

5-1-2013

Mitochondrial Genome Evolution in Pupillid Land Snails

Jason D. Marquardt

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

Recommended Citation

Marquardt, Jason D.. "Mitochondrial Genome Evolution in Pupillid Land Snails." (2013). https://digitalrepository.unm.edu/biol_etds/75

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.

Jason D Marquardt

Candidate

Biology

Department

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 (Avice *et al.* 1987, Avice 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 super families (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

ndl, *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 (1×10^{-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_n/D_s 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^{Lys} and tRNA^{Gln} to 61 bp in tRNA^{Thr} and tRNA^{Ser (AGN)}. In *P. muscorum* the size range is between 68 bp for tRNA^{Glu} and tRNA^{Ala} to 61 bp for tRNA^{His}. *Vertigo pusilla* is different from the others in that the range is from 69 bp in tRNA^{Ile} to 52 bp for tRNA^{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^{Ala} and *ND6* as well as 14 bp between tRNA^{Gln} and tRNA^{Leu (UUR)}. In *P. muscorum* the most significant non-coding regions are between tRNA^{Val} and *16S* (18 bp) and between tRNA^{Ala} and *ND6* which is 30 bp. *Vertigo pusilla* has one large non-coding region between tRNA^{Ala} and *ND6*, which is 23 bp. All 3 species have their POR between *COX3* and tRNA^{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 Panpulmonata 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^{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. vulcani* 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 Bayesian 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. These 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.

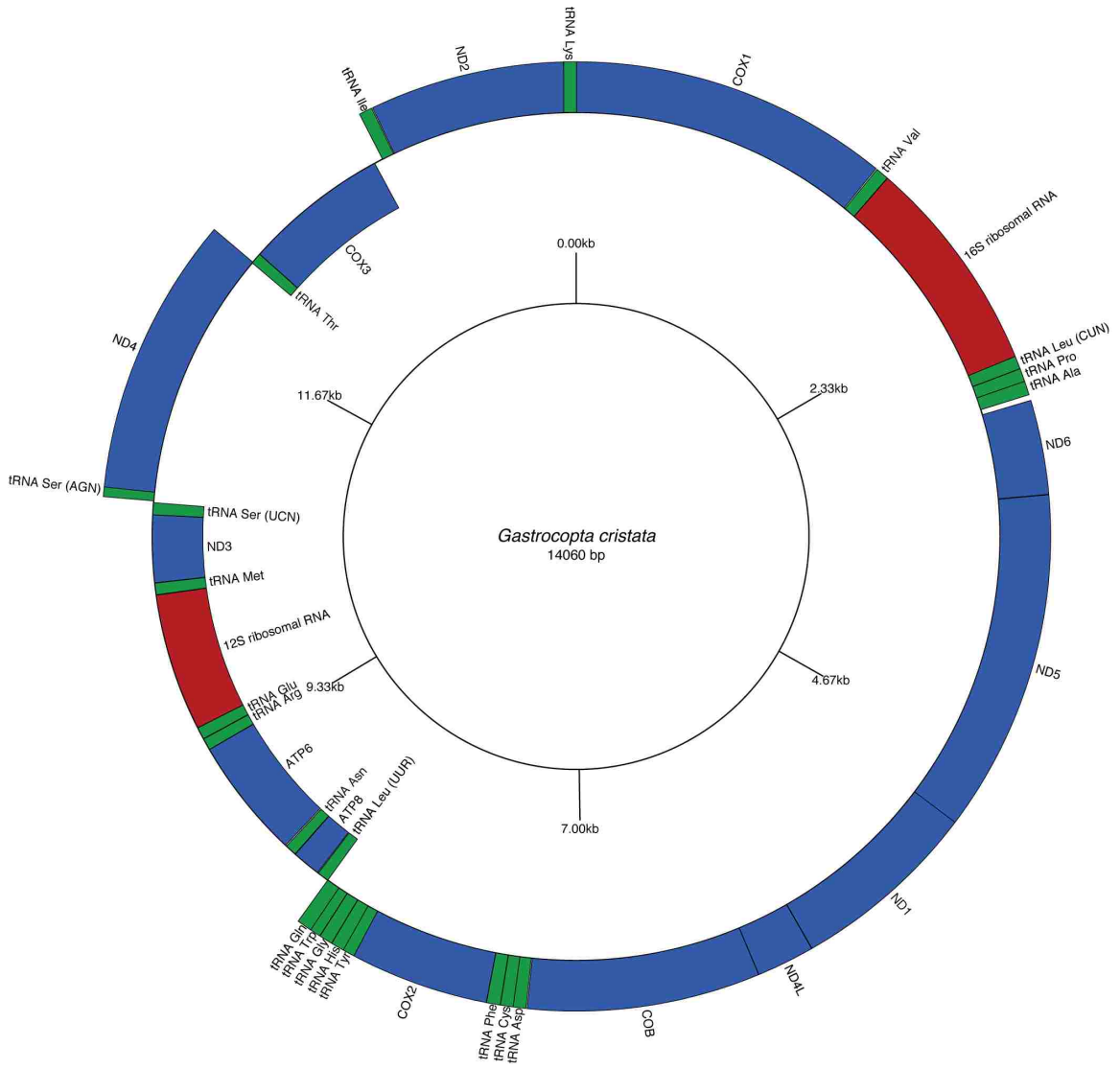


Figure 2: Mt genome map of *P. muscorum*. Protein coding genes are blue, tRNA are green and rRNA are red.

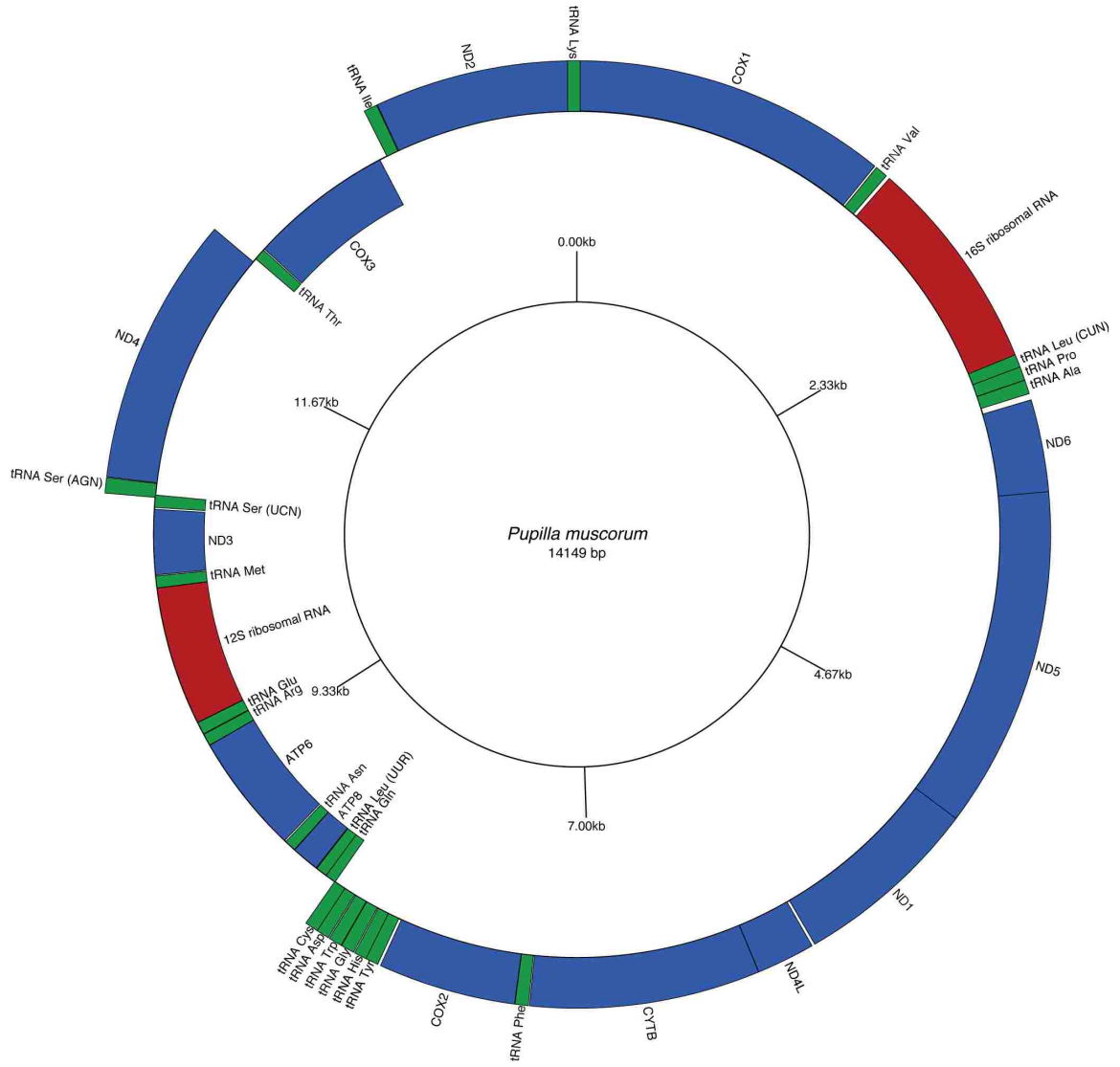


Figure 3: Mt genome map of *V. pusilla*. Protein coding genes are blue, tRNA are green and rRNA are red.

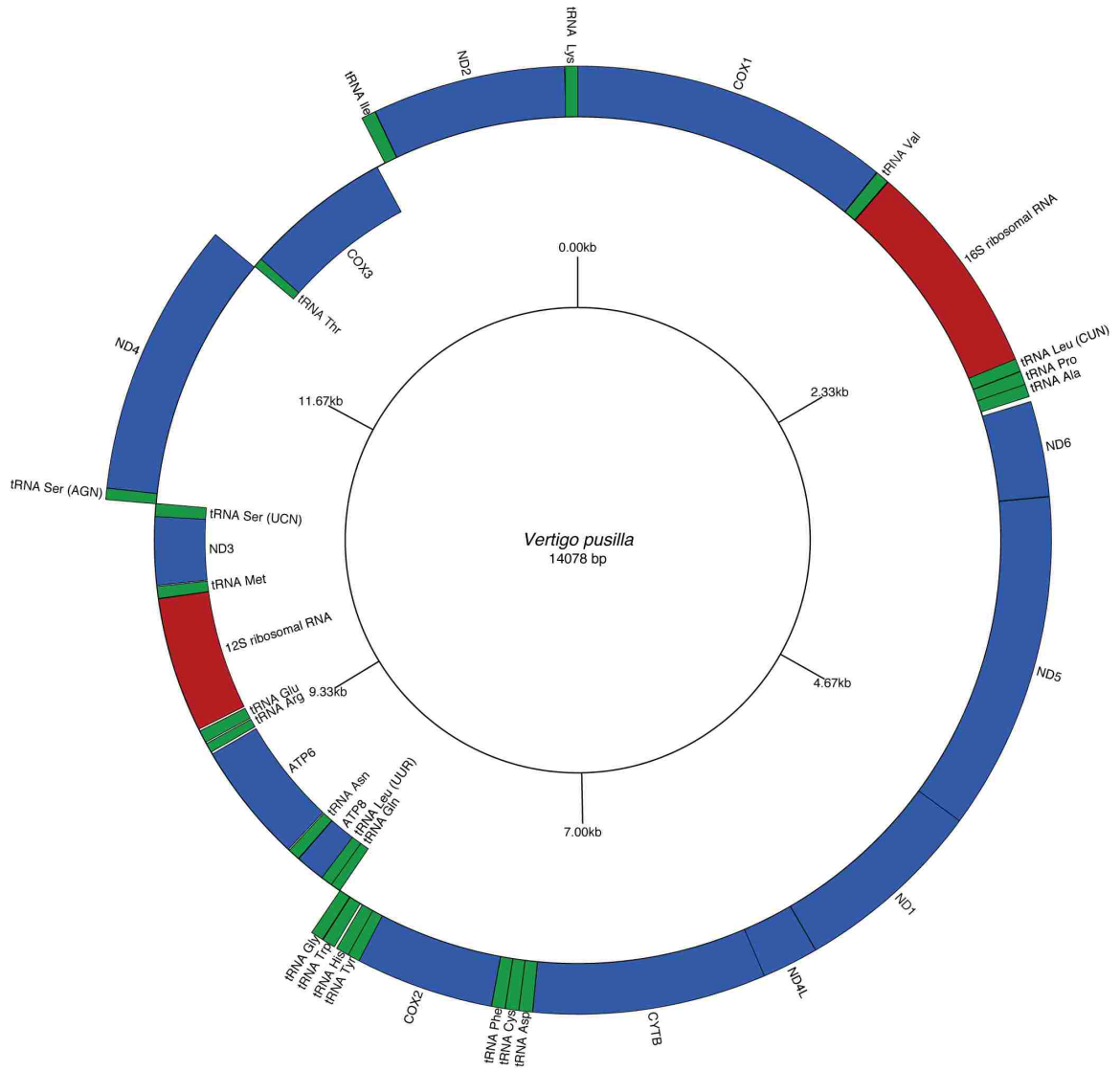


Figure 4: Comparison of genome length across data set.

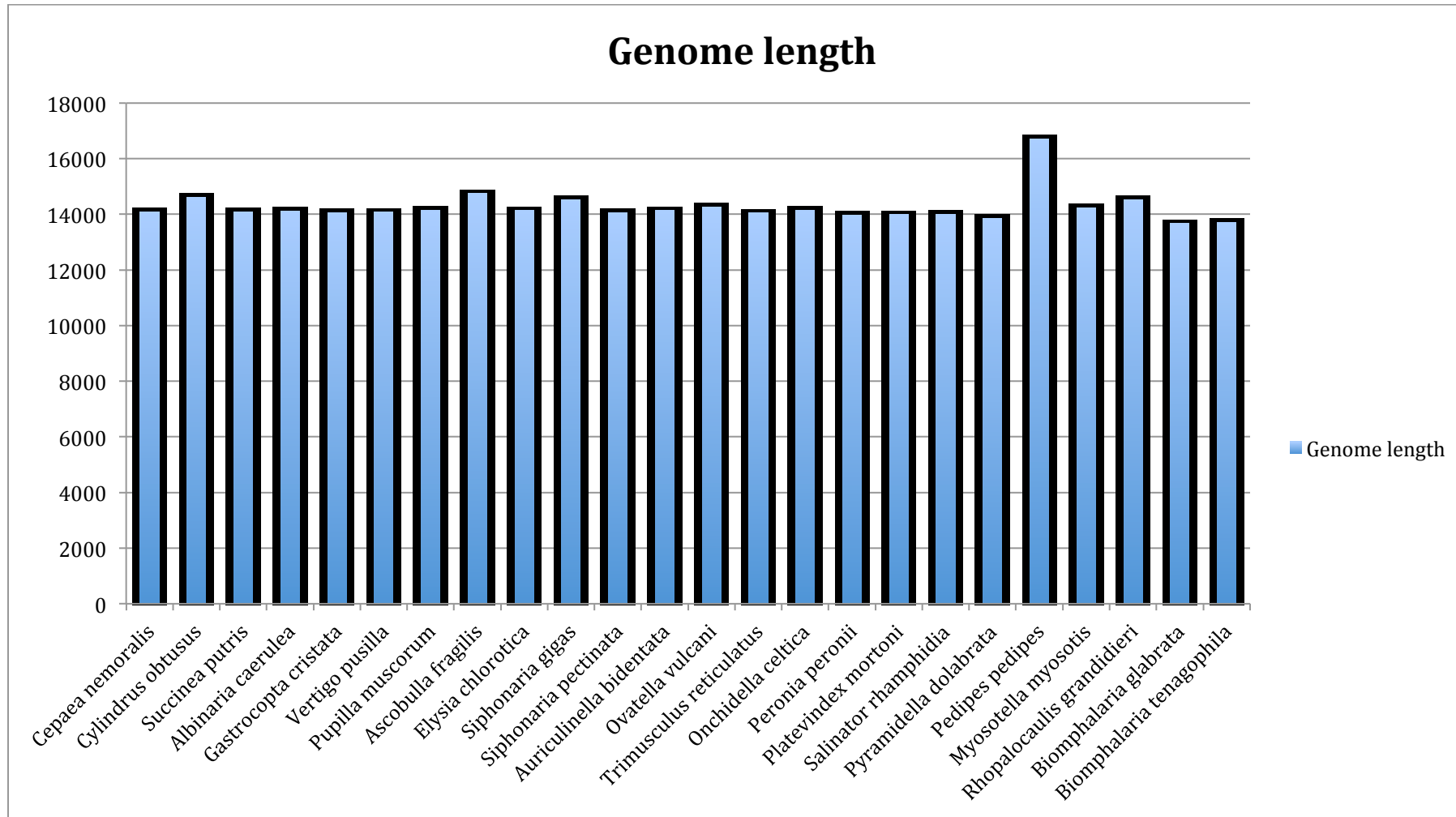


Figure 5: Nucleotide density plot for mt genome of *G. cristata*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*.

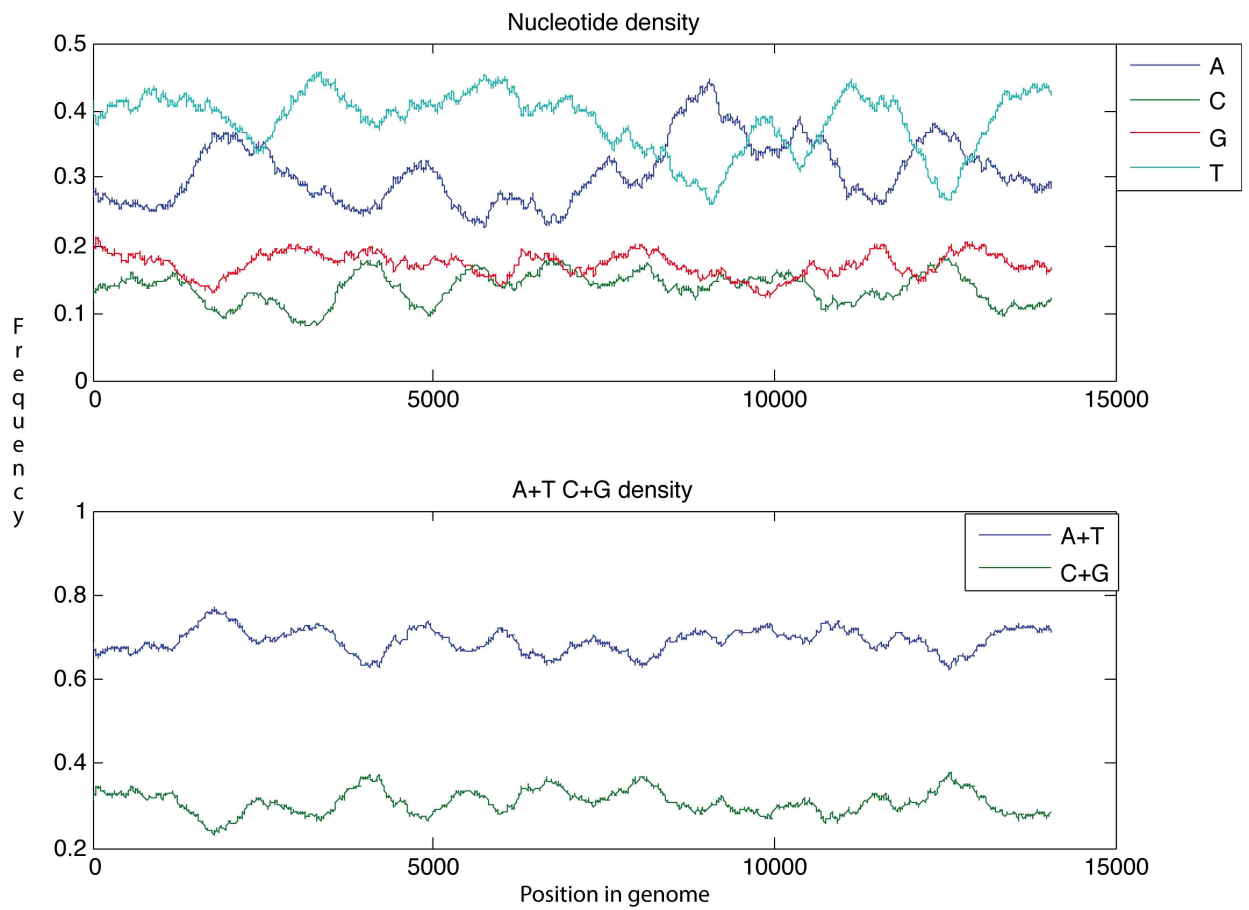


Figure 6: Nucleotide density plot for mt genome of *P. muscorum*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*.

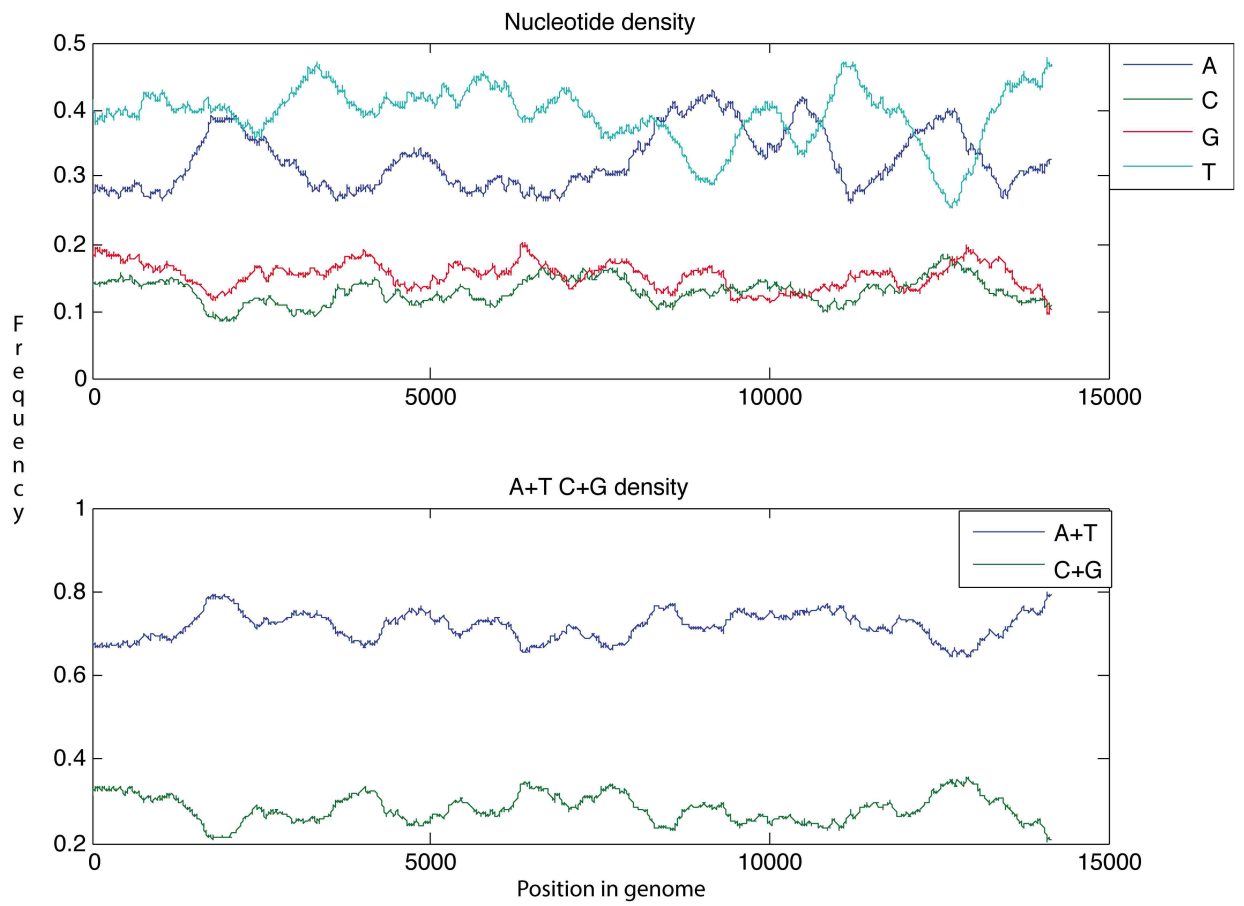


Figure 7: Nucleotide density plot for mt genome of *V. pusilla*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*.

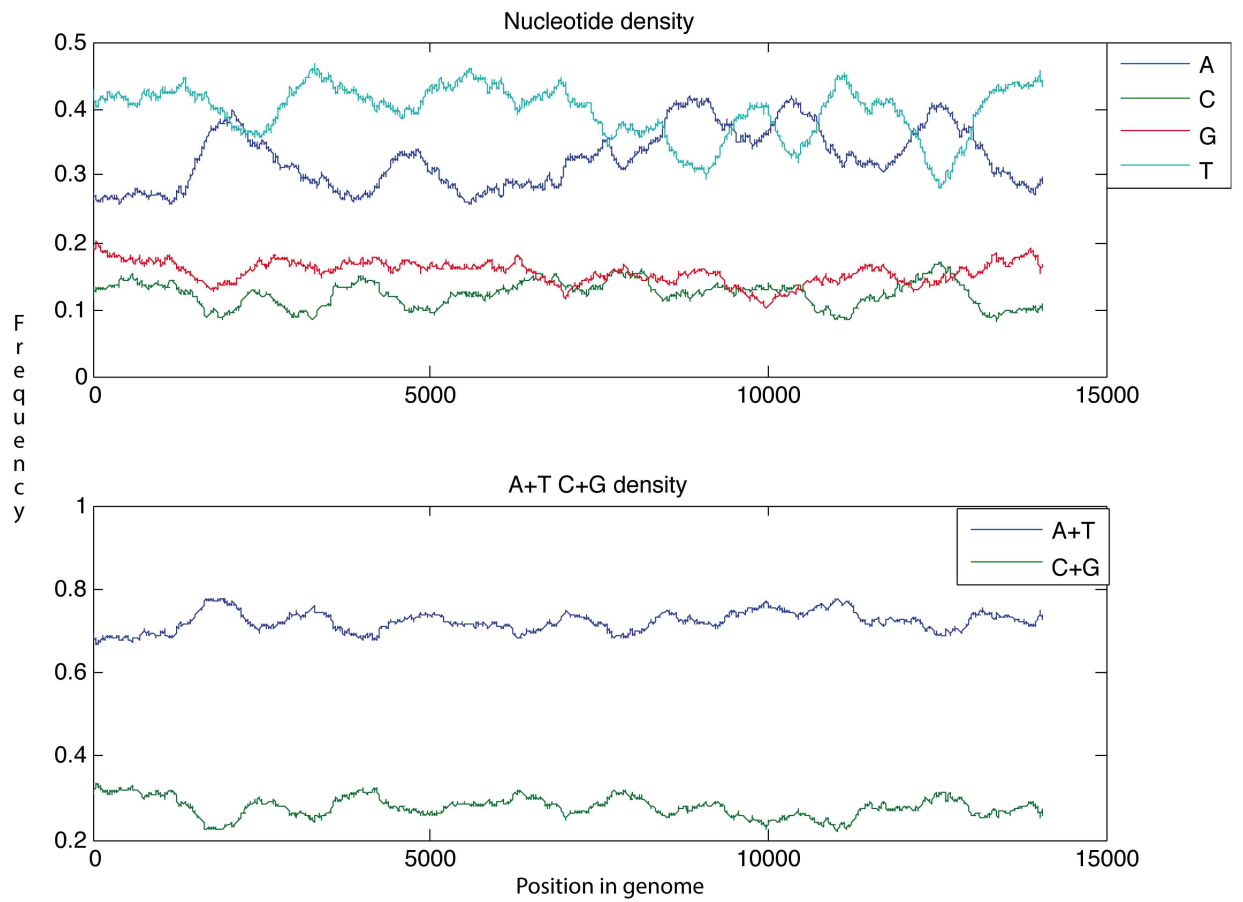


Figure 8: Predicted tRNA structures for *G. cristata*. non-standard tRNA are indicated with a black star.

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.

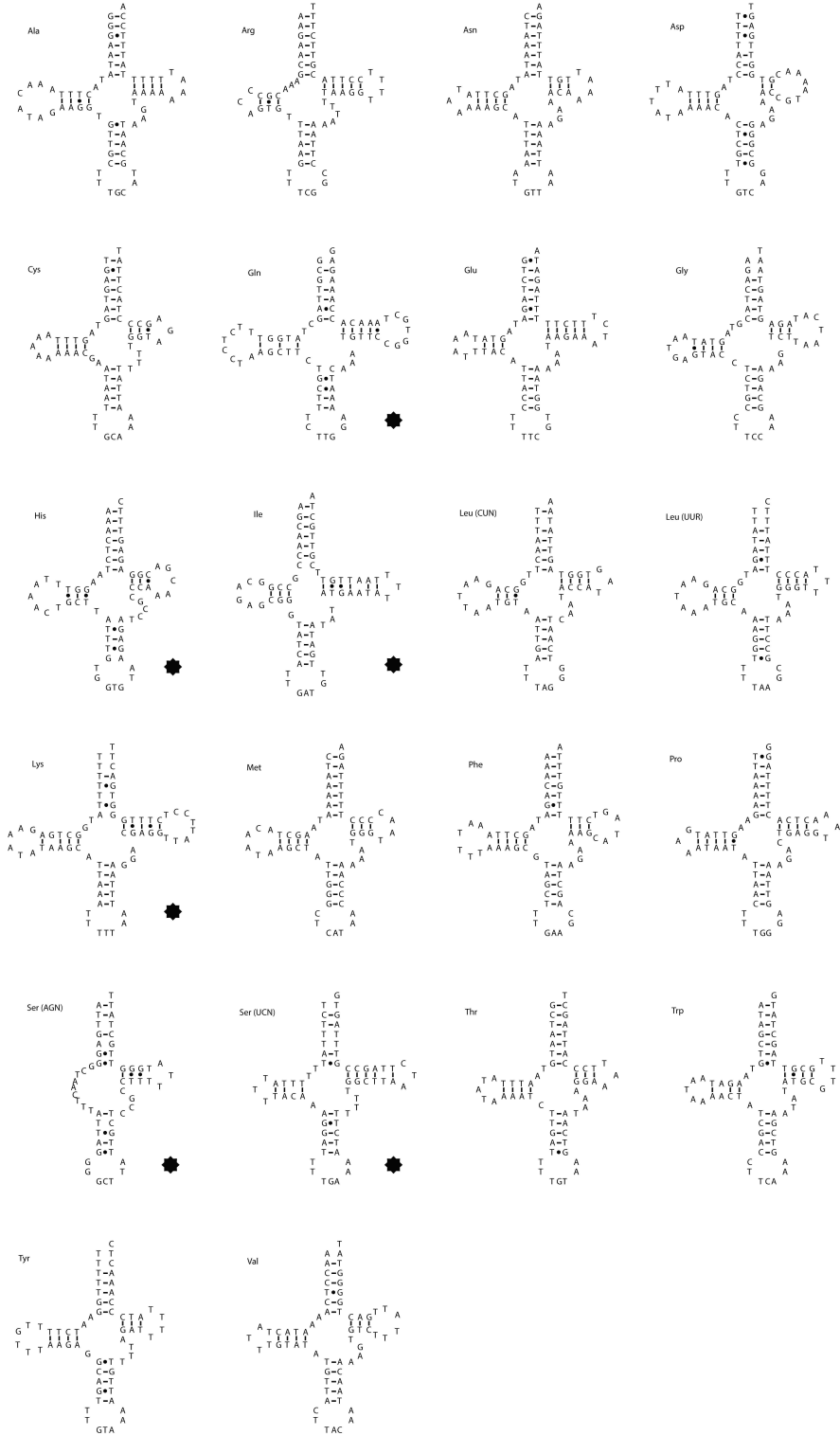


Figure 9: Predicted tRNA structures for *P. muscorum*. Non-standard tRNA are indicated with a black star. The number of missing bonds, missing arms and non-standard spacing determine which tRNA are considered non-standard. tRNA are in 5' -> 3', left to right and are based on DNA sequence.

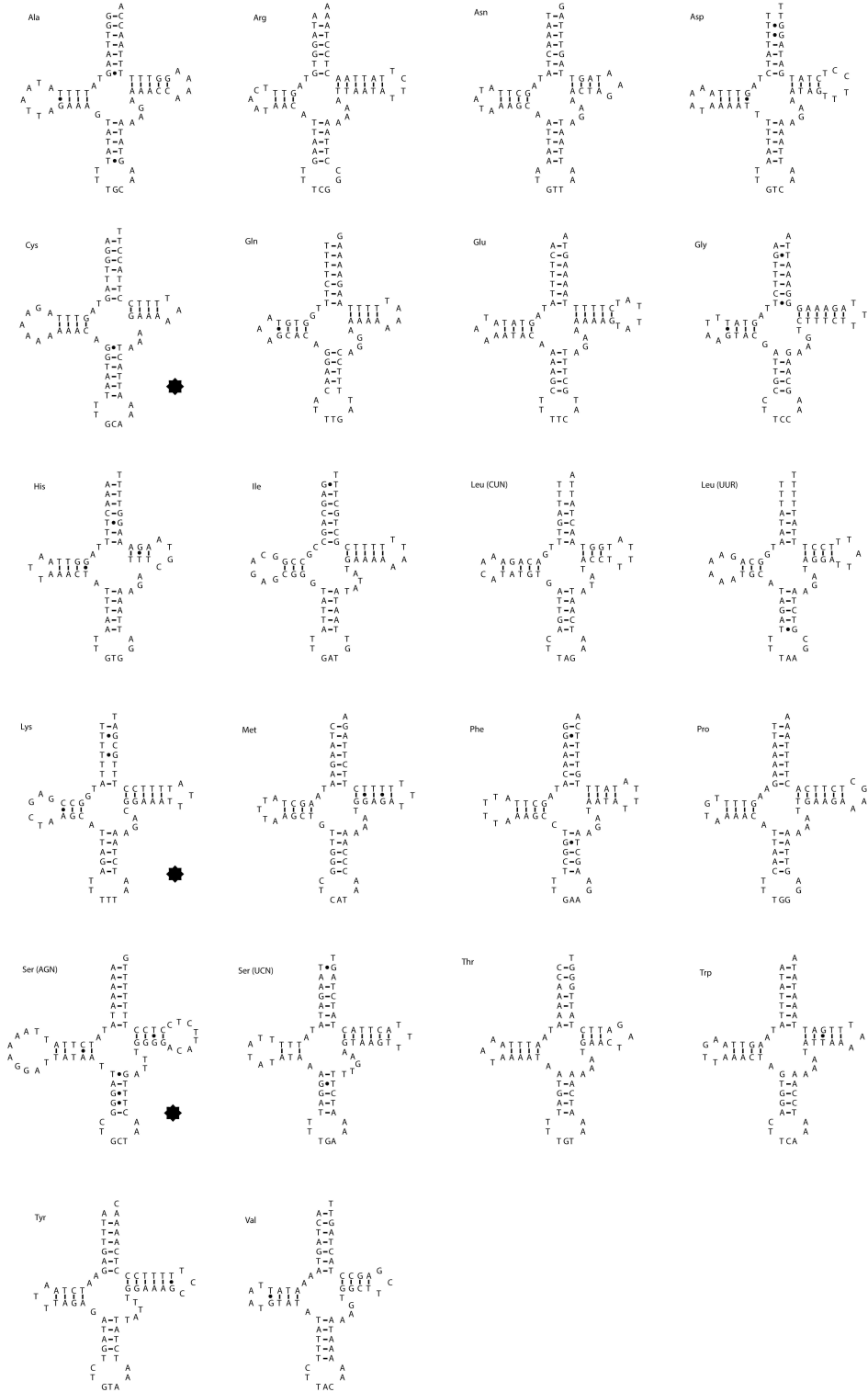


Figure 10: Predicted tRNA structures for *V. pusilla*. Non-standard tRNA are indicated with a black star.

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.

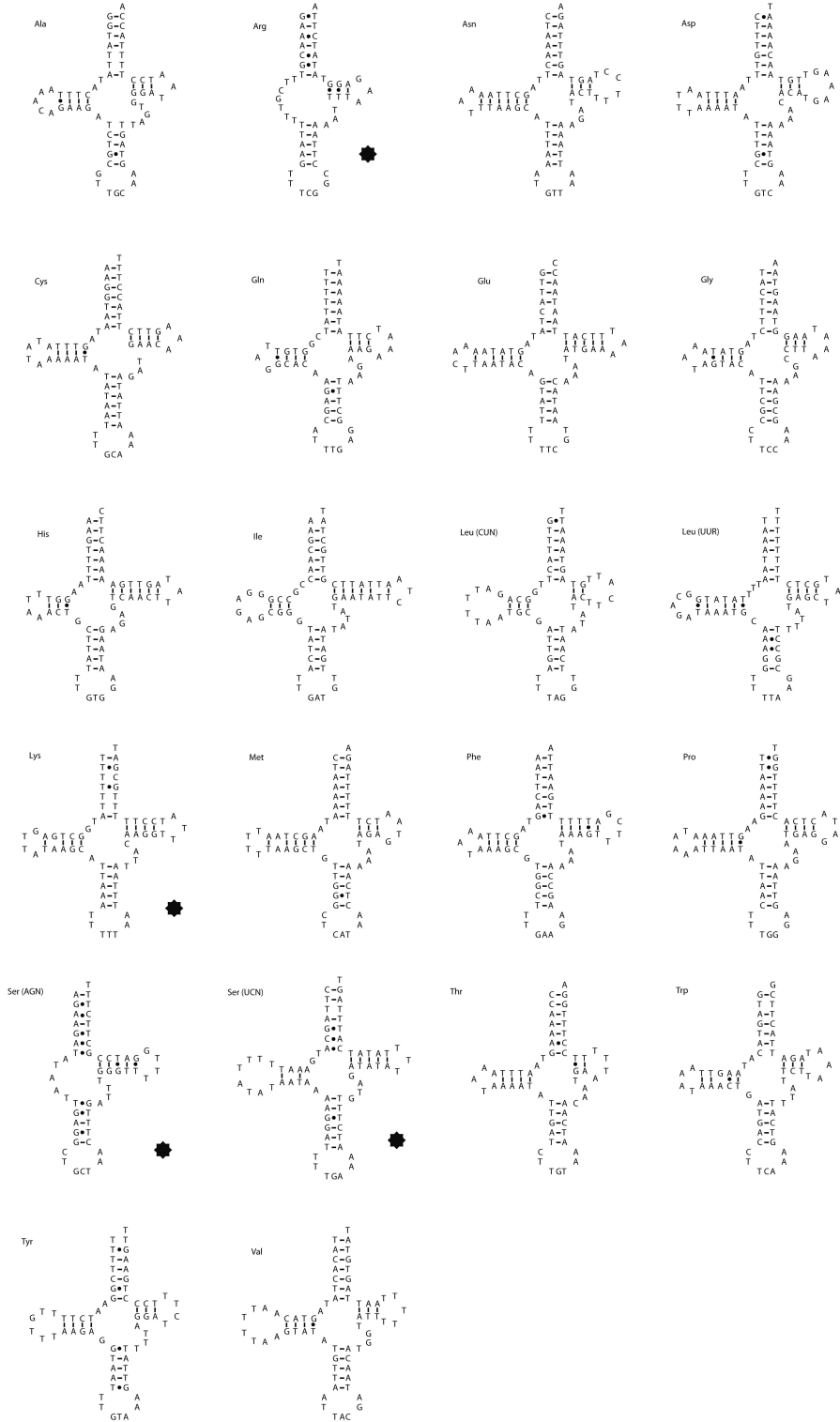


Figure 11: Nucleotide density plot for mt genome of *A. caerulea*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*.

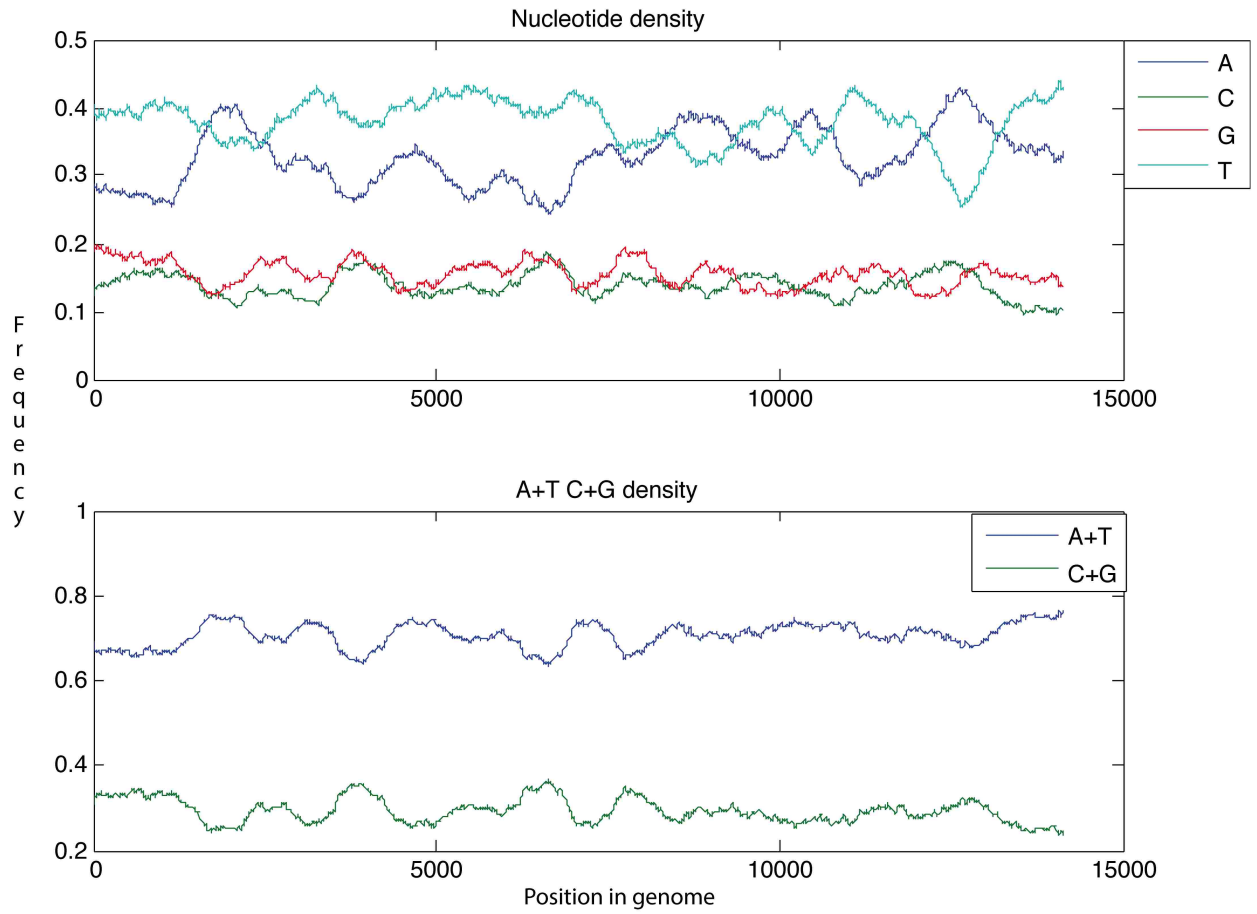


Figure 12: Nucleotide density plot for mt genome of *C. nemoralis*. Matlab automatically determined the bin size. Position 1 is at the beginning of *cox1*.

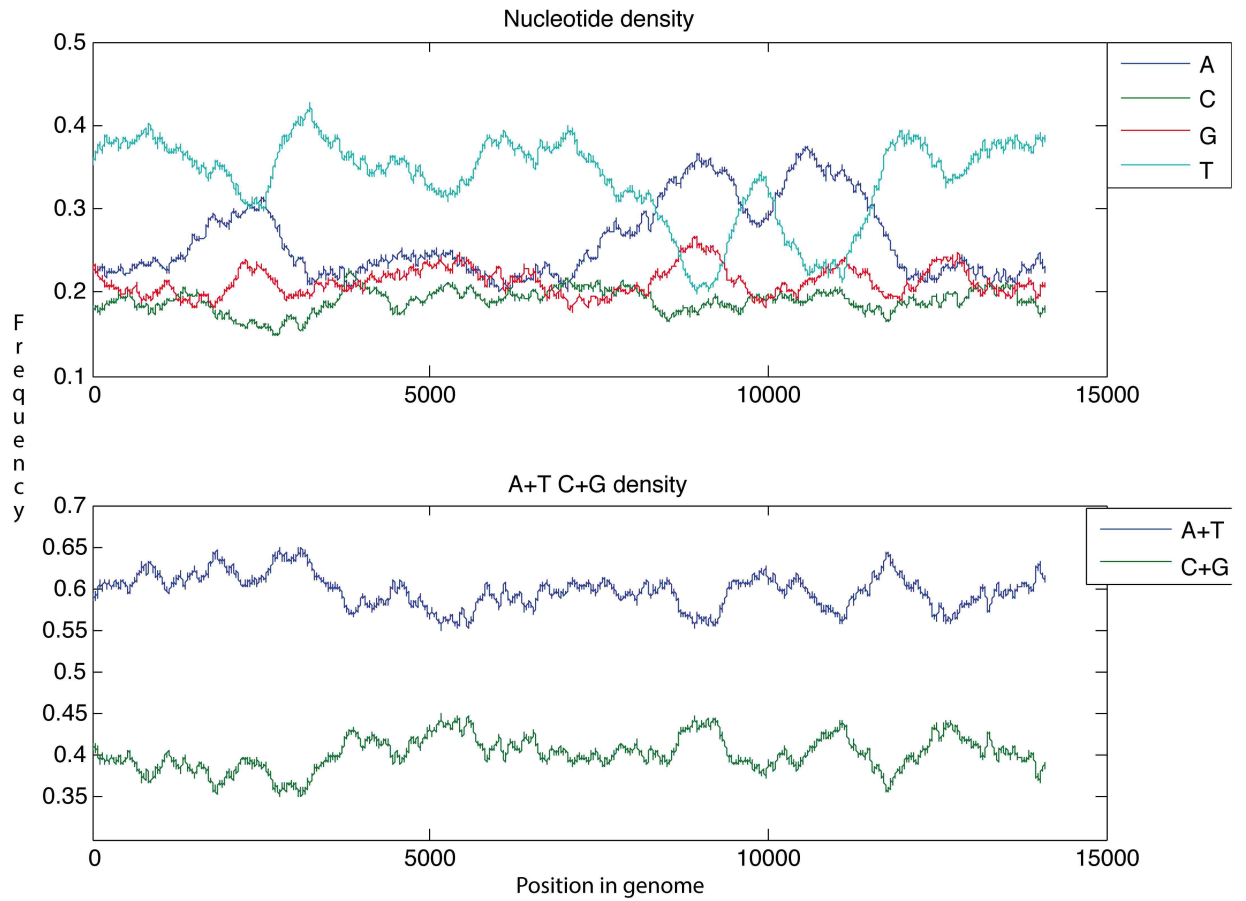


Figure 13: A+T and G+C percentages by mt genome.

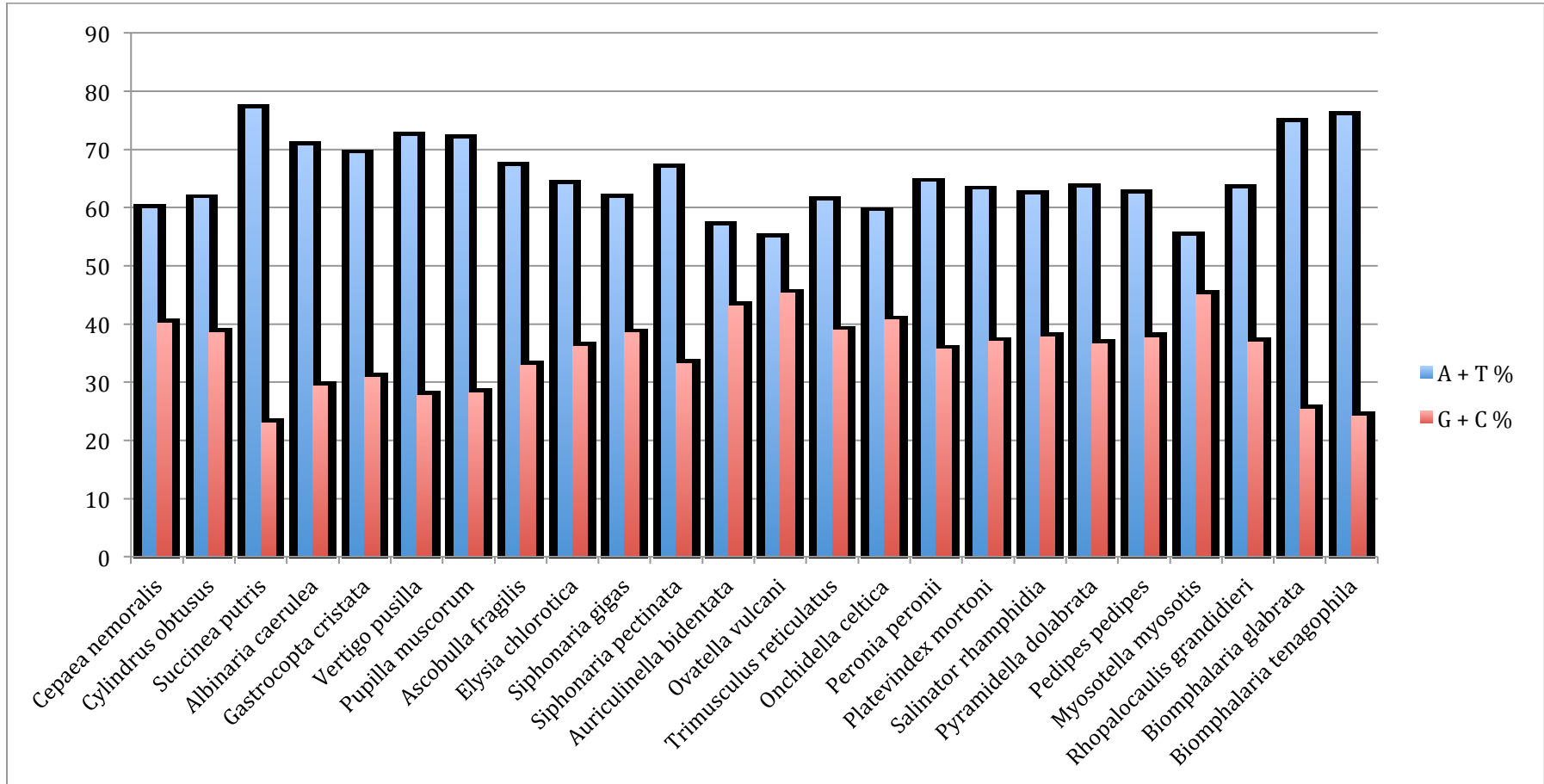


Figure 14: Gene order comparison across data set. Green boxes are tRNA, red boxes are rRNA and blue boxes are protein coding genes. Underlined genes are in the opposite orientation to non underlined genes.

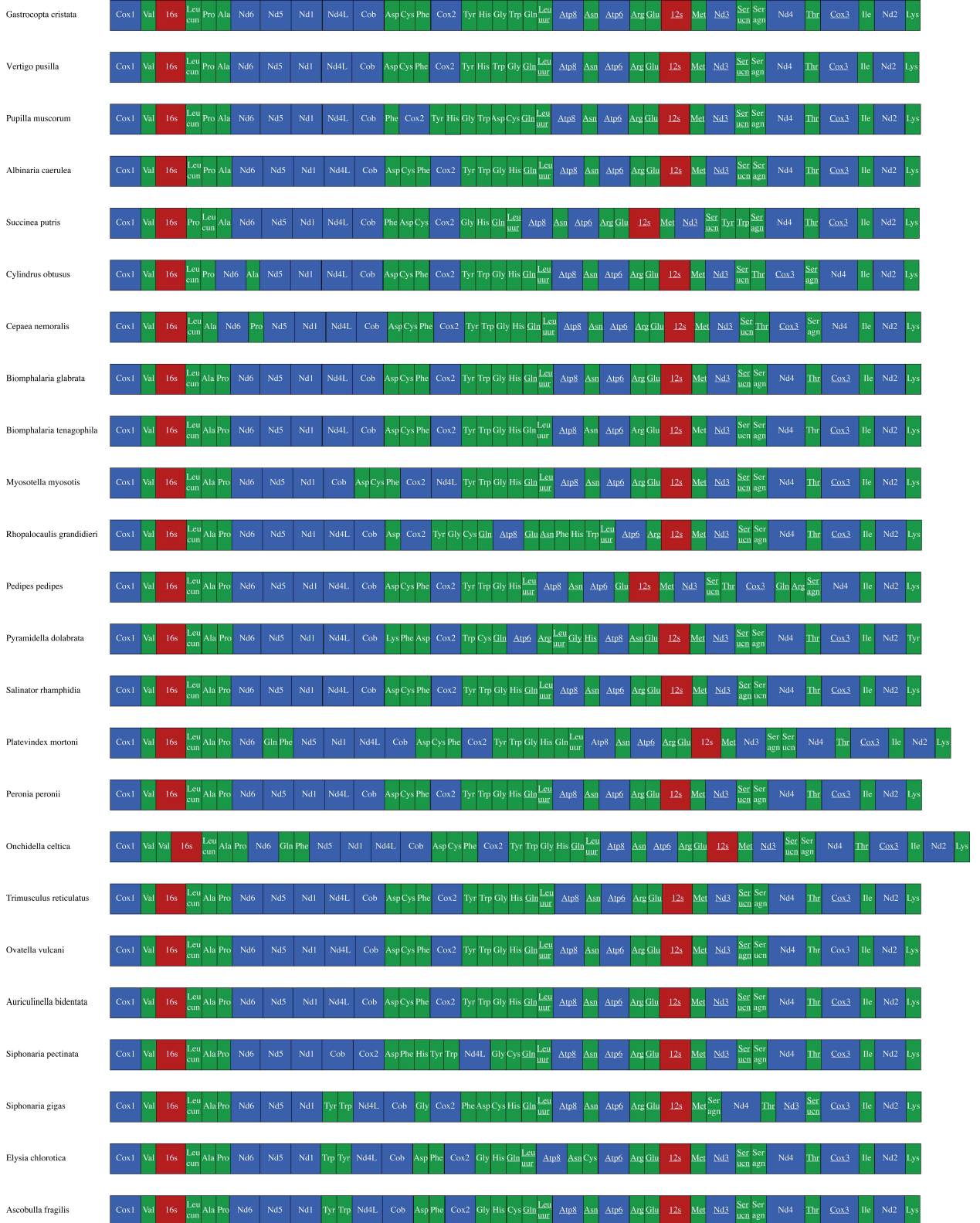


Figure 15: G/C skew comparison of *G. cristata*, *P. muscorum* and *V. pusilla*. Made using a coding sequence alignment. Upper half is ratio of G and lower half is ratio of C. Positive peaks mean G is more prevalent, negative peaks mean C is more prevalent. Bin size was 210 nucleotides, advancing 3 nucleotides at a time.

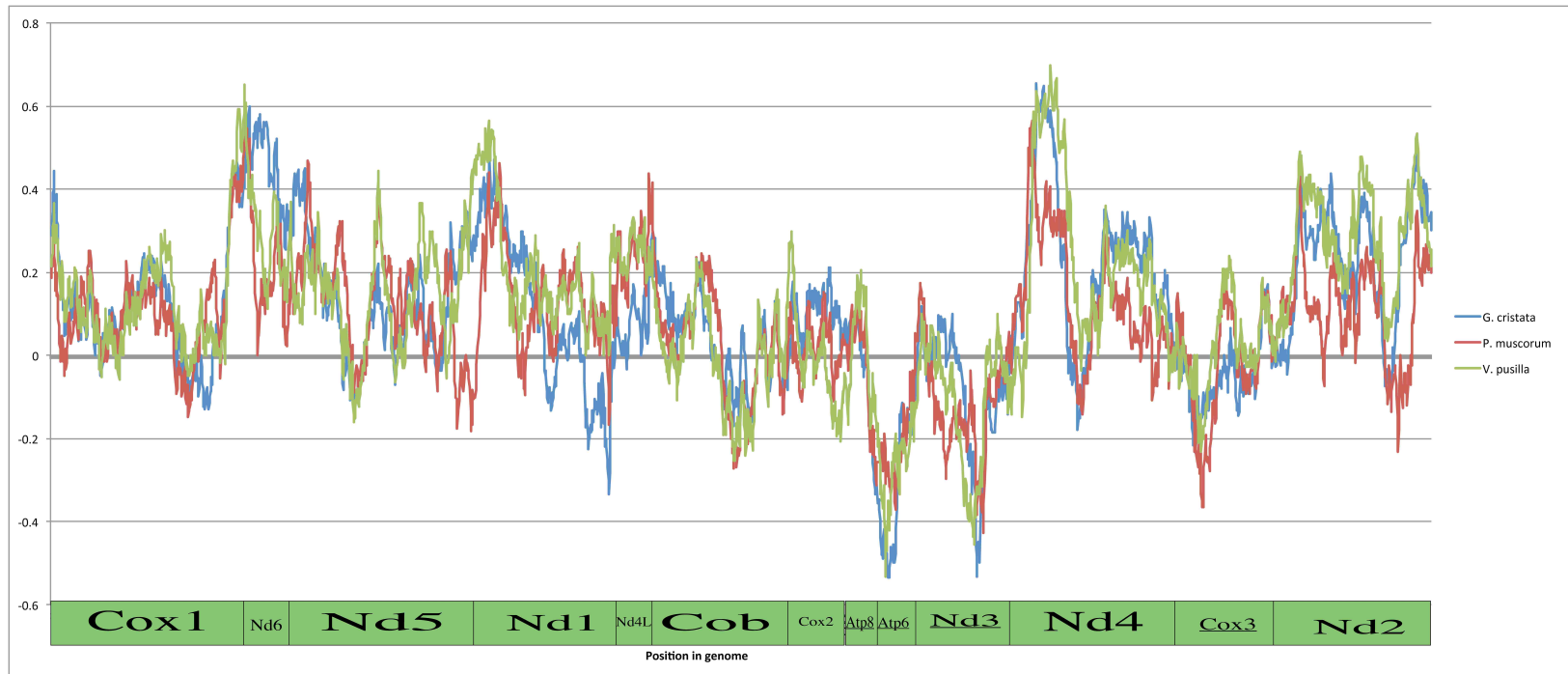


Figure 16: Codon usage pie chart for mt genome of *G. cristata* by amino acid. The two serine and two leucine tRNA's are grouped together under serine and leucine respectively.

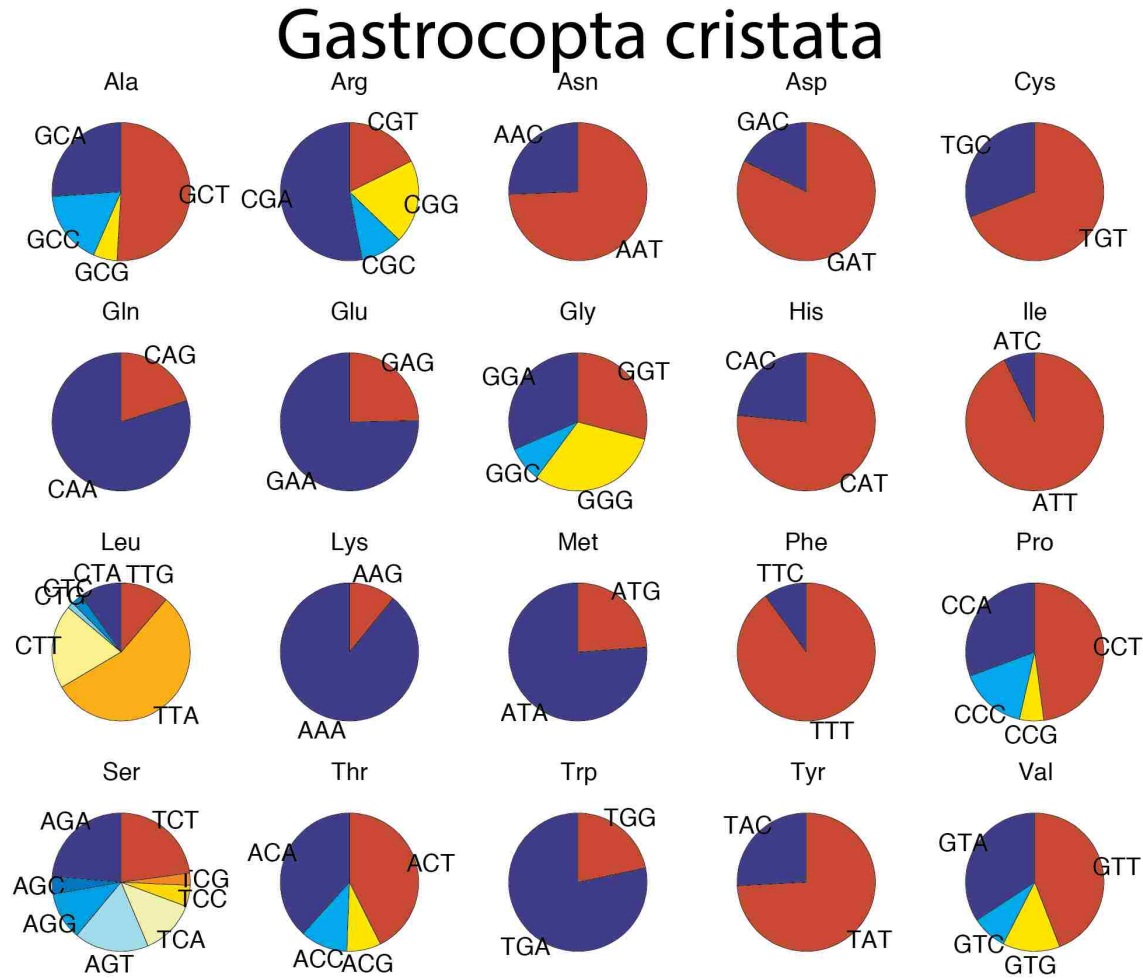


Figure 17: Codon usage pie chart for mt genome of *P. muscorum* by amino acid. The two serine and two leucine tRNA's are grouped together under serine and leucine respectively.

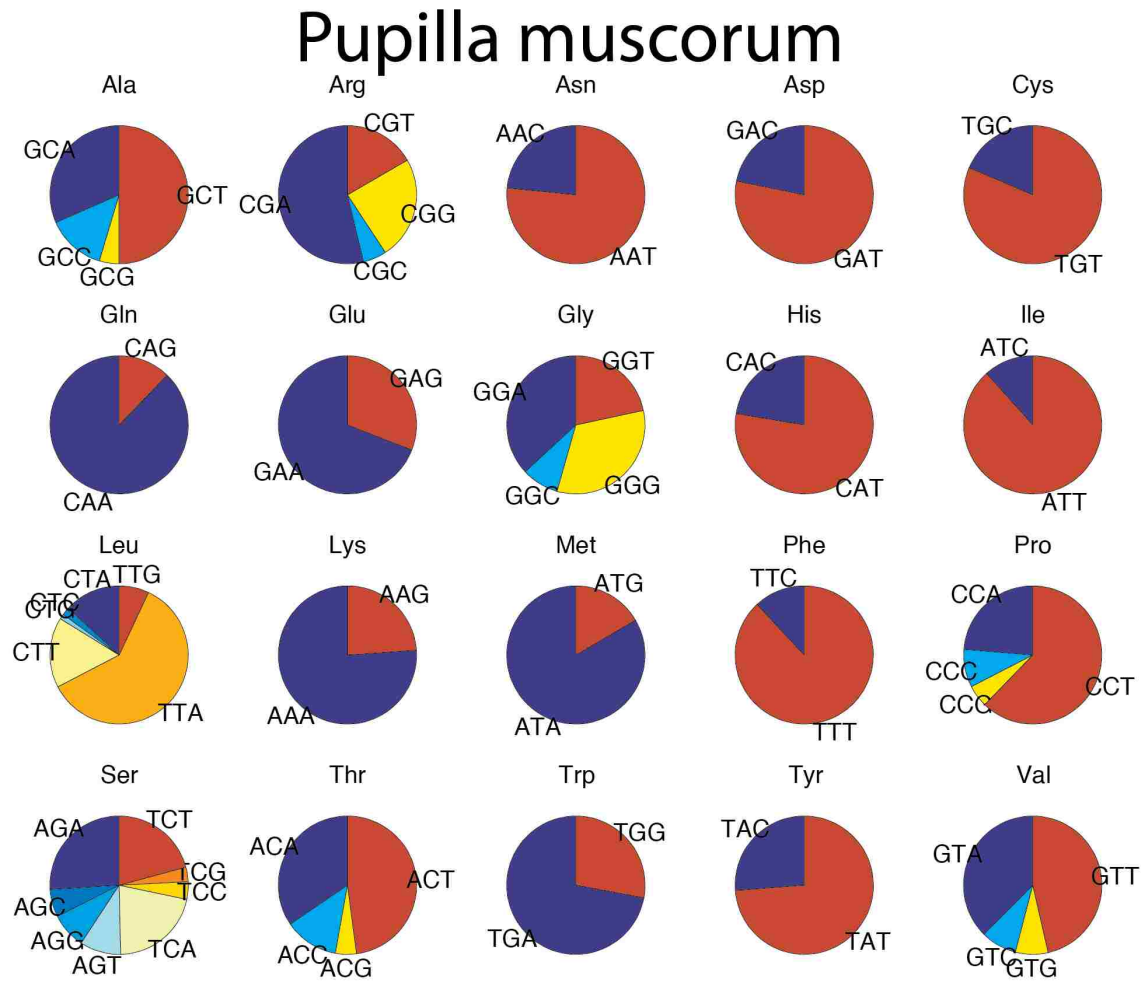


Figure 18: Codon usage pie chart for mt genome of *V. pusilla* by amino acid. The two serine and two leucine tRNA's are grouped together under serine and leucine respectively.

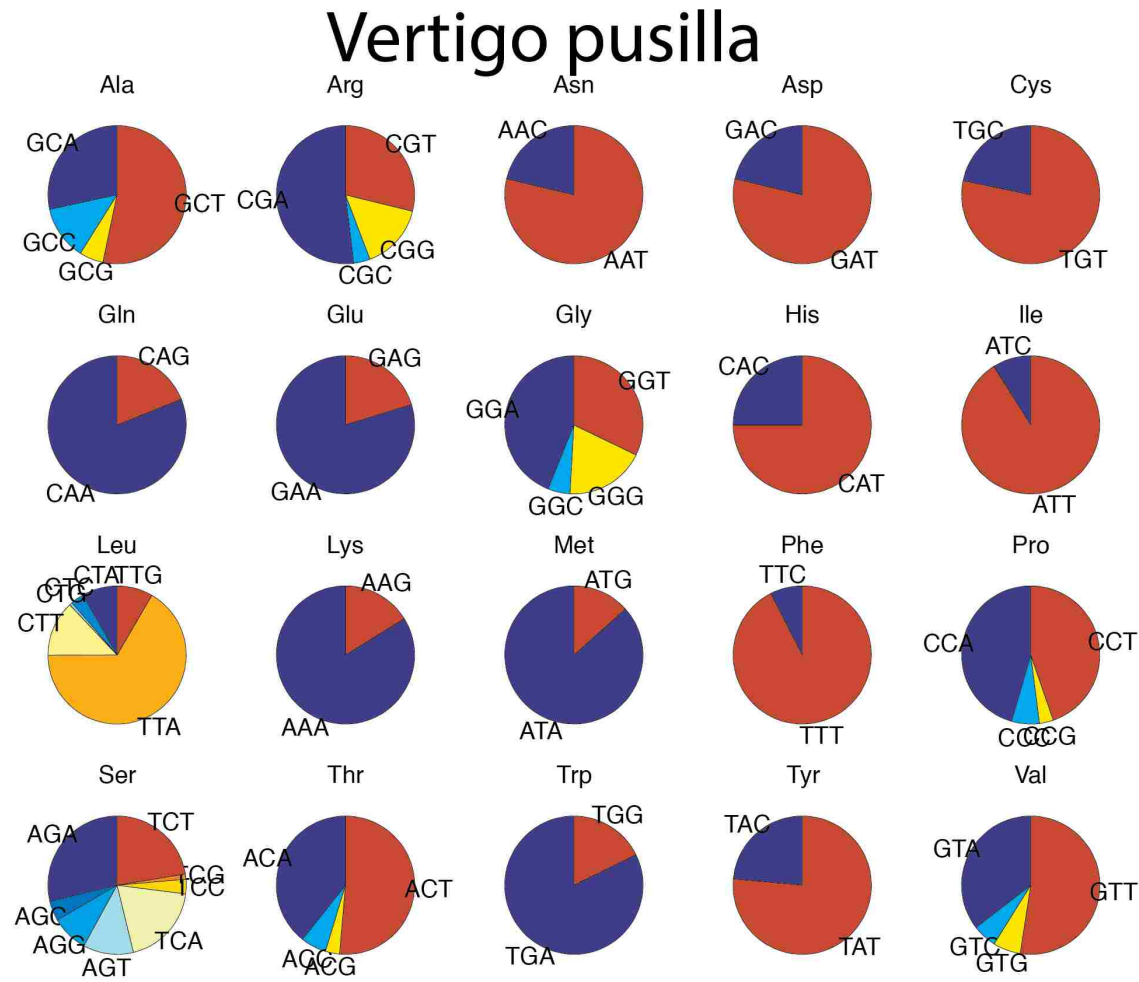


Figure 19: Codon usage pie chart for mt genome of *A. caerulea* by amino acid. The two serine and two leucine tRNA's are grouped together under serine and leucine respectively.

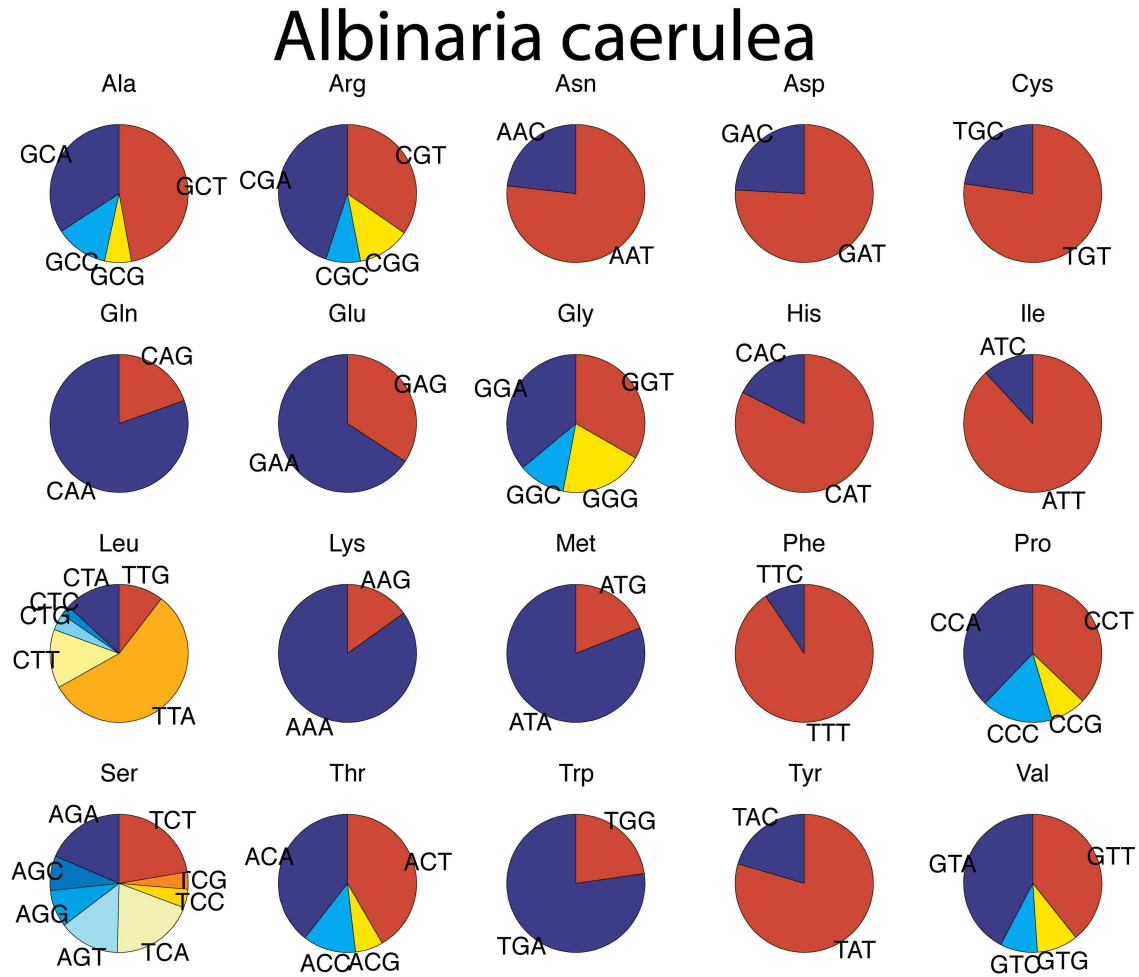


Figure 20: Codon usage pie chart for mt genome of *C. nemoralis* by amino acid. The two serine and two leucine tRNA's are grouped together under serine and leucine respectively.

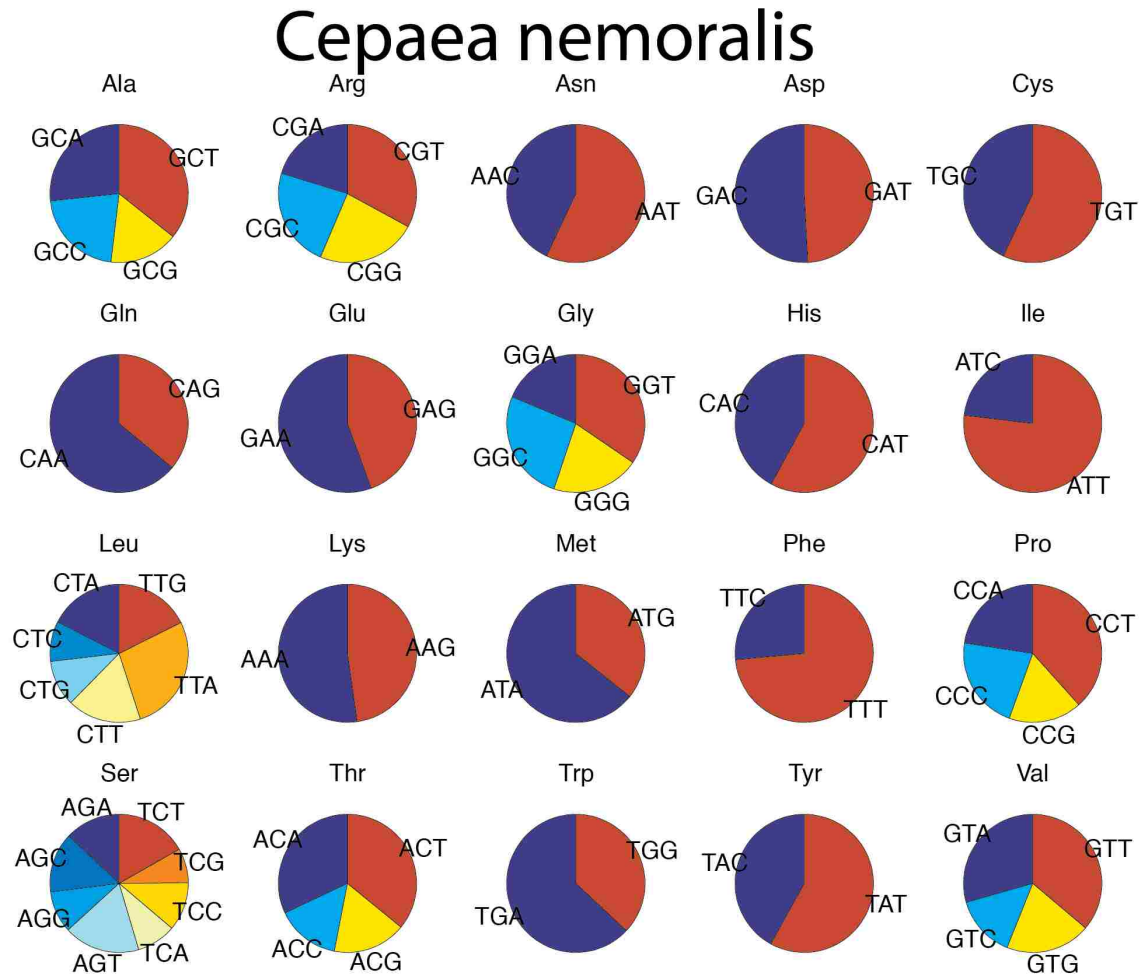


Figure 21: Bayesian phylogenetic tree for complete data set. Node labels are the posterior probabilities.

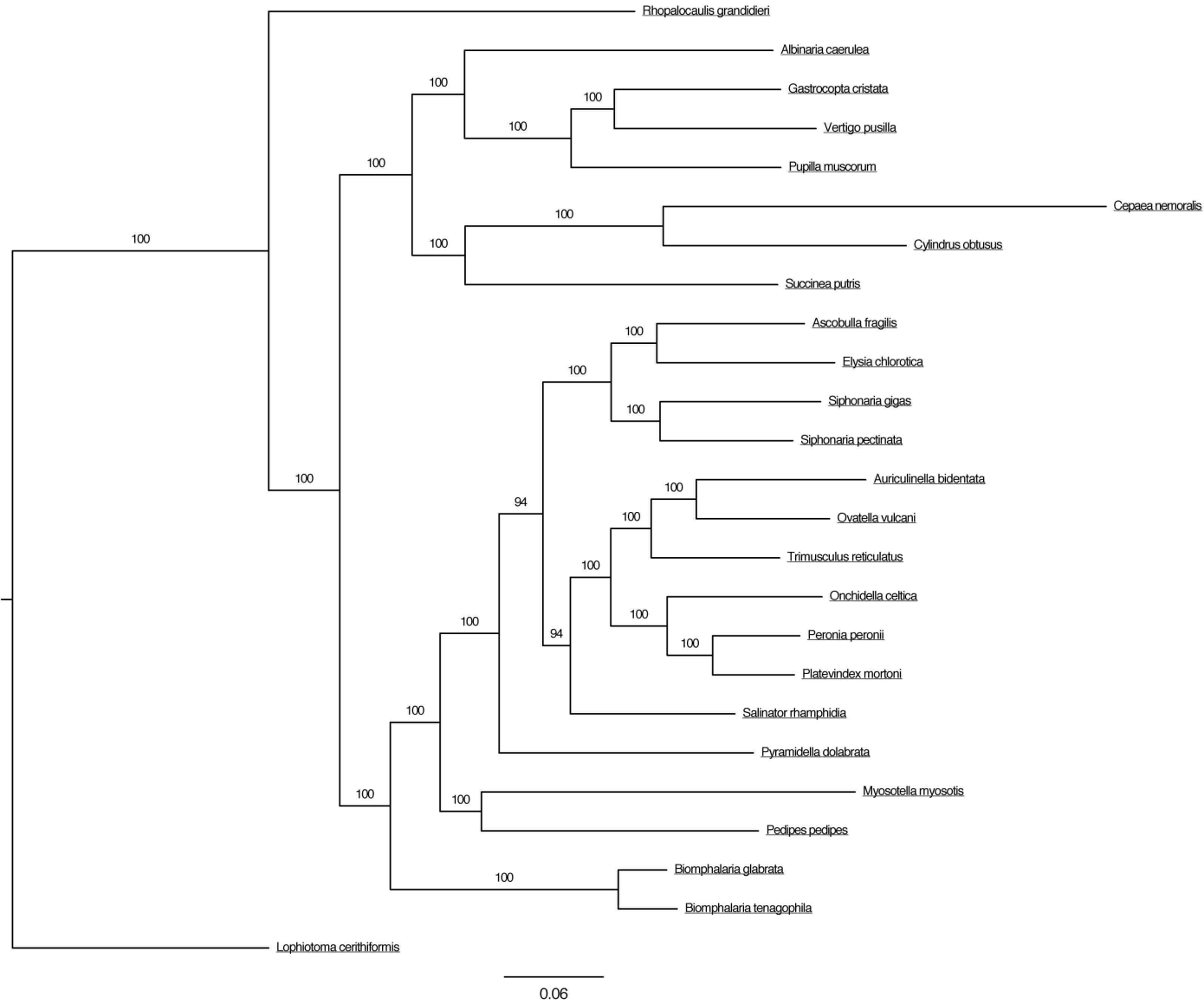


Figure 22: Maximum likelihood tree for complete data set. Node labels are bootstrap values.

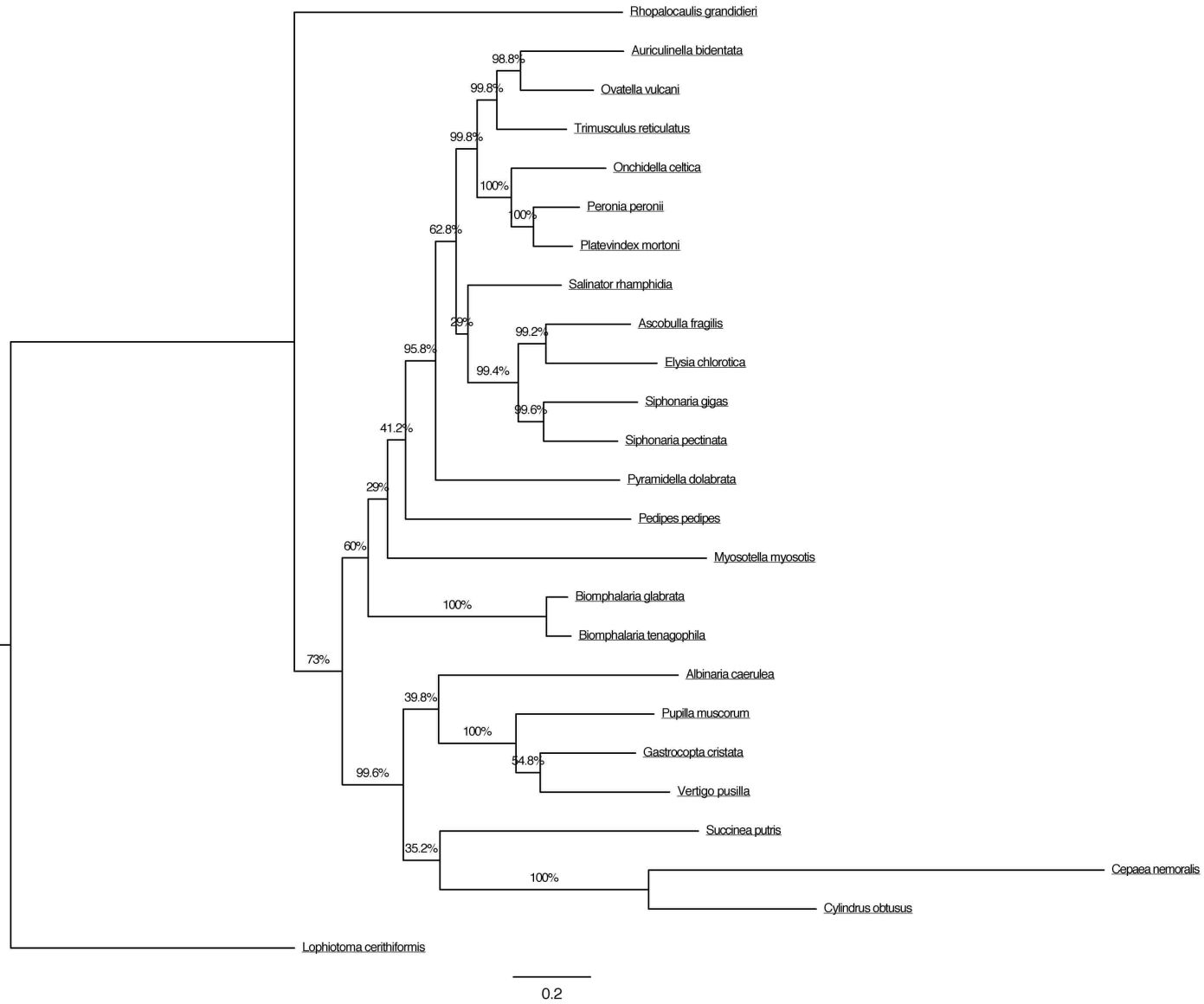


Figure 23: Maximum parsimony tree for complete data set. Node labels are bootstrap values.

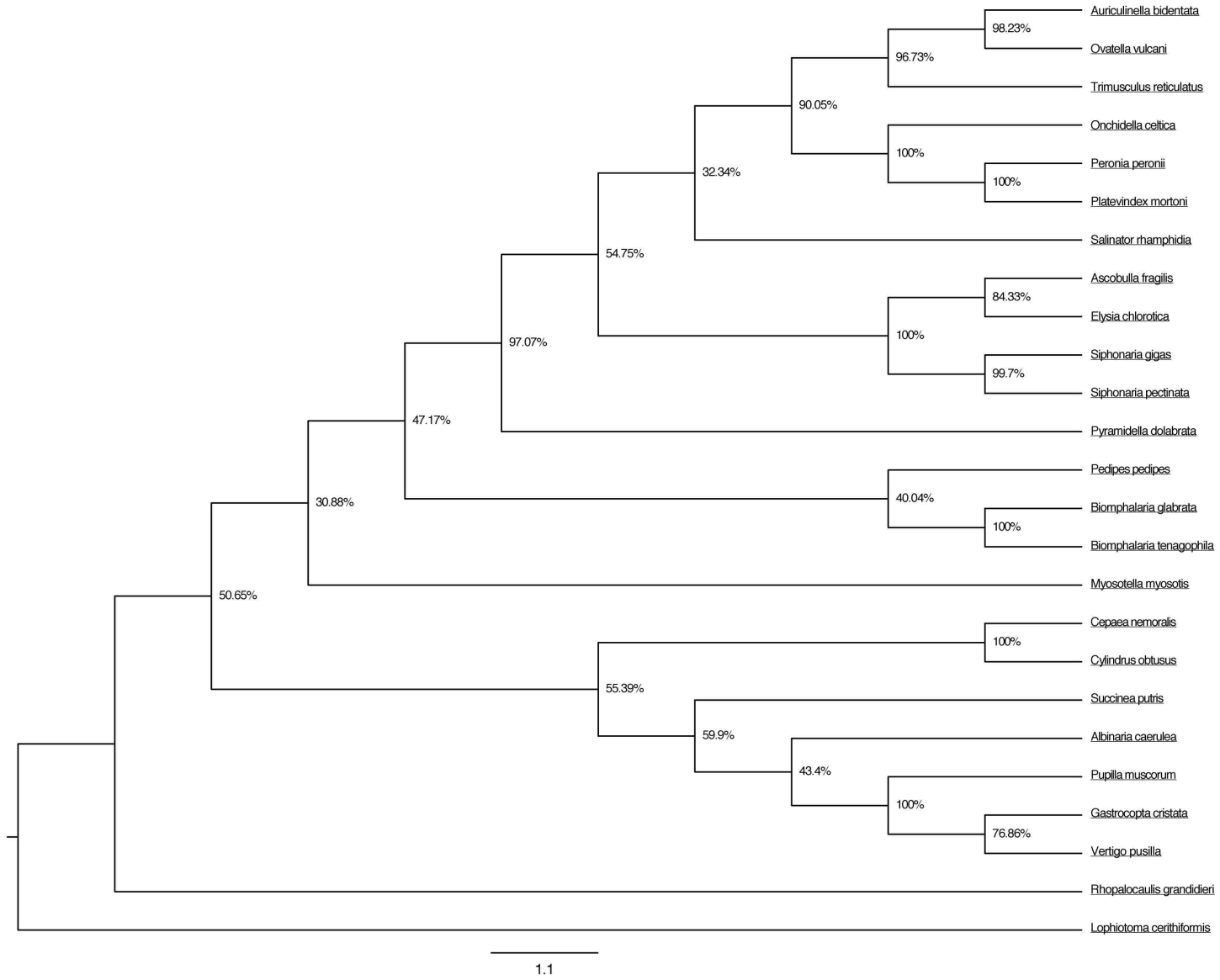


Figure 24: Nearest neighbor joining tree for complete data set. Node labels are bootstrap values.

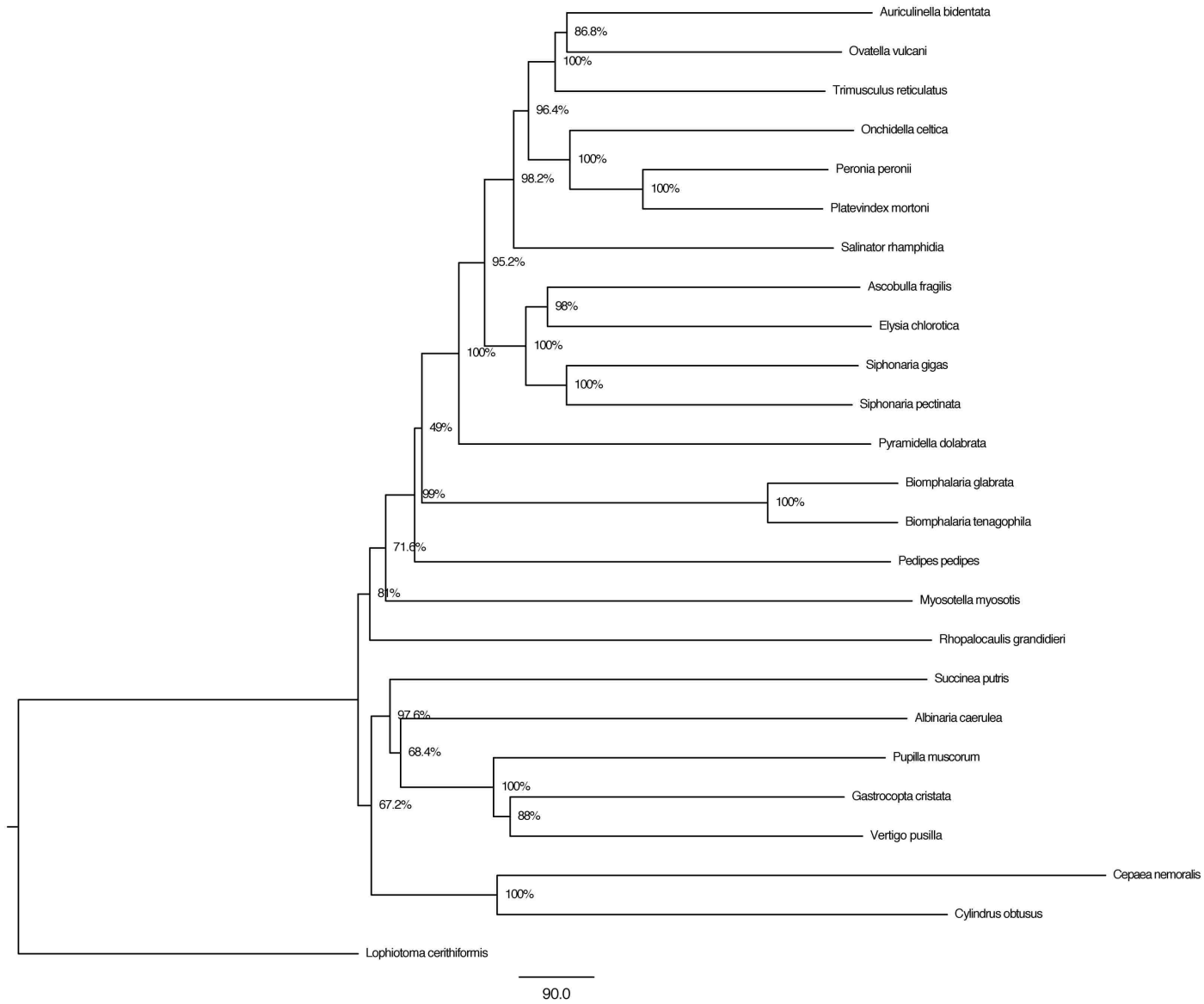


Table 1: List of species used for this project and their associated Genbank accession numbers.

Species name	Genbank accession number
<i>Albinaria caerulea</i>	NC_001761.1
<i>Ascobulla fragilis</i>	NC_012428.1
<i>Auriculinella bidentata</i>	NC_016168.1
<i>Biomphalaria glabrata</i>	NC_005439.1
<i>Biomphalaria tenagophila</i>	EF433576.1
<i>Cepaea nemoralis</i>	NC_001816.1
<i>Cylindrus obtusus</i>	NC_017872.1
<i>Elysia chlorotica</i>	EU599581.1
<i>Gastrocopta cristata</i>	KC185403
<i>Myosotella myosotis</i>	NC_012434.1
<i>Onchidella celtica</i>	AY345048.2
<i>Ovatella vulcani</i>	NC_016175.1
<i>Pedipes pedipes</i>	NC_016179.1
<i>Peronia peronii</i>	NC_016181.1
<i>Platevindex mortoni</i>	GU475132.1
<i>Pupilla muscorum</i>	KC185404
<i>Pyramidella dolabrata</i>	NC_012435.1
<i>Rhopalocaulis grandidieri</i>	NC_016183.1
<i>Salinator rhamphidia</i>	NC_016185.1
<i>Siphonaria gigas</i>	NC_016188.1
<i>Siphonaria pectinata</i>	NC_012383.1
<i>Succinea putris</i>	JN627206.1
<i>Trimusculus reticulatus</i>	NC_016193.1
<i>Vertigo pusilla</i>	KC185405
<i>Lophiotoma cerithiformis</i>	NC_008098.1

Table 2: List of all start codons, by gene, for *G. cristata*, *P. muscorum* and *V. pusilla*.

Yellow highlighted codons are alternative start codons.

	<i>G. cristata</i>	<i>P. muscorum</i>	<i>V. pusilla</i>
Atp8	ATG	ATG	ATG
Atp6	ATG	TTG	ATG
Cox1	TTG	TTG	TTG
Cox2	ATG	ATG	ATG
Cox3	ATG	ATG	ATG
Cytb	TTG	TTG	TTG
Nd1	ATG	ATG	GTG
Nd2	ATA	GTG	ATG
Nd3	GTG	TTG	TTG
Nd4	ATG	ATA	TTG
Nd4l	ATG	ATA	ATC
Nd5	ATA	ATG	ATA
Nd6	ATG	TTG	ATG

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 comparison was conducted using Mega 5.0 and coding only sequence with the divergent regions concatenated out.

Unique differences in <i>Gastrocopta cristata</i>	249
Unique differences in <i>Albinaria caerulea</i>	291
Unique differences in <i>Succinea putris</i> (outgroup)	373
p-value	0.07
value needed for significance using the sequential Bonferroni correction	0.0125
Unique differences in <i>Pupilla muscorum</i>	285
Unique differences in <i>Albinaria caerulea</i>	294
Unique differences in <i>Succinea putris</i> (outgroup)	358
p-value	0.7
value needed for significance using the sequential Bonferroni correction	0.05
Unique differences in <i>Vertigo pusilla</i>	261
Unique differences in <i>Albinaria caerulea</i>	275
Unique differences in <i>Succinea putris</i> (outgroup)	351
p-value	0.54
value needed for significance using the sequential Bonferroni correction	0.025
Unique differences in <i>Gastrocopta cristata</i>	187
Unique differences in <i>Pupilla muscorum</i>	237
Unique differences in <i>Albinaria caerulea</i> (outgroup)	501
p-value	0.015
value needed for significance using the sequential Bonferroni correction	0.01
Unique differences in <i>Pupilla muscorum</i>	224
Unique differences in <i>Vertigo pusilla</i>	209
Unique differences in <i>Albinaria caerulea</i> (outgroup)	469
p-value	0.47
value needed for significance using the sequential Bonferroni correction	0.0167
Unique differences in <i>Gastrocopta cristata</i>	190
Unique differences in <i>Vertigo pusilla</i>	244
Unique differences in <i>Pupilla muscorum</i> (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.

	G. cristata vs. V. pusilla			G. cristata vs. P. muscorum			P. muscorum vs. V. pusilla		
	D_n	D_s	D_n/D_s	D_n	D_s	D_n/D_s	D_n	D_s	D_n/D_s
Atp6	0.221	1.370	0.161	0.316	2.185	0.144	0.322	1.446	0.223
Atp8	0.699	0.841	0.831	0.810	1.292	0.627	0.697	0.715	0.975
Cytb	0.170	1.477	0.115	0.194	2.172	0.089	0.192	1.247	0.154
Cox1	0.049	1.234	0.040	0.050	1.205	0.042	0.055	1.182	0.047
Cox2	0.125	1.287	0.097	0.141	1.285	0.109	0.160	2.161	0.074
Cox3	0.161	1.685	0.095	0.144	2.249	0.064	0.181	1.038	0.174
Nd1	0.249	1.017	0.245	0.272	1.600	0.170	0.288	1.500	0.192
Nd2	0.450	1.202	0.374	0.402	1.163	0.346	0.463	1.185	0.390
Nd3	0.360	2.110	0.170	0.318	0.840	0.378	0.384	0.990	0.388
Nd4	0.287	1.247	0.231	0.328	1.167	0.281	0.345	0.950	0.363
Nd4l	0.444	1.142	0.389	0.522	1.084	0.481	0.668	1.138	0.587
Nd5	0.359	1.201	0.299	0.353	1.566	0.225	0.369	1.247	0.296
Nd6	0.334	0.840	0.398	0.413	1.757	0.235	0.380	0.715	0.532

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';

%how 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  $D_n/D_s$  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)
var ds1(:,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;
    %=====
    fasta1=fastaentry;
    fasta2=fastaentry;
    [header, genome]= fastaread(name,'Blockread', [fasta1
    fasta2], 'IgnoreGaps', 'True');
    len=length(genome);

    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));

end
```

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 :: Sequencher 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"
CDS 2..1531
/gene="COX1"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit I"
/translation="MRWLYSTNHKDIGTLYMIFGVWCGMVGTGLSLLIRLELGTSGVL
MDDHFFNVVVTAAHAFVMIFFMVMPIIMIGGFGNWMVPLLIGAPDMSFPRMNNMSFWLLP
PSFILLISSMVEGGAGTGWTVYPPLSGITGHSGASVDLAI FSLHLAGMSSILGAINF
ITTIFNM RAPGMTFERLSLFVWSILVTVFLLLLSLPVLAGAITMLLTD RNFNTSFFDP
AGGGDPILYQH LFWFFGHPEVYILILPGFGIISHILGNASVKQPFGLGMIYAMISIG
ILGFIVWAHMFVTGMDVDTRAYFTAATMIIAVPTGIKVFVSWLMTLYGMNSKFDASMY
WVLGFIFLFTLGGTLGIVLSNSSLDIVLHDTYYVVAHFHYVLSMGAVFAIFAGFYWF
PVM TGLV LQERLAKAQFIIMFIAVNLTFPPQHFLGLSGMPRRYS DYPDSYFKWNQISS
FGSLMSIFAVLLFVLIVWEAFLSQRSVMFSSAAPYSREWTE NNFPDFHG NLI PSISV
L"
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"
CDS 2864..3316
/gene="ND6"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 6"
/translation="MIVLFGLGMFLVFSFCFLRSPVSYGGALIALSTIGALLSLGCS
TWYGYLLFLVYVGGLLVFLYIIMLSSNFNLQVSFKLMGLIFLAFVSKLYNFSYPKN
SLGVSLSECEDFSLGLFLGLGGLLLL VFFAIVHIVFLKGPVQVKND"
gene 3318..4969

/gene="ND5"
 CDS 3318..4969
 /gene="ND5"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 5"
 /translation="MSRLPVLLLLSCLLLMSLFFYSVFKSGETYVIELNMLTLSSVSF
 STLFIFDKISLSFGAVVTLISFCVFSFANTYMSEDNFQFRFILILAMFVLSMNILIFS
 GSLFVLLLGWDGLGISSFALIIYYQNKTSLGAGYLTLTNRLGDDVVIICVPPFFLILG
 SFNIFPIMSSYETIIFILAI AALTKSAQYPFSAWLPAAAMAAPTVPVSALVHSSTLVTAG
 VYLIIRLSVNCMSNSTASLLLF CGSVTCVVGGLSALYENDLKKIIAFSTLSQLGLMM
 FSLGLKLPNLALLHL YAHAMFKALLFLSAGLILISGFGS QDLRLLGATLYTIPSVVVF
 FNISSLCLMGIPFLSAFYSKHVIVYEVVAMSETNAISFFLMLIGAILTTIYSIRTVKIL
 SWNSISGLSVFSRPLTYTVPLCILFFTSIHSKGTFSMLDISYLTTFMFTSSYYQLMLH
 FLMALGVMIGLFLS GEKKSHMLSSMFFLWPSNNLTKYFWPVLKSAKMLDYGWVVEPVV
 LLSSSLNKLGNKHLHFLFLNESKYLN SMVRGVFCSLTVFFLCI"
 gene 4961..5866
 /gene="ND1"
 CDS 4961..5866
 /gene="ND1"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 1"
 /translation="MYLFIMKAVILCLCVLLAVAFFTLERKVLGYMQIRKGNPKVGI
 WGIVQPLADALKLFIKEKVLLYMSNNLLFIFPMLGFFVSLSLWNLAPSNNMMVNYMIF
 GLLFFCLTGLNVYTTALAGWSSNSLYAFLGAIRASAQSSISYEVSLVMLLLFPALVSS
 SMCWFSALHTPPSSLLVFPLLAIWFASSLAETNRAPFDFAEGESELVSGFNVEYEGGL
 FALLFLAEYTSIIFMSMATTVWFLSSWFTIDFVLITLMICIWFLFARGVYPRYRYDLL
 MMLCWKSYLPFAISLLLYFCFSIMF"
 gene 5868..6143
 /gene="ND4L"
 CDS 5868..6143
 /gene="ND4L"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 4L"
 /translation="MSQLLFILYGFVILLYSFSSQSKHLISSLIILESLMVISIYYL
 FFSLGSVNSIGLLLLILTFAAAEAAALALSLLISILRVMKNDNMLNLSY"
 gene 6150..7262
 /gene="COB"
 CDS 6150..7262
 /gene="COB"
 /codon_start=1
 /transl_table=5
 /product="cytochrome b"
 /translation="MRKISLLEGLTSLPTPLNISIFWSGGSILGMMLGIQLVTGLFLS
 MHYTSDISISFDSVVHIMRDVPNGWLFRLFHANGASFFFIFLYLHIGRGLYQSYSVQ
 PKTWMVGVITFLISMATAFLGYVLPWGQMSFWGATVITNLLSAIPYWGGLVEVWVWG
 FSVGQATLNRFFSLHFILPFLIALLV MVHLIFLHEKGTNPLGQTHHINKVTFHPYFT
 YKDLVGFICVGLSLVLICAYFPFSLSDPENFIYANPMVTPTHIQPEWYFLFAYAILRS
 IPSKLGVVVALAMSIFILYFLPLSSQFMSVPASFNMIIQQMLFWWL VVVFLLLTWLGAC
 PIEDPYLSLAGPLTHIYFFSYPLL MIVGNFVSILK"
 tRNA 7267..7332
 /product="tRNA-Asp"
 tRNA 7328..7389
 /product="tRNA-Cys"
 tRNA 7390..7456
 /product="tRNA-Phe"
 gene 7457..8154
 /gene="COX2"
 CDS 7457..8154
 /gene="COX2"
 /codon_start=1
 /transl_table=5
 /product="cytochrome c oxidase subunit II"
 /translation="MSYWGSLNLM DPSSPIQIEMMLFHDHAMILLIGIFTLVALLGVK
 LCLNKFSSRA THEAQLLEIFWTHIIPALLLIWALPSLRLLFLLD EQSGNGLILKATGH
 QWYWSYEIPSSGTD SFD SYMVQE QALTNGE FRLLEVDNRPVLPYSNEISVLTSSDVI
 HAWALPSMGVKMDA VPGRLNMMGLYSLTPGVFYGQCSEICGANHSYMPIVVEFVHLND
 FINFFGIFCLSGSL"

tRNA 8121..8183
/product="tRNA-Tyr"
tRNA 8178..8241
/product="tRNA-His"
tRNA 8242..8306
/product="tRNA-Gly"
tRNA 8307..8370
/product="tRNA-Trp"
tRNA 8359..8431
/product="tRNA-Gln"
tRNA complement(8431..8493)
/product="tRNA-Leu"
gene complement(8498..8650)
/gene="ATP8"
CDS complement(8498..8650)
/gene="ATP8"
/codon_start=1
/transl_table=5
/product="ATP synthase F0 subunit 8"
/translation="MPQLSPASGLFIFCTVLLILSILIFALNFCPLPVKQPQIPSQKL
THKIFN"
tRNA complement(8652..8713)
/product="tRNA-Asn"
gene complement(8720..9367)
/gene="ATP6"
CDS complement(8720..9367)
/gene="ATP6"
/codon_start=1
/transl_table=5
/product="ATP synthase F0 subunit 6"
/translation="MMTDLFSSMDANTSFFLWLLPLAAAAFFIFNQSFMSFFSVLGN
SICSLWGTEETFFSKKLLLA VMSFLMLN NFLGLLPLVFSTTSLWINSSVALVLWGS
ILISGWVFSFKKSAHLAPAGAPGALIPFLVLIETVSILIRPITLTVRLVANISAGHI
ILGLIANVLASTLSFIPFSLAASVMIFYL FEMFVSIQAYIFLLISLYMAEHP"
tRNA complement(9368..9431)
/product="tRNA-Arg"
tRNA complement(9433..9497)
/product="tRNA-Glu"
rRNA complement(9498..10234)
/product="12S ribosomal RNA"
tRNA complement(10236..10298)
/product="tRNA-Met"
gene complement(10300..10647)
/gene="ND3"
CDS complement(10300..10647)
/gene="ND3"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 3"
/translation="MMYYCIFPFLTTILVLLYFFTLWKAASPPSSKESPFECGFDP
SSMRKPFSLRFFLLIILFLIFDVEVLLFPILTQMKMATSTVVLAAYSTFLMLLGG
FYEWAMGALDWIK"
tRNA complement(10660..10724)
/product="tRNA-Ser"
tRNA 10733..10793
/product="tRNA-Ser"
gene 10779..12119
/gene="ND4"
CDS 10779..12119
/gene="ND4"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4"
/translation="MGLAIVFLTILGLFFSSWDVVMTLVVGVLLSFLNMYLNFNFNL
FSDLFLETSLLSMLLFLSLLSILAILCTWWIKSTKYIMTVMALNLFVFAFSVKNLL
VFYFFFEASLIPTLALIIIWGYQPERLQAGSYMMLYTVLASFPLLFFILWVNWNMSSL
DFLLKMLGSSLGAMLIFSSLGFLVKLPMSVHLWLPKAHVEAPLSGSMILAGVLLK
LGGYGLYLLNYSFSLIVLGTSSMIICFIALWGGLFASLMCLQQQDMKALVAYSSVSHM
GLVIGGVFLNLNWGVVSAKITMVAHGFTSSALFVLVNMSYKKVLSRSFSASGGLLSIF
PKMALLWFLFCSINMAAPPSLNLVGEYLLPVLFVLSNIFVLIMALSLFSLAGYNMML

YTINGNHGATNNLMTPTSYSTADYFSLILHLLPFFFLKLNMFYCYSSTLLDES"

tRNA complement(12114..12174)
/product="tRNA-Thr"

gene complement(12175..12952)
/gene="COX3"

CDS complement(12175..12952)
/gene="COX3"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit III"
/translation="MFKTPFHLEVEFSPWPLLSALLTCMPLGLIYYIRIGTPYLLAYG
MACTAMVAYLWWRDIVREATYQGHHTTYVVKGLKLVMLFILSEVCFFFAFFWAYFHS
SLAPSVEIGSVWPPVGITTLAFAQVPLLNSTVLLSGVSVTWAHHAIEEGKYNSAISG
LLLTLLGLYFLFLQFGEYKETSFSIADSVYGSTFFMATGFHGLHVAVGATFLFVCLI
RMILYHFNTTHVGFLLAAAWYWHFVDVWVFLYVSIYWWGS"

tRNA 12999..13063
/product="tRNA-Ile"

gene 13068..14000
/gene="ND2"

CDS 13068..14000
/gene="ND2"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/translation="MMSWYGVLFVIVFLGPLISLSTSNWLLVWGGMELSLISLLPIL
NLAKSSIPVESAMKYFLVQASAVILITGLNIYMLEGSEYLILFILGLILKLVV
PLHFVWIPVVKGLKYIHMAFLGPKIVPLMLMTNMNLDSSIFLFLVSSIGTVLLG
ALLGNNMSSIRMLGASSISHSGWFLMGGLVGFMIWYFSIYMISLYVVLYLKVYVPSD
NYLVSIFLFSLSGLPFLMFPKMMVLYSLLSMGLSPIVFSFFIISIVSLNFYMKFS
FSFVLSSENQKNFGAGVMPGLLLMMGGFILFFI"

BASE COUNT 4331 a 1915 c 2417 g 5397 t

ORIGIN

1 ttgcggtga cttattcaa ctaatcataa agatattggg accttatata taattttgg
61 agtttgatg ggaatagtg ggaaccggatt atctttatta atcgtcttg aattaggaac
121 tctcggggtc ttaatagatg atcattttit taatgtgga gttacagctc atgcttttgt
181 tataatttt ttatagtaa tacctattat aattggagga ttggaaact gaatgggtcc
241 actattgatt ggtgctccc atataagatt tctcgaata aataataaa gattttgact
301 ttgcccctc tctttattt tactcattag atcgagaata gttgaagggt ggcaggaac
361 aggatgaact gtatacccc cattaagagg aattactgga cacagtgggg ctccggtgga
421 tctgctatt tttctctc actggctgg aatatctct attttaggag caattaatt
481 tattacaact attttaata tacgagctcc aggaataacc ttgaaacga tgagattatt
541 tgatgatct attttagta ctgtttttt attactctt tcttacctg ttttagctg
601 tgcattaca atgcttttaa cagatcgtaa tttaataca agatttttg atccctggtg
661 tggcggggat ccaattttat atcaacattt gttttgatt tttgacac ccgaagtta
721 tatttaatt ctctcgggt ttggaattat ttctcactt ctggtaatg ctacagtaa
781 acaaccattt ggaagattag gaataattta tctataatt tctattgaa tcttaggttt
841 tatttttga gcccatcata tgttactgt tgggatggat gttgacacac gggcttactt
901 tactgcagca acaatgatta ttccggtacc tacaggtatt aaagtttta ggtgattaat
961 aactctttat ggaataata gtaagtttga cgctctata tactgagat taggtttat
1021 tttctttt acittagggg gcttaacagg gattgtattg tcaactcct cttggatat
1081 ttgctacat gatacatatt atgtagtggc ccaatttcat tatgtcttt caataggagc
1141 agttttgct attttgcag gattgttta ctgattccct gttataactg gattgtttt
1201 acaagaacgg ctggccaaag ctcaatttat tattatatt attgctgta attaaactt
1261 tttctctca cafttttag gattatcggg gatgcccca cgatactcag actatccgga
1321 tagatatttt aaatgaaac aaatttctt tttgctca cttatgcta ttttctgt
1381 acttttatt gtcttaatcg tatgggaggc cttttaaga caacgatctg taatattag
1441 tccgctgct cttattcac gagaatgaac agaaaataat ttctcctg atttccatgg
1501 taactaatt ccaagaattt cagtttata atttgaacc tcaaaact attgtatat
1561 gttacttaca ataacaaagt gctttattg actggggtat attaaataa tegtattatg
1621 actattataa aaattaatta aaatgattat aaaagtagta ggtttgtct tacctttgc
1681 ataatggtta aataaaac tttgtttat tcaattacc gaaataagaa gatctatgat
1741 ttttaactg ttaggttgt aatttaag tagaataatt aattacaat gaaattatg
1801 gggtgaaacg cttcaattc ttaggataac tggattaat agaaaaacat ttagtgttt
1861 agtcaggcaa taagtataa gattattgc taaaaaagat ttaattttt aatatttaca
1921 ggaactaat atctatagt atatataac tagaattgtt ttaaatga tctttttaa
1981 aatatttaag ctttcatgaa tatctctgag ggttttact gtttaatta gtattaagt
2041 ttaatttta ataaggaact cggcaaaaat tttctcaa ctgtttatca aaaaatagc
2101 ttaagaaag ttagaataa tttaagttt cacctgccct gtgatttat ttaacggcc
2161 cagtacttg actgtctaa ggtagcataa taaattggc ttaattggg gtcgagtatg
2221 aaagggatta tgggaaata ctgtctttt aaaaattaca taaatttac tgaataggtg

2281 aaaataccta tagttacttt taagacgaga agaccctaag aattataaft aaaggatatt
2341 agttctcgt tttgttggg gcaaccattgt taggaaat aaccaacatt aataaagttt
2401 tacatgtcta ctataatag aaaggtttta attaccttag ggataacagc ataattata
2461 ttatagattg tgacctcgtat gttggactag ggactttagt tattatccat tactagagat
2521 agttctgttc gaactgtaaa accctacatg atctgagttc agaccggcgt aagccaggtt
2581 agtttctatc taatatagga ggtatagtag tacgaaagga ccctatatct acaaaftatt
2641 tgttttaact tggcagaatt aatgtaattg atttagggtc aatcaatacc atagttggtag
2701 ttataaftaa aagaagttat gaaataatat taactttggg agttaagac ttaggttaaac
2761 tcacttttag ggaatatact ttaacatag aagggtgtgc ttgcatgca ataagtaaaa
2821 aaatftttta ttccaaaat ttaatagatg aactcatcta gagatgattg ttcttttgg
2881 gttagggaata tttttagat ttagttttg ttcttacga agtccagctc catatggggg
2941 tgccttgatt gctcttagaa caaftgggtgc ttattatta agtttaggtt gcagtactg
3001 atatggttat ttattattt tagttatgt ggggggcctt ttggtattg tttgtatat
3061 tattatatta agaagtaatt tcaatctaca agtaagattt aagftaatag gtttaattt
3121 ttggctttt tttagttaga aattatataa ttttagtagc cctaaaaaca ggctagggtg
3181 aagattaagg gaatgtagag aagattttag tttaggttta ttttaggtt tagggggcct
3241 ttgtattta gtttttttg caattgtca tattgtttt ttgaagggcc aaccagtcca
3301 agtgaanaat gattaaaata tctcgggtgc cagtgttatt attattaagg tgtttgctt
3361 taatgagttt atftttttt agtgtcttta aaagagggtga aacatagta attgaactta
3421 atttaatac ctaagaaga gtaaggttta gaactttatt tattttgat aaaattagtc
3481 ttgatttgg agcgggtgta acatataatt cctttgtgt tttagtttt gctaatact
3541 atagagtga agataacttc cagttcgtc ttattttgat tctgtctata ttgttttat
3601 caataaacat ttaattttt aggggatcat ttttttctt attacttggg ttgagttggc
3661 tgggaatttc atcttttgc ttaattatc attatcaaaa taaactaga ttagggtctg
3721 gttatctaac tcttctaact aatcggcttg gtagtgggt tattattatc tgcgttccat
3781 ttttttaat tctaggaaga ttaataatt tcccatat atctagggtat gaaacafta
3841 tttttattt agccattgcg gctcttacta aaagggctca atactttt agagcttggc
3901 ttccggcagc tatggctgcc ccaactcctg tgagtgcact ggttcaactc tcaaacctg
3961 tgacggcagg cgtctatctt attatccggg taagggttaa ttgtagtata agcaattcta
4021 cggcttcaact gcttttatt ttgtggaggg taacctgcgt ttggggcggg ttaagagct
4081 ttatgagaa cgtactaaaa aaaattattg ctftttcaac cttaagaca ttagggttaa
4141 taatatttc tttagggta aaacttccaa acttagctct cttaacatta tatgccatg
4201 ctatattta agctctttg ttttaagtg cgggattaat cctaattagt gggtttgggt
4261 ctcaggattt acgtctccta ggggcaacgc tctacactat cccaagagt ttgtttttt
4321 ttaatttag aagtttatg ttaatgggaa tccctttct tagggcattt tactccaaac
4381 atgtaattta tgaagtgggt gctataagag agactaacgc aattagggtt ttttaatat
4441 taattggggc tatttactt acaattatt caattcgaac ggtaaaaaat ttaagatgaa
4501 attctattc tggctatcg gtttttagcc gattaccctt ttatacatat gtaccittat
4561 gcattttgt ttttacgta attattagag gaaaaacatt tagaatgta gacattagg
4621 acttaacatt catgtttact tcaaggattt atcagftaat actccattt ttaatggctt
4681 taggagtaat aattgggtta tttttatctg gtgaaaaaaa aaggcatatg cttagaagta
4741 ttttttttt atggccaagt tctaataact taacaaaaa tttttggccc gfttaaaga
4801 gagcaaaaat attagactac ggatgagtag agccagttgt tcttttaaga aggtcattga
4861 acaagctagg aataaattg cattttttt ttttaatga gtccaaatat taaatagaa
4921 tagttcgtg agttttttt agattaacag tttctttt atgtattat ttattataa
4981 agctgtaatt ttatgttat gttattatt agctgttct tttttactt tactgagcg
5041 aaaaattata ggttatatac aaattcgtaa aggtccaaat aaagtgggta ttgaggaa
5101 tgttcaacc ttggcagatg ctftaaaatt atttataa gagaaagtt ttgtatatat
5161 atctaaacat ttgctttta ttttacgcc tatactagga ttttttga gtttaaggct
5221 ttggaattta gctccaagta acataatagt aaattataa attttgggc tattattat
5281 tttttttg acggggttaa atgttacac aacggcacta gctggatgat catcaaatag
5341 cttttatct tttctgggg caattcgagc ttcagcccaa tcaatttct atgaagtaag
5401 gttagtata ctctcttt ttctgcttt agttagtaga tcaatattt gatttagagc
5461 cttcatacg cccctctt ctcttttag tttctttg ttggcaattt gattgtcag
5521 atctctagct gaaacaaac gagcgccttt tgattttgct gaagggtgagt ctagctagt
5581 ttcaggcttt aatgtgaat acgaagggtg gctttttgct ctcttttt ttggctgaata
5641 taccagattt atttttat caatggccac cacagtctga tttcttct ctgtattac
5701 tattgacttt gtttaafta catataat ttgtattga ttctattg ctagagggt
5761 ttatctcgc tatcgtatg acttataat gatactatgt tgaanaagct acctccctt
5821 tgcaattaga ctttactct actttgttt tagaattata ttttaatatg aggcaactt
5881 tattcattc tattttggg ttgtttatt tactttact ttttctct caaagaaaac
5941 atttaattc tagcttaac attttgaga gcttaatagt aattaagatt atttatttt
6001 tttttagatt ggggagtga aactccattg gctttttatt actaattta acttttgct
6061 cgcgtaagc ggcgttagct ctctcttt ttattagat ccttcgggtt ataaaaaatg
6121 ataataatt aaacttaagt tattaagct tgcgaaaaa tagattata gaaggtttga
6181 caagactcc aactcctta aatattagta tttttgaag cgggtgttca attttaggaa
6241 taatattagg tattcaacta gtaacaggat tattttgtc aatacattat acaagggata
6301 ttctattc atttgattc gtatgtcata ttatgcgaga ttttccaat ggttacttt
6361 ttgcttatt tcatgccaat ggggcttct tctttttt tttttatac ctccattg
6421 gtcgggggtt atactacaa aggtacagtg tgcacactaa aacatggata gtcgggggta

6481 caatftttct tattagaata gctactgctt ttttaggcta tgtgttacct tgggggcaga
6541 tatcatttg aggtgccaca gttattacca atcttttacc ggcaattcca tactgaggcg
6601 gaaccttggg tgagtgaagtc tgagggggat ttcggtagg gcagggccacc ctaaatcgat
6661 tttttccct acattttatt ctccatttt taatgcttt gctgttata gtccatttaa
6721 ttttttga tgaaaaaggc tcaacaaatc ctttagggca gactcatcat attaataagg
6781 ttacttttca ccctttttt acatataaag attttagtagg ttttattgt gtcgggctta
6841 gattagtitt aatttgcgct tattttccat ttagtgttc agaccagaa aattttatt
6901 atgctaacc tatagtaact cctactcaca ttcaacctga gggactctt ctgttgcatt
6961 atgctattt acgttctatt ccttctaagt tggggggggg tgtagctctt gctataagaa
7021 ttttatttt atattttctc cctttaagt ctcaatttat gtccgttcca gctagcttca
7081 acatgattca gcaaatgtta ttttgatgat tagctgtagt atttttact ttaacatggg
7141 taggggcatg tccattgaa gatccttatt tatctttgac ggggcttfa acaattttt
7201 attttttag ttaccctta ctaataatg ttggaaactt ttccgtgaga attcttaaat
7261 aaacctttt acctagtta ttataaaac actcgtttgt cagcggggag aaaccgtaaa
7321 acgtgttga gtagtagttt aaaaaaacg aataatttc aaaaatttt ttgtagagc
7381 cctacttata aacagatagc ttaatttta aagcgtagct ttgaagcagc taagaaaaag
7441 atagctttt tgtttaatga gttactagg ttcatataa ttaatagatc caagtctcc
7501 aattcaaat gaaataaac ttttcatga tcatgcaata attttataa ttggaattt
7561 tacgttagt gctttattag gtgtaaaact ttgcttaaat aaatttagaa ggcgggcaac
7621 acatgagcc caattgttag aaatttttg aacaattatt ccagcattt tattaattg
7681 gttagcttfa cctagctac gacttttatt tcttctgat gagcaaggga gcaattgggt
7741 aattttaaa gctacgggac atcaatgga ctgaagtat gaaattcca gaagaggta
7801 tgactcgtt gattcttata tagtacagga acaagcttta actaacgggt aattccggct
7861 tttagaagt gataaccgic cagttttacc ttatttcaat gaaattagag tgttaactac
7921 aagttgggat gtaattcatg cctgagcctt acctctatg ggagtaaaaa tagatcgagt
7981 tccaggcgca ttaaatataa taggtcttfa tctcttact cctgtgtat tctacggcca
8041 gtttctgaa atttcgggg caaatcattc ctatatacc attgtagttg aattgttca
8101 ctgaaacgac ttttaattt ttttggat cttttgta agaggcagt tgaaaattg
8161 ttttagatt ttatccaaa ctctaagggt taaactgctt atttgggtg taagatccc
8221 gccaaaacga cggagagtt cagactacgt agtataatga gtacctctc ctccaagca
8281 gaaagtctta atcatagagt agtaatatag ctgtaagata ataaaactat cgacctcaa
8341 agtcaaat atgcgtttgc gttagctatg gtttctccta agcttctgt ttttggaaa
8401 atcaaatgt cgggtgctaa acaccaaga gaaataaagg gtaaaacca ttaagggcgc
8461 taaaaccct tgcatttatt ctgccaatca taaaagtcta attaaaatt ttatagttta
8521 attttgaga aggaattga ggtgtttta ctggaagggg gcaaaagtt aaagcaaaa
8581 taaaattga taaaatgagc aaaactgtc agaaaataaa tagaccagat cgagggctaa
8641 gttggggcat actaaataa caatttgtt tcttttaatt aacatttaa tgcittttt
8701 aagctatatt tagtaaaatt taagggtgt cggctatata taaactaatt aaaagagtaa
8761 aaatgtaagc ttgaataaaa gaaacaaca ttcaataaa ataataaaa attataacag
8821 atgccgctaa agagaaagga ataaaacta aagtactgac taaaactta gcaattaac
8881 ctagaataat atgacctgct ctaattatg ctactaatgc gacagtcaa gtaattggtc
8941 gaattaaaat tctacggtt tcaattaga ctaaaaatgg aattaaagcc ccaggagctc
9001 ctgctgagc taaatgagca gcagacttt taaaagaaaa aactcaacca gaaattaaaa
9061 tagaacctca taaaactagg gccactctc tattaattca aagtctgtt gtagttctaa
9121 aaactaagg tagaagtct aaaaaattg taagtataag gaaacttatt actcgtgtta
9181 aaagtaatt tttgaaaaa aaagtctct ctgtcccca taagctcaa atagaattac
9241 ctagaaccga aaaaaagaa gatataaag attgattaaa aataaagaat gcgctcgtg
9301 caagaggtaa taatcataa aaaaaggaag ttttagcgtc catggaagaa aaaagatctg
9361 ttatcataa gaacgtaagg aaaaccttaa aatttaagg ccgaaacta aacctgggc
9421 gttcgttct ttatctaaa aagaagatt tctatttt accacgaaa ggtaatgtaa
9481 atttacta tctagacta gctttattt agctagagac agctccgct accttacta
9541 tgttacgact tctctattt tttaacgaga gtgacggcg atttctgac tattaaaaat
9601 attattcaaa tttataaac ttttttatt ttacttataa gtccacttt ataaaaatt
9661 taaattttt attccgaact aaaaaggtat tgaactcgc ctaaacataa ttatggctg
9721 cacctgac tgtgttatt aacaaaatta ttaattaaaa atcttctaa aaaagattt
9781 acagcagcgc atacaactt taaaattatg ttgatatagc ttggtggtta tcaattacta
9841 ggttaagtcc ctaaataggt tttaaaagcc gccatgattt itaagtcca acattaagat
9901 gtttactcc tcttatttt ttcttacc taggataa gggtatctaa tctagtctc
9961 tctccaggt taaaataaaa taaaacttt cctttgggtg atacttctct aatgtattc
10021 actactatta gtaagtataa agaacagttg taaaataaga ttttaagta attttactt
10081 aggtttgacc ggggtgctg gcacctaaaa taccataaaa tacactgtaa cctaaattta
10141 tttttcaata atttataaa atttttaact ggggttactt aattttccat aattaaat
10201 acagtgaact aactttacc atcaaatcag gtagttctaa aaaggggtta cccattttg
10261 ggttagagc ccaatagcct attgtagctt atttttagtt tatttaatic aatctaatg
10321 gcctatgct cactcataa aaagccccc caaaagtatt agcaagaatg ttctatagc
10381 agctaatac acagtctgg tggctattt tatttgagt aaaaaggaa acaaaaggc
10441 cactcaaca tcaaaaatta aaaataaat aattaataa aagaatcgt aagagaagg
10501 tttctgatt gatctatag gatcaaac acctcaaat gggactctt tagatgact
10561 cggacttct gctttcata gggtaaaaa atataataa actaaaatg ttgttaaa
10621 aaaagaaaa atacaatag atatcactt tagtgggtg actaaacggc taagattaag

10681 ccaaaaaa tttcaaaa ctttgtaaa ataaaaata aagaggttt taattgagc
10741 tgctaacttt gattagggc ttatgctcc gccttttt gggctctgct attgttttt
10801 taactatctt aggattatt ttagtagtt gggatgtggt agtaataca ttggtgttg
10861 gtgtattact tagatttta aatatatatt taaatttag gtttaattt ttagtgatt
10921 tgttttaac tgaagacta agaagaatat tattttctt aagattatta attagtatt
10981 tagctatfff gtgcacttga tgaattaaat ccacaaaata tattataact gtaatagcat
11041 taaatttatt tttagtttc gcctttagag taaaaaattt atgggtttt tactttttt
11101 ttgagccctc tcttattccc actttagctt taactattat ttggggatac caaccggaac
11161 gactacaagc tgggaagtat ataactttt atactgtact tgcctcattc cctttattat
11221 tttttactt atgagtaaat tgaacataa ggaggctgga ttcttttta ttaaaaatgc
11281 tagggagttc tttagggggg gcaatactta ttttagggag gctaggtttt cttgttaaat
11341 taccaatatt taggtccat ttatgactt ctaaagctca tgtggaagct cctttaagt
11401 ggtctatgat tctgtctggt gtattattaa agttaggggg ctatggattg tacttattaa
11461 attattcttt tcttttaatt gtactaggca ctagaaggat aattatttgt tttattgctt
11521 tatgggggtg attatttgc tcttaatat gtttacaaca gacagatata aaagctttgg
11581 tggcatattc ttcagtttca cacataggat tagttattgg aggtgtttt ttaatttaa
11641 actgaggggt tgtatcagct aaaattacga tagttgctca tgggttact tcatctgcat
11701 tattgttttt agtaaacata agctataaaa aagttttgag tctgtagattt tccgcaagag
11761 gtgggactact taggattttt ctaaaaatgg ctttattatg atttttttt tgcagaatta
11821 atatggcggc tctcctctct ctaaattag taggggaact ttatttact ccggtactat
11881 tttatttag taacattttt gttttaatta tagctttgct attattttg agggctgggt
11941 ataataaat actgtactat aatggaacc acggagctac aaataattta ataaccceaa
12001 gaacaagata tagaacagct gattattttt ctttaatttt acattattg ccaattttt
12061 tcttattaaa ataaacata ttttattgt atagtagaac actactggat gaaagctaat
12121 gggaaatttc tttattgact tacaaaatca agattttata taaattacat tagcaagatc
12181 ctcatcaata aattctcaca taaagaata atcagacgac atcaacaag tctcagatc
12241 aggcagctgc taggaacct acgtgatgag ttgtgtttaa gtgtagaga attattcgaa
12301 taaggcacac aaacaggaaa gttagcacca cggctacatg aagccatga aatccagttg
12361 ctataaaaa agtactacca taaactatc cagcaattga aaaactgtt tctttatatt
12421 caccaaatg aagaataaaa aaataaagt ctaacaatgt tgttaataag agtcctctaa
12481 tggcggaaft gtatttctc tctcaatcg catgggtggc tcatgtacc ctacacctg
12541 aaagtaaaa aactgaagta ttaaggagag gtacttgaaa ggctcttagc gttagtaatgc
12601 caacaggggg tctactctc ccaatttcta ctgaaggggc taaactctg tggaaataag
12661 ctcaaaaaaa gcaaaaaaaa aaacaaact ctgataaaat gaagagtatt acacctaatt
12721 ttatgctctt gactacatag gtatgtgat gtcttgata tttgtctcg cgaacaatat
12781 cgcgtacca taggtatgct actatagcag tacatgctat tccgtaagct agtaagtaag
12841 gtgttccaat tgaatataa tagattaac ctaaaggat gcaagttaa agtgcgacc
12901 ctagcaaggg tcaagggcta aattctaca gatggaaagg agttttaaac atattaagt
12961 aattgtaaaa gattttaatt ggttataaaa attcttaagc aaccgcccgc agagcgggta
13021 tcaatgatgt tgaatataa gaatattta attgttcgtt gctataata atgagttgat
13081 atgggtttt atttctgta attgttttt tagggccatt aattagtta tctactcca
13141 attggttatt agtttgagg gggatagaat taccactat taggttactt cctacttaa
13201 atctagctaa aaggcttatt cctgtagaaa gggcgataaa atatttttg gttcaggcaa
13261 gtgcaagagc agttattta attaccgat tgaatattt tatattaga gggagggaaft
13321 atttaattt gatttattc attttagggt taactttaa gctaggagta gtactttac
13381 atttttagt aattccggtg taaaaggac taaaatata tcatatagca ttctactag
13441 ggccattaaa aattgttctt taatattat taataactaa taaaatttg gataggctca
13501 tttttttt ttacttga aggagaattg gtaccgtttt gctaggtgct ttattagta
13561 ataataaag atctattcga ataatactg gagcactctc tattagtcat tctgggtgat
13621 ttttaatagg gggattagt ggctttat gaactattt taggattac ataattagct
13681 tgtatgtgt ctatattat ttaaaagtat atcctagfga taactatctt gtatctatt
13741 ttctattag tttaaggggt ctccctccat tttaattat tccagcaaaa ataattagtt
13801 tatattcgt ttaagaatg ggcttatctc caattgtttt tagtttttt atcatcagcg
13861 ceattgtgag gctaaattc tatataaaa ttagttttc attgtatta agaagaag
13921 aaaatcaaaa aaattttgg gctggtgtga tgcaggatt gctactaata atggggggt
13981 tcaatttatt ttttttagg ttttttagg ctgagaaata taagcattaa atttttaatt
14041 taaggacgag gttattcctc

//

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 Pupillid 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"
CDS 2..1531
/gene="COX1"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit I"
/translation="MRWLYSTNHKDIGTLYMIFGVWCGMVG TGLSLLIRLELGTSGVL
MDDHFYNNVIVTAHAFVMIFFMVPIMIGGFGNWMVPLLIGAPDMSFPRMNNMSFWLLP
PSFILLISSMVVEGGAGTGWTVYPPSSMVGHS GASVDL AIFSLHLAGMSSILGAINF
ITTIFNMRASGVTLERLSL FVWSILVTVFLLLSLPVLAGAITMLLTDRNFNTSFFDP
AGGGDPILYQHFWFFGHPEVYILILPGFGIISHILGNASSKQPFGLGMIYAMISIG
ILGFIVWAHHMFTVGMVDVDTRAYFTAATMIIA IPTGIKVF SWLMTLYGMNNKFDASMY
WVLGFIFLTLGGLTGIVLSNSSLDIVLHDTYYVVAHFHYVLSMGAVFAIFAGFNYWF
PVMTGLILQERLAKAQFIIMFIAVNLTFFPQHFLGLAGMPRRYS DYPDTYFKWNQISS
FGSLMSIFAVLLFCLLVWEAFLSERSVLF SKAAIYSREWVENNFPPEFHSNIIPSISV
R"
tRNA 1542..1603
/product="tRNA-Val"
rRNA 1622..2667
/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"
CDS 2888..3340
/gene="ND6"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 6"
/translation="MSLLFLILLASFIMFFTSEQPAALGGSLILMSFLSALSFSFMFS
SWYGLMLFLIYVGGLLVLFMYVIMLSSNFILSFLKTTIKGFFVLMFMWALKNFNLPS
KILAESMSSCSSLMMGLFSLG VFLLLVFFSIVHMVLFKGVILKLNE"
gene 3333..5000
/gene="ND6"

CDS 3333..5000
 /gene="ND5"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 5"
 /translation="MSKIYRLPFLLLFGLLFSALSSVMFFLSFKSFMVDIKLIISGS
 SWITIMLIFDKISISFMSVVLCSFCVFLFMCQYMEDEFKFRFGMILLFVLSMNIL
 IFSGSFFMLMLGWDGLGISSFALIIYSSASSKQAGFLTMTNRIGDVLIIASIPFFL
 SGGNFMFFSMEILPYWVVVLLCVAALTKSAQYPFSAWLPAAAMAAPTVPVSALVHSSTLV
 TAGVFLIIRVSMNFNLSDEICSLIFSGGVTCFLGGVSALYENDVKKIIAFSTLSQLG
 LMFVCLGMNMPFLALLHLFTHAMFKAMFLSAGLILLSFGVQDLRILGALQFRSPLL
 ICFFNVSSLCLMGAPFMSAFFSKHVIYEIMFMSSVNTLGVALMVVGSIFTSIYSVRLV
 KIISWKKISMCSINLSLTIFSYPMIILYFMSLVSGKMFSLMDISHNIIMFTPLTVQV
 LLNLLVVGILVGLMMKGERISHLFGSMLFLWPSHNINKFMTRSMKLTTSLDLWIE
 PKNILNNFYLESSFFLENLSNYMKFSLLSLFKGSAFMFMVIMWYFF"

gene 4985..5890
 /gene="ND1"

CDS 4985..5890
 /gene="ND1"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 1"
 /translation="MMFFLIKSLFFCLCVLIAVAFFTLLEKVLGFLQIRKGPKNKVGW
 WGIFQPLSDALKLFVKEKIFKYSSNNILFLIIPFPGFCLSLSMWNLIPSAYNMNFILY
 SLLFFFCIVGLNVYVTALAGWASNSIYAFLGAMRASAQTISYISLVMLLLPALISG
 SMCWFFNMKSFPVSLLMFPPLLITSSLAETNRAPDFAEGESELVSGFNVEYESGL
 FALLFLAEYTSILFMCMMTVWLMWSNTPLSFCLWTMFFSLIFLWSRGCFFRFRYDLL
 MMLCWKSFLPFSMCCLIYTGIFFLM"

gene 5905..6187
 /gene="ND4L"

CDS 5905..6187
 /gene="ND4L"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 4L"
 /translation="MIINFFSVFVIGFMVLLVMFFMQNKHIINSLIILEGMMLMVVSF
 IFFILSLASSGEEVMLLLFTFSASEAAIGLSLLISFIRLYGSDMMMNMFG"

gene 6191..7300
 /gene="CYTB"

CDS 6191..7300
 /gene="CYTB"
 /codon_start=1
 /transl_table=5
 /product="cytochrome b"
 /translation="MRKIKIIESLSALPSPVNISVWWSGGSLGMMGLIQLITGLFLS
 MHYVADTSLAFDSVTHIMRDVPMGWVFRLFHANGASFFIFLYLHVGRGLYYQSYILQ
 PKTWMAGVTIFLSMAAAFLGYVLPWQMSFWGATVITNLLSAIPYWGEILVQVWVGW
 FSVGQATLNRFFSLHFLVPFVILVMVLLHLVFLHEKGSSNPLGHNFHINKIAFHYPYFT
 YKDLVGFVLFVSLIFMCSFYPFYLSDPENFMNANPMVTPTHIQPEWYFLFAYAILRS
 IPSKLGIVIALAMSLILYFLPFLHLGKSPSSFNIFYQVNFVILVVFILLTWLGAC
 PIEDPYLSLAFPLTVWYFLSYFFLAFSNKLNFFIS"

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

gene 7372..8037
 /gene="COX2"

CDS 7372..8037
 /gene="COX2"
 /codon_start=1
 /transl_table=5
 /product="cytochrome c oxidase subunit II"
 /translation="MSHWGAINLLDPSSPVQTEMMLFHDHAMILLVGIFSLVALLGVK
 LSVNKLSTRTHEAQTLEILWTLPGFLLIWLALPSLRLLYLLDEQGNGLILKAIGH
 QWYWSYEIPSSSTSSFDSYMVQEKDLKGEYRLEVDNRPVLPYNTPISVLTTSDAVI
 HSWTLPSMGAKVDAVPGRLENLIGLSSYFPGVFYGCSEICGANHSYMPIVIEFIKLED
 FIQ"

tRNA 8052..8117
 /product="tRNA-Tyr"

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

tRNA 8180..8245

tRNA /product="tRNA-Gly"
 8248..8313
 tRNA /product="tRNA-Trp"
 8319..8385
 tRNA /product="tRNA-Asp"
 8381..8446
 tRNA /product="tRNA-Cys"
 complement(8444..8505)
 tRNA /product="tRNA-Gln"
 complement(8502..8565)
 tRNA /product="tRNA-Leu"
 complement(8569..8718)
 gene /gene="ATP8"
 CDS complement(8569..8718)
 /gene="ATP8"
 /codon_start=1
 /transl_table=5
 /product="ATP synthase F0 subunit 8"
 /translation="MPQLSPSSGLMTFLFMVFHIFINKLYFQNTPKSVSTFKISPQVK
 KFIYQ"
 tRNA complement(8720..8782)
 /product="tRNA-Asn"
 gene complement(8791..9438)
 /gene="ATP6"
 CDS complement(8791..9438)
 /gene="ATP6"
 /codon_start=1
 /transl_table=5
 /product="ATP synthase F0 subunit 6"
 /translation="MMTDLFSSMDANYSFLLWSFPLFMTTFILMKTYLNSLTSVFKA
 SLTNLWTFKESYFWSKLFLTLAMSFIMLCNFSGLLPYVYGSTTSLWVNSSMALLFWGI
 FLLSGYCFSFKKSLAHLTPAGAPLMLIPFLILIESILIRPLTLTVRLVANISAGHI
 ILALMANVLSLQALVPFTGVILVLMFYNFLFEMFVSFIQSYIFTLTLLSLYLAEHP"
 tRNA complement(9439..9504)
 /product="tRNA-Arg"
 tRNA complement(9507..9574)
 /product="tRNA-Glu"
 rRNA complement(9576..10325)
 /product="12S ribosomal RNA"
 tRNA complement(10326..10390)
 /product="tRNA-Met"
 gene complement(10396..10743)
 /gene="ND3"
 CDS complement(10396..10743)
 /gene="ND3"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 3"
 /translation="MIYYVAIPLTITMLFFILYLFLSWNPPSEIVSKGEPFECGFDPF
 SKIRKPFSLRFFLLIILFLIFDVEVVLLFPILMLLKSSCTTPQIISFTVFLIILIMGL
 FYEWKMGALEWIL"
 tRNA complement(10756..10822)
 /product="tRNA-Ser"
 tRNA 10809..10883
 /product="tRNA-Ser"
 gene 10887..12188
 /gene="ND4"
 CDS 10887..12188
 /gene="ND4"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 4"
 /translation="MVLWVFSISLVMGHWLLIFLMLVFFSVFSFLNLYKNFYSYDVFN
 SWIYLDSSFLVLLSVFVLFVFGTYWNKSVIYTKSLLLLVFFLFLAFSVKNLLCF
 YFFFEASLIPTLFLIIAWGYQPERLQAGTYMIMYTVLSSMPFLFFILYCFYSLNTLEV
 VLLMYSKMEISSIVFYIIVLAFLVKLPMYGLHLWLPKAHVEAPLSGSMILAGVLLKLG
 GYGLYLLNYILSLKNLNTNLVVLVLSLWGGLLATLMCFRQSDMKAMVAYSSVAHMSL
 VIVGMFLNSNWGLLSAKITMLAHGFTSSALFMLVNMSYKKIYSRSFNNTSGLLTYFPK
 MSLWLFIFCSINMAAPSMNLLGEMMVIPSLKLWGAIFVILMCIIMLLSAGYNMMLYV
 NNNHGGPSFYVTPSMNYSSEYYGLLMHVLPFFLLFKMELF"

tRNA complement(12209..12271)
/product="tRNA-Thr"

gene complement(12279..13058)
/gene="COX3"

CDS complement(12279..13058)
/gene="COX3"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit III"
/translation="MFKTPFHLVEFSPWLLSSLALTSMPGLGIYYLRFNYWVLLFG
MLLTAFIAYLWWRDIIRESTYQGYHTTYVVSIGIKLGVMLFVSEVCFFFAFFWAYFHS
SLAPAEVGSVWPPMGITTLPTFSVPLLNTSVLLSGVSVTWAHHAIEEGKLSATSG
LLVTVLLGLYFLFQGEYKETSFSIADSVYGSTFFMATGFHGMHVLVGATFLLICLI
RLMLYHFNKSHHVGFLLAAAWYWHFVDVWVWLFYVSIYWWGS"

tRNA 13101..13166
/product="tRNA-Ile"

gene 13170..14117
/gene="ND2"

CDS 13170..14117
/gene="ND2"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/translation="MSILMVFFSTIVFLGPILSGSVSSWFLVWAGMELSFISFLPVMN
YPKSSMSSESTIKYFLVQSCASAMILVSGSFLFIVEMTNMFYYSFLTGLMLKGLVLP
FHFVWIPVMKGCKFTDMIFILGPMKIIPFLLIISIPISSNLLLIGIFMSLFLVIGA
LLGNNSSSVRGMLAASSIVHSGWLLMGCLVGGMGLYFSIYLMQLYPFLFLCLLNQKFF
MSVLMLSFSGLPPFMLFFGKLYILQNIYKEVLVSLLFTHIIFSALISLNFYLKFSYS
FYLSSMLNLKNIKIFITNLMMISSLLGFILFIFYGRAKH"

BASE COUNT 4601 a 1830 c 2161 g 5557 t

ORIGIN

1 tttcggatga ttatattcaa cgaatcataa agatattggt actttatata taattttgg
61 agtttgatgt ggtatggttg gaactggtct ctcactact attcgcttg aactaggtag
121 cctcggagtt ttgatagatg atcattttta taatgttatt gttacagctc atgcttttgt
181 aataahtttt ttatagtaa tacctattat aattggagga ttcggaaatt gaatagtcct
241 tcttctcatt ggtgccccag atataagttt tcctcgaata aataatataa gtttttggct
301 tctctctcca tcttttatt tacttattag tcaagaatg gtagaaggag gggcaggaaac
361 aggttgaact gtatacccc cattaagaag aatagtagga catagaggag catcagtaga
421 tttgcaatt ttctcattac atttagctgg gatgtctca atttgggag ctattaatt
481 cattacaaca attttaaca tacgagcttc aggggtcact ttggaactg taaggctatt
541 tgtatgatct attttagtca cagtttttt acttcttta tcaatcctg ttctagcagg
601 ggctattact atactattaa cagatcgaaa ttttaacaca agatttttg atccagctgg
661 aggaggggat ccaatttat atcaacattt atttggitt tttggctc tcgaagtata
721 tattttaatt ctacctggtt tgggattat tccacatatt ttaggtaatg catcctcaaa
781 acaacctttt ggaagtttag gaataattta tgcataaatt tctattgga ttttaggtt
841 tattgtatgg gcccatcata tatttacagt aggtatagat gttgatacgc gagcttatt
901 tacagctgct actataatta ttgctatccc tacgggaatt aaagtttta ggtgattaat
961 aaccttatat ggaataaata acaaatitga tgcctcaata tactgagtc taggtttat
1021 tttttattt acattaggag ggtaacagg gattgttta tctaattctt cttagatata
1081 tgttctcat gatacttact atgtgtgtgc tcaattcac tatgtttat ctataggggc
1141 tgtatttga atctttgctg ggttcaacta ttgattccct gtaataacag ggttaattct
1201 tcaagaacga ttagctaaag cacaatttat tattatatt attcgagtt accctacttt
1261 ttttcgcaa catttttag gattagcagg aataccacgt cgttattcag actaccctga
1321 tacttatttt aaatggaac agatcctgic tttgggtca ctaatgtcta tctttgctg
1381 acttttatt tctctattgg ttgagaggc tttttaaga gaacgaagag tattatttag
1441 caaagctgca atttactcac gagagtgggt agaaaaaat ttccaccag aatttcacag
1501 gaatattatt ctagaatca gggctcgtta atataaact tactagtaaa atataatgt
1561 atataattt tacaaaaa aagtggcttc gagcctacta gttaitaagt agtattgagt
1621 gtttaactc atataatc cttgaaaaa ttaactctat aattattagg taatcata
1681 cctttgcat aatggtaaa taaaaataa ttttattaa agtttccga taataaaga
1741 tctatttatt ttctatatt agtctgtct ctaaatttta ttgtgcata aataagattt
1801 taataataa atagtagtga aatgtaaca atttattta tatctggata gttttaaaa
1861 gcaatttagg tttaggaac ttttagtga taagatctaa aagcttaaga atataaata
1921 aaacttatt ctaaaagatt tatgtttta tagaaatgat acttaatat aataactttt
1981 ataatcttt atagtaaaa cttttttat ataatttta tagaaattaa taatgttaa
2041 attagtaaaa atttattta actaataaa ggaactcagc aaaaagggtt ttggaactgt
2101 ttaacaaaa catagcttaa agtttatatt ttagggtgaga cctgctcaat gaagatgtag
2161 atattttaa tagccgcagt actttgactg tgcataagga gcataataa ttggctttta
2221 attgaagtct ggaatgaaag gtaaatitga aaaagattgt ctctaattag atttataa
2281 cttacttaat aggtgaaaat tctatgttt ttcataatag acgagaagac cctaagaatt

2341 attaaaaata agtttattt actttaattt ttgttgggg caacaacttt gaacataaaa
2401 ccaaaaggggt aaagatttcta agacaaaattt tattttgaaa aaaattaatt accttaggga
2461 taacagcata attttcttt tagattgtga cctcgaatgt ggattaagga ccccttaggt
2521 aagccactta agagaatagt tctgttcgaa ctttatatcc ttacatgatc tgagttcaga
2581 ccggcgtaag ccaggtcagt ttctatcttt tattttaaa ttgtagtac gaaaggacca
2641 ctttttagta ttttatatat tttagttga cagaacata tggattgac ttagaatcaa
2701 tatatacctt ttatgtaac tattaanaaga agttttgtaa aacattaact ttgggagta
2761 aaaattgaag aaagctcttc acttttaag gftaagtatt ttataattag aaagtatatt
2821 ttgcaagtat aaagaaaacc aaaaggtttt ttaaccattt gcttttaatt tgtatacaat
2881 ttaaaaattg agattactat ttttaatttt actagcctct ttcattatgt ttttacttc
2941 tgaacaaccg gcagctttag ggggggagatt aattttaata agttttctta gagcgttagg
3001 ttttaggttt atatttagaa ggtggtatgg ataatgtta ttttaattt atgtaggagg
3061 gttattagta ttatttatat atgtcattat actaagaagt aatttcattc ttagattttc
3121 tttaaagaca actatcaagg gttttttgt ataatattt atatgggctt taaagaattt
3181 taatcttct tcaaaaattt tagcagagag tatatctta tgtagctcaa gcctaataat
3241 aggggtattt ttaagcttag gggtttttt gtgttagtc ttctttcaa ttgtcacat
3301 ggttctattt aaaggagtaa tcttaaaatt aaatgagtaa aattatcga ttaccttttc
3361 tgttactttt tggcttttta tttccctag cactatctag agtgatattt ttcttaggt
3421 ttaagctttt tatagtatag attaaaattaa ttattcagg aaggatcga attactatta
3481 tattaatttt tgataaaatt agaattagg ttaggatagt agttctatgt atttctttt
3541 gtgtttctct ttttatatgc caatatatgg agaattgatga gttcaaat ttggtttgaa
3601 taattttact tttattgtt ctttccataa atactttaat ttctcaggt tctttttta
3661 tattaact ttgctgagac ggttaggaa ttcttcgt ttccittaatt atttattaca
3721 gaagagctag aagaaaacaa gcagggtttt taactttaat gactaatcga atgggtgatg
3781 tattaattat tgcctagaatt ccttttttt taaggggagg aaactttatg ttttttcaa
3841 tggaaatttt accatattgg gttgtagtac tctatgtgt tgcggccctt acaaaaagtg
3901 cccaatatcc ctttagtccc tggctcccgt ctgcaatagc agctcccacc cctgtgagag
3961 ctttagtca ctatctact ttagtacag ctgggggtatt ttttaattt cgagttagaa
4021 taaattttaa cttaaaggat gaaattttgt cattactat ttttagcggg ggagtactt
4081 gttttttagg gggagtaaga gcactatacg agaattgatg aaaaaaatt atgtcattt
4141 ccacattaag acaactagga cttatgtttt ttgtttagg gatgaatata ccaatttttag
4201 ccttacttca tttattact catgctatat ttaaggcaat attattttta agagcaggct
4261 taattctttt aagaagattt ggagtacaag atttacgaat tttaggagca ctcaatttc
4321 ggttccactt actaattttt ttttttaag ttaggagatt atgtttaata ggggcccctt
4381 tcataagtgc tttttttca aagcatgta tttatgaaat tatatttata agaagagta
4441 acacattagg agtagcccta atagtagttg gctctattt tacttcaatt tactctgttc
4501 gactagtaaa aattatttca tgaaaaaaga tctctatat ctctattaat ttaagtttaa
4561 ctatttttag gtattttctt ataattttt tatattttat gtcgttagtg agaggaaaaa
4621 tgttcagact gatagatatt tcacataata ttattatatt tactccttta actgtacaag
4681 tattattaaa ttttttacta gttgtcggaa tttagtagg ctttaataatg aaaggagaac
4741 gaatcagcca cttatttga agaataattt ttttagacc ttatcccatt aatattaaca
4801 aattttaaac ccgatcaata aaactcacga ctagtctga cttaggctga atggagccta
4861 aaaatattct taacaatttt tacttagaga gaagattttt tttagaaaaa ttaagtaatt
4921 atataaaatt ttctctttta agtcttttta aaggaagagc gtttatattt atagtactaa
4981 ttatatgata ttttttttaa tcaaacattt atttttttgt ttatgtgat taattgcagt
5041 tgcatttttt actttgttag aacggaaagt tttaggggtt ttacaattc gaaaaggacc
5101 taataaaagt ggggtatggg ggatttttca accttatca gatgcactaa aactttttgt
5161 taaagaaaaa atttcaagt atcttctaa taatatttta tttttatca ttccaccttt
5221 tggcttttgc ttaaggctat caatatgaaa ctttaattct agggcttaca acataaacct
5281 tattttatat aggttattat tattctttg tattgtagg ttgaatgat atgtaaccgc
5341 tttagcaggt tgagcttcaa ataggattta tgcattttta ggggcaatac gggcatctgc
5401 tcagactatt tctatgaaa tttagattgt taaactttta ctttcccag cattaattc
5461 tggatcgata tttttgtttt ttaataaaa aagttttctt gtatcggtat taattttcc
5521 atttttata attttaata ccagaagttt ggccgaaact aatcgggccc ctttcagtt
5581 tgcagaagga gaatctgagc tagtgcagg gtttaattgt gagtatgaaa gagggctatt
5641 tgccttatta ttttagctg aatacacaa tagtcttctc atagcttcta taataacagt
5701 gtgttataa tgaagaaata cccatlaag atttgtttta tgaacaatgt ttttagtct
5761 tatcttctt tgaagacgtg gttgcttcc tgggttctgt tatgatctt taataattt
5821 atgttgaag agattcttac cttttcaat atgtttttta atttactg gaattttctt
5881 tctatataa ttttaagag ttaataaatt ataacittt tttcagtttt tttatttgg
5941 ttcatagttt tactagtaat atttttata caaataaac atattattaa ttctttaatt
6001 attttagagg gtaataattt aatagtagta tctttattt tttttatctt aagtttggcc
6061 ttaggggggg aagaagtgat atttattata ttacctttt cagcctcgga agctgctatt
6121 ggcttattt ttgtaacag atttattcga ttatcggaa gggatataat aataaataa
6181 tttggittat tgcgaaaga ttaaaattat tgaagatta agagctttac cttcacctgt
6241 aaacattaga gtatgatgaa gagggggatc tctcttaggt ataactag gaattcagct
6301 aattactggc ttattttat caatacacta ttagcagac acaagattag cttttgactc
6361 ttgtactcat atcatcggg atgtccctat aggatgagta ttccggcttt tccatgcaa
6421 tggggcatca tttttttta ttttttata tcttcatgta ggtcgggggc ttattatca
6481 aagttattat cttaaccaaa aaacttggat agcaggtgta acaattttct tattaagaat

6541 agcggcagca tttctgggt atgtgctcc ctgaggca atactttt gagtgcaac
6601 agtaactact aatctttgt ctgctatcc ttattgagg gaaatttag tccagtggt
6661 gtgaggaggg ttctctgtgg ggcaagctac atgaaatcgg ttttttcat tacattttt
6721 agtacccttt gttatcttag taatgggttt actacaftta gttttttac acgaaaaagg
6781 atcatcta atccttagggc ataattcca tattaataaa atgcattcc atccttatt
6841 tacttaca aa gatttagttg gattgtatt agtttttaga ctacttatt ttatagttc
6901 tttttatct ttttactaa gagaccaga aaatttata aatgcaaat caatagtac
6961 cccaactcat attcaaccag agtgatactt ttatttgct tatgctatc ttcgatctat
7021 tccgagaaaa ctaggagggg tcattgcctt agcaataaga ttattaatt tatattttt
7081 acccttctt cacttaggga atcaattcc aagaagattt aatattttt atcaggtgaa
7141 cttttgaatt ttagtgtta tttttttt actaacttgg ctggggcct gccctattga
7201 agacccttat ttgagactag ctttctttt aacagttga tatttctta ggtattttt
7261 cctagctttt agtaataaat taaatttctt tattaggtaa tattaggaaa catagcttat
7321 tttaaagcct gcttttgaag aagctaagat aatatttata tttttttca catgagacat
7381 tgaggagcaa tcaattttt agaccaatc ctacctgttc aaacagagat aactttttc
7441 catgatcatg caataatttt attagttggg atttttagct tagtagcttt attaggagta
7501 aagttaaagag tcaataaaact tcaactcga accacacatg aggctcaaac attggaatt
7561 ctttgaacca ttttgcctgg ctttctctt attgactgg ctttaccag tcttgggtg
7621 ttataattac tagatgaaca aggaaataat ggggtgattt taaaggctat cggccatcaa
7681 tgatactgaa gttatgaaat tcttagaaga tctacttcat catttgactc atatatagt
7741 caagaaaaag acttaagaaa aggagaatat cgcttattag aagttgataa tccagctgtt
7801 ctccatata atactctat tagagttta actacaaggg cggatgttat tcaacttga
7861 acccttctt ctataggggc taaagtagat gctgtccctg gtcgattaaa ttgattggc
7921 ttgtcttctt actttctggg ggttttctat ggtcaatgtt cagaaattg tggtgcaaat
7981 cactgtata tacctattgt aatcgagttt attaaattag aagattttt ccaataaagt
8041 attagttttt aatttgagaa tctaatttag agatagtctg taaactatt atttggaaag
8101 cttttttccc tcaaaacttt taggttaatt aaactattta attgttgatt aaaagatttc
8161 gtaagaaagg tttcaaaaga gtttcttagt atttaagtag gattgccttc caagcaaga
8221 gtcttcttt tagaaaggga aattaaataa ttataaagt aagttaact agtggacttc
8281 aaatcaaaaa atattaaaa ttgattaaat ataagtagtt ttaictagt taaaaataa
8341 tttttaattg tcaattaaaa gaaatagttt cctctatgat aggttagtag tttagaaaa
8401 aaacaggtaa ttgcataat tactaaagaa aaatttctt taccctttt taaaaattt
8461 ttttcttga aaatcaaatg ttctgtgct ttacacaaa gaaaaataa gaaaataatc
8521 ctatctaga cgcttaaaat ctatgcaatt ttctgccatt ataaagtfff actgataaat
8581 aaatttttt acttgcgggtg aaatctaaa agtttctact gatttaggag taifttga
8641 gtatagttta ttaataaaaa taataaaaa taaaaataa aaagtatta atccactga
8701 agggctcaat tgaggcatgc taaactaact attctagttt ctattaata acatttaatt
8761 gttttata agctatagtt tagaatttaa ctatgggtgc tctgctaaat ataagcttaa
8821 tagtaaggta aaaatatatg attgaataaa tgaacaacac attcaataa aattgtaaaa
8881 tattaacact aagattcac ctgtaaatgg gactaaagct tgaagggatg ataaacatt
8941 tgctatcaaa gctaaaaataa tttggccggc gctaatgta gcaactagcc gcactgttaa
9001 ttttaattgg cgaaataagg tactgattct ttcaattaa ataaaaagg ggattaatat
9061 aaggggagcc ccagctgggg taagatgagc taacgattt taaaagaaa acaaatagcc
9121 agacaataaa aaaatacctc aaaataataa agctatagaa ctatttact atagcctagt
9181 agtggaccca taaactgaag gaagaaagcc actaaaata cataatataa tgaaccttat
9241 agctaaagtt aaaaagact tactccaaa gtaagactct taaaagttc ataaattgt
9301 taactagct taaaaaactc ttgtcaaaaga gtttagataa taagtttta taaaataa
9361 tttgtgtata aataaggga agcttcacag taaaaagaa taatttgcatt ctattgatg
9421 gaataaatca gttatcaatt taggagttaa taagaattt aatttttaa gcccgaact
9481 taattgtatt gaaactcac ctataact tttaaaaa taatacttt ttctaaagca
9541 tgaagagctt ttgtattta taaactataa aagtaataa ttgtcatta ttgaacag
9601 ctccgctac ttactttg ttacacttg tctatttc caacgagaat gacggcgat
9661 ttgtgacct taaataaaa ttgtcaaaag ttatttttaa tttaacttt tacaataaa
9721 tctatttaa aaaaagtata taaacttt tactaaca aaatctttt gtaactgcc
9781 taaaactcat aattattagc tgcacctga tctgatata taaaagatta ttctttaaa
9841 tttttttaa aaaaatttt ataacgacgg catacaact atttaaaat tatgagat
9901 agcttggta ttatcaatta ctggtaagt tctctgaag atttttaaag gccgcataa
9961 tctttgatt tcaatactt tatttacta ctactcagg ctgtctata ctcttctat
10021 aacagggtat ctaactctgg ttattatca agtgttactt aaaaattaa ttgtttaaa
10081 gggttaaatt gtttactac taaaatttc acttatctat ttgcaaatg tcttttagt
10141 aagccctagg taaaatttt ttataggt atgaccggg ttgtggcac ctattttacc
10201 aaagtaaaaa taaaactaa acttaaat cttagtttt taattttgt aactggataa
10261 gtaagattta tctccataa ttattttaa ttaataatt ttctatcag atttaataa
10321 aaatttctaa gaagaaaaaa atctccattt ttgggtatg agcccaacag cttaaataag
10381 ttattcttag atgattata aaattcactc taatgccct atctttcatt cataaaaaag
10441 tctataatt aaaaataa agaaaacggg aaagtaata attgaggcg tagtacagga
10501 tctttaa atgtataaaa taggaacaa tagaacaact tcaactcaa aaaaataaa
10561 tagaataatt aaaaaaaa aacgtaataa aaaaagctt cgaattttc taaaaggatc
10621 aaaccacac taaaagggt cacctttga gacaattca gagggcgggt ttacggataa
10681 aataaataa aaaaaaaa ataattttg aatagttaa ggaatagcga cataataaat

10741 caatttagtg tattcactag atagtaagta aaacttactt caaaagattt tcaaaatcct
10801 ttatatataa aaatatactt tattaaaaagg attataanta gggctgctaa ctttgatttg
10861 gggacattct cctccttttt ttgagaatag ttttatgggt attttctatt tcacttgttt
10921 ttatgggcca ttggctatta atttttttaa tattagtitt ttttcagtt tttagattt
10981 tgaactata taaaaatttt tcttatgatg tatttaatag atgaalltat ttatagtagga
11041 taagatcttt cttagtttta ttaagagttt ttattgtttt attctctttt gttgggacat
11101 attgaacaaa aagagttatc tacactaaaa gactgttgtt attagtittt tttttattt
11161 tagcttcag agtaaaaaat ttactttgtt ttattttttt ttttgaggct tccctatcc
11221 ctactctatt tttaattatt gcttgagggt accaaccaga acggctcaa gcaggaacat
11281 atataattat gtatcacagt ctactctcaa tgcctttttt gttttttatt ttatattgct
11341 tttattcatt aaatcgttta gagggttgtt tactaatata tagaaaaata gaaattagaa
11401 gtattgtttt ttatattatc gtattggctt ttttagtaaa gttaccatg tacgggcttc
11461 atttatgact tccaaaggct catgtggagg ccccttlaag cggatctata attttagctg
11521 gatttttact aaagcttga gगतatgggt tctatctatt aaattacatt cttctttaa
11581 aaaattttaa cactaattta gttgtcttaa ggtttttaa ttatagagga ggcctctag
11641 caactcttat atgttttct caatcagaca taaaagctat ggttgcctac tctctctag
11701 cacatataag tttagttatt gttagaatgt ttctaaatc aaactgagggt cttttatctg
11761 ctaaaattac aatattagca catggattta ctctacagc tctgtttata ttatgaata
11821 taagatataa aaaaatttat agacgaagat ttaatactt aggggggctt ttaacttatt
11881 ttctaaaa atctctttta tgalttatt ttttagaata taacatagct gctcctctt
11941 ctataaattt attaggcgaa ataatagtta tccctagctt aaagctttga ggggctattt
12001 ttgtttttt aatagtatt attatactat taagagcagg ctacaataata atattatag
12061 taaacaataa ccatggagggt ccttcgtttt atgttactcc tagaataaat tataacagggt
12121 cagaatatta cgggttatta atacatgtt tacctttttt cctattattt aaaatagaat
12181 tttttaaa ttatagtaga actactaac ccaatagaat ctagtctatt ttgatttac
12241 aaaatcatta tttttttaa atttttttg ggacaaaaat atgaacctca tcagtaaat
12301 ctacatata agaataacca tactacatct acaaaatgct aatcaagc cgcggctaaa
12361 aaccggacat ggtgggactt attaaagtgg tatagtatta agcgaattaa acaaaatagt
12421 aaaaatttag ctccaactaa acaatgtatt ccgtgaaaac ctgtagccat aaaaaaagt
12481 ctaccatata cgctatcggc aatagagaaa gaagtttctt tatattcccc aaattgaaga
12541 aataaaaaat ataaccctag taaaacagtt actagtagc ctctagtagc ggatttaagt
12601 tttcctctt caatgcatg gtgagcccat gttacgctta cccagataa aagcaataca
12661 gactgattaa gtaaaggctac tgaaaaaagta ggtagggtag taattcctat gggcgcccaa
12721 actgaccga ctcaatagc tggggctaaag ctgctatgaa agtaagccca aaaaaaagca
12781 aaaaagaagc agacttctga tacaataaaa agtataacc ctaattfaat cccgctaaca
12841 acataagttg tatgataacc ttgatatgct ctctctcgaa taatctctg tcatcagaga
12901 taagcgataa aagctgttaa tagtattcca aataatagta atactcaata attaaatoga
12961 aggtagtaaa ttaaacctaa aggcatttcta gttaaagcta gagagcttaa aagaggctag
13021 gggctaaatt ctacaagatg aaaaggagtt ttgaacatta ataataaaca aagtaaaact
13081 atttgatgag taattccttt gagcagccgc cggcagagcg ggtattattg atgttaatat
13141 atatgaaaaa attttttcgc tgccttatcg tctctatct tatagtittt tttcgaacta
13201 ttgtgtttt aggccttatt ttaagtgggt cgggtagaag ctggtttctt gttgagcgg
13261 ggatagagct atcattatc tctttcttc cggtaataaa ttatctaaa agctcaatgt
13321 ctccagaag aactattaa gacttttag ttcaatctg tcttagggca ataattctgg
13381 tttcaggatc atttttattc attgttgaaa taacaataat gttttactat aggcctttt
13441 taactggatt aatattaaag ttaggtttag ttcttttca tttttgggtt attcctgta
13501 taaaagggtg taagtttaca gacataaatt ttattttagg acctataaaa attattcctt
13561 ttttataat tattctatct attccaatca gatcaaaattt attattaatt ggtatttca
13621 tgagactatt tcttttgta attggggcat tattaggtaa caacagaaga agagtctgag
13681 gaattatagc agcttctct attgtccatt cagggtgact tctaattagg tctctagtag
13741 gaggatagg actttattt agtattatc ttatacaatt ataccattt ttgtatttt
13801 gttttataa ccaaaaattt ttatatctg tattaatatt aagttttagg ggtttacctc
13861 cttttact attttttgga aatattata ttttgcaaaa tattatttat aaagaggtcc
13921 tagtaagttt acttttaca ttattattt ttacgctct cattagacta aactctatc
13981 taaagttag ctactcttt tacttaagaa gaatatttaa ttaaaaaat aaaaatttta
14041 ttactaattt aataataatc tggagattat tagggtttat tcttttattc ttttttatg
14101 gccgagctaa gcattagatt tttaactaa agacggaaat ttattttcc

//

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 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..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="MRWLYSTNHKDIGTLYMIFGIWCGMVGTLGSLLRLELGTSGVL
FDDHFFNVVVTAAHAFVMIFFMVPVMIGGFGNWMVPLLAGPDMSPRMNNSFWLLP
PSFILLICSSMVEGGAGTGWTVYPLSSMVGHSGASVDLAIFSLHLAGMSSILGAINF
ITTFINMRSSGMTMERLSLFFVWSILVTVFLLLLSLPVLGAIITMLLTDNRNFTSFFDP
AGGGDPILYQHLLFWFFGHPEVYILILPGFGIVSHILGNASSKQPFGLGMIYAMISIG
ILGFIVWAHHMFTVGMVDVDRAYFTAATMVIAVPTGIKVFVWLMTLYGMNTKLDASMY
WVLGFIFLFTLGGLTGIVLSNSSLDIVLHDTYYVVAHFHYVLSMGAVFAIFAGFVYWF
PVMTGLLILNERLAKAQFIIMFSAVNLTFPQHFLGLAGMPRRYSYDPSYFKWNQISS
FGSLMSVFAVLFVFLIVWEAFLSQRSVAFSGAAIYSREWAENSYPDFHNSNVIPSIV
L"
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="MTTTCMALSLVSLICFLQSPVALGGLLIFLSIFLSVSVGFQMSS
WYGCILFMVYVGGLLVLFYVIMLSSNFYKASGKLLGLFLLMTSSLFMMYLLTSEFL
GKMVLFSGNECSLDLSLSLFLSLGLLLLLAFFSIVHVLLKKGKLSILVSV"
gene 3314..4972

/gene="ND5"
 CDS 3314..4972
 /gene="ND5"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 5"
 /translation="MGRLPVLLFFMLFIYLWMSFLFWSWQYETFLYEVSMLMVFVGSSKI
 TLMFIGDSISMSFSLVVMITISASVWFAMEYMSGDEFKFRFGMILFMFVLSMNLIYS
 GSVFLLLVGWDGLGISSFALIIYQSNSTSLKAGYLTLTNRIGDVLIIICSIPIFMIMSG
 NLYFFSVDHSLITLLIVLAALTKSAQYPFSAWLPAAAMAAPTVPVALVHSSTLVTAGV
 YLVRSLNLILEESVCQLLLFSGAITCILGGFSAIFENDIKKIIAFSTLSQLGLMMF
 CLGLGIPLLALLHLFTHAMFKALLFLSAGLILMSSFGTQDLRQLGSLFRSPLLVIIF
 NISSLCLVGAPFVSAFYSKHAIYEQMMMMGLNSFSMFLMMVGVILTTSWYVTRVLKALS
 WGKVSGLSVQSYLKLTSYFPLLLLFLGSIISGKMFSLLDISMTQSLFVASWFQLLLNF
 LLVFGVSMGLIWSPPKNSMFFSSMFFLWESSNKSSVFLKGILKQSKVLDYGWLEPKIY
 FEVSLFKSSFIIQKMFLGVSSYLSLIPGFSVGILIVSSCIYFW"
 gene 4954..5859
 /gene="ND1"
 CDS 4954..5859
 /gene="ND1"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 1"
 /translation="MYLFLMKSILVCLCVLIAVAFFTLERKVLGYIQIRKGPKNKVI
 MGIAQPLADALKLFIKEKISLYKSNMVFPLPLLGFFIALVLWGLSPSFFSIHFNLF
 GLLIFFCITGLNVYITSIAGWSSNSLYAFFGGIRASAQSISYEISLVMLLIFPAIISC
 SMGWLFQNMKDYWVFFLCFPLFIWLVSALAETNRAPFDFAEGESELVSGFNIEYESGL
 FALLFLAEYTCILFMATATTVWFCKSFVWLDFFIHSFIFSLFFLVARGVYPRFRYDLL
 MMLCWKSYLPFAISLLMYSLSKVF"
 gene 5860..6153
 /gene="ND4L"
 CDS 5860..6153
 /gene="ND4L"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 4L"
 /translation="MVIKLSYLFCLFCLFLIAFSAFRNQLMSNLIILEGLMLCIVFS
 VLSVAELGSKSLILMLLTFAAIEAALGLALLVNFLRSYGNDFCLNMSQFFAKN"
 gene 6140..7252
 /gene="CYTB"
 CDS 6140..7252
 /gene="CYTB"
 /codon_start=1
 /transl_table=5
 /product="cytochrome b"
 /translation="MRKISAVESMLSLPSPNLISIWWSGGSGILGMLLVQLLTGLFLS
 MHYASDIMLSFDSVIHIMRDVPAGWALRLIHANGASFFIFLYLHIGRGLYYQSYILQ
 PKTWMVGVTIYLLSMATAFLGYVLPWQGMSFWGATVITNLLSAVPYFGGSLVEWVWGG
 FSVGQATLNRFFSLHFILPFIMTVFILIHILFLHDKGSTNPLGHNYHLNKNINFHPYFT
 WKDMVGFVLLTLITICCFAPYVLSDPENFIYANPMLTPTHIQPEWYFLFAYAILRS
 IPSKLGIVIALAMSIFILYFLGIGSMKINIPSSFNLLYQVWFWVLVNVNFMLLTWLGAC
 PIENPYLLLPLTYLYFFSYPIILLSWLSFMVLK"
 tRNA 7253..7318
 /product="tRNA-Asp"
 tRNA 7319..7383
 /product="tRNA-Cys"
 tRNA 7384..7449
 /product="tRNA-Phe"
 gene 7450..8141
 /gene="COX2"
 CDS 7450..8141
 /gene="COX2"
 /codon_start=1
 /transl_table=5
 /product="cytochrome c oxidase subunit II"
 /translation="MSFWGAMNLM DPSSPIQSEMILFHDHAMILLIGIFVLVACLGVK
 LCVNQVSSRTTHEAQLLEILWTIIPAFLLVWLALPSLRLLYLLDEQSSNGLILKATGH
 QWYWSYEIPSIGISSFDSYMIQENSLQNGEFRLLLEV DNRPILPYNMTINTLTSTDVI
 HAWAIPSMGIKMDAVPGRNMMGIQSFSPGVFYGCCEICGANHSYMPIVVEFLSLED
 FINFFGIFCLSG"

tRNA 8114..8176
/product="tRNA-Tyr"

tRNA 8172..8235
/product="tRNA-His"

tRNA 8248..8311
/product="tRNA-Trp"

tRNA 8314..8375
/product="tRNA-Gly"

tRNA complement(8376..8435)
/product="tRNA-Gln"

tRNA complement(8431..8499)
/product="tRNA-Leu"

gene complement(8493..8651)
/gene="ATP8"

CDS complement(8493..8651)
/gene="ATP8"
/codon_start=1
/transl_table=5
/product="ATP synthase F0 subunit 8"
/translation="MPQLSPGSGMLTLLIFFLMSFLFTLLPFKPMMPKTIKMKSNFS
YLSTLIFK"

tRNA complement(8653..8718)
/product="tRNA-Asn"

gene complement(8725..9372)
/gene="ATP6"

CDS complement(8725..9372)
/gene="ATP6"
/codon_start=1
/transl_table=5
/product="ATP synthase F0 subunit 6"
/translation="MMTDLFSSMDGNHSIVIWAMPLVMTVIFISNSIFLTTSMSPITA
TITNLWESSDYFTRKSLTSLMLFHLNLFGLVPLVFSSTTSLVWNSSLALICWGT
ILVSGWVFSPOKSAHLAPSGAPGMLMPFLILLETISILIRPITLTVRLVANISAGHI
ILALLANVLSSSLIFSWSTCVSIMIFYLFFEMFVSIQA YIFLLISLYMAEHP"

tRNA complement(9385..9436)
/product="tRNA-Arg"

tRNA complement(9444..9511)
/product="tRNA-Glu"

rRNA complement(9523..10247)
/product="12S ribosomal RNA"

tRNA complement(10249..10313)
/product="tRNA-Met"

gene complement(10319..10672)
/gene="ND3"

CDS complement(10319..10672)
/gene="ND3"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 3"
/translation="MLYYSIFPGVLSVALIMLFFITYWSPIFSEDSEKSSPFECGFDP
MSNMRKPFSLRFFMLVILFLVFDVEVLLFPLLMLMNMAMSSTFLTLLIFLLALLG
LIYEWAMGALEWSVN"

tRNA complement(10684..10753)
/product="tRNA-Ser"

tRNA 10751..10805
/product="tRNA-Ser"

gene 10806..12128
/gene="ND4"

CDS 10806..12128
/gene="ND4"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4"
/translation="MISTLVFLSLASIFMFSWDMVLLSCFMMVLISFVNMHMNFSNLN
ESSVFMIESFNFLIFLFSVVIIFLSLTNSFWLKINSFSKVVMSLNLFLVLAFSWKDLL
GFYFFESSLIPTLILIMVWGYQPERLQAGTYMMLYTVFASLPLMLILYFFNKSSDL
SIYNFKLLGLEMSTFMLMIFVMAFLVKLPFSAHLWLPKAHVEAPLSGSMILAGVLLK
LGGYGLYLTNQCFNFSSLNLMVSVISFSLSMWGGFAALMCLQQTDMKAMVAYSSVAHM
SLVISGIFLNSVWGEYCAKVTMIAHGFTSSALFVLVNMSYKLLSRSFLTAGLLGMY
PKMSFLWFLFCINMAVPPSYNLVGEMLMLPSLYIYSVMLVIIMGLVMFFSATYNMFL"

tRNA YTQINHGMANPLINSGSAYYSADYLGLFLHLLTFIMLLKLNLFSTEIC"
 complement(12134..12193)
 /product="tRNA-Thr"
 gene complement(12184..12972)
 /gene="COX3"
 CDS complement(12184..12972)
 /gene="COX3"
 /codon_start=1
 /transl_table=5
 /product="cytochrome c oxidase subunit III"
 /translation="MIKTPFHLEVEFSPWPILGSTLLTCMPLGLVYYLRLHSINLMIFG
 IATALVAYLWWRDIVRESTYQGFHTSYVISGLKLGVFLFIVSEVCFFFAFFWAYFHS
 SLAPAVEIGSSWPPYGVSTLATAFQVPLLNTSVLLLSGVSVTWAHHAIEESKYLSAIYS
 LIFTCILGFYFLFQGEYEETSFSIADSVYGSTFFFIATGFHGLHVAVGATFLMVCLM
 RMYLYHFNSNHHVGFLLAAAWYWHFVDVWVWFLYVSIYWWGSFQK"
 tRNA 13019..13087
 /product="tRNA-Ile"
 gene 13089..14015
 /gene="ND2"
 CDS 13089..14015
 /gene="ND2"
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 2"
 /translation="MTVFSWMFLSLVFLGPVVSLSSSSWFLVWGGMELSLIGLLPLMS
 NDKSNIMTESAFKYFLVQALASLIVLISGINIFMFNSNSMMILGFFLIGLTLKLGIFP
 LHFVWIPVLKTSGYNLMMLLLGPMKIVPLSLLMKFFQVFNISNFSVLWFAVFSALIGA
 IIGNNVSSIRAVLGASSISHGWFVVGAWSSGMWSYFFIYLISLFLMLWYLEESFSKV
 TALLILALSGLPPFLFFGKVNILLSFLTIVNASLLFLVILHSAVLSNFYLFKFSYIF
 YLKEKKLGFTGRWNWFLFLSAWLVGAFVLLF"
 BASE COUNT 4585 a 1728 c 2181 g 5584 t
 ORIGIN

1 ttgcgatgat tgcattcaac aaatcataaa gatattggta ctttatatat aatttttgg
 61 atttgatgtg ggatagtagg aactggactt tcccttttaa ttcgattaga aftaggaact
 121 tctgtgtgac tgtttgatga tcattttttt aatgtttag tctactccca tgcctttgt
 181 ataattttt tcatagtcac gcctgttata atgggtggat ttggaattg aatagtgcc
 241 ttataattg gtgcgccaga tataagtitt ccacgaatga ataacataag attttgatta
 301 ttacctcctt cattttttt acctatttct tcaagtatag ttgaagggtg tgcctggaaca
 361 ggatgaacag tctatcccc tttaagtagg atagtgggcc atagtgggtc atctgtcgat
 421 ttactattt ttccattaca tttagctggt atatctcaa tcttaggtgc tattaattt
 481 attactaca ttttaatat acgctcttct ggtataacga tagaacgatt aaggttattt
 541 gtagatcaa tcttagtaac tgcatttctt ttactttat cttacctgt tttagcagg
 601 gcaattacta tacttttaac agaccgaaat tttaacta ggtttttga tccagctgga
 661 ggtggtgac caattttata tcaacatctt tttgatitt ttggacacc tgaagtatat
 721 attttaatt tacctggatt ttgaaattgt tcccatatc taggaaatgc atcatctaaa
 781 caacctttg gtagggttagg aataattac gcaataattt ctattggaat tttaggattt
 841 atgtatggg ctcatcatat atttactgta ggtatagatg ttgacactcg gcctatttt
 901 actgctgcta ctatagttat tgccttccct actggaatta aagttttcag ttgattaata
 961 acattatatg gaataaatac aaaattgat gcgtcaatat actgagtttt agggtttatt
 1021 tttttgta ctttaggagg attaactggt atgtttttt caaactctc tttagatatt
 1081 gtgttgcacg atacttata tgcagtgcca cactttcact atgtttgtc aatggggggc
 1141 gttttgcta ttttgctgg attgtatat tgattccctg tgataacagg attaatitta
 1201 aacgaacgat tagctaaagc tcaatttatt attataattt cagcagtaaa ctaactttt
 1261 ttctctcaac attttttagg tttagcgggt atgcctcgtc gatattctga ttaccagat
 1321 agataattta aatgaaatca aatttctct tttggtctt taatatcagt tttgcggta
 1381 ttatttttg ttttaattgt ctgagaagct ttittaagcc aacggtcagt tgcctttagt
 1441 ggtgcagcta ttattctcg agaattgact gagaattcct atcctcctga tttcatagc
 1501 aatgtcatcc caagtatttc tgcattgtaa attacactat agtacaattt taagtatatg
 1561 ttaattacga taacatgga ttttttaat tagtgatta ataataaaa tgcctatagt
 1621 ttttaattaag ttaacttaaa aagggttata aatttttga aggaggtcct tacttttgc
 1681 ataaggttt aattataata tttttgata aattttccc aacaagaaag atctataaac
 1741 taagtatttt taggtataaa tattgttgt gaataaacia tcaaaaatta tttataggg
 1801 gcgaaacacc tacaattttt ttgataactg gatttaattt aaatgtattt agtacattaa
 1861 gacagacaat aagtataag atttattgta atgaaattt taattttaa aattttggag
 1921 aaataaaatt tcaaaaattt ttaagagta taataaatct ctatagaaa tcttattta
 1981 aattaaaatt ttaagattt agttttgaaa ataaattatg tttaaatgag taataataaa
 2041 caatatattt taaggaactc ggcaaacata tttttcacct gtttaacaaa aacatagctt
 2101 taagattttt cattaaagt ttacctgcc cgggtggaatt tcttaacggc cgcagctatt
 2161 tgcactgtct aaggtagcat aataaattgg cttttaatta aagtcgggta tgaaggaggaa
 2221 ttggaaaaat aactgtctca agtatatttt attttaattt aattaatgag tgaaaattc

2281 catattttat tcattagac gagaagacc taagagttat aattaaatg aatttttfc
2341 atatacatt ttgttgggg caacaaaagac tggatagaac cgtcaaaaag taattttaca
2401 taaccaataa tatttaaaaa gaaataaact accttaggga taacagcata attaatttt
2461 agatttggac ctcgatgtg gattaaggac ttagttact agccgtaaac taagatagt
2521 ctgttcgaac ttagtagct tacatgatct gagttcagac cggcgtaac caggtcagt
2581 tctatctttt atatttaata ttctagtacg aaaggaccgt atatttatag atactatft
2641 gtttactgg cagatttaa tgcgattgat ttaggttcaa tatatacttc atgtagtaa
2701 attttaaaa gaagttaaat aaaattaata ttaactttgg gagttaaaga atgaggatac
2761 tcacttttgg tattatact taacagaag atctgcgttg caagtagtta gtgggataat
2821 cctttacca aactaatg atactatct gaaatgacta ctactgtat agccttaagt
2881 ttgtaagac ttatttttt ttacaaga cagtagctt taggggggtt attaatftt
2941 ttaagaattt ttatctgt aagagttgga ttcaataaa gaaggtgata tggatgtat
3001 ttatttatag ttatgtggg gggccttta gttttatt tatagttat tataitaaga
3061 agaaattttt atttaaaagc ctctgaaaa ttattaggtt tattttttt aataactct
3121 agactgtta taatatactt aacaacatct gaatttttag gcaaaaatag tttagggtt
3181 tctggaaatg aatgcagatt agatttaagc ttaagattat tttaagttt agggctatta
3241 ctattattag cattttttc cattgtcac gttgtccttt taaaaggaa aagaatttta
3301 gttagtgttt aaaataggtc gacttccgtt attattatc ttatacttt ttattttt
3361 gtgaataaga tttttttt ggtcttgaca atagaaact ttctttatg aggttagtct
3421 aatattttt ggttctcta aatctactt aatattttt ggtgattcaa ttagaataag
3481 attcagatta gtcgtaataa caatttcagc atctgtttt tgatttgcta tagagtatat
3541 atctggagat gaatttaagt tctgtttgg gataattct ttataattg ttctatcaat
3601 aatattttt atttactcag gatcagttt ttacttcta gttggatggg acgggttagg
3661 aatttctcc ttgctctaa ttatttacta tcaaaagaa actagattaa aagctggta
3721 tttaacttta ttaacaaatc gaattgggga tttttaatt atttcagga ttccattat
3781 aattataaga ggaaatttat attttttt ggttgatcac tctttctaa ttactttt
3841 aattgttta gctgcttaa ctaaaagggc tcaatatcca ttcagagctt gactccggc
3901 agccatgctt gcccacact ctgtaagggc tttagtacac tcttactat tagttacagc
3961 aggggtttat tttagtttc gtttaagttt aaactgatt ttagaagagt ctgtttgca
4021 gttgtgctt ttctgggg ctattactg tatttttaga ggatttaggg ctattttga
4081 aaatgatatt aagaaaaa ttgcctttc tacattaagg caattaggtt taataatatt
4141 ttgctaggg ttaggattc ccttattggc attacttcat cttttacc atgcaatatt
4201 taaagcttta ctattttaa gagctgctt aatttaata tcaagcttg gcacacaaga
4261 ttactgtag ttaggatctt tatttttcg aagaccttta ttagtgtta tcttaaat
4321 tagaagattg ttttagtag gagcccat ttagaagctt ttatttcta agcatgcat
4381 ttatgaacag atgataataa taggttaaa ttcttttag atactctaa taatagtagg
4441 cgttaattta acaagatgat atgtactcg tttttaaa gcttaagat gagggaaagt
4501 taggggatta agagttcaat ctatttaaa attaacaagc tacttccat tattattat
4561 atttttaggg tcaatcatta gaggaaaaat gttcagatta ttgatatta gtataacaca
4621 aagcctattt gtagcatctt ggtccaact actactaac ttcttttag ttttggagt
4681 aagaatgggg ctaatttgat cttcacctaa aaatagcata tttttcta gaattttt
4741 ttatgggaa agatccta aaagtctgt atttttaaa ggaatttaa aacaaggaa
4801 agtttttagat tatgggtgac tagaaccaa aattatttt gaggtttc tttaagag
4861 tagttttat atcaaaaaa tattttggg ggttaggaga tacttaagaa gattaattcc
4921 aggttttaga gttggtatc taattgtaag aaggtgtatt tattttgat aaaaactat
4981 ctgtatgtt tatgtgtgt aatcgtgta gctttttta cttattaga acgtaaggt
5041 ttagggtata ttcaattcg aaaaggccca aataagttt ctattatggg aattgctcaa
5101 ctttagctg acgctttaa attatttat aaagaaaaa ttgacttta taaactaat
5161 aggatagat ttttttaac cctttatta ggaatttta ttgcttagt ttgtgaggg
5221 ttaagaccaa gattttttt aattatttt aactatttt gttacttat tttttgt
5281 attactggat taaatgttta tattacaaga atggcgggtt ggtcctcaa tagttgtat
5341 gctttttg gtggaattcg agctcagct caaagaattt ctatgagat tagattagta
5401 atactattaa ttttccagc tattattct tttcaatag gctgattcca aaataaaaa
5461 gattattgag tattttttt atgtttctt ttactttta ttgctgtt aagagcatta
5521 gcagaacaaa accggggccc atttgattt gctgaagggg aactgagct agtatcagga
5581 ttaaatattg aatagaaaag tggctgttt gctttactt ttggctga atacttgc
5641 attctattta tagcaacagc cactacagtt ttatttga aaagattttg agttttgac
5701 ttattattt ttgatattt tttagctg tttttttag ttgcccagg agtttatccc
5761 cgtttctgt atgatctat aataactc ttttggaaa gatacttcc ttgtctatt
5821 agactattaa ttagtagact tttagtaaa gtttttaaa tctgattaa actaagata
5881 ttattttgt gttttgctt atttctatt gcttttcgg cttttcgtaa ttagttaata
5941 tcaatttaa ttattttaga aggttaata ttaatttga ttgttttag agtctcaga
6001 gtagctgaat taggatcaa aagtttaatt ttaactctt taacatttc gctattgaa
6061 gcagcttag gtttagctt actagtaaat ttttaccgt catatggaaa ttattttgt
6121 ttaaatataa gtcaatttt tgcgaaaaat tagtgcagt gaaagaatgc ttgattacc
6181 tccccatta aacattagta ttgatgaag agggagctca atcctaggaa tattattag
6241 ttccaactt ttgactgtt tatttttc tataactat gcaagagata ttatactat
6301 attgactca gtaattcaca ttacgaga cgttctgct ggttagcccc ttcttaaat
6361 tcatgctaat ggggtctt tttttttt ttttttat ttacattg gtcggggatt
6421 atactatcaa agatatttt tacaacaaa aactgtaata gttggagta caatttact

6481 gttagaatg gctacagcat tcttggata tgtttacca tgagtcaga tatcatftg
6541 aggtgcaaca gtaaitaca atctctctc agcagtcaca tattttggag ggtctttagt
6601 agaattgagtt tgaggtggat tttctgttg tcaagcaacg taaatcgat tttttctc
6661 tcattttatt ttaccttta ttataacagt gttatttta attcatttaa tttttctca
6721 tgataaagg tcaactaac cactgggaca taattaccat taaataaaa ttaatttca
6781 tcctatttt acttgaagg atagtagg atttgattg gttttattaa cttaattac
6841 aattgttgc ttgtccat acgtttgtc agatccagaa aattttatt atgccaatcc
6901 aatactgacc ccaactcata ttcaacctga atgataatt ttattgcat atgctatttt
6961 acgatcaatc ccatcaaat taggtggagt tattgctta gctataagaa tttttactt
7021 atatttctg ggtattggat ctataaaaat taatatccct tcaagattta atctttata
7081 tcaagatga ttttgatt tagtagtaaa tttatactt ttaactgac ttggggcttg
7141 ccctattgag aatccatacc tcttttagc tggaccatta acctatctat attttttag
7201 ctactctatt ttaataagt ccaggtggtt gagattata gtttataat agctttgtt
7261 aatttaattt aaaaatttg ctgtcaagt aaaaaacaca tgaagtgtt aaaaaataa
7321 ggtaatagt tatataaaa tatataatt gcaaaattat agatgaaca agttctacc
7381 ttaattcag tagctaaat aaagcgtggc ttggaaga ccaaaataa gtttcgatt
7441 ttgaattaa tgagttttg aggtgcaata aattaatag atcctaggc tccaattca
7501 tcagaaataa tttatttca tgaccatgct ataatttat taattggaat tttgtttg
7561 gtacatggt taggggtaaa actctgtgtt aatcaagtt catcacgtac gaccatgaa
7621 gctcaacttt tagaaactt atgaactat attccagctt ttttattag ttgattagct
7681 ttaccaagct tacggctct ctactctta gacgagcaga gtagtaatgg attaatfta
7741 aaagccactg gacatcaatg gtattgaaga tatgaaatcc caagaattgg aattagatca
7801 ttgattct atataatca agaaaattct ctcaaaatg tgaatttcg ctacttga
7861 gttgataacc gtcctattt accataaca ataaacta acaccctac aactagaaca
7921 gatgtattc atgctgagc tattcttca ataggaata aatagatgc cgtcccagga
7981 cgattaaata taataggtat tcaactttt agaccaggag tattttatgg tcaatgtct
8041 gaaatttgg gagtaacca ctctacata ccaattgtt ttgatttct ttcatagaa
8101 gatttatta atttttcgg aatctttgt taaaggga atttgaag ttattttag
8161 atcttccct gaagtftaa ggtttaaact gctattttt ggaataagag agtcaacta
8221 tagttgaaa acttcttaa ttttttagt gtacataagt taaataaact gatgacctc
8281 aaagcattt atcttaaat agattactc gtatacttc tagtataat agtacatcg
8341 cttccaagc gaaaagcctt aaataaggta agtaaatftt tataagattt cttctttaa
8401 gcctcaaatg ctctgtgccc taacaccgat aaaaaaaga gcatagctca taaaagcgcg
8461 taaaacctt tgcatttat tgcctataa aattattaa aaattagcgt agataaatc
8521 ctaaaattg atttatttt aattgtttt atggtagat gtttaaggg aagtaagta
8581 aatgaaaac tattaagaa aaaaattaa ataatgta taaacctc tctggacta
8641 agttgagga tatctaaact actagaaaa gtactttta ataacattt aatgcttaa
8701 attttaagct aagtttagat tttttatgg gtgtcggct atataaaggc taattagaag
8761 agtaaaaatg tatgctgaa taaatgatac aaatattca aataaatag taaaagatt
8821 aattgaaaca caagtgtac atgaaaaaat tcttaagaa gatcttaaa cattagctaa
8881 gagggctaaa ataatgtcc cagcactaat atttgcaact aatgcaactg ttaagtaat
8941 aggcagaatt aaaaactaa tagtctcaat taaaataaa aatggtatta gtatcccagg
9001 agccctgaa ggggctaaat gtctgcaga ttttgggc gaaaaaact atctgaaa
9061 taaaatagt cctcagcaaa taaagcaag gctctattt actcataatc tgggtgaga
9121 tctaaact agaggcacta accccaaaa atctttaa atataata atattagaga
9181 agttaataat gacttcgag taaaataagt atctgaagc tccataagt ttgtaatag
9241 tgcgttaatt ggagatattc tgggttaa aaagattgaa ttgaaataa aaataacat
9301 tataactaa ggtattgctc aaattacaat agagtatta cctctatag atgaaataa
9361 gtctgtatc atttgataag aatgtaagat atacctta aattttaag gccgaaact
9421 aaaaacgaaa cgttcttc ttgggttatt aatgaaatt tcatattgt attacgaaa
9481 aatactgtat taagtttat actatagtaa catgaattg ttttgcagt aaatgcaag
9541 aagcaattc cattactct actttgtac gactgtctc attttaac gagaatgac
9601 ggcgattgt gcactactaa aaaaatatt agattttatt tttattaa atcttactt
9661 caagtcate tcaaaatt agtacttaa tttatctga taaaataat atgtaactc
9721 gccttaatta aaactataag ttgcacctg atctgtctta ttgattact ttaaaaaat
9781 ttttaaca aattttaga cgacggcata caaacttaa agttttatg tttacttg
9841 tggattcaa ttattagga agttcccctg atttttta ttgcccga taacttga
9901 gtttatat taatattag actactatc aaataaggtt gaatatctag aataatagg
9961 tatctaacc tagtttctt tctagctta aataaaaa atagaagcaa aataaaaaa
10021 gaatatatt cactattat caaagtttg tctataaaa tctatttga ttattgta
10081 ttgttttt tataggtctg accgcggtg ctggcaccta taaacaaa aaagaataa
10141 tttactaagt taacattct gttatttatt aaaaattata actggtctt acaaaattc
10201 cataattata taaacattaa actaaattc tttatcaagt taggtaaatc taaaaagat
10261 tactctatt ttgattatg agcccaacag cttaaaaatt agctatttt tagataatt
10321 aattaacact tcaattcaa gctctatgg ctactcata aattaacct agtaaaagg
10381 ctaaaagaaa aattaagagt gtcaaaagta aaaaagtga tgatattga tttataaca
10441 ttaagagagg gaagagcaaa actactcaa cgtcaaaa taaaataaa atfactaata
10501 taaaaaacg gagtgaaaaa ggctacgca tattcttat tggatcaaaa ccactcaaa
10561 aagcgatct ttttctca tctctgaaa agatagggtc tcaagtaata aaaaaaaa
10621 gtataattaa tctacacta agaactcctg gaaaaattga gtagtatagc aatttagtga

10681 gctcactaaa tgatataaaa tatatctaca aagatttca aaatccttta ttatataaaa
10741 attcatgcta agaagatata atfgaggctg ctaactttga ttgggtttt tggatccgct
10801 tcttttgat ttctacttta gttttttaa gtttagcaag aatttttatg tttaggtgag
10861 atatagtttt attaagatgc ttataaatag ttttaattag atftgtaat atacataaa
10921 attttagatt aaatttagaa agaagagat ttataattga atcattta atggttttaa
10981 ttttttaag ggtagttatt attttttgt cattaacaaa tctttttga ttaaaaaa
11041 ataggtttag gaaagtagta atacattaa actgtttct agtttagct tttagtgaa
11101 aagatttatt agggttttt ttttttctg aatctcttt aatccgaca ttaatttaa
11161 ttatagatg aggggtaccag ccagagcgcg tacaagcggg aacttatata atactttaca
11221 cagtattgc gctttacc ttactaatat taacttata tttttta ataggagag
11281 attaagaat ctataattt aagtattag gattagaa atgactttt atattaata
11341 tttttgat agcttttta gttaaattac ctattttag agctcattg tgacttcaa
11401 aagcacatgt agaagctcct ttaagaggt ctataattt agctggagtt ttattaaat
11461 taggaggata tggctttat ttaacaaacc aatgtttaa ttttagctct cttaattaa
11521 taagagtaag gattaggttt ttaagaatat gaggaggatt atftgcagct ttaatgtgcc
11581 tcaacagac agatataaaa gcaatagttg ctactctc agtagctcat ataatgttag
11641 tttagtgg aatttttta aactctgtat ggggggaa ttgtccaaa gttaccataa
11701 ttgccatgg tttacttca tcagctttgt ttgtttagt aaatataaga tataaaaaat
11761 tacttagtcg gaggtttta acagcgggag gattactagg tatatatcca aaaaatctt
11821 ttttaggt tctttttgt agaattaata tagctgtccc gccatctat aattagtag
11881 gagagttaat acttctcca tctttata ttactcagt aatgttagtt atcttatag
11941 gtcttgaat atttttaga gctacttata atataattt atatactag ataatcacg
12001 ggatagctaa tctttaatt aattctgaa gggcttata ttcagcagat tatttagggc
12061 tafttttca tctttaact ttattatat ttaaaaa atattttt tcaacagaaa
12121 tctgttaaat ggttccaaag gaaaaattca ttgttattt acaagatcaa tafttttt
12181 aaftacttt tggaaagaac ctaccagta aattctacg tataagaata atcaaacac
12241 gtcaacaaa tgcaaatc aagctgcagc taaaagcct acatgggtt tgaattaaa
12301 atgataaag tatattcgt taagacatc ttttaaaat gttgtccaa cagcaacatg
12361 gagtccgtga aaactgttg caataaaaa agtcttcca taaactctat cagcgataga
12421 aaatgatgt tcttcatatt ctcaaaattg gagaataag aaataaaacc ctaaaata
12481 agtaaaaaat aagctataaa ttccggataa atftttctt tctcaatag catgatgagc
12541 tcatgttacc ctactcctg ataataaag aactctagta ttaagaggg gaacttgaaa
12601 agttgcaaga gttcttac cgtaagggtg tcatctagat ccaatttca cagcaggtgc
12661 aaatgttcta tgaataaag ctcaaaagaa agcaaaaaaa aagcaaaact ctgatacaat
12721 aaataaagaa actcctaatt ttaaccact aattacataa gatgtatgaa aacctgata
12781 tgttcttct cgaacaatat ctctcatca taaatagca accagtcagc ttgcaataat
12841 tcaaaaaat attaaattaa ttaggtgtaa tctgaagtaa tatactaac ctaaaagcat
12901 acaagtttag agagttgat ctaaaattg ccatggctta aattctata atfgaaaagg
12961 agtttaac atataat atataat ctattttt ttattagaa gtttctaaa
13021 gcaaccgcc ggagagcggg tatcttgat gttgatalat atgaatatt taatttctg
13081 ttgctataat gacagtttt aggtgaatat ttttaagt ttgattttta ggtcctgtg
13141 tttagattc aagatcaaga tgaattttg ttgaggggg aatagaacta tctttaattg
13201 gattattacc ttaataagc aacgataaaa gaaatattt aacagaaaaga gcttttaaat
13261 attttttag tcaagcttta gctagttaa ttgtttat tagtggatt aacatttta
13321 tatttaatg aaataggata ataattctg gttttttt aattggtta actttaaat
13381 taggtattt tctttacat ttttagtga tccagttt aaagactagc ggttataat
13441 taataatatt gttactagg cctataaaa ttgtacctt aaggttata ataaaaatt
13501 ttaggttt taacattaga aatttagag tactatgatt tgggtcttt agtgactaa
13561 ttgggcaat tattgtaat aatgttagaa gaattcgagc agttttagg gcttctcta
13621 tttctatc tgggtgatt atagtaggag ctggagagg agggatgtg tcttatttt
13681 taattactt aattagccta ttttaatat tatgatttt agaagaatc tttagaaaag
13741 taacagcatt gtttaattt gcattaaggg cccttcccc attctgtt tttttggaa
13801 aagtaaacat ttattgaga ttttaacag taaatgcctc ttgctttc tttagtaatt
13861 taattattg ggcagttct aggttaaac tctactgaa gttcagttt atttttatt
13921 taaaggaaaa aaaattagga ttacagggc gctgaaattg attttttt ttaagtctt
13981 ggttagtggg agcttttta gttttgctat ttttaaaagt ttttatggc tgagtataa
14041 gcaataaatt ttaatttaa ttacaaggtt tctcttt

//

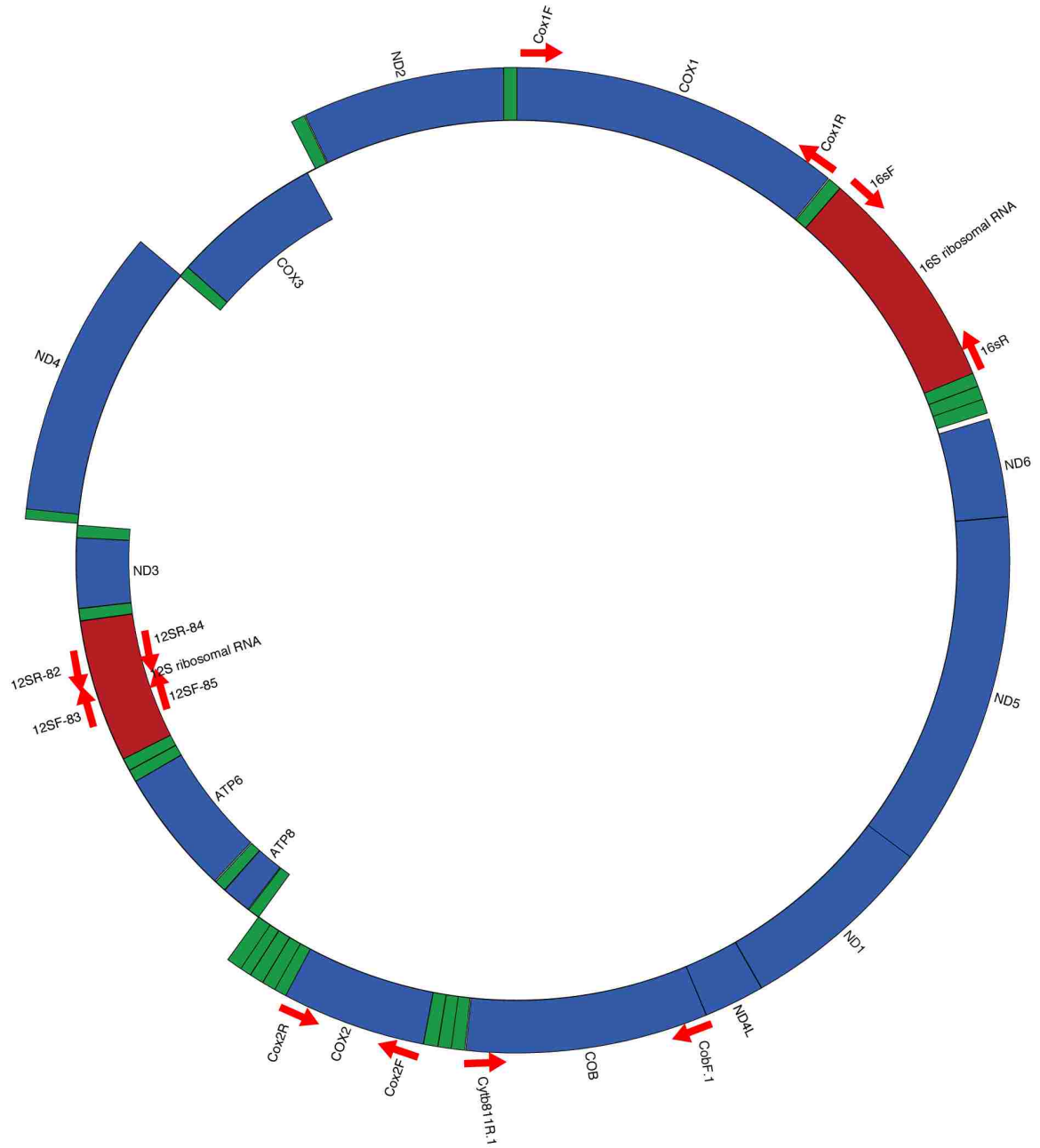
Appendix C: Section 1

12s universal primers designed in house from NCBI GenBank sequence alignment. *cytb* primers based on primers from vonProschwitz *et al.* 2009. *Cox1*, *cox2* and *16S* primers from the work of Gittenberger *et al.* 2004, Hugall *et al.* 2002, and Nekola *et al.* 2009. For relative locations and orientations see Appendix C: Section 2.

Primer name	Primer sequence
12sF-85	5-GACCGCGGTTGCTGGCACCTA-3
12sR-84	5-AGGTGCCAGCAACCGCGGTCA-3
12sF-83	5-AATTATATGGTAAGTTCCCCTAA-3
12sR-82	5-TYAAACTTAAAWRTTATGGCGG-3
CobF.1	5-CATTTTGAGGTGCAACAGTAATTAC-3
Cytb811R.1	5-GCRTAYAGAAAGTATCATTCWGG-3
Cox2F	5-AAATAATGCTATTTTCATGAYCAYGC -3
Cox2R	5-GCTCCGCAAATCTCTGARCAYTG -3
Cox1F	5-GGTCAACAAATCATAAAGATATTGG-3
Cox1R	5-TAAACTTCAGGGTGACCAAAAAATCA-3
16sF	5-GCGCTGTTTATCAAAAACAT-3
16sR	5-GCCGGTCTGAACTCAGATCAT-3

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 μ l of sodium acetate
2. Add 30 μ l of ice cold 100% ETOH
3. Invert 4X
4. Set in dark 4 $^{\circ}$ 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 $^{\circ}$ C, for 30 minutes
7. Invert into paper towel and flick GENTLY
8. Add 35 μ l of ice cold 70% ETOH
9. Centrifuge at 13,000 rpm, 4 $^{\circ}$ C, for 15 minutes
10. Invert into paper towel and flick GENTLY
11. Vacuumfuge at 45 $^{\circ}$ C for 10 minutes
12. Submit to MBF for sequencing

References

- Abascal, F., Zardoya, R., Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations *Nucleic Acids Research*, 38, W7-W13.
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Saunders, N. C. (1987). Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics*, 18, 489-522.
- Avise, J. C. (1991). Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annual review of genetics*, 25, 45-69.
- Bandyopadhyay, P. K., Stevenson, B. J., Cady, M. T., Olivera, B. M., & Wolstenholme, D. R. (2006). Complete mitochondrial DNA sequence of a Conoidean gastropod, *Lophiotoma (Xenuroturris) cerithiformis*: Gene order and gastropod phylogeny. *Toxicon*, 48, 29-43.
- Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, 27(8), 1767-1780.

- Boore, J. L., Medina, M., & Rosenberg, L. A. (2004). Complete Sequences of the Highly Rearranged Molluscan Mitochondrial Genomes of the Scaphopod *Graptacme eborea* and the Bivalve *Mytilus edulis*. *Molecular Biology and Evolution*, 21(8), 1492–1503.
- Conant, G. C., Wolfe, K. H. (2008). GenomeVx: Simple web-based creation of editable circular chromosome maps. *Bioinformatics*, 24(6), 861-862.
- Grande, C., Templado, J., Zardoya, R. (2008). Evolution of gastropod mitochondrial genome arrangements. *Bmc Evolutionary Biology*, 8(61).
- Groenenberg, D. S., Pirovano, W., Gittenberger, E., & Schilthuizen, M. (2012). The complete mitogenome of *Cylindrus obtusus* (Helicidae, Ariantinae) using Illumina next generation sequencing. *BMC Genomics*, 13(114).
- Gittenberger, E. P., W.H., Groenenberg, D.S.J. (2004). The Pleistocene glaciations and the evolutionary history of the polytypic snail species *Arianta arbustorum* (Gastropoda, Pulmonata, Helicidae). *Molecular Phylogenetics and Evolution*, 30, 64-73.
- Hatzoglou, E., Rodakis, G. C., Lecanidou, R. (1995). Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea*. *Genetics*, 140(4), 1353-1366.

- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, 17, 754-755.
- Hugall, A. M., Craig, Moussalli, Adnan, Stanistic, John. (2002). Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *PNAS*, 99(9), 6112-6117.
- Jex, A. R., Hall, R. S., Littlewood, D. T. J., & Gasser, R. B. (2010). An integrated pipeline for next-generation sequencing and annotation of mitochondrial genomes. *Nucleic Acids Research*, 38(2), 522-533.
- Jørgensen, A., Kristensen, T. K., & Stothard, J. R. (2004). An investigation of the "Ancyloplanorbidae" (Gastropoda, Pulmonata, Hygrophila): preliminary evidence from DNA sequence data. *Molecular Phylogenetics and Evolution*, 32, 778-787.
- Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451-1452.
- Lowe, T. M., Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of

transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25(5), 955-964.

Lynch, M. (1996). Mutation accumulation in transfer RNAs: Molecular evidence for Muller's ratchet in mitochondrial genomes. *Molecular Biology and Evolution*, 13(1), 209-220.

Misawa, K., & Nei, M. (2003). Reanalysis of Murphy et al.'s data gives various mammalian phylogenies and suggests overcredibility of bayesian trees. *Journal of Molelecular Evolution*, 57, S290-S296.

Moriyama, E. N., & Powell, J. R. (1997). Synonymous Substitution Rates in Drosophila: Mitochondrial Versus Nuclear Genes. *Journal of Molelecular Evolution*, 45, 378-391.

National Center for Biotechnology Information: Basic Local Alignment Search Tool. (2012). < <http://www.ncbi.nlm.nih.gov/> >

Nekola, J. C., & Coles, B. F. (2010). Pupillid land snails of eastern North America. *American Malacological Bulletin*, 28, 29-57.

Nekola, J. C., Coles, B. F., & Bergthorsson, U. (2009). Evolutionary pattern and process within the *Vertigo gouldii* (Mollusca: Pulmonata, Pupillidae) group of minute

North American land snails. *Molecular Phylogenetics and Evolution*, 53(3), 1010-1024.

Platt, A. R., Woodhall, R. W., George, Jr., A. L. (2007). Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *BioTechniques*, 43(1), 58-62.

Rambaut, A. (2012). Se-AL <<http://tree.bio.ed.ac.uk/software/seal/>>

Ronquist, F., & Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572-1574.

Schattner, P. B., Angela N. ; Lowe, Todd M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, 33, W686-W689.

Serb, J. M., & Lydeard, C. (2003). Complete mtDNA Sequence of the North American Freshwater Mussel, *Lampsilis ornata* (Unionidae): An Examination of the Evolution and Phylogenetic Utility of Mitochondrial Genome Organization in Bivalvia (Mollusca). *Molecular Biology and Evolution*, 20(11), 1854–1866.

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.

Tongkerd, P., Lee, T., Panha, S., Burch, J. B., & Foighil, D. O. (2004). Molecular phylogeny of certain Thai gastrocoptine micro land snails (Stylommatophora : Pupillidae) inferred from mitochondrial and nuclear ribosomal DNA sequences. *Journal of Molluscan Studies*, 70, 139-147.

von Proschwitz, T., Schander, C., Jueg, U., & Thorkildsen, S. (2009). Morphology, ecology and dna-barcoding distinguish *Pupilla pratensis* (clessin, 1871) from *Pupilla muscorum* (linnaeus, 1758) (pulmonata: pupillidae). *Journal of Molluscan Studies*, 75, 315-322.

Wade, C. M., Mordan, P. B., & Naggs, F. (2006). Evolutionary relationships among the Pulmonate land snails and slugs (Pulmonata, Stylommatophora). *Biological Journal of the Linnean Society*, 87(4), 593-610.

White, T. R., Conrad, M. M., Tseng, R., Balayan, S., Golding, R., de Frias Martins, A. M., & Dayrat, B. A. (2011). Ten new complete mitochondrial genomes of pulmonates (Mollusca: Gastropoda) and their impact on phylogenetic relationships. *BMC Evolutionary Biology*, 11.

Xia, X., Xie, Z. (2001). DAMBE: Data analysis in molecular biology and evolution.
Journal of Heredity, 92, 371-373.

Yamazaki, N., Ueshima, R., Terrett, J. A., Yokobori, S., Kaifu, M., Segawa, R., Thomas,
R. H. (1997). Evolution of pulmonate gastropod mitochondrial genomes:
Comparisons of gene organizations of *Euhadra*, *Cepaea* and *Albinaria* and
implications of unusual tRNA secondary structures. *Genetics*, 145(3), 749-758.

Youngblood, M., Masta, S. (2012) <<http://www.mt-trna-draw.org/index.html>>