Are ultraconserved elements an informative marker for reconstructing deep molluscan phylogeny?

Emily A. Pabst[1] & Kevin M. Kocot[1,2]

University of Alabama Department of Biological Sciences[1] and Alabama Museum of Natural History[2]

Introduction

Because of their incredible diversity, economic value, and ecological importance, there has been a great deal of interest in resolving evolutionary relationships among the major lineages of Mollusca (1). Studies in the last decade employing nuclear protein-coding genes have greatly advanced understanding of molluscan evolution (2-4), but important questions such as placement of Monoplacophora remain unanswered (Figure 1). The genomes of diverse animals have been shown to contain ultraconserved elements (UCEs), regulatory sequences with utility as phylogenetic markers (5-11). Thus, the goal of this work was to determine whether UCEs are present in mollusc and related lophotrochozoan (=spiralian) genomes, and if they have utility for inferring relationships among major molluscan lineages, which likely diverged in the late Precambrian (4).

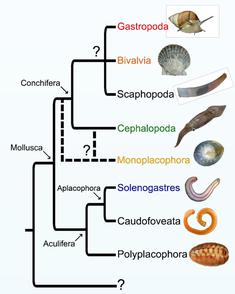

Figure 1. Consensus of molluscan phylogeny based on studies employing nuclear protein-coding genes (2-4).

Methods

A single specimen of *Neomenia megatrapezata* was collected in Antartica by K. Kocot. The genome was sequenced by New York Genome Center using one lane of an Illumina HiSeq X with 2 X 150 bp reads. We used Phyluce (6) to screen for UCEs in published and newly sequenced genomes from 21 molluscs and five outgroups. Resulting UCEs were aligned with MAFFT (12) and trimmed with Gblocks (13), keeping only alignments sampled for 20/26 taxa. UCEs were sorted by evolutionary rate and concatenated matrices were assembled for all UCEs and the slowest ½ of the UCEs, and additional matrices were created from the MARE-reduced versions of both of these datasets. Maximum likelihood (ML) trees were inferred for each partitioned matrix using RAxML 8 (14) with the GTR+G4 model and rapid bootstrapping with the number of replicates determined using the autoMRE criterion. We then used SWSC-EN (15) and PartitionFinder2 (16) to create a custom partitioning scheme for each matrix and ran RAxML 8 again with the new partitions. Finally, we inferred a ML tree with RAxML 8 using a matrix of the best ½ of UCEs based on average entropy score calculated from SWSC-EN. Bayesian inference (BI) analyses were conducted on the two MARE-reduced matrices using Phylobayes 4.1b with the CAT+GTR model (17). Convergence was indicated by a bpcomp maxdiff value <0.3.

Phylogenetic Trees

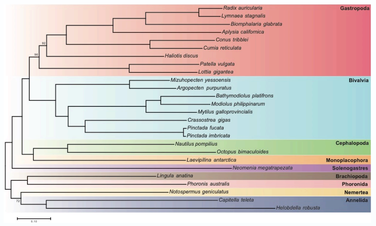

Figure 2. Data Matrix 1 containing all 982 UCEs analyzed in RAxML. Matrix length is 171, 1994 nt with 23.6% missing data.

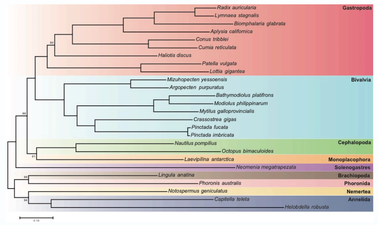

Figure 3. Data Matrix 2 containing the slowest evolving half of UCEs analyzed in RAxML. Matrix length is 88,361 nt with 22.6% missing data.

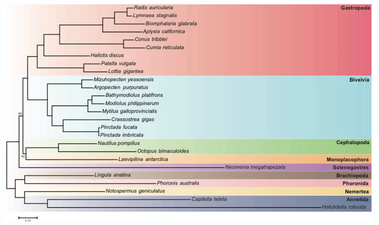

Figure 4. Data Matrix 2 containing the slowest evolving half of UCEs analyzed in Phylobayes. Matrix length is 88,361 nt with 22.6% missing data.

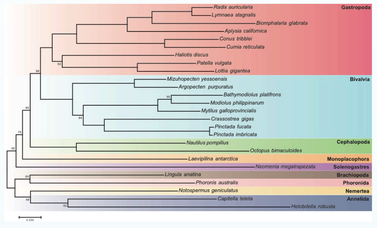

Figure 5. Data Matrix 3 containing the slowest-evolving half of UCEs after reduction in MARE analyzed in RAxML. Matrix length is 17,177 nt with 24.7% missing data.

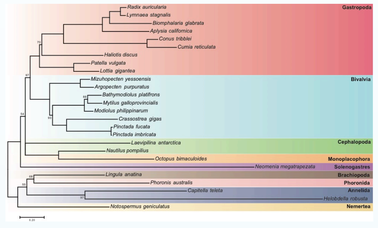

Figure 6. Data Matrix 3 containing the slowest-evolving half of UCEs after reduction in MARE analyzed in Phylobayes. Matrix length is 17,177 nt with 24.7% missing data.

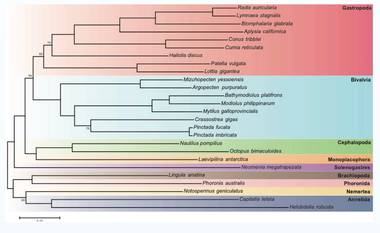

Figure 7. Data Matrix 4 containing the best 1/2 of UCEs according to average entropy score calculated by SWSC-EN analyzed in RAxML. Matrix length is 86,199 nt with 24.5% missing data.

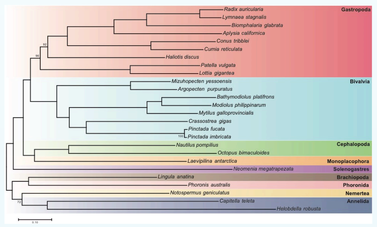

Figure 8. Data Matrix 1 containing all 982 UCEs analyzed in RAxML with custom partitions. Matrix length is 171, 1994 nt with 23.6% missing data.

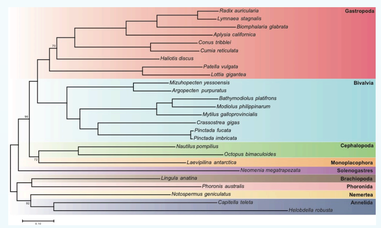

Figure 9. Data Matrix 2 containing the slowest evolving half of UCEs analyzed in RAxML with custom partitions. Matrix length is 88,361 nt with 22.6% missing data.

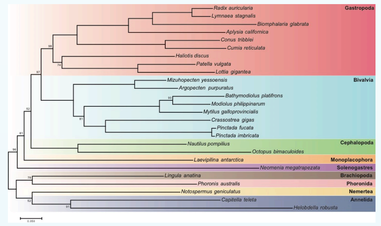

Figure 10. Data Matrix 3 containing the slowest-evolving half of UCEs after reduction in MARE analyzed in RAxML with custom partitions. Matrix length is 17,177 nt with 24.7% missing data.

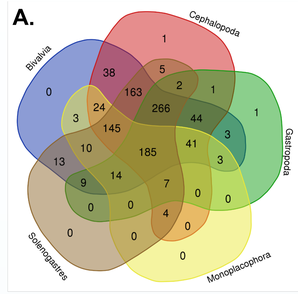

Figure 11. Venn diagrams showing number of UCEs shared among (A) major lineages of Mollusca and (B) phyla. Each UCE was included if represented by at least one member of the group.

Conclusions

UCEs have been shown to be an informative phylogenetic marker in vertebrates, arthropods, and even anthozoans (5-11). This work demonstrates the utility of UCEs in the lophotrochozoan clade Mollusca as well. Results are consistent with recent phylogenomic analyses supporting the Aculifera-Conchifera dichotomy and a clade of Gastropoda + Bivalvia to the exclusion of Cephalopoda. Placement of Monoplacophora remains contentious, but recovery of Monoplacophora sister to all other conchiferans in some analyses warrants further exploration. Future directions for this work are to include underway genomes for representatives of Scaphopoda and Polyplacophora and explore more advanced data filtering strategies.

Results and Discussion

Phylogenetic analyses (Figures 2-10) generally recovered Mollusca, Conchifera, and Bivalvia + Gastropoda with strong support (but see Figure 6). As previously observed (3) Monoplacophora was recovered sister to Cephalopoda in analyses of matrices 1 and 2 (Figures 2-4), but support decreased when the fastest-evolving UCEs were excluded (Figures 3-4). In Figure 5, Monoplacophora was recovered sister to the rest of Conchifera, albeit with weak support. This result, which is consistent with traditional morphological views (reviewed by 1), was also recovered in a RAxML analysis of the second-slowest quartile of the UCEs (data not shown). Matrix 4 (Figure 7) and matrices 1 and 2 analyzed in RAxML with custom partitions (Figures 8, 9) also recovered Monoplacophora sister to Cephalopoda. This relationship was recovered with maximal support in figures 7 and 8, but support decreased in Figure 9 when only the slowest-evolving half of UCEs were examined. The MARE-reduced datasets had varying results when analyzed with our custom partitioning scheme. RAxML analysis of matrix 3 recovered Monoplacophora sister to the rest of Conchifera (Figures 5, 10) which is still contested (1, 3). In the Phylobayes analysis of matrix 3 (Figure 6), Monoplacophora was recovered in a polytomy with Cephalopoda and Gastropoda + Bivalvia. Considering relationships within Bivalvia and Gastropoda, the molluscan clades with by far the most genomic resources to date, our results are consistent with recent phylogenomic investigations (2-3, 18-20). Within Bivalvia, all analyses recovered mussels in a clade with oysters to the exclusion of scallops. Within Gastropoda, most analyzes placed Patellogastropoda sister to all other gastropods but the analysis of the slowest ½ of UCEs after MARE reduction recovered Patellogastropoda + Vetigastropoda. The large number of UCEs shared across most or all major lineages of Mollusca (Figure 11A) and most or all sampled phyla (Figure 11B) indicate that these markers are broadly conserved and will be useful in future studies with improved taxon sampling.

Acknowledgements

We would like to thank Rebecca Varney for extracting DNA from *Neomenia megatrapezata* and Rebecca Varney and Meghan Yap-Chiongco for help assembling the Neomenia genome.

References

1. Kocot, K.M., 2013. Recent Advances and Unanswered Questions in Deep Molluscan Phylogenetics. Am. Malacol. Bull. 31, 195–208.
2. 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.
3. Smith, S.A., Wilson, N.G., Goetz, F.E., Feehery, C., Andrade, S.C., Rouse, G.W., Giribet, G. and Dunn, C.W., 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature, 480, 364-367.
4. Vinther, J., Sperling, E.A., Briggs, D.E. and Peterson, K.J., 2011. A molecular palaeobiological hypothesis for the origin of aplacophoran molluscs and their derivation from chiton-like ancestors. Proceedings of the Royal Society B: Biological Sciences, 279(1732), 1259-1268.
5. 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
6. Faircloth, B.C., 2017. Identifying conserved genomic elements and designing universal bait sets to enrich them. Methods Ecol. Evol. 8, 1103–1112.
7. 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.
8. 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.
9. 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.
10. 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.
11. Quattrini, A.M., Faircloth, B.C., Dueñas, L.F., Bridge, T.C., Brugler, M.R., Calixto-Botía, I.F., DeLeo, D.M., Foret, S., Herrera, S., Lee, S.M. and Miller, D.J., 2018. Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems. Molecular ecology resources, 18(2), pp.281-295.
12. 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.
13. Castresana, J., 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol. Biol. Evol. 17, 540–552.
14. Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
15. Tagliacollo, V., Lanfear, R.; Estimating Improved Partitioning Schemes for Ultraconserved Elements , Molecular Biology and Evolution, Volume 35, Issue 7, 1 July 2018, Pages 1798–1811, https://doi.org/10.1093/molbev/msy069
16. Lanfear, R., Frandsen, PB., Wright, AM., Senfeld, T., Calcott ,B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol. 34: 772–773.
17. Lartillot, N., Rodrigue, N., Stubbs, D., & Richer, J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Systematic Biology, 62(4), 611-615.
18. 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.
19. Lemer, S., González, V.L., Bieler, R. and Giribet, G., 2016. Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach. Proc. R. Soc. B, 283(1833), 20160857.
20. Zapata, F., Wilson, N. G., Howison, M., Andrade, S. C., Jörger, K. M., Schrödl, M., ... & Dunn, C. W. 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. Proc. R. Soc. B, 281(1794), 20141739.