Research Statement
Kyle T. David

Introduction. Arrow worms (chaetognaths) are a phylum of free-living predatory marine plankton. They are the second most abundant zooplankton group and represent a significant proportion of marine biomass[1]. Despite their abundance and ecological significance, arrow worms are very poorly understood. Charles Darwin[2] commented on the "remarkable… obscurity of their affinities" and in the 174 years since, arrow worms have been placed in many different bilaterian groups, including nematodes, annelids, molluscs, crustaceans, arachnids, and chordates[1]. Most modern molecular analyses place arrow worms within protostomes (Fig. 1), but a consensus has not yet been reached[3,4,5]. Internal relationships within the phylum are similarly ambiguous[1,6]. I will broadly sequence arrow worm transcriptomes to determine relationships within and outside the group and use these transcriptomic data to elucidate the evolution of development within animals.

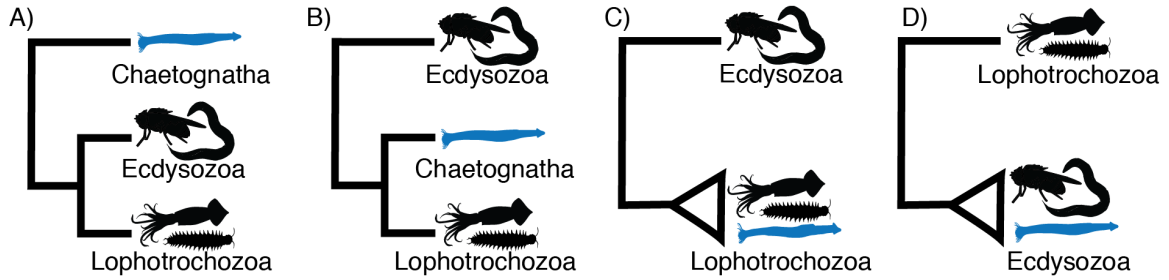

Figure 1. Four conflicting topologies that have all been recently recovered from molecular phylogenies[3,4,5]

Aim 1: Infer a Robust Chaetognath Species Tree. The inclusion of arrow worm transcriptomes to a larger protostome dataset will add significant power to phylogenetic analyses and resolve evolutionary relationships that have confounded biologists for hundreds of years[1].

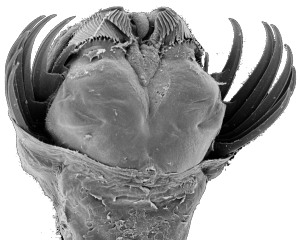
Photo: Michael Le Roux

Aim 2: Explore the Origins of Bilaterian Development. Though considered protostomes by most modern researchers, arrow worms possess deuterostome-like development (enterocoely, secondary mouth formation, radial cleavage). This makes arrow worms uniquely positioned to explore novel questions on the origins of development in bilaterians. If arrow worms are indeed the sister-group to protostomes (Fig. 1A), it is likely that deuterostomous development was present in the bilaterian ancestor. Alternatively, if arrow worms are instead nested somewhere within protostomes (Fig. 1B-D), it is likely these features are an example of convergent evolution.

Methods. The Halanych Lab has an established history of collecting and sequencing transcriptomes of non-model marine invertebrates. Our lab has sequenced and annotated 59 transcriptomes listed on NCBI's Sequence Read Archive (SRA). I will collect specimens of at least one representative from each of the 11 arrow worm families recognized by the World Register of Marine Species. I will be able to accomplish this through a previously established relationship with Dr. Janet Voight, an Associate Curator at the Field Museum of Natural History, who has access to these groups as well as with samples previously collected from my lab. I also aim to participate in the Graduate Research Internship Program (GRIP) available to GRFP fellows, which would allow me the opportunity to intern at the Smithsonian under Dr. Jon Norenburg in order to study and sample their collections. It may be necessary to collect from the field as well, which will be possible through research cruises like the Icy Inverts Antarctica

Cruises with which my lab has a history of participation. RNA samples will be extracted, prepared, and sequenced through previously validated Halanych lab protocols[3]. The generalized bioinformatics pipeline is represented in Figure 2. I will use the skills I have learned from my recent participation in the Workshop on Molecular Evolution to infer maximum likelihood and Bayesian gene and species trees while using a variety of model assumptions and parameters in a comparative approach. Several deuterostomes (sea urchin, acorn worm, mouse, human) will serve as an outgroup.

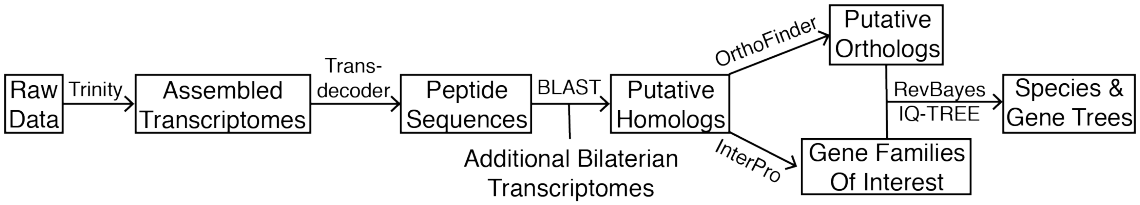

Figure 2. Simplified bioinformatics workflow for species and gene tree inference

Intellectual Merit. There is currently only a single arrow worm sequence on the SRA. This project will increase the amount of genetic data for this poorly understood group by an order of magnitude. A well resolved tree will also provide a phylogenetic framework for understanding the evolution of several key features in animal evolution and provide evidence for the ancestral bilaterian state. Arrow worms are known to have many unique features including lamellar photoreceptors[7] and mosaic *hox* genes[8] in addition to a putative whole genome duplication event[9]. Increasing the availability of coding sequences in this group will allow myself and others to explore expansions/losses of several significant gene families (e.g., *opsin* and *hox* genes) and test for evidence of whole genome duplication within this enigmatic group.

Broader Impacts. Results will be disseminated widely to expert (i.e., publications, symposia, talks) and non-expert (i.e., Skype a Scientist, outreach events, for details see **Personal Statement**) audiences. Through connections already established with faculty, I will also be able present my work as a guest lecturer through Auburn University's Alabama Prison Arts + Education Project, which provides pre-college classes to prisoners. In 2016, the New York Times reported that inmates who participate in college programs have a 4% re-offence rate, creating a 500% return on investment in prison education initiatives. Alabama law does not allow prisoners to take remote classes meaning courses must be run on-site and in-person, something that would only be possible for me to participate in with GRFP support. All assembled transcriptomes and raw reads from this project will be made publically available on the SRA. I am committed to open-source software and will continue to upload all scripts required to reproduce analyses to my public repository (github.com/KyleTDavid). I will also mentor students through the NSF Research Experience for Undergraduates (REU) program, of which my lab is a participating member in computational biology. Students will receive a primer in basic programming skills and an introduction to phylogenomic workflows, as well as an opportunity to pursue independent projects.

[1] Bone & Pierrot-Bults. (1991). Oxford University Press. **[2]** Darwin. (1844). *Journal of Natural History*. **[3]** Kocot et al. (2017). *Systematic biology*. **[4]** Marlétaz et al. (2006). *Current Biology*. **[5]** Matus et al. (2006). *Current Biology*. **[6]** Gasmi et al. (2014). *Frontiers in zoology*. **[7]** Goto et al. (1984). *Cell and tissue research.* **[8]** Papillon et al. (2003). *Development genes and evolution*. **[9]** Marlétaz et al. (2008). *Genome biology*.