Comparative Computational Analysis of Rhodopsin genes within the Animal Kingdom

By Camilo Acosta, Jayse Farrell, Jessica Kunder, Ryan Palm

April 2017

## Abstract

This study explores the prospective connection between rhodopsin gene similarity and circadian rhythm, as well as the evolutionary history of 31 species, including 8 Cephalopoda species. We performed a phylogenetic analysis of rhodopsin gene sequences that were sourced from either the National Center for Biotechnology Information (NCBI), or Uniprot. The analysis was carried out using the BioPython tool for biological computation, and alignment of the 31 genes was executed using the BLOSUM62 scoring matrix. The aligned sequences were then used to construct both a maximum parsimony phylogenetic tree as well as a Euclidean distance phylogenetic tree. Through visual comparison between the generated phylogenetic trees and either typical habitat light exposure or circadian rhythm (i.e. cephalopods and non-cephalopods, respectively), we have determined that there is not sufficient enough evidence to suggest that a correlation between rhodopsin and either of the two characteristics examined is present. It did, however, reveal unexpected relationships between particular species and their rhodopsin gene parallelism that might be worth further exploration.

## Introduction

Rhodopsin is a light sensitive receptor protein that plays a large role in the formation of images in eyes, especially in low-light environments. This protein can be found mainly in the rods within the retina where images of the visual world are developed after causing through the lens and cornea (William 2005). Its ubiquity within the animal kingdom has made it a target of numerous research attempts to better understand its evolutionary history and any differences that may have arisen in its genetics due to diversification of organisms over time and their adaptations to their environment. Due to its function within the visual system and its dependency on light levels, we decided to investigate genetic changes in the rhodopsin-coding gene in organisms that are exposed to low-light environments against those that frequently encounter increased to full UV light levels. Differences in light penetration down the water column influenced the decision to compare the rhodopsin gene from several species within the Cephalopoda class, as well as various other species across the Animalia kingdom to that of the humans. Due to the maximum absorption wavelength of rhodopsin being around 500 nm (Yau 2004), as well as the fact that the visible light spectrum does not penetrate past a depth of 300 m, one might expect there to be some sort of correlation between rhodopsin development and zone of the ocean inhabited (Pham 2012). Additionally, we analyzed animals with both diurnal and nocturnal habits, as well as a group of blind species (see the “Methods” section for a more detailed review of species used). We will answer these questions computationally and hope to shed light onto genetic changes that have occurred in the rhodopsin-coding gene and whether or not they correlate with circadian behavior and evolutionary history.

## Methods

### Data Sources

Rhodopsin gene sequences were collected from NCBI and Uniprot databases for 30 organisms, divided into two distinct categories of Cephalopoda and non-Cephalopoda species. Within the Cephalopoda group we sought to explore the relationship between the layers of the ocean inhabited by different species and the nuances of their respective rhodopsin genes. With regard to the non-Cephalopoda group, which spanned five different classes (i.e. Actinopterygii, Amphibia, Aves, Insecta, and Mammalia), we decided to investigate the relationship between differences in the rhodopsin gene and circadian rhythm type, also including a group of blind organisms.

### Rhodopsin Gene Alignments

Global alignment of all species’ rhodopsin genes were performed using the BioPython Pairwise2 Module. The Pairwise2 module implements a dynamic programming algorithm that provides functions for global and local alignments of genetic sequences (Chang 2017). The alignment uses the BLOSUM62 alignment scoring matrix with an open gap penalty of -10 and a gap extension penalty of -0.5. BLOSUM62 is the same matrix used by BLAST (Pearson 2013). Alignments were done using the *Homo sapiens* rhodopsin gene as a reference point in which all other genes were globally aligned to and then “trimmed” from the end of the aligned sequences so that all sequences were the same length for use in developing the phylogenetic trees. The length of all alignments was trimmed to a character sequence (including gap characters) of 1047 characters, the shortest observed length for an aligned rhodopsin gene within the dataset.

### Phylogenetic Analysis

For the phylogenetic reconstruction based on rhodopsin gene sequences, we used the BioPython TreeConstruction module. A distance matrix of the aligned gene sequences was created using a euclidean pairwise distance formula that utilized an ‘identity’ scoring matrix, which is the default scoring model used for both DNA and protein sequences (Chang 2017). That generated distance matrix was then used to develop a Matplotlib Distance and Parsimony phylogram. The Parsimony tree was generated using the Fitch algorithm for generating parsimony scores (Chang 2017). The phylogenetic analysis looked at 30 distinct species, 8 of which are members of the Cephalopoda class with various depth ranges they tend to reside in. The rest of the species were from various taxa with various circadian rhythms (e.g. nocturnal, diurnal) or whom are classified as blind.

### Species Analyzed

Cephalopoda

Epipelagic zone

*Octopus bimaculoides* = California two-spot octopus

> prefers tidal to subtidal depths of around 20m

*Sepia officinalis* = common cuttlefish

> 200m

*Todarodes pacificus* = Japanese flying squid

> typically 0-100m, occasionally up to 500m

Mesopelagic zone

*Doryteuthis pealeii* = Longfin inshore squid

> 0-500m

*Loligo vulgaris* = European squid

> 0-500m

*Sthenoteuthis oualaniensis* = purpleback squid

> 300-400m

*Vampyroteuthis infernalis* = vampire squid

> 600-900m

Bathypelagic zone

*Enteroctopus dofleini* = giant Pacific octopus

> immense range of intertidal zone to 2000m

Non-Cephalopoda

Diurnal

*Apis mellifera* = Western honeybee [Insecta]

*Drosophilia melangaster* = fruit fly [Insecta]

*Danio rerio* = zebrafish [Actinopterygii]

*Pan troglodytes* = common chimpanzee [Mammalia]

*Gallus gallus* = red junglefowl [Aves]

*Homo sapiens* = human [Mammalia]

*Anolis carolinensis* = Carolina anole [Reptilia]

*Athene cunicularia* = Western burrowing owl [Aves]

Nocturnal

*Apteryx australis mantelli* = brown kiwi [Aves]

*Ambystoma mexicanum* = axolotl [Amphibia]

*Xenopus tropicalis* = Western clawed frog [Amphibia]

*Heterocephalus glaber* = naked mole-rat [Mammalia]

*Mus musculus* = house mouse ("generally nocturnal") [Mammalia]

*Strix occidentalis* = spotted owl [Aves]

*Monodelphis domestica* = gray short-tailed opossum [Mammalia]

*Rattus norvegicus* = brown rat [Mammalia]

*Lithobates pipiens* = Northern leopard frog [Amphibia]

*Rhinella marina* = cane toad [Amphibia]

*Culex quinquefasciatus* = southern house mosquito [insectia]

Blind

*Astyanax mexicanus* = Mexican tetra [Actinopterygii]

*Spalax ehrenbergi* = Middle East blind mole-rat [Mammalia]

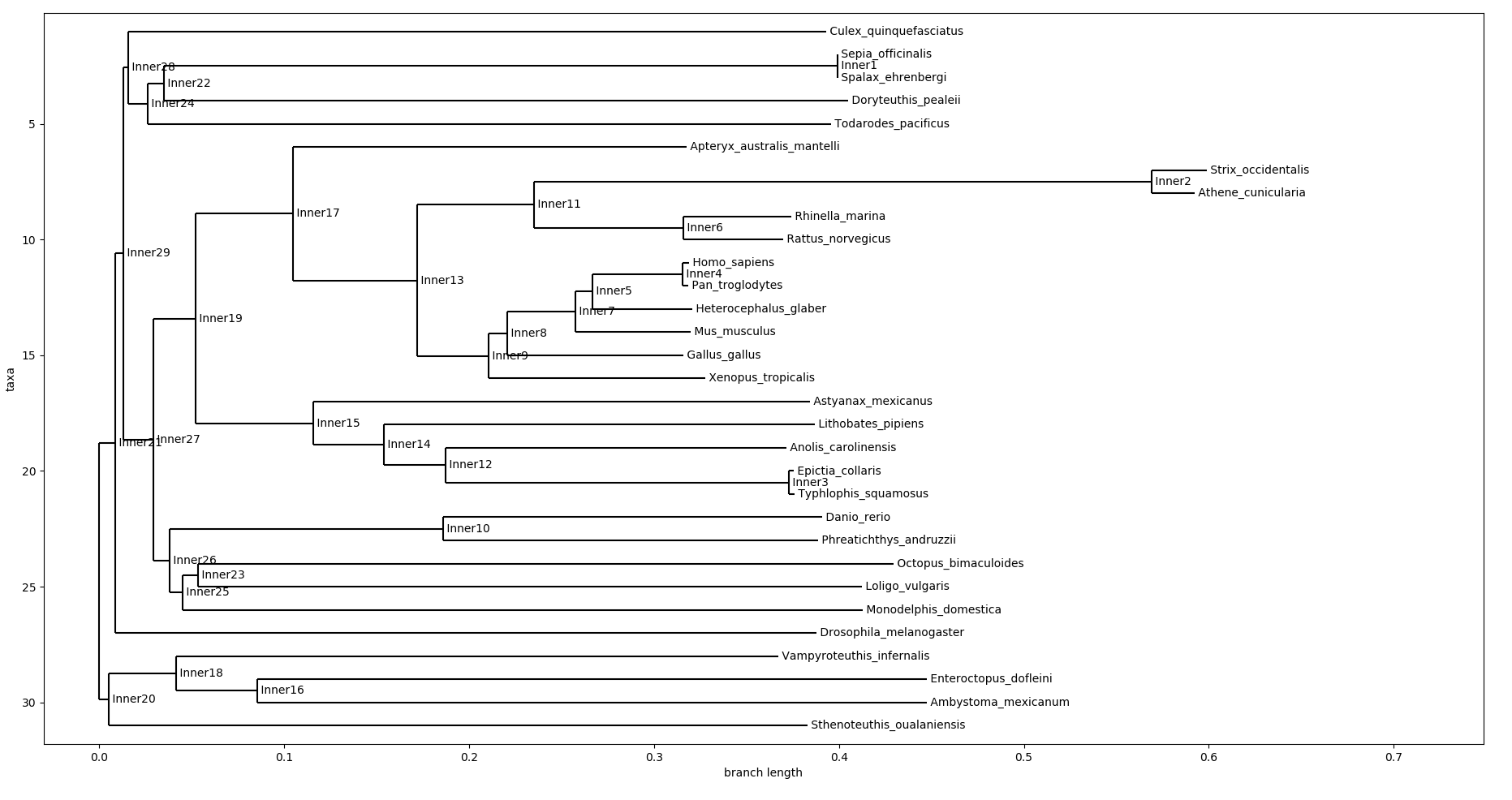
*Epictia collaris* = collared blind snake [Reptilia]

*Typhlophis squamosus* = Trinidad blind snake [Reptilia]

*Phreatichthys andruzzii* = Somalian cavefish [Actinopterygii]

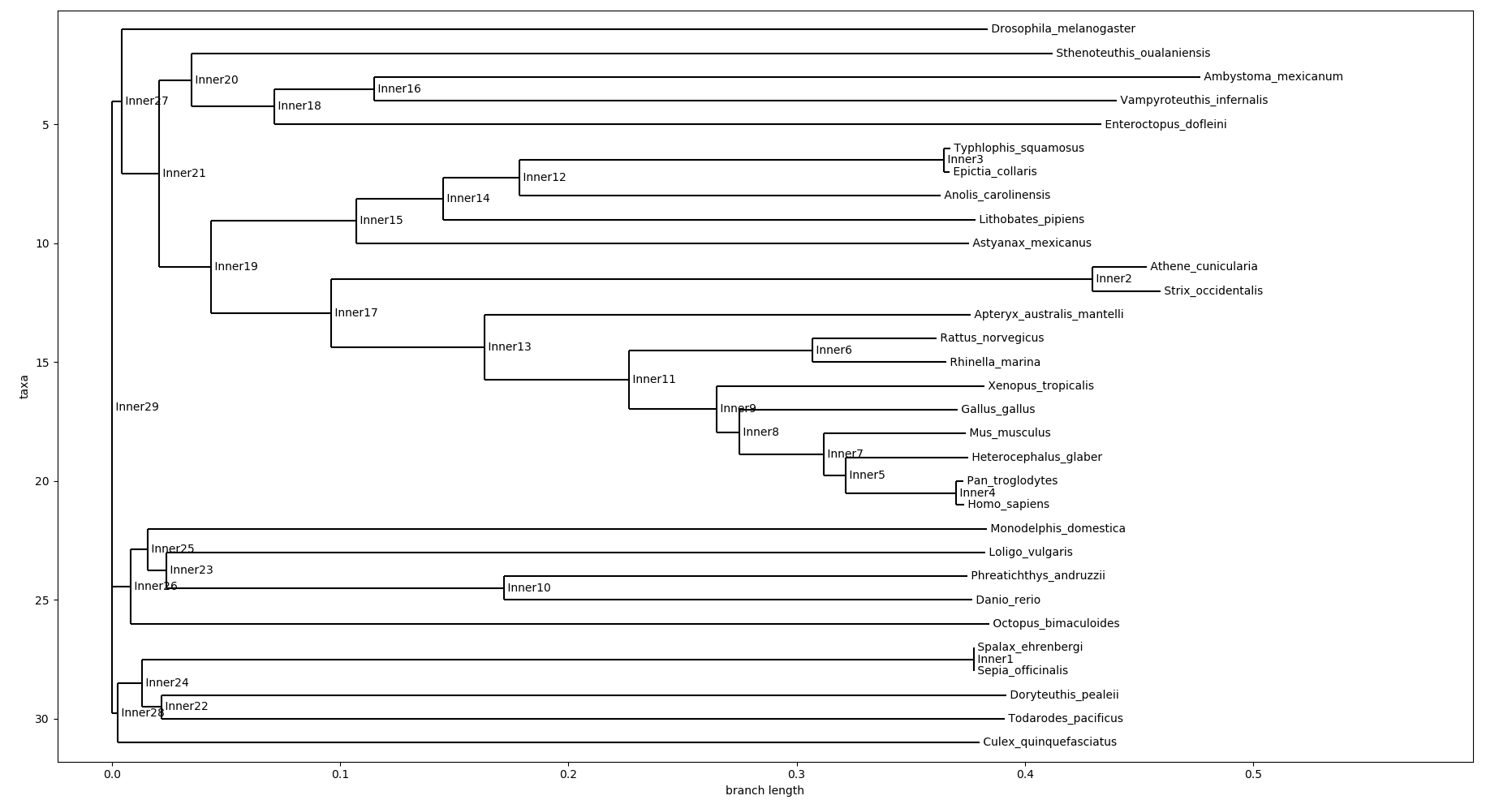
## Results & Discussion

#### Figure 1: Parsimony Tree



#### 

#### Figure 2: Euclidean Distance Phylogenetic Tree



The original goal of this analysis was to determine if genetic changes that have occurred in the rhodopsin-coding gene correlate with typical environmental light exposure and evolutionary history. Based on the analysis of 31 species, including 8 cephalopods, there appears to be no significant correlation between similarities in rhodopsin genes and circadian rhythm, at least compared to other potential links such as evolutionary distance. The result was contrary to what we were expecting, which was to see a correlation between rhodopsin gene similarity and similar sleep-wake cycles due to the gene’s role in visual phototransduction, or the converting of light into electrical signals (Rogers 2015). We believed this light-processing pigment might have accompanied the evolution of species as they adapted to best fit the light conditions of their respective habitats (e.g. a depth of 500 m into the ocean or active only during hours which allow it full UV light exposure).

The conclusions presented by this analysis relies on the use of ‘Inner nodes’, or common points that link several branches. Several Inner nodes present in the developed phylogenetic tree show the strongest evidence for this conclusion. For example, in Inner node 2 (see Fig. 2) the developed tree shows a strong relationship between two species, the spotted owl (*Strix occidentalis)* and the western burrowing owl (*Athene cunicularia)*, which have differing circadian rhythms, the former being nocturnal, while the latter is diurnal, though known to have crepuscular tendencies (Poulin and Todd 2006). These two species are, however, both members of the Strigidae family or “true owls”, with their densely-feathered bodies, front-facing eyes, and robust beaks (Parry-Jones 2004). Other tree nodes that show strong relationships between close cousins in evolutionary history are Inner node 6 (both nocturnal), Inner node 4 (both diurnal), Inner node 3 (both blind). A notable counterexample, however, is Inner node 16 which shows a relationship between the giant Pacific octopus *(Enteroctopus dofleini*), and the axolotl (*Ambystoma Mexicanum*), the former being a deep sea cephalopod, while the latter is a nocturnal salamander. The two species are not closely related on the evolutionary tree, however they both prefer low light environments. It is important to note, in this case, that while the tree shows a relationship between the two species, the two branches connecting the species are comparatively lengthy.

One node of particular interest was Inner node 1 in which the generated parsimony tree indicated a high degree of similarity between the rhodopsin genes of the naked mole-rat (*Heterocephalus glaber)* and the common cuttlefish (*Sepia officinalis).* Both of these species have “extreme” ways of interpreting light. According to a study in current Biology by Dr Shelby Temple, cuttlefish are extremely sensitive to light and have very acute polarization vision that allows them to see the direction in which light reflects off surfaces (Temple 2012). Comparatively, naked mole-rats have been shown to have higher sensitivity to white, blue, and green light waves, but a much lower sensitivity to red and UV light waves and the most remarkable aspect of their light sensitivity being their preference to low-light environments rather than high-light. Krott, a researcher in the study that determined these findings stated that the brain of a naked mole-rat only resolves a coarsely 'pixelated' image of the outside world. These two extreme examples of phototransduction and the similarity these two species share in rhodopsin gene expression might make it worth exploring this relationship further in future studies.

A larger sample of rhodopsin genes over a more broad range of species could possibly shed more light on the rhodopsin gene’s impact on circadian rhythm. A better understanding of the connection between rhodopsin genes and circadian rhythm may have been gained by the construction of a phylogenetic tree using circadian rhythm and then using it in an algorithmic comparison to the generated rhodopsin-based tree. Comparisons were, instead, done visually and backed through research. Hopefully further studies will incorporate more rigorous and thorough methods to determine if and how rhodopsin gene similarity affects behavior, as well as examining the unique points aforementioned in the previous paragraphs.

## References

Chang, Jeff, Brad Chapman, Iddo Friedberg, et al. Biopython Tutorial and Cookbook. 6 April

2017

J, Krause William (2005-07-01). Krause's Essential Human Histology for Medical Students.

Boca Raton: Universal Publishers. ISBN 9781581124682.

Kott O, Šumbera R, Němec P (2010) Light Perception in Two Strictly Subterranean Rodents

Life in the Dark or Blue?. PLOS ONE 5(7): e11810.

Parry-Jones, Jemima, and Christine Jeryan. "Owls (Strigidae)." *Grzimek's Animal Life*

*Encyclopedia*. Ed. Michael Hutchins, et al. 2nd ed. Vol. 9: Birds II. Detroit: Gale, 2004. 345-365. *Gale Virtual Reference Library*. Web. 29 Apr. 2017.

Pearson, William R. “Selecting the Right Similarity-Scoring Matrix.” Current protocols in

bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.] 43 (2013): 3.5.1–3.5.9.

PMC. Web. 26 Apr. 2017.

Pham, Long. "The Blue, the Bluer, and the Bluest Ocean." *NASA*. NASA, 8 May 2012. Web. 29

Apr. 2017.

Poulin, Ray G., and L. Danielle Rodd. "Sex and Nest Stage Differences in the Orcadian

Foraging Behaviors of Nesting Burrowing Owls." *The Condor* 108.4 (2006): 856-64.

*Elkhorn Slough Coastal Training Program*. The Cooper Orinthological Society, 7 Oct. 2014. Web. 28 Apr. 2017.

Rogers, Kara. “Rhodopsin.” Encyclopaedia Britannica. March 05, 2015. Accessed April 28,

2017

Temple S. E., Pignatelli V., Cook T., How M. J., Roberts N. W., and Marshall N. J.

High-resolution polarisation vision in a cuttlefish. *Current Biology* (Feb 21 2012). Accessed 28 April 2017

Yau, King-Wai. "Photoreceptors." *Encyclopedia of Biological Chemistry*. Ed. William J. Lennarz

and M. Daniel Lane. Vol. 3. Amsterdam: Elsevier, 2004. 326-329. *Gale Virtual Reference Library*. Web. 29 Apr. 2017.

Zhao H, Ru B, Teeling EC, Faulkes CG, Zhang S, et al. (2009) Rhodopsin Molecular Evolution

in Mammals Inhabiting Low Light Environments. PLOS ONE 4(12): e8326.

## Supplementary Data Files

The data, and source code used in this report are accessible to the public at the following GitHub repository:

<https://github.com/lightninglark/Bioinfomatics-Project-1>