

Another layer: The *Karenia brevis* plastid transcriptome expands the complex story of dinoflagellate RNA processing and tertiary endosymbiosis.

By:

Robert Newby

Committee:

Jeffrey Leblond (Adviser)

Aubrey Cahoon

Mohamed Salem

Department of Biology,
Middle Tennessee State University

May 2015

This work is dedicated to my second grade teacher Mrs. Marcum, who said I'd never be happy outside of science, Dr. Aubrey Cahoon for his mentorship and helping me to get out of my own way, Dr R. Stephen Howard for his friendship and advice, and finally my family for their love and support along the way.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Hyrum Carroll, Yu-wei Wang, Rafet Faeq Al-Tobasei, and Richard Clifton Ewool from the Computational Biosciences Department at MTSU for their help in assembling and vetting contigs from the Illumina data. I would also like to thank Delio Evangelista at MTSU and Conner Stanley from the University of Virginia Wise for their help in performing RT-PCR assays, as well as Dr. Manoj Khadka for providing the *Karenia brevis* cultures used herein. My deepest thanks goes to my committee for their help in shaping and directing my work. Lastly, my gratitude to Dr Richard Dorrell from Cambridge University for providing annotated *Lingulodinium* sequences, and a great deal of valuable advice.

ABSTRACT

Dinoflagellates are a diverse group of protists, which possess the ability to modify their genomic information at a rapid rate compared to other eukaryotes. When adding plastid organelles via endosymbiosis, the genome of the endosymbiont can also be modified by means such as fragmentation into circular chromosomes (minicircles), reorganization of protein structure, and transference of more of their photosynthesis genes to the nucleus than any other eukaryote. Due to being “suited” for endosymbiotic acquisition of organelles, the dinoflagellate family exhibits three distinct plastid types, two of which use chlorophyll *a* along with peridinin and fucoxanthin, respectively. The third plastid type is found in *Lepidodinium*, and appears to be derived from a chlorophyte inclusion, as it uses chlorophylls *a* and *b*. At this time, the peridinin dinoflagellates have been better studied than the fucoxanthin lineages such as *Karenia brevis* and *Karlodinium veneficum*. Although comparative studies have found phylogenetic evidence of similarities between the peridinin and fucoxanthin dinoflagellates, conflicting results and a significant body of empirical evidence suggests substantial differences between them in the areas of genome and membrane structure. In this study, we have used Illumina RNA sequencing to produce the most complete plastid transcriptome of *Karenia brevis* to date. We have analyzed the transcript structure and noted a number of dinoflagellate hallmarks, such as polyuridylation of the 3' end of mRNAs, and loss of consensus start codons. We have also identified evidence

of polyadenylation of the 5' end of protein coding transcripts, which has not been reported so far. Six of the protein sequences produced were used to conduct a comparative phylogenetic analysis, which has identified both previously known, and unreported similarities, between the peridinin and fucoxanthin dinoflagellates, including shared instances of consensus loss when compared to broad group of reference taxa. We have also found discrete instances of invariant site substitution preferences, which are shared between these groups. These findings do not support previous studies that have suggested a common lineage for these groups, but further elucidate a complex pattern of endosymbiont genome modifications associated with the dinoflagellate host cell.

TABLE OF CONTENTS

List of Tables	vii
List of Figures	viii
List of Appendices	ix
Chapter I: Introduction to the dinoflagellates and their plastids	1
Diversity among the dinoflagellates	2
Plastid structure and transcript processing	4
Purpose of this study	6
Chapter II: Sequencing, analysis and translation of <i>Karenia brevis</i> transcripts	7
Chapter Introduction	8
Methods	8
Results	13
Discussion	25
Chapter III: Phylogenetic analysis of 6 plastid proteins from the peridinin dinoflagellates and <i>Karenia brevis</i>	30
Chapter Introduction	31
Methods	32
Results	34
Discussion	46
References	50
Appendices	55

LIST OF TABLES

Table 2-1. Primers used to confirm <i>Karenia brevis</i> contigs.	11
Table 2-2. Contig numbers associated with assembly output and screening.	13
Table 2-3. List of all confirmed plastid derived mRNAs identified from <i>Karenia brevis</i> .	17
Table 2-4. Non-canonical start codons found in <i>Karenia brevis</i> plastid transcripts.	23
Table 2-5. Signal and transit peptide prediction.	24
Table 3-1. Taxa used in phylogenetic analysis.	35

LIST OF FIGURES

Figure 2-1. RT-PCR confirmation of six <i>Karenia brevis</i> plastid mRNAs.	18
Figure 2-2. Example of a mono-cistronic <i>Karenia brevis</i> plastid transcript, petD.	19
Figure 2-3. Example of a poly-cistronic <i>Karenia brevis</i> plastid transcript, psaB-psaC.	20
Figure 2-4. Non-canonical start codons.	22
Figure 3-1. Maximum likelihood phylogenetic tree, based on a 6 protein alignment.	37
Figure 3-2. Maximum likelihood phylogenetic tree, based on a 5 protein alignment.	38
Figure 3-3. Maximum likelihood phylogenetic trees built from the 6 protein concatemer with each dinoflagellate group added individually.	40
Figure 3-4. Multiple sequence alignment of the psbA proteins of 26 taxa.	41
Figure 3-5. Multiple protein sequence alignment of psbD from 26 taxa.	43
Figure 3-6. Multiple sequence alignment of complete petD proteins for 26 taxa.	44
Figure 3-7. Two selections from the MUSCLE multiple sequence alignment of psbB.	45
Figure 3-8. Protein sequence alignment of psbE.	49

LIST OF APPENDICES

Appendix A: <i>Karenia brevis</i> plastid transcripts	56
Appendix B: Translations	93
Appendix C: Substitution model selection	103
Appendix D: Maximum likelihood trees	105
D-1. Preliminary trees for plastid protein candidates	105
D-2. Five protein tree (petD-psbA-psbB-psbC-psbD)	110
D-3. Six protein tree (petD-psbA-psbB-psbC-psbD-psbE)	111
D-4. Five protein tree without <i>Karenia brevis</i> sequences	112
D-5. Six protein tree without <i>Karenia brevis</i> sequences	113
D-6. Five protein tree without peridinin dinoflagellate sequences	114
D-7. Six protein tree without peridinin dinoflagellate sequences	115
D-8. Unigene protein sequence trees	116

CHAPTER I
INTRODUCTION TO THE DINOFLAGELLATES AND THEIR PLASTIDS

Dinoflagellates are a highly diverse group of protists, united under the phylum Dinoflagellata. Most species live as marine or freshwater plankton, with a subset of these capable of producing toxic compounds and reproducing at a rapid rate under the right nutrient and environmental conditions (Hackett et al. 2004). This sudden and rapid reproduction and the corresponding increase in toxin concentrations, called a harmful algal bloom (HAB), is known to cause fish kills and illness in humans that consume fish harvested from these areas (reviewed in Reguera et al. 2012). Due to the economic and public health concerns related to the toxic species, many of these have been the focus of much dinoflagellate research. *Karenia brevis* is one such species, which is found in the warm waters of the Gulf of Mexico. HABs due to increases in *K. brevis* and similar dinoflagellates, have acquired the name “red-tides”, stemming from their characteristic red color. This pigment is due to accessory carotenoids from the red algal origin of their chloroplasts. The group of organisms causing such blooms is colloquially referred to as “red tide algae”.

Diversity among the dinoflagellates

All species of dinoflagellates are flagellated and contain some form of plastid, whether functional or vestigial. The source from which these plastids were derived, their degree of functionality within the organism, the structure of plastid genome (plastome) and level of reduction varies to a large degree between genera, but falls within a few semi-distinct groups (Gagat et al. 2014). Many species have retained the photosynthetic potential from their algal endosymbiote, while others have retained only the more basic

lipid synthesis and biochemical functions, reverting to a heterotrophic lifestyle. These heterotrophic dinoflagellates form one group within dinoflagellates (Jeong et al. 2010). The remaining photosynthetic species fall into 4 categories, based on plastid pigmentation or origin; peridinin dinoflagellates, non-peridinin or fucoxanthin dinoflagellates, green dinoflagellates, and dinotoms. Most of these autotrophic species have plastids that utilize chlorophyll *c* and the peridinin accessory pigment. Based on current evidence, these species are widely thought to have obtained their plastids by engulfing a eukaryotic algal cell already containing a plastid that itself captured as a prokaryotic algae (secondary endosymbiosis). This relationship is referred to as tertiary endosymbiosis (Keeling 2010). *Karenia* species, and other similar taxa, utilize a different characteristic pigment, 19'-hexanoyloxyfucoxanthin (fucoxanthin) (Bjornland et al. 2003). This pigment is associated with another group of algae called the haptophytes, which are secondary endosymbionts themselves (Zapata et al. 2004). The presence of fucoxanthin and comparative phylogenetic analysis of the plastid genes of fucoxanthin dinoflagellates and haptophyte algae, have led to the consensus that these fucoxanthin dinoflagellate species acquired their plastids as a haptophyte tertiary endosymbiote (Tengs et al. 2000). Dinoflagellates of the genus *Lepidodinium* have obtained their plastid from a green algae source, and therefore utilize chlorophylls *a* and *b*, with blooms being the bright green associated with the Chlorophyceae (Minge et al. 2010). A final group containing genera such as *Durinskia* and *Kryptoperidinium* contain plastids resulting from engulfing diatoms, and have been nicknamed "dinotoms" (Inagaki et al. 2000).

Plastid structure and transcript processing

Dinoflagellate cellular structure, especially regarding the host nuclei and plastid membrane structure, presents a number of difficulties in understanding the relationship between the nucleus of the host and the plastid organelle. Host cell effect may have provided a unique set of genomic pressures for ancestral plastids acclimating to their new endosymbiotic relationship, as many dinoflagellates exhibit atypical nuclear structure, such as a lack of histones, significantly higher DNA to protein ratios and liquid-crystal chromosomes which remain condensed throughout interphase (Nassoury et al. 2003; Wisecaver and Hackett 2011). Evidence of serial endosymbiosis and horizontal gene transfer further complicates the picture of plastid integration in dinoflagellates (Takano et al. 2008; Wisecaver and Hackett 2010). Studies of the plastome structure of peridinin dinoflagellates, such as *Heterocapsa triquetra*, have revealed that their plastomes are not structured in the whole-genome rings that are normally associated with plastids. The plastomes of these organisms exist as many discrete rings containing a distinctive “core region” sequence, as well as a single or small number of functional genes (Howe et al. 2008). Fucoxanthin dinoflagellates, on the other hand, seemed to display the conventional whole-plastome ring structure based on initial evidence, but a recent study of *K. veneficum* identified a minicircle present in its plastid, via presence of a characteristic form of RNA sometimes produced when a minicircle is continuously transcribed, called a rolling circle transcript (Dang and Green 2010). Rolling circle transcripts occur when transcription of a minicircle does not terminate upon

completion, resulting in a transcript that has areas of redundancy where the RNA polymerase (RNAP) began transcribing the 5' coding region again (Richardson et al. 2014).

The process of transcription of plastid genes in dinoflagellates also poses interesting questions. A recent study noted extensive editing, and the presence of polyuridylation of the 3' ends of *Karenia mikimotoi* plastid mRNAs. The addition of a poly-U tail to plastid mRNA had previously been thought to have been a characteristic unique to the peridinin dinoflagellates and *Chromera velia* (Dorrell and Howe 2012). Evidence has also been produced that indicates that the current plastids of peridinin and fucoxanthin dinoflagellates share a common haptophyte origin, although it is not sufficiently clear whether or not the differences, such as genome size and structure between the groups, have arisen since a single inclusion event, or are due to separate inclusions of haptophyte ancestors during the history of the dinophyceae (Shalchian-Tabrizi et al. 2006; Yoon et al. 2002; Yoon et al. 2005).

Purpose of this study

The purpose of this study was to identify and catalog putative plastid gene sequences from the fucoxanthin dinoflagellate *K. brevis*. Despite the aforementioned economic importance and interesting evolutionary history of this organism, no complete genome or list of confirmed plastid genes currently exists. With these sequences, I intend to address the following questions:

1. As mentioned before, dinoflagellate plastomes exhibit significant variability in overall structure. With no published plastid genome for any near relatives to rely on as a guide for the composition of the *Karenia brevis* plastid genome, will a plastid transcriptome be a sufficient alternative to identify plastid genes?
2. Since minicircle plastomes have been shown to exhibit unique rolling circle transcription, can the characteristics of the plastid transcripts of *Karenia brevis* be used to make inferences about the form of the genome they were transcribed from?
3. Higher plants extensively process their plastid transcripts after they are made. Is there any evidence of processing or editing in the transcripts identified by this study?

CHAPTER II

SEQUENCING, ANALYSIS AND TRANSLATION OF *KARENIA BREVIS*

TRANSCRIPTS

Chapter Introduction

This chapter describes the sequencing, assembly, and analysis of the *Karenia brevis* transcriptome. The intention was to identify transcripts made from the chloroplast transcriptome and use them to answer several questions: Is it possible to assemble and identify plastid mRNA's with no available genomic sequence? Can the overall structure of the *K. brevis* transcriptome be determined from assembled transcripts? Are plastid genes transcribed as mono-cistronic and/or as poly-cistronic transcripts? How are mRNAs post-transcriptionally processed? Can we distinguish nuclear-encoded, plastid-targeted, transcripts from plastid-encoded transcripts?

Methods

Karenia brevis culture

Cultures of *K. brevis* CCMP 2281 were inoculated from either fresh or frozen stock into a 2 liter Fernbach flask containing 1 liter of sterile seawater at 25°C. The inoculated cultures were then incubated under constant low light ($50\mu\text{E m}^{-2} \text{s}^{-1}$) conditions for 6-8 weeks, whereupon the cells were harvested for nucleic acid extraction by full volume (1L) pelleting in a clinical centrifuge. Cells not immediately processed were flash frozen in liquid nitrogen and stored at -80°C.

Nucleic acid extraction

Cell pellets were resuspended in extraction buffer, and processed by either the RNeasy or DNeasy kits (Qiagen) according to the manufacturer's protocol to extract RNA and DNA, respectively. Extracted nucleic acids were then analyzed with the NanoDrop Lite (ThermoScientific) spectrophotometer to determine extraction yield and purity.

Transcriptome sequencing

mRNA presence was enriched using Ribo-Zero to remove ribosomal RNAs from each sample. Samples were prepared with the TruSeq RNAseq kit (Illumina, San Diego, CA), and ran using the TruSeq SBS sequencing kit v.3 on the HiSeq2000 platform (Illumina, San Diego, CA) at the University of Illinois sequencing center (Springfield, IL). Inserts averaged 260 nucleotides in length, and were read for 100 bases from each end. Paired end reads were analyzed with Casava 1.8.

Transcript de novo assembly

Three software packages were used to assemble the end read data into contiguous sequences (contigs). SOAPdenovo (Luo et al 2012), Trinity (Grabher et al. 2011), and Velvet (Zerbino and Birney 2008) were each used to build a contig set from the *Karenia brevis* read data, from which the best matches could be selected as putative chloroplast genes. In each case, the default k-mer for each program was used (63, 25, and 31 respectively).

Screening for plastid transcripts

A standalone tBLASTx search was conducted for each contig. Every chloroplast genome in GenBank was downloaded (April 2013) and stored locally for faster searching. Each contig was translated into each of the 6 possible reading frames, and searched against the translated plastomes database constructed above. Contigs that matched a portion of a plastid coding region with E-values of 0.001 or less were considered significant and saved.

A sample from *Nicotiana* was run in the same lane, with incomplete data separation. Any contigs that had a strong ($E \leq 0.001$) match to tobacco in the top 16 hits were eliminated. Some bacterial contamination was also present due to xenic nature of *K. brevis* culture. Any contigs that produced a top match containing the string “bacter” in the taxon label, or produced a very close ($E \leq 1 \times 10^{-10}$) nucleotide BLAST match to a bacterial species were removed from consideration. The remaining contigs were sorted into files based on which gene they best matched via BLAST, and within that, which de novo assembly application had produced them.

Plastid transcript manual ID and translation

Unfortunately, the above screens could not eliminate all bacterial and tobacco contamination, so each candidate contig selected above was manually BLAST searched for bacterial homology, and removed from consideration when any of high significance was found. The same process was performed using tBLASTx for *Nicotiana* homology.

Table 2-1: Primers used to confirm *Karenia brevis* contigs.

gene name	Left	Right	Left Primer paired with PolyA	Right Primer paired with PolyT
groEL	GCAGCGGGTTATAATCTGGA	CAAAAGTCACCGCTCCATTT		
clpC	CGTCGAGGAAGCTGAGAAGT	TGGTGCTGCTAAGAGACCT		
petB	ATGTTCCACCCACATCAAT	GGTTAGCGCAGGCAATAAGA		
petD	CCCCCAGGATATGAGTAACG	ACTGCGAAAGCGGCATAG		
psaA	CTTTGGCTAACCGACGTAGC	CTGTGCGTTCCCAAGAAAAT	GAAGAGGAGGGACGTGTCAA	GGAACCAACCAGCTGTGAAT
psaB	TTAAAGCCTTCACGCCATCT	AGCCACATAGCTTGGGTTTG		
psaC	AAAACCGTTCGGTACAATCT	ATCGATCGACTCGTTTCAGG		
psaJ	GGAAC TTGGACGAGGAGACA	GGCGAATTCCTCCAGAAGAT		
psbB	CGGCAAGGAATGTTGTTTT	GCACCCCATGAAGGTGATAC	CCACCCCGAGGTTTTATT	CTCCATCGCCCTTTGTTAAA
psbC	GGGTGGGTTATTAGCGTCAA	GCCATGGCTGAATATCTCGT		
psbD	AAAGCAAAGTCCGCTCAAAC	CTAGCCTGTCCCAACGGATA		
psbE	AAAAACCACCCATTTTTGT	AGCCCAAGAAAGCTCGAAAT		
psbF			CGATGGTTGACAATTCATGC	CCCCAAAATAGAGACGTTCCG
psbH	AAAAACCACCCATTTTTGT	AGCCCAAGAAAGCTCGAAAT		
psbI	GGGGTGTGCTCTAACCCTG	TATTCGGGTTTCGAGATGGA		
psbT	AACACGTACACGGGAGGAGT	ATGATACGTGGCGGATCTCT		
psbV	AACAACAACAAGTCCAAGCA	CTTTTCCTCACCCATTTT		
psbZ	GCTTCAGGCATCCAGTCCT	ACGGTGTGATGGCTGTGTA		
rbcL	AGAATCCCGGTGGCTTATCT	CGTAGAATTACCGGCACGAT	ATCGTGCCGGTAATTCTACG	AGATAAGCCACCGGGATTCT
rpl2	TGCAAGACCACGAACTGAAG	GGACGCCTTCCAATAAACA		
rpoA	TTTAAGGTCACCCGTCAAAGA	TGTCAATAATCGTGCCCGTA		
rpoB	TTGGTGAAGCATGGATTTC	GCTAACTCGCAGGTCAGGAA		
rpoC1	CATGCTCTGCTGTCGAGGTA	GTGGTACTGCAGCAGCTTTG		
rpoC2	ACGAACCCGGTTACAATACG	CCAGCAGCATCTGGTAGGT		
rps2	TTGGTGAAGCATGGATTTC	TAACCTCGCAGGTCAGGAACC		
rps3	GTGAGCTTACCGCTGTGGTT	CATCCGGTCTCCTCCTCATA		
rps4	CGAAAGCTAAATGCCTCCAG	CGAAAGTGGGTTTCTCCAAA	GCCGAACGACGATAATCAAT	CTGGAGGCATTAGCTTTCCG
rps5	GAGGACGTCAACGTGTCAGA	GAGGACGTCAACGTGTCAGA		
rps14	GCAGTGTCTCTTGGACCACA	ATGGGAGTGACATCCTCGAT		
rps15	CATTAGCGTCTCCTCGTGATT	CAAAAGAAGAAGCCACGAT		
rps18	GGACACAAAGGTGAGCCAAC	GGCGAACCTGTAGTGGTGAT		
rps19	AAAAGGGGTTTCGTTTTCGT	TCTTTTCGTTTATGGGCATGT		
secA	TTACCGCTTCTGAACGTGTG	CCTTGACGTCCTGCTCTACC		
secY	ATTTGAAGCAGACGGAGTGC	TTCATCGTGTGCCATGAACT		
ycr4	GCGCGGTGCAAGAAATACTA	GGCAGTCCCTCTTTAAAGTC		

The remaining contigs were then compared to the GenBank database with tBLASTx. The contig for each gene that had highest algal homology and greatest length of coverage was selected as the “best candidate” for each gene, and used for all further analysis. The coding region of these putative transcripts was translated in-frame with Virtual Ribosome (Wernersson 2006) and aligned with protein sequences from taxa related to *K. brevis* for final confirmation of gene identification. Each confirmed putative transcript was then used for a template-based reassembly using Geneious software

(Biomatter Ltd., San Francisco, CA), in order to extend coverage and resolve ambiguous base calls from the initial *de novo* assembly.

RT-PCR confirmation of assemblies

35 if the identified transcripts were suitable for PCR confirmation. Sets of left and right primers for each of these transcripts were designed in Primer 3 (Rozen and Skaletsky 2000), and used either together to amplify internal regions of the transcript, or with polyA or polyT primers to detect the 3' and 5' ends, respectively (Table 2-1). 50ng of template was used for RT-PCR, along with 1nM dNTPs and 10pmol of each primer. RT-PCR was either carried out in two-step fashion with reverse transcription carried out separately with the M-MLV Reverse Transcriptase kit (Promega, Madison, WI) according to the manufacturer's protocol and ExTaq (Clontech Inc., Mountain View, CA) or Fisher brand (Thermo Scientific, Sand Diego, CA) taq polymerase and supplied buffer for PCR amplification, or with a single-step method using the qScript One-Step RT-PCR kit (Quanta Biosciences, Gaithersburg, MD). In either case, the cycling protocol used was: denaturation at 95°C for 10 minutes for 1 cycle, then 95°C for 30s, 55°C for 30s, and 2 mins at 70°C for 35 cycles, and 1 cycle at 70°C for 10 mins at end of run, with single-step reactions having a 10 minute reverse transcription incubation at 45°C before denaturation.

Results

Transcriptome sequencing

Whole transcriptome Illumina sequencing of *K. brevis* yielded 151,417,524 paired end reads, each of which were 100 bases in length. The average insert length of the library sequenced was 260 nucleotides. Depth of coverage averaged 17,883, with *psbA* being by far the most represented at 558,035. Illumina read data was processed by *de novo* assemblers to construct contiguous sequences (contigs) that could be analyzed as putative plastid transcript candidates. SOAPdenovo (Short Oligonucleotide Analysis Package), a program specially designed to assemble Illumina short reads, produced 211,213 contigs (Luo et al. 2012). Trinity (Broad Institute), which combines three independent software modules to reconstruct transcriptomes from RNA-seq data, yielded 244,993 contigs. Velvet (Zehrino and Birney 2008), designed to assemble very short read data, was also used, and provided 313,396 contigs (Table 2-2).

Table 2-2: *Contig numbers associated with assembly output and screening.*

	<u>SOAPdenovo</u>	<u>Trinity</u>	<u>Velvet</u>
Initial Contig Number	211,213	244,993	313,396
Contigs After chloroplast filter	8,506	13,920	12,748
Contigs After Nicotiana filter	7,953	12,587	12,223
Contigs After Bacterial Filter	6,835	10,432	9,186
BLASTx Identifications	431	480	461

Sequence filtering

Each of the three *de novo* assemblers produced far more contigs than the number of genes targeted, so a systematic approach to filtering contigs for quality was used to progressively eliminate contigs that represented sample contamination, and then select the best representative for each plastid gene transcript identified (Table 2-2).

Total mRNA samples from *Karenia brevis* were sequenced, and thus contained transcripts of nuclear and mitochondrial origin. In order to identify putative plastid sequences from the total cell transcriptome read, a standalone database was created by downloading all complete plastid genomes from GenBank (accessed April 2013). All contigs produced by *de novo* assembly were searched against the plastome database via local tBLASTx. This variant of BLAST translates nucleotide queries into all 6 reading frames, and compares them to a translated nucleotide database. Any contigs that matched a coding region of a known plastid genome ($E \leq 0.001$) was considered significant and saved.

A sample from *Nicotiana* was run on the same lane as the *Karenia brevis* sample used for this study. Unfortunately, complete data separation did not occur, so some tobacco reads contaminated the data pool for this study. In order to filter for, and remove, the reads from tobacco contamination, any presumed plastid contig that produced a significant ($E \leq 0.001$) match to *Nicotiana* in the top 16 tBLASTx results was removed from consideration.

Due to the xenic nature of the *Karenia brevis* culture, there was a significant amount of bacterial contamination in the contigs produced by each assembler used. In order to screen for and eliminate these sequences, each remaining contig was compared to GenBank's NR database with tBLASTx. Any contig that produced a very close ($E \leq 1 \times 10^{-10}$) match to a sequence containing the string "bacter" in the taxon label, was eliminated from the candidate pool.

The remaining sequences were sorted according to the names of the chloroplast genes they'd been matched with, and the assembler that produced them. The cumulative effect of the above screening process was to greatly reduce the number of transcripts in the candidate pool before manual screening, translation, and analysis.

Manual Screening of Transcripts

1,372 contigs remained after the above screens as putative chloroplast transcript candidates. They corresponded to 68 genes commonly associated with plastid genomes, some specifically found in the plastome of *K. veneficum*, a dinoflagellate closely related to *K. brevis*. However, many of the remaining contigs were sequences from tobacco or bacterial contamination. In order to select a best candidate for each gene, and remove any remaining contamination, each assembled contig was converted to a ".seq" file format in EditSeq (DNASTAR, Inc., Madison, WI), with all such files for each plastid gene aligned in SeqMan (DNASTAR, Inc., Madison, WI). The alignments produced were used to select contigs likely to have full coverage of the coding region of the gene in question. The selected contigs were then searched against GenBank with nucleotide BLAST and

tBLASTx to identify contaminating sequences and verify algal homology. Contigs that closely matched (E value of 1.0e-20 or lower) a related algal ortholog, and had the greatest degree of transcript coverage were considered the best candidate, and used as the putative transcript for that gene going forward. A list of all sequences can be found in Appendix A.

Most putative transcripts had the closest homology to *Phaeocystis antarctica*, *Phaeocystis globulosa*, and/or *Emiliania huxleyii*, all of which are haptophytes with sequenced plastomes. Although there is almost no plastid data for *Karenia brevis* on GenBank, the one sequence available, the 16S ribosomal region of the psbD polycistronic unit, was an exact match for one of the contigs in our *Karenia brevis* library.

Contigs corresponding to 43 individual genes associated with plastid genomes were identified by the above process, that were unlikely to be ambiguous to the nucleus or mitochondria of *Karenia brevis*, or due to contamination from tobacco or bacteria (Table 2-3). These *de novo* putative transcripts were then used in Geneious (Biomatters, Ltd., Auckland, New Zealand) to perform a template-based reassembly, which increased coverage and resolved ambiguous base calls in a number of candidate transcripts. Each transcript was considered confirmed if the RT-PCR amplicon was consistent with the size predicted from the assembled contig and the –RT control was negative (Figure 2-1).

Table 2-3: List of all confirmed plastid derived mRNAs identified from *Karenia brevis*.

Gene name	Number of Reads	Peak Depth of Coverage	Average Depth of Coverage	full length algal match	Non-canonical start codon	PCR Attempted	PCR confirmed	5' polyA	polyA confirmed	3' polyU	polyU confirmed	poly-distron	other genes	BLAST Lineage
cbhX	1,459	169	94.5	+				-						dino and/or haptophyte
dpc	3,450	271	107.8	+				-						dino and/or haptophyte
groEL	3,450	271	107.8	+				-						green algal lineage
petB	88,432	23,795	6,513	+	+	+	+	-					ND	dino and/or haptophyte
petD	81,884	21,937	6,102	+		+	+	+						dino and/or haptophyte
psaA	307,017	19,638	11,298	+	+	+	+	+	+					dino and/or haptophyte
psaB	652,908	224,765	10,245	+		+	+	-					psac	dino and/or haptophyte
psac	652,908	224,765	10,245	+		+	+	-					psab	dino and/or haptophyte
psaf	96,106	53,175	5,156	+				-						dino and/or haptophyte
psbA	13,279,467	3,709,871	558,035	+	+	+	+	-					ND	dino and/or haptophyte
psbB	559,538	44,730	17,758	+	+	+	+	+						dino and/or haptophyte
psbC	894,774	168,211	13,180.30	+		+	+	+					rpoA	dino and/or haptophyte
psbD	523,981	65,961	31,786	+		+	+	+					165	dino and/or haptophyte
psbE	20,030	7,559	1,334	+				-						dino and/or haptophyte
psbF	36,119	12,383	4352	+		+	+	+					possible	dino and/or haptophyte
psbH	96,276	45,814	1959	+	+	+	+	+					rps2 and rps4	dino and/or haptophyte
psbI	18,179	8,498	595	+		+	+	-						green algal lineage
psbN	23,701	8,779	1,894	+	+			-					ND	dino and/or haptophyte
psbT	12,585	11,595	1054	+		+	+	-					ND	dino and/or haptophyte
psbV	83,067	17,072	7,713	+		+	+	+						dino and/or haptophyte
tblL	1,014,081	78,309	35,720	+		+	+	+						dino and/or haptophyte
rbcs	244,957	70,155	14,542	+		+	+	+						dino and/or haptophyte
rpl2	663	107	39.3	+				+					rpl23	dino and/or haptophyte
rpl3	48,689	11,500	1,407	+				+						dino and/or haptophyte
rpl5	3,411	125	73.1	+				-						dino and/or haptophyte
rpl14	3,411	125	73.1	+				+						dino and/or haptophyte
rpl16	3,411	125	73.1	+				-						dino and/or haptophyte
rpl23	1,426	562	151	+				+						dino and/or haptophyte
rpl31	48,155	8198	4380	+				-						dino and/or haptophyte
rpl36	20	8	3.4	+				-						dino and/or haptophyte
rpoA	26,723	3,687	1,440	+		+	+	-						dino and/or haptophyte
rpoB	471,029	72,482	11,668	+		+	+	-						dino and/or haptophyte
rpoC1	2,964	109	51.7	+	+	+	+	-						apicomplexa
rpoC2	179,370	21,924	2922.2	+	+	+	+	+						dino and/or haptophyte
rps2	1,182	189	86.8	+		+	+	+						apicomplexa
rps4	1,182	189	86.8	+		+	+	+						dino and/or haptophyte
rps5	2,031	551	244.4	+		+	+	-						dino and/or haptophyte
rps8				+				-						dino and/or haptophyte
rps15	885	286	152.2	+		+	+	-						apicomplexa
rps18	107	20	9	+		+	+	-						apicomplexa
rps19	48,719	7,807	4150.3	+	+	+	+	+						dino and/or haptophyte
rps23	1,230	383	72.2	+		+	+	-						dino and/or haptophyte
secA	58,284	3,107	1679.2	+		+	+	-						dino and/or haptophyte
ycf4	13,018	2,253	423.7	+		+	+	-						dino and/or haptophyte
165													psbD	dino and/or haptophyte

Transcript Structure

Among the assembled, screened and confirmed genes, thirty-three appeared as single coding regions and were scored as mono-cistronic transcripts (Figure 2-2 and Table 2-3). The remaining ten coding regions appeared as members of poly-cistronic transcripts (Figure 2-3 and Table 2-3).

Transcripts were amplified by one or two step RT-PCR using internal primers (Internal positive control), an internal primer paired with a polyA primer (polyU tail), or an internal primer paired with a polyT primer (polyA leader). Reactions without reverse transcriptase (-RT) were conducted to confirm that the sequences were actually RNA (negative control)

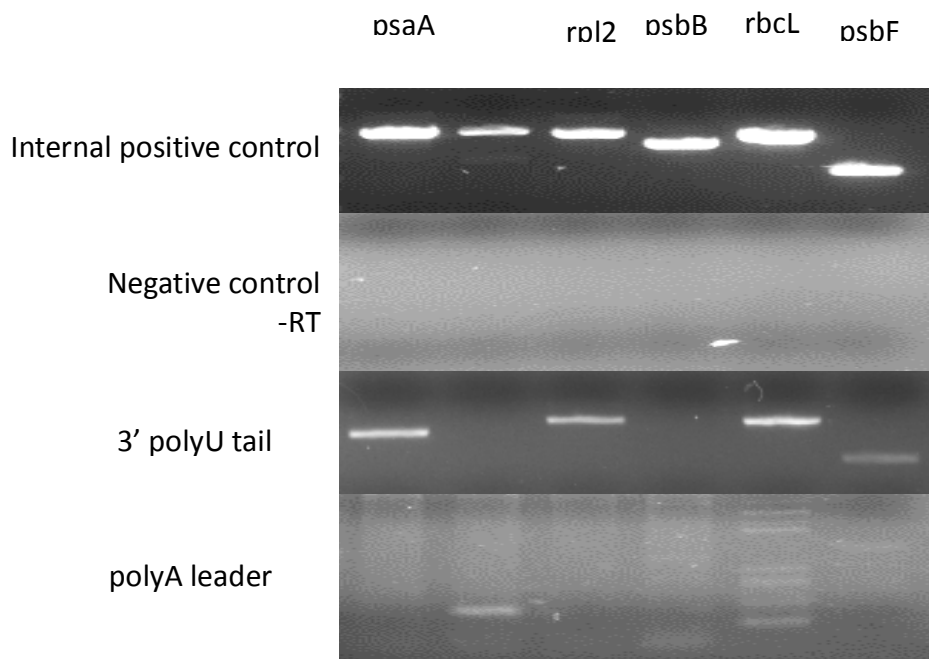


Figure 2-1: RT-PCR confirmation of six *Karenia brevis* plastid mRNAs.

petD

AA
AAAAAAAAAAAAAAAAAAAAAAAAAAGTTCCCCCAGGATATGAGTA
ACGTAAATCCCCTACCGAGCCAGGCTAAATCTCGGGCAAACCT
TTGTAAGGGGGCAGGTCATAATACATTCGGGGAACCGGCTTGG
CCTAACGATTTACTGTTCTTATTTCCCATCTGTATTACAGGTG
TACTTGCTTGTAGTTTAGGCTAGCAGTTCTATCTCCGCCGGC
AATCGACGCCCAATCAGACCCGTTTTCTACACCGCTGGAAATT
CTTCCGGAATGGTATTTCTTACCATCATTTAATTTGCTACGTC
TTCTTCCTAATAAACTTTTAGGTGTTGCTGCGATGGCCAGTAT
ACCTCTTGCATTAACGCAGGTAGCAGCTTCCGAGAACCAGACC
GCTTCTCAAATCCTAATCGACGCTCGCAAAGTTCACTGTTAT
ATATCCTGGGATCTTCGTCATCATCATGGCTTGGCGCAGGAGG
TGCTATGCCGCTTTCGCAGTCTTTAACTTTAGCACTTAATTAA
AAATT
TTT
TTTTTTTTTTTTTTTTTT

Figure 2-2: Example of a mono-cistronic *Karenia brevis* plastid transcript, *petD*, showing the 5' polyadenylation and 3' polyuridinilylation (underlined).

GTTCTTTGAAAATTAGAAGAGATTCGTTGGCGATTACGTTGCGGACTTTTCCTTCTTTCTCAAAGGACTAAG
 ACAAGATCCAACACTACGCGACGATTGTCATACTCTGTGGGTATTGCCCATGATTTGGAGTCACATGATTTTCGTTACG
 GAATCACTTTTGTATCAACGTATTTTGTCTCATTGTTGGGCACCTAGCAATAATATTTTATGGTCCGCTGGAAAT
 CTATTCCACGTAGCATGGCAAGGAAATTTCAAGAGTGGATTATAAATCCTCTGAAAACGTCACCAATTGCGCATG
 CGATATTCGATCCGCACTTTGGAGTAGACGCGCTTAAAGCCTTACGCCATCTGCCGGCTTATCCCGTTAATATA
 TCGACCTCAGGGTGTATCATTGGTGGTATACTATCGGCTTACGATCAAACACTGACTTGTATTCCGCTTCAATGTT
 CCTTATTCTATGCAGTCTTCTGTTTTATATGCAGGGTGGCTTCATTCTACACCTGGTCGGCAACCGAACCTAGCGT
 TTTTAAAAACAATGAATCAAGATTAATCATCATCTGTCGGGTCTTTTGGGGTAAAGTTCGCTGGCCTGGGCAGG
 GCATCTAATTCATGTTGCAATACCTGCGTCTCGTGGCAAGTGTAGTCTGGATCATCTCCGTTACCCCATCCCG
 CTGGGTAAACCCATTCTAACGTTAATTGGGGTATTTACGCCGAAAATCCTGATTCTCCAACCTCATGTGTTGCG
 ACTACAGAAGGTGCAGGTACGGCAATCCTCACTTTCCTAGGAGGATTTATCCTCAAACCCAAGCTATGTGGCTGA
 CTGATATTGCCATCATCATTTAGCAATCGCAATCATTTTATTGTTGCAGGGCATATGTATCGAACCACTATGCT
 TGGGGCCATCGCTTACCAGTGTGTCACGCGCACGTTCTCTAAAGGACGCTTCGCGCTGGTCATGTTGGCT
 TAATGGAAACGTTGACCAATAGTTTACACATGCAACTGGTTTACGCTGGCAGCGCTAGGCGTTGCAACTTCGCT
 AACGGCCCAGCATATGTATGCGTTACCTGCCTACGCATTTATCGCGAGCAGCCCCGTAACCAAGCAGCATTGTAC
 GTACACCATCAGTATATCGCGGGTTTCTAATGACGGGGCCCTTCGCCATGGTGCCATTTTTTTCGTCCTGATT
 ACGAGCCTTGAATGAAGGAAATGACTTGCACAGTACTTCTTCAAAAGAGGCTATAATTTCTCACTTGAGTTG
 GGTGAGTTATTTTTGGGTTTCCATACGTTGGGTCTCTATGTGCACAATGATACTTGCCTTGCATTGGCCTTCCCG
 AAAATCAAATATTATTGGATCCAATTTGCTCAACTCGTTCAAGCCGCATCAGGTAAGCAGGAAATGGAATTGA
 TATTCTTTCGGAACCTACAAGTCCCCTAGTATTGCAAGTAGTAAAGTTGGTTGCCTGGGTGGCTCGCGGGA
 ATTAACAGCCAGCGAAGTACCGCCTTCTACCAATTGGACCGGGTGAATTCCTTATTATCACGCAATTTCACTGG
 GATTGCATAACTACGCTAATCTTAGTAAAGGGTGTCTTGTATGCTCGAGCCTCAAGATTACTTCCAGATAAGAA
 AGATTTTGGTTATAGTTCCCGTGTGATGGTCTGGACGTGGTGGTACTTGTGACGTTTACGCTTGGGACGCCTT
 TATTGGCTGTCTTGGCAATTAATACTACGGCATGGACCACATTCTATTGGCACTGGAAAACATTTGACATTATG
 GGGTGGCAATAGCGCCCTTTCAATGAAGCTTCAACATACATTATGGGTTGGTTACGCGATTATTTATGGTTAAAC
 TCAGCTCCAACGATCACCGTTATACGGCACAAGGAATGAACGCACAGGCTGTTTGGACATGGATGTTCTGTTT
 GGACACCTTACTTGGGCTACCGTTTTATGTTCTTAATTTTATGCGAGGATATTGGCAAGAATTTATTGAAACCA
 TAGTTTGGGCACATGAAAGAACCCCACTAGCTAATTTGCTAAAATGGAATGACAAGCCTGTGGCTCTGTCAATTGT
 CCAAGCACGTCTCGTAGGGTTGACACATTTACGACAGGGTATATTTACTTATGCCCTTTTTAATTTCTAGTA
 CGTCCGCCAAAACAGCTGGAATATCTACTGCCGCCCTTATTTAAAGATCTACCTGGCGAAATTTATCTAGAACGT
 ACAAAAAAATGATTACTTATGATTTTCGCTTCAAGAAAGTTAATTGTTTTACAAATAGAAAAAAGTTAGTCCTTAC
 GCCTCAGCATATCAAGAAAAAATTTTTAATAAGAATTAGTCTGGTGAAGATACGAGGACAATCTACATTTTTAAT
 CTTTACTGAACCTGAACATAAATTTGGACTCTAATTTACGGAAGCTACTCTCACAAATCGTTGAACTTAAACA
 TATTTTACCCTACAGCTTGCAAAATCTTCTGGTGAACACTTTTTCTTCTCAATTTCTAATTTTAAAGGAGTCTAAG
 TCCAACTTTTTCACTAAACACTAATACGTTAGCCCCATCGATCGACTCGTTTCAGGTCCCAAGTATACACGTATG
 CTTAAGTAATCCGTGGGGCATGCCGATTACAACGCTTACAACCAATGCAGTCGGCAGGCCGCGGGGCAGCAGC
 AACTTGTCTGACTTATTCCATACGGTGTGGGTACCATTGCAAGGACATCAAAAGGACACGCACGGACACTG
 AGTACAGCCGATACATGTAGTATAAATCCTTAGGGTATGAGACATAAATTTAGATTGTACCGAACCGTTTTTTTTT
TTTTTTTTTTTT

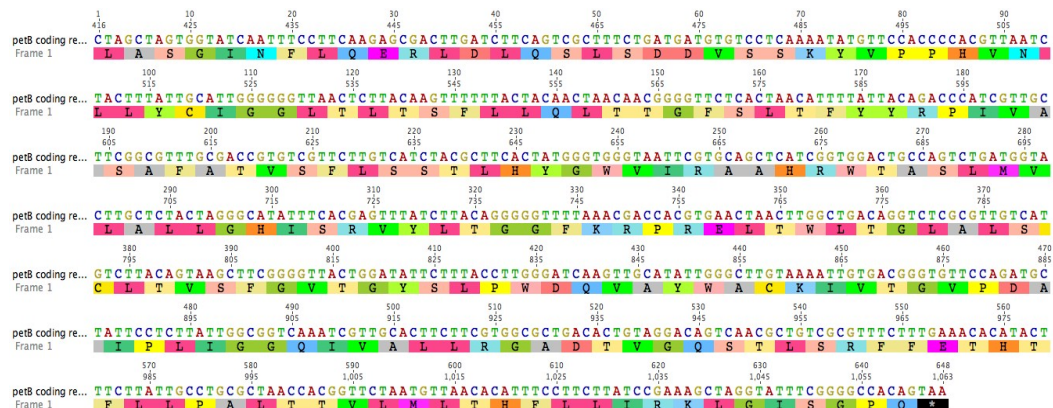
Figure 2-3: Example of a poly-cistronic *Karenia brevis* plastid transcript, *psaB-psaC*. *psaB* is colored red and *psaC* is green. 3' polyuridylation is underlined.

Since peridinin containing dinoflagellate plastomes are organized into mini-circles that undergo rolling circle transcription, evidence of mini-circles was sought as repeated sequences on the 3' and 5' ends of each assembled transcript. No repeating sequences were detected on the ends of the transcripts. Instead, twenty two of the assembled transcripts had a 5' polyA leader sequence and/or a 3' polyU tail sequence (Figures 2-1, 2-2, 2-3, and Table 2-3). RT-PCR was used to confirm the presence of these sequences (Figure 2-1). 3' PolyU tails were RT-PCR amplified from fifteen transcripts by pairing a polyA primer to a transcript specific primer. PolyA leader sequences were attempted by pairing a polyT primer with a transcript specific primer but the results were ambiguous except for two transcripts, *rps4* and *psbF* (Figure 2-1). Neither polyU or polyA sequences were detected from –RT control reactions, or PCR attempts from DNA templates.

Protein Translations

All forty three transcripts were translatable into full length protein sequences (Appendix B). Standard chloroplast and bacterial genetic codes identified protein sequences with clear homologous hits as determined by protein BLAST searches against GenBank's nr database. Ten of the transcripts did not have a canonical AUG-methionine start codon (Figure 2-4 and Table 2-3).

A



B

cytochrome b6 [Emiliania huxleyi]

Sequence ID: [ref|YP_277309.1](#) Length: 215 Number of Matches: 1[► See 4 more title\(s\)](#)Range 1: 1 to 214 [GenPept](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
318 bits(814)	1e-106	Compositional matrix adjust.	137/214(64%)	185/214(86%)	0/214(0%)
<i>Kbrevis</i>	Query	1	LASGINFLQERLDLQSLSDVSSKYVPPHVNLLYCIGGLTLSFLLQLTGFSLTFYFRP		60
<i>Ehuxleyi</i>	Sbjct	1	++S ++ QERL++Q+++DD++SKIVPPHVN+ YC+GG+TLT F++Q+ TGF++TFYFRP		60
	Query	61	IVASAFATVSFLSSTLHYGVVIRAAHRWTASLMVLALLGHISRVLTTGGFKRPRELWLT		120
	Sbjct	61	TVTEAFASVEYLMTEVNFVGLIRSVHRWSASMMVLMMLHVCRVYLTGGFKRPRELWVT		120
	Query	121	GLALSCLTVSFGVTGYSLPWDQVYWACKIVTGVDPDAIPLIGGQIVALLRGADTVGQSTL		180
	Sbjct	121	GVLASVTVSFGVTGYSLPWDQVYWACKIVTGVDPDAVPIVGGIVQVLRGGVSVGQSTL		180
	Query	181	SRFFETHFLPALTTVLMMLTHFLLRKLGISGP 214		648
	Sbjct	181	+RF+ HTF+LP + VLMLTHF++IRK GISGP 214		1,063

Figure 2-4: Non-canonical start codons. A – translation of *petB*, a *K. brevis* transcript with a non-canonical start codon, in this case a CTA coding for leucine. B – a comparison of the *Karenia brevis petD* translation with *petD* from the haptophyte *Emiliania huxleyi* showing the presumed *K. brevis* start codon as a leucine.

Table 2-4: *Non-canonical start codons found in Karenia brevis plastid transcripts.*

Transcript	Nan-Canonical Start	Amino Acid
petB	CUA	Leu
psaA	CGC	Arg
psaB	UUG	Leu
psbB	AAU	Val
psbF	CUA	Leu
psbH	UUG	Leu
psbN	UUG	Leu
rpoC2	CUU	Leu
rps19	GUG	Val

Translations were screened for signal and transit peptides to determine if any transcripts were mis-identified. Each peptide was screened with SignalP (Peterson et al 2011) and ChloroP (Emanuelsson et al 1999) along with *petF*, a known *K. brevis* nuclear-encoded, plastid-localized protein (Nosenko et al 2006). *petF* had clearly defined transit peptides even with the most stringent settings (Table 2-5). Several of the new *K. brevis* translations were identified as having possible signal and/or transit peptides but only *groEL* passed the strictest SignalP settings with a strong mitochondrial transit peptide. *groEL* is a general chaperonin found in many bacteria and necessary in eukaryotic mitochondria and plastids (Horwich et al 2007) so this sequence is likely a nuclear encoded, mitochondria-targeted protein. The translations *psaC*, *psaF*, *psbH*, and *psbV* had signal peptides of low confidence. Since these genes are found in the plastid genomes of red algae, at least one haptophyte, the alveolate *Chromera velia* and the fucoxanthin dinoflagellate *K. veneficum* (Gabrielsen et al 2011), I chose to dismiss the possibility that these are nuclear encoded.

Table 2-5: Signal and transit peptide prediction using the programs SignalP and ChloroP.

Protein	SignalP Strict	SignalP Sensitive	ChloroP
<i>petF*</i>	Y	Y	Y
<i>petB</i>	N	N	N
<i>petD</i>	N	N	N
<i>psaA</i>	N	N	N
<i>psaB</i>	N	N	N
<i>psaC</i>	N	Y	N
<i>psaF</i>	N	Y	Y
<i>psbA</i>	N	N	N
<i>psbB</i>	N	N	N
<i>psbC</i>	N	N	N
<i>psbD</i>	N	N	N
<i>psbE</i>	N	N	N
<i>psbF</i>	N	N	N
<i>psbH</i>	N	N	Y
<i>psbI</i>	N	N	N
<i>psbN</i>	N	N	N
<i>psbT</i>	N	N	N
<i>psbV</i>	N	Y	Y
<i>psbZ</i>	N	N	N
<i>rbcl</i>	N	N	N
<i>rbcS</i>	N	N	N
<i>rpl3</i>	N	N	N
<i>rpl6</i>	N	N	N
<i>rpl19</i>	N	N	N
<i>rpl23</i>	N	N	N
<i>rpl27</i>	N	N	N
<i>rpl31</i>	N	N	N
<i>rpl36</i>	N	N	N
<i>rpoA</i>	N	N	N
<i>rpoB</i>	N	N	N
<i>rpoC1</i>	N	N	N
<i>rpoC2</i>	N	N	N
<i>rps2</i>	N	N	N
<i>rps3</i>	N	N	N
<i>rps5</i>	N	N	N
<i>rps12</i>	N	N	N
<i>rps14</i>	N	N	N
<i>rps15a</i>	N	N	N
<i>rps18</i>	N	N	N
<i>rps19</i>	N	N	N
<i>rps23</i>	N	N	N
<i>ycf4</i>	N	N	N
<i>cbbX</i>	N	N	N
<i>clpC</i>	N	N	N
<i>secA</i>	N	N	N
<i>secY</i>	N	N	N
<i>groEL</i>	Y	Y	N

* positive control

Discussion

In this study, forty-three plastid transcripts were identified from a *K. brevis* transcriptome produced by Illumina sequencing. Assembled transcripts were identifiable by their homology to other algal sequences.

Transcript structure and Processing

Analysis of my data revealed no putative transcripts of rolling-circle structure, or any other indication of the use of minicircles in the plastid genome of *K. brevis*. However, mini-circles cannot be ruled out with my analysis since the methodology was skewed towards the assembly and analysis of fully processed transcripts. To date, most studies of dinoflagellate plastid genomes, transcription, and transcript processing have focused on peridinin-containing dinoflagellates. Peridinin-containing dinoflagellates have mini-circular plastome structures (Barbrook and Howe 2000; Howe et al. 2008; Zhang et al. 1999) that are transcribed by a rolling circle mechanism (Dang and Green 2010; Nisbet et al 2008). Transcripts are processed by RNA editing (Zauner et al 2004; Dang and Green 2009), the addition of a 3' polyU tail (Nelson et al 2007; Wang and Morse 2006), and an unknown 5' processing event. *K. brevis* is a fucoxanthin containing dinoflagellate, a group which has chloroplast gene sequences resembling haptophytes (Yoon et al 2005; Nosenko et al 2006). They appear to have been peridinin-containing secondary endosymbionts that later replaced the red-algal-like plastid with a haptophyte chloroplast in a tertiary endosymbiosis (Janouskavic 2010; Gabrielson 2011). The plastid genome of *K. brevis* has not been published but another fucoxanthin

dinoflagellate, *K. veneficum*, has a large circular chromosome (Gabrielson et al 2011) but also appears to have smaller circular fragments (Espelund et al 2012). The haptophyte, *Emiliana huxleyi*, also has a large circular plastome (Sanchez Puerta et al 2005). These observations suggest the possibility that other fucoxanthin dinoflagellates could have plastid genomes with these structural elements.

K. brevis transcripts are both mono- and poly-cistronic. This observation fits with observed plastid biology where mono- and poly-cistronic transcripts are common among plastids (Stern et al 2010; Zhelyazkova et al 2012). They have also been observed in the peridinin-containing dinoflagellate *Amphidinium carterae*, which has mini-circular plastome structure (Barbrook et al 2012)

K. brevis plastid transcripts are processed by RNA editing and the addition of a polyA 5' leader sequences and polyU 3' tail to each sequence. This feature appeared on mono- and poly-cistronic RNAs. Two recent studies have demonstrated that the fucoxanthin-containing dinoflagellates, *Karenia mikimotoi* and *K. veneficum* have maintained two RNA processing mechanisms found in the peridinin dinoflagellates but not haptophytes or diatoms, RNA editing and 3' polyuridylation (Dorrel and Howe 2012; Richardson et al 2014). In *A. carterae*, 3' polyuridylation has been shown to define the processed ends of both mono- and poly-cistronic transcripts cleaved from "multi-copy transcripts" produced from rolling circle transcription (Barbrook et al 2012). Our data is consistent with other RNA processing observations among dinoflagellates.

3' polyuridylation has been found associated with transcripts encoding proteins used in the photosynthetic electron transport chain in two chromerid algae, *Chromera velia* and *Vitrella brassicaformis* (Dorrell et al. 2014). The authors did find two exceptions suggesting this phenomenon is primarily but not exclusively associated with photosystem genes. It has also been reported in *K. mikimotoi* (Dorrell and Howe 2012) but there was no correlation to protein function reported. According to my data, eleven of the twenty-three polyU-containing transcripts encode proteins used in the photosynthetic electron transport chain suggesting there is no correlation between this processing event and protein function. It has also been suggested that the presence of the polyU tail may correlate to transcript abundance (Barbrook et al 2012; Dorrell et al 2014) but in the *K. brevis* data there is no correlation between the presence of a polyU tail and the average depth of coverage detected for any transcript.

In silico assemblies from this study suggest eighteen transcripts had a polyadenylated 5' end, fifteen of these also had a polyU 3' end. To the best of my knowledge there are no reports of 5' polyadenylation of dinoflagellate transcripts. Barbrook et al. (2012) sequenced circularized transcripts from the peridinin-containing *A. carterae*, which would provide both 3' and 5' ends. They report polyU 3' tails and 5' ends that varied greatly in the distance of the upstream processing site from the start codon, but did not report polyadenylation associated with the 5' end.

Translations

Translation of these transcripts was possible using standard plastid and bacterial genetic codes (code table 11) with the exception of nine transcripts with non-canonical start codons. Several coding regions in land plant plastomes have non-canonical start codons (Yang et al 2013; Yang et al. 2010; Cahoon et al. 2010; Raubeson et al. 2007; Gao et al. 2009) and they have also been reported in coding regions for several dinoflagellates. A single unusual start codon was found in the *K. veneficum* plastome as well as numerous premature stop codons and “highly divergent” coding sequences (Gabrielsen et al 2011). GUA has been shown to act as an alternative start codon in the dinoflagellate *Amphidinium operculatum* (Barbrook et al. 2001) and AUU and AUA are start codons in *Symbiodinium* sp. (Barbrook et al. 2014). Other algae include the haptophyte *E. huxleyi* could be translated using the standard plastid/bacterial genetic code but also uses GUG as an alternative start codon (Sanchez-Puerta et al. 2005). The diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, both secondary endosymbionts with red algal plastids use GUG and AUU as start codons (Oudot-Le Secq et al. 2007) but are otherwise translated with the standard plastid & bacterial genetic code.

RNA editing has been cited as a mechanism, which could change non-canonical start codons, premature stop codons, or places stop codons (Gabrielsen et al 2011; Cahoon et al 2010). Chloroplast mRNA editing is common among land plants and dinoflagellates (Dorrell and Howe 2012; Zauner et al 2004; Dang and Green 2009) but

not among red or green algae and their secondary endosymbionts. The transcripts assembled for *K. brevis* in this study appear to have been fully edited and processed due to their lack of premature stop codons and adherence to the standard plastid/bacterial genetic code to produce full-length recognizable proteins. Four of the nine non-canonical stops fit the 'NUG' start codon motif in the bacterial codon usage table, which leaves CUA, CGC, AAU, CUA, and CUU as putative non-canonical start codons.

CHAPTER III

PHYLOGENETIC ANALYSIS OF 6 PLASTID GENES FROM THE

PERIDININ DINOFLAGELLATES AND *KARENIA BREVIS*.

Chapter Introduction

The purpose of the work described in this chapter is to use the plastid transcriptome sequences described in Chapter 2 to perform a robust phylogenetic analysis of the *K. brevis*. The publicly available plastid transcript and protein data for *K. brevis* and other fucoxanthin dinoflagellates has been limited until recently. Most phylogenetic studies have focused on a limited number of gene products, in many cases incomplete ESTs and have noted considerable difficulty in using phylogenetic methods to place dinoflagellate plastid lineages and relationships (Yoon et al 2005, Shalchian-Tabrizi et al 2006).

For this phylogenetic inquiry, the protein products from the families *pet*, *psa*, *psb*, *rbc* and *rpo* were chosen for comparison to 21 representative taxa from the major eukaryotic algal groups and higher plants. Phylogenetic trees were built for each gene, as well as for the concatenated alignments. This analysis will be used to answer the following questions:

- Does phylogenetic analysis of the protein products of plastid genes obtained by this study agree with current evidence of the haptophyte origin of the fucoxanthin dinoflagellate plastid?
- Do the new *K. brevis* protein sequences presented in Chapter 2 allow a more rigorous phylogenetic analysis of the peridinin and fucoxanthin dinoflagellate plastid proteomes?

Methods

Protein sequence alignments

An alignment file was assembled for the putative amino acid sequences for each plastid gene of interest, as well as the corresponding sequences from a group of 26 taxa spanning the major groups of eukaryotic algae, and including higher plants retrieved from GenBank or UniProt (Table 3-1, Appendix Table 3-X). Protein sequence data was chosen due to greater availability of the necessary complete protein sequences of the related taxa on GenBank, and to mitigate the effects of atypical codon usage in dinoflagellates, as seen by others (Shalchian-Tabrizi et al. 2006). Sequence alignments were performed in Geneious (Biomatters, Ltd., Auckland, New Zealand) using MUSCLE (EMBL-EBI) with the default settings, and converted to FASTA and interleaved Phylip formats for phylogenetic analysis.

The concatenated sequences were assembled from the protein sequences produced in this study and those of all of the related taxa within the library. Concatemers were assembled in Geneious, and ordered in alphabetical fashion from left to right. Amino acid sequences from *Synechococcus sp.* WH 7803 was included in each uniprotein and concatenated alignment, and used below as the outgroup for each phylogenetic tree constructed.

Construction of phylogenetic trees

The “Find best DNA/Protein Model” feature in MEGA6 was used to determine the substitution model and parameters used to construct the phylogenetic trees. The model with the lowest Bayesian Information Criterion score was chosen as the most representative. The Le-Gascuel substitution model (LG) with gamma distribution (Γ) scored the lowest for most attempts (15 of 19), and was either the second or third choice in the remaining 3 instances. The output from these analyses is contained in Appendix C.

Phylogenetic trees were built using the genomic computing resources provided by the CIPRES portal (citation). Tree structure was determined using the Maximum Likelihood method in RAxML, based on the LG substitution model, with gamma distribution and allowance for invariant sites (LG + Γ + I), with the tree that had the highest log-likelihood being selected. Gamma distribution was used to model evolutionary rate differences among sites, with 4 rate site categories selected. All trees were bootstrapped with 1000 replications. Best-fit trees were edited in MEGA6 for formatting fit and addition of taxa information. The bootstrap values for each branch were added at the tree nodes.

Upon completion of a robust library of complete protein sequence data from over thirty taxa, including the pet, psa, psb, rbc and rpo gene families, the sequences were organized into files by gene, aligned with MUSCLE, and tested for the best substitution model and ML parameters in MEGA6 (citation).

Results

Sequence and substitution model selection

The taxa chosen for this study were selected to represent a broad range of eukaryotic photosynthetic organisms, based the availability of plastid gene/transcript/protein sequence data in the GenBank non-redundant protein database. Preliminary single protein phylogenetic analysis of alignments from the *pet*, *psa*, *psb*, *rbc* and *rpo* gene families, revealed that *petD*, *rbcL* and a number of the *psa* and *psb* family proteins produced empirically sound branching (Appendix D).

Each single protein alignment for the proteins selected above, was tested for the best-fit substitution model, using the model test function in MEGA6. The output from this analysis is collected in Appendix C. In the majority of cases, the Le-Gascuel substitution model with gamma distribution was shown to fit the data best, and allowing for invariant sites produced a non-significant change in BIC scores in most remaining cases. In light of this analysis, the Le-Gascuel substitution model with gamma distribution and allowance of invariant sites (LG + Γ + I) was chosen for construction of maximum likelihood (ML) phylogenetic trees used in this study. Invariant site consideration was included for empirical reasons, as all proteins contained some number of invariant sites across the breadth of taxa analyzed. The protein sequences of *petD*, *psbA*, *psbB*, *psbC*, *psbD*, and *psbE* were chosen to conduct a more thorough phylogenetic analysis, based on the amount sequence data for the taxa chosen.

Table 3-1. Taxa used in phylogenetic analysis

Group	Species	Order	TaxID
Embryophyta (higher plants)	<i>Arabidopsis thaliana</i>	Brassicales	3702
	<i>Pinus pinea</i>	Pinales	3346
	<i>Physcomitrella patens</i>	Funariales	3218
	<i>Zea mays</i>	Poales	4577
Chlorophyta (green algae)	<i>Chlamydomonas reinhardtii</i>	Volvocales	3055
	<i>Dunaliella salina</i>	Volvocales	3046
	<i>Leptosira terrestris</i>	Chaetophorales	34116
	<i>Pedinomonas minor</i>	Pedinomonadales	3159
Rhodophyta (red algae)	<i>Galdiera sulphuraria</i>	Cyanidiales	130081
	<i>Grateloupia taiwanensis</i>	Halymeniales	1260292
	<i>Porphyra umbicalis</i>	Bangiales	2786
	<i>Porphyridium purpureum</i>	Porphyridiales	35688
Phaeophyta (brown algae)	<i>Ectocarpus siliculosus</i>	Ectocarpales	2880
	<i>Fucus vesiculosus</i>	Fucales	49266
	<i>Saccharina japonica</i>	Laminariales	88149
Haptophyta	<i>Emiliania huxleyii</i>	Isochysidales	2903
	<i>Pavlova lutheri</i>	Pavlovales	2832
	<i>Phaeocystis globulosa</i>	Phaeocystales	33658
Bacillariophyta (diatoms)	<i>Asterionella formosa</i>	Pennales	210441
	<i>Lithodesmium undulatum</i>	Lithodesmiales	59812
	<i>Odontella sinensis</i>	Triceratiales	2839
	<i>Phaeodactylum tricorutum</i>	Naviculales	2850
	<i>Thalassiosira pseudonana</i>	Thalassiosirales	35128
Dinophyta (dinotom)	<i>Durinskia baltica</i>	Dinotrichales	400756
	<i>Kryptoperidinium foliaceum</i>	Dinotrichales	160619
Dinophyta (peridinin pigment)	<i>Amphidinium carterae</i>	Gymnodiales	2961
	<i>Heterocapsa triquetra</i>	Peridinales	66468
	<i>Lingulodinium</i>	Gonyaulacales	107758
Dinophyta (fucoxanthin pigment)	<i>Karenia brevis</i>	Brachidiniales	156230

Unstable topology of dinoflagellate taxa

A six protein concatemer was assembled from the single protein MUSCLE alignments of petD, psbA, psbB, psbC, psbD and psbE with Geneious, and used to construct a maximum likelihood tree in RAxML using the LG + Γ + I parameters allowing for 4 rate sites, which is shown in Figure 3-1. In this tree, all dinoflagellate taxa from the fucoxanthin and peridinin lines are grouped monophyletically with the haptophyceae, although with weak bootstrap support, which is congruent with the findings of Yoon et al. 2002. Non-canonical topologies such as this have been seen before, with psaA showing significant disruption to expected branching patterns when added to alignment concatenations (Shalchian-Tabrizi et al. 2006). Examination of the single protein alignments revealed patterns of consensus loss present in both peridinin and fucoxanthin dinoflagellate species, which are discussed in greater detail below.

The alignment of psbE contained the highest ratio of lost invariant sites to total sequence (figure 3-8), and at only 92 residues in length, could be eliminated from the concatemer without significant reduction in the total information thereof, and the corresponding drop in overall support for the tree. When reanalyzed without psbE sequence data the tree topology shifted, with the peridinin taxa branching from the heterokonts with weak support (Figure 3-2). Support for *K. brevis*'s monophyly with the haptophyceae is improved slightly by the removal of psbE, but remains low. In all cases, *Kryptoperidinium* sequences are congruent with the broader group of taxa, and show no sign of invariant site loss.

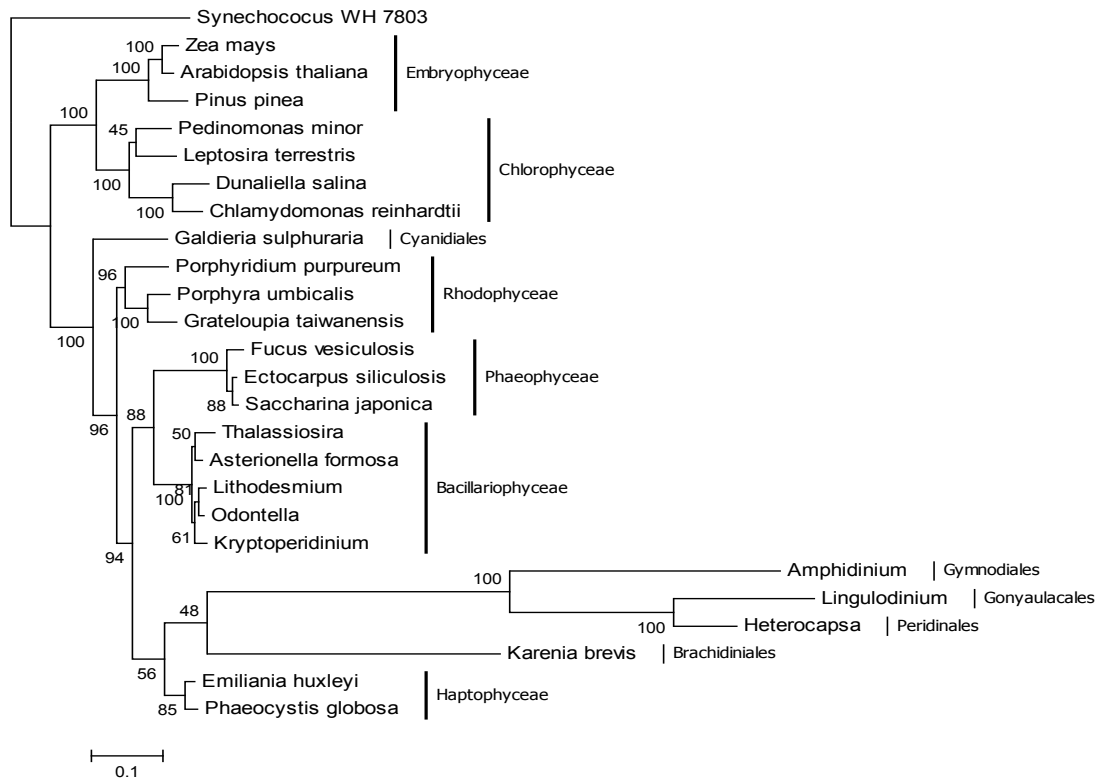


Figure 3-1. Maximum likelihood phylogenetic tree, based on a 6 protein alignment containing *petD*, *psbA*, *psbB*, *psbC*, *psbD* and *psbE* concatenated in that order. The tree shown was constructed in RAxML, using the Le-Gascuel substitution model, Γ distribution of site rates and allowance for invariant sites (LG + Γ + I). Bootstrap support values (1000 replications) for each node are given. All dinoflagellate taxa examined are grouped monophyletically with the haptophyceae and distinct from the heterokonts, albeit with weak support.

Since the loss of invariant sites seen in the alignments demonstrated evidence of a pattern preserved across the peridinin and fucoxanthin dinoflagellate taxa, two concatemers were reassembled containing *petD*, *psbA*, *psbB*, *psbC*, *psbD* and *psbE* proteins; one with the *K. brevis* sequence removed, and the other lacking *Amphidinium carterae*, *Heterocapsa triquetra* and *Lingulodinium polyedrum*. Maximum

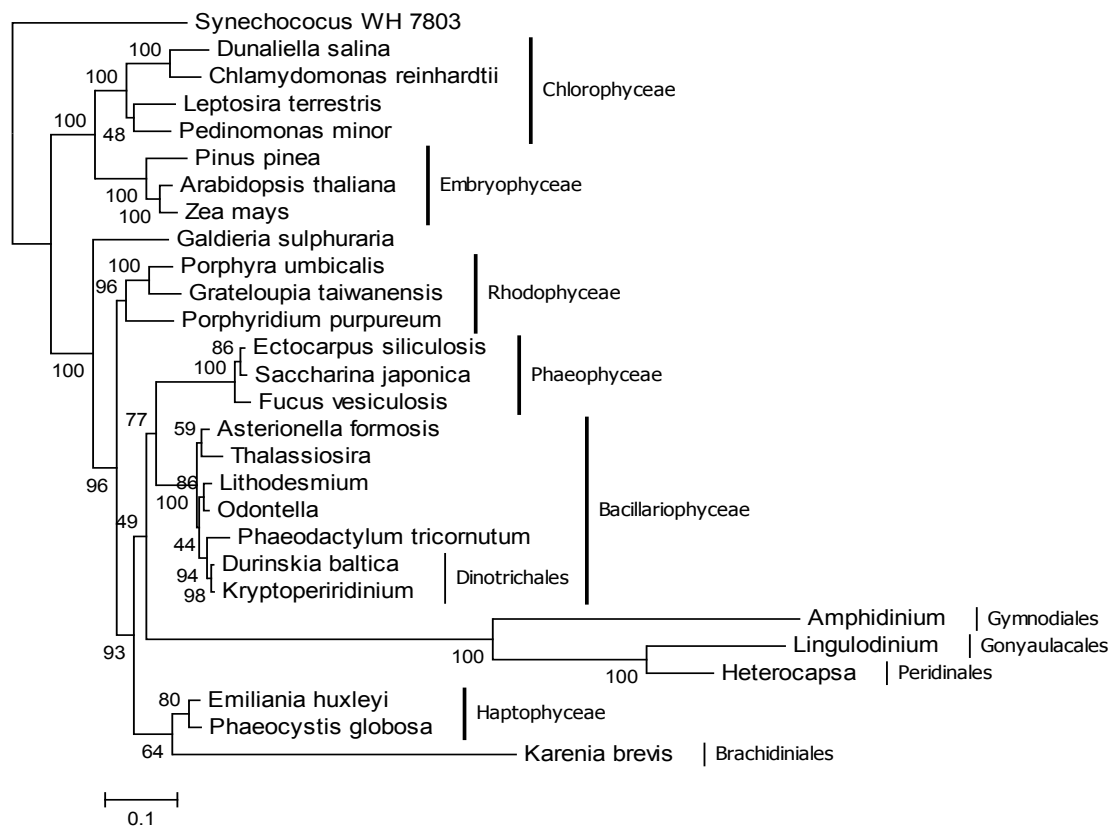


Figure 3-2. Maximum likelihood phylogenetic tree, based on a 5 protein alignment containing petD, psbA, psbB, psbC, and psbD concatenated in that order. The tree shown was constructed in RAxML, using the Le-Gascuel substitution model, and Γ distribution of site rates and allowance for invariant sites (LG + Γ + I). Bootstrap support values (1000 replications) for each node are given. *K. brevis* has moderate support for inclusion within the haptophyceae, while the dinoflagellate plastid proteins form a weakly supported monophyletic group with the heterokontophyceae.

likelihood trees for these concatemers are found in Figure 3-3. With the peridinin dinoflagellate taxa removed from the alignment, *K. brevis* remains grouped monophyletically with the haptophyceae, but the bootstrap support is increased to 100% at that node. With *K. brevis* removed, the peridinin dinoflagellates form a weakly supported monophyletic branch with the heterokontophyceae, and are distinct from the

haptophyceae by one level with weak support. In all cases, the proteins from the peridinin plastid taxa had significantly more loss of invariant sites than *K. brevis*.

Patterns in dinoflagellate invariant site loss

As mentioned above, when compared to the broader group of taxa, both peridinin and fucoxanthin dinoflagellate plastid proteins lack invariant sites that are preserved across all other photosynthetic eukaryotes. In a number of the proteins used in this study, these losses occur in somewhat localized regions that are shared across both dinoflagellate subgroups. Figure 3-4 shows the MUSCLE protein sequence alignment of the psbA protein sequences from the 26 taxa used in the phylogenetic analysis above. A sequence of the first 9 residues of the N-terminus, which are well conserved in the reference taxa, show substantial variation from reference consensus in both the peridinin and fucoxanthin dinoflagellates. Further into the sequence, at alignment position 233, all taxa of dinoflagellates have a glycine residue for the glutamic acid conserved across all reference taxa. The protein sequence of psbD, seen in figure 3-5, shows the same pattern as psbA, with a loss of a strongly conserved N-terminal motif, and position of uniform substitution preference. In this case, threonine is substituted for a position containing a consensus asparagine residue in all reference taxa, except the rhodophyceae.

The peridinin and fucoxanthin dinoflagellate petD protein has also lost a substantial amount of invariant sites when compared to the reference taxa. Whereas, in the previous proteins discussed, the loss of invariant sites is not prevalent in functional

regions of the protein, the petD protein in the peridinin and fucoxanthin dinoflagellates has been affected in 4 transmembrane regions, as shown in figure 3-6. The transmembrane regions affected show a number of areas of invariant site loss shared between the peridinin and fucoxanthin dinoflagellates.

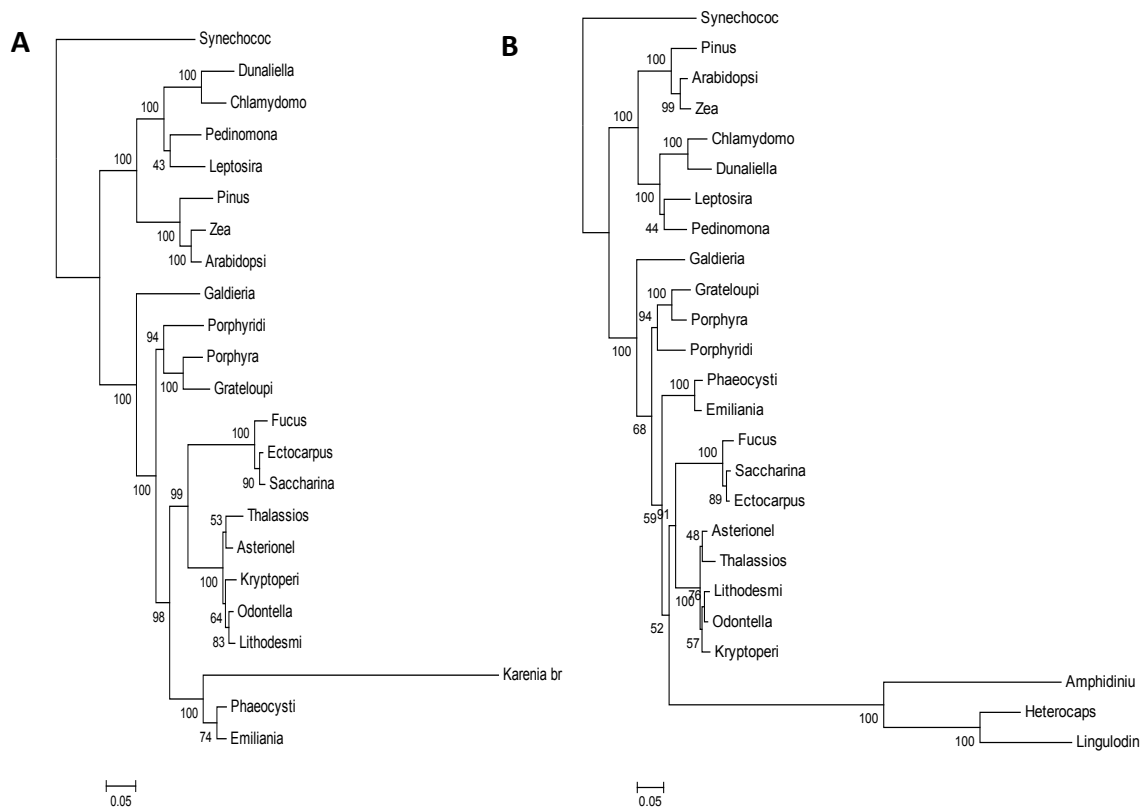


Figure 3-3. Maximum likelihood phylogenetic trees built from the 6 protein concatemer with each dinoflagellate group added individually. Trees were built with RAxML using LG + Γ + I parameters, and the node values given are bootstrap support (x1000). Panel **A** shows the tree constructed with the peridinin dinoflagellate taxa removed from the alignment. *K. brevis* is grouped monophyletically with the haptophyceae, showing strong (>98%) support at all relevant nodes. Removal of the *K. brevis* sequence data (**B**), yields a best fit tree that groups the peridinin dinoflagellates within heterokontophyceae, albeit with weak support (BS = 68, 59 and 52 respectively).

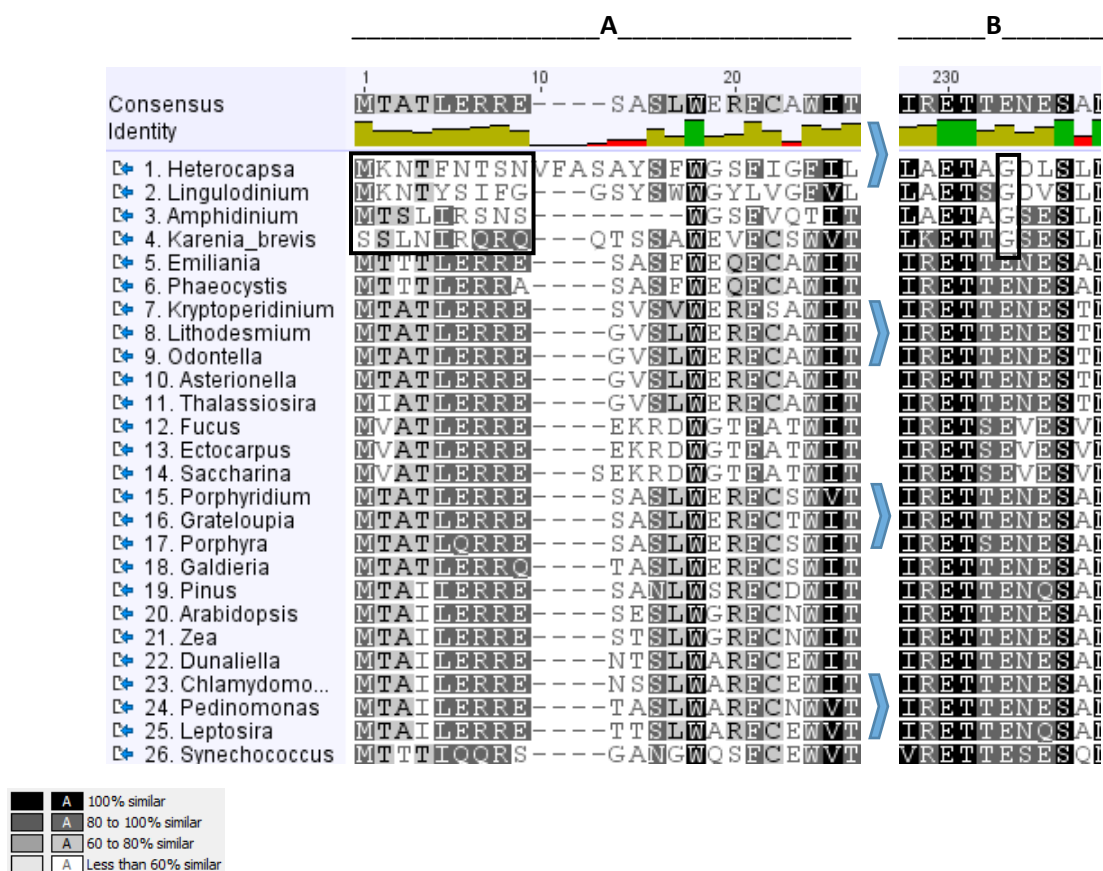


Figure 3-4. Multiple sequence alignment of the *psbA* proteins of 26 taxa, including representatives from 3 groups of peridinin-pigmented dinoflagellates, as well as the putative *K. brevis* sequence identified in Chapter 2. Section A shows the first 26 residues of the N-terminus of the aligned sequences, and highlights the lost region of consensus, and the missing methionine from the *K. brevis* protein. Section B shows a region of internal consensus loss, with a consistent preference for substituting the glycine residue for glutamic acid at the 233rd position of the alignment, which is present in the proteins of both peridinin and fucoxanthin plastid lineages. The sequence from *Kryptoperidinium*, which has a diatom endosymbiont, is congruent with non-dinoflagellate taxa in both respects. Similarity values according to Blosum62 scoring matrix.

The dinoflagellate protein sequences of *psbB* do not exhibit the shared loss of N-terminal consensus regions, as the previous proteins discussed have (Figure 3-7). There is a single site in this region of all peridinin and fucoxanthin taxa, where an isoleucine is substituted in place of a threonine. The threonine in this position is invariant across all reference taxa. Another position shown in panel B shows a preference for a serine substituted for an invariant aspartic acid. This preference is not uniform, however, as *A. carterae* has substituted asparagine at the same position. Five other positions of invariant site loss are shared across the peridinin and fucoxanthin dinoflagellate taxa, but no homogenous substitution preference is seen.

As discussed above, *psbE* was particularly disruptive to expected branching patterns when added to the concatenation. The protein alignment in figure 3-8 shows that much of its C-terminal sequence is affected by invariant consensus loss. There is one instance of shared site substitution preference, in which the dinoflagellates are seen to prefer a leucine over the serine seen in all reference taxa.

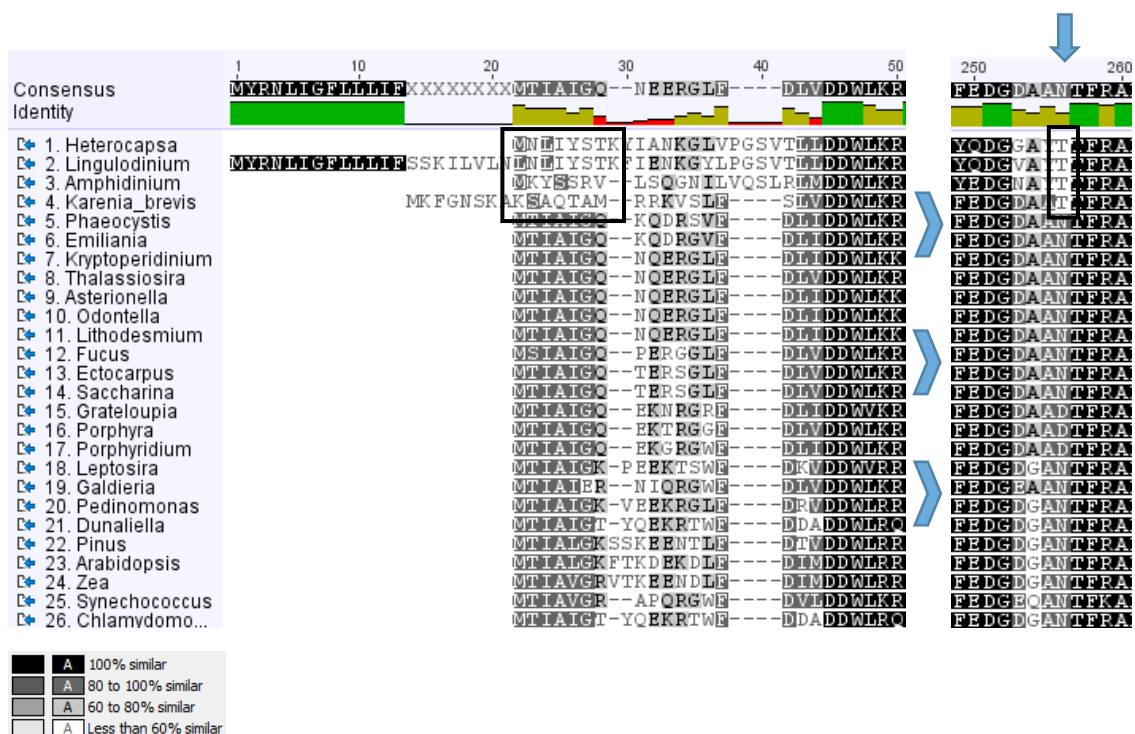


Figure 3-5. Multiple protein sequence alignment of psbD from 26 taxa, showing the N-terminal region, and an area ~250 residues into the alignment. Significant loss of consensus in the well conserved N-terminal region across all dinoflagellates, when compared to proteins from taxa representing the major groups of photosynthetic eukaryotes. The vertical arrow indicates an invariant site substitution of a threonine for asparagine, which is congruent among the peridinin dinoflagellates and the fucoxanthin-pigmented *K. brevis*. Sequences aligned with MUSCLE, with similarity values according to Blosum62 scoring matrix.

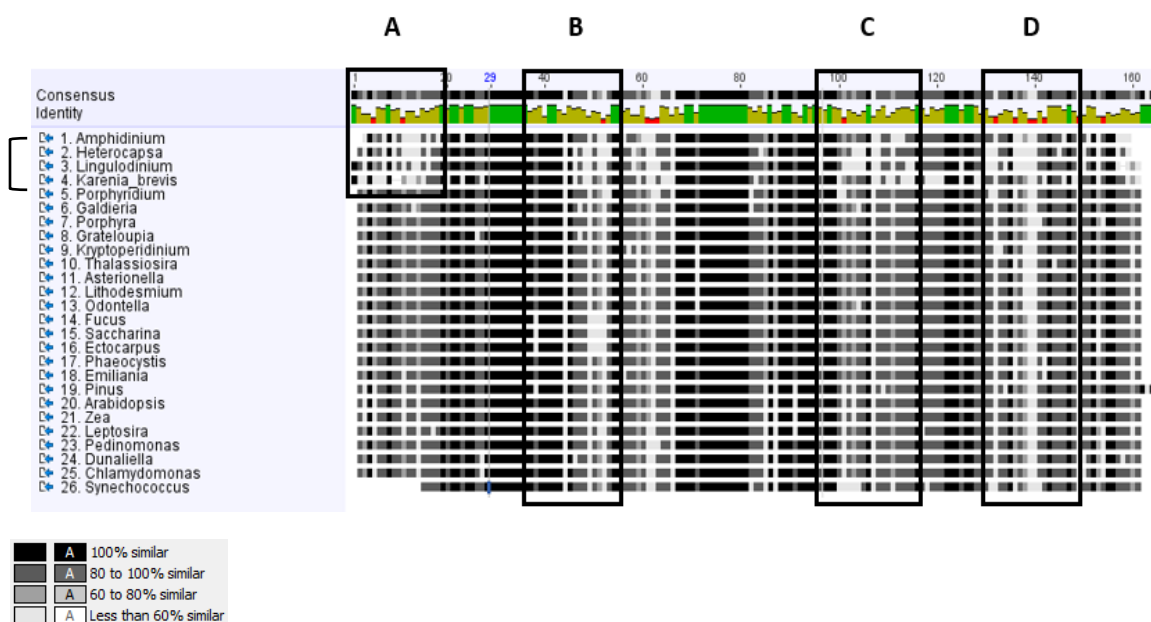


Figure 3-6. Multiple sequence alignment of complete *petD* proteins for 26 taxa, including representatives from most major photosynthetic groups, *K. brevis* and 3 taxa of dinoflagellates containing peridinin-pigmented plastids. Loss of invariant sites at the N-terminal region of the protein is highlighted in box *A*. *B*, *C* and *D* denote the approximate locations of transmembrane regions identified in the *petD* protein of *Chlamydomonas reinhardtii*. Similarity values according to Blosum62 scoring matrix.

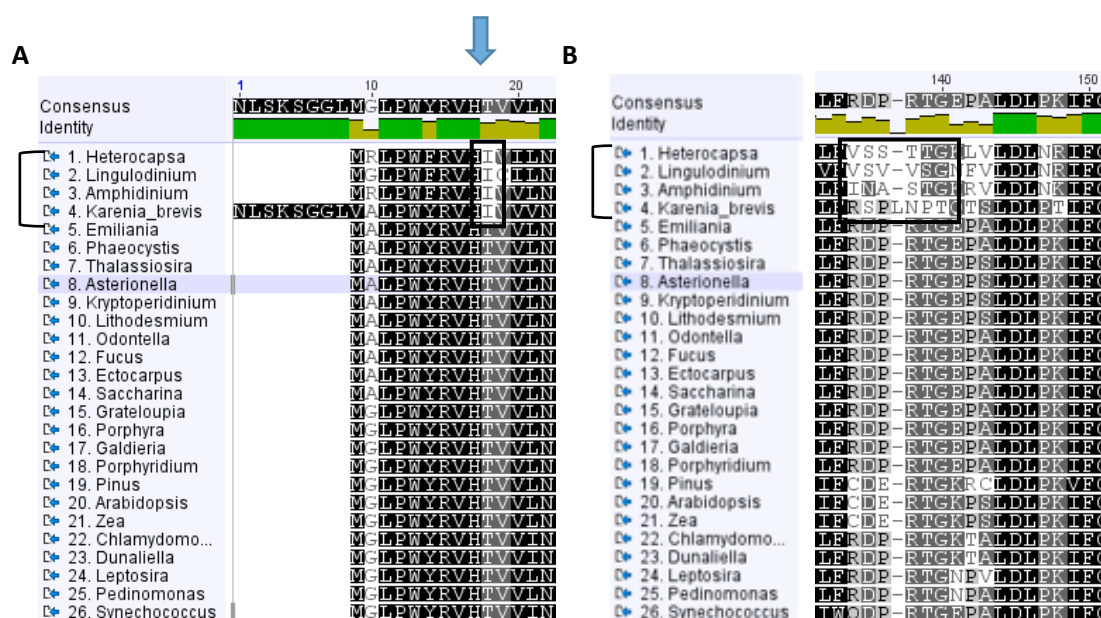


Figure 3-7. Two selections from the MUSCLE multiple sequence alignment of *psbB*, with dinoflagellate taxa containing peridinin and fucoxanthin pigmented plastids denoted by brackets. Arrows indicate positions with preferred substitution in both peridinin and fucoxanthin dinoflagellate taxa, boxes are shared lost invariant sites. The N-terminal region in panel **A** shows no significant loss of consensus compared to that of *petD* in Figure 3-X. A single position was found with a uniform preference for substituting the hydrophilic isoleucine for a hydrophobic threonine that is invariant across all other taxa examined. Panel **B** contains the first tetratricopeptide repeat (TPR) region, highlighting another instance of an invariant site substitution preferred across both plastid lineages, as well as other areas of reorganization in the TPR region shared amongst peridinin and fucoxanthin dinoflagellates. The protein from the diatom plastid in *Kryptoperidinium* is in agreement with the consensus regions in non-dinoflagellate taxa, as seen in the 9th sequence in the alignment. Similarity values according to Blosom62 scoring matrix.

Discussion

We have attempted to use the sequence data produced in Chapter 2 of this study to conduct a phylogenetic analysis comparing the *petD*, *psbA*, *psbB*, *psbC*, *psbD*, and *psbE* plastid protein sequences of the peridinin and fucoxanthin dinoflagellates. The maximum likelihood trees produced using the sequences of individual proteins were not similar to one another with respect to topology, and often showed low bootstrap support at nodes relevant to the placement of the peridinin and fucoxanthin dinoflagellates. Concatenation of the protein alignments for each gene resulted in more stable and well-supported topologies that took two forms, based on whether or not *psbE* was included in the concatemer. Addition of *psbE* to the other 5 proteins resulted in movement of the peridinin dinoflagellates from the heterokonts into a monophyletic group with the haptophytes and fucoxanthin dinoflagellates (Figures 3-1 and 3-2). This is similar to the findings regarding the unsuitability of *psaA* for phylogenetic analysis in a previous study (Shalchian-Tabrizi et al. 2006). They also identified synapomorphic characters in *psaA* protein sequences that are shared between the peridinin dinoflagellates and haptophytes. We have not found sufficient evidence to highlight that here, however we have noted broader regions of sequence that are dissimilar to the well-conserved (nearly invariant) motifs found in all other photosynthetic eukaryotes (Figures 3-4, 3-5, 3-6 and 3-7). In some instances these amino acid substitutions are uniform across both dinoflagellate plastid types. This was seen to some degree in each protein examined. The protein sequence of *psbE* shows a high degree of this type of general consensus loss related to total sequence length, which may be why it

significantly alters the topology of the concatenations, despite only comprising approximately 5% of the total information therein.

The loss of invariant sites in the peridinin and fucoxanthin lineages, while not highly congruent between them, infers that certain areas of contiguous invariant sites, both with and without defined function, are subject to loss in these taxa. The peridinin dinoflagellates have lost substantially more invariant sites than seen in the *K. brevis* proteins examined, however, as mentioned above, discrete instances of homologous invariant site substitution preference between these two groups are noted. The ML trees produced in this study with each putative plastid lineage removed infer that these areas of lost consensus similarity/invariant sites may be sufficient to cause the monophyletic grouping of peridinin and fucoxanthin dinoflagellates seen in this and other studies (Shalchian-Tabrizi et al. 2006, Yoon et al. 2002, and Yoon et al. 2005).

The environment within the dinoflagellate host cell provides for a substantially higher rate of evolution than seen in most eukaryotes (Gagat et al. 2014). When a new plastid is acquired by a dinoflagellate, the endosymbiont's genome becomes subjected to this rate increase. Heterotachy, which describes variation in the rates of substitution at a given site over time, as well as alternate codon preferences, have been shown to play a role in confounding DNA-based phylogenies of dinoflagellate plastid genes, causing confusing tree topologies (Lockhart et al. 2006, Shalchian-Tabrizi et al. 2006). This is mitigated to a large degree by using protein sequence data. However, even protein sequence trees have problems with long-branch attraction between

dinoflagellates containing tertiary plastids of different lineages (Yoon et al 2002 and Yoon et al 2005). The branch length for a given taxa or group in a phylogenetic tree is a representation of the amount of evolutionary change that has occurred since the previous split from the presumed common ancestor. The rapid evolution of the dinoflagellate plastid genome in both peridinin and fucoxanthin pigmented dinoflagellates causes branch lengths much longer than in the other taxa in the alignment (Figures 3-1, 3-2, 3-3, and Appendix D), which can confuse tree building algorithms, causing them to group the taxa possessing abnormally long branches together, despite having polyphyletic origins (Shalchian-Tabrizi et al. 2006, Yoon et al. 2005).

While protein phylogenies can minimize the impact of base heterotachy and codon convergence, they do so at the expense of reducing the length of sequence data available to be analyzed by a factor of 3. This is ordinarily dealt with by sequence alignment concatenation of multiple proteins, which increases the total sequence data provided to the phylogenetic software and the confidence of its decisions. This method, however, does not allow for one to examine the lineage of individual genes discretely from one another, and could potentially “average out” the impact of a single gene. This is important when more complex methods of gene acquisition mean direct assumptions of the lineage of the gene may be inaccurate, such as in the dinoflagellates.

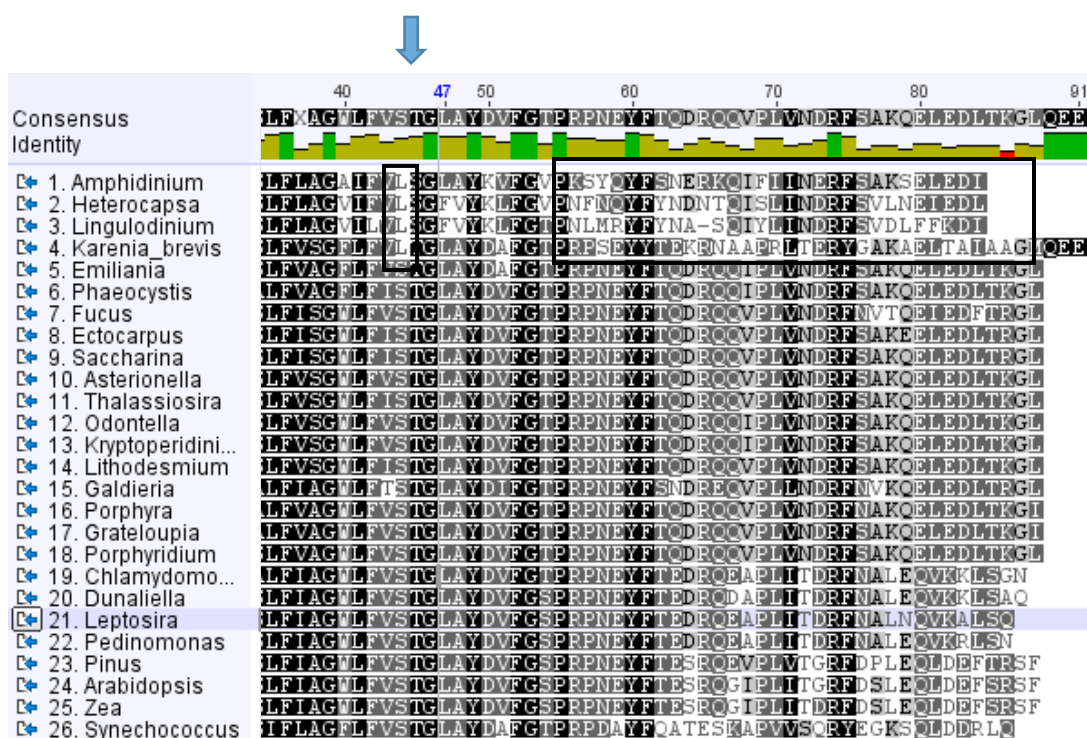


Figure 3-8. Protein sequence alignment of psbE. The box indicated by the arrow contains a substitution site that's homogenous across both groups of dinoflagellates. The second box highlights a large area of consensus loss which is shared between both dinoflagellate groups. Consensus determined by Blosum62 scoring matrix.

The combined results of this chapter's analysis open interesting questions about the nature of dinoflagellate host cell effect upon the genome of its plastid endosymbiont. Evidence of shared regions invariant site loss in functional and non-functional protein regions, and discrete sites of homologous substitution shared between both plastid types, may have significant implications upon discovering the relationship between the plastid genomes of these two dinoflagellate groups.

REFERENCES

- Barbrook AC, Symington H, Nisbet RER, Larkum A, Howe CJ (2001) Organisation and expression of the plastid genome of the dinoflagellate *Amphidinium operculatum*. *Mol Genet Genomics* 266:632-638.
- Barbrook AC, Voolstra CR, Howe CJ (2014) The chloroplast genome of a *Symbiodinium* sp. Clade c3 isolate. *Protist* 165:1-13.
- Barbrook AC, Dorrell RG, Burrows J, Plenderleith LJ, Nisbet RER, Howe CJ (2012) Polyuridylylation and processing of transcripts from multiple gene minicircles in chloroplasts of the dinoflagellate *Amphidinium carerae*. *Plant Mol Biol* 79:347-357.
- Bjørnland T, Haxo FT, Liaaen-Jensen S (2003) Carotenoids of the Florida red tide dinoflagellate: *Karenia brevis*. *Biochem Syst Ecol* 31:1147–1162.
- Cahoon AB, Sharpe RM, Mysaypohn C, Thompson EJ, Ward AD, Lin A (2010) The complete chloroplast genome of tall fescue (*Lolium arundinaceum*; Poaceae) and comparison of whole plastomes from the family Poaceae. *Am J Bot* 97:49-58.
- Dang Y, Green BR (2010) Long transcripts from dinoflagellate chloroplast minicircle suggest “rolling circle” transcription. *J Biol Chem* 285:5196-5203.
- Dorrell RG, Drew J, Nisbet RER, Howe CJ (2014) Evolution of chloroplast transcript processing in *Plasmodium* and its chromerid algal relatives. *PLoS Genetics* 10:e1004008.
- Dorrell RG and Howe CJ (2012) Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. *PNAS* 109:18879-18884.
- Edgar RC. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797.
- Espelund M, Minge MA, Garielsen TM, Nederbragt AJ, Shalchian-Tabrizi K, Otis C, Turmel M, Lenieux C, Jakobsen KS (2012) Genome fragmentation is not confined to the peridinin plastid in dinoflagellates. *PLoS ONE* 7:e38809.
- Gabrielsen TM, Minge MA, Espelund M, Tooming-Klunderud A, Patil V, Nederbragt AJ, Otis C, Turmel M, Shalchian-Tabrizi K, Lemieux C, Jakobsen KS (2011) Genome evolution of a tertiary dinoflagellate plastid. *PLoS ONE* 6:e19132.
- Gadagkar SR, Rosenberg MS, Kumar S. (2005) Inferring Species Phylogenies From Multiple Genes: Concatenated Sequence Tree Versus Consensus Gene Tree. *Mol Dev Evo* 304B:64–74.

Gagat et al. (2014) Book Chapter from *Endosymbiosis: Tertiary Plastid Endosymbioses in Dinoflagellates*. Edited by: Wolfgang Löffelhardt. Springer-Verlag Wien 2014.

Gao L, Yi X, Yang YX, Su YJ, Wang T (2009) Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC Evol Biol* 9:130.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38(Web Server issue):W695-9.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol.* 29(7):644-52.

Green BR (2004) The chloroplast genome of dinoflagellates – a reduced instruction set. *Protist* 155:23-31.

Hackett JD, Anderson DM, Erdner DL, Bhattacharya D (2004a) Dinoflagellates: a remarkable evolutionary experiment. *Am J Bot* 91:1523–1534.

Horwich A, Fenton W, Chapman E, Farr G (2007) Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol* 23:115-145.

Howe CJ, Nisbet RER, Barbrook AC (2008) The remarkable chloroplast genome of dinoflagellates. *J Exp Bot* 59:1035-1045.

Inagaki Y, Dacks JB, Doolittle WF, Watanabe KI, Ohama T (2000) Evolutionary relationship between dinoflagellates bearing obligate diatom endosymbionts: insight into tertiary endosymbiosis. *Int J Syst Evol Microbiol* 50:2075–2081.

Jackson CJ, Gornik SG, Waller RF (2012) A tertiary plastid gains RNA editing in its new host. *Mol Biol Evol* 30:788-792.

Janouskovec J, Horak A, Obornik M, Lukes J, Keeling PJ (2010) A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *PNAS* 107:10949-10954.

Jeong HJ, Du Yoo Y, Kim JS, Seong KA, Kang NS, Kim TH (2010) Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. *Ocean Sci J* 45:65–91.

Keeling PJ (2010) The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* 365:729–748.

- Kumar S, Stecher G, Peterson D, and Tamura K (2012) MEGA-CC: Computing Core of Molecular Evolutionary Genetics Analysis Program for Automated and Iterative Data Analysis. *Bioinformatics* 28:2685-2686.
- Lockhart P, Novis P, Milligan BG, Riden J, Rambaut A, and Larkum T. (2006) Heterotachy and Tree Building: A Case Study with Plastids and Eubacteria. *Mol Biol Evol* 23(1):40–45.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung D, Yiu SM, Peng S, Xiaoqian Zhu, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW and Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 2012 1:18.
- Minge MA, Shalchian-Tabrizi K, Torresen OK, Takishita K, Probert I, Inagaki Y, Klaveness D, Jakobsen KS (2010) A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum*. *BMC Evol Biol* 10:191.
- Nassoury N, Cappadocia M, Morse D (2003) Plastid ultrastructure defines the protein import pathway in dinoflagellates. *J Cell Sci* 116:2867–2874.
- Nisbet RER, Hiller RG, Barry ER, Skene P, Barbrook AC, Howe CJ (2008) Transcript analysis of dinoflagellate plastid gene minicircles. *Protist* 159:31-39.
- Nosenko T, Lidie KL, Van Dolah FM, Lindquist E, Cheng J-F (2006) Chimeric plastid proteome in the Florida “red tide” dinoflagellate *Karenia brevis*. *Mol Biol Evol* 23:2026-2038.
- Oudot-Le Secq M-P, Grimwood J, Shapiro H, Armbrst EV, Bowler C, Green BR (2007) Chloroplast genome of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: comparison with other plastid genomes of the red lineage. *Mol Genet Genomics* 277:427-439.
- Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK (2007) Comparative chloroplast genomics: analysis including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8:174.
- Reguera B, Velo-Suarez L, Raine R, Park MG (2012) Harmful Dinophysis species: a review. *Harmful Algae* 14:87–106.
- Richardson E, Dorrel RG, Howe CJ (2014) Genome-wide transcript profiling reveals the coevolution of plastid gene sequences and transcript processing pathways in the fucoxanthin dinoflagellate *Karlodinium venficum*. *Mol Biol Evol* 31:2376-2386.

- Rozen S, Skaletsky HJ: Primer3 on the WWW for general users and for biologist programmers. In *Methods Mol Biol* Edited by: Krawetz S, Misener S. Humana Press, Totowa, NJ; 2000:365-386.
- Sanchez Puerta MV, Bachvaroff TR, Delwiche CF (2005) The complete plastid genome sequence of the haptophyte *Emiliana huxleyi*: a comparison to other plastid genomes. *DNA Res.* 12:151-156.
- Shalchian-Tabrizi K, Skanseng M, Ronquist F, Klaveness D, Bachvaroff TR, Delwiche CF, Botnen, A Tengs T, Jakobsen KS. (YEAR)Heterotachy Processes in Rhodophyte-Derived Secondhand Plastid Genes: Implications for Addressing the Origin and Evolution of Dinoflagellate Plastids. *Mol Biol Evol* 23(8):1504–1515.
- Stern DB, Goldschmidt-Clermont M, Hanson MR (2010) Chloroplast RNA metabolism. *Annu Rev Plant Biol* 61:125-155.
- Tamura K, Stecher G, Peterson D, Filipowski A, and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol and Evol* 30:2725-2729.
- Takano Y, Hansen G, Fujita D, Horiguchi T (2008) Serial replacement of diatom endosymbionts in two freshwater dinoflagellates, *Peridiniopsis* spp. (Peridinales, Dinophyceae). *Phycologia* 47:41–53.
- Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF, Jakobsen KS (2000) Phylogenetic analyses indicate that the 19' hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol* 17:718–729.
- Wernersson R. (2006) Virtual Ribosome--a comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Res* 34:385-388.
- Wisecaver J, Hackett J (2010) Transcriptome analysis reveals nuclear-encoded proteins for the maintenance of temporary plastids in the dinoflagellate *Dinophysis acuminata*. *BMC Genomics* 11:366.
- Wisecaver JH, Hackett JD (2011) Dinoflagellate genome evolution. *Annu Rev Microbiol* 65:369–387.
- Yang JB, Tang M, Li HT, Zhang ZR, Li DZ (2013) Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol Biol* 13:84.
- Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, Zhao D, Al-Mssallem IS, Yu J (2010) The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS One* 5:e12762.

- Yoon HS, Hackett JD, Bhattacharya D. (2002) A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *PNAS* 99(18):11724-11729.
- Yoon HS, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D (2005) Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol Biol Evol* 22:1299-1308.
- Zapata M, Jeffrey S, Wright SW, Rodríguez F, Garrido JL, Clementson L (2004) Photosynthetic pigments in 37 species (65 strains) of Haptophyta: implications for oceanography and chemo-taxonomy. *Mar Ecol Prog Ser* 270:83–102.
- Zerbino DR and Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* 18:821-829.
- Zhang Z, Cavalier-Smith T, Green BR (2001) A family of selfish minicircular chromosomes with jumbled chloroplast gene fragments from a dinoflagellate. *Mol Biol Evol* 18:1558-1565.
- Zhelyazkova P, Sharma CM, Forstner KU, Liere K, Vogel J, Borner T (2012) The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* 24:123-136.

APPENDICES

APPENDIX A: *KARENIA BREVIS* PLASTID TRANSCRIPTS

Start codons are highlighted in green and stop codons in red. Ambiguous or non-canonical start codons are highlighted in blue.

petB

GTAAACTTCATTTTGAAGATCCTAATCTTTTATAAGCTGTGTACTTAAAACACTAGTTTTTTGCAA
TAAACTAGAATCCTTTGACCTAAAATTACCTATAACTGATTACGCGTCATCTGCAATTGATGGAA
CATCATTTGGCGGCTGCCTGCTACATCTTATCTACAACAGTAGGGGAGTCAGCAGGTTCCAACCGG
TTAGTTGTCGACTAATCAAATATCCTTCATTGTTACGGAATTTTATTGGGCCATCGTAGGTTG
AGGCCCATTTCTCGAAACGTTGTAAGTGGCTAAATTTGCCGCCAGAAGCTGTTCGAATAGTTTGCTA
CTGCTTCGAAGATATTGGATGGTTCCGTTGGCAACTAATAATATGATAGTATTACCTACTTGTTT
TCAATAGACATGATATTTGACCTAACTAGCTAGTGGTATCAATTTCCCTTCAAGAGCGACTTGATC
TTCAGTCGCTTTCTGATGATGTGTCTCAAATATGTTCCACCCACGTTAATCTACTTTATTGCT
ATTGGGGGGTTAACTCTTACAAGTTTTTTACTACAACCTAACAACGGGGTTCTCACTAACATTTTA
TTACAGACCCATCGTTGCTTCGGCGTTTTCGACCGTGTCTGTTCTTGTTCATCTACGCTTCACTATG
GGTGGGTAATTCGTGCAGCTCATCGGTGGACTGCCAGTCTGATGGTACTTGCTCTACTAGGGCAT
ATTTACAGAGTTTATCTTACAGGGGGTTTTAAACGACCACGTGAACCTAAGTGGCTGACAGGTTCT
CGCGTTGTCATGTCTTACAGTAAGCTTCGGGGTTACTGGATATTCTTTACCTTGGGATCAAGTTG
CATATTGGGCTTGTAATTTGTGACGGGTGTTCCAGATGCTATTCTCTTATTGGCGGTCAAATC
GTTGCACTTCTTCGTGGCGCTGACACTGTAGGACAGTCAACGCTGTCGCGTTTTCTTTGAAACACA
TACTTTCTTATTGCTGCGCTAACCACGGTTCTAATGTTAACACATTTCTTCTTATCCGAAAGC
TAGGTATTTTCGGGGCCACAGTAACTCAAAGAGTAGCTTCTATGTTTGAAGATTTCATAGATTTGAA
ATTCTGATTCTCTAGTTATAGTGGCAAGAAACCGAGCTCATCATGTTATGCATNGTATTTAGTGG
ACAATATCTATGTTTCAATTTATAACATTTTGAAGAGTTCATCGTAAGATTACCTTGTGAAAACT
CACCTCCCTCAATCACAGTTTTTACCCTTCAACTTTTTTTATTTGAAATTCAACTTTTGCTCTATTT
TTAAAGTGCTTTGTAGACTAACTCGTTCTAATGATCTCATGAATAAAAGTCCTTCTTTTTTTTAA
GAATTCAAACGTTACTGAACCTTCGCTTAAATCTATATCGTCGATGGTTCCGGGAAGGAAAAACCGT
AGCGAATACAGTGGCTACTTTTTGTCCAGAATCTATAGGAGATAATGGATCTTTACGAAGTCTAT

petD

AA
AGTTCCCCCAGGATATCAGTAACGTAAATCCCCTACCGAGCCAGGCTAAATCTCGGGCAAACCTT
TGTAAGGGGGCAGGTCATAATACATTCGGGGAACCGGCTTGGCCTAACGATTTACTGTTCTTATT
TCCCATCTGTATTACAGGTGTACTTGGTTGTAGTTAGGTCTAGCAGTTCTATCTCCGCCGGCAA
TCGACGCCCAATCAGACCCGTTTTCTACACCGCTGGAATTTCTTCCGGAATGGTATTTCTTACCA
TCATTTAATTTGCTACGTCTTCTTCTAATAAACTTTTAGGTGTTGCTGCGATGGCCAGTATAACC
TCTTGCATTAACGCAGGTAGCAGCTTCCGAGAACCAGACCGCTTCTCAAATCCTAATCGACGCT
CGCAAAGTTCACTGTTATATATCCTGGGATCTTCGTTCATCATCATGGCTTGGCGCAGGAGGTGCT
ATGCCGCTTTCGCAGTCTTTAACTTTAGCACTTAATTAATAATTTTTTTTTTTTTTTTTTTTTTTT

psaB - psaC di-cistronic mRNA

AAAGTTAACACAAAAGTTTCCTCGGGATTG
 GAGCGCAAATTGGGTGGTAAATTTTCATCTAAAGCTAAATATTGGTTCGAGACCGATAGCAAACAA
 GTACCATGAGGGAAAGGTGAAAAGGACTTTGAAAAGAGAGTTAAAAGTGCCTGAAATTGCTGAAA
 GGGAA GCGAATGGAACCGATTGTTCTTGGTTCGTTGGCGATTACGGTTGCGGACTTTTCCTTCTT
 TCTCAAAGGACTAAGACAAGATCCAACACTACGCGACGATTGTCATACTCTGTGGGTATTGCCCAT
 GATTTGGAGTCACATGATTTTCGTTACGGAATCACTTTTGTATCAACGTATTTTTGCTTCTCATT
 TGGGCACCTAGCAATAATATTTTTATGGTCCGCTGGAATCTATCCACGTAGCATGGCAAGGAA
 ATTTTCAAGAGTGGATTATAAATCCTCTGAAAACGTCACCAATTGCGCATGCGATATTCGATCCG
 CACTTTGGAGTAGACGCGCTTAAAGCCTTACGCCATCTGCCGGCTCTTATCCCGTTAATATATC
 GACCTCAGGGTTGTATCATTGGTGGTATACTATCGGCTTACGATCAAACACTGACTTGTATTCCG
 CTTCAATGTTCCCTTATTCTATGCAGTCTTCTGTTTTTATATGCAGGGTGGCTTCATTCTACACCT
 GGTTCGGCAACCGAACCTAGCGTTTTTTAAAAACAATGAATCAAGATTAATCATCATCTGTCCGG
 TCTTTTTGGGGTAAGTTCGCTGGCCTGGGCAGGGCATCTAATTCATGTTGCAATACCTGCGTCTC
 GTGGCCAAGTTGTTAGTCTGGATCATCTTCCGTTACCCATCCCGCTGGGTTAACCCATTCCCTA
 ACGTTTAATTGGGGTATTTACGCCGAAAATCCTGATTCTCCAACCATGTGTTTGCAGACTACAGA
 AGGTGCAGGTACGGCAATCCTCACTTTCCTAGGAGGATTTTCATCCTCAAACCCAAGCTATGTGGC
 TGACTGATATTGCCCATCATCATTTAGCAATCGCAATCATTTTTTATTGTTGCAGGGCATATGTAT
 CGAACACCTATGCTTGGGGCCATCGCTTTACCGAGTTGTTGCACGCGCACGTTCCCTCCTAAAGG
 ACGTCTTCGCGCTGGTCATGTTGGCTTAATGGAACGTTGACCAATAGTTTACACATGCAACTTG
 GTTTAGCGCTGGCAGCGCTAGGCGTTGCAACTTCGCTAACGGCCAGCATATGTATGCGTTACCT
 GCCTACGCATTTATCGCGAGCAGCCCCGTAACCAAGCAGCATTTGTACGTACACCATCAGTATAT
 CGCGGGTTTTCTAATGACGGGGCCCTTCGCCCATGGTGCCATTTTTTTTTCGTCCGTGATTACGAGC
 CTTCGAATGAAGGAAATGTACTTGCAGACAGTACTTCTTACAAAAGAGGCTATAATTTCTCACTTG
 AGTTGGGTGAGTTTATTTTTGGGTTCCATACGTTGGGTCTCTATGTGCACAATGATACTTGCGT
 TGCATTTGGCCTTCCCGAAAATCAAATATTATTGGATCCAATATTTGCTCAACTCGTTCAAGCCG
 CATCAGGTAAAGCAGGAAATGGAATTGATATTCTTCTTTTCGGAACCTACAAGTCCCGCTAGTATT
 GCAAGTAGTAAAGTTTGGTTGCCTGGGTGGCTCGCGGGAATTAACAGCCAGCGAAGTACCGCCTT
 CTTACCAATTGGACCGGGTGATTTCCCTTATTCATCACGCAATTTCACTGGGATTGCATACAAC
 CGCTAATCTTAGTAAAGGGTGCTCTTGATGCTCGAGCCTCAAGATTACTTCCAGATAAGAAAGAT
 TTTGGTTATAGTTTCCCGTGTGATGGTCTGGACGTGGTGGTACTTGTGACGTTTCAGCTTGGGA
 CGCCTTTTTATTTGGCTGTCTTCTGGCAATTAATACTACGGCATGGACCACATTCTATTGGCACT
 GGAAACATTTGACATTATGGGGTGGCAATAGCGCCCTTTTCAATGAAGCTTCAACATACATTATG
 GGTGGTTACGCGATTATTTATGGTTAAACTCAGCTCCAACGATCACCGGTTATACGGCACAAGG
 AATGAACGCACAGGCTGTTTGGACATGGATGTTCTGTTTGGACACCTTACTTGGGCTACCGGTT
 TTATGTTCTTAATTTTCATGGCGAGGATATTGGCAAGAACTTATTGAAACCATAGTTTGGGCACAT
 GAAAGAACCCCACTAGCTAATTTGCTAAAATGGAATGACAAGCCTGTGGCTCTGTCAATTGTCCA
 AGCACGTCTCGTAGGGTTGACACATTTACGACAGGGTATATATTTACTTATGCCCTTTTTTAA
 TTTCTAGTACGTCCGCCAAAACAGCTGGAATATCTACTGCCCGCCGCTTATTTAAAGATCTACCT
 GCGGAAATTATCTAGAACGTACAAAAAATGATTACTTATGATTTCGCTTCAAGAAAAGTTAATTG
 TTTTACAAATAGAAAAAAGTTAGTCCCTTGCAGCCTCAGCATATCAAGAAAAAAGTTTTTAATAA
 GAATTAGTCTGGTGAAGATACGAGGACAATCTACATTTTTTAAATCTTACTGAACTTGAACATAA
 ATTTGGACTCTAATTTACGGAAGCTACTCTCACAAATCGTTTCGAACACTTAAAACATATTTTTAC
 CCTACAGCTTGCAAAATCTTCTGGTGAACACTTTTTCTTCTCATTCTTAATTTTTAAAGGAGT
 CTAAGTCCAAACTTTTTCACTAAAACAATAATACGTTAGCCCCATCGATCGACTCGTTTCAGGTC
 CCAAGTATACCGTATGCTTAAGTAATCCGTGGGGCATGCCGATTCACAACGCTTACAACCAATG

CAGTCGGCAGGCCGCGGGCAGCAGCAACTTGTCTGACTTATTCCATACGGTGTGGGTACCAT
 TGCAAGGACATCACAAAGGACACGCACGGACACACTGAGTACAGCCGATACATGTAGTATAAATCC
 TTAGGGTATGAGACATAAATTTAGATTGTACCGAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
 TTTTTTTTTTTTTTTTTT

psaF

AAA
 TTTATACTATGTTACCTGCCGCTACGGATT
 CAAGAATTTGGCAGCAAAAATTCGCCACCCTTGCCCTAAGTTGCATGATATTCCTTCCTGAGAAA
 GCTCAGTCAAGCACTCCTCTCATCGTTATGTGTAAAGATTCGGCGGCTTTCAACAATCTAGCCAC
 GACAAGATTAACACGTCTGAACACCGCCTTGAGTAATTTGCAACCGCAACACCAGCTTATCGCG
 CTACAATGGAGAGAAAAGCGCGCCGCAAAAACCGATTCTCCCGGCAAGCACTCTTAATGTGTGGT
 CGGAATGATGGTTTACCTCGCTTATTTTCTGCTGACCTAATACGTTCTATAACTCTTTCGCCGT
 GCCCCACTACAGTTTTTGTATATAGCTACTTGCAATGGTTGGGCAGGCCGAGAGTATCTATCTC
 GCACGGCAAATTATAAGTCAGAAATCTAATCGATCTTACTCTAGCTGCTACACTAATGGGGTCT
 AGCCCATTCGAGCTGTAGGAGCTGCGTGGGACGACTTACTCCGTGGAAAGTTGTTAAATACCAC
 ACTTCAGTAAACTTTGTTTTCGTGCGAACAGAAGAAGATAAATTGCAAGATAATGGACATTGGCAA
 CCTAAAATATCTATCGAAATTAGTAGTGTTTTTGAAGTTTTATAACGAAAATCTTAGTAATTCTT
 ATGCCAATATACAACACCCAAATTTGGAATTACGTTTATTACGTCATAGAATATTTATTACGC
 AATTCCATGCTGAAAAAGTTGTGAAAGTAAAGTTTTTTAGTAACTAAATGAGAAAAATGACATCA
 ACAAATTAGAGTAAAAATCGCTATGCCAAAGATAAACAGAGTCGTACGTAATGTCAATAAAA
 GTATTATATTGACTAATCTACAAATGTACATGTTAAATGAGTTAACTCTTTTTGTGACGCGTACG
 ACGAAACGCGAGGTGTTACCTTACGATTAATAATGGATTACAGTGGTCTGGTATCTAGTTGTGAA
 ACTTTTATTGTCATTATTGAAACCATCATAAAATCACCAACTCCAATTGTACACAAGCTCACTAT
 TTCGACGGACTTGGTAATTTTTGGCGCTTGATTACATTTTATGAATCTGTTTGTCTTACTAGCAT
 TGCCAAAACCTTTAATTATAACAATGTTCTATTTCAAGCACTGGTACCTCTAGTTAGACCTTAT
 TCCTGTATTAGTACATGAACCTCTTTACCTCTAATCCGAACACTTTGAGTGAGACGATGAAGCT
 TATCCCACGCCGTCTCACTATTCTATATTATCTATAGTATTCTAAGTTTACCTAGACTTAGTACC
 GTTTACACGGCCCCGCGCCAAAATAGTCGCTTTACCCCCTAAGAATTTCAATAAGATGAATGCTGC
 GCCTCAACACATTTTCGGGGAGATCCAGCTAGCCTCAGATTCGAATAGAATTTCAACTAACC
 AAATTCATCCGTCATTTTTTCAACAATGGACGGTTGGCACCTCCGTTTAGTAATAACTAAGCTTCG
 TACTGAATTTGGTTAGATCATCTGAATTCGGGTCAAAAATCATGACTTAACAAAACCAAATCTT
 ATCACACGACTGTATGGAGACGCTTGCACTATGACGATTATTCATAGAAGAACGCCACTCTTGTT
 CTCGCTGAACATAACAAACAAAATGTATCTTCCATAATATCTAACTAAAGTCTTGCCACAACCCA
 TAAGTCGCTAGCACATTTCTTCAACAGGCACAGTTTCAATCGTTAAGTCTATCTTATCTTGCAAAA
 TAAGCATTACGTTTTTACATACTTTTTTCAATCCCCCTCCCGGGTTCTTTTCAACGTTCCCTCACG
 TACTTTCCG

psbA

ATGAATTATTATAATCCGAAATCATCACTAAACATTCGTCAACGTCAACAAACCTCATCTGCGTG
 GGAGTTTTTTTGTCTGTTGGTTACAACCTACAGAAAACAGACTATATCTAGGGTGGTTGGTTGTC
 TAATGGTACCGACCTTAGTTTTCAGCAACCTTTGCCATATATCATAGTTTTCATCGCGCACCGCCT
 GTCGACATAGACGGTATACGTGAACCTGTGTCAGGGTTCGTTGCTATATGGAAACAACATAATATC
 AGGAGCAGTAATTCAGTTCAAATGCTATCGGTGTACACTTCTACCCTATTTGGGAGCGGCAT

CTTTAGACGAGTGGCTTTACAACGGGGGAACCTACCAAATGGTGGTTTTCCATTTCTTCCTTGGT
 GTATGCAGTTATCTAGGGCGTGAGTGGGAACCTGTCGTATCGTCTAAGTATGCGTCCTTGGATATG
 CGTAGCTTTCTCAGCACCAGTTGCTGCAGCATCTGCAGTATTCATAGTATACCCGATAGGACAGG
 GTTCTTTCTCGGATGGTATGCCTTTAGGAATATCAGGGACGTTCAACTTCATGCTAGTTTTCCAA
 GCTGAGCACAACATATTAATGCATCCATTGCACATGTTTGGTGTAGCCGCTGTATTCGGGGGTTT
 ACTTTTCTCTGCAATGCATGGTTCAGTAGTAACCTCATCGCTTCTAAAAGAAACGACGGGTTCTG
 AGTCTCTGAACTACGGATACAGATTCGGTCAACATAATGAGACCTATAACATAGTAGCAGCGCAC
 GGATATTTTGGGCGACTAATCTTCCAATACGCTTCATTCAATAACTCAAGATCATTGCACTTCTT
 CCTAGGTGCGTGGCCAGTTATTGGTATATGGCTAACGTCTATGGGAGTTTCAACCATGGCGTTCA
 ATCTTAACGGGTTAAACTTCCAACAGTCAATCTTAGATGCTTGGCGGTGTCCAATAAGTACTTGG
 GCGGACATTCTAAACCGCGCAAACCTTGGTATGGAGGTAATGCATGAAAGAAATGCACACAACCTT
 CCCTCTGGATCTTGCTAGCACATCTAATAATAACACGAACACAACATGGCCTATAACTCTAGGAT
 AA

psbB - no obvious start codon

AAATTTTTTTTAAATTTTCAATTTAGA
 ATTTATCGAAGAGTGGTGGCCTGGTGGCTTTACCATGGTATAGGGTACATATTGTCGTAGTAAAT
 GACCCGGGACGACTAATTGCCGTTTCATCTAATGCATACGGCTCTCGTGTCCGGTGGGCTGGAGC
 CATGGCTTTGTATGAATTGGCTGTTTTTCGATCCTGGCGACCCAGTCTTGAATCCGATGTGGCGGC
 AAGGAATGTTTTGTTTTACCATTTCATGACGCGTCTTGGTGTACGCAATCATGGGGAGGTTGGAGT
 ATAACAGGCGAAAGCTTAGGTGGTAAATCGGTTGGTTTTATGGAGTTATGAGGGTGTGCTTTTGTG
 ACACATAGGTCTTTTCGGGCTTATGTTTCCTTGCTGCTTTATGGCACTGGCGTTTCTGGGATCTAG
 AACTGTTCCGATCACCATTAAATCCGACTCAAACATCGTTAGATTTACCGACAATTTTTGGTATT
 CATTTACTCCTGGCATCTCTTCTATGCTTTGGTTTTGGTTTCTATCATGTAACAGGGACATACGG
 TCCTGGTATTTGGGTCTCAGATGCATATGGTATTTGTGGAAGAGTACAAGGAGTATCACCTTCAT
 GGGGTGCTGCAGGTTTTAATCCATTCAATCCGGGTGGAATTGCAGCGCACCATATCGCGGCCGGA
 GTGCTAGGAACTCTAGCGCGGTTTTTCATATAACCAATCCACCCCCAGAGGTTTTATTTTTAGC
 CTTGCGAATGGCTAATATAGAAACGGTCTTTTCTAGTAGTATCGCTGCTGTTTTCTTCGCAGCTT
 TTGTTACCAGTGGAAACAATGTGGTATGGCGCAGCTGCAACGCCCATTGAGCTTTTTGGCCCAACC
 CGTTATCAATGGGACTCAGAATATTTCTTAAAAAGATCACGCAAACGGTTAACTACTATAAGGG
 ATTGTCCCTAACAGATTTCGGACGCATGGGCTCTTGTGCCTGAAAAATTAGCCTTTTATGATTATG
 TCGGAAATAATCCCGCAAAGGTGGACTTTTTCAGGGCAGGCCCTTTAACAAAGGGCGATGGAGTT
 GCAAAAGCTTGGCTTGGACACCCGCAATTTTATGGTCCAGATGGGTCAATATTGAGAGTGCGTGCG
 AATACCCGCTTTCTTTGAGACACTACCCGTACTTTTATTGGATACACGTGCCCTTGTTCGGGCGG
 ATATCCCATTCAGACGAGCTGAATCAAAGTTTAGTATCGAACAAATAGGCGTATCTGCAGTCATC
 TTAGGTGGAAGAGACGCCGGAGCCGAGTTTACGTCGGCAGCAAGCGTTAAAGCGCTTGCAGAGAAA
 AGCTCAATTCGGAGAACTTTTCGAATTTGATAGAGCGAAATTAATAATCGGATGGGGTGTTCGGTT
 CTAGTCCAAGAGCCTGGTATACTTATGCACACCTAAATTTTTCGTTCTTATTCTTCTTGGGACAT
 TTGTGGCACGGAGGACGTACACTTTTCCGTGATGATTTTGCGGGGATTGGTGTGACTCAATTGA
 ACAAGTTGAGTTTCGGGGCTTTTCGTAAAGTTGGGCGATGCTCGCACAAAGAAAAAACCGCGCAAT
 TGAAATCAATAATTT

psbC

AA
 AATTTTTTAAAAAAAAAAAAAAAAAAAAA
 AAAAAAAAAAGTTGGTTAAAAGTAAAAATCCCAGTAAATTTATGAAAAGGGAAAATCGAGAGTGTCC
 TCGAACGCTAGAAAAGTTTGCACCTCGGTGTGTATTTACTCAAACGTATCGGGGCAAAAATACCC
 GGACGGAAAATAACGGGATAGGGCGTGCAGCGGAGTCGAAAAACAGTAAATGGAAAAATCATGCT
 GAAAATTCCTCCGCTGCCGGTTCGTGACTTACTCACTACAGGGTATGCTTGGTGGTCTGGAAATGC
 TCGATTAATTTCTGTATCCGGGCGACTTTTGGGCGCTCACGTAGCACATGCCGGTCTTATGATTT
 TATGGTGTGGTGCAATGACGCTATTCGAAACAGCACATTATATTCCTGAGCTTCTCTCTATGAG
 CAAGGGTGTATTTTACTTCCACACATTGCCACATTAGGCTGGGGTGTGACCCTGAAGGGGAAAT
 ATCAGGAACCTTATGCAGCTTTCGTAATTGGTGTACTTCACTTATTAAGTGCTTCTGTAGTTGGCT
 GTGGTGGACTTTACCATGCACCTCTTGGCCCGGATACGCTCGAGGAAAACCTTGTCTATTATGGG
 TATGACTGGCGTGACAAAATAAAATGACAACAATTCAGGTTTTTCATCTAATACTTCTCGGTCT
 TGGATGTTATTTACTTGTGCTAAGGCTCTTTATTTTCGGAGGTATTTACGATACCTGGGCTGCAG
 GAGGTGGAGATGTAGTTCGAGTTACAGCACCGACCCTAAATCCTATAATAATAGCTAATTATGCG
 CTTAAGTCACCTTTTGGAGGAGATGGGTGGGTATTAGCGTCAATAATCTTCAAGACCTAGTAGG
 AGGTCACGCATGGATGGGATCGATCCTTATTTTAGGTGGTCTCTTCCATCAAAAATACAAAACCAC
 CTGCGTTTTGTAAGAAGAGCATTCAATTTGGTCTGGCGAAGCATATTTATCTTATTCACTAGGAGCA
 TTAACCATTTGTGGTTAACC CGCAGCAGTTTTTGTCTGGTATAATAATACCGCTTATCCTAGTGA
 ATTTCTTTGGTCCAACGCAGCCGAAGCTTCTCAAGCTCAGGCATTCACGTTCTTGTTCGAGACC
 AAAAATGGGAGCAAAAAGTAGCATCTGCTCAAGGACCTACGGGATTAGGTAAATATCTTATGAGA
 TCGCCGAGTGGTGAGATAATATTCGGAGGTGAAACTATGCGTTTTCTGGGATATGCGTGCACCGTG
 GGTGGAACCATTTGCGTGGGCCAACTGGTCTAGTTCTAACACGAATCCGACGAGATATTCAGCCAT
 GGCAGGTTTCGACGGGCAGCGGAATACATGACGCACGCACCTCTCGGTTTCAATAATTCAGTGGGA
 GGAGTTGCGACGGAAATTAATACTTTCAACTATGTTTCTCCACGTTTATGGTTATGTACTTCTCA
 TTTCTTCTTGGATTCTTTATGTGGATTGGCCATTTATGGCATGCCGGACGTGCTCGAGCCGCGAG
 GAGCTGGGTTTTGAAAAGGTATTTTCGAGAGAAAACGAACCGGTTTTTATACATGAGGCCGTTAGAT
 TAGTCAAAGTTATATTTTGAACCTACTTCGATCAGTGAGATTTTCCACGCGACATAAAGTTTCAA
 ACTTGTGTAATAGCTAGCAAAGAGAATTGAACTCTTGGCCTACTAATTACAAATCAGTTGCTCTA
 CCGGTTGAGCTACGCCAGCGCAAGATAACAAGACCTTATCTTAAACAGTCTATTTTGAATATAAAA
 ATACCCTTGAACCTCATACAAAAAGTCTAAATCAAGATAATGAGTAAAGTTTTGTTAGCAAAGTA
 TCGATGATTTTTTATAACAAGAAGAATTTTTGGAAAGATTAAAAATCACCACCAGATTAATTTA
 TCTTTTTACTTGTAAACGGTGAAGGATCTATTATGATCATTTTTAGCTTACAATTGATTGTAAGGT
 TAAGCGTCCTCAAAACTCTT

psbD

GCCTATTTAGCCGTTGGTGGTTGGTTGACCGGTATTACTTTTTGTAACGTCATGGTTTACACATGG
 TCAGGTCTTCTTCTATTCCCATGCGCCTATTTAGCCGTTGGTGGTTGGTTGACCGGAAATACTTT
 TGTTTTCGTTGGGTGGTCAGGTCTTCTTCTATTCCCATGCGCCTATTTAGCCAGGTAATCTAACG
 TTAACAGAAGATCAGTCATGTGAAACGATGTTATAATCGAAATTTGGTAACTCTAAAGCAAAGTCC
 GCTCAAACGGCTATGCGCCGAAAAGTATCGTTATTTAGTCTTGTGATGATTGGTTAAAACGAGA
 TCGTTTTGTTTTCGTTGGGTGGTCAGGTCTTCTTCTATTCCCATGCGCCTATTTAGCCGTTGGTG
 GTTGGTTGACCGGTATTACTTTTTGTAACGTCATGGTTTACACATGGTCTGGCAAGTTCTTTCTG

GAAGGTTGTAACCTCCTTACGGCTGCCGTTTCCACGCCACCAAATTGTATGGGTCATGCCCTACT
ATTATTATGGGGTCCAGAATTATCCGGTGATATTATAAGATGGTCCAAGCAGGTGGTCTGTGGA
CGTTTATCGCGCTACATGGTTCATTCGGAGTCGTTGGTTTTTGCCTACGGCAATTTGAAATAGCG
CGGCTAGTCGGTATCCGACCTTACAATGCTATAGCTTTCTCGGGTCCAATAGCTGTCTTCCTAAC
GGTTTTCTTGTATCCGTTGGGACAGGCTAGCTGGTTCCTCGCACCTAGTTTCGGGGTTCGCTG
CTATTTTCCGATTCTTCTTTTCATACAAGGGTTCACAATTATACCCTCAATCCGTTCCATATG
ATGGGGTAGCTGGAATCTTGGGTGGAGCTTTATTGTGTGCAATTCATGGTGCCACCGTCGAAAA
TACGCTTTTTTGAAGATGGAGACGCTGCAACTACTTTCGGTGCTTTCACACCAACACAAGCTGAAG
AAACGTATTCTATGGTAACAGCCAATCGTTTTTGGTCCAAATATTCGGCGTTGCATTTTCTAAT
AAACGCTGGCTACACTTCTTTATGCTTTTTTGTACCAGTCGCTGGAATGTGGACAAGTTC AATAGG
TATAGTGGGGTTAGCCTTAAATTTAAGATCTTACGACTTCGTTTCACAGGAAGTTCGCTGCTGCC
ATGACCCTGAGTTCGAAACCTTCTATACAAAAACCAACTTCTAAATGAAGGAATCCGTGCCTGG
ATGTCTGCACAAGATCAACCACACGAAAACCTTTATATTCCCCGAAGAAGTACTGCCTCGAGGAAA
TGGTCTG TAA AAATTTTTTTTTCACCTTTTCCATCGCAATAAATACCACTTCAGTACCTTGCAAA
AAGGGGATGGCTCCGATTCTTCGAAGTTAAACCTGTGAGGAGTATGAAGATAATCAAAAATCACT
GAAGATATGTATCAAATAATTAATAAAAACCTTTTTATTGTACGGAAAAATATTCATATGACCTT
TCTTCGAGATGAACATTAGATGTAACTTTACCCCATGGTTATTTTACTTTTTAAATCGACCCTTT
AAACATAAAATAGTTTTTTCATTATGATGGAACATAAATGGTTCGATAAAAAACAGTATATTTCTAT
ATCATAGAGGAATAATTGTTACCAAAACATTTTCGTTAAATAGCTAGATAAATATACTTGTTC AAG
TCACTGACGTTGGTTGTTTGGCTTGATTACGTTGGGTTACGATTAATGGGATCTTAAAAC TAGAT
TAAATTTGCTCTGGTAAATGGCGTTTATTTTTTTTAGATAACGAAGCAAACACCTTACCTGTCCGT
AGAAAAAGATAATTATGTCAATTTTTTATTGTGTTTTATATCTCTGACGTTGTTACGAACCGACAT
TGCCAGTTATGTTAACATTCGTTTTATAGGTTTTTAAATAAATCATTTAGGATTAATATAGAATA
TTAGTAAAACAATTCACAAGCAGTGACTGTTTTAAACTGTGATTTTGGTATTCTTAGAAGCTTGA
TCTTCGAAACAAGTATTTTCGCAGAAGAACAATCCGTTTCTATGTAGTTGTTAAATTTACGAACG
CTGGAAATTAATTGTAGTGTATTACCTTATTGCTGGATATTCATACAAGATGTTTCATAATTTTTA
GACTCACAGCGATTAATTATCAACAGTTTATATATGTGTAAGATATTAGACGACTTTATCTCTTG
CGTGTCCATTGTAACACCGGACGATTTCCGAAATGTCTTTTATTGGGGTAATTTTACTAATCC
TTATTAATTTTATAAAGCGAATTACATTATTGTATTTAATCGTGAACGTA AAAATGTAGAAATTCG
CCGGTAAATAAATGCTTTTTTACGCTTTAAAACCTTTGGTATAGAAGAAGTTTTAAATTTCTAAAAT
TTACAATTTAATTATGTAATATTTTTTAGACTAACTTGACGGCTTTCGCAAATCCAATTTTTTTA
TTGACTTAGCCACCGACTTAAAAATGAATTTAATCCTAGATATAACATTCGTTATAGTTTTCTCA
ATTAGACTTCGGGATCTAGAATAGTGGAGAAACGCTGAGCCGAACCTTTACCGCCGAAACACTCAA
GAAAAATCTAATACAGCATTTTTCACGGTTTTATCGTAATGGGTTTTATTTCATATTAAGTTGAAC
CAAAGCGCTAAAACATAAATAAAAACCCACTTATTTATAATGATAAAAACCTCTGCATACATAATTA
ACTAATATAAAAATTAAGGTGTTTTTGTATAAAAAATAATAAAACTTATTTATGGTATAACGAAG
TCAACTTCTATGCGACATGAAAGTGTTTTTGTATGAATCTTTTTGGTGAGTTTGATCCTGGCTC
GGGATGAACGTTGGAGGCATGCCTTGAGACATGCAAGTTGGTTCGAAATCATTTCAAATTTAATT
TCCGTAGCGGACGGGTGCGTAACACGTGAGAATCAACCCTTAGGGACGAGGATAAAGGTCGGTGA
CGATTTCTAAAACCTATAAGCCGTGCAATTACCCATGTAGTTGTCCGTGAAGATTTCTTAGGG
ACGAGCTTGCGCCCTGATTAGTTAGTTGGCGAGGTAATGGCTCACCAAGACTATGATCAGTAGCTG
GTTTGTAGAGAGCGGTCAGCCACATTGGGATTGAGATACGGCCCAAACCTTCTTCGGGGGGCAGCAG
TCGGGAATTTTCCGCAATGGGCGCAAGCCTGACGGAGCGATGCCTTGTGGAGGATGAAGGATCAT
AAAGGTTTCGTAAACTCCTTTTTATTAAGGAACATTAATTGTCGTTTCCGTATTTAGATAGTACTTA
AAGAATCAAGTATCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACGGGGGATGCAAACGTTGTC
CAGTTGTATTGGGCGTAAAGCGTCCGTAGGTGGACTGGTAAGTTTATTGTTAAAGATTCTGGCTC
AACCAGGAAAAAGCAATGAAAACCTGCTGGACTAGAGAGTGGACGGGGTGAAGAGAATATCTAGTG
GAGCGGTAAAATGCGATGATATTAGCGGAACACCAGTGGCGAAAGCGCTTCACTAGACCATTTTC
TGACACTGAGGGACGAAAGCTAGGGGAGCAAAAAGGATTAGGTACCCTTGTAGTCCATCGCTGT
AAACGCTGGGTACTAGATGTCGCAGTGGTGCGGTGTCTTAGCTAACCGGTTAAGTACCCCGCCTG

GGGAGTACGCTCGCAAGGGTGAAACTCAAAGGAATTGACGGGACCCCTCGCAAGCGGTGGAGCAT
 GCTGTTTAATTCGATCAACGCGAAGAACCTTACCTAGATTTGACCTTCTATAAATCTCCATGAAA
 ATGGAGGAAACCAAACCTTACATGGTGTAAAAGCTTCTTAGGGAGCAGTAGTTACAGGTGGTGCAT
 GGCTGTCGTCACTTCGTGTCTGAGATGTCAGGTAAAGTCTAGTAACGAACGCAACCCTTATCC
 TTAGTTAATTATATGTCTAGGGAAACGGCCAACGATCAAGTTGGAGGAAGGAGAGGATGACGTCA
 AGTCAGCATGCCCCGTATGTCTAGGGCTACAAGCGTGCTACAATGACTATTCAACAAAAAGTCGC
 AACTTTGTGAAAGGAAGCGAATCTACAAATGAGGTCACAGTTCGGATTGTAGGCTGCAACTCGGC
 TACATGAAGTCGGAATCGCTAGTAATCGCTGGTCAGCTATACAGCGGTGAACTTGTTTCAGGGGT
 TTGTACACACCCGCCCGTACACCCCTGAAGTTGACTATACCAGGCGTCGTTGTCTTAACCACACA
 TATTTTAAAGACGGAAGACAACGATAAAGGTTGTGTCGGTAACCGAGGTGAAGTCGTAACAAGGT
 AGCCGTACTGGAAGGTGTGGCTGGAAATCTTTTTTAAAGAAAAAACATCTAATTAAGTTCTCTA
 TCCTATATGTGCGACGGCTCATTTTTTCGGCGGTTTTAGCATTTAATATGCACGCATTACACTGC
 GACAACCATCTCGAATTCTGGAGTGAAGATATATAAAGTTTCGTACTGCTCAAATGCTTACTA
 TCCTACAAAAGGATAAGCATTCTAGAAACAAAAGAAATGAAAGACTATATTACCCGATAAACAT
 TCGTCAACGTCAACAAACCTCATCTGCGTGGGAGGTTTTTTGCTCGTGGGTTACAACACAGAAA
 ACAGACTATATCTAGGGTGGTTTCGGTTGTCTAATGGTACCGACCTTAGTTTTAGCAACCTTTGCC
 TATATCATAGGTTTCATCGCGGCACCGCCTGTGACATAGACGGTATACGTGAACCTGTGAGAAC
 GAAGCAGAAGAGCAATCGTAGCACAAAGAATCTTCTGAAGGCCCCAGAAATCCATCCAGAGGGCC
 GAAA
 AA

psbE

CAACATTGTCTCGTATGTACTAGTTAACAAAAAGTTTTTTTTGCACAAACGTGCTAAACTCAAAG
 AGATATTTGTTAACGATGCTCACTCGGCTTATTAACAAAATAAATGCTATTTCAAACCTTTACTC
 AATTTTTTGTAAAGTAAAATTAATGATACCATCGACTTTTTTAGAAAACGAGTAAAGAAGACGTGC
 AACTTACGAAGAGTTGTGATTCATCTAAAGTTAAGCTGAATTAATCGATTGCTCGAGTATTAAA
 AAATTTATCCCTGGAATCAAATACTTAATTGGAATATATCGCATTGAGCGTAACATATTGATT
 CTTTTAAACCCATAACTTATTTCCGTTAGAATAGAAAATCTTTATGCGAAGAAAACATAAATC
 TAAGATTGAGAAGCGTAGATAATATAAAGAGGGTTTGTAACTCAGTGGATAGAGTAGCGGACTCC
 GATTCCGTTGGTTCGTAGGTTTCAATCCTTCCAACTCAGATCACATATTTTTAATACAATAGTAG
 ACTAAAATAGCAGTTGTTTATTTGTATTTATATGAAGTAATTTGTTTCATATTATAAAATTTCA
 CTACTGGACGTTCCACATCATATACTTTTCTACAAATATTTCTTTGTAAACCCCTTCGACGTTTCGT
 TTTCTACATTAAGACTCGCAAAACCCTTGTGAGTCGCGTGGGCTCTTTACTTATTTTACGAAA
 AATAAAGCGATTGAAAATTACAAAAGGCGAAAAAATTTATTTCTTTCGTGTGCTTTGATTACAAT
 ACATATCAAGCTGAGGCTCAAATGGAATATTTTCACTATAAACGTGTTGATTAGAAAATACTTA
 TGCCCTTCGCGGACTAACAACTTTTTTATGTTACGTTTCGTGCTAGTGTTTAAGTATAAACCGG
 CAAAGTGTAGTTATTATTAACGACTACACAATTCCTTACCTTTTAAAAGAATCTCGTTCTAATTAG
 GCGATAAAATACGTTTTTTTTCGACCTCGTAACCGTTGTTTAAAGAAAGGGCATTTACAAGTATTC
 TAAATAAGGTGTTGCTTGTGAATTAATTTATAGTATGGCCGGTCCACAGGAGAACGACCGTTC
 ACAGATATTATAACCAGTGTACGCTATTGGATCATTCATAGTGTGACCATTCCATCACTTTTCGT
 CTCAGTTTTTTTTATTTGTATTAACCGGTTTAGCTTACGATGCATTTGGAACCCCTCGACCTTCTG
 AATATTACACCGAAAAGCGTAATGCTGCACCAAGATTAACCTGAGCGTTATGGTGCAAAGCGGAG
 TTAACAGCCATCGCGGCGGTATGCAAGAAGAAATAATGAAATCTTACGTTTTACTACGAACATC
 TTTTGTGAAGTTTAGACCAACTGAAAATAGTGCTGAATTAAGTAGGAAATATTTTCGTAAAATTAC
 TCTTAGTGCTATTACCAGAAGATTAATTTTATCAAAGTTATAACCATTGAATTTATTATTATACA
 CCAACTGTGTCTTTTGAATAATA

psbF

AAAAAAAAAAAAAAAAAAAAAAGTCTAAAGCCAGTGACCGGAGCTGTGACATATCCCATATTTACTTT
 TCGATGGTTGACAATTCATGCTCTTGCAGTCCCCACGGTATTCTTCCTTGGTGCGATCACTTGCA
 TGCAATTTATACAGAGAGCGTAAAGTACGGTGTTTAGATAACTAAACACTTCGAAAATTCGTGAT
 ACCCAAGTTCATACTTTCTAGTACCTTAAATTCTCGTCCCATGTTACAAAAAATCCCTATGAAG
 GAGAACCTGTAGAAGTGAATCGAACGTCTCTATTTGGGGATTGTTACTTTTCTTTGTACTAGCG
 GTTTTATTTTCAAGTTATTTATTAATTAGGAGTGACTTGTAATTTTTTTTTTTTTTTTTTTTTT
 TTTTT

psbH

AAA
 AAA
 AAA
 AAA
 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCCATTTTTGTACAAAGTCTTCTCTTATCAGCAAACGGC
 CGCAACGGGAAGAGGTTTAAAAACAAAGTTAGGGAGAGCCCTGTTGCCTCTTAATTCTCAAGCGG
 GCTTGGTAACGGCGGGGTGGGGTCAACGGGATTCATGGTATTTACTATGGTGGCATTCCCTATTG
 TTTTAAACAATTATTTACAATTATATAATTCTACAATATTATTGGATGATTTTCGAGCTTTCTTG
 GGCTGATTTACCAACAGGGTAAAGTTCTTGATTTCCGAATGTTTCAGTGTTTAATTATTGAAGGTT
 ATGGGAGTTAAATTTGGCATATGATCAAACATCATTAGGCAATCTAAGTGTGAACAATTTAGT
 CGTAGAGTACACTCAATAAGGATAATTCAGCAACTTATCTAATTTTTTAAACATATTTTACTTAG
 TAATAAAAAACACTGGTTGGAATTAATTTATCTTATGACCACGACATATGACAAATGTGATA
 TCAGTTGGTTAAATAATAACCCAAAACCAATGATGTTAGGCCTTACATTTAAACTCGTCTCGTA
 GTAAGATCATGTGAAGTTTTTCCGCATGATAATTTACTTTTATAAAATCCATCTTCAATTATTT
 TTGATTATCTCCATACTCCTCTACTTTTTCTTTTTAAACTTCTCGAAGAGCCTCGACAAGGAGCG
 TTAACATCGCCTTTTGGCAAGTTACTGAAGCGGTGTTTATTGTAATGGGAAAGGTAAATCCGCAA
 GGGTTTTCATCAGTATTGATGAAGCCTACTGTCCGAATTTTAAAGACGCTTACATTCTCTAAAGC
 TGGTTGGTTGACTTCGAGTTAAAAAAATTACCATTTTTGGTAAATAGGGGGTCAAGGGGCTG
 AAGACTCGCGGGTTCCCTCAATTCCTCAAATAAGCGGTTAAGTCGTTGAAATCTTCGACGAAAT
 TTAATAACGACTCTTCGATTCATTCTGGGGAATGGTTTCGAGTGACGACCACTGAGGCTTGTGCG
 TAAGAACTTAATTGGGTAGCACCCCTGAGAAATTTGGTTAGCGTTCGTTAATATCCCAGCAAGCC
 ACCGGAATTAACGAAACGAGATCCACACGCAATTGCAGCTTTTCGGGTTAAGTTCGCTGGTTTA
 ATAGATTTTCTAGGAGTAGGTTGCAAACCTTAAAGCGAAATACTTCCTTCCATAAAGGCTTCCTT
 TGAGGACCAAGATTTTTGAGGGACGAAAGCTAAATGCCTCCAGGAGCGTTTTGGTTTTAAACTTG
 GAACTGTTTTGGCTTTCGTCAACAGTTCGAGTGGGTGTAACCTCGAGTATCTACTAGAAGGACATCA
 GCTGCTTGGTAAGCACCTGTTTCTCTCTGAAGTCGCAAGAAAGTAAGAAGGCTGGATAATTCACT
 AGCCGAACGACGATAATCAATAATTAATCGTTCCCTTCTTATAACTTTGACTAGCTACATGCTTTA
 TCCATCCCTCAGAAGAATCTGACGGGCACCTAGCAAGAACTCTGTCATACATTTGAGGGTGTAGC
 TGACATCGTAAGGTACCCAGGTGACCACGTGTGTTAATAAACTTTCTCAAGGTGACGTACTTC
 ACGACTACGCGTGACCGATTGCTCAAGTCTTTTTTAAATCCTGGTAAAGATTGAGATGCGACTA
 CCTTTTCAGAGGGGAGTGACATATTTATATCAGCTCTTTTTGGAGAAACCCACTTTCGAAAATTA
 AGTACAGTCATAAAGTTATAAAGATGAGATAACAGCCTAATCGAAAAAATCACAAATTAATAAT
 AAGTTTTAAAGAAGGCGGAGTAAACTCCGCAATCAAGTTTTTGTTCACCCGAGGTTTTTTGGCA
 TCACGAAGAGTGCCACACGATACGATTCAAATTTATCCGAGTCTTCTTTTTGGAGACGATGCCGC

GATTTGTGACTGGCGAGCTTTTGTACCCCTCTGTTCCACAATCGCCTTTTTTATTTCCACAGTAT
 CTGTTGTTTGGCACGAATAACTGGGGATAGTTACAGGTACGCTATTAACCTTTCACATGTCCGTGA
 CTCACTATTTGACGAGCCGCACGAATTGATGTTCGTAGATAAAGCTCGATAAACCACGGTATCTAA
 ACGACGGTCTAAAGCTTCAAGTACAACATCTATTATGGAAGATCCTGCTTCTTTGAGCATCGTTG
 AAGACCGAGAAGAAGTAGCCTTTACATAAGCTCTTAGTTGCCGTTCCTCGTAGACCGTAATTATAC
 AAAAGTTGTTGTTTCGAAAGTAAACGTTTTCCGGCTAAAGTACGTTTCCTCGTGCTGCTGCTCC
 TCGTCTCCTCCTCCTTTTTGTTGGTCCGGCGCGATATCCATGCTCTCCTGGGACGTATGTACGCC
 GCGTTTGTTCCTTACAGAACCCTGGTAAAGGCCCTAATCTT

psbI

TCGATCAAAGTACATAATACCAGCACCAAGGTAAGCTGTGGTCCAACCAGAATCAGTCTTCATTT
 GAAATATACCTCCTCCTTCGGGAGTAGCTCCAATCCAGCACTGGTGCGCGGAAGCTAGGGTCTGT
 CTACCCCGGCCGGCACATGGGACGTCAAAGAGCGATGCTTGAAACTCGAGGGCAGCCGTATCTTC
 CCATCCTTCGGGAACCCAAACTCGAGCGGTAACGGTCATGTAGTTTATTATAGGGATATAATAGC
 CTGGGATTTCTCACTAATATTTAAAAGTTTTCAAATAATTAGACTTTACGAGATCCTAAGATAAATG
 GATAAAGCATGGATTTTTGCCTGCAATAAAGGCGTCTTTTTGTCTTTTTGGGCCCGCCTAAGAATCG
 AACCTAGGCCCCACCCTTGTCAAGGGTGTGCTTAACCACTGAGCTAGCGGCCGAAAGCCATA
 AAAAAATAACAAATTTTCACGAATAAGGGAAAAGGTGAAAAATTGTATTGAATCATATCAAGAAA
 AAAAAAAAAAAAAAAAAAACGTATAATAAGGCCCAAATGATCGCTTAAAGTTAGCTGTTTAT
 CTGGTTGTGATTTTCTTCTTAACTATTTTTACCCTAGGTTTCTTAAACAGGTGATCCATCTCGAAA
 CCCGAATAGAAAAGATATAGAGTAAAGGAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
 TTTTTTTTTTTTTT

psbN - no obvious start codon

TGCATGTTAACGAATATTACACTTAAAAGTCTCCAGAACGAAAACCTCAGTCAAACCTCCTCTA
 CTCTAATACTGTTCTAAAATCTTGTCTACGAGCACCAACTCCGCACCGGGGTCACCCACGCCCA
 CCAAGGTAGGAGCACAGCCTCTTGACAGACAGAATCTTACTCTTCTTCTGTCACTCCTGCCTCC
 TTATCGTTGGTGCCAACCTCTGGGAGATTCTCTCGCGCCAGCTGCAGGTCTTTGTTGTTCATCGTT
 ATCAGAAGGTGGTTACGGACAAGATGAATCGGAAAATACTTCTTTTCGGAACCAAGTAGAGTCAAAG
 GAATTTCTTCATTCTTGGCACGTCGCTTCTTGCGCAATATCTTCTTCTTCTCAAAAAACGACAA
 TTCGACTCTGTACAATTGCGTGCCGCAACGTACTAGTTTTGTGGTTCGTGTGAAAATATCCAATTT
 TGTTCAAGTTTTTATAAGAAAATTGGTTACTATTGCGATTTTATCCCGTACTTGCAGTCTTTAAA
 TATTTAAACGCTTATTCTTTTCATTTTCGCTGAGAGAAAGTTCACTTGTAATATTAGTCAAAGTTT
 ATTTTAGGGTTGTGCGCGTGCCTAAGTTATCATGCTATTACGTAGGAACGTTGGTAAGGTTGAAA
 CTTTTTGCAGATTTTAGTAAGTTGGAAGCTGCCGCTTTTGTAAACATACTTTGTCCTAGGCCTTTT
 AGTAGGTATCACAGGATATCCATATATACTGCTTTTGGAGCTGGTTCTTCAAATCTTCGCGATC
 CTTTTGAGGAACATGAGACAACGGAGGCTATTACATTACATACCCCTCGAATTTTTTACCC
 AGCCCTTTACTTTAATATTATGGTAATAGGCCATAACATGAGTCTTTTCCGATAAAAATTATGTA
 GAATGTGGGAATAAGATTTTTAACTTTGCCGATCTTTTAAAAGCTACCGACGCACTAAAACGTGT
 ATAAAATTATCACATTAGTCGTATATTTGTCAATTGTCGTACCTACTCACCTTCGTCGATGTTA
 TCAGTGTCTCACAATAGCTCGATTTTCTTAAAAGGTACTTCAGTAAGCTTTATTTACAACAAAA

GCTTAGAAGTTTGGTCGTGCTAATTTTATAAGATGGCTGTAAAACACTTTTTTATTTCAACAAGGA
 AAAAGTAAACAGAAACCTCAATTATTATACGTGCACTTCGTAGTAAACTACCCAGTATTCTCACG
 TATATACATAATAACAAAATAGAAGTCTTCATGCTTCATTTTGTAAATACGTTTTTAGCATATTT
 GA

psbT

GTTAATTTTCATAATTTTCAGAGTTATTTTATTTGATATCTTAGTAGTTTATGGAGGGGTAACGTCT
 GTTGCGCCATTTAAGAGATTGCTAAGTTAAAATTATTAACAATATTAGAGTAAGCAATGAAAAAG
 TTTGTAGTTTTCTAAGCACACCATCCCTCGAGATTGTAGTTCAATTGGTTGGAACGTCRCCCTGT
 CACGGCGAAGGTTGCGGGTTCGAGTCCCGTCAATCTCGATAGTAAGGTGACGTTCCAACCAATTG
 AACTACAATCTCGAGGGATGGTGTGCTTAGAAAAATWAAAAATTATTC AATTCTAATAAAAATTCT
 TGAGATTCGGGTCCTTTTTTATCTTTTATTTTAAAATTCTTAGTTTCCATAAAAAGGTTTTATAAA
 ACCCAATAATAATGGTAAAATTTAGTTCATTGGAAACATATCTATAGCTTTATAACGTGAATAAT
 TTTATTATCATAAAATAACACGTACATGGGAGTAGTTGTAAAAGTATTTATTACTCAAATAACAAA
 ATTTTTACTCTGTGCAAAAACCTTCAGAACTAAATTTAATGCTCTGAAAATTATAAGAATACGGGG
 CGTAACAAGACTAGATTTTTTCGATTTGATTAATGTCTTGAATAGCACTGCATATTACTAATAAAA
 TGAGGGTAAAATTTATATTTGTAGCCAGTTTAGTTTATAAAAACAAGTTTTTGGTATTGTATTTA
 ACGTTTTACTATCGTCTAAGTTCGATCGATTATATTAATAATATAATTGGCGATGTTATGAACTT
 TTGTACTCGTAAAGAGAGGGTAGTTCTCTATTCTACGGATATGAGGCCATTCTGTATACAGTTC
 TATTGATTGGCACTCTATTTGTTATATTTTTTTGCAATTTTTTTTTAGAGATCCGCCACGTCTTGT
 CGAAAAATAACGATTTTGGAGTTATTCATTGGTATTTATGCAACGTAAACACTAAGAGAACATT
 GCTAGAAGGTCTATTAACATTTTTGAGAGCGATCATGGTAACTTTGATTTTGGAGATAAAAATC
 TTTTCATACGACTTCAGAGAATAAATTTGAATTTTTTCGCCACTCTCTATATGGTTCAAATTAGTA
 GAAAAATTACGCTTTTTTCATGTATGGTTTTCAAATAAGATAATATATGATTATACCATAAAACAT
 ATTTCTTTTTCGACTGGAAGGTAATTTTACGCTAAAAAATTTATCTTCTACAATTTCCCTATCAG
 TTAATAAAGATCCTATTCTCTCAAAAATTCAGAAACACTTAAAAATAGACTGCTGGTAAA

psbV

AA
 AATTGTTACTAAAGCTTCA
 AAATGCCTAAAAAACAGATTTTTTCTAGGCAAGTAAGAACAACAACAAGTCCAAGCATCTCTT
 CTTTTTTCATTAGTTTTCTTAGCTTCCCCTGAGTTTGCATTTGCTCGCAAATTCACAGACTCGGT
 CAGAACCGTACAACCTGAATGCAGCGCGCGCAGATGTAGTACTATCACCAATCAAGTACGTCGAG
 GAAAGTTTCTATTCGAAAAGCTTGTGCGTCTTGTATGTAGGTGGGTTAACAAAACCTAATCCA
 AACGTGGGGTTAGATATCAAATCGTTACAAGTTGCACGACCTCCTAAGAACAATGTCGCCAATTT
 AATCGCATATATCATTGCACCGACCACATACGATGGATTAAGTATATCAGATATTCATCCTT
 GCACAAAACAATCCAGATTGTACACAAAGGTTAGGTCTCTAACTCGATTCGACTGTTTCTGTATT
 GGTGGTCACGTTCTACTACAAGCAAACTATTGGGAGAAAATGGGGTGGAGGAAAAGTTTATTA
 CTGA ACTTTCTCAGAATAAGTAATAAATTCGCTATTTTCATAGTGAAGGAAATCTTTATAAATA
 TAGAGCATGTCATTTCTTTTAGTGTAATTTTATCCAAAACAAGTTTTTAATTCTTCGATAAAAAG
 ATATGTCATAGGCAGTATAACGTTGAGGTCAATCTTATAGTTTTTTTTTACCTACTGAAAAATTAG
 AGTAAAGTATACAACGTTTTTTTATAGGTTAAGTTTGAACAATCATACTAAGAAATTTTTTTGG
 AATAATAAATTAACGAAGTTTTTTTAGATTAAGATTTTTTCACATAAACTAAAAAAGTAGTAT

TTAATTTTATGGCGACAAAGGGAAAAATTGAAACTGAGATGCGTCGAGAACGCTTAATAATTAGT
TATTACAACAAAAGAGAAAAGCTACGCGAGGCCGTAAAGTAAATAGACGTAATTTTAGACAAGG
TAAAGCTCCAACGCTCGTCCAACATGTTGAGTTATTGAAAAAATTCACAGTTTACCGAGGAATT
CTGCCCCACTCGGCATCGAAATCGCTGTTGGGCAACGGGTGGAAGTCGAGGGTTTTATAGAGAT
TTCGGGTTATCAAGACATGTTTTACGTGAAATGGGAAATGAGGGATTAATTCCTGGGCTTAAGAG
ATCAAGTTGGTAAAAACCTGTACGTAAAAGTACTTCGATAGAAAAAGTCGAGTAAATGTGTTT
AAAAATGTAGAATCTCCTGATTAAGTAGAAACTTTCTTGAAGCCATAAAATAATTACATCGAGAA
TAAAAACGCAGTTAGACACTTACTATTTTTCGACGACGACTCGTTCCATATCTTTTCGATTACTTT
TACACTTTCAAGATAAAGAGTTTTCTATCCGTTAAGATTATTATAGATTTTCATTATTTCTAGCCCTC
ATCTCTAATATTTTTATCTTCCTTGAGAACGGTCATGGTTAATCAAATTTTTAGTCACTTGGGTC
ATGAGTCAATTAAAAAAGTTTTAGAACGACAATATCTTTTTTTGGTTTTCTGGACCCATGGCACATA
GGTTATAGGTTTTTAGATAAAAGTCGAGAAAATACCCTATAAATAAATAAATTGGAAATCTCAA
TTGACTTGTA AAAATCATTGATTGTTTCAGATAAGATAACTTTATATTTATCGTAATATCTCCTT
TGTTTTATTAGATAAAGGATTTAATCAATAGATTTTAAAATCAAAAAAACAGGTCGCTTCTTG
TTAATGCGTTGTAAAAGAATACTATTTTAGTAAACTGCGATAGCTAACGAGCTTTTAAGAAGAAT
TCGTATAAGTAAGTTAATTAATAATGTGATAAACCATCGCATTAACTGGATACTTTTTTAATG
GAAATATTTAGCGTGAAGAAAGTTTTTAATACAGATTGTTTTTTCATTACTAAAGAATAAAAGT
AATCCTTTAATTTTATACGAACATTCCTTCTGTGCTTGTGACCCACGAGCGTAAAGTCTCGCAC
TCCAATGAAGTTTGTGCAAGTTGAATAATGGTTAGTGATGCTTTCTGTGTGTTAGAAATTA
ACGATAACGTGTGTTTATAGTAGTTATCTAAGTATACTTTATGTGTCCCAAACCTCGTCATTACTA
ACAACAATTTTATACCATTTTCGTATTTTCGAAAGAATACTTGGCGGAAGCGGAAGACTTAGTCCCG
ATAATTTTTGGCGCTCGATTACACACAAAATTTGGATATAGAGGTGGGCTACTACGCACTCGTTT
AAGGATTGTTGCTCTAGTCAAACCTTCCACCTGTATTAGTACGCGAACCTCCTTTACCACTAAG
CCTAGTACACTTTGGGGCCTTAATTTAACGGTTTAGGCTGTTTCCCATTTGACGATGAAGCTTAT
CCCACGCCGTCTCACTATTCTATATTATCTATAGTATTCTAAGTTTACCTAGACCTCGAATAGAA
TTTTACAACCTGACAAAAATTCACCCGTTATTTTTTCAATAATGAACGATTGGTATCGCCGTTTAG
TAAGAATTAATAATTCATCGCCAAACGATTCTTTTCGTCTTAAATGTTACTGTACATATTGTGCAAG
ATAACCCGAATAATATTTTTTTGGGCGTCTCTATCAACTCGCAATATTTTACTTGTATAGTAAGGA
TAAGAGTTCTAAATTAAGATCTTTGAGTAAGATCCTGAAGTTAGTAGTTATCTTTCCATTTTT
GGCAAAGCTAAACTCTAGGTATTCGTGTAAGGATCGGCCACAAATGTGAACATCCTTATTACCT
CATTATATTTAGCCAAAGAGCCAAGATTTTTAAATCCACTCTTTTCGGCCTAAATGGCGGTCTAC
CCGCCCAGAACCTCAAAAACCTCTCGTAGATGTTGGAAGATACGTCGTGCCGGCCTTAAAATTTTT
CCAACAAGCCCACCGAAATTTCCAATTCCTATGCTAAGTGATAACAATGCCCTTATTTTCGGCTTG
GATAATAGCTTCTCATAAGAGTATGATACTTAAATCCCTTATTACTAAATTATAAAAAAAGAGT
TGTTAATTAGTAAAAAATGTTTCACATAGCAATCTAAATAGGAAAAGATTTGAATCTGAAAATCCT
ACAGTAGCTATTCTCAAGTTTGGATAAATTAATAACACACACAGCTGTCAACTTTAATTTTTCTG
ATTTAATCGGAATTCATAGTGATTTTTTTTATTAATTACCGTTGTAAAAGTGTTTCAATTGATAAT
CCTATATGTTATCCTCAATGTGACGTCTCATGCTTCGTATAGTAGTGGGTACGCACTAGAAGAAA
TACAACAATTTTCTTTAATTTTACGAATGTGATTTATGTTTCAATTACTGCTGCAGCAGTTACGCC
TCACTGAAGGCTTTTTTTACAAAGTATTCTTACTGCCACCCTCCTTGTTGTATTACCTATTGCAGC
CGCATTAACCGCGGTTAGTTCTCTCGATAAAAATGCGAGATCATAAATGTGTCAAGTACACACCT
TTTTTCTTGAGCGTCTAATATTAGTAATTTATTGATTTCTCTAACCCAGCTTGAGATGTACGTTAC
TAAAAATTAAGTGTCTTCAACTTTTAATAACCAGCACTAAAAGTGTAAATAGATTTATTTCTTTAT
TTGTGCATTGCGCACATCGAGAGAAATTTTCCAAATACTAACCCGTTGTAAATTCGCAAAAGTTTA
AAATATATTTTACTATAACAATTAATACAAATTCAGCTTTCTTAACTATTATCAATTTCTGAAA

T

psbZ

CAAGATTGGMWYMAGTCAAGGATATACTGTTTTTCAGTTGGTTCGTCAAAGCTAGCCCCAATTCTTG
 CGTACTTTGTTAACCATGCGAGCAACTGCCTTCGTTCTGAGCTGCATAGCTTGCGCTCAAGCTCG
 CCAGAGCTCGCCAAGTCCGTTGCAATCGCTTGCAACACTGCTTCAGGCATCCAGTCTGAAGCTG
 GATGGCAACTGAGCGGTGCTGGCGCTGCTGCGAACAGTCAGGCTCCCCGGCAAGTGCCTTCGGTG
 GTTGCAATGAATGGCATTGGCAAGGAAGATGAATCTCGTGTATCAAGGCGTTCAGCAGTGGCGCA
 ACTTGCAACTGTAGCTGGTATGGCAAGCATGCTTGGTGCAATGCCAGCACTAGCCAAGGATGATG
 ATGACGATGACAGCGGCGTGAAGGCTCCTCCTCTCAGGACTGATGCCGTTTTGCAGGCGCGAGAC
 AAGTACGAATTCGGAAGGAAATACAAAGCCGACACGAAGCTCATGATTGACAACATGAAGATCGT
 TACAGACATGGGACGAGGCACTCCGCAGATGGAGGACATTGCAAAAAGCCACGCGAAAAGCAGATGA
 TTGAATACATCTCTCTCTACCGCGCGAGACCAGAGACGACCAAGTTGCCGTCGTTCTCTGTCATG
 TACACAGCCATCAACACCGTCGCCGGGCACTACGCATCGTATGGATACAAGTACCCAATTCCAGA
 GAAGCGCAGGGTGCAGACTTGCAGAGCAATACTTGGAGGTTGAGCGCGCGCTCGCTCGTGGCAAGT
 AAACAGCTGACCATATGAAAACAAAATTTCTCTCGACAATTTGAATCCTATGCCAGTTACATGATC
 TTTTCCGAAATGTTGA

rbcL

AA
 AAAATTCACGCTACGGCTATAATAGATTTTTATGTTTACAACGTTGCCTAAAAATTTGTCCACTT
 ATCAGATTATATAGATGTAAATTACAAAATAGAGAGAATGTATTATATGTCCTAAAGTTATGTTA
 ATAAAAAATTTTTCTGAGGACGTGAACTCACGAACTCGAATAAAAGCCGGACGATATAAATCAGG
 TGTACTTGCCTACAAAGACATGGGATATTGGGAACCTGACTATTCTATAAAAACAACAGATGTTT
 TAGCAACATTCCGATTCCAACCGCAGCCGAGTGTACACCCAATTGAATTAGCTGCTGCATTAGCA
 GGAGAATCTTCTACGGCTACATGGACCGTAGTTTGGACAGATCTTTAACC CGGAGTGATATTTA
 CCGTGCCAAAGCCTATAAAGTGATAAGTCTCCAAGATTCTTCAGCACTAGCTTACATTGCATATG
 ATCTAGATCTTTTTGAGGAAGGATCTATTGCAAATTTAACAGCCTCCATAAATTGGTAATATTTTT
 GGTTCAAAGCTGTAAAGTCTCTTCGGCTTGAGGATATGAGAATCCCGGTGGCTTATCTAAAAAC
 GTTCCAAGGTCCAGCTACTGGTTTAGTCGTTGAACGAGAACGCTAGATAAGTTCCGGTCGACCTC
 TTTTGGGAGCGACAGTTAAACCAAAAATTAGGTCTTTCGGTAAAAATTATGGAAGAGTTGTATAC
 GAAGGGTTAAAAGGGGGATTAGATTTCTTAAAGATGATGAAAATATTAATTTCGCAGCCTTTTCAT
 GCGGTATCGTGAACGTTTCTATACTCTATGGAAGGTGTCAATCATGCTGCCGCAAAAACCGGTG
 AAATCAAAGGACATTATTTAAATGTCACGGCAGCGACGATGGAAGACATGTATGAAAGAGCGAAT
 TTCGCTACGGAAC TAGGTTCCGTAATTGTAATGATAGATCTTGTAATCGGTTATACAGCGATACA
 ATCAATGGCTTACTGGGCTCGCAGTAATGATACTCTGCTTCACTTGCATCGTGCCGTAATTCTA
 CGTATTCTCGACAAAAAATCATGGTATGAATTTCCGCGTCATCTGTAATGGATGCGAATGGCT
 GGTGTTGATCATATTCATGCCGGAACCGTTGTGGGTAAACTAGAAGGAGATCCAAAAATGATTAA
 AGGGTCTATGAAACGTTATTGGGTCTGTCTACACCGGTTAACCTTAGTCTCGGGTTATTCTTTG
 CACAAGAATGGGCTAGTTTACGAAAGTGTGTACCTGTGCTAGTGGAGGAATTCACGCAGGACAA
 ATGCACTACTTATTACACTATTTGGGTGATGACGTTATTCTCCAGTTCGGAGGTGGAACAATAGG
 TCATCCTGATGGTATTCAGCTGGTGCAACAGCTAATCGAGTAGCGCTAGAAGCAATGGTTGCTG
 CGCGAAACGTCGGCTAGACTACGTTGCGCTTGGTAAAAGCATATTAATAATGCTGCAAAGCGG
 TGCGATCCATTACACGCTGCTATTTTCAGTTTGGAAAGGATATAGCCTTTAATTATGAATCAACAGA
 CACACCAGATGTAGGAATAGAAGCGGCTCATTTCGTCATCATCGACTTCTCAACATGCTTCTCAGC
 ACGATGTTGCCCTTCAATAAGTTCAAACCCCTTTACTTTTTTTTTTTTTTTTTTTTTTTTTTTT
 TTTTTTTTTT

rbcS

AAATGGCACACCGCCA
 TAAAAAAAAAGGCATTAGTAGCACCGTCATTATCATCATCAGTGTTTTTTGATATGGGGGATTTTAG
 TTGTGTATGTATATATATACATATATACATACATGTGGAAATATTTAATATGTAGCCAGGATTTCG
 CTAGTCGTTTTGTTTTGAAGTTGTACGAGTGTGTTATAGAAGTACTAGATTACTACGTAGGAACGCTGA
 TAAACTGGAACCTTTTACGATTCTTAAGGTTGTAACCTACCATACTTTTTTTGTTCTCCCTTAATT
 TAACTCGAGGTGACAGATATCTCTGTAATACCCTAAGAGGATAAGTATGTACGATGTACAAAATT
 TAGCTTCCGATGATTACACAAGGATGTTATTCTTATCTGCCTCCTCTTACGCCTCAGCAATTGGT
 AATGCAAATCGACAGTGCATTAACGAGGACTGGGTCTTGGTATCGAATACACTTATGATCCGC
 ATCCTAGAAATAGCTATTGGGAAATGTGGGAAACCCACGTTTACATTTGAAGGGGTTCATCG
 ATAGTTTATGAGCTGGCAGCTTGCCTAAAGCACACCCCGAAGCCTATATCAAACCTACTAGCTTT
 TGATCAAGCAAGAGGGACAGAGTCGTGTGTCTTGTCTTTCATCTATCATCGCCCATCTGTAGAAA
 CGGCATTTAGATTAGTTAGACAGGAAGCTAACTCAAGAAAAATTGTTTATACCTTAGAACCTACA
 ACTCCTTAAAGATT
 TT

rp12

AA
 AAATTTAAAAATCGTAAATGCTCGATCTT
 TAAGATTAGCTTCATCTGGATCTTCAGCACTACTTGAAGTTATTAAGTATCCTCTTTTCTCAGAA
 AAAGTCTATCGCTTAATCGAATCTAATCAATATTGTTTTGCAGTAGACGTGTATGCGACAAAGCC
 AGACATCAAATCTGCTGTGCAAAATCTTTTCGCTGTGAAGTTGATTCTGTGGCCACCGCACTTC
 CACCACCCAAAAAGTTCGGCCGTGCTTTGGCTAATGGGAGAAAATCAGCTTATAAACGAGCTTAC
 GTAAAATCTCTCCCGAAATAGTATATCGCTTATCACTGATTCTTAAATTTTCCCCTCTGCT
 TATTAAATTGAAATAATTAAATTTATGCTATAAAATACCACAATAATAAAATTGTGCGCAACTCAT
 GACTTAAATAACTTTAAGTCTTATAATTATTTTTGCTATTAGACGTAATTCATAATTAATATCAG
 CTAGAAATTTACGATACAAAATAAATAGTCGAGTACAACATAAATCCATTTCAACGTCGTTT
 TAAATAAATGATTTTCATGAATATTATCCTCAAGAATATATTTGAGTTGTTGTGAAAAACAGACA
 TGACTATACTTATGTTTTAATTTAATATGTGATCATGTACAAATATTCTATTTTTCAGATAAAA
 AATCATAAATGAAGCTATTTACCAAGATTAAAGTTATGTCTGACACATTTTATTAAAAAATGTA
 CATACTAGATTAAGGGTTTCCATACCTGTTTCGGTCCAGATAATTTAATAAAAAATCACCTCAA
 AATTCTTTTATGAAGTGAGGGTACAACTTCTACCTCTTAGCGCTAACTAAGTAAAAAGGACGA
 ATGCGGTTTTGAGCTACGTATCACCCCGTAGCCCGGAGTACGTCATTATAACCAGATTAAGCTT
 TGAAAATCTTTCTCAAAGCTCCTTTAAAAAGTTTAACTCGAGGTCTAGCGAAGACTGGCGGCC
 GAAATACTAGAGGCGTCATAACGGTACGAGGGCGTGGTGGTGGTGCAGAAAAAGTGTGCACTT
 ATAAACACCGTACGAAACAAGACGAAATTTTCAAGTGGACGCCTTCCACTAAAACAGTCCCTAAA
 TGCAAAGGTTATGTCACTAGAAATACGCTCCTCACAGAAATACAACCTCTCGCCTTAATTTTCTACG
 AAGATGGTCAACGGGAGTACATTATAGCACACCGTGCATTAAGGAGGGCTCGTCTTTTCTCTG
 GGACCTGGTTCAAGTCGTGCGATTGGTACAACCTCGGCCTATGGGCCATTACCAGCTGGTTGCGT
 TTTGCATAATGTAGAATTGAATGCGGGTGCCGGTGGTAAAATTGCACGGTCCGCAGGGACTTTTG
 CGCAATTAATTGCGAAAACGGCTTCTCATGTAACGCTGCGATTACCATCAGGTGAAGTTTCGTCTA
 TTATCAAACATTTGTCAAGCAACACTAGGGCAGGTGTGTAATTTTCAATGTGCAAAATAGATCACT
 GGGAAAATCCGGTAGAAAAGATGGTTAGGAAGAAGACCTTCAAGTTCGTGGTCTTGCATGAATC
 CTAACGACCATCCTCATGGAGGTGGTGAAGGGCGTTCACCCATTGGACGAAAAGCCCCCGTAACT
 CCTTGGGGTAAAATTGCGTTAGGTCTAAAACCTCGAAACACCCGTAAAAAAGTACCATGTACAT

ATTGCGTTCGCGTCATTCAAAAAC**TGA**TACTCCACAAAATGGGTGGGTTTTTTTTTTTTTTTTTT
 TT
 TTTTTTTTTTTTTTT

Rp13

AA**ATG**CCTCTCACCGCGGTAGGTTTA
 TTTGGAAAGAAGCTGGCAATGACACAATATGCATGCGATCGTTCACATTCTATAGCTACTTTTGT
 TCCCGCGACAGCGGTTGAAATTGGAGCTAACCAAGTTACGCAGATAAAAACGTGCAAACCTCTG
 GATATGGAGCGCTCAAGTTGCGTACAATGAAGTAAAGCCACGAGTGCTAAGCAAATCTTGTTTT
 GGTAAACTACGCAAGGCTGGGTCGAGTCGATTCTTTAAAAACTTTGGTGAGTTTCGTGTGGCACA
 ACCGAAAAGTATAGGATTGGAGAAATGTTTAGACGTAAGGCTTTTTTCATCCTGGGCAACCGGTAA
 CCGTCGTTGCGCCAGGGTCAGGGAAAGGATTTTTGGGTACTGTAAAACGACACAACCTTCAGTAGA
 GGTCCAATGACTCATGGTTCAAAAAATCATCGCTTACCTGGTAGTATTGGCTCGGGTACTACTCC
 GGGACGTGTCTATCTGGAAAGAAAATGGCTGGTCTTCTCTGGGCGCTCACGGGTACGATCAA
 CAGTCTTAGAAACAGATGAAAATAATAACATCGTTCCTTTACGTGGTTCCGCGCCGGGGAAGCGC
 GGACACTTTCTTAAGATAAGTGCAGAAAA**TAG**TAAAGTTTTTCCAATTTTTCTAAATTTAGAA
 GAATTAATTACGGTCATARATTAACGTTCTAAGTAGAAAATTGATAATAAACTTTTGCTTAATT
 TTGTAGACGTACGTTTCATGAGTACAAATTAACGTAAGCGAGTTTTGAGGGAACCTTAC
 ATTTAGCTGACTAAAATATTAATAACGTGGTGGTATATTTTACTACGTAAAACACGTACTCTATT
 TACAAATCGTAGTGATAATATAAAAACGGACTACTCGCTATCTAATTTATCTAAAGTTTTTTAGT
 ATGTTGACAAAAAAGTATGATTAATTACTTACAAAACCTTTTTTGATACCTTAAGTGTATTTCT
 AATTCGAGTATTTTGATAATTTAGGTCGTAGTTGCACGCGCTAAGAAGAAAGACCAGGTTGTTCC
 GATACCTCCAAGGAGATAATGAGCAACTCCAACGGAGCGACCCGTAGTGATGCTTAATGCTCGAG
 GAGAAATCGATGGAGAAAGGCGCACTTTACGGTGTGCCATACTAATGTGTCAATAAGCTCTTGC
 CAATACCCTCTACCAGAGAAAAGGAACATTAGACTGAAAGCCCAAATGAAATGAGCGCCAAGGAA
 TATTAACCATAACCGGACGCAGGTGTTCCATAACTTTGAATTACCTGTGAAGATTGCGACCATA
 GGAATCTCGAAGCCATCCGTTAACCGTAAGTGACCCAGAAGAGAAATTTCCATTCGTAATGTGC
 GATACCCGGCCGGAAGATGTTATGGTTCCCCAAACATCTGACTGCATCTTCCAATAGAAATGGAA
 TATAGTTATCGAAATACAATTGTACA

Rp16

CCCCGGTATGGTCAAGTCGAGCTGGTAAGGAGGGCATAGTATGAACGATCCTATCGGCGATATG
 CTCACCCGTATCCGCAACGCACAGATGCGCGGCAAGTCCACAACCATGACACCCGGCTTCCAAGCT
 GCGTGGCCGCGTTCCTCGCGGTGATGGCGGACGAAGGCTACATCCGCGGTTTTCGAAGAAACCAAG
 ACGTGAACGGCCACCCCGGTACGAAATCAGCCTCAAGTACTACGATGGCACCCCGGTGATCCGC
 GAACTCAAGCGCGTGTCCAAGCCGGGCGTCTGTGTATCTGGGTGTGGATGACATCCCGCAGGT
 CCGTCAGGGTCTCGGTGTGTGATCGTGTCCACGTCCAAAGGCGTCATGACCGATGCAAGTGCC
 GCAGCCACAATGTTGGCGGCAAGTACTCTGCACWGTCTTCTAAGGAGGTCTGGTA**ATG**TCTCGT
 ATTTGGTAAACAGCCGGTTCGAGCTGCCCTCGGGTGTCTCTGCATCCGTGTCTGGCCAGACCGTCA
 AGTGAAGGGCCCGAAAGGCGCCCGCAGCTTACCGCGACCGATGATATCACCATCACGGTTGAAG
 ATAACGTCATCACAGTCACGCCGCGTGGCAAGTCCAAGCGCGCACGCCAGCAGTGGGGCATGTCC

CGCACTATGGTTGCGAACCTGGTTCATGGTGTGACCAACACCTTCAAGAAAGAGCTTCAAATTCA
 GGGTGTGGTTACCGTGCTCAGCTTCAGGGCAATGTCCTGAAACTGAACCTTGGCCTGTTCGCATG
 ACGTGGATTTCCAAATTCCTGACGGTGTGACCGTCACTTGCCCCAAGCAGACAGAGATCGTGGTC
 GAAGGCCAAGATGCTCAGCTCGTGGGTCAAGTGGCTGCGAACATTCGCGACTGGCGTTCCTGA
 GCCCTACAAGGGCAAAGGCATTCGCTATAAGGGCGAATTCATCTCCGTAAGGAAGGCAAGAAGA
 AGTAAAGGACGCAAGGACATGGCAAACAGTAAACGTGATCTGTTCCAAAAGCGCCGTCTGCGCGTT
 CGGAACAAACTTCGCAAGGTCAATGCAGGGCGCGTGCGCCCTTAGCGTTCACCGTTCGAACAAGAA
 CATCAGCGTTCAGCTGATCGACGACGTTTACGGGCGTGACGCTTGCAGCGGCTTCGACAATGGAAA
 AAGACCTCGGCGGCAAGAATGATGTGCTTCGGCTGCAAAGGTTGGCGCGGCGATTGCCGAGCGG
 GCCAAGAAAGCTGGTGTAAACCGAAGCGTACTTCGACCGCGGTGGTTCCTTTTCCACGGTTCGCGT
 CAAGGCTCTCGCCGACGCAGCCCGCAAGGTGGTCTTAAGATCTAAGATAGCGGTGCGTGCAAGC
 ACGCACCTAGATATCTCGTAGGCCGGGCTTCTGCCCGGCATACATTTTGCGGAAGGGGTGACTC
 TTTCCCGATGATCCGGGAGCTCCGAAAGGCTCACCTGGATTGGACAATCGGGCCAATGGCCCACA
 GCACA

Rp119

GGGCTATAAAAATACCAAGAAAATGAGCGATTTAATCAAATTAATAGAAAACGAATATAATGCTG
 TAAGGGAGTCTATTCCTGCTTTTGGCGCTGGTGATACTGTAAACGTTACGTTAAAATCAAAGAG
 GGAAACAAAGAAAGGATTCAGCAGTTTCAAGGAACTGTAATGCAGCGAAGAAATTCGGGTAATAG
 CGAGACTTTTACCCTAAGAAAATGCTAGTGGTATCGCTGTAGAAAAGAAATCTTCCAGTTATTT
 CTCCTAACATCGAGAAGATAGAAGTATTGAGAAGAGGTAAGGTAAGAAGAGCTAGATTGTTCTAC
 CTGAGAGGAAGAAAAGGAAAAGCTGCTAGAATCAAAGAGAACTTTAAGATTCACTTTTTAAGAT
 ATTTTTCAAGAGCCATTCCTTCTCAGGTTTGGCTTTTTGTGTTTACAATACTTTCAGCTACGCGC
 AGCTAAAGTTAGTGTCTCACTAACTTTTTACAAGAGGTGTTTGTGAGGACACGAACCAAAGCGT
 AAGTTGATGACAAAACAACATTGATTCTTAAGCCTACAATGATTCTATTGTTGGTGTACTTCAT
 AACTTTGACGCAAAGAGAAGCTCATGACTATCAACTTTTTCAAATATCAGGGTACAGGCAATGAT
 TTTGTAATGATCGATGATCGMARCAATGGATTTCTGATGATATAAAATATGTACAGCAATTGTG
 TGATCGTAGGTTTGGTATCGGGGCTGATGGCCTCATCCTAATCCAGAACCATCCGACCTTGGATT
 TTGAGATGGTATACTTCAATGCTGATGGTAGCAAGAGCCTTTGTGGTAATGGCAGCAGAAGTGCT
 GTGAACCTCGCTAAGCAACTCGGTATTATAAAAAATGAAACGAAGTTCCTGACCATCGATGGGGA
 GTATGAAGCTACTATCGATGAGGATCAGATCGTTCACCTGCACATGAGGAATATTGAAGGGCTAA
 GTGCCAAAGGCTCTGATTTCT

rp123

ATTTTGGCTCAAGGTATATGGCTCAAGGAGAGTGCGCTTGATTGCCAGTCCCTTATAGCGTGGT
 GGTTCCATAGCTCCAGGCGAGATGAAACGTGGACGTGGCGGCAAGGGCGGTGGCAAGCTCAAGAT
 CTCGCTCGGTATGGGAGTGGCGTCTTGATGAATTGTGCGGACAACAGCGGTGCCAAGAATCTCT
 ACGTCGTCGCTGTGGCAGGCATCGGTGGACGCTTGAACAAGATACCAAAGTGCTGCCAGGAGAT
 ATGGTACTTTGCACCGTGAAGAAGGGCAAGCCGGAGCTCAGGAAGAAGATGTTGAAGGGTGTGGT
 GATCCGTCAGAAGAAGCAGTGGCGGCCCGCAAGGCCTCTTYGTCTACTTCGAAGACAACGCTG
 GCGTGATTGTCAATGACAAGGGCGAGATGAAGGGCTCAGCGATCACAGGCCCGCTGCCAAGGAG
 TGTGCSGAACCTCTGGCCAAAGGTGGCATCTACTGCACCTGCCATCTGCTAGAGGATATCTTGAAG
 AGCAAGAAGTGAGTGTGACCAGAAGCCGAAAAGAGAGCCGAAAGAGAGAGCTGAAGAATATCT
 TGAAGAGCCGWGCGCATTGTACTACCAGAAGCCGAAGAAGAGCCCCGCCCTTGAGGCCATGATCA

AGGATCGCTATGCGCGGGACTATGCAGACCGCGGGCGCGAAGTGCCTCCAGCACCCACCTTCTCA
 GATGCAGACCGGCGCCTCCAGTACGACGAGAATGCCGAGGTTTCATCTTGGTGTGGGTTGCGATGG
 TTGCGGCATCTGGCCAATCAGAGGACCAGCTTGGGTTGATGCAGACTGCCCAGAGCCAACCTGGTT
 TTCATCTCTGCGACCCTTGCTATCAACTGGGCTTTCACCGCCGCTCCTCACTGGCAGGTTCAAT
 CAGACACACATGCCGAAGAACAAGATGATCCCCT

rp127

ATTACAAATTATGTACGCAATAGTTGATATAGCCGGTAAGCAGTTCAAAGTAACGCAAGATCAGT
 ATGTTTACACTCCTCGTTTGGAGGGTGAAGCTGGCGCTCCGTAGAGTTTGACCGTGTATTGTTG
 ATTGACAACGATGGAAAGGTCAAAGTAGGTGCCCAACAGTGAAGGGTGCCAAGGTTTCAGGAAA
 AATCCTTGCTCAGGAGAAGGATGATAAAGTGATCGTATTCAAGAAGAAGAGAAGAAAAGGATACA
 AGGTAAAGAATGGTCACAGACAGCAGTTTTCTAAAATTTTAGTTGAAGACATTAAGGATATAAT
 ACCATCGCACATAAGAAAGGTGCAGGTAGTTCGAAAAACGGAAGAGAATCCCATAGTAAACGATT
 GGGAGTGAAGATATTTGGTGGACAAGACATTGTTGCTGGTAATATCATCGTAAGACAAAGAGGAA
 CGCAACACCATCCAGGAGAAAATGTTGGAATGGGGAAAGACCACACTTTGTTTGCATTGACAGAC
 GGTGTAGTTGAGTTCAAAAAGGAAGACAAAACAGATCTTTCGTATCAGTAGTTACTCCTGAGGC
 GAAGGCTTAAATTAGTCTGTTTAGAATATAAAAAAGCCATCTCGGTAAACGGGATGGCTTTTCTTG
 TTTGTAGTGTGGTACTAGACAGCATCCTCCTCTAGCTTTCTCAAGTCGAATTCCTTCAGAAAT
 TTTTGAGCTTCATTAGTATTTTGAAATTCGAAATTTTGGACCTGGCCATCAGTGGTGTSGATG

rp131

AAAGGTAACCCTGTCCCTCCCCAAGGGTCAGATCCCTCAATACTCTCACGGGATTCGACTTCATC
 GGAAATAATCAAGACAAATCCGTGCTATTTTAGACATTTAATAGATCCGTTTATCGCATTAGATT
 TGGATTATAAGGATTTTGACGTTTTTGTCTCTGTCAGAGGTGGGGGATTGATGGGGCAATCAGTA
 GCCATAGGTCTTGGGGTCGCTCGTGAACTTTGTACATGCACGCACAATTAATGAAACAGCTGA
 GGGCATCGAGTCGCCCTTCGTCTGCGGAAGGAACCTCTATGGAACTAAGAAAAAAGCTAAAAGTAC
 ATAAACTTTTAAACCCGAGACCCACGTTGTAAGGAACGAAAAAAATTTGGCCGGAAAAAAGCTAGA
 AAAAAGCCTCAATTTTCGAAAAGATGATCGCATTAATGCCAAAACCTGCAATTCACCCCACTTGG
 TATCCTCATGCAAAAAGTTTATCTCGATGGACAATTACTCCTTAAAGTTGGTGCACACTAAGCCAAA
 ATTAGAAGTGGATATTTGGTCCGGAACCTCACCTGCATATACAGGGCGTACAGAAGAAGCTCGATA
 CGGAAGGGCGAGCAAAAACGCTTTTTACGAAGATATGGTTGGGAGGAAGCTTCTCTAACAGAAGCT
 ACGATAGATGCCAACGATTCAACAACCTAGTTAGTCCCGTAGAAAACCGAAACCTCAACGTAGCA
 AAACCTCTGCTCTTAAAGGTTGTCCCAGCGTCGAGGTGTTTGCACGACTTATTGCACAGCT
 CCTAAAAACCAAATTCAGCTTTGCGAAAAGTTGCGCGCGTGAGATTATCCTCACGAGTCGAGGT
 ATCTGCGTCAATACCAGGTATTGGACATACTTTGACAGATCACTCTATTTGTATAGTAAGAGGTG
 GTCGGGTAAGATCTACCAGGTTGTCGTTATCGTGTGCTTCGAGGTGTTTTGGATTGGCGCCA
 GTACAAGGTCGAAAGCGTGCACGATCAAAATATGGATGCCCAAGGCCCGTTGGAAAAAAGTGAT
 TAATTAACCATTAATAATTTTATCGAATATTTTACGACTAAAAAACGTTGTGTACGCTCCCTT
 ATTTAAAAAGTATAAC

rp136 - weak homology

AAAAGTTC AAGACCTATTTT GATCGGCTCAATAAAAAATCTGGAGTTTGAATTTGTTTCGCAGTGAA
 GCCCCTCTTGTTACATTAGCCGAAAAATGAAGGTTTCGCTCCGCGGTCAAGACCATTTGCAGCGGC
 TGCACGCTGGTTATTCGCAATAGAGTTAGGTATGTAGTGTGCAAGAAGAACCCGAAGCACAAACA
 GCGTCAAGGATTTTACACCTTGGCGAATATTTCCCCGAGTGAATTTCCGTTCGCTGCCCTTTTGT
 TGAAGGGAAGCATAGGTCAGCAGAACTTGAATACCGCGGCCGCTCTCCATGGGAATCTCGGCCAA
 ATGCGGGCGAAAATATTTTGGAACGCACTCAAAAATTTGCTAAATAACTGGGTTCCACATTATATAT
 TTCCGTTTCTTTGGGTGGCATTTTGGGCTGTTTCGTTGCATTTTCGGCAGCTGAGCAAATGTGTCT
 TTATCTCCATATAGGCAGATCATAGACGAGGAGAAGGATCAGTGGCATTGAGTTACGCGCATTGA
 ATCGATTACCGCCAGTGTGGTTATTGTGGAGGCACCTTTCTTTATGTGTATTGTATATTGAAATT
 GTCTTTAAT

rpoA

GTGGGCCAACTGGTCTAGTTCTAACACGAATCCGACGAGATATTCAGCCATGGCAGGTTTCGACGG
 GCAGCGGAATACATGACGCACGCACCTCTCGGTTTCATTAAATTCAGTGGGAGGAGTTGCGACGGA
 AATTAATACTTTCAACTATGTTTCTCCACGTTTCATGGTTATGTACTTCTCATTTCTTCCTTGGAT
 TCTTTATGTGGATTGGCCATTTATGGCATGCCGGACGTGCTCGAGCCGCAGGAGCTGGGTTTTGAA
 AAAGTATTTTCGAGAGAAAACGAACCGTTTTTATACATGAGGCCGTTAGATTAGTCAAAGTTATA
 TTTTGAACCTTACTTCGATCAGTGAGATTTTCCACGCGACATAAAGTTTCAAACCTGTGTAATAGC
 TAGCAAAGAGAATTGAACCTTGGCCTACTAATTACAAATCAGTTGCTCTACCGGTTGAGCTACG
 CCAGCGCAAGATAACAAGACCTTATCTTAACAGTCTATTTTGAATATAAAAATACCCTTGAACCT
 CATAAAAAGTCTAAATCAAGATAATGAGTAAAGTTTTGTTAGCAAAGTATCGATGATTTTTTA
 TAACAAAGAAGAATTTTGGAAAGATTAAAAATCACCACCAGATTAATTTATCTTTTTACTTTGTA
 AACGGTGAAGGATCTATTATGATCATTTTAGCTTACAATTGATTGTAAGGTTAAGCGTCTCATAAA
 ACTCTTCAAAGTATAAAGATTTTCGTAAAGAATAACATGTATAATATTTTAAGGTCACCCGTCAA
 GAATTAATGGATCCTACGGGACAAAAGTTTAGTTCATGGAAAATATTATGTAGGCCCTTTATTAAG
 AGGGCAGTCGATTACACTTGCTTGTGCTTTGCGGCGAACTTTACTGGCGGAGATAGAAGGATTCC
 CTATAATAGGAATTCATGTTCCCTAACGTCAAACATGAATATGAGCATATTCCTGGACTAAAAGAA
 GATATGCTGGAACCTTCATTGAACCTAAAACAAATTAAGATTACATCCGAGAGCAACACGAATAC
 TACCTTGGCAAAAATTAAGTATCGGGGTCTAATAAAAAGTCAGGTTTTCTATGCGCGCAGGACC
 TGATACTTCCAGAAGGTTATACTGTTTCTAATCCACGAGCGTATATCGCGAGTTTAACTTCGACA
 GACGTCCCGGATTTTACCATGGAACCTTCTAATTGCTCGGGGAAAAGGTTTTGTGCTGGCTCAGGA
 CATACGGGCACGATTACCGACACGTTTCATTCTCTTGACGCTACTTTTACGCCAATACCTCGCG
 TTTTCGTTTTCTTTGAGGTCTCACGTGACAATAAATATGAAACACTCATATTTCGAGGTGTGGACT
 TTCGTTCGTTTTGACCCGAGATTTTAATTGCTGAAGCTTCTTTTATTTTGAACGTACATTTAG
 ACGTTTTTCGTCTGGTTCCTTTTCAGTCTACCATCAAACGTCCTGTTTCTCTTCCAACGTCTACAA
 AGTTAGAATCTGCTCCTACTGAACTTGACCAAACCTTCATTAAGAGAGCTTAAGTTAACAGTCCGA
 GCTCGTAACTGTTAATTCGAGCAGGCGTAAAAACCGTTGGTGACGTTACTCGTTTCTCACATAA
 ATCCTTACTACAGCTGCAGGGATGTGGTGTCCACTCTGTTAAAGAAATAGAAGCCAAAATGCTCG
 AATATTTTAAATTGGAAAATATCTTAAAGA

rpoB -rpl16-rpl14 poly-cistronic mRNA

AA
 AAATCTATTATCTAAGAATGCTTCCCCCTGG
 ACAGGACACGGTTGCCGCAGCATCTGTAGTTTCAGCTGGCCTAGCAATTGGATTATCATCAATCG
 GCCCCGGTCTCGCCCAAGGGGCCGGCAGCCCAAGCTGTTGATGGTCTTGCTCGTCAACCTGAG
 TCTGAAGGAAAAATAAGAGCGACATTATTGATCTCATTGGCTTTCATGGAATCACTAACAATTTA
 TGGATTAGTTGTTGCACTGTGTCTTATATTTGCAAATCCTTTCCTATGATGAATCATTTTATATT
 TAGCTTCGTACATAGTAATTTCTTTCATGGTATAGTTTAACTACTTCGAACGATTAACGGATT
 GGGTAAACAATCTTTCATAATTTTATAGAGCACTTCTACTAATTTTTTTAATAACTTAAGACTCTC
 CGTTACTGAACGAGCTTAAATCTAACTGCAGAATACGTTGATTTTACGTATGAGAAAGTTGACT
 TCTTATTGCTTCTTCCAAAAGAGTTAAGTTCGCACGGCAGCATCGTTCGCCCTAAAGCCAACG
 TCTTTGTCCCCCGAAATTCAAAATGGTTTTTGGTGATTTTGGGTACGGTTCGGTGCACCTAG
 TTGGGTTACTGCACGACAAATAGAAGCAACTCGTCGAGTTTTGACGCGTTATGTTTCGACGCACAG
 GAAAAATATGGATTTCGAATTTTTCCAGATTTACCCATTACTGAAAGGGCAGCAGAGTCCAGAATG
 GGCTCGGGAAAAGGAACAGTTTCTTATTGGGTTGCTTCTGTACCCGCTGGTTCCATGCTTTTTCGA
 ACTTAACGGACTCAAAAACCTATGGCGTATAAAGTACTAAAACCTGCTTCACATAAGCTACCTT
 TAAAAACGTGCGTTGTTGAAAAAGCGAATGTAACTCATAAAAGTCTATTTATAACCAGGAATT
 TGGATTTTGTAAGTCTGCGATGAAAGAAAAAAACCTAATAAGGCTCTTAAGGGTCGTGTGATC
 AACTGTTCCGCCGATGGTAGCAATATGTGCACATGACTGTGCGTGGTCCAGAGACGAATTTATGA
 CCCATTTTACAAGAAGGTTCTCGTTAAAAGCAAGAGGTTTTGCGTAGTTGATTTTACTCGTAAAG
 TAAGTTTTGGGGATACAGTCGAATTTATTCAAAGCAGACCACTCAGTAAACGAAAGAGTTGGAAA
 ATCAATCGAATTCTAAAAAGTTAAGTTAAATTCTATTGTTGCATGATACAACCTCAAACCTGTCT
 TTCTGTTGCGGATAAATGCGGTGCAACTCAGTTACAATGCATTTCGAATTCAGGTACCCGGCGCA
 AATTTGCTCGTGTGCGGTGATGAAATCGTAGGGGTGGTAAAGGCGGCGATGCCCTTGCTTACGGTA
 AAACGTAGAGAAATGTTAGAGCTGTTATTGTTGCACAAAAGCAAGCTCAAGCGCGAAAAGATGG
 TTCACATATACGATTTGATGATAATGCTGTGCGGATAGTTGACAAAATCGGGCAATCCTCTCGGAA
 CGCGGGTTTTCGGGCCTGTGGCTCGAGAAGTGAAAGTTAAAGGATTCACCGCTGTTGCTGCGCTG
 GCTCTGGATTTATTAFAACGACGAATGAAACTTGACTTTAGAGCGTTATATAGAGAGTCTATTGT
 TCCCAAAAATTTAAGTTCACACGCGTATTCTAATCCTCACCAAGTTCCAAAGATCCTGAAAATTT
 GCATTAATTGTGGCATAATCGCACGCTTCCCTCGAAAACGCAAAATCAGCCTAAGGAATTAGATTCC
 GTACTCCGTGAATTAAGCTTAATCGCAGGCTTGAAACCTAAAACAGTAGCTACAAAGAAATCGGT
 TGCTGGATTCAACATTCGTGATGGAATGATTGTTGGAGCCACTGTTACTCTTCGTGGTGTCTCGAA
 TGTATTCCTTCTTGAAAAATTAATTCACCTTGATTTACCCCGCGTTAGAGATTTCAAAGGCATT
 GACGCTCATGGTATGGAGCTGGAAATTAATCTTTTGGTATTACTGATCAGCTGGTCTTTCCTGA
 AATTTCTTATGAAAGCGTGGATCGCAGACGTGGTTTCGATATTTCCATAGTTACAAGCGCTTCGA
 CAGCCGAAGAAGGATGTGCATTGCTTCGTTCTTTTGAATGCCCTTTGTTACTTAGTTAAACTTA
 CGTAAAGAACAATTTTATAAAGATTATTTTCTAAATAGTAAAACTATGGTTGAATATTTCTC
 TAATATTCAGGAAAACGATCGTTGGAACGATCTGTGAATCCGAACGTAGTTATCTTAGCTCTAC
 TTAATTTAAATTTTACACCGTTCCATAAAAATTTCAACTCAAGTATTAATATAACGTTGTATGATT
 TTACCTCAAAAACGTATCTTAATTTACACAAGTTTCAGCGCGAATATAATGTTATATTTTTAGA
 GTAGTAAAAAGGTAGATTATTAATTTCTTATTCAATAGCGTCGTGAACATGAAGTTAAATTTCTT
 CTATGCCATAGTTTTAGAAGATGATAAATGTCTTCGTACAATGACTTTTCTCTTTATTTAAATA
 TTAGTCTTAAAGTATCCTTTAAAGACTAGGATGGTTTTCACGAAAAATCTGATCCTTCAAAAATAG
 CCGATAACTTCGTTTATTGGAATGCCATTTGTAAATCTCGCTTGAGTACAAGTTTCGTAATATAC
 GTAATAAATGATTCAGTTAGCGACATGTTTACACGCATCCGAAATGCAGCCATGCGCAAACT
 AAAGTACTTACGGTTCCGCGACACCAATAACACGTGGAATTGCTCAGGCCCTTTTATCTTACCG

TCTACTTGAGAATGTTCAAGAAAGTATCGAACAACTTCTCGCAGGACGCTTTACTTATCTTTGG
CTACGAAAAGAACTTCTCCTCAAGAGATTATCCTTAAAGATATCCAACGGTTGAGTAAACCGGGA
TCGAGAAAATATTGTAAAAGTAAAGCAATTCCTCACGCAATAAAAAGGGATAGGGATCCCTATAAT
ATCCACTTCAAAGGTGTAATGACAGGTTTGGATGCTCGAGAAAAAGGTCTTGGTGGAGAATTAC
TAGGATACTTTGTG **FAA**GTATCTCGTTAATATTTTTTCTAACCTAAATCAGTACGGTTAATTTAC
ATGTGATTTAATATTTCTTAGTGACCTTTCATTTGTCTTTAAATTAATGCTCTTTTCTGGAAATA
GCATGAAGGTTTTACCTTCGAGCATTAAACAATGTTCCGGTGAATGACTTATTCAAGTCGTGGAAT
AGGACTCCTTTTTTTGGTATGTGTAACGATAACCTGAATTATCAAATAATGTAAGTACTAGCTTTTA
TAAATGAGTTAAGAAAGTTTTTTATTTTGGGAAAAATTCAGTTGAACAAAAATAATTTAAGT
TTTTTGCAGTTAGAGATTGATTCATGACATAATTATAATCTTATAACTAATTCCTTTTGCTCATG
TTTATTTTTTTTATTCTAGACTCACTGTAAAGTTATAAAGACTTAAACT **GTG**TATAACCAAAAATC
AACTTTTTTACGGATCACAACGGGCTCTTCGCATTAATGAAATTCAGCACTTAAGTTTTTTTTGGT
TCTTAGAACGAGGATTCGGGCTTACCTTGGAAAAGTTTTCAAATTTCAATATATGCAAGGGCGT
GATTACATTTCTCGTAGAATACTCTACAGACTTTTACAACTTTGCCGGCCACCTTTATCGTTATC
TTCCGCAAAATCCTATGCACGTAGTTATTATACTACTTTATTTGTTAAAGCTCGTTGGACGAAAC
TTAAAGACACCAGTAATGCATGTTTACTTAATCCTTCATTATAGCGAGTGAAAGTCCAGATTTA
CATCATTCTCCTCGTCGTCCTTATCTTTACTTTAAAGTATGTGACTTACCTCTGTTAACACAGCA
GGGTATTTTTTATATTAATGGAGTAGCACGGGTTGTGATTTCCAGATTATCCGAGGTCCGGGCG
TCTACTATAATCAAGAACATCAGCCGCGTAAATCTTGTATACCGCTACATTGGTTCCCTTCTCC
GGGCTTGGTTAAAATTTAACCTCATTACTACGCATAAGGGGTCAAAGTGGTATTACACTATAA
AGGGATGGATATTCCGATGTATCATATTTTACATATATTTGGGTTTTCTGCTGTAAAAATTTCTGA
GCGTATTTCCAACAGACCCAACTATTTTTAAAATGGAAGCTGCAAAACCAACTTCGAAATTTGT
AACGAATTACGAGAAATTTTTACCCGTTGCTCTCAAACCCCTAAACAGTGATAAAGGTAAAAAAA
AATATTCAAGAATCAATTCTATCGATTTTTCTTTGATCCGCGAAATTACACTCTATCGGCGCTTG
GTAGACGAAAATTAACCTACATAACCACTTCAAAAACGGATGAAATTTGTAAGTACTTTGACCCCA
AACGATTTTCTAAACATTCTTATAAGTTTGTCTCGTTATGCTTATGATCCTAAATTTCCAGGCACC
CGATATCGATCATTTGGGAAATAAAAAGGATAAAAAGCATAGGAGAAATGCTACAATTAACAATTC
ATAAAGGTTTTTTCGCGCCATCAGCAAATATTAGTTACACGTGCCGGGAAATGGTTAAAGCCTCTT
GATAAACGATTAATGTACCTTAAATCATTAAATAAATGTCTCACTAATCGGCCTTTTGAAGCTCT
GCTTAAAGAGTTTTTCAATACTAATCCACTTTCACAATTTATGGATCAAGTAAACCCTTTATCTG
AATTAACGCATCTCAGAAAAGTTTCTCTCTTGGGACCTGGCGGACTAACTCGCGATCAGGCGCGG
TCAATCCAAGTCAGAAAGGTGCATCCATCGCAAACAGGAAGAATTTGTACAGTTGAAACTTCAGA
AGGTCAAATGCAGGATTGGTAAGCGCTATGTCTCTTGCAGCCATACCAGATGGGTTTGGTATTC
TTCGGGCTCCTTTTTTTCGTGTTCAATGGTAACATTATTTCGTAGACCAACATATCTCGATTCC
CATCAAGAAGAGGCATCTGTTGTATCTTCGACTGATCGATCGGTTAAATACGACAGTTTGCCAAC
TGTTTCTGTGCCGGCACGATTTTTTTCAGGTTACGTATGATGTCCCACCTGACATGCTTGAATATA
TTAGTATTTACCATTACAAAATTTATCGCTAGGAATATCTCTTGTCCCTTTTTCTAGAACATAAT
GATGTAACACGTTCAATTAATGGCATCGAATATGCAAAGACAGTCACTACCTTTAATGTTTCCGAG
TCGATCTATTATCGGGACGGGTTCTGAAAATCTTGTGGCTACTGAATCATTACGGGCTATTTCTA
GCGTCAGCTATGGAAAAGTTATTGCAGTTGACACAGCAGAGGTTTTTGTCAAAGATTTACACAAG
AATATCTTGAAATATTGTCTATGTGCCACGAACGATCAAATCAAGCAACAAATCTCAATCAAAG
AGTTTCAGTTTCTCTTAATGAAAAGTAGCTCCTGGGCAAGTAATTGCTGATAGTTTTTGTACAG
AAGGTGGGAACTAGCTTTGGGTCAAAGTGTTTTAGTAGCATAACATGCCATGGGATGGATATAAC
TTTGAAGATGCTGTCTTAATAAGCGAACGGTTAATTTCTGACAAATGTTTACACATCGATGCATGT
CGCAGAATTTACAGCGGATATTTGTTGTACTGCGTTGGGGTTAGAAGAAGTTACAAAACATTTAC
CTGGTATTGCTGCTTCAAAAACCTTCTCACCTCACCGATGAAGGTTTCGTCTCAATAGGAACATGG

GTAAATTCTCATGATATTCTTGTGGAAAAGTAACGCCGTTTGTATCCTGAAACGTTGTCTCCTTC
 CCAGCGACTACTAAAAGCTGTCATGGGCATTA AAAAATTGGCATATTACAGCGACACCGTTTTCGTT
 TACCCGTGGGACAAGATGGTGTGTGTGTAATATTAAGTTTTTAAAGCTATTCTCGAACCGTGG
 GAAATCGACATCCCAAACGTTTTCACATGCAAGTATTCGAGTTTGTCTTGACACAACGAAAAATACT
 TTGCGTAGGTGACAAAGTAGCCGGGCGTCATGGAAACAAGGGTGTATTTCCAAGATTTTACCTG
 TGGAAAGATATGCCTCATCTTTCGGATGGGACTGTCATCGATATGATTCTAAATCCAATGGGTGTG
 CCGTCGCGAATGAATGTTGGTCAAGTTTATGAGTGTACTAGGATGGGCCGAGATACACTTGG
 TTTACGACTTAAAGTATCCCCTTTTGTATGAAGGCTATGGAGCCAATGCCTCTTATAATTTGGTGC
 ATACAAAAATGGCGCAGGCTTCACTTCTTTCACCTCAAAATTTTTTCGGATATGACCCAGGTAAG
 ACGTTCCCTTTGGGATGGTCGAACGGGGTTGGCATTCACAGTGCAGTCACGGTTGGAAAAGCCTA
 CTTTTTAAAGCTTGTTCATTTGGCAGCGGGAAAATTACATGCACGTTCAACAGGTCCATATGCTC
 TATTGACACAACAACCTCTTTCACGGACGACGTA AAAATGGAGGACAGCGATTTGGTGAATGGAA
 GTATGGGCTTTACAGGCGTATGGAGCTGCATATACATTACATGAAC TATTGACAATCAAATCTGA
 CGATCTTTTAGGTGTA AAAAGACGGTAAAAGCACTTTTGC AAAGTTTTCCGGTTCCAAAGCCTG
 GTATTCGGAAACGTTTCGTACAATGGTGC GTGAATTACAAGCATTGTGTCTTGATATTGCAGCT
 TACCATAAACGAAATCATGATTCGTCTCTCGACGTAGAAGTTAATCTTTTTTCAGTACAATTTGGA
 AGTTTATTGCGTGAATATCAAGATATTGTA AACAGTAAAACAATAAATTATTAATGTGAAAATCT
 CAAATAAACTTTAAAGATATTTTTAACGTAAATATGTGCACACTTATTATTATCTCGAAGAACT
 ACGTCTTATAGAACAAGTTTTTTACAAACAAAAC TTTCTCTAATGCAACCGCTCATTTAAATTTA
 TTACATAAAAAAGGCCTGTTTCATAACA AAGAAAATTAGGATCCCTTTTTTGCTATATTCTTTTTT
 TTCGACGTCAATCTTACATAACCTAAAAACGTAATTTTAAACAAGAGTGATAGACCTTGTACAA
 TCAGCATGATTATCACAAATTCTCTAAAAC TTTTAGTACGAAAGGACCGAAGAAGCATTGGCAA
 TGGTCTTTCAGTTGTGCGCCAACGGTACACGCTGGGTAGCTAAC CAGTATTGTGCCAATATCCT
 AGGTTGTAAGCTTTTTTAAGCTCGCACTATGTAAAATATAACTGCTGAAAGCATCTAAGCGGGAAG
 TCAGGGCCAAGATTAGCAATCCTTTTTAAACGTATAATCATTTAATGTTTGACGAGGGTTGAAAGT
 TTTTTCAGGGACGCTGCTAGACAAGTTGTTTGATAGGCGTTAGTGTGAGGTTTGCATGTTGAGA
 TGCACTAATTGTCCCTTATCTTTGATTTTAGTAAAAGAATACTAGTGGTAGTGTGTCTTGCCCA
 TGTTGAGATGCACTAATTGTCCCTTATCTTTGATTTTATTA AAAAGAATACTAGTGGTAGTGTGT
 CTTGCAACTTGGTATCGTTTTTGCAGGTGCGTGTATATCACAGGATTTTATTA AAAAAACTGTTG
 TTTTGAAAAAAG

rpoC1

GCTCAAGGGACGCTACAAGCAGCTTTGTGCTCCCTCCCAGGGGCACCAAGGGGTTCCCGGCGGG
 GAAGGTTCGGCGGCACCTGCTGTGCAAGACGTGGGTGCGGCTGGCGAGTCTGGGAGGGCGTCGCCT
 GCCAGGCGGGCGCCGAAGTGTGAGCTGCGCGGTGCC TGGCCAGCCAGTAGTCTTCGGGGACACGGC
 CATCGGCAGGCGCAAAGAGCAGCGCGATGACGACCGCGTATGACCTGAATGTGCCATACTCCAGT
 GCGGAGCTGAAAAGAGTAAAGGCTGTGAGGTTTACGATTTTTTGTATGCGGACACGATTGTTGCATG
 TTCTGTGGCCGAGATCACTGAGTTAGACATCTACAAGAATGGTGAACCTATGTACGGGGGTGTAA
 ACGACCCACGCCTTGGCCCGCTTGATCCTCGGACACCATGCCACACTTGCGGCCTAGGCATCAAG
 GAGTGCCCTGGGCACTGGGGTTACGTCACCCTGGCAGGCCTATGTTCCATTGGGGCTTCATGAA
 AGCTACTACAATGTGCTGCGGTGTGTTTGTACTACTGCTCGAGATTGTTGGCTGACCCGAAGG
 ATCACCACATGATCAACGCAGAGGCGATGAAGAATGCAAAGAAGAGGTTGCAGGCTGTCATGCTC
 TGCTGTGAGGTAAGAAGCGATGTGCAGTCGCAGCAGATGCGGCCAATCCCAGGAGGTGCTTGG
 AAGGGATGGCGAGGAGGATCCAGGCAAGAACTTTGCTGGTGGCTGTGGGTATCTCCAGCCAAAGT
 TCACCGTAGATATGACCGTCTGTACGTGCTTTCCCGAAGGTGGAGAGGACGAGGGAGATCGA

AAGCGCGTCATGAGTGCTCAGGAAGCACTTGTGATTTTCAAGCGCATCACTGACACGGATGTGAG
ACGCATGGGCTTCGACCCCAAGATCAACCATCCCTCACAGATGATCCTGACTCATATGGCCATTC
CTCCACCAGCAGTGCGACCGTCCATCCAGATGGGAGCAGCGCGCGCCGAGGACGATGTGACTGCA
AAGTTGTCAGACATCATCAAAGTGAACCTAATGCTCAAGACGCAGGTGCAGTCTGGCGCCGGAGA
TCACATCTGCGCCGAGTTTGCAAAGCTGCTGCAGTACCACATCTACACGCTCAGCGACAACGGCA
TCATCGGCATCCCAACGGCGACGACAAAAGTGAAGCGACCACTGAAGTGCATCCGAGAACGACTG
AAGTCCAAGGAAGGGCGCTTACGAGGCAACCTCATGGGCAAGCGTGTGATTTCTGCGCACGGAG
TGTCATCGGCGGGGACGCCAATTTCAATACGGAGCAAGTGGGCGTCCCTAGGTCCATTGCTCTCA
ACCTAACATTCCCAGAGCGGGTGACACCTCATAACATACATTGGCTGAAGAAATAGTTGCGAAA
GGCCCGACACAGCACCCGGGTGCGAGGTACGTTGTCCGAGATGATGGCAGCCGGCTCGATTTGAG
ATATGTGCTGATACAAGCGCCATCACGCTGCAGTACGGCTGGAAGGTGCAGCGGCACATGATTG
ACGGCGACTATATCATTTTTCAACCGGCAGCCCTCATTACACAAGATGTCCATGATGGGACATCGC
GTGAAGGTGATGCCTTTCTCCACCCTTCGCTTCAACCTAGCAGTGACATCGCCATAACAACGCAGA
TTTTGATGGGGACGAGATGAATCTGCACCTTGGTCAGAGCCACGAGACTCGATCTGAGATCAAGC
ACATGATGTTGAACCCACGGATGGTTGTGTGCGCCGAGGGCAACAAGCCAGTGATGGGCATCGTG
CAGGACGCGCTGCTCGCCACGTGCAAGTTCACAAAGAGGGACAGCTTCCTGGAGAAGGACCTTGC
CATGAATATCCTTATGTGGGTTCTGTCTGGGACGGCTGTCTGACAATACCCACCATTTTGAGGC
CGAAGCCCATGTGGACTGGAAAACAAATTCCTTTCAATGATCATCCCAAAGGGCTTGTCTTGAAG
CGCGATGCTGCAATTGCTTCGAAGAACAAGAAGGACAACCCGGATTTTCCTGCGTCCGACTGCAA
GGTCATCGTCCAAGATGGCGAAATCCTAGCTGGCCAAGTCTGCAAGAAGACGGTTGGCTCCAGTG
GTGGCGGCTTAATCCACTTGATTTGGCTTGACTACGGGCCAGAGGTTTGCCGGAACCTTCATCTCG
TATTTGCAGAAGATTGTGAACCAGTGGCTGACGCACAACGGGTTCTCGTGTGGGGTGGCAGACAT
TATTGCCAACGACGAGACGTTACTCAATGTGAAAAGACATTGAAGCAGGCCAAAAGTCGAAGTGC
GAAAGATTTTGGCCGACGCGCAGCGTGGCAAGCTGGAGACGCAGCCTGGGAAGACCATGTACCAG
TCTTTCGAGGCCAAGGTCAATCAGAGATTGAACGCCGCCCGTGAGGATGCTGGGCGCATTGGGTC
TGACTCGTTGGATGAGCGCAACAACATCATCTCGATGGTGAACGCTGGGAGCAAGGGTTCTCCAA
TTAACATCGCCCAGATTATTGCTTGTGTGGGGCAGCAGAACGTCGAGGGATCCAGAATCAGGTAC
GGCTTCATGGATCGGACGCTGCCGCACTTCACCAAGGACGACTACGGTGCTGAGTCACGTGGCTT
TGTTGAGAACTCTTACCTTGCCGGCCTGACGCCGCAGGAGGTTTGGATGCACGCCATGGGTGGTC
GAGAAGGTGTCATTGACACCGCGTGAAGACTTCTGAGACAGGGTATATCCAGCGGCGCCTGGTC
AAGTCCATGGAGACGCTCAGGACTCACTATGACGGTACTTCACGGAATGCCAACACGAGATCAT
TCAGTTCATCTATGGTGAAGGATGGCATGGATGCCCTATGGATCGAGGACCAGACGCTGGAGATTA
TGACGTACGACAACAAAAAGCTCGAGAACACATTCAAGCATGATTATCAGTCTGAATCTTATGGT
GATTCATGGTTGCCGCCGGACGTGCAGCAAGCCATCAAGGCCAGCGGAGAGAATCAGCAAATCTT
AGATGACGAGTGGAAAAAATCAAAGATCTGAAAGAGATGATCTGTAACGAAGTCTTCCCCGACG
GCGACAGCAAGCAGCATTTTCCGATCAATCTTGCAAGAATGATGGGACGAGCCAAGCTGAAGTTG
GCCCAAGAGAATGATATCGACCCGGAACGATAAGTTTACGCCTATGGAGGTTGTTTCAAGGTTCAA
CAAATTACTCTCGGAGCTTGAGATTACAAGAGCGATCAGGGAGGGTACTCCATCGGCCGCGAGG
TCGAGGACAATGCAAAGATCGTCTTGAACGGACATCTCAGAGGCTGGCTCGGCTCCAAGAAGATC
TTGCAGCAGGAGAACTCAGCAAGCAGTCTTTCGATTGGCTGCTCGGCGAAGTCAAGCAGCGTTT
CATGAAATCTGTCGCGTTTGCAGGTGAGATGATTGGCACCTTGGCAGCACAGAGTGTGCGCGAGC
CTGCGACGCAGATGACTCTGAACACGTTCCATTTGCCGGGGTCCGCGCCAAGAACGTGACTCTG
GGTGTGCCCCGTTTGAAGGAGCTCATCAATGTGGCAAAGACCGTGAAGACTCCTTCATTGGCGGT
CTATCTGAATGGCGACCTGGGGAAAGACCAGGAAAAGAGCAAAGGACGTGCAGAGTATCCTCGAGC
ATACGACGTTGGAGAAGGTGACCTCCTTACACAGATCTTCTGGGACCCAGACCTGAGAATACC
CGTGTGGAAGAGGACAGGGAGTGGGTCGCTGAGTACTATGAGTTGCCCGATGATGATGAGAATCC

AAACAGATGTGGACCGTGGTGTCTGCGCATCCAACCTCAGCAACAAGGTCATGACGGATAAGAAGC
 TAACAGTCCGTGAGGTTGGTGAACGCATTCTTCAAGACTTTATGGGCGACTTGGATTGCATCTTC
 ACTGATGACAATGCAGAGGAGCTGGTCCCTCCGAATCCGACTCCTCAAGGAGGCGGCAGAGATTGA
 CCAGGTCCTCCACCTTTTCGACCCCTTGGACGACAAGGAGGACAAAGACTTCAAGTTCTTGAGAA
 GCATCGAGGCCAATATCTTGAAGGAGATGTCGCTACAGGGCATCCTCGGCATCAAGAAGGTGTTT
 ATGCGTGAGGACACGATGTCAGTGTACAACGCCGTCAAGGGAGGCTTCGAGCGCACAAAGGAGTG
 GGTCTCGACACTGACGGTGTCAACTTTGAGGAGGTCATGCAGATACCACAGGTGAGATTTGAGC
 GCTTGACAGAGCAATGACATTGTTGAGATCCTGAACGTATTGGGCATTGAAGCTGTTTCGTAATGCA
 CTCTTGTTCATGTTCGAATGGTGATCTCCTTCGATGGTTCATATGTCAATTACCGGCATTTGGG
 CACCTTGTGCGACGTGATGACTCAAAGAGGGCACCTCATGGCCATCACGCGCCACGGAATTAACA
 GAACCAATCAGGGGCCGCTCATGAAGTGCTCATTTCGAAGAAACGGTTGAGATCTTGATGGATGCT
 GCCATCTACGCAGAAGTTGACTACATGAGGGCAGTCTCAGAGAATGTCATCATGGGGCAGCTGGC
 GCCGATTGGTACTGGTGTCTTTGACCTCTACATGGATGACAAGGTCAACAAGGATGGAGACTGCA
 TGCTGAATGATGCAAGACCCGTGCTTCCCAATTCTGGGAAAGCAAACGAGATGTTCTCCATGGCT
 TCTCCAGCCGGAACGCCATCGAGGGGTTACCAATTCCGGCACTCACGCCGACGCGGATCAGGT
 TGAGCCTGACTTTGCACCGCCTTCAACAGAGGGGTGATTAGCGGCCTGCCGACGCCGAACGTC
 CCCC GAATGGTTCTCCAGAATCAATGGCCGGCGGACAGTCGCCTTTTACGCCCATGGACAGGTCA
 CCGGAGCAGCGAGGTTCTCACCAGCCTTCTCCGCATTGTCTAGCCAGTATCCGCAGGTGCTCG
 AACAAAGTCCGTGCTACTCGCCGGTCTTGGGTTATTCTCCCCAAGCTGACGAGCCTCAGAATGTGT
 CTCCACATACAGGCCAGACGACGGTCAGGGGTCCAGCAGCCCTCTCTACAAAACCTGGTGCTCAT
 GGTCCCTCGAGCTCCGCGGTGTACAGCGCTCCGTCTTACAGTCTGGCCAGCTGGTTACGCCCC
 CAGTAGTCCAGGGTACACACCGGGTAGAAGTCCGGCGTACGATTATTCGATGCCAGCTTCAACC
 AGTCTCCGTATCAAGCTGATTTTGGTGCTCGAACTAGAACGCGCCCATCTCCTTCAACGGCACCA
 GCGGACCACAGTCCGACGTGCGCAGCGGACACACCTGTAGACGGACCCGAGGCGCGTGGTCTTT
 TTCATCGTCACCGAGCTACACTCCTCATCCTGGGTGGATGGAGGGAACACTTTCGCCGAATTTA
 ACTATGATCCCGCAGCAGGGGGCTATGCACTGGACGGGGAAAGCTACTGATGACAATGCGAGCGAC
 GTCTTCGAGCCCAGCGATGATGAGATGGTACGATGAGGTCAGGTCCTCAGGAAGAGTCAATCCG
 CTCCGTGGCGAATGAGGTTTTTCAGGCCCGGCAGTCTGACAGCTGACGGGCACGATGCTGAAGACC
 ATGTGCGACGTGTGCATGGCCGGGGCAGTTATGCAGTGGCAGATTGGCTTCGGCACTCTACTAGA
 AATAGCTTGTACGACG

rpoC2

AAATTTTTTTTTCGCTGAAGGTGGAGGC
 TAGAACTTGTAGTTCCTCAAACAATTGCAGCTTTTAATTGCGCGCGGCTAAATAGATATTGTGCAT
 ACTCTTGGATCAGTAGTCTAATGAATCGCAAGCTGCAGCAAGCTGAGCTTTGGAAAAGGTACGA
 GCGCTTGCACATTTAAGTGCTGCCCGGCGTGCTTTAGGAGGAGTTATTTGTGCTACGTCACGGAT
 ACGCTCTTTCCAATACGACATTCGACCACCGGCGTGTAGTTCTAGGATCTGTTGATATGACT
 TTTTCGTTAATTTCTGTAGTCAGATAAGTTACTAATGCATATGCACACGGTGAGATGACCAAGATA
 AGATTAAACAGGTTTCGAATCACTAAGCGTGTCTGCTATCGAGAAAAGGGACACTTGCAAAAAAGTC
 GGCATATCGGTCAATGGCATTATTGTGAGGAATCCGAAGGGATTGATATCCCACCGAGTTAC
 CTAGTGCCAAGGGTATTGATGGACTTTGGAAAGGTATCCCTGCGGTTAAATTCCAGATCGTCCTT
 AAGGAATTTTCGGCCGTCTCGGAAACCGCTTGTACGGTATTTATATGGTCCAGGAAACCTCTTG
 CGCGCTATCTGCGGTGGTTTTGTGCAGTGTACTGGATGACATCTTTTGGCGTGAAGGACTGGTCAG
 AGATTGTTGCGGTACGGTGCGGGCCACATTTTGGTAAGAAAAAGCTTCGTGCAAAGTCTAAATG
 AAAAATGACTAAAATATCTTTATTTCATATTTCTCTATCAAAAATTATTATCTGCTGATTAAGAAA

TTTAATACACTTAAGATAATCGCGAAGGATATAAACTTTACAGAAAAGAAGTAAAAATGATGCAT
GAAATAATAGTTGAGTATATTGAGAGATTTTGATTAATAATGATAACAGCGAAATTTGTTTTACG
TACTTTAGTTTTGTTCTAGAACGACCTGCTCAAAAAACGAAGGTTTTTTAACTCATAATTACAAAC
AAGTTTTTCGTTGTAATGCTTAAATAATTCTCTAATGTTTCGTTATACTAATAATTAAATAACCT
ATTATTAATAGCAGTTTGGTTTCGCTAATATTATTTCCCAAGTTGGATGAAACGTAATAATTTTTT
ATTGTGATGGAACGCGCTAGTTGAATTATAATGCTGTCGAATTAATTAATAACAACTTCAATCA
AACTTGAAATACTCGAATAAAGCACTAGATAAAAAAGAGTTAAAAGCTTTTTTAACAATTACTGC
ACATAGCTTTGGTGCAACGCAGGCAACGATAATAAGTGACAAGGTGAAAAGCCTTGCTTTTCAA
GCTCAACTTATTCTGGCGCATCTTTGTCACTGCAAGATTTTACACCGGCTCCCTCTTTAATGCAT
TTAATACATTTAGCAAATGCAGGTGCGCGCCGAGCAAATAATGATTATTGTAATGGGGTAATTAC
ATTTAACGAACGACACAAGCTTATTACTACTTCGTGGCGAGCTGCGCTTGCTCCGTTAACAGATG
CTTTGCAAACGTATATGCAACGATGTGACTCTTTAAGTTCATTGTATGCCCTCGCATTTTTCCGGA
GCTCGAGGTAACATAACTCAGGCTAAACAACCTTGTGGTATGCGAGGGTTAATAATGGATGTGCA
GGGTAACTAGGAGAGCTACCAATCCGAACAAATTTGTTTTAGTGGTATGACTAGTTTAGAAATTG
CGATTGCTTCGCATGGTTCTCGAAAGGGTCTCGTTGATACAGGTCTACGGACTGCTGATGCGGGT
TATTTAACAAGACGTTTGGTTCGATTTCGGCACACGATATTCTAGTGCAGAGAAGAAGATTGTTTTGG
CGATGACTTTCTAACACTTGAGGAATTGATATCCCAGAGTAAAAATCGTAAGATTTTCATGGGGAG
AAAACTAGTAGGTCGCGTCTTAGCAACACCTTTATATAGCCCTGAAAACAGTTTGGTTTTTACA
AAGAACATGATAATAACAAAATCTAGCTTAAATAGTTTGAGGCGTGCAGGATTTTTGGAGTCTGC
AGAAAAACGCCGCTTATTAATATTCGTTCCCCACTTACGTGCACTACGCTTGCAAGGATATGTC
GAAATTGTTATGGTCTGATTTAGGAACCTCAAAAAAGTAAGCTTGGGATTAGCAGTGGGAGTC
ATAGCTGCGCAATCAATCGGTGAACCAGGGACGCAAATGATCTTACGAACTTTTCATACAGGTGG
AGCTGGTATAACTTTGGGTTATACGGCTAGAGATATCTGTTCTCCTTCGAGCGGTCTCGTAACGT
ATCACAAATATACAAGCAGCGCTCATAGTGAGATTTTCGCGATGGCTTACCAAGCTTGGTTTTGAAA
GAAGAAGGATCCCTTATTATTACGGGTAAACATTCTATTACTCAGTTAACTTTTCCTGCTGGAAC
TGCAATATTCCTTGCTAATAAATCTTTTGTCTATAGAAACCAAGTGATAGCAGATCTCCGTGTTT
CTCGACGTATGCGACCCCGCGTTTTTCATCCCAACATTCCTTGCTTCGCCATTACAATTAAAAAAG
CCCGTTATGAAAAAATAGTGCAAGAACGCTTACCAAAGTCTGGTGATCTTTTTTTATTGCAAAGC
GTCTTATACTGATCGTAGGTTAACGCATTCTTCGTCTCCTTCTTCGATTTTCATCGATACATAATC
CAGGCGTGCAAAATTAATTCATCGAATTTAGAAGATACCCGTTTGTCTCTGTGGGTTTTAGATGGC
TACCGTTATCAGCTTACGAAATTGGCGAAGGTACTTAAACCCGGGAAGACTTTTTCTCGCGGCAC
TTCCATTGCACATTTGAAGCTTATTAATAAGTTTTCTGGAATTCCATATTACCAAAAATTATCCA
AAGATAAAGGCCCTTTAATCATTAAATCTCTTGTGTTGCCGATAAATCTTTTCGTTAAGCAGTCA
TTAGAAAATTTTCAAGCAAAAAACCGTGTACTTCACTTACACACGTCGCGAAAAATTAATCCCT
TCATTCGCATATTTCCAATGAAACAATCATTTCGACCAGGTCATGTTATTTTTACACCTCAGCTCC
ACCCACACTGGTGTCTAGCGAAATGTGGAGGAATAGTTTACCGAATTTACGAAAAATAATTTATTG
CTTCTTTTACCAGAAGAACGTTACTATTTGCGATCTTTTTCGGGAACGTTGATTTCGACCGCTCGC
CAAAAGTATATACTTACGCGCTGGGACTAAGATAACATCTTCTATTGTTATGCCAACTCAAGGGA
TTCTACAGATTGACGCTAAAAAGCTCACATTAACAATTAAAACCGGGTTCGTTCATGCGCCCAAAG
ACGGCTAGCATTCAATCATGTGGAACATTATGGGAACCTACCCCTGTACACCGAACATGGTTTTCG
TGATGATTCTAAAAATAATCAAGATTATTCTACGTGGAAAAATTCATTTTTGAGGAAGTTTTCAT
ATGTTCTCGAACGAGTTCCCATCAAATATGTTAACCTTACTAAGTTGACGTCATATATACGTTAT
AACATAATCTCTGTGCCTTTAGTAAAAATACCGGATTATATATAAAAACCATATATTCTGCATT
GTACGTTTCTGGTGAACGCGTTTTTAAACACGATATTTTATTTGCAACACGCGTTATTATGGACT
CACCAGCAAAGCGATTTCTTCGAACAAGTACGCAAAGCACACCAGGGATTACAAAATTACAGCTT
CACCTTGTGCTACGAGATTAAGACGTCACCGAATAAAGAAATATATTCAGCTGATTTTTTTCGC

AACCAACCAGTATGCTACTTCTTCGCGTCTAGTTTCCCAAGAAACAAATTATAAACGACTTACGC
 AAGTTGTATCGCAATGGGATAAAAATTCGGCTACTAACTTGTGAGCATATTTAAAATTCCTAAGCTG
 GCCTACACTGCAACCAATCAAAGAGATGCAATTGTTTTGAAAAACCTGGCTATGCCTAGTACTTG
 TGTGGCACCAGGCACATTAATTTCTAAAGTTCTTTTGAATGCTTCGGTATCAGGCCAGGTGGTAC
 GACAGCAGTTAACGCCGTCTCATCTCGAAGTAATAACCGTTTCGTCTTGC GACTTGAGCTGTTTT
 CTATTTAATAGAGCTTGTGAAGTTCTTTACGTTGACGAGGGCGATTTTGTAAAATTAGGTACGCG
 TTAAACAACAAGTTTAGTTGCCAACTGCGTAGGTCAAGTTTATCGTGTTTTCCAAAACACGTATTT
 TTATACGTAAAGGTATGACACGTTTAATTTTACGCAACGCAACAGTTCCTTTACGTGACTCTATC
 AAAAGCAATTTTGCAAAAAAAGGGTCGTTTCATGTCCACAACCTGGTCAGAGTTGTTTGAAGCCAA
 AGACATCACGCAAGGAATTAAGATATTTCTAAAATTCCTGAGGTGGAAAAACTTCTCAACAG
 GACTTTTAGTGCCCTTCAATGGAATAATTTACGTTTCACGCCTCAGTTTTAAGGGTCGTGTACCT
 AATCGTAGAATCTGGAAATTTAACATCCCAGATTCGTTATGTCAAATGACATTTTACCAGCGAG
 TGGAAC TTATGTTTCGTCGATTAGATGTTCTCAGTGCTGGATTTGTGTCACTACGAACCCGGTTAC
 AAACGTTATATTACTATTGCCATGAAGAAAAATTTGAGTCGCGTGCGGCATGTCAAGAAAGT
 ATCAAAGATGCTCGTTTACATCTAACCGAGCAGGTAAAGTCAGCTTATTTAGAATATGGGATCAA
 TATTGCCGATAAAAAATCTGGAATTAATTGTTTCGACGTATGACATCAAAGTTACTATTCTCGATC
 CGGGTGCAACTGATCTAATGATTGGTGAATACATCGACTTTTACAACTTATGGCCATTGAAAAC
 GCTGTTGTGCAAAAAATCTTTAAAAAACCCCGCTACGAGCCACTTATTTTGGCCTAACAAAAAC
 CGCATTACACAGCAATAGTTTCTTATCAGCAGCGAGCTTTCAGTCCACAGTTAAGGTTCTTACAA
 GTGCTGCAATTC AAGGTAAAATCGATTGGTTTATTGGACTAAAAGAAGCGGTCATCGCAGGACGT
 ATGATTCCCACCTACCAAGATCGTGCTGGCTCTGCTTTACATCTACCTGTTTCAATCCAGTTTAAA
 AACTTCGATTTTGAATCTTCGAAAAACGTAGAGCAAAGGGTAAAGAGAAAATA **TAG** CCTATTA
 CGTTTTATCGGTGGTTGACTACGTCCCGTTTTGAGAATTTACCTGCATGTATTGACAGCTACACT
 TTTTTG TACTCAACTAGTAAGCGTGATAACGTCCTAGGTTACTTTGTTTACGTAGCTAATTGTTT
 TCATTATTTTTGTATAGACTCATCGTTAGAAATACCCAAAATTATTTTTTAAATTTTGTCAATCG
 GGTGATGGCCTGTCCGAAAAAGCTAAACACTTTGTAAGGGGGCAGGTCATAATAAATTCGGGG
 AACCGGCTTGGCCTAACGATTTACTGTTCTTATTTCCCATCTGTATTACAGGTGTACTTGCTTGT
 AGTTTAGGTCTAGCAGTCTATCTCCGCCGGCAATCGACGCCCAATCAGACCCGTTTTCTACACC
 GCTGAAAATTTCCGGAATGGTATTTCTTACCATCATTTAATTTGCTACGTCTTCTTCCTAATA
 AACTTTTAGGTGTTGCTGCGATGGCCAGTATACCTCTTGCATTAACGCAGGTAGCAGCTTCCGAG
 AACCAGACCGCTTCTCAAATCCTAATCGACGCTCGCAAAGTTCACTGTTATATATCCTGGGATC
 TTCGTCATCATCATGGCTTGGCGCAGGAGGTGCTATGCCGCTTTCGCAGTCTTTAACTTTAGCAC
 TTACTTAAAAATTT

rps2

AAGAAAGACGTGTGATTGCCAATATGAATAGTGGACGGTATGGAGCGTCACCCCGTGCAGCTGA
 TGCATCTGATTTCGGCAATGTTTTTCAACACCAGACTCCACTCATCCGAATAACCAGCTGTATTTCG
 CATAACCACATCCACGGTGACAGCATGAGAGACAACCTCCAAGCTGCTTGAATTACCATGAGGCAA
 ATACCTTGCCGTGTAAGTTTTGGTGCTGGTAAGCAGGAAATAACTTTCATCAATGCTGAGCTCCA
 GAGCATATTCACTGCTAGCCCTGTAGCAAACGCTATGCGTAAGCCCTTAGAAATGATTCCTCCCA
 TGTGGAAGATGAGAAAGATCTGGGTGGACAACAGGAAACCTAGCAACACTATGTGTAACAACCTGC
 AGAAGCCAAAACGGCTCAAGCCGTTTCGGTTCGAGTTTGGTTCCGGAGTTTAGGTGCAGGTGTGA
 TTGTGAGACAGGCCATCCTCCAGATCACTGTTTGTAGCA **ATG** GCTATGACTACCAAAGAGGAAGAC
 ATTCAGATGATGCTGGTGTGCAAGACCCACTTGGGCACACGCAATTGTGATCACCGTATGAAGCG
 CTACGTTTTTTCGTCGCACTGTTGACGGCATCCACATCATTACCTGGGCAAGACCTGGGAGAAGC

TCATGGCTGCCGCGCATCATCGTGGCCATCGAGAACCCCTCCGACATCATCGTGGCGTCCCAG
 CGTCCGTATGGTTCGCGGGCGGTGCTGAAGTTCTGCCAATACATCGGTGCACAAGCCATGGCGGG
 TCGCTGGATGCCGGGCACGCTCACCAACCAGATCACTACAAAGTTCTTGGAGCCACGCCTCCTCA
 TTGTCACTGACCCCCGACCGACAGCCAGGCTATCAAGGAGGTTCGGCTACTCCAGCATTCCMACG
 ATTGCCCTCTGCGACACGGACTCGCCCTTGGAGTACATTGACTTSGCCATCCCAGCAAACAACAA
 GGGCAAGGAGTCCATCGCCCTCCTCTATTGGCTCCTGGCTCGTGAGGTTCTCTACTTGCGGAATG
 TGCCTCTCGCTCCCAGCCTTGGGAAGTGATGGTGGACTCCTTCTTTTGGCGAGATCCAGAAGAG
 TTGGAGCGACAGGAGGAAATGGATCAGGGCTTTGCGGCAGCTCCCCGCCAGGCTTACACTGAGGA
 ATGGACAGCTGGTGGCGGTGAGTGGGAGGCCGGACCGCGGCCGAGAATTCAGCTGGGATGCGG
 CTGGCGGAGCGGACGACTGGAGTGCACCAGCTGCTGCAGGGGGAGAGGCCACTGAGGGCGGCAAC
 TGGTGA GTGTGCTGCTGTCTGCTTATCGTGTGAAGAAGATATGAGCATGACCTGCGTTG

rps3

ATTTTGGCTCAAGTCACTCTGGCTCACGGTAAAACACTTCAGTAGCTTCCAGCCAACATCCACCA
 CAGCTTTCACAGAGAATCAGGACCTGCAATATGGCGAATCAAGGACAGATCTCGAAGAAGCGA
 AAGTTTGTGCGACAGGCGTCTTCTACGCCGAATTGAACGAGTTCTTGATGCGAACGTTGGCCGA
 GGATGGCTACGCAGGTGTGGAGGTTCGCGTCACACCCATCCGCACAGAGATCATCATTCGCGCAA
 CGAAGACCCGGGAGGTGCTCGGAGATAAGGGACCGGCATCCGTGAGCTTACCCTGTGGTTTCA
 AAGCGTTACAACCTTCCGGAGAACAGCGTGGAACCTTTCGCGGAAAGGGTGCAAAACCGGGCTGC
 TTGCGCAATG GCACAGGCAGAGTCTCTCCGATACAACTCCTCGGTGGTCTCGCTGTGCGCCGCG
 CATGCTACGGCGTTCTTCGCTTCGTTCATGGAGAACAACGCAAGGGAGTCGAGATCATCATCAGT
 GGCAAGTTGCGGGCACAGCGAGCCAAGGCAATGAAATTCAAAGATGGCTACCTCATCTCTACTGG
 TGAGCCTAAGAGGCACTACATTGACGAAGCTGTGCGCCACGTGCTCATGCGCCAGGGTGTGCTCG
 GCATCAAAGTTAAGATCATGATGGGGTACGACCCTGAAGGAAAGATGGGAACTAAAGTGATCTTG
 CCGGATAACGTTGTTGTCCATGAGCCCAAGGACGAGGCGCCCGTCAATTTACGCGGGAGCAAAC
 TTATGAGGAGGAGACCGGATGA GAATGCTACCAAGCACCAGCCTCTCTACGGTAGGTGCTGTTGT
 GCTCAGGGCGCGGCAATGGGAWCTCAGCCGGTAACCCTTACCACCTGGAGGCCACGCCTTTGAC
 AAAGCTTGGCATACTTTCGAGTCAGCAGGCAGCTCATTGGCCAATGGTGTGCAAGCCTCTTCGT
 GCCGGGGGACGCTGCATATCCTGCGCACAGCGCTTCTGGCGGACCGTATGGCGCAAGATCATAAG
 GTGCGTCTTGGAAATGACGCTCGCTGGGACAGCGGAAGACGAGGTGGACAAGGAGAGGTTGAAAG
 ATCACTACGGGAAATTTCTCGGACATGTTCCCTATTTTAGACGAGGTGATGATGCTGATGAGCAAG
 TCCGGTGCAGAGTTCTTCTTACGCTACCCTGAGTTTGAAGAAGTCTTCATGAAGCCAT

Rps4

AAAAAAAAAAAAAAAAAAAAAGTTTATCTAATAATGTCTAGATATGTAGGCGCAAAGCTAAGAAT
 TACTCGAAGATTAGGGCCTTTACCAGGGTTCTGTAAAGAAACAAACGCGGCGTACATACGTCCCAG
 GAGAGCATGGATATCGCGCCGACCAACAAAAGGAGGAGGAGGACGAGGAGCAGCAGCAGGAGGA
 AACGTGACTTTAGCCGAAAACGTTTACTTTGAAAACAACAACCTTTTGTATAATTACGGTCTACG
 AGAACGGCAACTAAGAGCTTATGTAAAGGCTAGTTCTTCTCGGTCTTCAACGATGCTCAAAGAAG
 CAGGATCTTCCATAATAGATGTTGTACTTGAAGCTTTAGACCGTCGTTTAGATACCGTGGTTTAT

CGAGCTTTATCTACGACATCAATTCGTGCGGCTCGTCAAATAGTGAGTCACGGACATGTGAAAGT
 TAATAGCGTACCTGTAACATATCCCCAGTTATTCGTGCCAAACAACAGATACTGTGGAATAAAAA
 AGGCGATTGTGGAACAGAGGGGTACAAAAGCTCGCCAGTCACAAATCGCGGCATCGTCTCCAAAA
 GAAGACTCGGATAAAATTTGAATCGTATCGTGTGGCCACTCTTCGTGATGCCAAAAACCTCGGTT
 GAACAAAACCTTGATTGCGGAGTTTACTCGCGCCTTCTTTAAACTTATTTTTAATTTGTGAGTT
 TTTTCGATTAGGCTGTTATCTCATCTTTATAACTTTATGACTGTACTTAATTTTCGAAAGTGGGT
 TTCTCCAAAAAGAGCTGATATAAATATGTCACTCCCCCTGAAGAGGTAGTCGCATCTCAATCTT
 CACCAGGATTTAAAAAGAACTTGAGCAATTGGTCACGCGTAGTCGTGAAGTACCTCACCTTGAG
 AAAGTTTTATTAAACACACGTGGTCACCTGGGTACCTTACGATGTCAGCTACACCCTCAAATGTA
 TGACAGAGTTCTTACTGGGTGCCCGTCAGATTCTTCTGAGGGATCGATAAAGCATGTAGCTAGTC
 AAAGTTATAAGAAGGAACGATTAATTATTGATTATCGTCGTTTCGGCTAGTGAATTATCCAGCCTT
 CTTACTTTCTTGCGACTTCAGAGAGAAACAGGTGCTTACCAAGCAGCTGATGTCTTCTAGTAGA
 TACTCGAGTTACACCCTCGAACTGTTGACGAAAGCCAAACAGTCCAAGTTTAAACCCAAAAC
 GCTCCTGGAGGCATTTAGCTTTTCGTCCCTCAAAAATCTTGGTCTCAAAGGAAGCCTTTATGGAA
 GGAAGTATTTTCGCTTACGAGTTTGCAACCTACTCCTAGAAAATCCGTTAAACCAGCGACCTTAAC
 CCGACAAGCTGCAATTGCGTGTGGATCTCGTTTTCGTTAATTTCGCGGTGGCTTGCTGGGATATTAA
 CGAACGCTAACCAAAATTTCTCAGGGTGCTACCCAATTAAGTTTCTTACGCACAAGCCTCAGTGGT
 CGTCACTCGAAACCATTTCCCAGAATGAATCGAAGAGTCGTTAGTAAATTTTCGTCGAAGATTTCA
 ACGACTTAACCGCTTATTTGAGGGAATTGAGGGAACCCGCGAGTCTTCAGCCCCCTTCGACCCCCCT
 ATTTACCAAAAATGGTAATTTTTTTTTAACTCGAAGTCAACCAACCAGCTTTAGAGGAATGTAAG
 CGTCTTAAAATTCCGACAGTAGGCTTCATCAATACTGATGAAAACCCTTGC GGATTTACCTTTCC
 CATTACAATAAACACCGCTTCAGTAACTTGCCAAAAGGCGATGTTAACGCTCCTTGTCGAGGCTC
 TTCGAGAAGTTTAAAAAGAAAAAGTAGAGGAGTATGGAGATAATCAAAAATAATTGAAGATGGAT
 TTTATAAAAGTAAATTATCATGCGGAAAAAACTTCACATGATCTTACTACGAGACGAGTTTTTAAA
 TGTAAGGCCTAACATCATTGGTTTTGGGTTATTATTTAACCAACTGATATCACATTTGTCATATG
 TCGTGGTCATAAGAATAATTTTAAATCCAACCAGTGTTTTTTTTATTACTAAGTAAAATATGTTTA
 AAAAAATTAGATAAGTTGCTGAATTATCCTTATTGAGTGTACTCTACGACTAAATGTTCACACTT
 AGATTGCCTAATGATAGTTTGATCATATGCCAAAATTTAACTCCCATAACCTTCAATAATTAAC
 ACTGAACATTCGGAAATCAAGAACTTTACCCTGTTGGTAAATCAGCCCAAGAAAGCTCGAAATCA
 TCCAATAATATTGTAGAATTATATAATTGTAAAATAATTGTTAAAAACAATAGGAATGCCACCAT
 AGTAAATACCATGAATCCCGTTGAACCCACCCCGCGTTACCAAGCCCGCTTGAGAATTAAGAG
 GCAACAGGGCTCTCCCTAACTTTGTTTTTAAACCTCTTCCCGTTGCGGCCGTTTGCTGATAAGAG
 AAGACTTTGTACAAAAATGGGTGGGTT

rps5

ATTTTGGCTCAAGTTGGCTCGAATCAACTTTCGTGCGGTGATCTCCACCTTGCTCTGCCTACATC
 ACATCCACTCGCAATCTTCAAATATCCGTGATGTCAGACGTGAAGCTCTTCGGACGCTGGCCTTA
 CGAGGATGTCAATGTGTCAGACAACAGTCTTCTCGATTACATTGCAGTGACGCATGCAGCCCACC
 ATTTGCTTCCCTCCCACACACGCAGGGCCGCTACCAGAAGAAGCGCTTCCGCAAGGCGTTGTGCCCT
 ATCGTCGAGCGCCTGTGCTGCTCCATGATGATGCATGGCCGGAACAACGGCAAAAAGTTGATGGC
 CGTGCGCATTGTGAAGCACGCCTTCGAAATCATCCACCTTCTCACTGACAAGAACCCGATTACAG

TGTTTGTGACGCCGTGAAGAATGGTGGACCTCGTGAAGATTCAACGCGCATTGGTTCTGCTGGC
 GTTGTCCGCCGACAGGCAGTGGATGTGTCTCCTCTGCGCCGTGTGAACCAGGCAATCTACCTCAT
 CTGCACAGGAGCACGCAACAGCTCTTCCGCAACATCAAGTCCATTGCCGAATGCCTTGCTGATG
 AGATCATGAATTGCGCCAAGGAGTCATCCAACAGCTATGCCATCAAGAAGAAGGACGAGATCGAG
 CGTGTGCCAAGGCCAACCGCTGA GCTTGGCAAGCTAAGAGGCAAGCAGCATGCAGATCAGATGC
 TGCTGGCCAGCATTGGCGACAGAGCAGAGTATTGCTTCATGTGGAA

rps12

AACCTCTATTTGCCYCTGCCATTCTATGAGTCTCTTCCTTTTTGCGTATGGCATCGCCACTCCCT
 TTGGCAGCATCCACTAATTCGGAACCTAATTTGAAAGCCATATTTTCGACCCGGACGTTTTTCGGGA
 TGCCGCTAATAACCAACGAATGGCAAGTGCTTTTTCCTTGTGTGGATCCTATTTCAATGGGAACCTT
 GATGAGTCGATCCACCTACACGTCTTGCTTTTACTGTTATATCGGGAGTTACTCCACGTATTGCT
 TGACGTAAAACGGATAGTGGATTTGTTTCTGTCTTTTGTGAATCTTTTTTCACGGCTCGATAGAT
 AATTTGATAAGCCAATGATTTTTTTCCGTGTTTCAGAATACGGTTAACCAACATGTTAACTAATC
 GATTACGATAAATTGGATCGGATTTTGCTGTTTTTTTTTCTGCAGTACCTCGACGTGACATGAGC
 GTGAAAGGGTTCAAGAATCAGTTTTCTTTTTATAAGGGCTAAAATCACTTATTTTGGCTTTTTTT
 ACCCCATATTTAGAACGCCCTTGTGACGATCCTTACTCCGACAGCATCTAGGGTTCCTCGAAC
 AATGTGATATCTCACACCGGGTAAATCCTTAACCTTCCCCCTTACTAAGACTACAGAATGTT
 CTTGTAAATTATGGCCAATACCGGGTATATAAGCAGTGATTTCAAATCCAGAGGTTAATCGTACT
 CTGGCAACTTTACGTAAGGCAGAGTTTGGTTTTTTTTGGGGTGATAGTATACACCCTAGTACATGT
 TCCTCGACGCTGAGGACAGCCCCGAAGAGCGGGGATTTTCGTGACATTTCTGATTGGCTGTCTTG
 TATTTCTAATAAGTTGTTTAATAGTTGGCATGTTGAATCGTATACATAATATGATGGGTTGGTTT
 AGATTGATCCTAACCGAATGATGATGAATTACTTCTATTTAATAGAATATTCAATTCGAAGATAA
 AATCTCAAATCACAGATTTGCGCGAAATCCATGTTATTTTCATTCAACSCCTCAAGCCTGTTTTGG
 CTCCGGAGAAGAAAGCAACATTGCTGCATAGTCGCATCAATCTGGTAAGCAGTCATGGCGGAAG
 CAAGCCTCGCGGTATCCGCTCTGCTCGTAAGCAGGTAATCAGACGCCGAACACAGCGATGGGCTG
 ACAAAGACTACAAGAAGGCAAACCTCGGCACAAAGTACAAGTGCAATCCCTTTGGTGGCTCCTCC
 CACGCAAAGGCATTGTTGTGGAGAAGATCGGGATCGAGGCCAAACAGCCAAACTCAGCCATCCG
 AAAGTGCTGCCGCGTGCAGCTCATCAAAAACGGGAAGAAGATTGCTGCGTTTTGTCCCCGCGATG
 GCTGTTTGAACCTTCATCGACGAAAATGACGAGGTGCTCATTCTGGGTTTCGGCCGACGTGGCCAC
 GCTGTGGGCGATATCCCTGGTGTACGTTTTCAAGGTGGTGAAGGTTGCAGGCTGTGGCTTGGGTGC
 TTTGTACCGGGAGAAGAAGGAAAAGCCACGTTCA TGA AGGACCTCATCATCTGAGAGCGAATGCG
 TCGAGCAGTGCGGCATTGTGCTGGCGCTCTGGTTCTGCCACCTGCATGATCTTGGACAGAAGGAG
 GCGGATTTAGAGCAGTCTTGGCAAGWRMKGGCGCTCTGGTTCTGCCACCTGCATGATCTTGGAC
 AGAAGGAGGCGGG

rps14

ATTTTGGCTCAAGCCAATCTGGCGCGAGGAAAGGCATATACTTGCTGCGGCACACATTGAGCATT
 CTTACACCTGTGCATAACTTTACAAGATCGCGCCGAAAACAAAGGCGCCAGTGGAAGCAGTGTCT
 CTTGGACCACAGGTCCGAGAGGGAGAGACCGTCTTCGGAGTTGCTCACATCTTCGCATCGTTCAA
 TGACACTTTTGTCCACGTGACAGACGTCTCAGGGCGGGAGACTATCATCCGTGTCACAGGTGGCA
 TGAAGGTGAAGGCGGACCGTGATGAATCGTCACCATATGCAGCCATGCTTGCTGCCCAAGATGTT
 GCCGCTCGTTGCAAGGAATTGGGCATCTCCGCTTGCACATCAAGCTCCGCGCCACAGGTGGCAC

CAAGTCGCGCACGCCAGGCCAGGTGCGCAGAGCGCGCTCCGTGCGCTTGCCCCGCTCAGGCATGA
 AGATTGGACGCATCGAGGATGTCACTCCCATCCCCACAGATTGCACCCGTCGCAAGGCTGGCCGT
 CGCGGTGCGCGTCTG **TAA**GTGCGGATGTCCATGCTCAGGGACGTGCAATCAGAGTGCCTTGTGGT
 GTGTTGTTGCAGCTTGGCCAAAAGTGTGAC

Rps15a

TTTGGCTCAAGGTTCTTAGAGCAGGACTGTGCTGCTGCATCTTGATTGCACCCCGCTTCCCGTGT
 TCCCAAGCTAGACA **ATG**GTGCGCATTAGCGTCTCCGTGATTGCCTCAACAACATCAACAACGCC
 GAGAAGCGGGAAAGCGCCAGGTCTTGGTGAGGCCAATTCCAAAGTCGTGTCGTCAAATTCCTGCA
 GGTTCATGCAAAAGAAATGGTTACATTGGCGATTTGGAGATCATCGATGACCACCGAGCCGAAAAG
 TGGTCGTGGATCTCATCGGCCGCATCAACAAGTGGGTGTGATCTCACCCCGTTTTGATGTGCCA
 ATTGGTGGCATCGAGCAGTTGGCCAGTGACCTCCTTCCCTTCCCGTCAGTTCCGGACATGTGGTCCCT
 GACGACCCCATATGGCATCATGGACCATGAAGAAGCTAAGCGCAAGCATACTGGTGGCAAAATCG
 TGGGCTTCTTCTTT **TGA**GGATAGCTTCGACTCTGCTCATTGCAGGAAAGGGAGAGAGAGCAGAGG
 ACTTGGCAGATTGTGAGATCCAAGCTGGCAGCTGTGGATGCAGCAATGGCACAGAGCCCAG

rps18

GCCAATTCGGCTGAAGATAACCAGTGATCAACTGCAAGCTCTGCGTGCTTGCCAAGCCCCTACTCA
 TGCCTCCCCACGCAACGGTTGACTTGTGCGGGAAGCGGACGTAGGCAG **ATG**CGCTGGGAAATG
 CGCTCCGTGCAGCAGTCTGCCAGCACGCCACCGTGCTGGTGCCTTCATCGGACGAGGCTGGCCCT
 GCCTGCGCCAGGCATGCAGCCAACACGCCTGGCAAGGTGCCAAGTGGCAGTGGTCAGTCTGAAGA
 GGCCGTGCGGCTGATGCTGGACACAAAGGTGAGCCAACGAGGCACTGCCGATGCGCTGTTTCAGG
 AATGCATTGATGCCATTCGAGAGCGCAGGTTGAAGCTTGAGCAGGATCCCGTGCTGCAGGAGGGT
 GACACGCCAGCCATCACGAAGCTCAAGCAGGACGAGTTGGACCTCACGAAGAAGATCTTTTCAGCG
 CGTGGACCCGTCCAGGGACATCACGTCTGACCTTACACATGCAGGGCACAACATTGACCCGTATT
 GGAGCCCCTTCCACAAGGTGCGAGAGGTAAGGACCAGGTAGCGGCCGAGTTTCGACGAGTATGCA
 CGCCTTGTACAGTGGCTGAGCAGAAGCGTGTGTTAGAATCCAAATCAAGAAGAGCTTGAGAAGAGC
 AGCAGCGGTCTACAACCATACTCAAAGAGTTCCGGCAGTGGCATGGCAGAAAGTCTGGACAAG
 CACCGCCCAAGCCCTTCCGATTGCCGACAAGCACTGGGAGCCATCACCCTACAGGTTTCGCTG
 GCGCGGAGCGCATCACGTGGAGAGATGTTGACATCCTCCAGCATTTTCATCGCTGACAATGGCTA
 CATACTCCCAGGAGGACCACGATGCTGCCGAGGACAGAAGCAGCAAGCCTTGGTCCAAGCAGTGA
 AGACAGCGCAGAACATGTCTTTGCTGCCCTACCAGTGGAAAGTTAAAAGACTTCCAGGCCATGCCA
 CTCATGGACCCTCTGCAATGGATGGCTGATCGGCTCACGGACCGGTCATGGAATCTCGAGACCG
 CCGGTCACGAGCAATGTTGAGGGTCATGATGGAGCGCCACCCGGAGCTGAACTACCGAAACTTTC
 TGAGGCATGAAGCTCAGCGGGCAAAGGGTGCGGGCGATCCCAGTGATGCTGTT **TAG**CTATGATAC
 AGACTCCAGGAAAAAG

rps19

AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGTTAAAACCAAAAAAT **GTG**GGTTCGATCACGTAATA
 AATCTCCTTATGTTTCTGCTGCACTTTTAACTAATTTGGCGAAACAAGAAAAGACTCATTCTCGT
 CAAGCATCCGGAAGAAAGACACATCCTTCCCCAATAAAAACGTGGTCGCGAGCCTCAACTGTTCAT
 TCCCACCATGTTAGGGTATACAATATGCGTTTACAATGGTCGCGAACATATTCCCGTCTTTATCG
 ATACTCGAAAGCTTGGCTATAAACTGGGTGAATTTGCACCTACTAGAACATTCCGTGCCATAAA
 CGAAAAGATGCGAAGGCAAGTAA **TAG**AATAGAGATTTAATGGAGGCACCAGAAAGGATTATAGG

TCGGAGTGTATGGAGATATGCCAGGGTATCACCAAGTAAGCTTAAACGAGTGTATCTGAACTAA
 AGGGTTTAACTTTTTCTGATGCTAGTACAATATTAGCTTTCCTGCCTTATAAATCATGCTCAAAG
 ATCTACCGGCTTCTTAAATCCGCTCGAGCGAATTACAAATTCATTGCTAAGAAAAATGGGTTCTTC
 TACACCGCAGGCACCAGAGCAAGACAGTCTTTATACTCAGCTAGTTATTACAAAAGCAGCGTCAA
 CAAAGGGTCCCCCGATGAAACGCATGCGTCCCGCAGCTCGTGGAAAGAGGTTGGCAATTTCTTAGA
 AGAACATCACACGTAAC TATAGAAAGTCACCGGGCTAGCGTTTTAAATCGGAAGTCTGATCGATGGC
 ACAAAAAACACATCCGATTGGTTTTTCGACTTGGAAATAACGAAACAGCCGCAAGCGCATTGGTATG
 C

rps23

AACCTCTATTTGCCYCTGCCATTCTATGAGTCTCTTCCTTTTTTGGCGTATGGCATGCCACTCCCT
 TTGGCAGCATCCACTAATTCGGAAC TTAATTTGAAAGCCATATTTTCGACCCGGACGTTTTTCGGGA
 TGCCGCTAATAACCAACGAATGGCAAGTCTTTTTCTTGTGTGGATCCTATTTCAATGGGAACTT
 GATGAGTCGATCCACCTACACGTCTTGCTTTTACTGTTATATCGGGAGTTACTCCACGTATTGCT
 TGACGTAAAACGGATAGTGGATTTGTTTCTGTCTTTTGTGAATCTTTTTCACGGCTCGATAGAT
 AATTTGATAAGCCAATGATTTTTTTCCGTGTTTCAGAATACGGTTAACCAACATGTTAACTAATC
 GATTACGATAAATGGATCGGATTTTGCTGTTTTTTTTCTGCAGTACCTCGACGTGACATGAGC
 GTGAAAGGGGTTCAAGAATCAGTTTTCTTTTTATAAGGGCTAAAATCACTTATTTTGGCTTTTTT
 ACCCCATATTTAGAACGCCCTTGTGACGATCCTTTACTCCGACAGCATCTAGGGTTCCTCGAAC
 AATGTGATATCTCACACCGGGTAAATCCTTAACCTTCCCCCTCTTACTAAGACTACAGAATGTT
 CTTGTAATTTATGGCCAATACCGGGTATATAAGCAGTGATTTCAAATCCAGAGGTTAATCGTACT
 CTGGCAACTTTACGTAAGGCAGAGTTTGGTTTTTTTTGGGGTGATAGTATACACCCTAGTACATGT
 TCCTCGACGCTGAGGACAGCCCCGAAGAGCGGGGATTTTCGTGACATTTCTGATTGGCTGTCTTG
 TATTTCTAATAAGTTGTTTAATAGTTGGCATGTTGAATCGTATACATAATATGATGGGTTGGTTT
 AGATTGATCCTAACCGAATGATGATGAATTACTTCTATTTAATAGAATATTCAATTCGAAGATAA
 AATCTCAAATCACAGATTTGCGCGAAATCCATGTTATTTTCATTCAACSCCTCAAGCCTGTTTTGG
 CTCCGGAGAAGAAAGCAACATTGCTGCATAGTCGCATCAATCTGGTAAGCAGTCATCGGCGGAAG
 CAAGCCTCGCGGTATCCGCTCTGCTCGTAAGCAGGTAATCAGACGCCGAACACAGCGATGGGCTG
 ACAAAGACTACAAGAAGGCAAACCTCGGCACAAAGTACAAGTGCAATCCCTTTGGTGGCTCCTCC
 CACGCAAAGGCATTTGTTGTGGAGAAGATCGGGATCGAGGCCAAACAGCCAAACTCAGCCATCCG
 AAAGTGCTGCCGCTGCAGCTCATCAAAAACGGGAAGAAGATTGCTGCGTTTTGTGCCCCGCGATG
 GCTGTTTGAAC TTCATCGACGAAAAATGACGAGGTGCTCATTCTGGGTTTCGGCCGACGTGGCCAC
 GCTGTGGGCGATATCCCTGGTGTACGTTTTCAAGGTGGTGAAGGTTGCAGGCTGTGGCTTGGGTGC
 TTTGTACCGGGAGAAGAAGGAAAAAGCCACGTTCAATGAGGACCTCATCATCTGAGAGCGAATGCG
 TCGAGCAGTGCGGCATTGTGCTGGCGCTCTGGTTCTGCCACCTGCATGATCTTGACAGAAGGAG
 GCGGATTTTCAGAGCAGTCTTGGCAAGWRMKGCGCTCTGGTTCTGCCACCTGCATGATCTTGAC
 AGAAGGAGGCGGG

ycf4

CTTAGTTACTAAGTTTTTTATCGAGTAACAGCAACAAACGTGAAAATATTTTTACCTTTTTCAAAA
 AGTGAGTTCTTCACTCACTAGTATTTTTCCAGTTCTTTCTACCATTGAATAAACAATAATGAAAA
 GTTCGCTACTCTTAAATGTTATGGATCCAAACGTAAGTGGCGTTTTAATTACGGATGATTATGGT
 ATTGGAAAATCGACAGATAGCCATTATTTGATTGATTTACTTCCCAAATTAAGTTCGTAGCCGG
 AGATCCTTATATTTTCGCTAGTCACCGATAAAAATCTTATGAGCAATGAAGTTCGAATAACTCTTA

AAAATTGAAACTGTATTTGGTCAATTCCATTGGTAGAATTGCCCTTAGAAGCAACATAAGACCGT
GTTTATGGAACCTTTGGATATTTAAAAATCGTTATAGGACGGGGTCAAAGCGTTTAAATCTGACGT
TCGTTTCGGGTTAAAGGTATTCTTTATATCGATGCGATTTTTTTTATTGGATTCTAGAGCGTCCAGT
TGATATATTATTATTATTAACGATGAAATACTCTCTTTGTGCCAGACTTCGAGATCTATATAATT
TCAGGTAACCCCTCAAAAAGGCAATTTATGTCCTCAACTTTTTTGACCAATTTGGTATGTCGGTGG
GGTAGAAATAGAAAACCTCTCTTACGCTGTGGATCAAGTTTGTGAATAAGAAGTTTTTAGCAAAA
GATGAATCCGGTATACAGCGGATGCGAGATATAATTGTGTTCTGGTTATTACCCAAAGTCGTGCA
CTTCAACGAAGGGATGTAATTACCAGAAGTTTTTTAGAGACATTATACGTTTCTAATACCGCAAT
TACTCTTTATGAATAAATGGTATGAAAATAATAGAACTGAACCCGGTAGTATTTTTTTTACCATGC
GCGGAATTGATTAGTTAGCCCGTGCAACAAGGTTTGGCTACCGGTCACGTTAGAAGTCTATCCAAC
ATATTGAGTATGCTCGTAATACTGGAGCATCGATTGCCAAAAATAAGAGGGGGTTAGTCTTGTGA
ATCACAATCCCCACTCGTTGGCGATAAGTTGAGGATTTATGCCTTAACAAAGTACCACTACAGCC
TTTGCGAAGGTTTTCTCCTTCTAATTTATAACCGCCTATATATACTCCCTCCAACGATATGTAAA
TGTATCTGACAAATCATATTCGGTCGTAAAAAATAACGTATACGTACTACATTTCGCGTCATTCA
ATAACTTTTTATCTATGTTTGGGCATTTTTCAAAAAATTGACTAAATTTAGAATTATTACCAATAT
TTGAAAATTTTTGCGGTTTTGTAAAGTCATAAAAGTGGCAACTTTGCTCTAAAGTTTCTAAATAT
GATGTCCCTTAAAGTCGCAGCTCAATTGGTTGAACTACCGTTTTTTTTTTGGCCAGGTTTCGTATGA
GTTTACGATTAAGGTATCCTACTTAGTGTACGTTTTTAAATTTAAGGATTTAGATTACTCTC
TTTTGGTGAAACTATACCCAGGGTTCACGTAATCTGTTACTTAAAAACAGGAATATTAATCGGT
ATGTTCCCATAGTCTTGATAGTTTTGATCATATACCTTATAATCAGATTAGGCGAGGTGACTAT
AGAATCTCTTCTCATAGGCAAATAGGATATCTTTATGATCAGTACCTTGGAAATAGCTGTAAAG
AAACCCACGTTTTTATACTTGAGAGTGTGGACTGATAGCCTATAAAGTAGAAAGTTGCAAAAAAG
CACACGTATAACCCCTCCAGTATTAGACTTTGATCTGTTAAGAGGGACAGAAGAGTGTCTTATT
TTTCATCTACCATGGCGTATTTATAGAAATGCGAGTTAGATTAGTTAGAGTGAAGCCAACGAGC
GAAAAATACTATTAGCTAACATGAGTCGGTTGAATATATTTTTTTCTTAAATAAGATTGCCTTTTT
TGTCTTCGACTTTTTGGTTCTGTTCTCAATTGAATAGTACGTTTGTAGTAATCAACAAGGGAACA
CCTCGCAAATTATCGATTGTAGACATCTTGAAGCTAGAGTAAATTCATCTTCTGATAATTGTTAG
ATTTTGAAGATAGACGAGAATAAAAAATCTACGAAGCATTTTTCGTATTTTTTGATAAAGCGTTGAT
TTTATATTTTCATTAATAATATATTTAAAAATATTTGATCGAAGCTAAAAAGACCCATTTTAAAGC
AACAGTGATTTGAGTAGACGTTTATCTTAGTTATTATGTGACAACGTACAATATAGGGGAACGTT
AATGAAAATATCGTGTTAAAATTACGAAGAGAAACGTCCGGATTATGCAGTATGTAATACCCCA
GCGTGGTGCAAGAAATACTAGTAATTTAATTTTTTAGTATTCTGCTAACTCTAGGTGGTCTTGGAT
TTTTTCTTACTGGGTGTGCGAGTTTTTTTTAAATATAATTTGATATTCTTAAACAACGTTAGTGAG
ATTGCATTCTTACCCCAAGGTATCGTTTTACTTCTTTATGGTACTATCGGACTTACGTTAGGATC
GTTTCTTTTTTTAACTTTTTATTGGAATCTTGGTGCTGGTTATAATTTATATTGTACCAAGAAAA
AAGCAGTTCTTTTTTATCGGACTCGATTCCCGGGTAGAAAAACAAATAACCCCTTACTACTTAAA
TTTCCTTTTTGCAACAATTAATTCCTTTGGGACTTAAAAGAGGGACTGCCCTTTCACCTGACAACCTC
AGTTATTGCCTATACGTCAAATGGAAGTAAATTTCCATTAATGTACACCGACACGATGGAAGTGG
CTAGAAATCTGCTATGACACTGGCTTATTTAATGAACTGCCCTGTTGATTGAAAACACTTTTT
CCTAAAAACTTTTTAGTTTCCGATCTTCGAAAGTTTTTTAAATCATCCATATTAATTCTTTTTCG
ATAGTCATCTTACTACGAAAATATGTTAGACTTCCGTCATGGTTTAGTAGCTTCAGTGGTTTAG
AGCAGGGGATTCATAAGCCCAAGGACGCTGGTTCAAGTCCAGCCTGAACCATTTTCTGGCACACC
GCCATAATGGCATTAGTAGCACCCTCATTATCATCATCAGTGTTTTTGATATGGGGGATTTTAGT
TGTGTATGTATATATAT

cbbX

TTTTGGCTCAAGCATTCAAAGATCGTTTCGTTCGTGGGTTGCTTGTTCAGTTTCCATCTTTACATCT
 CCGACCATGCAGAGGATCGCCTTCTTGATTTGCTTGGCCTGCACAGGCTTTGGACGACGAGTGCA
 AGAGCAAGCGCAGAGGTCCACTGCATCTGATCCAGCTTCGGCGCTTGAGGAACTTCTCTTGTAC
 AAGTTCACATGCCGCTTTCAATGCACCCGGTCAAGGAGTATATTTTCCAGACAGCAAGGTTGCC
 ACTACTAGTTCCCGGGCAGCAGCTGCCCGATGATGGGGCAGGAAAGAAGGCAAAGTTGAAAGA
 GCAAAAGGAGGCTCTCGAGGCGAAGCTTGATAGAGAGCTGTCTTTGAAAAAGAAAGCGGCAGAGG
 ACAAAGCAATTCAGGACGAGTTGAAAGCGGATAAGCATTTCCAATCAATGAGCCTCCGAAGGATT
 GCTGGTACAGTAGACAAGTTCATGTACAAGGAAGGTGTCGCAGAGGTTCTGGACAAACTTGACAG
 AGAGATGATTGGCCTCAGAGAGGTCAAAGACCGTGTGCGGGCAGTTGCATCGATGTTGGTGATCG
 ACAAGATGCGCATGAAGCTTGGCTTAGAGACGAGCATGCCGTCGCTGCACATGAGCTTCCCGGT
 GCTCCTGGCACAGGCAAACTACCATTGCATTGCGCATGGGCCAGATCCTGCAAAAAATGGGCTA
 CTGTTCGACCCGGGCACGTCGTGATGGCCACCCGTGACGATTTGGTCGGCCAATATGTTGGCCACA
 CTGCTCCGAAGACAAAGGAGATGGTTAAGAAAGCAATGGGTGGCATCTTGTTTCGTTCGACGAGGCG
 TACTACTTGTATGATGAGGAAAACGAACGTGACTATGGGGCAGAAAGTTGTGAGATCCTTTTCGAG
 TGTGATGGAGACCATGCAGGATGATCTCATCGTGATTTTTGCAGGCTACAAGGAGTTGATGGACC
 ATTTCTATGCGGGCAGCCAGGCATCAAGTCGCGCTTGGCAACCATATTGATTTCCCAACTAC
 TCAGATGACGAGCTCATTGACATCGGCAAGGTGATGCTTAATGCATCTTCGATGACAATGACGCC
 TGATGCTATTGTGGCGTTGCGAGAATACATGGTAAAGCGCAGGCAGTTGCCATTTTTCTCAAATG
 CACGCACAGTCCGCAATGCTATGGACATGGCGCGTATGAAGGGGTCCGTCCGTACTTACATGGAG
 ATTTGCAAACCTATGATGGGATCAACGGTCAATCTCCCAATGTGCGGCTTACCGAGGTCTCAGC
 AAAAGACATTCCGACGATTGAAGAGATGCAGGGTGGTGGTGATGGCCAGAACAAGGTCATGGGC
 AGGCAAGGTTTAGAAGGATATTGAATGCCTCAAGCGCAATTGAAGGATATTGAATGCTTACACA
 TATTTGAAGGATACTGAATGCCTCAAAGTGATCCTCGAAAGAATATTGAATGCTTCAAGCATAGT
 TGAACAAAAGTTATATGACCCTTCTGGAAGCCTTTGTTACCAT

clpC

GTTTGAAGTACTTRGCTCAAGGGCATTCTCTGCTAAGTGTAGTTGCTCACATTAGTTCTTGGTG
 ACGATGCGCAAGATTGCTTTGGTCTTGGCTTGCCTGGCCTGCACGGGCCATGGACGGCGAGCGCA
 ACCGCTGCCTGAGCGGTTGTTGAACGCTGCATCTGCAGAGAGGAGCCAAGTAGCTCAAACAGACG
 AGGCTTCAGTTGCTGCACTCGCAAAGGTCTGAAGGCCTTTGAGCCTTCAACTGGCTTCAACTCG
 GCATCGACTCCGTTTGTGCAGCACGCAACTGTAAAGTCTCAGAGTCCAAGAACGCAGCTTCCCTC
 AGGCAGGCCACTCTTGGCGCATTGCCGTTCCGAGCCTCCCTACACAGGCGGCCGACGGTCACC
 ATTCAGATTTAGCCAGCCACAGATGATGTTTGAACGCTTTCAGTCAAAGCGCCATTCAGATTATC
 ATGATGGCTCAGAGAGAATCTCAGCGGCTTGGTCAACAATTACGTGGGAAACGAGATGATGCTTCT
 TGGCGTCGTTGCTCAGAAGAGCGGTGTTGGCTGCGAAGGTGCTGGCAAAGCTCGGCGTGAAGCTCG
 AGGAGATGCGCAAGGTTGTGAGGAATTGGTAGGCAAAGGAAACGGTATCACGGGAGTGGACATA
 CCGGTTCGCGAAGGAAGCAAGGAAGTCTGGAAGGTTGCGTCGAGGAAGCTGAGAAGTTAGGATC
 AAAAGCAGTTGACACTGTCCACATCCTTCTGGCGCTCATCAAGGACGAGGATGGAGATGCGGCGCA
 AGATCCTCAAAAAGCTGAGCGTTGATCCTGCCAAGGTGCGGACAGAGGTCTTAGATGCAGTGAAA
 AGCCAGGACAAGCCTCTTGTGCGGTGCTGTTTCTCTTCCGAGGGGCCAAAAGTGGCAAGTCCAA
 CACAGCCACACTCGAGCAGTATGCTCGTGACCTTACGAAGATGGCGGAAGCTGGGGAGCTAGATG
 TGGTCATTGGCCGCGAGAAAGAGATCAAGCGAGCTGTCCAAGTTTTGGCGCGCCGCCGTA AAAAC
 AATCCAGTTCTAATTGGTGAGCCTGGCGTAGGTAAGACAGCTATAGCAGAGGGGCTTGCAGAGAG

GATCGCCGATGGAAAAGGCCCAAGCTTTCTGCAGGGCAAGCGCATCATGCAGTTGGATTTGACAG
 GTCTCTTAGCAGGCACCAAATACCGAGGTGATTTTGAAGAGAGGTTGCAAAACGTCATCAAAGAA
 GTGGTCGCCTCGAACAGAGAAATCGTTCTGATGATTGATGAGCTTCACATGCTCGTTGGTGCAGG
 AAGCGCAGAAGGTTCCATGGATGCGGGGAACATGCTGAAGCCAATGCTCGCGAATGGAGAGCTCC
 AGCTTATCGGTGCCACAACCTTGGAGGAGTATCGCAAATATATTGAGAAAGACAAGGCATTGGAG
 CGGCGTTTTCCAAACAATCAATGTACCAGAGCCCACGGCAGATCAGGCTACAGACATTTTGAAGG
 ACTTACATACAAGTATGAAGAGCATCACGAGCTCCGGTATACACCAGAGGCGATTGAAGCATGCG
 TCAAGCTCGCCGACCGGTACATCACAGATCGTTACCTGCCAGATAAGGCAATTGACATACTTGAT
 GAGACGGGTGCTCGTGTGCGACTGCAGGCAGCCAAGGCCGTGCCAAAGGAGACCCTTGAGATCGA
 GACAGAGCTGAAGGAGGTTGAGAAACAGTTAGATGAGGCTGTGCGAAGTTCAAGACTTTGAGAAAG
 CTGCAGAGCTCAAGCCGAAGAAGGATAAGTTGAACGCAACGCTGATGTCTCTTGGCGAAGGCGAG
 GAGGAGAAAGCTGAGAAGACTTTTCGTTTTCCGAAAGCGACGTGGAGCGGATCGTTGCAGACATGAC
 TGGCATCCGAGTGGAGAAGGTCTCCTCCAATGAATCTGCACGCTTGCTGGGATTGGAGGATACCT
 TGCACGATCGCGTCATTGGTCAAGATGAGGCAGTGGTTGCTGTTTCAAAGGCTGTTCCGGAGATCA
 CGCTCAGGATTGAAGGATCCAAACCGACCCATTGCTTCCTTCATCTTCTGTGGACCGACAGGTGT
 GGGGAAGACTGAACTGTGCAAAGCATTTGTGCGATGCCTACTACGGCAGAAAGGATTCCATGATTC
 GTTTCGACATGTCTGAATTCATGGAACGACACACAGTTTCCAAGCTCATCGGCAGTCCTCCTGGC
 TATGTTGGTTACAACGATGAGAGCCAATTGACTGATAAGGTTCCGGCGAAATCCTTACAGTCTTAT
 CCTCTTCGATGAAATCGAGAAGGCTCATCCTGATGTGTTCAACCTCATGCTTCAAATTTTGGAGG
 ATGGGATGTTGACAGACTCCAAGGGTCAAGTGTCTTTTCAAGAATGCACTCATCATGACA
 TCAAATGTTGGCGCACAAAGGAATTGAGAAGACTRRTTGCTGGTGGCGGAGGCTTTGGCTTTCAAAG
 CGCGACAGATGATGTAGAGCAGGCGACCTACGAGAAGATGAAATCAGTCGTTGGAGATCAGCTTA
 AAAACAACCTTCAAGCCAGAGTTCATCAACCGTCTAGATGAGACGATTGTGTTCAAGCCACTGACG
 AAAAAGGAGATAGAGCAGATTGCAGAGTTGGAGTTGGCCAAAGTATTTGAGAGAGTCAAAGAACA
 AGGTTTGACAATTGAAATGACAGAGAGATTCAAGAAAAAGTGCGTAGATGACGGCTTCGACCCTA
 AGTTCGGAGCAAGGCCACTGCGACGTGCCATCGCCAAACTGCTAGAGGATGAATTAGCAGCTTCG
 GTTTTGCTGGAGCCCCTTAGGGAGAACGAGATCGCGATCGTAGACATTGACGACGACGGCAAGGT
 GAAAATTTTGCGCAACCAGGGAGAGGTCAAAGTCACTGAGGAAGAAGAGGCGCCCGTCTTAAGA
 GACCGAAGCATTACTGTCTGTACCTGGAAGCTGCTTCACGTCTTGACAATGAATAGACATCCGAC
 GACGATCATTGAGGAGCATGGTATGCTTAATTGTGGAGCCTAATTGTTTTCGGGCACAGATAGTA
 ATTTGCTTTTTGGACGCATCTGACATTGCAGTTTGGCCCTAGGAATCATAGTTGAGTTGTGTTGC
 GATTTTTTTTACAGATGTTTCACTGGCGAAGGCCAGCAGTTATACAGTTGAGGGATATTCCAS

groEL

TTCCTAGTTATATGAAAAGAGTTAAGTCACAAAATATTTGCGATGGTTGTAAAACGTATTTTTTTTCGTGA
 GGAAGCAAGAATAGCTATGGAACGTGGAATAGATATGCTTTCCGATGCTGTTTCCACGACGTTAGGGCCAA
 AAGGCCGAAATGTTGTGCTAGGACAAAAGTTTAGTACACCTCAAATTGTTAACGATGGAGTCACGATTGCT
 CGAGCAATAAATCCATTAGATAATATCGAAAACATGGGTGTTTGTGTTGATTTCGTCAAGCAGCATCAAAAAC
 AAATGATGTGCGGGGGGATGGTACTACTACCGCTACCGTTATAGCTCATGCAATAATAAAAACAAGGGTTAC
 GAAATATAGCAGCGGGTTCTAATCCCATTGTTTTGAAGCGTGGCATAGAAAAGGCCGCTCAATACATTGTT
 GATCGTATTGCAGAGGATGCAATGCCTATCGCTGGTTGGGAAGATATTTTACCAGTTGCGACAATTTCCGC
 TGGAACGATAAGGTAATGGGGAAGCTTATTGCCGATGCTATTTTGGGTGTCGGTAACGATGGGTTGATAT
 CTGTTGAAGAGAGTAATTCTACTGCAAATGAATTAGAGTTTACAGAAGGATTTGGCTTCGACAGAGGATTT
 CTCTCGGGATATTTTGTACAGACTCTGAAAAAGTAGAATGTCTTTTGCCTGATCCTTACATACTACTCAC
 AGATAAAAGACTCACTGTAGTTAAAGATTTATTATCAACTTTAGAGTTAATGACAAAACCTTAACAACAAAC
 CCTTATTGATTATTGCGGAAGATGTCCAACAAGAGGCACCTGCCACTCTTATAATAAATTTCTCTACGAGGT

ATATTTAAAAGTTGTAGCTGTGCGAGCGCCTGGTTTTGGCGAGAAACGGAGTACGCTGTTAGCAGACATGGC
 CGTAATGACAGGTGGACAAGTTATTTCTTCCGCGAGCGGGTTATAATCTGGAAAACATGCAAGTTGAACTAT
 TGGGACGTGCTCAACGCGTAATAATAAAAACGTGACTCTACGACAATAATAACTGACGCGAATAAACAGAAA
 GTTTTAGCGCGATGTGAACAACATAAAAAGAGAACTTGAAAATGTAGATTTCGACATTTGAAAAAGAAAAAT
 TCAAGAACGATTATCAAAATTAAGTGGTGGTATCGCGGTTATAAGAGTAGGGGCAGCAACTGAAACCGAAA
 TGAGGGATAAAAAATTGCGAATGGAAGACGCAGTTAATGCTACAAAAGCCGCTTTGGAGGAGGGCATTGTT
 CCGGGCGGTGGAATAACTTTAGCCCGACTAAATGAAGATCTTAAACGTTGGGTGAAATGTTCTTAAAAAA
 TGACGAAGTTACAGGAGGTTAATTGTTGGCGATGCGCTCTGTGCGCCTTTTAAAAAGATAGCAGCAAACG
 CAGGAAAAAATGGAGCGGTGACTTTTGGCAAGTTAATGAAAGAGTCTGACAAGCGAGTGGGTTATGATGCT
 GCTAGCGATGAAATGGTTCGACATGTATGATCGCGGAATAATTGATCCAGCCAAAGTCACGCGGGCAACACT
 CCAAACGCCACTTCAATCGCTAGCATGGTTTTAACGACGGAATGCATTATTGTGGATAGTAAAAATGACT
 ACCTTTGATCTACTGGTATTAACGGTTAATTTTTCGAATGTAATAGTACGAAATTGCTTGTGCTTAAATTT
 TATGTAGTGAATTACGCTCATGAAGATAACTACTTCTTTCAAATACATGAAAAACGGAGTTTTACTAGATT
 ACTCTTGATGTTTTCTTATAATCGCCCAAGTATTCAAACAACTTAATTTACAGAAAAACCAACACACTA
 CA

secA

GTTTTCAATTTTTAAGAGTTATTCGAACTTCATTGCTCATAAGATTTTTATCGGTGACTAGCGAAA
 TATAAGGATCTCCGGCTACGACTTTAATTTTTGGGAAGTAAATCAATCAAATAATGGCTATCTGTC
 GATTTTCCAATACCATAATCATCCGTAATTTAAACGCCACTTACGTTTGATCCATAACATTTAA
 GAGTAGCGAACTTTTTCATTATTGTTTATTCAATGGTAGAAAGAACTGGAAAAATACTAGTGAGTG
 AAGAACTCACTTTTTGAAAAGGTAAAAATATTTTACGTTTGTGCTGTTACTCGATAAAAAACT
 TAGTAATAAGTGCAGTTTTTTCTTCTAATTTAATTCAAATTTCCAAAATCTTAAATGTACATA
 TCATATTGAAAATGCTATTACCTACCAGTTGTTTAAACCGCAGATAAAAACGTCCAATCGAAAAATA
 ATGGTTTGCTTGATTTTTATCAACAAATTTCTCGAAAATTTTGATCTGGCAAATACAAACGTCAG
 GTTACGCGTATAAATGTGCTAGAACCGGTATTTTCCAAATATTGCAATCAAGATTTACGTGATAA
 AACAGTCCAGTTGAAAACCTTTTTGATGTCAGAAAACAAGGGCCACATCCAATGTTGTAGAAGAGG
 CGTTTGCTCTGGTTTCGGGAAGCAAGTAAGCGAGTTCTGGGATTACGACACTTTGATGTCCAAATA
 ATCGGTGGATTAGTAATTAATGATGATAAAATTTGCCGAAATGAAAACAGGAGAGGGAAAGACTTT
 GGCCTCTATTCTACCTGCATTCTTAAACGCATTGTATGGAAAAGCGTTCACATCATAACTGTAA
 ATGAATATTTAGCTCAACGGGATTCGCAGTATACTGGGCAGCTACATAACTTTTTAGGACTGAGC
 GTAGGTCTAGTTACAGAACTATGAGTTCTGTTGAAAAGAAAAAAATTAATCTATTGTGATATAAC
 TTATGTAACAAACAGTGAACCTAGGATTTGACTATTTACGAGATAATTTGGTCCAAAATGCAGAAG
 GTGTAGTGCAGAGATCTCTTTTTTATTGTGTCATTGATGAAATCGATTCTGTGTTAATCGATGAC
 GCAAGCACTCCCTAATAATTGCGGGTAGTCCTACGTCATTAGATGATGGGTTGACAGAAAAGTA
 TATTAAGTCAACTCGTGTGGCAAACATTTTAAAACGTGATAAAACATTACAAAACCTGAAGAAAAGA
 CGAAAAGTGTTTTTATTCTCTGACGCCGGAATATCCATGTGTGAAACATGTCTTGGAATAGAGAGG
 AGTGAGATGTATTCTTTACCAGATCCTTTCTGGTTCATTATTTTAAATGCAATAAAAAGCAAAAAGA
 ATTTTATCGTCGTAATAGAAATTATATTATTGATGCAAATCTAGAAATCGTTATTCTTGATGAAT
 TTACGGGTGCGCATTAGAAGGTTCGACGCTGGTTCGGAGGGAATGCAACAAGCGATGGAGTCTAAA
 GAGCGATTAACCATAAGTGGTTTTGACACCGCCGAGGCATCTATCAGTTACCAGAATTTATTCTT
 GTTGATGAACGTCTATCAGGTATGAGTGGTACAGCTGCAACAGAAAAACAAGAAATTTAGTACTA
 TTTATGATTTGCTGTCGTGGTGATACCAACAAATAAGCCTGGAAGACGAATAGATTTTTCTGAT
 TCGGTCTACACAAAACAACCTCTATAAATGGCAGGGTGTCAATTAACGAATGTCGTGACATGTATCG
 GGTAGGTGACCTGTTTTAATTGGGACAAAACCTATTACCCTTCTGAACGTGTGGCAGATTTTTCT
 TTAAATCACAGAATATCCCGTATCAACTTTTTGAATGCGCGTCTGAAAATTTCTCGACGCGAGTCG

GAAATTATTGCCAAGCTGGTTGTGAATCGACGGTCACCATCGCGACCAATATGGCAGGTCGAGG
TACTGATATTATTTTAGGAGGCAATTCCCAATATTTTCGCATCGACTACGATCGCGACATTTGTCTG
AAAACATAAACTTTTTCGAATGCGAGTATGAATTCGGCGTTACGAAATTATCACCAAGAATATGTT
CGTTCCTTTACAGTCTTCTAATGAAAAGAACTAGACTTAAATCTACGCCGTCTTCGAAAATTCTA
TAATCTTGTTTTATATAAAAAATCTTTATTAACCTCAGAAAAATCGTAAAACGTCTTACGTTTAG
GTGGACTTCATGTTATAGGTACTGAACGTCATGAGTCAAGGCGCATCGATAATCAGCTTCGAGGT
AGAGCAGGACGTCAAGGTGAGCCAGGGTCATCTCGCTTTTTTACATCATTAGATGACGATCTTTT
ACGGCAATTTGGGTCCGAAGCACTACAAAACCTTTTATTCTCGATTGGGCGTACCAGAAGATATAC
CTCTTCAATCAAACACGTTGACTACTCAATTAGATCAAGCGCAGCAAAAAGTTGAAAATTCATTT
TTTGACGCCAGGAAAAGACTTTTTGACTACGATCAAGTTTTAAATTCACAACGTAATTGTGTTTA
TAGCGAGCGAACACGGCAAGTTCAATCTCCAGATTTACCTAACTTAGTACGTAATCTCTAAAAC
TGACGATAGATGATTTATTTAAGTGTACTTTACTTAGTAAGTTTTTTACTTCCAAATCTGGAAAT
GGGATTTCAAACCTTTCAAACAAGAAATAAGAGAGATATTCTACGAAATTTTAGAAATTCATTT
AATGCCTTACCCCTTTTCTGATAATAGTATGACGGAACAAATAAGCAAGTTTTCGATAGAATAA
AAAGTCTAAGTAATACGTCACTTTTTCTTGACATACGCCAATTCAAGTAGGCCTATACCGTTATTC
GTAAATAAAATTGCGTCCAATGTGACACTGCGGGAATTGGATACTGTATGGTCCAATCATCTTCA
AAGAATGAGTGCTTTAAAAGAGTTGACGCAATGGTCTGCTTATGGTCGGAGAAATCCATTAACAG
AGTATAAAAGACTTAGCTTCTATTATTTCTTGCAAAATGATTAACATTTTTTCAGCAGAAGGTTTTA
CAAGCAGTTCGGGTCTTTTTGAAAGCATCGTAAATTTCTTAATTTTTGAACTATTTATAGA
GTTAATCAAGACAATAAGTACTTGAAATCGCGTTTTAGAAAATGGTCTAAAATTTTTCAAAGTCGG
AGTATGACTACATGGCGAATGTAGCCAAGAGGTTACGGCACTGGGTGTGATTTTCAGCATCCGCG
GGTTCGAATCCCGTCATTCGCCCATTAGCCAAGTTATTTCCGTAAATACTGTTATTTGACGAATT
AAGCAAGTTGAGGAGATAATTAATTTTTTACATCTCTTGATATCAGACTAATCGAATATTTTTAAC
GGAAGATTAATAAAAAGTTAGTTTTAATTTCTTGTAATAATAAAAAACGCGATCTTTTCAAATG
GTGGGTAATTTTGTACGGTTAATCTACACATTTAAACATTGCT

secY

GGAAGAACTCCATTACATACTTTTTAGAGCTGATATTGATTACGCTTTATCCGAAGCACAACTAC
TTATGGTAAAATAGGTATTAAGTTTGGATATGTAAAGGTGAAGTTTACGGTAAAAGAGATTTAT
CTCCAAATATTGGCGCAACTAAAGGCGGAAAAGGTGGAGGTAATAACCAACAAAGAGCACCAGAGA
AGAAAAAATAATACGATTAAGGTTAGATAAGATGTTACAACCAAAACGTACAAAATACAGAAA
GACACACAAAAACAAAGGAGCCGTTAAAAGAGGTGTTACTAAAAAAGGAGCATCTATAGCTTTTTG
GTTCTTTTGCAATTAAGTCTTTAGATACAGCATGGGTAAGTGCACGTCAAATTTGAAGCTGCAAGG
ATTGCTGTTACAAGATATATGAAAAGAGACGGGCAAATCTGGATTAGAATATTTCTGATAAACC
GTTGACAAAGAAACCGGCAGAAGTAAGGATGGGTAAAGGAAAAGGTAATCCGGAAATGTGGGTAG
CTACAGTTTTTACCTGGAAGAATTTTATTTGAAGCAGACGGAGTGCCTTTAGAAGTTGCTAAAGAA
GCAATGAGATTAGCAGCAAGTAAGCTTCCGGTAAGAACTAAATTTATTGTTAGAAGAGATTACTC
TGCATAATTTAAAGAAATAGTAAAACAATGAAACAGGCAGATATAAAAGGACTTTCAAATGACGA
CTTACAAGAGAAGTTGTTAGAGTCGAGAGACAGTGTAGCAAAGCTAAGACGTAGCCATGCTATAT
CTCCACTTGAAAACCCTGCGCAAATTAACCTTGCGAAAAAACAGTTGCTAGAATATTAAGTACTGAG
CTTAAAAAAGAGAGCTCGAACAACCTTAAAAATACTAAATAAGTATGAGCGCAACAAGATCACTT
AGAAAAGAAAGAACAGGTGTTGTGGTTAGCAACAAGATGGATAAGTCCATTGTTGTTTCTGTAGA
GAGAAAAGAGAAACATCCGATGTACGGAAAGTTTGTTAAGAAAACATCTAAGTTCATGGCACACG
ATGAAAAAACGAGTGTAACATTGGAGACACTGTGATGATAATGGAAACAAGACCATTAAGCAAA
AGCAAATGTTGGAGATTAGTAAACATAATTGAAAGAGCTAAGTAACCATGATACAACAAGAAAGT

AGATTATTAGTAGCAGACAATAGCGGAGCAAAAGAAGTGCTTTGCATCCGTGTATTAGGCGGTAC
TAAAAGAAGATACGCTTCCATTGGTGATACCATTGTAGTAAGTGTGAAATCAGCTTTACCTTCAG
GAAACGTAAAGAAGGGAGCTGTATCTAAAGCGGTTGTCTGTAAGAACAACAAAAAGAAATCAGAAGA
AATGATGGTTCTTACATCCGTTTTGATGATAACGCTGTTGTTCTTTTAAACACTCAAGGCGAGAT
GAGAGGTACAAGGATTTTTGGACCGGTAGCCAGAGAGCTTAGAGAAAAACAGTATATGAAAATTG
TTTCATTAGCACCCGGAGGTGCTTTAATATAAATTGAGTCATGAAATTTTCATGTAAAGTCAGGAGAT
AAAGTAGTAGTAATTGCTGGAAACAACAAGGCCAAAGAGGGTACAATACAAAAAGTATTAACCAA
AAAAAGCCGTGCAATTGTTGAAGGTGATGACATTAGAAAGGTTAAAAACATACAAGACCTTCTG
CTGCTTACCCGGAAGGAGGAATTGTAGAACTAGTCCTTCTATCCACATTTCTAATTTAATGGTA
ATTAACCTACTACAGGAAAACCTGCTAAGGTTGCCAAAAAAGAAATGAAAAAGGTAACTAGA
AAGGTATTTTAAAGAAAAAAATCTGAAGGAGGAAATAATAATGGCTTACTCACCTAGATTA
AGATAAGTACAGAAACGAAATCGTTTTCTAATCTAACTAAAAAGTTTTAGTTACTCTACAGTAATGC
AAGTTCCTAAGTTGCAAAAAATATGCATCAACCAAGGAGTTGGAAAAGCTGTAATTGATAAGAAA
ATTGTAGAACTGCTCTAGAAGAAATGACTACTATAGCCGGACAAAAAGCTGTTCCCTACAAAATC
TAAGAAAGATATCTCAAACCTCAAGTTAAGAAAAGGAGTTAACGTTGGAGTTAGAGTTACACTTA
GAGGAGAAAAAATGTATGAATTTTTAGACAGATTAATTGCTGTTTCTTTACCAAGAACCAGAGAC
TTTAGAGGAATTAGCCCTAAAGGATTTGACGGAAGAGGTAACTATACACTTGGTGTAAGAGCA
AATTATCTTTCCGGAAATCAATATAGACCAAGTTAAATACATGAATGGAGTTGATATTACTTTTTG
TTACTTCAGCTAACTCTGACGAAGAAGCAAAAGCCTTATTGACTGAATTCGGTTTTACCATTTAAG
AAAACTAAGGAGGAAAGTATGGCTAAAGAATCAATGAAAGCTAGAGAAAGAAAAAGACAAGCGT
TAGTTGCAAAATACGCAGCAAAAAGAGCTGCACTTAAAGAAGCAGAAGATTGGGAAGCTTTACAA
AACTTCCAAAAAATTTCTTAAAGTAAGATTGCACAACAGATGTCACTTACAGGAAGACCAAA
AGGTTATATCAGACAGTTTTGGCCTTTCAAGAAATACATTTAGAGAAATGGCTTCAAAGGTTTTAA
TACCTGGAGTAAAAAAGCAAGTTGGTAAAAAATATATATTACAATTATGACTGACGTAGTAGCA
GATTATTTAACTAGAATTAGAAATGCAGTATATGCCAAGCACAGAGTGGTGGATATACCTGGCTC
TAATTTAAAGAAAGAATTAACCTAAGATATTGATGCAACAAGGTTTTATCTTAACTATAAAGTAG
AAGACGATAAAGGACCTCAAGGAAACATTAATAATGCTTAAATATGATCCGGTAACTAAAGTT
CCTGCCATCAAAAACCTTAGAAAGAATTAGCTCACCTGGTTTTAAGAATTTACAAAAGTGTAGATGA
ATTGCCAAGAGTTATTAATGGATTGGGTATTGCAATTGTTTTCAACATCAAAGGCTTAATGACTG
ATAAGCAAGCTAGAAGAGAAGGCATTGGTGGTGAAGTAATGTGTACGTTTTATTAATTTATAAGA
TATGTCACGAATAGGAAATTTACCGATAGCAATTCACAAGGTGTTCAAGTTTTAGTGAACAACG
GTGAAGTAGTTGTTAAAGGTCCGAAAGGCGAATTAACAACAAAAAATGACAATGCTATTGGCATT
AAAGTGAAAATGATGAAGTAGTTGTTTTCAAGATCTACTGAAGAAAAAGACCACAAAGCCAAGCA
CGGTTTATACCGCTCACTTATCAGTAATATGATAGAAGGAGTTTTCTAATGGTTACAACTAGAAC
TTGAATTAGTGGGGTTGGTTACAGAGCAACTGTTAAAGGACAAATATTAGAATTAGCCTTAGGT
TACTCTCACACATTTGTGTTAGGATTACCAAAAAGGTTTTAGTAGCTGCGGTTCAAGAAAAAGG
TCAACCACCTATTATTACTCTTGAGTCTATTGACAAACAATTAGTAGGTCACGTTGCTGCTAAAA
TCAGGTCATTTAGAAAACCTGAACCTTACAAAGGAAAAGGTATCAAATTTAAAGGAGAACACATT
AGAAGAAAAGCTGGTAAATCAGCATCTGCTAAATAATAATAAATGAAGGGCTATGGCACTTACA
ACTAGAGAAAGAAGATTAAGAATTAATAATGCGTGTAGAAAGAAAAGTCTACGGGACTGCTGACTC
GCCAAGAATGTCTGTTTTTTAGAAGCAATAAAGAAAATTTATGTGCAACTAATTAATGATGATGATT
CTAAAGTAATAGTTTCTTCATCTTCAAGACAAAAAGATATAGCTTCTAAAAAAGGAACTAAATCA
GAAATAGCTGCAATGGTAGGAAAAGATGTTGCCGAAAAGCTATTAAGCTGGTGTACTACTGT
TGCTTTTGATAGAAAACGGATACTTATACCATGGTAGAATTAATAATCTCTTGACAGATGCTGCTAGAG
AAGGTGGATTAATAATTTAATAATAATAATATTATAATAGAATGGCTAGCGCAAACGTACAAAG
AGTAAAGTCAAGCGAAATAGAACTTAAAGACAACTAGTAGCGATAAACCGTGAACCTAAGGTTA

CTAAAGGTGGACGTCACCTTTAGTTTTCTGCAATAGTTGTTGTAGGTAACGAAAGAGGTATTGTA
GGGCAAGGGCTTGGAAAAGCAAATGAAGTTACTGACGCAATTTCTAAAGCAATTGACGATGCTAA
GAAAACTTAATTAAGGTTTTCAATTCCTAAAGGAACTATTCCTCATGAAGTTGAAGCTAAGTTTA
GTGGAGCTCAAGTTTTTATTAAGCCTGCTGCAAATGGTACCGGAGTTATTGCCGGTGGTGCAATG
CGTGCTGTGCTTGAAAGCGTTGGAGTTACTGACGTACTAGCTAAATCTAAAGGTTCTTCTAACCC
TCACAATGTGGTTAAAGCTACTTTATTAGCACTAATGAATGTTAGAGATGCTTCTACCGTAGCAA
AACACAGAGGCGTAGAATTAGAAAAAGTATTTAACGGTTAATTTACTGGGCAATGGCTAAAATTA
AAATAAAACAAGTACGAAGCATTATTAAGACCTAAAACACAAAAGCTTACTATGCAAGCTTTA
GGTCTTAGAAAAATTAACCATGTTGTTGAACATGATGACACTCCACAAATTTTGGGGATGGTTAA
CAAGGTAAAACACTTAGTAGAAATAGTAAATAATATTAAGAGATGAACTTACATAGTTTAAAC
CGGCAGCTGGTTCTACCAAAAACGGAAAGAAAAGATTAGGAAGAGGACAAGGTTCCGGAAAAGGT
GGTACTTCTACAAAAGGTCACAAGGGAGCTAAATCCAGATCTGGATATAAATCTAAAAGAGCTTT
TGAAGGAGGGCAAATGCCGCTTCAAAGAAGAGTACCTAAATTTGGTTTCAACAACATTAATAGAG
TTGAGTACAGACCCTTAATCTTGATATTATTCAAGAAATAGTGGATAAAACCAAAGCAGAAAAA
ATTGACTTTGCCTTTTTAGTAGAAAACGGATTAGCTAAAAGAAAAGACTTAGTTAAAATTTTAGG
TAGAGGTGAATTAAGCAAAGTAGAAGTTGAAGCACATGCTTTTTCTGCTTCTGCATCTAAAG
CTATTGAAGCTAATGGCGGAAAAGCCGTTATCATAAAATAAAAATGAAAAGATTTTACAGACCA
TAAAAACATTTGGAGTATTGAAGATCTTCGTAGCAGGATTTTAACTACCCTAGGACTTATTCTT
ATTTACAGAGTTGGTTCGTATGTAGTTTTACCTGGCGTTGACTCTGATGCTCTTGCTGCAGCTAA
TGCCGGAAGCGGTGGAGGAATTACAGACCTAATCAACCTATTTGCGGGTGGTGCTTTTTCTAGAG
CATCAATTTTTGCATTAGGAATTATGCCTTATATCTCTGCATCCATTGTAATKCAGTTAATGGGA
ATTGCGATTCTTACTTTCAAAAATTACAAAAGAAGGCGAAAGCGGTAGAAGAAAAATCAATCA
AATTACCAGATACCTTACTATTGCAATTACTGCAATGCAAGCTCTAGTTATATAGCTACTCAAG
TTCCAAATGAAGCTATTTATAATGTTGGTTTTTGGTCGTTTTACTTTCCCTGCTGTAATTAATTTG
GTTACAGGAACCTATGTTTCGTTATGTGGTTAGGTGAAAGAATTACTGAAAAGGAATTGGCAATGG
TATTTCAATTATTAATTATGATTGGTATTATAGCTAACTTACCATTTGCATTTACTGCTGAATTTG
CTTCCAGACTTTCAAACGGAGGTATGGTGGTATTCTTAATTGAGATGGTAGCATTAATACTAGTT
GTAATAGCATGTATCATGCTTGTACAAGGAACACGAAAAATTCCTGTACACTTTGCTAAACGAAT
AGTTGGTAAAGGCTCTAAAATGATGCAACAAGGTGGAGTAAGACAATACATTCCTTTAAAAGTTA
ATGCTGCTGGTGTAAATGCCAATCATCTTTGCTCAGGCAATTATGTTTGTGCCTATTACCTTAGCA
GGATTTTCTGATTCAATGAGTGGTATTGCAGCAGCATTTCAGACTTTACTGGATTTTGGTATAA
TTTTACTTTTTCGTTACTTTGTATCGCTTTTACATATTTCTATACTGCAGTTACAGTTAATCCTA
ATCAAATTGCAGATGATTTAAAAGAAACGGAGGCTTTATCCCTGGTGTAAACCTGGAAAAAGA
ACCTCTGAGTTCATTGATAATGTAATGTCAAGAATAACCCTTCCGGGTTCAATATTTTTAGCATT
CGTTGCAATATTACCTGCATTTGCCGGTATTTTAGAAGTAACTCTCAATTTGCTCAATTCCTTG
GCGGAACGTCACCTATTAATTATGGTGGGTGTTGTGCTAGATACATTGCAACAAATTGAATCTCAT
TTATTAATGAGAGAGTATGATGGATTAATGAAATCAGGTAGAATTAAGGAAGAAGCAGTACAGG
CGGAGTAGGAATGGCTGGAATAGTATGGAAAGATCTATCTGAAAACAGAAGATGAAATTGAATTA
ATAAGACAAAGTTCTTTACTTGTGGTAAAACCTTAGCAGAAGTTGCTAAACACATTGAGCCGGG
TGTAACTACACTTGAGCTTGACAAAAGTTGCTGAAGAGTTTATTCTTGATCATGGTGCAAAACCTG
GATTTAAAGGATTTACCGGTACGAATACACCCTATGTACATCTGTAATAGTGCAGTTGTTTCAT
GGCTTACCTACTTCAAACCCACTTCAAAGTGGAGATATAGTTTCAGTAGACTGTGGTGTTTTAAAT
GAACGGATACTATGGTGATGTTGCCTATACTTTTATGGTAGGTGAAGTTTCTGAAGATGTTAAGA
AACTCTTGAAAGTTACAAAAGAGTGCCTTGACCTTGCAATAGAACAAGCGGTTGCAGGAAATAGA
AT

APPENDIX B: TRANSLATIONS

petB

LASGINFLQERLDLQSLSDDVSSKYVPPHVNLLYCIGGLTLTSFLLQLTTGFSLTFYYRPIVASA
 FATVSFLSSTLHYGWVIRAAHRWTASLMVLALLGHISRVTYLTGGFKRPRELTLWTGLALSCLTVS
 FGVTGYSLPWDQVAYWACKIVTGVPDAIPLIGGQIVALLRGADTVGQSTLSRFFETHFTLLPALT
 TVLMLTHFLLIRKLGISGPQ

petD

MSNVNPLPSQAKSRAKLCCKGAGHNTFGEPAWPNDLLFLFPICITGVLACSLGLAVLSPPAIDAQS
 DPFSTPLEILPEWYFLPSFNLLRLLPNKLLGVAAMASIPALATQVAASENQATASQNPNRRSQSSL
 LYILGSSSSSWLGAGGAMPLSQSLTLALN

psaA

RFPTTFSAWRDPGHFSKTLAKGPKSTTWIWNLHADAHDFDAHTNSLQSLSRKIYSAHFGQLSLIL
 LWLAGMWFHGARFSNYIAWLSDPGTGLTPMAQSVWPVVGQDLLNSDGGGTLGLPFTAGWFLWRA
 CGITSEAQLFSTACSGLVLAALTAGGWYHYHRSAPSLDWFQHAQTMLVHHLTGLLGLGCLSWAG
 HQIHVSLPINSLLDAGVDIRILPTPAALLFDRGYLENLYPSFKQGLTPLLSGHWAAYTDFLTFKG
 GLNPITGGLWLTQVAHHHLALALLFLIAGHMYRTSFPIGHNISAIHAAHKGPLTGAGHNGLYTL
 TTSWHSQLALNLALLGSTTILVAHHMYAMPPYPFIALQYATQLSLFTHHMWIGGFLIVGAGAHAS
 IFMTRDYTPASPNGFSDTNLLERVLLHRTTILAHNLWVCIFLGTHSFGLYAHNDTMRALGRPTD
 TFGDGSIALSPIFAKCIQALHSVAPGNTAPHAISVSSYVFGAGTIAVNGRIAIQPLYLGTADFMI
 HHIHAFQIHVATLILLKGVLFARSSGLIPDKYRLGFRFPCDGPGRGGTCQVSPWDHVFLGLFWMY
 NCISITIFHFYWKMQSDVWGTITSSGRVSHITNGNFSSGSLTVNGWLRDFLWSQSSQVIQSYGTP
 ASAYGLIFLGAHFIWAFSLMFLFSGRGYWQELIDTLVWAHRKVRLSPSISPRALSITTRSVGVA
 HYLGGIGITTSFFLARATTT

psaB

KGSEWNQLFLVVGDSRLRTFPSFSKGLRQDPTTRRLSYSVGIAHDLESHDFVTESLLYQRIFASH
 FGHLAIIFLWSAGNLFHVAWQGNFQEWIINPLKTSPIAHAI FDPHFGVDALKAFTPSAGSYPVNI
 STSGLYHWWYTI GLRSNTDLYSASMFLILCSLLFLYAGWLHSTPGRQPNLAFFKNNESRLNHHLS
 GLFGVSSLAWAGHLIHVAIPASRGQVVSLDHLPLPHPAGLTPFLTFNWGIYAENPDSPTHVFATT
 EGAGTAILTFLGGFHPQTQAMWLT DIAHHHLAIAIIFIVAGHMYRTTYAWGHRFTELLHAHVPPK
 GRLRAGHVGLMETLTNSLHMQLGLALAALGVATSLTAQHMYALPAYAFIASSPVTQAALYVHHQY
 IAGFLMTGAFAHGAIFFVRDYEPSNEGNVLATVLLHKEAII SHLSWVSLFLGFHTLGLYVHNDTC
 VAFGLPENQILLDPIFAQLVQAASGKAGNGIDILLSEPTSPASIASKVVLPGLWLAGINSQRSTA
 FLPIGPGDFLIHHAISLGLHTTTLILVKGALDARASRLLPDKKDFGYSFPCDGPGRGGTCVSAW
 DAFYLAVFWQLNTTAWTTFYWHWKHLTLWGGNSALFNEASTYIMGWLRDYLWLN SAPTITGYTAQ

GMNAQAVWTWMFLFGHLTWATGFMFLISWRGYWQELIETIVWAHERTPLANLLKWNKPVALSIV
QARLVGLTHFTTGYIFTYAPFLISSTS AKTAGISTAAALFKDLPGEII

psaC

MSHTLRIYTTICIGCTQCVRACPCDVLAMVPDTVWNKSGQVAAAPRPADCIGCKRCESACPTDYL
IRVYLGPESTRSMGLTY

psaF

MLPAATDSRIWQQKFATLALSCMIFLPEKAQSSTPLIVMCKDSAAFNNLATTRLTRLNTALS
NFEPRTPAYRATMERKRAAKNRF SRQALLMCGRNDGLPRLFSADLNTFYNSFAVPALQFLYIATCIGW
AGREYLSRTANYKSEILIDLTLAATLMGSSPFRAVGAAWDDLLRGKLLNTTLQ

psbA

MNYYNPKSSLNIRQRQQTSSAWEVFCSWVTTTENRLYLGWFGCLMVPTLVSATFAYIIGFIAAPP
VDIDGIREPVSGLLYGNIIISGAVIPSSNAIGVHFYPIWEAASLDEWLYNGGTYQMVVHFHFLG
VCSYLGREWELSYRLSMRPWICVAFSAPVAAASAVFIVYPIGQGSFSDGMPLGISGTFNFMLVFQ
AEHNILMHPLMHFGVAAVFGGSLFSAMHGSLVTSSLLKETTGSSESLNYGYRFGQHNETYNIVAAH
GYFGRLIFQYASFNNSRSLHFFLGAWPVIGIWLTSMGVSTMAFNLNGLNFQQSILDACGCP
ISTWADILNRANLGMVMEHNAHNFPLDLASTSNNNTNTTWPITLG

psbB

NLSKSGGLVALPWYRVHIVVNDPGRLIAVHLMHTALVSGWAGAMALYELAVFDPGDPVLNPMWR
QGMFVLPFMTLRLGVTQSWGGWSITGESLGGKSVGLWSYEGVALSHIGLSGLCFLAALWHWR
FDL ELFRSPLNPTQTSLDLPTIFGIHLLLASLLCFGFGSYHVTGTYGPGI WVSDAYGICGRVQGVSPS
WGAAGFNPFNPGGIAAHHIAAGVLGTLAAVFHITNPPPEVLFLALRMANIETVLSSSIAAVFFAA
FVTSGMTWYGAAATPIELFGPTRYQWDSEYFLKKITQTVNYYKGLSLTDSDAWALVPEKLAFYDY
VGNNPAKGGLFRAGPLTKGDGVAKAWLGHPQFYGPDGSILRVRRIPAFFETLPVLLDTRALVAA
DIPFRAESKFSIEQIGVSAVILGGRDAGAEFTSAASVKALARKAQFGELFEFDRAKLKS
DGVFRSSPRAWYTYAHLNFAFLFLLGHLWHGGRTLFRDVFAGIGADSIEQVEFGAFVKLGDARTRKKPRQ
LKS

psbC

MECPRTLES LHLGVYLLKTYRGKNTRTENNGIGRRAESKNSKWNHAENSSAAGRDLTTGYAWW
SGNARLISVSGRLLGAHVAHAGLMILWCGAMTLFETAHYIPELPLYEQGCILLPHIATLWGVD
P EGEISGTYAAAFVIGVLHLLSASVVGCGGLYHALLGPDITLEENFAYYGYDWRDKNKMTTILGFHLI
LLGLGCYLLVAKALYFGGIYDTWAAGGGDVVRVTAPTLPNPIIIANYALKSPFGGDG
WVIVSNNLQLDVGGHAWMGSILILGGLFHQNTKPPAFVRRAFIWSGEAYLSYSLGALTICGLTAAV
FVWYNNTA

YPSEFFGPTAAEASQAQAFITFLVRDQKLGAKVASAQGPTGLGKYLMRSPSGEIIIFGGETMRFWDM
 RAPWVEPLRGPTGLVLRIRRDIQPWQVRRAAEYMTHAPLGSLNSVGGVATEINTFNYSVPRSWL
 CTSHFFLGFFMWIHLWHAGRARAAGAGFEKGISRENEPVLVYMRPLD

psbD

MKFGNSKAKSAQTAMRRKVSLSLVDDWLKRDRFVFGWSGLLLFPCAYLAVGGWLTGITFVTSW
 FTHGLASSFLEGCNFLTAAVSTPPNCMGHALLLLWGPELSGDIIRWFQAGGLWTFIALHGSFGVV
 GFCLRQFEIARLVGIRPYNIAIAFSGPIAVFLTVFLVYPLGQASWFFAPSGVAAIFRFLLLFIQGF
 HNYTLNPFHMMGVAGILGGALLCAIHGATVENTLFEDGDAATTFRAFTPTQAEETYSMVTANRFW
 SQIFGVAFSNKRWLHFFMLFVPVAGMWTSSIGIVGLALNLRSYDFVSQELRAAYDPEFETFYTKN
 QLLNEGIRAWMSAQDQPHENFIFFPEEVLPRGNGL

psbE

MAGSTGERPFTDIITSVRYWIIHSVTIPSLFVSGFLFVLTGLAYDAFGTPRPSEYYTEKRNAAPR
 LTERYGAKAELTAIAAGMQEE

psbF

SLKPVTGAVTYPIFTFRWLTIIHALAVPTVFFLGAITCMQFIQRA

psbH

HPFLYKVFSYQQTAATGRGLKTKLGRALLPLNSQAGLVTAGWGSTGFMVFTMVAFLFLTIILQL
 YNSTILLDDFELSWADLPTG

psbI

MYRLKLAVYLVVIFFLTIFTLGFLTGDPSRNPNRKDIE

psbN

TLILFISLRESSLVILVKVYFRVVRVRKLSYVGTLVRLKLFADFSKLEAAAFVTYFVLGLLVG
 ITGYSIYTAFGAGSSNLRDPFEEHETTEAITLHTPR

psbT

MEAILYTVLLIGTLFVIFFAIFFRDPRLVRK

psbV

MPKKQIFSRQVRTNKQVQASLLFSLVFLASPEFAFARKFTDSVRTVQLNAARADVLSPNQVRRG
 KFLFGKACASCHVGGLTKPNPNVGLDIKSLQVARPPKNNVANLIAYIIAPTYYDGLTDISDIHPC
 TKQSRLYTKVRSLTRFDCFCIGGHVLLQAKLLGEKWGGGKVYY

psbZ

MPALAKDDDDDDSGVKAPPLRTDAVLQARDKYEFGGRKYKADTKLMIIDNMKIVTDMGRGTPQMEDI
 AKATRKQMIIEYISLYRARPETTKLPSFSVMTAINTVAGHYASYGYKYPIPEKRRVRLAEQYLEV
 ERALARGK

rbcL

MLIKKFEDEVNSRTRIKAGRYKSGVLAYKDMGYWEPDYSIKTTDVLATFRFQPQPSVHPIELAAA
 LAGESSTATWTVVWTDLLTASDIYRAKAYKVISLQDSSALAYIAYDLDLFEEGSIANLTASIIIGN
 IFGFKAVKSLRLEDMRI PVAYLKTFQGPATGLVVERERLDKFRPLL GATVKPKLGLSGKNYGRV
 VYEGKGGDLFLKDDENINSQPFMRYRERFLYSMEGVNHAATAKTGEIKGHYLNVTAAATMEDMYER
 ANFATELGSVIVMIDLVI GYTAIQSMAYWARSNDTLLHLHRAGNSTYSRQKNHGMNFRVICWMMR
 MAGVDHIHAGTVVVGKLEGDPKMIKGFYETLLGLSTPVNLSLGLFFAQEWASLRKCVPVASGGIHA
 GQMHYLLHYLGDDVILQFGGGTIGHPDGIQAGATANRVALEAMVAARNVGLDYVALGKSILTNA
 KRCDPLHAAISVWKDIAFNYESDTDPDVGIEAAHSSSSTSQHASQHDVAPS

rbcS

MITQGCYSYLPPLTPQQQLVMQIDSALKRGLGLGIEYTYDPHPRNSYWEMWGNPRFTFEGVQSIVY
 ELAACAKAHPEAYIKLLAFDQARGTESCVLSFIYHRPSVETAFRLVLRQEANSRKIVYTLEPTTP

Rp13

MPLTAVGLFGKKLAMTQYACDRSHSIATFVPATAVEIGANQVTQIKTVQTSYGGALQVAYNEVKP
 RVLSKSCFGKLRKAGSSRFFKNFGEFRVAQPEKYRIGECLDVRLFHPGQPVTVVAPGSGKGLGT
 VKRHNFSRGPMTGSKNHRLPGSIGSGTTPGRVYPGKMGAGRSSGRSRVRSVLETDENNNIVLL
 RGSAPGKRGHFLKISAEK

Rp16

MSRIGKQPVELPSGVSASVSGQTVVEVKGPKGARSFTATDDITITVEDNVI TVTFRGKSKRARQQW
 GMSRTMVANLVHGVNTNFKKELEIQGVGYRAQLQGNVLLKLNGLSHDVFQIPDGVTVTCPKQTE
 IVVEGQDAQLVGQVAANIRDWRRPEPYKGGKIRYKGEFIFRKEGKKK

Rp119

MSDLIKLIENEYNAVRESIPAFGAGDTVNVHVKIKEGNKERIQQFQGTVMQRRNSGNSETFTVRK
 IASGIAVERIFPVISPNIKIEVLRRGKVRRLRFLYLRGRKGKAARIKEKL

rp123

MKRGRGGKGGGKLIKISLGMGVASLMNCADNSGAKNLYVVAVAGIGGRLNKIPKCCPGDMVLCTVK
 KGKPELRKKMLKGVVIRQKKQWRRREGLFVYFEDNAGVIVNDKGEMKGSAITGPVAKECAELWPK
 VASTAPAIC

rp127

MAHKKGAGSSKNGRESHKRLGVKIFGGQDIVAGNIIVRQRGTQHHPGENVGMGKDHTLFALTDG
 VVEFKKGRQNRSFVSVVTPEAKA

rp131

MIALMPKPAIHPTWYPHAKVYLDGQLLKVGATKPKLEVDIWSGTHPAYTGRTEELDTEGRAKRF
 LRRYGWEEASLTEATIDANDSTTS

rp136 - weak homology

MKVRSAVKTICSGCTLVIRNRVRYVVCKKNPKHKQRQGFHTLANISPSEFPSPAPFVKGSIGQQN
 LNTAAALHGNLGQMRKIFWNALKIC

rpoA

MYNILRSPVKELMDPTGQSLVHGKYYVGPLLRGQSITLACALRRTLLAEIEGFAIIGIHVPNVKH
 EYEHIPGLKEDMLELSLNLKQIKITSESNTNTTLAKLKVSGSNKKSGLCAQDLILPEGYTVSNP
 RAYIASLTSTDVPDFTMELLIARGKGFVLAQDIRARLPTRFIPLDATFTPIPRVSFSFEVSRDNK
 YETLIFEVWTFVGFDPFELIAEASFILERTFRFRFLVFPFQSTIKRPVSLPTATKLESAPTELDQT
 SLRELKLTVRARNCLIRAGVKTVDVTRFVSHKSLQLQCGVHVSVEIEAKMLEYFNWKIS

rpoB

MNEAGQPRKRAKAEPEASKRQD GASNLSVLGQACANHVESFNALLEGGDLGIVSQMDVAEVDPRE
 GTNDPAIQIQLTELYVTKPTNMLDSKTDHDGGKFAGRTLNI LPVDCRVGHFTYAAPLHGK FELRV
 TGKQPQTASATLEIGRIPIMVKS KACYLHGKSPQQLVELGEDQTECGGYFIVNGNERVVRMLIMPR
 ANYPMAIIRPSYESRGHLYTKYATLFRGMRRDGSTQTNTLHYLLDGSCWMMRFSHGREWFIPLVA
 IAYCFYPLTDEL FVQQLCGTGEHWSMTFLRERALVMLQHQR TKHKISNQRDAYSYIGKTFRITLG
 LMI PRRTWNEEVGKWMNR FVLVHTQDPVEKLQTLCLMWQKLMGLVREQLEPDNQDVMSSHEILL
 PGQLYGLVLKESLETMMERLKAI IAKSTRPRDDVKA EK KATDYIKNERLLEQVIKAASKVSN DLQ
 YFVATGNVNSRSGLDLMQKSGYTIVADKLNQARFSSHFAAVHRGQYFAEMKTTTVRKL LPETWGF
 LCPVHTPDGSPCGLLNHMARS AKAVIHFP SAEITQSVTRFLSTLGV DVYPGEKTDLPTYQNAKYA
 WVMLDGRPVGR LAFDR LDSVAQELRRSKVAGRNGIPRDM EIVPISPDKYLFPGFLFLGPSRLI
 RPVHCHSLGTVEWIGPLEQVFMNIAMTGKETEEARVAVAAAKDSAGKELVQRDEDLPEQKPIEHT
 HQEIKPTEIFSVLAAALTPFSNHNQSPRNM YQCQMLKQTMGTPYHNHPYRTDNKVYKIWCPQKPIV
 RTEMYNQYNFDEHPQGTNAIVAVITYTYGDMEDSMI INKHSFERGFHGHGIVYKTKIIEAGDKNMR
 AQEKAACMFTNLRPEFKGGGRYVKDVPATGEFRLQDDGLPPIGIKMENG DPLCCWAGPDGREKV

VKFDHDDLEAYVENVTRVHPEALPGAQAEPMHKVLIKLRYPVGVGDKFSSRHGQKGVMSRLWP
 AEDMPFSESGVTPDIMFNPHGFPSRMTIGMLIESIAAKAAAEGKPVVDGTTFRGYAGHLTGKHN
 NEKDFLQKESNRKRQHNEKDGPNVYKYFGDTLVKHGFQRLGTEKLYSGIHGTEMETDIFMGVI
 YYQRLRHMVGDKAQARMRGPVDRLTNQPVKGRQKHGGIRFGEMERDSSLAHGTAYLLHDRLCRSS
 DFDVGFVCPMCESILTPQANAHLKTGLHSQIAVSGQEWECPPCSRKMKRSVICHVPVPIPIWIFRFL
 TCELAAMNVRLQIKVKSREGREASMLPRSDVTAADAKKAGSGAAAGSVKK

rpoC1

MTTAYDLNVPYSSAELKRVKAVRFTIFDADTIVACSVAEITELDIYKNGEPMYGGVNDPRLGPLD
 PRTPTCHTCGLGIKECPGHWGIVTLARPMFHWGFMKATHNVLRCVCYYCSRLADPKDHHMINAEA
 MKNNAKRLQAVMLCCRGKKRCAVAADAANPEEVLGRDGEEDPGKNFAGGCGYLQPKFTVDMTVLY
 VSFPEGGEDEGDRKRVMSAQEALVIFKRITDTRRGMGFDPKINHPSQMILTHMAIPPPAVRPSI
 QMGAARAEDDVTAKLSDIKVNLMKLTQVQSGAGDHICAFAKLLQYHIYTLSDNGIIGIPTATT
 KSKRPLKSIRERLKSKEGRLRGNLMGKRVDFCARSVIGGDANFNTEQVGVPRSIANLTFPERVT
 PHNIHWLKKLVAKGPTQHPGARYVVRDDGSRDLRYVADTSAITLQYGWKVERHMI DGDYIIIFNR
 QPSLHKMSMMGHRVKVMPFSTLRFNLAVTSPYNADFDGDEMNLHLGQSHETRSEIKHMMLNPRMV
 VSPQGNKPVMGIVQDALLATCKFTKRDSFLEKDLAMNIMWVVPWDGCLTIPTILRPKPMWTGKQ
 ILSMIIPKGLSLKRDAAIASKNKKDNPDFPASDCKVIVQDGEILAGQVCKKTVGSSGGGLIHLIW
 LDYGPEVCRNFISYLQKIVNQWLTHNGFSCGVADI IANDETLNVEKTLKQAKVEVRKILADAQR
 GKLETQPGKTMYSFEAKVNQRLNAARE DAGRIGSDSLDERNNIISMVNAGSKGSPINIAQIIAC
 VGQQNVEGSRIRYGFMDRTLPHFTKDDYGAESRGFVENSYLAGLTPQEVMMHAMGGREGVIDTAC
 KTSETGYIQRRLVKSMETLRTHYDGTSRNANNEIQFIYGEDGMDALWIEDQTLIEMTYDNKKLE
 NTFKHQYQSESYGDSWLPDPVQQAIIKASGENQQILDDEWKKIKDLKEMICNEVFPDGDQKHFPI
 NLARMMGRAKLKLAQENDIDRNDKFTPMEVVQKVNKLLSELEITRAIREGDSIGREVEDNAKIVL
 NGHLRGWLGSKKILQQEKLSKQSFDWLLGEVQKRFMKSVAFAGEMIGTLAAQSVGEPATQMTLNT
 FHFAGVGAKNVTLGVPRLKELINVAKTVKTPSLAVYLNGLDGLKQDQERAKDVQSILEHTTLEKVT
 FTQIFWDPDPENTRVEEDREWVAEYELPDDDENPNRCGPWCLRIQLSNKVMTDKKLTVREVGER
 ILQDFMGDLDCIFTDNAEELVLRIRLLKEAAEIDQAPPPFDPLDDKEDKDFKFLRSIEANILKE
 MSLQGIILGKIKVFMREDTMSVYNAVKGGFERTKEWVLDTDGVNFEEVMQIPQVRFERLQSN DIVE
 ILNLVGLIEAVRNALLFHVRMVISFDGSYVNYRHLGTLCDVMTQRGHLMAITRHGINRTNQGPLMK
 CSFEETVEILMDAAIYAEVDYMRVSENVI MGQLAPIGTGVDLYMDDKVNKDGDCMLNDARPVL
 PNSGKANEMFMSASPAGTPSRGSPIPALTPHADQVEPDFAPPSTEGSISGLTPNVTPNGSPESM
 AGGQSPFTPMDRSPGAARFSPAFASSPVSAGARTSPSYSPVLGYSPQADEPQNVSPTYRPDDG
 QGSSSPLYKTGAHGPPSSAVYSAPSYSPGPAGYAPSSPGYTPGRSPAYDYSMPSEFNQSPYQADFG
 ARTRTRPSPSTAPGDHSPTSPADTPVDGPEARRSFSSSPSYTPHPGWMEGTTSPQFNYPDPAAGGY
 ALDGEATDDNASDVFEPSDDEMVR

rpoC2

LQSNLKYSNKALDKKELKAFLTITAHSGATQATIISDKVKS LAFQSSTYSGASLSLQDFT PAPS
 LMHLIHLANAGARRANNDYCNVITFNERHKLITTSWRAALAPLTDALQTYMQRCDLSLSLYALA
 FSGARGNITQAKQLVGMRLIMDVQKLGELPIRTNLFSGMSTLEIAIASHGSRKGLVDTGLRTA
 DAGYLTRRLVDSAHDILVREEDCFGDDFLTLEELISQSKNRKISWGEKLVGRVLATPLYSPENSL
 VFTKNMIITKSSLNSLRRAGFLES AEKRRLINIRSPLTCTTLAGVCRNCYGSDLGTSKKVSLGLA
 VGVIAAQSIGEPGTQMILRTFHTGGAGITLGYTARDICSPSSGLVTYHNIQAALIVRFRDGLPSL

VLKEEGSLIITGKHSITQLTFPAGTAIFLANKSFVYRNQVIADLRVSRMRPRVSSQHSSASPLQ
 LKKPVMKKIVQERLPKSGDLFYCKASYTDRRLTHSSSPSSISSIHNPVQNYSSNLEDTRLSLWV
 LDGYRYQLTKLAKVLKPGKTFSRGTSIAHLKLINKFSGIPYYQKLSKDKGPLIINSLVCPINLFV
 KQSLENFQAKNRVLHLHTSRKIIPLHSHISNETIIRPGHVIFTPQLHPHWCLAKCGGIVYRIYEN
 NLLLLLPEERYYLRSFRELLIRPLAKSIYLRAGTKITSSIVMPTQGILQIDAKKLTLTIKPGFVM
 RPKTASIQSCGTLWEPTPVHRTWFRDSDKNNQDYSYVEKFILRKVSYVLERVPIKYVNLTKLTSY
 IRYNIISVPLVKIPRLYIKTIYSALYVSGERVFKHDILFATRVIDMSPAKRFLRTSTQSTPGITK
 LQLHPCCYEIKTSPNKEIYSADFFATNQYATSSRLVSQETNYKRLTQVVSQWDKIRLLTCEHIKI
 PKLAYTATNQDAIVLKNLAMPSTCVAPGTLISKVLLNASVSGQVVRQQLTPSHLEVITVRPCDL
 SCFLFNACEVLYVDEGDFVVLGTRLTSLVANCVGQVYRVSKTRIFIRKGMTRLISATQQFPLR
 DSIKSNFAKKGSMSTTWSELFEAKDITQGIKDISKILEVGKTSSTGLLVFNGIIVHASVLRV
 VSPNRRIWKFNI PDSLQNDILPASGTYVRRLDVLSAGFVSLRTRLQTLYYYYCHEEKFESRAAC
 QESIKDARLHLTEQVKSAYLEYGINIADKNLELIVRRMTSKVTILDPGATDLMIGEYIDFYKLMA
 IENAVVQKSLKKPRYEPLIFGLTKTALHSNSFLSAASFQSTVKVLTSAAIQGKIDWFIGLKEAVI
 AGRMIPTYQDRAGSALHLPVHSSLKTSILNLRKNVEQKGKEKI

rps2

MAMTTKEEDIQMMLVCKTHLGTNRCDHRMKRYVFRRTVDGIHIIHLGKTWEKLMMAARIIVAIEN
 PSDIIVASQRPYGSRAVLKFCQYIGAQAMAGRWMPGTLTNQITTKFLEPRLLIVTDPRTDSQAIK
 EVGYSSIPTIALCDTDSPLEYID?AIPANNKGKESIALLYWLLAREVLYLRNVHSRSQPWEVMVD
 SFFWRDPEELERQEEMDQGFAAAPRQAYTEEWTAGGGeweagpaaensswdaaggaddwsapaaa
 GGEATEGGNW

rps3

MAQAESLRYKLLGGLAVRRACYGVLRFVMENCKGVEIIISGKLRAQRAKAMKFKDGYLISTGEP
 KRHYIDEAVRHVLMRQGVLGKVKIMMGYDPEGKMGTKVILPDNVVVHEPKDEAPVNFSAEQTYE
 EETG

rps5

MSDVKLFGRWPYEDVNVSDNSLLDYIAVTHAAHHCFLPHTQGRYQKKRFRKALCPIVERLCCSMM
 MHGRNNGKLMAVRIVKHAFEI IHLITDKNPIQVFDVAVKNGGPREDSTRIGSAGVVRQAVDVS
 PLRRVNQAIYLICTGARNSSFRNIKSIAECLADEIMNCAKESNSYAIKKKDEIERVAKANR

rps12

MGGSKPRGIRSARKQVIRRRRTQRWADKDYKKANLGTKYKCNPFGGSSHAKGIVVEKIGIEAKQPN
 SAIRKCCRVQLIKNGKKIAAFVPRDGCLNFIDENDEVLISGFGRRHAVGDI PGVRFKVVKVAGC
 GLGALYREKKEKPRS

rps14 -

MAPKTKAPVEAVSLGPQVREGETVFGVAHIFASFNDTFVHVTDVSGRETIIRVTGGMKVKADRDE

SSPYAAMLAAQDVAARCKELGISALHIKLRATGGTKSRTPGPGAQSALRALARSGMKIGRIEDVT
PIPTDCTRKAGRRGRRL

Rps15a

MVRI SVLRDCLNNINNAEKRGKRQVLVRPNSKVVVKFLQVMQKNGYIGDLEIIDDHRAGKVVVDL
IGRINKCGVISPRFDVPIGGIEQLASDLLPSRQFGHVVLTPPYGIMDHEEAKRKHTGGKIVGFFF

rps18

MRLGNALRAAVL PARHRAGAF IGRWPACARHAANTPGKVPSGSGQSEEAVRLMLDTKVSQRGTA
DALFQECIDAIRERRLKLEQDPVLQEGDTPAITKLKQDELDTKKIFQRVDPSRDITSDLTHAGH
NIDPYWSPFHKQREVVDQVAEFDEYARLVTVAEQKRVRIQIKKSLRRAAVNPNYSKEFRQWHG
RKSGQAPPKPFRLPDKHWEPSPLQVRLARERITWRDVDILQHFIADNGYILPRRTTMLPRQKQQA
LVQAVKTAQNMSLLPYQWKLKDFQAMPLMDPLQWMADRLTDRVMESRDRRSRAMLRVMMERHPEL
NYRNFLRHEAQRAGAGDPSDAV

rps19

NQKIVGRSRNKSPYVSAALLTNLAKQEKTHSRQASGKEAHPSPIKTWSRASTVIPTMLGYTICVH
NGREHIPVFIDTRKLGKLGEFAPTRTFRAHKRDKAKASK

rps23

MGGSKPRGIRSARKQVIRRRQTQRWADKDYKKANLGTKYKCNPFGGSSHAKGIVVEKIGIEAKQPN
SAIRKCCRVQLIKNGKIAAFVPRDGCLNFIDENDEVLISGFGRRGHAVGDI PGVRFKVVKVAGC
GLGALYREKKEKPRS

ycf4 -

MVIPQRGARNTSNLIFSILLTLGGLGFFLTGLSSFFKYNLIFLTVNSEIAFLPQGIVLLLLYGTIG
LTLGSFLFLTFYWNLGAGYNLYCTKKKAVLFYRTRFPGRKTNNPLLLKFPFATINSLGLKRGTA
SPDNSVIAYTSNGSKFPLMYTDTMEVARNAMTAYLMNCPVDSKTLFPKNFSVSDLRKVKSSI
LILFR

cbbX

MSLRRIAGTVDKFMYKEGVAEVLDKLDREMIGLREVKDRVRAVASMLVIDKMRMKLGLETSMPSL
HMSFTGAPGTGKTTIALRMGQILQKMGYCRTHVVMATRDDLQYVGHVAPKTKEMVKKAMGGI
LFVDEAYLYDEENERDYGAESCEILSSVMETMQDDLIVIFAGYKELMDHFYAGQPGIKSRVGNH
IDFPNYSDELIDIGKVMLNASSMTMPDAIVALREYMKRRQLPFFSNARTVRNAMDMARMKGS
VRTYMEIVKPMMSGTVNLPMFAFTEVSAKDIPTIEEMQGGGQNKVMG

clpC

MMFERFSES AIQIIMMAQRESQRLGHNYVGNEMMLLGVVAQKSGVAAKVLAKLGVKLEEMRKVVE
 ELVGKNGITGVDPVAKEAKEVLEGCVEEAELGSKAVDTVHILLALIKDEDGDAAKILKKLSV
 DPAKVRTEVLDAVKSQDKPLVGAVSSSGGAKSGKSNTATLEQYARDLTKMAEAGELDVVIGREKE
 IKRAVQVLARRRKNPNVLI GEPGVGKTAIAEGLAQRIADGKGPSFLQGKRIMQLDLTGLLAGTKY
 RGDFFERLQNVIKEVVASNREIVLMI DELHMLV GAGSAEGSMDAGNMLK PMLANGELQLIGATTL
 EYRKYIEKDKALERRFQTINVPEPTADQATDILKGLTYKYEHHHELRYTPEAIEACVKLADRYI
 TDRYLPDKAIDILDETGARVRLQAAKAVPKETLEIETELKEVQKQLDEAVEVQDFEKAAELKPKK
 DKLNATLMSLGE GEEKAEKTFVSESDVERIVADMTGIRVEKVSSNESARLLGLEDTLHDRVIGQ
 DEAVVAVSKAVRRSRSGLKDPNRP IASFI FCGPTGVGKTELCKALS DAYYGRKDSMIRFDMSEFM
 ERHTVSKLIGSPPGYVGYNDESQ LTKVRRNPYSLILFDEIEKAHPDVFNMLQLILEDGMLTDSK
 GRTVSFKNALIIMTSNVGAQGI EKT?AGGGGFGFQSATDDVEQATYEKMKSVVGDQLKNNFKPEF
 INRLDETIVFKPLTKKEIEQIAELELAKVFERVKEQGLTIEMTERFKKCVDDGDFDPKFGARPLR
 RAI AKLLEDELAASV LLEPVRENEIAIVDIDDDGKVKILRNQGEVVKVTEEEEAPVLR

groEL

MVVKRIFFREEARIAMERGIDMLS DAVSTTLGPKGRNVVLGQKFSTPQIVNDGVTIARAINPLDN
 IENMGVCLIRQAASKTNDVAGDGT TTTATVIAHAIKQGLRNIAAGSNPIVLKRGIEKAAQYIVDR
 IAEDAMPIAGWEDI FTVATI SAGNDKVMGKLIADAILGVGNDGLISVEESNSTANELEFTEGFGEF
 DRGFLSGYFVTDSEKVECLLRDPY ILLTDKRLTVVKDLLSTLELMTKLNKPLLI IAEDVQOEAL
 ATLIINSLRGILKVVAVRAPGFGEKRSTLLADMAVMTGGQVISSAAGYNLENMQVELLGRAQRVI
 IKRDSTTIITDANKQKVLARCEQLKRELENDSTFEKEKI QERLSKLSGGIAVIRVGAATETEMR
 DKKLRMEDAVNATKAAL EEGIVPGGITLARLNEDLKRWVKCSLKNDEVTGGLIVGDALCAPFKK
 IAANAGKNGAVTFAKLMKESDKRVGYDAASDEMVDMYDRGIIDPAKVTRATLQ NATS IASMVLT
 ECIIIVDSKNDYL

secA

MKNNGLLDFYQQI LENFDLAKYKRQVTRINVLEPVFSKYSNQDLRDKTVQLKTL MSETRATSNV
 VEEAFALVREASKRVLGRLRHFDVQIIGGLVINDDKIAEMKTGEGKTLASILPAFLNALYGKGVHI
 ITVNEYLAQRDSQYTGQLHNFLGLSVGLVTETMSSVEKKKNYYCDITYVTNSELGFDYLRDNLVQ
 NAEGVVQRSLFYCVIDEIDSVLIDDASTPLIIAGSPTSLDDGLTEKYIKSTRVANILKRDKHYKT
 EEKTKSVLFS DAGISMCETCLGIERSEMYSLPDPFWFIILNAIKAKEFYRRNRNYIIDANLEIVI
 LDEFTGRALEGRRWSEGMQQAMESKERLTISGLTPPQASISYQNLFLLYERLSGMSGTAATEKQE
 FSTIYDLPVVVIPTNKPGRRIDFSDSVYTKQLYKWQGVINECRDMYRVGRPVLI GTKTITASERV
 ADFLKSQNI PYQLLNARPENSRRESEIIAQAGCESTVTIATNMAGRGTDIILGGNSQYFASTTIA
 TFVENINFSNASMNSALRNYHQEYVRS LQSSNEKKLDLNLRLRKFYNLVLYKKSLLTQKNRKT
 LRLGGLHVIGTERHESRRIDNQLRGRAGRQGEPPGSSRFFTS LDDDLLRQFGSEALQNFYSRLGVP
 EDIPLQSNLTLTQLDQAQQKVENSF DANKRLFDYDQVLNSQRNCVYSERTRQVQSPDLPNLVRT
 SLKLTIDDLFKCTLLSKFFTSKSGNGISNLSKQEIREFIYEILEIPLMPYPFSDNSMTEQISKVS
 IELKSLSNTSLFLAYANSSRPIPLFVNKIASNVTLRELDTVWSNHLQRMSALKELTQWSAYGRRN
 PLTEYKRLSFYYFLQMINIFQQKVLQAVPVLFESIV

secY

MKRFLQTIKNIWSIEDLRSRILTTLGLILYRVGSYVVLPGVDS DALAAANAGSGGGITDLINLF
AGGAFSRASIFALGIMPYISASIV?QLMGIAIPYFQKLQKEGESGRRKINQITRYLTIAITAMQA
PSYIATQVPNEAIYNVGFWSFTFPAVIILVTGTMFVMWLGERITEKGIGNGISLLIMIGIIANLP
FAFTAEFASRLSNGGMVVFLIEMVALILVVIACIMLVQGTRKIPVHFAKRIVGKGSKMMQQGGVR
QYIPLKVNAAGVMPIIFAQAIMFVPITLAGFSDSMSGIAAAFADEFTGFWYNFTFALLCIAFTYFY
TAVTVNPNQIADDLKRNGGFIPGVKPGKRTSEFIDNVMSRITLPGSIFLAFVAILPAFAGILEVN
SQFAQFFGGTSLIMVGVVLDLTLQQIESHLLMREYDGLMKSGRIKGRSSTGGVGMAG

APPENDIX C: SUBSTITUTION MODEL SELECTION

Appendix Table 3-1. Results from MEGA6 "Find best protein model" analysis.

PetD		PsaA		PsaB		PsaC		PsaF		PsaJ	
Model	BIC	Model	BIC	Model	BIC	Model	BIC	Model	BIC	Model	BIC
LG+G	6570	LG+G	20808	LG+G	17234	LG+G	1803	LG+G	10252	LG+G	1938
LG+G+I	6575	LG+G+I	20815	LG+G+I	17242	JTT+G	1804	LG+G+I	10259	LG+G+I	1942
cpREV+G	6625	LG+G+F	20828	cpREV+G	17330	LG+G+I	1809	WAG+G	10282	cpREV+G+I	1943
cpREV+G+I	6629	LG+G+I+F	20835	cpREV+G+I	17330	WAG+G	1810	WAG+G+I	10290	cpREV+G	1945
rtREV+G	6638	cpREV+G	20840	LG+G+F	17351	JTT+G+I	1810	cpREV+G	10303	mtREV24+G+I	1946
rtREV+G+I	6643	cpREV+G+I	20840	WAG+G	17352	Dayhoff+G	1812	JTT+G	10309	mtREV24+G	1949
mtREV24+G	6644	cpREV+G+F	20850	LG+G+I+F	17358	WAG+G+I	1815	cpREV+G+I	10311	WAG+G	1950
mtREV24+G+I	6651	cpREV+G+I+F	20853	WAG+G+I	17359	Dayhoff+G+I	1819	JTT+G+I	10317	JTT+G+I	1951
WAG+G	6656	WAG+G	20952	JTT+G	17366	rtREV+G	1826	Dayhoff+G	10349	JTT+G	1953
WAG+G+I	6660	WAG+G+I	20958	JTT+G+I	17373	JTT+I	1831	LG+G+F	10353	WAG+G+I	1953
JTT+G	6674	rtREV+G+F	20962	rtREV+G	17390	rtREV+G+I	1831	Dayhoff+G+I	10357	LG+I	1954
JTT+G+I	6679	rtREV+G+I+F	20968	rtREV+G+I	17398	LG+I	1833	LG+G+I+F	10361	cpREV+I	1958
LG+I	6685	WAG+G+F	20968	cpREV+G+F	17422	WAG+I	1833	WAG+G+F	10367	WAG+I	1965
LG+G+F	6687	WAG+G+I+F	20975	cpREV+G+I+F	17423	Dayhoff+I	1841	WAG+G+I+F	10375	JTT+I	1967
rtREV+G+F	6691	JTT+G	21006	rtREV+G+F	17431	cpREV+G	1843	Dayhoff+G+F	10382	Dayhoff+G+I	1968
LG+G+I+F	6693	JTT+G+I	21012	rtREV+G+I+F	17438	cpREV+G+I	1847	JTT+G+F	10387	mtREV24+I	1969
rtREV+G+I+F	6698	JTT+G+F	21027	WAG+G+F	17457	mtREV24+G	1847	Dayhoff+G+I+	10390	Dayhoff+G	1970
Dayhoff+G	6730	JTT+G+I+F	21034	Dayhoff+G	17460	mtREV24+G+I	1854	rtREV+G	10394	rtREV+G	1975
cpREV+I	6733	Dayhoff+G+F	21049	WAG+G+I+F	17464	rtREV+I	1855	JTT+G+I+F	10395	rtREV+G+I	1978
Dayhoff+G+I	6734	mtREV24+G+F	21050	Dayhoff+G+I	17467	cpREV+I	1867	rtREV+G+F	10397	rtREV+I	1989

NOTE.-- Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best.

Δ BIC > 10 is considered significant.

PsbA		PsbB		PsbC		PsbD		PsbE		PsbF	
Model	BIC	Model	BIC	Model	BIC	Model	BIC	Model	BIC	Model	BIC
LG+G+I	7494	LG+G	15849	LG+G	13359	LG+G	7194	LG+I	2535	JTT+G	1279
LG+G	7494	LG+G+I	15859	LG+G+I	13365	cpREV+G	7201	LG+G	2538	rtREV+G	1281
cpREV+G	7496	cpREV+G	15892	LG+G+F	13410	LG+G+I	7204	LG+G+I	2540	LG+G	1282
cpREV+G+I	7499	cpREV+G+I	15902	LG+G+I+F	13416	cpREV+G+I	7210	cpREV+I	2545	cpREV+G	1282
WAG+G	7552	WAG+G	15945	cpREV+G	13437	JTT+G	7245	cpREV+G	2550	JTT+G+I	1285
WAG+G+I	7555	WAG+G+I	15954	rtREV+G+F	13442	WAG+G	7249	LG	2553	rtREV+G+I	1287
cpREV+I	7556	JTT+G	15968	cpREV+G+I	13442	JTT+G+I	7254	cpREV+G+I	2554	LG+G+I	1288
LG+I	7563	JTT+G+I	15977	rtREV+G+I+F	13447	WAG+G+I	7258	WAG+I	2555	cpREV+G+I	1289
JTT+G	7584	LG+G+F	16008	WAG+G	13490	rtREV+G	7289	WAG+G	2559	WAG+G	1292
JTT+G+I	7584	LG+G+I+F	16018	rtREV+G	13491	rtREV+G+I	7299	cpREV	2560	cpREV+I	1295
LG+G+I+F	7592	cpREV+G+F	16038	rtREV+G+I	13496	mtREV24+G	7302	WAG+G+I	2562	Dayhoff+G	1296
rtREV+G+I	7593	rtREV+G	16043	WAG+G+I	13497	mtREV24+G+I	7311	JTT+I	2565	WAG+G+I	1298
rtREV+G	7594	Dayhoff+G	16047	cpREV+G+F	13510	LG+G+F	7315	JTT+G	2568	JTT+I	1300
LG+G+F	7596	cpREV+G+I+F	16047	cpREV+G+I+F	13516	LG+G+I+F	7325	rtREV+I	2570	Dayhoff+G+I	1302
cpREV+G+F	7610	rtREV+G+I	16053	JTT+G	13523	cpREV+G+F	7340	JTT+G+I	2571	mtREV24+G	1303
cpREV+G+I+F	7612	rtREV+G+F	16055	JTT+G+I	13531	cpREV+G+I+F	7349	rtREV+G	2572	rtREV+I	1304
rtREV+G+I+F	7617	Dayhoff+G+I	16056	WAG+G+F	13547	WAG+G+F	7352	WAG	2573	LG+I	1307
WAG+I	7622	rtREV+G+I+F	16065	WAG+G+I+F	13554	mtREV24+G+I	7352	rtREV+G+I	2575	mtREV24+G+I	1309
rtREV+G+F	7622	WAG+G+F	16125	JTT+G+F	13595	rtREV+G+F	7353	JTT	2584	WAG+I	1312
WAG+G+I+F	7629	WAG+G+I+F	16134	JTT+G+I+F	13602	Dayhoff+G	7354	rtREV	2589	Dayhoff+I	1323

Appendix Table 3-1. Results from the MEGA6 "Find best Protein Models" analysis.

PsbH		PsbI		PsbN		PsbT		PsbV		RbcL	
Model	BIC	Model	BIC	Model	BIC	Model	BIC	Model	BIC	Model	BIC
LG+G	2433	JTT+G	1242	JTT+G	1806	LG+G	1209	LG+G	5974	LG+G	13578
LG+G+I	2440	LG+G	1247	JTT+G+I	1812	rtREV+G	1215	LG+G+I	5982	LG+G+I	13584
WAG+G	2444	JTT+I	1248	LG+G	1821	LG+G+I	1216	JTT+G	5984	rtREV+G	13661
JTT+G	2445	JTT+G+I	1249	JTT+I	1826	rtREV+G+I	1222	JTT+G+I	5992	rtREV+G+I	13667
cpREV+G	2451	JTT	1251	cpREV+G	1826	LG+I	1222	cpREV+G	5999	WAG+G	13685
WAG+G+I	2451	LG+G+I	1253	LG+G+I	1827	JTT+G	1222	WAG+G	6003	WAG+G+I	13691
JTT+G+I	2453	cpREV+G	1254	WAG+G	1827	LG	1225	cpREV+G+I	6007	rtREV+G+F	13735
cpREV+G+I	2458	LG+I	1254	Dayhoff+G	1828	rtREV+I	1226	WAG+G+I	6011	rtREV+G+I+F	13742
Dayhoff+G	2458	WAG+G	1255	cpREV+G+I	1831	WAG+G	1226	Dayhoff+G	6047	LG+G+F	13745
Dayhoff+G+I	2464	Dayhoff+G	1258	WAG+G+I	1834	rtREV	1228	Dayhoff+G+I	6055	LG+G+I+F	13751
rtREV+G	2472	LG	1259	Dayhoff+G+I	1834	cpREV+G	1228	rtREV+G	6066	JTT+G	13785
rtREV+G+I	2479	cpREV+G+I	1259	cpREV+I	1841	JTT+G+I	1229	LG+G+F	6071	JTT+G+I	13791
LG+I	2490	cpREV+I	1259	JTT	1842	cpREV+G+I	1231	rtREV+G+I	6074	Dayhoff+G	13794
WAG+I	2498	WAG+I	1261	LG+I	1843	WAG+G+I	1232	LG+G+I+F	6079	Dayhoff+G+I	13799
JTT+I	2502	WAG+G+I	1261	rtREV+G	1846	JTT+I	1235	JTT+G+F	6094	LG+I	13815
cpREV+I	2504	WAG	1264	WAG+I	1851	Dayhoff+G	1237	JTT+G+I+F	6102	cpREV+G	13836
mtREV24+G	2517	Dayhoff+G+I	1264	rtREV+G+I	1853	WAG+I	1237	WAG+G+F	6105	cpREV+G+I	13840
Dayhoff+I	2523	Dayhoff	1267	Dayhoff+I	1853	JTT	1237	rtREV+G+F	6106	WAG+G+F	13858
mtREV24+G+I	2525	cpREV	1267	LG	1860	mtREV24+G	1239	mtREV24+G	6113	WAG+G+I+F	13864
rtREV+I	2535	Dayhoff+I	1267	mtREV24+G	1862	WAG	1239	WAG+G+I+F	6113	WAG+I	13873

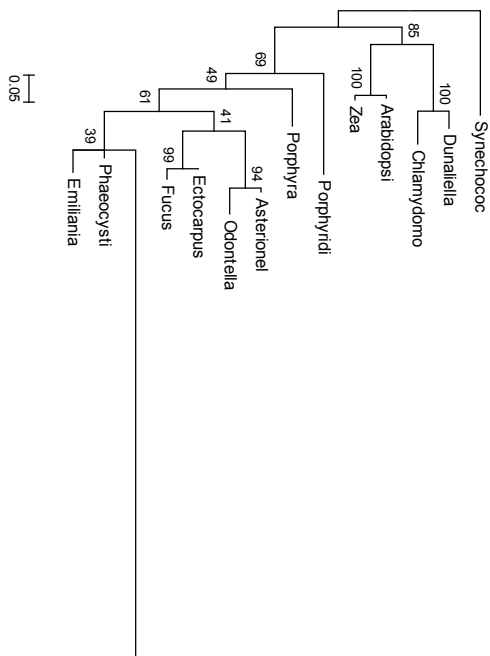
RpoB

Model	BIC
LG+G+F	61141
LG+G+I+F	61151
cpREV+G	61177
cpREV+G+I	61186
JTT+G+F	61243
JTT+G+I+F	61253
LG+G	61266
cpREV+G+F	61273
LG+G+I	61276
cpREV+G+I+F	61283
WAG+G+F	61386
WAG+G+I+F	61396
mtREV24+G+I	61458
mtREV24+G+I	61468
rtREV+G+F	61498
rtREV+G+I+F	61508
JTT+G	61540
JTT+G+I	61550
WAG+G	61740
WAG+G+I	61750

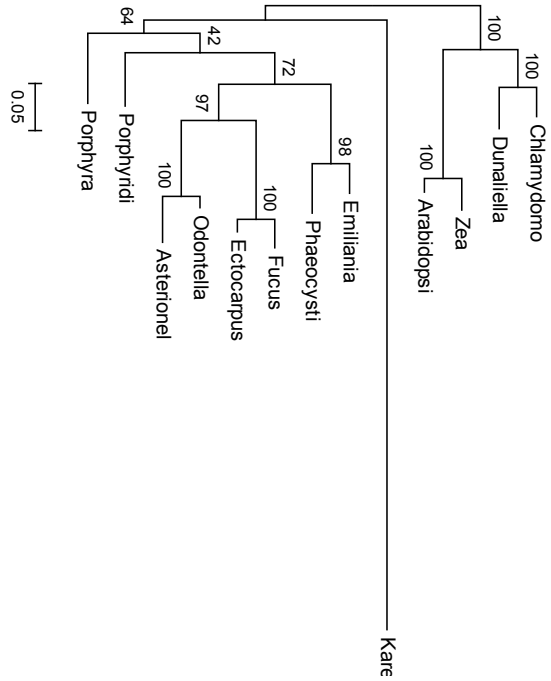
APPENDIX D: MAXIMUM LIKELIHOOD TREES

D-1. Preliminary trees for plastid protein candidates.

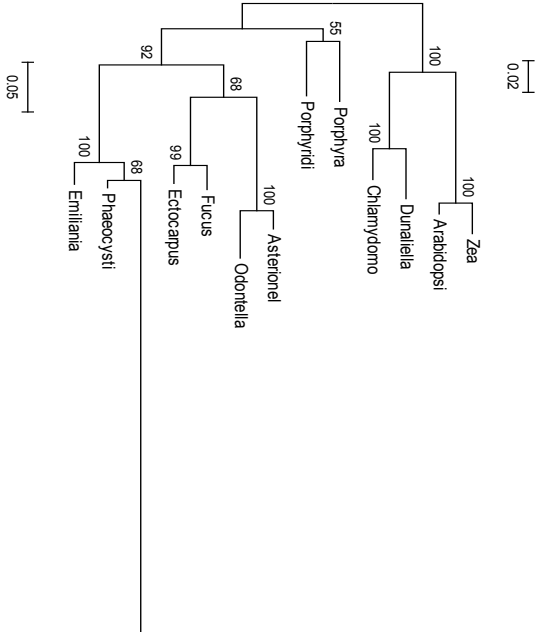
PetD



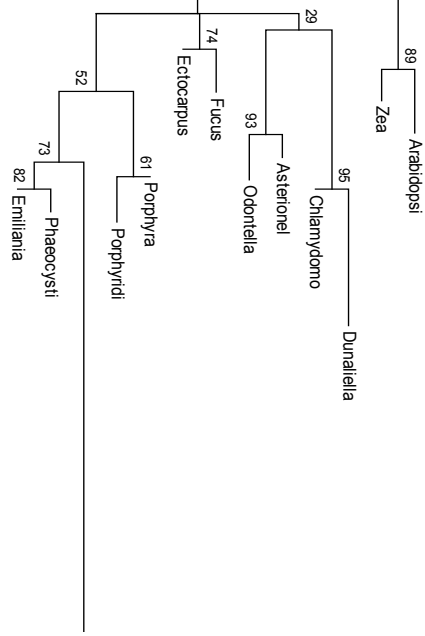
PsaA



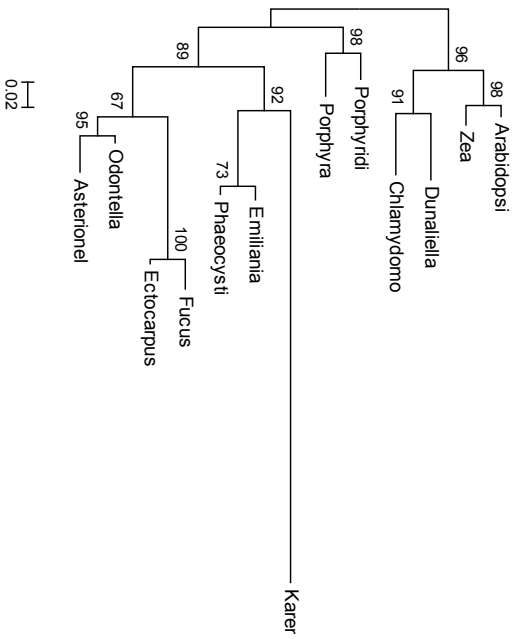
PsaB



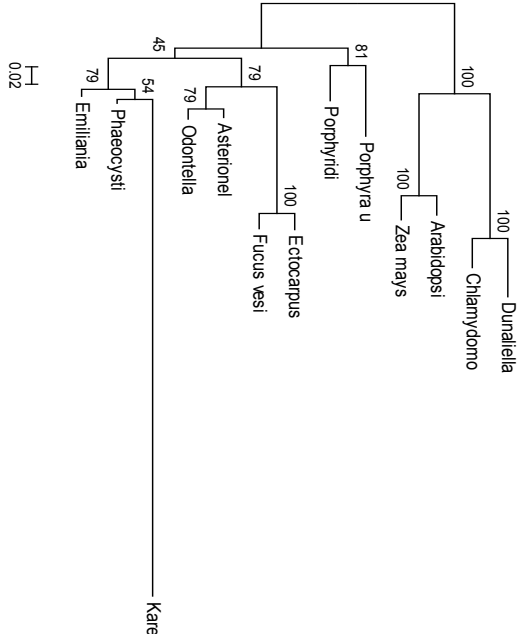
PsaC

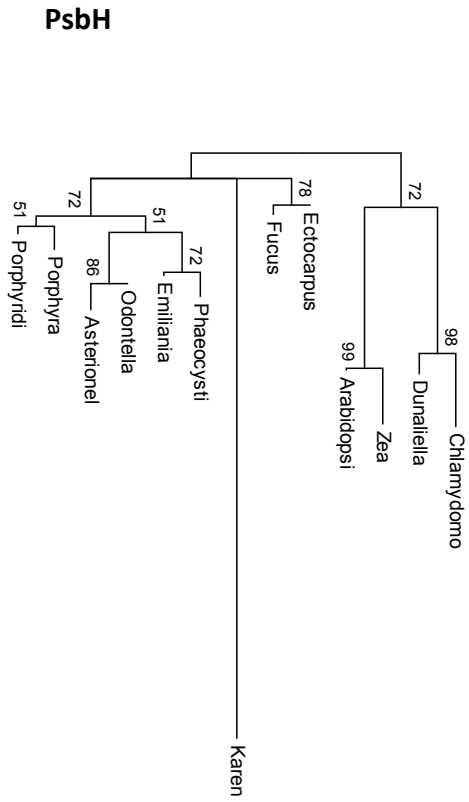
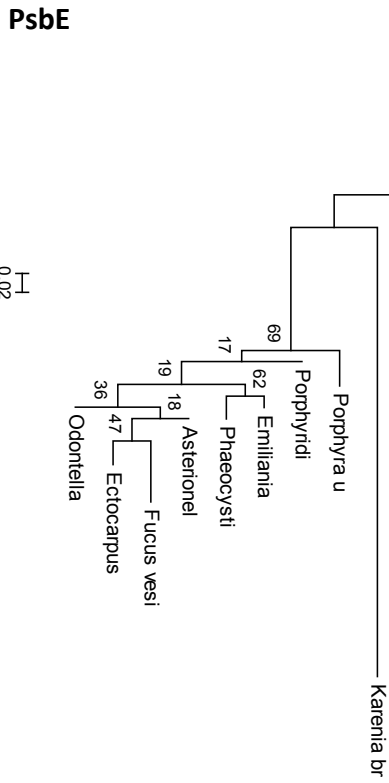
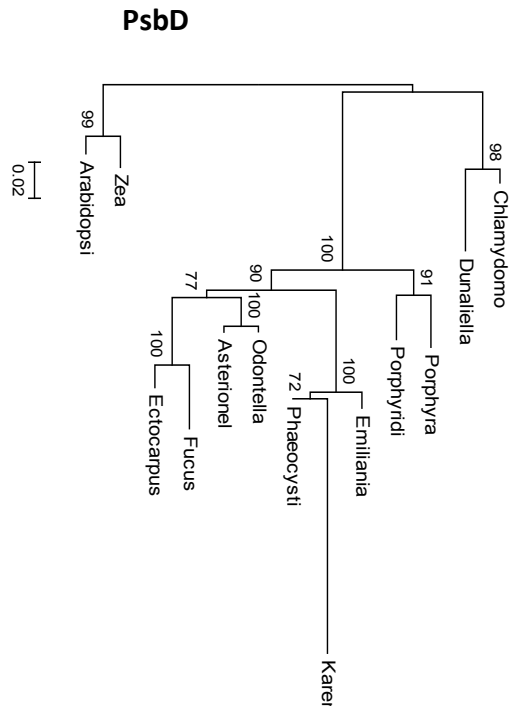
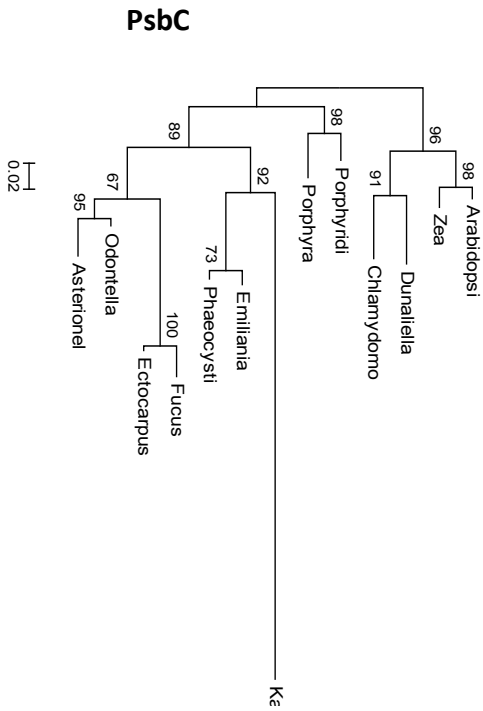


PsbA

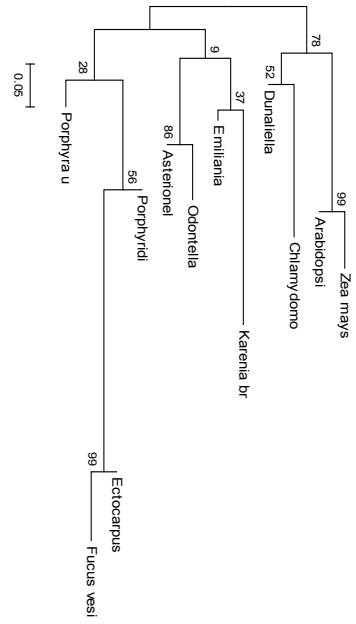


PsbB

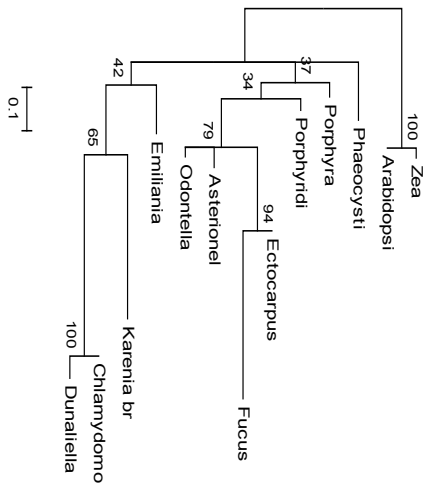




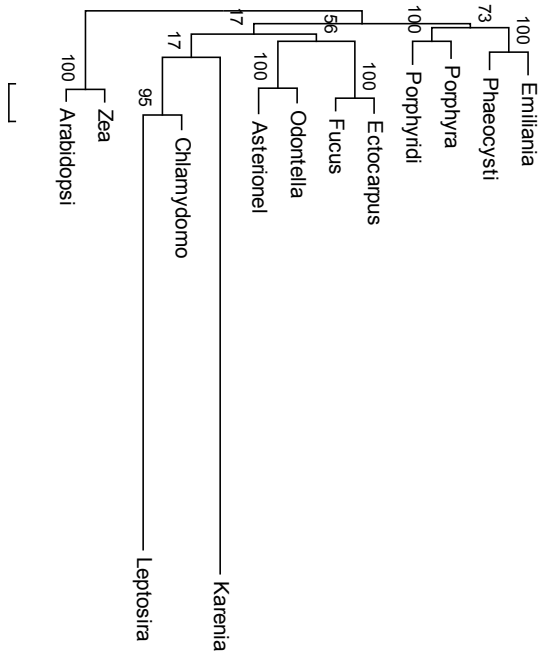
PsbT



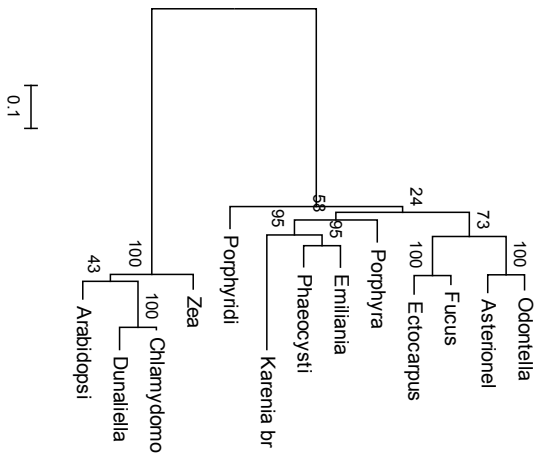
PsbN



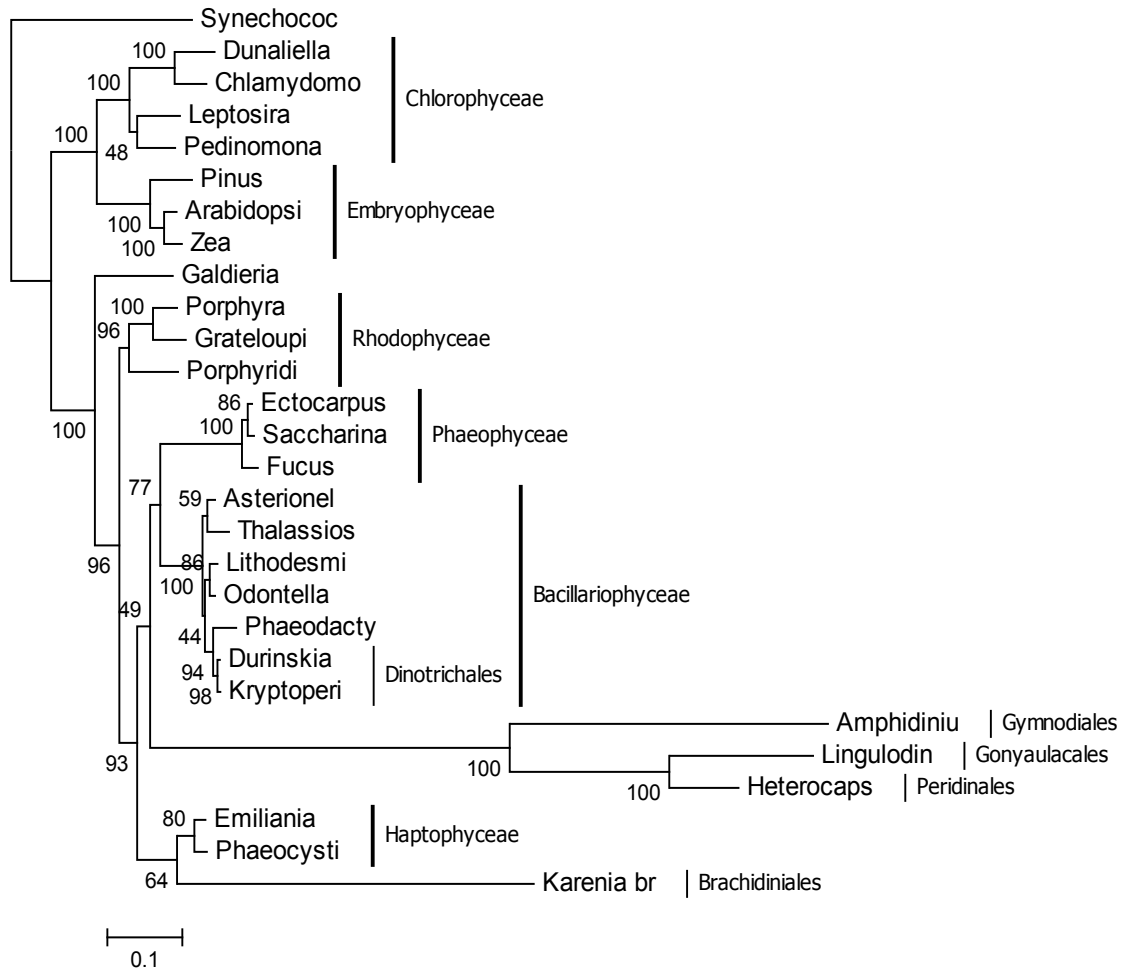
RpoB



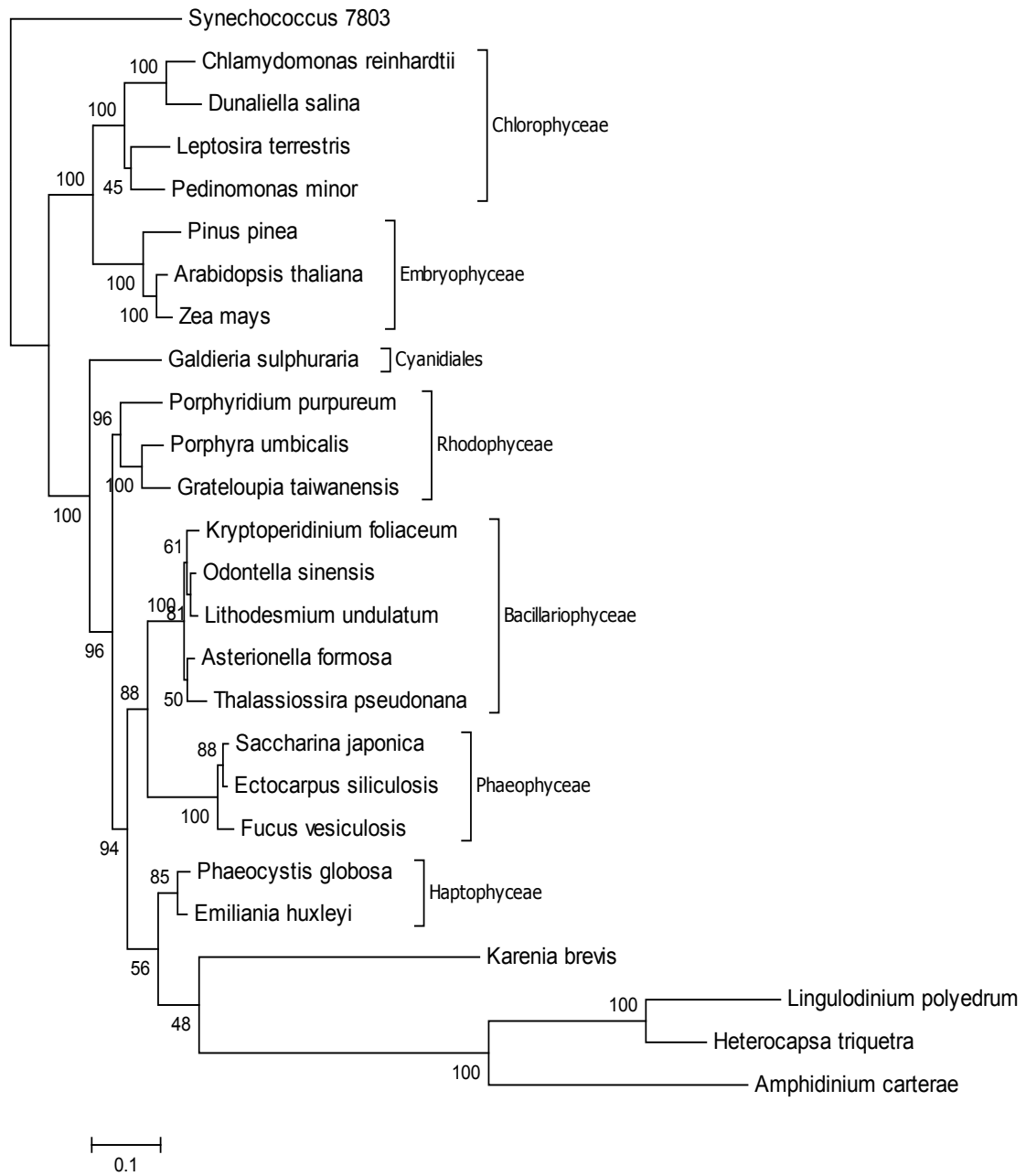
RbcL



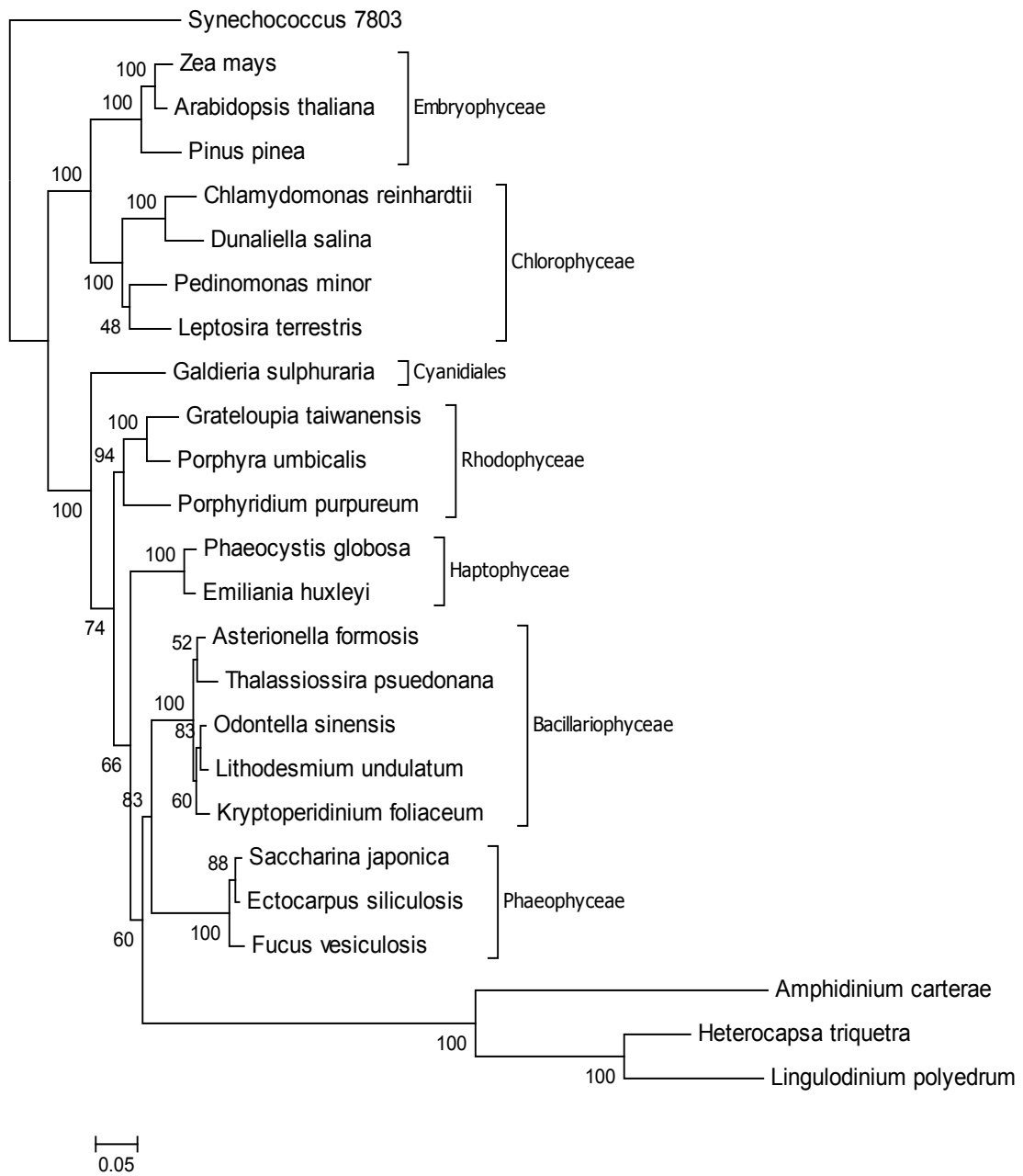
D-2. Five protein tree (petD-psbA-psbB-psbC-psbD)



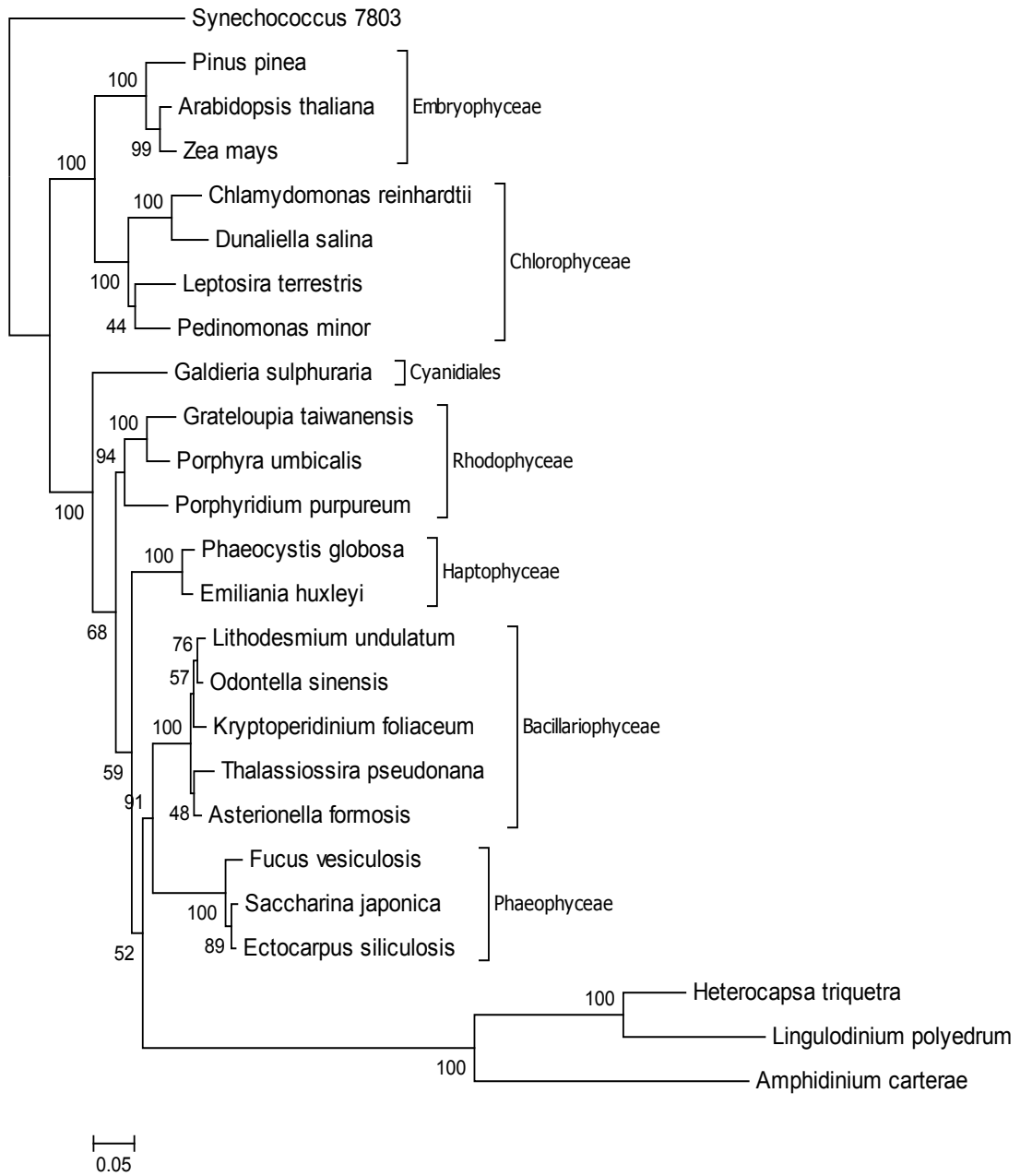
D-3. Six protein tree (petD-psbA-psbB-psbC-psbD-psbE)



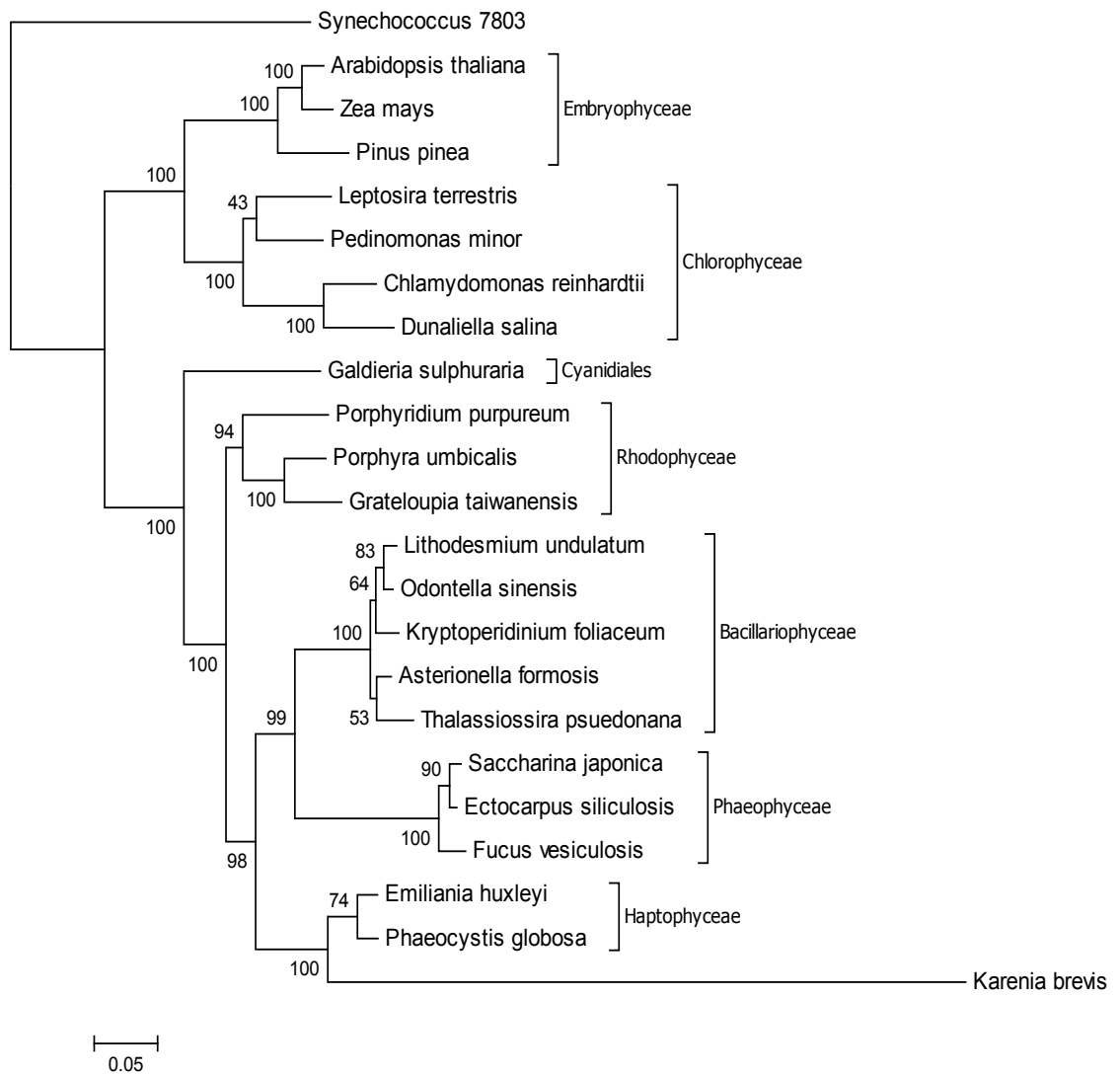
D-4. Five protein tree without *Karenia brevis* sequences



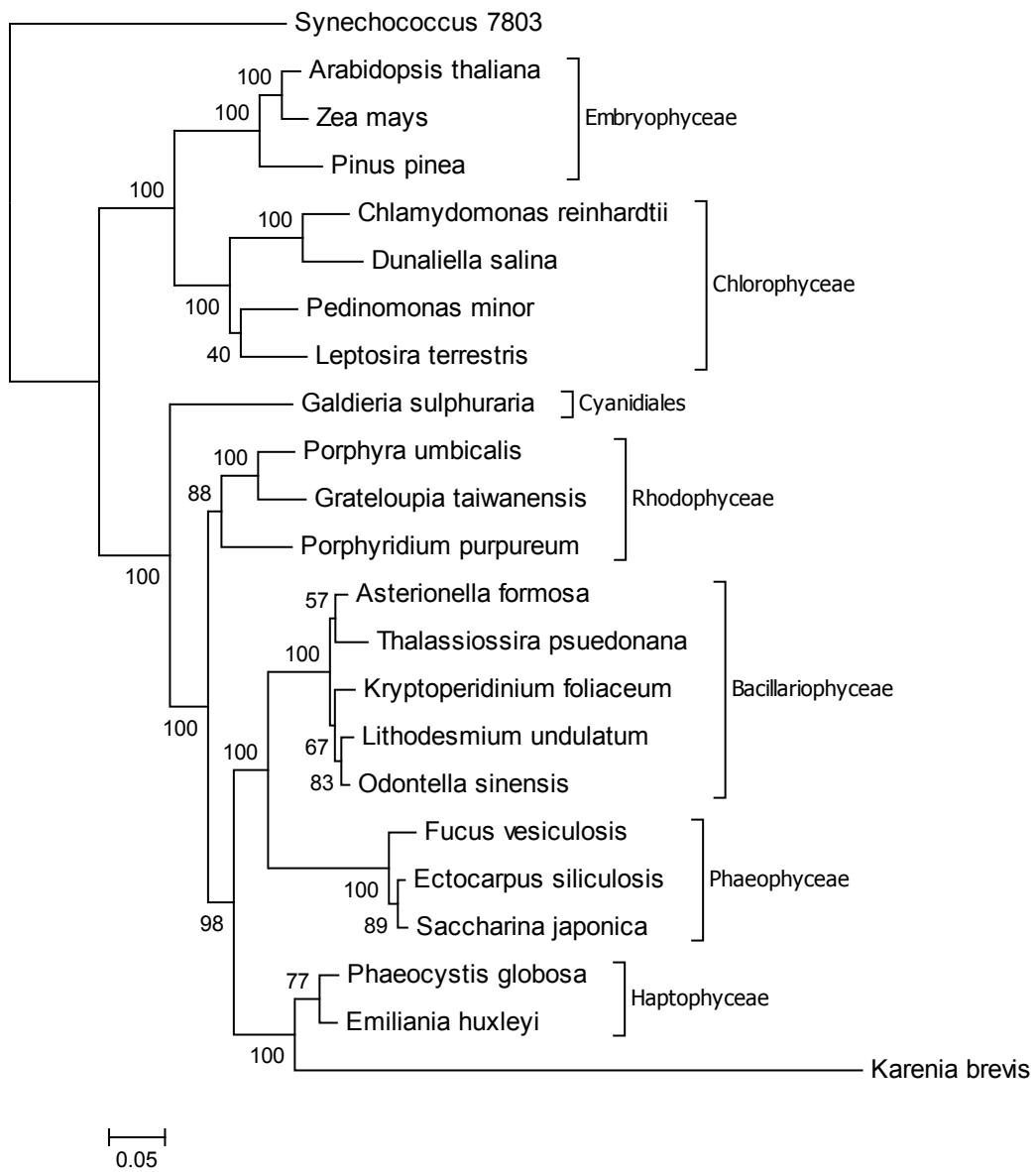
D-5. Six protein tree without *Karenia brevis* sequences



D-6. Five protein tree without peridinin dinoflagellate sequences

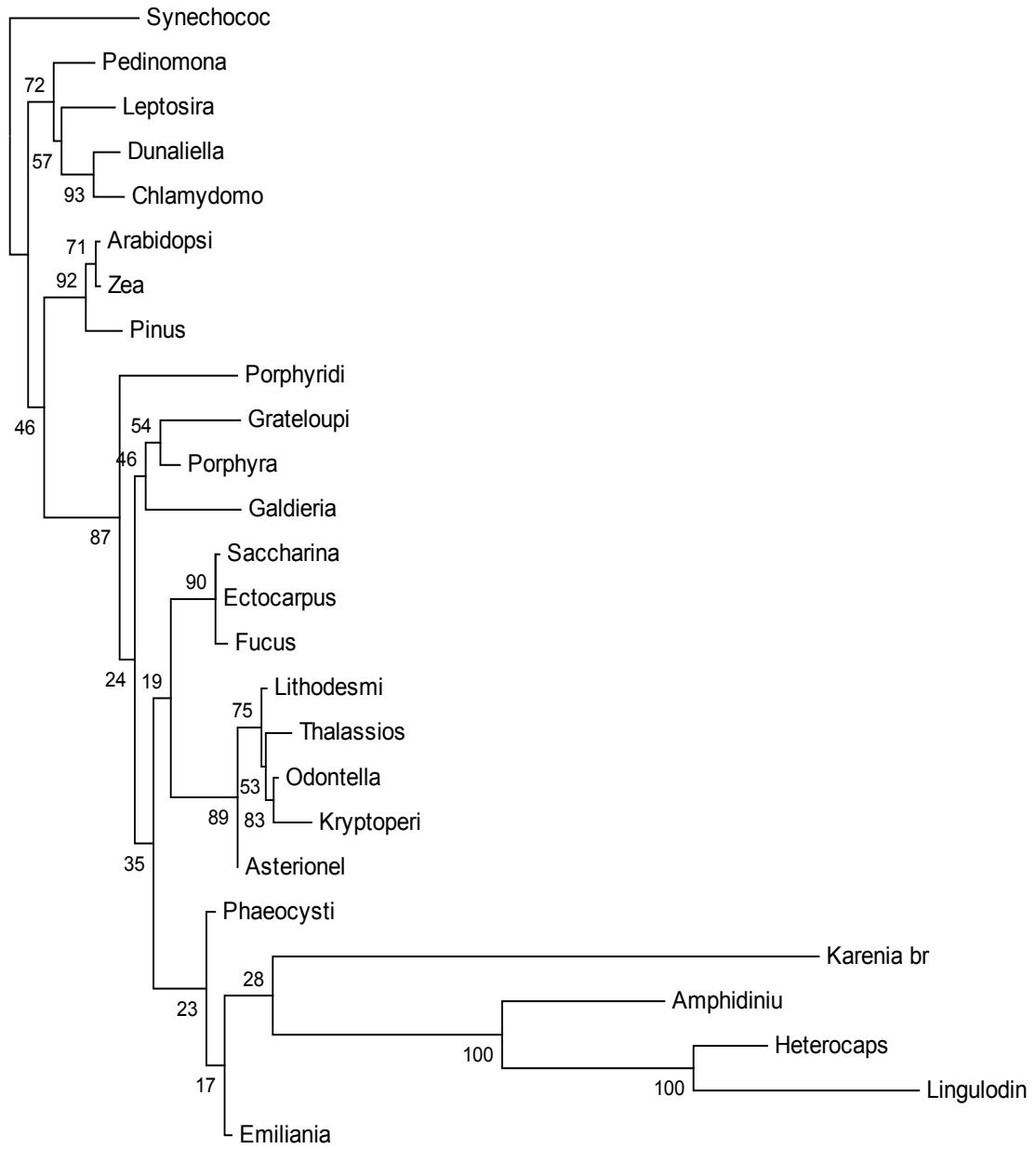


D-7. Six protein tree without peridinin dinoflagellate sequences



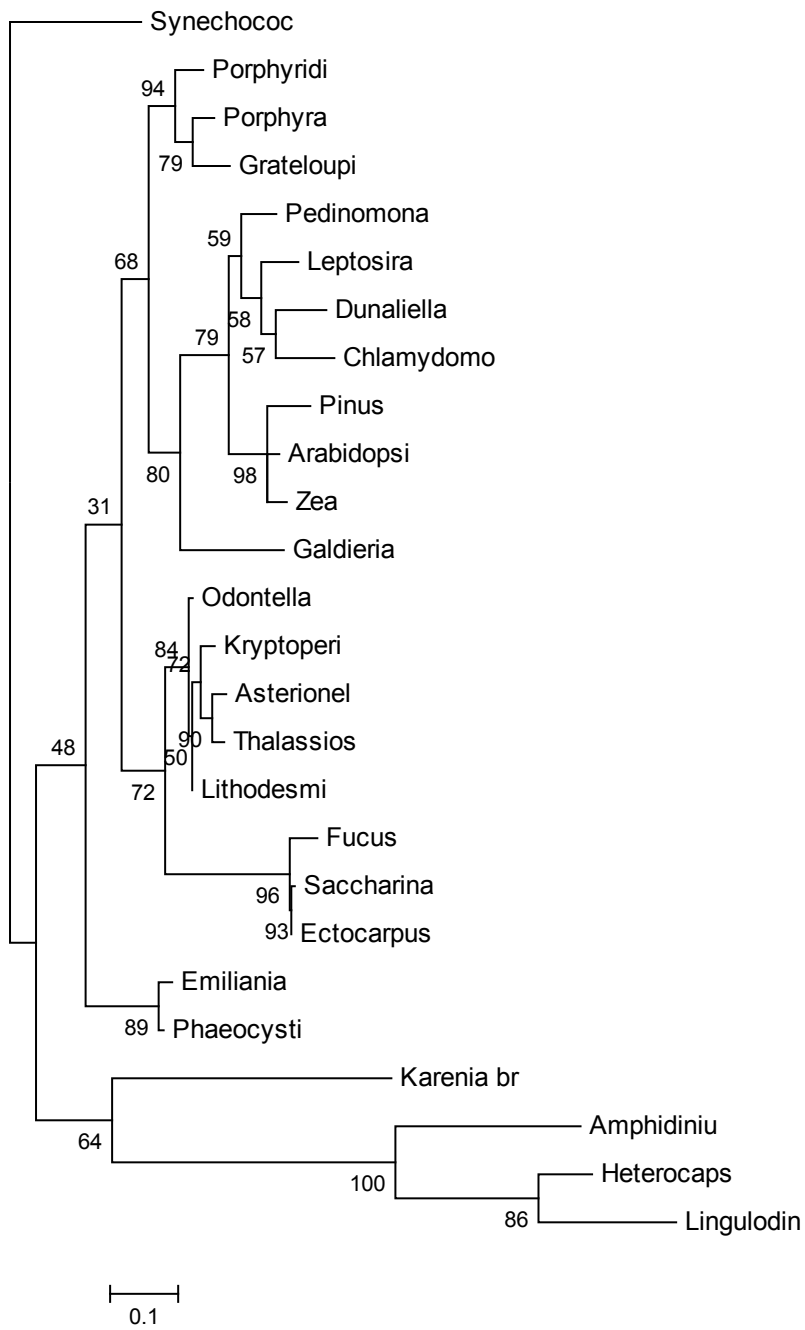
D-8. Unigene protein sequence trees

petD

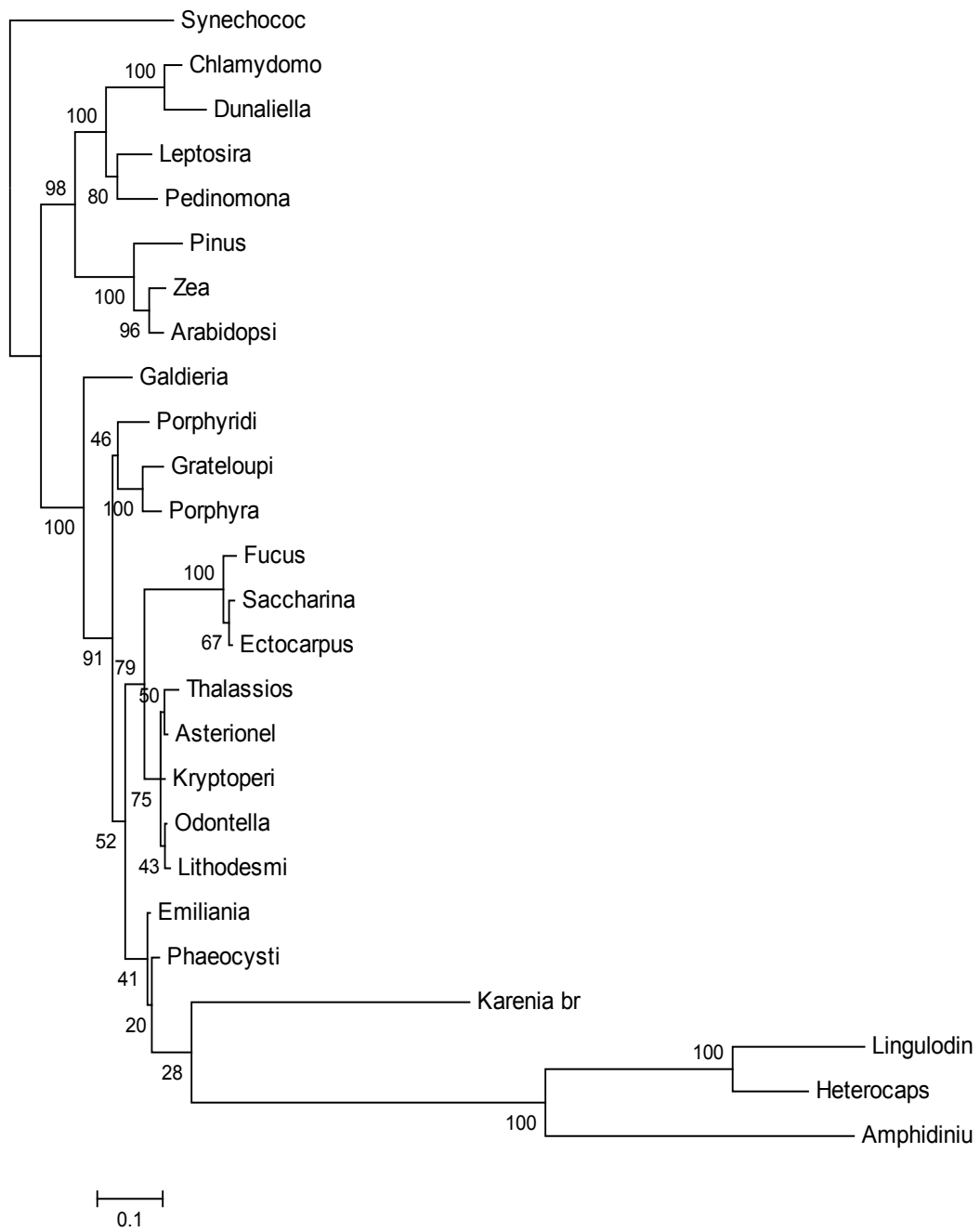


0.05

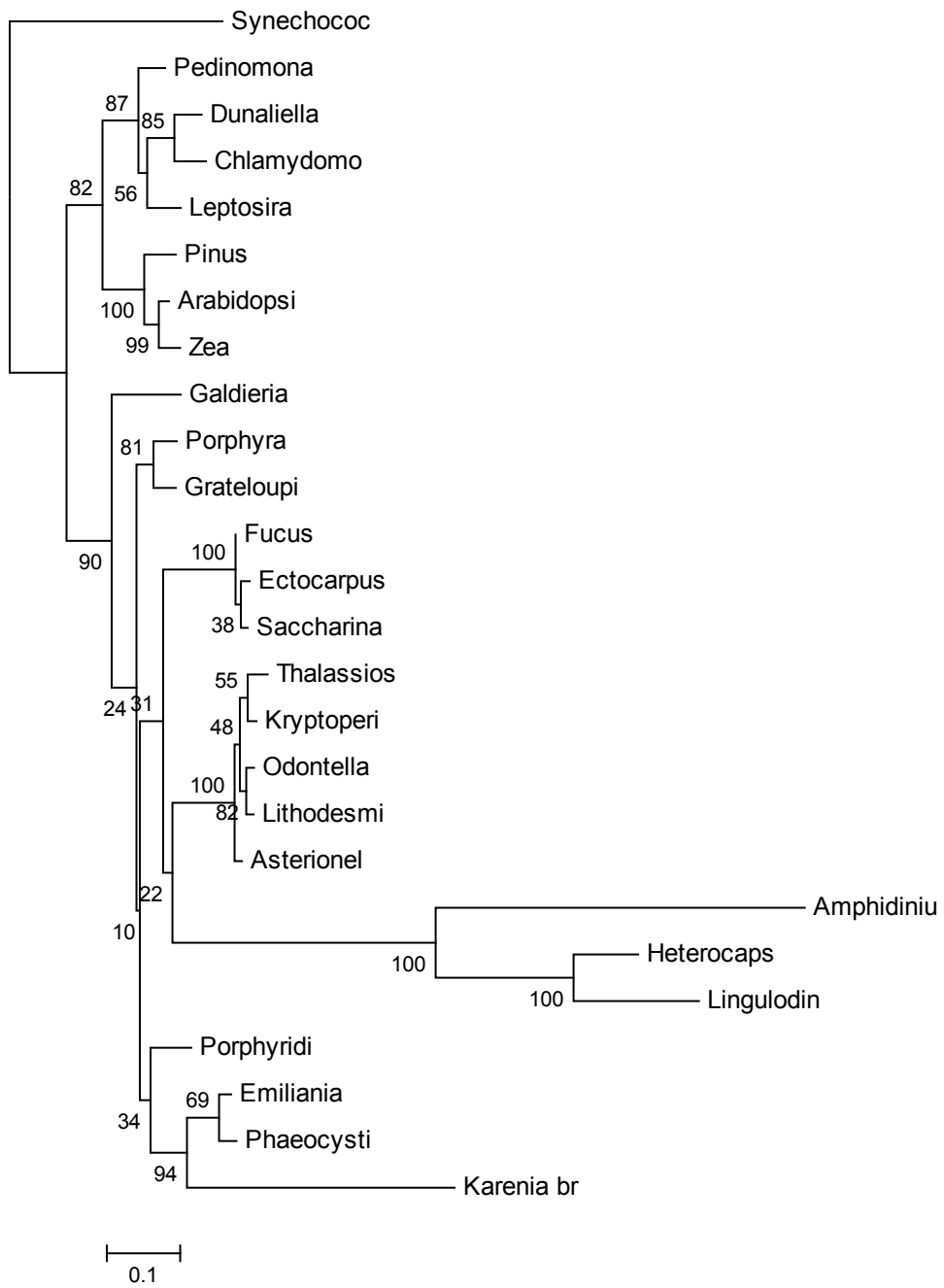
psbA



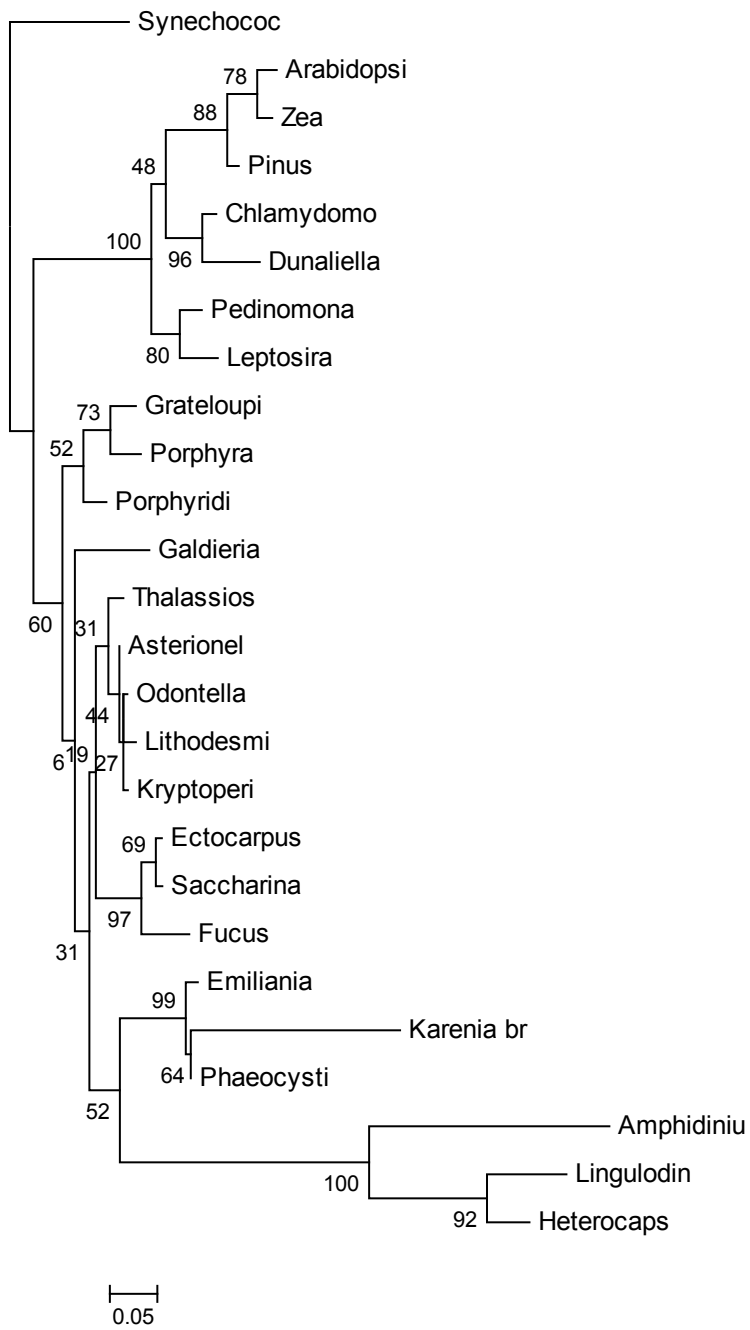
psbB



psbC



psbD



psbE

