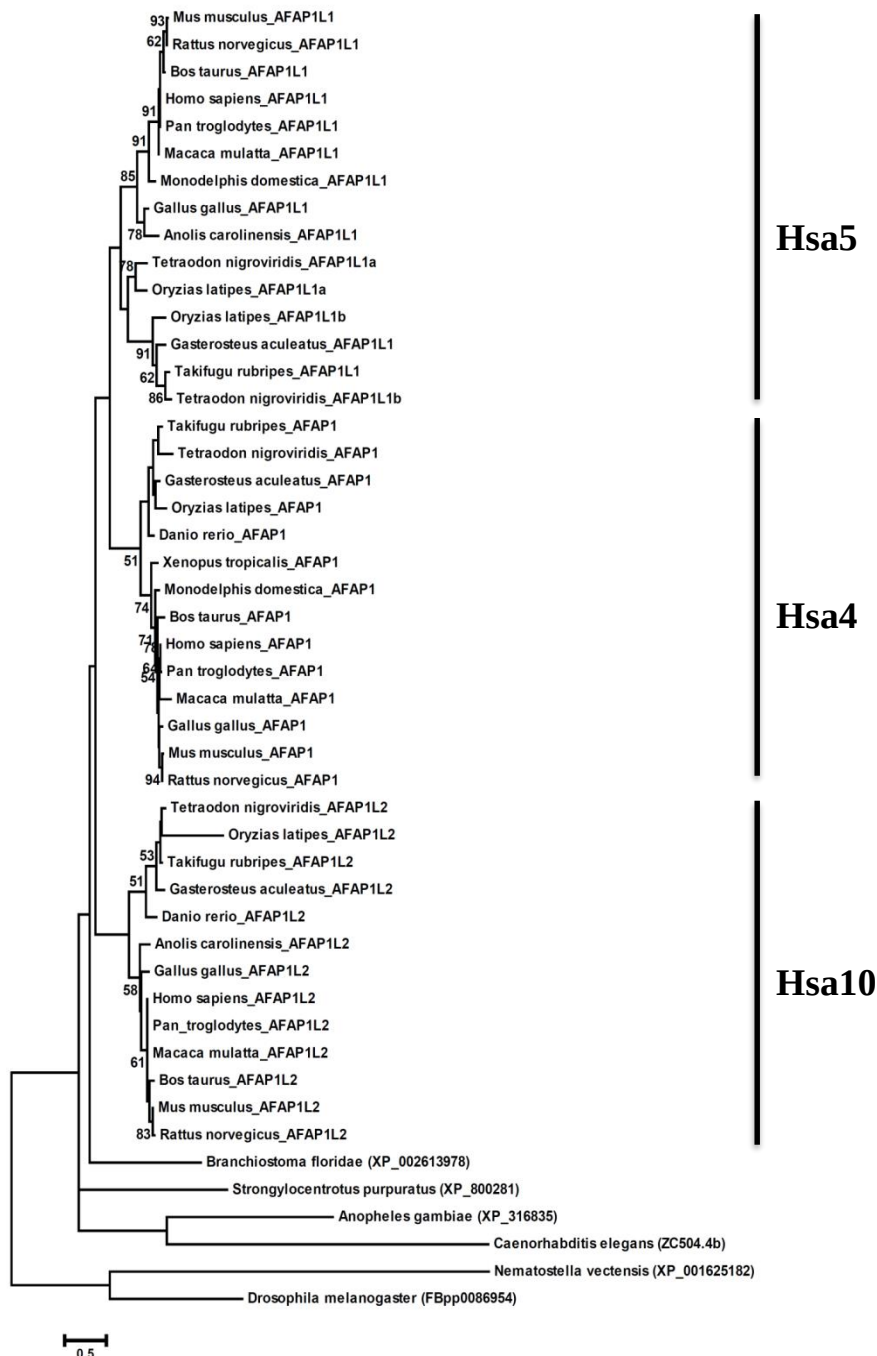


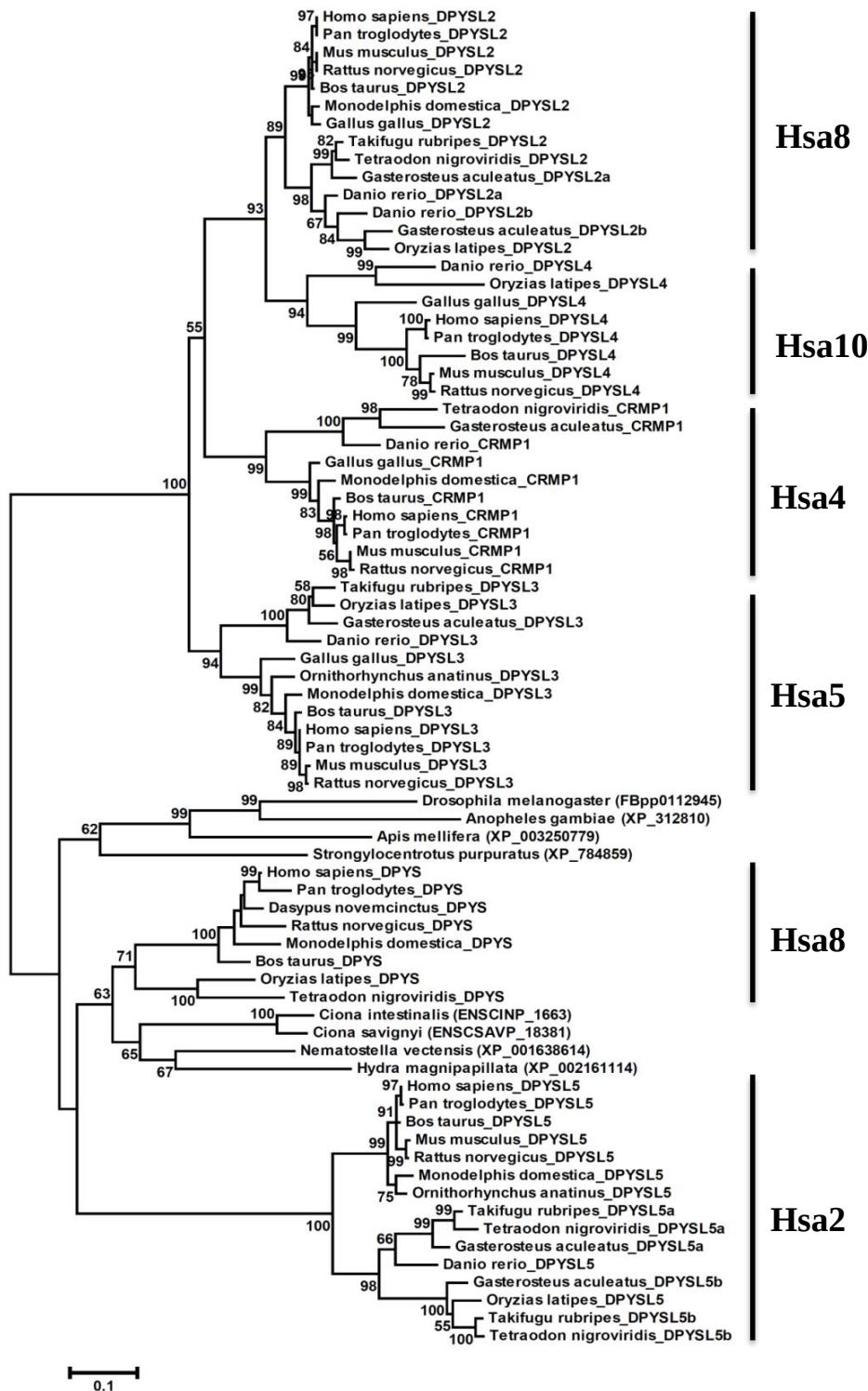
**Maximum Likelihood Trees of gene families residing on
chromosomes 4, 5, 8 and 10.**

Actin filament associated protein - AFAP



Phylogenetic tree of AFAP family.

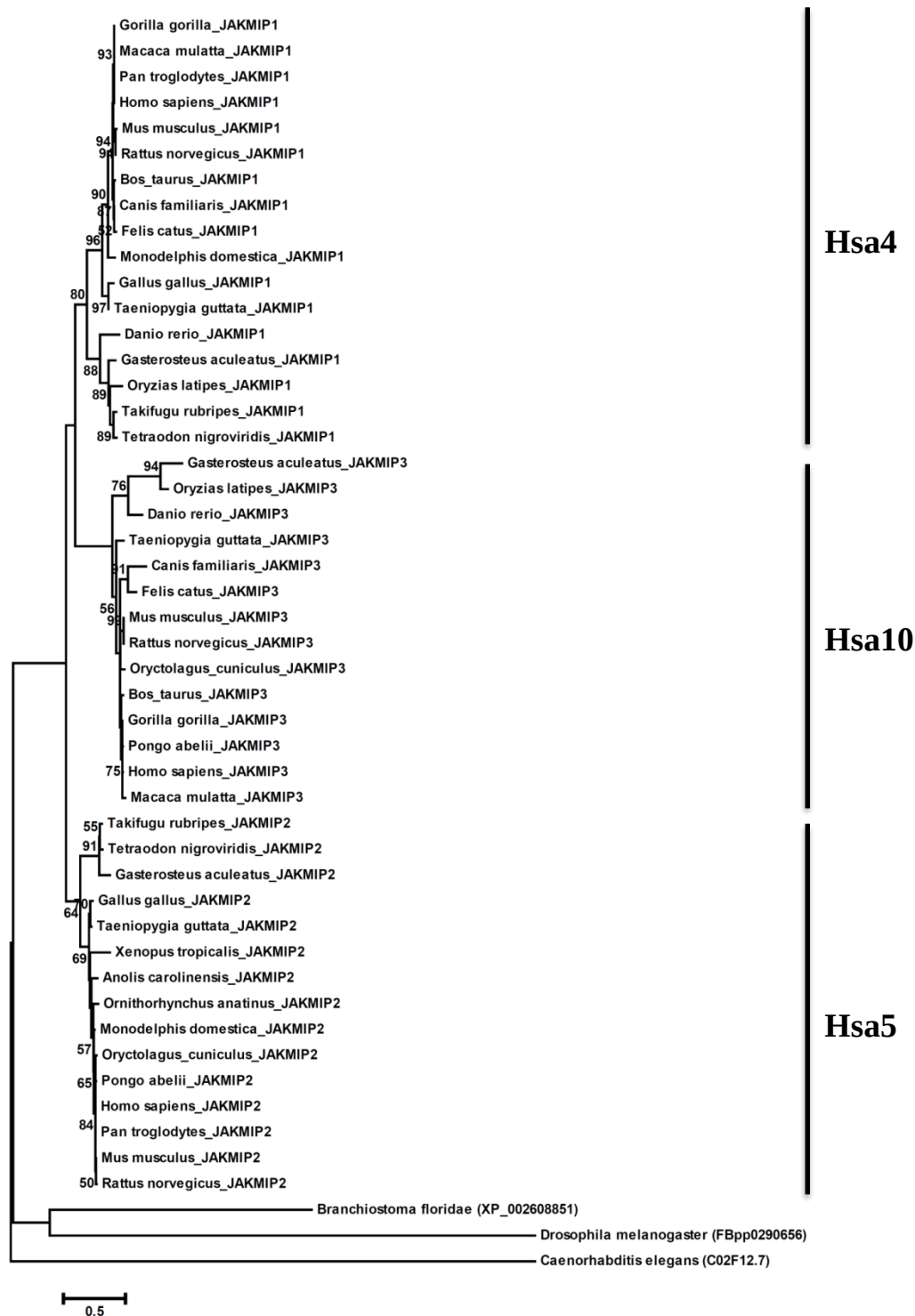
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-5455.6729) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 6.1444)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 48 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 117 positions in the final dataset.



Phylogenetic tree of DPYSL family.

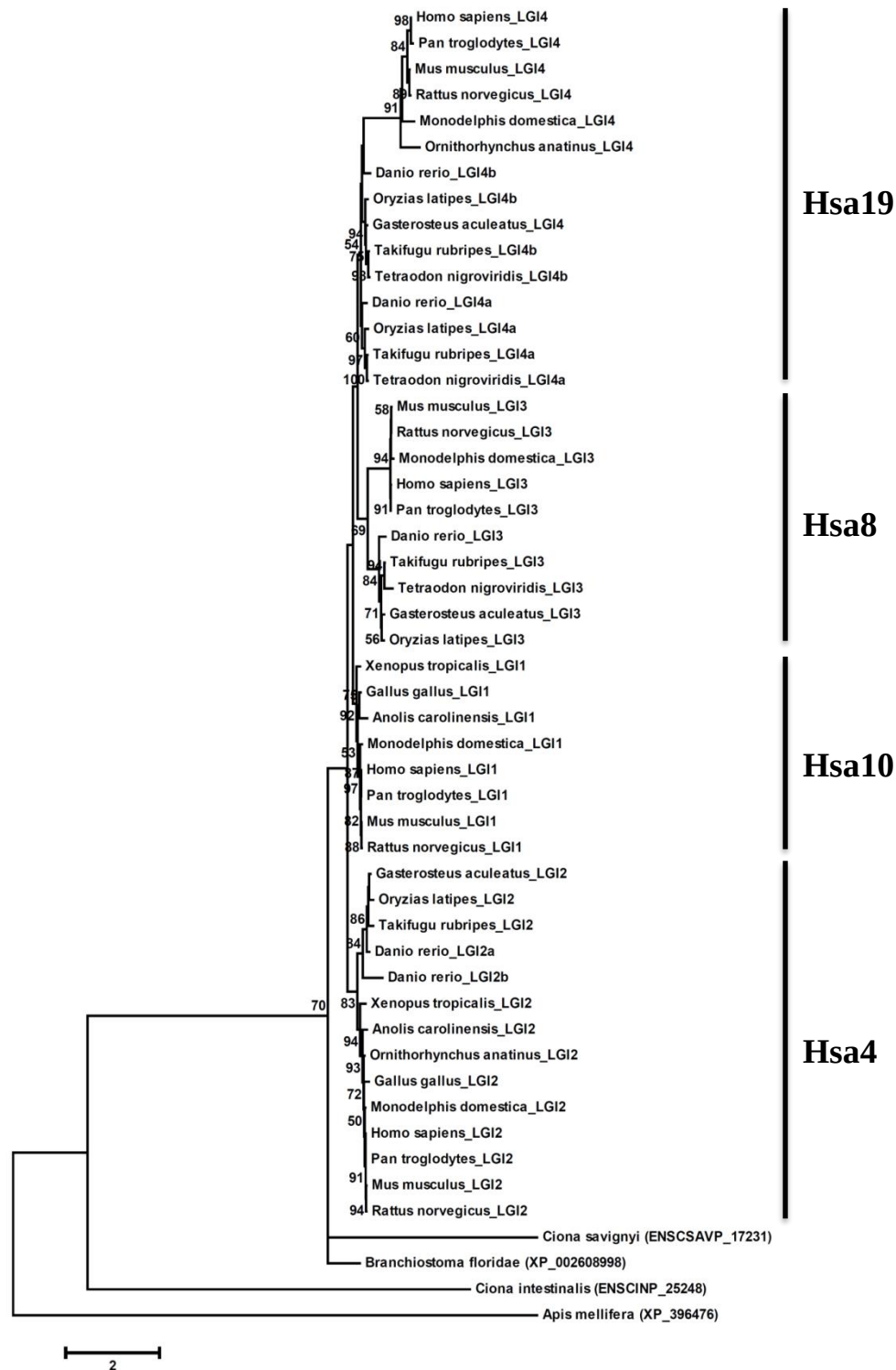
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-14764.6713) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.4959)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 75 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 419 positions in the final dataset.

Janus kinase and microtubule interacting protein – JAKMIP



Phylogenetic tree of JAKMIP family.

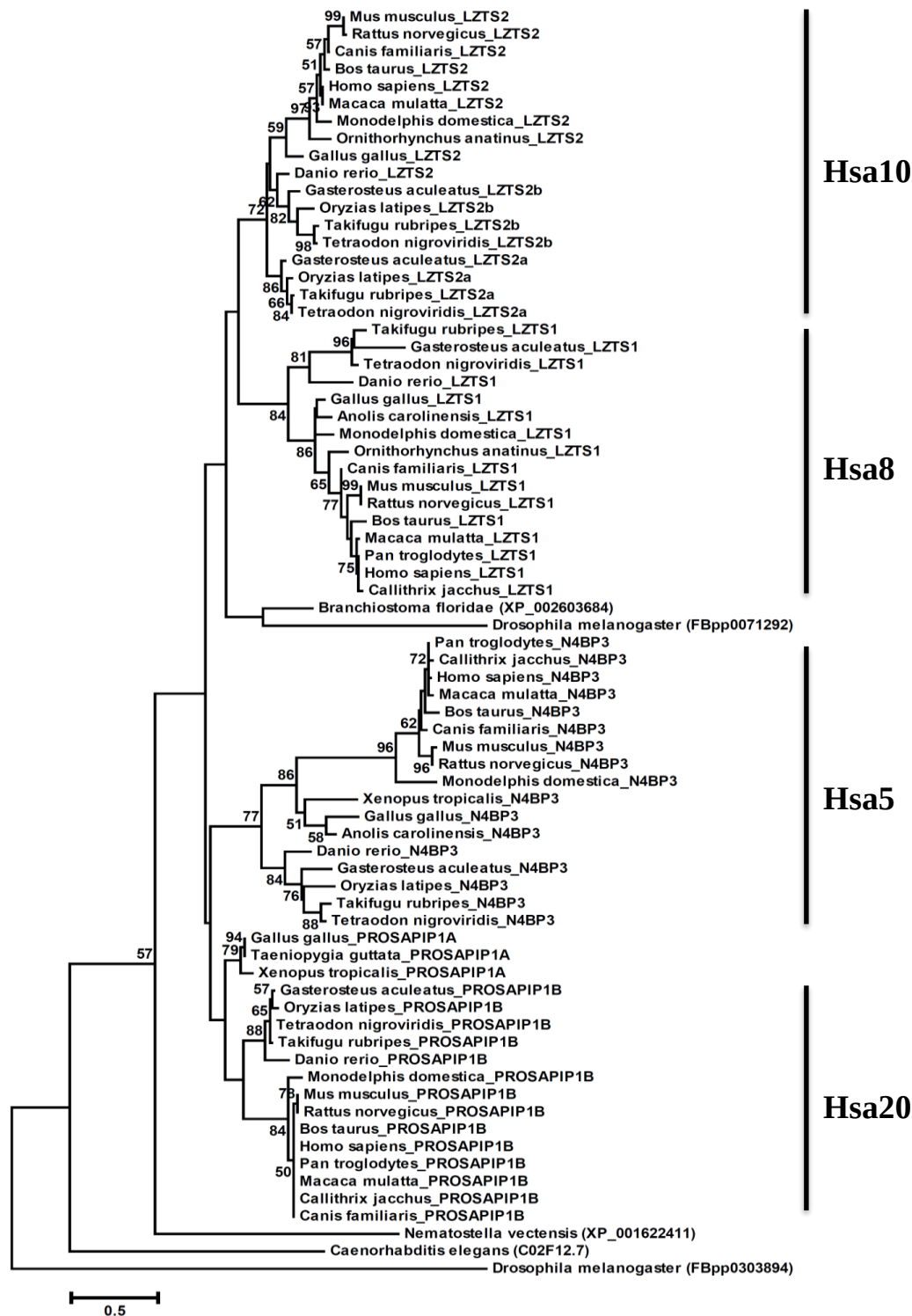
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-7290.3883) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 5.7161)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 49 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 262 positions in the final dataset.



Phylogenetic tree of LGI family

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-12291.9264) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 4.1213)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 51 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 313 positions in the final dataset.

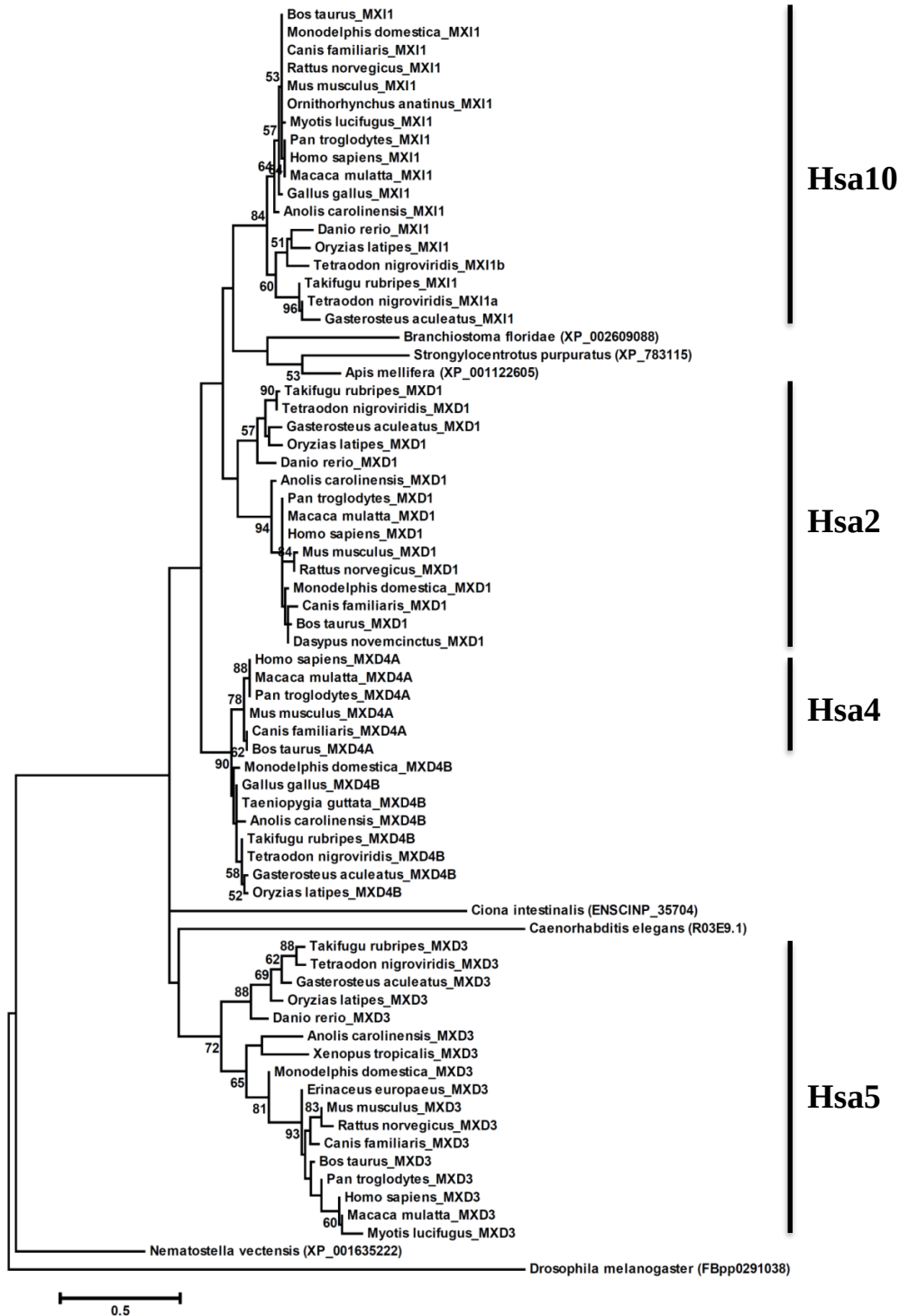
Leucine zipper, putative tumor suppressor – LZTS



Phylogenetic tree of LZTS family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-4727.5570) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 4.7711)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 73 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 89 positions in the final dataset.

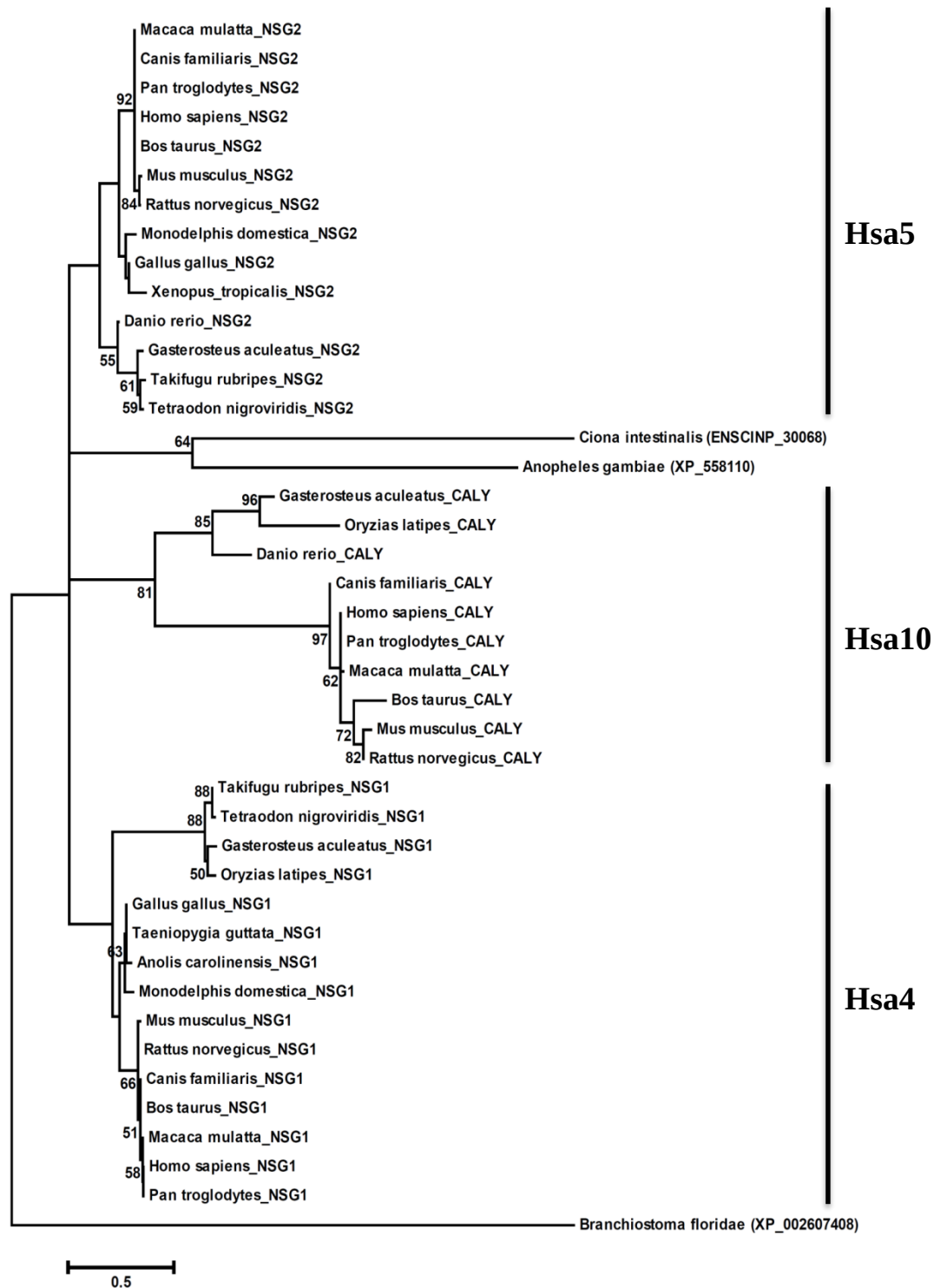
MAX dimerization protein – MXD



Phylogenetic tree of MXD family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-2965.2026) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.9291)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 71 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 83 positions in the final dataset.

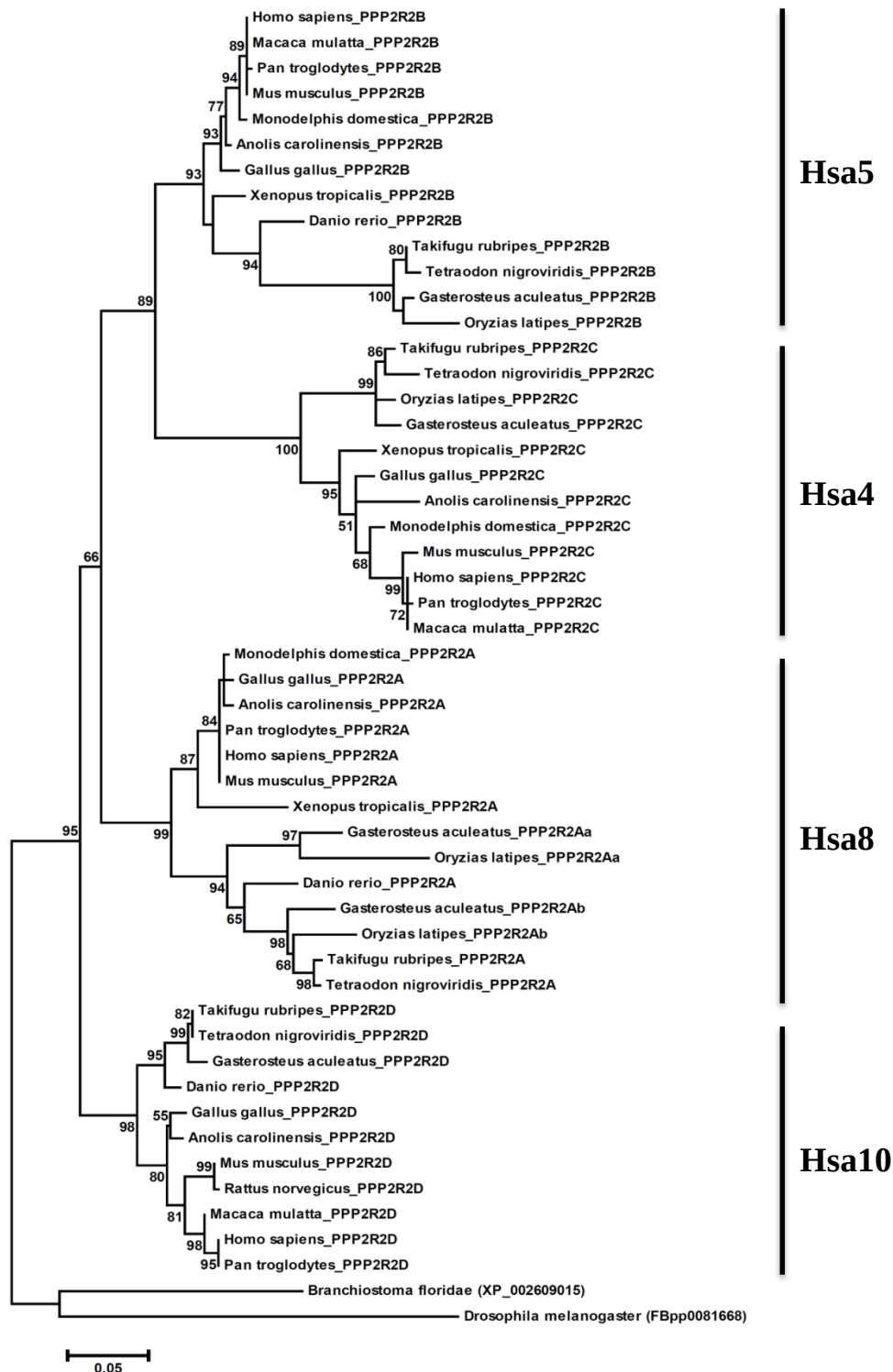
Neuron-specific protein family – NSG



Phylogenetic tree of NSG family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-2400.9740) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 10.3481)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 42 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 89 positions in the final dataset.

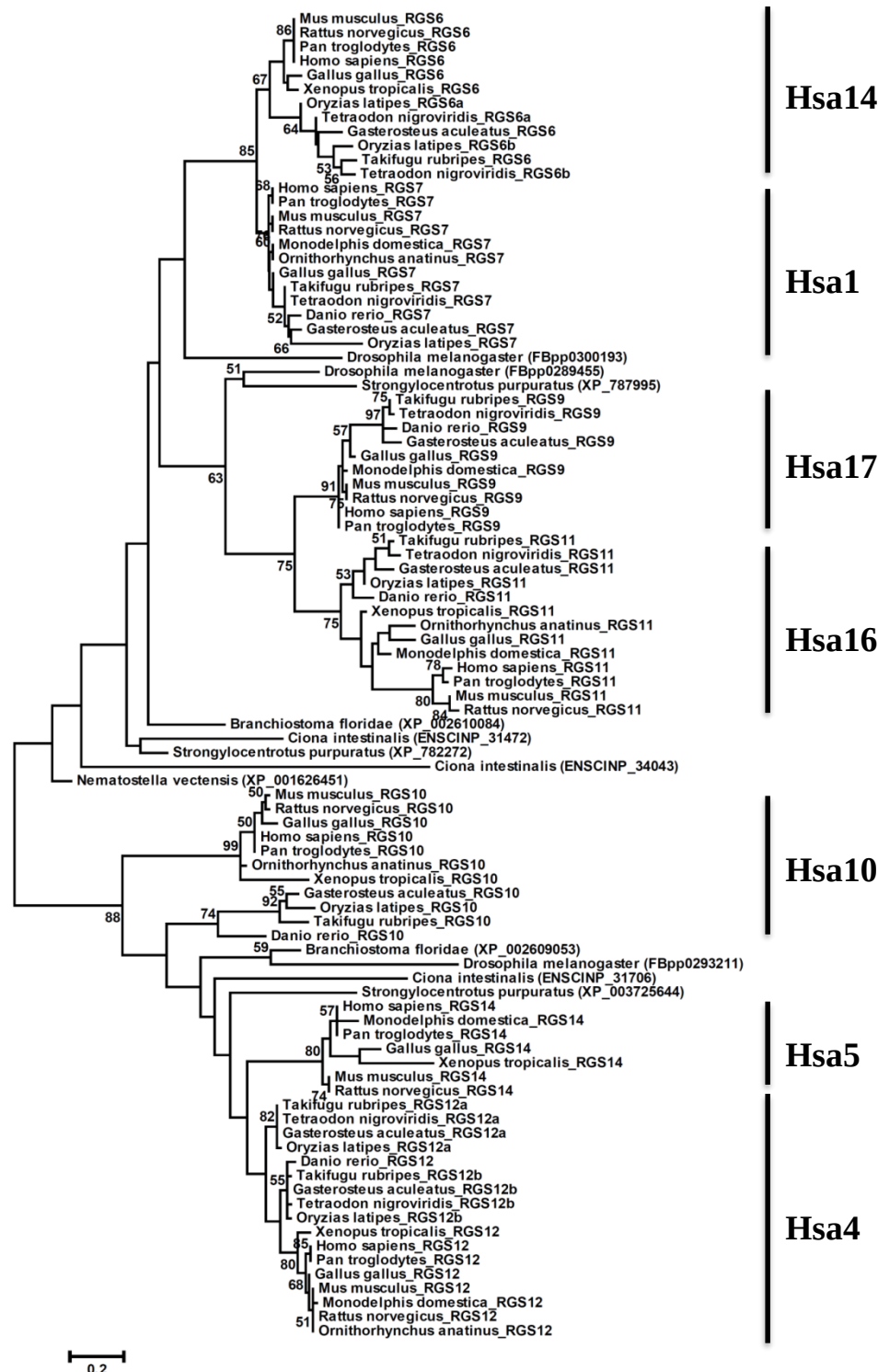
Protein phosphatase 2, regulatory subunit B – PPP2R2



Phylogenetic tree of PPP2R2 family.

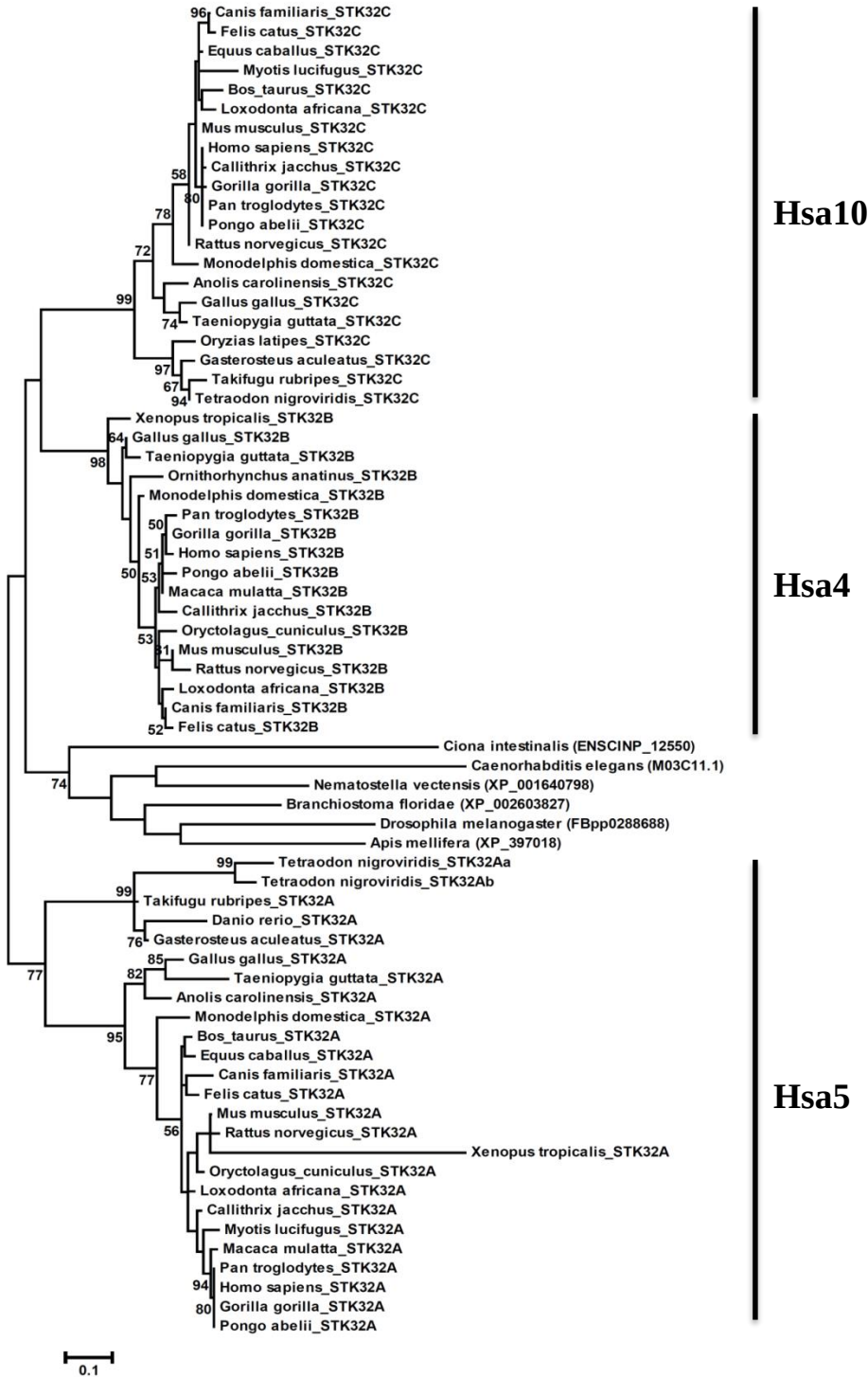
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-4665.5328) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3691)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 52 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 419 positions in the final dataset.

Regulator of G-protein signaling – RGS



Phylogenetic tree of RGS family.

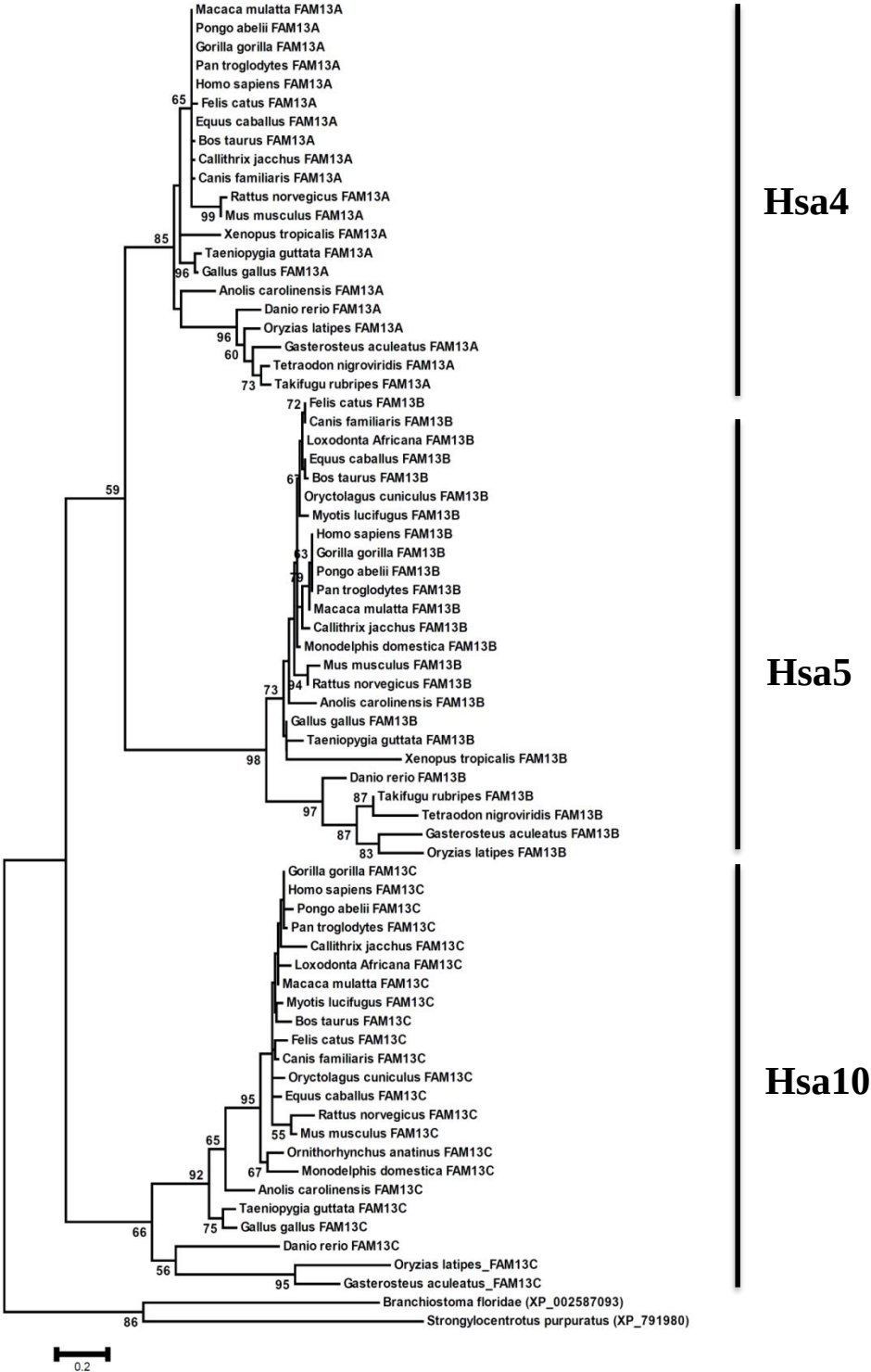
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-3713.9113) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.1437)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 94 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 65 positions in the final dataset.



Phylogenetic tree of STK32 family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-4442.2043) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.7230)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 69 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 135 positions in the final dataset.

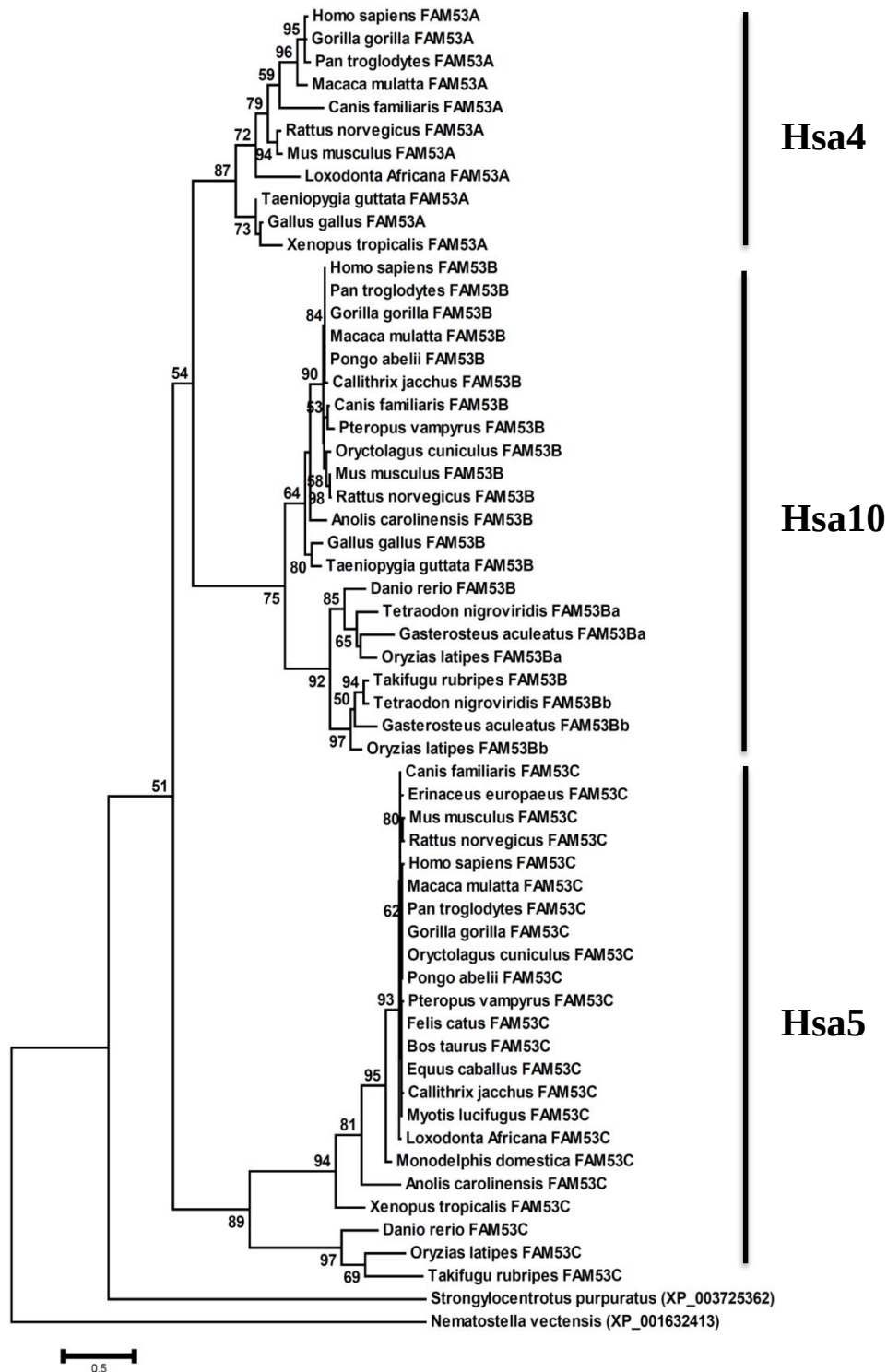
Family with sequence similarity 13-FAM13



Phylogenetic tree of FAM13 family.

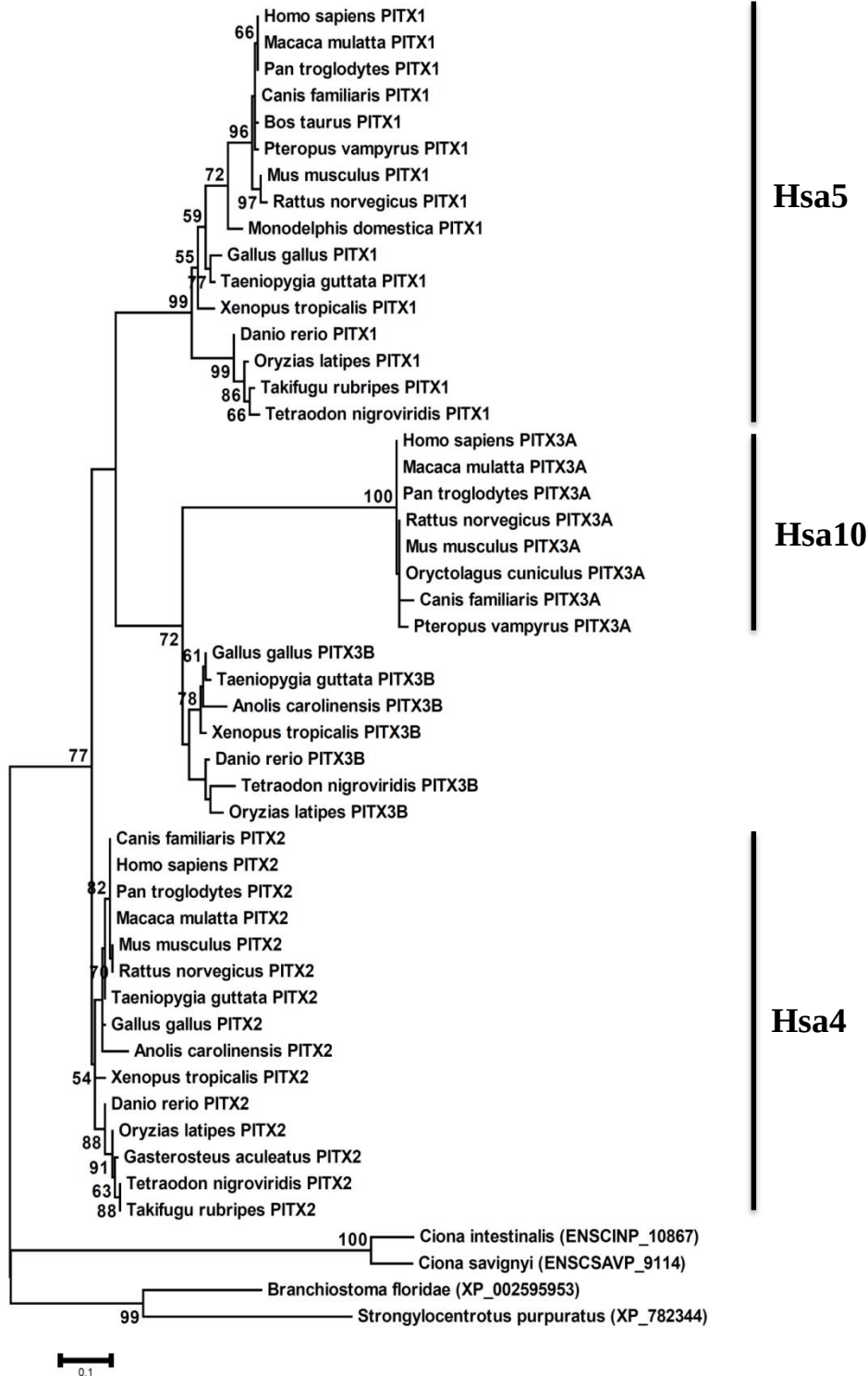
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-3726.2946) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.8374)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 71 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 110 positions in the final dataset.

Family with sequence similarity 53-FAM53



Phylogenetic tree of FAM53 family.

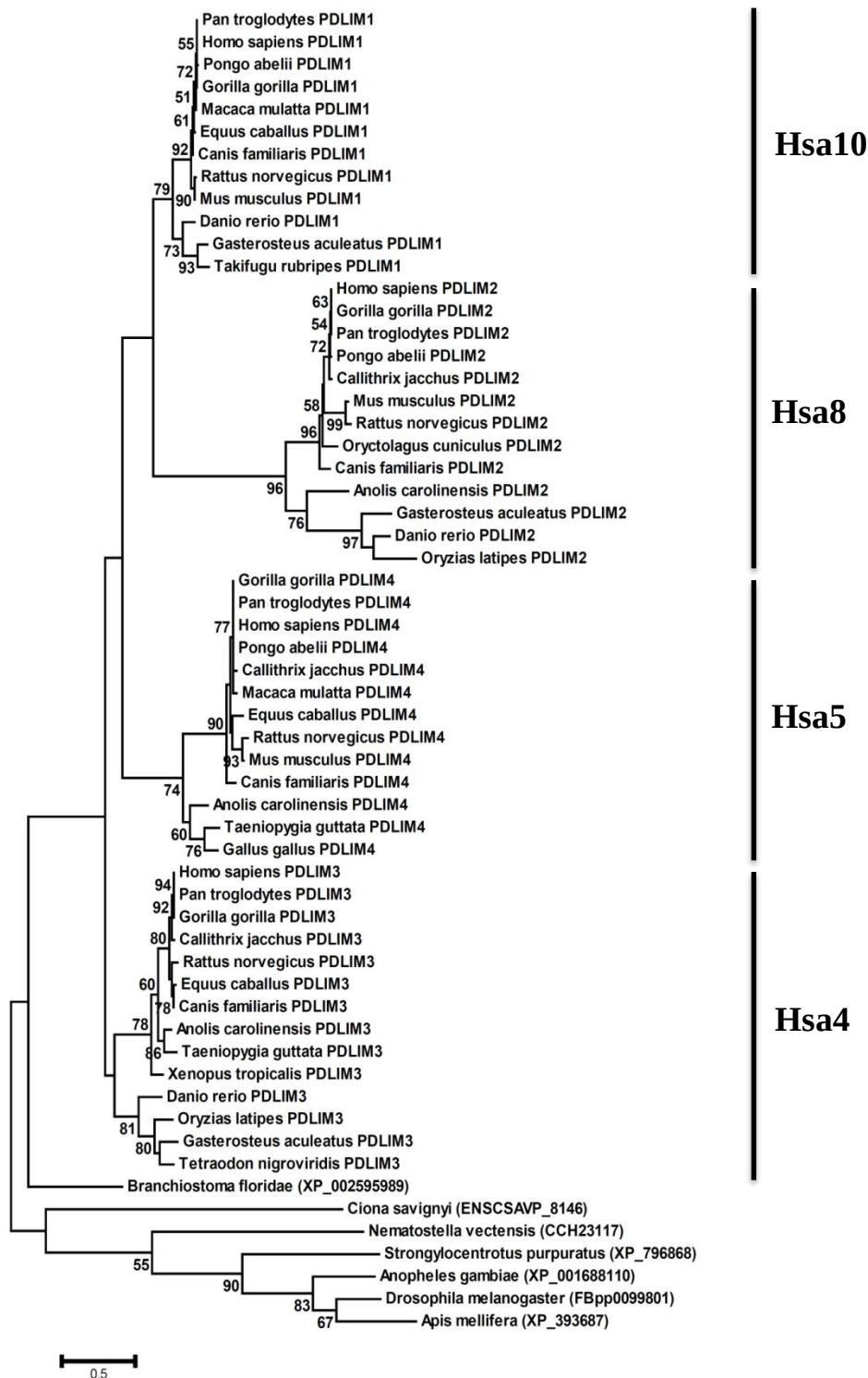
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-4856.7232) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.8016)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 58 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 137 positions in the final dataset.



Phylogenetic tree of PITX family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-3089.9337) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.0181)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 50 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 199 positions in the final dataset.

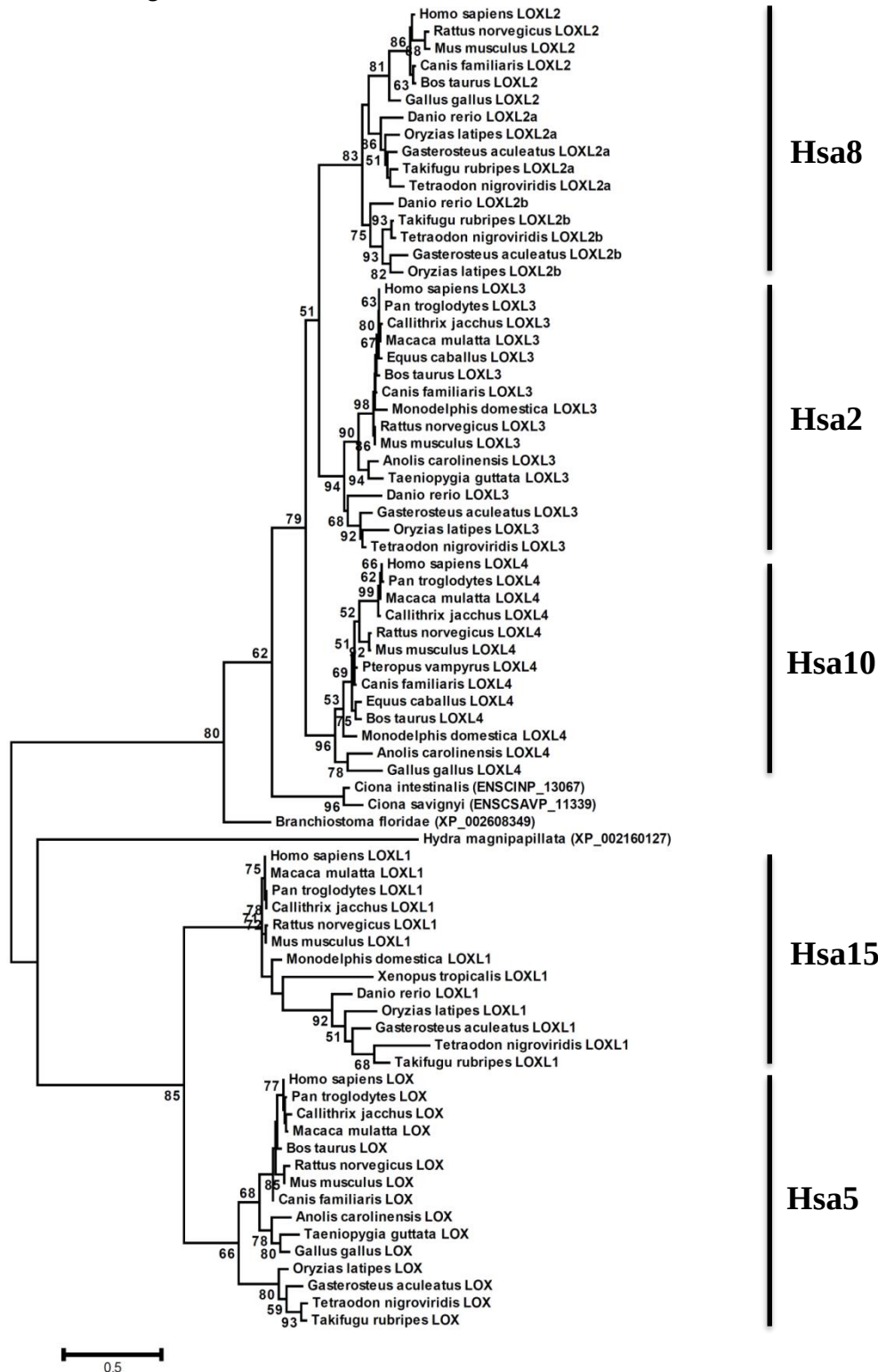
PDZ and LIM domain protein-PDLIM



Phylogenetic tree of PDLIM family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-6403.2852) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 6.1796)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 59 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 129 positions in the final dataset.

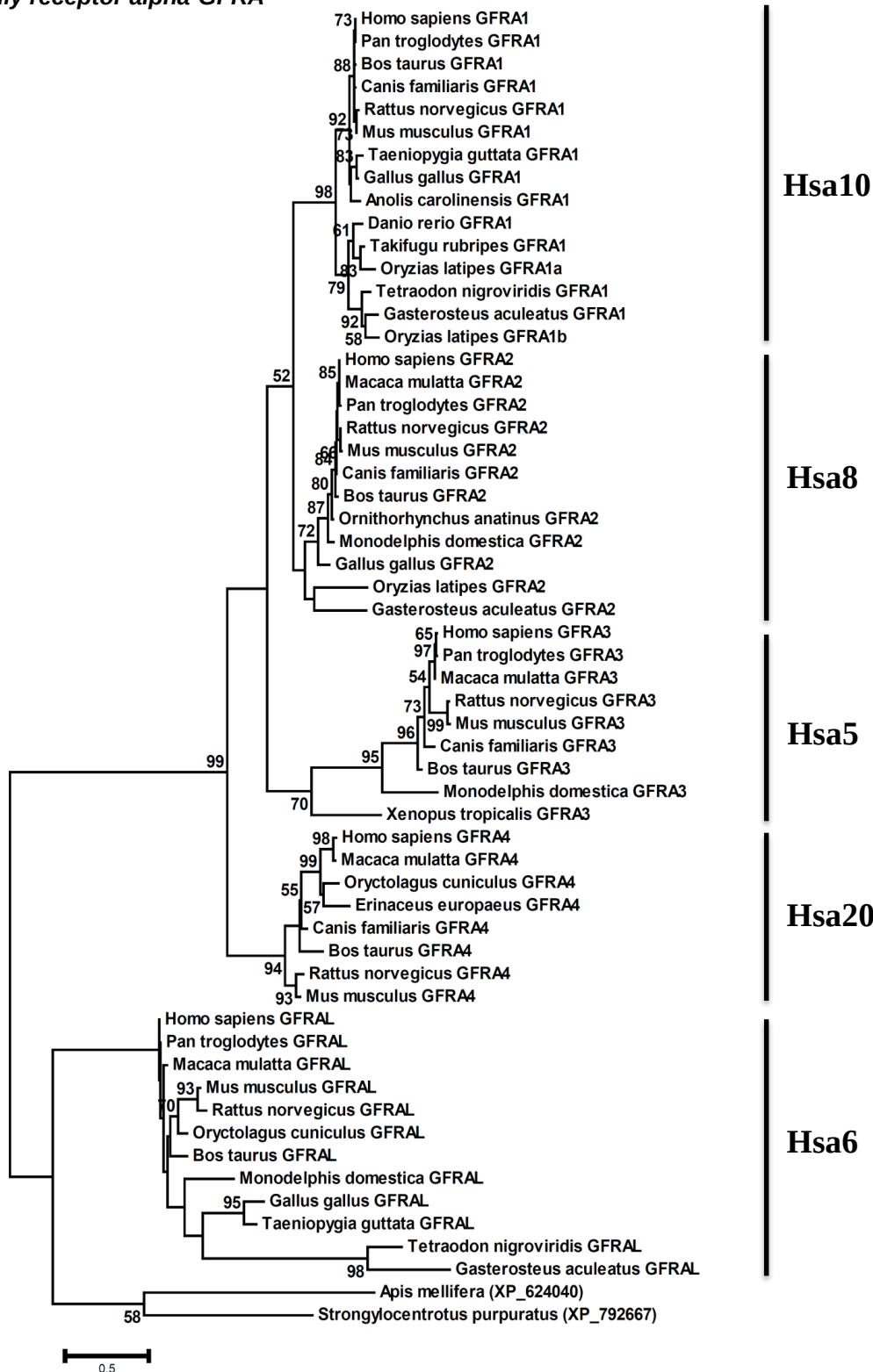
Lysyl oxidase homolog-LOXL



Phylogenetic tree of LOXL family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-7200.2009) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.3659)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 77 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 156 positions in the final dataset.

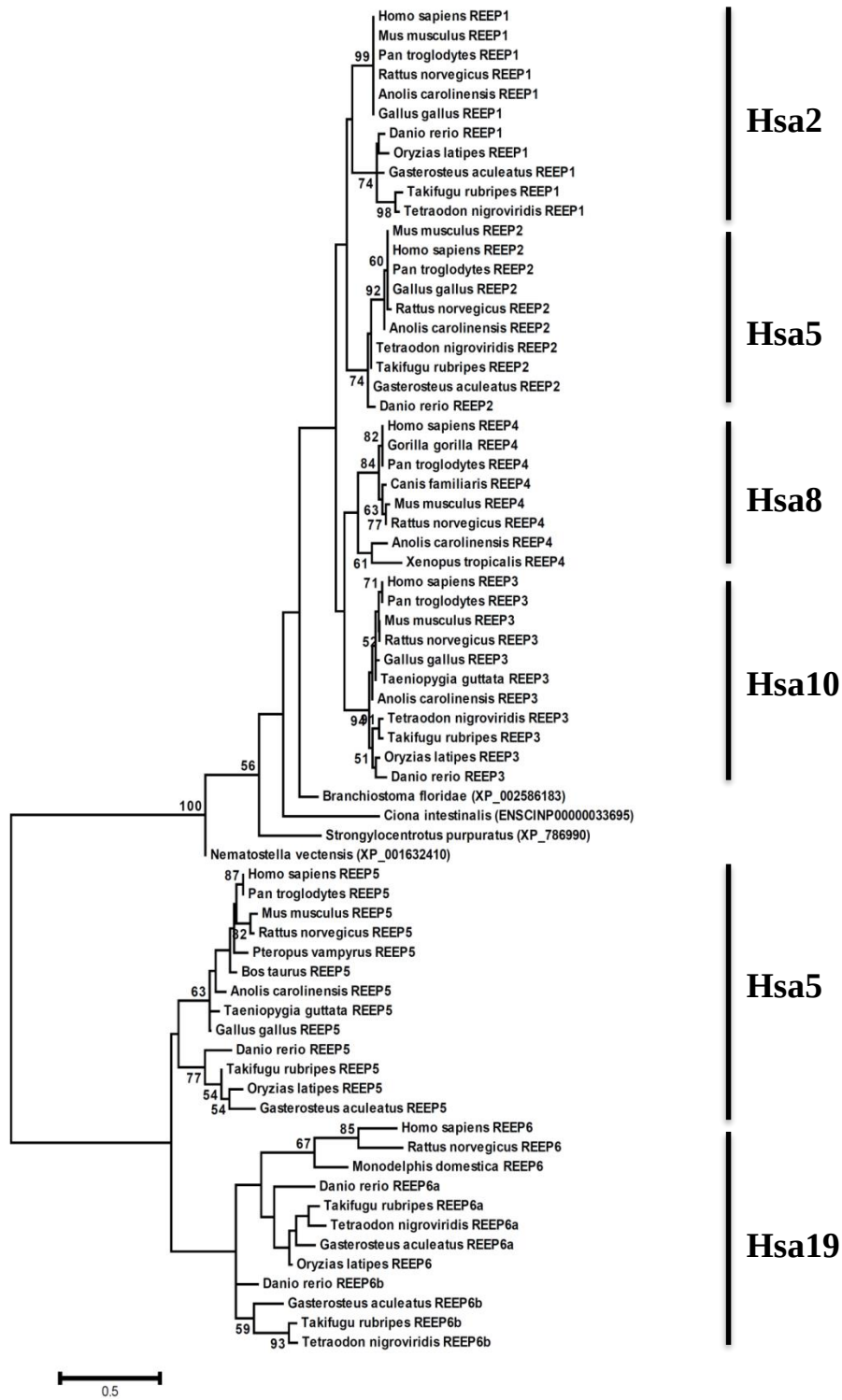
GDNF family receptor alpha-GFRA



Phylogenetic tree of GFRA family.

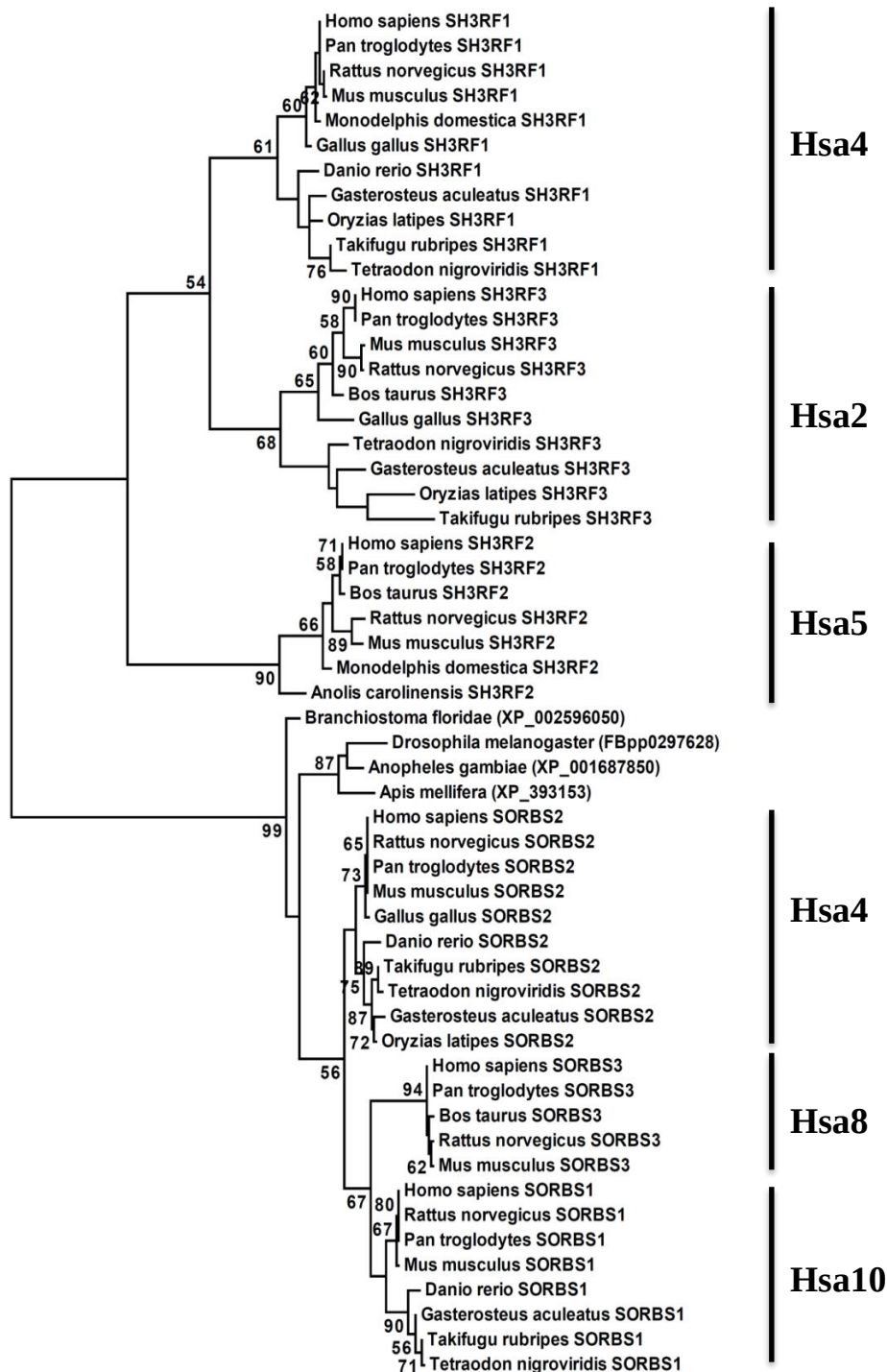
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-6078.0564) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.4353)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 58 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 139 positions in the final dataset.

Receptor expression-enhancing protein-REEP



Phylogenetic tree of REEP family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-2252.4975) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.9823)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 69 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 63 positions in the final dataset.

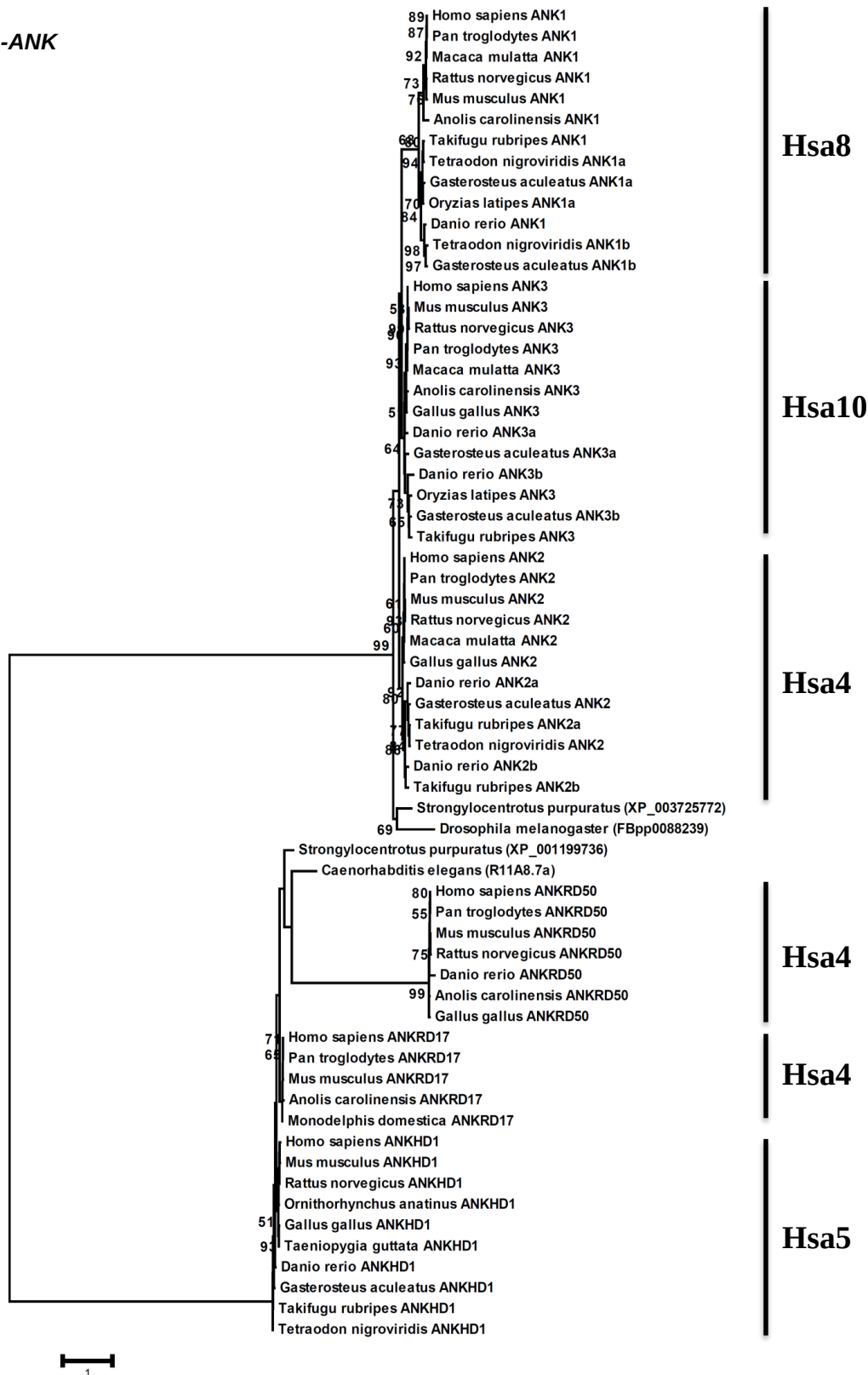


0.5

Phylogenetic tree of SORBS family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-3064.1154) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 3.1266)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 55 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 84 positions in the final dataset.

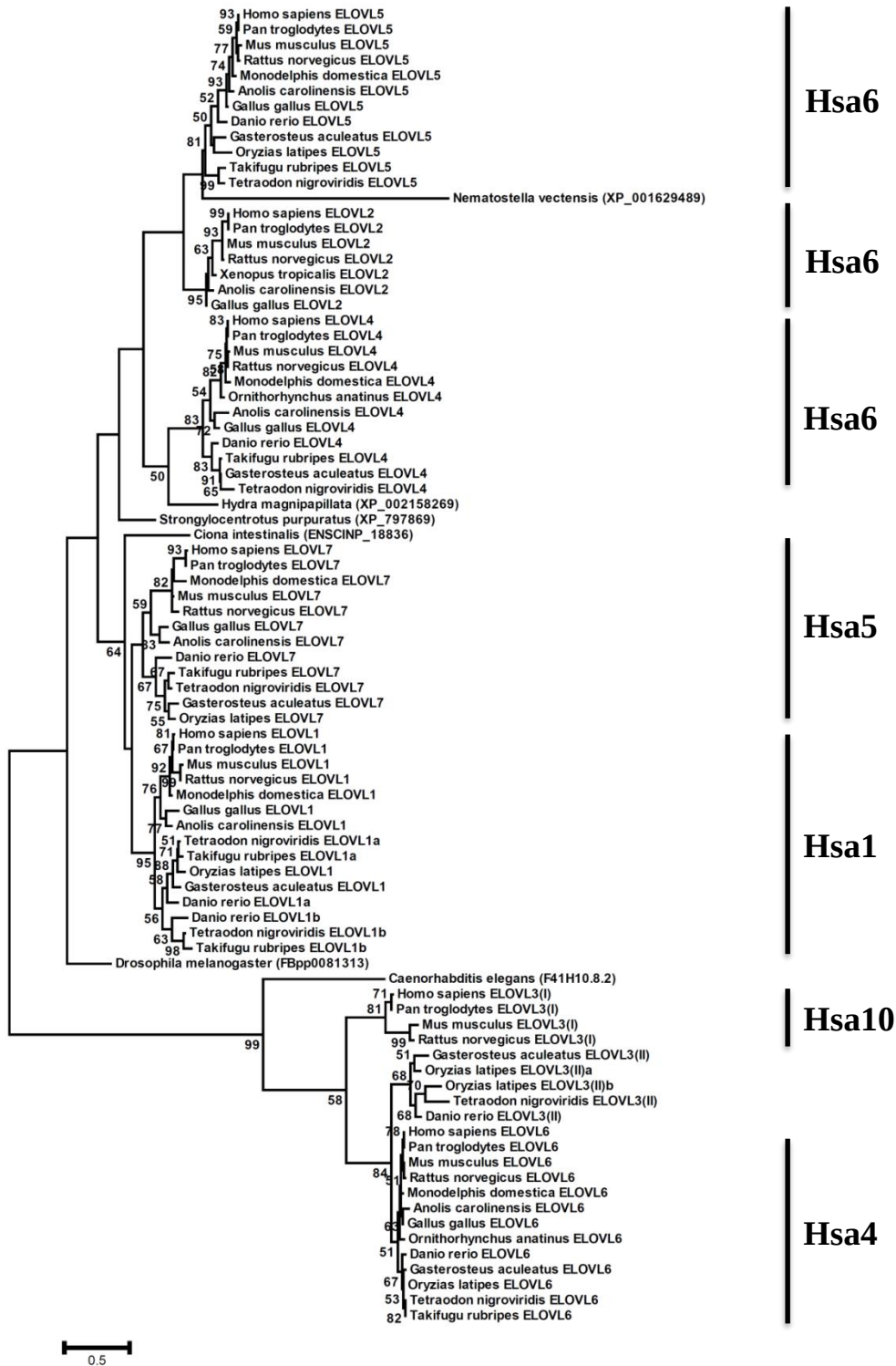
Ankyrin-ANK



Phylogenetic tree of ANK family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-10151.0828) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 3.3216)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 64 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 370 positions in the final dataset.

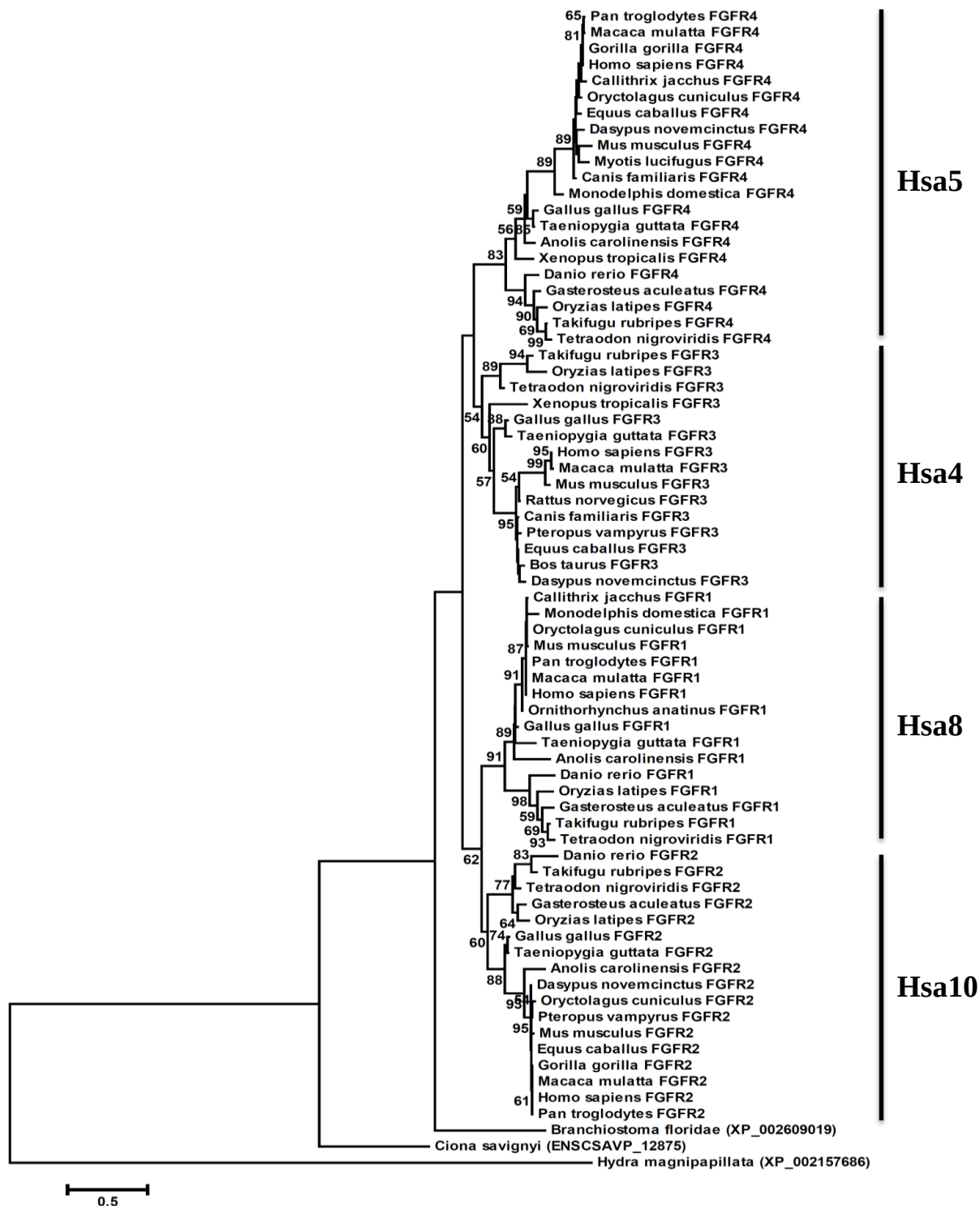
Elongation of very long chain fatty acids protein-ELOVL



Phylogenetic tree of ELOVL family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-6728.2914) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.9378)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 86 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 135 positions in the final dataset.

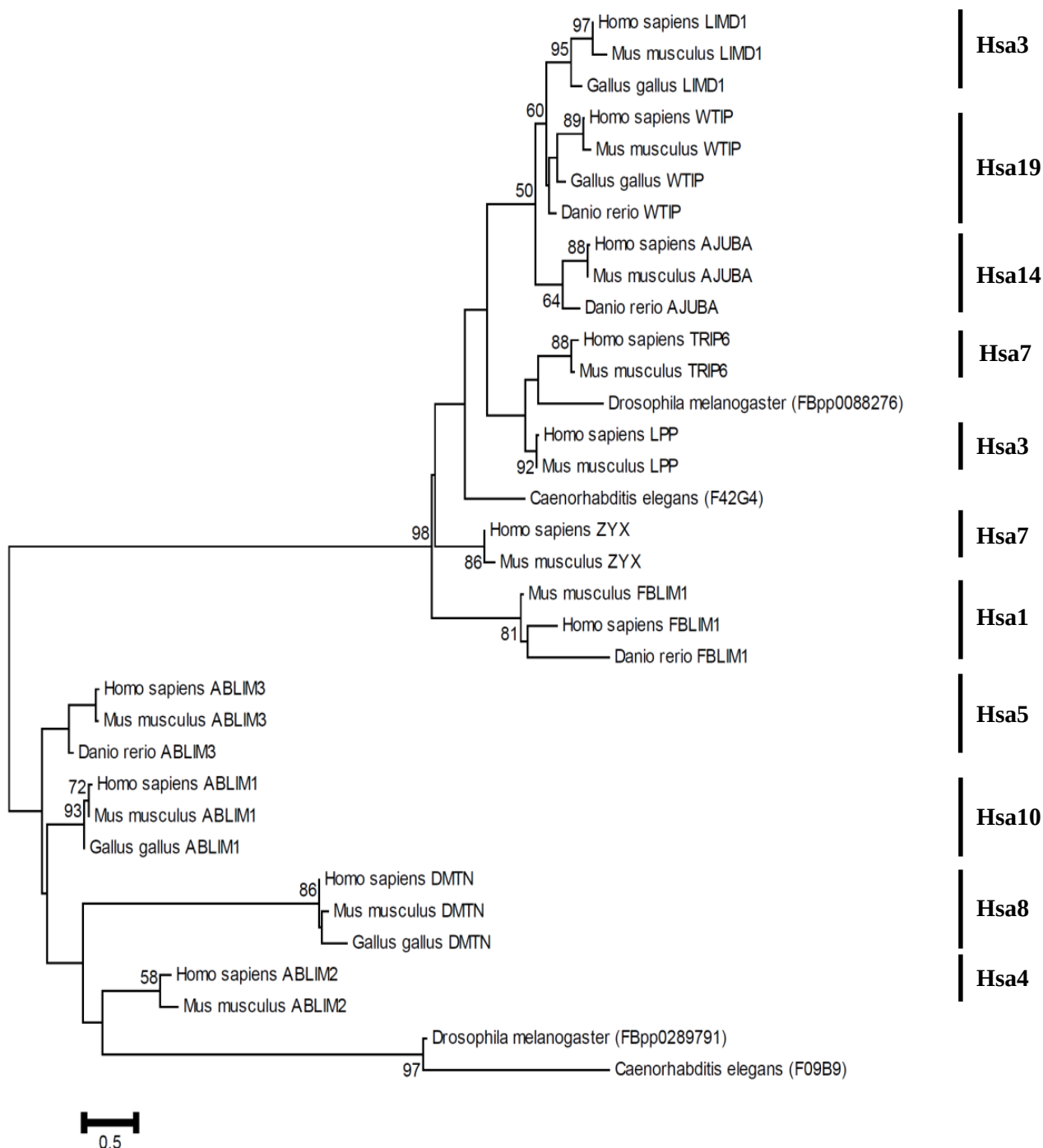
Fibroblast growth factor receptor-FGFR



Phylogenetic tree of FGFR family.

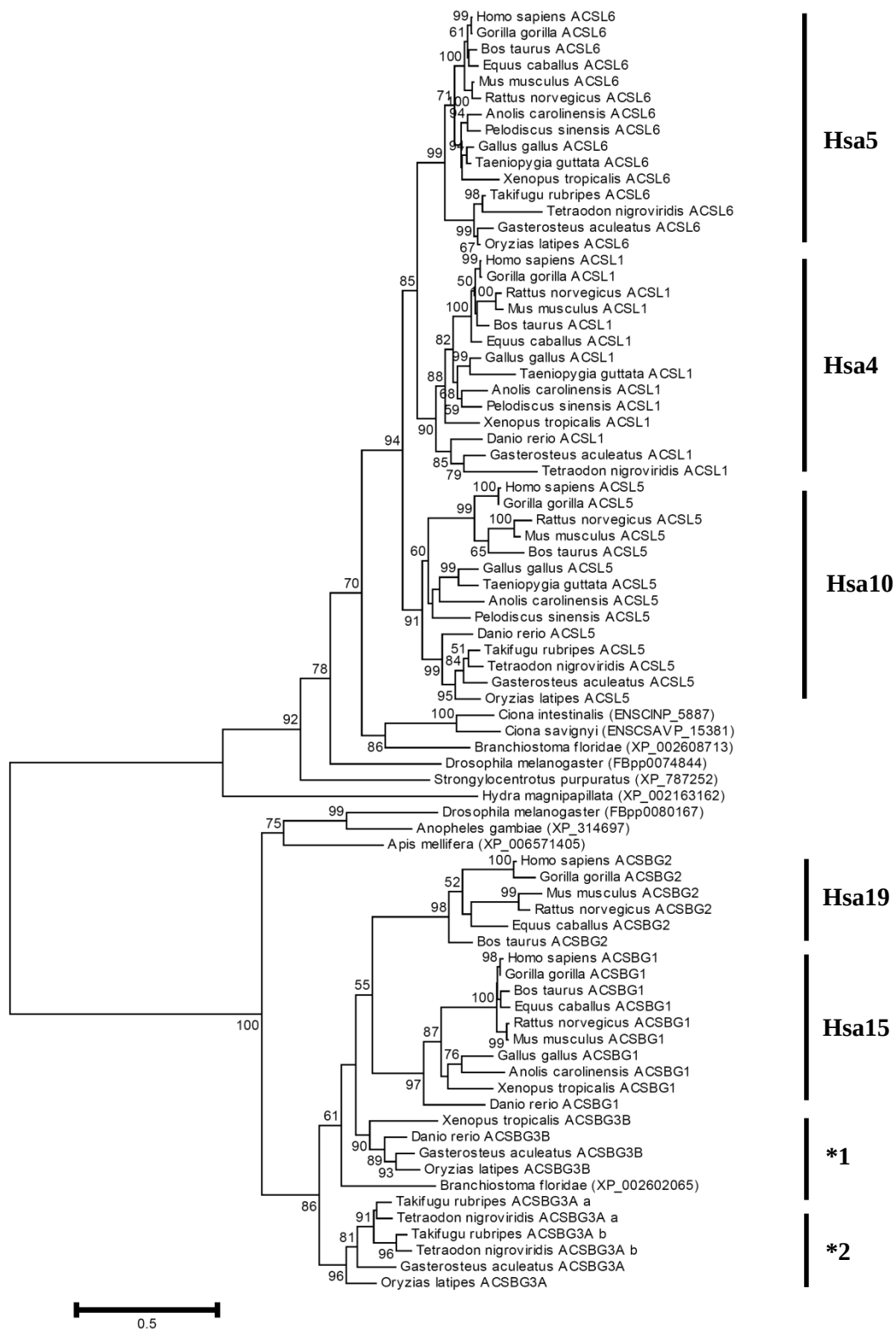
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model [1]. The tree with the highest log likelihood (-8688.1853) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.5569)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 72 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 255 positions in the final dataset.

Actin Bnding LIM protein-ABLIM



Phylogenetic tree of ABLIM family.

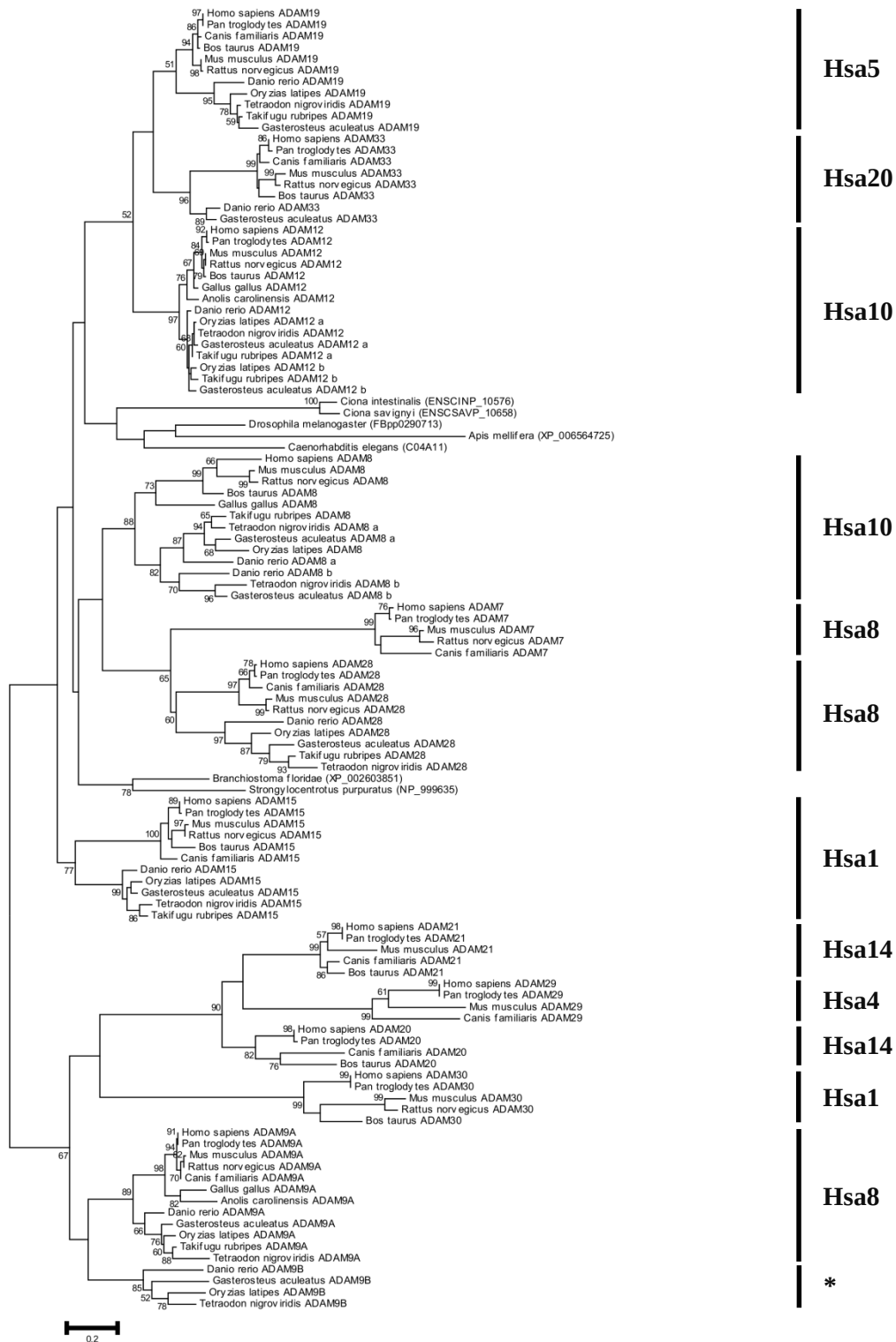
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-3334.3017) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 6.5443)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 34 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 74 positions in the final dataset.



Phylogenetic tree of ACSL family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (- 24273.8074) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.0241). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 79 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 416 positions in the final dataset *1 another paralog which is missing in mammals, *2 another paralog which is missing in tetrapods. These are named as ACSBG3A and ACSBG3B by our group.

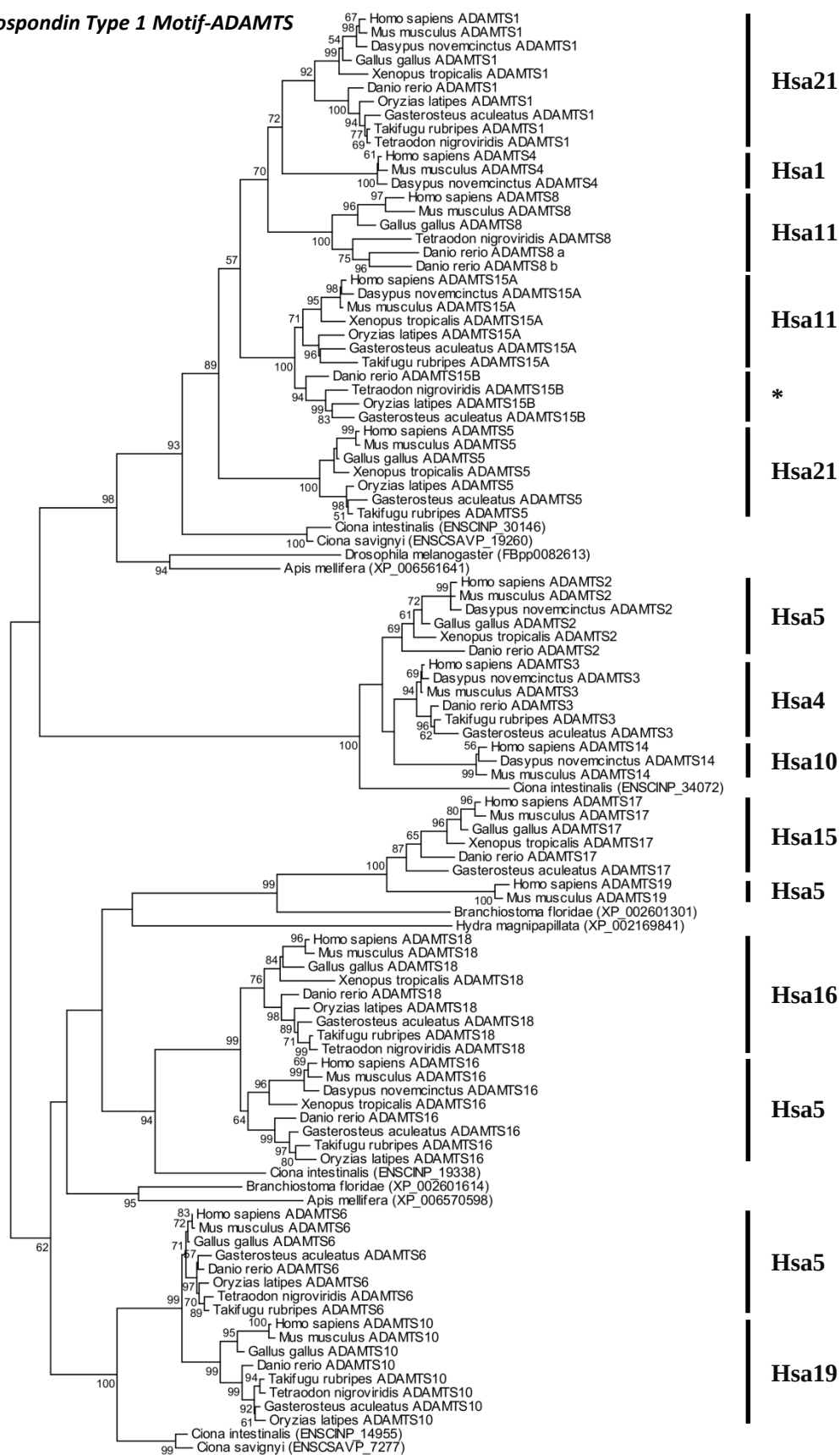
ADAM Metallopeptidase Domain- ADAM



Phylogenetic tree of ADAM family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-11932.4396) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.3087)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 114 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 136 positions in the final dataset. * represents new paralog which is missing in tetrapods.

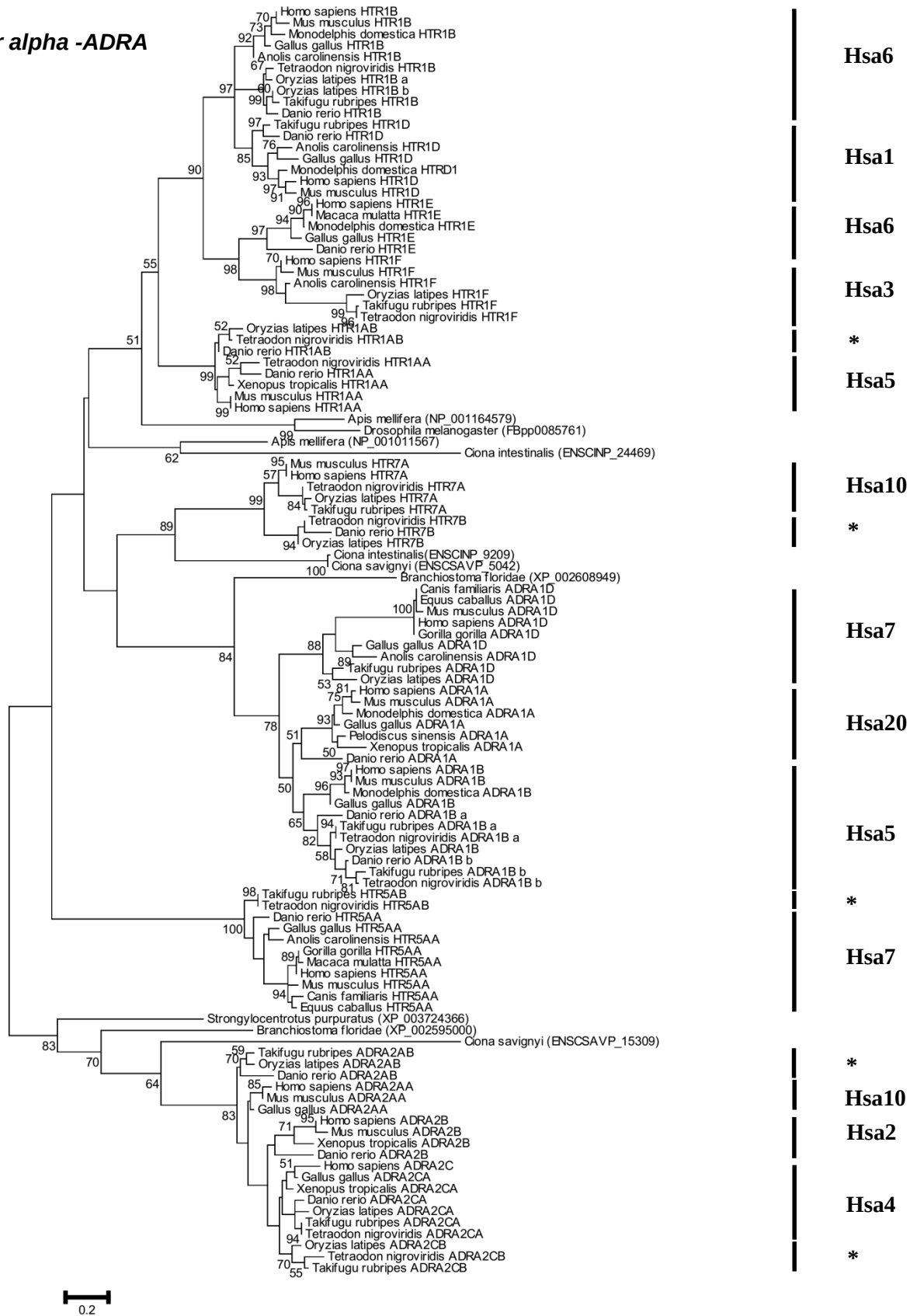
ADAM Metallopeptidase with Thrombospondin Type 1 Motif-ADAMTS



Phylogenetic tree of ADAMTS family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model. The tree with the highest log likelihood (-19039.7449) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.8369)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 105 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 264 positions in the final dataset. * shows another paralog which is missing in tetrapods.

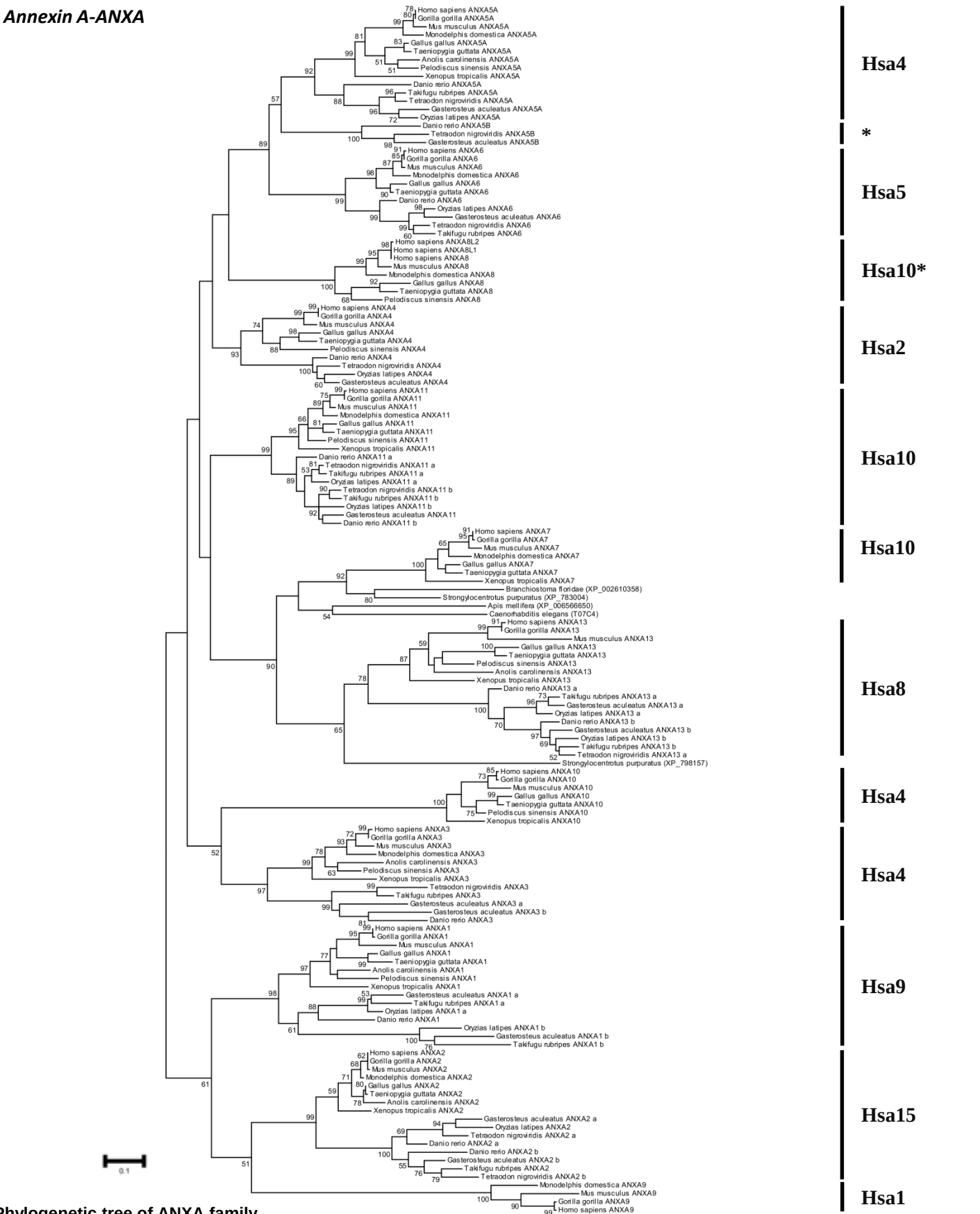
Adrenoceptor alpha -ADRA



Phylogenetic tree of ADRA family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-7843.8634) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 1.1388)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 112 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 136 positions in the final dataset. * shows another paralog which is missing in tetrapods.

Annexin A-ANXA



Phylogenetic tree of ANXA family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the WAG. The tree with the highest log likelihood (-18885.4229) is shown. The percentage in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.0321)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Analysis involved 58 amino acid sequences. The analysis involved 146 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 189 positions in the final dataset. 10* shows one human specific duplication on chromosome 10. * shows another paralog which is missing in tetrapods.

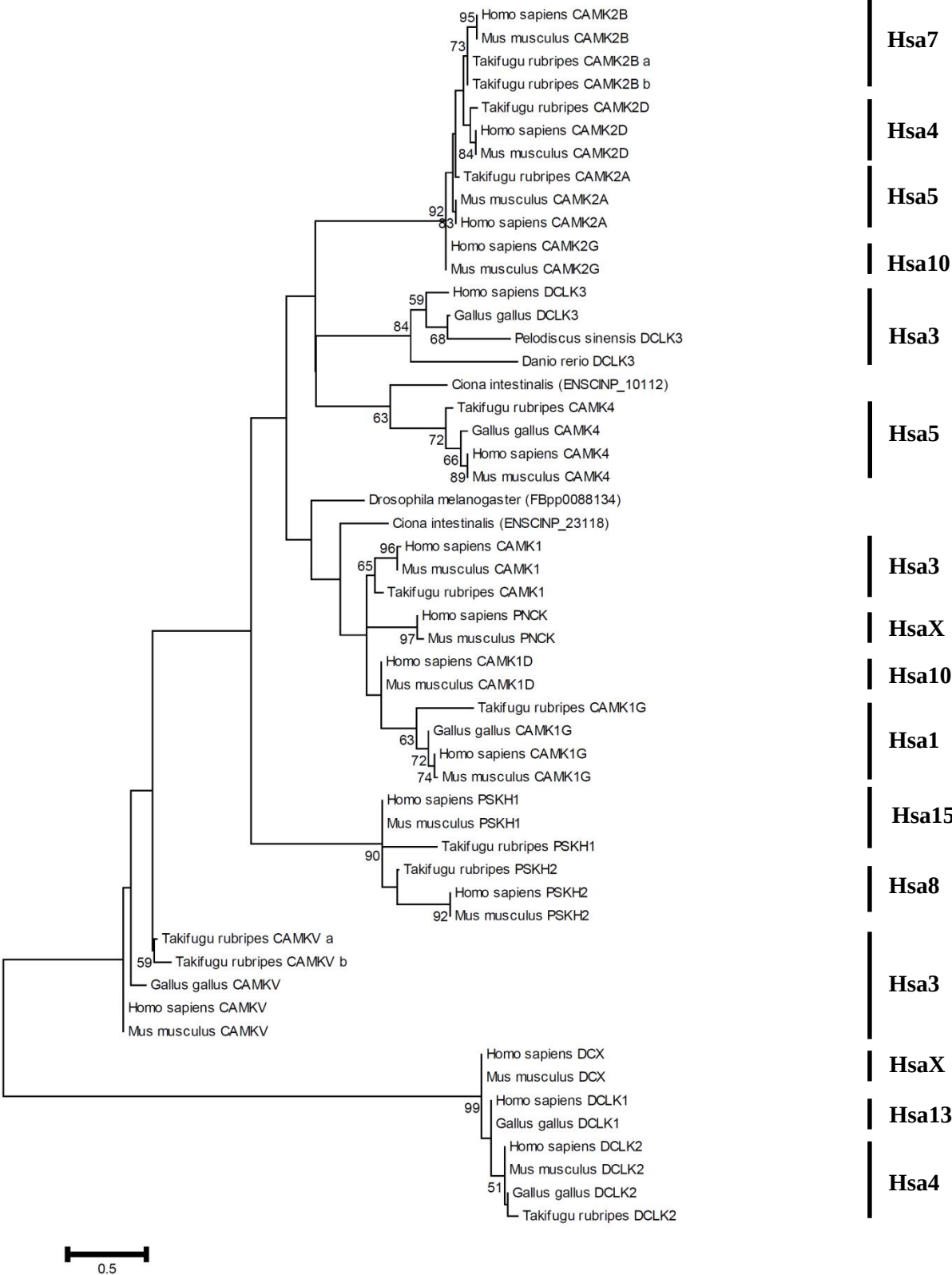
ArfGAP(Ankyrin Repeat and PH Domain)-ARAP



Phylogenetic tree of ARAP family.

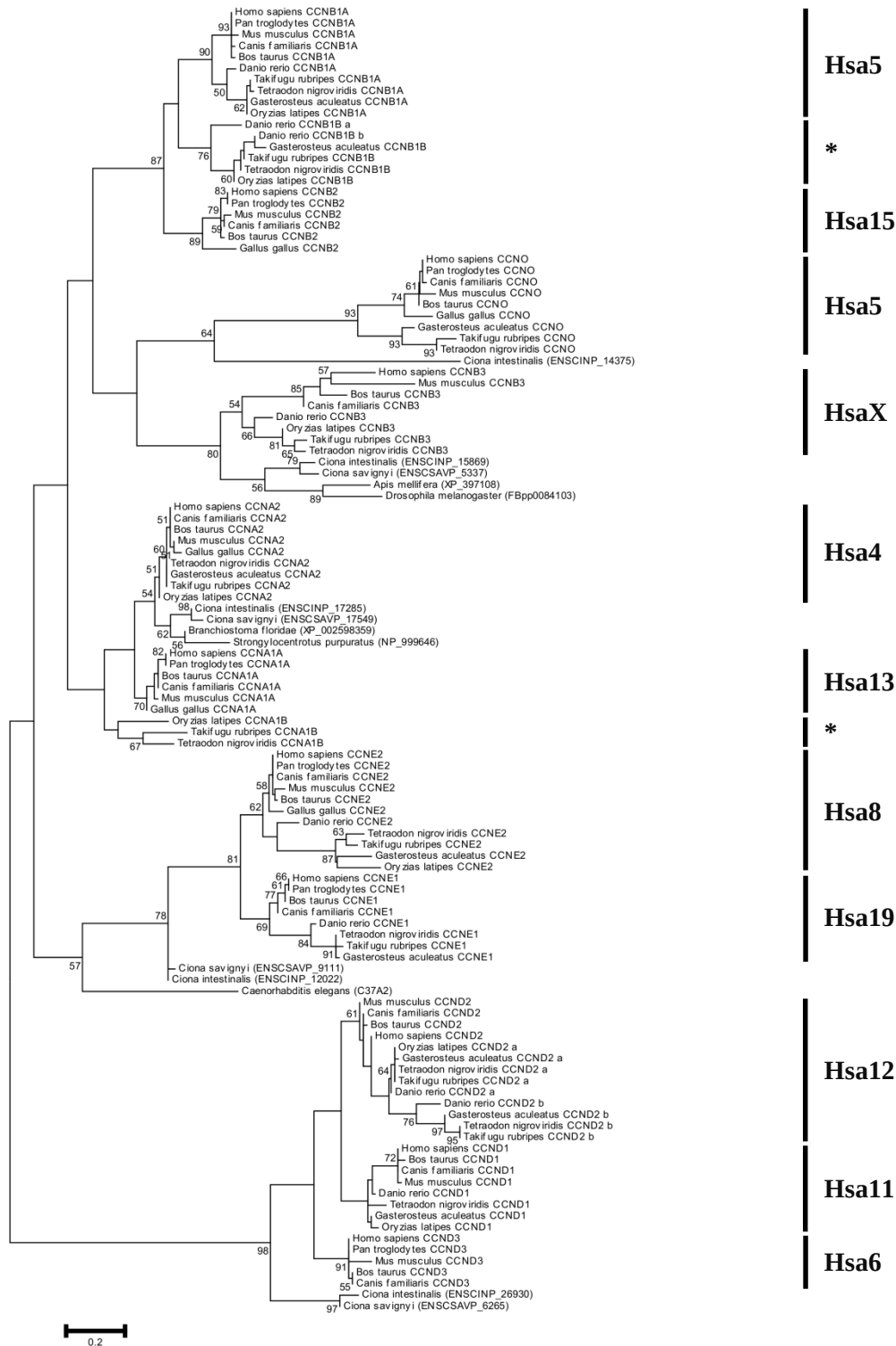
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-6494.4580) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.4507)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 116 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 101 positions in the final dataset. *10 shows human specific duplication on chromosome 10. * shows another paralogue with deleted mammals.

Calcium/calmodulin-dependent protein kinase -CAMK



Phylogenetic tree CAMK family.

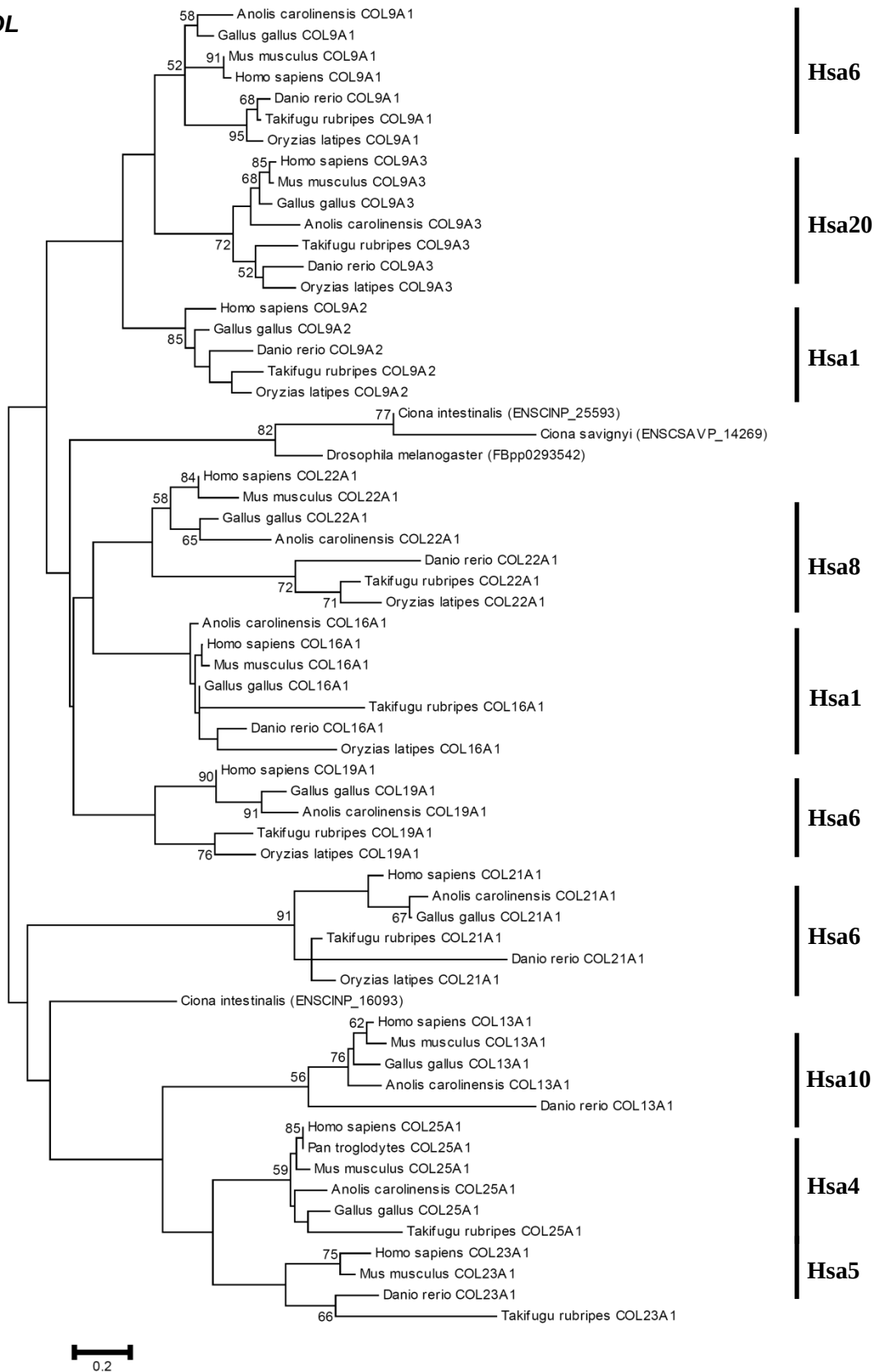
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-2131.3412)is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 3.7351). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 53 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 54 positions in the final dataset.



Phylogenetic tree of CNN family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-4294.8347) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.0254)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 116 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 75 positions in the final dataset. * is another paralog which is missing in tetrapods.

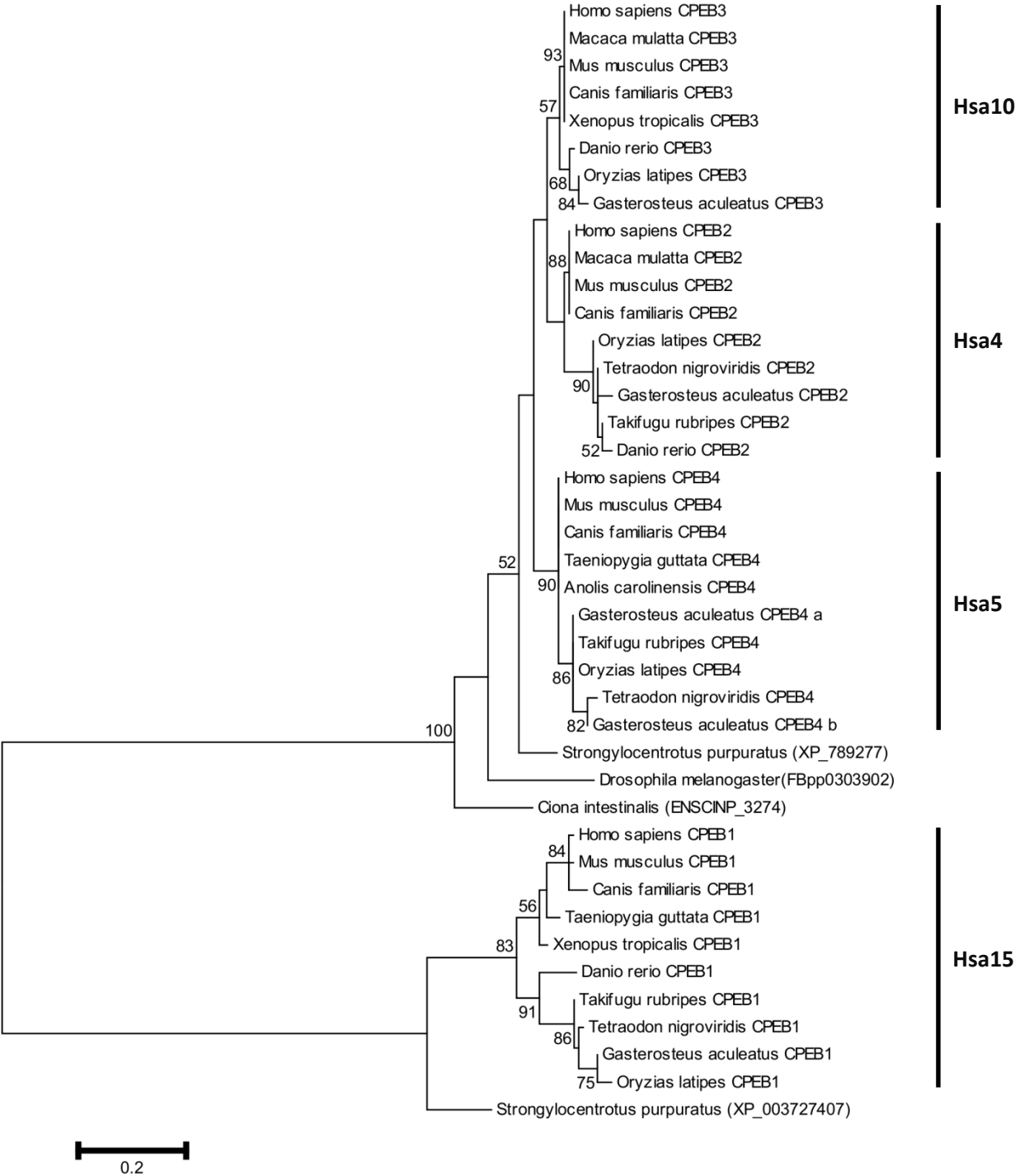
Collagen alpha-COL



Phylogenetic tree of COL family

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-2541.5152) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.1432)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 63 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 38 positions in the final dataset.

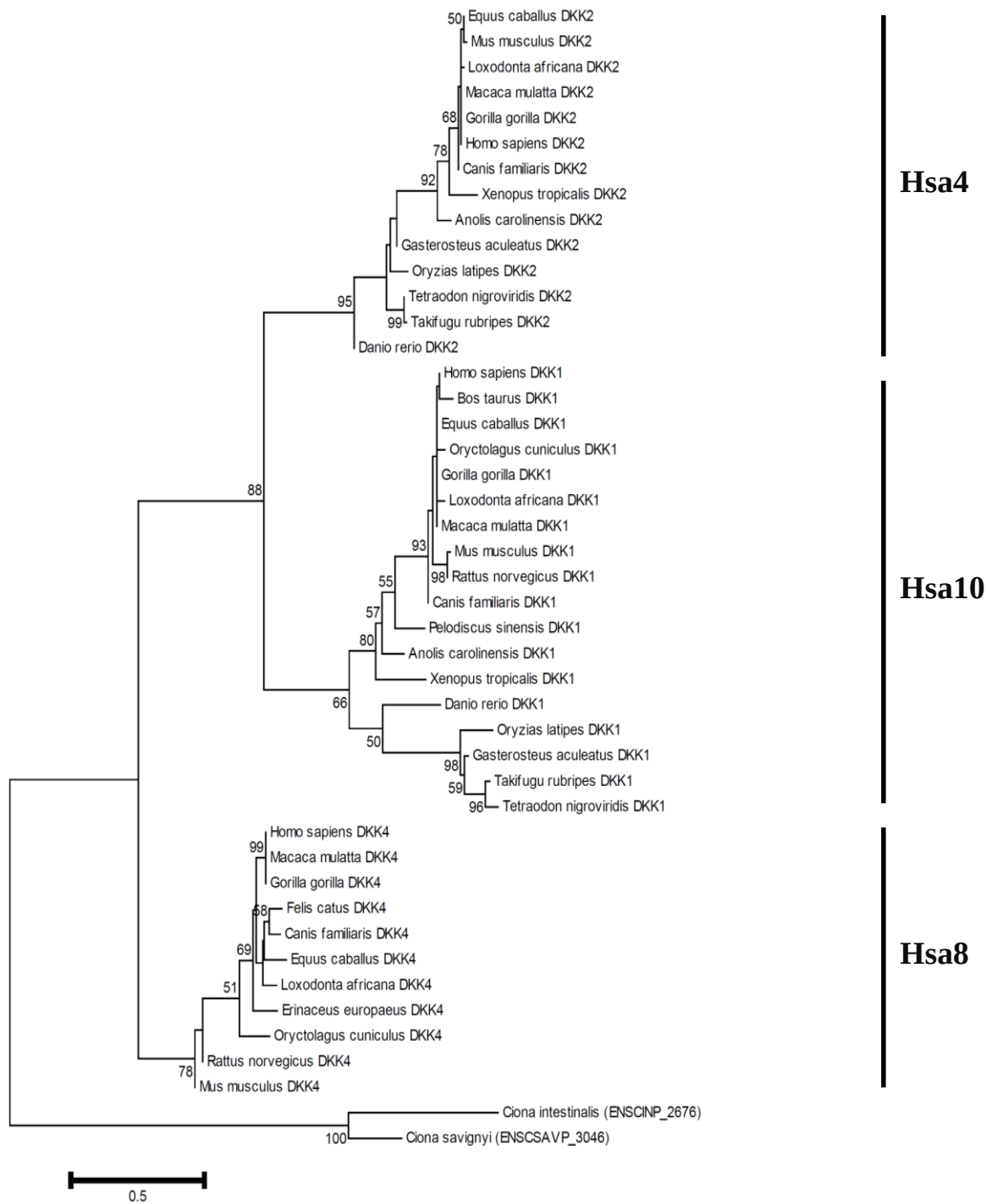
Cytoplasmic Polyadenylation Element Binding
Protein-CPEB



Phylogenetic tree of CEBP family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-1594.6120) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1473)). The analysis involved 41 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 126 positions in the final dataset.

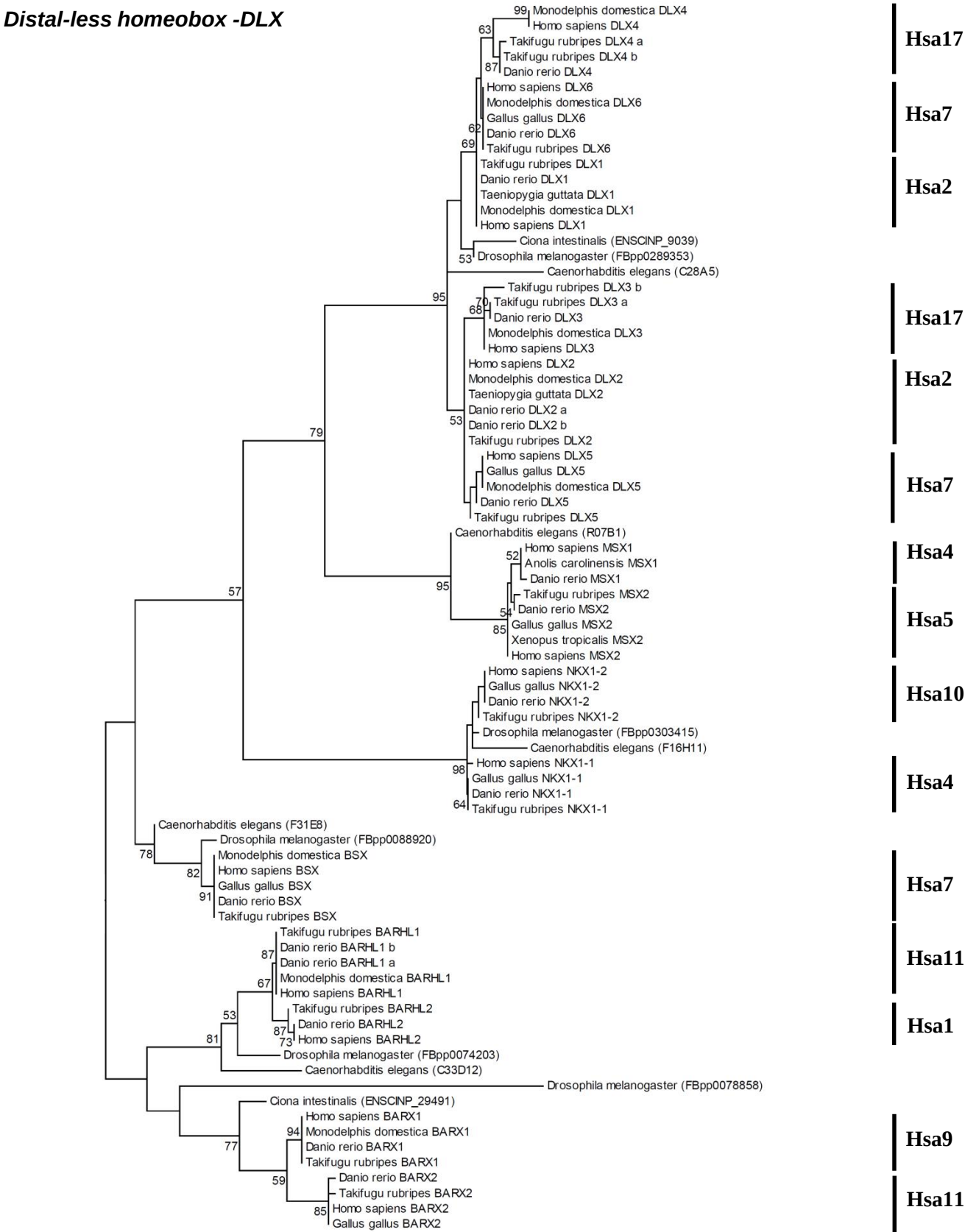
Dickkopf WNT Signaling Pathway Inhibitor -DKK



Phylogenetic tree of DKK family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-2419.3572) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1862)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 45 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 94 positions in the final dataset.

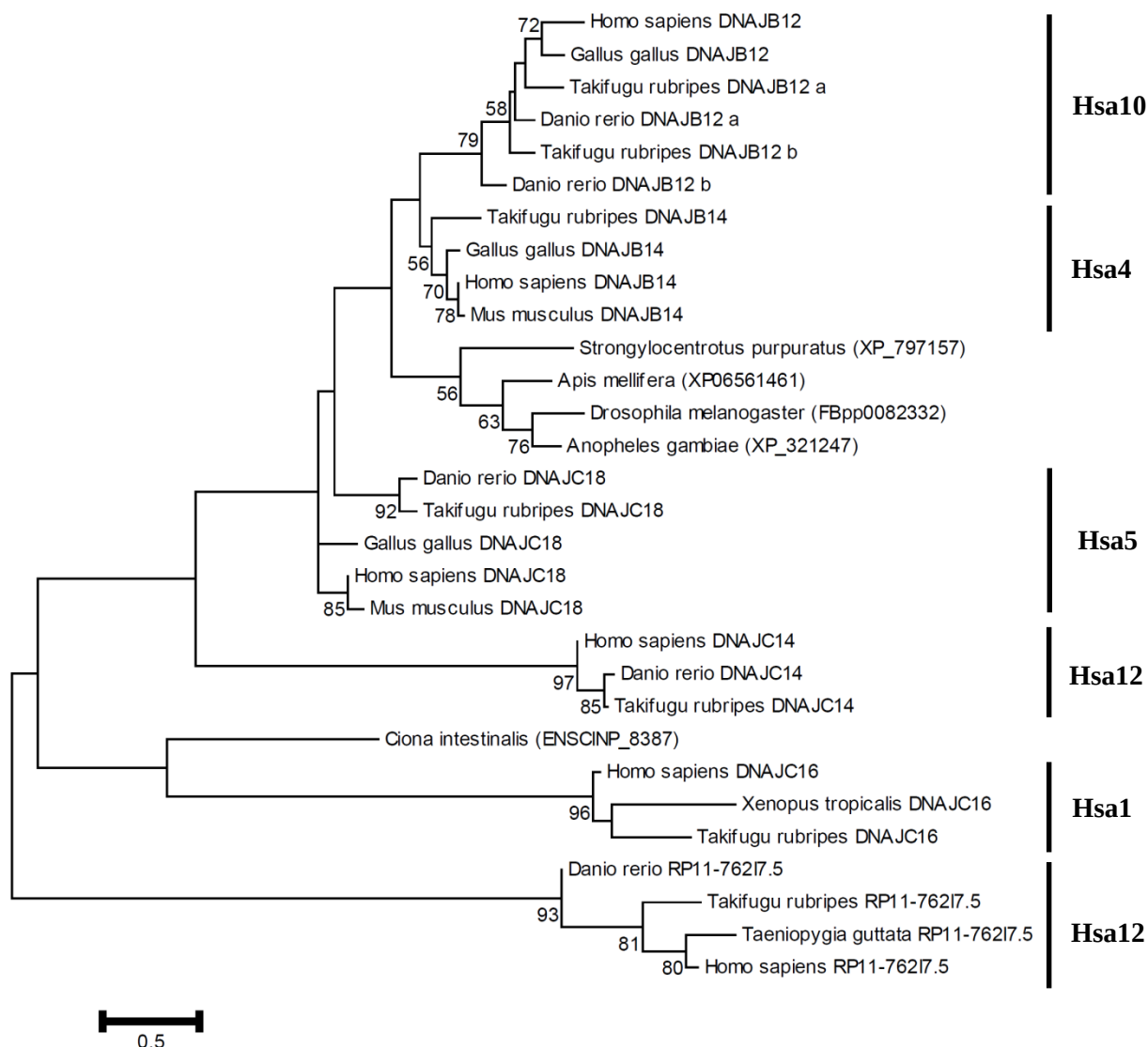
Distal-less homeobox -DLX



Phylogenetic tree DLX family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-1224.0126) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 0.9081). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 80 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 54 positions in the final dataset.

DnaJ (Hsp40) homolog -DNAJ



Phylogenetic tree DNAJ family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-2848.3983) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 6.9606)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 30 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 65 positions in the final dataset.

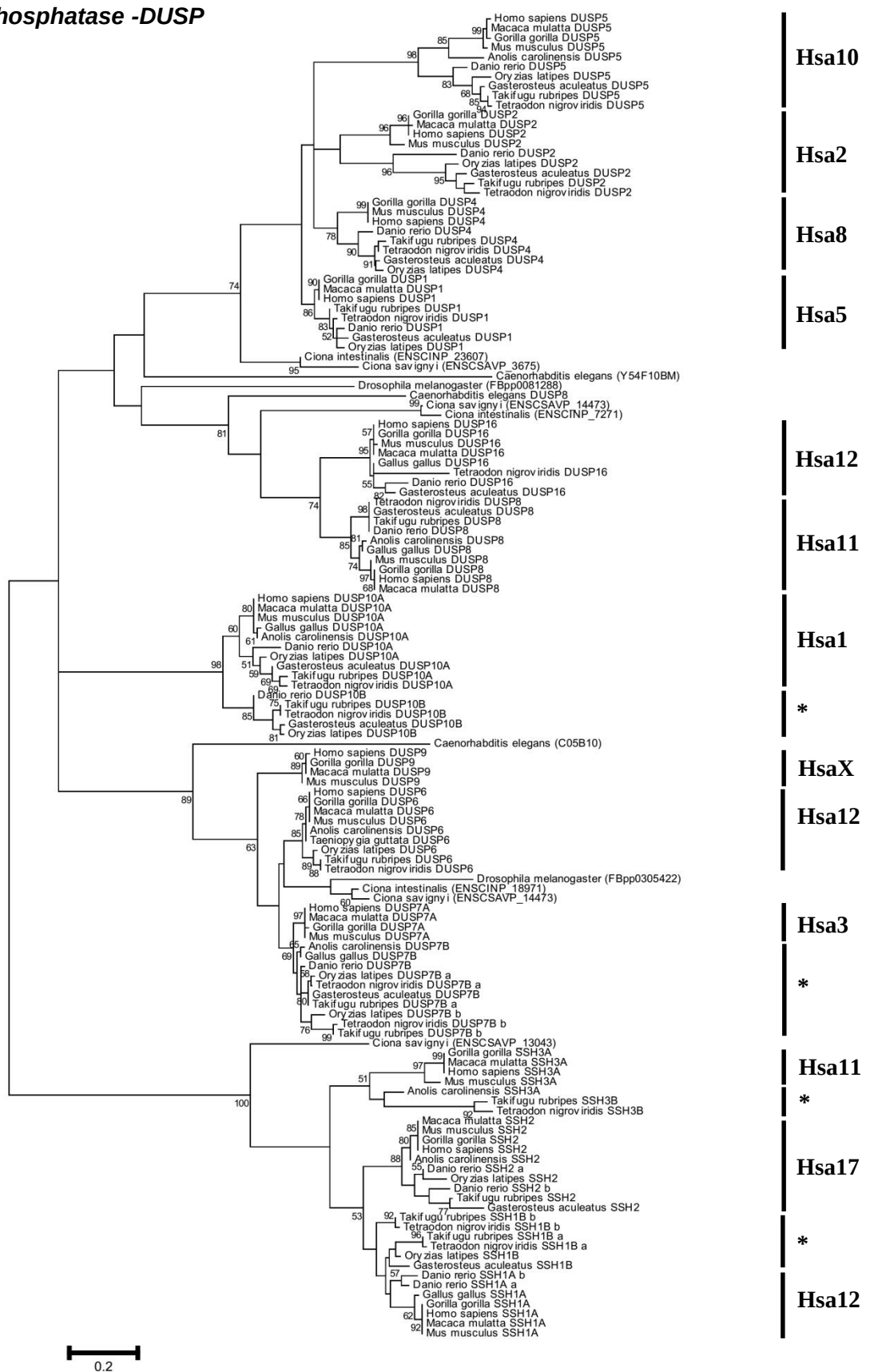
Dedicator of cytokinesis -DOCK



Phylogenetic tree of DOCK family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood ((-18205.3205) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 3.1766)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 111 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 342 positions in the final dataset.

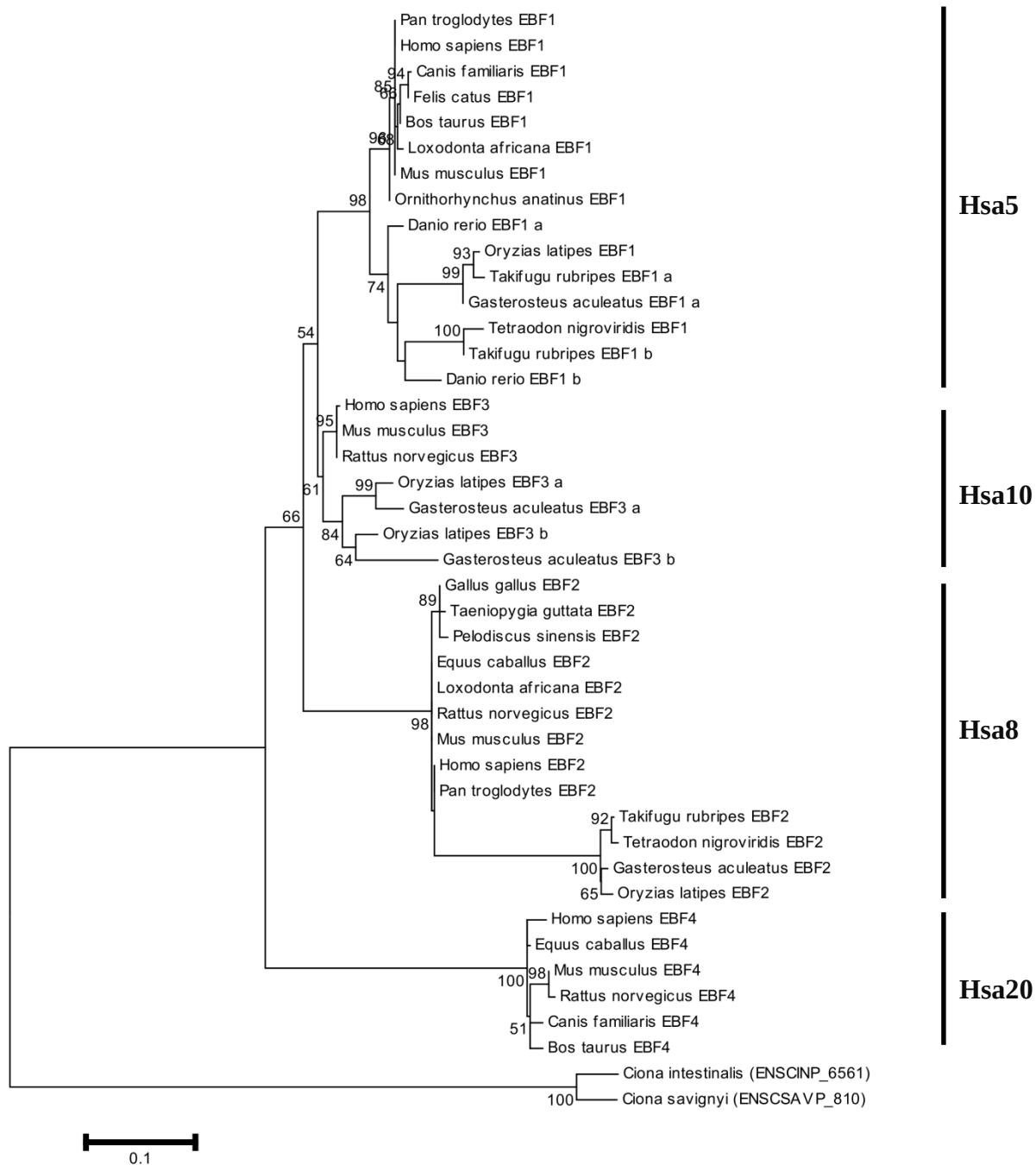
Dual specificity phosphatase -DUSP



Phylogenetic tree DUSP family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-5365.7665) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 1.9105)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 137 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 90 positions in the final dataset. * shows another paralog which is missing in tetrapods

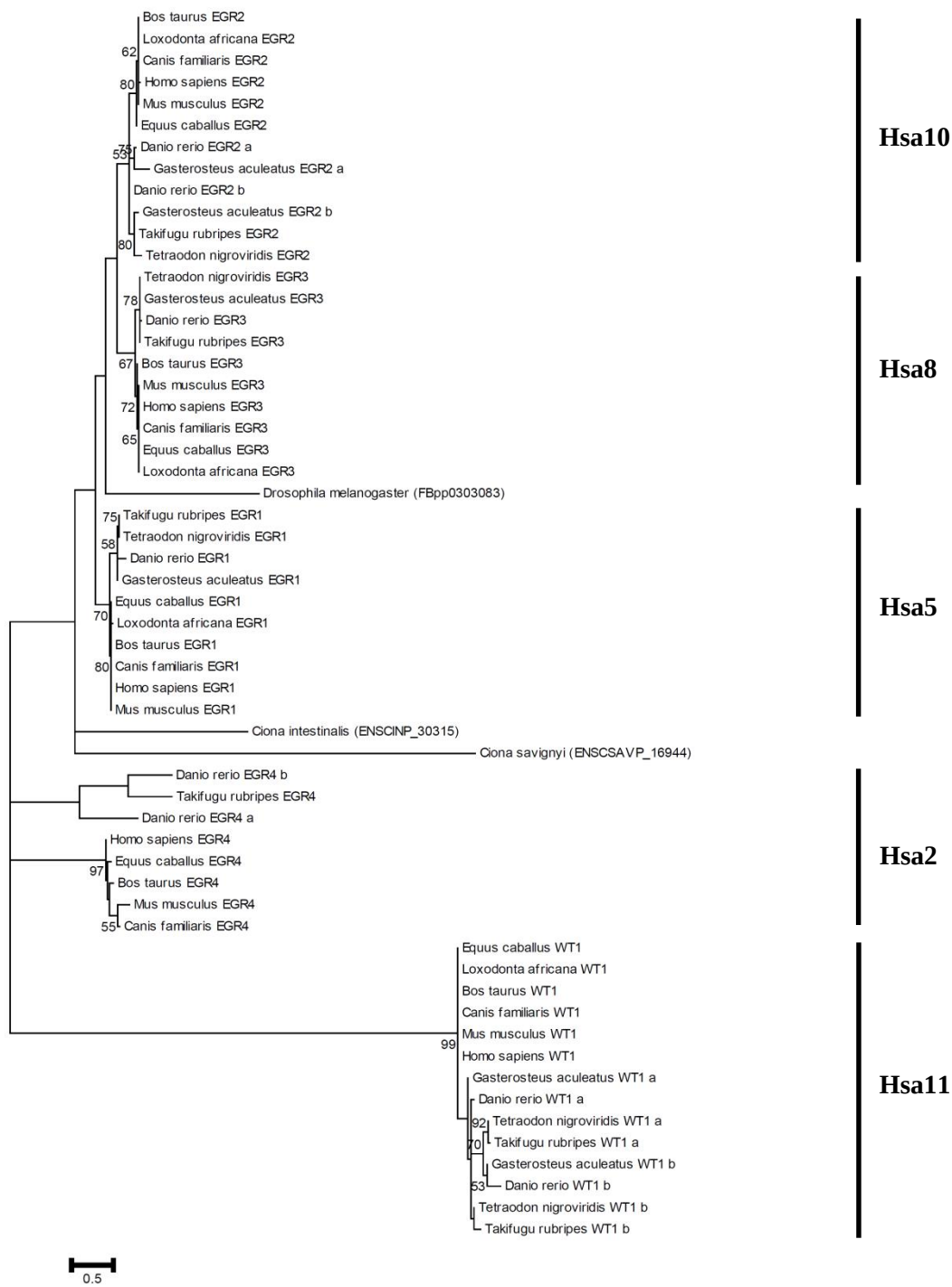
Early B-cell Factor -EBF



Phylogenetic tree of EBF family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-4301.3288) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.6392)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 34 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 406 positions in the final dataset.

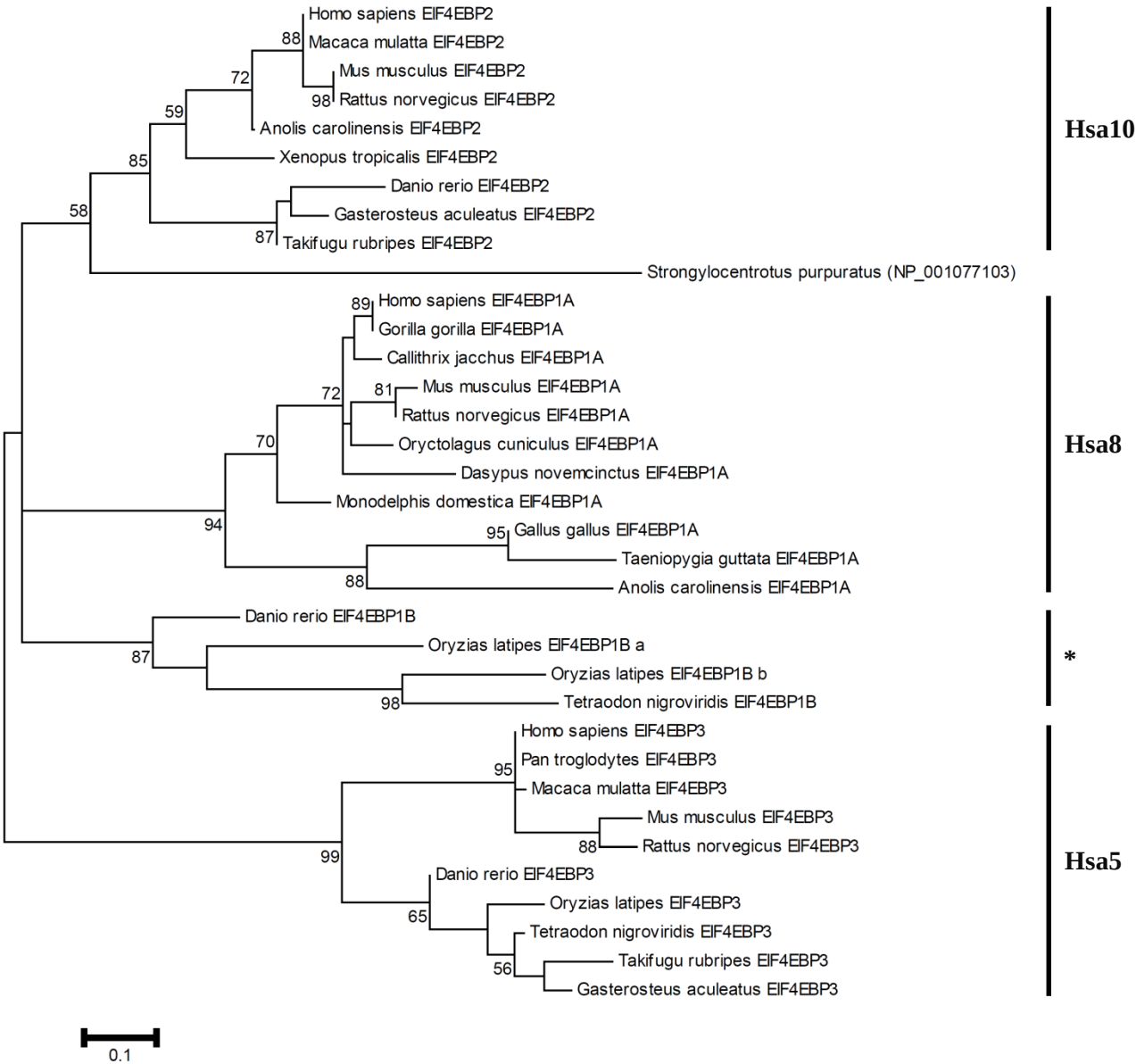
Early Growth Response-EGR



Phylogenetic tree of EGR family

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-1744.4131) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter =2.6053)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 58 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 54 positions in the final dataset.

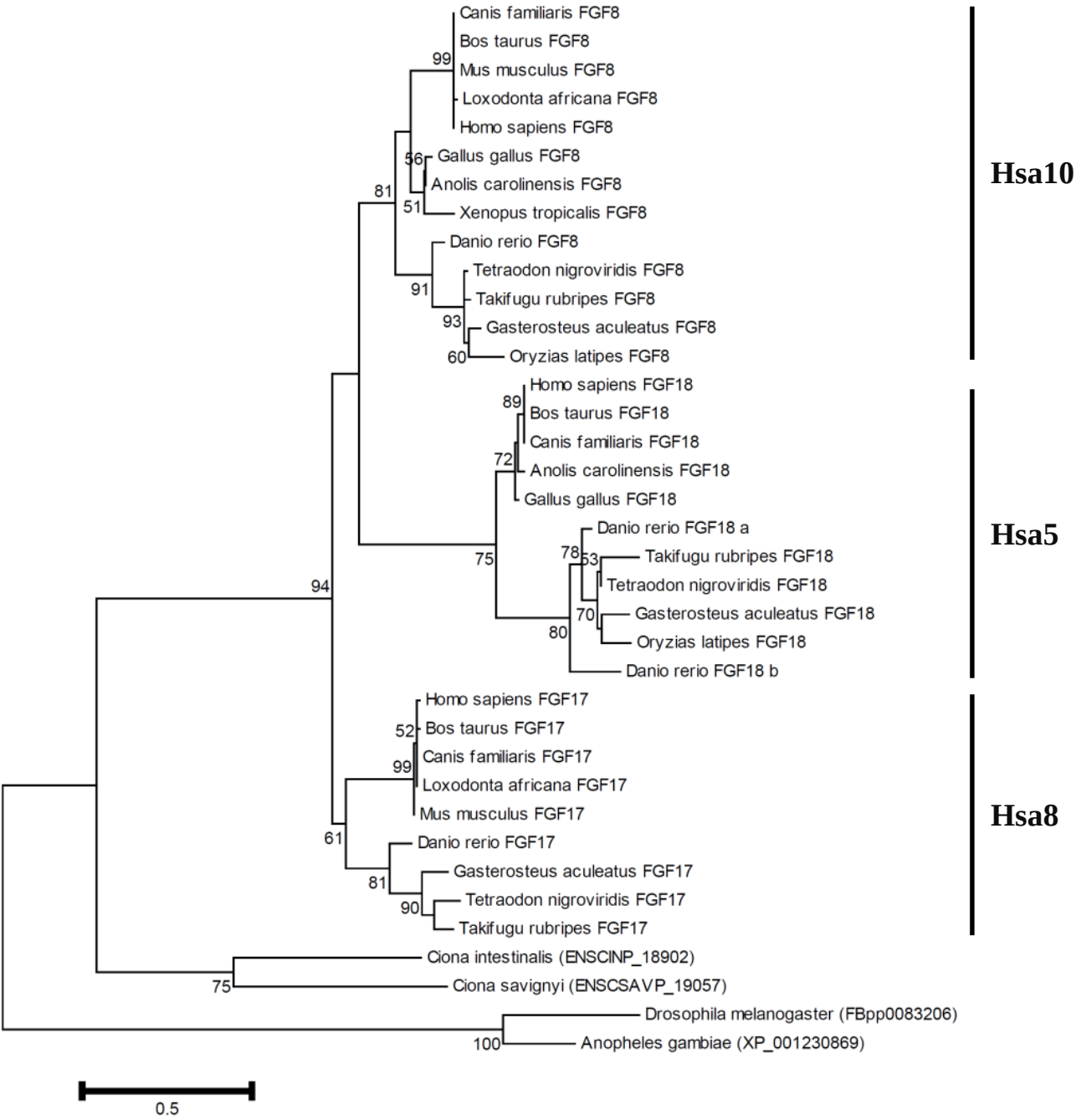
Eukaryotic Translation Initiation Factor 4E Binding Protein-EIF4EBP



Phylogenetic tree of EIF4EBP family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-1807.8739) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.6808)). The analysis involved 35 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 87 positions in the final dataset. * shows another paralog which is missing in tetrapods.

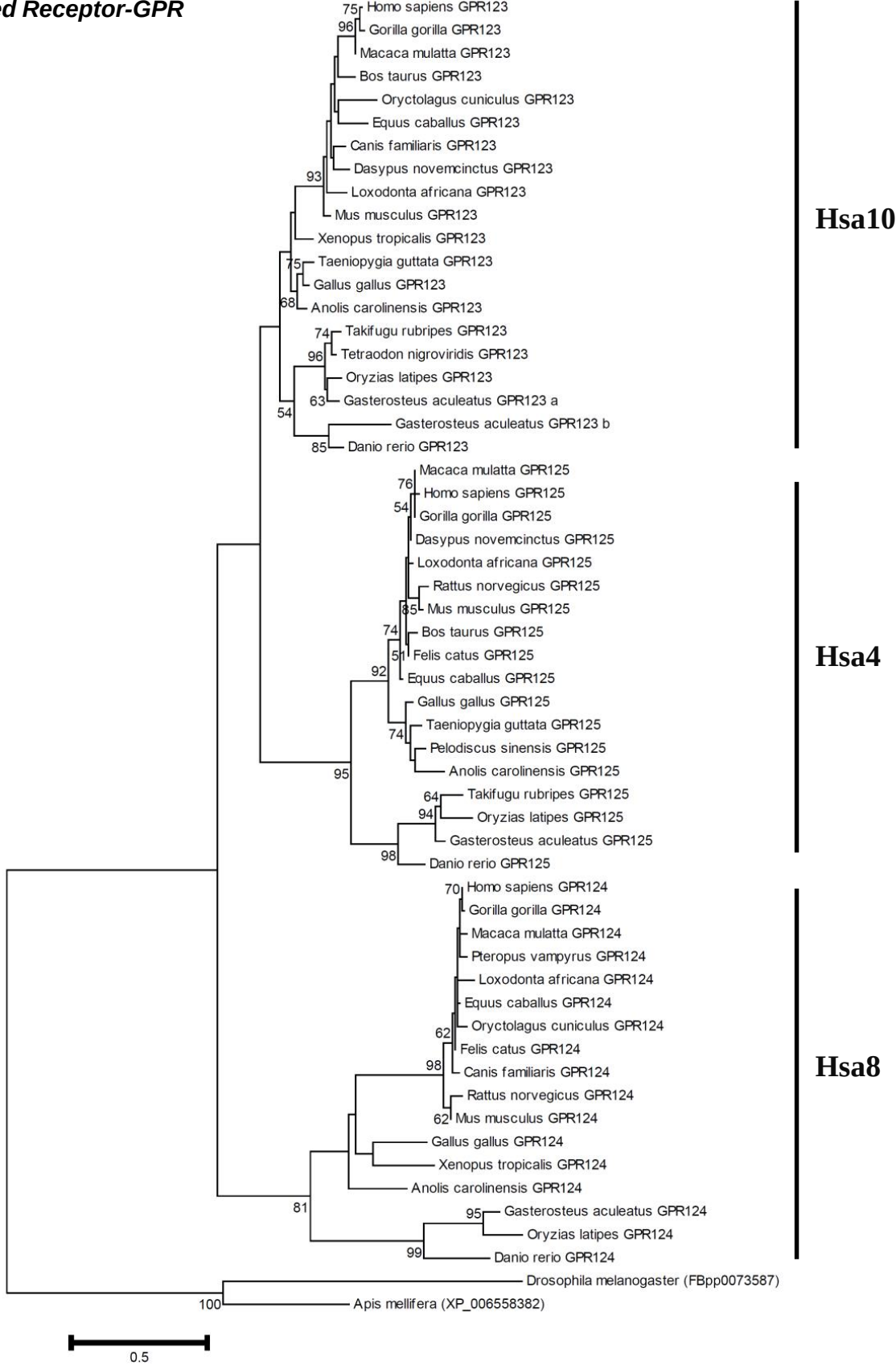
Fibroblast Growth Factor-FGF



Phylogenetic tree of FGF family

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood(-2888.5497) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.9702)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 37 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 117 positions in the final dataset.

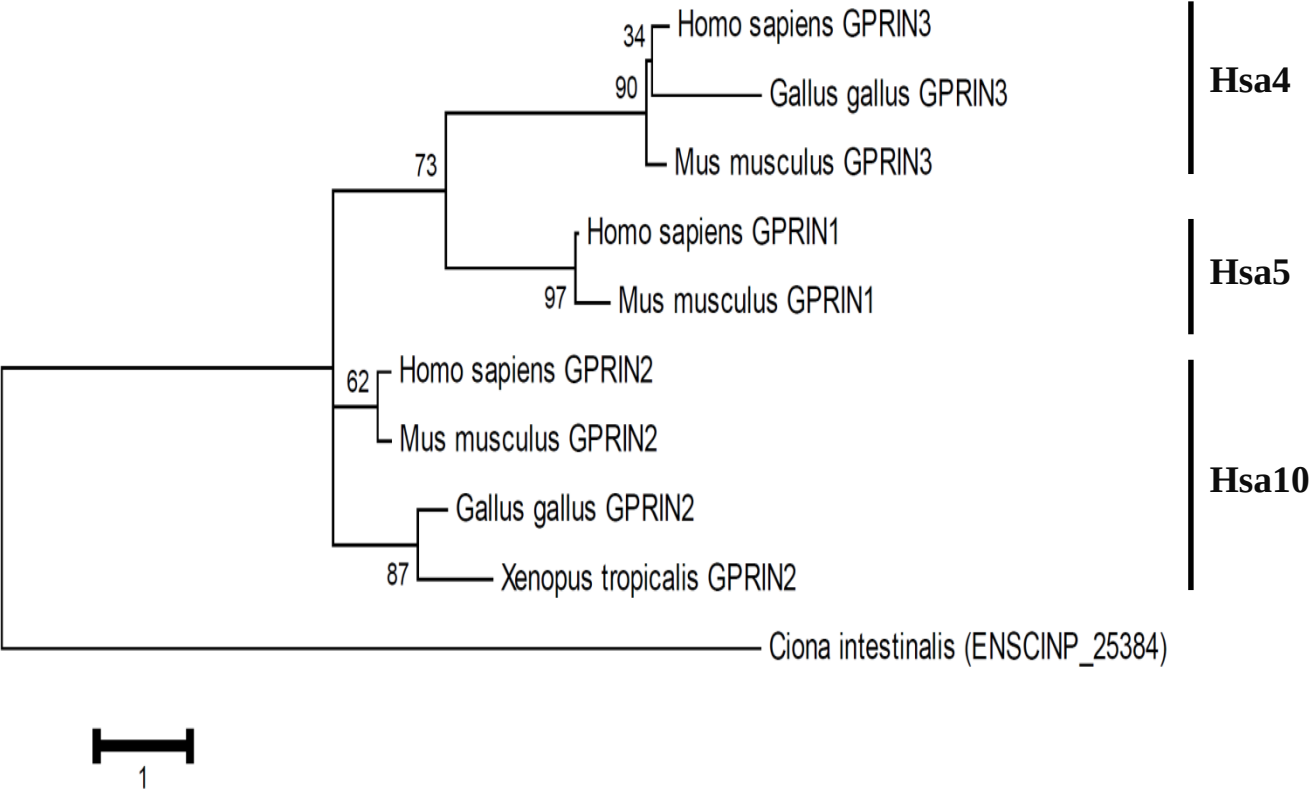
G Protein-Coupled Receptor-GPR



Phylogenetic tree of GPR family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-4572.9392) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.7798)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 57 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 139 positions in the final dataset.

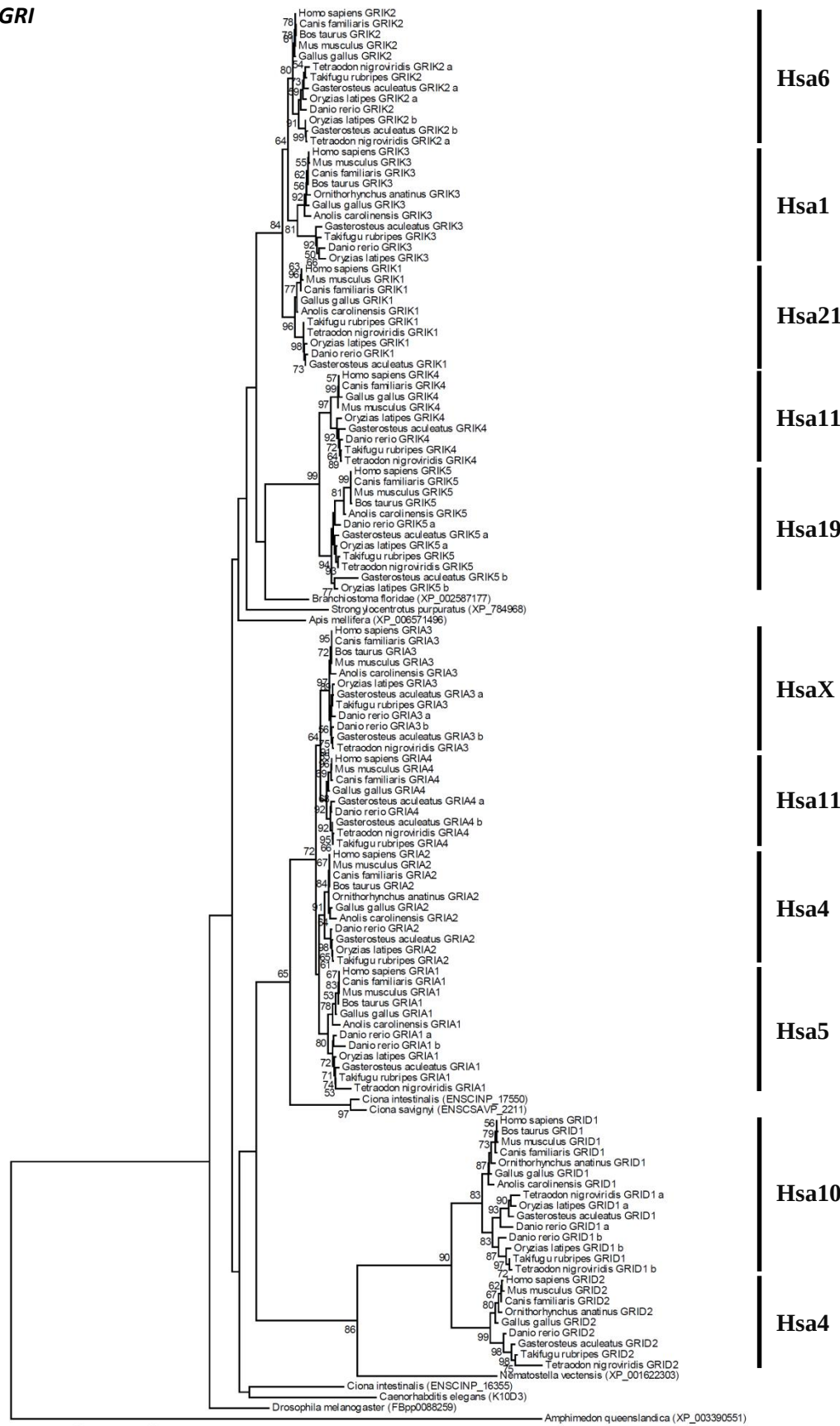
G protein Regulated Inducer of Neurite Outgrowth-GPRIN



Phylogenetic tree of GPRIN family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model(WAG). The tree with the highest log likelihood (-6159.2196) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 5.0172)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 10 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 289 positions in the final dataset.

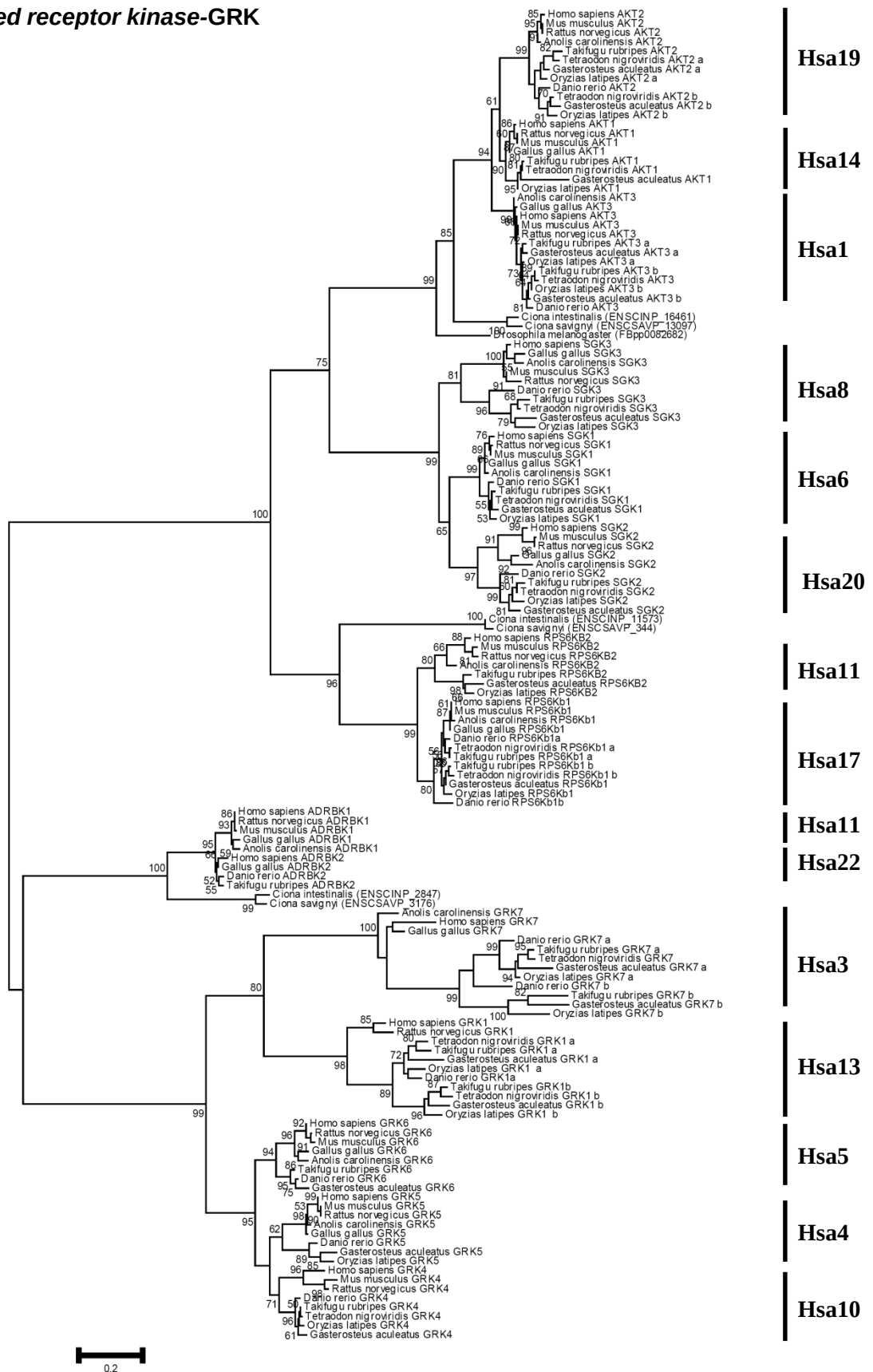
Glutamate Receptor Ionotropic-GRI



Phylogenetic tree of GRI family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-12184.6783) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.6056)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 133 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 258 positions in the final dataset.

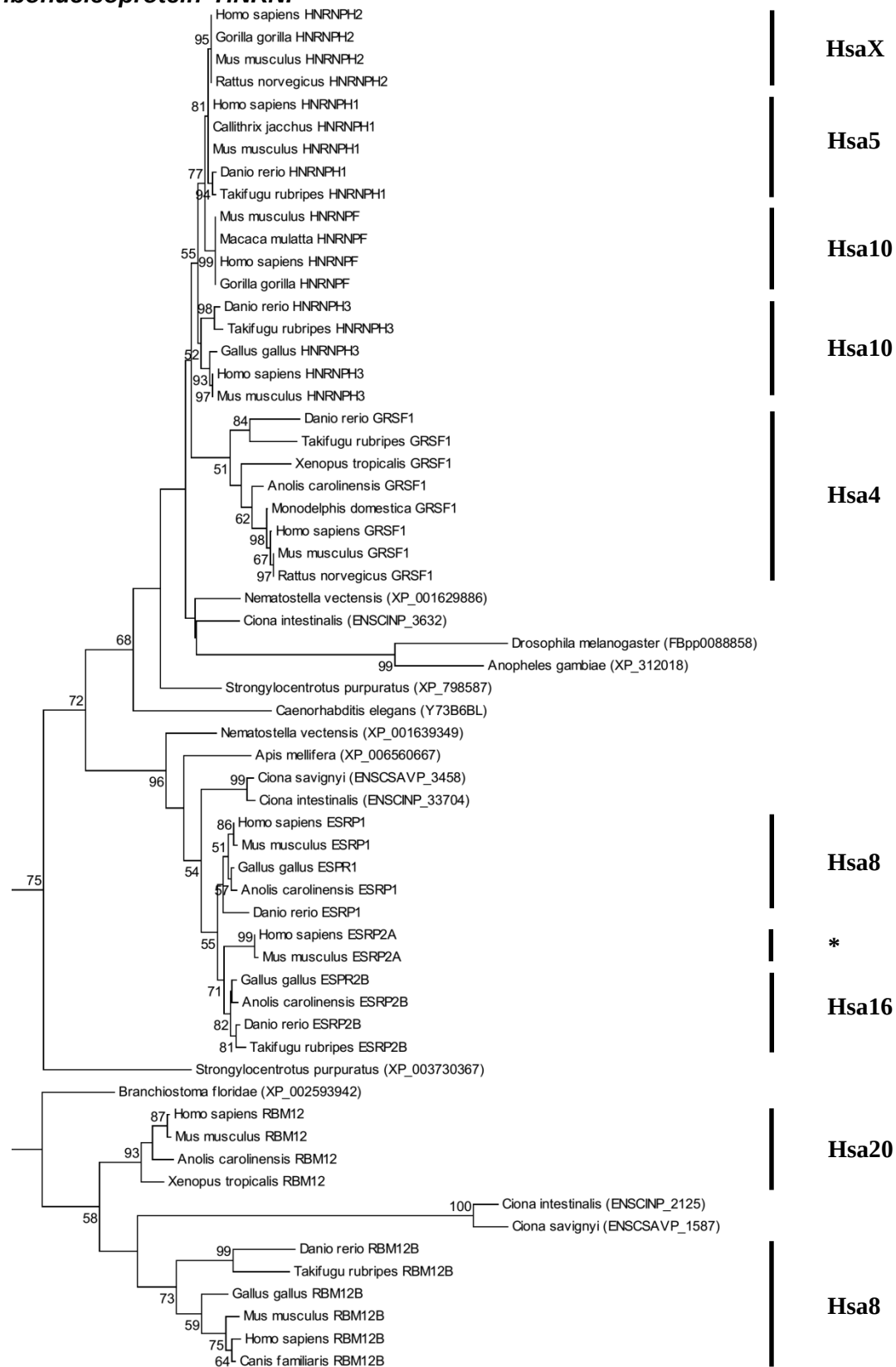
G protein-coupled receptor kinase-GRK



Phylogenetic tree of GRK family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-12617.6823) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1833)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 145 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 239 positions in the final dataset.

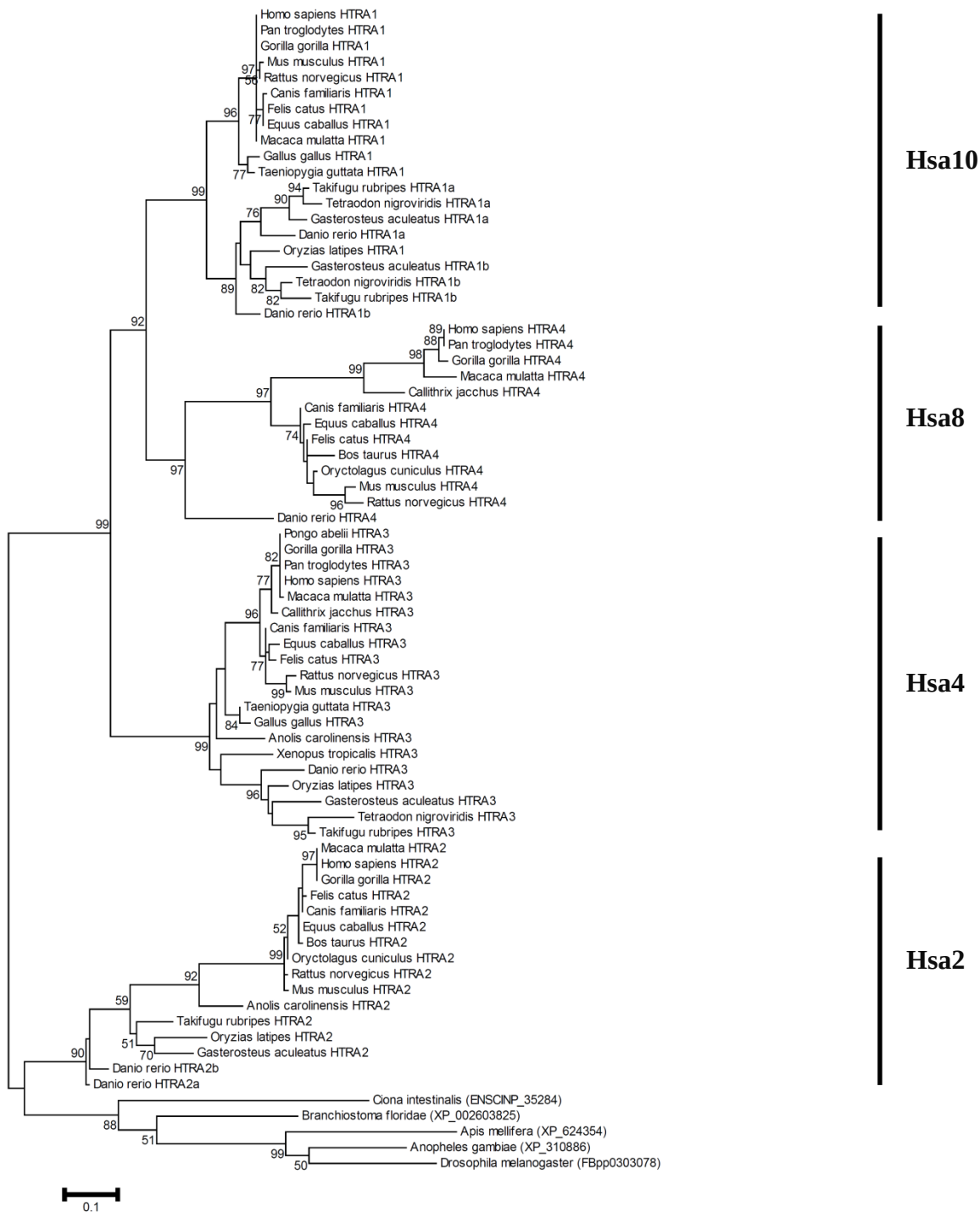
Heterogeneous nuclear ribonucleoprotein- HNRNP



Phylogenetic tree HNRNP family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-8043.6055) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 5.4700). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 61 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 146 positions in the final dataset. * shows another paralog which is missing in tetrapods.

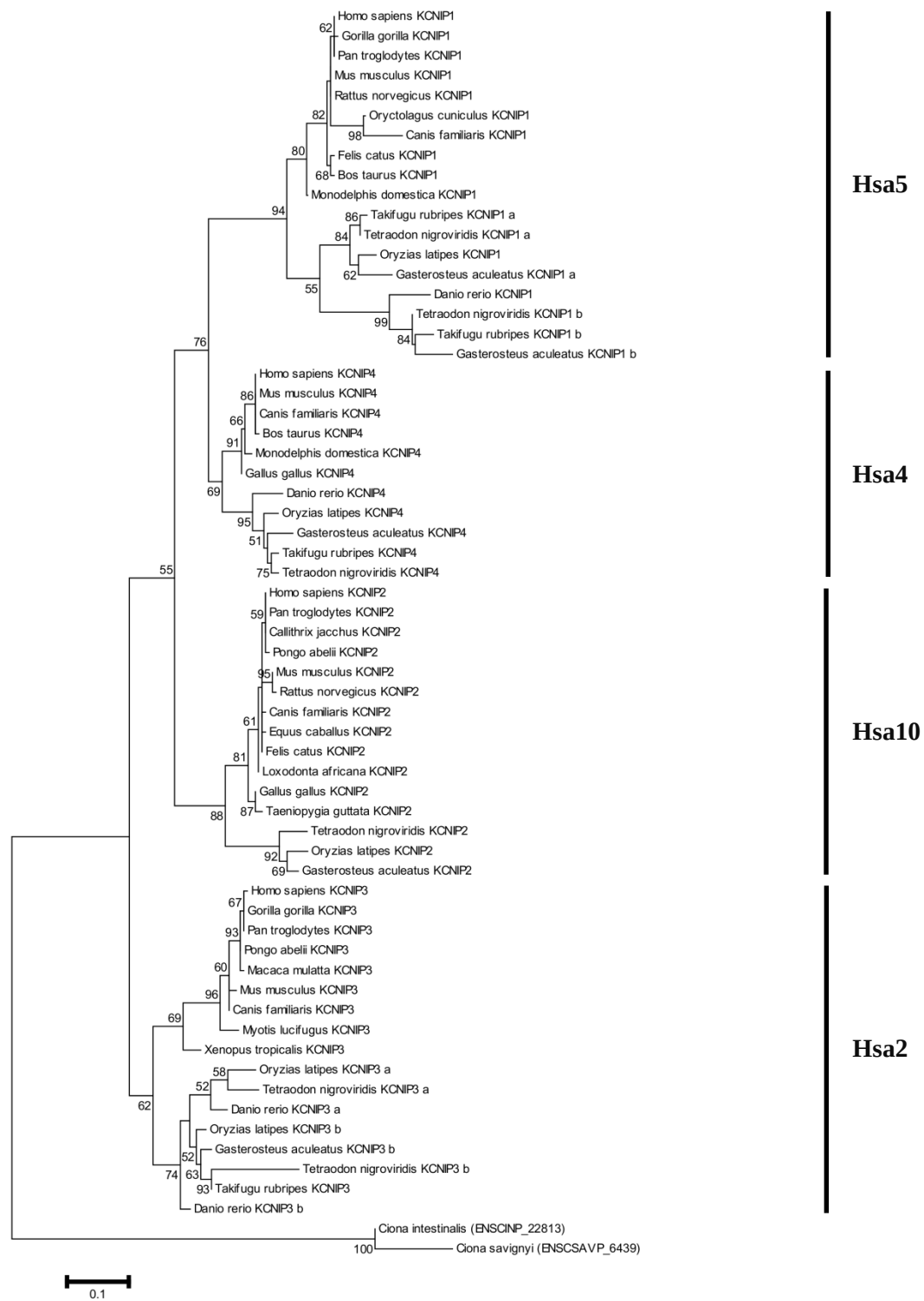
HtrA Serine Peptidase-HTRA



Phylogenetic tree of HTRA family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-4799.12708) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1483)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 74 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 154 positions in the final dataset.

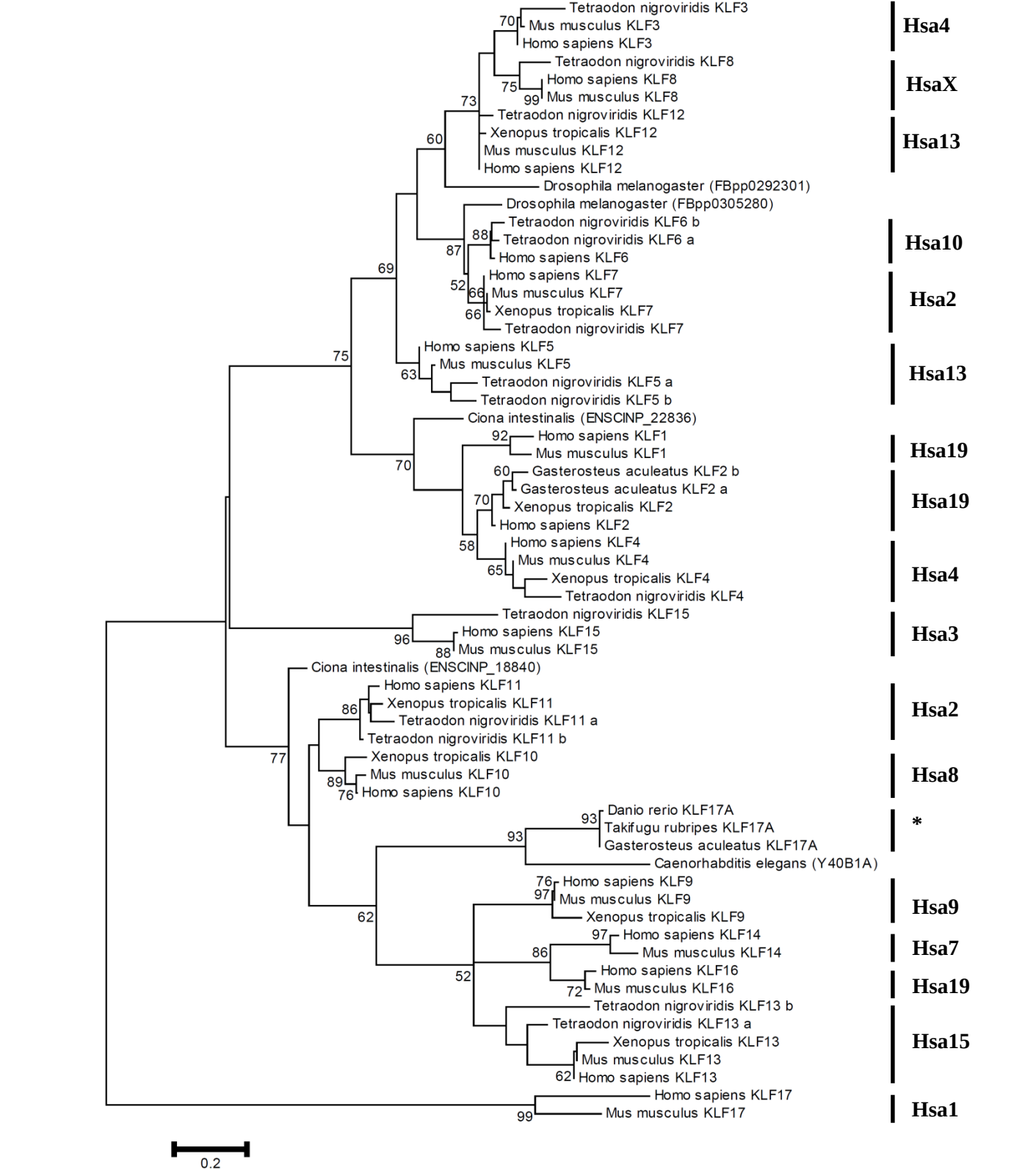
Kv Channel Interacting Protein-KCNIP



Phylogenetic tree of KCNIP family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-3289.5504) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1807)). The analysis involved 63 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 175 positions in the final dataset.

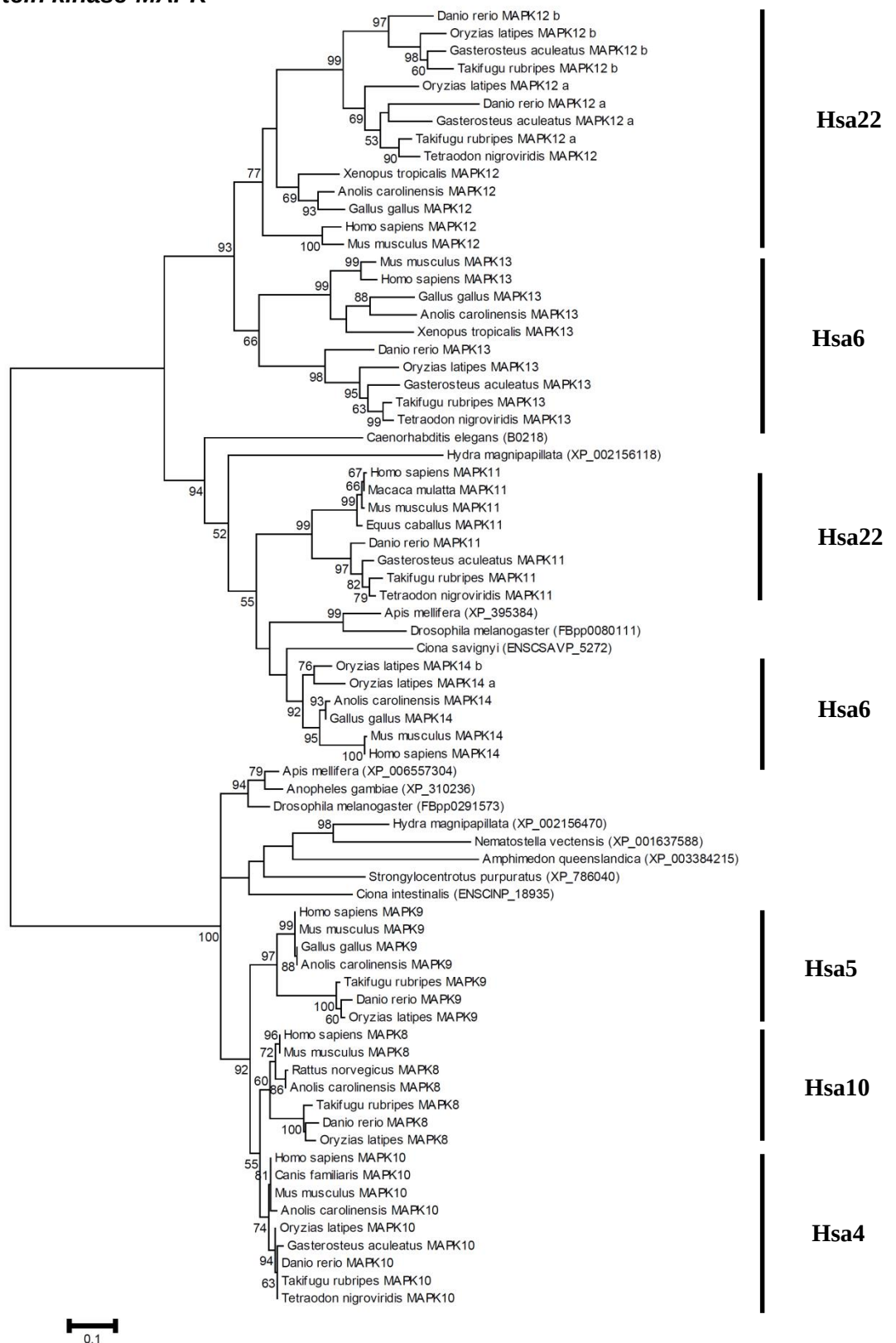
Kruppel-like factor-KLF



Phylogenetic tree KLF family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-3227.7658)is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 0.5983)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 63 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 86 positions in the final dataset. * shows another paralog which is missing in tetrapods.

Mitogen-activated protein kinase-MAPK

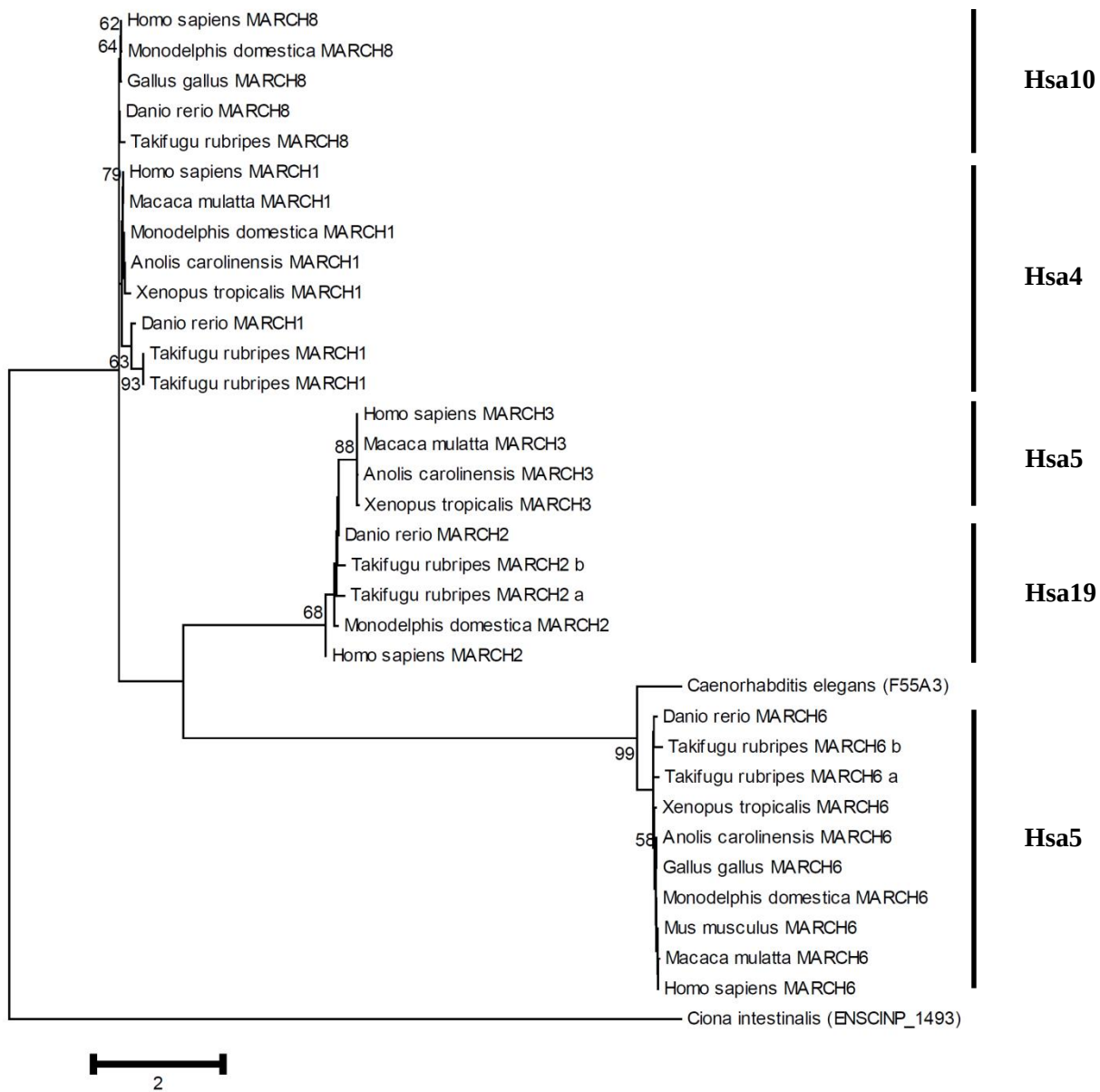


0.1

Phylogenetic tree of MAPK family.

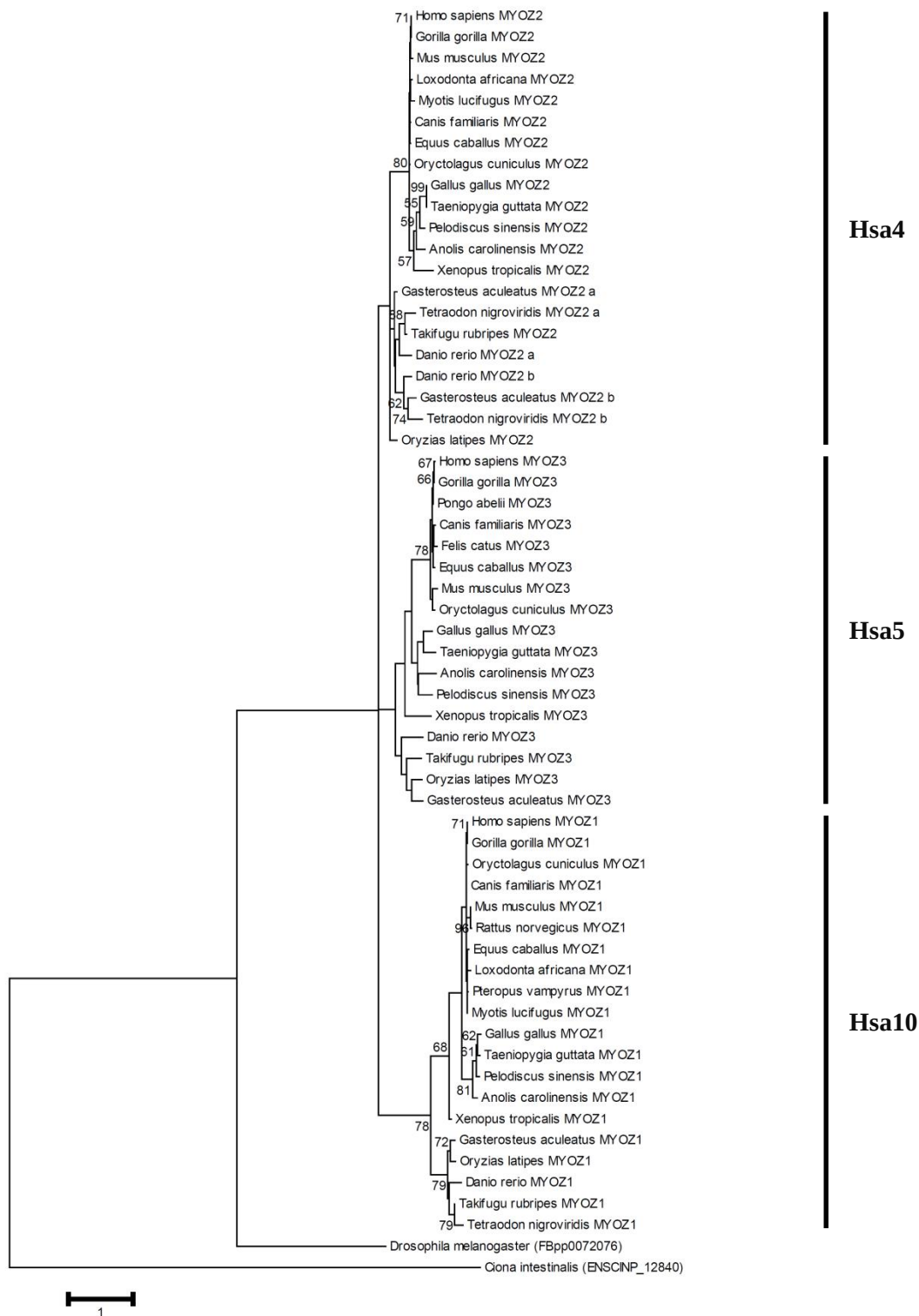
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-9050.8359) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 1.01277). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 74 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 246 positions in the final dataset.

Membrane-associated ring finger (C3HC4)-MARCH



Phylogenetic tree MARCH family.

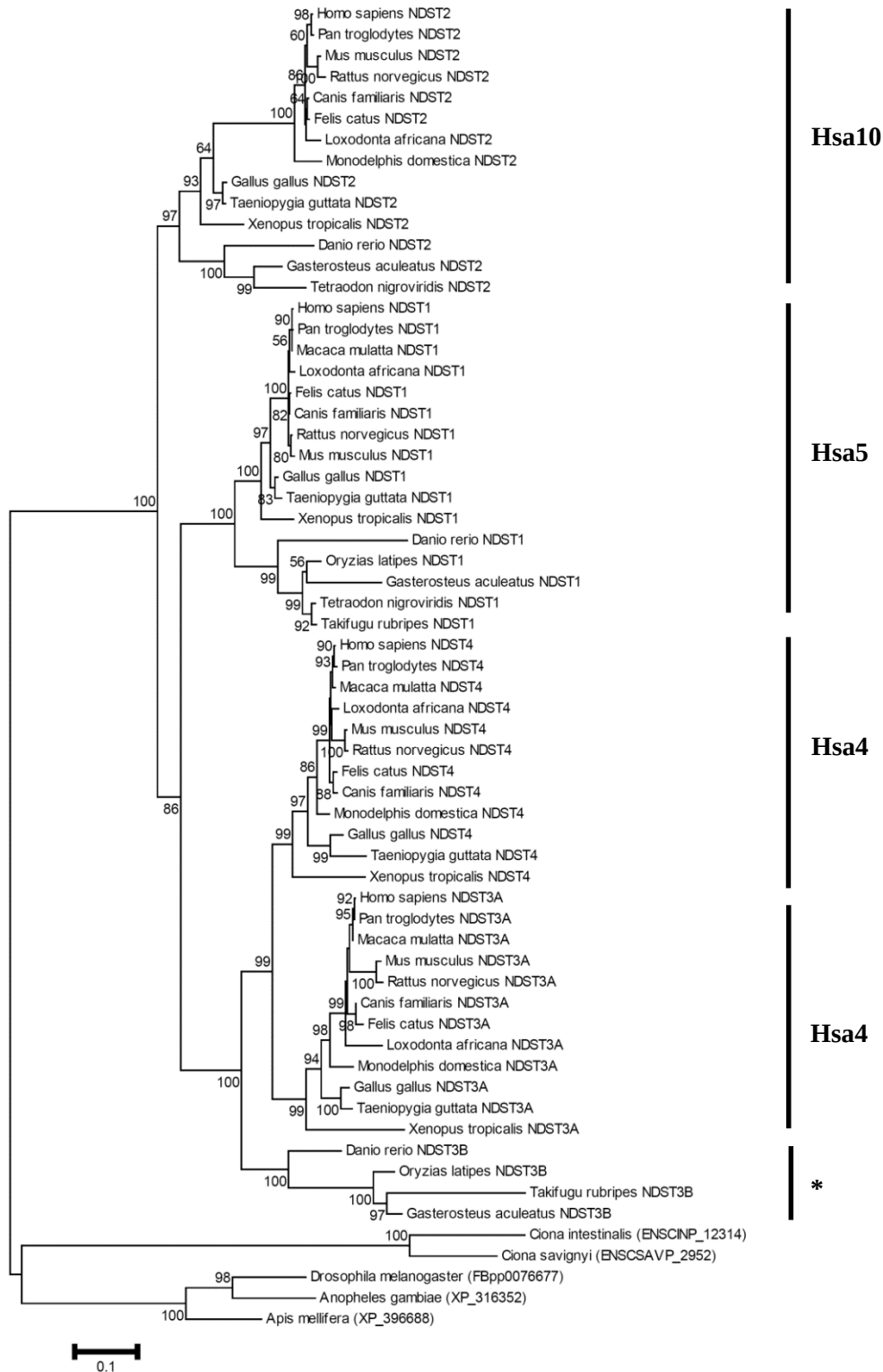
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-2937.2604)is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 5.5449). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 34 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 138 positions in the final dataset.



Phylogenetic tree of MYOZ family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-3847.0421) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.9890)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 60 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 83 positions in the final dataset.

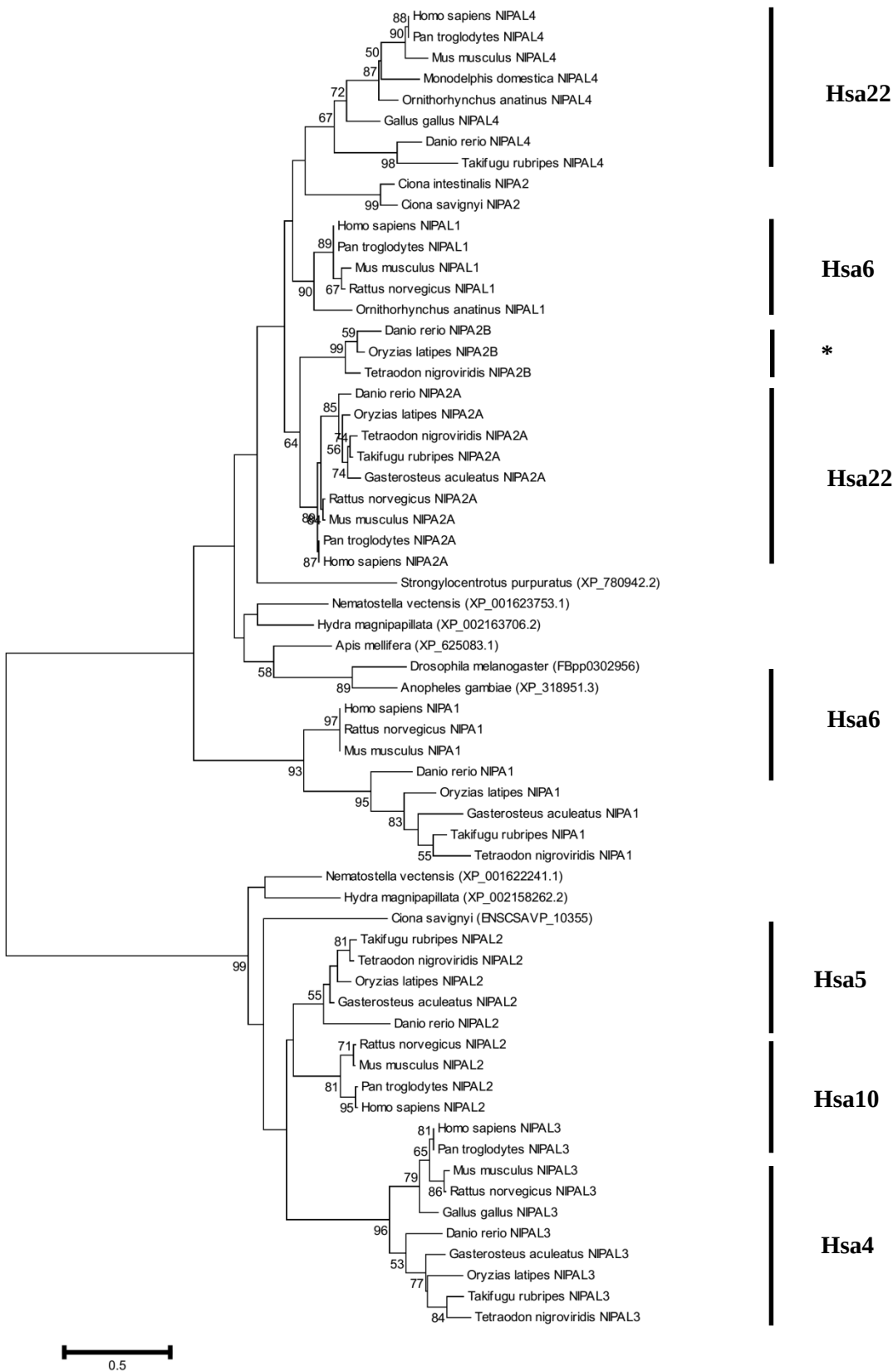
N-deacetylase/N-sulfotransferase-NDST



Phylogenetic tree of NDST family.

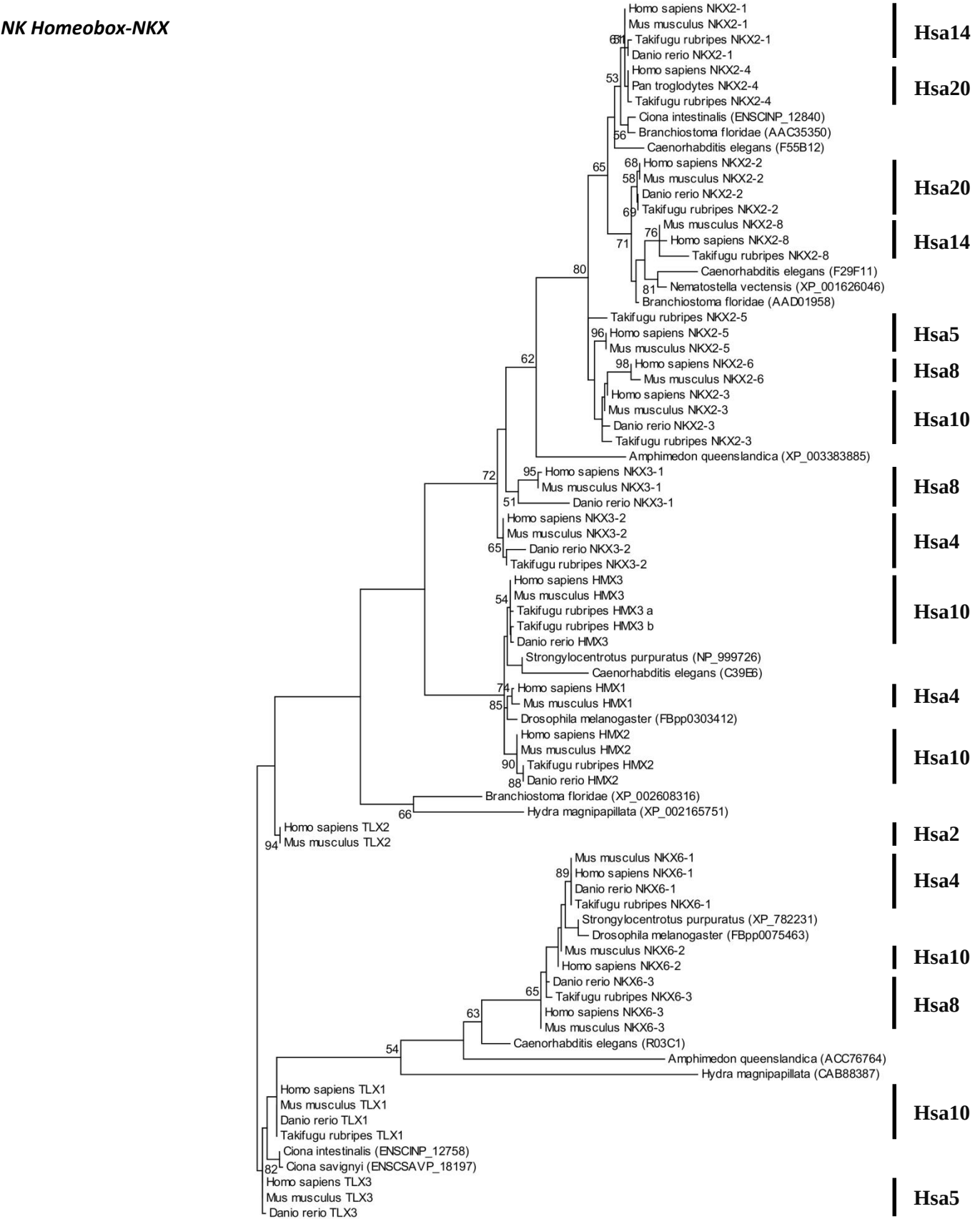
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-18402.9545) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter 1.0686)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 64 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 748 positions in the final dataset.. * is showing another paralog with deleted tetrapods.

NIPA-like domain



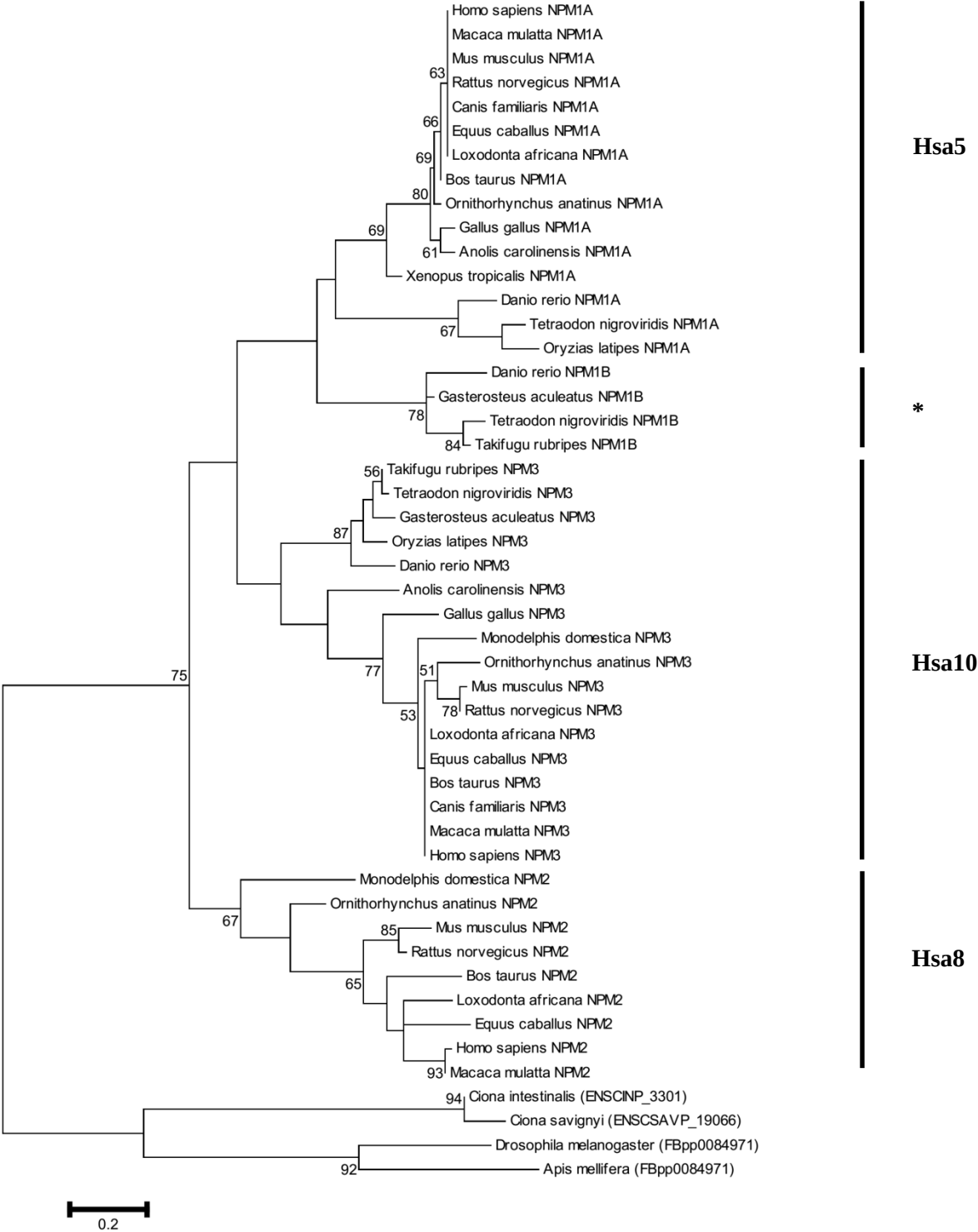
Phylogenetic tree of NIPAL family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-6809.5119) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 2.4774). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 63 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 134 positions in the final dataset. * shows another paralog which is missing in tetrapods.



Phylogenetic tree of NKX family.

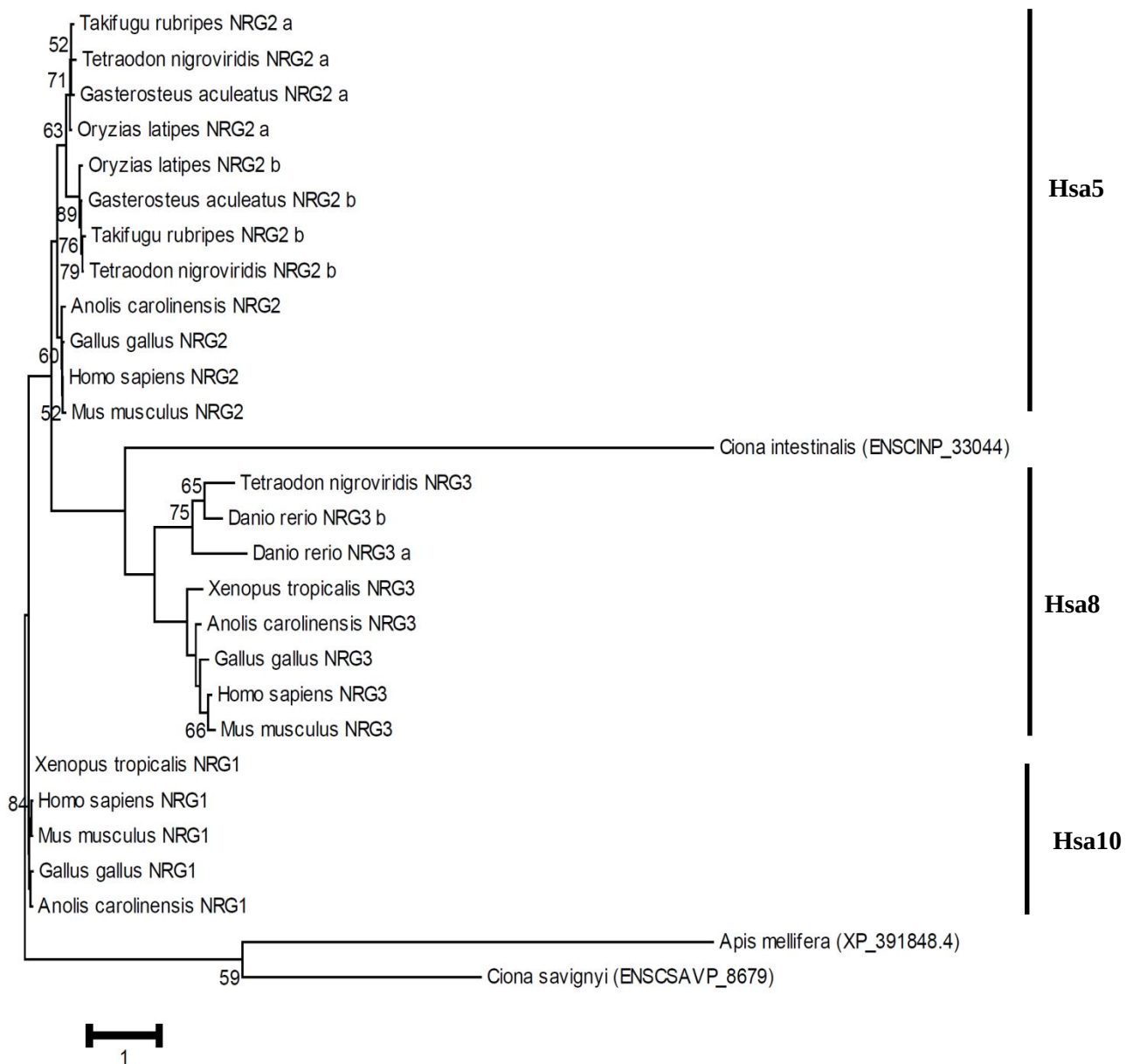
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-1876.7499) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.5369)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 79 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 58 positions in the final dataset.



Phylogenetic tree of NPM family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-1953.0165) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 6.0735)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 49 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 56 positions in the final dataset. * shows another paralog which is missing in tetrapods.

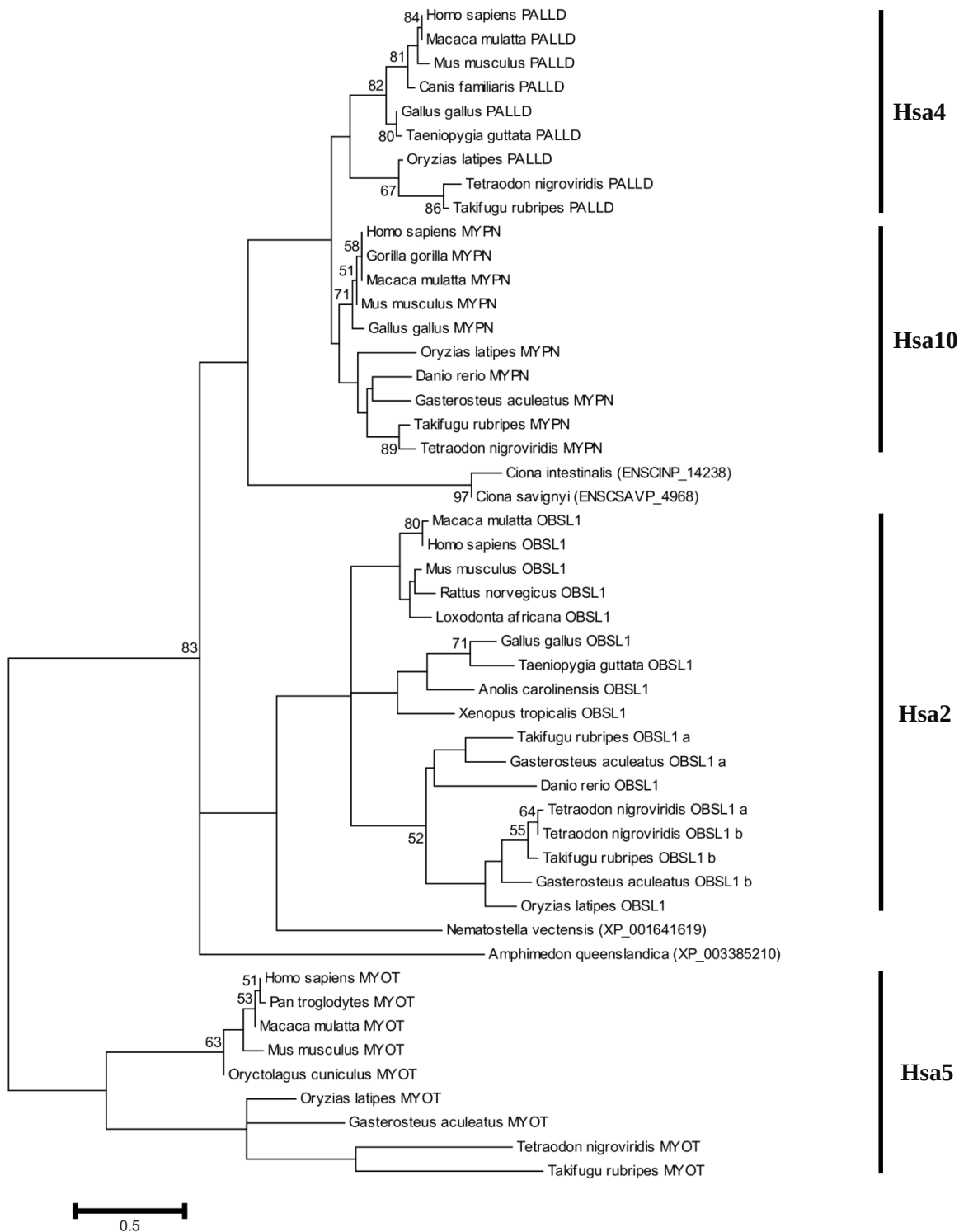
Neuregulin-NRG



Phylogenetic tree of NRG family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-3676.2485) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 41.8112)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 28 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 115 positions in the final dataset.

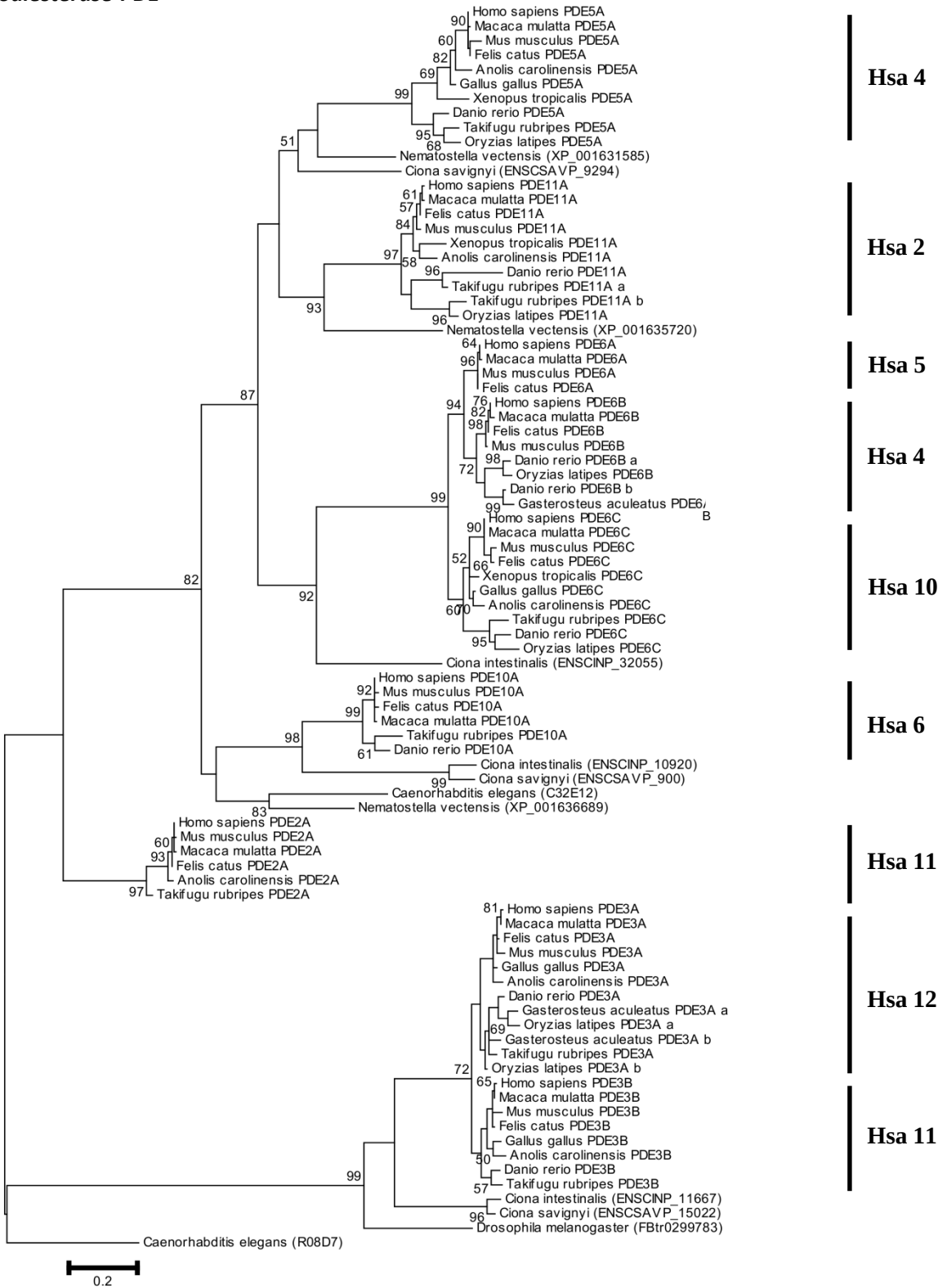
Palladin-PALLD



Phylogenetic tree of PALLD family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-2371.4347) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 9.6487)). The analysis involved 49 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 43 positions in the final dataset.

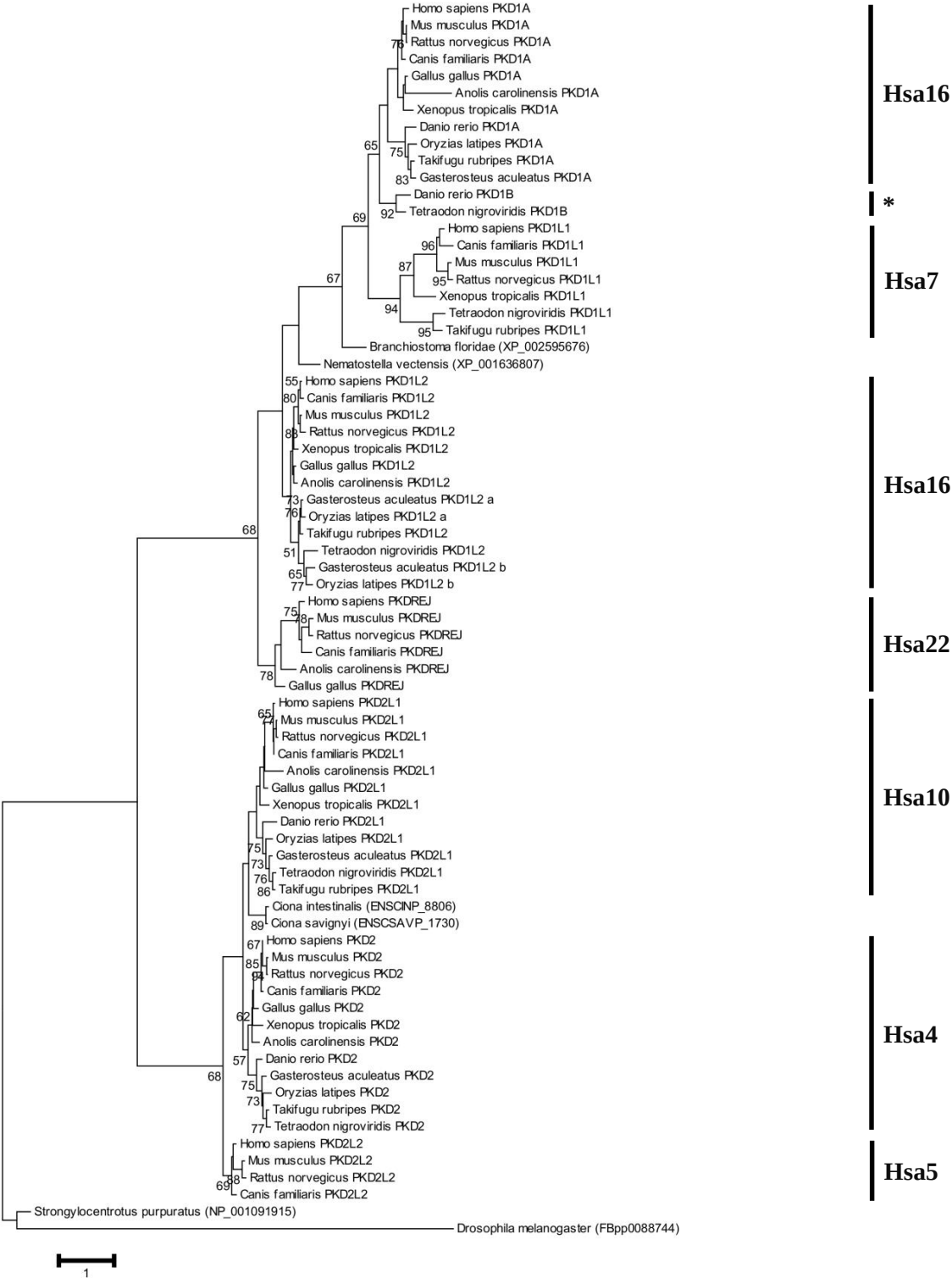
Phosphodiesterase-PDE



Phylogenetic tree of PDE family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-8864.4409) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.1697). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 86 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 181 positions in the final dataset.

Polycystic Kidney Disease-PKD



Phylogenetic tree of PKD family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-14217.2463) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 8.8533)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 73 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 179 positions in the final dataset. * shows another paralog which is missing in tetrapods.

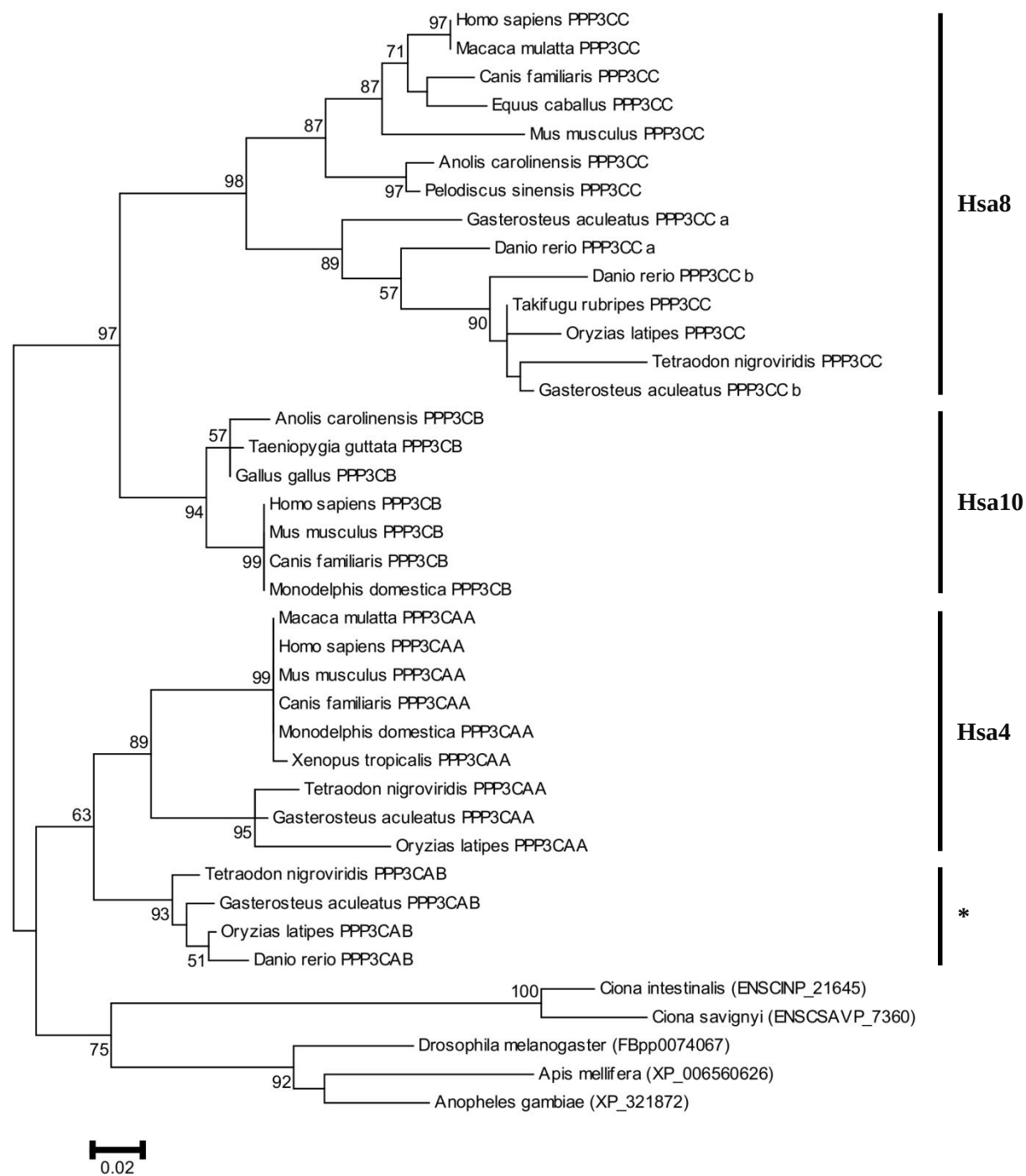
POU Family



Phylogenetic tree POU family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-2290.6445)is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 1.2866). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 61 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 64 positions in the final dataset. * shows another paralog which is missing in tetrapods.

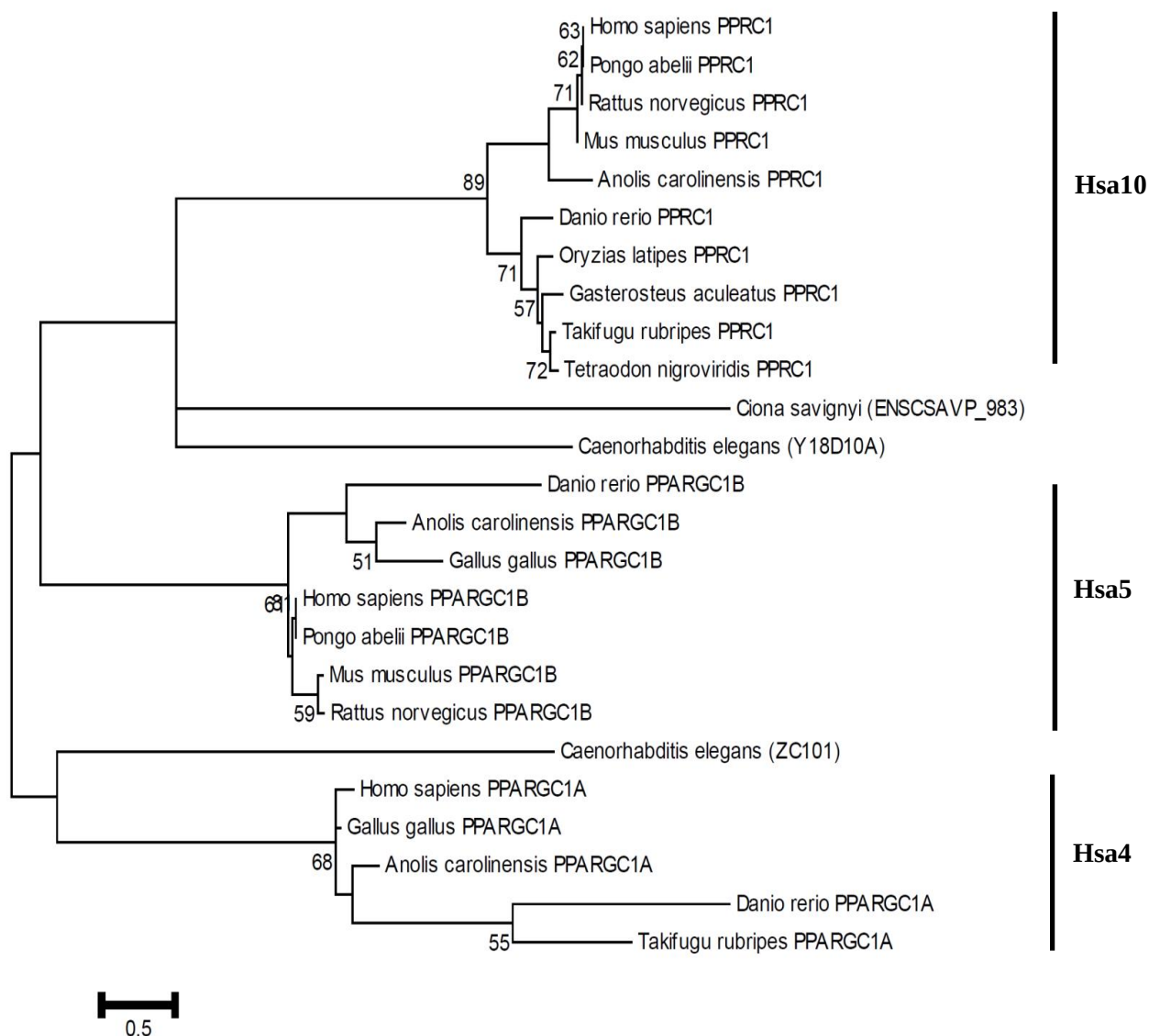
Protein Phosphate 3 Catalytic Subunit-PPP3C



Phylogenetic tree of PPP3C family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-2083.3348) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.4114)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 39 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 199 positions in the final dataset. * shows another paralog which is missing in tetrapods.

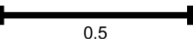
Peroxisome Proliferator-Activated Receptor Gamma, Coactivator -PPRGC



Phylogenetic tree of PPRGC family.

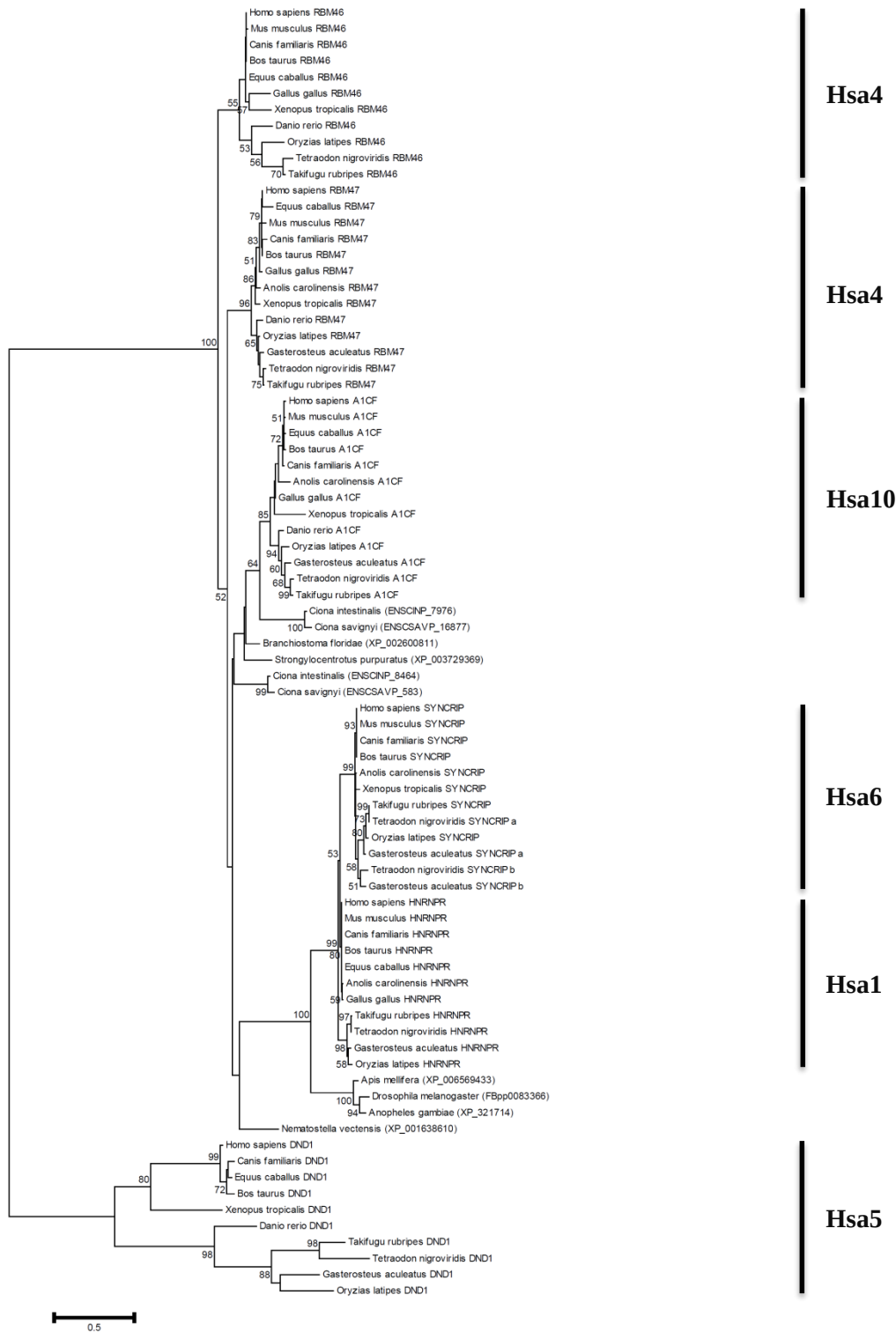
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-3989.7829) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter =20.3337)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 25 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 104 positions in the final dataset.

Phylogenetic tree of PSD family.



The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-2484.6346) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.9656)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 58 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 97 positions in the final dataset.

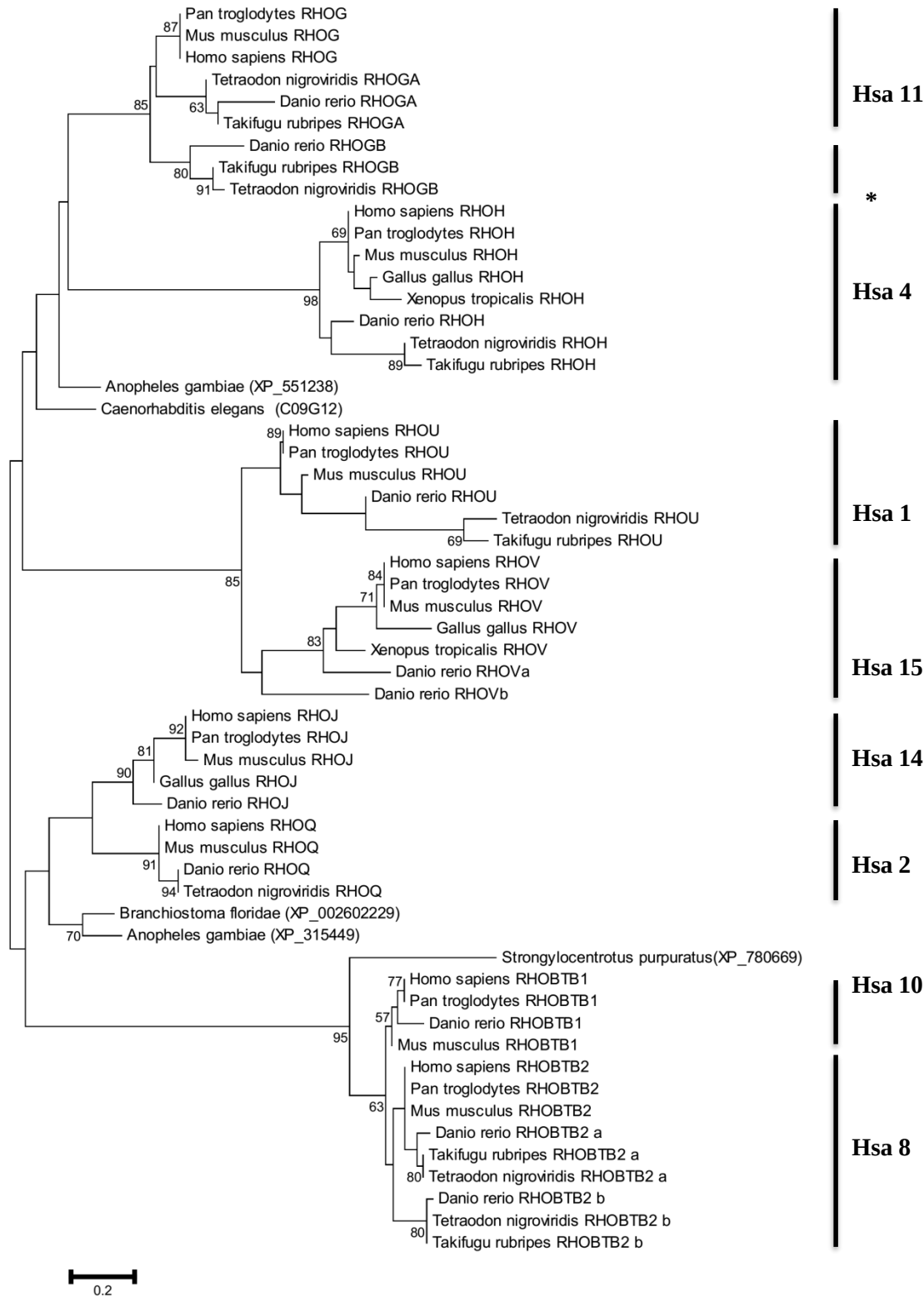
RNA Binding Motif-RBM



Phylogenetic tree of RBM family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-8698.9855) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.7095)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 80 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 219 positions in the final dataset.

Ras Homolog Family -RHO



Phylogenetic tree of RHO family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-1866.9600) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.3341)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 57 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 48 positions in the final dataset. * shows another paralog which is missing in tetrapods.

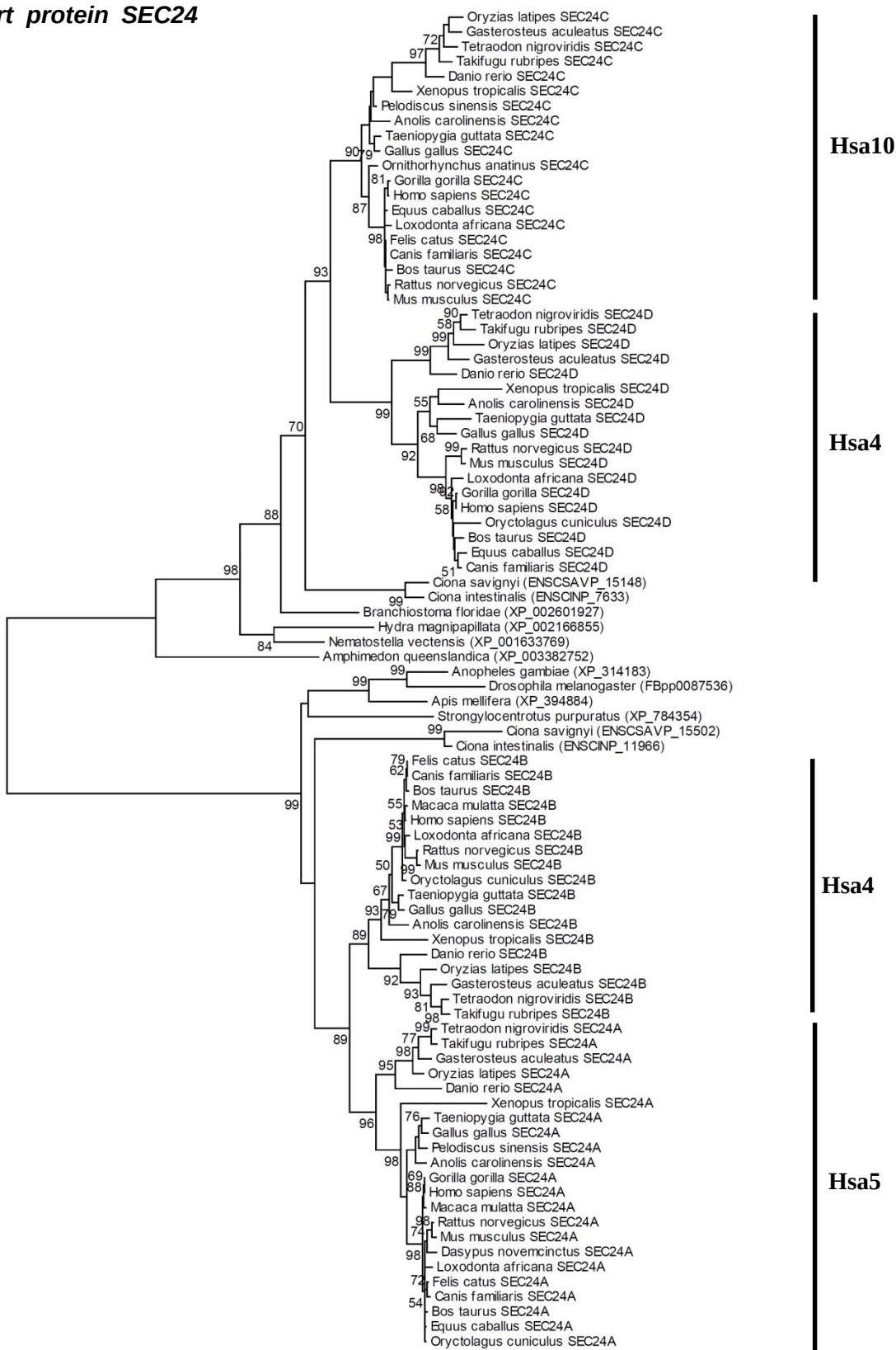
RUN and FYVE domain containing-RUFY



Phylogenetic tree of RUFY family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-5901.5586) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.4215)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 87 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 167 positions in the final dataset.

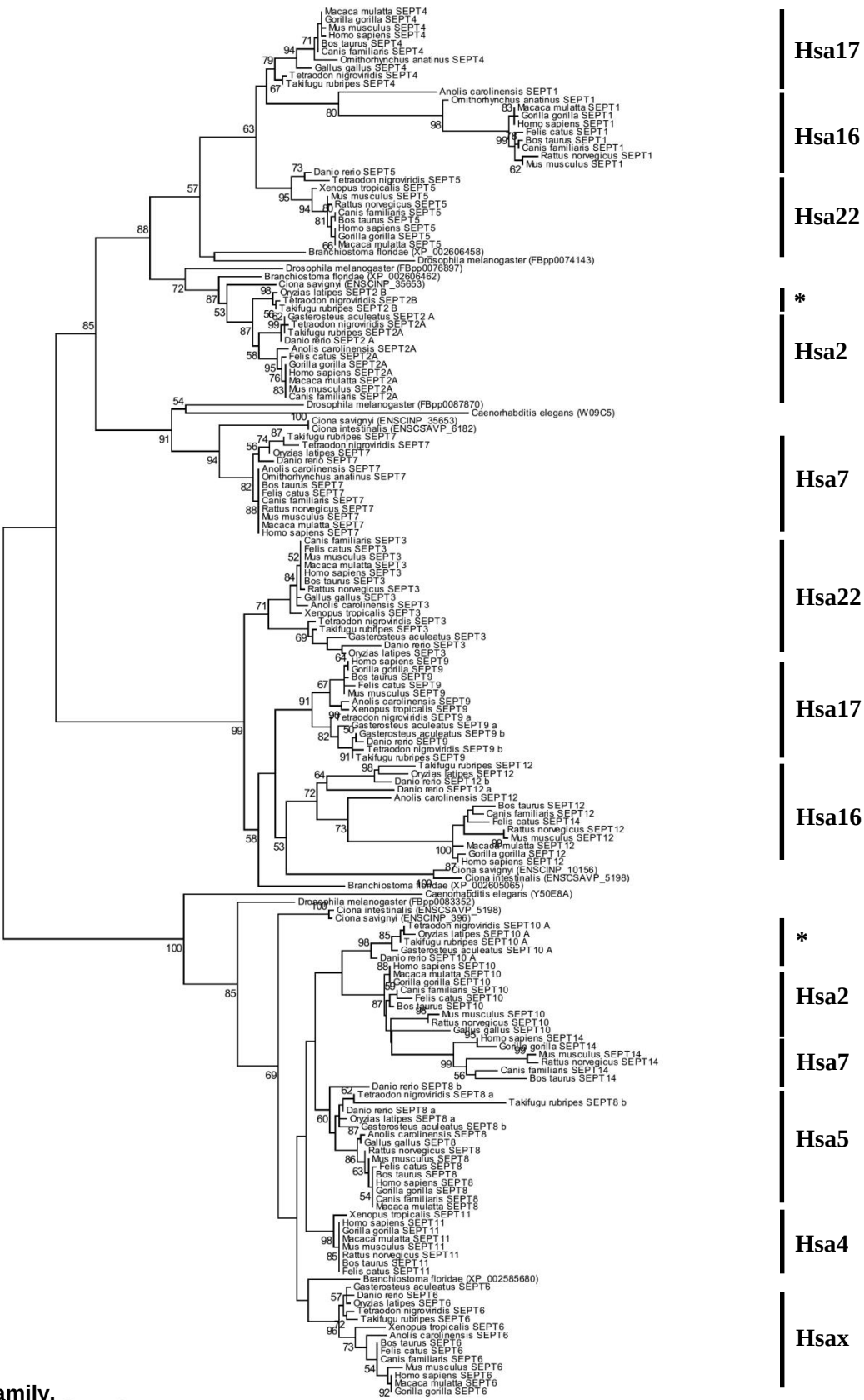
Protein transport protein SEC24



Phylogenetic tree of SEC24 family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-14500.0858) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter =2.2650)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 90 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 353 positions in the final dataset.

Septin-SEPT

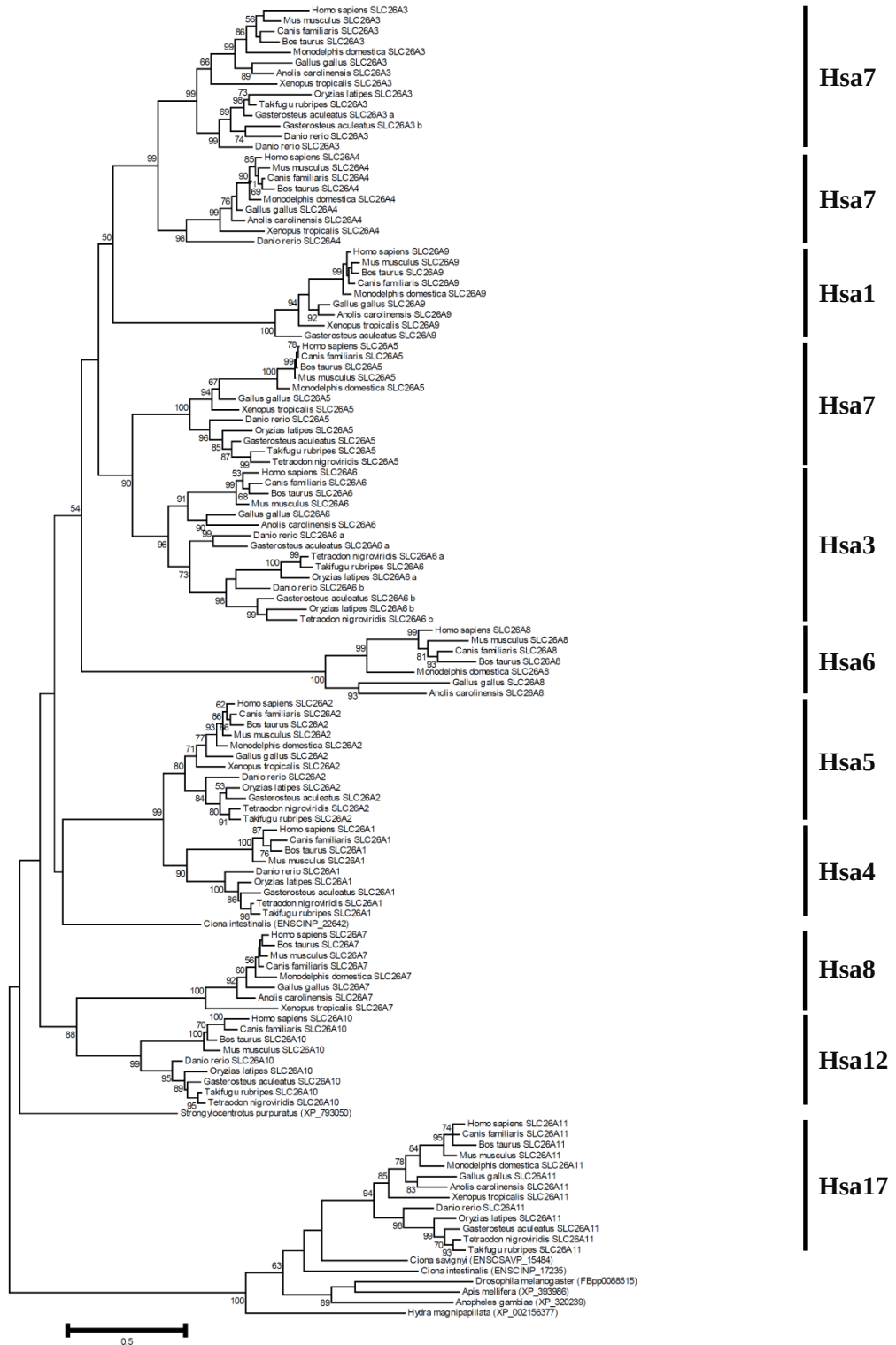


Phylogenetic tree of SEPT family.



The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-7280.2146) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.5061)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 173 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 142 positions in the final dataset. * shows another paralog which is missing in tetrapods.

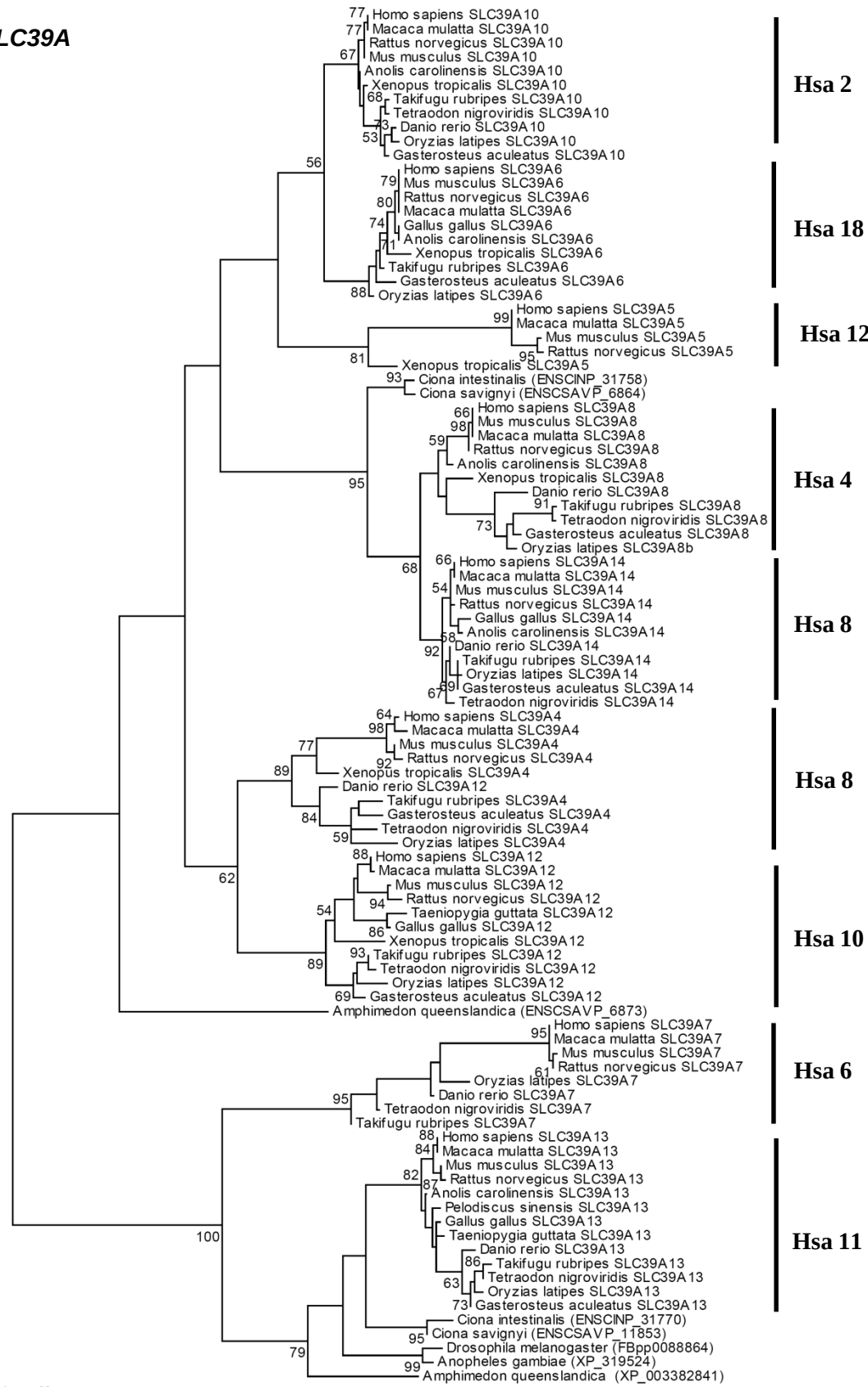
Solute Carrier Family 26-*SLC26A*



Phylogenetic tree of SLC26A family.

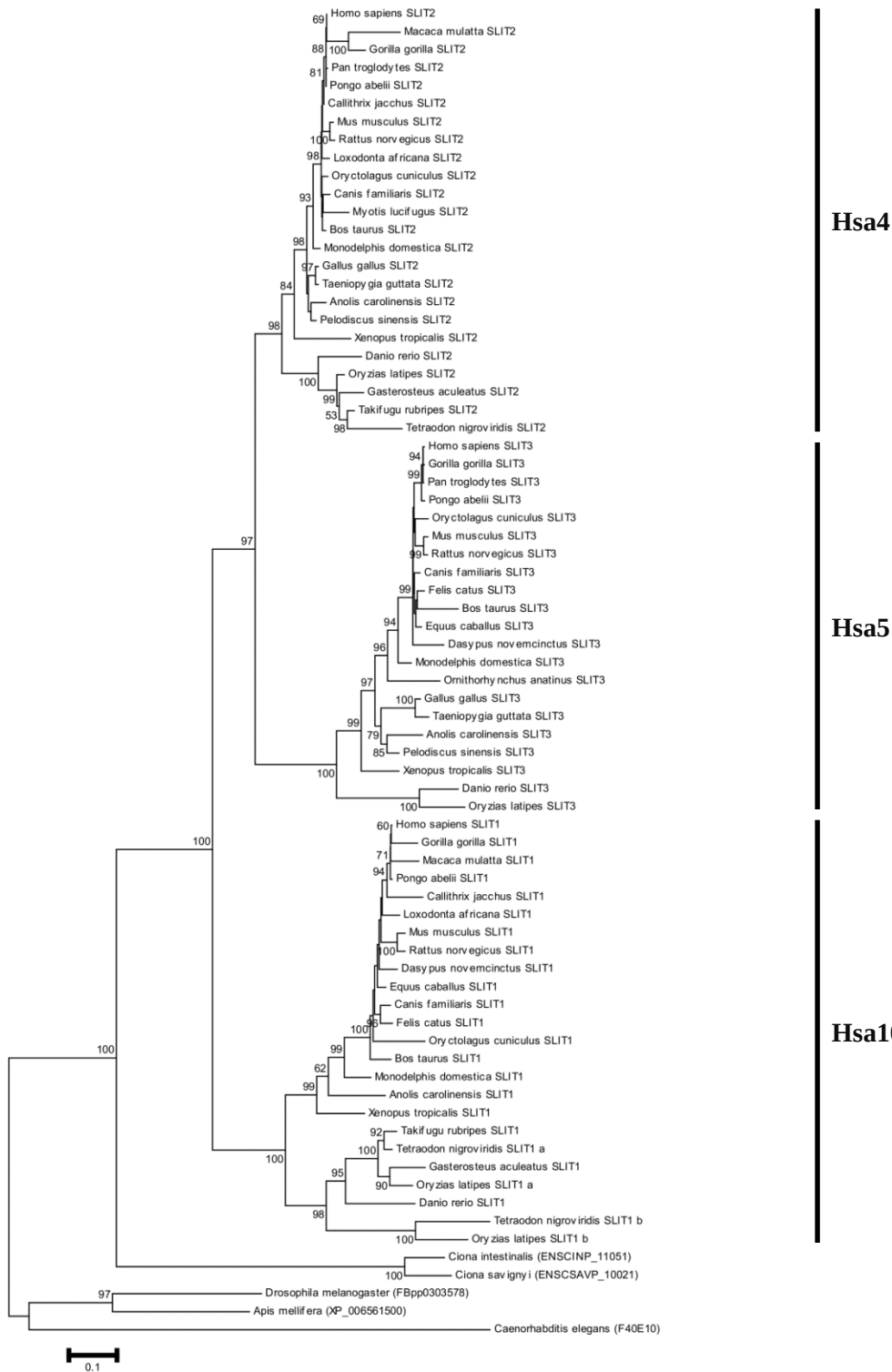
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-30358.3257) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 2.2419)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 125 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 307 positions in the final dataset.

Solute Carrier Family 39-SLC39A



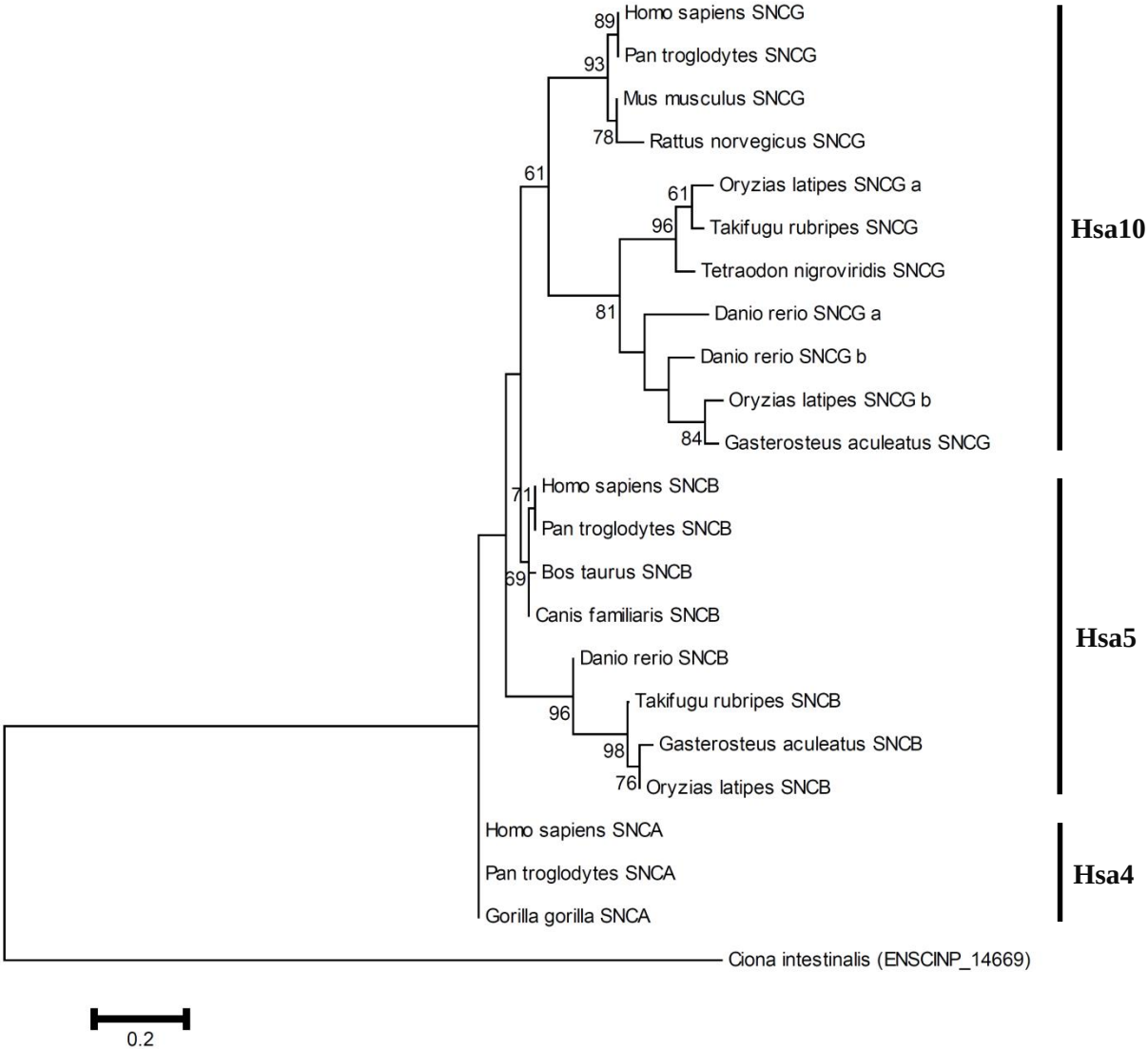
Phylogenetic tree of SLC39A family. 

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-4057.0553) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.4231)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 98 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 100 positions in the final dataset.



Phylogenetic tree of SLIT family.

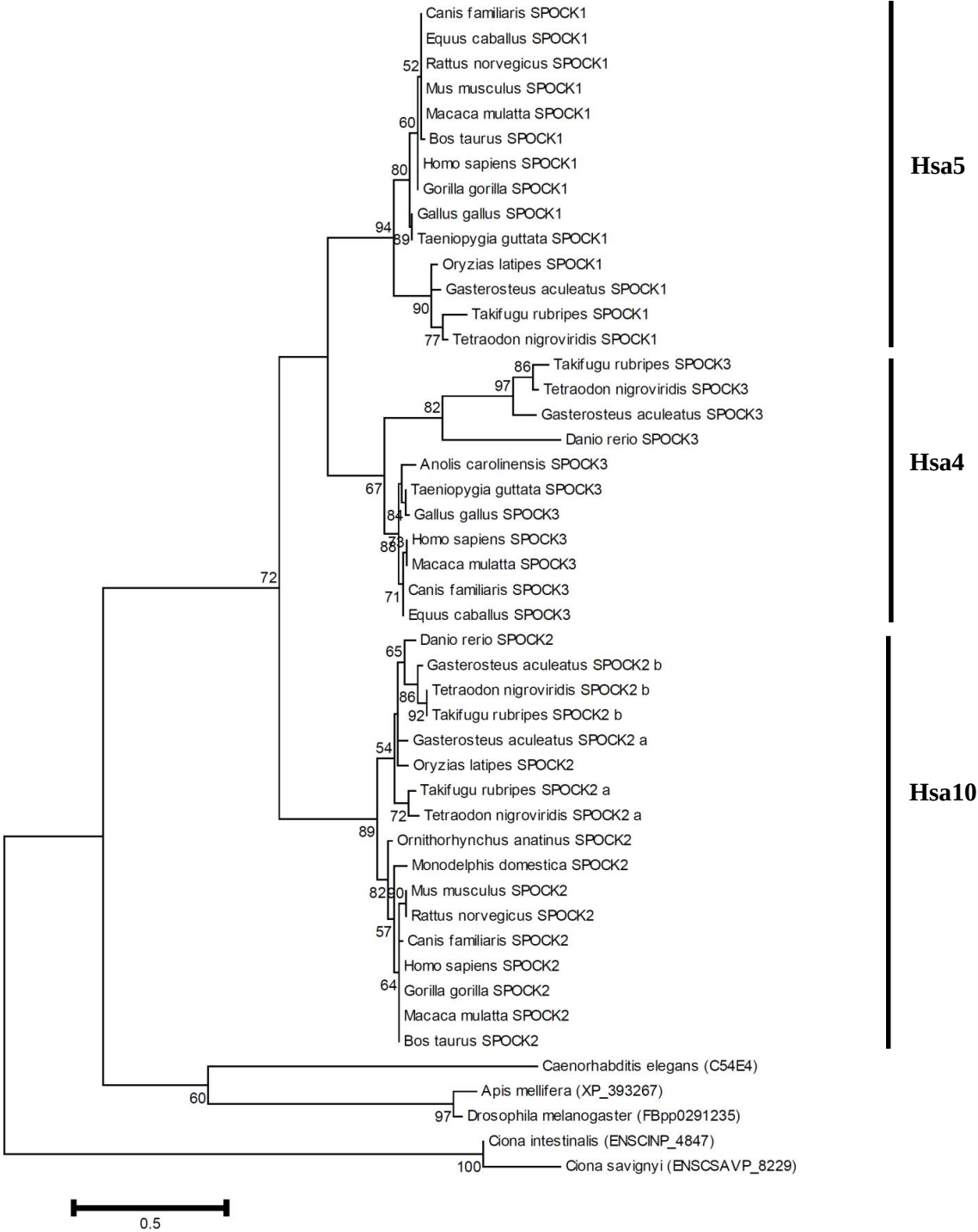
The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-25218.1039) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.5096)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 74 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 743 positions in the final dataset.



Phylogenetic tree of SNC family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-996.5533) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 4.8827)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 23 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 69 positions in the final dataset.

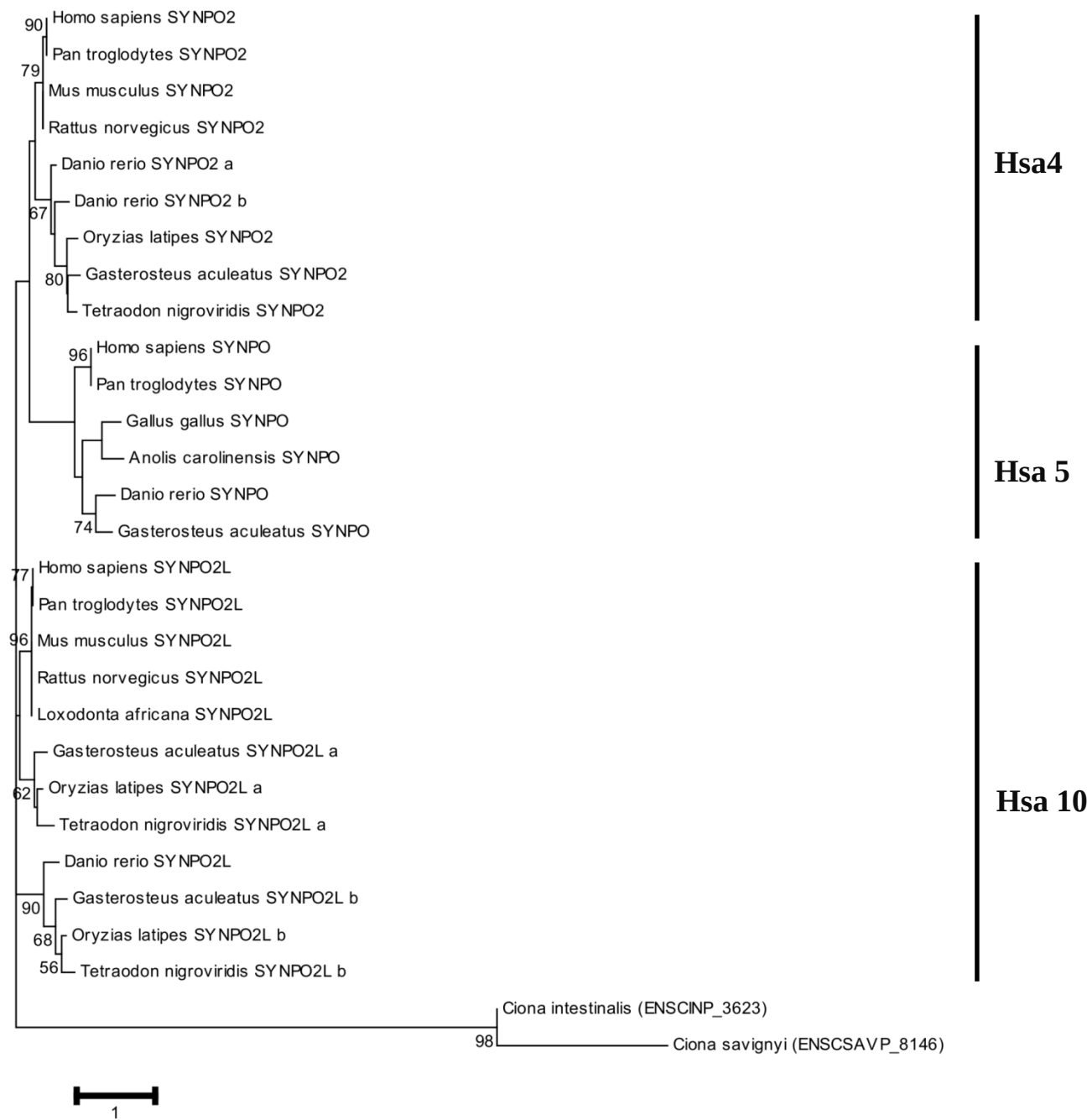
Sparc/Osteonectin, Cwcv and Kazal-like domains proteoglycan (testican)-SPOCK



Phylogenetic tree of SPOCK family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-2182.4066) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter =2.5411)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 47 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 92 positions in the final dataset.

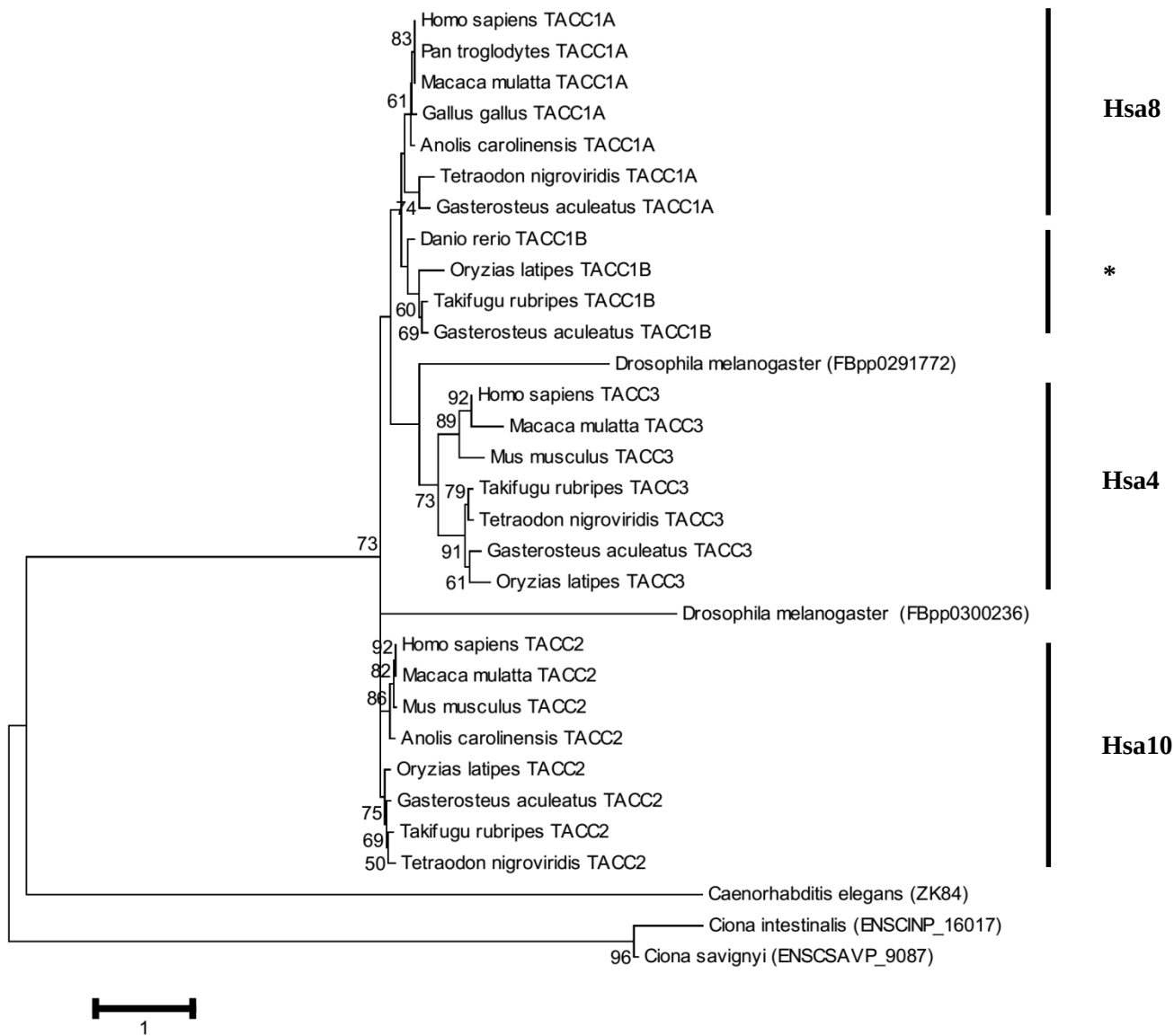
Synaptopodin-SYNPO



Phylogenetic tree of SYNPO family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood(-1966.2378) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 4.3076)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 29 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 64 positions in the final dataset.

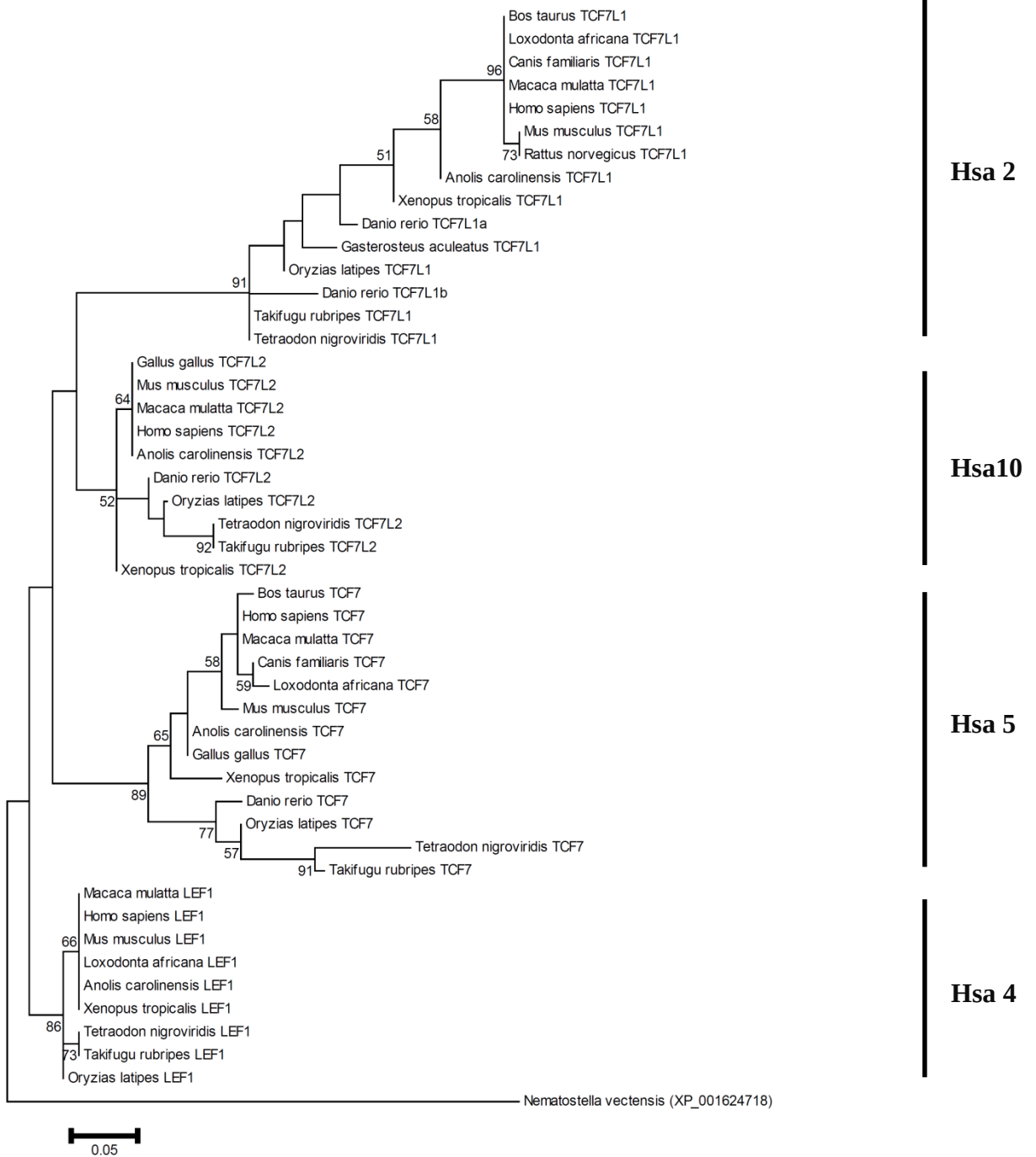
Transforming, acidic coiled-coil containing protein-TACC



Phylogenetic tree of TACC family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG) . The tree with the highest log likelihood (-6187.0554)is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter =8.0022)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 31 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 182 positions in the final dataset. * shows another paralog which is missing in tetrapods.

Transcription Factor 7-TCF7



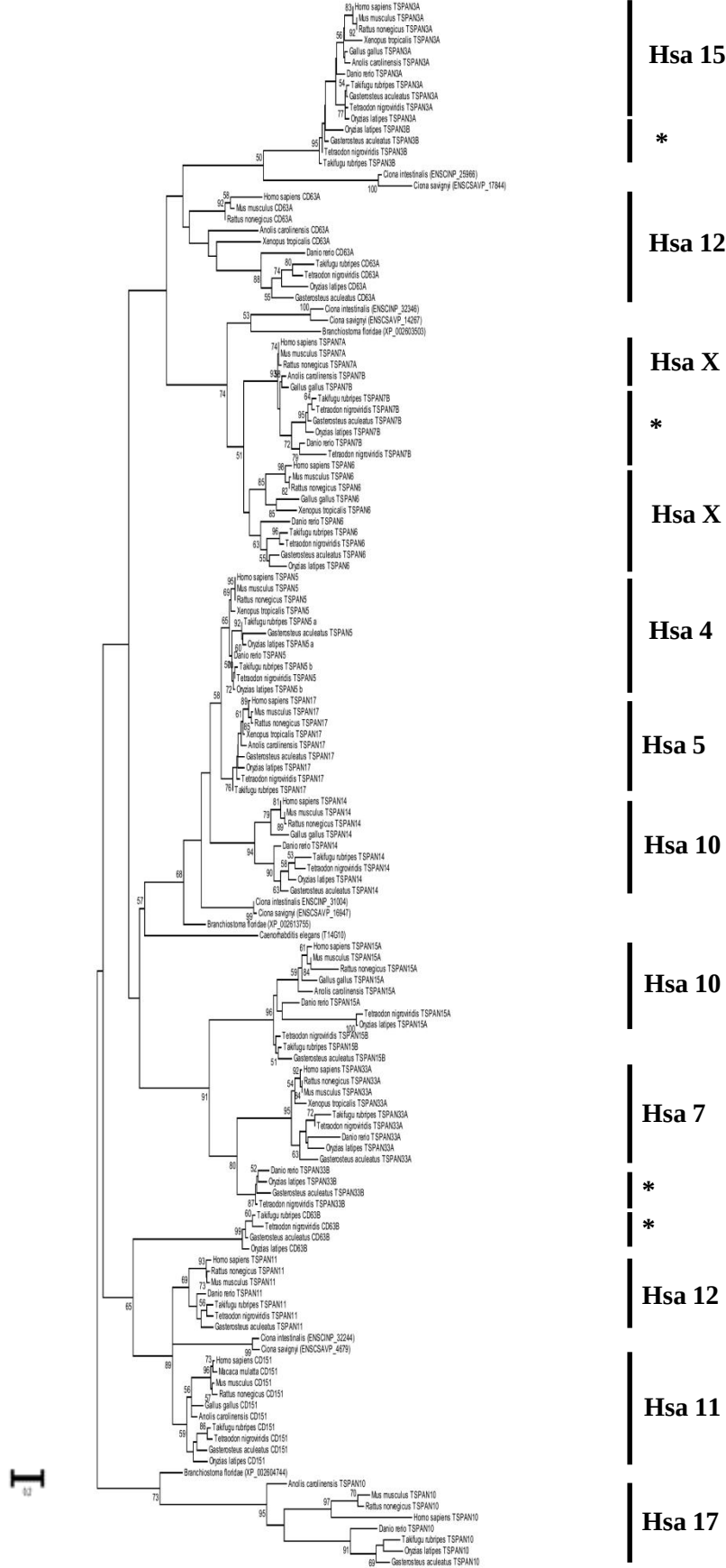
Phylogenetic tree of TCF7 family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-835.7891) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.2971)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 48 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 95 positions in the final dataset.

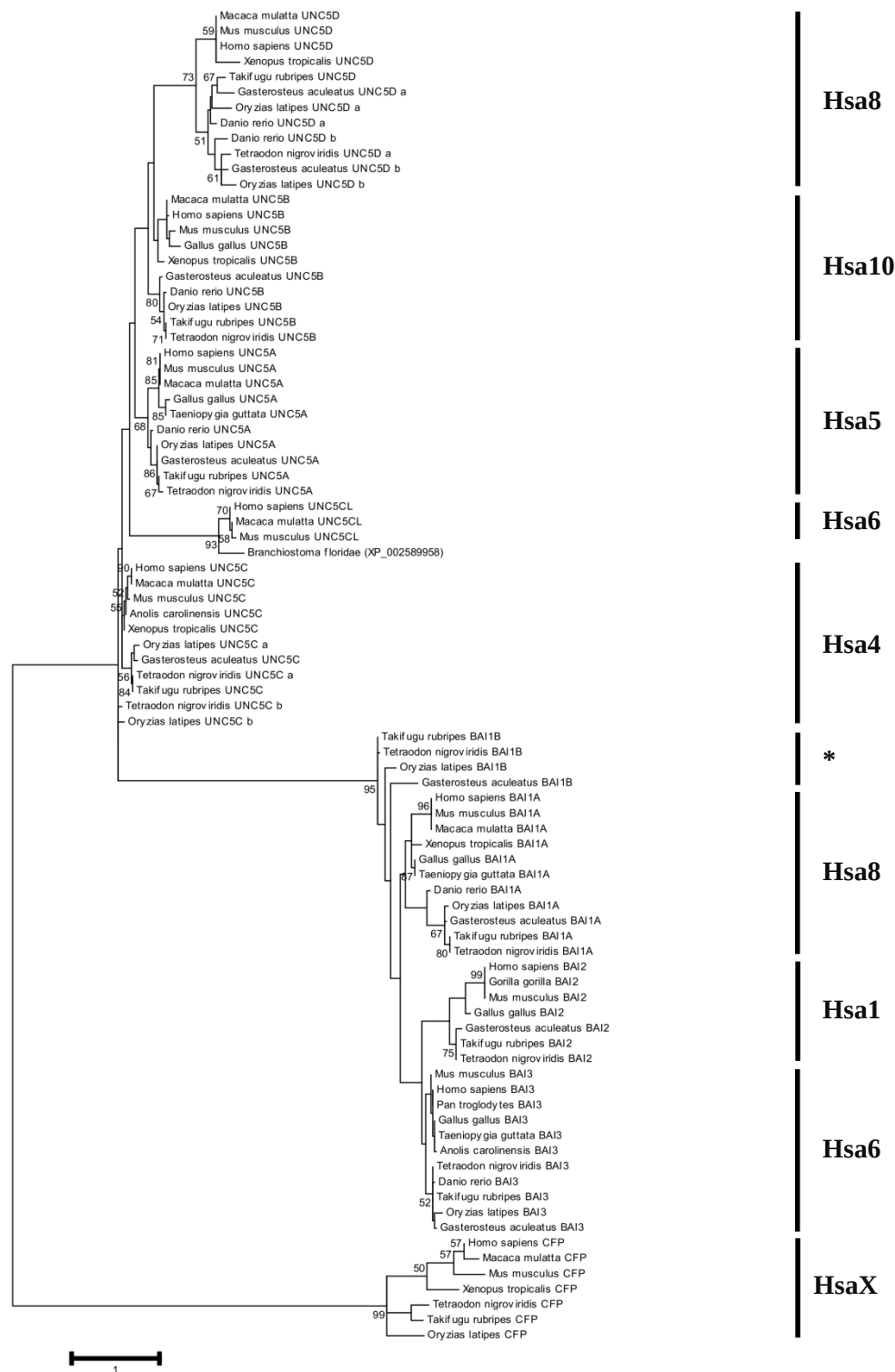
Tetraspanin-TSPAN

Phylogenetic tree of TSPAN family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the WAG. The tree with the highest log likelihood (-10402.3163) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 3.0009)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 140 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 97 positions in the final dataset. * shows another paralog with missing tetrapods.



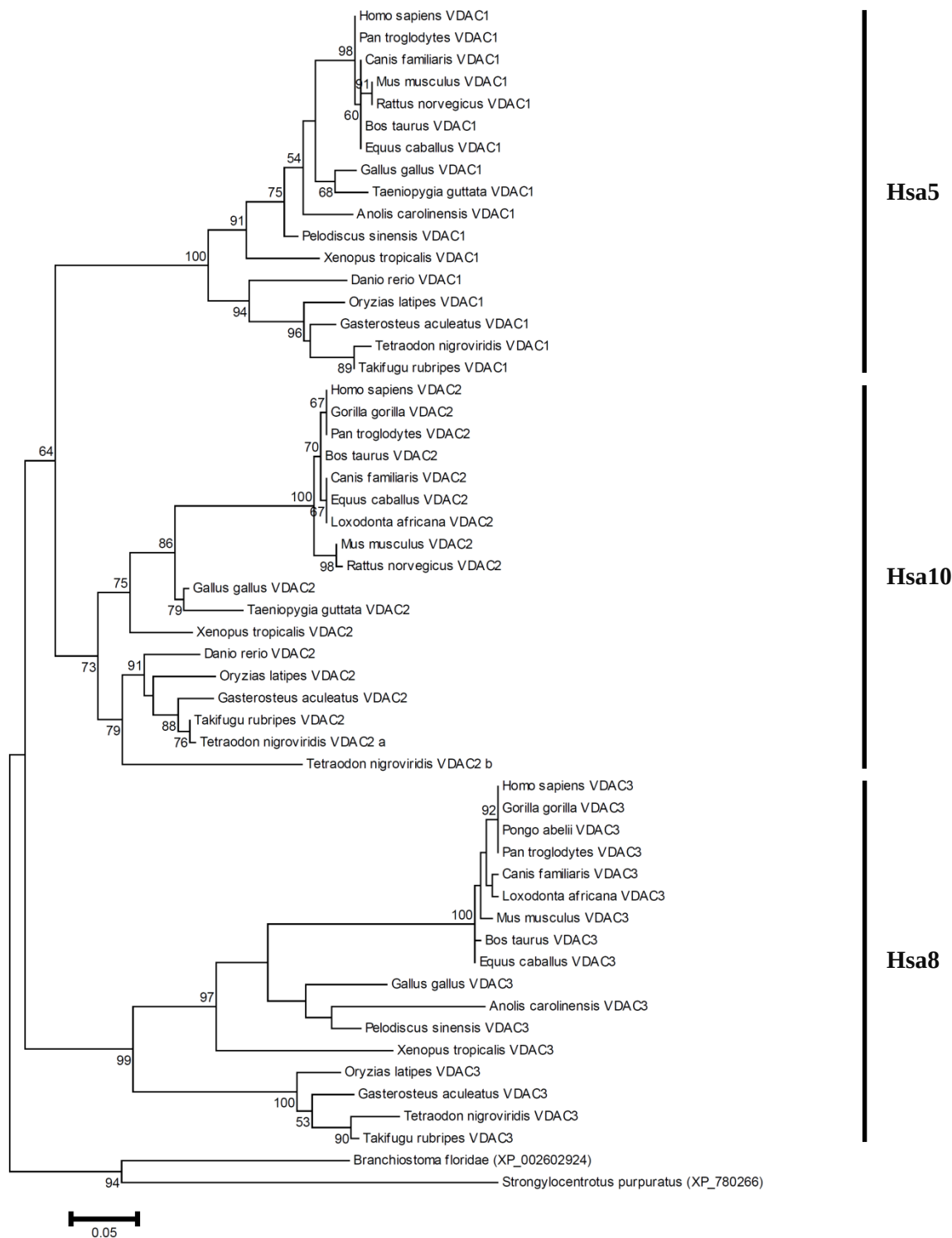
UNC-5



Phylogenetic tree of UNC5 family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-2773.3766) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 5.8922)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 87 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 50 positions in the final dataset. *shows another paralog with deleted tetrapods, named as BAI1B by our group.

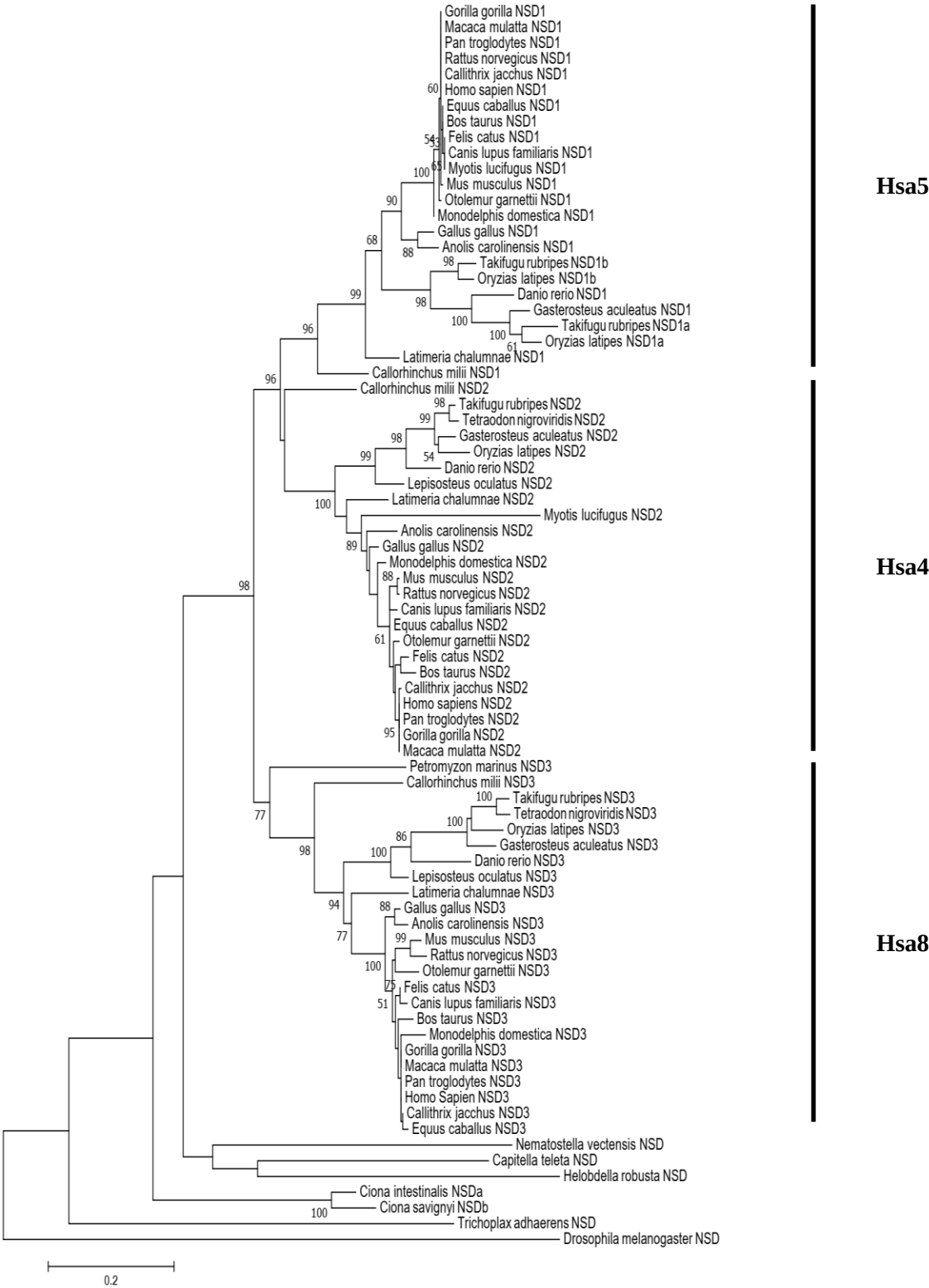
Voltage-Dependent Anion Channel-VDAC



Phylogenetic tree of VDAC family.

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model (WAG). The tree with the highest log likelihood (-3994.8623) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches; only the values $\geq 50\%$ are shown here.. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.9399)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 54 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 241 positions in the final dataset.

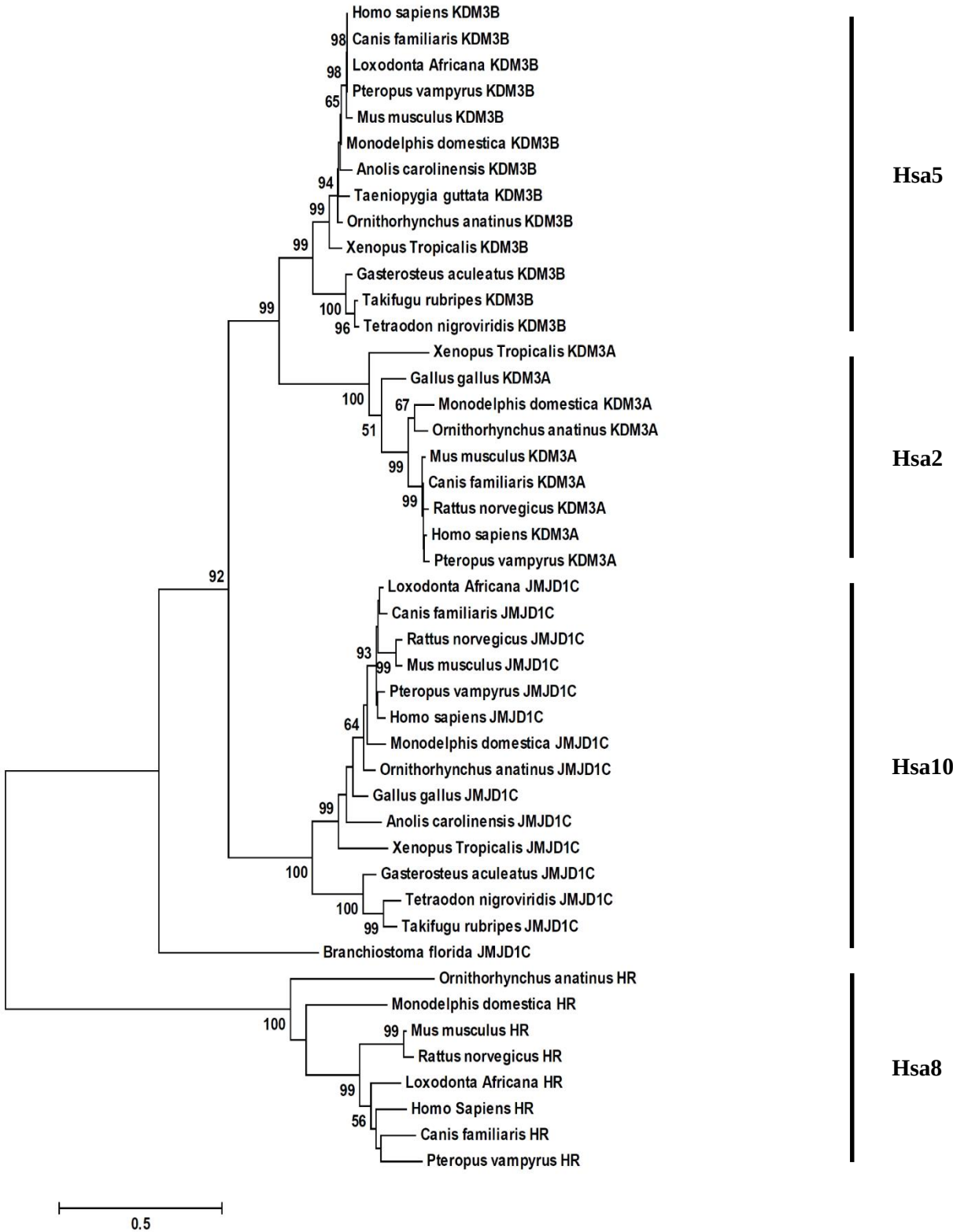
Nuclear Receptor Binding SET domain protein-NSD



Phylogenetic tree of NSD family:

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model. The tree with the highest log likelihood (-10713.81) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.3091)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 79 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 340 positions in the final dataset. Evolutionary analyses were conducted in MEGA7

HR lysine demethylase and nuclear receptor corepressor (HR)



Phylogenetic tree of HR Family:

The evolutionary history was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model. The tree with the highest log likelihood (-11085.6583) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches ; only the values $\geq 50\%$ are shown here. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.6609). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 46 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 391 positions in the final dataset.