
ATXN3 - CDS C. jacchus	94.3%	96.2%	64.6%	87.5%	95.0%		89.7%	61.0%	74.3%	78.2%	76.4%	78.7%	77.7%	77.7%	74.4%	78.3%	83.0%	83.0%	82.1%	80.1%	81.2%	78.2%	93.0%	93.6%	87.4%	79.4%	70.5%	77.2%	74.2%
ATXN3 - CDS N. leucogenys	88.5%	92.5%	60.9%	84.2%	92.4%	89.7%		61.8%	69.7%	74.9%	73.2%	75.4%	74.7%	74.5%	72.0%	75.1%	78.2%	78.2%	79.5%	77.8%	78.4%	76.0%	87.4%	89.1%	82.2%	76.1%	64.9%	74.9%	69.5%
ATXN3 - CDS T. syrichta	60.4%	62.1%	40.0%	55.4%	61.2%	61.0%	61.8%		44.1%	54.0%	52.6%	53.9%	53.9%	54.1%	53.4%	54.3%	55.3%	54.6%	55.0%	54.1%	54.3%	53.2%	58.7%	60.8%	57.2%	55.6%	51.1%	51.6%	53.9%
ATXN3 - CDS O. garnettii	72.0%	75.3%	48.7%	67.1%	74.4%	74.3%	69.7%	44.1%		65.6%	63.8%	66.2%	64.7%	65.3%	62.0%	66.0%	65.2%	65.2%	66.7%	64.4%	65.9%	63.9%	75.3%	72.4%	65.9%	67.2%	58.5%	67.4%	58.8%
ATXN3L1 - H. sapiens	76.4%	79.8%	54.0%	71.8%	79.5%	78.2%	74.9%	54.0%	65.6%		96.1%	98.4%	97.2%	95.9%	82.8%	93.0%	69.8%	69.9%	72.2%	70.4%	71.9%	68.8%	76.8%	76.7%	72.4%	73.6%	59.4%	69.2%	66.2%
ATXN3L1 - P. troglodytes	74.3%	77.6%	52.2%	69.7%	77.3%	76.4%	73.2%	52.6%	63.8%	96.1%		96.2%	94.9%	93.9%	80.8%	91.2%	67.6%	67.7%	70.0%	68.2%	69.8%	66.8%	74.9%	74.7%	70.7%	71.6%	57.2%	67.3%	64.5%
ATXN3L1 - G. gorilla	76.8%	80.2%	54.2%	72.3%	80.0%	78.7%	75.4%	53.9%	66.2%	98.4%	96.2%		97.7%	96.5%	83.4%	93.8%	70.1%	70.2%	72.6%	70.7%	72.3%	69.5%	77.2%	77.2%	72.9%	74.3%	59.6%	69.4%	66.3%
ATXN3L1 - P. abelii	75.9%	79.2%	53.6%	71.4%	78.9%	77.7%	74.7%	53.9%	64.7%	97.2%	94.9%	97.7%		95.7%	82.7%	93.1%	69.1%	69.1%	71.5%	69.7%	71.2%	68.2%	76.0%	76.4%	72.0%	73.5%	58.7%	68.8%	66.1%
ATXN3L1 - N. leucogenys	76.0%	79.2%	53.6%	71.4%	78.9%	77.7%	74.5%	54.1%	65.3%	95.9%	93.9%	96.5%	95.7%		83.5%	93.1%	69.6%	69.7%	72.0%	70.1%	71.7%	69.1%	76.0%	76.3%	72.1%	73.3%	58.8%	68.4%	65.4%
ATXN3L1 - C. jacchus	72.7%	75.7%	51.5%	68.1%	75.4%	74.4%	72.0%	53.4%	62.0%	82.8%	80.8%	83.4%	82.7%	83.5%		83.7%	66.1%	66.0%	68.4%	67.2%	68.0%	65.8%	73.2%	73.3%	69.4%	70.1%	57.8%	65.5%	64.2%
ATXN3L1 - M. mulatta	76.3%	79.6%	53.6%	71.8%	79.3%	78.3%	75.1%	54.3%	66.0%	93.0%	91.2%	93.8%	93.1%	93.1%	83.3%		70.1%	70.0%	72.5%	70.7%	72.0%	69.7%	76.7%	76.6%	72.6%	73.4%	59.2%	69.3%	66.5%
ATXN3L2 - H. sapiens	81.5%	82.6%	55.7%	74.9%	82.0%	83.0%	78.2%	55.3%	65.2%	69.8%	67.6%	70.1%	69.1%	69.6%	66.1%	70.1%		97.7%	94.7%	91.9%	92.3%	88.0%	78.9%	81.1%	74.9%	71.8%	63.5%	69.7%	65.3%
ATXN3L2 - P. troglodytes	82.3%	82.6%	55.8%	74.9%	82.0%	83.0%	78.2%	54.6%	65.2%	69.9%	67.7%	70.2%	69.1%	69.7%	66.0%	70.0%	97.7%		94.9%	92.2%	92.5%	88.4%	78.8%	81.2%	74.7%	71.8%	63.5%	69.8%	65.3%
ATXN3L2 - G. gorilla	80.2%	83.6%	56.8%	76.0%	83.5%	82.1%	79.5%	55.0%	66.7%	72.2%	70.0%	72.6%	71.5%	72.0%	68.4%	72.5%	94.7%	94.9%		94.8%	95.5%	91.2%	80.6%	81.0%	77.3%	72.3%	63.6%	71.7%	67.6%
ATXN3L2 - P. abelii	78.4%	81.8%	57.1%	74.1%	81.4%	80.1%	77.8%	54.1%	64.4%	70.4%	68.2%	70.7%	69.7%	70.1%	67.2%	70.7%	91.9%	92.2%	94.8%		94.8%	90.3%	78.7%	79.1%	77.5%	70.9%	64.6%	72.1%	66.7%
ATXN3L2 - N. leucogenys	79.4%	82.7%	57.1%	75.1%	82.6%	81.2%	78.4%	54.3%	65.9%	71.9%	69.8%	72.3%	71.2%	71.7%	68.0%	72.0%	92.3%	92.5%	95.5%	94.8%		91.3%	79.3%	80.2%	77.5%	71.9%	63.5%	72.1%	68.1%
ATXN3L2 - M. mulatta	76.4%	79.5%	54.9%	72.2%	79.7%	78.2%	76.0%	53.2%	63.9%	68.8%	66.8%	69.5%	68.2%	69.1%	65.8%	69.7%	88.0%	88.4%	91.2%	90.3%	91.3%		76.6%	77.6%	75.6%	69.1%	62.5%	69.4%	65.6%
ATXN3L2 - C. jacchus	88.9%	92.9%	61.8%	84.5%	92.5%	93.0%	87.4%	58.7%	75.3%	76.8%	74.9%	77.2%	76.0%	76.0%	73.2%	76.7%	78.9%	78.8%	80.6%	78.7%	79.3%	76.6%		89.4%	85.8%	78.1%	67.8%	75.7%	72.1%
ATXN3L? - M. mulatta	93.0%	95.1%	63.4%	87.3%	95.3%	93.6%	89.1%	60.8%	72.4%	76.7%	74.7%	77.2%	76.4%	76.3%	73.3%	76.6%	81.1%	81.2%	81.0%	79.1%	80.2%	77.6%	89.4%		84.9%	77.8%	70.1%	76.0%	72.3%
ATXN3L? - C. jacchus	83.9%	87.5%	60.4%	79.2%	87.0%	87.4%	82.2%	57.2%	65.9%	72.4%	70.7%	72.9%	72.0%	72.1%	69.4%	72.6%	74.9%	74.7%	77.3%	77.5%	77.5%	75.6%	85.8%	84.9%		72.3%	64.7%	73.9%	70.3%
ATXN3L?1 - T. syrichta	77.4%	80.6%	54.4%	72.4%	79.9%	79.4%	76.1%	55.6%	67.2%	73.6%	71.6%	74.3%	73.5%	73.3%	70.1%	73.4%	71.8%	71.8%	72.3%	70.9%	71.9%	69.1%	78.1%	77.8%	72.3%		62.5%	69.8%	65.2%
ATXN3L?2 - T. syrichta	69.3%	71.6%	60.1%	71.8%	72.0%	70.5%	64.9%	51.1%	58.5%	59.4%	57.2%	59.6%	58.7%	58.8%	57.8%	59.2%	63.5%	63.5%	63.6%	64.6%	63.5%	62.5%	67.8%	70.1%	64.7%	62.5%		62.4%	56.8%
ATXN3L? - O. garnettii - se...	75.2%	78.6%	55.6%	71.6%	78.6%	77.2%	74.9%	51.6%	67.4%	69.2%	67.3%	69.4%	68.8%	68.4%	65.5%	69.3%	69.7%	69.8%	71.7%	72.1%	72.1%	69.4%	75.7%	76.0%	73.9%	69.8%	62.4%		63.3%
ATXN3 - CDS G. gallus	72.3%	74.9%	53.8%	67.9%	74.8%	74.2%	69.5%	53.9%	58.8%	66.2%	64.5%	66.3%	66.1%	65.4%	64.2%	66.5%	65.3%	65.3%	67.6%	66.7%	68.1%	65.6%	72.1%	72.3%	70.3%	65.2%	56.8%	63.3%	

Table A2 -– Matrix of genetic distances among *ATXN3*, *ATXN3L1* and *ATXN3L2* of several primates, without the CAG tract. Cells with darker background represent higher similarity between two sequences, whereas lighter cells show lower similarity. Notably, a lighter gray means further apart while darker means closer together. The longest distance in the tree is white and the shortest distance is black (may not be zero).

	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3 - ...
ATXN3 - CDS H. sapiens		97.4%	68.6%	88.2%	96.4%	95.6%	90.9%	62.4%	73.2%	79.0%	76.8%	79.4%	78.6%	78.7%	75.5%	78.9%	82.3%	82.9%	82.6%	82.0%	82.4%	79.4%	91.3%	94.4%	88.3%	79.7%	73.6%	79.0%	75.2%
ATXN3 - CDS P. troglodytes	97.4%		68.8%	90.0%	98.2%	97.6%	92.7%	63.2%	74.7%	80.6%	78.4%	81.0%	80.1%	80.0%	76.9%	80.4%	83.7%	84.3%	84.0%	83.4%	83.8%	80.6%	93.1%	96.2%	89.7%	81.5%	74.5%	80.4%	75.9%
ATXN3 - CDS G.gorilla	68.6%	68.8%		74.3%	67.6%	67.7%	63.0%	41.6%	50.6%	55.5%	53.7%	55.8%	55.2%	55.1%	53.0%	55.1%	58.4%	58.8%	58.4%	58.0%	58.3%	56.0%	63.9%	66.2%	61.3%	56.6%	61.0%	56.3%	54.9%
ATXN3 - CDS P. abelii	88.2%	90.0%	74.3%		89.5%	88.8%	84.2%	56.2%	66.2%	72.4%	70.2%	72.9%	72.0%	71.9%	68.9%	72.3%	75.8%	76.3%	76.2%	75.6%	76.0%	73.2%	84.5%	87.6%	81.3%	73.1%	74.6%	73.4%	68.8%
ATXN3 - CDS M. mulatta	96.4%	98.2%	67.6%	89.5%		97.0%	92.3%	62.5%	74.1%	80.0%	77.7%	80.5%	79.5%	79.4%	76.3%	79.8%	83.3%	84.0%	83.7%	82.9%	83.5%	80.7%	92.5%	96.7%	89.2%	81.2%	75.2%	80.4%	75.7%
ATXN3 - CDS C. jacchus	95.6%	97.6%	67.7%	88.8%	97.0%		91.3%	62.6%	74.9%	80.2%	78.3%	80.7%	79.7%	79.6%	76.7%	80.3%	83.2%	83.7%	83.6%	82.8%	83.5%	80.4%	94.5%	94.7%	90.8%	81.1%	74.1%	80.0%	76.3%
ATXN3 - CDS N. leucogenys	90.9%	92.7%	63.0%	84.2%	92.3%	91.3%		63.1%	69.1%	75.1%	73.3%	75.6%	75.0%	74.7%	72.7%	75.3%	79.3%	79.8%	79.5%	79.0%	79.1%	76.7%	87.2%	90.1%	84.1%	77.0%	67.8%	76.5%	70.3%
ATXN3 - CDS T. syrichta	62.4%	63.2%	41.6%	56.2%	62.5%	62.6%	63.1%		44.8%	55.6%	54.1%	55.4%	55.4%	55.6%	54.8%	55.9%	56.6%	56.2%	56.4%	55.9%	56.0%	54.8%	59.9%	62.0%	59.3%	56.5%	52.4%	53.4%	55.6%
ATXN3 - CDS O. garnettii	73.2%	74.7%	50.6%	66.2%	74.1%	74.9%	69.1%	44.8%		65.8%	63.9%	66.5%	65.0%	65.5%	62.7%	66.3%	65.7%	66.2%	66.4%	65.4%	66.3%	64.4%	75.0%	72.5%	67.3%	67.6%	61.3%	68.9%	59.4%
ATXN3L1 - H. sapiens	79.0%	80.6%	55.5%	72.4%	80.0%	80.2%	75.1%	55.6%	65.8%		96.0%	98.4%	97.2%	95.7%	83.0%	92.8%	71.4%	72.0%	72.1%	71.3%	72.4%	69.2%	77.3%	78.3%	73.6%	75.0%	62.7%	70.5%	66.8%
ATXN3L1 - P. troglodytes	76.8%	78.4%	53.7%	70.2%	77.7%	78.3%	73.3%	54.1%	63.9%	96.0%		96.1%	94.9%	93.8%	81.1%	91.0%	69.2%	69.7%	69.8%	69.1%	70.1%	67.2%	75.3%	76.1%	71.9%	72.8%	60.3%	68.5%	65.0%
ATXN3L1 - G. gorilla	79.4%	81.0%	55.8%	72.9%	80.5%	80.7%	75.6%	55.4%	66.5%	98.4%	96.1%		97.7%	96.4%	83.7%	93.6%	71.8%	72.4%	72.5%	71.7%	72.7%	70.0%	77.7%	78.7%	74.1%	75.7%	62.9%	70.7%	66.9%
ATXN3L1 - P. abelii	78.6%	80.1%	55.2%	72.0%	79.5%	79.7%	75.0%	55.4%	65.0%	97.2%	94.9%	97.7%		95.7%	83.0%	92.9%	70.8%	71.3%	71.4%	70.7%	71.7%	68.7%	76.5%	78.0%	73.2%	75.0%	61.9%	70.1%	66.8%
ATXN3L1 - N. leucogenys	78.7%	80.0%	55.1%	71.9%	79.4%	79.6%	74.7%	55.6%	65.5%	95.7%	93.8%	96.4%	95.7%		83.7%	92.9%	71.2%	71.8%	71.9%	71.0%	72.1%	69.5%	76.4%	77.8%	73.3%	74.8%	62.1%	69.7%	65.9%
ATXN3L1 - C. jacchus	75.5%	76.9%	53.0%	68.9%	76.3%	76.7%	72.7%	54.8%	62.7%	83.0%	81.1%	83.7%	83.0%	83.7%		83.5%	68.0%	68.3%	68.6%	68.2%	68.7%	66.5%	74.1%	74.8%	70.8%	71.7%	60.9%	66.7%	64.7%
ATXN3L1 - M. mulatta	78.9%	80.4%	55.1%	72.3%	79.8%	80.3%	75.3%	55.9%	66.3%	92.8%	91.0%	93.6%	92.9%	92.9%	83.5%		71.7%	72.2%	72.4%	71.6%	72.5%	70.2%	77.1%	78.1%	73.8%	74.9%	62.4%	70.6%	67.1%
ATXN3L2 - H. sapiens	82.3%	83.7%	58.4%	75.8%	83.3%	83.2%	79.3%	56.6%	65.7%	71.4%	69.2%	71.8%	70.8%	71.2%	68.0%	71.7%		98.2%	97.2%	95.3%	95.3%	90.8%	80.0%	81.8%	77.8%	73.4%	66.7%	72.2%	67.2%
ATXN3L2 - P. troglodytes	82.9%	84.3%	58.8%	76.3%	84.0%	83.7%	79.8%	56.2%	66.2%	72.0%	69.7%	72.4%	71.3%	71.8%	68.3%	72.2%	98.2%		98.0%	96.3%	96.1%	91.9%	80.5%	82.4%	78.0%	73.7%	67.2%	72.9%	67.6%
ATXN3L2 - G. gorilla	82.6%	84.0%	58.4%	76.2%	83.7%	83.6%	79.5%	56.4%	66.4%	72.1%	69.8%	72.5%	71.4%	71.9%	68.6%	72.4%	97.2%	98.0%		96.0%	96.2%	91.9%	80.6%	82.2%	78.5%	73.2%	67.0%	72.7%	67.9%
ATXN3L2 - P. abelii	82.0%	83.4%	58.0%	75.6%	82.9%	82.8%	79.0%	55.9%	65.4%	71.3%	69.1%	71.7%	70.7%	71.0%	68.2%	71.6%	95.3%	96.3%	96.0%		95.3%	90.7%	79.8%	81.4%	77.6%	72.6%	67.0%	72.1%	66.9%
ATXN3L2 - N. leucogenys	82.4%	83.8%	58.3%	76.0%	83.5%	83.5%	79.1%	56.0%	66.3%	72.4%	70.1%	72.7%	71.7%	72.1%	68.7%	72.5%	95.3%	96.1%	96.2%	95.3%		91.2%	80.0%	82.0%	78.0%	73.3%	66.4%	72.5%	67.8%
ATXN3L2 - M. mulatta	79.4%	80.6%	56.0%	73.2%	80.7%	80.4%	76.7%	54.8%	64.4%	69.2%	67.2%	70.0%	68.7%	69.5%	66.5%	70.2%	90.8%	91.9%	91.9%	90.7%	91.2%		77.3%	79.5%	76.0%	70.5%	65.3%	69.8%	65.3%
ATXN3L?2 - C. jacchus	91.3%	93.1%	63.9%	84.5%	92.5%	94.5%	87.2%	59.9%	75.0%	77.3%	75.3%	77.7%	76.5%	76.4%	74.1%	77.1%	80.0%	80.5%	80.6%	79.8%	80.0%	77.3%		90.5%	87.8%	79.0%	70.8%	77.3%	73.0%
ATXN3L? - M. mulatta	94.4%	96.2%	66.2%	87.6%	96.7%	94.7%	90.1%	62.0%	72.5%	78.3%	76.1%	78.7%	78.0%	77.8%	74.8%	78.1%	81.8%	82.4%	82.2%	81.4%	82.0%	79.5%	90.5%		87.9%	79.0%	73.1%	78.7%	74.1%
ATXN3L?1 - C. jacchus	88.3%	89.7%	61.3%	81.3%	89.2%	90.8%	84.1%	59.3%	67.3%	73.6%	71.9%	74.1%	73.2%	73.3%	70.8%	73.8%	77.8%	78.0%	78.5%	77.6%	78.0%	76.0%	87.8%	87.9%		74.5%	67.0%	73.7%	70.8%
ATXN3L?1 - T. syrichta	79.7%	81.5%	56.6%	73.1%	81.2%	81.1%	77.0%	56.5%	67.6%	75.0%	72.8%	75.7%	75.0%	74.8%	71.7%	74.9%	73.4%	73.7%	73.2%	72.6%	73.3%	70.5%	79.0%	79.0%	74.5%		65.1%	71.8%	66.7%
ATXN3L?2 - T. syrichta	73.6%	74.5%	61.0%	74.6%	75.2%	74.1%	67.8%	52.4%	61.3%	62.7%	60.3%	62.9%	61.9%	62.1%	60.9%	62.4%	66.7%	67.2%	67.0%	67.0%	66.4%	65.3%	70.8%	73.1%	67.0%	65.1%		64.5%	59.4%
ATXN3L? - O. garnettii	79.0%	80.4%	56.3%	73.4%	80.4%	80.0%	76.5%	53.4%	68.9%	70.5%	68.5%	70.7%	70.1%	69.7%	66.7%	70.6%	72.2%	72.9%	72.7%	72.1%	72.5%	69.8%	77.3%	78.7%	73.7%	71.8%	64.5%		63.6%
ATXN3 - CDS G. gallus	75.2%	75.9%	54.9%	68.8%	75.7%	76.3%	70.3%	55.6%	59.4%	66.8%	65.0%	66.9%	66.8%	65.9%	64.7%	67.1%	67.2%	67.6%	67.9%	66.9%	67.8%	65.3%	73.0%	74.1%	70.8%	66.7%	59.4%	63.6%	

Table A3 - Matrix of genetic distances among *ATXN3*, *ATXN3L1* and *ATXN3L2* Josephin domain of several primates. Cells with darker background represent higher similarity between two sequences, whereas lighter cells show lower similarity. Notably, A lighter gray means further apart while darker means closer together. The longest distance in the tree is white and the shortest distance is black (may not be zero)

	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3 - ...
ATXN3 - CDS H. sapiens		99.4%	70.2%	82.1%	98.4%	97.4%	88.4%	77.1%	66.2%	86.0%	83.5%	86.5%	85.4%	85.8%	83.4%	85.6%	82.7%	82.5%	82.2%	81.8%	82.5%	79.8%	94.1%	97.0%	93.3%	87.6%	73.4%	80.0%	87.4%
ATXN3 - CDS P. troglodytes	99.4%		70.4%	82.3%	98.6%	98.0%	88.5%	77.3%	66.4%	86.5%	84.1%	87.1%	86.0%	86.3%	83.9%	86.1%	83.2%	83.1%	82.7%	82.4%	83.1%	79.9%	94.6%	97.2%	93.4%	88.0%	73.7%	80.2%	87.8%
ATXN3 - CDS G.gorilla	70.2%	70.4%		84.5%	68.6%	69.3%	59.1%	54.3%	40.3%	60.8%	58.2%	61.2%	60.3%	60.8%	57.9%	60.1%	59.1%	58.8%	58.2%	58.0%	58.5%	56.3%	66.0%	68.4%	65.8%	62.1%	74.1%	57.2%	63.8%
ATXN3 - CDS P. abelii	82.1%	82.3%	84.5%		80.6%	81.0%	71.0%	63.6%	50.8%	71.2%	68.8%	71.7%	70.8%	71.2%	68.8%	70.8%	68.1%	67.9%	67.9%	67.7%	68.3%	65.7%	77.6%	80.0%	77.0%	72.3%	73.7%	67.0%	73.4%
ATXN3 - CDS M. mulatta	98.4%	98.6%	68.6%	80.6%		96.9%	87.8%	76.4%	64.3%	86.2%	83.7%	86.8%	85.7%	86.0%	83.3%	85.9%	82.6%	82.6%	82.3%	81.9%	82.6%	80.3%	93.4%	96.9%	92.6%	87.8%	74.4%	79.8%	88.0%
ATXN3 - CDS C. jacchus	97.4%	98.0%	69.3%	81.0%	96.9%		86.9%	76.2%	66.5%	86.5%	84.7%	87.1%	86.0%	86.3%	83.9%	86.3%	82.4%	82.2%	82.2%	81.8%	82.5%	79.6%	96.1%	95.6%	94.7%	87.8%	73.0%	80.2%	88.5%
ATXN3 - CDS N. leucogenys	88.4%	88.5%	59.1%	71.0%	87.8%	86.9%		76.9%	56.4%	77.3%	75.6%	77.8%	77.4%	77.1%	76.5%	77.1%	75.0%	74.7%	74.3%	74.1%	74.3%	71.9%	83.5%	86.3%	83.2%	80.2%	57.0%	73.1%	77.4%
ATXN3 - CDS T. syrichta	77.1%	77.3%	54.3%	63.6%	76.4%	76.2%	76.9%		50.3%	70.2%	68.8%	70.4%	70.1%	70.6%	70.2%	70.8%	67.2%	66.7%	67.0%	66.1%	66.5%	65.0%	73.4%	75.4%	72.5%	72.6%	57.3%	64.4%	73.2%
ATXN3 - CDS O. garnettii	66.2%	66.4%	40.3%	50.8%	64.3%	66.5%	56.4%	50.3%		61.9%	60.6%	62.3%	61.4%	61.6%	60.6%	63.0%	56.8%	56.6%	57.1%	56.5%	57.1%	56.0%	68.1%	64.5%	62.7%	62.1%	47.0%	60.7%	60.6%
ATXN3L1 - H. sapiens	86.0%	86.5%	60.8%	71.2%	86.2%	86.5%	77.3%	70.2%	61.9%		96.9%	99.3%	98.9%	98.2%	86.1%	93.9%	74.7%	74.9%	75.0%	74.4%	75.8%	72.8%	84.1%	84.5%	82.3%	83.2%	67.9%	73.1%	80.8%
ATXN3L1 - P. troglodytes	83.5%	84.1%	58.2%	68.8%	83.7%	84.7%	75.6%	68.8%	60.6%	96.9%		96.9%	96.5%	96.1%	83.7%	91.9%	72.4%	72.5%	72.7%	72.1%	73.4%	70.5%	82.3%	82.3%	80.5%	81.1%	64.8%	71.5%	78.7%
ATXN3L1 - G. gorilla	86.5%	87.1%	61.2%	71.7%	86.8%	87.1%	77.8%	70.4%	62.3%	99.3%	96.9%		98.9%	98.2%	86.1%	94.3%	75.2%	75.4%	75.6%	75.0%	76.3%	73.4%	84.7%	85.0%	82.9%	83.7%	68.3%	73.7%	80.8%
ATXN3L1 - P. abelii	85.4%	86.0%	60.3%	70.8%	85.7%	86.0%	77.4%	70.1%	61.4%	98.9%	96.5%	98.9%		97.8%	85.6%	93.5%	74.2%	74.3%	74.5%	73.9%	75.2%	72.3%	83.5%	84.3%	81.8%	83.2%	66.9%	73.3%	80.4%
ATXN3L1 - N. leucogenys	85.8%	86.3%	60.8%	71.2%	86.0%	86.3%	77.1%	70.6%	61.6%	98.2%	96.1%	98.2%	97.8%		85.6%	93.9%	74.7%	74.9%	75.0%	74.4%	75.8%	73.2%	83.9%	84.3%	82.1%	83.0%	67.6%	73.1%	80.0%
ATXN3L1 - C. jacchus	83.4%	83.9%	57.9%	68.8%	83.3%	83.9%	76.5%	70.2%	60.6%	86.1%	83.7%	86.1%	85.6%	85.6%		86.0%	71.5%	71.3%	72.0%	71.1%	72.0%	69.8%	81.9%	82.3%	79.8%	80.6%	66.2%	70.4%	77.4%
ATXN3L1 - M. mulatta	85.6%	86.1%	60.1%	70.8%	85.9%	86.3%	77.1%	70.8%	63.0%	93.9%	91.9%	94.3%	93.5%	93.9%	86.0%		74.7%	74.9%	75.2%	74.6%	75.6%	73.0%	83.9%	84.1%	82.7%	82.4%	67.6%	74.0%	80.8%
ATXN3L2 - H. sapiens	82.7%	83.2%	59.1%	68.1%	82.6%	82.4%	75.0%	67.2%	56.8%	74.7%	72.4%	75.2%	74.2%	74.7%	71.5%	74.7%		98.6%	97.1%	95.9%	95.7%	90.0%	79.5%	81.8%	80.7%	76.5%	67.2%	69.9%	75.8%
ATXN3L2 - P. troglodytes	82.5%	83.1%	58.8%	67.9%	82.6%	82.2%	74.7%	66.7%	56.6%	74.9%	72.5%	75.4%	74.3%	74.9%	71.3%	74.9%	98.6%		97.5%	96.3%	96.1%	90.7%	79.3%	81.8%	80.4%	76.8%	67.6%	69.9%	75.9%
ATXN3L2 - G. gorilla	82.2%	82.7%	58.2%	67.9%	82.3%	82.2%	74.3%	67.0%	57.1%	75.0%	72.7%	75.6%	74.5%	75.0%	72.0%	75.2%	97.1%	97.5%		95.9%	95.7%	91.3%	80.0%	81.6%	81.4%	75.9%	67.2%	69.5%	76.3%
ATXN3L2 - P. abelii	81.8%	82.4%	58.0%	67.7%	81.9%	81.8%	74.1%	66.1%	56.5%	74.4%	72.1%	75.0%	73.9%	74.4%	71.1%	74.6%	95.9%	96.3%	95.9%		95.4%	90.3%	79.0%	81.0%	80.6%	75.7%	67.2%	69.5%	75.7%
ATXN3L2 - N. leucogenys	82.5%	83.1%	58.5%	68.3%	82.6%	82.5%	74.3%	66.5%	57.1%	75.8%	73.4%	76.3%	75.2%	75.8%	72.0%	75.6%	95.7%	96.1%	95.7%	95.4%		91.1%	79.7%	82.0%	80.9%	76.1%	67.2%	69.9%	76.8%
ATXN3L2 - M. mulatta	79.8%	79.9%	56.3%	65.7%	80.3%	79.6%	71.9%	65.0%	56.0%	72.8%	70.5%	73.4%	72.3%	73.2%	69.8%	73.0%	90.0%	90.7%	91.3%	90.3%	91.1%		77.1%	79.8%	79.4%	73.4%	68.6%	67.5%	73.7%
ATXN3L?2 - C. jacchus	94.1%	94.6%	66.0%	77.6%	93.4%	96.1%	83.5%	73.4%	68.1%	84.1%	82.3%	84.7%	83.5%	83.9%	81.9%	83.9%	79.5%	79.3%	80.0%	79.0%	79.7%	77.1%		92.2%	91.8%	84.7%	71.4%	78.9%	85.8%
ATXN3L? - M. mulatta	97.0%	97.2%	68.4%	80.0%	96.9%	95.6%	86.3%	75.4%	64.5%	84.5%	82.3%	85.0%	84.3%	84.3%	82.3%	84.1%	81.8%	81.8%	81.6%	81.0%	82.0%	79.8%	92.2%		91.4%	85.6%	72.4%	78.8%	86.0%
ATXN3L?1 - C. jacchus	93.3%	93.4%	65.8%	77.0%	92.6%	94.7%	83.2%	72.5%	62.7%	82.3%	80.5%	82.9%	81.8%	82.1%	79.8%	82.7%	80.7%	80.4%	81.4%	80.6%	80.9%	79.4%	91.8%	91.4%		83.1%	71.7%	77.5%	84.2%
ATXN3L?1 - T. syrichta	87.6%	88.0%	62.1%	72.3%	87.8%	87.8%	80.2%	72.6%	62.1%	83.2%	81.1%	83.7%	83.2%	83.0%	80.6%	82.4%	76.5%	76.8%	75.9%	75.7%	76.1%	73.4%	84.7%	85.6%	83.1%		67.6%	75.7%	80.8%
ATXN3L?2 - T. syrichta	73.4%	73.7%	74.1%	73.7%	74.4%	73.0%	57.0%	57.3%	47.0%	67.9%	64.8%	68.3%	66.9%	67.6%	66.2%	67.6%	67.2%	67.6%	67.2%	67.2%	67.2%	68.6%	71.4%	72.4%	71.7%	67.6%		62.1%	71.3%
ATXN3L? - O. garnettii	80.0%	80.2%	57.2%	67.0%	79.8%	80.2%	73.1%	64.4%	60.7%	73.1%	71.5%	73.7%	73.3%	73.1%	70.4%	74.0%	69.9%	69.9%	69.5%	69.5%	69.9%	67.5%	78.9%	78.8%	77.5%	75.7%	62.1%		71.5%
ATXN3 - CDS G. gallus	87.4%	87.8%	63.8%	73.4%	88.0%	88.5%	77.4%	73.2%	60.6%	80.8%	78.7%	80.8%	80.4%	80.0%	77.4%	80.8%	75.8%	75.9%	76.3%	75.7%	76.8%	73.7%	85.8%	86.0%	84.2%	80.8%	71.3%	71.5%	

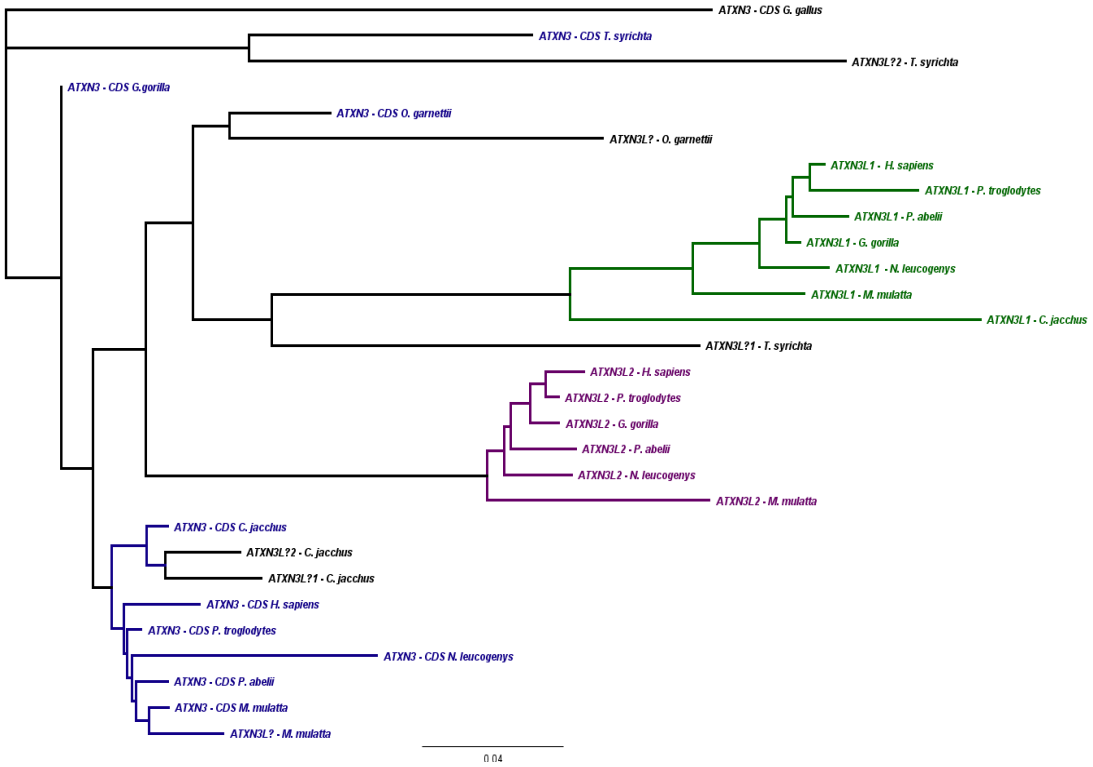

Figure A2 –Phylogenetic tree of *ATXN3*, *ATXN3L1* and *ATXN3L2* for primates, without the GAG tract, based on the genetic distances presented on Table A2. *Gallus gallus* was used as outgroup.

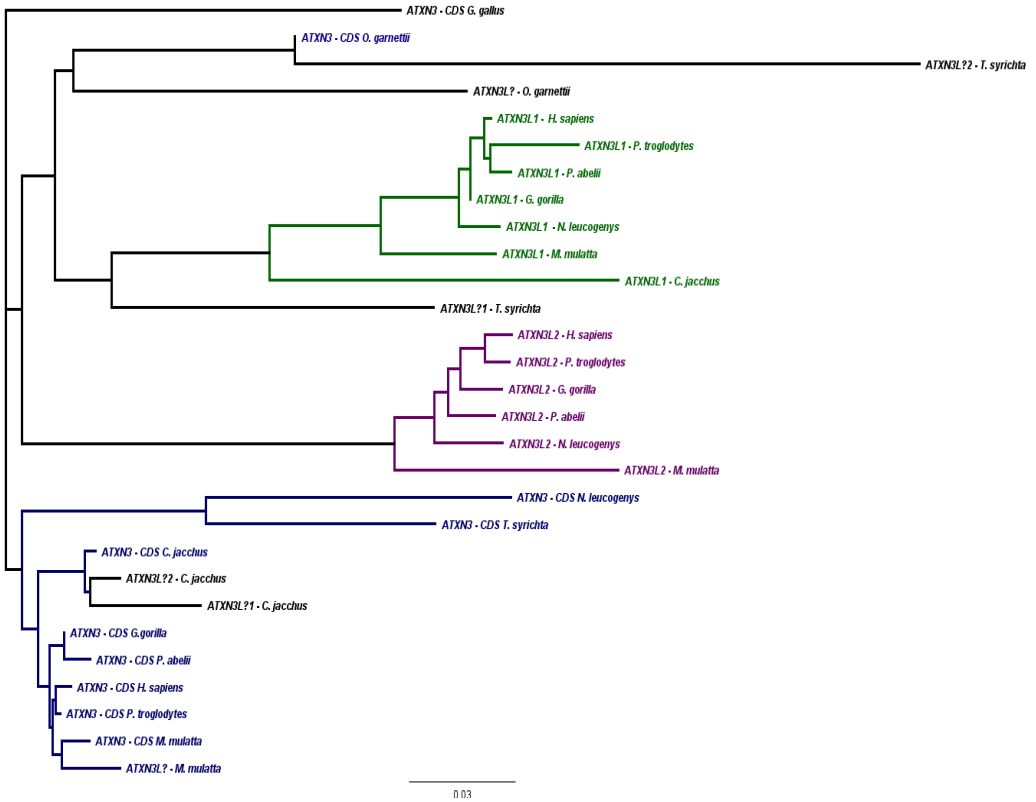

Figure A3 – Phylogenetic tree of *ATXN3*, *ATXN3L1* and *ATXN3L2* Josephin domain for primates, based on the genetic distances presented on Table A3. *Gallus gallus* was used as outgroup.

Table A4 - Matrix of genetic distances among ATXN3, ATXN3L1 and ATXN3L2 complete sequences of several mammals. Cells with darker background represent higher similarity between two sequences, whereas lighter cells show lower similarity. Notably, a lighter gray means further apart while darker means closer together. The longest distance in the tree is white and the shortest distance is black (may not be zero). Continue on the next page.

	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...
ATXN3 - CDS H. sapiens		94.9%	64.8%	86.6%	88.5%	93.7%	94.3%	60.4%	89.1%	80.7%	81.2%	80.7%	88.9%	84.4%	81.8%	72.0%	76.4%	74.3%	76.8%	75.9%	76.0%	72.7%	76.3%
ATXN3 - CDS P. troglodytes	94.9%		66.3%	90.1%	92.5%	97.7%	96.2%	62.1%	93.0%	83.9%	84.6%	83.6%	92.7%	84.8%	84.9%	75.3%	79.8%	77.6%	80.2%	79.2%	79.2%	75.7%	79.6%
ATXN3 - CDS G.gorilla	64.8%	66.3%		71.5%	60.9%	65.3%	64.6%	40.0%	63.0%	63.8%	63.9%	66.2%	62.9%	70.3%	63.2%	48.7%	54.0%	52.2%	54.2%	53.6%	53.6%	51.5%	53.6%
ATXN3 - CDS P. abelii	86.6%	90.1%	71.5%		84.2%	89.4%	87.5%	55.4%	85.2%	76.2%	76.7%	75.9%	84.0%	75.2%	77.1%	67.1%	71.8%	69.7%	72.3%	71.4%	71.4%	68.1%	71.8%
ATXN3 - CDS N. leucogenys	88.5%	92.5%	60.9%	84.2%		92.4%	89.7%	61.8%	87.4%	78.9%	78.8%	78.9%	87.5%	79.0%	80.2%	69.7%	74.9%	73.2%	75.4%	74.7%	74.5%	72.0%	75.1%
ATXN3 - CDS M. mulatta	93.7%	97.7%	65.3%	89.4%	92.4%		95.0%	61.2%	92.4%	84.2%	84.5%	83.4%	92.5%	84.8%	84.7%	74.4%	79.5%	77.3%	80.0%	78.9%	78.9%	75.4%	79.3%
ATXN3 - CDS C. jacchus	94.3%	96.2%	64.6%	87.5%	89.7%	95.0%		61.0%	90.7%	82.9%	83.1%	82.4%	90.2%	85.3%	83.8%	74.3%	78.2%	76.4%	78.7%	77.7%	77.7%	74.4%	78.3%
ATXN3 - CDS T. syrichta	60.4%	62.1%	40.0%	55.4%	61.8%	61.2%	61.0%		62.0%	56.0%	58.7%	56.1%	60.7%	57.6%	57.5%	44.1%	54.0%	52.6%	53.9%	53.9%	54.1%	53.4%	54.3%
ATXN3 - CDS P.vampyrus	89.1%	93.0%	63.0%	85.2%	87.4%	92.4%	90.7%	62.0%		82.6%	83.6%	83.8%	91.1%	84.2%	85.6%	73.9%	78.7%	76.4%	78.8%	78.1%	78.3%	75.3%	79.3%
ATXN3 - CDS C.porcellus	80.7%	83.9%	63.8%	76.2%	78.9%	84.2%	82.9%	56.0%	82.6%		83.3%	85.7%	82.8%	83.4%	83.6%	69.2%	71.8%	69.9%	72.1%	71.0%	71.3%	68.5%	72.1%
ATXN3 - CDS O. cuniculus	81.2%	84.6%	63.9%	76.7%	78.8%	84.5%	83.1%	58.7%	83.6%	83.3%		84.5%	83.9%	82.7%	82.2%	69.3%	73.3%	71.0%	73.5%	72.6%	72.5%	69.8%	73.3%
ATXN3 - CDS C. familiaris	80.7%	83.6%	66.2%	75.9%	78.9%	83.4%	82.4%	56.1%	83.8%	85.7%	84.5%		84.2%	85.2%	85.6%	69.8%	71.3%	69.3%	71.7%	70.6%	70.9%	68.6%	71.6%
ATXN3 - CDS C. haffmanni	88.9%	92.7%	62.9%	84.0%	87.5%	92.5%	90.2%	60.7%	91.1%	82.8%	83.9%	84.2%		84.4%	84.9%	73.4%	78.6%	76.4%	79.1%	78.2%	78.1%	75.1%	79.1%
ATXN3 - CDS M. domestica	84.4%	84.8%	70.3%	75.2%	79.0%	84.8%	85.3%	57.6%	84.2%	83.4%	82.7%	85.2%	84.4%		84.1%	69.1%	74.7%	72.6%	74.8%	73.8%	73.7%	72.5%	74.7%
ATXN3 - CDS B. taurus	81.8%	84.9%	63.2%	77.1%	80.2%	84.7%	83.8%	57.5%	85.6%	83.6%	82.2%	85.6%	84.9%	84.1%		71.0%	72.9%	71.1%	73.5%	72.3%	72.9%	70.9%	74.4%
ATXN3 - CDS O. garnettii	72.0%	75.3%	48.7%	67.1%	69.7%	74.4%	74.3%	44.1%	73.9%	69.2%	69.3%	69.8%	73.4%	69.1%	71.0%		65.6%	63.8%	66.2%	64.7%	65.3%	62.0%	66.0%
ATXN3L1 - H. sapiens	76.4%	79.8%	54.0%	71.8%	74.9%	79.5%	78.2%	54.0%	78.7%	71.8%	73.3%	71.3%	78.6%	74.7%	72.9%	65.6%		96.1%	98.4%	97.2%	95.9%	82.8%	93.0%
ATXN3L1 - P. troglodytes	74.3%	77.6%	52.2%	69.7%	73.2%	77.3%	76.4%	52.6%	76.4%	69.9%	71.0%	69.3%	76.4%	72.6%	71.1%	63.8%	96.1%		96.2%	94.9%	93.9%	80.8%	91.2%
ATXN3L1 - G. gorilla	76.8%	80.2%	54.2%	72.3%	75.4%	80.0%	78.7%	53.9%	78.8%	72.1%	73.5%	71.7%	79.1%	74.8%	73.5%	66.2%	98.4%	96.2%		97.7%	96.5%	83.4%	93.8%
ATXN3L1 - P. abelii	75.9%	79.2%	53.6%	71.4%	74.7%	78.9%	77.7%	53.9%	78.1%	71.0%	72.6%	70.6%	78.2%	73.8%	72.3%	64.7%	97.2%	94.9%	97.7%		95.7%	82.7%	93.1%
ATXN3L1 - N. leucogenys	76.0%	79.2%	53.6%	71.4%	74.5%	78.9%	77.7%	54.1%	78.3%	71.3%	72.5%	70.9%	78.1%	73.7%	72.9%	65.3%	95.9%	93.9%	96.5%	95.7%		83.5%	93.1%
ATXN3L1 - C. jacchus	72.7%	75.7%	51.5%	68.1%	72.0%	75.4%	74.4%	53.4%	75.3%	68.5%	69.8%	68.6%	75.1%	72.5%	70.9%	62.0%	82.8%	80.8%	83.4%	82.7%	83.5%		83.3%
ATXN3L1 - M. mulatta	76.3%	79.6%	53.6%	71.8%	75.1%	79.3%	78.3%	54.3%	79.3%	72.1%	73.3%	71.6%	79.1%	74.7%	74.4%	66.0%	93.0%	91.2%	93.8%	93.1%	93.1%	83.3%	
ATXN3L2 - H. sapiens	81.5%	82.6%	55.7%	74.9%	78.2%	82.0%	83.0%	55.3%	80.8%	72.8%	73.5%	73.1%	80.1%	74.4%	74.1%	65.2%	69.8%	67.6%	70.1%	69.1%	69.6%	66.1%	70.1%
ATXN3L2 - P. troglodytes	82.3%	82.6%	55.8%	74.9%	78.2%	82.0%	83.0%	54.6%	80.7%	72.8%	73.7%	73.2%	80.2%	74.8%	74.0%	65.2%	69.9%	67.7%	70.2%	69.1%	69.7%	66.0%	70.0%
ATXN3L2 - G. gorilla	80.2%	83.6%	56.8%	76.0%	79.5%	83.5%	82.1%	55.0%	82.0%	74.7%	75.6%	75.1%	82.8%	75.1%	76.3%	66.7%	72.2%	70.0%	72.6%	71.5%	72.0%	68.4%	72.5%
ATXN3L2 - P. abelii	78.4%	81.8%	57.1%	74.1%	77.8%	81.4%	80.1%	54.1%	80.0%	74.7%	75.9%	75.4%	80.9%	74.1%	75.7%	64.4%	70.4%	68.2%	70.7%	69.7%	70.1%	67.2%	70.7%
ATXN3L2 - N. leucogenys	79.4%	82.7%	57.1%	75.1%	78.4%	82.6%	81.2%	54.3%	80.9%	75.0%	76.0%	75.3%	82.0%	74.4%	76.8%	65.9%	71.9%	69.8%	72.3%	71.2%	71.7%	68.0%	72.0%
ATXN3L2 - M. mulatta	76.4%	79.5%	54.9%	72.2%	76.0%	79.7%	78.2%	53.2%	78.2%	72.3%	73.0%	73.2%	78.9%	72.0%	74.7%	63.9%	68.8%	66.8%	69.5%	68.2%	69.1%	65.8%	69.7%
ATXN3L?2 - C. jacchus	88.9%	92.9%	61.8%	84.5%	87.4%	92.5%	93.0%	58.7%	87.9%	80.4%	80.1%	80.1%	87.6%	82.7%	81.4%	75.3%	76.8%	74.9%	77.2%	76.0%	76.0%	73.2%	76.7%
ATXN3L? - M. mulatta	93.0%	95.1%	63.4%	87.3%	89.1%	95.3%	93.6%	60.8%	89.5%	81.4%	81.5%	80.9%	89.1%	83.4%	82.0%	72.4%	76.7%	74.7%	77.2%	76.4%	76.3%	73.3%	76.6%
ATXN3L?1 - C. jacchus	83.9%	87.5%	60.4%	79.2%	82.2%	87.0%	87.4%	57.2%	82.6%	78.9%	78.3%	78.7%	83.6%	77.9%	79.1%	65.9%	72.4%	70.7%	72.9%	72.0%	72.1%	69.4%	72.6%
ATXN3L?1 - T. syrichta	77.4%	80.6%	54.4%	77.4%	76.1%	79.9%	79.4%	55.6%	79.5%	71.4%	72.8%	72.5%	78.7%	76.9%	77.8%	67.2%	73.6%	71.6%	74.3%	73.5%	73.3%	70.1%	73.4%
ATXN3L?2 - T. syrichta	69.3%	71.6%	60.1%	71.8%	64.9%	72.0%	70.5%	51.1%	71.8%	64.9%	72.6%	66.5%	70.5%	69.5%	65.5%	58.5%	59.4%	57.2%	59.6%	58.7%	58.8%	57.8%	59.2%
ATXN3L? - O. garnettii	75.2%	78.6%	55.6%	71.6%	74.9%	78.6%	77.2%	51.6%	76.8%	73.8%	74.2%	73.1%	78.3%	72.9%	73.3%	67.4%	69.2%	67.3%	69.4%	68.8%	68.4%	65.5%	69.3%
ATXN3L?2 - P. vampyrus	78.2%	81.8%	63.2%	81.5%	77.1%	81.7%	79.7%	52.8%	83.9%	75.3%	76.0%	75.0%	82.3%	73.9%	76.4%	63.6%	70.4%	68.4%	70.7%	70.2%	70.2%	66.8%	70.9%
ATXN3L?1 - P. vampyrus	72.2%	75.6%	58.8%	73.9%	71.5%	75.9%	74.1%	48.9%	77.3%	71.6%	72.6%	70.8%	75.8%	68.2%	71.9%	58.6%	64.3%	62.5%	64.5%	63.7%	63.8%	61.0%	64.4%
ATXN3L? - C. porcellus	70.8%	73.6%	51.3%	67.3%	69.2%	74.0%	73.0%	46.8%	72.3%	72.2%	68.8%	68.2%	73.9%	66.1%	70.0%	62.9%	64.2%	62.9%	64.6%	63.9%	64.0%	63.2%	64.5%
ATXN3L? - O. cuniculus	77.5%	80.8%	56.4%	73.8%	76.8%	81.0%	79.6%	54.1%	80.1%	75.6%	80.5%	75.0%	80.9%	75.8%	76.6%	65.3%	69.6%	68.0%	70.2%	69.5%	69.1%	66.8%	69.9%
ATXN3L?1 - C. familiarirs	85.1%	88.8%	60.5%	81.2%	83.4%	88.2%	87.1%	58.8%	89.2%	78.9%	79.9%	84.0%	87.2%	80.6%	82.6%	70.9%	74.7%	72.5%	74.9%	73.8%	74.7%	71.4%	75.2%
ATXN3L?2 - C. familiaris	86.5%	89.5%	60.3%	81.2%	84.1%	88.9%	88.4%	60.2%	90.0%	79.7%	80.6%	86.3%	88.7%	81.8%	83.6%	71.5%	75.2%	73.2%	75.7%	74.4%	74.9%	72.1%	75.9%
ATXN3L?1 - C. haffmanni	85.3%	89.5%	69.1%	89.0%	83.0%	89.1%	86.9%	55.8%	87.7%	79.3%	81.4%	79.4%	94.9%	79.9%	81.1%	68.8%	75.0%	72.5%	75.5%	74.5%	74.4%	71.2%	75.2%
ATXN3L?2 - C. haffmanni	84.9%	84.9%	70.1%	77.0%	78.1%	86.8%	84.5%	58.4%	83.9%	84.2%	83.7%	84.7%	90.4%	78.4%	83.4%	65.1%	75.7%	73.5%	76.3%	75.3%	74.7%	73.2%	74.9%
ATXN3L?1 - B. taurus	75.1%	77.2%	58.1%	69.4%	72.2%	77.1%	76.2%	52.0%	77.6%	73.4%	75.0%	76.5%	76.7%	81.5%	79.8%	63.8%	68.1%	66.3%	68.6%	67.5%	67.8%	66.7%	69.5%
ATXN3L?2 - B. taurus	73.8%	75.9%	59.6%	68.1%	71.1%	75.5%	74.6%	51.2%	75.9%	75.0%	76.1%	77.9%	75.7%	84.3%	82.2%	65.6%	65.8%	64.2%	66.5%	65.4%	65.8%	64.7%	67.1%
ATXN3L?3 - B. taurus	56.8%	57.8%	44.5%	51.1%	53.9%	57.3%	57.0%	45.3%	57.3%	57.7%	57.8%	60.6%	57.3%	64.6%	60.4%	48.0%	51.7%	50.4%	51.6%	51.3%	52.2%	50.2%	52.1%
ATXN3L? - M. domestica	67.4%	69.4%	52.4%	61.8%	64.9%	69.3%	68.7%	48.1%	68.4%	66.4%	68.9%	68.5%	69.8%	92.3%	66.3%	60.4%	62.0%	60.6%	62.3%	61.2%	61.4%	60.6%	61.8%
ATXN3 - CDS G. gallus	72.3%	74.9%	53.8%	67.9%	69.5%	74.8%	74.2%	53.9%	74.5%	70.2%	70.8%	69.4%	74.0%	78.9%	71.2%	58.8%	66.2%	64.5%	66.3%	66.1%	65.4%	64.2%	66.5%

Table A4 - Continuation

	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3 - ...
ATXN3 - CDS H. sapiens	81.5%	82.3%	80.2%	78.4%	79.4%	76.4%	88.9%	93.0%	83.9%	77.4%	69.3%	75.2%	78.2%	72.2%	70.8%	77.5%	85.1%	86.5%	85.3%	84.9%	75.1%	73.8%	56.8%	67.4%	72.3%
ATXN3 - CDS P. troglodytes	82.6%	82.6%	83.6%	81.8%	82.7%	79.5%	92.9%	95.1%	87.5%	80.6%	71.6%	78.6%	81.8%	75.6%	73.6%	80.8%	88.8%	89.5%	89.5%	84.9%	77.2%	75.9%	57.8%	69.4%	74.9%
ATXN3 - CDS G.gorilla	55.7%	55.8%	56.8%	57.1%	57.1%	54.9%	61.8%	63.4%	60.4%	54.4%	60.1%	55.6%	63.2%	58.8%	51.3%	56.4%	60.5%	60.3%	69.1%	70.1%	58.1%	59.6%	44.5%	52.4%	53.8%
ATXN3 - CDS P. abelii	74.9%	74.9%	76.0%	74.1%	75.1%	72.2%	84.5%	87.3%	79.2%	72.4%	71.8%	71.6%	81.5%	73.9%	67.3%	73.8%	81.2%	81.2%	89.0%	77.0%	69.4%	68.1%	51.1%	61.8%	67.9%
ATXN3 - CDS N. leucogenys	78.2%	78.2%	79.5%	77.8%	78.4%	76.0%	87.4%	89.1%	82.2%	76.1%	64.9%	74.9%	77.1%	71.5%	69.2%	76.8%	83.4%	84.1%	83.0%	78.1%	72.2%	71.1%	53.9%	64.9%	69.5%
ATXN3 - CDS M. mulatta	82.0%	82.0%	83.5%	81.4%	82.6%	79.7%	92.5%	95.3%	87.0%	79.9%	72.0%	78.6%	81.7%	75.9%	74.0%	81.0%	88.2%	88.9%	89.1%	86.8%	77.1%	75.5%	57.3%	69.3%	74.8%
ATXN3 - CDS C. jacchus	83.0%	83.0%	82.1%	80.1%	81.2%	78.2%	93.0%	93.6%	87.4%	79.4%	70.5%	77.2%	79.7%	74.1%	73.0%	79.6%	87.1%	88.4%	86.9%	84.5%	76.2%	74.6%	57.0%	68.7%	74.2%
ATXN3 - CDS T. syrichta	55.3%	54.6%	55.0%	54.1%	54.3%	53.2%	58.7%	60.8%	57.2%	55.6%	51.1%	51.6%	52.8%	48.9%	46.8%	54.1%	58.8%	60.2%	55.8%	58.4%	52.0%	51.2%	45.3%	48.1%	53.9%
ATXN3 - CDS P.vampyrus	80.8%	80.7%	82.0%	80.0%	80.9%	78.2%	87.9%	89.5%	82.6%	79.5%	71.8%	76.8%	83.9%	77.3%	72.3%	80.1%	89.2%	90.0%	87.7%	83.9%	77.6%	75.9%	57.3%	68.4%	74.5%
ATXN3 - CDS C.porcellus	72.8%	72.8%	74.7%	74.7%	75.0%	72.3%	80.4%	81.4%	78.9%	71.4%	64.9%	73.8%	75.3%	71.6%	72.2%	75.6%	78.9%	79.7%	79.3%	84.2%	73.4%	75.0%	57.7%	66.4%	70.2%
ATXN3 - CDS O. cuniculus	73.5%	73.7%	75.6%	75.9%	76.0%	73.0%	80.1%	81.5%	78.3%	72.8%	72.6%	74.2%	76.0%	72.6%	68.8%	80.5%	79.9%	80.6%	81.4%	83.7%	75.0%	76.1%	57.8%	68.9%	70.8%
ATXN3 - CDS C. familiaris	73.1%	73.2%	75.1%	75.4%	75.3%	73.2%	80.1%	80.9%	78.7%	72.5%	66.5%	73.1%	75.0%	70.8%	68.2%	75.0%	84.0%	86.3%	79.4%	84.7%	76.5%	77.9%	60.6%	68.5%	69.4%
ATXN3 - CDS C. haffmanni	80.1%	80.2%	82.8%	80.9%	82.0%	78.9%	87.6%	89.1%	83.6%	78.7%	70.5%	78.3%	82.3%	75.8%	73.9%	80.9%	87.2%	88.7%	94.9%	90.4%	76.7%	75.7%	57.3%	69.8%	74.0%
ATXN3 - CDS M. domestica	74.4%	74.8%	75.1%	74.1%	74.4%	72.0%	82.7%	83.4%	77.9%	76.9%	69.5%	72.9%	73.9%	68.2%	66.1%	75.8%	80.6%	81.8%	79.9%	78.4%	81.5%	84.3%	64.6%	92.3%	78.9%
ATXN3 - CDS B. taurus	74.1%	74.0%	76.3%	75.7%	76.8%	74.7%	81.4%	82.0%	79.1%	72.8%	65.5%	73.3%	76.4%	71.9%	70.0%	76.6%	82.6%	83.6%	81.1%	83.4%	79.8%	82.2%	60.4%	66.3%	71.2%
ATXN3 - CDS O. garnettii	65.2%	65.2%	66.7%	64.4%	65.9%	63.9%	75.3%	72.4%	65.9%	67.2%	58.5%	67.4%	63.6%	58.6%	62.9%	65.3%	70.9%	71.5%	68.8%	65.1%	63.8%	65.6%	48.0%	60.4%	58.8%
ATXN3L1 - H. sapiens	69.8%	69.9%	72.2%	70.4%	71.9%	68.8%	76.8%	76.7%	72.4%	73.6%	59.4%	69.2%	70.4%	64.3%	64.2%	69.6%	74.7%	75.2%	75.0%	75.7%	68.1%	65.8%	51.7%	62.0%	66.2%
ATXN3L1 - P. troglodytes	67.6%	67.7%	70.0%	68.2%	69.8%	66.8%	74.9%	74.7%	70.7%	71.6%	57.2%	67.3%	68.4%	62.5%	62.9%	68.0%	72.5%	73.2%	72.5%	73.5%	66.3%	64.2%	50.4%	60.6%	64.5%
ATXN3L1 - G. gorilla	70.1%	70.2%	72.6%	70.7%	72.3%	69.5%	77.2%	77.2%	72.9%	74.3%	59.6%	69.4%	70.7%	64.5%	64.6%	70.2%	74.9%	75.7%	75.5%	76.3%	68.6%	66.5%	51.6%	62.3%	66.3%
ATXN3L1 - P. abelii	69.1%	69.1%	71.5%	69.7%	71.2%	68.2%	76.0%	76.4%	72.0%	73.5%	58.7%	68.8%	70.2%	63.7%	63.9%	69.5%	73.8%	74.4%	74.5%	75.3%	67.5%	65.4%	51.3%	61.2%	66.1%
ATXN3L1 - N. leucogenys	69.6%	69.7%	72.0%	70.1%	71.7%	69.1%	76.0%	76.3%	72.1%	73.3%	58.8%	68.4%	70.2%	63.8%	64.0%	69.1%	74.7%	74.9%	74.4%	74.7%	67.8%	65.8%	52.2%	61.4%	65.4%
ATXN3L1 - C. jacchus	66.1%	66.0%	68.4%	67.2%	68.0%	65.8%	73.2%	73.3%	69.4%	70.1%	57.8%	65.5%	66.8%	61.0%	63.2%	66.8%	71.4%	72.1%	71.2%	73.2%	66.7%	64.7%	50.2%	60.6%	64.2%
ATXN3L1 - M. mulatta	70.1%	70.0%	72.5%	70.7%	72.0%	69.7%	76.7%	76.6%	72.6%	73.4%	59.2%	69.3%	70.9%	64.4%	64.5%	69.9%	75.2%	75.9%	75.2%	74.9%	69.5%	67.1%	52.1%	61.8%	66.5%
ATXN3L2 - H. sapiens		97.7%	94.7%	91.9%	92.3%	88.0%	78.9%	81.1%	74.9%	71.8%	63.5%	69.7%	71.3%	65.8%	64.4%	72.8%	78.3%	78.8%	76.8%	73.8%	66.8%	65.2%	51.4%	61.2%	65.3%
ATXN3L2 - P. troglodytes	97.7%		94.9%	92.2%	92.5%	88.4%	78.8%	81.2%	74.7%	71.8%	63.5%	69.8%	71.2%	65.6%	64.7%	72.9%	78.3%	78.8%	76.9%	73.9%	66.7%	65.1%	51.2%	61.1%	65.3%
ATXN3L2 - G. gorilla	94.7%	94.9%		94.8%	95.5%	91.2%	80.6%	81.0%	77.3%	72.3%	63.6%	71.7%	72.9%	67.3%	66.8%	75.0%	79.5%	79.6%	79.4%	73.6%	68.6%	67.0%	52.6%	63.0%	67.6%
ATXN3L2 - P. abelii	91.9%	92.2%	94.8%		94.8%	90.3%	78.7%	79.1%	77.5%	70.9%	64.6%	72.1%	72.0%	67.9%	66.6%	75.1%	77.4%	77.5%	77.4%	73.5%	69.0%	67.2%	52.3%	63.5%	66.7%
ATXN3L2 - N. leucogenys	92.3%	92.5%	95.5%	94.8%		91.3%	79.3%	80.2%	77.5%	71.9%	63.5%	72.1%	72.9%	68.3%	67.1%	75.4%	78.5%	79.0%	78.6%	74.3%	69.4%	67.8%	52.4%	63.2%	68.1%
ATXN3L2 - M. mulatta	88.0%	88.4%	91.2%	90.3%	91.3%		76.6%	77.6%	75.6%	69.1%	62.5%	69.4%	70.4%	65.5%	65.1%	72.8%	76.7%	75.9%	75.9%	71.8%	67.5%	65.4%	52.1%	60.9%	65.6%
ATXN3L?2 - C. jacchus	78.9%	78.8%	80.6%	78.7%	79.3%	76.6%		89.4%	85.8%	78.1%	67.8%	75.7%	77.7%	71.6%	71.9%	77.3%	84.6%	85.3%	84.0%	81.7%	74.5%	72.9%	55.9%	68.5%	72.1%
ATXN3L? - M. mulatta	81.1%	81.2%	81.0%	79.1%	80.2%	77.6%	89.4%		84.9%	77.8%	70.1%	76.0%	78.4%	72.8%	71.2%	78.5%	86.1%	87.0%	85.2%	82.2%	75.0%	73.4%	56.5%	66.8%	72.3%
ATXN3L?1 - C. jacchus	74.9%	74.7%	77.3%	77.5%	77.5%	75.6%	85.8%	84.9%		72.3%	64.7%	73.9%	74.8%	69.8%	68.6%	76.1%	80.0%	80.9%	80.0%	77.2%	72.9%	71.2%	56.6%	65.3%	70.3%
ATXN3L?1 - T. syrichta	71.8%	71.8%	72.3%	70.9%	71.9%	69.1%	78.1%	77.8%	72.3%		62.5%	69.8%	70.6%	64.1%	63.7%	70.8%	76.4%	77.5%	75.0%	75.8%	68.1%	66.4%	51.9%	63.1%	65.2%
ATXN3L?2 - T. syrichta	63.5%	63.5%	63.6%	64.6%	63.5%	62.5%	67.8%	70.1%	64.7%	62.5%		62.4%	64.1%	66.3%	59.6%	64.6%	68.2%	67.8%	69.2%	72.5%	64.4%	62.4%	43.9%	56.4%	56.8%
ATXN3L? - O. garnettii	69.7%	69.8%	71.7%	72.1%	72.1%	69.4%	75.7%	76.0%	73.9%	69.8%	62.4%		70.6%	67.3%	65.6%	71.6%	73.9%	74.7%	75.4%	72.7%	68.2%	66.9%	52.3%	61.4%	63.3%
ATXN3L?2 - P. vampyrus	71.3%	71.2%	72.9%	72.0%	72.9%	70.4%	77.7%	78.4%	74.8%	70.6%	64.1%	70.6%		78.4%	67.6%	73.4%	78.9%	79.2%	80.9%	79.7%	69.4%	67.9%	54.4%	60.7%	66.4%
ATXN3L?1 - P. vampyrus	65.8%	65.6%	67.3%	67.9%	68.3%	65.5%	71.6%	72.8%	69.8%	64.1%	66.3%	67.3%	78.4%		61.2%	70.0%	73.4%	73.0%	81.1%	71.5%	66.9%	64.8%	50.5%	58.4%	62.2%
ATXN3L? - C. porcellus	64.4%	64.7%	66.8%	66.6%	67.1%	65.1%	71.9%	71.2%	68.6%	63.7%	59.6%	65.6%	67.6%	61.2%		67.2%	71.5%	70.8%	70.5%	65.9%	63.3%	61.3%	47.1%	56.6%	60.1%
ATXN3L? - O. cuniculus	72.8%	72.9%	75.0%	75.1%	75.4%	72.8%	77.3%	78.5%	76.1%	70.8%	64.6%	71.6%	73.4%	70.0%	67.2%		77.5%	78.0%	77.5%	76.5%	70.4%	69.4%	54.2%	64.5%	67.1%
ATXN3L?1 - C. familiarirs	78.3%	78.3%	79.5%	77.4%	78.5%	76.7%	84.6%	86.1%	80.0%	76.4%	68.2%	73.9%	78.9%	73.4%	71.5%	77.5%		90.3%	84.1%	80.3%	75.3%	73.1%	57.4%	66.0%	72.6%
ATXN3L?2 - C. familiaris	78.8%	78.8%	79.6%	77.5%	79.0%	75.9%	85.3%	87.0%	80.9%	77.5%	67.8%	74.7%	79.2%	73.0%	70.8%	78.0%	90.3%		84.1%	83.5%	76.7%	74.3%	58.0%	66.5%	72.8%
ATXN3L?1 - C. haffmanni	76.8%	76.9%	79.4%	77.4%	78.6%	75.9%	84.0%	85.2%	80.0%	75.0%	69.2%	75.4%	80.9%	81.1%	70.5%	77.5%	84.1%	84.1%		91.7%	71.8%	70.8%	53.2%	64.7%	70.3%
ATXN3L?2 - C. haffmanni	73.8%	73.9%	73.6%	73.5%	74.3%	71.8%	81.7%	82.2%	77.2%	75.8%	72.5%	72.7%	79.7%	71.5%	65.9%	76.5%	80.3%	83.5%	91.7%		82.2%	83.3%	64.7%	75.1%	72.4%
ATXN3L?1 - B. taurus	66.8%	66.7%	68.6%	69.0%	69.4%	67.5%	74.5%	75.0%	72.9%	68.1%	64.4%	68.2%	69.4%	66.9%	63.3%	70.4%	75.3%	76.7%	71.8%	82.2%		82.7%	62.1%	70.0%	67.0%
ATXN3L?2 - B. taurus	65.2%	65.1%	67.0%	67.2%	67.8%	65.4%	72.9%	73.4%	71.2%	66.4%	62.4%	66.9%	67.9%	64.8%	61.3%	69.4%	73.1%	74.3%	70.8%	83.3%	82.7%		62.2%	70.1%	64.6%
ATXN3L?3 - B. taurus	51.4%	51.2%	52.6%	52.3%	52.4%	52.1%	55.9%	56.5%	56.6%	51.9%	43.9%	52.3%	54.4%	50.5%	47.1%	54.2%	57.4%	58.0%	53.2%	64.7%	62.1%	62.2%		54.9%	52.5%
ATXN3L? - M. domestica	61.2%	61.1%	63.0%	63.5%	63.2%	60.9%	68.5%	66.8%	65.3%	63.1%	56.4%	61.4%	60.7%	58.4%	56.6%	64.5%	66.0%	66.5%	64.7%	75.1%	70.0%	70.1%	54.9%		65.7%
ATXN3 - CDS G. gallus	65.3%	65.3%	67.6%	66.7%	68.1%	65.6%	72.1%	72.3%	70.3%	65.2%	56.8%	63.3%	66.4%	62.2%	60.1%	67.1%	72.6%	72.8%	70.3%	72.4%	67.0%	64.6%	52.5%	65.7%	

Table A5 - Matrix of genetic distances among *ATXN3*, *ATXN3L1* and *ATXN3L2* of several mammals, without the $(CAG)_n$ tract. Cells with darker background represent higher similarity between two sequences, whereas lighter cells show lower similarity. Notably, a lighter gray means further apart while darker means closer together. The longest distance in the tree is white and the shortest distance is black (may not be zero). Continue on the next page.

	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...
ATXN3 - CDS H. sapiens		94.9%	64.8%	86.6%	88.5%	93.7%	94.3%	60.4%	89.1%	80.7%	81.2%	80.7%	88.9%	84.4%	81.8%	72.0%	76.4%	74.3%	76.8%	75.9%	76.0%	72.7%	76.3%
ATXN3 - CDS P. troglodytes	94.9%		66.3%	90.1%	92.5%	97.7%	96.2%	62.1%	93.0%	83.9%	84.6%	83.6%	92.7%	84.8%	84.9%	75.3%	79.8%	77.6%	80.2%	79.2%	79.2%	75.7%	79.6%
ATXN3 - CDS G.gorilla	64.8%	66.3%		71.5%	60.9%	65.3%	64.6%	40.0%	63.0%	63.8%	63.9%	66.2%	62.9%	70.3%	63.2%	48.7%	54.0%	52.2%	54.2%	53.6%	53.6%	51.5%	53.6%
ATXN3 - CDS P. abelii	86.6%	90.1%	71.5%		84.2%	89.4%	87.5%	55.4%	85.2%	76.2%	76.7%	75.9%	84.0%	75.2%	77.1%	67.1%	71.8%	69.7%	72.3%	71.4%	71.4%	68.1%	71.8%
ATXN3 - CDS N. leucogenys	88.5%	92.5%	60.9%	84.2%		92.4%	89.7%	61.8%	87.4%	78.9%	78.8%	78.9%	87.5%	79.0%	80.2%	69.7%	74.9%	73.2%	75.4%	74.7%	74.5%	72.0%	75.1%
ATXN3 - CDS M. mulatta	93.7%	97.7%	65.3%	89.4%	92.4%		95.0%	61.2%	92.4%	84.2%	84.5%	83.4%	92.5%	84.8%	84.7%	74.4%	79.5%	77.3%	80.0%	78.9%	78.9%	75.4%	79.3%
ATXN3 - CDS C. jacchus	94.3%	96.2%	64.6%	87.5%	89.7%	95.0%		61.0%	90.7%	82.9%	83.1%	82.4%	90.2%	85.3%	83.8%	74.3%	78.2%	76.4%	78.7%	77.7%	77.7%	74.4%	78.3%
ATXN3 - CDS T. syrichta	60.4%	62.1%	40.0%	55.4%	61.8%	61.2%	61.0%		62.0%	56.0%	58.7%	56.1%	60.7%	57.6%	57.5%	44.1%	54.0%	52.6%	53.9%	53.9%	54.1%	53.4%	54.3%
ATXN3 - CDS P.vampyrus	89.1%	93.0%	63.0%	85.2%	87.4%	92.4%	90.7%	62.0%		82.6%	83.6%	83.8%	91.1%	84.2%	85.6%	73.9%	78.7%	76.4%	78.8%	78.1%	78.3%	75.3%	79.3%
ATXN3 - CDS C.porcellus	80.7%	83.9%	63.8%	76.2%	78.9%	84.2%	82.9%	56.0%	82.6%		83.3%	85.7%	82.8%	83.4%	83.6%	69.2%	71.8%	69.9%	72.1%	71.0%	71.3%	68.5%	72.1%
ATXN3 - CDS O. cuniculus	81.2%	84.6%	63.9%	76.7%	78.8%	84.5%	83.1%	58.7%	83.6%	83.3%		84.5%	83.9%	82.7%	82.2%	69.3%	73.3%	71.0%	73.5%	72.6%	72.5%	69.8%	73.3%
ATXN3 - CDS C. familiaris	80.7%	83.6%	66.2%	75.9%	78.9%	83.4%	82.4%	56.1%	83.8%	85.7%	84.5%		84.2%	85.2%	85.6%	69.8%	71.3%	69.3%	71.7%	70.6%	70.9%	68.6%	71.6%
ATXN3 - CDS C. haffmanni	88.9%	92.7%	62.9%	84.0%	87.5%	92.5%	90.2%	60.7%	91.1%	82.8%	83.9%	84.2%		84.4%	84.9%	73.4%	78.6%	76.4%	79.1%	78.2%	78.1%	75.1%	79.1%
ATXN3 - CDS M. domestica	84.4%	84.8%	70.3%	75.2%	79.0%	84.8%	85.3%	57.6%	84.2%	83.4%	82.7%	85.2%	84.4%		84.1%	69.1%	74.7%	72.6%	74.8%	73.8%	73.7%	72.5%	74.7%
ATXN3 - CDS B. taurus	81.8%	84.9%	63.2%	77.1%	80.2%	84.7%	83.8%	57.5%	85.6%	83.6%	82.2%	85.6%	84.9%	84.1%		71.0%	72.9%	71.1%	73.5%	72.3%	72.9%	70.9%	74.4%
ATXN3 - CDS O. garnettii	72.0%	75.3%	48.7%	67.1%	69.7%	74.4%	74.3%	44.1%	73.9%	69.2%	69.3%	69.8%	73.4%	69.1%	71.0%		65.6%	63.8%	66.2%	64.7%	65.3%	62.0%	66.0%
ATXN3L1 - H. sapiens	76.4%	79.8%	54.0%	71.8%	74.9%	79.5%	78.2%	54.0%	78.7%	71.8%	73.3%	71.3%	78.6%	74.7%	72.9%	65.6%		96.1%	98.4%	97.2%	95.9%	82.8%	93.0%
ATXN3L1 - P. troglodytes	74.3%	77.6%	52.2%	69.7%	73.2%	77.3%	76.4%	52.6%	76.4%	69.9%	71.0%	69.3%	76.4%	72.6%	71.1%	63.8%	96.1%		96.2%	94.9%	93.9%	80.8%	91.2%
ATXN3L1 - G. gorilla	76.8%	80.2%	54.2%	72.3%	75.4%	80.0%	78.7%	53.9%	78.8%	72.1%	73.5%	71.7%	79.1%	74.8%	73.5%	66.2%	98.4%	96.2%		97.7%	96.5%	83.4%	93.8%
ATXN3L1 - P. abelii	75.9%	79.2%	53.6%	71.4%	74.7%	78.9%	77.7%	53.9%	78.1%	71.0%	72.6%	70.6%	78.2%	73.8%	72.3%	64.7%	97.2%	94.9%	97.7%		95.7%	82.7%	93.1%
ATXN3L1 - N. leucogenys	76.0%	79.2%	53.6%	71.4%	74.5%	78.9%	77.7%	54.1%	78.3%	71.3%	72.5%	70.9%	78.1%	73.7%	72.9%	65.3%	95.9%	93.9%	96.5%	95.7%		83.5%	93.1%
ATXN3L1 - C. jacchus	72.7%	75.7%	51.5%	68.1%	72.0%	75.4%	74.4%	53.4%	75.3%	68.5%	69.8%	68.6%	75.1%	72.5%	70.9%	62.0%	82.8%	80.8%	83.4%	82.7%	83.5%		83.3%
ATXN3L1 - M. mulatta	76.3%	79.6%	53.6%	71.8%	75.1%	79.3%	78.3%	54.3%	79.3%	72.1%	73.3%	71.6%	79.1%	74.7%	74.4%	66.0%	93.0%	91.2%	93.8%	93.1%	93.1%	83.3%	
ATXN3L2 - H. sapiens	81.5%	82.6%	55.7%	74.9%	78.2%	82.0%	83.0%	55.3%	80.8%	72.8%	73.5%	73.1%	80.1%	74.4%	74.1%	65.2%	69.8%	67.6%	70.1%	69.1%	69.6%	66.1%	70.1%
ATXN3L2 - P. troglodytes	82.3%	82.6%	55.8%	74.9%	78.2%	82.0%	83.0%	54.6%	80.7%	72.8%	73.7%	73.2%	80.2%	74.8%	74.0%	65.2%	69.9%	67.7%	70.2%	69.1%	69.7%	66.0%	70.0%
ATXN3L2 - G. gorilla	80.2%	83.6%	56.8%	76.0%	79.5%	83.5%	82.1%	55.0%	82.0%	74.7%	75.6%	75.1%	82.8%	75.1%	76.3%	66.7%	72.2%	70.0%	72.6%	71.5%	72.0%	68.4%	72.5%
ATXN3L2 - P. abelii	78.4%	81.8%	57.1%	74.1%	77.8%	81.4%	80.1%	54.1%	80.0%	74.7%	75.9%	75.4%	80.9%	74.1%	75.7%	64.4%	70.4%	68.2%	70.7%	69.7%	70.1%	67.2%	70.7%
ATXN3L2 - N. leucogenys	79.4%	82.7%	57.1%	75.1%	78.4%	82.6%	81.2%	54.3%	80.9%	75.0%	76.0%	75.3%	82.0%	74.4%	76.8%	65.9%	71.9%	69.8%	72.3%	71.2%	71.7%	68.0%	72.0%
ATXN3L2 - M. mulatta	76.4%	79.5%	54.9%	72.2%	76.0%	79.7%	78.2%	53.2%	78.2%	72.3%	73.0%	73.2%	78.9%	72.0%	74.7%	63.9%	68.8%	66.8%	69.5%	68.2%	69.1%	65.8%	69.7%
ATXN3L?2 - C. jacchus	88.9%	92.9%	61.8%	84.5%	87.4%	92.5%	93.0%	58.7%	87.9%	80.4%	80.1%	80.1%	87.6%	82.7%	81.4%	75.3%	76.8%	74.9%	77.2%	76.0%	76.0%	73.2%	76.7%
ATXN3L? - M. mulatta	93.0%	95.1%	63.4%	87.3%	89.1%	95.3%	93.6%	60.8%	89.5%	81.4%	81.5%	80.9%	89.1%	83.4%	82.0%	72.4%	76.7%	74.7%	77.2%	76.4%	76.3%	73.3%	76.6%
ATXN3L?1 - C. jacchus	83.9%	87.5%	60.4%	79.2%	82.2%	87.0%	87.4%	57.2%	82.6%	78.9%	78.3%	78.7%	83.6%	77.9%	79.1%	65.9%	72.4%	70.7%	72.9%	72.0%	72.1%	69.4%	72.6%
ATXN3L?1 - T. syrichta	77.4%	80.6%	54.4%	72.4%	76.1%	79.9%	79.4%	55.6%	79.5%	71.4%	72.8%	72.5%	78.7%	76.9%	72.8%	67.2%	73.6%	71.6%	74.3%	73.5%	73.3%	70.1%	73.4%
ATXN3L?2 - T. syrichta	69.3%	71.6%	60.1%	71.8%	64.9%	72.0%	70.5%	51.1%	71.8%	64.9%	72.6%	66.5%	70.5%	69.5%	65.5%	58.5%	59.4%	57.2%	59.6%	58.7%	58.8%	57.8%	59.2%
ATXN3L? - O. garnettii	75.2%	78.6%	55.6%	71.6%	74.9%	78.6%	77.2%	51.6%	76.8%	73.8%	74.2%	73.1%	78.3%	72.9%	73.3%	67.4%	69.2%	67.3%	69.4%	68.8%	68.4%	65.5%	69.3%
ATXN3L?2 - P. vampyrus	78.2%	81.8%	63.2%	81.5%	77.1%	81.7%	79.7%	52.8%	83.9%	75.3%	76.0%	75.0%	82.3%	73.9%	76.4%	63.6%	70.4%	68.4%	70.7%	70.2%	70.2%	66.8%	70.9%
ATXN3L?1 - P. vampyrus	72.2%	75.6%	58.8%	73.9%	71.5%	75.9%	74.1%	48.9%	77.3%	71.6%	72.6%	70.8%	75.8%	68.2%	71.9%	58.6%	64.3%	62.5%	64.5%	63.7%	63.8%	61.0%	64.4%
ATXN3L? - C. porcellus	70.8%	73.6%	51.3%	67.3%	69.2%	74.0%	73.0%	46.8%	72.3%	72.2%	68.8%	68.2%	73.9%	66.1%	70.0%	62.9%	64.2%	62.9%	64.6%	63.9%	64.0%	63.2%	64.5%
ATXN3L? - O. cuniculus	77.5%	80.8%	56.4%	73.8%	76.8%	81.0%	79.6%	54.1%	80.1%	75.6%	80.5%	75.0%	80.9%	75.8%	76.6%	65.3%	69.6%	68.0%	70.2%	69.5%	69.1%	66.8%	69.9%
ATXN3L?1 - C. familiarirs	85.1%	88.8%	60.5%	81.2%	83.4%	88.2%	87.1%	58.8%	89.2%	78.9%	79.9%	84.0%	87.2%	80.6%	82.6%	70.9%	74.7%	72.5%	74.9%	73.8%	74.7%	71.4%	75.2%
ATXN3L?2 - C. familiaris	86.5%	89.5%	60.3%	81.2%	84.1%	88.9%	88.4%	60.2%	90.0%	79.7%	80.6%	86.3%	88.7%	81.8%	83.6%	71.5%	75.2%	73.2%	75.7%	74.4%	74.9%	72.1%	75.9%
ATXN3L?1 - C. haffmanni	85.3%	89.5%	69.1%	89.0%	83.0%	89.1%	86.9%	55.8%	87.7%	79.3%	81.4%	79.4%	94.9%	79.9%	81.1%	68.8%	75.0%	72.5%	75.5%	74.5%	74.4%	71.2%	75.2%
ATXN3L?2 - C. haffmanni	84.9%	84.9%	70.1%	77.0%	78.1%	86.8%	84.5%	58.4%	83.9%	84.2%	83.7%	84.7%	90.4%	78.4%	83.4%	65.1%	75.7%	73.5%	76.3%	75.3%	74.7%	73.2%	74.9%
ATXN3L?1 - B. taurus	75.1%	77.2%	58.1%	69.4%	72.2%	77.1%	76.2%	52.0%	77.6%	73.4%	75.0%	76.5%	76.7%	81.5%	79.8%	63.8%	68.1%	66.3%	68.6%	67.5%	67.8%	66.7%	69.5%
ATXN3L?2 - B. taurus	73.8%	75.9%	59.6%	68.1%	71.1%	75.5%	74.6%	51.2%	75.9%	75.0%	76.1%	77.9%	75.7%	84.3%	82.2%	65.6%	65.8%	64.2%	66.5%	65.4%	65.8%	64.7%	67.1%
ATXN3L?3 - B. taurus	56.8%	57.8%	44.5%	51.1%	53.9%	57.3%	57.0%	45.3%	57.3%	57.7%	57.8%	60.6%	57.3%	64.6%	60.4%	48.0%	51.7%	50.4%	51.6%	51.3%	52.2%	50.2%	52.1%
ATXN3L? - M. domestica	67.4%	69.4%	52.4%	61.8%	64.9%	69.3%	68.7%	48.1%	68.4%	66.4%	68.9%	68.5%	69.8%	92.3%	66.3%	60.4%	62.0%	60.6%	62.3%	61.2%	61.4%	60.6%	61.8%
ATXN3 - CDS G. gallus	72.3%	74.9%	53.8%	67.9%	69.5%	74.8%	74.2%	53.9%	74.5%	70.2%	70.8%	69.4%	74.0%	78.9%	71.2%	58.8%	66.2%	64.5%	66.3%	66.1%	65.4%	64.2%	66.5%

Table A5 - Continuation

	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3 - ...
ATXN3 - CDS H. sapiens	81.5%	82.3%	80.2%	78.4%	79.4%	76.4%	88.9%	93.0%	83.9%	77.4%	69.3%	75.2%	78.2%	72.2%	70.8%	77.5%	85.1%	86.5%	85.3%	84.9%	75.1%	73.8%	56.8%	67.4%	72.3%
ATXN3 - CDS P. troglodytes	82.6%	82.6%	83.6%	81.8%	82.7%	79.5%	92.9%	95.1%	87.5%	80.6%	71.6%	78.6%	81.8%	75.6%	73.6%	80.8%	88.8%	89.5%	89.5%	84.9%	77.2%	75.9%	57.8%	69.4%	74.9%
ATXN3 - CDS G.gorilla	55.7%	55.8%	56.8%	57.1%	57.1%	54.9%	61.8%	63.4%	60.4%	54.4%	60.1%	55.6%	63.2%	58.8%	51.3%	56.4%	60.5%	60.3%	69.1%	70.1%	58.1%	59.6%	44.5%	52.4%	53.8%
ATXN3 - CDS P. abelii	74.9%	74.9%	76.0%	74.1%	75.1%	72.2%	84.5%	87.3%	79.2%	72.4%	71.8%	71.6%	81.5%	73.9%	67.3%	73.8%	81.2%	81.2%	89.0%	77.0%	69.4%	68.1%	51.1%	61.8%	67.9%
ATXN3 - CDS N. leucogenys	78.2%	78.2%	79.5%	77.8%	78.4%	76.0%	87.4%	89.1%	82.2%	76.1%	64.9%	74.9%	77.1%	71.5%	69.2%	76.8%	83.4%	84.1%	83.0%	78.1%	72.2%	71.1%	53.9%	64.9%	69.5%
ATXN3 - CDS M. mulatta	82.0%	82.0%	83.5%	81.4%	82.6%	79.7%	92.5%	95.3%	87.0%	79.9%	72.0%	78.6%	81.7%	75.9%	74.0%	81.0%	88.2%	88.9%	89.1%	86.8%	77.1%	75.5%	57.3%	69.3%	74.8%
ATXN3 - CDS C. jacchus	83.0%	83.0%	82.1%	80.1%	81.2%	78.2%	93.0%	93.6%	87.4%	79.4%	70.5%	77.2%	79.7%	74.1%	73.0%	79.6%	87.1%	88.4%	86.9%	84.5%	76.2%	74.6%	57.0%	68.7%	74.2%
ATXN3 - CDS T. syrichta	55.3%	54.6%	55.0%	54.1%	54.3%	53.2%	58.7%	60.8%	57.2%	55.6%	51.1%	51.6%	52.8%	48.9%	46.8%	54.1%	58.8%	60.2%	55.8%	58.4%	52.0%	51.2%	45.3%	48.1%	53.9%
ATXN3 - CDS P.vampyrus	80.8%	80.7%	82.0%	80.0%	80.9%	78.2%	87.9%	89.5%	82.6%	79.5%	71.8%	76.8%	83.9%	77.3%	72.3%	80.1%	89.2%	90.0%	87.7%	83.9%	77.6%	75.9%	57.3%	68.4%	74.5%
ATXN3 - CDS C.porcellus	72.8%	72.8%	74.7%	74.7%	75.0%	72.3%	80.4%	81.4%	78.9%	71.4%	64.9%	73.8%	75.3%	71.6%	72.2%	75.6%	78.9%	79.7%	79.3%	84.2%	73.4%	75.0%	57.7%	66.4%	70.2%
ATXN3 - CDS O. cuniculus	73.5%	73.7%	75.6%	75.9%	76.0%	73.0%	80.1%	81.5%	78.3%	72.8%	72.6%	74.2%	76.0%	72.6%	68.8%	80.5%	79.9%	80.6%	81.4%	83.7%	75.0%	76.1%	57.8%	68.9%	70.8%
ATXN3 - CDS C. familiaris	73.1%	73.2%	75.1%	75.4%	75.3%	73.2%	80.1%	80.9%	78.7%	72.5%	66.5%	73.1%	75.0%	70.8%	68.2%	75.0%	84.0%	86.3%	79.4%	84.7%	76.5%	77.9%	60.6%	68.5%	69.4%
ATXN3 - CDS C. haffmanni	80.1%	80.2%	82.8%	80.9%	82.0%	78.9%	87.6%	89.1%	83.6%	78.7%	70.5%	78.3%	82.3%	75.8%	73.9%	80.9%	87.2%	88.7%	94.9%	90.4%	76.7%	75.7%	57.3%	69.8%	74.0%
ATXN3 - CDS M. domestica	74.4%	74.8%	75.1%	74.1%	74.4%	72.0%	82.7%	83.4%	77.9%	76.9%	69.5%	72.9%	73.9%	68.2%	66.1%	75.8%	80.6%	81.8%	79.9%	78.4%	81.5%	84.3%	64.6%	92.3%	78.9%
ATXN3 - CDS B. taurus	74.1%	74.0%	76.3%	75.7%	76.8%	74.7%	81.4%	82.0%	79.1%	72.8%	65.5%	73.3%	76.4%	71.9%	70.0%	76.6%	82.6%	83.6%	81.1%	83.4%	79.8%	82.2%	60.4%	66.3%	71.2%
ATXN3 - CDS O. garnettii	65.2%	65.2%	66.7%	64.4%	65.9%	63.9%	75.3%	72.4%	65.9%	67.2%	58.5%	67.4%	63.6%	58.6%	62.9%	65.3%	70.9%	71.5%	68.8%	65.1%	63.8%	65.6%	48.0%	60.4%	58.8%
ATXN3L1 - H. sapiens	69.8%	69.9%	72.2%	70.4%	71.9%	68.8%	76.8%	76.7%	72.4%	73.6%	59.4%	69.2%	70.4%	64.3%	64.2%	69.6%	74.7%	75.2%	75.0%	75.7%	68.1%	65.8%	51.7%	62.0%	66.2%
ATXN3L1 - P. troglodytes	67.6%	67.7%	70.0%	68.2%	69.8%	66.8%	74.9%	74.7%	70.7%	71.6%	57.2%	67.3%	68.4%	62.5%	62.9%	68.0%	72.5%	73.2%	72.5%	73.5%	66.3%	64.2%	50.4%	60.6%	64.5%
ATXN3L1 - G. gorilla	70.1%	70.2%	72.6%	70.7%	72.3%	69.5%	77.2%	77.2%	72.9%	74.3%	59.6%	69.4%	70.7%	64.5%	64.6%	70.2%	74.9%	75.7%	75.5%	76.3%	68.6%	66.5%	51.6%	62.3%	66.3%
ATXN3L1 - P. abelii	69.1%	69.1%	71.5%	69.7%	71.2%	68.2%	76.0%	76.4%	72.0%	73.5%	58.7%	68.8%	70.2%	63.7%	63.9%	69.5%	73.8%	74.4%	74.5%	75.3%	67.5%	65.4%	51.3%	61.2%	66.1%
ATXN3L1 - N. leucogenys	69.6%	69.7%	72.0%	70.1%	71.7%	69.1%	76.0%	76.3%	72.1%	73.3%	58.8%	68.4%	70.2%	63.8%	64.0%	69.1%	74.7%	74.9%	74.4%	74.7%	67.8%	65.8%	52.2%	61.4%	65.4%
ATXN3L1 - C. jacchus	66.1%	66.0%	68.4%	67.2%	68.0%	65.8%	73.2%	73.3%	69.4%	70.1%	57.8%	65.5%	66.8%	61.0%	63.2%	66.8%	71.4%	72.1%	71.2%	73.2%	66.7%	64.7%	50.2%	60.6%	64.2%
ATXN3L1 - M. mulatta	70.1%	70.0%	72.5%	70.7%	72.0%	69.7%	76.7%	76.6%	72.6%	73.4%	59.2%	69.3%	70.9%	64.4%	64.5%	69.9%	75.2%	75.9%	75.2%	74.9%	69.5%	67.1%	52.1%	61.8%	66.5%
ATXN3L2 - H. sapiens		97.7%	94.7%	91.9%	92.3%	88.0%	78.9%	81.1%	74.9%	71.8%	63.5%	69.7%	71.3%	65.8%	64.4%	72.8%	78.3%	78.8%	76.8%	73.8%	66.8%	65.2%	51.4%	61.2%	65.3%
ATXN3L2 - P. troglodytes	97.7%		94.9%	92.2%	92.5%	88.4%	78.8%	81.2%	74.7%	71.8%	63.5%	69.8%	71.2%	65.6%	64.7%	72.9%	78.3%	78.8%	76.9%	73.9%	66.7%	65.1%	51.2%	61.1%	65.3%
ATXN3L2 - G. gorilla	94.7%	94.9%		94.8%	95.5%	91.2%	80.6%	81.0%	77.3%	72.3%	63.6%	71.7%	72.9%	67.3%	66.8%	75.0%	79.5%	79.6%	79.4%	73.6%	68.6%	67.0%	52.6%	63.0%	67.6%
ATXN3L2 - P. abelii	91.9%	92.2%	94.8%		94.8%	90.3%	78.7%	79.1%	77.5%	70.9%	64.6%	72.1%	72.0%	67.9%	66.6%	75.1%	77.4%	77.5%	77.4%	73.5%	69.0%	67.2%	52.3%	63.5%	66.7%
ATXN3L2 - N. leucogenys	92.3%	92.5%	95.5%	94.8%		91.3%	79.3%	80.2%	77.5%	71.9%	63.5%	72.1%	72.9%	68.3%	67.1%	75.4%	78.5%	79.0%	78.6%	74.3%	69.4%	67.8%	52.4%	63.2%	68.1%
ATXN3L2 - M. mulatta	88.0%	88.4%	91.2%	90.3%	91.3%		76.6%	77.6%	75.6%	69.1%	62.5%	69.4%	70.4%	65.5%	65.1%	72.8%	76.7%	75.9%	75.9%	71.8%	67.5%	65.4%	52.1%	60.9%	65.6%
ATXN3L?2 - C. jacchus	78.9%	78.8%	80.6%	78.7%	79.3%	76.6%		89.4%	85.8%	78.1%	67.8%	75.7%	77.7%	71.6%	71.9%	77.3%	84.6%	85.3%	84.0%	81.7%	74.5%	72.9%	55.9%	68.5%	72.1%
ATXN3L? - M. mulatta	81.1%	81.2%	81.0%	79.1%	80.2%	77.6%	89.4%		84.9%	77.8%	70.1%	76.0%	78.4%	72.8%	71.2%	78.5%	86.1%	87.0%	85.2%	82.2%	75.0%	73.4%	56.5%	66.8%	72.3%
ATXN3L?1 - C. jacchus	74.9%	74.7%	77.3%	77.5%	77.5%	75.6%	85.8%	84.9%		72.3%	64.7%	73.9%	74.8%	69.8%	68.6%	76.1%	80.0%	80.9%	80.0%	77.2%	72.9%	71.2%	56.6%	65.3%	70.3%
ATXN3L?1 - T. syrichta	71.8%	71.8%	72.3%	70.9%	71.9%	69.1%	78.1%	77.8%	72.3%		62.5%	69.8%	70.6%	64.1%	63.7%	70.8%	76.4%	77.5%	75.0%	75.8%	68.1%	66.4%	51.9%	63.1%	65.2%
ATXN3L?2 - T. syrichta	63.5%	63.5%	63.6%	64.6%	63.5%	62.5%	67.8%	70.1%	64.7%	62.5%		62.4%	64.1%	66.3%	59.6%	64.6%	68.2%	67.8%	69.2%	72.5%	64.4%	62.4%	43.9%	56.4%	56.8%
ATXN3L? - O. garnettii	69.7%	69.8%	71.7%	72.1%	72.1%	69.4%	75.7%	76.0%	73.9%	69.8%	62.4%		70.6%	67.3%	65.6%	71.6%	73.9%	74.7%	75.4%	72.7%	68.2%	66.9%	52.3%	61.4%	63.3%
ATXN3L?2 - P. vampyrus	71.3%	71.2%	72.9%	72.0%	72.9%	70.4%	77.7%	78.4%	74.8%	70.6%	64.1%	70.6%		78.4%	67.6%	73.4%	78.9%	79.2%	80.9%	79.7%	69.4%	67.9%	54.4%	60.7%	66.4%
ATXN3L?1 - P. vampyrus	65.8%	65.6%	67.3%	67.9%	68.3%	65.5%	71.6%	72.8%	69.8%	64.1%	66.3%	67.3%	78.4%		61.2%	70.0%	73.4%	73.0%	81.1%	71.5%	66.9%	64.8%	50.5%	58.4%	62.2%
ATXN3L? - C. porcellus	64.4%	64.7%	66.8%	66.6%	67.1%	65.1%	71.9%	71.2%	68.6%	63.7%	59.6%	65.6%	67.6%	61.2%		67.2%	71.5%	70.8%	70.5%	65.9%	63.3%	61.3%	47.1%	56.6%	60.1%
ATXN3L? - O. cuniculus	72.8%	72.9%	75.0%	75.1%	75.4%	72.8%	77.3%	78.5%	76.1%	70.8%	64.6%	71.6%	73.4%	70.0%	67.2%		77.5%	78.0%	77.5%	76.5%	70.4%	69.4%	54.2%	64.5%	67.1%
ATXN3L?1 - C. familiarirs	78.3%	78.3%	79.5%	77.4%	78.5%	76.7%	84.6%	86.1%	80.0%	76.4%	68.2%	73.9%	78.9%	73.4%	71.5%	77.5%		90.3%	84.1%	80.3%	75.3%	73.1%	57.4%	66.0%	72.6%
ATXN3L?2 - C. familiaris	78.8%	78.8%	79.6%	77.5%	79.0%	75.9%	85.3%	87.0%	80.9%	77.5%	67.8%	74.7%	79.2%	73.0%	70.8%	78.0%	90.3%		84.1%	83.5%	76.7%	74.3%	58.0%	66.5%	72.8%
ATXN3L?1 - C. haffmanni	76.8%	76.9%	79.4%	77.4%	78.6%	75.9%	84.0%	85.2%	80.0%	75.0%	69.2%	75.4%	80.9%	81.1%	70.5%	77.5%	84.1%	84.1%		91.7%	71.8%	70.8%	53.2%	64.7%	70.3%
ATXN3L?2 - C. haffmanni	73.8%	73.9%	73.6%	73.5%	74.3%	71.8%	81.7%	82.2%	77.2%	75.8%	72.5%	72.7%	79.7%	71.5%	65.9%	76.5%	80.3%	83.5%	91.7%		82.2%	83.3%	64.7%	75.1%	72.4%
ATXN3L?1 - B. taurus	66.8%	66.7%	68.6%	69.0%	69.4%	67.5%	74.5%	75.0%	72.9%	68.1%	64.4%	68.2%	69.4%	66.9%	63.3%	70.4%	75.3%	76.7%	71.8%	82.2%		82.7%	62.1%	70.0%	67.0%
ATXN3L?2 - B. taurus	65.2%	65.1%	67.0%	67.2%	67.8%	65.4%	72.9%	73.4%	71.2%	66.4%	62.4%	66.9%	67.9%	64.8%	61.3%	69.4%	73.1%	74.3%	70.8%	83.3%	82.7%		62.2%	70.1%	64.6%
ATXN3L?3 - B. taurus	51.4%	51.2%	52.6%	52.3%	52.4%	52.1%	55.9%	56.5%	56.6%	51.9%	43.9%	52.3%	54.4%	50.5%	47.1%	54.2%	57.4%	58.0%	53.2%	64.7%	62.1%	62.2%		54.9%	52.5%
ATXN3L? - M. domestica	61.2%	61.1%	63.0%	63.5%	63.2%	60.9%	68.5%	66.8%	65.3%	63.1%	56.4%	61.4%	60.7%	58.4%	56.6%	64.5%	66.0%	66.5%	64.7%	75.1%	70.0%	70.1%	54.9%		65.7%
ATXN3 - CDS G. gallus	65.3%	65.3%	67.6%	66.7%	68.1%	65.6%	72.1%	72.3%	70.3%	65.2%	56.8%	63.3%	66.4%	62.2%	60.1%	67.1%	72.6%	72.8%	70.3%	72.4%	67.0%	64.6%	52.5%	65.7%	

Table A6 - Matrix of genetic distances among *ATXN3*, *ATXN3L1* and *ATXN3L2* Josephin domain of several mammals. Cells with darker background represent higher similarity between two sequences, whereas lighter cells show lower similarity. Notably, a lighter gray means further apart while darker means closer together. The longest distance in the tree is white and the shortest distance is black (may not be zero). Continue on the next page.

	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3 - ...	ATXN3L?...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...
ATXN3 - CDS H. sapiens		99.4%	70.2%	82.1%	88.4%	98.4%	97.4%	77.1%	93.9%	94.0%	90.6%	94.5%	95.5%	92.1%	93.8%	66.2%	80.0%	86.0%	83.5%	86.5%	85.4%	85.8%	83.4%	85.6%
ATXN3 - CDS P. troglodytes	99.4%		70.4%	82.3%	88.5%	98.6%	98.0%	77.3%	94.5%	94.2%	91.1%	94.7%	95.7%	92.6%	93.9%	66.4%	80.2%	86.5%	84.1%	87.1%	86.0%	86.3%	83.9%	86.1%
ATXN3 - CDS G.gorilla	70.2%	70.4%		84.5%	59.1%	68.6%	69.3%	54.3%	67.8%	66.1%	65.3%	67.3%	67.6%	66.5%	66.5%	40.3%	57.2%	60.8%	58.2%	61.2%	60.3%	60.8%	57.9%	60.1%
ATXN3 - CDS P. abelii	82.1%	82.3%	84.5%		71.0%	80.6%	81.0%	63.6%	79.1%	77.6%	75.1%	78.3%	78.5%	76.3%	77.9%	50.8%	67.0%	71.2%	68.8%	71.7%	70.8%	71.2%	68.8%	70.8%
ATXN3 - CDS N. leucogenys	88.4%	88.5%	59.1%	71.0%		87.8%	86.9%	76.9%	84.5%	83.6%	80.2%	85.1%	85.5%	82.8%	84.4%	56.4%	73.1%	77.3%	75.6%	77.8%	77.4%	77.1%	76.5%	77.1%
ATXN3 - CDS M. mulatta	98.4%	98.6%	68.6%	80.6%	87.8%		96.9%	76.4%	94.2%	94.0%	91.1%	94.2%	95.5%	92.8%	93.3%	64.3%	79.8%	86.2%	83.7%	86.8%	85.7%	86.0%	83.3%	85.9%
ATXN3 - CDS C. jacchus	97.4%	98.0%	69.3%	81.0%	86.9%	96.9%		76.2%	93.9%	94.6%	90.8%	95.0%	94.8%	93.2%	94.1%	66.5%	80.2%	86.5%	84.7%	87.1%	86.0%	86.3%	83.9%	86.3%
ATXN3 - CDS T. syrichta	77.1%	77.3%	54.3%	63.6%	76.9%	76.4%	76.2%		76.7%	75.0%	76.5%	75.0%	75.8%	76.3%	76.3%	50.3%	64.4%	70.2%	68.8%	70.4%	70.1%	70.6%	70.2%	70.8%
ATXN3 - CDS P.vampyrus	93.9%	94.5%	67.8%	79.1%	84.5%	94.2%	93.9%	76.7%		92.1%	90.8%	94.3%	93.6%	92.8%	93.2%	65.8%	78.9%	85.8%	83.4%	86.0%	85.2%	86.0%	83.5%	86.1%
ATXN3 - CDS C.porcellus	94.0%	94.2%	66.1%	77.6%	83.6%	94.0%	94.6%	75.0%	92.1%		90.4%	92.3%	92.7%	90.2%	92.5%	64.9%	79.5%	84.4%	82.3%	85.0%	83.8%	84.2%	82.1%	85.2%
ATXN3 - CDS O. cuniculus	90.6%	91.1%	65.3%	75.1%	80.2%	91.1%	90.8%	76.5%	90.8%	90.4%		90.6%	90.2%	89.6%	90.4%	63.6%	75.8%	85.0%	82.3%	85.0%	84.2%	84.6%	82.1%	85.0%
ATXN3 - CDS C. familiaris	94.5%	94.7%	67.3%	78.3%	85.1%	94.2%	95.0%	75.0%	94.3%	92.3%	90.6%		94.2%	91.7%	95.4%	66.4%	79.8%	85.3%	83.3%	85.8%	84.7%	85.5%	82.7%	85.5%
ATXN3 - CDS C. haffmanni	95.5%	95.7%	67.6%	78.5%	85.5%	95.5%	94.8%	75.8%	93.6%	92.7%	90.2%	94.2%		92.4%	93.6%	64.7%	80.0%	85.5%	82.9%	86.0%	84.9%	85.3%	83.1%	86.2%
ATXN3 - CDS M. domestica	92.1%	92.6%	66.5%	76.3%	82.8%	92.8%	93.2%	76.3%	92.8%	90.2%	89.6%	91.7%	92.4%		90.3%	64.5%	77.5%	85.4%	83.7%	85.4%	84.7%	84.8%	82.8%	85.0%
ATXN3 - CDS B. taurus	93.8%	93.9%	66.5%	77.9%	84.4%	93.3%	94.1%	76.3%	93.2%	92.5%	90.4%	95.4%	93.6%	90.3%		65.3%	78.3%	83.6%	81.8%	84.2%	83.1%	84.2%	81.6%	85.5%
ATXN3 - CDS O. garnettii	66.2%	66.4%	40.3%	50.8%	56.4%	64.3%	66.5%	50.3%	65.8%	64.9%	63.6%	66.4%	64.7%	64.5%	65.3%		60.7%	61.9%	60.6%	62.3%	61.4%	61.6%	60.6%	63.0%
ATXN3L? - O. garnettii - se...	80.0%	80.2%	57.2%	67.0%	73.1%	79.8%	80.2%	64.4%	78.9%	79.5%	75.8%	79.8%	80.0%	77.5%	78.3%	60.7%		73.1%	71.5%	73.7%	73.3%	73.1%	70.4%	74.0%
ATXN3L1 - H. sapiens	86.0%	86.5%	60.8%	71.2%	77.3%	86.2%	86.5%	70.2%	85.8%	84.4%	85.0%	85.3%	85.5%	85.4%	83.6%	61.9%	73.1%		96.9%	99.3%	98.9%	98.2%	86.1%	93.9%
ATXN3L1 - P. troglodytes	83.5%	84.1%	58.2%	68.8%	75.6%	83.7%	84.7%	68.8%	83.4%	82.3%	82.3%	83.3%	82.9%	83.7%	81.8%	60.6%	71.5%	96.9%		96.9%	96.5%	96.1%	83.7%	91.9%
ATXN3L1 - G. gorilla	86.5%	87.1%	61.2%	71.7%	77.8%	86.8%	87.1%	70.4%	86.0%	85.0%	85.0%	85.8%	86.0%	85.4%	84.2%	62.3%	73.7%	99.3%	96.9%		98.9%	98.2%	86.1%	94.3%
ATXN3L1 - P. abelii	85.4%	86.0%	60.3%	70.8%	77.4%	85.7%	86.0%	70.1%	85.2%	83.8%	84.2%	84.7%	84.9%	84.7%	83.1%	61.4%	73.3%	98.9%	96.5%	98.9%		97.8%	85.6%	93.5%
ATXN3L1 - N. leucogenys	85.8%	86.3%	60.8%	71.2%	77.1%	86.0%	86.3%	70.6%	86.0%	84.2%	84.6%	85.5%	85.3%	84.8%	84.2%	61.6%	73.1%	98.2%	96.1%	98.2%	97.8%		85.6%	93.9%
ATXN3L1 - C. jacchus	83.4%	83.9%	57.9%	68.8%	76.5%	83.3%	83.9%	70.2%	83.5%	82.1%	82.1%	82.7%	83.1%	82.8%	81.6%	60.6%	70.4%	86.1%	83.7%	86.1%	85.6%	85.6%		86.0%
ATXN3L1 - M. mulatta	85.6%	86.1%	60.1%	70.8%	77.1%	85.9%	86.3%	70.8%	86.1%	85.2%	85.0%	85.5%	86.2%	85.0%	85.5%	63.0%	74.0%	93.9%	91.9%	94.3%	93.5%	93.9%	86.0%	
ATXN3L2 - H. sapiens	82.7%	83.2%	59.1%	68.1%	75.0%	82.6%	82.4%	67.2%	81.3%	80.1%	77.9%	81.6%	81.5%	79.9%	80.3%	56.8%	69.9%	74.7%	72.4%	75.2%	74.2%	74.7%	71.5%	74.7%
ATXN3L2 - P. troglodytes	82.5%	83.1%	58.8%	67.9%	74.7%	82.6%	82.2%	66.7%	81.1%	80.0%	77.9%	81.4%	81.5%	79.7%	80.1%	56.6%	69.9%	74.9%	72.5%	75.4%	74.3%	74.9%	71.3%	74.9%
ATXN3L2 - G. gorilla	82.2%	82.7%	58.2%	67.9%	74.3%	82.3%	82.2%	67.0%	81.5%	79.6%	77.7%	81.7%	81.5%	80.0%	80.1%	57.1%	69.5%	75.0%	72.7%	75.6%	74.5%	75.0%	72.0%	75.2%
ATXN3L2 - P. abelii	81.8%	82.4%	58.0%	67.7%	74.1%	81.9%	81.8%	66.1%	80.8%	79.3%	77.4%	81.1%	81.2%	79.4%	79.8%	56.5%	69.5%	74.4%	72.1%	75.0%	73.9%	74.4%	71.1%	74.6%
ATXN3L2 - N. leucogenys	82.5%	83.1%	58.5%	68.3%	74.3%	82.6%	82.5%	66.5%	81.5%	80.0%	78.3%	81.7%	81.9%	79.9%	80.9%	57.1%	69.9%	75.8%	73.4%	76.3%	75.2%	75.8%	72.0%	75.6%
ATXN3L2 - M. mulatta	79.8%	79.9%	56.3%	65.7%	71.9%	80.3%	79.6%	65.0%	78.5%	76.9%	75.0%	80.0%	79.4%	77.3%	77.9%	56.0%	67.5%	72.8%	70.5%	73.4%	72.3%	73.2%	69.8%	73.0%
ATXN3L2 - C. jacchus	94.1%	94.6%	66.0%	77.6%	83.5%	93.4%	96.1%	73.4%	90.9%	90.9%	87.3%	91.4%	91.5%	90.4%	90.6%	68.1%	78.9%	84.1%	82.3%	84.7%	83.5%	83.9%	81.9%	83.9%
ATXN3L? - M. mulatta	97.0%	97.2%	68.4%	80.0%	86.3%	96.9%	95.6%	75.4%	92.2%	92.1%	88.6%	92.6%	93.6%	90.8%	91.5%	64.5%	78.8%	84.5%	82.3%	85.0%	84.3%	84.3%	82.3%	84.1%
ATXN3L? - C. jacchus	93.3%	93.4%	65.8%	77.0%	83.2%	92.6%	94.7%	72.5%	89.8%	90.1%	85.8%	90.9%	90.5%	88.7%	90.2%	62.7%	77.5%	82.3%	80.5%	82.9%	81.8%	82.1%	79.8%	82.7%
ATXN3L?1 - T. syrichta	87.6%	88.0%	62.1%	72.3%	80.2%	87.8%	87.8%	72.6%	87.1%	85.5%	85.4%	86.8%	87.6%	86.1%	84.9%	62.1%	75.7%	83.2%	81.1%	83.7%	83.2%	83.0%	80.6%	82.4%
ATXN3L?2 - T. syrichta	73.4%	73.7%	74.1%	73.7%	57.0%	74.4%	73.0%	57.3%	74.7%	72.6%	75.4%	72.4%	75.4%	74.1%	73.0%	47.0%	62.1%	67.9%	64.8%	68.3%	66.9%	67.6%	66.2%	67.6%
ATXN3L?2 - P. vampyrus	83.4%	83.6%	77.3%	83.1%	72.6%	83.1%	82.9%	66.0%	84.6%	83.0%	81.3%	82.3%	83.1%	80.7%	82.0%	50.2%	70.6%	75.3%	72.6%	75.8%	75.6%	75.6%	72.1%	75.6%
ATXN3L?1 - P. vampyrus	71.4%	71.8%	62.9%	68.5%	63.1%	72.1%	71.4%	57.4%	72.7%	72.5%	70.6%	70.8%	72.0%	69.2%	71.0%	44.8%	60.7%	63.3%	61.3%	63.7%	63.1%	63.7%	61.8%	63.3%
ATXN3L? - C. porcellus	66.8%	66.8%	45.4%	54.1%	58.3%	66.3%	67.3%	50.6%	65.5%	68.2%	63.4%	66.8%	66.3%	65.7%	66.6%	54.2%	59.5%	62.0%	60.9%	62.5%	61.6%	62.2%	61.3%	62.0%
ATXN3L? - O. cuniculus	84.4%	85.0%	60.6%	70.9%	76.6%	85.0%	84.1%	68.0%	83.5%	83.6%	84.2%	83.8%	84.6%	82.1%	82.7%	58.5%	71.2%	77.8%	76.2%	78.2%	77.6%	77.3%	75.1%	77.5%
ATXN3L?1 - C. familiarirs	89.3%	89.5%	64.6%	74.9%	80.3%	89.2%	89.9%	71.6%	89.7%	87.3%	84.4%	93.2%	88.8%	87.3%	89.5%	62.0%	75.4%	80.6%	78.6%	80.8%	80.1%	81.2%	77.9%	80.6%
ATXN3L?2 - C. familiaris	93.5%	93.7%	66.2%	77.4%	84.3%	93.2%	93.9%	74.4%	93.7%	91.7%	88.4%	97.2%	93.4%	91.1%	93.8%	65.8%	79.1%	84.3%	82.3%	84.8%	83.7%	84.5%	81.9%	84.8%
ATXN3L?1 - C. haffmanni	91.6%	91.9%	88.7%	91.9%	77.7%	91.4%	90.8%	71.7%	89.8%	89.9%	88.1%	88.8%	93.7%	87.7%	89.4%	52.8%	76.0%	80.9%	77.5%	81.7%	80.4%	81.2%	77.7%	81.2%
ATXN3L?2 - C. haffmanni	87.6%	87.8%	65.1%	76.0%	77.9%	90.7%	87.3%	69.7%	86.7%	88.4%	86.5%	86.8%	93.2%	85.8%	86.2%	58.9%	73.9%	79.9%	77.5%	80.4%	79.3%	79.7%	77.1%	79.2%
ATXN3L?1 - B. taurus	92.6%	92.8%	65.6%	76.9%	82.8%	92.2%	92.6%	75.0%	92.1%	91.5%	87.9%	93.0%	92.2%	89.1%	97.1%	64.0%	76.8%	82.4%	80.6%	83.0%	81.9%	83.0%	81.1%	84.3%
ATXN3L?2 - B. taurus	94.3%	94.5%	67.1%	78.6%	84.5%	93.8%	94.3%	76.0%	93.7%	93.1%	89.4%	94.7%	93.8%	90.8%	98.7%	65.2%	78.4%	83.7%	81.9%	84.3%	83.2%	84.3%	81.7%	85.6%
ATXN3L?3 - B. taurus	81.0%	81.2%	59.3%	67.6%	73.1%	80.8%	81.2%	66.6%	80.1%	80.0%	77.3%	81.3%	80.2%	78.6%	82.4%	55.5%	68.5%	72.0%	70.2%	72.4%	71.8%	72.4%	69.4%	72.7%
ATXN3L? - M. domestica	86.7%	87.2%	61.9%	71.5%	78.0%	87.4%	87.8%	72.3%	86.9%	85.4%	84.4%	86.2%	87.0%	93.9%	84.6%	62.0%	72.4%	80.4%	78.9%	80.8%	79.7%	80.0%	77.4%	79.5%
ATXN3 - CDS G. gallus	87.4%	87.8%	63.8%	73.4%	77.4%	88.0%	88.5%	73.2%	86.7%	87.3%	85.9%	85.1%	86.2%	88.7%	84.9%	60.6%	71.5%	80.8%	78.7%	80.8%	80.4%	80.0%	77.4%	80.8%

Table A6 - Continuation

	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3L?...	ATXN3 - ...
ATXN3 - CDS H. sapiens	85.6%	82.7%	82.5%	82.2%	81.8%	82.5%	79.8%	94.1%	97.0%	93.3%	87.6%	73.4%	83.4%	71.4%	66.8%	84.4%	89.3%	93.5%	91.6%	87.6%	92.6%	94.3%	81.0%	86.7%	87.4%
ATXN3 - CDS P. troglodytes	86.1%	83.2%	83.1%	82.7%	82.4%	83.1%	79.9%	94.6%	97.2%	93.4%	88.0%	73.7%	83.6%	71.8%	66.8%	85.0%	89.5%	93.7%	91.9%	87.8%	92.8%	94.5%	81.2%	87.2%	87.8%
ATXN3 - CDS G.gorilla	60.1%	59.1%	58.8%	58.2%	58.0%	58.5%	56.3%	66.0%	68.4%	65.8%	62.1%	74.1%	77.3%	62.9%	45.4%	60.6%	64.6%	66.2%	88.7%	65.1%	65.6%	67.1%	59.3%	61.9%	63.8%
ATXN3 - CDS P. abelii	70.8%	68.1%	67.9%	67.9%	67.7%	68.3%	65.7%	77.6%	80.0%	77.0%	72.3%	73.7%	83.1%	68.5%	54.1%	70.9%	74.9%	77.4%	91.9%	76.0%	76.9%	78.6%	67.6%	71.5%	73.4%
ATXN3 - CDS N. leucogenys	77.1%	75.0%	74.7%	74.3%	74.1%	74.3%	71.9%	83.5%	86.3%	83.2%	80.2%	57.0%	72.6%	63.1%	58.3%	76.6%	80.3%	84.3%	77.7%	77.9%	82.8%	84.5%	73.1%	78.0%	77.4%
ATXN3 - CDS M. mulatta	85.9%	82.6%	82.6%	82.3%	81.9%	82.6%	80.3%	93.4%	96.9%	92.6%	87.8%	74.4%	83.1%	72.1%	66.3%	85.0%	89.2%	93.2%	91.4%	90.7%	92.2%	93.8%	80.8%	87.4%	88.0%
ATXN3 - CDS C. jacchus	86.3%	82.4%	82.2%	82.2%	81.8%	82.5%	79.6%	96.1%	95.6%	94.7%	87.8%	73.0%	82.9%	71.4%	67.3%	84.1%	89.9%	93.9%	90.8%	87.3%	92.6%	94.3%	81.2%	87.8%	88.5%
ATXN3 - CDS T. syrichta	70.8%	67.2%	66.7%	67.0%	66.1%	66.5%	65.0%	73.4%	75.4%	72.5%	72.6%	57.3%	66.0%	57.4%	50.6%	68.0%	71.6%	74.4%	71.7%	69.7%	75.0%	76.0%	66.6%	72.3%	73.2%
ATXN3 - CDS P.vampyrus	86.1%	81.3%	81.1%	81.5%	80.8%	81.5%	78.5%	90.9%	92.2%	89.8%	87.1%	74.7%	84.6%	72.7%	65.5%	83.5%	89.7%	93.7%	89.8%	86.7%	92.1%	93.7%	80.1%	86.9%	86.7%
ATXN3 - CDS C.porcellus	85.2%	80.1%	80.0%	79.6%	79.3%	80.0%	76.9%	90.9%	92.1%	90.1%	85.5%	72.6%	83.0%	72.5%	68.2%	83.6%	87.3%	91.7%	89.9%	88.4%	91.5%	93.1%	80.0%	85.4%	87.3%
ATXN3 - CDS O. cuniculus	85.0%	77.9%	77.9%	77.7%	77.4%	78.3%	75.0%	87.3%	88.6%	85.8%	85.4%	75.4%	81.3%	70.6%	63.4%	84.2%	84.4%	88.4%	88.1%	86.5%	87.9%	89.4%	77.3%	84.4%	85.9%
ATXN3 - CDS C. familiaris	85.5%	81.6%	81.4%	81.7%	81.1%	81.7%	80.0%	91.4%	92.6%	90.9%	86.8%	72.4%	82.3%	70.8%	66.8%	83.8%	93.2%	97.2%	88.8%	86.8%	93.0%	94.7%	81.3%	86.2%	85.1%
ATXN3 - CDS C. haffmanni	86.2%	81.5%	81.5%	81.5%	81.2%	81.9%	79.4%	91.5%	93.6%	90.5%	87.6%	75.4%	83.1%	72.0%	66.3%	84.6%	88.8%	93.4%	93.7%	93.2%	92.2%	93.8%	80.2%	87.0%	86.2%
ATXN3 - CDS M. domestica	85.0%	79.9%	79.7%	80.0%	79.4%	79.9%	77.3%	90.4%	90.8%	88.7%	86.1%	74.1%	80.7%	69.2%	65.7%	82.1%	87.3%	91.1%	87.7%	85.8%	89.1%	90.8%	78.6%	93.9%	88.7%
ATXN3 - CDS B. taurus	85.5%	80.3%	80.1%	80.1%	79.8%	80.9%	77.9%	90.6%	91.5%	90.2%	84.9%	73.0%	82.0%	71.0%	66.6%	82.7%	89.5%	93.8%	89.4%	86.2%	97.1%	98.7%	82.4%	84.6%	84.9%
ATXN3 - CDS O. garnettii	63.0%	56.8%	56.6%	57.1%	56.5%	57.1%	56.0%	68.1%	64.5%	62.7%	62.1%	47.0%	50.2%	44.8%	54.2%	58.5%	62.0%	65.8%	52.8%	58.9%	64.0%	65.2%	55.5%	62.0%	60.6%
ATXN3L? - O. garnettii - se...	74.0%	69.9%	69.9%	69.5%	69.5%	69.9%	67.5%	78.9%	78.8%	77.5%	75.7%	62.1%	70.6%	60.7%	59.5%	71.2%	75.4%	79.1%	76.0%	73.9%	76.8%	78.4%	68.5%	72.4%	71.5%
ATXN3L1 - H. sapiens	93.9%	74.7%	74.9%	75.0%	74.4%	75.8%	72.8%	84.1%	84.5%	82.3%	83.2%	67.9%	75.3%	63.3%	62.0%	77.8%	80.6%	84.3%	80.9%	79.9%	82.4%	83.7%	72.0%	80.4%	80.8%
ATXN3L1 - P. troglodytes	91.9%	72.4%	72.5%	72.7%	72.1%	73.4%	70.5%	82.3%	82.3%	80.5%	81.1%	64.8%	72.6%	61.3%	60.9%	76.2%	78.6%	82.3%	77.5%	77.5%	80.6%	81.9%	70.2%	78.9%	78.7%
ATXN3L1 - G. gorilla	94.3%	75.2%	75.4%	75.6%	75.0%	76.3%	73.4%	84.7%	85.0%	82.9%	83.7%	68.3%	75.8%	63.7%	62.5%	78.2%	80.8%	84.8%	81.7%	80.4%	83.0%	84.3%	72.4%	80.8%	80.8%
ATXN3L1 - P. abelii	93.5%	74.2%	74.3%	74.5%	73.9%	75.2%	72.3%	83.5%	84.3%	81.8%	83.2%	66.9%	75.6%	63.1%	61.6%	77.6%	80.1%	83.7%	80.4%	79.3%	81.9%	83.2%	71.8%	79.7%	80.4%
ATXN3L1 - N. leucogenys	93.9%	74.7%	74.9%	75.0%	74.4%	75.8%	73.2%	83.9%	84.3%	82.1%	83.0%	67.6%	75.6%	63.7%	62.2%	77.3%	81.2%	84.5%	81.2%	79.7%	83.0%	84.3%	72.4%	80.0%	80.0%
ATXN3L1 - C. jacchus	86.0%	71.5%	71.3%	72.0%	71.1%	72.0%	69.8%	81.9%	82.3%	79.8%	80.6%	66.2%	72.1%	61.8%	61.3%	75.1%	77.9%	81.9%	77.7%	77.1%	81.1%	81.7%	69.4%	77.4%	77.4%
ATXN3L1 - M. mulatta		74.7%	74.9%	75.2%	74.6%	75.6%	73.0%	83.9%	84.1%	82.7%	82.4%	67.6%	75.6%	63.3%	62.0%	77.5%	80.6%	84.8%	81.2%	79.2%	84.3%	85.6%	72.7%	79.5%	80.8%
ATXN3L2 - H. sapiens	74.7%		98.6%	97.1%	95.9%	95.7%	90.0%	79.5%	81.8%	80.7%	76.5%	67.2%	71.3%	61.3%	57.8%	76.7%	79.1%	81.3%	76.9%	75.3%	79.1%	80.4%	70.9%	75.6%	75.8%
ATXN3L2 - P. troglodytes	74.9%	98.6%		97.5%	96.3%	96.1%	90.7%	79.3%	81.8%	80.4%	76.8%	67.6%	71.1%	61.0%	58.2%	76.7%	79.0%	81.1%	76.6%	75.1%	79.0%	80.2%	70.5%	75.6%	75.9%
ATXN3L2 - G. gorilla	75.2%	97.1%	97.5%		95.9%	95.7%	91.3%	80.0%	81.6%	81.4%	75.9%	67.2%	70.2%	61.0%	57.8%	77.2%	79.0%	81.5%	76.1%	74.9%	79.3%	80.2%	70.9%	75.9%	76.3%
ATXN3L2 - P. abelii	74.6%	95.9%	96.3%	95.9%		95.4%	90.3%	79.0%	81.0%	80.6%	75.7%	67.2%	69.9%	60.7%	57.6%	76.5%	78.1%	80.8%	75.5%	74.8%	78.7%	79.9%	69.9%	75.3%	75.7%
ATXN3L2 - N. leucogenys	75.6%	95.7%	96.1%	95.7%	95.4%		91.1%	79.7%	82.0%	80.9%	76.1%	67.2%	70.9%	62.2%	58.7%	77.6%	79.1%	81.8%	76.9%	75.8%	80.0%	80.9%	70.3%	75.9%	76.8%
ATXN3L2 - M. mulatta	73.0%	90.0%	90.7%	91.3%	90.3%	91.1%		77.1%	79.8%	79.4%	73.4%	68.6%	67.7%	59.1%	57.1%	75.0%	77.8%	79.0%	74.8%	73.6%	76.9%	77.6%	69.4%	73.2%	73.7%
ATXN3L2 - C. jacchus	83.9%	79.5%	79.3%	80.0%	79.0%	79.7%	77.1%		92.2%	91.8%	84.7%	71.4%	79.5%	68.6%	66.4%	81.0%	86.2%	90.2%	86.4%	84.1%	89.1%	90.8%	78.1%	85.4%	85.8%
ATXN3L? - M. mulatta	84.1%	81.8%	81.8%	81.6%	81.0%	82.0%	79.8%	92.2%		91.4%	85.6%	72.4%	81.4%	69.7%	65.3%	83.0%	88.2%	92.2%	88.7%	85.8%	91.1%	92.1%	79.9%	85.4%	86.0%
ATXN3L? - C. jacchus	82.7%	80.7%	80.4%	81.4%	80.6%	80.9%	79.4%	91.8%	91.4%		83.1%	71.7%	78.4%	67.9%	64.7%	82.5%	86.3%	89.8%	86.7%	83.3%	88.7%	90.3%	78.6%	83.8%	84.2%
ATXN3L?1 - T. syrichta	82.4%	76.5%	76.8%	75.9%	75.7%	76.1%	73.4%	84.7%	85.6%	83.1%		67.6%	77.3%	65.1%	63.1%	80.0%	82.5%	85.8%	82.7%	80.1%	83.4%	85.0%	74.4%	82.1%	80.8%
ATXN3L?2 - T. syrichta	67.6%	67.2%	67.6%	67.2%	67.2%	67.2%	68.6%	71.4%	72.4%	71.7%	67.6%		67.9%	70.6%	50.9%	68.6%	68.9%	70.6%	74.4%	74.7%	71.3%	72.4%	63.5%	67.9%	71.3%
ATXN3L?2 - P. vampyrus	75.6%	71.3%	71.1%	70.2%	69.9%	70.9%	67.7%	79.5%	81.4%	78.4%	77.3%	67.9%		79.8%	58.6%	75.2%	79.0%	82.9%	83.2%	83.4%	81.2%	82.9%	76.5%	74.3%	77.5%
ATXN3L?1 - P. vampyrus	63.3%	61.3%	61.0%	61.0%	60.7%	62.2%	59.1%	68.6%	69.7%	67.9%	65.1%	70.6%	79.8%		48.3%	64.8%	69.6%	71.0%	84.6%	69.6%	71.4%	71.4%	64.2%	64.9%	66.6%
ATXN3L? - C. porcellus	62.0%	57.8%	58.2%	57.8%	57.6%	58.7%	57.1%	66.4%	65.3%	64.7%	63.1%	50.9%	58.6%	48.3%		60.6%	64.8%	66.8%	58.9%	61.3%	65.9%	67.0%	57.4%	62.5%	61.1%
ATXN3L? - O. cuniculus	77.5%	76.7%	76.7%	77.2%	76.5%	77.6%	75.0%	81.0%	83.0%	82.5%	80.0%	68.6%	75.2%	64.8%	60.6%		79.4%	83.3%	80.3%	77.9%	81.6%	83.2%	71.7%	77.6%	78.2%
ATXN3L?1 - C. familiarirs	80.6%	79.1%	79.0%	79.0%	78.1%	79.1%	77.8%	86.2%	88.2%	86.3%	82.5%	68.9%	79.0%	69.6%	64.8%	79.4%		92.1%	85.1%	81.6%	88.6%	89.9%	78.7%	82.3%	81.9%
ATXN3L?2 - C. familiaris	84.8%	81.3%	81.1%	81.5%	80.8%	81.8%	79.0%	90.2%	92.2%	89.8%	85.8%	70.6%	82.9%	71.0%	66.8%	83.3%	92.1%		87.7%	86.2%	93.0%	94.3%	81.0%	85.4%	84.5%
ATXN3L?1 - C. haffmanni	81.2%	76.9%	76.6%	76.1%	75.5%	76.9%	74.8%	86.4%	88.7%	86.7%	82.7%	74.4%	83.2%	84.6%	58.9%	80.3%	85.1%	87.7%		94.8%	88.0%	89.8%	78.3%	81.7%	83.5%
ATXN3L?2 - C. haffmanni	79.2%	75.3%	75.1%	74.9%	74.8%	75.8%	73.6%	84.1%	85.8%	83.3%	80.1%	74.7%	83.4%	69.6%	61.3%	77.9%	81.6%	86.2%	94.8%		84.9%	86.3%	73.0%	81.2%	79.5%
ATXN3L?1 - B. taurus	84.3%	79.1%	79.0%	79.3%	78.7%	80.0%	76.9%	89.1%	91.1%	88.7%	83.4%	71.3%	81.2%	71.4%	65.9%	81.6%	88.6%	93.0%	88.0%	84.9%		98.0%	81.8%	83.4%	83.5%
ATXN3L?2 - B. taurus	85.6%	80.4%	80.2%	80.2%	79.9%	80.9%	77.6%	90.8%	92.1%	90.3%	85.0%	72.4%	82.9%	71.4%	67.0%	83.2%	89.9%	94.3%	89.8%	86.3%	98.0%		82.7%	85.0%	85.0%
ATXN3L?3 - B. taurus	72.7%	70.9%	70.5%	70.9%	69.9%	70.3%	69.4%	78.1%	79.9%	78.6%	74.4%	63.5%	76.5%	64.2%	57.4%	71.7%	78.7%	81.0%	78.3%	73.0%	81.8%	82.7%		73.7%	75.0%
ATXN3L? - M. domestica	79.5%	75.6%	75.6%	75.9%	75.3%	75.9%	73.2%	85.4%	85.4%	83.8%	82.1%	67.9%	74.3%	64.9%	62.5%	77.6%	82.3%	85.4%	81.7%	81.2%	83.4%	85.0%	73.7%		83.5%
ATXN3 - CDS G. gallus	80.8%	75.8%	75.9%	76.3%	75.7%	76.8%	73.7%	85.8%	86.0%	84.2%	80.8%	71.3%	77.5%	66.6%	61.1%	78.2%	81.9%	84.5%	83.5%	79.5%	83.5%	85.0%	75.0%	83.5%	

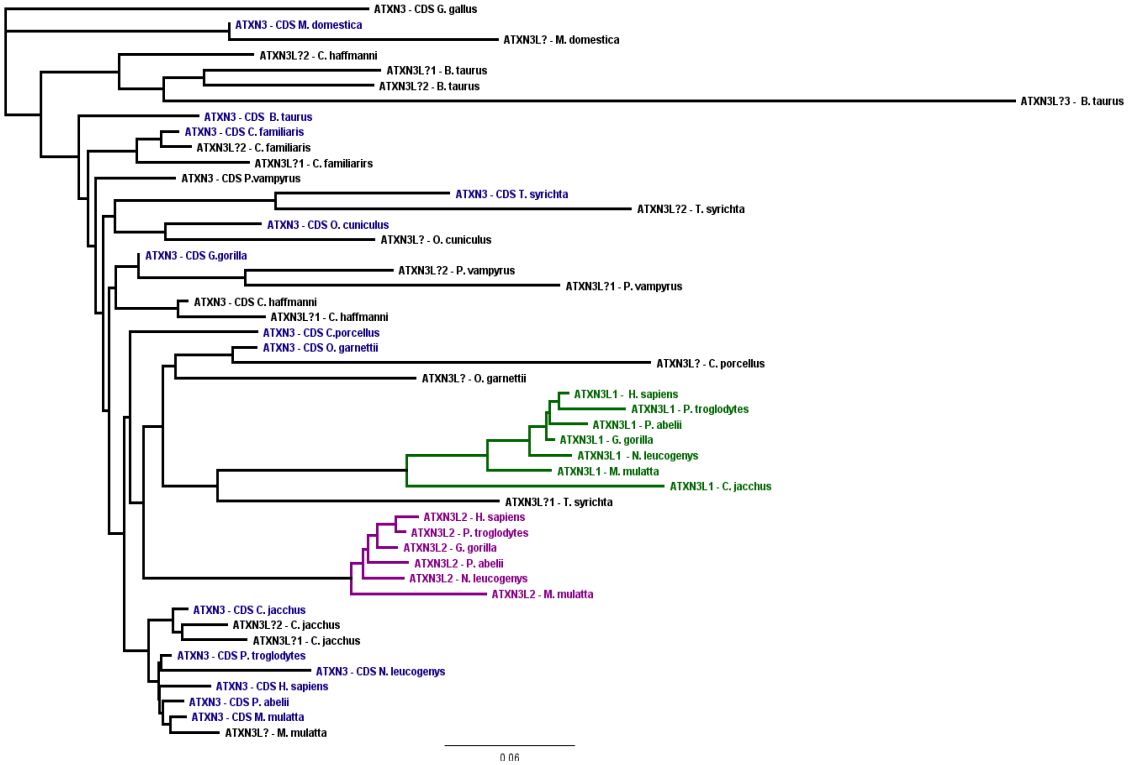

Figure A4 – Phylogenetic tree of *ATXN3*, *ATXN3L1* and *ATXN3L2* for mammals, without the GAG tract, based on the genetic distances presented on Table A5. *Gallus gallus* was used as outgroup.

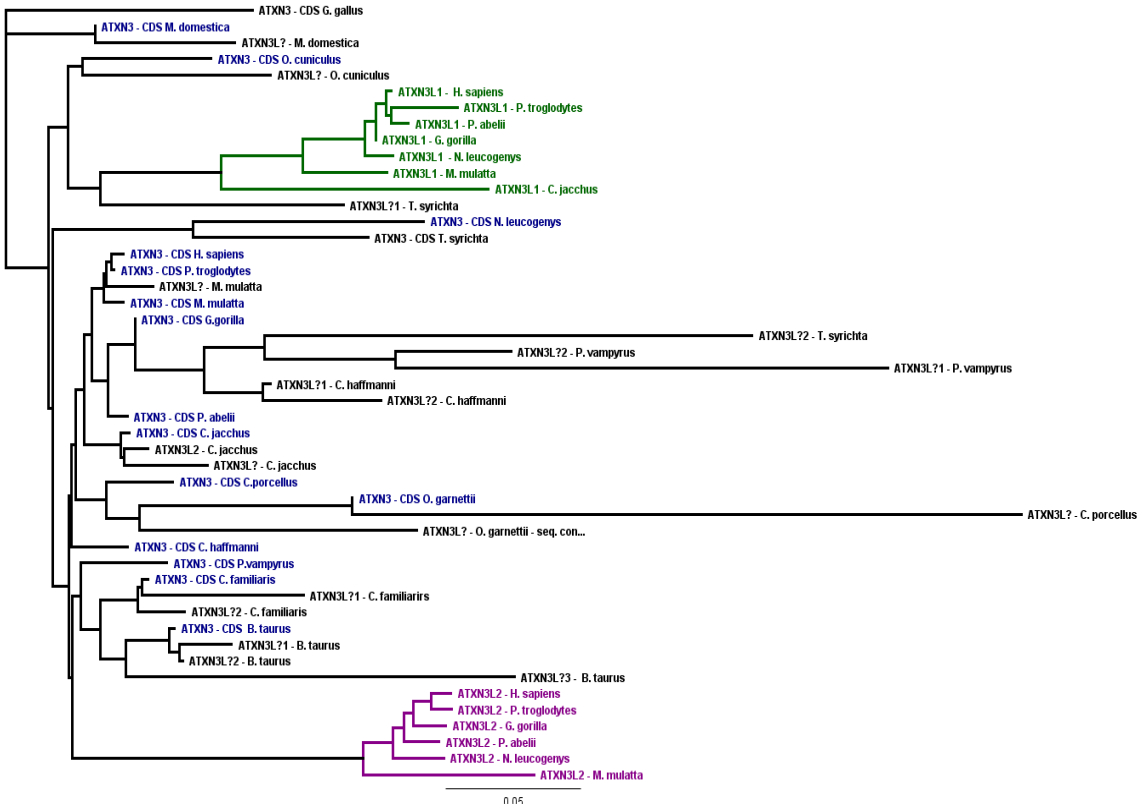

Figure A5 – Phylogenetic tree of *ATXN3*, *ATXN3L1* and *ATXN3L2* Josephin domain for mammals, based on the genetic distances presented on table A6. *Gallus gallus* was used as outgroup.

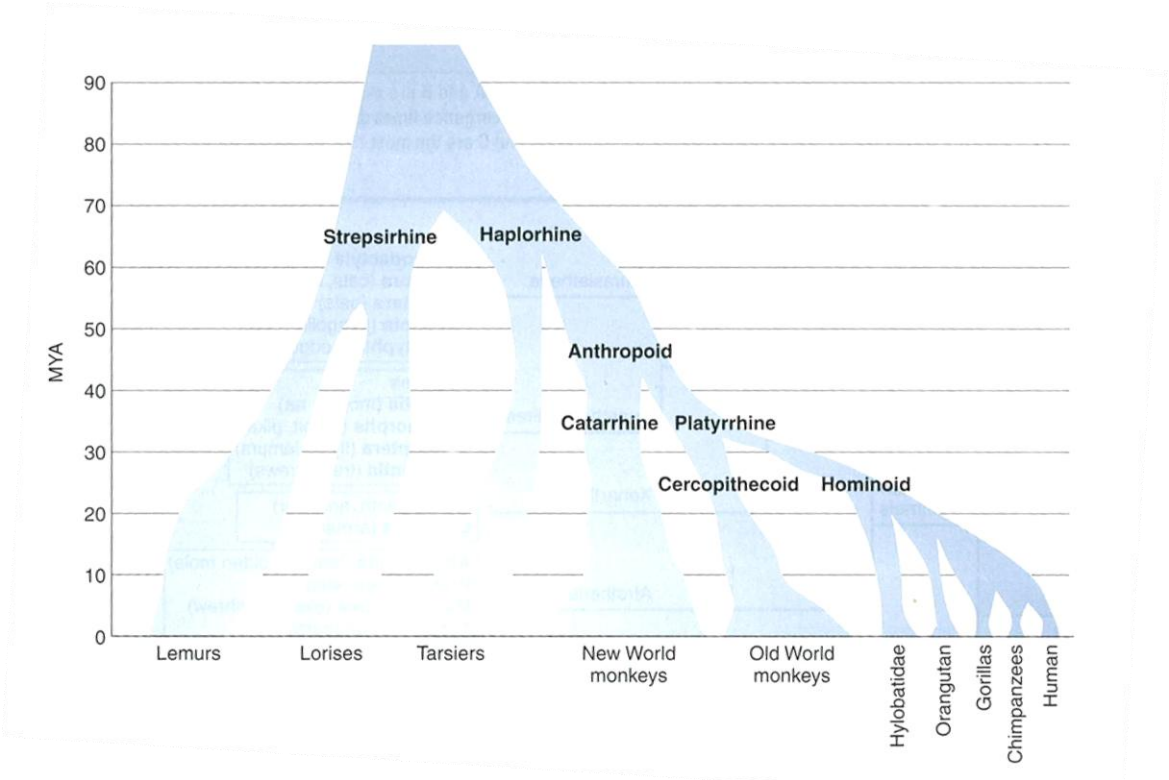

Figure A6 – Phylogeny of extant primate groups and corresponding date of branching splits.[79]

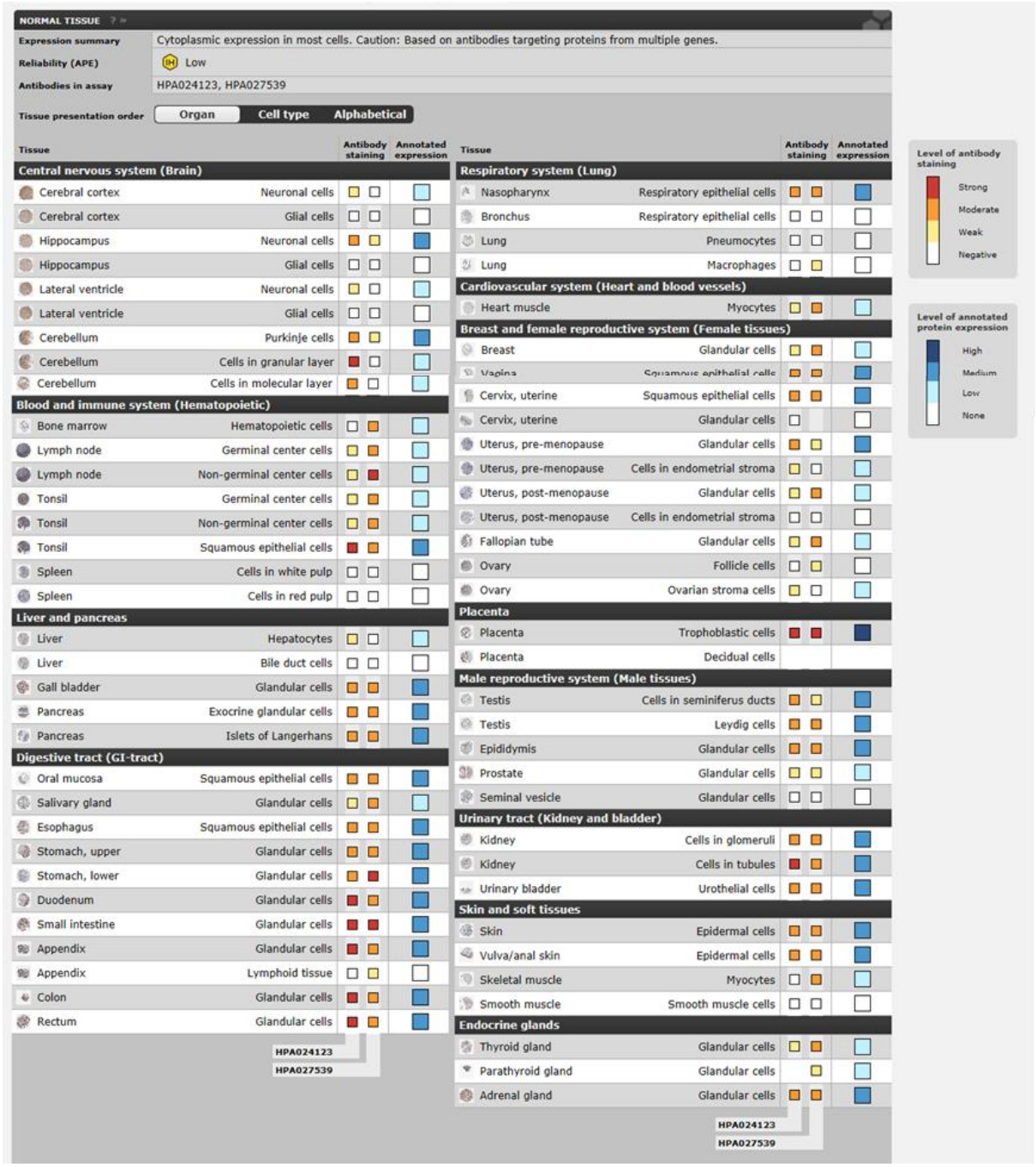

Figure A7 - ATXN3L1 expression levels in different human organs/tissues. (data from the Human Protein Atlas: http://www.proteinatlas.org/ENSG00000123594)

Hs.382641 - ATXN3L: Ataxin 3-like

Breakdown by Body Sites

		Hs.382641
adipose tissue	0	0 / 12865
adrenal gland	0	0 / 32921
ascites	0	0 / 39829
bladder	0	0 / 29856
blood	0	0 / 122262
bone	0	0 / 71609
bone marrow	0	0 / 48711
brain	1	2 / 1092524
cervix	0	0 / 48469
connective tissue	0	0 / 149048
ear	0	0 / 16098
embryonic tissue	0	0 / 212847
esophagus	0	0 / 20152
eye	0	0 / 208810
heart	0	0 / 89512
intestine	0	0 / 232030
kidney	0	0 / 210738
larynx	0	0 / 23489
liver	0	0 / 205232
lung	0	0 / 334751
lymph	0	0 / 44292
lymph node	0	0 / 89697
mammary gland	0	0 / 151228
mouth	0	0 / 66139
muscle	0	0 / 106323
nerve	0	0 / 15526
ovary	0	0 / 101482
pancreas	0	0 / 213410
parathyroid	0	0 / 20579
pharynx	0	0 / 40762
prostate	0	0 / 189585
salivary gland	0	0 / 20264
skin	0	0 / 210718
spleen	0	0 / 53365
stomach	0	0 / 95775
testis	21	7 / 327305
thymus	0	0 / 79668
thyroid	0	0 / 46584
tonsil	0	0 / 17016
trachea	0	0 / 51769
umbilical cord	0	0 / 13764
uterus	0	0 / 232051
vascular	0	0 / 51637

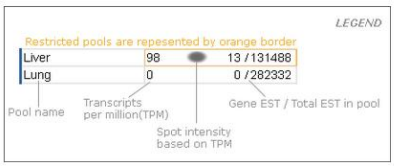

Figure A8 - ATXN3L1 expression levels in different human organs/tissues. (data from NCBI – UniGene – EST Profile data base: http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer.cgi?uglist=Hs.382641)