

A Comprehensive Phylogenetic Analysis of Deadenylases

Athanasia Pavlopoulou^{1,#}, Dimitrios Vlachakis^{1,#}, Nikolaos A.A. Balatsos² and Sophia Kossida¹

¹Bioinformatics and Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Soranou Efessiou 4, Athens 11527, Greece. ²Department of Biochemistry and Biotechnology, University of Thessaly, 26 Ploutonos st., 41 221 Larissa, Greece. #AP and DV contributed equally to this work. Corresponding author email: skossida@bioacademy.gr

Abstract: Deadenylases catalyze the shortening of the poly(A) tail at the messenger ribonucleic acid (mRNA) 3'-end in eukaryotes. Therefore, these enzymes influence mRNA decay, and constitute a major emerging group of promising anti-cancer pharmacological targets. Herein, we conducted full phylogenetic analyses of the deadenylase homologs in all available genomes in an effort to investigate evolutionary relationships between the deadenylase families and to identify invariant residues, which probably play key roles in the function of deadenylation across species. Our study includes both major Asp-Glu-Asp-Asp (DEDD) and exonuclease-endonuclease-phosphatase (EEP) deadenylase superfamilies. The phylogenetic analysis has provided us with important information regarding conserved and invariant deadenylase amino acids across species. Knowledge of the phylogenetic properties and evolution of the domain of deadenylases provides the foundation for the targeted drug design in the pharmaceutical industry and modern exonuclease anti-cancer scientific research.

Keywords: deadenylases, phylogenesis, molecular modelling, evolution

Evolutionary Bioinformatics 2013:9 491–497

doi: [10.4137/EBO.S12746](https://doi.org/10.4137/EBO.S12746)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article published under the Creative Commons CC-BY-NC 3.0 license.



Introduction

Shortening of the polyadenylated (poly(A)) tail at the mRNA 3'-end, referred to as deadenylation, is a key step in mRNA decay in eukaryotes.^{1,2} This process is catalyzed by the deadenylase enzymes. Poly(A) tails are the preferred substrates of deadenylases, although in some instances they are capable of degrading non-adenosine ribopolymers *in vitro* with reduced efficiency.³⁻⁶ According to Goldstrohm and Wickens,⁷ the known deadenylases are classified into 2 superfamilies, DEDD and exonuclease-endonuclease-phosphatase (EEP), which are defined by conserved exonuclease sequence motifs required for catalysis. Members of the EEP superfamily of deadenylases use a conserved glutamic acid (E) and a histidine (H) for catalysis.^{7,8} This superfamily includes the families carbon catabolite repressor 4 (CCR4), Nocturinin, ANGEL and 2' phosphodiesterase (2'PDE).⁷ The DEDD superfamily of deadenylases is named after the invariant catalytic acidic residues aspartic acid (D) and glutamic acid (E), which are distributed in 3 exonuclease motifs.^{7,9} The DEDD superfamily includes the families POP2, Poly(A)-specific ribonuclease (PARN), CAF1Z and PAB-dependent poly(A)-specific ribonuclease subunit 2 (PAN2).^{7,9} In the present study, we focus on the molecular evolution of these families, thus providing insights into the amino acid conservation patterns that may be subsequently used for further studying deadenylases as a promising and emerging anti-cancer pharmacological target.

Methods

Identification of deadenylase homologues

To identify homologous deadenylase protein sequences, the accession numbers of the characterized deadenylases reported in literature⁷ were used to retrieve their corresponding amino acid sequences from publically-available databases UniProtKB¹⁰ and GenBank.¹¹ These sequences were subsequently used as probes to search the sequence databases by applying reciprocal BLASTp and tBLASTn.¹² This process was reiterated until no new putative deadenylase homologues could be found.

Motifs construction

Representative DEDD and EEP peptide sequences were aligned and edited with Utopia suite's CINEMA

alignment editor.¹³ Sequence motifs were collected from the alignments, manually edited for insertions or gaps, and submitted to WebLogo3¹⁴ to generate consensus sequences.

Phylogenetic analyses

The deadenylase sequences under study were searched against InterPro¹⁵ in order to identify the boundaries of the core nuclease domain. In order to optimize the alignment and avoid unreasonable gap penalties, the amino acid sequences that correspond to the nuclease domain were collected from the entire deadenylase peptide sequences and aligned using CLUSTALW.¹⁶ The resulting multiple sequence alignment was first trimmed for gaps using Gblocks^{17,18} and manually edited. The trimmed alignment was then used to reconstruct phylogenetic trees by employing 2 different methods. The first one is the maximum-likelihood method implemented in PhyML,¹⁹ where an initial distance-based tree (BIONJ) is optimized using a hillclimbing algorithm. In our study, the nearest-neighbor-interchange (NNI) heuristic was used with 4 substitution-rate categories; the proportion of the invariable sites and the gamma shape parameter were estimated from the data. The number of amino acid substitutions per site was estimated with the LG²⁰ model. The second one is the Neighbor-net method²¹ implemented in SplitsTree4,²² a distance-based method which detects conflicting phylogenetic signals, presented in the form of reticulations; the Uncorrected P substitution model was used. Bootstrap analyses (1000 pseudo-replicates) were conducted in order to assess the robustness of the reconstructed trees. The inferred phylogenetic trees were visualized with the program Dendroscope.²³

Evolutionary rate shift analysis

A maximum-likelihood method²⁴ was employed for the identification of evolutionary rate differences at specific protein sites in DEDD families. Towards this end, a set of 19 protein sequences from the four DEDD families was analyzed in order to identify amino acid positions with significant 4 rate differences among the DEDD families as described in Knudsen and Miyamoto (2001).²⁴ The alignment was based on the core nuclease domain, and it was carried out using CLUSTALW.¹⁶

Results/Discussion

Phylogenetic analyses of deadenylases

In the present study, we performed comprehensive and updated phylogenetic analyses of the deadenylase homologs in all available genomes (Figs. 1, 2, S1 and

S2, Table S1). Collectively, 114 DEDD and 97 EEP homologous protein sequences were identified in the genomes of 38 and 37 species, respectively, which represent major eukaryotic taxonomic divisions (according to the NCBI taxonomy database; Table S1).²⁵

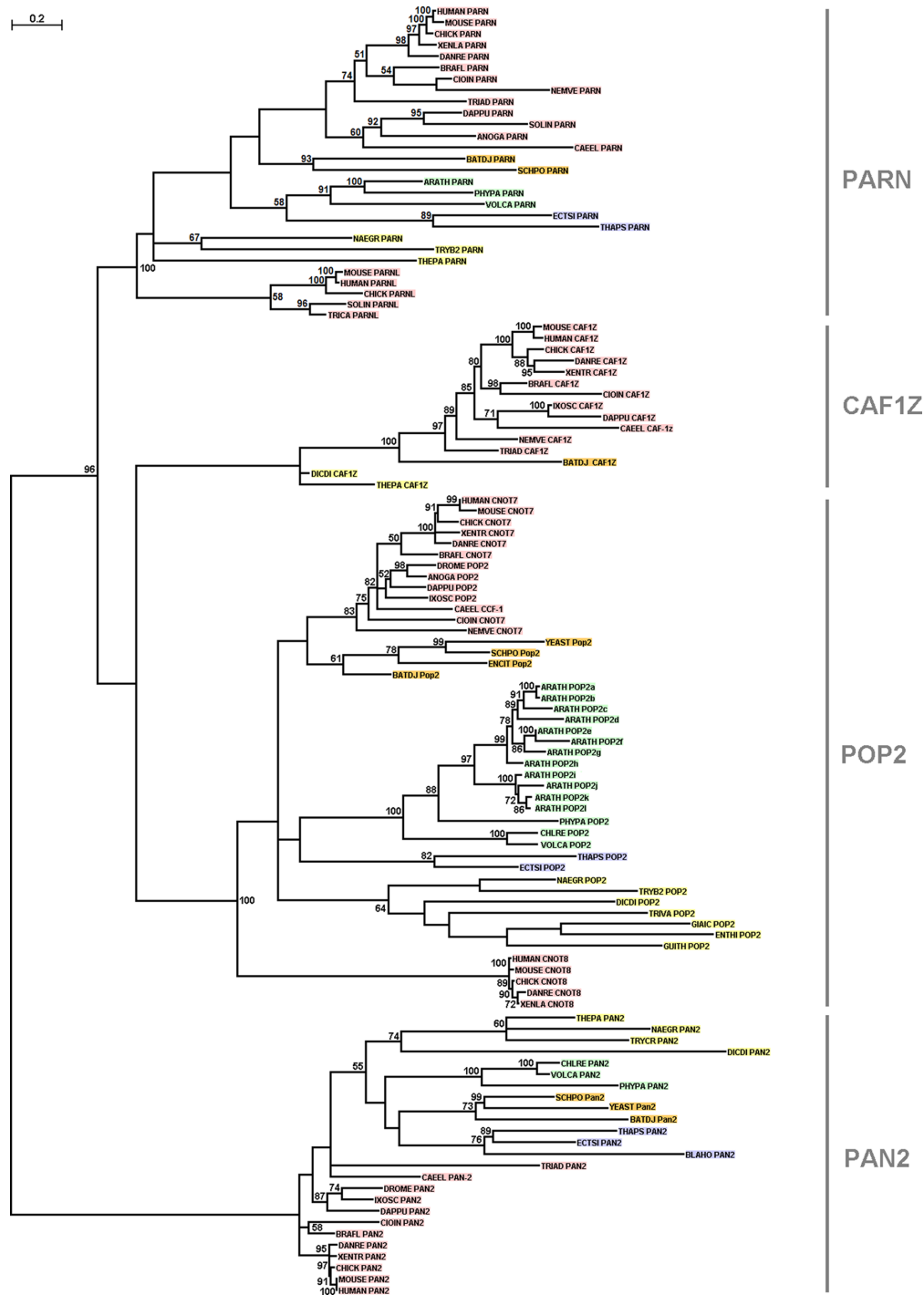


Figure 1. Phylogenetic tree of DEDD deadenylases. Bootstrap values (>50%) are shown at the nodes. The length of the tree branches reflects evolutionary distance. The scale bar at the upper left represents the length of amino acid substitutions per position. To minimize confusion, we used the protein names as described in Goldstrohm and Wickens;⁷ the UniProt 5-letter codes were used for the species names. The proteins derived from metazoa are shown in red, from viridiplantae in green, from fungi in orange and from protozoa in yellow.

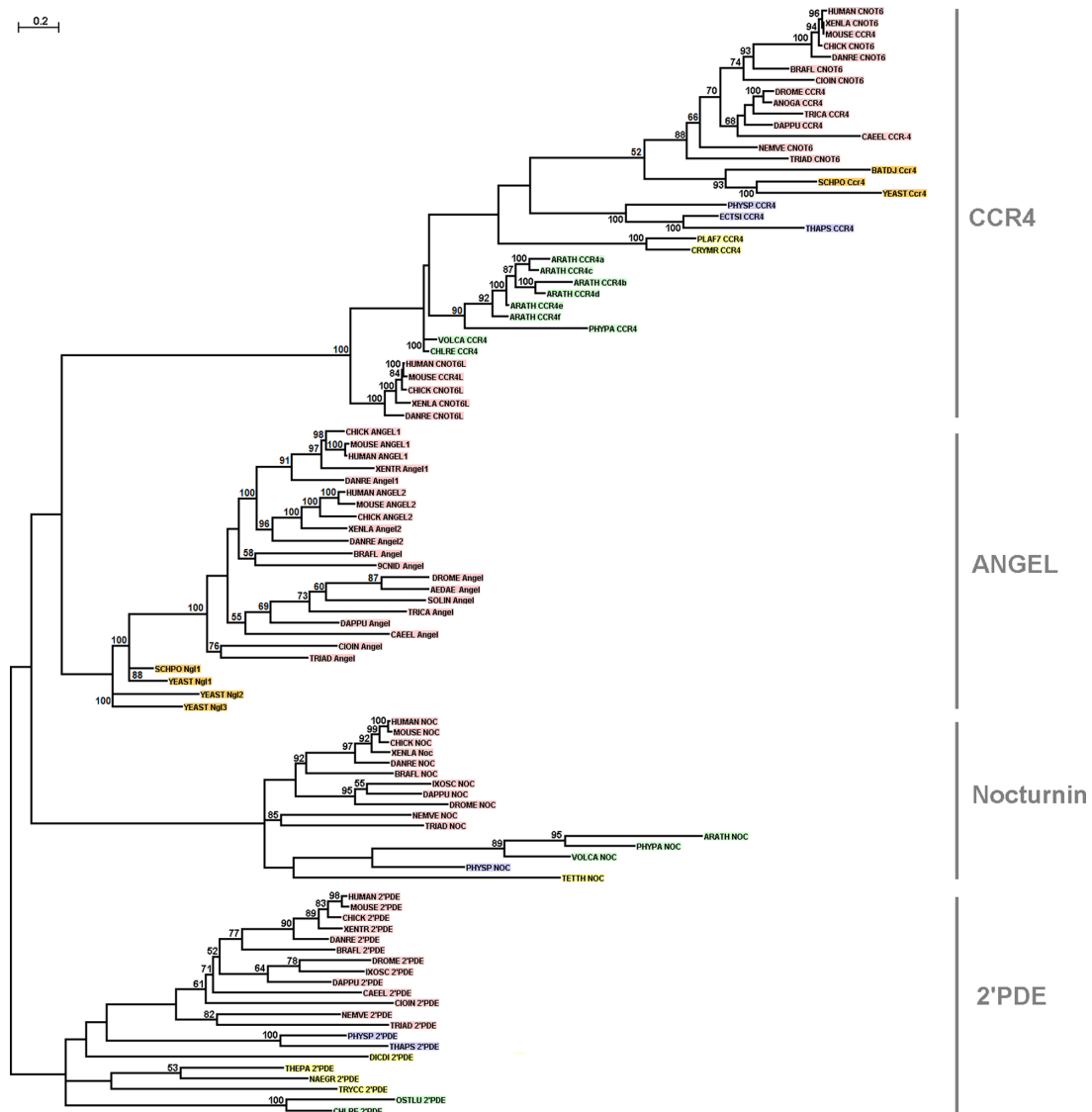


Figure 2. Phylogenetic tree of EEP deadenylases. Bootstrap values above 50% are shown at the nodes. The length of the tree branches depicts evolutionary distance. The scale bar at the upper left represents the length of amino acid substitutions per site. To minimize confusion, we used the protein names as described in Goldstrohm and Wickens;⁷ the UniProt 5-letter codes were used for the species names. The proteins derived from metazoa are shown in red, from viridiplantae in green, from fungi in orange and from protozoa in yellow.

In order to better resolve the evolutionary relationships between the deadenylase families, we applied 2 different methods for phylogenetic tree reconstruction. The phylogenetic trees reconstructed with both methods are congruent, since the overall topology is similar, and all main branches are supported by high bootstrap values (Figs. 1, 2, S1 and S2). 8 coherent, well-supported monophyletic branches that correspond to the 4 families of the DEDD superfamily (Figs. 1 and S1), and the 4 families that comprise the EEP superfamily (Figs. 2 and S2) are distinguished.

Based on our analysis, putative members of the families POP2, PARN, PAN2 and CCR4 were

identified in the major eukaryotic taxonomic divisions, ranging from metazoa to protozoa. POP2 appears to be the largest family in size with a wide distribution among taxa (Figs. 1 and S1).

Based on the phylogenetic analyses (Figs. 1, 2, S1 and S2), the deadenylase families POP2, PARN and CCR4 appear to have undergone gene duplications in metazoa giving rise to the, metazoan-specific, subfamilies CNOT8, PARNL and CNOT6 L, respectively. In POP2 and CCR4 families, gene duplications have rather occurred after the emergence of teleosts (bony fishes) (Figs. 1, S1, 2 and S2), since teleost (DANRE) homologs were detected in the

corresponding subfamilies CNOT8 and CNOT6L. In PARN, a duplication event has presumably followed the radiation of arthropods (Figs. 1 and S1), as arthropod (SOLIN and TRICA) homologs were identified in the PARNL subfamily. However, neither frog (XENTR) nor fish (DANRE) PARNL homologs were identified; we suggest that frog and fish *PARNL* genes might have existed that probably got deleted during the evolutionary course.

Of importance, PARN homologs were not detected in the fungus *Saccharomyces cerevisiae* (YEAST) and the arthropod *Drosophila melanogaster* (DROME), whereas putative PARN homologs were found in other fungi (BATDJ and SCHPO) and arthropods (DAPPU, ANOGA and SOLIN) (Figs.1 and 1S). This leads to the suggestion that alternative metabolic pathways might exist in yeast and *Drosophila* that might compensate for PARN's function. Furthermore, a series of species-specific gene duplications in the green plant *Arabidopsis thaliana* (ARATH) yielded 12 POP2 and 6 CCR4 paralogs (Figs. 1, 2, S1 and S2).

However, the deadenylase families CCR4-associated factor 1Z (CAF1Z), ANGEL, Nocturnin and 2'PDE are restricted to certain eukaryotic taxa (Figs. 1, S1, 2 and S2). CAF1Z is restricted to metazoa and protozoa (Figs. 1 and S1). Moreover, a putative CAF1Z homolog was detected in the chytrid fungus *Batrachochytrium dendrobatidis* (BATDJ), which infects frogs causing chytridiomycosis^{26,27} (Figs. 1 and S1), leading to the suggestion that this fungal parasite has presumably acquired CAF1Z from its amphibian host by horizontal gene transfer. Furthermore, members of the Nocturnin and 2'PDE families were detected in the main eukaryotic taxa, except fungi (Figs. 2 and S2). We suggest that either alternative metabolic pathways may exist in fungi for poly(A) degradation, or differences in the life-style or physiology of fungi led to the loss of the Nocturnin and 2'PDE families. The ANGEL family is restricted to opisthokonts (metazoa and fungi; Figs. 2 and S2). Based on the reconstructed phylogeny, the ANGEL1/ANGEL2 duplication should have occurred after the vertebrate-invertebrate separation (Figs. 2 and S2). This is supported by the observation that a single ANGEL1/2-like homologue was detected in the invertebrate chordate *Branchiostoma floridae* (BRAFL Angel), which lies in the

vertebrate-invertebrate evolutionary boundary, that appears to be basal to the ANGEL1 and ANGEL2 clades in the trees reconstructed with both methods (Figs. 2 and S2). In the ANGEL family, the fungi *Ngl* homologues form a separate, highly supported clade. 3 *S. cerevisiae* (*Ng1*, *Ng2* and *Ng3*) and one *S. pombe* homologue (*Ng11*) were identified. Based

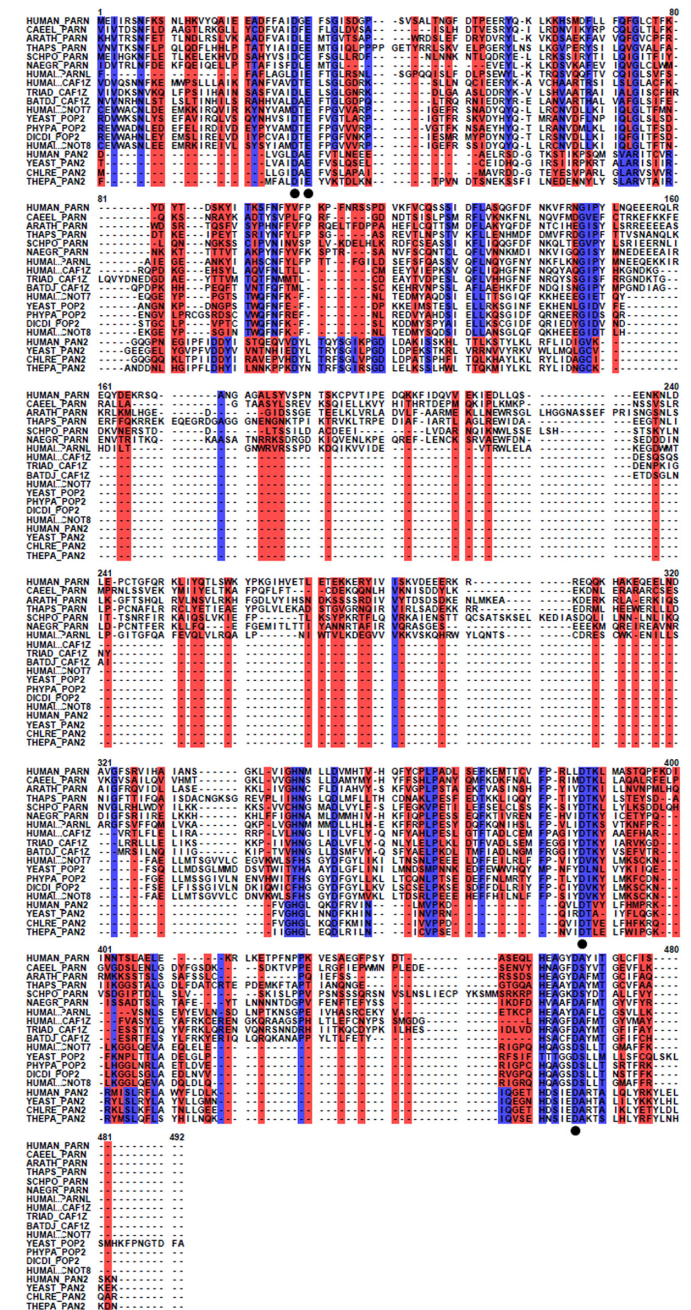


Figure 3. Results of the rate shift analysis for the 19 DEDD proteins. Sites with blue and red highlight correspond to those with slower and faster evolutionary rate, respectively. Sites with entirely blue or red highlight represent amino acid sites with the same evolutionary rate in all families, but with significantly slower or faster rates compared to the average of all sites, respectively.

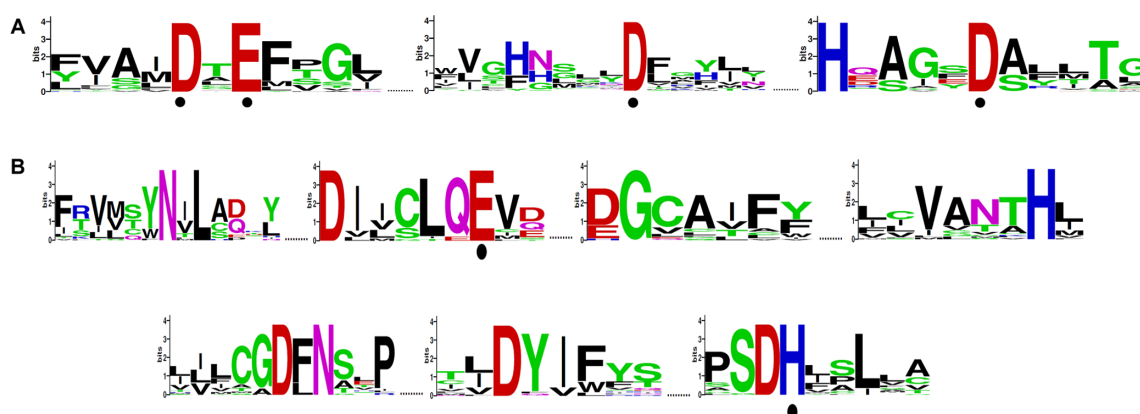


Figure 4. Sequence logos of the motifs identified in deadenylase protein sequences. (A) DEDD, numbered according to the human PARN nuclease domain (PDB code 2A1R) and (B) EEP, numbered according to the human CNOT6 nuclease domain. The height of each letter is relative to the frequency of the corresponding residue at that position, and the letters are ordered such as the most frequent is on the top. The invariant catalytic residues that define each superfamily are indicated with dots.

on the phylogenetic tree (Figs. 2 and S2), we suggest that tandem duplication events, apparently after the *S. cerevisiae* and *S. pombe* divergence, may have copied Ng12 and Ng13.

Based on the rate shift analysis, a total of 153 sites, distributed across the core domain, were detected with significant evolutionary rate differences. Among them, 29 (19%) sites were detected with significant rate shifts between the PAN2 family and the other DEDD families (Fig. 3, both blue and red highlighting). This is in agreement with the phylogenetic analyses results where PAN2 appears to be more distantly related to the other DEDD families. Moreover, 29 (19%) conserved sites were detected in all DEDD families, exhibiting slower evolutionary rates compared to the average of all proteins under investigation (Fig. 3, blue highlighting). This leads to the suggestion that must have been evolutionary pressure to these sites to evolve slowly because they have a critical role in the function or structure of DEDD enzymes; as expected, the 4 catalytic residues that define the DEDD superfamily are also included in this category (Fig. 3). Also, 95 sites, a significantly high percentage (62%), were detected with faster evolutionary rates compared to the average of all DEDD proteins (Fig. 3, red highlighting).

Furthermore, sequence logo analyses were generated in order to determine the consensus sequence of each of the conserved motifs that were deduced from the alignment of representative deadenylase sequences from both superfamilies (Fig. 4A and B). In this way, a set of structurally-conserved residues

were identified on both the DEDD and EEP deadenylases. More specifically, 3 major motifs were identified in DEDD deadenylases and 7 prime motifs in EEP deadenylases (Fig. 4A and B).

Importantly, apart from the known catalytic residues (Fig. 4A and B), several other residues were found to be evolutionary conserved across species in all various deadenylases. Therefore, these amino acids may serve important functional roles in the action of the deadenylase mechanism. They could also represent potential drug targets.

Author Contributions

Conceived and designed the experiments: AP, DV, NB, SK. Analyzed the data: AP, DV, NB, SK. Wrote the first draft of the manuscript: AP, DV, NB, SK. Contributed to the writing of the manuscript: AP, DV, NB, SK. Agree with manuscript results and conclusions: AP, DV, NB, SK. Jointly developed the structure and arguments for the paper: AP, DV, NB, SK. Made critical revisions and approved final version: AP, DV, NB, SK. All authors reviewed and approved of the final manuscript.

Funding

Author(s) disclose no funding sources.

Competing Interests

NB is supported by the Postgraduate Programs of the Department of Biochemistry and Biotechnology, University of Thessaly, and the Hellenic Thoracic Society. Other authors disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests.

Supplementary Files

Supplementary Figure 1

Phylogenetic tree of DEDD. Support values above 50% are indicated at the nodes within the major clades. The scale bar at the upper left denotes the length of amino acid substitutions per site.

Supplementary Figure 2

Phylogenetic tree of EEP. Support values (>50%) are indicated at the nodes within the major clades. The scale bar at the upper left indicates the length of amino acid substitutions per position.

Supplementary Table 1

Phylogenetic distribution of the deadenylases analyzed in the present study.

References

1. Garneau NL, Wilusz J, Wilusz CJ. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol.* 2007;8(2):113–26.
2. Chen CY, Shyu AB. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA.* 2011;2(2):167–83.
3. Thore S, Mauxion F, Séraphin B, Suck D. X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. *EMBO Rep.* 2003;4(12):1150–5.
4. Tucker M, Valencia-Sanchez MA, Staples RR, Chen J, Denis CL, Parker R. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell.* 2001;104(3):377–86.
5. Uchida N, Hoshino S, Katada T. Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. *J Biol Chem.* 2004;279(2):1383–91.
6. Bianchin C, Mauxion F, Sentis S, Séraphin B, Corbo L. Conservation of the deadenylase activity of proteins of the Caf1 family in human. *RNA.* 2005;11(4):487–94.
7. Goldstrohm AC, Wickens M. Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol.* 2008;9(4):337–44.
8. Dlakić M. Functionally unrelated signalling proteins contain a fold similar to Mg²⁺-dependent endonucleases. *Trends Biochem Sci.* 2000;25(6):272–3.
9. Zuo Y, Deutscher MP. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.* 2001;29(5):1017–26.
10. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford).* 2011;2011:bar009.
11. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2011;39(Database issue):D32–7.
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
13. Pettifer S, Thorne D, McDermott P, et al. Visualising biological data: a semantic approach to tool and database integration. *BMC Bioinformatics.* 2009;10 Suppl 6:S19.
14. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004;14(6):1188–90.
15. Hunter S, Jones P, Mitchell A, et al. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* 2012;40(Database issue):D306–12.
16. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics.* 2002;Chapter 2:Unit 2.3.
17. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 2000;17(4):540–52.
18. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;56(4):564–77.
19. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–21.
20. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol.* 2008;25(7):1307–20.
21. Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol.* 2004;21(2):255–65.
22. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics.* 1998;14(1):68–73.
23. 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.
24. Knudsen B, Miyamoto MM. A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins. *Proc Natl Acad Sci U S A.* 2001;98(25):14512–7.
25. Federhen S. The NCBI Taxonomy database. *Nucleic Acids Res.* 2012;40(Database issue):D136–43.
26. Berger L, Speare R, Daszak P, et al. Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proc Natl Acad Sci U S A.* 1998;95(15):9031–6.
27. Daszak P, Berger L, Cunningham AA, Hyatt AD, Green DE, Speare R. Emerging infectious diseases and amphibian population declines. *Emerg Infect Dis.* 1999;5(6):735–48.