Data in Brief 8 (2016) 142–152

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on four apoptosis-related genes in the colonial tunicate *Botryllus schlosseri*

Nicola Franchi, Francesca Ballin, Lucia Manni*,
Filippo Schiavon, Loriano Ballarin

Department of Biology, University of Padova, Italy

A R T I C L E   I N F O

Article history:
Received 13 April 2016
Received in revised form
3 May 2016
Accepted 11 May 2016
Available online 20 May 2016

Keywords:
Sequence databases
Alignments
BLAST analysis
Botryllus schlosseri
Apoptosis-related transcripts
Bax
AIF1
PARP1
IAP7

ABSTRACT

The data described are related to the article entitled "Recurrent phagocytosis-induced apoptosis in the cyclical generation change of the compound ascidian *Botryllus schlosseri*" (Franchi et al., 2016) [1]. Four apoptosis-related genes, showing high similarity with mammalian Bax (a member of the Bcl-2 protein family), AIF1 (apoptosis-inducing factor-1), PARP1 (poly ADP ribose polymerase-1) and IAP7 (inhibitor of apoptosis-7) were identified from the analysis of the trascriptome of *B. schlosseri*. They were named BsBax, BsAIF1, BsPARP1 and BsIAP7. Here, their deduced amino acid sequence were compared with known sequences of orthologous genes from other deuterostome species together with a study of their identity/similarity.

Specifications Table

Subject area	*Biology*
More specific subject area	*Developmental Biology*
Type of data	*Tables, figures*

DOI of original article: http://dx.doi.org/10.1016/j.dci.2016.04.011

* Corresponding author at: Department of Biology, University of Padova via U. Bassi 58/B 35131 Padova Italy. Tel.: +39 049 8276252; fax: +39 049 8276199.

E-mail address: lucia.manni@unipd.it (L. Manni).

http://dx.doi.org/10.1016/j.dib.2016.05.017

How data was acquired	*Bioinformatic analysis, RACE*
Data format	*Raw and analysed data*
Experimental factors	*The partial transcripts present in the transcriptome, identified by BLAST analysis, were elongated through 5'and 3' RACE according to the 2nd generation of 5'/3' RACE kit*
Experimental features	*Analysis with BLAST, LALIGN, SMART, Clustal Omega*
Data source location	*Padova, Italy*
Data accessibility	*Data are available in this article and at GenBank via accession numbers GenBank: KU948200 for BsBAX, GenBank: KU948201 for BsPARP1, GenBank: KU948202 for BsAIF1, GenBank: KU948203 for BsIAP7.*

Value of the data

- The data provide the full-length sequences of four apoptosis-related transcripts from the colonial ascidian *B. schlosseri* useful to study the phylogeny trees of the corresponding proteins in chordates.
- From the data, the protein primary structures can be deduced and, from that, three-dimensional models can be obtained, useful to compare the domain organization of the corresponding chordate proteins.
- Expression studies, exploiting the present data, can contribute to elucidate the dynamics of the cyclical apoptosis, which characterizes the colonial blastogenetic cycle of the ascidian *B. schlosseri*.

1. Data

The data reported include supporting information to the phylogenetic analyses of Franchi et al. [1]. They consist of transcript sequences, sequence alignments and comparisons of four apoptosis-related genes identified in the recently-obtained transcriptome of *B. schlosseri* [2]. The sequences show high similarity with mammalian transcripts for Bax, AIF1, PARP1 and IAP7 and were named BsBax, BsAIF1, BsPARP1 and BsIAP7, respectively. The expression of these genes was studied further in the above-reported paper [1].

2. Experimental design, materials and methods

Amplification and cloning of transcripts for BsBax, BsAIF1, BsPARP1 and BsIAP7 was achieved with specific primers designed on sequences found in our collection of transcriptomes [2]. In order to verify and complete the full length cDNA, PCR reactions were carried out with a denaturing step at 94 °C for 2 min, 40 cycles of 30 s at 94 °C, 40 s at 60 °C and 90 s at 72 °C, and a final extension at 72 °C for 10 min. Amplicons were separated using 1.5% agarose gel, purified, cloned and sequenced. The partial transcripts were elongated through 5′ and 3′ RACE according to the 2nd generation of 5′/3′ RACE kit (Roche). Supplementary Table 1 reports the specific primers used for amplicons production and their elongation through 5′- and 3′-RACE and for the *in situ* hybridisation experiments reported in [1].

The sequences from GenBank, reported in Supplementary Tables 2–5, were used for alignments and sequence comparisons with the sequences of BsBax, BsAIF1, BsPARP1, BsIAP7, respectively. The latter were deposited in GenBank and corresponding to the accession numbers GenBank: KU948200, GenBank: KU948201, GenBank: KU948202 and GenBank: KU948203, respectively (Figs. 1–8).

The predicted amino acid sequences of BsBax, BsAIF1, BsPARP1, BsIAP7 are reported in Figs. 2, 4, 6 and 8, respectively. They were aligned with known orthologous sequences from both vertebrate and invertebrates using the MUSCLE programme [3] and data are reported in Figs. 1, 3, 5, and 7, respectively. Identity analysis of the deduced amino acid sequences were performed using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and LALIGN (http://www.ch.embnet.org/software/LALIGN_form.html) [4] and are reported in Supplementary Tables 2–5.

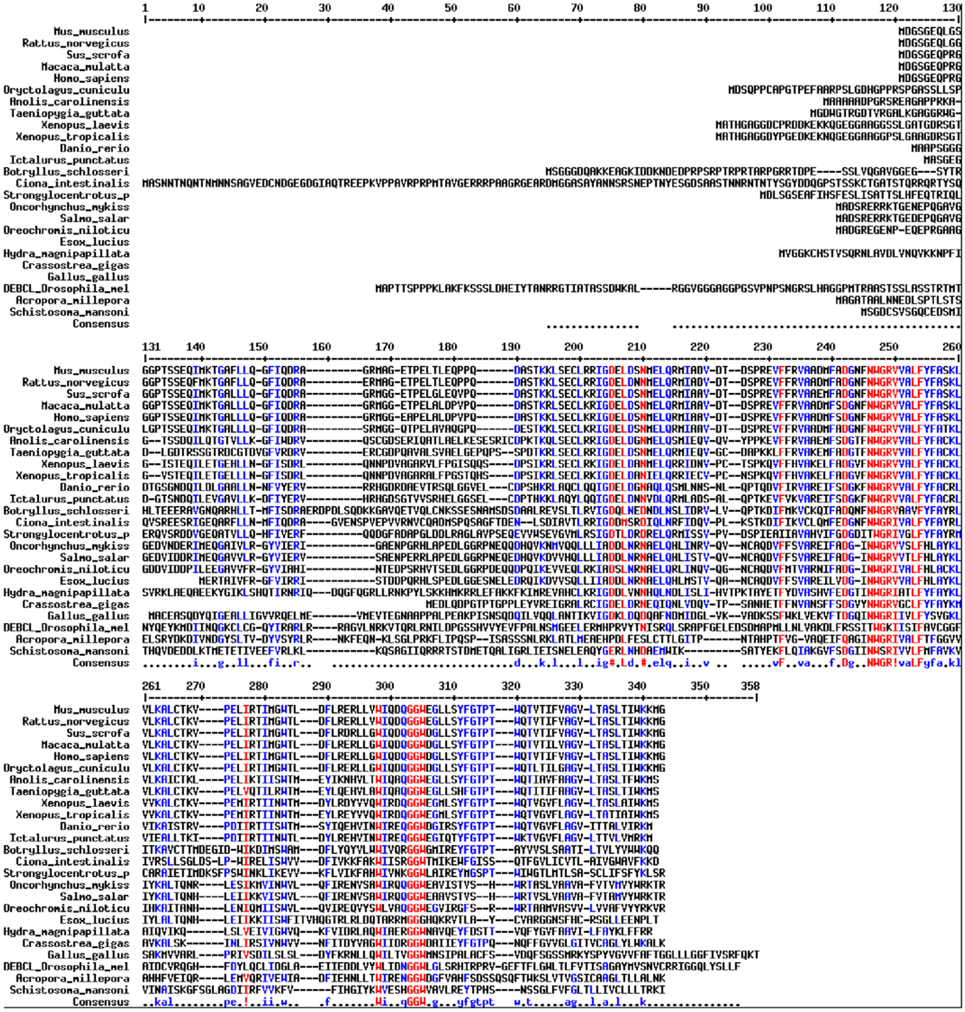
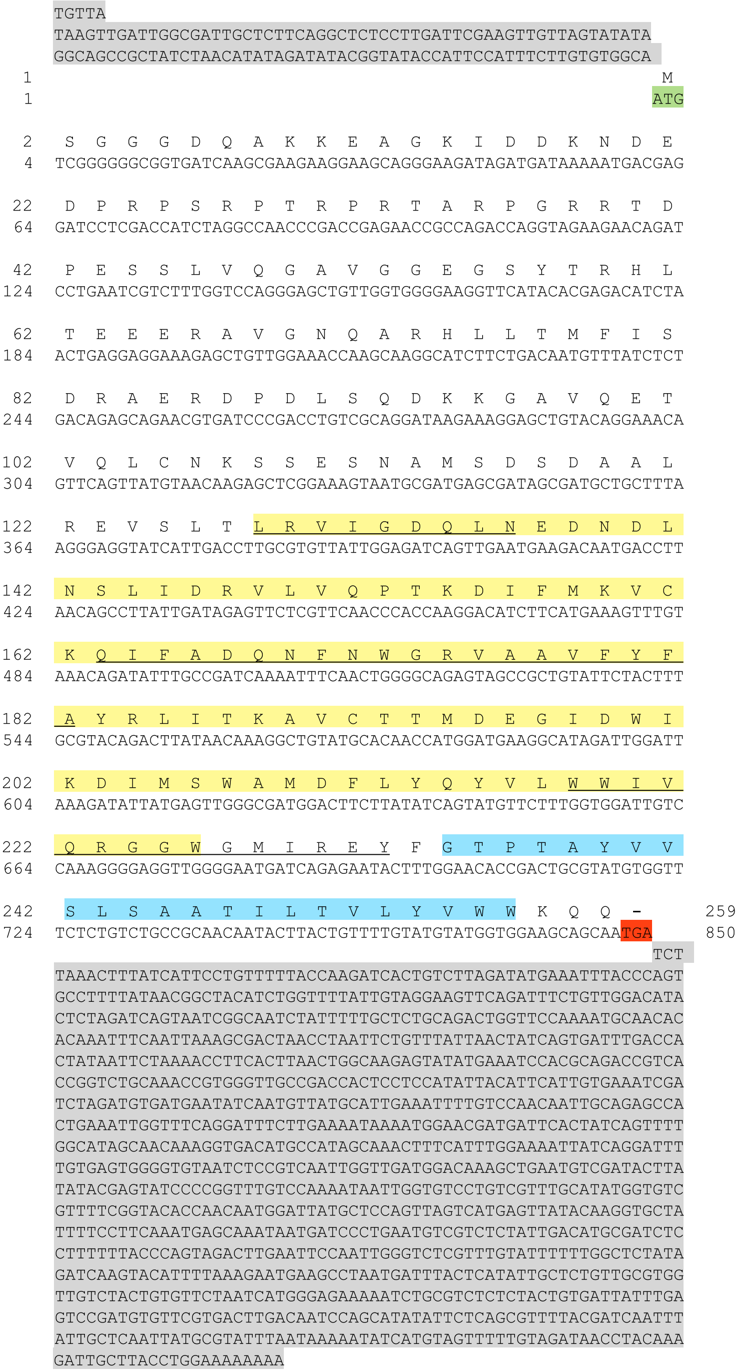


Fig. 1. Alignments of the deduced amino acid sequence of BsBax with known orthologous sequences from both vertebrate and invertebrates.



```
      TGTTA
      TAAGTTGATTGGCGATTGCTCTTCAGGCTCTCCTTGATTCGAAGTTGTTAGTATATA
      GGCAGCCGCTATCTAACATATAGATATACGGTATACCATTCCATTTCTTGTGTGGCA
  1                                                           M
  1                                                          ATG

  2    S  G  G  G  D  Q  A  K  K  E  A  G  K  I  D  D  K  N  D  E
  4   TCGGGGGGCGGTGATCAAGCGAAGAAGGAAGCAGGGAAGATAGATGATAAAAATGACGAG

 22    D  P  R  P  S  R  P  T  R  P  R  T  A  R  P  G  R  R  T  D
 64   GATCCTCGACCATCTAGGCCAACCCGACCGAGAACCGCCAGACCAGGTAGAAGAACAGAT

 42    P  E  S  S  L  V  Q  G  A  V  G  G  E  G  S  Y  T  R  H  L
124   CCTGAATCGTCTTTGGTCCAGGGAGCTGTTGGTGGGGAAGGTTCATACACGAGACATCTA

 62    T  E  E  E  R  A  V  G  N  Q  A  R  H  L  L  T  M  F  I  S
184   ACTGAGGAGGAAAGAGCTGTTGGAAACCAAGCAAGGCATCTTCTGACAATGTTTATCTCT

 82    D  R  A  E  R  D  P  D  L  S  Q  D  K  K  G  A  V  Q  E  T
244   GACAGAGCAGAACGTGATCCCGACCTGTCGCAGGATAAGAAAGGAGCTGTACAGGAAACA

102    V  Q  L  C  N  K  S  S  E  S  N  A  M  S  D  S  D  A  A  L
304   GTTCAGTTATGTAACAAGAGCTCGGAAAGTAATGCGATGAGCGATAGCGATGCTGCTTTA

122    R  E  V  S  L  T  L  R  V  I  G  D  Q  L  N  E  D  N  D  L
364   AGGGAGGTATCATTGACCTTGCGTGTTATTGGAGATCAGTTGAATGAAGACAATGACCTT

142    N  S  L  I  D  R  V  L  V  Q  P  T  K  D  I  F  M  K  V  C
424   AACAGCCTTATTGATAGAGTTCTCGTTCAACCCACCAAGGACATCTTCATGAAAGTTTGT

162    K  Q  I  F  A  D  Q  N  F  N  W  G  R  V  A  A  V  F  Y  F
484   AAACAGATATTTGCCGATCAAAATTTCAACTGGGGCAGAGTAGCCGCTGTATTCTACTTT

182    A  Y  R  L  I  T  K  A  V  C  T  T  M  D  E  G  I  D  W  I
544   GCGTACAGACTTATAACAAAGGCTGTATGCACAACCATGGATGAAGGCATAGATTGGATT

202    K  D  I  M  S  W  A  M  D  F  L  Y  Q  Y  V  L  W  W  I  V
604   AAAGATATTATGAGTTGGGCGATGGACTTCTTATATCAGTATGTTCTTTGGTGGATTGTC

222    Q  R  G  G  W  G  M  I  R  E  Y  F  G  T  P  T  A  Y  V  V
664   CAAAGGGGAGGTTGGGGAATGATCAGAGAATACTTTGGAACACCGACTGCGTATGTGGTT

242    S  L  S  A  A  T  I  L  T  V  L  Y  V  W  W  K  Q  Q  -      259
724   TCTCTGTCTGCCGCAACAATACTTACTGTTTTGTATGTATGGTGGAAGCAGCAATGA      850
                                                                 TCT
      TAAACTTTATCATTCCTGTTTTTACCAAGATCACTGTCTTAGATATGAAATTTACCCAGT
      GCCTTTTATAACGGCTACATCTGGTTTTATTGTAGGAAGTTCAGATTTCTGTTGGACATA
      CTCTAGATCAGTAATCGGCAATCTATTTTTGCTCTGCAGACTGGTTCCAAAATGCAACAC
      ACAAATTTCAATTAAAGCGACTAACCTAATTCTGTTTATTAACTATCAGTGATTTGACCA
      CTATAATTCTAAAACCTTCACTTAACTGGCAAGAGTATATGAAATCCACGCAGACCGTCA
      CCGGTCTGCAAACCGTGGGTTGCCGACCACTCCTCCATATTACATTCATTGTGAAATCGA
      TCTAGATGTGATGAATATCAATGTTATGCATTGAAATTTTGTCCAACAATTGCAGAGCCA
      CTGAAATTGGTTTCAGGATTTCTTGAAAATAAAATGGAACGATGATTCACTATCAGTTTT
      GGCATAGCAACAAAGGTGACATGCCATAGCAAACTTTCATTTGGAAAATTATCAGGATTT
      TGTGAGTGGGGTGTAATCTCCGTCAATTGGTTGATGGACAAAGCTGAATGTCGATACTTA
      TATACGAGTATCCCCGGTTTGTCCAAAATAATTGGTGTCCTGTCGTTTGCATATGGTGTC
      GTTTTCGGTACACCAACAATGGATTATGCTCCAGTTAGTCATGAGTTATACAAGGTGCTA
      TTTTCCTTCAAATGAGCAAATAATGATCCCTGAATGTCGTCTCTATTGACATGCGATCTC
      CTTTTTTACCCAGTAGACTTGAATTCCAATTGGGTCTCGTTTGTATTTTTTGGCTCTATA
      GATCAAGTACATTTTAAAGAATGAAGCCTAATGATTTACTCATATTGCTCTGTTGCGTGG
      TTGTCTACTGTGTTCTAATCATGGGAGAAAAATCTGCGTCTCTCTACTGTGATTATTTGA
      GTCCGATGTGTTCGTGACTTGACAATCCAGCATATATTCTCAGCGTTTTACGATCAATTT
      ATTGCTCAATTATGCGTATTTAATAAAAATATCATGTAGTTTTTGTAGATAACCTACAAA
      GATTGCTTACCTGGAAAAAAAA
```

Fig. 2. cDNA and deduced amino acid sequence of BsBax. START and STOP codon are in green and red, respectively; 5′ and 3′ UTR are in grey. In yellow, the Bcl-2 domain and, within this, BH1, BH2 and BH3 domains are underlined. Light blue: transmembrane domain.

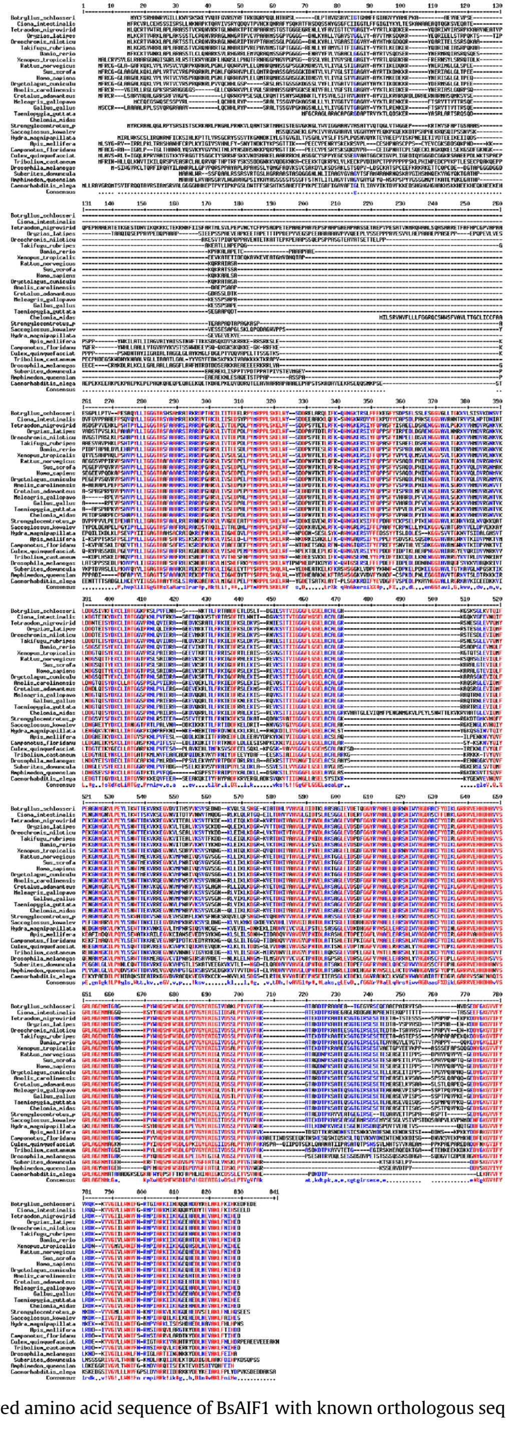
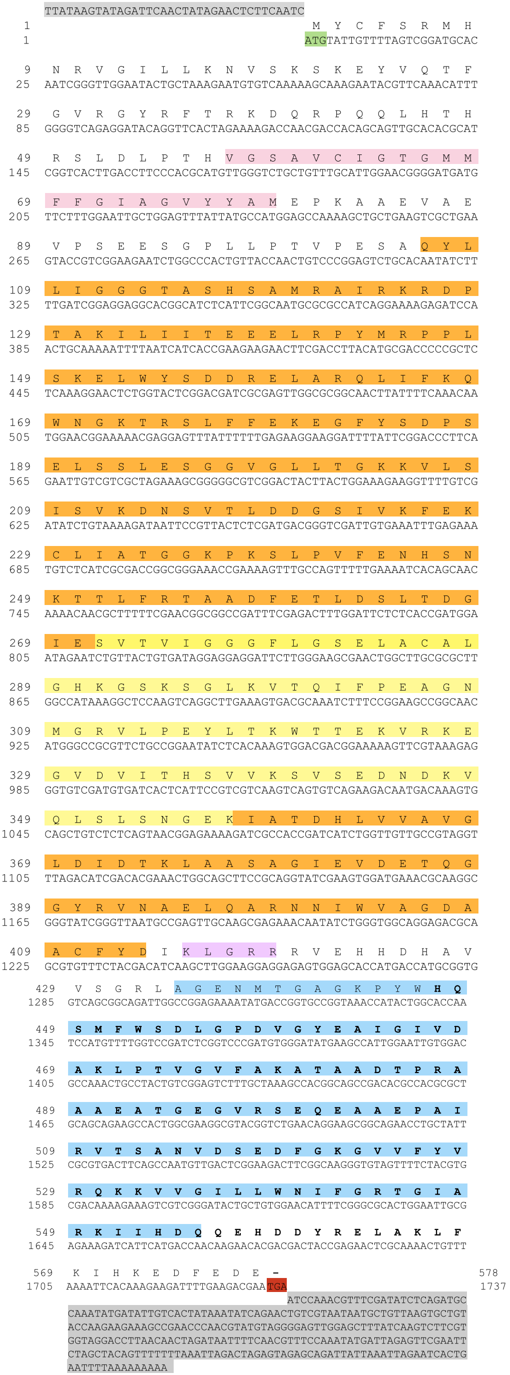


Fig. 3. Alignments of the deduced amino acid sequence of BsAIF1 with known orthologous sequences from both vertebrate and invertebrates.



```
     TTATAAGTATAGATTCAACTATAGAACTCTTCAATC
   1                                  M  Y  C  F  S  R  M  H
   1                                 ATGTATTGTTTTAGTCGGATGCAC

   9  N  R  V  G  I  L  L  K  N  V  S  K  S  K  E  Y  V  Q  T  F
  25 AATCGGGTTGGAATACTGCTAAAGAATGTGTCAAAAAGCAAAGAATACGTTCAAACATTT

  29  G  V  R  G  Y  R  F  T  R  K  D  Q  R  P  Q  Q  L  H  T  H
  85 GGGGTCAGAGGATACAGGTTCACTAGAAAAGACCAACGACCACAGCAGTTGCACACGCAT

  49  R  S  L  D  L  P  T  H  V  G  S  A  V  C  I  G  T  G  M  M
 145 CGGTCACTTGACCTTCCCACGCATGTTGGGTCTGCTGTTTGCATTGGAACGGGGATGATG

  69  F  F  G  I  A  G  V  Y  Y  A  M  E  P  K  A  A  E  V  A  E
 205 TTCTTTGGAATTGCTGGAGTTTATTATGCCATGGAGCCAAAAGCTGCTGAAGTCGCTGAA

  89  V  P  S  E  E  S  G  P  L  L  P  T  V  P  E  S  A  Q  Y  L
 265 GTACCGTCGGAAGAATCTGGCCCACTGTTACCAACTGTCCCGGAGTCTGCACAATATCTT

 109  L  I  G  G  G  T  A  S  H  S  A  M  R  A  I  R  K  R  D  P
 325 TTGATCGGAGGAGGCACGGCATCTCATTCGGCAATGCGCGCCATCAGGAAAAGAGATCCA

 129  T  A  K  I  L  I  I  T  E  E  E  L  R  P  Y  M  R  P  P  L
 385 ACTGCAAAAATTTTAATCATCACCGAAGAAGAACTTCGACCTTACATGCGACCCCCGCTC

 149  S  K  E  L  W  Y  S  D  D  R  E  L  A  R  Q  L  I  F  K  Q
 445 TCAAAGGAACTCTGGTACTCGGACGATCGCGAGTTGGCGCGGCAACTTATTTTCAAACAA

 169  W  N  G  K  T  R  S  L  F  F  E  K  E  G  F  Y  S  D  P  S
 505 TGGAACGGAAAAACGAGGAGTTTATTTTTTGAGAAGGAAGGATTTTATTCGGACCCTTCA

 189  E  L  S  S  L  E  S  G  G  V  G  L  L  T  G  K  K  V  L  S
 565 GAATTGTCGTCGCTAGAAAGCGGGGGCGTCGGACTACTTACTGGAAAGAAGGTTTTGTCG

 209  I  S  V  K  D  N  S  V  T  L  D  D  G  S  I  V  K  F  E  K
 625 ATATCTGTAAAAGATAATTCCGTTACTCTCGATGACGGGTCGATTGTGAAATTTGAGAAA

 229  C  L  I  A  T  G  G  K  P  K  S  L  P  V  F  E  N  H  S  N
 685 TGTCTCATCGCGACCGGCGGGAAACCGAAAAGTTTGCCAGTTTTTGAAAATCACAGCAAC

 249  K  T  T  L  F  R  T  A  A  D  F  E  T  L  D  S  L  T  D  G
 745 AAAACAACGCTTTTTCGAACGGCGGCCGATTTCGAGACTTTGGATTCTCTCACCGATGGA

 269  I  E  S  V  T  V  I  G  G  G  F  L  G  S  E  L  A  C  A  L
 805 ATAGAATCTGTTACTGTGATAGGAGGAGGATTCTTGGGAAGCGAACTGGCTTGCGCGCTT

 289  G  H  K  G  S  K  S  G  L  K  V  T  Q  I  F  P  E  A  G  N
 865 GGCCATAAAGGCTCCAAGTCAGGCTTGAAAGTGACGCAAATCTTTCCGGAAGCCGGCAAC

 309  M  G  R  V  L  P  E  Y  L  T  K  W  T  T  E  K  V  R  K  E
 925 ATGGGCCGCGTTCTGCCGGAATATCTCACAAAGTGGACGACGGAAAAAGTTCGTAAAGAG

 329  G  V  D  V  I  T  H  S  V  V  K  S  V  S  E  D  N  D  K  V
 985 GGTGTCGATGTGATCACTCATTCCGTCGTCAAGTCAGTGTCAGAAGACAATGACAAAGTG

 349  Q  L  S  L  S  N  G  E  K  I  A  T  D  H  L  V  V  A  V  G
1045 CAGCTGTCTCTCAGTAACGGAGAAAAGATCGCCACCGATCATCTGGTTGTTGCCGTAGGT

 369  L  D  I  D  T  K  L  A  A  S  A  G  I  E  V  D  E  T  Q  G
1105 TTAGACATCGACACGAAACTGGCAGCTTCCGCAGGTATCGAAGTGGATGAAACGCAAGGC

 389  G  Y  R  V  N  A  E  L  Q  A  R  N  N  I  W  V  A  G  D  A
1165 GGGTATCGGGTTAATGCCGAGTTGCAAGCGAGAAACAATATCTGGGTGGCAGGAGACGCA

 409  A  C  F  Y  D  I  K  L  G  R  R  R  V  E  H  H  D  H  A  V
1225 GCGTGTTTCTACGACATCAAGCTTGGAAGGAGGAGAGTGGAGCACCATGACCATGCGGTG
   429  V  S  G  R  L  A  G  E  N  M  T  G  A  G  K  P  Y  W  H  Q
  1285 GTCAGCGGCAGATTGGCCGGAGAAAATATGACCGGTGCCGGTAAACCATACTGGCACCAA

   449  S  M  F  W  S  D  L  G  P  D  V  G  Y  E  A  I  G  I  V  D
  1345 TCCATGTTTTGGTCCGATCTCGGTCCCGATGTGGGATATGAAGCCATTGGAATTGTGGAC

   469  A  K  L  P  T  V  G  V  F  A  K  A  T  A  A  D  T  P  R  A
  1405 GCCAAACTGCCTACTGTCGGAGTCTTTGCTAAAGCCACGGCAGCCGACACGCCACGCGCT

   489  A  A  E  A  T  G  E  G  V  R  S  E  Q  E  A  A  E  P  A  I
  1465 GCAGCAGAAGCCACTGGCGAAGGCGTACGGTCTGAACAGGAAGCGGCAGAACCTGCTATT

   509  R  V  T  S  A  N  V  D  S  E  D  F  G  K  G  V  V  F  Y  V
  1525 CGCGTGACTTCAGCCAATGTTGACTCGGAAGACTTCGGCAAGGGTGTAGTTTTCTACGTG

   529  R  Q  K  K  V  V  G  I  L  L  W  N  I  F  G  R  T  G  I  A
  1585 CGACAAAAGAAAGTCGTCGGGATACTGCTGTGGAACATTTTCGGGCGCACTGGAATTGCG

   549  R  K  I  I  H  D  Q  Q  E  H  D  D  Y  R  E  L  A  K  L  F
  1645 AGAAAGATCATTCATGACCAACAAGAACACGACGACTACCGAGAACTCGCAAAACTGTTT

   569  K  I  H  K  E  D  F  E  D  E  -                              578
  1705 AAAATTCACAAAGAAGATTTTGAAGACGAATGA                            1737
                                       ATCCAAACGTTTCGATATCTCAGATGC
       CAAATATGATATTGTCACTATAAATATCAGAACTGTCGTAATAATGCTGTTAAGTGCTGT
       ACCAAGAAGAAAGCCGAACCCAACGTATGTAGGGGAGTTGGAGCTTTATCAAGTCTTCGT
       GGTAGGACCTTAACAACTAGATAATTTTCAACGTTTCCAAATATGATTAGAGTTCGAATT
       CTAGCTACAGTTTTTTTAAATTAGACTAGAGTAGAGCAGATTATTAAATTAGAATCACTG
       AATTTTAAAAAAAAA
```

Fig. 4. cDNA and deduced amino acid sequence of BsAIF1. START and STOP codon are in green and red, respectively; 5′ and 3′ UTR are in grey. Orange: pyridine nucleotide-disulphide domain (PNDD); yellow: NADH-binding PNDD; light blue: C-terminal domain of dimerisation; light purple: nuclear localisation motif. Bold: the N-terminal transmembrane domain.

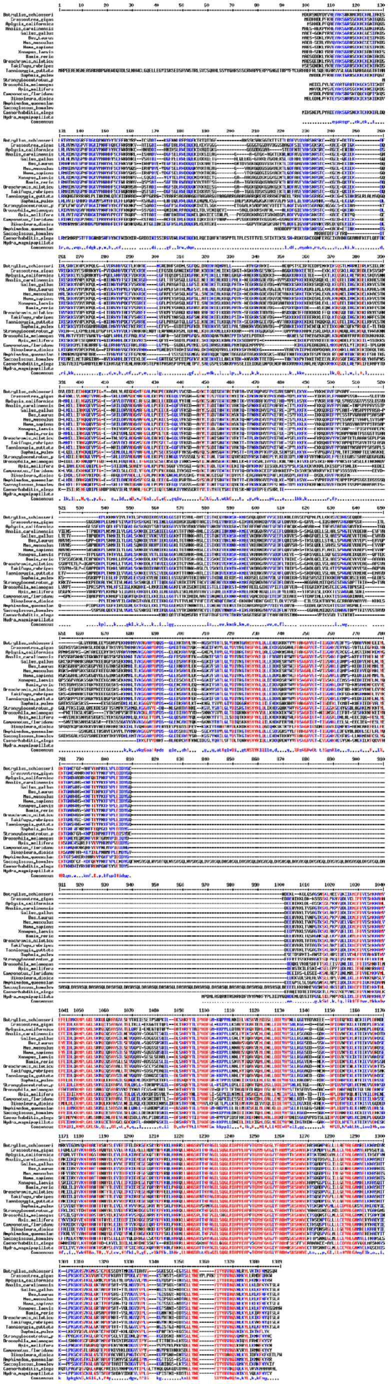
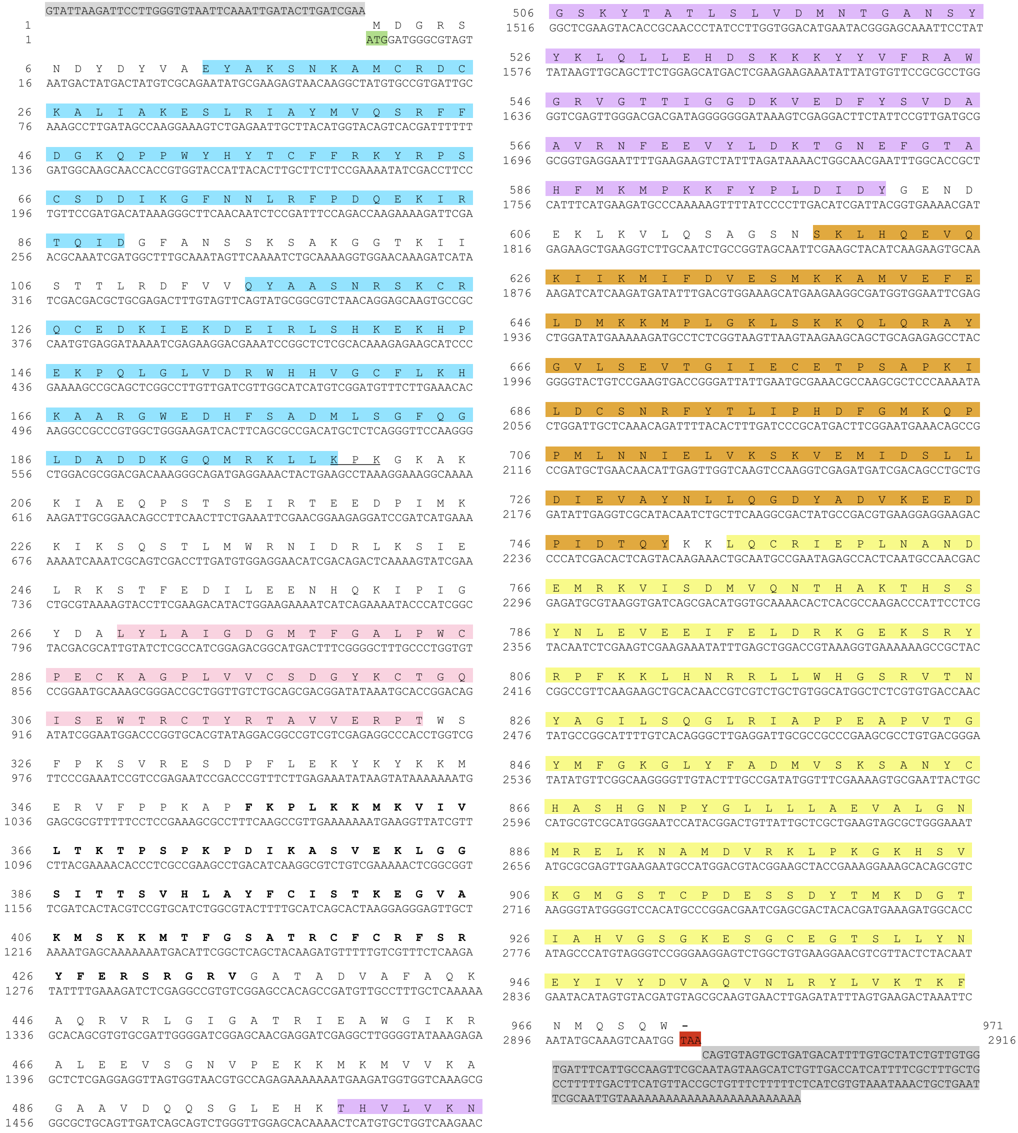


Fig. 5. Alignments of the deduced amino acid sequence of BsPARP1 with known orthologous sequences from both vertebrate and invertebrates.



```
      GTATTAAGATTCCTTGGGTGTAATTCAAATTGATACTTGATCGAA
   1                                                  M  D  G  R  S
   1                                                 ATGGATGGGCGTAGT

   6   N  D  Y  D  Y  V  A  E  Y  A  K  S  N  K  A  M  C  R  D  C
  16  AATGACTATGACTATGTCGCAGAATATGCGAAGAGTAACAAGGCTATGTGCCGTGATTGC

  26   K  A  L  I  A  K  E  S  L  R  I  A  Y  M  V  Q  S  R  F  F
  76  AAAGCCTTGATAGCCAAGGAAAGTCTGAGAATTGCTTACATGGTACAGTCACGATTTTTT

  46   D  G  K  Q  P  P  W  Y  H  Y  T  C  F  F  R  K  Y  R  P  S
 136  GATGGCAAGCAACCACCGTGGTACCATTACACTTGCTTCTTCCGAAAATATCGACCTTCC

  66   C  S  D  D  I  K  G  F  N  N  L  R  F  P  D  Q  E  K  I  R
 196  TGTTCCGATGACATAAAGGGCTTCAACAATCTCCGATTTCCAGACCAAGAAAAGATTCGA

  86   T  Q  I  D  G  F  A  N  S  S  K  S  A  K  G  G  T  K  I  I
 256  ACGCAAATCGATGGCTTTGCAAATAGTTCAAAATCTGCAAAAGGTGGAACAAAGATCATA

 106   S  T  T  L  R  D  F  V  V  Q  Y  A  A  S  N  R  S  K  C  R
 316  TCGACGACGCTGCGAGACTTTGTAGTTCAGTATGCGGCGTCTAACAGGAGCAAGTGCCGC

 126   Q  C  E  D  K  I  E  K  D  E  I  R  L  S  H  K  E  K  H  P
 376  CAATGTGAGGATAAAATCGAGAAGGACGAAATCCGGCTCTCGCACAAAGAGAAGCATCCC

 146   E  K  P  Q  L  G  L  V  D  R  W  H  H  V  G  C  F  L  K  H
 436  GAAAAGCCGCAGCTCGGCCTTGTTGATCGTTGGCATCATGTCGGATGTTTCTTGAAACAC

 166   K  A  A  R  G  W  E  D  H  F  S  A  D  M  L  S  G  F  Q  G
 496  AAGGCCGCCCGTGGCTGGGAAGATCACTTCAGCGCCGACATGCTCTCAGGGTTCCAAGGG

 186   L  D  A  D  D  K  G  Q  M  R  K  L  L  K  P  K  G  K  A  K
 556  CTGGACGCGGACGACAAAGGGCAGATGAGGAAACTACTGAAGCCTAAAGGAAAGGCAAAA

 206   K  I  A  E  Q  P  S  T  S  E  I  R  T  E  E  D  P  I  M  K
 616  AAGATTGCGGAACAGCCTTCAACTTCTGAAATTCGAACGGAAGAGGATCCGATCATGAAA

 226   K  I  K  S  Q  S  T  L  M  W  R  N  I  D  R  L  K  S  I  E
 676  AAAATCAAATCGCAGTCGACCTTGATGTGGAGGAACATCGACAGACTCAAAAGTATCGAA

 246   L  R  K  S  T  F  E  D  I  L  E  E  N  H  Q  K  I  P  I  G
 736  CTGCGTAAAAGTACCTTCGAAGACATACTGGAAGAAAATCATCAGAAAATACCCATCGGC

 266   Y  D  A  L  Y  L  A  I  G  D  G  M  T  F  G  A  L  P  W  C
 796  TACGACGCATTGTATCTCGCCATCGGAGACGGCATGACTTTCGGGGCTTTGCCCTGGTGT

 286   P  E  C  K  A  G  P  L  V  V  C  S  D  G  Y  K  C  T  G  Q
 856  CCGGAATGCAAAGCGGGACCGCTGGTTGTCTGCAGCGACGGATATAAATGCACCGGACAG

 306   I  S  E  W  T  R  C  T  Y  R  T  A  V  V  E  R  P  T  W  S
 916  ATATCGGAATGGACCCGGTGCACGTATAGGACGGCCGTCGTCGAGAGGCCCACCTGGTCG

 326   F  P  K  S  V  R  E  S  D  P  F  L  E  K  Y  K  Y  K  K  M
 976  TTCCCGAAATCCGTCCGAGAATCCGACCCGTTTCTTGAGAAATATAAGTATAAAAAAATG

 346   E  R  V  F  P  P  K  A  P  F  K  P  L  K  K  M  K  V  I  V
1036  GAGCGCGTTTTTCCTCCGAAAGCGCCTTTCAAGCCGTTGAAAAAAATGAAGGTTATCGTT

 366   L  T  K  T  P  S  P  K  P  D  I  K  A  S  V  E  K  L  G  G
1096  CTTACGAAAACACCCTCGCCGAAGCCTGACATCAAGGCGTCTGTCGAAAAACTCGGCGGT

 386   S  I  T  T  S  V  H  L  A  Y  F  C  I  S  T  K  E  G  V  A
1156  TCGATCACTACGTCCGTGCATCTGGCGTACTTTTGCATCAGCACTAAGGAGGGAGTTGCT

 406   K  M  S  K  K  M  T  F  G  S  A  T  R  C  F  C  R  F  S  R
1216  AAAATGAGCAAAAAAATGACATTCGGCTCAGCTACAAGATGTTTTTGTCGTTTCTCAAGA

 426   Y  F  E  R  S  R  G  R  V  G  A  T  A  D  V  A  F  A  Q  K
1276  TATTTTGAAAGATCTCGAGGCCGTGTCGGAGCCACAGCCGATGTTGCCTTTGCTCAAAAA

 446   A  Q  R  V  R  L  G  I  G  A  T  R  I  E  A  W  G  I  K  R
1336  GCACAGCGTGTGCGATTGGGGATCGGAGCAACGAGGATCGAGGCTTGGGGTATAAAGAGA

 466   A  L  E  E  V  S  G  N  V  P  E  K  K  M  K  M  V  V  K  A
1396  GCTCTCGAGGAGGTTAGTGGTAACGTGCCAGAGAAAAAAATGAAGATGGTGGTCAAAGCG

 486   G  A  A  V  D  Q  Q  S  G  L  E  H  K  T  H  V  L  V  K  N
1456  GGCGCTGCAGTTGATCAGCAGTCTGGGTTGGAGCACAAAACTCATGTGCTGGTCAAGAAC

 506   G  S  K  Y  T  A  T  L  S  L  V  D  M  N  T  G  A  N  S  Y
1516  GGCTCGAAGTACACCGCAACCCTATCCTTGGTGGACATGAATACGGGAGCAAATTCCTAT

 526   Y  K  L  Q  L  L  E  H  D  S  K  K  K  Y  Y  V  F  R  A  W
1576  TATAAGTTGCAGCTTCTGGAGCATGACTCGAAGAAGAAATATTATGTGTTCCGCGCCTGG

 546   G  R  V  G  T  T  I  G  G  D  K  V  E  D  F  Y  S  V  D  A
1636  GGTCGAGTTGGGACGACGATAGGGGGGGATAAAGTCGAGGACTTCTATTCCGTTGATGCG

 566   A  V  R  N  F  E  E  V  Y  L  D  K  T  G  N  E  F  G  T  A
1696  GCGGTGAGGAATTTTGAAGAAGTCTATTTAGATAAAACTGGCAACGAATTTGGCACCGCT

 586   H  F  M  K  M  P  K  K  F  Y  P  L  D  I  D  Y  G  E  N  D
1756  CATTTCATGAAGATGCCCAAAAAGTTTTATCCCCTTGACATCGATTACGGTGAAAACGAT

 606   E  K  L  K  V  L  Q  S  A  G  S  N  S  K  L  H  Q  E  V  Q
1816  GAGAAGCTGAAGGTCTTGCAATCTGCCGGTAGCAATTCGAAGCTACATCAAGAAGTGCAA

 626   K  I  I  K  M  I  F  D  V  E  S  M  K  K  A  M  V  E  F  E
1876  AAGATCATCAAGATGATATTTGACGTGGAAAGCATGAAGAAGGCGATGGTGGAATTCGAG

 646   L  D  M  K  K  M  P  L  G  K  L  S  K  K  Q  L  Q  R  A  Y
1936  CTGGATATGAAAAAGATGCCTCTCGGTAAGTTAAGTAAGAAGCAGCTGCAGAGAGCCTAC

 666   G  V  L  S  E  V  T  G  I  I  E  C  E  T  P  S  A  P  K  I
1996  GGGGTACTGTCCGAAGTGACCGGGATTATTGAATGCGAAACGCCAAGCGCTCCCAAAATA

 686   L  D  C  S  N  R  F  Y  T  L  I  P  H  D  F  G  M  K  Q  P
2056  CTGGATTGCTCAAACAGATTTTACACTTTGATCCCGCATGACTTCGGAATGAAACAGCCG

 706   P  M  L  N  N  I  E  L  V  K  S  K  V  E  M  I  D  S  L  L
2116  CCGATGCTGAACAACATTGAGTTGGTCAAGTCCAAGGTCGAGATGATCGACAGCCTGCTG

 726   D  I  E  V  A  Y  N  L  L  Q  G  D  Y  A  D  V  K  E  E  D
2176  GATATTGAGGTCGCATACAATCTGCTTCAAGGCGACTATGCCGACGTGAAGGAGGAAGAC

 746   P  I  D  T  Q  Y  K  K  L  Q  C  R  I  E  P  L  N  A  N  D
2236  CCCATCGACACTCAGTACAAGAAACTGCAATGCCGAATAGAGCCACTCAATGCCAACGAC

 766   E  M  R  K  V  I  S  D  M  V  Q  N  T  H  A  K  T  H  S  S
2296  GAGATGCGTAAGGTGATCAGCGACATGGTGCAAAACACTCACGCCAAGACCCATTCCTCG

 786   Y  N  L  E  V  E  E  I  F  E  L  D  R  K  G  E  K  S  R  Y
2356  TACAATCTCGAAGTCGAAGAAATATTTGAGCTGGACCGTAAAGGTGAAAAAAGCCGCTAC

 806   R  P  F  K  K  L  H  N  R  R  L  L  W  H  G  S  R  V  T  N
2416  CGGCCGTTCAAGAAGCTGCACAACCGTCGTCTGCTGTGGCATGGCTCTCGTGTGACCAAC

 826   Y  A  G  I  L  S  Q  G  L  R  I  A  P  P  E  A  P  V  T  G
2476  TATGCCGGCATTTTGTCACAGGGCTTGAGGATTGCGCCGCCCGAAGCGCCTGTGACGGGA

 846   Y  M  F  G  K  G  L  Y  F  A  D  M  V  S  K  S  A  N  Y  C
2536  TATATGTTCGGCAAGGGGTTGTACTTTGCCGATATGGTTTCGAAAAGTGCGAATTACTGC

 866   H  A  S  H  G  N  P  Y  G  L  L  L  L  A  E  V  A  L  G  N
2596  CATGCGTCGCATGGGAATCCATACGGACTGTTATTGCTCGCTGAAGTAGCGCTGGGAAAT

 886   M  R  E  L  K  N  A  M  D  V  R  K  L  P  K  G  K  H  S  V
2656  ATGCGCGAGTTGAAGAATGCCATGGACGTACGGAAGCTACCGAAAGGAAAGCACAGCGTC

 906   K  G  M  G  S  T  C  P  D  E  S  S  D  Y  T  M  K  D  G  T
2716  AAGGGTATGGGGTCCACATGCCCGGACGAATCGAGCGACTACACGATGAAAGATGGCACC

 926   I  A  H  V  G  S  G  K  E  S  G  C  E  G  T  S  L  L  Y  N
2776  ATAGCCCATGTAGGGTCCGGGAAGGAGTCTGGCTGTGAAGGAACGTCGTTACTCTACAAT

 946   E  Y  I  V  Y  D  V  A  Q  V  N  L  R  Y  L  V  K  T  K  F
2836  GAATACATAGTGTACGATGTAGCGCAAGTGAACTTGAGATATTTAGTGAAGACTAAATTC

 966   N  M  Q  S  Q  W  -                                            971
2896  AATATGCAAAGTCAATGG TAA                                           2916
                           CAGTGTAGTGCTGATGACATTTTGTGCTATCTGTTGTGG
        TGATTTCATTGCCAAGTTCGCAATAGTAAGCATCTGTTGACCATCATTTTCGCTTTGCTG
        CCTTTTTGACTTCATGTTACCGCTGTTTCTTTTTCTCATCGTGTAAATAAACTGCTGAAT
        TCGCAATTGTAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 6. cDNA and deduced amino acid sequence of BsPARP1. START and STOP codon are in green and red, respectively; 5′ and 3′ UTR are in grey. Light blue: N-terminal zinc-finger domains; pink: PADR1 (poly(ADP-ribose)-synthase 1) domain; bold: BRCT (BRCA1 (breast cancer susceptibility protein C-terminus) domain; light purple: WGR (tryptophane-, glycine-, arginine-rich) motif; orange; C-terminal regulatory PARP domain; yellow: C-terminal catalytic PARP domain.

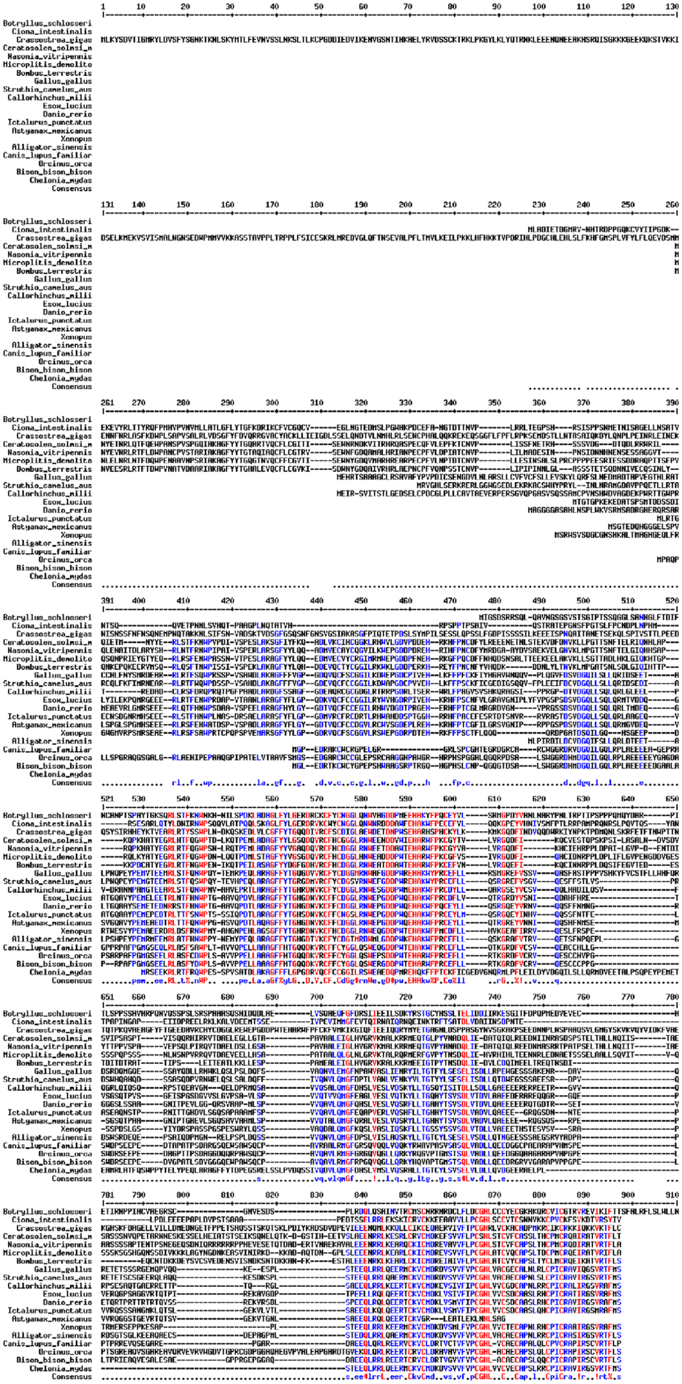


Fig. 7. Alignments of the deduced amino acid sequence of BsIAP7 with known orthologous sequences from both vertebrate and invertebrates.

```
     AAGGTATTCTTTGTACCTATTATATAGTGTATATATGGAGTTGGACGAGTAGA
     ACGCAGCAAAATACGTGCCAAGTTCTGTATGATAGTTCGCCTTGTGAAACTGGA


   1                                                        M  I
   1                                                       ATGATT

      G  S  D  S  R  R  S  Q  L  Q  A  V  N  G  S  G  S  V  S  T
     GGATCGGACAGCAGGAGGTCACAGCTTCAAGCTGTGAATGGAAGCGGAAGCGTGTCAACA

  23  S  G  I  P  T  S  S  Q  G  G  L  S  A  N  H  G  L  F  T  D
  67 TCTGGTATACCTACCAGTTCTCAAGGCGGACTATCAGCAAACCATGGTCTCTTCACGGAC

  43  I  F  N  C  A  N  P  T  S  P  A  Y  T  G  K  S  Q  R  L  S
 127 ATCTTCAACTGTGCAAATCCAACTTCTCCAGCATATACGGGGAAGTCGCAAAGGCTTTCA

  63  T  F  K  N  W  N  K  H  N  I  L  S  P  D  K  L  A  D  A  G
 187 ACCTTTAAAAACTGGAACAAGCACAACATACTGAGCCCAGATAAATTAGCCGACGCAGGA

      L  F  Y  L  G  E  R  D  R  C  K  C  F  Y  C  N  G  G  L  Q
 247 TTGTTTTATTTGGGCGAAAGAGACAGATGTAAATGCTTTTACTGCAATGGAGGATTGCAG

 103  N  W  V  A  G  D  D  P  M  E  E  H  A  K  Y  F  P  Q  C  E
 327 AATTGGGTGGCAGGCGACGACCCAATGGAGGAGCATGCCAAATATTTCCCACAATGCGAA

 123  Y  V  L  S  R  M  G  P  D  Y  V  A  N  L  N  A  R  Y  P  N
 387 TACGTTCTGTCAAGAATGGGACCCGATTACGTTGCAAACCTGAATGCAAGATATCCGAAT

 143  L  T  R  P  T  I  P  S  P  P  P  Q  M  Q  Y  D  A  R  P  I
 447 TTAACGCGACCGACGATTCCGTCACCACCCCCGCAAATGCAATATGACGCCCGGCCAATT

 163  T  L  S  P  P  S  S  A  V  H  R  P  Q  N  V  Q  S  S  P  S
 507 ACGCTGTCACCGCCCAGCAGCGCTGTTCACAGGCCACAAAACGTGCAATCATCGCCATCA

 183  L  S  R  S  P  A  A  A  S  Q  S  H  I  D  Q  D  L  A  E  L
 567 CTCAGCCGGTCGCCAGCAGCTGCATCACAAAGTCATATTGATCAAGACTTGGCGGAATTA

 203  V  S  Q  A  E  Q  F  G  F  D  R  S  I  I  E  E  I  L  S  D
 627 GTATCGCAAGCAGAACAATTCGGTTTTGATCGATCAATTATTGAGGAAATATTATCGGAT

 223  K  Y  R  S  T  G  C  Y  H  S  S  L  T  E  L  I  D  D  I  I
 687 AAGTACAGGTCAACGGGTTGTTATCACAGCAGCCTCACAGAACTTATTGATGACATCATT

 243  R  K  E  S  G  I  T  F  D  P  Q  P  M  E  D  V  E  V  E  C
 747 AGAAAAGAAAGTGGGATTACCTTCGATCCCCAGCCAATGGAAGACGTCGAAGTCGAATGT

 263  H  E  T  I  R  N  P  P  I  A  C  V  A  E  G  R  S  C  G  N
 807 CATGAAACAATACGGAATCCACCGATAGCATGCGTGGCAGAGGGTAGATCCTGTGGTAAT

 283  V  E  S  D  S  P  L  R  D  Q  L  Q  S  A  I  N  V  T  R  C
 867 GTAGAAAGCGACTCGCCGCTACGCGATCAGCTGCAATCAGCAATAAACGTAACCCGTTGT

 303  M  S  C  N  R  R  N  R  D  C  L  F  L  D  C  G  H  L  C  C
 927 ATGTCATGTAACAGGCGGAATCGTGATTGCTTATTCTTAGATTGCGGTCACTTGTGCTGC

 323  C  Y  E  C  G  K  A  K  Q  R  C  V  I  C  G  T  R  V  R  E
 987 TGTTACGAATGCGGGAAAGCAAAGCAAAGATGCGTGATTTGTGGGACCCGGGTTAGGGAA

 343  V  I  K  I  F  T  T  S  F  A  L  R  F  L  S  L  W  L  L  N
1047 GTGATAAAAATATTTACGACATCTTTCGCATTAAGGTTCCTTTCACTGTGGCTTCTAAAC

 363  Y  C  N  -
1107 TACTGCAATTAA
                 ATATAGTTATCTTCAAGATCTTATTTTCGTTTTACTATAATGAATAAA
     CATTCGTTGTGTTCTGATAAAAAAAA
```

Fig. 8. cDNA and deduced amino acid sequence of BsIAP7. START and STOP codon are in green and red, respectively; 5′ and 3′ UTR are in grey. Orange: BIR domain; blue: RING (really interesting new gene) finger domain.

Acknowledgements

This research was supported by grants from the CARIPARO Foundation (Excellence Project 2009) and the Italian M.I.U.R (P.R.I.N. grant 20109XZEPR to LB).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.05.017.

Appendix B. Transparency Document

Transparency data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2016.05.017.

References

[1] N. Franchi, F. Ballin, L. Manni, F. Schiavon, G. Basso, L. Ballarin, Recurrent phagocytosis-induced apoptosis in the cyclical generation change of the compound ascidian *Botryllus schlosseri*, Dev. Comp. Immunol. 62 (2016) 8–16.

[2] D. Campagna, F. Gasparini, N. Franchi, N. Vitulo, F. Ballin, L. Manni, G. Valle, L. Ballarin, Transcriptome dynamics in the asexual cycle of the chordate *Botryllus schlosseri*, BMC Genomics 17 (2016) 275.

[3] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.

[4] X. Huang, W. Miller, A time-efficient linear-space local similarity algorithm, Adv. Appl. Math 12 (1991) 337–357.

Table 1. PCR primers used in this study.

Primer	Sequence	Description	Amplicon Lenght (bp)
(dT)Anchor	GTTTTCCCACGACTTTTTTTTTTTTTTTTTT	cDNA synthesis	
Anchor	GTTTTCCCACGAC	3’ RACE Anchor	
AAP	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG	5’ RACE Anchor	
BAXF	GATCAAGCGAAGAAGGAAGCAGGG	PCR of BsBAX and riboprobe	725
BAXR	GTTGCGGCAGACAGAGAAACCAC	for ISH	
AIFF	GCGGGGGCGTCGGACTACTTAC	PCR of BsAIF and riboprobE	928
AIFR	GCCGCTTCCTGTTCAGACCGTAC	for ISH	
PARPF	GATCGGAGCAACGAGGATCGA	PCR of BsPARP and riboprobe	820
PARPR	TCCAGCAGGCTGTCGATCAT	for ISH	
IAPF	ATTCCGTCACCACCCCCG	PCR of BsIAP and riboprobe	396
IAPR	GGATCTACCCTCTGCCACGC	for ISH	
BAXF-3’	CCTTGCGTGTTATTGGAG	3' RACE of BsBAX	1526 with Anchor
BAXR-5’	CCTGGTCTGGCGGTTCTCGGTC	5’ RACE of BsBAX	249 with AAP
AIFF-3’	GGACGCCAAACTGCCTACTGTCGG	3' RACE of BsAIF	649 with Anchor
AIFR-5’	GTACGGTCTGAACAGGAAGCGGC	5' RACE of BsAIF	1580 with AAP
ParpF-3’	GGACTGTTATTGCTCGCTGAA	3’ RACE of BsPARP	501 with Anchor
PARPR-5’	CAAAGTACAACCCCTTGCC	5’ RACE of BsPARP	2628 with AAP
PARPR-5’N	GCTCCAACCCAGACTGCT	5’ RACE of BsPARP Nested	1551 with AAP
IAPF-3’	GGAAAGCAAAGCAAAGATGC	3' RACE of BsIAP	218 with Anchor
IAPR-5’	AAGCTGTGACCTCCTGCTGT	5' RACE of BsIAP	170 with AAP

Table 2. Percentage of identity between BsBAX and orthologous proteins.

Species	**Accession Number**	**% of identity with BsBAX**
Rattus norvegicus)	GenBank: AAC26327	42.3
Sus scrofa	GenBank: XP_003127338	41.5
Mus musculus	GenBank: NP_031553	42.3
Macaca mulatta	GenBank: NP_001247945	40.8
Oryctolagus cuniculus	GenBank: XP_008250583	39.7
Homo sapiens	GenBank: NP_620116	40.8
Homo sapiens	GenBank: NP_004315	41.8
Taeniopygia guttata	GenBank: XP_002199710	40.4
Anolis carolinensis	GenBank: XP_003226241	43.6
Xenopus laevis	GenBank: NP_001079104	43.4
Xenopus (Silurana) tropicalis	GenBank: NP_989185	42.6
Onchorynchus mykiss	GenBank: ACO08752	34.3
Danio rerio	GenBank: NP_571637	41.6
Ictalurus punctatus	GenBank: NP_001187866	41.4
Oreochromis niloticus	GenBank: XP_003456606	36.8
Salmo salar	GenBank: ACI68449	35.5
Esox lucius	GenBank: ACO13345	35.6
Ciona intestinalis	GenBank: XP_002123003	38.4
Strongylocentrotus purpuratus	GenBank: XP_791118	34.7
Crassostrea gigas	GenBank: EKC42310	37.2
Drosophila melanogaster	GenBank: NP_788278	17.9
Hydra magnipapillata	GenBank: XP_002157460	32.1
Acropora millepora	GenBank: ABX61041	29.9

Table 3. Percentage of identity between BsAIF1 and orthologous proteins.

Species	**GenBank Accession Number**	**% of identity with BsAIF1**
Homo sapiens	GenBank: NP_004199	57.5
Oryctolagus cuniculus	GenBank: XP_002720337	59.3
Sus scrofa	GenBank: NP_001284561	56.7
Rattus norvegicus	GenBank: NP_112646	57.1
Meleagris gallopavo	GenBank: XP_003208346	60.7
Taeniopygia guttata	GenBank: XP_002198740	60.7
Gallus gallus	GenBank: NP_001007491	60.5
Anolis carolinensis	GenBank: XP_008118433	60
Xenopus (Silurana) tropicalis	GenBank: NP_001017244	62.4
Danio rerio	GenBank: NP_956396	61.9
Takifugu rubripes	GenBank: XP_003971249	61.3
Oreochromis niloticus	GenBank: XP_003456194	61.7
Oryzias latipes	GenBank: XP_004076557	61.7
Tetraodon nigroviridis	GenBank: CAG02984	60.5
Ciona intestinalis	GenBank: XP_002131727	60.1
Saccoglossus kovalevski	GenBank: XP_006825404	58.6
Strongylocentrotus purpuratus	GenBank: XP_783530	62.6
Crassostrea gigas	GenBank: EKC36189	49.1
Apis mellifera	GenBank: XP_006563283	51.5
Tribolium castaneum	GenBank: XP_972831	53.7
Camponotus floridanus	GenBank: EFN62354	51
Culex quinquefasciatus	GenBank: XP_001844711	49.7
Drosophila melanogaster	GenBank: NP_608649	48.1
Loa loa	GenBank: EFO16709	46.9
Caenorhabditis elegans	GenBank: NP_499564	39.6
Hydra magnipapillata	GenBank: XP_002156723	56
Suberites domuncula	GenBank: CAL36989	47.8
Amphimedon queenslandica	GenBank: XP_003385522	46.8

Table 4. Percentage of identity between BsPARP1 and orthologous proteins.

Species	**Accession Number**	**% of identity with BsPARP1**
Homo sapiens	GenBank: NP_001609	49.1
Mus musculus	GenBank: P11103	48.6
Bos taurus	GenBank: NP_777176	48
Gallus gallus	GenBank: P26446	49.5
Taeniopygia guttata	GenBank: XP_002194832	48.3
Anolis carolinensis	GenBank: XP_003216114	48.7
Xenopus laevis	GenBank: NP_001081571	48.8
Oreochromis niloticus	GenBank: XP_003449904	50.4
Takifugu rubripes	GenBank: XP_003971725	49.5
Danio rerio	GenBank: NP_001038407	49.2
Ciona intestinalis	GenBank: XP_009860226	55.0
Oicopleura dioica	GenBank: CBY18306	42.6
Saccoglossus kovalevski	GenBank: XP_006812383	38.3
Strongylocentrotus purpuratus	GenBank: XP_001177436	42.9
Crassostrea gigas	GenBank: EKC34863	46.9
Aplysia californica	GenBank: NP_001191521	45.5
Camponotus floridanus	GenBank: EFN69632	43.8
Daphnia pulex	GenBank: EFX75651	43.4
Apis mellifera	GenBank: XP_624477	43
Drosophila melanogaster	GenBank: NP_001104452	39.2
Caenorhabditis elegans	GenBank: NP_491072	30.5
Hydra magnipapillata	GenBank: XP_002163577	60.1
Amphimedon queenslandica	GenBank: XP_003385071	43.8

Table 5. Percentage of identity between BsIAP7 and orthologous proteins.

Species	**Accession Number**	**% of identity with BsIAP7**
Canis lupus familiaris	GenBank: XP_543094.1	31.7
Orcinus orca	GenBank: XP_004282448.1	28.7
Bison bison bison	GenBank: XP_010834953.1	26.8
Gallus gallus	GenBank: XP_417413.4	30.0
Struthio camelus australis	GenBank: XP_009669159.1	30.7
Alligator sinensis	GenBank: XP_006021704.1	29.9
Chelonia mydas	GenBank: EMP35847.1	27.3
Xenopus laevis	GenBank: NP_001082290.1	30.7
Astyanax_mexicanus	GenBank: XP_007229030.1	28.7
Ictalurus punctatus	GenBank: AFL70282.1	29.0
Callorhinchus milii	GenBank: XP_007887106.1	29.0
Danio rerio	GenBank: XP_005162039.1	31.3
Esox lucius	GenBank: XP_010874145.1	28.1
Ciona intestinalis	GenBank: XP_002125780.2	30.7
Crassostrea gigas	GenBank: EKC32616.1	27.6
Ceratosolen solmsi marchali	GenBank: XP_011494114.1	26.4
Bombus terrestris	GenBank: XP_003393059.1	23.4
Nasonia vitripennis	GenBank: XP_001606042.2	26.7
Microplitis demolitor	GenBank: XP_008554578.1	24.7