Characterization, developmental expression and evolutionary features of the huntingtin gene in the amphioxus Branchiostoma floridae

Simona Candiani*1, Mario Pesterino1, Elena Cattaneo2 and Marzia Tartari2

Address: 1Department of Biology, University of Genoa, viale Benedetto XV 5, 16132, Genoa, Italy and 2Centre for Stem Cell Research and Department of Pharmacological Sciences, University of Milan, Via Balzaretti 9, 20133, Milan, Italy

Email: Simona Candiani* - candiani@unige.it; Mario Pesterino - pesta@unige.it; Elena Cattaneo - elena.cattaneo@unimi.it; Marzia Tartari - marzia.tartari@unimi.it

* Corresponding author

Abstract

Background: Huntington's disease is an inherited neurodegenerative disorder that is caused by the expansion of an N-terminal polyQ stretch in the huntingtin protein. In order to investigate the hypothesis that huntingtin was already involved in development of the nervous system in the last common ancestor of chordates, we isolated and characterised the huntingtin homologue from the amphioxus Branchiostoma floridae. In the present paper the amphioxus general term must be referred to Branchiostoma floridae.

Results: In this report, we show that the exon-intron organization of the amphioxus huntingtin gene is highly conserved with that of other vertebrates species. The AmphHtt protein has two glutamine residues in the position of the typical vertebrate polyQ tract. Sequence conservation is greater along the entire length of the protein than in a previously identified Ciona huntingtin. The first three N-terminal HEAT repeats are highly conserved in vertebrates and amphioxus, although exon rearrangement has occurred in this region. AmphHtt expression is detectable by in situ hybridization starting from the early neurula stage, where it is found in cells of the neural plate. At later stages, it is retained in the neural compartment but also it appears in limited and well-defined groups of non-neural cells. At subsequent larval stages, AmphHtt expression is detected in the neural tube, with the strongest signal being present in the most anterior part.

Conclusion: The cloning of amphioxus huntingtin allows to infer that the polyQ in huntingtin was already present 540 million years ago and provides a further element for the study of huntingtin function and its evolution along the deuterostome branch.

Background

Huntingtin is a completely soluble, ubiquitously expressed 350-kDa protein of 3144 aa which, once mutated, causes Huntington's disease (HD), a late-onset neurodegenerative disease characterised by movement disorders, dementia and psychiatric disturbances, and by preferential vulnerability of striatal and cortical neurons [1].
One obviously important portion of the mammalian protein is the polyQ tract, which is present in the normal protein with up to 36 glutamines, but becomes further elongated in the mutant protein as a consequence of the DNA CAG triplet repeat expansion in the gene [1]. The role of the polyQ region in huntingtin’s physiological function is currently unknown in mammals. The polyQ tract is found in many transcription factors [2] and, in huntingtin, is followed by a recently discovered polyP tract that may give it structural and biochemical advantages [3]. Recent studies have suggested that the polyP tract helps to maintain protein solubility [4]. It is also possible that, during evolution, an expanded polyQ has conferred important molecular function(s) partially because of its cooperation with the emerging polyP tract. An aberrantly expanded polyQ region in huntingtin is sufficient to cause HD.

Investigating the physiological functions of huntingtin involves a number of difficulties, there is a biological evidence indicating that the protein has individual beneficial activities in the brain (e.g. it is anti-apoptic and neuroprotective) [5]. Its primary amino acid sequence reveals little about its function because it is unlike any other known protein and contains only a few known sequence motifs. However, it does have HEAT repeat consensas (approximately 40-amino-acid-long sequences that occur multiple times) [6,7], whose presence indicates an ability to participate in multiple protein-protein interaction networks, as it has been further documented in subsequent studies [8]. However, the presence of these domains does not allow a complete definition of its biological function(s). Furthermore, the influence of the evolution of the HEAT repeats is far from being established and a more thorough knowledge of their presence in non-vertebrate huntingtin may help us to understand their role in the mammalian protein.

In the absence of information about its three-dimensional structure, comparisons of huntingtin homologues should help to define conserved or newly emerged functional domains in mammalian cells, although only limited information is available about huntingtin in other species. Furthermore, the comparative expression and distribution of huntingtin mRNA in different organisms may be instructive as to its role in mammals. As huntingtin homologues in the vertebrate subphylum are highly conserved, whereas Drosophila melanogaster huntingtin diverges substantially (particularly in its N-terminal portion and the absence of the polyQ-rich region) [9], we have recently concentrated our study on the cloning and comparative analysis of invertebrate deuterostome homologues, such as ascidians [10], echinoderms [11; Tartari et al., unpublished] and, as described in this paper, amphioxus. These molecules may share similar (but not identical) functions to those of the human protein and may help in reconstructing the evolution of huntingtin.

Before this study, some of us studied a complete huntingtin gene from the ascidians C. intestinalis and C. savignyi [10] and found that Ciona huntingtin contains regions that have specifically evolved in this genus and are concentrated in the central part of the protein, whereas major differences in the N-terminal part indicate the more recent evolution of this group-specific portion of the protein. Furthermore, C. intestinalis huntingtin transcript exhibits an alternative splicing in the 3' coding region and in the 3'UTR [10]. One further characteristic of ascidian huntingtin is the complete absence of the polyQ-rich region, whereas polyQ is described for the first time in zebrafish huntingtin that contains a QQQQ tract [12].

A partial huntingtin sequence is available from two sea urchin species [11], which shows that their nervous system organisation is profoundly different from that of chordates [13-15]. In situ hybridisation, using a probe from the 3’ region of the sea urchin Helicodaris erythrogramma huntingtin homologue, has shown that huntingtin expression is confined to non-neuronal compartments [11]. A similar experiment using the ascidian Halocynthia roretzi showed ubiquitous expression of the huntingtin homologue as in vertebrates [11]. On the contrary, vertebrate huntingtin is expressed throughout life and in all tissues, but it is particularly enriched in brain, suggesting that it may play a particular role in this district. Consistently, there is now considerable genetic and biological evidence indicating that huntingtin is important for the formation and maintenance of brain neurons, as it contributes to neuronal survival, neuronal gene expression and BDNF production [16].

Taking advantage of the newly available data from amphioxus B. floridae genome sequencing, we here describe the cloning of amphioxus huntingtin (AmphHtt), coming from an invertebrate chordate whose phylogenetic node of divergence is thought to go back 540 million years, while Ciona seems to have diverged more recently [17-19]. We also describe for the first time the distribution of huntingtin mRNA in this invertebrate chordate, whose nervous system development is particularly close to that of vertebrates as it includes vertebrate-like anatomical characteristics such as a dorsal nerve cord, a notochord and segmentally arranged muscles.

We show that AmphHtt protein has two glutamines in the same polyQ tract position of vertebrate homologues, thus suggesting that polyQ was already present 540 million years ago. We also report that the primary sequence around the QQ is highly conserved with respect to vertebrates and that sequence conservation along the entire
length of the protein is greater than in *C. intestinalis* huntingtin. The first three N-terminal HEAT repeats are highly conserved in vertebrates and amphioxus, although exon rearrangement has occurred in this region. We also show that amphioxus huntingtin is not exclusively neural, but mainly enriched in the neural compartment; this is a clear indication that huntingtin, in amphioxus, at least within the analyzed developmental stages, could have a specific neuronal function.

**Results**

**Cloning and characterisation of the amphioxus huntingtin sequence (AmphiHtt)**

Starting from the recently available *B. floridae* genomic sequencing data, two scaffolds were identified by means of TblastN similarity as possibly containing the huntingtin gene. A first messenger prediction was produced and used to design eight primer pairs (Figure 1 and Additional file 1). PCR assays with first-strand adult cDNA and a 5- to 24-hour *B. floridae* embryos cDNA library as templates, yielded eight overlapping clones that constituted the coding sequence of an amphioxus huntingtin gene, *AmphiHtt*. The *AmphiHtt* cDNA sequence (deposited in GenBank: Accession No. EF210456) is 9293 bp long and contains a putative 2972 bp open reading frame encoding a 3090 amino acid protein; an in-frame stop codon upstream from the putative start codon was found at 9 bp, and a stop codon at 9291 bp.

Sequence similarity analysis of the entire huntingtin sequences of several chordates (Figure 2A) showed that the amphioxus sequence has 46% percent identity with mammals, 46–48% with fish, and 34% with *Ciona*, whereas *Ciona* proteins have only 34–36% identity with vertebrate huntingtin. Furthermore, additional analysis led to the calculation of a 124–127 aa divergence between *Ciona* and vertebrates (Figure 2A), but only an 80–83 aa divergence between amphioxus and vertebrates (Figure 2A). Multiple sequence alignments were generated and huntingtin phylogenetic trees were constructed (Figure 2B, Additional files 2 and 3).

The results obtained using these methodologies indicate that AmphiHtt is more similar to vertebrate huntingtin than to *Ciona* proteins, which apparently conflicts with the current view that tunicates are the sister group of vertebrates [17-19] (probably likely due to the generally high rates of evolution of the tunicate genome [20] with long-branch attraction as a biasing factor in their phylogenetic position).

Qualitatively, the AmphiHtt sequence has two glutamines (Q17 and Q18) at the corresponding position to polyQ in vertebrates (Figure 3), whereas *Ciona* huntingtin has an aromatic amino acid group in this position. Amphioxus is therefore the first known non-vertebrate species to contain glutamine residues in huntingtin, thus dating the presence of glutamine in a non-vertebrate contest and indicating that the common ancestor of cephalochordates and vertebrates already possessed this characteristic. The polyP tract is only present in mammalian huntingtin and absent in non-mammalian vertebrates, *Ciona* huntingtin and AmphiHtt. In addition, the first 17 amino acids of AmphiHtt, with its three lysines that have been shown to participate in determining the intracellular distribution of the protein between the cytoplasm and nucleus in vertebrates [21], are also strongly conserved (Figure 3).

Finally, in order to look at a variability in the polyQ tract and at a somatic instability, we carried out a BLAST search of the NCBI amphioxus dbEST and Trace-Archives databases using AmphiHtt cDNA sequence as a query. No matching EST sequences were found. Furthermore, all shot gun sequences covering the polyQ tract confirmed the exclusive presence of two glutamines residues and consequently the absence of a somatic instability.

We next searched for AmphiHtt HEAT repeats by applying the REP program to the AmphiHtt sequence (Figure 4, Tables 1 and 2). HEAT repeats are bioinformatic consensas present in vertebrate huntingtin that may have molecular
activity. We therefore applied the score of the consensus, sequence similarity and relative position of the HEATs in a multiple alignment (containing a good number of informative protein sequences) in order to evaluate the most possible functionally active HEATs. Although this evaluation may need to be revised in the future, we found that the program identified six highly scored HEAT-AAA repeats (aa 75–113, aa 156–194, aa 198–236, aa 306–344, aa 802–840, and aa 2476–2784). Comparison of the primary sequences of amphioxus, human and Ciona huntingtin revealed five potential additional HEAT repeats (aa 1371–1409, aa 1556–1595, aa 1618–1656, aa 2324–2362, and aa 2785–2824).

Figure 2

A. Sequence identity matrix of huntingtin, reconstructed from protein sequences: percent divergence is calculated by comparing sequence pairs in relation to their relative positions in the alignment; percent identity is estimated by comparing percent sequence identity directly without considering phylogenetic relationships. B. Unrooted tree of huntingtin proteins. (The huntingtin proteins used to reconstruct the tree were from the following species: H. sapiens, R. norvegicus, M. musculus, S. scrofa, G. gallus, X. tropicalis, F. rubripes, T. negroviridis, D. rerio, C. savignyi, C. intestinalis, B. florideae). Phylogenetic analysis reveals that B. florideae huntingtin branching just basal to vertebrate huntingtin protein, and that it groups very robustly with vertebrate orthologues. The branch length in the tree is proportional to the number of amino acid substitutions, and the scale bar indicates 0.1 amino acid substitution per position in the sequences. The numbers on each node indicate the percentage confidence values based on 100 replicate bootstrap resamplings of the alignment data.
2927–2965, aa 3020–3059) that are consistent with the previously published consensas in human huntingtin [6,7,10], whereas the contrary, none of the consensas specific to the central part of the Ciona homologue (aa 867–905 and aa 1341–1378), seems to have a correspondent in amphioxus. Numbering the HEAT consensas of human huntingtin [10] from the N-terminal to the C-terminal end, the sequence and position of the first three amino terminal human HEAT repeats are very well conserved in AmphiHtt (Figure 4, Tables 1 and 2), as are the ninth and eleventh in the central region, and the fifteenth in the C-terminal portion. Amphioxus seems to have one HEAT consensus at aa 306–344 that has no correspondent in human huntingtin, although a similar sequence can be found in Ciona huntingtin. An opposite situation can be established for the fourteenth human HEAT, which seems to have a correspondent only in amphioxus and not in Ciona, whereas the fifteenth seems to be specifically lost in Ciona. Finally, two additional consensas (add1 and add2, see Tables 1 and 2) in the C-terminal portion of Ciona huntingtin (aa 2771–2809, aa 2864–2904) met a possible correspondence in both amphioxus (aa 2927–2965, aa 3020–3059) and human huntingtin.

Genomic organisation of the huntingtin gene: a comparative overview

The AmphiHtt cDNA sequence was superimposed on the genomic sequence available at JGI, and it was found that the genomic coordinates for the AmphiHtt cDNAs were from 29810 nt to 81840 nt in scaffold_613 (minus strand), and from 685048 nt to 743806 nt in scaffold_378 (plus strand). Taking advantage of the newly cloned cDNA sequence, we reconstructed the genomic organisation of the huntingtin gene in both scaffolds using two genomic mapping software programs: GMAP and Wise2. The amphioxus huntingtin gene contains 63 exons and spans a genomic region of over 50 kb, whereas vertebrate sequences have 67 exons and corresponding gene lengths ranging from 80 Kb (fishes) to 180 Kb (humans), and Ciona has 61 coding exons covering a genomic region of 33 kb [10]. The predicted sequence of our AmphiHtt corresponded with minor polymorphisms to the coding sequences predicted from the two genomic scaffolds. Nevertheless, as the two scaffolds have some assembly errors and several tandem repeat elements (deduced using the Tandem repeats Finder program, [22]), the information on exons 36 and 37 comes from the scaffold_613, and that on exon 57 from the scaffold_378 (Figure 5 and Additional file 4). Furthermore, in this preliminary genomic assembly, the information in some intron sequences is not conclusive but seems to match our exon mRNA data perfectly: 60 exons are correctly recognised in both scaffolds. We therefore suggest that there is only a single copy gene of huntingtin in the amphioxus genome, and that the two scaffolds represent the two alleles of the same gene.

Comparison of the genomic and cDNA sequences of AmphiHtt allowed us to determine its exon/intron structure, and to compare it to what is known for members of the same family in other chordates. Furthermore, analysis of the pattern of exon-intron junctions can provide important insights into the evolution of huntingtin genes. In particular, as shown in Figure 5, we compared the genomic organisation of the H. sapiens (Chr4:3103557–3288752; assembly version v35), B. floridiae and C. intestinalis (scaffold_31: 333864–386142; assembly version v 1.95) huntingtin genes. Comparison of the conservation of the exon/intron boundaries revealed the presence in amphioxus of 51 introns in con-
Figure 4
Comparison of the huntingtin HEAT repeats from *H. sapiens*, *B. floridae* and *C. intestinalis*. Schematic representation of HEAT repeats is drawn to scale. The repeats are indicated as black (high REP scores) and white arrows (low REP scores). The amino acid position of each HEAT repeat is also shown. The sequences relating to the HEAT repeats and their corresponding positions following multiple alignment in the three organisms are shown in Tables 1 and 2.
served positions, including 43 completely conserved introns (dashed lines outlined in black in Figure 5) and eight that are not in exactly the same position but have slipped of 4–18 bp (dashed lines outlined in blue in Figure 5). In addition, there are 38 orthologous exons in which the predicted amino acid sequences from amphibian and H. sapiens can be aligned over the entire length: 14 of different lengths and 24 of identical length (respectively shown as green and blue boxes in Figure 5). The other exons are grouped in 14 exon-clusters defined as groups of exons delimited by introns that are positionally conserved or which have slipped by 4–18 bp (dashed lines).

### Table 2: HEAT repeat consensus sequence

<table>
<thead>
<tr>
<th>consensus</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. sapiens</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>B. florideae</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>1</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>B. florideae</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>2</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>B. florideae</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>3</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>B. florideae</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>4</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>B. florideae</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>5</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>B. florideae</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>6</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>B. florideae</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>7</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>B. florideae</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>8</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>B. florideae</td>
<td>HEAT_AAA</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>HEAT_AAA</td>
</tr>
</tbody>
</table>

The consensus sequence of the most conserved HEAT repeats in H. sapiens, B. florideae and C. intestinalis huntingtin are shown.
lines outlined in blue in Figure 5), five of which are indicated as red box when the number of exons is the same in both species; whereas the other five exon-clusters have more exons in Homo sapiens than in amphioxus (grey boxes), and the remaining four show the opposite situation (yellow boxes) (Figure 5). Otherwise, Ciona has 23 orthologous exons (black boxes) and 17 exon-clusters (pink boxes) (Figure 5).

The AmphiHtt gene has a highly conserved distribution of exons and introns with respect to the human sequence (Figure 5), and a length range of 60–457 bp that does not substantially differ from that of the human gene (48–341 bp). The exon/intron splice sites in AmphiHtt correspond to the expected GT-AG intron consensus splicing sequences; the intron phases in AmphiHtt are identical to those in the human gene with the exception of intron 36...
and the majority of introns are in phase 0. In particular, the AmphiHtt gene shares with the human huntingtin gene 28 phase 0, 10 phase 1 and 12 phase 2 introns, thus indicating a more conserved gene structure than the Ciona gene homologue, that has 18 phase 0, 8 phase 1 and 6 phase 2 introns (Figure 5).

In conclusion, on the basis of such analysis, we found that: i) the exon-intron organisation of the huntingtin gene is remarkably conserved in the phylum Chordata, as both amphioxus and Ciona huntingtin genes have a very similar genomic organisation to that of other vertebrate species; ii) at least four reduction events in exon numbers (yellow exon-clusters) occurred between the amphioxus and human genes, which are preferentially located at the 3’ end and in the central region of the gene, whereas greater exon acquisition has mainly occurred in the 5’ end. This confirms previous observations that the greater exon acquisition corresponds to a larger difference in the N-terminal part of the protein between human and Ciona huntingtin [9]. However, the genomic organisation of the amphioxus huntingtin gene is more similar to that of vertebrates than to Ciona, including the larger number of positionally conserved introns (51 in amphioxus against 39 in Ciona), the smaller number of exon-clusters (14 in amphioxus against 17 in Ciona), and the conservation of intron phases. These differences could be also explained by a high evolutionary rate such as that observed in tunicate species.

**AmphiHtt expression in amphioxus**

Analyses of huntingtin expression in vertebrates have provided limited information concerning its potential physiological function as the protein is expressed ubiquitously and throughout the entire life of a vertebrate. A first attempt to evaluate huntingtin distribution in an invertebrate organism was made by Kauffman et al. [11], and we tested the expression of AmphiHtt mRNA during amphioxus development. Amphioxus and vertebrates share anatomical features such as a dorsal nerve cord, a notochord, segmentally arranged muscles (myomeres), pharyngeal gill slits and a post-anal tail (see Figure 6).

We performed whole mount in situ hybridisation on B. floridae developmental stages of 0–10 hours, 11-hour early neurula, 15-hour late neurula, 18-hour late neurula, 24-hour early larva, and 48-hour larva. In order to increase our confidence in the results (the presence of possible alternative transcripts that may be differentially expressed and that are not identified in the present work) we used three different >1000 bp probes mapping to the 5’, central and 3’ portions of the messengers. We obtained the same results using both mixed probes and one probe at a time in separate experiments.

No detectable transcripts of AmphiHtt were found between fertilisation and gastrula stage (0–10 hrs) (data not shown). The first visible expression was found at the most anterior neural plate of 11-hour early neurula (Figure 7A). At this stage, the AmphiHtt transcripts are mainly located at the anterior tip of the neural plate (Figures 7A and 7B), and in some more posterior cells at the neural plate borders (Figures 7A and 7C). As neurulation proceeds and the neural tube forms by the dorsal folding of the lateral edges of the neural plate, AmphiHtt expression extends along the antero-posterior axis of 15-hour neurula. At this stage, neural expression is found in the entire cerebral vesicle and in the most anterior two-thirds of the neural tube (Figure 7D).

In order to reveal differences in dorso-ventral distribution, we used cross-sections of the same embryo (15-hour neurula) and found transcripts in some ventrolateral (Figures 7E–G) and dorsolateral cells of the cerebral vesicle (Figure 7F), at the level of the precursor of the frontal eye complex and the infundibular organ. More posteriorly, we found labelled ventrolateral nerve cells of the hindbrain (Figures 7H–I), most of which consisted of paired neural cells located ventrolaterally in the neural tube and may correspond to differentiating DC motoneurons that innervate the dorsal compartment of the myomeres [23,24]. Furthermore, at the 15-hour stage (early neurula), non-neural expression appears in some endodermal cells of the tail bud around the neuroenteric canal (Figures 7D and 7K), and in some cells of Hatschek’s left diverticulum (Figures 7D and 7G). Neural tube expression is strongly maintained in late neurulae (18 hours), and new labelling appears in individual somite cells, which were only detected at the 18-hour late neurula stage, mainly confined between somite 3 and somite 10, and sometimes arranged as a row of cells at the most lateral margins of the somites, just near the epidermic layer (Figures 7N–S). At the 24-hour early larval stage, the expression was found in the neural tube (Figures 8A–H), being localised to some cells of the cerebral vesicle (Figures 8A–E) and cells of the most anterior two-thirds of the neural tube (Figures 8A,B,F,G). AmphiHtt-expressing cells can also be seen in the ventro-lateral position of the neural tube, just behind the first pigment spot (Figure 8H1). This pattern of expression is essentially maintained in the later stages of development (48-hour larva) (Figures 8I–P), but the highest expression of AmphiHtt mRNA is found at the level of the cerebral vesicle (Figures 8J–M).

In conclusion, our findings demonstrate that during amphioxus development huntingtin transcripts are detected into the neuronal compartment starting from early neurula stage. Except for endodermal (tail bud) and mesodermal structures (Hatschek’s left diverticulum and somites) (i.e. non neuronal cells in the 15- to 18-hour...
stages), the expression of AmphiHtt mRNA follows an antero-posterior gradient, and is enriched in the anterior neural tube. This result could reflect a specific neuronal function of huntingtin in the middle and later developmental stages of this non-vertebrate organism. Nevertheless, we cannot exclude that in situ hybridization, being a relatively insensitive technique, does not allow us to detect low levels of messengers in the early developmental stages and in non nervous structures of amphioxus larvae.

**Discussion**

In order to increase our understanding of normal huntingtin function(s) and reconstruct polyQ evolution along the deuterostome branch as an indication of possible protein activity, we have recently concentrated our study on the cloning and comparative analysis of non-vertebrate deuterostome homologues, such as ascidians [10], echinoderms (Tartari et al., unpublished) and amphioxus. Amphioxus shares the full suite of chordate characteristics with vertebrates and the nerve cord has dorso-ventral specialisation, but they lack the vertebrate typically extensive subcellular and tissue specialisation of the nervous system. At genetic level, amphioxus did not undergo the extensive gene duplication events that characterise vertebrate genomes [25,26], possibly lacking the newly-acquired gene innovation of vertebrates. All of these characteristics make this organism particularly useful to infer
features that were already present in the last common ancestor of chordates.

Along the deuterostome branch, the recently cloned ascidian huntingtin homologue [10], whose sequence conservation is greater than that of Drosophila, suggests a more recent evolution of the 5' end of the gene, which is also characterised by the lack of a polyQ tract. Only partial sequences are available from other invertebrates in the deuterostome branch, such the tunicate Halocynthia roretzi, and from two echinoderms, Strongylocentrotus purpuratus and Heliocidaris erythrogramma, for which no extensive details of sequence conservation are known [11].

Huntingtin cloning from amphioxus allowed us to discover an evolutionarily ancient point of emergence of the polyQ tract similar to that characterising huntingtin in vertebrates. AmphiHtt protein has two glutamines in the same position of the polyQ tract as that characterising the entire vertebrate subphylum. This indicates that a double Q was present in last common ancestor between cephalo-chordates and vertebrates, and that Ciona has differently and subsequently lost this characteristic.

A further biochemical indication of the possible molecular activity of the protein is the presence of HEAT repeat consensas. The AmphiHtt protein has 11 HEAT repeats, thus falling between the 8 of C. intestinalis and the 17 of human huntingtin. Our analyses identified the most conserved HEATs (the first three in the N-terminus, the ninth and eleventh in the central region, and the fifteenth in the C-terminus) in a homologue that precedes vertebrate genome duplication. Although this does not yet allow us to confirm a HEAT repeat-dependent evolutionary trend in huntingtin, or the impact of these sequences on protein function, we can report the strong maintenance of HEATs at the extreme N-terminus.
With respect to human huntingtin, conservation of the primary amino acid sequence in the remaining amphioxus protein is greater than that in C. intestinalis, and comparison of the gene structure of AmphiHtt and the human and ascidian homologues shows that the gene and exon boundaries are more conserved with respect to vertebrates than the ascidian huntingtin gene. This suggests that amphioxus huntingtin is closer to, and less divergent from vertebrate huntingtin than ascidian huntingtin, and leads us to hypothesise that its function is possibly also closer to that of vertebrate huntingtin.

In particular, as exon acquisition events are mainly located in the 5' portion of the gene, whereas the extreme N-terminal portion of AmphiHtt is highly conserved at protein level, we suggest that huntingtin refined its possible N-terminal corresponding function in the evolutionary transition between cephalochordates and vertebrates, and we postulate that this function can be linked to the emergence of a role of huntingtin in the nervous system (at least during development) as amphioxus huntingtin messenger RNA is enriched in neuronal tissues.

First expression analysis on deuterostome invertebrates the echinoderm H. erythrogramma and the ascidian H. roretzi [11], suggests an ubiquitous expression of huntingtin mRNA at all developmental stages of ascidian (as in vertebrates), and a non-neuronal signal in echinoderms. Moreover, by RT-PCR Drosophila huntingtin transcripts were found in all developmental stages [9].

The complete cloning of AmphiHtt also allowed us to analyse its expression in the embryonic and larval stages of amphioxus, and may help in inferring hypothesis on possible huntingtin function.

The pattern of huntingtin expression in amphioxus substantially differs from that at the corresponding stages of...
vertebrate development. It is mainly limited to the neuronal compartments from 11-hour neurula to 48-hour larva, and is not detectable at the early developmental stages until the gastrula stage, whereas it seems to be widely expressed in vertebrates. In humans, rodents and pigs, huntingtin is ubiquitously expressed but has the highest levels in brain and testis, followed by lung, heart, kidney and liver. Even lower vertebrates (fish) seem to express huntingtin at all developmental stages and in all tissues, particularly in the head of adults [12].

Huntingtin does not seem to be expressed until the end of gastrulation in amphioxus. Although this finding cannot exclude the possibility of simply undetectable low-messenger levels, it is possible that, unlike in vertebrates, huntingtin may not be required for gastrulation in this organism. Mammalian data indicate that huntingtin is required at different stages of development, and that its total absence causes embryo lethality at the gastrulation stage [27-29]. However, amphioxus embryos differ from mammalian embryos in early gastrulation insofar as they have a double-layered gastrula (ectoderm and mesendoderm) instead of the three-layered vertebrate gastrula (ectoderm, mesoderm and endoderm) [30].

As mammalian embryogenesis proceeds, huntingtin is required for epiblast formation and neurogenesis [31]. Finally, the removal of huntingtin from post-natal neurons causes cell death, which indicates that mammalian huntingtin plays an important role in nervous system formation and neuronal survival in adulthood. In any case, the expression of mammalian huntingtin is always ubiquitous at all these stages of development.

Closer analysis of amphioxus huntingtin expression in the nervous system shows an initial homogeneous localisation in the most anterior two-thirds of the neural tube, where the signal seems to become more intense at later stages of development [48 hours], thus indicating a possible antero-posterior gradient. This suggests that amphioxus huntingtin may play a role in events occurring at the time of neurogenesis.

In addition, serial cross-sections of whole-mounted labelled amphioxus embryos, showing dorso-ventral views of specific neural tube regions, revealed the presence of huntingtin throughout the most anterior cerebral vesicle, whereas it was restricted ventrally to the posterior cerebral vesicle. Moving caudally, huntingtin specifically marks some paired ventro-lateral cells in the hindbrain that can be assumed to be dorsal compartment (DC) motor neurons and, even more caudally (after the first pigment spot), huntingtin transcripts preferentially localise dorso-laterally in the neural tube. This is the first evidence of the preferential sub-regionalisation of huntingtin expression in the nervous system.

**Conclusion**

We have recently hypothesised that the different functions of huntingtin during mammalian development may possibly reflect evolutionary steps in the protein and that its early non-neuronal activity in mammals can be likened to its ancestral function in species with a poorly organised or no nervous system [5]. In this study, we found that the sequence of amphioxus huntingtin is not critically different from that of vertebrates, and that its expression is particularly enriched in the nervous system. In this view, it can be inferred that an ancestral neuronal function of huntingtin was present 540 millions years ago. The differences in the length of the polyQ tract between amphioxus and vertebrates suggest that the function of huntingtin may have evolved different biochemical properties in both lineages. In particular, we argue that the domain(s) involved in these ancestral function(s) are positioned in the extreme N-terminal portion as the protein’s primary sequence and the consensus of secondary structures (HEAT repeats) are highly conserved with respect to vertebrate huntingtin, and because the corresponding 5’ portion of the gene seems to be due to more recent evolution.

**Methods**

**Animal collection and RNA preparation**

Ripe specimens of the Florida amphioxus (*Branchiostoma floridae*) were collected in Old Tampa Bay, FL. Animals were induced to spawn by electric stimulation. Eggs obtained from electrically stimulated females were fertilized, and the developmental stages were raised in laboratory culture. Adult specimens were harvested and immediately submerged in RNA later (Ambion Europe Ltd., UK). Total RNA from a single adult was extracted using the TRIzol LS reagent (Invitrogen, San Diego, CA). Following extraction, RNA was treated with RNase-free DNase I (Ambion Europe Ltd., UK) according to the manufacturer’s recommendations in order to digest contaminating genomic DNA. First-strand cDNA was synthesised with 5 μg RNA using the SuperScript first-strand synthesis system (Invitrogen, San Diego, CA) and oligo(dT) primers.

**Retrieving sequence from the Branchiostoma floridae genome**

The *B. floridae* genome assembly (v1.0) was searched at JGI [32] using the TblastN algorithm and several vertebrate huntingtin protein sequences as queries. The identified sequences were analysed by means of two gene prediction programs (GenomeScan [33], GENSCAN [34]) in order to correct the preliminary annotation reported at JGI (Protein ID: 101261, 101262 and 252341). Then, a predicted coding sequence for the amphioxus huntingtin
gene was used to define the PCR amplification strategy (Figure 1). Finally, we reconstructed the genomic organisation of the amphioxus huntingtin gene using two genomic mapping software programs: GMAP [35] and Wise2 [36], which respectively re-align messengers and protein to genomic sequences.

Cloning of AmphiHtt mRNA
The resulting first-strand cDNA of adult amphioxus B. floridai and a 5- to 24-hour B. floridai embryo cDNA library (kindly provided by Jim Langeland) were used in PCR assays with specific primers designed on the basis of the predicted coding sequence. PCRs were carried out in a 50 μl reaction mixture using the Hot Master mix in accordance with the manufacturer’s instructions (Eppendorf Srl, Italy) and the primers specified in Additional file 1. The PCR products (Figure 1) were directly cloned using a TOPO TA cloning kit (Invitrogen, San Diego, CA). Ten clones for each amplified fragment were randomly chosen for automated sequencing using a 377 PerkinElmer sequencer and the universal or internal sequence-specific primers.

Sequence and phylogenetic analysis
The Vector NTI Suite (version 9.0, Informax, North Bethesda, MD) software package was used for sequence analysis. Multiple sequence alignments were carried out using the huntingtin sequences from Homo sapiens (P42858), Mus musculus (P42859), Rattus norvegicus (P51111), Sus scrofa (BA36752), Danio rerio (AAC63983), Fugu rubripes (P51112), Tetraodon nigroviridis (CAG03293), and Ciona intestinalis (AM162277). We also used the sequences of Xenopus tropicalis, Gallus gallus and Ciona savignyi predicted from genomic sequences by Gissi et al. [10]. The amino acid sequences from amphioxus and eleven other species were aligned using the CLUSTAL W program [37] and manually corrected. Amino acid sites with gaps in any sequence were excluded, and so a total of 2491 characters were considered for the analysis. The best-fitting model of evolution (JTT, with an estimated alpha parameter to 0.73 and a gamma distribution of rates between sites of 4.0) was inferred by means of the ProtTest [38]. Phylogenetic analysis was performed using a fast and accurate maximum likelihood heuristic method (PHYML v2.4.4) [39] starting from the BIONJ tree, under the parameters estimated by ProtTest. Tree stability was assessed by means of a bootstrap analysis with 100 cycles. Phylogenetic analysis was also performed by CLUSTAL W program and MEGA version 3.1 [40] (Additional file 3). The tree was produced using the neighboring method with Poisson correction and complete deletion of gaps and bootstrapped 1000 times. Such tree was rooted using the huntingtin from D. melanogaster (AF146362) as the outgroup. The phylogenetic trees were visualised using TREEVIEW. The sequence data were also analyzed using the MEGALIGN program from LASERGENE (DNASTAR, Madison, WI) in order to evaluate sequence similarities in the huntingtin proteins (Figure 2A).

HEAT repeat evolution analysis
HEAT repeat consensus were found by searching for the HEAT option with the REP program [41], and loading the individual human, amphioxus and Ciona intestinalis huntingtin amino acid sequences. The resulting highly scored consensus were listed, and additional human consensus previously published by Gissi et al. [10] were added to the list. By applying the REP program and searching the all consensus option, additional low-score HEAT consensus were found in the three sequences. Therefore, we considered and listed only those corresponding to our multiple alignment (Additional file 2, Tables 1 and 2).

Whole mount in situ hybridisation
To obtain riboprobes for whole mount in situ hybridisation, PCR was performed using adult amphioxus cDNA as a template and the HttA_F and HttA_R primers (Figure 1 and Additional file 1, probe A). The resulting 1345-bp fragment was subcloned into the pCR II TOPO vector and Additional file 1, probe C). Both sense and antisense RNA probes were generated using a digoxigenin (DIG) RNA labelling kit (Roche Diagnostics, Canada) in accordance with the manufacturer’s instructions. In order to detect AmphiHtt mRNA the probes were used singly and mixed in different experiments. The in situ hybridisation experiments were performed at different developmental stages from fertilisation to 48-hour larvae according to Holland et al. [42]. Labelled whole mount embryos were photographed using an Olympus IX71 microscope (Olympus Italia s.r.l., Italy), and then counterstained with 1% Ponceau S in 1% acetic acid, dehydrated in ethanol, embedded in Spurr’s resin, and serially sectioned at 3–4 μm. Moreover, samples were also examined without Ponceau S counterstaining in order to avoid masking of possible low-level signals. The signal was identical using single or mixed probes, but we only show the results obtained with the mixed probes because the signal was more intense. Negative control experiments were done using sense riboprobes and no specific signal was obtained (Additional file 5).

Authors’ contributions
SC carried out the bioinformatic and molecular analysis and in situ hybridization assays. MT participated to the bioinformatic analyses. SC and MT draft the manuscript. EC and MP provided technical assistance, supervised the
Additional material

Additional file 1
PCR primers. PCR primers used to isolate a contig of the AmphiHtt cDNA sequence. Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-213X-7-127-S1.pdf]

Additional file 2
Huntingtin amino acid alignment. Alignment in fasta format of the huntingtin proteins from chordates. Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-213X-7-127-S2.txt]

Additional file 3
Rooted tree of huntingtin. Phylogenetic tree created using the neighbor-joining method with Drosophila melanogaster huntingtin as the out-group. Numbers close to the nodes are percentage values represent 1000 bootstrapping. The scale bar of 0.2 at the bottom left corner of the tree indicates 0.2 substitutions for the site. Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-213X-7-127-S3.pdf]

Additional file 4
Exon size of chordate huntingtin genes. Size of protein-coding exons is indicated in basepairs (bp). For amphioxus both scaffolds (scf378 and scf613) are shown. Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-213X-7-127-S4.xls]

Additional file 5
ISH control. ISH control experiments on amphioxus embryos using sense riboprobes. Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-213X-7-127-S5.pdf]

Acknowledgements
We would like to thank Skip Pierce and John M. Lawrence (Department of Biology, USF, Tampa, FL) for the use of laboratory space and equipment; Linda Holland (Scripps Institution of Oceanography, La Jolla, CA) for her comments and criticisms; Jim Langland (Department of Biology, Kalamazoo College, Kalamazoo, MI) for providing the cDNA library; Ray Martinez and Marilyn Wetzel (Department of Biology, USF, Tampa, FL) for their logistic support. We are also grateful to Eileen Donovan-Wright (Dalhousie University, Halifax, Canada), and the anonymous reviewers for their very useful criticisms and suggestions. This research was supported by Fondazione Telethon (GGP06250) and PRIN 2006 (2006052993) to EC.

References
1. Huntington’s Disease Collaborative Research Group: A novel gene containing a trinucleotide repeat that is expanded and unsta-
3. Perutz MF, Johnson T, Suzuki M, Finch JT: Glutamine repeats as polar zippers: their possible role in inherited neurodegener-
7. Neuwald AF, Hirano T: HEAT repeats associated with con-
tebrates, cohesins, complexes involved in chromosome-
9. Li Z, Karlovich CA, Fish MP, Scott MP, Myers RM: A putative Dro-
12. Karlovich CA, John RM, Ramirez L, Stainer DTR, Myers RM: Char-
14. Cobb JLS: The significance of the radial nerve cords in Astero-
15. Cvey MJ, Markel K: Echinodina. In Microscopic Anatomy of Inverte-
brates Volume 14. Edited by: Harrison FW, Chia FS. New York: Wiley-
Liss; 1994:345-400.
audella T, Leavitt BR, Hayden MR, Timmusk T, Rigamonti D, Cattan-
18. Delsec F, Brinkmann H, Chourrout D, Philippe H: Tunicates and not cephalochordates are the closest living relatives of verte-
schner M, Lander ES, Thorniedy M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ: Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xeno-
22. Tandem repeats Finder program [http://tandem.bu.edu/trf/]
23. Lacalli TC, Kelly SJ: Somatic motoneurons in the anterior nerve cord of amphioxus larvae: cell types, cell position and innera-


32. B. floridae genome assembly (v1.0) [http://genome.jgi-psf.org/Brfl1/Brfl1.home.html]

33. GenomeScan [http://genes.mit.edu/GENESCAN.html]

34. GENSCAN [http://genes.mit.edu/GENSCAN.html]

35. GMAP [http://www.gene.com/share/gmap/]

36. Wise2 [http://www.ebi.ac.uk/Wise2/]


41. REP program [http://www.embl-heidelberg.de/andrade/papers/rep/search.html]


