Supporting Information

Touzot et al. 10.1073/pnas.0914918107

SI Text

Coemergence of the TRF1/TRF2 Paralogues and the TBM Motif of Apollo. Apollo (SNM1B), as well as its paralogues SNM1A and Artemis (SNM1C), is found in all genomes of metazoans, from the placozoan *Trichoplax adhaerens*, which represents the primitive metazoan form, to humans. The distinction between the different paralogous sequences can be made based on the alignment of their β-CASP domains (between the conserved motifs A and B), which are much more divergent than the metallo–β-lactamase domains (Fig. S5). In contrast, only one gene is found in unicellular organisms, including protozoans, such as the ciliate *Paramecium tetraulia* and the amoeba *Entamoeba histolytica*, and fungi (PSO2/SNM1). This suggests that the appearance of paralogous SNM1 sequences is an event linked to the evolution toward multicellular organisms. By using sensitive hydrophobic cluster analysis (HCA), we identified TBMs in the C-terminal regions of the Apollo sequences in mammals as well as in chicken and fish (*Danio rerio*) (Figs. S6 and S7). Although a complete Apollo sequence could not be directly identified among the available amphibian sequences, the presence of TBMs was highlighted by searching the *Xenopus tropicalis* EST database (Fig. S6) using the C-terminal sequence of human Apollo (481 to end) as bait. Because the sequence similarity extends outside the strict TBM toward an Apollo-specific C-terminal helix (Fig. S6), this result clearly indicates that Apollo is also present in amphibians. The C-terminal sequences of Apollo in the predicted sequences from nonvertebrate species, from the cephalochordate *Branchiostoma floridae* (Amphioxus or lancelet) toward lower metazoans, including echinoderms, arthropods, and nematodes, were examined by HCA and ScanProsite (Expasy) methods, but no TBMs or TBM-like motifs were identified (Fig. S7). Even if the results should be cautiously interpreted because data are based on predicted sequences that can only rarely be validated by searching EST databases, the systematic absence of the TBMs in nonvertebrate species strongly suggests that the TBM is vertebrate-specific.

Both TRF1 and TRF2 exist in vertebrates, from humans to fish (Fig. S8). TRF1 and TRF2 possess a C-terminal Myb domain (also known as the SANT domain) allowing DNA binding and an upstream TRFH domain for the dimerization and recruitment of various proteins to telomeres. Although TRF1 and TRF2 share a similar docking site in their TRFH domains, they bind different proteins, as shown by the differential recruitment of TIN2 and Apollo by TRF1 and TRF2, respectively (1). This versatility can be explained by substantial differences outside the TBM within the TIN2 and Apollo sequences but also within the TRF1 and TRF2 TRFH domains (1) (Fig. S8). In contrast, only one TRF-like protein is found in several nonvertebrate eukaryotes. In these cases, the Myb domain is preceded by a helical domain sharing distant relations with TRFH domains [e.g., in *Ciona intestinalis* (GenBank accession no. 198426240) and in *Nematostella vectensis* (GenBank accession no. 156407029)]. A highly divergent TRF-like domain was also proposed in a protein from *Trypanosoma brucei* (2) and in the fission yeast (*Schizosaccharomyces pombe*) TAZ1 (3). Like TRF2 in mammals, spTAZ1 recruits spRAP1 to *S. pombe* telomeres (4). In contrast, in *Saccharomyces cerevisiae*, no TRF orthologue is apparently present; however, in this case, the acquisition of direct DNA-binding activity by Rap1 enables the particular telomeric sequence of budding yeasts to function without a TRF module (5). A remnant of TRF, Tbf1, exists in *S. cerevisiae*, but this protein does not bind to telomeric DNA (6). This suggests that the absence of TRF1/TRF2 in *S. cerevisiae* coincides with a change in the sequence of telomeric DNA (5). A Tbf1 protein has also recently been identified in *S. pombe,* but this protein does bind telomeric DNA with high sequence specificity in vitro (7). This suggests that the fission yeast, like mammals, has two factors that bind double-stranded telomeric DNA and perform distinct roles in telomere length regulation. Although clearly distinct at the sequence level, spTAZ1 and spTbf1 might thus be functionally related to mammalian TRF2 and TRF1, respectively.

1. Chen Y, et al. (2008) A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science* 319:1092–1096.
2. Li B, Espinal A, Cross GA (2005) Trypanosome telomeres are protected by a homologue of mammalian TRF2. *Mol Cell Biol* 25:5011–5021.
3. Li B, Oestreich S, de Lange T (2000) Identification of human Rap1: Implications for telomere evolution. *Cell* 101:471–483.
4. Kanoh J, Ishikawa F (2001) spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast. *Curr Biol* 11:1624–1630.
5. Palm W, de Lange T (2008) How shelterin protects mammalian telomeres. *Annu Rev Genet* 42:301–334.
6. Koering CE, et al. (2000) Identification of high affinity Tbf1p-binding sites within the budding yeast genome. *Nucleic Acids Res* 28:2519–2526.
7. Pitt CW, Valente LP, Rhodes D, Simonsson T (2008) Identification and characterization of an essential telomeric repeat binding factor in fission yeast. *J Biol Chem* 283: 2693–2701.

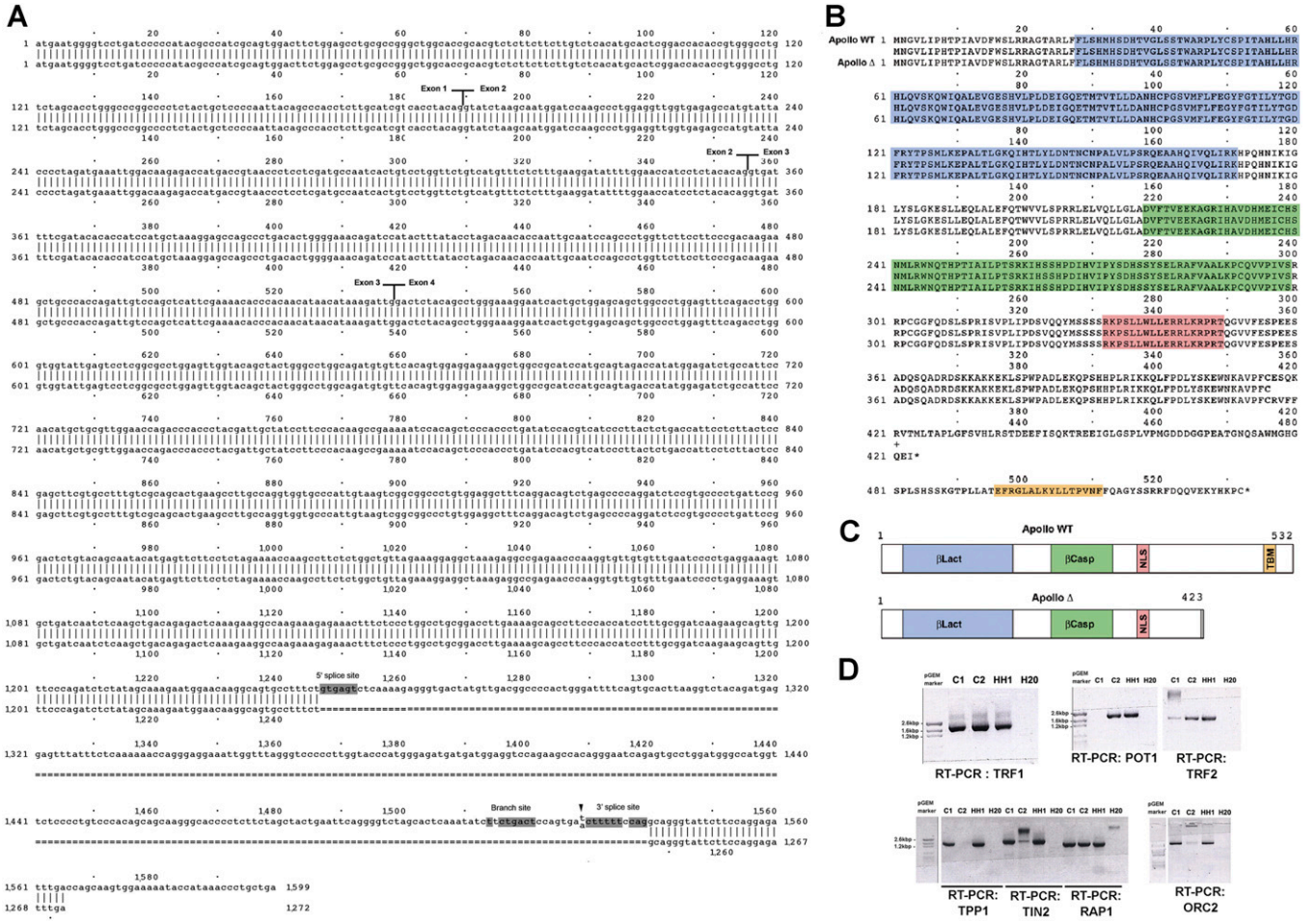

Fig. S1. *Apollo* splice variant in cells of HH1. Alignment of sequences of human Apollo and Apollo-Δ. (*A*) Alignment of coding sequences of human *Apollo* and *Apollo*-Δ. Gray boxes highlight the cryptic splice sites, including the 5′ splice site, the branch site, and the 3′ splice site. The junctions of the different exons are indicated. (*B*) Protein sequence (one-letter code) alignment of human Apollo (*Upper*) and Apollo-Δ (*Lower*). The metallo–β-lactamase fold (blue box), β-CASP domain (green box), nuclear localization domain (pink box), and TRF2-binding motif (TBM; orange box) are indicated. (*C*) Scheme representing WT and truncated Apollo resulting from the intraexonic splice. (*D*) Normal RT-PCR products of several genes involved in telomere protection demonstrate that there is not a general defect of the splicing machinery in cells of HH1.

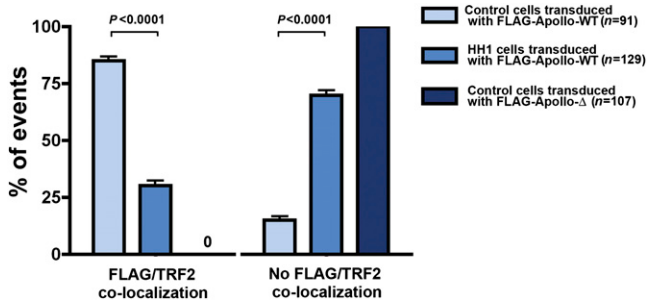

Fig. S2. Colocalization of FLAG-Apollo-WT and TRF2 in HH1 and control primary fibroblasts. Primary control fibroblasts and fibroblasts of HH1 were transduced with FLAG-Apollo-WT– or FLAG-Apollo-Δ–expressing vector. The percentage of cells presenting with colocalization of TRF2 with FLAG labeling, as measured by immunofluorescence using anti-FLAG and anti-TRF2 antibodies, is represented. Only a small fraction of cells of HH1 displayed colocalization of FLAG-Apollo-WT with TRF2, whereas the vast majority of control fibroblasts showed colocalization of FLAG-Apollo-WT with TRF2. Conversely, the FLAG-Apollo-Δ molecules expressed in control fibroblasts were systematically found diffuse in the nucleus without colocalization with TRF2. These results highlight the robust dominant negative effect of endogenous Apollo-Δ in cells of HH1.

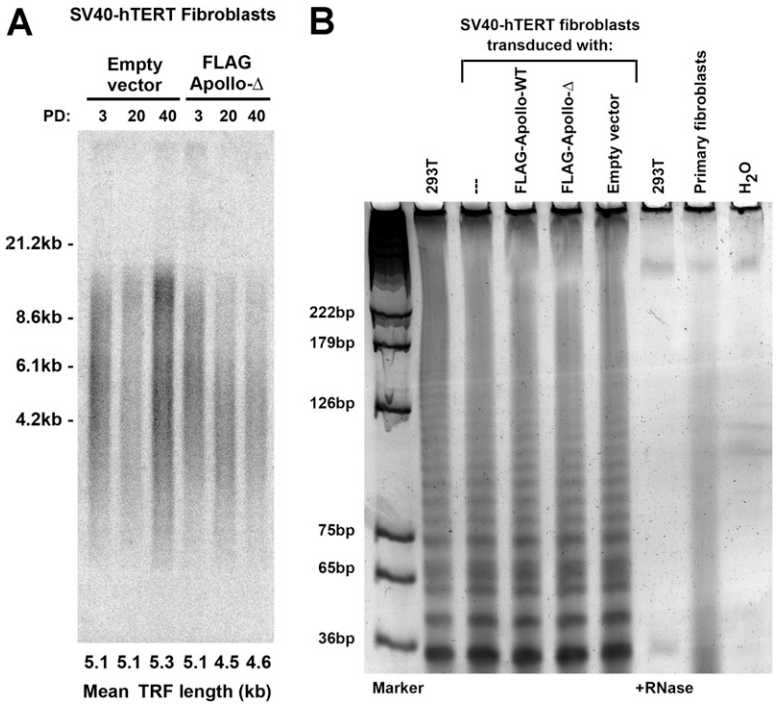

Fig. S3. FLAG-Apollo-Δ induces telomere shortening but not inhibition of the in vitro telomerase activity in SV40-hTERT fibroblasts. (*A*) Telomere length was measured by the TRF method in SV40-hTERT fibroblasts transduced with empty vector or FLAG-Apollo-Δ–expressing vector at PD3, PD20, and PD40. Telomere length was estimated by digital image analysis and revealed telomere shortening in cells expressing FLAG-Apollo-Δ. (*B*) In vitro telomerase activity was analyzed by telomere repeat amplification protocol (TRAP) assay with cell extracts from 293T cells, SV40-hTERT fibroblasts transduced or not transduced with empty vector and with FLAG-Apollo-WT– or FLAG-Apollo-Δ–expressing vector. As expected, no telomerase activity was detected with RNase-treated extracts from 293T cells, with extracts from primary fibroblasts, and without extract (H_20).

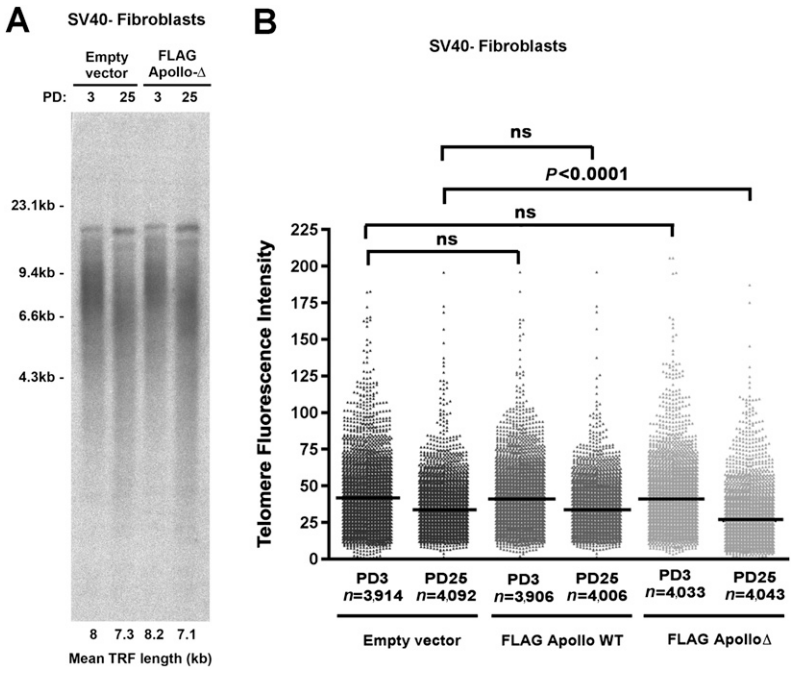

Fig. S4. Expression of FLAG-Apollo-Δ accelerates the telomere attrition of SV40 fibroblasts. (*A*) Telomere length was measured by the TRF method in SV40 fibroblasts transduced with empty vector or FLAG-Apollo-Δ–expressing vector at PD3 and PD25. Mean telomere length was estimated by digital image analysis. (*B*) Quantitative FISH analysis of telomeric signals obtained from metaphase spreads of SV40 fibroblasts transduced by an empty vector, a FLAG-Apollo-WT–expressing vector, or a FLAG-Apollo-Δ–expressing vector after PD3 and PD25. Individual and mean values are presented. Both telomere length measurement methods revealed that the telomere shortening normally occurring in SV40 fibroblasts (which do not express hTERT) is accelerated in the presence of FLAG-Apollo-Δ.

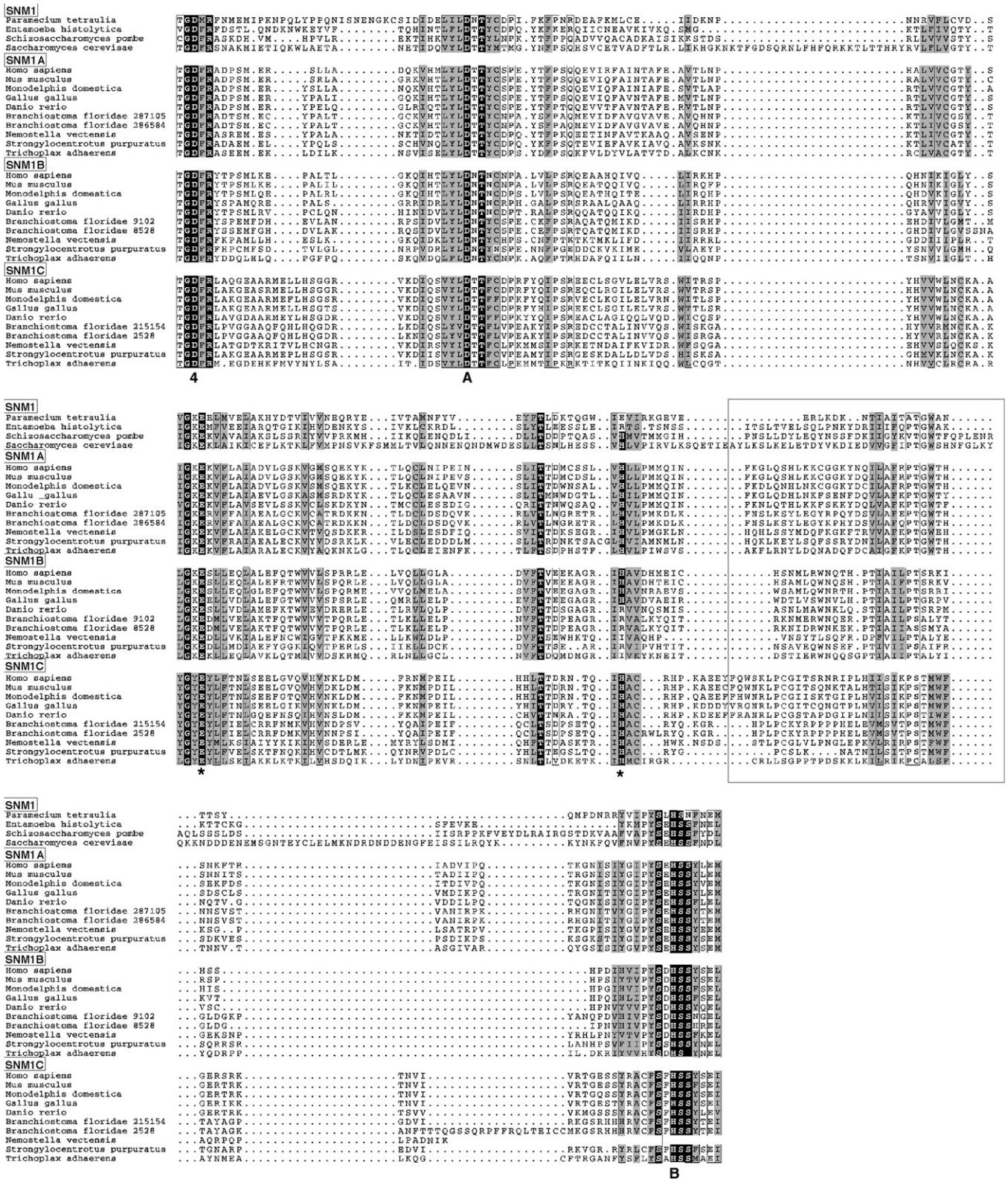

Fig. S5. Multiple alignment of SNM1 sequences focused on the β-CASP domain, typical of metallo–β-lactamases of the β-CASP family and located between the conserved motifs A and B. This alignment allows clear separation of the SNM1 sequences into three groups (SNM1A, SNM1B, and SNM1C) in metazoans (particularly in the boxed sequences), whereas only one sequence (SNM1) can be found in unicellular organisms. The two highly conserved residues common to the β-CASP domains of SNM1A, SNM1B, and SNM1C are shown with stars. GenBank accession numbers for SNM1A, SNM1B (Apollo), and SNM1C (Artemis), respectively, are as follows: 73620749, 73620756, 71153325 (*Homo sapiens*), 73620752, 73620758, 71153326 (*Mus musculus*), 126273412, 126311625, 126340452 (*Monodelphis domestica*), 73620743, 73620753, 71153324 (*Gallus gallus*), 189526049, 169145639, 92096533 (*Danio rerio*), 219491749 (BRAFLDRAFT_287105)/

Legend continued on the following page

219489377 (BRAFLDRAFT_286584), 219461437 (BRAFLDRAFT_91025)/219450530 (BRAFLDRAFT_85286), 219431795 (BRAFLDRAFT_215154)/219499922 (BRAFLDRAFT_252863) (*Branchiostoma floridae*), 115616464, 115923314, 115735528 (*Strongylocentrotus purpuratus*), 156390186, 156399461, 156404099 (*Nematostella vectensis*), 196013719, 196000276, and 195999584 (*Trichoplax adhaerens*). GenBank accession numbers for SNM1: 145553259 (*Paramecium tetraurelia*), 67473862 (*Entamoeba histolytica*), 267010 (*Saccharomyces cerevisiae*), and 19862928 (*Schizosaccharomyces pombe).*

Fig. S6. Sequence alignment of the conserved C-terminal sequences of Apollo proteins from vertebrates. The identical amino acids making up part of the TBM consensus are shown in white on a black background, whereas other similarities are boxed (shaded gray for hydrophobic amino acids). The segment of human Apollo whose 3D structure was analyzed in complex with TRF2 is shown with an arrow (Protein Data Bank ID code 3BUA, chain F). GenBank accession numbers are as follows: 73620756 (*Homo sapiens*), 73620758 (*Mus musculus*), 126311625 (*Monodelphis domestica*), 73620753 (*Gallus*), and 169145639 (*Danio rerio*). The *Xenopus tropicalis* partial sequence was extracted from the EST database (GenBank accession no. 50394267).

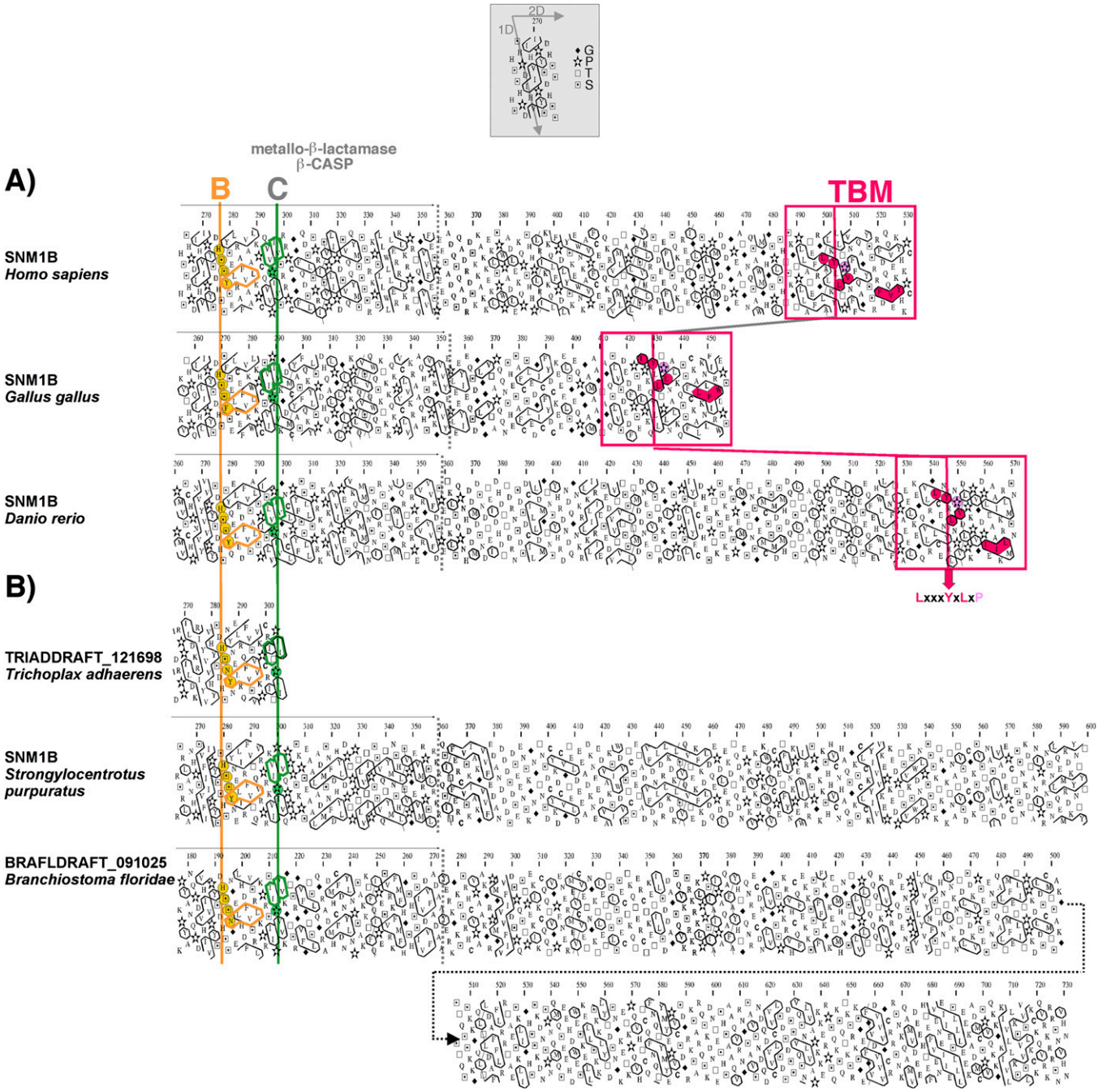

Fig. S7. HCA of the C-terminal part of Apollo in different species. HCA allows the comparison of highly divergent sequences by combining the comparison of primary structures with that of secondary structures, which are more conserved. The sequence is shown on a duplicated α-helical net, in which hydrophobic residues (VILFMYW) are surrounded. These form clusters, which mainly correspond to the internal faces of regular secondary structures. (*Inset*) How to read the sequences and secondary structures, as well as special symbols, is indicated. This methodology has previously been used to analyze the highly divergent sequences of the metallo–β-lactamase/β-CASP family. (*A*) HCA analysis of the C-terminal extension of the Apollo sequences in vertebrate sequences led to highlighting two conserved clusters (shaded pink) in the C termini, which are included in a small globular domain (boxed) bearing the TRF2-binding motif (TBM, LxxxYxLxP). The TBM is found in Apollo sequences from vertebrate species (*A*) and is apparently absent in Apollo sequences from nonvertebrate species, three of which are shown (*B*). Note that the Apollo sequences from some nonvertebrate species, such as *Trichoplax adhaerens* as well as *Nematostella vectensis*, have no extension C terminal to the metallo–β-CASP domain. The GenBank accession numbers of the sequences can be found in the legend for Fig. S5.

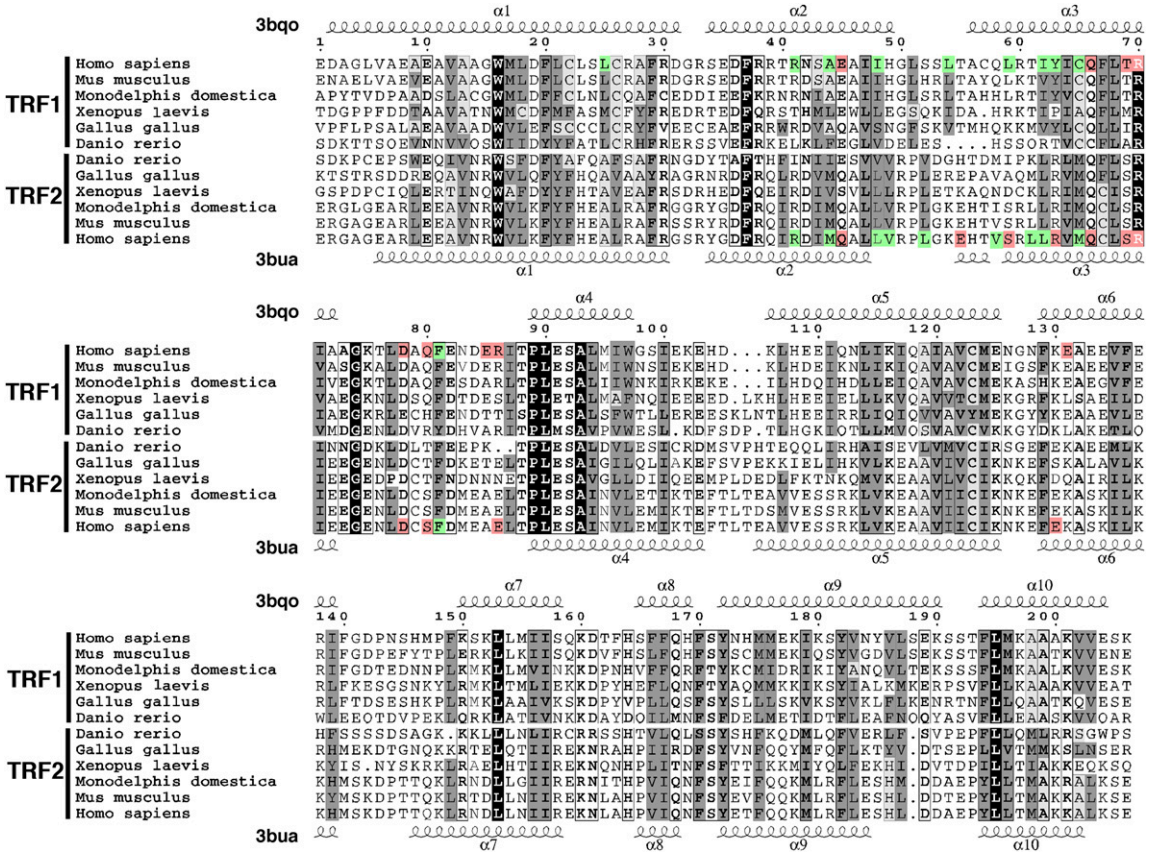

Fig. S8. Multiple alignment of TRF1/TRF2 sequences from vertebrates. Multiple alignment of the TRF1/TRF2 sequences from several vertebrates, from human to fish, focused on the TRFH domains. Identities are shown in white on a black background, whereas similarities between hydrophobic amino acids (V, I, L, M, F, Y, W) are shaded in gray [light gray for residues that can substitute them in a context-dependent way (A, C, T, S)]. Other striking similarities are boxed. Green and pink boxes highlight residues having hydrophobic and hydrophilic interactions with the TIN2 (for TRF1 and TRF2) and Apollo (for TRF2 only) TBMs, respectively, as reported [Chen Y, et al. (2008) A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. Science 319:1092–1096]. Observed secondary structures [Protein Data Bank ID codes 3bqo (TRF1) and 2bua (TRF2)] are reported above and below the human TRF1 and TRF2 sequences, respectively. GenBank accession numbers for TRF1 and TRF2 are, respectively, 206729904 and 21542277 (*Homo sapiens*), 2499054 and 158515400 (*Mus musculus*), 126321276 and 126305027 (*Monodephis domestica*), 213623663 and 184191025 (*Xenopus laevis*), 33317668 and 21542298 (*Gallus gallus*), and 126632156 and 67677850 (*Danio rerio*). The highly divergent TRF-like domains found in some hypothetical proteins from complete genomes of nonvertebrate species, such as *Nematostella vectensis* (GenBank accession no. 156407029) or *Ciona intestinalis* (GenBank accession no. 198426420), were not reported on this multiple alignment because they cannot be aligned in an accurate way over their whole length.

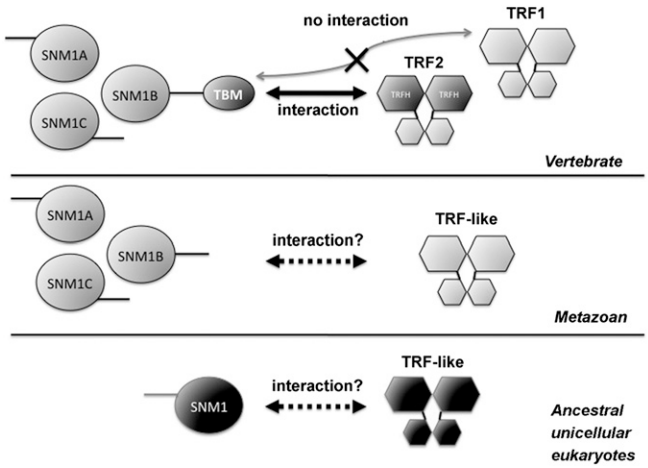

Fig. S9. Coemergence of the TBM motif of Apollo/SNM1B and the TRF1/TRF2 paralogues. The scheme illustrates the appearance of paralogous SNM1 and TRF1/TRF2 sequences and the coemergence of the TBM of Apollo/SNM1B and TRF2. Phylogenetic analysis suggests that the Apollo TBM domain may have coemerged with TRF1/TRF2 to recognize the vertebrate-specific TRFH domain and, at the same time, to distinguish between the two TRF paralogues, therefore enabling Apollo to interact specifically with TRF2.