

Research article

Open Access

## ***In silico* characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs)**

Helge Otto<sup>1</sup>, Pedro A Reche<sup>2,3</sup>, Fernando Bazan<sup>2,4</sup>, Katharina Dittmar<sup>5</sup>, Friedrich Haag<sup>1</sup> and Friedrich Koch-Nolte\*<sup>1</sup>

Address: <sup>1</sup>Institute of Immunology, University Hospital Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany., <sup>2</sup>DNAX Research Institute, Palo Alto, CA 94304, USA., <sup>3</sup>Dana-Farber Cancer Institute, Harvard University, Boston, MA 02115, USA., <sup>4</sup>Depts. of Molecular Biology and Protein Engineering, Genentech, SF, CA 94080, USA. and <sup>5</sup>Department of Integrative Biology, Brigham Young University, Provo, UT 84602, USA.

Email: Helge Otto - helge.otto@t-online.de; Pedro A Reche - reche@research.dfci.harvard.edu; Fernando Bazan - bazan.fernando@gene.com; Katharina Dittmar - katharinad@gmail.com; Friedrich Haag - haag@uke.uni-hamburg.de; Friedrich Koch-Nolte\* - nolte@uke.uni-hamburg.de

\* Corresponding author

Published: 04 October 2005

Received: 13 May 2005

BMC Genomics 2005, 6:139 doi:10.1186/1471-2164-6-139

Accepted: 04 October 2005

This article is available from: <http://www.biomedcentral.com/1471-2164/6/139>

© 2005 Otto et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### **Abstract**

**Background:** ADP-ribosylation is an enzyme-catalyzed posttranslational protein modification in which mono(ADP-ribosyl)transferases (mARTs) and poly(ADP-ribosyl)transferases (pARTs) transfer the ADP-ribose moiety from NAD onto specific amino acid side chains and/or ADP-ribose units on target proteins.

**Results:** Using a combination of database search tools we identified the genes encoding recognizable pART domains in the public genome databases. In humans, the pART family encompasses 17 members. For 16 of these genes, an orthologue exists also in the mouse, rat, and pufferfish. Based on the degree of amino acid sequence similarity in the catalytic domain, conserved intron positions, and fused protein domains, pARTs can be divided into five major subgroups. All six members of groups 1 and 2 contain the H-Y-E trias of amino acid residues found also in the active sites of Diphtheria toxin and Pseudomonas exotoxin A, while the eleven members of groups 3 – 5 carry variations of this motif. The pART catalytic domain is found associated in Lego-like fashion with a variety of domains, including nucleic acid-binding, protein-protein interaction, and ubiquitylation domains. Some of these domain associations appear to be very ancient since they are observed also in insects, fungi, amoebae, and plants. The recently completed genome of the pufferfish *T. nigroviridis* contains recognizable orthologues for all pARTs except for pART7. The nearly completed albeit still fragmentary chicken genome contains recognizable orthologues for twelve pARTs. Simpler eucaryotes generally contain fewer pARTs: two in the fly *D. melanogaster*, three each in the mosquito *A. gambiae*, the nematode *C. elegans*, and the ascomycete microfungus *G. zaeae*, six in the amoeba *E. histolytica*, nine in the slime mold *D. discoideum*, and ten in the cress plant *A. thaliana*. GenBank contains two pART homologues from the large double stranded DNA viruses Chilo iridescent virus and Bacteriophage AehI and only a single entry (from *V. cholerae*) showing recognizable homology to the pART-like catalytic domains of Diphtheria toxin and Pseudomonas exotoxin A.

**Conclusion:** The pART family, which encompasses 17 members in the human and 16 members in the mouse, can be divided into five subgroups on the basis of sequence similarity, phylogeny, conserved intron positions, and patterns of genetically fused protein domains.

## Background

ADP-ribosylation is a posttranslational protein modification in which the ADP-ribose moiety is transferred from NAD onto specific amino acid side chains of target proteins [1-4]. ADP-ribosylation was originally discovered as the pathogenic principle of Diphtheria toxin, a multidomain secreted protein which inactivates elongation factor 2 by ADP-ribosylation after translocation into eucaryotic cells [5]. Subsequently, numerous other bacterial toxins were shown to ADP-ribosylate target proteins in host cells. Moreover, endogenous toxin-like ADP-ribosylating enzyme activities were detected in eucaryotic cells. Several of these enzymes were purified to homogeneity, sequenced, expressed as recombinant proteins, and crystallized.

Sequence and structural analyses revealed the existence of two distinct families of toxin-related ADP-ribosyltransferases in mammals [6,7]: The RT6 family of GPI-anchored and secretory mono-(ADP-ribosyl)transferases (mARTs) catalyzes mono-ADP-ribosylation of cell surface and secretory proteins [8]. The PARP family of nuclear and cytoplasmic poly(ADP-ribosyl)transferases (pARTs) catalyzes poly-ADP-ribosylation of nuclear and cytosolic proteins [9-12]. While mARTs have been implicated to mediate signalling functions of extracellular NAD, pARTs have been shown to play important roles in DNA repair and maintenance of genome integrity [8,9,12].

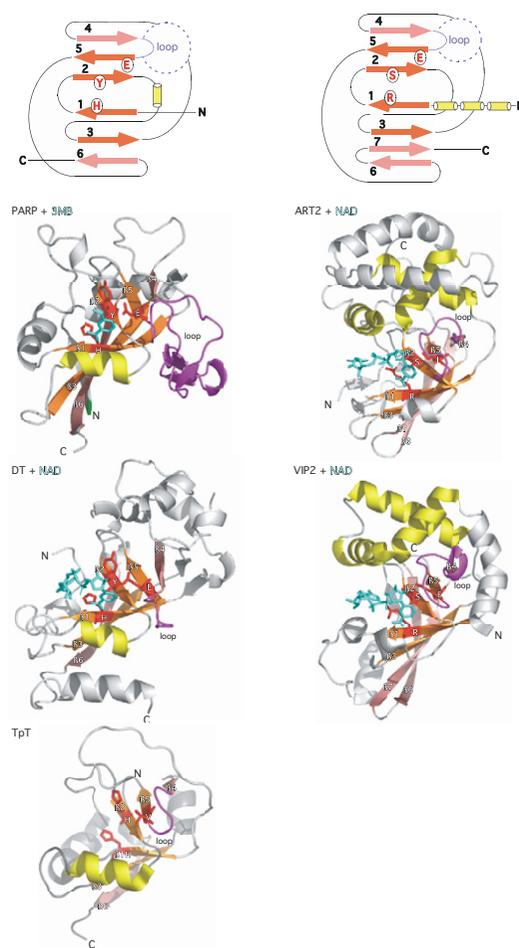
In this paper we use the term pART (poly ADP-ribosyltransferase) rather than the more established term PARP (poly-ADP-ribosyl-polymerase) for various reasons. Firstly, to emphasize the structural and functional similarities of the poly- and mono-ADP-ribosyltransferase subfamilies. Secondly, with respect to the biochemical classification of enzymes the term transferase is more appropriate than polymerase: ADP-ribosyltransferases belong to the family of glycosyltransferases; the term polymerase is more commonly used for template-dependent DNA or RNA synthesizing enzymes. Thirdly, use of the term PARP would have confounded comparison of our results with those of the recent review by Ame *et al.* [11], who used the term PARP and a numbering system without regard to structural similarities among gene family members.

The 3D-structures of rat ART.2 (PDB accession number 1og3), chicken PARP-1 (1a26, 3pax), mouse PARP-2 (1gs0), and numerous ADP-ribosylating toxins uncovered a common NAD binding fold with a conserved core of five  $\beta$  strands arranged in two abutting  $\beta$  sheets [13-19]. These two  $\beta$  sheets form the upper and lower jaws of a Pacman-like active site crevice (Figure 1). Remarkably, only a single amino acid residue, the catalytic glutamic acid residue at the front edge of the fifth conserved  $\beta$ -strand, is strictly

conserved in all known 3D structures of enzymatically active mARTs and pARTs. In a seminal study, Collier and co-workers pinpointed the corresponding glutamic acid residue in PARP-1 (before its 3D structure was solved) on the basis of barely detectable sequence similarity to Diphtheria toxin [20,21]. More recently, the 3D structures of anthrax lethal factor, VIP2, and iota toxin have been discovered to harbour ART-domains that lack a corresponding glutamic acid residue and may represent inactivated enzymes [16,22,23].

Comparative structure and amino acid sequence analyses revealed that PARP-1 and PARP-2 share additional secondary structure and conserved amino acids with Diphtheria toxin and Pseudomonas exotoxin A, which evidently are not conserved in other mARTs (Fig. 1) [6,7]. These additional elements include a sixth  $\beta$  strand, an alpha helix between  $\beta$  strands 2 and 3, and a triad of amino acids, the so-called H-Y-E motif, encompassing a histidine residue in  $\beta$  strand 1, a tyrosine residue in  $\beta$  strand 3 and the catalytic glutamic acid residue at the front edge of  $\beta$  strand 5. These features, highlighted in the 3D structures of PARP-1 and Diphtheria toxin in Figure 1, clearly distinguish the structures of PARP-1, PARP-2, and DT/ETA from those of a second major ART subfamily that includes rat ART2 and the *Bacillus cereus* VIP2 toxin. Distinguishing features of the ART2/VIP2 subfamily include a seventh  $\beta$  strand that displaces  $\beta$  strand 6, three conserved alpha helices preceding  $\beta$  strand 1, and an R-S-E triad of amino acid residues in place of the H-Y-E motif of PARP-1 and DT. Interestingly, the recently reported 3D-structure of a prototype member of the family of tRNA:NAD 2' phosphotransferases (TpT) [24] revealed a striking resemblance to the structures of the PARP-1/DT subfamily rather than to those of the ART2/VIP subfamily, including the sixth  $\beta$  strand, the alpha helix between  $\beta$  strands 2 and 3, and a variant H-Y-E motif (H-H-V). These enzymes catalyze removal of a splice junction 2' phosphate from ligated tRNA. This reaction resembles the reaction catalyzed by ARTs but yields ADP-ribose 1"-2" cyclic phosphate rather than ADP-ribosylated proteins [25].

The remarkable degree of plasticity of ART amino acid sequences poses a challenging problem for genome data base mining [7] and even the most sensitive database search programs fail to connect all known members of the ART gene family. Notwithstanding, the results of such *in silico* analyses can provide important insight into the structural and phylogenetic relationship of ART subfamilies. We have previously demonstrated that the known members of the mART gene family in the human and mouse could be faithfully connected with many known bacterial ADP-ribosylating toxins, but not with pARTs or Diphtheria toxin [26,27]. These analyses also pointed out the presence of mART-encoding genes in the

**Figure 1****Schematic illustration of the distinguishing structural features of the PARP-I/DT vs. the ART2/VIP2 sub-families of ADP-ribosyltransferases.**

Two abutting sheets of anti-parallel  $\beta$  strands form the upper and lower jaws of a Pacman-like NAD-binding crevice in all known structures of ADP-ribosyltransferases. The distinguishing structural features of the PARP/DT and ART2/VIP2 subfamilies are depicted schematically on top and are highlighted in the structures of chicken PARP-I (3pax), diphtheria toxin (DT) (1tox), an archaeal tRNA:NAD 2'-phosphotransferase (TpT) (1wfx), rat ART2 (1og3) and *B. cereus* VIP2 toxin (1qs2) below. The structures are depicted from the "front view" with a full view of the ligands bound in the active site crevice. The ligands NAD and 3MB are colored cyan and are depicted as stick models. The central four  $\beta$ -strands (from top to bottom:  $\beta$  5,  $\beta$  2,  $\beta$  1,  $\beta$  3, colored orange) are conserved in all mARTs and pARTs. The  $\beta$  strands at the edges of the respective sheets ( $\beta$  4 and  $\beta$  6, colored pink) show greater structural variation than the central  $\beta$  strands. The H-Y-E motif residues are depicted in red and their side chains are shown as sticks. The glutamic acid residue at the front edge of  $\beta$  5 is the critical catalytic residue in both diphtheria toxin and PARP-I – a corresponding glutamic acid residue is observed also in the 3D structures of rat ART2 and numerous bacterial mARTs. Diphtheria toxin (1tox), pseudomonas exotoxin A (1aer), PARP-I (3pax), and PARP-2 (1gs0) share the following structural features which are not conserved in either rat ART2 (1og3) or most other bacterial mARTs: the orientation of  $\beta$  6, the alpha helix between  $\beta$  2 and  $\beta$  3 (colored yellow) and the conserved histidine and tyrosine amino acid residues in  $\beta$  1 and  $\beta$  3. The loop between  $\beta$  4 and  $\beta$  5 (colored magenta) is thought to play a role in the recognition of target proteins and ADP-ribose polymers. Distinguishing features of ART2, VIP2, iota toxin (1gir), and the C3 exoenzymes (1g24, 1ojz) include three conserved alpha helices upstream of  $\beta$  strand 1, a seventh  $\beta$  strand that displaces  $\beta$  strand 6 and an R-S-E- motif instead of the H-Y-E motif of PARP-I and DT. (Note that the depicted ART2 structure carries a site directed mutation of the catalytic glutamic acid residue E189I). The recently determined 3D structure of the tRNA:NAD 2'-phosphotransferase (1wfx) bears striking resemblance to that of DT and PARP-I and carries an H-H-V variant of the H-Y-E motif. Note that the structure of the diphtheria toxin catalytic domain shown here in complex with NAD is truncated C-terminally at the proteolytic cleavage site that separates this domain from the translocation domain. The PARP-I catalytic domain shown here is truncated N-terminally at the position of the phase 0 intron that separates this domain from a neighboring helical domain. The TpT catalytic domain is truncated N-terminally at the point of fusion to a winged-helix domain.

genomes of many but not all other model organisms. Of note, no mART-encoding genes could be detected in plants, fungi, or archaea. Here we provide an in depth analysis of the pART gene family.

## Results and discussion

### Identification of human and mouse pART family members in the EST database

The human and mouse pART gene family members were identified using a combination of data base search tools. The human and mouse EST databases as well as the non-redundant GenBank database (nr) were screened with tBLASTn using as queries the amino acid sequences of the catalytic domains of the known and newly identified pART family members. Whenever possible, the full coding sequence of the catalytic domain and of the adjacent regions was assembled using the sequences of published cDNAs and overlapping ESTs. Screening of the EST and nr databases was initiated in 1997 and was repeated in regular intervals. The coding sequences were extended when suitable new sequences became available. When the sequences of the human, mouse and rat genomes were published in 2000, 2001, and 2004, respectively, the EST database searches were complemented with corresponding tBLASTn and BLASTn searches of the genome sequences [28-30]. Thereby, 17 pART family members were identified in the human. These genes were designated pART1-pART17. Numbering reflects the degree of amino acid sequence similarity to PARP-1 (= pART1) and the degree of similarity within each of the pART subgroups. An orthologue for each of these genes was detected in the mouse and in the rat, with the sole exception of pART7.

A complete list of human pART family members, including the common names and aliases of known genes, is presented in Figure 2. Based on the degree of amino acid sequence similarities, conserved intron positions, and fused protein domains, the mammalian pART family can be divided into five major subgroups. Group 1 (pART1-pART4) contains PARP and its closest relatives, PARP-2, PARP-3 and VPARP. Group 2 (pART5, pART6) contains tankyrase 1 and tankyrase 2. Group 3 (pART7-pART10) contains four proteins including the recently described B-Aggressive Lymphoma Protein (BAL = pART9) [31] and a myc-interacting protein with PARP activity (PARP-10) [32]. Group 4 (pART11-pART14) contains four proteins including the recently described Zinc-finger Antiviral Protein (ZAP = pART13) [33] and TCDD-inducible PARP (TiPARP) [34]. Group 5 (pART15-pART17) contains three proteins of unknown function.

The steady growth in the number of matching ESTs obtained for each of the human pART gene family members over the past 6 years is illustrated in additional file 1

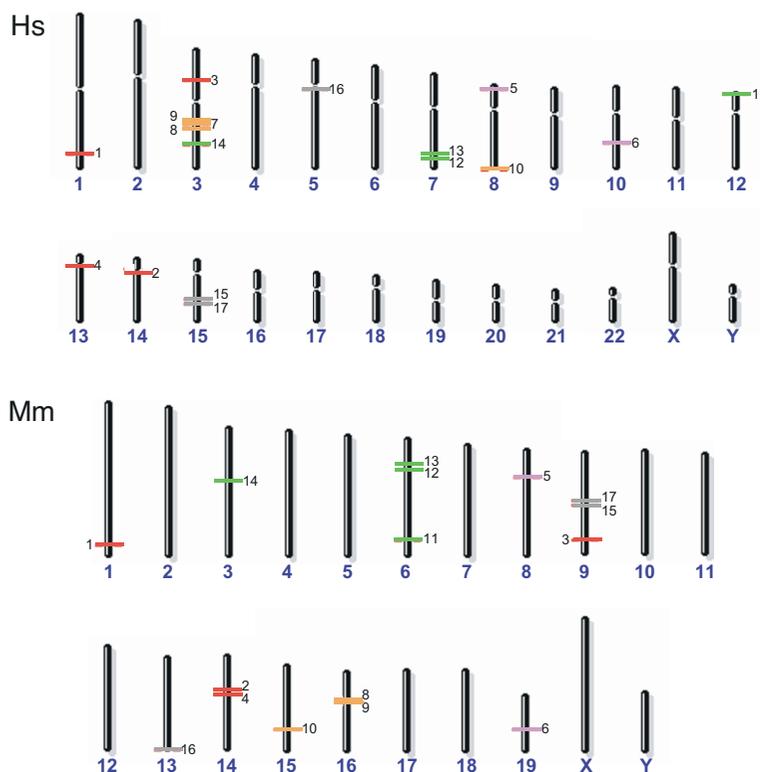
("Representation of pART gene transcripts in the database of expressed sequence tags"). By October 2004, each human pART except pART7 was represented by more than 100 ESTs. Interestingly, each pART except pART7 is represented by more ESTs than poly (ADP-ribose) glycohydrolase (PARG), the single known enzyme capable of removing poly-ADP-ribose from pART target proteins. The large number of ESTs corresponds to a large variety of tissues found to contain pART ESTs and presumably reflects an ubiquitous pattern of gene expression, i.e. akin to that of the house keeping enzymes hypoxanthine-guanine phosphoribosyltransferase (HPRT) and glyceraldehyde-3-phosphate dehydrogenase (GAPD). For comparison, the members of the mART gene family (ART1-ART5), which exhibit highly restricted patterns of expression, are each represented by much fewer ESTs than the pARTs. As of January 2005, the mammalian gene collection <http://mgc.nci.nih.gov> contains annotated full-length cDNA sequences for 10 of the 17 human pARTs and for 12 of 16 mouse pARTs (Fig. 2).

### Chromosomal localizations and exon/intron structures of the human and mouse pART gene family members

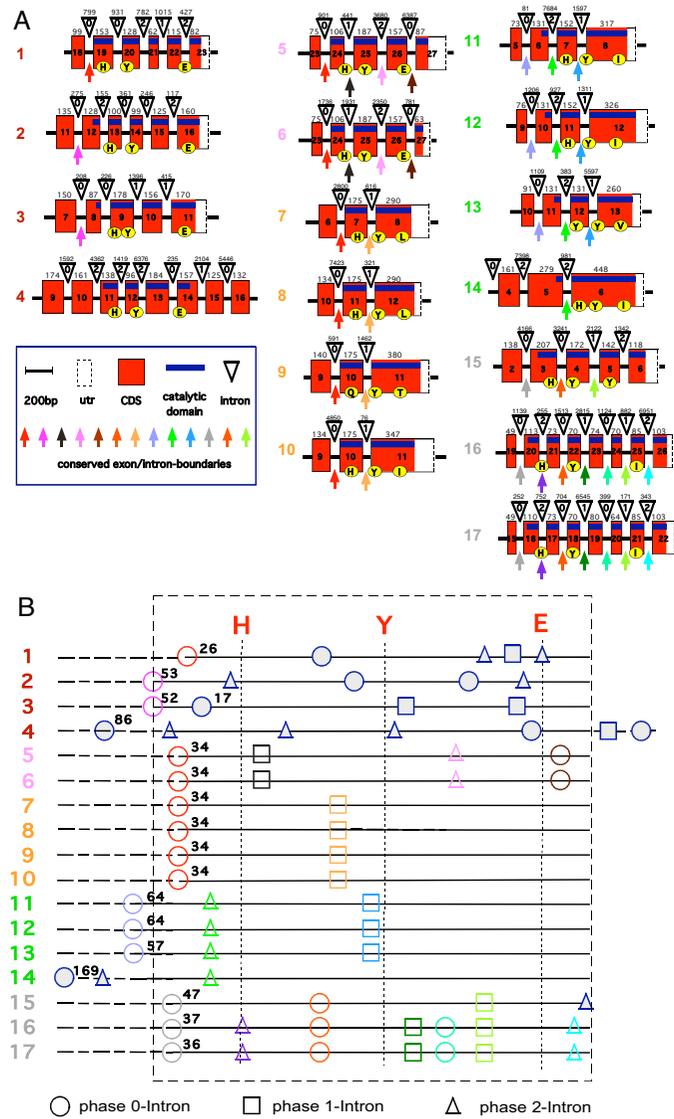
The results of tBLASTn and BLASTn searches of the human, mouse, and rat genome sequences yielded the chromosomal localization and the exon/intron structure of each pART gene family member. The chromosomal localizations of the pART genes are represented schematically in Figure 2. All human and mouse pART orthologues lie in regions of conserved synteny. There are three conserved pART gene clusters containing two related paralogues (pARTs 8 and 9; pARTs 12 and 13; pARTs 15 and 17). However, the two most closely related pairs of pARTs (pARTs 5 and 6; pARTs 16 and 17) each are located on different chromosomes. All other pARTs are distributed as single copy genes on different autosomes. In the human genome, the cluster containing pARTs 8 and 9 also contains pART7. Additional file 2 illustrates the local chromosomal environment of this pART gene cluster on human chromosome 3q and the syntenic region on mouse chromosome 16B3. The local order of genes is similar in the human and mouse. However, the region corresponding to pART7 is missing in the mouse. The corresponding region is also missing in the rat genome (not shown).

The total number of exons in each pART gene is depicted in Figure 2 and the exon structure of the catalytic domain is illustrated schematically for the human pARTs in Figure 3. All intron positions within the coding region are fully conserved in human and mouse orthologues. With the sole exception of pART4 (VPARP), the catalytic domain is encoded by the 3' terminal exons. Remarkably, in all pART genes, with the exception of pART4 (VPARP) and pART14 (TiPARP), the exons encoding the catalytic domain are separated from the rest of the respective

pART	gene symbol	aliases	Hs protein accession #	chrom. localization		# of exons		amino acids		MGC accession #	
				Hs	Mm	Hs	Mm	Hs	Mm	Hs	Mm
1	<i>PARP1</i>	PARP	P09874	1q41-q42	1 H5	23	23	1014	1014	BC037545	BC012041
2	<i>PARP2</i>	PARP-2	CAB41505	14q11.2-q12	14 C1	16	16	583	559	na	BC062150
3	<i>PARP3</i>	PARP-3	AAM95460	3p21.1-22.2	9 F1	11	11	540	528	(BC014260)	BC014870
4	<i>PARP4</i>	vaultPARP	AAD47250	13q11	14 C1	34	> 28	1724	> 1446	na	na
5	<i>TNKS</i>	Tankyrase	AAC79841	8p23.1	8 A4	27	27	1327	1320	na	BC057370
6	<i>TNKS2</i>	Tankyrase2	NP_079511	10q23.3	19 C2	27	28	1166	1337	na	na
7	<i>PARP15</i>		NP_689828	3q21.1	---	8	---	444	---	na	---
8	<i>PARP14</i>		AAN08627	3q21.1	16 B3	12	12	1518	1535	na	(BC021340)
9	<i>PARP9</i>	BAL	NP_113646	3q13.3-q21	16 B3	11	11	854	830	(BC039580)	BC003281
10	<i>PARP10</i>	PARP-10	BAB55067	8q24.3	15 D3	11	11	1025	960	na	na
11	<i>PARP11</i>		AAF91391	12p13.3	6 F3	8	9	331	331	BC017569	BC040269
12	<i>ZC3HDC1</i>		NP_073587	7q34	6 B1	12	12	701	711	BC081541	na
13	<i>ZC3HAV1</i>	ZAP	NP_064504	7q34	6 B1	13	> 11	902	996	(BC025308)	(BC029090)
14	<i>TIPARP</i>	TiPARP	NP_056323	3q25.31	3 E1	6	6	657	657	BC050350	BC068173
15	<i>PARP16</i>		AAH31074	15q22.2	9 C	6	7	322	322	BC006389	BC055447
16	<i>PARP8</i>		NP_078891	5q11.2	13 D2.3	26	26	854	852	(BC075801)	(BC021315)
17	<i>PARP6</i>		CAB59261	15q22.23	9 C	22	22	630	630	(BC026955)	BC062096



**Figure 2**  
**Chromosomal localizations and exon compositions of the human and mouse pART family members.** A) pART family members are sorted by subgroup on the basis of similarities in amino acid sequence, intron positions and associated protein domains. Color-coding of subgroups is as follows: 1 = red, 2 = pink, 3 = orange, 4 = green, 5 = grey. This color-coding is used in subsequent figures. Official gene designations, common aliases and accession numbers are shown. Exon compositions and lengths of open reading frames are given for the longest known or predicted gene transcripts. Available full length cDNAs from the Mammalian Gene Collection (MGC) are indicated with their respective accession numbers. MGC cDNAs which apparently do not contain the full open reading frame are indicated in parentheses. Hs = *Homo sapiens*, Mm = *Mus musculus*. B) Chromosomal localizations of pART genes were determined by tBLASTn searches of the respective genome sequences using the amino acid sequences of the catalytic domains of individual pARTs. Members of the five pART family subgroups are color-coded as in A).



**Figure 3**  
**Schematic diagram of the exon/intron structures of the regions encoding the catalytic domain of pART family members.** A) Exon/intron structures were determined by BLASTn searches of the human genome sequence with individual pART cDNA sequences. Only the exons corresponding to the catalytic domain of PARP-1 are shown. The coding region is marked in red, the 3' untranslated region (utr) is marked in white, and a blue bar marks the region corresponding to the catalytic domain. Exons are represented as boxes with the width of each box reflecting the size of the respective exon (the 3' utr is not drawn to scale). Exon numbers are given with exon 1 corresponding to the exon encoding the presumptive initiation methionine. In all cases except pART4 (VPARP) the catalytic domain is encoded by the 3' terminal exons. Exon sizes (or size of coding region in case of the 3' exons) in basepairs are indicated on top of the boxes. Introns are depicted as triangles and are not drawn to scale. Intron sizes in base pairs are indicated on top of the triangles. The position of each intron with respect to the reading frame is indicated in the triangles (0 = between codons, +1 = between codon positions 1 and 2, +2 = between codon positions 2 and 3). Conserved exon boundaries are marked by colored arrows. Codons corresponding to the H-Y-E motif in the NAD binding crevice of DT and PARP-1 (see Fig. 1) are marked by yellow circles. B) The catalytic domain as delineated in this paper is indicated by the dashed rectangle. For each pART the cDNA coding region within the catalytic domain is marked by a straight line, regions extending beyond this domain in the 5' direction (and in the 3' direction in case of pART4) are marked by dashed lines. The positions of the codons corresponding to the H, Y, E residues in the NAD-binding crevice are indicated by vertical lines. Intron phases are indicated by circles (phase 0), boxes (phase 1), and triangles (phase 2). Numbers indicate the distance in codons between the conserved histidine in  $\beta$  1 and the next upstream phase 0 intron. Color-coding of conserved introns corresponds to that shown in A). Nonconserved introns are indicated in blue (filled) icons.

coding exons by a phase 0 intron shortly upstream of the codon for the first residue of the conserved H-Y-E catalytic site motif, the conserved histidine in  $\beta$  1 (Fig. 3). For most pARTs, the amino acid sequences encoded by exons upstream of this phase 0 intron do not show any detectable similarities, except for members of a particular subgroup. We used the position of this phase 0 intron in pART1 to delineate the N-terminal border of the catalytic domain (e.g., see the green labeled end of the PARP-1-model in Figure 1 and the dashed rectangle in Figure 3B).

The exon/intron structures of the pART catalytic domains reveal a number of intriguing features (Fig. 3). The region encoding the catalytic domain is disrupted by a remarkable variety of introns with the number of introns varying from one in subgroup 3 and in pART14 to six in pARTs 16 and 17. The catalytic domain of pART1 (PARP-1) and those of its closest relatives in subgroup 1 are disrupted by three (pARTs 3 and 4) or four (pARTs 1 and 2) introns. Strikingly, not one of these 14 intron positions is conserved. The catalytic domains of the two closely related tankyrases in subgroup 2 each are interrupted by three conserved introns. In subgroup 3, the catalytic domains of pARTs 7–10 each contain a single conserved intron. The pARTs of subgroup 4 (pARTs 11–14) share a single conserved intron in their catalytic domains, pARTs 11–13 share a second conserved intron in the catalytic domain, which is missing in pART14. The pARTs of subgroup 5 (pARTs 15–17) share two conserved introns in their catalytic domains, pARTs 16 and 17 share four additional conserved introns in the catalytic domain, which are missing in pART15.

#### **Conserved structural features revealed by multiple amino acid sequence alignments and secondary structure predictions**

PSI-BLAST is a powerful, position sensitive iterative program designed to detect distantly related proteins in the protein database [35]. Initial matches in the first iteration correspond to those detected by classic BLASTp searches and typically reveal proteins with an amino acid sequence identity to the query sequence of > 30%. PSI-BLAST then derives a position specific scoring matrix from the aligned protein sequences obtained in the first iteration, which is then used for the subsequent search of the protein database. This process is repeated in an iterative fashion until no further matches are detected and the search 'converges'. We performed PSI-BLAST searches of the protein database using as query the amino acid sequences of the catalytic domain of each member of the pART gene family. Figure 4 schematically illustrates the tiling paths of PSI-BLAST searches obtained with the stringent default threshold setting (0.005 for the expect value) for a representative member of pART family subgroups 1, 3, 4 and 5. Typically, the other members of the same subgroup were

detected in the first iteration and obtained the highest scores. The pARTs of other subgroups were usually detected within two additional iterations, except in case of pART15. Here, five iterations were required to detect all pART family members.

The amino acid sequence alignments generated by PSI-BLAST typically contained the highest degree of sequence similarity in secondary structure motifs corresponding to the NAD-binding cores in the known 3D structures of chicken PARP-1 (1a26) and mouse PARP-2 (1gs0). Separate multiple amino acid sequence alignments were generated with T-Coffee for each of the pART subgroups using the orthologous sequences from human and mouse [36]. PSIPRED was used to predict secondary structure units and GenTHREADER was used to predict the optimal alignment of pART amino acid sequences with the 3D structures of chicken PARP-1 and mouse PARP-2 [37]. In all cases, predictions and alignments yielded consistent results with respect to the sole alpha helix and five of the six  $\beta$ -strands of the PARP-1 catalytic domain (see additional files 3, 4, 5, 6, 7: "Multiple amino acid sequence alignments, secondary structure predictions and threading results for pART subgroups 1–5"). The small  $\beta$  strand ( $\beta$  4) at the upper edge of the active site crevice was aligned and predicted congruently only for subgroups 1–4, and could not be predicted with confidence for the most distant relatives of PARP-1 (pARTs 15–17). Regions corresponding to connecting loops showed significant sequence identities only for members of a particular pART subgroup. Most likely, these regions fold similarly only in closely related pART family members.

A striking result of the alignment analyses is that the H-Y-E catalytic site motif is fully conserved only in subgroups 1 and 2 (pARTs 1–6). All other pARTs show deviations from this motif. The histidine in  $\beta$  1 is conserved in 9 of the 11 members of subgroup 3–5, the tyrosine in  $\beta$  3 is conserved in all family members, yet the presumptive catalytic glutamic acid at the N-terminal end of  $\beta$  6 is exchanged in each of the pARTs 7–17.

Moreover, the amino acid sequence of the loop immediately upstream of  $\beta$  5 and the active site glutamic acid residue deviates markedly from those of PARP-1 and PARP-2 in most other family members except for the tankyrases (pARTs 5 and 6). A growing body of evidence indicates that this region influences the target specificity of pARTs and mARTs [38–40]. In the 3D structure of PARP-1 with carba-NAD (3pax), the ligand was found to interact with this loop outside of the active site crevice, and it was proposed that this may reflect the binding of the ADP-ribose polymer in the target protein [14].

input	Hs.pART1	Hs.pART9	Hs.pART12	Hs.pART15	Ci.pART
iteration 1 = traditional Blastp searches	Ag.pARTa	Hs.pART8	Hs.pART11	Ag.pARTc	Ag.pARTa
	Dm.pARTa	Hs.pART10	Hs.pART14	Hs.pART17	Hs.pART1
	Hs.pART2	Hs.pART7	Hs.pART13		Hs.pART2
	Ce.pARTa	Hs.pART13	Hs.pART7		Dm.pARTa
	Hs.pART3	Hs.pART14	Hs.pART8		
	Ce.pARTb	Hs.pART12	Hs.pART10		
	Hs.pART4	Hs.pART11	Ag.pARTb		
	Ce.pARTc	Hs.pART6	Hs.pART16		
	Hs.pART5	Hs.pART5	Hs.pART6		
	Ci.pART		Dm.pARTb		
Ag.pARTb		Hs.pART9			
Dm.pARTb		Hs.pART5			
iteration 2	Hs.pART6	Ag.pARTb	Hs.pART3	Hs.pART16	Hs.pART3
	Hs.pART14	Dm.pARTb	Dm.pARTa	Hs.pART2	Ce.pARTa
	Hs.pART13	Hs.pART3	Ce.pARTc	Ag.pARTa	Hs.pART4
	Ba.pART	Hs.pART1	Hs.pART1	Ce.pARTa	Ce.pARTb
	Hs.pART16	Dm.pARTa	Hs.pART4		Ce.pARTc
	Hs.pART7	Ce.pARTb	Ce.pARTb		Dm.pARTb
	Hs.pART11	Hs.pART4	Ag.pARTa		Hs.pART5
	Ag.pARTc	Ce.pARTc	Hs.pART2		Ag.pARTb
	Hs.pART17	Ag.pARTa	Ce.pARTa		Hs.pART6
	Hs.pART12		Ci.pART		Hs.pART14
				Hs.pART16	
				Hs.pART7	
				Ba.pART	
iteration 3	Hs.pART8	Hs.pART2	Hs.pART17	Dm.pARTa	Hs.pART8
	Hs.pART10	Ce.pARTa	Hs.pART15	Hs.pART1	Hs.pART10
	Hs.pART15	Hs.pART15	Ag.pARTc	Hs.pART3	Hs.pART12
	Hs.pART9	Ci.pART		Ce.pARTc	Hs.pART11
		Ag.pARTc		Ce.pARTb	Hs.pART13
		Hs.pART16		Hs.pART4	Ag.pARTc
		Hs.pART17		Hs.pART5	Hs.pART17
		Ba.pART		Ag.pARTb	Hs.pART15
			Dm.pARTb		
iteration 4				Hs.pART6	Hs.pART9
				Hs.pART14	
				Hs.pART7	
				Hs.pART13	
				Hs.pART11	
				Ci.pART	
				Hs.pART12	
			Ba.pART		
iteration 5				Hs.pART8	
				Hs.pART10	
				Hs.pART9	
iteration 6	converged		Ba.pART		converged
iteration 7		converged	converged		
iteration 8				converged	

**Figure 4**  
**Representative tiling paths of PSI-BLAST searches initiated with the catalytic domain amino acid sequences of selected pART family members.** PSI-BLAST searches were initiated with the catalytic domain amino acid sequences of the pARTs indicated on top as query sequences with the default threshold setting for the expect value of 0.005. Matching sequences from selected model organisms are indicated at the iteration in which they first appeared above threshold. pART subgroups are color coded as in Figure 2. Accession numbers of the indicated pARTs are listed in Figures 2 and 9. Species of origin is color-coded in the two letter abbreviation of the organism as follows: *Homo sapiens* (Hs) red, *Drosophila melanogaster* (Dm) and *Anopheles gambiae* (Ag) purple, *Caenorhabditis elegans* (Ce) blue, Chilo iridescent virus (Ci) and Bacteriophage Aeh (Ba) brown.

The results of the secondary structure prediction and threading analyses were used to refine a multiple amino acid sequence alignment of the catalytic domains of all human pART family members. The resulting alignment is shown in Figure 5. The conserved secondary structure units corresponding to the catalytic NAD binding core (the six beta strands and one alpha helix marked in Figure 1) are indicated schematically below the alignment. The corresponding amino acid residues are highlighted in the alignment. Intron positions are projected onto the amino acid sequence in Figure 5. The positions of conserved introns are marked by colored arrows below the alignment. Note that the alignment diverges most strongly both in length and in sequence in the loops immediately downstream and upstream of  $\beta$  3.

Figure 6A shows a condensed version of the alignment in which the diverging intervening loops are indicated only by the number of amino acid residues. These 66 amino acid residues can be superimposed well in the 3D structures of PARP-1, PARP-2, DT, and ETA. The respective amino acid sequences of DT, ETA and the putative Chilo iridescent virus pART are also shown for these regions. Figure 6B shows the calculated amino acid sequence identities of the pART family members in this region. The percentage amino acid sequence identity in the aligned core region is higher among members of a particular subgroup than between members of different subgroups, lending support to the subgroup assignments. For each pART, the next most closely related paralogue is a member of the same subgroup. Note that two pairs of pART analogues show very close sequence similarity: pARTs 5 and 6 (94% identity in the aligned core region) and pARTs 16 and 17 (86% identity). This close similarity is reflected also in the conserved exon intron structures of the respective pART pairs (see Fig. 3).

Comparison of mouse and human pART orthologues shows that seven of such pairs exhibit 100% sequence identity in the aligned core region (pARTs 1, 5, 6, 11, 14, 16, and 17) and six show > 90% identity (pARTs 2, 3, 4, 10, 12, and 15). The mouse and human orthologues of pARTs 8, 9, 13 show the least degrees of sequence identity in this region (82%, 82%, and 70%, respectively) (Fig. 6B).

Phylogenetic analysis of the amino acid sequences of the catalytic cores of pARTs resulted in three very similar trees when using Maximum Parsimony (PAUP), Maximum Likelihood (PhyML), and Bayesian Markov Chain Monte Carlo (MrBayes) optimization criteria (Figure 7). All topologies showed moderate to high support values for the recovered relationships. All trees recovered five basic clades corresponding to the subgroups 1–5. The results indicate that pARTs of subgroups 1 and 2 are more closely

related (sistergroups) to one another than to members of the other subgroups. A similar relationship is seen for pARTs of subgroups 3 and 4. Note that the putative Chilo iridescent virus pART clusters with the mammalian pARTs of subgroup 1, suggesting that this large double stranded DNA virus may have acquired its pART by horizontal gene transfer.

#### **The pART catalytic domain has become genetically fused to a wide spectrum of protein domains**

With the exception of closely related members within a subgroup, the amino acid sequence similarity between pART family members breaks off upstream of  $\beta$  1. Interestingly, loss of sequence similarity correlates well with the presence of a phase 0 intron upstream of  $\beta$  1. All pART family members except pART4 and pART14 contain such a phase 0 intron 26–64 codons upstream of the conserved histidine in  $\beta$  1 (Fig. 3B).

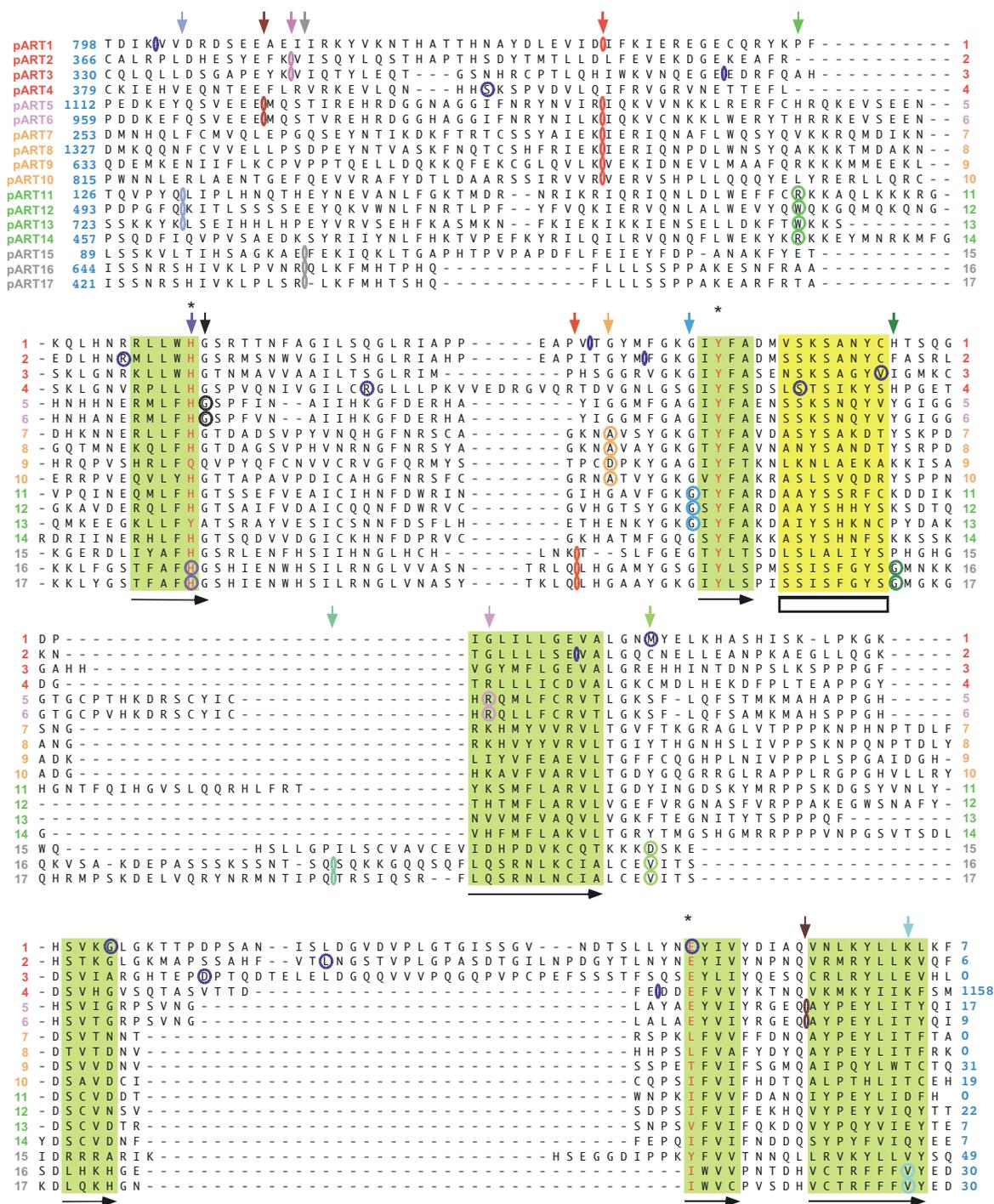
Using the sequences flanking the catalytic domain of each pART family member as queries, we performed further PSI-BLAST analyses and searches of the Conserved Domain Database [41]. The results, summarized in Figure 8, reveal that each of the 17 human pARTs with the possible exception of pART15 is a multi-domain protein. Strikingly, the pART catalytic domain is associated – in a Lego like fashion – with a broad spectrum of known protein domains. In all family members except pART4 the catalytic domain represents the C-terminal domain.

A number of associated domains occur in two or more human pART family members. Note that domain sharing generally is restricted to members of a particular pART subgroup. For example, all members of subgroup 1 contain a helical domain preceding the catalytic domain, whereas this domain is missing in members of other pART subgroups. The two members of subgroup 2 share SAM and ankyrin-repeat domains. Three of four pARTs in subgroup 3 share A1pp domains [42], all members of subgroup 4 share WWE domains, and two members of subgroup 5 contain a second, truncated pART domain, reminiscent of the duplicated inactive ART domain found in the VIP2 and iota mART toxins [16,23].

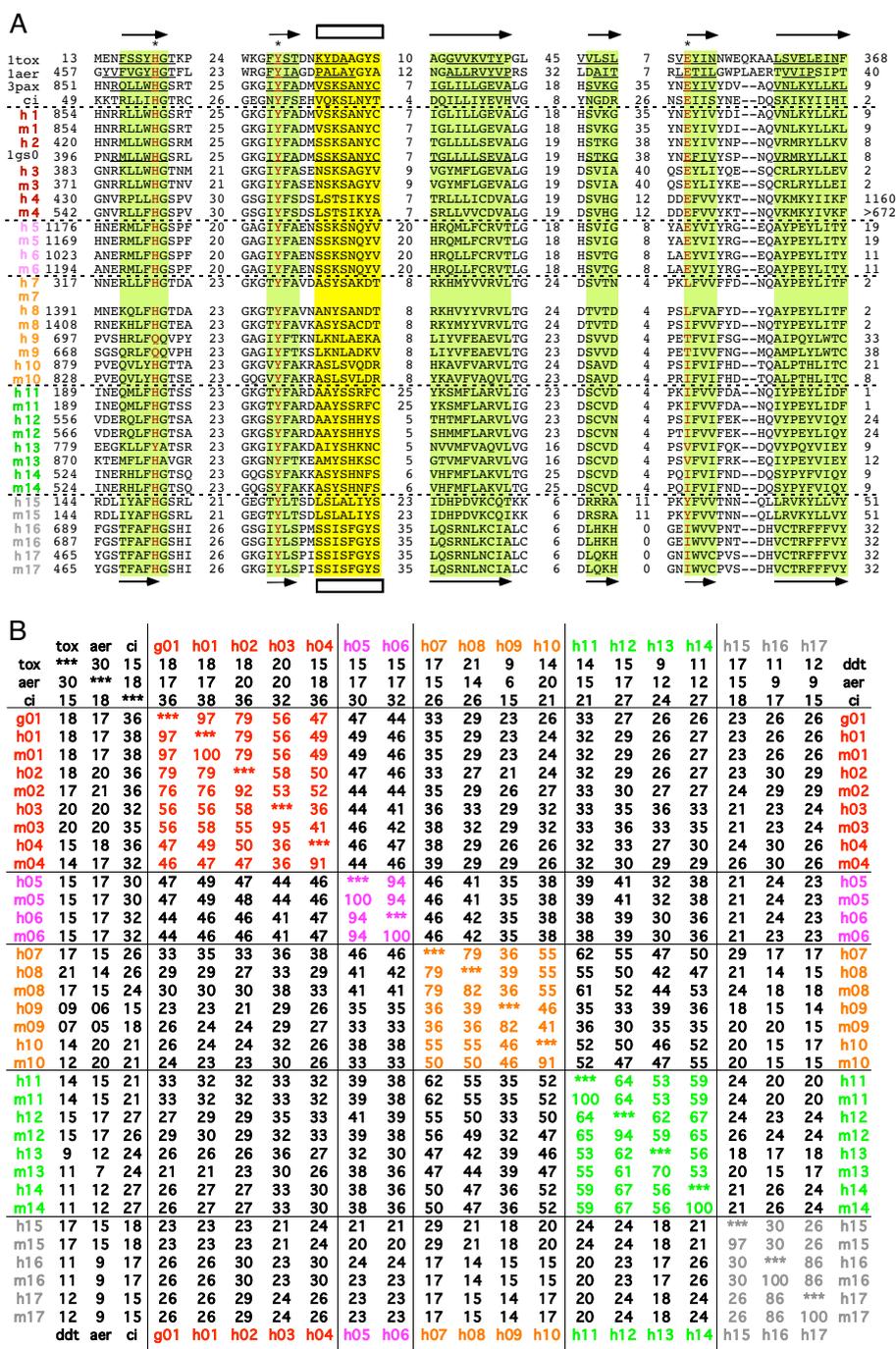
Several pARTs carry recognizable zinc-fingers containing putative RNA-, DNA-, or ubiquitin-binding domains (pART1, pART2, pART10, pART12, pART13). This indicates that the genetic fusion of a pART catalytic domain with zinc-fingers has occurred repeatedly in evolution.

#### **Representation of pARTs in other model organisms**

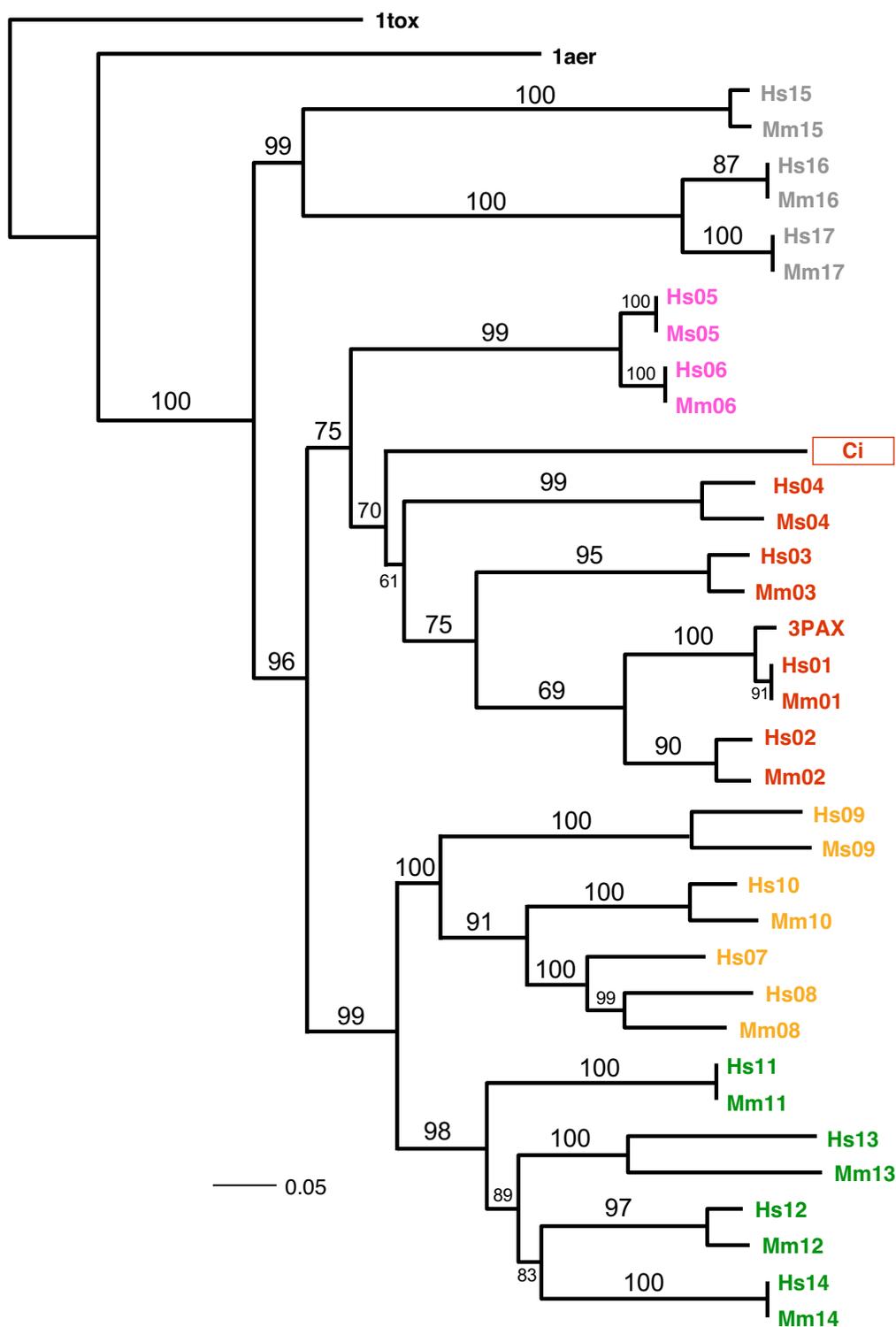
We also used PSI-BLAST to screen the protein database for recognizable pART family members in other organisms using as queries the amino acid sequences of catalytic domains of each of the 17 human pARTs (Figure 9). The



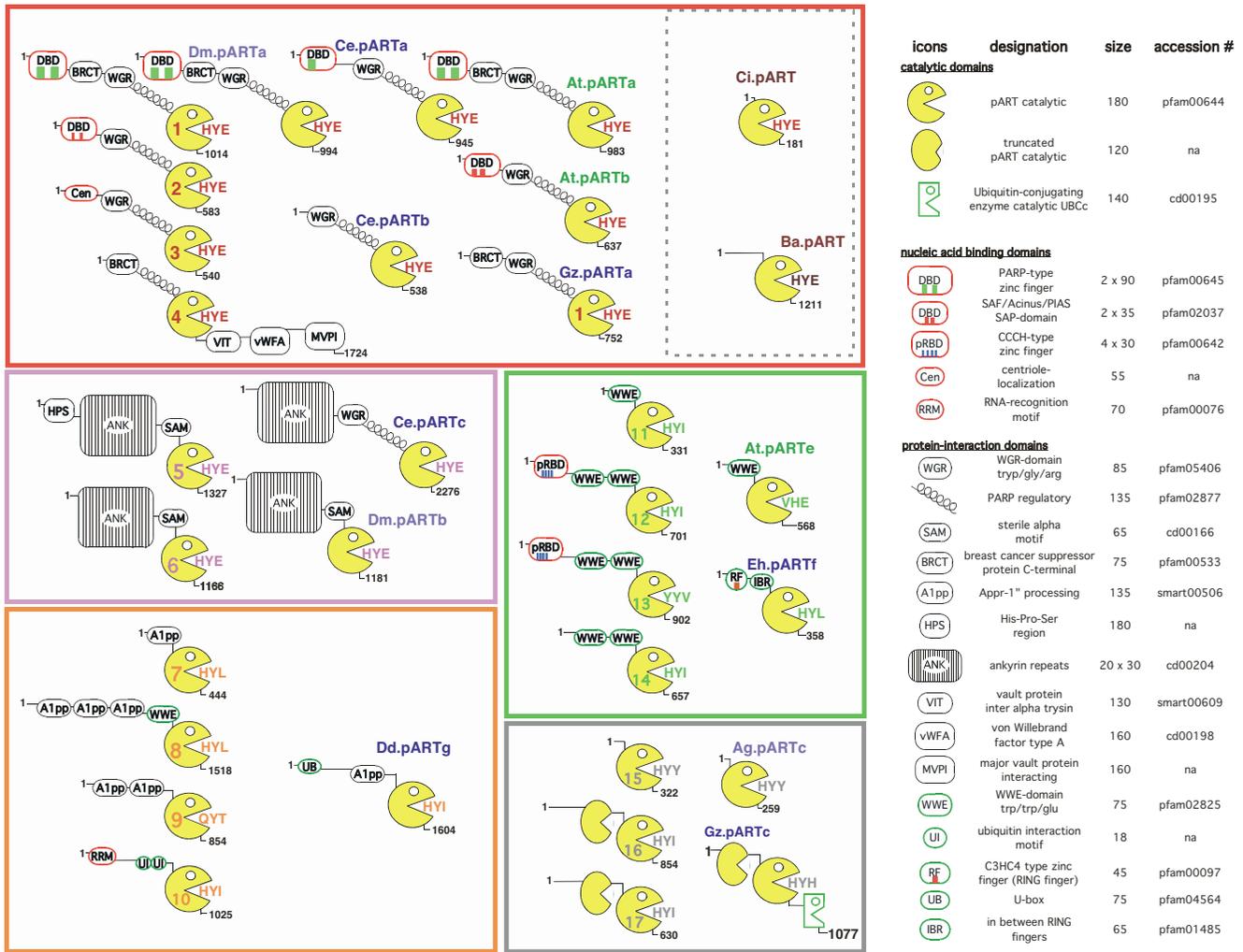
**Figure 5**  
**Multiple amino acid sequence alignment of the catalytic cores of the human pART family.** The multiple sequence alignment was generated with T-Coffee and manually adjusted using the results of the PSI-BLAST, PSIPRED, and GenTHREADER analyses. Numbers at the sequence ends indicate the number of additional residues upstream and downstream of the alignment shown. Residues corresponding to the HYE motif in the NAD binding crevice of diphtheria toxin are in red and marked by asterisks. The conserved  $\beta$  sheets and alpha helix are shaded in green and yellow. Conserved intron positions are marked in the multiple alignment using the same color-coding as in Figure 3. Conserved intron positions are indicated also above the alignment with arrows. Non-conserved intron positions are marked in blue in the alignment.



**Figure 6**  
**Structure based amino acid sequence alignment of the catalytic cores of the pART gene family.** A) The alignment is restricted to those regions corresponding to the conserved secondary structure units of PARP-I and DT as highlighted in Figure 1. The H Y E motif is marked by asterisks and is highlighted in red. Black numbers indicate amino acid residues from the N- and C-terminal ends of the protein and within the loops connecting the structure units shown. For proteins with known 3D structures the pdb accession number is given and the residues corresponding to respective secondary structure units are underlined. Itox = diphtheria toxin; Iaer = pseudomonas exotoxin A, 3pax = chicken PARP-I (pART1), Igs0 = mouse PARP-2 (pART2). Human and mouse pARTs are indicated by colored numbers. The sequence of the putative pART from Chilo iridescent virus is also shown for comparison (ci). B) Pairwise percentage sequence identities were calculated for the 66 amino acid residues shown in A), which correspond to the conserved core secondary structure units in Figure 1.



**Figure 7**  
**Phylogram of the evolutionary relationship of the pART family.** Evolutionary relationships of the amino acid sequences in the catalytic core of the pARTs shown in Figure 6 are illustrated as a maximum *a posteriori* phylogram (MAP) of Bayesian Markov Chain Monte Carlo analysis (pP = 0.92). Posterior probabilities were converted into percentages and are shown above the branches. Members of the five pART family subgroups are color-coded as in Figure 2: subgroup 1 = red, 2 = pink, 3 = orange, 4 = green, 5 = grey. Hs = *Homo sapiens*, Mm = *Mus musculus*.



**Figure 8**  
**Schematic diagram of the domain structures of human pARTs and pARTs from distantly related organisms.**  
 Recognizable protein domains in the pART family are represented by the icons defined on the right. The domain structures of human pARTs (on the left, numbered Pacman icons) and related pARTs from other species are illustrated schematically. Potential DNA binding domains are boxed in red, potential ubiquitylation motifs are boxed in green. Members of the five pART family subgroups are grouped within colored boxes using the color-coding as in Figure 2: subgroup 1 = red, 2 = pink, 3 = orange, 4 = green, 5 = grey. Amino acids corresponding to the HYE catalytic site motif of DT and PARP-1 are shown in the mouths of the Pacman icons. Black numbers indicate protein lengths in number of amino acids. Species of origin is color-coded in the two letter abbreviation of the organisms as in Figures 4 and 9: *Drosophila melanogaster* (Dm) and *Anopheles gambiae* (Ag) purple, *Caenorhabditis elegans* (Ce), *Dictyostelium discoideum* (Dd), *Entamoeba histolytica* (Eh), and *Gibberella zeae* (Gz) blue, *Arabidopsis thaliana* (At) green, Chilo iridescent virus (Ci) and Bacteriophage Aeh (Ba) brown. Protein database accession numbers for the illustrated pARTs are listed in Figures 4 and 9. On the right, the approximate size of each domain is indicated in number of amino acid residues. The accession numbers of the respective domain families in the pfam, cd, and smart databases are indicated. In case of zinc finger (zf) containing domains, the number of recognizable zinc fingers is indicated by colored bars within the icon.

order in which PSI-BLAST picked up putative pART sequences from the database in successive iterations was similar for different members of a particular pART sub-

group but differed markedly for members of different subgroups (see additional file 8: "Representative tiling paths of PSI-BLAST searches initiated with the catalytic domain

pART	protein	aliases	human	mouse	chicken	fish	fly	mosquito
1	PARP1	PARP	P09874	NP_031441	NP_990594	CAG09179	P35875_a	XP_312938_a
2	PARP2		Q9UGN5	NP_033762		CAF92030		
3	PARP3		AAM95460	NP_663594		CAG06805		
4	PARP4	vaultPARP	AAD47250	XP_283217	XP_417150	CAG08214		
5	TNKS	Tankyrase	AAC79841	AAH57370	NP_989671	CAG12585		
6	TNKS2	Tankyrase 2	NP_079511	XP_129246	NP_989672	CAG04910	AAF56487_b	XP_321116_b
7	PARP15		NP_689828	---				
8	PARP14		AAN08627	XP_488522	XP_422113	CAF98988		
9	PARP9	BAL	NP_113646	NP_084529	XP_422116	CAG12587		
10	PARP10		BAB55067	AAH24074		CAG05989		
11	PARP11		AAF91391	NP_852067	XP_416489	CAG01913		
12	ZC3HDC1		NP_073587	NP_766481	XP_416333	CAF98285		
13	ZC3HAV1	ZAP	NP_064504	BAB32047	XP_423977	CAF96305		
14	TIPARP	TiPARP	NP_056323	NP_849223	XP_422828	CAF96664		
15	PARP16		AAH31074	NP_803411	XP_413903	CAG05566		XP_308419_c
16	PARP8		NP_078891	AAH21881		CAG05573		
17	PARP6		CAB59261	XP_134863	XP_424786	CAF95416		

pART	protein	nematode	slime mold	fungi	amoeba	weed	viruses	bacteria
1	PARP1	AAM27195_a	CAD59237_a	EAA75569_a	EAL47198_a	NP_850165_a	AAB94432 AAQ17796	IAERA 760286A AAW80252
2	PARP2		CAD58666_c		EAL50270_b	CAA88288_b		
3	PARP3	Q09525_b	CAD59238_d			BAB09119_c		
4	PARP4		CAD59240_e					
5	TNKS	AAC04454_c	CAD59239_b					
6	TNKS2		AA051129_f					
7	PARP15							
8	PARP14		AAS38928_g					
9	PARP9							
10	PARP10							
11	PARP11				EAL43406_c			
12	ZC3HDC1				EAL50270_d	NP_849739_d		
13	ZC3HAV1				EAL49071_e	AAC36170_e		
14	TIPARP				EAL45174_f			
15	PARP16							
16	PARP8							
17	PARP6			EAA73885_c				

**Figure 9**  
**pARTs in distantly related species.** pART relatives were identified by PSI-BLAST searches as in Figure 4. Matching sequences from other organisms were sorted by group on the basis of sequence similarity and associated domains. Accession numbers are given for pARTs from *Homo sapiens* (human), *Mus musculus* (mouse), *Gallus gallus* (chicken), *Tetraodon nigroviridis* (puffer fish), *Drosophila melanogaster* (fruit fly), *Anopheles gambiae* (malaria mosquito), *Caenorhabditis elegans* (nematode), *Dictyostelium discoideum* (slime mold), *Gibberella zeae* (ear root microfungus), *Entamoeba histolytica* (amoeba), *Arabidopsis thaliana* (cress plant), Chilo iridescent virus and Bacteriophage AehI (viruses), *Pseudomonas aeruginosa*, *Corynebacterium diphtheriae* and *Vibrio cholerae* (bacteria). Lower case letters in black indicate the pART designations used in Figure 8.

amino acid sequences of selected pART family members"). In many instances, PSI-BLAST detected pART sequences from distantly related organisms in earlier iterations than the human pART paralogues from other subgroups.

Figure 9 summarizes the matches of pART-related proteins found in model organisms with completed genome sequences. On the basis of amino acid sequence similarity, conserved intron positions and/or conserved associ-

ated domains, pARTs from other vertebrates including fish and chicken, generally can be assigned to a particular human pART orthologue. In contrast, pARTs of lower eucaryotes can be assigned to a subgroup but not to a particular vertebrate pART.

pART homologues were found in many model organisms from the animal, plant, fungi, and protist kingdoms. The recently completed genome of the pufferfish *T. nigroviridis* contains recognizable orthologues for all pARTs except for

pART7 [43]. The nearly completed albeit still fragmentary chicken genome contains recognizable orthologues for all pARTs except for pARTs 2, 3, 7, 10, and 17 [44]. Simpler eucaryotes generally contain fewer pARTs (two in the fruit fly *D. melanogaster*, three each in the malaria mosquito *A. gambiae*, the nematode *C. elegans*, and the ascomycete *G. zaeae*; six in the amoeba *E. histolytica*, nine in the slime mold *D. discoideum*, and ten in the cress plant *A. thaliana*).

Remarkably, the yeast *S. cerevisiae* and the archaea lack detectable pARTs. Only two matches were found in the viral proteome: these derive from two double stranded DNA viruses: the insect virus Chilo iridescent virus and the bacteriophage Aeh1. Although PSI-BLAST initially failed to connect the pART family with Diphtheria toxin and Pseudomonas exotoxin A, these toxins were readily connected with the eucaryotic pARTs when using as query a chimera, e.g. of Diphtheria toxin and Chilo iridescent virus pART in which the sequences of three of the conserved structure units highlighted in Figures 1 and 6A were interchanged. These searches uncovered a DT/ETA-like putative ADP-ribosyltransferase in *V. cholerae*, but no other proteins in the microbial proteome in GenBank.

Of note, none of the known R-S-E motif bacterial or vertebrate mARTs were ever connected by PSI-BLAST with the DT/ETA/pART group. In several cases, however, we observed intriguing matches just slightly below threshold (in the region surrounding the conserved H in  $\beta$  1) to members of the family of RNA:NAD 2' phosphotransferases. These enzymes catalyze a reaction during tRNA splicing that is similar to the reaction catalyzed by ARTs, but in which ADP-ribose is transferred to the 2'-phosphate in immature tRNA rather than to an amino acid residue in a protein [25]. The 3D-structure of a prototype member of this gene family, indeed, reveals a structure closely resembling that of PARP-1 and Diphtheria toxin (see Fig. 1), providing strong support for the relevance of the matches detected by PSI-BLAST.

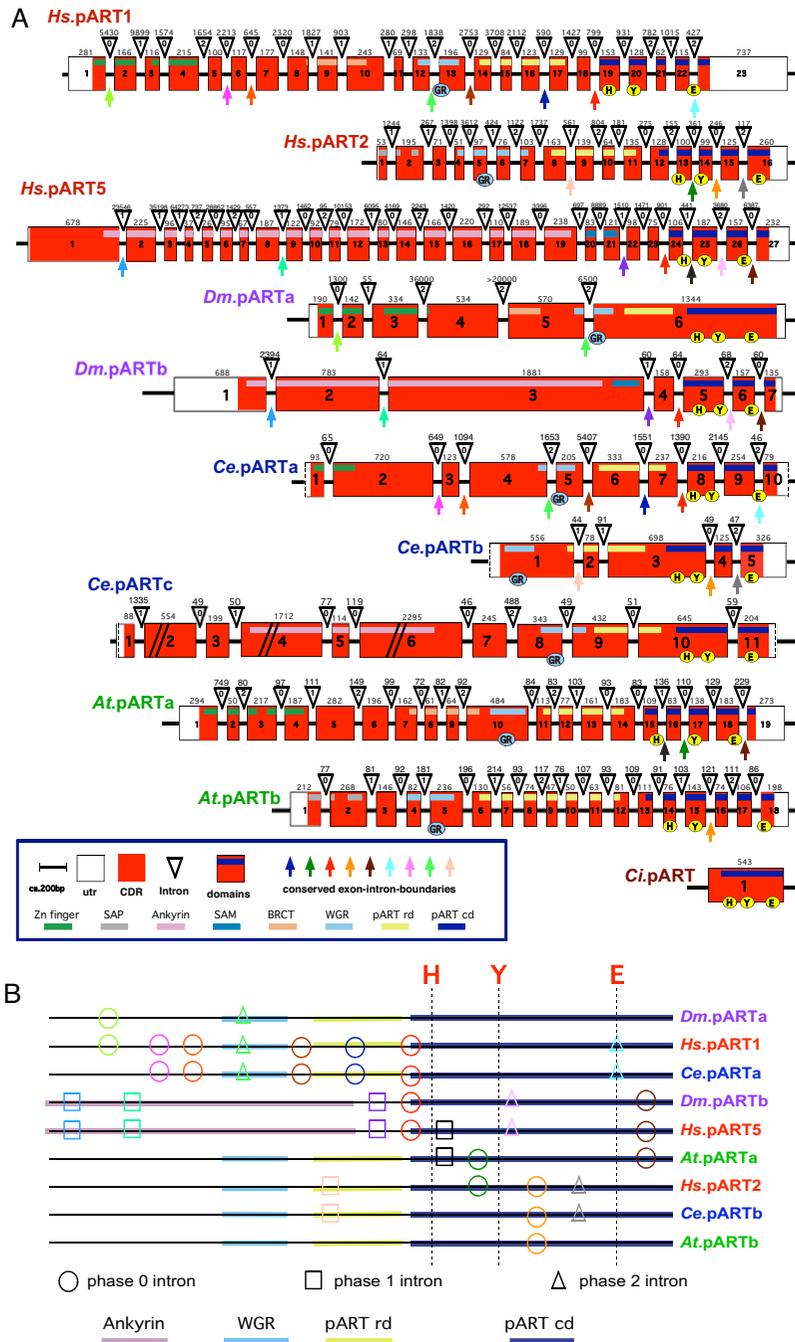
For the pART homologues shown in Figure 9 we also analyzed the sequences flanking the pART catalytic domain for associated conserved domains. The results reveal that many pARTs, even from very distantly related organisms, share domain associations found in human and mouse pARTs. Some of these are illustrated in Figure 8. For example, the association of regulatory, BRCT, and DNA binding domains observed in pART1 (PARP-1) is found also in similar proteins encoded by fruit fly, nematode, microfungi and cress plant genomes. Tankyrase-like association with ankyrin repeats is found in pARTs from the fruit fly and nematode. The association of a pART catalytic domain with an A1pp domain, as seen in human pART subgroup 3, is found also in a pART from the slime mold *Dictyostelium discoideum*. The combination with a WWE

domain, as seen in human pART subgroup 4, is found also in putative pARTs from cress plant. A domain corresponding to the unknown upstream region of the smallest human pART (pART15) is observed also in a pART from the malaria mosquito *Anopheles gambiae*, and a duplicated truncated pART catalytic domain as in pARTs 16 and 17 is observed also in a pART from the microfungus *Gibberella zaeae*. These results indicate that many of the domain combinations observed in human and mouse pARTs represent evolutionary ancient inventions.

Some pARTs of distantly related proteins are associated with domains not found in any of the human pARTs. A striking example is that of *G. zaeae* pARTc, which most closely resembles human pARTs 16 and 17, but is associated with a second potential catalytic, ubiquitin ligase domain (Fig. 8). A similar pART is found also in the related microfungus *Aspergillus nidulans* [GenBank: EAA66581]. These microfungus pARTs are the only examples found so far, in addition to vertebrate pART4, where a distinct domain(s) is genetically fused to the C-terminal end of the pART catalytic domain. The large domain(s) associated with the putative pART from bacteriophage Aeh1 does not bear any resemblance to pART-associated domains in vertebrates but shows distant similarity to viral coat proteins. The only organism containing an isolated pART domain reminiscent of the isolated ART domain found in vertebrate mARTs [27] is the Chilo iridescent insect virus. This "naked" viral pART catalytic domain contains the H-Y-E motif of PARP-1 and DT. It will be interesting to determine whether this protein exhibits the predicted pART activity.

A striking example of domain shuffling is observed in one of the three *C. elegans* pARTs: like the human tankyrases (pARTs 5 and 6), *Ce.pARTc* contains ankyrin repeats, but also harbors the regulatory and WGR domains typical of human group 1 pARTs instead of the SAM domain found in human pARTs 5 and 6 (Fig. 8). A similar variation of domains as in *Ce.pARTc* is found also in one of the ten pARTs of *D. discoideum* (*Dd.pARTb*).

Finally, we addressed the question whether the striking differences in exon/intron compositions of the closest PARP-1-homologues in groups 1 and 2 might be reflected in similar differences in pART orthologues of distantly related species. To this end we determined the exon/intron structures of distant pART orthologues by BLASTn searches of the respective genome databases using cDNA sequences as queries; and compared the results with those obtained for human pART genes. The results are illustrated schematically in Figure 10, with conserved intron positions highlighted. As in case of most other genes, the pART genes of 'lower' animals, protists, and plants in general contain fewer and shorter introns than the human



**Figure 10**

**Schematic diagram of the exon/intron structures of pART family members of distantly related organisms. A)** Exon/intron structures were determined by BLASTn searches of the genome browsers using the pART cDNA sequences. The positions of codons corresponding to the H Y E motif in the NAD-binding crevice of diphtheria toxin are marked by yellow circles. The position of the conserved glycine and arginine pair of residues within the WGR domain is marked in blue. Coding regions for catalytic and other domains are indicated by colored bars. Conserved introns are marked by colored arrows. **B)** The diagram contains only those introns that are conserved in at least two distantly related species. Color-coding of the introns corresponds to that shown in A). The position of codons encoding/corresponding to the H, Y, E residues in the NAD binding crevice are indicated by vertical lines. The position of each intron with respect to the codon is indicated by circles (phase 0 introns), boxes (phase 1 introns), and triangles (phase 2 introns). Coding regions for catalytic and other selected domains are indicated by colored lines as in A).

homologues. However, some of the introns found in human pART genes are found also in homologues of distantly related organisms. For example, all six introns observed in *D. melanogaster* pARTb are found in corresponding positions also in human pART5 (tankyrase 1); yet human pART5 contains 14 additional introns not found in the fruit fly pART. The other pART of the fruit fly shares two of its five introns with human pART1 (PARP-1). The three pARTs of the nematode *C. elegans* show a different, only partially overlapping set of conserved introns: Ce.pARTa shares seven of its nine introns with human pART1, Ce.pARTb shares three of its four introns with human pART2, whereas Ce.pARTc does not seem to share any of its introns with pART5, despite the similar domain organization on the protein level (see Fig. 8). The pARTs from the model plant *Arabidopsis thaliana* contain a fairly high number of introns, however only very few intron positions correspond to ones found also in human pARTs. For example, At.pARTa which is most closely related to human PARP-1 in terms of amino acid sequence similarity and organization of conserved protein domains, evidently does not share any of its 18 introns with human pART1. Strikingly, however, the introns found in the catalytic domain of this pART exhibit conserved positions with two different human pARTs: two of the four intron positions in the catalytic domain of At.pARTa are found in corresponding positions in human pART5 (tankyrase), another intron is found at a corresponding position in human pART2 (Fig. 10), whereas the fourth intron is not found in any human pART. At.pARTb which is most closely related to human pART2 in terms amino acid sequence similarity and domain organization, shares one of its 17 introns with human pART2. Note further, that in only two cases (Chilo iridescent virus pART and pARTa of the fruit fly), the pART catalytic domain lacks introns, i.e. is encoded by a single exon as in case of the vertebrate mARTs [27].

## Discussion

The results of our study illustrate the great power and utility of the public genome databases and database search programs. Moreover, they provide important novel insights into the molecular structure and evolution of the pART gene family.

Our results differ in some details from those of a recent report by Ame and coworkers [11]. These discrepancies can be explained by errors in the draft sequence of the human genome available at the time of the previous report. For example, the database entry AK023746 given by Ame *et al.* for PARP-5c evidently represents a truncated cDNA for pART6 (alias tankyrase 2 or PARP-5b). This entry contains two point mutations and a 65 bp deletion in the 3' utr vs. the cDNA and genomic sequences of pART6. Blast analyses of the high quality sequence of the

human genome and of the EST database with the AK023746 sequence provide no evidence for a distinct copy of this gene in the human genome. We conclude that the PARP-5c gene identified by Ame *et al.* represents an allelic variant or cloning/sequencing error rather than a genuine pART gene family member; i.e. that the total number of human pART genes is 17 rather than 18 suggested in the previous report. Large discrepancies exist also in the number of amino acids assigned in the two reports for pART7/PARP-15 (444 vs. 989) and for pART16/PARP-8 (854 vs. 501). The earlier database entries for PARP-8 (XM\_018395) and PARP-15 (XM\_093336) have hence been removed as a result of standard genome annotation processing because these entries evidently contained frameshift mutations and/or fused cDNA sequences that led to erroneous amino acid assignments. Similarly, the small differences in assignments for five other PARPs/pARTs can be accounted for by differences in the draft vs. high quality sequence of the human genome (Ame *et al.*/our study): pART2/PARP2 (583/570), pART3/PARP3 (540/533), pART10/PARP10 (1020/1025), and pART14/PARP7 (657/680).

We assigned the 17 human pARTs into five distinct subgroups (Fig. 2). This assignment is supported by several independent lines of evidence: Firstly, members of a particular subgroup show higher amino acid sequence identities to one another than to members of other subgroups (Fig. 6). This is reflected in the tiling paths of PSI-Blast searches, where members of the same subgroup were detected in the first iteration, whereas members of other subgroups generally were detected in later iterations (Fig. 4). Secondly, members of a particular subgroup typically share one or more associated domains not found in members of other subgroups (Fig. 8); pARTs 8, 10 and 15 pose exceptions to this rule. Thirdly, members of a particular subgroup typically share one or more intron positions not found in members of other subgroups (Fig. 3); pARTs 1–4 pose notable exceptions to this rule. Fourthly, when genes of two or more pARTs are physically linked in a cluster on the same chromosome, they belong to the same subgroup – possibly reflecting regional duplications (Fig. 2). Finally, results of all phylogenetic analysis converged in topologies with clearly distinct clades for each of the subgroups (Fig. 7). Members of subgroups 1 and 2 evidently are more closely related to one another than to other subgroups (Figs. 6 and 7). Similarly, members of subgroups 3 and 4 are sister-groups to one another, indicating a close relationship.

Members of the pART family are found fused to a striking variety of associated domains (Fig. 8). It is not farfetched to hypothesize that the associated domains direct the respective pARTs to subcellular structures and/or target proteins. Genetic fusion of group 1 and group 2 pARTs

with DNA-binding domains is in line with their established roles in DNA-repair, chromosome remodeling, and mitotic spindle formation [9,11,12]. Moreover, the SAM and ankyrin domains of pARTs 5 and 6 have been shown to mediate interactions with target proteins in telomere-associated protein complexes [45]. Similarly, the C-terminal domain of pART4 evidently plays a role in targeting pART4 to the major vault particles [46]. A flurry of domains implicated in the ubiquitination pathway point to a possible connection between ubiquitination and ADP-ribosylation. Indeed, it has recently been reported that ADP-ribosylation of TRF1 by tankyrase (pART5) results in the release of the protein from telomers and its subsequent ubiquitination [47]. Strikingly, pARTs from the microfungi *G. zeae* and *A. nidulans* provide examples for the genetic fusion of two enzyme domains catalyzing these post-translational protein modifications into a single polypeptide.

So far, only a single example of a 'naked' pART catalytic domain akin to the isolated catalytic domain of the vertebrate ecto-ARTs 1–5 [27] was recovered from the public database. This putative pART from Chilo iridescent virus clusters with the mammalian pARTs of subgroup 1 (Fig. 7), suggesting that this large double stranded DNA virus [48] may have acquired its pART by horizontal gene transfer.

The definition of the pART catalytic domain proposed in this paper is somewhat smaller than that commonly used in the field [11]. We used the position of the common phase 0 intron upstream of the first conserved  $\beta$  sheet to set the N-terminal end of the catalytic domain (e.g. see Figs. 1 and 3B). The pARTs of subgroup 1 are extended N-terminally of this position by an alpha helical domain (Fig. 8) which is often included as part of the PARP-1 catalytic domain. However, since other pART family members lack this region, we propose to omit it from the proper pART catalytic domain. Moreover, this N-terminal delineation of the catalytic domain corresponds well to the N-terminus of the 'naked' pART of Chilo iridescent virus as well as to those of Diphtheria toxin and Pseudomonas exotoxin A after proteolytic processing of the signal sequence or translocation domain (Fig. 1).

With the exception of pART4, the group 1 pARTs are extended upstream of this helical region by another domain named after its conserved motif of tryptophane (W) – glycine (G) – arginine (R) residues. This WGR domain is found also in poly-A-polymerases, its function is unknown. Many group 1 pARTs from distantly related organisms, e.g. plants, insects, nematodes, and microfungi, also contain these two domains. Interestingly, in *Drosophila melanogaster* pARTa these three domains (WGR, helical, catalytic) are encoded by a single, large exon (Fig.

10). Human pARTs 5–17 lack the WGR and helical domains. However, pART5/6 (tankyrase)-like pARTs from *C. elegans* (Ce.pARTc) and *D. discoideum* (Dd.pARTb) contain the WGR and helical domains whereas a SAM domain is found at this position in human pARTs 5 and 6 (Fig. 8).

A puzzling finding is the lack of conservation of the classic H-Y-E motif found in the catalytic cores of PARP-1, PARP-2, Diphtheria toxin and Pseudomonas Exotoxin A (Fig. 1). This motif is conserved only in members of subgroups 1 and 2. All other human pARTs carry notable variations from this motif. In particular, all other pARTs carry a replacement of the glutamic acid residue in  $\beta$  5, i.e. the residue that was shown to be critical for the catalytic activities of DT, PARP-1 and many other pARTs and mARTs [6,7,20,21]. In six cases, this glutamic acid is replaced by an isoleucine residue, in two cases by leucine, and in one case each by threonine, valine, or tyrosine. Enzyme activity has been reported recently for two of the six pARTs that carry an H-Y-I motif instead of the H-Y-E motif (pARTs 10 and 14) [32,34]. Thus, it is not unlikely that the four other pARTs carrying the H-Y-I motif turn out to be active enzymes (pARTs 11, 12, 16, and 17). Mouse pART8 also carries an H-Y-I motif, whereas its human orthologue, like pART7, carries an H-Y-L variant motif. H-Y-I and H-Y-L variant motifs are also found in pARTs from the slime mold (Dd.pARTg) and amoeba (Eh.pARTf) (Fig. 8). Human pART15 carries an H-Y-Y variant motif, which is conserved in its orthologues from mouse and the malaria mosquito (Fig. 8). It will be interesting to determine whether and how site directed mutagenesis of the H-Y-E motif in pARTs 1–6 to the variant motifs of pARTs 7–17 – and vice versa – affects their enzyme activities. Moreover, it remains to be determined whether the most striking variation of the H-Y-E motif – to Q-Y-T in human and mouse pART9 is compatible with enzyme activity.

The results of our PSI-BLAST and PSIPRED analyses (Figs. 4, 5, 9 and additional files 3, 4, 5, 6, 7, 8) support the conclusions that the pART gene family described here and the mART gene family described in our previous study [27] constitute two distinct ART subfamilies, and further, that the family of tRNA:NAD 2'-phosphotransferases [24,25] constitutes a branch that is more closely related to the pART subfamily than to the mART subfamily. Our results illuminate the power and limits of PSI-BLAST searches: PSI-BLAST readily connected members of the pART subfamily in many different species, while DT, ETA and TpTs were found at or below the threshold. In contrast PSI-BLAST searches never connected pART family members with members of the mART subfamily or vice versa. The results of PSI-BLAST searches, thus, are in accord with insights gained from the known 3D structures of representative ADP-ribosyltransferases (Fig. 1), i.e. that certain

conserved structural features clearly distinguish these two subfamilies. Is it possible that some of the pART gene family members described here actually possess mono-ADP-ribosyltransferase rather than poly-ADP-ribosyltransferase activity? Given the structural similarity to DT/ETA this is a possibility. Moreover, it cannot be excluded that some family members may have lost enzyme activity and have acquired a new function. In any case, the respective proteins clearly are more closely related to the pART than to the mART gene family, in line with the nomenclature proposed here. Have all ARTs encoded in the human genome been identified? A number of ADP-ribosylation reactions have been described in mammalian cells that cannot yet be accounted for by the ARTs identified in this study or our previous study, e.g. mono-ADP-ribosylation of actin, rho, glutamate dehydrogenase, and of the alpha and beta subunits of heterotrimeric G proteins [3,4,8]. Given the fact that the pART subfamily described here and the mART subfamily described in our previous study [27] could not be interconnected by PSI-BLAST, it remains an intriguing possibility that other ART subfamilies in the human genome still await to be identified.

## Conclusion

The family of proteins containing a PARP-like catalytic domain consists of 17 members in the human and 16 in the mouse, rat, and pufferfish. The vertebrate pART family can be divided into five subgroups on the basis of sequence similarity, phylogenetic relationships, conserved intron positions, and patterns of genetically fused protein domains. The four members of group 1 and the two members of group 2 each contain a conserved triad of residues (H-Y-E motif) also observed in Diphtheria toxin and Pseudomonas exotoxin A. The eleven other pART proteins carry variants of this motif (six H-Y-I, two H-Y-L, and one each Q-Y-T, Y-Y-V, H-Y-Y). All human pARTs are multi-domain proteins in which the pART catalytic domain is associated in a Lego-like fashion with other putative protein-protein interaction, DNA binding and ubiquitination domains. In all but one case (pART4) the catalytic domain represents the C-terminal end of the multi-domain protein. Most of the domain associations observed in human pARTs appear to be very ancient inventions since they can be found also in insects, plants, microfungi, and amoeba.

## Methods

### Database searches

Protein databases were searched using PSI-BLAST [35]. Genome databases were searched using BLASTn and tBLASTn [49]. Tissue distributions of pART-ESTs were analyzed using Electronic Northern calculations at the GeneCard website [50].

### Structure and sequence analyses

Amino acid sequence alignments were performed with T-Coffee [36]. Secondary structure predictions were performed with PSIPRED [37]. Threading of amino acid sequences onto known 3D structures in PDB were performed with GenTHREADER [37]. Sequence analyses were performed using DNA-Star software, 3D-images were prepared with PyMol [51] software.

### Phylogenetic analyses

Phylogenetic analyses were applied to the 36 catalytic core amino acid sequences using the dataset in Figure 6. Phylogenetic analyses were performed on the computational cluster of the College of Biology and Agriculture at Brigham Young University by using maximum parsimony and Bayesian Markov chain Monte Carlo approaches <http://babeast.byu.edu>. The topologies were reconstructed using equally weighted maximum parsimony (MP) analysis as implemented in PAUP\* 4.0b10 [52], maximum likelihood (ML) with simultaneous adjustment of topology, and branch length as implemented in PhyML [53], as well as Bayesian methods coupled with Markov Chain Monte Carlo inference (BMCMC, MrBayes) [54]. The best fit likelihood model for amino acid evolution was determined based on the lowest Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) score as implemented in ProtTest 1.2.6 [53,55,56].

The MP analysis was run using 5000 random addition replicates and tree bisection-reconnection branch swapping. Nonparametric bootstrap values were calculated for MP and ML analyses (10,000/100 bootstrap replicates, 100/1 heuristic random addition replicates) to assess confidence in the resulting relationships. ML analysis was run implementing the RtREV+I+G+F model of amino acid evolution (AIC= 4907.73; -lnL= 2800). The *a priori* information obtained by ProtTest 1.2.6 was incorporated into the BMCMC analysis. Bayesian phylogeny estimation was achieved using random starting trees, run for  $3 \times 10^6$  generations, with a sample frequency of 1000, and ten chains (nine heated, temperature= 0.2). Analyses were repeated three times to check for likelihood and parameter mixing and congruence. Likelihood scores were plotted against generation time to determine stationery levels. Sample points before reaching stationery were discarded as "burn-in". Repeated analyses were compared for convergence on the same posterior probability distributions [57]. The maximum *a posteriori* tree (MAP) is presented in this paper, showing to percentage converted posterior probabilities (pP%).

### Abbreviations used

ART = ADP-Ribosyltransferase, BLAST = basic local alignment search tool, 3MB = 3-methoxybenzamide, NAD =

nicotinamide adenine dinucleotide, PDB = protein database.

### Authors' contributions

This study was initiated in the summer of 1997 while FKN was a visiting scientist in FB's lab at DNAX. Initial database searches were performed by FKN and FB, later searches by HO, PAR, and FKN. KD performed the phylogenetic analyses. FKN supervised the study with essential contributions by FH. HO prepared the figures and FKN wrote the paper. The results represent the partial fulfillment of the requirements for the graduate thesis of HO.

### Additional material

#### Additional File 1

**Representation of pART gene transcripts in the database of expressed sequence tags** The public EST database was screened for ESTs encoding pARTs using tBLASTn and the amino acid sequences of the catalytic domain of known pART family members as queries at the dates indicated on top. Accession numbers of the corresponding Unigene clusters are indicated. Blank fields indicate lack of detectable ESTs encoding the respective pART catalytic domain. Tissue distribution analyses were performed for each cluster by "electronic Northern" analyses. For each family member, the two tissues with the highest numbers of ESTs are indicated. Tissue abbreviations: BMR bone marrow, BRN brain, HRT heart, MSL muscle, PNC pancreas, PST prostate, KDN kidney, LNG lung, LVR liver, LYN lymph node, SPC spinal chord, SPL spleen, TMS thymus, UTR uterus

Click here for file  
[<http://www.biomedcentral.com/content/supplementary/1471-2164-6-139-S1.pdf>]

#### Additional File 2

**Schematic illustration of the local human and mouse chromosomal environments of the pART subgroup 3 gene cluster** The figure schematically illustrates the local chromosomal environment of the syntenic cluster of pART genes and neighboring genes on human chromosome 3q (top) and mouse chromosome 16B3 (bottom). The order and orientation of all genes in the depicted cluster is conserved. Known transcripts in GenBank are indicated schematically with their respective accession number. Exons are indicated by boxes. The direction of transcription is marked by arrows. Grey vertical bars correspond to a scale of 10,000 base pairs. The figure was modified from the respective online UCSC human and mouse genome browsers <http://genome.ucsc.edu>.

Click here for file  
[<http://www.biomedcentral.com/content/supplementary/1471-2164-6-139-S2.pdf>]

#### Additional File 3

**Multiple amino acid sequence alignments, secondary structure predictions, and threading results for pART subgroup 1** A multiple sequence alignment was generated for the catalytic domains of pARTs 1–4 with T-Coffee. Each residue in the sequence is reported as a single letter code. Secondary structure units in the 3D structures of chicken PARP-1 (1a26) and mouse PARP-2 (1GS0) are indicated on top of the alignment. Positions with identical residues in all sequences are marked by asterisks, similarities are marked with colons and periods below the alignment. Residues corresponding to the HYE motif in the NAD binding crevice of diphtheria toxin are marked in red. Intron positions are projected onto the multiple alignment and are marked in grey (phase 0), blue (phase 1), and yellow (phase 2). Secondary structure predictions were generated for human pART1 with PSIPRED and are indicated in blue below the alignment (pr1); the confidence of the prediction is indicated in orange (highest confidence = 9). Secondary structure units are abbreviated as follows: H = helix; B = residue in isolated beta bridge; E = extended beta strand; G = 310 helix; I = pi helix; T = hydrogen bonded turn; S = bend.

Click here for file  
[<http://www.biomedcentral.com/content/supplementary/1471-2164-6-139-S3.pdf>]

#### Additional File 4

**Multiple amino acid sequence alignments, secondary structure predictions, and threading results for pART subgroup 2** A multiple sequence alignment was generated for the catalytic domains of pARTs 5 and 6 with T-Coffee. Residues, identities, intron positions, and secondary structure units are marked as in additional file 3. Indicated secondary structure predictions were generated for human pART5 (pr5) with PSIPRED.

Click here for file  
[<http://www.biomedcentral.com/content/supplementary/1471-2164-6-139-S4.pdf>]

#### Additional File 5

**Multiple amino acid sequence alignments, secondary structure predictions, and threading results for pART subgroup 3** A multiple sequence alignment was generated for the catalytic domains of pARTs 7–10 with T-Coffee. Residues, identities, intron positions, and secondary structure units are marked as in additional file 3. Indicated secondary structure predictions were generated for human pART7 (pr7) with PSIPRED.

Click here for file  
[<http://www.biomedcentral.com/content/supplementary/1471-2164-6-139-S5.pdf>]

#### Additional File 6

**Multiple amino acid sequence alignments, secondary structure predictions, and threading results for pART subgroup 4** A multiple sequence alignment was generated for the catalytic domains of pARTs 11–14 with T-Coffee. Residues, identities, intron positions, and secondary structure units are marked as in additional file 3. Indicated secondary structure predictions were generated for human pART11 (pr11) with PSIPRED.

Click here for file  
[<http://www.biomedcentral.com/content/supplementary/1471-2164-6-139-S6.pdf>]

### Additional File 7

Multiple amino acid sequence alignments, secondary structure predictions, and threading results for pART subgroup 5. A multiple sequence alignment was generated for the catalytic domains of pARTs 15–17 with T-Coffee. Residues, identities, intron positions, and secondary structure units are marked as in additional file 3. Indicated secondary structure predictions were generated for human pART15 (pr15) and for human pART16 (pr16) with PSIPRED.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-6-139-S7.pdf>]

### Additional File 8

Representative tiling paths of PSI-BLAST searches initiated with the catalytic domain amino acid sequences of selected pART family members. PSI-BLAST searches were initiated with the query sequences indicated on top at a threshold setting for the expect value of 0.005 as in Figure 2. pART subgroups are color coded as in Figure 2. Matching sequences from the slime mold (*D. discoideum*, blue) and from a model plant (*A. thaliana*, green) are indicated at the iteration in which they first appeared above threshold. The respective pART homologues from these species were arbitrarily numbered (pARTa-j) in the order in which they were detected in the search that was initiated with human pART1 (PARP-1). Protein data base accession numbers are listed in Figure 9. pARTs indicated in black include short possibly truncated coding sequences of pART homologues that could not be assigned to a particular subgroup with certainty.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-6-139-S8.pdf>]

### Acknowledgements

This work was supported by grant No310/3 from the Deutsche Forschungsgemeinschaft to FKN. HO was a grantee of the Studienstiftung des Deutschen Volkes. KD is funded by the NSF grants DEB-0120718 and DEB-9983195. DNAX is fully funded by the Schering Corporation. We thank Sahil Adriouch, Bernhard Fleischer, Stefan Kernstock, and Stefan Rothenburg (University Hospital Hamburg) for critical reading of the manuscript.

### References

- Aktories K, Just I: **Bacterial Protein Toxins**. Berlin, Springer Verlag; 2000.
- Althaus FR, Hilz H, Shall S: **ADP-ribosylation of proteins**. Berlin, Springer Verlag; 1985.
- Haag F, Koch-Nolte F: **ADP-Ribosylation in Animal Tissues: Structure, Function and Biology of Mono(ADP-Ribosyl)transferases and Related Enzymes**. Volume 419. New York, Plenum Press; 1997.
- Jacobson MK, Jacobson EL: **ADP-ribose Transfer Reactions: Mechanisms and Biological Significance**. New York, Springer Verlag; 1989.
- Honjo T, Nishizuka Y, Hayaishi O: **Diphtheria toxin-dependent adenosine diphosphate ribosylation of aminoacyl transferase II and inhibition of protein synthesis**. *J Biol Chem* 1968, **243**:3553-3555.
- Domenighini M, Rappuoli R: **Three conserved consensus sequences identify the NAD-binding site of ADP-ribosylating enzymes, expressed by eukaryotes, bacteria and T-even bacteriophages**. *Mol Microbiol* 1996, **21**:667-674.
- Bazan JF, Koch-Nolte F: **Sequence and structural links between distant ADP-ribosyltransferase families**. *Adv Exp Med Biol* 1997, **419**:99-107.
- Seman M, Adriouch S, Haag F, Koch-Nolte F: **Ecto-ADP-ribosyltransferases (ARTs): emerging actors in cell communication and signaling**. *Curr Med Chem* 2004, **11**:857-872.
- Ziegler M, Oei SL: **A cellular survival switch: poly(ADP-ribosylation) stimulates DNA repair and silences transcription**. *Bioessays* 2001, **23**:543-548.
- Smith S: **The world according to PARP**. *Trends Biochem Sci* 2001, **26**:174-179.
- Ame JC, Spenehauer C, de Murcia G: **The PARP superfamily**. *Bioessays* 2004, **26**:882-893.
- Meyer-Ficca ML, Meyer RG, Jacobson EL, Jacobson MK: **Poly(ADP-ribose) polymerases: managing genome stability**. *Int J Biochem Cell Biol* 2005, **37**:920-926.
- Ritter H, Koch-Nolte F, Marquez VE, Schulz GE: **Substrate binding and catalysis of ecto-ADP-ribosyltransferase 2.2 from rat**. *Biochemistry* 2003, **42**:10155-10162.
- Ruf A, Rolli V, de Murcia G, Schulz GE: **The mechanism of the elongation and branching reaction of poly(ADP-ribose) polymerase as derived from crystal structures and mutagenesis**. *J Mol Biol* 1998, **278**:57-65.
- Bell CE, Eisenberg D: **Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide**. *Biochemistry* 1996, **35**:1137-1149.
- Han S, Craig JA, Putnam CD, Carozzi NB, Tainer JA: **Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex**. *Nat Struct Biol* 1999, **6**:932-936.
- 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**:456-464.
- Menetrey J, Flatau G, Stura EA, Charbonnier JB, Gas F, Teulon JM, Le Du MH, Boquet P, Menez A: **NAD binding induces conformational changes in Rho ADP-ribosylating clostridium botulinum C3 exoenzyme**. *J Biol Chem* 2002, **277**:30950-30957.
- Li M, Dyda F, Benhar I, Pastan I, Davies DR: **The crystal structure of Pseudomonas aeruginosa exotoxin domain III with nicotinamide and AMP: conformational differences with the intact exotoxin**. *Proc Natl Acad Sci U S A* 1995, **92**:9308-9312.
- Carroll SF, Collier RJ: **NAD binding site of diphtheria toxin: identification of a residue within the nicotinamide subsite by photochemical modification with NAD**. *Proc Natl Acad Sci U S A* 1984, **81**:3307-3311.
- 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**:3247-3254.
- Pannifer AD, Wong TY, Schwarzenbacher R, Renatus M, Petosa C, Bienkowska J, Lacy DB, Collier RJ, Park S, Leppla SH, Hanna P, Lidington RC: **Crystal structure of the anthrax lethal factor**. *Nature* 2002, **414**:229-233.
- Tsuge H, Nagahama M, Nishimura H, Hisatsune J, Sakaguchi Y, Ito-gawa Y, Katunuma N, Sakurai J: **Crystal structure and site-directed mutagenesis of enzymatic components from Clostridium perfringens iota-toxin**. *J Mol Biol* 2003, **325**:471-483.
- Kato-Murayama M, Bessho Y, Shirouzu M, Yokoyama S: **Crystal structure of the RNA 2'-phosphotransferase from Aeropyrum pernix K1**. *J Mol Biol* 2005, **348**:295-305.
- Spinelli SL, Kierzek R, Turner DH, Phizicky EM: **Transient ADP-ribosylation of a 2'-phosphate implicated in its removal from ligated tRNA during splicing in yeast**. *J Biol Chem* 1999, **274**:2637-2644.
- Otto H, Tezcan-Merdol D, Girisch R, Haag F, Rhen M, Koch-Nolte F: **The spvB gene-product of the Salmonella enterica virulence plasmid is a mono(ADP-ribosyl)transferase**. *Mol Microbiol* 2000, **37**:1106-1115.
- Glowacki G, Braren R, Firner K, Nissen M, Kuhl M, Reche P, Bazan F, Cetkovic-Cvrlje M, Leiter E, Haag F, Koch-Nolte F: **The family of toxin-related ecto-ADP-ribosyltransferases in humans and the mouse**. *Protein Sci* 2002, **11**:1657-1670.
- Lander ES, Linton LM, Birren B, Nussbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N,

- Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubinfeld M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramseser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korfi I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowski J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
29. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley R, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Grauss TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korfi I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrum J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachtter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Raymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhauser AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.
30. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celera, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Foster C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fechtel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, De Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramson S, Nierman WC, Havlak PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodwark C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar Alba M, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelmayer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman MI, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyras E, Searle SM, Cooper GM, Batzoglou S, Brudno M, Sidow A, Stone EA, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachtter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution.** *Nature* 2004, **428**:493-521.
31. Takeyama K, Aguiar RC, Gu L, He C, Freeman GJ, Kutok JL, Aster JC, Shipp MA: **The BAL-binding protein BBAP and related Deltex family members exhibit ubiquitin-protein isopeptide ligase activity.** *J Biol Chem* 2003, **278**:21930-21937.
32. 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-Firzloff J, Luscher B: **PARP-10, a novel Myc-interacting protein with poly(ADP-ribose) polymerase activity, inhibits transformation.** *Oncogene* 2005.
33. Gao G, Guo X, Goff SP: **Inhibition of retroviral RNA production by ZAP, a CCHC-type zinc finger protein.** *Science* 2002, **297**:1703-1706.
34. Ma Q, Baldwin KT, Renzelli AJ, McDaniel A, Dong L: **TCDD-inducible poly(ADP-ribose) polymerase: a novel response to 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *Biochem Biophys Res Commun* 2001, **289**:499-506.
35. Altschul SF, Koonin EV: **Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases.** *Trends Biochem Sci* 1998, **23**:444-447.
36. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302**:205-217.
37. McGuffin LJ, Bryson K, Jones DT: **The PSIPRED protein structure prediction server.** *Bioinformatics* 2000, **16**:404-405.
38. Koch-Nolte F, Reche P, Haag F, Bazzan F: **ADP-ribosyltransferases: plastic tools for inactivating protein and small molecular weight targets.** *J Biotechnol* 2001, **92**:81-87.
39. Han S, Tainer JA: **The ARTT motif and a unified structural understanding of substrate recognition in ADP-ribosylating bacterial toxins and eukaryotic ADP-ribosyltransferases.** *Int J Med Microbiol* 2002, **291**:523-529.
40. Sun J, Maresso AV, Kim JJ, Barbieri JT: **How bacterial ADP-ribosylating toxins recognize substrates.** *Nat Struct Mol Biol* 2004, **11**:868-876.
41. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L:

- Database resources of the National Center for Biotechnology.** *Nucleic Acids Res* 2003, **31**:28-33.
42. Ladurner AG: **Inactivating chromosomes: a macro domain that minimizes transcription.** *Mol Cell* 2003, **12**:1-3.
  43. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthonard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H: **Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype.** *Nature* 2004, **431**:946-957.
  44. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MA, Delany ME, Dodgson JB, Chinwalla AT, Clifton PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J, Smith SM, Wallis JW, Yang SP, Romanov MN, Rondelli CM, Paton B, Smith J, Morrice D, Daniels L, Tempest HG, Robertson L, Masabanda JS, Griffin DK, Vignal A, Fillon V, Jacobsson L, Kerje S, Andersson L, Crooijmans RP, Aerts J, van der Poel JJ, Ellegren H, Caldwell RB, Hubbard SJ, Grafham DV, Kierzek AM, McLaren SR, Overton IM, Arakawa H, Beattie KJ, Bezzubov Y, Boardman PE, Bonfield JK, Croning MD, Davies RM, Francis MD, Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown WR, Rogers J, Buerstedde JM, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong GK, Wang J, Liu B, Yu J, Yang H, Nefedov M, Koriabine M, Dejong PJ, Goodstadt L, Webber C, Dickens NJ, Letunic I, Suyama M, Torrents D, von Mering C, Zdobnov EM, Makova K, Nekrutenko A, Elnitski L, Eswara P, King DC, Yang S, Tyekucheva S, Radakrishnan A, Harris RS, Chiaromonte F, Taylor J, He J, Rijnkels M, Griffiths-Jones S, Ureta-Vidal A, Hoffman MM, Severin J, Searle SM, Law AS, Speed D, Waddington D, Cheng Z, Tuzun E, Eichler E, Bao Z, Flicek P, Shteynberg DD, Brent MR, Bye JM, Huckle EJ, Chatterji S, Dewey C, Pachter L, Kouranov A, Mourelatos Z, Hatzigeorgiou AG, Paterson AH, Ivarie R, Brandstrom M, Axelsson E, Backstrom N, Berlin S, Webster MT, Pourquie O, Reymond A, Ucla C, Antonarakis SE, Long M, Emerson JJ, Betran E, Dupanloup I, Kaessmann H, Hinrichs AS, Bejerano G, Furey TS, Harte RA, Raney B, Siepel A, Kent WJ, Haussler D, Eyras E, Castelo R, Abril JF, Castellano S, Camara F, Parra G, Guigo R, Bourque G, Tesler G, Pevzner PA, Smit A, Fulton LA, Mardis ER, Wilson RK: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432**:695-716.
  45. Seimiya H, Smith S: **The telomeric poly(ADP-ribose) polymerase, tankyrase 1, contains multiple binding sites for telomeric repeat binding factor 1 (TRF1) and a novel acceptor, 182-kDa tankyrase-binding protein (TAB182).** *J Biol Chem* 2002, **277**:14116-14126.
  46. 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**:917-928.
  47. Chang W, Dynek JN, Smith S: **TRF1 is degraded by ubiquitin-mediated proteolysis after release from telomeres.** *Genes Dev* 2003, **17**:1328-1333.
  48. Jakob NJ, Darai G: **Molecular anatomy of chilo iridescent virus genome and the evolution of viral genes.** *Virus Genes* 2002, **25**:299-316.
  49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
  50. Rebhan M, Chalifa-Caspi V, Prilusky JLD: **GeneCards: encyclopedia for genes, proteins and diseases.** [<http://bioinformatics.weizmann.ac.il/cards>].
  51. DeLano WL: **The PyMOL User's Manual.** 2002 [<http://www.pymol.org>]. San Carlos, CA, USA., DeLano Scientific
  52. Swofford DL: **PAUP: Phylogenetic Analysis Using Parsimony (and other methods) version 4.** Sunderland, Massachusetts, Sinauer Associates Inc.; 2002.
  53. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
  54. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
  55. Drummond A, Strimmer K: **PAL: an object-oriented programming library for molecular evolution and phylogenetics.** *Bioinformatics* 2001, **17**:662-663.
  56. Abascal F, Zardoya R, Posada D: **Protest: selection of best-fit models of protein evolution.** *Bioinformatics* 2005.
  57. Huelsenbeck JP, Bollback JP: **Empirical and hierarchical Bayesian estimation of ancestral states.** *Syst Biol* 2001, **50**:351-366.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

