Contents lists available at ScienceDirect

Data in Brief

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

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 ortholo-gous genes from other deuterostome species together with a study of their identity/similarity.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Specifications Table

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

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    TCGGGGGCGGTGATCAAGCGAAGAAGGAAGCAGGGAAGATAGATGATAAAAATGACGAG

   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  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    CAAAGGGGGAGGTTGGGGAATGATCAGAGAATACTTTGGAACCACCGACTGCGTATGTGGTT

  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
         CCGGTCTGCAAACCGTGGGTGCCGACCACTCCTCCATATTACATTCATTGTGAAATCGA
         TCTAGATGTGATGAATATCAATGTTATGCATTGAAATTTTGTCCAACAATTGCAGAGCCA
         CTGAAATTGGTTTCAGGATTTCTTGAAAATAAAATGGAACGATGATTCACTATCAGTTTT
         GGCATAGCAACAAAGGTGACATGCCATAGCAAACTTTCATTTGGAAAATTATCAGGATTT
         TGTGAGTGGGGTGTAATCTCCGTCAATTGGTTGATGGACAAAGCTGAATGTCGATACTTA
         TATACGAGTATCCCCGGTTTGTCCAAAATAATTGGTGTCCTGTCGTTTGCATATGGTGTC
         GTTTTCGGTACACCAACAATGGATTATGCTCCAGTTAGTCATGAGTTATACAAGGTGCTA
         TTTTCCTTCAAATGAGCAAATAATGATCCCTGAATGTCGTCTCTATTGACATGCGATCTC
         CTTTTTTTACCCAGTAGACTTGAATTCCAATTGGGTCTCGTTTGTATTTTTTGGCTCTATA
         GATCAAGTACATTTTAAAGAATGAAGCCTAATGATTTACTCATATTGCTCTGTTGCGTGG
         TTGTCTACTGTGTTCTAATCATGGGAGAAAAATCTGCGTCTCTCTACTGTGATTATTTGA
         GTCCGATGTGTTCGTGACTTGACAATCCAGCATATATTCTCAGCGTTTTACGATCAATTT
         ATTGCTCAATTATGCGTATTTAATAAAAAATATCATGTAGTTTTTGTAGATAACCTACAAA
         GATTGCTTACCTGGAAAAAAA
```

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.
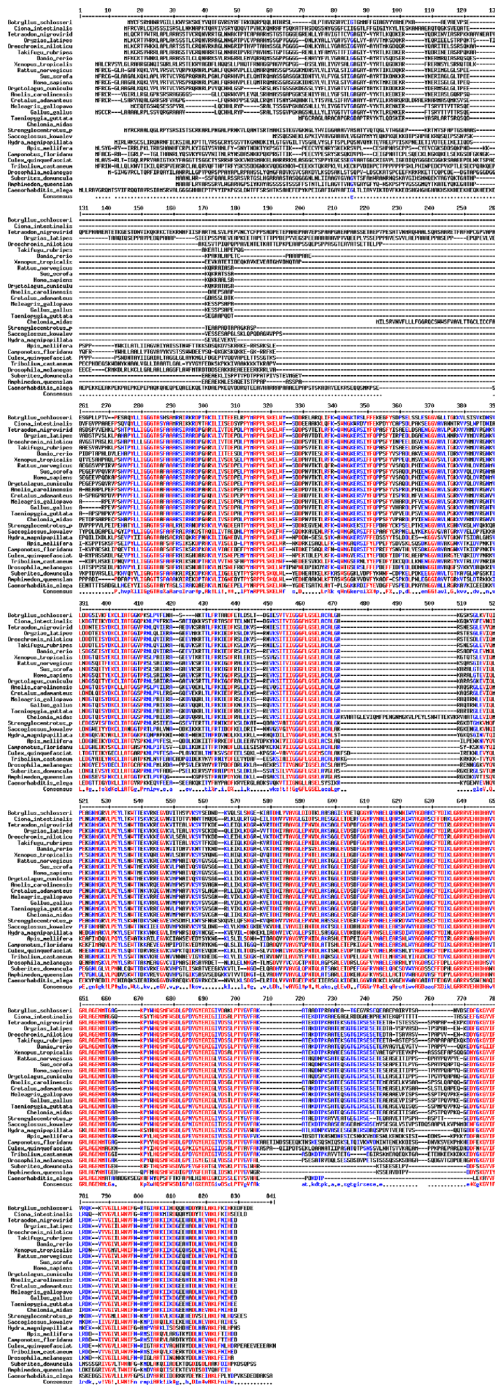
Fig. 3. Alignments of the deduced amino acid sequence of BsAIF1 with known orthologous sequences from both vertebrate and invertebrates.
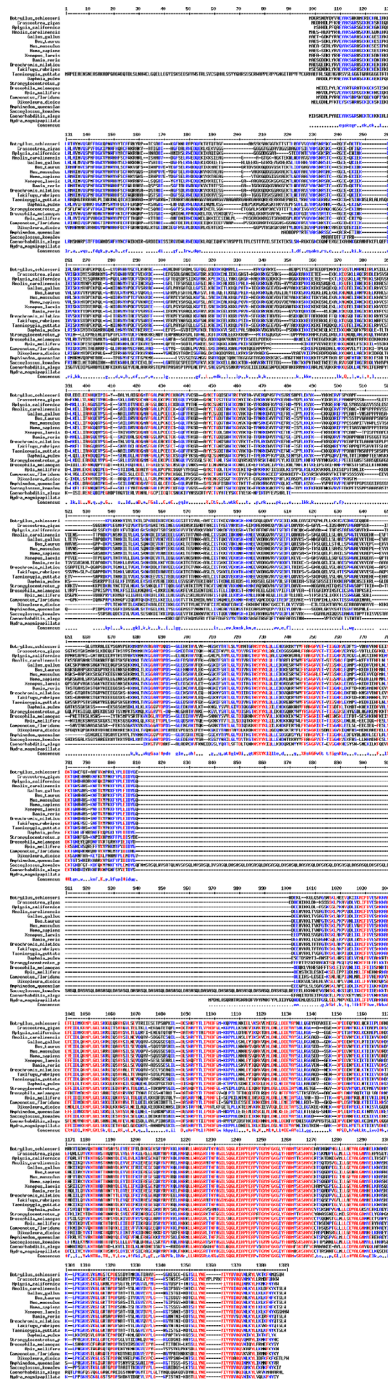
Fig. 4. cDNA and deduced amino acid sequence of BsAIF1. START and STOP codon are in green and red, respectively; 5′ and 3′ UTR are in grey. Orange: pyridine nucleotide-disulphide domain (PNDD); yellow: NADH-binding PNDD; light blue: C-terminal domain of dimerisation; light purple: nuclear localisation 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.
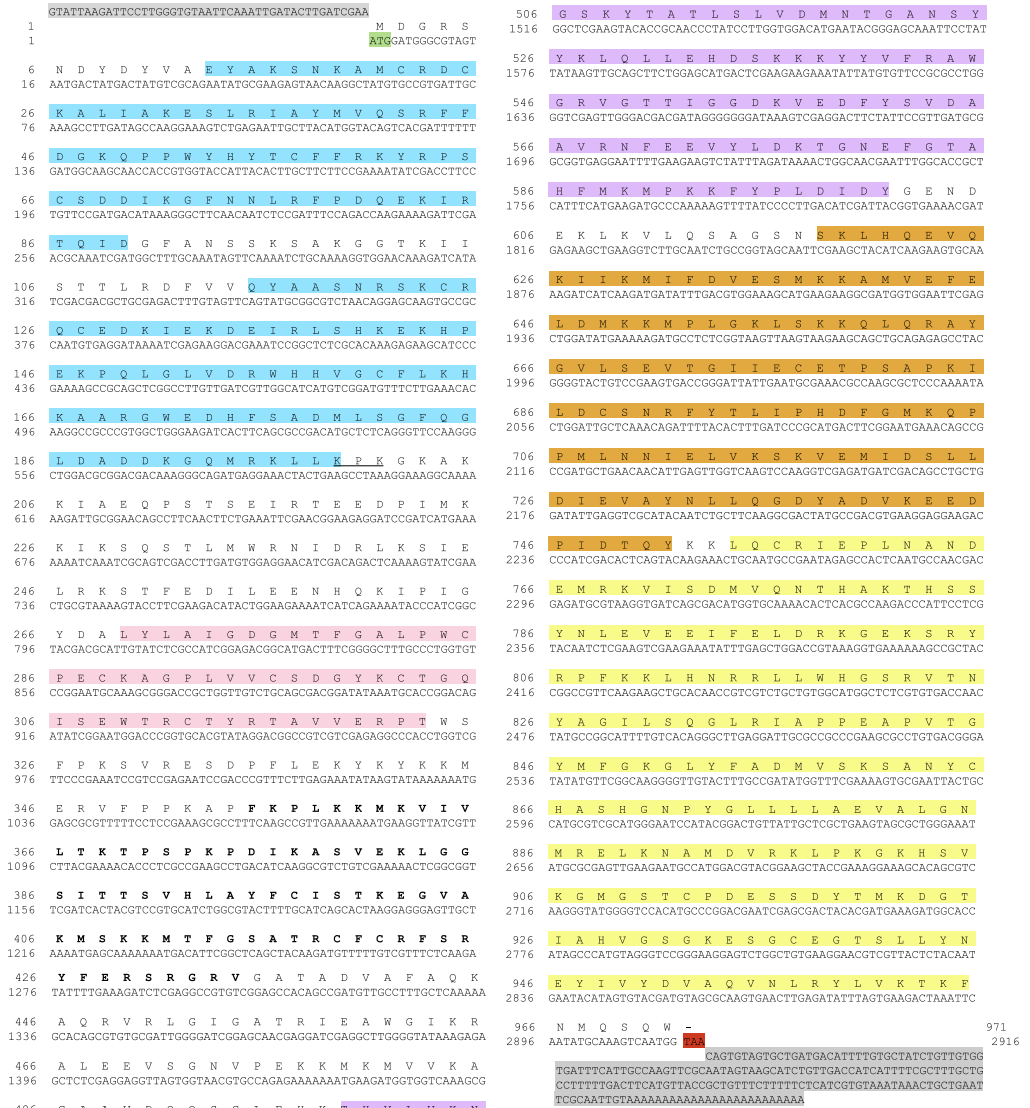
Fig. 6. cDNA and deduced amino acid sequence of BsPARP1. START and STOP codon are in green and red, respectively; 5′ and 3′ UTR are in grey. Light blue: N-terminal zinc-finger domains; pink: PADR1 (poly(ADP-ribose)-synthase 1) domain; bold: BRCT (BRCA1 (breast cancer susceptibility protein C-terminus) domain; light purple: WGR (tryptophane-, glycine-, arginine-rich) motif; orange; C-terminal regulatory PARP domain; yellow: C-terminal catalytic PARP domain.
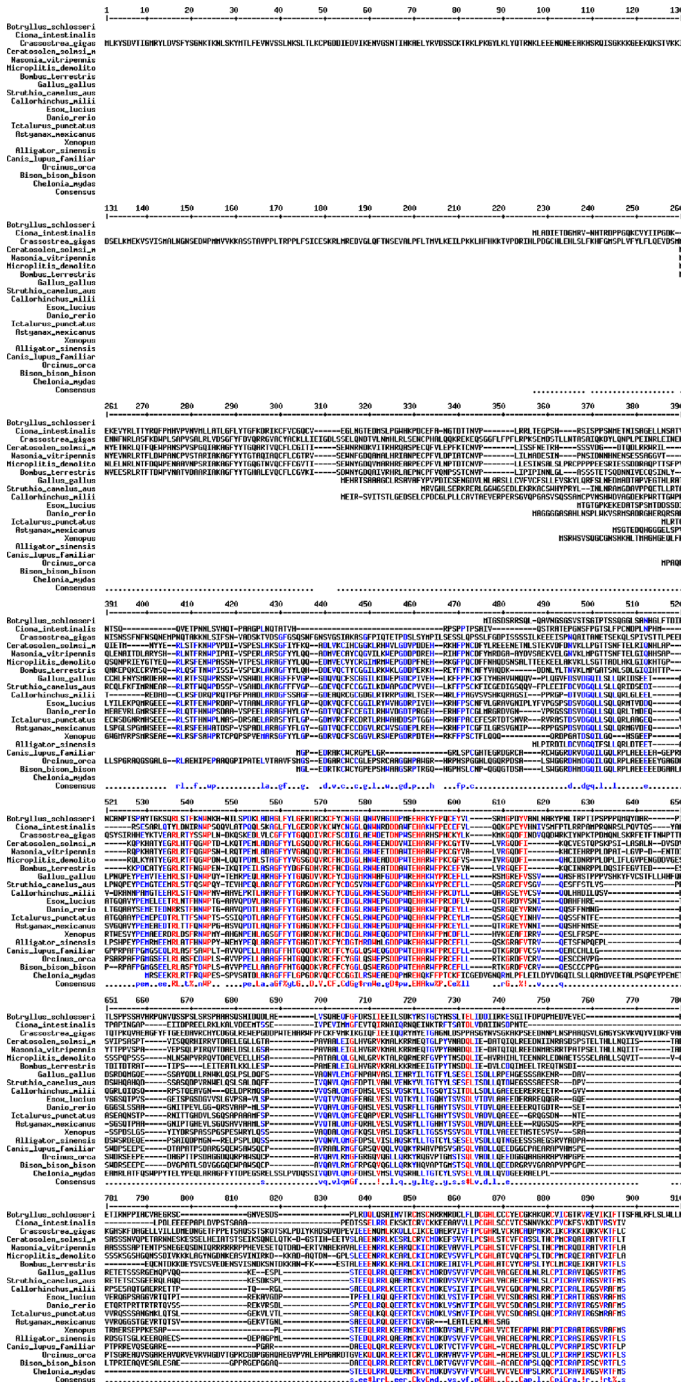
Fig. 7. Alignments of the deduced amino acid sequence of BsIAP7 with known orthologous sequences from both vertebrate and invertebrates.

```
AAGGTATTCTTTGTACCTATTATATAGTGTATATATGGAGTTGGGACGAGTAGA
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 AGAAAGAAAGTGGGATTACCTTCGATCCCCAGCCAATGGAAGACGTCGAAGTCGAATGT

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

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.