

RESEARCH ARTICLE

Open Access

Origins of amyloid- β

William G Tharp^{1,2} and Indra Neil Sarkar^{1,3,4*}

Abstract

Background: Amyloid- β plaques are a defining characteristic of Alzheimer Disease. However, Amyloid- β deposition is also found in other forms of dementia and in non-pathological contexts. Amyloid- β deposition is variable among vertebrate species and the evolutionary emergence of the amyloidogenic property is currently unknown. Evolutionary persistence of a pathological peptide sequence may depend on the functions of the precursor gene, conservation or mutation of nucleotides or peptide domains within the precursor gene, or a species-specific physiological environment.

Results: In this study, we asked when amyloidogenic Amyloid- β first arose using phylogenetic trees constructed for the Amyloid- β Precursor Protein gene family and by modeling the potential for Amyloid- β aggregation across species *in silico*. We collected the most comprehensive set of sequences for the Amyloid- β Precursor Protein family using an automated, iterative meta-database search and constructed a highly resolved phylogeny. The analysis revealed that the ancestral gene for invertebrate and vertebrate Amyloid- β Precursor Protein gene families arose around metazoic speciation during the Ediacaran period. Synapomorphic frequencies found domain-specific conservation of sequence. Analyses of aggregation potential showed that potentially amyloidogenic sequences are a ubiquitous feature of vertebrate Amyloid- β Precursor Protein but are also found in echinoderm, nematode, and cephalochordate, and hymenoptera species homologues.

Conclusions: The Amyloid- β Precursor Protein gene is ancient and highly conserved. The amyloid forming Amyloid- β domains may have been present in early deuterostomes, but more recent mutations appear to have resulted in potentially unrelated amyloid forming sequences. Our results further highlight that the species-specific physiological environment is as critical to Amyloid- β formation as the peptide sequence.

Keywords: Amyloid, Alzheimer disease, Phylogenetics, In silico, Aggregation, Maximum parsimony, Bayesian inference

Background

The Amyloid- β Precursor Protein (A β PP, APP) has been intensively studied due to its role in the generation of pathogenic cortical plaques in Alzheimer Disease [1]. It belongs to a gene family with deep evolutionary origins and is a member of a highly conserved protein family of type-1 transmembrane proteins [2-4]. The A β PP family consists of up to three homologues in vertebrate species: A β PP, amyloid precursor like protein 1 (APLP-1), and amyloid precursor like protein 2 (APLP-2). Invertebrate species genomes encode a single homologue referred to

as either amyloid precursor like 1 protein (APL-1) or A β PP-like 1 protein (APPL-1). Vertebrates and flatworms exhibit ubiquitous expression of at least one member of the A β PP family, while fruit flies express APPL-1 only in neurons. In all species, A β PP proteins are cleaved into multiple peptides and fragments by a series of proteases, but only vertebrate A β PP contains the sequence coding the pathological Amyloid- β (A β) peptide fragment.

The β -fold intrinsic to amyloid formation is a commonly observed biochemical property [5-7]. Amyloid formation is observed in non-pathological contexts from an efficient steric mechanism for storage of small peptide hormones to rudimentary forms of biological compartmentalization [5,8]. The neuropathological changes observed in the brains of patients with Alzheimer Disease led to the formation of the Amyloid

* Correspondence: neil.sarkar@uvm.edu

¹Center for Clinical and Translational Science, University of Vermont, Given Courtyard N309, 89 Beaumont Avenue, Burlington, VT 05405, USA

³Department of Microbiology and Molecular Genetics, University of Vermont, Given Courtyard N309, 89 Beaumont Avenue, Burlington, VT 05405, USA
Full list of author information is available at the end of the article

Hypothesis, which implicates both extracellular deposits of A β fibrils and low-order intracellular A β oligomers in the disruption of neuronal function, distortion of neural architecture, and induction of inflammation [1].

Mutations in the A β PP sequence and in associated proteases have been independently associated with familial early onset Alzheimer Disease characterized by rapidly progressive dementia and heavy A β plaque burden [9]. Recently, a protective mutation in A β PP reducing the formation of A β was identified [10]. However, >95% of sporadic Alzheimer Disease exhibits no mutation in the A β PP gene sequence. Further, deposition of A β is not limited to Alzheimer Disease. A β plaques have been observed in vascular dementias, Lewy body dementia, and Parkinson Disease with dementia, as well as in the brains of aged individuals without any cognitive deficits [11-14]. Together, these studies indicate that while the sequence of A β can contribute to the progression and severity of disease factors regulating the production by proteolysis and the degradation and clearance of A β , it also plays a critical role in generation of A β pathology.

Beyond the eponymous production of A β , A β PP processing produces other active peptides with functions ranging from hemostatic modulators to trophic factors to pro-apoptotic proteins [15-18]. There is a substantial body of knowledge focusing on the neural impacts of A β PP and A β . However, this family of proteins is also widely expressed in peripheral tissues of vertebrate species including skin, skeletal muscle, leukocytes, platelets, intestinal epithelia, pancreas, and adipose tissue. The function and regulation of non-neuronal A β PP are not fully understood [19-26].

The A β PP family is variably essential for viability among species. Experimental data show that the N-terminus of APL-1 is necessary for progression through molting stages by nematodes [27]. The C-terminus of at least one member of the A β PP family is necessary for viability in early parturition of knockout mouse models [28-30]. *Drosophila* models without APPL-1 show subtle neuronal patterning defects but remain viable and able to reproduce [31]. Zebrafish knockout models have impaired body development and synaptogenesis [32,33]. Each of these models can be rescued by expression of truncated portions of A β PPs, indicating that absences of different domains are responsible for the observed lethality or defects in each model. Thus, the persistence of this protein family appears domain-dependent among species despite high evolutionary conservation of the entire gene.

The major conserved regions found in A β PP family proteins include two ectoplasmic domains (E1 and E2), which contain extracellular matrix and divalent cation binding regions and a growth-factor-like domain (GFLD), and the cytoplasmic region (E3) that contains a

basolateral sorting signal (BLS) and an NPXY internalization sequence (YENPTY) (Figure 1). The corresponding nucleic acids coding the domains are termed D1, D2, and D3, respectively. Other important conserved domains include a Kunitz-protease inhibitor (KPI) domain found only in vertebrate A β PP and APLP-2, as well as the β A4 region that gives rise to A β in certain vertebrate species. Interestingly, it has been demonstrated that the corresponding region of APPL-1 in *Drosophila melanogaster* can form amyloid deposits when co-expressed in high levels with the *Drosophila* β -secretase [34]. Similar to the conservation of the E1, E2, and E3 domains, the β A4 region and corresponding regions may have arisen from a common ancestral domain [35]. It is not known when the amyloidogenic trait first appeared in this gene family nor why species with nearly identical A β sequences do not develop A β deposits.

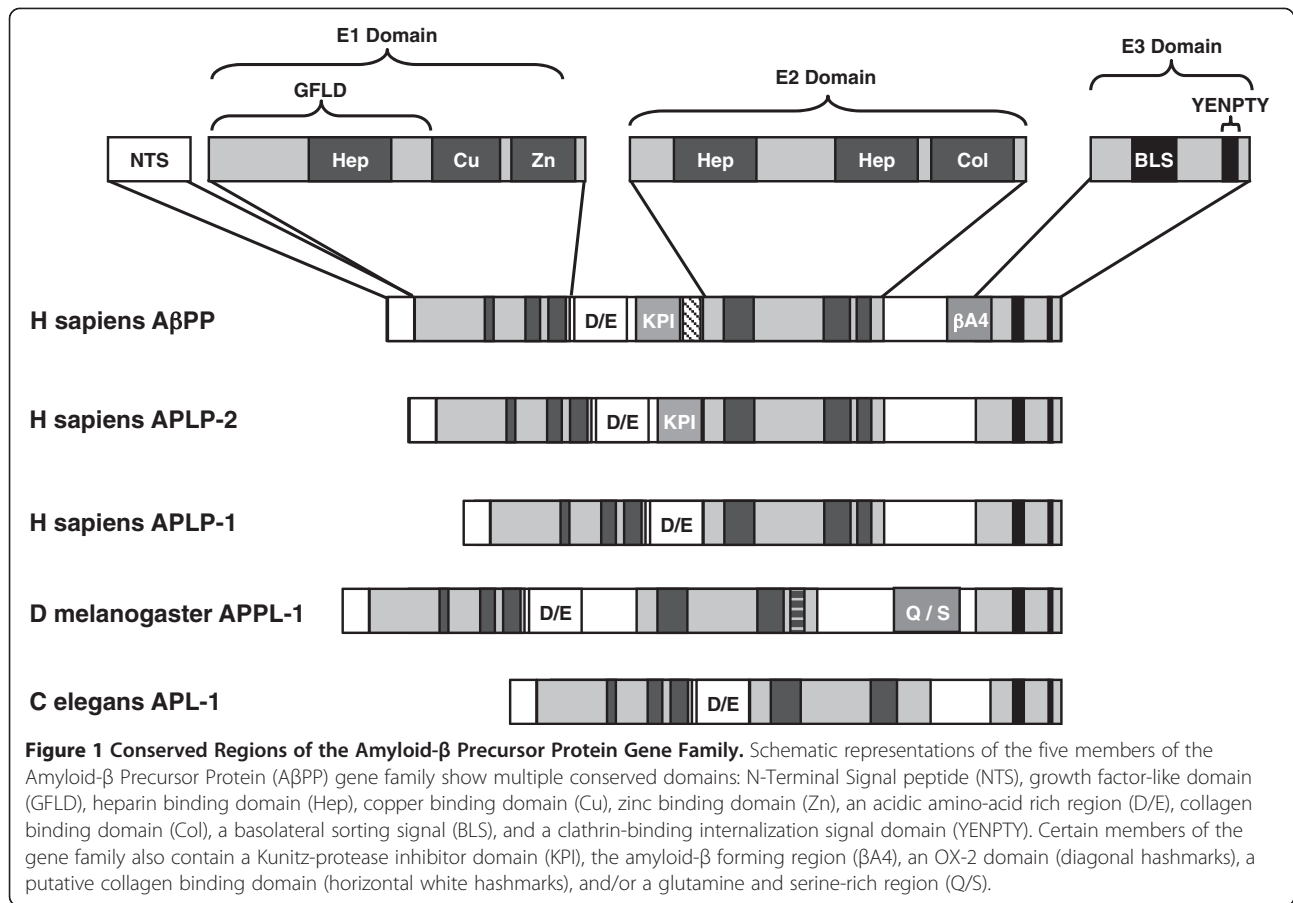
Previous phylogenetic studies showed that this ancient protein family has been widely distributed among multicellular eukaryotes since at least the divergence of protostomia and deuterostomia [36]. These studies and corresponding conclusions are based on at most ten sequences that were trimmed and concatenated, focusing solely on the major conserved domains (D1, D2, and D3 only). Use of trimmed sequences does yield cleaner sequence alignments and better branch supports on the phylogenetic tree, but ignores potentially valuable evolutionary data encoded in adjoining regions. For the A β PP gene family in particular, the omission of A β from the analyses occludes an understanding of the evolution of the pathological eponymous domain. Despite wide distribution of the A β PP family across species, it is not known when amyloidogenic A β peptides first evolved. This study uses the full complement of available molecular sequence data to provide an *in silico* model of the evolutionary history of this essential gene family and the origin of the A β peptide.

Results

Phylogenetics of A β PP gene family

Amino acid and nucleotide sequences were collected using an automated, iterative search method from *Entrez* Protein (GenPept) and *Entrez* Nucleotide (GenBank) (see Methods and Additional file 1: Table S1). Character matrices were generated in Mesquite 2.7.5 and aligned using Muscle 3.8.31) and the longest sequence for each species' homologue(s) was retained [37,38]. Amino acid and nucleotide trees were generated under maximum parsimony using TNT 1.1 and Bayesian inference using MrBayes 3.2 [39,40].

The overall topology of the nucleotide and amino acid trees is similar among trees generated by maximum parsimony (Figure 2a and b) and Bayesian inference (Additional file 1: Figure S1a and b). Branch support



data for the Bayesian analyses are found in Additional file 1: Figure S1 and support for the maximum parsimony analyses are found in Additional file 1: Figure S2.

The presence of an AβPP-like sequence in hydra (*Hydra magnipapillata*) and sea anemone (*Nematostella vectensis*) genomes suggests that the ancestral gene arose around metazoic divergence in the Ediacaran period, between 630–540 million years ago (Mya). No related sequences from single-celled organisms were found. A single member of the gene family has persisted across invertebrate species with a major divergence around the evolution of arthropods during the Cambrian period giving rise to APPL-1 (~500 Mya). Two gene duplication events occurred during the evolution of vertebrate species. Our search recovered a single AβPP-like gene for the cephalochordate lancelet (*Branchiostomidae floridae*) genome that was more closely related to mollusks and cnidarians than vertebrate sequences. The cartilaginous ray (*Narke japonica*) genome contains a single AβPP gene with high homology to human AβPP. The results indicate that AβPP and APLP-2 genes are present in the zebrafish (*Danio rerio*) but only AβPP was recovered for other members of class Osteichthyes (*Takifugu rubripes*, *Tetraodon fluviatilis*, and *Perca flavescens*). The majority

of tetrapod genomes in this study contained all three members of the vertebrate AβPP gene family. Sequences for all three genes were found for Xenopus species, but APLP-1 sequences were not found for any members of class Aves or Reptilia.

Within the gene family, the nucleotide sequence phylogenetic trees (Figure 2a) indicate that AβPP and APLP-2 are more closely related than APLP-1 and APLP-2. Furthermore, the placement in the nucleotide phylogenetic tree suggests that APLP-1 may be the original vertebrate sequence. However, placement of the AβPP branch on the nucleotide tree is weakly supported under both maximum parsimony (47% resampling support, 12% relative Bremer; see Additional file 1: Figure S2a, b) and Bayesian inference (60% posterior probability; see Additional file 1: Figure S1a). In the amino acid sequence phylogenetic trees, APLP-1 and APLP-2 are more closely related and AβPP appears to be the original vertebrate peptide (Figure 2b). This arrangement has higher support for the placement of AβPP (65% resampling support, 100% relative Bremer support [maximum parsimony] Additional file 1: Figure S2c, d; 100% posterior probability [Bayesian inference]; Additional file 1: Figure S1b).

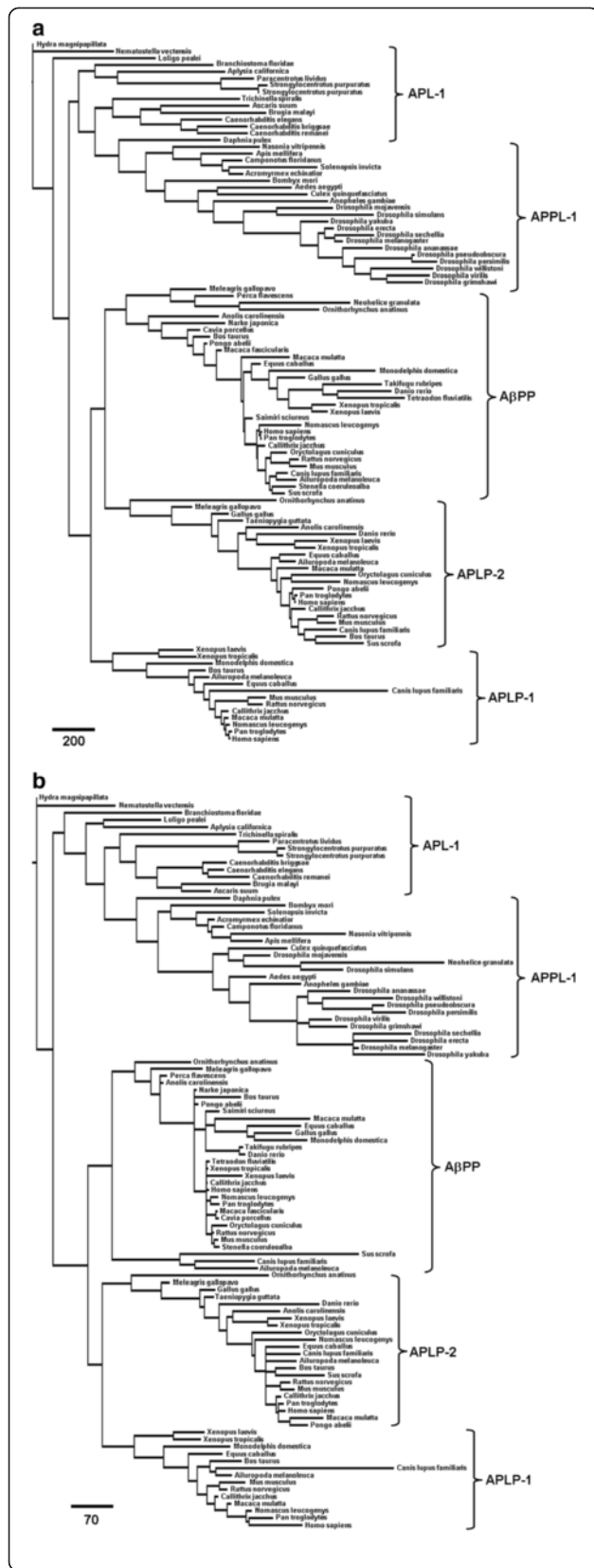


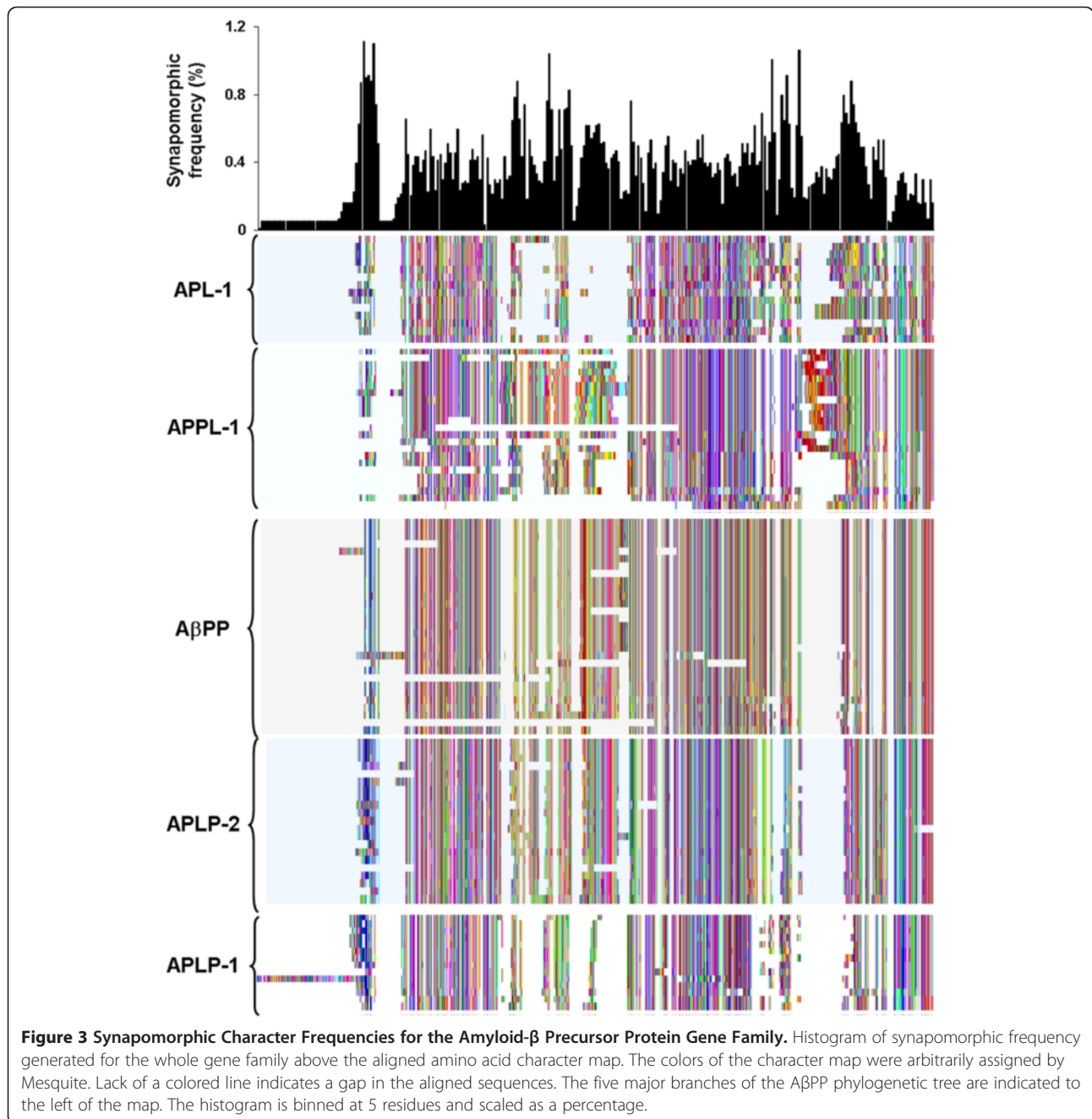
Figure 2 Phylogenetic Relationships of 103 members of the Amyloid-β Precursor Protein Gene Family. Shown are:

(a) Phylogram showing the evolutionary relationships among the nucleotide sequences of the AβPP gene family; (b) Phylogram for the corresponding protein sequences. Trees were generated by maximum parsimony methods. Scale bars indicate character changes contributing to branch lengths.

Persistence of amyloid-β

The variability in the essential nature of the AβPP gene family can be observed by analyzing the evolutionary differences between related genes and shared residues according to specific functional domains. This was accomplished using synapomorphy frequency histograms. A synapomorphy is a trait or character shared by sister taxa of a clade that was derived from a previous common ancestor but not shared by taxa from another clade. Thus, synapomorphies contribute to the topology of a phylogenetic tree as factors in defining nodes on the tree [41]. Using the TNT program we collected synapomorphies present at each node of the consensus amino acid tree and examined the frequency of synapomorphy for each character across the sequence matrix. High frequencies of synapomorphy indicate residue changes at a given position make large contributions to the topology of the phylogenetic tree (conversely, low frequencies on the plots are associated with highly conserved domains/characters present in all terminal taxa on the tree).

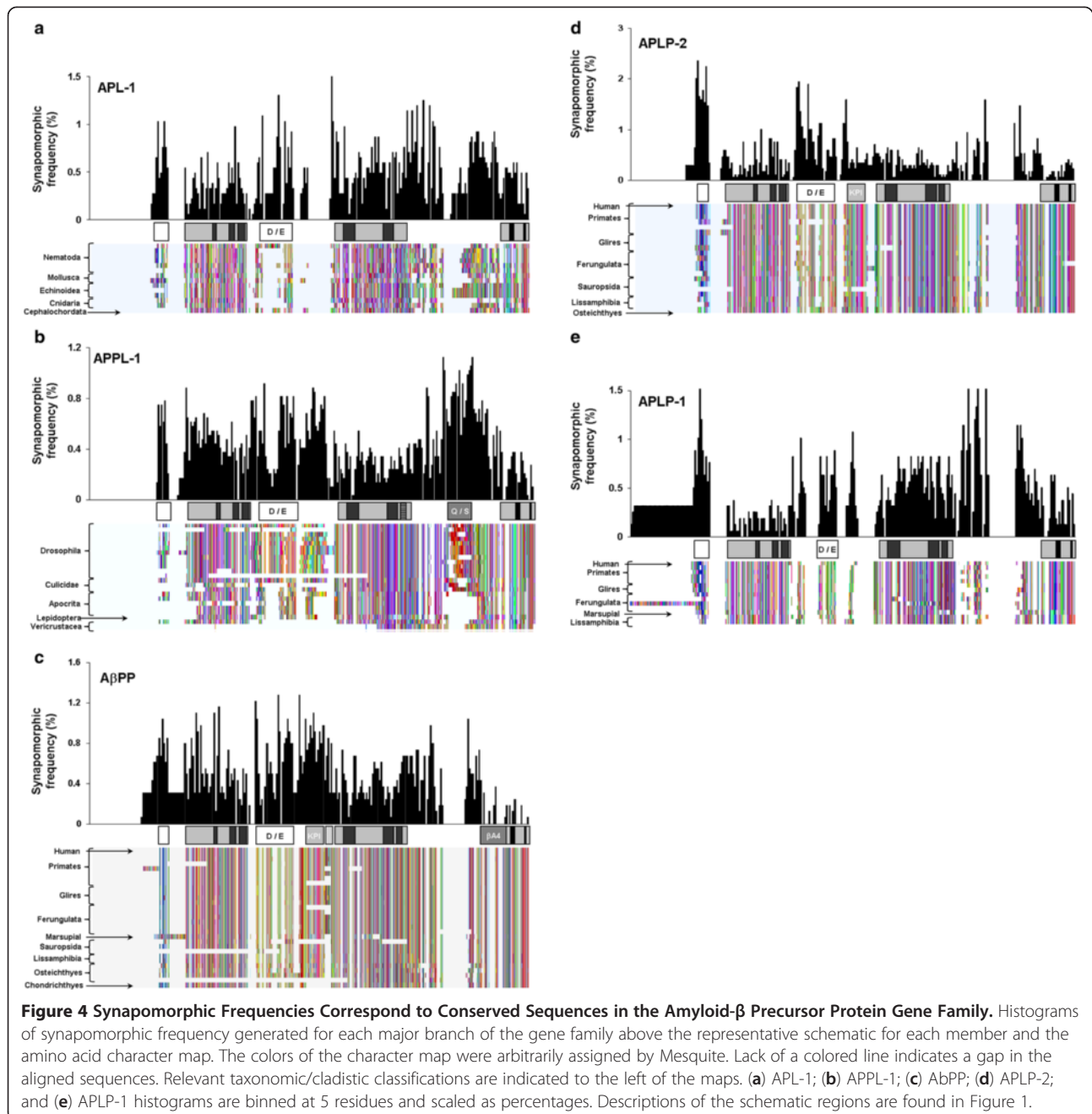
In this study, synapomorphies were first analyzed across all positions in the dataset (Figure 3) and then analyses were stratified by the major branches corresponding to APPL-1, APL-1, AβPP, APLP-2, and APLP-1 (Figure 4, Tables 1 and 2). The most highly variable region was the E2 domain, which accounted for 18.3% of synapomorphies on the whole tree, while the most highly conserved domain was the E3 domain (which had 6.9% of the residue synapomorphies). The E3 region of AβPP, encoded by exons 17 and 18, contributed 1.7% of the synapomorphic frequencies to that branch and 0.3% for the whole tree. By contrast, the same region of the other homologues contributed between 1.2 – 2.4% for the whole tree and 6.3 – 12.5% for each major branch. The βA4 domain, encoded by exons 16 and 17, contributed 0.5% and 3.1% of synapomorphies to the whole tree and AβPP branch, respectively. The exon 16–17 regions of APLP-1 and APLP-2 contributed 1.5% and 1% of synapomorphies to the whole tree, respectively, and 8.8% and 5.4% to their respective branches. These data showed the E3 region to be the most highly conserved part of the entire gene family and conservation of the βA4/E3 region is even stronger for vertebrate species.



Evolutionary relationships of amyloid- β formation potential

Deposition of A β has been well documented in mammals; the sequence is generally >95% identical across mammals and all vertebrates express β - and γ -secretases [42-44]. The Guinea pig rodent (*Cavia porcelus*) and the common hare (*Oryctolagus cuniculus*) have been shown to generate A β plaques, but neither the *Mus musculus* nor *Rattus norvegicus* rodents naturally produce A β plaques [45-48]. Evidence of A β accumulation in other

vertebrate species is sparse. Deposition of extracellular A β has only been documented in one member of class Osteichthyes: *Onchyrus* sockeye salmon; the sockeye A β PP gene has not been sequenced [49]. While some species of birds may generate A β plaques or vascular amyloid deposition, there is no evidence of plaque formation or extracellular deposition in reptiles and amphibians despite >90% sequence homology [47,50]. No natural invertebrate amyloid- β plaques have been documented. Recently it was shown that the corresponding



peptide from *Drosophila* can form an amyloid *in vivo* when co-expressed at high levels with the endogenous β -secretase gene [34].

In order to determine when A β formation first arose in evolution, we modeled β -sheet aggregation and amyloid formation probabilities for sequences corresponding to the human β A4 region using the AmylPred tool and PASTA server, both of which have been designed and validated using A β [51-54]. We found the C-terminal region of β A4 domains to have a high probability to form an amyloid or aggregate for nearly all sequences

(Figures 5, 6, and 7). Only sequence from the silkworm *Bombyx mori* had no amyloidogenic potential using both methods, although AmylPred predicted amyloidogenicity in a short C-terminal region. Very few sequences with amyloidogenic potential were found in the N-terminal region of invertebrate β A4 domains. The cnidarian sea anemone *Nematostella vectensis* and hydra *Hydra magnipapillata* exhibit strong C-terminal amyloid potential but little potential for amyloid formation in their N-terminal β A4 (Figure 6a and b). The nematode *Trichinella spiralis* is the only worm in our dataset with

Table 1 Synapomorphic frequencies within conserved domains

| Domain | Whole tree | APL-1 | APPL-1 | AβPP | APLP-2 | APLP-1 |
|---------------------------|------------|-------|--------|------|--------|--------|
| N-Terminal Signal Peptide | 11.61 | 1.25 | 1.49 | 2.05 | 3.31 | 3.57 |
| E1 | 14.86 | 2.71 | 5.71 | 3.50 | 2.43 | 1.09 |
| E2 | 18.33 | 4.95 | 4.18 | 3.22 | 2.83 | 4.17 |
| E3 | 6.89 | 2.44 | 1.96 | 0.30 | 1.19 | 1.39 |

APL-1: amyloid precursor like 1 protein; APPL-1: AβPP-like 1 protein; AβPP: amyloid-β precursor protein; APLP-2: amyloid precursor protein like protein 2; APLP-1: amyloid precursor protein like protein 1. This table shows the percentage of synapomorphy accounted for by the N-terminal signal peptide and the three major conserved regions of the peptide sequences for the entire phylogenetic tree and the relative contribution for the five major branches.

strongly amyloidogenic βA4 N-terminal sequences predicted by both methods (Figure 6d). The *Neohelice granulata* crab has a short N-terminal amyloid prone region (Figure 6e), but the water flea *Daphnia pulex* does not (Figure 6f). The *Drosophila* flies all express a potentially amyloidogenic N-terminal sequence predicted by AmylPred but with a PASTA energy of - 3.01 (Figure 6g), but no other members of Hymenoptera express amyloid prone sequence at the N-terminus. The squid *Loligo pealei* has PASTA energies < - 4 for long stretches of the N-terminal βA4 region but no consensus support from AmylPred (Figure 6i). The sea slug *Aplysia californica* has a short region with probable amyloid forming potential supported by AmylPred and with a PASTA energy of - 3.16 (Figure 6j). The sea urchins *Strongylocentrotus purpuratus* and *Paracentrotus lividus* also had a short N-terminal region predicted to form amyloid (Figure 6k). The cephalochordate lancelet *Branchiostoma floridae* had two long N-terminal regions with high amyloidogenic potential (Figure 6l). All AβPP sequences in the dataset exhibited a strongly amyloidogenic N-terminal region, though the rodent *M. musculus* and *R. norvegicus* sequences had reduced PASTA energies compared to other vertebrates for their N-terminal regions (Figures 5 and 7). Interestingly, *Danio rerio* APLP2 showed an N-terminal amyloidogenic region (Figure 7e) while all other APLP2 sequences were identical at these residues and had a lower probability of forming an amyloid (Figure 7f). The APLP1 sequences for *Xenopus laevis* and *Monodelphis domestica* showed long sections of

aggregation prone sequence (Figure 7g and h). The remaining APLP1 sequences, representing only placental mammals, show a region with lowered probability of aggregation or fibril formation (Figure 7i).

Discussion

This study provides the most comprehensive phylogeny of the AβPP gene family based on available data to date. The analysis reveals that the ancestral sequence evolved during metazoic divergence, which is much earlier than previously thought. The results further suggest that AβPP itself was the first vertebrate sequence and that APLP-1 and 2 are likely derived from gene duplication of AβPP.

It is possible that the vertebrate gene family arose as a duplication of APLP-1 followed by a second duplication to form APLP-2 and AβPP. However, it is also possible that the original duplication gave rise to APLP-2 and AβPP after which a duplication of APLP-2 gave rise to APLP-1. The search strategy used in this study found APLP-1 sequences only in tetrapods, AβPP in both cartilaginous and bony fish, and APLP-2 in one bony fish and most tetrapods.

We found that the E3 C-terminal region of the protein is essentially unchanged since the divergence of jawed vertebrates during the Ordovician period and that amyloidogenic Aβ was present around this evolutionary step. Its persistence is likely due to the overlap with the E3 domain. It has been shown that the E3 domain is essential for life in mammals and the βA4 domain contains an HD motif with evidence of positive selection, both of which may explain some of the persistence of amyloidogenic Aβ in the mammalian genome [30,35]. Our analysis also found evidence of aggregation prone C-terminal regions in nearly all sequences in the dataset, which is not surprising as this is part of the transmembrane region high in hydrophobic residues, but a stable β-fold requires two regions within the peptide. Sequences with two separate domains capable of forming and stabilizing an amyloid were rare in protostomes, suggesting the characteristic developed after the divergence of deuterostomes and protostomes or was subsequently lost through mutation (Figure 8). Of particular note, the *Drosophila* sequence is predicted to form an

Table 2 Synapomorphic frequencies of domains within each branch

| Domain | APL-1 | APPL-1 | AβPP | APLP-2 | APLP-1 |
|---------------------------|-------|--------|-------|--------|--------|
| N-Terminal Signal Peptide | 6.41 | 4.80 | 11.74 | 18.47 | 21.38 |
| E1 | 13.91 | 18.36 | 20.07 | 13.56 | 6.54 |
| E2 | 25.38 | 13.42 | 18.49 | 15.81 | 24.94 |
| E3 | 12.50 | 6.30 | 1.70 | 6.63 | 8.31 |

APL-1: amyloid precursor like 1 protein; APPL-1: AβPP-like 1 protein; AβPP: amyloid-β precursor protein; APLP-2: amyloid precursor protein like protein 2; APLP-1: amyloid precursor protein like protein 1. This table shows the percentage of synapomorphy accounted for by the N-terminal signal peptide and the three major conserved regions of the peptide sequences within each branch of the phylogenetic trees.

Cnidaria

N Vectensis EEDPENRERD-----FFGE-----PRSPATFAAV-----IGLSC-GALVIM IIVIVAMAMRRSR-----GSSATKTVL

H Magnipapillata IKVOTVDTNIVDPSTTTTFPONLALELKSSDEDVANVVOEPEKNEHRSKRSVLLTAI-----LGLSC-GIVVVM FVVVALIMRQR-----IVKTRVIV

Nematoda

T spiralis KQDIVHD-----FVFOEKSIG-----KSASKYIDHSF-----IFFSI-AGVSLIA AVVCGIFFMRHR-----ANHAGQGFI

B malayi KSVRELHVEIEPIVSEPVRVHDLPE-----PKPSYARHEPLQHFKMSAEDASASMLKSTN-----LLLIF-VSAATV IFIVFATVLRQR-----PRHRGFIE

C elegans SDEDEKNIKELRVDIEPIIDEPAS-----FYRDKLIOSPEVERS-----ASSVOPYVI-----ASAMF-----T PAICIIAFITNAR-----RRRAMR-GFIE

Crustacea

D pulex VDAQVQHMQ-ERPKVAHAQNHQIA-----HNEAGFSVRHE-----SGHRENRSVY-----FTLAF-AGVALI AAMVGLVLLKRRH---ARSPHQGFIE

N granulata RPEPLLVHHAEDVQVAHAQHQIS-----HSQASYVTRTE-----SSGSESRGVY-----FTLAF-AGVALI MAMVGVIVLRRRN---STNPHNQGFME

Neoptera

B mori IRAALEHAEERAPPPAHALKHELQ-----HSQPGYTVRGTGG-----AGSGGAGAL-----YPALCVGGAAL AAACVALAVARRR---DRAPQAQGFVQ

Hymenoptera

C floridanus VVAAMHSEG--VIRAAHSMTHDVS---HSEPGYSVRRE-----VYHRENRSVY-----FTLAF-AGIALMAAAVVAV AIFKRRS---ARSPHQGFIE

A aegypti VAAVAHQEA-E-EPKMQHILSHDTG---HSEPTFSVRREIYG-----PNGRENRSVY-----FTLMF-AFVALMGAVFVGV AVAKWKT---SRSPHQGFVE

D melanogaster VAAVAHQEA-E-PQVQHFMT HDLG---HRESSFSLRREFAQHA---HAAKEGRNSVY-----FTLSF-AGIALMAAVFVGV AVAKWRT---SRSPHQGFIE

Mollusca

β-secretase site

potential γ-secretase sites

L pealei KGYIAHVQQ-----QPMIADEVS---LDN-----LYANSHANSV-----LGIAT-GGVVVB I IIVVAVMLKRR---QRQVTHGFVE

A californica H DYERGSQFI-----AHRMDDKMHVRRQGFAS-----AATSSQMGST-----IGIAL-GGVSVB VIIVVAVFMVRRKN---RNQYPSPGYVE

Echinodermata

S purpuratus IDDELTAPLDSVVKPQTLLIAQKP---KAKVEKSAARMEKQRLVGRQDVNSIDNSVTKKRMSLRVEPMMAFGLAC-GVLA VAVLVIIVLRRT---RRTPVNSGFTE

Cephalochordata

B floridae ADRPEOASIVSQMMNHVWNSCYSS---SSNETROCSCYOBEHYLCLRVFHMHSHTNLIT-----LSLTV-AAATIAV IIVALIVLRKKHAPMRQPGVNHAVVE

Vertebrata

N japonica AβPP MEVEFOODSG---YEVHHOKLV---FFP-----KDVGSNKGAI-----IGLMV-GGVVIA FIVITVLMKKK---QYTSIHHGVVE

D rerio AβPP MEAEERH-----SEVYHOKLV---FFA-----EDVSSNKGAI-----IGLMV-GGVVIA FIVITVLMKKK---QYTSIHHGIE

H sapiens AβPP MDAEFRHDSG---YEVHHOKLV---FFA-----EDVGSNKGAI-----IGLMV-GGVVIA FIVITVLMKKK---QYTSIHHGVVE

β-secretase site

α-secretase site

γ-secretase sites

D rerio APLP2 NKINENVVID---ETL DVKEVI---LNGVDPLFSRGSEQESFRPLVDDFSPGSSAL-----IGLLV-I AVAIATVIVISLVLLRKR---QYGTISHGIVE

H sapiens APLP2 NKVDENMVID---ETL DVKEMI---FNAERVGGLEEBRESVGPTRREDFSLSSAL-----IGLLV-I AVAIATVIVISLVMLRKR---QYGTISHGIVE

β-secretase site

α-secretase site

γ-secretase sites

X laevis APLP1 RTLTFPGPKDNGSVKQVHYQSEDIQ-----RDQLEPD---TLVTFNRGAL-----IGLLV-VAVAVAVI IISLLLRKR---PYGTISHGIE

M domestica APLP1 LRVPMGFPERGEGLAVNSVPRGLP---FSPSEIORDELVS-----GGTGVARGAV-----AGLLV-VAVAVAVI VLSLLVARRRR---PYGAISHGVVE

H sapiens APLP1 RKVNASVPRG---FPFHSSEIQ-----RDELAP---AGTGVSREAV-----SGLLI-MGAGGSSLI VLSMLLLRKR---PYGAISHGVVE

Putative cleavage site

α-secretase site

γ-secretase sites

Figure 5 (See legend on next page.)

(See figure on previous page.)

Figure 5 Amyloidogenic Potential in the Amyloid- β Sequence. Aligned representative amino acid sequences for the regions corresponding to exons 16 – 17 of human A β PP. Sequences tested are marked with boxes. Residues with AmylPred consensus and PASTA energies < -4 are in red; residues with AmylPred consensus and PASTA energies between -3 and -4 are in blue. Known secretase cleavage sites are marked by arrows.

amyloid but at a lowered probability than mammalian A β and there is experimental evidence that it can form fibrillar A β *in vivo* [34]. As no other Hymenoptera species in the study show amyloid potential, this likely represents a new mutation in the development of the fruitfly species. Interestingly, non-vertebrate deuterostome species in this study have amyloidogenic sequence but little homology to the mammalian A β sequence, suggesting that early amyloid prone regions may have evolved prior to the divergence of echinoderms, hemichordate, and chordate species. The main sequence variations arise from the N-terminal region aligned to *Homo sapiens* A β PP exon 16. APLP-2 from the zebrafish *Danio rerio* also showed amyloidogenic potential and all other APLP-2 had reduced potential to form A β . This is may be a result of mutation or indels in the exon 16 region during or after the gene duplication events giving rise to APLP-1 and APLP-2.

Because the data used in this study were based on *in silico* search strategies from deposited sequences in public repositories (GenBank and GenPept), it cannot be assumed that these data are necessarily complete for each species (i.e., a *de novo* sequencing was not performed for each species studied). Nonetheless, these data support the hypotheses that A β PP is the ancestral sequence for vertebrates, gene duplication after the speciation of cartilaginous and bony fish gave rise to APLP-2, and a subsequent partial or degenerate duplication of APLP-2 following the speciation of tetrapods gave rise to APLP-1. Some species may have subsequently lost either APLP-1 or APLP-2 genes.

The sequence difference in *Mus musculus* and *Rattus norvegicus* results from three amino acid substitutions from three single nucleotide changes. Whether the lack of amyloidogenesis in these particular rodents comes from these three changes or from other physiological considerations is unclear, but the presence of identical sequence in other rodents and mammals in general suggests that the ancestral species to mice and rats evolved around amyloidogenic A β . The lack of data on A β deposition in fish, birds, reptiles, and amphibians also suggests unknown physiological adaptations may limit A β production or deposition. Recently a mutation encoding a change from alanine to threonine at position 673 of A β PP was found to be protective against developing Alzheimer Disease, likely through reduction of β -secretase processing at that site [10]. It is interesting to

note that all fish sequences in this study, with the exception of *Danio rerio*, have a threonine at this position, suggesting β -secretase processing may be reduced in these animals. In addition to processes that may increase or decrease A β production by regulating secretase efficiency or transcription, the presence of a β -secretase in the gene repertoire is an important consideration.

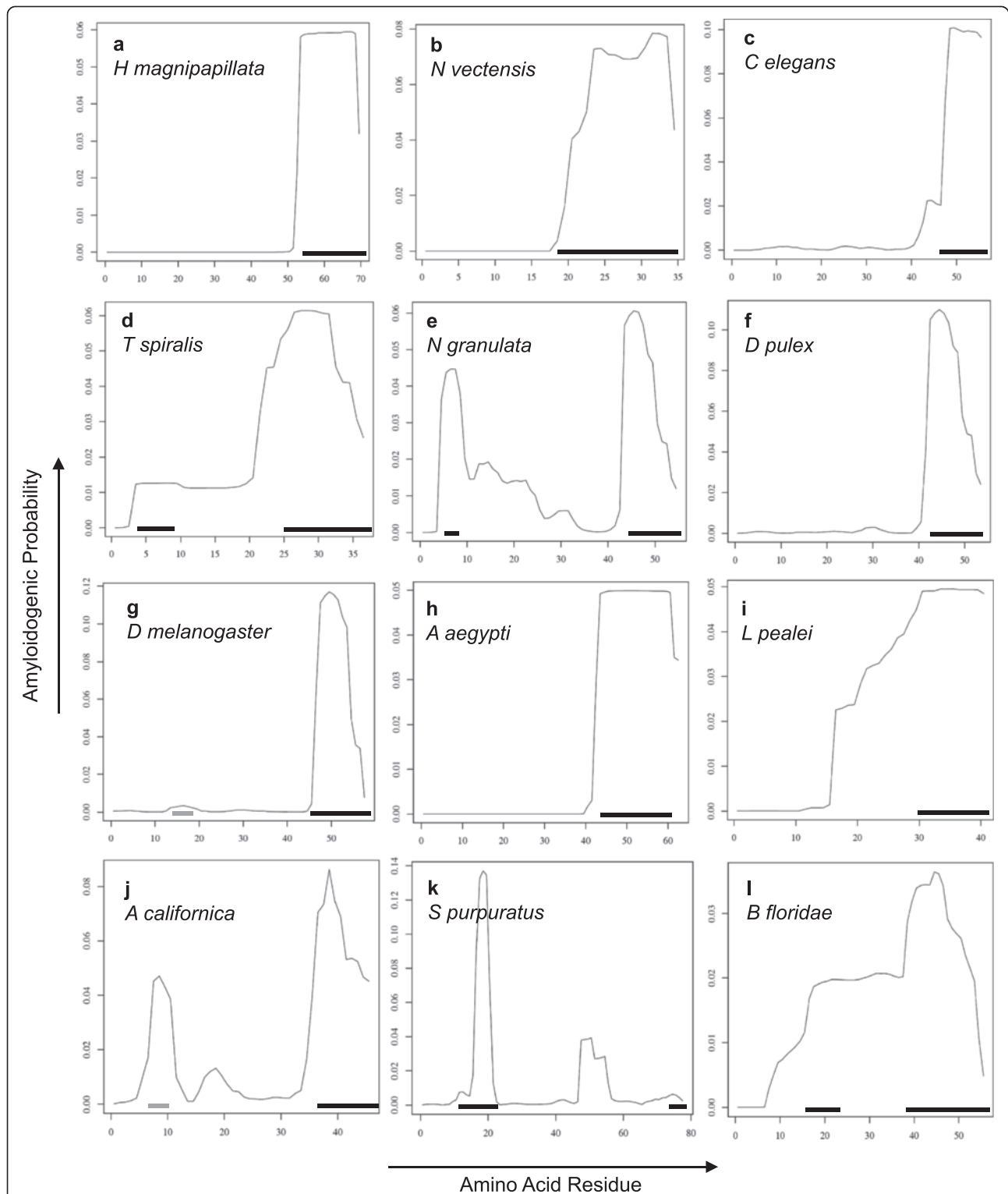
A whole genome assembly for *Nematostella vectensis* indicate the presence of the secretases but no studies have examined amyloid formation [55]. A genome for *Hydra magnipapillata* predicted the presence of a γ -secretase, but not a β -secretase (REFSEQ NW_002165109). Experimental evidence suggests that the nematode *Caenorhabditis elegans* does not express a β -secretase, although both α - and γ -secretases have been identified [56]. A search of Entrez Nucleotide returned no β -secretase sequences for other nematodes, crustaceans, hymenoptera, or lepidoptera in our dataset.

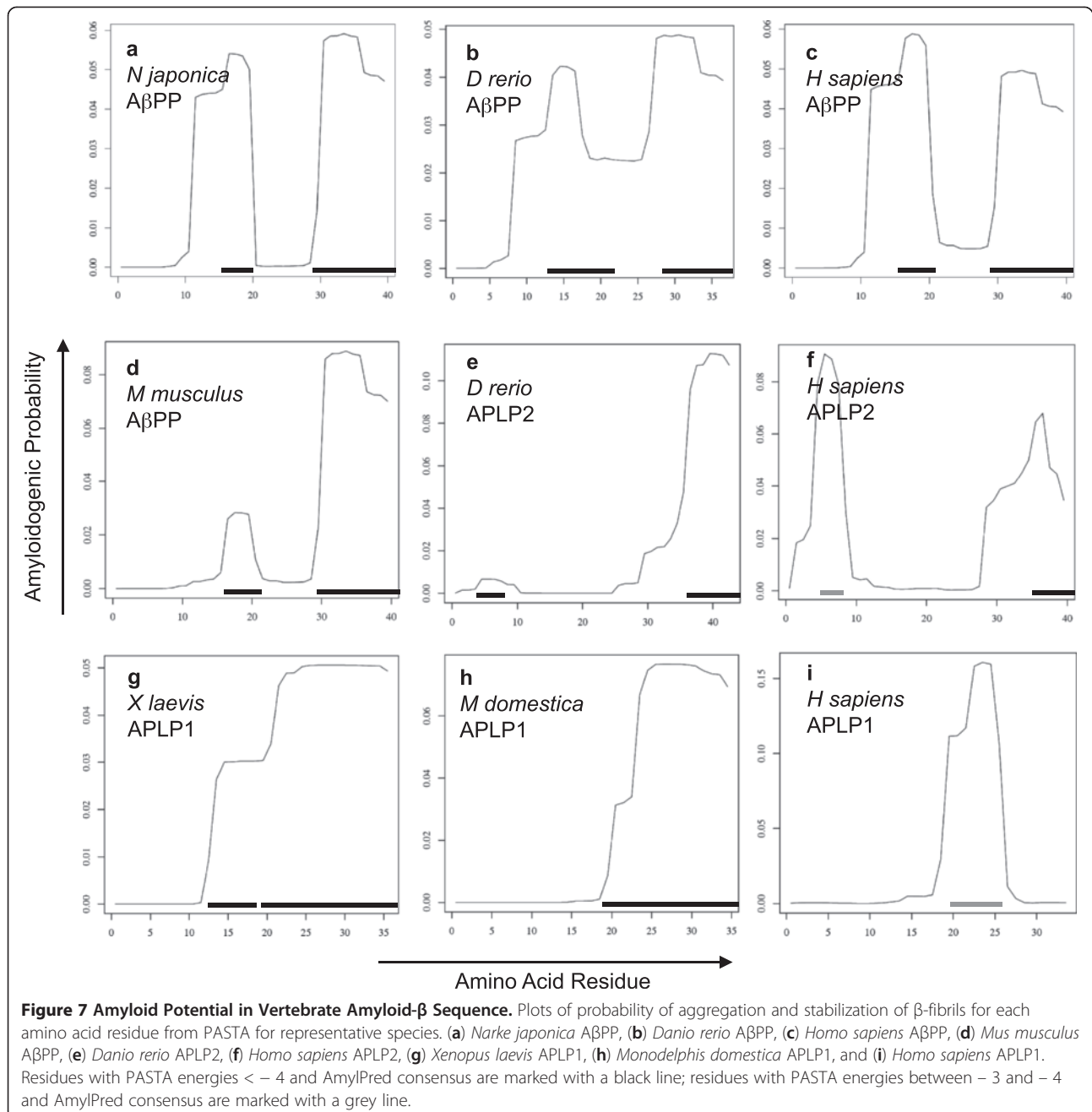
The increased understanding of disease genetics and increasing availability of molecular sequence data provide an opportunity to harness evolutionary approaches to provide deep insights pertaining to the etiology of disease. Using this approach we found the A β PP family to have origins in the speciation of the metazoic lineage and propose that ancestral A β may have arisen as deuterostomia and protostomia diverged. However, other mutations may continue to produce amyloidogenic sequences in this domain, as seen with *Drosophila* or unknown physiological factors may play a role in preventing A β formation as in mice and rats. The approach developed here may be widely applicable to the study of other critical disease genes and builds a foundation for further studies on the co-evolution of Alzheimer Disease associated proteins (e.g., co-evolution of ApoE or β -secretase with A β PP) that may yield novel approaches to treating or preventing A β formation.

Methods

Dataset collection and alignment

Amino acid sequences were collected through Entrez Protein using a combination of search terms and sequence similarity searches. First, based on previous studies of sequences from the Amyloid- β Precursor Protein [2-4,36] family five sets of metadata-based search terms developed and used to identify those sequences from across the Amyloid- β Precursor Protein family: (1) "App"[gene name] AND "animals"[porgn: __txid33208];





(2) "aplp1"[gene name] and "animals" [porgn:__txid33208]; (3) "aplp2"[gene name] and "animals"[porgn:__txid33208]; (4) "apl-1"[gene name] and "nematodes" [porgn:__txid6231]; and (5) "app_amyloid". Sequences for which the organism was either "Unknown" or listed as a "synthetic construct" were removed. Next, a stringent (E-value = 0.0) blastp (BLAST+ v.2.2.26) was used to search Entrez Protein for potential orthologous amino acid sequences for each of the sequences identified in the metadata-based search from the non-redundant protein

database. An additional stringent blastp search was then done iteratively for each new sequence identified, until no additional sequences were found. The resulting dataset (which contained 435 sequences) was then subjected to multiple sequence alignment using MUSCLE v.3.8.31 [38]. The multiple sequence alignment was manually inspected (by viewing the data in Mesquite 2.75 [37]) to identify the one longest representative sequence per taxon (e.g., only the sequence for human A β PP770 which contains all transcribed and translated exons was kept). As sequences were removed from the dataset, the multiple sequence alignment



Figure 8 (See legend on next page.)

(See figure on previous page.)

Figure 8 Evolutionary Relationship of Amyloid- β Formation Potential. Phylogram of the protein sequence for the A β PP family color-coded by prediction of Amyloid- β formation. Red: High potential (PASTA energies < -4 and AmylPred consensus); Blue: Low potential (PASTA energies between -3 and -4 and AmylPred consensus); Black: No potential. Species with unique sequences demonstrating potential for Amyloid- β formation are labeled.

was redone. The resulting dataset reflected 103 taxa corresponding to 67 species. Based on identifiers within GenPept records, corresponding nucleic acid sequences were then collected for each amino acid sequence. These nucleotide sequences were also subjected to multiple sequence alignment using MUSCLE. Character maps were generated using the Mesquite character matrices.

Generation of phylogenetic trees

Trees were obtained by maximum parsimony using TNT 1.1 and Bayesian inference using MrBayes 3.2.0 [39,40]. For analyses in TNT, the 'aquickie.run' script was used to guide the search, which aimed to find the optimal score 20 times independently, using defaults of "xmult" plus 10 cycles of tree-drifting. This resulted in 131 nucleotide trees from more than 8×10^8 rearrangements and 103714 amino acid trees from more than 7×10^8 rearrangements. For consensus tree calculation, trees were TBR-collapsed, Bremer group supports calculated by TBR-swapping, and bootstrap resampling by 100 replications of symmetric resampling with a single random addition (see Additional file 1).

For MrBayes, the Metropolis-coupled Markov chain Monte Carlo analysis was set for 2 runs with 4 chains each with a temperature of 0.2 degrees. A General Time Reversible (GTR) model with a Dirichlet (flat) probability distribution of nucleotide rate change parameters, stationary nucleotide frequencies, no specified shape parameter for the gamma distribution of rate variation, and no invariable sites was used for the nucleotide analyses; this is the default prior model for nucleotide matrices in MrBayes. All runs favored the WAG rate matrix as the prior model for the amino acid analyses [57].

Markov Chain analysis was continued until the runs converged, when the standard deviation of the split frequencies remained < 0.01 and likelihood analysis found the potential scale reduction factor approached 1.0 [58]. For the nucleotide modeling this took more than 3×10^6 generations; for the protein analysis this took more than 2×10^6 generations. Consensus trees were constructed using the 50% majority rule with 95% cumulative posterior probability from 925 nucleotide trees and 1,591 amino acid trees (see Additional

file 1). All tree diagrams were generated in either Dendroscope 3.1.0 or FigTree 1.3.1 [59,60].

Synapomorphic frequencies

Unambiguous synapomorphies at each node were generated in TNT for the maximum parsimony consensus trees. The frequency of a given character being synapomorphic at a given node was examined for the entire amino acid tree and for each of the five major branches. Probabilistic models of synapomorphy have been developed to address the confounding of homoplasy and lend statistical support to defining a character as synapomorphic as opposed to homoplasious [41]. While these are important considerations for higher resolution analysis of a gene family, use of simple statistical analysis for such a large and diverse dataset is a reasonable approach to defining areas of conservation or change, accepting internal error for random mutation producing homoplasy or loss of an actual synapomorphy.

Aggregation modeling

There are a number of programs available for modeling β -folding and aggregation of amyloidogenic peptides [61]. AmylPred is a consensus tool that predicts β -folding and aggregation based on a set of five published methods and uses agreement of 2 or more methods for determining consensus [54]. PASTA predicts stabilizing sequences in β -fibrillar structures using a calculation of the change of energy from pairing between amino acid sequences [53]. Regions that are known to form ordered β -fibril structures have a PASTA energy less than -4 . Using aligned amino acid sequences coded by *Homo sapiens* A β PP exons 16 and 17, we examined the corresponding β A4 region across all taxa and used known secretase cleavage sites to determine the aligned sequences for submission to AmylPred and PASTA [62-64]. Where cleavage sites are not known from previous studies, boundaries were chosen based on similar species and sequences. In cases where there was no clear similarity, boundaries were extended to correspond with *Homo sapiens* A β 42. PASTA energies were collected until greater than -2 by sequential truncation of the C-terminus for each sequence.

Additional file

Additional file 1: Figure S1. Phylogenetic Relationships among the Amyloid- β Precursor Protein Gene Family from Bayesian Inference. **a**, Phylogram showing the evolutionary relationships among the nucleotide sequences of the A β PP gene family. **b**, Phylogram for the corresponding protein sequences. Trees were generated by Bayesian inference methods and show posterior probability values at each node. **Figure S2.** Branch Supports for Phylogenetic Trees. Symmetric bootstrap re-sampling and Bremer supports, for nucleotide trees (**a** and **b**, respectively) and for amino acid trees (**c** and **d**, respectively). **Table S1.** Taxa Species Names and Sequence Accession Numbers.

Abbreviations

A β : Amyloid- β ; A β PP: APP, Amyloid- β Precursor Protein; APPL-1: Amyloid- β Precursor Protein-like 1 protein; APLP-1: Amyloid precursor like protein 1; APLP-2: Amyloid precursor like protein 2; APL-1: Amyloid precursor like 1 protein; BLS: Basolateral sorting signal; GFLD: Growth-factor-like domain; KPI: Kunitz-protease inhibitor; Mya: Million years ago.

Competing interests

INS and WGT do not have any conflicts of interest to disclose.

Authors' contributions

INS and WGT conceived of and designed the study together. INS collected and aligned the sequences. WGT conducted the tree building and aggregation analyses. Both INS and WGT interpreted the results and drafted the manuscript together. Both authors read and approved the final manuscript.

Acknowledgements

This work was supported in part by a grant to I.N.S. from the National Library of Medicine (R01 LM009725). W.G.T. is supported by an individual fellowship award from the National Institute of Diabetes and Digestive and Kidney Diseases (F30 DK084605).

Author details

- ¹Center for Clinical and Translational Science, University of Vermont, Given Courtyard N309, 89 Beaumont Avenue, Burlington, VT 05405, USA.
- ²Division of Endocrinology, Department of Medicine, University of Vermont, Given Courtyard N309, 89 Beaumont Avenue, Burlington, VT 05405, USA.
- ³Department of Microbiology and Molecular Genetics, University of Vermont, Given Courtyard N309, 89 Beaumont Avenue, Burlington, VT 05405, USA.
- ⁴Department of Computer Science, University of Vermont, Given Courtyard N309, 89 Beaumont Avenue, Burlington, VT 05405, USA.

Received: 16 November 2012 Accepted: 24 April 2013

Published: 30 April 2013

References

- Hardy J, Selkoe DJ: The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 2002, **297**(5580):353–356.
- Guo Q, Wang Z, Li H, Wiese M, Zheng H: APP physiological and pathophysiological functions: insights from animal models. *Cell Res* 2012, **22**(1):78–89.
- Jacobsen KT, Iverfeldt K: Amyloid precursor protein and its homologues: a family of proteolysis-dependent receptors. *Cell Mol Life Sci* 2009, **66**(14):2299–2318.
- Walsh DM, Minogue AM, Sala Frigerio C, Fadeeva JV, Wasco W, Selkoe DJ: The APP family of proteins: similarities and differences. *Biochem Soc Trans* 2007, **35**(Pt 2):416–420.
- Greenwald J, Riek R: Biology of amyloid: structure, function, and regulation. *Structure* 2010, **18**(10):1244–1260.
- Chernoff YO: Amyloidogenic domains, prions and structural inheritance: rudiments of early life or recent acquisition? *Curr Opin Chem Biol* 2004, **8**(6):665–671.
- Brack A, Orgel LE: Beta structures of alternating polypeptides and their possible prebiotic significance. *Nature* 1975, **256**(5516):383–387.
- Inge-Vechtomov SG, Zhouavleva GA, Chernoff YO: Biological roles of prion domains. *Prion* 2007, **1**(4):228–235.
- Rosenberg RN: The molecular and genetic basis of AD: the end of the beginning: the 2000 Wartenberg lecture. *Neurology* 2000, **54**(11):2045–2054.
- Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, Stefansson H, Sulem P, Gudbjartsson D, Maloney J, et al: A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012. advance online publication.
- Barrachina M, Dalfo E, Puig B, Vidal N, Freixes M, Castano E, Ferrer I: Amyloid-beta deposition in the cerebral cortex in Dementia with Lewy bodies is accompanied by a relative increase in AbetaPP mRNA isoforms containing the Kunitz protease inhibitor. *Neurochem Int* 2005, **46**(3):253–260.
- Yamada M, Naiki H: Cerebral amyloid angiopathy. *Prog Mol Biol Transl Sci* 2012, **107**:41–78.
- Liang WS, Duncley T, Beach TG, Grover A, Mastroeni D, Ramsey K, Caselli RJ, Kukull WA, McKeel D, Morris JC, et al: Neuronal gene expression in non-demented individuals with intermediate Alzheimer's Disease neuropathology. *Neurobiol Aging* 2010, **31**(4):549–566.
- Kotzbauer PT, Cairns NJ, Campbell MC, Willis AW, Racette BA, Tabbal SD, Perlmutter JS: Pathologic Accumulation of alpha-Synuclein and Abeta in Parkinson Disease Patients With Dementia. *Arch Neurol* 2012:1–6.
- Nikolaev A, McLaughlin T, O'Leary DD, Tessier-Lavigne M: APP binds DR6 to trigger axon pruning and neuron death via distinct caspases. *Nature* 2009, **457**(7232):981–989.
- Sinha S, Lieberburg I: Cellular mechanisms of beta-amyloid production and secretion. *Proc Natl Acad Sci U S A* 1999, **96**(20):11049–11053.
- Thinakaran G, Koo EH: APP trafficking, processing and function. *J Biol Chem* 2008, **283**(44):29615.
- Xu F, Davis J, Miao J, Previti ML, Romanov G, Ziegler K, Van Nostrand WE: Protease nexin-2/amyloid beta-protein precursor limits cerebral thrombosis. *Proc Natl Acad Sci U S A* 2005, **102**(50):18135–18140.
- Bush AI, Martins RN, Rumble B, Moir R, Fuller S, Milward E, Currie J, Ames D, Weidemann A, Fischer P, et al: The amyloid precursor protein of Alzheimer's disease is released by human platelets. *J Biol Chem* 1990, **265**(26):15977–15983.
- Joachim CL, Mori H, Selkoe DJ: Amyloid beta-protein deposition in tissues other than brain in Alzheimer's disease. *Nature* 1989, **341**(6239):226–230.
- Lee YH, Tharp WG, Maple RL, Nair S, Permana PA, Pratley RE: Amyloid precursor protein expression is upregulated in adipocytes in obesity. *Obesity (Silver Spring, Md)* 2008, **16**(7):1493–1500.
- Galloway S, Jian L, Johnsen R, Chew S, Mamo JC: Deta-amyloid or its precursor protein is found in epithelial cells of the small intestine and is stimulated by high-fat feeding. *J Nutr Biochem* 2007, **18**(4):279–284.
- Hansel DE, Rahman A, Wehner S, Herzog V, Yeo CJ, Maitra A: Increased expression and processing of the Alzheimer amyloid precursor protein in pancreatic cancer may influence cellular proliferation. *Cancer Res* 2003, **63**(21):7032–7037.
- Herzog V, Kirfel G, Siemes C, Schmitz A: Biological roles of APP in the epidermis. *Eur J Cell Biol* 2004, **83**(11–12):613–624.
- Kuo YM, Kokjohn TA, Watson MD, Woods AS, Cotter RJ, Sue LI, Kalback WM, Emmerling MR, Beach TG, Roher AE: Elevated abeta42 in skeletal muscle of Alzheimer disease patients suggests peripheral alterations of AbetaPP metabolism. *Am J Pathol* 2000, **156**(3):797–805.
- Kang J, Lemaire HG, Unterbeck A, Salbaum JM, Masters CL, Grzeschik KH, Multhaup G, Beyreuther K, Muller-Hill B: The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* 1987, **325**(6106):733–736.
- Hornsten A, Lieberthal J, Fadia S, Malins R, Ha L, Xu X, Daigle I, Markowitz M, O'Connor G, Plasterk R, et al: APL-1, a *Caenorhabditis elegans* protein related to the human beta-amyloid precursor protein, is essential for viability. *Proc Natl Acad Sci U S A* 2007, **104**(6):1971–1976.
- von Koch CS, Zheng H, Chen H, Trumbauer M, Thinakaran G, van der Plöeg LH, Price DL, Sisodia SS: Generation of APLP2 KO mice and early postnatal lethality in APLP2/APP double KO mice. *Neurobiol Aging* 1997, **18**(6):661–669.
- Heber S, Herms J, Gajic V, Hainfellner J, Aguzzi A, Rulicke T, von Kretschmar H, von Koch C, Sisodia S, Tremml P, et al: Mice with combined gene knock-outs reveal essential and partially redundant functions of amyloid precursor protein family members. *J Neurosci* 2000, **20**(21):7951–7963.

30. Li H, Wang Z, Wang B, Guo Q, Dolios G, Tabuchi K, Hammer RE, Sudhof TC, Wang R, Zheng H: **Genetic dissection of the amyloid precursor protein in developmental function and amyloid pathogenesis.** *J Biol Chem* 2010, **285**(40):30598–30605.
31. Poock B, Strauss R, Kretschmar D: **Analysis of amyloid precursor protein function in *Drosophila melanogaster*.** *Exp Brain Res* 2012, **217**(3–4):413–421.
32. Song P, Pimplikar SW: **Knockdown of amyloid precursor protein in zebrafish causes defects in motor axon outgrowth.** *PLoS One* 2012, **7**(4):e34209.
33. Joshi P, Liang JO, DiMonte K, Sullivan J, Pimplikar SW: **Amyloid precursor protein is required for convergent-extension movements during Zebrafish development.** *Dev Biol* 2009, **335**(1):1–11.
34. Carmine-Simmen K, Proctor T, Tschape J, Poock B, Triphan T, Strauss R, Kretschmar D: **Neurotoxic effects induced by the *Drosophila* amyloid-beta peptide suggest a conserved toxic function.** *Neurobiol Dis* 2009, **33**(2):274–281.
35. Miklos I, Zadori Z: **Positive evolutionary selection of an HD motif on Alzheimer precursor protein orthologues suggests a functional role.** *PLoS Comput Biol* 2012, **8**(2):e1002356.
36. Coulson EJ, Paliga K, Beyreuther K, Masters CL: **What the evolution of the amyloid protein precursor supergene family tells us about its function.** *Neurochem Int* 2000, **36**(3):175–184.
37. Maddison WP, Maddison DP: *Mesquite: a modular system for evolutionary analysis.* 275th edition; 2011.
38. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**(5):1792–1797.
39. Goloboff P, Farris J, Nixon K: **TNT, a free program for phylogenetic analysis.** *Cladistics* 2008, **24**:774–786.
40. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**(12):1572–1574.
41. Turjak M, Trontelj P: **A method for measuring support for synapomorphy using character state distributions on phylogenetic trees.** *Cladistics* 2012, **1**:1–12.
42. Finch CE, Sapolsky RM: **The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms.** *Neurobiol Aging* 1999, **20**(4):407–428.
43. Sarasa M, Gallego C: **Alzheimer-Like Neurodegeneration as a Probable Cause of Cetacean Stranding.** In *Poster session presented at: 5th FENS Forum: 2006; Vienna, Austria; 2006.*
44. Selkoe DJ, Bell DS, Podlisny MB, Price DL, Cork LC: **Conservation of brain amyloid proteins in aged mammals and humans with Alzheimer's disease.** *Science* 1987, **235**(4791):873–877.
45. Beach TG: **Physiologic origins of age-related beta-amyloid deposition.** *Neurodegener Dis* 2008, **5**(3–4):143–145.
46. Ionov ID, Pushinskaya II: **Amyloid-beta production in aged guinea pigs: atropine-induced enhancement is reversed by naloxone.** *Neurosci Lett* 2010, **480**(1):83–86.
47. Dayan AD: **Comparative neuropathology of ageing. Studies on the brains of 47 species of vertebrates.** *Brain* 1971, **94**(1):31–42.
48. Flood DG, Lin YG, Lang DM, Trusko SP, Hirsch JD, Savage MJ, Scott RW, Howland DS: **A transgenic rat model of Alzheimer's disease with extracellular Abeta deposition.** *Neurobiol Aging* 2009, **30**(7):1078–1090.
49. Maldonado TA, Jones RE, Norris DO: **Timing of neurodegeneration and beta-amyloid (Abeta) peptide deposition in the brain of aging kokanee salmon.** *J Neurobiol* 2002, **53**(1):21–35.
50. Nakayama H, Katayama K, Ikawa A, Miyawaki K, Shinozuka J, Uetsuka K, Nakamura S, Kimura N, Yoshikawa Y, Doi K: **Cerebral amyloid angiopathy in an aged great spotted woodpecker (*Picoides major*).** *Neurobiol Aging* 1999, **20**(1):53–56.
51. Trovato A, Seno F, Tosatto SC: **The PASTA server for protein aggregation prediction.** *Protein Eng Des Sel* 2007, **20**(10):521–523.
52. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L: **Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins.** *Nat Biotechnol* 2004, **22**(10):1302–1306.
53. Trovato A, Chiti F, Maritan A, Seno F: **Insight into the structure of amyloid fibrils from the analysis of globular proteins.** *PLoS Comput Biol* 2006, **2**(12):e170.
54. Frousios KK, Iconomidou VA, Karletidi CM, Hamodrakas SJ: **Amyloidogenic determinants are usually not buried.** *BMC Struct Biol* 2009, **9**:44.
55. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, et al: **Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization.** *Science* 2007, **317**(5834):86–94.
56. Link CD: **C. elegans models of age-associated neurodegenerative diseases: lessons from transgenic worm models of Alzheimer's disease.** *Exp Gerontol* 2006, **41**(10):1007–1013.
57. Whelan S, Goldman N: **A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach.** *Mol Biol Evol* 2001, **18**(5):691–699.
58. Gelman A, Rubin D: **Inference from Iterative Simulation using Multiple Sequences.** *Stat Sci* 1992, **7**:457–511.
59. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R: **Dendroscope: An interactive viewer for large phylogenetic trees.** *BMC Bioinformatics* 2007, **8**:460.
60. Rambaut A: *FigTree.* 131st edition; 2009.
61. Hamodrakas SJ: **Protein aggregation and amyloid fibril formation prediction software from primary sequence: towards controlling the formation of bacterial inclusion bodies.** *FEBS J* 2011, **278**(14):2428–2435.
62. Hognl S, Kuhn PH, Colombo A, Lichtenthaler SF: **Determination of the proteolytic cleavage sites of the amyloid precursor-like protein 2 by the proteases ADAM10, BACE1 and gamma-secretase.** *PLoS One* 2011, **6**(6):e21337.
63. Yanagida K, Okochi M, Tagami S, Nakayama T, Kodama TS, Nishitomi K, Jiang J, Mori K, Tatsumi S, Arai T, et al: **The 28-amino acid form of an APLP1-derived Abeta-like peptide is a surrogate marker for Abeta42 production in the central nervous system.** *EMBO Mol Med* 2009, **1**(4):223–235.
64. Minogue AM, Stubbs AK, Frigerio CS, Boland B, Fadeeva JV, Tang J, Selkoe DJ, Walsh DM: **Gamma-secretase processing of APLP1 leads to the production of a p3-like peptide that does not aggregate and is not toxic to neurons.** *Brain Res* 2009, **1262**:89–99.

doi:10.1186/1471-2164-14-290

Cite this article as: Tharp and Sarkar: Origins of amyloid- β . *BMC Genomics* 2013 **14**:290.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

