

Supplementary Document for manuscript titled “Evolution of MAF and OTX families during the emergence and stabilization of rod photoreceptors across Metazoans”

Soumitra Pal^{¶,*}, Noor D. White Carreiro^{¶,2}, Zachary Batz, Leo Goubet-McCall, Anand Swaroop^{*\$}

Supplementary Figure Legends

Figure S1. Presence of annotated MAF and OTX proteins across studied species.

Heatmap illustrating the presence/absence of annotated MAF and OTX transcription factors in diverse organism groups. This panel is identical to Fig. 2A except that de novo predicted proteins have been omitted to facilitate visualization of annotation gaps.

Figure S2. Evolution of large MAF genes through vertebrate whole-genome duplications (WGDs).

Synteny plots of chromosomal regions flanking the four large MAF transcription factor genes (*cMAF*, *NRL*, *MAFA*, and *MAFB*) in six jawed vertebrates (human, laboratory mouse, opossum, western clawed frog, chicken, spotted gar), two jawless vertebrates (sea lamprey, brown hagfish), and the invertebrate outgroup (amphioxus).

The phylogeny on the left illustrates whole-genome duplication (WGD) events (labeled 1R_v, 2R_{iv}, 2R_{cy}), and is manually constructed from chromosomal segment orthology. For jawless vertebrates and amphioxus, based on multiple lines of evidence: 1) phylogeny of large MAFs, 2) chromosomal locations and orthogroup assignments of large MAF loci and flanking genes, and 3) phylogeny of Chordate Linkage Group D (CLGD) from Marletaz et al. (2024). For jawed vertebrates, it simply recapitulates the established species topology.

Colored arrows and pentagons represent chromosomal segments and genes therein, with arrow direction indicating genome assembly orientation and pentagon shape denoting transcriptional

strand specificity. Colors match human orthologs of large MAF (red) neighboring genes: *ADCYx* in light blue, *CPNEx* in pink, *JPHx* in green, *TOXy* in dark blue, *CHDx* in teal, *NDRGx* in yellow. Genomic positions are not to scale. Dotted arrows and pentagonal borders highlight absences, including *NRL* and its flanking genes in chicken.

This analysis reveals that not only the large MAF genes themselves (see Fig. 3A) but also their flanking genes duplicated in parallel, confirming the whole-genome nature of the events. Although the large MAF phylogeny in Figure 3A resolved only two jawless-specific subclades after the shared first vertebrate WGD, synteny of the larger region uncovers jawless-specific triplications, consistent with recent studies.

Figure S3. Evolution of small MAF genes through vertebrate whole-genome duplications (WGDs).

Synteny plots of chromosomal regions flanking the three small MAF transcription factor genes (*MAFF*, *MAFG*, and *MAFK*) in six jawed vertebrates (human, laboratory mouse, opossum, western clawed frog, chicken, spotted gar), two jawless vertebrates (sea lamprey, brown hagfish), and the invertebrate outgroup (amphioxus).

The phylogeny on the left illustrates whole-genome duplication (WGD) events (labeled 1R_v, 2R_{jv}, 2R_{cy}), and is manually constructed from chromosomal segment orthology. For jawless vertebrates and amphioxus, based on multiple lines of evidence: 1) phylogeny of small MAFs, 2) chromosomal locations and orthogroup assignments of small MAF loci and flanking genes, and 3) phylogeny of Chordate Linkage Group H (CLGH) from Marletaz et al. (2024). For jawed vertebrates, it simply recapitulates the established species topology.

Colored arrows and pentagons represent chromosomal segments and genes therein, with arrow direction indicating genome assembly orientation and pentagon shape denoting transcriptional strand specificity. Colors match human orthologs of small MAF (red) neighboring genes: *SYNGRx*

in light blue, *RBFOXx* in pink, *ANKRDx* in green, *CARDx* in dark blue, *CYTHx* in teal, *RACx* in yellow). Genomic positions are not to scale.

This analysis reveals that not only the small MAF genes themselves (see Fig. 3B) but also their flanking genes duplicated in parallel, confirming the whole-genome nature of the events. Although the small MAF phylogeny in Figure 3B resolved only two jawless-specific subclades after the shared first vertebrate WGD, synteny of the larger region uncovers jawless-specific triplications, consistent with recent studies.

Abbreviations: WGD, whole-genome duplication; v, vertebrate; jv, jawed vertebrate; cy, cyclostome; *ANKRDx*, Ankyrin Repeat Domain x; *SYNGRx*, Synaptogyrin x; *CYTHx*, Cytohesin x; *RBFOXy*: RNA Binding Fox-1 Homolog y; *CARDx*, Caspase Recruitment Domain Family Member x; *RACx*, Ras-Related C3 Botulinum Toxin Substrate x.

Figure S4. Evolution of OTX genes through vertebrate whole-genome duplications (WGDs).

(A) Phylogenetic tree of OTX proteins from 86 representative species (out of total 181 studied), rooted with *Drosophila* CNC (Cap'n'collar) as outgroup. The tree highlights three distinct clades in jawed vertebrates (*OTX1*, *OTX2*, and *CRX*) and four distinct clades in jawless vertebrates. These clades appear to have arisen from a single common ancestral duplication event (yellow), followed by independent lineage-specific duplications in jawed vertebrates (blue) and jawless vertebrates (orange). Branches connecting the root to the outgroup and the ancestral OTX were manually shortened for clarity. (B) Synteny plots of chromosomal regions flanking the three OTX transcription factor genes (*OTX1*, *OTX2*, and *CRX*) in six jawed vertebrates (human, laboratory mouse, opossum, western clawed frog, chicken, spotted gar), two jawless vertebrates (sea lamprey, brown hagfish), and the invertebrate outgroup (amphioxus).

The phylogeny on the left illustrates whole-genome duplication (WGD) events (labeled 1R_v, 2R_{jv}, 2R_{cy}), and is manually constructed from chromosomal segment orthology. For jawless

vertebrates and amphioxus, based on multiple lines of evidence: 1) phylogeny of small MAFs, 2) chromosomal locations and orthogroup assignments of small MAF loci and flanking genes, and 3) phylogeny of Chordate Linkage Group A1 (CLGA1) from Marletaz et al. (2024). For jawed vertebrates, it simply recapitulates the established species topology.

Colored arrows and pentagons represent chromosomal segments and genes therein, with arrow direction indicating genome assembly orientation and pentagon shape denoting transcriptional strand specificity. Colors match human orthologs of OTX (red) neighboring genes: *VRKx* in light blue, *RTNx* in pink, *SLC8Ax* in green, *MAP4Kx* in dark blue, *STRNx* in teal, *LTBPx* in yellow). Genomic positions are not to scale.

This analysis reveals that not only the OTX genes themselves (panel A) but also their flanking genes duplicated in parallel, confirming the whole-genome nature of the events. Although the OTX phylogeny in panel A resolved only three jawless-specific subclades (two on one side and one on the other) after the shared first vertebrate WGD, synteny of the larger region uncovers jawless-specific triplications, consistent with recent studies.

Abbreviations: WGD, whole-genome duplication; v, vertebrate; jv, jawed vertebrate; cy, cyclostome; *VRKx*, Vaccinia Related Kinase x; *RTNx*, Reticulon x; *SLC8Ax*, Solute Carrier Family 8 Member Ax; *MAP4Kx*, Mitogen-Activated Protein Kinase x; *STRNx*: Striatin x; *LTBPx*, Latent Transforming Growth Factor Beta Binding Protein x.

Figure S5. Sequence evolution of vertebrate NRL reveals positive selection in mammals.

(A) HyPhy aBSREL analysis mapped onto the vertebrate NRL phylogeny. Branch colors indicate the proportion of sites evolving under one, two, or three distinct ω (dN/dS) categories: blue for mostly purifying selection ($\omega < 1$), gray/black for neutral evolution ($\omega \approx 1$), and red/orange for positive (diversifying) selection ($\omega > 1$). Branch thickness corresponds to the strength of statistical support (p-value). Branches with p-values < 0.05 are highlighted.

(B) Tabular summary of aBSREL results for the top 10 proteins, ranked by p-value.

Figure S6. Acquisition of novel mammal-specific cis-regulatory elements upstream of NRL containing OTX2/CRX binding sites that drive photoreceptor-specific expression.

(A) UCSC 100-vertebrate conservation tracks (hg38, Dec. 2013) showing three conserved regions (C in blue, B in green, A in red) upstream of NRL that represent mammal-specific evolutionary innovations. All three regions exhibit sparse or absent conservation in non-mammals; region B is exclusively mammal-specific among the analyzed vertebrates.

(B) Epigenomic tracks reproduced from Cherry et al. (2020) demonstrating regulatory activity of these regions in the retina. Tracks include ATAC-seq (chromatin accessibility), CRX and OTX2 ChIP-seq (transcription factor binding), and active enhancer histone marks (H3K4me2 and H3K27ac). Prominent CRX and OTX2 binding occurs at regions B and C.

(C) In vivo reporter assays reproduced from Kautzmann et al. (2011) showing that among these elements particularly region B is sufficient to drive *Nrl* promoter activity in photoreceptors. Neonatal mouse retinas were electroporated with constructs containing different combinations of the conserved *Nrl* promoter regions driving *Nrlp*-EGFP (green); CAG-mCherry (red) marks transfected cells. Representative P14 retinal sections are shown for: (a) region A only; (b) A + B plus intervening region; (c) A + B + C plus intervening regions. Merged images are on the right. Scale bar: 20 μ m.

Figure S7. Co-evolution analysis of OTX homeobox domains with large MAF domains across vertebrate clades.

Extended correlation matrices showing Pearson correlation coefficients (PCC) between similarity matrices of OTX homeobox domains and all combinations of large MAF domains (rows). The bottom row reproduces the results from Fig. 5B for reference. Matrices are presented separately for all jawed vertebrates (left column), mammals (middle column), and non-mammalian jawed

vertebrates (right column). Cell values indicate PCC; color intensity reflects $-\log_{10}(\text{p-value})$ of the PCC (shared color scale across all matrices).

Figure S1

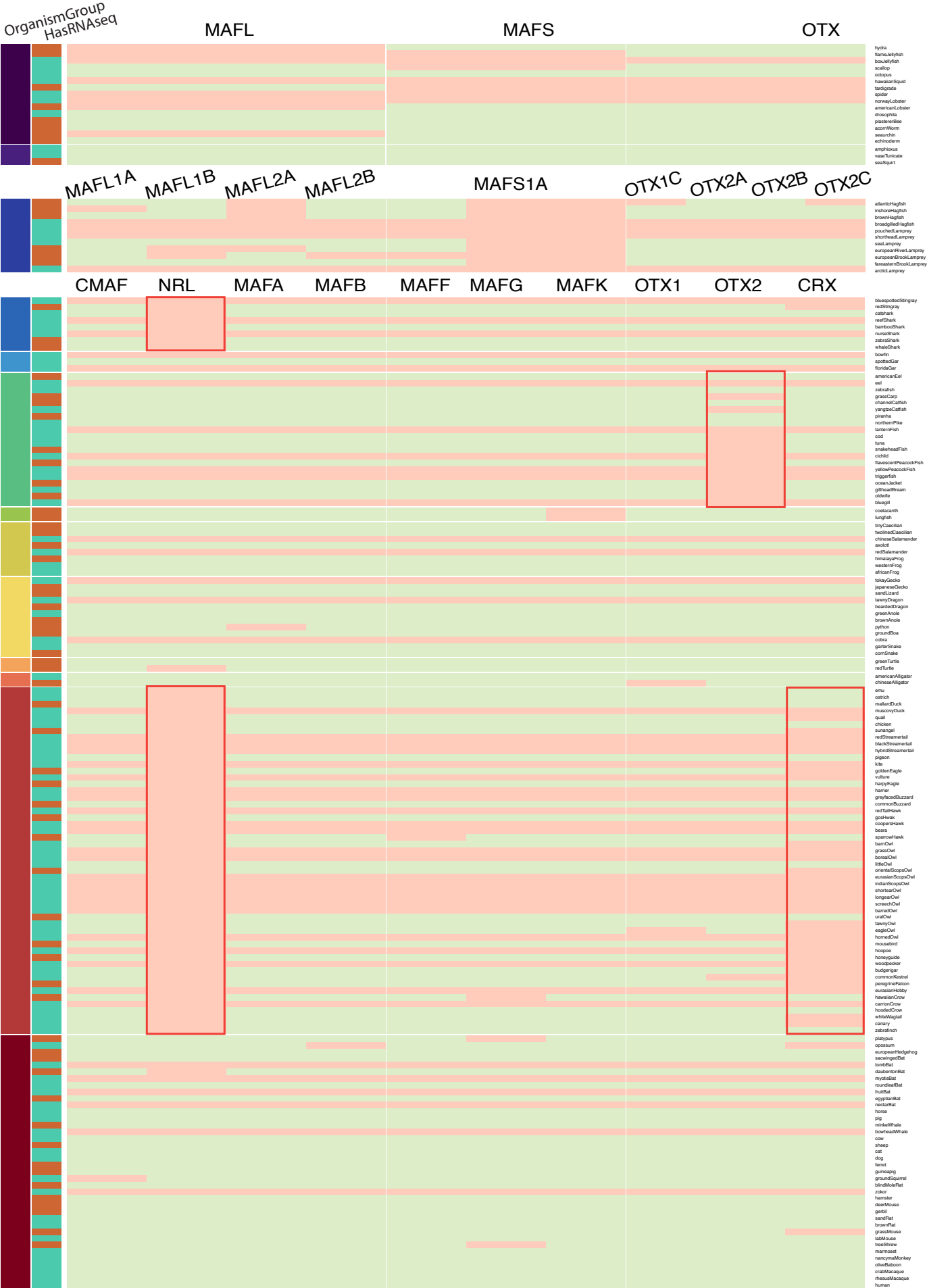


Figure S2

Human orthologs of large MAF neighbors

						
CMAF	ADCY7	CPNE7	JPH3	TOX3	CHD9	NDRG4
NRL	ADCY4	CPNE6	JPH4	TOX4	CHD8	NDRG2
MAFA	ADCY8	CPNE3	JPH1	TOX	CHD7	NDRG1
MAFB	ADCY3	CPNE1	JPH2	TOX2	CHD6	NDRG3

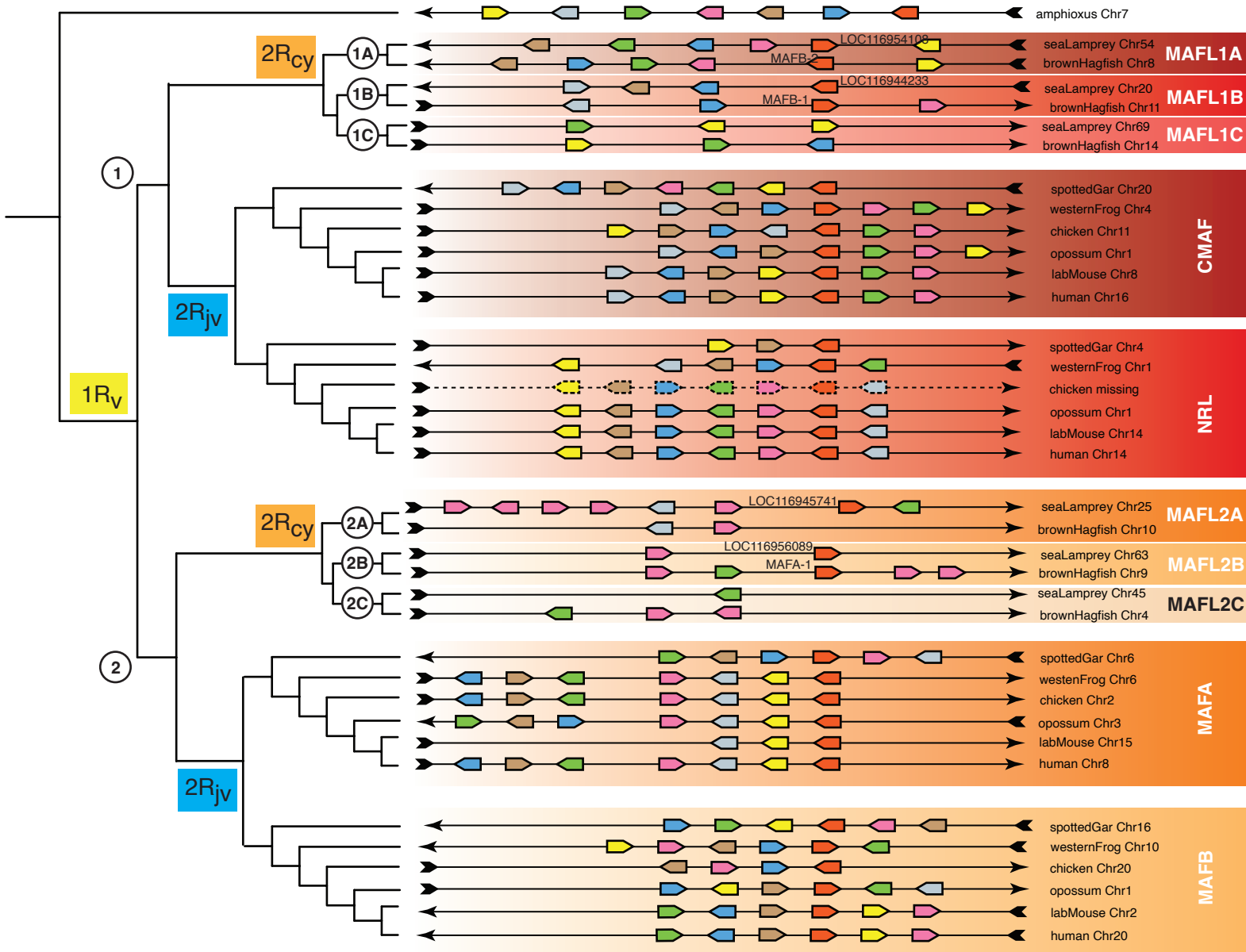
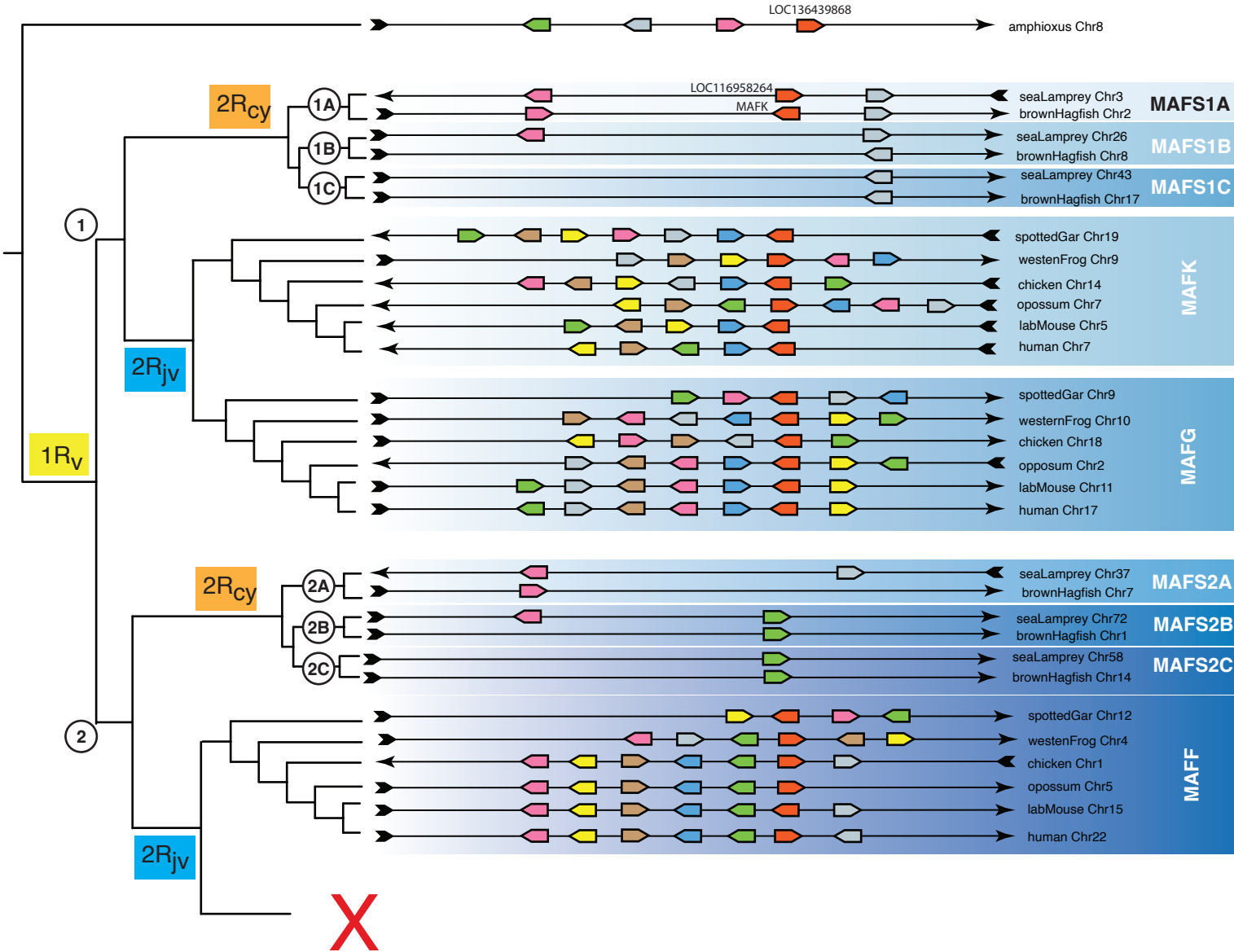


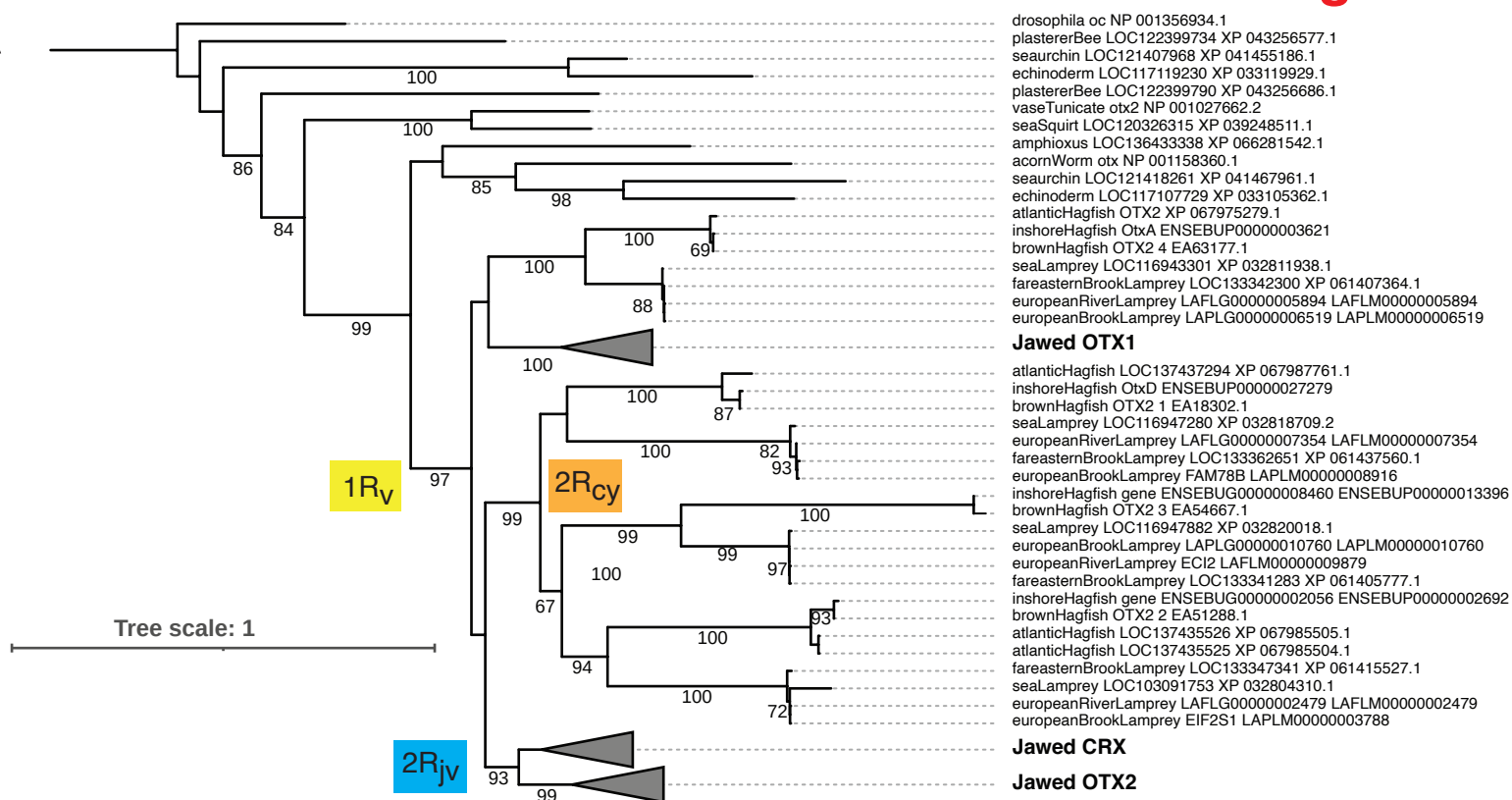
Figure S3

Human orthologs of small MAF neighbors

						
MAFF	SYNGR1	RBFOX2	ANKRD54	CARD10	CYTH4	RAC2
MAFG	SYNGR2	RBFOX3	ANKRD40	CARD14	CYTH1	RAC3
MAFK	SYNGR3	RBFOX1	ANKRD61	CARD11	CYTH3	RAC1



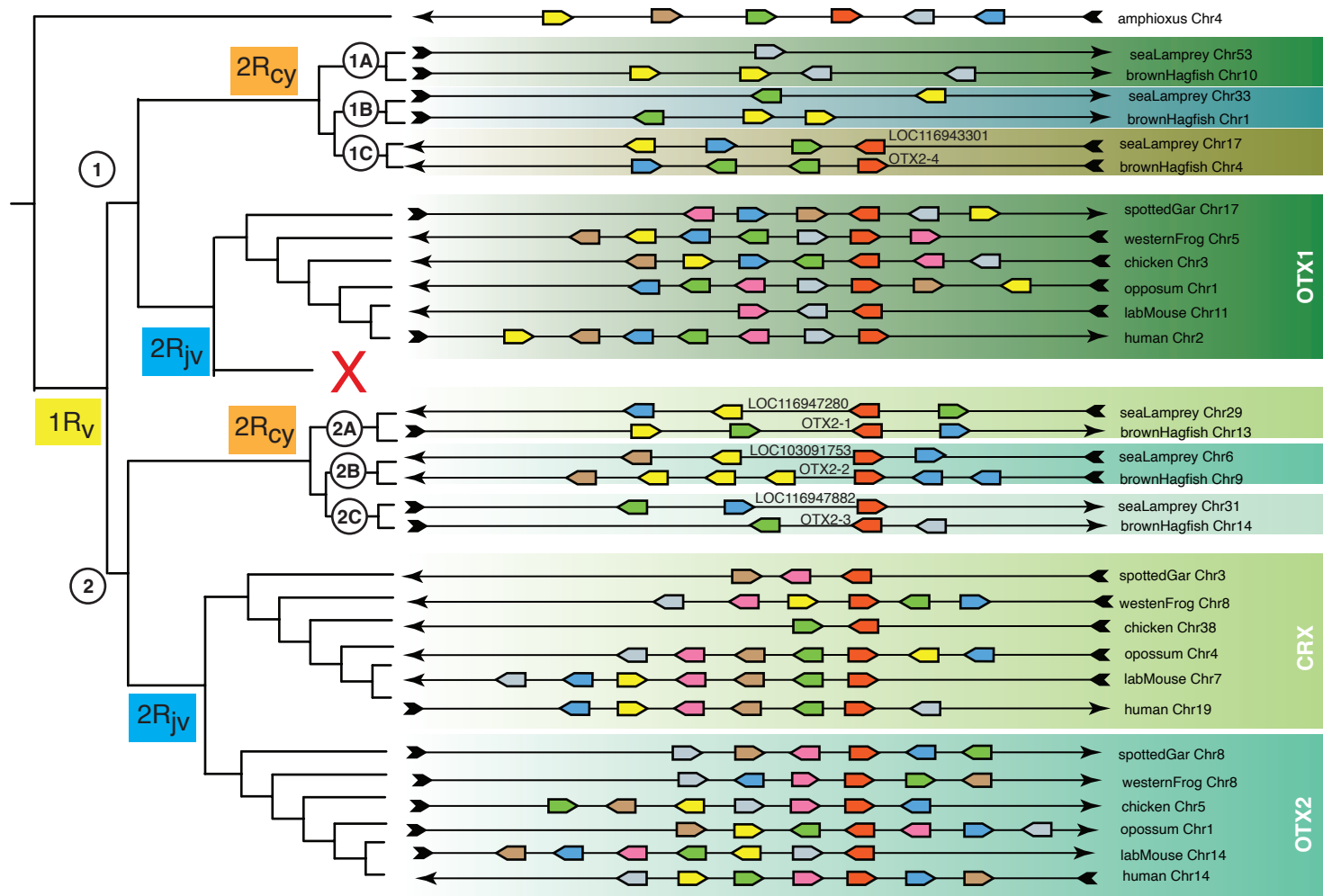
A



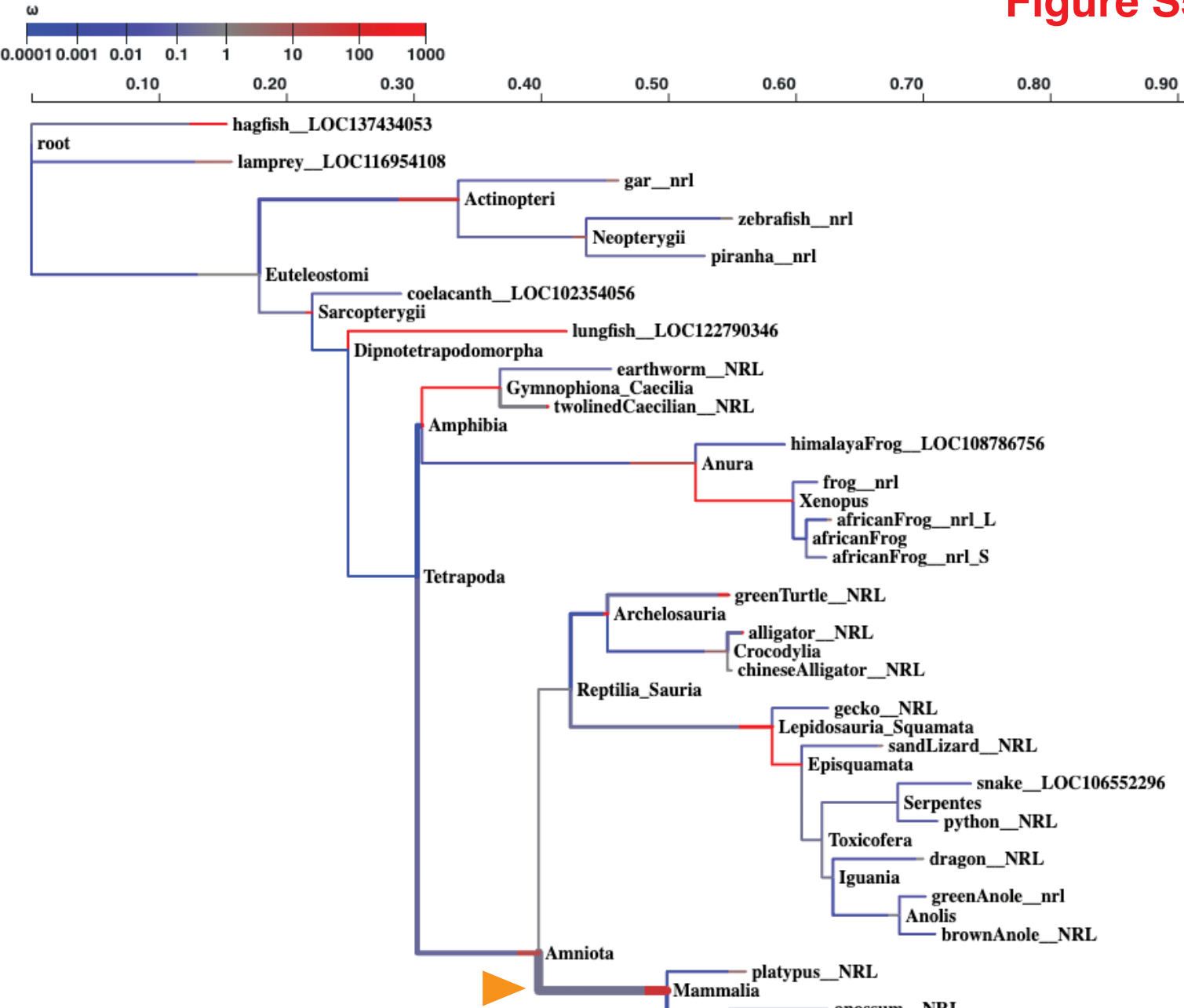
B

Human orthologs of OTX neighbors

OTX1	VRK2	RTN4	SLC8A1	MAP4K3	STRN	LTBP1
OTX2	VRK1	RTN1	SLC8A3	MAP4K5	STRN3	LTBP2
CRX	VRK3	RTN2	SLC8A2	MAP4K1	STRN4	LTBP4

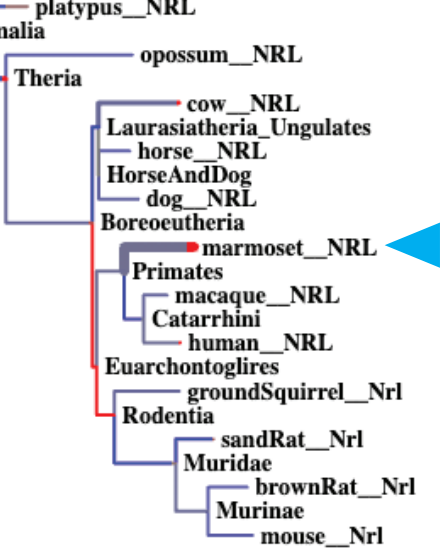


A

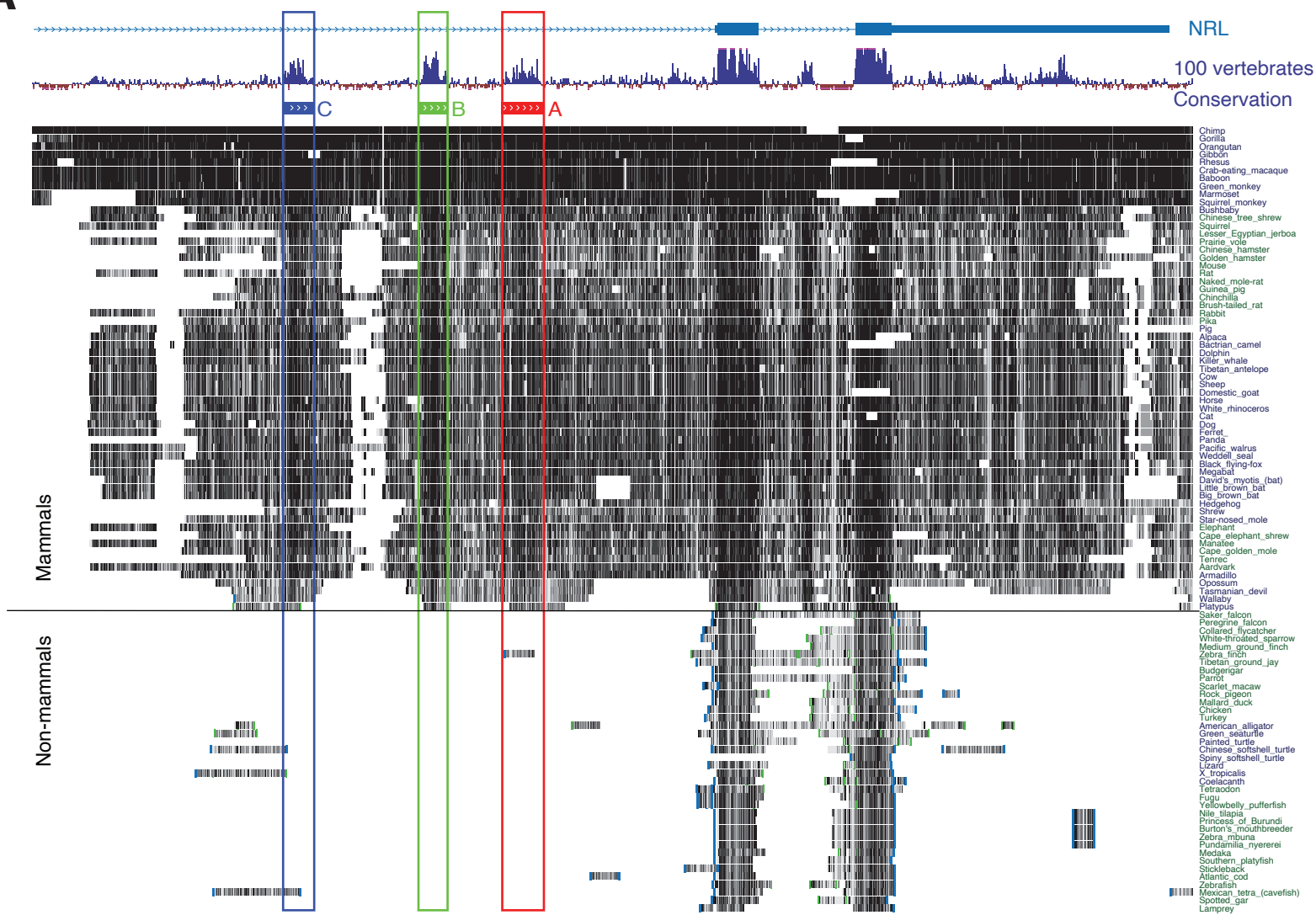


B

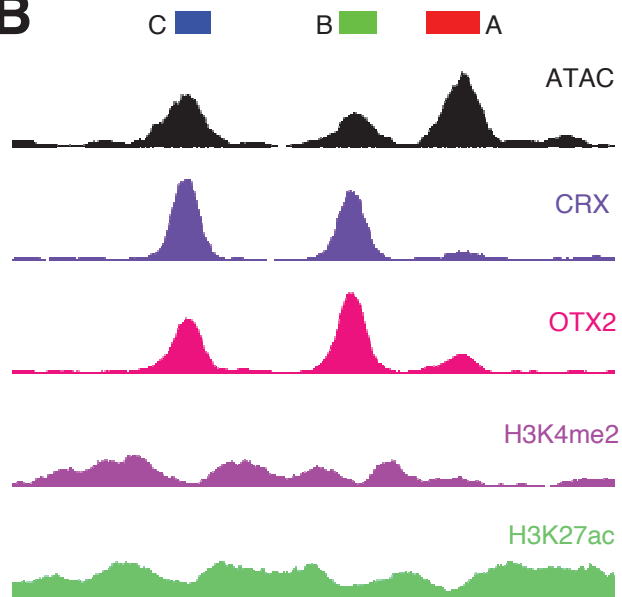
branch	▲ p-value	sites	rates	ω distribution
Mammalia	0	6	2	0.3524 (82.732%) 52.46 (17.268%) Mean = 9.350, CoV = 2.106
marmoset__NRL	0.001	4	2	0.3052 (90.330%) 340.2 (9.6696%) Mean = 33.17, CoV = 3.028
Amniota	0.063	3	2	0.1630 (82.619%) 18.70 (17.381%) Mean = 3.385, CoV = 2.075
Amphibia	0.079	3	2	0.000 (94.937%) 225.4 (5.0630%) Mean = 11.41, CoV = 4.330
Archelosauria	0.128	2	2	0.000 (88.143%) 1916 (11.857%) Mean = 227.1, CoV = 2.727
twolinedCaecilian__NRL	0.166	5	2	0.7572 (96.231%) 99.40 (3.7691%) Mean = 4.475, CoV = 4.198
alligator__NRL	0.188	1	2	0.1306 (99.023%) 513.2 (0.97703%) Mean = 5.144, CoV = 9.812
greenTurtle__NRL	0.217	2	2	0.1595 (90.748%) 86.87 (9.2523%) Mean = 8.182, CoV = 3.071
cow__NRL	0.3	5	2	0.1890 (96.695%) 1529 (3.3046%) Mean = 50.72, CoV = 5.389
Actinopteri	0.305	8	2	0.01572 (70.047%) 50.42 (29.953%) Mean = 15.11, CoV = 1.528



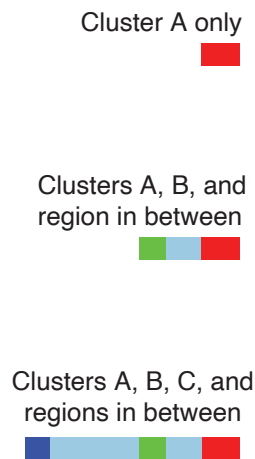
A



B



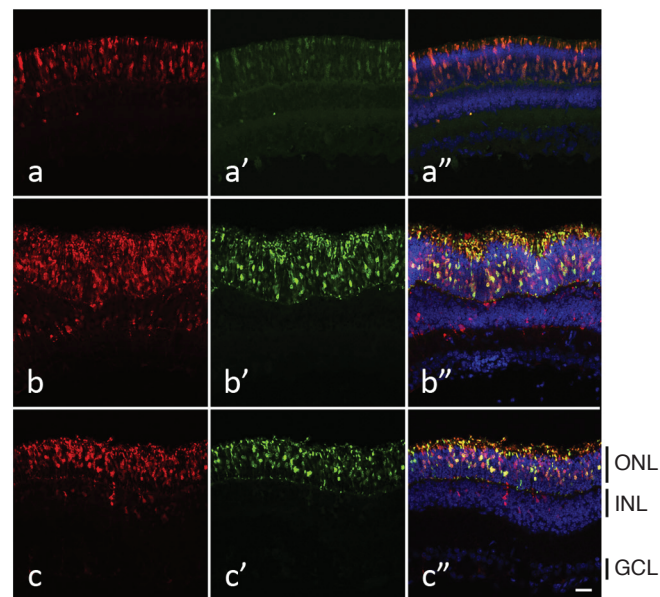
C



CAG-mCherry

Nr1p-EGFP

Merge



Non-mammal

