Assessing population-level variation in the mitochondrial genome of *Euphausia superba* using 454 next-generation sequencing

Mattias Johansson[1], Elizabeth Duda[1,2], Angela Sremba[1], Michael Banks[1], William Peterson[3]

[1]*Cooperative Institute for Marine Resources Studies, Hatfield Marine Science Center, Newport, OR*
[2]*Pomona College, Claremont, CA*
[3]*National Marine Fisheries Service, NOAA, Hatfield Marine Science Center, Newport, OR*

Abstract:

The Antarctic krill (*Euphausia superba* Dana 1852) is widely distributed throughout the southern ocean, where it provides a key link between primary producers and upper trophic levels and supports a major commercial fishery. Despite its ecological and commercial importance, genetic population structure of the Antarctic kill remains poorly described. In an attempt to illuminate genetic markers for future population and phylogenetic analysis, five *E. superba* mitogenomes, from samples collected west of the Antarctic Peninsula, were sequenced using new 454 next-generation sequencing techniques. The sequences, of lengths between 13310 and 13326 base pairs, were then analyzed in the context of two previously published near-complete sequences. Sequences revealed relatively well-conserved partial mitochondrial genomes which included complete sequences for 11 of 13 protein-coding genes, 16 of 23 tRNAs, and the large ribosomal subunit. Partial sequences were also recovered for Cox1 and the small ribosomal subunit. Sequence analysis suggested that the Cox2, Nad5, and Nad6 genes would be the best candidate for future population genetics analyses, due to their high number of variable sites. Future work to reveal the noncoding control region remains.

Keywords: *Euphausia superba*, mtDNA, mitogenome, variability, control region

Introduction:

Antarctic krill, *Euphausia superba* Dana 1852, is a common species of zooplankton vital to the marine ecosystem. Along with providing the key link between the primary producers and the animals in the upper trophic levels such as whales, seals, penguins, seabirds, and fishes [1,2], krill may also play an important role in maintaining the balance of dissolved carbon in the water [3]. Additionally, Antarctic krill is of major economic importance, with an average of 230,000 metric tons caught annually [4]. Recent evidence suggests that krill abundance may be decreasing, possibly in part due to lower levels of sea ice [5-7], making a more complete understanding of these organisms all the more vital.

E. superba has been the focus of a number of recent mtDNA studies, with a primary concentration on determining the level of heterogeneity in the circumpolar population by examining diversity of a single gene [8-11]. However, while there are extensive data on a few portions of the mitogenome, only two near-complete sequences have been published to date [12,13] and only one of these [13] shows the full gene set of 38 genes. Within these two studies there is some disagreement as to which genes show the highest level of variability, and the majority of the mitochondrial control region remains unsequenced. Therefore, the goal of this study was to leverage newly-available next generation sequencing techniques to sequence multiple mitogenomes in order to examine the intraspecific diversity of *E. superba* and pinpoint regions for population genetics analyses. This research should also provide data that may be useful in improving our understanding of phylogenetic relationships within the Malacostraca.

For this study, we chose to sequence the mitochondrial genome of *E. superba* due to its matrilineal inheritance, relatively high rate of mutation accumulation, and simplicity of isolation. A typical strand of metazoan mtDNA is a circular molecule around 15,000 bp long that contains 37 genes – 22 tRNA, 13 protein coding, and the large and small ribosomal subunits – along with at least one non-coding region [14-16]. In general, mitochondrial gene arrangements remain relatively stable, but enough significant variations have been found to exist above the generic level to allow them to be useful for phylogenetic analysis [17].

Materials and Methods:

Euphausia superba specimens were collected from the Antarctic Peninsula in 1994 and 1997 (Figure 1 and Supplemental Table S1) and dry frozen. Genomic DNA was extracted using a standard phenol-chloroform-isoamyl extraction protocol [18]. Extracted DNA was eluted in TE buffer and stored at 4˚C.

Mitochondrial genomes were amplified in five fragments via a long-range polymerase chain reaction (PCR) approach [19]. PCR and sequencing methods were adapted from those of Shen et. al. [13]. PCRs of a total volume of 20µl were prepared using six primers from Shen et. al. [13] and four novel primers designed to optimize PCR results (Supplemental Table S2). PCR reactions consisted of 1 µl of DNA, 11.8 µl of ultrapure water, 0.4 µl of 10mM dNTPs, 0.6 µl DMSO, 1.0 µl each of the forward and reverse primers, 4.0 µl of 5X Phusion HF reaction buffer, and 0.2 µl of Phusion polymerase (Finnzymes, Thermo Scientific, Vantaa, Finland). Thermal cycling conditions were an initial denaturation at 98ºC for 30 seconds ; 30 cycles of 98ºC for 10

seconds, 55-67ºC for 20 seconds, and 72ºC for one minute; and a final extension step at 72ºC for five minutes. Products were run through 1% agarose gels containing ethidium bromide and visualized under ultraviolet light prior to sequencing.

PCR products were combined in equimolar amounts and purified using a Qiaquick PCR Cleanup Kit (Qiagen, Valencia, CA) in preparation for 454 sequencing. Mitogenome libraries were prepared from PCR products for each individual using the Rapid Library Preparation Method protocol (Roche), with PCR products treated as genomic DNA, and using unique RL MID adaptors for each sample. Libraries were combined in equimolar amounts, and subjected to emulsion PCR following the emPCR Amplification Method Manual – Lib-L. Emulsion PCR products were then cleaned and prepared for sequencing according to the Sequencing Method Manual before being analyzed on a 454 GS Junior second generation sequencer (Roche).

Initial sequences were assembled using GS De Novo Assembler (Roche) and aligned using GS Reference Mapper (Roche), with the published mitogenome sequence from Shen et. al. [13] as a reference. The resulting contigs were aligned and edited manually in Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, MI) for quality control at variable sites. Benchmarks were set to establish high- and low- confidence variation, where low-confidence nucleotides with fewer than 30 unique reads and a quality score from sequencing output below 40 were marked as unknowns. Edited sequences were then analyzed with DOGMA [20] and tRNAscan-SE [21] to identify protein-coding genes, tRNA sequences, and the short and long ribosomal subunits. Amino acid sequence translations and A/T content analysis were performed using

BioEdit Sequence Alignment Editor [22]. Sequences were deposited in GenBank with accession numbers JQ286347 to JQ286351.

Results and Discussion:

Sequenced Regions

The final high-quality sequences varied in length from 13310 to 13326 bp after editing and included complete sequences for 11 out of 13 protein coding genes, 16 of 23 tRNA regions, and the long ribosomal subunit (Table 1, Supplemental Tables S3-S6). Partial sequences were obtained for one more protein-coding gene, *cox1* (95.7% complete), and the short ribosomal subunit (79.0% complete). The remainder of the mitogenome between these two regions was not recovered. This missing fragment should contain the control region, the protein-coding gene *nad2*, the tRNA sequences *trnI, trnN, trnQ, trnM, trnC, trnY,* and *trnW,* the missing beginning of *cox 1* and end of the short ribosomal subunit [13].

Although no sequencing data aligned with the region of the reference from *rrnS* to *cox1,* PCR product of the correct size, which was amplified by the primer pair Eus-*rRNAF* and Eus-*cox1R,* was nonetheless visible on agarose gels and a contig assembled by the GS De Novo Assembler (Roche) may account for it. NCBI BLAST results suggest that this contig represents a sequence originating in the *E. superba* nuclear genome that encodes a duplication of the short ribosomal subunit, where the primer was designed to sit down in the mitochondrial genome. It is also possible that the age and quality of DNA influenced primer annealing in an adverse manner, since Shen et. al. [13] obtained a portion of the desired mitochondrial fragment using the same primers as those used in

this study. Different primers, ideally outside the short ribosomal subunit, should be designed for future work to eliminate inconsistencies.

The difficulties encountered in sequencing the noncoding region, which were also a problem in the two previous studies of this species [12,13], may be the result of secondary stem and loop structures [12]. Although DMSO was used as a PCR additive to reduce issues with secondary structure [23], it is possible that these stem and loop structures disrupted the Phusion enzyme during long PCRs and led to premature termination of fragment extension. Shorter PCRs, possibly in combination with other PCR additives such as BSA [24], formamide [25], or non-ionic detergents such as Triton X-100 [26] may help to resolve PCR issues with the control region in this species.

Gene order for the successfully sequenced region, provided by DOGMA and tRNAscan, corresponded exactly with that published by Machida et. al. [12] and Shen et. al. [13] (Figure 2, Table 1, Supplemental Tables S3-S6). This includes two of the four previously-identified translocations relative to the pancrustacean ground pattern. The finding of Shen et. al. [13] that *E. superba* have twenty-three tRNA-encoding regions instead of the usual twenty-two could not be confirmed due to incomplete data.

A/T composition

Overall A/T composition for the five samples varied from 67.48% to 68.09% (Table 1, Supplemental Tables S3-S6). This variation partially reflects the actual variability of the sequence and partially reflects variable sequence quality, with unknown nucleotides (N) ranging from 0.24% to 1.43% of the total sequence. The sequences of Machida et. al. [12] and Shen et. al.[13] had A/T contents of 67.8% and 68.1%, respectively, falling close to or within this range.

Variable Sites

A total of 226 variable sites that passed quality-control benchmarks were found between the five samples analyzed in this study and the two already published sequences from Machida et. al. [12] and Shen et. al. [13]. Of these sites, 219 were present in protein coding regions, one was found in the tRNA for aspartic acid, two were in the short ribosomal subunit, and four were part of noncoding intergenic sequences. Most of the 219 variable sites in protein coding regions were silent, third-position changes, leaving only two mutations in *nad4* and one in *nad5* that actually affected the amino acid sequence. All start and stop codons remained unchanged across all samples (Table 2).

Broken down by region, the level of variability within protein-coding genes ranged from 1.41% for *nad*3 to 3.14% for *atp*8 (Table 3). Machida et. al. [12] also found the *atp8* region to be highly variable, although Shen. et. al. [13] had opposing results, with no variable sites found at all. Regardless, the *atp8* region remains inadequate for future population genetics purposes due to its small size. (The high percentage variability found in this study actually represents only five total changes out of 159 bases.) The two larger regions of next highest variability, *cox2* and *nad6*, are better candidates for these purposes, and *Nad*5, at 1731 bp and 44 total variable sites despite its lower 2.54% variability, is probably best of all. The sequence for *nad*2 was not recovered for the five samples from this study, but both Machida et. al. [12] and Shen et. al. [13] found it to be the most variable region, suggesting it might also be a good candidate for future research. Additionally, the noncoding control region has been found to accumulate a relatively high number of mutations in other species [27-30] and further work to illuminate this sequence in *E. superba* should be done.

Conclusion

This study characterized 13,000 base pairs of the 15,000+ bp mitochondrial genome for five *E. superba* individuals collected off the Antarctic Peninsula using a 454 GS Junior next-generation sequencing approach. A/T composition was found to be between 67.48% and 68.09%, in line with previously published results for this species [12,13]. Although the mitochondrial genome was highly conserved overall, 226 high-confidence variable sites were identified, with *cox2*, *nad5*, and *nad6* identified as the regions potentially most useful for population genetic analysis due to their high number of variable sites. Difficulties in PCR amplifying the remaining 2000+ bp of the mitogenome mean that sequencing the control region remains to be completed, if this highly-variable region is to be used in future population genetic or phylogenetic research.

Acknowledgements

The authors would like to thank D. Jacobson, E. Slikas, and A. Alexander for their support with the 454 sequencing and analysis, C. Shaw for the samples used in this study, and M. O'Connor for assistance constructing the sampling map. This research was co-funded by the ASSURE program of the Department of Defense in partnership with the National Science Foundation REU Site program under Grant No. NSF OCE-1004947.

References

1. Laws, R. M., (1985). The ecology of the southern ocean. *American Scientist*, **73**, 26-40. Print.

2. Watanuki, Y., Yoshihisa, M., Yasuhiko, N. (1994) *Euphausia superba* dominates in the diet of Adélie penguins feeding under fast sea-ice in the shelf areas of Enderby Land in summer. *Polar Biology*, **14**(6), 429-432. doi: 10.1007/BF00240264

3. Ruiz-Halpern, S., Duarte, C. M., Tovar-Sanchez, A., Pastor, M., Horstkotte, B., Lasternas, S., and Agustí, S. (2011). Antarctic krill as a source of dissolved organic carbon to the Antarctic ecosystem. *Limnol. Oceanogr.*, **56**(2), 521–528. doi:10.4319/lo.2011.56.2.0521

4. Jones, C. D. and Ramm, D. C. (2004). The commercial harvest of krill in the southwest Atlantic before and during the CCAMLR 2000 Survey. *Deep Sea Research Part II: Tropical Studies in Oceanography*, **51**(12-13), 1421-1434. doi: 10.1016/j.dsr2.2004.06.009

5. Siegel, V., and Loeb, V. (1995). Recruitment of Antarctic krill *Euphausia superba a*nd possible causes for its variability. *Marine Ecology Progress Series*, **123**, 45-56. doi: 10.3354/meps123045

6. Loeb, V., Siegel, V., Holm-Hansen, O., Hewitt, R., Fraser, W., Trivelpiece, W., and Trivelpiece S. (1997). Effects of sea-ice extent and krill or salp dominance on the Antarctic food web. *Nature* 387, 897-900. doi:10.1038/43174

7. Atkinson, A., Siegel, V., Pakhomov, E., and Rothery, P. (2004). Long-term decline in krill stock and increase in salps within the Southern Ocean. *Nature*, **432**, 100-103. doi: 10.1038/nature02996

8. Zane, L., Ostellari, L., Maccatrozzo, L., Bargelloni, L., Battaglia, B., and Patarnell, T. (1998). Molecular evidence for genetic subdivision of Antarctic krill (*Euphausia superba* Dana) populations. *Proceedings of the Royal Society, London, B*, **256**, 2387-2391. doi:10.1098/rspb.1998.0588

9. Batta-Lona, P. G., Bucklin, A., Wiebe, P., Patarnello, T., Copley, N. J., (2010). Population genetic variation of the Southern Ocean krill, *Euphausia superba*, in the Western Antarctic Peninsula region based on mitochondrial single nucleotide polymorphisms (SNPs). *Deep-Sea Research II*. **58**(13-16), 1652-1661. doi:10.1016/j.dsr2.2010.11.017

10. Goodall-Copestakw, W. P., Pérez-Espona, S., Clark, M. S., Murphy, E. J., Seear, P. J., and Tarling, G. A. (2010). Swarms of diversity at the gene cox1 in Antarctic krill. *Heredity*, **104**, 513-518. doi:10.1038/hdy.2009.188

11. Bortolotto, E. Bucklin, A., Mezzavilla, M., Zane, L., and Patarnello, T. (2011) Gone with the currents: lack of genetic differentiation at the circum-continental scale in the Antarctic krill *Euphausia superba*. *BMC Genetics*, **12**(32). doi:10.1186/1471-2156-12-32

12. Machida, R. J., Masaki, U. M., Mitsugu, M. Y., Mutsumi, N., and Nishida, S. (2004). Organization of the Mitochondrial Genome of Antarctic Krill Euphausia superba (Crustacea: Malacostraca) *Marine Biotechnology*, **6**, 238-250. doi: 10.1007/s10126-003-0016-6

13. Shen, X., Haiqing, W., Ren, J., Tian, M., Want, M. (2010). The mitochondrial genome of *Euphausia superba* (Prydz Bay) (Crustacea: Malacostraca: Euphausiacea) reveals a novel gene arrangement and potential molecular markers. *Molecular Biology Reports*, **37**(2), 771-784. doi:10.1007/s11033-009-9602-7

14. Wilson, A. C., Cann, R. L., Carr, S. M., George, E. M., Gyllensten, U. B., Helm-Bychowski, K. M., Higuchi, R. G., Palumbi, S. R., Prager, E. M., Sage, R. D. and Stoneking, M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, **26**, 375–40. doi: 10.1111/j.1095-8312.1985.tb02048.x

15. Moritz, C., Dowling, T. E., & Brown, B. M. (1987). Evolution of Animal Mitochondrial DNA: Relevance for Population Biology and Systematics. *Annual Review of Ecology and Systematics , 18*, 269-292. doi: 0.1146/annurev.es.18.110187.001413

16. Boore, J. L,.; Macey, J. R., and Medina, M. (2005). Sequencing and comparing whole mitochondrial genomes of animals. *Methods in Enzymology*, **395**, 311-348. doi: 10.1016/S0076-6879(05)95019-2

17. Boore, J. L. (1999). Animal Mitochondrial Genomes. *Nucleic Acids Research*, **27**(8), 1767-1780. doi:10.1093/nar/27.8.1767

18. Sambrook, J., and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

19. Cheng, S., Chang, S. Y., Gravitt, P. (1994). Long PCR. *Nature* **369**, 684 – 685. doi:10.1038/369684a0

20.Wyman, S.K., Jansen R.K., and Boore, J. K. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**(17), 2353-3255. doi:10.1093/bioinformatics/bth352

21. Lowe, T. M. and Eddy, S.R. (1997). tRNAscna-SE: a program for improved detection

of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, **25**(5), 955-964. doi: 10.1093/nar/25.5.955

22. Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.

23. Frackman, S., Kobs, G., Simpson, D. and Storts, D. (1998). Betaine and DMSO: enhancing agents for PCR. *Promega Notes*, **65**, 27.

24. Pääbo, S., Gifford, J.A., and Wilson, A. C. (1988). Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Research,* **16**(20), 9775-9787. doi: 10.1093/nar/16.20.9775

25. Sarkar, G., Kapelner, S., and Sommer, S.S. (1990) Formamide can dramatically improve the specificity of PCR**.** *Nucleic Acids Research,* **18**(24), 7465. doi:10.1093/nar/18.24.7465

26. Bachmann, B., Lüke, W., and Hunsmann, G. (1990). Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucleic Acids Research*, **18**(5), 1309. doi:10.1093/nar/18.5.1309

27. Chu, K. H., Li, C. P., Tam, Y. K. and Lavery, S. (2003). Application of mitochondrial control region in population genetic studies of the shrimp *Penaeus*. *Molecular Ecology Notes*, **3**(1), 120–122. doi: 10.1046/j.1471-8286.2003.00376.x

28. Diniz, F.M., Maclean, N., Ogawa, M., Cintra, I.H.A., and Bentzen, P. (2004). The hypervariable domain of the mitochondrial control region in Atlantic spiny lobsters and its potential as a marker for investigating phylogeographic structuring. *Marine Biotechnology*, **7**(5), 462-473. doi: 10.1007/s10126-004-4062-5

29. Liu, Y. (2010) Complete mitochondrial genome of the Chinese spiny lobster Panulirus stimpsoni (Crustacea: Decapoda): genome characterization and phylogenetic considerations. *Molecular Biology Reports*, **38**(1), 403-410. doi: 10.1007/s11033-010-0122-2

30. Mancini, E., De Biase, A., Mariottini, P., Bellini, A., Audisio, P. (2008) Structure and evolution of the mitochondrial control region of the pollen beetle Meligethes thaslassophilus (Coleoptera: Nitidulidae). *Genome*, **51**(3), 196-207. doi:10.1139/G07-116

Appendix:

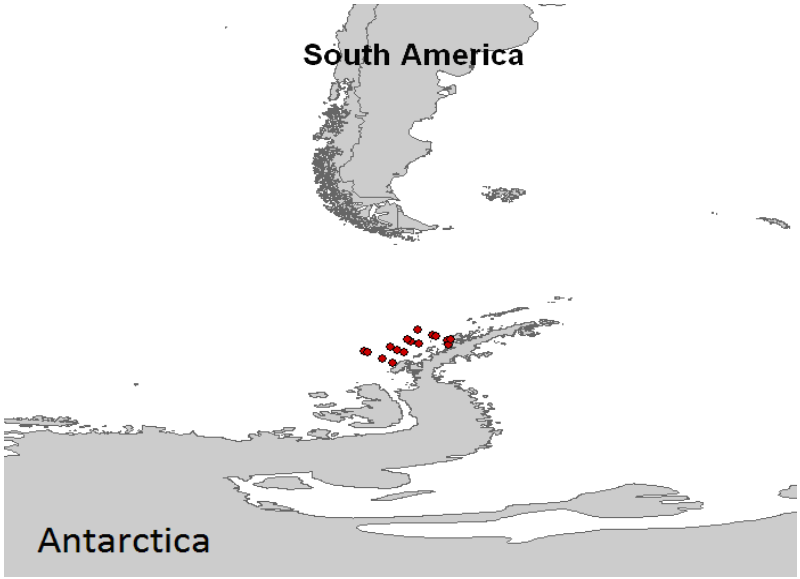

Figure 1. Distribution of krill samples off the Antarctic Peninsula

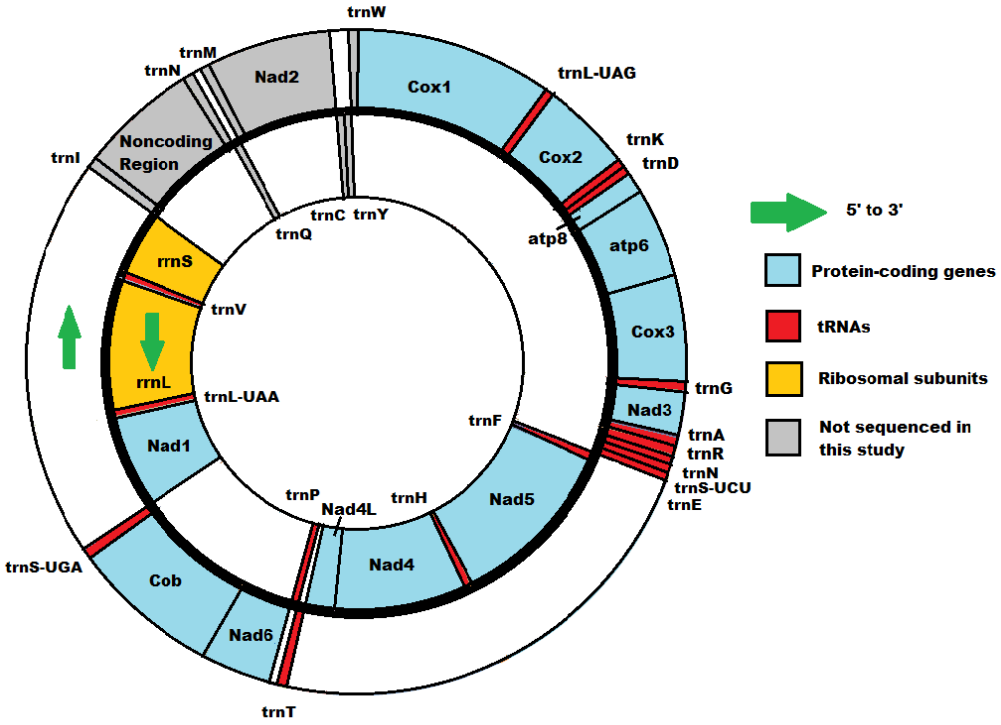

Figure 2. Arrangement of *E. superba* mitochondrial genome

Table 1. Profile of mitochondrial features of *E. superba* sample 813M

Accession Number: JQ286347

% A/T: 67.48 **% Unknown Bases:** 1.43 **Size:** 13326 bp

Feature	Size	Position From	Position To	Strand	Intergenic Nucleotides	% A/T
Cox1[a]	1486	1	1486	+		61.02
trnL-uag[b]	66	1503	1568	+	16	62.12
cox2	688	1569	2256	+	0	63.81
trnK-uuu	69	2257	2325	+	0	63.77
trnD-guc	68	2326	2393	+	0	76.47
atp8	159	2394	2552	+	0	75.47
atp6	675	2546	3220	+	-7	65.63
cox3	793	3220	4012	+	-1	62.55
trnG-ucc	67	4013	4079	+	0	67.16
nad3	354	4080	4433	+	0	64.12
trnA-ugc	66	4433	4498	+	0	66.67
trnR-ucg	67	4500	4566	+	1	61.19
trnN-guu	65	4567	4631	+	0	73.85
trnS-ucu	68	4632	4699	+	0	63.24
trnE-uuc	69	4701	4769	+	1	72.46
trnF-gaa	68	4869	4936	-	99	72.06
nad5	1731	4936	6666	-	-1	68.28
trnH-gug	66	6667	6732	-	0	69.70
nad4	1338	6733	8070	-	0	67.56
nad4L	300	8064	8363	-	-7	66.67
trnT-ugu	66	8366	8431	+	2	74.24
trnP-ugg	67	8433	8499	-	1	71.64
nad6	522	8503	9024	+	3	71.26
cob	1137	9024	10160	+	-1	65.61
trnS-uga	71	10181	10251	-	20	67.61
nad1	939	10269	11207	-	17	68.69
trnL-uaa	66	11224	11289	-	16	60.61
rrnL	1311	11308	12618	-	18	74.90
trnV-uag	72	12619	12690	-	0	72.22
rrnS	631	12696	13326	-	5	73.22

[a]Beginning of *cox1* and end of *rrnS* failed to be sequenced. Actual size of these regions may not be significantly longer than shown here.

[b]tRNA names are followed by their anticodon sequence

Table 2. Start and Stop Codons for five new sequences as well as published sequences by Machida et. al. (2004) and Shen et. al. (2010)

	Codon	
Feature	**Start**	**Stop**
cox1	Missing	TAA
cox2	ATA	T--
atp8	ATC	TAA
atp6	ATG	TAA
cox3	ATG	T--
nad3	ATT	TAA
nad5	ATG	TAA
nad4	ATG	TAA
nad4L	ATG	TAA
nad6	ATT	TAA
cob	ATG	TAA
nad1	ATA	TAG

Table 3. Variation within protein-coding sequences

Gene	Bases	Variable Sites	% Sites Variable in Sequence	Variable Sites Affecting AA[a] Sequence	% Sites Causing AA Sequence Change
cox1	1473	26	1.77	0	0
cox2	688	19	2.76	0	0
atp8	159	5	3.14	0	0
atp6	675	12	1.78	0	0
cox3	793	16	2.02	0	0
nad3	355	5	1.41	0	0
nad5	1731	44	2.54	1[b]	0.058
nad4	1338	33	2.47	2[c]	0.149
nad4L	300	6	2.00	0	0
nad6	522	14	2.68	0	0
cob	1137	19	1.67	0	0
nad1	939	19	2.02	0	0

[a]Amino Acid
[b]Leucine to Methionine at 6488 bp
[c]Histidine to Glutamine at 7530bp in sample 980349; Threonine to Tyrosine at 7824 bp in sample 841M and 7836 bp in sample 980349