

RESEARCH ARTICLE

Open Access

Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes

Matteo Citarelli¹, Sachin Teotia^{1,2}, Rebecca S Lamb^{1,2*}

Abstract

Background: The Poly(ADP-ribose)polymerase (PARP) superfamily was originally identified as enzymes that catalyze the attachment of ADP-ribose subunits to target proteins using NAD⁺ as a substrate. The family is characterized by the catalytic site, termed the PARP signature. While these proteins can be found in a range of eukaryotes, they have been best studied in mammals. In these organisms, PARPs have key functions in DNA repair, genome integrity and epigenetic regulation. More recently it has been found that proteins within the PARP superfamily have altered catalytic sites, and have mono(ADP-ribose) transferase (mART) activity or are enzymatically inactive. These findings suggest that the PARP signature has a broader range of functions than initially predicted. In this study, we investigate the evolutionary history of PARP genes across the eukaryotes.

Results: We identified *in silico* 236 PARP proteins from 77 species across five of the six eukaryotic supergroups. We performed extensive phylogenetic analyses of the identified PARPs. They are found in all eukaryotic supergroups for which sequence is available, but some individual lineages within supergroups have independently lost these genes. The PARP superfamily can be subdivided into six clades. Two of these clades were likely found in the last common eukaryotic ancestor. In addition, we have identified PARPs in organisms in which they have not previously been described.

Conclusions: Three main conclusions can be drawn from our study. First, the broad distribution and pattern of representation of PARP genes indicates that the ancestor of all extant eukaryotes encoded proteins of this type. Second, the ancestral PARP proteins had different functions and activities. One of these proteins was similar to human PARP1 and likely functioned in DNA damage response. The second of the ancestral PARPs had already evolved differences in its catalytic domain that suggest that these proteins may not have possessed poly(ADP-ribosyl)ation activity. Third, the diversity of the PARP superfamily is larger than previously documented, suggesting as more eukaryotic genomes become available, this gene family will grow in both number and type.

Background

Poly(ADP-ribosyl)ation activity was originally identified in the 1960s [1-5]; it is the rapid and reversible post-translational covalent attachment of ADP-ribose subunits onto glutamate, aspartate, and lysine residues of target proteins. The ADP-ribose polymer is formed by sequential attachment of ADP-ribosyl moieties from NAD⁺; the polymers can reach a length of over 200 units and can have multiple branching points. Overall, the ADP-ribose polymer is highly negatively charged

and has large physiological consequences on functional and biochemical properties of the proteins modified.

Poly(ADP-ribosyl)ation is done by enzymes called poly(ADP-ribose)polymerases (PARPs). The so-called PARP signature, a catalytic β -alpha-loop-B-alpha NAD⁺ fold [6,7], characterizes these enzymes. PARPs are found in diverse groups of eukaryotes [8,9], but are best studied in animals. PARPs have been shown to be involved in DNA damage repair, cell death pathways, transcription and chromatin modification/remodelling (reviewed in [10-13]). PARPs have been implicated in a wide range of human diseases (reviewed in [14]) and are important targets for anti-cancer therapies [15]. A polymorphism in human PARP1, which causes decreased enzymatic activity, has been reported to be associated with an

* Correspondence: lamb.129@osu.edu

¹Plant Cellular and Molecular Biology Department, Ohio State University, 500 Aronoff Laboratory, 318 W. 12th Ave., Columbus, OH 43210 USA
Full list of author information is available at the end of the article

increased cancer risk and a decreased risk of asthma [16,17], further underlining the importance of this class of enzymes and their complex roles in disease.

The first PARP purified and cloned, PARP1 from human, remains the best studied. PARP1 was long thought to be the only enzyme with poly(ADP-ribose)ylation activity until two PARP isoforms were identified in plants [18] and, simultaneously, tankyrase was identified as a PARP localized at the telomere in humans [19]. Subsequently, studies on PARP1 knock out mice demonstrated that the mutant mice still possessed poly(ADP-ribose)ylation capacity and developed normally [20,21], suggesting other enzymes existed. Since these studies, a number of genes containing the PARP signature have been identified, although a minority of them have been functionally characterized.

The PARP-like family has been best characterized in humans, where there are seventeen family members that share the PARP catalytic domain, but vary widely in other parts of the proteins [8,9]. It is postulated that different PARPs subfamilies participate in diverse events mediated by their variable domain structures. However, only some of the family members have been shown to have PARP activity, mostly in humans (PARP1 [22] and its orthologs from other species (for example, [23,24]), PARP2 [25,26], tankyrase1 [19,27], tankyrase2 [28,29], and vPARP [30]). Most of these enzymes contain an evolutionarily conserved catalytic glutamate residue in an "HYE" catalytic triad. This residue was shown to be essential for poly(ADP-ribose) chain elongation in human PARP1 [31]. It is clear that some proteins with PARP signatures missing the catalytic glutamate residue or other residues known to be important for chain elongation do not act in poly(ADP-ribose)ylation. For example, human PARP10 has transferase activity rather than polymerase activity, adding one ADP-ribose subunit to target proteins [32]. It is thought that other PARP-like proteins may actually function in mono(ADP-ribose)ylation [32-34] or even have non-enzymatic functions; human PARP9 appears to not have enzymatic activity [35]. Even enzymes that retain the catalytically important residues that have been identified may not act as PARPs. For example, conflicting reports about the catalytic activity of human PARP3 exist; it has been reported act in poly(ADP-ribose)ylation [36] and mono(ADP-ribose)ylation [37].

Our knowledge of the PARP gene family is principally based on animals, in particular mammals. This taxon is a member of the Opisthokonts, one of the six eukaryotic "supergroups" [38,39] and therefore represents only a portion of the evolutionary history and diversity of known eukaryotes. For the other five eukaryotic supergroups, studies on PARPs have been limited or non-existent. A previous study on PARPs identified new

members in more basal animals, amoebas, fungi and plants [40]. However, no representatives from Excavates or Chromalveolates were included in the analysis and only one member of Plantae (*Arabidopsis thaliana*). Here we use comparative genomics and phylogenetic analysis to investigate the distribution of PARP genes across almost the entire breadth of eukaryotes, to reconstruct the evolutionary history of this protein family and to gain insights into its functional diversification. Our results indicate that the last common ancestor of extant eukaryotes encoded at least two PARP proteins, one similar to human PARP1 and functioning in DNA repair and damage response, the other likely acting in mono(ADP-ribose)ylation; the cellular role of the last group is not known.

Results

Identification of PARP genes from eukaryotic genomes

We used the information obtained from the Pfam database [41-43] and Uniprot [44,45] along with BLAST searches [46] of sequenced eukaryotic genomes at the DOE Joint Genome Institute (JGI), the Broad Institute, the J. Craig Venter Institute, ToxoDB [47], NCBI, dictyBase [48] and the Arabidopsis Information Resource (TAIR) [49] to compile the sequences of over 300 PARP proteins. After preliminary alignment and phylogenetic analysis, we reduced the number of species representing animals; specifically we choose representative species of vertebrates since the genes from this group are shared by all and kept *Drosophila melanogaster* or *Anopheles gambiae* to represent insects, since all of our sequences were from Diptera. This left us with 236 sequences from 77 eukaryotic species (Additional file 1). In addition, another 46 sequences contained regions with high similarity to the PARP catalytic domain (Additional file 2); however, these sequences were incomplete and not included in the alignment. Nonetheless, these sequences likely represent *bona fide* members of the PARP catalytic domain. The PARP catalytic domain was extracted from the proteins sequences and aligned using MUSCLE [50]. This alignment can be found in Additional file 3.

Phylogenetic analysis of the PARP family suggests that the ancestral eukaryote had at least two PARP enzymes

We first analyzed all the PARP-like genes we identified in the eukaryotic lineage. We used the multiple sequence alignment of the PARP catalytic domain generated above (Additional file 3) to generate a maximum-likelihood phylogenetic tree of the PARP family (Additional file 4). We defined six clades of PARPs based on our maximum-likelihood tree, an examination of domains found outside of the PARP catalytic domain used to generate that tree and the evolutionary relationships of organisms within clades (Clades 1-6; Figure 1). Clades were defined as

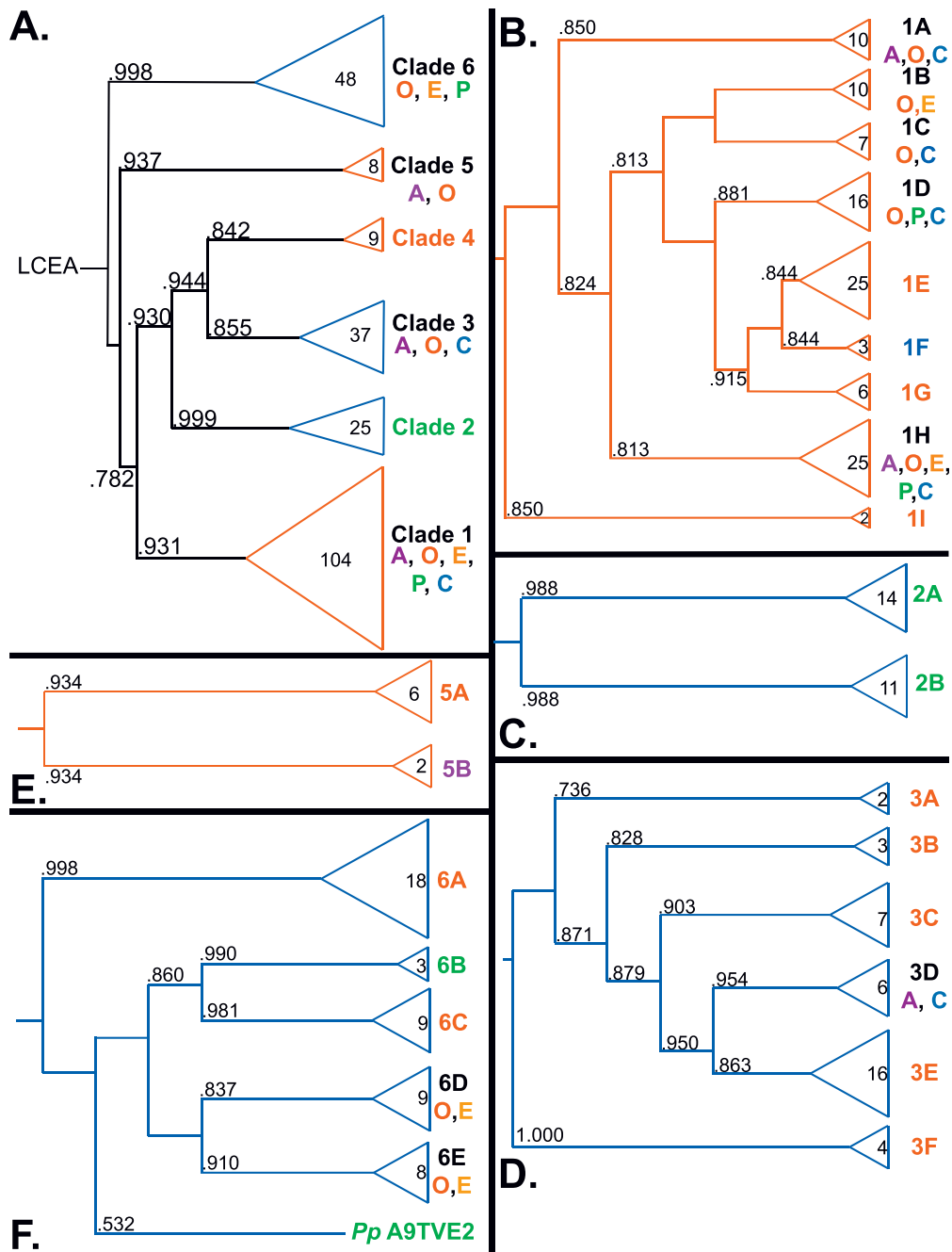


Figure 1 The PARP gene family forms six clades. A. Graphical representation of the maximum-likelihood (ML) phylogenetic tree of all identified eukaryotic PARPs indicating the relationships between the six clades as defined in the text. The full tree can be found in Additional file 4. The tree was based on an alignment of the PARP catalytic domains (Additional file 3). B. Graphical representation of the ML phylogenetic tree of Clade 1 PARPs indicating the relationships between nine subclades as defined in the text. C. Graphical representation of the ML phylogenetic tree of Clade 2 indicating the relationships between the two subclades as defined in the text. D. Graphical representation of the ML phylogenetic tree of Clade 3 PARPs indicating the relationships between the six subclades as defined in the text. E. Graphical representation of the ML phylogenetic tree of Clade 5 PARPs indicating the relationship between the two subclades as defined in the text. F. Graphical representation of the maximum-likelihood (ML) phylogenetic tree of Clade 6 PARPs indicating the relationships between the six subclades as defined in the text. Numbers in the clades or subclades indicate the number of proteins in each. Colors and letters indicate the eukaryotic supergroup or groups represented. A, Amoebozoa; O, Opithokonts; E, Excavata; P, Plantae; C, Chromalveolates. Purple, Amoebozoa; red, Opithokonts; orange, Excavata; green, Plantae; blue, chromalveolates. Branch support values are indicated at the nodes as computed in PhyML using an aLRT non-parametric Shimodaira-Hasegawa-like (SH) procedure and a midpoint rooting method. Triangle and branch colors indicate either the presence of the HYE (red) or variant (blue) catalytic triad in each group. Branch lengths do not indicate genetic distance.

having a bootstrap value of at least .8, one or more shared domains outside of the PARP catalytic domain, and having subbranches consisting of proteins from closely related species. Within each major clade one or more subclades were defined by similar reasoning; however, the branch supports for subclades were less stringent. Clade 5 contains proteins with almost the exact same domain structures all from closely related species; therefore, subclades were not defined for this clade. Four proteins (*Dictyostelium discoideum* DDB0232241, *Naegleria gruberi* 72525, *Naegleria gruberi* 80603 and *Caenorhabditis elegans* PME5) did not fall clearly into any clades; rather they fell between clades or next to proteins from widely divergent species (Additional file 4). Therefore, they have not been included in any of the defined clades. *Dictyostelium* DDB0232241 contains two WWE domains and a Cwf15/Cwc15 domain. WWE domains are postulated to be protein-protein interaction domains and are found in proteins involved in the ubiquitin/proteasome pathway and in PARPs [51]. Cwf15/Cwc15 domains are of unknown function and found in splicing factors [52]. *Naegleria gruberi* is a member of the Heterolobosea within the eukaryotic group Excavates (Figure 2). Heterolobosea are protozoa, many of which, including *Naegleria gruberi*, can transform between amoeboid, flagellate, and encysted stages. *Naegleria gruberi* is the only member of this group of organisms with a completed genome, making it impossible to determine if these genes are representative of ones found in a wide range of heterolobosea species or are more specific to *Naegleria* and its relatives. The two *Naegleria* PARP-like proteins are relatively short proteins with the PARP catalytic domain at their very C termini. Their N termini contain no known functional domains. The function of these proteins remains obscure, although they retain the “HYE” catalytic triad (Additional File 3), and may act as *bona fide* PARPs. *C. elegans* PME5 has been characterized as a tankyrase [53,54] and does share ankyrin repeats in its N terminus with those proteins, which are found in Clade 4. The placement of this protein outside of the defined clades likely reflects the large changes found in *C. elegans* PARPs (see below).

The PARP lineages (which will be detailed below) include one clade, Clade 1, which contains representatives from five of the six so-called eukaryotic supergroups: Plantae, Opisthokonta, Chromalveolates, Excavates, and Amoebozoa (Figures 1, 2 and 3; [38,39]). There is no completely sequenced species available from the sixth supergroup, Rhizaria. This broad distribution suggests that the last common ancestor of all extant eukaryotes encoded a gene similar to those of Clade 1. Clade 6 is only found in three of the eukaryotic supergroups; however, the position of this clade as sister group to all other members of

the PARP superfamily and the placement of these groups within eukaryotes supports the hypothesis that the last common eukaryote also encoded such a gene (Figure 2).

Clade 1: the PARP1 clade

Clade 1 is the most broadly distributed PARP clade among eukaryotes (Figures 1, 2 and 3 and [40]). The distribution of Clade1 proteins among eukaryotic species suggests that there was at least one Clade 1-like PARP protein encoded in the genome of their last common ancestor. This group of PARPs can be subdivided into nine subclades (A-H; Figures 1 and 3). Almost all members of Clade 1 are characterized by the presence of WGR and PARP regulatory domains (PRD) in addition to the PARP catalytic domains, one of the reasons we placed these proteins together (Figure 4). The WGR domain is found in PARPs as well an *Escherichia coli* molybdate metabolism regulator and other proteins of unknown function. Its exact function is unclear, but it is proposed to be a nucleic acid binding domain. The PRD domain is found only in Clade 1 PARP proteins and has been shown to increase the poly(ADP-ribosylation) activity of proteins that contain it. Consistent with the presence of PRD domains, many members of Clade 1 have been demonstrated to have poly(ADP-ribosylation) activity, making it likely that most if not all members have this activity; this is also supported by the finding that the so-called HYE catalytic triad is conserved in almost all of these proteins (Additional files 5 and 6). Another commonality between members of Clade 1 is that many of them have been shown to have roles in DNA repair. Other common domains found in Clade 1 proteins are zinc finger DNA binding domains, BRCT domains and PADR1 domains. The BRCT domain, originally identified in the C terminus of the BRCA-1 protein, is usually found in proteins involved in cell cycle regulation and/or DNA repair [55]. The PADR1 domain is found only in PARPs (specifically Clade 1 PARPs) and is of unknown function [56].

Clade 1A is found in Amoebozoa (*Dictyostelium*), Opisthokonta (fungi) and Chromalveolates (the ciliate *Paramecium tetraurelia*) and is the sister group to most of the other Clade 1 subclades (with the exception of Clade 1I; Figure 3). This subclade is unique within Clade 1 in containing proteins with ankyrin repeats, in addition to WGR, PRD and PARP catalytic domains. Clade 1B contains members from both the Opisthokonta (animals and Choanoflagellata) and the Excavata (the Heterolobosea member *Naegleria*). This subclade is typified by human PARP1, the founding member of the superfamily. This protein has three N terminal zinc fingers that contribute to DNA binding, a BRCT domain and a PADR1 domain in addition to WGR, PRD, and the catalytic domain (Figure 4; [22,57,58]).

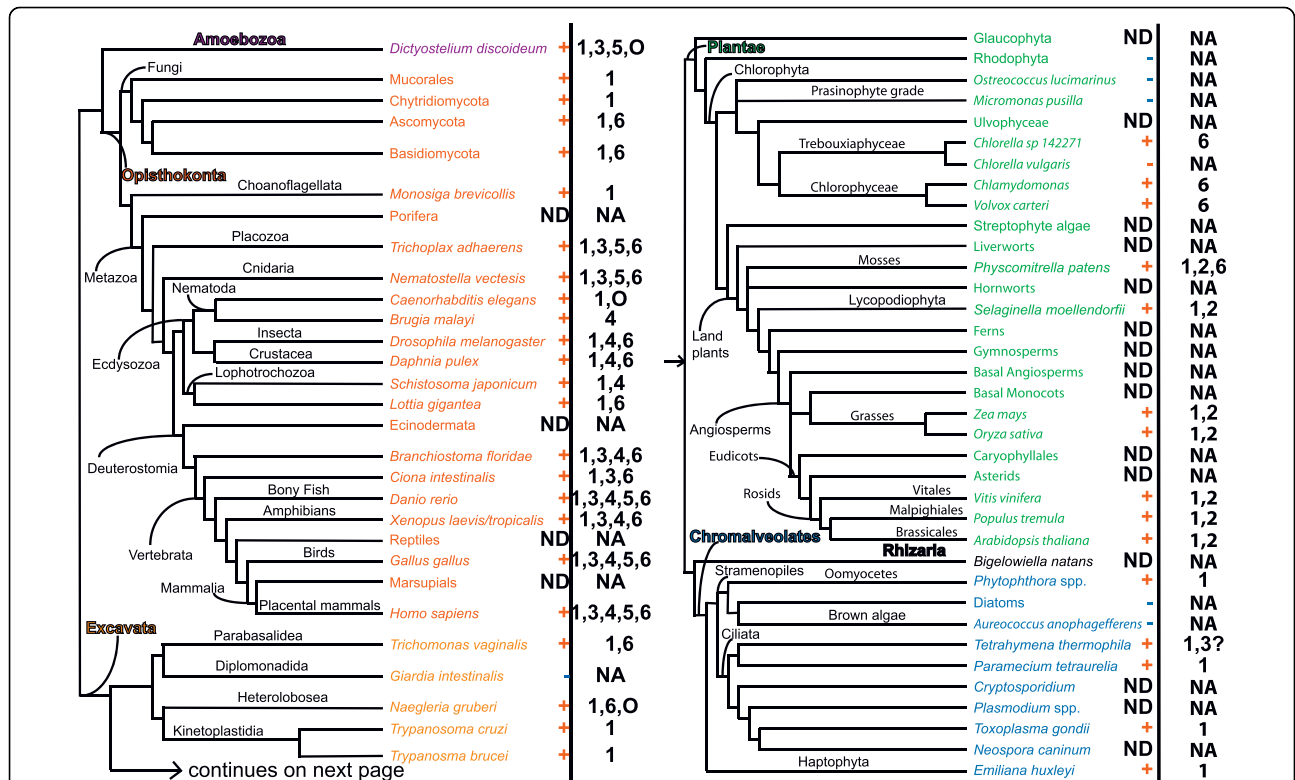


Figure 2 Phylogenetic distribution of the PARP family across eukaryotes. The topology of the schematic tree is based on recent evidence from single- and multi-gene phylogenetic analyses of eukaryotes or subgroups thereof. Some nodes, especially the deepest ones (e.g. monophyly of Excavates, Plantae or Chromalveolates + Rhizarians), remain controversial; these uncertainties do not affect the conclusions concerning the evolutionary history of the PARP family. Branch lengths do not reflect genetic distance. Presence or absence of PARP proteins are indicated by a red + or a blue -, respectively. For each species or group, PARP family members are listed with the clade numbers introduced in Figure 1. For an expanded phylogeny of the fungi, please see Figure 11. Accession numbers of the genes and details on the source of data for individual taxa is provided in Additional files 1 and 2. The six eukaryotic supergroups are indicated as follows: Amoebozoa, purple; Opisthokonta, red; Excavata, orange; Plantae, green; Rhizaria, black; Chromalveolates, blue. ND, no data; NA, not applicable; O, orphan PARPs, as discussed in the text; 3?, reflects the ambiguity of placement of the Tetrahymena proteins into this clade, as discussed in the text.

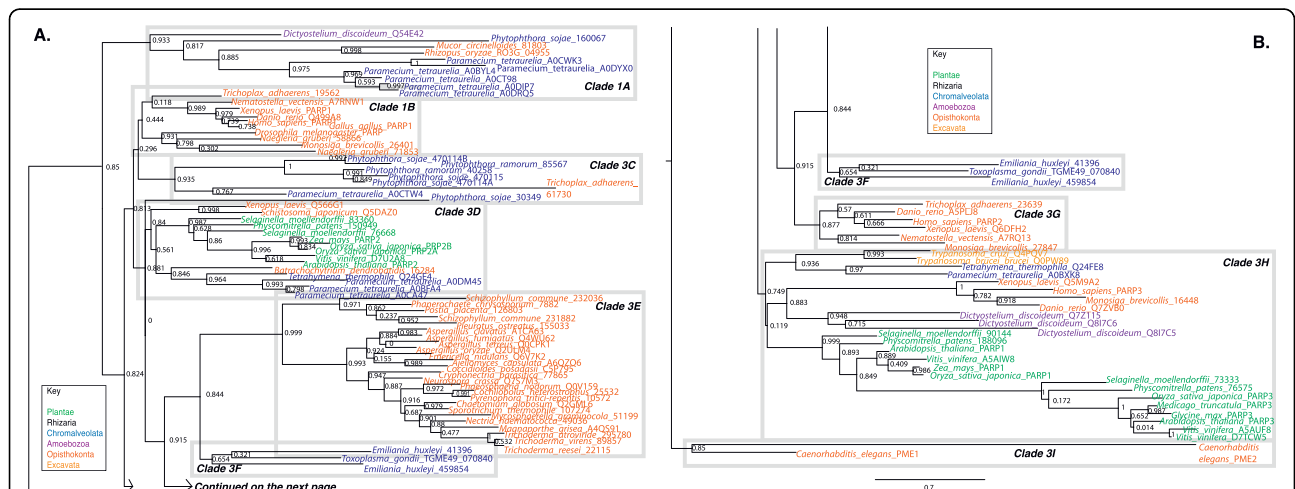


Figure 3 Phylogenetic analysis of Clade 1 PARP genes. The represented tree is a ML tree. This tree is based on a multiple alignment that includes the PARP catalytic domain (Additional file 5). Clade 1 proteins can be divided into nine subclades A-I, as indicated. Branch supports as in Figure 1. Scale bar indicates genetic distance reflected in branch length.

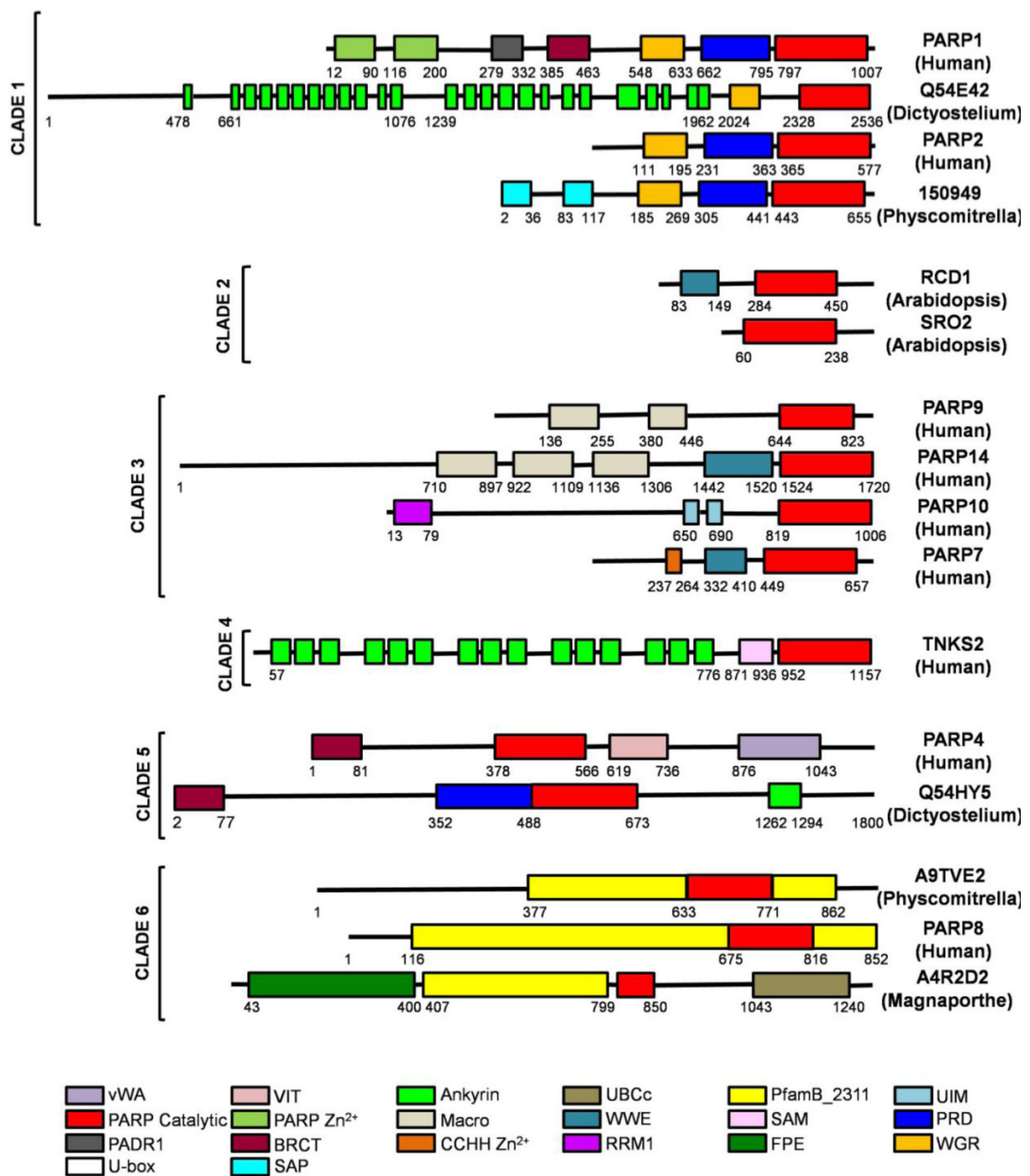


Figure 4 Schematic representation of domains found in PARP proteins. Proteins are arranged by clade as defined in Figure 1 and in the text (indicated on the left). The protein name is given on the right, with species in parenthesis. Numbers indicate amino acids. Protein domains are illustrated by coloured boxes and were defined according to Pfam 23.0. Although each protein is represented in scale, the proteins are not in scale between each other. vWA, von Willebrand factor type A; VIT, vault inter-alpha-trypsin domain; Ankyrin, ankyrin repeats; UBCc, ubiquitin conjugating enzyme catalytic domain; PfamB_2311, domain of unknown function found in Clade 6 PARPs; UIM, ubiquitin interaction motif; PARP catalytic, catalytic domain of PARPs, PARP Zn²⁺, DNA binding zinc finger; Macro, macro domain; WWE, WWE domain; SAM, sterile alpha domain; PRD, PARP regulatory domain; PADR1, domain of unknown function found in PARPs; BRCT, BRCA-1 C terminus domain; CCHH Zn²⁺, DNA binding zinc finger; RRM1, RNA-binding motif; FPE, Fungal PARP E2-like domain; WGR, domain defined by conserved Trp, Gly, and Arg residues; U-box, modified ring finger found in E3 ubiquitin ligases; SAP, DNA binding domain. Human, *Homo sapiens*; Dictyostelium, *Dictyostelium discoideum*; Arabidopsis, *Arabidopsis thaliana*; Magnaporthe, *Magnaporthe grisea*; Physcomitrella, *Physcomitrella patens*.

Both Clade 1C and 1D both contain proteins that have in common WGR, PRD and PARP catalytic domains and mostly do not contain other functional domains. Clade 1C is confined to several Oomyocete Phytophthora species (within the Excavata) and one basal animal. Clade 1D contains members from Opisthokonta (the animals *Xenopus laevis* (Q566G1) and *Schistosoma japonicum* (Q5DAZ0) and the fungus *Batrachomyces dendrobatidis*) and Plantae (land plants) as well as ciliate members of the Chromalveolates. Some of the land plant members of Clade 1D have acquired SAP domains DNA binding domains [59] N terminal to the other domains (Figure 4). In addition, the land plant members of this group have altered their catalytic triad, alone among Clade 1 members (Additional files 5 and 6). All the plant proteins have a cysteine in place of the histidine while all except for the moss protein have a valine instead of the tyrosine in the second position. However, the plant Clade 1D proteins have retained the glutamic acid in the third position. It is unclear what effect these changes might have on the catalytic activity of these proteins.

Clade 1E contains most of the fungal members of Clade 1 and is characterized by proteins with BRCT domains N terminal to WGR, PRD and PARP catalytic domains. Clade 1F is specific to the Excavata. The *Toxoplasma gondii* representative (TGME49_070840) has a similar domain structure to human PARP1, found in Clade 1B. Clade 1G is confined to the Opisthokonta (both animals and the Choanoflagellate *Monosiga brevicollis*), contains proteins with only WGR, PRD and PARP catalytic domains and includes human PARP2.

All five eukaryotic supergroups that contain sequenced species are represented in Clade 1H (Figures 1 and 3). This clade includes human PARP3. Interestingly, land plants have duplicated one of their Clade 1H genes; one duplicate lineage appears to be changing rapidly, based on the long-branch length in the phylogenetic tree (Figure 3). These proteins may have acquired a novel function or the original function may have been split between the two copies in these species (neofunctionalization or subfunctionalization), as these processes are hypothesized to increase the probability of retention of duplicate genes [60].

The final subclade in Clade 1, Clade 1I, consists of two *Caenorhabditis elegans* (*C. elegans*) proteins, PME1 and PME2, which have been characterized previously [61]. PME1 contains zinc fingers and PADR1, WGR, PRD and PARP domains, while PME2 only has WGR, PRD and PARP domains. As will be discussed further below, many of the nematode proteins are anomalous.

Clade 2: the RCD1 clade

Clade 2 of PARP-like genes consists of proteins identified only in land plants, with representatives found from

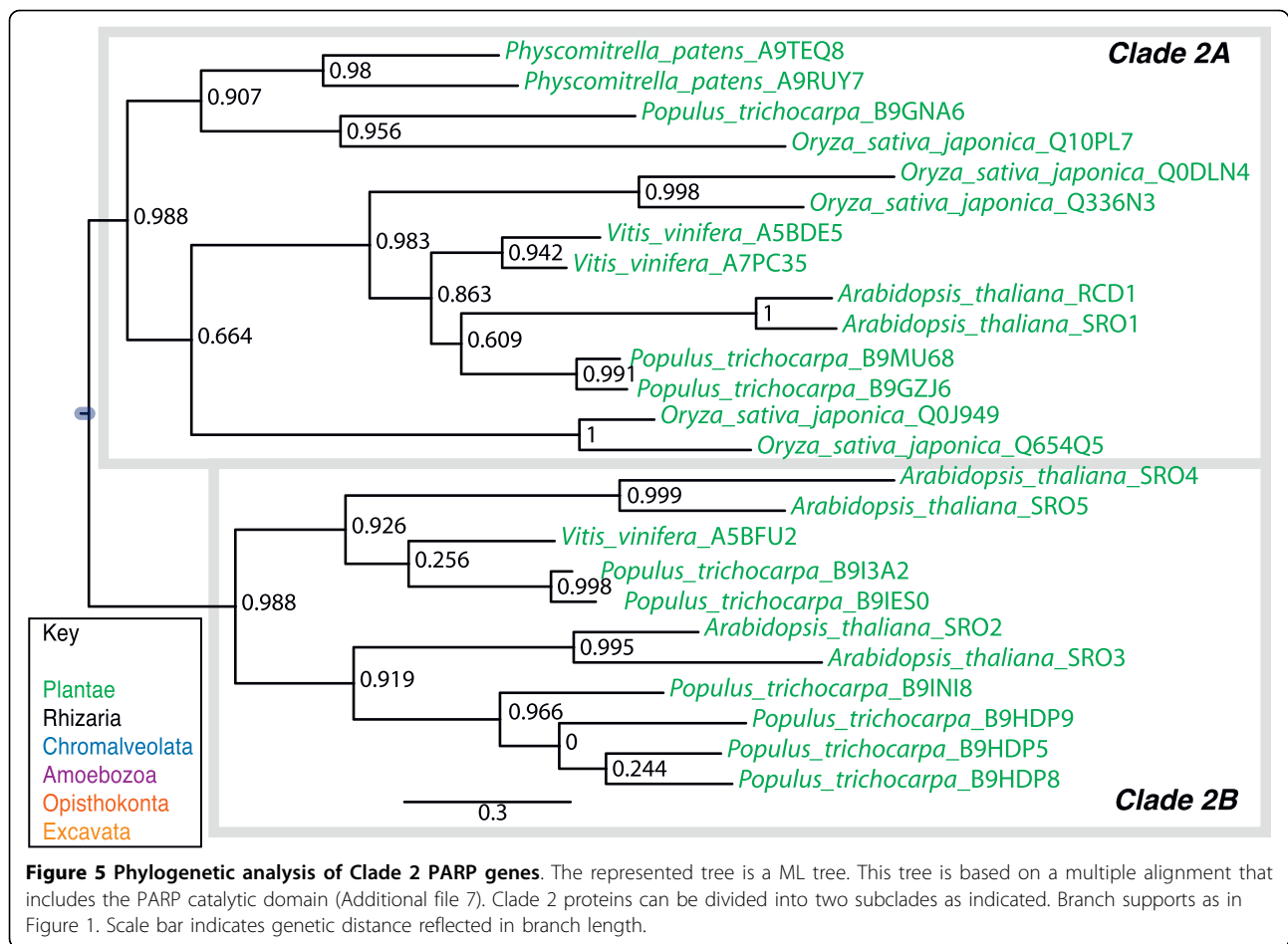
bryophytes to angiosperms (Figures 2 and 5), a finding that has also been made by another group [62]. However, there is no genomic information available for any member of the streptophyte algae, the sister group to land plants within Plantae, leaving open the possibility that members of this clade may be found in these organisms (Figure 2). All groups of land plants also contain members of Clade 1 PARPs, while the moss *Physcomitrella patens* contains Clade 6 proteins in addition (Figure 2).

The founding member of this type of PARP-like protein, RADICAL-INDUCED CELL DEATH 1 (RCD1), was identified in a genetic screen in the model plant *Arabidopsis thaliana* for genes involved in cell death in response to ozone [63] and has been shown to be involved in response to a number of abiotic stresses [64]. Other members of this clade have subsequently been identified based on sequence similarity and several are also involved in stress response [62,65,66]. Clade 2 is made up two subclades (Figure 5). Clade 2A consists of proteins that have, in common with RCD1, an N terminal WWE domain, the PARP signature and a C terminal extension (Figure 4) and are found throughout the breadth of the land plants (Figure 5). Clade 2B is apparently eudicot specific (Figure 5) and consists of relatively short proteins with only the PARP signature and the C terminal extension (Figure 4). Although Clade 2A proteins contain WWE domains, they do not group with another group of WWE containing PARPs, which fall into Clade 3, a clade with no plant representatives (see below). RCD1 has recently been shown to be enzymatically inactive, a result consistent with the lack of conservation of many of the catalytic residues within the PARP domain (Additional file 7; [62]).

One interesting observation we made concerning Clade 2 was the large number of independent gene duplications that have occurred within this gene lineage (Figure 5). While this is likely due to the propensity of plant genomes to undergo whole genome duplications (reviewed in [67]), the retention of many of the gene pairs suggests that Clade 2 proteins are undergoing neofunctionalization and/or subfunctionalization at a high rate [60,68]. This supposition is supported for a pair of Clade 2A paralogs in *Arabidopsis thaliana*, RCD1 and SIMILAR TO RCD ONE 1 (SRO1), which have been shown to be only partially redundant despite a relatively recent evolutionary origin [65,69].

Clade 3

Clade 3 contains proteins from three of the six eukaryotic supergroups: Opisthokonts (animals), Amoebozoa (*Dictyostelium discoideum*) and Chromalveolates (*Tetrahymena thermophila*) (Figures 1 and 6). This clade is likely to be somewhat artificial; the domain structures

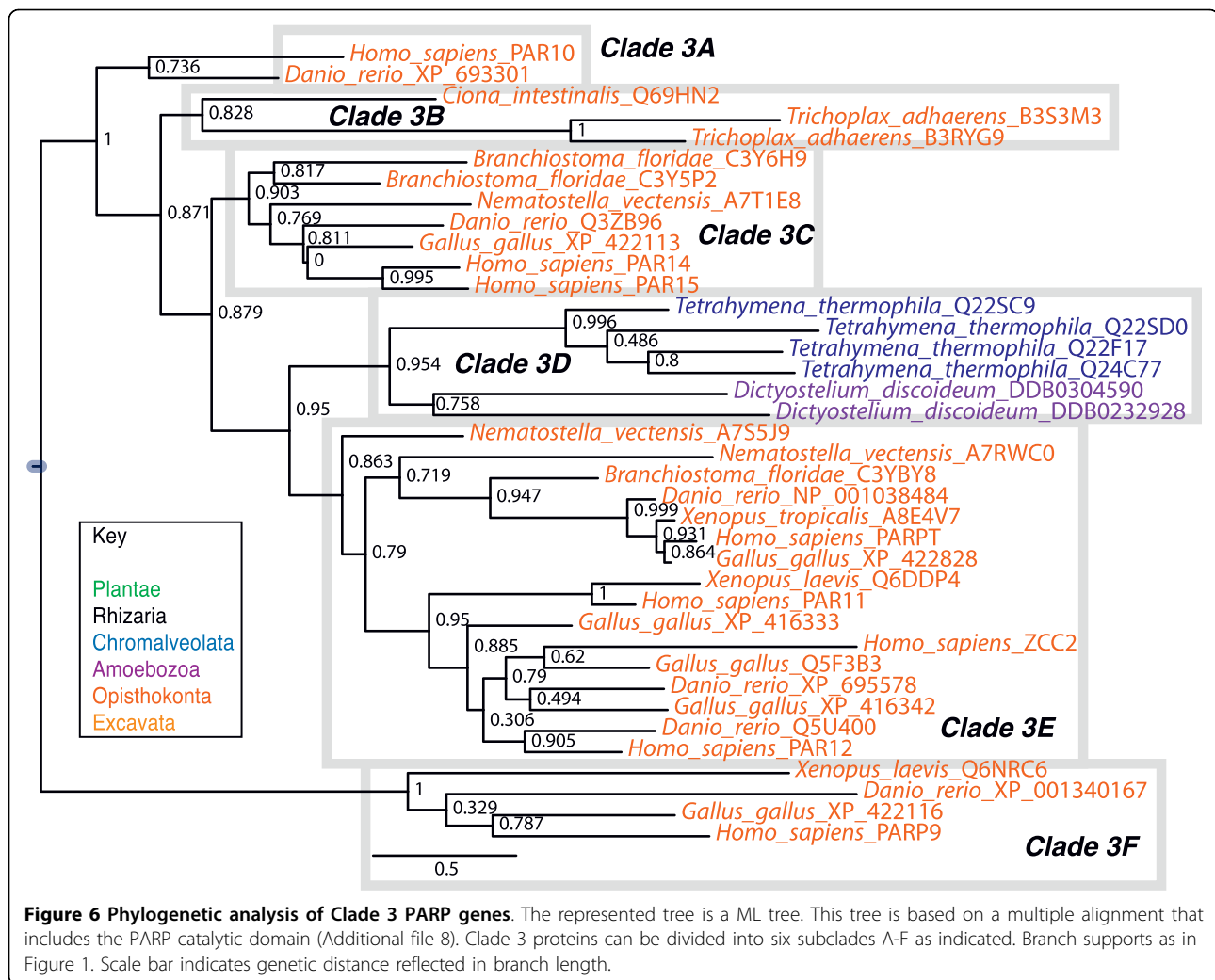


outside of the PARP catalytic domain are heterogeneous among Clade 3 proteins and the presence of *Tetrahymena thermophila* sequences within a group that otherwise contains Opisthokonts and Amoebozoa (which are sister groups) is unlikely to be real. These proteins do share certain characteristics in their catalytic domains suggestive of a switch from PARP activity to mART activity. PARP family members have catalytic domains containing the “HYE” catalytic triad conserved throughout the ADPr transferase superfamily [32]. The third residue, normally a glutamic acid, is not conserved in most Clade 3 members (Figure 7 and Additional file 8), with only one of its members retaining this residue (*Tetrahymena thermophila* Q22F17), while a second has a glutamine (*Tetrahymena thermophila* Q24C77). Most members of the clade have substituted the aliphatic amino acids isoleucine, valine, methionine or leucine for the glutamic acid, while one *Tetrahymena* protein (Q22SD0) as well as human PARP9 and its vertebrate orthologs have threonine or serine at this position. These substitutions have consequences for the catalytic activity of these proteins; these proteins likely do not have poly(ADP-ribosyl)ation activity [32]. It is likely that the grouping of at least the *Tetrahymena* proteins

into this clade is a result of convergent evolution of mART activity.

Given the heterogeneous composition of Clade 3, it is difficult to divide into subclades; however, we classified the proteins into six subclades as outlined below, partially for the purpose of discussion, and partially based on common domain structures and features of the catalytic domains (Figures 6 and 7 and Additional file 8). Clade 3A is composed of two proteins, including human PARP10, containing an RRM RNA binding domain [70], a glycine-rich region (GRD), and a UIM domain, known to bind monoubiquitin and polyubiquitin chains [71]. The proteins found in Clade 3B and 3C contain at least one Macro domain N terminal to their C terminal catalytic domain (Figure 4). Macro domains have been shown to bind to poly(ADP-ribose) (PAR) [72]. Clade 3B includes representatives from the most basal animal in our study *Trichoplax adhaerens*, while 3C includes two human proteins, PARP14 and PARP15. PARP10, PARP14 and PARP15 have been demonstrated to have mART activity [32].

Clade 3D consists of the two *Dictyostelium discoideum* and four *Tetrahymena thermophila* proteins. Unlike the

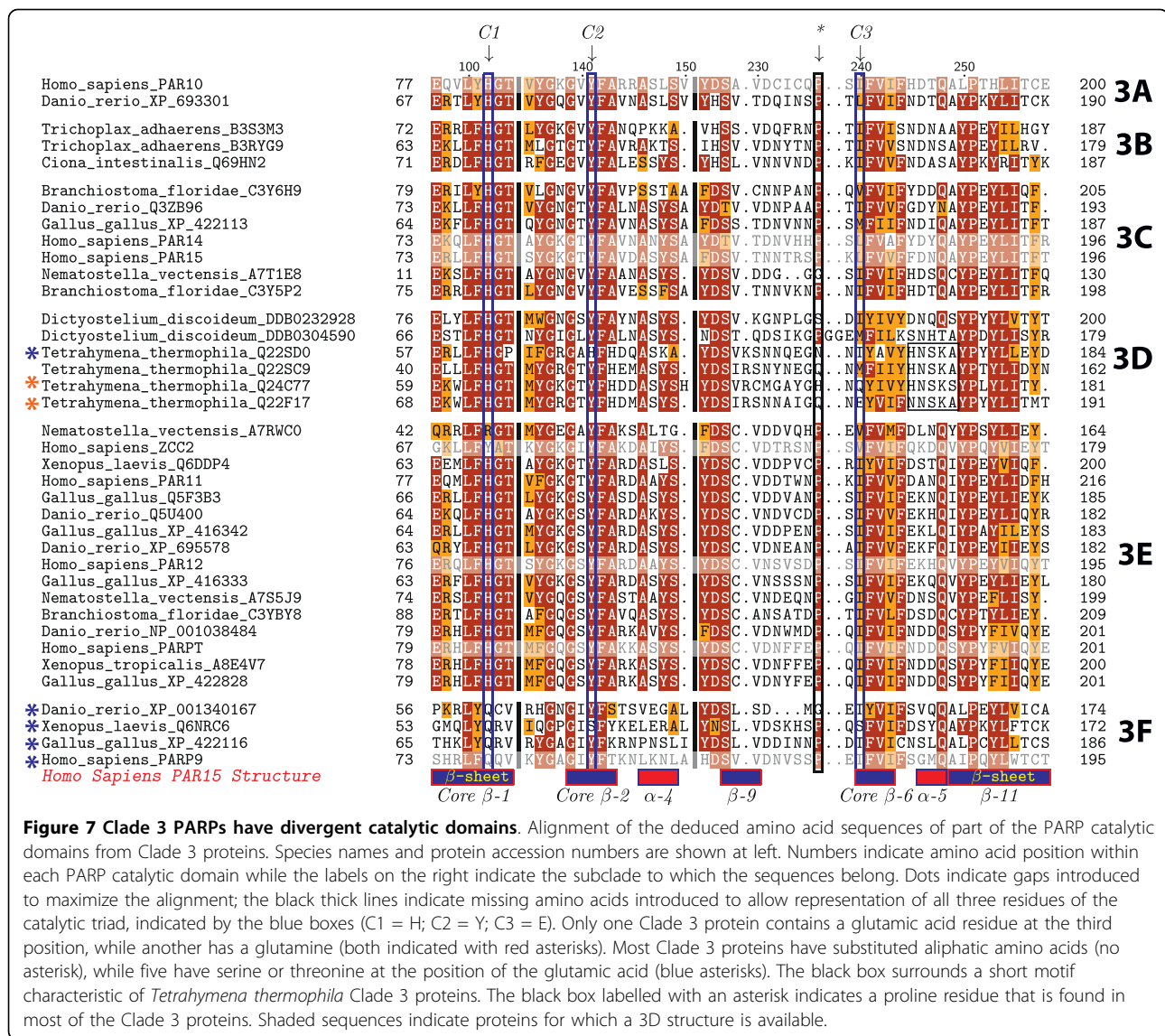


majority of animal proteins in Clade 3, only one of these proteins have a proline located one amino acid away from the third residue of the catalytic triad (Figure 7). The four proteins from the ciliate *Tetrahymena thermophila* have no known functional domains outside of their C terminal PARP catalytic domains and are only similar to one another in this region (data not shown), again supporting the idea that these proteins are not closely evolutionarily related to the other proteins in Clade 3. One of the *Tetrahymena* proteins has retained the glutamic acid of the “HYE” (Figure 7), again supporting this interpretation. All four proteins also share a H/NNSK motif just past the last amino acid of the putative catalytic triad not found in other members of Clade 3 (Figure 7). The *Dictyostelium* proteins in 3D do not show high similarity outside of the PARP domain. DDB0304590 is a relatively short protein with only the PARP catalytic domain and a short C terminal extension. DDB0232928 has a Macro domain and, at its very N terminus, a U-box (Figure 4). The U-box is a

modified RING finger [73] found in E3 ubiquitin ligases known to bind ubiquitin E2 enzymes [74]. As Amoebozoa is the sister group to Opisthokonts within eukaryotes and given that DDB0232928 contains a Macro domain as do some other members of Clade 3, it is possible that these proteins are orthologous to at least some of the animal Clade 3 proteins.

Clade 3E is confined to animals, but is not represented in Placozoa (Figure 6). Members of this subclade contain one to two WWE domains, alone or in combination with zinc fingers (either CCCH or CCCH types) in front of their PARP catalytic domains (Figure 4). All members of 3E have replaced the glutamic acid characteristic of PARPs with an isoleucine except for two (human ZCC2/PARP13 and *Nematostella vectensis* A7RWC0) that contain valines at that site (Figure 7). This subclade also contains human PARP12 and human PARPT/PARP7.

Clade 3F, which is sister group to all other Clade 3 subclades, contains human PARP9 and orthologs from vertebrates. These proteins contain two Macro domains N



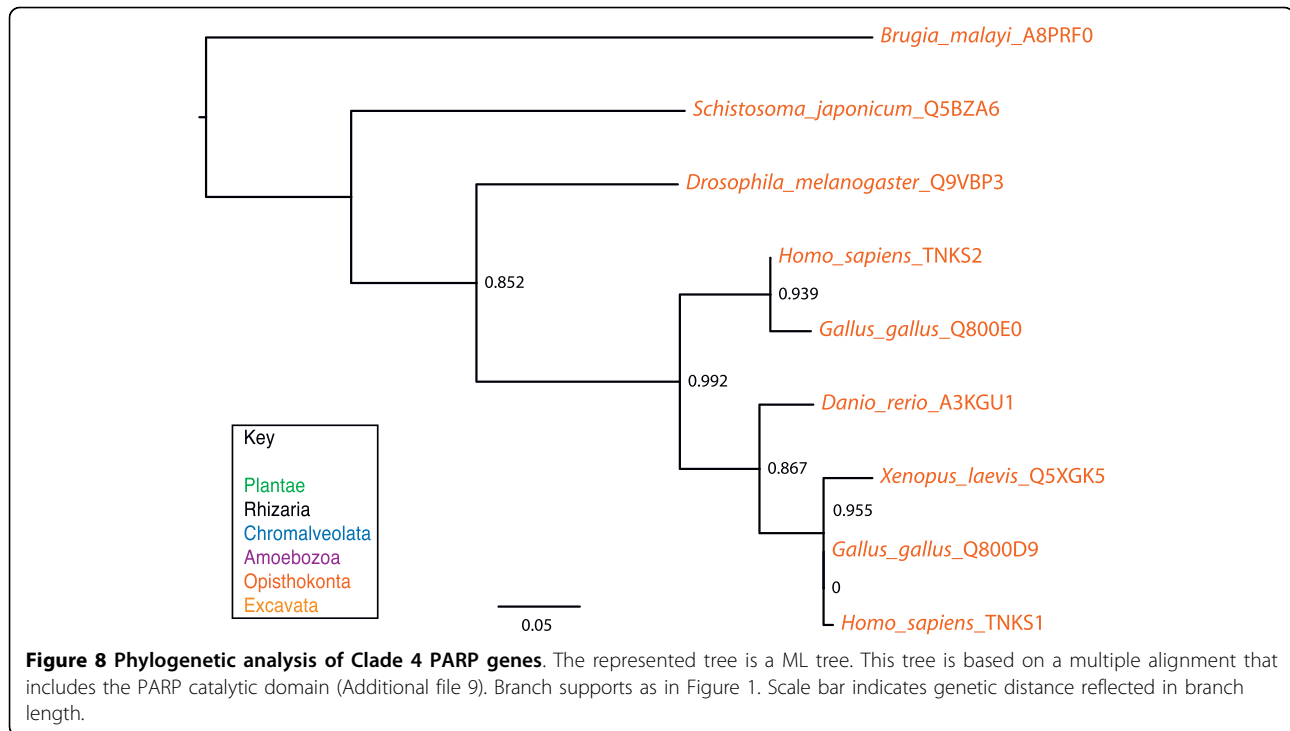
terminal to their PARP catalytic domains (Figure 4) and have a more divergent catalytic triad than the rest of Clade 3, having Q-Y/S-T/S instead of HYE (Figure 7 and Additional file 8). Human PARP9 has been shown to be inactive [35], suggesting that no Clade 3F proteins act as enzymes. PARP9 was originally identified as a gene conferring risk for diffuse large B-cell lymphoma and named BAL1 (B-aggressive lymphoma 1) [75]. Interestingly, two proteins identified by their similarity to BAL1, PARP14/BAL2 and PARP15/BAL3, although their domain structures resemble that of PARP9/BAL1, group in subclade 3C (Figures 6 and 7), and act as mARTs [35].

Clade 4: the tankyrase clade

Clade 4 proteins are characterized by fifteen to eighteen ankyrin repeats followed by a sterile alpha motif (SAM),

most likely a protein-protein interaction domain [76], and the PARP catalytic domain (Figure 4). These proteins are so similar to one another that we have not further subdivided them (Figure 8 and Additional file 9). The two human members of this clade, tankyrase1 and tankyrase2, have been shown to have poly(ADP-ribose)ation activity [19,27,77]. All proteins grouped in this clade retain the “HYE” catalytic triad (Figure 1 and Additional file 8), suggesting that they are likely to be active enzymes.

Our analysis indicates true tankyrases are confined to animals, and in fact do not appear to be found outside of the bilateria (Figures 2 and 8). A duplication event that generated two tankyrase-encoding genes appears to have occurred within the vertebrates, sometime after the separation of the amphibians. The absence of tankyrase



orthologs outside of the animals contradicts the report of such proteins in protozoa such as *Dictyostelium discoideum* and *Tetrahymena thermophila* [78]. However, these protozoan proteins differ from the canonical tankyrases in structure; although they have ankyrin repeats in their N terminal region, these are followed by WGR and PRD domains rather than a SAM motif (Figure 4). Consistent with the presence of the WGR and PRD domains and the low similarity between their PARP catalytic domain and that of tankyrases, these proteins fall into Clade 1A (Figure 3). This suggests that PARP proteins independently acquired ankyrin repeats at least twice.

Clade 5: The vPARP clade

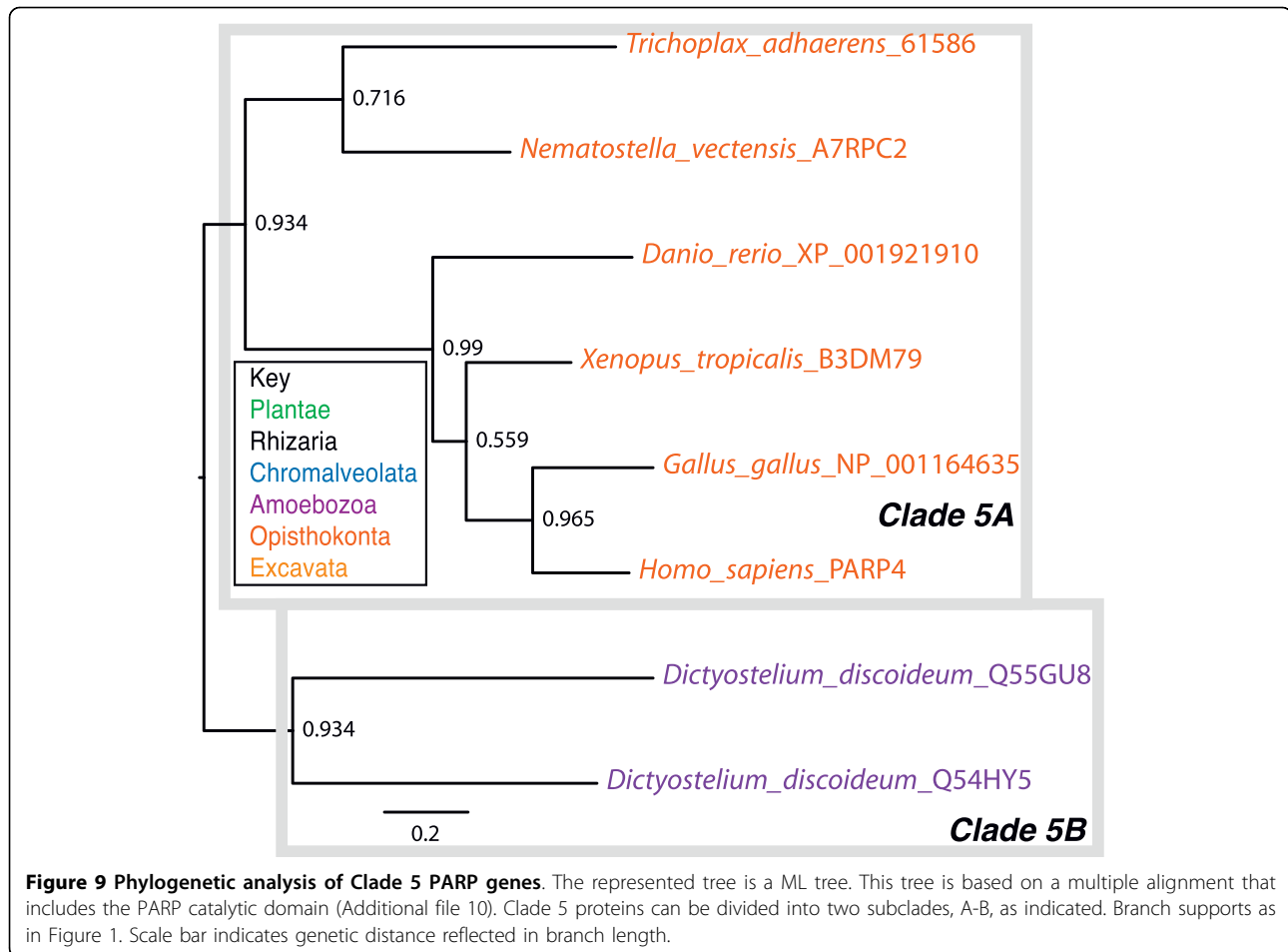
Clade 5 is found only in the Opisthokonts (animals) and Amoebozoa (Figure 9 and Additional file 10) and is characterized by the position of the PARP catalytic domain. In this group, the PARP signature is found in the middle of the protein, rather than at the C terminus and is typified by human vPARP/PARP4. vPARP has the catalytic domain preceded by a BRCT domain and followed by a vault protein inter-alpha-trypsin (VIT) domain, and a von Willebrand factor type A domain (vWA) (Figure 4; [30]). Both VIT and vWA domains are commonly found in proteins of multiprotein complexes and are structurally related to each other [79]. Clade 5 is further subdivided into two subclades (Figure 9). Clade 5A contains animal proteins while Clade 5B

contains two proteins from the amoeba *Dictyostelium discoideum* (Q54HY5 and Q55GU8). The amoeba proteins have a different protein structure than the animal members of this clade; they too have BRCT domains N terminal to their PARP catalytic domains and long C terminal extensions. However, there are no VIT or vWA domains found in these proteins.

vPARP is associated with vaults, very large cytoplasmic ribonucleoprotein particles first described in the 1980s whose function is unclear [80]. Vault has a patchy taxonomic distribution within eukaryotes. Our analysis suggests that the phylogenetic distribution of vPARP is also limited (Figures 2 and 9); members of Clade 5A with the vPARP domain structure are found only in animals that have been shown to contain vaults, while Clade 5B proteins are found in *Dictyostelium*, which also contains vaults [81]. However, although vaults have been identified in trypanosomes [82], no evidence of proteins sharing the domain structure of vPARP can be found in this group of organisms, although such proteins may be present in species with currently unsequenced genomes.

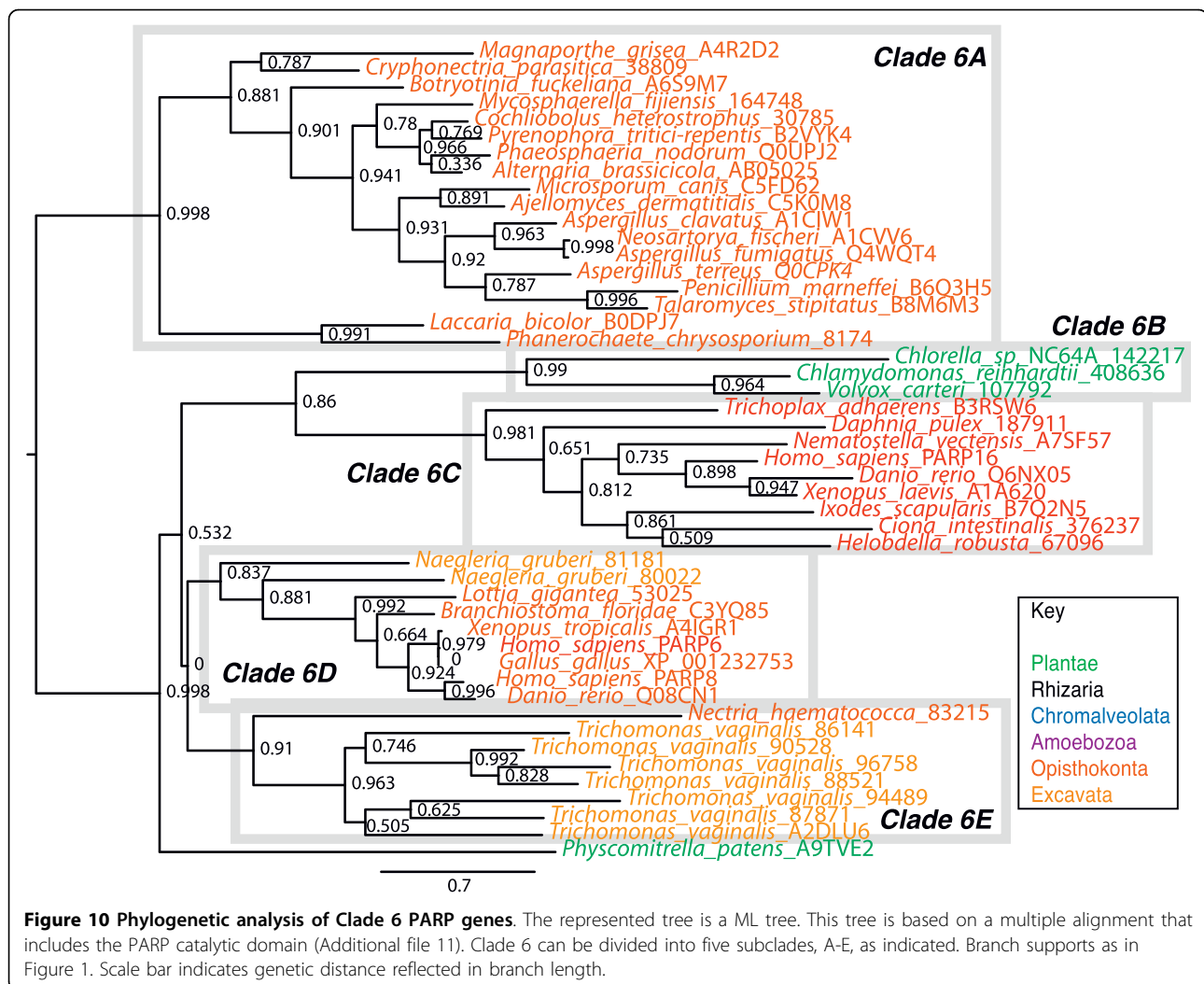
mART activity may be ancient

Clade 6 proteins are found in Opisthokonts (animals and fungi), Excavates (Parabasalids and Heterolobosa), and Plantae (chlorophyta and bryophytes) (Figures 1, 2 and 10 and Additional file 11). Based on its position as sister group to all other clades of PARPs (Figure 1) and



the distribution of species containing Clade 6 PARPs within the eukaryotes (Figure 2), it is likely that the last common eukaryotic ancestor had at least one Clade 6-like protein encoded in its genome. This clade is characterized by N termini with no known functional domains and C terminal extensions beyond the PARP catalytic domain of varying lengths. Almost all of these proteins contain a PfamB_2311 domain immediately before their PARP catalytic domain (Figure 4), although the function or significance of this domain is unknown, supporting the placement of these proteins in a single clade. Another characteristic of Clade 6 members is changes within the PARP catalytic domain. None of the Clade 6 proteins we identified contain the final glutamic acid of the HYE catalytic triad, although they mostly retain the histidine and tyrosine (Additional file 11). This might lead to an inability to catalyze poly(ADP-ribosyl)ation. In fact, the human proteins in this clade (PARP6, 8, and 16) have been predicted to have mono(ADP-ribosyl)ation activity based on structural models [32], although this awaits experimental confirmation. None of the Clade 6 PARPs have been functionally characterized.

Clade 6 can be subdivided into five groups (6A-E; Figure 10). Clade 6A contains fungal proteins exclusively (Figure 10; [40,83]). These proteins consist of a long N terminal region containing no known functional domains, a PfamB_2311 domain, the PARP catalytic domain, and a C terminal extension containing an UBCc (Figure 4 and Additional files 12 and 13). The UBCc domain is the catalytic domain contained in E2 Ub-conjugating enzymes (UBCs) [84]. These enzymes carry Ub and transfer it either directly to a substrate in cooperation with an E3 enzyme or to the E3 Ub-ligase. An active cysteine residue [84] characterizes the UBCc domain and is found in Clade 6A proteins (Additional files 12 and 13A-B). In addition, these proteins also share a number of residues conserved across a range of UBCc and UBCc-like domains (Additional files 12 and 13A-B). These include the residues making up the proline-hydrophobic side chain interaction at the top of the so-called E2 fold flap, and a chain of interacting residues at the bottom of the flap (see bolded residues in Additional file 13A-B). These residues have been implicated in the mechanical structure of the E2 fold [85].



Although it is unusual for E2 enzymes to have multiple functional domains, there is at least one other family of such enzymes, the BRUCE-like family, which has multiple domains. These proteins are large (between four and five thousand amino acids) and contain Baculovirus Inhibitor of apoptosis Repeats (BIR; [86,87]) in their N termini, followed by a large region of unknown function, and a UBCC domain at their C termini [88].

No other known functional domains can be identified in Clade 6A proteins; however, most of these proteins do share another PfamB domain, 30617, at their very N termini [41]. This domain is confined to fungal species and appears to only occur in Clade 6A family members with the exception of a protein from the fungus *Uncinocarpus reesii* (EEP82442.1) that consists only of this domain (Additional file 14). Pfam-B_30617 averages 360 amino acids in length and has some secondary structure similarity to the RWD domain when modelled using the Protein Homology/Analogy Recognition Engine (Phyre;

[89]), and is predicted to form an alpha helix/beta strand/alpha helix/beta strand/alpha helix structure (Additional file 13C). The RWA domain has some structural similarity to the UBCC domain [90], further providing a link between the Clade 6A proteins and Ub. The RWA domain is thought to mediate non-catalytic protein-protein interactions. We propose renaming the Pfam-B_30617 domain FPE, for Fungal PARP E2-associated.

Clade 6B proteins are found in a subset of green algae (Figure 10). These proteins have no other domains of known function but do contain PfamB_2311 domains as well as the PARP catalytic domain. Green algae have not previously been shown to have any PARP-like proteins encoded in their genomes. Clade 6C proteins are animal specific and are found in species from across this group, including human (PARP16; Figure 10). Again, other than a PfamB_2311 domain and a PARP catalytic domain, no other obvious protein motifs are present.

Clade 6D is confined to Deuterostomes with the exception of the mollusc *Lottia gigantea*. These proteins consist of no identifiable domains other than a PfamB_2311 domain and the PARP catalytic domain (Figure 4). Human PARP6 and PARP8 are found within this group of proteins.

Clade 6E consist of seven proteins encoded by *Trichomonas vaginalis*, the only member of the Parabasalids (Excavata) with a fully sequenced genome and one fungal protein (*Nectria haematocca* 83215). *Trichomonas* is the causative agent of the sexually transmitted disease trichomoniasis in humans; without other completed genomes available for the parabasalids, it is impossible to determine if members of Clade 6E are found elsewhere in this group. Besides the PARP catalytic domain, the only other identified domain in these proteins is a PfamB_2311 domain. The *Nectria haematocca* protein does not have a PfamB_2311 domain or any known functional domain.

Phylogenetic analysis suggest multiple independent losses of PARP genes across the eukaryotes

Although the five supergroups of eukaryotes with genome information contain organisms with PARP-encoding genes in their genome, some lineages appear to have lost all PARP genes (Figure 2 and Table 1). For example, in Plantae the sequenced genomes available for three red algae and a subset of green algae do not encode any PARP genes (Table 1), although it is possible that such genes may be present in other species not yet sequenced. The complement of PARP proteins present can differ even between closely related species; for example, the green algae *Chlorella sp. NC64A* contains a Clade 6 PARP representative while *Chlorella vulgaris* does not (Figure 2 and Table 1). Diatoms and brown algae (members of the Chromalveolates) do not appear to have PARPs, nor do the sequenced members of the Excavates group Diplomonads. While the sequenced species represent only a small amount of the diversity in these groups of organisms, the lack of PARP genes suggests that these lineages have lost PARPs and, further, demonstrate that these genes are not absolutely essential for eukaryotic life.

The fungal lineages within the Opisthokonts provide a particularly interesting pattern of gene loss. This group of organisms contain Clade 1 and 6 PARP proteins, and based on the phylogenetic distribution of these genes, the fungal ancestor contained proteins representing both clades (Figure 2). However, not all current fungal groups or species have both types of PARPs and some do not encode PARP genes at all (Figure 11A and Table 1). For example, the two major model fungal species, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, do not have PARPs. It appears that there have been at least five

independent losses of PARPs within the fungi. The basal fungi are not well represented by sequenced genomes, however within the Mucorales the genomes of three species have been sequenced and two have Clade 1 PARPs (*Rhizopus oryzae* and *Mucor circinelloides*) while the other has none (*Phycomyces blakesleeianus*). The Basidiomycota has had at least two losses of PARPs; one loss has occurred within the Pucciniomycotina and one within the Agaricomycotina. Only two species within the Pucciniomycotina are represented in our analysis and neither encodes PARP proteins (Table 1). Within the Agaricomycotina, there appear to have been two losses of PARPs. Both Clade 1 and 6 PARPs are found in some species within this group of Basidiomycota; however, *Postia placenta* (Polyporales) has retained only a Clade 1 PARP while *Heterobasidion annosum* (Russulales) has lost both types of PARPs (Figure 11B). The Ascomycota are the fungal group including the most species with sequenced genomes and have both Clade 1 and 6 PARPs (Figure 11A). This group has seen at least two independent losses of PARPs. The Taphrinomycotina (represented by *Schizosaccharomyces pombe*) contain no PARP genes while none of the Saccharomycotina has Clade 6 proteins and only a basal member of this group, *Yarrowia lipolytica*, retains Clade 1 proteins (Figure 11C). Interestingly, as previously noted by other groups [91], PARPs or PARP-like proteins are mostly retained in fungi that have multicellular hyphae and/or elaborate developmental programs, but not in yeasts (Figure 11).

Discussion

Evolutionary history of the PARP family

The broad distribution of PARPs across the eukaryotes indicates that the last common eukaryotic ancestor (LCEA) had genes encoding members of this protein family. Clade 1 PARPs are found in all five eukaryotic supergroups for which sequence information is available; this implies that the LCEA encoded at least one enzyme of this type, and may have had multiple members (Figures 2 and 12A). Based on the domain structure of modern Clade 1 proteins, we hypothesize that the Clade 1 enzyme or enzymes found in the LCEA consisted of WGR, PRD, and PARP catalytic domains (Figure 12B).

Members of Clade 1 have been characterized in a range of organisms, encompassing three of the six eukaryotic supergroups. While a wide range of functions has been described for these PARPs, most characterized members of Clade 1 have been implicated in or demonstrated to have roles in DNA damage response and repair. In Plantae, two of the *Arabidopsis thaliana* Clade 1 members, *AtPARP1* and *AtPARP2*, have been shown to be induced by DNA damage and be involved in the response to it [92,93]. In the Opisthokonts, several

Table 1 Eukaryotic organisms with no identifiable PARP genes in their nuclear genomes

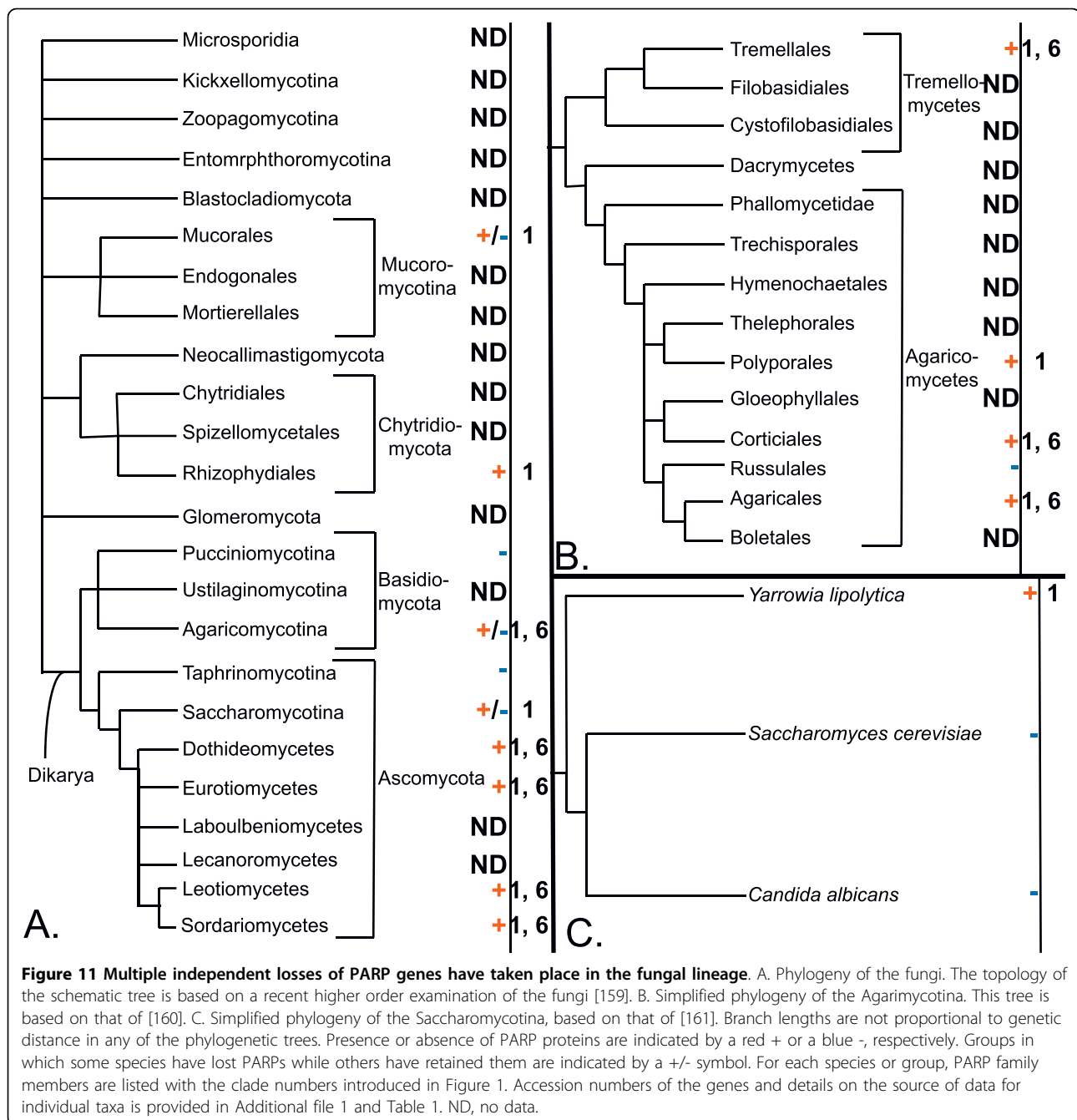
Eukaryotic Supergroup	Rank	Species	Citation or sequencing group	
Plantae	Chlorophyta	<i>Chlorella vulgaris</i>	JGI	
		<i>Ostreococcus lucimarinus</i>	[162]	
		<i>Micromonas pusilla</i>	JGI	
	Red algae	<i>Guillardia theta</i>	JGI	
		<i>Hemiselmis andersenii</i>	NCBI	
		<i>Cyanidioschyzon merolae</i>	[163]	
	Opisthokonts	Fungi	<i>Saccharomyces cerevisiae</i>	[164]
<i>Schizosaccharomyces pombe</i>			[165]	
<i>Pichia stipitis</i>			[166]	
<i>Heterobasidion annosum</i>			JGI	
<i>Melampsora laricis-populina</i>			JGI	
<i>Phycomyces blakesleeenanus</i>			JGI	
<i>Candida albicans</i>			[167]	
<i>Sporobolomyces roseus</i>			JGI	
Chromalveolates			Diatoms	<i>Phaeodactylum tricornutum</i>
	<i>Thalassiosira pseudonana</i>	[169]		
	Brown algae	<i>Aureococcus anophagefferens</i>	JGI	
	Excavates	Diplomonads	<i>Giardia intestinalis</i>	[170]
			<i>Spironucleus vortens</i>	JGI

Species listed have had their genomes individually searched by BLAST to determine if any PARP proteins were present. NCBI, National Center for Biotechnology Information; JGI, Department of Energy Joint Genome Institute.

animal Clade 1 members have been investigated and shown to be involved in DNA repair. This is a well-known function for the human Clade 1 members, PARP1, PARP2, and PARP3 [26,94,95]. In addition, a fungal protein, PrpA from *Aspergillus nidulans*, has been shown to act early in the DNA damage response [91], while loss of its ortholog from *Neurospora crassa*, NPO, causes sensitivity to DNA damage and acceleration of replicative aging [83]. Within the Excavates, a *Trypanosoma cruzi* Clade 1 member, TcPARP, has been shown to be induced in response by DNA damage, be enzymatically activated by nicked DNA and to require DNA for catalytic activity [96]. Clade 1 members in the Chromalveolates and the Amoebozoa have not been functionally characterized, but are also likely to function in DNA damage response. *Dictyostelium discoideum* in the Amoebozoa has at least four Clade 1 proteins encoded in its genome (Figure 3). Drug studies have

implicated PARP activity in oxidative stress response and DNA damage in this organism [97], but no direct evidence of which PARP or PARPs is involved has been published. The ubiquitous distribution of Clade 1 members and the consistent association of the proteins with DNA damage response suggests that this gene lineage is ancient and that the original function of this family was in DNA repair and genome integrity.

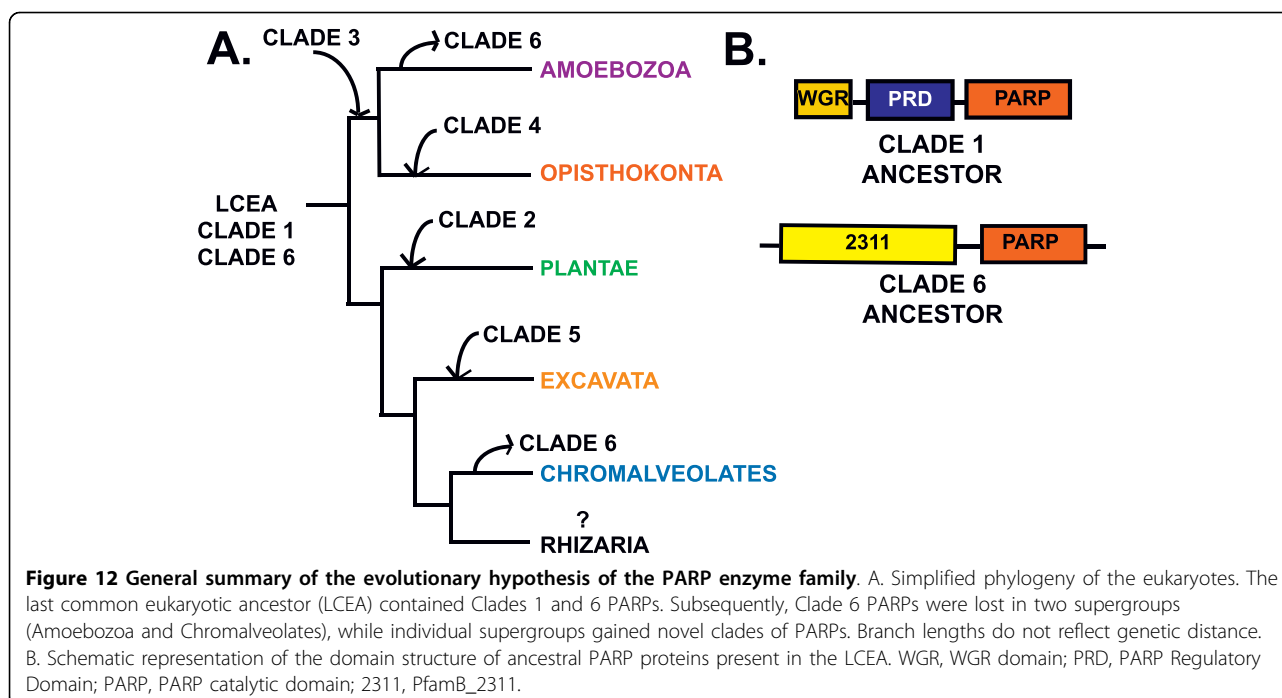
While Clade 6 is found in only three of the five eukaryotic supergroups with available genome information (Opisthokonta, Excavata, and Plantae), the phylogenetic relationship of these groups within eukaryotes suggests that a Clade 6-like protein was found in the LCEA (Figures 2 and 12A). Subsequently, during the eukaryotic radiation, Amoebozoa (or at least *Dictyostelium discoideum*) and Chromalveolates lost Clade 6 PARPs. The ancestral Clade 6 protein was likely to consist of a PfamB_2311 domain N terminal to the PARP catalytic



domain (Figure 12B). Members of Clade 6 were more difficult to identify than other PARPs; it was necessary to do supplemental BLAST searches with the human PARP6 catalytic domain to find most of these proteins (see Methods). This is consistent with the positioning of Clade 6 as sister group to the rest of the PARP superfamily. The fact that Clade 6 PARPs represent an ancient lineage further suggests that changes in the PARP catalytic domain likely to eliminate or change enzymatic activity evolved early in this protein family or,

alternatively, PARP activity evolved from mART activity. It is difficult to speculate on the possible function of the Clade 6 ancestral protein, as none of the extant Clade 6 members have been functionally characterized.

One group of PARPs defined in our study has an unusual distribution. Clade 3 is found in animals (*Opisthokonta*), *Dictyostelium discoideum* (Amoebozoa) and the ciliate *Tetrahymena thermophila* (Chromalveolates), but no other species in our analysis, including the ciliate *Paramecium tetraurelia*. Our phylogenetic tree is based



on the PARP catalytic domain. Clade 3 proteins have evolved to become either mARTs or non-enzymatic (Figure 7; [32]). We propose that the grouping of the Tetrahymena proteins in Clade 3 is an artefact caused by this group of proteins independently beginning to evolve similar changes in the PARP catalytic domain. Clades 3 and 6 independently acquired somewhat similar changes, supporting the idea that changes within the PARP catalytic domain may be constrained in order to preserve overall structure. The hypothesis that the Tetrahymena proteins are not closely related to the other Clade 3 proteins is supported by the fact that one of them (Q22F17) retains the glutamic acid of the PARP catalytic triad, while another (Q24C77) has a conservative substitution of a glutamine at that position and that they do not share any domains outside of the catalytic domain with other members of Clade 3. When more sequences within the ciliates become available, it may become possible to determine if this hypothesis is correct. The Dictyostelium proteins found in Clade 3 may be orthologous to the animal proteins, since one of them has a Macro domain, a domain found in other members of this clade (Figure 4).

In extant eukaryotes, the animal lineage within Opisthokonta appears to have the most diverse collection of PARPs. Most animal genomes encode representatives of at least two clades of PARPs. In addition, a PARP clade has been acquired in this lineage, Clade 4 (Figure 12A). Vertebrates contain the highest number and type of PARPs of any group examined within the

eukaryotes, containing members of Clades 1, 3, 4, 5 and 6; additionally they often encode more than one representative of each clade. However, within animals the nematodes are unusual. *C. elegans*, within the order Rhabditida, only encodes two Clade 1I proteins, PME1 and PME2 (Figure 3), and a protein (PME5) that did not clearly fall into any clade (Additional file 4). Within Clade 1, the nematode 1I PARPs do not group with other animal PARPs but rather are found as the sister group to all of the Clade 1 proteins. PME5 somewhat resembles tankyrases in domain structure but does not group with them. However, the branches leading to the *C. elegans* proteins are long. The length of these branches likely results in long-branch effects, causing misplacement of these proteins within the tree. Such long-branch effects can be caused by the independent acquisition of identical character states [98], phylogenetic signal erosion (“long branch repulsion”) [99], or by symplesiomorphy (retention of an old conserved character state) [100]. In contrast to the situation in *C. elegans*, we were unable to identify any Clade 1 PARPs in the nematode *Brugia malayi*, in the order Spirurida, but did identify a clear tankyrase (Figures 2, 3 and 8). The nematodes are clearly outliers within the animal lineage and a closer examination of the PARP family across a greater number of such species would be interesting.

Although PARPs are found throughout the eukaryotes, these proteins are not essential for eukaryotic life. This is illustrated most clearly in the fungal lineage within the Opisthokonta. In contrast to their fellow Opisthokont

lineage the animals, fungi encode members of only Clades 1 and 6 PARPs (Figures 2 and 11). Lineages within the fungi have independently lost PARPs at least five times, illustrating that eukaryotic organisms do not absolutely require this family of proteins. In addition, it should be noted that none of the fungal species examined retained Clade 6 PARPs in the absence of Clade 1 PARPs. This underscores the relative importance of the so-called "classical" Clade 1 PARPs in these organisms. Interestingly, many of the fungi that have lost all PARPs, including the model fungal systems *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, are yeasts. This suggests fungi with more complex life cycles may retain PARPs more readily than yeasts do. It is possible that a selective advantage is found in organisms with relatively rapid generation times in dispensing with this class of proteins. This is supported by the retention of Clade 1 PARPs in the basal Saccharomycia fungus *Yarrowia lipolytica* while the two other sequenced members of this fungal group have lost all PARPs (Figure 11C). *Yarrowia* can grow in three forms: as yeast, hyphae and pseudohyphae. *Candida albicans*, also a *Saccharomyces* member, is trimorphic but lacks PARPs; however, this diploid organism lacks a known sexual cycle, suggesting a simplification of its life cycle. *Saccharomyces cerevisiae* is only dimorphic, growing only as yeast or pseudohyphae (reviewed in [101]). Other groups have noted the association of retention of PARPs with filamentous growth [91]. This correlation is also found in the dimorphic human pathogen *Histoplasma capsulatum*, the cause of histoplasmosis, which grows as either yeast or hyphae. In this organism, we have found that its Clade 6A PARP gene is expressed only during the filamentous growth stage and not when the fungus is growing in the yeast form (Lee and Lamb, data not shown).

Our conclusions about the function and distribution of PARP proteins in the eukaryotes are limited by the availability of species with sequenced genomes. Currently, there is a dearth of sequences available in many groups of eukaryotes while animals, particularly vertebrates, and fungi are relatively well represented. A number of phylogenetically important groups such as streptophyte algae, glaucophytes, phaeophytes, dinoflagellates, and archamoeba have no sequenced genomes. The eukaryotic supergroup Amoebozoa is represented by only one species, *Dictyostelium discoideum*, while there are no representatives of Rhizaria sequenced. Despite the limitations of the available sequences, we have identified unique types of PARPs in *Naegleria gruberi*, *Trichomonas vaginalis* and green algae and clarified the phylogenetic distribution of tankyrases. There are likely to be additional variations of PARPs discovered as more eukaryotic genomes are sequenced and a further

advancement of our understanding of evolution of this important proteins superfamily.

Clade 5 and vaults

The Clade 5 PARPs have a limited phylogenetic distribution, found only in a subset of animals and amoeba (Figure 9). vPARP was originally identified in a two-hybrid screen using the major vault protein (MVP) protein as bait and shown to act as a *bona fide* PARP [30]. vPARP associates not only with the ribonucleoprotein vault complex, but also can be found in the nucleus, associated with the telomere and the mitotic spindle. The function of vPARP at any of its locations is unclear. Vaults have been best studied in mammals and in these organisms are composed of three proteins, MVP, TEL1 (also found at telomeres), and vPARP. In addition, several vault specific RNAs (vRNAs) are found. The function or functions of vaults are still unclear; they are associated with drug resistance and several signalling pathways (reviewed in [102]), as well as the nuclear pore complex [103,104]. vPARP-deficient mice are normal and fertile with no defects in telomeres or vaults [105]. More recently these mice have been found to develop more tumours in response to carcinogens, suggesting a role in chemically induced cancers [106].

Vaults have been identified in diverse animals and in other eukaryotes such as the amoeba *Dictyostelium discoideum*, flatworms, and trypanosomatids [81,82]. However, vaults appear to be missing from fungi, a number of model animals (*C. elegans* and *Drosophila melanogaster*) and in plants [107-109].

The fact that vPARP does not appear essential for normal development or vault structure in mouse [105] suggests that this protein is not essential for vault function. This may explain why organisms that have been demonstrated to contain vaults in their cells do not always encode proteins that look like vPARP.

Clade 2 plant-specific PARPs are involved in stress responses

In addition to containing three Clade 1 PARPs throughout and Clade 6 PARPs only in the bryophytes, the land plants contain a unique clade of PARP-like proteins. This clade can be subdivided into two subclades, one of which contains proteins with an N terminal WWE domain. Clade 2 is distinct from Clade 3, which also contains proteins with WWE domains. A group within Clade 2, confined to the eudicots within the angiosperms, consists of truncated proteins lacking the N terminal WWE domain. Examination of the phylogeny of Clade 2 clearly illustrates the importance of genome duplication during plant evolution [110-112]; plant species tend to encode gene pairs (Figure 5).

The plant Clade 2 proteins have only been investigated in the model angiosperm *Arabidopsis thaliana*. *Arabidopsis* has two genes, *RCD1* and *SRO1*, which encode full-length members of Clade 2A [64,113]. *RCD1* was originally identified as a stress response gene [63]. It is involved in the response to several abiotic stresses and shows altered hormone accumulation and gene expression [64,114,115]. *rcd1* mutants also display pleiotropic developmental defects including reduced stature, malformed leaves, and early flowering [64]. Loss of *SRO1* causes only minor defects; however *rcd1; sro1* double mutants are severely affected with a majority of individuals dying during embryogenesis [65,69], indicating that this clade of PARP proteins has essential functions in land plants. *RCD1* has been shown to bind to a number of transcription factors, suggesting that Clade 2 PARPs may function in transcriptional regulation [69,113]. *RCD1* does not appear to have catalytic activity, consistent with the absence of the HYE catalytic triad in this protein (Figure 1 and Additional file 7); however, other members of this clade do contain variant HYE motifs that may confer activity (Additional file 7). Therefore, it will be necessary to test individual members of this clade for activity.

Four genes in *Arabidopsis*, *SRO2-5*, encode proteins within Clade 2 that lack the N terminal WWE domain [64,113] and consist of two gene pairs: *SRO2/SRO3* and *SRO4/SRO5* (Figure 5). These genes may be involved in stress signalling; *SRO5* is necessary for response to both salt and oxidative stress [66] and can bind transcription factors [62] and *SRO2* is up regulated in chloroplastic ascorbic peroxidase mutants [116].

Multiple independent acquisitions of mART activity within the PARP superfamily

Although not closely evolutionarily related (Figure 1), the proteins belonging to Clades 3 and 6 have modified their catalytic domains, replacing the glutamic acid of the “HYE” catalytic triad with various other amino acids (Figure 7 and Additional files 8 and 11). The catalytic activity of several human members of Clade 3 has been experimentally investigated. PARP10, which falls into Clade 3A and has an isoleucine instead of a glutamic acid in its catalytic site, has been reported to have auto (ADP-ribosyl)ation activity and modify core histones [33,34]. More recently it was shown to have mono (ADP-ribosyl)ation activity, not poly(ADP-ribosyl)ation activity, and therefore function as a mono(ADP-ribosyl) transferase (mART) rather than a PARP [32]. Molecular modelling suggested that this enzyme uses substrate-assisted catalysis in order to activate the NAD⁺ substrate. This group further demonstrated that PARP14/BAL2, a Clade 3C member with a leucine in place of the glutamic acid, also has mART activity, consistent

with an earlier paper demonstrating auto(ADP-ribosyl)ation activity [35]. A human member of Clade 3F, PARP9/BAL1, has not only replaced the glutamic acid within the catalytic PARP signature but have also replaced the histidine (with a glutamic acid). This enzyme has been shown to be inactive [32,35]. Almost all of the proteins comprising both Clade 3 and Clade 6 have replaced at least the glutamic acid of the “HYE” triad. It is likely that none of these proteins function as bone fide PARPs but rather are either mARTs or are no longer enzymatically active. Clade 3 has a limited taxonomic distribution (Figures 2 and 6); Clade 6, on the other hand, is found in at least three of the six eukaryotic supergroups and was likely present in the LCEA (Figure 12A). This suggests that the evolution of mART activity within the PARP gene family occurred before the full complement of crown groups had formed. In addition, the changes in the catalytic domain of the Clade 2 proteins also suggest that these proteins have altered enzymatic activities (Additional file 6). Therefore, it is likely that mART activity and/or loss of enzymatic activity has evolved at least twice from PARP activity (in Clades 3 and 2) and that mART activity in extant Clade 6 proteins represents an even earlier acquisition of this enzymatic activity.

What functions do PARP-like/mART proteins play? While no members of Clade 6 have been characterized, several members of Clade 3 have, all in mammalian systems. PARP9/BAL1, PARP14/BAL2, and PARP15/BAL3 have been shown to interact with transcription factors and mediate transcriptional repression or activation [35,75,117,118]. PARP13/ZCC2/ZAP has been shown to bind to viral RNA through its zinc fingers and promote degradation of the RNA by the exosome [119-124]. PARP12 shares significant similarity to PARP13 and is thought to function similarly. PARP10 interacts with MYC and inhibits transformation; its overexpression leads to a loss of cell viability [33,34]. To date, no clear consensus about the function of Clade 3 proteins can be formulated.

True tankyrases are confined to animals

Human tankyrase1 was originally identified as a telomeric protein interacting with TRF1, a negative regulator of telomere length. It was shown to act as a PARP and automodify itself as well as TRF1 [19]. A second human tankyrase, tankyrase2 (Figure 4), was identified shortly after the initial discovery of tankyrase1 [28,29,125]. Human tankyrases can be found both in the nucleus [19], at the nuclear pore and centrosome [126], and in the cytoplasm associated with the Golgi or vesicles [127] or the plasma membrane [128]. Since their initial discovery, the known functions of these proteins have expanded to include spindle assembly and vesicle

trafficking (reviewed in [78]), sister chromatid segregation [129], and regulation of the WNT pathway [130-132]. Tankyrases have been identified in a number of animal species, including mouse. In this model organism, it appears tankyrase may not function in telomere length control [133], but its other functions are conserved and its function is essential [134]. Consistent with functions outside of the telomere, a tankyrase is found in *Drosophila melanogaster* (Figure 8; [78]), an organism with a highly divergent telomere consisting of transposons rather than the short repeats found in other eukaryotes [135].

Our phylogenetic tree places a number of proteins previously reported as tankyrases in Clade 1, rather than within Clade 4 (Figures 3 and 8). These proteins do have a different domain structure than tankyrases, sharing ankyrin repeats with tankyrases but having WGR and PRD domains rather than SAM motifs (Figure 4). It is likely that the Clade 1 ankyrin repeat proteins do not share functions with tankyrases.

PME5 from *C. elegans* was reported as a tankyrase and has been functionally characterized. As mentioned above, this protein does not clearly group with any clade, including Clade 4 (Additional file 4). In the original paper describing PME5, it was shown to be more closely related to a *Dictyostelium discoideum* protein we have placed in Clade 1A (Q54E42) and to have a higher similarity within the catalytic domain to human PARP1 than human tankyrase [54]. In addition, the induction of PME5 expression by DNA damaging agents, the increased apoptosis in *pme5(RNAi)* lines after DNA-damage, and the constitutively nuclear chromatin-associated localization of PME5 [53,136] is more consistent with a role in DNA damage. However, the difficulty in placing *C. elegans* PARPs into clades complicates the issue. Further work will need to be done to determine the function of PME5.

Connections between ubiquitination, SUMOylation and poly(ADP-ribosylation)

The attachment of ubiquitin to proteins is an important mechanism in regulating many cellular processes. Similarly to ADP-ribosylation, one to many ubiquitin units can be added to proteins, although only on lysine residues. A chain consisting of at least four ubiquitin linked together by Lys48 residues causes destruction of the protein via the 26S proteasome [137,138], while either monoubiquitination or polyubiquitination with chains linked at Lys63 serve as nonproteolytic signals in such processes as trafficking, DNA repair, and signal transduction [139,140]. Ubiquitination of proteins involves an enzymatic cascade involving ubiquitin-activating (E1), ubiquitin-conjugating (E2), and ubiquitin-ligating (E3) enzymes.

A number of connections between PARP proteins and ubiquitination have emerged. One connection involves the fact that both attachment of ubiquitin and ADP-ribose can be made at lysine residues, suggesting that these post-translational modifications could compete for substrates. In addition, several protein domains found in PARP proteins can also be found in proteins associated with the ubiquitin system (Figure 4). For example, many Clade 1 proteins have BRCT domains; these domains were originally identified in the BRCA1 protein. BRCA1 functions as an E3 ligase in a multi-protein complex in response to DNA damage [141-143]. Within Clade 6, Clade 6A proteins have a UBCc domain, similar to that found in ubiquitin E2s [144], at their C termini, as well as FPE domains at their N termini (Additional Figures 12, 13 and 14). This novel domain has some similarity to the RWD domain, which in turn is related to the UBCc domain, although thought to be non-catalytic. WWE domains are found in Clade 2 and 3 proteins and also in certain ubiquitin E3 ligases [51]. Some Clade 3 proteins have UIM domains, which can bind ubiquitin and polyubiquitin chains [145]; this domain is also found in the BRCA1-interacting protein Rap80 [141]. The *Dictyostelium discoideum* protein DDB0393590 contains a U-box (Figure 4), found in E3 ubiquitin ligases and known to bind E2 enzymes [74].

In addition to the structural similarities found between PARPs and classes of Ub enzymes, some functional connections are also known. Human PARP14/BAL2, a Clade 3E member, has been shown to bind to the multifunctional phosphoglucose isomerase/autocrine motility factor (PGI/AMF). This binding inhibits polyubiquitination of PGI/AMF, stabilizing the protein [146]. PARP1 in humans is regulated by ubiquitination [147] and has been shown to bind to the E2 enzyme hUBC9 [148]. Proteasome-mediated proteolysis of ubiquitinated tankyrase has also been documented; this is promoted by the auto-poly(ADP-ribosylation) of tankyrase, which releases the protein into the cytoplasm [128]. This is similar to the mechanism whereby tankyrase poly(ADP-ribosylates) the telomeric protein TRF1, releasing it from the telomere, allowing its ubiquitination and degradation [149] and the regulation of axin by tankyrase [130]. There are likely to be more connections found in the future between post-translational ADP-ribosylation and ubiquitination.

Recently, a connection between poly(ADP-ribosylation) and SUMOylation has also been demonstrated. PARP1 itself is SUMOylated [150,151], and this takes place within its automodification domain and does not regulate poly(ADP-ribosylation) activity [150]. Rather, PARP1's transcriptional co-activator activity is modified [150,151]. PARP1 can also form higher order complexes and influence SUMOylation of other proteins. In

response to both heat shock and DNA damage, human PARP1 associates with the SUMO E3 ligase PIASy [151,152] and this requires a PAR-binding motif in this protein [152]. Upon DNA damage, PIASy associates with PAR on PARP1 and subsequently its target NEMO binds and is SUMOylated by PIASy, leading to NF- κ B activation [152]. Clearly, the interplay between poly (ADP-ribosyl)ation and other post-translational modifications is just beginning to be explored.

Conclusions

We present here a large-scale phylogenetic analysis of the PARP gene family that extends previous examination of this family. Several main conclusions can be drawn from our study. First, the phylogenetic distribution of the PARP protein family is tremendously broad across the eukaryotes, consistent with the last common ancestor of modern eukaryotes containing at least two PARP-encoding genes. Second, two types of PARP-like proteins were present in the LCEA; one likely functioned in DNA repair and genomic maintenance and resembled modern members of Clade 1. The second probably had mART activity. Third, increasing numbers and types of PARP-like protein are likely to be found as more eukaryotic organisms have their genomes sequenced.

Methods

Retrieval of the PARP gene sequences

The initial sequence set was selected from the Pfam database (<http://pfam.sanger.ac.uk/>; [41-43]), using the sequences identified as members of the PARP family (PF00644). The full sequences of the proteins were retrieved from UniProt [44,45], using the links provided by Pfam. Additional sequences were retrieved from other eukaryotic organisms at the DOE Joint Genome Institute (JGI; <http://www.jgi.doe.gov/>), the Broad Institute <http://www.broadinstitute.org>, the J. Craig Venter Institute <http://www.jcvi.org/>, ToxoDB (<http://toxodb.org/toxo/>; [47]), and the Arabidopsis Information Resource (TAIR; <http://www.arabidopsis.org/>) using BLAST searches [46] based on human or *Arabidopsis thaliana* PARP catalytic domain sequences as search queries. Specific phylogenetically interesting genomes were also individually searched by BLAST to confirm the absence of PARP proteins (see Table 1). The catalytic domains of most retrieved sequences were delineated using Pfam. Sequences in Clade 6 have lower similarity to the classical PARPs (i.e. Clade 1) used to generate the Pfam HMM, so the PARP catalytic domains for these sequences were identified using BLAST searches based on human PARP6 catalytic domain as the query and identifying the region of retrieved sequences that had similarity to this PARP signature. In addition, many

sequences whose catalytic domain was incompletely identified by Pfam were completed by BLAST searches using closely related complete PARP catalytic domains from other closely related species, in order to provide as much sequence information as possible for the alignment and phylogeny inference. The identified PARP catalytic domains were extracted using the extract.pl tool in the Wildcat Toolbox set of Perl utilities (http://proteomics.arizona.edu/wildcat_toolbox; [153]). Sequences of less than 100 amino acids in length and many that were missing important structural elements of the PARP domain were discarded to allow better alignment and phylogenetic signal recovery. Many of these sequences were obtained from shotgun sequencing and are presumably incomplete.

Phylogenetic analyses

The collected PARP catalytic domains were aligned using the MUSCLE3.8.31 multiple alignment tool, using default settings [50]. The multiple alignment was subjected to a maximum-likelihood (ML) analysis using PhyML3.0 [154] using the computer facilities at the Ohio Supercomputer Center <http://www.osc.edu>. The substitution model parameters using for the PhyML analysis were the WAG substitution matrix, Γ 8+I correction to model site rate heterogeneity and empirical equilibrium frequencies. These parameters were selected as the optimal substitution model based on analysis by ProtTest v2.4 [155]. A parsimony-based starting tree was used. Branch supports were computed in PhyML using an aLRT non-parametric Shimodaira-Hasegawa-like (SH) procedure [156]. Once a tree with all PARP domains had been generated, it was used to identify the six clades referred to in the text in combination with examination of domains outside of the PARP catalytic domain. After the six clades were defined, sequences from each clade were aligned separately using MUSCLE. These alignments were used to generate individual clade trees using PhyML with identical parameters. The phylogenetic trees were generated for figures using FigTree <http://tree.bio.ed.ac.uk/software/figtree>. Alignment figures were generated using TEXshade [157] and Jalview [158].

Prediction of protein domains

After sequences of PARP family members were retrieved and placed into clades, the sequences were checked for other domains at the Pfam website [41]. Domains identified are shown in Figure 4. PfamB_30617 was identified in Clade 6A fungal proteins and extracted aligned as above. This domain was further analyzed using the Protein homology/analogy recognition engine (Phyre) [89] and renamed FPE (Fungal PARP E2-associated). Subsequently, a consensus FPE sequence was used in

BLAST searches to find other proteins containing this region. The UBCc domains from Clade 6A proteins were similarly processed.

Additional material

Additional file 1: List of all protein sequences used in our study.

The protein name is given as species followed by accession number. The eukaryotic supergroup, phylum, class and order are also indicated. The source of the sequence is given as well as the link to the sequence. The proteins are arranged alphabetically by species.

Additional file 2: Alignment of protein sequences removed from analysis due to incomplete PARP catalytic domains.

Additional file 3: Multiple alignment of the PARP catalytic domain. These alignments only show the conserved PARP catalytic domain. Only sequences containing complete domains were used to generate the alignment.

Additional file 4: Phylogenetic analysis of PARP genes in eukaryotes. The represented tree is a ML tree, based on an alignment of the PARP catalytic domain. Branch supports as in Figure 1. Scale bar indicates genetic distance reflected in branch length.

Additional file 5: Multiple alignment of the PARP catalytic domains of Clade 1 PARP proteins. These alignments only show the conserved PARP catalytic domain and the numbers indicate amino acids within the catalytic domain. Due to the large number of proteins in Clade 1, some needed to be removed in order to annotate the sequence. Dots indicate gaps introduced to optimize the alignment and identical amino acids indicated by red shading and similar amino acids indicated by orange shading. The structural elements present in *Gallus gallus* PARP1 are shown at the bottom of the alignment, with the six "core" β strands indicated [6]. Subclades are separated from one another by spaces with Clade 1A at the top. The amino acids of the HYE catalytic triad are boxed in blue and labelled C1 (H), C2 (Y) and C3 (E).

Additional file 6: Multiple alignment of the PARP catalytic domains of Clade 1 PARP proteins.

Additional file 7: Multiple alignment of the PARP catalytic domains of Clade 2 PARP proteins annotated with structural predictions.

These alignments only show the conserved PARP catalytic domain. The structural elements predicted to be present in *Arabidopsis thaliana* RCD1 by Phyre are shown at the bottom of the alignment [89]. Annotations as in Additional file 5.

Additional file 8: Multiple alignment of the PARP catalytic domains of Clade 3 proteins annotated with structural information. These alignments only show the conserved PARP catalytic domain. The structural elements present in *Homo sapiens* PARP15 are shown at the bottom of the alignment. The crystal structure of PARP15 is available in the Protein Data Bank (PDB; <http://www.rcsb.org/pdb/explore/explore.do?structureId=3GEY>; [171]). Annotations as in Additional file 5.

Additional file 9: Multiple alignment of the PARP catalytic domain from Clade 4 PARP proteins annotated with structural information.

These alignments only show the conserved PARP catalytic domain. The structural elements present in *Homo sapiens* TNK1 are shown at the bottom of the alignment [172]. Annotations as in Additional file 5.

Additional file 10: Multiple alignment of the PARP catalytic domain from Clade 5 PARP proteins.

These alignments only show the conserved PARP catalytic domain. The structural elements present in *Homo sapiens* vPARP are shown at the bottom of the alignment. Annotations as in Additional file 5.

Additional file 11: Multiple alignment of the PARP catalytic domain from Clade 6 PARP proteins annotated with structural information.

These alignments only show the conserved PARP catalytic domain. The structural elements predicted to be present in *Homo sapiens* PARP8 by Phyre are shown at the bottom of the alignment [89]. Annotations as in Additional file 6.

Additional file 12: Multiple alignment of the UBCc domain of Clade 6A PARPs. The entire UBCc domains as defined by Pfam from Clade 6A proteins are shown.

Additional file 13: Clade 6A PARP proteins contain FPE and UBCc domains. A. Clade 6A PARPs contain UBCc domains in their C termini. An alignment of the HMM consensus sequence of the UBCc domain from Pfam (UBCc) and the UBCc domain from *Phaeosphaeria nodorum* QOUPJ2 (Pn6F). The sequence similarity between the UBCc and UBCc-like domains is shown in red (CONS). +, similar amino acids; -, gaps introduced to maximize the alignment; ., any amino acid. Residues in bold have been shown to be diagnostic of UBCc domains as discussed in the text and [85]. B. Alignment of a region of the Clade 6A PARP UBCc-like domains, containing the catalytic cysteine. The names of the proteins and the amino acid positions (within the UBCc domain) are indicated at left. The blue asterisk marks a histidine and the red asterisk marks the catalytic cysteine, both shared with typical UBCc domains. C. The FPE domain consists of alpha helices and beta strands. The sequence of the FPE domain from the *Phaeosphaeria nodorum* Clade 6A member (QOUPJ2) is shown. Secondary structural characteristics as detected by Phyre are shown above the sequence. h, alpha helices; e, beta strands.

Additional file 14: Multiple alignment of the FPE from Clade 6A PARP proteins. The entire PfamB_30617/FPE domains as defined by Pfam from Clade 6A proteins are shown.

Acknowledgements

We thank Dr. Iris Meier (Ohio State University) and two anonymous reviewers for critical reading of the manuscript and members of the Lamb laboratory for discussions. This work was supported in part by an allocation of computing time from the Ohio Supercomputer Center. This work was supported by a grant from the Ohio Plant Biotechnology Consortium to RSL and by funds from the Ohio State University.

Author details

¹Plant Cellular and Molecular Biology Department, Ohio State University, 500 Aronoff Laboratory, 318 W. 12th Ave., Columbus, OH 43210 USA. ²Molecular, Cellular and Developmental Biology Program, Ohio State University, Columbus, OH 43210 USA.

Authors' contributions

MC retrieved the sequences, made the sequence alignments, and performed the phylogenetic analyses. RSL performed the domain analysis using Pfam and Phyre. ST, MC and RSL participated in data analysis and figure preparation. MC and RSL participated in the design and coordination of the study. RSL drafted the manuscript and all the authors participated in the editing of the manuscript. All the authors read and approved the final manuscript.

Received: 7 October 2009 Accepted: 13 October 2010

Published: 13 October 2010

References

1. Chambon P, Weill JD, Mandel P: Nicotinamide mononucleotide activation of new DNA-dependent polyadenylic acid synthesizing nuclear enzyme. *Biochem Biophys Res Commun* 1963, **11**:39-43.
2. Fujimura S, Hasegawa S, Shimizu Y, Sugimura T: Polymerization of the adenosine 5'-diphosphate-ribose moiety of nicotinamide-adenine dinucleotide by nuclear enzyme. I. Enzymatic reactions. *Biochim Biophys Acta* 1967, **145**(247-259).
3. Chambon P, Weill JD, Doly J, Strosser MT, Mandel P: On the formation of a novel adenylic compound by enzymatic extracts of liver nuclei. *Biochem Biophys Res Commun* 1966, **25**:638-643.
4. Nishizuka Y, Ueda K, Nakazawa K, Hayaishi O: Studies on the polymer of adenosine diphosphate ribose. I. Enzymic formation from nicotinamide adenine dinucleotide in mammalian nuclei. *J Biol Chem* 1967, **242**(13):3164-3171.

5. Doly J, Petek F: **Etude de la structure d'un composé "poly(ADP-ribose)" synthétisé par des extraits nucléaires de foie de poulet.** *CR Hebd Seanc Acad Sci Ser D Sci Nat* 1966, **263**:1341-1344.
6. Ruf A, Mennissier de Murcia J, de Murcia G, Schulz GE: **Structure of the catalytic fragment of poly(AD-ribose) polymerase from chicken.** *Proc Natl Acad Sci USA* 1996, **93**(15):7481-7485.
7. Oliver AW, Ame JC, Roe SM, Good V, de Murcia G, Pearl LH: **Crystal structure of the catalytic fragment of murine poly(ADP-ribose) polymerase-2.** *Nucleic Acids Res* 2004, **32**(2):456-464.
8. Hottiger MO, Hassa PO, Luscher B, Schuler H, Koch-Nolte F: **Toward a unified nomenclature for mammalian ADP-ribosyltransferases.** *Trends Biochem Sci* 2010, **35**(4):208-219.
9. Ame JC, Spenlehauer C, de Murcia G: **The PARP superfamily.** *Bioessays* 2004, **26**(8):882-893.
10. Kim MY, Zhang T, Kraus WL: **Poly(ADP-ribose)ylation by PARP-1: 'PAR-laying' NAD+ into a nuclear signal.** *Genes Dev* 2005, **19**(17):1951-1967.
11. Schreiber V, Dantzer F, Ame JC, de Murcia G: **Poly(ADP-ribose): novel functions for an old molecule.** *Nat Rev Mol Cell Biol* 2006, **7**(7):517-528.
12. Hassa PO, Hottiger MO: **The diverse biological roles of mammalian PARPs, a small but powerful family of poly-ADP-ribose polymerases.** *Front Biosci* 2008, **13**:3046-3082.
13. Hassa PO, Haenni SS, Elser M, Hottiger MO: **Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going?** *Microbiol Mol Biol Rev* 2006, **70**(3):789-829.
14. Pachter P, Szabo C: **Role of the peroxynitrite-poly(ADP-ribose) polymerase pathway in human disease.** *Am J Pathol* 2008, **173**(1):2-13.
15. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, Mortimer P, Swaisland H, Lau A, O'Connor MJ, Ashworth A, Carmichael J, Kaye SB, Schellens JH, de Bono JS: **Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers.** *N Engl J Med* 2009, **361**(2):123-134.
16. Cottet F, Blanche H, Verasdonck P, Le Gall I, Schachter F, Burkle A, Muirais ML: **New polymorphisms in the human poly(ADP-ribose) polymerase-1 coding sequence: lack of association with longevity or with increased cellular poly(ADP-ribose)ylation capacity.** *J Mol Med* 2000, **78**(8):431-440.
17. Tezcan G, Gurel CB, Tutluoglu B, Onaran I, Kanigur-Sultuybek G: **The Ala allele at Val762Ala polymorphism in poly(ADP-ribose) polymerase-1 (PARP-1) gene is associated with a decreased risk of asthma in a Turkish population.** *J Asthma* 2009, **46**(4):371-374.
18. Babiychuk E, Cottrill PB, Storozhenko S, Fuangthong M, Chen Y, O'Farrell MK, Van Montagu M, Inze D, Kushnir S: **Higher plants possess two structurally different poly(ADP-ribose) polymerases.** *Plant J* 1998, **15**(5):635-645.
19. Smith S, Giriati I, Schmitt A, de Lange T: **Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.** *Science* 1998, **282**(5393):1484-1487.
20. Wang ZQ, Auer B, Stingl L, Berghammer H, Haidacher D, Schweiger M, Wagner EF: **Mice lacking ADPRT and poly(ADP-ribose)ylation develop normally but are susceptible to skin disease.** *Genes Dev* 1995, **9**(5):509-520.
21. Wang ZQ, Stingl L, Morrison C, Jantsch M, Los M, Schulze-Osthoff K, Wagner EF: **PARP is important for genomic stability but dispensable in apoptosis.** *Genes Dev* 1997, **11**(18):2347-2358.
22. Uchida K, Morita T, Sato T, Ogura T, Yamashita R, Noguchi S, Suzuki H, Nyunoya H, Miwa M, Sugimura T: **Nucleotide sequence of a full-length cDNA for human fibroblast poly(ADP-ribose) polymerase.** *Biochem Biophys Res Commun* 1987, **148**(2):617-622.
23. Mahajan PB, Zuo Z: **Purification and cDNA cloning of maize Poly(ADP-ribose) polymerase.** *Plant Physiol* 1998, **118**(3):895-905.
24. Podesta D, Garcia-Herreros MI, Cannata JJ, Stoppani AO, Fernandez Villamil SH: **Purification and properties of poly(ADP-ribose)polymerase from *Crithidia fasciculata*. Automodification and poly(ADP-ribose)ylation of DNA topoisomerase I.** *Mol Biochem Parasitol* 2004, **135**(2):211-219.
25. Johansson M: **A human poly(ADP-ribose) polymerase gene family (ADPRTL): cDNA cloning of two novel poly(ADP-ribose) polymerase homologues.** *Genomics* 1999, **57**(3):442-445.
26. Ame JC, Rolli V, Schreiber V, Niedergang C, Apiou F, Decker P, Muller S, Hoger T, Mennissier-de Murcia J, de Murcia G: **PARP-2, A novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase.** *J Biol Chem* 1999, **274**(25):17860-17868.
27. Rippmann JF, Damm K, Schnapp A: **Functional characterization of the poly(ADP-ribose) polymerase activity of tankyrase 1, a potential regulator of telomere length.** *J Mol Biol* 2002, **323**(2):217-224.
28. Kuimov AN, Kuprash DV, Petrov VN, Vdovichenko KK, Scanlan MJ, Jongeneel CV, Lagarkova MA, Nedospasov SA: **Cloning and characterization of TNKL, a member of tankyrase gene family.** *Genes Immun* 2001, **2**(1):52-55.
29. Lyons RJ, Deane R, Lynch DK, Ye ZS, Sanderson GM, Eyre HJ, Sutherland GR, Daly RJ: **Identification of a novel human tankyrase through its interaction with the adaptor protein Grb14.** *J Biol Chem* 2001, **276**(20):17172-17180.
30. Kickhoefer VA, Siva AC, Kedersha NL, Inman EM, Ruland C, Streuli M, Rome LH: **The 193-kD vault protein, VPARP, is a novel poly(ADP-ribose) polymerase.** *J Cell Biol* 1999, **146**(5):917-928.
31. Marsischky GT, Wilson BA, Collier RJ: **Role of glutamic acid 988 of human poly-ADP-ribose polymerase in polymer formation. Evidence for active site similarities to the ADP-ribosylating toxins.** *J Biol Chem* 1995, **270**(7):3247-3254.
32. Kleine H, Poreba E, Lesniewicz K, Hassa PO, Hottiger MO, Litchfield DW, Shilton BH, Luscher B: **Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation.** *Mol Cell* 2008, **32**(1):57-69.
33. Yu M, Schreek S, Cerni C, Schamberger C, Lesniewicz K, Poreba E, Vervoorts J, Walsemann G, Grotzinger J, Kremmer E, Mehraein Y, Mertsching J, Kraft R, Austen M, Luscher-Firzlaff J, Luscher B: **PARP-10, a novel Myc-interacting protein with poly(ADP-ribose) polymerase activity, inhibits transformation.** *Oncogene* 2005, **24**(12):1982-1993.
34. Chou HY, Chou HT, Lee SC: **CDK-dependent activation of poly(ADP-ribose) polymerase member 10 (PARP10).** *J Biol Chem* 2006, **281**(22):15201-15207.
35. Aguiar RC, Takeyama K, He C, Kreinbrink K, Shipp MA: **B-aggressive lymphoma family proteins have unique domains that modulate transcription and exhibit poly(ADP-ribose) polymerase activity.** *J Biol Chem* 2005, **280**(40):33756-33765.
36. Augustin A, Spenlehauer C, Dumond H, Mennissier-De Murcia J, Piel M, Schmit AC, Apiou F, Vonesch JL, Kock M, Bornens M, De Murcia G: **PARP-3 localizes preferentially to the daughter centriole and interferes with the G1/S cell cycle progression.** *J Cell Sci* 2003, **116**(Pt 8):1551-1562.
37. Loseva O, Jemth AS, Bryant HE, Schuler H, Lehtio L, Karlberg T, Helleday T: **Poly(ADP-ribose) polymerase-3 (PARP-3) is a mono-ADP ribosylase that activates PARP-1 in absence of DNA.** *J Biol Chem* 2010, **285**(11):8054-8064.
38. Simpson AG, Roger AJ: **The real 'kingdoms' of eukaryotes.** *Curr Biol* 2004, **14**(17):R693-696.
39. Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW: **The tree of eukaryotes.** *Trends Ecol Evol* 2005, **20**(12):670-676.
40. Otto H, Reche PA, Bazan F, Dittmar K, Haag F, Koch-Nolte F: **In silico characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs).** *BMC Genomics* 2005, **6**:139.
41. Coggill P, Finn RD, Bateman A: **Identifying protein domains with the Pfam database.** *Curr Protoc Bioinformatics* 2008, **Chapter 2, Unit 2.5**.
42. Sammut SJ, Finn RD, Bateman A: **Pfam 10 years on: 10,000 families and still growing.** *Brief Bioinform* 2008, **9**(3):210-219.
43. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A: **The Pfam protein families database.** *Nucleic Acids Res* 2008, **36** Database: D281-288.
44. Jain E, Bairoch A, Duvaud S, Phan I, Redaschi N, Suzek BE, Martin MJ, McGarvey P, Gasteiger E: **Infrastructure for the life sciences: design and implementation of the UniProt website.** *BMC Bioinformatics* 2009, **10**:136.
45. The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res* 2009, **37** Database: D169-174.
46. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**(17):3389-3402.
47. Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X, Heiges M, Iodice J, Kissinger JC, Mackey AJ, Pinney DF, Roos DS, Stoeckert CJ Jr, Wang H, Brunk BP: **ToxoDB: an integrated *Toxoplasma gondii* database resource.** *Nucleic Acids Res* 2008, **36** Database: D553-556.
48. Fey P, Gaudet P, Curk T, Zupan B, Just EM, Basu S, Merchant SN, Bushmanova YA, Shauly G, Kibbe WA, Chisholm RL: **dictyBase—a Dictyostelium bioinformatics resource update.** *Nucleic Acids Res* 2009, **37** Database: D515-519.
49. Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E: **The Arabidopsis Information**

- Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res* 2008, **36** Database: D1009-1014.
50. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004, **32**(5):1792-1797.
51. Aravind L: The WWE domain: a common interaction module in protein ubiquitination and ADP ribosylation. *Trends Biochem Sci* 2001, **26**(5):273-275.
52. Ohi MD, Link AJ, Ren L, Jennings JL, McDonald WH, Gould KL: Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol Cell Biol* 2002, **22**(7):2011-2024.
53. White C, Gagnon SN, St-Laurent JF, Gravel C, Proulx LI, Desnoyers S: The DNA damage-inducible *C. elegans* tankyrase is a nuclear protein closely linked to chromosomes. *Mol Cell Biochem* 2009, **324**(1-2):73-83.
54. Gravel C, Stergiou L, Gagnon SN, Desnoyers S: The *C. elegans* gene pme-5: molecular cloning and role in the DNA-damage response of a tankyrase orthologue. *DNA Repair (Amst)* 2004, **3**(2):171-182.
55. Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV: A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J* 1997, **11**(1):68-76.
56. Staub E, Fiziev P, Rosenthal A, Hinzmann B: Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. *Bioessays* 2004, **26**(5):567-581.
57. Kameshita I, Matsuda Z, Taniguchi T, Shizuta Y: Poly (ADP-Ribose) synthetase. Separation and identification of three proteolytic fragments as the substrate-binding domain, the DNA-binding domain, and the automodification domain. *J Biol Chem* 1984, **259**(8):4770-4776.
58. Langelier MF, Servent KM, Rogers EE, Pascal JM: A third zinc-binding domain of human poly(ADP-ribose) polymerase-1 coordinates DNA-dependent enzyme activation. *J Biol Chem* 2008, **283**(7):4105-4114.
59. Aravind L, Koonin EV: SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* 2000, **25**(3):112-114.
60. Lynch M, Force A: The probability of duplicate gene preservation by subfunctionalization. *Genetics* 2000, **154**(1):459-473.
61. Gagnon SN, Hengartner MO, Desnoyers S: The genes pme-1 and pme-2 encode two poly(ADP-ribose) polymerases in *Caenorhabditis elegans*. *Biochem J* 2002, **368**(Pt 1):263-271.
62. Jaspers P, Overmyer K, Wrzaczek M, Vainonen JP, Blomster T, Salojarvi J, Reddy RA, Kangasjarvi J: The RST and PARP-like domain containing SRO protein family: analysis of protein structure, function and conservation in land plants. *BMC Genomics* 2010, **11**:170.
63. Overmyer K, Tuominen H, Kettunen R, Betz C, Langebartels C, Sandermann H Jr, Kangasjarvi J: Ozone-sensitive arabidopsis *rcd1* mutant reveals opposite roles for ethylene and jasmonate signaling pathways in regulating superoxide-dependent cell death. *Plant Cell* 2000, **12**(10):1849-1862.
64. Ahlfors R, Lang S, Overmyer K, Jaspers P, Brosche M, Tauriainen A, Kollist H, Tuominen H, Belles-Boix E, Phippo M, Inze D, Palva ET, Kangasjarvi J: Arabidopsis RADICAL-INDUCED CELL DEATH1 belongs to the WWE protein-protein interaction domain protein family and modulates abscisic acid, ethylene, and methyl jasmonate responses. *Plant Cell* 2004, **16**(7):1925-1937.
65. Teotia S, Lamb RS: The paralogous genes *RADICAL-INDUCED CELL DEATH1* and *SIMILAR TO RCD ONE1* have partially redundant functions during Arabidopsis development. *Plant Physiol* 2009, **151**(1):180-198.
66. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK: Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* 2005, **123**(7):1279-1291.
67. Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF: Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet* 2008, **42**:443-461.
68. Lawton-Rauh A: Evolutionary dynamics of duplicated genes in plants. *Mol Phylogenet Evol* 2003, **29**(3):396-409.
69. Jaspers P, Blomster T, Brosche M, Salojarvi J, Ahlfors R, Vainonen JP, Reddy RA, Immink R, Angenent G, Turck F, Overmyer K, Kangasjarvi J: Unequally redundant *RCD1* and *SRO1* mediate stress and developmental responses and interact with transcription factors. *Plant J* 2009, **60**(2):268-279.
70. Birney E, Kumar S, Krainer AR: Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 1993, **21**(25):5803-5816.
71. Swanson KA, Kang RS, Stamenova SD, Hicke L, Radhakrishnan I: Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. *EMBO J* 2003, **22**(18):4597-4606.
72. Karras GI, Kustatscher G, Buhecha HR, Allen MD, Pugieux C, Sait F, Bycroft M, Ladurner AG: The macro domain is an ADP-ribose binding module. *EMBO J* 2005, **24**(11):1911-1920.
73. Aravind L, Koonin EV: The U box is a modified RING finger - a common domain in ubiquitination. *Curr Biol* 2000, **10**(4):R132-134.
74. Hatakeyama SNK: U-box proteins as a new family of ubiquitin ligases. *Biochem Biophys Res Commun* 2003, **302**(4):635-645.
75. Aguiar RC, Yakushijin Y, Kharbanda S, Salgia R, Fletcher JA, Shipp MA: BAL is a novel risk-related gene in diffuse large B-cell lymphomas that enhances cellular migration. *Blood* 2000, **96**(13):4328-4334.
76. Ponting CP: SAM: a novel motif in yeast sterile and *Drosophila* polyhomeotic proteins. *Protein Sci* 1995, **4**(9):1928-1930.
77. Cook BD, Dynek JN, Chang W, Shostak G, Smith S: Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2 at human telomeres. *Mol Cell Biol* 2002, **22**(1):332-342.
78. Hsiao SJ, Smith S: Tankyrase function at telomeres, spindle poles, and beyond. *Biochimie* 2008, **90**(1):83-92.
79. Bork P, Rohde K: More von Willebrand factor type A domains? Sequence similarities with malaria thrombospondin-related anonymous protein, dihydropyridine-sensitive calcium channel and inter-alpha-trypsin inhibitor. *Biochem J* 1991, **279**(Pt 3):908-910.
80. Kedersha NL, Rome LH: Isolation and characterization of a novel ribonucleoprotein particle: large structures contain a single species of small RNA. *J Cell Biol* 1986, **103**(3):699-709.
81. Vasu SK, Rome LH: Dictyostelium vaults: disruption of the major proteins reveals growth and morphological defects and uncovers a new associated protein. *J Biol Chem* 1995, **270**(28):16588-16594.
82. Kedersha NL, Miquel MC, Bittner D, Rome LH: Vaults. II. Ribonucleoprotein structures are highly conserved among higher and lower eukaryotes. *J Cell Biol* 1990, **110**(4):895-901.
83. Kothe GO, Kitamura M, Masutani M, Selker EU, Inoue H: PARP is involved in replicative aging in *Neurospora crassa*. *Fungal Genetics and Biology* 2010, **47**(4):297-309.
84. Cook WJ, Jeffrey LC, Sullivan ML, Vierstra RD: Three-dimensional structure of a ubiquitin-conjugating enzyme (E2). *J Biol Chem* 1992, **267**(21):15116-15121.
85. Burroughs AM, Jaffee M, Iyer LM, Aravind L: Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. *J Struct Biol* 2008, **162**(2):205-218.
86. Birnbaum MJ, Clem RJ, Miller LK: An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. *J Virol* 1994, **68**(4):2521-2528.
87. Verhagen AM, Coulson EJ, Vaux DL: Inhibitor of apoptosis proteins and their relatives: IAPs and other BIRPs. *Genome Biol* 2001, **2**(7), REVIEWS3009.
88. Bartke T, Pohl C, Pyrowolakis G, Jentsch S: Dual role of BRUCE as an antiapoptotic IAP and a chimeric E2/E3 ubiquitin ligase. *Mol Cell* 2004, **14**(6):801-811.
89. Kelley LA, Sternberg MJ: Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 2009, **4**(3):363-371.
90. Doerks T, Copley RR, Schultz J, Ponting CP, Bork P: Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res* 2002, **12**(1):47-56.
91. Semighini CP, Savoldi M, Goldman GH, Harris SD: Functional characterization of the putative *Aspergillus nidulans* poly(ADP-ribose) polymerase homolog PrpA. *Genetics* 2006, **173**(1):87-98.
92. Vanderauwera S, De Block M, Van de Steene N, van de Cotte B, Metzclaff M, Van Breusegem F: Silencing of poly(ADP-ribose) polymerase in plants alters abiotic stress signal transduction. *Proc Natl Acad Sci USA* 2007, **104**(38):15150-15155.
93. Doucet-Chabeaud G, Godon C, Brutusco C, de Murcia G, Kazmaier M: Ionising radiation induces the expression of *PARP-1* and *PARP-2* genes in Arabidopsis. *Mol Genet Genomics* 2001, **265**(6):954-963.
94. Trucco C, Oliver FJ, de Murcia G, Menissier-de Murcia J: DNA repair defect in poly(ADP-ribose) polymerase-deficient cell lines. *Nucleic Acids Res* 1998, **26**(11):2644-2649.
95. Rouleau M, McDonald D, Gagne P, Ouellet ME, Droit A, Hunter JM, Dutertre S, Prigent C, Hendzel MJ, Poirier GG: PARP-3 associates with

- polycarbonyl group bodies and with components of the DNA damage repair machinery. *J Cell Biochem* 2007, **100**(2):385-401.
96. Fernandez Villamil SH, Baltanas R, Alonso GD, Vilchez Larrea SC, Torres HN, Flawia MM: **TcPARP: A DNA damage-dependent poly(ADP-ribose) polymerase from Trypanosoma cruzi.** *Int J Parasitol* 2008, **38**(3-4):277-287.
 97. Rajawat J, Vohra I, Mir HA, Gohel D, Begum R: **Effect of oxidative stress and involvement of poly(ADP-ribose) polymerase (PARP) in Dictyostelium discoideum development.** *FEBS J* 2007, **274**(21):5611-5618.
 98. Felsenstein J: **Cases in which parsimony or compatibility methods will be positively misleading.** *Syst Zool* 1978, **27**:401-410.
 99. Pol D, Siddall ME: **Biases in maximum likelihood and parsimony: a simulation approach to a 10-taxon case.** *Cladistics* 2001, **17**:266-281.
 100. Hennig W: **Phylogenetic Systematics.** Urbana, IL: University of Illinois Press 1966.
 101. Bastidas RJ, Heitman J: **Trimorphic stepping stones pave the way to fungal virulence.** *Proc Natl Acad Sci USA* 2009, **106**(2):351-352.
 102. Berger W, Steiner E, Grusch M, Elbling L, Micksche M: **Vaults and the major vault protein: novel roles in signal pathway regulation and immunity.** *Cell Mol Life Sci* 2009, **66**(1):43-61.
 103. Dickenson NE, Moore D, Suprenant KA, Dunn RC: **Vault ribonucleoprotein particles and the central mass of the nuclear pore complex.** *Photochem Photobiol* 2007, **83**(3):686-691.
 104. Vollmar F, Hacker C, Zahedi RP, Sickmann A, Ewald A, Scheer U, Dabauvalle MC: **Assembly of nuclear pore complexes mediated by major vault protein.** *J Cell Sci* 2009, **122**(Pt 6):780-786.
 105. Liu Y, Snow BE, Kickhoefer VA, Erdmann N, Zhou W, Wakeham A, Gomez M, Rome LH, Harrington L: **Vault poly(ADP-ribose) polymerase is associated with mammalian telomerase and is dispensable for telomerase function and vault structure in vivo.** *Mol Cell Biol* 2004, **24**(12):5314-5323.
 106. Raval-Fernandes S, Kickhoefer VA, Kitchen C, Rome LH: **Increased susceptibility of vault poly(ADP-ribose) polymerase-deficient mice to carcinogen-induced tumorigenesis.** *Cancer Res* 2005, **65**(19):8846-8852.
 107. Kickhoefer VA, Vasu SK, Rome LH: **Vaults are the answer, what is the question?** *Trends Cell Biol* 1996, **6**(5):174-178.
 108. van Zon A, Mossink MH, Scheper RJ, Sonneveld P, Wiemer EA: **The vault complex.** *Cell Mol Life Sci* 2003, **60**(9):1828-1837.
 109. Suprenant KA: **Vault ribonucleoprotein particles: sarcophagi, gondolas, or safety deposit boxes?** *Biochemistry* 2002, **41**(49):14447-14454.
 110. Bowers JE, Chapman BA, Rong J, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events.** *Nature* 2003, **422**(6930):433-438.
 111. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y: **Modeling gene and genome duplications in eukaryotes.** *Proc Natl Acad Sci USA* 2005, **102**(15):5454-5459.
 112. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH: **Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps.** *Genome Res* 2008, **18**(12):1944-1954.
 113. Belles-Boix E, Babiychuk E, Van Montagu M, Inze D, Kushnir S: **CEO1, a new protein from Arabidopsis thaliana, protects yeast against oxidative damage.** *FEBS Lett* 2000, **482**(1-2):19-24.
 114. Fujibe T, Saji H, Arakawa K, Yabe N, Takeuchi Y, Yamamoto KT: **A methyl viologen-resistant mutant of Arabidopsis, which is allelic to ozone-sensitive rcd1, is tolerant to supplemental ultraviolet-B irradiation.** *Plant Physiol* 2004, **134**(1):275-285.
 115. Overmyer K, Brosche M, Pellinen R, Kuitinen T, Tuominen H, Ahlfors R, Keinänen M, Saarna M, Scheel D, Kangasjarvi J: **Ozone-induced programmed cell death in the Arabidopsis radical-induced cell death1 mutant.** *Plant Physiol* 2005, **137**(3):1092-1104.
 116. Kangasjarvi S, Lepisto A, Hannikainen K, Piippo M, Luomala EM, Aro EM, Rintamaki E: **Diverse roles for chloroplast stromal and thylakoid-bound ascorbate peroxidases in plant stress responses.** *Biochem J* 2008, **412**(2):275-285.
 117. Goenka S, Cho SH, Boothby M: **Collaborator of Stat6 (CoaSt6)-associated poly(ADP-ribose) polymerase activity modulates Stat6-dependent gene transcription.** *J Biol Chem* 2007, **282**(26):18732-18739.
 118. Goenka S, Boothby M: **Selective potentiation of Stat-dependent gene expression by collaborator of Stat6 (CoaSt6), a transcriptional cofactor.** *Proc Natl Acad Sci USA* 2006, **103**(11):4210-4215.
 119. Muller S, Moller P, Bick MJ, Wurr S, Becker S, Gunther S, Kummerer BM: **Inhibition of filovirus replication by the zinc finger antiviral protein.** *J Virol* 2007, **81**(5):2391-2400.
 120. Zhang Y, Burke CW, Ryman KD, Klimstra WB: **Identification and characterization of interferon-induced proteins that inhibit alphavirus replication.** *J Virol* 2007, **81**(20):11246-11255.
 121. Guo X, Ma J, Sun J, Gao G: **The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA.** *Proc Natl Acad Sci USA* 2007, **104**(1):151-156.
 122. MacDonald MR, Machlin ES, Albin OR, Levy DE: **The zinc finger antiviral protein acts synergistically with an interferon-induced factor for maximal activity against alphaviruses.** *J Virol* 2007, **81**(24):13509-13518.
 123. Kerns JA, Emerman M, Malik HS: **Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein.** *PLoS Genet* 2008, **4**(1):e21.
 124. Zhu Y, Gao G: **ZAP-mediated mRNA degradation.** *RNA Biol* 2008, **5**(2):65-67.
 125. Monz D, Munnia A, Comtesse N, Fischer U, Steudel WI, Feiden W, Glass B, Meese EU: **Novel tankyrase-related gene detected with meningioma-specific sera.** *Clin Cancer Res* 2001, **7**(11):113-119.
 126. Smith S, de Lange T: **Cell cycle dependent localization of the telomeric PARP, tankyrase, to nuclear pore complexes and centrosomes.** *J Cell Sci* 1999, **112**(Pt 21):3649-3656.
 127. Chi NW, Lodish HF: **Tankyrase is a golgi-associated mitogen-activated protein kinase substrate that interacts with IRAP in GLUT4 vesicles.** *J Biol Chem* 2000, **275**(49):38437-38444.
 128. Yeh TY, Meyer TN, Schwesinger C, Tsun ZY, Lee RM, Chi NW: **Tankyrase recruitment to the lateral membrane in polarized epithelial cells: regulation by cell-cell contact and protein poly(ADP-ribosylation).** *Biochem J* 2006, **399**(3):415-425.
 129. Hsiao SJ, Smith S: **Sister telomeres rendered dysfunctional by persistent cohesion are fused by NHEJ.** *J Cell Biol* 2009, **184**(4):515-526.
 130. Huang SM, Mishina YM, Liu S, Cheung A, Stegmeier F, Michaud GA, Charlat O, Wiellette E, Zhang Y, Wiessner S, et al: **Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling.** *Nature* 2009, **461**(7264):614-620.
 131. Karlberg T, Markova N, Johansson I, Hammarstrom M, Schutz P, Weigelt J, Schuler H: **Structural basis for the interaction between tankyrase-2 and a potent Wnt-signaling inhibitor.** *J Med Chem* 2010, **53**(14):5352-5355.
 132. Karner CM, Merkel CE, Dodge M, Ma Z, Lu J, Chen C, Lum L, Carroll TJ: **Tankyrase is necessary for canonical Wnt signaling during kidney development.** *Dev Dyn* 2010, **239**(7):2014-2023.
 133. Muramatsu Y, Ohishi T, Sakamoto M, Tsuruo T, Seimiya H: **Cross-species difference in telomeric function of tankyrase 1.** *Cancer Sci* 2007, **98**(6):850-857.
 134. Chiang YJ, Hsiao SJ, Yver D, Cushman SW, Tessarollo L, Smith S, Hodes RJ: **Tankyrase 1 and tankyrase 2 are essential but redundant for mouse embryonic development.** *PLoS One* 2008, **3**(7):e2639.
 135. Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen FM: **Transposons in place of telomeric repeats at a Drosophila telomere.** *Cell* 1993, **75**(6):1083-1093.
 136. Dequen F, Gagnon SN, Desnoyers S: **Ionizing radiations in Caenorhabditis elegans induce poly(ADP-ribosylation), a conserved DNA-damage response essential for survival.** *DNA Repair (Amst)* 2005, **4**(7):814-825.
 137. Chau V, Tobias JW, Bachmair A, Marriott D, Ecker DJ, Gonda DK, Varshavsky A: **A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein.** *Science* 1989, **243**(4898):1576-1583.
 138. Thrower JS, Hoffman L, Rechsteiner M, Pickart CM: **Recognition of the polyubiquitin proteolytic signal.** *EMBO J* 2000, **19**(1):94-102.
 139. Hicke L, Schubert HL, Hill CP: **Ubiquitin-binding domains.** *Nat Rev Mol Cell Biol* 2005, **6**(8):610-621.
 140. Grabbe C, Dikic I: **Functional roles of ubiquitin-like domain (ULD) and ubiquitin-binding domain (UBD) containing proteins.** *Chem Rev* 2009, **109**(4):1481-1494.
 141. Yan J, Kim YS, Yang XP, Li LP, Liao G, Xia F, Jetten AM: **The ubiquitin-interacting motif containing protein RAP80 interacts with BRCA1 and functions in DNA damage repair response.** *Cancer Res* 2007, **67**(14):6647-6656.
 142. Wu W, Koike A, Takeshita T, Ohta T: **The ubiquitin E3 ligase activity of BRCA1 and its biological functions.** *Cell Div* 2008, **3**:1.
 143. Feng L, Huang J, Chen J: **MERIT40 facilitates BRCA1 localization and DNA damage repair.** *Genes Dev* 2009, **23**(6):719-728.
 144. Winn PJ, Religa TL, Battey JN, Banerjee A, Wade RC: **Determinants of functionality in the ubiquitin conjugating enzyme family.** *Structure* 2004, **12**(9):1563-1574.

145. Sato Y, Yoshikawa A, Mimura H, Yamashita M, Yamagata A, Fukai S: **Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80.** *EMBO J* 2009, **28**(16):2461-2468.
146. Yanagawa T, Funasaka T, Tsutsumi S, Hu H, Watanabe H, Raz A: **Regulation of phosphoglucose isomerase/autocrine motility factor activities by the poly(ADP-ribose) polymerase family-14.** *Cancer Res* 2007, **67**(18):8682-8689.
147. Wang T, Simbulan-Rosenthal CM, Smulson ME, Chock PB, Yang DC: **Polyubiquitylation of PARP-1 through ubiquitin K48 is modulated by activated DNA, NAD+, and dipeptides.** *J Cell Biochem* 2008, **104**(1):318-328.
148. Masson M, Menissier-de Murcia J, Mattei MG, de Murcia G, Niedergang CP: **Poly(ADP-ribose) polymerase interacts with a novel human ubiquitin conjugating enzyme: hUbc9.** *Gene* 1997, **190**(2):287-296.
149. Chang W, Dynek JN, Smith S: **TRF1 is degraded by ubiquitin-mediated proteolysis after release from telomeres.** *Genes Dev* 2003, **17**(11):1328-1333.
150. Messner S, Schuermann D, Altmeyer M, Kassner I, Schmidt D, Schar P, Muller S, Hottiger MO: **Sumoylation of poly(ADP-ribose) polymerase 1 inhibits its acetylation and restrains transcriptional coactivator function.** *FASEB J* 2009, **23**(11):3978-3989.
151. Martin N, Schwamborn K, Schreiber V, Werner A, Guillier C, Zhang XD, Bischof O, Seeler JS, Dejean A: **PARP-1 transcriptional activity is regulated by sumoylation upon heat shock.** *Embo J* 2009, **28**(22):3534-3548.
152. Stilmann M, Hinz M, Arslan SC, Zimmer A, Schreiber V, Scheiderei C: **A nuclear poly(ADP-ribose)-dependent signalosome confers DNA damage-induced IkkappaB kinase activation.** *Mol Cell* 2009, **36**(3):365-378.
153. Haynes PA, Miller S, Radabaugh T, Galligan M, Brezi L, Rohrbough J, Hickman F, Merchant N: **The wildcat toolbox: a set of perl script utilities for use in peptide mass spectral database searching and proteomics experiments.** *J Biomol Tech* 2006, **17**(2):97-102.
154. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**(5):696-704.
155. Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**(9):2104-2105.
156. Anisimova M, Gascuel O: **Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative.** *Syst Biol* 2006, **55**(4):539-552.
157. Beitz E: **TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon.** *Bioinformatics* 2000, **16**(2):135-139.
158. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ: **Jalview Version 2—a multiple sequence alignment editor and analysis workbench.** *Bioinformatics* 2009, **25**(9):1189-1191.
159. Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R, et al: **A higher-level phylogenetic classification of the Fungi.** *Mycol Res* 2007, **111**(Pt 5):509-547.
160. Hibbett DS: **A phylogenetic overview of the Agaricomycotina.** *Mycologia* 2006, **98**(6):917-925.
161. Spatafora JW, Sung GH, Johnson D, Hesse C, O'Rourke B, Serdani M, Spotts R, Lutzoni F, Hofstetter V, Miadlikowska J, et al: **A five-gene phylogeny of Pezizomycotina.** *Mycologia* 2006, **98**(6):1018-1028.
162. Palenik B, Grimwood J, Aerts A, Rouze P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S, et al: **The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation.** *Proc Natl Acad Sci USA* 2007, **104**(18):7705-7710.
163. Nozaki H, Takano H, Misumi O, Terasawa K, Matsuzaki M, Maruyama S, Nishida K, Yagisawa F, Yoshida Y, Fujiwara T, Takio S, Tamura K, Chung SJ, Nakamura S, Kuroiwa H, Tanaka K, Sato N, Kuroiwa T: **A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*.** *BMC Biol* 2007, **5**:28.
164. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG: **Life with 6000 genes.** *Science* 1996, **274**(5287):546, 563-547.
165. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, et al: **The genome sequence of *Schizosaccharomyces pombe*.** *Nature* 2002, **415**(6874):871-880.
166. Jeffries TW, Grigoriev IV, Grimwood J, Laplaza JM, Aerts A, Salamov A, Schmutz J, Lindquist E, Dehal P, Shapiro H, Jin YS, Passoth V, Richardson PM: **Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*.** *Nat Biotechnol* 2007, **25**(3):319-326.
167. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, Davis RW, Scherer S: **The diploid genome sequence of *Candida albicans*.** *Proc Natl Acad Sci USA* 2004, **101**(19):7329-7334.
168. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otilar RP, et al: **The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes.** *Nature* 2008, **456**(7219):239-244.
169. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, et al: **The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism.** *Science* 2004, **306**(5693):79-86.
170. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, et al: **Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*.** *Science* 2007, **317**(5846):1921-1926.
171. Dutta S, Burkhardt K, Young J, Swaminathan GJ, Matsuura T, Henrick K, Nakamura H, Berman HM: **Data deposition and annotation at the worldwide protein data bank.** *Mol Biotechnol* 2009, **42**(1):1-13.
172. Lehtio L, Collins R, van den Berg S, Johansson A, Dahlgren LG, Hammarstrom M, Helleday T, Holmberg-Schiavone L, Karlberg T, Weigelt J: **Zinc binding catalytic domain of human tankyrase 1.** *J Mol Biol* 2008, **379**(1):136-145.

doi:10.1186/1471-2148-10-308

Cite this article as: Citarelli et al.: Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes. *BMC Evolutionary Biology* 2010 **10**:308.

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

