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

Soumitra Pal[1,¶,*], Noor D. White Carreiro[1,2,¶], Zachary Batz[1], Leo Goubet-McCall[1], Caroline Duffie Judy[3-5], Anand Swaroop[1,*$]

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 $1R_v$, $2R_{jv}$, 2Rcy), 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 $1R_v$, $2R_{jv}$, 2Rcy), 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 $1R_v$, $2R_{jv}$, 2Rcy), 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 (ω > 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 μm.

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 -log10(p-value) of the PCC (shared color scale across all matrices).

Figure S1

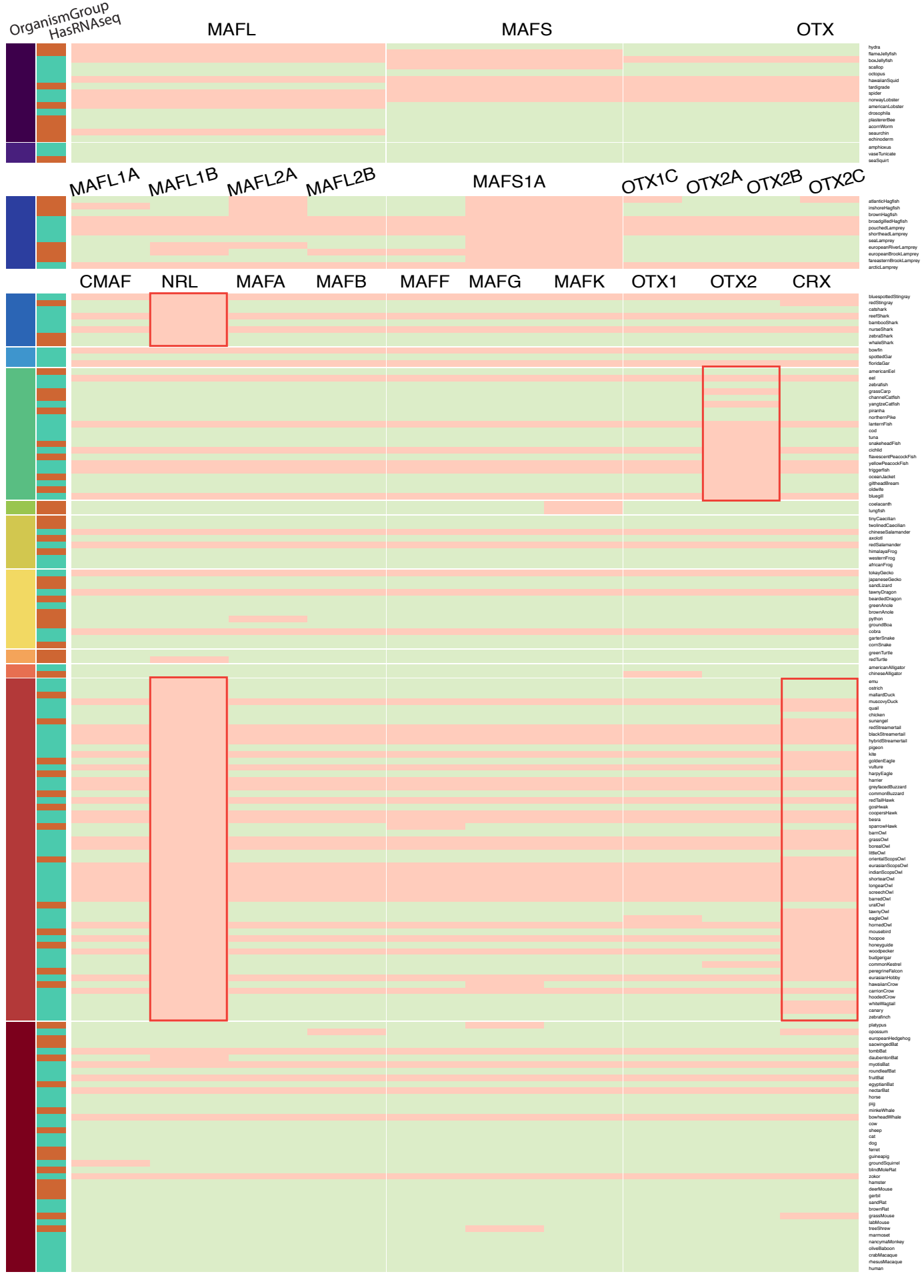

OrganismGroup
HasRNAseq
MAFL
MAFS
OTX
MAFL1A
MAFL1B
MAFL2A
MAFL2B
MAFS1A
OTX1C
OTX2A
OTX2B
OTX2C
CMAF
NRL
MAFA
MAFB
MAFF
MAFG
MAFK
OTX1
OTX2
CRX
OrganismGroup
Nonchordates
Protochordates
Agnathans
Cartilaginous
Ray–finned excl teleosts
Teleosts
Lobe–finned
Amphibians
Squamata
Turtles
Crocodiles
Birds
Mammals
HasRNAseq
Yes
No
Protein presence
Annotated
None

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

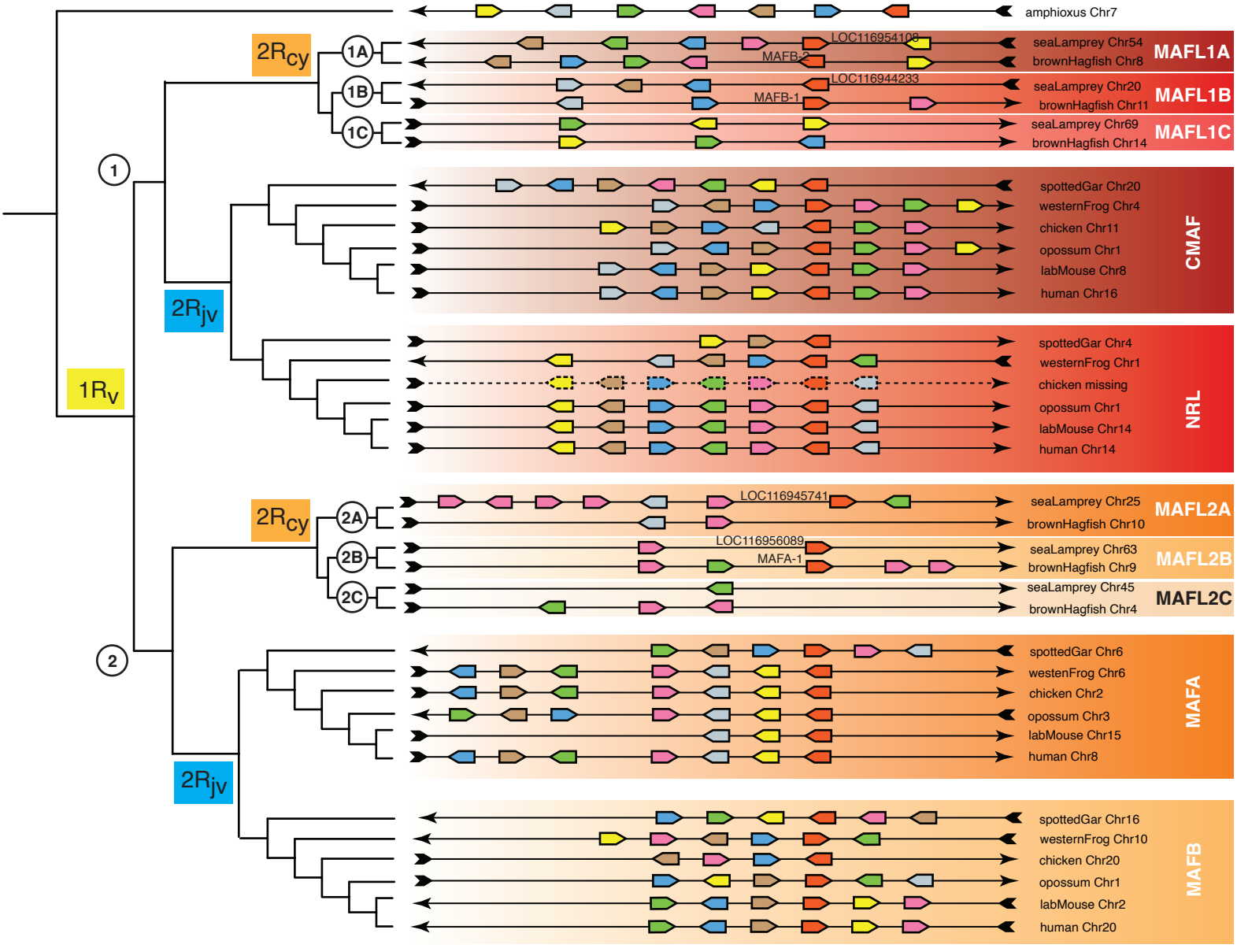


Figure S3

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

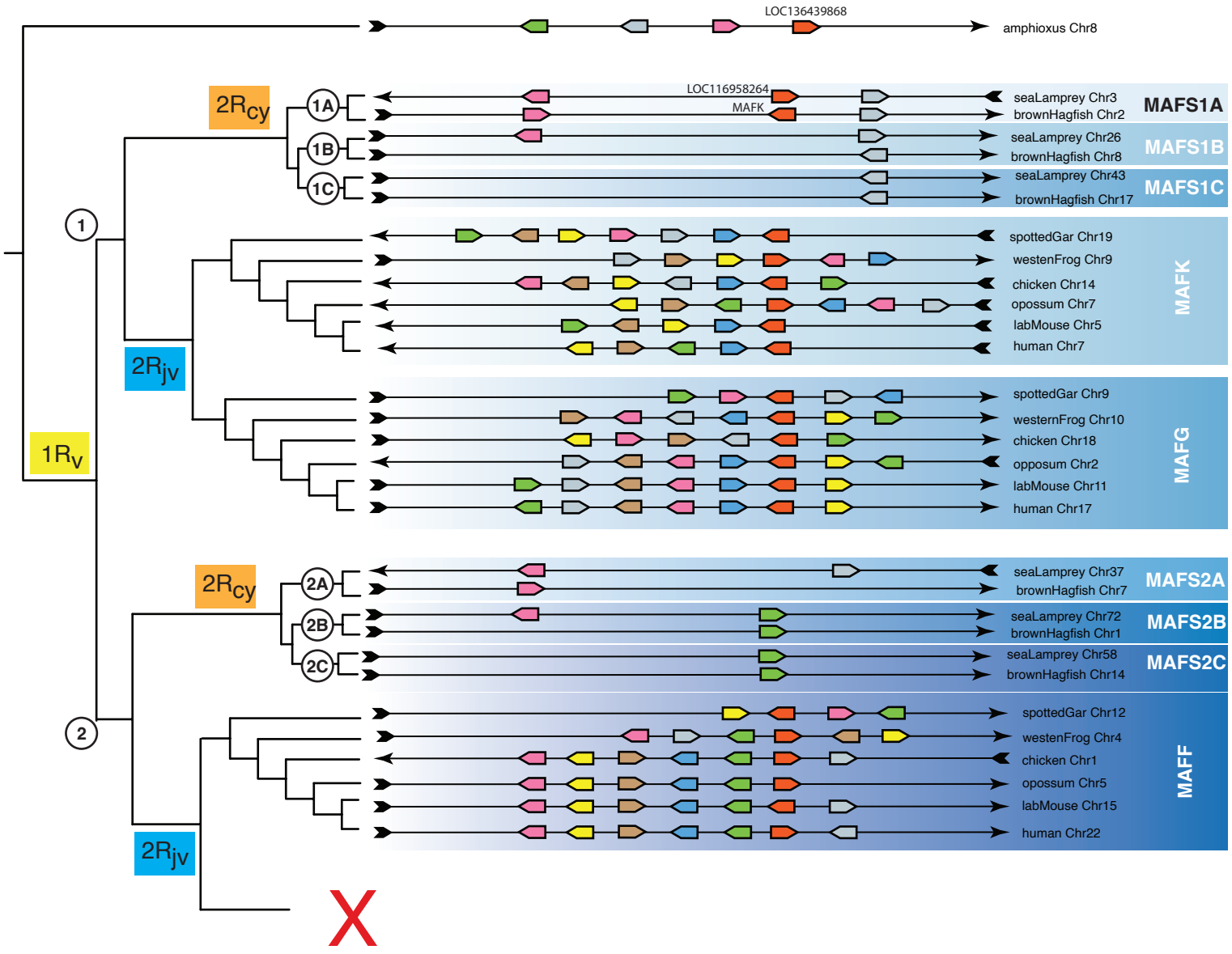


Figure S4

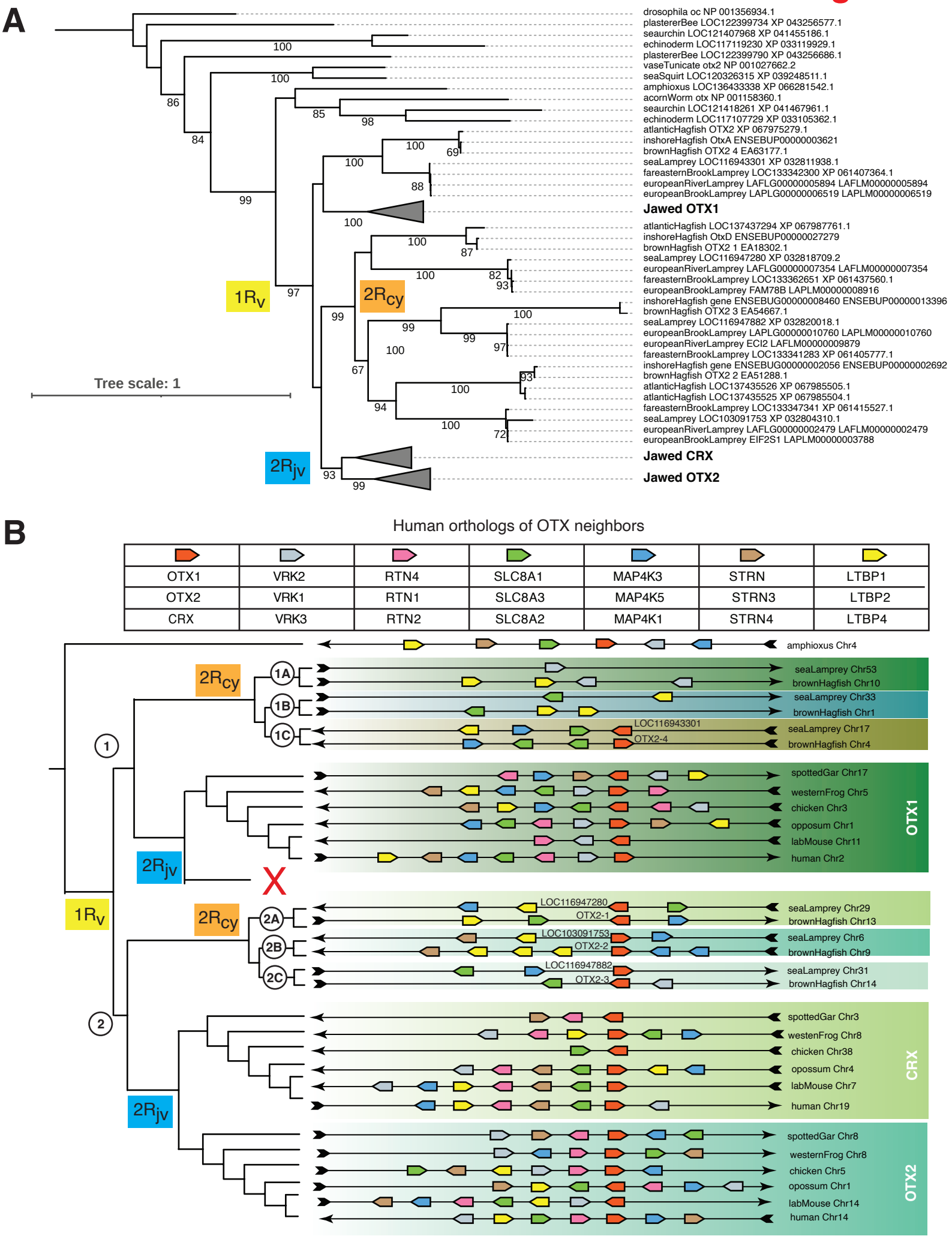
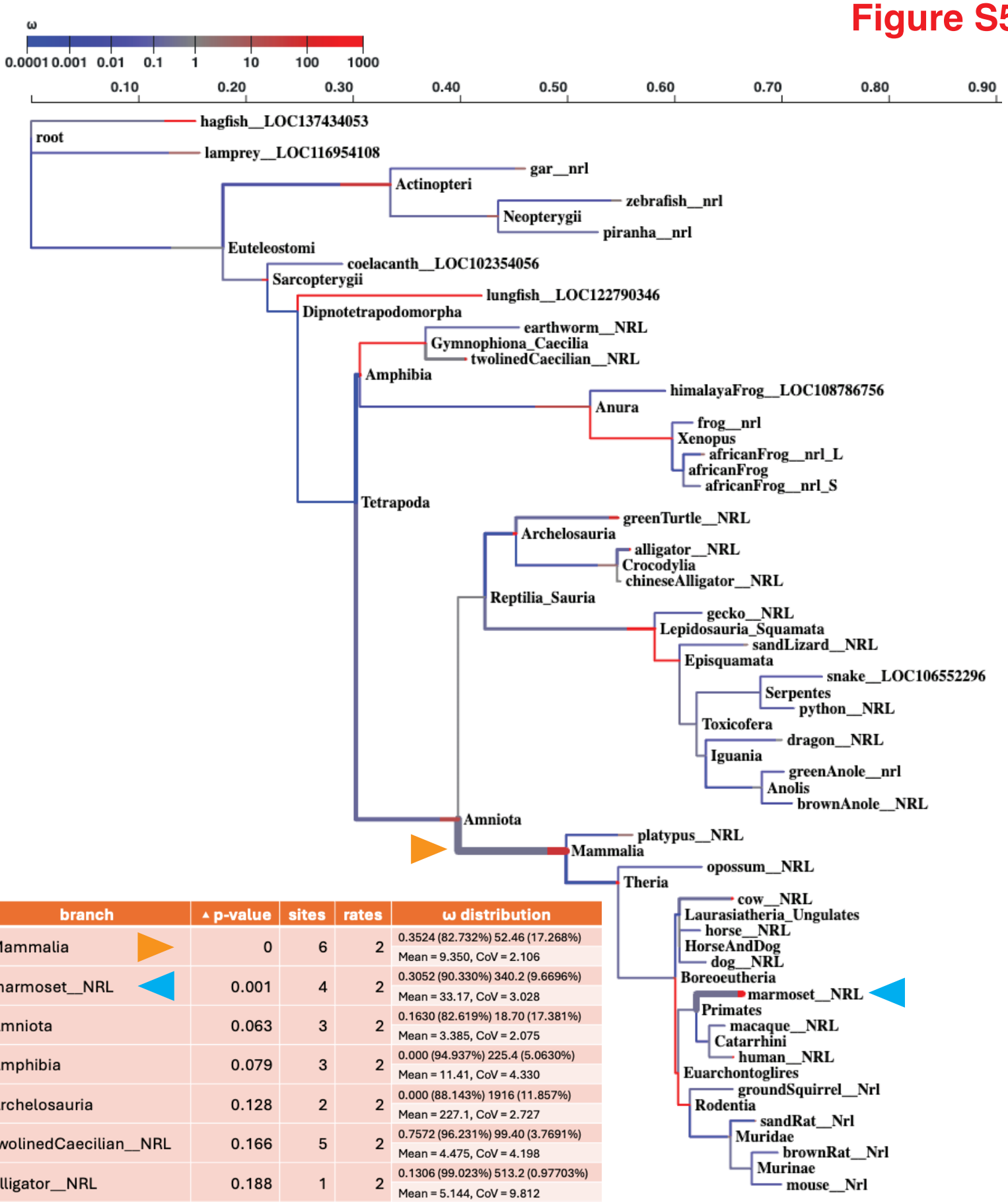

A
drosophila oc NP 001356934.1
plastererBee LOC122399734 XP 043256577.1
seaurchin LOC121407968 XP 041455186.1
echinoderm LOC117119230 XP 033119929.1
plastererBee LOC122399790 XP 043256686.1
vaseTunicate otx2 NP 001027662.2
seaSquirt LOC120326315 XP 039248511.1
amphioxus LOC136433338 XP 066281542.1
acornWorm otx NP 001158360.1
seaurchin LOC121418261 XP 041467961.1
echinoderm LOC117107729 XP 033105362.1
atlanticHagfish OTX2 XP 067975279.1
inshoreHagfish OtxA ENSEBUP00000003621
brownHagfish OTX2 4 EA63177.1
seaLamprey LOC116943301 XP 032811938.1
fareasternBrookLamprey LOC133342300 XP 061407364.1
europeanRiverLamprey LAFLG00000005894 LAFLM00000005894
europeanBrookLamprey LAPLG00000006519 LAPLM00000006519
Jawed OTX1
atlanticHagfish LOC137437294 XP 067987761.1
inshoreHagfish OtxD ENSEBUP00000027279
brownHagfish OTX2 1 EA18302.1
seaLamprey LOC116947280 XP 032818709.2
europeanRiverLamprey LAFLG00000007354 LAFLM00000007354
fareasternBrookLamprey LOC133362651 XP 061437560.1
europeanBrookLamprey FAM78B LAPLM00000008916
inshoreHagfish gene ENSEBUG00000008460 ENSEBUP00000013396
brownHagfish OTX2 3 EA54667.1
seaLamprey LOC116947882 XP 032820018.1
europeanBrookLamprey LAPLG00000010760 LAPLM00000010760
europeanRiverLamprey ECI2 LAFLM00000009879
fareasternBrookLamprey LOC133341283 XP 061405777.1
inshoreHagfish gene ENSEBUG00000002056 ENSEBUP00000002692
brownHagfish OTX2 2 EA51288.1
atlanticHagfish LOC137435526 XP 067985505.1
atlanticHagfish LOC137435525 XP 067985504.1
fareasternBrookLamprey LOC133347341 XP 061415527.1
seaLamprey LOC103091753 XP 032804310.1
europeanRiverLamprey LAFLG00000002479 LAFLM00000002479
europeanBrookLamprey EIF2S1 LAPLM00000003788
Jawed CRX
Jawed OTX2
1Rv
2Rcy
2Rjv
Tree scale: 1
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
amphioxus Chr4
1A
1B
1C
seaLamprey Chr53
brownHagfish Chr10
seaLamprey Chr33
brownHagfish Chr1
LOC116943301
OTX2-4
seaLamprey Chr17
brownHagfish Chr4
spottedGar Chr17
westernFrog Chr5
chicken Chr3
opposum Chr1
labMouse Chr11
human Chr2
OTX1
2A
2B
2C
LOC116947280
OTX2-1
seaLamprey Chr29
brownHagfish Chr13
LOC103091753
OTX2-2
seaLamprey Chr6
brownHagfish Chr9
LOC116947882
OTX2-3
seaLamprey Chr31
brownHagfish Chr14
spottedGar Chr3
westenFrog Chr8
chicken Chr38
opossum Chr4
labMouse Chr7
human Chr19
CRX
spottedGar Chr8
westernFrog Chr8
chicken Chr5
opossum Chr1
labMouse Chr14
human Chr14
OTX2

Figure S5

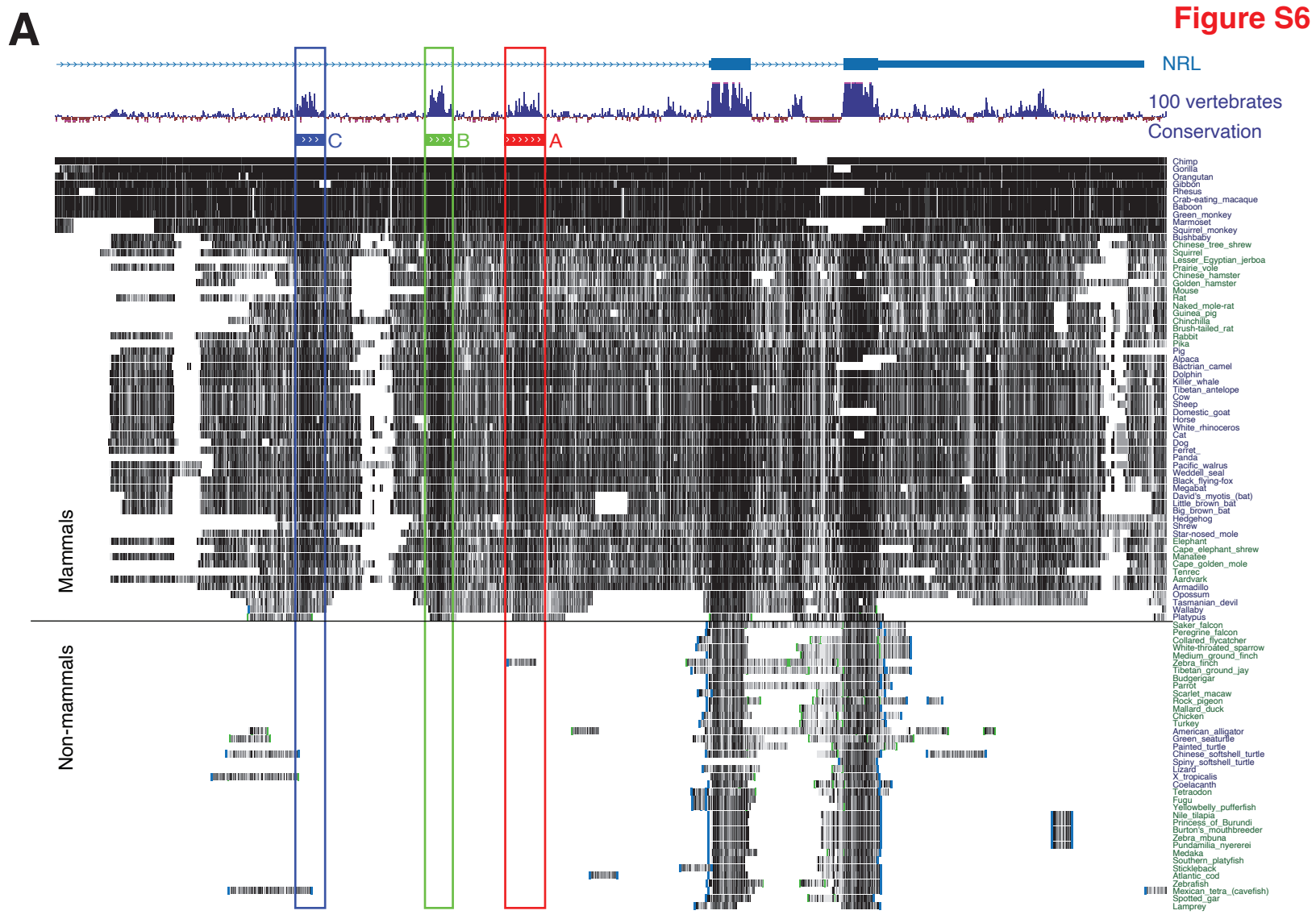
A

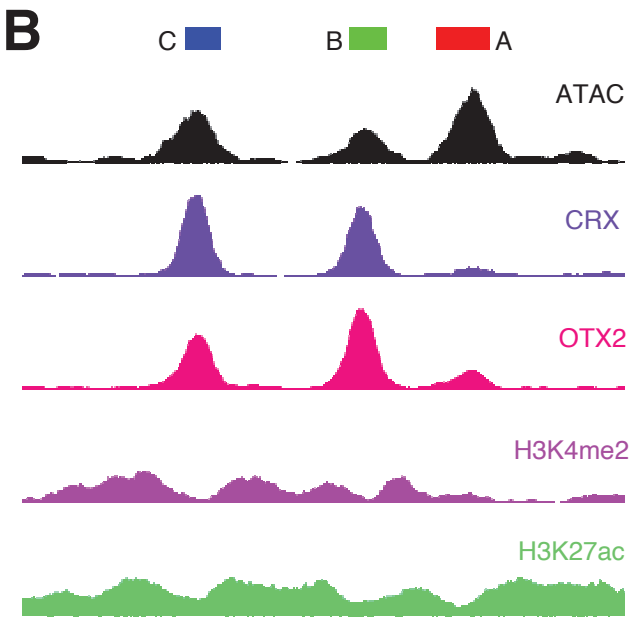


B

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

Figure S6


A
NRL
100 vertebrates
Conservation
C
B
A
Mammals
Non-mammals
Chimp
Gorilla
Orangutan
Gibbon
Rhesus
Crab-eating_macaque
Baboon
Green_monkey
Marmoset
Squirrel_monkey
Bushbaby
Chinese_tree_shrew
Squirrel
Lesser_Egyptian_jerboa
Prairie_vole
Chinese_hamster
Golden_hamster
Mouse
Rat
Naked_mole-rat
Guinea_pig
Chinchilla
Brush-tailed_rat
Rabbit
Pika
Pig
Alpaca
Bactrian_camel
Dolphin
Killer_whale
Tibetan_antelope
Cow
Sheep
Domestic_goat
Horse
White_rhinoceros
Cat
Dog
Ferret
Panda
Pacific_walrus
Weddell_seal
Black_flying-fox
Megabat
David's_myotis_(bat)
Little_brown_bat
Big_brown_bat
Hedgehog
Shrew
Star-nosed_mole
Elephant
Cape_elephant_shrew
Manatee
Cape_golden_mole
Tenrec
Aardvark
Armadillo
Opossum
Tasmanian_devil
Wallaby
Platypus
Saker_falcon
Peregrine_falcon
Collared_flycatcher
White-throated_sparrow
Medium_ground_finch
Zebra_finch
Tibetan_ground_jay
Budgerigar
Parrot
Scarlet_macaw
Rock_pigeon
Mallard_duck
Chicken
Turkey
American_alligator
Green_seaturtle
Painted_turtle
Chinese_softshell_turtle
Spiny_softshell_turtle
Lizard
X_tropicalis
Coelacanth
Tetraodon
Fugu
Yellowbelly_pufferfish
Nile_tilapia
Princess_of_Burundi
Burton's_mouthbreeder
Zebra_mbuna
Pundamilia_nyererei
Medaka
Southern_platyfish
Stickleback
Atlantic_cod
Zebrafish
Mexican_tetra_(cavefish)
Spotted_gar
Lamprey


B
C
B
A
ATAC
CRX
OTX2
H3K4me2
H3K27ac

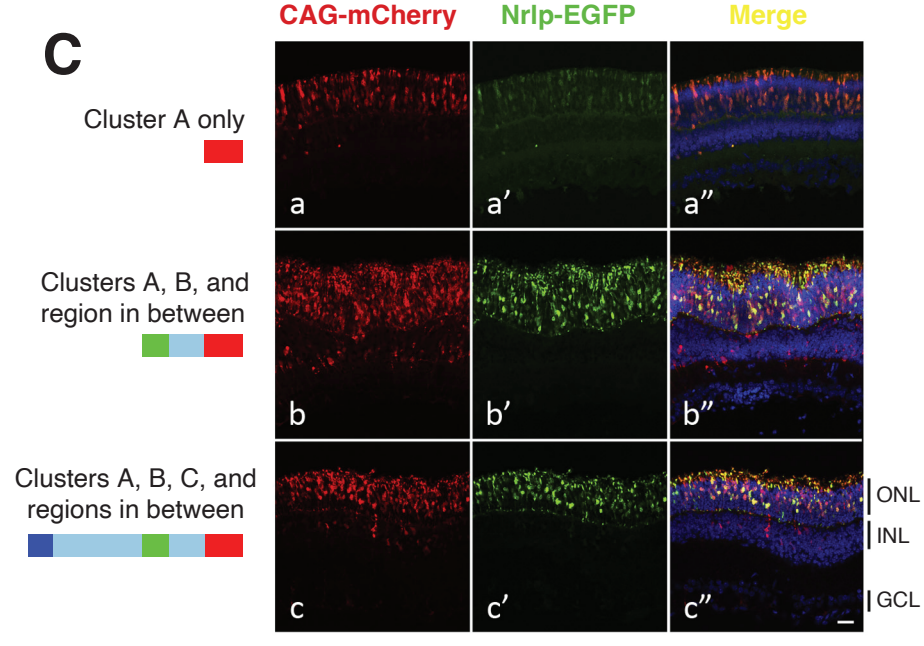
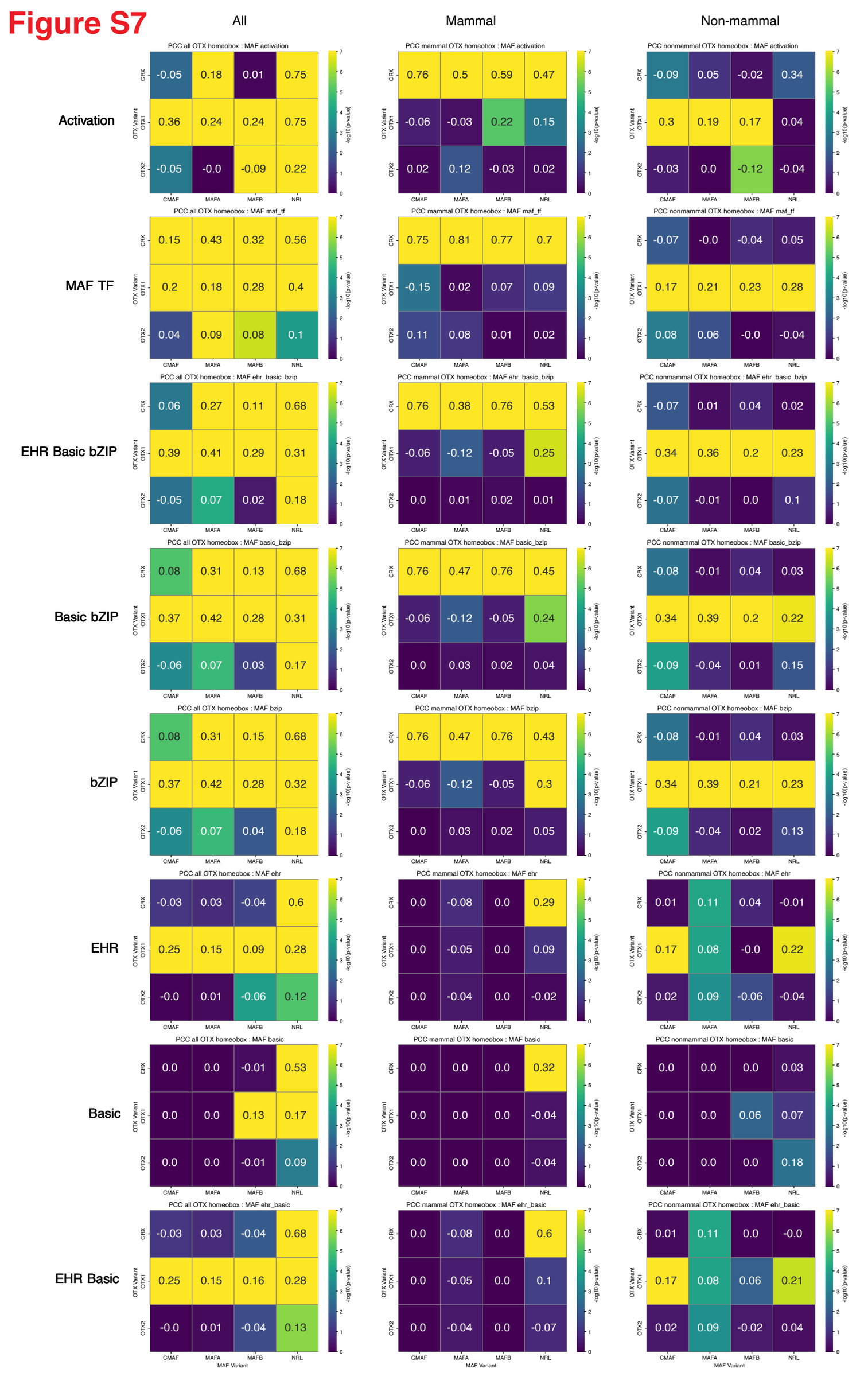

C
Cluster A only
Clusters A, B, and region in between
Clusters A, B, C, and regions in between
CAG-mCherry
Nrlp-EGFP
Merge
a
a′
a″
b
b′
b″
c
c′
c″
ONL
INL
GCL

Figure S7


All
Mammal
Non-mammal
Activation
MAF TF
EHR Basic bZIP
Basic bZIP
bZIP
EHR
Basic
EHR Basic
PCC all OTX homeobox : MAF activation
PCC mammal OTX homeobox : MAF activation
PCC nonmammal OTX homeobox : MAF activation
PCC all OTX homeobox : MAF maf_tf
PCC mammal OTX homeobox : MAF maf_tf
PCC nonmammal OTX homeobox : MAF maf_tf
PCC all OTX homeobox : MAF ehr_basic_bzip
PCC mammal OTX homeobox : MAF ehr_basic_bzip
PCC nonmammal OTX homeobox : MAF ehr_basic_bzip
PCC all OTX homeobox : MAF basic_bzip
PCC mammal OTX homeobox : MAF basic_bzip
PCC nonmammal OTX homeobox : MAF basic_bzip
PCC all OTX homeobox : MAF bzip
PCC mammal OTX homeobox : MAF bzip
PCC nonmammal OTX homeobox : MAF bzip
PCC all OTX homeobox : MAF ehr
PCC mammal OTX homeobox : MAF ehr
PCC nonmammal OTX homeobox : MAF ehr
PCC all OTX homeobox : MAF basic
PCC mammal OTX homeobox : MAF basic
PCC nonmammal OTX homeobox : MAF basic
PCC all OTX homeobox : MAF ehr_basic
PCC mammal OTX homeobox : MAF ehr_basic
PCC nonmammal OTX homeobox : MAF ehr_basic
OTX Variant
CRX
OTX1
OTX2
CMAF
MAFA
MAFB
NRL
MAF Variant
-log10(p-value)