5-1-2013

Mitochondrial Genome Evolution in Pupillid Land Snails

Jason D. Marquardt

Recommended Citation

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

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

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

Sequencing of mt genomes is informative for taxonomic identification as well as analysis of evolutionary changes. Pupillid land snails fall under the Pupilloidea super family. They are a wide-ranging geographic group for which completely assembled mitochondrial genome sequences are not available. Among the Stylommatophora, an infra-order of pulmonates (Gastropoda, Mollusca) that the pupillid land snails belong to, only four mitochondrial genomes have been sequenced: *Albinaria caerulea* (Hatzoglou *et al.* 1995)*, Cepaea nemoralis* (Yamazaki *et al.* 1997), *Succinea putris* (White *et al.* 2011) and *Cylindrus obtusus* (Groenenberg *et al.* 2012*)*; all representatives of other 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)*, 16S*F (Tongkerd *et al.* 2004) and 16SR (Jørgensen *et al.* 2004) were used to obtain initial amplicons which were then sequenced and used to design exact primers for long distance PCR and DNA sequencing. Universal primers for *12S* were designed in collaboration with Coenraad Adema by inspecting alignments, BLAST N, of all available alignable *12S, nd1, cox3* and *nd3* gene sequences under Panpulmonata (clade) and identifying the most conserved regions to place primers in. No conserved regions of sufficient length were identified in *nd1, cox3* or *nd3* leaving 12S with the only viable locations for primers. The 5 prime PCR extender system, high fidelity enzyme system (2200510, 5 PRIME, Inc., Gaithersburg, MD, USA), long distance PCR, was used per the following protocol to amplify the mt genome with initially; 5 within gene fragments then later in 2-3 larger fragments each being 6-8 kb covering the entire mt genome (Appendix C, Sec 3). The thermo cycler program was carried out according to LDPCR manufacturer protocol with individual modifications for *G. cristata* (Appendix D, sec. 1), *P. muscorum* (Appendix D, sec. 1) and *V. pusilla* (Appendix D, sec. 1). Each LDPCR reaction was 25 µl in volume containing: 2.5 µl buffer, 1.25 µl of 8 µM DNTP's, .2 µl of TAQ (5U/µl), 16.55 µl of molecular water, 1.5 µl of genomic DNA and 1.5 µl of the forward and reverse 10 µM primers. *Vertigo pusilla* used .5 µl of primer for both the forward and reverse primer and water was used to add up to the final volume. A 1% agarose gel with the stain ethidium bromide was used to ascertain that the LDPCR was producing an amplicon as well as amplifying fragments of the correct size. All reactions were treated with ExoSAP-IT (78201, Affymetrix, Inc., Santa Clara, CA, USA) prior to sequencing to eliminate unincorporated primers and free DNTPs. The amplicons were then completely sequenced via primer walking using exact primers that were designed using previously obtained sequences. All sequencing reactions used the BigDye Terminator v3.1 Cycle Sequencing Kits (4337455, Life Technologies Corporation, Carlsbad, California, USA). A standard thermo cycler protocol (Platt *et al*. 2007) was used except that the anneal temperature was modified between the different species. All sequencing reactions were 10 µl containing 1-6 µl of ExoSAPed LDPCR product, 1 µl of one 10 µM primer, 2 µl of 5X big dye buffer and 1µl of enzyme with any remaining volume being water. *Gastrocopta* *cristata* functioned optimally using 2 µl of ExoSAPed LDPCR product (Appendix D, sec. 2). *Pupilla muscorum* worked best using 4 µl of ExoSAPed LDPCR product (Appendix D, sec. 2). For *Vertigo pusilla* 2 µl of enzyme combined with 6 µl of ExoSAPed LDPCR product (Appendix D, sec. 2). All ethanol precipitations were according to standard procedure (Appendix D, sec. 3). The UNM Molecular Biology Facility ran all sequences, dissolving the DNA pellet in 10 µl of formamide for sequencing using a ABI 3130xl (Applied Biosystems by Life Technologies Corporation, Carlsbad, California, USA).

Genome annotation:

Chromatogram editing and contig assembly were done with Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, Michigan, USA). Genes were tentatively identified with a combination of BLAST X and BLAST N searches on NCBI, restricted to Panpulmonata (NCBI:BLAST 2012). Exact beginning and end points were identified by manually aligning the newly sequenced genes with genes from related snail species represented in NCBI, and then compared to find the correct start and stop codons based on length of the genes and the location of the start codon in related species (Rambaut 2012). The tRNAs were identified by a combination of the tRNAscan-SE Search Server with exceptionally low cutoff scores (1×10^{-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$^{\mathrm{Lys}}$ and tRNA$^{\mathrm{Gln}}$ to 61 bp in tRNA$^{\mathrm{Thr}}$ and tRNA$^{\mathrm{Ser\ (AGN)}}$. In *P. muscorum* the size range is between 68 bp for tRNA$^{\mathrm{Glu}}$ and tRNA$^{\mathrm{Ala}}$ to 61 bp for tRNA$^{\mathrm{His}}$. *Vertigo pusilla* is different from the others in that the range is from 69 bp in tRNA$^{\mathrm{Ile}}$ to 52 bp for tRNA$^{\mathrm{Arg}}$. 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$^{\mathrm{Ala}}$ and *ND6* as well as 14 bp between tRNA$^{\mathrm{Gln}}$ and tRNA$^{\mathrm{Leu\ (UUR)}}$. In *P. muscorum* the most significant non-coding regions are between tRNA$^{\mathrm{Val}}$ and *16S* (18 bp) and between tRNA$^{\mathrm{Ala}}$ and *ND6* which is 30 bp. *Vertigo pusilla* has one large non-coding region between tRNA$^{\mathrm{Ala}}$ and *ND6,* which is 23 bp. All 3 species have their POR between *COX3* and tRNA$^{\mathrm{Ile}}$ 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 Bayseian and maximum likelihood phylogenies *S. putris* is in a clade with *C. nemoralis* and *C. obtusus*. This is unlike the nearest neighbor joining phylogeny where *S. putris* is in a clade with *A. caerulea*, *G. cristata, P. muscorum,* and *V. pusilla* or like in the maximum parsimony phylogeny where *S. putris* is sister to *A. caerulea.* The other species move between clades in an analogous manner to *S. putris,* depending on the method analysis.

Discussion

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

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

All of the mitochondrial genomes examined have high A/T content and low G/C content (Figure 13). The three new genomes are on the higher end of that spectrum, as defined by my dataset at ~77%. The A/T richness is not particular to a certain region; the relative amount of A/T is equal from the protein coding genome to the remainder of the genome. There are several regions were 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 descrepancy 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 analysies have *Biomphalaria* as a clase nested within the non-Stylommatophoran clade.

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

Conclusion

Gastrocopta cristata, P. muscorum and *V. pusilla* genomes' are the first fully sequenced mitochondrial genomes in their families. These three genomes, *G. cristata*, *P.*

muscorum and *V. pusilla* have been sequenced to provide for the first time coverage of the Pupilloidea super family. These sequences add genera to improve the accuracy of phylogenetics in areas where there previously has been no data.

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

Figure 1: Mt genome map of *G. cristata*. Protein coding genes are blue, tRNA are green and rRNA are red.

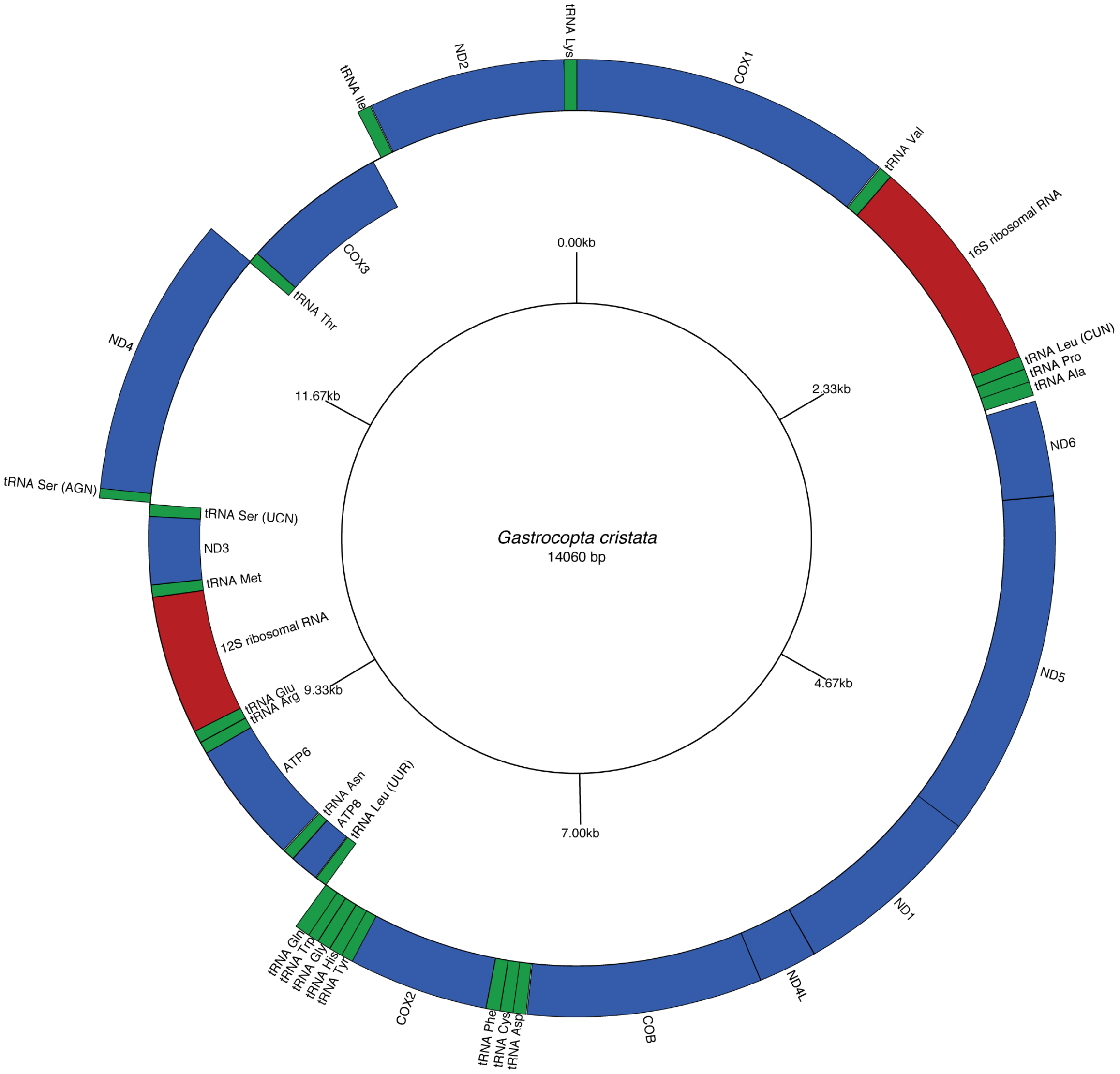


Figure 2: Mt genome map of *P. muscorum*. Protein coding genes are blue, tRNA are green and rRNA are red.

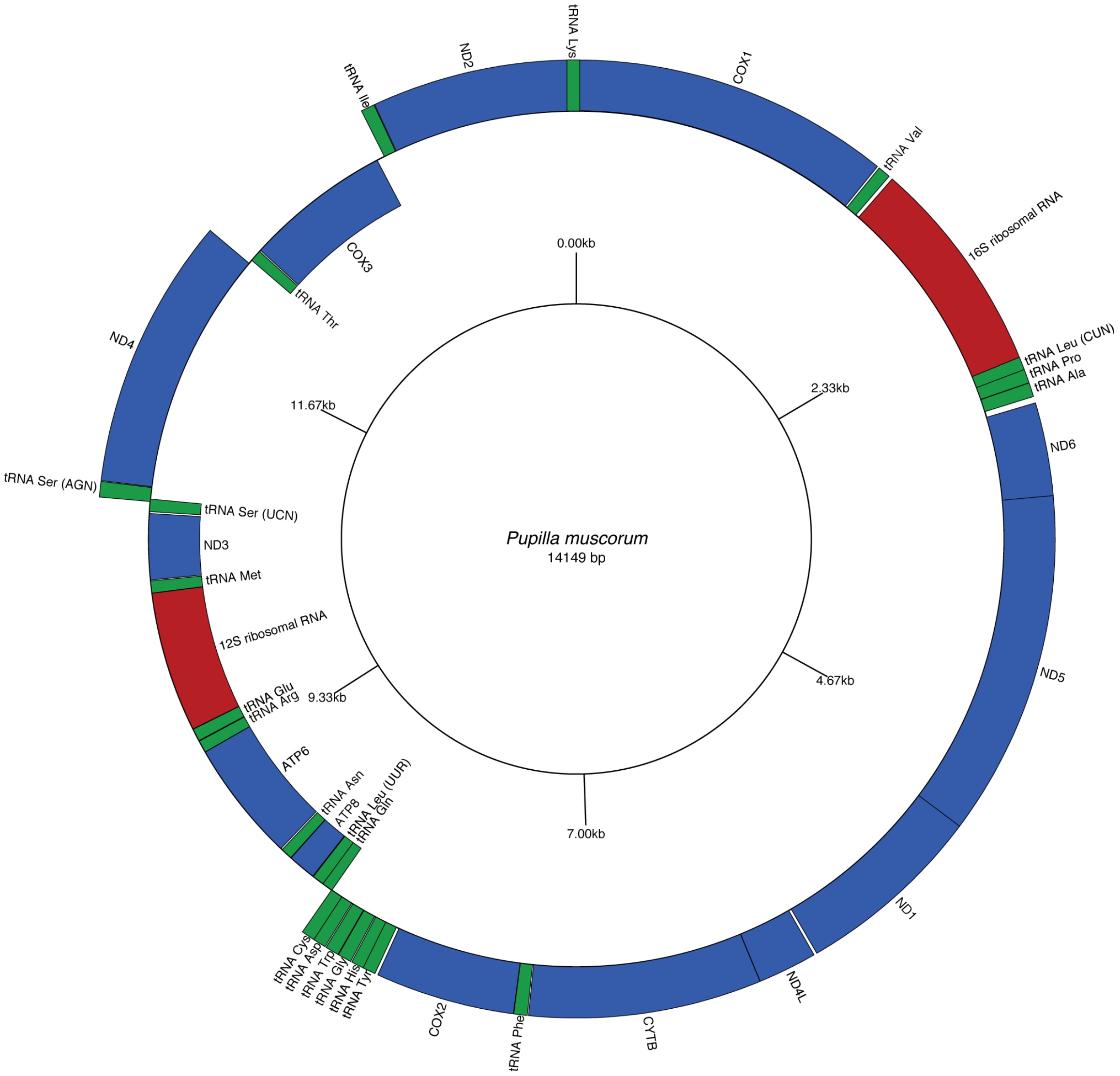


Figure 3: Mt genome map of *V. pusilla*. Protein coding genes are blue, tRNA are green and rRNA are red.

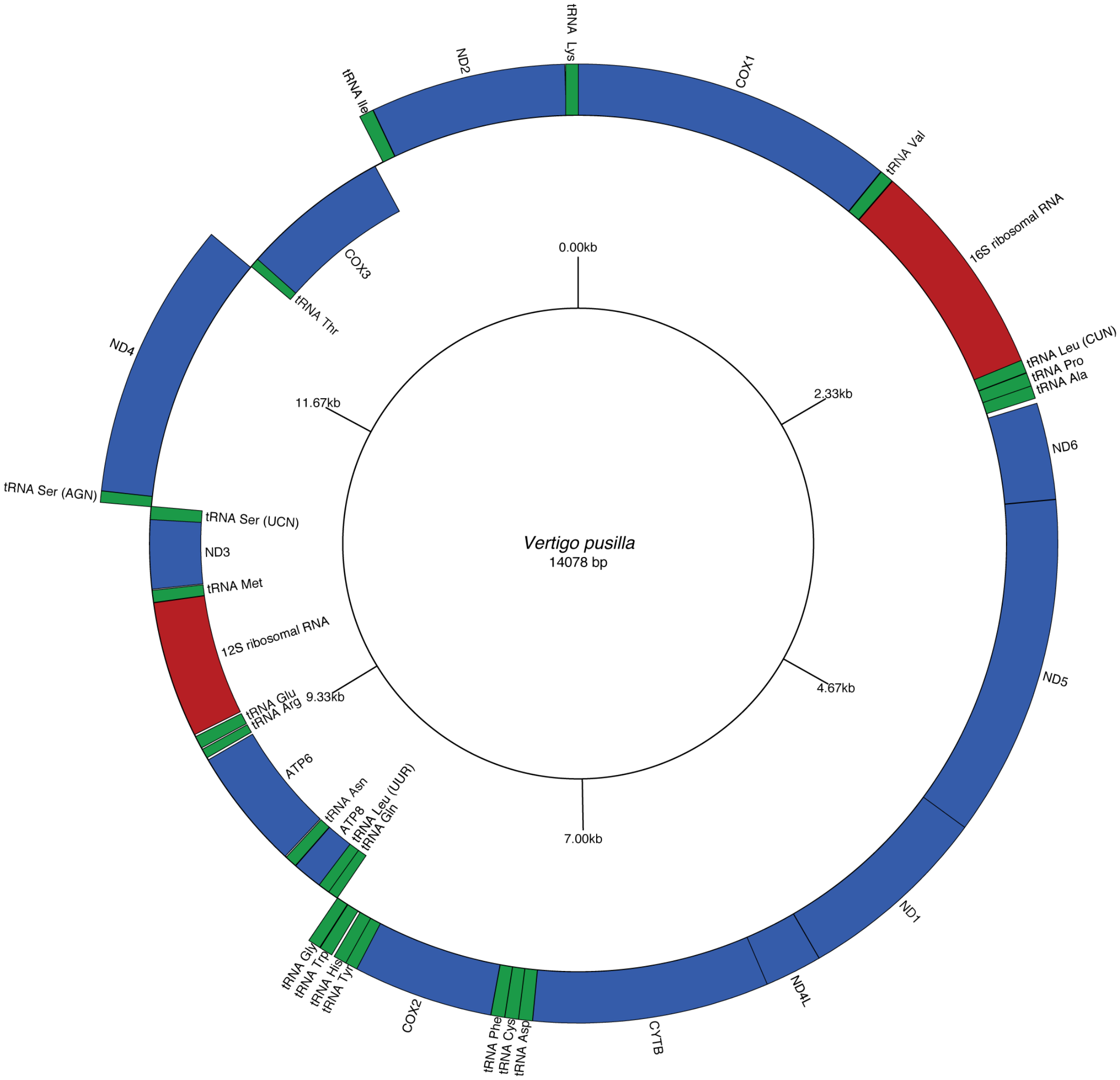


Figure 4: Comparison of genome length across data set.

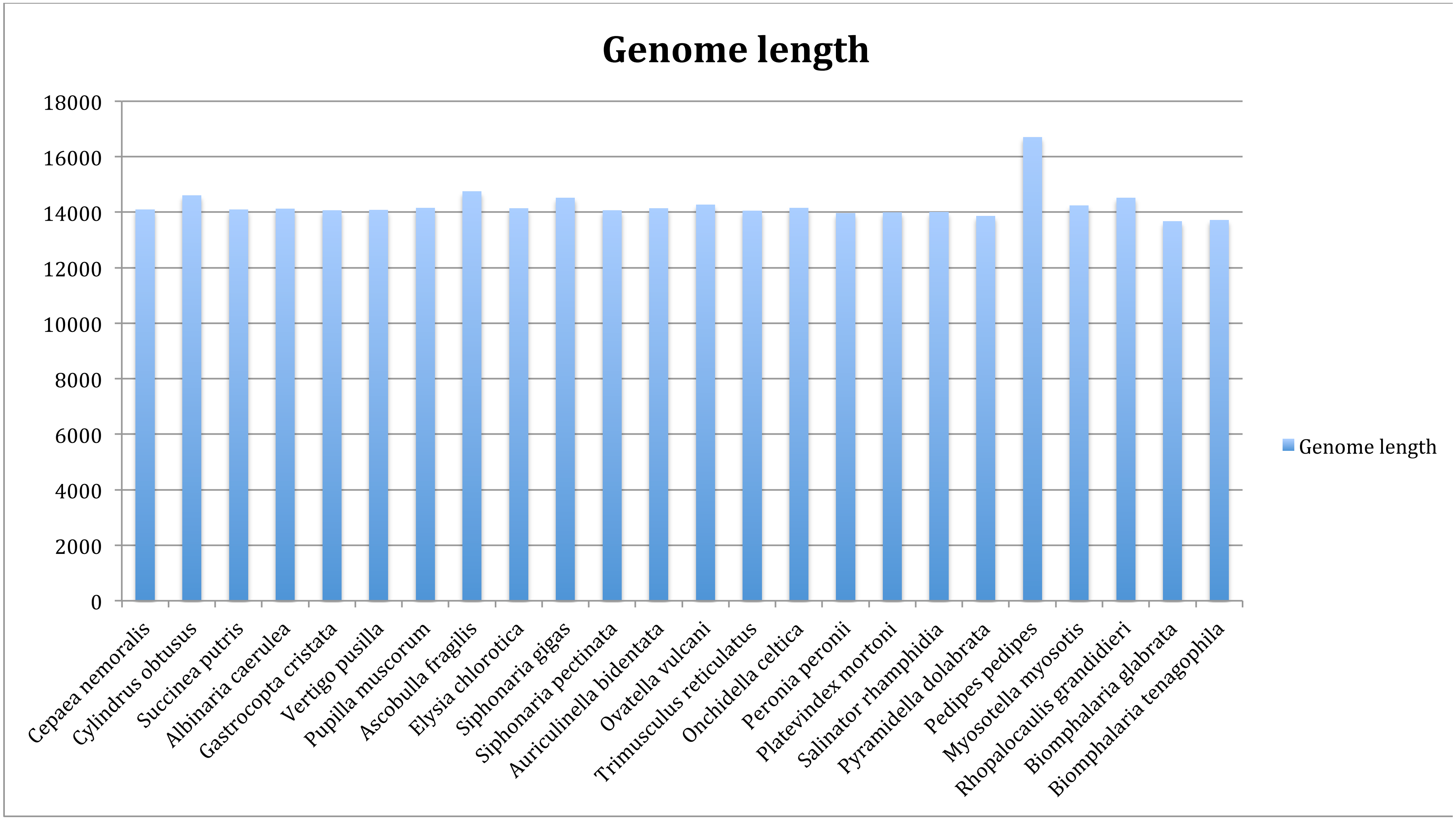


Figure 5: Nucleotide density plot for mt genome of *G. cristata*. Matlab automatically determined the bin size. Position 1 is at the begining of *cox1*.

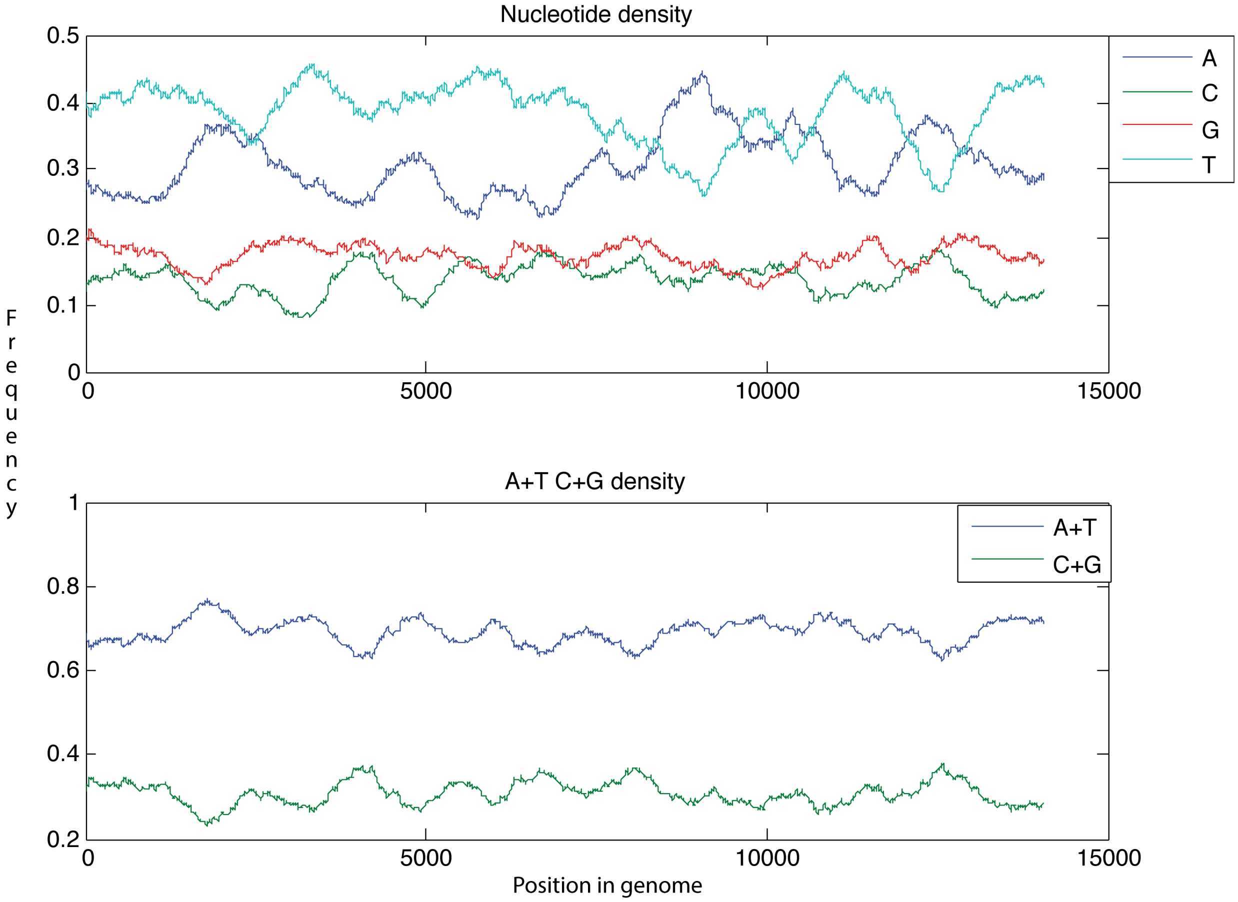


Figure 6: Nucleotide density plot for mt genome of *P. muscorum*. Matlab automatically determined the bin size. Position 1 is at the begining of *cox1*.

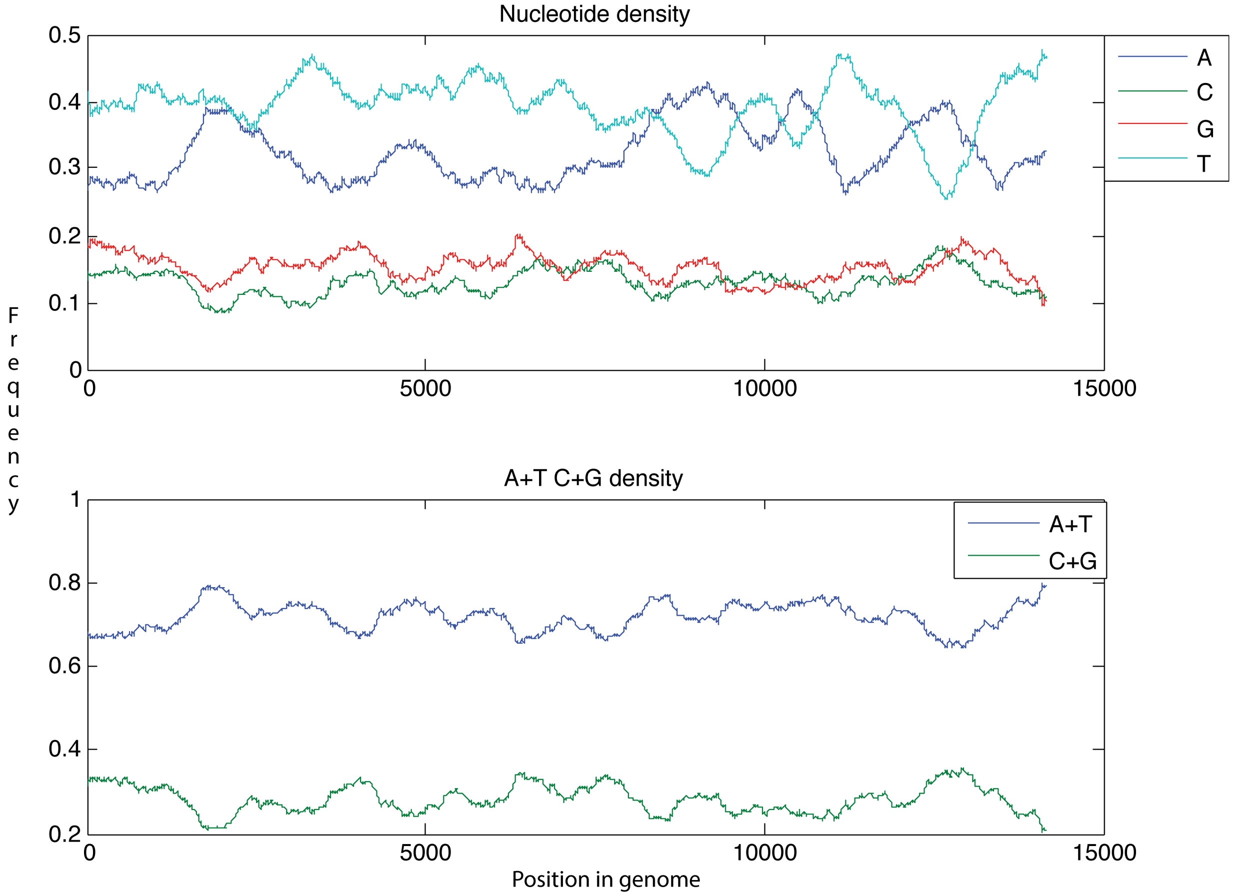


Figure 7: Nucleotide density plot for mt genome of *V. pusilla*. Matlab automatically determined the bin size. Position 1 is at the begining of *cox1*.

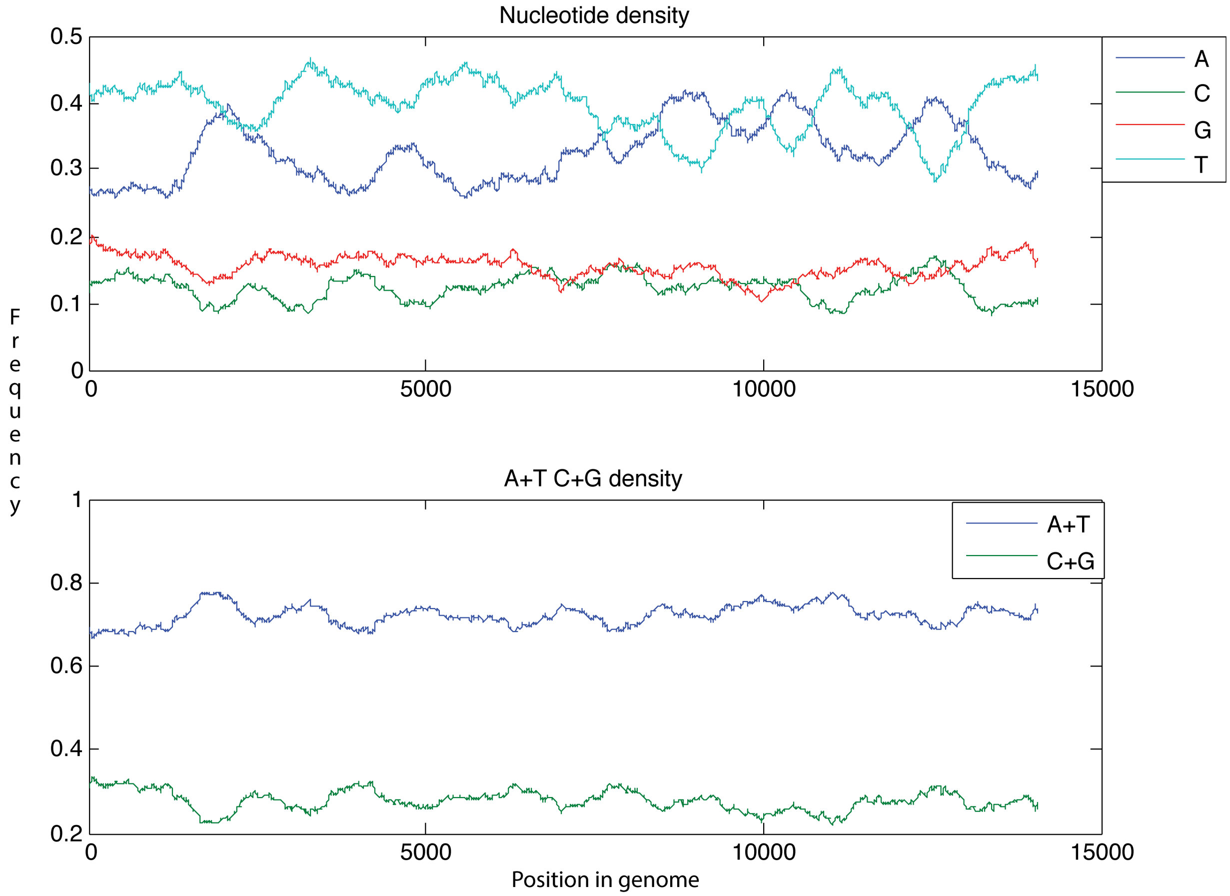


Figure 8: Predicted tRNA structures for *G. cristata.* non-standard tRNA are indicated with a black star. The number of missing bonds, missing arms and non-standard spacing determine which tRNA are considered non-standard. tRNA are in 5' -> 3', left to right and are based on DNA sequence.

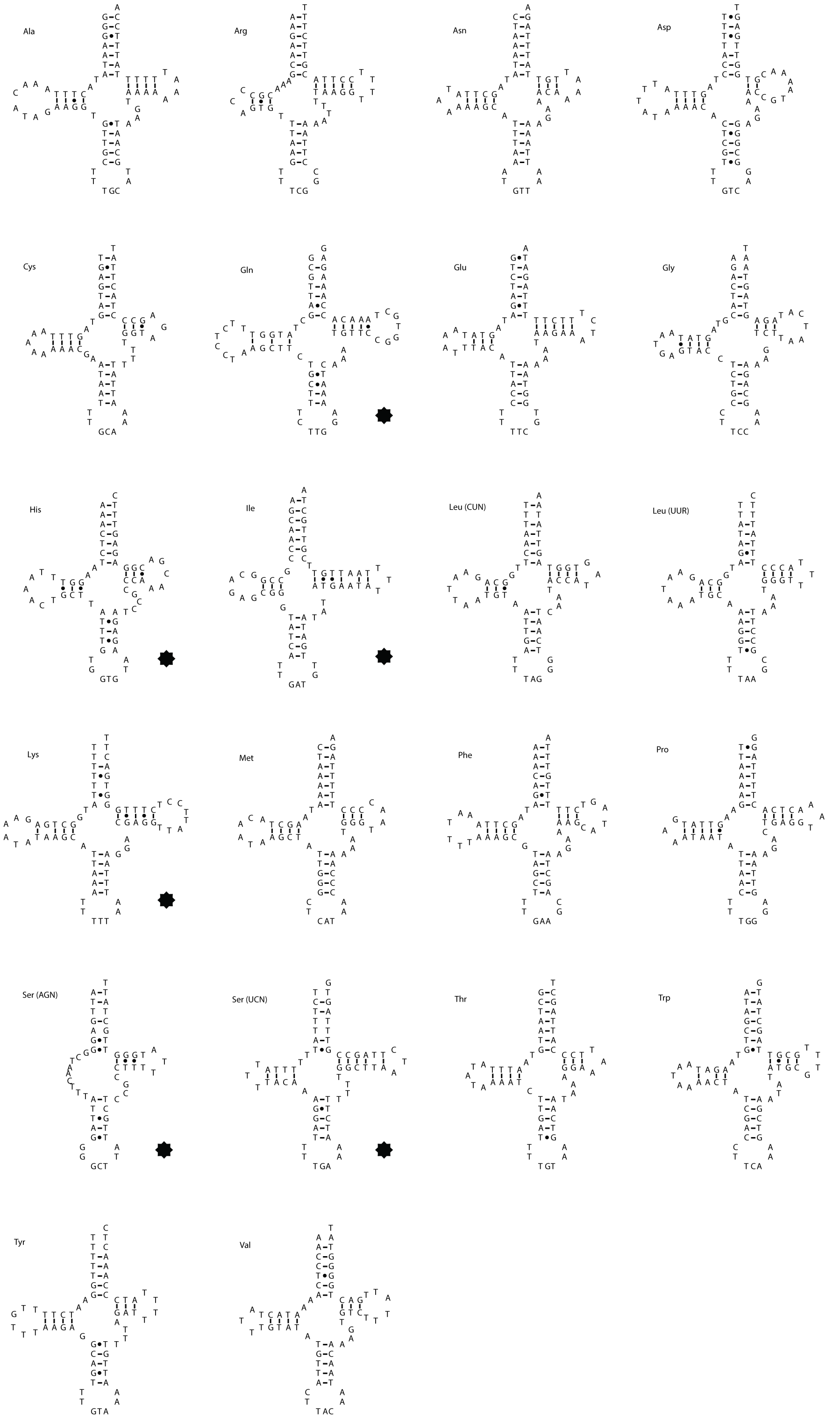


Figure 9: Predicted tRNA structures for *P. muscorum.* Non-standard tRNA are indicated with a black star. The number of missing bonds, missing arms and non-standard spacing determine which tRNA are considered non-standard. tRNA are in 5' -> 3' , left to right and are based on DNA sequence.

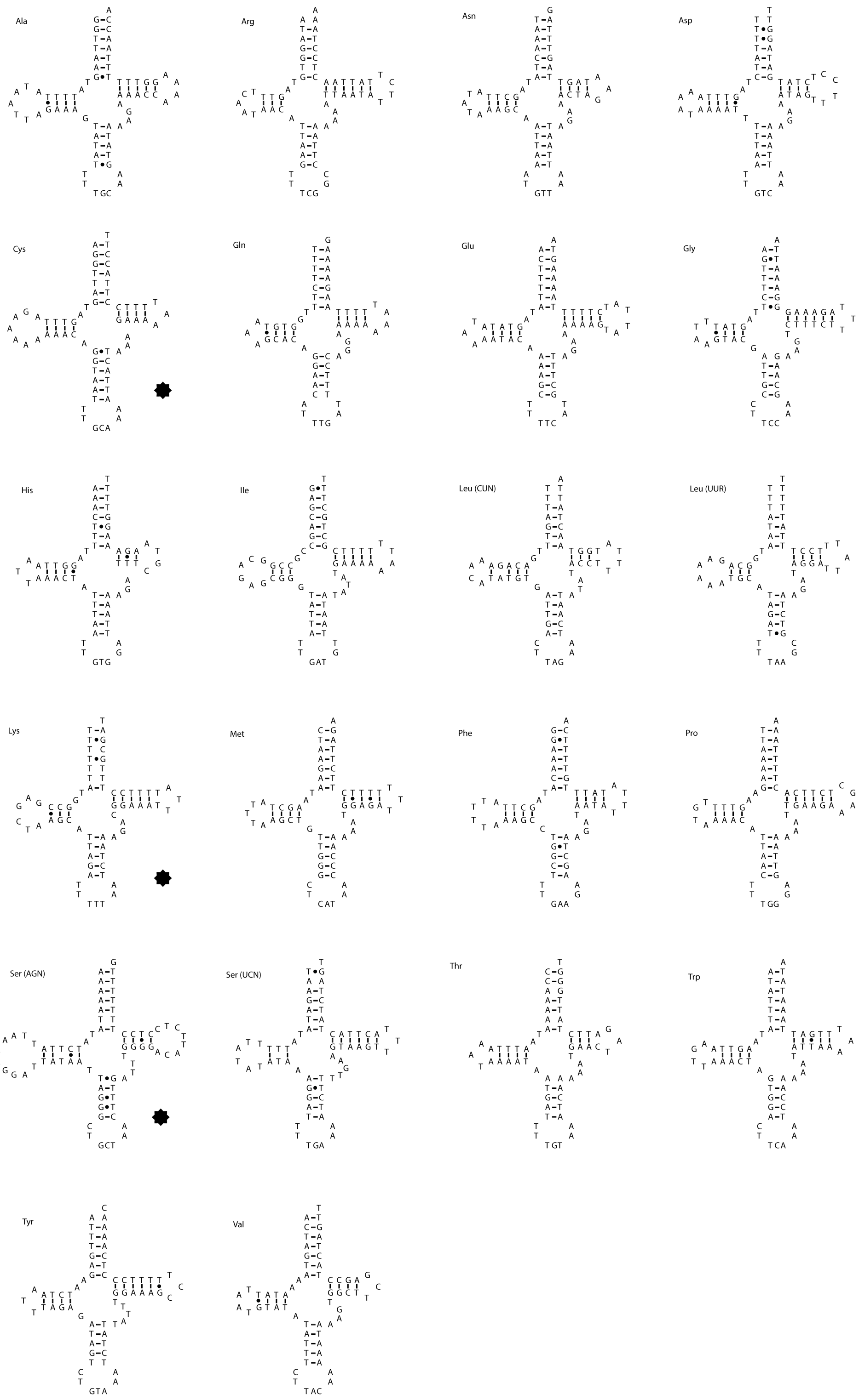


Figure 10: Predicted tRNA structures for *V. pusilla*. Non-standard tRNA are indicated with a black star. The number of missing bonds, missing arms and non-standard spacing determine which tRNA are considered non-standard. tRNA are in 5' -> 3' , left to right and are based on DNA sequence.

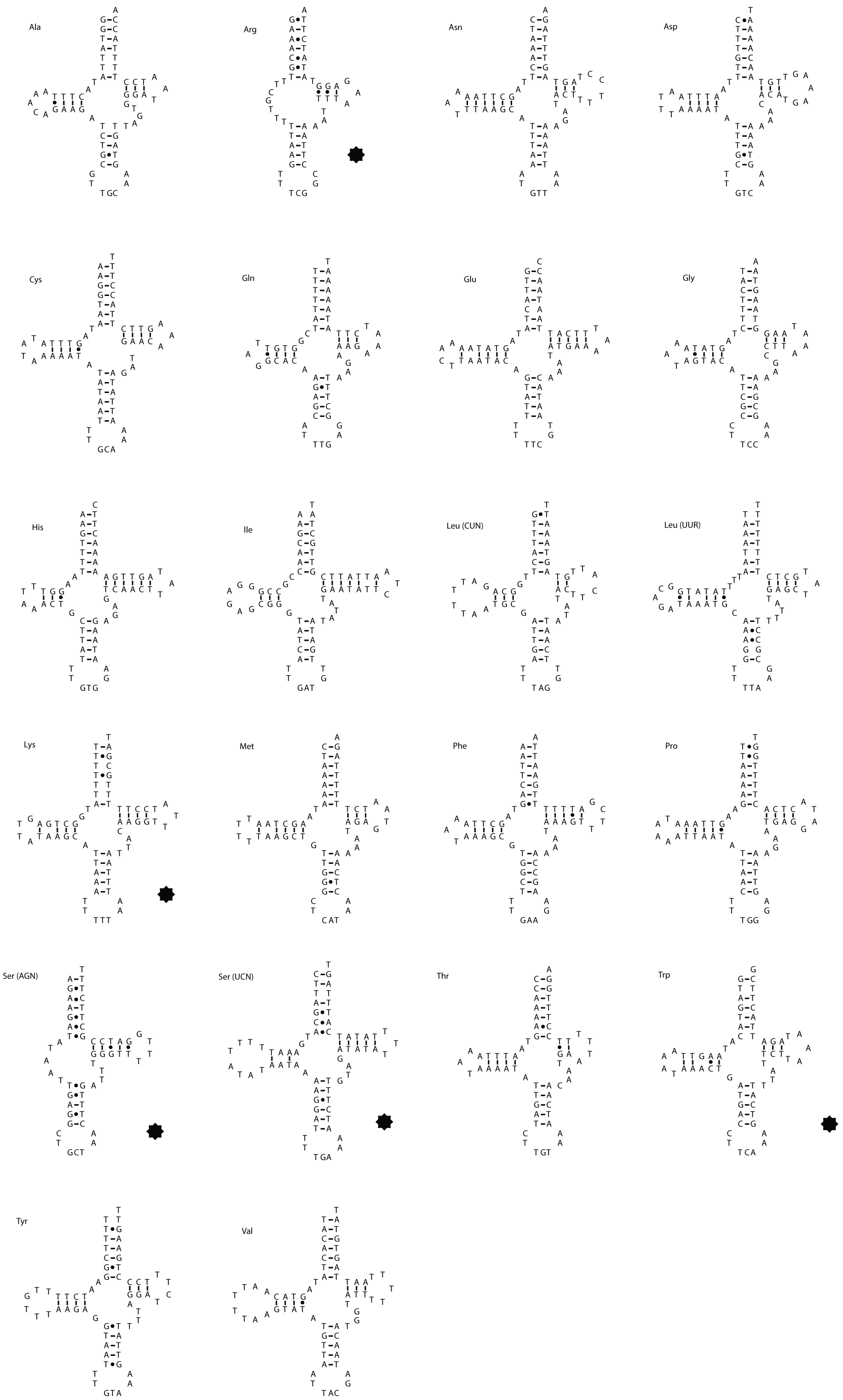


Figure 11: Nucleotide density plot for mt genome of *A. caerulea*. Matlab automatically determined the bin size. Position 1 is at the begining of *cox1*.

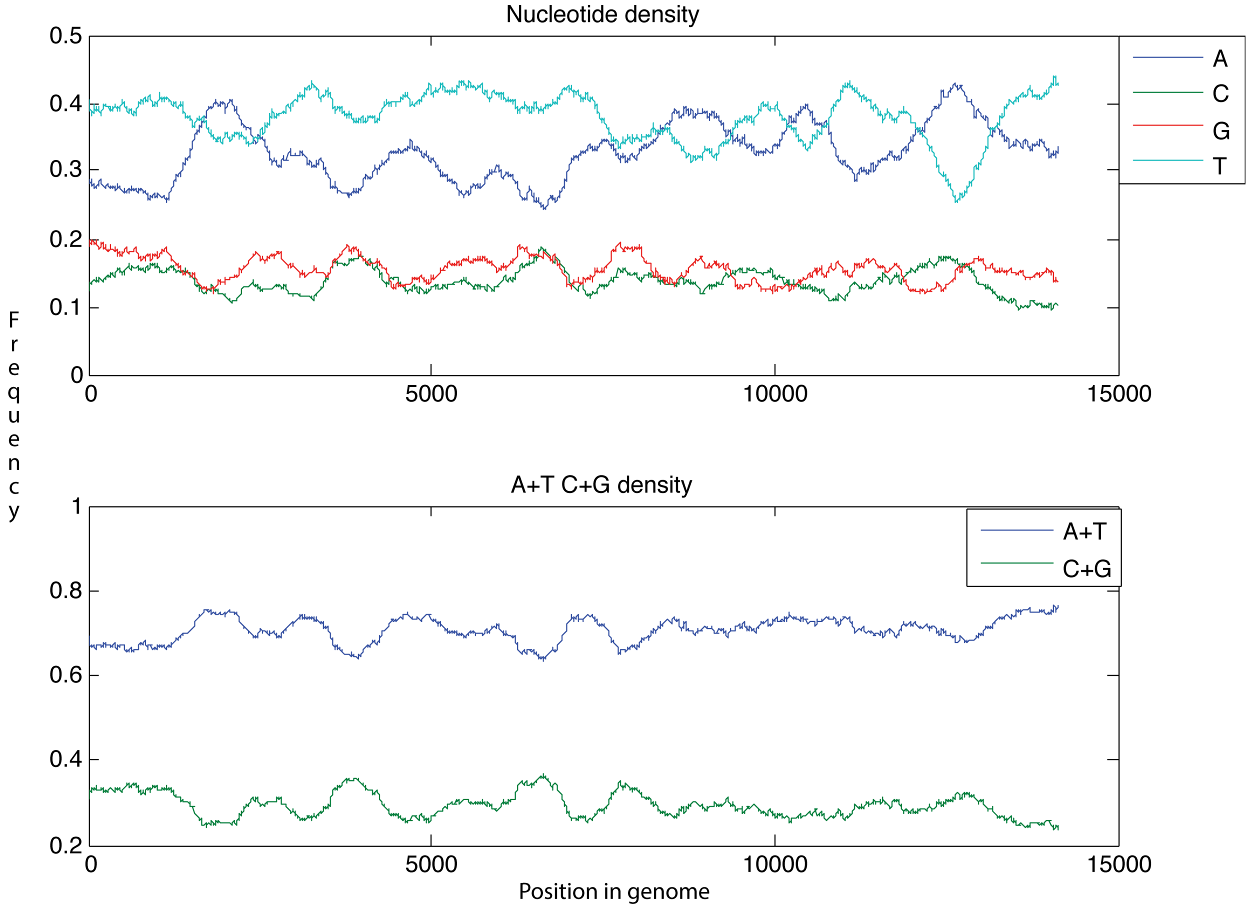


Figure 12: Nucleotide density plot for mt genome of *C. nemoralis*. Matlab automatically determined the bin size. Position 1 is at the begining of *cox1*.

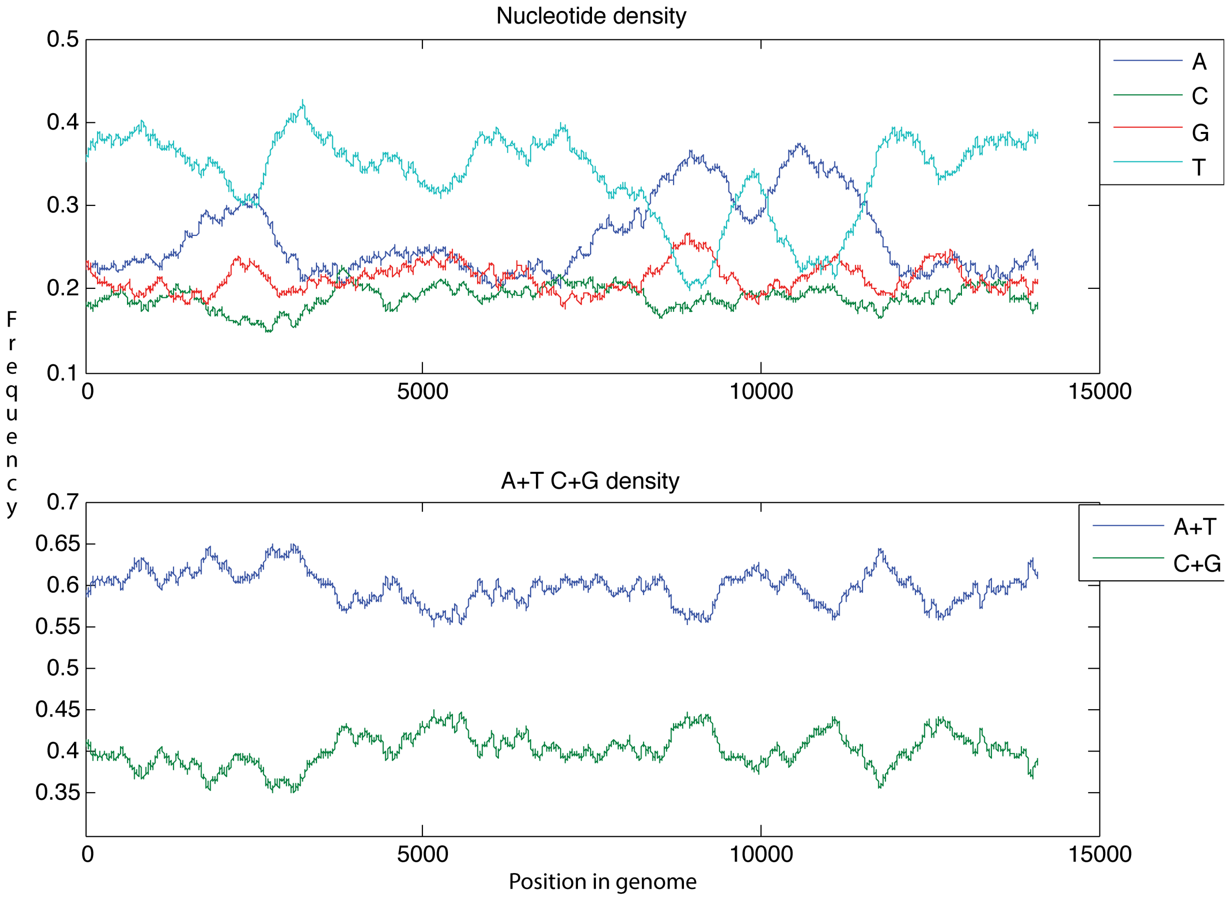


Figure 13: A+T and G+C percentages by mt genome.

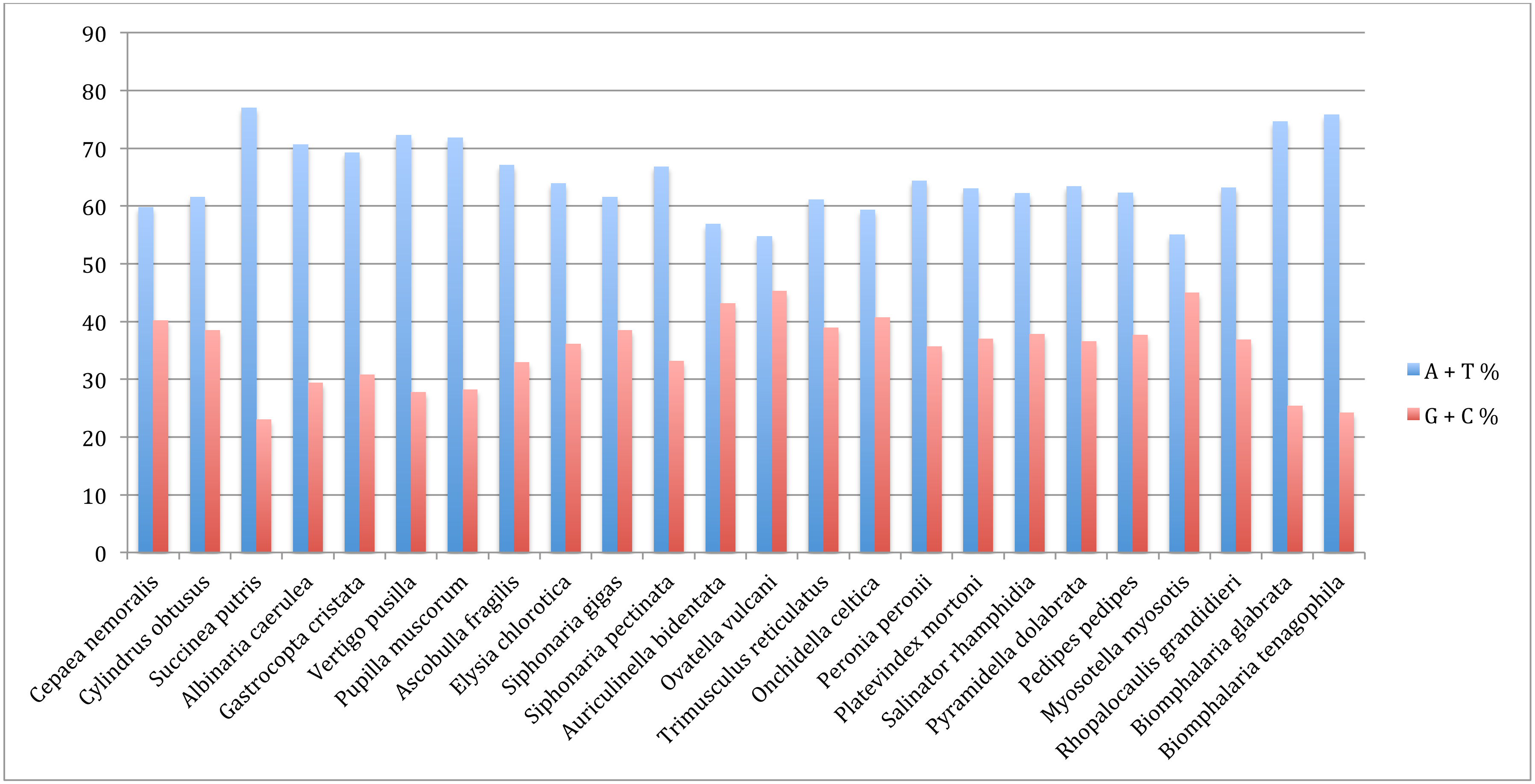


Figure 14: Gene order comparison across data set. Green boxes are tRNA, red boxes are rRNA and blue boxes are protein coding genes. Underlined genes are in the opposite orientation to non underlined genes.

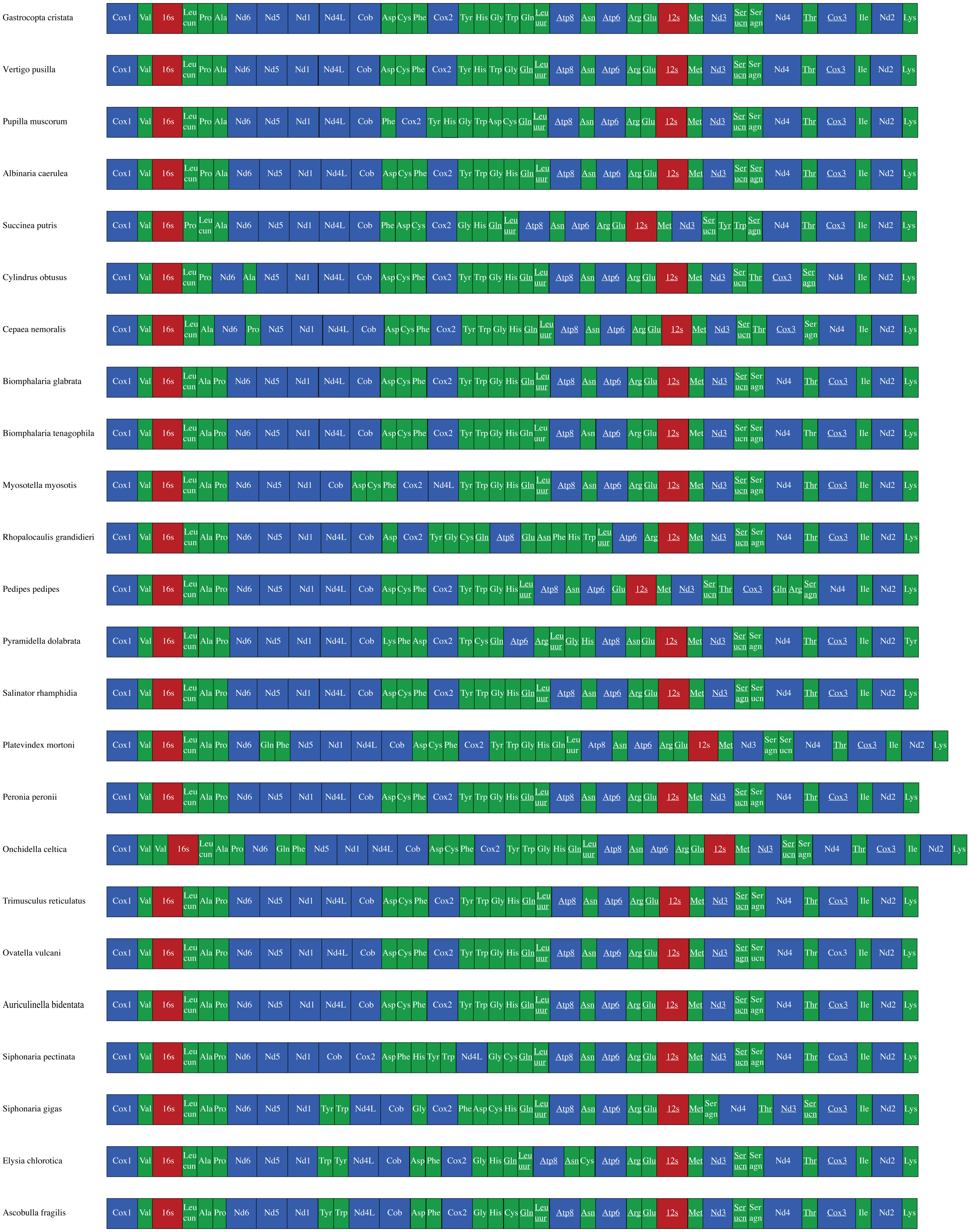


Figure 15: G/C skew comparison of *G. cristata*, *P. muscorum* and *V. pusilla*. Made using a coding sequence alignment. Upper half is ratio of G and lower half is ratio of C. Positive peaks mean G is more prevalent, negative peaks mean C is more prevalent. Bin size was 210 nucleotides, advancing 3 nucleotides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

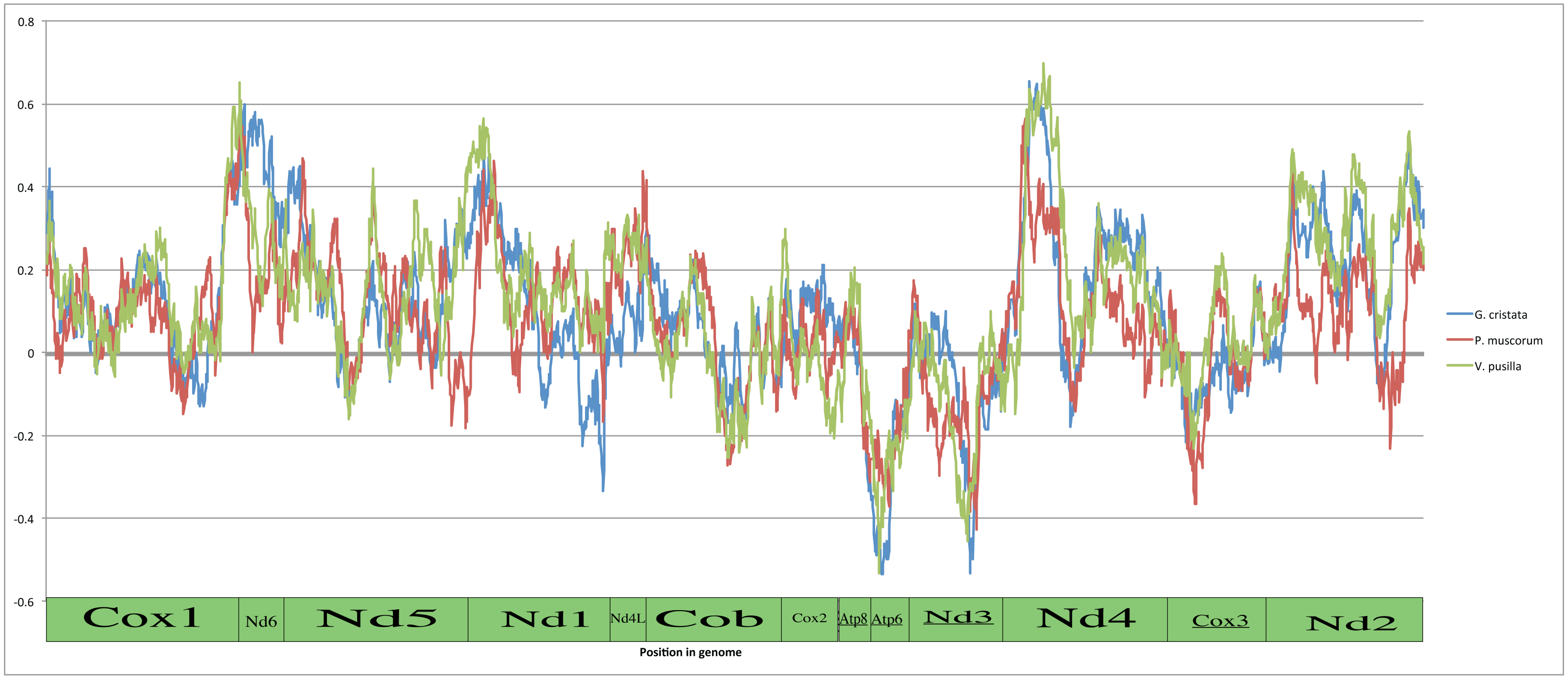


Figure 16: Codon usage pie chart for mt genome of *G. cristata* by amino acid. The two serine and two leucine tRNA's are grouped together under serine and leucine respectively.

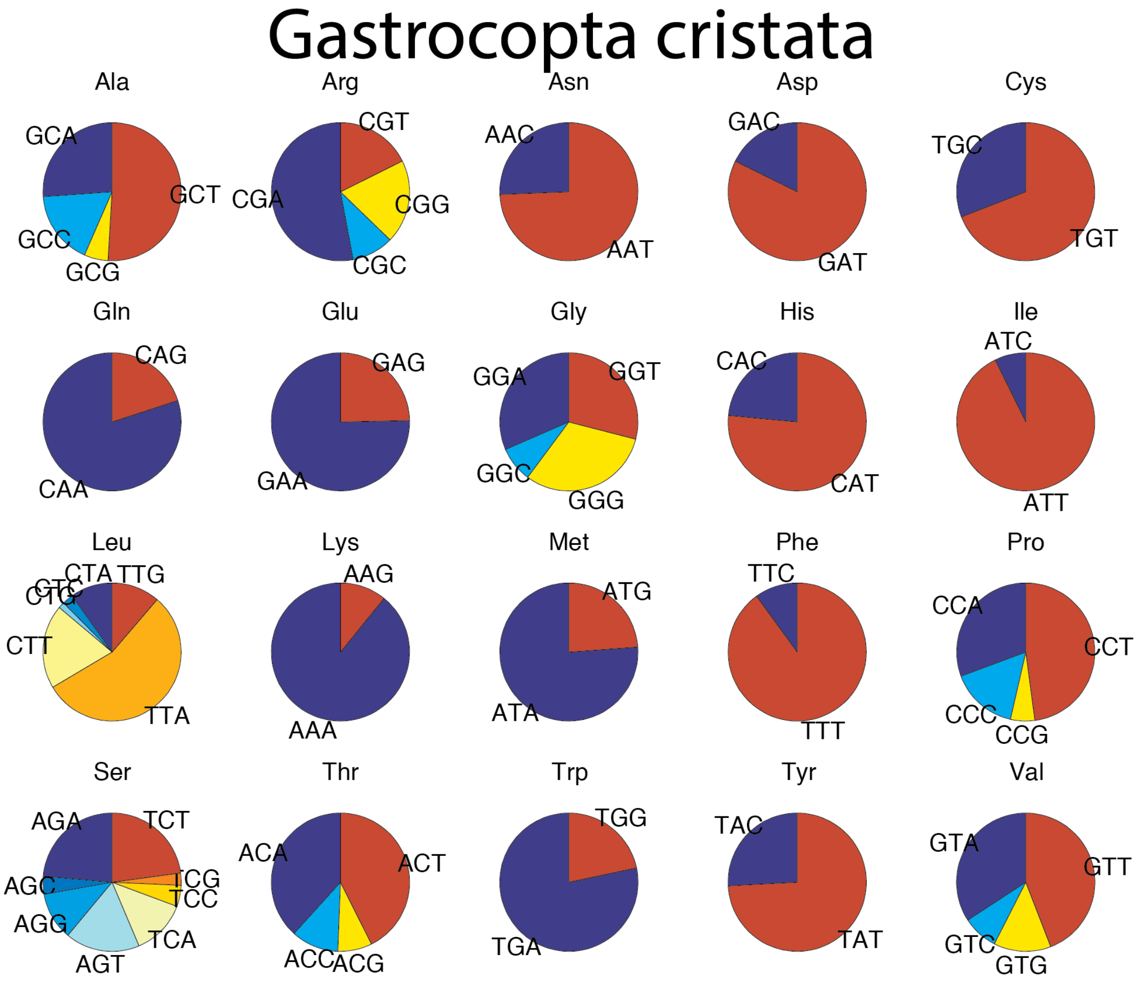


Figure 17: Codon usage pie chart for mt genome of *P. muscorum* by amino acid. The two serine and two leucine tRNA's are grouped together under serine and leucine respectively.

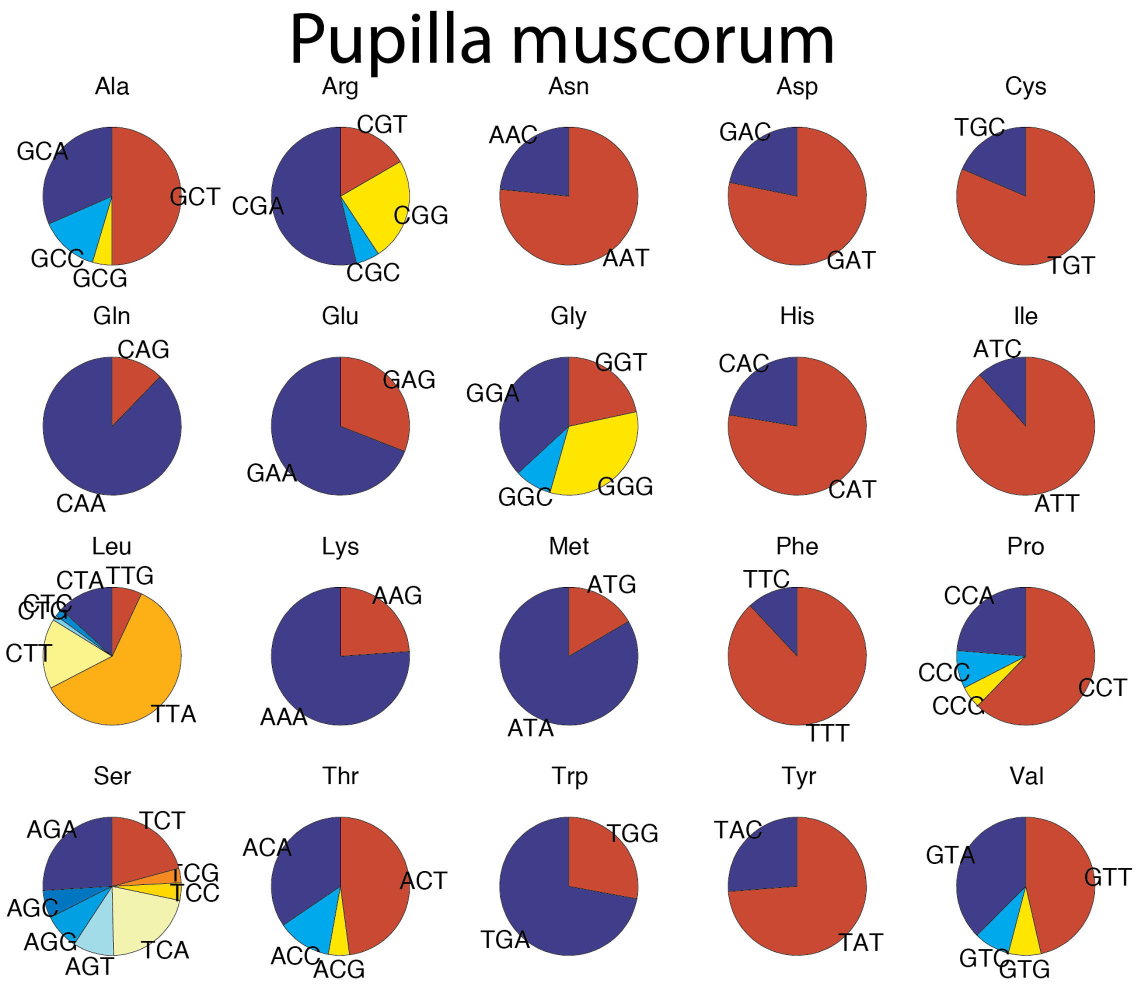


Figure 18: Codon usage pie chart for mt genome of *V. pusilla* by amino acid. The two serine and two leucine tRNA’s are grouped together under serine and leucine respectively.

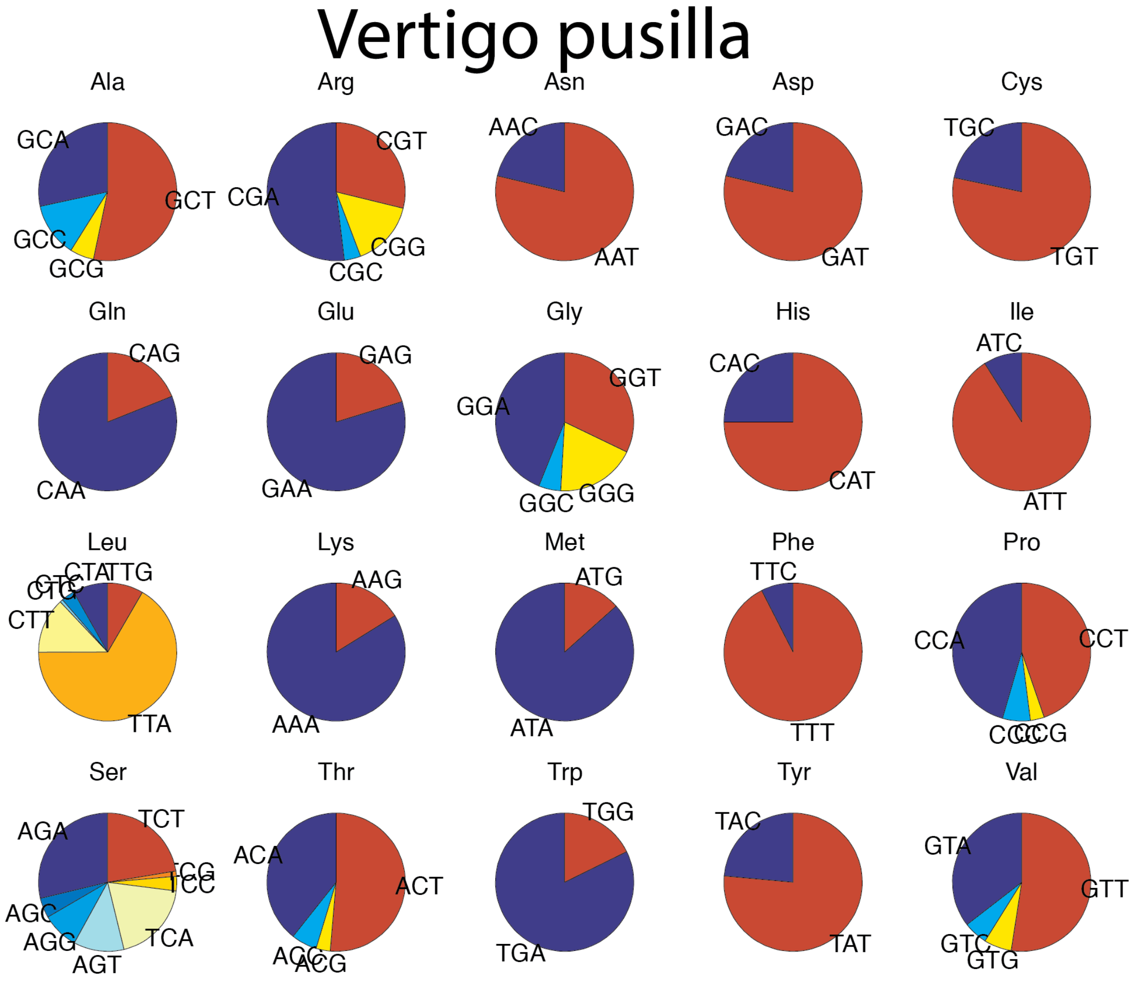


Figure 19: Codon usage pie chart for mt genome of *A. caerulea* by amino acid. The two serine and two leucine tRNA's are grouped together under serine and leucine respectively.

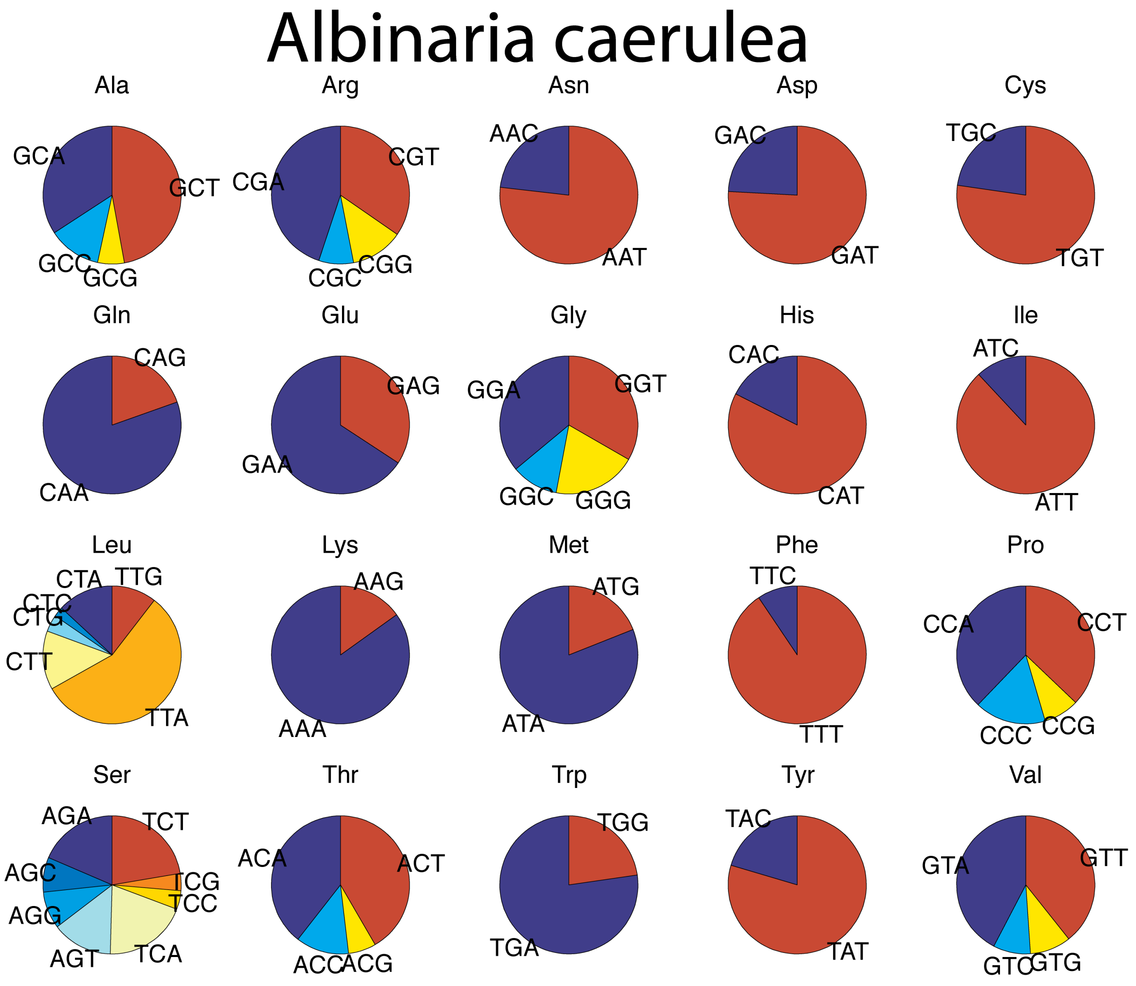


Figure 20: Codon usage pie chart for mt genome of *C. nemoralis* by amino acid. The two serine and two leucine tRNA's are grouped together under serine and leucine respectively.

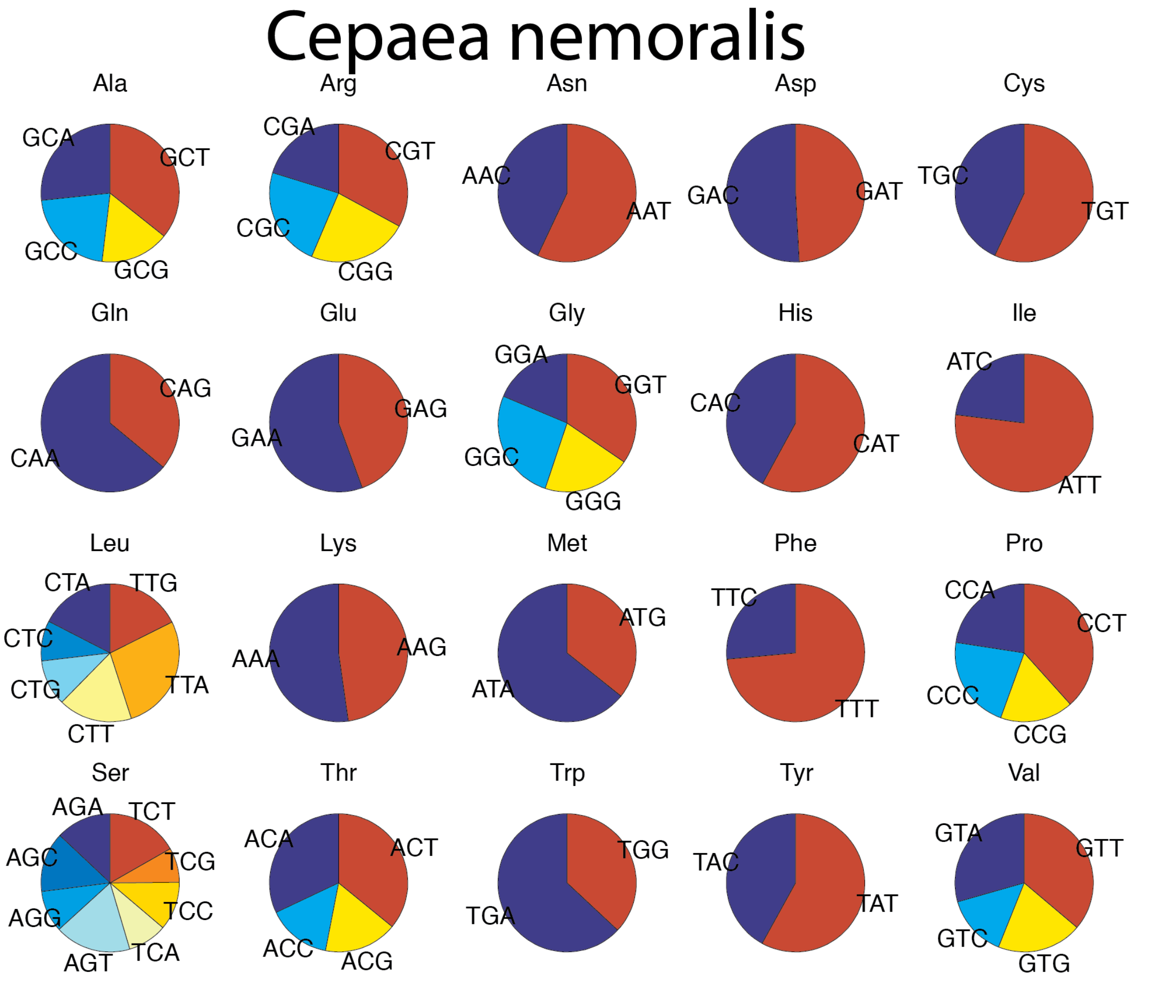


Figure 21: Bayesian phylogenetic tree for complete data set. Node labels are the posterior probabilities.

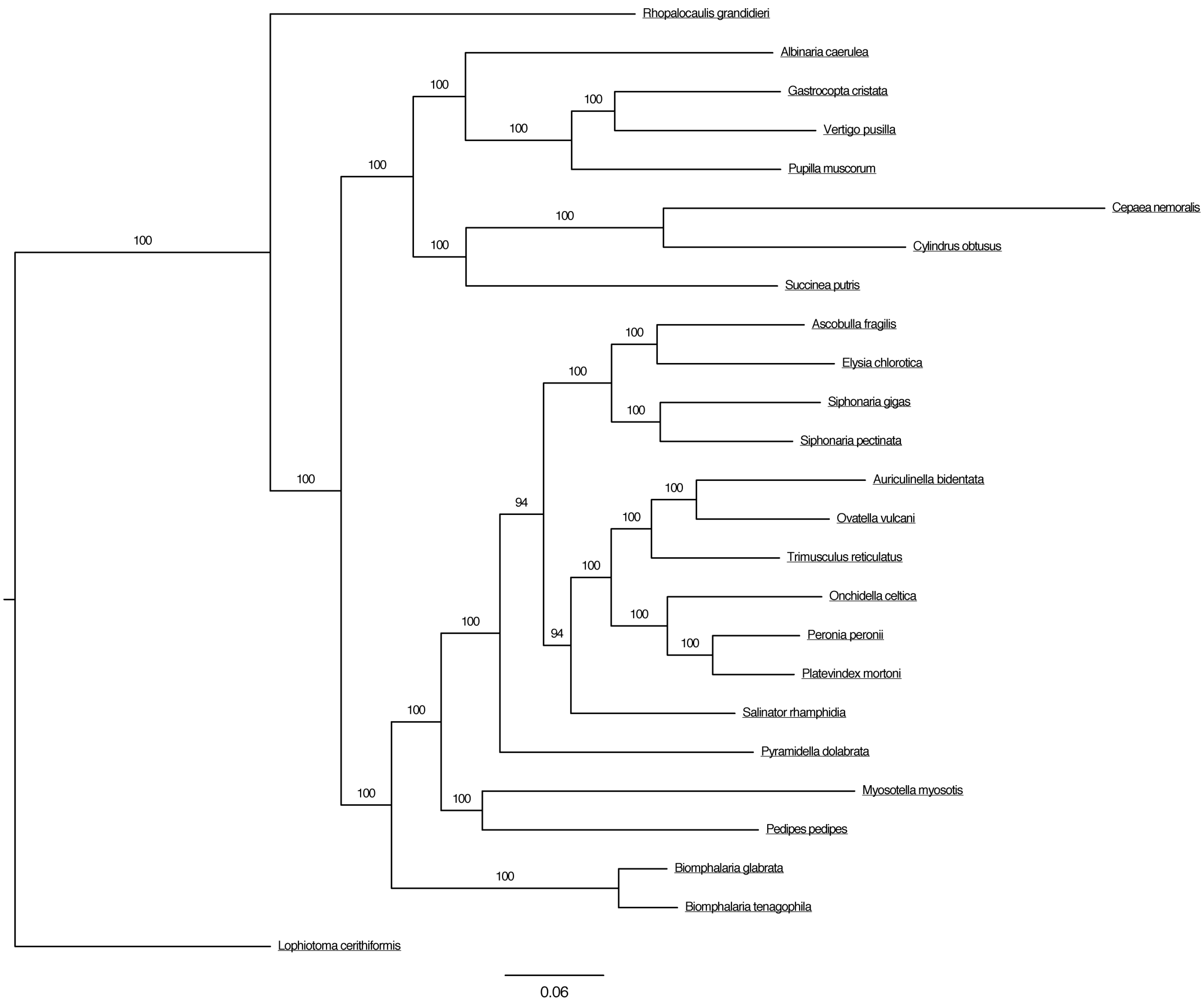


Figure 22: Maximum likelihood tree for complete data set. Node labels are bootstrap values.

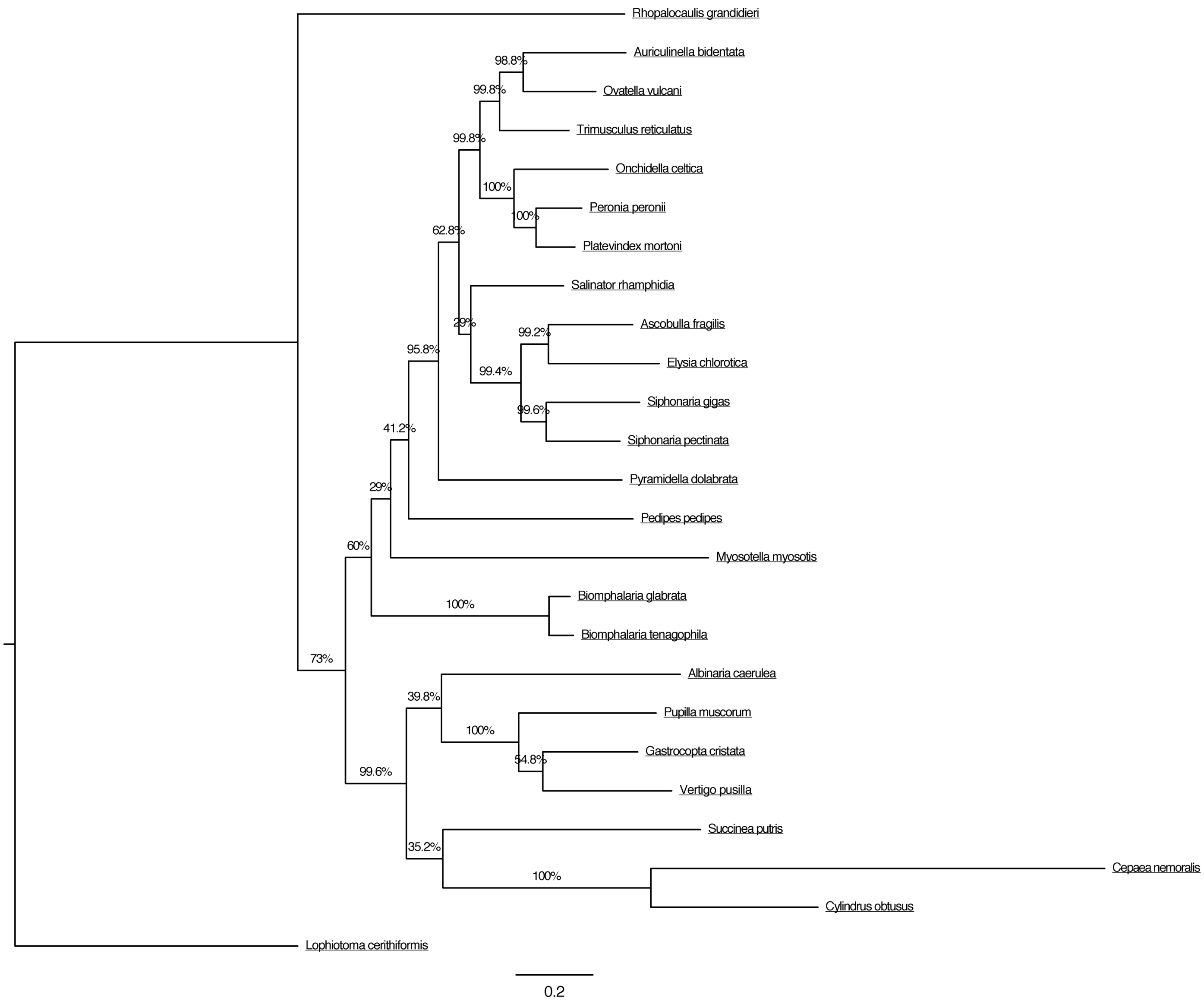


Figure 23: Maximum parsimony tree for complete data set. Node labels are bootstrap values.

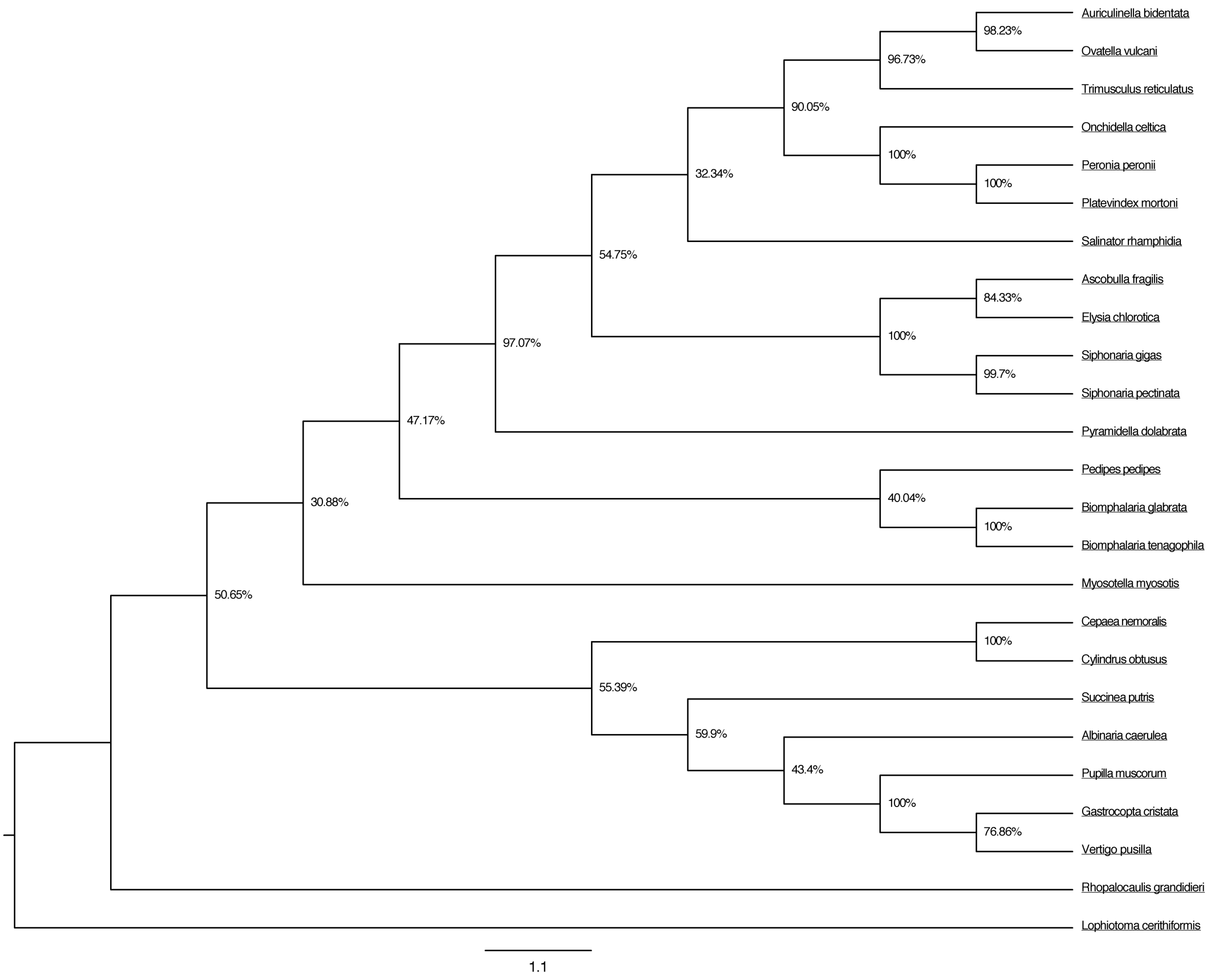


Figure 24: Nearest neighbor joining tree for complete data set. Node labels are bootstrap values.

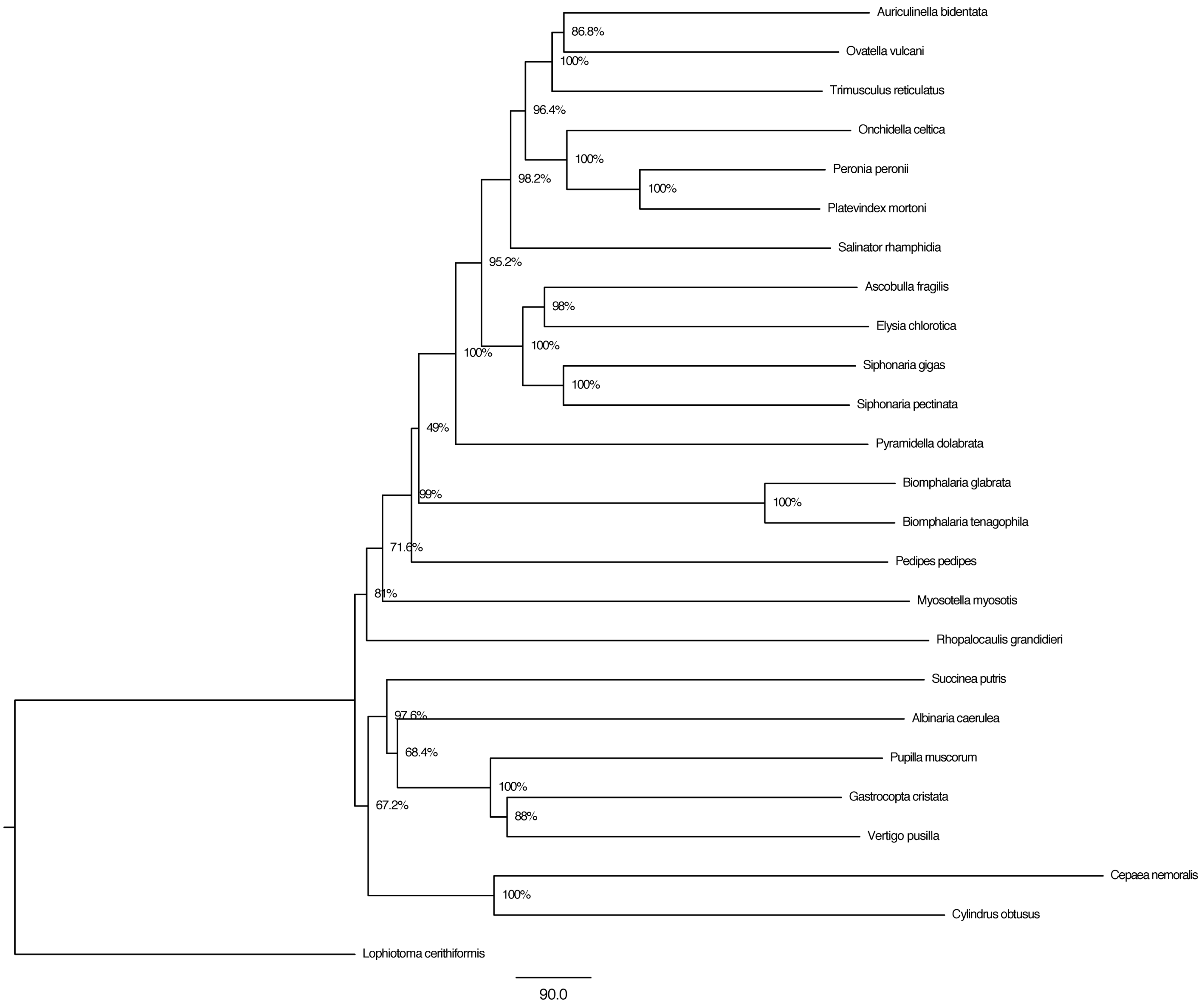


Table 1: List of species used for this project and their associated Genbank accession numbers.

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

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

Yellow hilighted codons are alternative start codons.

	G. cristata	*P. muscorum*	*V. pusilla*
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 comprision was conducted using Mega 5.0 and coding only sequence with the divergent regions concatenated out.

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

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

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

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

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

Unique differences in Gastrocopta cristata	190
Unique differences in Vertigo pusilla	244
Unique differences in Pupilla muscorum (outgroup)	308
p-value	0.009
value needed for significance using the sequential Bonferroni correction	0.0083

Table 4: D_n/D_s between *G. cristata*, *P. muscorum V. pusilla*. Calculated using a matlab script (Appendix A, sec. 2) and the coding sequence only with the divergent regions concatenated out.

	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 Dn/Ds by Jason Marquardt

clc
clear all

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

%number of sequences
x=24;

%size of gene
win=1554

count=1;

[Headers, Sequences]=multialignread(name);

for i=1:1:x

    for j=1:1:x

[dn1(:,count) ds1(:,count) vardn1(:,count)
vards1(:,count)]=dnds(Sequences{i}, Sequences{j}, 'window',
win,'GeneticCode',5);

count= count+1;

    end

end
```

Appendix A: Section 3

```
% various whole genome nucleotide analysis by Jason Marquardt

clc
clear all
format('short')

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

%how many entries
x=24;

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

for i=1:1:x

fastaentry=i;
%=========================
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
                     MDDHFFNVVVTAHAFVMIFFMVMPIMIGGFGNWMVPLLIGAPDMSFPRMNNMSFWLLP
                     PSFILLISSSMVEGGAGTGWTVYPPLSGITGHSGASVDLAIFSLHLAGMSSILGAINF
                     ITTIFNMRAPGMTFERLSLFVWSILVTVFLLLLSLPVLAGAITMLLTDRNFNTSFFDP
                     AGGGDPILYQHLFWFFGHPEVYILILPGFGIISHILGNASVKQPFGSLGMIYAMISIG
                     ILGFIVWAHHMFTVGMDVDTRAYFTAATMIIAVPTGIKVFSWLMTLYGMNSKFDASMY
                     WVLGFIFLFTLGGLTGIVLSNSSLDIVLHDTYYVVAHFHYVLSMGAVFAIFAGFVYWF
                     PVMTGLVLQERLAKAQFIIMFIAVNLTFFPQHFLGLSGMPRRYSDYPDSYFKWNQISS
                     FGSLMSIFAVLLFVLIVWEAFLSQRSVMFSSAAPYSREWTENNFPPDFHGNLIPSISV
                     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="MIVLFGLGMFLVFSFCFLRSPVSYGGALIALSTIGALLLSLGCS
                     TWYGYLLFLVYVGGLLVLFLYIIMLSSNFNLQVSFKLMGLIFLAFLVSKLYNFSYPKN
                     SLGVSLSECSEDFSLGLFLGLGGLLLLVFFAIVHIVFLKGQPVQVKND"
     gene            3318..4969
```

```
                     /gene="ND5"
     CDS               3318..4969
                     /gene="ND5"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 5"
                     /translation="MSRLPVLLLLSCLLLMSLFFYSVFKSGETYVIELNLMTLSSVSF
                     STLFIFDKISLSFGAVVTLISFCVFSFANTYMSEDNFQFRFILILAMFVLSMNILIFS
                     GSLFVLLLGWDGLGISSFALIIYYQNKTSLGAGYLTLLTNRLGDVVIIICVPFFLILG
                     SFNIFPIMSSYETIIFILAIAALTKSAQYPFSAWLPAAMAAPTPVSALVHSSTLVTAG
                     VYLIIRLSVNCSMSNSTASLLLFCGSVTCVLGGLSALYENDLKKIIAFSTLSQLGLMM
                     FSLGLKLPNLALLHLYAHAMFKALLFLSAGLILISGFGSQDLRLLGATLYTIPSVVVF
                     FNISSLCLMGIPFLSAFYSKHVIYEVVAMSETNAISFFLMLIGAILTTIYSIRTVKIL
                     SWNSISGLSVFSRLPLYTYVPLCILFFTSIISGKTFSMLDISYLTFMFTSSYYQLMLH
                     FLMALGVMIGLFLSGEKKSHMLSSMFFLWPSSNNLTKYFWPVLKSAKMLDYGWVEPVV
                     LLSSSLNKLGNKLHFLFLNESKYLNSMVRGVFCSLTVFFLCI"
     gene             4961..5866
                     /gene="ND1"
     CDS               4961..5866
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /translation="MYLFIMKAVILCLCVLLAVAFFTLLERKVLGYMQIRKGPNKVGI
                     WGIVQPLADALKLFIKEKVLLYMSNNLLFIFTPMLGFFVSLSLWNLAPSNMMVNYMIF
                     GLLLFFCLTGLNVYTTALAGWSSNSLYAFLGAIRASAQSISYEVSLVMLLLFPALVSS
                     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="MSQLLFILIYGFVILLYSFSSQSKHLISSLIILESLMVISIIYL
                     FFSLGSVNSIGLLLLILTFAAAEAALALSLLISILRVMKNDNMLNLSY"
     gene             6150..7262
                     /gene="COB"
     CDS               6150..7262
                     /gene="COB"
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome b"
                     /translation="MRKISLLEGLTSLPTPLNISIFWSGGSILGMMLGIQLVTGLFLS
                     MHYTSDISISFDSVVHIMRDVPNGWLFRLFHANGASFFFIFLYLHIGRGLYYQSYSVQ
                     PKTWMVGVTIFLISMATAFLGYVLPWGQMSFWGATVITNLLSAIPYWGGTLVEWVWGG
                     FSVGQATLNRFFSLHFILPFLIALLVMVHLIFLHEKGSTNPLGQTHHINKVTFHPYFT
                     YKDLVGFICVGLSLVLICAYFPFSLSDPENFIYANPMVTPTHIQPEWYFLFAYAILRS
                     IPSKLGGVVALAMSIFILYFLPLSSQFMSVPASFNMIQQMLFWWLVVVFLLLTWLGAC
                     PIEDPYLSLAGPLTIIYFFSYPLLMIVGNFSVSILK"
     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="MSYWGSLNLMDPSSPIQIEMMLFHDHAMILLIGIFTLVALLGVK
                     LCLNKFSSRATHEAQLLEIFWTIIPALLLIWLALPSLRLLFLLDEQGSNGLILKATGH
                     QWYWSYEIPSSGTDSFDSYMVQEQALTNGEFRLLEVDNRPVLPYSNEISVLTTSSDVI
                     HAWALPSMGVKMDAVPGRLNMMGLYSLTPGVFYGQCSEICGANHSYMPIVVEFVHLND
                     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="MMTDLFSSMDANTSFFLWLLPLAAAAFFIFNQSFMSSFFSVLGN
            SICSLWGTEETFFSKKLLLTAVMSFLMLNNFLGLLPLVFSTTTSLWINSSVALVLWGS
            ILISGWVFSFKKSAAHLAPAGAPGALIPFLVLIETVSILIRPITLTVRLVANISAGHI
            ILGLIANVLASTLSFIPFSLAASVMIFYYLFEMFVSFIQAYIFTLLISLYMAEHP"
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="MMYYCIFPFLLTTILVLLYFFTLWKAASPSSSKESPFECGFDPM
            SSMRKPFSLRFFLLIILFLIFDVEVVLLFPILTQMKMATSTVVLAAYSTFLLMLLGGL
            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="MGLAIVFLTILGLFFSSWDVVVMTLVVGVLLSFLNMYLNFSFNL
            FSDLFLTESLSSMLLFLSLLISILAILCTWWIKSTKYIMTVMALNLFLVFAFSVKNLL
            VFYFFFEASLIPTLALIIIWGYQPERLQAGSYMMLYTVLASFPLLFFILWVNWNMSSL
            DFFLLKMLGSSLGGAMLIFSSLGFLVKLPMFSVHLWLPKAHVEAPLSGSMILAGVLLK
            LGGYGLYLLNYSFSLIVLGTSSMIICFIALWGGLFASLMCLQQTDMKALVAYSSVSHM
            GLVIGGVFLNLNWGVVSAKITMVAHGFTSSALFVLVNMSYKKVLSRSFSASGGLLSIF
            PKMALLWFLFCSINMAAPPSLNLVGELYLLPVLFLFSNIFVLIMALSLFLSAGYNMML
```

```
                YTNGNHGATNNLMTPSTSYSTADYFSLILHLLPFFFLLKLNMFYCYSSTLLDES"
     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="MFKTPFHLVEFSPWPLLGSALLTCMPLGLIYYIRIGTPYLLAYG
                MACTAMVAYLWWRDIVREATYQGHHTTYVVKGLKLGVMLFILSEVCFFFAFFWAYFHS
                SLAPSVEIGSVWPPVGITTLSAFQVPLLNTSVLLLSGVSVTWAHHAIEEGKYNSAISG
                LLLTTLLGLYFLFLQFGEYKETSFSIADSVYGSTFFMATGFHGLHVAVGATFLFVCLI
                RMILYHFNTTHHVGFLAAAWYWHFVDVVWLFLYVSIYWWGS"
     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="MMSWYGVLFSVIVFLGPLISLSTSNWLLVWGGMELSLISLLPIL
                NLAKSSIPVESAMKYFLVQASASAVILITGLNIYMLEGSEYLILILFILGLILKLGVV
                PLHFWVIPVVKGLKYIHMAFLLGPLKIVPLMLLMTNMNLDSSIFLFLLVSSIGTVLLG
                ALLGNNMSSIRMMLGASSISHSGWFLMGGLVGFMWIYFSIYMISLYVVLYYLKVYPSD
                NYLVSIFLFSLSGLPPFLMFPAKMMVLYSLLSMGLSPIVFSFFIISAIVSLNFYMKFS
                FSFVLSSSENQKNFGAGVMPGLLLMMGGFILFFI"
BASE COUNT     4331 a   1915 c   2417 g   5397 t
ORIGIN
        1 tttgcggtga ctttattcaa ctaatcataa agatattggg actttatata taatttttgg
       61 agtttgatgt ggaatagttg ggaccggatt atctttatta attcgtcttg aattaggaac
      121 ttctggggtc ttaatagatg atcatttttt taatgtggta gttacagctc atgcttttgt
      181 tataattttt tttatagtaa tacctattat aattggagga tttggaaact gaatggttcc
      241 actattgatt ggtgctcccg atataagatt tcctcgaata aataatataa gattttgact
      301 tttgcccccct tcttttattt tactcattag atcgagaata gttgaaggtg gggcaggaac
      361 aggatgaact gtatacccccc cattaagagg aattactgga cacagtgggg cttcggtgga
      421 tcttgctatt ttttctcttc acttggctgg aatatcttct attttaggag caattaattt
      481 tattacaact atttttaata tacgagctcc aggaataacc tttgaacgat tgagattatt
      541 tgtatgatct attttagtga ctgttttttt attacttctt tctttacctg ttttagctgg
      601 tgctattaca atgcttttaa cagatcgtaa ttttaataca agatttttttg atcctgctgg
      661 tggcggggat ccaattttat atcaacattt gttttgattt tttggacatc ccgaagttta
      721 tattttaatt cttcctgggt ttggaattat ttctcacatt cttggtaatg cttcagttaa
      781 acaaccattt ggaagattag gaataattta tgctataatt tctattggaa tcttaggttt
      841 tattgtttga gcccatcata tgtttactgt tgggatggat gttgacacac gggcttactt
      901 tactgcagca acaatgatta ttgcggtacc tacaggtatt aaagtttttta ggtgattaat
      961 aactctttat ggaataaata gtaagtttga cgcttctata tactgagtat taggttttat
     1021 ttttcttttt actttagggg gcttaacagg gattgtattg tcaaactcct ctttggatat
     1081 tgtgctacat gatacatatt atgtagtggc ccattttcat tatgttcttt caataggagc
     1141 agtttttgct atttttgcag gatttgttta ctgattccct gttataactg gattagtttt
     1201 acaagaacgg cttgccaaag ctcaatttat tattatattt attgctgtta atttaacttt
     1261 ttttcctcaa catttttttag gattatcggg gatgccccga cgatactcag actatccgga
     1321 tagatatttt aaatgaaatc aaatttcttc ttttggctca cttatgtcta tttttgctgt
     1381 acttttattt gtcttaatcg tatgggaggc ctttttaaga caacgatctg taatatttag
     1441 ttccgctgct ccttattcac gagaatgaac agaaaataat ttttcctcctg atttccatgg
     1501 taatctaatt ccaagaattt cagttttata attttgaacc tcaaaatact atttgtatat
     1561 gttacttaca ataacaaagt gtctttattg actggggtat attaaataaa tgtattatgt
     1621 actattataa aaattaatta aaatgattat aaaagtagta ggtttgttct taccttttgc
     1681 ataatggtta aattaaaatc ttttgtttat tcaattaccc gaaataagaa gatctatgat
     1741 tttttaactg ttaggttgtt aattttaatg tagaataatt aatttacaat gaaattatag
     1801 gggtgaaacg ctttcaattc ttaggataac tggattaaat agaaaaacat ttagtgtttt
     1861 agtcaggcaa taagtgataa gatttattgc taaaaaagat tttaattttt aatatttaca
     1921 ggaactaaat attctatagt atatataaac tagaattgtt tttaaattga tctttttaaa
     1981 aatatttaag ctttcatgaa tatctctgag ggtttttact gtttaaatta gtattaagtt
     2041 ttaattttta ataaggaact cggcaaaaat ttttcctcaa ctgtttatca aaaacatagc
     2101 tttaagaaag ttagaataaa tttaaagttt cacctgccct gtgattttat ttaacggccg
     2161 cagtactttg actgtgctaa ggtagcataa taaattggcc tttaattggg gtcgagtatg
     2221 aaagggatta tggggaaata ctgtcttttt aaaaattaca taaaatttac tgaataggtg
```

2281 aaaataccta tagttacttt taagacgaga agaccctaag aattataatt aaaggatatt
2341 agttcttcgt ttttgttggg gcaacattgt taggaaatat aaccaacatt aataaagttt
2401 tacatgtcta ctataattag aaaggtttta attaccttag ggataacagc ataattatat
2461 ttatagattg tgacctcgat gttggactag ggactttagt tattatccat tactagagat
2521 agttctgttc gaactgtaaa accctacatg atctgagttc agaccggcgt aagccaggtc
2581 agtttctatc taatatagga ggtatagtag tacgaaagga ccctatatct acaaattatt
2641 tgttttaact tggcagaatt aatgtaattg atttagggtc aatcaatacc atagtggtag
2701 ttataattaa aagaagttat gaaataatat taactttggg agttaaagac tgaggtaaac
2761 tcacttttag ggaatatact ttaaacatag aaggtgttgc tttgcatgca ataagtaaaa
2821 aaattttta tttccaaaat ttaatagatg aactcatcta gagatgattg ttctttttgg
2881 gttaggaata tttttagtat ttagtttttg tttcttacga agtccagtct catatggggg
2941 tgccttgatt gctcttagaa caattggtgc tttattatta agtttaggtt gcagtacttg
3001 atatggttat ttattatttt tagtttatgt ggggggcctt ttggtattgt ttttgtatat
3061 tattatatta agaagtaatt tcaatctaca agtaagattt aagttaatag gtttaatttt
3121 tttggctttt ttagttagta aattatataa ttttaggtac cctaaaaaca ggctaggtgt
3181 aagattaagg gaatgtagag aagattttag tttaggttta tttttaggtt tagggggcct
3241 tttgctatta gtgtttttttg caattgttca tattgttttt ttgaagggcc aaccagtcca
3301 agtgaaaaat gattaaaata tctcggttgc cagtgttatt attattaagg tgtttgcttt
3361 taatgagttt attttttat agtgtcttta aaagaggtga aacatatgta attgaactta
3421 atttaataac cttaagaaga gtaaggttta gaactttatt tatttttgat aaaattagtc
3481 ttagatttgg agcggttgta acattaattt ccttttgtgt ttttagtttt gctaatactt
3541 atatgagtga agataacttc cagtttcgct ttattttgat tcttgctata tttgttttat
3601 caataaacat tttaattttt aggggatcat tgtttgttct attacttggt tgagatgggc
3661 tgggaatttc atcttttgcc ttaattatct attatcaaaa taaaactaga ttaggtgctg
3721 gttatctaac tcttctaact aatcggcttg gtgatgtggt tattattatc tgcgttccat
3781 ttttttaat tctaggaaga tttaatattt tccccattat atctaggtat gaaacaatta
3841 tttttatttt agccattgcg gctcttacta aaagggctca atatcctttt agagcttggc
3901 ttccggcagc tatggctgcc ccaactcctg tgagtgcact ggttcactct tcaacactgg
3961 tgacggcagg cgtctatctt attatccggt taagggttaa ttgtagtata agcaattcta
4021 cggcttcact gcttttattt tgtgggaggg taacctgcgt tttgggcggg ttaagagctc
4081 tttatgagaa cgatctaaaa aaaattattg ctttttcaac cttaagacaa ttagggttaa
4141 taatattttc tttagggtta aaacttccaa acttagctct cttacattta tatgcccatg
4201 ctatatttaa agctcttttg tttttaagtg cgggattaat cctaattagt gggtttgggt
4261 ctcaggattt acgtctccta ggggcaacgc tctacactat cccaagagtt gttgtttttt
4321 ttaatattag aagtttatgt ttaatgggaa tcccttttct tagggcattt tactccaaac
4381 atgtaattta tgaagtggtg gctataagag agactaacgc aattaggttt tttttaatat
4441 taattggggc tattcttact acaatttatt caattcgaac ggtaaaaatt ttaagatgaa
4501 attctatttc tggtctatcg gtttttagcc gattacccct ttatacatat gtacctttat
4561 gcattttgtt ttttacgtca attattagag gaaaaacatt tagaatgtta gacattaggt
4621 acttaacatt catgtttact tcaaggtatt atcagttaat actccatttt ttaatggctt
4681 taggagtaat aattgggtta tttttatctg gtgaaaaaaa aaggcatatg cttagaagta
4741 tgtttttttt atggccaagt tctaataact taacaaaata tttttggccc gtgttaaaga
4801 gagcaaaaat attagactac ggatgagtag agccagtggt tcttttaaga aggtcattga
4861 acaagctagg aaataaattg catttttat ttttaaatga gtccaaatat ttaaatagaa
4921 tagttcgtgg agttttttgt agattaacag ttttcttttt atgtatttat ttattataaa
4981 agctgtaatt ttatgtttat gtgtattatt agctgttgct ttttttactt tacttgagcg
5041 aaaagtatta ggttatatac aaattcgtaa aggtccaaat aaagtgggta tttgaggaat
5101 tgttcaaccc ttggcagatg ctttaaaatt atttattaaa gagaaagttt tgttatatat
5161 atctaacaat ttgcttttta tttttacgcc tatactagga tttttttgtga gtttaaggct
5221 ttggaattta gctccaagta acataatagt aaattatata atttttgggc tattattatt
5281 tttttgtttg acggggttaa atgtgtacac aacggcacta gctggatgat catcaaatag
5341 cctttatgct tttcttgggg caattcgagc ttcagcccaa tcaatttctt atgaagtaag
5401 gttagtgata cttcttcttt ttcctgcttt agttagtaga tcaatatgtt gatttagagc
5461 ccttcatacg ccccccttctt ctcttttagt ttttcctttg ttggcaattt gatttgctag
5521 atctctagct gaaacaaatc gagcgccttt tgattttgct gaaggtgagt ccgagctagt
5581 ttcaggcttt aatgttgaat acgaaggtgg gctttttgct cttctttttt tggctgaata
5641 taccagtatt atttttatat caatggccac cacagtctga tttctttctt cttgatttac
5701 tattgacttt gttttaatta cattaataat ttgtatttga ttcctatttg ctcgaggtgt
5761 ttatcctcgc tatcgttatg acttattaat gatactatgt tgaaaaagct acctcccctt
5821 tgcaattaga cttttactct acttttgttt tagaattata ttttaatatg aggcaacttt
5881 tattcattct tatttatggg tttgttattt tactttactc tttttcctct caaagaaaac
5941 atttaatttc tagcttaatc attttggaga gcttaatagt aattagaatt atttatcttt
6001 tttttagatt ggggagtgta aactccattg gtcttttatt actaatttta acttttgctg
6061 ccgctgaagc ggcgttagct ctttctcttt tgattagtat ccttcgggtt ataaaaaatg
6121 ataatatatt aaacttaagt tattaagtct tgcgaaaaat tagattatta gaaggtttga
6181 caagacttcc aactccttta aatattagta ttttttgaag cggtggttca attttaggaa
6241 taatattagg tattcaacta gtaacaggat tatttttgtc aatacattat acaagggata
6301 tttctatttc atttgattct gtagttcata ttatgcgaga tgttcctaat ggttgacttt
6361 ttcgcttatt tcatgccaat ggggcttctt tctttttat tttttatac ctccatattg
6421 gtcgggggtt atactaccaa aggtacagtg tgcaacctaa aacatggata gtcggggtta

6481 caatttttct tattagaata gctactgctt ttttaggcta tgtgttacct tgggggcaga
6541 tatcattttg aggtgccaca gttattacca atcttttatc ggcaattcca tactgaggcg
6601 gaaccttggt tgagtgagtc tgagggggat tttcggtagg gcaggccacc ctaaatcgat
6661 tttttccct acattttatt cttccatttt taattgcttt gcttgttata gtccatttaa
6721 tttttttgca tgaaaaaggc tcaacaaatc ctttagggca gactcatcat attaataagg
6781 ttacttttca cccttatttt acatataaag atttagtagg ttttatttgt gtcgggctta
6841 gattagtttt aatttgcgct tattttccat ttagtttgtc agacccagaa aattttattt
6901 atgctaaccc tatagtaact cctactcaca ttcaacctga gtggtacttt ctgtttgcat
6961 atgctatttt acgttctatt ccttctaagt tggggggggt tgtagctctt gctataagaa
7021 tttttatttt atattttctc cctttaagtt ctcaatttat gtccgttcca gctagcttca
7081 acatgattca gcaaatgtta ttttgatgat tagtcgtagt atttttactt ttaacatggt
7141 taggggcatg tcccattgaa gatccttatt tatctttggc ggggccttta acaattattt
7201 atttttttag ttacccctta ctaataattg ttggaaactt ttccgtgaga attcttaaat
7261 aaaccttttt acctagttta tttataaaac actcgtttgt caggcgggag aaaccgtaaa
7321 acgtggttga gtagtagttt aaaaaaaacg aataatttgc aaaattattt ttggtagagc
7381 cctacttata aacagatagc ttaaatttta aagcgtagct ttgaagcagc taagaaaagc
7441 atagtctttt tgtttaatga gttactgagg ttcattaaat ttaatagatc caagttctcc
7501 aattcaaatt gaaataatac tttttcatga tcatgcaata attttattaa ttggaatttt
7561 tacgttagtt gctttattag gtgtaaaact ttgcttaaat aaatttagaa ggcgggcaac
7621 acatgaggcc caattgttag aaattttttg aacaattatt ccagcacttt tattaatttg
7681 gttagcttta cctagtctac gacttttatt tcttctcgat gagcaaggga gcaatgggtt
7741 aattttaaaa gctacgggac atcaatggta ctgaagttat gaaattccga gaagaggtac
7801 tgactcgttt gattcttata tagtacagga acaagcttta actaacggtg aattccggct
7861 tttagaagtt gataaccgtc cagttttacc ttattctaat gaaattagag tgttaactac
7921 aagttcggat gtaattcatg cctgagcctt accttctatg ggagtaaaaa tagatgcagt
7981 tccagggcga ttaaatataa taggtcttta ttctcttact cctggtgtat tctacggcca
8041 gtgttctgaa atttgcgggg caaatcattc ctatataccg attgtagttg aatttgttca
8101 cttgaacgac tttattaatt tttttggaat cttttgttta agaggcagtt tgtaaaattg
8161 ttttagattt ttatcccaaa ctctaaggtt taaactgctt atttgtggtg taagagatcc
8221 gcccaaacga cggagagttt cagactacgt agtataatga gtacctctgc cttccaagca
8281 gaaagtctta atcatagagt agtaatatag ctgtaagata ataaaactat cgaccttcaa
8341 agtcgaatat atgcgtttgc gttagctatg gtttctccta agcttctgct ttcttggaaa
8401 atcaaatgtt ccggtgctaa acaccaaaga gaaataaagg gtaaaaccca tttaaggcgc
8461 ttaaaacctt tgcatttatt ctgccatcta taaaagtcta attaaaaatt ttatgagtta
8521 atttttgaga aggaatttga ggttgtttta ctggaagggg gcaaaagttt aaagcaaaaa
8581 ttaaaattga taaaatgagc aaaactgtgc agaaaataaa tagaccagat gcagggctaa
8641 gttggggcat atctaaataa caattttgtt tcttttaatt aacatttaaa tgctttttat
8701 aagctatatt tagtaaaatt taagggtgtt cggctatata taaactaatt aaaagagtaa
8761 aaatgtaagc ttgaataaaa gaaacaaaca tttcaaataa ataataaaaa attataacag
8821 atgccgctaa agagaaagga ataaaactta aagtacttgc taaaacatta gcaattaacc
8881 ctagaataat atgacctgct ctaatattag ctactaatcg gacagtcaaa gtaattggtc
8941 gaattaaaat tcttacggtt tcaattaaga ctaaaaatgg aattaaagcc ccaggagctc
9001 ctgctggagc taaatgagca gcagactttt taaaagaaaa aactcaacca gaaattaaaa
9061 tagaacctca taaaactagg gccactcttc tattaattca aagtcttgtt gtagttctaa
9121 aaactaaggg tagaagtcct aaaaaattgt taagtataag gaaacttatt actgctgtta
9181 aaagtaattt ttttgaaaaa aaagtctctt ctgttcccca taagctacaa atagaattac
9241 ctagaaccga aaaaaaagaa gatataaaag attgattaaa aataaagaat gcggctgctg
9301 caagaggtaa taatcataaa aaaaaggaag tgttagcgtc catggaagaa aaaagatctg
9361 ttatcataaa gaacgtaagg aaaaccttaa aattttaagg ccgaaactta aacactgggc
9421 gtttcgttct tttatctaaa aagaaagatt tcttattttt accacgaaaa ggtaatgtaa
9481 atttatacta tctagactta gctttatttt agctagagac agcttccgct acctctacta
9541 tgttacgact tgtctcattt tttaacgaga gtgacgggcg atttgtgcac tattaaaaat
9601 attattcaaa ttttataaac tatttttatt ttacttataa gtccatcttt ataaaaattt
9661 ttaaattttt attccgaact aaaaaggtat tgtaactcgc ctaaacataa ttataggctg
9721 caccttgatc tgtggttatt aacaaaatta ttaattaaaa atctttctaa aaaagatttt
9781 acacgacggc atacaaactt taaaattatg ttgatatagc ttggtggtta tcaattacta
9841 ggtaagttcc cctaataggt tttaaaagcc gccatgattt ttaagtttca acattaagat
9901 gttttactcc tccttatttt ttctttacca taggataata gggtatctaa tcctagtttc
9961 tcttccaggt taaaataaaa taaaaacttt ccttttggtg atactttgct aatgtatttc
10021 actactatta gtaagtaaaa agaacagttg taaaataaga ttttaagtta atttttactt
10081 aggtttgacc gcggttgctg gcacctaaaa taccataaaa tacactgtaa cctaaattta
10141 tgtttcaata atttattaaa atttttaact ggggttactt aaatttccat aattaaatat
10201 acagtgaact aactttaccc atcaaatcag gtagttctaa aaaggggtta cccatttttg
10261 ggttatgagc ccaatagctt attgtagctt atttttagtt tatttaattc aatctaatgc
10321 gcctatggct cactcataaa aaagcccccc caaaagtatt agcaagaatg ttctataggc
10381 agctaatact acagttctgg tggctatttt tatttgagtt aaaatgggaa acaaaaggac
10441 cacttcaaca tcaaaaatta aaaataaaat aattaataaa aagaatcgta aagagaaggg
10501 ttttcgtatt gatctcatag gatcaaaacc acactcaaat ggggactctt tagatgacct
10561 cggacttgct gctttccata gggtaaaaaa atataataaa actaaaatgg ttgttaaaag
10621 aaaaggaaaa atacaatagt atatcacttt tagtggttgc actaaacggc taagattaag

```
  10681 ccaaaaaaga ttttcaaaat cctttgtaaa ataaaaaata aagaggtttt taattgaggc
  10741 tgctaacttt gattaggggc ttattgctcc gcctttttat gggtcttgct attgtttttt
  10801 taactatctt aggattattt tttagtagtt gggatgtggt agtaataaca ttggttgttg
  10861 gtgtattact tagattttta aatatatatt taaattttag gtttaattta tttagtgatt
  10921 tgtttttaac tgaaagacta agaagaatat tattatttct aagattatta attagtattt
  10981 tagctatttt gtgcacttga tgaattaaat ccacaaaata tattataact gtaatagcat
  11041 taaatttatt tttagttttc gcctttagag taaaaaattt attggttttt tacttttttt
  11101 ttgaggcctc tcttattccc actttagctt taatcattat ttggggatac caaccggaac
  11161 gactacaagc tggaagttat ataatacttt atactgtact tgcctcattc cctttattat
  11221 tttttatctt atgagtaaat tgaaacataa ggaggctgga tttcttttta ttaaaaatgc
  11281 tagggagttc tttagggggg gcaatactta tttttaggag gctaggtttt cttgttaaat
  11341 taccaatatt tagggtccat ttatgacttc ctaaagctca tgtggaagct cctttaagtg
  11401 ggtctatgat tcttgctggt gtattattaa agttaggggg ctatggattg tacttattaa
  11461 attattcttt ttctttaatt gtactaggca ctagaaggat aattatttgt tttattgctt
  11521 tatggggtgg attatttgcc tctttaatat gtttacaaca gacagatata aaagctttgg
  11581 tggcatattc ttcagtttca cacataggat tagttattgg aggtgttttt ttaaatttaa
  11641 actgaggggt tgtatcagct aaaattacga tagttgctca tgggtttact tcatctgcat
  11701 tatttgtttt agtaaacata agctataaaa aagttttgag tcgtagattt tccgcaagag
  11761 gtgggctact taggattttt cctaaaatgg ctttattatg atttttattt tgcagaatta
  11821 atatggcggc tcctccttct ctaaatttag taggggaact ttatttactt ccggtactat
  11881 ttttatttag taacattttt gttttaatta tagctttgtc attatttttg agggctgggt
  11941 ataatataat actgtacact aatggaaacc acggagctac aaataattta ataaccccaa
  12001 gaacaagata tagaacagct gattattttt ctttaatttt acatttattg ccattttttt
  12061 tcttattaaa attaaacata ttttattgtt atagtagaac actactggat gaaagctaat
  12121 gggaatttcc tttattgact tacaaaatca agattttata taaattacat tagcaagatc
  12181 ctcatcaata aattctcaca taaagaaata atcagacgac atcaacaaag tgtcagtatc
  12241 aggcagctgc taggaaccct acgtgatgag ttgtgttaaa gtggtagaga attattcgaa
  12301 taaggcacac aaacaggaaa gtagcaccta cggctacatg aagcccatga aatccagttg
  12361 ctataaaaaa agtactacca taaacactat cagcaattga aaaacttgtt tctttatatt
  12421 caccaaattg aagaaataaa aaataaagtc ctaacaatgt tgttaataag agtcctctaa
  12481 tggcggaatt gtattttcct tcttcaatcg catggtgggc tcatgttacc cttacacctg
  12541 aaagtaaaag aactgaagta ttaaggagag gtacttgaaa ggctcttagc gtagtaatgc
  12601 caacaggggg tcatactctc ccaatttcta ctgaaggggc taaacttctg tggaaataag
  12661 ctcaaaaaaa ggcaaaaaaa aaacaaactt ctgataaaat gaagagtatt acacctaatt
  12721 ttagtccttt gactacatag gtagtgtgat gtccttgata tgttgcttcg cgaacaatat
  12781 cgcgtcacca taggtatgct actatagcag tacatgctat tccgtaagct agtaagtaag
  12841 gtgttccaat tcgaatataa tagattaatc ctaaaggtat gcaagttaaa agtgccgacc
  12901 ctagcaaggg tcaagggcta aattctacaa gatggaatgg agttttaaac atattaagtt
  12961 aattgtaaaa gatttttaat ggttataaaa attcttaagc aaccgccggc agagcgggta
  13021 tcattgatgt tgatatatat gaatatttta attgttcgtt gctataaata atgagttgat
  13081 atggtgtttt attttctgta attgtttttt tagggccatt aattagttta tctacttcca
  13141 attggttatt agtttgaggg gggatagaat tatcacttat taggttactt cctatcttaa
  13201 atctagctaa aaggtctatt cctgtagaaa gggcgataaa atatttttttg gttcaggcaa
  13261 gtgcaagagc agttatttta attaccggat tgaatattta tatattagaa gggagggaat
  13321 atttaatttt gattttattc attttagggt taatcttaaa gctaggagta gtacctttac
  13381 atttttgagt aattccggta gtaaaaggac taaaatatat tcatatagca ttcttactag
  13441 ggccattaaa aattgttcct ttaatattat taataactaa tataaatttg gataggtcta
  13501 tttttttatt tttacttgta aggagaattg gtaccgtttt gctaggtgct ttattaggta
  13561 ataatataag atctattcga ataatacttg gagcatcttc tattagtcat tctgggtgat
  13621 ttttaatagg gggattagtt ggctttatat gaatctattt taggatttac ataattagct
  13681 tgtatgttgt cttatattat ttaaaagtat atcctagtga taactatctt gtatctattt
  13741 ttctatttag tttaaggggt ctccctccat ttttaatatt tccagcaaaa ataatagttt
  13801 tatattcgct ttttaagaatg ggcttatctc ccattgtttt tagttttttt atcatcagcg
  13861 ccattgtgag gctaaatttc tatataaaat ttagtttttc atttgtatta agaagaagag
  13921 aaaatcaaaa aaattttggg gctggtgtga tgccaggatt gctactaata atggggggtt
  13981 tcattttatt ttttatttag ttttttatgg 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="MRWLYSTNHKDIGTLYMIFGVWCGMVGTGLSLLIRLELGTSGVL
                     MDDHFYNVIVTAHAFVMIFFMVMPIMIGGFGNWMVPLLIGAPDMSFPRMNNMSFWLLP
                     PSFILLISSSMVEGGAGTGWTVYPPLSSMVGHSGASVDLAIFSLHLAGMSSILGAINF
                     ITTIFNMRASGVTLERLSLFVWSILVTVFLLLLSLPVLAGAITMLLTDRNFNTSFFDP
                     AGGGDPILYQHLFWFFGHPEVYILILPGFGIISHILGNASSKQPFGSLGMIYAMISIG
                     ILGFIVWAHHMFTVGMDVDTRAYFTAATMIIAIPTGIKVFSWLMTLYGMNNKFDASMY
                     WVLGFIFLFTLGGLTGIVLSNSSLDIVLHDTYYVVAHFHYVLSMGAVFAIFAGFNYWF
                     PVMTGLILQERLAKAQFIIMFIAVNLTFFPQHFLGLAGMPRRYSDYPDTYFKWNQISS
                     FGSLMSIFAVLLFCLLVWEAFLSERSVLFSKAAIYSREWVENNFPPEFHSNIIPSISV
                     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
                     SWYGLMLFLIYVGGLLVLFMYVIMLSSNFILSFSLKTTIKGFFVLMFMWALKNFNLPS
                     KILAESMSSCSSSLMMGLFLSLGVFLLLVFFSIVHMVLFKGVILKLNE"
     gene            3333..5000
                     /gene="ND6"
```

```
CDS          3333..5000
             /gene="ND5"
             /codon_start=1
             /transl_table=5
             /product="NADH dehydrogenase subunit 5"
             /translation="MSKIYRLPFLLLFGLLFSLALSSVMFFLSFKSFMVDIKLIISGS
             SWITIMLIFDKISISFSMVVLCISFCVFLFMCQYMENDEFKFRFGMILLLFVLSMNIL
             IFSGSFFMLMLGWDGLGISSFALIIYYSSASSKQAGFLTLMTNRIGDVLIIASIPFFL
             SGGNFMFFSMEILPYWVVVLLCVAALTKSAQYPFSAWLPAAMAAPTPVSALVHSSTLV
             TAGVFLIIRVSMNFNLSDEICSLLIFSGGVTCFLGGVSALYENDVKKIIAFSTLSQLG
             LMVFCLGMNMPFLALLHLFTHAMFKAMLFLSAGLILLSSFGVQDLRILGALQFRSPLL
             ICFFNVSSLCLMGAPFMSAFFSKHVIYEIMFMSSVNTLGVALMVVGSIFTSIYSVRLV
             KIISWKKISMCSINLSLTIFSYFPMIILYFMSLVSGKMFSLMDISHNIIMFTPLTVQV
             LLNFLLVVGILVGLMMKGERISHLFGSMLFLWPSSHNINKFMTRSMKLTTSLDLGWIE
             PKNILNNFYLESSFFLENLSNYMKFSLLSLFKGSAFMFMVLIMWYFF"
gene        4985..5890
             /gene="ND1"
CDS          4985..5890
             /gene="ND1"
             /codon_start=1
             /transl_table=5
             /product="NADH dehydrogenase subunit 1"
             /translation="MMFFLIKSLFFCLCVLIAVAFFTLLERKVLGFLQIRKGPNKVGV
             WGIFQPLSDALKLFVKEKIFKYSSNNILFLIIPPFGFCLSLSMWNLIPSAYNMNFILY
             SLLLFFCIVGLNVYVTALAGWASNSIYAFLGAMRASAQTISYEISLVMLLLFPALISG
             SMCWFFNMKSFPVSLLMFPFLLILITSSLAETNRAPFDFAEGESELVSGFNVEYESGL
             FALLFLAEYTSILFMCSMMTVWLMWSNTPLSFCLWTMFFSLIFLWSRGCFPRFRYDLL
             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="MRKIKIIESLSALPSPVNISVWWSGGSLLGMMLGIQLITGLFLS
             MHYVADTSLAFDSVTHIMRDVPMGWVFRLFHANGASFFFIFLYLHVGRGLYYQSYILQ
             PKTWMAGVTIFLLSMAAAFLGYVLPWGQMSFWGATVITNLLSAIPYWGEILVQWVWGG
             FSVGQATLNRFFSLHFLVPFVILVMVLLHLVFLHEKGSSNPLGHNFHINKIAFHPYFT
             YKDLVGFVLVFSLLIFMCSFYPFYLSDPENFMNANPMVTPTHIQPEWYFLFAYAILRS
             IPSKLGGVIALAMSLLILYFLPFLHLGKSIPSSFNIFYQVNFWILVVIFILLTWLGAC
             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
             LSVNKLSTRTTHEAQTLEILWTILPGFLLIWLALPSLRLLYLLDEQGNNGLILKAIGH
             QWYWSYEIPSSSTSSFDSYMVQEKDLSKGEYRLLEVDNRPVLPYNTPISVLTTSADVI
             HSWTLPSMGAKVDAVPGRLNLIGLSSYFPGVFYGQCSEICGANHSYMPIVIEFIKLED
             FIQ"
tRNA         8052..8117
             /product="tRNA-Tyr"
tRNA         8114..8174
             /product="tRNA-His"
tRNA         8180..8245
```

```
          /product="tRNA-Gly"
tRNA         8248..8313
          /product="tRNA-Trp"
tRNA         8319..8385
          /product="tRNA-Asp"
tRNA         8381..8446
          /product="tRNA-Cys"
tRNA         complement(8444..8505)
          /product="tRNA-Gln"
tRNA         complement(8502..8565)
          /product="tRNA-Leu"
gene        complement(8569..8718)
          /gene="ATP8"
CDS          complement(8569..8718)
          /gene="ATP8"
          /codon_start=1
          /transl_table=5
          /product="ATP synthase F0 subunit 8"
          /translation="MPQLSPSSGLMTFLFMVFIIFINKLYFQNTPKSVSTFKISPQVK
          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="MMTDLFSSMDANYSFLLWSFPLFMTTFILMKTYYLNSLTSVFKA
          SLTNLWTFKESYFWSKLFLTLAMSFIMLCNFSGLLPYVYGSTTSLWVNSSMALLFWGI
          FLLSGYCFSFKKSLAHLTPAGAPLMLIPFLILIESISILIRPLTLTVRLVANISAGHI
          ILALMANVLSSLQALVPFTGVILVLMFYNLFEMFVSFIQSYIFTLLLSLYLAEHP"
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="MVLWVFSISLVFMGHWLLIFLMLVFFSVFSFLNLYKNFSYDVFN
          SWIYLDSMSSFLVLLSVFIVLFSFVGTYWNKSVIYTKSLLLLVFFLFLAFSVKNLLCF
          YFFFEASLIPTLFLIIAWGYQPERLQAGTYMIMYTVLSSMPFLFFILYCFYSLNTLEV
          VLLMYSKMEISSIVFYIIVLAFLVKLPMYGLHLWLPKAHVEAPLSGSMILAGVLLKLG
          GYGLYLLNYILSLKNLNTNLVVLSFLSLWGGLLATLMCFRQSDMKAMVAYSSVAHMSL
          VIVGMFLNSNWGLLSAKITMLAHGFTSSALFMLVNMSYKKIYSRSFNTSGGLLTYFPK
          MSLLWFIFCSINMAAPPSMNLLGEMMVIPSLKLWGAIFVILMCIIMLLSAGYNMMLYV
          NNNHGGPSFYVTPSMNYNSSEYYGLLMHVLPFFLLFKMELF"
```

```
    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="MFKTPFHLVEFSPWPLLSSLALTSMPLGLIYYLRFNYWVLLLFG
                    MLLTAFIAYLWWRDIIRESTYQGYHTTYVVSGIKLGVMLFIVSEVCFFFAFFWAYFHS
                    SLAPAIEVGSVWPPMGITTLPTFSVPLLNTSVLLLSGVSVTWAHHAIEEGKLKSATSG
                    LLVTVLLGLYFLFLQFGEYKETSFSIADSVYGSTFFMATGFHGMHVLVGATFLLICLI
                    RLMLYHFNKSHHVGFLAAAWYWHFVDVVWLFLYVSIYWWGS"
    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
                    YPKSSMSSESTIKYFLVQSCASAMILVSGSFLFIVEMTNMFYYSLFLTGLMLKLGLVP
                    FHFWVIPVMKGCKFTDMIFILGPMKIIPFLLIILSIPISSNLLLIGIFMSLFSLVIGA
                    LLGNNSSSVRGMLAASSIVHSGWLLMGCLVGGMGLYFSIYLMQLYPFLLFCLLNQKFF
                    MSVLMLSFSGLPPFMLFFGKLYILQNIIYKEVLVSLLFTFIIFSALISLNFYLKFSYS
                    FYLSSMLNLKNKIFITNLMMISSLLGFILFIFFYGRAKH"
BASE COUNT     4601 a   1830 c   2161 g   5557 t
ORIGIN
        1 tttgcgatga ttatattcaa cgaatcataa agatattggt actttatata taatttttgg
       61 agtttgatgt ggtatggttg gaactggtct ctcactactt attcgtcttg aactaggtac
      121 ctctggagtt ttgatagatg atcattttta taatgttatt gttacagctc atgcttttgt
      181 aataattttt tttatagtaa tacctattat aattggagga ttcggaaatt gaatagtccc
      241 tcttctcatt ggtgccccag atataagttt tcctcgaata aataatataa gtttttggct
      301 tcttcctcca tcttttattt tacttattag ttcaagaatg gtagaaggag gggcaggaac
      361 aggttgaact gtataccccc cattaagaag aatagtagga catagaggag catcagtaga
      421 tttggcaatt ttttcattac atttagctgg gatgtcttca attttgggag ctattaattt
      481 cattacaaca atttttaaca tacgagcttc aggggtcact ttggaacgtt taaggctatt
      541 tgtatgatct attttagtca cagttttttt acttctttta tcattacctg ttctagcagg
      601 ggctattact atactattaa cagatcgaaa ttttaacaca agattttttg atccagctgg
      661 aggaggggat ccaattttat atcaacattt attttggttt tttggtcatc ctgaagtata
      721 tattttaatt ctacctggtt ttgggattat ttcacatatt ttaggtaatg catcctcaaa
      781 acaacctttt ggaagtttag gaataattta tgctataatt tctattggta ttttaggttt
      841 tattgtatgg gcccatcata tatttacagt aggtatagat gttgatacgc gagcttattt
      901 tacagctgct actataatta ttgctatccc tacgggaatt aaagtttttta ggtgattaat
      961 aaccttatat ggaataaata acaaatttga tgcctcaata tactgagttc taggttttat
     1021 tttttattt acattaggag ggctaacagg gattgtttta tctaattctt ctttagatat
     1081 tgttcttcat gatacttact atgttgttgc tcattttcac tatgttttat ctataggggc
     1141 tgtatttgca atctttgctg ggttcaacta ttgattccct gtaataacag ggttaattct
     1201 tcaagaacga ttagctaaag cacaatttat tattatattt attgcagtta accttacttt
     1261 ttttccgcaa catttttttag gattagcagg aataccacgt cgttattcag actaccctga
     1321 tacttatttt aaatggaatc agatctcgtc ttttgggtca ctaatgtcta tctttgctgt
     1381 acttttattt tgtctattgg tttgagaggc ttttttaaga gaacgaagag tattatttag
     1441 caaagctgca atttactcac gagagtgggt agaaaataat ttccaccag aatttcacag
     1501 gaatattatt cctagaatca gggttcgtta attaataact tactagtaaa atattaatgt
     1561 atatatttct tacaaaaata aagtggcttc gagcctacta gttattaagt agtattgagt
     1621 gatttaatct atataatatc ctttgaaaaa ttaattctat aattattagg taatatcata
     1681 ccttttgcat aatggttaaa taaaaataaa tttttattaa agtttcccga taataaaaga
     1741 tctatttatt tttctatatt agtcttgtct ctaaatttta ttgttgcata aataagattt
     1801 taaataataa atagtagtga aatgctaaca attttattta tatctggata gtttttaaaa
     1861 gcatttagtg tttaggcaac ttttaggtga taagatctaa aagcttaaga atattaaata
     1921 aaacttattc tctaaagatt tatgtttttta tagaaatgat atcttaatat aataactttt
     1981 attaatcttt atagtaaaaa cttttttttat ataattttta tagaaattaa taatgtttaa
     2041 attagtaaaa attttattta actaattaaa ggaactcgac aaaaaggttt tttgaactgt
     2101 ttaacaaaaa catagcttaa agtttatatt ttaggtgaga cctgctcaat gaagatgtag
     2161 attatttaaa tagccgcagt actttgactg tgctaaggta gcataataaa ttggctttta
     2221 attgaagtct ggaatgaaag gtaaatttga aaaagattgt ctctaattag atttattaaa
     2281 cttacttaat aggtgaaaat tcctatgttt ttcataatag acgagaagac cctaagaatt
```

2341 attaaaaata agtttatttt actttaattt tttgttgggg caacaacttt gaacataaaa
2401 ccaaagggtt aaagattcta agacaaattt tattttgaaa aaaattaatt accttaggga
2461 taacagcata attttctttt tagattgtga cctcgatgtt ggattaagga ccccttaggt
2521 aagccactta agagaatagt tctgttcgaa ctttatatcc ttacatgatc tgagttcaga
2581 ccggcgtaag ccaggtcagt ttctatcttt tatttttaaa ttgttagtac gaaaggacca
2641 ctttttagta ttttatatat tttagtttga cagaaacata tgtgattgac ttagaatcaa
2701 tatatacctt ttatggtaac tattaaaaga agttttgtaa aacattaact ttgggagtta
2761 aaaattgaag aaagctcttc acttttaaag gttaagtatt ttataattag aaagtatatt
2821 ttgcaagtat aaagaaaacc aaaaggtttt ttaaccattt gcttttaaat tgtatacaat
2881 ttaaaaattg agattactat ttttaatttt actagcctct ttcattatgt tttttacttc
2941 tgaacaaccg gcagctttag gggggagatt aattttaata agttttctta gagcgttgag
3001 ttttaggttt atatttagaa ggtggtatgg attaatgtta tttttaattt atgtaggagg
3061 gttattagta ttatttatat atgtcattat actaagaagt aatttcattc ttagattttc
3121 tttaaagaca actatcaagg gttttttttgt attaatattt atatgggcct taaagaattt
3181 taatcttcct tcaaaaattt tagcagagag tatatcttca tgtagctcaa gcctaataat
3241 agggttattt ttaagcttag gggttttttt gttgttagtc ttcttttcaa ttgttcacat
3301 ggttctattt aaaggagtaa tcttaaaatt aaatgagtaa aatttatcga ttaccttttc
3361 tgttactttt tggtctttta ttttccctag cactatctag agtgatattt tttcttaggt
3421 ttaagtcttt tatagtagat attaaattaa ttatttcagg aaggtcatga attactatta
3481 tattaatttt tgataaaatt agaattaggt ttaggatagt agttctatgt atttcttttt
3541 gtgtgttcct ttttatatgc caatatatgg agaatgatga gttcaaattt cggtttggaa
3601 taattttact tttatttgtt ctttccataa atatcttaat tttctcaggt tcttttttta
3661 tattaatact tggctgagac ggattaggaa tttcttcgtt tgccttaatt atttattaca
3721 gaagagctag aagaaaacaa gcagggtttt taactttaat gactaatcga attggtgatg
3781 tattaattat tgctagaatt cctttttttt taaggggagg aaactttatg tttttttcaa
3841 tggaaatttt accatattgg gttgtagtac tcttatgtgt tgcggccctt acaaaaagtg
3901 cccaatatcc ctttagtgcc tggctccctg ctgcaatagc agctcccacc cctgtgagag
3961 ctttagttca ctcatctact ttagttacag ctggggtatt tttaattatt cgagttagaa
4021 taaattttaa cttaagggat gaaatttgtt cattacttat ttttagcggt ggagttactt
4081 gtttttagg gggagtaaga gcactatacg agaatgatgt aaaaaaaatt attgcatttt
4141 ccacattaag acaactagga cttatggttt tttgtttagg gatgaatata ccatttttag
4201 ccttacttca tttatttact catgctatat ttaaggcaat attattttta agagcaggct
4261 taattctttt aagaagattt ggagtacaag atttacgaat tttaggagca cttcaatttc
4321 ggtctccatt actaatttgt tttttaatg ttaggagatt atgtttaata ggggcccctt
4381 tcataagtgc tttttttca aagcatgtta tttatgaaat tatatttata agaagagtta
4441 acacattagg agtagcccta atagtagttg gctctatttt tacttcaatt tactctgttc
4501 gactagtaaa aattatttca tgaaaaaaga tctctatatg ctctattaat ttaagtttaa
4561 ctatttttag gtattttcct ataattattt tatattttat gtcgttagtg agaggaaaaa
4621 tgttcagact gatagatatt tcacataata ttattatatt tactccttta actgtacaag
4681 tattattaaa tttttacta gttgtcggaa ttttagtagg cttaataatg aaaggagaac
4741 gaatcagcca cttatttgga agaatattat ttttatgacc ttcatcccat aatattaaca
4801 aatttataac ccgatcaata aaactcacga ctagtcttga cttaggctga attgagccta
4861 aaaatattct taacaatttt tacttagaga gaagattttt tttagaaaat ttaagtaatt
4921 atataaaatt ttctctttta agtcttttta aaggaagagc gtttatattt atagtactaa
4981 ttatatgata tttttttaa tcaaatcatt atttttttgt ttatgtgtat taattgcagt
5041 tgcatttttt actttgttag aacggaaagt tttagggttt ttacaaattc gaaaaggacc
5101 taataaagtt ggggtatggg ggatttttca acctttatca gatgcactaa aacttttttgt
5161 taaagaaaaa attttcaagt attcttctaa taatatttta tttttaatca ttccaccttt
5221 tggcttttgc ttaaggctat caatatgaaa cttaattcct agggcttaca acataaactt
5281 tattttatat aggttattat tattcttttg tattgtaggg ttgaatgtat atgtaaccgc
5341 tttagcaggt tgagcttcaa ataggattta tgcattttta ggggcaatac gggcatctgc
5401 tcagactatt tcctatgaaa ttagattagt tatactttta cttttcccag cattaatttc
5461 tggatcgata tgttggtttt ttaatataaa aagttttcct gtatcgttat taatatttcc
5521 atttttatta attttaatta ccagaagttt ggccgaaact aatcgggcgc ctttcgattt
5581 tgcagaagga gaatctgagc tagtgtcagg gtttaatgtt gagtatgaaa gagggctatt
5641 tgctctatta tttttagctg aatacacaag tattcttttc atatgctcta taataacagt
5701 gtggttaata tgaagaaata ccccattaag attttgttta tgaacaatgt tttttagtct
5761 tatctttctt tgaagacgtg gttgctttcc tcggtttcgt tatgatcttt taataatatt
5821 atgttgaaag agattcttac ctttttcaat atgttgttta atttatactg gaattttctt
5881 tcttatataa ttttaatgag tttaataatt attaactttt tttcagtttt tgttattggt
5941 ttcatagttt tactagtaat atttttata caaaataaac atattattaa ttctttaatt
6001 attttagagg gtataatatt aatagtagta tcctttattt tttttatctt aagtttggcc
6061 tctagggggg aagaagtgat attattatta tttacctttt cagcctcgga agctgctatt
6121 ggcttatctt tgttaatcag atttattcga ttatacggaa gggatataat aataaatata
6181 tttggtttat ttgcgaaaga ttaaaattat tgaaagatta agagctttac cttcacctgt
6241 aaacattaga gtatgatgaa gagggggatc tctcttaggt ataatactag gaattcagct
6301 aattactggc ttatttttat caatacacta tgtagcagac acaagattag cctttgactc
6361 tgttactcat atcatacggg atgtccctat aggatgagta tttcggcttt tccatgccaa
6421 tggggcatca ttcttttta ttttttata tcttcatgta ggtcgggggc tttattatca
6481 aagttatatt cttcaaccaa aaacttggat agcaggtgta acaattttct tattaagaat

6541 agcggcagca tttcttgggt atgtgcttcc ctgaggtcaa atatcatttt gaggtgcaac
6601 agtaattact aatcttttgt ctgctattcc ttattgaggt gaaattttag tccagtgggt
6661 gtgaggaggg ttctctgtgg ggcaagctac attgaatcgg ttttttcat tacatttttt
6721 agtacctttt gttatcttag taatggtttt actacattta gtttttttac acgaaaaagg
6781 atcatctaat cctttagggc ataatttcca tattaataaa attgcattcc atccttattt
6841 tacttacaaa gatttagttg gatttgtatt agtttttaga ctacttattt ttatatgttc
6901 tttttatcct ttttacttaa gagacccaga aaattttata aatgcaaatc caatagtgac
6961 cccaactcat attcaaccag agtgatactt tttatttgct tatgctattc ttcgatctat
7021 tccgagaaaa ctaggagggg tcattgcctt agcaataaga ttattaattt tatatttttt
7081 accttttctt cacttaggga aatcaattcc aagaagattt aatatttttt atcaggtgaa
7141 cttttgaatt ttagttgtta tttttatttt actaacttgg cttggggcct gccctattga
7201 agacccttat ttgagactag catttccttt aacagtttga tatttcctta ggtatttttt
7261 cctagctttt agtaataaat taaatttctt tattaggtaa tattaggaaa catagcttat
7321 tttaaagcct ggctttgaag aagctaagat aatatttata tttgttttca catgagacat
7381 tgaggagcaa tcaatttatt agacccatct tcacctgttc aaacagagat aatacttttc
7441 catgatcatg caataatttt attagttggg atttttagct tagtagcttt attaggagta
7501 aagttaagag tcaataaact ttcaactcga accacacatg aggctcaaac attggaaatt
7561 ctttgaacca ttttgcctgg ctttcttctt atttgactgg ctttaccaag tcttcggttg
7621 ttatatttac tagatgaaca aggaaataat gggttgattt taaaggctat cggccatcaa
7681 tgatactgaa gttatgaaat tcctagaaga tctacttcat catttgactc atatatagtt
7741 caagaaaaag acttaagaaa aggagaatat cgcttattag aagttgataa tcgacctgtt
7801 ctcccatata atactcctat tagagtttta actacaaggg cggatgttat tcactcttga
7861 acccttcctt ctataggggc taaagtagat gctgtccctg gtcgattaaa tttgattggc
7921 ttgtcttctt actttcctgg ggttttctat ggtcaatgtt cagaaatttg tggtgcaaat
7981 cattcgtata tacctattgt aatcgagttt attaaattag aagattttat ccaataaagt
8041 attaggtttt aatttgagaa tctaatttag agatagtctg taaatctatt atttggaaag
8101 ccttttttccc tcaaaacttt taggttaatt aaactattta attgtggatt aaaagatttc
8161 gtaagaaagg ttttcaaaga gtttcttagt atttaagtac gattgccttc caagcaagaa
8221 gtctttcttt tagaaaggga aattaaaata tttataagtt aagttaaact agtggacttc
8281 aaatccaaaa atattaaaat ttgattaaat ataagtagtt ttatctagtt taaaataaaa
8341 tttttaattg tcaattaaaa gaaatagttt cctctatgat aggttagtag tttagaaaaa
8401 aaacaggtaa tttgcaaaat tactaaagaa aaattttcct taccttttct taaaaatttt
8461 ttttcctgga aaatcaaatg ttcctgtgct ttacaccaaa gaaaaataaa ggaaataatc
8521 ctatcttaga cgcttaaaat ctatgcattt ttctgccatt ataaagtttt actgataaat
8581 aaattttttt acttgcggtg aaatcttaaa agttcttact gatttaggag tattttgaaa
8641 gtatagttta ttaatgaaaa taataaaaac tataaataaa aaagttatta atccacttga
8701 agggctcaat tgaggcatgc taaactaact attctagttt ctattaatta acatttaatt
8761 gctttatata agctatagtt tagaatttaa ctatgggtgc tctgctaaat ataagcttaa
8821 tagtaaggta aaaatatatg attgaataaa tgaaacaaac atttcaaata aattgtaaaa
8881 tattaacact aagattacac ctgtaaatgg gactaaagct tgaagggatg ataaaacatt
8941 tgctatcaaa gctaaaataa tgtggccggc gctaatgtta gcaactagcc gcactgttaa
9001 tgttaatggg cgaataagga tactgattct ttcaattaaa attaaaaagg ggattaatat
9061 aaggggagcc ccagctgggg taagatgagc taacgatttt ttaaaagaaa aacaatagcc
9121 agacaataaa aaaatacctc aaaataataa agctatagaa ctatttactc atagcctagt
9181 agtggaccca taaacgtaag gaagaagccc actaaaatta cataatataa tgaaccttat
9241 agctaaggtt aaaaagagct tactccaaaa gtaagactct ttaaaagttc ataaatttgt
9301 taatctagct ttaaaaactc ttgtcaaaga gtttagataa taagttttta ttaaaataaa
9361 tgtggttata aataagggaa agcttcacag taaaaaagaa taatttgcat ctattgatga
9421 gaataaatca gttatcaatt taggagttaa taagaatatt aattttttaa ggccgaaact
9481 taatgttatt gaaactacac ctataatact tttaaaaaga taatactttt ttctaaagca
9541 tgaaaagctt ttgtatttta tatactataa aagtaaataa ttgttcatta ttagaaacag
9601 cttccgctac ttctactttg ttacgacttg tctcattttc caacgagaat gacgggcgat
9661 ttgtgcacct ttaaataaaa tgttcaaaag ttatttttaa tttaaacttt tacaatcaaa
9721 tcctatttaa aaaaagtata taaactttct tactaacaaa aatctttttt gtaactcgcc
9781 tataaatcat aattattagc tgcaccttga tctgatatat taaaagatta ttcttttaaa
9841 ttttttttaa aaaaattttt ataacgacgg catacaaact atttaaaaat tatgcgagat
9901 agcttggtga ttatcaatta cttggtaagt tcctctgaag atttttaaag gccgccataa
9961 tctttgagtt tcataatctt tatttactta ctactcaggg ctgtcttata cttcttgtat
10021 aacagggtat ctaatcctgg tttattatca agtgttactt aaaattttaa ttgtttaaaa
10081 gggttaaatt gtgttactac taaaaatttc acttatctat ttagcaaatg tctttttagt
10141 aagccctagg taaaattatt ttttataggt atgaccgcgg ttgctggcac ctattttacc
10201 aaagtaaaaa tataaactaa acttaaattt ctttagtttt taattttgct aactggataa
10261 gtaagattta tcttccataa ttatttttaa tttaataatt tttctatcag atttaataga
10321 aaatttctaa gaagaaaaaa atctccattt ttgggttatg agcccaacag cttaaatagc
10381 ttattcttag atgatttata aaattcactc taatgcccct atctttcatt cataaaaaag
10441 tcctataatt aaaataatta agaaaacggt aaagctaata atttgaggcg tagtacagga
10501 tctctttaat agtattaaaa taggaaacaa tagaacaact tcaacatcaa aaattaaaaa
10561 tagaataatt aataaaaaaa aacgtaatga aaaaggcttt cgaatttttc taaaaggatc
10621 aaacccacac tcaaaaggtt cacctttga gacaatttca gagggcgggt ttcaggataa
10681 aaataaatat aaaataaaaa ataatattgt aatagttaaa ggaatagcga cataataaat

```
   10741 caatttagtg tattcactag atagtaagta aaacttactt caaaagattt tcaaaatcct
   10801 ttatatataa aaatatatct tattaaaagg attataatta gggctgctaa ctttgatttg
   10861 gggacattct cctccttttt ttgagaatag ttttatgggt attttctatt tcacttgttt
   10921 ttatgggcca ttggctatta atttttttaa tattagtttt tttttcagtt tttagatttt
   10981 tgaatctata taaaaatttt tcttatgatg tatttaatag atgaatttat ttagatagga
   11041 taagatcttt cttagtttta ttaagagttt ttattgtttt attctctttt gttgggacat
   11101 attgaaacaa aagagttatc tacactaaaa gactgttgtt attagttttt tttttatttt
   11161 tagctttcag agtaaaaaat ttactttgtt tttatttttt ttttgaggct tcccttatcc
   11221 ctactctatt tttaattatt gcttgagggt accaaccaga acggcttcaa gcaggaacat
   11281 atataattat gtatacagtt ctatcttcaa tgcccttttt gttttttatt ttatattgct
   11341 tttattcatt aaatacgtta gaggttgtgt tactaatata tagaaaaata gaaattagaa
   11401 gtattgtttt ttatattatc gtattggctt ttttagtaaa gttacccatg tacgggcttc
   11461 atttatgact tcctaaggct catgtggagg ccccttttaag cggatctata attttagctg
   11521 gagttttact aaagcttgga ggatatgggt tgtatctatt aaattacatt ctttctttaa
   11581 aaaatttaaa cactaattta gttgtcttaa ggtttttaag tttatgagga ggccttctag
   11641 caactcttat atgttttcgt caatcagaca taaaagctat ggttgcttac tcgtctgtag
   11701 cacatataag tttagttatt gtaggaatgt ttctaaattc aaactgaggt cttttatctg
   11761 ctaaaattac aatattagca catggattta cttcatcagc tctgtttata ttagtaaata
   11821 taagatataa aaaaatttat agacgaagat ttaatacttc aggggggctt ttaacttatt
   11881 ttcctaaaat atctctttta tgatttattt tttgtagaat taacatagct gctcctcctt
   11941 ctataaattt attaggcgaa ataatagtta tccctagctt aaagctttga ggggctattt
   12001 ttgttatttt aatatgtatt attatactat taagagcagg ctacaatata atattatatg
   12061 taaacaataa ccatggaggt ccttcgtttt atgttactcc tagaataaat tataacaggt
   12121 cagaatatta cgggttatta atacatgttt tacctttttt cctattattt aaaatagaat
   12181 tattttaaag ttagtagaga actactaaac ccaatagaat ctagttcatt tttgatttac
   12241 aaaatcatta ttttatttaa attatttttg ggacaaaatt atgaacctca tcagtaaatt
   12301 cttacatata agaataacca tactacatct acaaaatgtc aatatcaagc cgcggctaaa
   12361 aacccgacat ggtgggactt attaaagtgg tatagtatta agcgaattaa acaaattagt
   12421 aaaaatgtag ctccaactaa aacatgtatt ccgtgaaaac ctgtagccat aaaaaaagtt
   12481 ctaccatata cgctatcggc aatagagaaa gaagtttctt tatattcccc aaattgaaga
   12541 aataaaaaat ataaccctag taaaacagtt actagtagtc ctctagtagc ggatttaagt
   12601 tttccttctt caatagcatg gtgagcccat gttacgctta ccccagataa aagcaataca
   12661 gacgtattaa gtaaaggtac tgaaaaagta ggtagggtag taattcctat gggcggccaa
   12721 actgacccga cttcaatagc tggggctaag ctgctatgaa agtaagccca aaaaaaagca
   12781 aaaaagaagc agacttctga tacaataaaa agtataaccc ctaatttaat cccgctaaca
   12841 acataagttg tatgataacc ttgatatgtc ctctctcgaa taatatctcg tcatcagaga
   12901 taagcgataa aagctgttaa tagtattcca aataatagta atactcaata attaaatcga
   12961 aggtagtaaa ttaaacctaa aggcattcta gttaaagcta gagagcttaa aagaggtcag
   13021 gggctaaatt ctacaagatg aaaaggagtt ttgaacatta ataataaaca aagtaaactt
   13081 atttgatgag taattccttt gagcagccgc cggcagagcg ggtattattg atgttaatat
   13141 atatgaaaaa attttttcgc tgctttatcg tgtctattct tatagttttt ttttcgacta
   13201 ttgtgttttt aggccctatt ttaagtggtt cggttagaag ctggtttctt gtttgagcgg
   13261 ggatagagct atcatttatc tcctttcttc cggtaataaa ttatcctaaa agctcaatgt
   13321 cttcagaaag aactattaag tactttttag ttcaatcttg tgctagggca ataattctgg
   13381 tttcaggatc atttttattc attgttgaaa taacaaatat gttttactat aggctctttt
   13441 taactggatt aatattaaag ttaggtttag ttccttttca tttttgggtt attcctgtta
   13501 taaaagggtg taagtttaca gacataattt ttattttagg acctataaaa attattcctt
   13561 ttttattaat tattctatct attccaatca gatcaaattt attattaatt ggtattttca
   13621 tgagactatt ttctttggta attggggcat tattaggtaa caacagaaga agagttcgag
   13681 gaatattagc agcttcctct attgtccatt cagggtgact tctaataggg tgtctagtag
   13741 gagggatagg actttatttt agtatttatc ttatacaatt atacccattt ttgttatttt
   13801 gtttattaaa ccaaaaattt tttatatctg tattaatatt aagttttagg ggtttacctc
   13861 cttttatact atttttgga aaattatata ttttgcaaaa tattatttat aaagaggtcc
   13921 tagtaagttt actttttaca tttattattt ttagcgctct cattagacta aacttctatc
   13981 taaagtttag ctactctttt tacttaagaa gaatattaaa tttaaaaaat aaaattttta
   14041 ttactaattt aataataatc tcgagattat tagggtttat tctttttatc tttttttatg
   14101 gccgagctaa gcattagatt tttaatctaa 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="MRWLYSTNHKDIGTLYMIFGIWCGMVGTGLSLLIRLELGTSGVL
                     FDDHFFNVVVTAHAFVMIFFMVMPVMIGGFGNWMVPLLIGAPDMSFPRMNNMSFWLLP
                     PSFILLICSSMVEGGAGTGWTVYPPLSSMVGHSGASVDLAIFSLHLAGMSSILGAINF
                     ITTIFNMRSSGMTMERLSLFVWSILVTVFLLLLSLPVLAGAITMLLTDRNFNTSFFDP
                     AGGGDPILYQHLFWFFGHPEVYILILPGFGIVSHILGNASSKQPFGSLGMIYAMISIG
                     ILGFIVWAHHMFTVGMDVDTRAYFTAATMVIAVPTGIKVFSWLMTLYGMNTKLDASMY
                     WVLGFIFLFTLGGLTGIVLSNSSLDIVLHDTYYVVAHFHYVLSMGAVFAIFAGFVYWF
                     PVMTGLILNERLAKAQFIIMFSAVNLTFFPQHFLGLAGMPRRYSDYPDSYFKWNQISS
                     FGSLMSVFAVLFFVLIVWEAFLSQRSVAFSGAAIYSREWAENSYPPDFHSNVIPSISV
                     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
                     WYGCILFMVYVGGLLVLFLYVIMLSSNFYLKASGKLLGLFFLMTSSLFMMYLTTSEFL
                     GKMVLGFSGNECSLDLSLSLFLSLGLLLLLAFFSIVHVVLLKGKSILVSV"
     gene            3314..4972
```

```
                     /gene="ND5"
     CDS             3314..4972
                     /gene="ND5"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 5"
                     /translation="MGRLPVLLFFMLFIYLWMSFLFWSWQYETFLYEVSLMFVGSSKI
                     TLMFIGDSISMSFSLVVMTISASVFWFAMEYMSGDEFKFRFGMILFMFVLSMNILIYS
                     GSVFLLLVGWDGLGISSFALIIYYQSNTSLKAGYLTLLTNRIGDVLIICSIPFMIMSG
                     NLYFFSVDHSFLITLLIVLAALTKSAQYPFSAWLPAAMAAPTPVSALVHSSTLVTAGV
                     YLVIRLSLNLILEESVCQLLLFSGAITCILGGFSAIFENDIKKIIAFSTLSQLGLMMF
                     CLGLGIPLLALLHLFTHAMFKALLFLSAGLILMSSFGTQDLRQLGSLLFRSPLLVVIF
                     NISSLCLVGAPFVSAFYSKHAIYEQMMMMGLNSFSMFLMMVGVILTSWYVTRVLKALS
                     WGKVSGLSVQSYLKLTSYFPLLLLFLGSIISGKMFSLLDISMTQSLFVASWFQLLLNF
                     LLVFGVSMGLIWSSPKNSMFFSSMFFLWESSNKSSVFLKGILKQSKVLDYGWLEPKIY
                     FEVSLFKSSFIIQKMFLGVSSYLSSLIPGFSVGILIVSSCIYFW"
     gene            4954..5859
                     /gene="ND1"
     CDS             4954..5859
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 1"
                     /translation="MYLFLMKSILVCLCVLIAVAFFTLLERKVLGYIQIRKGPNKVSI
                     MGIAQPLADALKLFIKEKISLYKSNSMVFFLTPLLGFFIALVLWGLSPSFFSIHFNLF
                     GLLIFFCITGLNVYITSIAGWSSNSLYAFFGGIRASAQSISYEISLVMLLIFPAIISC
                     SMGWFQNMKDYWVFFLCFPLLFIWLVSALAETNRAPFDFAEGESELVSGFNIEYESGL
                     FALLFLAEYTCILFMATATTVWFCKSFWVLDFIIFSFIFSLFFLVARGVYPRFRYDLL
                     MMLCWKSYLPFAISLLMYSLFSKVF"
     gene            5860..6153
                     /gene="ND4L"
     CDS             5860..6153
                     /gene="ND4L"
                     /codon_start=1
                     /transl_table=5
                     /product="NADH dehydrogenase subunit 4L"
                     /translation="MVIKLSYLFLCFCLFLIAFSAFRNQLMSNLIILEGLMLICIVFS
                     VLSVAELGSKSLILMLLTFAAIEAALGLALLVNFLRSYGNDFCLNMSQFFAKN"
     gene            6140..7252
                     /gene="CYTB"
     CDS             6140..7252
                     /gene="CYTB"
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome b"
                     /translation="MRKISAVESMLSLPSPLNISIWWSGGSILGMLLGVQLLTGLFLS
                     MHYASDIMLSFDSVIHIMRDVPAGWALRLIHANGASFFFIFLYLHIGRGLYYQSYILQ
                     PKTWMVGVTIYLLSMATAFLGYVLPWGQMSFWGATVITNLLSAVPYFGGSLVEWVWGG
                     FSVGQATLNRFFSLHFILPFIMTVFILIHLIFLHDKGSTNPLGHNYHLNKINFHPYFT
                     WKDMVGFVLVLLTLITICCFAPYVLSDPENFIYANPMLTPTHIQPEWYFLFAYAILRS
                     IPSKLGGVIALAMSIFILYFLGIGSMKINIPSSFNLLYQVWFWVLVVNFMLLTWLGAC
                     PIENPYLLLAGPLTYLYFFSYPILLSSSWLSFMVLK"
     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="MSFWGAMNLMDPSSPIQSEMILFHDHAMILLIGIFVLVACLGVK
                     LCVNQVSSRTTHEAQLLEILWTIIPAFLLVWLALPSLRLLYLLDEQSSNGLILKATGH
                     QWYWSYEIPSIGISSFDSYMIQENSLQNGEFRLLEVDNRPILPYNMTINTLTTSTDVI
                     HAWAIPSMGIKMDAVPGRLNMMGIQSFSPGVFYGQCSEICGANHSYMPIVVEFLSLED
                     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="MPQLSPGSGLMTLILIFFLMSFLFTLLPFKPMPMKTIKMKSNFS
             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
             TITNLWESSDTYFTRKSLLTSLMLFIILSNFLGLVPLVFSSTTSLWVNSSLALICWGT
             ILVSGWVFSPQKSAAHLAPSGAPGMLMPFLILIETISILIRPITLTVRLVANISAGHI
             ILALLANVLSSSLSIFSWSTCVSIMIFYYLFEMFVSFIQAYIFTLLISLYMAEHP"
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
             MSNMRKPFSLRFFMLVILFLVFDVEVVLLFPLLMLMNAMSSTFLTLTLLIFLLALLLG
             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="MISTLVFLSLASIFMFSWDMVLLSCFMMVLISFVNMHMNFSLNL
             ESSVFMIESFNSFLIFLSVVIIFLSLTNSFWLKINSFSKVVMSLNLFLVLAFSWKDLL
             GFYFFFESSLIPTLILIMVWGYQPERLQAGTYMMLYTVFASLPLLMLILYFFNKSSDL
             SIYNFKLLGLEMSTFMLMIFVMAFLVKLPIFSAHLWLPKAHVEAPLSGSMILAGVLLK
             LGGYGLYLTNQCFNFSSLNLMSVSISFLSMWGGLFAALMCLQQTDMKAMVAYSSVAHM
             SLVISGIFLNSVWGEYCAKVTMIAHGFTSSALFVLVNMSYKKLLSRSFLTAGGLLGMY
             PKMSFLWFLFCSINMAVPPSYNLVGELMLLPSLYIYSVMLVIIMGLVMFFSATYNMFL
```

```
                YTQINHGMANPLINSGSAYYSADYLGLFLHLLTFIMLLKLNLFSTEIC"
    tRNA            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="MIKTPFHLVEFSPWPILGSTLLTCMPLGLVYYLRLHSINLMIFG
                IIATALVAYLWWRDIVRESTYQGFHTSYVISGLKLGVFLFIVSEVCFFFAFFWAYFHS
                SLAPAVEIGSSWPPYGVSTLATFQVPLLNTSVLLLSGVSVTWAHHAIEESKYLSAIYS
                LIFTCILGFYFLFLQFGEYEETSFSIADSVYGSTFFIATGFHGLHVAVGATFLMVCLM
                RMYLYHFNSNHHVGFLAAAWYWHFVDVVWLFLYVSIYWWGSFQK"
    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
                LHFWVIPVLKTSGYNLMMLLLGPMKIVPLSLLMKFFQVFNISNFSVLWFAVFSALIGA
                IIGNNVSSIRAVLGASSISHSGWFMVGAWSGGMWSYFFIYLISLFLMLWYLEESFSKV
                TALLILALSGLPPFLLFFGKVNILLSFLTVNASLLFLVILIISAVLSLNFYLKFSYIF
                YLKEKKLGFTGRWNWLFFLSAWLVGAFLVLLF"
BASE COUNT     4585 a   1728 c   2181 g   5584 t
ORIGIN
        1 ttgcgatgat tgtattcaac aaatcataaa gatattggta ctttatatat aatttttggt
       61 atttgatgtg ggatagtagg aactggactt tcccttttaa ttcgattaga attaggaact
      121 tctggtgtac tgtttgatga tcattttttt aatgttgtag ttactgccca tgcttttgtt
      181 ataatttttt tcatagtcat gcctgttata attggtggat ttggaaattg aatagtgccc
      241 ttattaattg gtgcgccaga tataagtttt ccacgaatga ataacataag attttgatta
      301 ttacctcctt catttatttt acttatttgt tcaagtatag ttgaaggtgg tgctggaaca
      361 ggatgaacag tctatccccc tttaagtagg atagtgggcc atagtggtgc atctgtcgat
      421 ttagctattt tttcattaca tttagctggt atatcttcaa tcttaggtgc tattaatttt
      481 attactacaa tttttaatat acggtcttct ggtataacga tagaacgatt aaggttattt
      541 gtatgatcaa tcttagtaac tgtatttctt ttacttttat ctttacctgt tttagcaggg
      601 gcaattacta tacttttaac agaccgaaat tttaatacta ggtttttga tccagctgga
      661 ggtggtgatc caattttata tcaacatctt ttttgatttt ttggacaccc tgaagtatat
      721 attttaattt tacctggatt tggaattgtt tcccatatcc taggaaatgc atcatctaaa
      781 caacctttg gtaggttagg aataatttac gcaataattt ctattggaat tttaggattt
      841 attgtatggg ctcatcatat atttactgta ggtatagatg ttgacactcg ggcttatttt
      901 actgcggcta ctatagttat tgctgtccct actggaatta aagttttcag ttgattaata
      961 acattatatg gaataaatac aaaattagat gcgtcaatat actgagtttt agggtttatt
     1021 tttttgttta ctttaggagg attaactggt attgttttat caaactcttc tttagatatt
     1081 gtgttgcatg atacttatta tgtagtggca cactttcact atgttttgtc aatgggggcc
     1141 gtttttgcta tttttgctgg atttgtatat tgattccctg tgataacagg attaatttta
     1201 aacgaacgat tagctaaagc tcaatttatt attatatttt cagcagtaaa cttaactttt
     1261 tttcctcaac atttttttagg tttagcgggt atgcctcgtc gatattctga ttacccagat
     1321 agatatttta aatgaaatca aatttcttct tttggttctt taatatcagt ttttgcggta
     1381 ttattttttg ttttaattgt ctgagaagct tttttaagcc aacggtcagt tgcttttagt
     1441 ggtgcagcta tttattctcg agaatgagct gagaattcct atcctcctga ttttcatagc
     1501 aatgtcatcc caagtatttc tgtattgtaa attacactat agtacaattt taagtatatg
     1561 ttaattacga taacatggta ttttttaat tagtgtatta ataatataaa tgtctatagt
     1621 tttaattaag ttaatcttaa aaggtttata aattttttga aggaggtcct taccttttgc
     1681 ataatggttt aattataata tttttggata aatttttcccg aacaagaaag atctataaac
     1741 taagtatttt taggtataaa tattgttgtt gaataaacaa tcaaaaatta ttttataggg
     1801 gcgaaacacc tacaattttt ttgataactg gatttaattt aaatgtattt agtacattaa
     1861 gacagacaat aagtgataag atttattgta atgaaatttt taatttttaa aattttggag
     1921 aaataaaatt ttcaaaaatt tttaagagta taataaatct ctagtagaaa ttcttattta
     1981 aattaaaatt tttaagattt agttttgaaa ataaattatg tttaaatgag taataataaa
     2041 caatatattt taaggaactc ggcaaacata tttttcacct gtttaacaaa aacatagctt
     2101 taagattttt catttaaagt tttacctgcc cggtggaatt tcttaacggc cgcagtactt
     2161 tgactgtgct aaggtagcat aataaattgg cttttaatta aagtcgggta tgaaagggaa
     2221 tttggaaaat aactgtctca agtatatttt attttaattt aattaatgag tgaaaattct
```

2281 catattttat ttcattagac gagaagaccc taagagttat aattaaaatg aatttatttc
2341 atatatcatt tttgttgggg caacaaagac tggatagaac cgtcaaaaag taattttaca
2401 taaccaataa tatttaaaaa gaaataaact accttaggga taacagcata attaattttt
2461 agattgtgac ctcgatgttg gattaaggac ttagtttact agccgtaaac taagatagtt
2521 ctgttcgaac tgtagtacct tacatgatct gagttcagac cggcgtaagc caggtcagtt
2581 tctatctttt atatttaata ttctagtacg aaaggaccgt atatttatag atatctattt
2641 gtttacttgg cagattttaa tgcgattgat ttaggttcaa tatatacttc attgtagtaa
2701 atttttaaaa gaagttaaat aaaattaata ttaactttgg gagttaaaga atgaggatac
2761 tcacttttgg tattatactt taaacagaag atctgcgttg caagtagtta gtgggataat
2821 ccttttacca aactaagatg aattcatctt gaaatgacta ctacttgtat agccttaagt
2881 ttggtaagac ttatttgttt tttacaaagc ccagtagctt tagggggggtt attaattttt
2941 ttaagaattt ttttatctgt aagagttgga tttcaaataa gaaggtgata tggatgtatt
3001 ttatttatag tttatgtggg gggcctttta gttttatttt tatatgttat tatattaaga
3061 agaaattttt atttaaaagc ctctggaaaa ttattaggtt tatttttttt aataacttct
3121 agactgttta taatatactt aacaacatct gaatttttag gcaaaatagt tttagggttt
3181 tctggaaatg aatgcagatt agatttaagc ttaagattat ttttaagttt agggctatta
3241 ctattattag cattttttc cattgttcac gttgtccttt taaaagggaa aagaatttta
3301 gttagtgttt aaaataggtc gacttccggt attattattc tttatacttt ttatttattt
3361 gtgaataaga tttttatttt ggtcttgaca atatgaaact ttcctttatg aggttagtct
3421 aatatttgtt ggttcctcta aaattacctt aatatttatt ggtgattcaa ttagaataag
3481 attcagatta gtcgtaataa caatttcagc atctgttttt tgatttgcta tagagtatat
3541 atctggagat gaatttaagt ttcgttttgg gataattctt tttatatttg ttctatcaat
3601 aaatattttg atttactcag gatcagtttt tttacttcta gttggatggg acgggttagg
3661 aatttcctcc tttgctctaa ttatttacta tcaaagaaac actagattaa aagctggtta
3721 tttaacttta ttaacaaatc gaattgggga tgttttaatt atttgcagga ttccatttat
3781 aattataaga ggaaatttat attttttttc ggttgatcac tcttttctaa ttactctttt
3841 aattgtttta gctgctttaa ctaaaagggc tcaatatcca ttcagagctt gactcccggc
3901 agccatggct gccccaactc ctgtaagggc tttagtacac tcttctactt tagttacagc
3961 aggggtttat ttagttattc gtttaagttt aaacttgatt ttagaagagt ctgtttgtca
4021 gttgttgctt ttttctgggg ctattacttg tattttagga ggatttaggg ctatttttga
4081 aaatgatatt aagaaaatta ttgccttttc tacattaagg caattaggtt taataatatt
4141 ttgtctaggg ttaggtattc ccttattggc attacttcat ctttttaccc atgcaatatt
4201 taaagcttta ctatttttaa gagctggtct aattttaata tcaagctttg gcacacaaga
4261 tttacgtcag ttaggatcct tattatttcg aagaccttta ttagttgtta tctttaatat
4321 tagaagattg tgtttagtag gagccccatt tgtaagagct ttttattcta agcatgccat
4381 ttatgaacag atgataataa taggtttaaa ttcttttagt atattcctaa taatagtagg
4441 cgtaatttta acaagatgat atgttactcg tgttttaaaa gctttaagat gagggaaagt
4501 taggggatta agagttcaat cttatttaaa attaacaagc tacttcccat tattattatt
4561 atttttaggg tcaatcatta gaggaaaaat gttcagatta ttagatatta gtataacaca
4621 aagcctattt gtagcatctt ggttccaact actactaaac tttcttttag tttttggagt
4681 aagaatgggg ctaatttgat cttcacctaa aaatagcata tttttttcta gaatattttt
4741 tttatgggaa agatctaata aaagttctgt atttttaaaa ggaattttaa aacaaaggaa
4801 agttttagat tatgggtgac tagaaccaaa aatttatttt gaggtttctc tttttaagag
4861 tagttttatc atccaaaaaa tatttttggg ggttaggaga tacttaagaa gattaattcc
4921 aggttttaga gttggtatct taattgtaag aaggtgtatt tatttttgat aaaatcaatt
4981 cttgtatgtt tatgtgtgtt aatcgctgta gcttttttta ctttattaga acgtaaggtt
5041 ttaggttata ttcaaattcg aaaaggccca aataaagttt ctattatggg aattgctcaa
5101 cctttagctg acgctttaaa attatttatt aaagaaaaaa ttagacttta taaatctaat
5161 aggatagtat tttttttaac ccctttatta ggatttttta ttgctttagt tttgtgaggg
5221 ttaagaccaa gatttttttc aattcatttt aacttatttg gtttacttat tttttttttgt
5281 attactggat taaatgttta tattacaaga attgccggtt ggtcctcaaa tagtttgtat
5341 gctttttttg gtggaattcg agcttcagct caaagaattt cttatgagat tagattagta
5401 atactattaa tttttccagc tattatttct tgttcaatag gctgattcca aaatataaaa
5461 gattattgag tatttttttt atgttttcct ttactttttta tttggcttgt aagagcatta
5521 gcagaaacaa accgggcccc atttgatttt gctgaagggg aatctgagct agtatcagga
5581 tttaatattg aatatgaaag tggcttgttt gctttactct ttttggctga atacacttgc
5641 attctattta tagcaacagc cactacagtt tgattttgta aaagattttg agttttggac
5701 tttattattt ttagatttat ttttagcttg tttttttttag ttgcccgagg agtttatccc
5761 cgttttcgtt atgatctatt aataatactc tgttggaaaa gataccttcc ttttgctatt
5821 agactattaa tgtatagact ttttagtaaa gtttttttaaa tcgtgattaa actaagatat
5881 ttatttttgt gtttttgctt atttcttatt gctttttcgg cttttcgtaa tcagttaata
5941 tctaatttaa ttattttaga agggttaata ttaatttgta ttgtttttag agtcctcaga
6001 gtagctgaat taggatcaaa aagtttaatt ttaatgctct taacatttgc ggctattgaa
6061 gcagctttag gtttagcttt actagtaaat tttttacggt catatggaaa tgatttttgt
6121 ttaaatataa gtcaattttt tgcgaaaaat tagtgcagtt gaaagaatgc ttagattacc
6181 ttccccatta aacattagta tttgatgaag aggaggctca atcctaggaa tattattagg
6241 tgtccaactt ttgactggtt tatttctttc tatacactat gcaagagata ttatactatc
6301 atttgactca gtaattcaca ttatacgaga cgttcctgct ggttgagccc ttcgtttaat
6361 tcatgctaat ggggcttctt tcttttttat tttttatat ttacatattg gtcggggatt
6421 atactatcaa agatatattt tacaaccaaa aacttgaata gttggagtta caatttactt

6481 gttaagaatg gctacagcat ttcttggata tgttttacca tgaggtcaga tatcattttg
6541 aggtgcaaca gtaattacaa atcttctctc agcagttcca tattttggag ggtctttagt
6601 agaatgagtt tgaggtggat tttctgttgg tcaagcaacg ttaaatcgat tttttctct
6661 tcattttatt ttacctttta ttataacagt gtttatttta attcatttaa tttttcttca
6721 tgataaaggt tcaactaacc cacttggaca taattaccat ttaaataaaa ttaattttca
6781 tccttatttt acttgaaagg atatagtagg atttgtattg gttttattaa ctttaattac
6841 aatttgttgc tttgctccat acgttttgtc agatccagaa aattttattt atgccaatcc
6901 aatactgacc ccaactcata ttcaacctga atgatatttt ttatttgcat atgctatttt
6961 acgatcaatc ccatcaaaat taggtggagt tattgcttta gctataagaa tttttatctt
7021 atatttcttg ggtattggat ctataaaaat taatatccct tcaagattta atcttttata
7081 tcaagtatga ttttgagttt tagtagtaaa ttttatactt ttaacttgac ttggggcttg
7141 ccctattgag aatccatacc ttcttttagc tggaccatta acctatctat attttttag
7201 ctatcctatt ttattaagtt ccaggtggtt gagatttata gttttaaaat agctttgttt
7261 aatttaattt aaaatatttg cttgtcaagt aaaaaacaca tgaaagttgt aacaaaataa
7321 ggtaatagtt tatataaaaa tatataattt gcaaaattat agatgaacaa agttcttacc
7381 tttaattcag tagcttaaat aaagcgtggc tttgaagaag ccaaaataaa gtttcgattt
7441 tttgaattaa tgagtttttg aggtgcaata aatttaatag atcctaggtc tccaattcaa
7501 tcagaaataa ttttatttca tgaccatgct ataattttat taattggaat ttttgttttg
7561 gtagcatgtt taggggtaaa actctgtgtt aatcaagttt catcacgtac gacccatgaa
7621 gctcaacttt tagaaatctt atgaactatt attccagctt ttttattagt ttgattagct
7681 ttaccaagct tacggcttct ctatctctta gacgagcaga gtagtaatgg attaatttta
7741 aaagccactg gacatcaatg gtattgaaga tatgaaatcc caagaattgg aattagatca
7801 tttgattcct atataattca agaaaattct cttcaaaatg gtgaatttcg cttacttgaa
7861 gttgataacc gtcctatttt accatacaac ataacaatta acacccttac aactagaaca
7921 gatgttattc atgcttgagc tattccttca ataggaatta aaatagatgc cgtcccagga
7981 cgattaaata taataggtat tcagtctttt agaccaggag tattttatgg tcaatgttct
8041 gaaatttgtg gagctaacca ctcttacata ccaattgttg ttgagtttct ttcattagaa
8101 gattttatta attttttcgg aatcttttgt ttaagaggta atttgtaaag ttattttagg
8161 atctttccct gaagttttaa ggtttaaact gcttatttgt ggaataagag agtcaactta
8221 tagttgaaaa acttccttaa ttttttagta gtacataagt taaataaact gatgaccttc
8281 aaagtcattt attcttaaat agattacttc gtatactttc tagtataaat agtacattcg
8341 ccttccaagc gaaaagcctt aaataaggta agtaaatttt tataagattt ctttcttaaa
8401 gcctcaaatg ctcttgtgcc taacaccgat aaaaaaaaga gcatagctca taaaaggcgc
8461 ttaaaacctt tgcatttatc tgccatataa aattatttaa aaattagcgt agataaatac
8521 ctaaagtttg attttatttt aattgttttt attggtatag gtttaaaggg aagtaaagta
8581 aatagaaaac ttattaagaa aaaaattaaa attaatgtta ttaaaccact tcctggacta
8641 agttgaggca tatctaaact actaggaaaa gtatctttta attaacattt aaatgcttaa
8701 attttaagct aagtttagat ttttttatgg gtgttcggct atataaaggc taattagaag
8761 agtaaaaatg tatgcttgaa taaatgatac aaatatttca aataaatagt aaaagattat
8821 aattgaaaca caagttgatc atgaaaaaat tcttaaagaa gatcttaaaa cattagctaa
8881 gagggctaaa ataatgtgcc cagcactaat atttgcaact aatcgaactg ttaaagtaat
8941 aggacgaatt aaaatactaa tagtctcaat taaaattaaa aatggtatta gtatcccagg
9001 agcccctgaa ggggctaaat gtgctgcaga tttttgtggc gaaaaaactc atcctgaaac
9061 taaaatagtt cctcagcaaa ttaaagcaag gcttctattt actcataatc tggttgtaga
9121 tctaaacact agaggcacta aacccaaaaa atttcttaaa ataataaata atattagaga
9181 agttaataat gactttcgag taaaataagt atctgaagac tcccataagt ttgtaatagt
9241 tgctgtaatt ggagatattc tcgttgttaa aaagattgaa tttgaaataa aaataacagt
9301 tataactaaa ggtattgctc aaattacaat agagtgatta ccgtctatag atgaaaataa
9361 gtctgttatc atttgataag aatgtaagat atacctctta aaattttaag gccgaaactt
9421 aaaaacgaaa cgtttcattc ttgggttatt aatgaaattt tcatatttgt attacgaaaa
9481 aatactgtat taagttttat actatagtaa catgaatttg tttttgcagt aaatgcaaag
9541 aagcaatttc cattacttct actttgttac gacttgtctc atttttaaac gagaatgacg
9601 ggcgatttgt gcactactaa aaaaatattt agattttatt ttttatttaa atcttacttt
9661 caagtccatc ttcaaaattt agtacttaaa tttatctgta ttaaaataat attgtaactc
9721 gccttaatta aaactataag ttgcaccttg atctgtctta ttgattactc ttaaaaaaat
9781 tttctaaaca aatttttaga cgacggcata caaactttaa agttttattg ttttacttgg
9841 tggttatcaa ttattaggta agttcccctg attattttta ttagccgcca taactcttga
9901 gtttatatat taaatatagt actactcatc aaataaggtt gaatatctag aataataggg
9961 tatctaatcc tagttttctt tctagcttta aaataaaaat atagaagcaa aaataaaaaa
10021 gaatatattt cactattatt caaagttttg tctataaaaa ttctatttga tttattgtaa
10081 ttagtttttt tataggtctg accgcggttg ctggcaccta ttaaaccaaa aaagaaataa
10141 tttactaagt taacatttct gttatttatt aaaatttata actggttctt acaaaatttc
10201 cataattata taaacattaa actaaatttc tttatcaagt taggtaaatc taaaaaagat
10261 tactctattt ttgagttatg agcccaacag cttaaaaatt agcttatttt tagataattt
10321 aattaacact tcattctaaa gctcctatgg ctcactcata aattaaacct agtaaaaggg
10381 ctaaaagaaa aattaagagt gtcaaagtta aaaaagttga tgatattgca tttattaaca
10441 ttaagagagg gaagagcaaa actacttcaa cgtcaaaaac taaaaataaa attactaata
10501 taaaaaaacg gagtgaaaaa ggcttacgca tatttcttat tggatcaaaa ccacactcaa
10561 aaggcgatct tttttctcta tcttctgaaa agatagggct tcagtaagta ataaaaaaaa
10621 gtataattaa tgctacacta agaactcctg gaaaaattga gtagtatagc aatttagtga

```
   10681 gctcactaaa tgatataaaa tatatctaca aagattttca aaatccttta ttatataaaa
   10741 attcatgcta agaagatata attgaggctg ctaactttga tttgggtttt tggatccgct
   10801 tctttttgat ttctacttta gttttttaa gtttagcaag aatttttatg tttaggtgag
   10861 atatagtttt attaagatgc tttataatag ttttaattag atttgttaat atacatataa
   10921 attttagatt aaatttagaa agaagagtat ttataattga atcatttaat aggtttttaa
   10981 ttttttaag ggtagttatt attttttgt cattaacaaa ttctttttga ttaaaaatta
   11041 ataggtttag gaaagtagta atatcattaa acttgtttct agttttagct tttagttgaa
   11101 aagatttatt agggttttat ttttttcg aatcttcttt aattccgaca ttaattttaa
   11161 ttatagtatg agggtaccag ccagagcgac tacaagcggg aacttatata atactttaca
   11221 cagtatttgc gtctttaccc ttactaatat taattctata ttttttaat aagaggagag
   11281 atttaagaat ctataatttt aagttattag gattagaaat aagtactttt atattaataa
   11341 tttttgtgat agctttttta gttaaattac ctatttttag agctcatttg tgacttccaa
   11401 aagcacatgt agaagctcct ttaagaggtt ctataatttt agctggagtt ttattaaaat
   11461 taggaggata tggtctttat ttaacaaacc aatgttttaa ttttagctct cttaatttaa
   11521 taagagtaag gattaggttt ttaagaatat gaggaggatt atttgcagct ttaatgtgcc
   11581 ttcaacagac agatataaaa gcaatagttg cttactcttc agtagctcat ataagtttag
   11641 ttattagtgg aatttttta aactctgtat gaggggaata ttgtgccaaa gttaccataa
   11701 ttgcccatgg ttttacttca tcagctttgt ttgttttagt aaatataaga tataaaaaat
   11761 tacttagtcg gaggtttta acagcgggag gattactagg tatatatcca aaaatatctt
   11821 ttttatggtt tcttttttgt agaattaata tagctgtccc gccatcttat aatttagtag
   11881 gagagttaat acttcttcca tctttatata tttactcagt aatgttagtt atcattatag
   11941 gtcttgtaat attttttaga gctacttata atatattttt atatactcag attaatcacg
   12001 ggatagctaa tcctttaatt aattctggaa gggcttatta ttcagcagat tatttagggc
   12061 tatttcttca tcttttaact tttattatat tattaaaatt aaatttattt tcaacagaaa
   12121 tctgttaaat ggttccaaag gaaaaattca ttgttgattt acaagatcaa tattttattt
   12181 aaattacttt tggaaagaac ctcaccagta aattcttacg tataagaata atcaaaccac
   12241 gtcaacaaaa tgtcaatatc aagctgcagc taaaaagcct acatggtggt ttgaattaaa
   12301 atgatataag tatattcgta taagacatac tattaaaaat gttgctccaa cagcaacatg
   12361 gagtccgtga aaacctgttg caataaaaaa agttcttcca taaactctat cagcgataga
   12421 aaatgatgtt tcttcatatt ctccaaattg gagaaataag aaataaaacc ctaaaataca
   12481 agtaaaaatt aagctataaa ttgcggataa atattttctt tcttcaatag catgatgagc
   12541 tcatgttacc cttactcctg ataataaaag aactctagta tttaagaggg gaacttgaaa
   12601 agttgcaaga gttcttacac cgtaaggtgg tcatctagat ccaatttcta cagcaggtgc
   12661 tagtcttcta tgaaaataag ctcaaaagaa agcaaaaaaa aagcaaactt ctgatacaat
   12721 aaataagaaa actcctaatt ttaatccact aattacataa gatgtatgaa aaccttgata
   12781 tgttctttct cgaacaatat ctcgtcatca taaatatgca accagtgcag ttgcaataat
   12841 tccaaaaatt attaaattaa tcgagtgtaa tcgtaagtaa tatactaagc ctaaaggcat
   12901 acaagttaag agagttgatc ctaaaattgg ccatggtcta aattctacta aatgaaaagg
   12961 agttttaatc atatattaat atataatatt ctaatttttt ttattagaaa gtttctcaaa
   13021 gcaaccgccg ggagagcggg tatcattgat gttgatatat atgaatattc taattattcg
   13081 ttgctataat gacagttttt aggtgaatat ttttaagttt ggtattttta ggtcctgttg
   13141 ttagattatc aagatcaaga tgatttttag tttgaggggg aatagaacta tctttaattg
   13201 gattattacc tttaataagc aacgataaaa gaaatattat aacagaaaga gcttttaaat
   13261 attttttagt tcaagcttta gctagtttaa ttgttttaat tagtggtatt aacattttta
   13321 tatttaatag aaataggata ataattcttg gttttttttt aattggttta actttaaaat
   13381 taggtatttt tcctttacat ttttgagtga tcccagtttt aaagactagc ggttataatt
   13441 taataatatt gttactaggt cctataaaaa ttgtaccttt aaggttatta ataaaatttt
   13501 ttcaggtttt taacattaga aattttagag tactatgatt tgcggtcttt agtgcattaa
   13561 ttggggcaat tattggtaat aatgttagaa gaattcgagc agttttaggg gcttcttcta
   13621 tttctcattc tgggtgattt atagtaggag cttggagagg agggatgtgg tcttattttt
   13681 ttatttactt aattagccta tttttaatat tatgatattt agaagaatcc tttagaaaag
   13741 taacagcatt gttaatttta gcattaaggg gccttccccc attcttgttg ttttttggaa
   13801 aagtaaacat tttattgaga tttttaacag taaatgcctc tttgcttttc ttagtaattt
   13861 taattattag ggcagttctt aggctaaact tctacttgaa gttcagttat atttttttatt
   13921 taaaggaaaa aaaattagga tttacagggc gctgaaattg attatttttt ttaagtgctt
   13981 ggttagtggg agctttttta gttttgctat tttaaaaagt tttttatggc tgagttataa
   14041 gcattaaatt tttaatttaa ttacaaggtt tatccttt
//
```

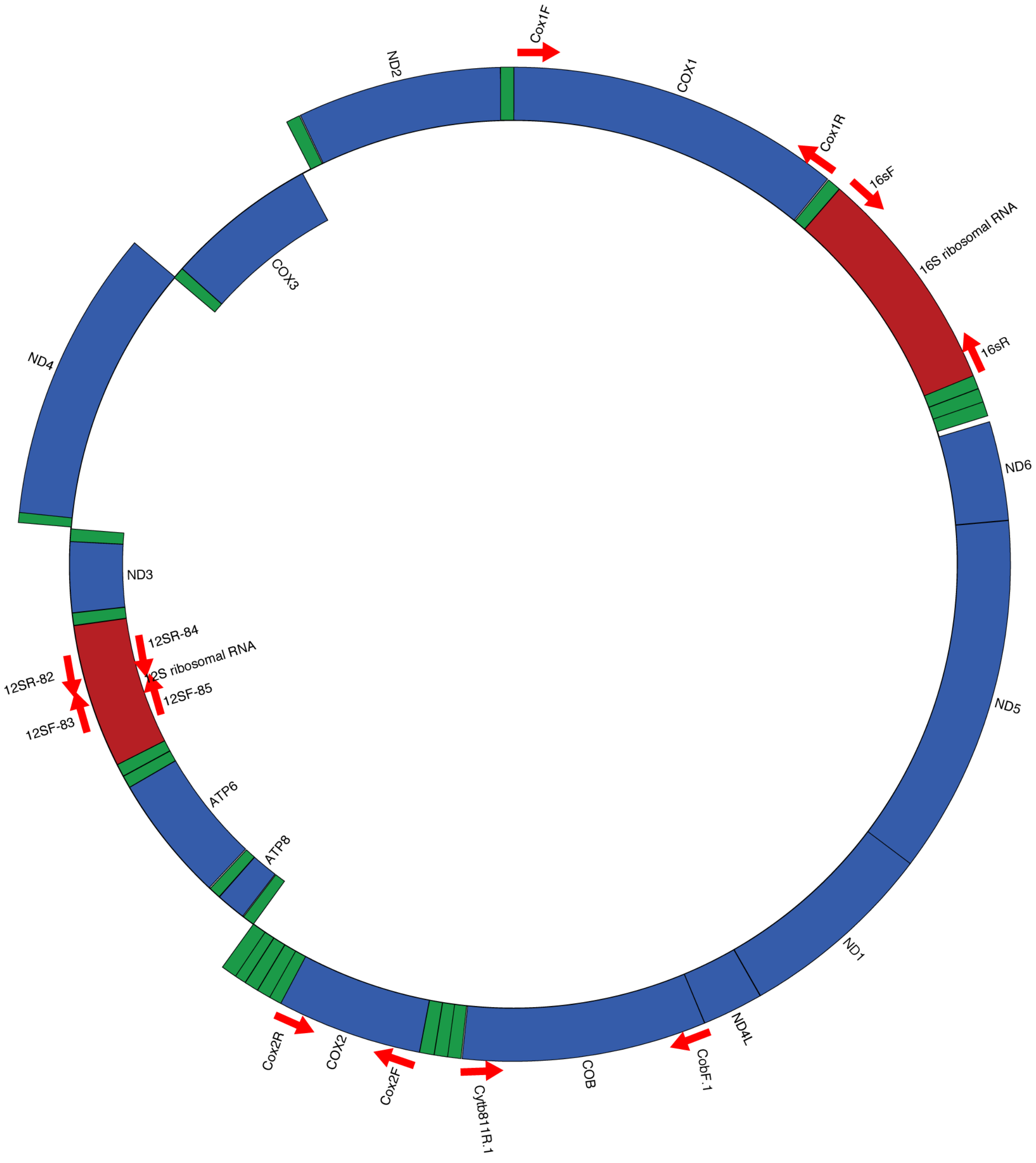
Appendix C: Section 1

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

References

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, Stanisic, 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>