Characterization of two novel EF-hand proteins identifies a clade of putative Ca^{2+}-binding protein specific to the Ambulacraria

Arisnel Soto-Acabá[1], Pablo A. Ortíz-Pineda[1#a], José E. García-Arrarás[1*]

[1]Department of Biology, University of Puerto Rico, Río Piedras Campus, San Juan, Puerto Rico

[#a]Current Address: Department of Pediatrics, Yale University, New Haven, Connecticut, United States of America

* Corresponding author

E-mail: jegarcia@hpcf.upr.edu (JEG)

Abstract

In recent years, transcriptomic databases have become one of the main sources for protein discovery. In our studies of nervous system and digestive tract regeneration in echinoderms, we have identified several transcripts that have attracted our attention. One of these molecules corresponds to a previously unidentified transcript (*Orpin*) from the sea cucumber *Holothuria glaberrima* that appeared to be upregulated during intestinal regeneration. We have now identified a second highly similar sequence and analyzed the predicted proteins using bioinformatics tools. Both sequences have EF-hand motifs characteristic of calcium-binding proteins (CaBPs) and N-terminal signal peptides. Sequence comparison analyses such as multiple sequence alignments and phylogenetic analyses only showed significant similarity to sequences from other echinoderms or from hemichordates. Semi-quantitative RT-PCR analyses revealed that transcripts from these sequences are expressed in various tissues including muscle, haemal system, gonads, and mesentery. However, contrary to previous reports, there was no significant differential expression in regenerating tissues. Nonetheless, the identification of unique features in the predicted proteins suggests that these might comprise a novel subfamily of EF-hand containing proteins specific to the Ambulacraria clade.

Introduction

Modern genome and transcriptome studies allow for the identification and discovery of hitherto unknown sequences that code for different types of proteins. This discovery process has been possible due to the ease by which DNA and/or RNA sequences are obtained, even from non-model organisms that make available millions of sequences for comparative analyses. Our group has focused on transcriptomes obtained from normal and regenerating tissues of an echinoderm, the sea cucumber *Holothuria glaberrima* [1–4]. Studies in this model have been done to explore gene expression of intestinal and nervous systems in an attempt to expand our knowledge of the Echinodermata, a phylum which lies on the evolutionary branch of chordates [5,6].

In this effort, we have constructed several transcriptomic libraries using high throughput sequence analyses, including EST (expressed sequence tag) analyses [3], 454 and Illumina sequencing [7,8]. Moreover, we have performed differential gene expression studies, particularly microarrays and transcriptomic comparisons between normal and regenerating tissues. The results from these experiments have been a large number of differentially expressed genes associated with the regenerating tissues.

Out of these hundreds of genes, we have focused on the study of unknown sequences that show increased expression during regeneration. One of these molecules corresponds to a previously unidentified transcript from *H. glaberrima* that was shown to be upregulated during the initial stages of intestine regeneration by microarray analyses

[3,8]. The sequence was annotated to public databases as *Orpin* (GU191018.1, ACZ73832.1) on 12-13-2009.

We now provide a full report on the putative Orpin sequence including the prediction of an N-terminal signal peptide, which is characteristic of secreted proteins. Moreover, we have discovered an additional *Orpin* isoform in *H. glaberrima* and provide a full description of both *Orpin* isoforms. Both sequences are newly discovered putative EF-hand coding proteins with structural characteristics that are evolutionarily related to this group of proteins. We have also probed other available databases and have found previously undescribed sequences whose similarities suggest they are part of the Orpin family, a protein family that appears to be restricted to the Ambulacraria clade.

Materials and methods

Ethics statement

This research deals only with invertebrate animals, thus the University of Puerto Rico IACUC waives ethical approval of research performed on invertebrates. Animals were sacrificed by immersion in ice cold water for 29-30 min and then sectioning the anterior part of the animal close to the oral nerve ring, which accounts for the main component of the nervous system.

Animals

Adult specimens (10–15 cm in length) of the sea cucumber *H. glaberrima* were collected in coastal areas of northeastern Puerto Rico and kept in indoor in aerated seawater aquaria at room temperature (RT: 22°C ± 2°C). Evisceration was induced by 0.35 M KCl injections (3–5 mL) into the coelomic cavity [1]. Eviscerated animals were let to regenerate for 3, 5, 7, 10, and 14 days before the dissection and tissue extraction. For the dissection, organisms were anesthetized by placement in ice-cold water for 1 h. [1,3,9]. Dissected tissues were rinsed in ice-cold filtered seawater and processed for RNA isolation.

RNA extraction and cDNA synthesis

RNA extraction was performed on tissue extracts of normal and 3, 5, 7, 10, and 14 dpe animals. Extracted tissues included gonads, mesentery, haemal system, respiratory tree, longitudinal muscle, and radial nerve cords. After dissection, tissues were placed in 1 mL of TRIzol reagent (Invitrogen), homogenized with a PowerGen Model 125 Homogenizer (Thermo Scientific) and incubated 30 min on ice. These samples were mixed vigorously with 200 µL of chloroform and incubated 10 min at RT. After centrifuged at 12,000 rpm at 4°C, the aqueous RNA phase was separated, mixed with 70% ethanol, and transferred to an RNeasy Mini Kit column (QIAGEN) for deoxyribonuclease (DNase) treatment (QIAGEN). Total RNA was extracted following the manufacturer's protocol. The concentration and purity of the total RNA was measured using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). The cDNA was synthesized from 1 µg of the total

RNA using the ImProm-II Reverse Transcription System (Promega) and oligo $(dT)_{23}$ primers.

Semi-quantitative RT-PCR

RT-PCR reactions were performed using cDNAs prepared from extracted RNA. These reactions were set up in a reaction volume of 25 µL with the final concentration of the PCR primers of 100 nM. Specific primers for the most variable regions between *Orpin A* and *Orpin B* sequences were designed using OligoAnalizer tools from the Integrated DNA Technology webpage (www.idtdna.com). The primers used were: *Orpin B* forward: 5'-ACAGGGAGTACAAACAGTCGTCAA-3' and *Orpin B* reverse: 5'-CTATTTACTCTGCAACTGACACTTTCT-3'; *Orpin A* forward: 5'-ACTTCTGCAGAATCAGTTGTTAAGA-3' and *Orpin A* reverse: 5'-TTCAGTGGAGTCGCCAAC-3'. RT-PCR reactions were performed on three independent RNA samples purified from each of the regeneration stages (previously mentioned) as well as from the normal intestines. The PCR amplification was done by an initial denaturation step of 94°C (45 s), a primer annealing step of 50.2°C (45 s), and an extension step of 72°C (45 s) with a final additional 72°C (10 min) for 28 cycles for *Orpin A*, 26 cycles for *Orpin B*, and 26 cycles for *NADH*, as the amplification parameters for each pair of primers. All samples were analyzed in triplicate. Additional tissues were amplified for 35 cycles (2–4 replicates). The relative expression of *Orpin A* and *Orpin B* was normalized relative to the expression of the housekeeping gene *NADH dehydrogenase subunit 5* using ImageJ software [10] from the optical density values from electrophoresed sample bands on 1% agarose gels, using a Molecular Imager ChemiDoc

XRS+ (BioRad). The primers used for the *NADH* sequence amplification were: forward: 5'-CGGCTACTTCTGCGTTCTTC-3' and reverse: 5'-ATAGGCGCTGTCTCACTGGT-3'. The *Orpin A* and *Orpin B* sequences were confirmed by sequencing excised electrophoresed sample bands at the Sequencing and Genotyping Facility (UPR-RP).

Bioinformatics analyses

Homolog sequences were identified and retrieved from the NCBI GeneBank protein database [11] using the original Orpin sequence previously identified [8] as a query. BLASTp [12,13] were performed against the public non-redundant protein database in GeneBank. Conserved domain identification and UTR analysis were performed using CDD [14], RegRNA [15], UTRScan [16] and PSIPRED [17,18], ScanProsite [19], InterProScan 5 [20,21], Phobos [22], SignalP 5.0 [23], and Phobius [24] on Geneious 11.1.5 software (https://www.geneious.com). Sequence alignments were carried out with MUSCLE [25] (10 iterations) and the Blosum62 matrix and edited with Geneious software 11.1.5 (https://www.geneious.com). Note: It is possible that there are N-terminal sequencing artifacts on two annotated sequences from *A. japonicus* sequences (ARI48335.1 and PIK49419.1). If we delete the residues from the predicted cytoplasmic N-terminal region from the ARI48335.1 sequence and from PIK49419.1 up until their next methionine, they also show a predicted signal peptide of 21 residues each.

Phylogenetic analysis

EF-Hand proteins and other similar sequences were retrieved from literature and protein database as mentioned in results section and the multiple sequence alignment was performed on MAFFT v7.309 [26] with BLOSUM62 scoring matrix, gap open penalty of 1.57, and offset value of 0.123. For the tree building, the Maximum-Likelihood analysis was done using JTT model of sequence evolution with 1000 bootstraps using PhyML 3.0 [27] plugin using Geneious 11.1.5 software (https://www.geneious.com). The corresponding sequences are included in S1 Table. The tree was edited for better visualization and colors in iTOL v4 online tool [28]. The (frog) *X. laevis,* (mouse), *M. musculus,* and (human) *H. sapiens* calcineurin A sequences were selected as outgroups and does not contain EF-Hand motifs. In addition, the Orpin homologs from *A. japonicus* ARI48335.1 and PIK49419.1 were edited for the analyses by deleting the N-terminal residues down to the second predicted methionine for the reason mentioned above.

Statistical analyses

Statistical significance of the resulting data was evaluated through one-way ANOVA using the JMP®, Version 12. SAS Institute Inc., Cary, NC, 1989-2019. The multiple comparison procedure and statistical test Tukey-Kramer HSD (honestly significant difference) was used to determine significant differences between means from optical densities determined by ImageJ software as mentioned before [10]. The Tukey-Kramer results are displayed as small circles for high number of data points and large circles for low number of data points. The large red circle shows significant differences to small grey circles

sample means. All values were reported as the mean ± standard mean error, including mean diamond with confidence interval ([1 – alpha] x 100), and outlier box plot from a quantiles report. While a $P < .05$ and $P < .001$ were considered to indicate statistical significance difference between groups.

Results

Identification of the original *Orpin* (*Orpin A*) sequence and characterization of a second *Orpin* isoform (*Orpin B*)

The original report [3] described a contig sequence (4766-1) which was later annotated as *Orpin*. This contig was used as a template to identify the remaining nucleotides upstream from the open reading frame (ORF) region through RACE-PCR analysis [29]. The *Orpin* sequence is composed of 106 nucleotides from the 5' UTR and 291 nucleotides from the 3' UTR with a 366 nucleotide ORF (plus stop codon) that encodes a putative 122 amino acid peptide followed by a stop codon (Figs 1 and 2). The nucleotide composition of this gene sequence was validated by sequencing the RT-PCR products amplified from a normal intestine tissue cDNA sample (Fig 2). At the time it was annotated in the NCBI database (ACZ73832.1; 12/13/2009), there was no match with other sequences. Two similar sequences from the hemichordate *Saccoglossus kowalevskii* were later added as *Orpin*-like sequences (XP_006824981.1 and XP_002736736.1).

After performing further in-depth analyses of the available transcriptome libraries from regenerating and non-regenerating intestine and regenerating and non-regenerating radial nerve, we discovered an additional highly similar sequence that was identified as a putative *Orpin* isoform. This new putative protein shared 90% identity and 98% similarity with the original *Orpin* sequence but displayed different UTR's from the original sequence. We refer to this sequence as *Orpin B* to differentiate it from the original *Orpin* which we refer from now on as *Orpin A*.

The sequence corresponding to *Orpin B* was also validated through RT-PCR amplification and sequencing (Fig 3 and S1 Fig). *Orpin B* mRNA sequence is composed of 103 nucleotides from the 5' UTR and 364 nucleotides from the 3' UTR (Figs 1 and 3). Its ORF is 369 nucleotides (plus stop codon) long and encodes a putative 123 amino acid protein.

Fig 1. *Orpin A* and *Orpin B* are isoforms. Differences between sequences are highlighted. White bars: 5' UTR and 3' UTR regions of both sequences; pink bar: predicted signal peptides; orange bar: ORF regions; purple bars: predicted EF-hand motifs; green: conservation level; top sequences: nucleotide and amino acid consensus sequences. Differences between nucleotide sequences are highlighted. It is shown a significant difference, especially between both 3' UTR sequences. Analysis was done using the MAFFT plugin in Geneious 11.1.5.

Fig 2. Primers for sq-RT-PCR of *Orpin A*. *Orpin A* UTRs regions (blue boxes) and coding region (green box) of the *Orpin A* gene. Primer sequences designed to specifically amplify *Orpin A* (light red letters).

Primer sequences used for the identification of the original *Orpin* sequence in previous reports (green letters). These primers were designed prior to identification *Orpin* isoform.

Fig 3. Primers for sq-RT-PCR of *Orpin B.* *Orpin B* UTR sequences (blue boxes) and coding region (green box). Primer sequences designed to specifically amplify *Orpin B* (blue letters).

Sequence comparisons among the two *Orpins* from *H. glaberrima* and the two *Orpin*-like sequences from *S. kowalevskii* show that the latter shared 46–50% identity and 76–77% similarity with the *Orpin A* (Fig 4). Similarly, *Orpin B* translated amino acid sequence shared 46–50% identity and 66–67% similarity with the sequences from *S. kowalevskii* (Fig 4). Furthermore, we identified three additional putative *Orpin* homologs from another sea cucumber species, *Apostichopus japonicus*, one from the starfish *Acanthaster planci*, and two from the sea urchin *Strongylocentrotus purpuratus* with expected values (*E*-value < 0.001 and total scores > 47.8). All *Orpin*-like sequences contain one domain that is predicted to be a calcium-binding domain composed of two EF-hand motifs at their carboxy-terminal (Figs 5 and 6).

Fig 4. Orpin homologs pairwise sequence divergence. Translated amino acid sequences comparison by (A) identity% and (B) similarity%. The alignments were done using Muscle with 50 iterations using Geneious 11.1.5.

Fig 5. Orpin homologs alignment. The most conserved residues are indicated by letters in black boxes, green identity regions, and large cartoon letters at the sequence Logo. The exception is PIK49419.1 because 20 amino acid residues from the N-terminal portion are not compared to other sequences. We can see the additional N-terminal regions from *A. japonicus* sequences ARI48335.1 and PIK49419.1 that did not match to the other homologs. Blue box: signal peptide prediction; red box: EF-Hand motif pair prediction. This alignment was done by Muscle plugin with 50 iterations using Geneious 11.1.5.

Domain analyses

Orpin A and Orpin B amino acid sequences were analyzed using different bioinformatics tools (refer to methodology) for evidence that could point towards a possible function. After evaluating these sequences for domain composition using InterProScan and NCBI's CDD, and Phobos [14,20–22], we identified that both sequences contain putative calcium-binding domain regions. Both Orpin isoforms shared identical calcium-binding loops residue composition. The key residue positions that participate in calcium chelation within these loops are conserved when compared with other known EF-hand proteins. The X, Y, Z, −X, −Z positions from each loop of the EF-hands are Asp, Asp, Asp, Asp, Glu ("odd loop") and Asp, Asp, Asp, Ser, Glu ("even loop"), respectively (Fig 6). The only difference is located downstream to the "odd" loop. There are two consecutive amino acids, Leu87 and Ile88, immediately after the Trp86 (−Z+1) of the first calcium-binding loop from Orpin A which are changed to Ser87 and Met88 in Orpin B. Interestingly, Orpin A and Orpin B included a Cys residue at −Z−1 position which is particular to both sea cucumber sequence homologs and is an unusual feature in EF-hand proteins.

When we compared *H. glaberrima* Orpin EF-hand sequences to the other identified putative homologs from *S. kowalevskii*, *A. japonicus, A. planci,* and *S. purpuratus*, we found that additional positions are highly conserved as well. All Orpin sequences share conserved positions at X−4 (Phe), −Z (Glu), −Z+1 (Trp) and −Z+6 (Gly) positions from the "odd" EF-hands, and X−8 (Phe), X (Asp), Z (Asp), −X−1 (Ile), −X (Ser), −Z (Glu) and −Z+1 (Phe) positions from the "even" EF-hands. Alternatively, there are residues particular to the EF-hands from *H. glaberrima* Orpin isoforms, such as Ala at X+1, Ser at Y, Ala at −X+1 from the odd EF-hand, and Val at X+1 and Asn at −X+2 from the even EF-hand. Moreover, there are residues that are particular to the holothurians such as Lys at X−3, Cys at −Z−1, Lys at −Z+9 from the odd EF-hand, and Lys at −Y from the even EF-hand (Fig 6).

Fig 6. Orpin homologs EF-hand motifs alignment. The predicted odd EF-hands match with the canonical EF-hand pattern and the predicted even EF-hands were identified as non-canonical motifs (14 residues vs 12 residues) (red boxes). The non-canonical motifs are similar to vertebrates S100s. Predicted calcium coordinating residues from the identified EF-hands patterns are indicated by blue triangles. Holothurian Orpins contain a Cys residue at the −Z−1 position of the predicted calcium-binding loop (left orange box). The characteristic residues from Orpin residues are highlighted by orange boxes. Alignment was done using the MAFFT plugin in Geneious 11.1.5.

Additional bioinformatics analyses revealed the presence of a signal peptide in the N-terminal of both isoforms (Fig 7). These signal peptides are 20 amino acids long each

and are mainly composed of hydrophobic residues. The predicted signal peptides of Orpin A and Orpin B are nearly identical, with the exception of two residues at positions 3 (Arg/Lys) and 15 (Ala/Ser). Furthermore, InterProScan and Phobius identified the same region of 20 residues as a possible transmembrane region. In both cases, a high probability of a cleavage site was identified at the Cys21 residue of each isoform.

If the signal peptide is eliminated, the remaining sequence is predicted to be localized outside the cytoplasm (Fig 7). This strongly suggests that these peptides could be secreted to the extracellular space and not targeted to the membrane of other cell organelles.

The average length for the predicted signal peptides of the Orpin homologs is 20–22 residues based on SignalP and Phobius predictions [23,24,30], from the initial Met residue. The predicted signal peptides from the two *S. kowalevskii* sequences are longer (22 residues) than the other Orpins. The predicted signal sequence of a sea urchin homolog (XP_011664021.1) is the shortest (18) of the Orpins (Fig 5).

Fig 7. Orpin A and Orpin B bioinformatics characterization. (A) Orpin A and (B) Orpin B have predicted signal peptides at their transmembrane N-terminal regions including cleavage sites. Also, the two isoforms have predicted EF-hand motifs in their non-cytoplasmic regions. These were predicted by various bioinformatics plugin tools using Geneious 11.1.5.

The UTR's of the *H. glaberrima Orpin* sequences were also analyzed. Even though the 5' UTR's from *Orpin A* and *Orpin B* are 80.2% identical, there are 20 nucleotide differences between them, mainly SNPs. A ribosome binding site within the 5' UTR of each isoform sequence was identified. In contrast, the retrieved 3' UTR's of both *Orpin* isoforms were completely different. Polyadenylation sites were identified in both *Orpin A* [8] and *Orpin B* downstream to their corresponding stop codons. Interestingly, these analyses revealed the presence of two putative Musashi binding elements (MBEs) within the 3'UTR of *Orpin A*. Even though the available retrieved 3' UTR from *Orpin B* is longer than its paralog, no MBEs were identified within this sequence. Surprisingly, two putative MBEs were also found within the coding sequence of each *Orpin* isotype.

Orpin phylogenetic analysis

In order to determine the relationship of the different Orpin homologs among themselves and with other EF-hand proteins, a phylogenetic tree was constructed with the PhyML program using a MAFFT alignment as input [26,27]. Orpin A and Orpin B amino acid sequences were used as probes to identify the closest sequences through BLAST searches against the public databases. In addition, representative sequences from different EF-Hand subfamilies of various organisms were obtained from the scientific literature and available databanks. These sequences included members from the following protein families: S100s, calcineurin, recoverin, calbindin, parvalbumin, oncomodulin, osteonectin, SPARC, troponin C, calmodulin, centrin, Spec, and recoverin

(S1 Table). Thus, these sequences were used for the final alignment to generate the phylogenetic tree.

The results from this analysis cluster Orpin and the identified hypothetical homologs from *S. kowalevskii* (acorn worm), *A. japonicus* (sea cucumber), *A. planci* (sea star), and *S. purpuratus* (sea urchin) together with a bootstrap value of 92, separately from other subfamilies of EF-Hand proteins (Fig 8). The other EF-Hand protein sequences cluster together as individual groups. The Orpin-like cluster was the most distant group after the outgroup sequences of mouse and frog calcineurin A, which do not contain EF-Hand motifs, suggesting that Orpins have evolved separately and are not direct homologs of EF-Hand proteins from other species. As expected, *H. glaberrima* Orpins were close to the other sea cucumber *A. japonicus* Orpin-like sequences. The most distant Orpin homologs were those from sea urchin *S. purpuratus.* The closest protein cluster was the osteonectins, BM-40, or SPARC proteins, which comprise a group of secreted CaBP modulators with a single pair of EF-Hand motifs. After these, the other group of proteins that appeared close by were the S100s, which also are small secreted proteins with two EF-hand motifs. The tree also showed the other outgroup EF-hand lacking protein, calcineurin A from humans, was placed separately from the other EF-Hand proteins of a high number of motifs (3 to 6 EF-Hands).

Fig 8. Orpin isoforms are specific to the Ambulacraria clade. EF-hand protein representative sequences from different subfamilies were aligned to build a phylogenetic tree. Orpin homologs were clustered together as a group, separated to the other EF-hand proteins. The tree was made using the

PhyML plugin ran through Geneious 11.1.5. The parameters used for this analysis were JTT model of amino acid substitution and 1000 bootstraps. Scale bar: 1. Protein sequences accession numbers are included in S1 Table.

Orpin gene is expressed in several tissues of *H. glaberrima*

In order to determine the distribution of *Orpin* expression, mRNA was obtained from different tissues or organs of normal (non-regenerating) *H. glaberrima* specimens and processed for PCR analysis. The tissues and organs selected were: small intestine, large intestine, mesentery, radial nerve complex, longitudinal body wall muscle, gonads, and respiratory tree. Primers were designed for the specific detection of *Orpin A* and *Orpin B* mRNA sequences (Figs 2 and 3). Transcript levels were evaluated relative to the expression of *NADH subunit 5*, a constitutively expressed housekeeping gene. The results showed that *Orpin A* and *Orpin B* shared similar tissue specificity (Figs 9 and 10). Transcripts were detected in the gonads, muscle, mesentery, and haemal system but not in the respiratory tree nor in the nerve. Tissue expression varies significantly, with higher expression levels in the mesentery followed by the expression in muscle and gonads where it is slightly higher than in other tissues. Interestingly, a faint second lighter band was detected from *Orpin B* samples from gonads and muscle tissues.

Fig 9. *Orpin A* expression in different tissues. Composite image from RT PCR amplification of *Orpin A* from *H. glaberrima* tissues. *Orpin A* expression (top band) was detected in haemal system (H), muscle (Mu), gonads (G), and mesentery (Me) relative to the expression of *NADH*. *Orpin A* was detected neither

in the nerve (N) nor in the respiratory tree (RT). The image is a composite from different gels and is divided by a white line.

Fig 10. *Orpin B* expression in different tissues. Composite image from RT PCR amplification of *Orpin B* from *H. glaberrima* tissues. *Orpin B* expression (top band) was detected in haemal system (H), muscle (Mu), gonads (G), and mesentery (Me) relative to expression of *NADH*. *Orpin B* was detected neither in the nerve (N) nor in the respiratory tree (RT). The faint band below the *NADH* band corresponded to primer dimers. The image is a composite from different gels and is divided by a white line.

Orpin expression during intestinal regeneration in the sea cucumber *H. glaberrima*

Previous results from our laboratory have shown that *Orpin* was differentially expressed in regenerating intestinal tissues when compared to normal intestinal tissues [3,8]. In order to validate the upregulation of this novel sequence during regenerative processes, *Orpin* transcript levels were measured during different stages of intestine regeneration. In contrast to previous experiments where no particular effort was made to separate the intestine of normal animals from the attached mesenteries, in the present experiments we measured separately the intestine (a mixed portion from the small intestine and from the large intestine) and the mesentery that attaches the intestine to the body wall, for the normal (non-regenerating) samples. Orpin transcript levels were measured relative to the housekeeping gene *NADH subunit 5*. The gene expression levels were monitored using

406 semi-quantitative RT-PCR of tissue extracts from 3 days post evisceration (dpe), 5 dpe,
407 7 dpe, and 10 dpe along with tissues from normal intestine and normal mesentery.
408
409 Previously it was found that the expression levels of *Orpin A* increased after 3 days of
410 intestine regeneration and then gradually returned to the basal levels at 14 days of
411 regeneration. In contrast, to those findings [3,8], there was no statistically detected
412 difference found between the transcript expression of *Orpin A* from normal intestine
413 samples and those from any of the studied regenerative days (Data not shown). This was
414 also true for *Orpin B* (Data not shown). However a high differential expression was
415 detected between tissues from *Orpin A* from normal mesentery and tissues from 7–10
416 dpe sample group, with a $P < .05$ (P=.002) (Fig 11). *Orpin B* exhibited a high differential
417 expression between tissues from normal mesentery and tissues from 3–5 dpe with a $P <$
418 0.05 (P=.02), and 7–10 dpe with a $P < .001$ (Fig 12).
419
420 Interestingly, we found a different expression profile between *Orpin A* and *Orpin B*
421 transcript levels. While both *Orpin* forms show subsequently decreases in their
422 expression to similar levels at 7-dpe to 10-dpe, the decrease of *Orpin B* seems to occur
423 much faster than that of *Orpin A*.
424
425
426 **Fig 11. *Orpin A* expression during intestine regeneration grouped tissues.** Semi-quantitative RT-
427 PCR amplification of *Orpin A* transcripts from mRNA samples from different intestine regenerative days
428 compared to the corresponding expression in samples from normal intestine (NI) and normal mesentery
429 (NM). A statistical high differential expression was found between NM and 7–10 days post evisceration

(dpe) ($P < .05$; $P = .002$); and between 3–5 dpe and 7–10 dpe ($P < .05$; $P = .03$) as indicated in the all pairs Tukey-Kramer HSD test. The large red circle (low number of data points) displays the significant difference between the small grey circles (high number of data points) group means. Red boxes: outlier box plots summarizing the distribution of points at each factor level from the quantiles report. Green diamonds: sample mean and confidence interval ([1 − alpha] x 100). Blue lines: standard mean error. JMP®, Version 12 software was used for the statistical analyses.

Fig 12. *Orpin B* expression during intestine regeneration grouped tissues. Semi-quantitative RT-PCR amplification of *Orpin B* transcripts from mRNA samples from different intestine regenerative days compared to the corresponding expression in samples from normal intestine (NI) and normal mesentery (NM). A statistical high differential expression was found between NM and 3–5 dpe ($P < .05$; $P = .02$), and to 7–10 dpe ($P < .001$) as indicated in the all pairs Tukey-Kramer HSD test. The large red circle (low number of data points) displays the significant difference between the small grey circles (high number of data points) group means. Red boxes: outlier box plots summarizing the distribution of points at each factor level from the quantiles report. Green diamonds: sample mean and confidence interval ([1 − alpha] x 100). Blue lines: standard mean error. JMP®, Version 12 software was used for the statistical analyses.

Discussion

We have now described the presence of two predicted EF-hand domain-containing proteins from the sea cucumber *H. glaberrima*. These putative proteins apparently belong to a unique group that is present in echinoderms and hemichordates. According to the mRNA distribution in the sea cucumber, the translated proteins are expressed in multiple

organs. Moreover, they are highly represented within the mesentery of the normal and regenerating intestine. The possibility that these are Ca^{2+}-binding proteins is discussed in the following sections.

Orpins are novel genes

When the first *Orpin* sequence from *H. glaberrima* was identified, no other sequence that showed significant similarity to it could be found within databases [3,8]. A few months later, two highly similar sequences (and possible homologs) were identified in the hemichordate *Saccoglossus kowalevskii* (acorn worm) and were added to the databases. These sequences were annotated with accession numbers XM_006824918.1, XP_006824981 (*E*-value: 9E-31) and XM_002736690.2, XP_002736736 (*E*-value: 2E-29). Later on, several homologs from closely related organisms of the Echinodermata phylum were added to the public databases: two from the sea urchin *Strongylocentrotus purpuratus*, three from other sea cucumber *Apostichopus japonicus*, and one from the starfish *Acanthaster planci*. The finding of an additional *Orpin* sequence in *H. glaberrima* increased to ten the known sequences and suggested that these sequences belong to a novel family of proteins within a group of metazoans. In all cases, the sequences have been annotated with little or no descriptive information other than their tissue/organism from where they originated. At present, *Orpins* appear to be restricted to the Ambulacraria clade (the group that encompasses echinoderms and hemichordates), however, it remains to be seen if, with the sequencing of other animal genomes, the specificity of *Orpins* to the Ambulacraria still stands.

Although the *H. glaberrima* A and B *Orpin* variants share a high percentage of similarity at the nucleotide and protein levels, our data suggest that they correspond to distinct genes. First, the nucleotide and putative amino acid differences are distributed throughout the complete sequence of both variants. Second, even though the 5' UTR sequences of both *Orpin* sequences share a large similarity, they are not identical, and similar to the coding region, have multiple nucleotide differences distributed along the nucleotide sequences. Third, their 3' UTR sequences are very different, sharing minimal similarities. Finally, other species also have more than one *Orpin*-like gene. For example, sequence information from *Orpin* homologs from *S. kowalevskii* and *S. purpuratus* were annotated as located in different loci. These differences are characteristic of different genes rather than allele variants or products from differential splicing. Nonetheless, in spite of these results that suggest two different isoforms originating from two different genes, it is necessary to have the genome information as conclusive evidence. Moreover, it should be emphasized that while *Orpin A* and *B* mRNAs have been identified and sequenced from various tissues (see below), the *S. kowalevskii*, *A. japonicus*, *A. planci*, and *S. purpuratus* sequences are hypothetical mRNA/protein-coding sequences that remain to be characterized. Even though it was expected that *H. glaberrima Orpin* sequences would be more similar to those from the other sea cucumber *A. japonicus,* the results showed that they shared higher similarity to the hemichordate *S. kowalevskii* sequences. This can be attributed to longer N-terminal regions from two of the *A. japonicus* sequences (ARI48335.1 and PIK49419.1) that did not match with any of the other homologs. These additional regions were annotated as part of the corresponding ORFs due to an identified

methionine upstream to the one that matched the other homologs. Given the fact that these sequences were not validated, it has to be considered the possibility that these first encoding methionine residues could be the result of a PCR artifact, and could be in fact part of their corresponding 5' UTR regions. Hopefully, the characterization of *Orpin* isoforms will provide essential insights that eventually would make feasible the characterization of these homolog sequences.

Are Orpins calcium-binding proteins?

Rigorous analysis of the residue composition of the EF-hand domains coupled to structural and functional experimentation is the mainframe of the study of uncharacterized CaBPs. Thus, understanding the architecture of EF-hand domains provides a hint of the role that a particular EF-hand protein might have.

One of the most prominent characteristics of the putative protein Orpin is the presence of EF-hand motifs that are a distinctive signature of calcium-binding proteins. The EF-hand motif has been used as a standard of reference for the description of the calcium-binding loops from the corresponding domain by the residues at key positions for the chelation of each calcium ion. Even though the EF-hand motifs are a characteristic feature of many CaBPs, few of them consist of less than four EF-hands. As mentioned before, Orpin A and Orpin B paralogs comprise a single putative calcium-binding domain composed of two EF-hands. The canonical EF-hand motif topology is a helix-loop-helix conformation, which regularly binds calcium ions [31,32]. Usually, this conformation is composed of a highly conserved 12 residues calcium-binding loop flanked on both sides by alpha-helices

[33]. The residues that participate in calcium coordination were labeled as X, Y, Z and −X, −Y, −Z [32,34], and those conserved positions are conventionally used as a reference frame to analyze the calcium-binding potential and dynamics of the EF-hand CaBPs. A typical EF-hand domain is composed of two calcium-binding loops containing motifs flanked by alpha-helices. The adjacent alpha helices are named incoming and exiting helices from the odd (N-terminal) and even (C-terminal) EF-hand motifs [35–38]. *H. glaberrima* Orpin isoforms odd calcium-binding loops are composed of highly conserved key amino acids of the canonical domain structure. We showed that these domains shared a high level of conservation in the residues that participate in the chelation of calcium ions. However, the “even” calcium-binding loop from sea cucumber Orpins slightly deviates from the canonical pattern. The conserved Gly6 residue that provides for loop flexibility is substituted by a Glu6. This substitution is well conserved throughout all the available Orpin homologs with the exception of the starfish sequence. Also, they share a conserved Trp13 at position −Z+1 of the “odd” EF-hand. Furthermore, this conserved residue seemed to be particular to Orpins after comparison to the other 84 EF-Hand sequences from this study. Interestingly, a Cys11 residue is particular to the holothurian Orpins.

In view of these facts, Orpin homologs can be classified as novel EF-hand proteins. Although the bioinformatics analysis strongly suggests that they might be a new type of CaBPs, in order to assure this, experimental confirmation of the actual binding of calcium ions will be required along with the phylogenetic evidence provided in this study.

Orpin relationship with other Ca^{2+} binding proteins

Whether they are indeed CaBPs or not, the sequence comparisons show that Orpins share several characteristics with CaBP subfamilies and that in fact, CaBPs account for the most similar proteins in the database. Several patterns have been developed to accurately classify newly discovered EF-hand proteins. The main pattern is the calmodulin canonical EF motif mostly known as the DXDXDG pattern [39,40], which contrasts with the pseudo-EF-hand binding loops from the most recent vertebrate S100 family of proteins.

One of the main EF-hand protein subfamilies are the S100 proteins. These proteins are small CaBPs containing only two EF-hand motifs. Nevertheless, their N-terminal motifs are considered pseudo-EF-hands, which is the main characteristic of this protein family. At the moment of this study, this subfamily has only been found in vertebrates [41]. Although there were no pseudo-EF-hand predicted from both Orpin isoforms sequences, they share several characteristics with S100 proteins, such as the small size, acidic composition, secretion to extracellular location, and EF-hand motif number. Thus, Orpin sequences share structural features with the main CaBP subfamilies, making it difficult to classify them as any of them. Moreover, we have shown that Orpin and Orpin-like sequences clustered together more closely to osteonectins (BM-40/SPARC) proteins, a group of secreted CaBPs with only two EF-Hand motifs, suggesting that these two groups shared a common ancestral origin. In addition, the best-known CaBPs that grouped closely to Orpin-like sequences were mouse (*M. musculus*, AAA37432.1) and frog calcineurin A isoforms (*X. laevis*, AAC23449.1). These two sequences and the human

569 calcineurin A (*H. sapiens*, AAC37581.1) do not contain EF-hands and were included as
570 outgroups for this analysis. We emphasize the fact that the calcineurin isoform B from
571 zebrafish does contain an EF-hand domain, thus it is separated from the outgroup
572 calcineurin A sequences and is included with the other aligned CaBPs. These results
573 suggest that Orpin sequences comprise a specific EF-hand protein group that is different
574 from the other known subfamilies of calcium-binding proteins. These data suggest that
575 we can be dealing with a new subfamily of EF-hand proteins that is specific to the
576 Ambulacraria clade.
577

578 Orpins as secreted proteins

579 In addition to their EF-hand motif, an additional feature of Orpins is the presence of a
580 signal peptide. In this respect, Orpin isoforms strongly resemble the groups of EF-hand
581 proteins that are secreted, namely the osteonectins, oncomodulins, and S100s. Similar
582 to Orpin sequences, osteonectins (BM-40/SPARC), oncomodulin, and S100 proteins are
583 small peptides containing two EF-hand motifs each and are secreted to the extracellular
584 milieu. Interestingly, these proteins display calcium-mediated dimerization either as
585 heterodimers as in the case of S100A8/S100A9 [42] or homodimers as in the case of
586 S100P [43], osteonectins [44], and oncomodulin [45]. Usually, EF-hand proteins
587 containing signal peptides are targeted to the outer plasma membrane. There, these
588 CaBPs can act as growth factors recognizing binding targets located on other cell
589 surfaces, thus activating different signaling pathways. Such is the case of osteonectin
590 (BM-40/SPARC), which promotes changes in cell morphology, disrupt cell adhesion,
591 inhibit cell cycle, regulate extracellular matrix, and modulate cell proliferation and

592 migration [46], thus underlying the process of wound repair. Similarly, the secreted
593 (although lacking a signal peptide) oncomodulin and S100 proteins, are involved in a
594 variety of biological processes including: cell proliferation, differentiation, survival, nerve
595 regeneration, interaction with transcription factors, and calcium homeostasis [47–55]
596 among other functions.
597

598 *Orpins* and regeneration

599 Previous results from our laboratory had shown differential expression of the *Orpin*
600 transcript when mRNAs levels from 3- and 7- day regenerating intestines were compared
601 to normal intestinal tissues. Thus, we had concluded that *Orpin* was over-expressed
602 during intestinal regeneration. Our new data, where we detect high levels of the transcript
603 in the mesentery region, questions our previous interpretation. Thus, if we consider that
604 the samples containing the 3- and 5- day regenerating intestinal rudiments contain a large
605 proportion of the remaining mesentery (that remains attached to the body wall), then the
606 high expression of *Orpin A* and *B* transcript levels that were detected in the 3- or 5- day
607 regenerating tissues can be interpreted as representing the expression in the mesenteric
608 portion and not necessarily in the rudiment itself. As the rudiment itself grows and
609 encompasses a larger proportion of the dissected tissues (in relation to the mesentery)
610 then the *Orpin* expression would appear to decrease. This is why there is no difference
611 between normal mesentery and early regenerating rudiments. The previously observed
612 difference between regenerating rudiments and “normal” intestine would be merely a
613 reflection of the proportion of mesenterial tissue present in both samples; low in “normal”
614 intestines and high in regenerating ones.

In this respect, the isolated “normal” mesentery is a more appropriate control to compare the relative expression of *Orpin* transcript sequences between regenerating and normal tissues. This is particularly true in early regenerating stages (3–5 days) when the proportion of tissues corresponding to the mesentery is quite high.

While the lack of *Orpin* differential expression argues against a possible role in the intestinal regeneration process, we cannot completely exclude this possibility. The presence of Musashi binding elements within *Orpin* isoform sequences suggests that the post-transcriptional regulation of their mRNAs might be controlled by RNA-binding proteins. This element is present in genes that are post-transcriptionally regulated in a spatial and temporal dependent manner [56,57]. Moreover, this type of regulation has been implicated in the self-renewal of epithelial, neural and hematopoietic stem and progenitor cells [58–64]. Such is the case of the target transcript encoding the transcription factor TTK69 in *Drosophila,* where translational activation is mediated by the neural *Drosophila* Musashi. In this way, the Musashi protein induces the differentiation of *Drosophila* IIb cells as neural precursor cells by repressing the translation of the mRNA of this neural differentiation inhibitory factor [58]. Furthermore, the expression of mammalian Numb protein (m-Numb) induces the expression of regeneration-related genes such as prostate stem cell antigen (PSCA) and metallothionein-2 (Mt2) in gastric mucosal regeneration in mice. Musashi protein (Msi1) enhances the expression of m-Numb during this regenerative process through post-transcriptional regulation. Having stated this, we cannot disregard the possibility that *Orpin* isoforms play a role during the

initial stages of intestine and nerve regeneration. Nonetheless, to explore this possibility, we need to determine if a Musashi protein is present in the *H. glaberrima* proteome during regenerative processes of the sea cucumber and that it binds to *Orpins* mRNA.

It is of interest that *Orpin* isoforms are found to be differentially expressed in a transcriptomic library of regenerating nerve from *H. glaberrima* [7], also suggesting a possible regeneration-associated function. The two sequences displayed an increase in expression during the regeneration of the radial nerve complex after induced injury. By day 2 and also by day 20 after nerve injury, *Orpin A* expression was significantly higher than non-regenerating radial nerve ($P < .001$). In addition, *Orpin B* was higher in the same samples after day 2, 12, and 20 after nerve injury ($P < .001$). However, in view of our findings in the intestinal system, it remains to be determined whether this differential expression is also the product of the gene is expressed preferentially in the remaining tissues following injury, and not necessarily of increasing its expression.

In summary, we have identified and characterized a group of Orpin-like proteins from a particular group of invertebrate deuterostomes and shown they all share similarities in size, domain composition, and little significant similarities to other known EF-hand protein sequences. We provide bioinformatics evidence for the presence of signal peptides and cleavage sites in these proteins that suggest secretion of the putative proteins to the extracellular environment. Together, with the identification of predicted EF-hand domains with unique features, we can suggest that these might comprise a novel subfamily of EF-hand containing proteins specific to the Ambulacraria clade. Finally, we studied their

661 expression in normal and regenerating tissues, with the surprise finding that they are
662 highly expressed in the intestinal mesentery.
663
664

665 Acknowledgments

666 The authors thank Dr. Vladimir S. Mashanov for editorial comments on the manuscript
667 as well as to kindly provide sequence differential expression information from *Holothuria*
668 *glaberrima* regenerating nerve.

References

1. García-Arrarás JE, Estrada-Rodgers L, Santiago R, Torres II, Díaz-Miranda L, Torres-Avillán I. Cellular mechanisms of intestine regeneration in the sea cucumber, Holothuria glaberrima Selenka (Holothuroidea:Echinodermata). J Exp Zool. 1998;281: 288–304. Available: http://www.ncbi.nlm.nih.gov/pubmed/9658592

2. García-Arrarás JE, Greenberg MJ. Visceral Regeneration in Holothurians. Microsc Res Tech. 2001;55: 438–451. doi:10.1002/jemt.1189

3. Rojas-Cartagena C, Ortíz-Pineda P, Ramírez-Gómez F, Suárez-Castillo EC, Matos-Cruz V, Rodríguez C, et al. Distinct profiles of expressed sequence tags during intestinal regeneration in the sea cucumber Holothuria glaberrima. Physiol Genomics. 2007;31: 203–15. doi:10.1152/physiolgenomics.00228.2006

4. San Miguel-Ruiz JE, García-Arrarás JE. Common cellular events occur during wound healing and organ regeneration in the sea cucumber Holothuria glaberrima. BMC Dev Biol. 2007;7: 115. doi:10.1186/1471-213X-7-115

5. Mashanov VS, Zueva OR, Garcia-Arraras JE. Organization of glial cells in the adult sea cucumber central nervous system. Glia. 2010;58: 1581–93. doi:10.1002/glia.21031

6. Mashanov VS, Zueva OR, García-Arrarás JE. Radial glial cells play a key role in echinoderm neural regeneration. BMC Biol. 2013;11: 49. doi:10.1186/1741-7007-11-49

7. Mashanov VS, Zueva OR, García-Arrarás JE. Transcriptomic changes during

691 regeneration of the central nervous system in an echinoderm. BMC Genomics.
692 2014;15: 1–21. doi:10.1186/1471-2164-15-357
693 8. Ortiz-Pineda PA, Ramírez-Gómez F, Pérez-Ortiz J, González-Díaz S, Santiago-
694 De Jesús F, Hernández-Pasos J, et al. Gene expression profiling of intestinal
695 regeneration in the sea cucumber. BMC Genomics. 2009;10: 262.
696 doi:10.1186/1471-2164-10-262
697 9. Suárez-Castillo EC, García-Arrarás JE. Molecular evolution of the ependymin
698 protein family: a necessary update. BMC Evol Biol. 2007;7: 23. doi:10.1186/1471-
699 2148-7-23
700 10. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of
701 image analysis. Nat Methods. 2012;9: 671–675.
702 11. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al.
703 Database resources of the National Center for Biotechnology Information. Nucleic
704 Acids Res. 2019;47: D23–D28. doi:10.1093/nar/gky1069
705 12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment
706 search tool. J Mol Biol. 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2
707 13. Bias C, Gish W. Combined Use of Sequence Similarity and Codon Bias for
708 Coding Region Identification. J Comput Biol. 1994;1: 39–50.
709 14. Marchler-bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD /
710 SPARCLE : functional classification of proteins via subfamily domain
711 architectures. 2017;45: 200–203. doi:10.1093/nar/gkw1129
712 15. Huang H-Y, Chien C-H, Jen K-H, Huang H-D. RegRNA: an integrated web server
713 for identifying regulatory RNA motifs and elements. Nucleic Acids Res. 2006;34:

W429-34. doi:10.1093/nar/gkl333

16. Bengert P, Dandekar T. A software tool-box for analysis of regulatory RNA elements. 2003;31: 3441–3445. doi:10.1093/nar/gkg568

17. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. 2019;47: 402–407. doi:10.1093/nar/gkz297

18. Jones DT. Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices. 1999;292: 195–202.

19. Castro E De, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-genevaux PS, Gasteiger E, et al. ScanProsite : detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. 2006;34: 362–365. doi:10.1093/nar/gkl124

20. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan : protein domains identifier. 2005;33: 116–120. doi:10.1093/nar/gki442

21. Jones P, Binns D, Chang H, Fraser M, Li W, Mcanulla C, et al. Sequence analysis InterProScan 5 : genome-scale protein function classification. 2014;30: 1236–1240. doi:10.1093/bioinformatics/btu031

22. Yao X, Liu Y, Tan Y, Song Y, Corlett RT. The complete chloroplast genome sequence of Helwingia himalaica (Helwingiaceae, Aquifoliales) and a chloroplast phylogenomic analysis of the Campanulidae. PeerJ. 2016;4: e2734. doi:10.7717/peerj.2734

23. Juan J, Armenteros A, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat

737 Biotechnol. 2019;37: 420–423. doi:10.1038/s41587-019-0036-z

738 24. Madeira F, Park Y mi, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The
739 EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res.
740 2019;47: W636–W641. doi:10.1093/nar/gkz268

741 25. Edgar RC, Drive RM, Valley M. MUSCLE : multiple sequence alignment with high
742 accuracy and high throughput. 2004;32: 1792–1797. doi:10.1093/nar/gkh340

743 26. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7 :
744 Improvements in Performance and Usability Article Fast Track. Mol Biol Evol.
745 2013;30: 772–780. doi:10.1093/molbev/mst010

746 27. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New
747 algorithms and methods to estimate maximum-likelihood phylogenies : assessing
748 the performance of PhyML 3 . 0 New Algorithms and Methods to Estimate
749 Maximum-Likelihood Phylogenies : Assessing the Performance of PhyML 3.0.
750 Syst Biol. 2010;59: 307–321. doi:10.1093/sysbio/syq010

751 28. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new
752 developments. Nucleic Acids Res. 2019;47: W256–W259.
753 doi:10.1093/nar/gkz239

754 29. Ortíz-Pineda PA. Analysis and Characterization of Genes Associated with
755 Intestinal Regeneration in the Sea Cucumber (Echinodermata: Holothuroidea):
756 University of Puerto Rico (Rio Piedras Campus); 2010. 364 p.

757 30. Petersen TN, Brunak S, Heijne G Von, Nielsen H. Correspondence SignalP 4.0 :
758 discriminating signal peptides from transmembrane regions. Nat Publ Gr. 2011;8:
759 785–786. doi:10.1038/nmeth.1701

760 31. Persechini A, Moncrief ND, Kretsinger RH. The EF-hand family of calcium-
761 modulated proteins. Trends Neurosci. 1989;12: 462–467. doi:10.1016/0166-
762 2236(89)90097-0
763 32. Kretsinger RH, Tolbert D, Nakayama S, Pearson W. The EF-Hand, Homologs and
764 Analogs. Nov Calcium-Binding Proteins. 1991; 17–37. doi:10.1007/978-3-642-
765 76150-8_3
766 33. Nelson MR, Chazin WJ. Structures of EF-hand Ca2+-binding proteins: Diversity in
767 the organization, packing and response to Ca2+ binding. BioMetals. 1998;11:
768 297–318. doi:10.1023/A:1009253808876
769 34. Malmendal A, Carlström G, Hambraeus C, Drakenberg T, Forsén S, Akke M.
770 Sequence and Context Dependence of EF-Hand Loop Dynamics. An 15 N
771 Relaxation Study of a Calcium-Binding Site Mutant of Calbindin D9k.
772 Biochemistry. 1998; 2586–2595. doi:10.1021/bi971798a
773 35. Gifford JL, Walsh MP, Vogel HJ. Structures and metal-ion-binding properties of
774 the Ca2+-binding helix-loop-helix EF-hand motifs. Biochem J. 2007;405: 199–221.
775 doi:10.1042/BJ20070255
776 36. Moncrief ND, Kretsinger RH, Goodman M. Evolution of EF-hand Calcium-
777 modulated Proteins. I. Relationships Based on Amino Acid Sequences. J Mol
778 Evol. 1990;30: 522–562. doi:Doi 10.1007/Bf02101108
779 37. Grabarek Z. Structural Basis for Diversity of the EF-hand Calcium-binding
780 Proteins. J Mol Biol. 2006;359: 509–525. doi:10.1016/j.jmb.2006.03.066
781 38. Kawasaki H, Nakayama S, Kretsinger RH. Classification and evolution of EF-hand
782 proteins. BioMetals. 1998;11: 277–295. doi:10.1023/A:1009282307967

783 39. Rigden DJ, Galperin MY. The DxDxDG motif for calcium binding: Multiple
784 structural contexts and implications for evolution. J Mol Biol. 2004;343: 971–984.
785 doi:10.1016/j.jmb.2004.08.077
786 40. Denessiouk K, Permyakov S, Denesyuk A, Permyakov E, Johnson MS. Two
787 Structural Motifs within Canonical EF-Hand Calcium- Binding Domains Identify
788 Five Different Classes of Calcium Buffers and Sensors. PLoS One. 2014;9: 1–14.
789 doi:10.1371/journal.pone.0109287
790 41. Morgan RO, Martin-Almedina S, Garcia M, Jhoncon-Kooyip J, Fernandez MP.
791 Deciphering function and mechanism of calcium-binding proteins from their
792 evolutionary imprints. Biochim Biophys Acta - Mol Cell Res. 2006;1763: 1238–
793 1249. doi:10.1016/j.bbamcr.2006.09.028
794 42. Edgeworth J, Gorman M, Bennett R, Freemont P, Hogg N. Identification of p8,14
795 as a highly abundant heterodimeric calcium binding protein complex of myeloid
796 cells. J Biol Chem. 1991;266: 7706–7713.
797 43. Vivo SPH, Koltzscher M, Gerke V. Identification of Hydrophobic Amino Acid
798 Residues Involved in the Formation of. 2000; 9533–9539.
799 44. Maurer P, Hohenadl C, Hohenester E, Göhring W, Timpl R, Engel J. The C-
800 terminal Portion of BM-40 (SPARC/Osteonectin) is an Autonomously Folding and
801 Crystallisable Domain that Binds Calcium and Collagen IV. J Mol Biol. 1995;253:
802 347–357. doi:10.1006/jmbi.1995.0557
803 45. Mutus B, Palmer EJ, MacManus JP. Disulfide-Linked Dimer of Oncomodulin:
804 Comparison to Calmodulin. Biochemistry. 1988;27: 5615–5622.
805 doi:10.1021/bi00415a033

806 46. Phan E, Ahluwalia A, Tarnawski AS. Role of SPARC-matricellular protein in
807 pathophysiology and tissue injury healing: Implications for gastritis and gastric
808 ulcers. Med Sci Monit. 2007;13: RA25-30.
809 47. Yin Y, Cui Q, Li Y, Irwin N, Fischer D, Harvey AR, et al. Macrophage-Derived
810 Factors Stimulate Optic Nerve Regeneration. J Neurosci. 2003;23: 2284–2293.
811 doi:10.1523/JNEUROSCI.23-06-02284.2003
812 48. Yin Y, Cui Q, Gilbert H-Y, Yang Y, Yang Z, Berlinicke C, et al. Oncomodulin links
813 inflammation to optic nerve regeneration. Proc Natl Acad Sci U S A. 2009;106:
814 19587–19592. doi:10.1073/pnas.0907085106
815 49. Kurimoto T, Yin Y, Habboub G, Gilbert H-Y, Li Y, Nakao S, et al. Neutrophils
816 express oncomodulin and promote optic nerve regeneration. J Neurosci. 2013;33:
817 14816–24. doi:10.1523/JNEUROSCI.5511-12.2013
818 50. Yin Y, Cui Q, Li Y, Irwin N, Fischer D, Harvey AR, et al. Macrophage-derived
819 factors stimulate optic nerve regeneration. J Neurosci. 2003;23: 2284–2293.
820 doi:23/6/2284 [pii]
821 51. Donato R. S100: A multigenic family of calcium-modulated proteins of the EF-
822 hand type with intracellular and extracellular functional roles. Int J Biochem Cell
823 Biol. 2001;33: 637–668. doi:10.1016/S1357-2725(01)00046-2
824 52. Donato R. Functional roles of S100 proteins, calcium-binding proteins of the EF-
825 hand type. Biochim Biophys Acta - Mol Cell Res. 1999;1450: 191–231.
826 doi:10.1016/S0167-4889(99)00058-0
827 53. Donato R, Cannon BR, Sorci G, Riuzzi F, Hsu K, Weber DJ, et al. Functions of
828 S100 proteins. Curr Mol Med. 2013;13: 24–57.

829 doi:10.2174/1566524011307010024

830 54. Donato R, Sorci G, Riuzzi F, Arcuri C, Bianchi R, Brozzi F, et al. S100B's double
831 life: Intracellular regulator and extracellular signal. Biochim Biophys Acta - Mol
832 Cell Res. 2009;1793: 1008–1022. doi:10.1016/j.bbamcr.2008.11.009

833 55. Schaub MC, Heizmann CW. Calcium, troponin, calmodulin, S100 proteins: From
834 myocardial basics to new therapeutic strategies. Biochem Biophys Res Commun.
835 2008;369: 247–264. doi:10.1016/j.bbrc.2007.10.082

836 56. MacNicol MC, Cragle CE, MacNicol AM. Context-dependent regulation of
837 Musashi-mediated mRNA translation and cell cycle regulation. Cell Cycle.
838 2011;10: 39–44. doi:10.4161/cc.10.1.14388

839 57. Takahashi T, Suzuki H, Imai T, Shibata S, Tabuchi Y, Tsuchimoto K, et al.
840 Musashi-1 post-transcriptionally enhances phosphotyrosine-binding domain-
841 containing m-Numb protein expression in regenerating gastric mucosa. PLoS
842 One. 2013;8: e53540. doi:10.1371/journal.pone.0053540

843 58. Okano H, Kawahara H, Toriya M, Nakao K, Shibata S, Imai T. Function of RNA-
844 binding protein Musashi-1 in stem cells. Exp Cell Res. 2005;306: 349–356.
845 doi:10.1016/j.yexcr.2005.02.021

846 59. Nakamura M, Okano H, Blendy JA, Montell C. Musashi, a neural RNA-binding
847 protein required for Drosophila adult external sensory organ development.
848 Neuron. 1994;13: 67–81. doi:10.1016/0896-6273(94)90460-X

849 60. Sakakibara S, Nakamura Y, Yoshida T, Shibata S, Koike M, Takano H, et al.
850 RNA-binding protein Musashi family: roles for CNS stem cells and a
851 subpopulation of ependymal cells revealed by targeted disruption and antisense

ablation. Proc Natl Acad Sci U S A. 2002;99: 15194–15199. doi:10.1073/pnas.232087499

61. Asai R, Okano H, Yasugi S. Correlation between Musashi-1 and c-hairy-1 expression and cell proliferation activity in the developing intestine and stomach of both chicken and mouse. Dev Growth Differ. 2005;47: 501–510. doi:10.1111/j.1440-169X.2005.00825.x

62. Wang X-Y, Yin Y, Yuan H, Sakamaki T, Okano H, Glazer RI. Musashi1 modulates mammary progenitor cell expansion through proliferin-mediated activation of the Wnt and Notch pathways. Mol Cell Biol. 2008;28: 3589–99. doi:10.1128/MCB.00040-08

63. Kharas MG, Lengner CJ, Al-Shahrour F, Bullinger L, Ball B, Zaidi S, et al. Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. Nat Med. 2010;16: 903–8. doi:10.1038/nm.2187

64. Hope KJ, Cellot S, Ting SB, MacRae T, Mayotte N, Iscove NN, et al. An RNAi screen identifies msi2 and prox1 as having opposite roles in the regulation of hematopoietic stem cell activity. Cell Stem Cell. 2010;7: 101–113. doi:10.1016/j.stem.2010.06.007

Supporting information

S1 Fig. RT-PCR primers test. *Orpin A* and *Orpin B* RT-PCR amplification using primers for semi-quantitative analysis. cDNA used was from a pool of cDNAs including normal intestine, mesentery, and regenerating days tissues. Run cycles: 35. *Orpin B* primers show amplification of dimers, also shown in cDNA (-) control.

S1 Table. EF-hand proteins sequences for the phylogenetic analyses


signal peptide
Identity
5' UTR
1. Orpin A c...
2. Orpin B c...
signal peptide
EF hand 1
EF hand 2
3' UTR

Figure 1

ORPIN A

AAAACTTCTGCAGAATCAGTTGTTAAGAATTGTACACAGGCCGAAACTAGTCGTTCAACATCGCAA
CTCCTCCAGAATCATCCAGAACAAGCATTTACATCAATC

ATGGCTAGACTTCTTGTACTCGTCGCTACACTTGCTTTGGTCGCTGTCATTAGTGAAGCGTGCACCT
CATCATCTGGGCGCACGCCATCTTGGGACGTTGGCGACTCCACTGAAATCTACCCCACTTACGATCC
TGCTACCGAATCAGTTGGAGTCGGAGTCACTGTTAGGTTTAAAAGAGAAACCGACGATAGTAAACT
GCAGGTCTTCAAATTAATTGACGCATCTGGTGACGGTTACATTGACGCTTGCGAGTGGTTAATAGAG
GGCGGTATTGTCAAAAACTTTGTCCAGTTCCTGACAGACGATGATGTCGATGGTGATGAGAAGATT
TCTTGGAATGAGTTCCAGAAAGTGTCAGTCGCATAG

TAGATAGAAATAGCGCCCCCAGTTCCAGTTATTTAATATAGTCTATTTTAGAGACCTGTAGGGTAAGG
AACATTTACTTCAAAGATGACAGTCTTTCCAAACCTTTCTTCAAAAGCTCAACGAACTGCACCTATC
ATGTCACCTTAAGCTTAGAGTAGGATCTATCGACATTTTGAAACACAATATTGGTATAACGCTCTTTG
AATACCGATAATACCGGATGCATGGGTATATATGCAAGCAGAAAATAAATACATTGTCTCTATGTGAG
CCATTGTGAAAAACCGTGAAAA

Figure 2

ORPIN B

ACAGGGAGTACAAACAGTCGTCAAGAATTGTTCACGGGCCGAAACTAGTTGTTCAACATCGTAACT
CTTCCAGAATCACCCAGAACAAGTATTTACATCAATC

ATGGCTAAACTTCTTGTACTCGTTGCTACACTTGCTTTGGTCTCTGTTATTAGTGAAGCGTGCACCT
CATCATCTGGGCGCACACCAGCTTGGGACATTGGAAACTCAGCTGAAATTTACCCCACTTTTGATC
CTGCTACCAAGTCAGTTGGAGTCGGAGTCACTATTAGGTTTAAAAGAGAAACCGGCGATAGTAAAC
TGCAGGTCTTCAAATTAATCGACGCATCTGGTGACGGTTACATTGACGCTTGCGAGTGGTCAATGG
AGGGCGGTATTGTCAAAAACTTTGTCCAATTCCTGACAGACGATGATGTCGATGGTGATGAAAAGA
TTTCTTGGAATGAGTTCCAGAAAGTGTCAGTTGCAGAGTAA

ATAGAAATAGCGCGCTCACCTCAGTTATTCACATTGTCTATTTTAAAGGGAATATTGCGTTTTAAATG
CCCCAACATGGCATACAAAGAAAATACCTTTAATCCTTATTAAACCATTCGTTCTCACTAATAAACTC
TAGTAATTTTTTCTTCTCCTCAAAAATTATGCATTACAAGAGTAATGTATTGTTGCATCATTTGAATTT
TTGAAAGTTTGGGTGAACAAATTACTTTTGTGCAACAAATAATCGAATTACTTCCTGTGAGAGCCTT
ATTTCTTGAAAAGAAGCTATCGTACTGCCTTATTTTGAATTGTCTATCATACTGTATCTAACATATTG
GTTTAAAATATATGAAAATTTTCT

Figure 3

A

	ORPIN A OR...	ORPIN B OR...	S. kowalevsk...	S. kowalevsk...	A. japonicus ...	A. japonicus ...	A. japonicus ...	A. planci LO...	S. purpuratu...	S. purpuratu...
ORPIN A ORF translation		90.16%	43.08%	45.38%	46.81%	46.81%	35.00%	40.48%	34.35%	25.52%
ORPIN B ORF translation	90.16%		43.51%	42.75%	47.52%	47.52%	33.89%	37.80%	31.82%	22.80%
S. kowalevskii LOC100375...	43.08%	43.51%		86.23%	30.87%	32.21%	23.40%	25.74%	28.06%	20.00%
S. kowalevskii LOC100369...	45.38%	42.75%	86.23%		29.53%	30.87%	23.40%	25.00%	31.65%	22.50%
A. japonicus ARI48335.1	46.81%	47.52%	30.87%	29.53%		81.07%	49.72%	37.32%	30.20%	22.86%
A. japonicus PIK49419.1	46.81%	47.52%	32.21%	30.87%	81.07%		49.72%	36.62%	30.20%	22.86%
A. japonicus PIK49418.1	35.00%	33.89%	23.40%	23.40%	49.72%	49.72%		25.27%	20.74%	16.47%
A. planci LOC110976159	40.48%	37.80%	25.74%	25.00%	37.32%	36.62%	25.27%		26.09%	18.09%
S. purpuratus LOC105438...	34.35%	31.82%	28.06%	31.65%	30.20%	30.20%	20.74%	26.09%		47.09%
S. purpuratus LOC100890...	25.52%	22.80%	20.00%	22.50%	22.86%	22.86%	16.47%	18.09%	47.09%	

B

	ORPIN A OR...	ORPIN B OR...	S. kowalevsk...	S. kowalevsk...	A. japonicus ...	A. japonicus ...	A. japonicus ...	A. planci LO...	S. purpuratu...	S. purpuratu...
ORPIN A ORF translation		98.36%	76.15%	76.92%	68.79%	67.38%	54.44%	67.46%	59.54%	45.31%
ORPIN B ORF translation	98.36%		77.10%	77.86%	69.50%	68.09%	52.78%	66.14%	57.58%	44.04%
S. kowalevskii LOC100375...	76.15%	77.10%		97.10%	56.38%	55.03%	45.21%	59.56%	56.12%	40.00%
S. kowalevskii LOC100369...	76.92%	77.86%	97.10%		58.39%	57.05%	46.28%	60.29%	57.55%	41.50%
A. japonicus ARI48335.1	68.79%	69.50%	56.38%	58.39%		89.94%	69.06%	63.38%	51.68%	40.95%
A. japonicus PIK49419.1	67.38%	68.09%	55.03%	57.05%	89.94%		69.61%	64.08%	51.68%	40.95%
A. japonicus PIK49418.1	54.44%	52.78%	45.21%	46.28%	69.06%	69.61%		47.25%	44.15%	35.74%
A. planci LOC110976159	67.46%	66.14%	59.56%	60.29%	63.38%	64.08%	47.25%		51.45%	38.69%
S. purpuratus LOC105438...	59.54%	57.58%	56.12%	57.55%	51.68%	51.68%	44.15%	51.45%		56.80%
S. purpuratus LOC100890...	45.31%	44.04%	40.00%	41.50%	40.95%	40.95%	35.74%	38.69%	56.80%	

Figure 4

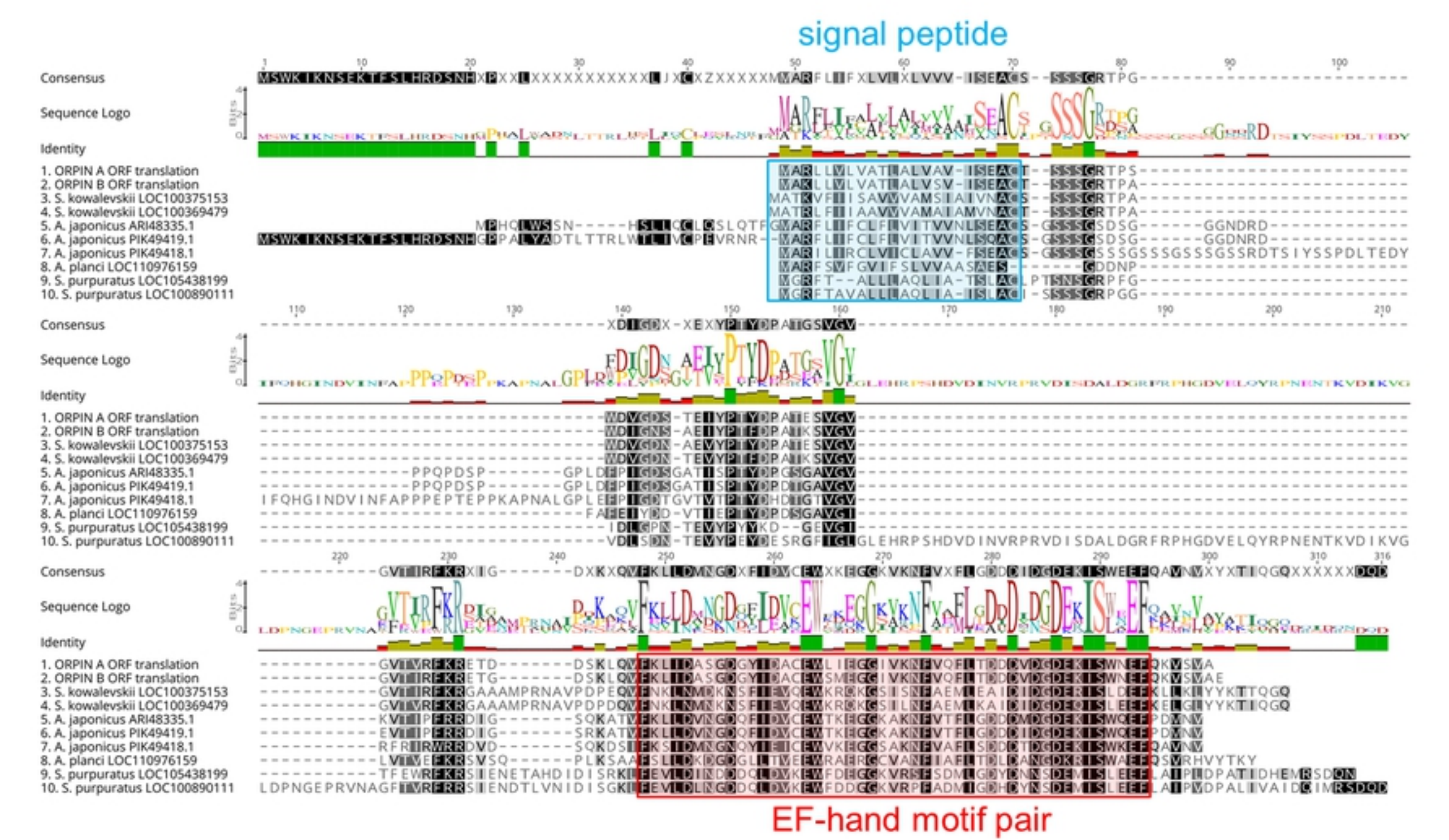
signal peptide
Consensus
Sequence Logo
Identity
1. ORPIN A ORF translation
2. ORPIN B ORF translation
3. S. kowalevskii LOC100375153
4. S. kowalevskii LOC100369479
5. A. japonicus ARI48335.1
6. A. japonicus PIK49419.1
7. A. japonicus PIK49418.1
8. A. planci LOC110976159
9. S. purpuratus LOC105438199
10. S. purpuratus LOC100890111
EF-hand motif pair

Figure 5

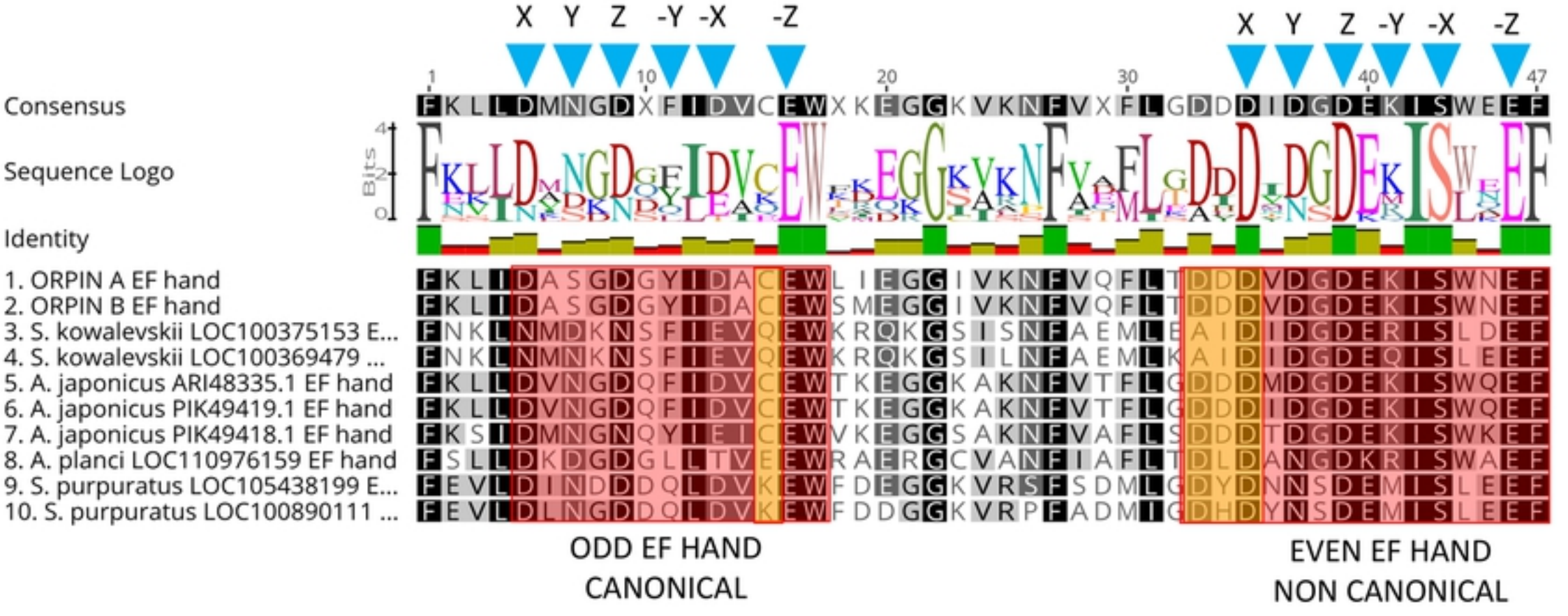
X Y Z -Y -X -Z
X Y Z -Y -X -Z
1
10
20
30
40
47
Consensus
Sequence Logo
Bits
Identity
1. ORPIN A EF hand
2. ORPIN B EF hand
3. S. kowalevskii LOC100375153 E...
4. S. kowalevskii LOC100369479 ...
5. A. japonicus ARI48335.1 EF hand
6. A. japonicus PIK49419.1 EF hand
7. A. japonicus PIK49418.1 EF hand
8. A. planci LOC110976159 EF hand
9. S. purpuratus LOC105438199 E...
10. S. purpuratus LOC100890111 ...
ODD EF HAND
CANONICAL
EVEN EF HAND
NON CANONICAL

Figure 6


A
Hg_Orpin_A
SignalP-noTM
SignalP-TM
Transmembrane (potential)
Signal peptide C-region
Signal peptide H-region
Signal Peptide
Cytoplasmic (potential)
signal cleave region (Predicted)
Signal peptide N-region
unknown
EF-hand_5
EF_HAND_1
EF-hand
Extracellular (potential)
Non cytoplasmic domain
signal cleave region (Predicted)
100%
out
TM Prediction tm
in
0%
B
Hg_Orpin_B
signal cleave region (Predicted)
SignalP-noTM
SignalP-noTM
SignalP-TM
Transmembrane (potential)
Signal Peptide
Cytoplasmic (potential)
Signal peptide N-region
Signal peptide H-region
Signal peptide C-region
unkrown
EF-hand_5
EF_HAND_1
EF-hand
Extracellular (potential)
Non cytoplasmic domain
100%
out
TM Prediction tm
in
0%

Figure 7

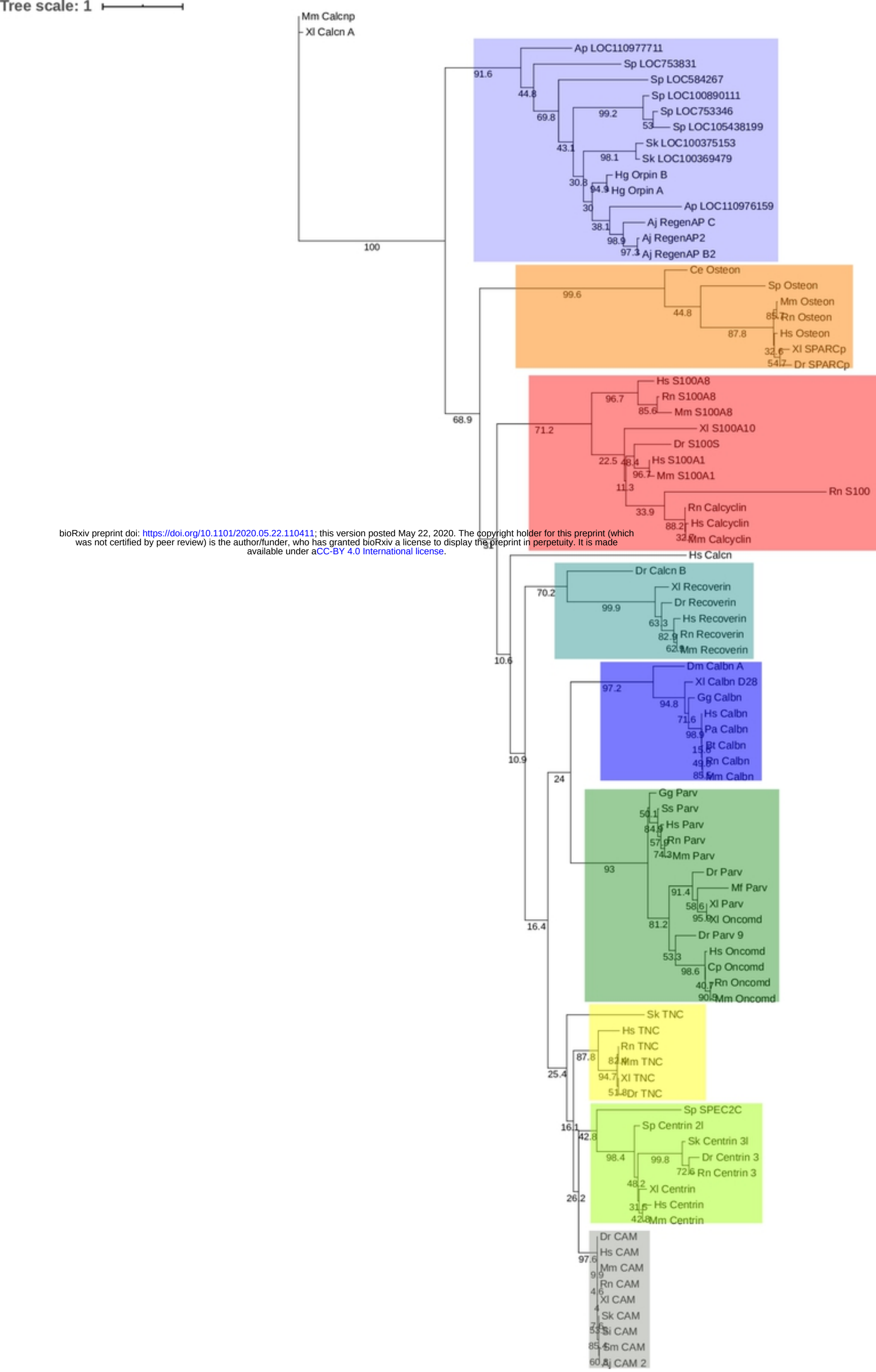

Figure 8

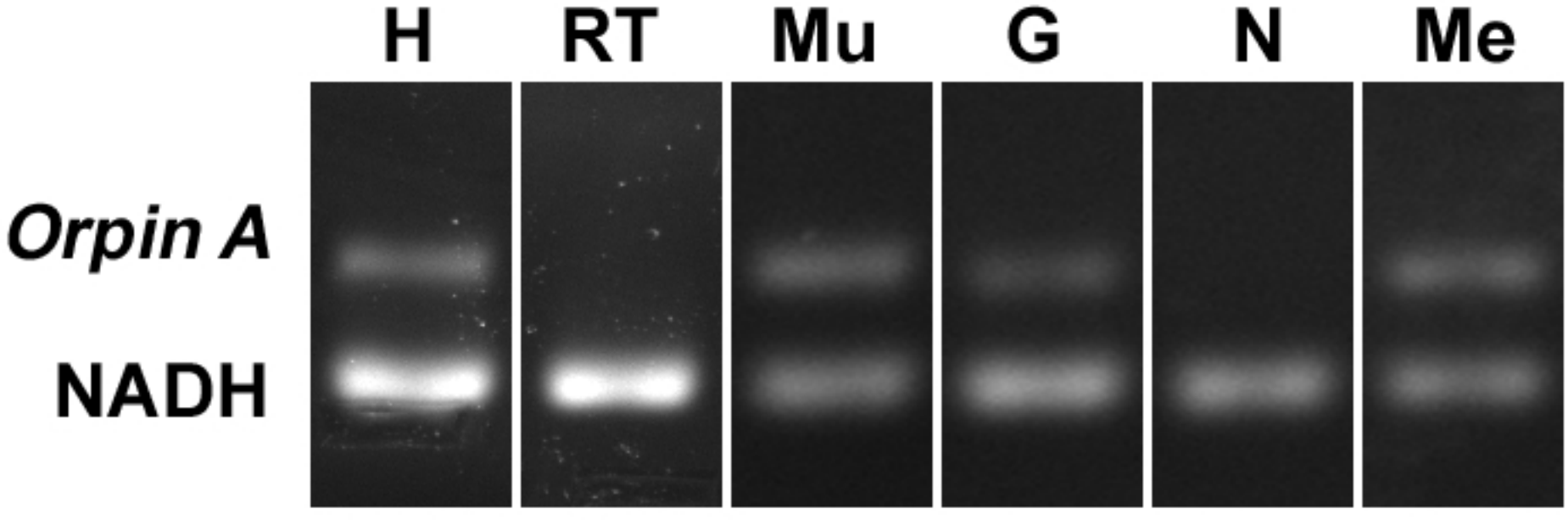

Figure 9

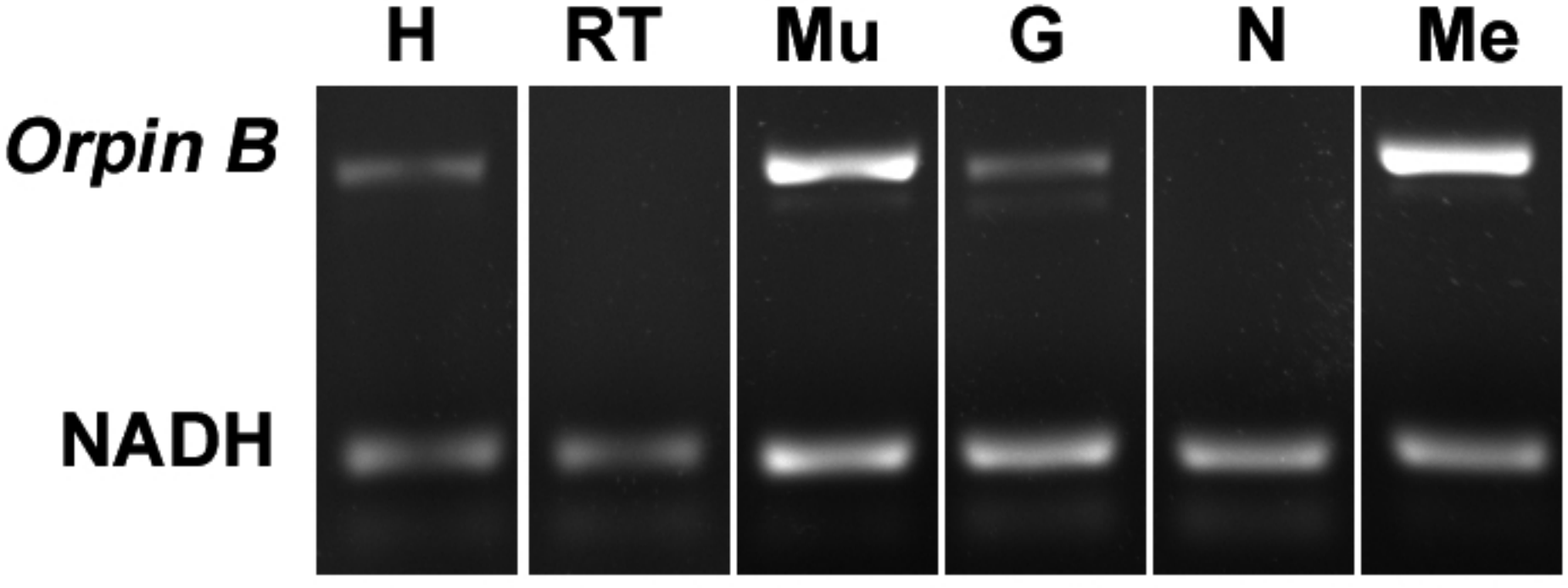

Figure 10

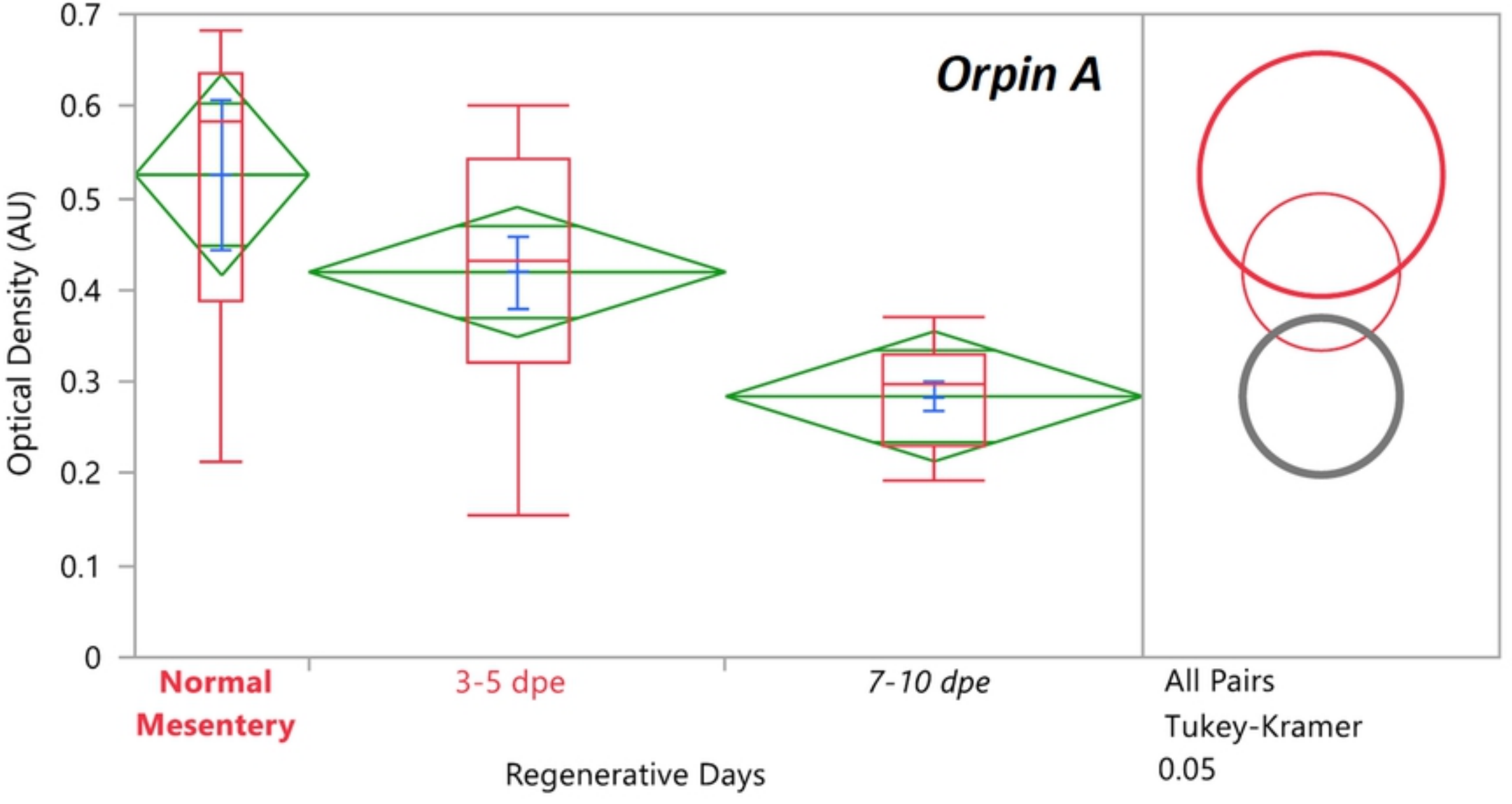
Orpin A
Optical Density (AU)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Normal Mesentery
3-5 dpe
7-10 dpe
All Pairs
Tukey-Kramer
0.05
Regenerative Days

Figure 11

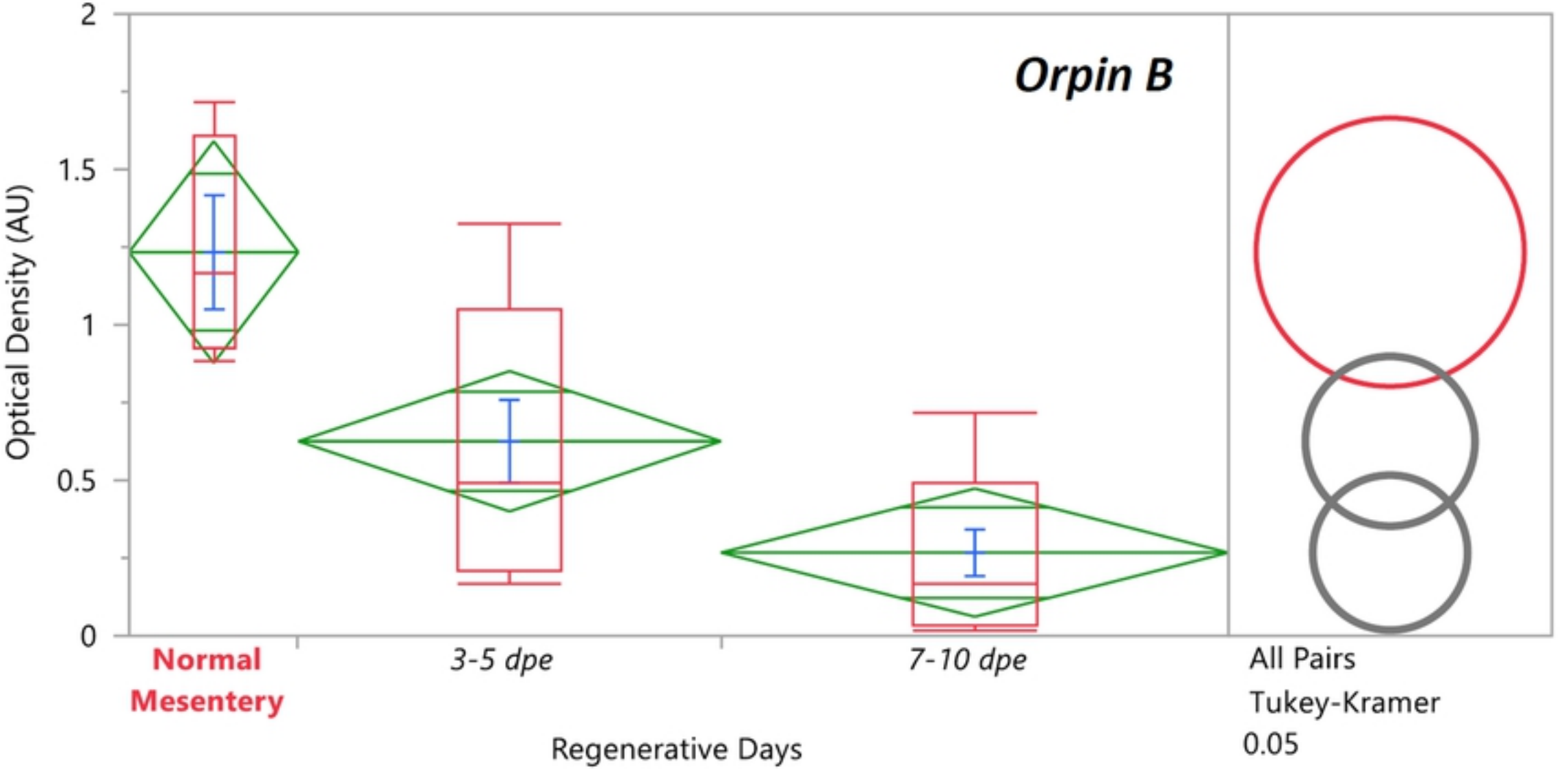
Orpin B
Optical Density (AU)
2
1.5
1
0.5
0
Normal Mesentery
3-5 dpe
7-10 dpe
Regenerative Days
All Pairs
Tukey-Kramer
0.05

Figure 12