

Supplemental Information for:

Movement of transposable elements contributes to cichlid diversity

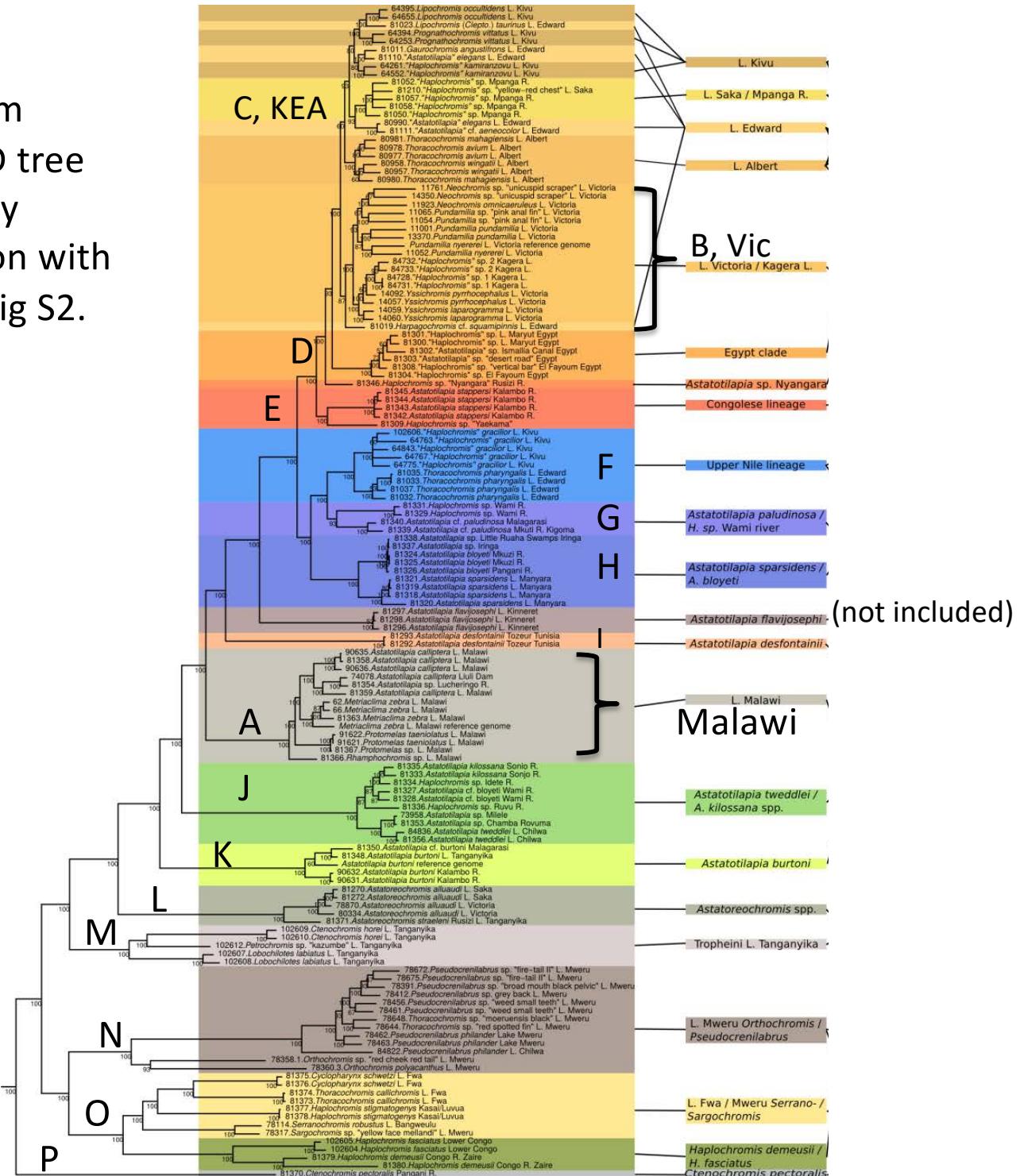
Karen L. Carleton¹⁺, Matthew A Conte¹, Milan Malinsky^{2,3}, Sri Pratima Nandamuri^{1*}, Benjamin A.

Sandkam^{1&}, Joana I Meier^{4,5,6,%}, Salome Mwaiko^{4,5}, Ole Seehausen^{4,5}, Thomas D Kocher¹

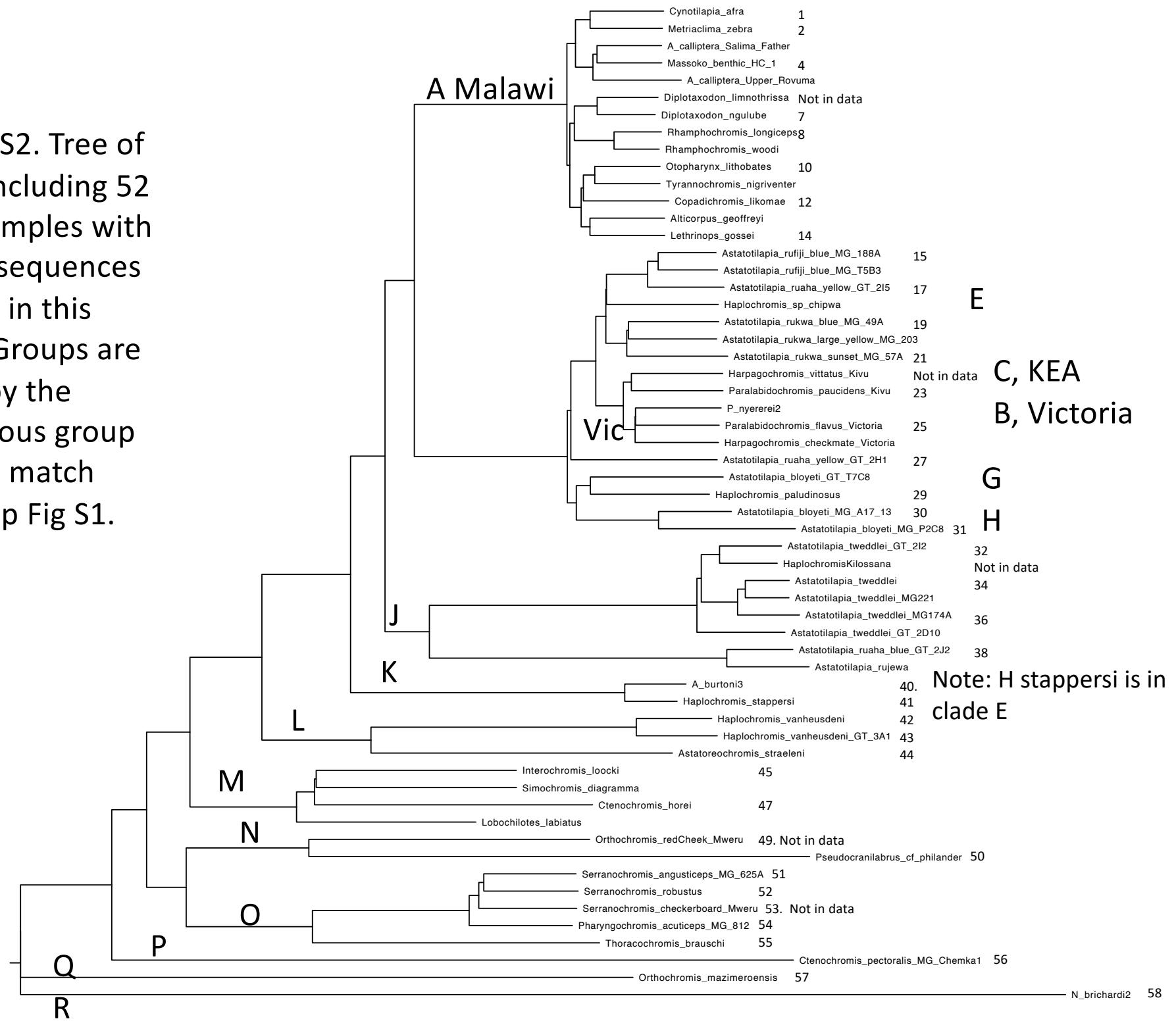
Table of Contents:

Supp Fig S1 Victoria RAD tree	Page 2
Supp Fig S2 Malawi plus tree	Page 3
Supp Fig S3A-D Indels in large genomic regions for 4 loci	Page 4-7
Supp Fig S4A-D TE annotation for 4 indel regions	Page 8-11
Supp Fig S5A-C Comparison of Rx1 indel in <i>C pectoralis</i>	Page 12-14
Supp Fig S6A-D TEs in sequence for dating indels	Page 15-18
Supp Fig S7A-D Trees dating indel origins	Page 19-22
Supp Table S1	Separate file
Supp Table S2 Sequencing primers	Page 23
Supp Table S3 Genomic location of 4 regulatory indels	Page 24-25
Supp Table S4 Genomic sequence of 4 regulatory indels	Page 26-27
Supp Table S5 Genomic frequency of 3 TE families	Page 28
Supp Table S6 Divergence time estimate for 4 indels	Separate file
Supp Table S7 Summary of opsin gene correlations	Page 29

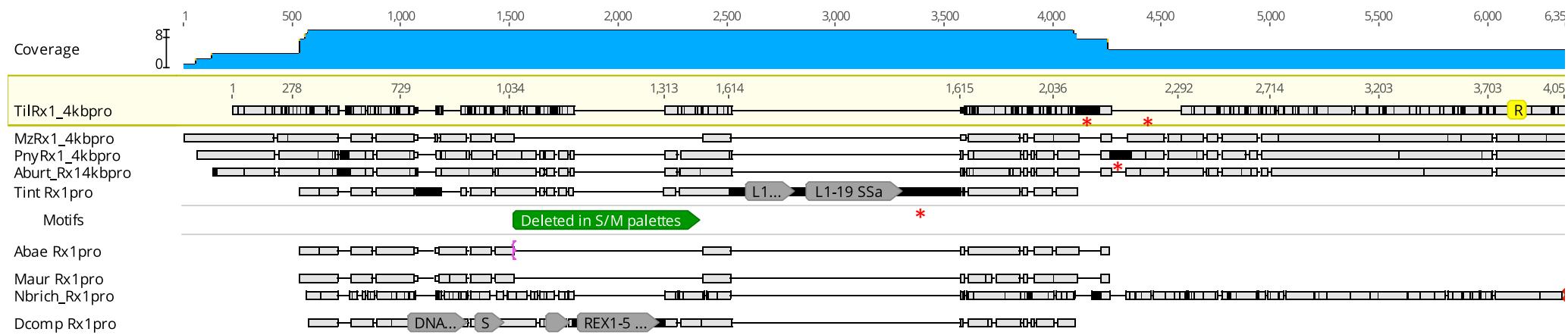
Supp Fig S1. Tree from Meier et al 2017 RAD tree with nodes labeled by groups for comparison with cichlid taxa in Supp Fig S2.



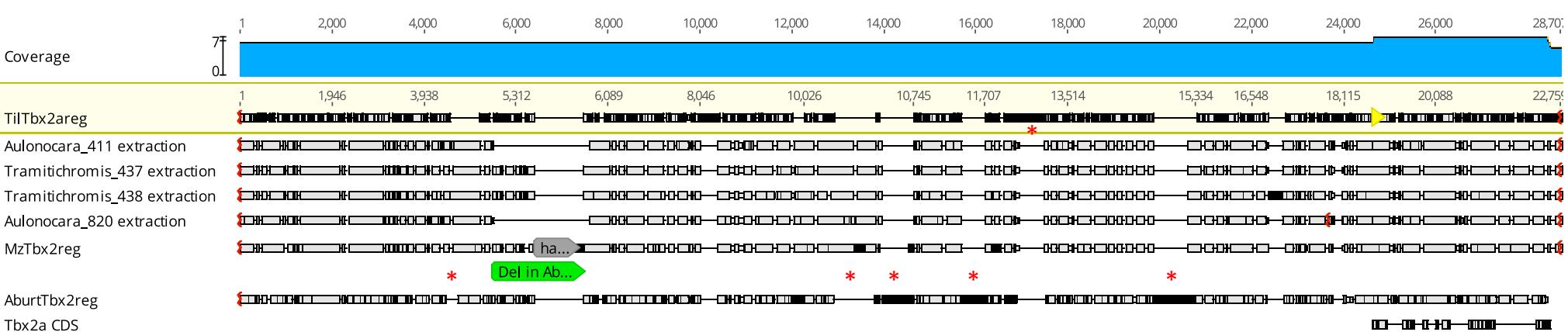
Supp Fig S2. Tree of 58 taxa including 52 cichlid samples with genome sequences searched in this project. Groups are labeled by the homologous group that they match from Supp Fig S1.



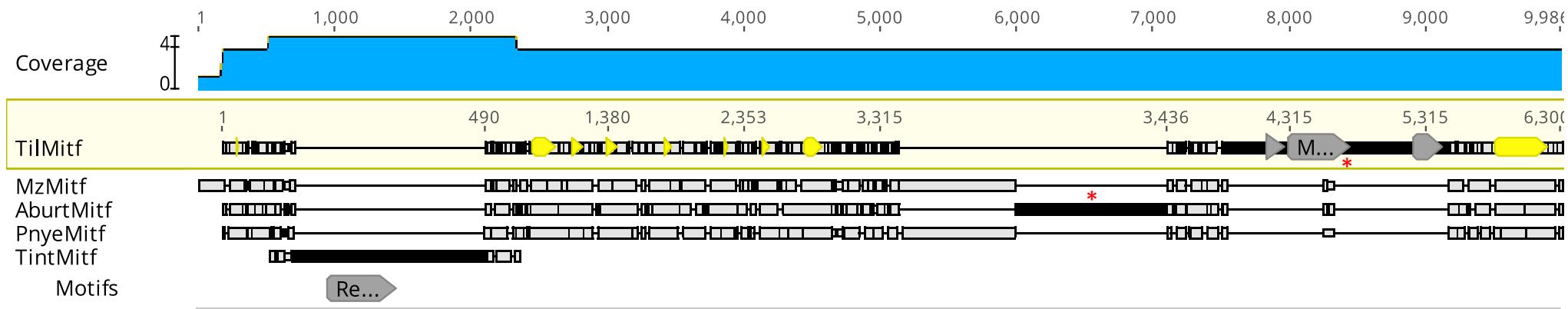
Supp Fig S3A. Transposable element and indel locations for the broader Rx1 region. Species include tilapia *O. niloticus* (Til), *M. zebra* (Mz), *P. nyererei* (Pny), *A. burtoni* (Aburt), *Tramitichromis intermedius* (Tint), *Aulonocara baenschi* (Abae), *Melanochromis auratus* (Maur), *Neolamprologus brichardi* (Nbrich) and *Dimidiochromis compressiceps* (Dcomp). Translational start site is at 3795 bp in tilapia (yellow box). The key regulatory variation is the insertion in *D. compressiceps* and the corresponding deletion (green box). However, there is an additional large insertion in *T. intermedius* and smaller indels (*) in tilapia and *P. nyererei*



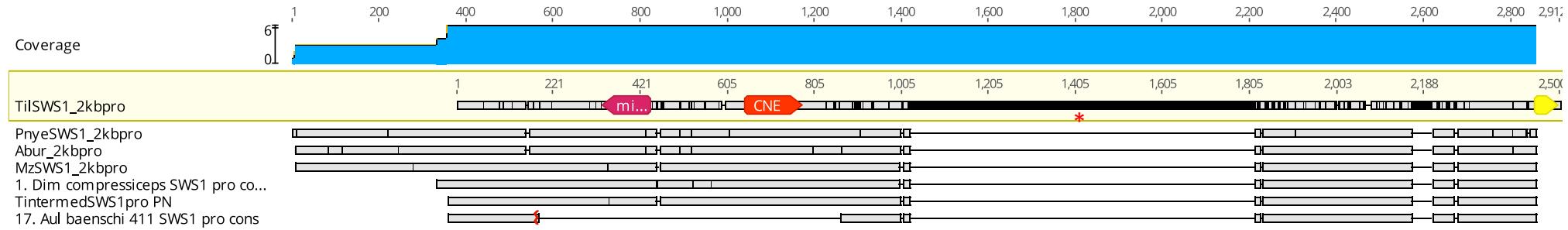
Supp Fig S3B. Transposable element and indel locations for the broader Tbx2a region. Species include tilapia *O. niloticus* (Til), *A. burtoni* (Aburt), *M. zebra* (Mz), *Aulonocara baenschi* (Aulonocara) and *Tramitichromis intermedius* (Tramitichromis). Tbx2a exons are shown as black boxes in the bottom row, which start near 18731 bp of tilapia (yellow arrow). The key regulatory variation is the insertion in *M. zebra* (gray arrow) and the corresponding deletion in *Aulonocara* (green arrow). However, there are small insertions in *A. burtoni*, and tilapia (*).



Supp Fig S3C. Transposable element and indel locations for the broader Mitf region. Species include tilapia *O. niloticus* (Til), *M. zebra* (Mz), *A. burtoni* (Aburt), *P. nyererei* (Pnye), and *Tramitichromis intermedius* (Tint). Mitf exons are shown in yellow starting at tilapia base 117. The key regulatory variation is the insertion in *T. intermedius* in intron 1 (gray arrow). However, there are other indels in *A. burtoni* and tilapia (*).

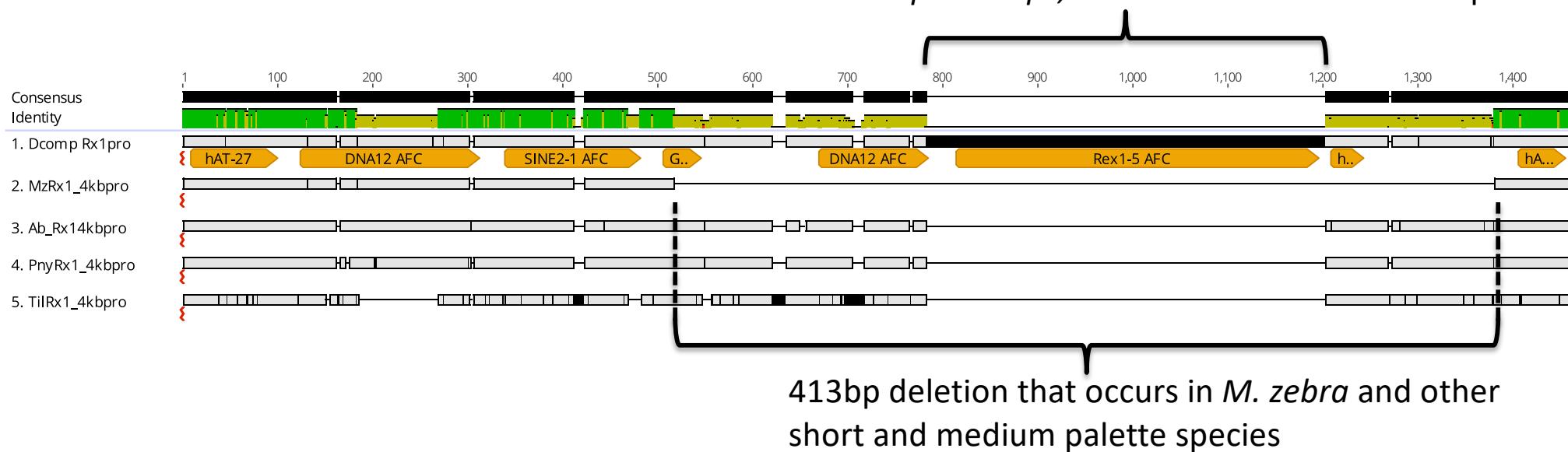


Supp Fig S3D. Indel locations for the broader SWS1 region. Species include the tilapia *O. niloticus* (Til), *P. nyererei* (Pnye), *A. burtoni* (Abur), *M. zebra* (Mz), *D. compressiceps* (Dim compressiceps), *Tramitichromis intermedius* (Tintermed) and *A. baenschi* (Aul baenschi). The first SWS1 exon is shown in yellow starting at tilapia base 2443. The key regulatory change is the deletion in *A. baenschi* which overlaps the key regulatory elements shown in red, miRNA-729 and the CNE. However, there is a large insertion in tilapia (*).



Supp Fig S4A. Transposable element and indel locations for the Rx1 region for *D. compressiceps* (Dcomp) aligned with *M. zebra* (Mz), *A. burtoni* (Ab), *P. nyererei* (Pny) and the tilapia, *O. niloticus* (Til). Translational start site is at 4189 bp (not shown).

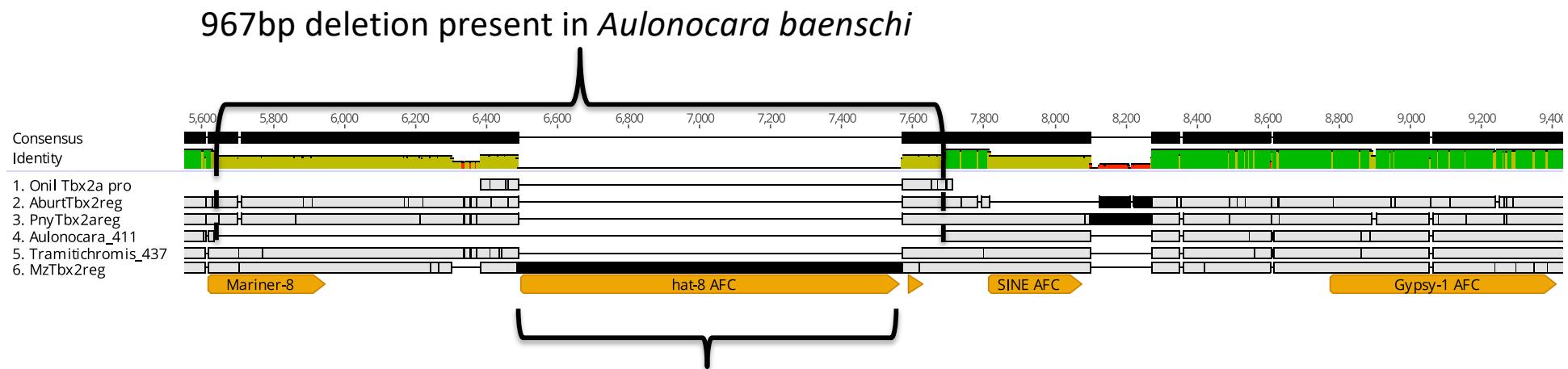
384 bp is of the 421 bp insertion in *D. compressiceps*, is a Rex1 nonLTR retrotransposon



Giri CENSOR table

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
Dcomp_Rx1	10	103	<u>hAT-27_DR</u>	2652	2741	DNA/hAT	c	0.7444	1.3333	250
Dcomp_Rx1	124	311	<u>DNA12_AFC</u>	1	214	DNA	d	0.7801	1.4783	669
Dcomp_Rx1	336	471	<u>SINE2-1_AFC</u>	7	163	NonLTR/SINE /SINE2	d	0.7857	1.2222	533
Dcomp_Rx1	493	534	<u>Gypsy-26_Ad-I</u>	794	834	LTR/Gypsy	d	0.8333	2.0000	241
Dcomp_Rx1	642	746	<u>DNA12_AFC</u>	1553	1676	DNA	d	0.7850	2.1111	338
Dcomp_Rx1	774	1157	<u>Rex1-5_AFC</u>	450	761	NonLTR/Rex1	c	0.9019	6.0000	2093
Dcomp_Rx1	1163	1205	<u>hAT-N138B_DR</u>	1406	1449	DNA/hAT	d	0.8182	1.7500	230
Dcomp_Rx1	1363	1415	<u>hATm-11_HM</u>	1795	1847	DNA/hAT	d	0.8148	1.7500	235

Supp Fig S4B. Transposable element and indel locations for the Tbx2a upstream regulatory region for tilapia, *O. niloticus* (Til), *A. burtoni* (Aburt), *P. nyererei* (Pny), *Aulonocara baenschi* (Aulonocara), *Tramitichromis intermedius* (Tramitichromis), and *M. zebra* (Mz). Translational start site is at 27,130 bp (not shown).

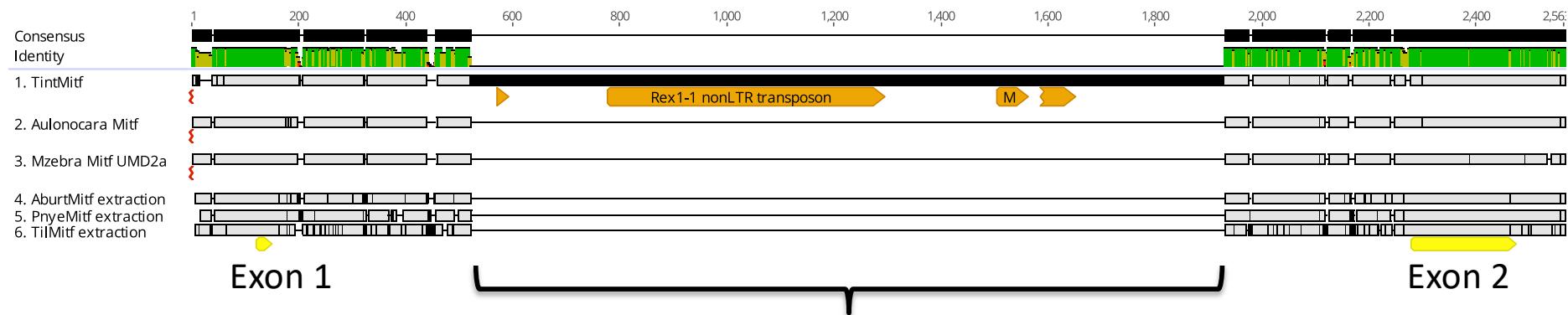


Of the 1081 bp insertion in *M. zebra*, 1073 bp is an haT-8 DNA transposon known in African cichlids (AFC)

Giri CENSOR table

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
MzTbx2indel	45	378	Mariner-8 SSA	1279	1630	DNA/Mariner	d	0.8018	2.0000	1631
MzTbx2indel	844	1916	hAT-8 AFC	1	1128	DNA/hAT	d	0.9498	2.0000	7577
MzTbx2indel	1936	1982	Mariner-12 LCh	432	476	DNA/Mariner	c	0.8043	1.6000	237
MzTbx2indel	2160	2431	SINE AFC	41	328	NonLTR/SINE/SINE2	c	0.8199	1.7727	1227
MzTbx2indel	2945	3585	Gypsy-1 AFC-I	1090	1720	LTR/Gypsy	d	0.7405	1.6915	2539

Supp Fig S4C. Transposable element and indel locations for the Mitf region for *Tramitichromis intermedius* (Tint), *Aulonocara baenschi* (Aulonocara), and *M. zebra* (Mzebra), *A. burtoni* (Aburt), *P. nyererei* (Pnye), and tilapia, *O. niloticus* (Til). Exons shown as yellow arrows.

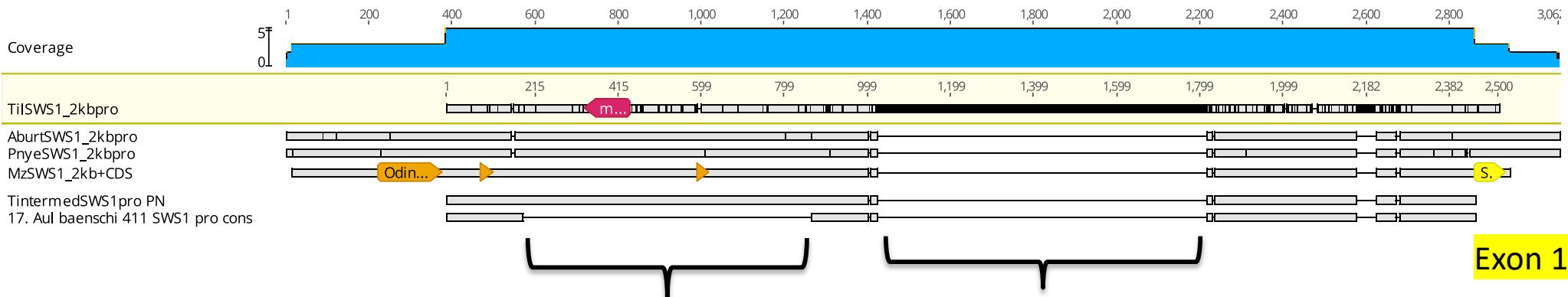


A 1408 bp insertion in the first intron of Mitf includes a 523bp Rex1 TE in *Tramitichromis intermedius*. Intron 1 in the other species is 678 bp in length.

Giri CENSOR table

<u>Name</u>	<u>From</u>	<u>To</u>	<u>Name</u>	<u>From</u>	<u>To</u>	<u>Class</u>	<u>Dir</u>	<u>Sim</u>	<u>Pos/Mm:Ts</u>	<u>Score</u>
Tint_Mitf	510	563	<u>L1-1_NV</u>	5157	5213	NonLTR/L1	c	0.8148	2.0000	207
Tint_Mitf	716	1238	<u>Rex1-1_PM</u>	2644	3173	NonLTR/Rex1	c	0.7524	1.8333	1948
Tint_Mitf	1444	1505	<u>Mariner-11_DF</u>	1382	1447	DNA/Mariner	d	0.8065	1.5000	227
Tint_Mitf	1524	1594	<u>Copia-71_MLP-I</u>	2054	2122	LTR/Copia	c	0.7500	1.3000	256

Supp Fig S4D. TE vs indel locations. SWS1 region for the tilapia *O. niloticus* (Til), *A. burtoni* (Aburt), *P. nyererei* (Pnye), *M. zebra* (Mz), *Tramitichromis intermedius* (Tintermedius), and *Aulonocara baenschi* (Aul baenschi). Exon 1 is shown as yellow arrow while miRNA-729 is shown as red arrow.



A 692 bp deletion in *Aulonocara baenschi* shows little repetitive sequence. The repetitive sequence in this region does not have high scores.

795 bp insertion in the tilapia, *O. niloticus* mostly a NonLTR L2 repeat

Giri CENSOR table

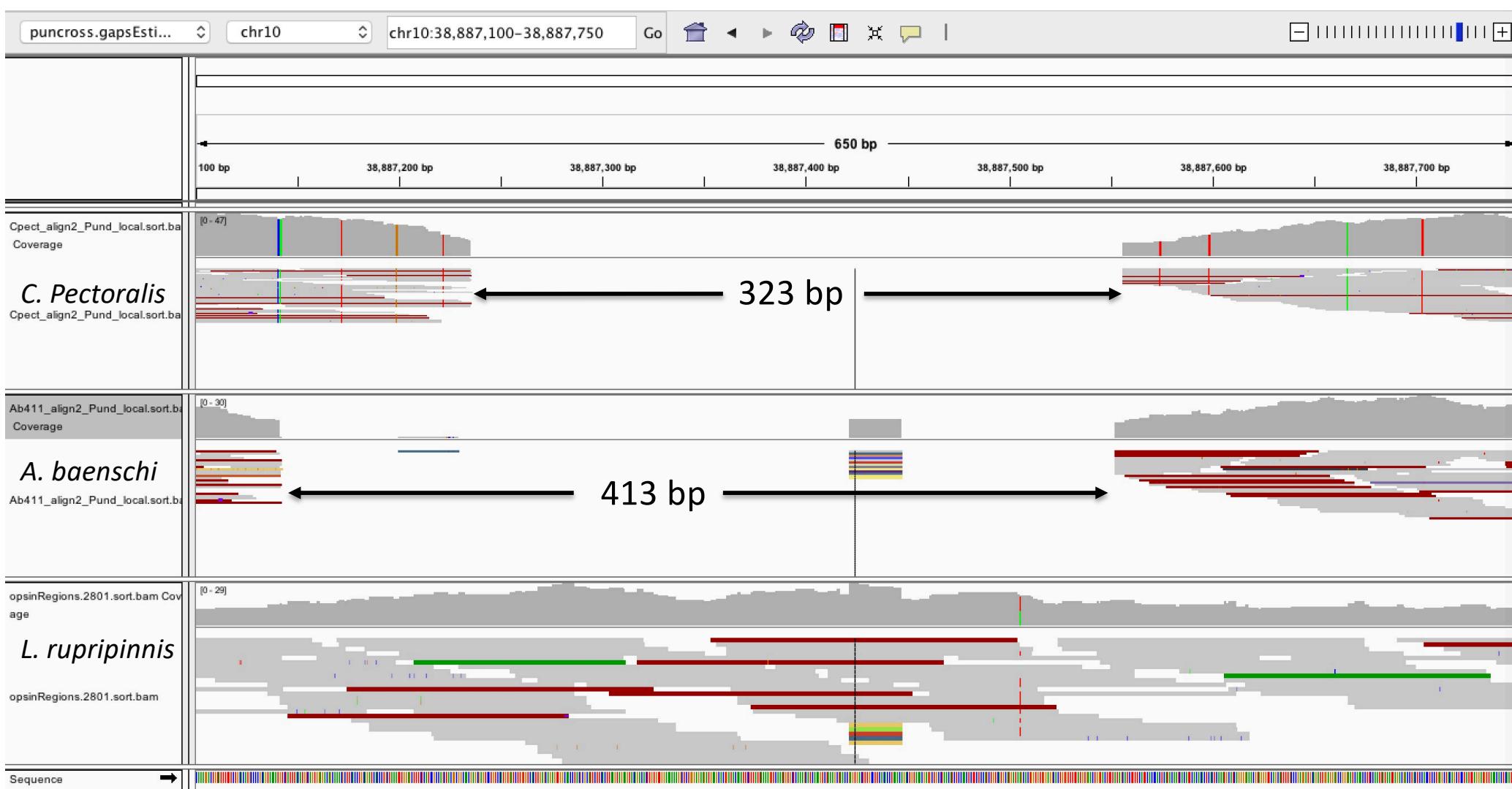
<u>Name</u>	<u>From</u>	<u>To</u>	<u>Name</u>	<u>From</u>	<u>To</u>	<u>Class</u>	<u>Dir</u>	<u>Sim</u>	<u>Pos/Mm:Ts</u>	<u>Score</u>
Mz_SWS1	101	261	<u>Odin-20_OD</u>	1787	1932	NonLTR	d	0.6954	1.4000	221
Mz_SWS1	345	383	<u>Gypsy-33_LSal-I</u>	1024	1062	LTR/Gypsy	d	0.8462	2.0000	207
Mz_SWS1	868	901	<u>Copia-2_SPDB-I</u>	4994	5025	LTR/Copia	d	0.9091	1.0000	232

<u>Name</u>	<u>From</u>	<u>To</u>	<u>Name</u>	<u>From</u>	<u>To</u>	<u>Class</u>	<u>Dir</u>	<u>Sim</u>	<u>Pos/Mm:Ts</u>	<u>Score</u>
TilSWS1ins	20	75	<u>SINE_FR2</u>	102	167	NonLTR/SINE/SINE2	d	0.8448	2.3333	257
TilSWS1ins	129	414	<u>L2-6_SSA</u>	2873	3157	NonLTR/L2	d	0.6972	1.9750	739
TilSWS1ins	434	658	<u>L2-11_EL</u>	3257	3497	NonLTR/L2	d	0.7500	1.7143	473

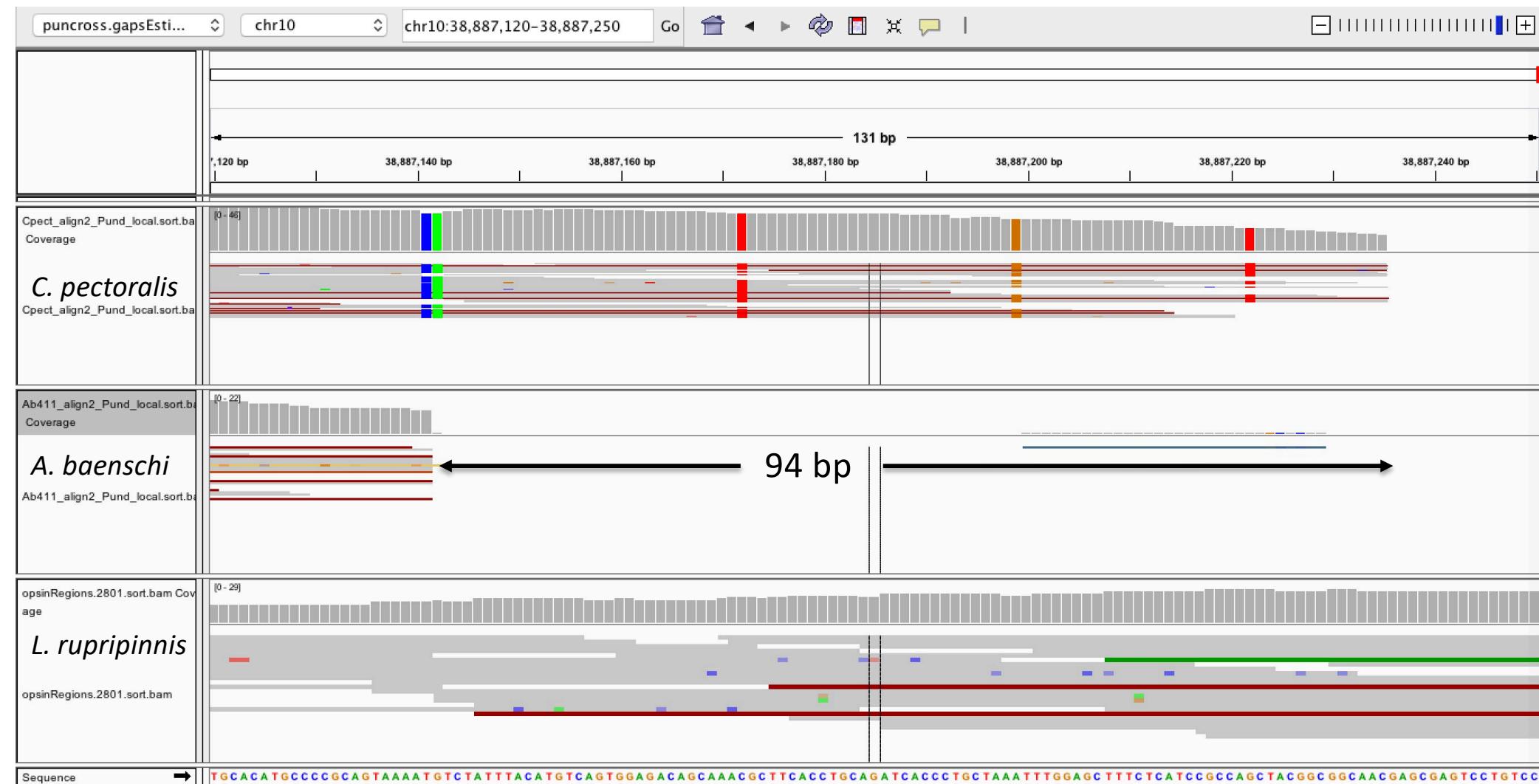
Supp Fig S5. Comparison of deletion in the Rx1 enhancer. Reads for *Ctenochromis pectoralis* (Cpect), *Aulonocara baenschi* (Ab411), and *Lithochromis rubripinnis* (species 2801) aligned against the *Pundamilia nyererei* genome. Region on LG10: 38,887,100-38,887,750).

p.12

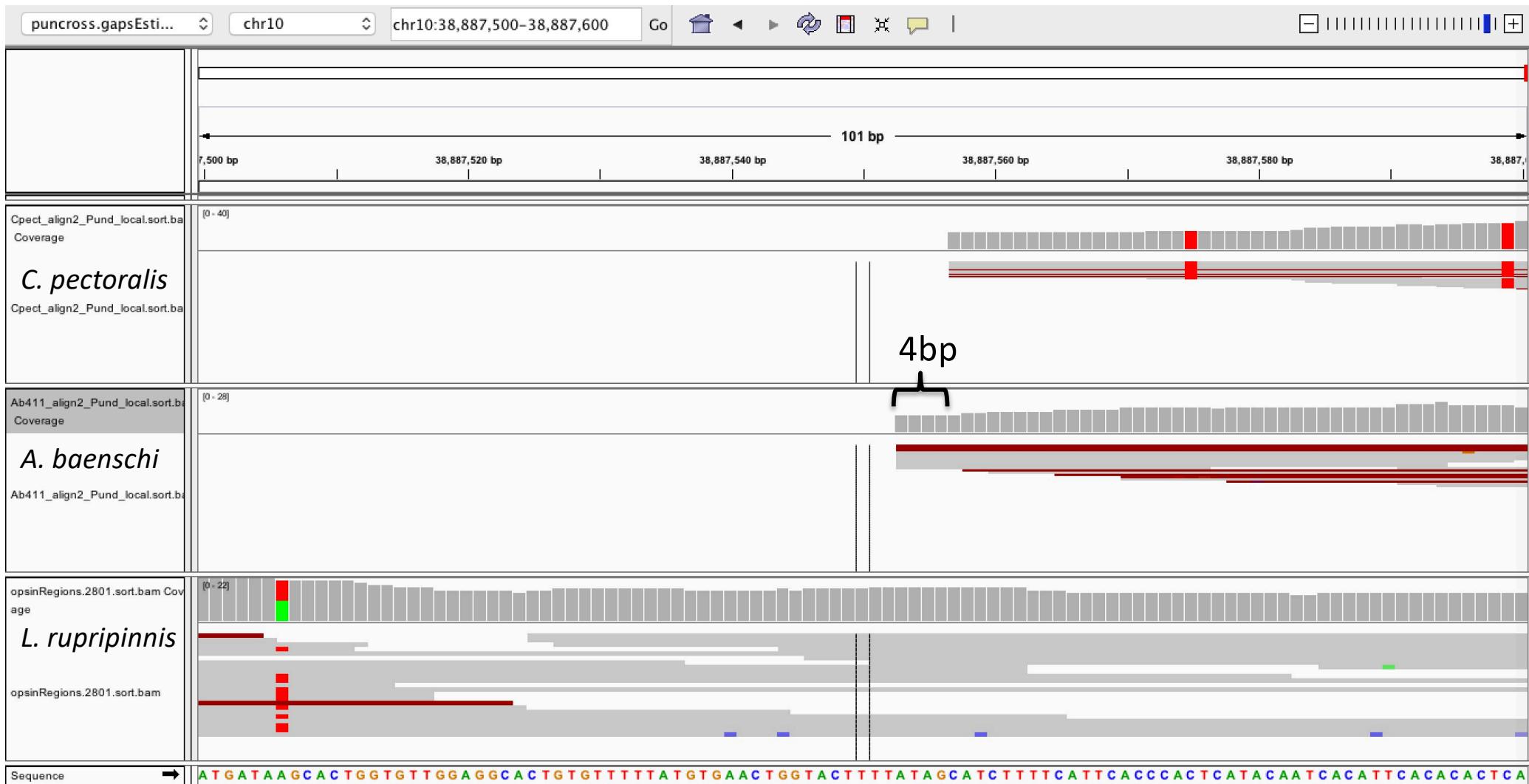
A) Read mapping overview showing the contiguous region for *L. rubripinnis* as opposed to the unique deletions for *C. pectoralis* (323bp) as compared to *A. baenschi* (413 bp).



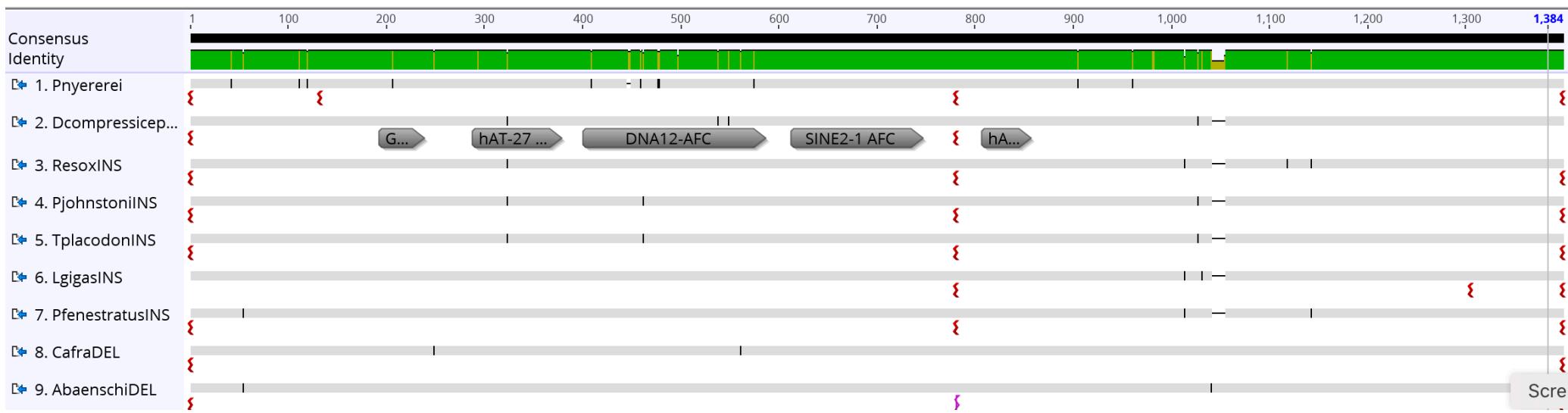
Supp Fig S5B. Close up of read mapping that shows the left hand side of the deletion where the *C. pectoralis* sequence has 94 bp that are missing in *A. baenschi*.



Supp Fig S5C. Close up read mapping that shows the right hand side of the deletion where the *A. baenschi* sequence has 4 bp that are missing in *C. pectoralis*.



Supp Fig S6A. Transposable elements annotated in the Rx1 sequence surrounding the indel used to date the indel origin (note indel sequences are removed in this alignment). A subset of the 18 species studied are shown here including *Rhamphochromis esox*, *Placidochromis johnstoni*, *Trematocranus placodon*, *Labidichromis gigas*, *Protomelas fenestratus*, *Cynotilapia afra* and *Aulonocara baenschi*. The TE sequences are annotated in *D. compressiceps* (Dcomp), shown in gray, and shared across species.



Giri CENSOR table

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Screens
DcompRxlenh	192	241	Gypsy-23_LMi-I	3383	3432	LTR/Gypsy	d	0.7600	1.7143	200
DcompRxlenh	287	380	hAT-27_DR	2652	2741	DNA/hAT	c	0.7444	1.3333	250
DcompRxlenh	401	588	DNA12_AFC	1	214	DNA	d	0.7801	1.4783	669
DcompRxlenh	613	748	SINE2-1_AFC	7	163	NonLTR/SINE/SINE2	d	0.7857	1.2222	533
DcompRxlenh	806	858	hATm-11_HM	1795	1847	DNA/hAT	d	0.8148	1.7500	235

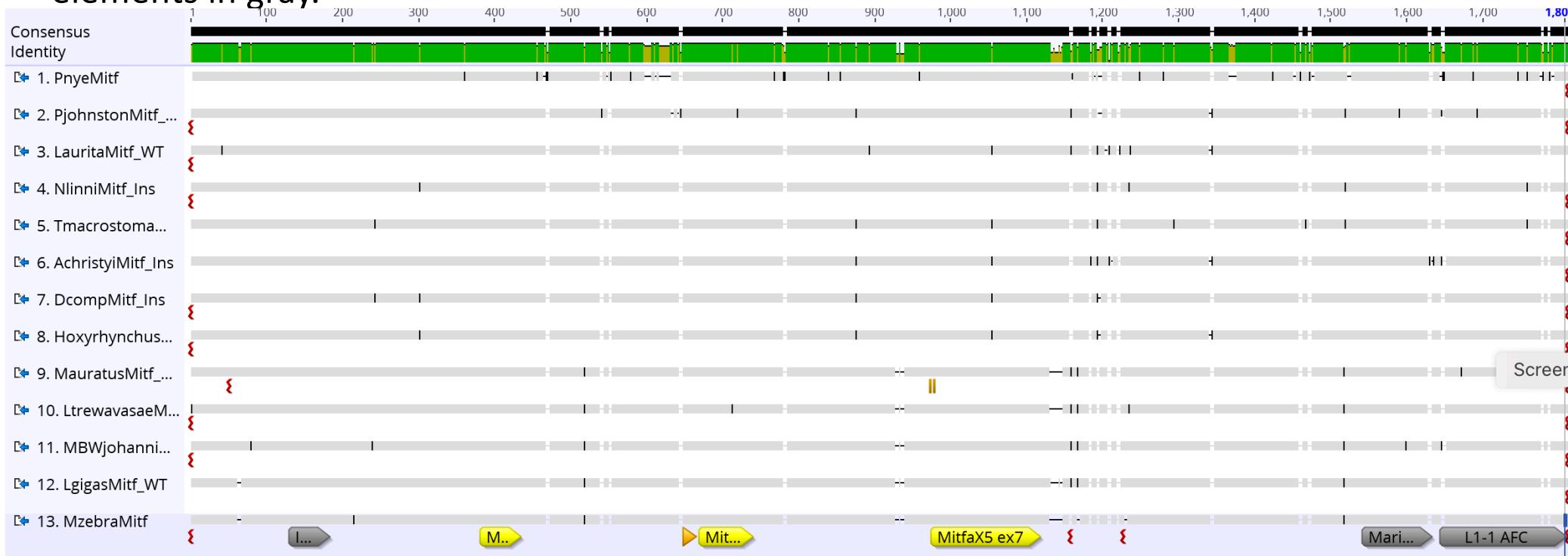
Supp Fig S6B. Transposable elements annotated in the Tbx2a sequence surrounding the indel used to date the deletion origin (note indel sequences are removed in this alignment). Region shown for *P. nyererei* (Pny), *M. zebra* (Mz), *Aulonocara baenschi* (Aulonocara), and *Tramitichromis intermedius* (Tramitichromis). The TEs are noted in gray and shared across species.



Giri CENSOR table

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
AulTbx2pro	10	332	RTE-2_AFC	13	347	NonLTR/RTE	d	0.7785	1.5476	1573
AulTbx2pro	351	500	IC-Anja	977	1151	DNA/Mariner	d	0.8182	2.0000	694
AulTbx2pro	971	1244	SINE_AFC	39	328	NonLTR/SINE/SINE2	c	0.8212	1.7727	1244
AulTbx2pro	1731	1811	Gypsy-33_GA-I	1067	1148	LTR/Gypsy	d	0.7439	1.4286	358

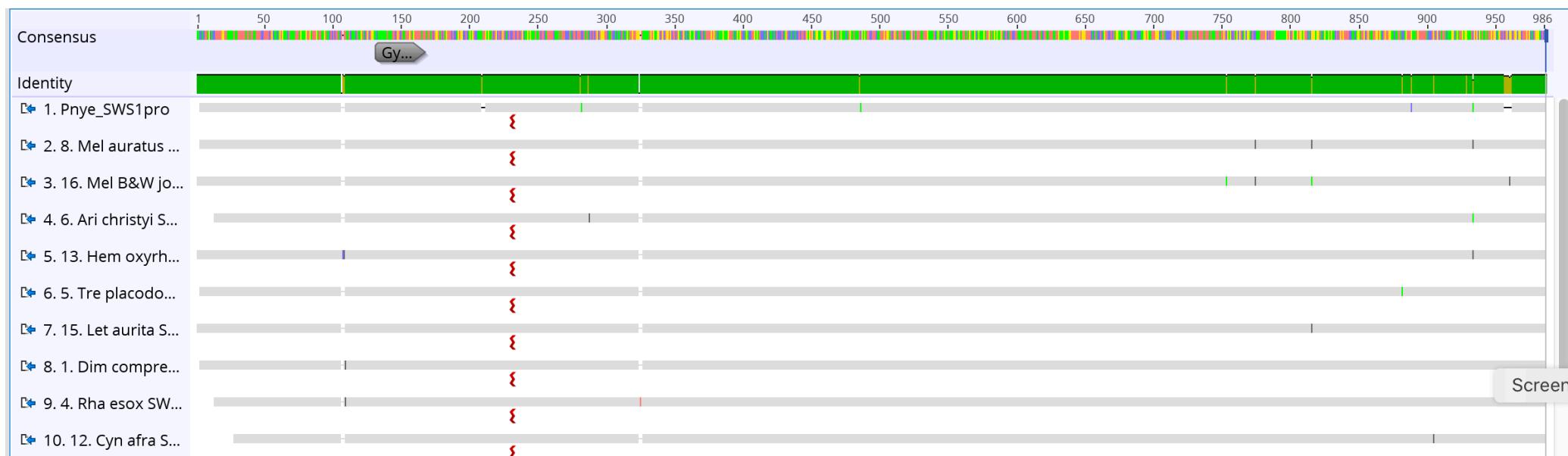
Supp Fig S6C. Transposable elements annotated in the *Mitf* sequence surrounding the indel used to date the insertion origin (note inserted sequences are removed in this alignment). The alignment includes *Pundamilia nyererei* and 12 Lake Malawi species including *Placidochromis johnstoni*, *Lethrinops aurita*, *Nimbochromis linni*, *Tyrannochromis macrostoma*, *Aristichromis chrysty*, *Dimidiochromis compressiceps*, *Hemichromis oxyrhynchus*, *Melanochromis auratus*, *Labeotropheus trewavasae*, *M 'black & white' johanni*, *Labidochromis gigas*, *Metriaclima zebra*. Exons shown as yellow arrows and TE elements in gray.



Giri CENSOR table

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Screensh
MzMitfa	127	183	ISL2EU-1_HSal	2669	2720	DNA/ISL2EU	d	0.7818	1.1250	455
MzMitfa	1502	1596	Mariner-16_DR	1135	1239	DNA/Mariner	c	0.8125	1.5455	498
MzMitfa	1604	1767	L1-1_AFC	6935	7092	NonLTR/L1	c	0.9125	1.7143	1131

Supp Fig S6D. Transposable elements annotated in the SWS1 sequence surrounding the indel used to date the deletion's origin (note deleted region are removed in this alignment). SWS1 region alignment for *P. nyererei* (Pnye), and 10 of the studied Lake Malawi cichlid species including *Melanochromis auratus*, *M 'black & white' johanni*, *Aristichromis chryssty*, *Hemichromis oxyrhynchus*, *Trematocranus placodon*, *Lethrinops aurita*, *Dimidiochromis compressiceps*, *Rhamphochromis esox*, and *Cynotilapia afra*. The one TE is noted in gray and shared across all species.

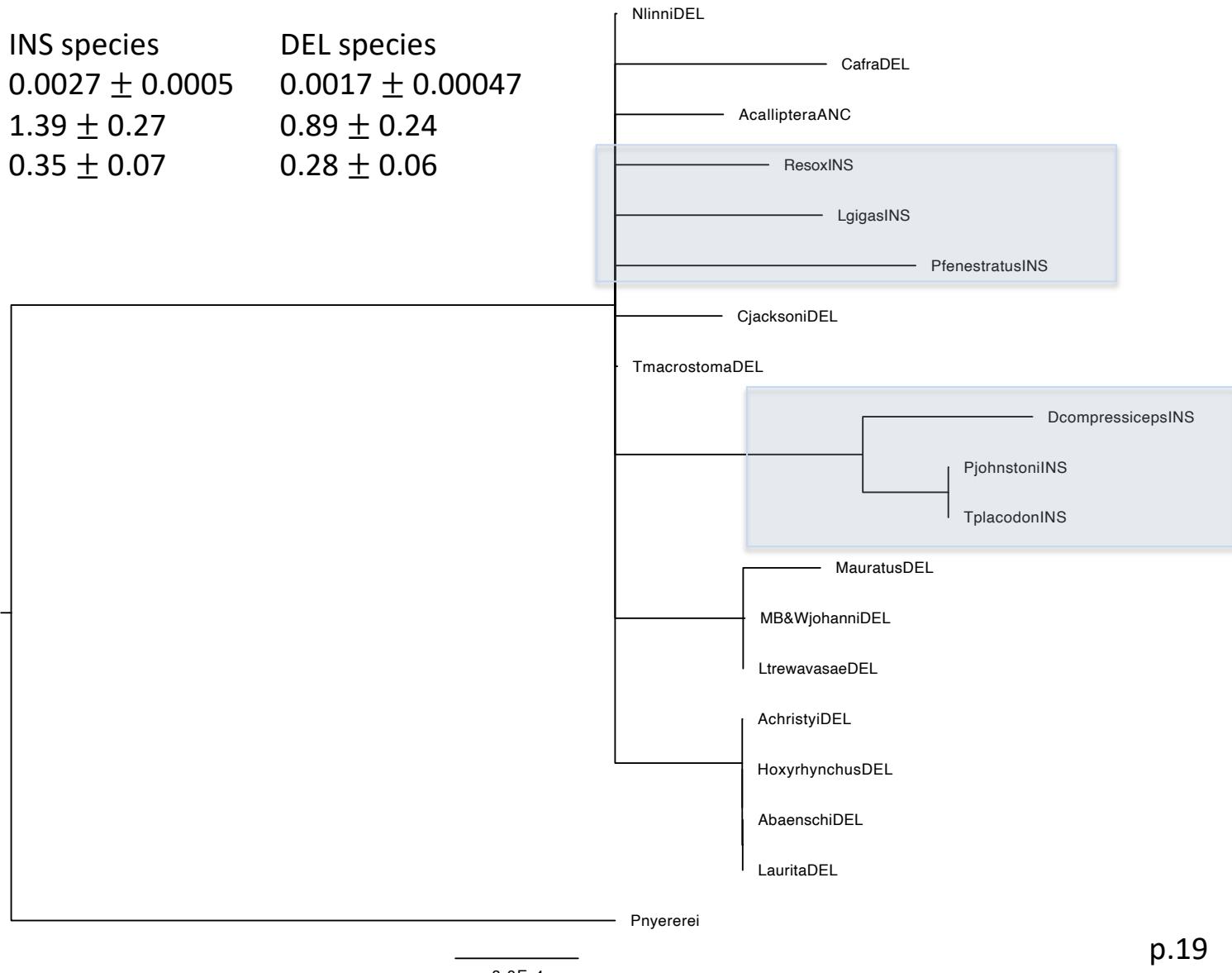


Giri CENSOR table

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
CafraSWS1pro	104	142	Gypsy-33_LSal-I	1024	1062	LTR/Gypsy	d	0.8462	2.0000	207

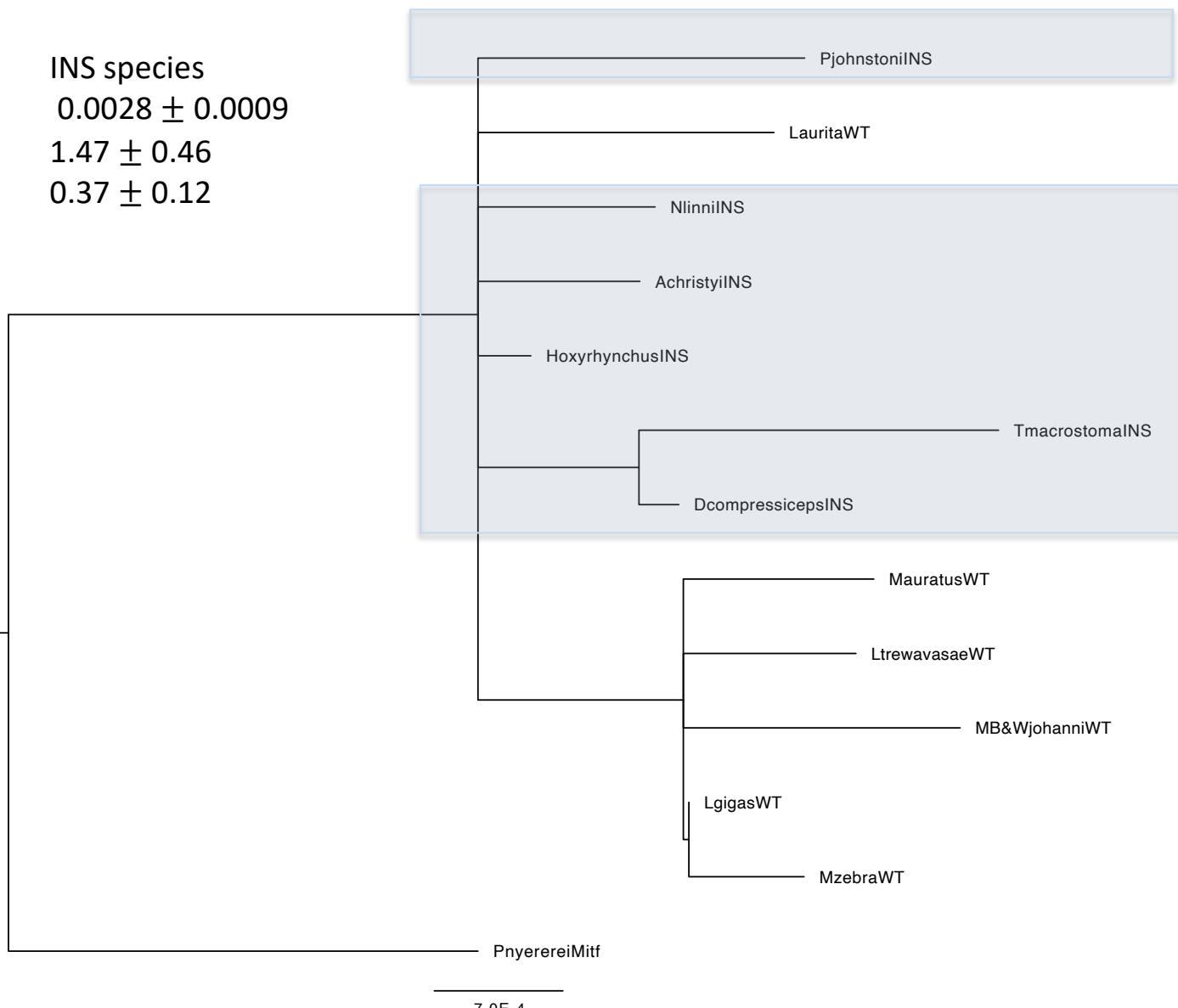
Supp Fig S7A. Tree based on 1400 bp of sequence surrounding the Rx1 indel. This enables us to compare the distance between species either with an insertion (INS; boxed species) or with the deletion (DEL) to the ancestral *A calliptera* sequence. The ratio of this distance to the pelagic Malawi species or the Victorian species, *P. nyererei* is then determined.

Comparison	INS species	DEL species
Ave distance to <i>A calliptera</i>	0.0027 ± 0.0005	0.0017 ± 0.00047
Ratio to Malawi pelagics	1.39 ± 0.27	0.89 ± 0.24
Ratio to <i>P. nyererei</i>	0.35 ± 0.07	0.28 ± 0.06



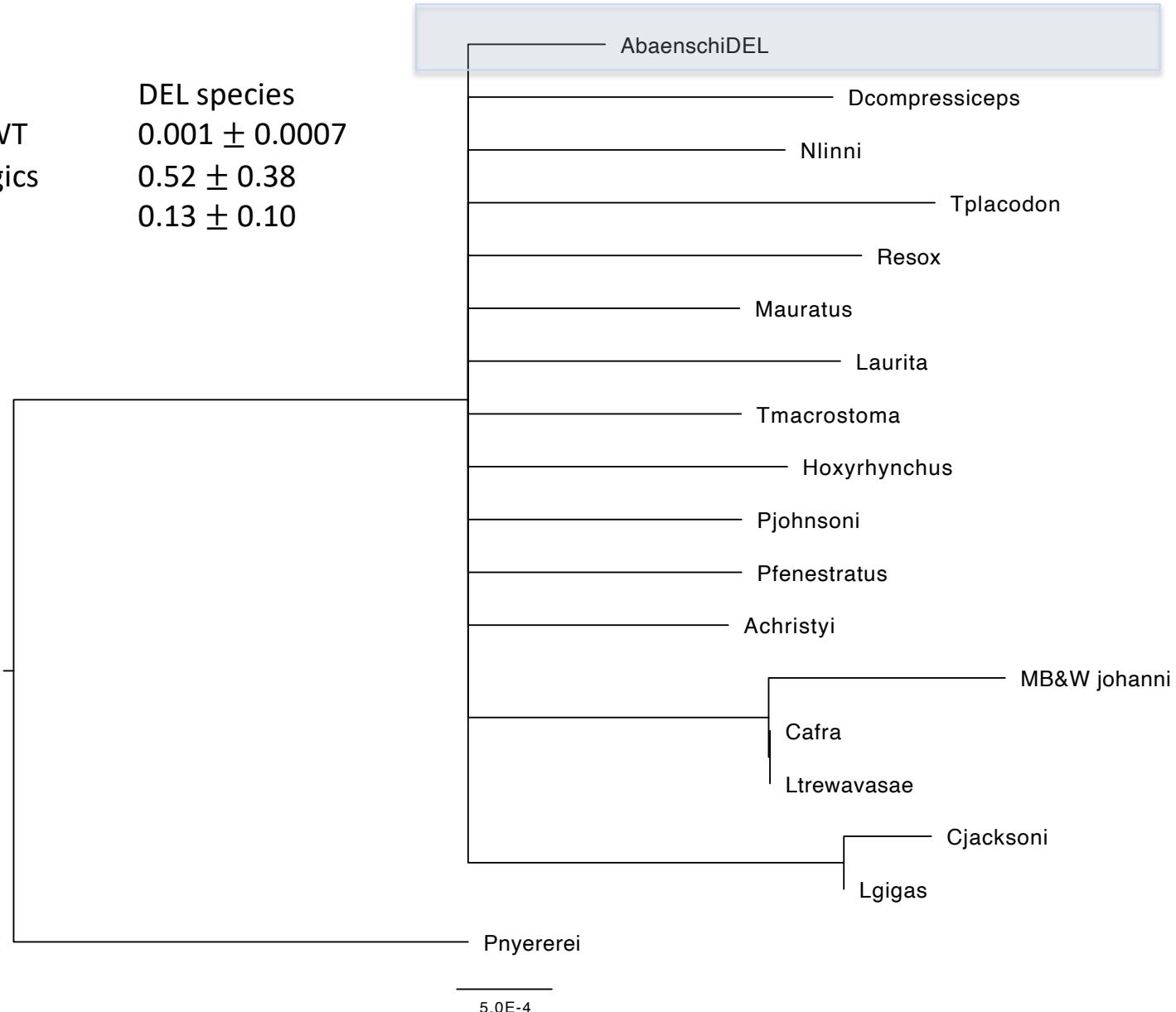
Supp Fig S7B. Tree based on 1800 bp of sequence downstream of the *Mitf* insertion. This enables us to compare the distance between species with an insertion (Ins; boxed species) to those without (wild type, WT). The ratio of this distance to the pelagic Malawi species or the Victorian species, *P. nyererei* is then determined.

Comparison	INS species
Ave INS distance to WT	0.0028 ± 0.0009
Ratio to Malawi pelagics	1.47 ± 0.46
Ratio to <i>P. nyererei</i>	0.37 ± 0.12



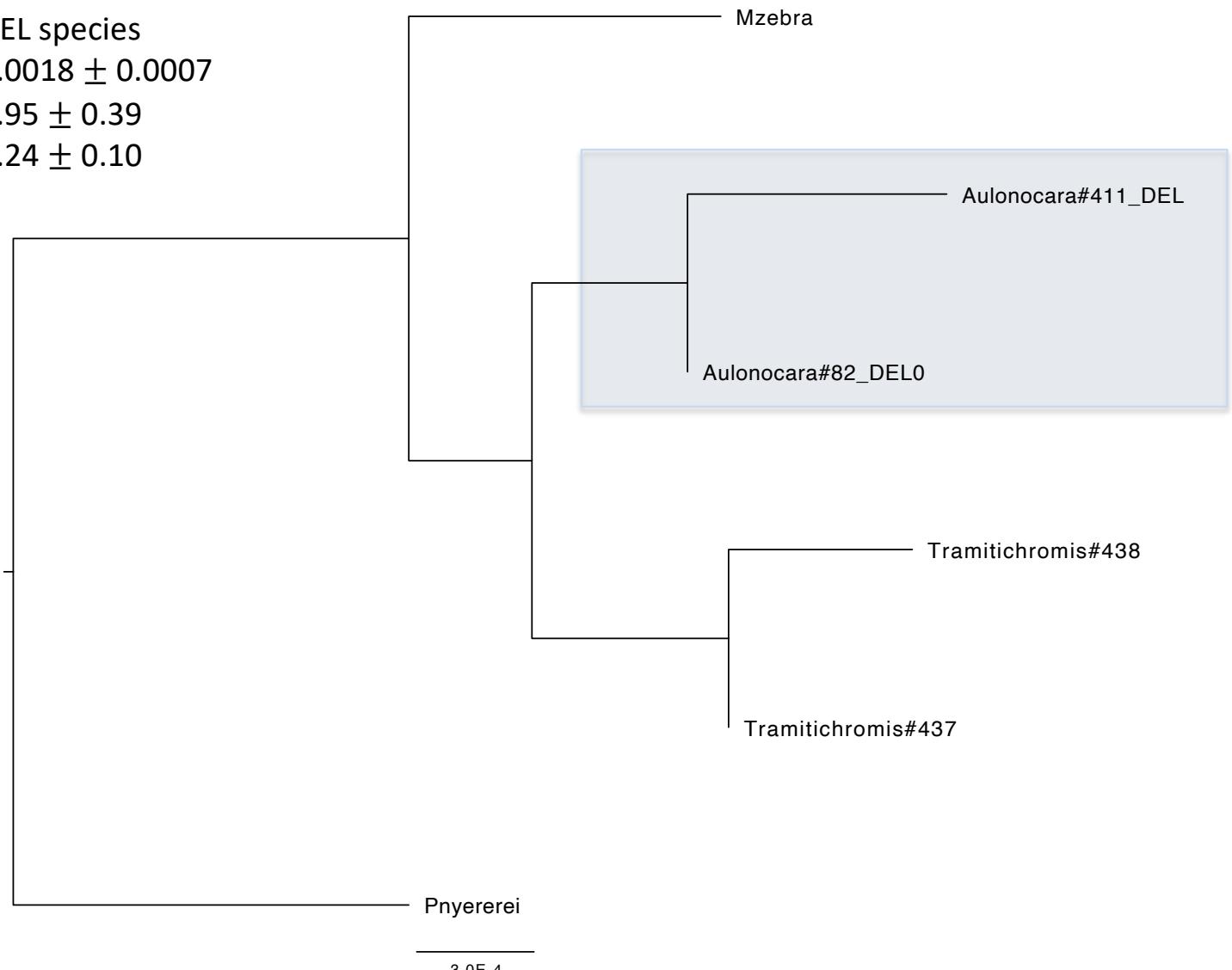
Supp Fig S7C. Tree based on 1000 bp of sequence surrounding the SWS1 deletion in *A. baenschi* (in box) to other Lake Malawi species. The ratio of this distance to the pelagic Malawi species or the Victorian species, *P. nyererei* is then determined.

Comparison	DEL species
Ave DEL distance to WT	0.001 ± 0.0007
Ratio to Malawi pelagics	0.52 ± 0.38
Ratio to <i>P. nyererei</i>	0.13 ± 0.10



Supp Fig S7D. Tree based on 1800 bp of sequence surrounding the *Tbx2a* deletion in *A. baenschi* ((in box) as compared to *T intermedium* and *M zebra*. The ratio of this distance to the pelagic Malawi species or the Victorian species, *P. nyererei* is then determined.

Comparison	DEL species
Ave DEL distance to WT	0.0018 ± 0.0007
Ratio to Malawi pelagics	0.95 ± 0.39
Ratio to <i>P nyererei</i>	0.24 ± 0.10



Supp Table S2. Primers for PCR screening for indels and for sequencing nearby regions

Gene	Forward		Reverse	
Primers for PCR screening indel presence/absence				
Rx1 ¹	Rx1P3F Rx1KpnF1	TGGGTTTAATGGCTCCACAG GCGCGGGTACCCCTGCACTCTTGGCC TTGTT	Rx1P3R Rx1XhoR1	CGTTTTCTTCGCTGGTTACTT GCGCGCTCGAGTGAGAAGAACTGT CCCGTGA
Tbx2a ²	Tbx2_FD-F1	GGTCAATGATGCTTCCTCTCAG	Tb13kb_R	GAGGTTTGATCAGAGTAAGAACCT
Mitf ³	MitfP6F MitfP7F	TGTGTTTGGGTGTTCAGAAC GTTGGACAGGTGCTGAAATGT	MitfP6R MitfP7R	ACGTGCTTGCTTGCAACAT ACATCGTCGTTGAGCTTGAC
SWS1 ⁴	mnpromoterF	CCTCAATGCTCCAAAAAGGA	promoter_3R	GCCAACCAGCTCACTGTTGT
Primers for sequencing region around indel to date origin				
Rx1	Rx1p3F5	ATGCATCAGGGGAAACGTAG	Rx1p4LF	TGCGTCAGCACTTAATTATGG
Mitf	Mitf1_P10F Mitf1_P11F mitf1_71F Mitf11F2	CTGTGGCCAATATmGAAAGCTA GACAGAGGAAGCTGGAGCAC TGAACGAAGACGGAGATTCA CTACTTGATCCCAGATGTACAGT	Mitf1_P10R Mitf1_P11R Mitf1_71R Mitf11R2 Mitf11R3	GAATCTCCGTCTCGTTCAA CGCTTTCTTTCCCCAGTTA CAACATGAGATGCCGATTG TGACCCTTTATTTGGTGGCA AACTTATCGAAACAATGCCAGAG

¹Schulte et al 2014

²Sandkam et al 2019

³Nandamuri 2018

⁴Nandamuri et al 2018

Supp Table S3. Location of regulatory sequences in *Metriaclima zebra* UMD2a genome (Conte et al 2019)

Nearby gene	Cross	Region	Indel location in <i>M zebra</i> genome*	Start site of gene and NCBI Accession #	<i>M zebra</i> status	<i>A baenschi</i> status	Other cross parent status
Rx1	TA	Regulatory region 2.5kb upstream	LG23: 1374765<> 1374766	LG23: 1377294 XM_004555105.1	Deletion	Deletion	Insertion heterozygous, <i>T. intermedius</i>
Tbx2a	TA	Regulatory region 13kb upstream	LG10: 8403573.. 8402302	LG10: 8389309 XM_004557464.1	TE insertion	Deletion	Ancestral
Mitfa	TA	Intron 1	LG5: 7206112<>720 6113	LG5: 7225844 XM_024802633.1	Ancestral, no insertion	Ancestral	Insertion, <i>T. intermedius</i>

SWS1	MA	Regulatory region upstream of SWS1	000578F_pilon _quiver: 239435..24012 4	000578F_pilon_qu iver: 240877 NM_001310074.1	Ancestral, no deletion	Deletion	Ancestral, <i>M. 'mberjii</i>
------	----	---	---	--	-------------------------------	----------	-------------------------------

* Regions deleted in other species listed from start to end. Location of insertion is marked with <>

Supp Table S4. Sequences for four genomic indel regions. For the SWS1 and Tbx2a loci, this lists the sequence that is present in *M. zebra* genome (UMD2a; Conte et al 2019) but missing in *A. baenschi*. For the Mitf locus, it denotes the location of the insertion in different species including *T. intermedius*. For the Rx1 locus, it shows the location of the deletion which is shared by *M. zebra* and other short or medium palette species.

>MzUMD2a_SWS1pro_deleted_in_Abaenschi

```
ACAGAAACCTAACATTAAACAGTGAAGCAGCAGACATAATCTCCGGTCTTCATACCTGAACCACCAGGAGCTTAATCAATAATCCATAGTAGCAA  
GTAATCAATAGTCATAACATCAGTGATTAATCAGTCTTTAAGGCTGCAAGTCACCTCAAACACTGAAGTCGTATCATACCCATGCTCCATGCAGCTCA  
CAGCCTGGTTATGATACAGCCCCGCTCTAAAGGTCAAGCAATCATTACAGCCATCGATCAATCGATCAGTCAATAATGTGTCTCCTCTTCAGTCTGG  
TCCAGGATGAAAATGGGCAACAGAAGTGGAGACAGTGCAGTTCTTGTCCGCACGTTCTCCGGACTCTGAGCAGGAGAAGTGTGAAGGTT  
GGAGGACAGCTCCTCTAAAGGCTCAGTGTCTCCTCCTCACAGTCTCCTCACGCTGAACCCCTTAACCGTTAGTGGTCGTGAG  
TCTTCTAATCTGACTGTGTGACGATGTTAAGGATTGGAGGATTGAGGAGGATCACCTGGTCAGGTAAATCTGAAATATCCGGATTACATCGGA  
AGTGAGCACACGGAAAAACAAAAGACTCTTATTGGATTAGATCCGTAGCCACCTGCTGCTCTTCATCATCAGGCGTCTCATGCCCTGCA  
GTGGGCCTGAC
```

>MzUMD2a_Tbx2aEnhancer_deleted_in_Abaenschi

```
AAAAAACTATTATGCATTAATTACTTCATTCACAGTGTAGACTGTTGTAAC TGCCGCAACAGACAATTGAAATTACAGCCAGCTGGGA  
GTAAGTGATGTGACCAAGATTGGAGCAGGATTAGAGGTATGGTAGGTAGGGCTGGGTATCGTCACTGATTCTAGAATCGATTCAATTCCGATT  
CACAAGGTCCTCGAATCGATTGATCCACGATTGATTAATTCGATTGATTAAATCTGGGAAATTGACAGTCAGAGATAATTAGAT  
CAGGACATTACATATTTGTATCTATAAAAAGGAAGCTGACACTCTCAAGACTTATCCAAGGTGTGAGCATTACAGCAGATGCCCTGTCAA  
AGTAGCTGAAGATAAAACACAGAAAAACATGTAGGTGATTTCTGGCCTGGGATTATAAAAATATTCTGAGTACATCAAAAAGGAAGAAAA  
CCATTAATTAAATCAACATTACCTCTGACGTTACAGCGGTTTATTAGAGACATGGCTAAGCATTGCAATTGAAATAATTAAAAAGTTCAATTG  
TTCAGTATTGAACAGCAGAAATGAGGTTTCTTCGGAAGAATGTAAGGGAAAAACAGCGGCCGACAGCGCTGTAAGCAACAGTAGACTT  
GTGTAACAAGCAAGCGAATAATGAAGAAAGCGAAACTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGCTGCGATTGCAAT  
GGAAGCAAAACAGTAAAGAGTGAATTGACGATGTTATGTGAAGCGTTGGATCTCTTGTGCTGATTCGGTCGATATTGTTGGAGAG  
AGATCAAACACTAACAGCTTACAATCAGCGCATAAAGGGCGTGAACACAAAGCGCGGACCCGGATCAGCGAGCTGCGCTTCAGCCCCACTGT  
GAGAAAGGCGACATCTAATGATTGATCCGGGGTTGCGCTGTGTTAAGTCTATGTGAGAATCCCTGACCTGCTCTCATTTACTGTTGG  
TGGTTCTGAAATTGAGATGTCACCAGAGATTCTGGCATTGAGAATGAACTAACCCACCCCTAATGGTGAGGTTAAGATTCTTAACGTCTCAAGTCC  
TTGATTGTTACCTGAAAATCTTTCACCTCACACACATTACGAACAGTTGAAGGCAGATTGACCTGCCAGG
```

>MzUMD2a_Mitf_intron1_Tramitichromis_insertion

GAGTAGATGCGATTTCACACTAAAAGAGTCAAAATATCTTGCTGCGCTGAAAACAGGACTAAGGTTGAGGATTAGGGTTGAGTTCTAAATAC
TTCATTATGTGTACCCAGTTACAAGTCTTGGTTAAATGAAGACAGGGAAAGCTGTTTACACACAGTCGTATTAAATTGCTGCACGAGTTGGACA
GGTGCCTGAAATGTTATATAGTGCTCTTGTAGAAAAAAAGAGCTTGATGAAAAAAATGTAAGAAATGTTGTTGGACATTCAACATCTTATCAC
CGTA<insertion>ATGTTGCAAAGCAAGCACGTTAGATTATTGGTTTAAAGGCTGTTTCAGGGAACTTTAACATGTAACAGAGTAGTATAGT
GTGATTAATTGGATTGTAAAGTTGACTTGAAGCTGGAAACATTGAACATTAAAGTCTGCATTAAATTGTGCTTTACCACACCGGGACA
GATGAATTATTATAATGAAAACATTCACTTTA

>MzUMDa_Rx1pro_deleted_in_short_medium_palettes

CTGCATCCTCGGGCAAAATGCTGAACCCCTACGTTCCCTGATGCATTGAGTGAGTGATGATTGATGAGTGGGTGAATGAAAAGATG
CTAT<deletion_here>TTTACTGCGGGGCATGTGCAGAAAAAAATACAGTCCTTGTAGCACAGAAGTGTAAACATAAAAACAAAAAGACTGTGACGA
GGAATTGAAATGATCACGGGACA

Supp Table S5. Frequency and location of recent insertions for three TE families in the genomes of *Oreochromis niloticus* (On) and *Metriaclima zebra* (Mz).

TE_family in assembly	# insertions (div < 2.0)	#total length of insertions, div < 2.0)	% promoter	% exon	% intron	% intergenic	# promoter	# exon	# intron	# intergenic
DNA/TcMar in O_niloticus_UMDNMBU	8,127	7,470,032	35.73	0.32	21.94	42.01	2904	26	1783	3414
DNA/TcMar in M_zebra_UMD2	13,467	13,922,869	24.07	0.09	22.51	53.32	3242	12	3031	7181
Mz - On	5,340	6,452,837	-11.66	-0.23	0.57	11.31	338	-14	1248	3766
Ratio Mz / On	1.66	1.86	0.67	0.28	1.03	1.27	1.12	0.47	1.70	2.10
<hr/>										
DNA/hAT in O_niloticus_UMDNMBU	8,785	3,839,634	38.80	0.25	21.43	39.51	3409	22	1883	3471
DNA/hAT in M_zebra_UMD2	9,316	6,931,987	23.78	0.12	24.79	51.32	2215	11	2309	4781
Mz - On	531	3,092,353	-15.02	-0.13	3.36	11.81	-1193	-11	427	1310
	1.06	1.81	0.61	0.48	1.16	1.30	0.65	0.51	1.23	1.38
<hr/>										
LINE/Rex in O_niloticus_UMDNMBU	3,331	1,946,563	25.01	0.51	11.86	62.62	833	17	395	2086
LINE/Rex in M_zebra_UMD2	8,219	6,291,320	27.44	0.21	21.55	50.81	2255	17	1771	4176
Mz - On	4,888	4,344,757	2.43	-0.30	9.69	-11.81	1422	0	1376	2090
	2.47	3.23	1.10	0.41	1.82	0.81	2.71	1.02	4.48	2.00

Supp Table S7. Significant correlations between opsin genes from previous work. Correlational data includes ontogenetic series for *Oreochromis niloticus*¹, F₂ genetic crosses for *Tramitichromis intermedius* x *Aulonocara baenschi*² and *Metriaclima mbenji* x *Aulonocara baenschi*³, and data for 54 species from Lake Malawi⁴. For each gene combination, the correlation coefficient (R) and the p value are given. Only correlations with a p value <0.0033 (0.05 / 15 comparisons) were included to correct for 15 multiple tests within each dataset.

	SWS2B	SWS2A	RH2B	RH2A	LWS	Ref
SWS1	-0.99 <1x10 ⁻²⁵ -0.41 2.1x10 ⁻³	-0.77 1.1x10 ⁻⁵ 0.45 6.4x10 ⁻⁴	0.90 2.9x10 ⁻⁹ -0.44 9.2x10 ⁻⁴		-0.89 7x10 ⁻⁹ -0.30 1.7x10 ⁻⁴	1 3 4
SWS2B		-0.61 3.2x10 ⁻¹⁷			-0.30 1.7x10 ⁻⁴	2
SWS2A			-0.89 1x10 ⁻⁸ -0.28 3.0x10 ⁻⁴ -0.74 1.2x10 ⁻¹⁰	-0.59 2.5x10 ⁻³	0.89 6.5x10 ⁻⁹ 0.72 1.1x10 ⁻⁹	1 3 4
RH2B				-0.39 4.3x10 ⁻⁷ -0.46 5.7x10 ⁻¹⁰	-0.97 2.2x10 ⁻¹⁴ -0.31 6.1x10 ⁻⁵ -0.81 1.1x10 ⁻¹³	1 2 3 4
RH2A					-0.67 3.9x10 ⁻⁴ -0.68 8.3x10 ⁻²³ -0.85 <1x10 ⁻²⁵ -0.48 2.2x10 ⁻⁴	1 2 3 4

1 (Carleton et al. 2008)

2 (O'Quin et al. 2012)

3 (Nandamuri et al. 2017)

4 (Hofmann et al. 2009)

Carleton, K. L., T. C. Spady, J. T. Streelman, M. R. Kidd, W. N. McFarland and E. R. Loew (2008). "Visual sensitivities tuned by heterochronic shifts in opsin gene expression." *BMC Biol* 6(1): 22.

Hofmann, C. M., K. E. O'Quin, N. J. Marshall, T. C. Cronin, O. Seehausen and K. L. Carleton (2009). "The eyes have it: Regulatory and structural changes both underlie cichlid visual pigment diversity." *PLoS Biol* 7(12): e1000266.

- Nandamuri, S. P., B. E. Dalton and K. L. Carleton (2017). "Determination of the Genetic Architecture Underlying Short Wavelength Sensitivity in Lake Malawi Cichlids." J Hered **108**(4): 379-390.
- O'Quin, K. E., J. E. Schulte, Z. Patel, N. Kahn, Z. Naseer, H. Wang, M. A. Conte and K. L. Carleton (2012). "Evolution of cichlid vision via trans-regulatory divergence." BMC Evol Biol **12**: 251.