Mitochondrial cytochrome b sequence data are not an improvement for species identification in scleractinian corals

There are well-known difficulties in using the cytochrome oxidase I (COI) mitochondrial gene region for population genetics and DNA barcoding in corals. A recent study of species divergence in the endemic Caribbean genus *Agaricia* reinforced such knowledge. However, the growing availability of whole mitochondrial genomes may help indicate more promising gene regions for species delineation. I assembled the whole mitochondrial genome for *Agaricia fragilis* from Illumina single-end 250bp reads and compared this sequence to that of the congener *A. humilis*. Although these data suggest that the cytochrome *b* (CYB) gene region is more promising, comparison of available CYB sequence data from scleractinian and other reef-building corals indicates that multilocus approaches are still probably necessary for phylogenetic and population genetic analysis of recently-diverged coral taxa.

J. P. Wares[1]

[1]**Department of Genetics, University of Georgia, Athens, GA 30602; jpwares@uga.edu**

ABSTRACT

Keywords: coral, barcoding, mitochondrial DNA

INTRODUCTION

Coral reefs are widely recognized as being important representatives and biogenic harbors of biodiversity (Plaisance et al., 2011). At the same time as coral reefs are in crisis due to disease and habitat change, there is still new diversity being explored (Breedy et al., 2013; Breedy and Guzman, 2014). With the incredible phenotypic diversity that may be found in an individual taxon (Bruno and Edmunds, 1997; Veron, 2011), as well as the potential for hybridization among taxa (Vollmer and Palumbi, 2002), biologists often turn to molecular techniques for separating taxa.

In many animal species, this approach has been relatively straightforward and has often relied on a single gene that is both highly variable at silent nucleotide positions as well as highly conserved for amino acid sequence (Folmer et al., 1994; Hebert et al., 2003). This combination allows the sequencing of this gene with 'universal' primers, yet the discovery of tremendous amounts of nucleotide variation that may be used to distinguish taxa. However, this gene region has proven nearly useless in corals (Shearer and Coffroth, 2008). Researchers have also tried using other protein-coding loci such as *atp*6 (Bongaerts et al., 2013) to explore phylogenetic diversity in Agariciid corals, but still tended to recover nonmonophyletic taxa. Similarly, Meyers (2013) showed that using intron regions within the mitochondrial ND5 locus (Concepcion et al., 2006) could not resolve many species in the genus *Agaricia*.

The focus for such work has often been mitochondrial regions because the DNA is abundant in animal tissues, often variable within and among populations, and the lower effective size of the mitochondrial genome – a haploid genome that is typically maternally inherited – tends to result in diagnostic nucleotide characters for a population in less time than for a nuclear locus (Avise, 2000). For both historical and empirical reasons, some groups of systematists and population geneticists have widely used other mitochondrial regions with success. Population genetics in fishes, for example, frequently explore cytochrome b or ND4, and some have used non-coding regions (e.g. ribosomal or the 'D-loop' origin of replication)(Muss et al., 2001; Taylor and Hellberg, 2003; Hyde and Vetter, 2007).

The brief goal of this study is to attempt to identify another useful mitochondrial region for population genetics and systematics studies in scleractinian corals. Here I focus on the Agariciidae; no complete phylogeny yet sufficiently resolves the endemic Caribbean genus *Agaricia* (Meyers, 2013; Bongaerts et al., 2013), and overall the family is an important one for reef development but needs further exploration of its biogeographic and phylogenetic history (Luck et al., 2013). Thus, this study first compares whole mitochondrial sequences between two divergent taxa of *Agaricia* (*A. fragilis* and *A. humilis*), and identifies the most divergent protein-coding region (using coding regions for increased likelihood of conserved primer development). I then analyze divergence in this region (cytochrome *b*, CYB) across available scleractinian data to show that this region alone is unlikely to improve our ability to separate taxa using DNA sequence-based methods.

METHODS

To identify potential regions on the easily-sequenced mitochondrial locus, a single individual of A. fragilis (AS1943, collected in the Upper Florida Keys and detailed in Meyers et al. (2013)) was shotgun sequenced with a single Illumina MiSeq library preparation as in Wares (2013). Resultant single-end 250nt reads were trimmed and mapped to the *A. humilis* mitochondrial genome (GenBank DQ643831) using Geneious 7.1.4 (Biomatters). The alignment process included up to 5 iterations, with maximum gapped sites per read of 10%, maximum mismatches per read of 20% and a minimum overlap of 20 nucleotides. Annotation of

this assembled genome was initiated using MITOS (Bernt et al., 2013) and corrected via re-alignment with the *A. humilis* sequence.

Aligned coding sequences were evaluated for K2P divergence between the two genomes using PAUP*4.0b10 (Swofford, 2000) as in Shearer and Coffroth (2008); a sliding-window measure of divergence was calculated for 500-bp regions in 25-bp increments along the whole mitochondrial genome.

Subsequently, sequence data for the CYB locus from coral studies that included 'Scleractinia' (these studies often include other taxa from the Anthozoan subclass Hexacorallia) were downloaded from GenBank using Geneious 7.1.4 and aligned in CodonCode Aligner v4.2.2. Again, K2P distances among all sequences were obtained using PAUP*, and all pairwise distances were coded as conspecific, congeneric (excluding conspecific), or "other". The distances observed for these 3 classes of comparison were density plotted using ggplot2 in the R computational environment.

RESULTS

Illumina sequencing of the *A. fragilis* genomic DNA library resulted in a total of 31,957,468 reads. Mapping these reads to the *A. humilis* mitochondrial genome generated a single contig of 18,667 bp. The completed *A. fragilis* mitochondrial genome (Genbank KM051016) had no observed gene rearrangements and is consistent with the standard type SII for scleractinian corals (Lin et al., 2014).

Sliding window comparison of the two mitochondrial genomes is shown in Figure **??**. The coding region with highest divergence between the two sequences is cytochrome b (CYB) with a mean divergence of 0.024 substitutions per site. All other coding regions exhibit lower divergence per nucleotide, with COI only about 1.6 percent divergent.

Evaluation of CYB sequences across the Scleractinia and other corals for their utility in separating congeneric and more distant species included 249 pairs of intraspecific contrasts, 2098 intrageneric contrasts, and a total of 64981 contrasts that included 203 species from 94 genera of corals. The sequence alignment of these data is provided as Supplemental File S1; a neighbor-joining tree generated from these data is provided as Supplement S2. These contrasts, shown in Figure **??**, indicate that a divergence comparable to intraspecific diversity can be observed between members of the same genus or even more distantly related taxon pairs.

DISCUSSION

The results of this study do less than hoped to advance molecular methods for species identification in corals. Mitochondrial DNA is often an optimal solution for metazoan species barcoding and a first attempt at species delineation (Hebert et al., 2003). Yet, in corals - or anthozoans in general - the processes of mutation and DNA error correction in mitochondrial DNA (Hellberg, 2006), along with the propensity for hybridization among some taxa, has led to the need for more laborious locus development for such goals. Many studies are relying on microsatellite development (Concepcion et al., 2010, 2014), which enables additional variation and the benefits of a multi-locus study; however, the direct identification and analysis of shared, derived characters that distinguish populations is sometimes more complicated with such data.

Ultimately the goal of species delineation is identification of character states that are diagnostic. Finding gene regions that provide sufficient information, above and beyond the variation found within a population, is the challenge. Some nuclear gene regions have shown promise. For example, Concepcion et al. (2008) identified the SRP54 exon-primed intron-crossing locus as being a single-copy locus that is typically more variable than non-coding regions such as the ribosomal internal transcribed spacer (ITS) regions. Other authors are combining data from several loci to attain the same goal (McFadden et al., 2014). Certainly it is now common to approach phylogenetic and population genetic questions with multi-locus data where possible, and the same rigor appears to be necessary for species delimitation in coral taxa.

A somewhat circular problem of using available data for consideration of barcode locus efficiency is that Genbank itself is rife with poorly identified data (Meier et al., 2006; Kwong et al., 2012). When a nucleotide sequence is labeled/described only to genus, as in 'Discosoma sp.', it may or may not be conspecific with *Discosoma nummiforme*. In these analyses, such taxa would be compared at the genus, not the species level; however, in the data included in this study this would affect at most 5 taxon pairs. Additionally, there is the occasional problem of species that phenotypically appear to be similar to the focal taxon, but genetically divergent (for example, in a phylogeography study), and are labeled with '*cf.*'

as indication of this uncertainty. These may or may not be the same actual species; in this study, they are treated the same as with underdescribed taxa noted above. Other taxonomic concerns among the data could similarly affect the inclusion of a contrast as within- or among-genera, but are beyond the scope of this study to resolve.

The premise of this study, that more divergent regions could be found by comparing mitochondrial genomes, is directly relevant only to the genus *Agaricia* from which these sequences derive. Using only a single genome from each species presents an incomplete picture of overall net nucleotide divergence (Nei and Li, 1979). However, given the typical problem of developing such markers in corals it may make sense as a general strategy to first explore available genomic data - whether mitochondrial or whole-genome - rather than blindly tackle the problem with available primer regions or use the same gene region that has proven useful in other Metazoans. Here, for example, we see that while the divergence between members of *Agaricia* is low at most protein-coding regions of the mitochondrion, the 'control region' exhibits high divergence; Luck et al. (2013) used another novel non-coding region to separate the Agariciid taxa *Leptoseris* and *Pavona*. It remains to be seen whether using next-generation approaches, as in this study, to generate whole mitochondrial genome sequences, may be more informative (but see Fukami and Knowlton (2005)) and nearly as cost-effective as attempting to capture several distinct gene regions via PCR.

ACKNOWLEDGMENTS

Sequence data were generated at the Georgia Genomics Facility (dna.uga.edu). The specimen of *Agaricia fragilis* was collected by M. K. Meyers. Thanks to WriteLatex.com for aiding my escape from the hegemony of Microsoft.

REFERENCES

Avise, J. C. (2000). *Phylogeography, the History and Formation of Species.* Harvard Univ. Press, Cambridge, MA.

Bernt, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., Putz, J., Middendorf, M., and Stadler, P. F. (2013). Mitos: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2):313–319.

Bongaerts, P., Frade, P. R., Ogier, J. J., Hay, K. B., van Bleijswijk, J., Englebert, N., Vermeij, M. J. A., Bak, R. P. M., Visser, P. M., and Hoegh-Guldberg, O. (2013). Sharing the slope: depth partitioning of agariciid corals and associated *Symbiodinium* across shallow and mesophotic habitats (2-60 m) on a caribbean reef. *BMC Evolutionary Biology*, 13:205.

Breedy, O. and Guzman, H. M. (2014). A new species of alcyonacean octocoral from the peruvian zoogeographic region. *Journal of the Marine Biological Association of the United Kingdom*, 94(3):493–498.

Breedy, O., Williams, G. C., and Guzman, H. M. (2013). Two new species of gorgonian octocorals from the tropical eastern pacific biogeographic region (cnidaria, anthozoa, gorgoniidae). *Zookeys*, (350):75–90.

Bruno, J. F. and Edmunds, P. J. (1997). Clonal variation for phenotypic plasticity in the coral *Madracis mirabilis*. *Ecology*, 78(7):2177–2190.

Concepcion, G. T., Baums, I. B., and Toonen, R. J. (2014). Regional population structure of *Montipora capitata* across the hawaiian archipelago. *Bulletin of Marine Science*, 90(1):257–275.

Concepcion, G. T., Crepeau, M. W., Wagner, D., Kahng, S. E., and Toonen, R. J. (2008). An alternative to its, a hypervariable, single-copy nuclear intron in corals, and its use in detecting cryptic species within the octocoral genus *Carijoa*. *Coral Reefs*, 27(2):323–336. 291IV Times Cited:23 Cited References Count:51.

Concepcion, G. T., Medina, M., and Toonen, R. J. (2006). Noncoding mitochondrial loci for corals. *Molecular Ecology Notes*, 6:1208–1211.

Concepcion, G. T., Polato, N. R., Baums, I. B., and Toonen, R. J. (2010). Development of microsatellite markers from four hawaiian corals: *Acropora cytherea, Fungia scutaria, Montipora capitata* and *Porites lobata*. *Conservation Genetics Resources*, 2(1):11–15.

Folmer, O., Black, M., Hoeh, W., Lutz, R., and 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.

Fukami, H. and Knowlton, N. (2005). Analysis of complete mitochondrial dna sequences of three members of the *Montastraea annularis* coral species complex (cnidaria, anthozoa, scleractinia). *Coral Reefs*, 24(3):410–417.

Hebert, P., Cywinska, A., Ball, S., and deWaard, J. (2003). Biological identification through dna barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270:313–321.

Hellberg, M. E. (2006). No variation and low synonymous substitution rates in coral mtdna despite high nuclear variation. *BMC Evolutionary Biology*, 6:24.

Hyde, J. R. and Vetter, R. D. (2007). The origin, evolution, and diversification of rockfishes of the genus *Sebastes* (cuvier). *Mol Phylogenet Evol*, 44:790–811.

Kwong, S., Srivathsan, A., and Meier, R. (2012). An update on dna barcoding: low species coverage and numerous unidentified sequences. *Cladistics*, 28(6):639–644.

Lin, M. F., Kitahara, M. V., Luo, H., Tracey, D., Geller, J., Fukami, H., Miller, D. J., and Chen, C. A. (2014). Mitochondrial genome rearrangements in the scleractinia/corallimorpharia complex: implications for coral phylogeny. *Genome Biol Evol*, 6(5):1086–95.

Luck, D. G., Forsman, Z. H., Toonen, R. J., Leicht, S. J., and Kahng, S. E. (2013). Polyphyly and hidden species among hawai'i's dominant mesophotic coral genera, *Leptoseris* and *Pavona* (scleractinia: Agariciidae). *PeerJ*, 1:e132.

McFadden, C. S., Reynolds, A., and Janes, M. P. (2014). Dna barcoding of xeniid soft corals (octocorallia: Alcyonacea: Xeniidae) from indonesia: species richness and phylogenetic relationships. *Systematics and Biodiversity*, 12:247–257.

Meier, R., Shiyang, K., Vaidya, G., and Ng, P. K. L. (2006). Dna barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology*, 55(5):715–728.

Meyers, M., Porter, J. W., and Wares, J. P. (2013). Genetic diversity of fluorescent proteins in caribbean agariciid corals. *Journal of Heredity*, 104(4):572–577.

Meyers, M. K. (2013). *From molecules to communities: an examination of multiple scales of diversity in reef building corals*. PhD thesis.

Muss, A., Robertson, D. R., Stepien, C. A., Wirtz, P., and Bowen, B. W. (2001). Phylogeography of *Ophioblennius*: the role of ocean currents and geography in reef fish evolution. *Evolution Int J Org Evolution*, 55:561–72.

Nei, M. and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA*, 76:5269–5273.

Plaisance, L., Caley, M. J., Brainard, R. E., and Knowlton, N. (2011). The diversity of coral reefs: What are we missing? *Plos One*, 6(10).

Shearer, T. L. and Coffroth, M. A. (2008). Barcoding corals: limited by interspecific divergence, not intraspecific variation. *Mol. Ecol. Res.*, 8:247–255.

Swofford, D. L. (2000). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland.

Taylor, M. S. and Hellberg, M. E. (2003). Genetic evidence for local retention of pelagic larvae in a caribbean reef fish. *Science*, 299:107–109.

Veron, J. E. N. (2011). Coral taxonomy and evolution. *Coral Reefs: An Ecosystem in Transition*, pages 37–45. Bsw81 Times Cited:1 Cited References Count:17.

Vollmer, S. V. and Palumbi, S. R. (2002). Hybridization and the evolution of reef coral diversity. *Science*, 296:2023–2025.

Wares, J. P. (2013). Mitochondrial evolution across lineages of the vampire barnacle *Notochthamalus scabrosus*. *Mitochondrial DNA*, pages 1–3.

Figure 1

Divergence of mitochondrial genomes in *Agaricia*

Sliding window divergence between mitochondrial genomes of *A. fragilis* and *A. humilis*. Window size was 500bp, measured every 25bp. Coding regions are shaded in green; ribosomal regions in red. Other non-coding regions (tRNA and the intron region for ND5) not indicated. Cytochrome b is shaded blue and harbors highest mean divergence of 0.024 (dotted line).
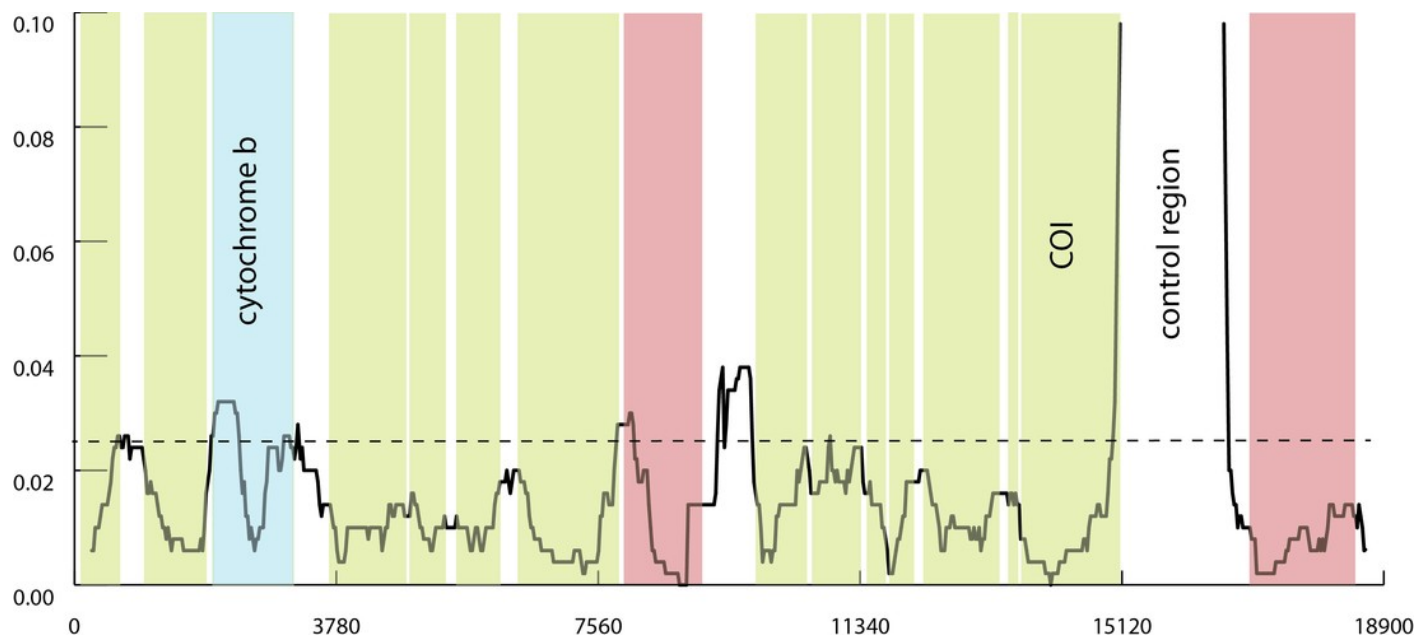
Figure 2

Cytochrome b divergence among scleractinian and related corals

Cytochrome b divergence, using K2P genetic distances, among taxon pairs. Plots are separated by intraspecific contrasts, intra-generic contrasts, and all other observed distances. Plot is truncated at 0.05 for clarity; all intraspecific and almost all intra-generic contrasts are shown.