Published online 27 September 2008 *Nucleic Acids Research, 2008, Vol. 36, No. 18* ***6021–6034***
doi:10.1093/nar/gkn605

Evolutionary conservation supports ancient origin for Nudt16, a nuclear-localized, RNA-binding, RNA-decapping enzyme

Melissa J. Taylor and Brenda A. Peculis*

Department of Biochemistry, University of Missouri, Columbia, MO 65211, USA

Received August 6, 2008; Revised September 4, 2008; Accepted September 5, 2008

ABSTRACT

Nudt16p is a nuclear RNA decapping protein initially identified in *Xenopus* (X29) and known to exist in mammals. Here, we identified putative orthologs in 57 different organisms ranging from humans to Cnidaria (anemone/coral). *In vitro* analysis demonstrated the insect ortholog can bind RNA and hydrolyze the m^7G cap from the 5′-end of RNAs indicating the Nudt16 gene product is functionally conserved across metazoans. This study also identified a closely related paralogous protein, known as Syndesmos, which resulted from a gene duplication that occurred in the tetrapod lineage near the amniote divergence. While vertebrate Nudt16p is a nuclear RNA decapping protein, Syndesmos is associated with the cytoplasmic membrane in tetrapods. Syndesmos is inactive for RNA decapping but retains RNA-binding activity. This structure/function analysis demonstrates evolutionary conservation of the ancient Nudt16 protein suggesting the existence and maintenance of a nuclear RNA degradation pathway in metazoans.

INTRODUCTION

Modulation of RNA stability *in vivo* is a major regulatory level of control of gene expression in eukaryotes. In the cytoplasm, two major pathways for RNA turnover are known to exist, depending on whether RNA degradation occurs from the 5′-end or from the 3′-end (1–5). Cytoplasmic degradation of RNA from the 5′-end requires the removal of the 5′ m^7GpppG cap that confers stability to mRNAs in the cell. In the cytoplasm this cap-removal activity is performed by the decapping enzyme, Dcp2. Dcp2 is a NUDIX protein with RNA-dependent cap hydrolysis activity responsible for initiating RNA turnover. The Dcp2 protein is conserved among eukaryotes from yeast to human and functions in coordination with a variety of proteins (6–11). Dcp2 is a member of the NUDIX family; these proteins catalyze the hydrolysis of NUcleoside DIphosphates that are linked to some moiety X (12–15). In the cytoplasm, Dcp2 cleaves the cap releasing m^7GDP and leaving a 5′-monophosphate on the 5′-end of the RNA. This decapped RNA is a substrate for the 5′–3′ exonuclease Xrn1. While Dcp2 has the catalytic hydrolysis activity for cap removal, little is known about the specificity directing the binding activity of this protein.

RNA turnover also occurs in the nucleus (16–18). There are nuclear exonucleases with both 5′–3′ and 3′–5′ degradatory activity such as the exosome and Xrn2/Rat1p (19–23). Since 5′-caps are added to RNAs very soon after transcription is initiated (24–28), there must be nuclear decapping machinery that parallels the cytoplasmic machinery. Nudt16, first characterized as X29, is a nuclear-localized NUDIX protein with metal-dependent RNA decapping activity (29–31). It has high affinity and specificity for binding U8 snoRNA based on electrophoretic mobility shift assays (EMSAs) and can be cross-linked to U8 snoRNA (29–31).

The X29/Nudt16 protein is distinct from Dcp2; other than sharing the catalytic NUDIX domain, the two proteins show very little sequence similarity. While both proteins display RNA decapping activity *in vitro*, the decapping repertoire of Nudt16 is broader than that of Dcp2. Dcp2 can use Mg^{2+} or Mn^{2+} to decap m^7GpppG capped RNAs, but X29/Nudt16 protein can utilize Mg^{2+}, Mn^{2+} or Co^{2+} to remove both the m^7G cap or the $m^{227}G$ hypermethylated cap, present on many of the small nuclear-limited RNAs comprising nuclear RNPs (29,31).

The family of NUDIX proteins eliminates potentially toxic nucleotide derivatives from cells. These proteins have been described in bacteria (MutT was a primary member of this family) and over 20 different family members with different cellular functions exist in humans (13,14). Nudt16p was identified in vertebrates but the evolutionary distribution of the protein was not known. Phylogenetic relationships and protein conservation data have been useful in providing functional and structural information which have assisted in assigning biological

*To whom correspondence should be addressed. Tel: +1 573 884 1424; Fax: +1 573 884 4812; Email: peculisb@missouri.edu

roles *in vivo*. Since the information within various available databases has expanded tremendously in the past few years we undertook a search for Nudt16 orthologs using publicly available databases, including those for cDNAs, ESTs, genes and genomes. We identified putative Nudt16 orthologs in 57 different organisms, including vertebrates and invertebrates. Nudt16 ortholog sequences were aligned and characterized by phylogenetic trees, which showed the relationship between organisms with respect to this gene. The phylogenetic trees presented here are consistent with Nudt16 being well conserved in the metazoan lineage (32–34). Biochemical analysis indicated there was a functional conservation of the Nudt16 protein as well; the insect ortholog was active for RNA decapping *in vitro*. The searches also revealed evidence for the gene duplication that resulted in closely related genes on different chromosomes, followed by functional divergence. The paralog of Nudt16p, previously described as Syndesmos (35,36) is a cytoplasmic protein incapable of nucleoside hydrolysis but which can still bind RNA. All metazoans examined here have Nudt16 protein but only a subset have Syndesmos indicating that Nudt16p is the more ancient and functionally conserved protein that plays an essential biological role.

MATERIAL AND METHODS

Identifying orthologs

Putative Nudt16 orthologs were identified using the BLAST algorithm against a variety of databases including GenBank (http://www.ncbi.nlm.nih.gov/), Ensembl (http://www.ensembl.org/index.html) and the Joint Genome Institute (http://www.jgi.doe.gov/). Initially genomes were analyzed with the sequence for X29 (NCBI: NP_001084713.1). As more Nudt16 sequences were identified in other organisms, those sequences were also used to screen each database and identify more diverged Nudt16 orthologs. Detailed sequence information and accession numbers are included in Supplementary Table 1.

Phylogenetic analysis

Sequences were aligned using the ClustalX1.83 alignment program. Alignments were analyzed with the PHYLIP 3.67 package. Initially, Seqboot was used to generate bootstrap values for the alignment generating 100 random data sets. Those data sets were then analyzed by applicable programs. Although all programs available in the package were used to analyze the phylogenetic relationships, reported data were generated using distance algorithms, ProtDist and DNADist. Distance matrices were then analyzed by Fitch (in the PHYLIP package), which uses the Fitch–Margoliash algorithm without a molecular clock to generate possible trees from the data sets. PHYLIP program Consense was used to generate a consensus tree using the anemone sequence as the outgroup for each tree. The trees were visualized by TreeView or PhyFi. Branch lengths are included as a ruler at the bottom of the figures. Bootstrap values are placed at each node where the bootstrap was less than 80 out of 100 trees.

Cloning of cDNAs—sharpshooter Nudt16

Template-encoding full-length cDNA for sharpshooter (NCBI #: DN199400.1) was a generous gift from Wayne Hunter's lab at the USDA in Florida. Deoxyoligonucleotide primers were generated by PCR to clone the cDNA into pET19b in frame with the vector-encoded His-Tag. Primer sequences were: Sense: 5′-CGCGCA TATGTCCTCTGATACAGG-3′ (NdeI) and Anti-sense: 5′-CGCGCTCGAGCTATTATCAGGACTTGAGTGG AAG-3′ (Xho). The restriction site in parenthesis was designed into the primer and assisted with directed, in-frame cloning. The protein was overexpressed in *Escherichia coli* and purified on a sequential Ni-NTA resin (Qiagen, Valencia, CA, USA) followed by Heparin Sepharose or Q Sepharose (Amersham/Pharmacia, Piscataway, NJ, USA)

Cloning of cDNAs—human Syndesmos

Template-encoding full-length cDNA for human Syndesmos (NCBI #: BE799462) was purchased from Open Biosystems, Huntsville, AL, USA (MHS1011-61470). The Clone ID (or IMAGE ID) was 3944355. Deoxyoligonucleotide primers were generated to PCR and clone the cDNA into pET19b in frame with the vector-encoded His-Tag. Primer sequences were: Sense: 5′-CGCGCATATGTCGACGGCGGCGG-3′ (Nde) and Anti-Sense: 5′-CGCGGGATCCCTATTATCAAGAGG AGGCCGGGAGCAAC-3′ (Bam).

The restriction site in parenthesis was designed into the primer and assisted with directed, in-frame cloning. The protein was overexpressed in *E. coli* and purified on a sequential Ni-NTA resin (Qiagen, Valencia, CA, USA) followed by Heparin Sepharose (Amersham/Pharmacia, Piscataway, NJ, USA).

RNA–protein cross-linking

Uniformly ^{32}P-labeled, 4-thioU containing RNAs were transcribed as previously described (30). Cross-linking reactions were performed essentially as described (30). To summarize with modifications included, 0.08 pmol of the 4-thioU RNA was combined with cold competitor tRNA (250-M fold excess), in 45 mM HEPES, pH 8.5. Protein was added in the amounts indicated to yield cross-links: Human: 14.2 pmol; X29: 1.64 pmol; Sharpshooter: 11.8 pmol; Syndesmos: 12.9 pmol. The reaction was incubated at room temperature for 15 min then exposed to 365-nm UV light source at a distance of 3 cm for 15 min on ice. RNase A was then added to each reaction and incubated for 30 min at 37°C. The proteins in the reactions were resolved on a 10% Bis–Tris NuPAGE gel. Dried gels were exposed to Kodak phosphorimager plates to visualize proteins via ^{32}P-label transfer from the RNA.

Protein–protein cross-linking

Protein–protein cross-linking was performed using the water-soluble, reversible cross-linker 3,3′-dithiobis (sulfosuccinimidylpropionate) (DTSSP) (Pierce/Thermo, Rockford, IL, USA) at final concentration of 0.5 mM with 40 pmol of proteins for 30 min. NuPAGE 4 × loading

dye was added to each sample, with or without β-mercaptoethanol (BME) as indicated. Reactions were resolved on NuPAGE gels and detected via coomassie blue staining.

Molecular modeling

The molecular modeling was performed using the DeepView/Swiss-PdbViewer program. The sequences were aligned to the *Xenopus* Nudt16p amino acid sequence, then the modeling program generated corresponding folded molecules based upon the X29 crystal structure (PDB: 2A8P). Single conserved amino acids and a beta sheet were highlighted in the ribbons diagrams to facilitate correct orientation of the structures before electrostatic potential was calculated.

RESULTS

Identification and functional analysis of Nudt16 in vertebrates

We had previously identified the *Xenopus laevis* and human homologs of the Nudt16 protein (29). Alignments of these protein sequences demonstrated that X29/Nudt16 and the putative mammalian orthologs of Nudt16 contain three well-conserved regions, outlined by boxes in Figure 1. (Supplementary Table 1 provides details about sequences and accession numbers.) These three regions were designated the 'hallmarks' of authentic Nudt16 sequences because they were well conserved among the *Xenopus*, human and rat Nudt16 proteins, all of which are functional orthologs [(29,31); rat: our unpublished data].

The three conserved regions were mapped onto the crystal structure of the X29 protein (PDB: 2A8P) (37) to reveal a probable functional basis for each. The sequence outlined in the green box of Figure 1 forms a β-strand passing deep within the interior core of the protein, presumably required for organization and proper folding of the protein. The sequence in the red box is the NUDIX consensus sequence. This functional domain is critical for hydrolysis activity and defines members of the NUDIX family. The third conserved region is the sequence outlined in the blue box in Figure 1. This is the putative protein dimerization domain. In the crystal structure of X29, this domain forms a significant part of the dimerization interface between the monomers in the crystallographic unit of X29 (37). All putative Nudt16 orthologs were required to have all three conserved regions and specifically to contain the catalytic residues within the NUDIX sequence.

Identification of putative Nudt16 orthologs in other organisms

Previous BLAST results revealed Nudt16 existed in *X. laevis* and a variety of mammals, but orthologs were not readily identified in the 'model nonvertebrate' organisms including *Caenorhabditis elegans*, yeast, or *Drosophila melanogaster*. Therefore, Nudt16 was initially proposed to be a vertebrate protein (29). We expanded the search for Nudt16p orthologs using publicly available databases; we have identified putative Nudt16 orthologs in 57 different organisms, both vertebrate and invertebrate (see below). Approximately half of the represented classes are invertebrates. Curiously, exhaustive searches of the *C. elegans*, yeast and *D. melanogaster* genomes have not yet revealed orthologs of Nudt16 on the basis of amino acid sequence alone, but convincing sequences have been identified in other worms and insects. Since these organisms do contain several NUDIX proteins but the Nudt16 proteins are a distinct and identifiable subset, we acknowledge that convincing functional orthologs may not be identified through sequence similarity alone.

Phylogenetic relationships of Nudt16 protein orthologs

To appreciate the relatedness of these proteins phylogenetic trees were generated based on sequence alignments; algorithms were used to generate trees. It was necessary to first identify a standard tree to which all subsequent trees could be compared, since different trees can be generated using different algorithms. We created a 'generic' tree demonstrating phylogenetic relationships of organisms using information from the Tree of Life Web Project (http://www.tolweb.org/tree/). Taxonomic information for each organism was obtained from TaxBrowser on the NCBI website (www.ncbi.nlm.nih.gov) and can be found in Supplementary Table 1. This taxonomic tree was the standard for comparison with all trees generated from sequence alignments; it is provided here for reference (Figure 4).

Potential Nudt16 orthologs were analyzed using several standard phylogenetic programs and methods (38–41). We aligned the sequences of Nudt16 orthologs with ClustalX1.83 and generated trees with the PHYLIP program package (see Methods section). A phylogenetic tree (Supplementary Figure 1) was generated from aligned protein sequences using the anemone (a cnidarian) Nudt16 ortholog as the outgroup. Cnidarians were an appropriate choice for the outgroup because they are thought to be the most ancient of the organisms included in this analysis (42). Sequence alignments were used to generate trees based upon Nudt16p sequence distance, parsimony or maximum likelihood algorithms. Most results matched the predicted phylogenetic relationships. Species within the same class or phylum were clustered on the Nudt16 tree consistent with phylogenetic relationships predicted by other molecular analyses (32,34).

Nud16p orthologs are functional for decapping

Alignments of the Nudt16 orthologs demonstrated that this was a conserved metazoan protein and that all orthologs contained a functional NUDIX domain. Phylogenetic conservation of the nuclear decapping protein implied the existence of a conserved nuclear pathway for RNA turnover; thus it was important to confirm that the sequences identified were functional orthologs. Since we had previously characterized functional (decapping) activity in several vertebrates, it was desirable to examine a nonvertebrate protein and chose an insect ortholog for this purpose. The glassy-winged sharpshooter (*Homalodisca coagulata*) was selected as a representative insect because the sequence of a full-length cDNA could be obtained. This insect is of economic interest because it is responsible for spreading the bacteria causing Pierce's disease in grapevines and citrus trees (43).

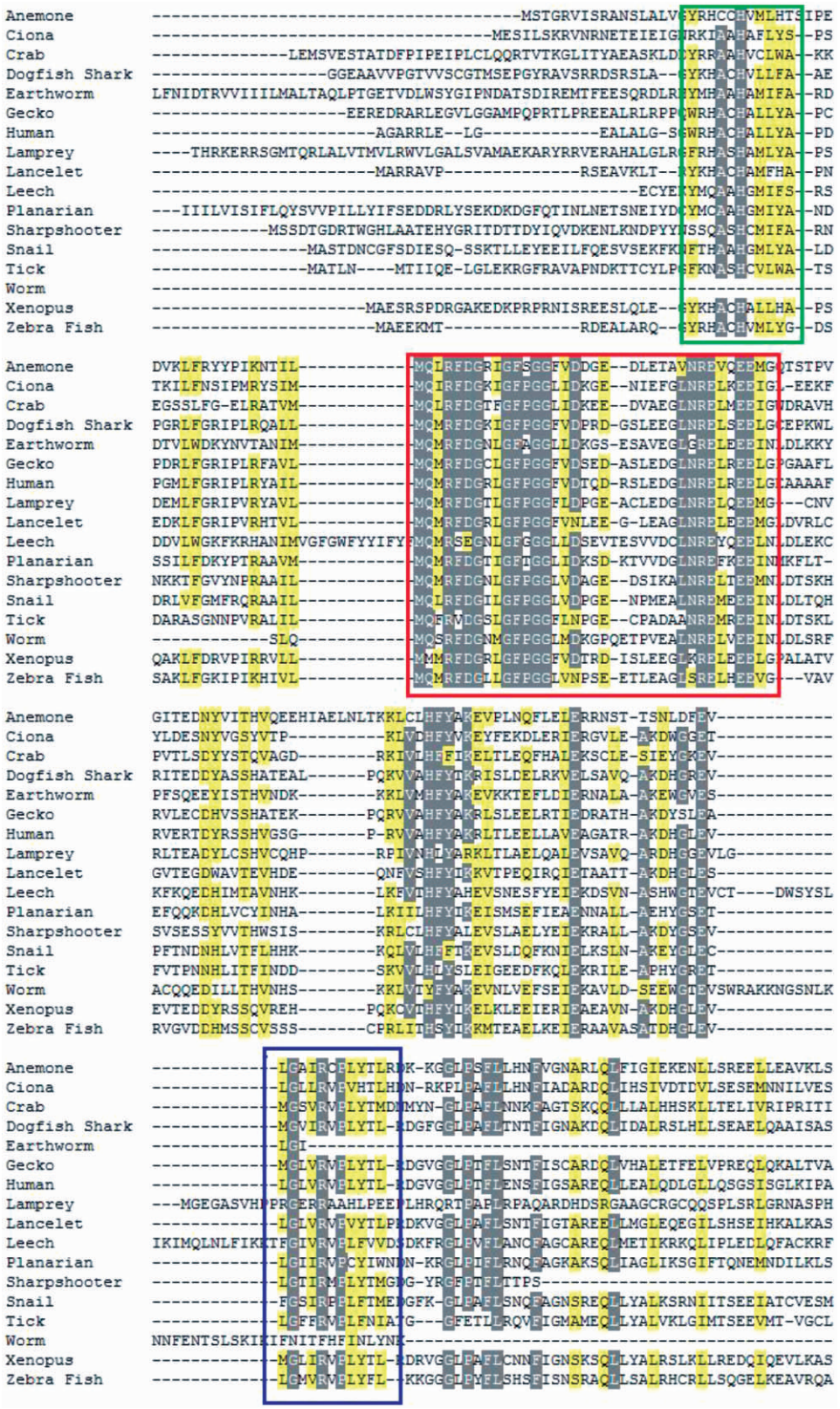

Figure 1. Amino acid alignment of putative Nudt16 orthologs from a sampling of the organisms found to have this protein. The alignment reveals three areas of highly conserved sequence, indicated by the green, red and blue boxes. A functional role based on structure for each conserved region is described in more detail in the text. To be designated as a Nudt16 homolog, sequences had to have the NUDIX domain (in red box) and the other two domains. Gray residues indicate identity; yellow highlighted residues show similarity. The accession number for each sequence used is listed in Supplementary Table 1.

We amplified the region encoding the predicted *H. coagulate* (sharpshooter) Nudt16p ortholog via PCR, and cloned the product into an expression vector (see Methods section). The His-tagged fusion protein was overexpressed in bacteria, purified and used for biochemical analysis. Decapping reactions combined ^{32}P-cap-labeled U8 snoRNA, metal and purified protein, as in our standard assay (29,31), product formation was examined via thin layer chromatography (Figure 2). The results indicated the sharpshooter Nudt16p ortholog is active for cap hydrolysis, releasing m^{7}GDP from the 5′-end of U8 cap-labeled RNA in a metal-dependent manner. Two forms of the human Nudt16 protein and the *Xenopus* protein are present as positive controls. The product of the decapping reaction was confirmed to be m^{7}GDP by incubation of the reaction product with nucleotidediphosphate kinase (29); the kinase activity altered the mobility of this material which then comigrated with m^{7}GTP (data not shown). Similar to the other Nudt16 orthologs previously characterized by this lab, decapping activity by the insect protein was higher in the presence of Mn^{2+} than in Mg^{2+} or Co^{2+}, but all three metals were capable of mediating cap hydrolysis from U8 snoRNA (data not shown). Protein concentrations were optimized for the decapping assays; the amount of protein required for 50% decapping within 10 min was comparable to that in the better characterized vertebrate orthologs. Thus, the insect Nudt16p protein was catalytically active for decapping and a true functional ortholog of the vertebrate nuclear decapping protein (Figure 2).

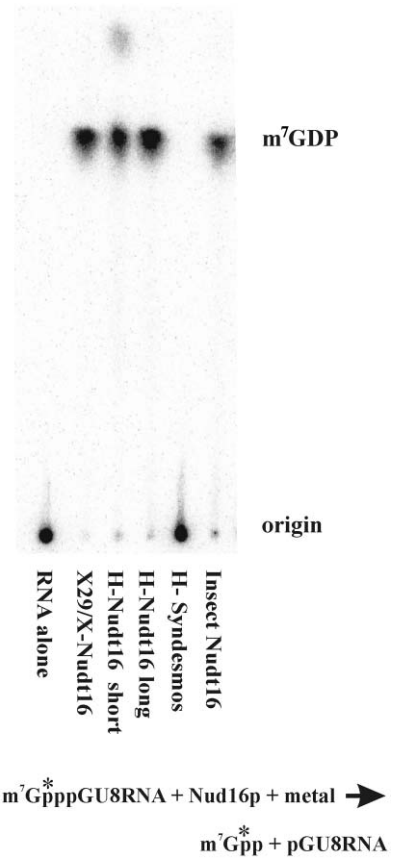

Figure 2. Insect Nudt16 protein is active for decapping RNA. Cap-labeled U8 RNA was incubated in the presence of buffer, Mn^{2+} and protein (as indicated) for 30 min at 37°C (reaction indicated below). Samples of the reactions were spotted on a TLC, which was developed and then visualized on a phosphorimager. The *Xenopus* (X29/X-Nudt16) protein, present as a positive control, released m^{7}GDP from the cap-labeled RNA. The two human proteins, which varied by an unrelated amino terminal extension, were both efficient at decapping the U8 RNA. Insect (sharpshooter) Nudt16 protein hydrolyzed the RNA to release the m^{7}GPD cap comigrating with that cleaved by the other orthologs. The human H-Syndesmos protein displays no decapping activity.

Identification of closely related non-NUDIX proteins in mammals and birds

In addition to identifying functional orthologs of Nudt16p in metazoans, the BLAST searches also identified a closely related family of vertebrate proteins called 'Syndesmos' (Table 1 and Figure 3, and Supplementary Table 1 and Supplementary Figure 3). Syndesmos was first reported in chicken and further characterized in human. Syndesmos is a cytoplasmic protein that interacts with Syndecan-4 and the Paxillin family of proteins, including Hic-5 (35,36). When overexpressed in human tissue culture cells Syndesmos localizes to the ventral plasma membrane and enhances cell spreading and focal contact formation (36). There are no *in vivo* functional data describing a biological role for Syndesmos.

Nucleotide (Supplementary Figure 3) and amino acid (Figure 3) alignments of the Nudt16 proteins and the Syndesmos proteins demonstrated that they were closely related but were distinct protein families. Within mammals, overall the two proteins were 50–60% identical and 60–75% similar at the amino acid level (Table 1). The highest identity occurred within the three regions that were conserved among the Nudt16p proteins (see Figure 1 and compare with the partial sequence shown in Figure 3). Syndesmos maintains the domain that constitutes the center of the Nudt16 protein (green box, Figure 1, not shown in Figure 3) and also contains the 'interface interaction' domain (blue boxes, Figures 1 and 3) found in Nudt16. Interestingly, in Nudt16, the interaction interface is the contact surface for the two monomers seen in the X-ray crystal structure (37). In Syndesmos, this protein domain is reported to be involved in facilitating interactions with other cytoplasmic proteins (35).

The Nudt16 proteins were easily distinguished from Syndesmos orthologs by critical sequence changes within the catalytic NUDIX domain (red box, Figures 1 and 3). While the NUDIX-like domain is well conserved in Syndesmos compared to Nudt16p (over 90%), the Syndesmos proteins contained a pair of glycine–leucine repeats (outlined in the orange box in Figure 3) in place of four amino acids that include the two glutamic acid residues required for catalysis in Nudt16 proteins (outlined in the heavy red box of Figure 3) (12,13,15,29,44). Thus the cytoplasmic Syndesmos family of proteins is clearly a paralog of the catalytic nuclear Nudt16p family of proteins, but was predicted to lack decapping activity. When used in a decapping reaction it demonstrated no cap hydrolysis activity (Figure 2).

Table 1. Conservation and divergence of Nudt16 and Syndesmos proteins

Organism	S:Nud Percent identity (%)	S:Nud Percent similarity (%)	humNudt versus Nudt (%)	HumSynd versus Synd (%)	Chromo# Nudt	Chromo# Synd
Bat	44	54	73	49	–	–
Bovine	62	73	82	97	–	25
Chimpanzee	61	70	99	70	3	16
Dog	53	60	78	88	23	6
Galago	50	60	86	56	–	–
Horse	51	60	40	64	16	–
Human	61	75	100	100	3	16
Macaque	64	74	49	85	–	–
Mouse	58	72	79	95	9	16
Opossum	54	69	65	92	X	6
Pig	57	68	66	84	–	–
Platypus	48	58	51	83	–	–
Rabbit	60	69	57	40	–	–
Rat	60	73	68	97	8	10
Rhesus Monkey	64	74	71	99.5	2	16
Averages	56	67	71	80		

Columns 2 and 3 provide the percent identity and similarity (respectively) of Nudt16 compared to the Syndesmos protein within each organism indicated in Column 1 (i.e. bat Nudt16 compared to bat syndesmos). The percent identity of the Nudt16 in the indicated organism when compared to human Nudt16 is shown in Column 4. The percent identity of Syndesmos in the indicated organism compared to human Syndesmos is shown in Column 5. These values were calculated by BL2SEQ (53) or MultiView Alignment Display (54) using the Biology Workbench application (workbench.sdsc.edu). Columns 6 and 7 indicate the chromosomal location of the genes-encoding Nudt16 or Syndesmos, respectively, obtained through GenBank and Ensembl. Some organisms have not been fully annotated or chromosome information was not available. X indicates the X sex chromosome.

Phylogenetic relationship between Nudt16p and Syndesmos

To examine the phylogenetic relationship between Syndesmos and Nudt16p, we aligned the sequences of these proteins and generated phylogenetic trees from the available data. Although the nucleic acid sequences were used in most of these analyses, Figure 5 shows the amino acid sequences of the representative subset of those organisms listed in Supplementary Table 1. The trees did not differ significantly when nucleotide sequences were aligned. When arranged by degree of sequence divergence, the data indicated that the Syndesmos orthologs were grouped together as a distinct branch, diverging from the amniota branch of the Nudt16 tree (Figure 5). Figure 4 is a schematic indicating the branches where Syndesmos (S) and Nudt16 (N) proteins have been found on a more simplified phylogenetic tree.

One striking observation arising from the alignment of Nudt16p and Syndesmos paralogs is that to date, Syndesmos has been reported only in birds and mammals (Figure 4B; Supplementary Materials). Our analysis identified several cases of incorrectly annotated proteins ('NUDIX-like' proteins that are Syndesmos and vice versa). There are birds, amphibians, reptiles and mammals for which only one paralog has been annotated. Since Nudt16p is found across metazoans, we presume that incomplete coverage/sequencing is the explanation where only Syndesmos had been identified. The salamander sequencing project (www.abystoma.org) yielded a partial salamander Syndesmos sequence. Similarly, sequencing of PCR products amplified from chicken genomic DNA demonstrated that chickens have the Nudt16 gene (M.J.T. and B.A.P.; data not shown). Clearly, another possible explanation for organisms where Nudt16p is present but the Syndesmos paralog is absent is that the organism (or its evolutionary predecessor) never underwent the gene duplication event or that stochastic genetic loss has occurred.

Phylogenetic relationships at the nucleic acid level

The genetic code is redundant; most amino acids are encoded by more than one triplet. Thus comparisons of proteins on the amino acid level may mask some of the evolutionary drift or changes that occur at the genomic level (45–47). Since we are interested in the phylogeny and determining the relatedness of these genes across a variety of organisms, we also performed phylogenetic comparisons at the nucleic acid level. The programs within the PHYLIP package were used, including DNAdist, DNAml, DNAmlk and DNApars, to obtain phylogenetic trees using alignments of the nucleic acid sequences with the same methods and constraints that were used for the proteins. Different programs resulted in very slight rearrangements of the branches (compared to the amino acid trees), but demonstrated a similar phylogeny. Vertebrates were separated from the invertebrate organisms and Syndesmos appeared at the same place on the trees generated from either the amino acid or the nucleotide sequences.

While single nucleotide polymorphism is typically a standard for determining relatedness between organisms, both nucleotide and amino acid alignments were examined because of the very broad spectrum of organisms compared. The ability to obtain similar trees despite the diversity of the organisms strengthens the conclusions about relatedness and demonstrates the ancient origin of the Nudt16 protein and the relatively recent appearance of Syndesmos.

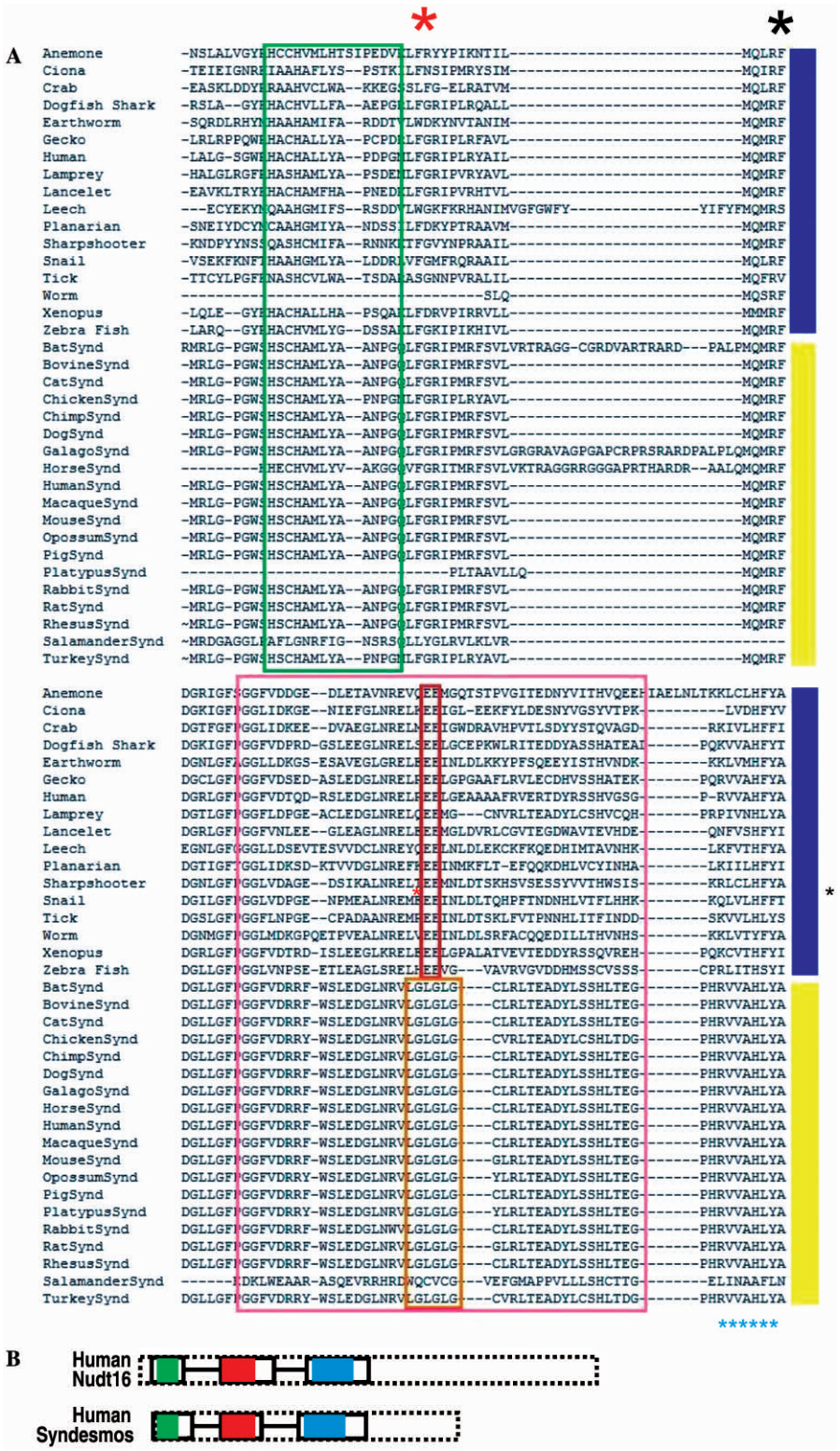

Figure 3. Nudt16 and Syndesmos are closely related members of a gene family. **(A)** Alignment of a portion of the Syndesmos protein from 17 organisms (denoted by yellow box on right edge) and the corresponding region of Nudt16 from 10 organisms (marked with the blue box on right edge). All Syndesmos orthologs have a repeated glycine/leucine sequence (orange box) in place of the glutamic acid residues (heavy red box) required for catalysis in the NUDIX domain (red box). The green box is the conserved region in Nudt16 proteins, equivalent to the green box in Figure 1. Syndesmos paralogs have all three conserved regions but lack a functional NUDIX domain. The central parts of the proteins are aligned here. The accession numbers are in Supplementary Table 1. (**B**) Genomic organization of the orthologs in human. Dashed boxes indicate the transcription unit, solid boxes are exons, horizontal lines are introns and the conserved domains are color coded as per Figure 1. Note Nudt16 has a longer 5′-UTR and 3′-UTR than Syndesmos.

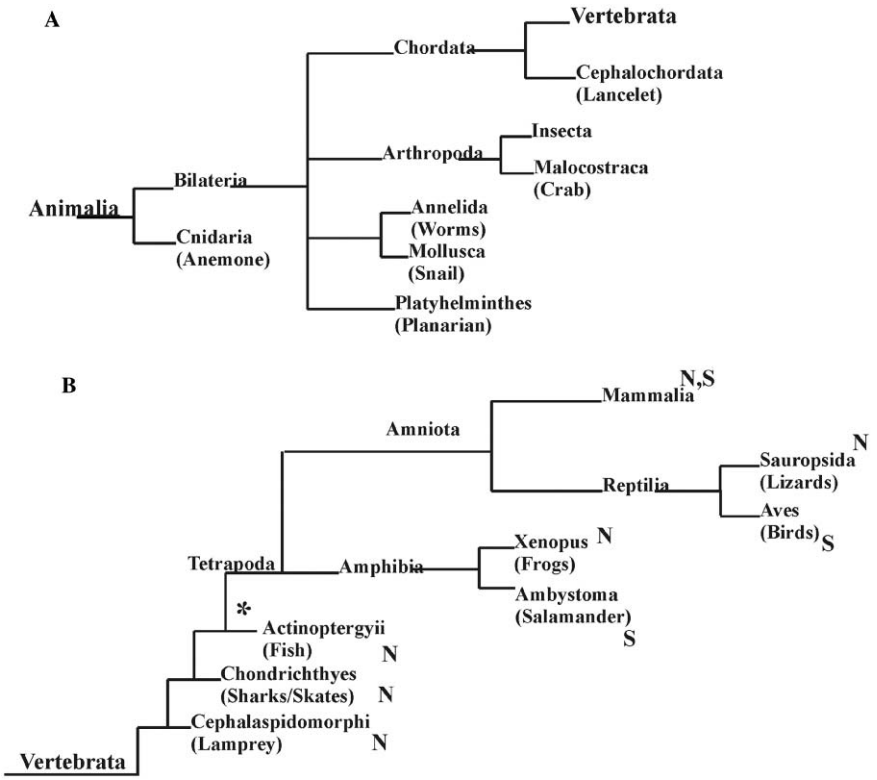

Figure 4. A 'standard' phylogenetic tree. This tree illustrates the relative position of the organisms examined here and is provided as a quick reference for taxonomic relationships, based on the Tree of Life Web Project. (**A**) The Kingdom Animalia/Metazoa was searched. Only those branches for organisms containing a Nudt16 ortholog are shown. (**B**) An expansion of the vertebrate branch reveals the conservation of Nudt16 and appearance of Syndesmos. The 'N' indicates branches with organisms containing Nudt16 orthologs, 'S' indicates branches containing Syndesmos and 'asterisk' indicates the likely gene duplication event; acknowledging the possibility that it may have occurred earlier with high rates of subsequent gene loss. Sequences 'below' the 'asterisk' have only the Nudt16 protein.

Nudt16 and Syndesmos are related by gene duplication, not RNA processing

Searches of the genomic databases provided additional supportive data about the gene organization and chromosomal localization of the protein paralogs. In mammals, the two proteins are encoded by single copy genes located on distinct chromosomes. Both genes contain two introns; the relative location of each intron was conserved between the two gene families (Figure 3B). Each exon contains one of the functional/conserved domains identified in Figure 1. The conserved intron positions indicate that the loss in Syndesmos of the amino acids critical for cap hydrolysis was not be due to alternative splicing but from mutation of the DNA (see Figure 3). Likewise, the nucleotide sequences indicate that the sequence changes altering the amino acids required for catalysis in the NUDIX domain cannot be due to frame shifting or an alternative reading frame. Table 1 lists the chromosomal locations of the genes for these two proteins. In the cases where chromosomal synteny is known, the genes have maintained similar relative positions (human and chimp).

Functional similarity between Nudt16p and Syndesmos

To assess which (if any) of the functional biochemical characteristics of Nudt16p proteins were shared by the Syndesmos paralogs, we cloned the human Syndesmos protein. The Syndesmos protein sequences lacked the amino acids critical for catalysis, so they were not expected to have decapping activity but the other properties could be tested. We used PCR to amplify the coding region of the human Syndesmos cDNA. The purified protein was assayed for decapping activity, ability to homodimerize and for RNA binding. Since Syndesmos proteins lack the amino acids critical for cap hydrolysis, our results indicating that the human Syndesmos protein did not have decapping activity (Figure 2) were not surprising.

Xenopus Nudt16p protein forms a homodimer in solution, readily visible in protein–protein chemical cross-linking assays (31) and in the X-ray crystal structure. It is not yet clear whether the monomer or dimer is the functional form of the protein. The human Nudt16p protein can also form a dimer but less readily; a smaller portion of the human Nudt16p protein is involved in dimer formation (Figure 6A) demonstrated by more of the monomer form being present despite incubation with cross-linker relative to the *Xenopus* protein. Human Syndesmos did not form detectable homodimers under the conditions used in this assay (Figure 6A).

MOLECULAR MODELING

Since Nudt16p and Syndesmos displayed differential ability to form dimers in the cross-linking assay, the question

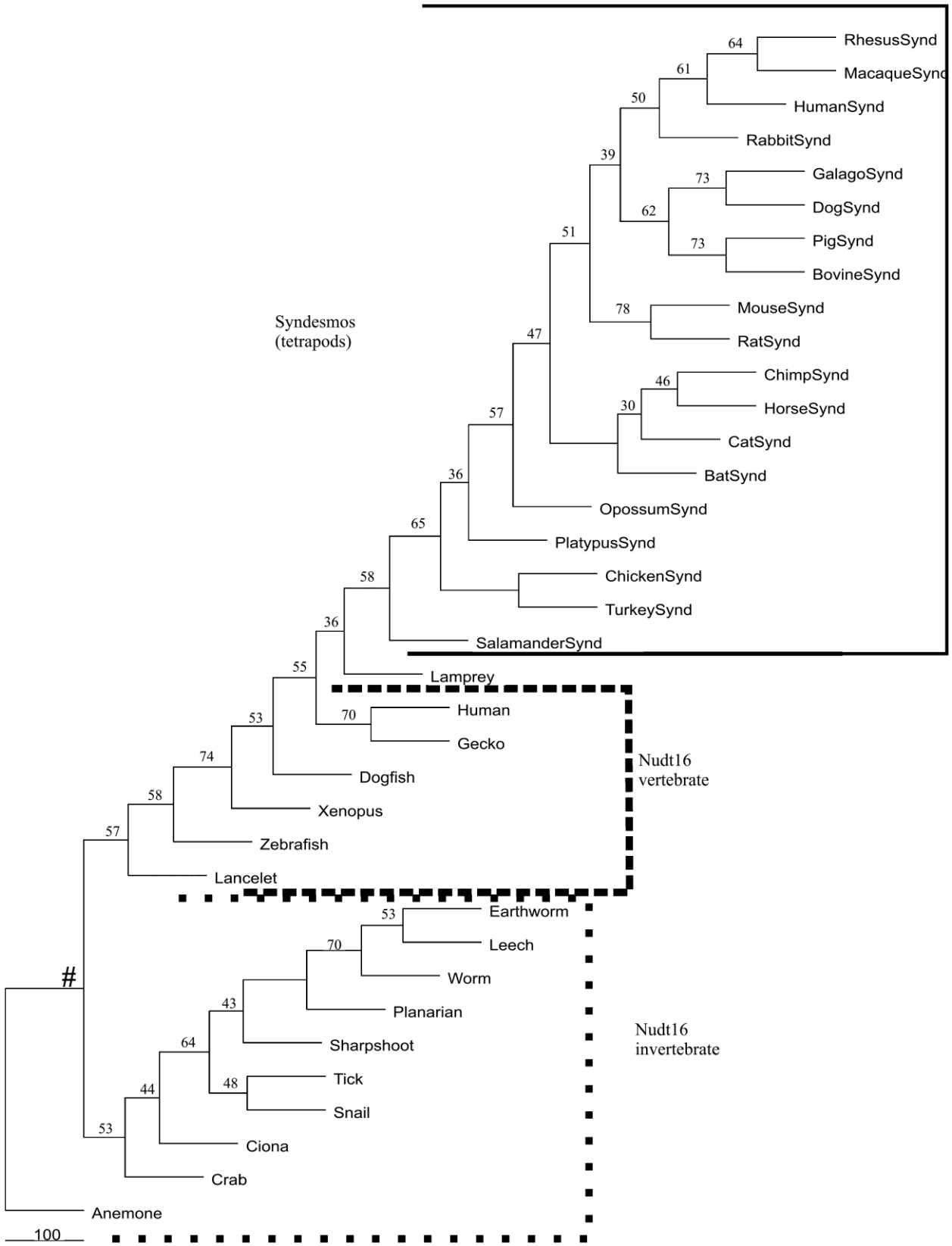

Figure 5. A tree of Nudt16 and Syndesmos identifies them as paralogs. Amino acid sequences were aligned, graphed and analyzed as in Methods section. The tree was visualized with TreeView. Bootstrap values are shown. Vertebrates and invertebrates are separated and phylogenetic relationships are comparable to the generic tree of evolution. Syndesmos sequences branch from the tetrapod lineage of the Nudt16 sequences.

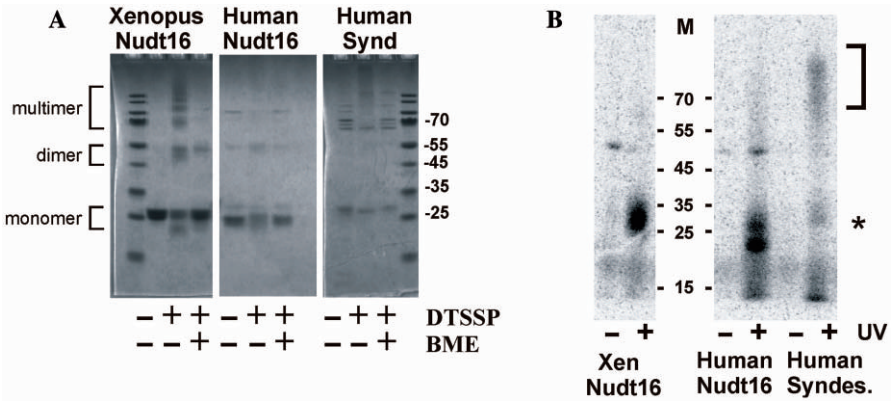

Figure 6. Functional comparison of paralogous proteins. (**A**) Chemical cross-linking to examine homodimer formation. Over-expressed, purified proteins were untreated or incubated with the reversible chemical cross-linker DTSSP (as indicated). One half of the treated samples were heated in the presence of reducing agent, BME to reverse the cross-link. All samples resolved on a NuPAGE gel. The 30 kDa *Xenopus* protein cross-linked to form a ~60 kDa dimer and larger multimers, as indicated. These resolved to monomers under reducing conditions. Human Nudt16 formed a dimer which also resolved to a monomer under reducing conditions. Human Syndesmos did not appreciably alter mobility in the presence of the cross-linker, although some of the higher molecular weight contaminating proteins did alter mobility. Molecular weight markers are present on the far right and left with values indicated on the right. (**B**) RNA–protein cross-linking examined RNA binding. RNA and protein were incubated, then exposed (+) or not (−) to UV. After RNaseA digestion samples were resolved on NuPAGE gels and exposed to a phosphorplate. U8 snoRNA was cross-linked to *Xenopus* Nudt16 protein in a UV-dependent manner, resulting in labeled protein at 30 kDa. Human Nudt16 forms UV-dependent cross-links, migrating slightly faster, consistent with the smaller size of the human protein (see panel A). The human Syndesmos protein forms a faint UV-dependent cross-link migrating at 30 KDa, indicated by the asterisk and a higher order cross-link (indicated by the bracket) over 70 kDa. Molecular weight markers (M) are as indicated.

of electrostatic surface charges of each of the proteins of interest was explored by molecular modeling. Amino acid sequences for the human and sharpshooter Nudt16 and the human Syndesmos protein were aligned to the *Xenopus* Nudt16p (X29) sequence (PDB: 2A8P). Conserved amino acids or short sequences were used to ensure correct alignment and orientation of the modeled structures (Figure 7A). The blue strand is the CVTHFY sequence conserved in the Nudt16p sequences (see Figure 3, blue asterisks). The red residue was F49 in *Xenopus*, a well-conserved residue that is located very near the catalytic site (red asterisk in Figures 3 and 7A). A second phenylalanine residue near the active site was highlighted in black (black asterisk at F62 in X29 in Figures 3 and 7A). The Nudt16 and Syndesmos proteins were modeled as monomers. X-ray crystal data revealed the X29 protein was a dimer, but the structure is shown with a yellow line to indicate the dimer interface. The monomer unit of the Nudt16p protein above the yellow line is positioned and aligned with the orthologs in the ribbon diagrams (Figure 7A). Once the molecules were rotated to correctly orient them relative to X29, the electrostatic potential surface charge was calculated for each structure (Figure 7B).

The alignments shown by ribbons in Figure 7A provided further evidence that the proteins are orthologs/paralogs and closely related. The calculated electrostatic potential of X29 (Figure 7B) indicated that the protein was highly polarized with the predicted RNA-binding site on the positively (blue) charged face (37). The predicted surface of the human Nudt16p protein was not as extensively charged but positive patches were present. The corresponding face of the Sharpshooter Nudt16 protein was even less charged but there was a strongly positively charged patch just proximal to the NUDIX

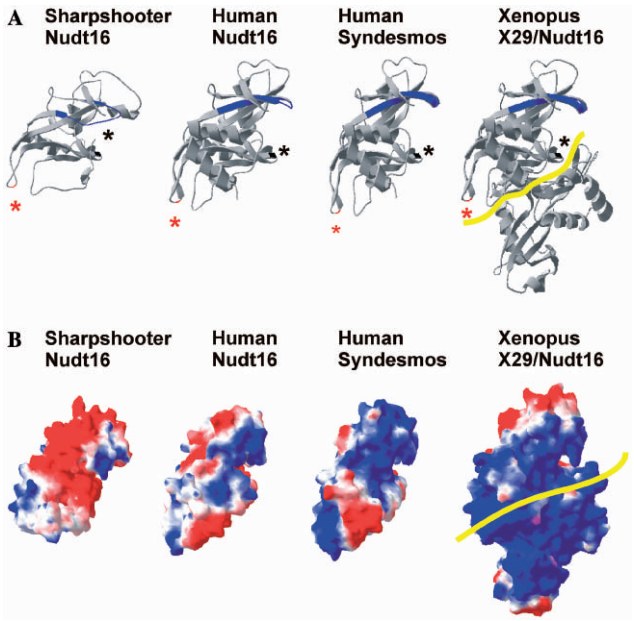

Figure 7. Structural comparison of paralogous proteins. (**A**) Molecular modeling of the orthologs and paralogs of *Xenopus* Nudt16p. The Xray crystal structure of *Xenopus* X29/Nudt16p (PDB: 2A8P) was used to model the human and sharpshooter Nudt16p proteins and human Syndesmos. Selected conserved residues were highlighted in the ribbon models to correctly align the models relative to each other. The blue strand (CVTHFY) is indicated with blue asterisks in Figure 4. The black asterisk is a conserved Phe residue near the catalytic site, while the red asterisk is a second conserved Phe residue indicated in Figure 4. The yellow line in the *Xenopus* protein is oriented across the dimer interface; the monomer above the line is oriented and positioned relative to the other three proteins. (**B**) Electrostatic charge potential of structures. Once the proteins were correctly oriented, electrostatic charge potential was calculated. Blue is positively charged surface and red is negatively charged. The yellow line in the *Xenopus* protein is oriented across the dimer interface as in (A).

site where the RNA would bind. Interestingly, the human Syndesmos protein had an electrostatic charge potential that most resembled that of the *Xenopus* Nudt16p protein monomer, X29.

PROTEIN–RNA CROSS-LINKING

Another characteristic of the *Xenopus* protein was that it could efficiently bind U8 snoRNA, as evidenced both by EMSA and by UV-mediated RNA–protein cross-linking (30). The *Xenopus* Nudt16p protein was first found because of its high affinity for U8 snoRNA and an ability to generate an EMSA (30). However with the possible exception of the weak affinity demonstrated by the rat ortholog (M.J.T. and B.A.P.; data not shown), no other Nudt16p protein has been found to have sufficiently high affinity for detectable complex formation in EMSA.

Another method for examining U8 RNA–X29 protein interactions involved the use of 4-thioU-labeled RNA (30). This resulted in a ~30 kDa UV-dependent cross-link. The RNA–protein cross-link assay was performed using the human and sharpshooter Nudt16 proteins and the human Syndesmos protein. Proteins were pre-incubated with U8 RNA labeled with 4-thioU and ^{32}P-rATP. The reaction was then exposed to UV light to cross-link molecules interacting within 4–10 Å. Samples were treated with RNase A, resolved on SDS gels and exposed to phosphorplates. The expected 30 kDa cross-link was seen for the *Xenopus* protein (30) (Figure 6B). The human Nudt16p protein is slightly smaller; a UV-dependent cross-link was detected migrating near 25 kDa. The sharpshooter protein also generated a UV-dependent cross-link at the predicted size. The human Syndesmos protein generated a UV-dependent RNA–protein cross-link at a size consistent with the full-length Syndesmos protein. The cross-links were not seen in comparable reactions containing large amounts of bovine serum albumen and could be competed with 100-M fold excess unlabeled U8 RNA or 2000-M fold excess tRNA, but not by lower amounts of tRNA (data not shown).

Collectively these data indicated that the Syndesmos protein and Nudt16 protein are closely related. Syndesmos, like the Nudt16p protein orthologs, can bind RNA. Unlike the Nudt16p proteins, the Syndesmos protein cannot decap RNA (Figure 2) nor form homodimers under the conditions used here (Figure 6).

DISCUSSION

We found orthologs of Nudt16 in 57 metazoans, both invertebrates and vertebrates. All putative orthologs contained an intact NUDIX domain plus the two additional structural/functional regions that are conserved among vertebrate Nudt16 proteins. Phylogenetic trees were generated from sequence alignments of the orthologs. Overall, the trees were consistent with the accepted phylogeny: the division between vertebrates and nonvertebrates was apparent, the two Cnidarians were grouped together closest to the ancient ancestor and the mammals were grouped with the marsupials and monotremes, being more closely related than to the eutherian mammals.

There were a few notable exceptions in the placement of some organisms on the tree. Some of these sequences were known to have partial or incomplete cDNA or EST sequences. The others possibly resulted from imprecise manual editing of whole-genome shotgun sequences. Trees generated from the alignment of protein orthologs do not always mirror trees obtained by phenotypic traits or rRNA sequence (34,48–51). Sequence drift is informative, however, as it can identify the natural sequence variation tolerated at any single amino acid residue since the selective parameter has the ability to function efficiently *in vivo*. The low bootstrap values, represented by the numbers on the schematics, reflect the exceedingly wide variety of organisms being aligned and compared (most of one of the three arms of the Tree of Life). As such, these values are not inconsistent with other proteins that are conserved across such a wide diversity of organisms.

FUNCTIONAL ORTHOLOGS OF NUDT16

Nudt16 orthologs were identified in many organisms. A representative of the insects, the Sharpshooter Nudt16 protein, was cloned, purified and characterized here. The insect Nudt16 protein was shown to be active for both RNA binding and 5′-cap hydrolysis *in vitro*, demonstrating it to be a functional ortholog of the vertebrate nuclear decapping proteins.

To date, convincing orthologs have not been identified in yeast or plants; this is not to say the protein does not exist there. Database searches indicated that Nudt16 was rapidly diverging at the amino acid level; true functional homologs may not be distinguished by sequence comparisons alone. Putative orthologs in other species may have been missed in these searches because of their high dissimilarity in regions that we would have allowed to be 'variable', but which the programs disallowed because of the stringency of the search parameters employed. On-going and future searches with custom search parameters dictating high stringency requirements within the 'conserved' regions and lesser stringency in 'variable' regions may lead to the identification of orthologs of Nudt16 in other organisms. Functional homologs may be more effectively identified through structural similarity and validated in *in vitro* functional assays.

SYNDESMOS IS A PARALOG OF NUDT16P

Previous database searches for Nudt16-related proteins identified the avian and mammalian protein Syndesmos alternately (and sometimes incorrectly) annotated in databases as both Syndesmos- and Nudt16-like. This work demonstrates Syndesmos is closely related to but differs from Nudt16 primarily in that Syndesmos is not a NUDIX protein (Figure 3), and is reported to be cytoplasmic.

The similarity between Nudt16 and Syndesmos protein sequence, gene organization and evolutionary distribution suggests that syndesmos probably originated as a gene

duplication event of the more ancient Nudt16. Subsequent divergence of the Syndesmos protein allowed it to acquire new cellular functions (due to or perhaps assisted by the loss of the NUDIX domain). Data supporting this hypothesis include the different chromosomal location of the genes but shared number and placement of introns within the coding region (Table 1 and Figure 3B). The presence of introns makes retroposition via an RNA intermediate unlikely. Intron/exon borders that flank but do not interrupt the NUDIX domain makes alternative splicing less likely (Figure 3B). Finally, careful sequence analysis indicated that neither an alternate splicing nor frame shifting are feasible explanations for the origin of the two proteins from a single gene. Thus, Syndesmos arose from a gene duplication of the more ancient Nudt16 hydrolase gene early in the tetrapod line, resulting in the cytoplasmic Syndesmos protein.

The X-ray crystal structure of the *Xenopus* X29/Nudt16p protein (37) demonstrated that the residues comprising the conserved Domain 3 (in blue on Figures 1 and 3) were involved in the dimerization interface. The comparable, conserved region in the Syndesmos protein is involved in the interactions with the Paxillin family of proteins and Syndecan-4 *in vivo* (35,36). Thus, it is likely that while the *Xenopus* Nudt16 can homodimerize *in vitro*, this interaction surface may be involved in heterologous protein interactions *in vivo*. Additional insights into particle assembly and *in vivo* binding partners could be gleaned from examining the structural interactions of Syndesmos and its *in vivo* binding partners.

The human Syndesmos protein, like the Nudt16 proteins, displays RNA-binding activity, although we have not yet mapped the residues involved in RNA interactions on either protein. The biological role—if any—of the RNA-binding activity for Syndesmos is not known, in part because the *in vivo* function of Syndesmos is not clearly defined.

This phylogenetic study provided strong evidence that the Nudt16 gene is the more ancient protein, and that Syndesmos is a newer acquisition, having lost both the catalytic residues in the NUDIX domain and its nuclear/nucleolar location. The lack of Syndesmos in reptiles, such as the gecko and the anole, was somewhat surprising and may reflect a lack of data (rather than a 'real' evolutionary trait) (52) or may indicate a stochastic loss of the Nudt16 gene in some—or perhaps all organisms in some branches of the evolutionary tree. As more genomes are fully sequenced and released, it will be interesting to discover whether other reptiles and amphibians have a Syndesmos ortholog.

AN EVOLUTIONARY CONSERVED NUCLEAR DECAPPING MACHINERY

Our early studies identified the X29/Nudt16 protein through its ability to bind U8 snoRNA with high affinity and specificity (30). Once cloned and over-expressed, the nuclear localization and decapping activity of the protein was discovered and a role in negative regulation of ribosome biogenesis was proposed (29). Recent studies have demonstrated that several mammalian orthologs (29,31) and the insect ortholog of this protein (this work) can remove the m^7G cap from a variety of RNA substrates in a metal-dependent manner. The nuclear enzyme has broader substrate specificity.

This work was initiated to determine the extent of evolutionary conservation of the Nudt16 protein. The findings revealed unexpected information about the divergence of an ancient protein family. The cytoplasmic protein Syndesmos, present in many vertebrates, was demonstrated to have RNA-binding activity. Syndesmos is a paralog of the more ancient nuclear hydrolase Nudt16 based on sequence and functional data. How—or whether—this RNA-binding activity is used by the protein *in vivo* remains to be demonstrated.

The distribution of Nudt16 across metazoans implies an evolutionary conserved biological role for this protein. This nuclear decapping protein has the ability to remove the m^7G cap (present on pre-mRNAs and mature mRNA prior to export) and the $m^{227}G$ cap (present on nuclear-limited snRNAs and snoRNAs). Thus it is likely to be the catalytic component in an evolutionary conserved (but as yet uncharacterized) nuclear pathway present in all metazoans. Nudt16 may act upon a variety of RNA substrates including the stable, nuclear-limited RNAs (snoRNA, snRNA and other hypermethylated RNAs) and m^7G-capped pre-mRNAs and mRNAs prior to their export to the cytoplasm giving it a very early position in the complex pathway for the control of gene expression. These evolutionary data have facilitated new analyses of the structure–function relationship within the protein family and can provide additional insight into other biological roles for these proteins.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Wayne Hunter for his generous donation of the construct-encoding Sharpshooter Nudt16p and Chad Raw for assistance in cloning. We are grateful to Kristen Reynolds and Frank Schmidt for scientific discussions and to Frank Schmidt for critical reading of the article.

FUNDING

National Institutes of Health training (grant T90 DK070105 to M.J.T.); an National Science Foundation (grant #0718256 to B.A.P.); internal funding from MU (to B.A.P.). Funding for open access charge: NSF grant #0718256 to B.A.P.

Conflict of interest statement. None declared.

REFERENCES

1. Decker,C.J. and Parker,R. (1994) Mechanisms of mRNA degradation in eukaryotes. *Trends Biochem. Sci.*, **19**, 336–340.

2. Coller,J. and Parker,R. (2004) Eukaryotic mRNA decapping. *Annu. Rev. Biochem.*, **73**, 861–890.
3. Gu,M., Fabrega,C., Liu,S., Liu,H., Kiledjian,M. and Lima,C. (2004) Insights into the structure, mechanism, and regulation of scavenger mRNA decapping activity. *Mol. Cell*, **14**, 67–80.
4. Meyer,S., Temme,C. and Wahle,E. (2004) Messenger RNA turnover in eukaryotes: pathways and enzymes. *Crit. Rev. Biochem. Mol. Biol.*, **39**, 197–216.
5. Parker,R. and Song,H. (2004) The enzymes and control of eukaryotic mRNA turnover. *Nat. Struct. Mol. Biol.*, **11**, 121–127.
6. Dunckley,T. and Parker,R. (1999) The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. *EMBO J.*, **18**, 5411–5422.
7. Dunckley,T., Tucker,M. and Parker,R. (2001) Two related proteins, Edc1p and Edc2p, stimulate mRNA decapping in Saccharomyces cerevisiae. *Genetics*, **157**, 27–37.
8. Van Dijk,E., Cougot,N., Meyer,S., Babajko,S., Wahle, E. and Seraphin, B. (2002) Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J.*, **21**, 6915–6924.
9. Wang,Z., Jiao,X., Carr-Schmid,A. and Kiledjian,M. (2002) The hDcp2 protein is a mammalian mRNA decapping enzyme.[comment]. *Proc. Natl Acad. Sci. USA*, **99**, 12663–12668.
10. Piccirillo,C., Khanna,R. and Kiledjian,M. (2003) Functional characterization of the mammalian mRNA decapping enzyme hDcp2. *Rna*, **9**, 1138–1147.
11. Bail,S. and Kiledjian,M. (2006) More than 1 + 2 in mRNA decapping. *Nat. Struct. Mol. Biol.*, **13**, 7–9.
12. Bessman,M.J., Frick,D.N. and O'Handley,S.F. (1996) The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J. Biol. Chem.*, **271**, 25059–25062.
13. McLennan,A.G. (1999) The MutT motif family of nucleotide phosphohydrolases in man and human pathogens (review). *Int. J. Mol. Med.*, **4**, 79–89.
14. Mildvan,A.S., Xia,Z., Azurmendi,H.F., Saraswat,V., Legler,P.M., Massiah,M.A., Gabelli,S.B., Bianchet,M.A., Kang,L.W. and Amzel,L.M. (2005) Structures and mechanisms of Nudix hydrolases. *Arch. Biochem. Biophys.*, **433**, 129–143.
15. Galperin,M.Y., Moroz,O.V., Wilson,K.S. and Murzin,A.G. (2006) House cleaning, a part of good housekeeping. *Mol. Microbiol.*, **59**, 5–19.
16. Das,B., Butler,J.S. and Sherman,F. (2003) Degradation of normal mRNA in the nucleus of Saccharomyces cerevisiae. *Mol. Cell Biol.*, **23**, 5502–5515.
17. Das,B., Das,S. and Sherman,F. (2006) Mutant LYS2 mRNAs retained and degraded in the nucleus of Saccharomyces cerevisiae. *Proc. Natl Acad. Sci.*, **103**, 10871–10876.
18. Kuai,L., Das,B. and Sherman,F. (2005) A nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae. *Proc. Natl Acad. Sci.*, **102**, 13962–13967.
19. Cougot,N., Babajko,S. and Seraphin,B. (2004) Cytoplasmic foci are sites of mRNA decay in human cells. *J. Cell Biol.*, **165**, 31–40.
20. Johnson,A.W. (1997) Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Mol. Cell Biol.*, **17**, 6122–6130.
21. Lejeune,F., Li,X. and Maquat,L.E. (2003) Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Mol. Cell*, **12**, 675–687.
22. Kastenmayer,J.P. and Green,P.J. (2000) Novel features of the XRN-family in Arabidopsis: evidence that AtXRN4, one of several orthologs of nuclear Xrn2p/Rat1p, functions in the cytoplasm. *Proc. Natl Acad. Sci.*, **97**, 13985–13990.
23. West,S., Gromak,N. and Proudfoot,N.J. (2004) Human 5′→3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, **432**, 522–525.
24. Lindstrom,D.L., Squazzo,S.L., Muster,N., Burckin,T.A., Wachter,K.C., Emigh,C.A., McCleery,J.A., Yates,J.R. and Hartzog,G.A. (2003) Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell Biol.*, **23**, 1368–1378.
25. Mandal,S.S., Chu,C., Wada,T., Handa,H., Shatkin,A.J. and Reinberg,D. (2004) Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc. Natl Acad. Sci.*, **101**, 7572–7577.
26. Moteki,S. and Price,D. (2002) Functional coupling of capping and transcription of mRNA. *Mol. Cell*, **10**, 599–609.
27. Speckmann,W.A., Terns,R.M. and Terns,M.P. (2000) The box C/D motif directs snoRNA 5'-cap hypermethylation. *Nucleic Acids Res.*, **28**, 4467–4473.
28. Aguilera,A. (2005) Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Curr. Opin. Cell Biol.*, **17**, 242–250.
29. Ghosh,T., Peterson,B., Tomasevic,N. and Peculis,B.A. (2004) Xenopus U8 snoRNA binding protein is a conserved nuclear decapping enzyme. *Mol. Cell*, **13**, 817–828.
30. Tomasevic,N. and Peculis,B. (1999) Identification of a U8 snoRNA-specific binding protein. *J. Biol. Chem.*, **274**, 35914–35920.
31. Peculis,B.A., Reynolds,K. and Cleland,M. (2007) Metal determines efficiency and substrate specificity of the nuclear NUDIX decapping proteins X29 and H29K (Nudt16). *J. Biol. Chem.*, **282**, 24792–24805.
32. Stuart,G.W., Moffett,K. and Leader,J.J. (2002) A comprehensive vertebrate phylogeny using vector representations of protein sequences from whole genomes. *Mol. Biol. Evol.*, **19**, 554–562.
33. Wu,W., Niles,E., Hirai,H. and LoVerde,P. (2007) Evolution of a novel subfamily of nuclear receptors with members that each contain two DNA binding domains. *BMC Evol. Biol.*, **7**, 27–42.
34. Meyer,A. and Zardoya,R. (2003) Recent advances in the (molecular) phylogeny of vertebrates. *Annu. Rev. Ecol. Evol. Syst.*, **34**, 311–338.
35. Denhez,F., Wilcox-Adelman,S.A., Baciu,P.C., Saoncella,S., Lee,S., French,B., Neveu,W. and Goetinck,P.F. (2002) Syndesmos, a syndecan-4 cytoplasmic domain interactor, binds to the focal adhesion adaptor proteins paxillin and Hic-5. *J. Biol. Chem.*, **277**, 12270–12274.
36. Baciu,P.C., Saoncella,S., Lee,S.H., Denhez,F., Leuthardt,D. and Goetinck,P.F. (2000) Syndesmos, a protein that interacts with the cytoplasmic domain of syndecan-4, mediates cell spreading and actin cytoskeletal organization. *J. Cell Sci.*, **113**(Pt 2), 315–324.
37. Scarsdale,J.N., Peculis,B.A. and Wright,H.T. (2006) Crystal structures of U8 snoRNA decapping nudix hydrolase, X29, and its metal and cap complexes. *Structure*, **14**, 331–343.
38. Hall,B.G. (2001) *Phylogenetic Trees Made Easy: A How-to Manual for Molecular Biologists*. Sinauer Associates, Inc., Sunderland, Massachusetts.
39. Baldauf,S.L. (2003) Phylogeny for the faint of heart: a tutorial. *Trends Genet.*, **19**, 345–351.
40. An,D.J., Roh,I.S., Song,D.S., Park,C.K. and Park,B.K. (2007) Phylogenetic characterization of porcine circovirus type 2 in PMWS and PDNS Korean pigs between 1999 and 2006. *Virus Res.*, **129**, 115–122.
41. Palma,A.C., Araujo,F., Duque,V., Borges,F., Paixao,M.T. and Camacho,R. (2007) Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect. Genet. Evol.*, **7**, 391–398.
42. Abouheif,E., Zardoya,R. and Meyer,A. (1998) Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. *J. Mol. Evol.*, **47**, 394–405.
43. Bi,J.L., Castle,S.J. and Toscano,N.C. (2007) Amino acid fluctuations in young and old orange trees and their influence on glassy-winged sharpshooter (Homalodisca vitripennis) population densities. *J. Chem. Ecol.*, **33**, 1692–1706.
44. Abdelghany,H.M., Bailey,S., Blackburn,G.M., Rafferty,J.B. and McLennan,A.G. (2003) Analysis of the catalytic and binding residues of the diadenosine tetraphosphate pyrophosphohydrolase from Caenorhabditis elegans by site-directed mutagenesis. *J. Biol. Chem.*, **278**, 4435–4439.
45. Simmons,M.P. and Ochoterena,H. (2000) Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.*, **49**, 369–381.
46. Simmons,M.P., Ochoterena,H. and Carr,T.G. (2001) Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Syst. Biol.*, **50**, 454–462.

47. Simmons,M.P., Ochoterena,H. and Freudenstein,J.V. (2002) Amino acid vs. nucleotide characters: challenging preconceived notions. *Mol. Phylogenet. Evol.*, **24**, 78–90.
48. Hillis,D.M. (1987) Molecular versus morphological approaches to systematics. *Annu. Rev. Ecol. Syst.*, **18**, 23–42.
49. Baldauf,S.L., Roger,A.J., Wenk-Siefert,I. and Doolittle,W.F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*, **290**, 972–977.
50. Woese,C.R. (2000) Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. USA* **97**, 8392–8396.
51. Philippe,H., Lartillot,N. and Brinkmann,H. (2005) Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.*, **22**, 1246–1253.
52. Philippe,H., Snell,E.A., Bapteste,E., Lopez,P., Holland,P.W.H. and Casane,D. (2004) Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.*, **21**, 1740–1752.
53. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
54. Brown,N.P., Leroy,C. and Sander,C. (1998) MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics*, **14**, 380–381.