Online Supplementary Materials

Title: Molecular cloning of metal-responsive transcription factor-1 (MTF-1) and transcriptional responses to metal and heat stresses in Pacific abalone, *Haliotis discus hannai*

Author: Sang Yoon Lee, Yoon Kwon Nam (Corresponding author)

Legends for Supplementary Materials

Suppl. Table S1.

Summarized information on oligonucleotide primers used in this study.

Suppl. Fig. S1.

Full-length cDNA and deduced amino acid sequences of abalone *Haliotis discus hannai* metal responsive transcription factor-1 (MTF-1). In the nucleotide sequence, stop codon is indicated by an asterisk and putative polyadenylation signal is underlined. On the other hand, in the amino acid sequence (in a singlet code), presumed nuclear localization signal (NLS) is underlined (in the front of the first zinc finger), while six putative C_2H_2-zinc fingers in the DNA binding domain are boxed.

Suppl. Fig. S2.

Multiple sequence alignments of abalone MTF-1 along with its representative orthologs. Putative NLS, six C_2H_2-zinc fingers, NES embedded in transactivation domain, and C-terminal Cys-cluster are indicated.

Suppl. Fig. S3.

Neighbor-joining phylogenetic tree based on amino acid sequences of zin-finger DNA binding domain and NLS region in MTF-1 orthologs. Tree was computed by poisson correction method using MEGA 7.0 program. Confidence level of each clade was evaluated with bootstrap testing (1,000 replicates).

Suppl. Fig. S4.

Putative transcription factor binding motifs predicted in the 5′-flanking upstream region of the abalone MTF-1 gene. Transcription factor sites were predicted with perfect and imperfect matches to consensus core sequences. TRANSFAC® search (GeneXplain GmbH, Germany) was carried out with the cut-offs scores > 0.99 (core score) and > 0.95 (matrix score). Abbreviations for transcription factors are as below.

AhR: aryl hydrocarbon receptor to bind xenobiotic response element (XRE; TNGCGTG)
HIF-1α: hypoxia-inducible factor-1 alpha to bind hypoxia response element (HRE; RCGTG)
CREBP: cAMP response element binding protein (TGACGY)
C/EBP: CCAAT-enhancer binding protein (TTDNGNAA)
E-box: E-box binding protein (CANNTG)
ER-h: estrogen receptor half site (AGGTCANNNTGACCT)
HNF: hepatocyte nuclear factor (TRTTKRYTY)
HSF: heat shock factor to bind heat shock element (HSE; GAANRTTC)
MTF-1: metal-responsive transcription factor-1 to bind metal responsive element (MRE; TGCRCNC)
NF-AT1: nuclear factor for activated T-cells (WGGAAA)
Smad4: Mothers against decapentaplegic homolog 4 (TGTCTGN)

Suppl. Table S1. Summarized information on oligonucleotide primers used in this study

Primer name	Sequence (5' to 3')	Thermal cycling conditions	Purpose
HHcMTF-1 1R	GCCAGGTGTACTGTAAGTTCTTGAAC	Initial denaturation step at 94°C for 3 m and 35 cycles (1st PCR and 2nd PCR) at 94°C for 30 s, 58°C for 30 s and 72°C for 30 s	5' RACE
HHcMTF-1 2R	ATTGAAACCTCTTCACCTCCTTGGCA		
HHcMTF-1 1F	TGCAGATGGCTGCCTGCTTTGTGCAC	Initial denaturation step at 94°C for 3 m and 35 cycles (1st PCR and 2nd PCR) at 94°C for 30 s, 58°C for 30 s and 72°C for 2 m	3' RACE
HHcMTF-1 2F	CTGATGGCAGATAGGAGGTATGTGAA		
HHcMTF-1 FW1	CAAGAGTTTACGTCCGGAAA	Initial denaturation step at 94°C for 2 m and 35 cycles at 94°C for 30 s, 58°C for 30 s and 72°C for 2 m	Isolation of full length cDNA
HHcMTF-1 RV1	CCAGATCAATACACAGACTA		
HHgMTF-1 FW1	TCTAAGATGGGGCACAATCT	Initial denaturation step at 94°C for 2 m and 35 cycles at 94°C for 30 s, 58°C for 30 s and 72°C for 2 m	Isolation of 5'-upstream region
HHgMTF-1 RV1	GAATCGGCCACTTGGTTTTG		
HHMTF-1 GW1R	GTACACTTGGGTCACTGCACTGCTGT	Initial denaturation step at 94°C for 2 m and 7 (1st PCR) or 5 (2nd PCR) cycles at 94°C for 25 s and 68°C for 3 m, followed by 32 (1st PCR) or 20 (2nd PCR) cycles at 94°C for 25 s and 72°C for 3 m, followed by a final elongation at 72°C for 7 m	Genome walking to 5'-upstream region
HHMTF-1 GW2R	TGGGTCCATCTTCGTCTACAAACGTC		
qHHMTF-1 FW1	CACATTCGTACACACACTGG	Initial denaturation step at 94°C for 2 m and 45 cycles at 94°C for 20 s, 60°C for 20 s and 72°C for 20 s	RT-qPCR assay (for MTF-1, MT and normalization genes)
qHHMTF-1 RV1	CATGTCTGTTCTTGTGAGAC		
qHHRPL5 FW	AGATGAGGATGGCAAACCAG		
qHHRPL5 RV	TCGCTGCTCTCAGAGTCAAA		
qHHRPL7 FW	CAAGCTGAACACTCCAAACG		
qHHRPL7 RV	TCCACAGCACTGATGTTTCC		
qHHRPL8 FW	TGGAAACTACGCCACAGTCA		
qHHRLP8 RV	GTCCTGCCTTCAACATTGGT		
qHHMT 1F	GGTACCGACTGCAAGTGTAA		
qHHMT 1R	TCATCGGAAGTCATGTGAGC		

```
   1  caagagtttacgtccggaaaattctgttcggcccattcggttttcgctATGGAATTTACTGATAGCAACGACAGCAATAAAGGA
                                                         M  E  F  T  D  S  N  D  S  N  K  G   12
  85  ACCGATTCTTTTGAAGATTTAATGACGTTTGTAGACGAAGATGGACCCACTGACAGCAGTGCAGTGACCCAAGTGTACATAGAA
         T  D  S  F  E  D  L  M  T  F  V  D  E  D  G  P  T  D  S  S  A  V  T  Q  V  Y  I  E   40
 169  AGACTAGGATCTCACTCAAAAGATGACCAAAACCAAGTGGCCGATTCGGGGTCAGATACCATCCCTCCTGGTATGGAGAACTGT
         R  L  G  S  H  S  K  D  D  Q  N  Q  V  A  D  S  G  S  D  T  I  P  P  G  M  E  N  C   68
 253  CAAAATGAAAGTTCTCTTTCAACCATCGAACATGAAGGGTACATACATAATACAATATCTGACGACCAGATATTGATGACCTTA
         Q  N  E  S  S  L  S  T  I  E  H  E  G  Y  I  H  N  T  I  S  D  D  Q  I  L  M  T  L   96
 337  AACCCTGGAAATGAGCGGATGCCTATCAACCCGTCACATGCGACCATTACACTGGAGACACAAGATCCATACACTAATGCCAAG
         N  P  G  N  E  R  M  P  I  N  P  S  H  A  T  I  T  L  E  T  Q  D  P  Y  T  N  A  K  124
 421  GAGGTGAAGAGGTTTCAATGTAACTTCCAAGACTGTTCAAGAACTTACAGTACACCTGGCAACCTGAAAACATCTGAAGACC
         E  V  K  R  F  Q  C  N  F  Q  D  C  S  R  T  Y  S  T  P  G  N  L  K  T  H  L  K  T  152
 505  CACCGAGGAGAATATACCTTTGTATGTGACCAGCACGGCTGTGGGAAGGCCTTCCTCACCTCATACAGTCTCAAAATACATGTT
         H  R  G  E  Y  T  F  V  C  D  Q  H  G  C  G  K  A  F  L  T  S  Y  S  L  K  I  H  V  180
 589  CGAGTACATACAAAAGAAAAACCTTATGAGTGCGACACCACCGGCTGTGAGAAGTCTTTCAACACACTTTACAGGTTACGTGCT
         R  V  H  T  K  E  P  Y  E  C  D  T  T  G  C  E  K  S  F  N  T  L  Y  R  L  R  A  208
 673  CACAAAGTCTTCACTCCGGGAATACTTTCAACTGTGATGAGAGTGGATGCACAAAGTACTTTACAACTTTGAGTGATCTTCGT
         H  K  R  L  H  S  G  N  T  F  N  C  D  E  S  G  C  T  K  Y  F  T  T  L  S  D  L  R  236
 757  AAACACATTCGTACACACTGGAGAGAAACCTTATGTGTGCAGTGAAACAGGCTGTCAAAAAGCTTTTGCTGCAAGTCACCAT
         K  H  I  R  T  H  T  G  E  K  P  Y  V  C  S  E  T  G  C  Q  K  A  F  A  A  S  H  H  264
 841  CTAAAAACACATTCTCGAACTCATTCAGGTGAGAAGCCATACACATGTTCCCAGGAGGGATGTCACAAGTCTTTCACAACCAAC
         L  K  T  H  S  R  T  H  S  G  E  K  P  Y  T  C  S  Q  E  G  C  H  K  S  F  T  T  N  292
 925  TATAGCCTTAAGTCTCACAAGAACAGACATGACAAGGGCGGAGGACAGTCTGATCCGTCAGGGACACACGAAGCAGCAGAGACA
         Y  S  L  K  S  H  K  N  R  H  D  K  G  G  G  Q  S  D  P  S  G  T  H  E  A  A  E  T  320
1009  CATGACAGTGGAGGGTCCATGACAGCTGAACAGCTCTTCAACACTATCTATGTGAATCCTACAAGTACGGACCATGTGAGCCTG
         H  D  S  G  G  S  M  T  A  E  Q  L  F  N  T  I  Y  V  N  P  T  S  T  D  H  V  S  L  348
1093  GACGAGGCTGCTCTGCAACAGACAGATGTTGTGCCAGGTATACAAACAGTTCCTGTTCAAGAAATACTGCAGCCAGTGATTCCT
         D  E  A  A  L  Q  Q  T  D  V  V  P  G  I  Q  T  V  P  V  Q  E  I  L  Q  P  V  I  P  376
1177  GTAGTGGATACTGGTGCATCAGGAGCTCCCACGCCCTCAGAGGGCAGTGGGTGTGTCCAGCATGTTATTCTAAACCAGTCAGCT
         V  V  D  T  G  A  S  G  A  P  T  P  S  E  G  S  G  C  V  Q  H  V  I  L  N  Q  S  A  404
1261  ATACCCACCCTCTCAGATGCAGCCACAGACTTCCTTCTTCCCAGCAGTCTCAGCAACTCCACACACACAGGCACACTTCCTGTC
         I  P  T  L  S  D  A  A  T  D  F  L  L  P  S  S  L  S  N  S  T  H  T  G  T  L  P  V  432
1345  AGCAACCGCTTGCAAGGGGAAGTAACCCAGCCCCAGAACATTGTCCCAGCCCCCACAACAGATGCAGTGCAGCTCCAGATTCAG
         S  N  R  L  Q  G  E  V  T  Q  P  Q  N  I  V  P  A  P  T  T  D  A  V  Q  L  Q  I  Q  460
1429  ACAGCTTCAGGCAGCACTGTTCCAGTCAGTCAGATCTTTGTGCCAGTGGTTTCTAACACGGACAAGGGTCCAGTCATAGAGCTT
         T  A  S  G  S  T  V  P  V  S  Q  I  F  V  P  V  V  S  N  T  D  K  G  P  V  I  E  L  488
1513  GTGCCACTTCAGAACAGCATTTCAGTGAACGATGAAAGCCGCACTgaaatgccacataaaccctatgcttcaagattttcacc
         V  P  L  Q  N  S  I  S  V  N  D  E  S  R  T  *                                      503
1597  tttttgtggaatttatgtgtttgataattttcatgttgtttgtcttcacaccccaggacaatgtgtgcatatttctatgtgtc
1681  agcttgttgcagatggctgcctgctttgtgcactgtgtttaatacctactatctactacacattcataacttgccaaatcttta
1765  gttcaaattgagccattcaccataaactgcttttgtcttcaggatgcatggttccatgtgtcacattactgatggcagatagga
1849  ggtatgtgaatacatctaaatgaaaacttgcctgaaaagtaaatgttctggaataatctacttacacatttcatttgcaatatt
1933  atctttattcattagcagctctgaaccttatgagtttccgctgtagttttgtgtttgaaagtgcagttttcggtttttcactgtg
2017  tgtaagatagtggttctagataacaatgtaagatcgtacaagttagtacaccaactgtatatagtctgtgtattgatctggggga
2101  ataaatacacttctttttttaaaaaaaaaaaaaaaaaaaa
```

Species abbreviations

Biomphalaria glabrata **(BG)**	XP_013093604.1	*Mus musculus* **(MM)**	NP_032662.3
Aplysia californica **(AC)**	XP_005098642.1	*Alligator mississippiensis* **(AM)**	KYO35612.1
Haliotis discus hannai **(HH)**	AMS38481.1	*Callorhinchus milii* **(CM)**	XP_007893150.1
Crassostrea gigas isoform-1 **(CG1)**	EKC32469.1	*Poecilia reticulata* **(PR)**	XP_008429346.1
Crassostrea gigas isoform-2 **(CG2)**	XP_011433432.1	*Xiphophorus maculatus* **(XM)**	XP_005796106.1
Crassostrea gigas isoform-3 **(CG3)**	Meng et al., 2015	*Oreochromis aurea* x *O. nilotica* **(OAN-1)**	AAP93663.1
Crassostrea gigas isoform-4 **(CG4)**	XP_011433433.1	*Takifugu rubripes* **(TR)**	NP_001027866.1
Octopus bimaculoides **(OB)**	XP_014784404.1	*Danio rerio* **(DR)**	AJF36548.1
Homo sapiens **(HS)**	EAX07312.1	*Lingula anatina* **(LA)**	XP_013403275.1

```
BG   -----------------------------------------------------------------------------------------------MEPYSDFESDS--NFVDFLQFED--
AC   -----------------------------------------------------------------------------------------------MEPYTDFVNDSNMNFDDFLHFEVED
HH   -----------------------------------------------------------------------------------------------MEFTDSNDSNKGTDSFEDLMTFVD--
CG1  ------------------------------------------------------------------------------------------------------------------MDDTSNGL
CG2  ------------------------------------------------------------------------------------------------------------------MDDTSNGL
CG4  ------------------------------------------------------------------------------------------------------------------MDDTSNGL
CG3  ------------------------------------------------------------------------------------------------------------------MDDTSNGL
OB   MEISSFVTMEEKNSPRPCTSTDPNRENAAIISATNSNHTPSNNDNSIANSSDRAGGGGGGGSCGGGSCGCVVAGGYYHKNEEDRHDITQGQSHYHHHHHNGEDEDGDDDDDDDERLPFISYVGENPIGYDE
HS   -----------------------------------------------------------------------------------------MGEHSPDNNIIYFEAEEDELTPDDKMLRFVDKNGLV
MM   -----------------------------------------------------------------------------------------MGEHSPDDNIIFFKGEEDDLTPHDKMLRFVDDNGLV
AM   ---------------------------------------------------------------------------------------------------MMRFVDKNGLV
CM   --------------------------------------------------------------------------MGDNDPEKGELLSENNNETYYETEDDDEEEMTREERMGFSGKNGLI
PR   ----------------------------------------------------------------------MSENGPRTEAPMYFEVEVDQLERE---DEEEDGKIHYDKDDDLIAEP
XM   ----------------------------------------------------MVLLAKEGKATGTMSENGPRTEAPMYFKVEVDQLERE---DEEEDDKIHYDKDDDLIAEP
OAN1 -----------------------------------------------------------------------MSENGPHTEAPMYFEVEVAPLERD---DEE-DDKIHFGKDGDLIAEP
TR   -----------------------------------------------------------------------MSGNGPHSEVPMYFEVEVGHLDQED--EEEEDKIHFSKGDDLIPET
DR   -----------------------------------------------------------------------MGENGPLSETAMLFEDEEEEDVDKLGREEEEDEDKMANFDKDGNMISGP
LA   --------------------------------------------------------------------------------------------------------MAEVGE
```

```
BG   DENMEDMSSITQVCVDEDPEEDTDDKTNRSRSITRSKSEGDISFESDGSKKTSEINDRATRSRSMKSNASNS-------------------NFIIENDIDEINDNDHDGYIHHTISEDEILMQINPGSS
AC   DENLQDMSSITQVCVDHDEDEPPDDGQIHVSEGVCNKEQGGCSSSSRDDNSDFEARYTQEEGVKMEHTITS-------------------------------DNGDGYIHHTISEDEILMQINPGSS
HH   EDGPTDSSAVTQVYIERLGSHSKDDQNQVADSGSDTIPPGMENCQNESSLSTIEH-------------------------------EGYIHNTISDDQILMTLNPGNE
CG1  GMDPEDFTSVTQVYMEYDN----DDDCSIGNPTCSRFDDHGHSFGNDNSMSASEHDKQFD------------------------------------------DDE-------RIEMSINPGMR
CG2  GMDPEDFTSVTQVYMEYDN----DDDCSIGNPTCSRFDDHGHSFGNDNSMSASEHDKQFD------------------------------------------DDESRYIHNTISDDRIEMSINPGMR
CG4  GMDPEDFTSVTQVYMEYDN----DDDCSIGNPTCSRFDDHGHSFGNDNSMSASEHDKQFD------------------------------------------DDESRYIHNTISDDRIEMSINPGMR
CG3  GMDPEDFTSVTQVYMEYDN----DDDCSIGNPTCSRFDDHGHSFGNDNSMSASEHDKQFD------------------------------------------DDESRYIHNTISDDRIEMSINPGMR
OB   EDEENDMSSMTAVFIETDMDEEDYGAMVKFKVEGEEDEEGNESGERKRGEAGAGCDDNKVDERQTGDESDDCQRQQRRRLLLPHLFDVRKGGDGNIIQTSTANLETSVGGYIHNTISEHQILMQINPGDS
HS   PSSSGTVYDRTTVLIEQDPGTLEDEDDD-------GQCGEHLP---FLVGGEEG--FHLIDH------------------------------------------EAMSQGYVQHIISPDQIHLTINPGST
MM   PSSSGTVYDRTTVLIEQDPGTLEDEDDD-------GQCGEPLP---FLVEGEE---GFLIDQ------------------------------------------EAMSQGYVQHIISPDQIHLTINPGST
AM   PSSSGTVYDRTTVLIEQDHGILEDDEDE-------GQCGDHLS---FLPEGHEEGFHLIADH------------------------------------------EGMSQGYVQHIISPDQIHLTINPGST
CM   SSSPGTVYDRTTVLIEQDPISLEEEEEEDEEEEEGGNNNSLVS---FLSHGGEDNFFLIGDP------------------------------------------DGMSQGYVHHTISPDQIQFTINPGST
PR   SSSSGRVYDRTTVLIERDPIRLDEEGEEEGHCGGDEDGVTFLTEGEGDADEEEGSLAFMADP------------------------------------------DGMSQGYVHHTISPDQIQFTINPGST
XM   SSSSGRVYDRTTVLIERDPIRLDEEGEEEGHCGGDEDGVTFLTEGEGDADEEEGSLVFMADP------------------------------------------DGMSQGYVHHTISPDQIQFTINPGST
OAN1 SSSSGRVYDRTTVLIERDPIRLDEEGEEEGHCGGDEDGVTFLTEGEGDGDEEEGSLAFMTDP------------------------------------------DGMSQGYVHHTISPDQIQFTINPGST
TR   SSPSGRLYDRATVLIERDPIRLDEEGEEGHCGGEDDGVTFLTEGEGDGDEEEGTLTFINDP------------------------------------------DGMSQGYVHHTISADQIQFTINPGST
DR   SSSSGTVYDRTTVLIEQDPIRLDEEGEE--------DGMTFLPEGEGD----EGSLAFMGDP------------------------------------------DGMSQGYIHHTISPDQIQFIINPGST
LA   NSSNGTTEENVSQAMRYEFMKFDDDAEDG-----YEDTSHSTPVFIEVDDSDDSQPQFQDRG------------------------------------------GEEHEGYIHHTISDNQIMMHIHPGSN
```

aNLS | 1st ZnF C₂H₂ | 2nd ZnF C₂H₂ | 3rd ZnF C₂H₂ | 4th ZnF C₂H₂

```
BG   KMPS--NPSHATLTIESQNPRTKAKEVTRFKCTFSGCARTYSTPGNLKTHEKTHRGEYTFVCSEMGCGKRFLTSYSLKIHVRVHTNEKPYECDKPGCEKSFNTIYRLRAHERLHTGETFKCGSDGCTKYF
AC   RLPL--NPSHATLTVESQNPRTKAKEVTRFKCTFAGCARTYSTQGNLKTHEKTHRGEYTFVCNESSCGKRFLTSYSLKIHVRVHTNEKPYECDISGCEKSFNTIYRLRAHKRLHTGETFKCESDGCTKYF
HH   RMPI--NPSHATILETQDPYTNAKEVKRFQCNFQDCSRTYSTPGNLKTHLKTHRGEYTFVCDQHGCGKAFLTSYSLKIHVRVHTKEKPYECDTTGCEKSFNTLYRLRAHKRLHSGNTFNCDESGCTKYF
CG1  GVPANFNPSHATITVETCDPDTKEKEYRRFQCDYKGCTRTYSTAGNLRTHQKTHKGEYTFICDQHGCGKAFLTSYSLKIHVRVHTKEKPYECEVKGCAKNFNTLYRLRAHQRIHTGDTFDCNEDGCTKFF
CG2  GVPANFNPSHATITVETCDPDTKEKEYRRFQCDYKGCTRTYSTAGNLRTHQKTHKGEYTFICDQHGCGKAFLTSYSLKIHVRVHTKEKPYECEVKGCAKNFNTLYRLRAHQRIHTGDTFDCNEDGCTKFF
CG4  GVPANFNPSHATITVETCDPDTKEKEYRRFQCDYKGCTRTYSTAGNLRTHQKTHKGEYTFICDQHGCGKAFLTSYSLKIHVRVHTKEKPYECEVKGCAKNFNTLYRLRAHQRIHTGDTFDCNEDGCTKFF
CG3  GVPANFNPSHATITVETCDPDTKEKEYRRFQCDYKGCTRTYSTAGNLRTHQKTHKGEYTFICDQHGCGKAFLTSYSLKIHVRVHTKEKPYECEVKGCAKNFNTLYRLRAHQRIHTGDTFDCNEDGCTKFF
OB   PMPLN--PTHATLTIERQDPQTQAKEVKRFRCTFTGCTRTYSTAGNLRTHQKTHKGDYQFLCNLEGCGKTFLTSYSLKTHVRVHTKEKPYECDMKGCAKAFNTLYRLRAHQRLHTGDTFNCDENGCTKYF
HS   PMPR--NIEGATLTLQSECPETKRKEVKRYQCTFEGCPRTYSTAGNLRTHQKTHRGEYTFVCNQEGCGKAFLTSYSLRIHVRVHTKEKPFECDVQGCEKAFNTLYRLKAHQRLHTGKTFNCESEGCSKYF
MM   PMPR--NIEGATLTLQSECPETKRKEVKRYQCTFEGCPRTYSTAGNLRTHQKTHRGEYTFVCNQEGCGKAFLTSYSLRIHVRVHTKEKPFECDVQGCEKAFNTLYRLKAHQRLHTGKTFNCESQGCSKYF
AM   PMPR--NIEGATLTLQSECPETKLKEVKRYQCTFEGCPRTYSTAGNLRTHQKTHRGEYTFVCNQEGCGKAFLTSYSLKIHVRVHTKEKPFECDVQGCEKAFNTLYRLKAHQRLHTGKTFNCDTQGCSKYF
CM   PMPR--NIEGATLTLHSECPDTKLKEVKRYQCTFEGCTRTYSTAGNLRTHQKTHRGEYTFVCNQQGCGKAFLTSYSLKIHVRVHTKEKPFECDVQGCEKAFNTLYRLKAHQRLHTGKTFNCESEGCTKYF
PR   RMPR--NIEGATLTLHSECPETKQREVKRYQCTFEGCTRTYSTAGNLRTHQKTHRGEYTFVCNQQGCGKAFLTSYSLKIHVRVHTKEKPFECDVQGCEKAFNTLYRLKAHQRLHTGKTFNCESEGCTKYF
XM   PMPR--NIEGATLTLHSECPETKQREVKRYQCTFEGCTRTYSTAGNLRTHQKTHRGEYTFVCNQQGCGKAFLTSYSLKIHVRVHTKEKPFECDVQGCEKAFNTLYRLKAHQRLHTGKTFNCESEGCTKYF
OAN1 PMPR--NIEGATLTLHSECPETKQREVKRYQCMFEGCTRTYSTAGNLRTHQKTHRGEYTFVCNQQGCGKAFLTSYSLKIHVRVHTKEKPFECDVQGCEKAFNTLYRLKAHQRLHTGKTFNCESEGCTKYF
TR   PMPR--NIEGATLTLHSECPETKQRENNRYKCMFEGCTRTYSTAGNLRTHQKTHRGEYTFVCNQQGCGKAFLTSYSLKIHVRVHTKEKPFECDVQGCEKAFNTLYRLKAHQRLHTGKTFNCESEGCTKYF
DR   PMPR--NIEGATLTLHSECPETKQRENKRYQCLFEGCTRTYSTAGNLRTHQKTHRGEYTFVCNQQGCGKAFLTSYSLKIHVRVHTKEKPFECDVQGCEKAFNTLYRLKAHQRLHTGKTFNCESEGCTKYF
LA   RMPS--NPSHATLTVESRNPETKATEVKRYCCSFENCDRTYSTAGNLKTHQKTHKGEYTFVCNQENCGKSFLTSYSLKIHVRVHTKEKPYECSISNCEKAFNTLYRLKAHQRLHTGNTFNCEASDCSKAF
```

5th ZnF C₂H₂ | 6th ZnF C₂H₂

```
BG   TTLSDLRKHIRTHTGEKPFICNENGCGKAFAASHHLKSHNRIHTGDKPYECTQDGCCKAFTSVYSLKSHVSKHGKESEKG----------SQLSSTKGCSGGCHD-----DSCENMSTLQNVILVNP---
AC   TTLSDLRKHIRTHTGEKPFVCHENGCGKAFAASHHLKSHNRIHTGGRPFECTQDGCLKAFTSIYSLKSHISRHERDSEKDKGQQPSSEKKSTAGGGQCSGECHGNDNSADTGPDIVNFQNVVLVQQGDV
HH   TTLSDLRKHIRTHTGEKPYVCSETGCGKAFAASHHLKTHSRTHSGEKPYTCSQEGCHKSFTTNYSLKSHKNRHDKGGGQS------------------------------------DP
CG1  TTLSDLRKHIRTHTGEKPYQCDENGCGKAFAASHHLKTHQRTHTGEKPYTCQEDGCSRAFSTSYSLKTHKSKHEKSLG----------------------------DQTEGILSTLGLFEMGEQ
CG2  TTLSDLRKHIRTHTGEKPYQCDENGCGKAFAASHHLKTHQRTHTGEKPYTCQEDGCSRAFSTSYSLKTHKSKHEKSLGDQNS-------KDDTATIYNEDGRFQYNISDSVPHRTEGILSTLGLFEMGEQ
CG4  TTLSDLRKHIRTHTGEKPYQCDENGCGKAFAASHHLKTHQRTHTGEKPYTCQEDGCSRAFSTSYSLKTHKSKHEKSLGDQNS-------KDDTATIYNEDGRFQYN-----ISDTEGILSTLGLFEMGEQ
CG3  TTLSDLRKHIRTHTGEKPYQCDENGCGKAFAASHHLKTHQRTHTGEKPYTCQEDGCSRAFSTSYSCKTHKSKHEKSLGDQNS-------KDDTATIYNEDGRFQYN-----ISDTEGILSTLGLFEMGEQ
OB   TTLSDLRKHIRTHTGERPYKCSENGCGKAFAASHHLKTHTRTHTDGIPELNSKPRIIF-----------------------------------------
HS   TTLSDLRKHIRTHTGEKPFRCDHDGCGKAFAASHHLKTHVRTHTGERPFFCPSNGCEKTFSTQYSLKSHMKGHDNKGHSYNALPQHNGSEDTNHSLCLSDLSLLSTDSELRENSSTTQGQDLSTISPAII
MM   TTLSDLRKHIRTHTGEKPFRCDHDGCGKAFAASHHLKTHVRTHTGERPFFCPSNGCEKTFSTQYSLKSHMKGHDNKGTAYSALPQHNGSEDTNHSLYLSELGLLSTDSELQENSSSTQGQDLSTISPAII
AM   TTLSDLRKHVRTHTGEKPFRCDHDGCGKAFAASHHLKTHVRTHTGERPFFCPTDGCEKTFSTQYSLKSHMKGHEKG-HSYNALPNSGVSEDTNHSLCLSDLSLMSTDSELRENANTTQGQDLSTISPASI
CM   TTLSDLRKHIRTHTGEKPFRCDHNGCGKAFAASHHLKTHVRTHTGERPFLCPSDGCEKTFSTQYSLKSHMKGHDKENPCIVSTQN-SLNEETNQSLCLSDLSLMSTDSELRENSHSNNRRDLSTISAESI
PR   TTLSDLRKHIRTHTGEKPFRCDHDGCGKAFAASHHLKTHVRTHTGEKPFNCPSDGCEKTFSTQYSLKSHIRGHDKG-QPFTVTLTRPHSEDANHSLCLSDLSLISTDSELRENINNAQNLDLNNVTPIKI
XM   TTLSDLRKHIRTHTGEKPFRCDHDGCGKAFAASHHLKTHVRTHTGEKPFNCPSDGCEKTFSTQYSLKSHIRGHDKG-QPFTVTLTHPHSEDANHSLCLSDLSLISTDSELREINNAQNLDLNNVTPIKI
OAN1 TTLSDLRKHIRTHTGEKPFRCDHDGCGKAFAASHHLKTHVRTHTGEKPFNCPSDGCEKTFSSQYSLKSHIRGHDKG-QFSVSLTHPLSEDANHSLCLSDLSLISTDSELREINHNAQNLDLNNTTPVKI
TR   TTLSDLRKHIRTHTGEKPFRCDHDGCGKAFAASHHLKTHVRTHTGEKPFNCPSDGCEKTFSSQNSLKSHIRGHDKV-QPFTVTLTHPISEDANHSLCLSDLSLISTDSELRENLNNAQDLDLNNMTPVKI
DR   TTLSDLRKHLRTHTGEKPFRCDHDGCGKAFAASHHLKTHVRTHTGEKPFNCPSDGCEKTFSSQYSLKSHIRGHDKG-PSFVSS-HPLSMMPEQEKFQQQLKQSQQQQQSPGQGASSAASADMQTPCISSQNL
LA   TTLSDLRKHLRTHTGEKPYKCEEGGCGKAFAASHHLKTHIRTHTGEKPYHCPTDGCSKAFTTQYGLKTHVGRHEQSAGSESDSSNQMMPEQEKFQQQLKQSQQQQQSPGQGASSAASADMQTPCISSQNL
```

aNLS
NES
acidic domain

```
BG    NPGLEPVKNEPEDIPLSVEQVIKSMYSNQIVSENYSQEITSGEDSVSSG----------------------------MVQEILRPVISLGTSPQQNSVCLDMSNQSVAMLPSTLQVGAVNSLVQTPQVG
AC    SSSLPAESLQQGKVTLSAEQLFNSLYTTPVLNENLVQHGKGDRDCVGCD----------------------SVQEIMQPVIPLESCGEPQPIPIGLMTSSVPVCHSHSAGHLVAPTVSAQHIL
HH    SGTHEAAETHDSGGSMTAEQLFNTIYVNPTSTDHVSLDEAALQQTDVVPGIQTV----------------------PVQEILQPVIPVVDTGASG
CG1   NNSNTNDYVDRPAITEIVDELKDTDNCNMCVYNHTHPSVMGDTDVNYSNQNQQP--------------------PHHQVFTELK-KDNPGSQQPIKDMVQLTPVPVIPINLPVVQDIHSLQNTAPN
CG2   NNSNTNDYVDRPAITEIVDELKDTDNCNMCVYNHTHPSVMGDTDVNYSNQNQQP--------------------PHHQVFTELK-KDNPGSQQPIKDMVQLTPVPVIPINLPVVQDIHSLQNTAPN
CG4   NNSNTNDYVDRPAITEIVDELKDTDNCNMCVYNHTHPSVMGDTDVNYSNQNQQP--------------------PHHQVFTELK-KDNPGSQQPIKDMVQLTPVPVIPINLPVVQDIHSLQNTAPN
CG3   NNSNTNDYVDRPAITEIVDELKDTDNCNMCVYNHTHPSVMGDTDVNYSNQNQQP--------------------PHHQVFTELK-KDNPGSQQPIKDMVQLTPVPVIPINLPVVQDIHSLQNTAPN
OB    ------------------------------------------------------------------
HS    FESMFQNSDDTAIQEDPQQ------TASLTESFNGDAESVSDVPPSTGNSASLS----------------------LPLVLQPGLSEPPQPLLPASAPSAPPPAPSLGPGSQQAAFGNPPALLQPPE
MM    FESMFQNSDDPGIQDDPLQ------TAALIDSFNGDAESVIDVPPPAGNSASLS----------------------LPLVLQSGISEPPQPLLPATAPSAPPPAPSLGPGSQPAAFGSPPALLQPPE
AM    FESMFQSPDHTANQEDSQQ------TAALIESFNGEDDAATTVQSSVGDSASLS----------------------LPLS
CM    FESMFHNTDSCTSHDGSQARAEDGLTGTYGSDVPPAPVVNPIVAEPGSLSFTIP----------------------LQTVSQPSVESHSQPSTTAAALPHITLNAAVISNTHQTVFGNPTSVVQTPQ
PR    FELMFQSPENSISQDDPHPTKNLGETFCIETSTQPGVTDVSSVISFSVNPTSSSSCSNNAAVLEASSQHIQNSSPQAHTPVPVTASVSSTQNPPFMQLSGAQQISDVPPAPVQAPNHVTSQHYVALPPTF
XM    FELMFQSPENSISQDDPHPTKNLGETFCIETSIQPRVTDVSSVISFSVNPTSSSSCSNNAAVLEASSQHIQNSSPQAHTPVPVTASVSSTQNPPFMQQSGAQQISDVPPAPVQAPNHVTSQHYVALPPTF
OAN1  FELMFQSPENSLSQDDAHLGESLAERFSLETPTQPGVTETSSLISFSVNPTSSSSCSRNAAVMEASSQLNQSSSSQTSTPASITANVSTTQPPAFMQLNGPQQISDVPQASVQPQNHVTPQQYVALPPTF
TR    FERMFQSPENSLSQDDAHPKESLAETFSLKNSTKPIATDASSLISFSLDPTASS----------------------THTTTVNSTPALPFMQLTGPQ-----QPSVQANHIGPQHYVSLSPTF
DR    FELMFQSPENSVSEDDPKPTESLAESFGLEPSPQTAPADASTHPAFPQPPPSTCSSS----------------------CSITAPAQDAQTPPTTQQAPPPAVSSSSQTSSFPSAPPSSSQPAEVS--S
LA    AGMSAEQLLTSMYLG-------QQPVAVNLPGIGPVVVAACSVDALQAANVTTHP----------------------QPVESVDENPSAGTQPQQ
```

← → proline-rich domain

```
BG    ILQSKPGEQGTVSPNVHLVGAVN-----------
AC    GSQLAVEAQGGVSQGTSLLLDTQGGVSQGTSLLLDTQGGVSQGTSLLLDTQGGVSQGTSLLLDTQGGMSQGTSVLLENQGGVSQGPSVLLETQGCVSQGQPVLLENQGGVSQGTSVLLESGGSSAILGTG
HH    -----------
CG1   LQPLDTQPMSIPAVQTTPKGDSS-----------
CG2   LQPLDTQPMSIPAVQTTPKGDSS-----------
CG4   LQPLDTQPMSIPAVQTTPKGDSS-----------
CG3   LQPLDTQPMSIPAVQTTPKGDSS-----------
OB    -----------
HS    VPVPHSTQFAANHQEFLPHPQAP-QP-----------IVPGLSVVAGASA
MM    VPVPHSTQFAANHQEFLPHPQAPPQT-----------IVPGLSVVAGAPA
AM    -----------
CM    VPVAQTAQFSANQQEFIQPTQMP-----------QP
PR    LQSESATQTSPLPQPVSAPALT-----------
XM    LQSESATQTSPLPQPISSPALT-----------
OAN1  LQQDSTTQTTPLPPPITTPAPAP-----------ALTAT
TR    QHKESTIRTTSLQPPIPAPPPVAS-----------VLTATT
DR    PSAPSATQHYMMAQPVSSPSAAS-----------
LA    LAVNNSDLSTLAAEAAEIFSQVP-----------
```
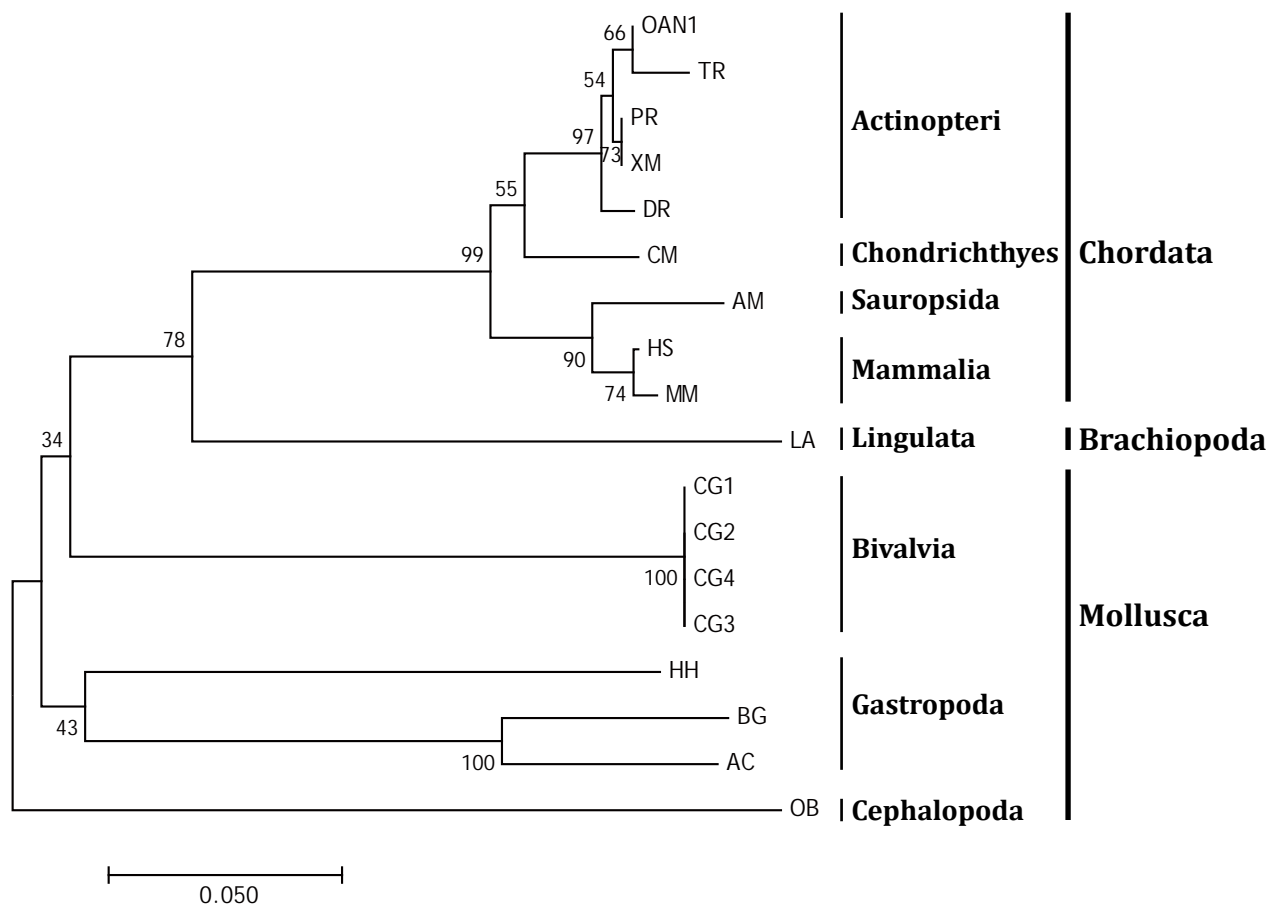
→

```
BG    -----ANEAADTLTLPAGISDLNMLTSELVKSLLGN-NTVMVQSGDGGLVQIPTSVLLRNDSVSSLAP-CSAACAMLDHSSLTASHSMAANSGTLGSMPGLG-----------
AC    GVEAGGSTTTATTVPDPNSLANIQLLRSELVKNILGTSNTIVVQTADGSLVQVPSSALLDQAGLASLVPGCLGANQTLAAANNSVCLPAAACHSTFLELSKPSEGGEVIASTMAANTALNGLSELVAGSS
HH    -----------
CG1   -------LQTVTALPVLQSGSDVVNSSSMFVEHYLIT----------KVIEKTPAGQQVKSESTLQLTP-----------
CG2   -------LQTVTALPVLQSGSDVVNSSSMFVEHYLIT----------KVIEKTPAGQQVKSESTLQLTP-----------
CG4   -------LQTVTALPVLQSGSDVVNSSSMFVEHYLIT----------KVIEKTPAGQQVKSESTLQLTP-----------
CG3   -------LQTVTALPVLQSGSDVVNSSSMCVEHYLIT----------KVIEKTPAGQQVKSESTLQLTP-----------
OB    -----------
HS    SAAAVASAVAAPAPPQSTTEPLPAMVQTLPLGANSVLTNNPTITITPTPNTAILQSSLVMGEQNLQWIL-----------
MM    SAATVASAVAAPAPPQSTTEPLPAMVQTLPLGANSVLTNNPTITITPTPNTAILQSSIVMGEQNLQWIL-----------
AM    ----------------TAILQSSIVMGEQNLQWIL-----------
CM    PSQGFPVMPSNSLPPSGNTEPLPTLVQTVPLG-----TNPPTITITPSQNTTILQPSIVMADPNLQWII-----------
PR    ---SAAPASAAVTADGLSAAPQPVPLANSAVTNTGPGQATTPATITVASTPNVLQPSIVMSDQNFQWIL-----------
XM    ---SAAPVSAASTADGLSAAPQPVPLANSAVTNTGPGQASAPATITIASTPNVLQPSLVMSDQNLQWIL-----------
OAN1  APGPGPAVVAATTTDALAPVAQPVPLANNPGPNSGPGLATTPATITIAPTQNLLQPSLVMSDQNLQWIL-----------
TR    ASTDAVSLVAAATTDTLAAVAQPVPLVNHPVPESGASFPTTSATYTVTPTHNLLQPNLVMSDQNLQWIL-----------
DR    -----VSSVPAGTAEVTAAVTHTVPLAAPPTISIAPTLG-------------LQPSLVMSDQNLQWIL-----------
LA    -----------------ENIEPSVHVSDGQEVAQVQ-----------
```

← → serine/threonine-rich domain

```
BG    -------MVTSQNPSTETSVLLPTTPAYSVAHVLPTLQSGPIQTVADLVTLKGLPSQGVPSIQLNSQPNVSFSFNSDNLDQSCGASFSSLTLMDPAACSAVPSMDEASTSVSLNLTDNSVPPVILKCPPV
AC    GGASQNFLLQPSGGSSSSSEATSGVFLPPTTMLVPNVASQSGQQQHLLRGQEGMGSGLSTAIFEVAATGQAVAPAAAVVGSIPNQTVVSNSVKRESNDGSGGSSLSPLTFVDPAAMSASQPNVDLTASSL
HH    ----------------APTPSEGSGCVQHVILNQSAIPTLSDAATDFLLPSSLSNSTHTGTLPVSNRLQGEVTQPQNIVPAPTTDAVQLQIQTASG--
CG1   ----------SLPPTQVIDAPQSVIDLSTTGTGIGLNPHQTQQPAQFINTNAANENPQILEKSNTDYMSMSAPAPTIMMPSASALPPQVQSLQSQHG--ILDDFSSNC
CG2   ----------SLPPTQVIDAPQSVIDLSTTGTGIGLNPHQTQQPAQFINTNAANENPQILEKSNTDYMSMSAPAPTIMMPSASALPPQVQSLQSQHG--ILDDFSSNC
CG4   ----------SLPPTQVIDAPQSVIDLSTTGTGIGLNPHQTQQPAQFINTNAANENPQILEKSNTDYMSMSAPAPTIMMPSASALPPQVQSLQSQHG--ILDDFSSNC
CG3   ----------SLPPTRVIDAPRSVIDLSTTGTGIGLNPHQTQQPAQFINTNAANENPQILEKSNTDYMSLSSPAPTIMMPSANALPPQVQSLQSQHG--ILDDFSSNG
OB    -----------
HS    -----------------------NGATSSPQNQEQ--IQQASKVEKVFFTTAVPVASSPGSSVQQIGLSVPVIIIKQEEACQCQCACRDSAKERASSRRKGCSSPPPPEPSPQAPDGPSLQLPAQTFSS
MM    -----------------------NGATSSPQNQEQ--IQQASKVEQVFATAVPVASGTGSSVQQIGLSVPVIIIKQEEACQCQCACRDSAKERAAGRRKGCSSPPPPEPNPQPPDGPSLQLPP-----
AM    -----------------------NGATSSPQNQEQ--MPQQVPKVEKVFFTTALPVAGNTGNSVQQI--SVPVIIIKQEEACQCQCACRDAAKDKATSKKRCSSPELKPSEKPDAKLQPQTFPSSS----
CM    -----------------------NGATVQQNSEH--MQQGPKVEKVFFTTAIPVSGNTGGSVQQIGLSVPVIIIKQEESCQCQCACRDSNKQKVPSSKESSTQEAGSSREPTTVQEQSRPEAQTFTST
PR    -----------------------SSAANSQQNPEQASHQGAPKVEKVFFTTAIPVGGNAGNSVQQIGLSLPVIIIKQEESCQCQCACRDSVKDKTAKSASSIISAPSPQQTVEPTPQLSSEPPP--HST
XM    -----------------------SSAANSQQNPEQASHQGAPKVEKVFFTTAIPVGGTAGNSVQQIGLSLPVIIIKQEESCQCQCACRDSVKDKTAKSASSVISAPSQQQTVEPTPQLSSEPPP--HST
OAN1  -----------------------SSAANSQQNPEQAAHQGAPKVEKVFFTTAIPVGGNAGNSVQQIGLSLPVIIIKQEESCQCQCACRDSAKDKSAKSASSIVSAPAQQQQPEPPSPPLSEPQH--HSA
TR    -----------------------STAANSQQNAEQQG-APKVEKVFFTTAIPMGGNAGNSVQQIGLSLPVIIIKQEESCQCQCACRDSAKEKNSKSSSSSMSAQKEPYTSKPLPSPLSLQPEPQHNP
DR    -----------------------SSAASAQQNPE----QQGPKVEKVFFTTAIPVQGNSGNAVQQIGLSLPVIIIKQEESCQCHCARDSAKDKSTSSSQDKTKNTSPPPPPAPPPPPPALPKE-----
LA    -----------------------DKDRNSEDSIEDALNLFSQFLVQNSGGTGPGVPACVPCTENPTMDTPPIPMTDSQGQLIPQCLRHHQGGQMAMQGNGQMMGQGGAWLCHGTEHGDTDTAYD-----
```

← → cysteine cluster

```
                                                                                                          Accession #      AA no.    MW
BG    AHTHVASSGAACPKSCTPSPSLANS----AAVSGLALAPDGTLISAIQKGIPISMLTGTYPDSIVLNQVFVPIYSNTDKGPIIELVPIKTNT--------      XP_013093604.1    802   85886.9
AC    MSDSCCHASSSSSEVFNSTVSTVSSGMVQAGPSNLVLPEDGTLTSAIQKGIPVSILPGANPESIVLNQVFVPVSNTDKGPVIELVPLQSNISVNDESRT        XP_005098642.1    962   100524.3
HH    --------STVPVSQIFVPVVSNTDKGPVIELVPLQNSIVNDESRT                                                             AMS38481.1        503   54861.3
CG1   VPRQENANKTLVTPAEFNVRMEHIPAIIAANPDKKITILPSTDGAMQPSIVKISTDKGYIELPCGMEGKLVPIVSTSNFTFSSTNAINTT----------      EKC32469.1        670   74046.4
CG2   VPRQENANKTLVTPAEFNVRMEHIPAIIAANPDKKITILPSTDGAMQPSIVKISTDKGYIELPCGMEGKLVPIVSTSNFTFSSTNAINTT----------      XP_011433432.1    708   78474.0
CG4   VPRQENANKTLVTPAEFNVRMEHIPAIIAANPDKKITILPSTDGAMQPSIVKISTDKGYIELPCGMEGKLVPIVSTSNFTFSSTNAINTT----------      Meng et al., 2015 707   77851.3
CG3   VPRQENANKTLVTPAEFNVRMEHIPAIIAANPDKKITILPSTDGAMQPSIVKISTDKGYIELPCGMEGKLVPIVSTSNFTFSSTNAINTT----------      XP_011433433.1    703   77897.4
OB                                                                                                             XP_014784404.1    446   49740.2
HS    APVPGSSSSTLPSSCEQSR----------------------QAETPSDPQTETLSAMDVSEFLSLQSLDTPSNLIPIEALLQGEEEMGLTSSFSK----      EAX07312.1        753   80956.9
MM                                                                                                             NP_032662.3       675   72602.9
AM    --------SPLRERSGQKE----------------------GLENELMPGVRIWDCWGISALRAVIGLWSSLEI----------                      KYO35612.1        559   61761.0
CM    LEDSSQVTQPLPQTDPQTG----------------------LNSIDVTEFLSLESSGTPSNLAGIEANDIQALLEKVKEEINMNSRFPK----------      XP_007893150.1    752   82565.0
PR    TSSSSSCCLPKSSSKVGEPRPEAPTSSSSSSSSSTAAQSFSPIASNTATHPPTSDGLANMDVSDFLSMQSPETAANIEALLLVADDFSMTTDSGP-------    XP_008429346.1    804   86424.1
XM    TSSSSSCCLPKSSSKVGEPRPEAPTSSSSSSS-TAAQSFSPIVSNTATHPPTSDGLANMDVSDFLSMQSPETAANIEALLLVADDFNMTTDSGP-------     XP_005796106.1    816   87811.9
OAN1  TSSSSSCCPKSSSKVGEVTLEAPSSSSSSSS-STAQTFSTIVSSTATNPPSSDGLASMDVSDFLSLMAVDDFNMATDGNP----------              AAP93663.1        811   86819.7
TR    ANSPSCCFPASSAKVGEMRPGAHSSSS-------AQTFSTVVGSTATIFFPSSDGLANVSVSDFLSLQNPEAAANIEALLLVAEDINMATNCNAYKS----      NP_001027866.1    780   84604.8
DR    --PPSSCCPSKPHTQPQQDLPTAP--------------------LTPNPPTDALGSMDVSDLLSP---ETTANIEALLMVADGFSMADGGGSGAP----      AJF36548.1        722   77266.6
LA    --------------------------ARSGDRPAYDAGSRSKSGHRSTYDARSR----------------------                              XP_013403275.1    584   63861.9
```

Phylogenetic tree:

- OAN1 — 66
- TR
- 54
- PR — 97
- XM — 73
- DR — 55
- CM — 99
- AM
- HS — 90
- MM — 74
- 78
- LA — 34
- CG1
- CG2
- CG4 — 100
- CG3
- HH
- BG
- AC — 100
- 43
- OB

Groupings: Actinopteri, Chondrichthyes, Sauropsida, Mammalia — Chordata; Lingulata — Brachiopoda; Bivalvia, Gastropoda, Cephalopoda — Mollusca

0.050

Species	Accession
Haliotis discus hannai **(HH)**	AMS38481.1
Biomphalaria glabrata **(BG)**	XP_013093604.1
Aplysia californica **(AC)**	XP_005098642.1
Crassostrea gigas isoform-1 **(CG1)**	EKC32469.1
Crassostrea gigas isoform-2 **(CG2)**	XP_011433432.1
Crassostrea gigas isoform-3 **(CG3)**	Meng et al., 2015
Crassostrea gigas isoform-4 **(CG4)**	XP_011433433.1
Octopus bimaculoides **(OB)**	XP_014784404.1
Oreochromis aurea x *O. nilotica* **(OAN-1)**	AAP93663.1
Danio rerio **(DR)**	AJF36548.1
Poecilia reticulata **(PR)**	XP_008429346.1
Xiphophorus maculatus **(XM)**	XP_005796106.1
Takifugu rubripes **(TR)**	NP_001027866.1
Callorhinchus milii **(CM)**	XP_007893150.1
Homo sapiens **(HS)**	EAX07312.1
Mus musculus **(MM)**	NP_032662.3
Alligator mississippiensis **(AM)**	KYO35612.1
Lingula anatina **(LA)**	XP_013403275.1

-1691 tctaagatggggcacaatctactttgaaataatcatgataggactaatgtaa<u>attagcaaa</u>tttaaataagtata
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　CEBP

-1616 tcaggttctgttagttattccataatccagaatacccctgcaagttgtttagctggaggacaaaatgttcataatt

-1541 acttgttggctctccaattttttaattcatatacaattgtatatgctgttcactgctatgtacatggtgtacagaa

-1466 cataatgctcattatctgaaacttagacatgttagaaacttacaaagacaccaccaagtttgtccagcatccatg

-1391 actttggagccgtaactcttttaagcac<u>tttcct</u>ccgtcctcgataaccctgaaaacatgacaaattggtaagaa
　　　　　　　　　　　　　　　　　NF-AT1

-1316 atatttcactcatcaacatgatcaaag<u>caggtg</u>gagccttcgtattttaatcaaatcgtcaaa<u>acgtcaccaact</u>
　　　　　　　　　　　　　　　　　***E*-box**　　　　　　　　　　　　　　　　　　　　　　　**CREBP**

-1241 tagct<u>tgtctgt</u>aacatagcgctatcccatgaagatc<u>caggtca</u>taaatggcgtttagctacccatgcttgtctt
　　　　　Smad4　　　　　　　　　　　　　　　　**ER-h**

-1166 atgtgacaacttataggatcgggtggtaaggctcatggactaggctgatgcatgtcatcatattccaattgtata

-1091 aattgatgcctgtgatcctaatcaca<u>ggtctgtctg</u>gttcatacttgattactt<u>acagacc</u>ttcattggtgttaa
　　　　　　　　　　　　　　　　　Smad4　　　　　　　　　　　　　　　　**Smad4**

-1016 ccaaaacaaggct<u>gaatattc</u>taaaatccctgtaccaaccctggatataacatgttg<u>tgcacact</u>acagaaagta
　　　　　　　　　HSF (HSE)　　　　　　　　　　　　　　　　　　　　　**MTF-1 (MRE)**

 -941 <u>tggaaacc</u>ttttggggccattgttatttggaactgatatggcattg<u>cacg</u>ttatcacgacacagctaactttgtt
　　　　　NF-AT1　　　　　　　　　　　　　　　　　　　　**HIF-1α (HRE)**

 -866 ttagcagctacattaacattattaaatttgatatttatttttagactgacagtcattcatgg<u>ttgcttt</u>atgaaa
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　CEBP

 -791 agaaagaatcatgttagat<u>ttacaaac</u>atgcttaaaaatctgatagttaatgaat<u>tcgcgt</u>gagcctccaacata
　　　　　　　　　　　　HNF3　　　　　　　　　　　　　　　　　**AhR (XRE)**

 -716 tccttagttacatcaatgaaatattaagaaaccagtgtacgagagatccggaggcatacctgtcataagtact<u>ag</u>

 -641 <u>gtca</u>ctgttgatagcatattaaatacaggccccaatggcgaaagtatttgcgcgtatatgtggtgtcgtccgtcc
　　　　ER-h

 -566 accactctctcccctcctaattatgttacaaaaggtgttccacttattcgtggcaact<u>tgtttacac</u>actgctag
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　HNF3

 -491 aggatgtaaatttgaga<u>tataaag</u>acttcaaaagacgataactaacgtaactaaaatgcttgatatattagtggt
　　　　　　　　　　　　***TATA*-box**

 -416 atctatctgctaatctgtgacaacacatgtgaatacaatcgaaacgtatccaacatgggcccaaaaaaatcaacc

 -341 atatcacacatctggtttacattacactcacaatatggagtaatcataaaagaaga<u>cgtg</u>taaatgaacatta
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　HIF-1α (HRE)

 -266 aaaacaaaattatgtcctatttgtctatattttaaactggatgcatcag<u>attagcaag</u>acgggtatctgtatgtc
　　　　　　　　　　　　　　　　　　　　　　　　　　　　CEBP

 -191 ctcaccattttggatggctacc<u>ggaaaaa</u>ggtcgaatatgcgatcagcaattct<u>gtttcc</u>gcttccgctatggcc
　　　　　　　　　　　　　NF-AT1　　　　　　　　　　　　　　　　**NF-AT1**

 -116 tgcgttatcgtctctttgattggttatt<u>cgtctgg</u>atagtgaaccaataggatacattctaacacaggcaagagt
　　　　　　　　　　　　　　　　　Smad4

  -41 ttacgtcc<u>ggaaaattc</u>tgttcggcccattcggttttcgct**ATG**
　　　　　　HSF (HSE), **NF-AT1**