Global Biodiversity of Aplacophora (Mollusca): A DNA Barcoding Approach

Joshua Goble and Kevin M. Kocot
The University of Alabama Department of Biological Sciences

Abstract:

Aplacophorans are a group of spiny, worm-like animals that are closely related to snails and other molluscs. Their bodies are covered by a cuticle with calcareous structures called sclerites, structures that are typically used in morphological identification. Often found in remote, deep-sea benthic environments, these organisms are difficult to sample. Tedious and often ambiguous morphological identification, frequent diet contamination in PCR products, and a lack of representation in existing sequence databases all result in this group being one of the least studied of all Mollusca. This work seeks to alleviate a few of these difficulties by greatly expanding the number of taxa represented genetically through the production of a DNA barcode database. PCR amplifications of COI and 16S sequences were conducted on over 70 specimens representing the known – and unknown – diversity of Aplacophora. These sequences were combined with existing sequences to create a DNA barcode database to be used for future specimen identification and phylogenetic analyses. The DNA barcode database, in conjunction with existing transcriptome data, was used to build a phylogenetic tree representing the evolutionary history of Aplacophora. Our results greatly increase the existing number of COI and 16S sequences for this group and provide a roadmap for future taxonomic and phylogenetic work.

Introduction:

Aplacophorans are a group of spiny, worm-like animals that are closely related to snails and other molluscs. Although aplacophorans lack shells, their bodies are covered by a cuticle bearing sclerites, which give them their characteristic shiny appearance under light microscopy. The calcareous sclerites, wide ranging in appearance, are commonly used for morphological identification and taxonomy (García-Álvarez and Salvini-Plawen 2007). There are two subgroups of Aplacophora: the Solenogastres (=Neomeniomorpha) and Caudofoveata (=Chaetodermomorpha). The two groups of aplacophorans, along with chitons (Polyplacophora), form the sister group of Mollusca, and thus are interesting with respect to understanding early molluscan evolution (reviewed by Kocot 2013). Solenogastres are characterized by a narrow, ciliated foot located in a mid-ventral groove, which is used to move along the sea floor. Caudofoveates lack a foot but have a lip-like structure called the oral shield. It is known that Solenogastres feed primarily on cnidarians (and sometimes annelids and possibly other soft-bodied invertebrates) while Caudofoveates are known to feed on detritus and foraminiferans preferentially (Todt et al. 2008). These organisms are exclusively marine and predominantly inhabit subtidal environments. Most species are <10 mm in length fully-grown although several species are on the order of a few centimeters and the largest specimen ever collected was over 400 mm in length (e.g., Ivanov and Scheltema 1997). Roughly 415 aplacophoran species have been described, but estimates place the total diversity at over 4,000 (Todt 2013). This level of diversity is relatively low compared to most other molluscan classes. Despite this, these two

groups remain among the least known higher taxa within Mollusca, and little is known of their evolutionary relationships (but see Mikkelsen et al. 2018a, 2018b).

This lack of knowledge can be attributed to several factors. First, due to their remote, deep-sea habitats, aplacophorans have proven difficult to collect (Todt 2013). However, recent scientific explorations have drastically increased the number of available specimens. Additionally, aplacophorans are often difficult to identify and describe (Ostermair et al. 2018). Often, data on both external and internal anatomy are required for morphological identification. Histology is often required to characterize internal anatomy (García-Álvarez and Salvini-Plawen 2007), but this precludes other techniques such as scanning electron microscopy (SEM) or DNA extraction. Aplacophorans also frequently cause issues in both PCR amplification and DNA sequencing procedures. Because the diet of solenogaster aplacophorans consists of other animals, PCR contamination is a common problem (Todt 2013). This makes designing aplacophoran-specific primers difficult due to sequence similarity between the aplacophoran and its metazoan prey. Finally, genetic approaches are often hindered by limited availability of data. For example, sequences from more genetically distinct organisms may fail to be amplified by PCR primers not specifically designed for them or, if PCR and sequencing are successful, resulting sequences may not have any close matches in any sequence databases.

DNA barcoding is a common practice to simply yet reliably identify unknown specimens through gene sequencing. Genes used for DNA barcoding share a few common features. First, they are often ubiquitous among a vast array of organisms, allowing for protocol standardization. Second, they are highly conserved genes with vital roles to play in cellular functioning, such that the time-scale for mutations is significantly longer than that of a non-essential gene (Hajibabaei et al. 2007). When comparing multiple genes, it is often productive to use genes with significantly different mutation rates to diversify the level of phylogenetic signal to span both deep and recent radiations.

This work uses COI and 16S sequences for comparative phylogenetic analyses of more than 100 representative taxa spanning the known – and unknown – diversity of Aplacophora. The COI mitochondrial gene encodes the cytochrome *c* oxidase protein subunit I, an essential part of the electron transport chain responsible for the generation of energy in the cell (Vrijenhoek et al. 1994). The 16S small subunit ribosomal RNA gene encodes an RNA molecule forming part of the ribosome responsible for protein production (Palumbi et al. 1991). Both genes are essential for cellular function in virtually all animals and thus are good metazoan barcoding genes. This study greatly expands the existing database of aplacophoran DNA barcode sequences and provides a phylogenetic framework making possible at least general (e.g., family-level) identifications for many specimens that could not be identified based on external anatomy alone. Although DNA barcoding alone should not be used for proper specimen identification (Todt 2013), it is a powerful tool that can be combined with morphological data for specimen identification purposes and to help place organisms in a general evolutionary context.

Material and Methods:

Specimens were collected from waters off Antarctica, the U.S., Iceland, and elsewhere at depths ranging from ~70 m to >4 km using grabs, box corers, epibenthic sleds, and trawls. Large animals were picked from trawl and epibenthic sled samples while smaller animals were picked from fresh or bulk ethanol-fixed sieved material. Specimens were preserved and stored in 95-100% ethanol.

Specimens were first imaged using an Olympus SZX16 stereo microscope. Specimens or fragments of specimens were then imaged using a Phenom Pro SEM with an accelerating voltage of 5 kV for detailed inspection of sclerite structure, an important trait used in morphological characterization. Following SEM imaging, each sample was placed in a 1.5 mL microcentrifuge tube with 200 µL of TL buffer (Omega Bio-Tek lysis buffer) and stored at -80°C until DNA extraction.

DNA extractions were conducted using the Omega Bio-Tek E.Z.N.A. MicroElute Genomic DNA kit following the manufacturer's provided protocol. Following DNA extraction, sample concentration and purity were quantified using a Nanodrop Lite spectrophotometer and extracted DNA was stored at -20°C for later PCR amplification.

PCR reactions were conducted using VWR Life Science Hot Start Taq PCR Master Mix 2X using either published or custom primers that are modified versions of existing primers to improve aplacophoran specificity (Table 1). A reaction size of 25 µL (half reaction) was run in an Eppendorf Mastercycler pro following the thermal cycling protocol presented in Table 2. In some instances, the DNA concentration of the initial extraction was insufficient to yield perceptible PCR product. For these samples, the template volume was either doubled or tripled to provide adequate template amount for the PCR reaction.

Primer Name:	Primer Sequence (5' – 3')	Citation
HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. 1994
LCO_Apl	TTTCTACTAAYCATAARGATATTGG	This study
16S_arL	CGCCTGTTTATCAAAAACAT	Palumbi et al., 1991
16S_brH_Apl	CCGGTCTGAACTCAGATCATGT	This study
16S_fSoleno	RRGAGTYAGRCCTGCCCAGT	Bergmeier et al. 2017
16S_rSoleno	YYTAATCCAACATCGAGGTC	Bergmeier et al. 2017

Table 1: Primer names and sequences

Primer Type:	PCR Reaction Protocol
COI – HCO + LCO_Apl	30s at 95°C (30s at 95°C, 15s at 50°C, 30s at 72°C) x 35, 7 min at 72°C
16S – arL + brH_Apl	30s at 95°C, (30s at 95°C, 30s at 54°C, 60s at 65°C) x 35, 7 min at 65°C
16S – fSoleno + rSoleno	30s at 98°C, (30s at 98°C, 5s at 50°C, 20s at 72°C) x 35, 60s at 72°C

Table 2: Primer set and PCR protocol

For COI, gel electrophoresis was conducted using a 1% agarose gel with fresh TAE buffer. A volume of 20 µL of PCR product was run out on the gel and the band of greatest fluorescence at ~650 bp in length was excised from the gel. For 16S, 3 µL of each PCR product was first run out on a 1% agarose gel made with SB buffer to check for the presence of multiple product bands. If multiple bands were detected, the rest of the sample was run out in a 1% agarose gel with TAE buffer and the appropriate band was excised, as was done for COI PCR products. If not, 16S PCR products with a single band were directly purified using the Omega Bio-Tek E.Z.N.A. Cycle Pure Kit. Those COI and 16S PCR products encased in agarose gel were extracted and purified using the Omega Bio-Tek E.Z.N.A. Gel Extraction Kit. Following purification, PCR product concentration and purity was quantified again using a Nanodrop Lite spectrophotometer.

PCR products were sent to the University of Arizona Genetics Core for Sanger sequencing using an Applied Biosystems 3730 DNA Analyzer. Sequence correction and contig

creation was done using Sequencher 5.4.6 DNA sequence analysis software. Contig alignment was conducted using MEGA7 (Kumer et al. 2016) and sequence identification was done using NCBI nucleotide BLAST (Basic Local Alignment Search Tool). RAxML 8 (Stamatakis 2014) was used to make maximum-likelihood phylogenetic trees based on the general time reversible model with an autocorrelated gamma distribution (Waddell et al. 1997) with four rate categories with over 100 bootstrap replicates. For trees involving multiple genes per sample, concatenated sequence files were made using FASconCAT version 1.11 (Kück and Meusemann 2010).

Results:

Figure 1: Geographical location of samples

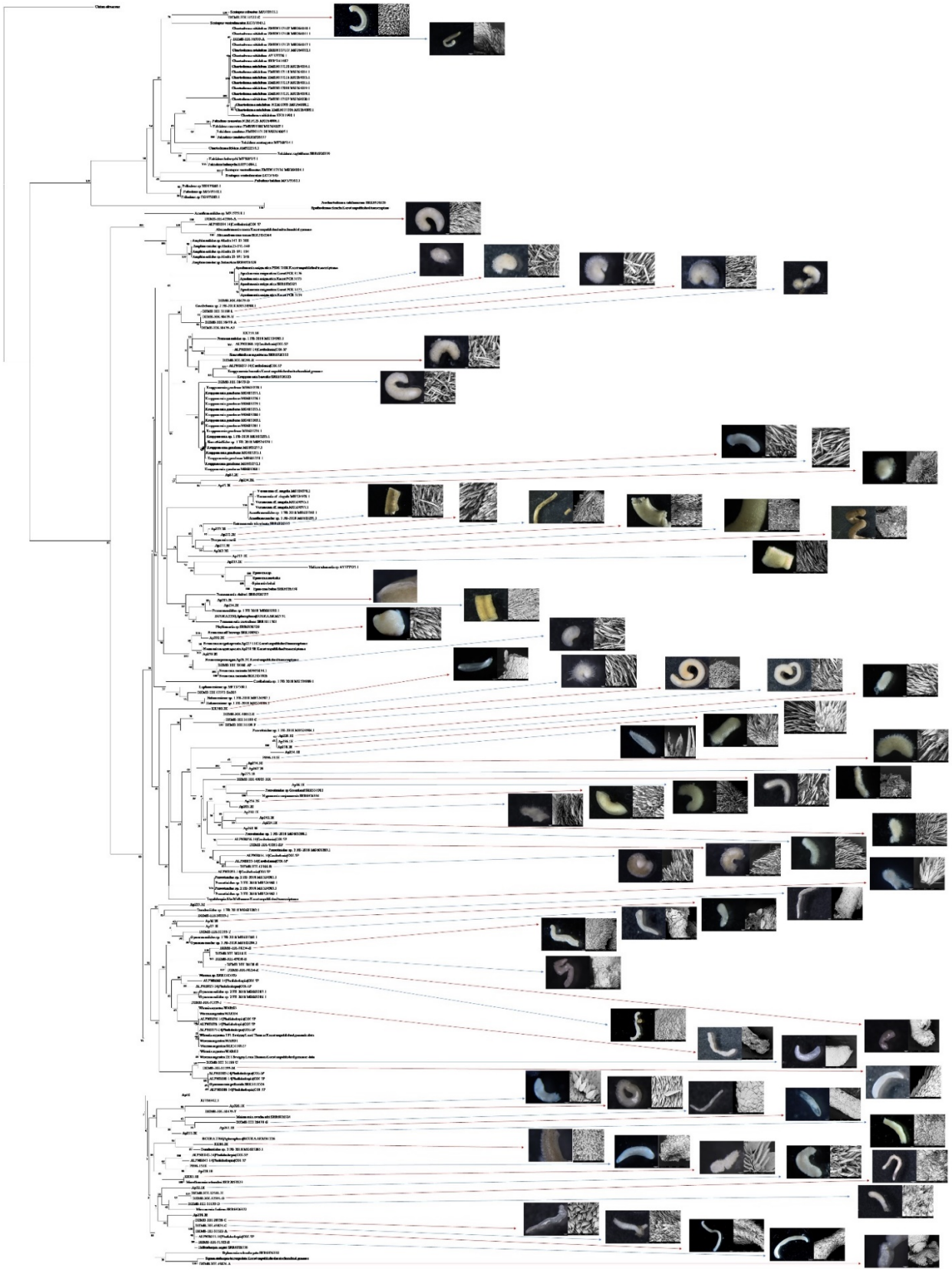

Figure 2: Phylogenetic tree based on COI, 16S, and transcriptome sequence data (in supplementary materials).

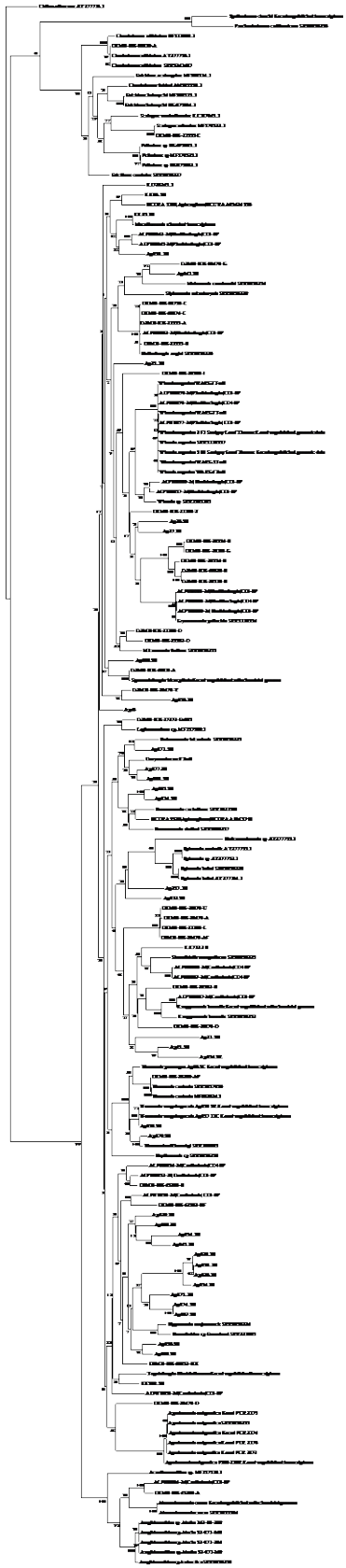

Figure 3: Phylogenetic tree based solely on COI data (in supplementary materials).

Figure 4: Phylogenetic tree based solely on 16S sequence data (in supplementary materials).

Sample Name	PCR No.	Gene	Closest Contamination	% Match	Higher Classification
Ap216.1E	3	COI	*Vulcanonemertes rangitotoensis*	86%	Nemertea
Ap268.3E	29	COI	*Peripatoides* sp.	78%	Onychophora
DZMB-HH-38218-E	31	COI	*Jullienia rolfbrandti*	77%	Gastropoda
DZMB-HH-49838-B	33	COI	*Tricula hortensis*	77%	Gastropoda
Ap223.1E	34	COI	*Cyclocanna welshi*	91%	Cnidaria
Ap213.2E	38	COI	*Chondropoma pictum*	78%	Gastropoda
Ap56.1E	42	COI	*Halopsis ocellata*	91%	Cnidaria
DZMB-HH-38218-G	48	COI	*Cyanoplax keepiana*	76%	Polyplacophora
DZMB-HH-38254-E	49	COI	*Cyanoplax keepiana*	76%	Polyplacophora
DZMB-HH-49874-A	54	COI	*Paruroctonus becki*	81%	Arachnida
DZMB-HH-28479-G	55	COI	*Chondropoma pictum*	77%	Gastropoda
DZMB-HH-31199-T	57	COI	*Novalena intermedia*	78%	Arachnida
DZMB-HH-31199-M	58	COI	*Hamana incita*	81%	Insecta
DZMB-HH-31199-U	76	COI	*Rhipicephalus appendiculatus*	78%	Arachnida
DZMB-HH-49923-HX	77	COI	*Chondropoma gnote*	78%	Gastropoda
Ap255.2E	78	COI	*Clytia folleata*	88%	Cnidaria
DZMB-HH-38479-A	81	COI	*Sinocoelotes pseudoterrestris*	76%	Arachnida
DZMB-HH-38293-B	92	COI	*Stiphidion* sp.	77%	Arachnida
DZMB-HH-31199-I	95	COI	*Dipoena washougalia*	80%	Arachnida
Ap32.1E	99	COI	*Tegenaria tlaxcala*	79%	Arachnida
Ap203.2E	103	COI	*Lutzomyia walkeri*	77%	Insecta
DZMB-HH-39380-BY	107	COI	*Cephalothrix simula*	80%	Nemertea
KK579.2E-3	115	COI	*Eochionelasmus ohtai*	90%	Crustacea
KK579.2E-3	156	16S	*Neveritis aridorum*	81%	Gastropoda
KK579.2E-2	160	16S	*Neveritis aridorum*	80%	Gastropoda
DZMB-HH-31199-AD	264	16S	*Calcinus californiensis*	77%	Crustacea
DZMB-HH-38479-AC	266	16S	*Loxoblemmus appendicularis*	92%	Insecta
DZMB-HH-31199-AE	315	16S	*Calcinus haigae*	78%	Crustacea
DZMB-HH-31466-A	332	COI	*Manania gwilliami*	85%	Cnidaria
Ap31.2E	336	COI	*Balloniscus sellowii*	77%	Crustacea
Ap254.1E	337	COI	*Orthocladiinae* sp.	77%	Insecta
KK579.2E-1	338	COI	*Elaver excepta*	76%	Arachnida
KK579.2E-2	339	COI	*Elaver excepta*	75%	Arachnida
KK713.3E	350	COI	*Diodora cayenensis*	76%	Gastropoda

Table 3: Summary of Sequence Contamination

Discussion:

Our results show a 74% increase in the number of existing COI sequences and an 87% increase in the number of existing 16S sequences based on the data currently available on NCBI and BOLD. The COI samples had significantly greater rates of contamination, which could be due to a lack of primer specificity, and future studies would benefit from further modifying the universal primers or designing new primers to better target aplacophorans. Despite the prevalence of contamination, our results provide a more comprehensive understanding of the evolutionary history of Aplacophora while allowing for more nuanced specimen identification.

Of note are the members of Prochaetodermatidae selected for this study. Of the samples selected, none were able to be PCR amplified. Mikkelsen et al. (2018) had similar issues amplifying template from this group, indicating that these difficulties are more likely based on a lack of primer specificity rather than laboratory procedures. Therefore, more group-specific primers will need to be developed in the future to better characterize this family.

Figure 1 details the exact location where each sample used for this study was collected (link included below). Most of the samples employed in this study were sourced from the Northeastern Atlantic from regions surrounding Iceland during the IceAGE cruises (Brix et al. 2014) and the Antarctic (collected during cruises led by Ken Halanych and *Polarstern* cruise PS96). Samples were also collected from the western Atlantic off the coast of North Carolina, the eastern Pacific off San Juan Island and in the Clarion-Clipperton Zone. Samples collected from regions other than off Iceland and Antarctica were more difficult to PCR amplify, possibly due to a lack of primer specificity. As expected, samples sequenced from outside the two primary regions (Ap31.2E from the northern Pacific, Ap32.1E from the mid-Atlantic, etc.) showed distinct genetic and morphological differences from the samples located in the more polar regions. They were dissimilar from all other samples and had no similar sequences in the NCBI database with which to compare.

The phylogenetic tree built solely from COI sequences (Figure 3) has stronger support for more shallow nodes relative to the tree built from only 16S sequences (Figure 4). However, when combined with existing transcriptome data, the resulting tree has more overall support (Figure 2), especially along the ‘backbone’ of the tree. These results highlight the value of multi-gene analyses, as the addition of more gene sequences from the transcriptomes provides improved support for relationships among those taxa.

The stereomicroscope and scanning electron microscope images provide additional information for use in specimen identification. DNA barcoding alone is not sufficient for specimen characterization, but when used with morphological data, in the form of microscope images, it becomes easier to place organisms in an evolutionary framework. Sclerite structure is an important morphological character that is compatible with molecular approaches. Of note, however, are the differing sclerite structures throughout an organism, and the problems that can arise when comparing sclerites from different areas of the body.

Despite the possible confounding variable associated with different types of sclerites in different regions of the body, morphological data like those presented in Figure 2 play an important role in specimen identification. This is evidenced by the case study involving samples KK85.1E and *Macellomenia schanderi*. Based on initial stereomicroscope and SEM images, it was hypothesized that KK85.1E was a juvenile form of *Macellomenia schanderi* but it could have also been a representative of the cogener *Macellomenia morseae* that co-occurs with this species. To test this hypothesis, COI and 16S sequences from KK85.1E were compared to

existing sequences of *Macellomenia schanderi*, with the resulting trees placing them as very closely related with near certainty (bootstrap value of 100). This is just one of many examples of unknown specimens being closely identified through the DNA barcoding approach, outlining the power of this method for specimen identification and the value in increasing the representative diversity of samples in our DNA barcode database.

These results also have the potential to identify incorrectly identified sequences in public databases. For example, Zapata et al. (2014) uploaded transcriptome data from "*Chaetoderma* sp." to the NCBI SRA database under accession number SRR1505105. We mined COI from this transcriptome and showed that this organism was misidentified and is probably an undescribed species of the solenogaster *Wirenia*. Likewise, a sequence uploaded to the NCBI Nucleotide database under accession number AY377725.1 was labelled as *Helicoradomenia* sp. (Simrothiellidae), but this sequence clusters with sequences from *Epimenia* (Epimeniidae) species rather than the monophyletic Simrothiellidae clade.

The different gene sequences provide additional, and sometimes confounding, information about the evolutionary history of Aplacophora. The existing COI sequence from a member of Acanthomeniidae places it in a clade sister to all other solengasters and closely related to *Epimenia*. However, the existing 16S sequences from this group place it far from this location, all on its own, closer to Amphimeniidae rather than *Epimenia*. Sample DZMB-42366-A remains in roughly the same location in both trees built from COI and 16S sequences respectively and can be used as a landmark to better recognize this distinction. This could be due to a mis-identified specimen, as was most likely the case with the previously-mentioned *Helicoradomenia* sp., or this could be due to differences in phylogenetic signal content of 16S and COI. Unfortunately, this sample sequence was not generated as a part of this work, and thus morphological data are not readily available to help confirm or deny any suspicions as to the identity of this specimen.

Although most of the PCR reactions were successful, many were contaminated either through bad reads or inadvertent amplification of the wrong template. The bad reads were most likely due to multiple PCR products of similar size that were unable to be separated through gel electrophoresis prior to sequencing, leading to a noise-filled result. Table 3 details the sequences that had a BLAST hit to an organism other than an aplacophoran. Of note are the four very close results to members of Cnidaria. Aplacophorans, specifically the Solenogastres, are known to feed on benthic cnidarian colonies, and thus these results suggest inadvertent amplification of gut contents providing more insight into the diets of these organisms. Additionally, results closest in similarity to other common marine invertebrates (Gastropoda, Crustacea, Nemertea, Polyplacophora) are most likely unknown aspects of the aplacophoran diet with no similar sequence currently present in the NCBI database. It is also possible that the non-aplacophoran mollusc BLAST hits represent actual aplacophoran sequences that are highly divergent from other species sequenced to date, as was shown to be the case for COI mined from the transcriptome of *Dondersia todtae* (Kocot unpublished data). Finally, the apparent insect-related sequences commonly found among our PCR products were most likely not insect-related, as evidenced by the low sequence similarity with their top hits in the NCBI Nucleotide database. A lack of representation of many lineages of microeukaryotes in the NCBI database and the excess of insect sequences both skew BLAST results in favor of insect sequences being the best hit due to random change. However, these sequences are most likely of organisms living on or in aplacophorans or being consumed by aplacophorans as a part of their diets. These results provide

a compelling rationale for future study into the relatively unknown diets of these benthic organisms.

This work represents a substantial advancement toward understanding the true diversity of Aplacophora, though this group remains poorly known relative to most other classes within Mollusca. As shown in Figure 2, there are many newly formed clades with no previously characterized specimens. DNA barcoding is hampered by the nature of the database to which they are being compared. With a robust database of characterized individuals, DNA barcoding provides a simple and efficient way of roughly identifying organisms. At this stage, however, the database is sorely lacking in representative diversity, especially in the number of characterized genera and species. Fortunately, there are additional specimens of nearly every species amplified through this work, providing an avenue for future detailed morphological identification to then pair with the sequence data for a more complete understanding of the phylogeny of this group.

References:

1. Bergmeier, F. S., Brandt, A., Schwabe, E., & Jörger, K. M. (2017). Abyssal Solenogastres (Mollusca, Aplacophora) from the Northwest Pacific: Scratching the Surface of Deep-Sea Diversity Using Integrative Taxonomy. *Frontiers in Marine Science*, *4*, 410.
2. Brix, S., Meissner, K., Stransky, B., Halanych, K.M., Jennings, R.M., Kocot, K.M. and Svavarsson, J., 2014. Preface: the IceAGE project-a follow up of BIOICE. *Polish Polar Research*, *35*(2), pp.141-150
3. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. (1994). DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3, 294–299.
4. García-Álvarez, O., & Salvini-Plawen, L. V. (2007). Species and diagnosis of the families and genera of Solenogastres (Mollusca). *Iberus*, 25(2), 73-143.
5. Hajibabaei, M., Singer, G. A., Hebert, P. D., & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *TRENDS in Genetics*, 23(4), 167-172.
6. Ivanov, D. L., & Scheltema, A. (2007). Chaetoderma felderi: A new giant caudofoveate species from the Gulf of Mexico (Mollusca: Aplacophora). *Ruthernica*, 17(1-2), 7-12.
7. Kocot, K.M. (2013). Recent advances and unanswered question in deep molluscan phylogenetics. *American Malacological Bulletin*. 31(1). 195-208.
8. Kück, P., & Meusemann, K. (2010). FASconCAT: Convenient handling of data matrices. *Molecular Phylogenetics and Evolution*, 56(3), 1115-1118.
9. Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870-1874.
10. Mikkelsen, N. T., Todt, C., Kocot, K. M., Halanych, K. M., & Willassen, E. (2018). Molecular phylogeny of Caudofoveata (Mollusca) challenges traditional views. *Molecular phylogenetics and evolution*.
11. Ostermair, L., Brandt, A., Haszprunar, G., Jörger, K. M., & Bergmeier, F. S. (2017). First insights into the solenogaster diversity of the Sea of Okhotsk with the description of a new species of Kruppomenia (Simrothiellidae, Cavibelonia). *Deep Sea Research Part II: Topical Studies in Oceanography*.
12. Palumbi, S. (1991). Simple fool's guide to PCR.
13. Sequencher® version 5.4.6 DNA sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com
14. Schaeffer, N.V., Missoup, A.D., Kennis, J., Colyn, M., et al. (2012) Assessment of Three Mitochondrial Genes (16S, Cytb, CO1) for Identifying Species in the Praomyini Tribe (Rodentia: Muridae). *PLOS ONE* 7(5): e36586.
15. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313.
16. Todt, C. (2013). Aplacophoran mollusks—still obscure and difficult?. *American Malacological Bulletin*, 31(1), 181-187.
17. Todt, C., Okusu, A., Schander, C., & Schwabe, E. (2008). Solenogastres, Caudofoveata, and Polyplacophora. *Phylogeny and Evolution of the Mollusca*, 4, 71-96.

18. Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), 294-9.
19. Waddell, P. J., & Steel, M. A. (1997). General time-reversible distances with unequal rates across sites: mixing Γ and inverse Gaussian distributions with invariant sites. *Molecular Phylogenetics and Evolution*, 8(3), 398-414.
20. Zapata, F., Wilson, N.G., Howison, M., Andrade, S.C., Jörger, K.M., Schrödl, M., Goetz, F.E., Giribet, G. and Dunn, C.W., 2014. Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc. R. Soc. B*, 281(1794), p.20141739.

Link to specimen sampling locations:
https://drive.google.com/open?id=1U2HGHRFrFQ7tXFeokLZsYuWC6-wwuQ48&usp=sharing