# Methods

Previously published opsin sequences were downloaded and publicly available assembled and raw RNAseq data were captured from the National Center for Biotechnology Information (NCBI) for crustacean orders. Taxonomic definitions for all crustaceans followed those set by Bracken-Grissom and Wolfe [1]. Raw sequence data was trimmed using Trimmomatic default parameters [2], and assembled using Trinity (v2.11.0) software [3] on the National Center for Genome Analysis Support’s Carbonate computing cluster at Indiana University. Preassembled and Trinity assembled data were loaded to the University of California Santa Barbara Galaxy instance for analysis with the Phylogenetically Informed Annotation (PIA) pipeline [4]. In brief, open reading frames (ORFs) of at least 30 amino acids were generated from transcripts, and searches were performed using NCBI’s Basic Local Alignment Tool (BLAST) against the Light Interacting Toolkit (LIT), a collection of genes containing known opsin sequences. Newick tree files were generated for proteins matching rhabdomeric opsin genes (r-opsins). When over 200 opsin hits were found, ORF cutoffs were increased to 150 amino acids to reduce the number of opsin fragments in the analysis.

Opsins generated from each transcriptome were analyzed using Geneious (R10.2.6) software [5] for alignment using MAFFT G-INS-i [6] algorithms to check for redundancy. Duplicate opsin sequences within single transcriptomes with over 98% protein similarity were discarded, and remaining opsin sequences were searched using BLAST against the NCBI nr/nt database to confirm opsin identity. Opsins were checked for specific motifs including the DRY sequence and the terminal lysine residue which acts as a retinal binding site in opsin proteins as well as the R(E/D)QAKK sequence found in arthropod visual opsins.

An initial maximum likelihood phylogeny was generated using RaxML [7] on the CIPRES platform [8] to determine approximate opsin placement and identify opsin fragments which may have originated from the same protein sequence. Fragments under 150 amino acids with overlap were then combined. When sequence overlap did not occur for sequence fragments under 150 amino acids within the same opsin clade, fragments were merged and ‘X’ was used to denote the missing interior portion of the sequence. Remaining sequence fragments under 150 amino acids were discarded if they were not the only representative per species or opsin group. A final opsin phylogeny was generated from remaining sequences in the manner described above. Closely related GPCR sequences from the placozoan *Trichoplax adhaerens* were used to root the crustacean opsin phylogeny. These GPCRs are opsin-like in sequence but lack the terminal lysine residue in the seventh transmembrane helix.

**References**

1. Bracken‐Grissom HD, Wolfe JM. 2020 *The Pancrustacean Conundrum: A Conflicted Phylogeny with Emphasis on Crustacea*. 8th edn. Oxford University Press. See https://books.google.com/books?id=ix7ZDwAAQBAJ.

2. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)

3. Grabherr MG., Brian J. Haas, Moran Yassour Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W. N, Friedman and AR. 2013 Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**, 644–652. (doi:10.1038/nbt.1883.Trinity)

4. Speiser DI *et al.* 2014 Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinformatics* **15**, 1–12. (doi:10.1186/s12859-014-0350-x)

5. Kearse M *et al.* 2012 Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649. (doi:10.1093/bioinformatics/bts199)

6. Katoh K, Misawa K, Kuma KI, Miyata T. 2002 MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066. (doi:10.1093/nar/gkf436)

7. Stamatakis A. 2014 RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)

8. Miller MA, Pfeiffer W, Schwartz T. 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gatew. Comput. Environ. Work. GCE 2010* (doi:10.1109/GCE.2010.5676129)