Marine Biological Laboratories
Workshop in Molecular Evolution

Adaptive protein evolution: Detecting changes in selection

Belinda Chang

Department of Ecology & Evolutionary Biology
Department of Cell & Systems Biology
University of Toronto

@opsinlab

Adaptive protein evolution

1

Phylogenetic approaches to the study of protein structure and function

- Ancestral protein reconstruction

- Computational analyses of selection (dN/dS)

-> Combining computational with experimental approaches allows us to test hypotheses of selection in protein evolution

Codon models of evolution -
# Ratio of nonsynonymous/synonymous substitution rates
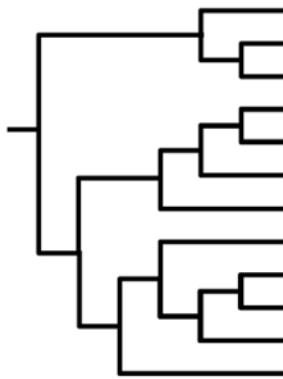
$\omega = dN/dS$

A measure of evolutionary constraint which can provide insight into the form and strength of selection on proteins.
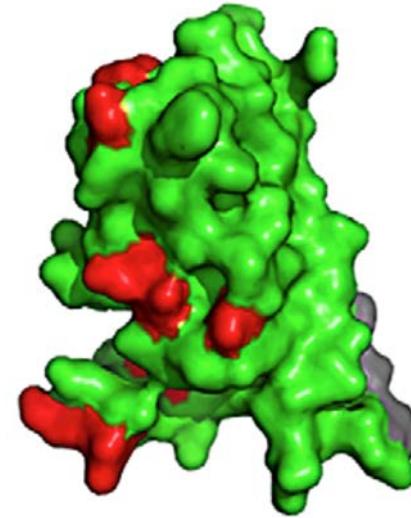
- purifying selection (AA changes are purged) $\longrightarrow$ $\omega < 1$
- neutral evolution $\longrightarrow$ $\omega = 1$
- positive selection (AA changes are favoured) $\longrightarrow$ $\omega > 1$

ADAPTIVE EVOLUTION:
## Combining computational and experimental approaches to test hypotheses of selection in protein evolution

```
GLELHPDYKTWSPEQVCSFLRRGGF
GPELHPDHKTWGPEQVCSFLRRGGF
GLELHPDYKTWGPEQVCSFLRRGGF
GLELHPDYKTWDPEQVCSFLRRGGF
GLELHPDYKTWGPEQVCFFLRGGGF
GLELHLDYKTWDPEQVCFFLRRGGF
GLELHPDYKTWGPEQVCFFLRRGGF
GLELHPDYKTWGPEQVCFFLRRGGF
GLELHPDYKTWDPEQVCFFLRRGGF
GLELHPDYKTWDPEQVCFFLRRGGF
GLELHPDYKTWDPEQVCFFLRRGGF
GLELDPDYKTWDPEQVCSFLGRGGF
```

- Testing hypotheses of selection in a phylogenetic framework
- Identification of positively selected sites/regions
- Experimental studies of shifts in protein function

4

ADAPTIVE EVOLUTION:
Combining computational and experimental approaches to test hypotheses of selection in protein evolution

```
GLELHPDYKTWSPEQVCSFLRRGGF
GPELHPDHKTWGPEQVCSFLRRGGF
GLELHPDYKTWGPEQVCSFLRRGGF
GLELHPDYKTWDPEQVCSFLRRGGF
GLELHPDYKTWGPEQVCFFLRGGGF
GLELHLDYKTWDPEQVCFFLRRGGF
GLELHPDYKTWGPEQVCFFLRRGGF
GLELHPDYKTWGPEQVCFFLRRGGF
GLELHPDYKTWDPEQVCFFLRRGGF
GLELHPDYKTWDPEQVCFFLRRGGF
GLELHPDYKTWDPEQVCFFLRRGGF
GLELDPDYKTWDPEQVCSFLGRGGF
```

- Testing hypotheses of selection in a phylogenetic framework
- Identification of positively selected sites/regions
- Experimental studies of shifts in protein function

5

Testing hypotheses of selection in protein evolution: Combining computational and experimental approaches

1. Formulate hypothesis of selection acting on evolution of particular group of sequences
2. Collect target group of sequences
3. Test hypothesis of selection using codon models in a phylogenetic framework
4. Identify amino acid sites/regions that have been targets of selection
5. Formulate hypothesis of selection acting on particular aspect of protein function
6. Select proteins/mutants for experimental studies
7. Test hypothesis of selection using experiments investigating shifts in protein function

Testing hypotheses of selection in protein evolution: Combining computational and experimental approaches

1. Formulate hypothesis of selection acting on evolution of particular group of sequences
2. Collect target group of sequences
3. Test hypothesis of selection using codon models in a phylogenetic framework
4. Identify amino acid sites/regions that have been targets of selection
5. Formulate hypothesis of selection acting on particular aspect of protein function
6. Select proteins/mutants for experimental studies
7. Test hypothesis of selection using experiments investigating shifts in protein function

Testing hypotheses of adaptive evolution in protein function: Combining computational and experimental approaches

Computational analyses
- Statistical test for positive, or divergent selection in a phylogenetic context
- Identifies amino acid residues that may have been the targets of positive selection
- Does NOT prove adaptive evolution -> Need function

Experimental studies
- Empirical test for shifts in protein function as a consequence of substitutions at sites targeted by selection
- Does NOT prove anything about selection & adaptive evolution, need evolutionary context
- Ideally combined with ancestral reconstruction and mutagenesis studies

Testing hypotheses of adaptive evolution in protein function: Combining computational and experimental approaches

Computational analyses
- Statistical test for positive, or divergent selection in a phylogenetic context
- Identifies amino acid residues that may have been the targets of positive selection
- Does NOT prove adaptive evolution -> Need function
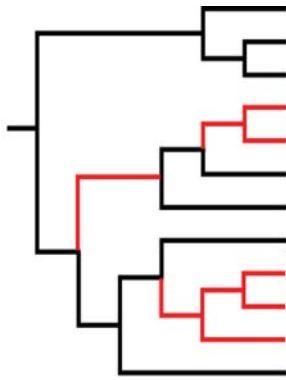
Experimental studies
- Empirical test for shifts in protein function as a consequence of substitutions at sites targeted by selection
- Ideally combined with ancestral reconstruction and mutagenesis studies

-> Taken together, computational and experimental approaches can be used to test for ADAPTIVE EVOLUTION in protein function as a consequence of natural selection!
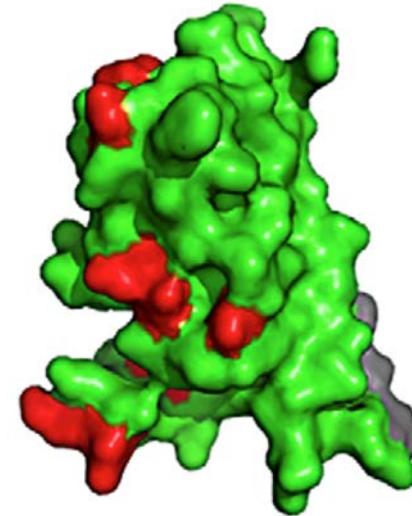
Formulating specific and testable hypotheses of adaptive evolution

Pervasive vs. episodic selection

Sites targeted by selection

Effect on protein function

-> Answering the question, not only when/where but WHY did selection occur?

-> The role of a priori information in formulating hypotheses about function

Examples of positive selection on proteins: Experimental models

Host-viral arms races

- Primate intracellular immune response TRIM5a, protein kinase R (Sawyer et al. 2005, Elde et al. 2009)
- Potato virus Y (Moury & Simon 2011)
- Neutrophil-mediated host response to infection (Loughran et al. 2012)

Sperm-egg recognition proteins

- Abalone lysin, egg coat (Aagaard et al. 2013)
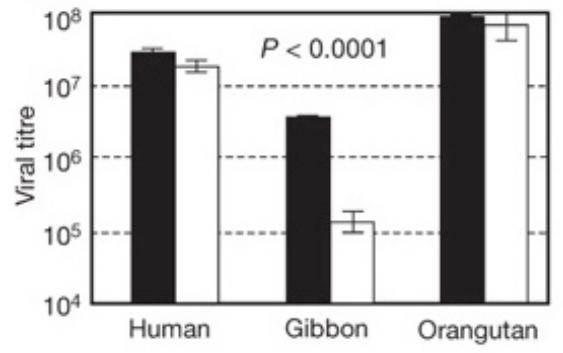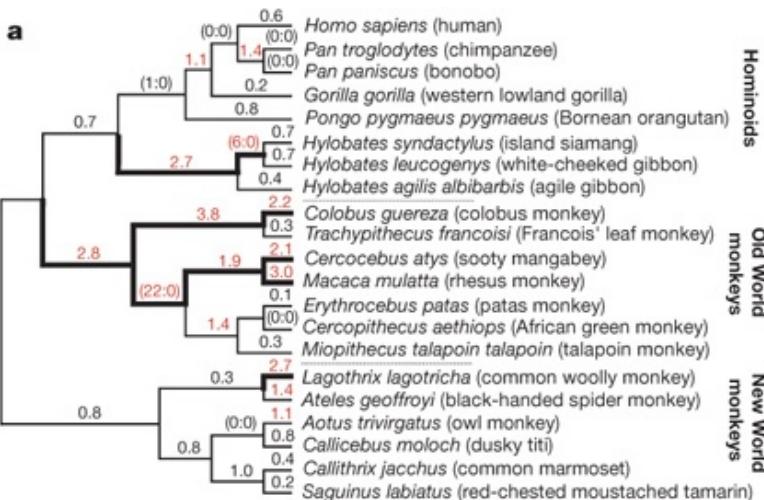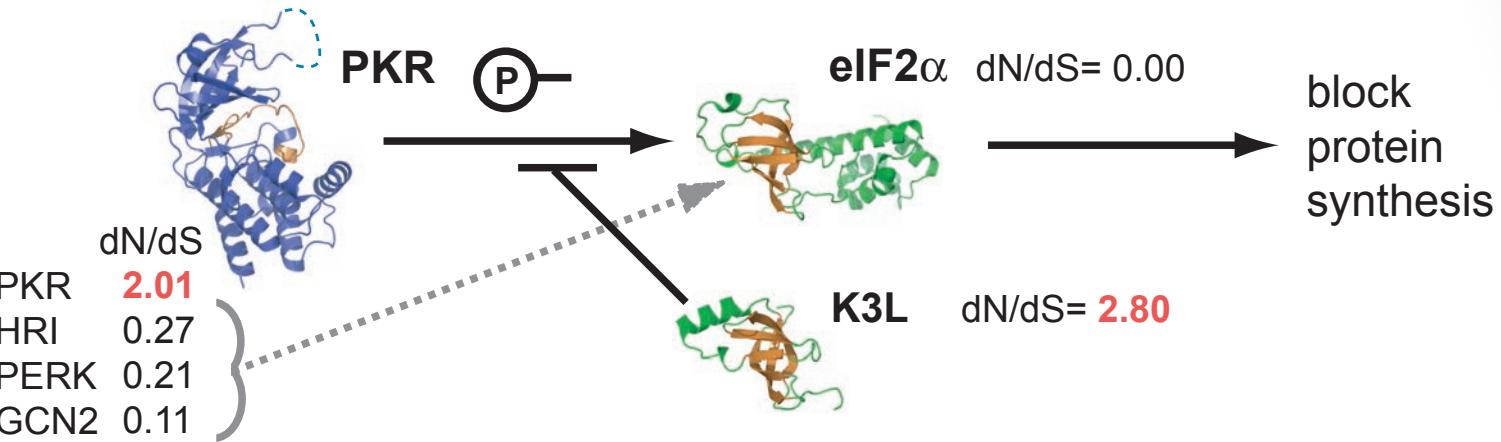
Sensory proteins

- Visual pigments in whales, fishes, birds, bats (Dungan & Chang 2017, Hauser et al. 2017)

Transcription factors in development
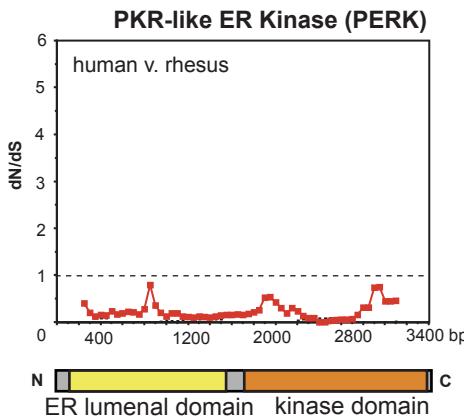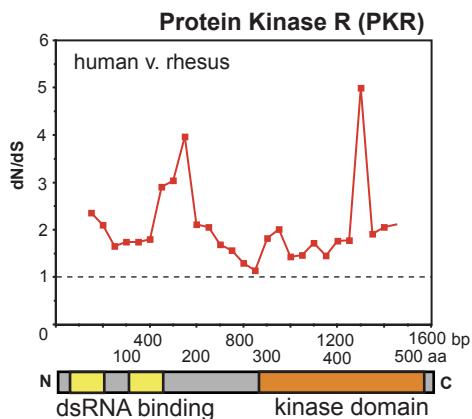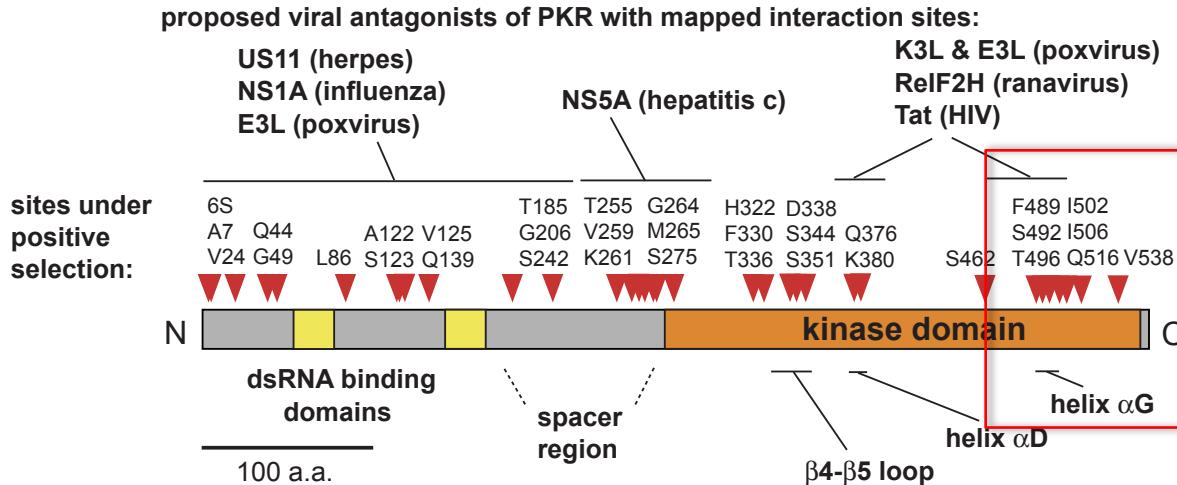
- Stem cell pluripotency (Baker et al. 2016)

Primate protein kinase R locked in arms race with viral protein mimic K3L

PKR → (P) — eIF2α dN/dS= 0.00 → block protein synthesis

K3L dN/dS= **2.80**

dN/dS
PKR **2.01**
HRI 0.27
PERK 0.21
GCN2 0.11

a

		0.6	Homo sapiens (human)
(0:0)	(0:0)	(0:0)	Pan troglodytes (chimpanzee)
1.1	1.4		Pan paniscus (bonobo)
(1:0)	0.2		Gorilla gorilla (western lowland gorilla)
0.7	0.8		Pongo pygmaeus pygmaeus (Bornean orangutan)

0.7
(6:0) 0.7   Hylobates syndactylus (island siamang)
2.7   0.7   Hylobates leucogenys (white-cheeked gibbon)
       0.4   Hylobates agilis albibarbis (agile gibbon)
3.8   2.2   Colobus guereza (colobus monkey)
       0.3   Trachypithecus francoisi (Francois' leaf monkey)
1.9   2.1   Cercocebus atys (sooty mangabey)
(22:0) 3.0   Macaca mulatta (rhesus monkey)
2.8   0.1   Erythrocebus patas (patas monkey)
1.4   (0:0) Cercopithecus aethiops (African green monkey)
       0.3   Miopithecus talapoin talapoin (talapoin monkey)
       2.7   Lagothrix lagotricha (common woolly monkey)
0.3   1.4   Ateles geoffroyi (black-handed spider monkey)
       1.1   Aotus trivirgatus (owl monkey)
0.8   (0:0) 0.8 Callicebus moloch (dusky titi)
1.0   0.4   Callithrix jacchus (common marmoset)
       0.2   Saguinus labiatus (red-chested moustached tamarin)

Hominoids
Old World monkeys
New World monkeys

$P < 0.0001$

Viral titre: 10^4 – 10^8

Human   Gibbon   Orangutan

Elde et al. Nature 2009

Positive selection has shaped PKR in primate evolution

Elde et al. Nature 2009
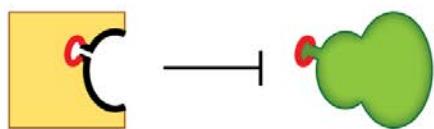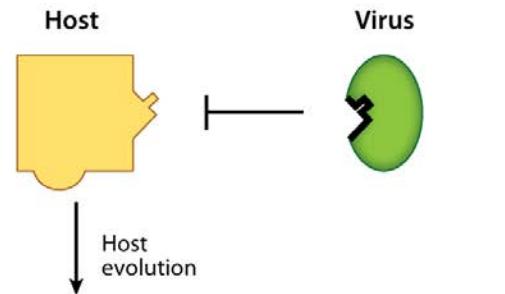
Positively selected sites on distinct surfaces of the PKR kinase domain are crucial to K3L resistance

14

Elde et al. Nature 2009

a Offense

Host   Virus
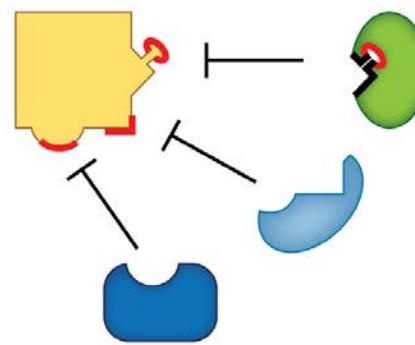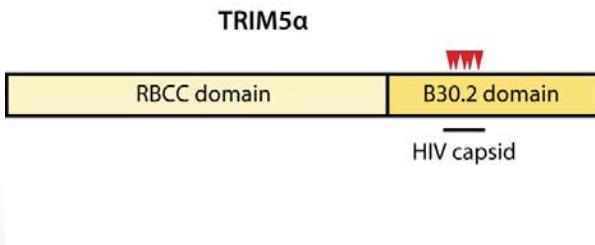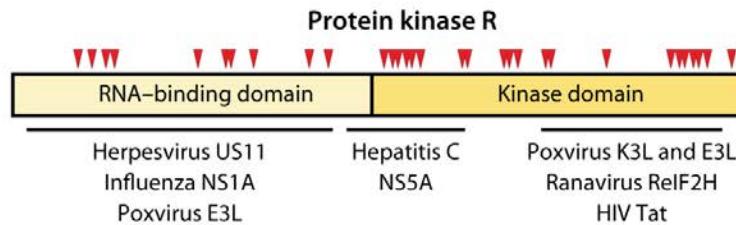
Viral evolution

Host evolution

b Defense

Host   Virus

Host evolution

Viral evolution

TRIM5α

RBCC domain	B30.2 domain

HIV capsid

Protein kinase R

RNA–binding domain	Kinase domain

Herpesvirus US11   Hepatitis C   Poxvirus K3L and E3L
Influenza NS1A   NS5A   Ranavirus ReIF2H
Poxvirus E3L   HIV Tat
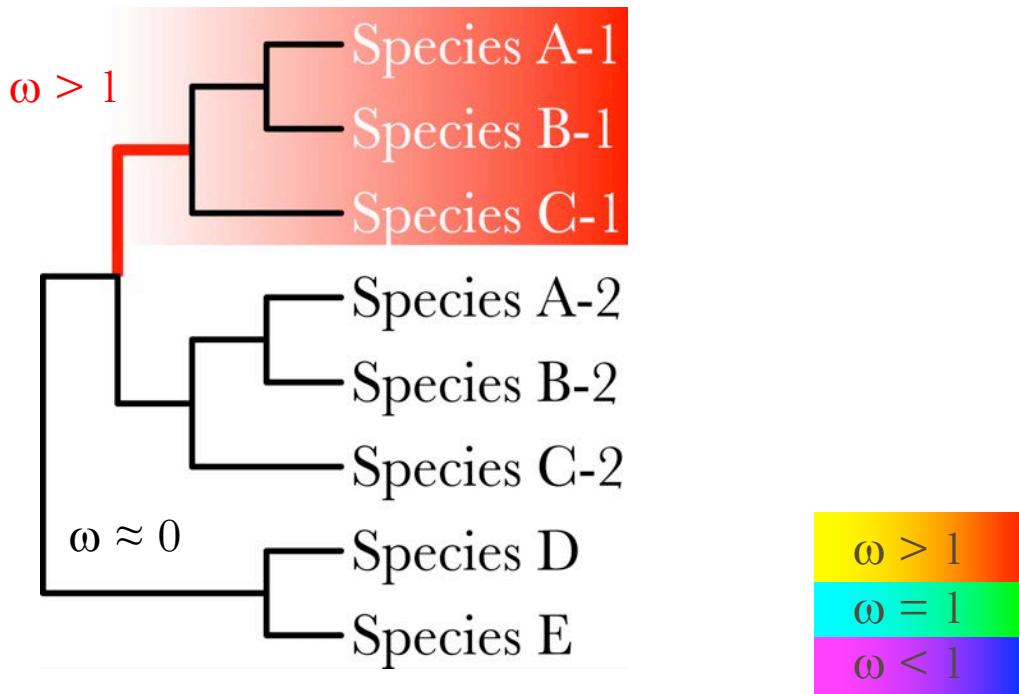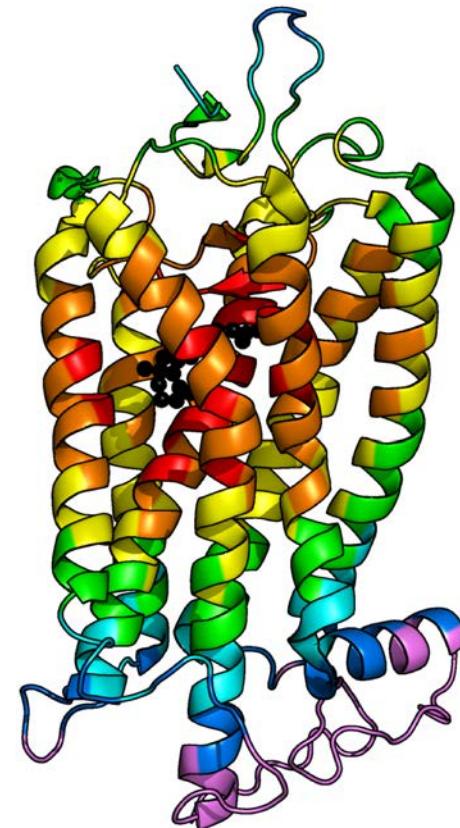
Beyond random sites and branch-sites models: Clade models

The form and strength of selection may vary across both the phylogeny and the protein.
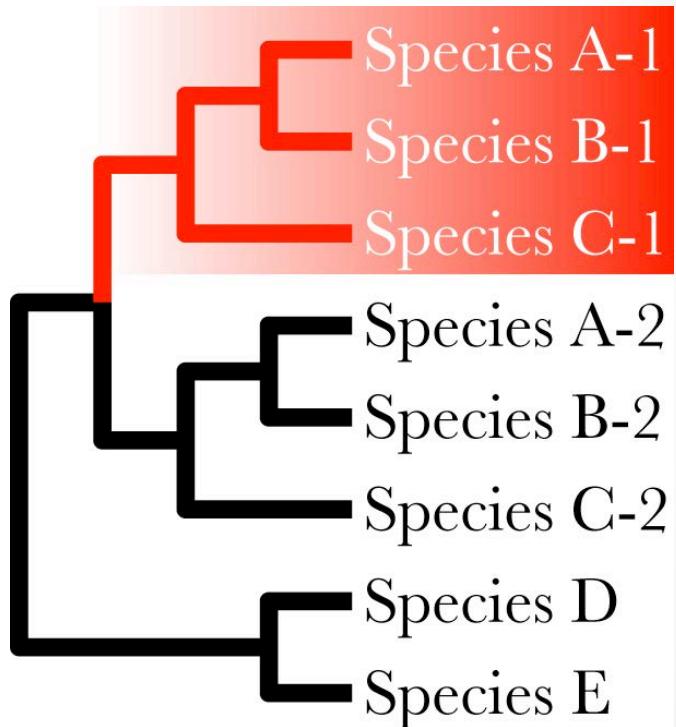
Codon models can incorporate variation across sites and lineages, and can be used to detect positive and divergent selection.

Bielawski & Yang (2004)

16

Codon models of divergent selection:
Clade model C (CmC)

CmC assumes sites evolve in three ways:

1. <u>purifying selection</u> ($0 < \omega_0 < 1$) constrains evolution at some sites, operating consistently across the tree.

2. some sites evolve free of constraint (<u>neutrally</u>: $\omega_1 = 1$) across the entire tree.

3. the remaining sites evolve <u>divergently</u> ($\omega_2 \neq \omega_3$) between particular sections of the tree.

Bielawski & Yang (2004)

Codon model tests of divergent selection
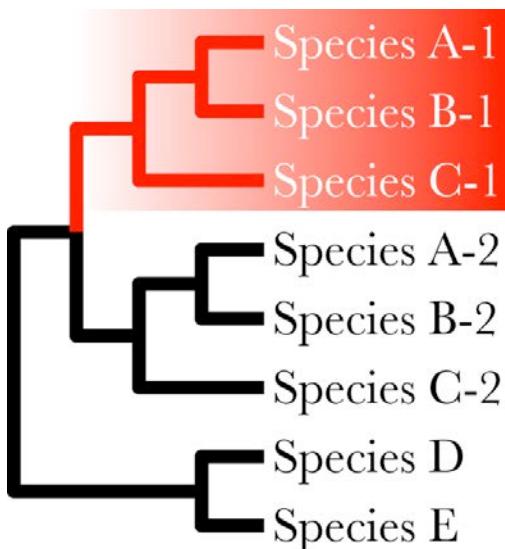
CmC assumes sites evolve in three ways:

1. underline always purifying selection
   $(0 < \omega_0 < 1)$

2. always neutral
   $(\omega_1 = 1)$

3. divergent
   $(\omega_2 \neq \omega_3)$
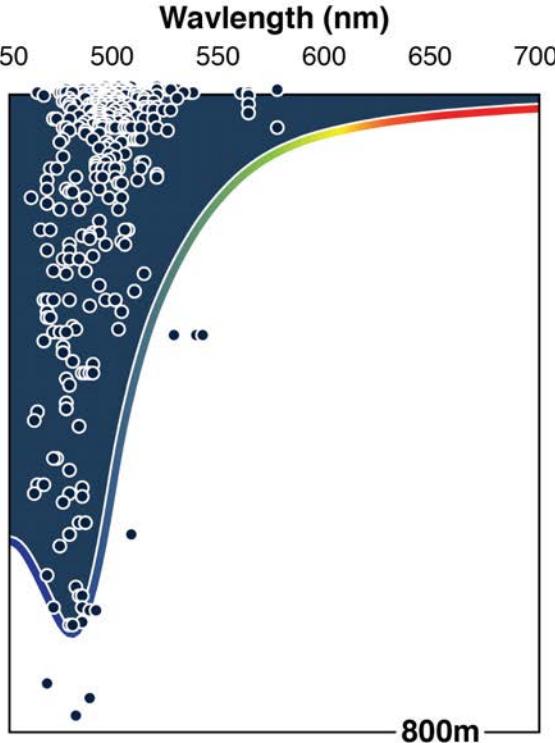
3 (alt). no divergence
   $(\omega_2 > 0)$
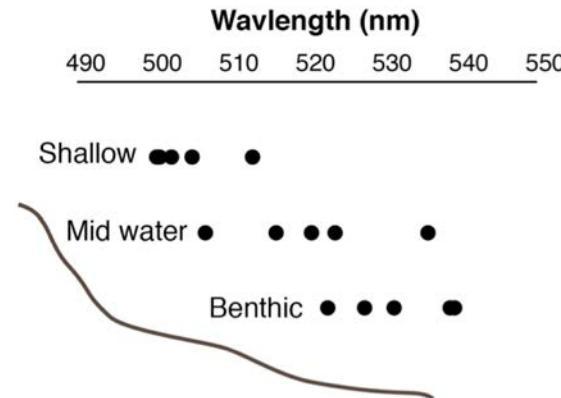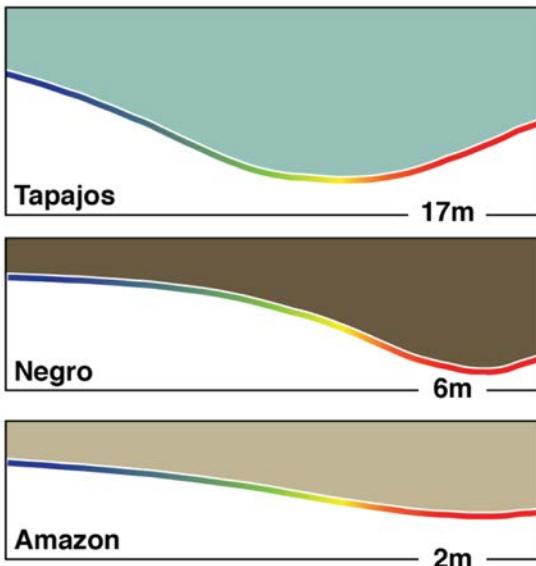
An improved null model for testing for divergent selection, M2a_rel model (Weadick & Chang 2012).

M2a_rel retains the third site class, but simply estimates a single ω ratio that applies to all branches of the phylogeny.

Species A-1
Species B-1
Species C-1
Species A-2
Species B-2
Species C-2
Species D
Species E

Vision in aquatic environments

- Marine environments become increasingly **blue** with depth

- Deep-sea fishes have **blue-shifted** Rhodopsin

- Rivers are **red-shifted** and dimmer than marine
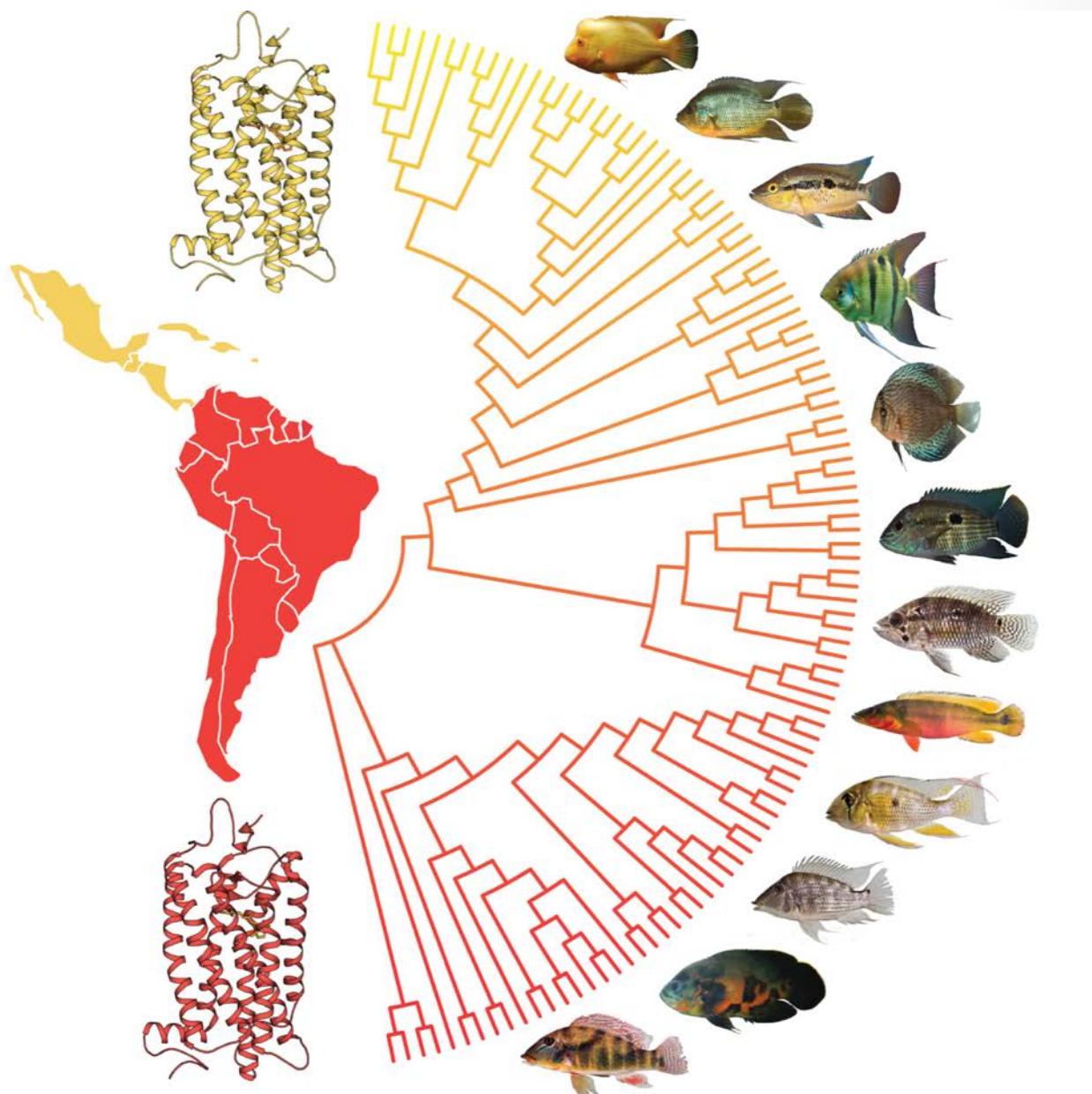
Freshwater fishes have **red-shifted** rhodopsin pigments

19

Molecular evolution of rhodopsin in Neotropical cichlids
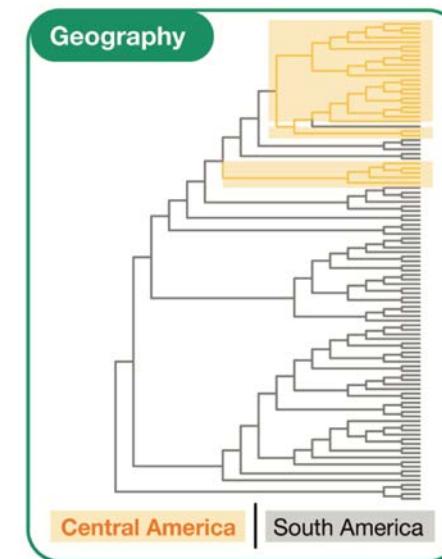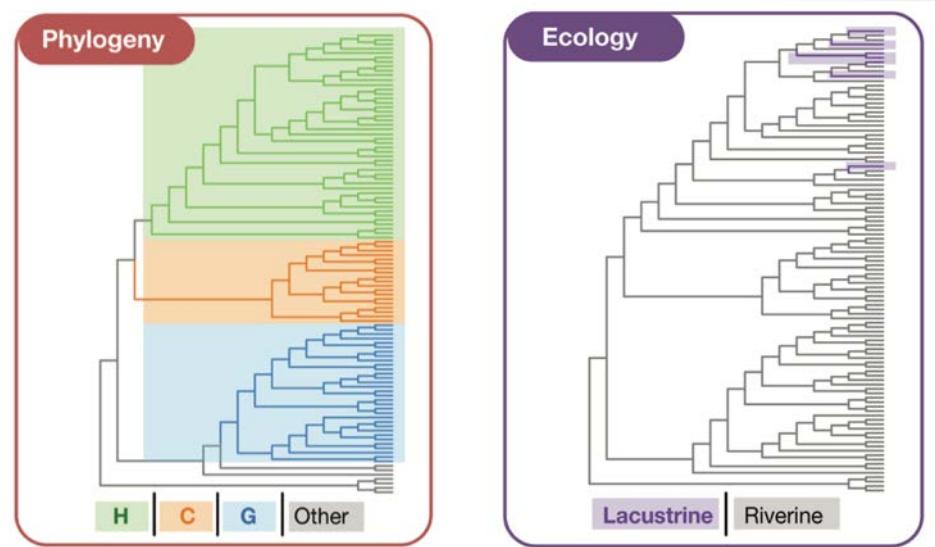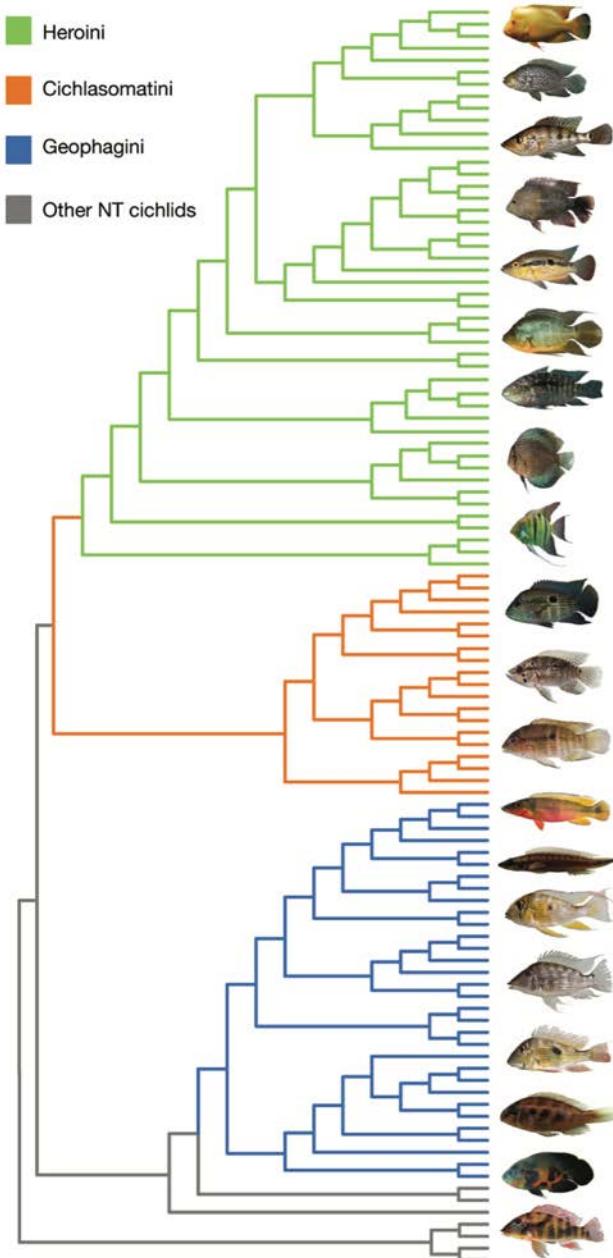
Frances Hauser

Hernan Lopez-Fernandez, ROM

# Clade models used to test different hypotheses of Neotropical cichlid rhodopsin evolutionary divergence
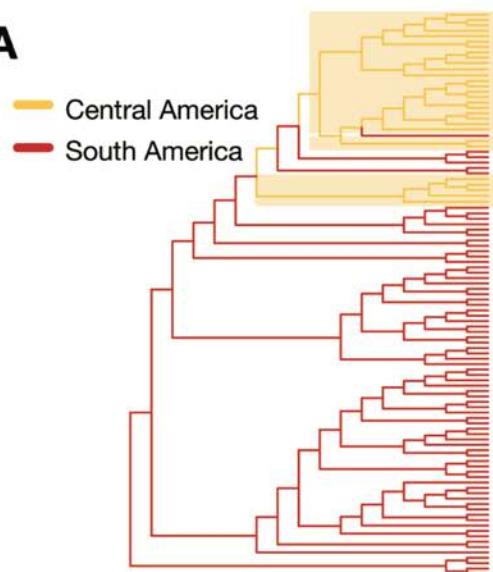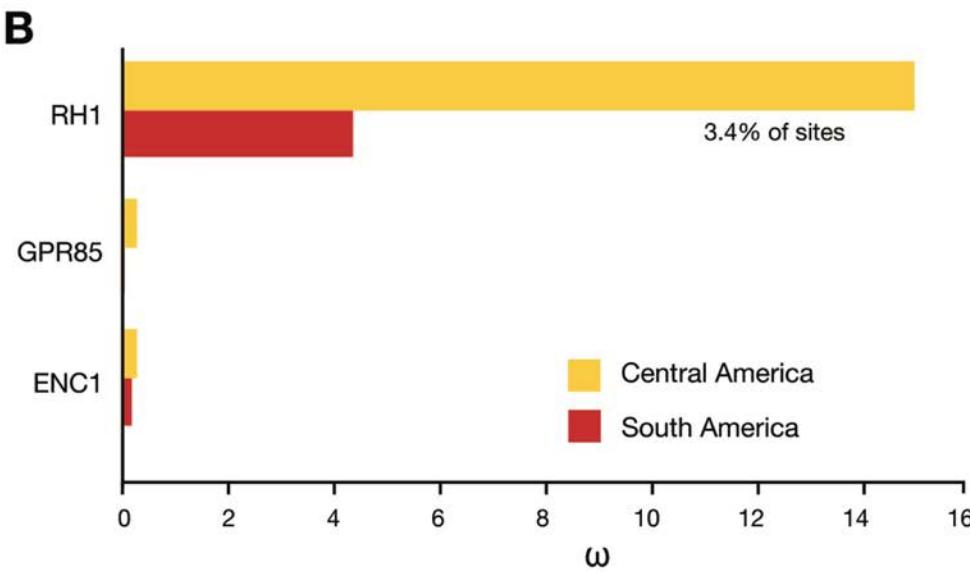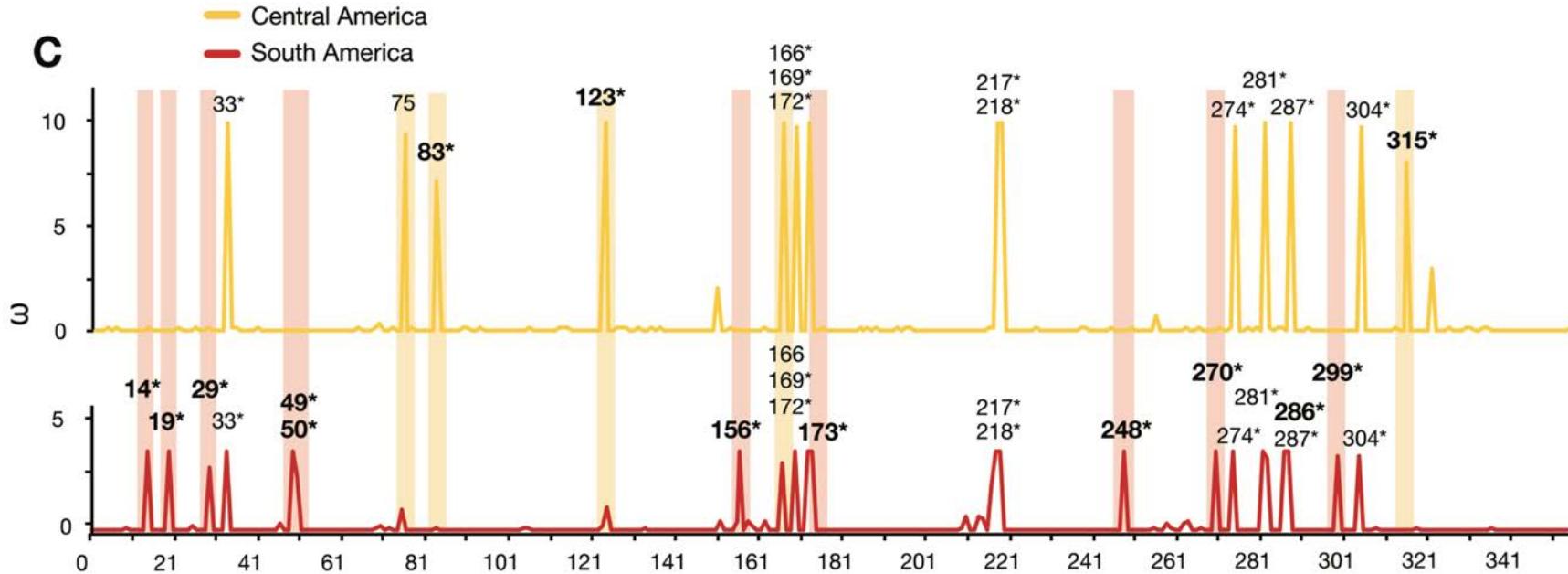
Hauser et al. 2017
MBE
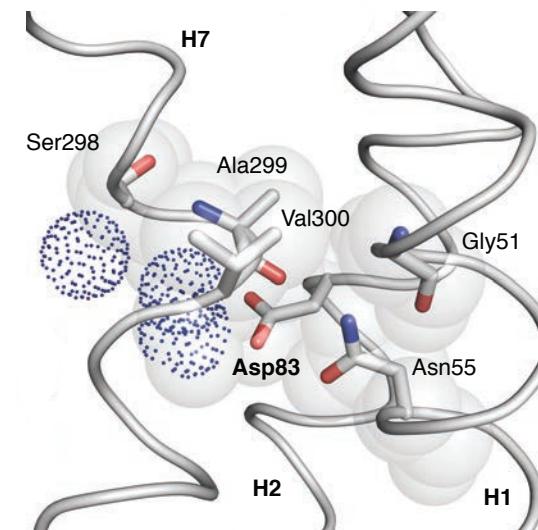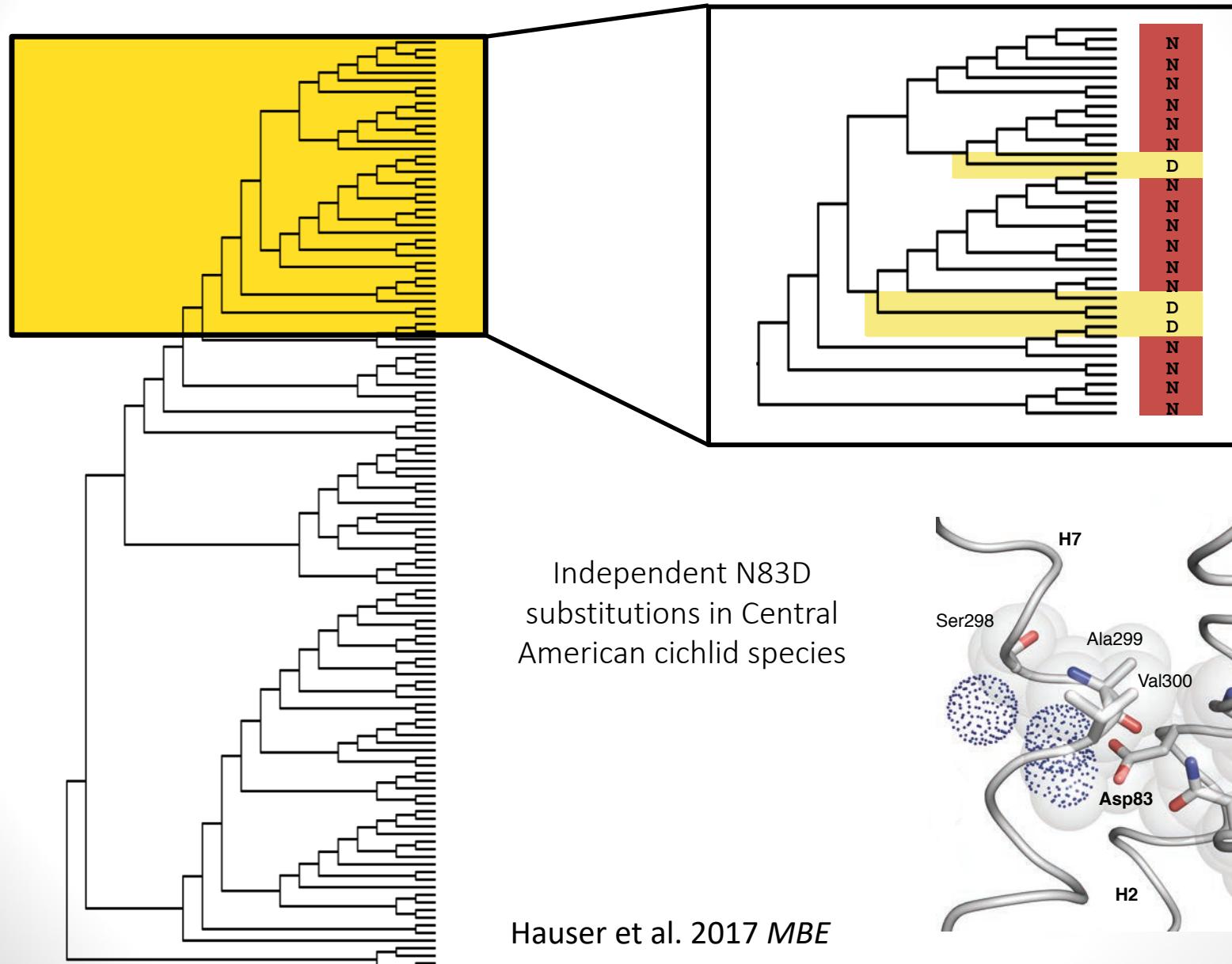
Model (foreground clade)	np	lnL	k	AIC	ΔAIC	Parameters				Null	LRT	df	P
M2a_rel	210	-5909.85	3.12	12239.70	17.46	p:	0.868	0.091	0.041	M1a	250.160	2	**0.0000**
						w:	0.009	1	5.470				
C+H+G	213	-5904.67	3.06	12235.34	13.10	site	0	1	2	M2a_rel	10.360	3	0.0157
						proportion	0.868	0.097	0.034				
						background	0.009	1	3.979				
						Cichlasomatini	0.009	1	6.357				
						Heroini	0.009	1	8.710				
						Geophagini	0.009	1	4.100				
Central America (clade)	211	-5900.12	2.41	12222.24	7.34	site	0	1	2	M2a_rel	19.464	1	**0.0000**
						proportion	0.868	0.097	0.034				
						background	0.009	1	4.476				
						foreground	0.009	1	11.660				
Cichlasomatini	211	-5909.84	3.12	12241.68	19.44	site	0	1	2	M2a_rel	0.020	1	0.8875
						proportion	0.868	0.091	0.040				
						background	0.009	1	5.430				
						foreground	0.009	1	5.710				
Heroini	211	-5906.21	3.09	12234.42	12.18	site	0	1	2	M2a_rel	7.280	1	**0.0070**
						proportion	0.868	0.095	0.037				
						background	0.009	1	4.630				
						foreground	0.009	1	8.020				
Geophagini	211	-5907.70	3.10	12237.40	15.16	site	0	1	2	M2a_rel	4.300	1	**0.0381**
						proportion	0.868	0.094	0.038				
						background	0.009	1	6.520				
						branch	0.009	1	4.200				
Central America	**211**	**-5896.45**	**3.07**	**12214.90**	**0.00**	**site**	**0**	**1**	**2**	**M2a_rel**	**26.800**	**1**	**0.0000**
						proportion	**0.868**	**0.097**	**3.440**				
						background	**0.009**	**1**	**4.500**				
						Central America	**0.009**	**1**	**14.800**				
Lake-dwelling	211	-5909.02	3.11	12240.03	25.13	site	0	1	2	M2a_rel	1.670	1	**0.1963**
						proportion	0.868	0.092	0.040				
						background	0.009	1	5.410				
						foreground	0.009	1	10.250				

Adaptive protein evolution

Hauser et al. 2017 *MBE*

23

Rhodopsin site 83 identified as under positive selection and undergoes independent transitions in Central American cichlids

Independent N83D substitutions in Central American cichlid species

Hauser et al. 2017 *MBE*

Spectroscopic assays of rhodopsin function

*In vitro* expression
& purification

Plasmid

transfection

HEK293T cells
or SK-N-SH cells

immunoaffinity
purification

Purified
protein

reconstitution

Spectroscopic assays

Relative Fluorescence

Time (min)

Light-activated
retinal release

$h\nu$

Absorbance

Wavelength (nm)

400

450 500

Dark-light
difference
spectrum

light

11-*cis*

1

all-*trans*

2

opsin

+

all-*trans* retinal

The N83D mutation produces a significant shift in rhodopsin function

N83D in Central American fishes mediates faster kinetics

Hauser et al. 2017 *MBE*

Accelerated rhodopsin divergence during invasion of Central America

Central America

Combination of computational and functional studies suggest **adaptation** to increased light levels in clear streams of Central America

Adaptive protein evolution

Positive selection in RH1
3.4% of sites

Background (South America)	ω = 4.5
Foreground (Central America)	ω = **14.0**
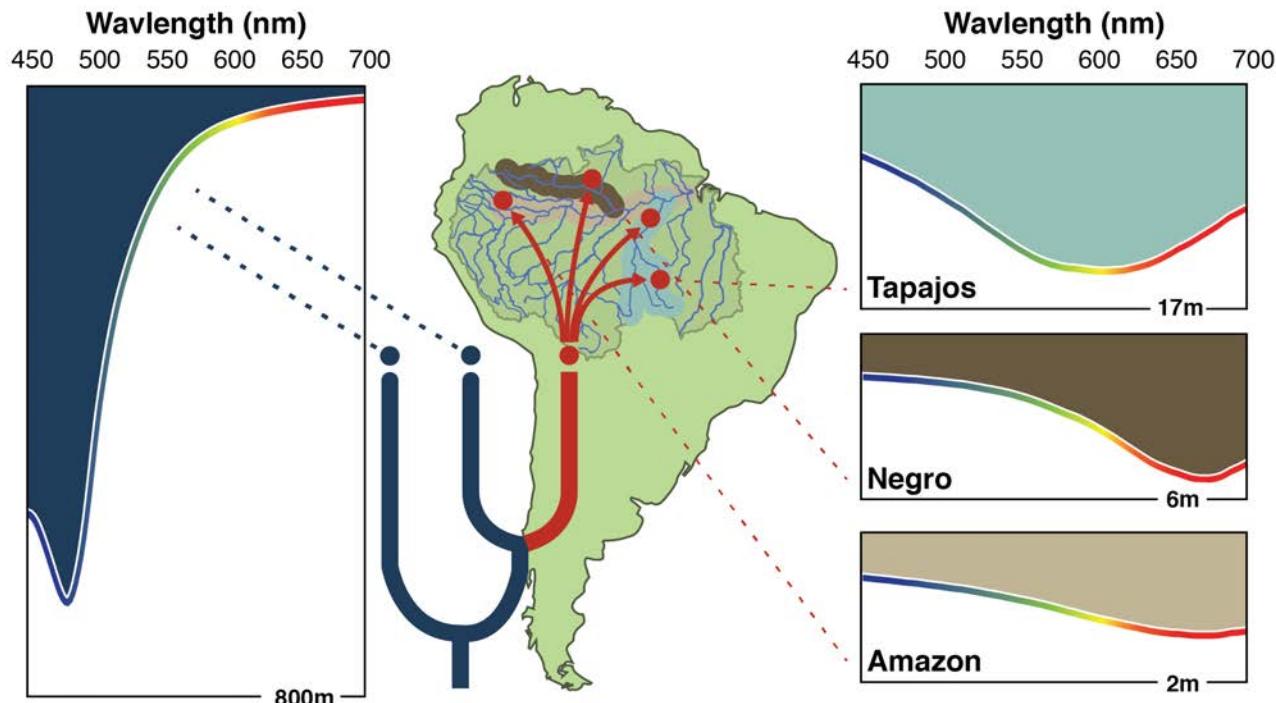
Hauser et al. 2017 *MBE*

Rhodopsin evolution in marine-derived Amazonian anchovies

Hypothesis: Adaptive evolution in rhodopsin gene of freshwater anchovies

Alex Van Nynatten

Nathan Lovejoy
U of Toronto
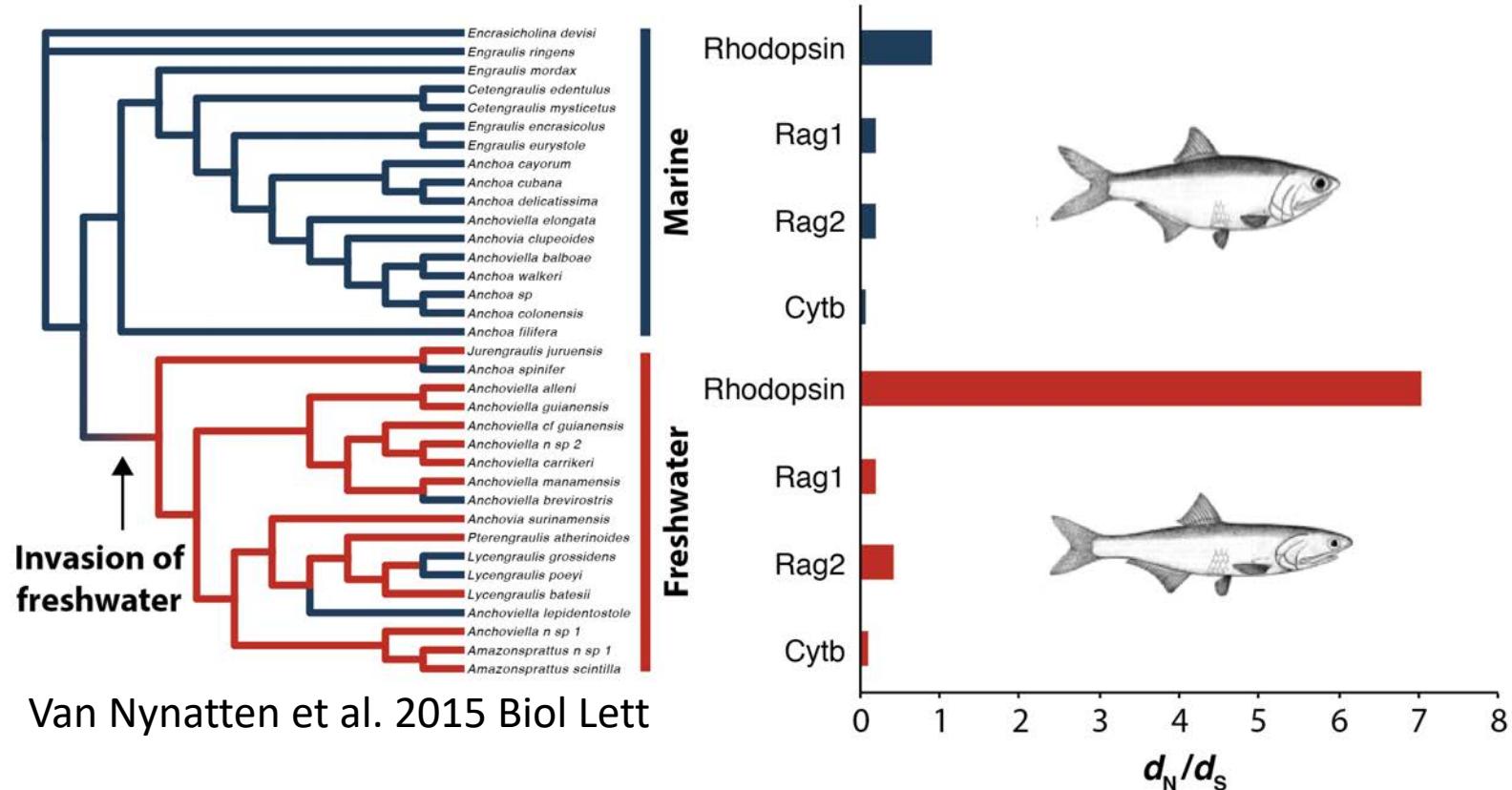
- Closely related marine and freshwater clades
- **Ancestrally marine**

Van Nynatten et al. 2015 Biol Lett

28

Rhodopsin evolution in marine-derived Amazonian anchovies
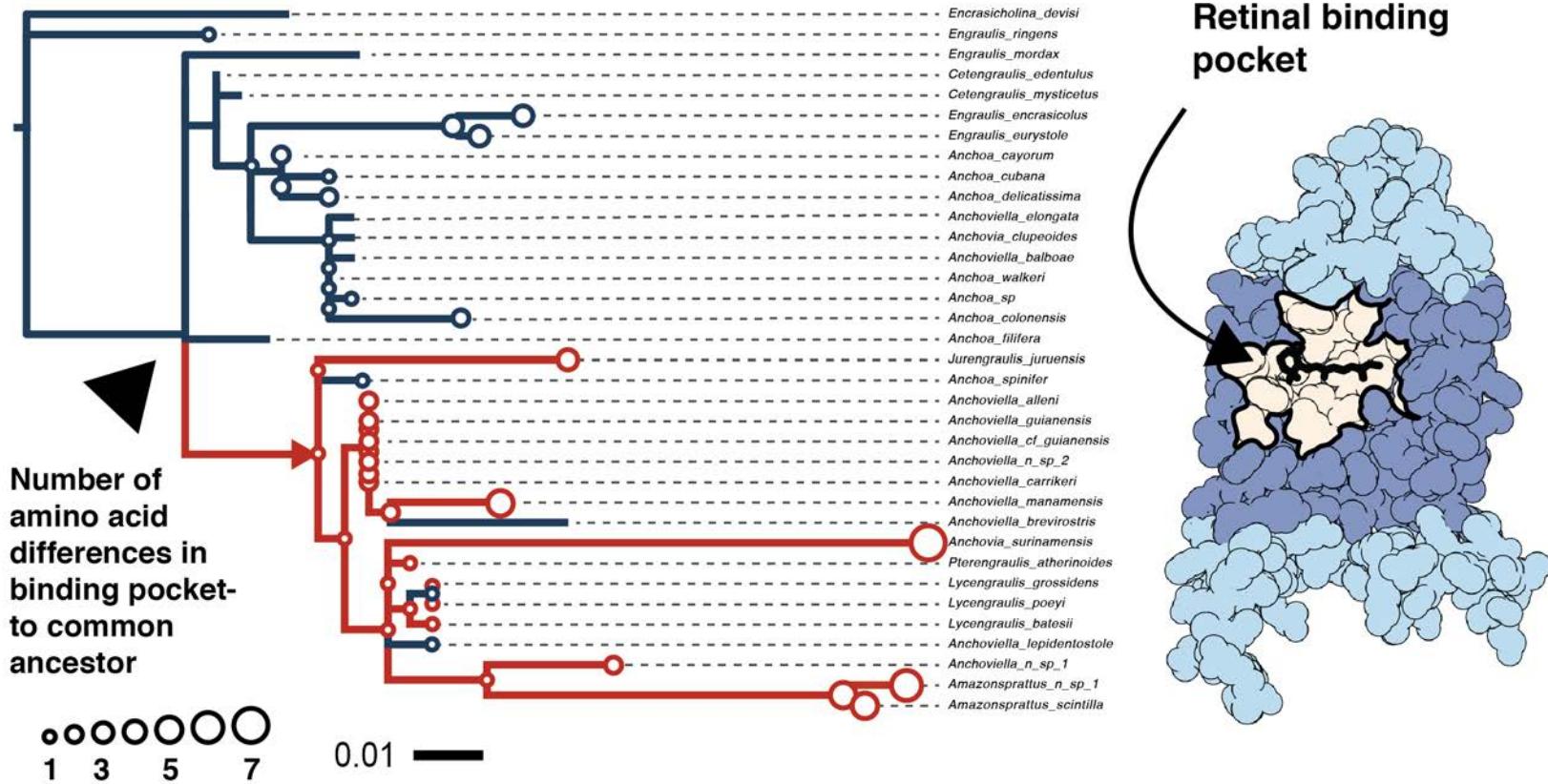
Hypothesis: Adaptive evolution in rhodopsin gene of freshwater anchovies

Van Nynatten et al. 2015 Biol Lett

29

- Increased dN/dS (CmC) in rhodopsin of the freshwater clade
- No difference in non-visual control genes

Amino acid substitutions in anchovy rhodopsin
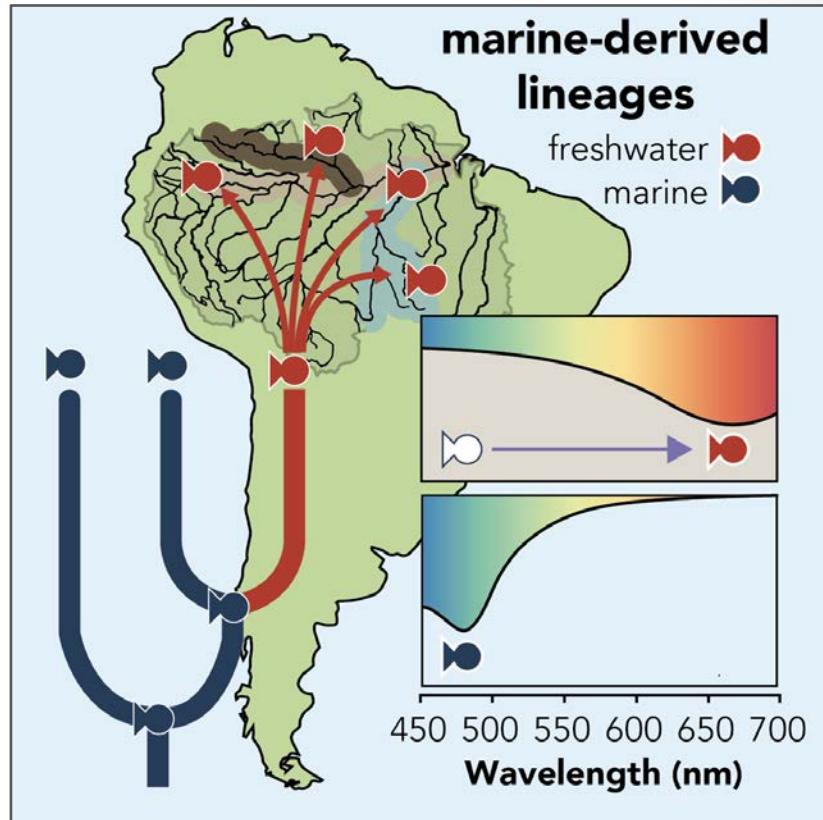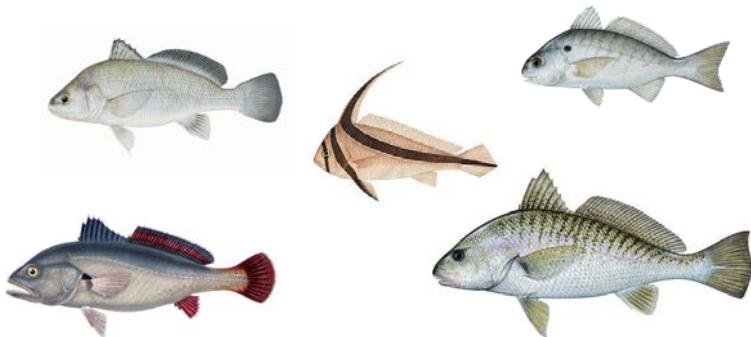
- Longer branch lengths in freshwater clade when scaled by number of amino acid substitutions in rhodopsin
- More remodelled chromophore binding pocket in freshwater clade
  Larger circles at nodes represent more amino acid differences in retinal binding pocket residues with common ancestor (arrowhead)
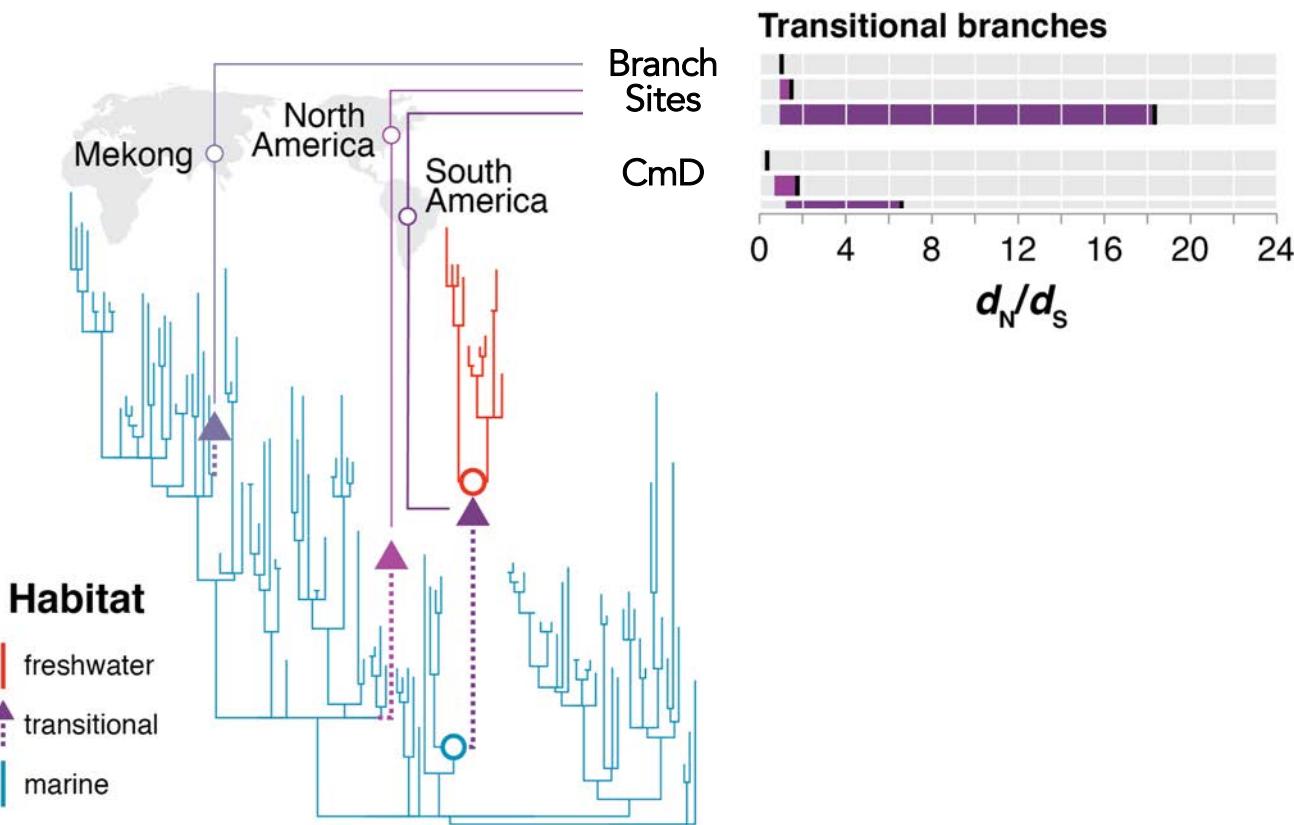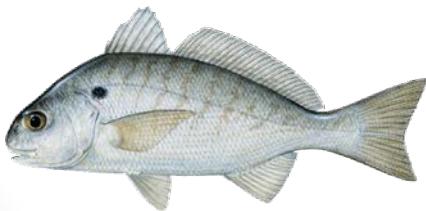
Drum and croakers (family: Sciaenidae)

- Miocene marine incursion events resulted in dramatic radiation
- Diverse morphologies, adaptations to dim light

31
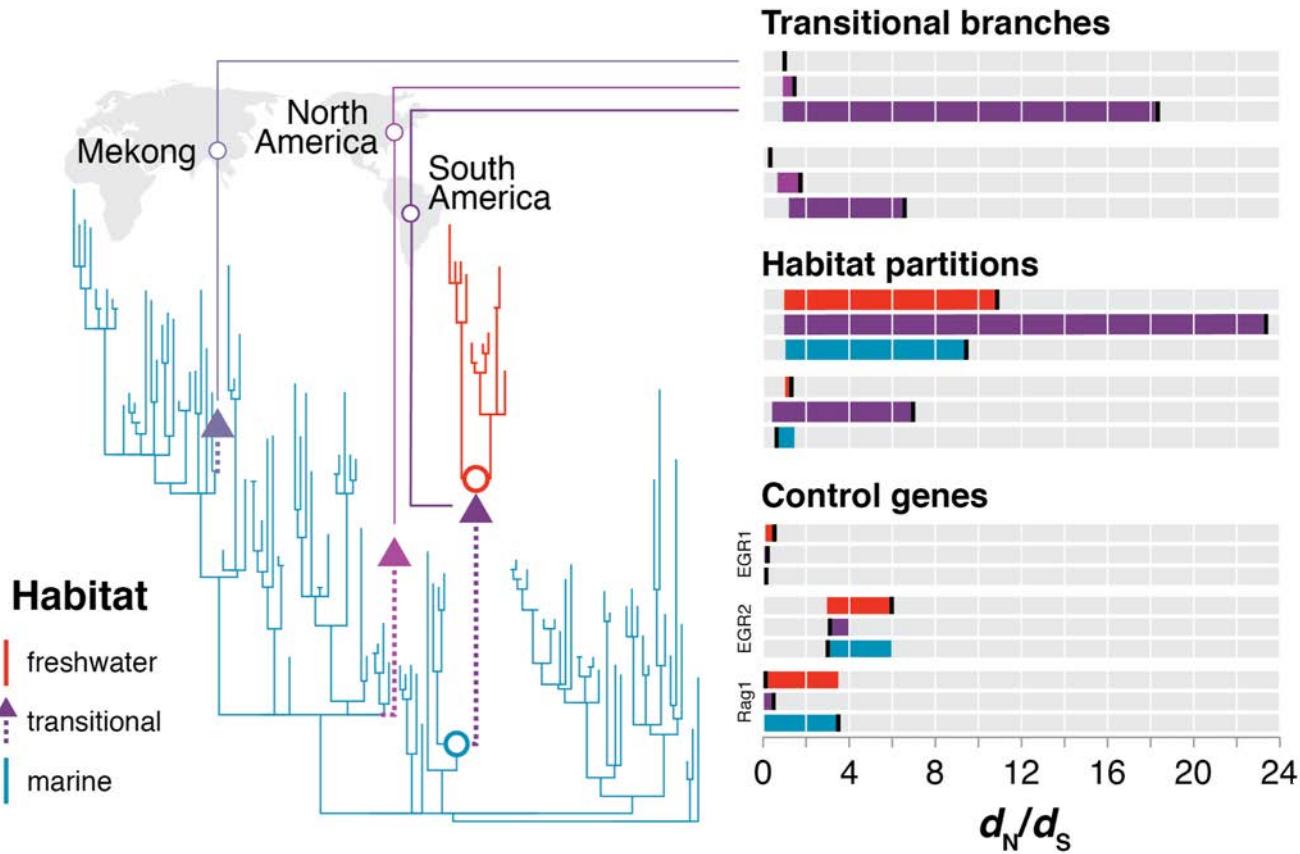
Positive selection in rhodopsin during marine to freshwater transition in South America

- Significantly higher rates of evolution on transitional branch
- Not on other transitional branches

Van Nynatten *et al.* MBE 2021
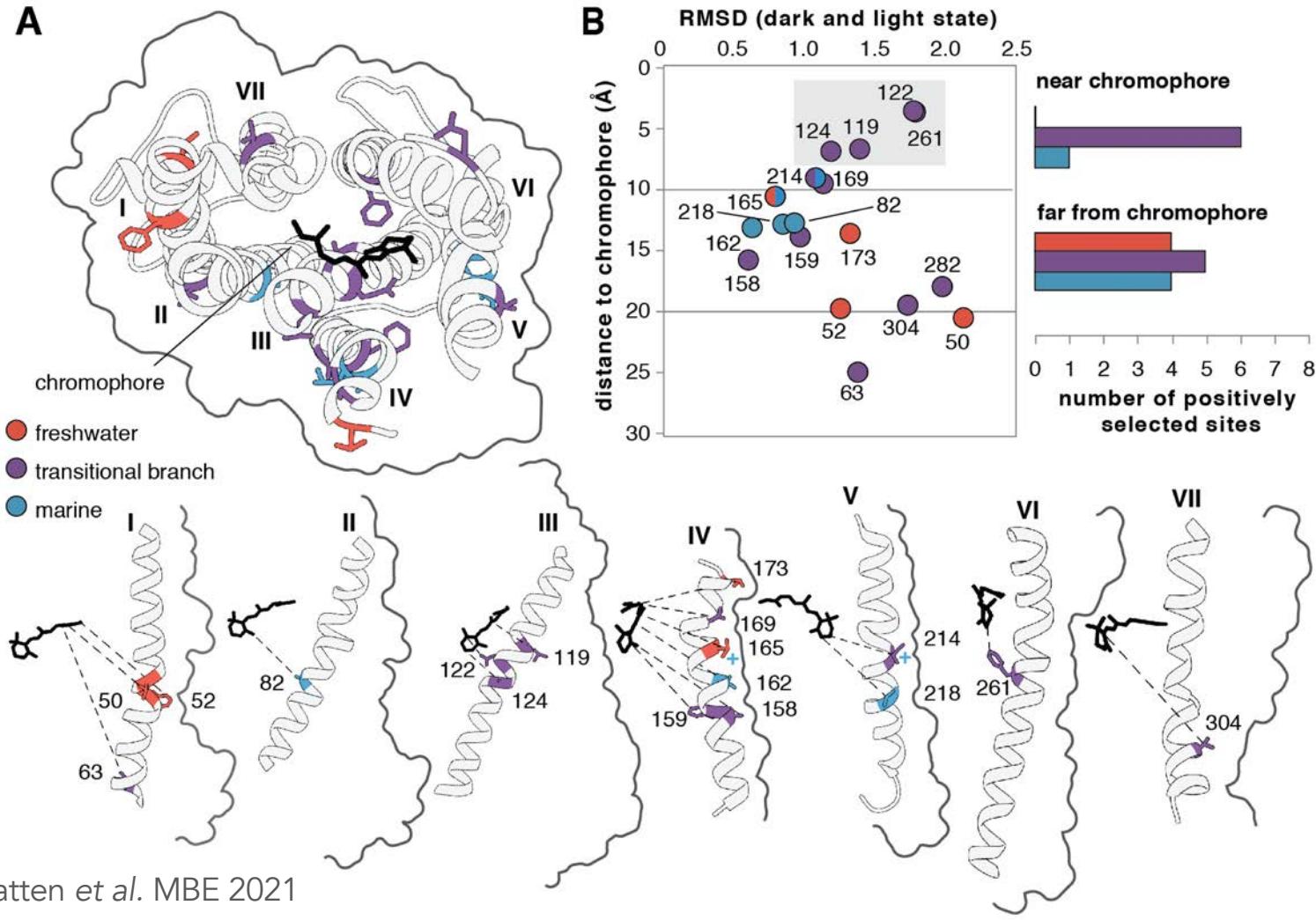
# Positive selection on all ecological partitions, but highest on transitional branch
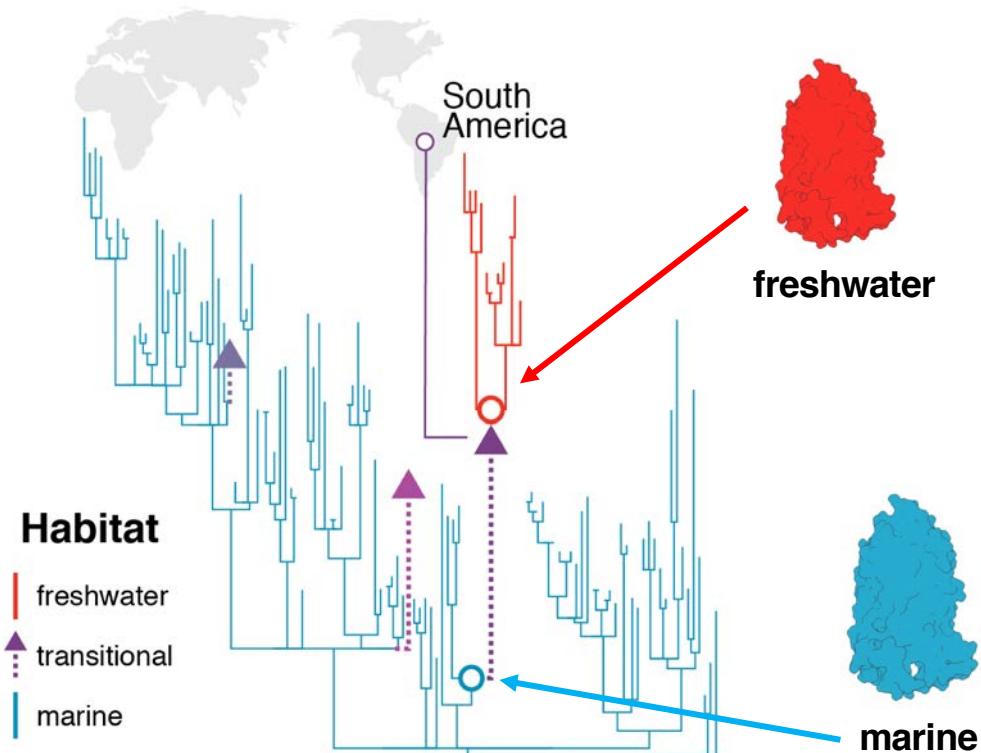
- Much higher d_N/d_S on the transitional branch than marine or freshwater clades

- No difference in control genes – seems specific to rhodopsin

Habitat
- freshwater
- transitional
- marine

Transitional branches

Habitat partitions

Control genes

EGR1

EGR2

Rag1

d_N/d_S

Van Nynatten *et al.* MBE 2021
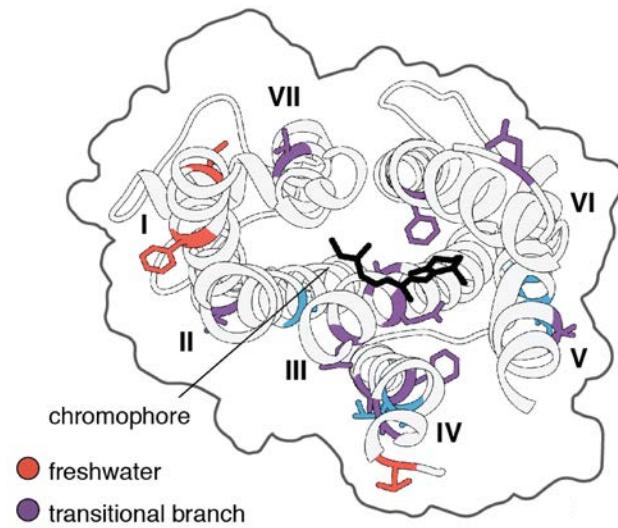
Positively selected sites are tend to be in the binding site on the transitional branch


Van Nynatten *et al.* MBE 2021
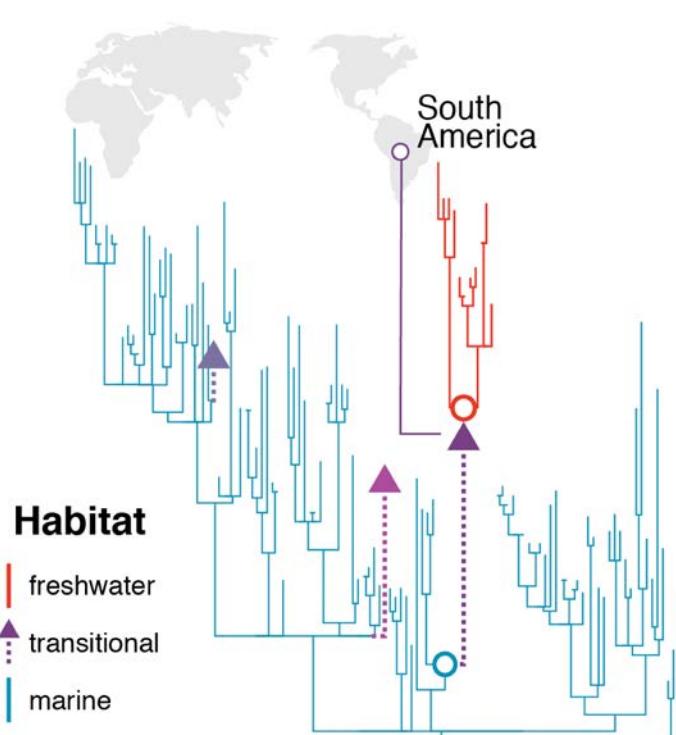
Using ancestral reconstructions to test evolutionary hypotheses of adaptation

Positively selected sites
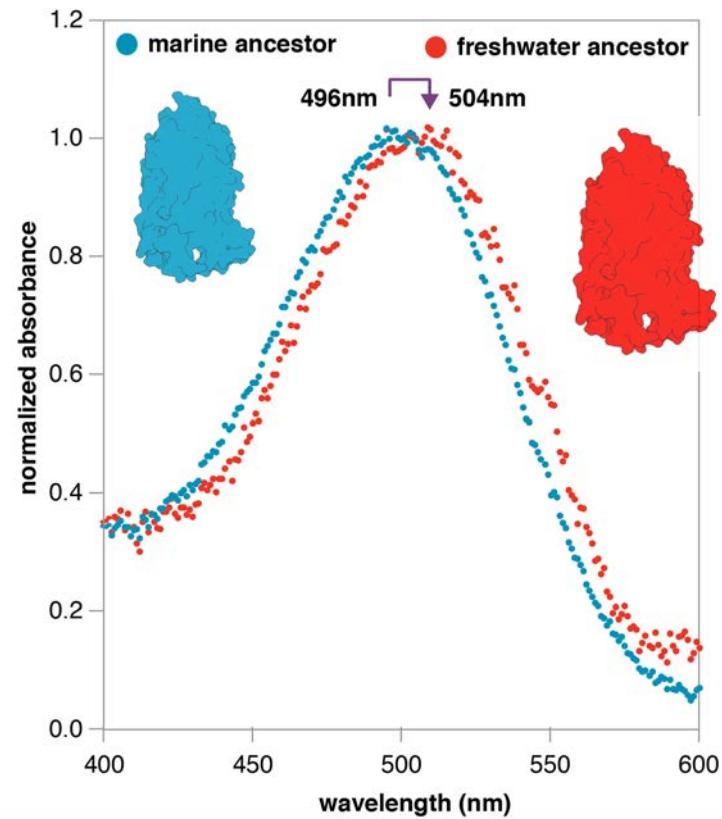

Van Nynatten *et al.* MBE 2021
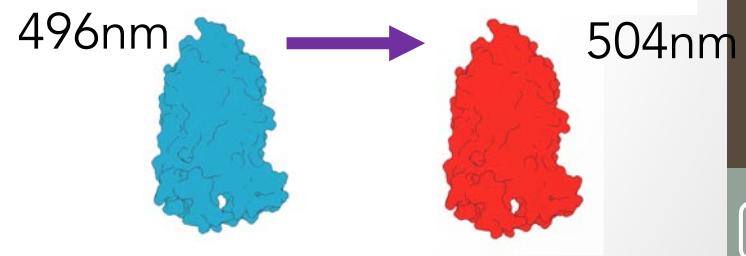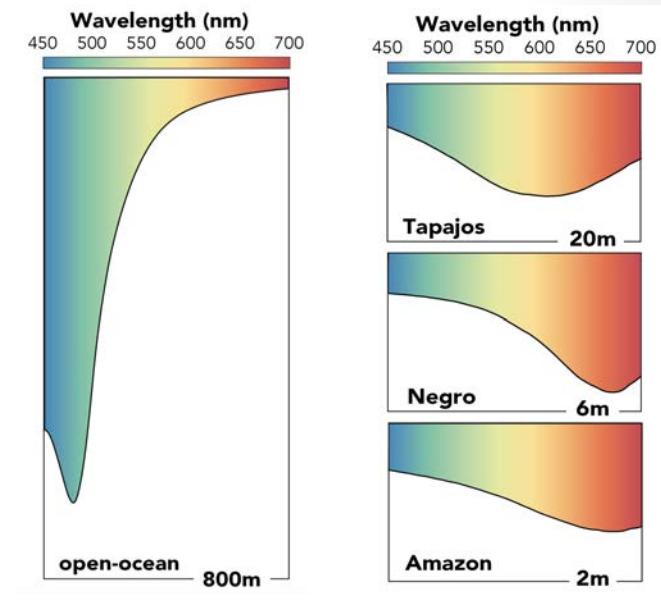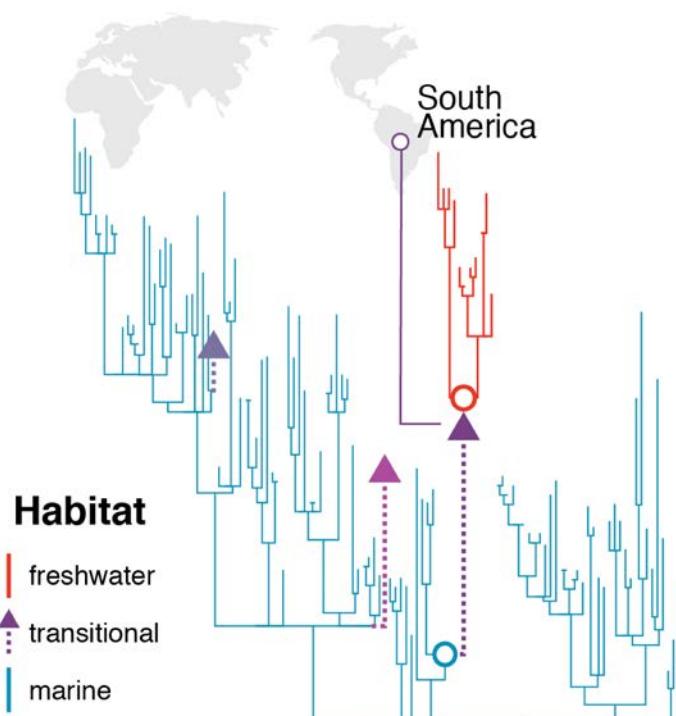
# Freshwater rhodopsin has red-shifted spectral sensitivity
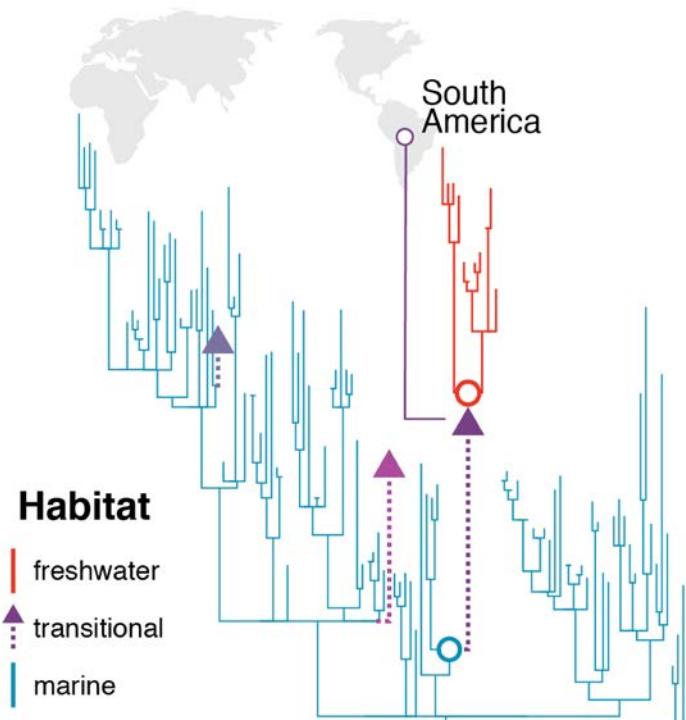

Van Nynatten *et al*. MBE 2021
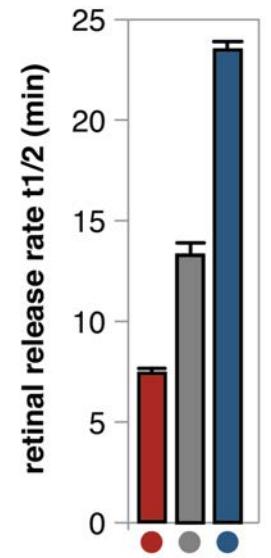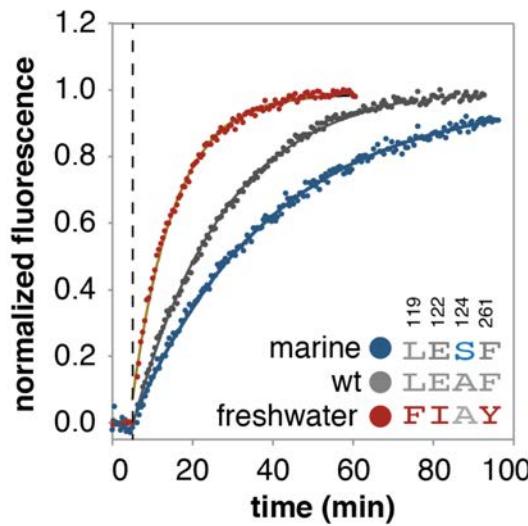
Freshwater croaker rhodopsin matches Amazonian underwater environment

496nm → 504nm

Measuring dark adaptation in rhodopsin variants


Habitat
- freshwater
- transitional
- marine
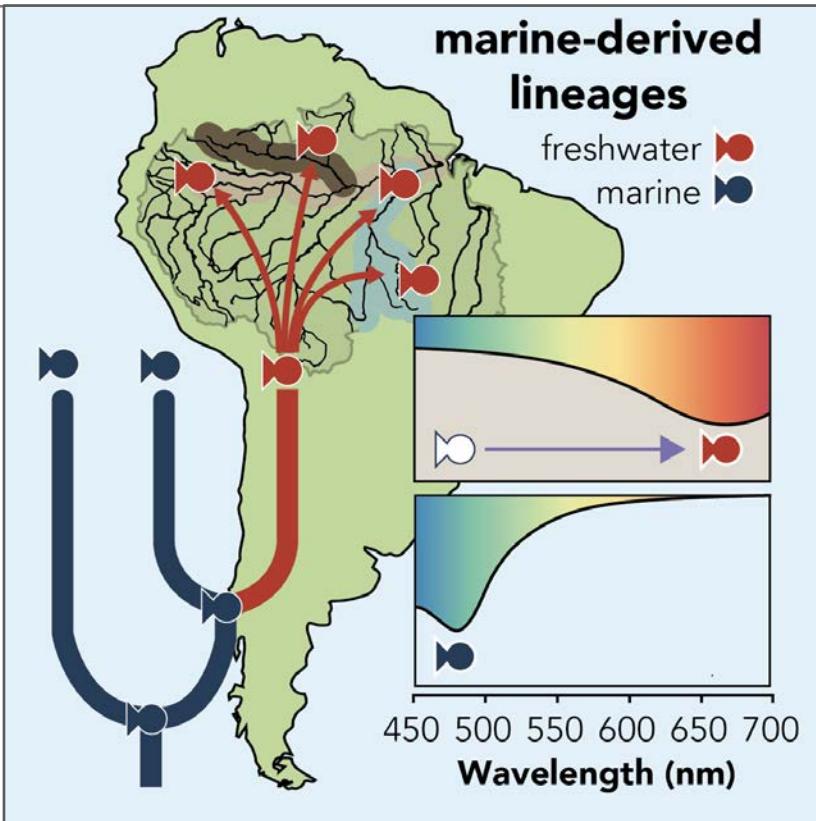
- Marine vs. freshwater variants show differences in retinal release rate
- Freshwater variant has faster kinetics, similar to cone opsins

marine ● LESF
wt ● LEAF
freshwater ● FIAY

Van Nynatten *et al.* MBE 2021
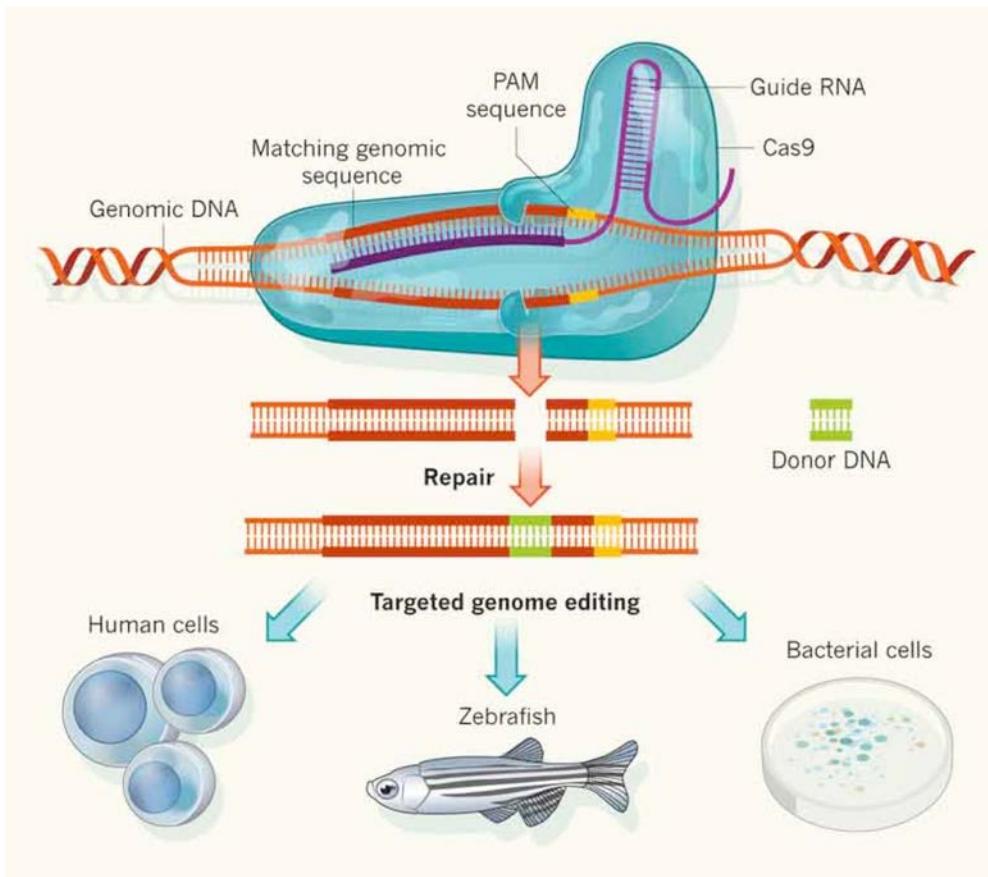
Ecological significance of faster dark adaptation in freshwater environments

High dynamic range of light intensities in freshwater necessitates rapid dark adaptation?

The future of studies in adaptive molecular evolution?

Combining computational and experimental approaches to study adaptive evolution within a hypothesis testing framework enables us to draw conclusions about adaptive molecular evolution that would not otherwise be possible with either method alone

CRISPR-Cas system

The End