

# Are Ultraconserved Elements an Informative Phylogenetic Marker in Molluscs? Emily Pabst<sup>1</sup> & Kevin Kocot<sup>1,2</sup>

INTRODUCTION

Mollusca is an important phylum of invertebrates with over 100,000 described species in eight morphologically disparate major lineages (Kocot et al., 2011). Molluscs have many significant uses to humans as food, producers of pearls and shells, and as model organisms in studies of brain organization, learning, and memory. However, they can also be harmful as pests of agriculture, invasive species that can damage ecosystems, and vectors of parasites. Despite their importance and prevalence, there is not a clear consensus on the evolutionary relationships among the eight major lineages of molluscs (Kocot et al., 2011; Kocot et al., 2013; Kocot, 2013). Earlier research has found that the genomes of many animals contain ultraconserved elements (UCEs), typically regulatory sequences that allow targeted sequencing (Stephen et al., 2008; Faircloth et al., 2012; Faircloth et al., 2013; Faircloth et al., 2015; Starrett et al., 2017). The use of UCEs as molecular markers for reconstructing evolutionary relationships has expanded possibilities for phylogenetic studies because they have recovered well-supported phylogenies that provide information on both shallow and deep relationships.



Figure 1. (A) Octopus (Getty Images). (B) Clam (HamaHamaOysters). (C) Snail (Mother Naure Network). (D) Aplacophran (Washington.edu).

### OBJECTIVES

The goal of this project was to test whether or not UCEs are present in mollusc genomes and whether they have utility for higher-level molluscan phylogenetics. Inferring evolutionary relationships among the major lineages of Mollusca will allow us to deduce which species likely share similar characteristics. This will aid in identification of additional biomedically important species or others that might be potentially problematic.



Office of the Vice President for Research & Economic Development Undergraduate Research

## The University of Alabama <sup>1</sup>Department of Biological Sciences, <sup>2</sup>Alabama Museum of Natural History

### METHODS

### UCE identification:

We used the Phyluce (Faircloth, 2017) pipeline with assistance from the developer, Dr. Brant Faircloth, to screen for potential UCEs in the genomes of 10 molluscs and five other taxa representing putative close relatives of Mollusca. We downloaded the genome assemblies for the 15 organisms from NCBI SRA. ART (Huang et al., 2012) was used to simulate reads from the genomes to mimic what would result from Illumina DNA sequencing. We then aligned these reads to the base genome, Lottia gigantea. Initial conserved locus identification involved finding loci that were potentially orthologous between Lottia gigantea and the rest of the taxa. This involved merging together potential conserved regions adjacent in the genomes and removing duplicates from the set. Sequences where the alignment to the base genome was shorter than 80 bp were removed in addition to sequences where greater than 25% was identified as repetitive DNA. At this point, sequences present across multiple genomes were considered UCEs (Figure 2). Next, FASTA sequences of 160 bp were extracted from the base genome for 4,759 of the sequences (those shared by *Lottia gigantea* and at least 10 other taxa).

### Preliminary bait set:

Using these sequences, two baits were selected per UCE from the Lottia gigantea genome with 3x tiling that overlapped in the center of the UCE. Baits with over 25% repeat content and a GC percentage outside of the range of 30-70 were removed and so were duplicate baits. Next we located these baits in the genomes of the other taxa by aligning the baits and extracting FASTA sequences of 180 bp for each UCE. Again, we checked to see which sequences were present in all of the taxa (Figure 2).

### Final bait set:

Baits were selected again as described above for Lottia gigantea but all taxa were used. Once we had the bait set, we selected the baits for the 10 taxa in Mollusca to be synthesized. This final bait set was aligned to all genomes, and FASTA sequences (400 bp on either side of loci center) were extracted for the remaining sequences to identify the baits in the contigs. We then constructed one matrix of all of the filtered sequences in a single FASTA file. The sequences were aligned with MAFFT (Katoh et al., 2005), trimmed with Gblocks (Castresana, 2000), and then a 75% complete matrix was generated with 2,818,367 nucleotide positions.

### Tree construction:

Using the resulting matrix, we inferred a UCE-based phylogeny by reconstructing a maximum likelihood tree with raxmlHPC-PTHREADS-SSE3 (Stamatakis, 2014) using the best-fitting model for each partition with a GAMMA model of rate heterogeneity. Topological robustness was assessed using rapid bootstrapping with the optimal number of replicates as determined using the RAxML autoMRE criterion.

## FIGURES

Lottia gigantea +	Number of	Number of	
Number of	Shared Loci	Shared Loci	
Additional Taxa	(UCEs)	(Bait Set)	
1	805,830	1,736	
2	122,159	1,643	
3	71,653	1,560	
4	48,025	1,484	
5	32,998	1,419	
6	22,720	1,355	
7	15,721	1,292	
8	10,741	1,215	
9	7,214	1,114	
10	4,759	985	
11	3,078	804	
12	1,945	587	
13	1,166	339	
14	690	113	
15	325	16	



**Figure 2.** Number of taxa that share sequences with base genome in UCE identification stage and bait targeting stage.

**Figure 3.** Maximum likelihood tree with support values shown. The nonmolluscs were used to root the tree.



		2
Bathymodiolus platifrons Modiolus philippinarum Mytilus galloprovincialis Mizuhopecten yessoensis Pinctada fucata Crassostrea gigas	BIVALVIA Pteriomorphia	Mollusca
Octopus bimaculoides	CEPHALOPODA	
Phoronis australis Lingula anatina Notospermus geniculatus Capitella teleta Helobdella robust	Outgroup	

We found 4,759 UCEs shared among at least 10 taxa sampled in our study and 325 shared across all 15 organisms used in the analysis. Using a final set of 1,000 UCEs, we assembled and analyzed a matrix with 2,818,367 nucleotide positions and 75% completeness, generating a tree with generally well-supported relationships that are consistent with the current understanding of molluscan evolution (Figure 3). We recovered Mollusca with maximal support (bs = 100). Within Mollusca, Pleistomollusca (Bivalva + Gastropoda) was recovered sister to Cephalopoda (bs = 90) in agreement with other studies to date (Kocot et al., 2011; Kocot, 2013). Within Bivalvia we recovered Pteriomorpha (bs = 100), represented here by Bathymodiolus platifrons, Modiolus philippinarum, Mytilus galloprovincialis, Mizuhopecten yessoensis, Pinctada fucata, and Crassostrea gigas. These relationships are consistent with a previous phylogenomic investigation of bivalve evolution (González et al., 2015). However, the relationships we recovered within Gastropoda are in conflict with previous studies (Colgan et al., 2000; Kocot et al., 2011). We would have expected a clade of Patellogastropoda (*Lottia gigantea*) + Vetigastropoda (*Haliotis* discus), but instead recovered Heterobranchia (Radix auricularia) + Patellogastropoda (Lottia gigantea). One possible explanation is that this is due to algal contamination in the genome of the herbivorous *Haliotis discus*.

These results demonstrate that UCEs are an informative phylogenetic marker in molluscs. Our future plans include sequencing the genome of an aplacophoran, a group of molluscs not represented in our dataset, and re-constructing a tree with this and other additional molluscs to further discern how these taxa are related. By applying our methods to Mollusca as a whole, we can find other model organisms for biomedical studies. Additionally, we can also learn how to better protect ourselves against additional molluscs that could have detrimental effects for humans.



### **RESULTS & DISCUSSION**

### CONCLUSION

## REFERENCES

Castresana, J., 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol. Biol. Evol. 17, 540–552.

https://doi.org/10.1093/oxfordjournals.molbev.ao26334 Colgan, D.J., Ponder, W.F., Eggler, P.E., 2000. Gastropod evolutionary rates and phylogenetic relationships assessed using partial 28S rDNA and histone H3 sequences. Zool. Scr. 29, 29–63. https://doi.org/10.1046/j.1463-6409.2000.00021.x Faircloth, B.C., 2017. Identifying conserved genomic elements and designing universal bait sets to enrich them. Methods Ecol. Evol. 8, 1103–1112.

https://doi.org/10.1111/2041-210X.12754 Faircloth, B.C., Branstetter, M.G., White, N.D., Brady, S.G., 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Mol. Ecol. Resour. 15, 489–501. https://doi.org/10.1111/1755-0998.12328 Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. Syst. Biol. 61, 717–726. https://doi.org/10.1093/sysbio/syso04 Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E., 2013. A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted

Sequencing of Ultraconserved Elements (UCEs). PLOS ONE 8, e65923. https://doi.org/10.1371/journal.pone.oo65923 Huang, W., Li, L., Myers, J.R., Marth, G.T., 2012. ART: a next-generation sequencing read simulator. Bioinformatics 28, 593–594. https://doi.org/10.1093/bioinformatics/btr708 González, V.L., Andrade, S.C.S., Bieler, R., Collins, T.M., Dunn, C.W., Mikkelsen, P.M., Taylor, J.D., Giribet, G., 2015. A phylogenetic backbone for Bivalvia: an

RNA-seq approach. Proc R Soc B 282, 20142332. https://doi.org/10.1098/rspb.2014.2332 Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518. https://doi.org/10.1093/nar/gki198 Kocot, K.M., 2013. Recent Advances and Unanswered Questions in Deep Molluscan Phylogenetics. Am. Malacol. Bull. 31, 195–208.

https://doi.org/10.4003/006.031.0112 Kocot, K.M., Cannon, J.T., Todt, C., Citarella, M.R., Kohn, A.B., Meyer, A., Santos, S.R., Schander, C., Moroz, L.L., Lieb, B., Halanych, K.M., 2011. Phylogenomics reveals deep molluscan relationships. Nature 477, 452–456. https://doi.org/10.1038/nature10382 Kocot, K.M., Halanych, K.M., Krug, P.J., 2013. Phylogenomics supports Panpulmonata: Opisthobranch paraphyly and key evolutionary steps in a major radiation of gastropod molluscs. Mol. Phylogenet. Evol. 69, 764–771. https://doi.org/10.1016/j.ympev.2013.07.001 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

https://doi.org/10.1093/bioinformatics/btu033 Starrett, J., Derkarabetian, S., Hedin, M., Bryson, R.W., McCormack, J.E., Faircloth, B.C., 2017. High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. Mol. Ecol. Resour. 17, 812–823. https://doi.org/10.1111/1755-0998.12621 Stephen, S., Pheasant, M., Makunin, I.V., Mattick, J.S., 2008. Large-Scale Appearance of Ultraconserved Elements in Tetrapod Genomes and Slowdown of the Molecular Clock. Mol. Biol. Evol. 25, 402–408. https://doi.org/10.1093/molbev/msm268

## THE UNIVERSITY OF ALABAMA®