

**A DRAFT *DE NOVO* MITOCHONDRIAL GENOME ASSEMBLY FOR
THE HYACINTH MACAW**

An Undergraduate Research Scholars Thesis

by

AUSTIN CHRISTIANSEN WANG

Submitted to Honors and Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor:

Dr. Christopher Seabury

May 2015

Major: Biomedical Sciences
Psychology

TABLE OF CONTENTS

	Page
ABSTRACT.....	1
DEDICATION.....	3
ACKNOWLEDGEMENTS.....	4
CHAPTER	
I INTRODUCTION.....	5
II METHODS.....	8
III RESULTS.....	9
IV CONCLUSION.....	13
REFERENCES.....	14
APPENDIX.....	17

ABSTRACT

A Draft *de novo* Mitochondrial Genome Assembly for the Hyacinth Macaw. (May 2015)

Austin Christiansen Wang
College of Veterinary Medicine & Biomedical Sciences
Department of Psychology
Texas A&M University

Research Advisor: Dr. Christopher Seabury
Department of Veterinary Pathobiology

Establishing genome assemblies is important for comparative and population studies that examine genetic diversity¹. This is especially vital for threatened animal species. The hyacinth macaw (*Anodorhynchus hyacinthinus*), native to the countries of Bolivia, Paraguay, and Brazil, is a vulnerable species currently in decline, as a result of illegal trade and habitat loss². Maintaining species and genomic diversity is crucial to the biological and ecological balance of many environments. Thus, genome-wide data for a threatened species are particularly useful for population analyses and future conservation efforts. Despite dwindling population numbers, no hyacinth mitochondrial draft genome assembly currently exists. Therefore, the goal of this study was to produce a hyacinth draft *de novo* mitochondrial genome assembly with annotation by using BLAST and other published tools. CLC Genomics Workbench was used to map and extract reads and perform a *de novo* assembly to yield a complete mitochondrial genome for the hyacinth macaw. The mitochondrial sequence was 17,000 bp long and contained 13 protein coding genes, 2 rRNA genes, 22 tRNA genes, and a control region. Comparative analysis showed that many mitochondrial gene regions were conserved between the hyacinth macaw and several highly similar species found through a BLAST search. Continuation of such studies will allow for further population analyses involving the diversity of maternal lineages for this species.

This work will directly facilitate modern genomics research for the hyacinth macaw and other avian species.

DEDICATION

To my brother, Aaron Stephen Wang, for all his years of company, and my sister, Dr. Amy Michelle Wang. Also to my parents, Dr. Yu Feng Wang and Mrs. Huey Ying Wang, for all their love and support in helping me get to where I am today.

ACKNOWLEDGEMENTS

I would like express my deepest gratitude to my professor and research advisor, Dr. Christopher Seabury, for guiding me and allowing me to work on this project in his lab. To research associate Mr. Eric Bhattarai, it has been a pleasure getting to know you during my time in the Seabury Lab. Also, to graduate students Ms. Yvette Halley and Mr. David Oldeschulte, I would like to thank you both for your help in this process. Finally, a massive thank you goes to graduate student Mrs. Maral Molaei for mentoring me throughout this project and putting up with all of my many questions. I could not have done this without the support of these fine individuals.

CHAPTER I

INTRODUCTION

Even though many non-model and non-agricultural animal species around the world hold biological significance, current research initiatives for many of these species only include minimal genome-wide sequence and polymorphism data, thus hindering the application of genomic approaches for addressing biological questions in these species^{1,3}. The avian order Psittaciformes is an example of a taxonomic group with relatively few genomic resources. Only recently have some genomic resources become available through the completion of the Puerto Rican parrot genome (*Amazona vittata*)⁴, and the Budgerigar (*Melopsittacus undulatus budgerigar*) sequencing initiative (<http://aviangenomes.org/budgerigar-raw-reads/>). Within the order Psittaciformes, there are three families: the Psittacidae (true parrots), Cacatuidae (cockatoos), and Strigopidae (New Zealand parrots)⁵. Within the Psittacidae family alone, there are over 300 divergent species that exhibit considerable variation in terms of geography, phenotype, cognition, and behavior⁶⁻⁹, yet little is known about the individual genomes of these unique creatures. In comparison to other avian families, Psittacidae has been documented to have the highest number of threatened or endangered avian species¹⁰⁻¹¹, making their study important for future conservation efforts. Furthermore, the conservation status of this family has promoted research in many biological disciplines that focus on genetics, conservation biology, nutrition, and natural history^{1, 12-16}.

Currently, the most commonly studied avian genomes provide representation from the orders Galliformes (*Gallus gallus*, chicken; *Meleagris gallopavo*, turkey) and Passeriformes

(*Taeniopygia guttata*, Zebra finch)¹⁷⁻¹⁹, with recent efforts providing some genomic insight into the Psittacidae⁴ (<http://aviangenomes.org/budgerigar-raw-reads/>), which are expected to be essential for examining key features of this family, such as longevity and intelligence^{7-9, 20}. Evolutionarily, members of Psittaciformes and Galliformes diverged about 117.3 million years ago, while members of Psittaciformes and Passeriformes diverged more recently, about 96.4 million years ago (for review, see <http://www.timetree.org>)²¹⁻²². Notably, the production of new avian genome assemblies, such as those representative of Darwin's finch (*Geospiza fortis*; <http://gigadb.org/darwins-finch/>), the Budgerigar (<http://aviangenomes.org/budgerigar-raw-reads/>), and other species of Psittacidae will provide tangible insight into avian traits of interest, including longevity, intelligence, body size, and adaptability^{1, 6-9, 11, 14-15, 19-20}.

The hyacinth macaw (*Anodorhynchus hyacinthinus*; see Figure 1) is a species of conservation concern within the family Psittacidae, and ranges throughout regions of South America². However, because of declining population trends and relatively small (i.e., extant) population sizes², conservation efforts will be important for maintaining genetic diversity in this species. For these reasons, we selected the hyacinth macaw for mitochondrial genome sequencing, assembly, and annotation. The objective of this research project is to produce a high-coverage hyacinth draft *de novo* mitochondrial genome assembly with annotation for future population analyses, and more specifically, for examining the diversity of maternal lineages present within extant populations. Importantly, monitoring levels of genetic diversity over time will be useful for developing and assessing the impact of modern management practices. It was hypothesized that a complete, high-coverage hyacinth draft *de novo* mitochondrial genome assembly can be generated from Illumina paired-end genomic DNA libraries by using a well-annotated and

established avian mitochondrial assembly from a similar species (i.e., scarlet macaw)¹, both for initial sequence read capture prior to *de novo* assembly, and for subsequent comparative annotation.



Figure 1. Hyacinth macaw. An image of a hyacinth macaw (*Anodorhynchus hyacinthinus*).

From Wikimedia Commons.

CHAPTER II

METHODS

All data analyses were performed using existing high performance compute clusters and software available in the Seabury Lab¹. First, the hyacinth macaw Illumina Hi-Seq 2000 sequence reads were mapped onto a published scarlet macaw mitochondrial chromosome using the CLC reference mapping algorithm¹. The reads were then extracted and *de novo* assembled using a pipeline employing CLC Genomics Workbench¹. A 2636 bp partial hyacinth macaw mitochondrial sequence (Genbank Accession EF104124.1), specifically a fragment encompassing tRNA-Glu, DLoop, and tRNA-Phe, was utilized to help manually resolve three contigs from the *de novo* assembly into one single contig representing the complete mitochondrial sequence. Thereafter, we used both the scarlet macaw¹ and the annotated conure mitochondria (Genbank Accession JQ782214.1) as a guide to annotate the hyacinth macaw mitochondrion using the full suite of available BLAST tools (blastn, bl2seq, blastp; <http://blast.ncbi.nlm.nih.gov/>). The annotation included 13 protein coding genes (*ND1*, *ND2*, *COX1*, *COX2*, *ATP8*, *ATP6*, *COX3*, *ND3*, *ND4L*, *ND4*, *ND5*, *ND6*, *CYTB*), two ribosomal RNA genes (*12S*, *16S*), and the DLoop. Afterwards, tRNAscan-SE (<http://lowelab.ucsc.edu/tRNAscan-SE/>) was used to predict tRNA genes in the mitochondrial sequence, as recently described¹. The ExPASy translate tool (<http://web.expasy.org/translate/>) was used to translate nucleotide sequences of the 13 protein coding genes from the mitochondrial sequence into amino acid sequences.

CHAPTER III

RESULTS

The sequencing and assembly process was successful at reconstructing the mitochondrial genome of the hyacinth macaw at an average coverage of 227× from over 42,000 mapped reads. Annotation of the assembled mitochondrial sequence was comparatively executed using the complete mitochondrial sequence of *Aratinga acuticaudata* (blue-crowned parakeet), as it had the highest overall identity among all curated sequences in a blastn search of the NCBI nr/nt database. Using the *A. acuticaudata* mitochondrial genome refseq (GenBank Accession JQ782214.1) in conjunction with BLAST (blastn, bl2seq, blastp; <http://blast.ncbi.nlm.nih.gov/>), we annotated 13 hyacinth macaw mitochondrial protein coding genes (*ND1*, *ND2*, *COX1*, *COX2*, *ATP8*, *ATP6*, *COX3*, *ND3*, *ND4L*, *ND4*, *ND5*, *ND6*, *CYTB*) and two ribosomal RNA genes (*12S*, *16S*). Nine of the protein coding genes utilize ATG as the start codon. *ND2* and *ND3* use ATA as the start codon, while *COX1* and *ND5* use GTG. *COX2*, *ATP8*, *ATP6*, *ND4L*, *ND5*, and *CYTB* utilize TAA as the stop codon. *ND1* and *COX1* end with AGG, while *ND6* ends with TAG; additionally, four genes (*ND2*, *COX3*, *ND3*, and *ND4*) utilize truncated stop codons (see Table 1). Also, the *ND3* gene was found to have one extra, untranslated cytosine nucleotide. This finding has been observed in some species of birds and turtles²³.

Table 1. Predicted genes in the complete mitochondrial genome of *A. hyacinthinus*.

Gene	Position		Intergenic nucleotides	Size		Codon	
	From	To		(nn)	(aa)	Start	Stop
12S rRNA	1	969	0	969			
tRNA^{VAL}	970	1039	0	70			
16S rRNA	1040	2612	0	1573			

Table 1. Continued.

Gene	Position		Intergenic nucleotides	Size		Codon	
	From	To		(nn)	(aa)	Start	Stop
tRNA ^{LEU}	2613	2687	6	75			
ND1	2694	3674	-2	981	326	ATG	AGG
tRNA ^{ILE}	3673	3744	5	72			
tRNA ^{GLN}	3750	3820	0	71			
tRNA ^{MET}	3821	3888	0	68			
ND2	3889	4928	0	1040	346	ATA	TA_
tRNA ^{SEC}	4929	4999	1	71			
tRNA ^{ALA}	5001	5069	1	69			
tRNA ^{ASN}	5071	5144	2	74			
tRNA ^{CYS}	5147	5213	0	67			
tRNA ^{TYR}	5214	5283	9	70			
COX1	5293	6840	0	1548	515	GTG	AGG
tRNA ^{SER}	6841	6906	7	66			
tRNA ^{ASP}	6915	6983	2	69			
COX2	6986	7669	1	684	227	ATG	TAA
tRNA ^{LYS}	7671	7742	1	72			
ATP8	7744	7911	-10	168	55	ATG	TAA
ATP6	7902	8585	-1	684	227	ATG	TAA
COX3	8585	9368	0	784	261	ATG	T__
tRNA ^{GLY}	9369	9437	0	69			
ND3	9438	9788	0	351	116	ATA	TA_
tRNA ^{ARG}	9789	9858	1	70			
ND4L	9860	10156	-7	297	98	ATG	TAA
ND4	10150	11542	0	1393	464	ATG	T__
tRNA ^{HIS}	11543	11611	0	69			
tRNA ^{SER}	11612	11677	0	66			
tRNA ^{LEU}	11678	11747	0	70			
ND5	11748	13562	11	1815	604	GTG	TAA
CYTB	13574	14713	0	1140	379	ATG	TAA
tRNA ^{THR}	14714	14781	4	68			
tRNA ^{PRO}	14786	14854	3	69			
ND6	14858	15370	2	513	170	ATG	TAG
tRNA ^{GLU}	15373	15441	1	69			
D-loop	15443	16933	0	1491			
tRNA ^{PHE}	16934	17000	0	67			

Analyses using tRNAscan-SE (<http://lowelab.ucsc.edu/tRNAscan-SE/>) and BLAST estimated and revealed evidence for 22 mitochondrial tRNA genes (see Table 1). The genomic locations and boundaries for the tRNA genes were supported through use of both methods. The tRNA-Phe gene was manually predicted and annotated. The consensus mitochondrial genome spanned 17,000 contiguous base pairs and possessed an average GC content of 46.39%. The overall base composition of *A. hyacinthinus* is as follows: A (30.39%), T (23.21%), G (13.97%), C (32.43%).

The hyacinth mitochondrial sequence was compared to those from several related species that were identified using a BLAST search (see Table 2). Query coverage was 99%, and identity values ranged from 89% – 91% for all species that were compared to the hyacinth macaw (Table 2). Interestingly, in all compared bird species (Table 2), the stop codon TAA is predicted to be completed by addition of a 3' A residue for *ND2* and *ND3*. That same stop codon is also predicted to be completed by addition of 3' A residues for *COX3* and *ND4* in all compared species (Table 2). The sizes of the mitochondrial genomes for all examined species ranged from 16,970 bp – 17,030 bp.

Table 2. A BLAST search table comparing the mitochondrial sequence of the hyacinth macaw to those of other highly similar bird species.

Description	Max Score	Total Score	Query Cover	E- Value	Identity	Accession
<i>Aratinga acuticaudata</i> mitochondrion	22426	22426	99%	0.0	91%	JQ782214.1
<i>Aratina brevipes</i> mitochondrion	22223	22223	99%	0.0	90%	KC936100.1
<i>Ara glaucogularis</i> mitochondrion	22220	22220	99%	0.0	90%	JQ782215.1
<i>Primolius couloni</i> mitochondrion	21955	21955	99%	0.0	90%	KF836419.1

Table 2. Continued.

Description	Max Score	Total Score	Query Cover	E-Value	Identity	Accession
<i>Rhynchopsitta terrisi</i> mitochondrion	20452	20538	99%	0.0	89%	KF010318.1
<i>Aratina pertinax chrysogenys</i> mitochondrion	18231	22125	99%	0.0	90%	HM640208.1
<i>Ara macao</i> mitochondrion	19848	22355	99%	0.0	91%	CM002021.1

CHAPTER IV

CONCLUSION

This project demonstrates that a high-coverage draft *de novo* mitochondrial genome assembly can be generated with relative ease, with few data, and at low cost, despite the fact that the hyacinth macaw has no existing genomic maps. The hyacinth macaw mitochondrial genome assembly was supported by the high values of coverage and identity that were observed following comparison to mitochondrial sequences from related species using BLAST.

Conservation of mitochondrial nucleotide identity was evident between the hyacinth macaw and the avian species listed in Table 2, with the other avian mitochondrial sequences displaying high identity values compared to the hyacinth macaw draft *de novo* mitochondrial genome assembly. Incomplete stop codons were observed in all examined birds, as the *ND2*, *COX3*, *ND3*, and *ND4* genes were predicted to require some level of polyadenylation to complete the stop codon. This supports the concept that similar genetic mechanisms are conserved between related species.

This project yielded a complete mitochondrial genome for the hyacinth macaw, which will facilitate examination of the diversity of maternal lineages within this species. Furthermore, this resource will aid in studying mitochondrial haplotype variation within hyacinth macaw populations. Future studies could attempt to sequence and assemble the entire nuclear genome for the hyacinth macaw. This would certainly facilitate and promote conservation efforts for this and possibly other threatened avian species. Increased availability of genome-wide information will positively augment management and conservation plans, help refine phylogenetic relationships, and literally provide millions of variable genetic markers for future genetic studies.

REFERENCES

1. Seabury CM, Dowd SE, Seabury PM, Raudsepp T, Brightsmith DJ, et al. (2013). A Multi-Platform Draft *de novo* Genome Assembly and Comparative Analysis for the Scarlet Macaw (*Ara macao*). PLoS ONE 8(5): e62415.
2. BirdLife International 2014. *Anodorhynchus hyacinthinus*. The IUCN Red List of Threatened Species. Version 2014.2. <www.iucnredlist.org>.
3. Seabury CM, Bhattarai EK, Taylor JF, Viswanathan GG, Cooper SM, et al. (2011). Genome-wide polymorphism and comparative analyses in the white-tailed deer (*Odocoileus virginianus*): a model for conservation genomics. PLoS ONE 6: e15811.
4. Oleksyk TK, Pombert J-F, Siu D, Mazo-Vargas A, Ramos B, et al. (2012). A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education. GigaScience 1: 14.
5. Christidis L, Boles WE (2008). Systematics and Taxonomy of Australian Birds. Canberra: CSIRO Publishing. p. 200. ISBN 9780643065116.
6. Frynta D, Liškova S, Bültmann S, Burda H (2010). Being Attractive Brings Advantages: The Case of Parrot Species in Captivity. PLoS ONE 6: e12568.
7. Munshi-South J, Wilkinson GS (2006). Diet influences life span in parrots (Psittaciformes). The Auk 123: 108–118.
8. Huber L, Gajdon GK (2006). Technical intelligence in animals: the kea model. Anim Cogn 9: 295–305.
9. Pepperberg IM (2006). Grey parrot numerical competence: a review. Anim Cogn 9: 377–391.
10. Collar NJ (1997). Family Psittacidae (Parrots) In: Del Hoyo J, Elliot A, Sargatal J, editors. Handbook of the Birds of the World. Vol. 4. Lynx Edicions; Barcelona. 280–477.

11. Bennett PM, Owens PF (1997). Variation in extinction risk among birds: chance or evolutionary predisposition? *Proc R Soc B* 264: 401–408.
12. Tavares ES, Baker AJ, Pereira SL, Miyaki CY (2006). Phylogenetic relationships and historical biogeography of neotropical parrots (Psittaciformes: Psittacidae: Arini) inferred from mitochondrial and nuclear DNA sequences. *Syst Biol* 55: 454–470.
13. Presti FT, Oliveira-Marques AR, Caparroz R, Biondo C, Miyaki CY (2011). Comparative analysis of microsatellite variability in five macaw species (Psittaciformes, Psittacidae): Application for conservation. *Genet Mol Biol* 34: 348–352.
14. Manning AD, Lindenmayer DB, Barry SC, Nix HA (2007). Large-scale spatial and temporal dynamics of the vulnerable and highly mobile superb parrot. *J Biogeogr* 34: 289–304.
15. Ribas CC, Moyle RG, Miyaki CY, Cracraft J (2007). The assembly of montane biotas: linking Andean tectonics and climatic oscillations to independent regimes of diversification in Pionus parrots. *Proc Biol Sci* 274: 2399–2408.
16. Cornejo JE, Dierenfeld E, Bailey CA, Brightsmith DJ (2011). Predicted metabolizable energy density and amino acid profile of the crop contents of free-living scarlet macaw chicks (*Ara macao*). *J Anim Physiol Anim Nutr.*
17. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, et al. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.
18. Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, et al. (2010). Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol* 8: e1000475.
19. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, et al. (2010). The genome of a songbird. *Nature* 464: 757–762.
20. Montgomery MK, Hulbert AJ, Buttemer WA (2011). Does the oxidative stress theory of aging explain longevity differences in birds? I. Mitochondrial ROS production. *Exp Gerontol* (Epub ahead of print).

21. Hedges SB, Dudley J, Kumar S (2006). TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22: 2971–2972.
22. Kumar S, Hedges SB (2011). TimeTree2: species divergence times on the iPhone. *Bioinformatics* 27: 2023–2024.
23. Mindell, D. P., et al. (1998). "An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles." *Molecular Biology And Evolution* 15(11): 1568-1571.

APPENDIX

HYACINTH MACAW MITOCHONDRIAL SEQUENCE AND ANNOTATION

>Hyacinth_Mitochondrial_Genome

```
AAAAGACTTAGTCCTAACCTTATCGTTAACTCTTGCTCAGCATATACATGCAAGTGTCCGCGTCCCAGTG
TAAACGCCCAAACCCCTTACCAAGGTACAGGGAGCAGGCATCAGGCACACCCACCCACCGTGGCCTAAA
ACGCCCCGCCCGCCACACCCACGGGTACTCAGCAGTAATTAACCTTAAGCAATAAGCGAAAGCCTGA
CTTAGTTAGAGCAACCAGGGTTGGTAAATCTTGTGCCAGCCACCGCGGTCATACAAGAAACCCAAGTTAA
CCATACACGGCGTAAAGAGTGGGCTCAAACCTATCATACCAAATAAGATCAAACCCTGTCCAAGCCGTCA
TAAGCCCAAGACACCCCTAAGTCCAACCTGAAGACGATCTTAACACCCCCGATCCACTCCACCCACTA
AAGCCAGGACACAAACTGGGATTAGATACCCACTATGCCCGGCCCTAAATCTTAATGCCCCAAACACA
AGCATTGCGCTGAGAACTACGAGCACAAACGCTTAAACTCTAAGGACTTGGCGGTGCTCTAAACCCACC
TAGAGGAGCCTGTTCTATAATCGATAACCCACGATCCACCCGACCCTCCTTGCCAAAACAGCCTATATA
CCGCCGTGCCAGCCACCTCATGAGAGCACAGCAGTGAGCCCAACAGCCACAGCCACTAATAAGACAG
GTCAAGGTATAGCCTATGAAGTGAAGAAAATGGGCTACATTTTCTAAAATAGAAAACCCAACGAAAGGG
GCCTGAAACCTACCCCGAGAAGGCGGATTTAGCAGTAAAGCAGGACAATTGAGCCTTCTTTAAACCGGTC
CTAGAGCAGTACACACCCGCCGTCACCTCCTCACAAACCCCAAGCACACTAACTAATATCACACCAAC
CACCAAAGACGAGGTAAGTCGTAACAAGGTAAGTGTACCGGAAGGTGCACTTAGCACATCAGGGCGTAGC
TACAACACAAAGCATTGAGCTTACACCTGAAAGATATCTGCCACCCAGATCACCCTGAAAGCCCACCCT
AGCTCCACCAACCACAATCAAAAATCCACTACTTCTACCCAATAAAACATTACCCCAACTCAGTATAG
GTGATAGAAAAGTACGACCTGGACGCCATAGAGACAGTACCGTAAGGGAAAGGTGAAATAACAATGAAAA
CCAAGCACTACATAGCAAAGATAACCCCTTGTACCTTTTGCATCATGATCTAGCAAGAACAACCAGGCAA
AGTGAACCTAAGCCTGCCACCCCGAAACCCAAGCGAGCTACTACAAGCAACTGTCACGAGCTAACCCGT
CTCTGTTGCAAAAGAGTGGGATGACTTGTTAGTAGCGGTGAAAAGCCAACCGAGCTGGGTGATAGCTGGT
TGCCTGCAAAATGAATCTAAGTTCTCCCTTAACCTCTATCCCCAGACAAGCCTAACTCAAATGTAACAG
CTAAGGGCTACTTAAAGGGGTACAGCCCCTTTAAAAAAGGACACAACCTCCACCAGCGGATAACCCCA
TACACCAACCCCGTGGGCCTTAAAGCAGCCATCCCTAAAGATTGCGTCAAAGCTCCATCAACTAAAAAA
TCCCAAAAATAATGCGACTCCCTACATTCATAGCAAGCTAACCTATGACAATAGATGAATCAATGCTAG
AACGAGTAACCAGGACACCATGTCCCCCTAAGCGCCAGCCTACACACCATTAACAGCAAAACCTCAATAA
TAACCACCCCCCTCCCCACTATTGAACATGTCTGTCAATCCAACCTCAGGGGCGCACACTGGAGTGAT
TAAAGTCTACAAAAGGAACTCGGCAAACCCAAGACCCGACTGTTTACCAAAAACATAGCCTTCAGCCAAA
CAAGTATTGGAGGTGATGCCTGCCAGTGACACCATGTTTAAACGGCCGCGGTATCCTAACCGTGCAAAG
TAGCGCAATCAATTGTCCATAAATCGAGACCTGTATGAACGGCTAAACGAGGTCTTAAGTGTCTCCTGT
AGACAATCGGTGAAACTGATCTTTCTGTACAAAAGCAGAGATAAAACGCATAAGACGAGAAGACCCTGTGG
AACTTCAAAATCAGTAGCCACCCACCCAACCTCACAAACCCACCCAGGCCACCACCCAAGACACTGGCT
AACATTTTTAGGTTGGGGCGACCTTGGAGAAAAACAGACCCTCCAAAAACAAGGCCAAACCCCTTAACTA
AGAGCAACCCCTCAACGTGCCAACAGCACCCAGACCCAGTAAATCTGATCAATGAACCAAGCTACCCAG
GGATAACAGCGCAATCTCCTCCAAGAGCCCTATCGACGAGGAGGTTTACGACCTCGATGTTGGATCAGG
ACATCCTAGTGGTGCAGCCGCTACTAAGGGTTCGTTTGTTC AACGATTAACAGTCTACGTGATCTGAGT
TCAGACCGGAGCAATCCAGGTGCGTTTCTATCTATGACCAACCTTCCCCAGTACGAAAGGACCCGAAAG
TGGGGCCAATACTCCAAGCACGCCCCCTCCCCAGTGATGCCCCCACTTAATCACCAAAGGATCACCCC
CACTACCCCAACGAAAAGGTTGCTAGTGTAGCAGAGCCTGGCAAATGCAAAAGACTTAAACCCCTTTACC
TCAGAGGTTCAAATCCTCTCTCTAGCTCCTACCATGACCTGACCAAGCACCCCTCCCATCCACCTTATCA
TAACACTAGCTTACATAATCCCCATTCTAATTGCCGTGGCATTCCCTAACACTCGTTGAACGAAAGGTCT
```

AAGCTATATACAATCCCGAAAAGGCCCAAATATCGTTGGTCCATTTCGGACTACTTCAACCCGTCGCTGAT
GGTGTCAAGCTATTTATCAAAGAACCAATCCGCCCCTCCACCTCTTCCCCCTTCTATTCTCACAACCC
CAATACTCGCCCTCCTCCTTGCAATTAACAATTTGAGCCCCCTCCCTCTCCCCTTCTCTCTCGTAGACCT
AAACCTTGGATTTCTCTTCTCCTAGCAATATCTAGCCTAGCAGTTTACTCAATCCTATGATCAGGGTGG
GCCTCAAACCTCAAAGTACGCCCTAATTGGCGCACTGCGAGCAGTGTACAAACTATCTCCTACGAAGTAA
CTTTAGCCATCATTCTCCTATCCGTGGTCATACTAAGCGGAACTACACCCTAACCGCCCTCATCACCAC
ACAAGAACCCCTATAACCTAATATTCTCCTCCTGACCTCTAGCAATAATATGGTACATCTCAACGCTAGCT
GAAACAAACCGCTCTCCATTTGACCTCACAGAAGGGGAGTCAGAACTTGTATCAGGTTTTCAACGTAGAAT
ACTCTGCAGGGCCATTTGCTCTATTCTTCTCCTGGCTGAATATGCAAACATCATACTAATAAACACACTAAC
AAGCCTCCTATTCTAAGCCCAAGCGCACTCAACCCACCCCTAGAATTTTTCCCACTCATCCTGGCCACA
AAAACCCCTACTCCTCTCCTCAAGCTTCTTATGAATCCGAGCCTCCTACCCACGATTCCGATACGACCAAC
TCATGCACCTACTTTGAAAAAATTTCTCCCACCTAACACTATCCCTTACCTTTGACACACCAGCATAACC
AATCTCTTACGCGGGCTACCTCCTTACCTAAGGAAATGTGCCTGAATGTCAAGGGTCACTATGATAAAG
TGAACATAGAGGTACACCAACCCTCTCATTTCTAACACTTAGAAAAGCAGGAATCGAACCTACACAGGA
GGAATCAAACCCCTCCATACTTCCCTTATATTATTTCTAGTAAGGTGAGCTAACAAAGCTATCGGGCCC
ATACCCCGAAAATGATGGTTCAACTCCCTCCCCTACTAATAACCCCCCTCGAAAACCTAATCTCTGCCTC
AAGCATCCTCCTAGGAACAACGATCACAATCACAAGTAGCCACTGAATAGCAGCTTGAATTGGACTAGAA
ATCAACACCCCTATCAATTATCCCCCTAATTTCAAATCCCACCACCCACGAGCTATCGAAGCTACAACCTA
AATACTTCTCGTACAAGCAGCTGCCTCTACACTACTACTCTTCTCAGGCATAACTAACGCATGGTGCAC
TGGGCAATGAGACATCACCAACTCTCCTACCCCCCATCATGCCTCCTGCTAACAGCTGCAATTGCTATC
AAACTAGGTCTAACCCCTTCCACTTCTGATTCCCAGAGGTACTTCAAGGATCATCCCTTATCACAGCTC
TACTCCTCTCAACAGTAATAAAACTCCCACCAACCACCATCCTACTCATCACATCCCCTCACTAAACCC
CACACTACTTGCCTCTATATCCATCATATCCATTGCCCTAGGTGGTTGAATAGGACTAAACCAAACACAA
ACCCGAAAAATCATAGCCTTCTCATCCATCTCTCACATTGGTTGAATAACCATCATTATTATCTACAACC
CAAACCTAACCCCTACTAGCTTTCTACATTTACACCTAATAACAATCTCCATCTTTCTCACTATAAACAC
AACCAACACCTTAAACCTCCCAACATTAATAGCCTCTTGAACCAAACCCCCATACTAAGCACGACCCCTT
ATACTGACACTACTATCACTAGCAGGCTCCCCCACTAACAGGCTTCTTACCCAAATGACTCATCATCC
AAGAACTCATTAACAAGAGCTAACCACAATAGCCACAACCATCTCCCTACTATCACTCCTAAGCCTATT
TTTTTACCTACGCCTGGCATACTGTTCAACAATCACACTTCCCCCAACCCCTCAAACAAGATAAAAACAG
TGGTCTACTAAAAACCAACCAATTCGCTAATCCCATACTCACCCCACTATCTACCTCACTACTACCCC
TCGCCCCCATAATACTCACCATCACTTAAGAAGCTTAGGATAATATTAACCAAAGGCCTTCAAAGCCTT
AAACAAGAGTTAAACCCCTTTAGTTTCTGCTAAGATCCGTAGGACACTAACCTACATCTCCTGAATGCAA
CTCAGATGCTTTTCAATTAAGCTAGGGTCTCTCTAGACAGGTGAGCTTCGATCCCACAATGCTCCTAGTTAA
CAGCTAGGTGCCTAAACCAACAGGCCTCTGCCTAAAAGACTCTGATGTGCTTCAAGCACATATCAATGAG
CTTGCAACTCAACATGAACCTTCACTACAGAGTCGATAAGAAGGGGAATTAACCCCTGTAAAAAGGACTA
CAGCCTAACGCTTAAACATTCAGCCATCTTACCAGCTTACCTGTGACCTTAAATCGATGACTCTTCTCAA
CCAACCACAAAGACATTGGCACCCCTCTACCTAATCTTCGGCGCATGGGCGGGCATAGTCGGCACCCGCCCT
AAGCCTACTTATTCGCGCAGAGCTTGGTCAACCGGGGACCCCTCCTAGGAGACGACCAAATCTACAATGTA
GTCGTACACAGCCCATGCCTTCGTAATAATCTTCTTTCATGGTAATACCAATCATGATTGGGGGGTTTGAA
ACTGATTAGTTCCCTCATAATTGGTGCCCTGACATGGCCTTTCCACGTATGAACAACATAAGCTTCTG
ACTACTTCCCCCATCCTTCTCCTCCTATTAGCCTCCTCCACAGTAGAGGCAGGTGCTGGTACGGGCTGA
ACAGTCTACCCCCCTTTAGCTGGGAACCTAGCCCATGCTGGAGCATCAGTAGACCTAGCTATCTTCTCCC
TTCACCTAGCAGGGGTATCCTCCATTCTAGGGGCAATCAACTTTATCACCCACGGCCATCAACATAAAAACC
ACCTGCACTATCACAATACCAAACCCCACTATTTCGTCTGATCCGTCTAATCACAGCCGATTACTTCTA
CTATCCCTACCAGTTCTAGCTGCTGGAATCACTATACTCCTTACAGATCGCAATCTAAACACCACATTCT
TCGACCCTGCTGGGGGAGGAGACCCAATCTTGTATCAACACCTTTTCTGATTCTTCGGTCAACCAGAAAT
ATACATCCTCATCCTACCAGGATTTGGAATCATCTCCCACGTAGTAGCCTACCATGCAGGTAAAAAGAG
CCATTTGGCTACATAGGCATGGTATGAGCAATACTATCAATCGGGTTCCCTAGGATTTATTGTATGAGCCC
ACCATATAATCACAGTAGGAATAGACGTAGACACTCGAGCATACTTACATCCGCCACCATAATCATCGC
CATTCCAACAGGAATTAAGTTTTAGCTGACTAGCTACGCTACATGGGGGGACCATCAAATGGGACCCC
CCTATATTATGAGCCCTCGGATTCATCTTCTGTTACCATCGGAGGCCTTACAGGAATCGTCTTAGCAA
ACTCCTCACTAGACATTGCCCTACACGACACATACTATGTAGTAGCACACTTCCACTATGTCTTATCAAT
AGGTGCTGTCTTTGCCATCCTAGCAGGACTCACCCACTGATTCCCCCTATTCACAGGGTACACCTTAAAC
CAAACATGGGCCAAAGCACACTTTGGAGTTATATTACAGGTGTAAACCTTACCTTCTTCCCCAACACT

TTCTAGGGCTAGCAGGTATGCCACGACGATACTCCGACTATCCAGACGCCTATACATTATGAAACACCCCT
ATCGTCTATTGGATCCCTAATTTCAATAACAGCTGTAATTATACTAACCTTCATTATCTGAGAGGCCTTT
GCCTCTAAACGAAAAGTTGTACAGCCAGAATAACCCCAACCAACGTTGAGTGAATCCACGGCTGCCAC
CCCCATAACCACACCTTCGAAGAACCTGCCTTCGTCCAAGTACAAGAGAGGAAGGAGTCGAACCCCTCATA
ACTAGTTTCAAGCCAGCCGATACTAAACCCTTATGCTTCTCTCTCCCTATAGGGGCGTTAGTAAACT
AATTACATGGCCCTGTCAAAGCCAAACTACAGGTGAAAACCCTGTACACCCCTACATGGCCAACCCCTCA
CAACTAGGATTCCAAGATGCCTCATCCCAATCATAGAAGAAGTATCGAATTCACGACCACGCCCTAA
TAGTCGCCCTAATAATCTGTAGCCTTGACTCTACCTACTCACACTTATGCTAAAGGAAAAACTATCCTC
AACCGCCGTTGACGCTCAAGAAATTGAGCTAATTTGAACAATTCTACCAGCCATCGTCTTATCCTACTA
GCCCTCCCATCCCTGCAGATCCTATACATAATAGACGAAGTACGACGAGCCAGACCTAACCTTGAAAGCCA
TCGGACACCAATGATACTGATCCTATGAGTACACGGACTTCAAAGACCTCTCATTTGATTTCCTACATGAC
CCCCACAACAGACCTCCTGCCCCGCCACTTCCGACTCCTAGAAGTCGACCACCGCGTAATCATCCCAATA
GAATCCCAATCCGCATCATCATTACCGCCGACGACGTACTCCACTCATGAAGTGTCCCCACACTGGGAG
TAAAAACCGACGCCATTCCAGGACGACTTAACCAAACATCATTACCTACCCTCGCCAGGAATCTTTTA
CGGCAATGCTCAGAGATCTGCGGGGCTAACCCACAGCTTCATGCCCATCGTAGTAGAATCCACCCCTC
GGCCACTTCGAAGCCTGATCCTCCCTACTAACCTCCTAATCATTAAGAAGCTATGTACCAGCACTAGCCT
TTTAAGCTAGATAAAGAGGGAAACCCCTCCCCCTCCTAATGACATGCCTCAACTAAACCCTAGCCCTGA
CTCTTCATTATAATCATATCATGACTCACATTCCTCCCTAATCTTCCAACCAAGACATTATCCTTTACCT
CAACAAACCCCCCATCAATAAGGCGCCCGTAACCGCAAAAACACCCCTGAGCCTGACCATGATCCTA
ACCTTCTTCGACCAATTCTCAAGCCATATCTCCTAGGAGTCCCCTAATCTTCCTCTCAATACTATTAC
CTACCCTCCTCCTCCCATAACCAACAACCGATGAATCACTAACCGCTATCCACCTTACAACCTATGAAC
CATTAAACATAATCACAAACAACCTTATAACCCCATTAACAAACCGGGCCACAATGAGCCATTATCCTC
ACATCACTAATAATACTACTACTGACAATCAACCTCCTAGGCTTACTACCCTATACATTACCCCAACCA
CCCAACTATCAATAAACATAGCCCTTGCTTTCCACTATGACTTGCAACCCTGCTCACAGGCCTACGAAA
TCAACCCACAATCTCCCTAGGACACCTTCTACCCGAAGGCACACCTACCCCACTAATCCCGGCCCTAATC
ATTATTGAAACCATCAGCCTACTGATCCGCCCTCTAGCCTTAGGGGTCCGCCTTACAGCCAACCTCACTG
CAGGGCACCTACTCATCAACTCATCTCAACAGCTACCATCACACTACTCCCCTATACCTACAGTATC
CGCCCTCACTACCACAGTCCCTCCTACTGACAATCCTAGAAGTAGCAGTAGCTATAATCCAGGCTTAC
GTATTTCGTCTCTTATTAAGCCTCTATCTACAAGAAAACATCTAATGGCCCAACAGCACACTCCTACCA
CATAGTAGACCCAGCCCATGACCCATCTTCGGAGCAACTGCTGCCCTGCTCACCACATCAGGACTAATT
ATGTGATTCCACTATAACTCCTCGCACCTCCTAACCCCTCGGACTGGTATCAATCCTCATAGTCATACTCC
AATGATGACGAGATATTGTACGAGAGGGGACATTCGAAGGCCACCACACACCAACAGTACAAAAAGGCCCT
TCGATATGGAATAATCCTCTTATTACATCAGAAGTGTTCTTCTTTCTTGGCTTCTTCTGAGCCTTCTTC
CACTCTAGCCTAGCACCCACCCCAAGAACTAGGAAACCAATGACCCCCAACTGGAGTTACACCCCTAAACC
CCTTAGAAGTGCCACTACTAAACACAGCAATCCTCCTAGCCTCTGGAGTCACCGTCACCTGAACACATCA
TAGCATCACCGAAGGAGGCCAAAAACAAGCCACCCAAGCACTAACCCCTCACCATCTTATTAGGCTTATAC
TTTACCATCCTACAAGCAACAGAGTACTATGAAGCACCCCTTCTCAATCGCTGATAGCGTTTATGGATCAA
CTTTCTTCGTAGCTACAGGATTCCATGGACTTCACGTTATAATTGGATCTTCTTCTCCTACTGATCTGCCT
CTTACGACTAATTAAATTCCACTTCACACCGGGTCACCACTTCGGGTTCGAAGCAGCCGCTGATACTGA
CACTTCGTAGACGTTATCTGATTGTTCTCTACTTAACCATCTACTGATGAGGATCTTGCTCTTCTAGTA
TATCCATTACAATAGACTTCCAATCTCTAGAATCTGGTACAACCCCAAGAGAAGGCAATAAACATAATCA
TATTCACTTACCTCCTCCATTATCCTTAGCATAACCCCTAACCACTAACTTCTGACTCACCCAAAT
AACCCCAAGACTCAGAAAACTATCACCCCTACGAATGTGGATTCGACCCCTAGGATCTGCTCGACTTCCA
TTCTCCATCCGATTCTTCTCAGTAGCCATCCTATTCCTCCTATTTGACCTAGAAATTGCCCTCCTACTG
CCCCTACCATGAGCTACCCAACTAAAAACCCCAACCGTCACCCCTAATCTGAGCCTCTACTATCATCTCC
TACTAACCCCTAGGACTAATTTACGAATGAGCCCAGGGAGGACTAGAATGGGCAGAATAAGAGAGTTAGTC
TAAAAACAAGACAGTTGATTTGACTCAACAAACCATAGTCCACCCCTATGACTTTCTTCTCATGTCATTCT
CCGCCTAAGCTTCTGCTCAGCGTTACCCCTAAGTAGCCTAGGGCTGGCCTTTCACCGAGTACACCTTATC
TCTGCCCTACTCTGCCTAGAGAGCATAATACTATCGATATACATCGCCCTATCAACATGACCAATCGAAA
ACCAAGCACCCCTCCTCCTCCCTCACACCAATCCTCATACTAACATTTTCTGCATGTGAGGCAGGACTGG
ACTAGCACTACTAGTAGCCTCTACACGAACCCACGGCTCTGATCACTTACAAAACCTAAACCTCCTACAA
TGCTAAAAATCATCTTACCAACTCTAATACTACTTCCAACAACCTCTCCTCTCGCCCCAAAACCTCATATG
AACAAACACCACAACACATAGTCTATTAATCGCCACCCTAAGTCTACAATGACTAACACCCCTCGTACTAC
CCATACAAAAACCTCACTCAATATACAGGCATCGACCAAGTATCTTCCCCGCTACTAGTCTATCCTGCT

GACTTACGCCCTTATAATCTTAGCAAGTCAAATCATCTACAACACGAGCCACCAACACGAAAACAGAT
CTTACAGCAACCCTAGTCACAGTGCAACCTCTTATCATTCTAGCCTTCTCAACTACAGAACTTATAATA
TTCTACATCTCCTTCGAAGCAACCTTAATCCCCACACTAATTTTAATCACACGGTGAGGAAACCAACCGG
AACGCCTAAGTGCAGGCATCTACCTTCTCTTTTACACCCTCATCAGCTCCCTCCCCCTCCTAATTGCAAT
CCTATATCTACATTCACAAACGGGCACCCTCCACTTCCCCGCTCTAAAACCTCACCCCTCACACCCACAA
CCCATCTCAACCAACACTGATCCCCCTCCTCCTAAACATCGCCCTCCTCATAGCCTTCATAGTAAAAG
CACCTCTCTACGGACTTCACCTATGACTACCCAAAGCCCATGTAGAAGCCCAATCGCAGGATCAATACT
ACTCGCCGCCCTCCTCCTTAAACTCGGCGGATATGGCATCATGCGCATTACCCACCTAACAAGCCCCCCC
ACAAACAACCTCCTTCACTACCCATTATCACCCTCGCCTTATGAGGGGCCCTAATAACCAGCTCCATTT
GCTTACGTCAAATTGACCTAAAATCACTCATTGCCTACTCCTCCGTAAGCCACATAGGCCTAGTCATTGC
TGCATGTATAATCCAAACCCACTGATCCTTTTCAGGAGCTATAATCCTTATAATCTCCCACGGCCTAACC
TCCTCAATGCTATTCTGCTTAGCCAACACAAACTACGAACGCACGCACAGCCGCATCCTCCTACTAACTC
GAGGACTACAACCACTCCTCCCCCTAATAGCTACCTGATGATTATTAGCCAACCTAACAACATAGCCCT
GCCCCCACCACAAACCTAATAGCAGAACTAAGCATCATAATCACACTATCAACTGATCCCCCCCCAACA
ATCATCTTAACTGGGGCTGCCACCTTACTAACCCTCATAACACTATCCATGCTTACAATAACCCAAC
GAGGAGTCTCCCCCTCACATCACAACTTCAAAGCTCCACTACACGAGAACACTTACTGATGACCCT
ACACCTCCTCCCCACTCCTCCTAATCCTAAAACCTGAACCTAATCTCCGGACCCCTCTCATGCAAGTAT
AGTTTAAACCAACATTAGACTGTGACCCTAAAATAGAAGTTAGACTCTTCTTACCTGCCGAGGGGAGG
TTCAACCAACAAGAAGTCTAATTCCTGTATCTGAGTCTAAAACCTCAGCCCCCTTACTTTTAAAGGATA
ACAGCAATCCACTGGTCTTAGGAGCCACCCTCCTGGTCAAATCCAAGTAAAAGTAGTGAAATAGCCC
TACTCCTCAACACTCACACTACTCACACTAACAATATCCTAACCCTACACTCCTTCTATCCTCCT
AAAAACCTTCAAAAACCTCCCCAAAACCAATTACCTTAAACCTCAAACCGCCTTCTAATCAGCCTAGCA
CCAATAACACTCTTACATACTCAGGGTTAGAAAGCATCACCTCTTACTGAGAATGAAAATTCACCATAA
ACTTCAAAAATCCCCTCAGCTTCAAAAATAGATCAATACTCCCTCCTATTCTTCCCCATTGCATTATTCTG
AACATGATCTATCCTGCAATTCTCGACATACTACATAGCATCTGACCACATATTACAAAATTCCTTTCC
TACCTAACCAACATTCCTAATCGCAATGCTCACACTAACCATCGCCAACAACATCTTCATGCTCTTCATTG
GCTGGGAAGGAGTCGGCATCATATCCTTTCTACTCATCAGCTGATGACATGGACGGGCAGAGGCCAACAC
AGCAGCCTTACAAGCCGTGCTCTACAATCGGATTGGAGACATCGGACTTATCCTAAGCATGGCATGACTT
GCCTCCACCTTAACTCCTGAGAGATACAACAAATATTTTCCCCCACAACCAACACTCCCCCTAC
TAGGCCTCATCCTAGCTGCCACAGGAAAACTGCCCAATTCGGCCTCCACCCATGACTACCCGCTGCCAT
AGAGGGCCCCACCCCTGTCTCAGCCCTACTACACTCAAGCACAATAGTAGTTGCTGGAATCTTCTTATTA
ATCCGAACTCACCCCTTACTCACTAACAACAAAACCTGCCCTCACACTATGCCTCTGCCTAGGCGCCATGT
CCACACTATTTGCTGCTACCTGTGCCCTCACACAAAATGACATCAAAAAAATCATTGCATTCTCCACATC
TAGCCAGCTAGGGCTAATAATAGTCACCATCGGACTAAACCTCCCGCAACTAGCCTTCTTACACATCTCA
ACCCACGCCTTCTTCAAAGCCATACTATTCTATGTTTCCAGGATCAATCATCCACAGCCTAGGCGGAGAAC
AAGATATCCGAAAAATAGGGGGCTTGCAAAAAATACTTCCAACAACCACCTCCTGTTTAAACAATTGGAAA
CTTAGCATTAATAGGAACCCCTTCTTGGCAGGATTCTTCTCAAAAAGATCTCATCATTGAAAGCCTAAAC
ACCTCTCACCTAAATGCTTGAGCCCTGACCTTAACTACTCGCCACAGCCTTCCCGCCACATACAGCC
TACGAATAATCCTTCTAGTACAAACAAAATTCACCCGAACACCAACAATCACTCCAATGGACGAAAAACA
CCCACAAAATCATTAAACCAATCACTCGTTTAGCTCTAGGGAGCATCATAGTTGGTCTACTAATCACATCA
TACATAACCCCAACACAAATCCCACCAATGACAATACCCCTGCTAACAAAAGACCACCGCCATCCTAGTAA
CAGCTACAGGCATCATCCTCGCCTTAGAACTCACAGCCACTATCCACACTCTGACCCAACCCAAAACAAAA
TCCTTACTCAAACCTTCTCCCTAACACTAGGGTACTTTAACCCCTAGCCCACCGACCAAGCTCCACAGCC
CTATTAACCTCAGGACAAAAAATTGCCAACCAACCTTACCACACTACACAAAGGACTGATCAAAGC
CTATTTAGGATCATTGCGCTATCTATCCTAATCATCCTATTACTACTATAACCCAAAACCTAATGGCCC
CCAACCTACGAAAACACCACCCCTTCTAAAAATAGTAAACAACCTCTAATCGACCTACCAACCCCTC
AAACATTTGAGCCTGATGAAACTTCGGATCCCTTCTAGGAATCTGCCTAACAAACACAAATCCTAACTGGC
CTGCTCCTAGCCGCCACTACACTGCAGATACCTCCTTAGCCTTCTCATCCGTGGCTAATATATGCCGAA
ACGTACAATATGGTTGACTAATCCGAAACCTCCATGCAAACGGAGCCTCATTCTTCTCATCTGCATTTA
CCTCCACATTTGCCGAGGTTTCTACTACGGCTCATACTATACAAAGAAACCTGAAATACAGGCATCATC
CTTCTACTTACCCTCATAGCCACAGCCTTTGTTGGCTACGTTCTGCCCTGAGGTCAAATATCCTTCTGAG
GGGCTACAGTAATCACAAACCTATTCTCCGCCATCCCCTACATCGGCCACACCCTAGTAGAATGAGCCTG
AGGTGGATCTCTGTAGATAACCCACCCCTAACCCGATTCTTACCCTACATTTCTCCTCCCATTACATA

ATCGCTAGCTTAGTCCTTATTCATCTAACCTTCTTGCACGAATCAGGATCAAATAACCCCTTAGGTATCT
CCTCAAACGTGACAAAATTCCATTCCACCCTTATTTCTCCTTAAAAGACCTGTTAGGGTTCACAATTAT
GCTATCCCTACTCACCACCCTCGCCCTATTTCCCCAACCTACTAGGAGACCCTGAAAACCTCACCCCA
GCAAACCCCTAGTAACCCCCACATATCAAACCAGAATGATACTTCTCTTTGCATATGCAATTCTAC
GCTCAATCCCAACAAACTAGGGGGCGTCTTGGCCCTAGCCGCTCCGTAAGTCTAGTCTATTCTTAAGCCC
CCTGCTACATAAATCCAACAACGAGCTATAACCTTTCGCCCTATGTCCAACCTCCTATTCTGAGACTA
GCTGCCAACCTATTTATTCTAACATGGATTGGAAGCCAACCAGTAGAACACCCTTCATCATCATCGGAC
AACTAGCTTCACTAACCTACTTACCATCATCCTAATCCTACTCCCCACCACCTCCTTCTAGAAAACAA
AATCCTCAACTAACTCTAATAGTTTATAAAAAACATTGGTCTTGTAACCAAAAAGAAGAAGGCTACCAC
TTCTTAGAGTTCTTATCAGAAAAAGAGGACTAAACCTCCATCACCACCTCCAAAGCTGGCATTTTACAT
TAACTATTTCTGACCCTAACTGCCGAATAACCCCCCGAGATAAACCTCGCACAAGTTCTAGCACAA
CAAACAAAGTCAACAACAGCCCCAACCGGTATCAAAAACACACCAGCCCCGAGGAGTAGAACAAAGC
CACCCCGCTAAAGTCCAACCGAGCAAAGAGTACCCCAACACTGTCAACAGTATCCACCTTAAACCTCCA
CCCCATCCCCAAACAACAAGCCCCACACCAGCAGGCACAAACCCCAAAACATACCCCATAACGTGCCAAC
CCCCTCAAGCCTGAGGAAACGGATCAGCTGCCAACGAAACAGAATACACAAAACCCTAACATCCCACC
TAAATACACCATAAAAAGCACCAGAGACACAAAAGAAGCCCTAAACTTGCTAGCCACCCACACCCCA
ACAGACCCAAAATCAACCCAACAACCCCGTAATGAGGAGAGGGGTTGGACGCCACCAACAATGACGCCA
AAACAAAATAATCCCCAAAATAACAAAAATAAATCATAATATTCTTACTTGGACCTAACCAAGACCT
ATGGTCTGAAAAACCATCGTTGTTATCTCAACTACAAGAACCCACGCCCTCAGGGTAGCCCCCCTACC
CCCCACTCTTTAAACCGTGGGGTTATTTACTTAGCTCTAGGCTATGTATATCGTGCATTTAATAATTAT
GCCCTAGACATTATATTAACCACCAAAGACTAATAAATTCATGCTCTAATAACATATATTGTATTATCG
GATTAACCTATGGTACCGCTCGGTTCTACCCCAAGGGCTGAAAACCTCATGATTGATGATAAGCAAGCT
TTATGGTTCAGGTCAAAGTACCGCTACCTTAACTTTACTGGTTGAGTATACGGAAGTGTCTAGTACAA
GAACCTAATGCTCCAAGCCATACACTGGGACTCTTTACATTAAGTATAACGGAAGTGTCTCTA
GTACAAGGGACTTATCGGCCCGCCATAATGGCCCGGAACCTTCTTATCCCGTAAGACTCGTTTCAG
GAGCCCGGTTATTTATTAGTCTGACTACTCACGAGAGATCACCAACCCGGTGAAGTAAGGTTCCGGCCTT
CCCAGCGTCAGGTCCATTCTTTCCCCCTACACCCTTACACATCTTGCCTTTTGCCTCTGGTTCCTCG
GTCAGGCACATAAATCGGTTCACTCACCTTACATTTCTATCGGGTCATCTGGTTCGCTATCCTGGTACTC
CCTACCTCGTAATCGCGACATCTCCAGTGCTCTTGCCTTTGGTTCCTTTTTGGGGTCAACTCACTC
TACATTTCCAGTGCAATGGGTACACAATTGGTTGACATGGACATACAATCCATAGCGAACCCCTTTTTTG
CCTTCTAGAACAATTAATGATACGGTTTCAGGTGTGGGTTGCTCTCATTTTCGCACTGATGCACTTGCT
CCCCATTTCGGTTATGCCCGTTTTCACATCTCTTCCCGCAGTATTTCTTAATTAACGGTTGTTGGACT
GTCTCTCATACCCGGGCAATTCGGTTTGAATCTCAGGGGTTACTTAATGCGACGGTTGAAGTATATTTCCG
TCTCACTTTTTCCCTGATGCACTTTGTTTTACTTGGTTATGGGATCCCCGCAACCTCTCTACTATGAG
ACTATTCGGTCAATGATCGCAGGACATAATTCTTCACTTTTCTTCTTAAACACCTTTACTTAAACT
CCAATTGTCTTAAACGAACCAGGAAAAATCCAAAAACAAACCCCTCACCCCAACAAACACTTACAAGCA
CTTACAAACACTTACGAACAAACAACTCCACCACCTAATTTACGAATCATTACACTAAGCATCACTTTG
TTTAAACACAACTTTTTTTCTTTCTTCAACCACCACCATCATGACCACCTATATCCAACCACCTACCC
AAAACATCCAACCCCTTTAAATTTTACTATTTATTTACTTGAATTTATCACCCCTACTTAAACACACAC
ACCGCCATATCTCCAACACTCATACTCCACCCACCTGACAAACAATAAACAAACAAACCTCGTCTTG
TAGCTCAAGCTAAAGCATGGCACTGAAGACGCCAAGACGGACACCATCACTCCCGAGGAC

12 sRNA

**AAAAGACTTAGTCCTAACCTTATCGTTAACTCTTGCTCAGCATATACATGCAAGTGTCCGGTCCAGTG
TAAACGCCCAAACCCCTTACCAAGGTACAGGGAGCAGGCATCAGGCACACCCACCCTGGCCTAAA
ACGCCCGCCCGCCACACCCCAACGGTACTCAGCAGTAATTAACCTTAAGCAATAAGCGAAAGCCTGA
CTTAGTTAGAGCAACCAGGGTTGGTAAATCTTGTGCCAGCCACCGGGTCATACAAGAAACCAAGTTAA
CCATACACGGCGTAAAGAGTGGGCTCAACTATCATACCAAACCTAAGATCAAACCCCTGTCCAAGCCGTC
TAAGCCCAAGACACCCCTAAGTCCAACCTGAAGACGATCTTAACACCCCGGATCCACTCCACCCCACTA
AAGCCAGGACACAACTGGGATTAGATACCCCACTATGCCCGGCCCTAAATCTTAATGCCCCCAACACA
AGCATTGCGCTGAGAATAACGAGCACAACCGCTTAAACTCTAAGGACTTGGCGGTGCTCTAAACCCACC
TAGAGGAGCCTGTTCTATAATCGATAACCCACGATCCACCCGACCTCCTTGCCAAAACAGCCTATATA
CCGCGTGCAGCCACCTCATGAGACACAGCAGTGAAGCCCAACAGCCACAGCCACTAATAAGACAG
GTCAGGTATAGCCTATGAAGTGAAGAAATGGGCTACATTTTCTAAATAGAAAACCAACGAAAGGG**

GCCTGAAACCTACCCCAGAAGGCGGATTTAGCAGTAAAGCAGGACAATTGAGCCTTCTTTAAACCGGTC
CTAGAGCAGTACACACCGCCCGTCACCTCCTCACAACCCCAAGCACACTAACTAATATCACACCAAC
CACCAAAGACGAGGTAAGTCGTAACAAGGTAAGTGTACCGGAAGGTGCACCTAGCACAT

16sRNA

AAGCCCACCTAGCTCCACCAACCACAATCAAAAATCCACTACTTCTACCCAACATAAACATTACCCCCA
ACTCAGTATAGGTGATAGAAAAGTACGACCTGGACGCCATAGAGACAGTACCGTAAGGGAAAGGTGAAAT
AACAATGAAAACCAAGCACTACATAGCAAAGATAACCCCTTGTACCTTTTGCATCATGATCTAGCAAGAA
CAACCAGGCAAAGTGAACCTAAGCCTGCCACCCCGAAACCCAAGCGAGCTACTCACAAGCAACTGTCAG
AGCTAACCCGTCTCTGTTGCAAAGAGTGGGATGACTTGTAGTAGCGGTGAAAAGCCAACCGAGCTGGG
TGATAGCTGGTTGCCTGCAAATGAATCTAAGTCTCCCTTAACCTCTATCCCCAGACAAGCCTAACTC
AAATGTAACAGCTAAGGGCTACTTAAAGGGGGTACAGCCCTTTAAAAAGGACACAACCTCCACCAGCG
GATAACCCCATACACCAACCCCGTGGGCCTTAAAGCAGCCATCCCTAAAGATTGCGTCAAAGCTCCAT
CAACTAAAAATCCCAAAAATAATGCGACTCCCTACATTCATAGCAAGCTAACCTATGACAATAGATGA
ATCAATGCTAGAACGAGTAACCAGGACACCATGTCCCCCTAAGCGCCAGCTACACACCATTAACAGCAA
AACCTCAATAATAACCACCCCCCTCCCCACTATTGAACATGTCTGTCAATCCAACCTCAGGGGCGCAC
ACTGGAGTGATTAAGTCTACAAAAGGAACTCGGCAAACCCAAGACCCGACTGTTTACCAAAAACATAGC
CTTCAGCCAAACAAGTATTGGAGGTGATGCCGCCCAGTGACACCATGTTTAAACGGCCGCGGTATCCTAA
CCGTGCAAAGGTAGCGCAATCAATTGTCCATAAATCGAGACCTGTATGAACGGCTAAACGAGGTCTTAA
CTGTCTCTGTAGACAATCGGTGAAACTGATCTTTCTGTACAAAAGCAGAGATAAACGCATAAGACGAGA
AGACCCTGTGGAACTTCAAATCAGTAGCCACCCACCCAACCTCACAACCCACCCAGGCCACCACCCA
AGACTGCTAACATTTTTTAGGTTGGGGCGACCTTGGAGAAAAACAGACCTTCAAAAACAAGGCCAAA
CCCCTTAATAAGAGCAACCCCTCAACGTGCCAACAGCACCCAGACCCAGTAAATCTGATCAATGAACCA
AGCTACCCAGGGATAACAGCGCAATCTCTCCAAGAGCCCTATCGACGAGGAGGTTTACGACCTCGAT
GTTGGATCAGGACATCCTAGTGGTGCAGCCGCTACTAAGGGTTCGTTTGTCAACGATTAACAGTCTAC
GTGATCTGAGTTCAGACCGGAGCAATCCAGTTCGTTTTCTATCTATGACCAACCTTCCCCAGTACGAAA
GACCGGAAAGGTGGGGCCAATACTCCAAGCAGCCCCCTCCCCAGTGATGCCCCCACTTAATCACCAA
AGGATCACCCCACTACCCCAACGAAAAAGTT

ND1

ATGACCTGACCAAGCACCCCTCCCATCCACCTTATCATAACACTAGCTTACATAATCCCCATTCTAATTG
CCGTGGCATTCCTAACACTCGTTGAACGAAAGTCCTAAGCTATATACAATCCCGAAAAGGCCCAAATAT
CGTTGGTCCATTTCGGACTACTTCAACCCGTCGCTGATGGTGTCAAGCTATTTATCAAAGAACCAATCCGC
CCCTCCACCTCTCCCCCTTCTATTCCTCACAACCCCAATACTCGCCCTCCTCCTTGCATTAACAATTT
GAGCCCCCTCCCTCTCCCCTTCTCTCTCGTAGACCTAAACCTTGGATTTCTCTCCTCCTAGCAATATC
TAGCCTAGCAGTTTACTCAATCCTATGATCAGGGTGGGCCTCAAACCTCAAAGTACGCCCTAATTGGCGCA
CTGGGAGCAGTGTCACAACTATCTCTACGAAGTAACTTTAGCCATCATTCCTATCCGTGGTTCATAC
TAAGCGGAAACTACACCCCTAACCGCCCTCATCACCACACAAGAACCCCTATACCTAATATTCTCTCCTG
ACCTTAGCAATAATATGGTACATCTCAACGCTAGCTGAAACAAACCGCTCTCCATTTGACCTCACAGAA
GGGGAGTCAGAACTTGTATCAGGTTTCAACGTTAGAATACTCTGCAGGGCCATTTGCTCTATTCTTCTGG
CTGAATATGCAAACATCATACTAATAAACACACTAACAAAGCCTCCTATTCCTAAGCCCAAGCGCACTCAA
CCCACCCCTAGAATTTTTCCCACTCATCTGGCCACAAAACCCCTACTCCTCTCCTCAAGCTTCTTATGA
ATCCGAGCCTCCTACCCACGATTCCGATACGACCAACTCATGCACCTACTTTGAAAAAATTCTCTCCAC
TAACACTATCCCCTCACCTTTGACACACCAGCATACCAATCTCTTACGCGGGCCTACCTCCTTACCTAAG

G

ND1 AA

MTWPSTPPIHLIMTLAYMIPILIAVAFLTLVERKVLQSYMQSRKGNIVGPFGLLPVADG
VKLFIKEPIRPSSTSSPLLFLTTPMLALLLALTIWAPLPLPFLSLVDLNLGFLFLLAMSSLA
VYSILWSGWASNSKYALIGALRAVSQTSYEVTLAIILLSVMVLSGNYTLTALITTTQEP
YLMFSSWPLAMMYYISTLAETNRSFPDLTEGESELVSGFNVEYSAGPFALFFLAEYANIM
LMNTLTLSSLFLSPSALNPPLEFFPLILATKTLTLLSSFLWIRASYPRFRYDQLMHLLWKN
FLPLTSLHLWHTSMPI SYAGLPPYL-

ND2

ATAACCCCTCGCAAACTAATCTCTGCCTCAAGCATCCTCCTAGGAACAACGATCACAATCACAAGTA
GCCACTGAATAGCAGCTTGAATTGGACTAGAAATCAACACCCTATCAATTATCCCCCTAATTTCAAATC
CCACCACCAGGCTATCGAAGCTACAATAAATACTTCCCTCGTACAAGCAGCTGCCTCTACACTACTA
CTCTTCTCAGGCATAACTAACGCATGGTGCCTGGGCAATGAGACATCACCAACTCTCCTACCCCAT
CATGCCTCCTGCTAACAGCTGCAATTGCTATCAAAGTGGTCTAACCCCTTCCACTTCTGATTCCCAGA
GGTACTTCAAGGATCATCCCTTATCACAGCTCTACTCCTCTCAACAGTAATAAACTCCACCAACCACC
ATCCTACTCATCACATCCCCTACTAAACCCACACTACTTGCCTCTATATCCATCATATCCATTGCC
TAGGTGGTGAATAGGACTAAACCAAACAAACCCGAAAAATCATAGCCTTCTCATCCATCTCTCACAT
TGGTTGAATAACCATCATTATTATCTACAACCCAAAATAACCTACTAGCTTTCTACATTTACACCTTA
ATAACAATCTCCATCTTTCTACTATAAACACAACCAACACCTTAAACCTCCCAACATTAATAGCCTCTT
GAACAAAACCCCATACTAAGCAGCACCCTTATACTGACACTACTATCACTAGCAGGCCTCCCCCACT
AACAGGCTTCTACCCAAATGACTCATCATCAAGAAGCTATTAAACAAGAGCTAACCACAATAGCCACA
ACCATCTCCCTACTATCACTCCTAAGCCTATTTTTTTTACCTACGCTGGCATACTGTTCAACAATCACAC
TTCCCCCAACCCCTCAAACAAGATAAAACAGTGGTCTACTAAAAACCAACCAATTCGCTAATCCCAT
ACTACCCCACTATCTACCTCACTACTACCCCTCGCCCCATAATACTCACCATCACTTA

ND2_AA

MTPLAKLISASSILLGTTITITSSHWMAAWIGLEINTLSIIPLIKSHHPRAIEATTKYF
LVQAAASTLLLFSGMNTAWCTGQWDITQLSYPPSCLLLTAAIAIKLGLTPFHFVPEVLQ
GSSLITALLLSTVMKLPPTTILLITSHSLNPTLLASMSIMSIALGGWMLNQTQTRKIMA
FSSISHIGWMTIIIIYNPKLTLAFYIYTLMTISIFLTMNTNTLNLPTLMSWTKPML
STTLMLTLLSLAGLPLTGFLPKWLI IQELIKQELTMTATTISLLSLLSLFFYLRLAYCS
TITLPPNPSNKMKQWSTKNPTNSLIPMLTPLSTSLPLAPMMLTIT

COX1

GTGACCTTAAATCGATGACTCTTCTCAACCAACCACAAGACATTGGCACCCTCTACCTAATCTTCGGCG
CATGGGCGGGCATAGTCGGCACCGCCCTAAGCCTACTTATTTCGCGCAGAGCTGGTCAACCGGGGACCCT
CCTAGGAGACGACCAAATCTACAATGTAGTCGCACAGCCCATGCCTTCGTAATAATCTTCTTCATGGTA
ATACCAATCATGATTGGGGGGTTGGAACTGATTAGTTCCCTCATAATTGGTGGCCCTGACATGGCCT
TTCCACGTATGAACAACATAAGCTTCTGACTACTTCCCCCATCCTTCTCCTCCTATTAGCCTCCTCCAC
AGTAGAGGCAGGTGCTGGTACGGGCTGAACAGCTTACCCCTTTAGCTGGGAACCTAGCCCATGCTGGA
GCATCAGTAGACCTAGCTATCTTCTCCCTCACCTAGCAGGGGTATCCTCCATTCTAGGGCAATCAACT
TTATCACCACGGCCATCAACATAAAACCACTGCATATCACAATACCAACCCCACTATTCTGCTGATC
CGTCTAATCACAGCCGTATTACTTCTACTATCCCTACCAGTTCTAGCTGCTGGAATCACTATACCTT
ACAGATCGCAATCTAAACACCACATTCTGACCCCTGCTGGGGGAGGAGACCCAATCTTGATCAACACC
TTTTCTGATTCTTCGGTCAACCAGAAGTATACATCCTCATCTACCAGGATTTGGAATCATCTCCACGT
AGTAGCTACCATGCAGGTAAAAAGAGCCATTTGGCTACATAGGCATGGTATGAGCAATACTATCAATC
GGGTTCCTAGGATTTATTGTATGAGCCACCATATATTACAGTAGGAATAGACGTAGACACTCGAGCAT
ACTTCACATCGCCACCATAATCATCGCCATTCCAACAGGAATTAAGTTTTAGCTGACTAGCTACGCT
ACATGGGGGACCATCAAATGGGACCCCTATATTATGAGCCCTCGGATTCATCTTCTGTTACCATC
GGAGGCCTTACAGGAATCGTCTAGCAAACCTCCTCACTAGACATTGCCCTACACGACACATACTATGTAG
TAGCACACTTCCACTATGTCTTATCAATAGGTGCTGTCTTTGCCATCTAGCAGGACTCACCCACTGATT
CCCCCTATTACAGGGTACACCTTAAACCAAACATGGGCAAAGCACACTTTGGAGTTATATTACAGGT
GTAAACCTTACCTTCTTCCCCAACACTTCTAGGGCTAGCAGGTATGCCACGACGATACTCCGACTATC
CAGACGCTATACATTATGAAACACCCATATCGTCTATTGGATCCCTAATTTCAATAACAGCTGTAATTAT
ACTAACCTTCAATTATCTGAGAGGCCTTTGCCCTTAAACGAAAAGTTGTACAGCCAGAACTAACCCCAACC
AACGTTGAGTGAATCCACGGCTGCCACCCCATACCACACCTTCAAGAACCTGCCTTCGTCCAAGTAC
AAGAGAGG

COX1_AA

VTLNRWLFSTNHKDIGTLYLIFGAWAGMVGTAALSLLIRAELGQPGTLLGDDQIYNVVVTA
HAFVMIFFVMVPIIMIGGFNWLVLPMIGAPDMAFPRMNNMSFWLLPPSFLLLLASSTVEA
GAGTGWTVYPPLAGNLAHAGASVDLAI FSLHLAGVSSILGAINFITTA INMKPPALSQYQ
TPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRLNNTTFDPAGGGDPILYQHLFWFFGH
PEVYILILPGFGIISHVVAYHAGKKEPFGYMGVMWAMLSIGFLGFIVWAHMHMFTVGMVDV
TRAYFTSATMIIAIPGTIKVFSWLATLHGGTIKWDPMLWALGFIFLFTIGGLTGIVLAN
SSLDIALHDTYYVVAHFHYVLSMGAVFAILAGLTHWFPLFTGYTLNQTWAKAHFGVMFTG

VNLTFFPQHFLGLAGMPRRYSYDYPDAYTLWNTLSSIGSLISMTAVIMLTFIIWEAFASKR
KVVQPELTPTNVEWIHGCPPPYHTFEEPAFVQVQE-

COX2

ATGGCCAACCCCTCACAACCTAGGATTCCAAGATGCCTCATCCCAATCATAGAAGAAGCTGATCGAATTCC
ACGACCACGCCCTAATAGTCGCCCTAATAATCTGTAGCCTTGACTCTACCTACTCACACTTATGCTAAA
GGAAAACTATCCTCAACCGCCGTTGACGCTCAAGAAATTGAGCTAATTTGAACAATTCTACCAGCCATC
GTCCTTATCCTACTAGCCCTCCCATCCCTGCAGATCCTATACATAATAGACGAAGCTCGACGAGCCAGACC
TAACCTTGAAAGCCATCGGACACCAATGATACTGATCCTATGAGTACACGGACTTCAAAGACCTCTCAT
TGATTCTTACATGACCCCAACAACAGACCTCTGCCCGGCCACTTCCGACTCCTAGAAGTCGACCACCGC
GTAATCATCCCAATAGAATCCCAATCCGCATCATCATTACCGCCGACGACGTACTCCACTCATGAACTG
TCCCCACACTGGGAGTAAAAACCGACGCCATTCAGGACGACTTAACCAAACATCATTATTACTACTCG
CCCAGGAATCTTTTACGGCCAATGCTCAGAGATCTGCGGGGCTAACCACAGCTTCATGCCCATCGTAGTA
GAATCCACCCCTCGGCCACTTCGAAGCCTGATCCTCCCTACTAACCTCCTAA

COX2_AA

MANPSQLGFQDASSPIMEELIEFHDHALMVALMICSLVLYLLTLMLKEKLSSTAVDAQEI
ELIWTILPAIVLILLALPSLQILYMMDELDEPDLTLKAIGHQWYWSYEYTFDKDLSFDSY
MTPTTDLLPGHFRLLEVDHRV IIPMESPIRI IITADDVLHSWTVPTLGVKTD AIPGRLNQ
TSFITTRPGIFYGQCSEICGANHSFMP IIVVESTPLGHFEAWSSLLTS-

ATP8

ATGCCTCAACTAAACCCCTAGCCCTGACTCTTCATTATAATCATATCATGACTCACATTCTCCCTAATCT
TCCAACCAAGACATTATCCTTTACCTCAACAAACCCCCCATCAATAAGGCGCCCGTAACCGCCAAAAA
CCACCCCTGAGCCTGACCATGATCCTAA

ATP8_AA

MPQLNPSPLFIMIMSWLTFSLIFQPKTLSFTSTNPPINKAPVTAKNHPWAWPWS-

ATP6

ATGATCCTAACCTTCTTCGACCAATTCTCAAGCCATATCTCCTAGGAGTCCCACTAATCTTCCCTCTCAA
TACTATTACCTACCCTCCTCCTCCCCATACCAACAACCGATGAATCACTAACCGCCTATCCACCTTACA
ACTATGAACCATTAACATAATCACCAACAACCTTATAACCCCATTAACAACCGGGCCACAAATGAGCC
ATTATCCTCACATCACTAATAATACTACTACTGACAATCAACCTCCTAGGCTTACTACCCTATACATTCA
CCCCAACCCCAACTATCAATAAACATAGCCCTTGCCCTTCCACTATGACTTGCAACCCTGCTCACAGG
CCTACGAAATCAACCCACAATCTCCCTAGGACACCTTCTACCCGAAGGCACACCTACCCCACTAATCCCG
GCCCTAATCATTATTGAAACCATCAGCCTACTGATCCGCCCTTAGCCTTAGGGGTCCGCCCTTACAGCCA
ACCTCACTGCAGGGCACCTACTCATCCAACCTCATCTCAACAGCTACCATCACACTACTCCCCATCATACC
TACAGTATCCGCCCTCACTACCACAGTCCCTCCTCCTACTGACAATCCTAGAAGTAGCAGTAGCTATAATC
CAGGCTTACGTATTTCGTCCTTATTAAAGCCTCTATCTACAAGAAAACATCTAA

ATP6_AA

MILTFFDQFSSPYLLGVPLIFLSMLLPTLLLMPNPNRWITNRLSTLQLWLTINMITKQLMT
PLNKPGHKWAIILTSMLMLLLTINLLGLLPTFTPTTQLSMNMAAFPLWLATLLTGLRN
QPTISLGHLLPEGTPPLIPALII IETISLLIRPLALGVRLTANLTAGHLLIQLISTATI
TLLPIMPTVSALTTTVLLLLLILEVAVAMIQAYVFVLLLSLYLQENI-

COX3

ATGGCCACCAAGCACACTCCTACCACATAGTAGACCCAGCCCATGACCCATCTTCGGAGCAACTGCTG
CCCTGCTCACCACATCAGGACTAATTATGTGATTCCTACTATAACTCCTCGCACCTCCTAACCCCTCGGACT
GGTATCAATCCTCATAGTCATACTCCAATGATGACGAGATATTGTACGAGAGGGGACATTCCAAGGCCAC
CACACACCAACAGTACAAAAGGCCTTCGATATGGAATAATCCTCTTATTACATCAGAAGTGTTCTTCT
TTCTTTGGCTTCTTCTGAGCCTTCTTCCACTCTAGCCTAGCACCCACCCAGAAGCTAGGAAACCAATGACC
CCCAACTGGAGTTACACCCCTAAACCCCTTAGAAGTGCCACTACTAAACACAGCAATCCTCCTAGCCTCT
GGAGTACCGTCACCTGAACACATCATAGCATCCCGAAGGAGGCCAAAAACAAGCCACCCAAGCACTAA
CCCTCACCATCTTATTAGGCTTATACTTTACCATCCTACAAGCAACAGAGTACTATGAAGCACCCCTTCTC
AATCGCTGATAGCGTTTATGGATCAACTTTCTTCGTAGCTACAGGATTCATGGACTTCACGTTATAAT
GGATCTTCTTCTACTGATCTGCCTCTTACGACTAATTAATTCCTTACACCCGGGTCACCACTTCG

GGTTCGAAGCAGCCGCTGATACTGACACTTCGTAGACGTTATCTGATTGTTCTCTACTTAACCATCTA
CTGATGAGGATCTT

COX3_AA

MAHQAHSYHMVDPSWPWPIFGATAALLTTSGLIMWFHYNSSHLLTLGLVSIILMVMLQWWRD
IVREGTFQGHHTPTVQKGLRYGMILFITSEVFFFLGFFWAFHSSLAPTPELGNQWPPTG
VTPLNPLEVPLLNTAILLASGVTVTWTHHSITEGGQKQATQALTTLTILLGLYFTILQATE
YYEAPFSIADSVYGSTFFVATGFHGLHVMIGSSFLICLLRLIKFHFTPGHHFGFEAAAW
YWHFVDVIWFLFLYLYWWS

ND3 (removed 1 nucleotide)

ATAAACATAATCATATTCATACTTACCTCCTCCATTATCCTTAGCATAACCCTAACCACA
CTAAACTTCTGACTCACCCAAATAACCCCAGACTCAGAAAACTATCACCTACGAATGT
GGATTTCGACCCCTAGGATCTGCTCGACTTCCATTCTCCATCCGATTCTTCCTAGTAGC
CATCCTATTCCTCCTATTTGACCTAGAAATGCCCCTCTACTGCCCTACCATGAGCTAC
CCAATAAAAACCCCAACCGTCACCCTAATCTGAGCCTCTACTATCATCCTCCTACTAAC
CCTAGGACTAATTTACGAATGAGCCAGGGAGGACTAGAATGGGCAGAATA

ND3_AA

MNMFMLTSSIIILSMTLTLNFWLTQMPDSEKLSPEYECGFDPLGSARLPFSIRFFLVA
ILFLFLDLEIALLLPLPWATQLKHPTVTLIWASTITLLTLGLIYEWAOGGLEWAE

ND4L

ATGTCATTCTCCGCCTAAGCTTCTGCTCAGCGTTCACCCTAAGTAGCCTAGGGCTGGCCTTTCACCGAG
TACACCTTATCTCTGCCCTACTCTGCCCTAGAGAGCATAATACTATCGATATACATCGCCCTATCAACATG
ACCAATCGAAAACCAAGCACCCCTCCTCCTCCCTCACACCAATCCTCATACTAACATTTTCTGCATGTGAG
GCAGGTACTGGACTAGCACTACTAGTAGCCTCTACACGAACCCACGGCTCTGATCACTTACAAAACCTAA
ACCTCCTACAATGCTAA

ND4L_AA

MSFLRLSFCSAFTLSSLGLAFHRVHLISALLCLEMMLSMYIALSTWPIENQAPSSSLTP
ILMLTFSACEAGTGLALLVASTRTHGSDHLQNLNLLQC-

ND4

ATGCTAAAAATCATCTTACCAACTCTAATACTACTTCCAACAACCTCCTCTCGCCCCAAAACCTCATAT
GAACAAACACCACAACACATAGTCTATTAATCGCCACCCTAAGTCTACAATGACTAACACCCTCGTACTA
CCCATACAAAAACCTCACTCAATATACAGGCATCGACCAAGTATCTTCCCGCTACTAGTCCTATCCTGC
TGACTTACGCCCTTATAATCTTAGCAAGTCAAAATCATCTACAACACGAGCCACCAACACGAAAAACAGA
TCTTTCACAGCAACCCTAGTCACAGTGCAACCTCTTATCATTCTAGCCTTCTCAACTACAGAACTTATAAT
ATTCTACATCTCCTTCGAAGCAACCTTAATCCCCACACTAATTTAATCACACGGTGAGGAAACCAACCG
GAACGCCTAAGTGCAGGCATCTACCTTCTCTTTTACACCCTCATCAGCTCCCTCCCCCTCCTAATTGCAA
TCCTATATCTACATTCACAAACGGGCACCCTCCACTTCCCCGCTCTAAAACCTCACCCCTCACACCCCA
ACCCATCTCAACCAAAACACTGATCCCCCTCCTCCTAAACATCGCCCTCCTCATAGCCTTCATAGTAAAA
GCACCTCTCTACGGACTTACCTATGACTACCCAAAGCCCATGTAGAAGCCCCAATCGCAGGATCAATAC
TACTCGCCGCCCTCCTCCTTAAACTCGGCGGATATGGCATCATGCGCATTACCCACCTAACAAGCCCCC
CACAAACAACCTCCTTCACTACCCATTCATCACCCCTCGCCTTATGAGGGGCCCTAATAACCAGCTCCATT
TGCTTACGTCAAATTGACCTAAAATCACTCATTGCCTACTCCTCCGTAAGCCACATAGGCCTAGTCATTG
CTGCATGTATAATCCAAACCCACTGATCCTTTTTCAGGAGCTATAATCCTTATAATCTCCCACGGCCTAAC
CTCCTCAATGCTATTCTGCTTAGCCAACACAAACTACGAACGCACGCACAGCCGCATCCTCCTACTAACT
CGAGGACTACAACCACTCCTCCCCCTAATAGCTACCTGATGATTATTAGCCAACCTAACAACATAGCCC
TGCCCCCACCACAAACCTAATAGCAGAATAAGCATCATAATCACACTATTCAACTGATCCCCCACAAC
AATCATCTTAACTGGGGCTGCCACCTTACTAACCGCCTCATAACACTATCCATGCTTACAATAACCCAA
CGAGGAGTCTCCCCCTCACATCACAACTTCAAAGCTCCACTACACGAGAACACTTACTGATGACCC
TACACCTCCTCCCCACACTCCTCCTAATCCTAAAACCTGAACTAATCTCCGGACCCCTCTCAT

ND4_AA

MLKIILPTLMLLPTLLSPPKLMWTNTTTHSLLIATLSLQWLTPSYYPYKNLTQYTGIDQ
VSSPLLVLSCWLTPMLILASQNLQHEPPTRKQIFTATLVTVQPLIILAFSTTELMMFYI
SFEATLIPTLILITRWGNQPERLSAGIYLLFYTLISSPLLIAILYLHSQTGTLHFPALK

LHPHTPQPISTKHWSPLLLNIALLMAFMVKAPLYGLHLWLPKAHVEAPIAGSMLLAALLL
KLGYGIMRITHLTSPPTNNLLHYPFITLALWGALMTSSICLRQIDLKSLIAYSSVSHMG
LVIAACMIQTHWSFSGAMILMISHGLTSSMLFCLANTNYERTHSRILLLLTRGLQPLLPLM
ATWVLLANLNTMALPPTTNLMAELSIMITLFWNSPPTIILTGAATLLTASYTLSMLTMTQ
RGVLPPIHTTLQSSTTREHLLMTLHLLPTLLLLLILKPELISGPLS

ND5

GTGGAAATAGCCCTACTCCTCAACACACTCACACTACTCACACTAACAACCTATCCTAACCCTTACACTCC
TTCTATCCTCCTAAAAACCTTCAAAAACCTCCCCAAAACCATTAACCTTAACCATCAAACCCGCTTCTCT
AATCAGCCTAGCACCAATAACACTCTTACATACTCAGGGTTAGAAAAGCATCACCTCTTACTGAGAATGA
AAATTACACATAAACTTCAAAAATCCCACTCAGCTTCAAAAATAGATCAATACTCCCTCCTATTCTTCCCCA
TTGCATTATTCGTAACATGATCTATCCTGCAATTCTCGACATACTACATAGCATCTGACCCACATATTAC
AAAATTCTTTTCTACCTAACAACATTCCTAATCGCAATGCTCACACTAACCATCGCCAACAACATCTTC
ATGCTCTTCATTGGCTGGGAAGGAGTCGGCATCATATCCTTTTCTACTCATCAGCTGATGACATGGACGGG
CAGAGGCCAACACAGCAGCCTTACAAGCCGTGCTCTACAATCGGATTGGAGACATCGGACTTATCCTAAG
CATGGCATGACTTGCCTCCACCTTAAACTCCTGAGAGATAACAACAAATATTTTCCCCCACAAAAACCCCA
ACACTCCCCCTACTAGGCCTCATCCTAGCTGCCACAGGAAAATCTGCCCAATTCGGCCTCCACCCATGAC
TACCCGCTGCCATAGAGGGCCCCACCCCTGTCTCAGCCCTACTACACTCAAGCACAATAGTAGTTGCTGG
AATCTTCTTATTAATCCGAACCTACCCCTTACTACTAACAACAAAACCTGCCCTCACACTATGCCTCTGC
CTAGGCGCCATGTCCACACTATTTGCTGCTACCTGTGCCCTCACACAAAATGACATCAAAAAAATCATTG
CATTCTCCACATCTAGCCAGCTAGGGCTAATAATAGTCACCATCGGACTAAACCTCCCGCAACTAGCCTT
CCTACACATCTCAACCCACGCCTTCTTCAAAGCCATACTATTCTATGTTTCAGGATCAATCATCCACAGC
CTAGGCGGAGAACAAAGATATCCGAAAAATAGGGGGCTTGCAAAAAATACTTCCAACAACCACCTCCTGTT
TAACAATTGGAACTTAGCATTAAATAGGAACCCCTTCTGGCAGGATTTCTTCTCAAAGATCTCATCAT
TGAAAGCCTAAACACCTCTCACCTAAATGCTTGAGCCCTGACCTTAAACTACTCGCCACAGCCTTACC
GCCACATACAGCCTACGAATAATCCTTCTAGTACAACAAAATTCACCCGAACACCAACAATCACTCCAA
TGGACGAAAAACAACCCACAAATCATTAACCCAATCACTCGTTTAGCTCTAGGGAGCATCATAGTTGGTCT
ACTAATCACATCATAACATAACCCCAACACAAATCCCACCAATGACAATACCCCTGCTAACAAAGACCACC
GCCATCCTAGTAACAGCTACAGGCATCATCTCGCCTTAGAACTCACAGCCACTATCCACACTCTGACCC
AACCCAAACAAAATCCTTACTCAAACCTTCTCCCTAACACTAGGGTACTTTAACCCCTAGCCCACCGACC
AAGCTCCACAGCCCTATTAAACTCAGGACAAAAAATTGCCAACACCTAATCGATCTGTTCTGATATAAA
AAAATAGGACCAGAAGGACTTGCCAACCTACAAACCATGGCAACCAAAACCTTACCACACTACACAAAG
GACTGATCAAAGCCTATTTAGGATCATTTGCGCTATCTATCCTAATCATCCTATTACTACTATAA

ND5_AA

VEMALLNLTLLTLLTLLTLLTLLTLLPILLKTFKNSPKTITLTIKTAFLISLAPMTLFTYSG
LESITSYEWKFTMNFKIPLSFKMDQYSLFFPIALFVWTSILQFSTYYMASDPHITKFF
SYLTTFLIAMLTLTIANNIFMLFIGWEGVIMSFLLSIWWHGRAEANTAALQAVLYNRIG
DIGLILSMAWLASTLNSWEMQMFSPKTPPLPLLGLILAATGKSAQFGLHPWLPAAAMEG
PTPVSALLHSSTMVAGIFLLIRTHPLLTNNKALTLCCLGAMSTLFAATCALTQNDIK
KIIAFSTSSQLGLMMVTIGLNLPLQLAFLHISTHAFFKAMFLFCSGSIIHSLGGEQDIRKM
GGLQKMLPTTTSCLTIGNLALMGTPFLAGFFSKDLIIESLNTSHLNAWALTLTLLATAFT
ATYSLRMILLVQTKFTRPTITPMDENNPIINPITRLALGSIMVGLLITSYMTPTQIPP
MTMPLLTKTTAILVTATGIILALELTATIHTLTQPKQNPYSNFSLTLYFNPLAHRPSST
ALLNSGQKIANHLIDLFWYKKMGPEGLANLQTMATKTSTTLHKGLIKAYLGSFALSILII
LLLL-

CYTB

ATGGCCCCAACCTACGAAAACACCACCCCTTCTAAAAATAGTAAACAACCTCTCTAATCGACCTACCAA
CCCCCTCAAACATTTTCAGCCTGATGAAACTTCGGATCCCTTCTAGGAATCTGCCTAACAAACACAAATCCT
AACTGGCCTGCTCCTAGCCGCCACTACACTGCAGATACCTCCTTAGCCTTCTCATCCGTGGCTAATATA
TGCCGAAACGTACAATATGGTTGACTAATCCGAAACCTCCATGCAAACGGAGCCTCATTCTTCTTCATCT
GCATTTACCTCCACATTGCCCGAGGTTTCTACTACGGCTCATACCTATACAAAGAAACCTGAAATACAGG
CATCATCCTTCTACTTACCCTCATAGCCACAGCCTTTGTTGGCTACGTTCTGCCCTGAGGTCAAATATCC
TTCTGAGGGGCTACAGTAATCACAAACCTATTCTCCGCATCCCCTACATCGGCCACACCCTAGTAGAAT
GAGCCTGAGGTGGATTCTCTGTAGATAACCCACCCCTAACCCGATTCTTACCCTACATTTCTCCTCCTCC

ATTCATAATCGCTAGCTTAGTCCTTATTCATCTAACCTTCTTGCACGAATCAGGATCAAATAACCCCTTA
GGTATCTCCTCAAACGTGACAAAATTCCATTCACCCTTATTTCTCCTTAAAAGACCTGTTAGGGTTCA
CAATTATGCTATCCCTACTCACCACCTCGCCCTATTCTCCCCAACCTACTAGGAGACCCTGAAAACTT
CACCCCAGCAAACCCCTAGTAACCCCCCACATATCAAACCAGAATGATACTTCCTCTTTGCATATGCA
ATTCTACGCTCAATCCCCAACAACTAGGGGGCGTCTTGGCCCTAGCCGCCTCCGTAAGTCTTCTATTCT
TAAGCCCCCTGCTACATAAATCCAAACAACGAGCTATAACCTTTGCCCCTATGTCCCAACTCCTATTCTG
AGCACTAGCTGCCAACCTATTTATTCTAACATGGATTGGAAGCCAACCAGTAGAACACCCCTTCATCATC
ATCGGACAACCTAGCTTACTAACCTACTTCACCATCATCTAATCCTACTCCCCACCACCTCCTTCTAG
AAAACAAAATCCTCAACTAA

CYTB_AA

MAPNLRKHHPLLKMNNSLIDLPTPSNISAWWNFGSLLGICLTTQILTGLLLAAHYTADT
SLAFSSVANMCRNVQYGWLRNLHANGASFFFICIIYLHIARGFYYSYLYKETWNTGIIL
LLTLMATAFVGYVLPWQMSFWGATVITNLFSAIPYIGHTLVEWAWGGFSVDNPTLTRFF
TLHFLLPFMIASLVLHILTFHESGSNNPLGISSNCDKIPFHPYFSLKDLLGFTIMLSLL
TTLALFSPNLLGDPENFTPANPLVTPPHIKPEWYFLFAYAILRSIPNKLGGVLALAASVL
VLFSLPLLHKSQRAMTFRPMSQLLFWALANLFIPTWIGSQPVEHPFIIIGQLASLTYF
TIILILLPTTSFLENKILN-

ND6

CTAAACTGCCGAATAACCCCCCGAGATAAACCTCGCACAAAGTTCTAGCACAAACAAAGTCAACAAC
AGCCCCCAACCGGCTATCAAAAACACACCAGCCCCGCAGGAGTAGAACAAAGCCACCCCGCTAAAGTCCA
ACCGAGCAAAGAGTACCCCAACACTGTCAACAGTATCCACCTTAAACCCCTCCACCCCATCCCCAAACAAC
AAGCCCCACACCAGCAGGCACAAACCCCAAAACATAACCCATAACGTGCCAACCCCTCAAGCCTGAGGA
AACGGATCAGCTGCCAACGAAACAGAATACACAAAAACCACTAACATCCCACCTAAATACACCATAAAAA
GCACCAGAGACACAAAAGAAGCCCCCTAAACTTGCTAGCCACCCACACCCCAACAGACCCCAAAAATCAA
CCCAACAACCCCGTAATGAGGAGAGGGGTGGACGCCACCAACAATGACGCCAAAACAAAATAATCCCC
AAAAATACAAAAAATAAATCAT

ND6

MIYFLVFLGISFVLASLLVASNPSPHYGVVGLIFGSVVGCWLASLGASFVSLVLFMVYL
GGMLVVFVYSVSLAADPFPQAWGGWHVMGYVLGFVPAGVGLVVWGWGGGFKVDTVDSVGV
LFARLDFSGVALFYSCGAGVFLMAGWLLLLTLFVVLELVRGLSRGVIRAV-

tRNAscan-SE:

Search Mode: Organellar
Searching with: Cove only
Covariance model: TRNA2.cm
tRNA Cove cutoff score: 15

Max intron + var. length: 40
Pseudogene checking disabled

Cove Stats:

Sequences read: 1
Cove-confirmed tRNAs: 19
Bases scanned by covels: 34000
% seq scanned by covels: 100.0 %

Script CPU time: 0.02 s
 Cove CPU time: 68.53 s
 Scan speed: 496.1 bp/sec

Cove analysis of tRNAs ended: Fri Nov 21 10:16:03 PST 2014

Overall scan speed: 496.0 bp/sec

tRNAs decoding Standard 20 AA:	18
Selenocysteine tRNAs (TCA):	1
Possible suppressor tRNAs (CTA,TTA):	0
tRNAs with undetermined/unknown isotypes:	0
Predicted pseudogenes:	0

Total tRNAs:	19

tRNAs with introns: 0

|

Isotype / Anticodon Counts:

Ala : 1	AGC:	GGC:	CGC:	TGC: 1	
Gly : 1	ACC:	GCC:	CCC:	TCC: 1	
Pro : 1	AGG:	GGG:	CGG:	TGG: 1	
Thr : 1	AGT:	GGT:	CGT:	TGT: 1	
Val : 1	AAC:	GAC:	CAC:	TAC: 1	
Ser : 1	AGA:	GGA:	CGA:	TGA: 1	ACT:
GCT:					
Arg : 1	ACG:	GCG:	CCG:	TCG: 1	CCT:
TCT:					
Leu : 2	AAG:	GAG:	CAG:	TAG: 1	CAA:
TAA: 1					
Phe : 0	AAA:	GAA:			
Asn : 1	ATT:	GTT: 1			
Lys : 0			CTT:	TTT:	
Asp : 1	ATC:	GTC: 1			
Glu : 1			CTC:	TTC: 1	
His : 1	ATG:	GTG: 1			
Gln : 1			CTG:	TTG: 1	
Ile : 1	AAT:	GAT: 1		TAT:	
Met : 1			CAT: 1		
Tyr : 1	ATA:	GTA: 1			
Supres: 0			CTA:	TTA:	
Cys : 1	ACA:	GCA: 1			
Trp : 0			CCA:		
SelCys: 1				TCA: 1	

Predicted tRNA Secondary Structures:

