Mitochondrial Genome Evolution in Pupillid Land Snails

Jason D. Marquardt

Follow this and additional works at: http://digitalrepository.unm.edu/biol_etds

Recommended Citation

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Biology ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.
This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Ulfar Bergthorsson, Chairperson

Jeffery C. Nekola

Coenraad M. Adema
Mitochondrial Genome Evolution in Pupillid Land Snails

by

JASON MARQUARDT

B.A., BIOLOGY, UNIVERSITY OF NEW MEXICO, 2010
M.S., BIOLOGY, UNIVERSITY OF NEW MEXICO, 2013

THESIS
Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
Biology

The University of New Mexico, Albuquerque, New Mexico
MAY 2013
Acknowledgements

First, I would like to acknowledge my advisor, Ulfar Bergthorsson. His support in this project has been irreplaceable. I am appreciative of his time answering questions and explaining concepts.

Secondly, I want to thank my committee members Coenraad Adema and Jeff Nekola for their help in answering questions and providing samples and supplies whether they were required to or not.

Finally, I would like to thank James Farslow for his constant availability to answer questions and his erudite understanding of Matlab.
Mitochondrial Genome Evolution in Pupillid Land Snails

by

Jason Marquardt

B.A., Biology, University of New Mexico, 2010
M.S., Biology, University of New Mexico, 2013

Abstract

Pupillid land snails (Family Pupillidae) are small (<6mm) snails with a broad geographic distribution and often found in very high density in their habitat. The sequencing of the mitochondrial genome (mt genome) of three pupillid taxa has been undertaken to understand their genome evolution more fully. The Gastrocopta cristata mitochondrial genome is 14,060 bp in length and contains 13 protein coding genes, 2 rRNA genes and 22 tRNA genes. The Pupilla muscorum, and Vertigo pusilla genomes contain all of the same genes but are of differing total sizes, 14,149 bp and 14,078 bp respectively. The AT content of the three genomes is similar at ~71% A+T which is comparable to their closest sequenced relatives. There are no major gene rearrangements among the mt genomes of the three pupillids, but the positions of many tRNA genes differ from those of Albinaria caerulea. There is a genomic sequence region of high concentration of thymine on the leading strand of the mt genome that is shared among all three genomes as well as their relatives.
# TABLE OF CONTENTS

Introduction .................................................................................................................. 1

Methods .......................................................................................................................... 2

Results .............................................................................................................................. 8

Discussion ...................................................................................................................... 14

Conclusions .................................................................................................................. 17

Figures ........................................................................................................................... 19

Tables .............................................................................................................................. 43

Appendix A ................................................................................................................... 47

Appendix B .................................................................................................................... 51

Appendix C .................................................................................................................... 72

Appendix D .................................................................................................................... 75

References ..................................................................................................................... 78
Introduction

The mitochondrial genomes (mt genomes) of land snails are double-stranded circular DNA molecules and encompass 13 protein coding genes, 22 transfer RNA genes and two ribosomal RNA genes. The sizes of these mitochondrial genomes are presumed to be ~14kb, as *Albinaria caerulea* (Hatzoglou *et al.* 1995), *Cepaea nemoralis* (Yamazaki *et al.* 1997), *Succinea putris* (White *et al.* 2011) and *Cylindrus obtusus* (Groenenberg *et al.* 2012) are all approximately 14kb. The taxonomically informative changes in gene order between the mt genomes of different lineages make gastropods superb candidates for inferring evolutionary relationships (Grande *et al.* 2008). Mt genomes have many features which make them ideal for extrapolating taxonomic relationships: they are abundant in animal cells and are easy to work with, they have a higher mutation rate than the nuclear genome, and they lack recombination due to their uni-parental inheritance (Avise *et al.* 1987, Avise 1991, Moriyama *et al.* 1997). All these characteristics make the mitochondria adequate to capture recent taxonomic history despite the issues with introgression and linkage between the nuclear and mitochondrial genomes.

Sequencing of mt genomes is informative for taxonomic identification as well as analysis of evolutionary changes. Pupillid land snails fall under the Pupilloidea super family. They are a wide-ranging geographic group for which completely assembled mitochondrial genome sequences are not available. Among the Stylommatophora, an infra-order of pulmonates (Gastropoda, Mollusca) that the pupillid land snails belong to, only four mitochondrial genomes have been sequenced: *Albinaria caerulea* (Hatzoglou *et al.* 1995), *Cepaea nemoralis* (Yamazaki *et al.* 1997), *Succinea putris* (White *et al.* 2011)
and *Cylindrus obtusus* (Groenenberg *et al.* 2012); all representatives of other superfamilies (White *et al.* 2011, Wade *et al.* 2006). The complete characterization and annotation of additional mitochondrial genomes will provide data for interpretation of evolution among snails in Pupilloidea as well as outside this super family, adding to previous research by Von Proschwitz *et al.* 2009 and Hatzoglou *et al.* 1995 among others.

Three genomes from the Pupilloidea super family were sequenced. Specifically, the entire mitochondrial genomes of *Gastrocopta cristata*, *Pupilla muscorum*, and *Vertigo pusilla* were sequenced with Sanger sequencing. These particular species were chosen because each are type species within their genus, or as close to type species as easily available. These three genomes were compared with the 21 complete and annotated genomes from Panpulmonata (Gastropoda, Mollusca) from the National Center for Biotechnology Information (NCBI) database. I elucidate in this thesis the genome features, taxonomy and codon bias, G/C skew, $D_n/D_s$, gene order changes, and nucleotide composition. Some comparative analysis was also undertaken to look at possible evolutionary factors: $D_n/D_s$ comparisons, relative rate of evolution comparisons, nucleotide density comparisons, codon bias comparisons and gene order comparisons.

**Materials and Methods**

**Sample collection:**

The *Gastrocopta cristata* (35.0727 N., 106.6160 W.) sample was collected by Jeff Nekola in Albuquerque, NM, while the *Pupilla muscorum* and *Vertigo pusilla* were
acquired by Michal Horsak from (49.2509 N; 16.5738 E.) and (48.8586 N., 15.8960 E.) in Moravia, Czech Republic. All samples were collected using standard protocols (Nekola et al. 2010) and were single whole snails in estivation that were cleaned with molecular water prior to DNA extraction.

**DNA extraction:**

Total DNA was extracted from whole body tissue of individual snails, including shells, with the E.Z.N.A. mollusk DNA extraction kit (D3373, Omega Bio-Tek, Inc., Norcross, GA, USA) with the following modifications to manufacturer’s protocol. The initial incubation, in lysis buffer, was increased from 4 to 7 hours and vortexing was replaced with inversion mixing. When bonding the DNA to the column, after protein precipitation, instead of filtering once with 750 µl of lysate, filter twice with 350 µl of lysate each time ending with ~100 µl of product.

**Sequencing and genome assembly:**

Universal primers for cox1 based on (Gittenberger et al. 2004), 12S (Appendix C, Sec 1), cox2 (Hugall et al. 2002), cob (von Proschwitz et al. 2009), 16SF (Tongkerd et al. 2004) and 16SR (Jørgensen et al. 2004) were used to obtain initial amplicons which were then sequenced and used to design exact primers for long distance PCR and DNA sequencing. Universal primers for 12S were designed in collaboration with Coenraad Adema by inspecting alignments, BLAST N, of all available alignable 12S, nd1, cox3 and nd3 gene sequences under Panpulmonata (clade) and identifying the most conserved regions to place primers in. No conserved regions of sufficient length were identified in
nd1, cox3 or nd3 leaving 12S with the only viable locations for primers. The 5 prime PCR extender system, high fidelity enzyme system (2200510, 5 PRIME, Inc., Gaithersburg, MD, USA), long distance PCR, was used per the following protocol to amplify the mt genome with initially; 5 within gene fragments then later in 2-3 larger fragments each being 6-8 kb covering the entire mt genome (Appendix C, Sec 3). The thermo cycler program was carried out according to LDPCR manufacturer protocol with individual modifications for G. cristata (Appendix D, sec. 1), P. muscorum (Appendix D, sec. 1) and V. pusilla (Appendix D, sec. 1). Each LDPCR reaction was 25 µl in volume containing: 2.5 µl buffer, 1.25 µl of 8 µM DNTP’s, .2 µl of TAQ (5U/µl), 16.55 µl of molecular water, 1.5 µl of genomic DNA and 1.5 µl of the forward and reverse 10 µM primers. Vertigo pusilla used .5 µl of primer for both the forward and reverse primer and water was used to add up to the final volume. A 1% agarose gel with the stain ethidium bromide was used to ascertain that the LDPCR was producing an amplicon as well as amplifying fragments of the correct size. All reactions were treated with ExoSAP-IT (78201, Affymetrix, Inc., Santa Clara, CA, USA) prior to sequencing to eliminate unincorporated primers and free DNTPs. The amplicons were then completely sequenced via primer walking using exact primers that were designed using previously obtained sequences. All sequencing reactions used the BigDye Terminator v3.1 Cycle Sequencing Kits (4337455, Life Technologies Corporation, Carlsbad, California, USA). A standard thermo cycler protocol (Platt et al. 2007) was used except that the anneal temperature was modified between the different species. All sequencing reactions were 10 µl containing 1-6 µl of ExoSAPed LDPCR product, 1 µl of one 10 µM primer, 2 µl of 5X big dye buffer and 1µl of enzyme with any remaining volume being water. Gastrocopta
*cristata* functioned optimally using 2 μl of ExoSAPed LDPCR product (Appendix D, sec. 2). *Pupilla muscorum* worked best using 4 μl of ExoSAPed LDPCR product (Appendix D, sec. 2). For *Vertigo pusilla* 2 μl of enzyme combined with 6 μl of ExoSAPed LDPCR product (Appendix D, sec. 2). All ethanol precipitations were according to standard procedure (Appendix D, sec. 3). The UNM Molecular Biology Facility ran all sequences, dissolving the DNA pellet in 10 μl of formamide for sequencing using a ABI 3130xl (Applied Biosystems by Life Technologies Corporation, Carlsbad, California, USA).

**Genome annotation:**

Chromatogram editing and contig assembly were done with Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, Michigan, USA). Genes were tentatively identified with a combination of BLAST X and BLAST N searches on NCBI, restricted to Panpulmonata (NCBI:BLAST 2012). Exact beginning and end points were identified by manually aligning the newly sequenced genes with genes from related snail species represented in NCBI, and then compared to find the correct start and stop codons based on length of the genes and the location of the start codon in related species (Rambaut 2012). The tRNAs were identified by a combination of the tRNAscan-SE Search Server with exceptionally low cutoff scores (1 x 10^-66) and by identifying the anti-codon of each possible tRNA in likely locations and with surrounding sequence that could attain the appropriate secondary structure (Lowe *et al.* 1997, Schattner *et al.* 2005). The predicted secondary structures for the tRNA were made using mt-tRNA-Draw and FASTA sequence exports (Youngblood *et al.* 2012). The ribosomal RNA genes were identified using both the sequence similarity between genomes plus BLAST N searches, in addition
to identifying the abutting genes. Genome maps were made with GenomeVx (Conant et al. 2008) and the genomes available from GenBank and the experimentally derived genomes from pupillid snails.

**Data analyses:**

Initially 21 mt genomes from Panpulmonata were acquired from NCBI GenBank (Table 1). These genomes were aligned with the three newly sequenced genomes on a gene-by-gene basis via an amino acid nucleotide alignment with manual correction afterwards (Abascal et al. 2010). The tRNA’s and rRNA’s for the above genomes were visually compared but not sequence aligned. The full mt genomes were analyzed for nucleotide diversity and nucleotide content with a custom Matlab script (Appendix A, sec. 3, 4). The gene-by-gene alignments were concatenated to create protein-only genomes, preserving non-aligning sites. The putative origin of replication was identified by three factors: the region was 40 nucleotides or longer, the region’s A+T % was elevated compared to the surrounding sequence, and the region was not overlapping with any genes (Grande et al. 2008). The D_α/D_δ between these genomes was calculated via a custom Matlab script (Appendix A, sec. 2), which used a window equal to each genes’ size, and then graphed in Excel for comparison. The script counts the raw number of synonymous and non-synonymous differences per pairwise comparison for a coding sequence aligned data set, with regions of no alignment, for all possible comparisons while treating gaps as missing data. The codon usage was calculated using a custom Matlab script (Appendix A, sec. 1) and graphed as pie charts per amino acid for
comparison. The GC skew was calculated using the concatenated set of aligned protein encoding codons, the non-aligning sites removed, with a fixed 210 bp window and a 3 bp step in DAMBE (Xia et al. 2001). The non-synonymous and synonymous divergences between various protein coding sequences of the genomes were calculated using DnaSP set for haploid, mitochondrial, and entire sequence coding DNA (Librado et al. 2009). The raw number of non-synonymous and synonymous changes were divided by the number of sites and multiplied by the percentage of sites considered from the total number of coding sites. The relative rate was calculated using Tajima's relative rate test in Mega 5.0 with substitution model A and the coding-only data set with areas of non-alignment eliminated (Tamura et al. 2011).

**Phylogenetic analyses:**

For phylogenetic analysis the gene-by-gene alignments were re-edited and joined to eliminate areas of non-alignment inside of the genes. Four different reconstruction methods were used: maximum parsimony, nearest neighbor joining, maximum likelihood and Bayesian. The maximum parsimony, nearest neighbor joining, and maximum likelihood analyses were all performed using Mega 5.0 (Tamura et al. 2011). The maximum parsimony and nearest neighbor joining analyses both incorporated amino acid sequence with support values based on 500 bootstrap replicates. The maximum parsimony search method was to use close neighbor interchange on random trees with 15 initial trees and a search level of one. The nearest neighbor joining analysis was based on the number of differences.
The maximum likelihood analysis employed the mitochondrial reversible model in addition to the frequency F+ model, a gamma distribution with invariant sites and 12 categories for the rates among sites and support values based on 500 bootstrap replicates. The heuristic method for the maximum likelihood analysis was nearest neighbor interchange with an automatic initial tree. The Bayesian analysis was generated using Mr.Bayes (Huelsenbeck et al. 2001, Ronquist et al. 2003). The Bayesian analysis is amino acid general time reversible with variable rates and 1,000,000 generations with a burn-in of 250,000 generations. The tree is a consensus tree with a 25% burn-in on the number of trees (~5000). The out-group is Lophiotoma cerithiformis (Bandyopadhyay et al. 2006), a non-pulmonate venomous gastropod.

Results

Complete mt-genomes of three Pupillid species

Genome organization and nucleotide composition

The exact sizes of G. cristata (Figure 1), P. muscorum (Figure 2) and V. pusilla (Figure 3) are 14,060 bp, 14,149 bp and 14,078 bp, all which fall into the typical size range for pulmonate mt genomes (Figure 4). All three mt genomes consist of 13 protein-coding genes, 22 tRNA genes and two rRNA genes with no duplications. Sequencing coverage was three-fold for the majority of the genome and two-fold everywhere else. The genomes’ A+T richness is similar at: 69.2% G. cristata, 71.8% P. muscorum and 72.3% V. pusilla. Gastrocopta cristata (Figure 5), P. muscorum (Figure 6) and V. pusilla (Figure 7) have parallel regions of A+T and G+C richness across their genomes.
Protein coding genes

Pupillid mitochondrial genomes have 13 protein coding genes. Of those genes, cox1, nd6, nd5, nd1, nd4L, cytb, cox2, nd4 and nd2 are on the L-strand while atp8, atp6, nd3 and cox3 are on the H-strand. The protein-coding genes make up ~77% of their mt genomes. Non-standard starting codons are also moderately common and which start codon is used for which gene sometimes varies between different species (Serb et al. 2003, Boore et al. 2004). The genes cox1 and cytb have consistently begun with alternative start codons among the Pupillids. Cox1 starts with the alternative start codon for leucine but for cytb and the rest of the genes the particular start codon utilized varies (Table 2). In G. cristata the common genes, cox1 and cytb, as well as nd3 all commence with non-standard start codons. In P. muscorum the common genes plus nd6, atp6, nd3 and nd2 begin with alternative start codons. In V. pusilla the common genes, in addition to nd1, nd4L, nd3 and nd4, start with alternative start codons. I hypothesize that one gene, nd5, has a stop codon completed via polyadenylation in G. cristata. Furthermore, some protein coding genes also appear to overlap. Gastrocopta cristata has an overlap between nd5 and nd1 of 10 bp. The same overlap occurs in P. muscorum, except it is 16 bp, and additionally there is another overlap between nd6 and nd5 that is 8 bp. Vertigo pusilla has the same overlap as G. cristata, except it is 34 bp in length and there is also an overlap between nd4L and cytb of 15 bp.

Transfer and ribosomal RNA genes

There are 22 tRNA genes and 2 rRNA genes in pupillid land snail mitochondrial genomes and tRNA genes make up ~10% of the mt genomes of the three species. Rather
than on the H-strand (~ 8) most of the tRNA’s (~14) for all three species are on the L-strand with the majority of the genes. There are frequent overlaps between different tRNA’s and between tRNA’s and protein coding genes in all three species. For *G. cristata* the tRNA sizes range from 73 bp in tRNA\textsuperscript{Lys} and tRNA\textsuperscript{Gln} to 61 bp in tRNA\textsuperscript{Thr} and tRNA\textsuperscript{Ser(AGN)}. In *P. muscorum* the size range is between 68 bp for tRNA\textsuperscript{Glu} and tRNA\textsuperscript{Ala} to 61 bp for tRNA\textsuperscript{His}. *Vertigo pusilla* is different from the others in that the range is from 69 bp in tRNA\textsuperscript{Ile} to 52 bp for tRNA\textsuperscript{Arg}. Most, but not all, of the tRNAs for all three species fit the “cloverleaf” secondary structure model. In *G. cristata* there are six tRNAs which do not fit the model (Figure 8), for *P. muscorum* there are three tRNAs (Figure 9), and for *V. pusilla* there are five tRNAs (Figure 10). These tRNAs are referred to as non-standard tRNAs and are identified by the following criteria: missing leaves, the number of missing bonds, number of mispaired bases, and non-standard secondary structure.

Pulmonate snail mitochondria have two rRNA genes, *12S* and *16S*, which is the same with the new genomic sequences. The rRNAs are arranged in the typical manner, *16S* between *cox1* and *nd6* with the tRNAs on either side and *12S* between *atp6* and *nd3* with it’s own tRNAs on either side (Boore 1999). rRNA makes up ~12.5 % of the three pupillid mt genomes. The rRNA in *G. cristata* is mostly composed of A/T with a G/C content of only ~27%. The same is true of the other two Pupillidae except their GC content is ~25%.
Non-coding regions

Most of the mitochondria genome sequences in pupillid snails is taken up by protein coding genes, but there are still gaps between the coding sequence and differences in the distribution of those regions. While the percentages of tRNA genes, rRNA genes and protein coding genes are approximately the same across genomes, the percentage of non-coding DNA varies more. *Gastrocopta cristata* has .01% intergenic DNA, *P. muscorum* has .7% and *V. pusilla* has .06%. This difference is probably not significant however it is interesting to note that the difference in mt genome size between *V. pusilla* and *G. cristata* versus *P. muscorum* can be accounted for almost entirely by an increase in the amount of intergenic space in *P. muscorum*. All three mt genomes have many non-coding regions with most of them being in the 1-14 bp range, the size of the putative origin of replication (POR) being an exception. In *G. cristata* there are two regions that are longer: 26 bp between tRNA\textsuperscript{Ala} and ND6 as well as 14 bp between tRNA\textsuperscript{Gln} and tRNA\textsuperscript{Leu (UUR)}. In *P. muscorum* the most significant non-coding regions are between tRNA\textsuperscript{Val} and 16S (18 bp) and between tRNA\textsuperscript{Ala} and ND6 which is 30 bp. *Vertigo pusilla* has one large non-coding region between tRNA\textsuperscript{Ala} and ND6, which is 23 bp. All 3 species have their POR between COX3 and tRNA\textsuperscript{Ile} and they are all in the 40 bp range. Specifically *G. cristata*’s POR is 46 bp long with 87% A/T, *P. muscorum*’s POR is 42 bp with 81% A/T and *V. pusilla*’s POR is 45 bp long with 89% A/T.

Comparative analyses

Mt genome comparisons
The mt genome lengths compared across Panulmonata are fundamentally the same with one exception, *Pedipes pedipes* (Figure 4). The patterns of nucleotide usage demonstrated by the Pupillids are preserved across genera (Figure 5,6,7,11,12). T is the most favored nucleotide followed by A, C, and G on the leading strand. A/T content is between 55% and 75% for all of the genera reviewed (Figure 13), and the G/C content is between 25% and 45%. All these gastropod mt genomes have the majority of genes on the L-strand rather than the H-strand.

The gene order across species is largely consistent sans the frequent changes in the organization of tRNA (Figure 14). The $D_s$ is greater than one for most comparisons regardless of which species are being compared (Table 4). The $D_n$ was fairly consistent (~.2) across all gene and species comparisons (Table 4). Correspondingly, the $D_n/D_s$ ratios for those comparisons are unreliable due to silent site saturation. While the newly sequenced genomes are not unique in having non-standard tRNA’s, which tRNAs are non-standard is inconsistent across genera except for tRNA$^{\text{Lys}}$ which was non-standard in every species.

**Protein coding genes**

The patterns of nucleotide usage in the protein coding genes differ slightly from the whole genome although the ratios of nucleotides remain the same. The universally favored nucleotide in protein coding genes is A, and T never overtakes it in overall prevalence. The A/T richness of the protein coding genes is not significantly different from that of the overall genome. When comparing the protein coding portion of the mitochondrial genomes against each other, the non-synonymous divergence between
them is 18% while the synonymous divergence is 14%. The G/C skew indicates what nucleotide is mutationally preferred on the L-strand, in this case it is G most of the time with C being usually preferred when genes switch strands (Figure 15). All three genomes show some form of codon bias although slight in some cases. When all three are compared to each other the differences between them are minor. On the whole, codons ending in A or T in addition to codons that are more A/T rich are preferred (Figure 16, 17, 18). More distantly related species are only slightly A/T biased as demonstrated by their more equal codon usage (Figure 19, 20), seeming to indicate that A/T selection is stronger, or at least more apparent, in G. cristata, P. muscorum, and V. pusilla.

When examining the relative evolutionary rates between the three species as compared to A. caerulea, with S. putris as the out-group, the results are non-significant and confirm the null hypothesis (Table 3). The null hypothesis states that all species in the comparison are evolving at the same rate.

**Comparative phylogenetics**

The phylogenetic relationships of the pupillid land snails were inferred using the three newly sequenced genomes and 22 existing genomes. All of the phylogenetic analyses were somewhat similar in results but they did contain significant differences. As such the Bayesian (Figure 21), maximum likelihood (Figure 22), maximum parsimony (Figure 23), and nearest neighbor joining (Figure 24) consensus trees were all included for comparison. Overall the outer clades, clades close to the taxa, are more supported than the inner clades, clades closer to the root.
Despite the fluctuating support values for various clades the groupings themselves are reasonably consistent across analyses. One of the consistent clades was that all three of the genomes recently sequenced clustered together. *Cepaea nemoralis* also persistently groups as a sister with *C. obtusus*. *Ascobulla fragilis* and *E. chlorotica* are sister species and consistently clade with the two *Siphonaria* sister species. The *A. bidentata*, *O. vulcans*, and *T. reticulatus* clade consistently comes out as a sister to the *P. mortoni*, *P. peronii*, and *O. celtica* clade. On average all the species belonging to a particular super family group together.

*Succinea putris*, *M. myosotis*, *P. pedipes*, *P. dolabrata* and the *Biomphalaria* sister species, however, all move depending on the method of analysis. For example, in the Bayseian and maximum likelihood phylogenies *S. putris* is in a clade with *C. nemoralis* and *C. obtusus*. This is unlike the nearest neighbor joining phylogeny where *S. putris* is in a clade with *A. caerulea*, *G. cristata*, *P. muscorum*, and *V. pusilla* or like in the maximum parsimony phylogeny where *S. putris* is sister to *A. caerulea*. The other species move between clades in an analogous manner to *S. putris*, depending on the method analysis.

**Discussion**

When comparing the mt genomes of *G. cristata*, *P. muscorum*, and *V. pusilla* to previously sequenced mt genomes it is apparent that there is a great deal of similarity between them and that there are also some minor differences. The gene orientation, despite re-arrangements, usually remains consistent; the genes start and end in approximately the same locations across all genomes referenced. The protein and rRNA
gene order remains constant while the tRNA’s frequently move around across genera. Despite all of the movement, the tRNA’s are mainly located in the same regions across the genome rather than scattered.

Seventy percent of the genome is protein coding. Of the changes that occur in the protein coding genes most are synonymous. This makes sense based on the expectation that most mutations are synonymous in order to preserve fitness by reducing the number of deleterious changes. An explanation for the amount of non-standard tRNA may be that the higher mutation rates in the mitochondrion compared to the nucleus, in combination with the absence of recombination, results in the accumulation of deleterious mutations by Muller’s ratchet (Lynch et al. 1996). The number of changes, synonymous or non-synonymous between species, also means that designing primers for LDPCR is difficult because species-specific primers are necessary for large amplicons.

All of the mitochondrial genomes examined have high A/T content and low G/C content (Figure 13). The three new genomes are on the higher end of that spectrum, as defined by my dataset at ~77%. The A/T richness is not particular to a certain region; the relative amount of A/T is equal from the protein coding genome to the remainder of the genome. There are several regions where T content is extremely high (Figure 5,6,7). This region is conserved among G. cristata, P. muscorum, and V. pusilla as well as with other species. A. caerulea has a similar region (Figure 11), while C. nemoralis has another similar region but on a different scale (Figure 12). The region of high T content potentially varies between the different species because the comparisons are between distinctive super families. Codons with more A/T are usually favored over codons with more G/C (Figure 16,17,18,19,20). While this relationship exists in the other species of
the data set, they don’t have the same amount of bias towards particular codons (Hatzoglou et al. 1995). Some amino acids which particularly highlight this trend are valine, glutamic acid and tyrosine.

The G/C skew indicates that G is favored by the three genomes over C on the L strand. The fact that the $D_n$ estimates are usually higher than one suggests that the sequences are already saturated at silent sites. The saturation of synonymous sites indicates that $D_n$ and $D_n/D_s$ are unreliable in cases where $D_s$ is greater than one.

The relative rate test which supports the null, that there is no difference in evolutionary rates between the compared species, indicates that any differences in the effective population sizes or phenotypic differences between these snails have little to no effect on the mutation rates.

The phylogeny recovered from this data set is not well conserved across different methods. Any clade containing *S. putris*, *M. myosotis*, *P. pedipes*, *P. dolabrata* or the *Biomphalaria* sister species varies across analyses with low (55% <) support values. This uncertainty about where to place these species might be due to lack of data for intervening taxa, skewing the data based on a few informative sites or perhaps the data set itself does not contain enough informative sites to place these particular species. Another possibility is that the data set is too divergent, despite being an amino acid alignment of protein coding genes with the indels in the alignment concatenated out.

The Bayesian phylogeny is an exception to low support values on the clades containing the above species, however there is reason for the values to be suspect. The maximum likelihood phylogeny doesn’t agree with the support the Bayesian phylogeny assigns to identical nodes by as much as 65%. It has also been suggested that the
posterior probabilities from Bayesian phylogenetics are consistently overconfident (Misawa et al. 2003). The nearest neighbor joining and maximum parsimony phylogenies similarly do not agree with the Bayesian phylogeny’s support values, although the discrepancies are smaller.

When comparing the retrieved phylogenies to previous work by Wade et al. 2006 and White et al. 2011 some striking similarities and differences are revealed. Both authors recovered the same groupings by super family as is retrieved in the four phylogenies above. Interestingly enough among the list of species with low support values and uncertain positions that White uses he also has low support values for those clades but the positions are fixed across his analyses while they varied in mine. There is also a discrepancy in the location of the Biomphalaria clade, White has Biomphalaria as a clade sister to but outside of the Stylommatophorans and the Eupulmonates, while my analyses have Biomphalaria as a clade nested within the non-Stylommatophoran clade.

All of the phylogenies support a few conclusions. The monophyly of the Stylommatophoran snails is supported by all of the phylogenies. The groupings of species by superfamily are supported across all phylogenies. The relations among the super families when using mt DNA seem to depend on the data set and analyses run. Theses phylogenies support the previous work by Wade and White but they don’t provide additional resolution on the relations between the super families and deeper nodes.

**Conclusion**

*Gastrocopta cristata, P. muscorum* and *V. pusilla* genomes’ are the first fully sequenced mitochondrial genomes in their families. These three genomes, *G. cristata, P.*
*muscorum* and *V. pusilla* have been sequenced to provide for the first time coverage of the Pupilloidea super family. These sequences add genera to improve the accuracy of phylogenetics in areas where there previously has been no data.

In addition to improving future taxonomic efforts these three genomes provide improved insight into the general features and organization of pupillid mt genomes. The *G. cristata* (KC185403), *P. muscorum* (KC185404) and *V. pusilla* (KC185405) genomes are not yet released on NCBI GenBank but they can be found in Appendix B. Interesting features such as the region with an increased prevalence of thymine need more work and characterization of mt genomes of comparable genera to more fully understand. This work has hopefully improved the ease of sequencing pupillid mt genomes and adds to the growing body of knowledge about mitochondrial features as well as snail mitochondrial genomic features. Some future extensions of this work will be to cover all of the families, if not genera, that lack representative sequences using a next generation sequencing approach similar to the one suggested by Jex et al. 2010. This will expand the taxonomy and provide additional data on the evolutionary processes shaping pupillid mitochondrial genomes.
Figure 1: Mt genome map of *G. cristata*. Protein coding genes are blue, tRNA are green and rRNA are red.
Figure 2: Mt genome map of *P. muscorum*. Protein coding genes are blue, tRNA are green and rRNA are red.
Figure 3: Mt genome map of *V. pusilla*. Protein coding genes are blue, tRNA are green and rRNA are red.
**Figure 4:** Comparison of genome length across data set.

**Genome length**

![Genome length chart](chart.png)
**Figure 5:** Nucleotide density plot for mt genome of *G. cristata*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*. 
Figure 6: Nucleotide density plot for mt genome of *P. muscorum*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*. 
Figure 7: Nucleotide density plot for mt genome of *V. pusilla*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*.
**Figure 8:** Predicted tRNA structures for *G. cristata*. Non-standard tRNA are indicated with a black star.

The number of missing bonds, missing arms and non-standard spacing determine which tRNA are considered non-standard. tRNA are in 5' -> 3', left to right and are based on DNA sequence.
Figure 9: Predicted tRNA structures for *P. muscorum*. Non-standard tRNA are indicated with a black star. The number of missing bonds, missing arms and non-standard spacing determine which tRNA are considered non-standard. tRNA are in 5' -> 3', left to right and are based on DNA sequence.
Figure 10: Predicted tRNA structures for *V. pusilla*. Non-standard tRNA are indicated with a black star.

The number of missing bonds, missing arms and non-standard spacing determine which tRNA are considered non-standard. tRNA are in 5’ -> 3’, left to right and are based on DNA sequence.
**Figure 11:** Nucleotide density plot for mt genome of *A. caerulea*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*. 
Figure 12: Nucleotide density plot for mt genome of *C. nemoralis*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*. 
Figure 13: A+T and G+C percentages by mt genome.
Figure 14: Gene order comparison across data set. Green boxes are tRNA, red boxes are rRNA and blue boxes are protein coding genes. Underlined genes are in the opposite orientation to non underlined genes.
Figure 15: G/C skew comparison of *G. cristata*, *P. muscorum* and *V. pusilla*. Made using a coding sequence alignment. Upper half is ratio of G and lower half is ratio of C. Positive peaks mean G is more prevalent, negative peaks mean C is more prevalent. Bin size was 210 nucleotides, advancing 3 nucleotides at a time.
Figure 16: Codon usage pie chart for mt genome of *G. cristata* by amino acid. The two serine and two leucine tRNA’s are grouped together under serine and leucine respectively.
**Figure 17:** Codon usage pie chart for mt genome of *P. muscorum* by amino acid. The two serine and two leucine tRNA’s are grouped together under serine and leucine respectively.
**Figure 18:** Codon usage pie chart for mt genome of *V. pusilla* by amino acid. The two serine and two leucine tRNA’s are grouped together under serine and leucine respectively.
Figure 19: Codon usage pie chart for mt genome of *A. caerulea* by amino acid. The two serine and two leucine tRNA’s are grouped together under serine and leucine respectively.
Figure 20: Codon usage pie chart for mt genome of *C. nemoralis* by amino acid. The two serine and two leucine tRNA’s are grouped together under serine and leucine respectively.
**Figure 21:** Bayesian phylogenetic tree for complete data set. Node labels are the posterior probabilities.
Figure 22: Maximum likelihood tree for complete data set. Node labels are bootstrap values.
Figure 23: Maximum parsimony tree for complete data set. Node labels are bootstrap values.
**Figure 24:** Nearest neighbor joining tree for complete data set. Node labels are bootstrap values.
**Table 1:** List of species used for this project and their associated Genbank accession numbers.

<table>
<thead>
<tr>
<th>Species name</th>
<th>Genbank accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Albinaria caerulea</em></td>
<td>NC_001761.1</td>
</tr>
<tr>
<td><em>Ascobulla fragilis</em></td>
<td>NC_012428.1</td>
</tr>
<tr>
<td><em>Auriculinella bidentata</em></td>
<td>NC_016168.1</td>
</tr>
<tr>
<td><em>Biomphalaria glabrata</em></td>
<td>NC_005439.1</td>
</tr>
<tr>
<td><em>Biomphalaria tenagophila</em></td>
<td>EF433576.1</td>
</tr>
<tr>
<td><em>Cepaea nemoralis</em></td>
<td>NC_001816.1</td>
</tr>
<tr>
<td><em>Cylindrus obtusus</em></td>
<td>NC_017872.1</td>
</tr>
<tr>
<td><em>Elysia chlorotica</em></td>
<td>EU599581.1</td>
</tr>
<tr>
<td><em>Gastrocopta cristata</em></td>
<td>KC185403</td>
</tr>
<tr>
<td><em>Myosotella myosotis</em></td>
<td>NC_012434.1</td>
</tr>
<tr>
<td><em>Onchidella celtica</em></td>
<td>AY345048.2</td>
</tr>
<tr>
<td><em>Ovatella vulcana</em></td>
<td>NC_016175.1</td>
</tr>
<tr>
<td><em>Pedipes pedipes</em></td>
<td>NC_016179.1</td>
</tr>
<tr>
<td><em>Peronia peroni</em></td>
<td>NC_016181.1</td>
</tr>
<tr>
<td><em>Platevindex mortoni</em></td>
<td>GU475132.1</td>
</tr>
<tr>
<td><em>Pupilla muscorum</em></td>
<td>KC185404</td>
</tr>
<tr>
<td><em>Pyramidella dolabrata</em></td>
<td>NC_012435.1</td>
</tr>
<tr>
<td><em>Rhopalocaulis grandidieri</em></td>
<td>NC_016183.1</td>
</tr>
<tr>
<td><em>Salinator rhamphidia</em></td>
<td>NC_016185.1</td>
</tr>
<tr>
<td><em>Siphonaria gigas</em></td>
<td>NC_016188.1</td>
</tr>
<tr>
<td><em>Siphonaria pectinata</em></td>
<td>NC_012383.1</td>
</tr>
<tr>
<td><em>Succinea putris</em></td>
<td>JN627206.1</td>
</tr>
<tr>
<td><em>Trimusculus reticulatus</em></td>
<td>NC_016193.1</td>
</tr>
<tr>
<td><em>Vertigo pusilla</em></td>
<td>KC185405</td>
</tr>
<tr>
<td><em>Lophiotoma cerithiformis</em></td>
<td>NC_008098.1</td>
</tr>
</tbody>
</table>
Table 2: List of all start codons, by gene, for *G. cristata*, *P. muscorum* and *V. pusilla*.

Yellow highlighted codons are alternative start codons.

<table>
<thead>
<tr>
<th>Gene</th>
<th>G. cristata</th>
<th>P. muscorum</th>
<th>V. pusilla</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atp8</td>
<td>ATG</td>
<td>ATG</td>
<td>ATG</td>
</tr>
<tr>
<td>Atp6</td>
<td>ATG</td>
<td>TTG</td>
<td>ATG</td>
</tr>
<tr>
<td>Cox1</td>
<td>TTG</td>
<td>TTG</td>
<td>TTG</td>
</tr>
<tr>
<td>Cox2</td>
<td>ATG</td>
<td>ATG</td>
<td>ATG</td>
</tr>
<tr>
<td>Cox3</td>
<td>ATG</td>
<td>ATG</td>
<td>ATG</td>
</tr>
<tr>
<td>Cytb</td>
<td>TTG</td>
<td>TTG</td>
<td>TTG</td>
</tr>
<tr>
<td>Nd1</td>
<td>ATG</td>
<td>ATG</td>
<td>GTG</td>
</tr>
<tr>
<td>Nd2</td>
<td>ATA</td>
<td>GTG</td>
<td>ATG</td>
</tr>
<tr>
<td>Nd3</td>
<td>GTG</td>
<td>TTG</td>
<td>TTG</td>
</tr>
<tr>
<td>Nd4</td>
<td>ATG</td>
<td>ATA</td>
<td>TTG</td>
</tr>
<tr>
<td>Nd4l</td>
<td>ATG</td>
<td>ATA</td>
<td>ATC</td>
</tr>
<tr>
<td>Nd5</td>
<td>ATA</td>
<td>ATG</td>
<td>ATA</td>
</tr>
<tr>
<td>Nd6</td>
<td>ATG</td>
<td>TTG</td>
<td>ATG</td>
</tr>
</tbody>
</table>
Table 3: Relative rate comparisons, using Tajima's relative rate test, between *G. cristata*, *P. muscorum* *V. pusilla* and *A. caerulea* with *S. putris* as the out group. The rate comprison was conducted using Mega 5.0 and coding only sequence with the divergent regions concatenated out.

Unique differences in Gastrocopta cristata 249
Unique differences in Albinaria caerulea 291
Unique differences in Succinea putris (outgroup) 373
p-value 0.07
value needed for significance using the sequential Bonferroni correction 0.0125

Unique differences in Pupilla muscorum 285
Unique differences in Albinaria caerulea 294
Unique differences in Succinea putris (outgroup) 358
p-value 0.7
value needed for significance using the sequential Bonferroni correction 0.05

Unique differences in Vertigo pusilla 261
Unique differences in Albinaria caerulea 275
Unique differences in Succinea putris (outgroup) 351
p-value 0.54
value needed for significance using the sequential Bonferroni correction 0.025

Unique differences in Gastrocopta cristata 187
Unique differences in Pupilla muscorum 237
Unique differences in Albinaria caerulea (outgroup) 501
p-value 0.015
value needed for significance using the sequential Bonferroni correction 0.01

Unique differences in Pupilla muscorum 224
Unique differences in Vertigo pusilla 209
Unique differences in Albinaria caerulea (outgroup) 469
p-value 0.47
value needed for significance using the sequential Bonferroni correction 0.0167

Unique differences in Gastrocopta cristata 190
Unique differences in Vertigo pusilla 244
Unique differences in Pupilla muscorum (outgroup) 308
p-value 0.009
value needed for significance using the sequential Bonferroni correction 0.0083
Table 4: \( D_n/D_s \) between *G. cristata*, *P. muscorum* *V. pusilla*. Calculated using a matlab script (Appendix A, sec. 2) and the coding sequence only with the divergent regions concatenated out.

<table>
<thead>
<tr>
<th></th>
<th>G. cristata vs. V. pusilla</th>
<th>G. cristata vs. P. muscorum</th>
<th>P. muscorum vs. V. pusilla</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( D_n )</td>
<td>( D_s )</td>
<td>( D_n/D_s )</td>
</tr>
<tr>
<td>Atp6</td>
<td>0.221</td>
<td>1.370</td>
<td>0.161</td>
</tr>
<tr>
<td>Atp8</td>
<td>0.699</td>
<td>0.841</td>
<td>0.831</td>
</tr>
<tr>
<td>Cytb</td>
<td>0.170</td>
<td>1.477</td>
<td>0.115</td>
</tr>
<tr>
<td>Cox1</td>
<td>0.049</td>
<td>1.234</td>
<td>0.040</td>
</tr>
<tr>
<td>Cox2</td>
<td>0.125</td>
<td>1.287</td>
<td>0.097</td>
</tr>
<tr>
<td>Cox3</td>
<td>0.161</td>
<td>1.685</td>
<td>0.095</td>
</tr>
<tr>
<td>Nd1</td>
<td>0.249</td>
<td>1.017</td>
<td>0.245</td>
</tr>
<tr>
<td>Nd2</td>
<td>0.450</td>
<td>1.202</td>
<td>0.374</td>
</tr>
<tr>
<td>Nd3</td>
<td>0.360</td>
<td>2.110</td>
<td>0.170</td>
</tr>
<tr>
<td>Nd4</td>
<td>0.287</td>
<td>1.247</td>
<td>0.231</td>
</tr>
<tr>
<td>Nd4l</td>
<td>0.444</td>
<td>1.142</td>
<td>0.389</td>
</tr>
<tr>
<td>Nd5</td>
<td>0.359</td>
<td>1.201</td>
<td>0.299</td>
</tr>
<tr>
<td>Nd6</td>
<td>0.334</td>
<td>0.840</td>
<td>0.398</td>
</tr>
</tbody>
</table>
Appendix A: Section 1

% Codon bias pie graphs by Jason Marquardt

clc
clear all
format('short')

% genome to import
name='no crop protein alignment copy.fasta';

%xhow many entries
x=24;

for i=1:1:x

fastaentry=i;
%=================================
fasta1=fastaentry;
fasta2=fastaentry;
[header, genome]= fastaread(name,'Blockread', [fasta1
fasta2], 'IgnoreGaps', 'True');
len=length(genome);
cb(i)= codonbias(genome, 'pie', true, 'GeneticCode', 5);
h=uicontrol('units', 'normalized', 'style', 'text', 'position',[.2 .95 .6 .05], 'string', header, 'fontsize', 20);
set(h, 'BackgroundColor', [1 1 1]);
saveas(gcf,[header, '.ai'], 'ai')
end
Appendix A: Section 2

% Calculate Dn/Ds by Jason Marquardt

clc
clear all

% name of file to import
name='no crop protein alignment.aln';

% number of sequences
x=24;

% size of gene
win=1554

count=1;
[Headers, Sequences]=multialignread(name);

for i=1:1:x
    for j=1:1:x
        [dn1(:,count) ds1(:,count) vardn1(:,count) vards1(:,count)]=dnds(Sequences{i}, Sequences{j}, 'window', win, 'GeneticCode',5);
        count= count+1;
    end
end
Appendix A: Section 3

% various whole genome nucleotide analysis by Jason Marquardt
clc
clear all
format('short')

% genome to import
name='no crop protein alignment copy.fasta';

%how many entries
x=24;

disp(sprintf('name       len   A   T   G   C'));

for i=1:1:x
    fastaentry=i;
    %=================
    fastal=fastaentry;
    fasta2=fastaentry;
    [header, genome]= fastaread(name,'Blockread', [fastal fasta2], 'IgnoreGaps', 'True');
end

base=basecount(genome);
A=base.A;
T=base.T;
G=base.G;
C=base.C;
disp(sprintf('%s     %d %d %d %d %d',header,len,A,T,G,C));
Appendix A: Section 4

% Nucleotide composition & base counts by Jason Marquardt

clear all
clc

name='vert complete genome copy.txt';
[header, genome] = fastaread(name);

base=basecount(genome);
length=length(genome)
A=base.A
T=base.T
G=base.G
C=base.C

figure
ntdensity(genome)
Appendix B: Section 1

LOCUS Gastrocopta_cristata 14060 bp DNA circular 16-NOV-2012
DEFINITION Gastrocopta cristata.
ACCESSION
VERSION
KEYWORDS
SOURCE mitochondrion Gastrocopta cristata
ORGANISM Gastrocopta cristata Unclassified.
REFERENCE 1 (bases 1 to 14060)
AUTHORS Marquardt,J.D., Adema,C.M., Nekola,J.C. and Bergthorsson,U.
TITLE Mitochondrial Genome Evolution in Pupillid Land Snails
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14060)
AUTHORS Marquardt,J.D., Adema,C.M., Nekola,J.C. and Bergthorsson,U.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2012) Department of Biology, University of New Mexico, 1 University of New Mexico, Albuquerque, New Mexico 87131-0001, USA
COMMENT ##Assembly-Data-START##
Assembly Method :: Sequencer v. 5.0
Sequencing Technology :: Sanger dideoxy sequencing
##Assembly-Data-END##
FEATURES Location/Qualifiers
source 1..14060
  /organism="Gastrocopta cristata"
  /organelle="mitochondrion"
  /mol_type="genomic DNA"
tRNA join(1..13,13999..14060)
  /product="tRNA-Lys"
gene 2..1531
  /gene="COX1"
  /codon_start=1
  /transl_table=5
  /product="cytochrome c oxidase subunit I"
  /translation="MRWLYSTNHKDIGTLYMIFGVWCGMVGTGLSLIRLELGTSGVLMDDHFFNVVVTAAHFVFMMFVFNMIPMGGFGNWMVPLLIGAPDMSPRNMMNWFLLP
PSILLISSMVEGGAGTGWTVYVPLSAGTGSHGASVSDLAIIFSLHLAGMSISLGAINFITTTIFMRAPGMTFERLSFWSILVTVFLLLLSPLVLAGAITMLLTDRPNFTSFDPAGGDDIPILYHFWFGHPFVYILLPFGFIHSLHGNASVKQPFDSLGMJYAMISIG
ILGFIVWAHHVFYMVDVITRAFYTAATMIIAIVPTIKVE5WLMTLYGMSKFDASMYWGLSFVWLAGTGLGGGTGIVLSNSSLIIVLHDTTYVVAFHFLVLSMGAAYFAJFAGFVYWF
PVMGLVLQERLAKAFIQMFIAVNLTFPQPHGLSGMPRSYDPSYFKNQIWSS
FGSLMSIFAVLLFLVILWAEFLSQRPSVMFSAAPYSREWTENFPDHFGLNPSISVL""
tRNA 1537..1600
  /product="tRNA-Val"
rRNA 1601..2650
  /product="16S ribosomal RNA"
tRNA 2644..2706
  /product="tRNA-Leu"
tRNA 2707..2771
  /product="tRNA-Pro"
tRNA 2770..2836
  /product="tRNA-Ala"
gene 2864..3316
  /gene="ND6"
  /codon_start=1
  /transl_table=5
  /product="NADH dehydrogenase subunit 6"
  /translation="MIVLFGLMGMLVFSFCFLRSPVSYGGALIALSTIGALLLSLGSCT
TWYGLYLLFLVYYVGGGVLFLFYIMLSSSNFLQVSVFSKLMGLFLAVLSKLYNFSYPKN
SLGVSLSCECSDFLGLFLGLGGLLLVLFFAVHVFLGQPQVVKND"
CDS 2864..3316
  /gene="ND6"
  /codon_start=1
  /transl_table=5
  /product="NADH dehydrogenase subunit 6"
  /translation="MIVLFGLMGMLVFSFCFLRSPVSYGGALIALSTIGALLLSLGSCT
TWYGLYLLFLVYYVGGGVLFLFYIMLSSSNFLQVSVFSKLMGLFLAVLSKLYNFSYPKN
SLGVSLSCECSDFLGLFLGLGGLLLVLFFAVHVFLGQPQVVKND""
CDS 3318..4969
/product="NADH dehydrogenase subunit 5"
/translation=MSRLPVLLLSCLMLSSLMLSLFYYFVSFKSGETYVIELNLMTLSVSFSFSTLFIIDKLSFSGAVVTLISFCVFSFANTYMSEDNQFRFPLMLAMFVLSMNIFSGLSFVILLLGDWDLGISFLALYQNYQNKTSLGAGYLLTLTNRGDVVIVCVPPFLSVGFLPNIPPMSSYETIIFPLIAALLTSQYPPSAWLPAAMAAAPTYPANHSLVHSTSLVTAGVYLRIILSVCSNSNFASSLSCFGSVCSTVCLGGSLSAEYDLKIAIFSTLSQLGLMMFSLGLKLPNALLHILYAHAMFKALLFLSAGLILISGFSQDLRLGATLTYTPSVVFNFNSSLCLGMPFLSAFSYKHVIEYVAMSETNAISFFLMLIAGALTITISIRTVKLWSNSISGLVFSRPLPMYTVPILCILFTSISGKTSFVMGLIYFTFMTSSYQMLHFMLALGVIMGFLGSGEKSHMLSMMFLWPSNNNTKLYFPWLVKSAKMLDYGWVEPVLSSLNLKQLKHLFLFLENSKYLNSMRVGFCSTLVFCLT"  
gene 4961..5866
/product="NADH dehydrogenase subunit 1"
/translation=MYLFLMKAVILLCVILLAVAFITLLERKVLGYMOKRKGPNKVGIWGIVQVOPLADKLFIEKKEEVLLYMSNNLLEFFFTMLGFVSLSWLNLAPNSMWVNYMFGLLFLCLGTLYVNTALAGWSSNSLYAFLGAIRASAQISYSVEYVLMLLFLPALVSMSMCWSALHTPPSSLVFPLAILWASSLAETNRPFDFAEGESELVSFGHVNEYEGGLFALLLFALYETSHISFMATTTWVFLSWFTDFVLILMICWIFLARGVYPRYRDLLMMLCWKSYLPFAISLLLLYFCFSIMF"  
gene 5868..6143
/product="NADH dehydrogenase subunit 4L"
/translation=MSQLLFILIYGFVILLYSFSSQKHLSLMLISLMVISIYLFFSLGVSNSIGLLLILITFAAEEAALALSISLISILRVMKNDNMMLNSY"  
gene 6150..7262
/product="cytochrome b"
/translation=MRKISLLEGLTLSTPTPLNISIFWSGGSISGLMMGLGIQLVTLGFLSMyTSDISISFSVSVHMRDPVNGWFLRFHANGASFIFELHHLIGCRLYQSYSVQPKTWVMGVTIFLSMATAFVLSWPGQSFVGFATVINLLSAIPYWGVTLVEWVWGGFSGVQATLNRFSLHFLILFIALLVMHLFLEHGGTNPGLQTTHINKVTQFYFTYKDLVGFICVSLVLICAYFFPSLDLDPENFYIAPMVTFPQTPFVFYFLAYAILRIPSCLGQVVALAMSIFILYFLPLSSQMSVPAFNMIQQMLFWVLVVFLLLLSTWLGACPIEDPVSLAGPLTHIIFSYPPMLMIVGNFSVSLK"  
tRNA 7267..7332
/product="tRNA-Asp"  
tRNA 7328..7389
/product="tRNA-Cys"  
tRNA 7390..7456
/product="tRNA-Phe"  
gene 7457..8154
/product="COX2"  
CDS 7457..8154
/product="COX2"
/translation=MSYWSGLNLMDDPSSTIEMMLFHDDHAMILIGIFTLVALLGVKLCNLKSSRTAHEAQLEIFWTPIALLWLALPRLIFLLDEQOSNGIILKATGHQWYWSYEIPSSGTDFSDFYSMVQEQALTNGEFRLLEVDNARPVLPYSESJNVLTTSDDVIHAWALPSGVMKMDAVPGRLNNMGMLYSLTPGVFYGQCEICGANGHSYMPIVVFEVHLNDFINFFGIFCLGS"
tRNA  complement(12114..12174)
  /product="tRNA-Thr"
gene  complement(12175..12952)
  /gene="COX3"
  /transl_table=5
  /product="cytochrome c oxidase subunit III"
  /translation="MFKTPHFLVFSVPWPLLSALLTTCMPGLIGYIIRTPYPYLAYG
MACTAMVAYLWVRDIREATYQGHHTTVYVKGLKGVMLILSFECVFVFAPFHWFS
SLAPSVEGSVWPPVGITTLSAQVPPLLNTSVLLLSGVSTVAHHAIIEEGKNSAISG
LLLLLTLGLYFLQLFQGEKYEKTFSIASDVSYGSTFFMATGFGHLHAVAVGATFLFVCL1
RMILYHFTNTHHVFGLAAAWYWHFVDVVVLFLYVSIFWGS"

CDS  complement(12175..12952)
  /gene="COX3"
  /codon_start=1
  /transl_table=5
  /product="cytochrome c oxidase subunit III"
  /translation="MFKTPHFLVFSVPWPLLSALLTTCMPGLIGYIIRTPYPYLAYG
MACTAMVAYLWVRDIREATYQGHHTTVYVKGLKGVMLILSFECVFVFAPFHWFS
SLAPSVEGSVWPPVGITTLSAQVPPLLNTSVLLLSGVSTVAHHAIIEEGKNSAISG
LLLLLTLGLYFLQLFQGEKYEKTFSIASDVSYGSTFFMATGFGHLHAVAVGATFLFVCL1
RMILYHFTNTHHVFGLAAAWYWHFVDVVVLFLYVSIFWGS"

tRNA  12999..13063
  /product="tRNA-Ile"
gene  13068..14000
  /gene="ND2"
  /codon_start=1
  /transl_table=5
  /product="NADH dehydrogenase subunit II"
  /translation="MMSWYQGVLSFVIELGLPLSLTSNWLWVGGMELSLILLPILO
NLAKSIVPESAMKYFLVQASASAVILITGLNYMEGSEYLLLILFLGLILKLVG
PLFWVIPVVKGLKLYHMAFLLPGKIVPLLMNTMNLDSSFILFLVSSIGTVLLG
ALLGNMNSSIRMLGLASSHSGWFLMGVLGFMYFSYGMLISLYLWYLYLKYPD
NYLVSFLFSLPLPFMLFMPIKMLVLYSLLSMGLPSIVSFSSIAISVLNFYMKFS
FSVLSSSENQKNGAVGVMPLLLLLMGGFILFFI"

BASE COUNT
  4331 a  1915 c  2417 g  5397 t

ORIGIN
  1 tttgggga cttatcaca ctatcatata atagatggtg aactttatat ttatttttgg
  61 agttgtagt ggaagagggt gacagttag ttttgggaa atagtttatt ttttaattttt
  121 tcctgggc ttaatagagc atctattggtgc cttttatctt ctgtttactt cttctttactt
  181 tataatttt tatttttggg gatttgtaat gtttttacttt ttatttttactt tttttcttttt
  241 actatttatt ttattttt tatttttactt ctttttttacttt ctttttttacttt tttttcttttt
  301 ttcttttaatt tatttttactt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  361 aggatttttata tatttttactt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  421 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  481 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  541 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  601 tcctatttatt tatttttactt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  661 actatttttatt tatttttactt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  721 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  781 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  841 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  901 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
  961 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1021 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1081 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1141 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1201 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1261 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1321 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1381 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1441 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1501 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1561 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1621 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1681 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1741 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1801 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1861 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1921 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
1981 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
2041 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
2101 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
2161 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
2221 ttttttttttt ttttttttttt ctttttttacttt ctttttttacttt tttttcttttt tttttcttttt
**Appendix B: Section 2**

LOCUS       Pupilla_muscorum       14149 bp    DNA     circular     16-NOV-2012
DEFINITION Pupilla muscorum.
ACCESSION
VERSION
KEYWORDS    .
SOURCE mitochondrion Pupilla muscorum
ORGANISM  Pupilla muscorum
Unclassified.
REFERENCE   1 (bases 1 to 14149)
AUTHORS   Marquardt,J.D., Adema,C.M., Nekola,J.C. and Bergthorsson,U.
TITLE Mitochondrial Genome Evolution in Pupilled Land Snails
JOURNAL Unpublished
REFERENCE   2 (bases 1 to 14149)
AUTHORS   Marquardt,J.D., Adema,C.M., Nekola,J.C. and Bergthorsson,U.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2012) Department of Biology, University of New Mexico, 1 University of New Mexico, Albuquerque, New Mexico 87131-0001, USA
COMMENT  ##Assembly-Data-START##
Assembly Method :: Sequencher v. 5.0
Sequencing Technology :: Sanger dideoxy sequencing
##Assembly-Data-END##
FEATURES             Location/Qualifiers
source          1..14149
    /organism="Pupilla muscorum"
    /organelle="mitochondrion"
    /mol_type="genomic DNA"
tRNA            join(1..8,14092..14149)
    /product="tRNA-Lys"
gene            2..1531
    /gene="COX1"
    /transl_table=5
    /product="cytochrome c oxidase subunit I"
    /translation="MRWLYSTNHKDIGTLYMIFGVWCGMVGTGLSLIRLELGTSGVLMDDH
FNYVNTAFAHVFMIMFFMMVPKMIGGFNWMVPLLGAIGAPDSFPRMNMSSFWL
LPFSFLISSSMVVEGAGTGWTVPYLSSMVGHGASVDAIFFSLHLAGMSCSC
INFITTFNMRASGVTTLERLSLFVWSILVTVELLLSLPVLAGAIFMLLTD
RNENTSSFDPA
AGGEGDPEIYQLFLWFPGHEPVVILILPFGIIHSLGMASKQPPFSLGMIYAMISIG
ILGFIWVAHMHFTVGMVDTRAYAATMTIAIPGTGKVFSLMLTYGLGMNKF
FDASMYYWLVGFILFTLGGLTGILVLSNSLDLVHLHTYYVVAHVHLYVLS
MGAVFAIFAGFYNYWP
PVMGTGLIQERLAKAQNFIMIAVNLTFPPQHFLGLAMPRSYDSPDYTF
WKQISSFGSLSMIFAYLVLCILVVEAFLSERSVLFSKAAIYSREWVENNP
FEFHSNIIPSIVR"
tRNA            1542..1603
    /product="tRNA-Val"
tRNA            1622..2266
    /product="16S ribosomal RNA"
tRNA            2661..2725
    /product="tRNA-Leu"
tRNA            2723..2789
    /product="tRNA-Pro"
tRNA            2790..2857
    /product="tRNA-Ala"
gene            2888..3340
    /gene="ND6"
    /transl_table=5
    /product="NADH dehydrogenase subunit 6"
gene            3333..5000
    /gene="ND6"
CDS             2888..3340
    /gene="ND6"
    /codon_start=1
    /transl_table=5
    /product="NADH dehydrogenase subunit 6"
    /translation="MSLLFLILASFIMFTSEQPAALGGSLILMSFLALSFSFMSF
SWYGLMLFLIYVGLLVLFLMYVMILSSNIFLSKLSTTNGFVFLMMFWALKKN
LPSKI"
CDS  3333..5000
    /gene="ND5"
    /codon_start=1
    /transl_table=5
    /product="NADH dehydrogenase subunit 5"
    /translation="MSKIYRLPFLLFLFSLALSSVMFSFKSMVMDIKLJISGS
SWTITMLFDFKISKISFSMVCLISCFVLFMCQYMEKDEFKFRFGMILLLVFLSMNIL
IFGSFSGMLMLGWDGLGIIIASSFFQAGFLLMIMTNRIDGFLIIASIPFFL
SGGNMFFMSMEILFYVWWVLAVAILLKSQAQFPSSWLPAAMAPPTPVSAVLSHSTLT
VTAGVFKLWVSMFNLSDECSLILLISFGGGVTCFLGVSALYENDVKKIIAFTSLILQOFG
LMVFCFLGMNMFLALLIIFTHAMFKAMLFLSAGILSISFSGVQDRLRLGALQFRSPLL
ICFNVSSCLCMGAPFMSAFAKSHYIVIEIMFMSSVNTLGVAMLVGVISFTISYVRLV
KIJSWKJISCMSNLSTLTIFSYPFMIYLFMSLYSGKMFLMDSHINIFMTPLTQLV
LLNFLVVGILVLGMKGERISHLFGSLFWSHNNINKFMTRSMKLTTSDLGWE
PKNLINNYLFISSFLENLSNYMKFSLLSLFKGSAFMVLMWYFF"

gene  4985..5890
    /gene="ND1"
CDS  4985..5890
    /gene="ND1"
    /codon_start=1
    /transl_table=5
    /product="NADH dehydrogenase subunit 1"
    /translation="MMFFLIKSLFFCLCVLIAVAFFTLLERKVLGFLQIRKGPNKVGV
WGIFQPLSDLKLFVKKEFKYSSNNVLFLJIPPGFLCSLMKMLDNLPSAYNMFLY
SLLLFECVTGLNYYVTALAGWASNSIYAFGLAMRASAOQTSYEISLVMMLLPALJISG
SMCWFFNMKSFPSFVSMPPFLLLIITSLAETRNPADFAEGESEYLSGFDNEYESGL
FALLFLAEYTLIFMCMSMTVVLWMSNTPLSFLWTMFFFFLIFLWRSRFCFPRFYDDL
MMLCWKSFLPPSMCCLYTGFLML"

gene  5905..6187
    /gene="ND4L"
    /transl_table=5
    /product="NADH dehydrogenase subunit 4L"
    /translation="MIINFFSVFVIFMVLLVMFMMQNKHIINSLILEGMMLMVVSF
IFFILASSGEELVMVLLFTFSASEAAAAGLSLISIFSLRYGDSMNMMNF"

gene  6191..7300
    /gene="CYTB"
CDS  6191..7300
    /gene="CYTB"
    /codon_start=1
    /transl_table=5
    /product="cytochrome b"
    /translation="MRKIKIIESLSALPSPVNISVWWSGSSLGLMLLGILQITFLGFLS
MHYYADTLASFDSVTMRDPVWGVVFLRFLHANGASFFFFLFYLVHVRGGLYQSYLQ
PKTMWAGVTFIFILSSAAALFLGYYLPWQMSFGWATVTINLLSAIPYWGELVQWVWG
FSVQQATLNRFFSLHFLFLVFLVMVSLLHLVLHFLHEKGSSNLPHGFHINKIAHFYPFT
YKDVLGFVGLFLFSLLMIFCMSFYPPFLSLDPENFMNANPMVPTPHQPSFYFLFAYAILRS
IPSKLGQVIALAMLMLILYFLPFLHLGKISPSFNIFYQVFNVWILVVFILLTWLGGAC
PIEPYFVLSAFLTVWYFFSFKLNFII""

tRNA  7306..7370
    /product="tRNA-Phe"
    /product="tRNA-Phe"
    /product="tRNA-Phe"
    /product="tRNA-Phe"
    /product="tRNA-Phe"

tRNA  7372..8037
    /product="tRNA-Tyr"
    /product="tRNA-Tyr"
    /product="tRNA-Tyr"
    /product="tRNA-Tyr"
    /product="tRNA-Tyr"

tRNA  8114..8174
    /product="tRNA-His"
    /product="tRNA-His"
    /product="tRNA-His"
    /product="tRNA-His"
    /product="tRNA-His"

tRNA  8180..8245
tRNA complement(12209..12271)  
/product="tRNA-Thr"  
gen complement(12279..13058)  
gen="COX3"  
CDS complement(12279..13058)  
gen="COX3"  
codon_start=1  
/transl_table=5  
/product="cytochrome c oxidase subunit III"  
/translation="MFKTPFHLVEFSPWPLLSSLALTSMPLGLIYYLRFNYWLLLFGMLTAFAYLWWRDIRESTYQGYHTTVVSGIKLGVMFLFVSEVCFFAFAWWFVSHSLAPAVFVWVPPMGTTLPTFVPLNTSVVLLSGVSYTVSWAIJHEEGKLSATSSLLVTVLGLYFLFQLFEGYKETSFISADSVYGSTFFMATGFGHMHVLVGATFLLLICLIRLMLYHFNKSHVHGVFLAAAYWHVFVDVWFLFVSYIWWG"  
tRNA 13101..13166  
/product="tRNA-Ile"  
gen 13170..14117  
gen="ND2"  
codon_start=1  
/transl_table=5  
/product="NADH dehydrogenase subunit 2"  
/translation="MSILMVFFSTIVFLGPILSGSVSSWFLVAGMELSFISIFPVMYNPKSSMSESTIKYFLQVQSCASAMILVSGLFIVEMTMNFYSLFLTGLMLKGLVPMFHFWVIPMVKCGFTDMIFILGPKMIIPFLLIIILSISSIPNLLLIGIFMSLFSVIAGALGNNNSSVRGMALSAVISVHSGWLLMCGLVMGMYLSIYLQMYPLFLLFFCLNQKKFMSVLMFLSGGPLFMPMFLFKLYLQNIYKEVLSVLLFTIFALISLNYFELKSYSFYLSMNLIKIFITNLMSLLGFILFIFFFFYGRAKH"  
BASE COUNT  
4601 a   1830 c   2161 g   5557 t  
ORIGIN  
1 tttgcgatga ttatattcaa cgaatcataa agatattggt actttatata taatttttgg  
61 agtttgatgt gg tatggttg gaactggtct ctcactactt attcgtcttg aactaggtac  
121 ctctggagtt ttgatagatg atcattttta taatgttatt gttacagctc atgcttttgt  
181 aataattttt tttatagtaa tacctattat aattggagga ttcggaaatt gaatagtccc  
241 tcttctcatt ggtgccccag atataagttt tcctcgaa ta aataatatga gtttttggct  
301 tcttcctcca tcttttattt tacttattag ttcaagaatg gtagaaggag gggcaggaac  
361 aggtgaaat gcatacctcc ctaaaagagc aagtagagc acgagagctg ctaaaagctgct  
421 cattacaca attttttat caggtctcct ttgggacttt gccattaactg aatgctttttgct  
481 ctgatcgtt attttttata tc acttttttct aaagtagagc aagtagagc gc aatcaactg  
541 tattgttatttt aatatggtgg gatttttttt catttttttt tatttttttt tatttttttt  
601 ggttatttattt cttctctatt ctttttttagg gcttttttagg aatgctttttg ctttttttagg  
661 aggagagggc ataaatttttt actttttttt tatttttttt tatttttttt tatttttttt  
721 tatttttttta tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
781 aatatcgtt aatgctttttg catttttttt tatttttttt tatttttttt tatttttttt  
841 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
901 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
961 aatattttttt gataatctttt ggtattttttt gtttttttttt gtttttttttt gtttttttttt  
1021 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1081 ttctttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1141 ttctttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1201 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1261 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1321 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1381 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1441 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1501 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1561 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1621 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1681 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1741 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1801 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1861 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1921 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
1981 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
2041 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
2101 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
2161 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
2221 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt  
2281 tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt tatttttttt
Appendix B: Section 3

LOCUS Vertigo_pusilla 14078 bp DNA circular 16-NOV-2012
DEFINITION Vertigo pusilla.
ACCESSION
VERSION
KEYWORDS .
SOURCE mitochondrion Vertigo pusilla
ORGANISM Vertigo pusilla
Unclassified.
REFERENCE 1 (bases 1 to 14078)
AUTHORS Marquardt,J.D., Adema,C.M., Nekola,J.C. and Bergthorsson,U.
TITLE Mitochondrial Genome Evolution in Pupillid Land Snails
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14078)
AUTHORS Marquardt,J.D., Adema,C.M., Nekola,J.C. and Bergthorsson,U.
TITLE Mitochondrial Genome Evolution in Pupillid Land Snails
JOURNAL Submitted (16-NOV-2012) Department of Biology, University of New Mexico, 1 University of New Mexico, Albuquerque, New Mexico 87131-0001, USA
COMMENT #/Assembly-Data-START##
Assembly Method :: Sequencher v. 5.0
Sequencing Technology :: Sanger dideoxy sequencing
#/Assembly-Data-END#
FEATURES Location/Qualifiers
source 1..14078
/organism="Vertigo pusilla"
/organelle="mitochondrion"
/mol_type="genomic DNA"
tRNA join(1..7,14020..14078)
/product="tRNA-Lys"
gene 1..1530
/gene="COX1"
CDS 1..1530
/gene="COX1"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit I"
/translation="MRWLYSTNHKDIGLYMFSTIIWCMGVTGLILLRELELGTSGVFLDDHFPWVVTATAHFMIFVMVMPFMGFGNVWMVPLIGAPDMSFPRMNMSFWLLPFSILLICSSMVGGAATGWTVVYPLSSMVGHGASVDLAIFSLHLAGMSSILGAINFDLSFGMSGGTMERLSLFVWMSLVTVFLLLSLPVLAGAITMLTDRNFTSFDFPAGGDDPILYQHLFWSFGHPFEYVILLPFGGISVLGNASSKQPPFSGLMYAMISGLILGFIWAWHMFTVGMVDTRAYFTAAATMVIAVPITKIFWSLMTLYGMNTKLDASMYWVGLFGLFLLTGLLTSVMSSLVHDTYVVAFHYVVMGAVFAIFAGFYVWFPVMGTLINLERLAKAQFIFMSAVNLTFPPHQFHLGAMPPRYSDYPSYYFKWNQISSFGSLMSSVFAVFLFVLIWVEAFLSQRSVAFSGIAIYSSREAENYPPDFHSVNPISSIVLI"
tRNA 1533..1598
/product="tRNA-Val"
rRNA 1601..2647
/product="16S ribosomal RNA"
tRNA 2641..2703
/product="tRNA-Leu"
tRNA 2705..2771
/product="tRNA-Pro"
tRNA 2769..2830
/product="tRNA-Ala"
gene 2854..3312
/gene="ND6"
CDS 2854..3312
/gene="ND6"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 6"
/translation="MTTTCMALSLVLSCFLQSPVALGGLFLSIFLSVSVGFQMSSWYGCLIFMRYYYGVGLVLFLYMLMSNFYKLASKGKLFLFLMTSSLFMYLFTTSEFLGKTVLGFSQNECSLDSLFLSFLSLGLLLLAFFISIVHVVLKGSILVSV"
/gene="ND5"
    3314..4972
/gene="ND5"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 5"
/translation="MGRLPVLLFMYFLYLYMSFLFLWSWQYETFLYEVSLMFVGSKJ
    TLMFKGDSMSMSFLVYVTISAVFVWFAMEYMSGDEFKRFEMILFMFVLVSMNLILYS
    GSFVLLVGYWDGLGSSFAHYIQSNSTSLLKAYGTYLTLTNRIGDIVLICISPMMMSG
    NLYFFSVDHSLFLTLVVLALAKTSAQYPFSWLPAAAMAPTVPVSAIUHVSSTLVTAGV
    YLVSRSLNLNILEEVCQLLFLSFSGATKCIILGFSAIFENDIKKAIFSTLSQGLLMMF
    CLGGLGIPULLALHLHFTAMFKALFLSAGLILMSMFSTQDQLRQLGLLFRSPPLLVLVF
    NMISSCLVGAPFVSANYKSHAIYEQMMMGMLNSFSFLMMGVIYVTVTRLKALS
    WGVKVSGLVSQYLKLLSFYPLLFLGLSIESGKMFSLDLSMTQSLFVASWQQLLNF
    LLVGFVSMGLWSSPKNSMFFSSMFLWSSWNSKSSVFLGILKQSKVLDYGWLEPKIY
    FEVSLFSSFIQJKMFLGVSSLILPGFSVGLIVSSCIFYW"
gene
    4954..5859
/gene="ND1"
/CDS
    4954..5859
/gene="ND1"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 1"
/translation="MYLFLMKSILVCLCFCLFLIAFSAFRNQLMSNLIILEGLMLICIVFVS
    VLLSVAELGSKSILMLLTLFAAIIEAALGLLALLVFNSLRSYGNDCLNMSQFFAKN"
gene
    5860..6153
/gene="ND4L"
/CDS
    5860..6153
/gene="ND4L"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4L"
/translation="MVIKLSYLFLCFCLFLIAFSAFRNQLMSNLIILEGLMLICIVFVS
    VLLSVAELGSKSILMLLTLFAAIIEAALGLLALLVFNSLRSYGNDCLNMSQFFAKN"
gene
    6140..7252
/gene="CYTB"
/CDS
    6140..7252
/gene="CYTB"
/codon_start=1
/transl_table=5
/product="cytochrome b"
/translation="MSFWGAMNLMDPSSPIQSEMILFHDHAMILLIGIFVLVACLGVKL
    LCVNQVSSRTTHEAQLLEILWTIIPAFLLVWLALPSLRLLYLLDEQSSNGLILKATGH
    QWYWSYEIPSIGSSFDSSMYIQENSLOQENGFEFRLEVDNPILPYNMTINTLTTSTDV
    IHAWAPSMLGKMADVPGRLNMGIQSFSPGVFYGCSEICANHYSYMPIVVEFLSLED
    FINFFGIFCLSG"
tRNA            complement(12134..12193)
    /product="tRNA-Thr"
    /gene            complement(12184..12972)
    /gene="COX3"
    /transl_table=5
    /product="cytochrome c oxidase subunit III"
    /translation="MIKTPFHVLVESFPWPLGSTLTCPMLGLVYYLYLHISNLGMFG
IIATALVAYLWDRVESTYGFHSTSIFGSLKLGFLFVI5VEVCFFAFFWAFYHFS
SLAPAVGSGSPPGYGVS3LTAFQPVLNTRSLGLGSVSTVWAHJAIIESKYLSAIYS
LIFTCLGFYFLEFQGEYEEFSD6SIDSYGVSTFIATGFHGLHVAVGATFLMVCLM
RMYLYHFSNHHYGVFLA2AAYWYHFVDVVWLFLLYFYYWGWSPQKP"

CDS             complement(12184..12972)
    /gene="COX3"
    /codon_start=1
    /transl_table=5
    /product="cytochrome c oxidase subunit III"
    /translation="MIKTPFHVLVESFPWPLGSTLTCPMLGLVYYLYLHISNLGMFG
IIATALVAYLWDRVESTYGFHSTSIFGSLKLGFLFVI5VEVCFFAFFWAFYHFS
SLAPAVGSGSPPGYGVS3LTAFQPVLNTRSLGLGSVSTVWAHJAIIESKYLSAIYS
LIFTCLGFYFLEFQGEYEEFSD6SIDSYGVSTFIATGFHGLHVAVGATFLMVCLM
RMYLYHFSNHHYGVFLA2AAYWYHFVDVVWLFLLYFYYWGWSPQKP"

BASE COUNT     4585 a   1728 c   2181 g   5584 t
Appendix C: Section 1


<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>12sF-85</td>
<td>5-GACCGCGGTTGCTGGCACCTA-3</td>
</tr>
<tr>
<td>12sR-84</td>
<td>5-AGGTGCCAGCAACCGCGGTCA-3</td>
</tr>
<tr>
<td>12sF-83</td>
<td>5-AATTATATGGTAAGTTCCCTAA-3</td>
</tr>
<tr>
<td>12sR-82</td>
<td>5-TYAAACTTAAAWRTTATGGCGG-3</td>
</tr>
<tr>
<td>CobF.1</td>
<td>5-CATTTCGAGGTCAACAGTAATTAC-3</td>
</tr>
<tr>
<td>Cytb811R.1</td>
<td>5-GCRTAYAGAAAAGTATCATTTCWGG-3</td>
</tr>
<tr>
<td>Cox2F</td>
<td>5-AAATAATGCTATTTTATGAYCAYGC-3</td>
</tr>
<tr>
<td>Cox2R</td>
<td>5-GCTCGCAAAATCCTCCTGARCYTG-3</td>
</tr>
<tr>
<td>Cox1F</td>
<td>5-GGTCACAACAAATCATAAAGATATTGG-3</td>
</tr>
<tr>
<td>Cox1R</td>
<td>5-TAAACTTCAAGGTGACCAAAAATCA-3</td>
</tr>
<tr>
<td>16sF</td>
<td>5-GCGCTGTTTATCAAAAACAT-3</td>
</tr>
<tr>
<td>16sR</td>
<td>5-GCCGGTGCTGAACCTCATCAT-3</td>
</tr>
</tbody>
</table>

Appendix C: Section 2

The relative locations and orientations of universal primers.
Appendix C: Section 3
A list of working amplicons by species and region. Use of custom primers is indicated with a (C). Universal primers were used to get initial sequence inside of the genes then that sequence was used to design the custom primers.

**G. cristata**
- cox1F --> 16SR
- cox2F --> cox1R
- 16SF --> cobR
- cobF(C) --> cox2R(C)
- cox2F(C) --> 12SR-82/12SR-84
- cox1R(C) --> 12SF-85/12SF-83
- cox1F --> cox1R
- 16SF --> 16SR
- cox2F --> cox2R
- cobF --> cobR

**P. muscorum**
- cobF --> cox2R
- cox1F --> 16SR(C)
- 16SF(C) --> cobR(C)
- cox2F(C) --> cox1R(C)
- cox1F --> cox1R
- 16SF --> 16SR
- cox2F --> cox2R
- cobF --> cobR
- 12sF --> 12SR (used only for sequencing)

**V. pusilla**
- cox1R(C) --> cox2F(C)
- cox1F(C) --> cox2R(C)
- cox1F --> cox1R
- 16SF --> 16SR
- cox2F --> cox2R
- cobF --> cobR
- 12sF --> 12SR (used only for sequencing)
Appendix D: section 1

LDPCR thermo-cycler program for *G. cristata* (a), *P. muscorum* (b) and *V. pusilla* (c).

1. 93°C for infinity (hot-start, manually continue after reactions loaded)
2. 93°C for 3 minutes
3. 93°C for 15 seconds
4. TM for 30 seconds
   a. 55°C
   b. 55°C
   c. 50°C
5. 68°C for 10 minutes
6. Go to #3 - 9 times
7. 93°C for 15 seconds
8. TM for 30 seconds
   a. 55°C
   b. 55°C
   c. 50°C
9. 68°C for 10 min +20 seconds a cycle
10. Go to #7 – 10 times
11. 4°C for infinity
Appendix D: section 2

Sequencing thermo-cycler program for *G. cristata* (a), *P. muscorum* (b) and *V. pusilla* (c).

1. 96°C for 1 minute
2. 96°C for 10 seconds
3. TM °C for 5 seconds
   a. 55°C
   b. 51°C
   c. 51°C
4. 60°C for 1 minute and 15 seconds
5. Go to 2, 14 times
6. 96°C for 10 seconds
7. TM °C for 5 seconds
   a. 55°C
   b. 51°C
   c. 51°C
8. 60°C for 1 minute and 30 seconds
9. Go to 6, 4 times
10. 96°C for 10 seconds
11. TM °C for 5 seconds
    a. 55°C
    b. 51°C
    c. 51°C
12. 60°C for 2 minutes
13. Go to 10, 4 times
14. 4°C for infinity
Appendix D: section 3

Ethanol precipitation protocol for all BigDye sequencing reactions.

1. Add 10 µl of sodium acetate
2. Add 30 µl of ice cold 100% ETOH
3. Invert 4X
4. Set in dark 4°C space for 15 minutes
5. Spin down and transfer contents from PCR tube to 1.7 ml eppendorf tube
6. Centrifuge at 13,000 rpm, 4°C, for 30 minutes
7. Invert into paper towel and flick GENTLY
8. Add 35 µl of ice cold 70% ETOH
9. Centrifuge at 13,000 rpm, 4°C, for 15 minutes
10. Invert into paper towel and flick GENTLY
11. Vacufuge at 45°C for 10 minutes
12. Submit to MBF for sequencing
**References**


Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood,
Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution, 28*(10), 2731-2739.


83

*Journal of Heredity, 92*, 371-373.
