PNAS
www.pnas.org

Supplementary Information for

Ancient whale rhodopsin reconstructs dim-light vision over a major evolutionary transition: Implications for ancestral diving behaviour

Sarah Z. Dungan and Belinda S-W Chang

Corresponding author: Belinda S-W Chang
Email: belinda.chang@utoronto.ca

This PDF file includes:

Supplementary text
Figs. S1 to S6
Tables S1 to S4
References for SI citations

Robustness of ancestral sequence reconstructions

Codon models and posterior probability. We reconstructed the ancestral sequences using the codon-based likelihood models available in the *codeml* program of PAML4.9(1) (Supplementary Table S1, Supplementary Table S2), and a tree topology comprised of representative cetartiodactyl rhodopsin coding sequences (Supplementary Table S3). Though there were no disagreements as to the most probable sites in either ancestor across the codon models, we generally recovered higher site-by-site posterior probabilities for the best-fitting model (Clade Model D with diving class partitions(2)). All the codon models revealed at least 12 sites transitioning between Whippomorpha and Cetacea (Supplementary Fig. S2). This finding was unsurprising given cetacean rhodopsin is known to be affected by d_N/d_S-related heterogeneity(2).

For the ancestral Cetacea rhodopsin, we reconstructed the translated amino acid sequence with marginal posterior probabilities > 0.80 for all sites under the best-fitting model and 96.5% of sites were certain (posterior probability = 1.0). These results were consistent across all the codon models we tested (Supplementary Fig. S2), indicating a highly robust reconstruction. The ancestral Whippomorpha sequence reconstruction was slightly less certain. Under Clade Model D, marginal posterior probabilities were > 0.80 except for one site, V300 (0.583), and only 23.3% of sites were certain. The other codon models we tested showed similar levels of uncertainty for either V300 or I300 (posterior probabilities of 0.50 – 0.75). To determine whether the uncertainty at this site would have an effect on our experimental results, we mutated the residue in the synthesized Whippomorpha coding sequence and functionally compared the two variants. Despite the uncertainty, the identity of this site as I or V had no significant effect on λ_{max} or retinal release $t_{1/2}$ (Supplementary Fig. S3).

Posterior distribution sampling. Even with a well-fit model, the most probable ancestral sequences using optimality-based models are known to be biased toward more frequent amino acid states in the dataset(3, 4). For example, 10 sequences randomly sampled from the posterior distribution in the reconstruction of the ancestral archosaur rhodopsin sequence showed variation when compared with the most probable sequence(3, 5). On the other hand, sampling ancestral sequences from the posterior distribution can be used to assess potential bias in ancestral protein function(6, 7). For example, in Bickelmann and colleagues(7) reconstruction of the ancestral mammal rhodopsin, a single randomly sampled sequence from the posterior distribution differed from the most probable sequence at 7 sites, yet expression experiments revealed it did not vary significantly in function from the most probable sequence. For our dataset, we inferred ancestral sequences from weighted random samplings of our best-fitting Clade Model D posterior probability distribution(7, 8). Of 10,000 random samplings, at least 50% matched the most probable sequence for both Cetacea and Whippomorpha, a result that contrasts the archosaur and ancestral mammal rhodopsin case studies. This difference is probably a reflection of the generally high certainty of our reconstructed sequences; in the ancestral mammal sequence, for example, 8 sites were reconstructed with < 0.8 probability(7), whereas only one site in our Whippomorpha sequence (and none in the Cetacea sequence) fell below this standard.

Nucleotide and amino acid models. Though an increasing number of protein evolution studies are making use of ancestral sequence reconstruction (and less frequently ancestral protein resurrection), few provide thorough comparisons across multiple methods, and most preferentially rely on amino acid models(9, 10). The codon models in PAML use a marginal reconstruction process, which assigns the combination of nucleotide states to each node sequence on the tree that maximizes the likelihood of the node sequence by working upward from the terminal sequences(11). This approach is considered more suitable when the goal of the study is to reconstruct specific ancestor sequences in their entirety. Alternatively, joint reconstruction methods assign ancestral character states so as to maximize the global (joint) likelihood of the tree/dataset(12), and are more suitable for mapping the evolution of sites across the whole tree. Joint reconstruction methods are computationally more complex, and so are not yet available for evolutionary models that incorporate rate heterogeneity (*e.g.* gamma-distributed in nucleotide models, variable d_N/d_S in codon models)(1).

Nevertheless, to observe the consistency of our results even when using less suitable reconstruction methods, we ran our dataset through the ASR program implemented on the Datamonkey web server(13), which includes methods for joint reconstruction(12, 14), and marginal reconstruction using nucleotide models(11). We also used two amino acid-based models: marginal reconstruction using *aaml* in PAML (with the JTT and WAG amino acid matrices, applied model frequencies +F, and gamma-distributed among-site rate heterogeneity +G), and the newly available ProtASR(15), which uses a mean-field substitution model and associated PDB file (dark-state bovine rhodopsin in our case, PDB: 1U19(16)) to better account for protein structural constraints. While these methods produced results that were generally consistent with the codon models, the amino acid models disagreed at one site each in Cetacea (195) and Whippomorpha (270), and supported I300 in Whippomorpha (Supplementary Fig. S2). These results implicate the transitioning substitution K195S, but the absence of V300I and G270S.

Nevertheless, the codon models calculated very low posterior probabilities for the nucleotide substitutions underlying these alternative amino acid states (Supplementary Figures S4 – S6). We thus recommend cross-checking results with codon models, even when amino acid methods return sequences with high site-by-site posterior probabilities.

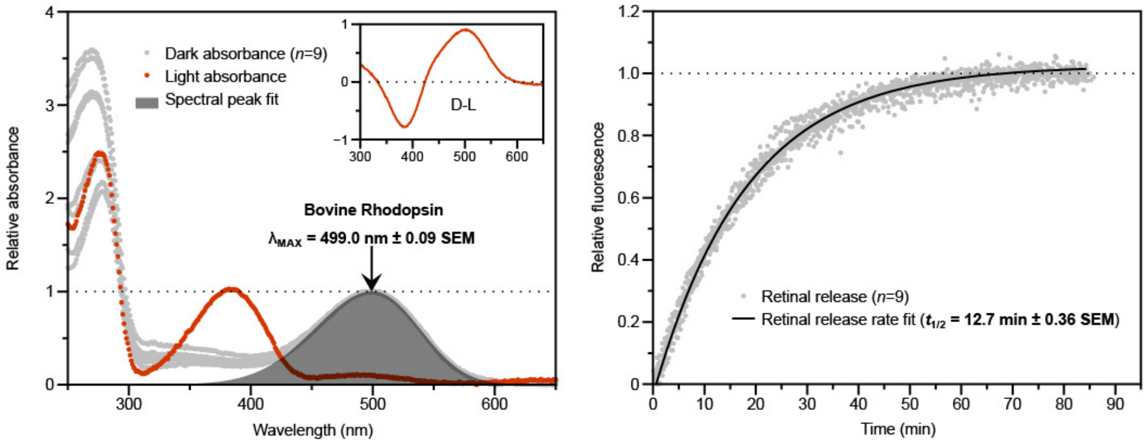

Fig. S1. Functional characteristics of bovine rhodopsin (positive control pigment). The left panel shows dark and light-activated absorption spectra, and the right panel shows light-activation fluorescence time series. The indicated λ_{max} value is the mean (± standard error) of estimates calculated for separately eluted samples (where n is the number of elutions per pigment). The light-activated spectral peak is 380 nm, which is characteristic of the light-activated intermediate, and the inset shows the dark-light difference spectrum. The indicated $t_{1/2}$ for retinal release is the mean (± standard error) of estimates calculated for separate fluorescence time series (where n is the number of time series).

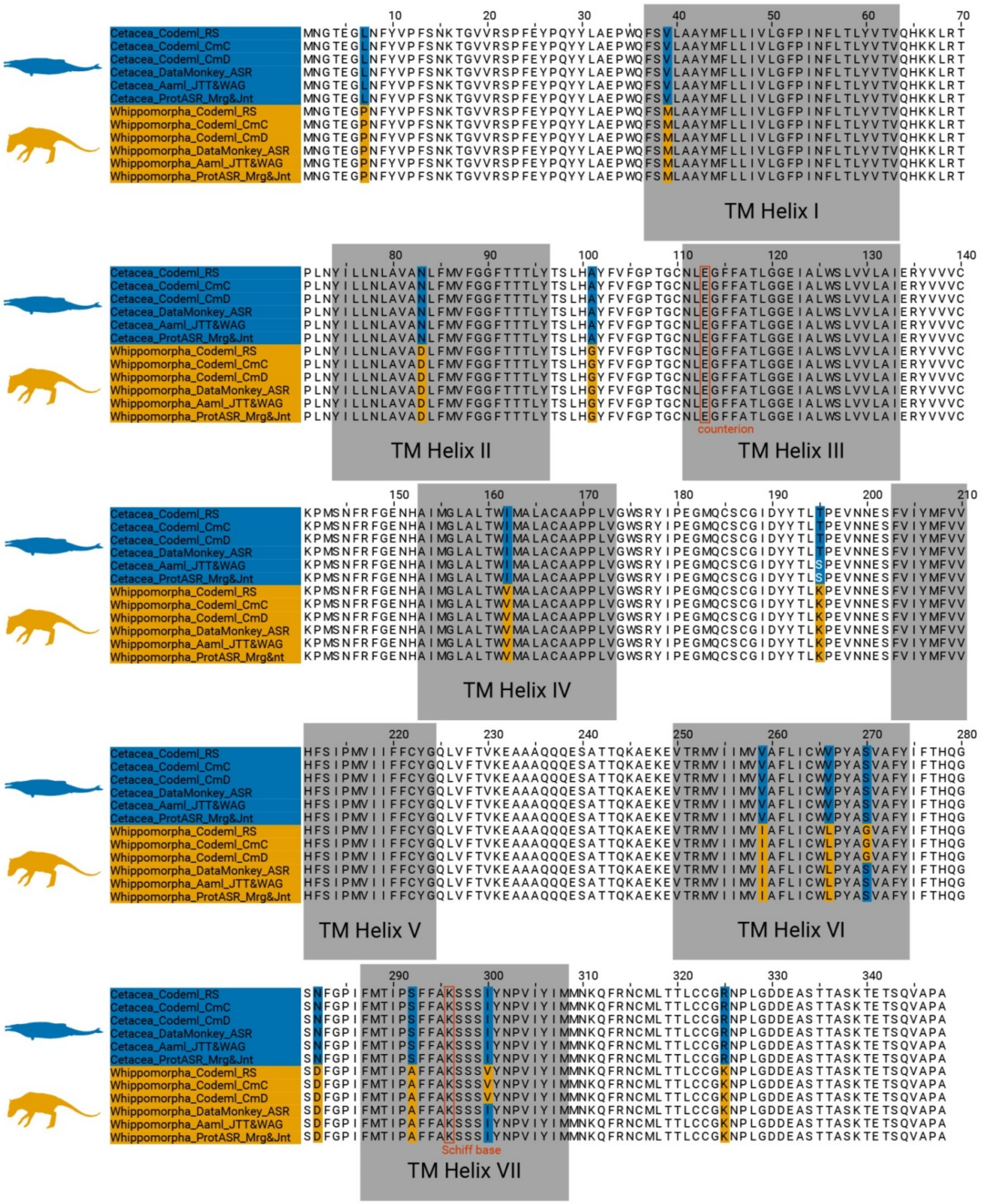
10 20 30 40 50 60 70
Cetacea_Codeml_RS MNGTEGLNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSVLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Cetacea_Codeml_CmC MNGTEGLNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSVLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Cetacea_Codeml_CmD MNGTEGLNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSVLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Cetacea_DataMonkey_ASR MNGTEGLNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSVLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Cetacea_Aaml_JTT&WAG MNGTEGLNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSVLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Cetacea_ProtASR_Mrg&Jnt MNGTEGLNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSVLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Whippomorpha_Codeml_RS MNGTEGPNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Whippomorpha_Codeml_CmC MNGTEGPNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Whippomorpha_Codeml_CmD MNGTEGPNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Whippomorpha_DataMonkey_ASR MNGTEGPNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Whippomorpha_Aaml_JTT&WAG MNGTEGPNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
Whippomorpha_ProtASR_Mrg&Jnt MNGTEGPNFYVPFSNKTGVVRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTLYVTVQHKKLRT
TM Helix I
80 90 100 110 120 130 140
Cetacea_Codeml_RS PLNYILLNLAVANLFMVFGGFTTTLYTSLHAYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Cetacea_Codeml_CmC PLNYILLNLAVANLFMVFGGFTTTLYTSLHAYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Cetacea_Codeml_CmD PLNYILLNLAVANLFMVFGGFTTTLYTSLHAYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Cetacea_DataMonkey_ASR PLNYILLNLAVANLFMVFGGFTTTLYTSLHAYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Cetacea_Aaml_JTT&WAG PLNYILLNLAVANLFMVFGGFTTTLYTSLHAYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Cetacea_ProtASR_Mrg&Jnt PLNYILLNLAVANLFMVFGGFTTTLYTSLHAYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Whippomorpha_Codeml_RS PLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Whippomorpha_Codeml_CmC PLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Whippomorpha_Codeml_CmD PLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Whippomorpha_DataMonkey_ASR PLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Whippomorpha_Aaml_JTT&WAG PLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
Whippomorpha_ProtASR_Mrg&Jnt PLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVC
counterion
TM Helix II
TM Helix III
150 160 170 180 190 200 210
Cetacea_Codeml_RS KPMSNFRFGENHAIMGLALTWIMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLTPEVNNESFVIYMFVV
Cetacea_Codeml_CmC KPMSNFRFGENHAIMGLALTWIMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLTPEVNNESFVIYMFVV
Cetacea_Codeml_CmD KPMSNFRFGENHAIMGLALTWIMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLTPEVNNESFVIYMFVV
Cetacea_DataMonkey_ASR KPMSNFRFGENHAIMGLALTWIMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLTPEVNNESFVIYMFVV
Cetacea_Aaml_JTT&WAG KPMSNFRFGENHAIMGLALTWIMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLSPEVNNESFVIYMFVV
Cetacea_ProtASR_Mrg&Jnt KPMSNFRFGENHAIMGLALTWIMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLSPEVNNESFVIYMFVV
Whippomorpha_Codeml_RS KPMSNFRFGENHAIMGLALTWVMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVV
Whippomorpha_Codeml_CmC KPMSNFRFGENHAIMGLALTWVMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVV
Whippomorpha_Codeml_CmD KPMSNFRFGENHAIMGLALTWVMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVV
Whippomorpha_DataMonkey_ASR KPMSNFRFGENHAIMGLALTWVMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVV
Whippomorpha_Aaml_JTT&WAG KPMSNFRFGENHAIMGLALTWVMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVV
Whippomorpha_ProtASR_Mrg&nt KPMSNFRFGENHAIMGLALTWVMALACAAPPLVGWSRYIPEGMQCSCGIDYYTLKPEVNNESFVIYMFVV
TM Helix IV
220 230 240 250 260 270 280
Cetacea_Codeml_RS HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVVAFLICWVPYASVAFYIFTHQG
Cetacea_Codeml_CmC HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVVAFLICWVPYASVAFYIFTHQG
Cetacea_Codeml_CmD HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVVAFLICWVPYASVAFYIFTHQG
Cetacea_DataMonkey_ASR HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVVAFLICWVPYASVAFYIFTHQG
Cetacea_Aaml_JTT&WAG HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVVAFLICWVPYASVAFYIFTHQG
Cetacea_ProtASR_Mrg&Jnt HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVVAFLICWVPYASVAFYIFTHQG
Whippomorpha_Codeml_RS HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQG
Whippomorpha_Codeml_CmC HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQG
Whippomorpha_Codeml_CmD HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQG
Whippomorpha_DataMonkey_ASR HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYASVAFYIFTHQG
Whippomorpha_Aaml_JTT&WAG HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYASVAFYIFTHQG
Whippomorpha_ProtASR_Mrg&Jnt HFSIPMVIIFFCYGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYASVAFYIFTHQG
TM Helix V
TM Helix VI
290 300 310 320 330 340
Cetacea_Codeml_RS SNFGPIFMTIPSFFAKSSSIYNPVIYIMMNKQFRNCMLTTLCCGRNPLGDDEASTTASKTETSQVAPA
Cetacea_Codeml_CmC SNFGPIFMTIPSFFAKSSSIYNPVIYIMMNKQFRNCMLTTLCCGRNPLGDDEASTTASKTETSQVAPA
Cetacea_Codeml_CmD SNFGPIFMTIPSFFAKSSSIYNPVIYIMMNKQFRNCMLTTLCCGRNPLGDDEASTTASKTETSQVAPA
Cetacea_DataMonkey_ASR SNFGPIFMTIPSFFAKSSSIYNPVIYIMMNKQFRNCMLTTLCCGRNPLGDDEASTTASKTETSQVAPA
Cetacea_Aaml_JTT&WAG SNFGPIFMTIPSFFAKSSSIYNPVIYIMMNKQFRNCMLTTLCCGRNPLGDDEASTTASKTETSQVAPA
Cetacea_ProtASR_Mrg&Jnt SNFGPIFMTIPSFFAKSSSIYNPVIYIMMNKQFRNCMLTTLCCGRNPLGDDEASTTASKTETSQVAPA
Whippomorpha_Codeml_RS SDFGPIFMTIPAFFAKSSSVYNPVIYIMMNKQFRNCMLTTLCCGKNPLGDDEASTTASKTETSQVAPA
Whippomorpha_Codeml_CmC SDFGPIFMTIPAFFAKSSSVYNPVIYIMMNKQFRNCMLTTLCCGKNPLGDDEASTTASKTETSQVAPA
Whippomorpha_Codeml_CmD SDFGPIFMTIPAFFAKSSSVYNPVIYIMMNKQFRNCMLTTLCCGKNPLGDDEASTTASKTETSQVAPA
Whippomorpha_DataMonkey_ASR SDFGPIFMTIPAFFAKSSSIYNPVIYIMMNKQFRNCMLTTLCCGKNPLGDDEASTTASKTETSQVAPA
Whippomorpha_Aaml_JTT&WAG SDFGPIFMTIPAFFAKSSSIYNPVIYIMMNKQFRNCMLTTLCCGKNPLGDDEASTTASKTETSQVAPA
Whippomorpha_ProtASR_Mrg&Jnt SDFGPIFMTIPAFFAKSSSIYNPVIYIMMNKQFRNCMLTTLCCGKNPLGDDEASTTASKTETSQVAPA
Schiff base
TM Helix VII

Fig. S2. Alignment of reconstructed ancestral rhodopsin amino acid sequences according to codon models (codeml), nucleotide models (DataMonkey), and amino acid models (aaml and ProtASR). Sites that transition along the branch separating the Cetacea and Whippomorpha nodes are highlighted. All the models returned sequences that were highly consistent with each other, but note the inconsistencies between codon and amino acid models at sites 195, 270, and 300.

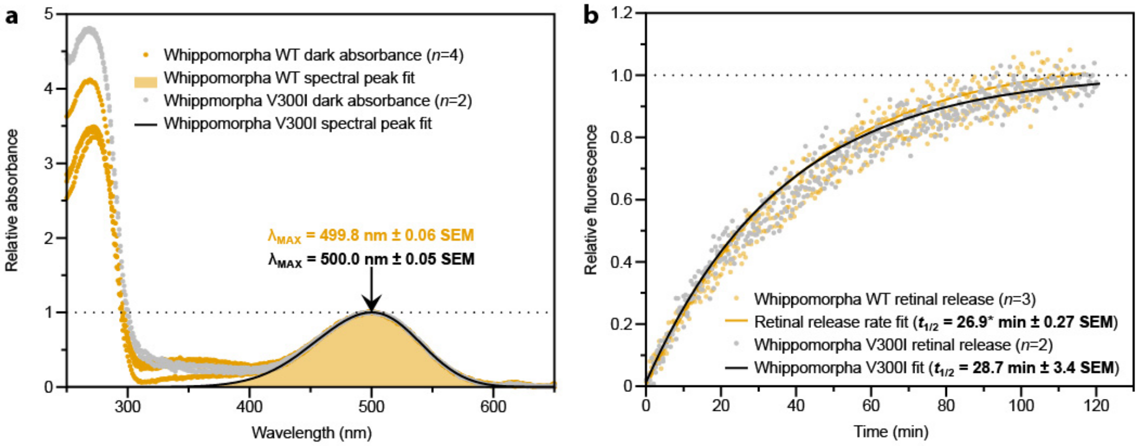

Fig. S3. The effect of uncertain site 300 on Whippomorpha rhodopsin. **a**, spectral tuning. **b**, retinal release. Mutating between the two most likely residues at this site (V300I) did not significantly affect either $\lambda_{\max}$ (t = 2.18, df = 3.64, p = 0.102; Welch's two-tailed t-test) or $t_{1/2}$ (t = 0.52, df = 1.02, p = 0.694; Welch's two-tailed t-test). *This value excludes an outlier (t = 0.10, df = 1.85, p = 0.928 with the outlier).

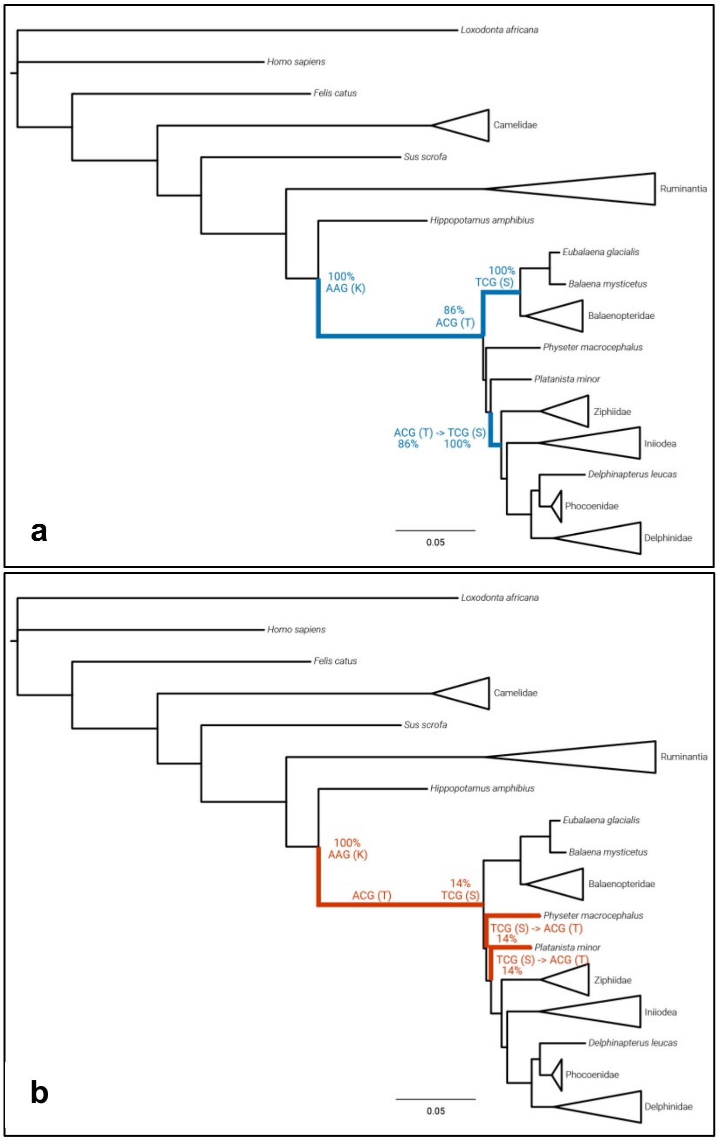

Fig. S4. Contrasting evolutionary scenarios for site 195. a, codon models. **b**, nucleotide substitutions implied by amino acid models. Despite high posterior probabilities (>0.95) under amino acid models, the nucleotide substitutions that would be required are both less probable and less parsimonious than the substitutions indicated by the codon models.

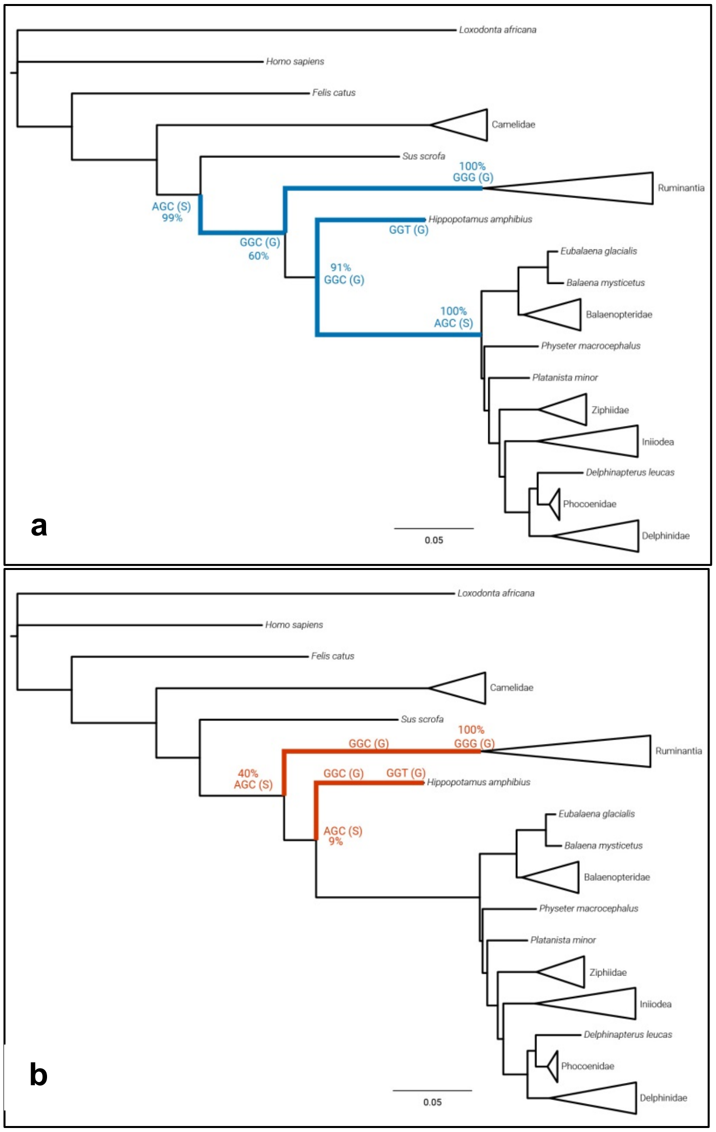

Fig. S5. Contrasting evolutionary scenarios for site 270. a, codon models. **b**, nucleotide substitutions implied by amino acid models. The scenario implied by the amino acid models suggests highly improbable nucleotide substitutions (*e.g.* 9% at the Whippomorpha node).

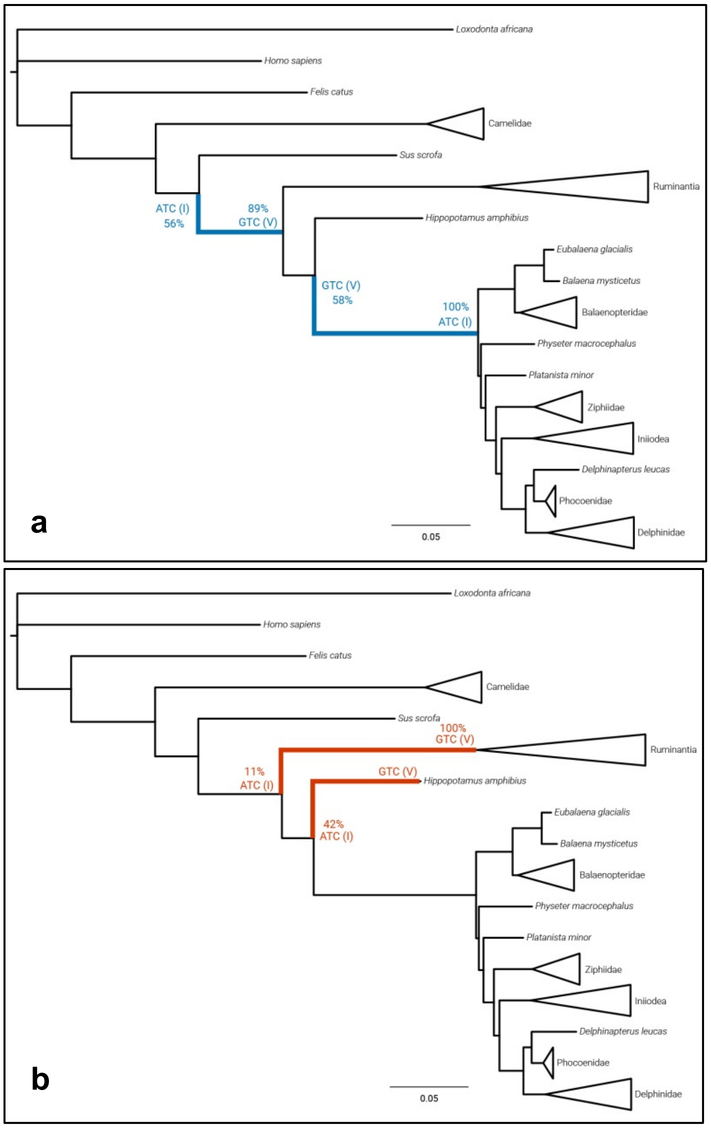

Fig. S6. **Contrasting evolutionary scenarios for site 300. a**, codon models. **b**, nucleotide substitutions implied by amino acid models. The scenario implied by the amino acid models suggests highly improbable nucleotide substitutions (*e.g.* 11% at the Cetruminantia node).

Supplementary Table S1. Likelihood ratio tests for random-sites models (PAML) of the cetacean *Rh1* species tree

Model	**np**	**ln *L***	**κ**	**Parameters[a]**			**Null**	**LRT**	**df**	**p**	**AIC**	**ΔAIC[b]**
				ω_0/p	**ω_1/q**	**ω_2/ω_p**						
M0	71	-4884.73	4.42	0.066							9911.46	371.06
M1a	72	-4755.02	4.62	0.026 (90.9%)	1.000 (9.1%)						9654.04	113.64
M2a	74	-4755.02	4.62	0.026 (90.9%)	1.000 (5.3%)	1.000 (3.8%)	M1a	0	2	1.000	9658.04	117.64
M3	75	-4723.42	4.51	0.000 (68.9%)	0.100 (22.0%)	0.575 (9.1%)	M0	322.62	4	0.000	9596.84	56.44
M7	72	-4724.02	4.52	0.099	1.104						9592.04	51.64
M8a	73	-4723.18	4.53	0.109	1.475	1.000 (1.5%)					9592.36	51.96
M8	74	-4722.99	4.54	0.108	1.405	1.295 (1.0 %)	M7	2.06	2	0.357	9593.98	53.58
							M8a	0.38	1	0.538		

Note: np, number of parameters; ln *L*, ln likelihood; κ, transition/transversion ratio; df, degrees of freedom. [a]For models M0-M3, the ω values for each site class (ω_0 - ω_2) are shown. For models M7-M8, *p* and *q* describe the shape of the beta distribution, and ω_p refers to the positively selected site class (with proportion in parentheses) for models M8 and M8a (where it is constrained to one). [b]ΔAIC is relative to the best-fitting codon model, CmD (see Supplementary Table S2).

Supplementary Table S2. Likelihood ratio tests for clade models (PAML) of the cetacean *Rh1* species tree

Model[a]	**np**	**ln *L***	**κ**	**Parameters**[b]			**Null**	**LRT**	**df**	**p**	**AIC**	**ΔAIC**
				ω_0	ω_1	ω_2/ω_d						
M2a_rel	74	-4725.25	4.54	0.007 (82.0%)	1.000 (2.3%)	0.309 (15.8%)					9598.50	58.10
M3	75	-4723.42	4.51	0.000 (68.9%)	0.100 (22.0%)	0.575 (9.1%)					9596.84	56.44
CmC_Null	75	-4700.53	4.57	0.009 (82.0%)	1.000 (1.2%)	B: 0.175 (15.7%)					9551.06	10.66
						Meso: 0.751						
						Non-Meso: 1.000						
CmC	76	-4700.30	4.57	0.010 (83.8%)	1.000 (1.4%)	B: 0.173 (14.8%)	M2a_rel	49.90	2	0.000	9552.60	12.20
						Meso: 0.790	CmC_Null	0.46	1	0.498		
						Non-Meso: 1.189						
CmD_Null	76	-4695.89	4.56	0.009 (83.5%)	0.500 (4.1%)	B: 0.136 (12.4%)					9543.78	3.38
						Meso: 0.953						
						Non-Meso: 1.000						
CmD	77	-4693.20	4.58	0.012 (84.7%)	0.500 (5.4%)	B: 0.130 (9.9%)	M3	60.44	2	0.000	9540.40	0.00
						Meso: 1.094	CmD_Null	5.38	1	0.020		
						Non-Meso: 1.821						

Note: np, number of parameters; ln L, ln likelihood; κ, transition/transversion ratio; df, degrees of freedom. [a]The clade models test the set of foreground partitions that best fit the cetacean *Rh1* dataset in Dungan et al. (2016) where there was significant evidence for divergence according to foraging depth zones that distinguish mesopelagic from non-mesopelagic (epipelagic, bathypelagic) divers. [b]The ω values for each site class (ω0 - ω2) are shown with their proportions in parentheses. For clade models, ω_d refers to the divergent site class.

Supplementary Table S3. Rhodopsin sequences used in ancestral sequence reconstruction

Common name	**Binomen**	**Accession number**
African elephant	*Loxodonta africana*	AY686752.1
Human	*Homo sapiens*	NM_000539.3
Domestic cat	*Felis catus*	NM_001009242.1
Bactrian camel	*Camelus bactrianus*	XM_010953086.1
Wild Bactrian camel	*Camelus ferus*	XM_006180073.1
Alpaca	*Vicugna pacos*	XM_006206787.1
Wild boar	*Sus scrofa*	NM_214221.1
Sheep	*Ovis aries*	XM_004018534.3
Tibetan antelope	*Pantholops hodgsonii*	XM_005955745.1
Goat	*Capra hircus*	XM_018066700.1
Water buffalo	*Bubalis bubalis*	XM_006078900.1
Plains bison	*Bison bison*	XM_010862448.1
Cattle	*Bos taurus*	NM_001014890.1
Hippo	*Hippopotamus amphibius*	KC676928.1
Bowhead whale	*Balaena mysticetus*	KC676921.1
N. Atlantic right whale	*Eubalaena glacialis*	JQ730751.1
Pygmy right whale	*Caperea marginata*	KC676926.1
N. Atlantic minke whale	*Balaenoptera acutorostrata acutorostrata*	KC676922.1
Blue whale	*Balaenoptera musculus*	KC676923.1
Fin whale	*Balaenoptera physalus*	KC676924.1
Sperm whale	*Physeter macrocephalus*	XM_007126220.1
South-Asian river dolphin	*Platanista minor*	KC676936.1
Sowerby's beaked whale	*Mesoplodon bidens*	AF055316.1
Baird's beaked whale	*Berardius bairdii*	KC676925.1
Cuvier's beaked whale	*Ziphius cavirostris*	KC676938.1
Yangtze river dolphin	*Lipotes vexillifer*	XM_007461564.1
Franciscana	*Pontoporia blainvillei*	KC676937.1
Amazon river dolphin	*Inia geoffrensis*	KC676929.1
Beluga	*Delphinapterus leucas*	KC676927.1
Finless porpoise	*Neophocaena phocaenoides*	KC676932.1
Harbour porpoise	*Phocoena phocoena*	KC676933.1
Dall's porpoise	*Phocoenoides dalli*	KC676934.1
Killer whale	*Orcinus orca*	XM_004284305.1
Bottlenose dolphin	*Tursiops truncatus*	AF055456.1
Pilot whale	*Globicephala melas*	AF055315.1
Common dolphin	*Delphinus delphis*	AF055314.1

Supplementary Table S4. Power analysis for protein assay sample sizes

Spectral tuning				**Retinal release $t_{1/2}$**			
1 nm effect	Cohen *d*=3.3	2 nm effect	Cohen *d*=6.7	4 min effect	Cohen *d*=3.1	5 min effect	Cohen *d*=3.8
n	**Power**	***n***	**Power**	***n***	**Power**	***n***	**Power**
2	0.4473	2	0.8912	2	0.4012	2	0.5381
3	0.8566	3	0.9999	3	0.8014	3	0.9328
4	0.9727	4	1.0000	4	0.9487	4	0.9937
5	0.9955	5	1.0000	5	0.9883	5	0.9995

Note: Cohen's *d* is calculated as the difference of means divided by pooled standard deviation, which we estimated as 0.3 and 1.3 for spectral tuning and retinal release respectively. These values are from bovine rhodopsin data (our positive control) in a prior publication (Morrow et al. 2017), and so provide a reasonable baseline for power analysis. Power is 1 - the type II error rate for a two-sample, two-tailed t-test given effect size (Cohen's *d*), type I error (0.05), and sample size (*n*). To detect biologically significant differences between rhodopsin samples (at least 2 nm spectral tuning and 5 min retinal release half-time), we used sample sizes of at least n = 2 for spectral tuning and n = 3 for retinal release (power at least 0.8, indicated by underlines).

References

1. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24** (2007).
2. S. Z. Dungan, A. Kosyakov, B. S. W. Chang, Spectral tuning of killer whale (Orcinus orca) rhodopsin: Evidence for positive selection and functional adaptation in a cetacean visual pigment. *Mol. Biol. Evol.* **33** (2016).
3. D. D. Pollock, B. S. W. Chang, "Dealing with uncertainty in ancestral sequence reconstruction: sampling from the posterior distribution" in *Ancestral Sequence Reconstruction*, (2008) https:/doi.org/10.1093/acprof:oso/9780199299188.003.0008.
4. H. Bar-Rogovsky, *et al.*, Assessing the prediction fidelity of ancestral reconstruction by a library approach. *Protein Eng. Des. Sel.* **28** (2015).
5. B. S. W. Chang, K. Jönsson, M. A. Kazmi, M. J. Donoghue, T. P. Sakmar, Recreating a functional ancestral archosaur visual pigment. *Mol. Biol. Evol.* **19** (2002).
6. B. S. W. Chang, *et al.*, "The future of codon models in studies of molecular function: Ancestral reconstruction and clade models of functional divergence" in *Codon Evolution: Mechanisms and Models*, (2012) https:/doi.org/10.1093/acprof:osobl/9780199601165.003.0011.
7. C. Bickelmann, *et al.*, The molecular origin and evolution of dim-light vision in mammals. *Evolution (N. Y).* **69** (2015).
8. E. A. Gaucher, S. Govindarajan, O. K. Ganesh, Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451** (2008).
9. M. J. Harms, J. W. Thornton, Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20** (2010).
10. R. Merkl, R. Sterner, Ancestral protein reconstruction: Techniques and applications. *Biol. Chem.* **397** (2016).
11. Z. Yang, S. Kumar, M. Nei, A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141** (1995).
12. T. Pupko, I. Pe'er, R. Shamir, D. Graur, A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* **17** (2000).
13. W. Delport, A. F. Y. Poon, S. D. W. Frost, S. L. Kosakovsky Pond, Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26** (2010).
14. R. Nielsen, Mapping mutations on phylogenies in *Systematic Biology*, (2002).
15. M. Arenas, C. C. Weber, D. A. Liberles, U. Bastolla, ProtASR: An Evolutionary Framework for Ancestral Protein Reconstruction with Selection on Folding Stability. *Syst. Biol.* **66** (2017).
16. T. Okada, *et al.*, Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **99** (2002).