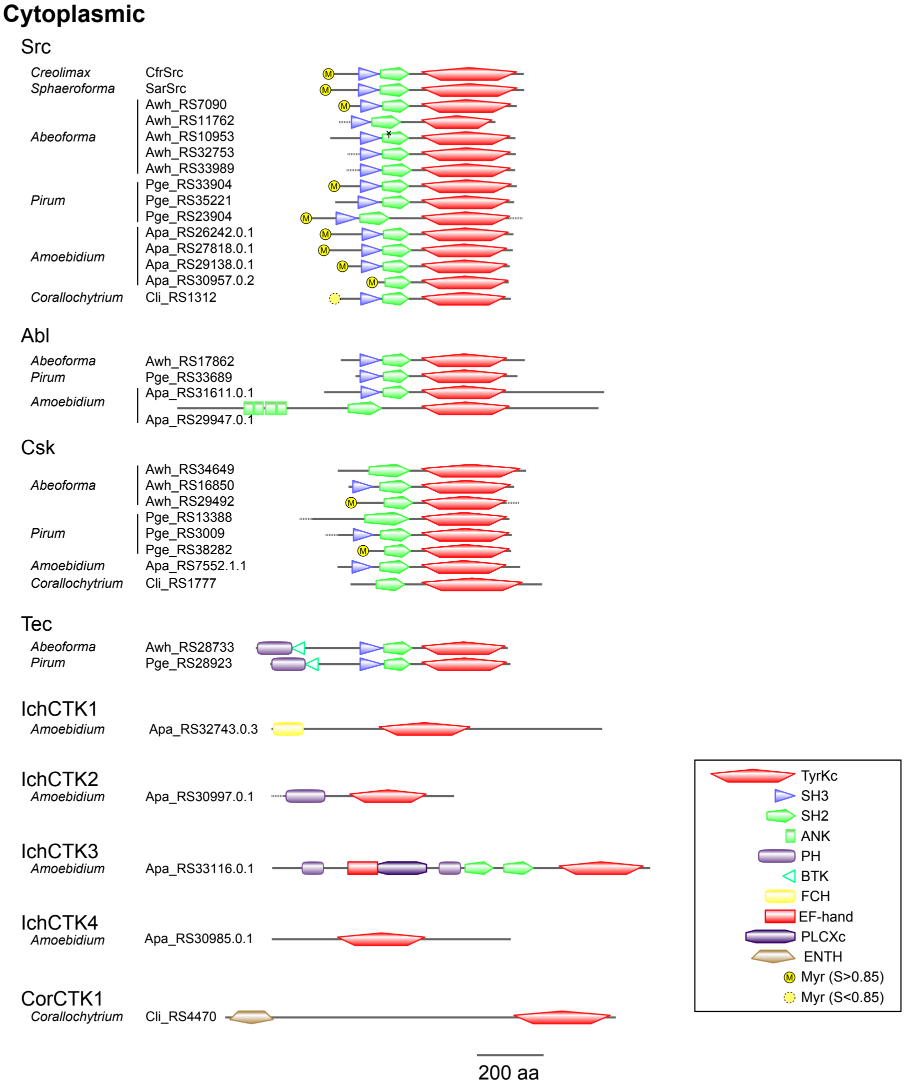


Figure S1. continued

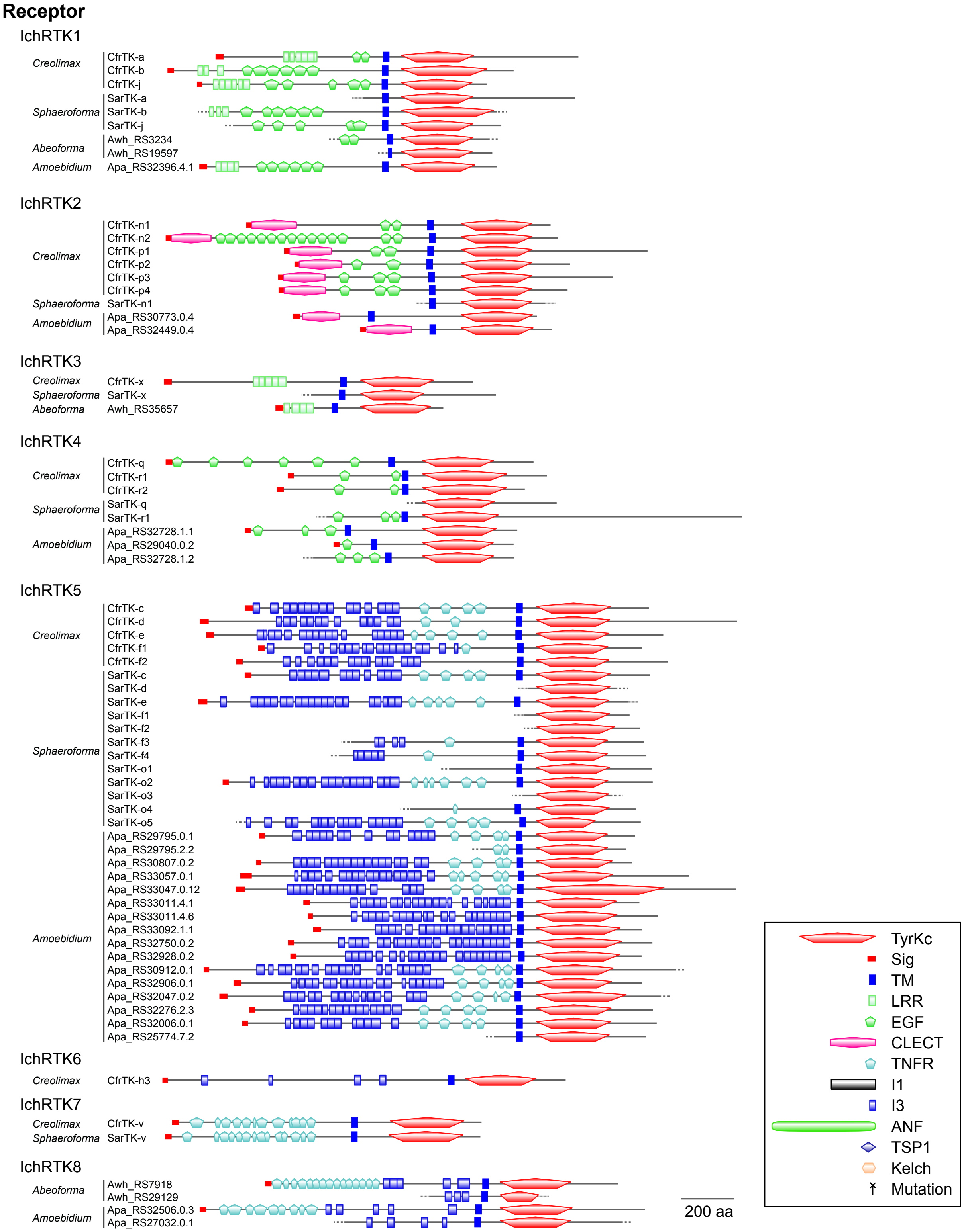


Figure S1. continued

**Receptor** (continued)

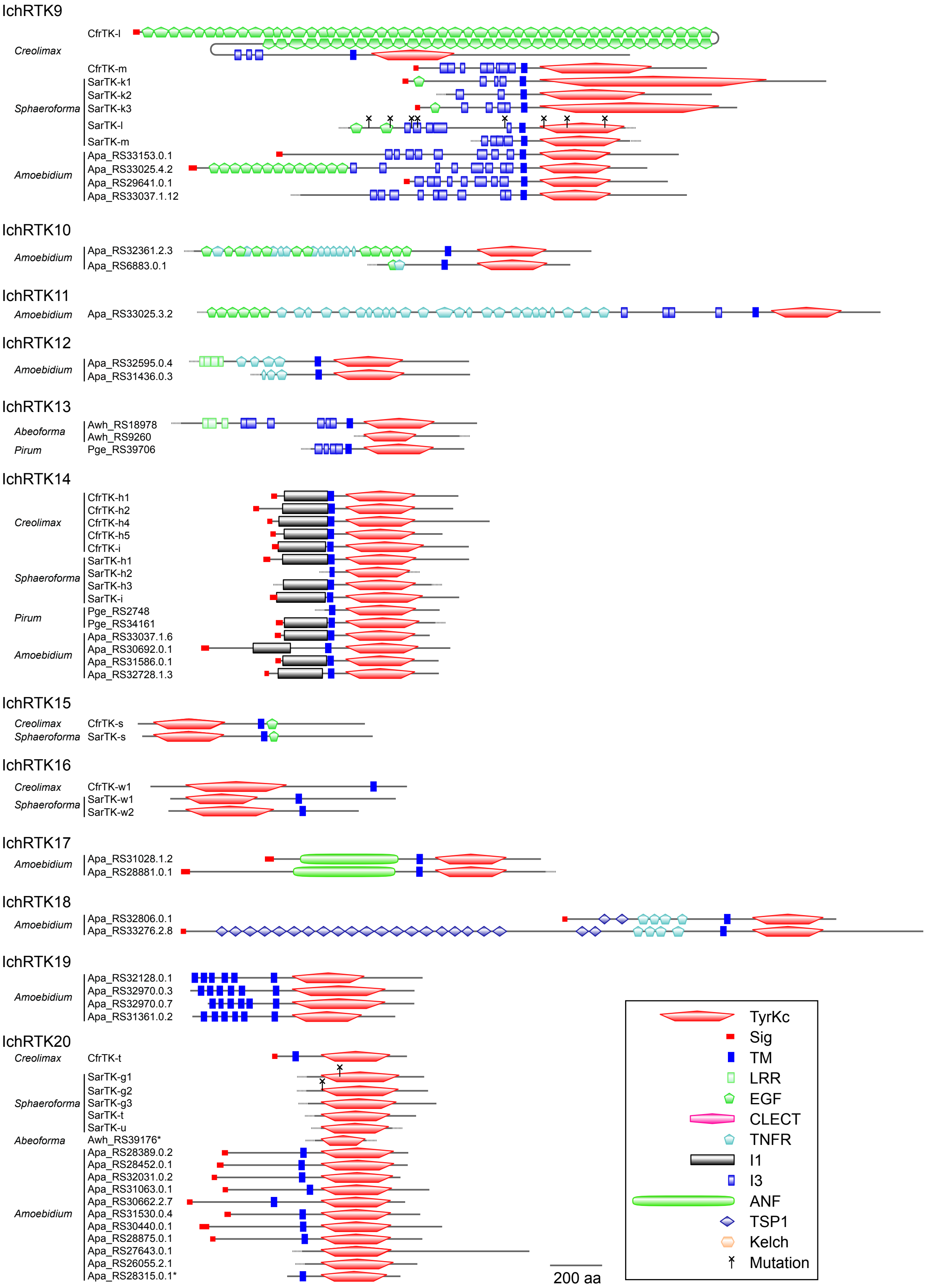


Figure S1. continued

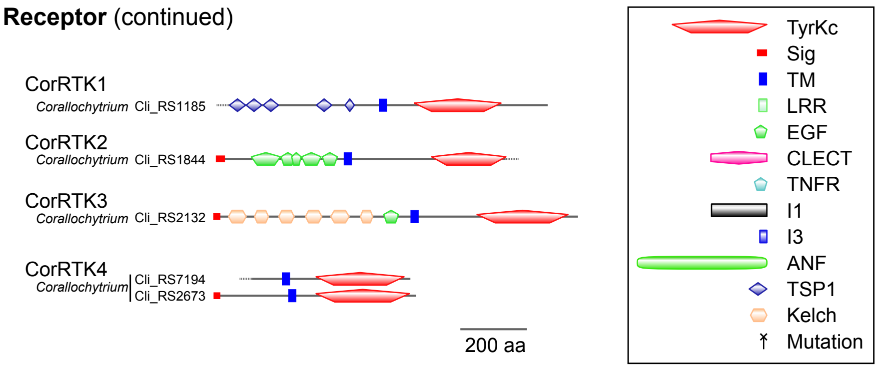
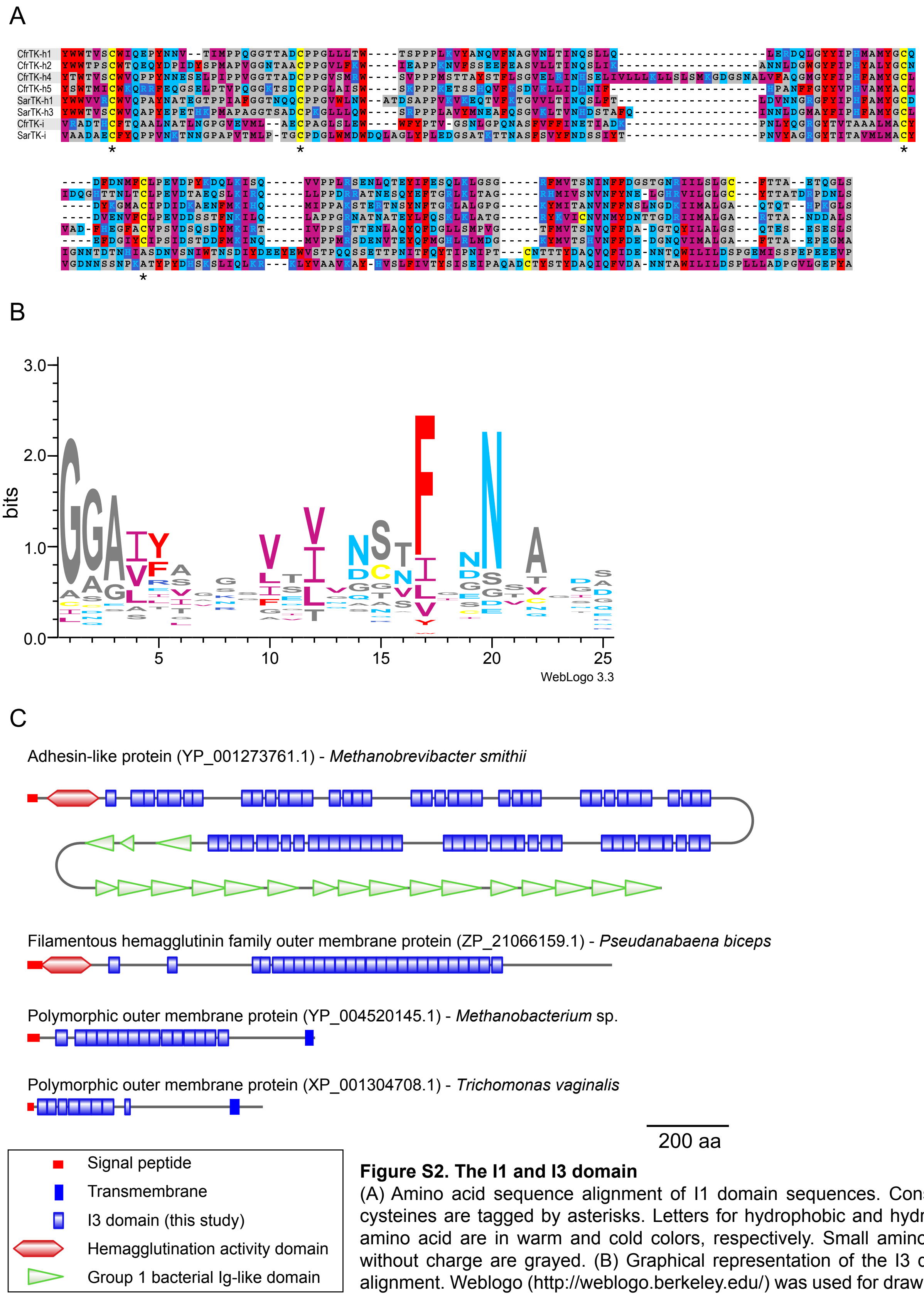

**Receptor** (continued)

**Figure S1. Ichthyosporean and corallochytrean TKs**
All TKs encoded by the genomes of *Creolimax* and *Sphaeroforma* and those identified in the RNAseq data of *Corallochytrium, Pirum, Abeoforma*, and *Amoebidium* are schematically represented. Point mutations introducing stop codons, frame shifts, and mis-splicings in the coding region are indicated by crosses with vertical bars. Asterisks indicate the two IchRTK20 family genes that were excluded from the phylogenetic analyses due to their highly divergent catalytic domain sequences. ANK, ankyrin repeats; BTK, Bruton's tyrosine kinase Cys-rich motif; CLECT, C-type lectin (CTL) or carbohydrate-recognition domain (CRD); EF-hand, EF-hand-like domain; EGF, epidermal growth factor-like domain; ENTH, epsin N-terminal homology domain; FCH, Fes/CIP4 homology domain; I1, I1 domain; I3, I3 domain; Kelch, Kelch motif; LRR, leucine-rich repeat; Myr, predicted myristoylation site (dotted if S < 0.85); PH, Pleckstrin homology domain; PLCXc, phospholipase C catalytic domain X; SH2, Src homology 2 domain; SH3, Src homology 3 domain; Sig, signal peptide; TM, transmembrane segment; TNFR, tumor necrosis factor receptor / nerve growth factor receptor repeat; TSP1, thrombospondin type 1 repeats; TyrKc, tyrosine kinase catalytic domain. Incomplete sequences are shown by dotted lines.

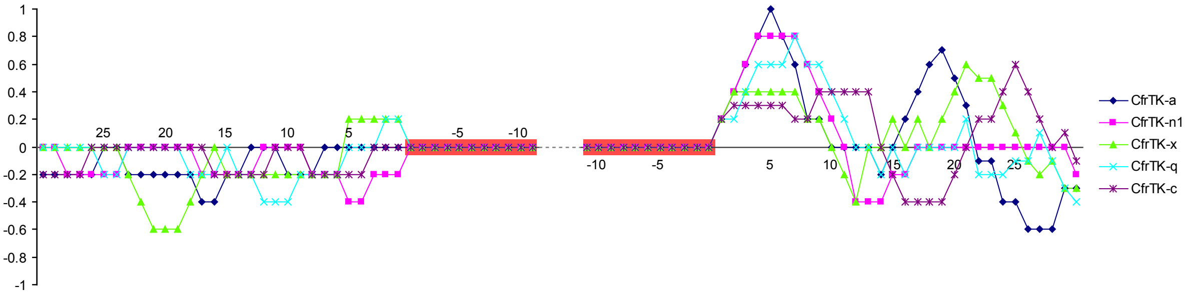


```
CfrTK-h1 YWWTVSCWIQEPYNNV--TIMPPQGGTTADCPPGLLLTW----TSPPPLKVYANQVFNAGVNLTINQSLLQ-------------------LERDQLGYYIPHMAMYGCQ
CfrTK-h2 YWWTPSCWTQEQYDPIDYSPMAPVGGNTAACPPGVLFKW----IEAPPKNVFSSEEFEASVLLTINQSLIK------------------ANNLDGWGYFIPHYALYGCN
CfrTK-h4 YTWTVSCWVQPPYNNESELPIPPVGGTTADCPPGVSMRW----SVPPPMSTTAYSTFLSGVELRINHSELIVLLLKLLSLSMKGDGSNALVFAQGMGYFIPHFAMYGCL
CfrTK-h5 YSWTMICWKQRRFEQGSELPTVPQGGKTSDCPPGLAISW----SKPPPKETSSHQVFKSDVKLLIDHNIF--------------------HPANFFGYYVPHVAMYACL
SarTK-h1 YWWVVRCWVQPAYNATEGTPPIAFGGTNQQCPPGVWLNW--ATDSAPPVKVKEQTVFKTGVVLTINQSLFT------------------LDVNNGRGFYIPHFAMYGCD
SarTK-h3 YWWTVSCWVQAPYEPETHKPMAPAGGTSADCPKGLLLQW----SRPPPLAVYMNEAFQSGVKLTVNHDSTAFQ----------------INNLDGMAYFIPHFAMYGCL
CfrTK-i  VRADTHCFTQAALNATLNGPGVEVML--AECPAGLSLEW----WFYPTV-GSNLGPQNASFVFFINETIADK-----------------PNLYQGRGYTVTAAALMACY
SarTK-i  VAADAECFYQPPVNKTNNGPAPVTMLP-TGCPDGLWMDWDQLAGLYPLEDGSATKTTNASFSVYFNDSSIYT-----------------PNVYAGRGYTITAVMLMACY
               *                       *                                                                            *

         ----DFDNMFCLPEVDPYKDQLKISQ-----VVPPLRSENLQTEYIFESQLKLGSG-----RFMVTSNINFFDGSTGNRIILSLGC---FTTA--ETQGLS
         IDQGHTTNLTCLPEVDTAEQSLKIRQ-----LLPPDRRATNESQYEFTGRLKLTAG-----RHMIVSNVNFYNE-LGHRVILGLGC---YTTATDRPDNLS
         ----DYKGMACIPDIDKAENFMKIHQ-----MIPPAKSTERTNSYNFTGKLALGPG-----KYMITANVNFFNSLNGDKIIMALGA---QTQT--HPKGLS
         ----DVENVFCLPEVDDSSTFNKILQ-----LAPPGRNATNATEYLFQSKLKLATG-----RYKVICNVNMYDNTTGDRIIMALGA---RTTA--NDDALS
         VAD-FHEGFACVPSVDSQSDYMKIRT-----IVPPSRTTENLAQYQFDGLLSMPVG-----TFMVTSNVQFFDA-DGTQYILALGS---QTES--ESESLS
         ----EFDGIYCIPSIDSTDDFMKINQ-----MVPPMRSDENVTEYQFMGHLRLMDG-----KYMVTSHVNFFDE-DGNQVIMALGA---FTTA--EPEGMA
         IGNNTDTNHIASDNVSNIWTNSDIYDEEYEWVSTPQQSSETTPNITFQYTIPNIPT---CNTTTYDAQVQFIDE-NNTQWILILDSPGEMISSPEPEEEVP
         VGDNNSSNPKATYPYDHSKSLIQLKR---KLYVAAVKAY-HVSLFIVTYSISEIPAQADCTYSTYDAQIQFVDA-NNTAWILILDSPLLLADPGVLGEPYA
                   *
```

**Figure S2. The I1 and I3 domain**
(A) Amino acid sequence alignment of I1 domain sequences. Conserved cysteines are tagged by asterisks. Letters for hydrophobic and hydrophilic amino acid are in warm and cold colors, respectively. Small amino acids without charge are grayed. (B) Graphical representation of the I3 domain alignment. Weblogo (http://weblogo.berkeley.edu/) was used for drawing the amino acid frequency diagram. (C) Four examples of I3 domain-containing proteins from archaea and eubacteria (the first three proteins) and from the protist *Trichomonas vaginalis* (bottom).

A Canonical RTKs



B Inverted RTKs

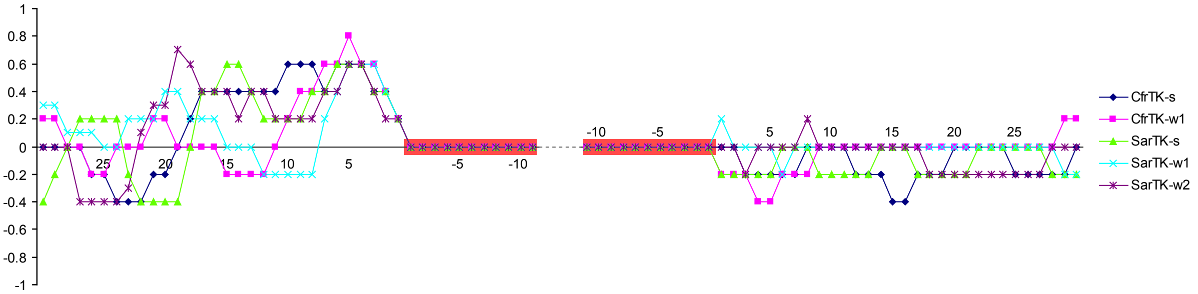


**Figure S3. Charges around the TM of RTKs**

Charges of RTK proteins were plotted around TM for the RTKs with the normal membrane topology (A) and those with the inverted topology (B). Red boxes schematically represent the positions of TMs. N-termini (left) and C-termini (right) of the TMs were separately analyzed, and shown in single illustrations. Y-axis, the averaged charge of the five amino acids in the scanning window; x-axis, the sequence position from the boundary between TM and non-TM.

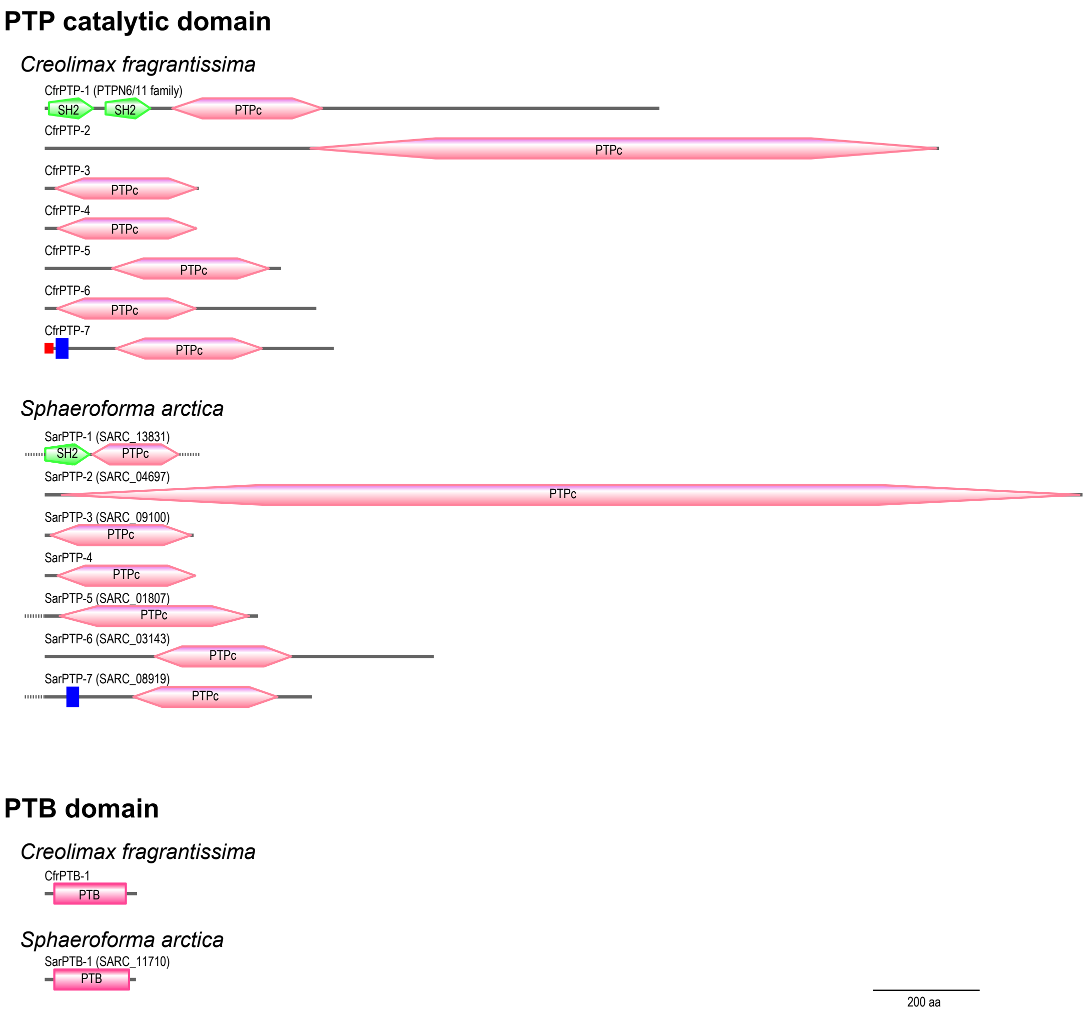


Figure S4. continued

## SH2 domain

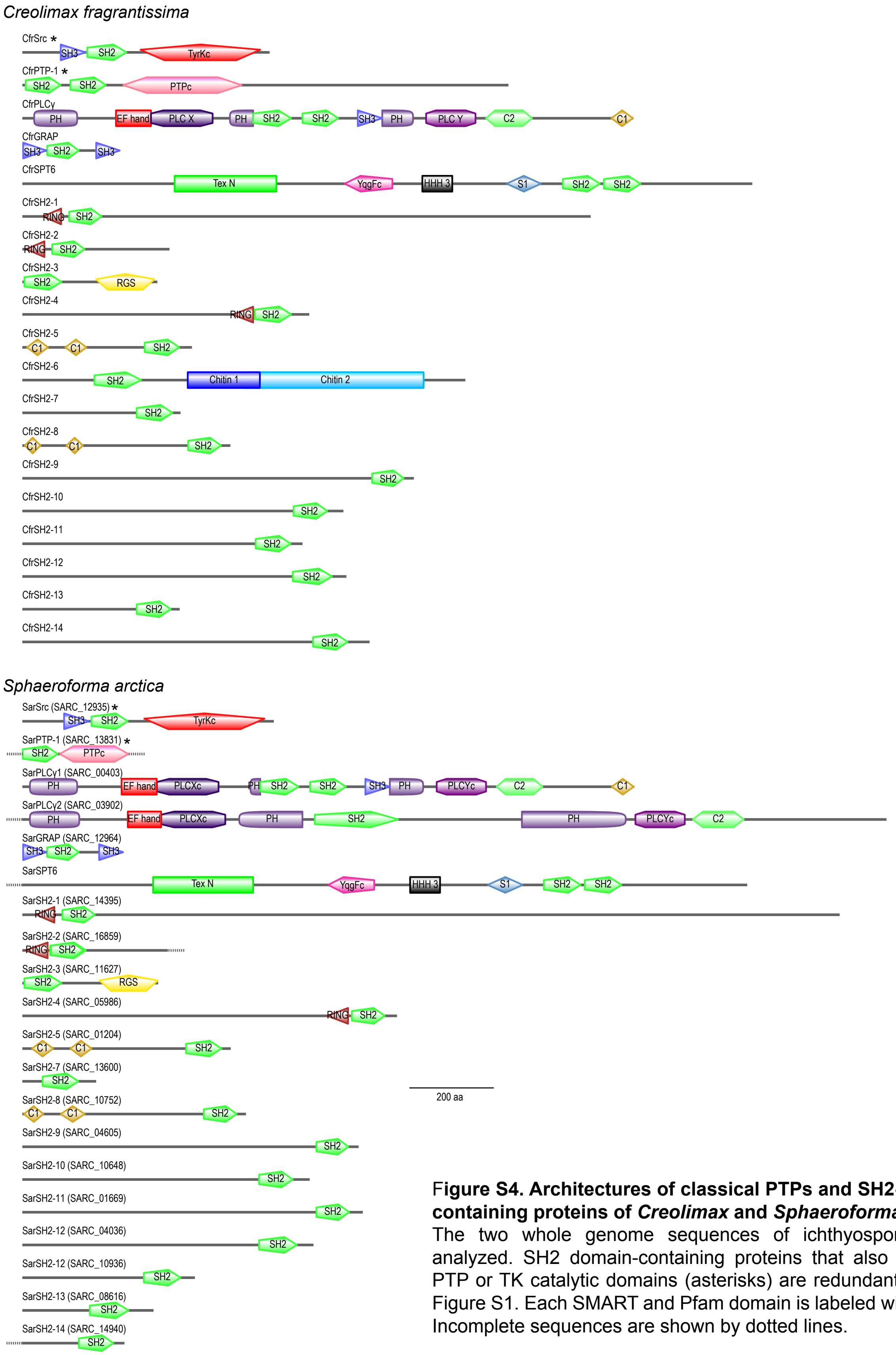


**Figure S4. Architectures of classical PTPs and SH2- and PTB-containing proteins of *Creolimax* and *Sphaeroforma***
The two whole genome sequences of ichthyosporeans were analyzed. SH2 domain-containing proteins that also have either PTP or TK catalytic domains (asterisks) are redundantly shown in Figure S1. Each SMART and Pfam domain is labeled with its name. Incomplete sequences are shown by dotted lines.

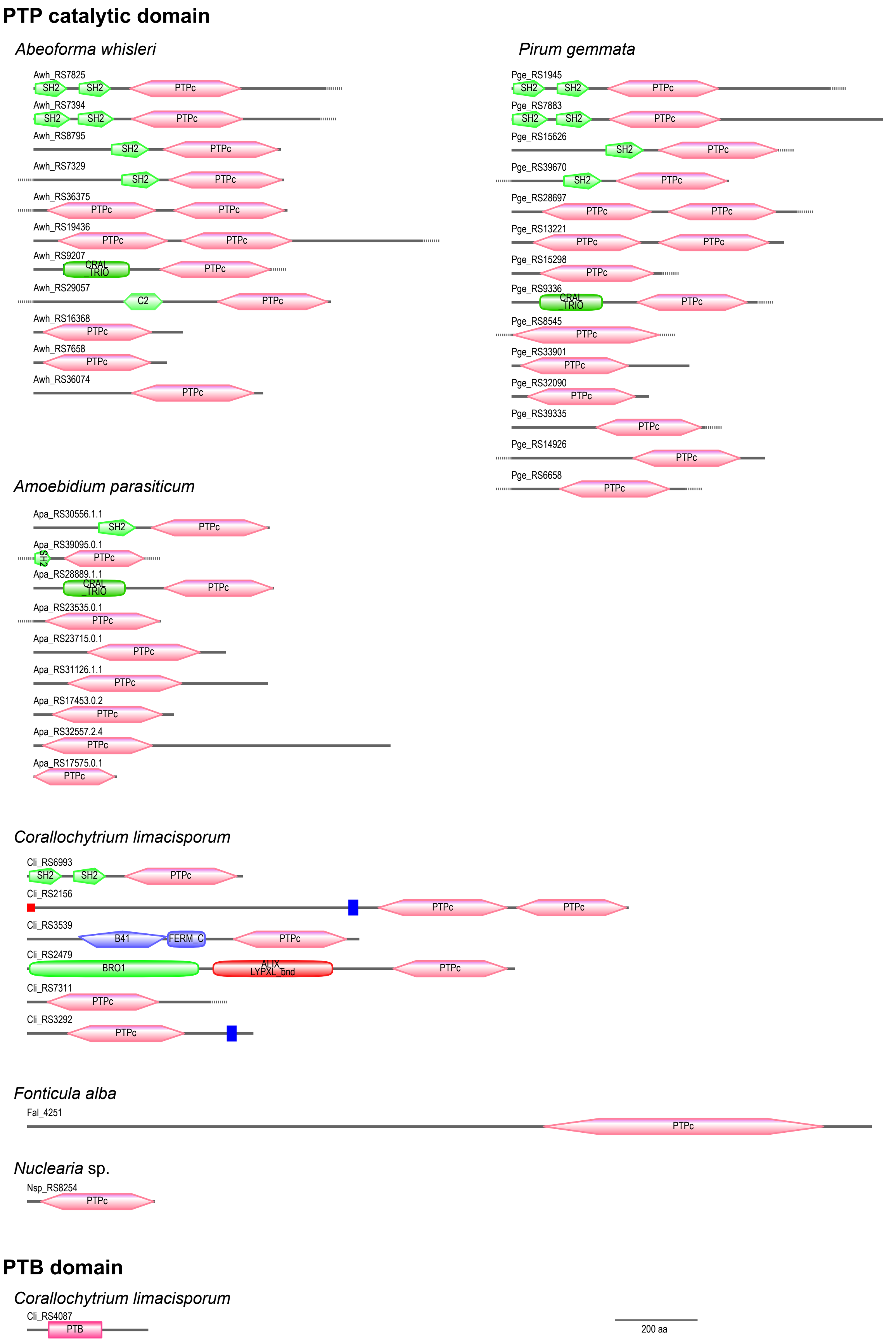


Figure S5. continued

## SH2 domain

*Abeoforma whisleri*

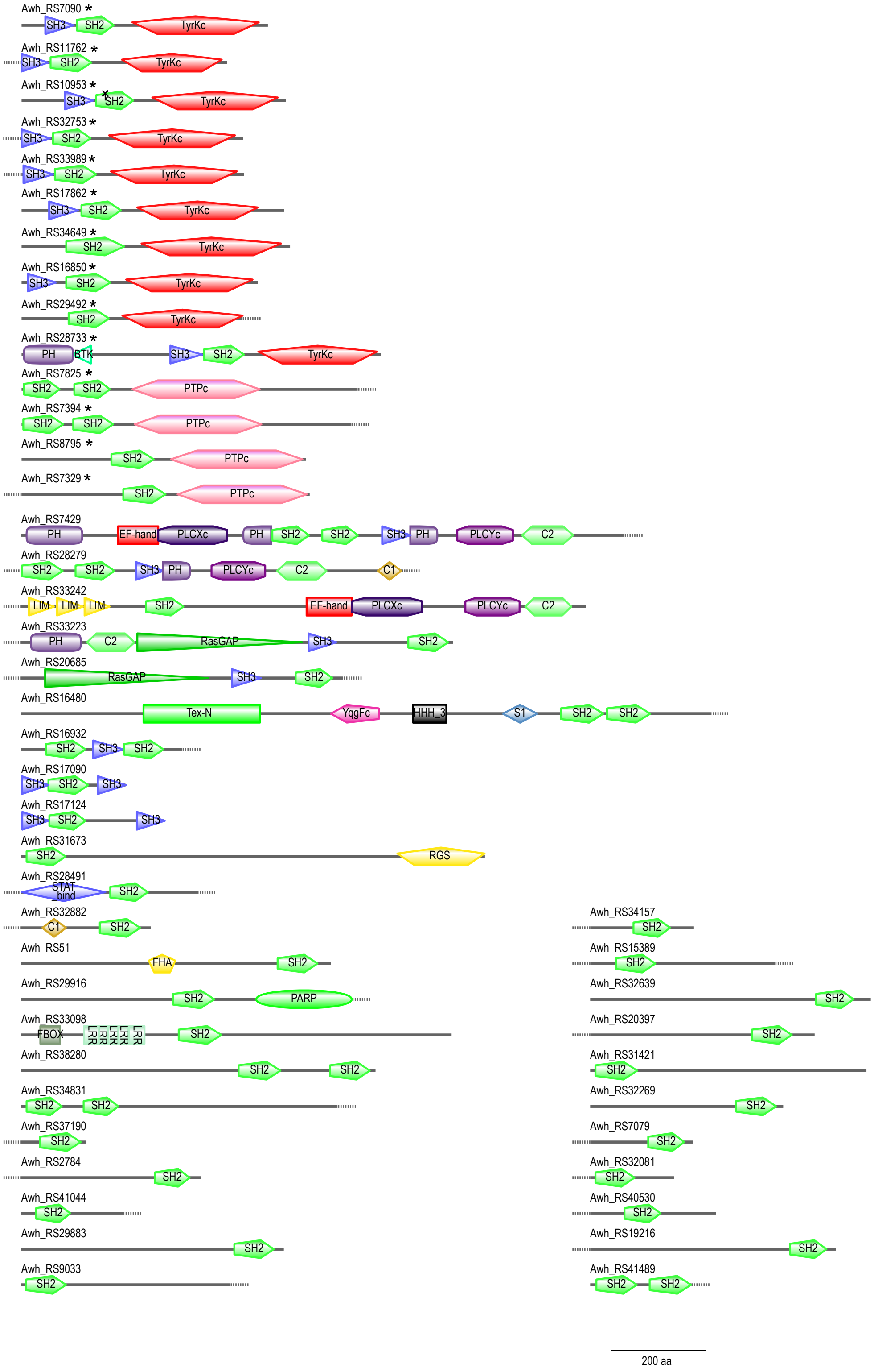


Figure S5. continued

## SH2 domain (continued)

### *Pirum gemmata*

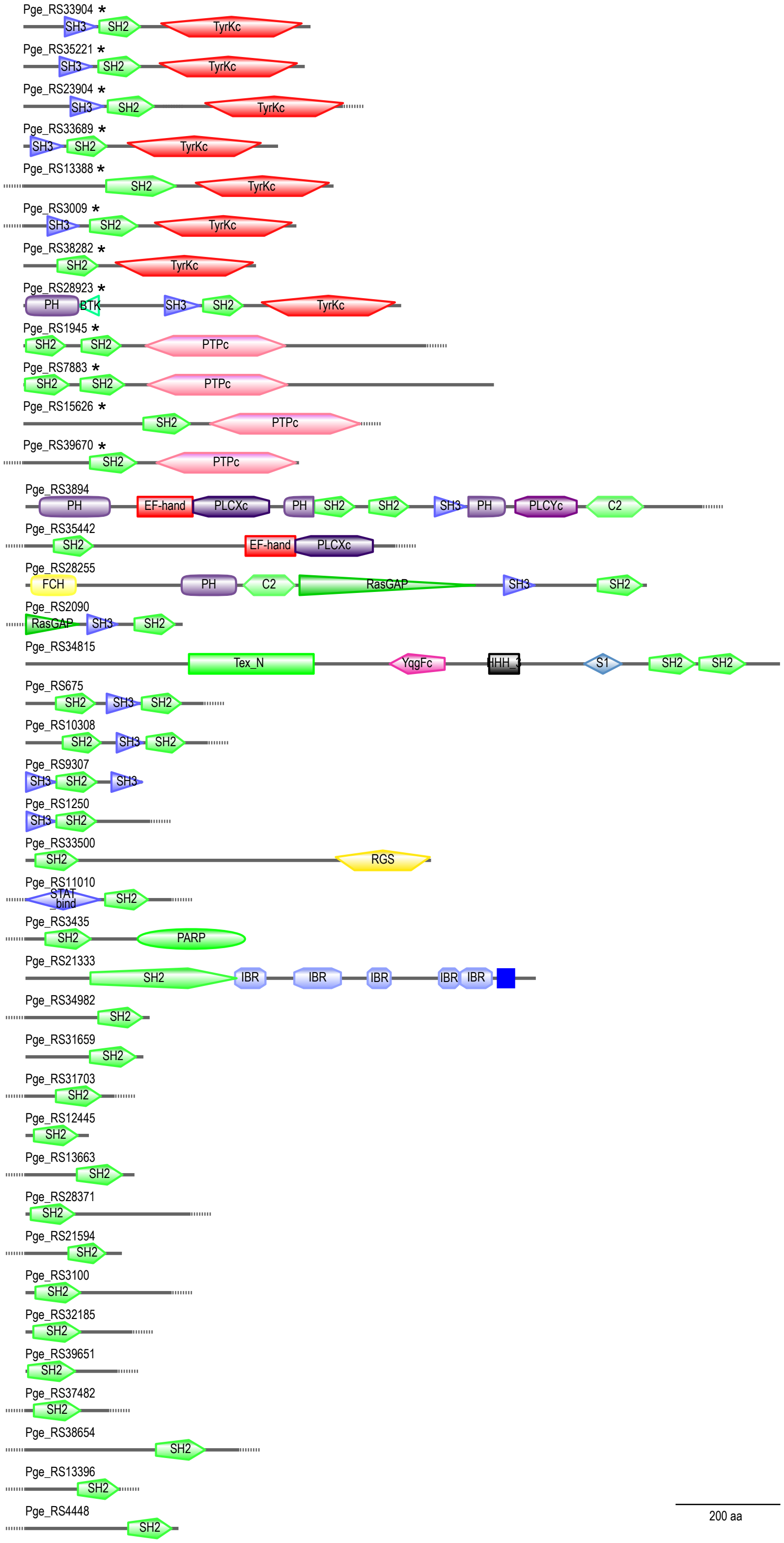


Figure S5. continued

## SH2 domain (continued)

*Amoebodium parasiticum*

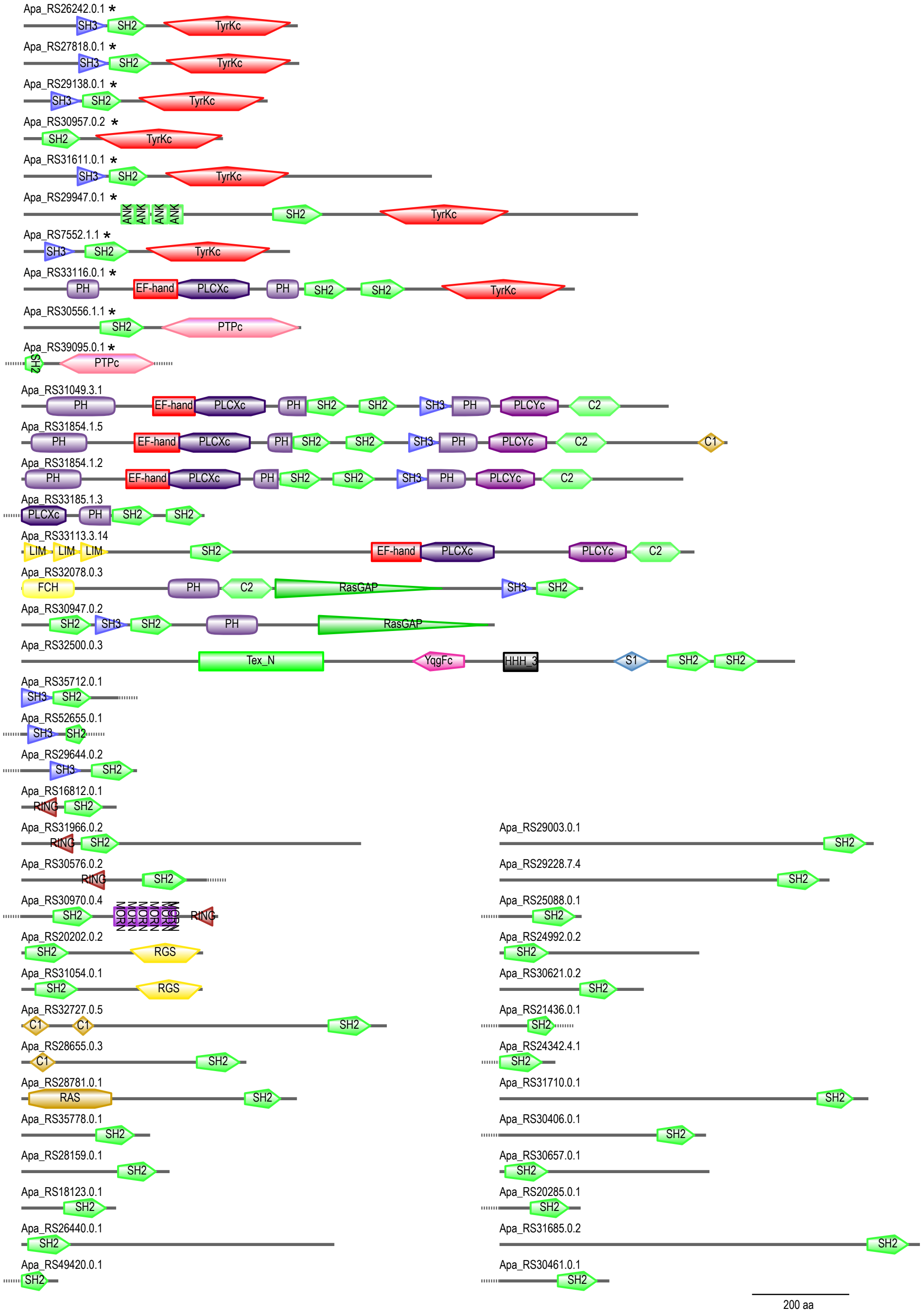


Figure S5. continued

## SH2 domain (continued)

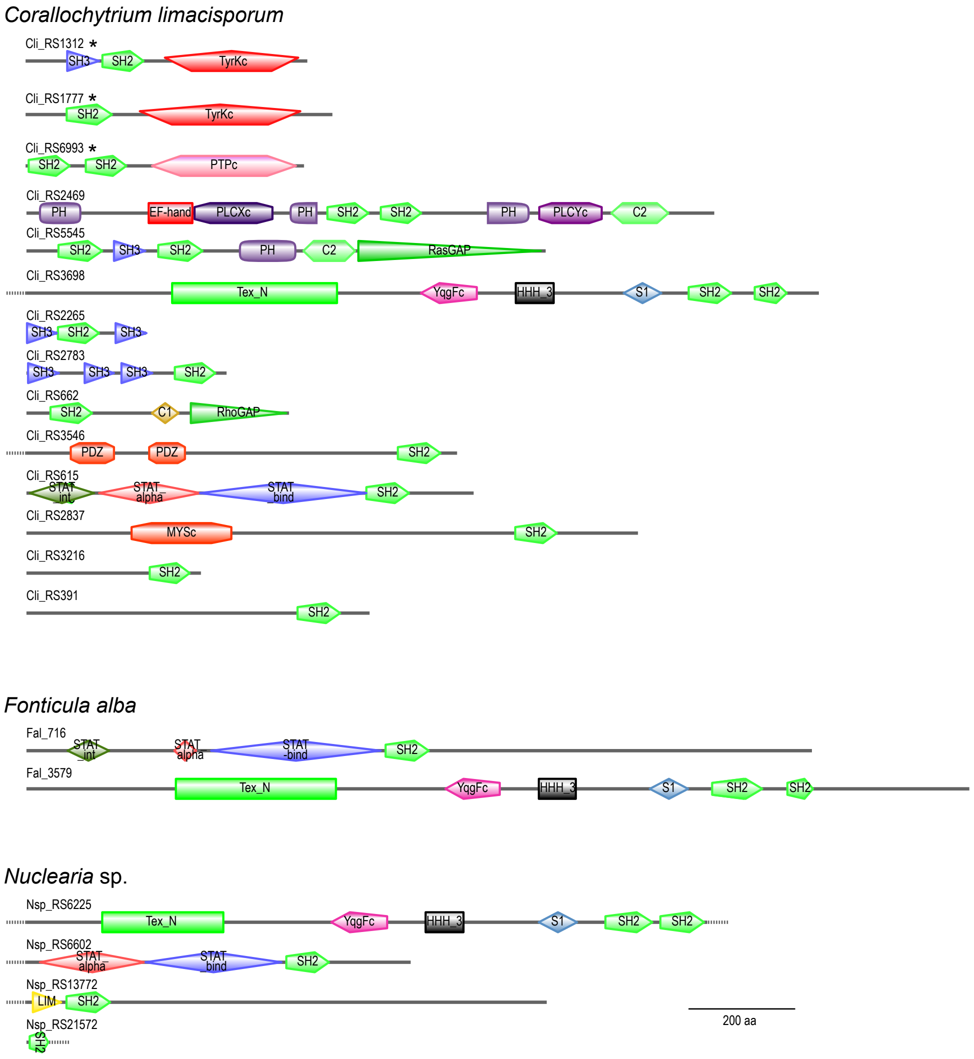


**Figure S5. Architectures of classical PTPs and SH2- and PTB-containing proteins of *Abeoforma, Pirum, Amoebidium, Corallochytrium, Fonticula*, and *Nuclearia***
Five RNAseq data of three ichthyosporeans, a corallochytrean, and a nucleariid and the whole genome data of *Fonticula* were analyzed. SH2 domain-containing proteins that also have PTP or TK catalytic domains (asterisks) are redundantly shown in Figure S1. Each SMART of Pfam domain is labeled with its name. Incomplete sequences are shown by dotted lines.

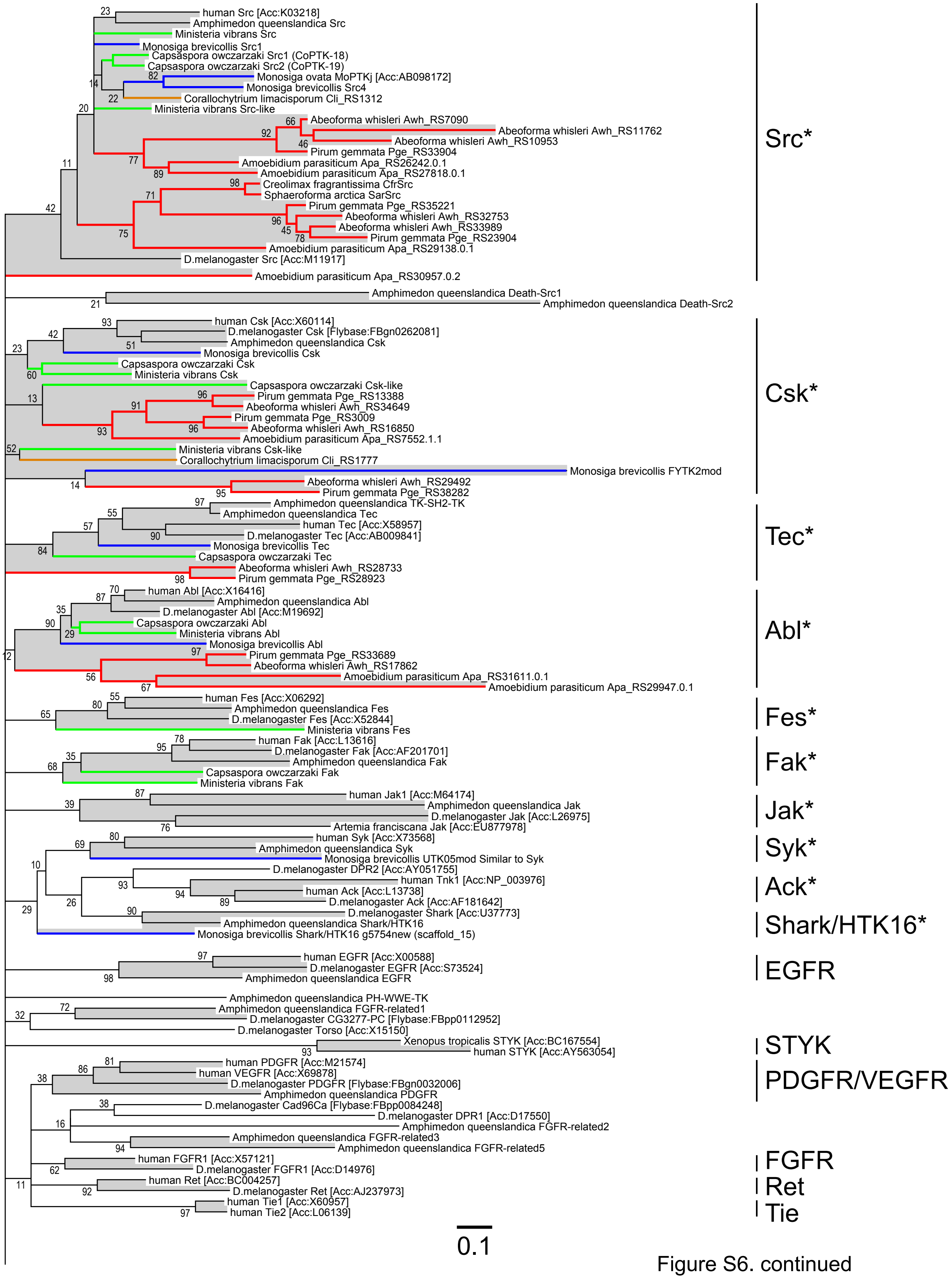


Figure S6. continued

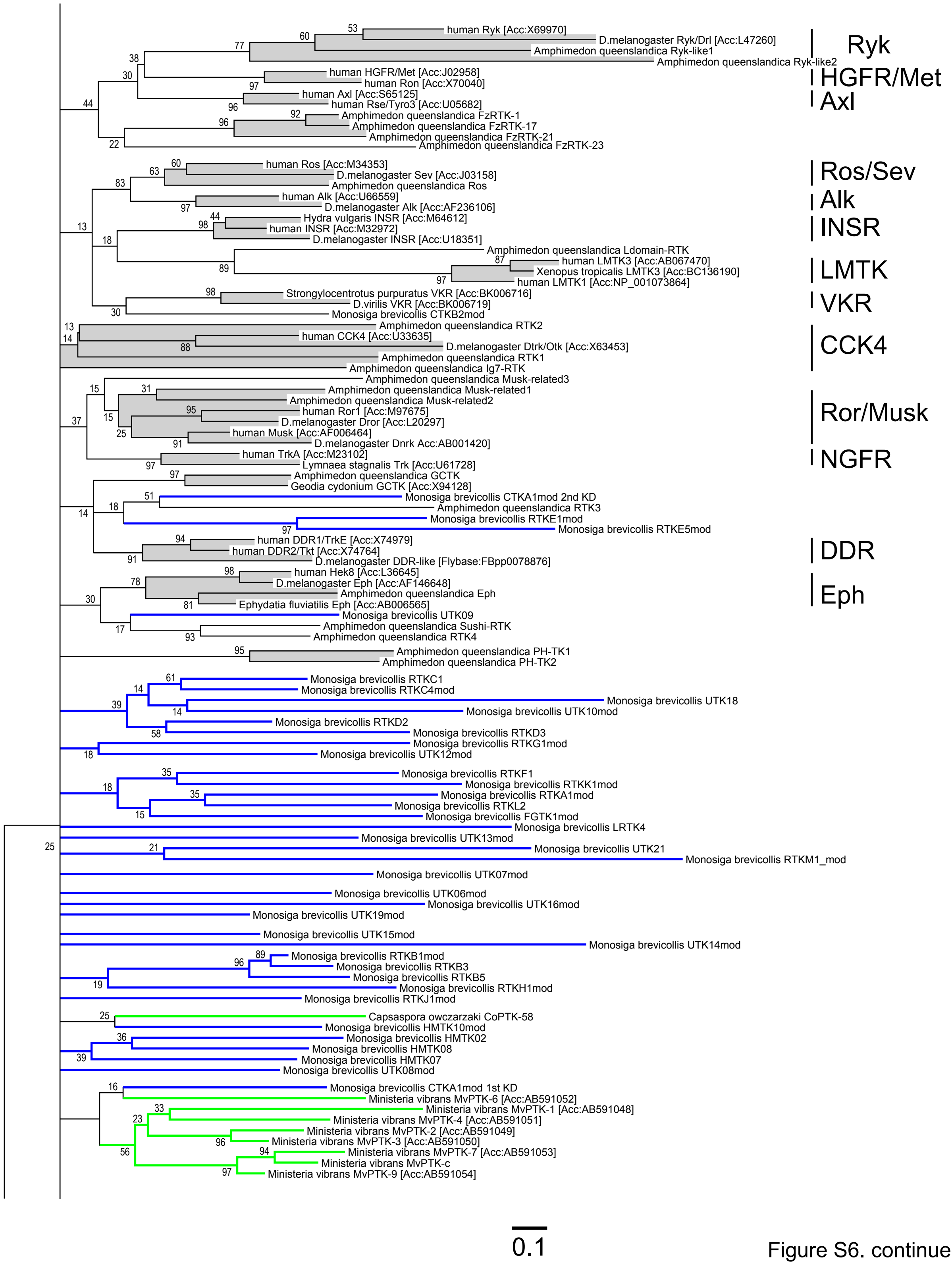


Figure S6. continued

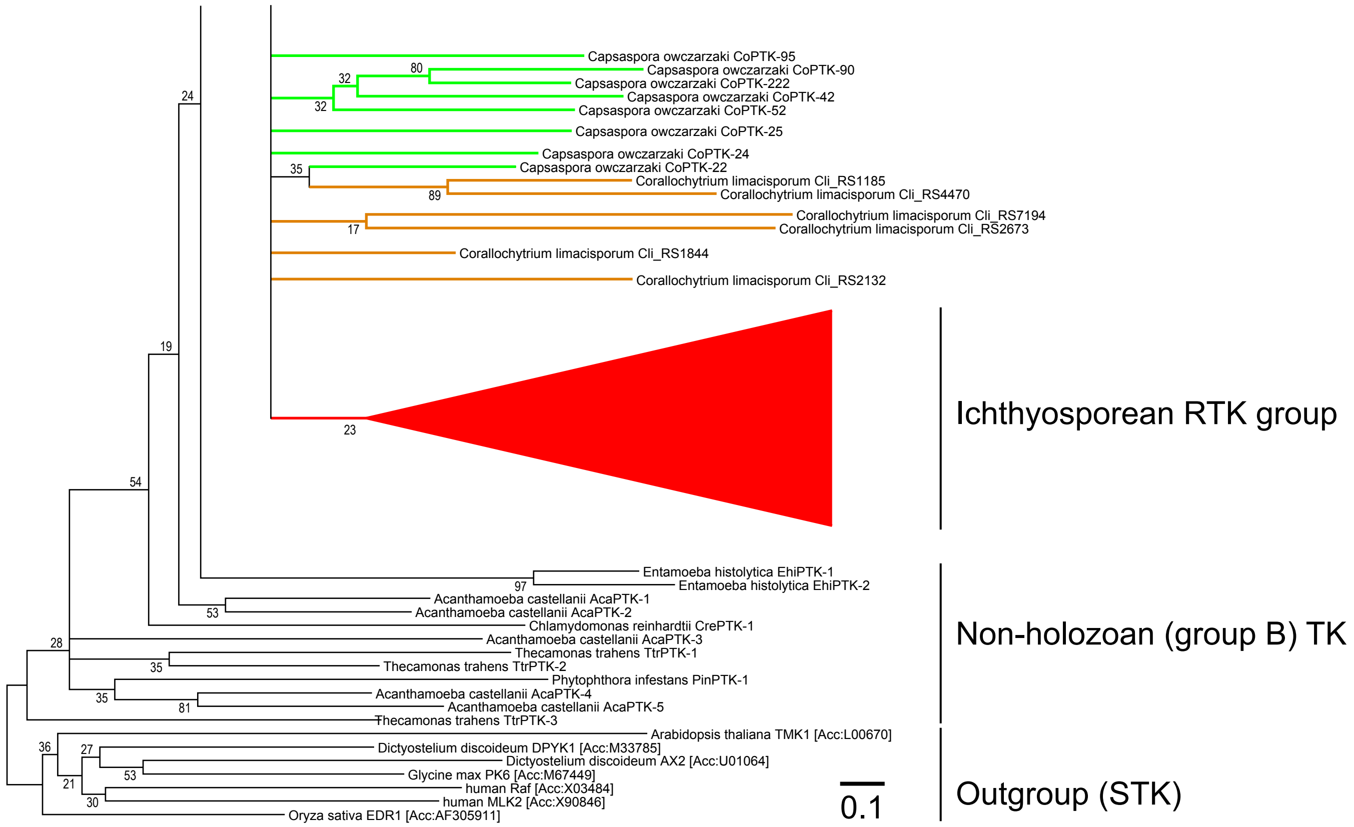


**Figure S6. Phylogenetic tree of the TK family**

Seven selected STKs are used as an ourgroup. Bootstrap values by 500 replicates are shown. Branches with low bootstrap values (< 10%) are collapsed and represented by multifurcated nodes. The names of 29 common metazoan TK families are shown on the right. CTK families are labeled by asterisks. Genes from ichthyosporeans, a corallochytrean, filastereans, and choanoflagellates are shown by red, orange, green, and blue lines, respectively.

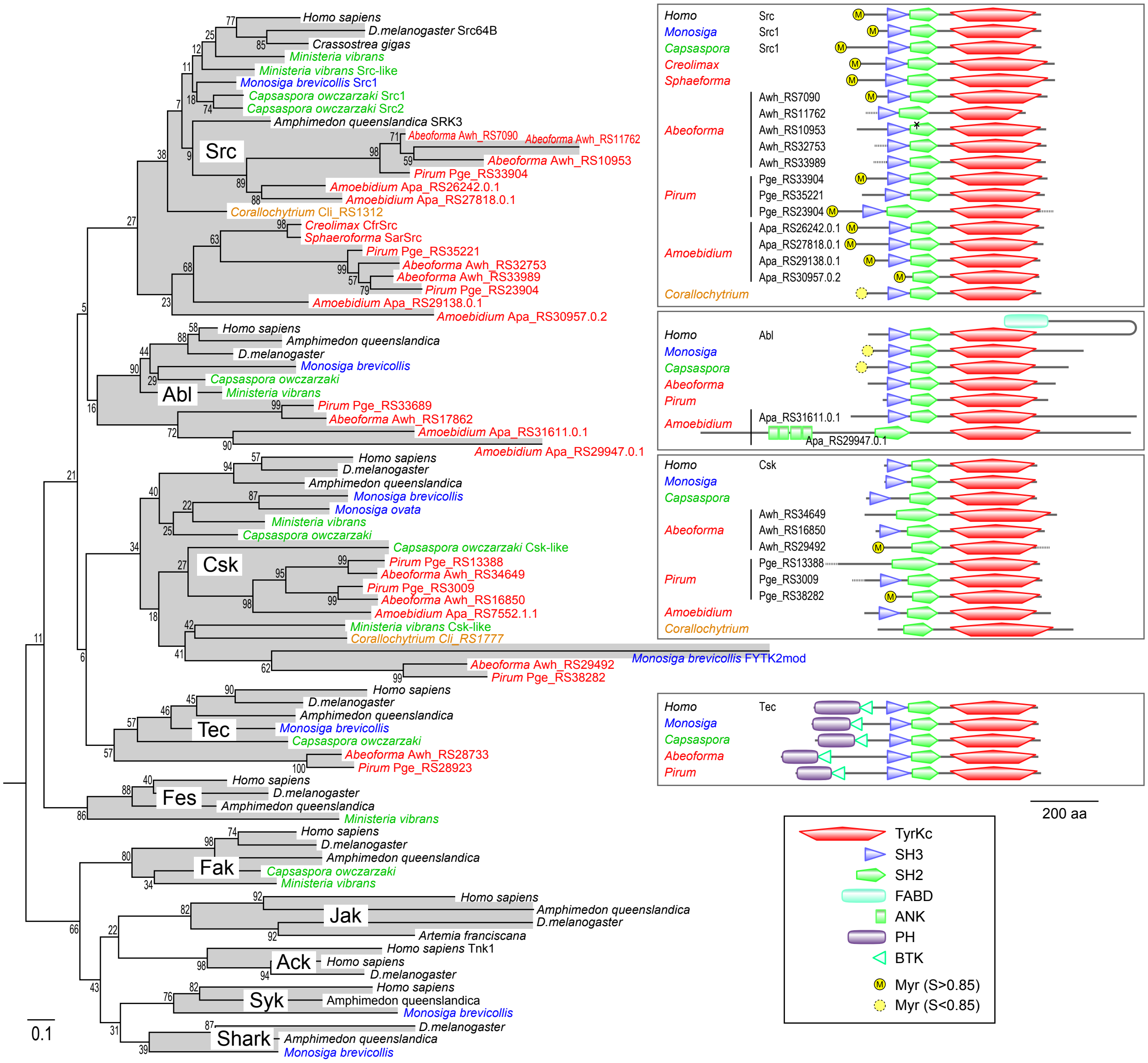


**Figure S7. Early diversification of CTK families in the Holozoa**
The ML phylogenetic tree comprised of 10 common holozoan CTK families (grey shades) was inferred. Group-B TKs were used as an outgroup (not shown). Genes of choanoflagellates, filastereans, ichthyosporeans, and *Corallochytrium* are in blue, green, red, and orange, respectively. Gene duplications within the Metazoa and the Choanoflagellida are not explicitly shown. Bootstrap values by 300 replicates are shown. The domain architectures of ichthyosporean and corallochytrean genes of the Src, Abl, Csk, and Tec families are schematically shown on the right, together with other family members from human, *Monosiga*, and *Capsaspora*. ANK, ankyrin repeats; BTK, Bruton's tyrosine kinase Cys-rich motif; FABD, F-actin binding domain; Myr (S>0.85), predicted myristoylation site with a high confidence (score S > 0.85); Myr (S<0.85), predicted myristoylation site with a low or medium confidence (score S < 0.85); PH, Pleckstrin homology domain; SH3, Src-homology 3 domain; SH2, Src-homology 2 domain; TyrKc, catalytic domain of tyrosine kinase.

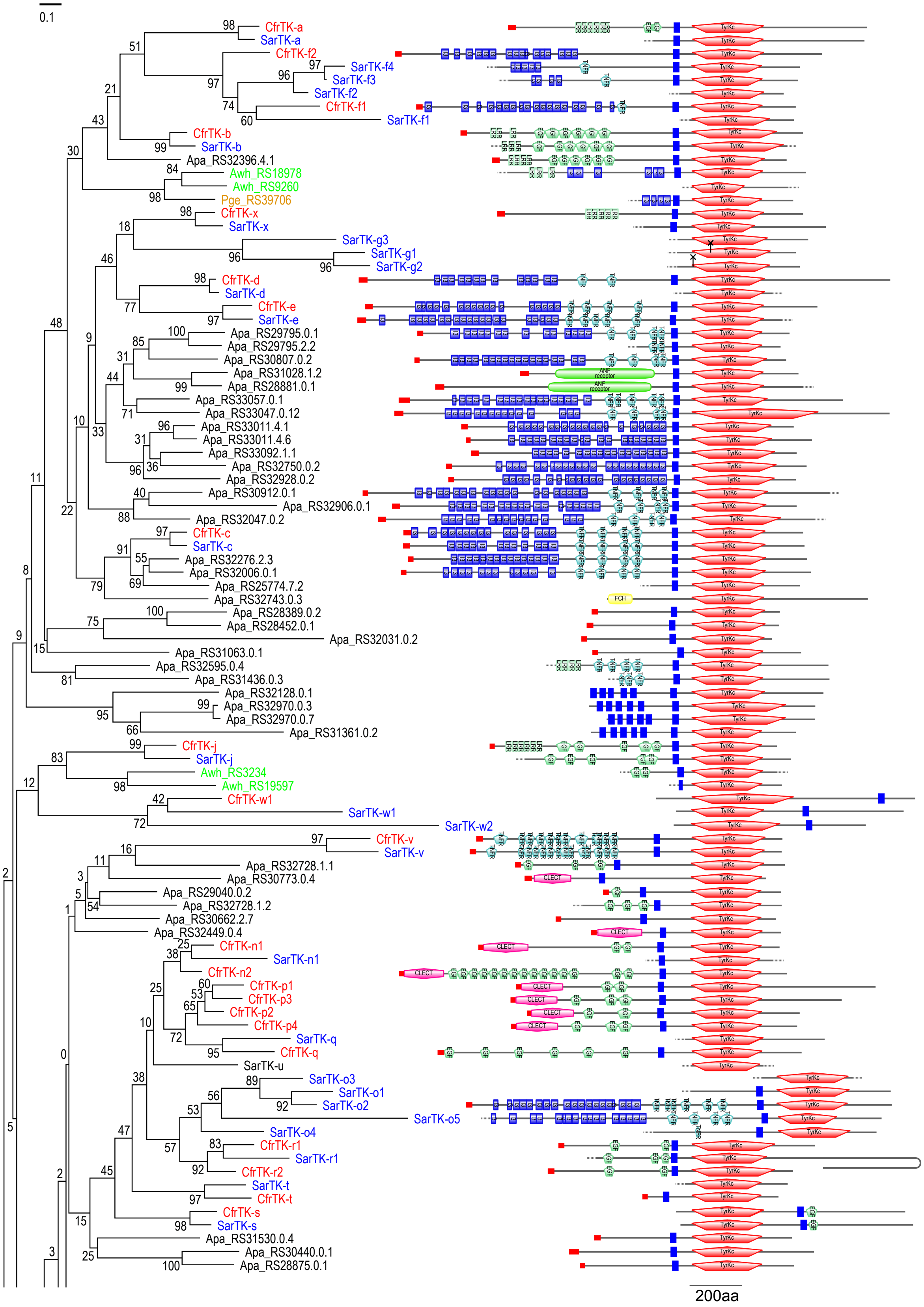


Figure S8. continued

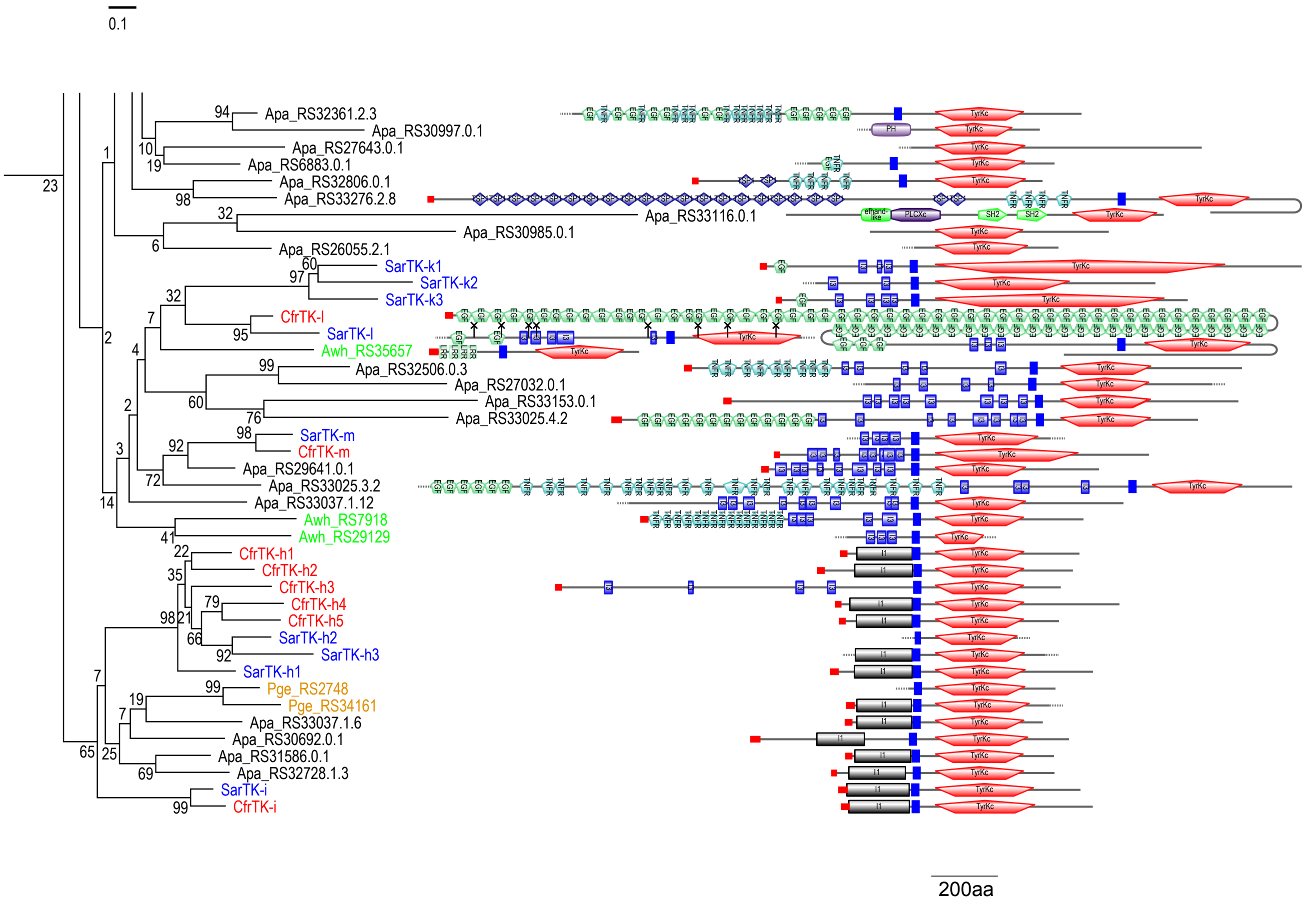


**Figure S8. The ichthyosporean RTK cluster**

The monophyletic cluster (red triangle in the Figure 4 tree) including all the ichthyosporean TKs, except those belonging to the Src, Csk, Tec, and Abl families, is shown in detail. Bootstrap values by 500 replicates are indicated. The domain architectures of proteins are schematically shown on the right. The SMART and Pfam domain names are labeled directly in the domain schemes. Note that all the ichthyosporean RTKs and only a few *Amoebidium* CTKs are included in this cluster. Names of genes from *Creolimax, Sphaeroforma, Abeoforma, Pirum*, and *Amoebidium* are in red, blue, black, green, and orange, respectively.

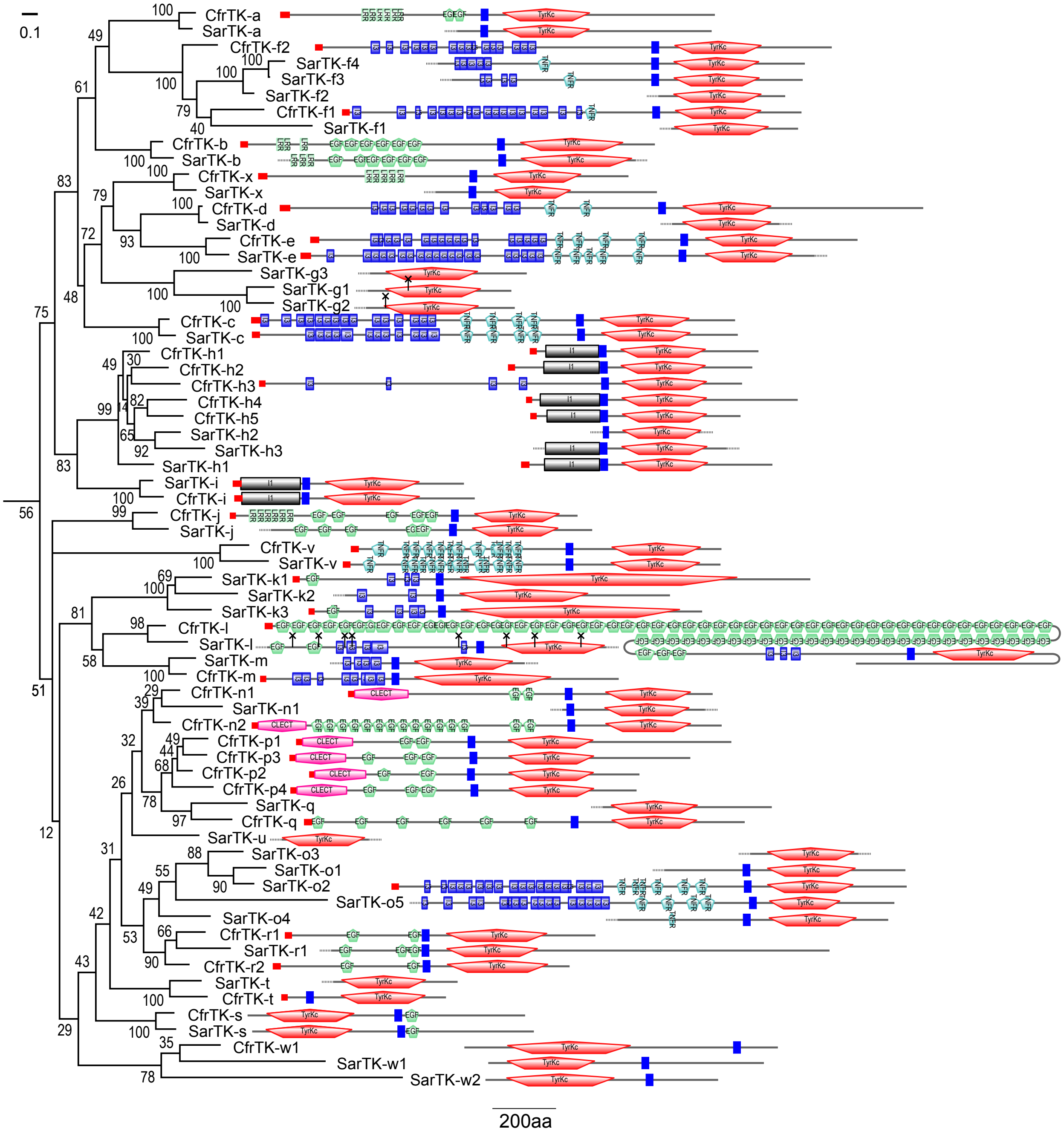


**Figure S9. Phylogeny of *Creolimax* and *Sphaeroforma* RTKs**
The ichthyosporean-specific TK subtree is shown in detail. This tree only contains the data of the two genome-sequenced ichthyosporeans, *Creolimax* and *Sphaeroforma*. Branches without support (bootstrap values < 10%) are represented as multifurcated nodes. The domain architectures of all TKs are schematically shown on the right. The SMART and Pfam domain names are directly shown in the domain schemes.