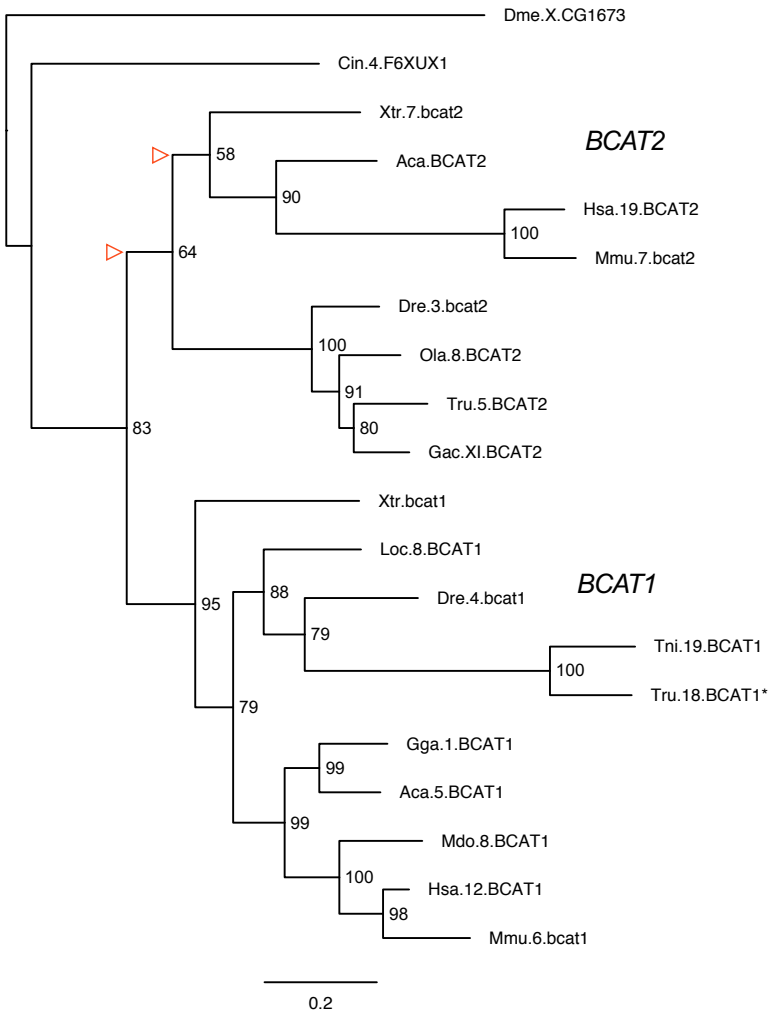
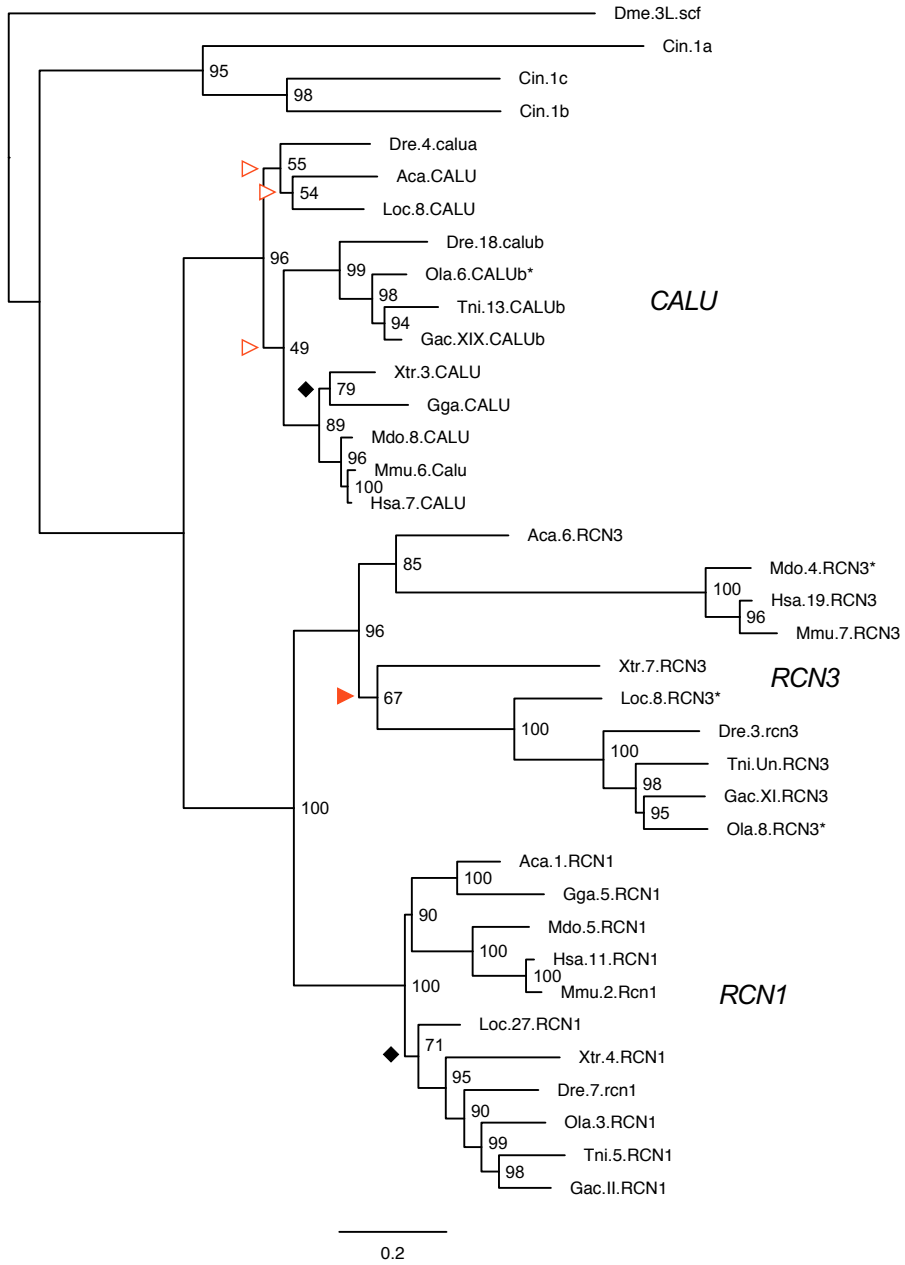
Supplementary Figures S2 – S32. Maximum likelihood phylogenies of neighboring gene families

Human, mouse, grey short-tailed opossum, chicken, Carolina anole lizard, Western clawed frog, spotted gar, zebrafish, medaka, three-spined stickleback, green spotted pufferfish, vase tunicate and fruit fly sequences were sought as described in Methods. Some neighboring family phylogenies were complemented with sequences from Mexican cave tetra (*Astyanax mexicanus*, accessed through Ensembl), Japanese pufferfish, elephant shark, Florida lancelet and/or purple sea urchin (*Strongylocentrotus purpuratus*, found through HMMER searches in EnsemblMetazoa). Phylogenetic tree topologies are supported by non-parametric Ultra-Fast Bootstrap (UFBoot) and approximate Likelihood-Ratio Test (aLRT) analyses with 1000 replicates. Only UFBoot values are shown. Filled arrowheads indicate nodes with low support in both UFBoot (≤70%) and aLRT (≤50%) tests. Outlined arrowheads indicate nodes with low UFBoot support only, and diamonds indicate nodes with low aLRT support only. Approved gene symbols and nomenclature guidelines for human, mouse, chicken, *Xenopus tropicalis* and zebrafish genes were used. Numbers following species abbreviations refer to chromosome or linkage group assignments for mapped genes. Asterisks denote partial sequences that do not span the full length of the alignment. Species abbreviations: Aca (Carolina anole lizard), Amf (honey bee), Amx (Mexican cave tetra), Bfl (*Branchiostoma floridae*), Cin (vase tunicate), Cmi (elephant shark), Dme (fruit fly), Dre (zebrafish), Gac (three-spined stickleback), Gga (chicken), Hsa (human), Hro (sea pineapple), Loc (spotted gar), Mdo (grey short-tailed opossum), Mmu (mouse), Ola (medaka), Spu (purple sea urchin), Tca (red flour beetle), Tni (green spotted pufferfish), Tru (Japanese pufferfish), Xtr (*Xenopus tropicalis*).

Supplementary figure S2. Branched chain amino acid transaminase family (BCAT). The gene family is made up of the *BCAT1* and *BCAT2* genes. The *BCAT1* branch is well-supported, however the *Xenopus tropicalis bcat1* sequence clusters in the basal-most position rather than with the other tetrapod *BCAT1* sequences. The *BCAT2* branch has low support in the UFBoot test (64). The *BCAT1* and *BCAT2* branches diverge in the time window of the basal vertebrate tetraploidizations (1R/2R).

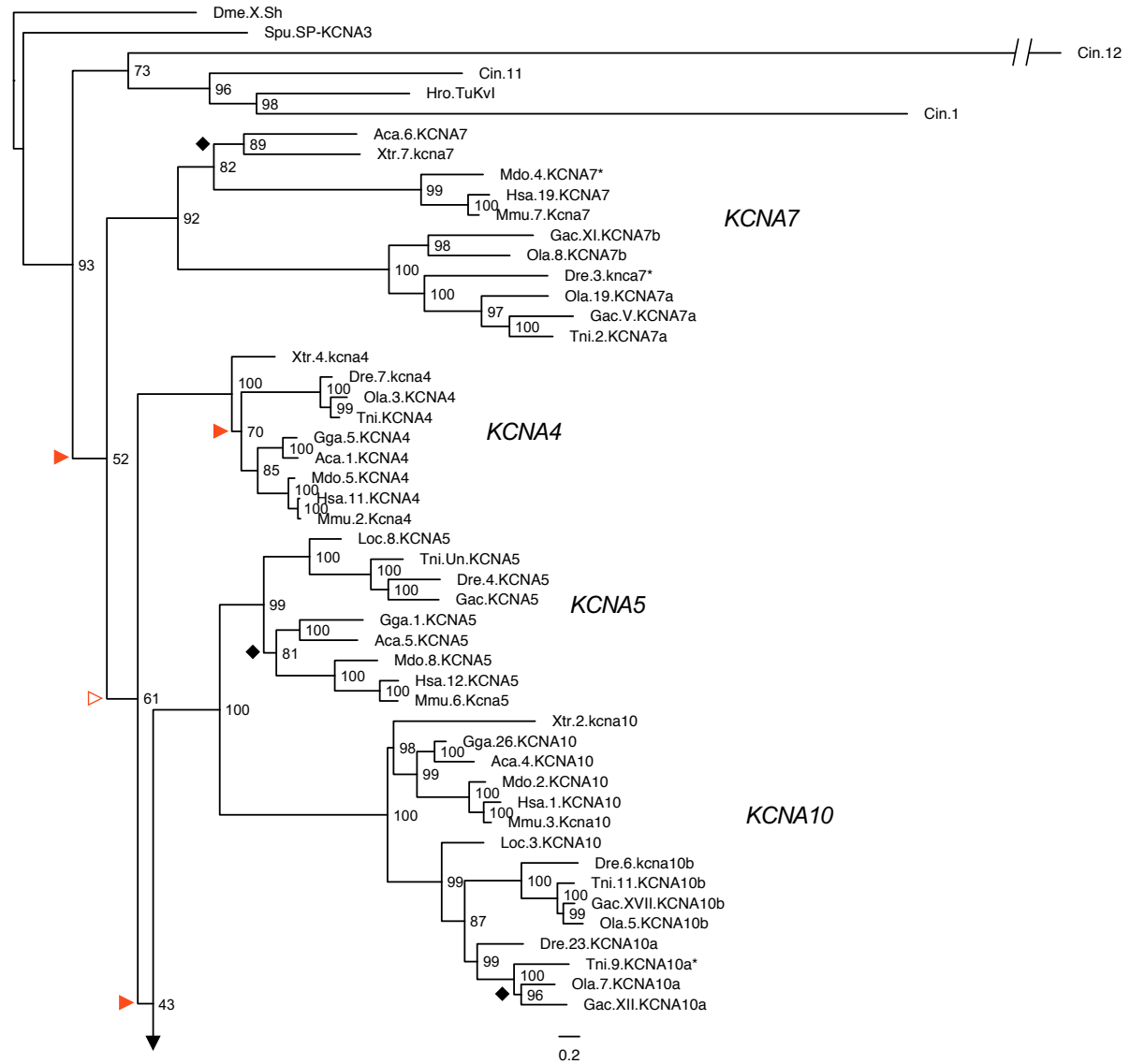


Supplementary figure S3. Reticulocalbin and calumenin family (CALU). The gene family is made up of the *CALU, RCN1* and *RCN3* genes. All three branches are well-supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). There are duplicate genes in the zebrafish genome (*calua* and *calub*), however the topology of the *CALU* branch is not well-defined and there is no support for a divergence in the time window of the basal teleost tetraploidization (3R).

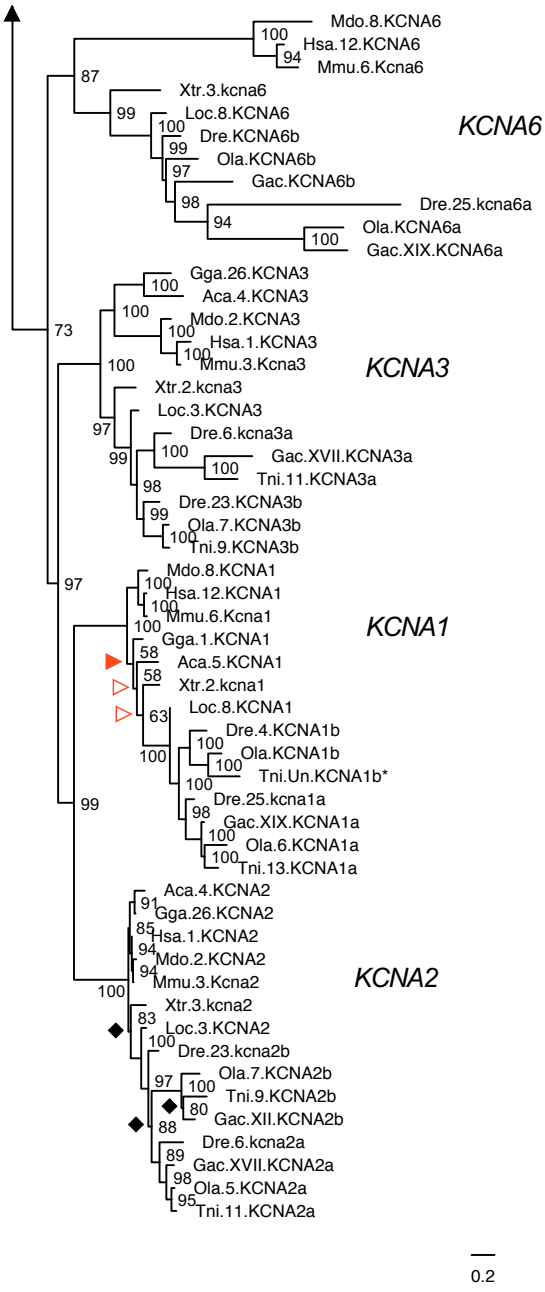


Supplementary figure S4. Voltage-gated potassium channel subfamily A (KCNA). The gene family is made up of the *KCNA1, -2, -3, -4, -5, -6, -7* and *-10* genes. It is also known as the Kv1 subfamily or the *Shaker*-related subfamily. All eight branches are well-supported and diverge in the time-window of the basal vertebrate tetraploidizations (1R/2R). The *KCNA1, -5, -6* genes, and *KCNA2, -3, -10* genes, form two syntenic clusters (located on chromosomes 12 and 1 respectively in the human genome). Our phylogenetic and conserved synteny analyses are consistent with an origin of these syntenic clusters through local duplications followed by the 2R tetraploidization. The *KCNA1, -2, -3, -6, -7* and *-10* branches have teleost duplicates diverging in the time window of the teleost tetraploidization (3R), however the *KCNA2b* and *KCNA6b* branches are not resolved in our phylogeny. These results are consistent with previous phylogenomic studies of the KCNA family (Hoegg and Meyer, 2007; Qiu et al., 2011). We could identify multiple member sequences in the vase tunicate *Ciona intestinalis*, as well as a sequence from the tunicate sea pineapple (*Halocynthia* roretzi, abbreviated Hro) named *TuKv1* (GenBank ID: BAA78383.1).

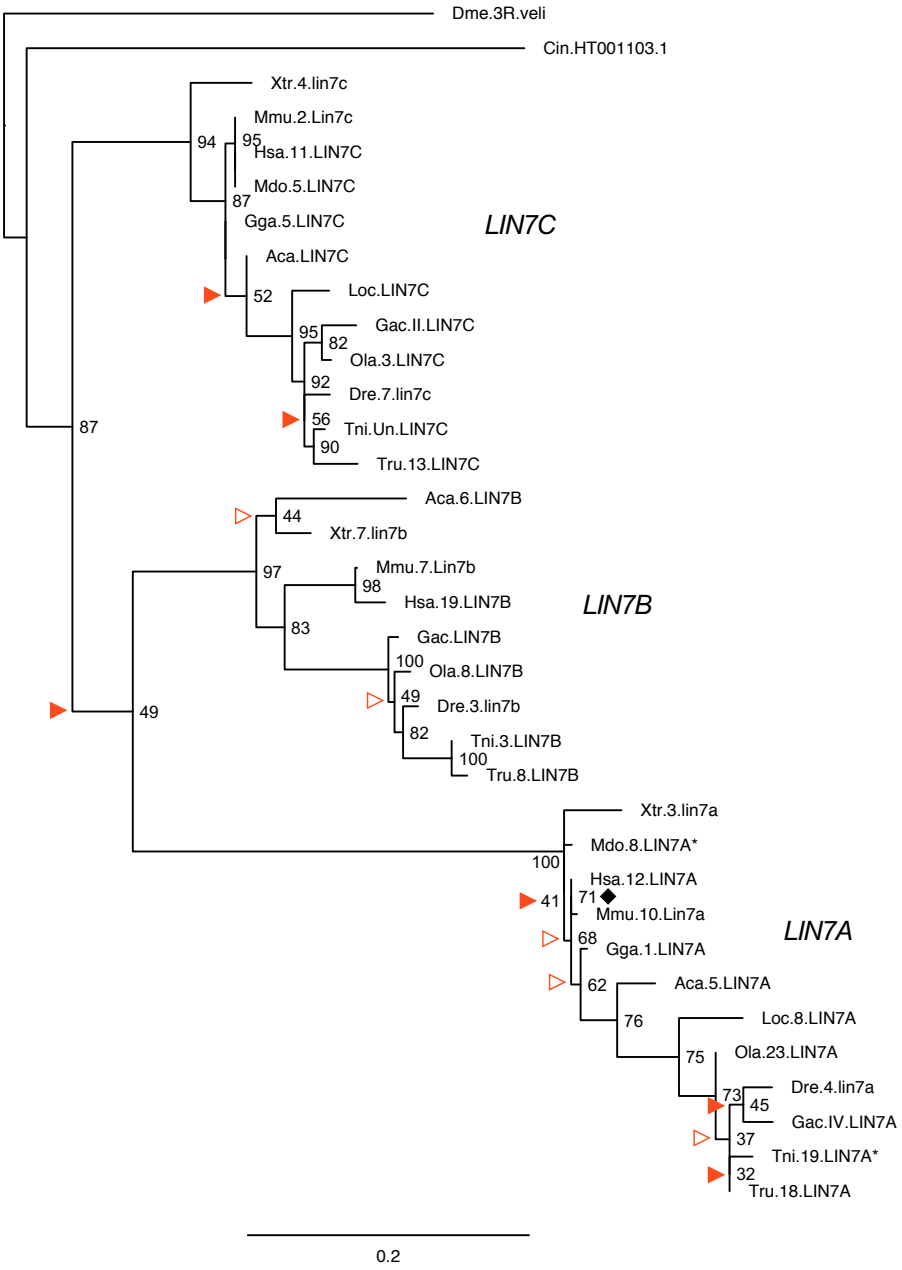
A.



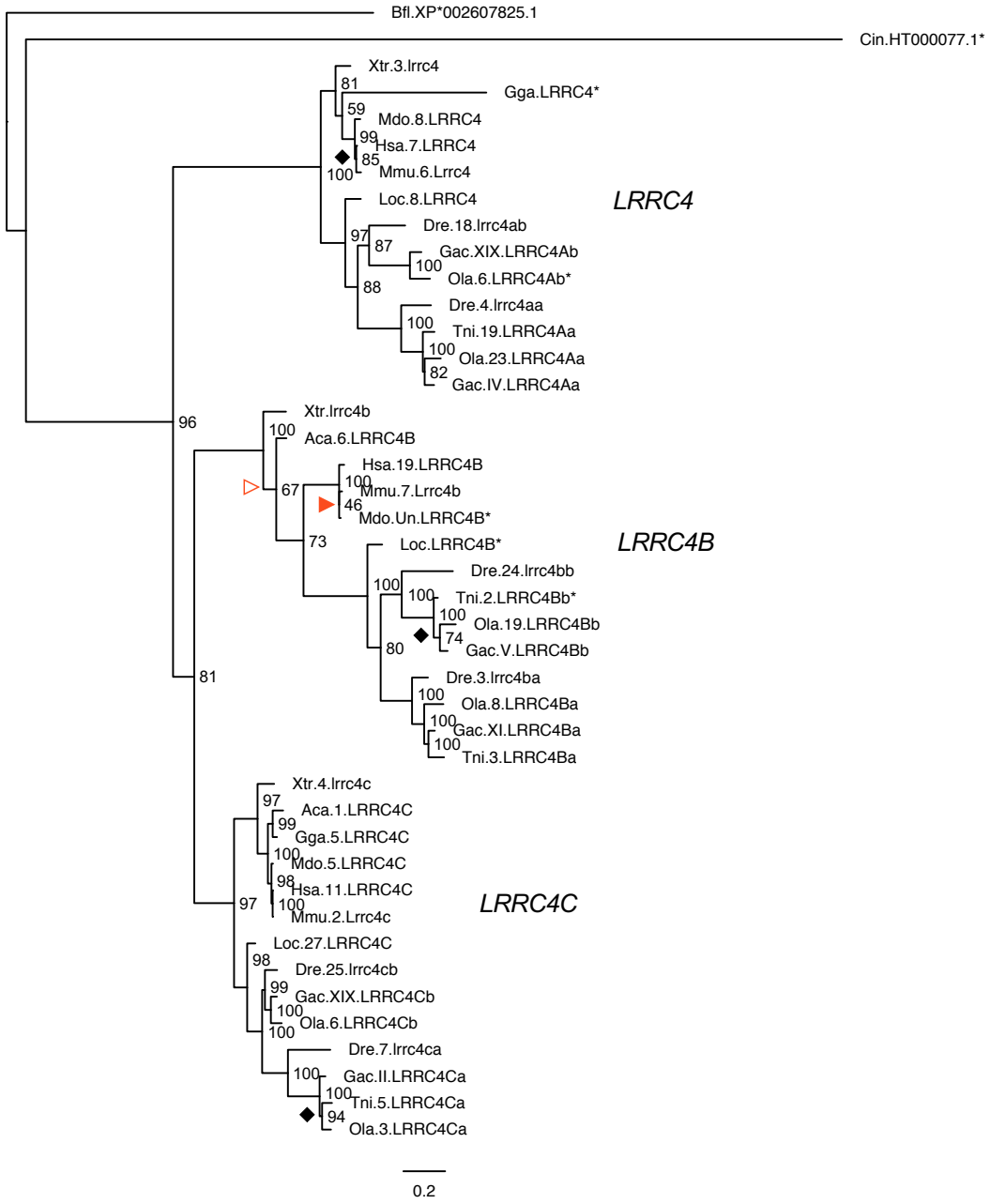
B.


Mdo.8.KCNA6
Hsa.12.KCNA6
Mmu.6.Kcna6
Xtr.3.kcna6
Loc.8.KCNA6
Dre.KCNA6b
Ola.KCNA6b
Gac.KCNA6b
Dre.25.kcna6a
Ola.KCNA6a
Gac.XIX.KCNA6a
KCNA6
Gga.26.KCNA3
Aca.4.KCNA3
Mdo.2.KCNA3
Hsa.1.KCNA3
Mmu.3.Kcna3
Xtr.2.kcna3
Loc.3.KCNA3
Dre.6.kcna3a
Gac.XVII.KCNA3a
Tni.11.KCNA3a
Dre.23.KCNA3b
Ola.7.KCNA3b
Tni.9.KCNA3b
KCNA3
Mdo.8.KCNA1
Hsa.12.KCNA1
Mmu.6.Kcna1
Gga.1.KCNA1
Aca.5.KCNA1
Xtr.2.kcna1
Loc.8.KCNA1
Dre.4.KCNA1b
Ola.KCNA1b
Tni.Un.KCNA1b*
Dre.25.kcna1a
Gac.XIX.KCNA1a
Ola.6.KCNA1a
Tni.13.KCNA1a
KCNA1
Aca.4.KCNA2
Gga.26.KCNA2
Hsa.1.KCNA2
Mdo.2.KCNA2
Mmu.3.Kcna2
Xtr.3.kcna2
Loc.3.KCNA2
Dre.23.kcna2b
Ola.7.KCNA2b
Tni.9.KCNA2b
Gac.XII.KCNA2b
Dre.6.kcna2a
Gac.XVII.KCNA2a
Ola.5.KCNA2a
Tni.11.KCNA2a
KCNA2
0.2

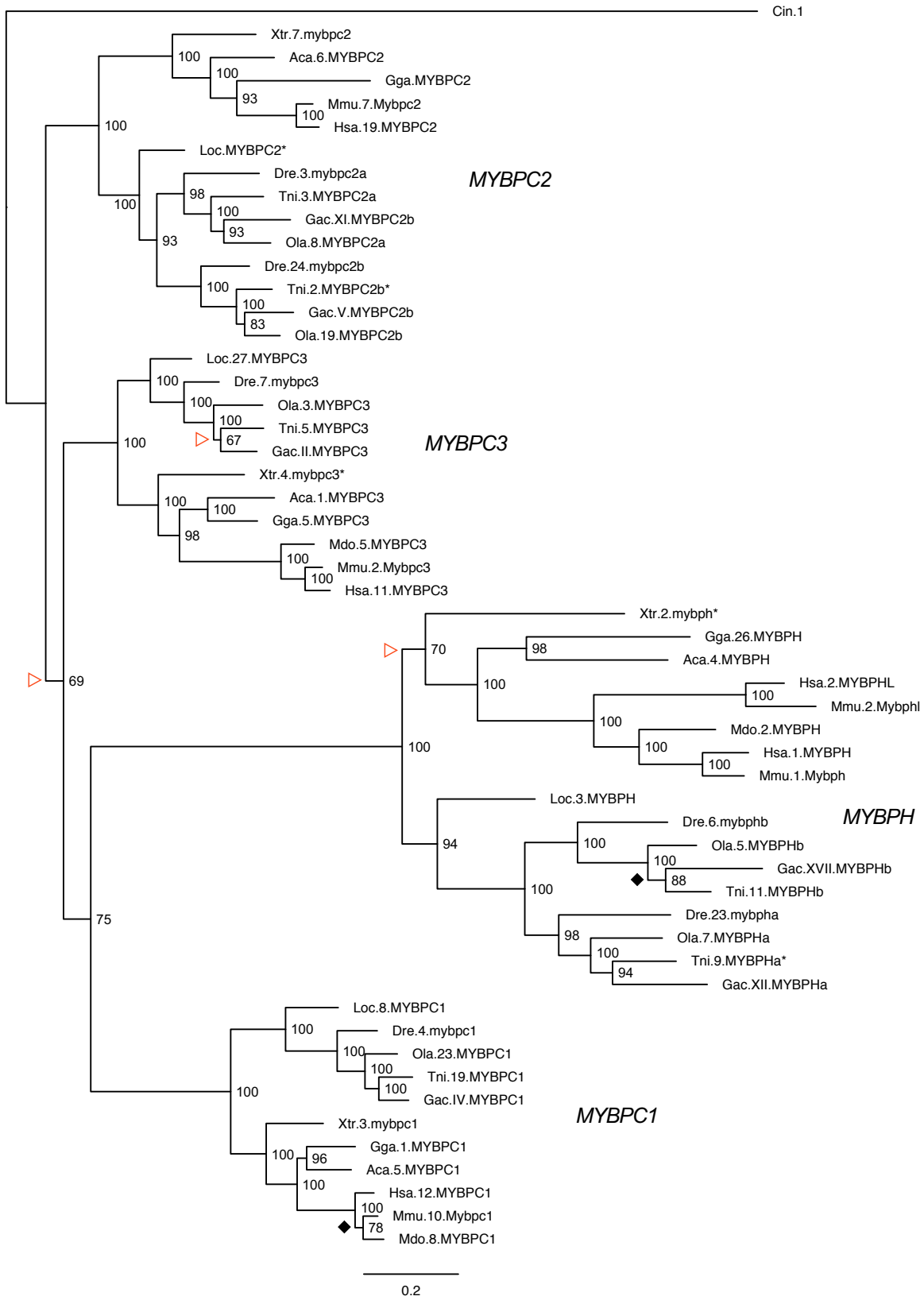
Supplementary figure S5. *Lin-7* homolog family (LIN7). The gene family is made up of the *LIN7A, LIN7B* and *LIN7C* genes. All three branches are well-supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). Due to the high degree of conservation and shortness of these sequences, the topologies within each branch diverge from the accepted species taxonomy.



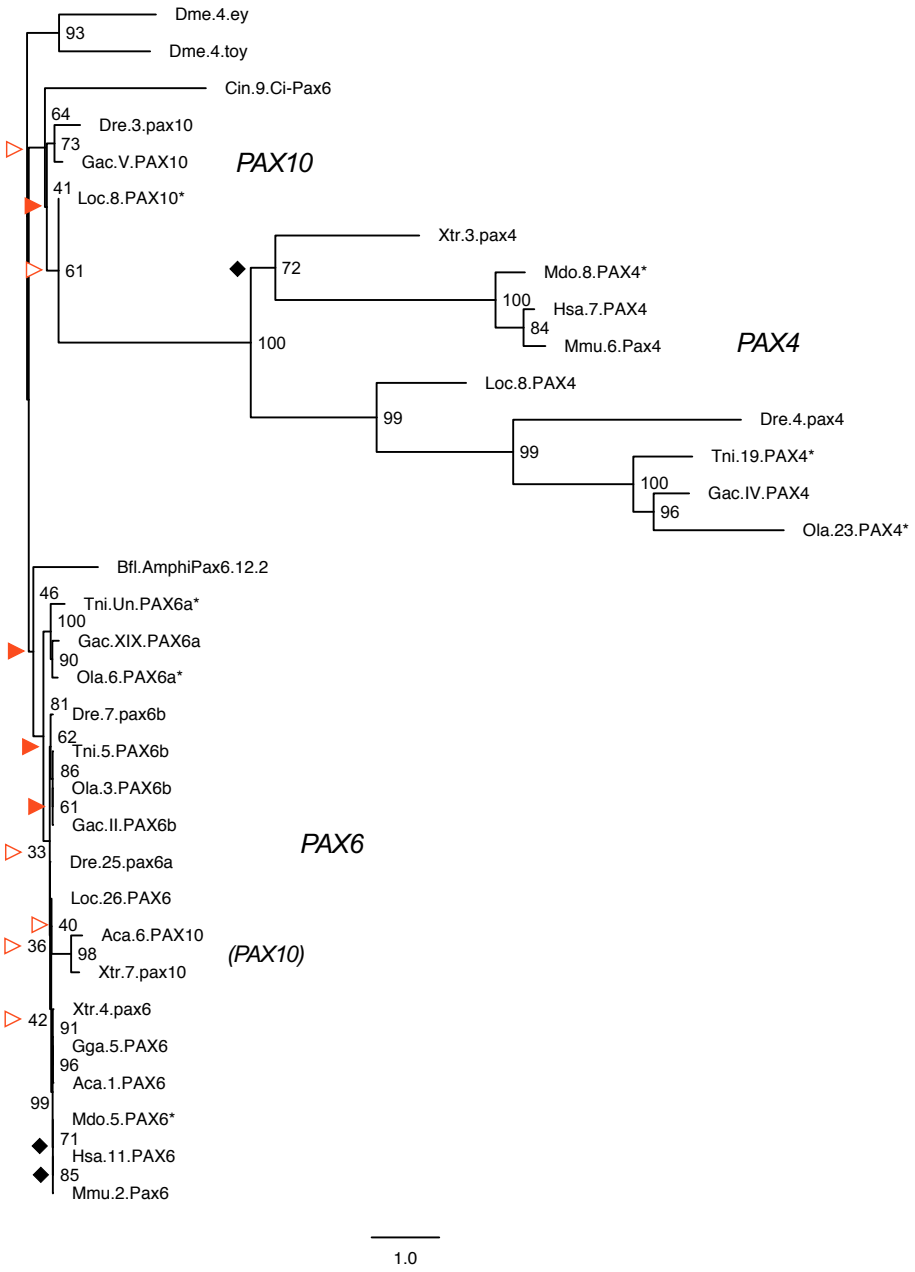
Supplementary figure S6. Leucine rich repeat containing 4 family (LRRC4). The gene family is made up of the *LRRC4* (*LRRC4A* in teleosts), *LRRC4B* and *LRRC4C* genes. All three branches are well supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). All three branches also have teleost duplicates diverging in the time window of the basal teleost tetraploidization (3R).



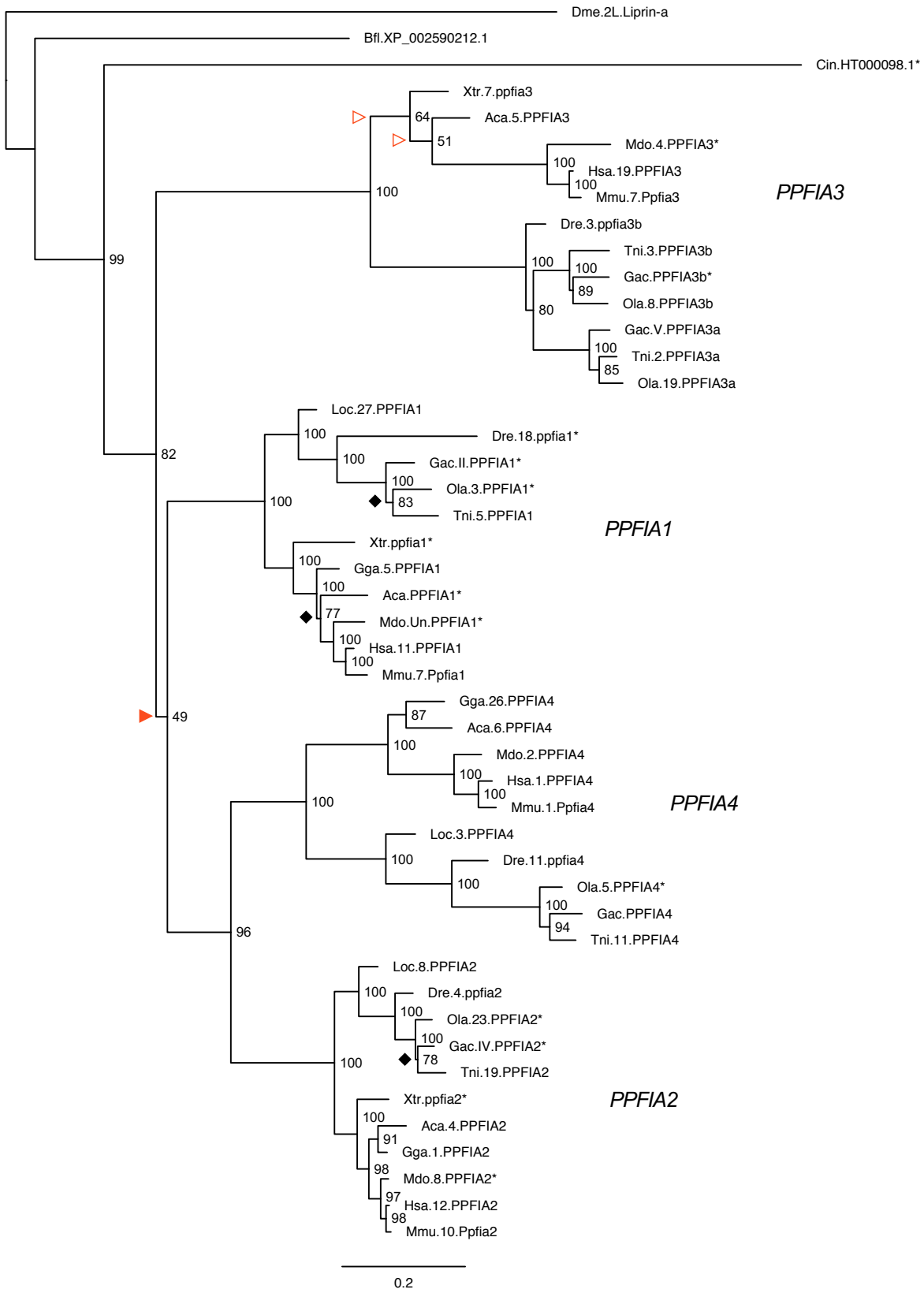
Supplementary figure S7. Myosin binding protein C family (MYBPC). (See following page.) The gene family is made up of the *MYBPC1, MYBPC2, MYBPC3* and *MYBPH* genes. All four branches are well supported, and the phylogeny is consistent with origins in the basal vertebrate tetraploidizations (1R/2R). However, since no other outgroup could be identified and the phylogeny is rooted with the vase tunicate sequence, the relative dating of the early divergences is uncertain. The *MYBPC2* and *MYBPH* branches have teleost duplicates diverging in the time window of the teleost tetraploidization 3R. The alignment used to create this phylogeny was edited to remove the N- and C-terminal ends, which could not be identified for many of the sequences. The final alignment consists of 1169 amino acid positions.


Cin.1
Xtr.7.mybpc2
Aca.6.MYBPC2
Gga.MYBPC2
Mmu.7.Mybpc2
Hsa.19.MYBPC2
Loc.MYBPC2*
Dre.3.mybpc2a
Tni.3.MYBPC2a
Gac.XI.MYBPC2b
Ola.8.MYBPC2a
Dre.24.mybpc2b
Tni.2.MYBPC2b*
Gac.V.MYBPC2b
Ola.19.MYBPC2b
MYBPC2
Loc.27.MYBPC3
Dre.7.mybpc3
Ola.3.MYBPC3
Tni.5.MYBPC3
Gac.II.MYBPC3
Xtr.4.mybpc3*
Aca.1.MYBPC3
Gga.5.MYBPC3
Mdo.5.MYBPC3
Mmu.2.Mybpc3
Hsa.11.MYBPC3
MYBPC3
Xtr.2.mybph*
Gga.26.MYBPH
Aca.4.MYBPH
Hsa.2.MYBPHL
Mmu.2.Mybphl
Mdo.2.MYBPH
Hsa.1.MYBPH
Mmu.1.Mybph
Loc.3.MYBPH
Dre.6.mybphb
Ola.5.MYBPHb
Gac.XVII.MYBPHb
Tni.11.MYBPHb
Dre.23.mybpha
Ola.7.MYBPHa
Tni.9.MYBPHa*
Gac.XII.MYBPHa
MYBPH
Loc.8.MYBPC1
Dre.4.mybpc1
Ola.23.MYBPC1
Tni.19.MYBPC1
Gac.IV.MYBPC1
Xtr.3.mybpc1
Gga.1.MYBPC1
Aca.5.MYBPC1
Hsa.12.MYBPC1
Mmu.10.Mybpc1
Mdo.8.MYBPC1
MYBPC1
0.2

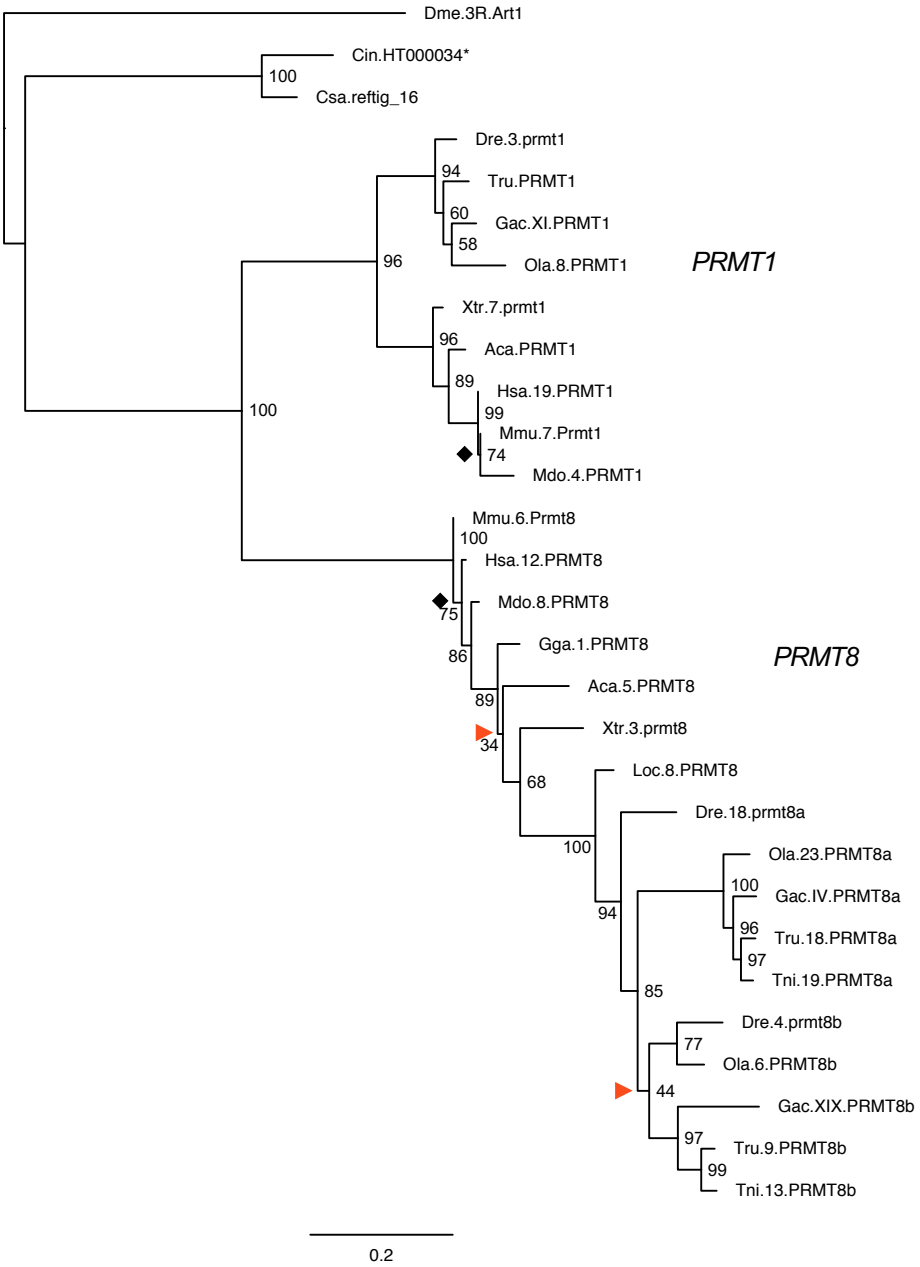
Supplementary figure S8. Paired box 4, -6 and -10 family (PAX4/6/10). The gene family is made up of the *PAX4, PAX6* and *PAX10* genes, a subfamily of the larger *paired box* gene family that includes the fruit fly *eyeless* (*ey*) and *twin of eyeless* (*toy*) genes. The phylogeny and evolution of this *paired box* subfamily has been studied previously, see for instance Wang et al., 2010 and Manousaki et al., 2011. Notably, in the study that first named the *PAX10* genes, the authors concluded that the three vertebrate genes had arisen in the basal vertebrate tetraploidizations (1R/2R), based on the chromosomal locations (Feiner et al., 2014). However, they also declared that their phylogenetic analysis did not support this conclusion. Our phylogeny was rooted with the fruit fly *ey* and *toy* genes, and also includes the vase tunicate *Ci-Pax6* (Irvine et al., 2008) and Florida lancelet *AmphiPax6* (Glardon et al., 1998) sequences. If the weakly supported tunicate and lancelet branches are collapsed, our phylogeny supports the origin of *PAX4* and *PAX6* in 1R/2R. However, the *PAX10* branch is not resolved. This is likely due to the uneven evolutionary rates within this gene family. There are duplicate *PAX6* genes in teleost fishes, however the time window for this divergence is not resolved in our phylogeny, likely due to the relatively slow rate of evolution for *PAX6*.



Supplementary figure S9. PTPRF-interacting protein alpha (PPFIA). (See following page.) The gene family is made up of the *PPFIA1, -2, -3* and *-4* genes. All three branches are well supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). There are *PPFIA3* duplicates in teleost fishes. Our phylogeny suggests that these duplicates arose within teleosts, after the divergence of the lineage leading to zebrafish, however the chromosomal locations are consistent with origins in the basal teleost tetraploidization (3R) – See Fig. 4. Accordingly, we have renamed the single *PPFIA3* gene in zebrafish *ppfia3b*.

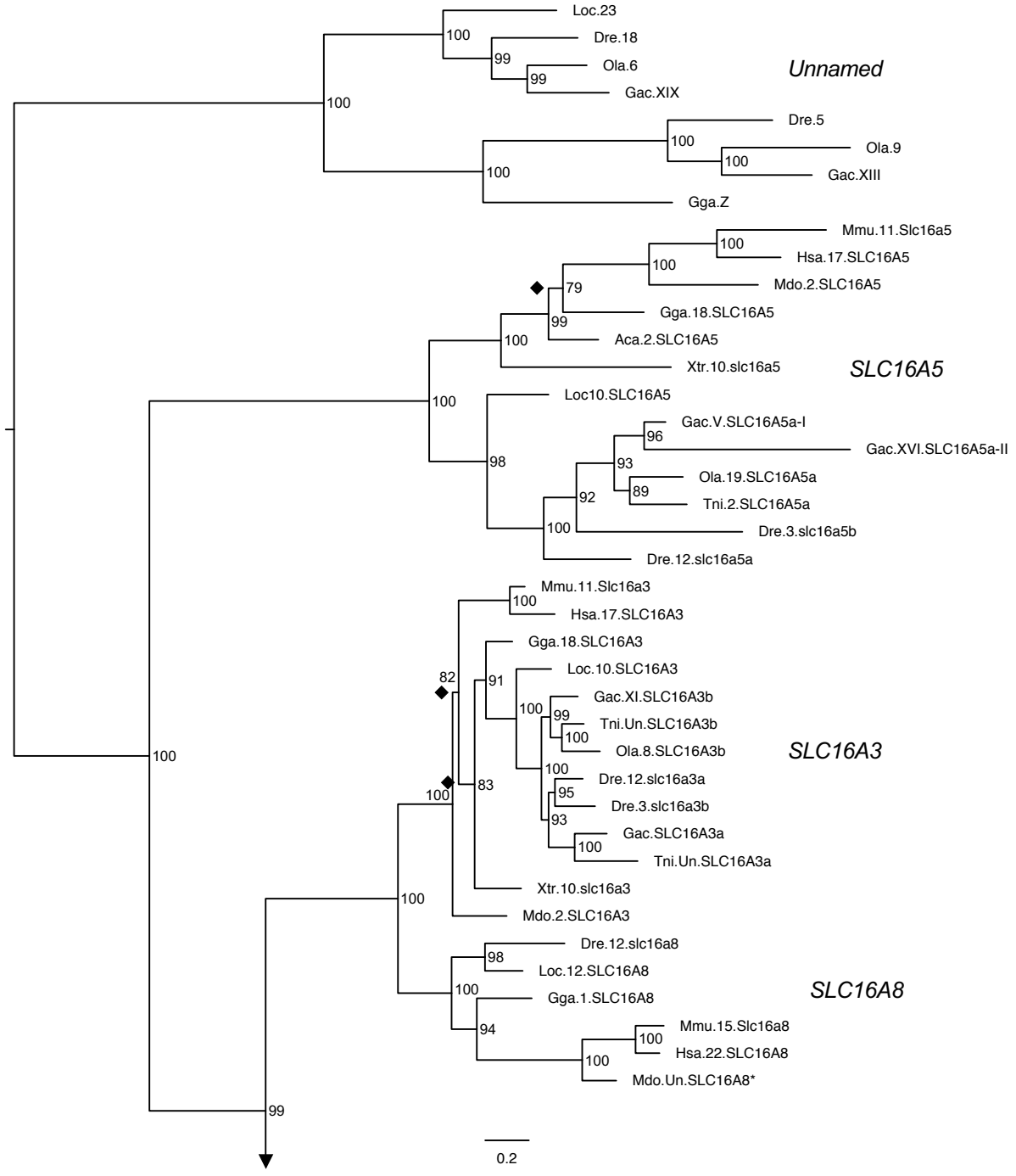

Dme.2L.Liprin-a
Bfl.XP_002590212.1
Cin.HT000098.1*
Xtr.7.ppfia3
Aca.5.PPFIA3
Mdo.4.PPFIA3*
Hsa.19.PPFIA3
Mmu.7.Ppfia3
PPFIA3
Dre.3.ppfia3b
Tni.3.PPFIA3b
Gac.PPFIA3b*
Ola.8.PPFIA3b
Gac.V.PPFIA3a
Tni.2.PPFIA3a
Ola.19.PPFIA3a
Loc.27.PPFIA1
Dre.18.ppfia1*
Gac.II.PPFIA1*
Ola.3.PPFIA1*
Tni.5.PPFIA1
PPFIA1
Xtr.ppfia1*
Gga.5.PPFIA1
Aca.PPFIA1*
Mdo.Un.PPFIA1*
Hsa.11.PPFIA1
Mmu.7.Ppfia1
Gga.26.PPFIA4
Aca.6.PPFIA4
Mdo.2.PPFIA4
Hsa.1.PPFIA4
Mmu.1.Ppfia4
PPFIA4
Loc.3.PPFIA4
Dre.11.ppfia4
Ola.5.PPFIA4*
Gac.PPFIA4
Tni.11.PPFIA4
Loc.8.PPFIA2
Dre.4.ppfia2
Ola.23.PPFIA2*
Gac.IV.PPFIA2*
Tni.19.PPFIA2
Xtr.ppfia2*
Aca.4.PPFIA2
Gga.1.PPFIA2
Mdo.8.PPFIA2*
Hsa.12.PPFIA2
Mmu.10.Ppfia2
PPFIA2
0.2

Supplementary figure S10. Protein arginine methyltransferase 1 and -8 (PRMT1/8). The family is made up of the *PRMT1* and *PRMT8* genes. Both branches are well-supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). The tetrapod *PRMT8* branch is not resolved, however this does not affect our conclusions. There are *PRMT8* duplicates in teleosts. While the topology of these duplicate branches is not resolved, the phylogeny is consistent with origins in the time window of the basal teleost tetraploidization (3R).



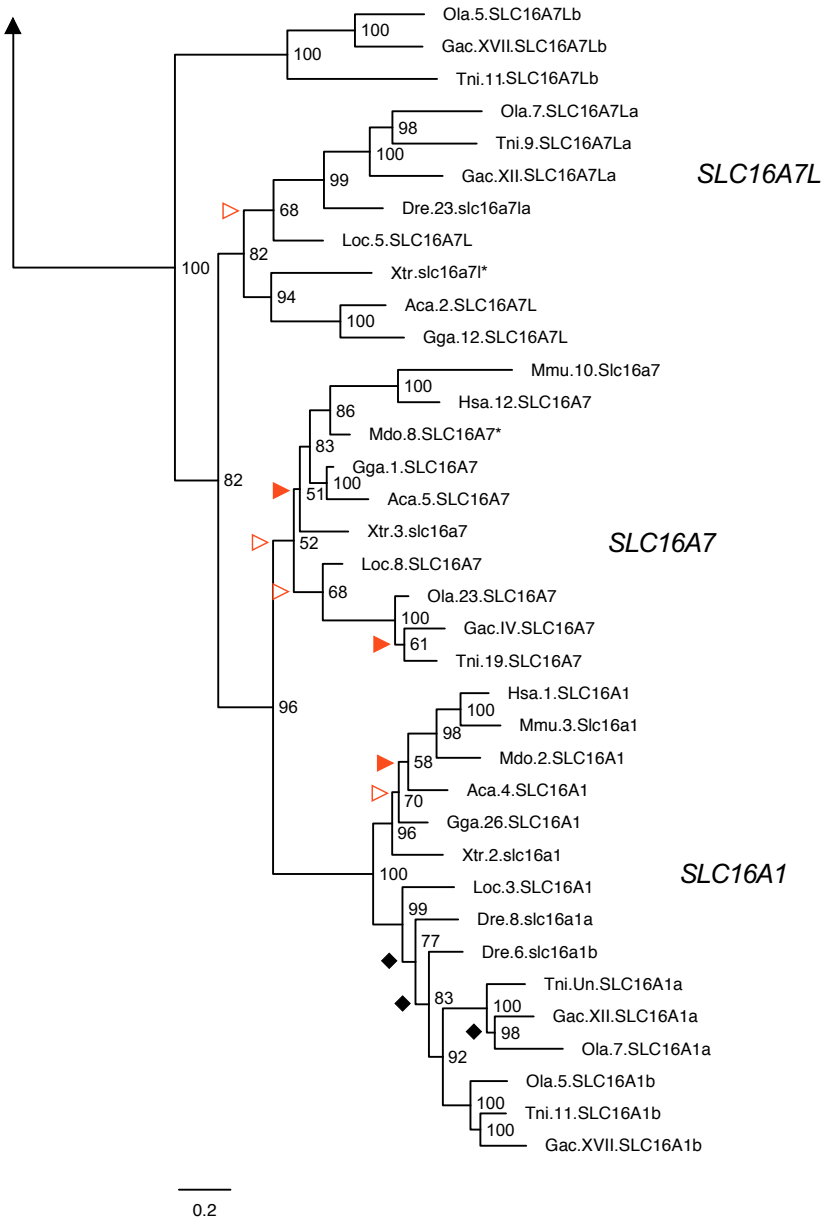
Supplementary figure S11. Solute carrier family 16, subfamily 1 (SLC16A1). This subfamily of the larger SLC16A family is made up of the *SLC16A1, -3, -5, -7* and *-8* genes, as well as two clades of unnamed genes. We have named one of these clades *SLC16A7L* following the accepted name for the chicken gene in this branch, *MCT2L* (*MCT2* is a synonym of *SLC16A7*). All branches, except the *SLC16A7L* branch, are well-supported. No outgroups could be identified (Liu et al., 2008), and the phylogeny is displayed as a midpoint-rooted tree. Thus, it is not possible to infer a relative time window for the origin of these genes from the phylogeny alone. The chromosomal locations of the *SLC16A1, -3, -5, -7* and *-8* genes suggest that they arose through a combination of local gene duplications and the basal vertebrate tetraploidizations (1R/2R). (Continued on next page.)

A.



There are putative duplicates of the *SLC16A7L* gene in teleost fishes. One such clade clusters basal to the other *SLC16A7L,* and the *SLC16A7* and *SLC16A1* branches, however their chromosomal locations are consistent with the duplication of an *SLC16A7L* gene in the basal teleost tetraploidization (see Figs. S3 and S4). The chromosomal locations of these genes in the medaka, three-spined stickleback and green spotted pufferfish also suggest that the *SLC16A7L* gene might have arisen through a local gene duplication that also gave rise to an ancestral *SLC16A1/7* gene before 1R/2R. There are also likely 3R-generated duplicates of *SLC16A1, SLC15A3* (with good phylogenetic support), and *SLC15A5*. Finally, the origin of the basal-most unnamed clade of genes remains unknown. These genes seem to be located on completely unrelated chromosome regions (see Figs. S2 – S4).

B.

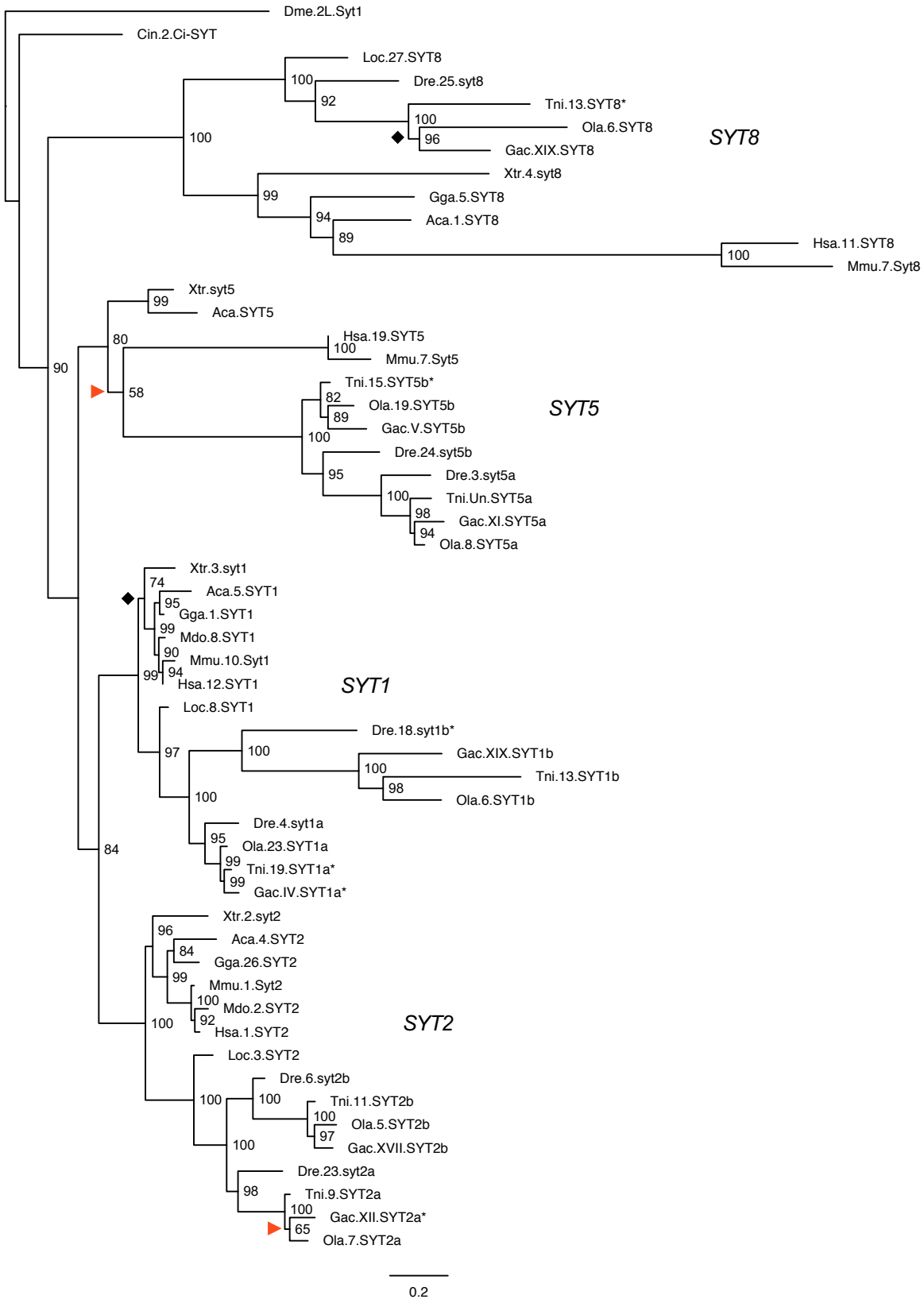


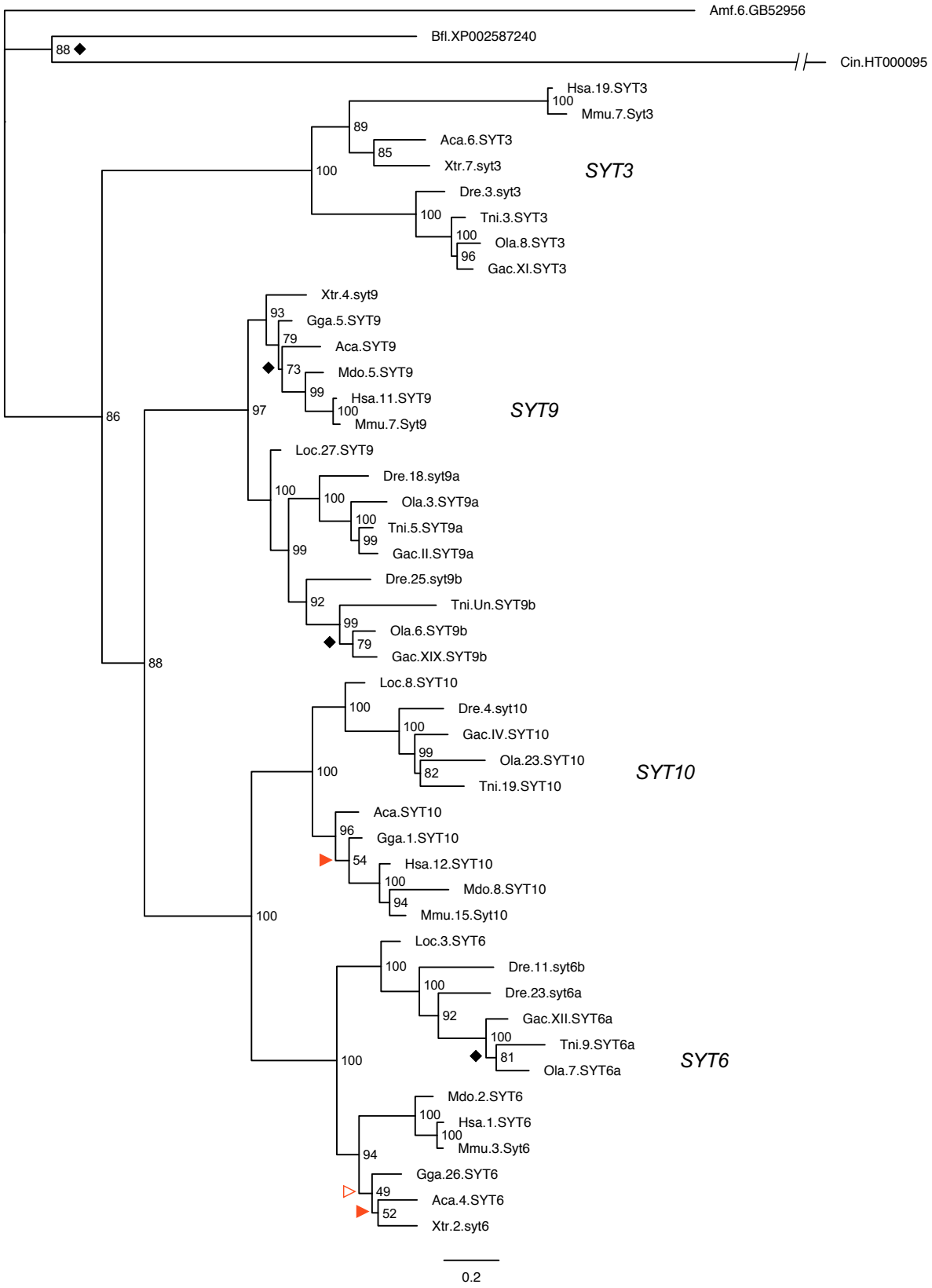
Supplementary figure S12. Synaptotagmin 1, -2, -5 and -8 family (SYT1/2/5/8). (See following page.) The gene family is made up of the *SYT1, SYT2, SYT5* and *SYT8* genes. All four branches are well supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). The *SYT1* and *-2* branches have teleost duplicates diverging in the time window of the basal teleost tetraploidization (3R). There are likely 3R-generated duplicates of *SYT5* as well, however the topology of this branch is less clear.

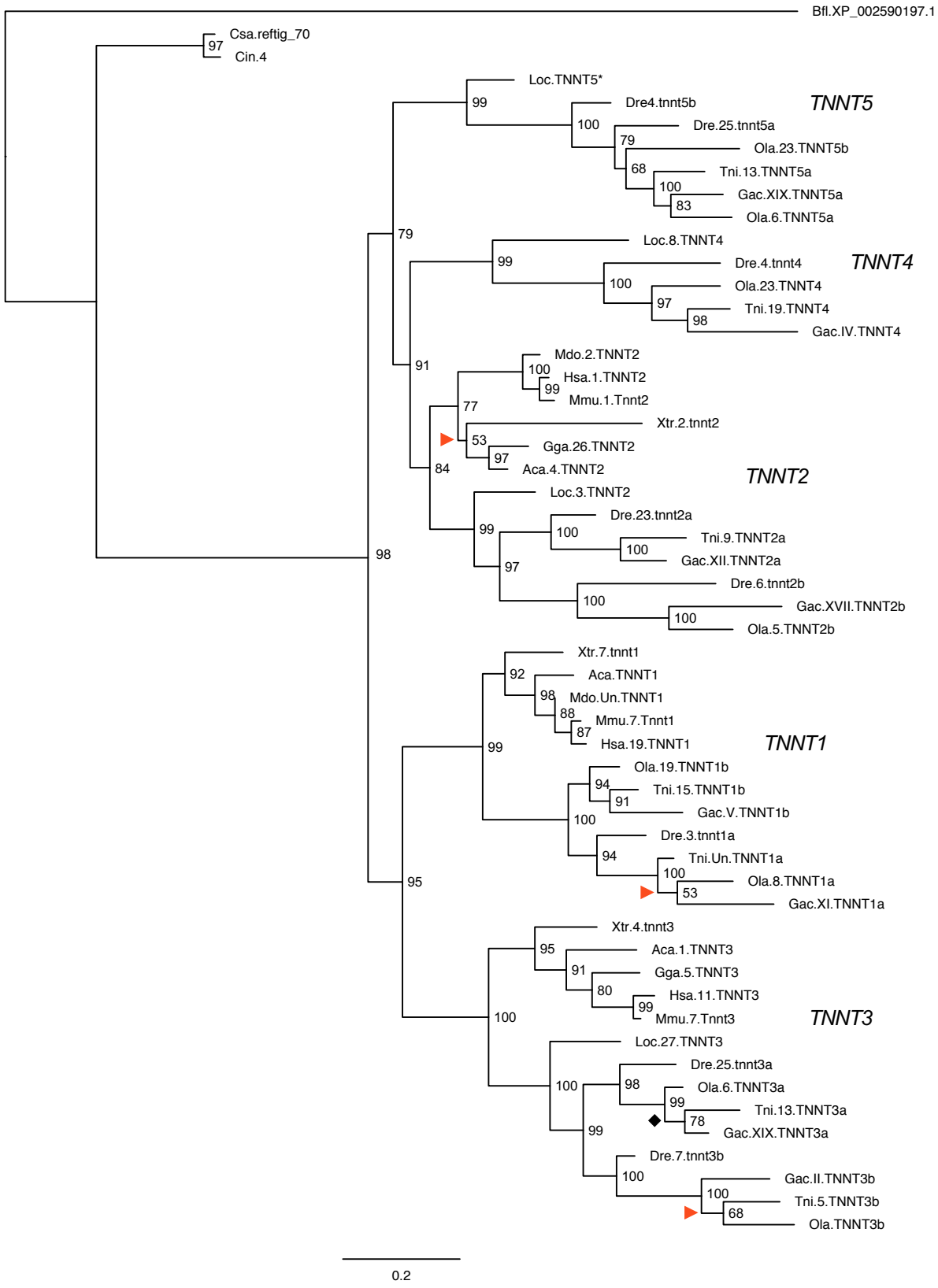
Supplementary figure S13. Synaptotagmin -3, -6, -9 and -10 family (SYT3/6/9/10). (See page 18.) The gene family is made up of the *SYT3, SYT6, SYT9* and *SYT10* genes. All four branches are well supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). The *SYT6* and *-9* branches have teleost duplicates diverging in the time window of the basal teleost tetraploidization (3R), however *SYT6b* could only be identified in the zebrafish. The phylogeny is rooted with the honey bee *GB52956* sequence, based on a previously published study (see Additional file 8 in Craxton, 2010). The branch of the vase tunicate sequence has been shortened for clarity.

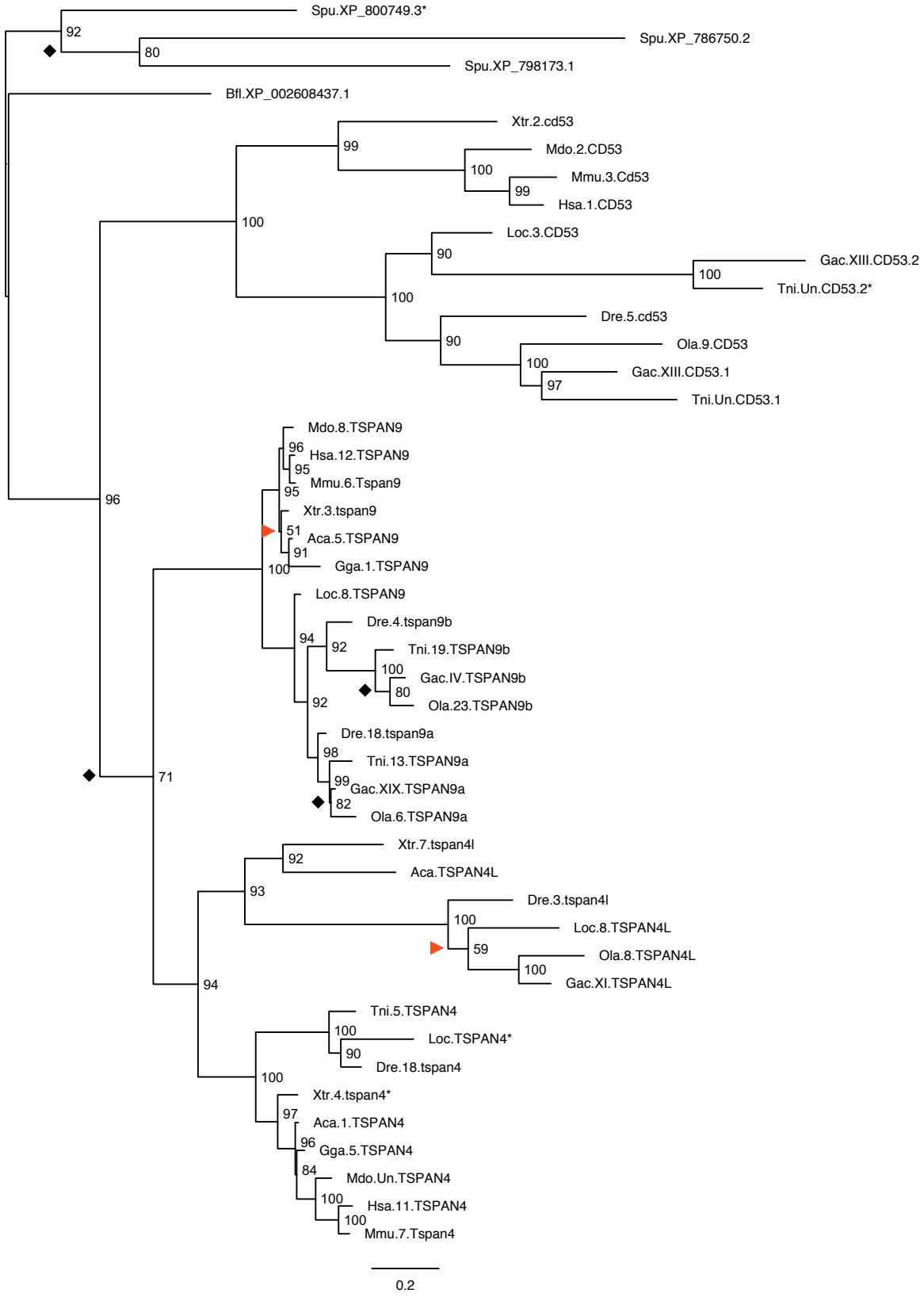
Supplementary figure S14. Troponin T family (TNNT). (See page 19.) The gene family is made up of the *TNNT1, -2* and *-3* genes, as well as two clades of previously undescribed troponin T genes we have named *TNNT4* and *-5*. The *TNNT1, -2, -3* branches are well supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). The *TNNT4* and *-5* branches also diverge in this time window, and the chromosomal locations of the *TNNT4* genes suggest that they make up the fourth branch generated in 1R/2R. However, the *TNNT4* genes either lack introns (spotted gar, zebrafish and medaka) or have an exon/intron structure that diverges from that of the *TNNT1, -2* and *-3* genes (three-spined stickleback and green spotted pufferfish). This suggests a different origin: through the retrotransposon-mediated duplication of a related gene, likely *TNNT2*. There is also some support that the *TNNT5* branch makes up the fourth 1R/2R branch, however the chromosomal locations are ambiguous in this regard (see Fig. 4). The *TNNT1, -2, -3* and *-5* branches have teleost duplicates that diverge in the time window of the basal teleost tetraploidization (3R). The *TNNT4* and *TNNT5* clades have not been fully described previously. A study focusing on zebrafish identified the *tnnt4, tnnt5a* and *tnnt5b* genes and named them *tnnt2c, -2d* and *-2e*, respectively (Ferrante et al., 2011). We propose a change of the nomenclature suggested by Ferrante et al., naming these genes *TNNT4* and *TNNT5* to better reflect the evolutionary scenario for this family, and to avoid confusion with the 3R-generated duplicates *TNNT2a* and *-2b*. The phylogeny is rooted with a Florida lancelet sequence identified through BLAST. We also found two putative purple sea urchin family members, *TropT* and *Trop1*; however, these sequences produced very long outgroup branches when used to root preliminary phylogenies. This indicates a high degree of divergence from the chordate *TNNT* sequences. Thus, the purple sea urchin *TropT* and *Trop1* sequences were not used to root this phylogeny. The alignment used to create this phylogeny was edited to remove the N-terminal encoding exons, which could not be identified for many of the sequences. The final alignment consists of 246 amino acid positions.

Supplementary figure S15. Tetraspanin 4, -9 and -25 (CD53) family (TSPAN4/6/25). (See page 20.) The gene family is made up of the *TSPAN4, TSPAN4L, TSPAN9* and *TSPAN25* genes. *TSPAN25* is a synonym of *CD53*, which is the accepted gene name in most species. All four branches are well supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). Our phylogeny recapitulates the evolutionary relationships demonstrated by a previously published study of the *Tetraspanin* superfamily in vertebrates (Huang et al., 2010). The *TSPAN9* branch has teleost duplicates that diverge in the time window of the basal teleost tetraploidization (3R). In the three-spined stickleback and green spotted pufferfish, there are local duplicates of *CD53,* which we have called *CD53.1* and *CD53.2*.

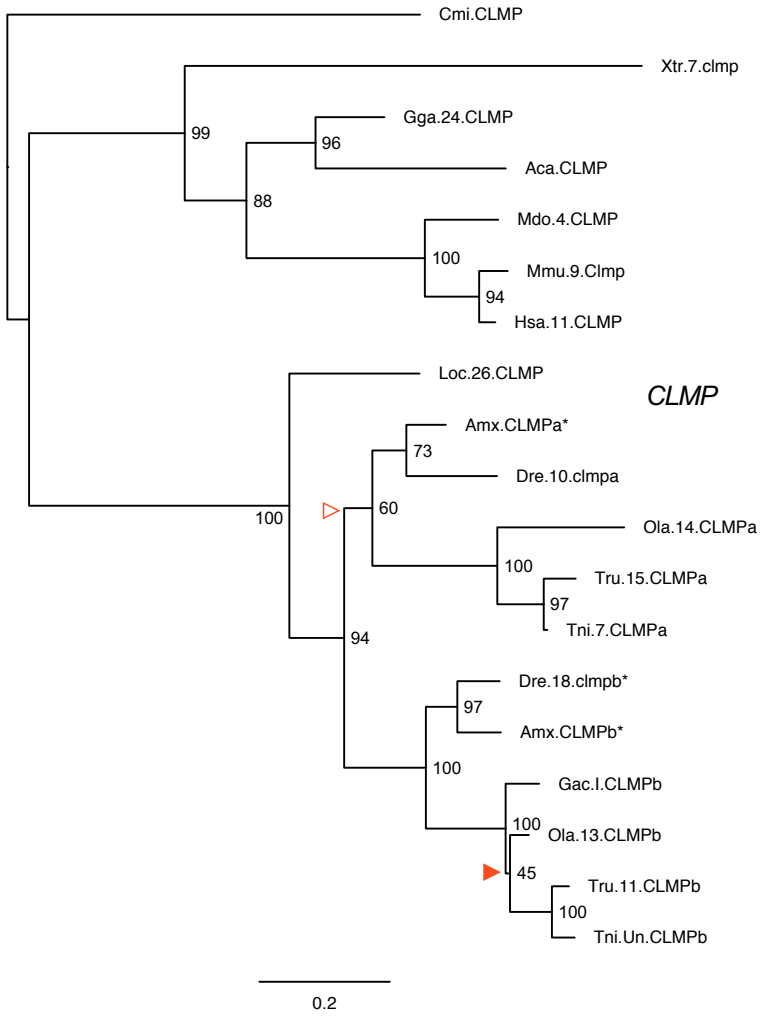

Dme.2L.Syt1
Cin.2.Ci-SYT
Loc.27.SYT8
Dre.25.syt8
Tni.13.SYT8*
Ola.6.SYT8
Gac.XIX.SYT8
SYT8
Xtr.4.syt8
Gga.5.SYT8
Aca.1.SYT8
Hsa.11.SYT8
Mmu.7.Syt8
Xtr.syt5
Aca.SYT5
Hsa.19.SYT5
Mmu.7.Syt5
Tni.15.SYT5b*
Ola.19.SYT5b
Gac.V.SYT5b
SYT5
Dre.24.syt5b
Dre.3.syt5a
Tni.Un.SYT5a
Gac.XI.SYT5a
Ola.8.SYT5a
Xtr.3.syt1
Aca.5.SYT1
Gga.1.SYT1
Mdo.8.SYT1
Mmu.10.Syt1
Hsa.12.SYT1
SYT1
Loc.8.SYT1
Dre.18.syt1b*
Gac.XIX.SYT1b
Tni.13.SYT1b
Ola.6.SYT1b
Dre.4.syt1a
Ola.23.SYT1a
Tni.19.SYT1a*
Gac.IV.SYT1a*
Xtr.2.syt2
Aca.4.SYT2
Gga.26.SYT2
Mmu.1.Syt2
Mdo.2.SYT2
Hsa.1.SYT2
SYT2
Loc.3.SYT2
Dre.6.syt2b
Tni.11.SYT2b
Ola.5.SYT2b
Gac.XVII.SYT2b
Dre.23.syt2a
Tni.9.SYT2a
Gac.XII.SYT2a*
Ola.7.SYT2a
0.2


Amf.6.GB52956
Bfl.XP002587240
Cin.HT000095
Hsa.19.SYT3
Mmu.7.Syt3
Aca.6.SYT3
Xtr.7.syt3
SYT3
Dre.3.syt3
Tni.3.SYT3
Ola.8.SYT3
Gac.XI.SYT3
Xtr.4.syt9
Gga.5.SYT9
Aca.SYT9
Mdo.5.SYT9
Hsa.11.SYT9
Mmu.7.Syt9
SYT9
Loc.27.SYT9
Dre.18.syt9a
Ola.3.SYT9a
Tni.5.SYT9a
Gac.II.SYT9a
Dre.25.syt9b
Tni.Un.SYT9b
Ola.6.SYT9b
Gac.XIX.SYT9b
Loc.8.SYT10
Dre.4.syt10
Gac.IV.SYT10
Ola.23.SYT10
Tni.19.SYT10
SYT10
Aca.SYT10
Gga.1.SYT10
Hsa.12.SYT10
Mdo.8.SYT10
Mmu.15.Syt10
Loc.3.SYT6
Dre.11.syt6b
Dre.23.syt6a
Gac.XII.SYT6a
Tni.9.SYT6a
Ola.7.SYT6a
SYT6
Mdo.2.SYT6
Hsa.1.SYT6
Mmu.3.Syt6
Gga.26.SYT6
Aca.4.SYT6
Xtr.2.syt6
0.2

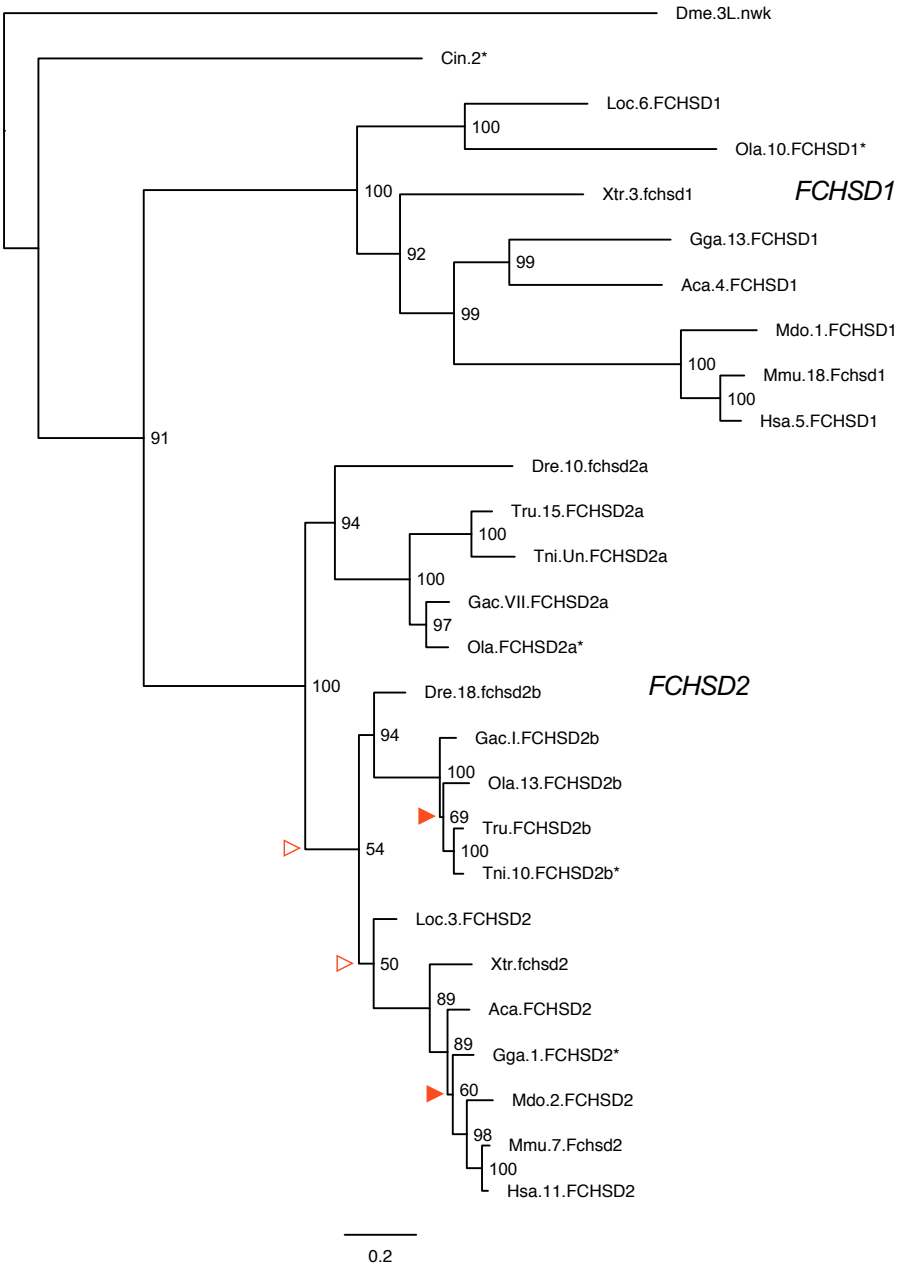

Bfl.XP_002590197.1
Csa.reftig_70
Cin.4
Loc.TNNT5*
Dre4.tnnt5b
Dre.25.tnnt5a
Ola.23.TNNT5b
Tni.13.TNNT5a
Gac.XIX.TNNT5a
Ola.6.TNNT5a
TNNT5
Loc.8.TNNT4
Dre.4.tnnt4
Ola.23.TNNT4
Tni.19.TNNT4
Gac.IV.TNNT4
TNNT4
Mdo.2.TNNT2
Hsa.1.TNNT2
Mmu.1.Tnnt2
Xtr.2.tnnt2
Gga.26.TNNT2
Aca.4.TNNT2
Loc.3.TNNT2
Dre.23.tnnt2a
Tni.9.TNNT2a
Gac.XII.TNNT2a
Dre.6.tnnt2b
Gac.XVII.TNNT2b
Ola.5.TNNT2b
TNNT2
Xtr.7.tnnt1
Aca.TNNT1
Mdo.Un.TNNT1
Mmu.7.Tnnt1
Hsa.19.TNNT1
Ola.19.TNNT1b
Tni.15.TNNT1b
Gac.V.TNNT1b
Dre.3.tnnt1a
Tni.Un.TNNT1a
Ola.8.TNNT1a
Gac.XI.TNNT1a
TNNT1
Xtr.4.tnnt3
Aca.1.TNNT3
Gga.5.TNNT3
Hsa.11.TNNT3
Mmu.7.Tnnt3
Loc.27.TNNT3
Dre.25.tnnt3a
Ola.6.TNNT3a
Tni.13.TNNT3a
Gac.XIX.TNNT3a
Dre.7.tnnt3b
Gac.II.TNNT3b
Tni.5.TNNT3b
Ola.TNNT3b
TNNT3
0.2


Spu.XP_800749.3*
92
Spu.XP_786750.2
80
Spu.XP_798173.1
Bfl.XP_002608437.1
Xtr.2.cd53
99
Mdo.2.CD53
100
Mmu.3.Cd53
99
Hsa.1.CD53
100
Loc.3.CD53
90
Gac.XIII.CD53.2
100
Tni.Un.CD53.2*
100
Dre.5.cd53
90
Ola.9.CD53
100
Gac.XIII.CD53.1
97
Tni.Un.CD53.1
Mdo.8.TSPAN9
96
Hsa.12.TSPAN9
95
Mmu.6.Tspan9
95
Xtr.3.tspan9
51
Aca.5.TSPAN9
91
Gga.1.TSPAN9
100
Loc.8.TSPAN9
Dre.4.tspan9b
94
92
Tni.19.TSPAN9b
100
Gac.IV.TSPAN9b
80
Ola.23.TSPAN9b
92
Dre.18.tspan9a
98
Tni.13.TSPAN9a
99
Gac.XIX.TSPAN9a
82
Ola.6.TSPAN9a
96
71
Xtr.7.tspan4l
92
Aca.TSPAN4L
93
Dre.3.tspan4l
100
Loc.8.TSPAN4L
59
Ola.8.TSPAN4L
100
Gac.XI.TSPAN4L
94
Tni.5.TSPAN4
100
Loc.TSPAN4*
90
Dre.18.tspan4
100
Xtr.4.tspan4*
97
Aca.1.TSPAN4
96
Gga.5.TSPAN4
84
Mdo.Un.TSPAN4
100
Hsa.11.TSPAN4
100
Mmu.7.Tspan4
0.2

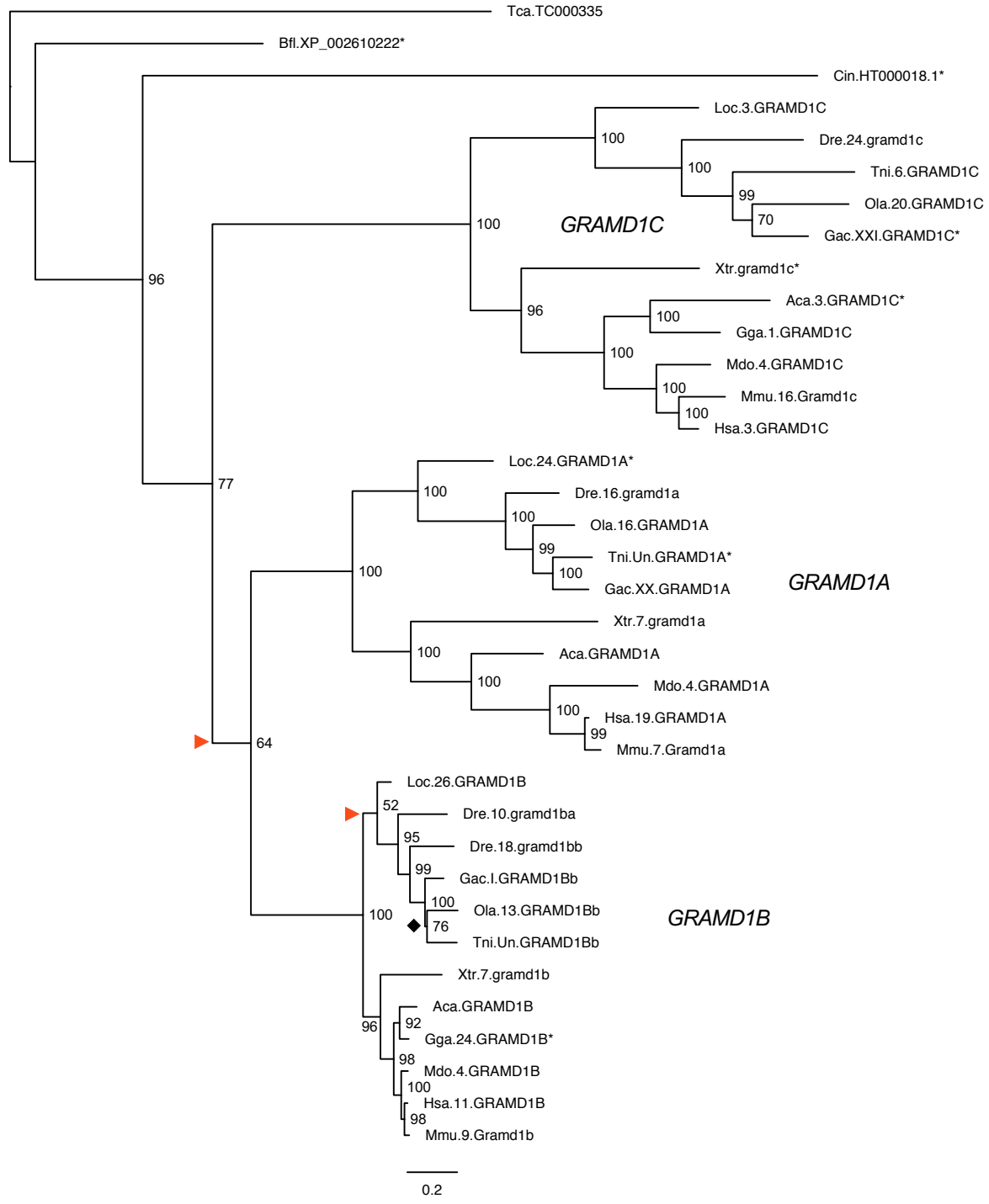
Supplementary figure S16. Coxsackie- and adenovirus receptor-like membrane protein (CLMP). The gene family consist of only one branch in jawed vertebrates, made up by the *CLMP* genes. No putative invertebrate family members could be identified, and the phylogeny is rooted by the elephant shark *CLMP* sequence. The duplicate *CLMPa* and *-b* branches in teleost fishes diverge in the time window of the basal teleost tetraploidization (3R). The *CLMPa* branch has somewhat low support in the UFBoot test. The alignment used to create this phylogeny was edited to remove the short N-terminal encoding exon 1, which could not be identified for many of the sequences. The final alignment consists of 389 amino acid positions.



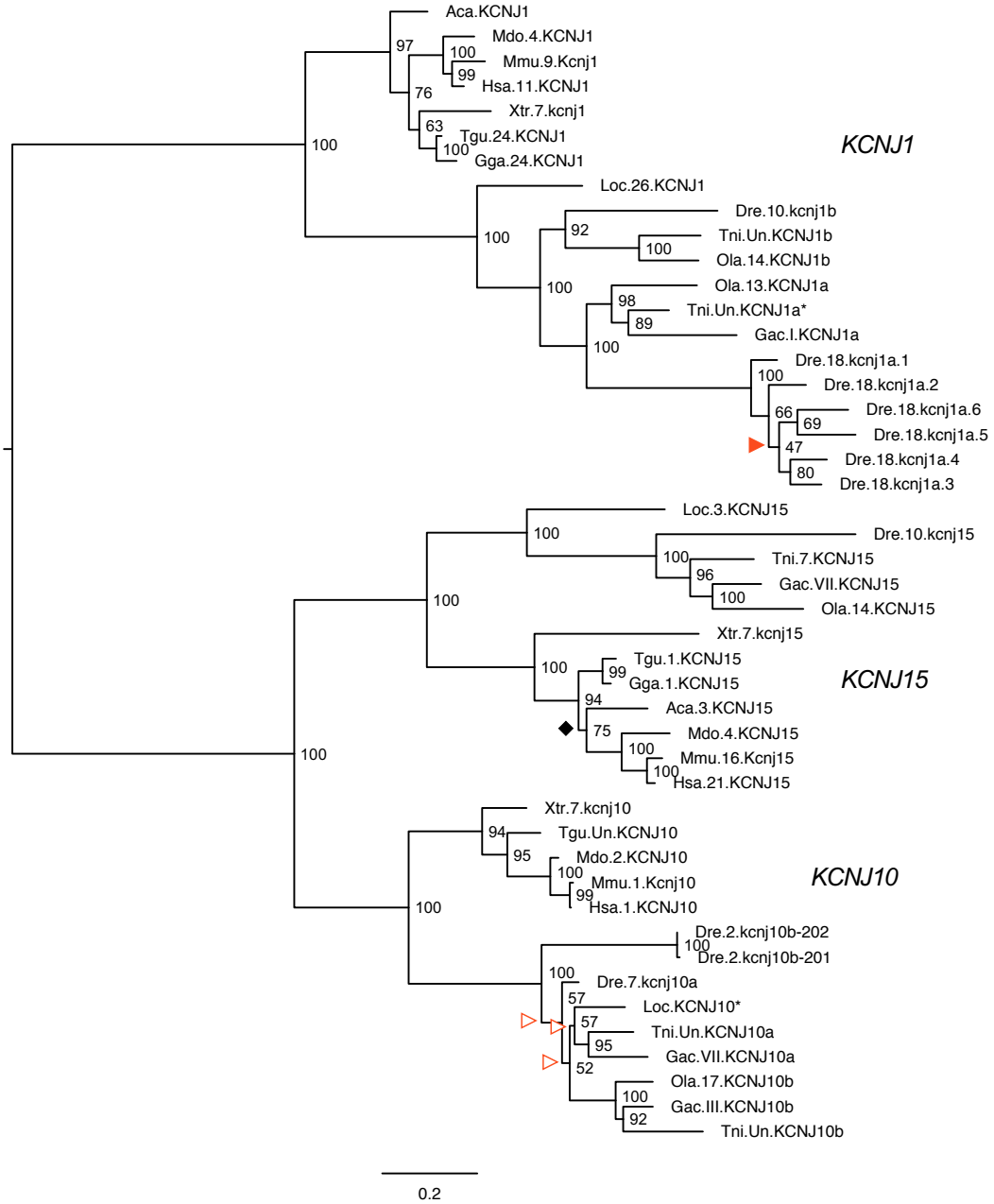
Supplementary figure S17. FCH and double SH3 domain protein (FCHSD). The gene family is made up of the *FCHSD1* and *FCHSD2* genes. Both branches are well-supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). The *FCHSD2* branch has teleost duplicates, however the topology of this branch is not well-resolved. The chromosome locations of the *FCHSD2a* and *-2b* genes nonetheless suggest an origin in the basal teleost tetraploidization (3R).



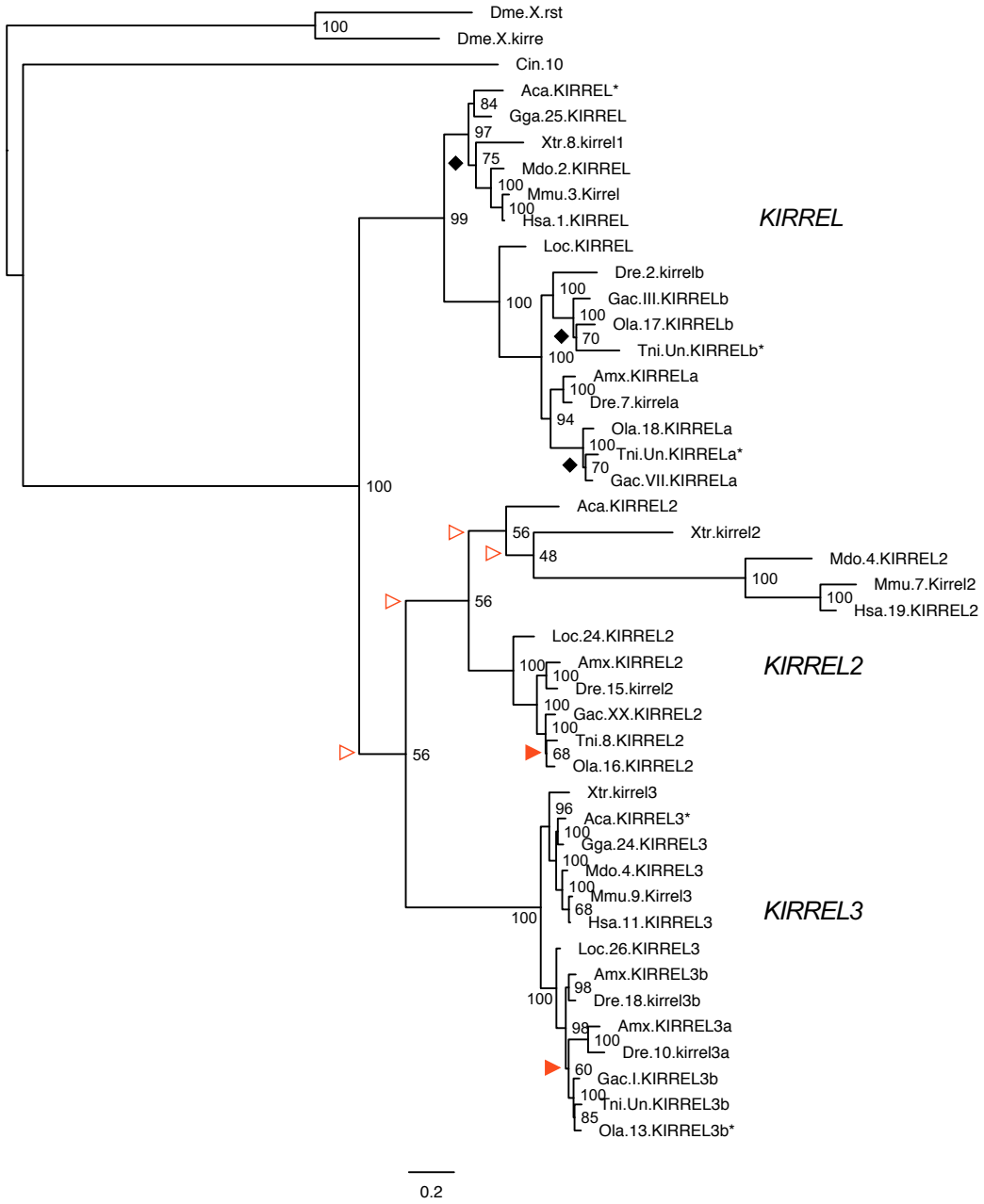
Supplementary figure S18. GRAM domain-containing protein 1 family (GRAMD1). The family is made up of the *GRAMD1A, GRAMD1B* and *GRAMD1C* genes. All three branches are well-supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). There are duplicates of *GRAMD1b* in zebrafish that diverge in the time window of the basal teleost tetraploidization (3R), however duplicates could not be found in other teleost species. The phylogeny is rooted with a red flour beetle sequence annotated as "*GRAM domain-containing protein 1B-like"*, identified through HMMER search. There is a putative fruit fly family member (*CG34394*); however, all predicted transcripts of this gene are considerably longer than the vertebrate sequences, and therefore it was not used.



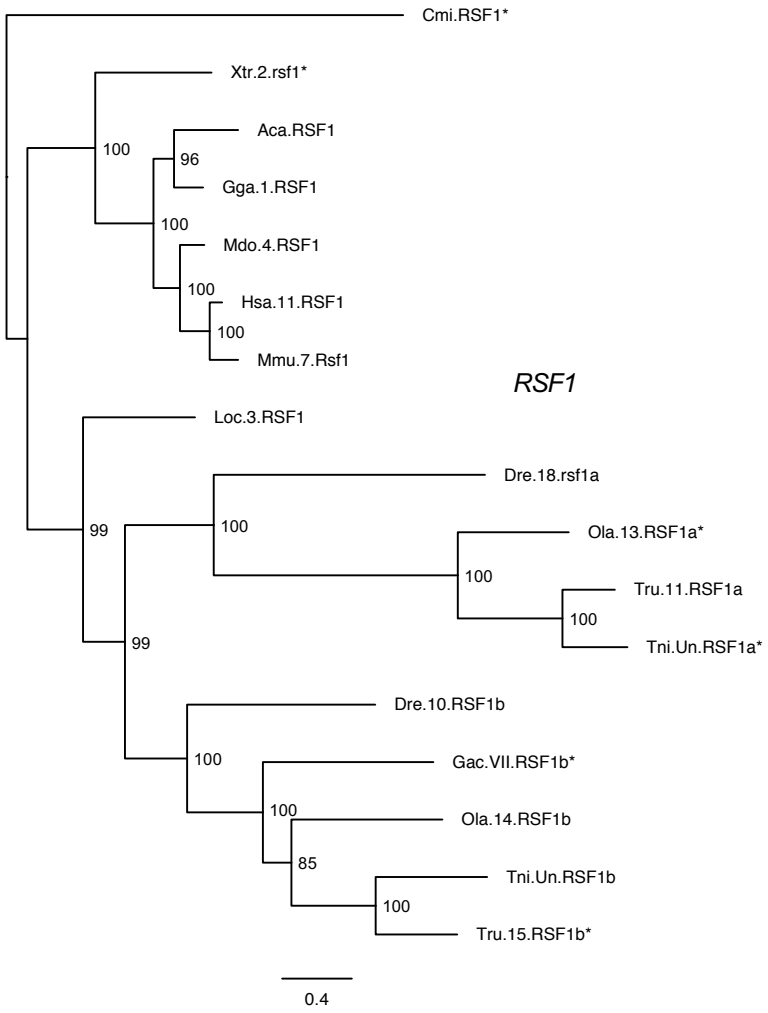
Supplementary figure S19. Voltage-gated potassium channel J, member 1, -10 and -15 family (KCNJ1/10/15). The gene family is made up of the *KCNJ1, KCNJ10* and *KCNJ15* genes. All three branches are well supported; however, no outgroups could be found for this family. Thus, it is not possible to infer a relative time window for the origin of these genes from the phylogeny alone. The phylogeny is displayed as a midpoint-rooted tree. There are teleost duplicates of *KCNJ1* that diverge in the time window of the basal teleost tetraploidization (3R). The zebrafish has 6 tandem duplicates of the *kcnj1a* gene on chromosome 18. There are teleost duplicates of *KCNJ10* as well, however the topology of this branch is not well resolved. The chromosomal locations of the *KCNJ1a* and *KCNJ1b* genes are nonetheless consistent with an origin in 3R (Fig. 4). We could identify two *kcnj10b* tandem duplicates in the genome of the zebrafish. These are annotated as copy number variants *kcnj10b-201* and *kcnj202* in the Ensembl database.



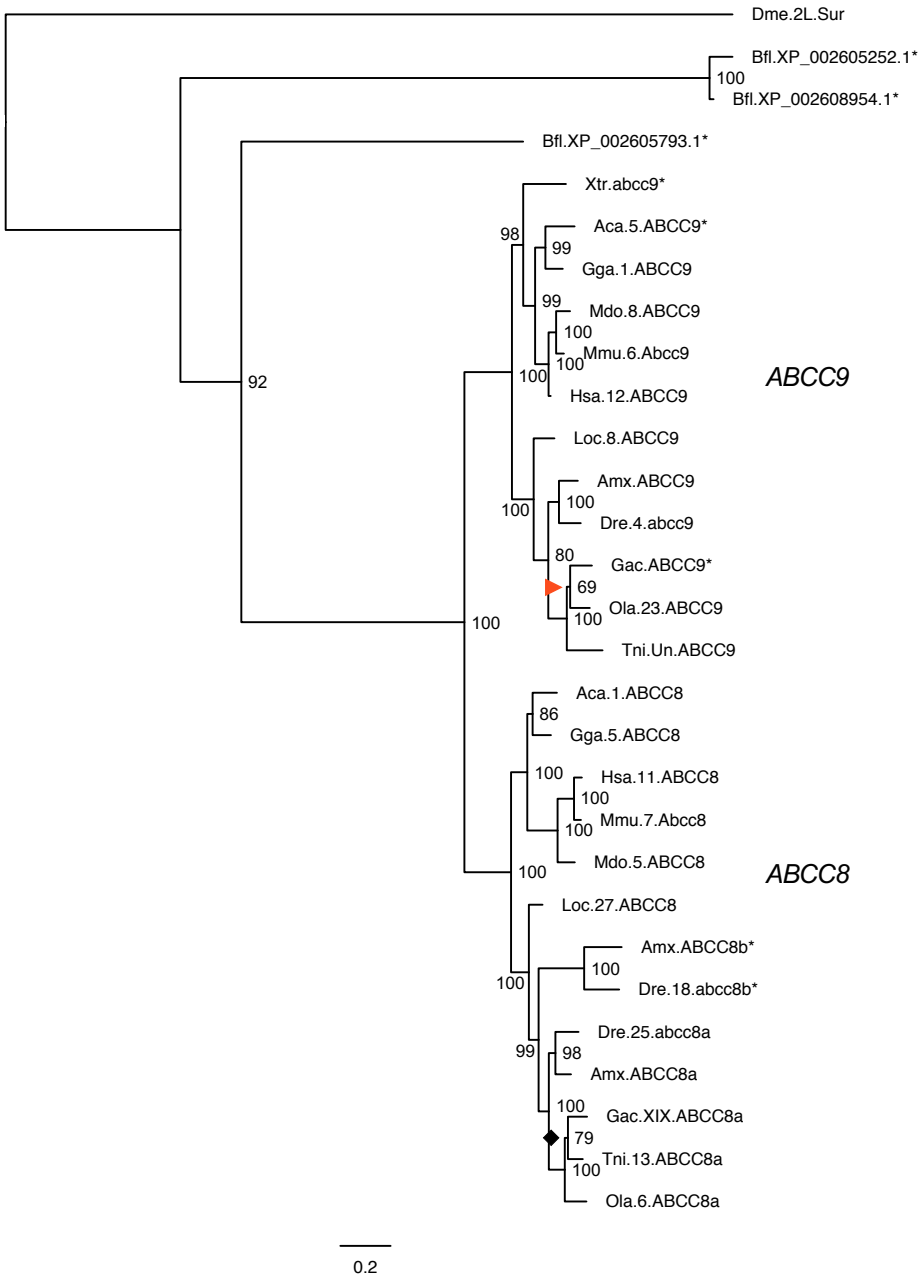
Supplementary figure S20. Kin of irregular chiasm-like protein family (KIRREL). The gene family is made up of the *KIRREL, KIRREL2* and *KIRREL3* genes. The *KIRREL* and *KIRREL3* branches are well supported, while the *KIRREL2* branch is poorly supported in the UFBoot test. This is likely due to the relatively faster evolutionary rate of the tetrapod *KIRREL2* sequences. Nonetheless, all three branches diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). The *KIRREL* and *KIRREL3* branches have duplicates in teleost fishes, however only the *KIRREL* branch shows the duplicate *KIRRELa* and *KIRRELb* branches diverging in the time window of the basal teleost tetraploidization (3R). The topology of the *KIRREL3* branch is not well resolved in this respect. The chromosomal locations of the *KIRREL3a* and *KIRREL3b* genes are nonetheless consistent with an origin in 3R (Fig. 4). The alignment used to create this phylogeny was edited to remove the N-terminal ends, which could not be identified for many of the sequences. The final alignment consists of 1018 amino acid positions.



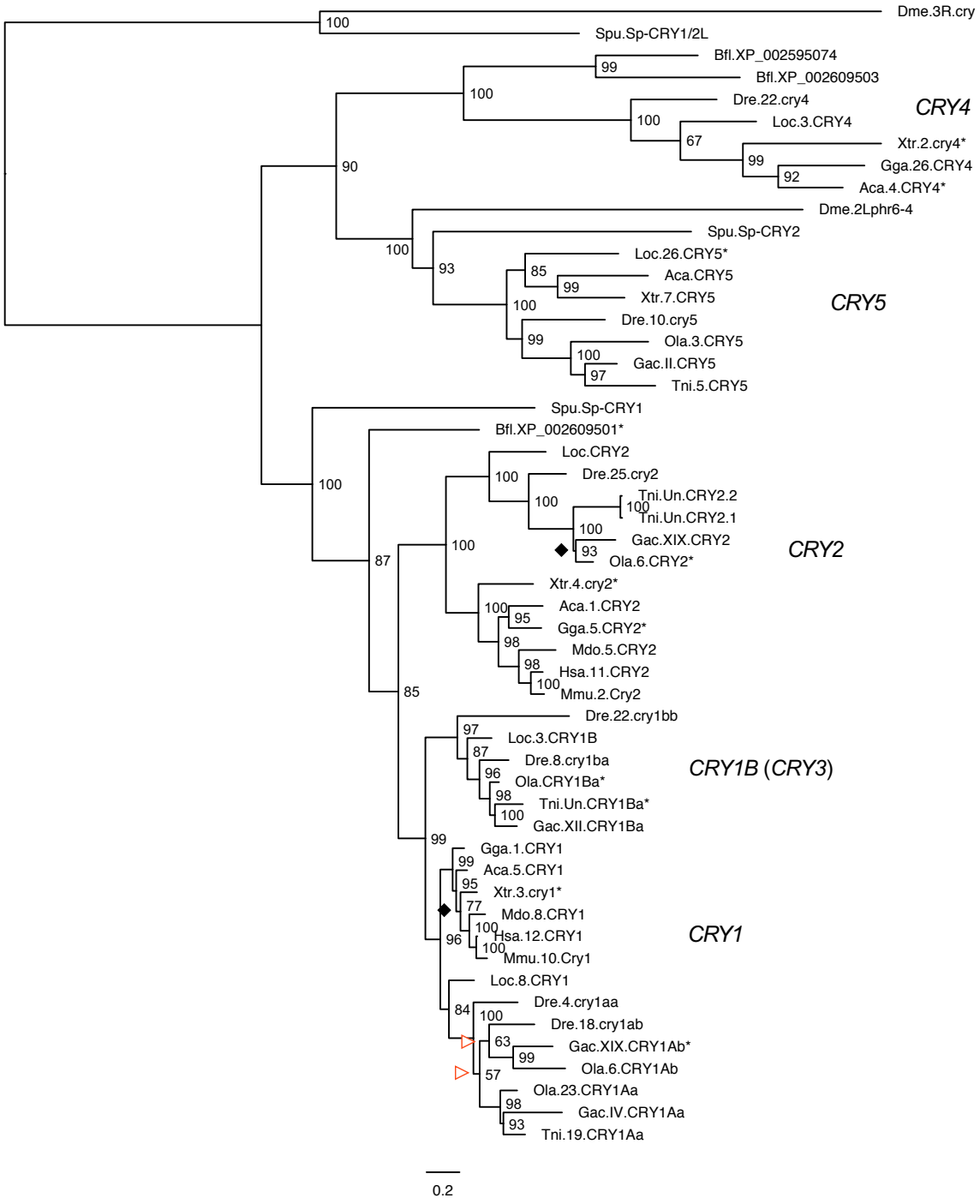
Supplementary figure S21. Remodeling and spacing factor 1 family (RSF1). The gene family consist of only one branch in jawed vertebrates, made up by the *RSF1* genes. No putative invertebrate family members could be identified, and the phylogeny is rooted by the elephant shark *RSF1* sequence. The duplicate *RSF1a* and *-b* branches in teleost fishes diverge in the time window of the basal teleost tetraploidization (3R).



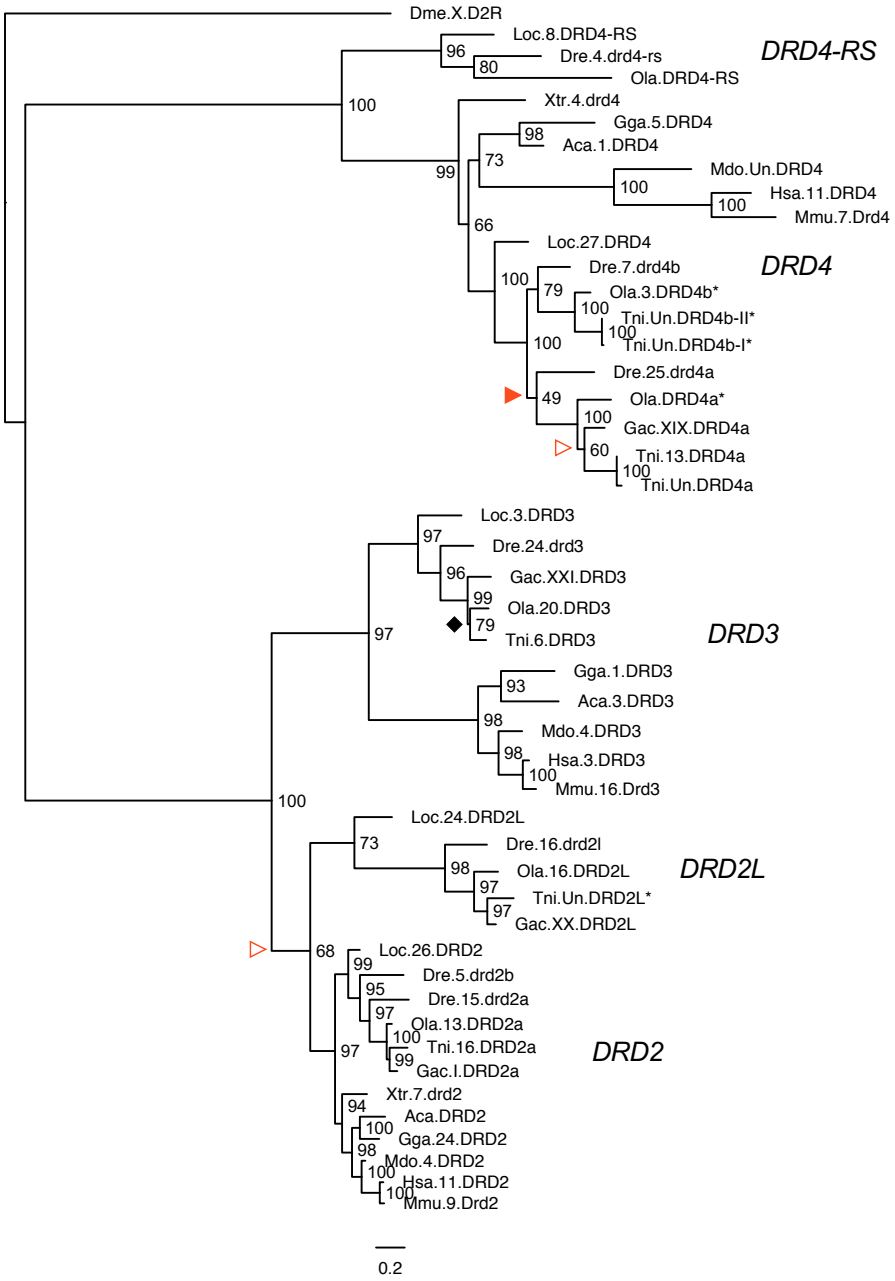
Supplementary figure S22. ATP-binding cassette subfamily C, member 8 and -9 (ABCC8/9). The gene family is made up of the *ABCC8* and *ABCC9* genes. Both branches are well-supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). The *ABCC8* branch has teleost duplicates diverging in the time window of the basal teleost tetraploidization (3R).



Supplementary figure S23. Cryptochrome family (CRY). (See following page.) The gene family is made up of the *CRY1* (*CRY1A* in teleost fishes)*, CRY1B* (also known as *CRY3*)*, CRY2, CRY4* and *CRY5* genes. All five branches are well supported, and our phylogeny supports the divergence of the *CRY1, CRY1B* (*CRY3*) and *CRY2* genes in the time window of the basal vertebrate tetraploidizations. The *CRY1* branch has duplicates in teleost fishes named *CRY1Aa* and *CRY1Ab* that diverge in the time window of the basal teleost tetraploidization. Although the topologies of these duplicate branches are not well-resolved, the chromosomal locations of the *CRY1Aa* and *CRY1Ab* genes are nonetheless consistent with an origin in 3R (Fig. 4). Our results and conclusions are consistent with previously published phylogenomic studies of this gene family (Haug et al., 2015; Liu et al., 2015). Our phylogeny suggests a much earlier time window for the divergences of *CRY4* and *CRY5*, in line with the above cited study by Haug et al. The chromosomal locations of the *CRY4* and *CRY5* genes in vertebrate genomes, in the same chromosomal regions as *CRY1B* and *CRY2* (see Figs. S2 – S4), suggests that the ancestral *CRY1/1B/2, CRY4* and *CRY5* genes were located on the same ancestral chromosome pre-1R/2R, possibly even further back. After 1R/2R, only 1 duplicate was preserved in each of the *CRY4* and *CRY5* lineages. In our phylogeny, the fruit fly *cry* and *phr6-4* sequences serve as outgroups. However, since they don't cluster together, our phylogeny is presented as a midpoint-rooted tree.

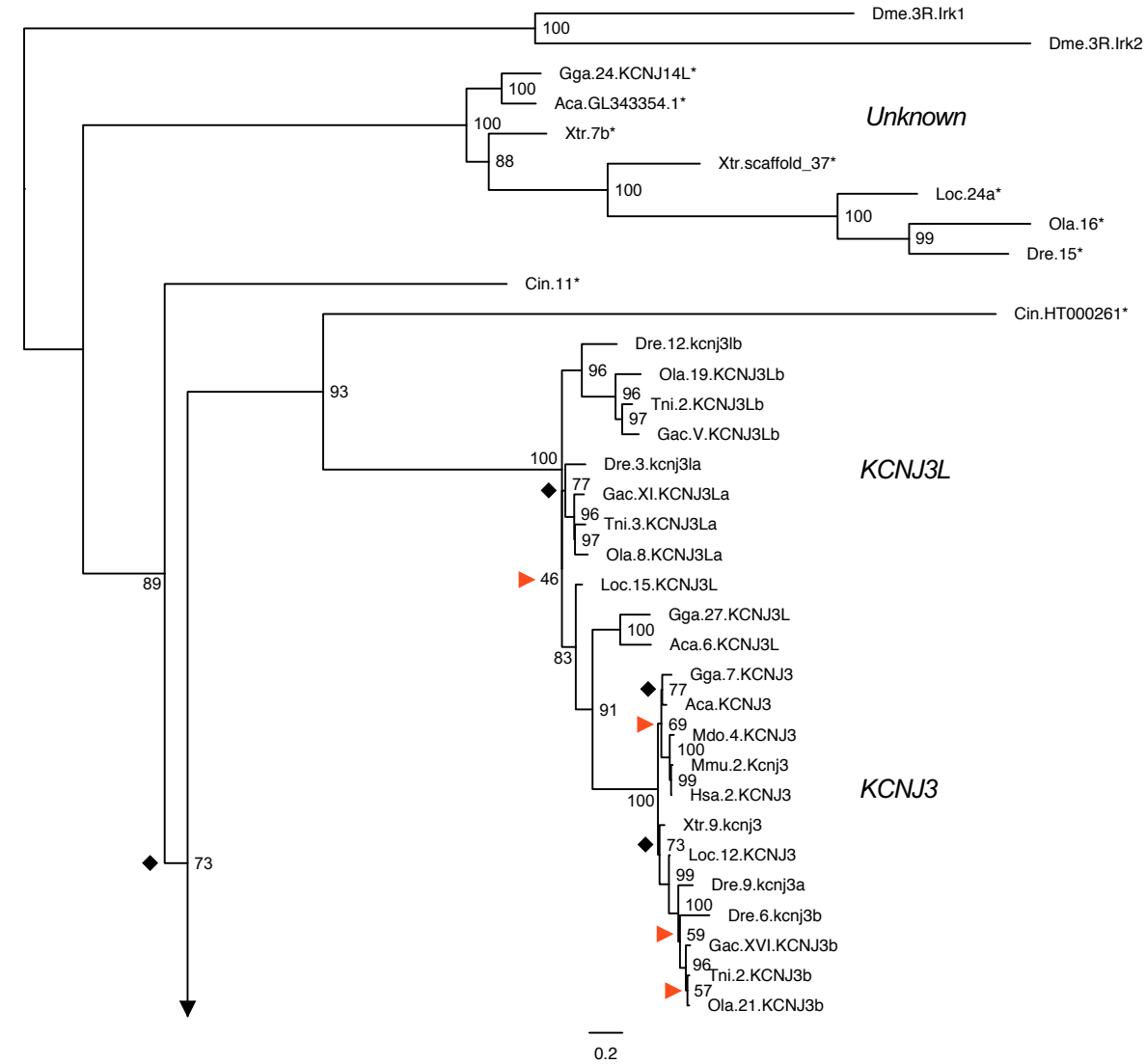

Dme.3R.cry
Spu.Sp-CRY1/2L
Bfl.XP_002595074
Bfl.XP_002609503
Dre.22.cry4
Loc.3.CRY4
Xtr.2.cry4*
Gga.26.CRY4
Aca.4.CRY4*
CRY4
Dme.2Lphr6-4
Spu.Sp-CRY2
Loc.26.CRY5*
Aca.CRY5
Xtr.7.CRY5
Dre.10.cry5
Ola.3.CRY5
Gac.II.CRY5
Tni.5.CRY5
CRY5
Spu.Sp-CRY1
Bfl.XP_002609501*
Loc.CRY2
Dre.25.cry2
Tni.Un.CRY2.2
Tni.Un.CRY2.1
Gac.XIX.CRY2
Ola.6.CRY2*
Xtr.4.cry2*
Aca.1.CRY2
Gga.5.CRY2*
Mdo.5.CRY2
Hsa.11.CRY2
Mmu.2.Cry2
CRY2
Dre.22.cry1bb
Loc.3.CRY1B
Dre.8.cry1ba
Ola.CRY1Ba*
Tni.Un.CRY1Ba*
Gac.XII.CRY1Ba
CRY1B (CRY3)
Gga.1.CRY1
Aca.5.CRY1
Xtr.3.cry1*
Mdo.8.CRY1
Hsa.12.CRY1
Mmu.10.Cry1
Loc.8.CRY1
Dre.4.cry1aa
Dre.18.cry1ab
Gac.XIX.CRY1Ab*
Ola.6.CRY1Ab
Ola.23.CRY1Aa
Gac.IV.CRY1Aa
Tni.19.CRY1Aa
CRY1
0.2

Supplementary figure S24. Dopamine receptor D2 family (D2). The gene family is made up of the *DRD2, DRD2L, DRD3, DRD4* and *DRD4-RS* genes. All five branches are well-supported; however, the lack of invertebrate chordate sequences in this family means that it is not possible to infer a relative time window for the origin of the genes from the phylogeny alone. The chromosomal locations indicate that the *DRD2, DRD2L* and *DRD3* genes on one side, and the *DRD4* and *DRD4-RS* genes on the other, arose from ancestral genes in the basal vertebrate tetraploidizations (1R/2R). Notably, the *DRD2, DRD2L* and *DRD3* genes are located in the *SL* gene-bearing chromosome regions (Fig. 5), while the *DRD4* and *DRD4-RS* genes are located in the *GH/PRL/PRL2* gene-bearing chromosome regions (Figs. S2 – S4). The *DRD2* and *DRD4* branches have duplicates in teleost fishes diverging in the time window of the basal teleost tetraploidization (3R). Our results and conclusions are consistent with previously published detailed analyses of the dopamine receptor family (Haug-Baltzell et al., 2015; Yamamoto et al., 2015).

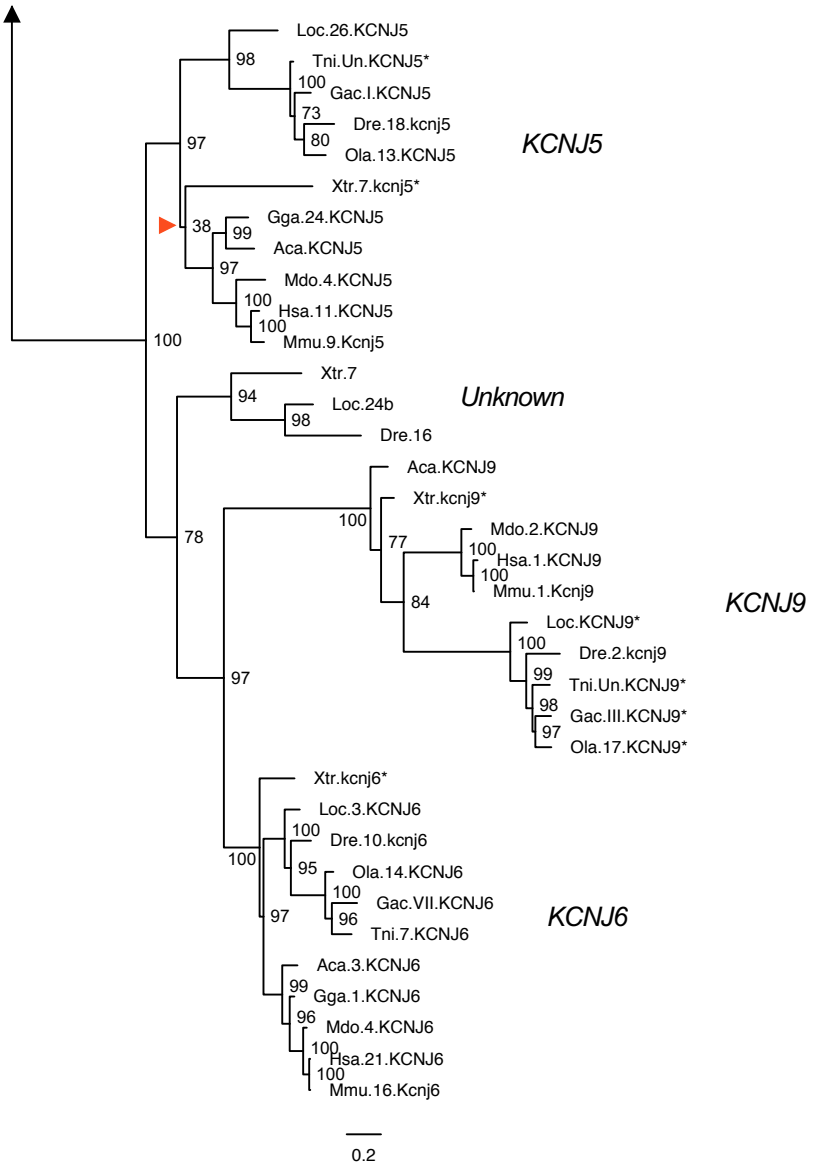


Supplementary figure S25. Voltage-gated potassium channel J, member 3, -5, -6 and -9 family (KCNJ3/5/6/9). The gene family is made up by the *KCNJ3, KCNJ5, KCNJ6* and *KCNJ9* genes, as well as three previously undescribed clades. One of these "unknown" clades makes up the most basal branch in our phylogeny (excluding the root), indicating that it diverged from the rest of the family before the split between tunicates and the lineage leading to vertebrates. The chicken gene annotated as *KCNJ14L* belongs to this branch. Another of these "unknown" clades clusters with the *KCNJ3* genes, and together they have a vase tunicate outgroup. We have named these *KCNJ3*-like sequences *KCNJ3L*. The *KCNJ3L* branch is not resolved in our phylogeny. Both *KCNJ3* and *KCNJ3L* genes are located outside our chromosomal regions of interest. Thus, they were not considered in our conserved synteny analyses. In the context of our analyses, the *KCNJ5, KCNJ6* and *KCNJ9* genes are of interest. All three branches are well supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). None of these branches seem to have preserved duplicates from the basal teleost tetraploidization (3R). The third "unknown" clade in our phylogeny clusters close to the *KCNJ6* and *KCNJ9* branches, and is made up of only *Xenopus tropicalis,* spotted gar and zebrafish sequences. The *Xenopus tropicalis* gene in this clade (located on chromosome 7) has wrongly been identified as *kcnj9*. In fact, we could find the correct *kcnj9* gene on scaffold_7147. Our phylogeny, as well as the chromosomal locations (Fig. 5), support this branch as the fourth duplicate generated in 1R/2R.

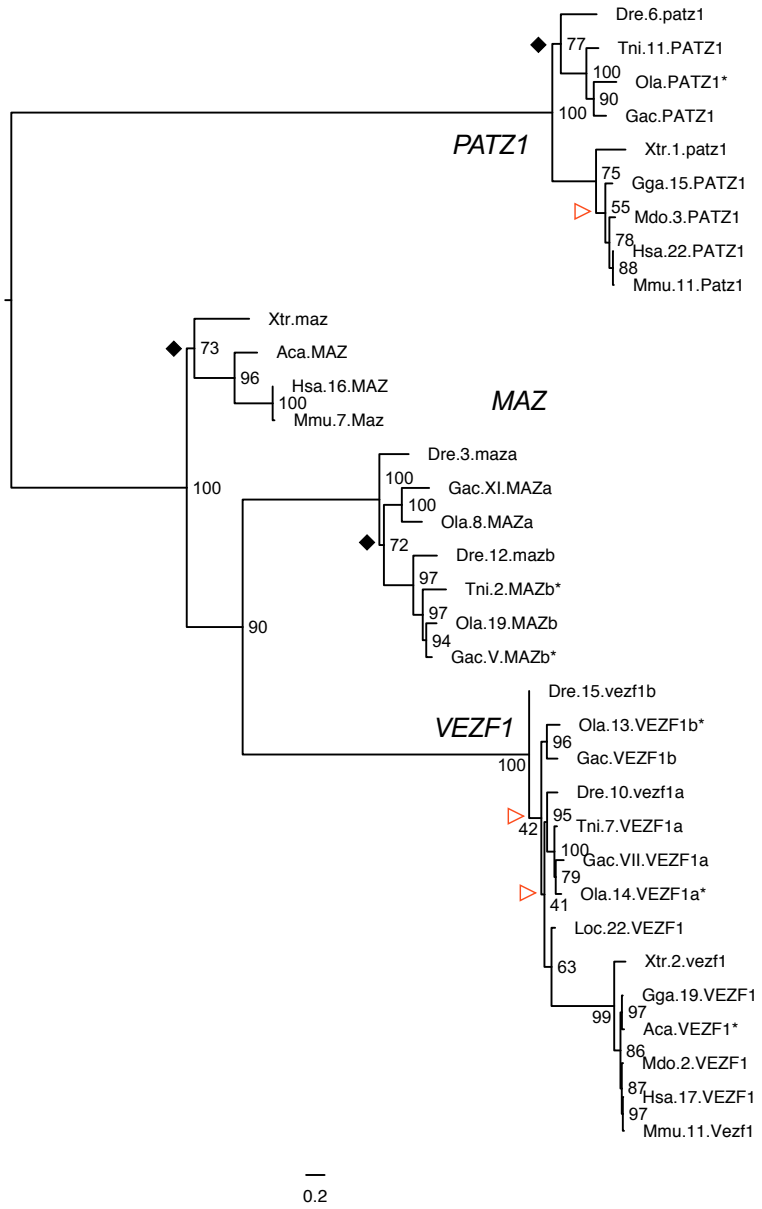
A.



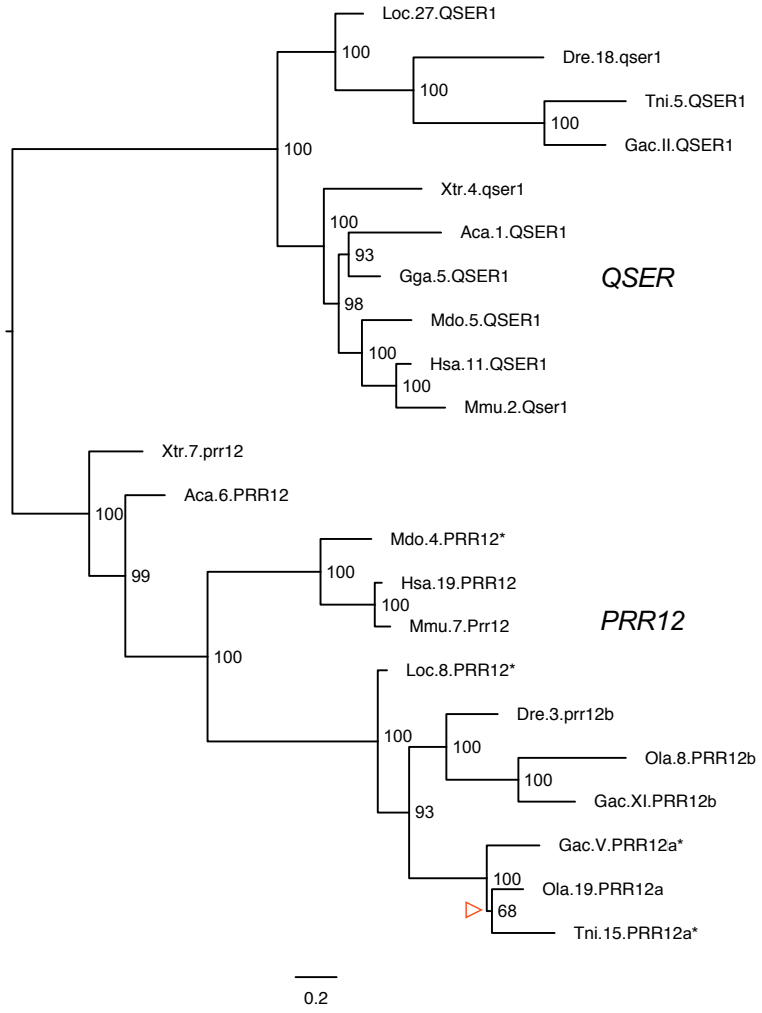
B.


Loc.26.KCNJ5
98
Tni.Un.KCNJ5*
100
Gac.I.KCNJ5
73
Dre.18.kcnj5
80
Ola.13.KCNJ5
KCNJ5
97
Xtr.7.kcnj5*
38
Gga.24.KCNJ5
99
Aca.KCNJ5
97
Mdo.4.KCNJ5
100
Hsa.11.KCNJ5
100
Mmu.9.Kcnj5
100
Xtr.7
94
Loc.24b
Unknown
98
Dre.16
Aca.KCNJ9
Xtr.kcnj9*
100
Mdo.2.KCNJ9
77
100
Hsa.1.KCNJ9
100
Mmu.1.Kcnj9
78
84
KCNJ9
Loc.KCNJ9*
100
Dre.2.kcnj9
99
Tni.Un.KCNJ9*
98
Gac.III.KCNJ9*
97
Ola.17.KCNJ9*
97
Xtr.kcnj6*
Loc.3.KCNJ6
100
Dre.10.kcnj6
100
95
Ola.14.KCNJ6
100
Gac.VII.KCNJ6
97
96
Tni.7.KCNJ6
KCNJ6
Aca.3.KCNJ6
99
Gga.1.KCNJ6
96
Mdo.4.KCNJ6
100
Hsa.21.KCNJ6
100
Mmu.16.Kcnj6
0.2

Supplementary figure S26. MYC-associated zinc finger transcription factor family. The gene family is made up of the *MAZ, PATZ1* (POZ/BTB and AT hook containing zinc finger 1) and *VEZF1* (vascular endothelial zinc finger 1) genes. No outgroups could be found for this family, and the phylogeny is displayed as a midpoint-rooted tree. Thus, it is not possible to infer a relative time window for the origin of these genes from the phylogeny alone. Additionally, the *MAZ* branch is not resolved in our phylogeny. Nonetheless, the chromosomal locations of these genes are consistent with origins in the basal vertebrate tetraploidizations (1R/2R) (Figs. S2 – S4). There are teleost duplicates in the *MAZ* and *VEZF1* branches; however, these duplicate teleost branches are not well-supported. The chromosomal locations of the *MAZa* and *MAZb* genes are consistent with an origin in the basal teleost tetraploidization (Figs. S3 and S4).

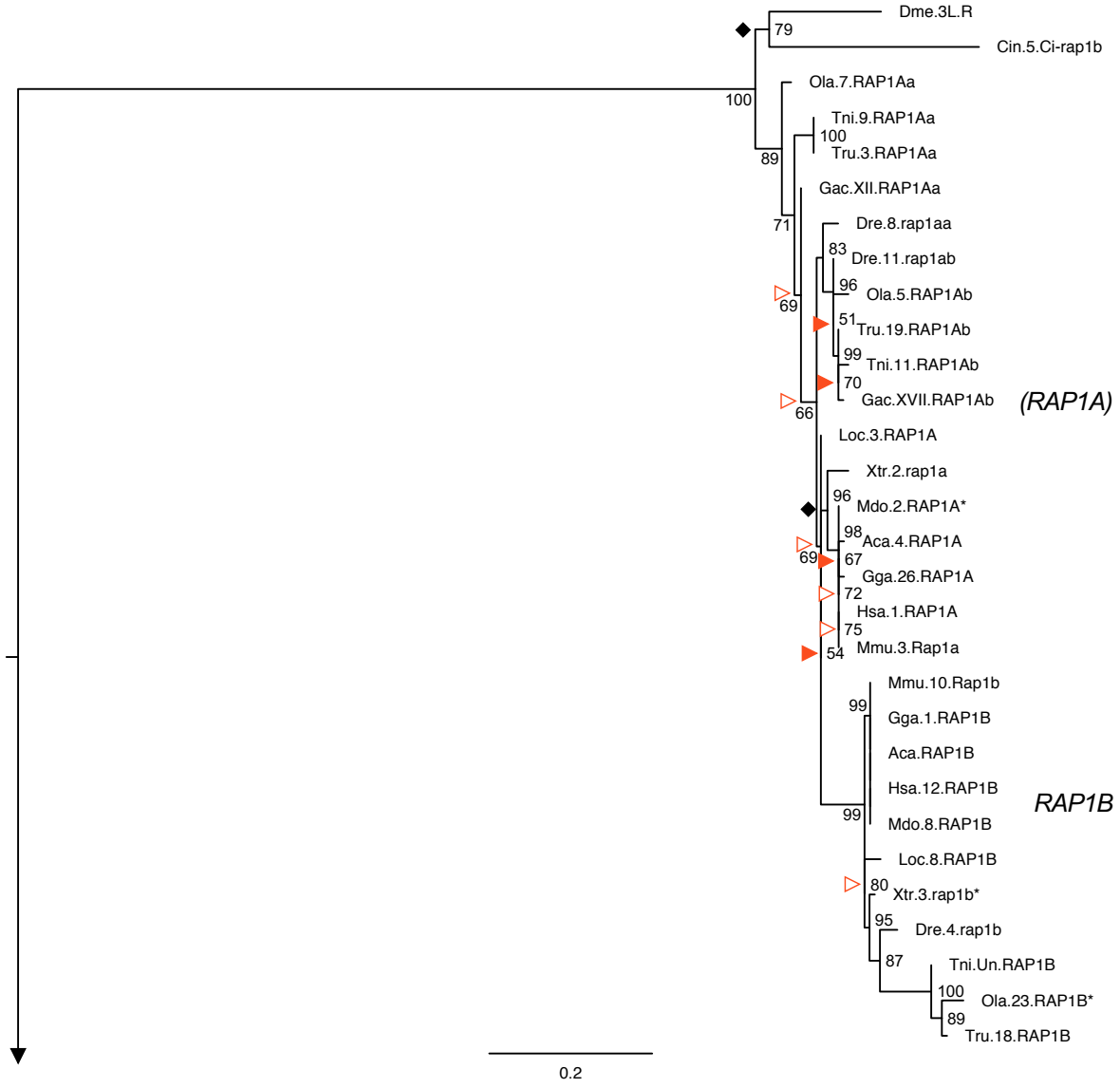


Supplementary figure S27. Glutamine and serine rich 1 protein and proline rich 12 protein family (QSER/PRR12). The gene family is made up of the *QSER* and *PRR12* genes. Both branches are well-supported, however the tetrapod *PRR12* branch is not resolved. No outgroups could be found for this family, and the phylogeny is displayed as a midpoint-rooted tree. Thus, it is not possible to infer a relative time window for the origin of these genes from the phylogeny alone. The chromosomal locations are nonetheless consistent with an origin in the basal vertebrate tetraploidizations (1R/2R) (Figs. S2 – S4). The *QSER* branch has teleost duplicates that diverge in the time window of the basal teleost tetraploidization (3R).



Supplementary figure S28. RAS oncogene family (RAP). Our phylogeny shows two clearly defined subfamilies: One made up of the *RAP1A* and *RAP1B* genes, and one made up of the *RAP2A, RAP2B* and *RAP2C* genes. The latter form three well-supported branches diverging in the time window of the basal vertebrate tetraploidizations (1R/2R). However, these genes are located outside our chromosomal regions of interest. Thus, they were not considered in our conserved synteny analyses. In the context of our analyses, only the *RAP1A* and *RAP1B* are of interest. In our phylogeny, only the *RAP1B* branch is resolved and well-supported. Additionally, the vase tunicate family member clusters with the fruit fly outgroup. Thus, the phylogeny of the *RAP1A* and *RAP1B* subfamily is inconclusive with respect to 1R/2R. Nevertheless, the chromosomal locations are consistent with origins in 1R/2R.

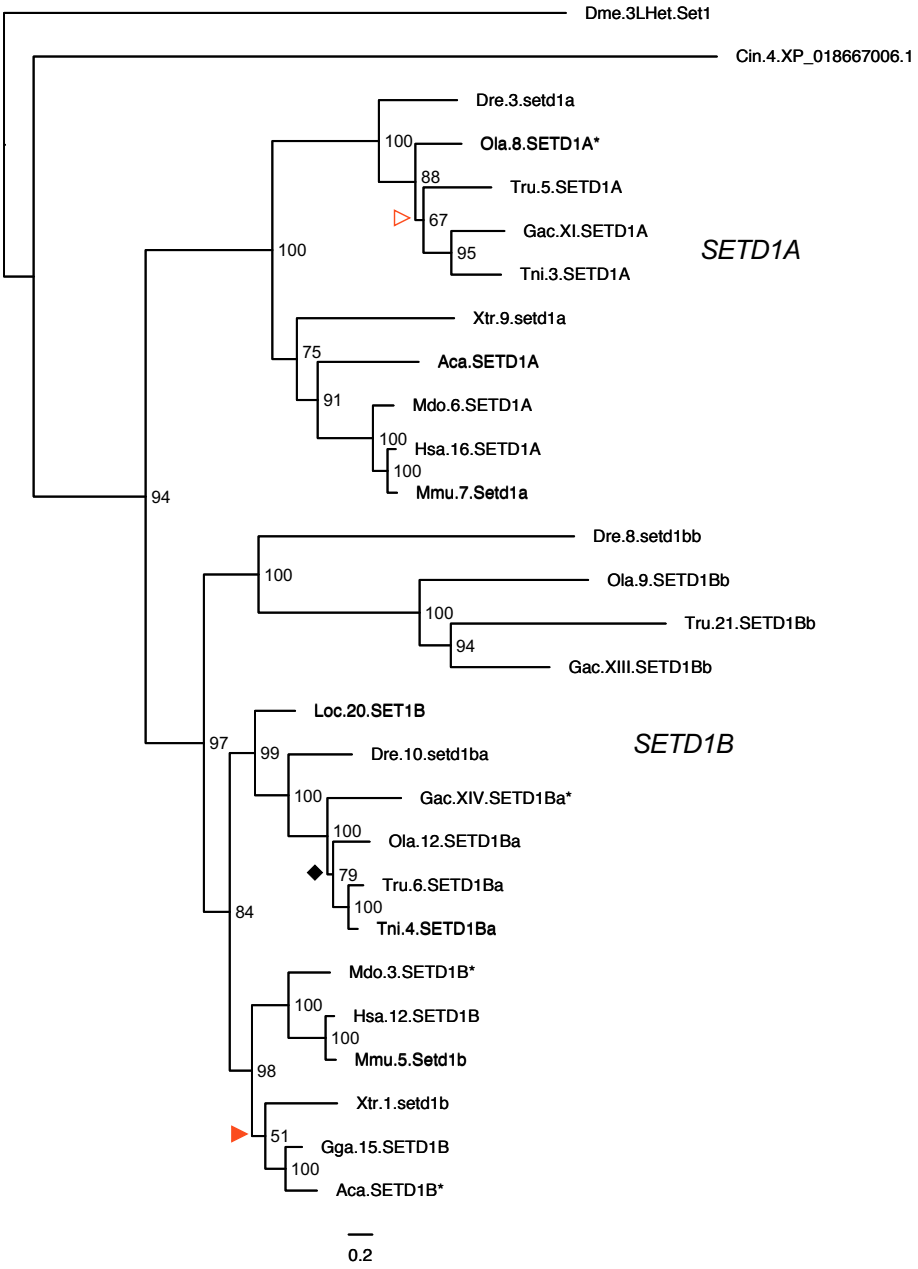
A.



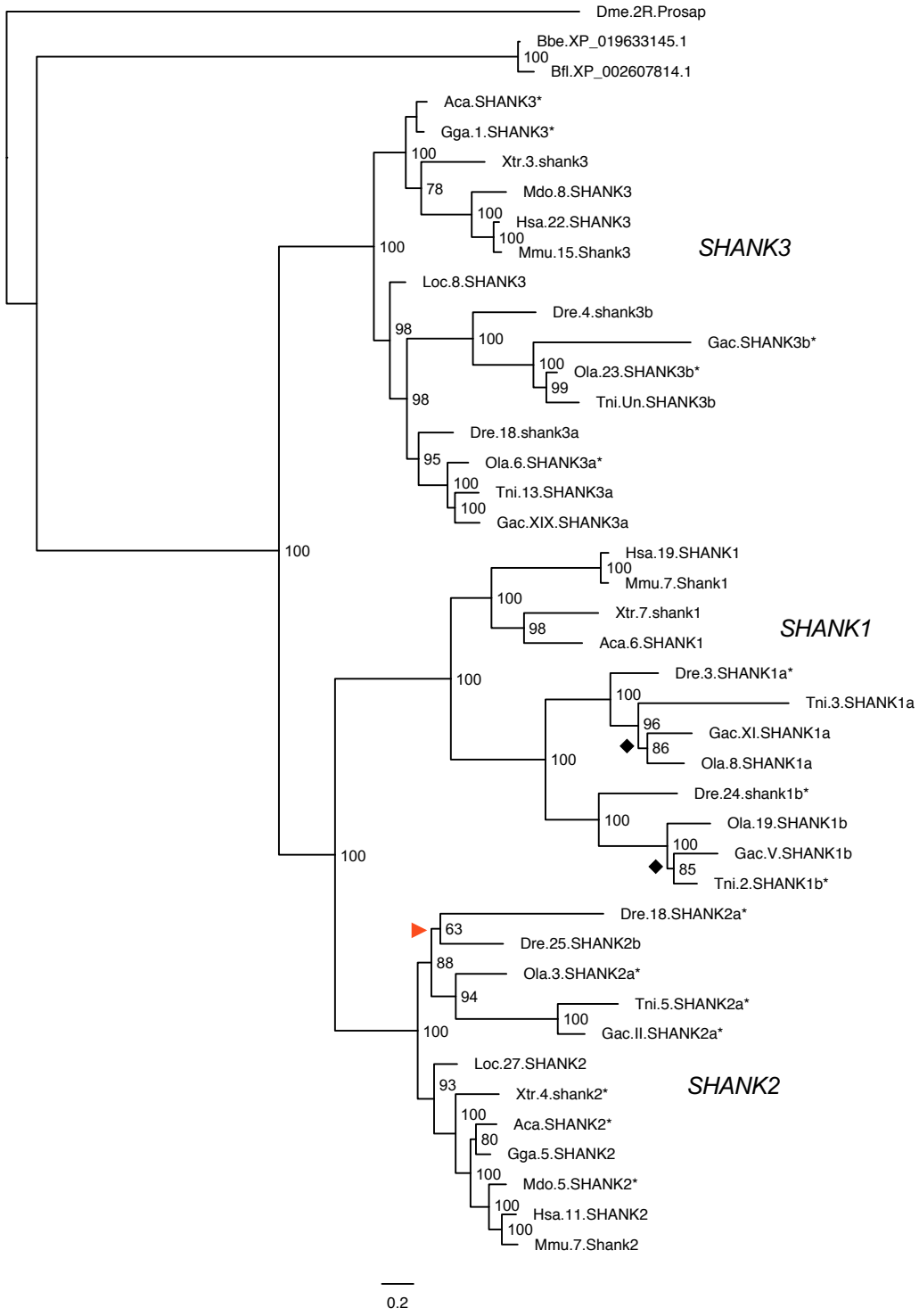
B.


Dme.2R.Rap2l
Cin.HT000116.1
Tni.16
Ola.13
Gac.I
Dre.18.rap2b
Loc.14.RAP2B
Aca.3.RAP2B
Mdo.7.RAP2B
Gga.9.RAP2B
Mmu.3.Rap2b
Hsa.3.RAP2B
Xtr.5.rap2b
Tni.Un.RAP2B-I*
Tni.Un.RAP2B-II
Tru.22.RAP2B
Gac.III.RAP2B
Ola.17.RAP2B
RAP2B
Mdo.X.RAP2C
Gga.4.RAP2C*
Xtr.8.rap2c
Mmu.X.Rap2c
Hsa.X.RAP2C
Loc.7.RAP2C
Ola.10.RAP2C
Dre.14.rap2c
Gac.IV.RAP2C
Tru.14.RAP2C
Tni.1.RAP2C
RAP2C
Hsa.13.RAP2A
Mdo.7.RAP2A
Mmu.14.Rap2a
Xtr.2.rap2a
Aca.3.RAP2A
Gga.1.RAP2A
Tru.1.RAP2Aa
Tni.2.RAP2Aa
Dre.9.rap2aa
Loc.17.RAP2A
Ola.21.RAP2Aa
Dre.1.rap2ab
Ola.2.RAP2Ab
Gac.I.RAP2Ab
Tni.3.RAP2Ab
Tru.8.RAP2Ab
RAP2A
100
91
86
91
95
90
0.2

Supplementary figure S29. SET domain-containing 1 family (SETD1). The gene family is made up of the *SETD1A* and *SETD1B* genes. Both branches are well-supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). There is a third branch diverging in this same time window; however, the species representation as well as the chromosomal positions (Fig. S3 and S4) are consistent with this being a teleost duplicate branch that arose in the basal teleost tetraploidization (3R). The position of this branch, basal to the rest of the *SETD1B* sequences, is likely cause by its relatively faster rate of evolution, as indicated by the branch lengths.

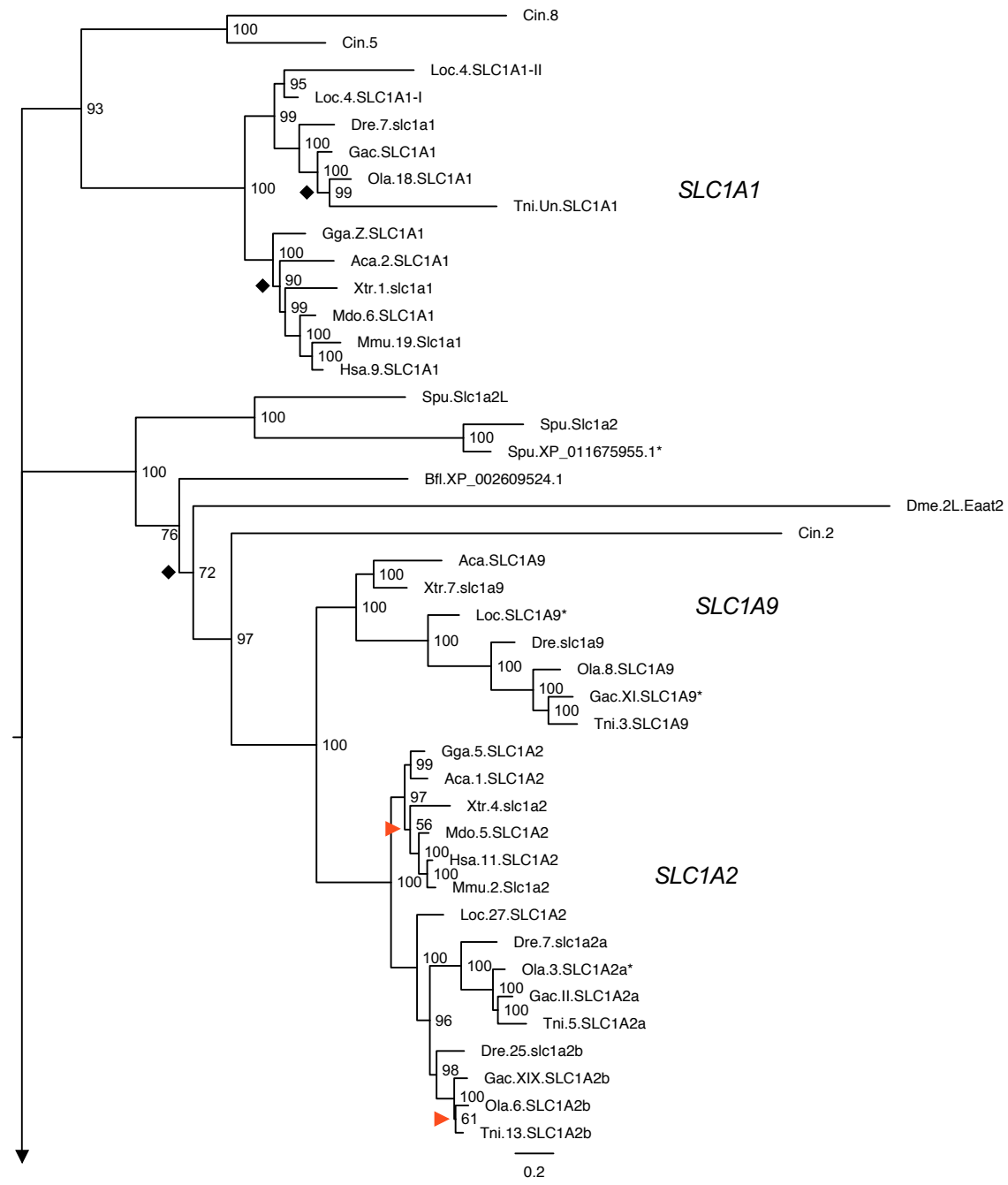


Supplementary figure S30. SH3 and multiple ankyrin repeat domains family (SHANK). The gene family is made up of the *SHANK1, SHANK2* and *SHANK3* genes. All three branches are well-supported and diverge in the time window of the basal vertebrate tetraploidizations. There are teleost duplicates in all three branches; however, only the topology of the *SHANK3* branch conclusively shows the duplicate branches diverging in the time window of the basal vertebrate tetraploidization (3R). The species distributions and chromosome locations nonetheless support the duplication of all three branches in 3R (Figs. S3 – S4). The Florida lancelet sequence used in this phylogeny could also be found in Belcher's lancelet (*Branchiostoma belcheri*, abbreviated Bbe).

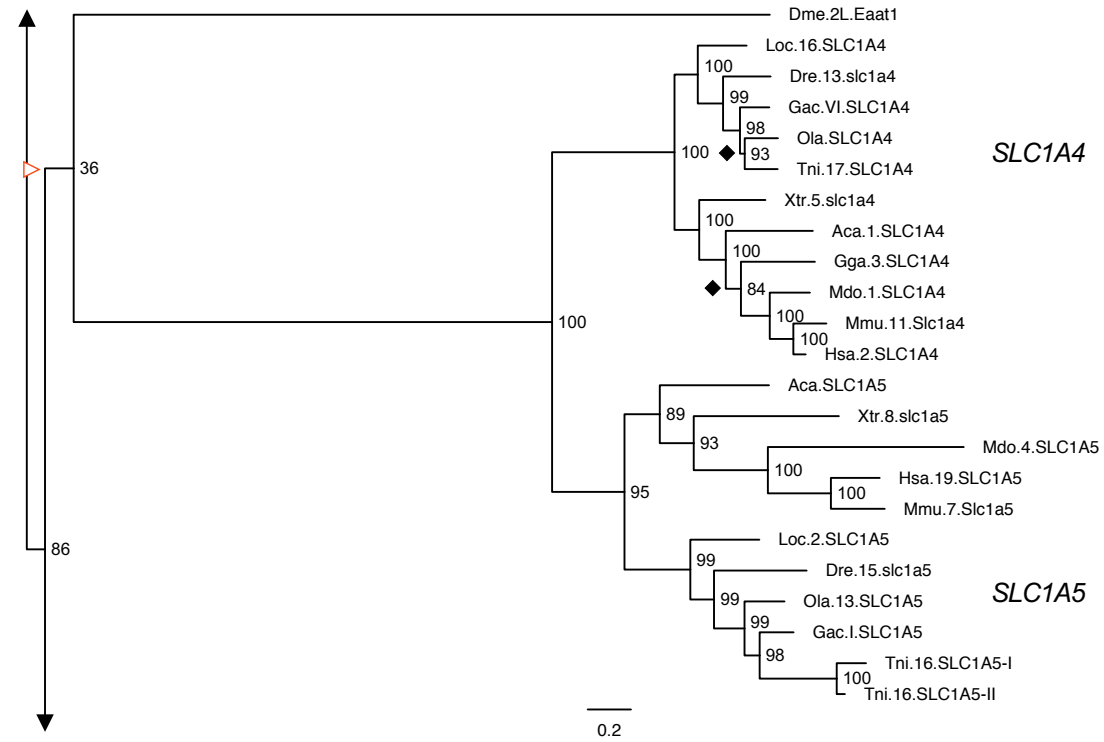


Supplementary figure S31. Solute carrier family 1 (SLC1A). (See page 40 – 42.) This large family of excitatory/neutral amino acid transporters is made up of 9 subtype genes in vertebrates (Gesemann et al., 2010). The majority of these genes are located outside our chromosomal regions of interest. Thus, they were not considered in the conserved synteny analyses of this study. Our midpoint-rooted phylogeny of the full SLC1A family is shown below. In the context of our analyses, only the *SLC1A2* and *SLC1A9* genes are of interest (Fig. S31A below). Both these branches are well-supported and diverge in the time window of the basal vertebrate tetraploidizations (1R/2R). In addition, the *SLC1A2* branch has teleost duplicates that diverge in the time window of the basal teleost tetraploidization (3R). We have analysed this family previously: The *SLC1A3, SLC1A6, SLC1A7* and *SLC1A8* genes (Fig. S31C below) are located in the chromosomal regions bearing the growth hormone and prolactin receptor genes *GHR* and *PRLR*, as well as the genes for the erythropoietin and thrombopoietin receptors *EPOR* and *TPOR.* In this study, we could show that these chromosomal regions arose in the 1R/2R tetraploidizations, and duplicated further in the 3R tetraploidization (Ocampo Daza and Larhammar, 2017).

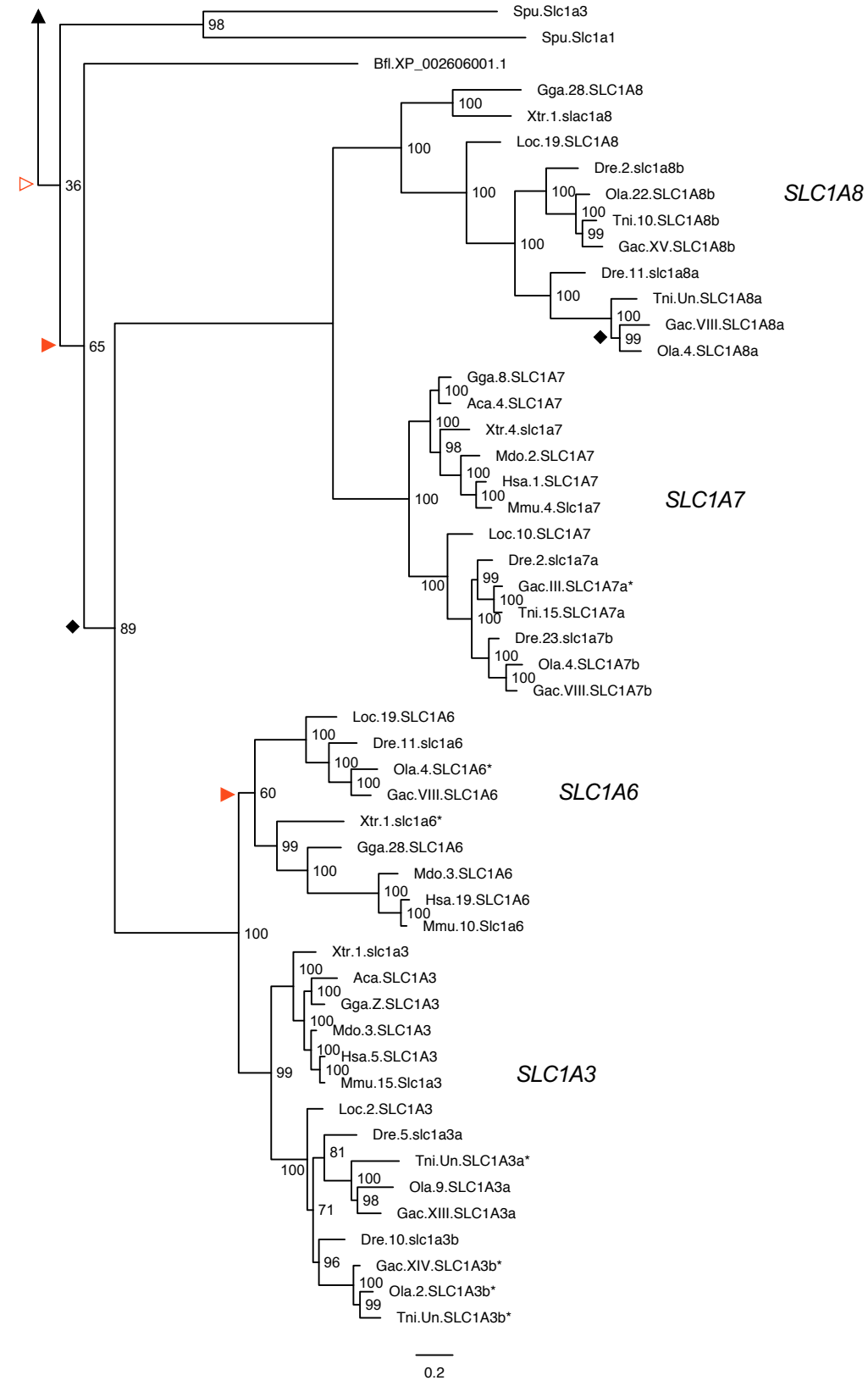
A.


Cin.8
Cin.5
Loc.4.SLC1A1-II
Loc.4.SLC1A1-I
Dre.7.slc1a1
Gac.SLC1A1
Ola.18.SLC1A1
Tni.Un.SLC1A1
SLC1A1
Gga.Z.SLC1A1
Aca.2.SLC1A1
Xtr.1.slc1a1
Mdo.6.SLC1A1
Mmu.19.Slc1a1
Hsa.9.SLC1A1
Spu.Slc1a2L
Spu.Slc1a2
Spu.XP_011675955.1*
Bfl.XP_002609524.1
Dme.2L.Eaat2
Cin.2
Aca.SLC1A9
Xtr.7.slc1a9
Loc.SLC1A9*
Dre.slc1a9
Ola.8.SLC1A9
Gac.XI.SLC1A9*
Tni.3.SLC1A9
SLC1A9
Gga.5.SLC1A2
Aca.1.SLC1A2
Xtr.4.slc1a2
Mdo.5.SLC1A2
Hsa.11.SLC1A2
Mmu.2.Slc1a2
SLC1A2
Loc.27.SLC1A2
Dre.7.slc1a2a
Ola.3.SLC1A2a*
Gac.II.SLC1A2a
Tni.5.SLC1A2a
Dre.25.slc1a2b
Gac.XIX.SLC1A2b
Ola.6.SLC1A2b
Tni.13.SLC1A2b
0.2

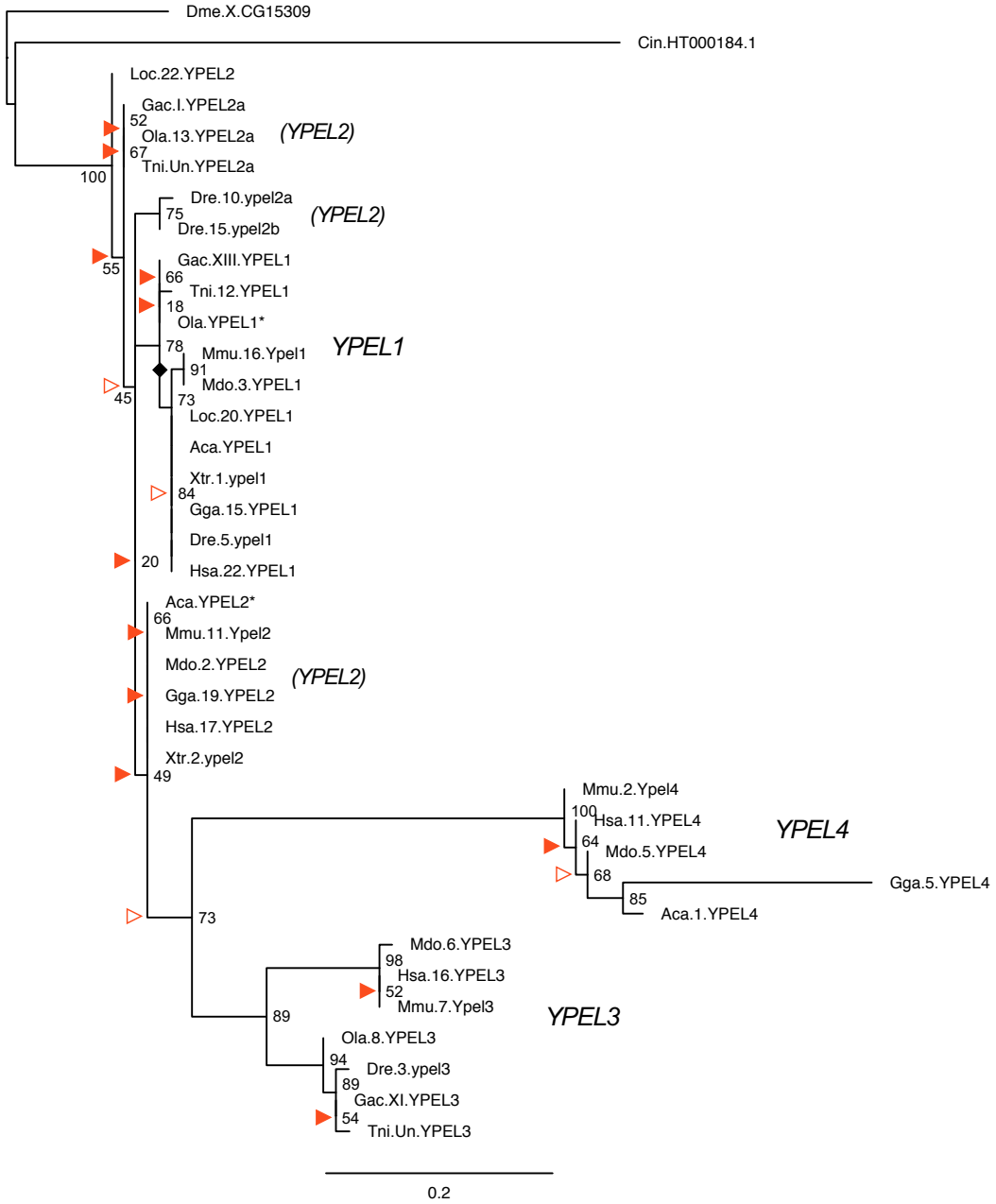
B.


Dme.2L.Eaat1
Loc.16.SLC1A4
Dre.13.slc1a4
Gac.VI.SLC1A4
Ola.SLC1A4
Tni.17.SLC1A4
SLC1A4
Xtr.5.slc1a4
Aca.1.SLC1A4
Gga.3.SLC1A4
Mdo.1.SLC1A4
Mmu.11.Slc1a4
Hsa.2.SLC1A4
Aca.SLC1A5
Xtr.8.slc1a5
Mdo.4.SLC1A5
Hsa.19.SLC1A5
Mmu.7.Slc1a5
Loc.2.SLC1A5
Dre.15.slc1a5
Ola.13.SLC1A5
Gac.I.SLC1A5
Tni.16.SLC1A5-I
Tni.16.SLC1A5-II
SLC1A5
36
86
100
100
99
98
100
93
100
100
100
84
100
100
89
93
100
100
95
99
99
99
98
100
0.2

C.


Spu.Slc1a3
Spu.Slc1a1
Bfl.XP_002606001.1
Gga.28.SLC1A8
Xtr.1.slac1a8
Loc.19.SLC1A8
Dre.2.slc1a8b
Ola.22.SLC1A8b
Tni.10.SLC1A8b
Gac.XV.SLC1A8b
Dre.11.slc1a8a
Tni.Un.SLC1A8a
Gac.VIII.SLC1A8a
Ola.4.SLC1A8a
SLC1A8
Gga.8.SLC1A7
Aca.4.SLC1A7
Xtr.4.slc1a7
Mdo.2.SLC1A7
Hsa.1.SLC1A7
Mmu.4.Slc1a7
Loc.10.SLC1A7
Dre.2.slc1a7a
Gac.III.SLC1A7a*
Tni.15.SLC1A7a
Dre.23.slc1a7b
Ola.4.SLC1A7b
Gac.VIII.SLC1A7b
SLC1A7
Loc.19.SLC1A6
Dre.11.slc1a6
Ola.4.SLC1A6*
Gac.VIII.SLC1A6
Xtr.1.slc1a6*
Gga.28.SLC1A6
Mdo.3.SLC1A6
Hsa.19.SLC1A6
Mmu.10.Slc1a6
SLC1A6
Xtr.1.slc1a3
Aca.SLC1A3
Gga.Z.SLC1A3
Mdo.3.SLC1A3
Hsa.5.SLC1A3
Mmu.15.Slc1a3
Loc.2.SLC1A3
Dre.5.slc1a3a
Tni.Un.SLC1A3a*
Ola.9.SLC1A3a
Gac.XIII.SLC1A3a
Dre.10.slc1a3b
Gac.XIV.SLC1A3b*
Ola.2.SLC1A3b*
Tni.Un.SLC1A3b*
SLC1A3
0.2

Supplementary figure S32. Yippee-like family. The gene family is made up of the *YPEL1, YPEL2, YPEL3* and *YPEL4* genes. While our phylogeny shows that this family diverges in the time window of the basal vertebrate tetraploidizations (1R/2R), only the *YPEL1, YPEL3* and *YPEL4* branches are well-supported. The *YPEL1* branch is not resolved with regard to the species taxonomy. This is likely caused by a combination of short sequence lengths (the alignment used to create this phylogeny consist of only 145 amino acid positions), and a high degree of sequence conservation within the family. Additionally, the *YPEL2* genes, and to some extent the *YPEL1* genes, are located outside the identified paralogous chromosomal regions in several of the genomes we investigated (Figs. 3 and 4). There are putative teleost duplicates of *YPEL2* in the zebrafish genome; however, neither our phylogeny nor the chromosomal locations of these genes give conclusive support for an origin in the basal teleost tetraploidization (3R). The syntenic localization of the *YPEL2a* and *VEZF1b* genes in all teleost genomes we investigated was used to distinguish the *YPEL2a* and *YPEL2b* genes.



References

Craxton, M., 2010. A manual collection of Syt, Esyt, Rph3a, Rph3al, Doc2, and Dblc2 genes from 46 metazoan genomes -an open access resource for neuroscience and evolutionary biology. BMC Genomics 11, 37. doi:10.1186/1471-2164-11-37

Feiner, N., Meyer, A., Kuraku, S., 2014. Evolution of the vertebrate Pax4/6 class of genes with focus on its novel member, the Pax10 gene. Genome Biol. Evol. 6, 1635–1651. doi:10.1093/gbe/evu135

Ferrante, M.I., Kiff, R.M., Goulding, D.A., Stemple, D.L., 2011. Troponin T is essential for sarcomere assembly in zebrafish skeletal muscle. J. Cell Sci. 124, 565–577. doi:10.1242/jcs.071274

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39 Suppl 2, W29-37. doi:10.1093/nar/gkr367

Gesemann, M., Lesslauer, A., Maurer, C.M., Schönthaler, H.B., Neuhauss, S.C.F., 2010. Phylogenetic analysis of the vertebrate excitatory/neutral amino acid transporter (SLC1/EAAT) family reveals lineage specific subfamilies. BMC Evol. Biol. 10, 117. doi:10.1186/1471-2148-10-117

Glardon, S., Holland, L.Z., Gehring, W.J., Holland, N.D., 1998. Isolation and developmental expression of the amphioxus Pax-6 gene (AmphiPax-6): insights into eye and photoreceptor evolution. Development 125, 2701–2710. doi:10.1109/CCECE.2010.5575132

Haug-Baltzell, A., Jarvis, E.D., McCarthy, F.M., Lyons, E., 2015. Identification of dopamine receptors across the extant avian family tree and analysis with other clades uncovers a polyploid expansion among vertebrates. Front. Neurosci. 9, 1–16. doi:10.3389/fnins.2015.00361

Haug, M.F., Gesemann, M., Lazović, V., Neuhauss, S.C.F., 2015. Eumetazoan cryptochrome phylogeny and evolution. Genome Biol. Evol. 7, 601–619. doi:10.1093/gbe/evv010

Hoegg, S., Meyer, A., 2007. Phylogenomic analyses of KCNA gene clusters in vertebrates: why do gene clusters stay intact? BMC Evol. Biol. 7, 139. doi:10.1186/1471-2148-7-139

Huang, S., Tian, H., Chen, Z., Yu, T., Xu, A., 2010. The evolution of vertebrate tetraspanins: gene loss, retention, and massive positive selection after whole genome duplications. BMC Evol. Biol. 10, 306. doi:10.1186/1471-2148-10-306

Irvine, S.Q., Fonseca, V.C., Zompa, M.A., Antony, R., 2008. Cis-regulatory organization of the Pax6 gene in the ascidian Ciona intestinalis. Dev. Biol. 317, 649–659. doi:10.1016/j.ydbio.2008.01.036

Liu, C., Hu, J., Qu, C., Wang, L., Huang, G., Niu, P., Zhong, Z., Hong, F., Wang, G., Postlethwait, J.H., Wang, H., 2015. Molecular evolution and functional divergence of zebrafish (Danio rerio) cryptochrome genes. Sci. Rep. 5, 8113. doi:10.1038/srep08113

Liu, Q., Dou, S., Wang, G., Li, Z., Feng, Y., 2008. Evolution and functional divergence of monocarboxylate transporter genes in vertebrates. Gene 423, 14–22. doi:10.1016/j.gene.2008.07.003

Manousaki, T., Feiner, N., Begemann, G., Meyer, A., Kuraku, S., 2011. Co-orthology of Pax4 and Pax6 to the fly eyeless gene: Molecular phylogenetic, comparative genomic, and embryological analyses. Evol. Dev. 13, 448–459. doi:10.1111/j.1525-142X.2011.00502.x

Ocampo Daza, D., Larhammar, D., 2017. Evolution of the receptors for growth hormone, prolactin, erythropoietin and thrombopoietin in relation to the vertebrate tetraploidizations. Gen. Comp. Endocrinol. doi:10.1016/j.ygcen.2017.06.021

Qiu, H., Hildebrand, F., Kuraku, S., Meyer, A., 2011. Unresolved orthology and peculiar coding sequence properties of lamprey genes: the KCNA gene family as test case. BMC Genomics 12, 325. doi:10.1186/1471-2164-12-325

Wang, W., Zhong, J., Wang, Y.-Q., 2010. Comparative genomic analysis reveals the evolutionary conservation of Pax gene family. Genes Genet. Syst. 85, 193–206. doi:10.1266/ggs.85.193

Yamamoto, K., Fontaine, R., Pasqualini, C., Vernier, P., 2015. Classification of Dopamine Receptor Genes in Vertebrates: Nine Subtypes in Osteichthyes. Brain. Behav. Evol. 86, 164–175. doi:10.1159/000441550