UNIVERSITY OF CALIFORNIA
SANTA BARBARA

Origins, Evolution, and Integration of GPCR Photosensitivity

A DISSERTATION SUBMITTED IN PARTIAL
SATISFACTION OF THE REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHILOSOPHY
IN
ECOLOGY, EVOLUTION, AND MARINE BIOLOGY
BY
ANDREW J. M. SWAFFORD

COMMITTEE IN CHARGE:
Professor Todd H. Oakley, Chair
Professor Scott Hodges
Professor Stephen Proulx

June 2019

The Dissertation of Andrew J. M. Swafford is approved.

Professor Todd H. Oakley, Committee Chair

Professor Stephen Proulx

Professor Scott Hodges

June 2019

Origins, Evolution, and Integration of GPCR Photosensitivity

by

Andrew J. M. Swafford

Acknowledgements

I would first like to thank my family, partner, and friends -- your endless support, patience, and empathy has been the foundation upon which I have grown as both a person and a scientist. I would also like to thank the Oakley lab members, past and present, who were a wellspring of incredible ideas, improvements, and friendship. Lastly, I would like to thank my thesis advisor, Todd Oakley, for his support throughout this process and unabashed enthusiasm whichever direction my research took us.

Andrew J. M. Swafford

Education

2019 PhD Ecology, Evolution, and Marine Biology - UC Santa Barbara.
2016 MA Molecular Evolution, Phylogenetics, Bioinformatics - UC Santa Barbara.
2011 BA Biology - Bowdoin College.

Publications

2019 **Swafford, Andrew J.M.**, and T. H. Oakley., Light-Induced Stress as a Primary Evolutionary Driver of Eye Origins, Accepted Int. Comp. Biol., doi: 10.20944/preprints201904.0107.v2.

2018 **Swafford, Andrew J.M.**, and T. H. Oakley., Multimodal sensorimotor system in unicellular zoospores of a fungus, Journal of Experimental Biology, doi: 10.1242/jeb.163196.

2018 Picciani N, Kerlin, J.R., Sierra, N.C., **Swafford A.J.M.**, Ramirez, M.D., Cannon J.T., ... & Oakley T.H., Prolific origination of eyes in Cnidaria with co-option of non-visual opsins, Current Biology, doi: 10.1016/j.cub.2018.05.055.

2017 Boyero, L., Graça, M. A., Tonin, A. M., Pérez, J., **Swafford, A.J.M.**, Ferreira, V., ... & Albariño, R. J., Riparian plant litter quality increases with latitude, Scientific Reports, 7, 10562.

2016 Ramirez, M.D., Pairett, A. N., Pankey, S.M., Serb, J. M., Speiser, D. I., **Swafford, A. J.M.**, Oakley, T. H., The last common ancestor of most bilaterian animals possessed at least 9 opsins, Genome Biology and Evolution, doi: 10.1093/gbe/evw248.

2015 Lenkov, K., Lee, M. H., Lenkov, O. D., **Swafford, A.J.M.** & Fernald, R. D., Epigenetic DNA Methylation Linked to Social Dominance, PLOS ONE, doi: 10.1371/journal.pone.0144750.

Grants and Awards

2018 Faculty Minigrant - UC Santa Barbara
Funds for phylogenetics pedagogy development

2018 Undergraduate Research and Creative Activities Grant - UC Santa Barbara

2018 Ellen Schamberg Burley Graduate Award - UC Santa Barbara

2018 Summer Undergraduate Research Fellowship - UC Santa Barbara (declined)

2017 Block Grant - UC Santa Barbara
Funds for writing and research development

2017 Wake Award Honorable Mention - Society for Integrative and Comparative Biology
Given to the best student presentation in the Division of Phylogenetic and Comparative Biology

2016 Outstanding Teaching Assistant Award - UC Santa Barbara Academic Senate
Awarded to 3 TAs across all departments. Recognition for inclusive pedagogy and involvement of minorities in STEM.

2016-19 Charlotte Magnum Support Award
Society for Integrative and Comparative Biology

2014 Sigma Xi GIAR - Selected from applicants without a PI in Sigma Xi

2013 Rosemary Grant Award - Society for the Study of Evolution

Presentations

2019† **Swafford, A.J.M.**, Van De Wyngaerde, K.R., Oakley, T.H., Insights into Early Sensory Evolution from Sensorimotor Systems in Unicellular Fungus Zoospores, SICB, Presentation.
Finalist for Rising Star in Organismal Botany Award

2019 **Swafford, A.J.M.**, Oakley, T.H., Opsin Family Macroevolution and the Origin of Light Sensitivity in GPCRs, SICB, Poster.

2018† **Swafford, A.J.M.**, Oakley, T.H., Multisensory Systems in Unicellular Fungi, Gordon Conference: Cellular and Molecular Fungal Biology, Presentation.

2018† **Swafford, A.J.M.**, Van De Wyngaerde, K.R., Locker-Cameron, T, Oakley, T.H., Co-option as a driver of early multisensory integration., Gordon Conference: Cellular and Molecular Fungal Biology, Poster.

2018 **Swafford, A.J.M.**, Oakley, T.H., Multimodal Sensorimotor System in Unicellular Zoospores of a Fungus, SICB, Presentation.

2017 **Swafford, A.J.M.**, Oakley, T.H., Evolution at the Speed of Light: Opsin Family Diversity, SICB, Presentation.

2017 **Swafford, A.J.M.**, Oakley, T.H., The Origin and Tempo of Opsin Family Evolution, SSB, Presentation.

2016 Swafford, A.J.M., Oakley, T.H., Sensory Suites in Unicellular Fungus Spores, SICB, Poster. 2015 ***Swafford, A.J.M.**, *Wightman, H., Duplication and Ecology of Opsin in Select Odonates, Tree Thinkers Workshop, Presentation.

*Authors contributed equally
† Invited

Select Computer Skills & Software

Basic: OpenCV, Machine Learning (Python)
Intermediate: SQL, LATEX, OpenGL, Jekyll, Gscript, LUA, Arduino (C/C++)
Advanced: Python, R, BASH, Unix, Microsoft Windows, Galaxy, Excel/GSheets

2018 Swafford, A.J.M., KRPYGrabbler: Automated Sequence Matching and Retrieval Algorithm and Tool, Python.

2018 Swafford, A.J.M., LOIC: Iterative, Automated Gene Family Delineation, Python.

2017 Swafford, A.J.M., corView: iTOL visualization pipeline for Ancestral State Reconstruction, Python/R.

2016-8 Swafford, A.J.M., Pyphy: A Useful Collection of Python Scripts for Phylogenetic Data Management & Manipulation, Python.

2016 Swafford, A.J.M., PIPS: Open Source, Phylogenetically Informed Probe Selection for Gene-Family Specific Target Enrichment, Python/R.

2016 *Swafford, A.J.M., *Ramirez M.D., *Oakley, T.H., opSQL: Database for unified reference of opsin genes and related studies, Python/SQL.

2014 Swafford, A.J.M., Welch, Z.S., Spekit: GUI and algorithm for growth rate calculations from optical density measurements of microbes, Python.

*Authors contributed equally

Outreach & Mentorship

2019 Press Interview
Smithsonian: What Scallops' Many Eyes Can Teach Us About the Evolution of Vision

2019 Data Carpentry Instructor Training
Training to lead R, Python, Git, and Bash data carpentry workshops.

2019 "Ask an Expert Booth" SICB program
Help booth for all things phylogenetics, bioinformatics, comparative methods, and evolution.

2018 Always Open Source Initiative
Founder. Link to A.O.S. Site

2018 UCSB Open-Access Week Workshop
An Introduction to Command line, Git, and Open-Source Hosting.

2018 FUSE mentor
Teach science to parents and children in classrooms of local elementary schools; focused on involving underrepresented minorities.

2018 Van De Wynguarde, K., Swafford A.J.M., Visual Systems in Unicellular Fungi Diversify Through Co-Option of Ion Channels, Undergraduate Research Symposium, UCSB.

2017-8 Scientific Communication Workshop.
Developed and taught an open workshop on communicating science to diverse audiences.

2016-7 R coding bootcamp for Undergraduate Biologists.

2016 TALES Teaching & Learning Excellence Series: Interview

2015 Mayner, E., Swafford A.J.M., Is Sugar Killing You? The Effects of Sugar and Aspartame on the Longevity of C. elegans, 1st in California State, CA Association of Prof. Scientists.

2015 Independent Mentorship Program
Mentored two high school students in examining the effects of oil spills on the marine microbial community.

2014-8 Independent Mentorship Program
Mentor for six undergraduates from underrepresented minorities investigating the evolution of multisensory integration in fungi.

2014 Independent Mentorship Program
Mentored a high school student investigating the effects of sugar and sugar substitutes on C. elegans lifespan

2013 Research Mentorship Program
Mentored two high school students in designing and completing a research experiment on senses in fungi.

Teaching Experience

2017-8 Faculty/Instructor of Record, Research in Biology, Cornell University, SML.

2017 Guest Lecture, Guiding Principles of Macro & Microevolution, Intro. to Marine Science - UCSB
2017 Guest Lecture, Evolution of Sensory Genes and Proteins, Cell Physiology - UCSB
2016-17 Guest Lecture, An Introduction to R: Data Manipulation and Logic, Investigative Marine Biology Laboratory - Cornell, SML
2016-17 Guest Lecture, Statistics and Data Analysis in R, Investigative Marine Biology Laboratory - Cornell, SML
2012-18 Teaching Assistant, UC Santa Barbara

Courses: Invertebrate Zoology, Introductory Biology, Principles of Evolution, Macroevolution, Plant Diversity.

Summary: Teach 4 one-hour sections or 2 four-hour lab sections each week and maintaining course websites and online resources. Lecture courses rely heavily on TAs to teach the fundamentals of tree thinking, phylogenetics, and evolutionary theory.

2009-17 Teaching Assistant, Cornell University, SML

Courses: Investigative Marine Biology Lab, Marine Biology, Marine Approach to Introductory Biology, Evolution and Marine Diversity, Whales Seals and Sharks, Anatomy and Function of Marine Vertebrates, Ecology of the Marine Environment, Ecology of Animal Behavior.

Summary: Supervision of students, aid with coursework and management of online course data, grading, lectures. Write quizzes and exams, and contribute to the creation, execution, analysis, and presentation of independent research projects for each student.

Recent Employment History

2017-19 Faculty,Cornell University, SML
Contact: David Buck; (603) 862-1202
2012-19 Teaching Assistant, UC Santa Barbara
Contact: Todd Oakley; (805) 893-4715
2009-17 Teaching Assistant, Cornell University, SML
Contact: David Buck; (603) 862-1202

Recommendations/References

Prof. Todd H. Oakley Professor - UC Santa Barbara 4101 Life Sciences Building
University of California, Santa Barbara Santa Barbara, CA 93106
Oakley@lifesci.ucsb.edu

Prof. Douglas S. Fudge
Associate Professor - Chapman University 405 N. Center Street
Orange, CA 92866
fudge@chapman.edu

Dr. James A. Coyer
Affiliate Research Professor - University of New Hampshire
kelpbed@aol.com

Abstract

Origins, Evolution, and Integration of GPCR Photosensitivity

by

Andrew Swafford

Complex sensory systems such as vision shape the way organisms perceive, interact, and adapt to the environment. Of these systems, sight is a unique sense that has arisen multiple times from a deceptively small cadre of genes given the complex coordination of molecular machinery it requires. In animals, almost every instance of vision relies on a single gene family, opsin, which is widely regarded as the evolutionary lynchpin that triggered the evolution of complex eyes. However, our understanding of when opsins first originated and how the gene family has diversified throughout time remains largely unexplored. I develop new methods to reveal opsins as an ancient gene family that first appeared in early, unicellular animals. My broad synthesis of existing literature reveals that the genes crucial to vision and eye development share the common theme of mitigating UV specific stressors both in and out of eyes. Opsins' function as a sensory protein in UV-mitigation pathways highlights the multitude of effects that light cues have on organismal biology. In opposition to current hypothesis that opsins represent the only photosensitive clade of G-protein coupled receptors (GPCRs), I find evidence of photosensitive non-opsin GPCRs. Additionally, I discover a staggering amount of convergent evolution in GPCRs on specific amino acids at positions known to enable light sensitivity. Lastly, I develop the *Allomyces* fungi as a system which can be used to further study sensory system remodeling and integration after the gain and loss of sensory modalities. As a whole, my research suggests that the origins of complex traits may be rooted in the elaboration of stress mitigation and developmental capture, that the integration of light

sensitivity into organismal physiology and behavior is likely more common than currently expected, and presents a system in which we can test these hypotheses.

Contents

Chapter 1:
Introduction

Sensory systems and their integration into behavior form the foundation from which incredible diversity can arise and are the conduit through which natural selection shapes the evolution of complexity. The way organisms perceive the world fundamentally shapes how they use, adapt, and interact with their environment, which in turn creates the basis for species ranges, interspecies interactions, and ecosystems dynamics. Modifications, origins, and losses of sensory systems represent pivotal moments in evolutionary history and the processes creating these events are central to understanding our evolving world.

Of all the senses to which humankind can relate, none is more well studied than vision. Our fascination with eyes in both their function and complexity is pervasive across cultures, lexicons, and disciplines. It is this seemingly improbable complexity that first earned Darwin's comment on their evolution, starting in motion our inexorable journey towards the origin of vision:

> "To suppose that the eye with all its inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, absurd in the highest degree. [...] Reason tells me, that if numerous gradations from a simple and imperfect eye to one complex and perfect can be shown to exist, each grade being useful to its possessor, as is certainly the case; if further, the eye ever varies and the variations be inherited, as is likewise certainly the case; and if such variations should be useful to any animal under changing conditions of life, then the difficulty of believing that a perfect and complex eye could be formed by natural selection, though insuperable by our imagination, should not be considered as subversive of the theory. How a nerve comes to be sensitive to light, hardly concerns us more than how life itself originated; but I may remark that, as some of the lowest organisms, in which nerves cannot be detected, are capable of perceiving light, it does not seem impossible that certain sensitive elements in their sarcode should become aggregated and developed into nerves, endowed with this special sensibility."
>
> - Charles Darwin, *The Origin of Species*

Over a century later, Darwin's intuitions about the many intermediate forms of eyes have been validated. Relatively simple eyes have been described across the tree of life [1–3], and we now know that complex, image-forming eyes evolved multiple times within animals [4]. In

turn, our obsessive cataloging of eye diversity and development has led to a preliminary hypothesis about the process by which complex eyes evolve from simple photoreceptors [5,6] and an estimate on how long this may take [7].

Our increasingly exhaustive knowledge of the gradations in eye complexity has created a framework from which we can begin to understand the evolution of sensation at a genetic scale. The advent of genetic sequencing has unified previously disparate studies of eye diversity by identifying common molecular components which underlie vision. In the post genomic era, this accumulation of sequence data has passed a critical point where we can now begin to investigate the evolution of photosensitivity itself. Leveraging comparative methods on genes which underlie vision across the tree of life will bring into focus how sensory systems originate, evolve, and integrate into behavior. Through our studies of these pivotal moments in evolutionary history, we hope to better understand processes which guide the evolution of complexity and forge the links between sensation and organismal behavior. It is with this goal in mind that I arrive at the overarching theme of my dissertation: The molecular evolution of vision and guiding processes of sensory system macroevolution.

In chapter 2, I begin by examining the timing, tempo, and origin of opsin evolution. In animals, almost every instance of vision relies on a single photosensitive protein, opsins, which is widely regarded as the lynchpin that triggered the evolution of complex eyes. However, our understanding of when opsins first originated and how the gene family has diversified throughout time remains largely unexplored. Here we bring the timing and tempo of opsin family evolution into sharp focus, revealing opsins as an ancient gene that likely emerged in the genomes of the first animals. We produce the first fossil-calibrated time tree of opsins' evolution, which illuminates changes in their evolutionary trajectories indicating the

colonization of new adaptive space and the co-option into previously unoccupied functional niches.

In chapter 3, I propose stress as a primary driver in the origin of vision and eyes, completing the existing framework describing eye evolution. The existing hypothesis posits that selection for increased visual function alone is the primary driver behind the eye evolution. However, this leaves notable gaps in our understanding about how the first photosensitive eyespot evolved before visual tasks were possible, and why we see similar genes assimilated in almost every convergent evolution of eyes. I conduct a broad review of existing literature and find that genes now crucial to vision, eye development, and eye evolution share the common theme of mitigating or avoiding stressors created specifically from UV light exposure. Their function in light-induced stress response pathways outside of eyes suggest that their role as a transient, co-expressed protective network preceded their assimilation under developmental control to become a permanently expressed, functional unit. This synthesis completes the existing framework surrounding eye evolution by proposing light-induced stress as a driver of eye origins coupled with selection for visual function shaping eyes' subsequent elaboration.

In chapter 4, I expand the search for light sensitive proteins beyond opsins, revealing potentially unknown intersections between light, physiology, and behavior. Opsins' function as a sensory protein in UV-mitigation pathways highlights the multitude of effects that light cues have on organismal biology. Opsins belong to much larger superfamily of sensory proteins dubbed G-protein coupled receptors (GPCRs), but are regarded as the only photosensitive GPCR clade to exist. Here, we find that recently discovered photosensitive GPCRs are not opsins, and reveal an incredible amount of convergent evolution towards

potential light sensitivity across the GPCR superfamily. Our results demonstrate that light sensitivity is not as rare as previously hypothesized, and is likely more thoroughly integrated into organismal biology than we had predicted.

Finally, in chapter 5 I develop a framework which future studies can use to begin understanding how sensory systems evolve and integrate into organismal behavior. To study integration of light sensitivity into organismal biology, I turned to a simple system: the *Allomyces* genera of fungi. Here I demonstrate that these fungi, during their unicellular life stage, have gained and lost crucial sensory modalities across a very narrow taxonomic range. *Allomyces* represent one of the closest relatives of animals that integrates multiple senses to orient and inform their movement. They embody the balance of trait complexity and organismal simplicity needed to effectively study how sensory systems and behavior remodel after the gain or loss of sensory modalities. Understanding the origins, evolution, and integration of sensory systems and behavior is a multidisciplinary goal that unites researchers and is central to future studies striving to build a holistic understanding of the natural world.

Chapter 2:
The Timing and Tempo of Opsin Family Macroevolution

I. Introduction

Eyes are one of the most diverse and well-studied complex traits in the animal kingdom, with phenotypes ranging from rudimentary directional photoreceptors to camera and compound eyes capable of image forming vision [5]. Regardless of eye complexity, almost all animal vision relies on a single light-sensitive protein family: opsins [8]. Although our knowledge of opsin genes expressed specifically in eyes is steadily increasing, a large number of opsins are expressed elsewhere [9,10]. Expressed in the skin and internal organs of many animals, non-ocular opsins are often found in places where light cues may be heavily filtered, monochromatic, or not present at all. The multifaceted role of opsins, from image formation to stress-mitigation pathways [6], makes them excellent candidates to study the processes surrounding gene family macroevolution. Understanding how the timing and tempo of opsin family diversification is key to understanding the emergence processes guiding the underlying molecular evolution that gives rise to complex traits.

Current work has shown that environmental light profiles and protein function can change the complement of opsins expressed in eyes and may influence the mutations preserved in opsin genes [11–15]. We can think of the combined pressures of environmental factors and functional requirements creating a "genic environment", which defines the available niche space that existing genes and new duplicates can occupy. The requirements on opsin function imposed by organismal behavior influence opsin expression in eyes to help preserve visual accuracy [11–13]. At the molecular level, the environment also places constraints on opsins themselves, for example, opsins in deep sea fish adapted to function more efficiently under extreme pressures [14]. While these examples show that the genic environment effect can guide

microevolution, there is mounting evidence that the macroevolutionary trajectory of gene families can also be influenced as well through the preferential maintenance of duplicate genes [15].

Gene duplication is a major driver of molecular macroevolution and, by proxy, complex trait diversity. The idea of gene duplication as a process that guides molecular evolution at a macro scale was first cemented by Ohno's 1970 book, "Evolution by Gene Duplication". In this text, he describes a number of 'fates' that befall recently duplicated genes: namely non-functionalization, sub-functionalization, and neo-functionalization. There are a number of current models describing ways in which duplications allow genes to explore new adaptive landscapes and become fixed in a population [16]. However, these models largely focus on duplicate maintenance at a microevolutionary scale, leaving the processes that contribute to macroevolutionary patterns of duplicate maintenance poorly understood. The shifts in duplicate maintenance rates that occur within/between gene families is a crucial component in understanding how novel traits and complexity arise. Although our knowledge of how gene family diversification influences complex trait evolution is improving, we have largely neglected investigating the feedback between the genic environment created by traits and the underlying genes driving their evolution -- a potentially critical feature of evolutionary processes guiding diversity, novelty, and complexity.

Examining the genic environment that arises with the eye complexity required for vision reveals a system that may influence opsin family evolution at a macro-scale. As eyes evolved, their complexity increased, from directional photoreceptors to enclosed compound/camera type eyes that shield opsins from UV and excessive light[5,6]. In most eyes, opsins are exposed to a broad spectrum of visible light and used in visual tasks where non-UV

color differentiation is important, which may create a much broader adaptive landscape to explore than their non-ocular counterparts. Outside of eyes, opsins may still be exposed to the entirety of the visible spectrum, but they typically function in stress-mitigation pathways -- restricting their function to detecting damage-causing UV and short wavelength light [6,17–20]. In stress-related functions, detecting a broad spectrum of light yields the most comprehensive protection, and duplicate opsin genes with peak sensitivities shifted away from the UV wavelengths could be perceived as disadvantageous by lowering the sensitivity and magnitude of responses to stressors. Conversely, in image forming vision, subdividing available wavelengths through neo- or subfunctionalization gives an organism new and potentially advantageous information about the surrounding environment. Thus, co-option of opsins from stress mitigation into vision may lift functional constraints on opsin diversification.

In this study, we use opsins to examine the effect that a genic environment plays in guiding gene family macroevolution. We hypothesize that the processes which influence the maintenance or loss of opsin duplicates at the microevolutionary level may shape the macroevolution of the opsin gene family. By constructing a fossil-calibrated ultrametric phylogeny of opsins we quantify the timing and tempo of opsin subfamilies throughout their evolutionary history. An ultrametric phylogeny of opsins allows direct comparison with historical geologic and atmospheric data as well as existing ultrametric phylogenies of animal evolution. Sampling from completely sequenced genomes across all animals allows us a previously unattainable understanding about the genic environment surrounding the origin of crucial opsin clades and how rates of duplicate gene maintenance have shifted across the opsin tree. We find that the first ancestral opsin likely arose in the early Cryogenian and that

organismal complexity and functional demand which define microevolutionary fates of duplicate genes may influence gene family evolution on a macro scale.

II. Methods

Data aggregation

We first compiled a list of organisms with sequenced genomes from NCBI, limiting our initial search to metazoa. We then removed all organisms with less than 10,000 predicted proteins for their genome, further refining this list by removing all but the two species in each genus that had the most predicted proteins. The final list consisted of 246 species across 13 Phyla. We then compiled a bait file of opsin sequences, including representatives from each notable opsin subfamily [8] with as broad a taxonomic representation as possible. We then designed and used a sequence discovery, ranking, inventory and value estimation script (SeDRIVE) to aggregate opsins from candidate species. SeDRIVE uses BLAST[21–23] to search each candidate for matches above a user defined bitscore, automatically generating individual storage files for each candidate species. We used a bitscore cutoff of 210. Next, the user can invoke customized filter steps using SeDRIVE. We performed a custom removal for the following keywords: octopamine, chemokine, somatostatin, adrenergic, and hGC as they are not part of the ingroup and were retained due to similarities to our melatonin outgroup. In addition, we removed all sequence fragments less than sixty amino acids long to avoid protein fragments. We then batch executed CD-HIT through SeDRIVE to remove sequences of 98% similarity or higher from each candidate species. Lastly, we removed isoforms after initial tree building, to better account for splice variants occupying different positions in a functional landscape. The final dataset after isoform removal and filtering scripts totaled 4,482 sequences (Table S1).

Phylogenetic reconstruction & dating

Although we observed best practices, our options for reconstruction were limited due to the size of the dataset. Sequences were aligned using SATe II [24] with under the gtr model. The tree was reconstructed under the LG4X model using IQtree [25,26] with approximate bayes test [27] and SH-like approximate likelihood ratio test [28] with 1000 bootstrap replicates. The initial tree was used to identify and remove all but one isoform from the same organism that were each other's closest relatives. After the removal of isoforms, the original sequence set was trimmed using the supercuts script [8]. This final set of sequences was first put through the LOIC script (https://bitbucket.org/swafford/loic). LOIC is an iterative script that removes sequences which fall outside a particular group. LOIC uses RAxML read-placement to identify sequences that fall within a known outgroup in a precomputed tree. We used the opsin tree from Porter et al. 2012 together with the landmarks from the PIA tool [29] to remove spurious sequences that did not fall within opsins. A robust set of melatonin receptors were added as an outgroup [8,30] as they currently appear to be part of a clade that is opsins' closest relative. This final dataset was then re-aligned and reconstructed into a phylogeny using the methods detailed above once again.

We curated a set of constraints for important speciation events from fossils in the current literature (Table S2) [31–52], using this as input for the constraint application on massive phylogenies (CAMP) script. In short, CAMP takes fossils constraints then identifies all possible descendants of each daughter of the most recent common ancestor (MRCA) represented by the fossil using either a user specified or, in the case of this publication, the NCBI common taxonomy tree. It then iterates over a gene tree with appropriately labelled tips, identifies all internal nodes which represent both orthologous nodes (predicted via ete3 [53]) and

the fossil MRCA, and automatically creates the appropriate r8s, treePL, or phylobayes, or PATHd8 constraint file output at the users discretion. Through this method, we applied constraints to 582 nodes on our final tree. Due to the size of this tree, we used TreePL [54] and PATHd8 [55] to generate to calculate divergence time estimates on each node with automatically calculated rate smoothing and 500 subsampled alignments to create confidence intervals and summary statistics for each node estimate. Because TreePL is more accurate than PATHd8 [54], the output from TreePL on the topology from the unresampled alignment was used in all analyses downstream. Because many analyses downstream do not take confidence intervals, no single tree fits within every confidence interval (CI) for every node, thus we use the tree built from the most complete dataset in analyses but also report CI's and average times from a "consensus tree" built from the average of bootstrap ages at node. Statistics were done using several in-house scripts available in the bitbucket repo for this publication (https://swafford@bitbucket.org/swafford/opsin_timetree.git), which includes a fully annotated and searchable consensus tree. Additionally, we created a spreadsheet with the min/max values for each node, available in the supplement (Table S3).

Lineage over time modeling

To investigate the tempo of opsin evolution, and possible models of evolution opsins and sub-clades may be following, we constructed lineage through time plots for several focal clades. We extracted the subtrees from our unresampled TreePL output and imported them into R. Using the *ltt* command from the *ape* package, we created lineage through time plots of each subtree. We used the *gammaStat()* function to test our hypothesis through a one-tailed z-test to determine if the model of evolution that best describes the accumulation of lineages over time

was different from a null model of linear accumulation. We merged, cleaned, and recolored plots for figures in Adobe Illustrator.

III. Results

The most recent common ancestor of opsins originated in the Cryogenian, about 828-822 million years ago.

From our consensus tree, we see that the most recent common ancestor of all opsins arose around 825 million years ago. Calculating the 95% confidence interval of all 500 bootstraps places this origin between 822-828 million years ago, with a oldest-youngest range between 876-677 million years ago.

Visual opsins arose in the rhabdomeric clade first, though the both R and C-opsins are equally old.

Our consensus tree reveals that the rhabdomeric and ciliary opsins both arose around 821 million years ago, with the 95% CI placing the origin between 824-818 million years and a oldest estimated age of 874.5 and youngest estimated age of 670 million years ago. However, R and C opsin clades recruited into extant eyes for visual processes show more asymmetrical origins, with visual R opsins appearing at ~762MYA (95% CI 758-766, max:833 min:608) and visual C opsins appearing ~ 645MYA (95% CI 648-642, max:728 min:535).

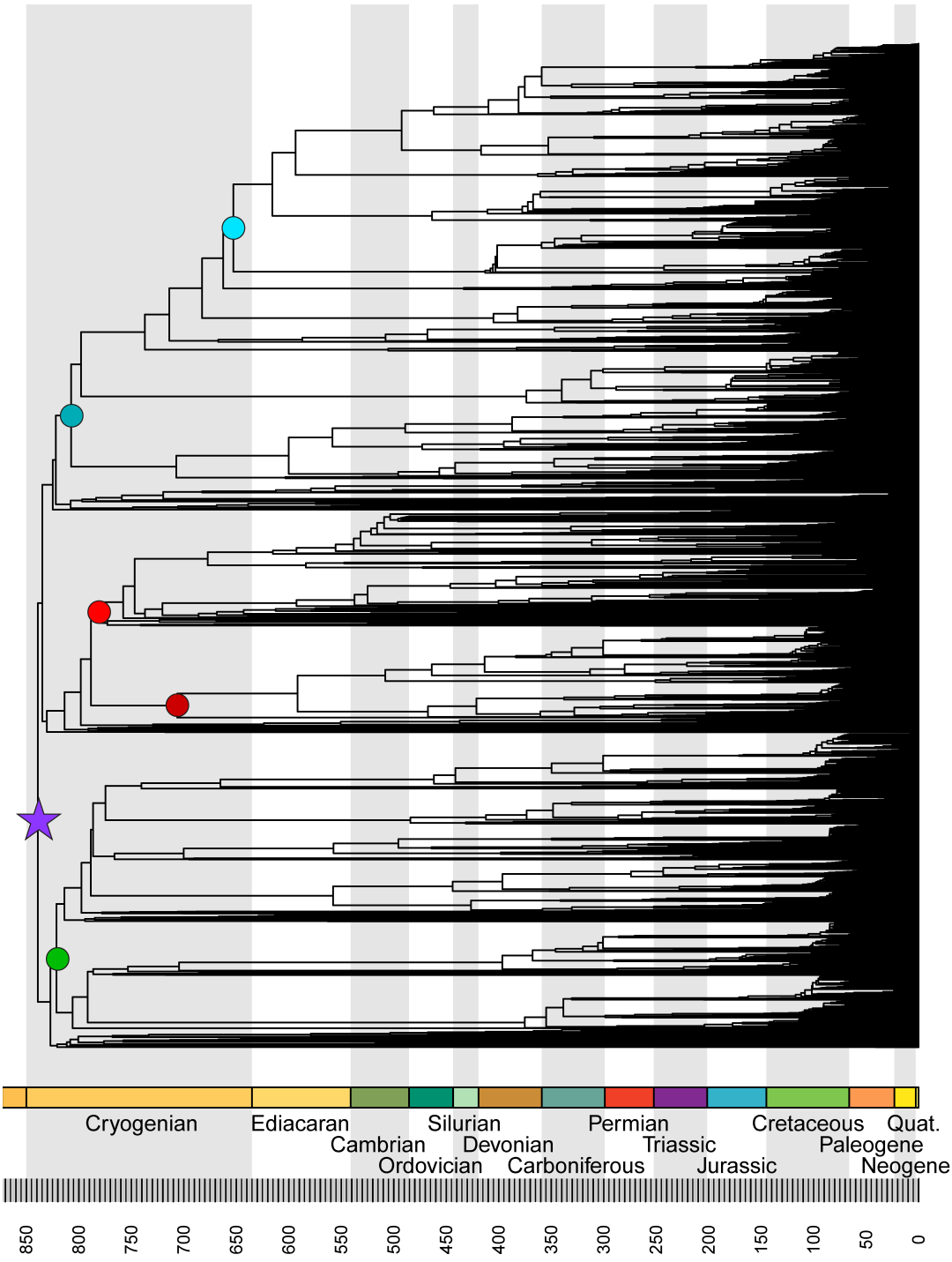

Figure 1. A ultrametric phylogeny of 4,482 opsins spanning the entire opsin family. We recover the origin of all opsins (**star**) at ~825MYA. Particular clades of interest are also marked: (**Green**) Non-visual group 4 opsins including RGR, Go, and Neuropsin clades. (**Red**) R opsins showing the origin of visual R opsins (**bright**) used by invertebrates and the surviving origin of non-visual R opsins (**muted**) found largely throughout vertebrates. (**Cyan**) C opsins showing the origin of visual C opsins (**bright**) used largely in the eyes of vertebrates, and the non-visual C opsins (**muted**), which include pinopsin, vertebrate ancient-long opsins, and a number of invertebrate C-like opsin sequences.

Opsin subclades follow different models of evolution.

We calculated gamma values on a number of subclades as well as the entire opsin family. Gamma values are an indicator of lineage accumulation, with positive values indicating that lineage expansion has occurred closer to the root than expected, 0.0 values indicating regular, exponential accumulation, and negative values indicating lineage expansion closer to the tips. The entire opsin family has a gamma statistic of +10.95, significantly deviating from the null hypothesis of a linear increase in duplicates ($P<0.0001$). By subdividing the opsin family tree into visual R, visual C, non-visual R, and non-visual C, we find differences in the patterns of evolution emerge. Visual R opsins follow an early burst model (gamma: +2.94, $P = 0.001$) while non-visual R opsins show evidence of high rates of extinction early on (gamma: -1.59), but barely fail to reject the null hypothesis of linear accumulation ($P = 0.055$). Visual C (gamma: -0.24, $P > 0.1$) opsins fit the model of linear accumulation well, though their non-visual counterparts show significant evidence of high rates of extinction early in their evolutionary history (gamma: -1.11, $P > 0.1$). The non-visual group 4 opsins show an incredibly robust signal for following an early burst model (gamm: 6.56, $P < 0.00001$).

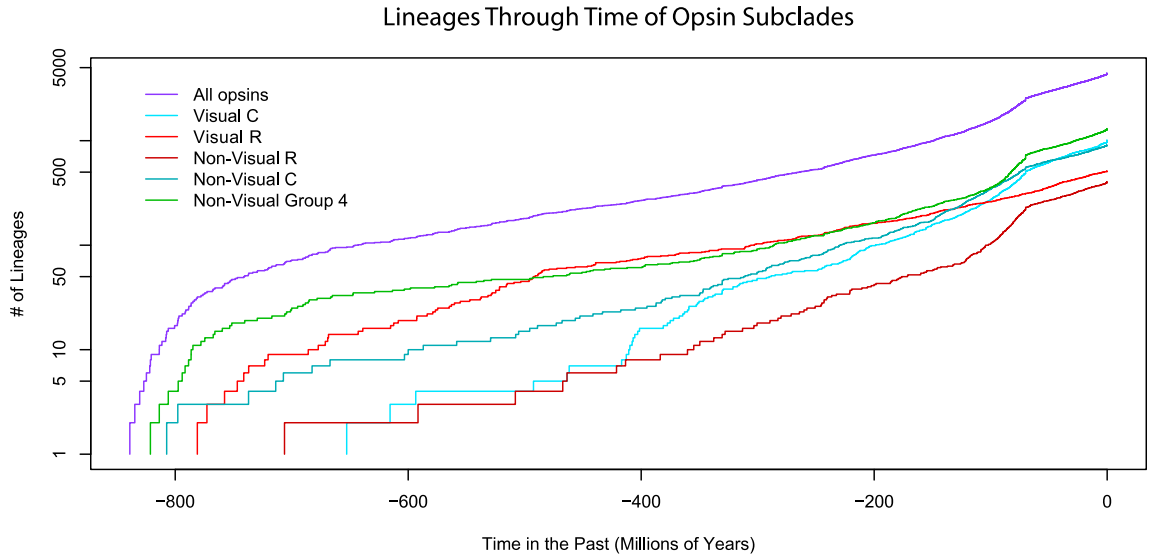

Figure 2. Lineage through time plots of opsin subfamilies. (**Purple**) Duplicate accumulation of all opsins since their origin strongly follows an early burst model. (**Cyan**) Duplicate accumulation in C-opsins since their origin, showing accumulation of visual (**bright**) opsins follows a linear model while non-visual (**muted**) opsins show evidence of high rates of gene loss early on. (**Red**) Duplicate accumulation in R-opsins showing that visual R (**bright**) strongly follow an early burst model, while non-visual R (**muted**) show evidence of high rates of gene loss followed by linear accumulation. (**Green**) Duplicate accumulation in non-visual group 4 opsins strongly follow an early burst model.

IV. Discussion

The earliest known fossils of eyes appear in the Cambrian, approximately 535-520 million years ago [56]. While it is reasonable to expect opsins to be older than this first occurrence, eyes theoretically require very little evolutionary time to reach the levels of complexity observed in the fossil record [7]. We find that opsins are much older than would be previously anticipated based on fossil evidence and relative presence/absence comparisons [8]. First appearing ~825 million years ago (MAT), the first opsins would have arisen in organisms living in the early Cryogenian period, during one of the most extreme glaciation events on earth. While the lack of fossilizable tissues and distinct chemical signatures make hard to form a satisfactory hypothesis of what organisms these first opsins may have appeared in, we can be reasonably sure they resembled unicellular, planktonic organisms not unlike choanoflagellates or zoospores alive today [57].

Regardless of its morphology, the common ancestor of Eumetazoans likely already possessed or evolved opsins before further diversification. Recent studies have placed the most recent common ancestor (MRCA) of all metazoans at approximately 833.5-649.8 MYA, overlapping well with our min-max estimates of opsin origins at 878-677 MYA [35]. We believe our results support opsins first emerging in eumetazoan ancestors because even though our estimates place opsins fairly confidently in the 830-815 MYA range, opsins are not found in either Placozoa or Porifera, two groups which diversified prior to Eumetazoa. The selective

drive to conserve opsins throughout evolutionary time appears strong enough to have ensured at least a single copy in nearly every eumetazoan, hinting that the true origin of opsins may be at the younger end of the recovered range while the true origin of crown Eumetazoans may be at the older end of their predicted range.

The expansion of opsin clades used in vision preceded the first evidence of eyes by over a hundred million years. Although we find that R and C opsins originated at about the same time (~821 MYA) the visual clades of these opsins began diversifying at different times. There is an approximately 117 million year gap between the origin of the first visual R opsin and the origin of the first visual C opsin (Fig 1), with the first fossil evidence of eyes occurring an additional ~115 million years after that. By the time eyes were first fossilized, both clades had begun to diversify, although visual C-opsins show a much lower initial rate of duplicate retention. This suggests that the processes initially shaping the maintenance of opsin duplicates for hundreds of millions of years was likely independent of their ability to function in a purely visual context and functionality in an eye as proposed in Nilsson, 2013 [5].

The complexity of these guiding processes is mirrored in the varying models of evolution which describe early duplicate accumulation and maintenance in opsins. We show that visual R and non-visual group 4 opsins strongly follow an early burst model, while all C-opsins and non-visual R opsins follow either a linear growth model or show evidence of high loss rates early in their history (Fig 2). There are a number of factors which we believe may influence the diversification of opsins through time, most notably the light environment opsins are exposed to and the role in which opsins function at the organismal level. Group 4 opsins are the result of high levels of duplicate maintenance soon after opsins' appearance. The group 4 opsins consist largely of RGR, Go, and Neuropsin subclades. These clades all detect UV

light, and are all involved in regulating highly conserved light-induced stress mitigation behaviors and pathways like the circadian rhythm. We find it compelling that, the large majority of R opsins are UV and short wavelength sensitive, suggesting it is likely an ancestral function which may have been selected for eventual co-option into vision [58].

Although shifts in peak wavelength sensitivity may have been selected against, maintaining additional duplicates with the same function is an evolutionary 'strategy' that can result in quickly increasing expression levels [59]. In opsins, initial retention of these duplicates driven by the selective pressure to increase UV sensitivity may have provided the basis for subfunctionalization as eyes and selection for non-UV photoreception emerged. Further synthesis reconstructing the ancestral sensitivity of opsins and their role in UV stress mitigation is needed to assess these hypotheses.

It is worth noting that there is a correlation between subfamilies which follow early burst models of evolution and the phyla they are present in today. With the exception of group 4 opsins, all subclades we tested that did not follow early burst models are found primarily in chordates. Although the existing dataset was skewed heavily towards sampling from chordate and arthropod genomes, the connection between phyla and evolutionary model presents an opportunity for future research. Interestingly, this division also highlights a difference rooted in the ancestral mechanisms used by opsins and their interaction with the genic environment at the time of divergence. All opsin proteins function exclusively as either a bistable (R-opsins) or a unistable (most C-opsins) light sensors. Bistable opsins regenerate by absorbing light at a different wavelength, functioning as standalone light sensor, depth sensor, and photon stoichiometer able to detect relative levels of two wavelengths [60]. Unistable opsins must rely on several additional proteins to regenerate light sensitivity, one such protein is a UV sensitive,

bistable, group 4 opsin: RGR [60]. Of the two, bistable opsins are less complicated and may inherently be useful at a unicellular level [61]. The ancestral opsin is predicted to be bistable [58], suggesting that the molecular support network required to support C-opsin functionality may only have evolved in chordates, harnessing the large number of UV sensitive RGR opsin copies to regenerate the substrate consumed by C-opsin photoreception. It is possible that in the time before this exaptation occurred, duplicate copies of C-opsins and non-visual R opsins present in the early ancestors of vertebrates may have been easily lost because visual R-opsins performed more efficiently.

In summary, we find that opsins are a much older gene family than previously anticipated and follow diverse tempos of evolution across the phylogeny. We place the origin of opsins in the early cryogenian where they quickly accumulate duplicate genes, as expected for a lineage exploring new adaptive space. We find particular clades of opsins follow this pattern as well, namely UV sensitive opsins and those which are now the basis of vision in arthropod/invertebrates, a lineage whose ancestors show the first fossilized evidence of complex eyes. We make particular note of the sudden increase in visual C-opsin diversity at the end of the silurian, around 419-390 million years ago. Although we have no satisfying explanation for this deviation from a linear model, it stands out as a potential area for future inquiry. Overall, our results reveal the timing and tempo of opsin evolution and suggest that processes which influence the maintenance or loss of gene duplicates at the microevolutionary level may shape the macroevolutionary trajectory of gene families.

Chapter 3:
Light-induced stress as a primary evolutionary driver of eye origins

doi: 10.20944/preprints201904.0107.v2

Abstract

Eyes are quintessential complex traits and our understanding of their evolution guides models of trait evolution in general. A long-standing account of eye evolution argues natural selection favors morphological variations that allow increased functionality for sensing light [4,5,7,62]. While certainly true in part, this focus on visual performance does not entirely explain why diffuse photosensitivity persists even after eyes evolve, or why eyes evolved many times, each time using similar building blocks. Here we briefly review a vast literature indicating most genetic components of eyes historically responded to stress caused directly by light, including UV damage of DNA, oxidative stress, and production of aldehydes. We propose light-induced stress had a direct and prominent role in the evolution of eyes by bringing together genes to repair and prevent damage from light-stress, both before and during the evolution of eyes themselves. Stress-repair and stress-prevention genes were perhaps originally deployed as plastic responses to light and/or as beneficial mutations genetically driving expression where light was prominent. These stress-response genes sense, shield, and refract light but only as reactions to ongoing light stress. Once under regulatory-genetic control, they could be expressed before light stress appeared, evolve as a module, and be influenced by natural selection to increase functionality for sensing light, ultimately leading to complex eyes and behaviors. Recognizing the potentially prominent role of stress in eye evolution invites discussions of plasticity and assimilation and provides a hypothesis for why similar genes are repeatedly used in convergent eyes. Broadening the drivers of eye evolution encourages consideration of multi-faceted mechanisms of plasticity/assimilation and mutation/selection for complex novelties and innovations in general.

I. Introduction

Understanding the evolutionary origins of complex structures and innovative functions are foundational goals of biology. Because we know a lot about their structure-function relationships and genetics, eyes serve as models for understanding complex trait evolution [63]. A commonly accepted explanation for eye evolution is that natural selection acted on variation in morphology, in turn increasing functional capabilities of photoreceptors, simple eyes, and lens-eyes [5,64]. Even though it explains many aspects of eye evolution, here we discuss how this focus on visual function is incomplete and leaves substantial features of eye evolution unexplained. Furthermore, recent work—mainly on other complex traits—increasingly examines how multiple engines of novelty, especially plasticity, contribute to the complexity of life [65–68]. We extend causes of eye evolution beyond visual function to include stress responses as another critical driver. This focus on stress invites discussions about the relative roles of plasticity and mutation in the origins and elaboration of eyes and other complex traits.

The diversity of animal photoreception can be approximated by four rough, functional categories describing a stepwise increase in complexity of both form and function [5]. First, non-directional photoreceptors use only photosensitive cells, which can only measure the intensity of ambient light for use in multiple behaviors. Second, directional photoreceptors pair light-blocking pigments with photosensitive proteins. Third, low-resolution spatial vision usually uses lens-like material to begin to focus light on the retina. Finally, high-resolution spatial vision uses a lens to focus light finely and precisely onto photoreceptors. This framework hypothesizes that transitions between these morphological categories may be driven by natural selection on morphologies that allow increasingly complex behaviors [4,5,62]. In fact, explicit calculations of the physical and optical requirements for the different sensory tasks match well

with physical capabilities inferred from the increasingly complex morphologies [5]. These functional categories also hold outside animals, in various single-celled organisms [61,69,70]

While the gradual elaboration of eyes may be explained by selection on visual function, it is unclear how each of these parts originated before selective pressures to refine visual acuity could shape their evolutionary trajectories [71]. Besides gradual modifications like deepening of pigment cups and elaboration of lenses, the complexification of eyes required discrete steps, including origins of photoreception, origins of pigmentation adjacent to photoreceptors, and origins of lens-like material in the path of light. While these discrete origins could be explained by purely random mutations that direct expression of components to evolving eyes, the randomness of this mutation-selection model does not account for some salient features of eye evolution. First, despite evolving many times separately, eyes use functionally similar components [4,72]. Second, genetic components of eyes often have dual or ancestral roles in responding to stress. Third, for fitness based on visual function to be dramatically higher, morphological novelties should be paired with behavioral innovations, and it may be unlikely that compatible and purely random mutations in both behavior and vision would occur simultaneously. Here we summarize the role of stress in evolution of the components of eyes and we hypothesize that responses to stress induced by light was an instrumental force in the evolution of eyes, especially in discrete origins of lenses, eye-pigments, and photoreception.

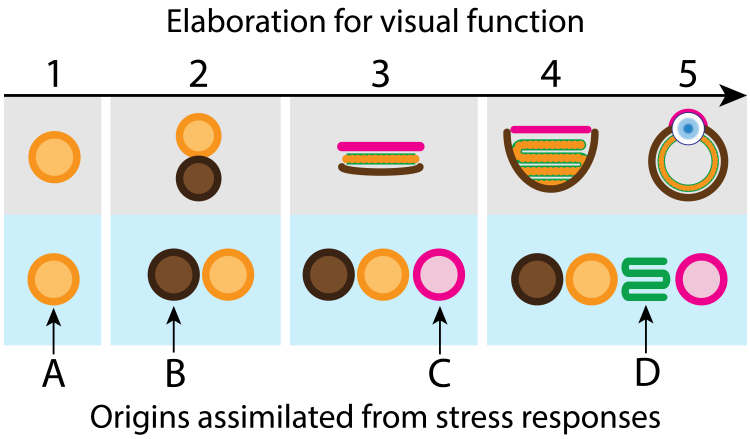

Figure 1. We propose a model of eye evolution with origins rooted in responses to light-induced stress. This builds upon previous ideas (**grey background**) that natural selection for increased visual function is the main driver of eye evolution beginning at (**1**) general photoreception and moving through (**2**) directional photoreception, (**3**) high sensitivity directional photoreception, (**4**) low-resolution image forming vision, and ending with **(5)** high-resolution image forming vision[5]. We agree natural selection acts to elaborate visual complexity. In addition, we propose origins of parts of eyes are rooted in responses to photostress (**blue background**) and were later co-opted for vision once coexpressed under developmental control (**grey background**). (**A**) indicates the origin of photosensitive proteins (**yellow**), which may have originated to repair and predict damage to light-induced stress. (**B**) pigments (**brown**), such as melanin also originated to respond to light-stress and were assimilated into photostress mitigation under (**A**). Proteins of lenses (**C**, **Pink**) are varied and have numerous linkages to light-induced stress, and layers of 'lens-like' material are often found associated with non-image forming eyes[5]. Finally, (**D**, **green**) cilia, which increase surface area of photoreceptive membranes, are driven by UV stress [73].

II. Light-Induced Stress and Eye Evolution

A. Responses to light-induced stress were critical for multiple levels of eye evolution

Image-forming eyes evolve incredible complexity optimized for fine scale resolution, object detection, and motion tracking. However, we propose canonical visual functions need not be the only driver of eye evolution, instead arising as an emergent property of stress induced evolutionary innovation in response to a pervasive environmental toxin: light. Ultraviolet (UV) light destroys lipids, proteins, and DNA. We hypothesize that components of eyes arise, often repeatedly, from stress response networks that evolve to predict, preempt, and avoid UV damage — facilitating multiple, independent origins of eyes. Even before the most rudimentary eyes originate, evolution may link independently evolving stress responses to create complex and effective networks to mitigate UV damage. The eventual co-option and

genetic-regulatory control of these networks may lead to the origins of crucial advances in functional complexity: lenses, retinas, and pigment shields. We propose a macroevolutionary history of vision in which stress response pathways are a primary driver of eye origins (Fig 1).

B. Response to light-stress could involve both plasticity and mutation under selection.

Phenotypic plasticity in response to the environment is one possible engine of innovation, whose contributions to complex trait evolution we are just beginning to appreciate more widely[67,68]. Plasticity is quite different from more typical mutation driven explanations of evolution where the primary source of variation is the random mutation of genes that underlie a focal trait. In contrast, phenotypic plasticity may be a faster generator of variation within populations than mutation, creating strikingly labile systems that allow organisms to tolerate a range of environmental regimes. Unlike mutations, which are rare and often detrimental, plastic expression or development may allow immediate and appropriate responses to different environments. Over time, plastic responses may become assimilated into developmental genetic programs, which can be elaborated to create complex traits [66]. Therefore, phenotypic plasticity changes the order of events in the origins of novelty compared to mutation-selection. Under mutation-selection, random, undirected mutations create the variation in traits that selection acts upon. With phenotypic plasticity, variation first arises through plastic expression of stress/developmental programs, while mutations may occur later, allowing advantageous, yet previously plastic patterns of expression to become heritable.

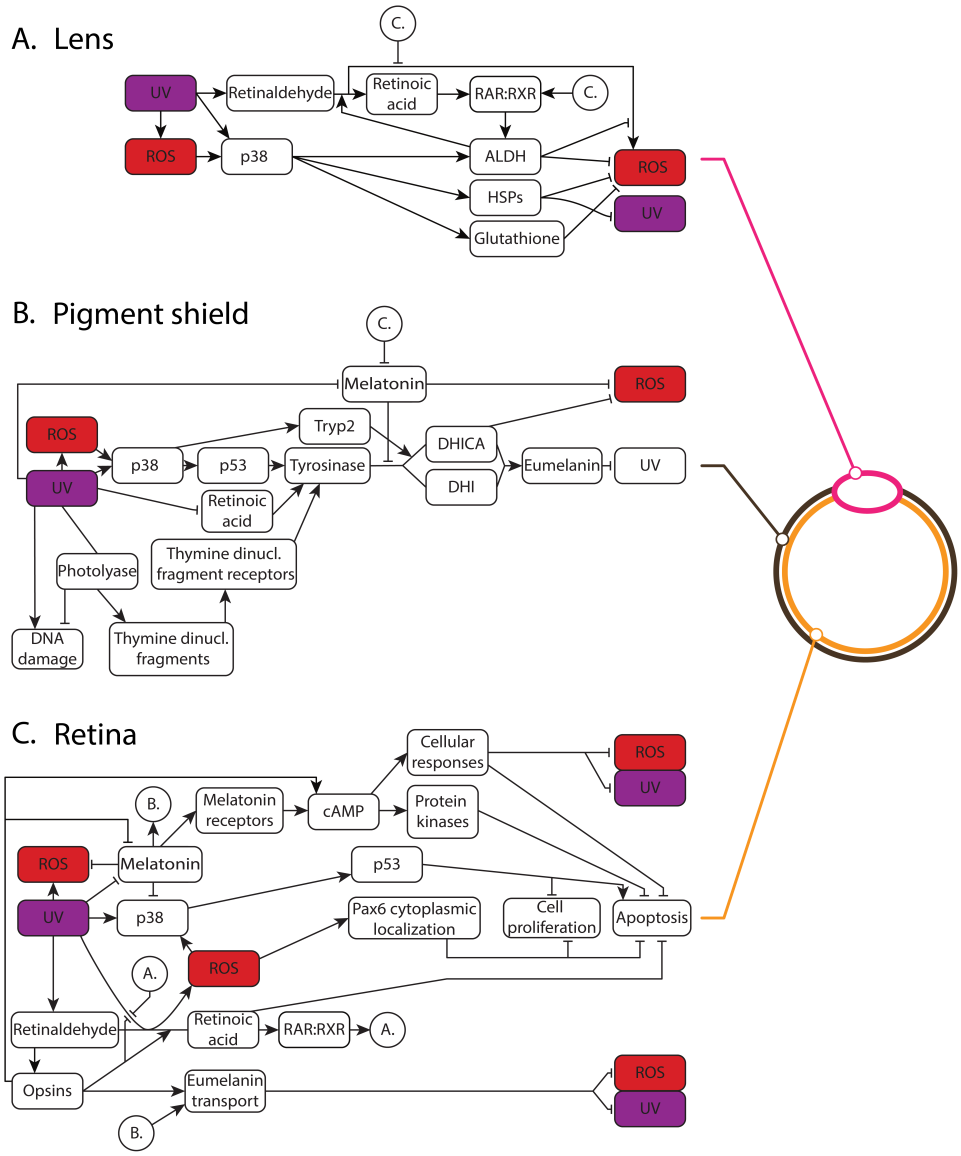

Figure 2. Diagram showing the UV- and ROS-initiated stress response pathways that underlie the origin of each functional part of vertebrate eyes. The lens (A) shows overexpression of UV-blocking, chaperone, and reducing enzymes tied to the expression of light sensitive molecules in the retina (C). The pigment cup (B) is an elaboration of the original pigment shield and is a full-spectrum light blocking membrane formed from specialized portions of stress-related pathways which mitigate ROS and UV stress. The retina (C) shows overexpression of light sensitive proteins and polypeptides evolved to control pathways that rely on pathways (A) to block excess UV light and (B) to produce melanin and metabolize cytotoxic compounds in order to mitigate UV and ROS stressors.

III. Stress-related innovations underlie crucial components of eye evolution

In the following sections, we discuss in more detail the role stress responses may have played in the evolution of eye components associated with discrete transitions in functional

ability. Beginning with the origin of lenses, we discuss the benefits of light-protection as eyes transitioned from directional light sensors to organs capable of object resolution. We next discuss the evolution of photosensitive proteins that make up retinas, revealing another example of a stress-induced evolutionary innovation, tying stress-mitigation pathways to increasingly accurate UV sensors. We then outline a possible evolutionary path behind the molecular roots of pigments and their integration with networks governed by photosensitive proteins. Finally, we describe how these separate components might have evolved to be regulated together genetically. The following sections lay out a hypothesis that addresses some gaps of acuity-driven eye evolution, showing that stress-responses to light function are a primary driver of eye evolution.

A. Lenses as Emergent Properties of Stress Mitigation

Because lenses allow higher visual acuity, their multiple origins define crucial evolutionary transitions between directional and image-forming vision, while at the same time illustrating connections between stress and eye evolution. The genes recruited to form lenses are quite variable in origin yet very commonly are stress-response proteins. This trend has been noted for some time, but has been explained largely because of proteins' ability to remain transparent and not interfere with the visual function of the eye [5,74]. However, by examining these proteins and their roles in UV-stress responses, we see possible origins of lenses by upregulation and concentration of particular proteins in front of photoreceptors to guard against toxins and mitigate UV stress. Lens proteins absorb UV light, reduce protein-protein interactions, chaperone protein folding, and metabolize cytotoxic compounds that arise as a product of both UV exposure and photoreception [75–77] (Fig. 2A). In the stepwise framework of eye evolution, the shift between directional photoreception and low resolution, image-forming

vision is marked not only by the deepening of a pigment cup, but also the appearance of 'lens-like' material between the retina and the epithelial lining [5]. We hypothesize that the evolution of these specialized lens-like cells originates from the pressure to filter UV light from reaching an increasingly sensitive retina, which would limit external generators of ROS and aldehydes and therefore create a milder environment for photoreceptors. Thus, the stress-related origins of these crucial novelties suggest that lenses, and by extension image forming vision, originated as a protective mechanism against damage from UV and reactive oxygen species and were later elaborated by natural selection on visual function (Fig. 2A).

Lenses are constructed of stress-response proteins in multiple species (Table 1) [74,78,79]. As one example, the lenses of octopuses recruited several proteins into lenses, the most common of which are aldehyde dehydrogenases and glutathione S-transferases [80]. The aldehyde dehydrogenases are particularly interesting because they belong to a class of proteins upregulated in response to ROS stress in animals. Aldehyde dehydrogenases aggressively find and neutralize dangerous reactive aldehydes such as retinaldehyde (i.e. retinal, the chromophore of opsin-based light sensitivity), 4-hydroxy-2-nonenal, and malondialdehyde, all of which are produced during UV exposure [75]. The second protein family in octopus lenses, glutathione S-transferase, detoxifies lipid molecules damaged by free radicals and ROS and acts as a free radical scavenger [75,81]. In mammals, lenses are comprised mostly of retinaldehyde dehydrogenase and diverse crystallin proteins. The incorporation and hyperexpression of aldehyde dehydrogenases and crystallins once again shows lenses are derivations of an ancestral function mitigating damage from UV stress [74]. A second protein in mammalian lenses, alpha-crystallin, arises from a larger gene family known by a different name: small heat shock proteins. These proteins, similar in function to glutathione S-transferase, work as

chaperones to prevent protein-protein interactions and refold damaged proteins in response to ROS, heat, and UV stressors [82]. These properties are exceptionally useful in a lens and cornea not only because of their tolerance for environmental stressors, but also because the clarity of the lens suffers when protein-protein interactions occur.

The repeated evolution of lenses as focusing optics creates opportunities for animals to perform high-resolution spatial vision, but the advantage of improved image resolution may not be what began lens evolution each time. Photoreceptor cells of any organism are exposed to extreme amounts of UV, ROS, and heat stress in order to perform their intended function. In addition, the molecular machinery required for photoreception generates toxic aldehydes and ROS. As eyes became more sensitive by packing more photosensitive proteins into each cell, the amount of free aldehydes and the rate of ROS production would have quickly increased. In order to mitigate damage incurred by this increase in visual acuity, expression of stress response proteins would have been useful, perhaps as a protective layer inside the evolving eye that inhibits protein-protein interactions and maximizes optical clarity of the proteins. We see these proteins co-opted into lenses. Aldehyde dehydrogenases, heat shock proteins, and homologs of glutathione S-transferase are all found highly expressed in lenses. The paralogs of these genes expressed outside lenses share the same protective functions as their lens-specific counterparts, indicating stress related functions predate co-option into lenses and do not represent adaptation for the specific stresses encountered in lens. Thus, examining the origins genes used in lenses and high-resolution spatial vision reveal that stress may have driven co-option in response to lethal UV and ROS stress as eyes became specialized for vision. Elaboration of lenses for fine focusing and a graded refractive index in water [83] probably evolved later.

Table 1. Common lens crystallins in vertebrates and invertebrates along with their non-lens functional classification. Table reconstructed from [79] with additional information from [84].

Lens Crystallins		
Vertebrates		
	α	small heat shock proteins/chaperones; all vertebrates
	βγ	members of microbial stress protein superfamily; all vertebrates
	ε	lactate dehydrogenase B; ducks, crocodiles
	δ	argininosuccinate lyase; birds, reptiles
	τ	α-enolase; turtles, ducks, other vertebrates
	ζ	novel quinone oxioreductase; guinea pig, camel, degu, llama, rock cavy
	μ	relative of bacterial ornithine cyclodeaminase; Australian marsupials
	η	retinaldehyde dehydrogenase; elephant shrews
	ρ	relative of aldo-keto reductase; frogs
	λ	relative of hydroxyl CoA dehydrogenase; rabbits, hares
	π	glyceraldehyde-3-phosphate dehydrogenase; geckos
Invertebrates		
	S	glutathione S-transferase and its relatives; cephalopods
	Ω/L	relative of aldehyde dehydrogenase; cephalopods, especially octopus; scallops; squid light organ
	J	novel protein, potentially saposin/swaposin homologs; jellyfish
	Drosocrystallin	novel protein; Drosophila

B. Protective pigments linked to light exposure

Shielding pigments are a crucial step in the evolution of eyes, allowing for directional photoreception by shading photoreceptors from one direction [64,85]. While pigment shields are needed for directional photoreception to evolve, the link between photosensitive proteins and antioxidant pigments may have been forged initially not for the sake of vision, but in order to protect cells from UV and ROS damage. Although animal eyes use different pigments in

different taxonomic groups [63], melanin is one of the best-studied light absorbing pigments recruited into eye spots, eyes, and extraocular photoreceptors. Besides its role in vision, melanin plays an integral role in protecting cells from UV damage as the end result of highly elaborate and specific protective pathways. Upon exposure to UV light, cells undergo immediate pigment darkening through the photooxidation of preexisting melanin, specifically blocking UV light and quenching ROS, while balancing the costs of *de novo* melanin synthesis [86–88] (Fig. 2B). A closer examination of melanin production reveals intimate connections with UV stress consistent with a deep and ancient association that predates the origins of eyes.

In vertebrates, melanin synthesis responds to the byproducts of UV-damage and stress (Fig 2b), often by regulating tyrosinase, the rate limiting enzyme in melanin synthesis. Here, we mention four different ways melanin production is connected to light, even outside of eyes. First, tyrosinase is upregulated by UV damage to DNA through a series of specific mechanisms. UV light damages DNA by causing neighboring thymines in the DNA to bond to each other. These damaged pieces of DNA are repaired by photolyase proteins (themselves activated by UV light), which excise and discard damaged segments, creating small fragments of DNA. The discarded fragments then upregulate tyrosinase, leading to melanin production [89]. A second way tyrosinase activity is related to light is through two proteins involved in every pathway discussed in this paper (Fig. 2A,B,C): p38 and p53. The MAPK protein p38 is activated in response to UV light and phosphorylates the inactive form of the p53 tumor suppressor protein. In addition to its many roles repairing DNA damage, controlling cell division, and regulating apoptosis of damaged cells, p53 also upregulates tyrosinase activity to create additional melanin [90,91]. Circadian rhythms, themselves controlled by light, further modulate p53 activity, making the interaction of p38 and p53 much more prevalent during

daylight hours [92,93]. A third connection between tyrosinase and light is through the photosensitive vitamin-a derivative, retinoic acid, which plays a significant role in regulating tyrosinase activity, controlling melanin synthesis in response to retinoic acid pathways that sense light [94–96]. Finally, while the previous mechanisms all increased melanin synthesis, melatonin—a ROS and photosensitive molecule—inhibits synthesis, potentially creating a "pool" of melanin precursors during dark hours, later available for quick assembly at low metabolic costs [97](Fig 2B). The clear links between melanin production/deployment, UV exposure, and ROS stress highlight the damage-mitigation properties that may have led to melanin's initial co-option into early photoreceptors.

C. Origins of photoreceptors from sensors of oxidative stress

Retinas can detect minute changes in light by densely packing photosensitive proteins into photoreceptor cells, thereby increasing sensitivity to light and allowing increased visual acuity [5,98]. In most animal eyes, the photosensitive proteins are opsins [but see 99], usually assumed to be present in animal eyes for reasons of visual function, including quick response time and use of a chromophore that can be efficiently regenerated [5]. However, we suggest the driving process behind repeated co-option of opsins into retinas may also be rooted in its notable predisposition to endure and mitigate photostress. In addition to the stressors discussed earlier in this paper, UV light produces the toxic compound retinaldehyde, a necessary ligand for opsin photosensitivity [100]. Extant opsin proteins tightly bind this free aldehyde and upregulate secondary messengers that perform a myriad of regulatory tasks in the cell [58,101–103].

The close ties between opsins and other initiators of stress response allows them to withstand the stresses of dedicated, high density photoreceptors, and could explain their co-

option into each independent evolution of a retina. As eyes transition along Nilsson's [5] stepwise evolutionary framework, the amount of opsin proteins in each photoreceptive cell retina is constantly increasing, leading to an increase in intracellular stress due to higher concentrations of retinal needed to achieve peak efficiency. Light-sensing of this caliber is a double-edged sword. The cells that are hypersensitized to light, and therefore the quickest to succumb to high levels of exposure [104], are also the cells that must be exposed to the brightest light levels in order to fulfill their function. To survive these conditions, mitigation responses to light stress must be linked to immediate detection of UV light. Opsins perform as part of a stress response network initiated by light exposure: Nuclear retinoic acid receptors detect a drop in RA associated with UV exposure, melatonin receptors respond to the increase in ROS, and opsins detect the increase in retinaldehyde that accompanies vitamin-a metabolism/photodegradation [100]. The speed and precision of opsin's photosensitivity may have been co-opted as a regulator for other UV stress responses, allowing immediate and comprehensive UV protection. Because of this, opsins may represent one of the only photosensitive proteins that could be expressed in high enough quantities to perform visual tasks yet also mitigate photo-stress quickly enough so that photoreceptors aren't immediately killed by cytotoxic byproducts of UV exposure. Thus, we find it possible that opsin's repeated use in eyes could have initially been related to its role as a cytosolic retinaldehyde and UV light receptor, while its use as a visible light receptor is a secondary elaboration on these ancestral functions. This hypothesis predicts the ancestral function of opsins to be UV-detectors, which could be tested with comparative methods.

We suggest the light sensitivity of opsins and their subsequent use in retinas is derived from an ancestral function tracking retinaldehyde levels to coordinate pathways that mitigate stress (Fig 2C). This hypothesis is supported by several lines of evidence: 1) Melatonin receptors (close relatives of opsins) track ligand concentrations as a proxy for light exposure; 2) Retinoic acid has clear advantages for mediating photostress and is used by existing photostress pathways, and 3) Rhodopsin-like GPCR 161 binds retinoic acid and is tied to crucial cell processes [105–107]. Melatonin acts as a powerful scavenger of ROS but it is quickly broken down by both ROS and UV light, leading to intracellular melatonin concentrations that closely track oxidative and UV stress levels [108]. Melatonin receptors regulate intracellular adenylyl cyclase in response to local melatonin concentrations, and are integral in many ROS and UV damage mitigating pathways like melanin production [109–111]. Retinoic acid can also be used as a sensor for UV stress, but unlike melatonin, retinoic acid fluctuations are more specifically tied to UV exposure, dropping more than fifty percent in the presence of UV light [100]. Nuclear retinoic acid receptors use the low intracellular concentrations of retinoic acid that accompany light to stall cell division and inhibit apoptosis [112–114]. We see very similar function through the rhodopsin-like GPCR-161, which acts as a retinoic acid receptor controlling cell division, migration, and growth [105,107]. Ancestral opsins likely also evolved as a retinoid receptor, initially regulating intracellular cyclic nucleotide levels (cAMP or cGMP) in response to retinaldehyde, the photosensitive metabolic precursor to retinoic acid [58,115]. In addition, opsin's evolutionary history likely primed their downstream signalling pathways to intertwine with existing GPCR initiated stress response pathways [116–119]. Nuclear retinoic acid receptors, melatonin receptors, and opsins all play roles in directing stress responses to UV exposure, but opsins have evolved as both the most accurate photodetector and a potent activator of UV-

protective pathways. Without the co-expression of each of these sensors into a coordinated ROS and UV quenching stress pathway, now found in retinas, reaching the concentration of photoreceptors needed for visual tasks could quickly kill cells through accumulated photooxidative stress and cytotoxic products of photodegradation [120–123].

D. Genomic regulation transforms stress response networks into a single evolutionary unit

The concept that selection for improved visual function drives eye evolution rests on the assumption that eye morphology is heritably expressed as a module. This evolutionary step could be achieved by movement away from transient, plastic expression to coordinated developmental processes divorced from stressors and driven by transcription factors [66]. One of the largest gaps in our knowledge of eye evolution is how these regulatory relationships evolved between transcription factors and the stress response pathways outlined above [124]. In animals, UV/ROS response pathways are largely controlled by the Pax family of transcription factors, which contribute to development and patterning of lenses, pigment shields, and retinas [125].

The evolution of genetic-regulatory control through Pax would have allowed disparate, plastic responses to be expressed as cohesive, permanent photosensitive modules. By examining literature surrounding a vertebrate copy of Pax, Pax-6, we hypothesize that ROS and UV stress responses may have been involved in the assimilation of transient stress pathways into predictable developmental programs [76,126,127]. Because of its direct role in UV-specific protection and ROS quenching [76,127], Pax-6 would have likely co-occurred alongside the other pathways it now regulates in eye development (Fig. 1). In addition, sequence similarities between Pax-6 promoters and promoters of genes involved in crucial UV-response networks could have facilitated the co-option of Pax-6 as an ‘unified’ activator. Heat shock

elements (found in lenses), antioxidant response elements (found throughout the eye), and p53 binding sites (cell cycle arrest and pigment expression) are all exceedingly similar to the 'optimal' Pax-6 binding site (Fig 2). As a result, very few mutations are needed to change any one of these promoter regions into one that would recognize Pax-6 as an activator [128]. Under Pax-6 regulation, exposure to UV light would no longer be necessary, though still sufficient, to initiate the pathways discussed in sections A-D -- setting in motion the evolution of complex eyes.

IV. Summary and Conclusions

A. Stress-induced origins complete a framework of eye evolution

Common hypotheses for the evolution of eyes rely on selection for increased visual function to explain the origins and elaboration of structures and functions of eyes. While this framework provides a reasonable explanation for the elaboration of existing traits (e.g. increasing precision of lens focusing after its origin) — it remains incomplete in its explanation of the origins of these traits. Attempting to justify the origin of eyes based solely on selection for improved visual acuity creates circular reasoning, leaving no obvious evolutionary starting point for eyes or parts of eyes to independently evolve. By focusing on eye evolution through the lens of mitigating stress from light, we suggest that the repeated co-option of genes from particular functional categories into each origin of lenses, retinas, and pigment shields could be driven by selection to mitigate photostress. Therefore, stress, not vision, may have often created *initial* selection to maintain co-expression and evolve co-regulation of genetic mechanisms, bringing together parts of eyes before the behavioral connections to light sensitivity could select for improved visual acuity.

B. Future directions & Acknowledging limitations

Thinking about stress-induced origins as an engine of novelty inspires significant hypotheses that relate to processes that potentially drive complex trait evolution as a whole. Yet we acknowledge much work needs to be done to test these ideas. In constructing this framework, we face a number of restrictions that represent excellent avenues for future research. Next, we discuss several limiting factors and potential investigations that could address them, including explicit comparative work to infer the history of function and the potential confirmation bias of studies that summarize vast amounts of literature.

First, our hypotheses need to be tested with explicit comparative methods. At present, the hypotheses rely mainly on findings from basic biology and clinical research that describe protein functions in the present day. This presents two unique problems which future comparative studies would address: the timeline of stress network evolution and ancestral protein function. Without specific comparative studies, including those that take advantage of time-calibrated molecular phylogenies, it is impossible to place an order to the evolutionary events that occurred as stress networks co-opted new actors and were in turn co-opted into innovations and novelties. At the same time, with the incredible volumes of untapped sequence data, reconstructing well supported histories of genes is fast becoming a matter of streamlining computational pipelines. Combined, investigating the history of gene function using comparative techniques will allow testing explicit hypotheses, such as responses to stress being more ancient functions than visual functions.

A particularly compelling application of comparative techniques would be to examine the ancestral function of opsins. A critical point in the evolutionary history of animals was the ability for opsins to tightly bind retinaldehyde, forming the first bistable opsins that could be

used for reliable, quick, light detection. Through experimental mutation, comparative studies, and ancestral sequence reconstruction [129], we may be able to unravel how this specific function evolved. Perhaps the drive to detect smaller concentrations of retinaldehyde created a selective pressure to increase binding efficiency without the need for any form of visual system in place. Assuming the ancestral opsin was a retinaldehyde receptor, the longer each molecule of retinaldehyde remained bound, the more sensitive the cell would be to smaller and smaller concentrations of this cytotoxic compound. This pressure may have culminated in an opsin with the ability to permanently bind retinaldehyde, no longer tracking intracellular concentrations, but instead detecting the isomer of retinaldehyde trapped in the binding pocket, which changes in response to UV light. This scenario makes particular predictions, (1) that the ancestor of all opsins may lack the ability to permanently bind retinal, but will function as a retinoid receptor, and (2) that close relatives of the opsin family may have preserved their ancestral function as retinoid receptors. Testing these predictions will require a combination of additional sequencing, further comparative phylogenetic methods, and experiments examining binding properties of extant ocular and nonocular opsins.

A second concern is that because stress responses occur throughout the body, we might expect under our model of stress-induced novelty, to see eyes evolve on organisms in more broadly distributed patterns than we do. In other words, why don't eyes evolve everywhere an animal is exposed to light? While this concern is tempting, we point out two counter arguments. First, we are not claiming that light-stress acts alone to fully evolve eyes, instead suggesting it acts in tandem with elaboration of visual structures through natural selection. Therefore, it is logical that natural selection will affect different body regions differently, only sometimes

resulting in evolution of elaborate eyes. Second, components of eyes, and sometimes even elaborated eyes, do show a broader distribution on animal bodies than complex eyes that are often only on the head. For example, chitons show a distributed system of eyes around their shells [130], ciliated cells in octopus skin sense and respond to light without a nervous system [9], hydra has light sensitive neurons around the body [131], sea urchins likely function as a single, large eye [132], and even human skin has melanin that might respond to light using opsins [133]. Each of these examples show that eyes or their components can evolve in many places, outlining intriguing questions about historical constraints on the evolution of eye placement [134]. Ultimately, these examples suggest that stress responses and selection for visual acuity function in tandem as drivers of eye evolution.

A third concern that inspires future research is that our synthesis of literature from a wide array of unrelated fields risks confirmation bias. The breadth and depth of study required to unravel the molecular mechanisms of cell physiology at the level required for this paper exists mainly in a few model organisms. Because of this, our conclusions are drawn from a small subset of the tree of life, relying on examples from vertebrates and a select few invertebrates. This critique highlights the need for future work and opens the door for future comparative studies to reinforce the support we report in a handful of well researched organisms. Interestingly, although we summarized a limited selection of pieces from the developmental network tied to eye evolution, we find that origins from light-induced stress responses likely extend far beyond our restricted exploration [74,99,135–139].

The advent of reliable single cell sequencing and advances in theory surrounding cell type evolution [66,140] provides another way to address issues surrounding both confirmation bias

and restrictive sampling. By sequencing genes expressed in individual cells, we can examine relative enrichment of particular stress networks in particular organs. We predict cells that make up eyes will have greater representation of stress-response genes than many other cell types. Perhaps counter to this prediction, we acknowledge how *shockingly common* it is to find stress-related genes involved in almost any innovation or novelty [65,73,141,142]. Pleiotropic function of developmental genes and novelties tracing back to stress-responses have been superficially noted during countless studies in nearly every field, a pattern that, from an evolutionary perspective, becomes more interesting the more evidence is gathered [68,143–145].

In conclusion, broadening our interpretations of the processes that drive eye evolution reveals an expanded framework that explains the origin and evolution of eyes, vision, and photoreception. Examining stress as a potential engine of plasticity and variation elucidates, once again, a surprisingly common occurrence of stress responses underlying many complex traits and developmental modules. We believe this near-ubiquitous co-option of stress genes in novelties may highlight an underappreciated macroevolutionary pattern that will illuminate, in part, the relative contributions of *de novo* evolution and co-option to evolution as a whole. We are excited to see future research on cell motility, sensory systems, and cell type evolution test these hypotheses and begin to delve into larger evolutionary questions about the roles of *de novo* evolution and co-option.

Acknowledgements
We would like to acknowledge J. Goodheart, N. Hensley, E. Ellis, and N. Picciani for their discussions and critiques of this manuscript. We would also like to acknowledge previous members of the Oakley lab: M.S. Pankey, D. Plachetzki, J. Serb, and especially D. Speiser,

who contributed the seeds of these ideas in years past. We acknowledge support from NSF grant IOS-1754770.

Chapter 4:
Convergent Evolution of Retinal-based Photosensitivity Across GPCRs

Abstract

Membrane receptors of the G-protein coupled receptor (GPCR) superfamily are indispensable as drug targets [146], optogenetic tools [147,148], and sensory proteins [149]. Within the riotous molecular diversity of GPCRs, only a single family is known to be light sensitive: the opsins. Understanding and leveraging opsin-based light-sensitivity has facilitated advancements in diverse fields such as medicine, ecology, and evolution. Despite the vast uses for light-sensitive GPCRs, no studies to date have recovered a photosensitive GPCR outside of opsins using comparative methods. Here we report convergent evolution of photosensitive GPCRs. Our broad phylogenetic analyses of over 18,000 GPCRs find two independent origins of photosensitivity outside of opsins and identify an additional seventy instances of convergent evolution on amino acid residues that may enable light-detection. These results reveal swaths of the GPCR superfamily as potentially light-sensitive, highlighting undiscovered functions and unknown biological interactions. Our evolutionary framework provides rich potential for future functional studies that will yield new possibilities for therapeutic targets, optogenetic tools, and a better understanding of basic biological processes.

I. Introduction

The G-protein coupled receptor (GPCR) superfamily is the largest clade of membrane-bound receptors, is responsible for a major portion of all cellular signaling, and is crucial to the identification and treatment of diseases [150,151]. GPCRs play critical roles in developmental and physiological processes surrounding vision, smell, taste, secretion, metabolism, neurological signaling, immune responses, cellular differentiation and growth [146,152–154]. They are the targets of approximately 40-50% of all drugs for applications ranging from psychiatric

medication, to cancer treatments, to experimental optogenetic therapies. However, our ability to discover new targets, therapies, and molecular tools is limited by our understanding of the stimuli used by uncharacterized GPCRs. Despite their importance to modern medicine, just 10% of all GPCRs in the human genome currently considered as possible drug targets [149]. In the remaining 90%, the large number of orphan receptors with unknown ligands and functions highlight the incredible untapped potential of the GPCR superfamily.

The rhodopsin-like GPCRs contain 80% of all GPCR diversity including the only clade known to detect light: opsins [30,155]. Because of their importance in vision and assumed uniqueness within GPCRs, incredible focus has been placed on understanding and harnessing the function of opsin proteins. However, the assumption that opsins represent the singular origin of photosensitivity in GPCRs has led to a myopic view, likely hindering progress to understand how light affects critical processes such as circadian rhythm, cognition, mood, and hunger. Here we present evidence counter to the opsin-centric paradigm of light sensitivity, identifying the first clades of non-opsin, light sensitive GPCRs, and an additional seventy convergent origins of amino acid residues that facilitate light sensitivity. Our evolutionary framework provides rich potential for future functional studies that will yield new possibilities for therapeutic targets, optogenetic tools, and a better understanding of basic biological processes.

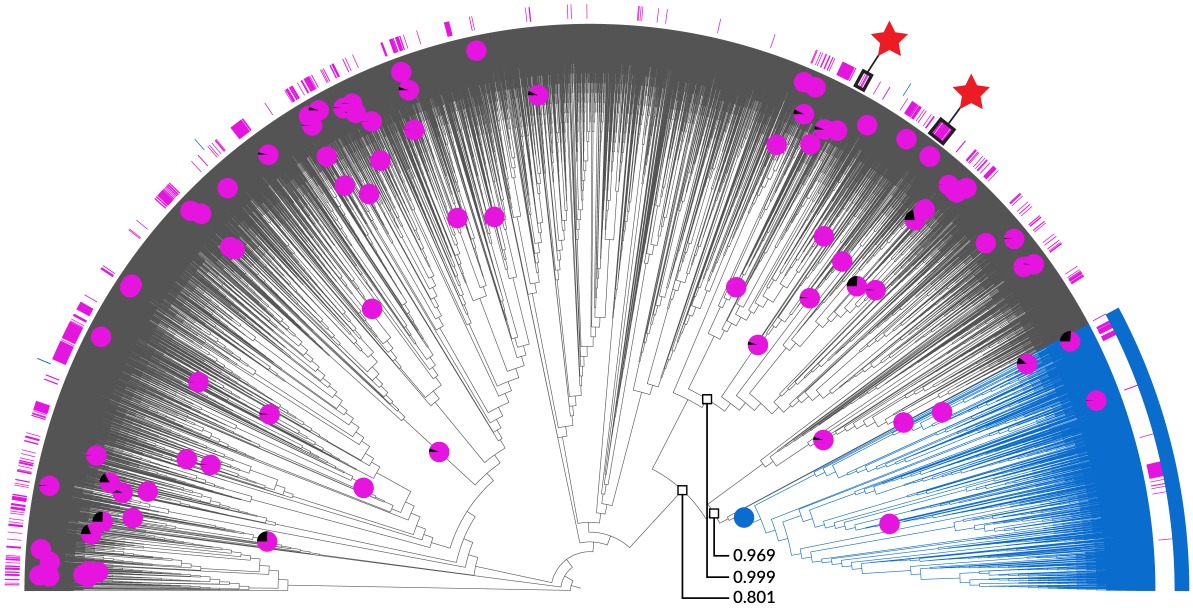

Figure 1 Phylogenetic tree of 18,634 GPCRs. **Blue** -- The canonical opsin family and the extent of previously known photoreception within GPCRs. **Magenta** -- All instances of a lysine at an alternate binding site (90, 91, 117, and 186 combined). **Tip color strips** - Each strip marks the position of an extant protein with a lysine at either an alternate binding site or the canonical K296. **Internal pie charts** - Each chart represents a predicted origin of a lysine at either an alternate binding site or the canonical site K296. Within each chart, black represents the likelihood of observing any other amino acid and the color section represents the likelihood of observing a lysine. Charts with no black indicate near certainty of a lysine first appearing at that node (Methods). **Red stars** indicate two clades which have been experimentally verified as light sensitive, **internal black boxes** indicate significant support values (Methods) for nodes separating these origins from opsins and each other. Additional support values, tip labels, and internal node labels are available in the supplement and online: (https://itol.embl.de/tree/1692313219255131559153215).

II. Scallop GPCRs represent first non-opsin, photosensitive GPCRs

Distantly related to canonical opsins and falling within the broader rhodopsin-like GPCR family, we find the first phylogenetic evidence of non-opsin, light-sensitive GPCRs (*CCAPR1, AAR1*) that lack canonical chromophore binding sites (*likelihood ratio test*, $P<0.0001$). Opsins' ability to detect light is derived from their binding affinity for the chromophore, retinal. It is held in the binding pocket of the opsin protein by a Schiff-base bond made possible via a lysine in the seventh-transmembrane helix. This lysine (K296) is highly conserved in opsins and is often used by scientists as a diagnostic residue, inferring loss of

photosensitivity in its absence. This result has justified removing sequences without K296 from opsin databases and phylogenies in the effort to filter out non-photosensitive genes [10,30,156].

Recently, several alternate sites in GPCRs were proposed that would enable retinal binding in the absence of the canonical lysine [157]. We discovered two GPCRs (*CCAPR1* and *AAR1*) that possess a lysine that resides in the binding pocket in homology models, bind retinal and exhibit photosensitivity in *in vitro* assays, and are expressed in photosensitive tissues of the bay scallop, *Argopecten irradians* [158]. Despite these multiple lines of evidence for photosensitive function, both sequences represent proteins that would have been excluded from previous studies of opsins because they do not have the canonical K296, leaving their phylogenetic position in the GPCR superfamily a critical, but unknown question. To answer this, we reconstructed and compared topologies of GPCR trees including opsins and outgroup GPCRs, and recovered these two photosensitive GPCRs far outside of opsins (Methods).

After accounting for topological uncertainty (Methods), we find that *CCAPR1* and *AAR1* were always recovered outside of canonical opsins in well supported clades. To test the sensitivity of our results to strategies of multiple sequence alignment, we aligned GPCRs using two separate programs and four different algorithms (Methods). While the topologies recovered were slightly different in each instance, no rearrangement led to the recovery of either *CCAPR1* or *AAR1* within the canonical opsin clades, supporting convergent origins of light sensitive GPCRs.

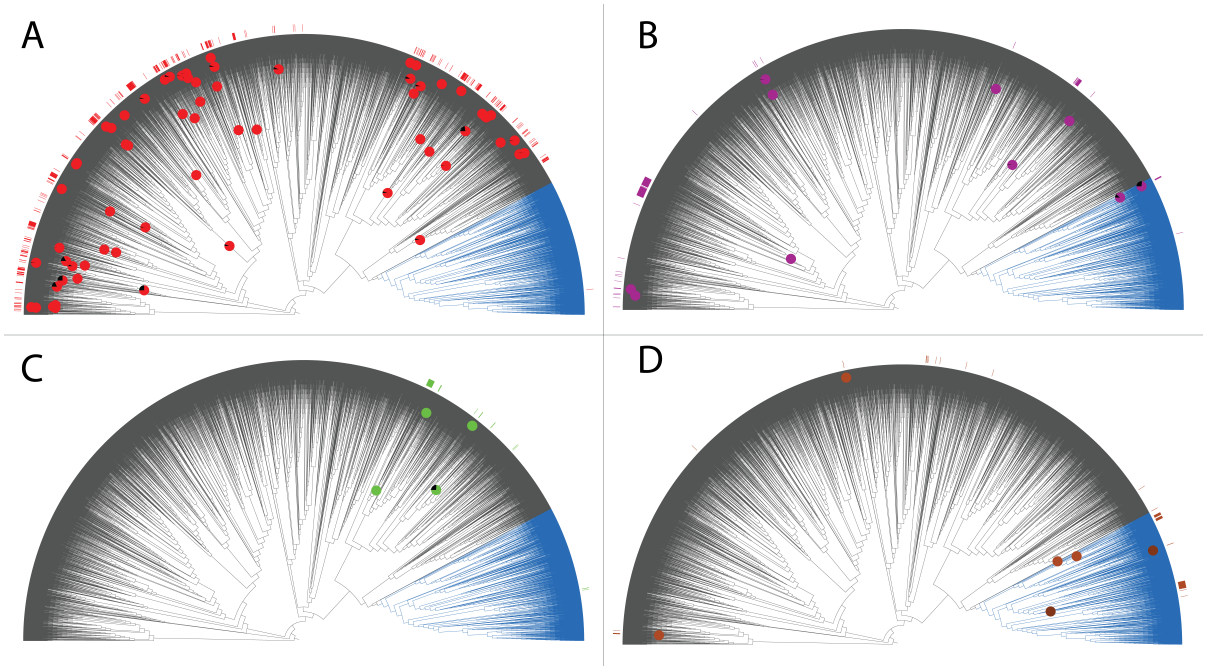

Figure 2. Origins of lysine at each alternate binding site. In every tree, **blue branches** represent the canonical opsin clade and extent of currently known photosensitivity in GPCRs. (**A**) Site 186, verified as functional in *AAR1*; (**B**) Site 94, verified functional in *AAR1*; (**C**) Site 117, verified as light sensitive in *CAPR1*; (**C**) Sites 90-91, unverified if functional on their own, but they are known to associate with retinal in the opsin binding pocket and shift opsin sensitivity towards UV.

III. Multiple origins of alternative retinal-binding sites across GPCRs

Using the alternate retinal-binding sites in *CCAPR1* and *AAR1*, we identify an average of 85 convergent origins of alternate retinal binding across the GPCR superfamily, some with origins predating Bilateria. Based on previous in vitro assays, there are five identified alternate binding sites that can potentially facilitate light sensing in non-opsin GPCRs when a lysine is present at the site. These alternate binding sites are referred to by the residue number that they align with on the bovine rhodopsin (NP_001014890): 90K, 91K, 94K, 117K, 186K, and 293K. *CCAPR1* has no lysine homologous with K296 but has 117K, and *AAR1* is also lacking K296 but has 94K and 186K alternate sites. We find that each gene represents an independent origin of light sensing in non-opsin GPCRs.

With both *CCAPR1 and AAR1* able to bind retinal with alternate binding sites, we test if these sequences represent a larger clade of genes with similar alternate binding sites, or are restricted to only *Argopecten.* We survey our GPCR tree to identify amino acids and reconstruct ancestral states at the alternate binding sites present in both *CCAPR1* (site 117) and *AAR1* (site 94,186). We used conservative models of prior and root states, favoring an inflation of type II error in order to reduce false positives and only recover well supported origins (Methods). While this approach allows us to confidently discuss the ancestral states recovered, its conservative estimate of clade depth may also split origins close to one another, slightly elevating the number of origins recovered. Close relatives of *CCAPR1* share the 117K mutation, and close relatives of *AAR1* share the 186K, 94K, or both mutations. We find *CCAPR1* and close relatives are only represented by paralogs within Bivalvia, placing their origin before the origin of Bivalves. *AAR1* and close relatives are much older, with paralogs containing the 94K mutation conserved in taxa across Bilateria, although the 186K mutation only appears in protostomes.

This expanded search also recovered clades distantly related to *CCAPR1*, *AAR1*, and canonical options. Across the GPCR superfamily, we estimate seventy-three independent, convergent origins of lysine at crucial alternate binding sites that facilitate light sensitivity. To account for the effects that uncertainty in our topology may have on ancestral state reconstruction, we created a distribution describing the number of independent origins over a set of 100 bootstrapped topologies for each alignment strategy (Methods). This number combines origins for all sites predicted to enable light sensitivity without K296 (90/91K, 94K, 117K, and 186K). While three sites 94K, 117K and 186K are functionally verified to enable photosensitivity [158], the number of origins added by including 90/91K is negligible. Across all

trees and alignment strategies the origins of 186K were consistently four to eight times higher than origins for the other binding sites.

It is important to note that outside of the two clades containing *CCAPR1* and *AAR1*, functional retinal binding and light sensitivity still must be confirmed. Given the large number of independent origins and genes involved, it is unlikely that each clade represents functional photosensitivity. By the same token, it is also unlikely that *CCAPR1* and *AAR1* are unique in their photosensitivity, and belong to clades which span bilateria. Future work must focus on streamlining the process of screening these putative light sensors for retinal binding and functionaling light sensitivity [158], with an effort to sample from a diversity of origins identified in this publication and species in which those origins appear.

IV. Conclusions

Overall, our evidence shows that light sensitivity has evolved more than once in GPCRs, and forces us to reevaluate the role light may play in many physiological pathways. While *CCAPR1* and *AAR1* represent the first functionally verified non-opsin photosensitive GPCRs, they each share the mutations that enable light sensitivity with close relatives found in a broad range of animal phyla. Convergence upon these specific mutations at crucial binding sites is widespread across the GPCR superfamily, revealing that light may play a much more pervasive role in biology than previously hypothesized. While our evolutionary framework presents unique opportunities for future work to investigate how these newly identified GPCRs use light to connect processes previously perceived as disparate and independent, to advance our understanding of their evolution, to help develop optogenetic tools, and to present new therapeutic approaches to pathologies that arise from their malfunction.

V. Methods

Data aggregation

Sequences were acquired from genbank's *nr* database using the Sequence Discovery, Recovery, Inventory, and Value Estimation (SeDRIVE) pipeline. This pipeline uses a bait file and bitscores to query a user defined set of organisms from genbank, remove duplicate, short, and spurious sequences, then produce both organism-specific and whole dataset fasta files for alignment. The advantage of this pipeline is the objective manner in which sequences are filtered, fixing the subjective bias that has plagued opsin studies since their inception. We used a bait file that contained opsins, rhodopsin-like, cAMP, frizzled, and adhesion GPCRs in order to test the assumption of clear delineations between these groups. We only used protein sequences from organisms with fully sequenced genomes available on genbank. A total of 18,634 sequences were recovered from 462 species (Table S4).

Phylogenetic reconstruction

Due to the size of this dataset, the selection of feasible reconstruction methods was somewhat limited. While subsetting the dataset and using supertree reconstruction is possible, it requires *a priori* definition of related proteins and clade delineation, a practice that would defeat the purpose of this study.

Because alignments impact both tree and ancestral state reconstruction, we used several alignment methods to guard against algorithmic biases and uncertainty. Due to the size of the dataset, we used the MAFFT alignment program. Within MAFFT, we tested several algorithms: L-INS-i (slow, but highly accurate), FFS-NS-1 (very fast, but very rough), and AUTO, which selected FFS-NS-2 (fast, but rough).

We then used each alignment to build a tree using FastTree under the LG+CAT model (20 rates), and pseudo replicated each alignment 100 times using RAxML's "-f j" option. We

used the initial best tree from each alignment as a starting point for each pseudo replicated alignment, and reconstructed a tree based on the pseudoreplicate in FastTreeMP under the LG+CAT model. Overall, this created a set of 100 alignments and trees for each alignment strategy for a total of ~700 alignments and trees. For each tree, support on nodes are reported as the result of 1,000 SH tests on each node, default in FastTree v2.1 and identical to PhyML's "SH-like local supports".

To statistically test the phylogenetic position of the two experimentally verified scallop photosensitive GPCRs, we used the best tree from the most accurate alignment strategy (MAFFT L-INS-i) to reconstruct the tree in FastTree with a constrained starting topology and compared the likelihood to that of our unconstrained reconstruction. Our constrained starting topology was identical to our unconstrained topology with the exception of 1) all canonical opsins were connected to a single, massive polytomy, and 2) the two scallop sequences were moved from their original position to within this polytomy. We enforce this polytomy by setting -constraintWeights to a very large number (10,000.0), effectively allowing us to test the alternative hypothesis that these light-sensitive scallop GPCRs are actually opsins and may have been incorrectly placed in our best tree. We use the approximately unbiased (AU) test as well as the Shimodaira-Hasegawa (SH) from CONSEL [159] test to compare if the two topologies were significantly different. We verified that this result by checking if the position of *CCAPR1* and *AAR1* were well supported outside of canonical opsins and independent from each other in each of the other 699 trees, ensuring a robust result insensitive to alignment or reconstruction methods.

Ancestral state reconstruction

To prepare for ancestral state reconstruction, we inspected the alignment to locate our target residues (90/91, 94, 117, 186, and 293) and developed a number of scripts for maneuvering large-scale data and extracting character states at user-defined residues, available here (bitbucket repo). Due to differences in alignment algorithms and the taxonomic breadth of our ingroup, it is possible that our alignment inserted gaps which resulted in sequences distantly related to opsins no longer sharing residue homology, despite our confidence in homology based on more detailed alignment analyses [158]. To account for this, if such a misalignment occured, we extracted the sequence and re-aligned against bovine rhodopsin alone under the previous publication's alignment algorithm. If this realignment fixed the misalignment, we assumed that the two columns in the alignment represent homologous sites, and the difference was due to algorithmic stochasticity with such a large dataset. We then scored each alternate site and any homologous columns for the presence of lysine, building a binary dataset for ancestral state reconstruction. We conducted ancestral state reconstruction using the *rayDISC* command from the corHMM *R* package [160]. Within this command, we used the *maddfitz* root prior on each trait reconstructed. This leads to a conservative interpretation of extant data and places a data-driven prior at the root, allowing for a more objective reconstruction of origins informed by our large aggregation of existing data [161].

For a transition to an alternative state to be considered successful, we required that reconstructed likelihoods of alternative state over ancestral state be greater than 2 for the following equation as described in Pagel, 1999: $-2 * ln(H_{a=i}/H_{a=j})$ where $H_{a=i}$ is the smaller of the two likelihoods and $a$is any node [162]. This system has commonly been used as a statistical benchmark in two-state ancestral state reconstruction studies to indicate a significant difference between the two states. To account for reconstruction of an incorrect topology and its effects

on the number of origins, we applied the ancestral state reconstruction methods detailed above to each of the 100 bootstrapped trees for each of the seven alignment strategies. To count origins, we designed an algorithm implemented in the script corIGINS, which takes a strict stance on the interpretation in counting the origins and regains across a tree (Figure S1). We also examined the effect each alignment algorithm had on the contribution of each binding site to the total number of origins under the original alignment for each algorithm (Table S5-7). We transformed these results into meaningful visuals by developing the corVIEW script, which creates the pie chart files compatible with iTOL [163] and allows for user-defined visibility thresholds on both binary and multistate characters. We reconstructed ancestral states for each of the focal residues independently.

Data Availability

All supplementary data and scripts can be found at the bitbucket repository for this publication: bitbucket.org/swafford/light-sensitive-gpcrs.git.

Chapter 5:
Multimodal Sensorimotor System in Unicellular Zoospores of a Fungus

doi: 10.1242/jeb.163196

Abstract

Complex sensory systems often underlie critical behaviors, including avoiding predators or locating prey, mates, and shelter. Multisensory systems that control motor behavior even appear in unicellular eukaryotes, such as *Chlamydomonas*, which are important laboratory models for sensory biology. However, we know of no unicellular opisthokonts that control motor behavior using a multimodal sensory system. Therefore, existing single-celled models for multimodal sensorimotor integration are very distantly related to animals. Here, we describe a multisensory system that controls the motor function of unicellular, zoospores of a fungus. We find zoospores of *Allomyces arbusculus* exhibit both phototaxis and chemotaxis. While swimming, they move towards light and settle on cellulose membranes exuding combinations of amino acids. Furthermore, we report that closely related *Allomyces* species do not share this multisensory system. Instead, these species respond to either chemical or light stimuli, but not both. This diversity of sensory systems within *Allomyces* provides a rare example of a comparative framework that can be used to examine the evolution of dramatic changes in sensory systems. The tractability of *Allomyces* and related fungi as laboratory organisms will allow detailed mechanistic investigations into how sensory systems may have functioned in early opisthokonts before multicellularity allowed for the evolution of specialized cell types.

I. Introduction

All organisms rely on sensory systems to gather information about their surroundings from external stimuli. The integration of individual sensory modalities into multisensory

systems, or sensory systems, greatly increases the amount of information an organism can use to form responses and behaviors. Although multimodal sensory systems are common in multicellular, motile organisms, there are significantly fewer multisensory systems known from unicellular eukaryotes [1]. The relative rarity of studies characterizing these systems in unicellular, laboratory-tractable organisms has resulted in a significant taxonomic gap between current model systems and animals.

To address this deficit, we focused on fungi characterized in part by motile, zoosporic life stages. Zoosporic fungi collectively form a clade within the 'early-diverging lineages' of fungi, outside the better known Ascomycota and Basidiomycota. They are largely found in freshwater ecosystems with a global distribution [164]. Zoosporic fungi are typically characterized as saprobes, such as *Allomyces*, although parasitic life strategies on both plant and animal hosts also do exist [165]. Similar to all fungi, colonies of *Allomyces* use mycelia to absorb nutrients and ultimately grow reproductive structures. Unlike most fungi, *Allomyces* produce zoosporangia, terminations of mycelial branches that make, store, and ultimately release a multitude of single-celled, flagellated propagules, termed zoospores [166]. When the appropriate environmental cues are present, zoospores are produced *en masse*, eventually bursting from zoosporangia [164]. Once in the water column, the zoospores rely on a single, posterior flagellum to propel themselves away from the parent colony and towards suitable substrates or hosts [166].

During dispersal of the zoosporic life stage, interpretation of environmental cues is critical for the survival and success of the future colony [167]. Zoospores have a finite amount of endogenous energy reserves, and no zoospore is known to metabolize energy from external sources [168]. This energetic constraint places significant pressure on the zoospore to efficiently

locate a favorable environment for settlement and growth. The evolution and maintenance of a sensory system within the unicellular zoospore allows it to evaluate external conditions, move towards suitable habitats, and avoid hazards [164]. Previous studies across zoosporic fungi led to the discovery of a number of sensory modalities that guide zoospore dispersal including chemotaxis, phototaxis, and electrotaxis [2,169,170]. However, these studies have neither tested a single species for multiple sensory modalities, nor posited the possibility that these single senses may only be a portion of a more complex sensorimotor system guiding zoospores.

In the fungus *Allomyces,* zoospores use chemotaxis or phototaxis to guide dispersal and settlement [2,171]. Chemotaxis towards the source of amino acid gradients allows zoospores to congregate at the site of an injury or on decaying material in the water column [169]. *Allomyces macrogynus* zoospores possess refined chemosensation, settling on substrates at varied rates in response to different amino acids [169]. Alternatively, the zoospores of *Allomyces reticulatus* display positive phototaxis, potentially leading spores to swim towards the air-water interface [2]. Studies to date have not tested for the presence of chemotaxis in *A. reticulatus* or phototaxis in *A. macrogynus*. In animals, positive phototaxis and subsequent 'rafting' on floating debris work to considerably increase the dispersal range of planktonic larvae [172]. Similarly, spores attracted to the surface may encounter floating debris, algal hosts, or currents that aid their dispersal.

While little is known about the molecular mechanisms of chemotaxis in fungal zoospores, the underpinnings of their photosensitivity are beginning to come to light. The protein responsible for light-detection in zoospores of *Blastocladiella emersonii*, a close relative of *Allomyces*, is a bacteriorhodopsin gene called *CyclOps (Beme-Cycl*) [173]. Unlike many bacteriorhodopsins that regulate ion channels, *Beme-Cycl* acts through regulating

intracellular cGMP [174]. A *CyclOps* gene is present in the genome of *A. macrogynus (Amag-Cycl*), a species that has been anecdotally described as having phototactic zoospores [166]. A recent study by Gao et. al., 2015, however, contradicts claims of phototaxis in *A. macrogynus* by revealing the proteins encoded by *Amag-Cycl* are orders of magnitude less sensitive to dark/light transitions than *Beme-Cycl* proteins [175]. This raises questions about the sensitivity of *CyclOps* proteins needed for phototaxis and mutations potentially responsible for shifts in photosensitivity.

The uncertainty surrounding the sensory systems of *Allomyces* zoospores demands experimental evidence to clarify number and types of modalities used during dispersal and settlement. Addressing the current deficits in our understanding of fungal sensory systems is also motivated by the potential to discover a system that will further our knowledge of multisensory evolution and function in early opisthokonts. Here, we investigate the responses to chemical and light gradients in three species of *Allomyces*; revealing previously unknown variation in fungal sensory systems, and discovering a novel multisensory system in a zoosporic fungus.

II. Materials and Methods

Culture conditions

We used *Allomyces arbusculus* str. ATCC 10983, *Allomyces reticulatus* str. California 70 from ATCC (cat. No. 42465), and *A. macrogynus* from the Roberson lab (Arizona State University). We kept cultures of *A. macrogynus* and *A. arbusculus* in both solid and liquid media. For solid media, we used [176] Emerson YSS (HiMedia M773) at half strength. Colonies transferred aseptically in a laminar flow hood every 4 weeks by moving a chunk of mycelia from the leading edge of the colony onto a new plate. For liquid media, we followed the protocol for

Machlis' medium B. We inoculated liquid cultures via sterile transfer of sporangia and mycelia into a 125mL Erlenmeyer flask containing 50mL of liquid media and antibiotics [176] for the first generation. For all subsequent generations kept in liquid culture, we used dilute salts solution to initiate sporulation of the previous generation's colonies. We then added 1 mL of this zoospore-dilute salts solution (referred to a sporulation product from here on) into new liquid media. Both liquid and solid cultures were grown on an orbital shaker at 140rpm and kept at room temperature (~24C). Cultures in liquid media were kept for a maximum of 5 days, and were considered ready for sporulation after 72 hours. Because *A. reticulatus* did not grow well in liquid media, we cultured *A. reticulatus* on full strength Emerson YSS media for no more than 6 weeks. Propagation of *A. reticulatus* cultures in solid media was performed identically to the other species.

Sporulation conditions

Liquid cultures of *A. macrogynus* and *A. arbusculus* were considered for sporulation after 72 hours. We visually inspected colonies under a microscope to confirm the absence of gametangia. Using a stainless steel sterile mesh, we strained the colonies out of growth media and rinsed them 5 times with dilute salts solution to remove the growth media from the colonies. We then placed the rinsed colonies and strainer in a pyrex dish with 10 mL of dilute salts solution and allowed them to sporulate for no more than 90 min. Once either sufficient spore density had been reached ($5x10^5$ spores mL^{-1} for chemotaxis [169], $1x10^6$ spores mL^{-1} for phototaxis [173]) or 90 minutes had elapsed, the mesh and colonies were lifted out of the dish [169]. Because *A. reticulatus* was only grown on solid media, we took a surface scraping to lift sporangia from the agar and placed it into a pyrex dish with 10 mL dilute salts solution [177]. If no zoospores were present, we replaced dilute salts solution every 20 minutes for the first hour.

Sporulation typically occurred within 8 hours, after which the colonies were strained from the dilute salts solution.

Phototaxis trials

We conducted phototaxis trials in a custom 1x5x3 cm (WxHxL) plexiglass chamber. We added 10 mL of sporulation product, diluted to $1x10^6$ spores mL^{-1} [173], to the test chamber and allowed the solution 15 minutes in total darkness to dark adapt and randomize spore distribution. Preliminary trials showed that an adjustment period of 15 minutes, rather than 30 minutes [173], was sufficient for zoospore randomization and dark adaptation. After the adjustment period, spores were exposed to a white light (USHIO halogen bulb) through a 5 mm diameter fiber optic cable positioned 5 cm from the leading edge of the test chamber. To calibrate the intensity of light, we used a JAZ Oceanoptics light sensor with Spectrasuite v2.0.162. We adjusted the intensity of the light to $1.8\text{-}1.0x10^{13}$ mol of photons cm^{-2} on the edge closest to the light source. The chamber was designed with thick plastic to limit heat transfer to the zoospore suspension, no appreciable thermal turbulence or change in temperature was observed after 15 minutes of light exposure. After 15 minutes of light exposure we divided the test chamber into 4 sequential sub-chambers (10x25x15mm) using sterile glass slides. This resulted in 4 sub chambers (1, 2, 3, and 4) arranged linearly so that sub chamber 1 was closest to the light source, while subchamber 4 was the farthest away (Fig. S1A). We gently agitated the liquid in each subchamber to homogenize swimming spore distribution and counted swimming spore density in four, 10µl samples from each subchamber using a hemocyometer. To account for the variation in spore counts between each trial in both control and light treatments, we quantified the number of swimming spores in each chamber as a proportion of the total number of spores counted in that trial. This resulted in less inter-trial variation. A total

of 10 control treatments (no light exposure) and 18 experimental treatments were conducted for each species.

Chemotaxis trials

Chemotaxis trials followed the protocol established by Machlis [169]. The amino acids and combinations thereof we tested were Lysine (K), Leucine (L), Proline (P), L+K, L+P, K+P, L+K+P, and Buffer ($5x10^{-3}$M KH_2PO_4) solution (referred to as 'treatment solutions' from here on). All amino acid concentrations were $5x10^{-4}$M for each amino acid in all treatments. We created a chemical dispersal apparatus by drilling a hole through the lid of a 60x15mm petri dish and inserting a 5mm inner diameter glass pipette. We secured dialysis membrane (3500 MWCO) to the tip of the pipette and positioned it 3mm above the bottom of the petri dish (Fig. S1B). This creates a gradient in the petri dish of whatever solution is placed behind the dialysis membrane, allowing zoospores to navigate to the membrane, where they settle and can later be counted. The dialysis membrane was soaked and rinsed with DI water for 24 hours to remove potential contaminants and bubbles that would affect results [178]. Turbulence in the petri dish would greatly affect the strength of the gradient. To avoid this, we placed the modified petri dishes in a larger, sealed container to avoid drafts and the treatment solution was added last. This resulted in no movement of any container between the addition of treatment solution and the end of the trial.

To test chemotactic and settlement response to these treatments, we added 10 mL of sporulation product (diluted to $5x10^5$ spores mL^{-1}) to the petri dish and 300 µl of treatment solution into the pipette. As a control, we used 300 µl of buffer alone. We allowed the spores to react to the gradient in total darkness for 90 minutes. At the end of the trial time, we removed the pipette and dialysis membrane from the dish and gently shook it to remove excess liquid

[169]. We counted spores settled on the membrane under an Olympus szx7 at 400x or greater magnification.

Molecular Methods:

A) PCR, cDNA synthesis & sequencing: Because genomic data existed for *A. macrogynus* but expression data did not, we used PCR to attempt to identify if CyclOps genes are expressed in *A. macrogynus* zoospores. mRNA was extracted from zoospores using a NucleoSpin RNA XS kit. We synthesized cDNA using the Clontech cDNA synthesis kit. Primers were designed from putative rhodopsin/guanalyl-cyclase fusion proteins identified from the BROAD institute's Allomyces macrogynus genome, using the CyclOps protein from *Blastocladiella emersonii* as bait sequences [173]. Sets of primers were designed in IDT PrimerQuest. All PCRs products were visualized using a 1% agarose gel with 100bp ladder. PCR for CyclOps was only done on *A. macrogynus* as transcriptomes for *A. reticulatus* and *A. arbusculus* would yield expression data.

RNA was isolated from zoospores of *A. reticulatus* and *A. arbusculus* using Nucleospin xsRNA kit, and cDNA was synthesized using the NEBNext RNA First and Second Strand Synthesis modules. cDNA was sequenced using a multiplexed Illumina HiSeq lane at approximately 50x coverage.

B) Bioinformatics and Statistics: We trimmed Illumina data using Trimmomatic [179], assembled using Trinity 2.0, and analyzed on the UCSB Osiris bioinformatics platform [180,181]. Putative CyclOps proteins were identified using *Beme-Cycl* as a bait sequence in BLASTn searches against the *A. macrogynus* genome, NCBI bioproject 20563, and the new *A. reticulatus* and *A. arbusculus* zoospore transcriptomes. Any sequence with an e-score lower than 1e-40 was considered as a candidate. We then used the 'get orf' feature from Trinity to produce predicted

proteins from candidate genes, selecting the longest orfs with the highest similarity to *Beme-Cycl* protein when reciprocally BLASTed using blastp. We identified the putative *CyclOps* gene in both *A. reticulatus* and *A. arbusculus* (*Aarb-Cycl*). Reads from *A. arbusculus* transcriptome were mapped back to the putative *Aarb-Cycl* using Bowtie2 [182] and visualized using IGV viewer [183]. Because the bacteriorhodopsin and guanylate-cyclase domains appeared in different orfs on the same strand, each base was manually examined for uncertainty and low support. A guanine at site 378 was manually removed due to low coverage and low support in the reads, implying that the addition of guanine at position 378 most likely an assembly artifact. The manually edited *Aarb-Cycl* gene produced a single predicted orf with the appropriate bacteriorhodopsin-guanylyl cyclase domains (Fig. S2).

Candidate proteins were aligned using MAFFT under the L-INS-i strategy. Outgroups were selected based on a previous analysis [173] (Fig. S2). The alignment was used to create a phylogeny of candidate genes with RAxML 8 [184] and 100 bootstrap replicates using the GTR + Γ model. Trees were visualized in Evolview [185] and annotations were added in Adobe Illustrator.

Comparisons of zoospore phototaxis behavior were analyzed using JMP v12.0. Average spore counts per subchamber per trial were analyzed using pairwise Tukey's HSD between control and experimental sub-chambers. Although sample size was low, each sample represents an average of four replicates for each treatment. Zoospore chemotaxis was analyzed using Wilcoxon each pair due to the nonparametric distribution of results and small sample size (n=10).

III. Results

Allomyces reticulatus *relies on phototaxis*

Zoospores of *A. reticulatus* showed no significant deviation from the control when exposed to any amino acid treatment ($P > 0.05$ for all treatments) (Fig. 1). In accordance with existing literature, *A. reticulatus* showed a significant response to a directional light source [2]. The number of zoospores swimming in the subchamber closest to the light source (Fig. 2) was significantly higher than when no light source was present ($P = 0.0005$).

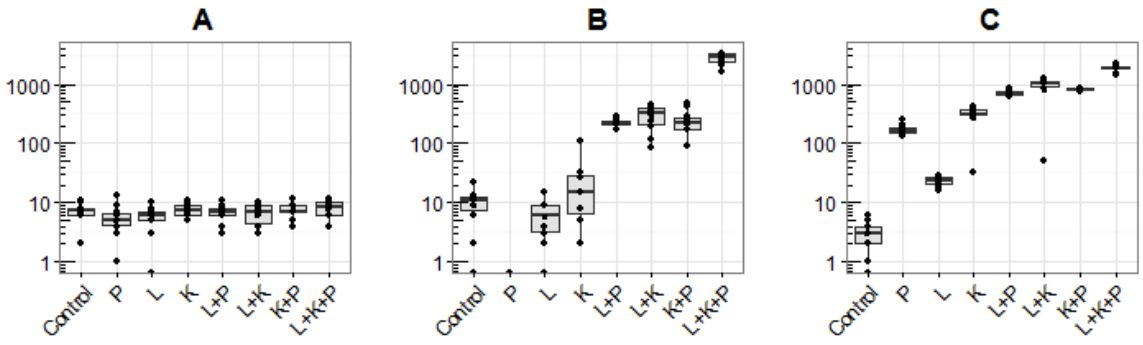

Figure 1. The number of zoospores settled on the dialysis membrane in response to varying amino acid treatments. (**A**) *Allomyces reticulatus,* (**B**) *A. arbusculus,* (**C**) *A. macrogynus*. Each column represents the number of zoospores settled on $2mm^2$ dialysis membrane after 90 minutes. Squares represent means. **Control**, Proline (**P**), Leucine (**L**), Lysine (**K**), and combinations thereof. N=10 for all treatments, Y axis in log scale.

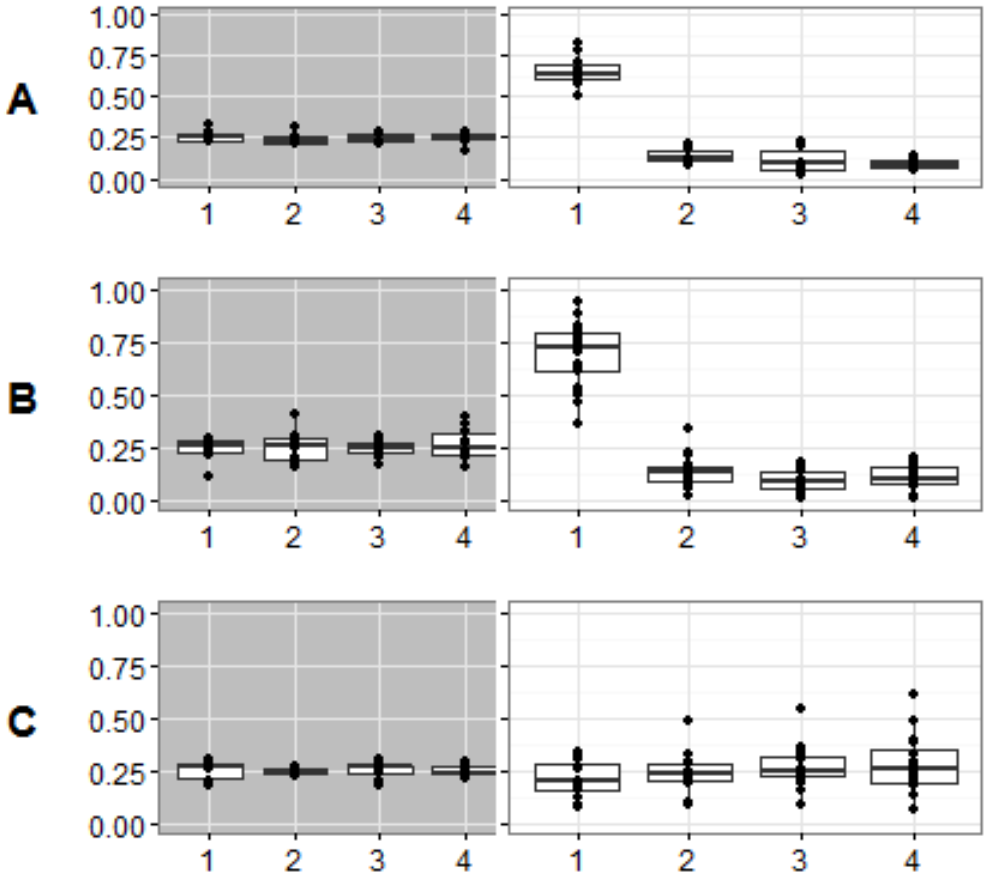

Figure 2. The proportion of swimming zoospores found in each subchamber in phototaxis trials. (**A**) *A. reticulatus* (**B**) *A. arbusculus* or (**C**) *A. macrogynus* zoospore distribution after 30 minutes of darkness (**grey background**) or 15 minutes of darkness followed by 15 minutes exposure to directional light (**white background**). The directional light source was positioned so light intensity was strongest in sub-chamber **1** and lowest in sub-chamber **4**. Data points in each subchamber represents the proportion of swimming zoospores counted in that subchamber relative to the total number of swimming zoospores counted in that trial. This method reduces the effect of variable inter-trial numbers of swimming zoospores on the overall trend.

Allomyces macrogynus *relies on chemotaxis*

As seen in previous experiments, *A. macrogynus* zoospores displayed a significant response to all amino acid treatments compared to the control [169]. Proline ($P = 0.0294$), leucine ($P=0.0275$), lysine ($P=0.0294$), and any combination of two or three amino acids when compared to a control ($P < 0.001$) (Fig. 1). Zoospore response to increasing treatment complexity was non-linear though roughly equal for all unique combinations of equal

complexity. *Allomyces macrogynus* zoospores showed no response when exposed to a directional light source (Fig. 2). The number of zoospores in all sub-chambers were the same for both light and dark trials ($P = 0.9976$).

Allomyces arbusculus *uses both chemotaxis and phototaxis in a multisensory system*

As expected from existing literature, *A. arbusculus* zoospores responded similarly to the zoospores of *A. macrogynus* when exposed to amino acid treatments. The number of spores settled increased in a nonlinear fashion as the complexity of the treatment increased, though at a lower average number of settled spores when compared to *A. macrogynus*: K+P ($P = 0.0014$), K+L ($P = 0.0014$), L+P ($P = 0.0008$), K+L+P ($P = 0.0004$) (Fig. 1). As opposed to *A. macrogynus, A. arbusculus* does not respond to, or cannot detect, gradients of single amino acids ($P > 0.05$ for all single A.A. treatments) with the possible exception of Proline. We found that no zoospores settled when *A. arbusculus* was exposed to trials of Proline alone, potentially indicating the potential for negative chemotaxis or inhibition of settlement in response to gradients of Proline by itself. However, we can not definitively resolve this reaction with our sample size (Proline v. Control: $P = 0.072$; $N=10$). When exposed to a directional light source, *A. arbusculus* display positive phototaxis (Fig. 2). The number of zoospores in the sub-chamber closest to the light was significantly higher ($P < 0.0001$) in light vs. dark trials and was comparable to the response of *A. reticulatus* zoospores.

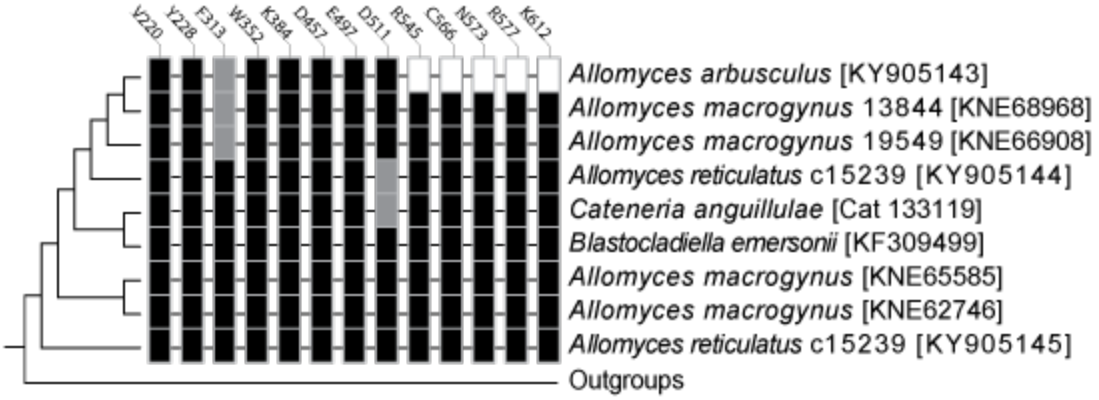

Figure 3. Cladogram showing the relationships and conserved amino acid residues in CyclOps proteins. Boxes indicate amino acid residues critical for binding [173], with residue numbers based on position in unaligned Beme-Cycl *(*AIC07007.1). Black boxes indicate an amino acid matching at that residue when compared to Beme-Cycl, grey boxes indicate a mutation, and white boxes indicate a gap.

CyclOps is present in all species, may not be expressed in A. macrogynus zoospores

Phylogenetic analysis of putative *CyclOps* genes reveals *CyclOps* presence in all three samples (*A. arbusculus & A. reticulatus*: transcriptome data. *Allomyces macrogynus*: previously available genome data). The single copy of *CyclOps* recovered from the *A. arbusculus* transcriptome revealed a possible truncation of the Guanylyl cyclase domain (Fig. 3). Both *Aarb-Cycl* and *Amag-Cycl* share a mutation at the putative functional residue F313 to I313. Despite the success of positive controls indicating successful PCR amplification, no primers successfully amplified *Amag-Cycl* from *A. macrogynus* zoospores.

IV. Discussion

Understanding how sensory modalities evolve and integrate with other behavioral circuits remains an open question in neurobiology and evolutionary biology. The *Allomyces* genus, with variation in sensory systems first discovered in this study, will aid in answering these questions. Previous studies showed *Allomyces* spores use either chemotaxis or phototaxis

to guide dispersal. Here, we reveal that the sensorimotor system in *A. arbusculus* is multimodal - able to process both chemical and light cues. Additionally, our results reveal the previously unknown complexity and variation of sensorimotor systems across species of *Allomyces*.

Variation in sensory modalities across Allomyces:

We report unknown variation in the distribution of sensory modalities across the genus. This variation manifests in two ways: the types of sensory modalities used by each species of *Allomyces* and the responses of *A. arbusculus* and *A. macrogynus* zoospores to the same amino acids. The lack of phototaxis, coupled with the inability to amplify *CyclOps* from zoospore mRNA in *A. macrogynus* is quite interesting due to conflicting reports between previous studies. These studies conflictingly report that *A. macrogynus* zoospores are either phototactic (through anecdotal evidence) [166] or that the photosensitive protein, Amag-Cycl, does not respond effectively to light [175]. Our findings support the hypothesis that *CyclOps* in *A. macrogynus* no longer effectively differentiates between light and dark, and suggest that *CyclOps* expression has been lost in zoospores. With the loss of phototaxis, the distribution and settlement of *A. macrogynus* zoospores likely deviated greatly from both *A. reticulatus* and *A. arbusculus* in natural settings.

The variation in *Allomyces* sensory systems discovered in this study coupled with the convergent function of *CyclOps* and animal opsins make it critical to our understanding of the evolutionary history of light sensing. Animal photoreception, mediated through Type II opsins, operates through intracellular regulation of cyclic nucleotides. Non-animal photoreception, mediated through Type I opsins, operates through channel and sensory rhodopsins (SRII). However, unlike Type II opsins, SRIIs do not regulate cyclic nucleotide concentrations. Instead, SRIIs indirectly regulate CheY, a protein that controls flagellar motion [186]. Type I and

II opsins are considered a spectacular example of convergent evolution [187]. *CyclOps* represents a functional evolutionary convergence, where the protein may modulate intracellular cGMP levels like some Type II phototransduction cascades, yet the sequence of *CyclOps* indicates its origin as a Type I opsin [174]. Studying the evolution of *CyclOps* sequence function through the variation in *Allomyces* will yield further insights into the evolution of this novel photosensory mechanism.

The failure of *A. reticulatus* to move and settle on the cellulose membrane does not necessarily indicate a complete lack of chemotaxis, but does highlight variation in *Allomyces* sensory system evolution. The amino acids tested here are frequently associated with wounded tissue and decaying organic matter, thus making them appropriate for guiding the distribution and settlement of sessile saperobes [178]. While previous studies uncovered differences in the combinations of amino acids *A. arbusculus* and *A. macrogynus* zoospores move toward [169], no one tested the possible loss of this chemotactic response in a close relative. The absence of chemotaxis to these amino acids in *A. reticulatus* may reveal a turnover in sensory capabilities responsible for controlling a vital behavior across the genus. Future studies will use the variation in both chemotaxis and phototaxis across *Allomyces* as a platform to understand behavioral integration, multisensory evolution, and sensory remodeling in an organism closely resembling ancestral opisthokonts.

The Multisensory System of Allomyces arbsuculus:

Allomyces arbusculus zoospores present an easily culturable, laboratory tractable system for investigating multimodal sensation in unicellular systems and its underlying mechanisms. Previous studies have independently confirmed that zoospores use a diversity of senses, but fungal zoospores have only been tested for a single sense per species [2,169,170,173].

Our findings represent the first study of zoospore multimodal sensing, and a concrete example of multimodal sensorimotor control in a unicellular opisthokont (Fig. 1-2). Though choanoflagellates fall within Opisthokonta [188] and potentially exhibit both aerotaxis and pH-taxis, it remains unknown if the sensory modalities in colonies (aerotaxis) are also used to direct the dispersal stage (pH-taxis) [189,190].

Zoospores, as unicellular, flagellated cells, might closely represent the ancestral opisthokont phenotype [188]. Specialized cell types and functions may often evolve through subfunctionalization followed by elaboration of the ancestral cell's functions [140]. This implies that as multicellular opisthokonts evolved, the foundation for specialized sensory modalities/cells already existed in 'pluripotent', generalized pathways. Under the subfunctionalization hypothesis, multimodal systems in unicellular organisms, such as we report in *A. arbusculus*, must have evolved prior to subfunctionalization in ancestral, multicellular opisthokonts. Future studies of the multimodal sensorimotor system in *A. arbusculus* zoospores and the variation in modalities across *Allomyces* may uncover how multiple senses became integrated into behavioral responses in ancestral opisthokonts. Understanding these mechanisms in the context of the cellular subfunctionalization hypothesis will further our understanding of sensorimotor evolution, elaboration and individuation through cellular specialization.

Conclusions

We present a multimodal sensorimotor system in a unicellular life stage of a fungus. The multisensory system of *A. arbusculus* zoospores is an excellent system to study how sensory modalities integrate into existing behavioral regimes. Together with existing models of unicellular sensory mechanisms, the variation in sensory modalities in *Allomyces* and other

early diverging fungi will allow us to formulate more accurate conclusions about the evolution of complex sensory systems and multisensory systems in ancestral eukaryotes. Lastly, the relatively narrow taxonomic breadth associated with the multiple transitions in sensory systems during *Allomyces* evolution will allow testing of broad questions in evolution; such as the role multimodal cell types play in the origin and evolution of specialized sensory systems, and how emergent behaviors evolve during sensory remodeling.

Acknowledgements

The authors would like to thank Robert Roberson for supplying *A. macrogynus* and Jason Stajich for his advice and guidance on molecular work in fungus systems. Additionally, the authors would like to thank Oren Malvey for his help in maintaining cultures.

Author Contributions

A.S. and T.O. designed experiments. A.S. carried out experiments and analyzed data. A.S. and T.O. wrote and edited the manuscript.

Funding

This work was supported by a Sigma-Xi Grant In Aid of Research and Rosemary Grant Award awarded to A.S. and NSF #1456859 awarded to T.O.

Data Availability

Alignment, newick, and behavioral data collected for this study is available for download at https://bitbucket.org/swafford/multisensory_systems_JEB_2017.

References Cited

1. Govorunova, E. G. & Sineshchekov, O. A. Chemotaxis in the green flagellate alga Chlamydomonas. *Biochemistry* **70,** 717–725 (2005).
2. Robertson, J. A. Phototaxis in a new Allomyces. *Arch. Mikrobiol.* **85,** 259–266 (1972).
3. Moldrup, M., Moestrup, Ø. & Hansen, P. J. Loss of phototaxis and degeneration of an eyespot in long-term algal cultures: evidence from ultrastructure and behaviour in the dinoflagellate Kryptoperidinium foliaceum. *J. Eukaryot. Microbiol.* **60,** 327–334 (2013).
4. v. Salvini-Plawen, L. & Mayr, E. in *Evolutionary Biology* (eds. Hecht, M. K., Steere, W. C. & Wallace, B.) 207–263 (Springer US, 1977).
5. Nilsson, D.-E. Eye evolution and its functional basis. *Vis. Neurosci.* **30,** 5–20 (2013).
6. Swafford, A. J. M. & Oakley, T. H. Light-Induced Stress as a Primary Evolutionary Driver of Eye Origins. *BIOLOGY* (2019). doi:10.20944/preprints201904.0107.v1
7. Nilsson, D. E. & Pelger, S. A PESSIMISTIC ESTIMATE OF THE TIME REQUIRED FOR AN EYE TO EVOLVE. *Proceedings of the Royal Society B-Biological Sciences* **256,** 53–58 (1994).
8. Ramirez, M. D., Pairett, A. N., Pankey, M. S., Serb, J. M., Speiser, D. I., Swafford, A. J. & Oakley, T. H. The last common ancestor of most bilaterian animals possessed at least 9 opsins. *Genome Biol. Evol.* (2016). doi:10.1093/gbe/evw248
9. Ramirez, M. D. & Oakley, T. H. Eye-independent, light-activated chromatophore expansion (LACE) and expression of phototransduction genes in the skin of Octopus bimaculoides. *J. Exp. Biol.* **218,** 1513–1520 (2015).
10. Porter, M. L., Blasic, J. R., Bok, M. J., Cameron, E. G., Pringle, T., Cronin, T. W. & Robinson, P. R. Shedding new light on opsin evolution. *Proc. Biol. Sci.* **279,** 3–14 (2012).
11. Wang, F. Y., Yan, H. Y., Chen, J. S.-C., Wang, T. Y. & Wang, D. Adaptation of visual spectra and opsin genes in seabreams. *Vision Res.* **49,** 1860–1868 (2009).
12. Spady, T. C., Seehausen, O., Loew, E. R., Jordan, R. C., Kocher, T. D. & Carleton, K. L. Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Mol. Biol. Evol.* **22,** 1412–1422 (2005).
13. Chang, C.-H. & Yan, H. Y. Plasticity of opsin gene expression in the adult red shiner (Cyprinella lutrensis) in response to turbid habitats. *PLoS One* **14,** e0215376 (2019).

14. Porter, M. L., Roberts, N. W. & Partridge, J. C. Evolution under pressure and the adaptation of visual pigment compressibility in deep-sea environments. *Mol. Phylogenet. Evol.* **105,** 160–165 (2016).
15. Musilova, Z., Cortesi, F., Matschiner, M., Davies, W. I. L., Patel, J. S., Stieb, S. M., de Busserolles, F., Malmstrøm, M., Tørresen, O. K., Brown, C. J., Mountford, J. K., Hanel, R., Stenkamp, D. L., Jakobsen, K. S., Carleton, K. L., Jentoft, S., Marshall, J. & Salzburger, W. Vision using multiple distinct rod opsins in deep-sea fishes. *Science* **364,** 588–592 (2019).
16. Innan, H. & Kondrashov, F. The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* **11,** 97–108 (2010).
17. Rapp, L. M. & Ghalayini, A. J. Influence of UVA light stress on photoreceptor cell metabolism: decreased rates of rhodopsin regeneration and opsin synthesis. *Exp. Eye Res.* **68,** 757–764 (1999).
18. Ozdeslik, R. N., Olinski, L. E., Trieu, M. M., Oprian, D. D. & Oancea, E. Human nonvisual opsin 3 regulates pigmentation of epidermal melanocytes through functional interaction with melanocortin 1 receptor. *Proc. Natl. Acad. Sci. U. S. A.* (2019). doi:10.1073/pnas.1902825116
19. Sato, K., Yamashita, T., Kojima, K., Sakai, K., Matsutani, Y., Yanagawa, M., Yamano, Y., Wada, A., Iwabe, N., Ohuchi, H. & Shichida, Y. Pinopsin evolved as the ancestral dim-light visual opsin in vertebrates. *Commun Biol* **1,** 156 (2018).
20. Kojima, D., Mori, S., Torii, M., Wada, A., Morishita, R. & Fukada, Y. UV-sensitive photoreceptor protein OPN5 in humans and mice. *PLoS One* **6,** e26388 (2011).
21. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).
22. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
23. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
24. Liu, K., Warnow, T. J., Holder, M. T., Nelesen, S. M., Yu, J., Stamatakis, A. P. &

Linder, C. R. SATé-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence Alignments and Phylogenetic Trees. *Syst. Biol.* **61,** 90–106 (2012).

25. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32,** 268–274 (2015).
26. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35,** 518–522 (2018).
27. Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C. & Gascuel, O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* **60,** 685–699 (2011).
28. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).
29. Speiser, D. I., Pankey, M. S., Zaharoff, A. K., Battelle, B. A., Bracken-Grissom, H. D., Breinholt, J. W., Bybee, S. M., Cronin, T. W., Garm, A., Lindgren, A. R., Patel, N. H., Porter, M. L., Protas, M. E., Rivera, A. S., Serb, J. M., Zigler, K. S., Crandall, K. A. & Oakley, T. H. Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinformatics* **15,** 350 (2014).
30. Feuda, R., Hamilton, S. C., McInerney, J. O. & Pisani, D. Metazoan opsin evolution reveals a simple route to animal vision. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 18868–18872 (2012).
31. Benton, M. J. & Donoghue, P. C. J. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24,** 26–53 (2007).
32. Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L. & Disotell, T. R. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* **75,** 165–183 (2014).
33. Pyron, R. A. A likelihood method for assessing molecular divergence time estimates and the placement of fossil calibrations. *Syst. Biol.* **59,** 185–194 (2010).
34. Murphy, W. J., Pringle, T. H., Crider, T. A., Springer, M. S. & Miller, W. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res.* **17,**

413–421 (2007).

35. dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M. J., Donoghue, P. C. J. & Yang, Z. Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. *Curr. Biol.* **25,** 2939–2950 (2015).
36. dos Reis, M., Inoue, J., Hasegawa, M., Asher, R. J., Donoghue, P. C. J. & Yang, Z. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. Biol. Sci.* **279,** 3491–3500 (2012).
37. Edgecombe, G. D., Giribet, G., Dunn, C. W., Hejnol, A., Kristensen, R. M., Neves, R. C., Rouse, G. W., Worsaae, K. & Sørensen, M. V. Higher-level metazoan relationships: recent progress and remaining questions. *Org. Divers. Evol.* **11,** 151–172 (2011).
38. Dunlop, J. A. & Selden, P. A. Calibrating the chelicerate clock: a paleontological reply to Jeyaprakash and Hoy. *Exp. Appl. Acarol.* **48,** 183–197 (2009).
39. Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L. & Rasnitsyn, A. P. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. Biol.* **61,** 973–999 (2012).
40. Blair, J. E. & Hedges, S. B. Molecular phylogeny and divergence times of deuterostome animals. *Mol. Biol. Evol.* **22,** 2275–2284 (2005).
41. Wiegmann, B. M., Trautwein, M. D., Winkler, I. S., Barr, N. B., Kim, J.-W., Lambkin, C., Bertone, M. A., Cassel, B. K., Bayless, K. M., Heimberg, A. M., Wheeler, B. M., Peterson, K. J., Pape, T., Sinclair, B. J., Skevington, J. H., Blagoderov, V., Caravas, J., Kutty, S. N., Schmidt-Ott, U., Kampmeier, G. E., Thompson, F. C., Grimaldi, D. A., Beckenbach, A. T., Courtney, G. W., Friedrich, M., Meier, R. & Yeates, D. K. Episodic radiations in the fly tree of life. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 5690–5695 (2011).
42. Nakatani, M., Miya, M., Mabuchi, K., Saitoh, K. & Nishida, M. Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaean origin and Mesozoic radiation. *BMC Evol. Biol.* **11,** 177 (2011).
43. Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A. J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T. R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L. S., Kawahara, A. Y.,

Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D. D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J. L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B. M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N. U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M. G., Wiegmann, B. M., Wilbrandt, J., Wipfler, B., Wong, T. K. F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D. K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K. M. & Zhou, X. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346,** 763–767 (2014).

44. Johnson, G. D., Ida, H., Sakaue, J., Sado, T., Asahida, T. & Miya, M. A 'living fossil' eel (Anguilliformes: Protanguillidae, fam. nov.) from an undersea cave in Palau. *Proc. Biol. Sci.* **279,** 934–943 (2012).
45. Near, T. J., Dornburg, A., Eytan, R. I., Keck, B. P., Smith, W. L., Kuhn, K. L., Moore, J. A., Price, S. A., Burbrink, F. T., Friedman, M. & Wainwright, P. C. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 12738–12743 (2013).
46. Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 13624–13629 (2011).
47. Martin, M. W., Grazhdankin, D. V., Bowring, S. A., Evans, D. A., Fedonkin, M. A. & Kirschvink, J. L. Age of Neoproterozoic bilatarian body and trace fossils, White Sea, Russia: implications for metazoan evolution. *Science* **288,** 841–845 (2000).
48. Moreau, C. S., Bell, C. D., Vila, R., Archibald, S. B. & Pierce, N. E. Phylogeny of the ants: diversification in the age of angiosperms. *Science* **312,** 101–104 (2006).
49. Benton, M. J. Phylogeny of the major tetrapod groups: morphological data and divergence dates. *J. Mol. Evol.* **30,** 409–424 (1990).
50. Peterson, K. J., Lyons, J. B., Nowak, K. S., Takacs, C. M., Wargo, M. J. & McPeek, M. A. Estimating metazoan divergence times with a molecular clock. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 6536–6541 (2004).

51. Hagadorn, J. W., Dott, R. H. & Damrow, D. Stranded on a Late Cambrian shoreline: Medusae from central Wisconsin. *Geology* **30,** 147–150 (2002).
52. Müller, J. & Reisz, R. R. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *Bioessays* **27,** 1069–1075 (2005).
53. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33,** 1635–1638 (2016).
54. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28,** 2689–2690 (2012).
55. Britton, T., Anderson, C. L., Jacquet, D., Lundqvist, S. & Bremer, K. Estimating divergence times in large phylogenetic trees. *Syst. Biol.* **56,** 741–752 (2007).
56. Lee, M. S. Y., Jago, J. B., García-Bellido, D. C., Edgecombe, G. D., Gehling, J. G. & Paterson, J. R. Modern optics in exceptionally preserved eyes of Early Cambrian arthropods from Australia. *Nature* **474,** 631–634 (2011).
57. Knoll, A. H., Javaux, E. J., Hewitt, D. & Cohen, P. Eukaryotic organisms in Proterozoic oceans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361,** 1023–1038 (2006).
58. Shichida, Y. & Matsuyama, T. Evolution of opsins and phototransduction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364,** 2881–2895 (2009).
59. Conant, G. C. & Wolfe, K. H. Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* **9,** 938–950 (2008).
60. Terakita, A. & Nagata, T. Functional properties of opsins and their contribution to light-sensing physiology. *Zoolog. Sci.* **31,** 653–659 (2014).
61. Colley, N. J. & Nilsson, D.-E. Photoreception in Phytoplankton. *Integr. Comp. Biol.* **56,** 764–775 (2016).
62. Darwin, C. On the origins of species by means of natural selection. *London: Murray* (1859). at <http://sciencestudies.pbworks.com/f/OoS.pdf>
63. Oakley, T. H. & Speiser, D. I. How Complexity Originates: The Evolution of Animal Eyes. *Annu. Rev. Ecol. Evol. Syst.* **46,** null (2015).
64. Nilsson, D.-E. The evolution of eyes and visually guided behaviour. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364,** 2833–2847 (2009).
65. Wagner, G. P., Erkenbrack, E., Maziarz, J. & Chavan, A. The uterine decidual stromal cell: A stabilized stress response? *Journal of Reproductive Immunology* **128,** 48 (2018).

66. Wagner, G. P., Erkenbrack, E. M. & Love, A. C. Stress-Induced Evolutionary Innovation: A Mechanism for the Origin of Cell Types. (2019). doi:10.20944/preprints201809.0193.v1
67. Moczek, A. P., Sultan, S., Foster, S., Ledón-Rettig, C., Dworkin, I., Nijhout, H. F., Abouheif, E. & Pfennig, D. W. The role of developmental plasticity in evolutionary innovation. *Proc. Biol. Sci.* **278,** 2705–2713 (2011).
68. West-Eberhard, M. J. *Developmental Plasticity and Evolution*. (Oxford University Press, 2003).
69. Gavelis, G. S., Keeling, P. J. & Leander, B. S. How exaptations facilitated photosensory evolution: Seeing the light by accident. *Bioessays* (2017). doi:10.1002/bies.201600266
70. Swafford, A. J. M. & Oakley, T. H. Multimodal sensorimotor system in unicellular zoospores of a fungus. *J. Exp. Biol.* **221,** (2018).
71. Oakley, T. H. & Pankey, M. S. Opening the 'Black Box': The Genetic and Biochemical Basis of Eye Evolution. *Evo Edu Outreach* **1,** 390–402 (2008).
72. Picciani, N., Kerlin, J. R., Sierra, N., Swafford, A. J. M., Ramirez, M. D., Roberts, N. G., Cannon, J. T., Daly, M. & Oakley, T. H. Prolific Origination of Eyes in Cnidaria with Co-option of Non-visual Opsins. *Curr. Biol.* **28,** 2413–2419.e4 (2018).
73. Chavali, P. L. & Gergely, F. Cilia born out of shock and stress. *EMBO J.* **32,** 3011–3013 (2013).
74. Piatigorsky, J. *Gene Sharing and Evolution: The Diversity of Protein Functions*. (Harvard University Press, 2009).
75. Lassen, N., Black, W. J., Estey, T. & Vasiliou, V. The role of corneal crystallins in the cellular defense mechanisms against oxidative stress. *Semin. Cell Dev. Biol.* **19,** 100–112 (2008).
76. Ou, J., Walczysko, P., Kucerova, R., Rajnicek, A. M., McCaig, C. D., Zhao, M. & Collinson, J. M. Chronic wound state exacerbated by oxidative stress in Pax6+/- aniridia-related keratopathy. *J. Pathol.* **215,** 421–430 (2008).
77. Leiers, B., Kampkötter, A., Grevelding, C. G., Link, C. D., Johnson, T. E. & Henkle-Dührsen, K. A stress-responsive glutathione S-transferase confers resistance to oxidative stress in Caenorhabditis elegans. *Free Radic. Biol. Med.* **34,** 1405–1415 (2003).

78. True, J. R. & Carroll, S. B. Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* **18,** 53–80 (2002).
79. Piatigorsky, J. Multifunctional lens crystallins and corneal enzymes. More than meets the eye. *Ann. N. Y. Acad. Sci.* **842,** 7–15 (1998).
80. Land, M. F. The evolution of lenses. *Ophthalmic Physiol. Opt.* **32,** 449–460 (2012).
81. Reiter, R. J. Interactions of the pineal hormone melatonin with oxygen-centered free radicals: a brief review. *Braz. J. Med. Biol. Res.* **26,** 1141–1155 (1993).
82. de Jong, W. W., Leunissen, J. A. & Voorter, C. E. Evolution of the alpha-crystallin/small heat-shock protein family. *Mol. Biol. Evol.* **10,** 103–126 (1993).
83. Sweeney, A. M., Des Marais, D. L., Ban, Y.-E. A. & Johnsen, S. Evolution of graded refractive index in squid lenses. *J. R. Soc. Interface* **4,** 685–698 (2007).
84. Piatigorsky, J., Norman, B., Dishaw, L. J., Kos, L., Horwitz, J., Steinbach, P. J. & Kozmik, Z. J3-crystallin of the jellyfish lens: similarity to saposins. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 12362–12367 (2001).
85. Walne, P. L. & Arnott, H. J. The comparative ultrastructure and possible function of eyespots: Euglena granulata and Chlamydomonas eugametos. *Planta* **77,** 325–353 (1967).
86. Bustamante, J., Bredeston, L., Malanga, G. & Mordoh, J. Role of melanin as a scavenger of active oxygen species. *Pigment Cell Res.* **6,** 348–353 (1993).
87. Brenner, M. & Hearing, V. J. The protective role of melanin against UV damage in human skin. *Photochem. Photobiol.* **84,** 539–549 (2008).
88. Borovansky, J. & Riley, P. A. *Melanins and Melanosomes: Biosynthesis, Structure, Physiological and Pathological Functions*. (John Wiley & Sons, 2011).
89. Eller, M. S., Ostrom, K. & Gilchrest, B. A. DNA damage enhances melanogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 1087–1092 (1996).
90. Khlgatian, M. K., Hadshiew, I. M., Asawanonda, P., Yaar, M., Eller, M. S., Fujita, M., Norris, D. A. & Gilchrest, B. A. Tyrosinase gene expression is regulated by p53. *J. Invest. Dermatol.* **118,** 126–132 (2002).
91. Cui, R., Widlund, H. R., Feige, E., Lin, J. Y., Wilensky, D. L., Igras, V. E., D'Orazio, J., Fung, C. Y., Schanbacher, C. F., Granter, S. R. & Fisher, D. E. Central role of p53 in the suntan response and pathologic hyperpigmentation. *Cell* **128,** 853–864 (2007).

92. Lau, A. W., Liu, P., Inuzuka, H. & Gao, D. SIRT1 phosphorylation by AMP-activated protein kinase regulates p53 acetylation. *Am. J. Cancer Res.* **4,** 245–255 (2014).
93. Volonte, D., Zou, H., Bartholomew, J. N., Liu, Z., Morel, P. A. & Galbiati, F. Oxidative stress-induced inhibition of Sirt1 by caveolin-1 promotes p53-dependent premature senescence and stimulates the secretion of interleukin 6 (IL-6). *J. Biol. Chem.* **290,** 4202–4214 (2015).
94. Paterson, E. K., Ho, H., Kapadia, R. & Ganesan, A. K. 9-cis retinoic acid is the ALDH1A1 product that stimulates melanogenesis. *Exp. Dermatol.* **22,** 202–209 (2013).
95. Orlow, S. J., Chakraborty, A. K. & Pawelek, J. M. Retinoic acid is a potent inhibitor of inducible pigmentation in murine and hamster melanoma cell lines. *J. Invest. Dermatol.* **94,** 461–464 (1990).
96. Roméro, C., Aberdam, E., Larnier, C. & Ortonne, J. P. Retinoic acid as modulator of UVB-induced melanocyte differentiation. Involvement of the melanogenic enzymes expression. *J. Cell Sci.* **107 (Pt 4),** 1095–1103 (1994).
97. Logan, A. & Weatherhead, B. Post-tyrosinase inhibition of melanogenesis by melatonin in hair follicles in vitro. *J. Invest. Dermatol.* **74,** 47–50 (1980).
98. Skalicky, S. E. *Ocular and Visual Physiology - Clinical Application*. (Springer, 2016).
99. Rivera, A. S., Ozturk, N., Fahey, B., Plachetzki, D. C., Degnan, B. M., Sancar, A. & Oakley, T. H. Blue-light-receptive cryptochrome is expressed in a sponge eye lacking neurons and opsin. *J. Exp. Biol.* **215,** 1278–1286 (2012).
100. Tolleson, W. H., Cherng, S.-H., Xia, Q., Boudreau, M., Yin, J. J., Wamer, W. G., Howard, P. C., Yu, H. & Fu, P. P. Photodecomposition and phototoxicity of natural retinoids. *Int. J. Environ. Res. Public Health* **2,** 147–155 (2005).
101. Peirson, S. & Foster, R. G. Melanopsin: another way of signaling light. *Neuron* **49,** 331–339 (2006).
102. Wicks, N. L., Chan, J. W., Najera, J. A., Ciriello, J. M. & Oancea, E. UVA phototransduction drives early melanin synthesis in human melanocytes. *Curr. Biol.* **21,** 1906–1911 (2011).
103. Provencio, I. & Foster, R. G. Circadian rhythms in mice can be regulated by photoreceptors with cone-like characteristics. *Brain Res.* **694,** 183–190 (1995).
104. Organisciak, D. T. & Vaughan, D. K. Retinal light damage: mechanisms and protection.

Prog. Retin. Eye Res. **29,** 113–134 (2010).

105. Feigin, M. E., Xue, B., Hammell, M. C. & Muthuswamy, S. K. G-protein-coupled receptor GPR161 is overexpressed in breast cancer and is a promoter of cell proliferation and invasion. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 4191–4196 (2014).
106. Pal, K., Hwang, S.-H., Somatilaka, B., Badgandi, H., Jackson, P. K., DeFea, K. & Mukhopadhyay, S. Smoothened determines β-arrestin-mediated removal of the G protein-coupled receptor Gpr161 from the primary cilium. *J. Cell Biol.* **212,** 861–875 (2016).
107. Li, B. I., Matteson, P. G., Ababon, M. F., Nato, A. Q., Jr, Lin, Y., Nanda, V., Matise, T. C. & Millonig, J. H. The orphan GPCR, Gpr161, regulates the retinoic acid and canonical Wnt pathways during neurulation. *Dev. Biol.* **402,** 17–31 (2015).
108. Xu, X.-R., Li, X.-Y., Li, X.-Z. & Li, H.-B. Degradation of melatonin by UV, UV/H2O2, Fe2+/H2O2 and UV/Fe2+/H2O2 processes. *Sep. Purif. Technol.* **68,** 261–266 (2009).
109. Yan, G., Yu, L., Jiang, S. & Zhu, J. Melatonin antagonizes oxidative stress-induced mitochondrial dysfunction in retinal pigmented epithelium cells via melatonin receptor 1 (MT1). *J. Toxicol. Sci.* **43,** 659–669 (2018).
110. Galano, A., Tan, D. X. & Reiter, R. J. Melatonin as a natural ally against oxidative stress: a physicochemical examination. *J. Pineal Res.* **51,** 1–16 (2011).
111. Fischer, T. W., Slominski, A., Zmijewski, M. A., Reiter, R. J. & Paus, R. Melatonin as a major skin protectant: from free radical scavenging to DNA damage repair. *Exp. Dermatol.* **17,** 713–730 (2008).
112. Konta, T., Xu, Q., Furusu, A., Nakayama, K. & Kitamura, M. Selective roles of retinoic acid receptor and retinoid x receptor in the suppression of apoptosis by all-trans-retinoic acid. *J. Biol. Chem.* **276,** 12697–12701 (2001).
113. McNamara, P., Seo, S.-B., Rudic, R. D., Sehgal, A., Chakravarti, D. & FitzGerald, G. A. Regulation of CLOCK and MOP4 by Nuclear Hormone Receptors in the Vasculature: A Humoral Mechanism to Reset a Peripheral Clock. *Cell* **105,** 877–889 (2001).
114. Mangelsdorf, D. J., Ong, E. S., Dyck, J. A. & Evans, R. M. Nuclear receptor that identifies a novel retinoic acid response pathway. *Nature* **345,** 224–229 (1990).
115. Koyanagi, M., Takano, K., Tsukamoto, H., Ohtsu, K., Tokunaga, F. & Terakita, A.

Jellyfish vision starts with cAMP signaling mediated by opsin-G(s) cascade. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 15576–15580 (2008).

116. Peirson, S. N., Halford, S. & Foster, R. G. The evolution of irradiance detection: melanopsin and the non-visual opsins. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364,** 2849–2865 (2009).
117. Bertolesi, G. E., Hehr, C. L. & McFarlane, S. Melanopsin photoreception in the eye regulates light-induced skin colour changes through the production of α-MSH in the pituitary gland. *Pigment Cell Melanoma Res.* **28,** 559–571 (2015).
118. Iyengar, B. The melanocyte photosensory system in the human skin. *Springerplus* **2,** 158 (2013).
119. Semo, M. 'ayan, Gias, C., Ahmado, A., Sugano, E., Allen, A. E., Lawrence, J. M., Tomita, H., Coffey, P. J. & Vugler, A. A. Dissecting a role for melanopsin in behavioural light aversion reveals a response independent of conventional photoreception. *PLoS One* **5,** e15009 (2010).
120. Du, Y., Veenstra, A., Palczewski, K. & Kern, T. S. Photoreceptor cells are major contributors to diabetes-induced oxidative stress and local inflammation in the retina. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 16586–16591 (2013).
121. Organisciak, D. T., Darrow, R. A., Barsalou, L., Darrow, R. M. & Lininger, L. A. Light-induced damage in the retina: differential effects of dimethylthiourea on photoreceptor survival, apoptosis and DNA oxidation. *Photochem. Photobiol.* **70,** 261–268 (1999).
122. Moreno-Manzano, V., Ishikawa, Y., Lucio-Cazana, J. & Kitamura, M. Suppression of apoptosis by all-trans-retinoic acid. Dual intervention in the c-Jun n-terminal kinase-AP-1 pathway. *J. Biol. Chem.* **274,** 20251–20258 (1999).
123. Kitamura, M., Ishikawa, Y., Moreno-Manzano, V., Xu, Q., Konta, T., Lucio-Cazana, J., Furusu, A. & Nakayama, K. Intervention by retinoic acid in oxidative stress-induced apoptosis. *Nephrol. Dial. Transplant* **17 Suppl 9,** 84–87 (2002).
124. Nishina, S., Kohsaka, S., Yamaguchi, Y., Handa, H., Kawakami, A., Fujisawa, H. & Azuma, N. PAX6 expression in the developing human eye. *Br. J. Ophthalmol.* **83,** 723–727 (1999).
125. Ashery-Padan, R. & Gruss, P. Pax6 lights-up the way for eye development. *Curr. Opin. Cell Biol.* **13,** 706–714 (2001).

126. Mikkola, I., Bruun, J. A., Bjorkoy, G., Holm, T. & Johansen, T. Phosphorylation of the transactivation domain of Pax6 by extracellular signal-regulated kinase and p38 mitogen-activated protein kinase. *J. Biol. Chem.* **274,** 15115–15126 (1999).
127. Laggner, M., Pollreisz, A., Schmidinger, G., Schmidt-Erfurth, U. & Chen, Y.-T. Autophagy mediates cell cycle response by regulating nucleocytoplasmic transport of PAX6 in limbal stem cells under ultraviolet-A stress. *PLoS One* **12,** e0180868 (2017).
128. Cvekl, A., Zhao, Y., McGreal, R., Xie, Q., Gu, X. & Zheng, D. Evolutionary Origins of Pax6 Control of Crystallin Genes. *Genome Biol. Evol.* **9,** 2075–2092 (2017).
129. Chang, B. S. & Donoghue, M. J. Recreating ancestral proteins. *Trends Ecol. Evol.* **15,** 109–114 (2000).
130. Serb, J. M. & Eernisse, D. J. Charting evolution's trajectory: using molluscan eye diversity to understand parallel and convergent evolution. *Evolution: Education and Outreach* **1,** 439–447 (2008).
131. Plachetzki, D. C., Fong, C. R. & Oakley, T. H. Cnidocyte discharge is regulated by light and opsin-mediated phototransduction. *BMC Biol.* **10,** 17 (2012).
132. Blevins, E. & Johnsen, S. Spatial vision in the echinoid genus Echinometra. *J. Exp. Biol.* **207,** 4249–4253 (2004).
133. de Assis, L. V. M., Moraes, M. N., Magalhães-Marques, K. K. & Castrucci, A. M. de L. Melanopsin and rhodopsin mediate UVA-induced immediate pigment darkening: Unravelling the photosensitive system of the skin. *Eur. J. Cell Biol.* **97,** 150–162 (2018).
134. Ramirez, M. D., Speiser, D. I., Pankey, M. S. & Oakley, T. H. Understanding the dermal light sense in the context of integrative photoreceptor cell biology. *Vis. Neurosci.* **28,** 265–279 (2011).
135. Galibert, M., Carreira, S. & Goding, C. R. The Usf-1 transcription factor is a novel target for the stress-responsive p38 kinase and mediates UV-induced Tyrosinase expression. *EMBO J.* **20,** 5022–5031 (2001).
136. Tamaru, T., Hattori, M., Ninomiya, Y., Kawamura, G., Varès, G., Honda, K., Mishra, D. P., Wang, B., Benjamin, I., Sassone-Corsi, P., Ozawa, T. & Takamatsu, K. ROS stress resets circadian clocks to coordinate pro-survival signals. *PLoS One* **8,** e82006 (2013).
137. Engelen, E., Janssens, R. C., Yagita, K., Smits, V. A. J., van der Horst, G. T. J. &

Tamanini, F. Mammalian TIMELESS is involved in period determination and DNA damage-dependent phase advancing of the circadian clock. *PLoS One* **8,** e56623 (2013).

138. Sancar, A. Cryptochrome: the second photoactive pigment in the eye and its role in circadian photoreception. *Annu. Rev. Biochem.* **69,** 31–67 (2000).
139. Weber, S. Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase. *Biochim. Biophys. Acta* **1707,** 1–23 (2005).
140. Arendt, D., Musser, J. M., Baker, C. V. H., Bergman, A., Cepko, C., Erwin, D. H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M. D. & Wagner, G. P. The origin and evolution of cell types. *Nat. Rev. Genet.* **17,** 744–757 (2016).
141. Feder, M. E. & Hofmann, G. E. Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu. Rev. Physiol.* **61,** 243–282 (1999).
142. Kültz, D. Molecular and evolutionary basis of the cellular stress response. *Annu. Rev. Physiol.* **67,** 225–257 (2005).
143. Palmer, A. R. Symmetry breaking and the evolution of development. *Science* **306,** 828–833 (2004).
144. Ramaswami, M. Network Plasticity in Adaptive Filtering and Behavioral Habituation. *Neuron* **82,** 1216–1229 (2014).
145. Chevin, L.-M., Lande, R. & Mace, G. M. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biol.* **8,** e1000357 (2010).
146. Latek, D., Modzelewska, A., Trzaskowski, B., Palczewski, K. & Filipek, S. G protein-coupled receptors--recent advances. *Acta Biochim. Pol.* **59,** 515–529 (2012).
147. Gaub, B. M., Berry, M. H., Visel, M., Holt, A., Isacoff, E. Y. & Flannery, J. G. Optogenetic Retinal Gene Therapy with the Light Gated GPCR Vertebrate Rhodopsin. *Methods Mol. Biol.* **1715,** 177–189 (2018).
148. Spangler, S. M. & Bruchas, M. R. Optogenetic approaches for dissecting neuromodulation and GPCR signaling in neural circuits. *Curr. Opin. Pharmacol.* **32,** 56–70 (2017).
149. Sriram, K. & Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol. Pharmacol.* **93,** 251–258 (2018).
150. Strotmann, R., Schröck, K., Böselt, I., Stäubert, C., Russ, A. & Schöneberg, T.

Evolution of GPCR: change and continuity. *Mol. Cell. Endocrinol.* **331,** 170–178 (2011).

151. Fang, Y., Kenakin, T. & Liu, C. Editorial: Orphan GPCRs As Emerging Drug Targets. *Front. Pharmacol.* **6,** 295 (2015).
152. Hu, G.-M., Mai, T.-L. & Chen, C.-M. Visualizing the GPCR Network: Classification and Evolution. *Sci. Rep.* **7,** 15495 (2017).
153. Schonenbach, N. S., Hussain, S. & O'Malley, M. A. Structure and function of G protein-coupled receptor oligomers: implications for drug discovery: Studying GPCR Oligomer Function. *WIREs Nanomed Nanobiotechnol* **7,** 408–427 (2015).
154. Lefkowitz, R. J. The superfamily of heptahelical receptors. *Nat. Cell Biol.* **2,** E133–6 (2000).
155. Schiöth, H. B. & Fredriksson, R. The GRAFS classification system of G-protein coupled receptors in comparative perspective. *Gen. Comp. Endocrinol.* **142,** 94–101 (2005).
156. Desmond Ramirez, M., Pairett, A. N., Sabrina Pankey, M., Serb, J. M., Speiser, D. I., Swafford, A. J. & Oakley, T. H. The last common ancestor of most bilaterian animals possessed at least 9 opsins. *bioRxiv* 052902 (2016). doi:10.1101/052902
157. Devine, E. L., Oprian, D. D. & Theobald, D. L. Relocating the active-site lysine in rhodopsin and implications for evolution of retinylidene proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 13351–13355 (2013).
158. Faggionato, D. & Serb, J. M. Strategy to Identify and Test Putative Light-Sensitive Non-Opsin G-Protein-Coupled Receptors: A Case Study. *Biol. Bull.* **233,** 70–82 (2017).
159. Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17,** 1246–1247 (2001).
160. Beaulieu, J. M., O'Meara, B. C. & Donoghue, M. J. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Syst. Biol.* **62,** 725–737 (2013).
161. Maddison, W. P. & FitzJohn, R. G. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst. Biol.* **64,** 127–136 (2015).
162. Pagel, M. The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. *Syst. Biol.* **48,** 612–622 (1999).

163. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gkz239
164. James, T. Y., Porter, T. M. & Wallace Martin, W. in *Systematics and Evolution* 177–207 (Springer Berlin Heidelberg, 2014).
165. Longcore, J. E., Pessier, A. P. & Nichols, D. K. Batrachochytrium Dendrobatidis gen. et sp. nov., a Chytrid Pathogenic to Amphibians. *Mycologia* **91,** 219–227 (1999).
166. Olson, L. W. Allomyces- A Different Fungus. *Opera Botanical* **73,** 1–96 (1984).
167. Dukes, P. D. & Apple, J. L. Chemotaxis of zoospores of Phytophthora parasitica var. nicotianae by plant roots and certain chemical solutions. *Phytopathology* **51,** 195–197 (1961).
168. Suberkropp, K. F. & Cantino, E. C. Utilization of endogenous reserves by swimming zoospores of Blastocladiella emersonii. *Arch. Mikrobiol.* **89,** 205–221 (1973).
169. Machlis, L. Zoospore Chemotaxis in the Watermold Allomyces. *Physiol. Plant.* **22,** 126–139 (1969).
170. Morris, B. M., Reid, B. & Gow, N. A. R. Electrotaxis of zoospores of Phytophthora palmivora at physiologically relevant field strengths. *Plant Cell Environ.* **15,** 645–653 (1992).
171. Pommerville, J. C. & Olson, L. W. Negative Chemotaxis of Gametes and Zoospores of Allomyces. *Microbiology* **133,** 2573–2579 (1987).
172. Epifanio, C. E., Masse, A. K. & Garvine, R. W. Transport of blue crab larvae by surface currents off Delaware Bay, USA. *Marine ecology progress series. Oldendorf* **54,** 35–41 (1989).
173. Avelar, G. M., Schumacher, R. I., Zaini, P. A., Leonard, G., Richards, T. A. & Gomes, S. L. A rhodopsin-guanylyl cyclase gene fusion functions in visual perception in a fungus. *Curr. Biol.* **24,** 1234–1240 (2014).
174. Avelar, G. M., Glaser, T., Leonard, G., Richards, T. A., Ulrich, H. & Gomes, S. L. A Cyclic GMP-Dependent K+ Channel in the Blastocladiomycete Fungus Blastocladiella emersonii. *Eukaryot. Cell* **14,** 958–963 (2015).
175. Gao, S., Nagpal, J., Schneider, M. W., Kozjak-Pavlovic, V., Nagel, G. & Gottschalk, A. Optogenetic manipulation of cGMP in cells and animals by the tightly light-regulated guanylyl-cyclase opsin CyclOp. *Nat. Commun.* **6,** 8046 (2015).

176. Machlis, L. Growth and Nutrition of Water Molds in the Subgenus Euallomyces. II. Optimal Composition of the Minimal Medium. *Am. J. Bot.* **40,** 450–460 (1953).
177. Saranak, J. & Foster, K. Rhodopsin guides fungal phototaxis. *Nature* **387,** 465–466 (1997).
178. Carlile, M. J. & Machlis, L. A Comparative Study of the Chemotaxis of the Motile Phases of Allomyces. *Am. J. Bot.* **52,** 484–486 (1965).
179. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
180. Oakley, T. H., Alexandrou, M. A., Ngo, R., Pankey, M. S., Churchill, C. K., Chen, W. & Lopker, K. B. Osiris: accessible and reproducible phylogenetic and phylogenomic analyses within the Galaxy workflow management system. *BMC Bioinformatics* **15,** 230 (2014).
181. Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J. & Others. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15,** 1451–1455 (2005).
182. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
183. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14,** 178–192 (2013).
184. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).
185. He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W.-H. & Hu, S. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* **44,** W236–41 (2016).
186. Klare, J. P., Gordeliy, V. I., Labahn, J., Büldt, G., Steinhoff, H.-J. & Engelhard, M. The archaeal sensory rhodopsin II/transducer complex: a model for transmembrane signal transfer. *FEBS Lett.* **564,** 219–224 (2004).
187. Larusso, N. D., Ruttenberg, B. E., Singh, A. K. & Oakley, T. H. Type II opsins: evolutionary origin by internal domain duplication? *J. Mol. Evol.* **66,** 417–423 (2008).
188. Cavalier-Smith, T., Chao, E. E., Snell, E. A., Berney, C., Fiore-Donno, A. M. & Lewis,

R. Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. *Mol. Phylogenet. Evol.* **81,** 71–85 (2014).

189. Miño, G. L., Koehl, M. A. R., King, N. & Stocker, R. Finding patches in a heterogeneous aquatic environment: pH-taxis by the dispersal stage of choanoflagellates. *Limnol. Oceanogr.* **2,** 37–46 (2017).
190. Kirkegaard, J. B., Bouillant, A., Marron, A. O., Leptos, K. C. & Goldstein, R. E. Aerotaxis in the closest relatives of animals. *Elife* **5,** (2016).