Table S1. Random-sites analyses (PAML) of the world-wide croaker rhodopsin dataset using the Bayesian species tree

Model	np	lnL	Parameter estimates: d_N/d_S (proportion of sites)	Null	LRT	p value
M0	225	-7129.66	0.27(1.00)			
M1a	226	-6627.71	0.02(0.85) 1.00^F(0.15)	M0	1003.91	< 0.0001
M2a	228	-6525.91	0.02(0.84) 1.00^F(0.13) 4.53(0.04)	M1a	203.60	< 0.0001
M7	226	-6618.10	β (p = 0.01, q = 0.04)			
M8a	227	-6608.15	β (p = 0.13, q = 2.94), 1.00^F(0.12)			
M8	228	-6517.02	β (p = 0.06, q = 0.38), 4.16(0.04)	M7	202.16	< 0.0001
				M8a	182.26	< 0.0001

Note: lnL, ln likelihood; ^{F}, d_N/d_S estimate fixed at reported value; *LRT, likelihood ratio test result*

Table S2. Random-sites analyses (PAML) of the world-wide croaker rhodopsin dataset using the maximum likelihood species tree

Model	np	lnL	Parameter estimates: d_N/d_S (Proportion of sites)	Null	LRT	p value
M0	227	-7075.70	0.26(1.00)	n/a		
M1a	228	-6586.36	0.02(0.85) 1.00^F(0.15)	M0	978.7	< 0.0001
M2a	230	-6490.44	0.03(0.84) 1.00^F(0.13) 4.54(0.03)	M1a	191.8	< 0.0001
M7	228	-6579.96	β (p = 0.06, q = 0.30)			
M8a	229	-6569.04	β (p = 0.13, q = 3.07) 1.00^F(0.12)	M7	21.8	< 0.0001
M8	230	-6483.41	β (p = 0.06, q = 0.39) 4.13(0.04)	M7	193.1	< 0.0001
				M8a	171.2	< 0.0001

Note: lnL, ln likelihood; ^{F}, d_N/d_S estimate fixed at reported value; *LRT, likelihood ratio test result*

Table S3. Random-sites analyses (PAML) of the world-wide croaker rhodopsin dataset using the maximum likelihood rhodopsin gene tree

Model	np	lnL	Parameter estimates: d_N/d_S (Proportion of sites)	Null	LRT	p value
M0	227	-6673.45	0.24(1.00)			
M1a	228	-6277.64	0.02(0.84) 1.00^F(0.16)	M0	791.6	< 0.0001
M2a	230	-6206.98	0.02(0.84) 1.00^F(0.16) 8.94(0.01)	M1a	141.3	< 0.0001
M7	228	-6271.43	β (p = 0.06, q = 0.32)			
M8a	229	-6263.12	β (p = 0.12, q = 2.61) 1.00^F(0.11)	M7	16.6	< 0.0001
M8	230	-6202.27	β (p = 0.01, q = 0.34) 8.29(0.01)	M7	138.3	< 0.0001
				M8a	121.7	< 0.0001

Note: lnL, ln likelihood; ^{F}, d_N/d_S estimate fixed at reported value; *LRT, likelihood ratio test result*

Table S4. Positively selected sites identified in statistically significant tests of positive selection

Full rhodopsin dataset, Bayesian phylogeny

M2a (BEB) – $\boldsymbol{d_N/d_S}$ = 4.53

82, 162, 165, 169, 173, 209, 214, 218, 262

M8 (BEB) – $\boldsymbol{d_N/d_S}$ = 4.16

33, 82, 112, 162, 165, 169, 173, 209, 214, 218, 262, 299

FUBAR

33, 37, 39, 40, 112, 115, 151, 165, 168, 173, 209, 214, 218, 256, 263, 266, 271, 290, 299

South American marine to freshwater transitional branch as foreground – Branch-Site (BEB) – $\boldsymbol{d_N/d_S}$ = 0.02/18.42, 1.00^F/18.42

119, 122, 159, 261, 282

Dataset with freshwater lineages removed, Bayesian phylogeny

M2a (BEB) – $\boldsymbol{d_N/d_S}$ = 4.50

82, 112, 162, 165, 169, 209, 214, 218, 262

M8 (BEB) – $\boldsymbol{d_N/d_S}$ = 4.31

33, 82, 112, 162, 165, 169, 173, 209, 214, 218, 262, 299

South American dataset, Bayesian phylogeny

M2a

162, 165, 214, 218

Marine lineages as foreground – Branch-Site (BEB) – $\boldsymbol{d_N/d_S}$ = 0.02/9.53, 1.00^F/9.53

82, 162, 165, 214, 218

Freshwater clade as foreground – Branch-Site (BEB) – $\boldsymbol{d_N/d_S}$ = 0.02/10.75, 1.00^F/10.75

50, 52, 165, 173

Marine to freshwater transitional branch as foreground – Branch-Site (BEB) – $\boldsymbol{d_N/d_S}$ = 0.02/23.52, 1.00^F/23.52

63, 119, 122, 124, 158, 159, 169, 214, 261, 282, 304

Note: d_N/d_S estimates for the positively selected site class reported for each model (background/foreground); all sites reported have posterior probabilities greater than 0.7

Table S5. Branch-sites and clade model analyses (PAML) comparing d_N/d_S estimates for the transitional branches of the world-wide croaker rhodopsin dataset using the Bayesian species tree

Model	np	lnL	Parameter Estimates: d_N/d_S (proportion of sites): background d_N/d_S / foreground d_N/d_S	Null	LRT	p value
M2aREL	228	-6525.91	0.02(0.84) 1.00^F(0.13) 4.53(0.04)			
M3	229	-6525.91	0.02(0.83) 0.99(0.13) 4.51(0.04)			
All Freshwater transitions as Foreground						
Br-site alt	228	-6620.94	0.02(0.82) 1.00^F(0.14) 0.02/5.62(0.04) 1.00^F/5.62(0.01)	Br-site null	4.98	0.0256
Br-site null	227	-6623.43	0.02(0.79) 1.00^F(0.13) 0.02/1.00^F(0.07) 1.00^F/1.00^F(0.01)			
CmC	229	-6525.88	0.02(0.84) 1.00^F(0.13) 4.55/4.12(0.04)	M2a_rel	0.05	0.8168
CmD	230	-6523.47	0.04(0.86) 8.81(0.01) 1.44/3.95(0.13)	M3	4.88	0.0272
North American Transition as Foreground						
Br-site alt	228	-6627.46	0.02(0.83) 1.00^F(0.15) 0.02/1.42(0.02) 1.00^F/1.42(0.00)	Br-site null	0.01	0.9031
Branch Site (null)	227	-6627.47	0.02(0.82) 1.00^F(0.15) 0.02/1.00^F(0.03) 1.00^F/1.00^F(0.00)			
CmC	229	-6525.39	0.02(0.84) 1.00^F(0.13) 4.55/2.16(0.04)	M2a_rel	1.03	0.3100
CmD	230	-6525.14	0.02(0.83) 4.42(0.04) 0.96/1.94(0.13)	M3	1.54	0.2143
Mekong River Transition as Foreground						
Br-site alt	228	-6627.71	0.02(0.85) 1.00^F(0.15) 0.02/1.00(0.00) 1.00^F/1.00(0.00)	Br-site null	0.00	1.0000
Br-site null	227	-6627.71	0.02(0.85) 1.00^F(0.15) 0.02/1.00^F(0.00) 1.00^F/1.00^F(0.00)			
CmC	229	-6524.26	0.02(0.84) 1.00^F(0.13) 4.56/0.00(0.04)	M2a_rel	3.30	0.0693
CmD	230	-6525.69	1.00(0.13) 4.52(0.04) 0.03/0.00(0.84)	M3	0.44	0.5071
South American Transition as Foreground						
Br-site alt	228	-6617.67	0.02(0.82) 1.00^F(0.14) 0.02/18.42(0.04) 1.00^F/18.42(0.01)	Br-site null	9.64	0.0019
Br-site null	227	-6622.49	0.02(0.75) 1.00^F(0.13) 0.02/1.00^F(0.10) 1.00^F/1.00^F(0.02)			
CmC	229	-6524.93	0.02(0.84) 1.00^F(0.13) 4.36/9.67(0.04)	M2aREL	1.95	0.1622
CmD	230	-6522.59	0.03(0.85) 6.86(0.02) 1.30/6.42(0.13)	M3	6.63	0.0100

Note: lnL, ln likelihood; ^{F}, d_N/d_S estimate fixed at reported value; *LRT, likelihood ratio test result;* Br-site, Branch-site; CmC, Clade model C; CmD, Clade model D

Table S6. Branch-sites and clade model analyses (PAML) comparing d_N/d_S estimates for the transitional branches of the world-wide croaker rhodopsin dataset using the maximum likelihood species tree

Partitions	np	lnL	Parameter Estimates: d_N/d_S (proportion of sites) (background d_N/d_S / foreground d_N/d_S)	Null	LRT	p value
M2aREL	230	-6490.44	0.03(0.84) 1.00^F(0.13) 4.54(0.03)			
M3	231	-6490.39	0.03(0.84) 1.04(0.13) 4.65(0.03)			
All transitional branches						
Br-site alt	230	-6579.40	0.02(0.82) 1.00^F(0.14) 0.02/5.49(0.04) 1.00^F/5.49(0.01)	Br-site null	4.45	0.0349
Br-site null	229	-6581.63	0.02(0.78) 1.00^F(0.13) 0.02/1.00^F(0.08) 1.00^F/1.00^F(0.01)			
CmC	231	-6489.73	0.03(0.84) 1.00^F(0.13) 4.64/2.33(0.03)	M2aREL	1.43	0.2317
CmD	232	-6486.78	0.03(0.86) 8.71(0.01) 1.38/3.71(0.13)	M3	7.21	0.0072
North American transition						
Br-site alt	230	-6586.12	0.02(0.83) 1.00^F(0.15) 0.02/1.42(0.02) 1.00^F/1.42(0.00)	Br-site null	0.01	0.9057
Br-site null	229	-6586.13	0.02(0.82) 1.00^F(0.15) 0.02/1.00(0.02) 1.00^F/1.00^F(0.00)			
CmC	231	-6489.95	0.03(0.84) 1.00^F(0.13) 4.55/2.19(0.03)	M2aREL	0.98	0.3219
CmD	232	-6490.10	0.04(0.86) 8.75(0.01) 1.47/2.58(0.13)	M3	0.57	0.4498
Mekong River transition						
Br-site alt	230	-6586.36	0.02(0.85) 1.00^F(0.15) 0.02/1.00^F(0.00) 1.00^F/1.00(0.00)	Br-site null	0.00	1.0000
Br-site null	229	-6586.36	0.02(0.85) 1.00^F(0.15) 0.02/1.00^F(0.00) 1.00^F/1.00^F(0.00)			
CmC	231	-6488.84	0.03(0.84) 1.00^F(0.13) 4.56/0.00(0.03)	M2aREL	3.20	0.0735
CmD	232	-6490.12	0.04(0.86) 8.74(0.01) 1.48/3.59(0.13)	M3	0.54	0.4631
South American transition						
Br-site alt	230	-6575.55	0.02(0.82) 1.00^F(0.14) 0.02/23.12(0.03) 1.00F/23.12(0.01)	Br-site null	9.52	0.0020
Br-site null	229	-6580.31	0.02(0.72) 1.00^F(0.12) 0.02/1.00^F(0.13) 1.00^F/1.00^F(0.02)			
CmC	231	-6490.25	0.03(0.84) 1.00^F(0.13) 4.48/8.02(0.03)	M2aREL	0.38	0.5370
CmD	232	-6487.31	0.03(0.86) 8.57(0.01) 1.40/5.78(0.13)	M3	6.16	0.0130

Note: lnL, ln likelihood; ^{F}, d_N/d_S estimate fixed at reported value; *LRT, likelihood ratio test result;* Br-site, Branch-site; CmC, Clade model C; CmD, Clade model D

Table S7. Branch-sites and clade model analyses (PAML) comparing d_N/d_S estimates for the transitional branches of the world-wide croaker rhodopsin dataset using the maximum likelihood rhodopsin gene tree

Partitions	np	lnL	Parameter Estimates: d_N/d_S (proportion of sites) (background d_N/d_S / foreground d_N/d_S)	Null	LRT	p value
M2aREL	230	-6206.98	0.02(0.84) 1.00^F(0.16) 8.94(0.01)			
M3	231	-6204.77	0.03(0.84) 1.21(0.15) 9.50(0.01)			
All transitional branches						
Br-site alt	230	-6271.17	0.02(0.82) 1.00^F(0.14) 0.02/6.30(0.03) 1.00^F/6.30(0.01)	Br-site null	4.4	0.0359
Br-site null	229	-6273.37	0.02(0.77) 1.00^F(0.14) 0.02/1.00^F(0.07) 1.00F/1.00^F(0.01)			
CmC	231	-6206.35	0.02(0.84) 1.00^F(0.16) 9.25/3.26(0.01)	M2aREL	1.25	0.2635
CmD	232	-6200.74	0.03(0.84) 9.49(0.01) 1.15/3.12(0.15)	M3	8.05	0.0045
North American (NA) transition						
Br-site alt	230	-6277.4	0.02(0.82) 1.00^F(0.15) 0.02/1.00(0.03) 1.00^F/1.00(0.01)	Br-site null	0.00	1.0000
Br-site null	229	-6277.4	0.02(0.82) 1.00^F(0.15) 0.02/1.00^F(0.03) 1.00^F/1.00^F(0.01)			
CmC	231	-6205.51	0.02(0.84) 1.00^F(0.16) 9.15/0.00(0.01)	M2aREL	2.94	0.0863
CmD	232	-6204.66	0.03(0.84) 9.50(0.01) 1.20/1.59(0.15)	M3	0.21	0.6457
Mekong River transition						
Br-site alt	230	-6277.64	0.02(0.84) 1.00^F(0.16) 0.02/1.00(0.00) 1.00^F/1.00(0.00)	Br-site null	0.00	1.0000
Br-site null	229	-6277.64	0.02(0.84) 1.00^F(0.16) 0.02/1.00^F(0.00) 1.00^F/1.00^F(0.00)			
CmC	231	-6206.32	0.02(0.84) 1.00^F(0.16) 9.04/0.00(0.01)	M2aREL	1.31	0.2519
CmD	232	-6204.23	0.03(0.85) 9.50(0.01) 1.20/2.88(0.15)	M3	1.07	0.3006
South American (SA) transition						
Br-site alt	230	-6267.03	0.02(0.82) 1.00^F(0.15) 0.02/25.27(0.03) 1.00^F/25.27(0.01)	Br-site null	10.3	0.0013
Br-site null	229	-6272.18	0.02(0.72) 1.00^F(0.13) 0.02/1.00^F(0.12) 1.00^F/1.00^F(0.02)			
CmC	231	-6206.81	0.02(0.84) 1.00^F(0.16) 8.85/17.73(0.01)	M2aREL	0.34	0.5606
CmD	232	-6200.16	0.03(0.84) 9.47(0.01) 1.16/5.42(0.15)	M3	9.21	0.0024

Note: lnL, ln likelihood; ^{F}, d_N/d_S estimate fixed at reported value; *LRT, likelihood ratio test result;* Br-site, Branch-site; CmC, Clade model C; CmD, Clade model D

Table S8. Branch-sites and clade model analyses (PAML) comparing d_N/d_S estimates in different ecological partitions in the South American croaker rhodopsin dataset using the Bayesian species tree

Model	np	lnL	Parameter Estimates: d_N/d_S (proportion of sites): background d_N/d_S / foreground d_N/d_S	Null	LRT	p value
Unpartitioned null models						
m0	92	-3535.69	0.24(1.00)			
M2aREL	95	-3328.05	0.02(0.84) 1.00^F(0.14) 8.99(0.01)			
M3	96	-3326.68	0.03(0.86) 1.37(0.13) 13.34(0.01)			
Freshwater clade and transitional branch						
Two-ratio	93	-3533.54	0.22/0.36(1.00)	m0	4.30	0.0381
Br-site alt	95	-3360.78	0.02(0.84) 1.00^F(0.12) 0.02/6.28(0.04) 1.00/6.28(0.01)	Br-site null	15.92	0.0001
Br-site null	94	-3368.74	0.01(0.81) 1.00^F(0.12) 0.01/1.00^F(0.07) 1.00/1.00^F(0.01)			
CmC	96	-3327.21	0.02(0.84) 1.00^F(0.14) 9.81/5.42(0.02)	M2a	1.68	0.1949
CmD	97	-3323.74	0.01(0.80) 4.28(0.04) 0.40/1.44(0.16)	M3	5.88	0.0153
Transitional branch						
Two-ratio	93	-3525.69	0.22/2.24(1.00)	m0	20.0	< 0.0001
Br-site alt	95	-3362.17	0.02(0.82) 1.00^F(0.12) 0.02/23.52(0.05) 1.00/23.52(0.01)	Br-site null	12.66	0.0004
Br-site null	94	-3368.50	0.01(0.70) 1.00^F(0.11) 0.01/1.00^F(0.16) 1.00/1.00^F(0.03)			
CmC	96	-3328.03	0.02(0.84) 1.00^F(0.14) 8.94/10.95(0.01)	M2a	0.04	0.8415
CmD	97	-3317.23	0.00(0.77) 4.12(0.05) 0.39/6.91(0.19)	M3	18.9	< 0.0001
Freshwater clade only						
Two-ratio	93	-3535.64	0.25/0.23(1.00)	m0	0.10	0.7518
Br-site alt	95	-3366.34	0.02(0.85) 1.00^F(0.14) 0.02/10.75(0.01) 1.00/10.75(0.00)	Br-site null	13.42	0.0002
Br-site null	94	-3373.05	0.02(0.84) 1.00^F(0.14) 0.02/1.00^F(0.02) 1.00/1.00^F(0.00)			
CmC	96	-3327.01	0.02(0.84) 1.00^F(0.14) 9.73/4.93(0.01)	M2a	2.08	0.1492
CmD	97	-3326.67	0.03(0.86) 13.32(0.01) 1.35/1.42(0.13)	M3	0.02	0.8875
Marine lineages only						
Br-site alt	95	-3337.09	0.02(0.83) 1.00^F(0.15) 0.02/9.53(0.02) 1.00/9.53(0.00)	Br-site null	74.26	< 0.0001
Br-site null	94	-3374.22	0.02(0.85) 1.00^F(0.15) 0.02/1.00^F(0.00) 1.00/1.00^F(0.00)			
Transitional branch and freshwater clade (separate foregrounds)						
Two-ratio	94	-3525.68	0.21/2.23/0.23(1.00)	m0	20.02	< 0.0001
CmC	97	-3327.01	0.02(0.84) 1.00^F(0.14) 9.74/9.48/4.93(0.01)	M2a	2.08	0.1492
CmD	98	-3315.91	0.01(0.78) 4.19(0.05) 0.36/8.21/0.72(0.17)	M3	21.54	< 0.0001
				Freshwater clade and transitional branch	15.66	< 0.0001
				Transitional branch	2.64	0.1042
				Freshwater clade only	21.52	< 0.0001

Note: lnL, ln likelihood; ^{F}, d_N/d_S estimate fixed at reported value; *LRT, likelihood ratio test result;* Br-site, Branch-site; CmC, Clade model C; CmD, Clade model D

Table S9. Branch-sites analyses (PAML) comparing d_N/d_S estimates in different ecological partitions in the control gene dataset using the Bayesian species tree

Foreground	np	lnL	Parameter Estimates: d_N/d_S (proportion of sites) (background d_N/d_S / foreground d_N/d_S)	LRT	p value
Rh1					
marine	62	-2587.14	0.03(0.86) 1.00(0.12) 0.03/10.23(0.02) 1.00/10.23(0.00)	54.38	< 0.0001
marine (null)	61	-2614.33	0.03(0.88) 1.00(0.12) 0.03/1.00(0.00) 1.00/1.00(0.00)		
Transitional branch	62	-2604.32	0.02(0.84) 1.00(0.10) 0.02/18.53(0.06) 1.00/18.53(0.01)	8.97	0.0027
Transitional branch (null)	61	-2608.81	0.02(0.69) 1.00(0.09) 0.02/1.00(0.20) 1.00/1.00(0.02)		
South American freshwater	62	-2609.18	0.03(0.88) 1.00(0.10) 0.03/7.55(0.02) 1.00/7.55(0.00)	8.10	0.0044
South American freshwater (null)	61	-2613.24	0.03(0.87) 1.00(0.10) 0.03/1.00(0.03) 1.00/1.00(0.00)		
Rag1					
marine	62	-3365.04	0.04(0.89) 1.00(0.09) 0.04/3.51(0.01) 1.00/3.51(0.00)	6.29	0.0121
marine (null)	61	-3368.19	0.03(0.88) 1.00(0.10) 0.03/1.00(0.02) 1.00/1.00(0.00)		
Transitional branch	62	-3367.71	0.03(0.58) 1.00(0.07) 0.03/1.00(0.31) 1.00/1.00(0.04)	0.00	1.0000
Transitional branch (null)	61	-3367.71	0.03(0.58) 1.00(0.07) 0.03/1.00(0.31) 1.00/1.00(0.04)		
South American freshwater	62	-3368.11	0.03(0.87) 1.00(0.11) 0.03/1.00(0.02) 1.00/1.00(0.00)	0.00	1.0000
South American freshwater (null)	61	-3368.11	0.03(0.87) 1.00(0.11) 0.03/1.00(0.02) 1.00/1.00(0.00)		
EGR1					
marine	62	-1793.39	0.00(0.93) 1.00(0.07) 0.00/1.00(0.00) 1.00/1.00(0.00)	0.00	1.0000
marine (null)	61	-1793.39	0.00(0.93) 1.00(0.07) 0.00/1.00(0.00) 1.00/1.00(0.00)		
Transitional branch	62	-1793.39	0.00(0.93) 1.00(0.07) 0.00/1.00(0.00) 1.00/1.00(0.00)	0.00	1.0000
Transitional branch (null)	61	-1793.39	0.00(0.93) 1.00(0.07) 0.00/1.00(0.00) 1.00/1.00(0.00)		
South American freshwater	62	-1792.15	0.00(0.91) 1.00(0.05) 0.00/1.00(0.04) 1.00/1.00(0.00)	0.00	1.0000
South American freshwater (null)	61	-1792.15	0.00(0.91) 1.00(0.05) 0.00/1.00(0.04) 1.00/1.00(0.00)		
EGR2					
marine	62	-1867.65	0.05(0.96) 1.00(0.04) 0.05/4.99(0.00) 1.00/4.99(0.00)	0.00	1.0000
marine (null)	61	-1867.65	0.05(0.96) 1.00(0.04) 0.05/1.00(0.00) 1.00/1.00(0.00)		
Transitional branch	62	-1867.65	0.05(0.89) 1.00(0.03) 0.05/2.06(0.08) 1.00/2.06(0.00)	0.00	1.0000
Transitional branch (null)	61	-1867.65	0.05(0.95) 1.00(0.04) 0.05/1.00(0.01) 1.00/1.00(0.00)		
South American freshwater	62	-1867.65	0.05(0.96) 1.00(0.04) 0.05/1.00(0.00) 1.00/1.00(0.00)	0.00	1.0000
South American freshwater (null)	61	-1867.65	0.05(0.96) 1.00(0.04) 0.05/1.00(0.00) 1.00/1.00(0.00)		

Note: lnL, ln likelihood; *LRT, likelihood ratio test result;*

Table S10. Clade model analyses (PAML) comparing d_N/d_S estimates in different ecological partitions in the control gene dataset using the Bayesian species tree

Foreground	np	lnL	Parameter Estimates: d_N/d_S (proportion of sites) (background d_N/d_S / foreground d_N/d_S)	null	LRT	p value
Rh1						
m2aREL (null)	62	-2577.2	0.03(0.87) 1.00(0.11) 8.95(0.02)			
m3 (null)	63	-2576.12	0.05(0.90) 1.61(0.09) 12.70(0.01)			
CmC / Marine lineages	63	-2576.63	0.03(0.87) 1.00(0.11) 5.91/9.92(0.02)	M2aREL	1.15	0.2843
CmD / Marine lineages	64	-2570.42	0.03(0.86) 5.73(0.03) 2.27/0.39(0.12)	M3	11.41	0.0007
CmC / Transitional branch	63	-2577.19	0.03(0.87) 1.00(0.11) 8.89/10.25(0.02)	M2aREL	0.02	0.8846
CmD / Transitional branch	64	-2569.00	0.03(0.85) 7.43(0.02) 0.68/11.88(0.13)	M3	14.25	0.0002
CmC / Freshwater clade	63	-2576.48	0.03(0.87) 1.00(0.11) 9.78/5.32(0.02)	M2aREL	1.43	0.2318
CmD / Freshwater clade	64	-2575.34	0.04(0.90) 9.59(0.01) 1.18/2.11(0.09)	M3	1.56	0.2111
Rag1						
m2aREL (null)	62	-3367.07	0.04(0.89) 1.00(0.08) 2.08(0.02)			
m3 (null)	63	-3367.04	0.04(0.89) 0.80(0.08) 1.89(0.04)			
CmC / Marine lineages	63	-3364.94	0.04(0.89) 1.00(0.10) 0.00/3.55(0.01)	M2aREL	4.26	0.0389
CmD / Marine lineages	64	-3364.92	0.04(0.89) 1.04(0.09) 0.00/3.55(0.01)	M3	4.24	0.0396
CmC / Transitional branch	63	-3367.03	0.04(0.90) 1.00(0.08) 2.08/0.00(0.02)	M2aREL	0.08	0.7740
CmD / Transitional branch	64	-3366.56	0.09(0.48) 1.40(0.09) 0.00/0.74(0.43)	M3	0.96	0.3269
CmC / Freshwater clade	63	-3365.07	0.04(0.89) 1.00(0.10) 3.55/0.00(0.01)	M2aREL	4.01	0.0453
CmD / Freshwater clade	64	-3365.05	0.04(0.89) 1.05(0.09) 3.56/0.00(0.01)	M3	3.98	0.0460
EGR1						
m2aREL (null)	62	-1792.98	0.00(0.91) 1.00(0.00) 0.69(0.09)			
m3 (null)	63	-1792.98	0.00(0.20) 0.00(0.72) 0.69(0.09)			
CmC / Marine lineages	63	-1792.03	0.00(0.00) 1.00(0.05) 0.03/0.00(0.95)	M2aREL	1.89	0.1696
CmD / Marine lineages	64	-1791.97	0.00(0.00) 0.85(0.06) 0.03/0.00(0.94)	M3	2.04	0.1527
CmC / Transitional branch	63	-1792.86	0.00(0.91) 1.00(0.00) 0.70/0.00(0.09)	M2aREL	0.24	0.6223
CmD / Transitional branch	64	-1792.86	0.00(0.49) 0.00(0.42) 0.70/0.00(0.09)	M3	0.26	0.6068
CmC / Freshwater clade	63	-1791.96	0.00(0.00) 1.00(0.05) 0.00/0.04(0.95)	M2aREL	2.04	0.1534
CmD / Freshwater clade	64	-1791.89	0.00(0.00) 0.85(0.06) 0.00/0.03(0.94)	M3	2.19	0.1390
EGR2						
m2aREL (null)	62	-1866.40	0.06(0.99) 1.00(0.00) 4.01(0.01)			
m3 (null)	63	-1866.40	0.06(0.99) 4.01(0.01) 4.01(0.00)			
CmC / Marine lineages	63	-1866.19	0.06(0.99) 1.00(0.00) 6.00/3.12(0.01)	M2aREL	0.42	0.5185
CmD / Marine lineages	64	-1866.19	0.06(0.40) 0.06(0.59) 6.00/3.12(0.01)	M3	0.42	0.5185
CmC / Transitional branch	63	-1866.40	0.06(0.99) 1.00(0.00) 4.01/4.49(0.01)	M2aREL	0.00	1.0000
CmD / Transitional branch	64	-1866.40	0.06(0.92) 0.06(0.07) 4.01/3.14(0.01)	M3	0.00	1.0000

CmC / Freshwater clade	63	-1866.19	0.06(0.99) 1.00(0.00) 3.12/6.00(0.01)	M2aREL	0.42	0.5185
CmD / Freshwater clade	64	-1866.19	0.06(0.31) 0.06(0.68) 3.12/6.00(0.01)	M3	0.42	0.5185

Note: lnL, ln likelihood; *LRT, likelihood ratio test result;* CmC, Clade model C; CmD, Clade model D

Table S11. Clade model analyses (PAML) comparing d_N/d_S on the South American transitional branch of the South American rhodopsin dataset with and without highly positively selected sites removed using the Bayesian species tree

Model	np	lnL	Parameter Estimates: dN/dS (proportion of sites) (background dN/dS / Foreground dN/dS)	Null	LRT	p value
Full rhodopsin dataset						
M3	96	-3326.68	0.03(0.86) 1.37(0.13) 13.34(0.01)			
M3 - 4 site classes	98	-3315.31	0.00(0.74) 0.40(0.20) 2.63(0.05) 13.64(0.01)	M3	22.74	< 0.0001
M2aREL	95	-3328.05	0.02(0.84) 1.00^F(0.14) 8.99(0.01)			
CmC	96	-3328.03	0.02(0.84) 1.00^F(0.14) 8.94/10.95(0.01)	M2a	0.04	0.8415
CmD	97	-3317.25	0.00(0.77) 4.12(0.05) 0.39/6.91(0.19)	M3	18.86	< 0.0001
				CmC	21.56	< 0.0001
Rhodopsin dataset with sites 165 and 214 removed (sites in 4th site class in the M3 analysis with four site classes)						
M3	96	-3145.94	0.00(0.75) 0.40(0.20) 2.63(0.05)			
M3 - 4 site classes	98	-3145.94	0.00(0.75) 0.40(0.20) 2.63(0.05) 3.82(0.00)	M3	0.00	1.0000
M2aREL	95	-3150.45	0.02(0.85) 1.00^F(0.11) 3.05(0.04)			
CmC	96	-3145.42	0.01(0.81) 1.00^F(0.11) 0.24/14.76(0.08)	M2a	10.06	0.0015
CmD	97	-3131.94	0.00(0.75) 2.76(0.05) 0.34/6.32(0.20)	M3	28.04	< 0.0001

Note: lnL, ln likelihood; ^{F}, d_N/d_S estimate fixed at reported value; *LRT, likelihood ratio test result;* CmC, Clade model C; CmD, Clade model D

Table S12. Substitutions on the South American transitional branch from marginal ancestral reconstructions and the frequency of their occurrence on other branches in the tree

Substitution on transitional branch (WAG)	WAG (PPm|PPf)	Dayhoff (PPm|PPf)	JTT (PPm|PPf)	M0 (PPm|PPf)	Number of substitutions at the same site	Number of identical substitutions	Conserved in the freshwater clade (*)
S33N	1|1	1|1	1|1	0.999|0.998	15	13	
G39A	1|0.794	1|0.792	1|0.794	1|1(G)	2	2	
F50L	1|0.901	1|0.797	1|0.924	1(L)|1	6	1	
F52L	1|0.999	1|0.999	1|0.999	1|0.992(F)	3	1	
L63I	1|1	1|1	1|1	1|0.999	4	3	*
L119F	1|1	1|1	1|1	0.999|0.997	1	1	*
E122I	1|1	1|1	1|1	1|0.994	1	1	*
S124A	1|0.999	1|0.999	1|0.999	1|0.999	3	3	*
G158A	1|1	1|1	1|1	1|0.999	14	8	*
F159L	1|1	1|1	1|1	1|1	3	2	*
S165V	0.996|0.805	0.995|0.901	0.996|0.801	0.363|0.605(F)	14	1	
V169G	1|1	1|1	1|1	1|1	17	1	*
V173I	1|0.999	1|0.999	1|0.999	1|0.995	7	2	
V209T	1|1	1|1	1|1	0.851|0.997	21	3	*
V214T	1|1	1|1	1|1	0.85|0.998	33	4	
V218I	1|0.999	1|0.999	1|0.999	0.929|0.999	16	12	
V259I	1|0.999	1|0.999	1|0.999	0.93|0.997	13	6	*
F261Y	1|0.999	1|0.999	1|1	1|0.999	1	1	*
E282D	1|1	1|1	1|1	1|0.999	2	1	
M304A	0.998|1	0.997|1	0.998|1	0.999|0.998	6	4	

Note: PPm, best supported amino acid identity in marine ancestor; PPf, best supported amino acid identity in freshwater ancestor; If the identity differs from the identity according to the WAG reconstruction the alternative amino acid is shown in parentheses

Table S13. Peak spectral sensitivity of croaker rhodopsin from the literature and replicate spectroscopic assays of ancestrally reconstructed rhodopsin sequences

Pigment	Spectral sensitivity of rhodopsin
Bovine rhodopsin	498.7, 498.8, 498.3
Bovine rhodopsin (A124S)	496.4 (Castiglione and Chang 2018)
Bovine rhodopsin (L119F, E122I, F261Y)	505.6, 505.7, 505.6
Ancestral marine croaker rhodopsin	498.1, 497.3, 499.6
Ancestral freshwater croaker rhodopsin	503.8, 504.5, 504.6
Published MSP estimates of croaker rhodopsin	
Cynoscion regalis	496.0 (Beatty 1973)
Micropogon undulatus	496.5 (Beatty 1973)
Leiostomus xanthursus	499.0 (Beatty 1973)

Table S14. Primers used in this study

Primer name	Sequence (5' to 3')	Gene	Reference
RH 193F	CNTATGAATAYCCTCAGTACTACC	rh1	Chen et al. (2003)
RH 1039R	TGCTTGTTCATGCAGATGTAGA	rh1	Chen et al. (2003)
Rho124F	GCCTACATGTTCTTYCTCATC	rh1	this study
Rho934R	GCTTGTTCATGCAGATGTA	rh1	this study
E1 225F	CCTGAYATCCCCTTCAACTGTG	egr 1	Lo et al. (2015)
E1 284F	CCCCCATCTCYTACACAGG	egr 1	Lo et al. (2015)
E1 290F	TMTCTTACACAGGCCGYTTCAC	egr 1	Lo et al. (2015)
E1 1104R	CGCAGGTGGATCTTRGTGTG	egr 1	Lo et al. (2015)
E1 1118R	CTTCTTGTCCTTCTGCCGYAGRT	egr 1	Lo et al. (2015)
E2B ex1PcoF	CYAAAACTTTGGAGAAAGTGC	egr 2	Lo et al. (2015)
E2B ex2intR2	ACTGGGARTCGATGGAGAACT	egr 2	Lo et al. (2015)
E1B 1117R	AGGTGGATTTTGGTGTGTCTYTT	egr 2	Lo et al. (2015)
E2B 1121R	CCTCAGGTGGATTTTAGTGTGTC	egr 2	Lo et al. (2015)
Rag1 f2	CTGAGCTGCAGTCAGTACCATAAGATGT	Rag1	Lopez et al. (2004)
Rag1 r1	GTGTAGAGCCAGTGGTGYTT	Rag1	Lopez et al. (2004)
GLUDGL	CGAAGCTTGACTTGAARAACCAYCGTTG	cyt b	Ward et al. (2009)
cyt bR	CTCCGATCTTCGGATTACAAG	cyt b	Ward et al. (2009)
co1fishF1	TCAACYAATCAYAAAGATATYGGCAC	coi	Baldwin et al. (2009)
co1fishR1	ACTTCYGGGTGRCCRAARAATCA	coi	Baldwin et al. (2009)

Table S15. Genbank accession numbers and museum catalog numbers for sequences used in phylogenetic reconstructions and selection analyses.

Species	taxa code	Isolate number	Voucher number	cyt b	COI	RAG1	RH1	EGR1	EGR2	Dataset usage
Cynoscion acoupa	1	480	NA	MT879777		MT879894	MT879935	MT879852	MT879975	
Cynoscion albus	2	13730	ROM T13765	MT879778	MT879814	MT879895	MT879936	MT879853	MT879976	Rho
Cynoscion arenarius	3	11630	NA	MT879779	MT879815	MT879896	MT879937	MT879854	MT879977	Rho
Cynoscion nebulosus	4	11628	NA		MT879816	MT879897	MT879938	MT879855	MT879978	Rho
Cynoscion nothus	5	11626	NA	MT879780		MT879898	MT879939	MT879856		Rho
Cynoscion praedatorius	6	13731	ROM T13978	MT879781	MT879817	MT879899	MT879940	MT879857	MT879979	
Isopisthus parvipinnis	7	10532	ROM T7713	MT879782	MT879818	MT879900	MT879941	MT879858	MT879980	Rho
Larimus fasciatus	8	11625	NA	MT879783	MT879819	MT879901	MT879942	MT879859		Rho SAC
Macrodon ancylodon	9	10534	ROM T7717	MT879784		MT879902	MT879943	MT879860	MT879981	
Menticirrhus paitensis	10	13732	ROM T13731	MT879785	MT879820	MT879903	MT879944	MT879861	MT879982	Rho
Menticirrhus sp. 1	11	10531	NA	MT879786	MT879821	MT879904	MT879945	MT879862	MT879983	Rho
Micropogonias furnieri	12	10540	ROM T8290	MT879787		MT879905	MT879946	MT879863	MT879984	
Nebris microps	13	10539	ROM T7728	MT879788	MT879822	MT879906	MT879947	MT879864	MT879985	
Nebris sp. 1	14	13733	ROM T13741	MT879789	MT879823	MT879907	MT879948	MT879865	MT879986	Rho SAC Con
Ophioscion scierus	15	13734	ROM T13732	MT879790	MT879824	MT879908	MT879949	MT879866	MT879987	Rho SAC Con
Ophioscion sp. 1	16	13735	ROM T08723	MT879791	MT879825	MT879909	MT879950	MT879867		Rho SAC
Pachypops fourcroi	17	11599	ANSP 40562	MT879792	MT879826	MT879910	MT879951	MT879868	MT879988	
Pachypops sp. 1	18	11601	ANSP 197648	MT879793	MT879827	MT879911	MT879952	MT879869		Rho SAC
Pachyurus bonariensis	19	11595	NA	MT879794	MT879828	MT879912	MT879953	MT879870	MT879989	
Pachyurus cf. paucirastrus	20	11596	ANSP 199599	MT879795	MT879829	MT879913	MT879954	MT879871	MT879990	Rho SAC Con
Pachyurus junki	21	11597	ANSP 193039	MT879796	MT879830	MT879914	MT879955	MT879872		
Pachyurus junki	22	11604	ANSP 198701	MT879797	MT879831	MT879915	MT879956	MT879873		Rho SAC
Pachyurus schomburgkii	23	528	NA		MT879832	MT879916	MT879957	MT879874	MT879991	Rho SAC
Paralonchurus dumerilii	24	13736	ROM T13891	MT879798	MT879833	MT879917	MT879958	MT879875	MT879992	
Paralonchurus dumerilii	25	13737	ROM T13728	MT879799	MT879834	MT879918	MT879959	MT879876	MT879993	Rho SAC Con
Paralonchurus sp. 1	26	10538	ROM T7723	MT879800	MT879835	MT879919	MT879960	MT879877	MT879994	Rho SAC Con
Petilipinnis grunniens	27	11591	ANSP 187423	MT879801	MT879836	MT879920	MT879961	MT879878	MT879995	Rho SAC Con
Plagioscion auratus	28	11602	ANSP 197653	MT879802	MT879837	MT879921	MT879962	MT879879	MT879996	
Plagioscion montei	29	698	NA	MT879803	MT879838	MT879922	MT879963	MT879880	MT879997	Rho SAC Con
Plagioscion squamosissimus	30	448	NA	MT879804	MT879839	MT879923	MT879964	MT879881	MT879998	
Plagioscion squamosissimus	31	447	NA	MT879805	MT879840	MT879924	MT879965	MT879882	MT879999	Rho SAC Con
Plagioscion squamosissimus	32	728	SIUC 37984		MT879841	MT879925	MT879966	MT879883	MT880000	
Plagioscion squamosissimus	33	377	NA	MT879806	MT879842	MT879926		MT879884	MT880001	
Plagioscion squamosissimus	34	529	NA		MT879843	MT879927		MT879885	MT880002	
Plagioscion squamosissimus	35	632	INHS 54286		MT879844	MT879928		MT879886		
Pogonias cromis	36	11634	NA	MT879807	MT879845	MT879929	MT879967	MT879887	MT880003	Rho
Sciaenops ocellatus	37	11629	NA		MT879846		MT879968			
Stellifer chrysoleuca	38	13738	ROM T13888	MT879808	MT879847	MT879930	MT879969	MT879888	MT880004	Rho SAC Con
Stellifer lanceolatus	39	11631	NA	MT879809	MT879848	MT879931	MT879970	MT879889		Rho SAC
Stellifer sp. 1	40	13739	ROM T13734	MT879810	MT879849	MT879932	MT879971	MT879890	MT880005	Rho SAC Con
Stellifer sp. 2	41	10536	ROM T7720	MT879811		MT879933	MT879972	MT879891	MT880006	Rho SAC

Stellifer stellifer	42	10535	ROM T7719	MT879812	MT879850	MT879934	MT879973	MT879892	MT880007	Rho SAC Con
Umbrina sp. 1	43	10530	NA	MT879813	MT879851		MT879974	MT879893	MT880008	Rho SAC
Aplodinotus grunniens	44	NA		KP722606	KP722699	KP722882	KP722970	KP722788	KP723062	Rho
Argyrosomus japonicus	45	NA		KP722607	KP722700	KP722883	KP722971	KP722789	KP723063	Rho
Argyrosomus regius	46	NA		KP722608	KP722701	KP722884	KP722972	KP722790	KP723064	Rho
Atractoscion nobilis	47	NA		KP722609	EU547246	KP722885	KP722973	KP722791	KP723065	Rho
Atrobucca nibe	48	NA		KP722610	KP722702	KP722886	KP722974	KP722792	KP723066	Rho
Austronibea oedogenys	49	NA		KP722611		KP722887	KP722975	KP722793	KP723067	Rho
Bahaba taipingensis	50	NA		KP722612	KP722703	KP722888	KP722976	KP722794	KP723068	Rho
Bairdiella cf armata	51	NA		KP722613	KP722704	KP722889	KP722977	KP722795	KP723069	Rho SAC Con
Bairdiella ronchus	52	NA		KP722614	KP722705		KP722978	KP722796	KP723070	Rho SAC
Boesemania microlepis	53	NA		KP722615	KP722706	KP722890	KP722979	KP722797	KP723071	Rho
Cheilotrema saturnum	54	NA		KP722616	KP722707	KP722891	KP722980	KP722798	KP723072	Rho SAC Con
Chrysochir aureus	55	NA		KP722617	KP722708	KP722892	KP722981	KP722799	KP723073	Rho
Cilus gilberti	56	NA		KP722618	KP722709	KP722893	KP722982	KP722800	KP723074	Rho SAC Con
Collichthys lucidus	57	NA		KP722619	KP722710	KP722894	KP722983	KP722801	KP723075	Rho
Corvula macrops	58	NA		KP722620	KP722711	KP722895	KP722984	KP722802	KP723076	
Corvula macrops	59	NA			KP722712	KP722896	KP722985	KP722803	KP723077	Rho SAC
Cynoscion acoupa	60	NA		KP722621	KP722713	KP722897	KP722986	KP722804	KP723078	Rho
Cynoscion guatucupa	61	NA		KP722622	KP722714	KP722898	KP722987	KP722805	KP723079	Rho
Cynoscion parvipinnis	62	NA		KP722623	KP722715	KP722899	KP722988	KP722806	KP723080	Rho
Cynoscion praedatorius	63	NA		KP722624	KP722716	KP722900	KP722989	KP722807	KP723081	Rho
Cynoscion regalis	64	NA		KP722625	KP722717	KP722901	KP722990	KP722808	KP723082	Rho
Cynoscion reticulatus	65	NA		KP722626	KP722718	KP722902	KP722991	KP722809	KP723083	Rho
Daysciaena albida	66	NA		KP722627	KP722719	KP722903	KP722992	KP722810	KP723084	Rho
Dendrophysa russelii	67	NA		KP722628	KP722720	KP722904	KP722993	KP722811	KP723085	Rho
Dicentrarchus labrax	68	NA		AP009166	AP009166	KP722969	KP723061	KP722881	KP723151	
Equetus lanceolatus	69	NA		KP722629	KP722721	KP722905	KP722994	KP722812	KP723086	Rho SAC Con
Genyonemus lineatus	70	NA		KP722630	KP722722	KP722906	KP722995	KP722813		Rho SAC
Isopisthus remifer	71	NA		KP722631	KP722723	KP722907	KP722996	KP722814	KP723087	Rho
Johnius amblycephalus	72	NA		KP722632	KP722724	KP722908	KP722997	KP722815	KP723088	Rho
Johnius belangerii	73	NA		KP722633	KP722725	KP722909	KP722998	KP722816	KP723089	Rho
Johnius borneensis	74	NA		KP722634		KP722910	KP722999	KP722817	KP723090	Rho
Johnius carouna	75	NA		KP722635	KP722726	KP722911	KP723000	KP722818	KP723091	Rho
Johnius distinctus	76	NA		KP722636		KP722912	KP723001	KP722819	KP723092	Rho
Johnius heterolepis	77	NA		KP722637		KP722913	KP723002	KP722820	KP723093	Rho
Johnius macropterus	78	NA		KP722638	KP722727	KP722914	KP723003	KP722821	KP723094	Rho
Johnius majan	79	NA		KP722639	KP722728	KP722915	KP723004	KP722822	KP723095	Rho
Johnius trewavasae	80	NA		KP722640	KP722729	KP722916	KP723005	KP722823	KP723096	Rho
Larimichthys crocea	81	NA		KP722641	KP722730	KP722917	KP723006	KP722824	KP723097	Rho
Larimichthys polyactis	82	NA		KP722642	KP722731	KP722918	KP72300	KP722825	KP723098	Rho
Larimus pacificus	83	NA		KP722643	KP722732	KP722919	KP723008	KP722826	KP723099	Rho SAC Con
Leiostomus xanthurus	84	NA		KP722644	KP722733	KP722920	KP723009	KP722827	KP723100	Rho
Macrodon ancylodon	85	NA		KP722645	KP722734	KP722921	KP723010	KP722828	KP723101	Rho
Megalonibea fusca	86	NA		KP722646	KP722735	KP722922	KP723011	KP722829	KP723102	Rho
Menticirrhus americanus	87	NA		KP722647	KP722736	KP722923	KP723012	KP722830	KP723103	Rho
Menticirrhus undulatus	88	NA		KP722648	KP722737	KP722924	KP723013	KP722831	KP723104	Rho
Micropogonias ectenes	89	NA		KP722649	KP722738	KP722925	KP723014	KP722832	KP723105	Rho
Micropogonias furnieri	90	NA		KP722650	KP722739	KP722926	KP723015	KP722833	KP723106	Rho

Micropogonias undulatus	91	NA	KP722651	KP722740	KP722927	KP723016	KP722834	KP723107	Rho
Miichthys miiuy	92	NA	KP722652	KP722741	KP722928	KP723017	KP722835	KP723108	Rho
Monotaxis grandoculis	93	NA	EU036430	FN689114	EF095651	Y18673	KC442100	KC442134	
Nebris microps	94	NA	KP722653	KP722742	KP722929	KP723018	KP722836	KP723109	Rho SAC
Nibea albiflora	95	NA	KP722654	KP722743	KP722930	KP723019	KP722837	KP723110	Rho
Nibea microgenys	96	NA	KP722656	KP722745	KP722931	KP723020	KP722838	KP723111	Rho
Nibea soldado	97	NA	KP722657	KP722746	KP722932	KP723021	KP722839	KP723112	Rho
Nibea squamosa	98	NA	KP722658	KP722747	KP722933	KP723022	KP722840	KP723113	Rho
Odontoscion xanthops	99	NA	KP722659	KP722748		KP723023	KP722841		Rho SAC
Ophioscion punctatissimus	100	NA	KP722660	KP722749	KP722934	KP723024	KP722842	KP723114	Rho SAC Con
Ophioscion scierus	101	NA	KP722661	KP722750		KP723025	KP722843	KP723115	
Ophioscion vermicularis	102	NA	KP722662	KP722751	KP722935	KP723026	KP722844	KP723116	Rho SAC
Otolithes ruber	103	NA	KP722663	KP722752	KP722936	KP723027	KP722845	KP723117	Rho
Pachypops fourcroi	104	NA	KP722664	KP722753	KP722937	KP723028	KP722846	KP723118	Rho SAC Con
Pachyurus bonariensis	105	NA	KP722665	KP722754	KP72293	KP723029	KP722847	KP723119	Rho SAC Con
Panna microdon	106	NA	KP722666	KP722755	KP722939	KP723030	KP722848	KP723120	Rho
Paralonchurus brasiliensis	107	NA	KP722667	KP722756	KP722940	KP723031	KP722849	KP723121	Rho SAC Con
Pareques sp. 1	108	NA	KP722668	KP722757		KP723032	KP722850		Rho SAC
Pennahia argentata	109	NA	KP722669	KP722758	KP722941	KP723033	KP722851	KP723122	Rho
Pennahia macrocephalus	110	NA	KP722670	KP722759	KP722942	KP723034	KP722852	KP723123	Rho
Pennahia pawak	111	NA	KP722671	KP722760	KP722943	KP723035	KP722853	KP723124	Rho
Plagioscion auratus	112	NA	KP722672	KP722761	KP722944	KP723036	KP722854	KP723125	Rho SAC
Plagioscion squamosissimus	113	NA	KP722673	KP722762	KP722945	KP723037	KP722855	KP723126	
Plagioscion surinamensis	114	NA	KP722674	KP722763	KP722946	KP723038	KP722856	KP723127	Rho SAC Con
Plagioscion ternetzi	115	NA	KP722675	KP722764			KP722857		
Pogonias cromis	116	NA	KP722676	KP722765	KP722947	KP723039	KP722858	KP723128	
Protonibea diacanthus	117	NA	KP722677	KP722766	KP722948	KP723040	KP722859	KP723129	Rho
Pseudotolithus brachygnathus	118	NA	KP722678	KP722767	KP722949	KP723041	KP722860	KP723130	Rho
Pseudotolithus elongatus	119	NA	KP722679	KP722768	KP722950	KP723042	KP722861	KP723131	Rho
Pseudotolithus senegallus	120	NA	KP722680	KP722769	KP722951	KP723043	KP722862	KP723132	Rho
Pseudotolithus typus	121	NA	KP722681	KP722770		KP723044	KP722863	KP723133	Rho
Pteroscion peli	122	NA	KP722682	KP722771	KP722952	KP723045	KP722864	KP723134	Rho
Pterotolithus maculatus	123	NA	KP722683	KP722772	KP722953	KP723046	KP722865	KP723135	Rho
Roncador stearnsii	124	NA	KP722684	KP722773	KP722954	KP723047	KP722866	KP723136	Rho SAC Con
Sciaena deliciosa	125	NA	KP722685	KP722774	KP722955	KP723048	KP722867	KP723137	Rho SAC Con
Sciaena umbra	126	NA	KP722686	KP722775	KP722956	KP723049	KP722868	KP723138	Rho
Sciaenops ocellatus	127	NA	KP722687	KP722776	KP722957	KP723050	KP722869	KP723139	Rho
Seriphus politus	128	NA	KP722688	KP722777	KP722958	KP723051	KP722870	KP723140	Rho SAC Con
Sparus aurata	129	NA	AF240735	FN689315	EF095657	Y18665	KC442100	KC442134	
Stellifer ericymba	130	NA	KP722689	KP722778	KP722959	KP723052	KP722871	KP723141	Rho SAC Con
Stellifer microps	131	NA	KP722690	KP722779	KP722960	KP723053	KP722872	KP723142	Rho SAC Con
Stellifer oscitans	132	NA	KP722691	KP722780	KP722961	KP723054	KP722873	KP723143	Rho SAC Con
Stellifer rastrifer	133	NA	KP722692	KP722781	KP722962	KP723055	KP722874	KP723144	Rho SAC Con
Totoaba macdonaldi	134	NA	KP722693	KP722782	KP722963	KP723056	KP722875	KP723145	Rho
Umbrina bussingi	135	NA	KP722694	KP722783	KP722964	KP723057	KP722876	KP723146	Rho SAC
Umbrina canariensis	136	NA	KP722695	KP722784	KP722965		KP722877	KP723147	
Umbrina cirrosa	137	NA	KP722696	KP722785	KP722966	KP723058	KP722878	KP723148	Rho
Umbrina roncador	138	NA	KP722697	KP722786	KP722967	KP723059	KP722879	KP723149	Rho SAC
Umbrina xanti	139	NA	KP722698	KP722787	KP722968	KP723060	KP722880	KP723150	Rho SAC Con

NOTE. Rho, Rhodopsin sequence included in dataset for analyses of molecular evolution; Con, Sequences used in control gene dataset; SAC, Rhodopsin sequences used in analyses of just the South American clade and marine sister species.

Table S16. Results for Partition Finder analyses of concatenated dataset.

Gene	Length	Mr. Bayes		RaxML	
		Partition	Model	Partition	Model
Rh1	827	1	K80+I+G	1	GTR+I+G
Cytb	1104	2	GTR+I+G	2	GTR+I+G
COI	648	3	GTR+I+G	3	GTR+I+G
EGR1	896	4	HKY+I+G	4	GTR+I+G
EGR2	1124	4	HKY+I+G	4	GTR+I+G
EGR2 intron	422	5	K80+I+G	5	GTR+G
Rag1	1439	6	SYM+I+G	6	GTR+I+G

Supplementary figures

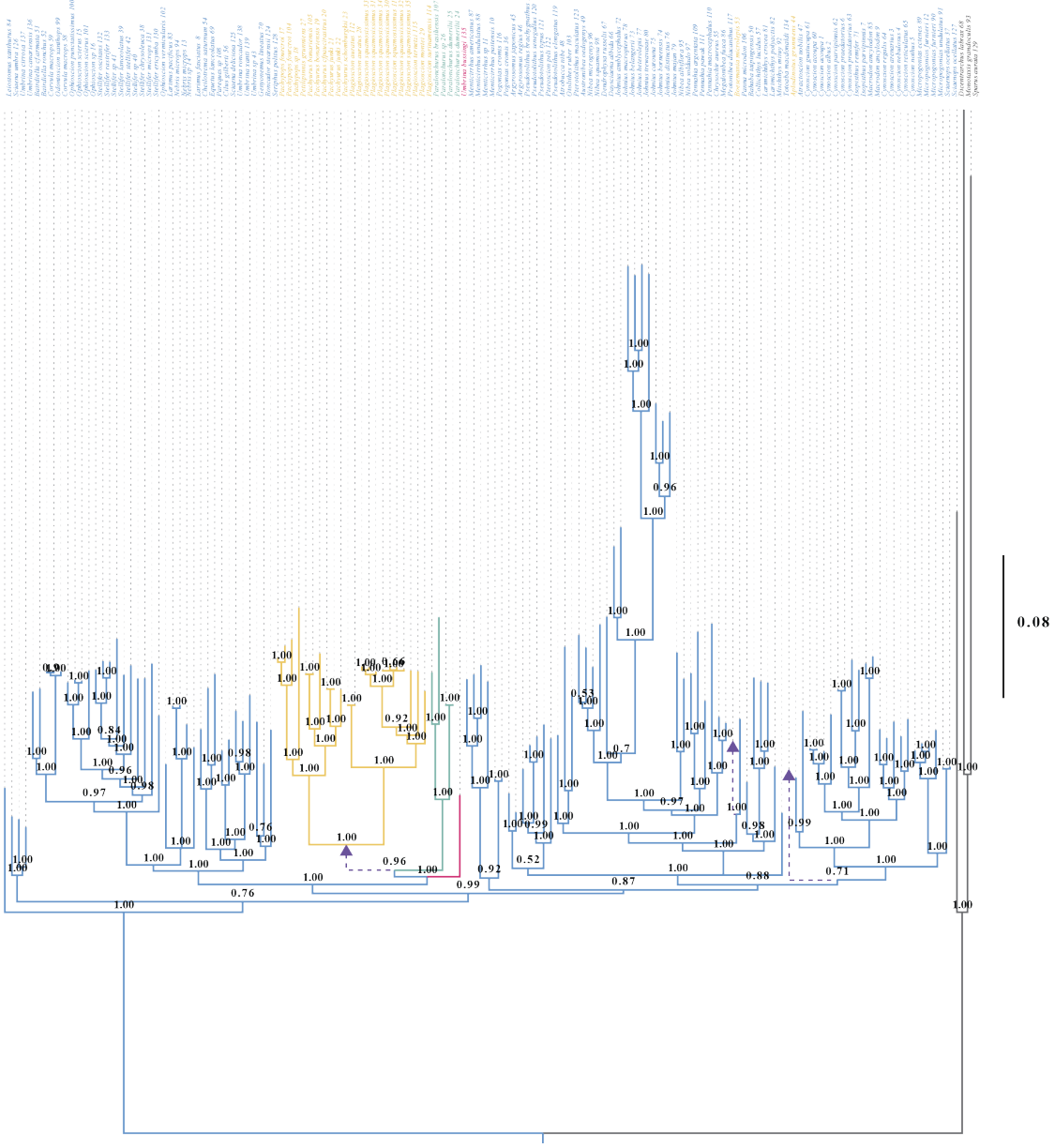

Fig. S1. Bayesian species tree reconstructed using a concatenated alignment of four nuclear loci and two mitochondrial loci. Posterior probability values displayed at nodes. Marine lineages in blue, freshwater lineages in yellow, transitional branches shown as purple arrows, and outgroup taxa shown in grey. Green and pink branches show the two possible sister groups to the South American freshwater clade. Branches proportional to number of substitutions per site.

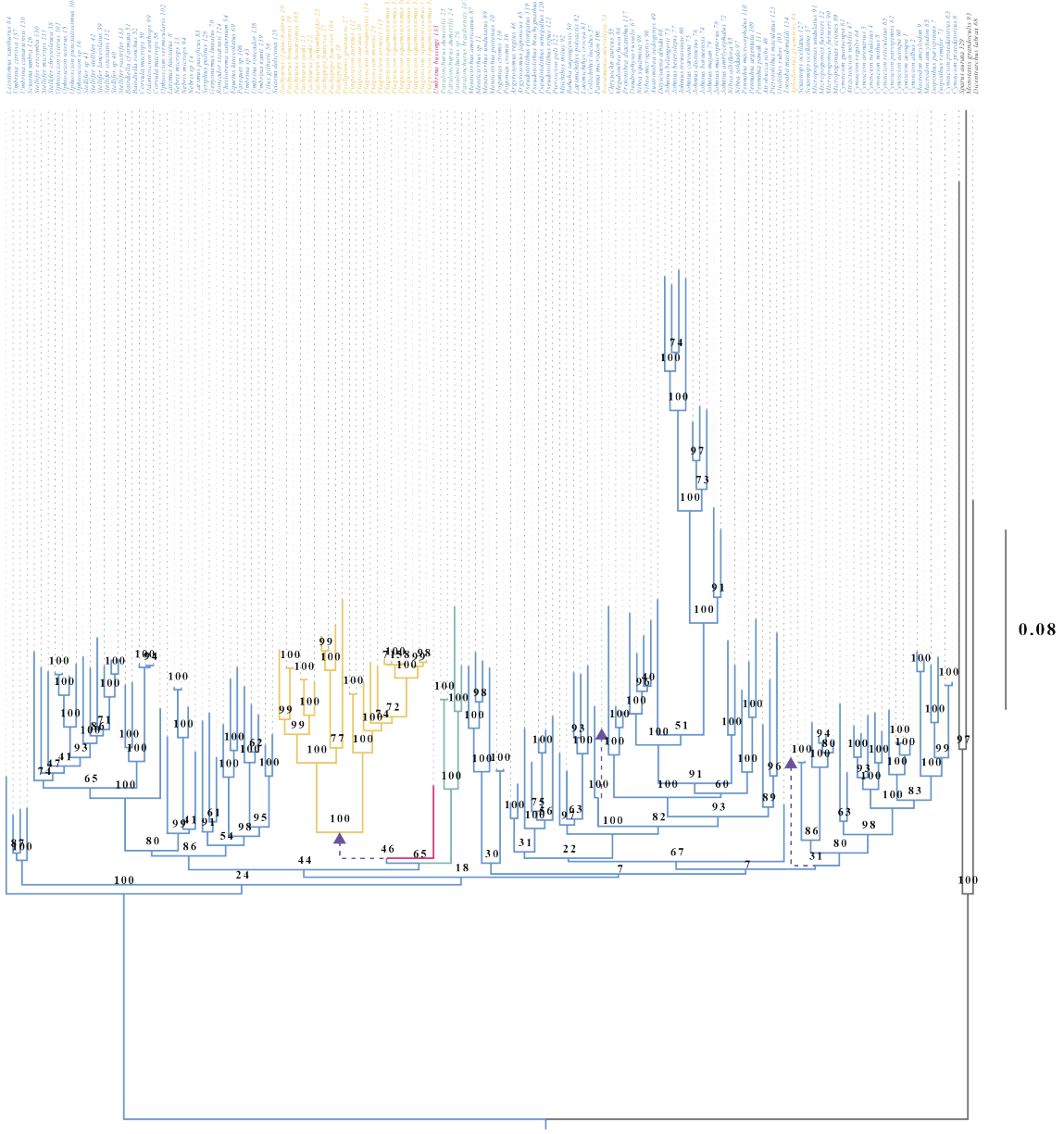

Fig. S2. Maximum likelihood species tree reconstructed using a concatenated alignment of four nuclear loci and two mitochondrial loci. Bootstrap support values from 1000 replicate analyses displayed at nodes. Marine lineages in blue, freshwater lineages in yellow, transitional branches

shown as purple arrows, and outgroup taxa shown in grey. Green and pink branches show the two possible sister groups to the South American freshwater clade. Branches proportional to number of substitutions per site.

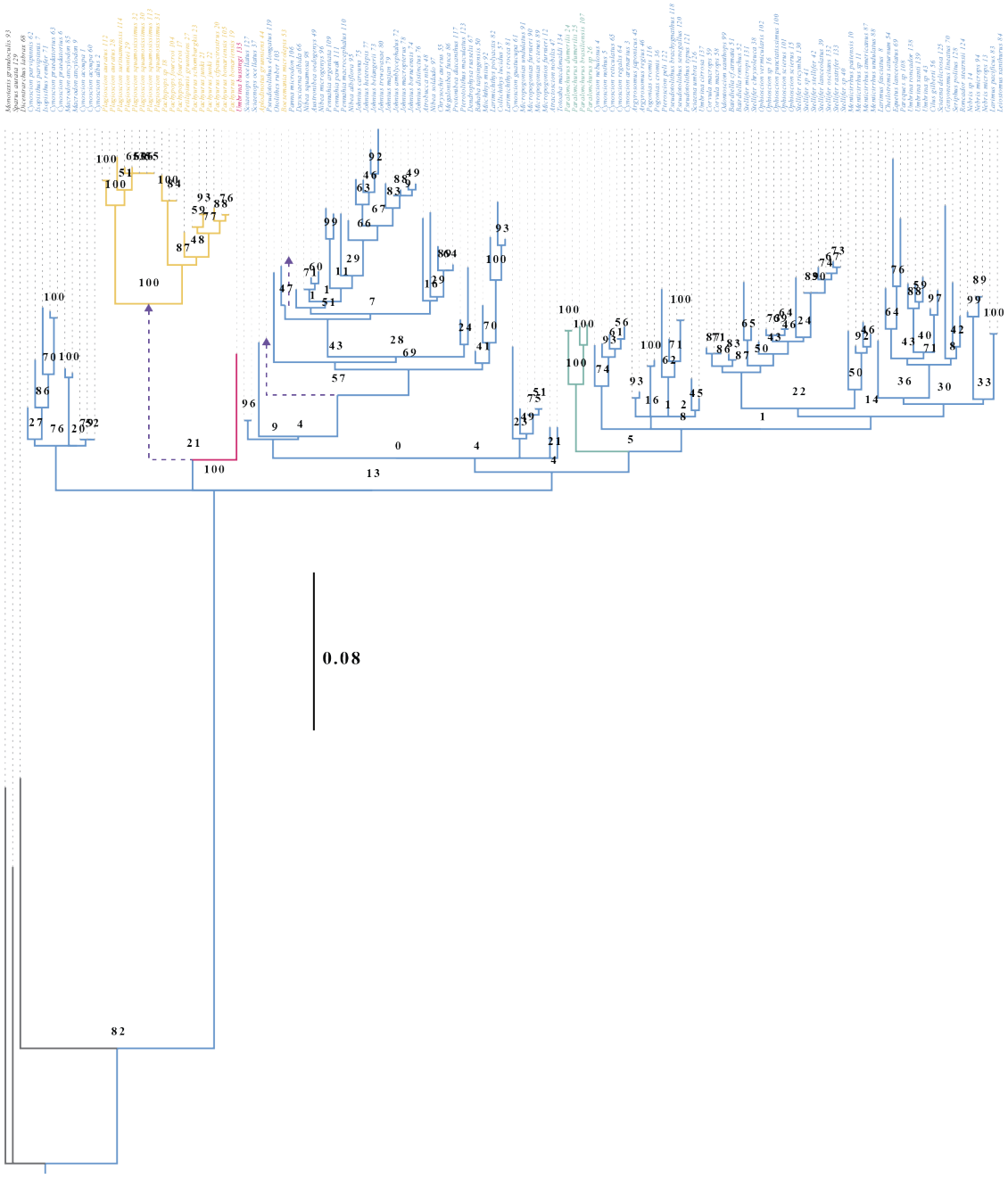

Fig. S3. Maximum likelihood rhodopsin gene tree. Bootstrap support values from 1000 replicate analyses displayed at nodes. Marine lineages in blue, freshwater lineages in yellow, transitional

branches shown as purple arrows, and outgroup taxa shown in grey. Green and pink branches show the two possible sister groups to the South American freshwater clade. Branches proportional to number of substitutions per site.

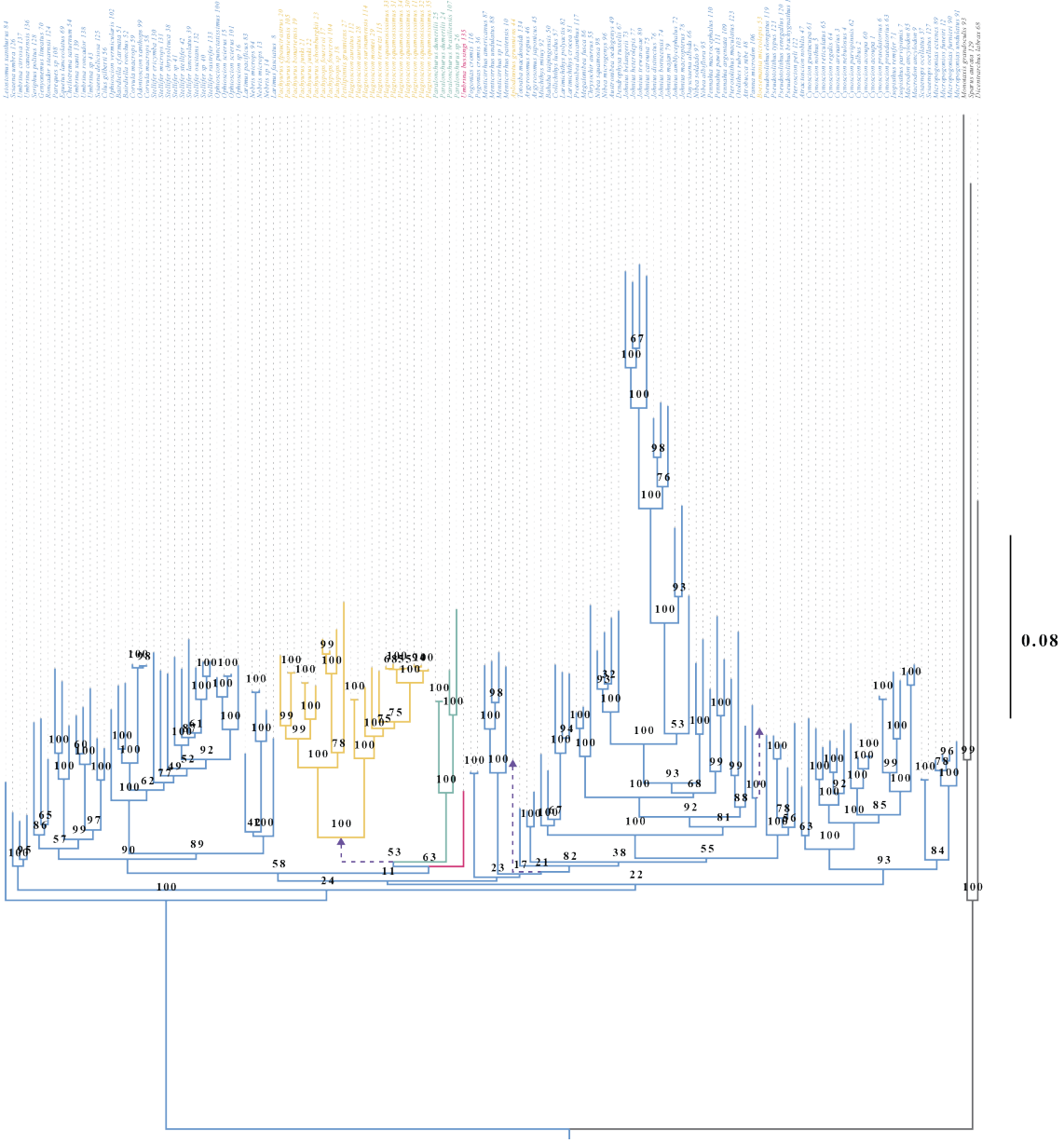

Fig. S4. Maximum likelihood species tree reconstructed using a concatenated alignment of four nuclear loci and two mitochondrial loci but with site 165 and 214 removed from rhodopsin dataset. Bootstrap support values from 1000 replicate analyses displayed at nodes. Marine lineages in blue, freshwater lineages in yellow, transitional branches shown as purple arrows, and

outgroup taxa shown in grey. Green and pink branches show the two possible sister groups to the South American freshwater clade. Branches proportional to number of substitutions per site.

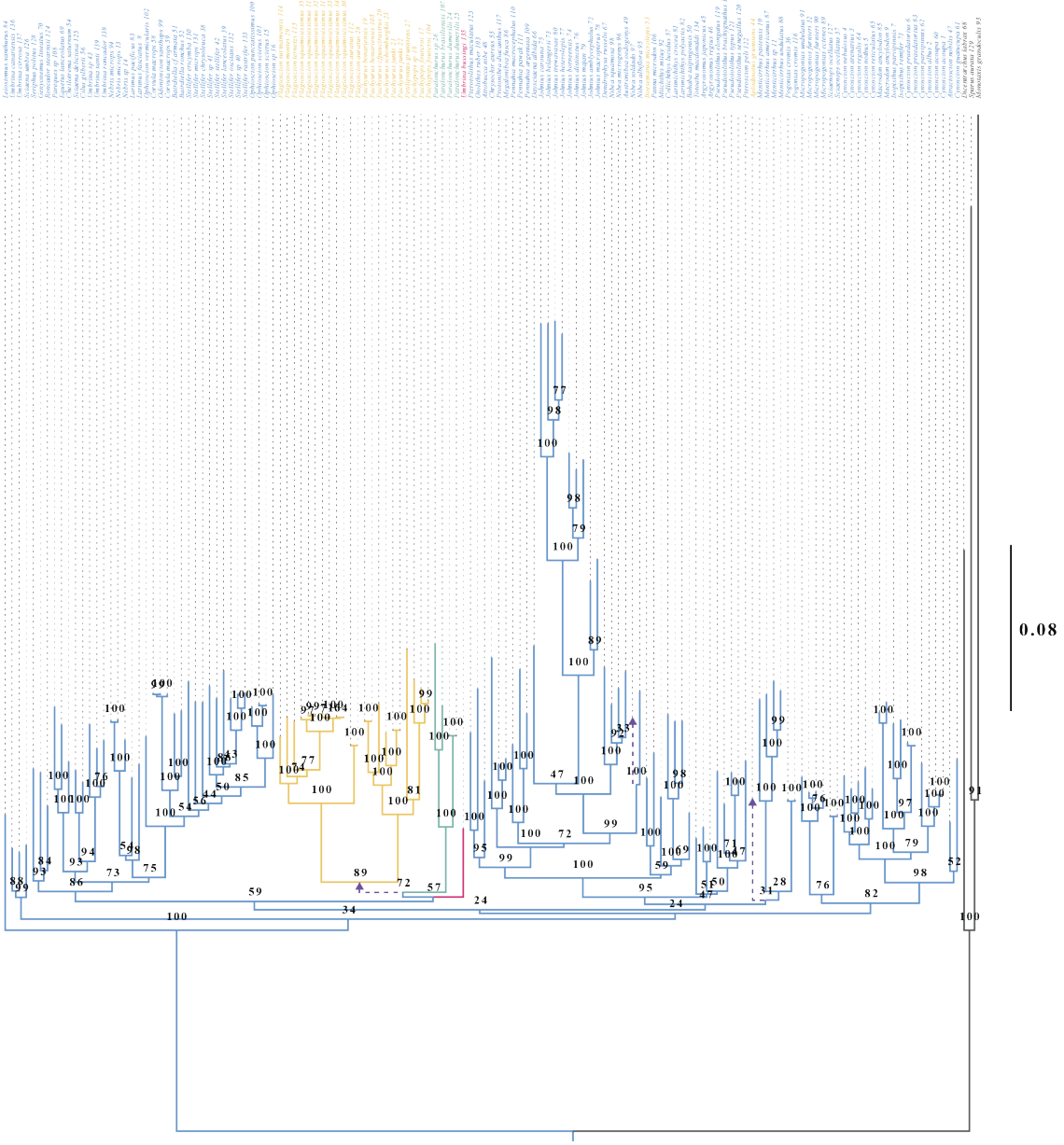

Fig. S5. Maximum likelihood species tree reconstructed using a concatenated alignment of four nuclear loci and two mitochondrial loci but with the first and second codon position removed from rhodopsin. Bootstrap support values from 1000 replicate analyses displayed at nodes. Marine lineages in blue, freshwater lineages in yellow, transitional branches shown as purple arrows, and outgroup taxa shown in grey. Green and pink branches show the two possible sister

groups to the South American freshwater clade. Branches proportional to number of substitutions per site.

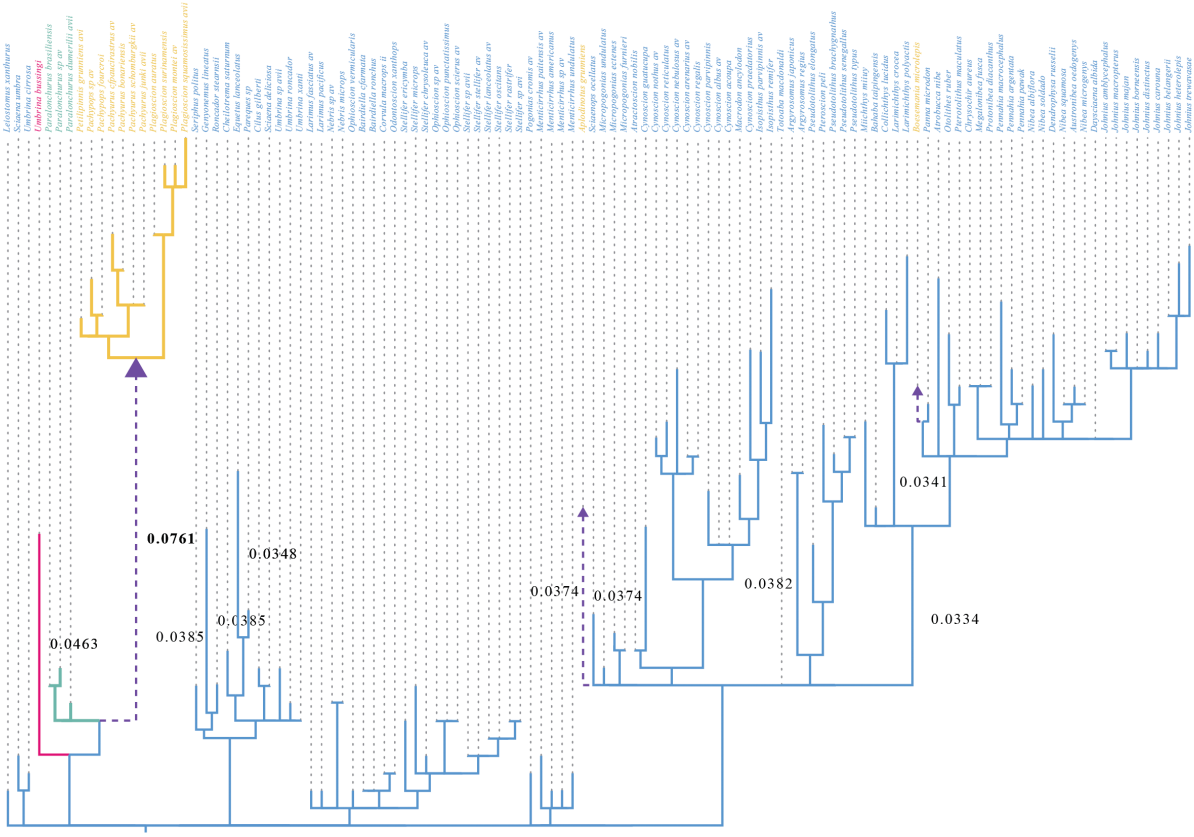

Fig. S6. Bayesian phylogeny with branch lengths scaled by number of amino acid substitutions. Branch lengths estimated in PAML using the WAG substitution matrix. Branch lengths above 0.03 shown on tree. Marine lineages in blue, freshwater lineages in yellow, transitional branches shown as purple arrows. Green and pink branches show the two possible sister groups to the South American freshwater clade.

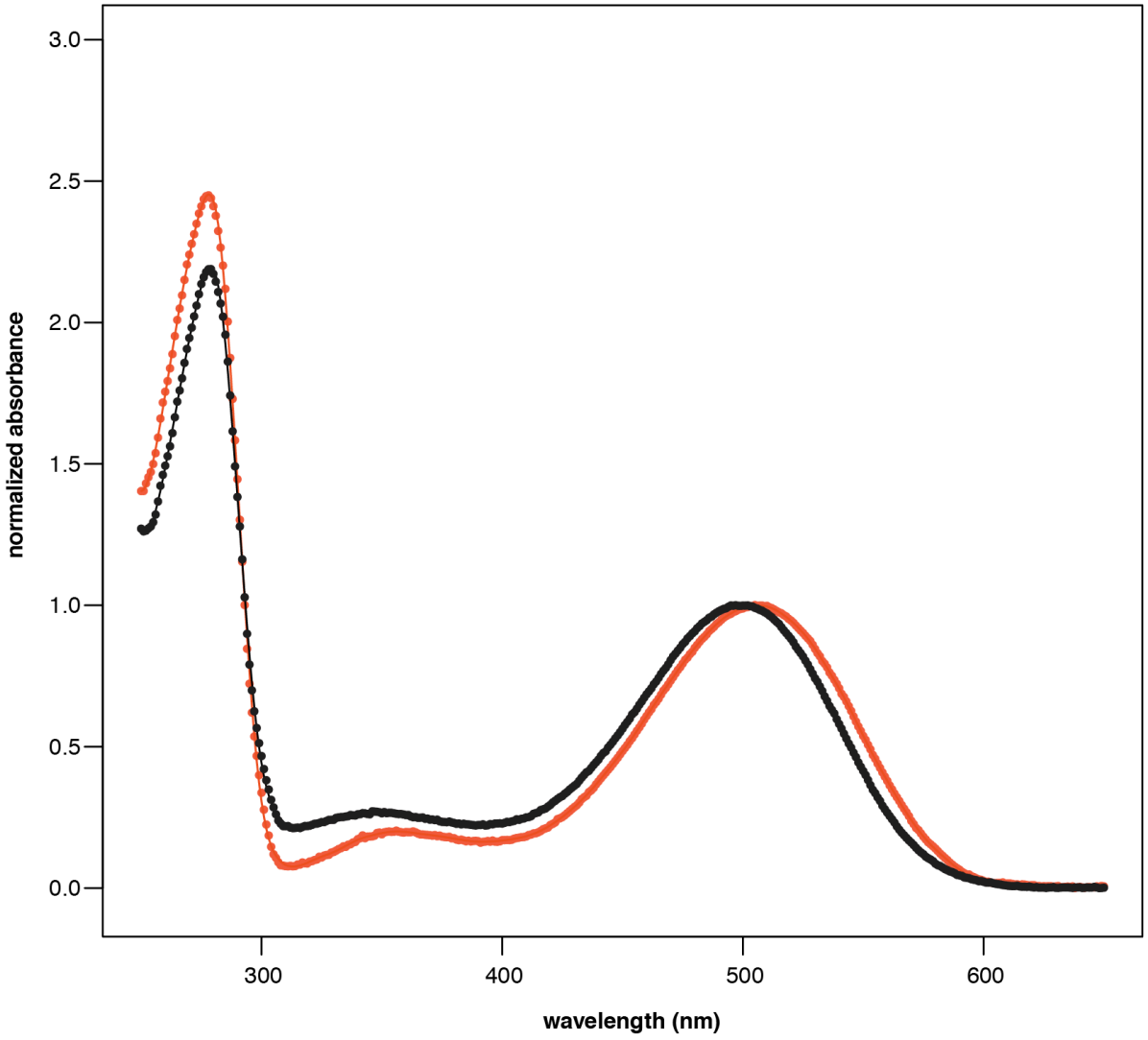

Fig. S7. Spectral absorbance curve of dark-state wild-type bovine rhodopsin (black) and bovine rhodopsin made to have freshwater like croaker rhodopsin residues at site 119, 122, and 165 (red) using site-directed mutagenesis.

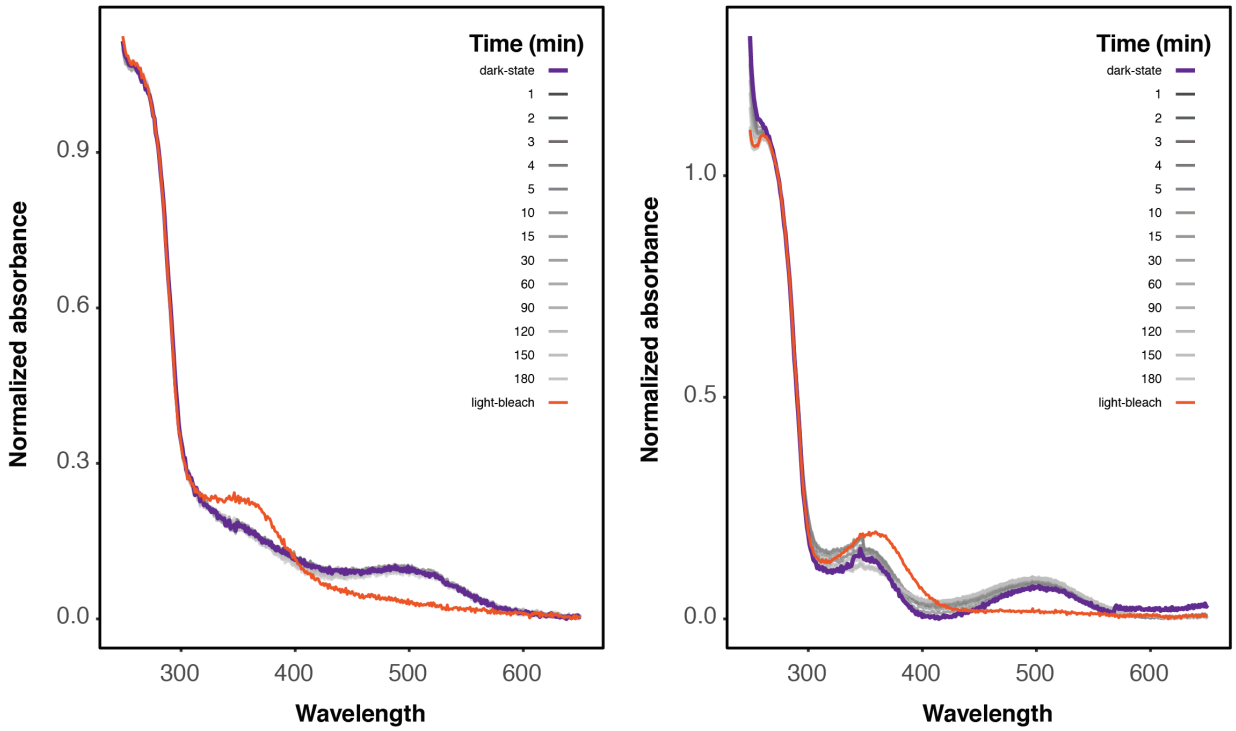

Fig. S8. Hydroxylamine assay of ancestral marine and freshwater rhodopsin pigments. Spectral absorbance curve of freshwater (left) and marine (right) ancestral croaker rhodopsin pigments. Measurements taken before the addition of 50mM hydroxylamine (purple line) and every minute subsequently for 5 minutes, every 5 minutes from 5 to fifteen minutes and every 30 minutes thereafter (shades of grey). Lastly, rhodopsin was bleached for thirty seconds and measured again (orange line).

Supplementary Text

Sequencing Methods

DNA was extracted from 43 tissue samples preserved in ethanol, representing 36 species using a QIAGEN DNeasy kit (Qiagen Inc, Santa Clara CA, USA). PCR primers and protocols have been previously published in (Lo et al. 2015; Van Nynatten et al. 2015) (**supplementary table S14**). For cytochrome b (ctyb), each PCR reaction contained 2 µL of template DNA, 2.5 µL of 10X PCR buffer, 1.5 µL of 25 mM $MgCl_2$, 2.0 µL of 10 mM dNTPs, 1.0 µL of forward and reverse primers, 0.5 µL of 5 U/µL Taq DNA polymerase, and ddH2O for a final volume of 25 µL. For cytochrome c oxidase subunit I (coi), each PCR reaction contained 2 µL of template DNA, 2.5 µL of 10X PCR buffer, 1.5 µL of 25 mM $MgCl_2$, 0.5 µL of 10 mM dNTPs, 0.5 µL of forward and reverse primers, 0.2 µL of 5 U/µL Taq DNA polymerase, and ddH2O for a final volume of 25 µL. For the nuclear genes, rhodopsin (rho), recombination activating factor 1 (rag1), and early growth response protein 1 and 2 (egr-1 and egr-2), each PCR reaction contained 3 µL of template DNA, 3.0 µL of 10X PCR buffer, 2.0 µL of 25 mM $MgCl_2$, 2.4 µL of X10 mM dNTPs, 1.5 µL of forward and reverse primers, 0.3 µL of 5 U/µL Taq DNA polymerase, and ddH2O for a final volume of 25 µL. Rho, rag1, and egr-1 were amplified using an initial denaturation phase at 95°C for 4 minutes, followed by 35 cycles of 95°C (40s), 55°C (40s), and 72°C (90s), followed by a 7-minute extension phase at 72°C. Egr-2 was amplified using an initial denaturation phase at 95°C for 4 minutes, followed by 35 cycles of 95°C (40s), 52°C (40s), and 72°C (90s), followed by a 7-minute extension phase at 72°C. Cytb was amplified using an initial denaturation phase at 95°C for 2 minutes, followed by 35 cycles of 95°C (30s), 49°C (60s), and 72°C (90s), followed by a 5-minute extension phase at 72°C. Coi was amplified using an initial denaturation phase at 95°C for 2 minutes, followed by 35 cycles of 94°C (30s), 52°C (40s), and 72°C (90s), followed by a 5-minute extension phase at 72°C. PCR products were visualized on a 1.5% agarose gel and were compared to an 100 bp ladder to check that amplicons were the correct size. PCR products were purified using ExoSAP-I PCR Product Cleanup Reagent (Applied Biosystems, Foster City, CA, USA) and sequenced by Sanger sequencing at the Hospital for Sick Children TCAG Sequencing Facility. Chromatograms were compared for forward and reverse reads in Geneious (Kearse et al. 2012). These sequences (form 43 individuals) were combined with additional sequences for the same six molecular markers generated in a previous phylogenetic reconstruction of croakers (Lo et al. 2015) to form a dataset

of 139 taxa (Genbank accession number in **supplementary table S15**, Supplementary Material online). Sequences for each gene in the final dataset were aligned using MUSCLE independently (Edgar 2004). Terminal gaps were removed when present in more than half of the samples. We concatenated the alignments of all six molecular markers for a total dataset length of 6460bp.

Phylogenetic Reconstructions

Model searches in Partition Finder (Lanfear et al. 2016) were constrained to models available in RAXML and Mr. Bayes, and the best fitting partitioning scheme was chosen using a Bayes Information Criterion (BIC) approach (**supplementary table S16**). Tracer was used to visualize log-likelihood values for each parameter estimate across successive samples in the Bayesian phylogenetic reconstruction to determine if stationarity was reached and to ensure that each parameter estimate had an effective sample size of greater than 200.

Molecular Evolutionary Analyses

We pruned our dataset to include only species with rhodopsin sequence data and removed all terminal gaps, outgroups and duplicate taxa. This resulted in an 825 bp long dataset, spanning all seven transmembrane helices of rhodopsin with sequences from 114 taxa. Branch-sites models in PAML include a parameter allowing positive selection in a subset of sites on the foreground, but constrains all other branches to be less than or equal to one. Explicit support for positive selection on the foreground lineage can be established with likelihood ratio tests (LRTs) with a nested null branch-site model where d_N/d_S estimates for the positively selected site class are constrained to equal one on foreground lineages (Yang and Nielsen 2002).

We also employed clade model C and D (CmC, CmD) in PAML (Bielawski and Yang 2004). These models allow multiple independent foreground lineages and do not constrain background lineages to be less than or equal to one, allowing positive selection in the background. Support for the divergent site class parameter included in CmC can be established with LRTs to the nested null model M2aREL, which assumes uniform d_N/d_S estimates for each site class parameter across the entire tree. Support for positive selection in the foreground divergent site class estimate can also be tested explicitly using a modified CmC null model where the foreground is constrained to equal one. CmD differs from CmC in that the second site class, with a uniform

dN/dS estimate across the phylogeny, is freely estimated and more flexible than the comparable site class in CmC that is constrained to equal one. This allows for more than one class of positively selected sites in CmD. The nested null model for CmD is M3, which has three freely estimated site classes but assumes uniform selection across the tree.

CmC and CmD can be used to determine the number of foreground partitions that best fit the data by comparing nested partitioning schemes (Schott et al. 2014). We partitioned our Croaker phylogeny so that freshwater invading lineages were the foreground. We tested this scheme using branch-sites and clade model analyses. We also isolated each individual freshwater invasion event as individual foreground lineages using both aforementioned models in PAML. We compared the clade model results for each freshwater invasion event with a more parameter rich partitioning scheme with each freshwater invasion event as its own isolated foreground partition with independent dN/dS estimate.

Protein expression and functional characterization

Ancestral sequences at the nodes bracketing the marine to freshwater transition in South America were synthesized using GeneArt (Invitrogen). To improve expression levels in HEK cells we appended the N and C terminal segments of the gene with human sequence and converted codon identities to match those of human rhodopsin where possible. The first two and last codon in the sequence were converted to match 5' and 3' restriction sites for insertion into the P1D3-hrGFP II expression vector (Morrow and Chang 2010). Because expression levels of these ancestral pigments were not sufficient for kinetic assays we used site-directed mutagenesis primers to induce single amino acid substitutions to convert the amino acid identities in bovine rhodopsin at sites 119, 122, 124 and 261 to match the marine and freshwater croaker sequences *via* PCR (QuickChange II, Agilent). All sequences were verified using a 3730 DNA Analyzer (Applied Biosystems) at the Centre for Analysis of Genome Evolution and Function (CAGEF) at the University of Toronto. All croaker and bovine rhodopsin sequences were transferred to the pIRES-hrGFP II expression vector (Stratagene) for subsequent transient transfection of HEK293T cells (8 μg per 10 cm plate) using Lipofectamine 2000 (Invitrogen). 30 plates were used to express each ancestral rhodopsin pigments. 14 plates were used to express the bovine rhodopsin pigments (wild-type and site-directed freshwater mutants). Media was changed after

24 hours, and cells were harvested 48 hours post-transfection. Cells were washed twice with harvesting buffer (PBS, 10 μg/mL aprotinin, 10 μg/mL leupeptin), and rhodopsins were regenerated for 2 hours in the dark with 5 μM 11-*cis*-retinal generously provided by Dr. Rosalie Crouch (Medical University of South Carolina). After regeneration the samples were incubated at 4ºC in solubilization buffer (50 mM Tris pH 6.8, 100 mM NaCl, 1 mM CaCl2, 1% dodecylmaltoside, 0.1 mM PMSF) for 2 hours and immunoaffinity purified overnight using the 1D4 monoclonal antibody coupled to the UltraLink Hydrazide Resin (ThermoFisher Scientific). Resin was washed three times with wash buffer 1 (50 mM Tris pH 7.0, 100 mM NaCl, 0.1% dodecylmaltoside) and twice using wash buffer 2 (50 mM sodium phosphate, 0.1% dodecylmaltoside; pH 7.0). Rhodopsins were eluted from the UltraLink resin using 5 mg/mL of a 1D4 peptide, consisting of the last 9 amino acids of bovine rhodopsin (TETSQVAPA).

The UV-visible absorption spectra of purified rhodopsin samples were recorded in the dark at 20 °C using a Cary 4000 double-beam absorbance spectrophotometer (Agilent). All peak spectral sensitivities were determined by fitting dark spectra to a standard template curve for A1 visual pigments (Govardovskii et al. 2000). Rhodopsin samples were light-activated for 30 seconds using a fiber optic lamp (Dolan-Jenner), resulting in a shift in peak spectral sensitivity to ~ 380 nm, characteristic of the biologically active metarhodopsin II intermediate (Van Eps et al. 2017). Pigments were also exposed to hydroxylamine (NH_2OH; 50mM) to test accessibility of the rhodopsin binding pocket before and after light activation, as previously described (Sakmar et al. 1989). Retinal release following rhodopsin photoactivation was monitored using a Cary Eclipse fluorescence spectrophotometer equipped with a Xenon flash lamp (Agilent), according to a protocol modified from previous studies (Farrens and Khorana 1995; Schafer et al. 2016). Rhodopsin samples (0.1-0.2 μM) were bleached for 30 seconds at 20ºC with a fiber optic lamp (Dolan-Jenner) using a filter to restrict wavelengths of light below 475 nm to minimize heat. Fluorescence measurements were recorded at 30-second intervals with a 2 second integration time, using an excitation wavelength of 295 nm (1.5 nm slit width) and an emission wavelength of 330 nm (10 nm slit width). There was no noticeable activation by the excitation beam prior to rhodopsin activation. This assay detected increasing fluorescence as a result of decreased quenching of intrinsic tryptophan fluorescence at W265 by the retinal chromophore (Farrens and Khorana 1995), and is a reliable proxy for the tracking the decay of MII (Schafer et al. 2016). Data was fit to a three variable, first order exponential equation ($y = y_0 + a(1\text{-}e^{-bx})$), and half-life values were

calculated using the rate constant b ($t_{1/2} = \ln 2/b$). All curve fitting resulted in r^2 values greater than 0.95.

Comparing CmC with CmD when two classes of positively selected sites present

CmD can also be directly compared with CmC, which differ only in how they estimate the second site class, to establish support for a second site class parameter different than one. M3 can also be set to have more than three site classes, and can be used to determine the number of site classes required to fit the data effectively by comparing models with more site classes with models with fewer (Bielawski and Yang 2004). We used this approach to test if the positively selected sites in the marine dataset (pervasive positively selected sites) were interfering with our estimation of episodic selection using CmC, which was not significantly better fitting than M2aREL (its requisite null model) and significantly worse fitting than CmD (**supplementary table S11**). Setting the number of site classes in M3 to four was significantly better fitting than three site classes (**supplementary table S11**). The fourth site class was occupied by two sites, 165 and 214, and had a very high d_N/d_S estimate (13.64, **supplementary table S11**). We suspected that these highly positively selected sites were shifting the estimate for the third site class in CmC, the only site class in this model allowed to be above one, beyond what was appropriate for modelling the divergent selection detected on the transitional branch using Branch-sites, CmD, Branch-sites REL and even the Two-Ratio model. Removing these two sites and re-running the analyses returned parameter estimates consistent with those reported for the other branch and clade models (**supplementary table S11**) This suggests that in models with more than one class of positively selected sites, CmD should be used, as it has the flexibility to model more than one class of positively selected sites.

Positively selected sites and phylogenetic incongruence

In the Bayesian phylogeny, the sister group to the freshwater clade is the Paralonchurus clade, previously identified as the sister group to the freshwater Claude in (Lo et al. 2015). However, this previous study contained only one Paralonchurus species. In the maximum likelihood species tree and maximum likelihood rhodopsin gene tree *Umbrina bussingi* was reconstructed as the sister species, which disagrees with this previous and Bayesian topology.

When the two highly positively selected sites 165 and 214 were removed from the rhodopsin partition of our concatenated alignment this topological incongruence in Bayesian and Maximum likelihood species trees disappeared and the maximum likelihood species tree was reconstructed with Paralonchurus as the sister group (**supplementary figure S4**). The same is true when the first and second codon position are removed from the rhodopsin partition (**supplementary figure S5**).

Supplementary Reference

Baldwin CC, Mounts JH, Smith DG, Weigt LA. 2009. Genetic identification and color descriptions of early life-history stages of Belizean Phaeoptyx and Astrapogon (Teleostei: Apogonidae) with comments on identification of adult Phaeoptyx. *Zootaxa* 2008: 1–22.

Beatty DD. 1973. Visual pigments of several species of teleost fishes. *Vision Res.* 13:989–992

Bielawski JP, Yang Z. 2004. A Maximum Likelihood Method for Detecting Functional Divergence at Individual Codon Sites, with Application to Gene Family Evolution. *J. Mol. Evol.* 59:121–132.

Castiglione GM, Chang BSW. 2018. Functional trade-offs and environmental variation shaped ancient trajectories in the evolution of dim-light vision. *eLife* **7,** 348

Chen WJ, Bonillo C, Lecointre G. 2003. Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Mol. Phylogenetics Evol.* 26:262–288.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Farrens DL, Khorana HG. 1995. Structure and Function in Rhodopsin. J Biol Chem 270:5073–5076.

Govardovskii VI, Fyhrquist N, Reuter T, Kuzmin DG, Donner K. 2000. In search of the visual pigment template. *Vis. Neurosci.* 17:509–528.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2016. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol. Biol. Evol. 34:msw260–msw773.

Lo P-C, Liu S-H, Chao NL, Nunoo FKE, Mok H-K, Chen WJ. 2015. A multi-gene dataset reveals a tropical New World origin and Early Miocene diversification of croakers (Perciformes: Sciaenidae). *Mol. Phylogenetics Evol.* 88:132–143.

López JA, Chen WJ, Orti G. 2004. Esociform Phylogeny.Wood RM, editor. *Copeia* 2004:449–464.

Morrow JM, Chang BSW. 2010. The p1D4-hrGFP II expression vector: A tool for expressing and purifying visual pigments and other G protein-coupled receptors. *Plasmid* 64:162–169.

Sakmar TP, Franke RR, Khorana HG. 1989. Glutamic acid-113 serves as the retinylidene Schiff base counterion in bovine rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* 86:8309–8313.

Schafer CT, Fay JF, Janz JM, Farrens DL. 2016. Decay of an active GPCR: Conformational dynamics govern agonist rebinding and persistence of an active, yet empty, receptor state. *Proc. Natl. Acad. Sci. U.S.A.* 113:11961–11966.

Schott RK, Refvik SP, Hauser FE, López-Fernández H, Chang BSW. 2014. Divergent Positive Selection in Rhodopsin from Lake and Riverine Cichlid Fishes. *Mol. Biol. Evol.* 31:1149–1165.

Van Eps N, Caro LN, Morizumi T, Kusnetzow AK, Szczepek M, Hofmann KP, Bayburt TH, Sligar SG, Ernst OP, Hubbell WL. 2017. Conformational equilibria of light-activated rhodopsin in nanodiscs. *Proc. Natl. Acad. Sci. U.S.A.* 114:E3268–E3275.

Van Nynatten A, Bloom DD, Chang BSW, Lovejoy NR. 2015. Out of the blue: adaptive visual pigment evolution accompanies Amazon invasion. *Biol. Lett.* 11:20150349.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. 2005. DNA barcoding Australia's fish species. *Phil. Trans. R. Soc. Lond. B* 360:1847–1857.

Yang Z, Nielsen R. 2002. Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Mol. Biol. Evol.* 19:908–917.