

THE CORRELATION OF EVOLUTIONARY RATE AND PATHWAY POSITION IN PLANT TERPENOID
SYNTHESIS

by

Heather Ramsay

B.Sc., The University of British Columbia, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES

(Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

December 2008

© Heather Ramsay, 2008

Abstract

Genes are expected to face stronger selective constraint and to evolve more slowly if they encode enzymes upstream as opposed to downstream in metabolic pathways, since upstream genes are more pleiotropic, being required for a wider range of end-products. However, few clear examples of this trend in evolutionary rate variation exist. We examined whether genes involved in plant terpenoid biosynthesis exhibit such a pattern, using data for 45 genes from four fully-sequenced angiosperms, *Vitis*, *Arabidopsis*, *Populus* and *Ricinus*. Taking a bioinformatic approach, we used phylogenetic trees of proteins from each genome to determine orthologs and count paralogs in each genome, using proteins from the *Oryza* genome as an outgroup. Our results show that dN/dS does in fact correlate with pathway position along pathways converting glucose to the terpenoid phytohormones abscissic acid, gibberellic acid and brassinosteroids. Upstream versus downstream rate variation is particularly strong in the gibberellic acid pathway, but absent in the pathway to lutein, another terpenoid derivative. In contrast, we found no apparent variation of dN/dS with gene copy number. We also introduce a new measure of pathway position, the Pathway Pleiotropy Index (PPI), which counts groups of enzymes between pathway branchpoints. We found that this measure is superior to pathway position in explaining variation in dN/dS along each pathway. Kendall's τ for the correlation of PPI with dN/dS was 0.24 - 0.54 with a mean of 0.4. We further show that variation in dN/dS is due to differences in selective constraint, not positive selection. Therefore, our results are consistent with the prediction that selective constraint is progressively relaxed along metabolic pathways, showing plant terpenoid synthesis to be a robust example of positional rate variation.

Table of contents

Abstract.....	ii
Table of contents	iii
List of tables.....	v
List of figures.....	vi
Acknowledgements	vii
Co-authorship statement	ix
CHAPTER 1: Introductory remarks on molecular evolution and network biology	1
Background on molecular evolution	1
Review of literature on network structure and evolutionary rate variation	5
Networks: definition and structure	5
Node connectivity and selective constraint.....	6
Metabolic flux and selective pressure.....	7
Effects of position in metabolic networks	7
Proposal for further research.....	10
Choice of study system	10
Improved methods for detecting positional rate variation	11
Choice of taxa.....	11
Research objectives	12
References.....	13
CHAPTER 2: The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis.....	17
Introduction.....	17
Methods	19
Study species	19
Finding orthologs and counting paralogs.....	20
Measuring evolutionary rates.....	20
Variability of dN/dS	21
Relationship of dN/dS to pathway structure and gene copy number	21
Tests for selection.....	22
Results.....	22
Ortholog availability and rate variation among enzymes	22
Variation of dN/dS with pathway position	23
Variation of dN/dS with the Pathway Pleiotropy Index	23
Gene copy number.....	23
Selection analyses.....	23
Discussion	24
Gene evolution and metabolic pathways	25
Pathway-specific patterns	26
Lineage-specific patterns	26
Gene copy number.....	27
Future directions.....	28
Acknowledgements	29
References.....	33
CHAPTER 3: Conclusions	37
Summary and evaluation of Ramsay <i>et al.</i>	37
Comparison of correlations among studies.....	38
Context and significance of Ramsay <i>et al.</i>	39

Future directions.....	39
References.....	42
Appendix 1. <i>Arabidopsis</i> loci used as queries in ortholog determination.....	43
Appendix 2. Orthologs used to estimate dN/dS.....	44
Appendix 2.1. <i>Arabidopsis</i> orthologs.....	44
Appendix 2.2. <i>Oryza</i> orthologs.....	45
Appendix 2.3. <i>Populus</i> orthologs.....	46
Appendix 2.4. <i>Ricinus</i> orthologs.....	47
Appendix 2.5. <i>Vitis</i> orthologs.....	48
Appendix 3. Graphs of dN/dS for <i>Ricinus</i>	50
Appendix 4. Graphs of dN/dS for <i>Arabidopsis</i>	51
Appendix 5. Graphs of dN/dS for <i>Vitis</i>	52
Appendix 6. Alignments used for estimating dN/dS.....	53

List of tables

Table 2.1. Kendall's τ and p values for the correlation of dN/dS with pathway position.	30
Table 2.2. Kendall's τ and p values for the correlation of dN/dS with the Pathway Pleiotropy	
Index	30
Table 3.1. Correlation values for measures of evolutionary rate with measures of connectivity, flux and gene expression.	41

List of figures

Figure 2.1. Plant terpenoid synthesis	31
Figure 2.2. The phylogenetic relationships of the taxa	31
Figure 2.3. Graphs of dN/dS for <i>Populus</i>	32

Acknowledgements

Science is as much a social process as an intellectual process. I am very much indebted to my co-supervisors Kermit Ritland and Loren Rieseberg. Thank you both for giving me free rein and ample time to develop my ideas. Your feedback throughout this project has been of great value. Thanks, Kermit, for your generous funding and for giving me the push I needed to finish. Thank you, Loren, for your calm encouragement and for the opportunity to join your lab. I have benefited greatly from the excellent group of scientists you have gathered around you.

Thanks also to Keith Adams, who served on my committee, and to Steven Hallam who supervised my directed studies. The metagenomics project allowed me to gain skills and write scripts that served me well in my thesis research also.

Many thanks to Mike Barker and Nolan Kane, members of the Rieseberg lab, who first suggested the idea of a pleiotropy index. Mike Barker generously provided a number of perl scripts, as well as much advice on ortholog determination and other issues in bioinformatics. Nolan Kane took part in several key discussions and provided his unpublished work on positional rate variation. That material provided much food for thought as I developed my plans for the research presented here. I am also grateful to Matt King, a postdoc in the Rieseberg lab, for his encouragement and his help in exploring an alternative research plan during the early stages of this work. Though that plan was abandoned, Matt's help allowed me to focus on the direction my research would ultimately take. Thanks also to Katrina Dlugosch for her help in learning perl.

Other members of the UBC evolutionary biology community also provided insightful comments at various stages of this work: Ligia Mateiu, Dilara Ally, Allen Liu, Itay Mayrose, Sally Otto, Brian Ellis, Sean Graham, Anne Dalziel, and Crispin Jordan, as well as Michael Nachman during a visit to UBC. Not all of your ideas appear in this thesis, but they were nevertheless of great value in bringing my own ideas to maturity.

Thanks to Vivienne Lam and Holly Rail who edited my thesis draft. Any remaining errors are mine.

Throughout my graduate studies, NSERC has funded my work through grants to me and Kermit. I benefited greatly from the freedom to focus on my research, without need to work at the same time.

Research, in my experience, is like descending into the earth to bring out some gem of knowledge. The search for that gem can be dark, lonely and discouraging. I have been tempted more than once to give up, and I owe many thanks to both Dilara Ally and Moira Scascitelli for their encouragement at key times. This project may never have come to completion, had I not run into Dilara on the bus one fateful day.

Dedication

*The world is charged with the grandeur of God.
It will flame out, like shining from shook foil;
It gathers to a greatness, like the ooze of oil
Crushed.*

-from "God's Grandeur" by Gerard Manley Hopkins

To the ones who taught me that beauty, science and faith all belong together.

Co-authorship statement

Chapter 2 resulted from a collaboration among Kermit Ritland (research supervisor), Loren Rieseberg (research supervisor), and me. Kermit first suggested looking at positional rate variation, and proposed the test for variability of dN/dS among genes. Loren proposed the comparison of likelihoods to test for selection, and provided insightful comments on the role of gene duplication. Both Loren and Kermit provided detailed comments on the manuscript. I designed the experiment, collected and analyzed the data, and wrote the manuscript. I also wrote several perl scripts used in processing the data, besides those provided by Mike Barker.

CHAPTER 1: Introductory remarks on molecular evolution and network biology

Within the diversity of living things, there is yet another diversity: comparisons of gene sequences among species reveal that while some genes are nearly identical, others differ wildly (Nei 1987). Since all organisms share a common ancestor, we understand that interspecies variation in gene sequences represents mutations accumulated over time. Yet why, over the same time span, do some genes accumulate more mutations than others? In other words, why do genes evolve at different rates?

The answer lies in the elusive link between phenotype and genotype, between the physical structure and functioning of the organism and the DNA sequence of genes and regulatory regions of the organism's genome. In short, genes evolve at different rates (in part) because they have different effects on the overall phenotype in the course of natural selection. When a mutation occurs in a gene, it is initially present in only one individual. If the mutation is detrimental to the individual's survival and reproduction, it will not be passed on to future generations. On the other hand, a mutation neutral or beneficial to the individual's reproductive fitness may actually be fixed. That is, it may become part of the normal sequence of that gene in the population or species. Thus, genes differ in their freedom to evolve; they differ in the selective pressure they face.

While many early studies of evolutionary rate variation focused on a few unrelated genes, research over the last decade has shifted to examining groups of interacting genes. No gene acts in isolation. Rather, all participate in networks or pathways of gene interaction that brings about the overall phenotype (Proulx *et al.* 2005). The structure of these networks often explains differences in evolutionary rates and selective pressures (Cork and Purugganan 2004). The question of how network or pathway structure shapes gene evolution remains an active area of research, and Chapter 2 presents a recent advance in this field. In order to introduce the research presented there, this chapter includes both an overview of how rates of gene evolution are measured and other general concepts in molecular evolution, as well as a literature review on the role of network structure in gene evolution.

Background on molecular evolution

Addressing the question of how network structure shapes gene evolution requires a measurement of the rate of gene sequence evolution that reflects the selective pressure shaping

DNA sequence variation. Measuring evolutionary rates begins with comparing gene sequences among related organisms. Gene sequences can be aligned so that nucleotide bases stack on top of one another, suggesting their correspondence to the probable ancestral sequence. Some columns may contain all the same nucleotide, while others may vary among the species represented, reflecting the substitution of one nucleotide for another in the course of evolution. Amino acid sequences can be similarly aligned, and used to guide nucleotide alignment. This protein-guided DNA alignment is more realistic; it is unlikely nucleotides would be inserted or deleted in a protein coding region except in multiples of three. Indels create a shift in the reading frame for translating nucleotide sequences to protein, causing a change in the entire protein sequence, rather than a single amino acid in the case of a substitution.

Nucleotide and amino acid alignments are the basis on which evolutionary rates are estimated. Within a nucleotide alignment, two types of differences among or between sequences can be observed: those that cause a change in the amino acid product and those that do not. These can be translated into measures of mutation rates: the nonsynonymous and synonymous substitution rates, respectively. The nonsynonymous substitution rate is the number of nonsynonymous changes per nonsynonymous site (that is, the number of sites in which a synonymous change could occur whether it actually did or not). The synonymous substitution rate is calculated likewise. Common abbreviations for the nonsynonymous substitution rate are K_a or dN , while K_s and dS stand for the synonymous substitution rate. The overall mutation rate can be estimated by dS (Li 1997). Although synonymous substitutions can in fact affect organism function and therefore can be subject to selective constraint (Lynch 2007), this is not generally the case. It is the nonsynonymous substitutions that face the test of natural selection, because they affect protein composition and therefore enzyme performance.

While dN and dS can be useful alone, it is the ratio of these two that provides an indicator of the selective pressures shaping gene evolution. The nonsynonymous substitution rate alone does not suffice, because genes could differ in dN simply because mutation rates vary across the genome (Lynch 2007). The ratio of the nonsynonymous and synonymous substitution rates, however, provides an estimate of how readily amino acid changes are incorporated relative the underlying mutation rate. The dN/dS ratio can therefore be used to describe how natural selection is acting on a gene.

The dN/dS ratio (also denoted ω) is interpreted in terms of selective pressure as follows. If dN/dS is close to zero, the gene is evidently under strong selective constraint; virtually no amino acid changes are tolerable to organism fitness, though synonymous changes may be retained. A

dN/dS value of 1 suggests the gene is evolving neutrally: mutations are fixed by chance, not as a result of natural selection, and so nonsynonymous and synonymous substitutions are retained with equal frequency. When the dN/dS value lies between 0 and 1, the gene is under some degree of selective constraint, closer to 0 being stronger constraint and closer to 1 being weaker constraint. Such a gene is under purifying selection; most mutations are removed as detrimental, but a few remain because they are benign or beneficial. A dN/dS ratio over 1 suggests that amino acid changes are actually beneficial and that the gene is under strong positive selection, because it is unlikely that more nonsynonymous than synonymous substitutions would be fixed by chance. In this way, dN/dS can be interpreted in terms of selective pressure.

Although dN/dS is a useful indicator of selective pressure, there is some oversimplification here. It may be that only certain codons in a gene can be changed in a way that enhances fitness whereas all others cannot accept substitutions without a cost to fitness. Also, the overall selective pressure on a gene is not necessarily constant through time: as various aspects of the organism's environment change, and indeed as the cellular environment changes through alterations to other genes, our gene of interest may actually face greater or weaker selective constraint than before. A dN/dS ratio measured across a whole gene, then, is a summary statement across all codons (though codons do vary in their dN/dS) over the span of time involved in the evolutionary relationships of the sequences being compared.

Problems of rate variation among codons and among lineages or periods in evolutionary history are particularly problematic in early methods of estimating dN/dS, which are based on pairwise comparisons across whole genes. More recent methods (notably the PAML package; Yang 1997, Goldman and Yang 1998, Yang 1998) use multiple sequence alignments and allow the user to test for selection on specific codons and to measure evolutionary rates along specific branches of a phylogenetic tree representing species relationships. In this way, rate variation among codons and lineages can be taken into account. Nevertheless, these more sophisticated methods also use the dN/dS ratio as an indicator of selective pressure, whether on the whole gene or on the individual codon.

The use of dN/dS as a gauge of selective pressure touches on two great debates in molecular evolution: the debate over Neutral Theory, and the debate over the importance of mutations in coding sequences (exons) versus those in the adjacent (cis) non-coding regulatory regions as the basis of phenotypic evolution. Neutral Theory, championed by Motoo Kimura, holds that nucleotide substitutions do not become fixed because they confer improved fitness, but rather that they are fixed by chance because they are neutral with respect to fitness (Kimura 1983,

Kimura 1991). Ample evidence of adaptive gene evolution has called Neutral Theory into question, but it remains the default null model for gene evolution (Hahn 2008). The second debate flows out of the first: if coding sequences usually evolve neutrally, mutations important to the phenotype may come in the cis regulatory sequences rather than the coding sequences. Others argue that positive selection on coding sequences means phenotypic evolution is based more heavily on mutations in exons, and a great debate between these two sides rages on (e.g., Hoekstra and Coyne 2007). The focus on coding sequences here is not intended to suppose a position within this debate. Coding sequences are relevant to phenotype but may or may not be where key evolutionary changes occur.

Before returning to the main topic of measuring and explaining variation in evolutionary rates and selective pressures, we must make one additional digression. It would be wholly misleading to claim that variation in selective pressure on the gene in question is the only influence on a gene's evolution. One other effect is linkage, the physical connection of a gene to other elements on the same chromosome. Notably, linkage to another gene that is under strong positive or purifying selection can either force fixation of detrimental mutations at the gene in question, or hamper fixation of beneficial mutations. The effects of linkage are noticeable only within a certain distance from a strongly selected site (Lynch 2007). Also, the size and structure of the breeding population can influence gene evolution. In large populations, the effects of selective pressure are more fully worked out, whereas in a small population the role of mere chance in fixation of a mutation has a greater effect. This demographic effect influences the whole genome (Lynch 2007). Linkage and demography are both important to gene evolution, but it is nevertheless often possible to discern effects of selective pressure on the gene as such, beyond the noise produced by linkage and demography.

Having described measures of evolutionary rates and selective constraint, we can now turn to an earlier question: what is it about the relationship between gene and phenotype that sets up the selective pressures laid on the gene? One factor is the importance of the gene's role to the overall phenotype. For example, genes that produce floral pigments face relatively weak selective constraint (Rausher *et al.* 2008), whereas genes for histones, on which the whole structure of chromosomes rests, barely diverge between distantly related species, showing very strong selective constraint on histone proteins (Li 1997). Similarly, essential genes, those unlikely to be lost in the course of evolution, tend to evolve slowly (Krylov *et al.* 2003) and genes that affect multiple cellular processes or multiple aspects of the phenotype (multifunctional or pleiotropic genes, respectively) tend to be under greater selective constraint than those that are

not multifunctional or pleiotropic (Waxman and Peck 1998, Otto 2004, Salathé *et al.* 2006). With this we return to the original suggestion: gene interaction networks, which are the basis of how gene gives rise to phenotype, are expected to shape evolutionary rates and selective pressures. The position of a gene within an interaction network, for example, may determine how important it is to the overall phenotype, and to what degree it is pleiotropic and/or multifunctional. Therefore, we expect that network structure (topology) will show noticeable relationships to measures of evolutionary rates and selective constraint.

Review of literature on network structure and evolutionary rate variation

Literature published over the last decade demonstrates a growing interest in relationships between network topology and gene evolution (e.g., Cork and Purugganan 2004). Three advances have enabled a new network-oriented trend in the study of molecular evolution: widely available DNA sequence data, use of the mathematical theory of networks (graph theory), and increased computational power (Lehner *et al.* 2005, Proulx *et al.* 2005).

Research carried out at the intersection of network biology and molecular evolution has covered a wide variety of network types, most prominently networks of physical protein-protein interactions, but also regulatory pathways, developmental networks, signaling pathways and metabolic pathways/networks. Work with all types reveals relationships between network structure and gene evolution (Cork and Purugganan 2004). Metabolic networks, though less studied than protein-protein interaction networks, offer some of the greatest insights. Metabolism is more clearly linked to phenotype than protein-protein interactions are, offering a clearer path to understanding the link between gene/genome and phenotype, a key goal in biology (Vitkup *et al.* 2006). Protein-protein interaction networks, though more often studied, offer less insight into critical biological questions because they lump together all physical protein interaction, whether regulatory, metabolic, or structural, thereby obliterating much opportunity to relate network properties to specific aspects of phenotype and fitness (Vitkup *et al.* 2006). Therefore, I will orient this review to studies of metabolic pathways/networks, but make reference to studies on other systems to demonstrate what is specific to metabolism and what is common to biological networks generally.

Networks: definition and structure

Before considering how network structure shapes gene evolution, we must begin with a solid definition of networks. A network is a collection of units or nodes that are connected by

directional or non-directional edges (Higgins 2007), a group of actually or potentially interacting elements (Proulx *et al.* 2005). In a metabolic network, nodes are enzymes and edges are the metabolites that are the product of one enzyme and the substrate of the next. In protein-protein interaction networks nodes are simply proteins of any kind, and edges represent physical contact.

Cellular networks, whether metabolic, protein-protein interaction or otherwise tend to be small-world (or scale-free) networks (Jeong *et al.* 2001, Wagner and Fell 2001, Jordan *et al.* 2004; but see Bader 2006). These networks have a few highly-connected nodes and many nodes with few connections. Oddly enough, this is exactly the same structure as many non-biological networks, including the Internet and social connections (Higgins 2007). These highly connected nodes result in a short average distance between nodes compared to the number of nodes in the network (Proulx *et al.* 2005). In metabolic networks, highly connected nodes tend to also be central to the network, having a short average distance to other nodes (Vitkup *et al.* 2006). This structure lends a certain degree of robustness. Most nodes, if disrupted, cause little overall perturbation (though targeted removal of highly connected nodes is more problematic; Albert *et al.* 2000).

This scale-free structure may have evolved through preferential node attachment (Eisenberg and Levanon 2003), with or without selection for robustness as such, around a core metabolism shared by virtually all organisms (Proulx *et al.* 2005, Prachumwat and Li 2006). Thus, network structure may have arisen for non-adaptive reasons, or due to selection for some other role than it now plays, as with many major events in evolution (Lynch 2007).

Node connectivity and selective constraint

Scale-free network structure results in stronger selective constraint on nodes with greater connectivity. Highly-connected nodes are vulnerable to attack (Albert *et al.* 2000), and their loss has effects that reverberate through wide sectors of the overall network, reducing overall fitness. These predictions are borne out quite well in the yeast (*Saccharomyces cerevisiae*) metabolic network; using more than 670 ORFs identified by Forster *et al.* (2003), Vitkup *et al.* (2006) show that dN/dS correlates with node degree (connectivity). This arises from functional constraints, since highly connected enzymes are required for more metabolic processes (Vitkup *et al.* 2006). Protein-protein interaction networks also show a similar correlation of connectivity and evolutionary rate, but due to structural constraints (Fraser *et al.* 2002, Fraser *et al.* 2003, Jordan *et al.* 2003a, b, Hahn *et al.* 2004, Makino *et al.* 2006).

Node connectivity is also correlated with retention of gene duplicates. Vitkup *et al.* (2006) show that highly connected nodes are more likely to retain duplicates, reminiscent of work showing that genes with critical functions tend to retain duplicates after whole genome duplication (Chapman *et al.* 2006). Critical nodes in metabolic networks, therefore, not only evolve slowly but are often protected from failure through gene duplication. In contrast, Prachumwat and Li (2006) found a negative correlation between protein connectivity and duplicability in the yeast protein-protein interaction network, showing that gene duplication may have a different relationship to connectivity in different network types.

Metabolic flux and selective pressure

Besides enzyme connectivity, rates of reaction and flux through the network have also been shown to correlate with evolutionary rates. Vitkup *et al.* (2006) show that under normal growth conditions for yeast (aerobic growth on glucose or fructose), enzymes with high metabolic flux have low K_a/K_s and are more likely to retain gene duplicates, suggesting greater selective constraint on these enzymes. On the other hand, enzymes at pathway branchpoints, which are important for controlling flux, may come under positive selection. This has been shown in *Drosophila* glycolysis (Flowers *et al.* 2007, Eanes 1999) and *Zea mays* starch synthesis (Whitt *et al.* 2002). Selection acts on the total phenotypic outcome of network performance (Proulx *et al.* 2005); these examples show that critical proteins can both bear the burden of selective constraint and/or evolve adaptively in response to needs for a metabolic shift.

Effects of position in metabolic networks

Another feature of metabolic networks relevant to evolutionary rate variation is the position of each enzyme relative other nodes in the network. One aspect of position is centrality, the average number of nodes that lie between one node and any other in the network. Vitkup *et al.* (2006) have shown that highly connected nodes are also central nodes. This suggests that centrality is related to selective constraint, but this is not formally analyzed (Vitkup *et al.* 2006). Hahn and Kern (2005) show that in yeast, worm and fly protein-protein interaction networks, central proteins evolve more slowly and are more essential to organism survival, regardless of protein connectivity.

While centrality provides one measure of position within a network, another measure is the position along directional paths through the network. A pathway in the metabolic sense is a

series of biochemical reactions that produces a particular end product from a particular substrate; more generally, a biological pathway is an ordered series of reactions or changes (Oxford English Dictionary), which may be metabolic, developmental or involved in signal transduction. In term of network theory, a path is a route traced through the network from any one node to another, traversing any nodes between them via the edges between nodes (Higgins 2007).

The order of biochemical steps, or enzymes, along a pathway should influence the evolutionary rates of genes for those enzymes (Rausher *et al.* 1999). Arguments for such a hypothesis can be made based on centrality, flux control and pleiotropy. Central elements in metabolism tend to be those that provide substrates for many other enzymes (Fell and Wagner 2000, Jeong *et al.* 2000). Since the more central enzymes lie upstream in metabolic flow, genes for upstream enzymes should evolve more slowly than those for downstream enzymes. Likewise, enzymes controlling the rate of flux are often, roughly speaking, concentrated in the more upstream areas of metabolic pathways (Rausher *et al.* 2008). Either of these alone could produce lower evolutionary rates upstream and higher rates downstream, a pattern called positional rate variation (Rausher *et al.* 1999). A third and very convincing argument for the positional rate variation hypothesis rests on the relationship between pleiotropy and the structure of metabolic pathways. Metabolic pathways often branch in such a way that upstream enzymes produce substrates used to synthesize a wide variety of metabolites, whereas enzymes downstream are required for only a few products. Both pleiotropy (multiple aspects of phenotypic influenced by one gene) and multifunctionality (multiple biochemical processes influenced by one gene) produce selective constraint (Waxman and Peck 1998, Otto 2004, Salathé *et al.* 2005). Therefore, upstream enzymes should evolve more slowly than downstream enzymes due to a gradient of selective constraint along the pathway.

Positional rate variation was first demonstrated in the anthocyanin biochemical pathway. Nonsynonymous substitution rates (dN) vary with pathway position for pairwise comparisons among *Zea*, *Antirrhinum* and *Ipomoea* for six genes in sequence starting from the first committed step in flavonoid synthesis (Rausher *et al.* 1999). This trend holds with dN/dS for the most upstream enzyme, CHS and the two downstream enzymes ANS and UFGT in comparisons among various *Ipomoea* species (Lu and Rausher 2003, Rausher *et al.* 2008). Various tests for positive selection show no evidence of positive selection driving this gradient of evolutionary rates. Rather, selective constraint is relaxed along the pathway (Lu and Rausher 2003, Rausher *et al.* 2008).

Apart from the anthocyanin pathway, there is little evidence for positional rate variation. Two studies suggest a similar trend but do not rigorously examine it. First, Sharkey *et al.* (2005) performed a phylogenetic analysis of three methylerythritol phosphate (MEP) pathway genes (which produce isoprenoid precursors) and isoprene synthase (which produces isoprene, thought to play a role in plant thermotolerance). Percent amino acid identity and percent nucleotide identity (rough indicators of selective constraint) are both higher in the upstream MEP proteins/genes than in the downstream isoprene synthase protein/gene, with 80-90% amino acid identity for the MEP genes but 57% for isoprene synthase.

A second study to suggest further evidence for positional rate variation examines 55 loci involved in sialic acid biology, covering the synthesis, activation and recognition of molecules involved in establishing cell identity (Altheide *et al.* 2006). Genes involved in biosynthesis and activation of these recognition molecules had lower pairwise Ka/Ks between human and chimpanzee than those involved in recognition of specific sialic acids (0.465 for average Ka/Ks for recognition proteins, versus 0.293 for biosynthesis). Altheide *et al.* (2006) suggest that this is an example of Rausher *et al.*'s (1999) phenomenon, but strictly speaking it is not, because pathways of sialic acid biology include interactions beyond metabolism. Although Sharkey *et al.* (2005) and Altheide *et al.* (2006) suggest evolutionary rate variation along pathways involving metabolism, neither provides a rigorous confirmation of positional rate variation.

A recent study of *Arabidopsis* phenylpropanoid metabolism, however, does rigorously explore the possibility of positional rate variation. Ramos-Onsins *et al.* (2008) survey nine phenylpropanoid genes in *Arabidopsis thaliana* and *A. lyrata* as candidate genes for adaptive evolution, since the phenylpropanoid pathway produces many metabolites important for adaptation to specific environments. The study suggests that the loci were all under selective constraint, not positive selection. While Ramos-Onsins *et al.* (2008) did not calculate dN/dS, tests for a correlation of pathway position and five other measures of gene sequence divergence revealed no significant correlation. Ramos-Onsins *et al.* (2008) conclude that demographic as opposed to selective forces are shaping variation at these loci, but suggest that their use of partial gene sequences may also be at fault.

In summary, the literature on positional rate variation is limited to work on four pathways. Only work on the anthocyanin pathway offers strong evidence (Rausher *et al.* 1999, Lu and Rausher 2003, Rausher *et al.* 2008), while excellent work on phenylpropanoid metabolism reveals none (Ramos-Onsins *et al.* 2008). Research on isoprene synthesis and sialic acid biology each offer anecdotal evidence of positional rate variation, but without the rigor of earlier work by

Rausher's group. The positional rate variation hypothesis offers an intriguing addition to the study of molecular evolution in gene interaction networks, but further evidence is needed to establish positional rate variation as a general phenomenon.

Proposal for further research

Further research on positional rate variation is needed to better understand the role of metabolic networks in shaping gene evolution. In order to carry out a sufficiently rigorous study of positional rate variation, three points need to be carefully considered: choice of metabolic pathway, method for testing for positional rate variation, and species from which to draw gene sequence data.

Choice of study system

An effective study of positional rate variation must examine an appropriate metabolic pathway. Because a pathway is simply a route from a chosen beginning to a chosen end, not all pathways should be expected to show positional rate variation. The anthocyanin pathway, the only solid example of PRV to date, is highly branched (Winkel 2006) so that each of the six enzymes in series represents a successive decrease in pleiotropy. Where pathway branchpoints are not so densely packed, pathway position may not correspond to pleiotropy. In fact, a hypothetical linear pathway with no branchpoints should in fact have no difference in pleiotropy among pathway members, and any rate variation should come from demographic effects or from differences in protein folding constraints among the enzymes, not from pathway structure. Therefore, an appropriate study system will be a long pathway that crosses several pathway branchpoints.

Besides choosing a long pathway with many branchpoints, an improved study of positional rate variation must consider not one but many branches of a metabolic system. If positional rate variation is truly a general effect, then the many downstream branches of the upstream pathway elements should all show the same trend of higher evolutionary rates downstream. This will prove that any evidence for positional rate variation is not simply the vagaries of one particular pathway branch. Plant terpenoid synthesis, described in detail in Chapter 2, is ideal for studying positional rate variation because of its length, many branchpoints, and many downstream branches with genetic data available for analysis.

Improved methods for detecting positional rate variation

In testing for positional rate variation, Ramos-Onsins *et al.* (2008) examined correlations of nucleotide divergence with pathway position. Pathway position, however, may not be the optimal metric for detecting positional rate variation. As described above, pleiotropy should decrease not from enzyme to enzyme, but from one pathway branch point to the next. Therefore, the order of *groups of enzymes* between pathway branchpoints should correlate with dN/dS more strongly than pathway position itself, even in cases where pathway position reveals no significant correlation. I call this metric the Pathway Pleiotropy Index (PPI). Besides tracking changes in pleiotropy along pathways, I expect the PPI to have a second advantage: some metabolic pathways are cyclical or reticulate, making pathway position difficult to assign. With the PPI, all enzymes between branchpoints are counted as a single unit regardless of whether they act in a predictable order or not.

Choice of taxa

A study of positional rate variation requires study species that have adequate genetic resources and that span diverse lineages. In terms of genetic resources, species with whole genome sequences are an excellent choice. Full coding sequences will normally be available for all genes, and the electronic format allows a bioinformatic approach, which is appropriate for a first look at a new pathway. Also, having a total genome sequence allows more reliable ortholog determination, since the total sequence space can be screened. Orthologs are genes from different taxa that are similar due to common descent, as opposed to paralogs, which are similar due to a gene duplication shared by those taxa. Accurate determination of orthology is essential for measuring rates of evolution along the chosen lineages, without unintentionally incorporating divergence of duplicates that occurred before the lineages separated.

Choosing several diverse species provides a basis for commenting on the generality of positional rate variation. Even if positional rate variation is a general phenomenon in genome evolution, not all lineages may exhibit such an effect. First, as Ramos-Onsins *et al.* (2008) found, demographic effects can outweigh selection in a species with small population sizes. Second, different lineages have different histories of genome duplication and retention of duplicates. Additional gene copies could relax the constraining effects of pleiotropy and obscure effects of pathway position if gene copy number is higher upstream than downstream in a particular lineage.

Oryza, *Vitis*, *Arabidopsis*, *Ricinus* and *Populus* are the first five angiosperms to have full genome sequences, and offer an appropriate setting for studying positional rate variation. First, the whole genome sequences allow for exhaustive searches during ortholog determination. The whole genome also allows a total count of gene copy number for examining whether this variable can explain cases in which positional rate variation is not found. Second, these species span a variety of historical effective population sizes (Bergelson *et al.* 1998, Tuskan *et al.* 2006) and histories of genome duplication. *Vitis*, *Arabidopsis*, *Populus* and *Oryza* all have genome duplication events after their divergence from the other species in this group (Barker *et al.* 2007).

Research objectives

Based on the recommendations above, my collaborators and I have carried out a study of positional rate variation in the plant terpenoid synthesis pathway, addressing the following four objectives:

1. To test the strength of correlation between pathway position and dN/dS and between the Pathway Pleiotropy Index and dN/dS
2. To evaluate the merit of the Pathway Pleiotropy Index over pathway position for detecting positional rate variation
3. To observe whether evolutionary lineages vary in exhibiting positional rate variation among any given pathway of enzymes
4. To consider whether gene copy number can explain why certain lineages do not show positional rate variation

Detailed methods and results of this study are described in chapter 2.

References

- Albert R., Jeong H, Barabasi AL. 2000. Error and attack tolerance of complex networks. *Nature*. 406: 378-382.
- Altheide TK, Hayakawa T, Mikkelsen TS, Diaz S, Varki N, Varki A. 2006. System-wide genomic and biochemical comparisons of sialic acid biology among primates and rodents - evidence for two modes of rapid evolution. *J Biol Chem*. 281: 25689-25702.
- Bader J. 2006. The *Drosophila* protein interaction network may be neither power-law nor scale-free. In: Koonin EV, Wolf YI, Karev GP, editors. *Power Laws, Scale-Free Networks and Genome Biology*. New York: Springer. p. 53-64.
- Barker MS, Kane N, Rieseberg L. 2007. Widespread paleopolyploidy across the Viridiplantae [poster]. In: *Botany & Plant Biology Joint Congress; 2007 July 7-11; Chicago, IL*.
- Bergelson J, Stahl E, Dudek S, Kreitman M. 1998. Genetic variation within and among populations of *Arabidopsis thaliana*. *Genetics*. 148:1311-1323.
- Chapman BA, Bowers JE, Feltus FA, Paterson AH. 2006. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicity to angiosperm genome duplication. *P Natl Acad Sci USA*. 103: 2730-2735.
- Cork J, Purugganan M. 2004. The evolution of molecular genetic pathways and networks. *BioEssays*. 26: 479-484.
- Eanes WF. 1999. Analysis of selection on enzyme polymorphisms. *Annu Rev Ecol Syst*. 30: 301-326.
- Eisenberg E, Levanon EY. 2003. Preferential attachment in the protein network evolution. *Phys Rev Lett*. 91: 138701.
- Fell DA, Wagner A. 2000. The small world of metabolism. *Nat Biotech*. 18: 1121-1122.
- Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, Eanes WF. 2007. Adaptive evolution of metabolic pathways in *Drosophila*. *Mol Biol Evol*. 24: 1347-1354.
- Forster J, Famili I, Fu P, Palsson BO, Nielsen J. 2006. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*. 13: 244-253.
- Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. *Science*. 296: 750-752.
- Fraser H, Wall D, Hirsh A. 2003. A simple dependence between protein evolution rate and the number of protein-protein interactions. *BMC Evol Biol*. 3: 11.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 11: 725-736.

- Hahn, MW. 2008. Toward a selection theory of molecular evolution. *Evolution*. 62: 255-265.
- Hahn MW, Conant GC, Wagner A. 2004. Molecular evolution in large genetic networks: does connectivity equal constraint? *J Mol Evol*. 58: 203-211.
- Hahn MW, Kern AD. 2005. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol*. 22: 803-806.
- Higgins PM. 2007. *Nets, Puzzles and Postmen: an exploration of mathematical connections*. Oxford; New York: Oxford University Press.
- Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution*. 61: 995-1016.
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. 2000. The large-scale organization of metabolic networks. *Nature*. 407: 651-654.
- Jeong H, Mason SP, Barabasi AL, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature*. 411: 41-42.
- Jordan IK, Wolf Y, Koonin E. 2003. No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. *BMC Evol Biol*. 3: 1.
- Jordan IK, Wolf Y, Koonin E. 2003. Correction: No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. *BMC Evol Biol*. 3: 5.
- Jordan IK, Marino-Ramirez L, Wolf YI, Koonin EV. 2004. Conservation and coevolution in the scale-free human gene coexpression network. *Mol Biol Evol*. 21: 2058-2070.
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge; New York: Cambridge University Press.
- Kimura M. 1991. The neutral theory of molecular evolution - a review of recent evidence. *Jpn J Genet*. 66: 367-386.
- Krylov DM, Wolf YI, Rogozin IB, Koonin EV. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res*. 13: 2229-2235.
- Lehner B, Tischler J, Fraser A. 2005. Systems biology: where it's at in 2005. *Genome Biol*. 6:338.
- Li WH. 1997. *Molecular evolution*. Sunderland, Massachusetts: Sinauer.
- Lu YQ, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. *Mol Biol Evol*. 20: 1844-1853.

- Lynch M. 2007. The origins of genome architecture. Sunderland, MA: Sinauer Associates.
- Makino T, Suzuki Y, Gojobori T. 2006. Differential evolutionary rates of duplicated genes in protein interaction network. *Gene*. 385: 57-63.
- Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
- Otto SP. Two steps forward one step back: the pleiotropic effects of favoured alleles. *P Roy Soc B-Biol Sci*. 271: 705-714.
- Prachumwat A, Li WH. 2006. Protein function, connectivity, and duplicability in yeast. *Mol Biol Evol*. 23: 30-39.
- Proulx SR, Promislow DEL, Phillips PC. 2005. Network thinking in ecology and evolution. *Trends Ecol Evol*. 20: 345-353.
- Ramos-Onsins SE, Puerma E, Balañá-Alcaide, Salguero D, Aguadé M. 2008. Multilocus analysis of variation using a large empirical data set: phenylpropanoid pathway genes in *Arabidopsis thaliana*. *Mol Ecol*. 17: 1211-1223.
- Rausher M.D, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol*. 16: 266-274.
- Rausher MD, Lu YQ, Meyer K. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *J Mol Evol*. 67: 137-144.
- Salathé M, Ackermann M, Bonhoeffer S. 2006. The effect of multifunctionality on the rate of evolution in yeast. *Mol Biol Evol*. 23: 721-722.
- Sharkey TD, Yeh S, Wiberley AE, Falbel TG, Gong DM, Fernandez DE. 2005. Evolution of the isoprene biosynthetic pathway in Kudzu. *Plant Physiol*. 137: 700-712.
- Tuskan GA, DiFazio S, Jansson S, *et al.* [107 coauthors]. 2006. The genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 313: 1596-1604.
- Vitkup D, Kharchenko P, Wagner A. 2006. Influence of metabolic network structure and function on enzyme evolution. *Genome Biology*. 7: R39.
- Wagner A, Fell DA. 2001. The small world inside large metabolic networks. *P Roy Soc B-Biol Sci*. 268: 1803-1810.
- Waxman D, Peck JR. 1998. Pleiotropy and the preservation of perfection. *Science*. 279: 1210-1213.
- Whitt SR, Wilson LM, Tenailon MI, Gaut BS, Buckler ES. 2002. Genetic diversity and selection in the maize starch pathway. *P Natl Acad Sci USA*. 99: 12959-12962.

- Winkel BSJ. 2006. The biosynthesis of flavonoids. In: Grotewold E, editor. The science of flavonoids. Columbus (OH): Springer. p. 71-96.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13: 555-556.
- Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 15: 568-573.

CHAPTER 2: The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis¹

Introduction

As species diverge through evolution, their genes diverge at unequal rates, and genomes become a mosaic of conservation and divergence. These differences in divergence can be explained as differential selective forces acting on the various loci, but what sets up these differences in selective pressure among genes?

Networks of gene interactions offer an intriguing context for answering this question. Functional genomics and whole-genome sequencing have opened the way to construct accurate representations of cellular networks and to more fully understand interactions among genes. Studies on a variety of network types (metabolic, protein-protein interaction, signaling, development) have shown that network properties such as node connectivity and centrality within the network correlate with evolutionary rate variation in a variety of species (Cork and Purugganan 2004). For example, highly connected nodes (i.e., enzymes or other proteins) consistently undergo stronger selective constraint than nodes with fewer connections in both metabolic and protein-protein interaction networks (Fraser *et al.* 2002, Fraser *et al.* 2003, Jordan *et al.* 2003a, b, Hahn *et al.* 2004, Makino *et al.* 2006, Vitkup *et al.* 2006).

A less explored aspect is the structure of metabolic pathways. Many known metabolic pathways are structured such that the initial substrates in the pathway are ultimately transformed into several end products, based on the channeling of substrate into various branches of the pathway downstream. Therefore, upstream genes should be more pleiotropic than those downstream, and so should face greater selective constraint (Waxman and Peck 1998, Otto 2004). This gradient of selective constraint should then produce low evolutionary rates upstream and high rates downstream, a pattern of evolutionary rate variation that has been termed "positional rate variation" (Rausher *et al.* 1999).

Positional rate variation (PRV) has been rigorously demonstrated in the anthocyanin pathway of higher plants. Rausher *et al.* (1999) show that among six enzymes involved in anthocyanin synthesis, the three upstream enzymes evolve more slowly than the three downstream enzymes. Lu and Rausher (2003) and Rausher *et al.* (2008) confirm that this is due to a relaxation of selective constraint, not stronger positive selection, on downstream genes.

¹ A version of this chapter has been submitted for publication.
Ramsay H, Rieseberg L, Ritland K. 2008. "The correlation of evolutionary rate and pathway position in plant terpenoid biosynthesis."

These studies establish an occurrence of PRV that holds true in three plant lineages and that can be demonstrated over both recent and ancient evolutionary time scales.

Yet, is positional rate variation truly a general effect in genome evolution? A study of four genes involved in plant isoprene synthesis suggests a similar trend of stronger selective constraint on upstream enzymes (Sharkey *et al.* 2005), but work on nucleotide variation in *Arabidopsis thaliana* phenylpropanoid metabolism reveals no correlation of pathway position and various measures of nucleotide divergence (Ramos-Onsins *et al.* 2008). Clearly, positional rate variation requires further study to be established as a general phenomenon.

In this study, we examine whether positional rate variation occurs among 45 genes involved in plant terpenoid synthesis, using data from five fully sequenced angiosperms. To capture meaningful changes in pleiotropy along the pathway, we trace the flow of metabolites from glycolysis through to the synthesis of four terpenoid derivatives. As the most upstream portion of this system, glycolysis plays a critical role in producing small precursors that are required for nearly every primary biosynthetic process (Plaxton 1996, Heldt and Heldt 2005). One of these processes is the synthesis of terpenoid precursors such as farnesyl diphosphate, which is also used in the production of a wide range of metabolites needed for growth and plant defense (Cheng *et al.* 2007, Liu *et al.* 2005). These include lutein, a carotenoid important for the assembly and function of the photosynthetic light harvesting complex (Pogson *et al.* 1996), as well as the phytohormones abscissic acid, gibberellic acid and brassinosteroids (Buchanan *et al.* 2000). The enzymes involved in the transformation of glucose to these four end products are therefore less pleiotropic downstream than they are upstream, and offer an ideal setting to study positional rate variation. The entire pathway is depicted in Figure 2.1 (information derived from Plaxton 1996, Clouse and Sasse 1998, Szekeres and Koncz 1998, Buchanan *et al.* 2000, Hedden and Phillips 2000, Naik *et al.* 2003, Sakamoto *et al.* 2004, Heldt and Heldt 2005, Liu *et al.* 2005, Nambara and Marion-Poll 2005, Grennan 2006, Cheng *et al.* 2007, KEGG 2008).

To improve detection of positional rate variation, we use a new measure termed the "Pathway Pleiotropy Index." Rather than counting enzymes in sequence from most upstream to most downstream, the Pathway Pleiotropy Index (PPI) numbers *groups of enzymes* relative pathway branchpoints. We compare this measure with the simpler approach of correlating evolutionary rates with pathway positions. We expect that the PPI will correlate more strongly with evolutionary rates because it tracks changes in pleiotropy along the pathways; enzymes between pathway branchpoints are responsible for the same set of products and therefore do not differ in pleiotropy. In fact, the anthocyanin pathway may show such strong PRV precisely

because virtually every enzyme acts as a pathway branch point and therefore each enzyme is successively less pleiotropic (Harbourne 1988, Winkel 2006).

In addition, if dN/dS does not correlate with pathway position, it may be that rate variation is more strongly influenced by another factor that mitigates the role of pathway structure. For example, after gene duplication, selective constraint is relaxed for a time, ostensibly due to redundancy (Lynch and Conery 2000). Also, gene duplication may offer an escape from the constraining effects of pleiotropy (Hoekstra and Coyne 2007), resulting in a negative correlation of copy number with pathway position, rather than a positive correlation of pathway position and dN/dS. Therefore, variation in gene copy number may explain instances where PRV is not found.

Here, we examine whether plant terpenoid synthesis exhibits positional rate variation by addressing four specific questions. First, we test whether statistically significant variation in dN/dS exists among enzymes. Second, we test whether this variability is explained by pathway position and/or by PPI. Third, we examine whether dN/dS varies with gene copy number, and whether copy number varies with pathway position. Finally, we use likelihood-based methods (Church *et al.* 2007) to explore whether differences in dN/dS arise from variation in selective constraint versus positive selection.

Methods

Study species

We used all angiosperm genomes that were available at the time of study: *Oryza sativa*, *Vitis vinifera*, *Arabidopsis thaliana*, *Ricinus communis* and *Populus trichocarpa*. Protein and coding sequences for each species were downloaded from the following websites on March 27, 2008:

http://www.genoscope.cns.fr/externe/Download/Projets/Projet_ML/data/annotation/

ftp://ftp.Arabidopsis.org/home/tair/Genes/TAIR7_genome_release/TAIR7_blastsets

http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.download.ftp.html

http://castorbean.tigr.org/castorbean_downloads.shtml

ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_5.0/all.chrs

Finding orthologs and counting paralogs

Using the KEGG Pathway database (2008), 47 genes involved in the conversion of glucose to lutein, abscissic acid (ABA), gibberellic acid (GA), and brassinosteroids (BR) were selected. *Arabidopsis* protein sequences for the enzymes were gathered from the KEGG Pathway database February 27, 2008 (glycolysis enzymes were gathered May 23, 2008). A list of genes and accession numbers for these queries are provided in Appendix 1. The *Arabidopsis* protein queries were BLASTed against all proteins from all five genomes (BLASTP; Altshul *et al.* 1990). A multi-sequence query was used if more than one sequence was given in the KEGG Pathway database (2008). Blast results were parsed for 25% amino acid identity over 50% of the query sequence length (if more than one query was used, this was taken to be 50% of the length of the shortest query). Parsed hits for all species were aligned using MUSCLE (Edgar 2004). If the alignment quality was poor, BLAST results were re-parsed by increasing the percent identity cutoff by 5% until the resulting multiple alignment was acceptable. The initial methionine was deleted before alignment in all cases. Next, these alignments were used to construct maximum likelihood protein trees with PhyML (Guindon and Gascuel 2003). For each alignment PhyML was run with the following parameters: JTT with 4 gamma categories; estimated proportion of invariable sites and gamma distribution parameter; 100 bootstrap replicates; neighbor-joining starting tree; and optimized both topology and branch lengths.

The resulting trees were visually inspected, and the orthologous group was chosen as that clade with 50% or greater bootstrap support which contained the *Arabidopsis* query as well as at least one representative of each of the other four taxa. If no such clade could be found, the gene was not included in any further analysis. If any taxon had more than one sequence in this clade, the sequence on the shorter branch was chosen. For *Arabidopsis*, the chosen ortholog was always one of the query sequences. Paralogs were counted as the number of sequences for each lineage in the ortholog clade. Names of orthologs, as used in the genome file for each species, are provided in Appendix 2.

Measuring evolutionary rates

After choosing one ortholog from each of the five taxa, protein-guided DNA alignments of coding sequences were constructed using an alignment pipeline provided by Mike Barker. This pipeline was used because not all coding sequences were in frame. These alignments were visually inspected. Only four genes contained ambiguously aligned regions; analyses were carried out with both the original alignments and with the ambiguous region were masked in

these four genes in order to ensure that alignment ambiguity did not sway results (alignments are provided in Appendix 6).

Using this alignment of five orthologs, dN/dS ratios along each lineage were calculated in CODEML (Yang 1997) using the tree (*Oryza*, *Vitis*, (*Arabidopsis*, (*Ricinus*, *Populus*))), with model = 1 (free ratios), and getSE = 1 (obtain standard errors of dN/dS estimates). Five replicate runs were carried out and the run with greatest absolute value of lnL was selected. From this selected run, dN/dS values for *Vitis*, *Arabidopsis*, *Ricinus* and *Populus* were estimated (*Oryza* served as an outgroup; see Figure 2.2; tree topology from Soltis *et al.* 2005). Values of dN/dS for which dS was less than 0.005 or more than 2.00 were discarded, as were values for which S*dS was less than 10.

Variability of dN/dS

Using standard errors from the CODEML output, we tested whether dN/dS varies among genes of each pathway for each lineage based on the following test statistic, summed over all n genes in a pathway:

$\Sigma[(\omega_i - \mu)^2 / \sigma_i^2]$ where ω_i is the dN/dS value for gene i , σ_i is the standard error for ω_i (from CODEML), and μ is the mean ω across all genes sampled (calculated separately for each lineage, summed over all genes). This statistic is distributed as χ^2 with $n-1$ degrees of freedom.

Relationship of dN/dS to pathway structure and gene copy number

To determine the relationship of dN/dS to each of two measures of pathway position, Kendall's τ rank correlation coefficient was calculated using JMP 7.0. First, Kendall's τ was calculated using pathway position. Enzymes were sequentially numbered from the first enzyme involved in the conversion of glucose to pyruvate through to the last enzyme involved in the production of each of the four end products. Second, Kendall's τ was calculated using our Pathway Pleiotropy Index (PPI). Groups of enzymes that act between pathway branchpoints were numbered from most upstream to most downstream (see Figure 2.1). If an enzyme (such as CrtL-b or CrtR-b; see KEGG Pathway Database 2008) was involved in more than one path downstream of a branch point, it was counted as part of the group above that branch point. (Creating a separate category for such genes made no difference to our results). Since the synthesis of abscissic acid, gibberellic acid, brassinosteroids and lutein from glucose is a branched pathway, the correlation of dN/dS with pathway position and with the PPI was

calculated separately for the path from glucose to each end product. The robustness of these correlations was tested by removing the highest dN/dS value and recalculating the correlation to ensure that significant results were not driven by such outliers. The relationship of dN/dS to gene copy number was examined visually, and since no apparent relationship was seen, Kendall's τ was not calculated for this variable. The relationship of gene copy number and pathway position was also examined visually.

Tests for selection

To verify whether any pattern of rate variation is due to relaxation of selective constraint, and not positive selection, we tested for selection using the following comparisons of models in CODEML (Goldman and Yang 1994, Yang 1997, Yang 1998), after Church *et al.* (2007). First, we tested for variable rates among codons by comparing likelihood values under M3 and M0. Second, we tested for a class of codons with dN/dS > 1 by comparing likelihood values under M2a and M1a. Likelihood values were compared between each pair of models, but the dN/dS values for sites were ignored as these may not be reliable (Church *et al.* 2007, and references therein). The same alignments as for dN/dS calculation were used to test for selection, except that the outgroup *Oryza* was omitted. This test of selection is the only point in the analysis in which lineages were not treated separately.

Results

Ortholog availability and rate variation among enzymes

Of 47 genes examined, orthologs were found for 45. After discarding dN/dS values with dS above 2.00 or below 0.005, all 45 genes remained for *Ricinus* and *Populus*. *Arabidopsis* retained 29 and *Vitis* 39 dN/dS values. All four ingroup lineages showed statistically significant variation in dN/dS for enzymes along the abscissic acid, gibberellic acid, and brassinosteroid pathways. No statistically significant variation was found for the lutein pathway, except in *Vitis*. In each lineage, dN/dS varied by an order of magnitude (dN/dS ranged from 0.033-0.28 for *Arabidopsis*, 0.047-0.28 for *Populus*, 0.031-0.29 for *Ricinus*, and 0.048-0.59 for *Vitis*).

Variation of dN/dS with pathway position

Variability in dN/dS correlated significantly with pathway position for 9 of 12 pathway-lineage combinations (robust to removal of the highest dN/dS value for 5 of 12; Table 2.1). The strongest correlations are for synthesis of gibberellic acid and abscissic acid (mean Kendall's τ 0.46 and 0.38, respectively, versus 0.30 for brassinosteroid synthesis). Scatterplots of data for *Populus* are given in Figure 2.3(a-d); plots of data from other lineages can be found in Appendices 3-5(a-d).

Variation of dN/dS with the Pathway Pleiotropy Index

Variability in dN/dS correlated significantly with PPI for 10 of 12 pathway-lineage combinations (robust to removal of the highest dN/dS value for 9 of 12; Table 2.2). Here, strong correlations are seen for synthesis of all three phytohormones, though synthesis of gibberellic acid has by far the highest rank correlation coefficients (mean Kendall's τ of 0.54 for synthesis of gibberellic acid, versus 0.42 and 0.40 for abscissic acid and brassinosteroid synthesis, respectively). Data for *Populus* are shown in Figure 2.3(e-h); plots of data from other lineages can be found in Appendices 3-5(e-h).

Gene copy number

Gene copy number varied among genes in all four lineages. *Vitis*, *Arabidopsis*, and *Ricinus* had mostly one copy per gene, while most *Populus* genes had two copies. Copy number was 1, 2 or 3 for *Vitis*; 1, 2, 3, 4, or 11 for *Arabidopsis*; 1 or 2 for *Ricinus*; and 1, 2, or 3 for *Populus*. Genes with two or more copies appeared to be concentrated in the downstream end of gibberellic acid synthesis (for *Vitis* and *Arabidopsis*), brassinosteroid synthesis (for *Populus* and *Ricinus*) and abscissic acid synthesis (for *Vitis* and *Ricinus*), but in all other cases duplicate and singleton genes were equally distributed along the pathways. Gene copy number bore no apparent relationship to dN/dS.

Selection analyses

All enzymes showed variable dN/dS among codons, but models including a class of codons with dN/dS > 1 were not significantly more likely than models without. In all the above analyses, results using alignments altered to exclude ambiguously aligned region produced qualitatively similar results.

Discussion

Positional rate variation (PRV) is a pattern of evolutionary rate variation wherein genes evolve under relaxed selective constraint if they code for enzymes downstream in a metabolic pathway, but face greater selective constraint if the enzyme acts upstream (Rausher *et al.* 1999). The data from plant terpenoid synthesis presented here clearly demonstrate PRV. This is the only clear example of PRV to date besides that for anthocyanin synthesis (Rausher *et al.* 1999, Lu and Rausher 2003, Rausher *et al.* 2008).

The research presented here is so far the most extensive survey of PRV, the only comparable studies being the work of Rausher and colleagues (Rausher *et al.* 1999, Lu and Rausher 2003, Rausher *et al.* 2008). We have incorporated sequence data for enzymes along not one but four branches of a metabolic system, examined 45 genes (more than seven times as many as previous studies) and used explicit ortholog determination by phylogenetic methods (considerably more reliable than reciprocal best hit methods; Poptsova and Gogarten 2007). Because Kendall's τ is a conservative test of correlation, our results are very robust. This work is made possible by advent of whole genome sequences, and by advances in functional genomic studies that elucidate pathways of gene interactions (e.g., Rodríguez-Concepción and Boronat 2002).

In this study, we expected that (1) longer pathways would show stronger PRV, and that (2) an index that accounts for the position of an enzyme relative pathway branchpoints, "PPI," would better correlate with dN/dS than pathway position as such. The long pathways in our study do show strong upstream vs. downstream trends in dN/dS, and examination of the graphs reveals that such a relationship is not generally evident in the separate "pathways" (such as glycolysis) that this longer system can be subdivided into. Notably, the variation in dN/dS in terpenoid synthesis is equivalent to that in the anthocyanin pathway. In the four lineages studied here, dN/dS varied from 0.04 upstream to 0.30 downstream, an order of magnitude difference. Rausher *et al.* (2008) found dN/dS to vary from 0.027 upstream to 0.220 downstream in the anthocyanin pathway, also an order of magnitude difference. This confirms our initial prediction, that the anthocyanin pathway exhibits strong PRV because of the close packing of pathway branchpoints.

Our "pathway pleiotropy index" (PPI) offers modest but noticeable improved detection of trends, compared to the method of simply considering the position of enzymes in the pathway.

First, the PPI produces values of Kendall's τ that are 0.07 units higher on average than those for pathway position. Second, PPI identifies statistically significant correlations that pathway position does not, as in the case of the brassinosteroid pathway in the *Ricinus* lineage. Third, the significance of correlations between the PPI and dN/dS is robust to removal of high-dN/dS values whereas the significance of correlations with pathway position is often not, as in the case of gibberellic acid synthesis in *Vitis*, and brassinosteroid synthesis in *Arabidopsis*, *Populus* and *Vitis*. The PPI may be particularly helpful for regions of the metabolic network that are cyclical or reticulate, such as fatty acid synthesis (Buchanan *et al.* 2000) or the later stages of brassinosteroid and gibberellic acid synthesis (Szekeres and Koncz 1998, Hedden and Philips 2000).

Since glycolytic enzymes are known to be highly conserved, our significant correlations of dN/dS with pathway position and with PPI could be driven solely by the low dN/dS values of the glycolytic genes. For the relationship of dN/dS and pathway position, only two correlations remain significant if the glycolytic genes are removed from the analysis. However, all significant correlations of dN/dS and PPI remain significant when glycolytic genes are removed. This demonstrates both that plant terpenoid synthesis is truly a robust example of PRV, and that our Pathway Pleiotropy Index is superior to pathway position for detecting positional rate variation.

Gene evolution and metabolic pathways

Both pathway position and PPI indicate that functional constraints are increasingly relaxed along a metabolic pathway. First, there are significant correlations of both indices with dN/dS values, and dN/dS values are generally higher downstream in these pathways, though always much less than 1. Second, there is no evidence of positive selection on specific codons. These points together reflect a relaxation of selective constraint on downstream genes.

Therefore, our data suggests that the directionality and branching structure of metabolic pathways shape the evolution of genes coding for enzymes in these pathways. While the current data do not prove that PRV is driven by variation in pleiotropy, or that variation in pleiotropy results from the branching of the pathway, the data presented here certainly fit such a scenario. An alternate hypothesis is that PRV is driven by differences in connectivity rather than pleiotropy along the pathway. High connectivity also produces selective constraint (Vitkup *et al.* 2006) and many upstream enzymes used here are highly connected (KEGG Pathway Database

2008), but pleiotropy likely still plays a role apart from connectivity. A hypothetical pathway in which all nodes have four connections (one in, three out) should still show PRV even though all nodes have equal connectivity, since the upstream nodes are more pleiotropic. Nevertheless, further work is required to disentangle the roles of pathway position, pleiotropy and connectivity in shaping evolutionary rate variation.

Pathway-specific patterns

Gibberellic acid synthesis provides a particularly strong example of pathway position-dependent evolution across all four lineages, with high Kendall's τ values and statistical significance generally robust to removal of the highest dN/dS value. Synthesis of abscissic acid and brassinosteroids also show strong positional rate variation in more than one lineage. Even the lutein pathway, which did not show significant correlations, suggests a positive relationship of PPI and dN/dS. Nevertheless, these individual pathways cannot be taken as independent examples of PRV, since all share the same upstream components. Rather, they together form evidence that plant terpenoid synthesis in general is a solid example of PRV.

The lack of significant variation in dN/dS among enzymes involved with the lutein pathway may be simply a sampling issue: only one lutein-specific enzyme was available for the correlation analysis; all other enzymes along this path from glucose to lutein are involved in glycolysis, synthesis of terpenoid precursors, and early carotenoid synthesis. Although this group of enzymes ought to be of unequal pleiotropy, it may be that all these enzymes are so highly pleiotropic that there is little difference in selective constraint among them. This is particularly obvious in that lutein serves a critical function in the assembly and functioning of the light harvesting complex used in photosynthesis (Pogson *et al.* 1996). That being said, *Arabidopsis* mutants in *LUT1* and *LUT2*, which are required for lutein synthesis, have no visibly aberrant phenotype (Pogson *et al.* 1996).

Lineage-specific patterns

There is some mystery as to why *Arabidopsis* and *Vitis* do not show significant PRV as often as *Ricinus* and *Populus* do. One possibility is that since several values were discarded from *Arabidopsis* and *Vitis* due to high synonymous substitution rates, little data was left to flesh out PRV in these lineages. Perhaps these lineages are simply too long. Hahn *et al.* (2004) also found that for the correlation of protein connectivity with evolutionary rate in the yeast protein

interaction network, results varied with choice of dataset for longer lineages but not for short lineages. Future studies should include more species in phylogenetic trees with shorter branches, in order to discern whether it is common to find much variability among species in detecting PRV, or whether the better results with *Ricinus* and *Populus* are simply due to using lineages of an appropriate age.

The role of historical effective population size is also worth considering. A lineage with low effective population size may not show PRV, as inefficient selection in a small population may cause all genes along the pathway to face essentially the same degree of selective constraint (Lynch 2007). The role of historic effective population size however is difficult to ascertain; estimates of synonymous site diversity (an indicator of effective population size) are very similar for *Arabidopsis* and *Populus* (Lynch 2006) although these species currently have very different effective population sizes. Perhaps differences in lineage-specific mutation rates results in similar estimates of synonymous site variation, despite different effective population sizes.

As with individual pathways, however, the dN/dS values for separate lineages presented here come from the same likelihood analysis and therefore are not truly independent. Therefore, our results should be taken first as general evidence of PRV in plant terpenoid synthesis, and second as a suggestion of the kinds of lineage-specific effects that can be more fully understood in future studies.

Gene copy number

At the outset of this study, we hoped that examining gene copy number could explain why certain lineages show PRV and others do not. Our observation that high copy genes appear to be concentrated downstream in some cases deserves further attention in future studies. This trend runs counter to the idea that gene duplication serves to release genes from the selective constraint arising from pleiotropy, since the most pleiotropic genes seem to be retained in duplicate as much as or even less than the less pleiotropic genes at the end of the pathway. Gene copy number, however, cannot explain why PRV is observed in plant terpenoid synthesis, since many of the strong positive correlations of dN/dS and pathway position correspond to an obvious lack of copy number variation along the pathway.

Several factors could explain why gene copy number shows no apparent relationship with dN/dS here. Duplicate genes experience relaxed selection for only a short period after duplication (Lynch and Connery 2006), and even then purifying selection is not wholly relaxed

(Hughes and Hughes 1993). Also, divergence of paralogs may be asymmetric (Brunet *et al.* 2006). Since the longer of two branches for any one species was rejected in the ortholog choice, the dN/dS values presented here are highly conservative and may not capture real variation. These are necessary evils: bioinformatic identification of paralogs, without functional verification, will necessarily miss real copies of some genes while overcounting copy numbers of others.

Another issue is that gene duplication represents only one form of redundancy, "homologous redundancy." Another form, "non-homologous redundancy," is the existence of parallel paths that offer an alternate route through the network. This latter form of redundancy may actually cause greater relaxation of selective constraint than gene copy number does (Wagner 2000, Kitami and Nadeau 2002). While gene copy number cannot explain why certain lineages do not exhibit PRV, this network-based non-homologous redundancy strengthens the case for plant terpenoid synthesis as an example of PRV. Those parts of this system known to have alternate metabolic routes all lie in the upstream part of these pathways (Rodríguez-Concepción and Boronat 2002, Plaxton 1996), suggesting that the real effect of pathway position is even stronger here than calculated correlations suggest. Even greater selective constraint would be seen upstream in the absence of these alternate pathways.

Future directions

Overall, the structure of metabolic pathways deserves greater consideration as a factor shaping genome evolution. Ideally, future studies in this area will move towards whole-network analysis. Studies of gene expression networks have shown that the strength of selective constraint on component genes is shaped by network topology (Jordan *et al.* 2004), as have studies of enzyme connectivity in metabolic networks (Vitkup *et al.* 2006). Evidence for PRV in metabolic pathways presented here and elsewhere (Rausher *et al.* 1999, Lu and Rausher 2003, Sharkey *et al.* 2005, Rausher *et al.* 2008) strongly suggests that pathway branching and direction of flux may also shape the evolution of metabolic gene sequences across the whole network.

Network-wide studies of PRV will require an organism for which the total metabolic network is well-understood, plus relatives with full genome sequences. Metabolic networks have been constructed for several organisms (Jeong *et al.* 2000, Wagner and Fell 2001), but these often consider only a core metabolism, representing a set of genes that may not vary sufficiently in pleiotropy or multifunctionality to produce observable variation in selective constraint.

A further challenge for moving to the whole-network level is to find a measure of position within that network that has biological meaning and/or would be expected to correlate with dN/dS. Plant metabolic networks are ideal in this regard. Non-photosynthetic organisms can absorb carbon in many forms, and therefore lack a true upstream that applies to the whole metabolic network. In green plants, however, all carbon flow begins with photosynthesis, which can then offer a reference point for counting global pathway positions or Pathway Pleiotropy Index values. For non-plants, centrality may offer a proxy to pathway position. Centrality can be measured as the mean distance of a node to all other nodes in the network (Vitkup *et al.* 2006). The tricarboxylic acid cycle is upstream of most metabolism, and is also central to the metabolic network (Fell and Wagner 2000, Jeong *et al.* 2000), whereas reactions like the glycosylation of an anthocyanin or the final steps in brassinosteroid synthesis would have a much larger average distance to other nodes (enzymes) in the network.

As more and more genomes are sequenced, it will be possible to offer increasingly sophisticated answers to questions of how network topology shapes evolutionary rate variation. These answers may address the relative roles of connectivity and pathway position, clarify the effect of pathway position on retention of gene duplicates, or give insight to how the effects of pathway position play out at the population genomic level. We particularly hope that future studies will harness the power of our Pathway Pleiotropy Index to determine how network structure shapes genome evolution.

Acknowledgements

Many thanks to Mike Barker for his advice on ortholog determination, and for providing several perl scripts, including his alignment pipeline. Many thanks also to Nolan Kane for sharing insightful comments and unpublished work on positional rate variation. Sally Otto, Dilara Ally, Itay Mayrose and Allen Liu also took part in helpful discussions. NSERC funded this project through grants to H. Ramsay and K. Ritland.

Table 2.1. Kendall's τ and p values for the correlation of dN/dS with pathway position.

	Abscissic acid		Gibberellic acid		Brassinosteroid		Lutein	
	τ	p	τ	p	τ	p	τ	p
<i>Arabidopsis</i>	0.43	0.054	**0.63	0.012	**0.63	0.003	0.21	0.327
<i>Populus</i>	**0.41	0.016	*0.45	0.020	0.31	0.060	0.20	0.240
<i>Ricinus</i>	**0.48	0.005	0.38	0.053	0.15	0.363	0.31	0.075
<i>Vitis</i>	0.24	0.160	*0.43	0.044	0.25	0.132	0.16	0.364

* $p < 0.05$. ** $p < 0.05$ after removal of highest dN/dS value in downstream genes.

Table 2.2. Kendall's τ and p values for the correlation of dN/dS with the Pathway Pleiotropy Index.

	Abscissic acid		Gibberellic acid		Brassinosteroid		Lutein	
	τ	p	τ	p	τ	p	τ	p
<i>Arabidopsis</i>	0.38	0.114	**0.53	0.047	**0.61	0.007	0.11	0.646
<i>Populus</i>	**0.45	0.012	**0.50	0.017	*0.37	0.039	0.24	0.214
<i>Ricinus</i>	**0.55	0.002	**0.47	0.023	0.21	0.236	*0.39	0.041
<i>Vitis</i>	0.30	0.109	**0.53	0.020	0.32	0.081	0.25	0.207

* $p < 0.05$. ** $p < 0.05$ after removal of highest dN/dS value in downstream genes.

Figure 2.1. Plant terpenoid synthesis. Early enzymes are involved in producing a wider array of metabolites and directly participate in a wider array of phenotypic functions than the downstream enzymes do. Arrows represent groups of enzymes that convert one metabolite to another (only metabolites that exist at pathway branchpoints or endpoints are noted). Italics denote some functional roles of the various metabolites. Numbers indicate values of the Pathway Pleiotropy Index (see Methods).

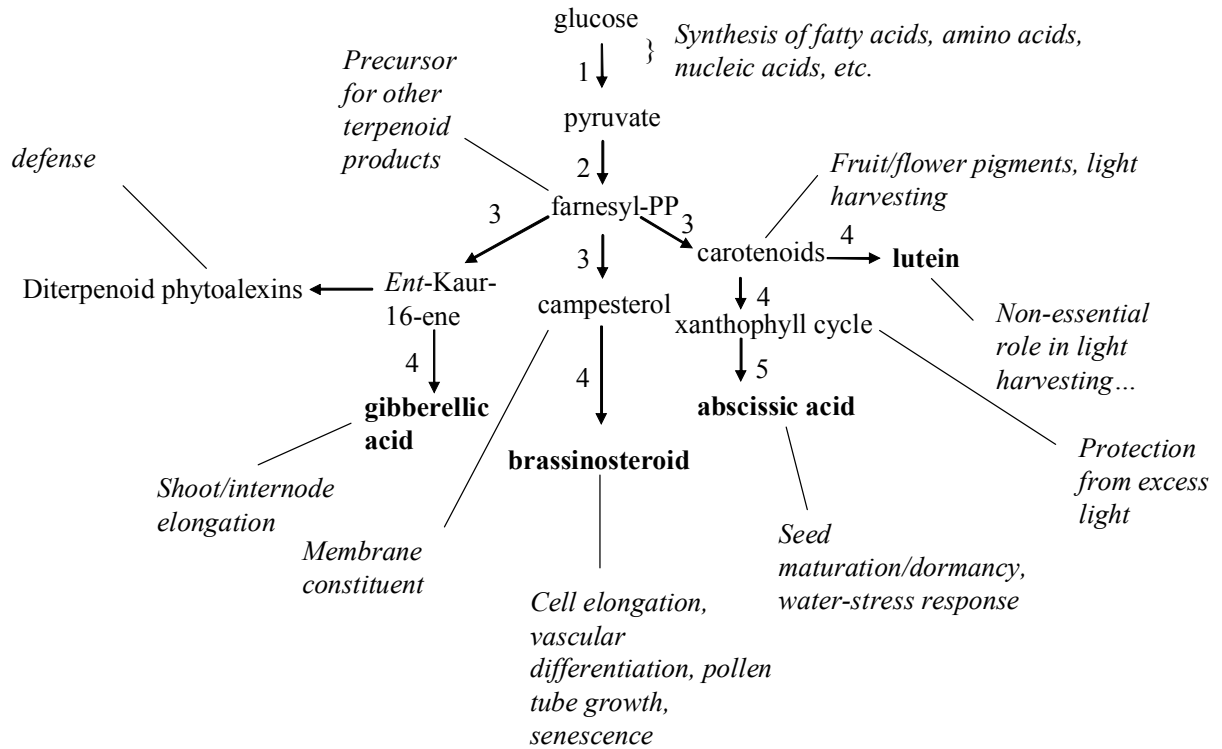


Figure 2.2. The phylogenetic relationships of the taxa used in this study. Bold branches are the lineages along which dN/dS was estimated.

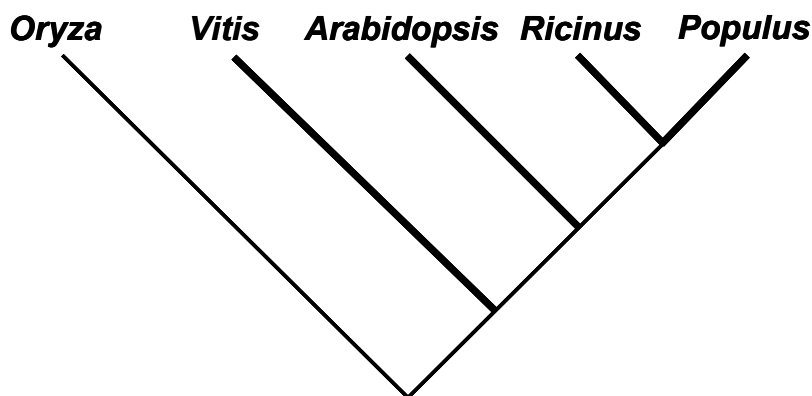
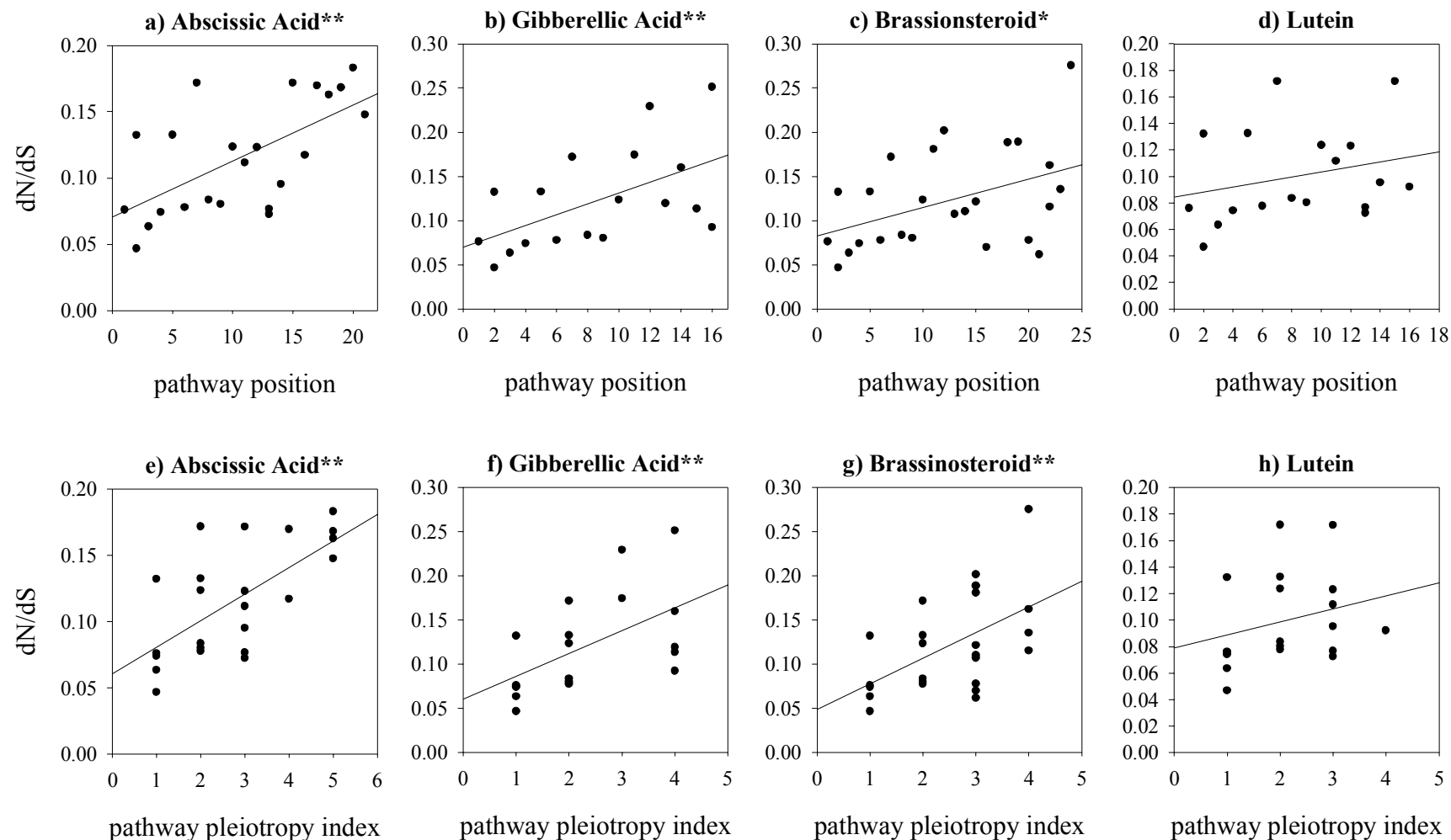


Figure 2.3. Graphs of dN/dS for *Populus*. The relationship between dN/dS and pathway position (a-d), and between dN/dS and the Pathway Pleiotropy Index (e-h), for *Populus* in four branches of terpenoid synthesis. Trend lines are given as a visual aid, and were not used to assess the strength of these relationships. **Correlation is significant and robust to removal of the highest dN/dS value. *Correlation is significant but not robust to removal of the highest dN/dS value. Kendall's τ rank correlation coefficients and p -values are given in Tables 2.1 and 2.2.



References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.
- Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M. 2006. Gene Loss and Evolutionary Rates Following Whole-Genome Duplication in Teleost Fishes. *Mol Biol Evol.* 23: 1808-1816.
- Buchanan BB, Gruissem W, Jones RL. 2000. *Biochemistry and molecular biology of plants.* Rockville (MD): American Society of Plant Physiologists.
- Cheng AX, Lou YG, Mao YB, Lu S, Wang LJ, Chen XY. 2007. Plant terpenoids: biosynthesis and ecological functions. *J Integr Plant Biol.* 49: 179-186.
- Church S, Livingstone K, Lai Z, Kozik A, Knapp S, Michelmore R, Rieseberg L. 2007. Using variable rate models to identify genes under selection in sequence pairs: their validity and limitations for EST sequences. *J Mol Evol.* 64: 171-180.
- Clouse SD, Sasse JM. 1998. Brassinosteroids: essential regulators of plant growth and development. *Annu Rev Plant Phys.* 49: 427-451.
- Cork J, Purugganan M. 2004. The evolution of molecular genetic pathways and networks. *BioEssays.* 26: 479-484.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32: 1792-1797.
- Fraser H, Wall D, Hirsh A. 2003. A simple dependence between protein evolution rate and the number of protein-protein interactions. *BMC Evol Biol.* 3: 11.
- Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. *Science.* 296: 750-752.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 11: 725-736.
- Grennan AK. 2006. Gibberellin metabolism enzymes in rice. *Plant Physiol.* 141: 524-526.
- Guindon S, Gascuel O. 2003. A simple fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52: 696-704.
- Hahn MW, Conant GC, Wagner A. 2004. Molecular evolution in large genetic networks: does connectivity equal constraint? *J Mol Evol.* 58: 203-211.
- Harbourne J. *The flavonoids.* 1988. London: Chapman and Hall.

- Hedden P, Phillips AL. 2000. Gibberellin metabolism: new insights revealed by the genes. *Trends Plant Sci.* 5: 523-530.
- Heldt HW, Heldt F. 2005. *Plant biochemistry*. Boston: Elsevier Academic Press.
- Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution*. 61: 995-1016.
- Hughes MK, Hughes AL. 1993. Evolution of duplicate genes in a tetraploid animal *Xenopus laevis*. *Mol Biol Evol.* 10: 1360-1369.
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. 2000. The large-scale organization of metabolic networks. *Nature*. 407: 651-654.
- Jordan IK, Marino-Ramirez L, Wolf YI, Koonin EV. 2004. Conservation and coevolution in the scale-free human gene coexpression network. *Mol Biol Evol.* 21: 2058-2070.
- Jordan IK, Wolf Y, Koonin E. 2003. Correction: No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. *BMC Evol Biol.* 3: 5.
- Jordan IK, Wolf Y, Koonin E. 2003. No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. *BMC Evol Biol.* 3: 1.
- KEGG PATHWAY database [Internet]. 2008. Kyoto: Kyoto University Bioinformatics Centre (Japan); [last updated 2008 Oct 1; cited 2008 Oct 23]. Available from: <http://www.genome.jp/kegg/pathway.html>.
- Kitami T, Nadeau JH. 2002. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nat Genet.* 32: 191-194.
- Liu Y, Wang H, Ye, HC, Li GF. 2005. Advances in the plant isoprenoid biosynthesis pathway and its metabolic engineering. *J Integr Plant Biol.* 47: 769-782.
- Lu YQ, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. *Mol Biol Evol.* 20: 1844-1853.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science*. 290: 1151-1155.
- Lynch M. 2006. The origins of eukaryotic gene structure. *Mol Biol Evol.* 23: 450-468.
- Lynch M. 2007. *The origins of genome architecture*. Sunderland, MA: Sinauer Associates.
- Makino T, Suzuki Y, Gojobori T. 2006. Differential evolutionary rates of duplicated genes in protein interaction network. *Gene*. 385: 57-63.

- Naik PS, Chanemougasoundharam A, Khurana SMP, Kalloo G. 2003. Genetic manipulation of carotenoid pathway in higher plants. *Curr Sci India*. 85: 1423-1430.
- Nambara E, Marion-Poll A. 2005. Absciscic acid biosynthesis and catabolism. *Annu Rev Plant Biol*. 56: 165-185.
- Otto SP. Two steps forward one step back: the pleiotropic effects of favoured alleles. *P Roy Soc B-Biol Sci*. 271: 705-714.
- Plaxton WC. 1996. The organization and regulation of plant glycolysis. *Annu Rev Plant Phys*. 47: 185-214.
- Pogson B, McDonald KA, Truong M, Britton G, DellaPenna D. 1996. *Arabidopsis* carotenoid mutants demonstrate that lutein is not essential for photosynthesis in higher plants. *Plant Cell*. 8: 1627-1639.
- Poptsova M, Gogarten JP. 2007. BranchClust: a phylogenetic algorithm for selecting gene families. *BMC Bioinformatics*. 8: 120.
- Ramos-Onsins SE, Puerma E, Balañá-Alcaide, Salguero D, Aguadé M. 2008. Multilocus analysis of variation using a large empirical data set: phenylpropanoid pathway genes in *Arabidopsis thaliana*. *Mol Ecol*. 17: 1211-1223.
- Rausher M.D, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol*. 16: 266-274.
- Rausher MD, Lu YQ, Meyer K. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *J Mol Evol*. 67: 137-144.
- Rodríguez-Concepción M, Boronat A. 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol*. 130: 1079-1089.
- Sakamoto T, Miura K, Itoh H, *et al.* (9 co-authors). 2004. An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiol*. 134: 1642-1653.
- Sharkey TD, Yeh S, Wiberley AE, Falbel TG, Gong DM, Fernandez DE. 2005. Evolution of the isoprene biosynthetic pathway in Kudzu. *Plant Physiol*. 137: 700-712.
- Soltis P, Soltis D, Edwards C. 2005. Angiosperms. Flowering Plants [Internet]. The Tree of Life Web Project; [last updated 2005 June 3, cited 2008 Oct 24]. Available from: <http://tolweb.org/Angiosperms/20646/2005.06.03>.
- Szekeres M, Koncz C. 1998. Biochemical and genetic analysis of brassinosteroid metabolism and function in *Arabidopsis*. *Plant Physiol Bioch*. 36: 145-155.

- Vitkup D, Kharchenko P, Wagner A. 2006. Influence of metabolic network structure and function on enzyme evolution. *Genome Biology*. 7: R39.
- Wagner A, Fell DA. 2001. The small world inside large metabolic networks. *P Roy Soc B-Biol Sci*. 268: 1803-1810.
- Waxman D, Peck JR. 1998. Pleiotropy and the preservation of perfection. *Science*. 279: 1210-1213.
- Winkel BSJ. 2006. The biosynthesis of flavonoids. In: Grotewold E, editor. *The science of flavonoids*. Columbus (OH): Springer. p. 71-96.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13: 555-556.
- Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 15: 568-573.

CHAPTER 3: Conclusions

Summary and evaluation of Ramsay *et al.*

The last decade has seen growing interest in the role of gene interaction networks in determining rates of gene/protein evolution. An early hypothesis was that evolutionary rates would correlate with the metabolic pathway position of enzymes, due to a relaxation of selective constraint along metabolic pathways (Rausher *et al.* 1999). This effect, called positional rate variation, was until recently proven only by studies of three to six enzymes from the anthocyanin pathway in plants (Rausher *et al.* 1999, Lu and Rausher 2003, Rausher *et al.* 2008). Ramsay *et al.*, however, have shown that statistically significant positional rate variation also exists in plant terpenoid synthesis. This study, presented as Chapter 2 of this thesis, examined dN/dS values for 45 genes along four branches of terpenoid synthesis, based on phylogenetic orthologs of four ingroup and one outgroup species. Ramsay *et al.* found that dN/dS was correlated with pathway position, and also with a new metric called the Pathway Pleiotropy Index, which they showed to be superior to pathway position for capturing effects of enzyme order on evolutionary rates. Tests for codon-specific selection did not detect any positive selection, supporting the idea that these trends are in fact driven by relaxation of selective constraint downstream. Further, gene copy number did not drive results.

The research by Ramsay *et al.* is both creative and methodologically sound. First, the Pathway Pleiotropy Index is a novel suggestion for understanding evolutionary rate variation in pathways, based on an understanding of the theoretical relationship between pathway structure and pleiotropy, predicted from first principles. Second, Ramsay *et al.* practiced sound methodology by their use of phylogenetic ortholog determination, by explicitly testing for significant variation in dN/dS, and by testing the robustness of correlations to the removal of extreme values. Also, the use of four pathway branches, numerous genes, and five species allows Ramsay *et al.* to make generalizable statements about positional rate variation. These strengths all distinguish Ramsay *et al.* as one of only three independent studies of positional rate variation (others being Rausher *et al.* 1999, Lu and Rausher 2003, Ramos-Onsins *et al.* 2008, Rausher *et al.* 2008).

Ramsay *et al.* is not without faults. The data set was not large enough to justify a full statistical analysis of the role of gene copy number on evolutionary rates. Also, two of the species used, *Vitis vinifera* and *Arabidopsis thaliana*, lie on very long branches relative the other

ingroup species. Along these long branches, several data points had to be discarded because synonymous substitutions had reached saturation.

While Ramsay *et al.* examine a far larger number of genes than previous studies, not every relevant enzyme was considered. All heteromeric genes involved in glycolysis were excluded, and of the two parallel paths in terpenoid synthesis (MEP and MVA; see Buchanan *et al.* 2000) only the MEP pathway was adequately represented. Since metabolites can be exchanged between these plastidial versus cytosolic compartments (Cheng *et al.* 2007), I assume that the lack of MVA pathway genes is not a serious problem.

One final weakness is that substitution rates are only one aspect of evolutionary lability. This weakness should not be overstated, since studies of molecular evolution commonly focus on substitution rates. With the advent of new methods that do provide estimates of indel rates (e.g., Lunter 2007), future studies may incorporate this additional aspect of molecular evolution. I predict that downstream genes will have higher rates of indel incorporation, as well as higher dN/dS.

Comparison of correlations among studies.

Ramsay *et al.* did not assess how strong the effects on pathway position are relative other known effects on evolutionary rates of proteins. Table 3.1 provides correlation coefficients for evolutionary rate with various variables. All studies except Ramsay *et al.* and Ingvarsson 2007 were performed in yeast. Direct comparison of Kendall's τ values from Ramsay *et al.* to correlation coefficients from other studies are challenging both because it is difficult to directly compare different measures of correlation, and because the other studies are based on fuller networks and larger numbers of genes. However, a few comments can be made: first, all correlations except for those of dN/dS with expression level are less than 0.5, suggesting that all three network properties (Pathway Pleiotropy Index, connectivity, and flux) have real but moderate effects on gene evolution, leaving much rate variation unexplained. The strongest rank correlations for PPI do reach 0.54, but it is unlikely that such a high value would be found over the whole metabolic network. It can be noted, however, that flux does produce stronger correlations than either metabolic connectivity or number of protein-protein interactors.

The relative importance of pathway position for determining evolutionary rate awaits further study. The analyses published to date do not allow us to discern whether effects of pathway position and connectivity are truly weaker than correlations with expression level, or

whether earlier studies did not properly control for confounding variables (e.g., see Bertin *et al.* 2007 for an attempt to account for different types of highly connected nodes). Also, none of these studies has analyzed the true total network. Studies of metabolic networks, for example, often focus on a core metabolism. Thus, they may be restricted to a set of genes that is too similarly pleiotropic to indicate the true range of rate variation across the whole network.

Context and significance of Ramsay *et al.*

The work of Ramsay *et al.* joins a growing body of literature on evolutionary systems biology, a recent trend in research. Like many new fields, there is now enough data to suggest real patterns, but not enough to resolve certain controversies (Koonin 2005). Current data confirms that gene interaction networks are typically scale-free, with a few highly-connected nodes acting as hubs of the network. Node connectivity, flux and position all affect evolutionary rate in metabolic networks, as for connectivity in protein-protein interaction and gene coexpression networks (see chapter 1 for a review of this research). It is yet unresolved how strong these effects are relative one another and relative other known effects on evolutionary rate. It also remains unknown whether effects such as expression level and dispensability are independent of, or arise from, effects of network structure. Relatively weak correlations of evolutionary rate with network properties suggest that they may not be the primary effect on evolutionary rate variation (if one primary effect exists), but that they certainly play a role in explaining why genes evolve at different rates.

Within the published literature, the role of position within the network is little explored. While a few studies examine centrality, not many address the position of enzymes along pathways through the metabolic network. If published, the work of Ramsay *et al.* will be the first study to empirically establish positional rate variation as a plausible reality for the metabolic network as a whole. (At the time of writing, a manuscript essentially identical to Chapter 2 is in submission to *Molecular Biology and Evolution*).

Future directions

Evolutionary systems biology has been born of the genomic era (Medina 2005) and will continue to grow as more genomic and functional data become available. I propose four important directions for future research. First, we need computational studies of the theoretical effects of pathway position, node connectivity and clustering coefficient (a measure of

modularity in a network) on evolutionary rates. I know of no such studies so far, but they would be useful for generating null distributions of results to comparison with empirical studies. Second, future research must include empirical whole-network studies comparing measures of centrality to measures of pathway position (including PPI). This will help to determine whether these are interchangeable concepts, and whether pathway position effects play out over the whole network. In order to carry out such studies, there is a need to think about what constitutes upstream and whether there are true endproducts at the network level, since "end products" can be degraded to metabolites used in a broad range of biosynthetic reactions. At present, *Drosophila*, bacteria and yeast have the best resources for such studies, with whole genome sequences for many closely related species (Medina 2005), plus well-understood metabolic networks. However, studies of network effects on evolutionary rate variation have relied almost exclusively on yeast; to make general statements on molecular evolution, research must necessarily move beyond *Saccharomyces*.

Third, future research must include population-level studies, alongside interspecies comparisons. In the coming age of population genomics, it will be possible to empirically examine the interplay of demographic and network effects on gene evolution. Finally, future studies must integrate metabolic flux information with measures of connectivity and position to explain rate variation. Topology is but one aspect of network behavior, and flux through the network is also important for shaping selective pressures (Vitkup *et al.* 2006, Flowers *et al.* 2007). By simultaneously examining several network properties, future research can begin to disentangle the roles of each property in shaping gene evolution. In closing, Ramsay *et al.* have put pathway position forward as one of many network properties worthy of future study. It remains for future researchers to understand the true links of genome, network and phenotype in the course of evolution.

Table 3.1. Correlation values for measures of evolutionary rate with measures of connectivity, flux and gene expression.

Variables	Correlation	Measure	Source
PPI vs. dN/dS	0.24-0.54	Kendall's τ	Ramsay <i>et al.</i>
metabolic connectivity vs. Ka/Ks	-0.2	Spearman rank correlation r	Vitkup <i>et al.</i> 2006
metabolic connectivity vs. Ka/Ks	-0.18	Pearson's correlation r	Vitkup <i>et al.</i> 2006
metabolic flux vs. Ka/Ks	-0.3	Spearman rank correlation r	Vitkup <i>et al.</i> 2006
metabolic flux vs. Ka/Ks	-0.24	Pearson's correlation r	Vitkup <i>et al.</i> 2006
protein-protein interactors vs. protein substitution rate	-0.24	Spearman rank correlation r	Fraser <i>et al.</i> 2003 Drummond <i>et al.</i> 2005
expression level vs. dN	0.28	r^2	
expression breadth* vs. dN/dS	-0.877	Spearman rank correlation r	Ingvarsson 2007

*Number of tissues in which a gene is expressed

References

- Bertin N, Simonis N, Dupuy D, Cusick ME, Han JDJ, Fraser HB, Roth FP, Vidal M. 2007. Confirmation of organized modularity in the yeast interactome. *P Roy Soc B-Biol Sci.* 5: e153.
- Buchanan BB, Gruissem W, Jones RL. 2000. *Biochemistry and molecular biology of plants.* Rockville (MD): American Society of Plant Physiologists.
- Cheng AX, Lou YG, Mao YB, Lu S, Wang LJ, Chen XY. 2007. Plant terpenoids: biosynthesis and ecological functions. *J Integr Plant Biol.* 49: 179-186.
- Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. *P Natl Acad Sci USA.* 102: 14338-14343.
- Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, Eanes WF. 2007. Adaptive evolution of metabolic pathways in *Drosophila*. *Mol Biol Evol.* 24: 1347-1354.
- Ingvarsson PK. 2007. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Mol Biol Evol.* 24:836-844.
- Koonin EV. 2005. Systemic determinants of gene evolution and function. *Mol Syst Biol* 1: 2005.0021.
- Lu YQ, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. *Mol Biol Evol.* 20: 1844-1853.
- Lunter G. 2007. Probabilistic whole-genome alignments reveal high indel rates in the human and mouse genomes. *Bioinformatics* 23: i289-296.
- Medina M. 2005. Genomes, phylogeny, and evolutionary systems biology. *P Natl Acad Sci USA.* 102: 6630-6635.
- Ramos-Onsins SE, Puerma E, Balaña-Alcaide, Salguero D, Aguadé M. 2008. Multilocus analysis of variation using a large empirical data set: phenylpropanoid pathway genes in *Arabidopsis thaliana*. *Mol Ecol.* 17: 1211-1223.
- Rausher M.D, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol.* 16: 266-274.
- Rausher MD, Lu YQ, Meyer K. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *J Mol Evol.* 67: 137-144.
- Vitkup D, Kharchenko P, Wagner A. 2006. Influence of metabolic network structure and function on enzyme evolution. *Genome Biology.* 7: R39.

Appendix 1. *Arabidopsis* loci used as queries in ortholog determination
(see chapter 2)

Name or EC number	Query locus	Name or EC number	Query locus
1.1.1.267	AT5G62790	5.3.3.2	AT3G02780
1.14.11.12	AT4G25420	5.3.3.2	AT3G02780
1.14.11.12	AT5G07200	5.4.99.8	AT1G78960
1.14.11.12	AT5G51810	5.4.99.8	AT2G07050
1.14.11.13	AT1G78440	5.4.99.8	AT3G45130
1.14.11.13	AT2G34555	5.4.99.8	AT4G15340
1.14.11.15	AT1G15550	5.4.99.8	AT5G36150
1.14.13.78	AT5G25900	5.4.99.8	AT5G42600
1.14.13.79	AT2G32440	5.5.1.9	AT5G50375
1.14.13.93	AT2G29090	AAO3	AT2G27150
1.14.13.93	AT3G19270	ABA2	AT1G52340
1.14.13.93	AT4G19230	Br6ox	AT3G30180
1.14.13.93	AT5G45340	Br6ox	AT5G38970
1.14.99.7	AT1G58440	Cpd	AT5G05690
1.14.99.7	AT1G58440	CrtB	AT5G17230
1.14.99.7	AT5G24140	CrtISO	AT1G06820
1.14.99.7	AT5G24140	CrtL-b	AT3G10230
1.14.99.7	AT5G24160	CrtL-e	AT5G57030
1.14.99.7	AT5G24160	CrtR-b	AT4G25700
1.17.1.2	AT4G34350	CrtR-b	AT5G52570
2.2.1.7	AT3G21500	CYP51G1	AT1G11680
2.2.1.7	AT4G15560	Det2	AT2G38050
2.2.1.7	AT5G11380	DWF1	AT3G19820
2.5.1.10	AT2G18620	Dwf4	AT3G50660
2.5.1.10	AT2G18620	DWF5	AT1G50430
2.5.1.10	AT3G14530	FK	AT3G52940
2.5.1.10	AT3G14530	HYD1	AT1G20050
2.5.1.10	AT3G14550	LUT1	AT3G53130
2.5.1.10	AT3G14550	NCED	AT1G30100
2.5.1.10	AT3G20160	NCED	AT1G78390
2.5.1.10	AT3G20160	NCED	AT3G14440
2.5.1.10	AT3G29430	NCED	AT3G24220
2.5.1.10	AT3G29430	NCED	AT4G18350
2.5.1.10	AT3G32040	PDS	AT4G14210
2.5.1.10	AT3G32040	SMT1	AT5G13710
2.5.1.10	AT5G47770	STE1	AT3G02580
2.5.1.21	AT4G34650	Vde	AT1G08550
2.5.1.21	AT4G34650	ZDS	AT3G04870
2.7.1.48	AT2G26930	Zep	AT5G67030
4.2.3.19	AT1G79460		

Appendix 2. Orthologs used to estimate dN/dS

Appendix 2.1. *Arabidopsis* orthologs

Pathway Segment	Gene Name or EC	
	Number	Locus Identity
glucose – pyuvate	4.1.2.13	AT2G01140.1
glucose – pyuvate	2.7.2.3	AT3G12780.1
glucose – pyuvate	5.3.1.9p	AT4G24620.1
glucose – pyuvate	5.3.1.9c	AT5G42740.1
glucose – pyuvate	5.4.2.2	AT5G51820.1
pyuvate – fanesyl-PP	2.5.1.10	AT2G18620.1
pyuvate – fanesyl-PP	2.7.1.148	AT2G26930.1
pyuvate – fanesyl-PP	5.3.3.2	AT3G02780.1
pyuvate – fanesyl-PP	2.2.1.7	AT4G15560.1
pyuvate – fanesyl-PP	1.17.1.2	AT4G34350.1
pyuvate – fanesyl-PP	1.1.1.267	AT5G62790.1
fanestyl-PP - campesteol	CYP51G1	AT1G11680.1
fanestyl-PP - campesteol	HYD1	AT1G20050.1
fanestyl-PP - campesteol	DWF5	AT1G50430.1
fanestyl-PP - campesteol	1.14.99.7	AT1G58440.1
fanestyl-PP - campesteol	5.4.99.8	AT2G07050.1
fanestyl-PP - campesteol	STE1	AT3G02580.1
fanestyl-PP - campesteol	DWF1	AT3G19820.1
fanestyl-PP - campesteol	2.5.1.21	AT4G34640.1
fanestyl-PP - campesteol	SMT1	AT5G13710.1
fanestyl-PP - campesteol	5.5.1.9	AT5G50375.1
fanestyl-PP - caotenoids	CtISO	AT1G06820.1
fanestyl-PP - caotenoids	ZDS	AT3G04870.1
fanestyl-PP - caotenoids	CtL-b	AT3G10230.1
fanestyl-PP - caotenoids	PDS	AT4G14210.2
fanestyl-PP - caotenoids	CtB	AT5G17230.1
fanestyl-PP - caotenoids	Ct-b	AT5G52570.1
fanestyl-PP - ent-kau-16-ene	4.2.3.19	AT1G79460.1
fanestyl-PP - ent-kau-16-ene	5.5.1.13	AT4G02780.1
campesteol - steoid	Det2	AT2G38050.1
campesteol - steoid	Dwf4	AT3G50660.1
campesteol - steoid	Cpd	AT5G05690.1
campesteol - steoid	B6ox	AT5G38970.1
caotenoids - xanthophyll cycle	Vde	AT1G08550.2
caotenoids - xanthophyll cycle	Zep	AT5G67030.1
caotenoids - lutein	LUT1	AT3G53130.1
ent-kau-16-ene - gibbeellic acid	1.14.11.15	AT1G15550.1
ent-kau-16-ene - gibbeellic acid	1.14.13.79	AT2G32440.1
ent-kau-16-ene - gibbeellic acid	1.14.11.13	AT2G34555.1
ent-kau-16-ene - gibbeellic acid	1.14.11.12	AT5G07200.1
ent-kau-16-ene - gibbeellic acid	1.14.13.78	AT5G25900.1
xanthophyll cycle - abscissic	ABA2	AT1G52340.1

acid		
xanthophyll cycle - abscissic acid	AAO3	AT2G27150.1
xanthophyll cycle - abscissic acid	NCED	AT3G24220.1
xanthophyll cycle - abscissic acid	1.14.13.93	AT5G45340.1

Appendix 2.2. *Oryza* orthologs

Pathway Segment	Gene Name or EC	
	Number	Locus Identity
glucose – pyuvate	4.1.2.13	Os01g02880.1 12001.m06927
glucose – pyuvate	2.7.2.3	Os05g41640.2 12005.m08338
glucose – pyuvate	5.3.1.9c	Os06g14510.3 12006.m091669
glucose – pyuvate	5.3.1.9p	Os09g29070.1 12009.m06058
glucose – pyuvate	5.4.2.2	Os10g11140.1 12010.m04443
pyuvate - fanesyl-PP	1.1.1.267	Os01g01710.1 12001.m06815
pyuvate - fanesyl-PP	2.7.1.148	Os01g58790.1 12001.m12021
pyuvate - fanesyl-PP	1.17.1.2	Os03g52170.1 12003.m10194
pyuvate - fanesyl-PP	2.2.1.7	Os05g33840.1 12005.m07616
pyuvate - fanesyl-PP	5.3.3.2	Os07g36190.1 12007.m07872
pyuvate - fanesyl-PP	2.5.1.10	Os07g39270.1 12007.m08169
fanesyl-PP - campesteol	HYD1	Os01g01369.1 12001.m150882
fanesyl-PP - campesteol	STE1	Os01g04260.1 12001.m07061
fanesyl-PP - campesteol	5.4.99.8	Os02g04710.2 12002.m33547
fanesyl-PP - campesteol	DWF5	Os02g26650.1 12002.m07808
fanesyl-PP - campesteol	1.14.99.7	Os03g12910.1 12003.m06738
fanesyl-PP - campesteol	2.5.1.21	Os03g59040.1 12003.m10805
fanesyl-PP - campesteol	SMT1	Os07g10600.2 12007.m28956
fanesyl-PP - campesteol	DWF1	Os10g25780.1 12010.m065197
fanesyl-PP - campesteol	5.5.1.9	Os11g19700.1 12011.m06045
fanesyl-PP - campesteol	CYP51G1	Os11g32240.1 12011.m07126
fanesyl-PP - caotenoids	CtL-b	Os02g09750.1 12002.m06272
fanesyl-PP - caotenoids	Ct-b	Os03g03370.1 12003.m05872
fanesyl-PP - caotenoids	PDS	Os03g08570.1 12003.m06365
fanesyl-PP - caotenoids	CtB	Os06g51290.3 12006.m31942
fanesyl-PP - caotenoids	ZDS	Os07g10490.1 12007.m05504
fanesyl-PP - caotenoids	CtISO	Os11g36440.1 12011.m07496
fanesyl-PP - ent-kau-16-ene	5.5.1.13	Os02g17780.1 12002.m07022
fanesyl-PP - ent-kau-16-ene	4.2.3.19	Os04g52210.1 12004.m10116
campesteol - steoid	Det2	Os01g63260.1 12001.m12442
campesteol - steoid	Dwf4	Os03g12660.1 12003.m06715
campesteol - steoid	B6ox	Os03g40540.1 12003.m09132
campesteol - steoid	Cpd	Os12g04480.1 12012.m04442
caotenoids - xanthophyll cycle	Vde	Os04g31040.3 12004.m35285
caotenoids - xanthophyll cycle	Zep	Os04g37619.1 12004.m08750

caotenoids - lutein	LUT1	Os10g39930.2 12010.m065331
ent-kau-16-ene - gibbeellic acid	1.14.11.15	Os01g08220.1 12001.m07447
ent-kau-16-ene - gibbeellic acid	1.14.11.13	Os01g55240.1 12001.m11683
ent-kau-16-ene - gibbeellic acid	1.14.11.12	Os03g63970.1 12003.m11264
ent-kau-16-ene - gibbeellic acid	1.14.13.79	Os06g02019.1 12006.m04938
ent-kau-16-ene - gibbeellic acid	1.14.13.78	Os06g37330.1 12006.m08268
xanthophyll cycle - abscissic acid	1.14.13.93	Os02g47470.1 12002.m33831
xanthophyll cycle - abscissic acid	NCED	Os03g44380.1 12003.m09477
xanthophyll cycle - abscissic acid	AAO3	Os03g57680.1 12003.m10682
xanthophyll cycle - abscissic acid	ABA2	Os03g59610.1 12003.m10858

Appendix 2.3. *Populus* orthologs

Pathway Segment	Gene Name or EC Number	Locus Identity
glucose – pyuvate	5.3.1.9p	411695 gw1.II.3030.1
glucose – pyuvate	5.3.1.9c	421146 gw1.VIII.2574.1
glucose – pyuvate	5.4.2.2	824892 estExt_fgenes4_pg.C_LG_XV1054
glucose – pyuvate	4.1.2.13	832739 estExt_fgenes4_pm.C_LG_VIII0532
glucose – pyuvate	2.7.2.3	833719 estExt_fgenes4_pm.C_LG_X0655
pyuvate - fanesyl-PP	2.7.1.148	181114 gw1.I.9714.1
pyuvate - fanesyl-PP	1.17.1.2	557296 eugene3.00090556
pyuvate - fanesyl-PP	1.1.1.267	728585 estExt_Genewise1_v1.C_LG_XII0355
pyuvate - fanesyl-PP	2.5.1.10	741258 estExt_Genewise1_v1.C_570227
pyuvate - fanesyl-PP	2.2.1.7	766757 fgenes4_pg.C_LG_VIII001771
pyuvate - fanesyl-PP	5.3.3.2	809590 fgenes4_pm.C_LG_XIX000155
fanesyl-PP – campesteol	SMT1	202903 gw1.IX.3368.1
fanesyl-PP – campesteol	5.5.1.9	252617 gw1.XV.2053.1
fanesyl-PP – campesteol	HYD1	554665 eugene3.00031298
fanesyl-PP – campesteol	CYP51G1	554675 eugene3.00031308
fanesyl-PP – campesteol	DWF1	564011 eugene3.00080606
fanesyl-PP – campesteol	5.4.99.8	717351 estExt_Genewise1_v1.C_LG_VI1514
fanesyl-PP – campesteol	2.5.1.21	818128 estExt_fgenes4_pg.C_LG_IV1345
fanesyl-PP – campesteol	STE1	831136 estExt_fgenes4_pm.C_LG_IV0218
fanesyl-PP – campesteol	1.14.99.7	831444 estExt_fgenes4_pm.C_LG_V0246
fanesyl-PP – campesteol	DWF5	833904 estExt_fgenes4_pm.C_LG_X0950
fanesyl-PP – caotenoids	PDS	246440 gw1.XIV.3183.1
fanesyl-PP – caotenoids	CtISO	256155 gw1.XVI.2094.1
fanesyl-PP – caotenoids	CtB	289129 gw1.44.342.1
fanesyl-PP – caotenoids	CtL-b	652630 grail3.0024027901
fanesyl-PP – caotenoids	Ct-b	826998 estExt_fgenes4_pg.C_440224
fanesyl-PP – caotenoids	ZDS	836745 estExt_fgenes4_pm.C_700037
fanesyl-PP - ent-kau-16-ene	4.2.3.19	564162 eugene3.00080757

fanesyl-PP - ent-kau-16-ene	5.5.1.13	761211 fgenes4_pg.C_LG_V001263
campesteol – steoid	Det2	576916 eugene3.00161051
campesteol – steoid	B6ox	745693 estExt_Genewise1_v1.C.1 550060
campesteol – steoid	Dwf4	820024 estExt_fgenes4_pg.C_LG_VII1057
campesteol – steoid	Cpd	820409 estExt_fgenes4_pg.C_LG_VIII0571
caotenoids - xanthophyll cycle	Vde	729935 estExt_Genewise1_v1.C_LG_XIII1884
caotenoids - xanthophyll cycle	Zep	760467 fgenes4_pg.C_LG_V000519
caotenoids – lutein	LUT1	589522 eugene3.00280258
ent-kau-16-ene - gibbeellic acid	1.14.13.78	412545 gw1.II.3 880.1
ent-kau-16-ene - gibbeellic acid	1.14.11.15	756811 fgenes4_pg.C_LG_III000353
ent-kau-16-ene - gibbeellic acid	1.14.11.13	758699 fgenes4_pg.C_LG_IV000327
ent-kau-16-ene - gibbeellic acid	1.14.11.12	824891 estExt_fgenes4_pg.C_LG_XV1053
ent-kau-16-ene - gibbeellic acid	1.14.13.79	835007 estExt_fgenes4_pm.C_LG_XIV0527
xanthophyll cycle - abscissic acid	ABA2	547844 eugene3.00010285
xanthophyll cycle - abscissic acid	NCED	757941 fgenes4_pg.C_LG_III001483
xanthophyll cycle - abscissic acid	AAO3	767106 fgenes4_pg.C_LG_IX000157
xanthophyll cycle - abscissic acid	1.14.13.93	783291 fgenes4_pg.C_scaffold_86000151

Appendix 2.4. *Ricinus* orthologs

Pathway Segment	Gene Name or EC	
	Number	Locus Identity
glucose – pyuvate	4.1.2.13	29660.m000779
glucose – pyuvate	5.3.1.9p	29908.m006236
glucose – pyuvate	5.4.2.2	30147.m014138
glucose – pyuvate	2.7.2.3	30169.m006297
glucose – pyuvate	5.3.1.9c	30170.m014025
pyuvate - fanesyl-PP	2.5.1.10	29646.m001058
pyuvate - fanesyl-PP	2.2.1.7	29726.m003963
pyuvate - fanesyl-PP	2.7.1.148	29785.m000940
pyuvate - fanesyl-PP	1.17.1.2	29851.m002455
pyuvate - fanesyl-PP	5.3.3.2	29929.m004583
pyuvate - fanesyl-PP	1.1.1.267	30147.m013949
fanesyl-PP – campesteol	5.4.99.8	29667.m000347
fanesyl-PP – campesteol	SMT1	29864.m001470
fanesyl-PP – campesteol	DWF1	30100.m000783
fanesyl-PP – campesteol	CYP51G1	30128.m008568
fanesyl-PP – campesteol	HYD1	30128.m008583
fanesyl-PP – campesteol	2.5.1.21	30131.m007117
fanesyl-PP – campesteol	5.5.1.9	30147.m014529
fanesyl-PP – campesteol	DWF5	30169.m006262
fanesyl-PP – campesteol	1.14.99.7	30170.m013663
fanesyl-PP – campesteol	STE1	30205.m001612

fanesyl-PP – caotenoids	CtB	28611.m000105
fanesyl-PP – caotenoids	CtISO	29333.m001086
fanesyl-PP – caotenoids	CtL-b	29941.m000224
fanesyl-PP – caotenoids	PDS	30065.m001161
fanesyl-PP – caotenoids	ZDS	30128.m008623
fanesyl-PP – caotenoids	Ct-b	30169.m006624
fanesyl-PP - ent-kau-16-ene	4.2.3.19	28670.m000104
fanesyl-PP - ent-kau-16-ene	5.5.1.13	29647.m002075
campesteol – steoid	Det2	27798.m000584
campesteol – steoid	Cpd	29188.m000051
campesteol – steoid	Dwf4	29634.m002158
campesteol – steoid	B6ox	29790.m000806
caotenoids - xanthophyll cycle	Vde	29950.m001179
caotenoids - xanthophyll cycle	Zep	30115.m001209
caotenoids – lutein	LUT1	30078.m002224
ent-kau-16-ene - gibbeellic acid	1.14.11.13	29693.m002012
ent-kau-16-ene - gibbeellic acid	1.14.13.79	29709.m001228
ent-kau-16-ene - gibbeellic acid	1.14.11.15	29841.m002759
ent-kau-16-ene - gibbeellic acid	1.14.11.12	30147.m014140
ent-kau-16-ene - gibbeellic acid	1.14.13.78	30170.m013942
xanthophyll cycle - abscissic acid	ABA2	29588.m000888
xanthophyll cycle - abscissic acid	1.14.13.93	29801.m003223
xanthophyll cycle - abscissic acid	AAO3	29854.m001100
xanthophyll cycle - abscissic acid	NCED	29904.m002987

Appendix 2.5. *Vitis* orthologs

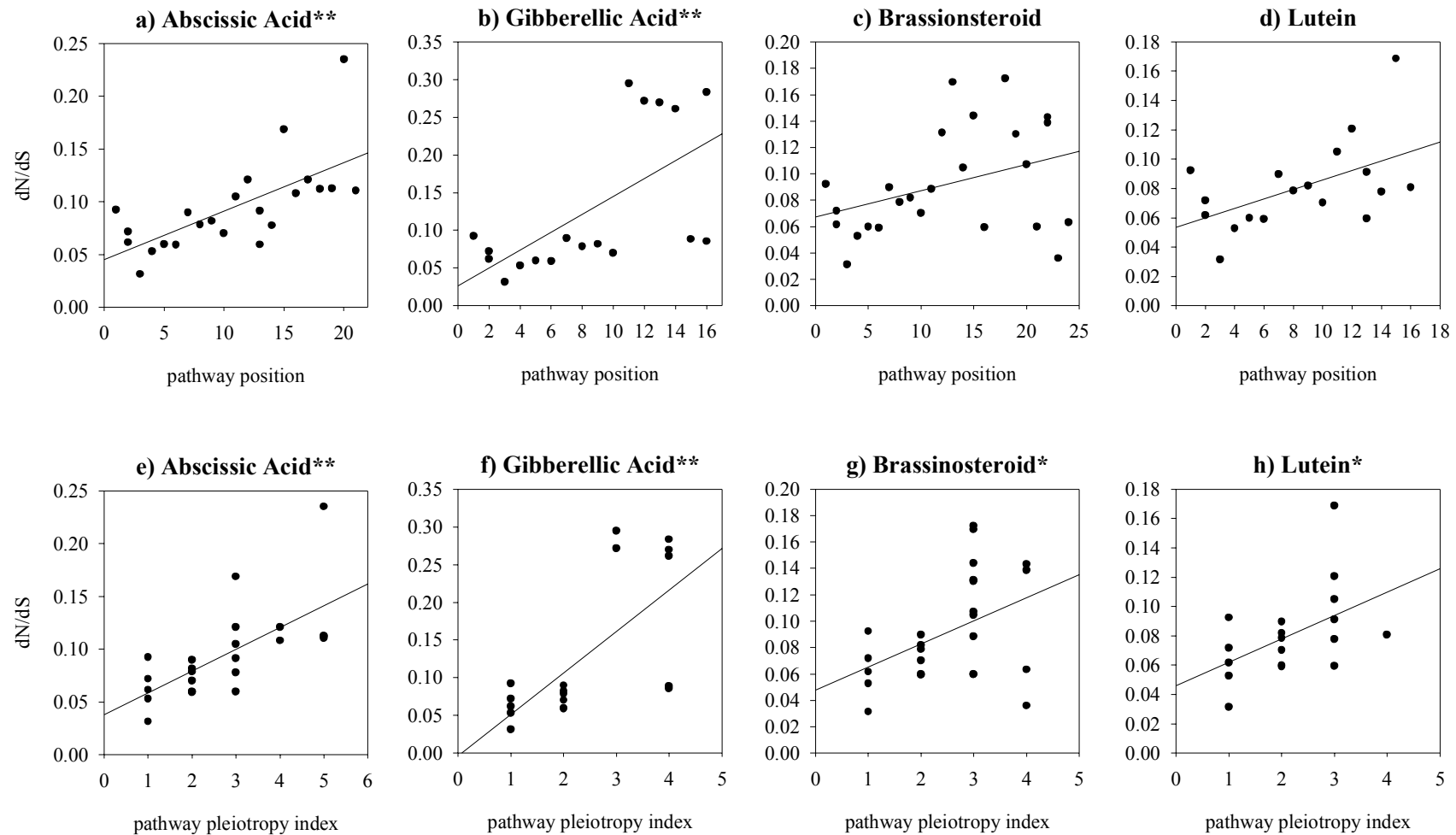
Pathway Segment	Gene Name or EC	
	Number	Locus Identity
glucose – pyuvate	5.4.2.2	GSVIVP00010600001
glucose – pyuvate	5.3.1.9c	GSVIVP00014624001
glucose – pyuvate	5.3.1.9p	GSVIVP00015125001
glucose – pyuvate	4.1.2.13	GSVIVP00030426001
glucose – pyuvate	2.7.2.3	GSVIVP00035675001
pyuvate - fanesyl-PP	1.1.1.267	GSVIVP00016085001
pyuvate - fanesyl-PP	2.7.1.148	GSVIVP00017298001
pyuvate - fanesyl-PP	2.2.1.7	GSVIVP00019845001
pyuvate - fanesyl-PP	2.5.1.10	GSVIVP00025000001
pyuvate - fanesyl-PP	5.3.3.2	GSVIVP00025071001
pyuvate - fanesyl-PP	1.17.1.2	GSVIVP00036436001
fanesyl-PP - campesteol	CYP51G1	GSVIVP00001689001
fanesyl-PP - campesteol	1.14.99.7	GSVIVP00015002001
fanesyl-PP - campesteol	5.4.99.8	GSVIVP00016985001

fanesyl-PP - campesteol	5.5.1.9	GSVIVP00017916001
fanesyl-PP - campesteol	DWF1	GSVIVP00021217001
fanesyl-PP - campesteol	HYD1	GSVIVP00024612001
fanesyl-PP - campesteol	SMT1	GSVIVP00028748001
fanesyl-PP - campesteol	DWF5	GSVIVP00029627001
fanesyl-PP - campesteol	2.5.1.21	GSVIVP00035946001
fanesyl-PP - campesteol	STE1	GSVIVP00038254001
fanesyl-PP - caotenoids	CtB	GSVIVP00023494001
fanesyl-PP - caotenoids	CtL-b	GSVIVP00025310001
fanesyl-PP - caotenoids	CtISO	GSVIVP00027139001
fanesyl-PP - caotenoids	ZDS	GSVIVP00031131001
fanesyl-PP - caotenoids	PDS	GSVIVP00033980001
fanesyl-PP - caotenoids	Ct-b	GSVIVP00037718001
fanesyl-PP - ent-kau-16-ene	4.2.3.19	GSVIVP00002095001
fanesyl-PP - ent-kau-16-ene	5.5.1.13	GSVIVP00023723001
campesteol - steoid	Det2	GSVIVP00021458001
campesteol - steoid	Dwf4	GSVIVP00024952001
campesteol - steoid	Cpd	GSVIVP00029273001
campesteol - steoid	B6ox	GSVIVP00038070001
caotenoids - xanthophyll cycle	Vde	GSVIVP00007662001
caotenoids - xanthophyll cycle	Zep	GSVIVP00019494001
caotenoids - lutein	LUT1	GSVIVP00000577001
ent-kau-16-ene - gibbeellic acid	1.14.11.12	GSVIVP00010598001
ent-kau-16-ene - gibbeellic acid	1.14.13.78	GSVIVP00014740001
ent-kau-16-ene - gibbeellic acid	1.14.13.79	GSVIVP00018965001
ent-kau-16-ene - gibbeellic acid	1.14.11.13	GSVIVP00028281001
ent-kau-16-ene - gibbeellic acid	1.14.11.15	GSVIVP00034482001
xanthophyll cycle - abscissic acid	1.14.13.93	GSVIVP00001184001
xanthophyll cycle - abscissic acid	ABA2	GSVIVP00005501001
xanthophyll cycle - abscissic acid	AAO3	GSVIVP00017483001
xanthophyll cycle - abscissic acid	NCED	GSVIVP00020467001

Appendix 3. Graphs of dN/dS for *Ricinus*

The relationship between dN/dS and pathway position (a-d), and between dN/dS and the Pathway Pleiotropy Index (e-h), for *Ricinus* in four branches of terpenoid synthesis. Trend lines are given as a visual aid, and were not used to assess the strength of these relationships.

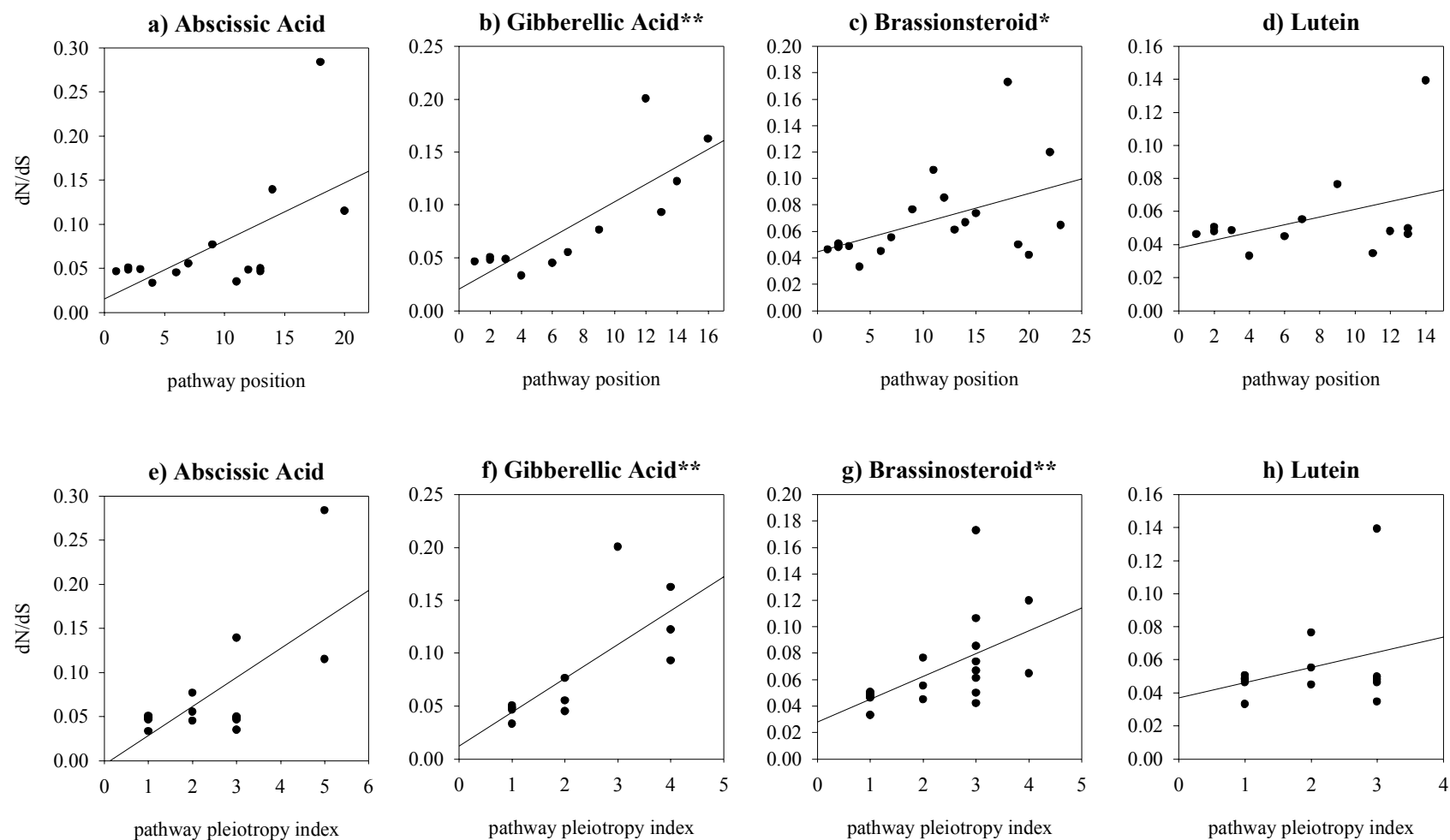
**Correlation is significant and robust to removal of the highest dN/dS value. *Correlation is significant but not robust to removal of the highest dN/dS value. Kendall's τ rank correlation coefficients and p -values are given in Tables 2.1 and 2.2.



Appendix 4. Graphs of dN/dS for *Arabidopsis*

The relationship between dN/dS and pathway position (a-d), and between dN/dS and the Pathway Pleiotropy Index (e-h), for *Arabidopsis* in four branches of terpenoid synthesis. Trend lines are given as a visual aid, and were not used to assess the strength of these relationships.

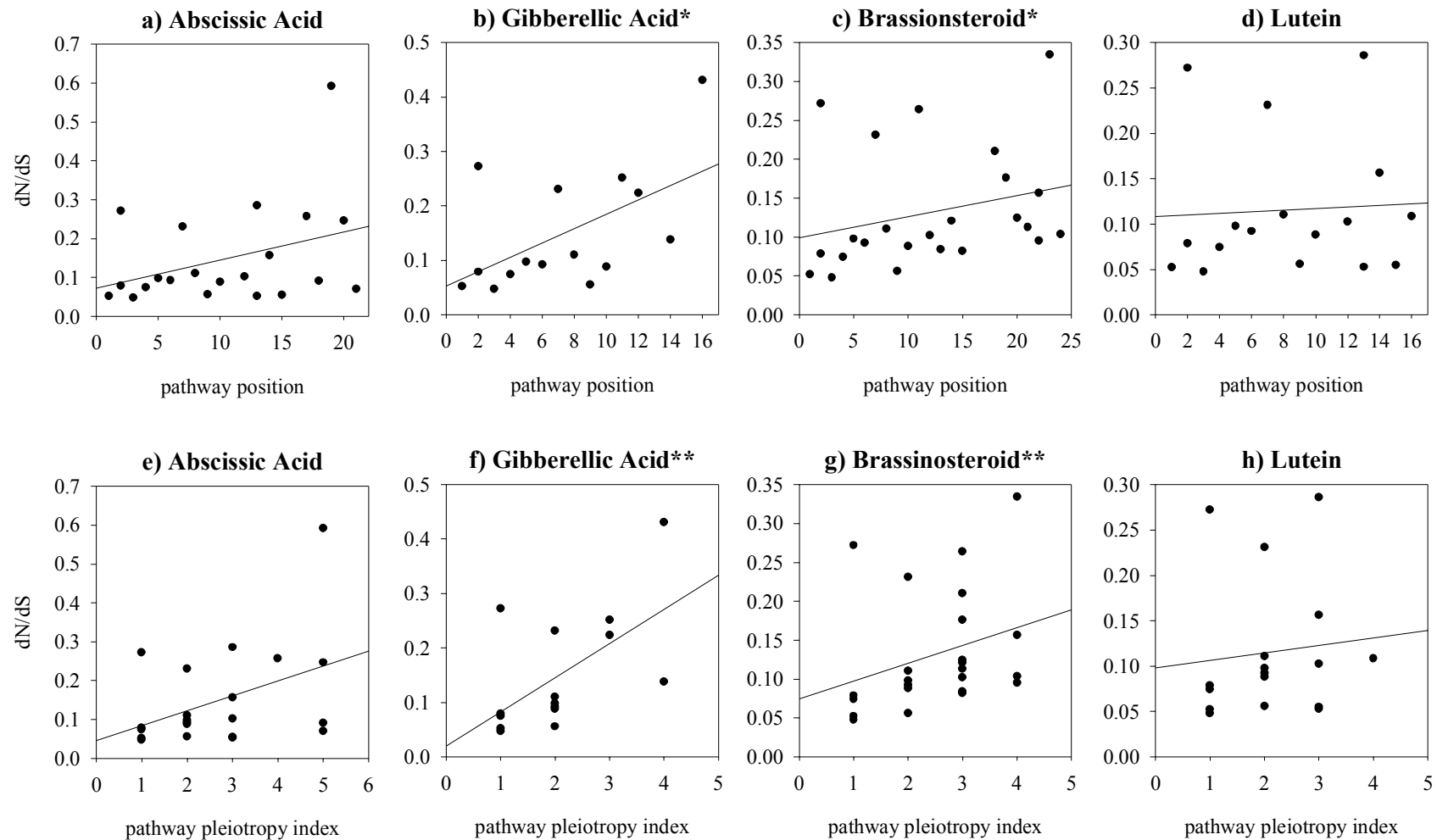
**Correlation is significant and robust to removal of the highest dN/dS value. *Correlation is significant but not robust to removal of the highest dN/dS value. Kendall's τ rank correlation coefficients and p -values are given in Tables 2.1 and 2.2.



Appendix 5. Graphs of dN/dS for *Vitis*

The relationship between dN/dS and pathway position (a-d), and between dN/dS and the Pathway Pleiotropy Index (e-h), for *Vitis* in four branches of terpenoid synthesis. Trend lines are given as a visual aid, and were not used to assess the strength of these relationships.

**Correlation is significant and robust to removal of the highest dN/dS value. *Correlation is significant but not robust to removal of the highest dN/dS value. Kendall's τ rank correlation coefficients and p -values are given in Tables 2.1 and 2.2.



Appendix 6. Alignments used for estimating dN/dS

The following fasta-format alignments were used for analysis in CODEML (Yang 1997; see Chapter 2). These can be cut and pasted into separate files for viewing in an alignment editor. Taxa are denoted by the first letter of the genus name. File names suggested are pathway segment.gene name or EC number.fasta. Pathway segment names refer to the following enzymatic conversions: ABA = carotenoid – abscissic acid, BR = campesterol – brassinosteroid, GA = farnesyl-PP – gibberellic acid, LUT = carotenoid – lutein, CRT = farnesyl-PP – carotenoid, pre-BR = farnesyl-PP – campesterol, TRP = pyruvate – farnesyl-PP, GLY = glucose – pyruvate

ABA.1.14.13.93.fasta

```
>A
-----
-----GATTCTCCGGTTTGTTC
ACTCTCTCC-----GCGCGGCTCTGTTT
TCTGT-----TACTCCGATTATCGCCGGAGTCCGC
CGTAGCTCTCCACGAACTCCCTCTTCTCCGGGAACAATGGGTATCC
TTACGTCGGCGAAACATTCCAACCTTACTCA--CAAGACCTAATGTGT
TCTTTGCAGCAAAACAGAGAAGATACGGATCGGTGTTCAAGACTCATGTA
TTGGGATGTCCATGTGTGATGATCTCGAGCCCTGAAGCAGCGAAATTCGT
ATTGGTTACAAAGTCTCATTTTGTAAACCGACTTTTCCGGCCAGTAAAG
AGAGGATGCTTGGAAAACAAGCCATCTTCTCCATCAAGGAGATTATCAT
TCCAAACTTAGAAAGCTTGTTTAAGAGCTTTCATGCCTGATGCAATCAG
AAACATGGTCCCTCACATTGAATCAATTGCTCAAGAATCACTCAATTCTT
GGGATGGAACCTCAACTCAACACTTACCAGGAAATGAAAACATACACTTTC
AATGTTGCGTTAATCTCAATACTCGGCAAGACGAAGTTTATTACCGAGA
AGATCTAAAACGATGCTACTACATTCTAGAGAAAGGTTACAATTCGATGC
CGATTAATCTTCCAGGAACATTATCCACAAAGCCATGAAAGCTCGCAAG
GAGCTAGGCTCAAATCCTCGCTAACATCTTATCCAAAAGAAAGCAAAACCC
ATCATCACACACAGATCTCCTCGGATCATTCTGGAAGACAAAGCAGGAT
TAACCGACGAACAAATCGCCGATAACATCATCGGAGTAATCTTCGCCGCA
AGAGACACGACGGCGAGTGTCTGACGTGGATCCTCAAGTACTTAGCTGA
TAATCCAACTGTTCTAGAAGCTGTCACTGAAGAGCAAAATGGCAATAAGGA
AAGATAAAAA-----GAAGGAGAGAGTCTCACTTGGGAAGAT
ACAAAGAAGATGCCATTAACTTATAGAGTAATCCAAGAGACATTAAGAGC
TGCTACAATCTTATCTTTCACATTTAGAGAAGCTGTGGAAGATGTGCAAT
ACGAAGGATATTTGATACCAAAGGGATGGAAGTACTGCCACTATTTCAGA
AATATTCATCACAATGCTGATATATTTTCGGATCCGGGGAAATTCGATCC
GTCGAGATTTCGAAGTTGCGCCGAAACCGAATACATTATGCCTTTTGGTA
GTGGGATTCACTTGTGCCAGGCAATGAGTTAGCTAACTTGAAATCTCT
GTTCTAATCCATCATCTCACCCTAAGTACAGATGGTCAATCGTAGGGCC
TAGCGATGGAATTCAGTATGGGCCGTTGCTCTTCTCAGAATGGATTGC
CTATTGCCTTGAACGAAAACCA-----
>P
-----
-----GAATTCTCCCCCTCTGTGTTT
TTCTATCA-----TTTGCTTCCATTCTATTGCTTG
TAGGAGGACTTCTCTTCAAGTCCCTCCTCAAATCTTTGCTTCT-----
---GATGCCCCCAGAGTTTGCCACTCCCTCCGGCACCATGGGG---CC
TTATATGGGTGAAACATTCCAACCTACTCT---CAAGACCCAATGTCT
TCTTTGCCTCTAAGAGAAAGAGGTATGGCTCTATCTTCAAGACCCACATC
TTGGGCTGCTCCCTGTGTCATGATATCCAGCCCTGAAGCTGCAAAATTTGT
CCTTGTAACCAATCCCATCTCTTCAAGCCAACATTTCCGGCTAGTAAAG
AAAGGATGTTAGGCAAAAGAGCCATATCTTTTACCAGGGAGCCTACCAC
ATGAAATTGAGAAAGCTTGTCTTCTGCTCTTCTTGCCTGAAGCAATAAA
AAACATCGTCCCTGATATTCAAACATTGCCAAGGACTCTCTTCAATATT
GGGAAGGAAGATTAATTAACACTTTCCAAGAAATGAAATCATACACATTC
AATGTTGCATTACTTTCAATATTTGGGAAGGATGAAGTTCTATATAGAGA
AGATCTTAAGAGGTGCTACTACATTCTCGAGAAGGGATACAATTCATGCG
CCATTAACCTCCAGGAACACTCTTCAACAAATCAATGAAAGCAAGAAAA
GAGCTGGCTAGGATCTTGGCTAAATCTTGTCCACAAGGAGGCAATGAA
GCTTGATCACAATGACTTACTTGGATCTTTCATGGGTGACAAGGAAGGTC
TCACTGACGATCAAATTGCCGACAACATCATTGGAGTAATCTTGTGCTGT
CGTGACACCACAGCTAGTGTGTTTGACATGGATTCTTAAATACCTTGGAGA
GAATCCAAGTGTCTTCAAGCTGTCACTGAAGAGCAGGAGGCCATAATGA
```

GGAGTGAAGAA-----AAGGGTGATGAGGAGAAGCTCCTTACGGCAGAT
ACCAAGAAGATGCCGATAACTTCAAGGGTCATTCAAGAACTCTAAGAGT
TGCTCGATTTTATCTTTTACCTTCAGAGAGGCGGTGGAGGATGTTGAAT
ATGAAGGATATTTGATCCCTAAGGGTTGGAAAGTCTTGCCACTTTTCAGA
AACATTCACCAAGCCAGCCAGAACTGTTTCCAGATCCTGAGAAGTTCGATCC
TTCAAGATTTGAGGTTGCACCGAAGCCCAATACATTTATGCCATTTGGGA
ATGGGACCCACTCATGTCTGGGAACGAGCTAGCCAAGGTGGAGATTTTG
GTGCTACTCCATCACCTAACCAAGTACAGGTGGTCTATTGTGGGTGC
AAATAATGGGATTGAGTACGGCCCTTTTGCCTTCCCCAGAATGGTTTGC
CCATCATTTTATCCACAAGTCA-----
>R

-----GATTTCTTCGCCTTCTTCTTT
CTATTGTT-----
-----CTTCCATTTCTTCGTCTTCC
TTCTC-----ATTTCTCTCCTTAATTTCTTCTTTCTCTCC---
---AATTCATCCGCCAAATTGCCTCTCCCTCCCGGCTCATTGGGTTGGCC
TTACATTGGAGAGACCTTCCAGCTCTACTCC---CAAAACCCGAATGTCT
TCTTTGCCCTCAAACAGAAAAGGTATGGCTCCGTCTTCAAGACCCACATC
TTGGGTTGCCCTTGTGTGATGATTTCTAGTCCCGAAGCTGCTAGATTGCT
ACTTGTGAGCAAGCTCATCTCTTCAAGCCAACATTCCCGGCGAGCAAG
AAAGGATGTTAGGCAACAAGCCATCTTTTTTACCAAGGAGACTACCAC
AACAAATTGAGAAAACCTGTTCTTTCGCGCTTTTGTACCTGATGCAATCAA
AAGTATCGTCTCTGATATCGAGTCCATTGCTAAAGAATCTGTACACTCTT
GGGAAGGAAGGGACATTAATACTTTTCAAGAAATGAAAACATATGCATT
AATGTTGCACTGCTCTCAATATTTGGAAAGGACGAATTTCAATATAGAGA
AGAACTAAAAAGATGTTACTACATTCTTGAAAAGGGATACAATTCAATGC
CCATTAATCTACCAAGGACACTCTTTCACAAAGCAATGAAAGCAAGAAAA
GAGCTAGCTCAGATCCTAGCCAAAACCTGTGCGAGCAGGAGACAAATGAA
GCTCGACCGAAATGACTTGGTCTTTCATGGAAGACAAGGAAGGAT
TGAGTGATGAACAAATTGCAGATAACATTATTGGAGTGATATTGCGAGCT
CGTGATACTACAGCTAGTGTCTCACATGGATTCTAAAGTATCTTGGAGA
AAACCCAGATGTTCTTCAAGCTGTAACCGAAGAGCAAGAGGAAATAGTAA
GAAGTAAAGAG--AAGAGTGGTGAGCAGAAAGTTCTGAGTTGGGCAGAT
ACAAAGAAGATGCCAGTAACATCAAGGGTTATTCAAGAAACATTAAGGGT
TGCTTCAATTTGTCTTTTACTTTTTCAGAGAAGCTGTTGAGGATGTTGAGT
ATGAAGGATATCTGATCCCAAAAGGATGGAAAGTCTTGCCACTTTTTCAGA
AACATTCACCATAGCCAGAAATCTTTCCAGATCCTGAAAAGTTTGATCC
TTCAAGATTTGAGGTTGCACCAAAACCAATACATTGATGCAATTTGGCA
ATGGGACCCATTCATGTCCAGGGAATGAATTAGCTAAGCTGGAAATCTT
GTCTTCTCCTACCTGACCTGACCAAGTACAGATGGACA--GTGAGTAC
AGATAATGGGATTGAGTATGGCCCTTTTGCCTTCCCCAGAATGGTTTGC
CCATCAAGTTGTCCAGAAATCG-----
>O

-----GGTGCTTTTCTT
CTGTTCTGTG-----
-----TGCGTGCTCGCGCTTTTCTTGCTTG
TCTGCGCC-----GTCCGCGGCCGCGCGGCGAGGCGGGCTCGTCTG
GAAGCGGCGGCGTGCAGGCTGCGGCTGCGCGCGGGGTCGATGGGGTGCC
GTACGTCGGGGAGACGTTCCAGCTGTACTCGTCCAAGAACCCCAACGTGT
TCTTCAACAAGAAGCGGAACAAGTACGGTCCCATCTTCAAGACGACATC
CTGGGATGCCCCCTGCGTGATGGTGTCCAGCCCGGAGGCGGCGGTTCTGT
GCTGGTGACGCGAGGCGACCTCTTCAAGCCACCTTCCGCGCGAGCAAGG
AGCGGATGCTGGGTCCCAGGCCATCTTCTTCCAGCAGGGCGGACTACCAC
GCCACCTCCGCGCATCGTCTCCGCGGCTTCTCCCGAGTCCATCCG
CGCTCCGTCCCGGCTCGAGGCCATCGCGCTCCGCTCCCTCCACTCCT
GGGACGGCCAGTTGCTCAACACCTTCCAAGAGATGAAGACTTACGCGCTG
AATGTGGCATTGCTGTCCATCTTGGGGAGGAGGAGATGCGCTACATCGA
GGAGCTGAACAGTGTCTACCTGACGCTGGAGAAGGGGTACAACCTCGATGC
CGGTGAACCTGCCGGGACCTGTTCCACAAGGCCATGAAGGCCGGAAG
AGGCTGGGCGCCATTGTGGCCACATCATCTCTGCCCAGGCGGAGCGGCA
GCGGGGG--AACGACCTGCTAGGGTCTGTTCTGTGAGCGGCGGAGGCCC
TCACCGACGCCAGATCGCCGACAACGTCATCGGCTCATCTTCCGCGCC
CGCGACACCGCCAGCGTCTTCACTGGATGGTCAAGTTCTTCCGCGCA
CCACCCCGCGTCTCAAGGCCGTCACCGAAGAGCAGCTGCAGATTGCCA
AGGAGAAAGAG-----GCGTCGGGCGAGCCGCTGTCTATGGGCGGAC
ACGCGCGGATGAAGATGACGAGCCGGTCTATCCAGGAGACG--AGGGT
GGCGTCCATCTCTCTTCACTTTCAGGAGGCGGTGGAGGACGTGGAAT
ACCAAGGGTACCTGATCCCCAAGGGCTGGAAAGTGTCTACCTCTGTTCCGC
AACATCCACCACAACCCGACCACTTCCCTGCCCGGAAAAGTTCGACCC
GTCCCGTTGAGGTGGCGCCCAAGCCCAACAGTTTCATGCCGTTCCGGA
ACGGGACCACTCGTGCCCGGCAACGAGCTCGCCAAGCTGGAGATGCTC

GTGCTCTTCCACCACCTCGCAACCAAGTACAGGTGGTCCACGTCCAAGTC
CGAGAGCGGGCTCCAGTTCGGCCCCCTTCGCGCTGCCGCTCAACGGCCTCC
CCAGCTTCACCCGCAAGAACACCGAGCAGGAG
>V
AGCTGTAGAGAGATTCCATTTGTGGGATTGAGCTGTATTTTCTGATATT
TCATTGCCCCCATCTCCCTCATATCCCTCCTTTCTTATTATAAATACCC
CCCTTTTCTTCTCCACTCCTCACACCCATCCTCCACCTCCTCTCTTCTTC
ACTCTCTCCACCATCATCAACCACAACCTATTCAAACCACTACTACTTCT
AATGGACTCCTCCTTACCACCATATGCTTCTCCATTGCTCTCCTCCTCT
TTCTCATCCACTTTCTCCTAAAGCTCTTCTCCTTCTTCTCTGCT-----
---TCCACTGCCCGGAAACTCCCTCTCCACCTGGCTCTCTGGGCTGGCC
CTACATAGGAGAAACCTTTTCAGCTTTACTCT---CAGAACCCCAATGTCT
TCTTTGCCCTCCAAGCAAGAGAGGTATGGAACCATATTCAAACCCACATA
TTGGGGTGCCCATGTGTGATGATTTCAAGTCTGAAGCTGCAAAGCTGGT
GCTGGTCACAAAGGCTCCTCTCTTTAAGCCTACATTTCCGGCTAGCAAAG
AGAGGATGTTGGGGAAACAAGCAATTTTCTTTACCAAGGTGACTACCAT
GCCAAGTTGAGGAAGCTTGTCTCCGAGCCTTCATGCCAGGAGCCATCAA
AAACATCGTCTTAACATCGACTCCATCGCCACCCAGACCTGCACTCCT
GGGAAGGCCGGTCGATCAATACTTTCCAAGAAATGAAGACATACACTTTC
AATGTTGCACTGCTGTCCATATTTGGAAGAGTGAATTTCTGTACAGAGA
GGAATCAAGAAATGCTACTACATTCTTGAGAAGGGGTATAATTCAATGC
CCATAAACATTTCTGGAACGCTCTTTAACAAATCAATGAAAGCAAGGAAG
GAATTGGCTCAGATCTTGGCTAAAATCCTCTCAACAAGGAGGCAGGCGAA
CGGTGATCACAACGACTTGTGGGATCTTTCATGGGAGACAAAGAAGGGC
TCACAGATGAACAAATTGCAGACAACATCATAGGCGTCATCTTCGCGGCC
CGAGACACTACTGCCAGTGTGCTGACATGGATCCTCAAGTATCTGGGCGA
AAACCCGAGTGTCTGCAAGCTGTACGGAAGAGCAAGAGTCCATAATGA
GGAGTAAAGAAGACAAGTGTGGTGAAGAGAAGTCTCTGACTTGGGAAGAT
GCCAAGAAG---CCAATAACTTCAAGGGTGATTCAAGAAACACTTAGAAT
TGCTTCAATTTTATCTTTACCTTTAGAGAGGCTGTTGAAGATGTTGAAT
TTGAAGGGTACCTGATTCCAAAAGGGTGGAAAGTTTTACCACTTTTCAGA
AACATTCACCACAGTCCAGACATATATCCAGAGCCAGAGAAGTTTGATCC
CTCAAGATTTGAGGTTGCTCCAAGGCCAATACATTTATGCCATTTGGAA
ATGGGGTCCACTCATGCCCTGGGAATGAGTTGGCCAAACTAGAGATTTTG
GTGCTTCTACATCATCTGACTACAAAGTACAGGTGGTCTATGGTGGGCAC
ACAGAATGGTATTCAAGTATGGCCCTTTTGCTCTTCCCAGAATGGTTTGC
CCATCACAATCTCCCTCAAGTCA-----

ABA.AA03.fasta

>A
-----GATTTGGAGTT
TGCAGTTAATGGGGAACGATTCAAGATTGATTCT---GTTGATCCATCAA
CGACCTTACTTGAGTTCTTGCATTGAACACTCCTTTCAAGAGTGTCAAG
CTCGGCTGTGGCGAAGGAGGGTGTGGTCTTGTCTTGTAGTGTATCTAA
ATATGATCCAGAGTTAGATCAAGTGAAGGAATGTTGTATAAACTCTTGTC
TCACTCTTCTCTGTAGTGTAAATGGATGTTCCATTACTACATCTGAAGGT
CTTGGAAACACGAAGAAAGGATTTTCATCCGATTGATAAGCGGTTTCGCGGG
TTTTACGCTCTCAATGCGGGTTTTGTACACCAGGGATGTGTATTTCTC
TCTACTCATCTCTTGCAAATGCTGAGAATAACAGC---TCTAAAGAT---
-----TTTACGGTTTCTGAAGCTGAGAAATCTGTTTC
GGGTAATCTTTGTCGATGTACTGGTTATCGTCCCATTGTGCGATGCTTGTA
AGAGTTTTGCTTCTGATGTTGATATTGAGGACTTAGGGTTAAATTCCTTT
TGGAAAAAAGGA---GAAAGCAAAGAGGTTATGTTTAAAGATTTGCCTCC
TTATAACCCCAAAGACCATCTTGTAACATTCCCTGAGTCTTGAAGAAGA
AAGAGAAGGTTGATAATGGTTCGGATCATTGGAAGTATCGT-----
-----TGGACCACTCCTTTTGTGTGGCTGAGCTTCATAA
CATAATGGAAGCAGCTAACTCT-----GGAGACTCGTTGAAGTTAGTTG
TAGGCAACACGGGTACAGGTTACTATAAAGACGAGGAACGGTTT---AGA
TATATTGATATCAGTAACATTCGGAGATGTCAATGATCAAGAAAGATGA
GAAAGGAATTGAAATAGGAGCAGCTGTTACTATATCGAATGCTATTGATG
CTTTGGAGAAGGAAAGCAAATCTTCTTAT-----GTT
TTCAAGAAAATGGCCACTCATATGGAGAAAATTGGCAACCGTTCTATCCG
AAACTCGGGTAGTATTGGTGGTAATCTAGTGATGGCAGAGAGCAGAAAGT
TTCTTCTGATGTTACCACTTTGTTGCTCGCTGTAGATGCCTCGGTGAT
ATGTTAAATGGTAGGAAAACCGAGAAGGTTACGCTACAGGAGTTTCTTGA
ATTGTCTCCGGTGTAGACTCCAAACGCGTTCTTTAAAGGTTGAGATTC
CATCATGGAAGTCT-----CCTTCTGGTGATGACACTGAATTT
CTCTTTGAAAGCTATCGAGCCGCACCTCGTTCTATTGGAATGCATTGCC
TTACTTGAACGCTGCTTTCTTGGCGCTAGTGCTCGCCAAAGAAGCATCAA
GAAAAGGTGTAACGGTAGAAAAATGCTTCTGGCATTGTTGTTCTTATGGC
GGGGATCATTGATCAGGGCGATAGAAGTGGAGACTTTCTTGACAGGTAA
ATTACTTAGCTACAGTGTCTATATGAAGCTGTTGGTTTACTTAAAGGAA
TCATAGTCCCTGGCAAAGACACCTTGCACTCGGAATATAGAAAAAGCTTG

GCTGTTGGATATCTTTTTGAGTTTTTTTACCCACTGATCGAGAGT-----
 -----GGCCATAGAATATGTAGTCTTG
 ATTCAGGGAAT---AAGCATAACAACAGCCATGTTGATACCGTAAAACTCT
 CTTCCTCTCTCATCTTCACAGCAAGTATTAGAG---AGCAATGAATT
 TAAACCTATTGGTGAAGCAGTTATCAAAGTTGGAGCTGCCTTACAAGCTT
 CTGGTGAGGCGGTTTTTGTGAT--ATACCGACTTTACCAGACTGTCTT
 CATGGAGCATTCAATTTATAGCACAGAGCCTTTGGCTAAGATAAAAAGCTT
 AAGCTTCAGAGAGAATGTAACCTCTACTGGAGTTTTTGTCTTCTACTT
 TCAAGGATATTCACAACAAGGACAAAACATTGGT---TCCAAGACCTT
 TTTGGACCAGGACCTTTATTTGCAGATGAACTAACTCGATGTGCAGGTCA
 AAGAATTGCTCTTGTGGTTGCAGACACACAGAAACATGCAGACATGGCTG
 CAAAACCTTGCAAGTAGTTGAGTATGATACAAAGAATTTAGAACAACCGATA
 TTAACGTGCGAAGATGCAAGTTAAAGATCTAGCTTTTTTCGAGGTTTCATCC
 GATGTTTTACCCGAACCAAGTAGGTGATGTTATAAAGGAATGGAAGAAG
 CTGAGCGAAAGATAATCTCCTCTGAGTTGAGGCTAGGGTCACAGTACTTC
 TTCTATATGGAGCCACAACAGCACTTGCCCTTGCCAGACGAAGACAATTG
 TGTCAAGGTATTTAGTTCATCTCAAGCACCTGAGTACGTACATTCGGTTA
 TTGCTACATGTCTTGGCATTCAAGAGCATAACGTGAGAGTCATCACTAGA
 AGAGTTGGTGGTGGCTTTGGTGGTAAAGCTGTAAAGTCAATGCCTGTTGC
 AACCGCATGCGCACTAGGCGCATACAAGTTGCAACGCTCTGTTAAGATGT
 TTTTGAACCGCAAGACGGATATGATAATGGCTGGAGGAAGACATCCGATG
 AAGATAAATTACAATGTGGGATTTAGATCAGATGGGAAGCTCACAGCACT
 GGAGCTTACAATGCTTATAGATGCAGGGCTTGAGCCAGACGTAAGTCCAA
 TCATGCCCGGAATATAATGGGTCCGCTGAGGAAGTATGACTGGGGAGCT
 TTGTCATTTGATGTAAAAGTGTGTAAGACGAATTGTCTGAGTAGAACAGC
 GATGAGAGCTCCCGGGGAAGTTCAAGGTTCTTACATAGCGGAATCCATTA
 TCGAGAATGTAGCTTCTCTCTTCAAATGGATGTTGATGCAGTGAGAAAA
 ATAAACCTTCATACTTATGATAGCCTTAGAAAAATTCTACAATCACATTGC
 TGGTGATCCTGATGAATATACACTGCCCTTATTATGGGAGAACTTGAGA
 TATCTTCAAAGTTCAAGGAAAGATCG---ATGGTGAAGGAGTTTAATCTT
 TGTAAATGTATGGCGAAAGAGAGGGATTTCTCGAGTACCTATAGTCCATCA
 AGTTATGCAGAGACCGACCCCGGGGAAGTAAGCATACTGAGTGATGGTT
 CTGTTGTAGTTGAGGTTGGAGGAATTGAGATAGGGCAAGGATTGTGGACT
 AAAGTACACAGATGTTGCTTATGGTCTCGGTATGGTTAAATGCGAAGG
 AAACGAAAAGCTCTTGGATAGAATACGTGTAGTTCAAGTCCGATACACTAG
 GTATGATCCAAGGAGGTTTCACTGCTGGTAGCACGACATCCGAGAGCAGT
 TGTGAAGCCGTTAGGCTTTGCTGTGTTATCTTGGTGGAGAGATTGAAACC
 TATAAATGGATCAGATGATGATGAGAGAAGTCAGGTTCCGTGACATGGAACA
 TACTCATTCAACAAGCGTATGGTCAGTATATTAATTTATCGGCGAGTACA
 -----TTGCCAGAGTATTCTTCCATGGAATATCTCAACTATGGAGTTGG
 AGTC---AGCGAGGTGGAAGTGGACCTTGTGACAGGAAAAACCGAGATTT
 TAAGATCCGATATTATTTATGACTGTGGAAGAGTCTTAATCCTGCGATT
 GATTTAGACAGACCGAAGGAGCATTGTTCAAGGCATCGGGTTTTTCAT
 GATGGAAGAGTACACTACTGATGAGAAGGGGCTTGTGGTGCAACAAGGCA
 CTTGGGACTACAAAATACCGACTGTGGACACAATCCCAAAACACTTCAAC
 GTTGAGATTCTCAACACTGGTCATCATAAAAACCGCGCTTCTCTCTCCAA
 AGCGTCGGGTGAGCCGCCATTGCTTTTAGCAGCTTCTGTTTATTGTGCAA
 CAAGATCAGCCATTAGGGAAGCTCGGAAACATTCCCTTTTCATCGAACTTC
 ATTGATGGGTCTGATTCCGAATTTGAGTTACCTGTTCCAGCGACTATGCC
 TGTGGTGAAGTCTCTTTGCGGATTGTATAGTGTAGAGAAATACTTACAAG
 GAAAGATCAAAGGACAA
 >P
 GAAGAGAAGCAGAGG-----GAAACAGGCAGAGGGAGTCTAGTTTT
 TGCCGTCAACGGACAGAGGTTGAGGTTGTCGTCGAGATTAGACCTTCAA
 CAACCTTGCTTGAGTTCTTGCGAACCAGGACATCTTTCAAGAGTGTCAAG
 CTCGGTTGTGGCGAAGGTGGTTGTGGTGCTTGCAATTGTTCTACTCTCCAA
 GTATGATCCTGTGCGTGACCAAGTTGAGGATTTACAGTCAAGTTCATGTC
 TCACACTGCTTTGCAGTGTGAATGGATGTTGCGTTACAACATCTGAAGGC
 CTTGGAATAGCAAAGATGGATTCCATCCGATTCACCAGAGATTCAAGTGG
 TTTCCATGCTTCTCAATGTGGCTTTTGTACTCCTGGAATGTGTGTTTAC
 TCTTCGGAGCCCTTGTTAAAGCTGAAAAGAATGAC---CAAAGGGAGCCT
 TCTCCAGGATCTCCAAGCTGACAGTCGTTGAAGCTGAAAAGGCTATCTC
 AGGAAATCTTTGTCGCTGTAAGTATCGACCCATTGCTGATGCGTGTA
 AGAGTTTTGCGGGTGATGTTGATATTGAAGATTTGGGATTAACCTCTTT
 TGGAAAGAGGAA---GAGAGTCCCGAAGCGAAGATGAGTAGGCTACCTTT
 ATATGACCATAATCATGAGATATGCACTTTCCCTGAATTTTTGAAAAGGG
 AAATAAAATCTTCTTGTCTTTGGATTCTGAAAGATATTCT-----
 -----TGGTGACACCTGCTACTGTTGAGGAGCTTCAGAG
 CTTATTA AAAAGCATTGATGCCGACTGCAAAACAGGATGAAACTAGTGG
 TTGGTAACACAGGTATGGGTTATTACAAGGACCTAGACCACCATGACAAG
 TACATAGATCTCAGATGTGTTCTGGAGCTCTCGAGTATTAGGAGGGAAGA
 A---GGAATCGAAATCGGGGCAGCTGTCACTATTTCTAAAACATTGGAAG
 CTCTGAAGGAAGAAATTAACAGCGAGTTAATTCAGAAATGCAAGATTGTG
 TTTAAAAGAATTGCATTGCACATGGAGAAGATTGCTTCTGAATTTGTTTCG
 AAATACAGGCAGTGTAGGGGGGAATTTGGTGATGGCACAAAGGAAACATT

TTCCATCAGATATTGCCACGATACTGCTGGCAGCAGGTGCATTGTTCAC
 ATTCTAACTGGTACCTTGCGATGAAAAGCTTACTTTAGATGAGTTTCTGGA
 AAGGCCTCCA - - - TTGGATTCCAAAAGTGTGCTATTAATATTAAGATT
 CAAATATGCGAGCATCCAAGAATATCTTCTGAAATGGACAGCAAGTTG
 TTATTGAAACTTATCGTGTGCGACCGGACCCCTTGAAATGCATTGCC
 CTATTTAAATGCAGCTTTCTTGTCTGAAGTTTCTGTTGAAATCTTCT -
 - - GGTTCAGCCGTGTTAAATAAATGCCGGTCTTTTTGGTGCATTGGA
 ACCAAACATGCTATCAGAGCAAAGGAAGTTGAGAAATTTTGTCTGGAAA
 AATACTAACCATTTGGTGTCTATATGAAGCTGTTAACTGGTTAAAGCCA
 ATGTGGTTCTGAAGATGGCACGCCGAGTCCGGCTACAGGTCAAGCTTA
 GCTGCTGGTTATCTCTTTGACTTTCTCTACCCCTTGGTAGACATTAACCT
 TAAAATTTCTGGTGTGGTGGTGGATGAATATTGTAATACTTCATTGTTCA
 AGGATGCAAAAATAAAACAGAAGTACAGCCAGCTTGATCATGTGCAATTA
 CCCACTTTGCTGTGCATCATCAGAGCAGGTGCTTGAATTAACAATGATCA
 TCATCCTGTTGGTCAGCCTACTAAGAAAGTTGGAGCCGCCCTTCAAGCTT
 CTGGGGAGGCTGTTTTGTGGATGACATTCCTCTCTCAAAATTTGCTCA
 CATGGAGCATTCATTCATAGCATGAAGCCTTATGCAAGGGTCAAGGATAT
 CAAATTCAGTCTAAATTAATACCAGATGGAGTTTCTGGGCTGATTTCTGG
 TCAGAGACATTCAAAAGGTGGGAGAACAGAGGT - - - TGTACGACTAGT
 TTTGGCACTGAATCTTTGTTCCGAGATGAGCTTACGCAATGCTGGAGA
 GCGTCTTGCTTTTGTGGTTGCAGATACACAGAAACATGCTGATATAGCAT
 CCAACCTTGGAGTTGATTATGACATTGAAATCTAGAACCCACTT
 TTAATGTAGAAGAGGCCATTAAGAGATCTAGCCTTCTCGAGGTTCTCT
 TCTCTCTACCCCAAACAGTTGGTGATATATCAAAAGGATTGGCTGAGG
 CTGATCACAAGATTCTCTGCAAGATAAACTCGGTTACAGTACCAT
 TTCTATATGGAGACTCAAACCTGCGCTTCCGCTTCTGATGAAACAACTG
 CATGGTGGTTTACAGTTCAACCCAGTGTCCGAATATGCACATGTTAACA
 TTGCAAAATGCTTGGTATTCTGAGCATAATGTGCGTGTGATTACAAGG
 AGAGTCGGAGGAGGCTTTGGTGGAAAGGCCATGAAAGCAATTCCTGTTGC
 TACAGCATGTGCTTGTGAGCGCACAGTTTCTGCGCCTGTGAGAACGT
 ATTTGAATCGCAAGACCGATATGATAATGGCAGGAGGAAGGCATCCCATG
 GAAATAACTTATAATGTAGGATTTAAATCAAATGGGAAAGTTACAGCCTT
 ACAGCTTGATATATTAATCAATGCTGGAATATCTTTCGATATAAGTCCAC
 TGATGCCAAAACCATTTCTGAGTGGGTAAAAAAGTATGACTGGGGTCT
 TTATCTTTTGATATAAAGGTATGCAAAACAAATCATTCAAGTAAACTGC
 AATGCGAGGCCCTGGGGAGGTACAAGGATCATACATTGCAGAAACCGTAA
 TCGAACATGTAGCGTCTACCTTTCCATGGATGTAGATTCTGTGAGAAAC
 ATAAATTTTACAGATATGATAGCCTTAAATTTCTATGATGTTGCCTC
 AGGTGATTCTGTAGAGTATACTCTAACATCAATATGGAATAAGTTAGCAG
 AATCTTCAAGCTTCAAACAAAGGGTTGAAATAATAAAGGAGTTCAATAGG
 TGTAAGGTGTGGAAGAAAGAGGTATTTCTCGAGTGCCTATTGTGCATCA
 AGTGTTCGTGGGACCAACTCCGGGGAAGTAAGCATTCTAAGTGATGGGT
 CGGTTGTGCTTGAAGTCGGCGGAATAGAGTTGGGTCAAGGGCTCTGGACA
 AAGGTAACAAATGGCTGCATTGCTCTCAGCTCAATCAAATGTGACGG
 GGTAGAAAATCTCTGGACAAAGTAAGGGTAACAAGCTGATACTCTGA
 GTTTAACTCAAGGAGGAATGACTGCAGGAGCACCACATCAGAGTCAAGC
 TGTGAATCAGTTAGACTTTGCTGTGCAGTCTTGGTTGAGAGACTGGCACC
 TCTAAAGGAAACGTTG - - - CAGGGGCAAATGGGTTCCGTAACATGGGATG
 CGCTCATTTGCAAGGCATATGTGGAATCACTGAATTTATCAGCATCTTCA
 CATTACATCTCGTACTTTACTTTCGATGCACTACTTAACTATGGCGTGC
 AGTA - - - AGTGAGGTAGAGGTAACCTTCTGACAGGGGAAACAACTATTT
 TGAGATCAGATATTATATACGATTGTGGACAAAGTCTCAACCTGCGGTG
 GATTTAGGACAGATTGAAGGAGCCTTTGTCCAAGGGATTGGTTTTTTAT
 GCTTGAAGAGTACACGACGAATTCTGATGGACTAGTGGTTGCAGACAGCA
 CATGGACATATAAGATCCCTACAATAGACACCATAACAAAACAATTCAAT
 GTGGAAATACACAACAGTGGACATCACCAGAAACGTGTTCTCTCTTCAAA
 AGCTTCCGGGGAGCCACCGCTACTCTTGCAGCATCAGTTCACTGTGCTG
 CAAGAGCAGCTATAAGAGATGCTAGACAACAGCTTCATTATGCGGTTGC
 ATGGATGAGTCTTACTCAACATTCAACTTGGAGGTCCCTGCCACCATGCC
 TAAGGTGAAGGAACTTTGTGGCTGGACAACGTGGAAGGTACTTAGGGT
 GGAAAATGGGTAGAAAAG
 >R
 GAGGATCATGAAAGTACAGCAACAGAAACAGAAAGGAGCAATCTTCTTTT
 TGCTGTTAATGGGAGAGGTTTGAGCTCTCAAGT - - - GTTGACCCTTCTA
 CTACTTTGCTCGAGTTCTTGCCTACCCAGACAGCTTTCAAGAGTGTTAAG
 CTCAGTTGCGGTGAAGGTGGTTGTGGTCTTGTATTGCGCTGCTCTCCAA
 GTATGATCCGTTCAAGTACGAAGTTGAGGATTTTACTGTAAGTTTATGTC
 TCACATTGCTTTGCAGTATAAATGGATGTTCAATTACAACAAGTGAAGGC
 CTTGGTAATAGCAAAGATGGCTTCCACTCAATTCACCAGAGGTTTACAGG
 ATTCCATGCTTCTCAATGCGGCTTTTGTACTCCAGGAATTTGCATCTCAC
 TCTATGGGGCCCTTGTAAATGCTGAAAAGACAGAT - - - CGACCAGAGCCC
 TCACCAGGATTCTCAAGCTTACAGTGGTTGAAGCTGAGAAAGCTGTTGC
 AGGAAATCTTTGTCGCTGACTGGATATAGACCATTTGCCGATGCGATGCA
 AAAGTTTTGCAGCTAATGTTGATATGGAGGACTTAGGATTCAACTCTTTC
 TGGAAAGAGGAA - - - GACATTCAGGAAGCAAAGATAAGCAAGTTACCTGT

GTATAATCATAATCACAAATAGTTGTAAGTTTCCCTGACTTCTTAAAAAGGG
 AAGTCAAGATTCTTGGCTTTTGGATTCTAAAAGATACCAT-----
 -----TGGTACAAACCTGCTAAAAATTGAGGAGCTTCATGA
 CTTATTAATAATCTAGTGATGCTGAT---GGAGTTCGTAGGAACTAGTGG
 TTGGTAATACAGGTGTGAGTTACTATAAGGAAGTAGAGTATTATGACACG
 TACATAGACCTCAGAAATATTCCTGAGCTCTCAATCATTAGAAGGGAGCA
 GTCAGGAGTTGAAATTGGGGCTGCTGTCACTATTTGAAAGCTATTGAAG
 CTTTGAAGGAGGAAAGTAAAGGTGAGTTTCTCTCAGAATGTAAATGATA
 TATGAAAAAATTGCTATTCATATGGAGAAAATTGCTGCTGCATTTGTTAG
 AAATACAGGTAGTGTGGGGGAAATTTAGTGATGGCACAGAGGAAACATT
 TTCCTTCAGATATTGCTACTATACTTCTAGCAGCTGGTTCATCAGTTGAG
 ATAATGACAGGCATTATTCGTAAGAAAGCTTACACTGGAGGAGTTTTTGGG
 AAGGCTCCCA---TTGGATTCCAAAAGTGTACTTTTAAAGTGTAGGATCC
 CTAATTGTGAATCAATTAAGAATGTATCTCTTGAAAGAGATAATAAGTTG
 TTATTTGAAACCTATCGAGCTGCACCGAGGCCCTTGGAATGCTTTGTC
 CTATCTGAATGCTGCTTTCTTGGCTGATGTTGCCTGCTCTAAACAATCG-
 --GGTGAATTTGTTAAATAGCTGTCGGTTGGCTTTTGGTGGCTTTTGGG
 ACCAAACATGCAATTAGGGCGAGGAAAGTTGAGGAATTTTGTAGTGGAAA
 ATTGTTAACCATTGGTGTCTTTACGAAGCTATTAATTTGGTTAAATCTA
 CTGTGATACCTGAAGAAGGGACTCGACATCTGCATATAGGACAAGCTTA
 GCTGTTGGTTTTCTTTTGGACTTTCTTGGCCAGTATCAGTTACT-----
 ----CTTGGTAGTGGTTGGTTAGATGGAGGTATAAATCTTCAATATTCA
 ACGGTGCGATACTAAACCAAAATCAAGCGTGGCTGGACCAAGTCAAATTC
 CCCACTTTGCTATCTTCATCAAGCAGGTGGTCCAAATAAACAAAGATTA
 TCATCCCATTGGCGAGCCAGTTACAAAATCTGGAGCTGCTCTTCAAGCTT
 CTGGTGAAGCTGTTTATGTGGATGACATTCCCTCTCCAGAAATTTGTCTA
 CATGGTGCATTCGTTTATAGCAAAAAGCCTTTTGCAGGGTAAAGGATAT
 AGAACTCAATTCCTAAATCCATATCAGTGGAGTCACTGCATTTATTACTT
 TTAGAGACATCCCAAGGCGGGGAAACATAGGT---TCTAAACTATT
 TTTGGTTTGGAGCCTCTGTTTGTCTGATGAGCTTACTCGATGTTGTGGAGA
 GCGTCTTGCTCTAGTGGTTGCAGATACACAGAAACATGCAGAATTAGCAT
 CAAACCTTGCAGTGGTTGACTATGACCTGGAAATCTAGACTCTCCCAT
 TTAACGTAGAGGATGCTATAAAGAGATCTAGCCTTTTTGATGTCCCTCC
 CTTTCTCTACCCTAAACAAAGTTGGTGATATATAAAGGGAATGGCTCAAG
 CTGATCACAAAGATTCTCTCAGCTGAGATAAACTTGGGTACAACTACTAT
 TTCATATATGGAGAATCAAACCTGCCCTTGCTGTGCCAGATGAAGACAATTG
 CATAGTAATTTATAGTTCAATTCAGTGCCCTGAGTTTGCACATGCTGTTA
 TTTCAAGATGCTTTGGTCTTCTGAACATAATGTACGTGTGATTACAAGA
 AGGGTTGGAGGAGGCTTTGGAGGAAAGGCCATAAAAGCCATGCCTGTTGC
 TACAGCCTGTGCACTAGCAGCATACAAGTTACAGCGCCCTGTGAGATTGT
 ATCTTAATCGCAAAGTAGATATGATAATGGCAGGAGGAAGGCATCCTATG
 AAAATAACATACAGTGTAGGGTTCAAATCTAATGGGAAAATTACAGCCTT
 ACAGCTTGATATATTGATTGATGCTGGCATATTTCCAGACATAAGTCCAA
 TAATGCCAATAAATATTCTGGGCTCGCTGAAAAAATATGACTGGGGTGT
 TTATCTTTTGATATAAAGGTATGCAAAACAACTTCCAAGTAGATCAGC
 TATGCGGGCCCTGGAGAGGTGCAAGGATCATATATCGCAGAAGCTGTAA
 TTGAACATGTAGCATCTTCTCTTCTGTGGATGCAGATTGAGTGAAGGCC
 ATAAACCTCCACACTTATGATAGCATTAACTTATTCTACGATAATATTGT
 TGGTGAGCCTCTAGAATACACTTTAACTTCAATATGGGATAAGTTAGTGA
 CATCTTCAAGTTTATCCAAAGGACTAAAATGATAAAAGAGTTCAATAAG
 TGTAAATTTGGGAAAAAAGAGGTATTTCTCAAATACCTATTGTACATCA
 AGTTACATTGAGACCAACTCCAGGGAAAGTAAGCATTCTAAGTGATGGGT
 CTGTTGTTGTGCAAGTTGGAGGAATAGAGCTGGGCCAGGGGCTCTGGACG
 AAGGTGAACACAGATGGCAGCATTGCTCTTAGTTCAATCAAGTGATGG
 AGTAGGAGACCTCTTGGATAAAGTAAGAGTCATACAAGGTGACACCTTGA
 GTTTAATTCAAGGGGGTTTACTTCTGGGAGCACTACATCCGAGTCAAGC
 TGTGAAGTAGTTAGACTTTGCTGTAAAGATTTGGTCGACAGACTGACACC
 TCTGAAAGAAAAGGTAA---CAAGGGCAAATGGGTTCTATAAGATGGGAGG
 TGCTAATTCACCAGGCTTATTTGGAAGCTGTCAACTTATCAGCAAGTTCT
 TACTTCGTCCCTGACTTTGCTTCTATGCAATACCTAAATTATGGTGTGTC
 TTCG---AGCGAGGTGGAGATAGACCTTCTGACAGGACAAACAATTT
 TGAGATCAGATATTATATATGACTGTGGACAAAGTCTCAACCTGTGTG
 GATCTAGGCCAGATCGAAGGGGCTTTCTGTCAGGAATTTGGCTTTTTCAT
 GCTTGAAGAGTACACAACAACTCGGATGGATTAGTGATTGAGGATGGCA
 CATGGACATACAAGATCCCTACACTAGACACTATACCAAAACAGTTCAAT
 GTGGAGATTTTAAACAGTGGCCATACCAAAAGCGTGTCTATCATCCAA
 AGCTTCTGGTGAACCGCCATTGCTCCTAGCAGCCTCAGTTCACTGTGCCA
 TAAGAGCTGCTATCAGGGATGCCCGACAGCAGCTTCATTTGTGGGGTTC
 CTAGATGATAGTCTACAACGTTGCACTTGGAGGTCCCTGCCACCATGCC
 TGTGTAAAAGAGCTCTGCCGGCTGGACATTGTGGAGAGGCACTTGCAGT
 GGAATAATGAACAGCAAG
 >0
 GGGTCAGAGGCGGCG-----GCGGCGGCGAGGGCGGTGGTGGT
 GCGGTGAACGGGAGAGGTACGAGGCGGTGGG---GTCGACCCGTCGA
 CGACGCTGCTCAGATTCTCCGACGCGGACCCCGTCAGGGGCCCAAG

CTCGGCTGCGGCGAAGGTGGATGTGGCGCATGCGTGGTCGTCGTGTCCAA
GTACGACGCCGTCGCCGACGAGGTGACCGAGTTCTCGGCGAGCTCCTGCC
TGACGCTGCTCGGCAGCCTGCACCACTGCGCGGTACACCACGAGGAGG
ATCGGCAACTCCAGGGATGGCTTCCACGCTGTGCAGCGGCGGCTCTCCGG
CTTCCACGCCCTCACAGTGGGCTTCTGCACGCCGGGATGTGCATGTCCA
TCTACTCCGCGCTCGCCAAGGCTGACAAGGCATCCGGTCCGCGCGGCCG
CCGACGGGGTTCTCAAGATCACCGCCGCCGAGGCAAGAAAGCCGTCTC
CGGCAACCTGTGCCGCTGCACTGGCTACAGGCCCATCGTCGACGCTGCA
AGAGCTTCGCCGCGGACGTCGACCTCGAGGATCTTGGCTCAACGCCCTT
TGGAAAGAAAGGCGTCGACGACGAGCAGCTGACATCAACAAGCTGCCTGC
ATACTCCGGTGGCGCCGCGTCTGTACGTTCCCGGAGTTCTCAAGTCGG
AGATAAGGTGTCATGGGTCAAGCGAATGGTGACACTTCGGCGGTGGTG
GTCACCGGCGACGGCTGGTTCCACCCCAAGAGCGTCGAGGAGTTCCACAG
GTTGTTTGTACTCTAACCTGTTTCGAT --- GAAAGGTCTGTGAAGATTGTGG
CATCGAACACCGGCTCCGGAGTGACAAAGGATCAAGACCTCCATGACAAG
TACATCAACATCTCGCAGATTCCGGAACCTCTCAGCCATCAATAGAAGCAG
CAAGGGCGTAGAGATCGGAGCTGTGTGTCTATCTCCCAAGCCATCGACA
TACTGTGCGATGGAGGCGG-----GTG
TTTAGAAAGATCGCTGATCACCTGAGCAAAGTGGCTCGCCGTTCTGTCG
GAACACGGCGACCATTTGGAGGCAACATCATCATGGCGCAGAGGTTGTCTT
TCTGTCGGACATCGCAACTGTTCTTCTGCTGCAGGATCACTGTCCACC
ATCCAGGTGGCCGCCCAAGAGGATGTGCATCACTCTGGAGGAGTTCTTGAA
GCAGCCTCCA---TGTGATTCCAGGACATTGCTGGTCAGCATCTCTATCC
CGGATTGG-----GGTTCAAGATGATGGCATC
ACCTTTTCAGACGTTTCGTGCGGCTCCTCGTCCACTTGGCAATGCTGTGTC
ATATGTTAATTCTGCATTCTTGGCAAGAAGTTCAAGTGGATGGATCATCA-
--GGGAGCCATCTCATTGAGGATGTTTGTGGCATTCCGCCCTTTGGG
GCCAAGCACGCCATCAGAGCTAGGGAGGTTGAGAAGTTCTTGAAGGGAAA
ATTGGTTAGTGCACCACTGATACTTGAAGCAGTTCTGTTCTGAAAGGTG
TTGTTTCAGAGCTGAAGGCACAAACATCCTGAATACAGAGTCAGCTTG
GCCGTCAAGTTATCTGTTCAAATTCCTGTCTTCCCTTACTAATGGCTTGG
TGAGCCTGAAAATGCTAATGTTCCCAATGGTTCAATTTACTAATGGAACCG
CAAATGGTATTGTTGATTCTCGCCAGAGAAGCACTCAAATGTTGACAGC
TCTTATTGCGCAATAAAATCAAGGCAAGAGATGGTTTTCTCTGATGAATA
CAGACCTATTGGAAAACCGATCGAGAAAACGGGAGCTGAGCTCCAAGCTT
CTGGGGAGGCGAGTGATGTCGATGACATCTCTGCTCCAAAGGATTGCCCT
TATGGGGCATTTATCTACAGCACACCCCTCACGTCACATAAAAGGTGT
AAACTTCAGATCATCTTTGGCTTCACAGAAGGTCATCACAGTTATCACTT
TAAAGGACATTCTACCAATGGAAAAAATATTGGATCATGCTCCCAATG
CTAGGAGATGAAGCGCTTTTCGTTGATCCTGTTTCTGAATTTGCTGGACA
GAATATTGGTGTGTGATTGCTGAAACTCAAAGTATGCCTATATGGCGG
CGAAGCAATCTGTGATCGAATATAGTACTGAAAATCTTCAGCCACCAATT
CTGACAGTAGAAGACGCTGTCCAGCATAACAGCTACTTCCAAGTTCCCC
ATTTTTAGCTCCTACGCCAATTGGTGAGTTCAACCAAGCCATGTCTGAAG
CTGATCACAAGATCATAGATGGGGAGGTGAAACTGAATCCCAATACTAT
TTCTACATGGAGACACAGACGGCCTTAGCTATCCCTGATGAAGATAACTG
TATAACCCTCTATGTGTCGGCGCAACTACCTGAGATTACTCAAATACTG
TTGCGAGGTGCCCTTGGCATTCCATATCACAACGTCGGTATCATCACGAGA
AGAGTTGGAGGAGGTTTTGGTGGAAAGGCAATGAAAGCAATACATGTTGC
AACTGCTTGTGCTGTGCGCGCATTCAAGCTGCGGCGTCCGGTTCGAATGT
ACCTCGATCGCAAGACAGACATGATCATGGCGGGAGGGCGGCATCCGATG
AAGGTAAGTACTCCGTCGGTTTCAAGTCCGACGGCAAGATCACGGGTCT
GCACGTTGATCTCAGAATCAACTGCGGAATATCGCCGGACTGCAGCCCGG
CGTTGCCAGTCCGCAATTGTGCGGGCTCTGAAGAAGTACAACCTGGGGGCT
CTCTCCTTCGACATCAAGCTCTGCAAGACGAACGCTCTGTCGAAGTCGGC
GATGCGGGCTCCCGGCGACGCGCAGGGCTCGTTCATCGCCGAGGCCATCG
TCGAACACATCGCGTCTACTCTCTCAGTGGACACGAACGCCATCAGGAGG
AAGAACCCTCCATGACTTCGAGAGCCTCAAGGTGTTTTACGGGAATAGCGC
CGGTGATCCTTCGACGTACAGCCTCGTACCATCTTCGACAAGCTGGCGT
CGTCCGCCGAGTACCAGCAGAGGGCCGCGTGGTTCGAGCATTTCAACGCC
GGCAGCAGGTGGAAGAAGCGGGGCAATTTCTGCGTGCCGATCACCTACGA
CGTGAGGCTCCGGCCATCTCCGGGGAAGGTGTCCATCATGAACGACGGCT
CCATCGCGCTCGAGGTCGGCGGCGTCGAGATCGGGCAGGGGCTGTGGACG
AAAGTGAAGCAG---ACAGCGTTGCGGCTGGGCCAGCTCTGCGATGATGG
CGGCGAGGGGCTCCTCGACAAGGTCGGTGTATCCAGGCCGACACCCTGA
GCATGATCCAGGGAGGCTTACCAGGCGGGAGCAGACCTCCGAGACTAGC
TGCAGGGCTGTTGGAAGTCGTGCGCGGCCCTCGTGGAGAGGTTGAAGCC
CATCAAGGAGAAGGCT-----GGCACGCTACCATGGAAT
CCTTGATTGCCCAGGCAAGCATGGCGAGCGTGAAGCTGACAGAGCACGCG
TACTGGAGCCTGACCCACGTTCACTAGCTACTTGAATATGGAGCTGC
AATT---AGTGAGGTGGAAGTCGATGTGCTGACGGGAGAAACCACCATCC
TGAGGAGCGACCTCGTGTACGACTGTGGGCAGAGCCTGAATCCTGCTGTG
GACTTGGGCCAGGTGGAAGGCGCATTCTGTGAAGGGATCGGCTTCTTAC
GAACGAGGAGTACACGACCAACTCCGATGGGCTGGTCATCAACGACGGCA
CGTGACGTACAAGATCCCAACCGTGGACACCATCCCAAGCAGTTCAAC

GTCGAGCTCATCAACAGCGCCCGGATCACAAGCGCGTCTCTCTCCAA
 AGCGTCCGGTGAGCCGCCGCTGCTGCTGGCGTCGTCGGTGCACTGCGCGA
 TGAGGGAGGCGATCAGGGCGGCCAGGAAGGAG-----TTCGCCGGC
 GCCGGCGGCTCGTCGCTGACGTTCCAGATGGACGTGCCGGCGACGATGCC
 GATCGTCAAGGAGCTCTCGGCCCTCGACGTCGTGGAGAGAGATCTCGAGA
 GCTTTGCTGCCAAAGCC
 >V
 GAGCAGTCTGAACCA-----ACTGTCAACGACTGTCTGGTTTT
 TTCTGTTAATGGGGAGAGGTTTGAGGTCTCCACC---ATCCACCCTTCCA
 CTACTTTGCTTGAATTCTTGCCTTCTCACACTCCTTTCAAAGGTGCCAAG
 CTCAGCTGTGGTGAAGGTGGCTGTGGTCTTGTGTTGTTCTTCTGTCCAA
 GTATGACCTGTGCTTGACCAGGTGGATGATTTTGCAGTGAGTTCATGTC
 TTACACTGCTTTGCAGTATAAATGGATGTTCAATTACTACAACTGAAGGC
 CTTGGGAACATAAAAAATGGGTTCCACCCAATTCATGAAAGGTTCTCTGG
 ATTCCATGCTTCTCAGTGTGGCTTTTGTACTCCTGGAATGTGCATGTCTAT
 TCTTCTCGGCTCTT-----CGGCCAGAGGCC
 CCTCTGGGATTCTCCAAGCTTAAA-----
 ----GGCTATTGCAGGAAATCT-----CTGT
 CGCTTTTGCAGCAGATGTTGATATGGAGGATTTGGGGTTTAATTCTTTT
 TGGAGAAAGGGA---GATAGTAATGAAGTGAAAATAAGCAGCTTACCCCTT
 GTATAACCATAATGATAAGATTAC-----
 -----TCTAGAAGATACTCC-----
 -----TGGAATAATCCTGTAGCCTTGAGGAACCTCAAAG
 CTTATTGGAAAGC-----
 -----TATGACAAA
 TATATTGATCTAAGATATATCCCTGAACCTCTCAATGATTAGAAGGGATAA
 TAATGGGATAAAAAATTGGAGCAACTGTAACAATTTCTAAAGCTATTGAAG
 CTCTAAGGGAATACAGCAAAGGAGAC-----GTC
 TATAAAAAAATTGCTGACCATATGGAAAAAATTGCTTCAGGGTTCATCCG
 GAACTCAGCTAGTCTAGGTGGAATCTTGTGATGGCACAAAGGAACCAT
 TTCCATCTGACATTGCTACAGTACTTCTTGCAGTAGGTTCAACAGTAAAT
 ATAATGAATGGTCTCAAAGTGAAGAGCTTACATTGGAGGAATTC-----
 -----CAAAAGTATACTCTTAAGTGCAAAATC
 CT-----ATTACAGGCATTTCTTCCGGGGCCCAAGATGAAGTTA
 CTGTTTGAACGTATCGAGCTGCACCACGACCCCTAGGAAATGCATTGCC
 CTATTTAAATGCTGCTTTAATGGCTGAAGTTTTCTATTGTAAACTTCT-
 --AATGGGATTATTATAAGTAGCTGCCAGTTTGTCTTTGGTGCTTATGGG
 ACTAAACATCCAATAAGAGCTGCAAAAGTTGAAGAATTTTAACTGGAAA
 AATGCTAAGTGTTGGTGTCTATATGAAGCTATTAATAGTTAGAGGCA
 TTGTGGTACCTGATGATGGCACTTCAAGTCCAGCTTACAGGGCAAGCTTG
 GCTGTCAGTTTTCTTTTGAAGTTCTTTAGCCATTTGGTTGAGCCT-----
 -----AATCCTGAATCTCATG
 ATGCTTCTGAACTAAAAAGAATTAGCAACCAGTTAGATCATGGTAAATC
 CCAACCTTGCTATCTCTGCAAAAGCAGGTAGTGGAATTGAATAGACAGTA
 TCATCCAGTTGGTGAGCCAATTGCAAAATCTGGAGCTGCACTCCAAGCTT
 CTGGTGAGGCTGTTTATGTGGATGACATTCTTCTCCAATGAATTGCCTA
 CACGGGGCTTTTATTTATAGTACGAAGCCTTACGCTCGAGTAAAGGGTAT
 CAAATTTAAGCCTAAATCACTGCCAGATGGAGTCAGTTCACTTATTTCTT
 TCAAAGATATCCCA-----GGGGAGAACATAGGC---TCTAAACTATA
 TTTGGTATTGAGCCTTTATTTGCTGATGATTTTACTCGATGTGCCGGACA
 GTATATTGCTTTTGTGGTTGCAGACACACAGAAACATGCAGATATGGCTG
 CAAACCTTGCAGTGGTTGATTATGATGTGGGAAATTTGGAAGTGCCTATT
 TTATCTGTAGAAGAGGCTGTTAGGAGATCTAGCTTTTTTGAGGTTCTTCT
 TATCTTGAACCCAAAAAGGTTGGTGATTTCTCGAGAGGAATGGCTGAAG
 CTGATCACAAGATTCTCTCTGCTGAGATCAAACCTGGGTACAGTACTAT
 TTCTATATGGAGACACAACTGCCCTTGCAATTCCAGATGAAGACAACCTG
 CATTGTGGTTTACAGTTCAATACAATGCCCTGAATATGCACATAGTACTA
 TTTTCGAGATGCTTGGTATTCTGAACATAATGTTCTGTGTATCACAAAGA
 AGGGTTGGAGGTGGTTTTGGTGGAAAGGCCATAAGAGCAATGCCTGTTGC
 TACAGCATGTGCACTTGCAGCATACAAATTACGCCGCCCTGTCCGGATAT
 ATATGAATCGCAAAACTGATATGATAATAGCAGGAGGAAGACATCCAATG
 AAAATAACCTATAGTGTAAGGATTCAAATCTGATGGCAAGATTACTGCCTT
 ACACCTTGATATATTAATCAATGCGGGGAATAGCTGCAGATATAAGCCCAA
 TTATGCCTCACAACTTCTGGGTGCCCTTAAAAAGTATGATTGGGGTGCT
 TTATCTTTTGATATAAAAGTTTGCAAAACAAACCATTCGACTAAATCAGC
 AATGCGTGCTCCTGGGGAGGTACAAGCTACATTTATTTCTGAAGCTGTAA
 TTGAACATGTAGCATCTACTCTTTCCATGGATGTGGACTCTGTGAGAAGC
 AAAAACCTCCATACTTTCAACAGCCTCAAATTTTCTATGAGGGCAGTGC
 AGGTGAACCTGTTGATTACATTTACCTTCAATCTGGGATAAGTTGGCTA
 GTTCTTCAAGACTTAAACAAAGAACTGAAATGATAAAGCAGTTTAATATG
 TGTAAATAAATGGCAGAAAAGAGGTATTTCTCAAGTACCGATTGTGCATGA
 AGTTTCATTGAGGCCAACTCCAGGGAAAGTAAGCATTCTTAGTGATGGAT
 CTGTTGCTGTTGAAGTTGGAGGGATTGAGCTAGGCCAGGGGCTGTGGACC
 AAAGTGAAACAGATGGCTGCATTTGCTCTCAGCTCAATTCAGTGTGATGG
 AATGGGAGACTTCTTGGAGAAAGTAAGGGTCATACAATCAGATACATTAA

GCTTAATACAGGGGGTTTTACTGCTGGAAGCACTACATCAGAGTCCAGC
TGTGAAGCTATTAGACTCTGCTGCAATATTTTGGTTGAAAGACTAACTCC
TACCAAGGAAAGATTG - - - CAGGAGCAAATGGGTTCTGTTGAATGGGGCA
CACTGATTCTTCAGGCACAAAGTCAAGCTGTGAACCTATCAGCAAGTTCT
TACTATGTCCCTGACTTTTCTTCCATGAAATACCTAAACTATGGTGCTGC
TGTAATGTCACAGGTGGAGGTAATCTTCTCACTGGAGAAACCACCATTT
TGCAATCTGATATTATATATGACTGTGGGCAGAGCCTCAACCTGCTGTG
GATTTAGGACAGATAGAAGGAGCTTTTGTCCAAGGGATTGGGTTTTTAT
GCTCGAGGAATACACAACATAATTCAGAGGGACTGGTGGTCACTGAAGGCA
CATGGACATACAAGATCCCAACAATTGACACCATAACCAAGCAGTTCAAT
GTTGAGATACTGAACAGCGGACATCATACAAAACGTGTTCTTTCTTCTAA
AGCTTCTGGTGAGCCACCATTACTTCTAGCAGTTTCAGTTCACTGTGCTA
CAAGGGCAGCTATTAGAGAAGCCGACAACAGCTTCTTCTGAGACTGGG
CTATGCAAGTCTGATTTGACATTCCAGTTGGAGGTCCCGGCAACAATGCC
TGTTGTGAAGAATTTGTGCGGACTGGAGAATGTGGAGAGCTACTTACAAA
GCTTGCTCTCT - - - - -

ABA.ABA2.fasta

>A

TCAACGAACACTGAATCTTCTTCTTATTCTTCTCTTCTAGTCAAAGGCT
TTTGGGTAAAGTGGCATTGATCACTGGAGGAGCCACAGGGATAGGTGAGA
GCATTGTTCTGCTGTTCCACAAGCACGGTGCCAAAGTCTGCATTGTTGAT
CTGCAAGAT - - - CTCGGAGGTGAGGTGTGTAAGTCTGCTTCGTGGTGA
GTCCAAGGAGACGGCTTTTTTTCATCCATGGCGATGTTAGAGTGGAAGATG
ACATTAGCAATGCGGTTGACTTTGCAGTCAAAAATTTTGGGACGCTTGAT
ATACTTATCAACAATGCAGGATTATGTGGAGCACCGTGCCCTGATATTCG
TAATTATAGTTTGAGTGAGTTCGAGATGACCTTTGATGTGAAT - - - GAAG
GAGCTTTTCTAAGCATGAAACATGCAGCTCGTGTAATGATACCGGAGAAG
AAAGGGTCGATAGTTTCTTATGTAGTGTGGGAGGTGTTGTGGGAGGCGT
TGGTCCACATTCTTATGTTGTTCCAAGCATGCTGTTCTAGGCTTGACTA
GGAGTGTGTCAGCGGAGCTTGGACAGCACGGGATACGTGTGAAGTGTGTT
TCGCCTTACGCGGTTGCAACTAAACTCGCTTTGGCTCATTTGCGGAGGA
AGAAAGAACGGAGGATGCATTTGTTGGTTTCAGGAATTTTGTGCTGCAA
ACCGCAATCTAAAAGGGGTGGAAGTGCAGGTTGATGATGTAGCGAACGCT
GTTCTGTTTTTGGCTAGCGATGACTCGCGGTACATAAGCGGA - - - GATTT
GATGATTGATGGAGGATTCACTTGCACTAACCCTCCTTTAAAGTCTTCA
GA - - -

>P

- - - - -GCCACCGCAAGCTCCACTGACTCATCCCTCTCTCCTCTAAGGTT
ATTGGGCAAGTAGCATTG - - - GCAGGTGGAGCTACCGGTATTGGGGAGA
GCATTGCCCGGCTATTTTGCAGACATGGTGCAAAAGTTTGATAGCTGAT
TTGCAAGACAACCTTGGACAAAATGTTTGCGAGTCGCTT - - - - -GGTGG
TGAGCCAAAACTTGTTATATC - - - CATTGTGATGTCACAATAGAGGATG
ATGTTCCGCAAGCTGTTGACTTTACTGTTGATAAATTTGGAACATTGGAT
ATTATGGTTAACAATGCTGGGCTGGGAGGTCCCCCTGTCCAGATATCCG
TAAGGTGGCCTTATCAGACTTTGAAAAGGTGTTGATGTAATGTGAAGG
GTGTTTTTCTTGGAATGAAGCATGCAGCACGGATTATGATCCCCCTAAAT
AAGGGTTCAATTGTTTTCTCTCTGTAGTGTGCCAGTGCCATAGGGGGTAG
C - - - CCACATGCTTACACAGGGTCAAAGCATGCTGTATTGGGTCTGACCA
GAAGTGTGTCAGCTGAAGTAGGAAAACATGGGATACGCGTAAACTGTGTC
TCTCCTTATGGAGTTGCGACAAGCTTGGCTGTGGCTCATTTGCCCGAGGA
TGAGAGGACTGAGGATGCCTTAATAGGTTTCCGCTCTTTTATTGGGAGGA
ATGCTAAGTTGCAAGGTGTGGAATTGACAGTTGATGATGTGGCCAATGCC
GTGCTCTTCTAGCAAGTATGAAGCAAGGTACATAAGTGGGGATAATCT
AATGCTTGATGGGGGCTTACGTGCACAAACCACTCACTCCGTGCTTCA
GA - - -

>R

- - - - -GCCACAACAGCTCCGTTGCTTCAACTCTCTCTTGCCAAAGGTT
ATTGGGGAAGTAGCTTTGGTAACAGGCGGATCGACTGGAATAGGAGAAA
GCATTGTCCGTCTATTTTACAAACATGGTGCAAAAGTCTGCTTAGTTGAT
TTGGAGGACAACCTAGGGCAAAATGCTGCGAATCCCTT - - - - -GGTGG
TGAACCAACATTTGCTATTTT - - - CACTGCGATGTCACAGTAGAAGATG
AGGTTCAACGTGCAAGTTGAGTTCACTGTTGATAAATTTGGAACACTGGAT
ATTATGGTCAATAATGCAGGTTTGTGAGGTCCACCTTGTCCAGACATCCG
GTATACGGAAATTATCCGACTTTTCAAGAGGATTTTGTGTAATGTGAAGG
GCACGTTTATTGGAATGAAGCACGCAGCGAGAATTATGATCCCACTAAAT
AAGGGTTCAATTATTTCTCTGTAGTGTGGCAAGCACCATTGGAGGCTT
AGGCCCTCATGGATATACCGGGTCAAAGCATGCTGTTTTGGGACTGACCA
GAAATGTTGAGCTGAGCTAGGAAAACATGGGATACGTGTGAAGTGTGTT
TCTCCATATGCAGTTCCAACAACTTGGCCTTGGCGCACTTGCATGAGGA
TGAGAGAACTGAGGATGCCATGGCAGGTTTTAAAGCATTTGCCAGGAAGA
ATGCTAATTTACAAGGGGTAGAATTGACAGCTGATGATGTTGCTAATTCT
GTTCTCTTCTTGCGAGTGAAGAATCAAGATACATTAGCGGGGAAATCT

AATGATAGATGGTGGTTTTACTAATTCAAATCACTCGCTCAGAGTGTTCA
GA ---
>0
-----TCCGCCGCCGCCGCCGCCGCCGATCCTCCCCGCTCCCCGGTT
GGAAAGCAAGGTTGCGCTGGTTACCGGTGGTGCTTCAGGTATTGGTGAAG
CAATTGTTGCGCTCTTTAGAGAGCATGGTGCAAAGGTATGTATTGCAGAT
ATCCAAGATGAAGCAGGTCAGAAAGCTCCGGGACTCCCTT-----GGAGG
TGACCAAGATGTCTATTGTGTC---CACTGCGATGTTTCGGTGGAAGAGG
ATGTAGCCCGAGCGGTCGATGCAACAGCTGAAAAGTTTGGTACTCTTGAC
ATCATGGTCAACAATGCTGGCTTTACAGGCCAGAAAATCACAGATATCCG
AAACATCGACTTTTCTGAAGTCAGGAAGGTAATCGACATCAATTTAGTTG
GTGTATTCCACGGGATGAAACACGCAGCGCGCATCATGATCCCCAATAAG
AAGGGGTCATCATCTCATTGGGAAGTGTCTTAGTGTCATTGGAGGGTT
GGGACCTCATTCATACACAGCAACCAAGCATGCTGTGGTGGGTCTAACCA
AGAATGTAGCTGGGGAATTGGGGAAGCATGGGATACGCGTGAAGTGCCTA
TCTCCCTATGCAAGTCCACAGGCTCTCTCCATGCCGTATCTGCCACAGGG
CGAGCGCAAGGATGATGCCCTGAAAGACTTTTTGCGCTTTGTTGGTGGTG
AAGCAAACTGAAAGGTGTGGATCTGCTACCTAAGGATGTTGCTCAAGCA
GTGCTCTACTTGGCAAGCGATGAAGCGAGGTACATCAGCGCGCTCAACCT
CATGGTGGATGGTGGCTTTACCTCTGTGAATCACAATTTGAGAGCATTTG
AAGAT
>V
-----CCTATGCTTAATAGAGTATT
ATTAGGGAAAGTGGCGCTGGTCACTGGTGGAGCCACTGGAATTGGAGAGA
GTATCGTGCGTCTATTCTTAAGCAAGGTGCAAAAGTTTGATTGTTGAT
GTGCAAGATGACCTTGGCCAGAAGCTCTGTGACACGCTT-----GGCGG
TGACCCAAATGTAAGTTTTTTC---CATTGTGATGTCACAATAGAGGATG
ATGTTTGGCATGCAAGTGCAGTTTCAAGTAAATTTGGCACACTTGAC
ATTATGGTCAACAATGCTGGCATGGCAGGCCACCTTGTCTGATATTG
TAACGTGGAAAGTATCAATGTTTGAGAAGGTATTTGATGTAATGTAAGG
GTGTCTTCTTGAATGAAGCATGCGGCTCGTATCATGATCCCATTGAAG
AAAGGAACAATCATATCCTTATGCAAGTGTTCGAGTGCCATAGCAGGTGT
TGGCCACATGCATACACAGGTTCTAAGTGTGCTGTTGCTGGCCTGACAC
AGAGTGTTCGCGCGGAGATGGGGGACATGGGATACGAGTCAACTGCAATT
TCACCCTATGCAATTGCTACAGGCTTGGCTTTGGCTCACCTGCCTGAGGA
TGAGAGGACGGAGGATGCCATGGCTGGTTTCCGTGCTTTCGTAGGAAAAA
ATGCAAACTTGCAGGGGGTGAATGACAGTTGATGACGTTGCTCATGCT
GCAAGTCTTCTTAGCTAGCGACGAGGCGAGGTACATAAGCGGCCTTAATCT
CATGCTTGATGGGGGCTTCTCTGCACAAATCACTCGCTTCGGGTCTTCA
GA ---

ABA.NCED.fasta

>A
-----CAAACTCTCTTCGT--
-----TCTGATCTTCTCCGACGAAGACTTCTCCTCGTTCTC
ATTTACTTCCACAACCCAAAAATGCAAAATTTTCTCGACGAATTCTCATT
AACCCCTTTC---AAGATACCGACACTT-----CCTGA
TCTCACTTCTCCGGTTCCGTCAACGGTTAAGCTCAAACCAACGTATCCAA
AC-----TTAAACCTTCTTCAGAAGCTA
GCGGCTACGATGCTCGACAAGATTGAG---TCCTCTATCGTTATTCCTAT
GGAGCAGAATCGCCGCTTCTAAACCGACCGACCGCGGTTCAATTAT
CAGGTAACTTCGCTCCGGTTAATGAATGTCCGGTTTCAGAACGGTTTAGAA
GTGGTTGGTCAGATTCTTCTTGTCTAAAAGGAGTTTACATCCGTAACGG
TGCAAACCTATGTTTCCGCCGTTAGCCGGACATCATTATTTGACGGTG
ACGGAATGATTACGCCGTTAGTATCGGTTTTGATAACCAAGGTTAGTTAC
AGCTGCCGGTACACTAAAACAACCGGCTTGTTCAAGAAACCGCGCTTGG
ACGATCGGTTTTCCCTAAACCAATCGGCGAGCTTACAGGCCATTCCGGTC
TAGCTCGACTCGCTCTCTTACGGCTCGAGCTGGGATCGGTCTAGTGGAC
GGGACACGTGGCATGGGCGTAGCTAACGCCGGTGTGGTTTTCTTAAACGG
CAGGTTATTAGCCATGTCAGAAGATGATCTTCCTTACCAAGTGAAGATCG
ACGGTCAAGGAGATCTTGAGACGATCGGACGGTTTCGGATTTCGAC---CAG
ATTGACTCTTCAGTGATAGCGCATCCTAAGGTGGACGCGACACAGGAGA
TCTCCATACACTGAGCTACAACGTTTTGAAGAAACCTCATCTCAGGTATC
TTAAATTCACACGTCGCGGAAAAAGACACGTGACGTGGAGATCACGCTC
CCTGAACCAACGATGATTGATGATTTTCGCGATAACCGAGAATTTTGTCTG
TATACCGGATCAGCAAATGGTATTCAAATTATCCGAAATGATTCGGGGCG
GGTCAACCGTTATCTACGTTAAAGAAAAATGGCGAGATTTGGAGTTTTG
TCAAAGCAGGATCCCGG---TCGGATATAAATTGGGTTGATGTACCGGA
TTGTTTTCTGTTTCCATCTATGGAATGCGTGGGAAGAGAGAACCGAAGAGG
GAGACCCAGTTATCGTCGTAATCGGGTCATGTATGAGCCACCCGACACG
ATCTTTAGTGAATCAGGAGAACCAACCGGGTTGAATTAAGTGAGATCCG
GTTAAACATGCGTACAAAAGAAATCGAACCGTAAGGTTATCGTA-----
--ACCGGAGTGAATTTAGAAGCGGGTCACATAAACCGTAGTTACGTGGG

-----GAAGGAGAAGAAGACAAAGGTTATGTAATGGGGTTTGTGA
GAGACGAAGAGAAAGACGAGTCGGAGTTTGTGGTGGTGACGCGACGGAT
ATGAAGCAAGTCGCGGCGGTGCGCTTGCCGGAGAGGGTACCTTATGGTTT
CCATGGAACGTTCTGTGAGCGAGAATCAGTTGAAGGAACAAGTTTTC--
>P

This image shows a full page of primary-ruled paper. It features ten sets of horizontal dashed lines, each set consisting of three parallel lines. These lines are evenly spaced vertically across the entire page, providing a guide for letter height and placement in handwriting practice. The background is white, and there are no margins or other markings present.

-----GCTCGAGCTGCAGTCGGCTTTGTTGAT
GGGACACGTGGCAGCGGTGTAGCCAATGCCGGGCTTGTCTATTTTAACGG
TAGATTACTGGCGATGCTCAGAGGATGACTCTCCTTTATAATGTATAAGTCA
AGAGTGACGGCGATATGGAAAGCATCGGAGCGTTCAATTTTGTATGATCAG
CTTGATTGTCCCATGATTGCACACCCTAAGGTGGACCTGTAACTGGTGA
GCTCCATGCATCTTAGTTACAACGTTATCAAGAAACCCATCTCAAAGTACT
TTAGATTTTATGATGCATGCGCTAAGAAGTCAATGACTTTGAGTATTTACATTT
GACCAAGCCAACTAATGATCCACGACTTTGCAATCACAAAAGATTTATGGT
GATCCCGGATCACCAAGTCGTGTTCAAGTTATCGAGATGATTCGAGGCG
GGTCAACCCGTAAATTTATGATCAAAGCAAGATTTACGGTTTGGAGTCTTG
TCAAAAAAAGCTGTCGATGATTTCAAGAATTCAAGTGATGACGTTCCGGA
TTGCTTTTGTTCATTTGTTGCAATGCATCGGAAGAGAAATCTAGTAGCA
GAGACAAGTATTATGTTGTTATCGGGTCGTGCATGGACCCGCCGTATTC
ATCTTCAACCAATCTGAACACCCGCTTCGAAGTGAGTTGTCTGAAATCCG
TTTGAATCTGAGGACAGGAGAGTCAACTCGAAGGGTTATTGTC-----
-GGGGTATGAATTTGGAAGCGGGTCAAGTAAACCGAAGGTTTCTAGGT
CAAAAACCCCGTTTCGTTTCTACTAGCAATTGCAGAGCCTTGGCCAAAGTG
TAGTGAATTGCAAAAGTTGATTTGGAGACTGACGAGGTGACCAAATTTA
TTTATGGTGCTGGTAGGTTTTGGTGGTGAGCCATGTTACGTGCCCAAAAT
-----GGAAATGTTGGCGA
TAATGGAAGAAGTGATGATGACGGCGAAGGCTTCATAATGGGTTTCGTGA
GAGATGAGGAAAAGGGGAGGTGAGAATTTGGTGATAGTGAATTCATCGAGC
ATGATGCAAGTAGCTTCAGTGAAAAATGCCCACTAGAGTGCCCTATGGCTT
CCATGGTACTTTTGTGACGGAAGCTGATTTAAACAACAATCTGTG---
>R

63

ATCTTCAATGAATCCGAAACAGCAATTCAAAGTGAACATCAGAGATTG
 CTTAAACATGAGGACCCGAGAGTCAACCCGGAAGCCATAGTC-----
 --AGGGGCATGAATTTAGAAAGCGGGTCAAGTAAACCGAAAGTTTGTGGGT
 CAGAAAAACCCGGTATGTATACTTGGCAATAGCGGAACCATGGCCGAAGTG
 TAGTGATTGCTAAGGTTGATTGGAGACTAAGGAAGTGAATAAATTTA
 TGTATGGTAAAGGAAGGTATGGAGGAGAGCCCTTTTTCGTGGCAAAAAAT
 AGAGAAGGAGAAGAAAGAAGATGATGATGATGATGGTAAGAAAAATTGG
 CAATGTTATTAATGATGAAGAAGGTGAAGGATACATTATGGGTTTTGTAA
 GAGATGAGATAAAGGAAAGGTCTGAGTTGGTGATAGTGAATGCAACAGGC
 ATGGAGCAAGTAGCTTCCGTAAGTTTGCCCACTCGAGTGCCTTATGGATT
 TCATGGTACTTTTGTAGTGAAGATGAATTAAGGACAGGCTTTGTCT
 >0
 GCGACGATCACGACGCCAGGATATGCTCACATACAGCGGCAGCACGGCAG
 GTGCTCGACGACGGCGGGAAGGCGTGGGGCGTCCAATTCGGTGAGATTCT
 CCGCGCGCGCGGTAGCTCCGTGCCGACGCGGCGGCGGCGTCATCGGCG
 CCGGCGTTCTCGCGGTGCCGTTCTGTG-----CCTGG
 GGCCGACGACCGTCCGCGTGGGGAAGAGTGCCATTGGCGTCCCGAAGG
 CGCCGAGGAAGGGGAGGAGGGGAAGAGGCTCAACTTCTTCCAGCGGCC
 GCGGCGATGGCGCTCGACGCTTCGAGGAGGGGTTTGTGGCGAATGTCCT
 CGAGCGCCCGCACGGGCTGCCGAGCACGGCCGACCCCGCGGTGCAGATCG
 CCGGCAACTTCGCGCGGTCGGTGAGACGCCGCGCGCGCGCTGCCG
 GTGTGCGGGCGCATCCCGCCCTTCATCAACGGCGTCTACGCGCGCAACGG
 CGCCAACCCGCACTTCGACCCCGTCCGCGGACACCTGTTGACGGCG
 ATGGC---GTGACGCGCTCAGGATACGAACGGCGCCGCGAGTCGTAC
 GCGTGCCGGTTACGGAGACCGCGGCTGCGGCAGGAGCGCGCATGGG
 GCGGCCCATGTTCCCAAGGCCATTGGGAGCTCCATGGCCACTCCGGCA
 TCGCGCGCTTGTCTGTTCTACGCGCGCGCGCTGCGGCTCCTCGAC
 CCGTCACACGGCACCGCGTCCGCAACGCCGCGCTCATCTACTTCAACGG
 CAGGCTCCTCGCCATGTGCGAGGACGACCTCCCTACAGGTGCGCGTCA
 CCGCGGACGGCGACCTCGAGACCGTCCGCGCTACGACTTCGACGGGCG
 CTCGGCTGCGCCATGATCGCGCACCCCAAGCTCGACCCGGCCACCGGAGA
 GCTCCACGCGCTCAGTACGACGTGATCAAGAAGCCGTACCTCAAAGTACT
 TCTACTTCGCGCCGACGGCACCAAGTCGGCCGACGTCGAGATCCCGCTC
 GACCAAGCCACCATGATCCACGACTTCGCCATCACGAGAATACGTGGT
 GGTACCCGACCAACAGGTGGTGTTCAGCTCCAGGAGATGCTCCGCGCG
 GCTCGCCCGTGGTGCTCGACAAGGAGAAGACGTCGCGGTTGCGGGTGCTC
 CCCAAGTCACGCCGCGGACGCGTGGAGATGGTGTTGGTGACGTCCCGGA
 CTGCTTCTGCTTCCACCTCTGGAACGCGTGGGAGGAGCGGACACCGACG
 AG-----GTGGTGGTGATCGGCTCGTGCATGACCCCGCGGACTCC
 ATCTTCAACGAGTCCGACGACCGCTCGAGAGCGTCCCTACCGAGATCCG
 CCTCAACACCCGACCGGCGAGTCGACGCGCGCGCCATCCTGCCGCGT
 CGAGCCAGGTCAACCTCGAGGTGGGCATGGTCAACCGCAACCTCCTCGGC
 CGCAAGACGCGGTACGCCCTACCTCGCCGTGGCCGAGCCGTGGCCAAAGGT
 GTCGGGCTTCGCCAAGGTGGACCTCGCCACGGGTGAGCTACCAAAGTTG
 AGTACGGCGAGGGCCGGTTCGGCGGCGAGCCCTGCTTCGTCCCGACGCC
 -----GC
 CGCCGCCACGCCCCGCGGCGAGGACGACGGCTACATCCTGTCTTCTGTC
 ACGACGAGCGCGCGGGACCTCCGAGCTCCTCGTCTGTAATGCCGCGAC
 ATGCGCCTTGAGGCCACCGTGCAGTGCCGTCCCGGTGCCGTACGGCTT
 CCACGGCACGTTATCACGCGGACGAGCTCACCAACCGAGCC-----
 >V
 -----CAAGCTTCTCCTGGCTT
 CTCCTCCACCGCATCACCTCTGCTCCTGAATCACCTGCTCCTACTAAAC
 TTCACCCACCGCATCC-----AACTTCTCATCAAACATCAAAC
 CCACCACTC---CCCTCAAACCTATT-----GCTCC
 ATTAACCTCTCCCTTGCAACT-----
 -----AAATGGCCTTCTCACCTCAACCCACTTCAGAAACTC
 GCTGCCTTGCCCTAGACATGGTTGAG---TGCTTCTGATCCACCGATT
 GGACAACAAACACACCTCCCCAAGCCCTCGACCCCGCTGTTCAACTTG
 TTGGCAACTTCGCCCCAGTTGCGGAGTGCTGTTTACGACGACCTCCAG
 GTGCTGGGTGAGCTTCCGCCAGCTCTCCGTGGCGTCTACTTGCAGAACGG
 CGCAATCCATGCTCTCTCAGCCGGTGGTACCACCTCTTCGACGGTG
 ACGGCATGATACATGCTGTTACCTTGGGCCAGGCAACAGAGCCAGTAC
 AGCTGCCGGTTACCCGCACGAGCCGGCTGGTGACGGAAGCCGATTAGG
 AAGGCCGCTGTTTCCAAACCGATCGGTGAGCTGCATGGCCACTCTGGGA
 TAGCTCGGCTGGCGCTGTTCTACGCCGCGCGGCAAGGCTTGGTGAT
 GGATCACGAGGACAGGTGTGCAAAATGCTGGCCTCGTCTATTTCAACGG
 TCGGCTTCTCGCCATGTCAGAGGATGACCTGCCCTACCAGTGAAGATCA
 AAGGCGATGGAGACCTGGAGACCACGGACGCTTCGACTTCTCAGGACAG
 ATGGACCTCCGATGATAGCGCATCCCAAGGTGGACCCATAACAGGGGA
 GCTTTTCTCTCAGCTACAACGTGGTGAAGAAGCCCTACCTCAAAGTATT
 ATAAGTTCCGACGTCGCGGAGAAAAGTACGTGAGGTATCGATTTCCCTC
 CAACAGCCACGATCATCCAGACTTCGCTTTAACGGAGACCTCCGTGGT
 GATCCAGACCAACAGGTGGTTTTCAAACATCGGAAATGGTTGCGGCG
 GCTCACCTGTTATCCAGACCCGAATAAGATCTCCCGCTTCGAGATCCTA

GTTGCATTTTCATGAAAGTAGTCGGTGTACTAGCTTGACAATGTTGCTTA
TTCCTTCAGCCAATGCTGTTGATGCTCTCAAACATGTACTTGCTTACTG
AAGGAATGCAGGTTGGAACCTCGCCAAGTGCATTGCAAACCCATCTTGCGC
AGCCAATGTTGCTTGTCTCCAGACCTGCAACAACAGACCTGATGAGACTG
AATGCCAG-----ATTAAATGTGGAGACCTTTTTGAAAACAATGTG
GTAGATGAATTCACGAGTGTGCTGTCTCACGAAAGAAATGTGTGCCTAG
AAAATCTGATATAGGAGAATTTCTGTCCAGACCCAGCTGTTCTTGTTA
AGAACTTTAACATTGCAGATTTTCAGTGGGAAGTGGTTCATAACTAGTGGC
CTAAATCCTACATTTGATACTTTTGATTGCCAATTGCATGAATTCACAC
AGAGTCCAAC---AAACTTGTGGGAAATTTGTCATGGCGAATACGAACCTC
CAGACGGTGGGTTTTTACCCGATCAACAATGCAGAGATTTGTACAAGAT
CCAAACAGCCTGGAATACTCTACAATCACGATAATGAGTATCTTCACTA
CCAGGATGACTGGTATGTTTTGTCATCCAAGATAGAGAACAACAGATG
ACTACGTATTTGTATACTATCAGGGAAGGAATGATGCATGGGATGGATAT
GGCGGTGCTGTTGTGTACACAAGAAGTCAACTTTGCCTGAAAGTATTGT
ACCAGAACTAGAAAAGGCGGCTAAAAGTGTAGGACGAGACTTCAACAAT
TCATTAGAACAGACAATACTTGTGGCCTGAGCCTCCTTGTGTGAGAGA
CTTGAAAAGACTGTGGAGGAAGGAGAGAAGACAATTATAAGGGAGGTTGA
AGAAATTGAAGGGGAG-----TTAGGGAAGACTG
AAAAGACCTTTTTTGAACATTGATGGAAGGATTTAAAGAGCTCAGAAAA
GATGAGGAATACTTCTTAAAGGAATTGAGCAAAGAGGAGATGGATCTTTT
GAGTGAACATAAAATGGAAGCAAGCGAGGTAGAAAAGCTCTTTGGACGAG
CACTGCCACTGAGGAAGCTAAGA

ABA.Zep.fasta

>A
---GGTTCAACTCCGTTTTGCTACTCTATCAATCCATCTCCATCAAAGCT
TGATTTTCACGAGGACCCATGTGTTAGTCCTGTTTCTAAACAGTTTTACT
TAGATTTATCATCG-----TTTTCCGGAACCC
GGAGGAGTATCTGGGTTTAGGAGCCGTCGAGCTTTGCTCGGAGTAAAGGC
GGCGACG-----GCGTTAGTTGAGAAGGAGGAGAAGA
GAGAGCGCGTGACGGAAGAAGAAGAAGAAATCGAGGGTTTTAGTTGCCGA
GGTGAATCGGAGGATTGGTGTGTTGCTTTAGCGGCTAAGAAGAAAGGATT
CGATGTGTAGTGTGTTGAGAAAGATTTGAGTGTATAGAGGAGAAGGAA
AATACAGAGGCCGATTCAAATACAGAGCAACGCT---TCTGCTTTGGAA
GCTATTGTATATTGAAGTTGCTGAACAAGTTATGGAAGCTGGGTGTATCAC
TGGTGATCGGATTAACGGTCTCGTTGATGGTATCTCTGGTACTTGGTATG
TAAAGTTTGATACTTTCACTCCTGCGGCGTCACGGGGACTTCTGTGACT
AGAGTAATTAGTAGAATGACTCTGCAGCAGATTCTAGCACGTGCGGTTGG
AGAAGATGTGATTAGAAAACGAGAGTAATGTTGTTGATTTTGAAGATTCTG
GAGATAAGGTTACTGTGGTACTCGAGAATGGTCAACGCTATGAAGGTGAT
CTGCTTGTGGGTGCAGATGGCATTGTTGCTAAGGTGAGAAATAATTTGTT
TGGCCGTAGTGAAGCTACTTATTCAGGCTACACTTGTACACGGGGATTG
CAGATTTTATACAGCGGATATCGAGTCTGTTGGCTACCGGGTTTTCTTG
GGACACAACAGTACTTTGTTTCTTCGGATGTTGGTGGTGGAAAAATGCA
ATGGTATGCATTTACGAGGAACCAAGCTGGTGGGGCTGATGCTCAAATG
GTATGAAGAAAAGGTTGTTTGAATATTTGACGGTGGTGCGACAATGTA
CTCGACTTGTGTCATGCGACTGAGGAGGAAGCCATTCTGAGAAGAGATAT
TTATGATAGAAGTCTGGTTTTACTTGGGGTAAAGGGCGTGTACGCTGC
TCGGGGATTCTATCCATGCGATGCAGCCAAATATGGGTCAAGGTGGATGC
ATGGCCATTGAGGATAGTTTTCACTAGCATTGGAGCTTGATGAAGCATG
GAAACAGAGTGTTGAAACGACTACACCTGTTGATGTTGTTTCTCTTTGA
AAAGATATGAGGAATCAAGA---CTGAGAGTCGCTATTATCCATGCAATG
GCGAGGATGGCTGCAATTATGGCTTCCACTTACAAAGCATACTTAGGTGT
TGGGCTTGGTCTCTGTCTTTCTTGACAAAGTTAGAGTACCACATCCAG
GAAGAGTTGGTGGTAGATTCTTCGTTGACATTGCTATGCCATCGATGCTT
GACTGGGTCTTGGAGGTAACAGTGAAAACTCCAAGGAAGGCCACCTAG
TTGCAGACTCACTGACAAAGCCGATGACCGGCTTCGAGAGTGGTTTGAAG
ATGACGATGCTCTTGAACGTACTATAAAGGGAGAATGGTATCTAATTCCA
CACGGCGAGATTGTTGCGTTTTCGGAAACATTATGTCTAACCAAGATGA
AGATCAACCTTGCATCGTGGGAAGCGAACCAGATCAAGATTTTCTGGAA
TGCGCATTGTGATCCCTTCGTCTCAGGTTTGAAGATGCATGCTCGTGTG
ATTTACAAAGACGGAGCTTTCTTCTTGATGGATCTTGAAGCGAACACGG
AACCTATGTGACCGATAACGAAGGAAGAAGATATAGAGCAACACCGAATT
TTCCGCGCGGTTTAGATCGTCCGACATCATCGAGTTTGGTTCAGATAAG
AAG-----
-----GCGGCGT
TTAGGGTGAAGTAATCAGGAAACTCCGAAATCGACGAGGAAGAATGAG
AGTAACAACGATAAAATTACTTCAGACAGCT
>P
GCATCTTCAACTTTGTTTTGCAACACT-----CCAACGGCTGTT--
---TTCTCAAGAACCCAGTTTCCAGTTCCAATCTTTAGCAACTCATCCG
TAGAATTCTCGTCTTCCACGCACTATAACTACAACCTTCAAAACCAAGACA

GGTAGT-----GCTAAGAAATTGAAACATGTGAATGC
TGTAGTG-----ACCGAGGCACCTGCTGTGAGTGAAT
CAAGAGGGAAACAGTCAGAACAAAGGAAGTTAAAGGTTCTTGTGCTGGA
GGTGGAATTGGAGGGCTGGTTTTTGCCTTGGCAGCAAAAAGGAAAGGGTT
TGAGGTGATGGTTTTTGGAGAAGGATTTGAGTGCGGTAAGAGGAGAAGGGC
AATATAGAGGTCCAATTCAGATACAGAGCAATGCATTAGCTGCTTTGGAA
GCTATTGATTTAGAGGTTGCTGAAGAGGTTATGAGAGCTGGCTGTATCAC
GGGTGATAGGATTAATGGACTTGTGTGAGGGTTTCTGGTACTTGGTATG
TCAAGTTTGTATACCTTCACTCCAGCAGCAGAAAGAGGGCTTCCTGTACG
AGAGTTATTAGCCGAATGACTTTGCAACAGATCCTAGCTCGTGCAGTTGG
AGATGATGTGATTCTTAATGATAGTAATGTTGTAGTTTTTCAGGATGAAG
GGAATAAGATCACTGTGGTGCTCGAGAATGGACAACAATTTGAAGGTGAT
CTCCTAGTGGGTGCTGATGGAATATGGTCAAAGGTGAGGAAGAACTTATT
TGGACCCAAGGAGGCAGTGTACTCAGGCTATACATGTTATACTGGTATAG
CAGATTTTGTGCTGTGATATCGAGACTGTTGGGTATCGTGTATTTTTG
GGTCACAAACAATACTTCGTGTCTTCAGATGTGGGTGCTGGAAGATGCA
GTGGTATGCAATTTCAACAAGGAACAACCTGGTGGCATGGATGGCCCCGTG
GTAAAAAGGATCGGTTGTTGAAAATTTTTGAGGGTTGGTGTGATAATGTG
ATAGATTTGATACTGGCCACAGATGAAGATGCCATTCTTCGACGTGATAT
ATATGACAGGGAACCCATTCTTACTTGGGGAAGGGGTGCTGTGACCTTGC
TTGGGGATTCTGTCCATGCCATGCAACCGAATATGGGTCAAGGGGGGTGC
ATGGCCATTGAGGATAGCTACCAACTTGCATTGGAGCTTGATAAAGCATG
GAAACAAAGTGTGAATCAGGAACCTCTGTTGACGTTATTTCACTACTAA
GGAGCTATGAGAATGCTAGAAGACTTCGAGTTGCCATTATCCATGGAATG
GCAAGAATGGCGGCAATTATGGCTTCAACATACAAGGCGTATCTGGGTGT
AGGGCTTGGTCCACTGTCGTTCTTGACAAAGTTTCGGATACCTCACCCAG
GAAGAGTTGGTGGCAGATTTTTCTGTTGACATAGCAATGCCTGTGATGCTC
AATTGGGTCTTGGGCGGAAATAGTTCCAAACCTGAAGGCAGGTCTCTAAG
TTGCCGACTCTCAGACAAAGCCAGTGACCAAGTTCGAAGATGGTTGAAG
ATGATGATGCTCTAGAGCGTGCTCTTGATGGAGAGTGGTTTTCTTTTACCA
TGTGGAATGAGGCTGTTGCTTCAACACCTATTGGTTTAAGTAGGGATGA
GAACAAACCTGTGTGGTTGGGAGTGTGTACATGATGATTTTCCGGGAA
TGTCCATAGTGATACCCGACCTGAGGTTTCAGAAATGCATGCTCGTATC
TCTTGTAAAAATGGTGCAATTTTACCCTAATTGATTTGCGGAGTGAGCATGG
TACCTTTATCACAGAT---GAGGGGAGACGGTATAGGGCAACTCCGAACCT
TTCCTGCCCGATTCCATCCATCAGATATGATTGAGTTTGGCTCGGACAAG
AAGGTGACAAATAATCCCTGCCCTGTTTCATAGCTTTTCTTCTGAAATCAG
ATTAAATCCATATTCCAAATGCAATCTAGGAATCAGAACCAACGCAACAT
TTCGTGTTAAGGTGATGAGGTCTCTCCGAAGATTTTCAGAAAAGAAGGAA
GAAAGC-----CAAGTTCTTCGTTCCGTA
>R
GCATCTTCAGCTTTCTTTTGCAACTCAATCAATCCATCAACTTCAGTT--
----TTTTCAAGAACCCATTTCTCGTTTCCAATTTTTAGTACCTCTACCG
TGGAAATTCATCTTTTGACAAATACAACTTTTCATTTCAAAACAAAGAAA
AGTGAC-----CACCAAAACAAGAGATTTACACAAGTCAAAGC
TGTGGTG-----ACTGAGTCACCTACTGTGGCCGAGT
CAAAATGGGAACTTTTCAGAGCAAAAGAACTCAGGATTCTTGTGCTGGA
GGTGGGATTGGAGGGTTGGTTTTTGCTTAGCAGCTAAAAGGAAGGGCTT
TGAAGTATTGGTTTTTGAGAAAGATTTGAGTGTATAAGAGGAGAAGGAC
AGTATAGAGGTCTTATTCAGGTGCAGAGTAATGCATTAGCTGCTCTAGAG
GCTATTGATTTAGAGGTTGCTGAAGAGGTTATGAGAGCTGGGTGATTAC
TGGTGATAGAATTAATGGCCTGGTTGATGGTGTCTGCTGACTTGGTACT
GCAAGTTTGATACCTTCACTCTGCAGCAGAACGGGGGCTTCCAGTCA
AGAGTCATTAGCCGAATGACTTTGCAACAAATCTTAGCGTGTGCAATGG
AGAAGATGTGATTATGAATGCAAGTAATGTTATTTTTCAGGATAATG
AGGATAAGGTACCGTGACGCTTGAGAATGGACAGCAATTCGAAGGTGAT
CTCCTAGTTGGTGCTGATGGAATTTGGTCAAAGGTGAGGAAGAACTTATT
CGGGCCAAAGGAAGCCACATACTCCGGCTACACTTGTTATACTGGCATTG
CAGATTTTGTGCTGTTGATATTGAGTCTGTTGGGTATCGAGTATTTCTT
GGTCATAACAATACTTTGTTTCTTCGGATGTGGGTGCTGGAAGATGCA
GTGGTATGCATTTCAATGAACCACTGGTGGCGTGGATAGCCCCAACG
GTAAAAAGGAAAGGTTGCTGAAAATATTGAGGGTTGGTGTGATAATGTA
ATAGATCTGTTGCACGCCACAGACGAAGATGCTATTCTTCGACGTGACAT
ATATGACAGAGAACCTGTTTTTACATGGGGAAGGGTGCCTGACATTGC
TTGGTGATTCAATTCATGCCATGCAGCCTAATATGGGTCAAGGAGGGTGC
ATGGCCATTGAGGATAGTTATCAACTCGCATTGGAGCTTGATAAGGCATG
GAAACAAAGCATCGAGTCAGGAACCCCGTTGATGTTGTTTTCATCTCTAA
AGAGCTATGAGAGAACTAGAAGACTTCGAGTTGCCATCATCCATGGAATG
GCAAGAATTGGCTGCTATTATGGCTTCGACTTATAAAGCTTATTTGGGTGT
AGGCCTTGGTCTTTGTCGTTCTTGACAAAGTACCGGATACCATCCAG
GAAGAGTCGGTGGGAGATTCTTTATTGACATAGCGATGCCGTGAATGCTC
AATTGGGTCTTAGGTGGTAATAGTTCAAACCTGAAGGAAGACCTCTAAG
TTGCAGACTCTCAGACAAAGCTAGTGACCAATTGCAGACGTGGTTTGAAG
ATGATAATGCGTTAGAGCGTGCTCTTAATGGAGAATGGTTTCTGTTACCA
TTTGAGATGATGCTGTT--CAGGAACCTATCTGTTTAAAGTAGGGATGA

GAACATACCTTGTATGGTTGGGAGTGAATCACAGGAAGACTTTCTCGGGA
 AGTCTATAGTGATTTCTCACCCAGGTCTCCAAATGCATGCGAGAATT
 AGTTACAAAGATGGTGGTTTTATGTGATAGATTTGCAGAGCGAACACGG
 CACCTTTATCACAGATAATGATGGGAGACGATCTAGGGTCCCTCCAACT
 TTCTACCTATTTTCATCCCTCGGAAGCAATAGAGTTTGGCTCTGCTGGG
 AAG-----GCAAAAT
 TCCGAGTCAAAGTGATGAAATCTCCTGCAAAGATTAAGAGAAGGGAGGA
 AAT-----GAAATTCTCCAATCTGTG
 >0
 -----GCGCTCCTCTCCGCGACC-----GCCCGCCCAAGAC
 GCGCTTCTCG-----CTCTTCTCCCACGAGGAGG
 CGCAGCACCCGCAC-----CCGCACGCGCTCTCGGCGTGCTGC
 GGCGGCGGCCAGCGGCAAGAGGCAGCGGGCGGGCCAGGGTGGCGGC
 GGCAATGCGGGCGGCGGACGCGGCCCTCCGTGCGCAGGCGGCGTGC
 CGGGCGGCGGCGGAGGGCACGCGGAGGCGCGGGTGCTCGTGCCGCG
 GCGGCATCGGGGGGCTGGTGTGGCGTGCGGCGAGGCGGAAGGGTA
 CGAGGTGACGGTGTTCGAGCGCGACATGAGCGCGGTGCGGCGAGGGG
 AGTACCGCGGGCCGATACAGATCCAGAGCAACGCGCTGCGGCGCTGGAG
 GCCATCGACATGTCGTCGCGGAGGAGGTGTCGCGCAAGGCTGCGTCAC
 CGGCGACCGCATCAACGGCCTCGTCGACGGCATCTCCGGTCTCTGGTACA
 TCAAGTTTGATACATTTACTCCTGCAAGTGAAGGAGGTCTCCAGTTACA
 AGGGTTATTAGCCGAATGACGCTGCGAGCAGATTCTTGCTGTCGGTTGG
 TGATGATGCTATACTGAACGATAGCCATGTTGTTGATTTTCATAGATGATG
 GCAACAAGGTAAGTGAATTTTGGAGGATGGCCGAAATTTGAAGGTGAC
 CTTTTGGTTGGTGTGCTGATGGAATATGGTCAAAGGTGAGGAAGGTGTTTT
 CGGGCAATCAGAAGCAACTTATTCAGAATATACTTGCTACACTGGCATTG
 CAGACTTTGTGCTCCTGACATTGACACAGTTGGGTACCGTGTATTTCTT
 GGTCACAAACAATTTTCGTCTCCTCAGATGTCGGTGTGAAAAATGCA
 GTGGTATGCGATTTTCATAAGGAACCTGCTGGTGGCACTGATCCTGAAATG
 GTAAAAATAAAGATTGCTCGAGATATTTAATGGTTGGTGGGATAACGTC
 GTTGATCTGATAAATGCAACTGATGAGGAAGCAATCTTCGCGGGGATAT
 ATATGACCGACCACTACTTTTAACTGGGGAAAGGTGCTGTCACTTTGTC
 TAGGTGACTGTGATGCTATGCAAGCAATCTGGGTCAAGGTGGCTGC
 ATGGCTATTGAGGATGGTTACCAGCTGGCTGTAGAACTTGAGAAGTCTTG
 GCAAGGAGAGTGCGAAAGTCTGGAACCTCTATGGATATAGTTTCCTCCTGA
 GGAGATATGAGAAGGAGAGAATACTACGTGTTTCTGTTATACATGGGTTG
 GCAAGAATGGCAGCAATC---GCTACCACTTATAGACCATACTTGGGTGT
 GGGTCTGGGACCATTGTCGTTTTTAACGAAGTTGAGGATACCACATCCTG
 GAAGAGTTGGTGGGAGATTCTTCATCAAGTATGGAATGCCTTTGATGTTG
 AGCTGGGTTCTAGGAGGAAACAGCACGAAGTTAGAAGGAAGACCGTTAAG
 CTGTAGGCTTTCTGACAAGGCAACAGCACAGCTTCGTGATGTTTGAAG
 ATGACGATGCATTGGAACAAGCCATGGGTGGAGAGTGGTACCTCCTCCCC
 ACAAGTTCTGGAGACTCG-----CAACCCATTGATTAATCAGGGATGA
 AAAAAAGTCACTCTCCATTGGAAGCGGGTCAGATCCCAGCAATTCGACTG
 CTTCCCTGGCATTGCCCTTGCCACAGATATCAGAAAACCATGCTACTATC
 ACATGCAAGAATAAGGCCTTTTATGTGACTGATAATGGAAGTGAACATGG
 TACATGGATTACCGACAACGAAGGTAGACGCTATCGCGTACCTCCGAAC
 TCCCTGTCCGTTTCCATCCCTCGGATGCCATTGAGTTTGGTTTCAGATAAA
 AAG-----GCCGTGT
 TCCGTGTGAAGGTGCTGAGCACGCTCCCATATGAATCTGCAAGAGGAGGG
 CCG-----CAAATCCTGCAGGCAGCA
 >V
 GCTTCAGCAGTGTGTTTTATAGTTCAAGT-----CAGCCTTCGATT--
 ----TTCTCAAGAACCACATCCCAATTCCAATTTGAAAGACTCCTTTG
 AAGAGTTTCGGCCACTCTATCAACTATAAGCACTACTTCAGAAGCAACCCA
 TGTGGG-----CAAAAGAAGAGGGTGGCACAGGTCAAAGC
 T-----ACTTTGGCTGAAGCCACACCAAGCGC
 CGCGGCGCGCTGCTGCGAGTAAGAAGGTTGCGATACTTGTGGCCGGA
 GGTGGGATTGGTGGGCTGGTTTTGGCGCTGGCAGCGAAGAAGAAGGGTTT
 TGATGTGGTGGTGGTTTGAGAAAGACATGAGTGCAATTAGGGGAGAGGGTC
 AGTACAGGGGTCCAATTCAGATACAGAGTAATGCATTGGCGGCTTTGGAA
 GCTGTTGATATGGAGGTTGCTGAGGAGGTTATGAGGGCTGGCTGCATCAC
 TGGTGACCGGATTAATGGGCTGGTTGATGGGGTTTCTGGTGATTGGTATG
 TCAAGTTTGATACATTCACTCCTGCGGCAAGCGGGGCTTCCTGTCAAC
 AGAGTTATAGCCGAATGACTTTGCAAGCAATCCTGGCTCGTGCAAGTAGG
 GGAAGATATCATTATGAATGGCAGTAATGTTGTTGATTTTGAAGATGATG
 GAAATAAGGTTACTGTATCCTTGAGAATGGACAACGTTATGAAGGTGAT
 CTTCTTATCGGTGCTGATGGAATATGGTCAAAGGTGAGAAAGGTTTGGT
 TGGGCTAAGGAAGCCACATACTCAGGCTACACTTGTACACTGGCATTG
 CAGATTTTGTGCTGCTGATATTGACTCTGTTGGGTACCGGGTATTTTTG
 GGACACAAACAATACCTCGTGTCTTCAGATGTGGGTGACGAAAAATGCA
 GTGGTATGCATTTTACAATGAACAGCTGGTGGTGGTGTGATGGTCTGAAG
 GTAAGAAGGAAAGGTTGCTTAAATATTTGGTGGCTGGTGTGATAATGTA

ATTGATCTCATACTTGCCACGGATGAAGAAGCAATTCTTCGACGTGACAT
 ATATGACCGGACACCCACCTTTACTTGGGGAAGGGCCGTGTAACCTTGC
 TTGGGGATTTCAGTCCATGCCATGCAACCAAATATGGGTCAAGGGGGATGC
 ATGGCGATTGAGGATAGCTATCAGCTTGCAATGGAACCTTGATAAAGCCTG
 GGAACAAAGCATCAAATCAGGGACTCCTATTGATGTTGTTTCTTGCCATA
 AGAGCTATGAGAAGGCTAGAAGAATACGAGTTGCCGTTATCCATGGAATG
 GCAAGAATGGCTGCAATTATGGCTTCAACTTACAAGGCTTATTGGGTGT
 AGGACTTGGTCCATTGTCGTTTTTGACAAAACCTCAGGATCCCACATCCTG
 GGAGAGTTGGTGGGAGATTTTTTATTGACATAGCAATGCCTTTAATGCTG
 AGTTGGGTCTAGGTGGTAACAGTTCAAACTTGAAGGGAGGCCACCATC
 TTGCAGACTCTCAGATAAAGCAAGTGATCAATTACGGAGATGGTTTGAAG
 ATGATGATGCTCTGGAGCGTGCTATTGGTGGAGAGTGGTTCTTTTGCCA
 TCAGAGAGAGTGGATTA-----CAACCTATTGTTTGAGTAAGGATGA
 GAACAAACCTGCATAATCGGGAGTGTGTCCCATACGGATTTTCTGGAA
 TATCAACAGTCATACCTTACCCAAGGTTTCCAAAATGCACGCTCGAATT
 AGTTGTAAAGGATGGTGCCTTTTTCTTAACCGATTTCAGAGTGAACATGG
 TACCTGGATCACTGATAATGTGGGGAGGCGGCAACGGGTGTCTCCAAATT
 TTCTACTCGATTCCATCCATCAGAAGTGATTGATTTTGGTTCTGAAAAG

 -----GCTTCAT
 TTCGAGTCAAGGTGGTGAGGACTCCCCAGACAATGCAGCTAAGAACGAA
 GAGAGC-----AAGCTCTTCCAGGCGGT

BR.Br6ox.fasta

>A
 ---GGAGCAATGATGGTGATGATGGGTCTTCTTGTATCATCGTGTCTCT
 CTGTTCCGCTCTCCTTCGATGGAATCAGATGCGATAT--ACCAAGAATG
 GTCTTCTCTGGGAACCATGGGCTGGCCAATCTTTGGCGAAACCAACCGAG
 TTTCTCAAACAAGGCCCAACTTCATGAGAAACCAAAGACTCCGATACGG
 GAGTTTCTTCAAATCTCATCTTCTAGGTTGTCCAACGTTAATCTCAATGG
 ACTCAGAAGTAAACAGATACATTTTAAAGAACGAATCAAAAGGTTTGGTT
 CCTGGTTACCCACAATCGATGCTTGATATACCTTGGGACTTGTAAACATGGC
 TGGCGTTTACGGTTTCGAGCCACCGGCTTATGAGAGGCTCGCTTCTGTCTC
 TCATAAGCTCGACCATGATGAGAGATCATATCTTGCCTAAAGTTGATCAC
 TTCATGAGAAGCTATCTTGATCAGTGGAAATGAGCTTGAG-----GT
 TATTGATATCCAAGATAAGACCAAACATATGGCATTTTTATCTTCACTGA
 CACAAATCGCTGGGAACCTTAAGAAAACCATTTGTTGAAGAATTCAAACT
 GCATTCTTCAAGCTTGTGTTGGGACTTTATCCGTTCCGATTGATCTTCC
 GGGCACAAATATCGTTGCGGAATCCAAGCAAGAAATAACATTGATAGGT
 TGTGAAGAGAGCTGATGCAAGAACGTAGAGATTCTGGAGAAACATTACAC
 GACATGTTAGGTTACTTGATGAAG--AAGGAAGGTAACCGATACCCGTT
 AACCGATGAAGAGATAAGAGACCAAGTTGTGACGATTTTGTATTCCGGTT
 ACGAAACTGTCTCTACGACCTCAATGATGGCTCTTAAGTACCTTCAT---
 CACCCAAAAGCTCTTCAAGAACCTAAGAGCTGAGCATTGGCATTTCAGGGA
 AAGAAAACGACAGGACGAACCACTCGGTCTTGAGGACGTGAAGTCAATGA
 AGTTCACTCGAGCTGTGATTTATGAGACATCAAGATTGGCAACGATCGTT
 AATGGGGTCTTAAGGAAAACCTACTCGTGACTTGGAAATCAACGGTTATTT
 AATCCCAAAAGGATGGAGAATTTATGTATACAGGAGGAAATTAATTACG
 ATGCAAATCTTTATGAAGACCCATTGATCTTTAATCCATGGAGATGGATG
 AAGAAGAGCTTGGAGTCACAAAACCTCATGCTTTGTGTTTGGAGGTGGGAC
 AAGGCTTTGTCTGGTAAGGAACTAGGGATTGTGCGAGATCTCGAGCTTTC
 TCCATTACTTTGTTACGAGATACAGATGGGAGGAAATAGGAGGGGATGAA
 TTAATGGTGTTCGAGAGTTTTTGACCAAAAAGGCTTCCATCTTAGGAT
 TTCACCTAC
 >P
 ATGGCAGTTCTCTTGATGGTTCTTGTGGCAGTTCTCTTTTGTCTGTAT
 CTCCTCTGCTTTGTTGAGGTTGAACGAGGTGAGATAT--AGGAAGAAAG
 GGTTGCCTCCAGGTACTATGGGATGGCCAGTCTTTGGAGAGACCACTGAG
 TTTCTAAAGCAAGGTCCAACTTCATGAAGAATCAGAGAGCAAGGTATGG
 GAGTATTTTCAAATCCACATTCTGGGGTGTCTTACCATTGTGTCATGG
 ATCCAGAGCTCAATCGATACATCCTAATGAACGAGGGAAGGGCCTTGTT
 CCTGGTTACCCTCAGTCCATGCTGGATATCTTAGGCAATCGCAACATTGC
 AGCAGTTTCATGGCTCCACTCACAAGTACATGAGAGGGGCATTATTATCCC
 TCATTAGCCCCACCATGATCAGAGAACAACCTTTGCCAACAAATTGATGAG
 TTCATGAGAACCACCTCAGCTACTGGGATACCAAA-----AT
 TATTGACATTCAACAAATGACTAAGGAGATGGCACTTCTCTCTGCACTTA
 AGCAAAATGCTGGCACTGATTCTGTCTCAATATCTCAAGCATTTCATGCCT
 GAGTTTTTCAAGCTGGTTTTAGGCATTTGTCTTCCCAATTGACCTTCC
 TGGCAAAATTCGACAAAGGAGTCCAGGCGAGAAAAAATATTGTACGCA
 TGTTAAGGCAGCTAATAGACGGGAGGAGGGCATCGAAATTATACCACGAG
 GACATGCTTGGTCGACTTATGAGA---ACTGAAGAAAACAAATTTAACT
 AACAGATGAAGAGATAATTGATCAAATAATCACAATTTGTACTCTGGCT
 ACGAAACGGTTTTGCACCACTTCAATGATGGCAGTCAAGTATCTGCATGAT

CACCCAAGAGTTCTTCAGGAGCTAAGAAAAGAGCATTTCGCAATTAGAGA
 AAAGAAAAGGCCTGAGGATCCAATCGATTTAAATGACCTTAAATCG --- C
 GTTTTACTCGTGCCGTGATTTTTGAGACCTCAAGATTGGCTACAATAGTA
 AATGGGGTTTTGAGGAAGACTACTAAAGAAATGGAACATAATAGATTGTG
 GATTCCAAAAGGATGGAGAATCTACGTTTACACAAGGGAGATAAACTATG
 ATCCATATTTATATCCTGACCCATTCTCCTTTAACCCATGGAGATGGCTG
 GACAAAAGTTTTGGAGTCTCAAACTATCTCTTCATTTTTGGAGGAGGTAC
 CAGGCAGTGTCCAGGAAGGAG --- CGAATAGCTGAGATTTCAACTTTCC
 TTCATTATTTTGTAACTAGATACAGATGGGAAGAGGTTGGAGGAGACTCA
 TTAATGAAATTTCCAAGAGTTGAAGCACCAAATGGGCTACACATTAGGGT
 CTCATCTCAC
 >R

 -----AAGAAGG
 GTTTGCCTCCTGGCACTATGGGTTGGCCAGTTTTTGGAGAACTACTGAG
 TTTCTTAAACAAGGTCCTCAACTTCATGAAAAACCAAGAGCAAGGTATGG
 TAATTTTTTTAAATCGCACATTCTTGGGTGTCCTACCATTTGTTCAATGG
 ATCCAGAGGTAATAGGTACATTCTCATGAATGAGGCAAAAGGGCTTGTT
 CCTGGTTATCCTCAATCCATGCTGGATATATTAGGCAATGTAATATTGC
 AGCAGTTTCATGGCTCCACTCACAAGTACATGAGAGGTGCACATTAAAGCC
 TCATTAGCCCTACCATGATCAGAGAACAACTTTTGCCAACAATCGATGTC
 TTCATGAGAACCACCTCAGCAATTGGCATGACAAA -----AT
 TATTGACATTCAACAGAAAACCAAGAGATGGCACTTCTCTCATCACTTA
 AGCAAAATGCTGGCCCTGATTCTAGCTCAATATCTCAAGCATTTCATGCCT
 GAGTTTTTCAAGCTTGTTTTAGGCACCTTGTCGTTGCCAATTGATCTTCC
 TGGAAACAATTCGTAAACGGATTTAGGCTAGGCAAAATATTGTGAGCT
 TATTGAGACAACCTTATAGAAGAGAGAAGGGCATCAAAAGAAACCCATAAA
 GACATGCTTGGTTGCCTCATGAAA --- ACTGATGAAAACAGATACAACT
 CAATGATGATGAGATAATTGATCAAATCATCACAATTTTATATTCTGGGT
 ACCGAAACAGTCTCAACAACCTTCAATGATGGCTATTAAGTATTGTCATGAT
 CACCCTCAAATCCTTCAAGAACTAAGAAAAGAACATTTAGCAATTAGAGA
 AAAGAAAATGCCTGAGGATCCAATTAACCTTAACGACCTTAAATCGATGT
 GTTTTACTCGTGCGGTGATTTTTGAGACCTCAAGATTAGCTACAATAGTT
 AATGGGGTTCTGAGAAAAGACTACTAAAGAAATGGAAATAAATGGATTGT
 GATTCTGAAGGATGGAGAATCTATGTTTATACTAGGGAGATAAACTATG
 ATCCCTACTTGTACCCAGATCCATTGTCTTTCAACCCATGGAGATGGATG
 GATAGAAGTTTGGAGTCTCAAACTATCTCTTCATTTTTGGAGGGGGTAC
 CAGACAGTGTCTCGGAAAGGAACTAGGAATAGCAGAAATTTCAACTTTCC
 TTCATATTTTGAACCAAGATACAGATGGGAAGAGTTGGTGGAGACACA
 CTAATGAAATTTCCAAGAGTAGAAGCACCCAATGGACTACACATTAGGGT
 TTCATCTTAC
 >O

 -----GTGTTGGTGGCGATTGGGGTGGTTGTGGCGGGCGGGTGGTGGT
 GAGCAGCCTGCTGCTGCGGTGGAACGAGGTGCGGTACAGCCGGAAGCGCG
 GCCTGCCGCCGGGACAAATGGGGTGGCCGCTCTTCGGCGAGACCACCGAG
 TTCCTCAAGCAGGGACCCAGTTTCATGAAGGCCCGGAGGCTCAGGTACGG
 GAGCGTGTTCAGGACGCACATCCTGGGGTGCCCGACGGTGGTGTGTATGG
 AGGCGGAGCTGAACCGGCGGGCGCTGGCCAGCGAAGGGCGCGGGTTCGTC
 CCGGGCTACCCGCGAGTCGATGCTGGACATCCTGGGGCGGAACAACATCGC
 CGCCGTGCGAGGGCCCCCTCCACCGCGCCATGCGCGGCGCCATGCTCTCCC
 TCGTCCGCCCCGCCATGATCCGCTCCTCCCTCCTCCCAAGATCGACGCC
 TTCATGCGCTCCACCTCGCCGCTGGTCTCTCTCTCTCTCCGCCGT
 CGTCGACATCCAGGCCAAGACCAAGGAGATGGCCTTGCTATCTGCACTCA
 GGCAGATTGCGCGGCTCTCCGCTGGCCACTCTCTGACGCTCTCAAGGCA
 GAGCTCTACACCCTTGCTGTTGGCACCATCTCCCTGCCCATCAACCTTCC
 TGGAAACCACTACTACCAAGGCTTCAAGGCAAGGAAGAAGCTTGTTGCAA
 TGCTAGAGCAGATGATCGCGGAACGGCGATCTCCGGTCAGGTACACGAC
 GACATGCTGGATGCGCTCTTGACCGGTGTCGAGGGCACAGGGAGAAGCT
 CACAGATGAGCAGATCATTGACCTGATCATCACCTTATATACTCTGGAT
 ATGAAACCATGTCGACGACCTCGATGATGGCTGTCAAGTACCTGTGACAC
 CATCCCAAAGCTCTTGAGCAACTCAGGAAAGAACATTTTGATATCAGGAA
 AGGTAAGCGCCCGAAGATGCCATCGACTGGAATGATTTCAAGTCCATGA
 CCTTCACTCGAGCTGTTATCTTCGAGACATTAAGATTAGCTACAGTTGTG
 AATGGGCTGCTGAGGAAAACACCAAGATGTTGAAATGAATGGGTATGT
 TATCCCAAAGGTTGGAGAATATATGTTTACACAAGGGAAATAAATTATG
 ATCCATTCTGTACCTGATCCCATGACATTCAATCCATGGAGGTGGCTG
 GAGAAGAACATGGAATCACATCCACACTTCATGCTGTTTGGAGGAGGTAG
 TCGAATGTGCCGGGGAAGGAAGTAGGCACCGTAGAAATTGCAACATTCC
 TTCATATTTCTGTAACATACAGATGGGAGGAAGAGGTAAACAACACA
 ATATTGAAGTTCCCCGAGTTGAAGCTCCCAACGGGTTACATATCCGCT
 TCAAGATTAC
 >V

 -----GCTGTTTTTCATGGCGGTTTTTGGTGTGGTTCTTATTGGGCTCTGTAT
 TTGCACTGCTTTGTTGAGATGGAATGAGGTGAGGTAC --- AGGAAGAAAG
 GATTGCCCCAGGTACAATGGGTTGGCCAGTCTTTGGGGAGACCACTGAG

[illegible]

$\geq R$

-----AAGCAACTAATGAGCTTTGATCCAGGG
GAGTGGACTGAGAGTTTAAAGGAAAGAGTACGTTCTTGTTATTGAAGGATT
TTTCACTGTGCTTTGCCTATA-----TTCTCCACCACAT
ACCGAAGGATCTACAAAGCTAGGACTAAGGTGGCAGAGGCATTAAAGCTT
ATAGTGATGCAAAAGACGAAGAGAGAGT-----
---GAGTCAGGA-----GAAAGAAAGAAAGACATGTTAGGAGCTCTTC
TAGCCGCCGATGAT-----GGGTTTCCCGATGAGGAAATAGTGGAATTT
TTGGTGGCATCTACTTGTGGCCGCTATGAAACTCCTACTATTATGAC
TCTTGCTGTAAAGTTCCTAACTGAGACACCTCTCGCTCTGGCTCAGCTCA
AGGAGGAGCATGAAGGGGATTAGGGCAAAGAAAGTGAAGGAGAGAAGCTCTA
GAGTGGAGTGATTACAAGTCAATGCCTTTCAGTCAATGTGTTGTAAATGA
GACATGAGATAGACAAACATATTAAGTGGGGTATTAGACAGGCAATGA
CAGACATCAATATAAAAGGTTACACAATCCCTAAGGGATGGAAGGTTTTT
GCTTCATTTTCGAGCTGTACATTTAGATCATGATCACTTCAAAGATGCTCG
ATCTTTTCAATCCATGGAGATGGCAGACAACATTCAGGAGTAACATGTCGAG
GG-----AATGTGTTCACTCTATTGTGAGAGGAGGACCAAGACTCTGC
CCAGGTTATGAAGTACGTAGAGTAGAACTCTCTGTTTTCTGCATCATCT
AGTCACTCGGTTGAGTTGGGTTCTGACAGGAAAGATAAGTTGGTTTTCT
TTCCGACGACCAGACACAGAAACGGTACCCAATCAACGTGCAGCGACGA
AGCGATGTAAGAGT-----

 ≥ 0

73

CTGATATTCATTATAAAGATTACACAATTCCAAAGGGATGCAAGATTTT
GCTTCATTCCGAGCTGTGCACCTTAATAATGAACACTACGAGAATGCTCG
GACATTTAAACCCTTGGAGATGGCAGATCAACAATAAACTTCAGAATGCGG
TA---GGGGCCAATATATTTACTCCATTTGGTGGTGGACCTCGGTTGTGT
CCTGGCTATGAGCTTGCCCGGGTTGTCGTTTCTATCTTCTCCATCATCT
TGTAACGCGCTTTAGCTGGGAAGAAACCGAAGAAGATAGACTTGTCTCT
TCCCCACCACACGAACTCTCAAAGGATACCCTATCAATCTTCGGCTGCTT
TCAGAATCAATTTGC-----
>V
GACTGGTCTCTGGCCGTTACCATCGGTGCGGCTCTCCTCGCCATCTTCTT
CTTGTTCTGCGCCTC-----A
CTCGCCCCAGGGGCCACCGCCTGCCTCCGGGCAACCTGGGGCTCCCCCTG
GTGGCGGAGACTCTGCAGCTCATCTCCGCTACAAGTCGGCCAACCTGA
ACCCTTCATCGACGAGAGGGTGACTCGGTACGGGCCCTTGTTACCAACCC
ATGTGTTGCGCGAGCCACGGTGTCTCGGCGGACCCGGAGACGAATCGG
TACATATTGCAAGCAGAGGGGAAGCTGTTGAGTGACGCTACCTGGCTC
CATCTCCAACCTTGCTTGAAGGCATTCCCTGTTGTTAATGAAGGGTAATC
TCCACAAAAAATGCATTCCCTCACCATGAGTTTCGGTAATTCTTCCATC
ATCAAGGACCACTTGCTGCTCGACATCGACGCTTGATCCGCTTCAACAT
GGACTCATGG-----ACGAGCCGGATCTTACTTATGGAAGAGGCCAAGA
AGATAACATTCGATTTGACAGTGAAGCAGCTGATGAGCTTTGATCCAGGG
GAGTGGACTGAGTCTCTGAGAAAAGAATATGTGCTTGTGATTGAAGGCTT
CTTTACCGTCCCTTTCCCTCTC-----TTCTCCGCCACCT
ACCGCAGAGCCATCCAAGCAAGGACGAAGGTAGCAGAGGCACTGAACCTG
GTAGTGAGGGAAGGAGGAAAGCCAAG-----
----GAAGAAAGGAGAAGAGGAGAAGAAGAAGACGACATGCTGGCCGCCCTCC
TCGACAGCGCGAT-----AACTTCTCCGACGAGCAGATCGTGGATTTC
ATCCTGGCCCTGCTGGTGGCTGGCTACGAGACCACCTCCACCATCATGAC
CCTCGCGATCAAGTTCCTCACAGAGACCCCTCTCGCTTGGCTCAGCTCA
GGGAGGAGCATGACGAGATTAGGGCAAGGATAAGTGACCTAATGTTCTG
GAATGGAGTGATTATAAATCCATGCCTTTCACTCAATGTGTTGTAATGA
GACTTTGAGGGTGGCAACATAATAAGGTGGGATTTTCAGGCGAGTAACCA
CTGACATCCATGTCAAAGGTTACACAATTCCTAAGGGATGGAAGGTGTTT
GCATCATTTGAGCTGTACACCTGGACCATGAATACTTCAAGGATGCTCG
TACTTTTAACTCCCTGGAGATGGCAGAATAATTCTGGACCCACGGTTCAG
GGAGTGGTGTGAATGTGTTACACCATTGAGGAGGGGCCGCGCTGTGC
CCTGGCTATGAGCTTGAAGGGTGGTGTCTCGGTTTTCTTACCCGAT
GGTCACTCGATACAGCTGGGCTCCTGCAGAGGAGGATAAGTTGGTGTCT
TTCCAACCACTAGGACTCAGAAGCGATACCCCATCAATGTGGAGAGGCGA
AAAGTGTATCCGCCGAGAAAAGGTGAGGGA

BR.Det2.fasta

>A
---GAAGAAATCGCGATAAAACCTTCTTCCGATACTGTCTCCTCACTCT
TATTTTCGCGGGCCCAACCGCCGCTCTTCTGAAATTCCTCCAAGCTC
CTTACGGTAAACACAACCGTACCGGATGGGGTCCACCGTATCTCCACCG
ATTGCTTGGTTCGTATGGAGAGCCCAACCTTGTTGGCTCACTCTCCTCCT
CTTCCCTTTGGTGTGTCACGCTCTCAACCCCTAAATCTCTACTTCTATTC-
-----TCTCCTTATCTCATTCTTACTTCCACCGCACCATCATTTACCCT
CTTCGCCTCTTC-----CGCAGCTCCTTCCCCGCCGGTAA
AAACGGATTTCCGATCACCATCGCCGCTTGGCTTTACCTTTAATCTCC
TCAATGGTTATATCCAGGCGAGG---TGG-----GTTTCG
CATTACAAGGATGACTACGAAGACGGAACCTGGTTCTGGTGGCGGTTTGT
TATCGGTATGGTGGTTTTATACCGGATGTATATAAATATCACGTCGG
ACCGGACTTTGGTACGATTGAAGAAAGAGAACCGGGGAGGTTATGTGATA
CCGAGAGGAGGCTGGTTCGAGTTGGTAAGCTGTCCGAATTATTTGGAGA
GGCGATTGAGTGGTTGGGCTGGGCTGTTATGACTTGGTCTTGGGCCGGTA
TTGGATTTTTCTGTACAGGTGTTCCAATTTGTTTCCGCGTGACGTGCG
AGTCACAAGTGGTACATTGCCAAGTTCAAGGAAGAGTATCCCAAGACTCG
TAAAGCTGTTATTCCTTTGTGTAC
>P
-----GCCCTATTAGATCAGAGCCTCTTTCCTACAGTCTCCTATCCAT
TTACCTCATAGGACCCCTACATTCATTTCCCTGAAATTTCTTCAAGCCC
CGTATGGGAAGCACCAACCGTCCAGGGTGGGGTCTACAATCTCTCCACCC
TTGGCTTGGATCCTCATGGAAGGCCCACTCTTTGGCTCACACTCCTCCT
CTTCCCCTATGGCCAACACTTCACCAATCCCAAAGCTATCATCTCATG-
-----TCCCCCTTCTCCTCCACTACTTCCACCGCACCTGCATCTACCCA
CTTCGGATTTAT-----CGGAATTCAAAACCACT-----
-GGTGGTTTCCCGGTGAGTGTGGCTGTTATGGCATTTGGGTTAATCTGC
TGAATGCTTACCTGCAGGCTAGG---TGG-----GTGTCT
CATTACAAGAGTGATTATGAC---GGGGGACTGTTTTGGTGAAGTTTTT
TGGTGGGTTAGTGGTGTGTTGTGGGGCATGAGGATCAATACGTGGGCCG
ATGGGGTTTTGTTAGGTTTGAAGAGGGAAGGT---GGTGGGTATAAGGTG

```

CCGAGAGGAGGGTGGTTTGTAGTTGGTGAGCTGCCCAACTACTTCGGGGA
GATTGTGGAGTGGCTTGGATGGGCAGTGATGACTTGGTCTTGGGCTGGGT
TTGGTTTTTACTCTATACATGTGCTAACTTGGTGCCTAGGGCATGTGCC
AATCACAAATGGTACCTGGAGAAATCAAAGAGGATTATCCCAGGAGCAG
GAAAGCAGTGATCCCTTTCTGTGT
>R
-----GCTTTCTCAGATGGAAGCCTCTTTCATAAAAGCCTCCTAGCTCT
CTACCTCATAGGGCCCCCTACCTTCATCTCCCTCAAGTTCCTTCAAGCAC
CGTATGGAAACACAACCGCCCTGGATGGGGCCCCACCATCTCGCCACCT
TTGGCCTGGTTCCTCATGGAGAGCCCCACTCTCTGGCTTACATTTTTACT
CTTCCCTTTTGGCCAACATGCTTCCAATCCCAAAGCTTGGGTCTTATG-
-----TTCCCTTATCTCTTCCATTATTTCCACCGCACCTGTCTCTACCT
CTACGCCTTTACTGGAATACAAATCAGCGGAATACCAAATCACT-----
-AGCGGTTTTCTGTGAGCATGGCCCTTCTGGCATTITTTGTTCAACTTCT
TGAATGCTTACTTGCAAGCCAGG---TGG-----GTATCT
CATTACAAGGAGGATTATGACCAGGATAGGTGGTCTGGTGGAGGTTTTG
GGGTGGGCTAGTGGTGTTCGTGGGGTATGTGACTAATGTATGGTCGG
ACAGGGTTTTGGTTCGGTTAAAGAGAGAAGGT---GATGGGTATAAGGTG
CCAAGAGGAGGGTTGTTTGATCTGGTGAGCTGCCCAACTACTTTGGTGA
GATTGTGGAATGGCTTGGTTGGACTCTGATGACATGGTCTGGGTGGGT
TTGGGTTTTTCTCTATACATCTGCCAACTTGGTGCCGAGGGCAGTGGCT
AACCAAGAGGTATCTGGACAAGTTTGGAAAGGATTATCCAAGAGGTAG
AAAAGCTGTATCCCTTTCTTGAT
>O
GGCGCGGCGGCGGCGACGACGCGCTGTACGCGGATGCCTGGTGATCCT
GTACCTGATCAGCCCGATCACCCTGTTCTGCTCCGGTTCGTGTGCGCGC
CGTACGGGAAGCTCTCCCGCCCGGGTGGGGTCCCGCGTGCCGGCGGCG
CTCGCGTGGTTCCTCATGGAGAGCCCCACCATGTGCTCCCGCCCTCGT
CCTCTCCGCGCGGCTCCTCGTCGCGCTGCGCGCGCGGCTCTCTTCC
CGCGCGGCTGTACGCGCTCCACTACGTGAACCGCACGCTCGTCCACCG
CTCCGGCTCTC-----CGCTCCGCGCGCGCG-----
-GCGCGGTTCCGATCCTCGTCGCGCTTCCGCTTCGGCTTCAACCTCC
TCAACGCCTACGTCCAGGCGGCTCCTGGGCGCTCGACGCGCGGCTCCC
CACAGCACCGCCACCGCCACCGCCACGCGCGCGGCGCTCGCGCTGCT
CGTCGGGCTAGCCCTCTTCGCGTGGGGGATGCGGACCAACATCGCGGCG
ACAAGGCGCTCCTGAGGCTGAAGGAGGCGGG---AAAGGTACAGATT
CCCCGCGGCGGGCTGTTTCGAGTGGTCACTGCCCAACTACTTCGGCGA
GGCGGTGGAGTGGCTCGGCTACGCGCTGGTGGCGTGGACGCGCGGCGCT
GGGCTTCTTCTCTACACCTGCTCAACCTCGGGCCGAGGGCCAGGGAT
CACCGCGGTTGGTACGTGCGCAAGTTCGGCGACAAGTACCGCGCTCGCG
CAAGGCGTTCTGTCCTGATCTAC
>V
-----GATTGGGACCGAACCTGTTCCACAACCTCTCTCATCGCCCT
CTACCTCATCGCCCCACCCACATTCTCTCCCTCGTTTCTTCCAGCCC
CCTATGGCAAGCACCACGATCGGGGTGGGGTCCCACTGTGCCGCTATCC
TTGGCATGGTTCCTCATGGAATCTCCTACTGTGTGGTTGACGCTCCTCAT
CTACCCCTTGGGTGATCGTTTCTCCAATCCCAGATCCCTAATCCTCATT-
-----TCATCTACCTCTTCCACTACATCCATCGTACTCTCATCTACCT
CTTACCTC-----CGCAACTCACTGCCGCTAAACG
GAATGGGTTCGGGTGAGCGTGGCGCTCATGGCTTTGGATTCAACCTCC
TGAACGCCTACTTGCAAGCCAGA---TGG-----GTGTCT
CATTAT---GCTAATTACGAGGACGATGGGTGGTCTGGTGGCGGTTCTG
GGTGGGACTGGTAATCTTCTGGATGGGATGTTGGTGAATGTAAGATCGG
ATTTGGCGTTGGTGGGTCTGAAAAGCCAGGGC---GGAGGGTACAAGGTG
CCGAGAGGAGGGTGGTTCGAGTTAGTGAGCTGTGCCAACTATTTGGGGGA
GGTGGTGGAGTGGTTGGGATGGGCGGTGATGACGTGGTCTTGGGTGGGT
TAGGGTTTCTTCTGTACACATGTGCCAATTTGGTGCCAGGGGCAGTGCC
AACCAAGTGGTATTTGGAAAAGTTTGGGAGGATTATCCAAGGGGAG
AAAAGCTGTTATTCCCTTCTCTAT

```

BR.Dwf4.fasta

```

>A
---TTCGAAACAGAGCATCATACTCTTACCTCTTCTTCTCTCCATC
GCTTTGTCTCTTCTTCTTCTTGTATCTTGAAG-----
-----AGAAGAAATAGAAAAACCAGA
TTCAATCTACCTCCGGGTAATCCGGTTGGCCATTTCTTGGTGAACCAT
CGGTTATCTTAAACCGTACACCGCCACAACACTCGGTGACTTCATGCAAC
AACATGTCTCCAAGTATGGTAAGATATATAGATCGAACTTGTGGAGAA
CCAACGATCTGATCAGCTGATGCTGGACTTAATAGATTATATACAAAA
CGAAGGAAGGCTCTTTGAATGTAGTTATCCTAGAAGTATAGGTGGGATTC
TTGGGAATGGTCGATGCTTGTCTTGTGGTGACATGATAGAGATATG
AGAAGTATCTCGCTTAACCTTCTAAGTCACGCACGCTTAGAACTATTCT
ACTTAAAGATGTTGAGAGACATACTTTGTTTGTCTTGATTCTTGGCAAC

```

AAAACTCTATTTTCTCTGCTCAAGACGAGGCCAAAAAGTTTACGTTTAAT
 CTAATGGCGAAGCATATAATGAGTATGGATCCTGGAGAAGAAGAAACAGA
 GCAATTAAGAAAGAGTATGTAACTTTC - - - AAAGGAGTTGTCTCTGCTC
 CTCTAAATCTACCAGGAAGCTTATCATAAAGCTCTTCAGTCACGAGCA
 ACGATATTGAAGTTCATTGAGAGGAAAAATGGAAGAGAGAAAAATGGATAT
 CAAGGAAGAAGAT - - - CAAGAAGAAGAAGAAGTGAACACAGAGGATGAAG
 CAGAGATGAGTAAGAGTGATCATGTTAGGAAACAAAGAACAGACGATGAT
 CTTTTGGGATGGGTTTTGAAACATTGCAATTTATCGACGGAGCAAATTC
 CGATCTCATTTCTAGTTTGTATTTGCCGGACATGAGACTTCTCTGTAG
 CCATTGCTCTCGCTATCTTCTTCTTGCAAGCTTGCCCTAAAGCCGTTGAA
 GAGCTTAGGGAAGAGCATCTTGAGATCGCGAGGCAAGAAGGAAGTGG
 AGAGTCAGAATTAATTTGGGATGATTACAAGAAAAATGGACTTTACTCAAT
 GTGTTATAAATGAAACTCTTCGATTGGGAAATGTAGTTAGGTTTTTGCAT
 CGCAAAGCACTCAAAGATGTTCCGTTACAAAGGATACGATATCCCTAGTGG
 GTGGAAAGTGTTACCGGTGATCTCAGCCGTACATTTGGATAATTCTCGTT
 ATGACCAACCTAATCTCTTTAATCCTTGAGAGTGGCAACAGCAAAACAAAC
 GGAGCGTCATCTCAGGAAGTGGTAGTTTTTCGACGTGGGGAACAACTA
 CATGCCGTTTGGAGGAGGGCCAAAGGCTATGTGCTGGTTGAGAGTAGCCA
 AGTTAGAAATGGCAGTGTTTATTCATCATCTAGTTCTTAAATTCAATTGG
 GAATTAGCAGAAGATGATAAACCATTTGCTTTTCTTTTGTGATTTTCC
 TAACGGTTTGCTATTAGGGTTTCTCGTATTCTG-----

 >P
 -----ATGTCTCACTCAGAGCTTGTGTCTTTCTCCTTCCATC
 GATTTTATCACTACTCTTGCTCTTCATTCTCGTGCAA-----
 -----AGAAAGCAAGTAAGA
 TTTAATCTCCACCAGGCAACATGGGGTGGCCATTTCTTGAGAAACCAT
 TGCTAC - - - CAGCCTTACTCTGCTACTTCAATAGGAGAATTCATGGAAC
 AGCAGATATCAAGGTATGGAAGATTTACAAGTCCAATTTGTTGGGGAG
 CCAACAATATGCTATCCGCAGATGCTGGACTTAACAGATTATACACAGAA
 TGAGGGAAGATTATTTGAATGCAGCTATCCAAAAAGTATTGGTGAATTC
 TTGGAATATGTTCCATGTTGGTTCTTGTGGAGACATGCATAGAGACATG
 AGGATTATATCTTAACTTTTTGAGCCATGCCAGGTTAAGAACTCATCT
 ATTTGAAAGAAAGTGGAAGCAACCCCTGCTTGTCTTAGCTCTTGGAAGG
 AGAATTGTACATTTTCAGCTCAAGATGAAGCAAGAAGTTTACCTTCAAT
 TTGATGGCAAAACATATCATGAGCTTGGATCCTGGAAGCCAGAGACTGA
 GCAGCTGAAAAAAGAGTATGTTACTTTTCATGAAAGGAGTAGTTTCAGCTC
 CTATAAATTTGCTTGGAACCCATATAGAAAAGCCTTGAAAGTCTCGATCA
 ATCATCTTGAAATTTATAGAGCGAAAGATGGAGGAGAGAATTGCTGAAAC
 GAAGGGTGGAGTA - - - GAAAACTTGGAAGAC-----
 -----GATGAT
 CTTCTTGATGGGCTTTGAAGCATTCAAATCTTTCTACGGAGCAAATCCT
 TGACTTAATCTTAAGCTTGCTCTTTGCTGGCCACGAACTTCTTCTGTGT
 CCATAGCTCTAGCCATCTACTTCTTGCAAGCTTGCTCCTGGTGCTATTCAA
 CAGTTAAAGAAGAACATATTCAAATCTCCAGAGCCAAGAAACGGTCAGG
 AGAGACGGAATTGAAGTGGGATGATTACAAAAAATGGAATTCACCCAAT
 GCGTTATAAGCGAGACACTGAGGCTTGGAACGTAAGTCAGGTTTCTTAC
 AGAAAAAGCTCTAAAAGATGTTCCGTTACAAAGGTTACGACATTCATGTGG
 ATGGAAGTGCTACCAAGTATCTCAGCCGTCCATTTAGATTCAACTCTTT
 TCGACCAACTCAACACTTCAATCCATGGAGATGGCAGCAGCATAAACAAT
 GCTCGTGGTGCTTCTACAAGTTCGAGTGGGACGACGAGTAGTAATTACTT
 CATGCCATTTGGGGAGGACCGGACTTTGTGCAGGATCGGAATTGGCAA
 AACTAGAAATGGCCGTGTTTCATCCACCATTTGGTCTCAACTTCCACTGG
 GAACTGGTTGATACCGATCAAGCCTTTGCTTTTCTTTTGTGATTTTCC
 CAAGGGCTTGCCATAAGAGTCAAGCACCTACTCACAGTTACA-----

 >R
 -----TCTGACTCAGAGCTTATTCCTTTTATTCTTCTTCTC
 AATCTTATCTCTAATTTTAGTCTTCATTCTCATATAC-----
 -----AGAAAGCAAGCAAGA
 TTCAATCTTCCACCAGGCAACATGGGCTGGCCCTTTCTTGAGAAACTAT
 TGGCTACTTTGAAGCCTTACTGTGCCACTACAATTGGCGAATTCATGGAAC
 AACACATCTCAAGGTATGGGAAGATATACAAGTCTAATTTATTTGGGGAG
 CCAACGATTGTCTCTGCAGATCCTGGACTAAACAAGTTCATACTACAGAA
 TGAAGGGAGGTTATTTGAATGCAGCTATCCGAGAAGCATAGGTGGAATTC
 TTGGAATATGTTCTATGCTTGTCTAGTTGGAGACATGCATAGAGACATG
 AGGACTATATCTCTCAATTTTTTAAGCCATGCAAGGCTCAGGACTCATCT
 ATTGAGAGAGGTAGAGAAGCATACCTTGCTTGTCTCGACTCTTGAGTG
 ACAACTCTACATTTTCAGCTCAACATGAGGCGAAGAAGTTCACCTTCAAT
 TTGATGGCGAAACATATAATGAGCATGGATCCTGGAAGCCAGAGACTGA
 ACAGCTAAAGAAAGAGTATGTTACTTTTATGAAAGGAGTTGTTTCTGCTC
 CTCTGAATTTTCTGGAAGTGCATACAGAAAGGCTTTGCACTCTCGATCG
 AGCATCTTAAAGTTCATAGAGAGGAAATGGAGGAAAGAGTAGGAAAAAT
 GGAGGAAGAA-----
 -----GAGGAT

CTTCTTGGATGGGTATTGAAGCATTCTAATCTGTCTAGGGAGCAAATCT
GGATTTAGTACTCAGTTTGCTGTTTGGCTGGCCATGAACTTCCTCGGTAT
CCATAGCTTTAGCTATGATTTTCTTGCAAGGTTGTCCAGGGCTACTCAG
CAGTTGAGGGAAGAACACATGGCAATTGCAAGAGCAAAGAATGAATCAAG
CCGGAGCGGAATTGAACTGGGAAGATTACAAGAAAATGGAATTTACTCAAT
GCGTTATTAGTGAGACACTTAGACTTGGAAACGTGGTGAGATTTCTTCAC
AGAAAGGCTGTGGAAAATGTTTCGATATAAAGGGTATGACATTCACGTGG
ATGGAAAGTGCTCCAGTGATTGCAGCCGTGCATTTGGATTCTAGTCTGT
TTGACCAGCCTCAACTATTCAATCCATGGAGATGGCTGCACCACTCG
AATAGGGGAGGTTGT-----GGAATTACTT
CATGCCGTTTGGTGGAGGACCGCGGTATGCGCGGATCGGAGTTGGCAA
AATTAGAGATGGCAATATTATCCACCATTTGGTTCTGAATTCAGCTGG
GAGTTGGCGGATAATGATCAAGCTTTTGTCTTCCCTTTGTGCACTTTCC
TAAAGGCTTACCCATAACTATAAGAGCCAACCACCAAGAAACAAGAAA
CTTCA

>0

GCCGCCATGATGGCGTCCATAACCAGCGAGCTGCTCTTCTTTCTCCCTT
CATCCTCCTTGCCCTGCTCACGTTCTACACCACCACCGTGGCCAAATGCC
ACGGCGGGCACTGGTGGCGAGGTGGGACGACGCGGCGAAGAGGAAGCGG
ATGAACCTGCCGCCCGCGCGCGGGTGGCCGCTCGTGGCGAGACGTT
CGGCTACCTCCGCGCCACCCCGCCACCTCCGTCGGCCGCTTCATGGAGC
AGCACATCGCACGGTACGGGAAGATATACCGGTCGAGCCTGTTGCGGGAG
CGGACGGTGGTGTGGCGGACGCGGGGCTCAACCGGTACATCTGCAGAA
CGAGGGGAGGCTGTTTCGAGTGACGCTACCCGCGCAGCATCGGCGGCATCC
TGGGCAAGTGGTCCATGCTGGTCTCGTCGGGGACCCGACCGCGAGATG
CGCGCCATCTCCCTCAACTTCTCTCCTCCGTCGCGCTCCGCGCGTCTCT
CCTCCCCGAGGTCGAGCGCCACACCCTCTCGTCTCCGCGCTGGCCCC
CTTCTCCACCTTCTCCGCTCAGCACCAAGCCAAGAAGTTCACGTTCAAC
CTGATGGCGAAGAACATAATGAGCATGGACCCGGGGGAGGAAGAGACGGA
GCGGCTGCGGCGGGAGTACATCACCTTCATGAAGGGCGTGGTCTCCGCGC
CGCTCAACCTGCCCGGACGCCCTACTGGAAGGCTCTCAAGTCGCGTGCT
GCCATTCTCGGAGTAATAGAGAGGAAA---GAAGAGCGGGTTGAGAAGCT
GAGCAAGGAGGAT---GCAAGCGTAGAGCAA-----
-----GACGAT
CTTCTCGGATGGGCTCTGAAACAATCTAACCTTTCAAAAGAGCAAATCCT
GGACCTCTTGCTGAGCTTGCTCTTCCGCGGGCACGAGACGTCGTCATGG
CGCTCGCCCTCGCCATCTTCTTCTTGAAGGCTGCCCCAAGGCTGTCCAA
GAACTGAGGGAGGAGCATCTTGGGATTGCAAGGAGACAAGGCTAAGAGG
GGAGTGCAAATTGAGCTGGGAAGACTACAAAGAGATGGTTTTACGCAAT
GTGTCTAAACGAGACGTTGCGGCTAGGAAACGTGGTCAGGTTCTGCAC
CGGAAGGTCTCAAGGACGTGCACTACAAGGGTTATGACATTCGAAGCGG
ATGGAAGATCTGCCGGTGTAGCCGCGGTGCACTTGGAAGTCTGCTCCGT
ACGAGGACCCCCAGCGCTTCAATCCCTGGAGATGGAAGAGTAGCGGATCA
TCCGCGGCTTGCTCAGAGCAGCAGC-----TT
CATGCCGTACGGCGGCGGGACGCGGCTGTGCGCCGGGTCGGAGCTCGCGA
AGCTGGAAGATGGCCGTGTTCTTGCAACCCTGGTGCTCAACTTCAGGTGG
GAGCTCGCCGAGCCGACCAAGCCTTCGTCTTCCCTTCGTGCACTTCCC
CAAGGGCTTCCCATTAGGGTTCATAGAATTGCACAGGATGATGAGCAGG
AG---

>V

-----TCTGACTTGGAGTTTTTCTCCTCCTCTTGCCTGCGTC
AATCTCAGTTCCCTTTGTGATTCTCAATCTGATCAAA-----
-----AGAAAGCAAAGCGGA
CTTAATCTACCTCCGGGAACACGGGTTGGCCTTTTCTGGGTGAAACCAT
CGGCTATTTAAAGCCGTAAGTCCGCGCACTTCAATAGGAGACTTCATGGAGC
AACACATATCAAGGTTTCGGAGAAATCTACAAGTCCAATCTGTTTGGCGAG
CCAACCATAGTCTCAGCGGATTCTGGGCTGAACAGATTCTACTACAGAA
CGAAGGAAAATTGTTTGAATGCAGCTATCCAGAAAGCATAGGTGGAATTC
TTGGGAAATGGTCCATGCTGGTTTTAGTTGGAGACATGCATAGAGACATG
AGAACCATCTCCCTCAACTTCTTGAGCCATGGCAGGCTTAGGACTCATCT
CCTACCAGAGGTGGTGAAGCACACTTTGCTTGTCTAAGCTCTTGAAGG
AGAATTGTACATTTTCTGCTCAAGATGAAGCTAAGAAGTTTACCTTCAAT
CTGATGGCAAAACATATCATGAGCTTGGATCCCGGAAAGCCGGAGACTGA
GCAGCTTAAGAAAGAGTATGTTACTTTTCATGAAAGGAGTGGTATCTGCTC
CTTTGAATTTCCCTGGAAGTGCATACAGAAAAGCCCTACAGTCTCGGTCC
ACCATCTTAAAAATTTATCGAGCTGAAGATGGAAGAGAGGATTCAGAACT
GAGGGGAGGAGGTTTGAAGATATGGAGGAT-----
-----GACGAT
CTTCTTGGATGGGTGCTGAAGCATTCCAACCTTTCAACTGAGCAAATCCT
GGACTTGGTACTGAGCTTGCTCTTGGTGGCCATGAACTTCATCAGTAG
CAATAGCTTTAGCTATCTACTTCTTGAAGGCTGTCCCAACGCGGTTCAA
CAGCTAAGAGAAGAACACCTTGAAATTTCCAGGGCCAAGAAGCAATCAGG
AGAGTTGGAATTGAGTTGGGAAGACTACAAAAAATGGAATTCATCAAT
GTGTTATTAGTGAGACACTGAGGCTTGGAAATGTGGTTAGGTTTCTACAC
AGAAAGGCCCTGAAAGATGTCCGGTACAAAGGATATGACATTCATGTGG

GTGGAAGTGCTTCCGGTGATTGCAGCAGTGCCTTGGATCCTTCCCTTT
 TCGACCAACCTCAACACTTCAATCCTTGGAGATGGCAGCAGAATAATGGT
 AACCGAGGAAATTCGACTAGCTGGAGCACAGCCACGAATCAAAC---TT
 CATGCCATTTGGGGGAGGACACGGCTATGTGCGGGCTCTGAATTGGCCA
 AACTGGAATGGCGGTGTTTCATCCACCATTTGGTTCTCAACTACAAGTGG
 GAGTTGGTTGATAAGGATCAGGCTTTTGCCTTCCCATTCGTCGATTTTCC
 GAAGGGCTTACCATCAAGGTCCGACACCACAATTTACG-----

CRT.CrtB.fasta

>A
 TCTTCTTCTGTAGCAGTGTTATGGGTTGCTACTTCTTCTCTAAATCCAGA
 CCCAATGAACAATTGTGGGTTGGTA-----AGGG
 TTCTAGAAATCTTCTAGACTGTTCTCTCCTTGTGAGAATCAGAGACTAAAC
 ---AAAGGTAAGAAGAAGCAGATACCAACTTGGAGTTCTTCTTT-----
 -----GTAAGGAACCGA-----A
 GTAGAAGAATTGGTGTTGTCTTCAAGCTTAGTAGCAAGTCCTTCTGGA
 GAGATAGCTCTTTCATCTGAAGAGAAGGTTTACAATGTTGTGTTGAAACA
 AGCTGCTTTGGTGAACAAACAGCTAAGGTCTTCTTCTTATGACCTTGATG
 TGAAGAAACCAACAAGATGTTGTTCTTCTGGGAGTTTGAGTTTGTGGGT
 GAAGCTTATGATCGATGCGGTGAAGTTTGCCTGAATATGCTAAGACGTT
 TTATCTTGGAACTTTGCTTATGACACCCGAAAGGCGAAAGGCGATTTGGG
 CAATCTACGTTTGGTGTAGAAGAACTGATGAAGTTGTGGATGGGCCAAAT
 GCTTACACATATAACTCCCATGGCTTTAGATAGATGGGAAGCAAGGTTAGA
 AGATCTTTTCCGTGGTCGTCTTTCGATATGCTTGATGCTGCTCTCGCTG
 ATACAGTTGCTAGATACCCGGTCGATATTCAGCCATTTGAGACATGATC
 GAAGGAATGAGAATGGACTTGAAGAAATCGAGATACCAAGCTTCGATGA
 TCTATACCTTTACTGCTACTACGTCGCTGGAACCGTCGGATTGATGAGCG
 TTCCGGTTATGGGAATCGATCCTAAGTCGAAAGCAACAACCGAAAGTGT
 TACAACGCTGCCTTGGCCCTTGGTATAGCCAATCAGCTTACTAACATACT
 CAGAGACGTAGGCGAAGATGCGAGAAGAGGAAGGTTTATCTGCCTCAGG
 ATGAATTGGCTCAGGCTGGTCTTTCAGATGAAGACATATTCGCCGGA
 GTAACTGATAAATGGAGAACTTCATGAAAATGCAGCTTAAACGAGCAAG
 AATGTTCTTCGACGAAGCTGAGAAAGGCGTCACCGAGCTCAGTGCCGCTA
 GCAGATGGCCTGTATGGGCTTCATTGCTATTGTACAGGAGAATACTGGAC
 GAGATTGAAGCGAATGATTACAACAATTTTACTAAGAGAGCTTATGTGGG
 GAAAGTCAAGAAAATTGCAGCTTTGCCATTGGCTTATGCTAAATCAGTAC
 TAAAGACTTCAAGTTCAAGACTATCGATA-----
 >P

 -----GTGATATCAAGCATGGTAGCAAGCCATGCTGGA
 GAAATGGCTGTTTTGTATCCGAGGAGATGGTTTATATGGTGCTGAAGCA
 GGCAGCTCTGGTTAAGAAAGAACTGAAGTCCAGAGGT---GATCTGGATG
 TGA---CCA---GATGTTGTTCTTCCAGGGACTTTGAGCTTGTGAGT
 GAAGCTTATGATCGGTGTGGAGAAGTTTGTGCTGAGTATGCCAAGACTTT
 TTACCTGGGAACCTCTGCTCATGACCCCTGAACGGAGAAGGGCTATTTGGG
 CCATATATGTGTGGTGTAGGAGGACCGATGAGCTTGTGATGGGCTAAT
 GCTTACACACATTACACCAACAGCTTTAGATAGGTGGGAGGCAAGGTTGGA
 AGATCTTTTCAAGGGCGTCCATTTGATATGATGGATGCTGCTCTAGCAG
 ACACAGTTGGAAAATTTCTGTTGATATTCAGCCATTCAAAGATATGATT
 CAAGGAATGAGGATGGACTTGAGGAAATCAAGATACCAAACTTTGATGA
 GCTCTATCTTTACTGTTATTATGTTGCTGGGACTGTTGGACTGATGAGTG
 TTCCAGTCATGGGCATTGCACCTGAATCACAGGCTTCGACCGAGAGTGTT
 TATAATGCTGCCTTGGCATTGGGGATAGCCAATCAACTCACTAACATACT
 CAGGGATGTAGGAGAAGATGCAAGAAGAGGAAGGTTTATCTACCACAAG
 ATGAGCTGGCACAGGCAAGGCTTTCAGATGATGACATATTGTGGAAG
 GTGACAGATAAGTGGAGAAACTTTATGAAGAATCAAATTAAGAGAGCAAG
 GATGTTCTTTGATGAGGCTGAGAAAGGAGTGACAGAATTGAGTGCTGCAA
 GTCGATGGCCGGTTTGGGCATCCTTGTGCTATATCGCCAAATACTAGAC
 GAGATAGAGGCTAACGATTACAACAGCTTCACAAAGAGGGCATATGTGAG
 AAAAGCCAAGAAGATAGTTGCTTTGCCAGTTGCGTATGCAAAATCCCTCA
 TTAGTCCATCATCG-----
 >R

-----ACTGTAGCATTACTATGGATTGCC-----AACCCCACTACAGA
 GGTTTCCAGTTCTTTGGGTTGCTCCATTCAATC-----AGGG
 TTTTAGATTTCATAAAATTTGGTTTTATCGATCGAAAT---TTGTTAAG
 GGAAGAGAAAAGAGGATAAGAAGCAGAAATGGAATCAAGTTCTGTGAA
 TATTGATTTGAGGAATCATTGCACAGGT-----A
 GTGGAAGCAAAATTGCCTCTGATATCTAGCATGGTAGCTAGCTCAGTGG

GAAATGGCTATTTCTTCTGAGGAGAAGGTATATAATGTGGTGGTGAAGCA
AGCAGCTTTGATTAAGGCAATTGAGGACTAGCGGT - - - GATCTTGATG
TGAAA - - - CCA - - - GATATTGTTCTTCTGGAACCTCTGGGATTGTTGAGT
GATGCTTATGACAGATGTGGAGAAGTTTGTGCTGAGTATGCGAAAACATT
TTATTTGGGAACCTTTGCTAATGACACCGGAAAGGCGAAGAGCTATCTGGG
CAATATATGTGTGGTGTAGGAGGACGGATGAGCTTGTGCTGATGGGCCTAAT
GCTTCGCACATAACTCCAACAGCTTTAGATAGGTGGGAGGCACGGTTAGA
AGATCTTTTCCAAGGTCGTCCATTTGATATGCTGGATGCTGCTTTGGCAG
ATACTGTTGCCAATTTTCTGTTGACATTGACCCATTCAAAGATATGATA
GAAGGAATGAGGTTGGACCTGAAGAAATCAAGATATAAGAACTTTGATGA
GCTTTATCTTTACTGTTATTATGTTGCTGGGACTGTTGGACTAATGAGTG
TTCCAGTCATGGGATTGACCTGAATCACAAGCAGCAACTGAGAGTGTT
TACAATGCTGCCTTGGCATTAGGGATAGCAAAACAGCTCACCAACATACT
TAGGGATGTTGGAGAGGATGCAAGAAGAGGAAGGGTGTATTTACCACAGG
ATGAGCTAGCGCAGGCAGGGCTTTCAGACGATGACATATTTGCTGGAAGA
GTGACAGATAAATGGAGAAATTTATGAAAAATCAAATAAGAGAGCAGC
GATGTTCTTCAATGAGGCGGAGAAAGGAGTGACTGAACTGAGTGCTGAAA
GTAGATGGCCGGTCTGGGCGTCTTACTGCTATACCGGCAAACTACTAGAC
GAGATCGAAGCAAATGACTATAACAATTCACAAGGAGAGCTTATGTAGG
CAAAGCCAAGAAGCTAGCTTTCTTGCCAATGGCATATGCAAGATCAGCTG
TTGGGCCCGCAAGACTGCCTTCTTTGACAAAGGCA

>0

- - - GCGGCCATCACGCTCCTACGTTGACGCTCTCTTCCGGGCCCTCTCCGA
CGCCCTCGCCCGG - - - - -
- - - - GACGCTGCTGCCGTCCAACATGTCTGCTCTCTACCTGCCAAC
AACAAAGGAGAAGAAGAGGAGGTGGATCCTCTGCTCG - - - - -
- - - - - CTCAAGTACGCTGCCTTGGCGTCGACCCTGCCCCGGGCGAGA
TTGCCCGGACCTCGCCGGTGTACTCCAGCCTCACCCTCACCCTGCTGGA
GAGGCCGTATCTCCTCGGAGCAGAAGGTGTACGACGTCGTCTCAAGCA
GGCAGCATTTGCTCAAACGCCACCTGCGCCCAACAACACACACCATTCCTCA
TCGTT - - - CCCAAGGACCTGGACCTGCCA - - - - - AGAAACGGCCTCAAG
CAGGCCTATCATCGCTGCGGAGAGATCTGCGAGGAGTATGCCAAGACCTT
TTACCTTGGAACTATGCTCATGACGGAGGACCGACGGCGGCCATATGGG
CCATCTATGTGTGGTGTAGGAGGACAGATGAGCTTGTAGATGGACCAAT
GCCTCGCACATCACACCGTCAGCCCTGGACCGGTGGGAGAAGAGGCTTGA
TGATCTCTTACCAGGACGCCCTACGACATGCTTGATGCTGCACTTTCTG
ATACCATCTCCAAGTTTCTATAGATATTCAGCCTTTTCAAGGACATGATA
GAAGGGATGCGGTCAGACCTCAGAAAGACTAGATACAAGAATTCGACGA
GCTCTACATGTACTGCTACTATGTTGCTGGAAGTGTGGGGCTAATGAGTG
TTCTGTGATGGGTATTGACCCGAGTCGAAGGCAACAACCTGAAAGTGTG
TACAGTGCTGCTTTGGCTCTCGGCATTGCAAAACAGCTCACAATATACT
CCGTGACGTTGGAGAGGACGCGAGAAGAGGGAGGATATATTTACCACAAG
ATGAAGTTGAGAGGACAGGGCTCTCTGATGAGGACATCTTCAATGGCGTT
GTGACTAACAATGGAGAAGCTTCATGAAGAGACAGATCAAGAGAGCTAG
GATGTTTTTTGAGGAGGAGAGAGAGGGGTGACCGAGCTCAGCCAGGCAA
GCCGGTGCCCGGTCTGGGCGTCTCTGTTGTTATACCGGCAAACTCCTTGAC
GAGATAGAAGCAAACGATTACAACAATTCACAAGAGGGGCTACGTTGG
GAAGGCCAAGAAATTGCTAGCGCTTCCAGTTGCATATGGTAGATCATTGC
TGATGCCCTACTCACTGAGAAATAGCCAGAAG - - -

>V

- - - - - TCTGTTGCTCTGTTGTGGATTGTT - - - - - CCCCCCAACAAGA
GGCCTCCAATTCCTTTGGCTTTTTGGAATCAATCCGAGATGGAACCGGT
CTTTAGATTACCATGGTTCATTCTAGAGACCAGAATTTGATATTTAAT
GACAGACCCAAGAAAGGTAGGAGGCAGAGGTGCAATTCGGCTCATCTGG
GGCAGATTTGAGGTATTCATACTTGGGT - - - - - GGATCAGGCATGGAGA
GTAGAACCAATTTTCTATATTTTCAAGTATGGTAGTGAACCCAACCTGGA
GAAATGGCAGTGTCGTCGGAGCAGAAGGTTTATGATGTGGTGCTGAAGCA
GGCAGCCCTTGTTAAGAAGCAATTGAGATCTAGTGAA - - - GAACCTGATG
TGAAA - - - CCT - - - GATATTATTCTTCTGGGACTCAGAGCTTGTGAGT
GAAGCCTATGATCGATGTGGTGAAGTTTGTGACAGATGCCAAGACATT
TTACTTGGGAACAATGCTGATGACACCTGAGAGACGAAGAGCTATCTGGG
CAATATATGTGTGGTGTGCGGAGAACAGATGAGCTTGTGATGGGCCTAAT
GCATCACACATAACCCCAACAGCCTTGGATAGGTGGGAGTCAAGGTTGGA
AGATCTTTTTCAGGGCCGCCCTTTTATGATGCTTGATGCTGCTTTATCAG
ATACAGTTGCCAAGTTTCTGTTGACATTGACCCATTCAAAGAT - - - ATT
GAAGGTATGAGGATGGACCTTAGAAAGTCAAGATACAGAAATTTGATGA
ACTATATCTGTACTGTTACTATGTGGCGGAACTGTTGGGTTAATGAGCG
TTCCAGTTATGGGCATTGCACCTGAATCACAGGCAACCACGGAGAGTGTC
TATAAGGCAGCCTTGGCATTAGGGATCGCAAATCAGCTGACTAACATACT
CAGGGATGTTGGAGAAGATGCAAGGAGAGGAAGGATTTATTTACCACAGG
ATGAGCTGGCACAAGCAGGGCTCTCAGATGAGGACATTTTTATGGGAAGG
GTAACAGATAAATGGAGAAGCTTCATGAAAGATCAAATTAAGAGAGCAAG
GATGTTCTTCGATGAGGCGGAAAAAGGAATTACGGAGCTCAGCTCAGCTA
GTAGATGGCCGGTATGGGCATCTTTCTGTTGTATCGGCAAACTACTGGAT
GAGATTGAAGCTAACGATTACAACAACCTTCACTAGGAGAGCTTATGTGAG

CAAAGCTAAGAAGATAGTTGCTTTGCCACTAGCCTATGCAAGGTCCTTG
TGAGGCCATCAAGAACAACATCTTTAGTCAAGGCA

CRT.CrtIS0.fasta

>A

GATTTGTGTTTTCAAATCCCGTAAAGTGTGGTGATCGTTTTG-----
-----TTCTCCGCATTGAATACCTCTACGTATT
ACAAGCTGGGTACTTTCGAATTTAGGGTTTAATGGTCCAGTTTTAGAGAAT
CGGAAGAAGAAGAAGAAGCTGCCT-----
-----CGTATGGTTACAGTGAAATCCGTTTCTTCAA
GTGTGGTTGCTAGTACCGTACAAGGGACGAAGAGAGATGGAGGAGAGAGT
CTGTACGACGCCATTGTTATCGGTTCTGGGATTGGTGGATTAGTTGCTGC
GACTCAGCTAGCTGTGAAAGAAGCTAGGGTTTTAGTTCTG---AAGTATT
TGATTCCTGGTGGCAGCTCTGGTTTTACGAGAGAGATGGATATACTTTT
GATGTTGGCTCTTCTGTCATGTTTGGTTTCAGCGATAAGGGGAACCTAAA
CTTGATAACTCAGGCATTGAAGGCAGTTGGTCGTAAGATGGAGGTTATAC
CTGATCCTACCACTGTCCATTTCCATCTCCCCAATAATCTTTCTGTTCCG
ATTCATAGGGAATACGATGATTTTCATCGCG---CTTACGAGCAAGTTTCC
TCACGAAAAGGAAGGGATCCTTGGATTCTATGGTGACTGTTGGAAGATCT
TCAACTCATACTAC---TCCCTGGAACATAAGTCACTAGAAGAGCCTATCTAC
CTTTTTGGACAGTTCTTTTCAAGCCGCTTGAATGCTTGACACTCGCTTA
TTACTTGCCCCAAAATGCTGGGGCTATAGCTCGGAAGTACATAAAGGATC
CTCAATTATTGTCTTTTCATT---GCAGAGTGTTCATTGTGAGCACAGTC
AATGCTTTTGCAGACGCCAATGATCAATGCAAGTATGGTTTTATGTGACAG
GCACTATGGAGGAATTAACCTACCCGTTGGTGGAGTTGGTGGGATTGCCA
AGTCTTTTGGCAGAAGGGCTTGTGATCAAGGCAGTGAAATACAATACAAG
GCTAATGTAAAAAGTATAATACTTGATCATGGAAGGCAGTTGGTGTAAAG
GCTAGCAGATGGAAGAGAATTTTTCGCTAAGACGATAATTTCAAATGCTA
CAAGATGGGATACGTTTGGGAAGCTGTAAAAGGAGAAAAGCTTCTTAAA
GAAGAAGAAAATTTCCAGAAAAGTACGTCAAGGCTCCATCATTTCTTTTC
AATTCACATGGGTGTTAAAGCAGAGGTTCTCCCTCCAGATACAGATTGCC
ATCATTTTGTGCTCGAGGACGATTGGAAGAATCTGGAGGAGCCTTATGGC
AGTATTTTCTCAGCATTCCAACCATTTCTTGATTCATCTTTGGCTCCAGA
TGGTCGACATATACTTCACATATTTACAACCTTCGTCCATTGAGGATTGGG
AGGGACTCCCTCCAAAAGAGTATGAGGCGAAAAAGGAAGATGTGGCAGCT
AGAATAATCCAGAGACTAGAGAAAAAATTTGTTTCCAGGGCTCAGTTTCATC
TATTACGTTCAAGGAGGTGGGAACACCAAGAACACACAGACGATTTCTTG
CGAGGGATAAAGGGAACCTTATGGGCCAATGCCTAGAGGAACCCCGAAAGGT
TTATTAGGCATGCCATTCAATACAACGGCTATAGATGGTCTATACTGCGT
GGGGGATAGTTGTTTCTTGGCCAGGGAGTTATAGCTGTGGCTTTTTTCAG
GAGTGATGTGTGCTCATCGGGTAGCTGCTGATATTGGGCTTGAGAAGAAA
TCACGAGTACTTGATGTTGGCCTTCTTGGTTTACTTGGTTGGCTAAGGAC
ACTCGCA

>P

-----AGC
AATTACGATGCAATTGTTATTGGGTCTGGTATCGGTGGGTGGTTGCTGC
TACACAGTTGGCA---GAGGGAGCTAAAGTTTTGGTCTTGGAGAAGTATG
TGATTCCTGGTGGGAGTTCTGGGTATTATGAGAGGGATGGATTACTTTT
GATGTTGGCTCTTCTGTTATGTTTGGTTTCAGTGATAAG-----
-----GCATTGGCAGCAGTTGGCTGTGAAATGGAGGTGATTC
CTGACCCAACCTACTGTCCATTTTCATCTACCCAATGACCTTTCTGTTCAA
GTTACAGAGAGTACATCGACTTCATCTCAGAACTTGCTGCTAAATTTCC
CCATGAAAAGGATGGGATCCTTAAATTCTACGGTGAATGCTGGAAGATAT
TCAATGCCCTGAACTCTCTGGAAGTGAAGTCACTTGAGGAGCCAATCTAC
CTGTTTGGACAAATTTTTTTCAGAAACCTTTGAATGCTTGACACTGGCTTA
TTATCTACCTCAAAATGCTGGAGATATTGCTCGAAAGTACATAAAGGATC
CTCAGTTATTGTCTTTTCATTGATGCAGAGTGTTCATAGTGAGCACAGTC
AATGCTTTGCAGACTCCAATGATCAATGCGGGCCATGGTTCTATGTGATAG
GCATTTTGGAGGGATCAATTACCCTGTTGGTGGGTTGGTGGAAATTGCGA
AGTCCTTATCAAAAGGTCTGGTTGATCAGGGCAGTGAAATACTTTACAGG
GCAAATGTGACCAACATCATTCTTGAGCATGGAAAGGCTGTAGGAGTAAG
GCTTTCAGATGGAAGGGAGTTCTTTGAAAAACCATAATCTCAAATGCTA
CTAGATGGGATACCTTCGGGAAGTTGTTAAAAGGAGAAACCTTCCAAAA
GAAGAAGAAAATTTCCAAAAAGTTTATGTTAAGGCTCCATCTTTTCTTTT
CATTTCATATGGGTGTTAAAGCTGAGGTTCTACCACCAGATACAGATTGCC
ACCACTTTGTGCTTGAGGATGACTGGGCAAGATTAGAGGAGCCTTATGGA
AGCATATTTCTAAGCATTCCAACATTTCTTGATTCGTGCTATTGGCTCCGGA
AGGCCATCATATACTTCACATATTTACAACATCTTCCATTGAAGACTGGG

AGGGACTCTCGACAAAGGACTATGAGGCCAAGAAGAAGGTTGCGGCTGAT
 GAGATTATAAGCAGATTGGAGAAGAACTTTTTCCAGGGATCAGATCATC
 AATTGCTTTTCATGGAGGTGGGTTACCCAAAGACACAGGCGGTACCTGG
 CTCGTGATAAAGGTACTTATGGACCATGCCACGCAGAACTCCTAAAGGT
 TTATTGGGAATGCCATTCAATACGACAGCTGTAGATGGTCTTTACTGTGT
 TGGCGATAGCTGCTTTCCAGGACAGGGCGTGATAGCTGTAGCTTTTTTCAG
 GAGTAATGTGTGCTCATCGAGTAGCTGCTGATATTGGCATTGAGAAAAAG
 TCTCCTGTACTGGATGCTGCTCTCCTTCGGCTTCTTGGCTGGTTAAGGAC
 CTTGGCA
 >R
 AGTTTAGCTTTTCAGTAGCACAGTTCAATTTGGTAGTTTTGGACCCAGAAG
 CCAGAAACATGGATTTTTCTCTTTTGAACCTTCCAAAGTCTCATACCTTT
 TCAATCTCAGTAACATCAATTACAGGTGGTTATGCAATTCACGTCCAGA
 AATTTAGAAATAGAAAGTGCAAGTTTCAATAGTAATAACAAGTTTAGGTC
 TTGCAAGGTGAAGAGGAACTTGGCTTTGAGGACAAAGTCAGTGTGAGTG
 TGAATAAAAAAGTGGATATACAGGGGAATAGA-----GGAGTAAGC
 AAATTTGATGCTATTGTTATTGGTTCTGGGATTGGTGGTTTGGTTGCTGC
 TACACAGCTGGCTGTGAAGGGAGCTAGGGTTTTGGTTTTGGAGAAATG
 TGATTCCTGGTGGGAGCTCTGGTTATTATGAGAGGGATGGTTATACTTTT
 GATGTTGGGTCTTCAGTTATGTTTGGCTTCAGTGATAAGGGCAACCTAAA
 TTTGATAACCCAAAGCATTGGCAGCAGTTGGTCGTAAGATGGAGGTGATTC
 CTGATCCAGTACTGTCCATTTCCTATCTACCAATAATCTTTCTGTTGCA
 GTTCACAGAGAGTACAGTGACTTTGTCTCGGAACCTATTACTAGATTTCC
 CCATGAAGAGCAAGGGATTCTTAAATTCATGGTGAATGTTGGAAGATAT
 TCAATGCCTTAAATTCATTGGAATTGAAGTCACTGGAGGAGCCAATTTAC
 CTGTTTGGGAGTTCTTTTCAGAAGCCTCTTGAATGCTTGACACTGGCTTA
 TTACCTGCCGCAAAATGCTGGAGACGTAGCTCGAAAAATACATAAAGGACC
 CTGAGTTATTATCTTTTATTGACGCGGAGTGTTTTATTGTGACGACAGTT
 AATGCGTTGCAGACACCAATGATTAATGCAGCCATGGTCCTATGTGATAG
 GCATTTTGGAGGAATTAACATCTCTGTTGGTGGTGGTGGTGGGAATTGCAA
 AGTCTTTGGTGAACGGTCTGGTTGATCAGGGGAGTGAAATTCGTACAAG
 GCAAACGTGACTAACATTATACTGGAGCATGGAAGGCTGTGGGAGTGAG
 GCTGTGAGATGGAAGAGAGATCTTTGCCAAAACCATATCAATGCTA
 CTAGATGGGATACCTTTGGGAAGTTGTTAAAGGTGAAAGCTTCCACAA
 GAAGAAGAGAACCTTTCAGAAAGTTTATGTCAAGGCTCCATCTTTCTTTC
 CATTACATGGGTGTTAAAGCTGAGGTTCTGCCTCCAGATACAGATTGCC
 ATCATTTTGTGCTTGAGAATGACTGGGCGAGATTACAGGAGCCTTATGGC
 AGTATATTTTAAAGCATTCCAACGTGCTCGATCCATCGCTAGCCCTGC
 AGGGCACCATATTCTTTCACATATTCACGACTTCTTCAATTGAAGACTGGG
 AGGGACTCCCTGTAAAGGATTATGAGGCAAGAAGCAGCTGGTTGCAGAC
 AAAATTATAGGCAGATTGGAGAAGGAAGTGTTCAGGGCTCAGATCATC
 AATTGCTTTTATGGAGGTGGGGTCACCCAAAGACACAGGCGGTATCTAG
 CTCGTGATAAAGGTACTTATGGACCA---CCATGCAGAACTCCTAAAGGC
 TTATTGGGAATGCCATTCAATACAACAGCTATAGATGGTCTTTACTGTGT
 TGGAGATAGCTGTTTTCCAGGACAGGGTGTAATAGCTGTAGCCTTTTCAG
 GAGTAATGTGTGCTCATCGTGTAGCTGCTGATATTGGACTGGAGAAAAAG
 TCCCCTGTAATGGATGCTGCTCTCCTTCGACTTCTTGGCTGGTTAAGGAC
 ATTGGCA
 >O
 CCGCTCCTCCTCTCGCCGCCGCGCGCCCGTGCGCCCCGCTC-----
 -----CTCGCTC
 CCTCCGCCGCCGAGCAGTAGTAGGCCGCCGTCCACCGCCCCGCTG
 GGACGAGGGACAGGAGAGGCGCG-----
 -----GGCGCGCGCGCGCGCGCTGCGGTGCGCGCGG
 AGAAGACGGTGGTGAAGCGGAGGAGGAGGAGGGTGGGGAGAG---
 ---TACGACGCGATCGTGGTGGGTCGGGGATCGGGGG---GTGGCGG
 GACGCGAGCTGGCGGCGAAGGGGCGAGGGTGCTGGTGGAGAAAGTACG
 TCATCCCCGGGGGAGCTCCGGGTACTACCGCCGCGACGGGTTACCTTC
 GACGTGGCTCCTCCGTCATGTTGGGCTTCTCCGACAAGGGGAACCTTGAA
 TTTGATTACACAAGCACTAGAAGCAGTTGGGCATAAGATGGAAGTTATAC
 CGGATCCTTCTACGGTTTCATTTCCACCTACCTGGTGACCTCTCAGTTCTT
 GTGCATAGAGAGTATGATGACTTCGTACGGAGCTGGTTAATAAATTTCC
 TCATGAAAAGGAAGGAATCCTGAAATTCATGGCACATGTTGGAAGATCT
 TCAATTCATTAAATTTCTTGGAGCTGAAGTCCCTTGAGGAGCCTCTATAT
 CTTTTTGGGCAATTTTTCAAAAGCCTCTGGAATGTTTGACGCTAGCTTA
 TTATCTGCCACAGAATGCTGGGATATTGCTCGCAAGTTCATAAAAGATC
 AGCAGTTACTCTCCTTCATAGATGCTGAGTGCTTTATTGTGAGCACAGTC
 AATGCCCTTGCAAAACCCATGATTAATGCAAGCATGGTCTTGTGCGATAG
 GCACTTCGGGGGAATTAACATCCAGTTGGTGGTGGTGGTGGTATTGCT-
 --TCCTTGGCAGATGGGCTTGTGACAAAGGCAGTGAAATACGCTACAAG
 GCTAATGTGACCAATGTAATTTAGAAAATGGCAAAGCTGTTGGAGTCAG
 GTTATCAAAATGGGAAGGAGTTCTTTGCTAAAACAGTGATATCGAATGCTA
 CCAGATGGGATACATTTGGTAAACTCTTGAAAGTTGAAGAGCTTCAGAA
 GAGGAGAAAAAATTCAAAAGAATTATGTTAAAGCACCATCATTCCTTTC
 CATTCATATGGGTGTCAAAGCCTCAGTTTTGCTGCTGATACCTGATTGCG

ACCATTTTGTACTTGAGGATGACTGGGCTAATTTGGAGAAGCCTTATGGA
 AGCATATTTTTAAGTATCCCTACAGTGCTTGACCCATCCTTGGCTCCCGA
 AGGACACCATATACTTTCATATTTTTACAACGTCAGGGATAGAGGATTGGG
 AAGGTCTGTCCAGGAAGGATTATGAGAAGAAAAAGGAGCTTGTGGCGACA
 GAGATAATAAAGAGGCTTGAAAAAAGCTGTTCCCTGGTCTTCAGGATTTC
 AATAGTCCTCAAGGAGGTAGGCTCACCAAAAACACACCGAAGATTTCTTG
 CGCGAAATGATGGTACATATGGACCCATGCCACGGGGTAAACCAAAGGGA
 TTGCTAGCA--CCTTTCAATACTACTTCAATAGATGGCTTGTACTGTGT
 TGGCGACAGCTGTTTTCTGGGCAAGGTGTAATTGCTGTGGCATTTTCAG
 GGATCATGTGTGCTCATCGAGTTGCAGCTGATATTGGACTGGAACAAGA
 TCTCCTGTACTAGATGCGGGTCTTCTTGGGCTCCTCAGATGGTTGAGAAC
 GCTTGCT

>V
 -----GGATTGGAGGGT-----
 -----T
 ACAAGCATGGTGGTGTGAATTAGGGGTC-----
 TATAGGTGGAGTAGGAGTGGAAGC-----
 -----TGTGTTGTGAGGCCTAGATCAGTGCTGAGTG
 TGGATGAAGTGGTAGAGAGGGAAGGGAATGGAAGAACAGGAGAGGAAAAAG
 CAATACGATGCCATTGTTATCGGGTCGGGAATTGGAGGGTTGGTTGCGGC
 AACACAGCTGGCAGTGAAGGGAGCTAGGGTGTGGTGCTGGAGAAGTATG
 TGATTCTCTGGTGGGAGCTCTGGGTTCTACCAGAGAGATGGGTATACGTTT
 GATGTTGGGTCTTCTGTGATGTTTGGTTTCAGCGATAAGGGAAATTTAAA
 TTTAATAACCCAAGCACTGGAAGCAGTTGGATGTAAGATGCAGGTGATAC
 CTGATCCAACACTGTCCATTTCCATTTACCCAGTAACCTTTTCGGTCCGA
 GTACATAGACAATACGGTGAATTCATTGCAGAACTTACCAAGAATTTTCC
 CCATGAAAAGGAAGGGATCCTCAAATTTACAATGAATGTTGGAAGATCT
 TCAATGCCCTAACTCATTGGAATTGAAGTCACTTGAGGAGCCAATCTAC
 CTTTTCGGACAGTTCTTTTCAAGAGCTCTTGAATGCTTGACACTTGCCTA
 TTATTTGCCCCAAAATGCTGGAGACATAGCCCGCAAGTACATAAAGGATC
 CCCAGCTGTTATCTTTCATAGATGCTGAGTGTATATAGTGAGCACAGTT
 AATGCTTTGCAGACACCAATGATCAATGCGAGCATGGTTCTATGTGATAG
 GCATTTTGGGGGAATTAACCTACCCTGTTGGTGGTGTGGTGGGAATTGCTA
 AGTCCTTGGCAAAAGGTCTAGTTGAAAATGGCAGTGAATACTTTATAAG
 GCAAATGTGACTAGCATTATATTGGCGCAGGGAAGGCCGTGGGAGTGAA
 GCTTTGAGATGGAAGGGAGTTCTTTGCCAAAACCATAGTATCTAATGCTA
 CTAGATGGGATACATTTGGAAAGCTTTTAAAGGGTGAAGAACTTCAAAA
 GAAGAGGTGAACCTTTCAAAGGTCTATGTTAAGGCCCATCATTTCTTTTC
 CATTACATGGGGTTAAAGCTGAGGTTTGGCACCTGATACAGATTGCC
 ACCATTTTGTCTTGAGAAATGACTGGTCAAGGTTAGAGGAGCCATATGGA
 AGCATATTTTGAGCATTCCAACATGCTTGATGCATCATTGGCTCCAGA
 AGGTGCGCATATTCTTCATATTTTACAACCTCTCCCATAGAGGACTGGA
 AGGGACTCCCTCTAAAAGACTATGAGGCAAAGAAAGAACTTGTGGCAGAT
 GAAATCATAAAGCAGATTGGAACAAAAAATATTTCCGGGGCTCAAATCATC
 CATTGTTTTTAAGGAGGTAGGGACACCAAAAAACACACAGGCGTTATCTTG
 CTCGAGATAATGGCACGTATGGACCAATGCCTCGCAGAACACCTAAGGGA
 TTATTGGGAATGCCATTCAATACAACCTGCTATACATGGTCTCTACTGTGT
 TGGGGATAGCTGCTTTCCAGGACAAGGTGTTATAGCTGTAGCCTTTTCAG
 GTGTAATGTGTGCACACCGAGTGGCTGCAGATATTGGGCTTGAGAAGAAG
 TCCCAAGTGCTGGATGCTGCCCTTCTTCGACTACTTGGTTGGTTAAGGAC
 ATTGGCA

CRT.CrtL.b.fasta

>A
 -----GATACTCTGTTGAAAACACCCAACAAGCTCGATTTTTTCAT
 CCCTCAGTTTCATGGGTTT---GAGAGA-----
 -----TTATGCAGCAATCCATAC-----CATTCAAGGGTT
 AGGCTTGGTGTGAAGAAAAGGGCTATCAAAATTGTC-----TCTAGTGT
 AGTGAGTGGTAGCGCTGCTCTTTTGGATCTTGTTCTCGTAAACTAAGAAGG
 AGAATCTTGACTTTGAGCTTCTTTGTACGACACTTCCAAGAGTCAAGTT
 GTTGATTTGGCTATTGTTGGTGGTGGTCTGCTGGTTTAGCCGTGGCTCA
 GCAGGTTTCTGAAGCTGGACTCTCTGTTGTTCCATTGATCCTTCTCCTA
 AGCTCATATGGCCTAACAATATGGAAGTTTGGGTTGATGAGTTTGAAGCT
 ATGGATTTACTAGACTGCCTGGATACCACATGGTCTGGTGCTGTTGTCTA
 TGTCT---GATGAAGGTGTCAAGAAGGATTTGAGCCGGCCTTATGGGAGAG
 TTAACCGGAAACAGCTCAAATCCAAAATGCTTCAGAAATGTATTACCAAC
 GGTGTTAAATTTTCATCAGTCTAAGGTCACTAATGTGGTTTACAGAGGAGGC
 AAACCTCACTGTGGTCTGCAAGTGACGGTGTAAAGATTGAGGCTTCCGTGG
 TTCTTGATGCCACTGGGTTTTTCCCGATGCTTGGTTTCAATATGACAAACCT
 TACAACCTGGGTACCAAGTAGCTTACGGGATTGTAGCTGAAGTTGATGG
 TCACCCATTTCGATGTAGACAAAATGGTGTTCATGGATTGGAGAGACAAAC
 ATCTGGACTCATATCCTGAGCTGAAAGAACGGAACAGCAAGATCCCAACG
 TTCTTGTACGCTATGCCATTTTCTTCCAACCGAATATTTCTTGAAGAAAC

TTCTTTAGTTGCTAGACCTGGTCTGAGAATGGAAGATATCCAAGAAAGAA
 TGGCTGCTAGACTGAAACATCTGGGGATCAATGTGAAGAGGATTGAGGAA
 GACGAGCGTTGTGTGATCCCGATGGGCGGTCTTTACCACTTACCTCA
 ACGGGTTGTGGGGATTGGTGGGACAGCAGGAATGGTTCATCCTTCAACTG
 GTTACATGGTTGCTAGGACTCTTGCAGCTGCACCAATAGTTGCAAAATGCC
 ATTGTGAGATACCTC-----GGTTCACCAAGTAGTAATAGCCTGAGAGG
 AGATCAACTCTCTGCTGAGGTTTGGAGAGACTTGTGGCCTATCGAACGGC
 GTAGACAGAGGGAGTCTTCTGTTTTGGAATGGATATTCTGCTGAAACTC
 GATTTAGAGCTACTAGAAGTTCTTTGATGCATTCTTTGATCTGCAACC
 TCATTACTGGCACGGATTCTTGTCTTCCAGGCTGTTTCTCCCGAACTGT
 TGGTCTTCGGGTTGTGCTCTTCTCACACGCTTCCAATACCTCAAGATTG
 GAGATCATGACAAAGGGGACTGTTCTCTTGCTAAGATGATCAACAATTT
 GGTACAAGATAGAGAC---
 >P
 -----ATGGATACTTTGTTGAAAACACATAACAAGATTGAA---TTTTT
 GCCTCAATTTTCATGGGTTTTTCGGAGAAA-----
 -----GTGAGTAATTCGAGCTCCGTAAAGATTCAAATCCAGGAACCT
 AGGTTTGGTCCCAAAAAGTTTTTTTCAAAGTGGGGAGGAATGGTTGTGT
 TGAGGCTAGTAGTAGTGCACCTTTTGGAGCTTGTACCAGAAACCAAGAAGG
 AAAATCTTGAGTTTGATCTTCCATGTATGACCCATCGAAAGGCTTGTGA
 GTTGACCTTGCAGTTGTGGGGGAGGTCCTGCAGGGCTTGTGTTGCTCA
 GCAAGTTTCGGAGGCAGGGCTCTCGGTTTGTTCGATTGATCCATCTCCTC
 AGTTGATTTGGCCTAATAATTATGGTGTGGGTTGATGAATTTGATGCC
 ATGGGTTTGTGCTGACTGCCCTTGATACCACTTGGTCTGGTGTGTTGCTA
 TGT---AATGACAAGACAAAAAAGATCTTGATAGGCCTTATGGGAGGG
 TTAATAGGAAGCAGCTGAAGTCCAAAATGTTACAGAAATGCATATCCAAT
 GGTGTTAAGTTTACCAAGCTAAAGTTATCAAGGTTATTCATGAGGAGTC
 CAAATCTCTATTGATTTGCAATGATGGTGTACAATCCAAGCTGTCGTGG
 TTCTTGATGCAACTGGTTTTCTAGATGCCTTGTTCATATGACAAGCCA
 TATAATCGGAGTTTACCAAGTGGCTTATGGAATTTTGGCAGAGGTAGAAGA
 GCACCCGTTTGTGTTGATAAGATGGTTTT---GATTGGAGAGATTAC
 ATCTGAACAACATCTGGAACATAAAGAGAGAAACAGCAAGACCCCTACT
 TTCCTCTACGCAATGCCCTTTTCATCGGACAGGATATTTCTGGAAGAAAC
 TTCCTCTCGATGCTAGGCCGTGGAGTACCCATGAAAGATATACAGGAAAGGA
 TGGTGGCCAGATTAAGGCACCTAGGTATAAAGTGAAAAGCATAGAGGAA
 GATGAGCGTTGTGTCATTCCAATGGGGGGACCACTTCTGTGCTCCCTCA
 AAGAGTAGTTGGAATTGGTGGTACAGCCGGTATGGTGCACCTTCAACTG
 GGTATATCGGTAGCAAGAAGCTTAGCAGCTGCTCCGATTGTTGCTAATTCT
 ATCGTTCAAGTATCTT-----GGTTCT-----GATAGAAGCTTTTCAGG
 AAGTGAGTTATCAGCTAAAGTTTGGAAAGATTATGGCCCGTCGAGAGGA
 GGAGACAAAGAGAGTTCTTCTGTTTTGGTATGGATATTTGCTTAAGCTT
 GATTTACCTGACCAAGAAGATTTTTTGTGATGCAATTTTTGATCTGGAACC
 TCATTATTGGCACGGATTCTATCATCTCGACTGTTTCTACCTGAGCTTG
 TACTTTTTGGGCTTTCCTGTTCTCCCATGCTTCTAATACTTCTAGGTTA
 GAGATTATGGCAAAGGGAACCTCTCTTTGGTTAATATGACCAACAATTT
 AATACAGGACAGAGAA---
 >R
 -----GATACTTTATTGAGAACGCATAATAAACTTGAA---TTCTT
 GCCTCAACTTCATGGTTTTCTCGGAAAAAGTCAGTATTTTGAACCAAAAA
 AAGAGAAAAATTAGCAATTCGAATTTTATAAAGTATCAGAACTCTGAGTT
 AGGTTTTGTTCCAAAGAACTTATCCAAAAGGTGCAGAAATGGATGTGT
 TAAGGCTAGTAGTGGTGGCCTTTTGGAGCTAGTACCTGAAACCAAGAAGG
 AAAATCTTGAATTTGAGCTTCCATGTATGACTCATCAAAGGTGGTGT
 GTTGATCTTGCTGTTGTGGGTGGTGGCCGGCAGGCCTCGCAGTTGCCA
 GCAAGTTTCAGAGGCAGGCCTTTCTGTTTGTCAATCGACCCATCTCCTA
 AATTGATTTGGCCAAATAATTATGGTGTGGGTGGATGAATTTGAGGCC
 ATGGATTTGCTTGATTGTCTTGATACCACTGGTCTGGTGTGTTGTCTA
 CATT---GATGATAAATCGAAGAAGGATCTTGGTAGGCCTTATGGGAGGG
 TTAACAGGAAGCAGCTAAAGTCGAAAATGCTACAAAATGTATATCCAAT
 GGTGTTAAGTTTACCAAGCCAAAGTTATCAAGGTTATTCATGAGGAATC
 AAAATCCCTATTGATTTGTAATGATGGTGTACAATCCAGGCTGCTGTGG
 TTCTTGATGCAACTGGATTTTCTAGATGCCTTGTTCAGTATGATAAGCCA
 TACAATCCAGGTTACCAAGTGGCTTATGGAATATTGGCTGAGGTAGAAGA
 TCACCCATTTGATGTAGATAAAATGCTTTTTATGGACTGGAGAGACTCAC
 ATTTGAAAAAATATCCGGAGCTACAAAAGAGAAATAGTAAGATCCCACT
 TTTCTGTATGCA---CCTTTTTCATCAGAGAGGATATTTCTCGAAGAAAC
 TTCCCTAGTTGCTCGGCCTGGTGTACGTATGGAAGATATCCAGGAAAGAA
 TGGTAGCTAGGTTGAAGCACTTAGGAATAAAGTGAAAAGTATTGAAGAA
 GATGAGCATTGCGTCATCCCAATGGGTGGCCCCCTCCCGTCTTCTCA
 AAGAGTTGTTGGCATTGGTGGCACAGCTGGCATGGTTCACTTCAACTG
 GGTACATGGTAGCAAGAAGCTTAGCGGCAGCTCCAGTTGTTGCCAATTCC
 ATAGTCCAGTACCTT-----GGTTCC-----AGTAGAAGTCTTTCAGG
 AACCGAATTGTGACGAGAAGTTTGGAAAGATTATGGCCAATAGAGAGGA
 GGAGACAAAGAGAGTTCTTCTGTTTTGGCATGGACATTTTGTCAAGCTT
 GATTTACAAGCCACAGGAAGTTCTTCGATGCATTTTTGATCTTGAACC

```

CCATTATTGGCACGGTTTCTTGTCTATCTCGTCTATTTCTCCCCGAGCTTA
TATTTTTTGGGCTGTCATTGTTTTCCCATGCTTCTAATACTTCCAGGTTA
GAGATTATGGCAAAGGGAACGCTTCTTTGGTTAACATGGTCAACAATTT
AATACAAGATAGTGAA---
>0
GCCACCACGCCCTCCTCTCCGCGCCACCCCTCCTGCAAG-----
-CCCCCT-----
-----CCCCCTCCTCGCCGTCG-----
-----CCGCGCCACGCGCGCTCGTCTGCCGC-----GC
CGCCGCCCGCGGCGAGGCGCTGCGGTGCTGGCCCCGCCGTACGCCCCG
AGCTGCTCTCCCTCGACCTCCCCGCTACGACCCCGCCGCTCCACCCCG
GTCGACCTCGCCGCTGCTCGGCGGCGGCCCGCCGCTCGCCGTCGCGCA
GCGCGTCCGCGGAGGCGGGCTCTCCGTGCGCCATCGACCCCTCCCCG
CCCTCGTCTGGCCCAACAACACTACGGCGTCTGGGTCGACGAGTTCGACGCC
ATGGGACTCTCCCACTGCTCGACGCCGTCTGGCCCTCGGCCACCGTCTT
CACCCACGACGACGGCGCGCCAAAGTCGCTCCACCGCCCTACGCCCGCG
TCGCCCGCGCAAGCTCAAGTCCACCATG---GACCGTGCCTCGCCCAT
GGCGTCACGTTCCACAAGGCCAGGGTCGTCAAGGCCGTCCACGGCGAGGC
ATCTCCCTCCTCATCTGCGACGACGGCGTCGCCGTCCCGGCCACCGTCG
TGCTCGACGCCACGGGGTTCTCCCGGTGCCTCGTCCAGTACGACAAGCCG
TACGACCCGGGTACCAAGTGCCTATGGCATCCTCGCCGAGGTGGACGG
ACACCCGTTTCGACATCGACAAGATGCTGTTTCATGGACTGGCGCGACGCGC
ACCTCCCCGAGGGGTCCGAGATCAGGGAGCGCAACCGCCGCATCCCGACG
TTCTCTACGCCATGCCCTTCTCCCGACGAGGATCTTCTCGAGGAGAC
CTCCCTCGTGGCGCGCCCGGGCTCGCCATGGACGACATCCAGGAGCGC-
--GCGGCGGAGGCTGCGCCACCTCGGGATACGCGTCCGCGCCGTGGAGGAG
GACGAGCGGTGCGTCATCCCATGGGCGGCCGCTCCCGGTGCTCCCGCA
GCGGGTCGTGCGCATCGGCGGACCGCCGGGATGGTGACCCGTCACGG
GCTACATGGTGGCGCGCACCCCTCGCCACTGCGCCCATCGTGGCGGACGCC
ATCGTGCCTTCTCTCGACACCGGCAAGC---GGCGACAGCGCGTTCGCCGG
CGACGCGCTGTCGGCGGAGGTGTGGAGGGAGCTGTGGCCGGCGCAGAGGA
GGAGGCAAGAGGAGTTCTTCTGCTTCGGCATGGACATCTCTCAAGCTC
GACCTCGACGGCACGCGGCGATTCTTCGACGCCCTTCTCGACCTGGAGCC
GCGCTACTGGCACGGCTTCTGTGCTCGAGGCTCTTCTTCCGGAGACTCG
CCATGTTTGGCCTCTCCCTCTTCGCCAAGGCCCTCCAACACGTGCGGCTC
GAGATCATGGCCAAGGGCACGCCCTCTCGCCAAGATGATCGGCAACCT
CATCCAGGACAGAGATAGG
>V
-----GATACTTTACTCAAGACTCATAATAAGCTTGAA---TTTCT
GCACCCACTTCATGGGTTTGCGGAGAAA-----
-----CTGGGCAATTTGACCTTTCCAAAGCTCCAAAACCAGGAGTTT
AGATTGTTCCGAAGAAGTCCAATCTGAAATGGGGTAGAAATGGGTGTGT
TAAGGCCAGTAGTAGTGCCCTTTTGGAGCTTGTTCAGAACTAAGAAGG
AGAATCTTGAGTTTGAGCTTCCATGTATGACCTTCAAAGGGCCTTGTA
GTCGACCTTGCAAGTTGTGGGAGGTGGCCCTGCCGGGCTTGCCGTGCGCA
GCAAGTTTCAGAGGCAAGGCTTTCAAGTCTGCTCGATTGACCATCTCCCA
AATTGATTTGGCCCAATAACTATGGTGTTTGGGTGGATGAGTTTGAGGCC
ATGGATTTGCTTGATTGTCTTGACACTACTTGGTCTGGTGCCGTTGTTTT
CATT---GATGATCACTCAAAGAAGGATCTTGGTAGGGCTTATGCAAGGG
TTAACAGGAAGCAGCTGAAATCGAAAATGATGCAGAAATGCATATTAAT
GGTGTAAGTTTCATCAAGCTAAGGTTATAAAGGTTATTCATGAGGAATC
CAAATCTCTGTTGATTTGCAATGATGGGTCACAATTCAGGCTGCTGTAG
TTCTTGATGCTACGGGTTTTCTAGATGTCTTGTTCAGTATGATAAGCCC
TATAATCCAGGTTACCAAGTTGCTTATGGGATTTTGGCAGAAGTGGAAGA
GCACCCATTTGATGTGGATAAG---GTTTTTCATGGACTGGAGAGATTCTC
ATCTGAACAACAATATGGAAGTAAAAATAGAAATAGTAGGATCCCCACT
TTTCTATATGCAATGCCTTTTTTCATCTAACGGGATATTTCTTGAAGAAAC
TTCTCTAGTAGCTCGACCAGGAGTGCCATGGAAGATATTCAGGAAAAGGA
TGGTAGCTCGGCTGAGGCACTTAGGCATAAAAGTCAAAGCATTGAAGAG
GATGAGCGTTGCATCATTCCAATGGGTGGGCCCTCCAGTGCTCCCTCA
AAGAGTTGTTGGCATTGGTGGAAACAGCTGGAATGGTCCACCCGTCGACTG
GGTACATGGTAGCAAGGACTTAGCAGCAGCTCCAATTGTTGCAAATTCT
ATAGTTCAGTACCTT-----GGCTCT-----GACAGAAGCTTTTTTG
CAATGAATTGTCTTCTGAAGTTTGGAGAGATCTATGGCCAATCGAAAGGA
GGCGTCAAAGGGAATCTTTTGTGTTGGTATGGATATCCTTCTTAAGCTT
GATTTACAAGGAACAAGAAGTTTTTGTGATTCATTCTTGTATCTAGAACC
TCGTTATTGGCATGGATTCTTGTCTATCTCGACTGTTTCTTCTGAGCTCA
TATTTTTTGGGCTTCCCTGTTCTCTCATGCATCTAATACTTCTAGGATA
GAGATTATGGCAAAGGGTACTCTTCTTTGATAAACATGATCAACAACCT
AATACAGGACAAGGAT---

```

CRT.CrtR.b.fasta
>A

```

---GCAGCAGGACTATCAACAATCGCCGTAACACTCAAACCA---CTCAA
CCGCTCTTCTTTCTCCGCAAACCAC-----CCTATA---TCCACCG
CAGTTTTCCCTCCATCTCTAAGATTCAACGGCTTCCGA-----
-----CGTCGAAAGATTCTACCGTTTGTTCGTCGTTGGAAGAACG
GAAACAA-----AGCTCTCCATGGACGATGACAACAACCAGAAAGCA
CGACAAGCTCTTCTGAGATCTTGATGACGTCGCGGCTTCTCAAGAAAGCG
GAGAAGAAGAAATCAGAGAGGTTCACTTATTTAATAGCAGCTGTGATGTC
GAGCTTTGGTATCACTTCTATGGCTATAATGGCTGTTTATTACCGATTTT
CTTGGCAAATGAAGGGA---GGTGAAGTGTCTGTGTTAGAGATGTTTGGT
ACATTTGCTCTTTCCGTTGGTGTGCCGTTGGGATGGAGTTTGGGCAAG
ATGGGCTCATAGAGCTCTTTGGCATGATTCTTTATGGAACATGCATGAGT
CACATCACAACCAAGAGAAGGAGCGTTTGAGTTAAACGATGTGTTTGCT
ATAACAACCGCGGTTCTGCGATTGGTCTACTCTATTATGGATTCTTGAA
TAAAGGACTTGTCTGCTGCTTCCGTTGCGGACTGGGAATAACGA
TGTTTGGAAATGGCTTACATGTTTCGTTACGATGGACTTGTGCACAAGAGA
TTCCCTGTAGGTCCTTGGCAACGTTCTTACCTTCGAAAGGTCGCCGC
CGCTCACCAGCTACACCACAGACAAATTCAAAGGTGTACCATATGGGC
TGTTTCTTGGACCAAGGAAGTAGAAGAAGTGGGAGGAAAAGAGATTG
GAGAAAGAGATAAGCCGGAGAATCAAATTATACAACAAGGGTCTTCTAC
CTCT
>P
ATCGCGGCAAGGACTCACCGCCGCCACTGTACCCAAACCTTCCGGTACAA
CTCTGTCTCTCACCTCCTTCCAAA-----CCAGTAACAGCAGCCT
CACTTTTCTTCCCAACATTAGACACCAAGCTTTTACACTATGGAACA
AAAGTCCAAGAAAAACAAGCCTTGCTGTATGTTTGTAGTTGAAGATCA
AACAAAACCGAGCAGTGCCATATCGAGAATCAACAAGAAGAGGTGCCA
AAGATGTTAATGAAAATCAGATCTCGACACCTCGCGTGGCAGAAAGATTG
GAAAGAAAGAGAAAAGAAAGAGTTACTTATTTGATTGCTGCTGTTATGTC
TAGTCTTGGGATTACTTCCATGGCTGTTATGGCTGTTTATTACAGATTTT
ATTGG---CTTGAGGGA---GGGGAAGTGCCCTTGCTGAAATGTTTGGT
ACATTTGCTCTTTCAGTAGGTGCTGCTGTGGGATGGAATTTGGGCGAG
ATGGGCTCATAGAGCACTTTGGCATGCTTCTTTGTGGCATATGCATGAGT
CTCACCATAGACCAAGAGATGGGCCGTTTGAGCTAAACGATGTTTTGCC
ATAACCAATGAGTCCAGCAATTTCCCTTCTCTCTATGTTTCTTCAA
CAAGGGCCTTGACCTGGTCTTTGTTTGGTGTGGTCTTGAATTACAG
TGTTTGGCATAGCCTATATGTTTGTCCATGATGGCCTTGTTTACAAGAGA
TTCCAGTGGGGCCATTGGCGATGTCCCATACTTCAAGAAAGTAGCAGC
AGCTCACAAAGATCCACCACTCGGACAAATTCAATGGTGTCCCATATGGGT
TGTTTCTAGGACCTAGGGAATTGAGGAAGTTGGAGGCCTTGAAGAATTG
GAAAGGGAGATAAGTAGGAGAACCAAATCATACAAGGAGTTG-----
-----
>R
---GCGGCTGGACTATCCGCCGCGCCAGTCTTTAAGCCTTTCCGTTATAT
AAACGCCTCCTACATCCTCCCAAAG-----CCTCTG---CCCACCTT
CACTTGTCCTCCTCTACTCGATACAACAGCTTATTTCTTGGCGGTGCA
-----AGAAAACTGGCTTCGCTATCTGTTTTGTTCTTGAAGATTT
GAAACAG-----AGCGTCGAAACTGAGAATCCCGTAGAGGAAGAGAATG
AGAAAGTAAAC---TACCAGATCTTGACGCCACGAGTAGCTGAGAGATTG
GCTAGGAAGAGATCGGAGCGGTTTACTTACTTGGTTGCGGCTGTTATGTC
TAGTTTTGGGATTACTTCTATGGCTGTTATGGCTTGTACTACAGTTTTT
ATTGGCAGATGGAGGGT---GGAGAAGTTCCTTTGGCAGAAATGTTTGGT
ACATTTTCTCTATCCGTAGGGGCTGCTGTGGGTATGGAATTTGGGCAAG
ATGGGCTCATAGAGCTCTTGCGCATGCTTCCCTGTGGCATATGCACGAGT
CTCACCATCGACCAAGAGAAGGTCCGTTGAGCTAAACGATGATTTTGCA
ATTACAAATGCAGTTCAGCAATTGGCCTTCTCTACTATGGATTCTTCAA
CAAAGGCCTTTTTCTGGTCTTTGTTTGGAGCTGGTCTTGGCATTACCG
TGTTTGGGATGGCTACATGTTTGTCCACGATGGTCTTGTTCACAAGAGA
TTCCAGTGGGTCCCATTTGCCAACGTGCCCTATTTAGAAAAGGTAGCCGC
TGCCCATCAGCTTCATCACTCTGAGAAATTCAATGGCGTCCCATATGGGT
TGTTTCTAGGACCAAGGAAATTGAAGAAGTTGGAGGCCTAGAAGAATTG
GAAAAGGAGACAAGTAGGAGAAAAAATCATACAATGGTAGA-----
-----
>O
GCCGTGCGGAGGCTGGTGGCCGCG-----
-CGGGCCCTCTCCTCTCCCCCGCC-----GCGGTGGCGGCGGCAC
ACCGATCGCCCCCGCGCTGCTGCGGCTGGCGTTGCGCCGCTCCCGCG
-----CGGCGGCTGGCCGTTCCCTTGCGCGTGGCGGTGGGTGAGCC
GGAGCCG-----GAGGAGGATGCGC
GGAGGGCG-----GTGGCGGAGCGCGC
GCGCGGAAGCAGTCGGAGCGGCGGACGTACCTGGTGGCAGCGGTGATGTC
CAGCCTCGGGATCACGTCCATGGCCGCCGCGCGCTACTACCGCTTCG
CCTGGCAAATGGAGGGCAGCGCGAGATCCCGGTCACGGAATGTTCCGGC
ACCTTCCGCTATCGGTGGGCGCGCGGTGGGATGGAATTTGGGCGCG
GTGGGCGCACCGGGCGCTGTGGCACGCGTGTGGTGGCACATGCACGAGT
CGCACACCGCCGCGAGACGGGCGTTGAGCTCAACGACGCTTTCGCC

```

ATCACCAACGCCGTCCCGGCCATGTCCCTCCTCGCCTACGGCTTCTTCAC
CCGCGGCCTCGTCCCGGCCTCTGCTTCGGCGCGGGACTCGGGATCACGC
TGTTTCGGGATGGCGTACATGTTTCGTCCACGACGGCCTGGTCCACCGCCGC
TTCCCGGTGGGGCCCATCGCCAACGTGCCCTACTTCCGGCGGGTCCCGGC
CGCCACACAGATACATCAC --- GACAAGTTCGAAGGTGTGCCATATGGGC
TGTTCTTGGTCCCAAGGAGCTGGAGGAGGTGGGTGGGATTGAGGAGCTG
GAGAAGGAGATCAAGAGGAGGATTAAGAGGAAAGAGACCTTAGATGCGAT
CCAA

>V

--- GCGGTGCAAAATCTCGGTTGCCACCTCCTCCCGCTCTCCCGCTTGGG
CCGGAACCCCTTTCTCGGCCCAAACCCACCTCCCTTACCCCCACTT
CTCTCTTCATCCCTTCCATTTCGGCGGCACGAAACATTTTCCGGTGTCCG
----- AAGAAGACAAGACTGACGGTTTGTTCGTGGTGGAGGAGGA
GAAGCTG-----AGCACAGAGGTG-----GTGGAGAATCGGAGCG
AGGAGACTCTGGCGAGTCAGATCTCCGCGCCCGGGTGGCGGAGAAGCTC
GCCAGGAAGAGATCAGAGCGGCTAACGTACCTTGTTCGGGCTGTGATGTC
GAGCTTTGGGATCACTTCCATGGCTGTTATGGCAGTTTACTACCGTTTTT
CATGGCAAATGGAGGGC---GGGGAGGTTCCCTTGTCCGAGATGTTCCGG
ACATTTGCTCTGTTCGGTTGGAGCCGCTGTAGGCATGGAGTTCTGGCGAG
ATGGGCTCACAAAGCTCTCTGGCACGCCCTCTTATGGCACATGCACGAGT
CCCACATAAACAAGAGAAGGTCCATTGAGCTCAACGATGTTTTCCGCG
ATCACAATGCAGTTCCAGCCATAGCCCTCCTCTCTATGGCTTCTTCCA
CAAAGGCCTCGTCCCTGGTCTATGTTTTGGTGCAGGCCTTGAATTACCG
TGTTTGGAAATGGCTACATGTTTGTCCACGATGGTCTCGTCCACAGGAGA
TTTCCAGTGGGGCCATTGCGAACGTGCCCTATTTAGACGCGTAGCTGC
AGCACACCAACTCCACCACTCAGACAAATTCATGGGGTCCCATATGGGC
TGTTCTTGGGCCCTAAGGAAGTGAAGAGGTGGGAGGCAAGGATGAGTTG
GAGAAAGAGATCAATCGCAGAGCCAGATTATCCAACGGTCCAAGA-----

CRT.PDS.fasta

>A

GTTGTGTTTGGGAATGTTTCTGCGGCGAATTTGCCTTATCAAAACGGGTT
TTTGGAG-----
-----GCACTTTCATCTGGAGGTTGTGAATAATGGGA
CATAGCTTTAGGGTTCCCACTTCTCAAGCGCTTAAGACAAGAACAAGGAG
GAGGAGTACTGCTGGTCTTTGCAGGTAGTTTGTGTGGATATTCCAAGGC
CAGAGCTAGAGAACACTGTCAATTTCTTGAAGCTGCTAGTTTATCTGCA
TCCTTCCGTAGTGCTCCTCGTCTGCTAAGCCTTTGAAAGTTGTAATTGC
TGGTGTGCTGATTGGCTGGATTGTCAACTGCAAAGTACCTGGCTGATGAG
GCCACAAACCTCTGTTGCTTGAAGCAAGAGATGTTCTTGGTGGAAAGATA
GCTGCATGGAAGGATGAAGATGGGGACTGGTATGAGACTGTTTACATAT
TTTCTTCGGTGCTTATCCGAATGTGCAGAATTTATTTGGAGAAGTTGGGA
TCAATGATCGGTTGCAAGTGAAGGAACACTCCATGATTTTTGTATGCCA
AGTAAACCTGGAGAATTTAGTAGATTTGACTTCCCAGATGTCCTACCAGC
ACCCTTAAATGGTATTTGGGCTATTTTGCAGAACACGAGATG---CCAT
GGCCAGAGAAAAAAAGTTTGCTATTGGACTTTTCCAGCCATGGTCGGC
GGTCAGGCTTATGTTGAGGCCCAAGATGGTTTATCAGTCAAAGAATGGAT
GGAAAAGCAGGCAATACCTGAGCGCGTGACCGACGAGGTGTTTATTGCCA
TGTCAAAGGCGCTAAACTTTATAAACCTGATGAACTGTCAATGCAATGC
ATTTTGATAGCTTTGAACCGGTTTCTTCAGGAAAAACGT---TCCAAGAT
GGCACTTCTGGATGGTAATCCTCCGGAAGGCTTTGTATGCCAGTAGTGG
ATCATATTGATCACTAGGTGGGGAAGTGCAACTTAATTCTAGGATAAAG
AAAATTGAGCTCAATGACGATGGCACGGTTAAGAGTTTCTTACTCACTAA
TGGAAGCACTGTGGAAGGAGACGCTTATGTGTTTGCCGCTCCAGTCGATA
TCCTGAAGCTCCTTTTACCAGATCCCTGGAAGAAAATACCGTACTTCAAG
AAATTGGATAAATTAGTTGGAGTACCAGTTATTAATGTTTCATATATGGTT
TGATCGAAAACCTGAAGAACACATATGATCACCTACTCTTATGAGCAAGTA
ACCTTCTGAGCGTGTATGCCGACATGTCCTTAACTTGTAAAGGAATATTAC
GATCCTAACCGGTCAATGCTGGAGCTAGTATTTGCACCAGCAGAGGAATG
GATATCACGGACTGATTCTGACATCATAGATGCAACAATGAAAGAACTTG
AGAAACTCTTCCCTGATGAAATCTCAGCTGACCAAAGCAAAGCTAAAATT
CTGAAGTACCATGTCGTTAAGACTCCAAGATCTGTGTACAAGACCATCCC
AAACTGTGAACCATGTCGTCCTCTACAAAGATCACCTATTGAAGGATTCT
ACTTAGCTGGAGATTACACAAAACAGAACTTAGCTTCCATGGAAGGC
GCTGTCTCTCTGGCAAATTCTGCTCTCAGTCTATTGTTGAGGATTACGA
GCTACTGGCTGCGTCTGGACCAAGAAAGTTGTGCGAGGCAACAGTATCAT
CATCA-----

>P

-----GCACTAGCTTTTAGAGGCAAGTGAATCAATGGGC

CATTCTTTGAAATTCCTTTGGAATTTCTTCTGCTAAACAAGACTAAG
 GAATCATATCCGCCCTCCTTTGCGGGTTGCTGTGGACTATCCAAGAC
 CGGACCTTGATAACACGGTGAATTTTTAGAAAGCTGCCTTGTTATCTTCA
 TCCTTTTCGTTCTTCTCCGCGTCCAGCTAAACCATTAAATGTTGTCTATTGC
 TGGTGCAGGTTTGGCGGGTTTATCGACTGCAAAATATTGGCAGATGCAG
 GCCATAAGCCTATATTGCTTGAAGCAAGAGATGTTTTAGGTGGAAGGTG
 GCTGCATGGAAAGAC - - - GATGGAGACTGGTACGAGACAGGCTTGCATAT
 ATTCTTTGGGGCATATCCAATGTGCAGAATCTTTTTGGTGAAGTTGGTA
 TCAATGATAGGTTGCAATGGAAGGAGCATTCTATGATATTTGCAATGCCA
 AATAAGCCAGGAGAATTCAGTCGATTTGATTTTCTGAAGTTCTCCCTGC
 ACCATTAATGGGATATTAGCCATTTTAAAGAACAATGAAATGCTGACTT
 GGCCAGAGAAAGTGAAGTTTGCATTTGGGCTACTGCCAGCAATTGTTGGT
 GGACAAGCTTATGTTGAGGCTCAAGATGGTTTAAAGTGTCAAGAGTGGAT
 GAGAAAGCAGGGTGTACCTGATAGAGTGAAGTGTGAGGTGTTTATTGCCA
 TGTCAAAGGCTCTAAACTTTATTAACCCAGATGAGCTTTCAATGCAATGC
 ATTTTGATAGCTTTGAACAGATTTCTTCAGGAGAAACATGGTTCAAAGAT
 GGCTTTCTTGGATGGTAATCCCCAGAGAGGCTCTGCATGCCAATTGTTG
 ATCATATTCAGTTGCGTGGTGGTGAAGTCAAGCTTAATTCTCGGATAAAG
 AAAATTGAGCTAAATGATGATGGAACAGTGAAGAGTTTTTACTAAATAC
 CGGGGATGTGATTGAAGGGGATGTTTATGTGTTTCCCACTCCAGTTGATA
 CCCTGAAGCTTCTTTGCCTGATAACTGGAAAGAGATTCTTACTTCAAG
 AAAGTGGAGAAATTAGTTGGAGTTTCTGTTATTAATGTTTACATATGGTT
 TGACAGGAACTGAAGAATACATACGATCACCTACTTTTTCAGCAGGAGTC
 CTCCTTCTCAGTGTGATGCTGACATGTCTCTGACATGTAAGGAGTATTAT
 GACCCAAATAAATCTATGCTGGAGTTAGTTTTTGGCCCTGCTGAAGAATG
 GATTTTCACGCAGTGACTCAGAGATTATTAATGCTACAATGGGGAACTTG
 CAAAACCTTTTCTGATGAAATATCCGCAGATCAAAGCAAAGCAAAAATC
 GTGAAGTATCATGTTGTTAAACTTCAAGGTCGGTTTACAAGACTGTCCC
 AGATTGTGAACCTTGGCGTCCCTTGAAGATCTCCGATAGAGGGTTTCT
 ATTTAGCTGGTGACTACACAAACAAAGTACTTGGCTTCAATGGAAGGT
 GCTGTTCTATCAGGGAAGCTTTGTGCACAGGCAATTGTACAGGATTATGA
 GTTCCTGGTTGCTCGGGGGCAAGGAAGGTTGACTGAGGCAACCATTACT -

 >R
 GCTCTATATGGGGGTGTTTCTGCTTTGAATTTAAGCTGGCATAGTGATGT
 CTTAGACACTAGAAATCTGCAATCAGCCCTTAGATGTGGTTACGCTACCT
 GTTCTAATGAAACCAATGTAAGCTTTTAGAGGCAGTGAATCTATGGGC
 CATGCTTTGAGA - - - - - AATTCTTCTAAACAAGATTTAG
 GAATACTGGTAGCTGCCCTTTGAAGGTAGTTTGTGTGGACTATCCTAGAC
 CAGACCTTGATAACACAGTGAATTTCTTGAAGCTGCCTACTTATCATCA
 TCTTTTTCGATCTTCTTCCCTCCAGATAAACCATTGAAGGTTGTAATTGC
 TGGTGCAGGATTGGCTGGTTTATCAACTGCAAAATATTTGGCAGATGCAG
 GACACAAGCCTTTATTGCTGGAAGCAAGAGATGTTCTAGGTGAAAGGTG
 GCTGCATGGAAAGATGATGATGGGGACTGGTACGAGACAGGCTTGCATAT
 ATTCTTTGGAGCATACCCAAATGTGCAGAACCCTGTTTGGAGAAGTTGGTA
 TAAATGATAGATTGCAGTGAAGGAGCATTCTATGATATTTGCGATGCCA
 AACAAGCCTGGAGAATTGAGCCGATTTGATTTCCAGATGTTCTTCTGC
 ACCATTAATGGGATATGGGCAATTCTGAGAAACAATGAGATGCTGACAT
 GGCCAGAGAAAGTGAATTTGCAATTGGAAGTCTGCCAGCGATGGTTGGT
 GGACAGGCTATGTTGAGGCTCAAGATGGTTTAAAGTGTCAAGAGTGGAT
 GAGAAAGCAGGGCTACCTGATAGAGTGAAGGAGTTTTTATTGCTA
 TGTCAAAGGCGCTAAACTTTATTAACCTGATGAGCTTTCAATGCAATGT
 ATATTGATAGCATTGAACAGATTCTTCAGGAGAAACATGGTTCAAAGAT
 GGCTTTCTTAGATGGAAATCCCCAGAGAGACTCTGCATGCCAATTGTTG
 ACCATGTGCAGTCACTTGGTGGTGAAGTCCGGCTAAATTCACGAATAAAG
 AAAATTGAATTAATAATGATGGAGCAGTGAAGAAGCTTTTACTAAATAA
 TGGGGAAGTGATTGAAGGAGATGTTTATGTGGTTGCTACTCCAGTTGATA
 TCCTGAAGCTTCTTTGCTGATAACTGGAAAGAGATTCCATACTTCAAG
 AAGCTGGATAAATTAGTTGGAGTTCTGTTATTAATGTTTACATATGGTT
 TGACAGGAAGCTGAAGAATACATATGATCACCTACTTTTTCAGCAGAAGTC
 CCCTTCTTAGTGTTTATGCGGACATGTCTGTAACATGTAAGGAATATTAT
 AATCCAATCAGTCTATGCTGGAGTTAGTTTTTGCACCTGCAGAAGAATG
 GGTATCACGCAGCGATGAAGAAATTTGAGGCTACAATGATGGAAGTAG
 CAAAACCTTTCTGATGAAATATCTGCAGATCAGAGCAAAGCAAAAAT
 GTTAAATACCATGTTGTCAAAACTCCAGGCTGTTTACAAGACTGTCCC
 AAATTGTGAACCTTGGCGACCTTGCAGAGATCTCCTATAGAGGGCTTCT
 ATTTGGCTGGTGAACACAAAACAAAATATTTGGCTTCGATGGAGGGT
 GCTGTTCTATCCGGGAAGTATTGTGCACAAGCCATTGTACAGGATTATGG
 GTTGCTTATCGCTCGCAAGCAAAAAAGTTGGCTGAGACTTCTGGTATGC
 ACATTACTGAGCTTTACTGCAGAAGGCTGAGTTCTGTTGTCTATGGACTT
 TAT
 >0
 --- GATACTGGCTGCCTGTCTATGACATAACTGGAACCAAGCAAGC
 AAGATCTTTTGGGGACAACCTTCTACTCATAGGTGC-----

->A
 -----GCTTCTTCAGTCGCTCTTCGCGGCAACTGG
 GTCTCTCTCC-----
 -----GTTCTCTCCATTGAAGTCTCGGAGA-----TTCTAT
 GTTAATTCCTCACTCGACTCC-----GATGTTTCCGACATGAGTGT
 TAACGCTCCAAGGGGTTATTTCTCTCCAGAGCGTGTACCATACAAGGGGA
 CTAAGCTGAAAGTAGCCATTATAGGAGCTGGACATGACGCATGTCTGAC
 GCTGTAGAGCTTCTGGATCAGGGTACAGAGGTGGATATATGATTCCGAG
 GACATTCATTGGTGGCAAAGTGGGTCTTTTGTAGATAGACGTGGGAACC
 ATATA--ATGGGATTACATGTTTCTTCTCGGATGCTACAACAATCTCTTTC
 CGTTTGTAGAAAAAGTGGGTCTGAGAAAAACCTTCTAGTCAAAGACCA
 TACTACATGACCTTTATAACAAAGACGGCAACTTGGAGAGCTTGATTTTC
 GGTTTCTGTTGGAGCTCCAATCCATGGTATTCGCGCATTTTTGGTCACA
 AATCAACTCAAGCCCTATGACAAACTAAGGAATTCAGTGGCTCTTGCAAT
 AAGCCCTGTTGTTAAAGCTTCTTGTCGACCCTGAGGACCATGAGAGACA
 TTCGCAACCTTGACAGCATTAAGTTTCTCGGATTGGTTCTGTCAAAGGT
 GGGACACGGGCAAGTATCCAGAGAATGTGGGATCCTGTTGCTTATGCTCT
 AGGTTTTCATAGACTGCGATAACATGAGTGCAGCATGATGCTTACAATAT
 TCTCATTTACTCGCTACCAAAACAGAGAAGCTTCTGTGTCGATGCTCAA
 GGCTCTCCTGATGTTTACTTGTAGTGGTCTCATCAACAATATATCACAGA
 TAGAGGTGGCAGAATCCATCTAAGGTGGGGTTCGAGGGAGATACTTTATG
 ATAAATCAGCAGATGAGAGAACTTATGTCACTGGACTAGCAATTTCTAAG
 GCTACGACAAGAAGATTGTGAAGCTGATGTATGTTGCAGCATGTGA
 GTGCCGGGAATCAAAGGTTACTTCAAAGAAATGGAGGGAATCTCGGT
 TTTTCAACGATATATATGAAGTCTGAGGGAGTACCTGTGGTTACAGTACAG
 CTCAGATACAATGGTTGGGTCAACGAGTTCACAGACATGAACTTGCTAG
 CGAGTTGAAACGAGCGGTTGGCTTAGATAATCTTCTACACTCCTGATG
 CTGATTTCTCGTGCTTTGCGGATCTTGCACTAGCTTCGCTGCAAGATTAT
 TACATCGAAGGACAAGGCACTTTACTTCAAGTGCCTTCTAACACCGAGAGA
 TCCGATATAGTCGAATGCCAATGATAAGATCATAGAGAAGTAGCGGATGC
 AGGTACATGAGTTGTTTCCATCGTCAAGAGCCGTAGAAGTCACTTGGTCTG
 TCGGTTGTAAAAATCGTCAAGTCTGTACCGTGAAGACCCCGAAAGGA
 TCCATTACAGACCCGATCAGAAGACGCCCCATAAAGAAATTTCTTCTCGCCG
 GTTCATACACAAAGCAGGATTACATAGATAGTATGGAAGGAGCAACATTA
 TCAGGGAGACAAGCTTCTGCATACATCTGTGATGCTGGTGAAGAACTAGC
 AGAGCTAAACAAGAAAGCTTTCTTCT-----TCTG
 CAAGTGCAGTTTCTGATGAGCTAAGTCTGGTC-----

-----AGTGT
TAATGCTCCAAGGGGTTATTTCCACCAGAACCTGCTCATTATAGGGGAC
CAAACTGAAGGTGGCTATTATTGGAGCTGGGTTGGCAGGCATGTCAACT
GCAGTGGAGCTATTGGATCAAGGCCATGAGGTGGATATATACGATTCCGAG
GCTTTTCATTGGTGGTAAAGTGGGTTCTGTTTGTGATAGACGTGGAACCC
ATATTGAAATGGGACTTCACGTTTTCTTGGTTGCTATAATAATCTTTTC
CGTTTGATGAAGAAGGTGGGCGCAGATAAAAATCTTCTTGGAAGGATCA
TACTCACACATTTGTTAAACAAGGGAGGTGAAATTGGTGAGCTTGATTTTC
GGTTTCTAATTGGAGCTCCGTTGCATGGGATTCTGTCATTTTTGTGCACA

[illegible]

-----ATGGCTTCATCGATTCTTCTCCGGCCAAT--
-TCTGTTGCTGGAAGCTAGAAGGAGGACTCCAGGATCTTGTCTTCTGTGTG
ATCGCCCGCTCGTGCAGCGCTCAGGTGGTGAGGACTCAGAGGTTGTTCTTT
TGCTGCTCGCTTTGGATTCCCTGGAAACAAAAGTGTCGTATAGTGT
TAATGCTCCAAGGGGTTGTTTCCACCTGAACCTGAGCATTATAGAGGAC
CAAACTGAAGTGCGCTATTATTGGAGCTGGGCTGCAGGGGATGTCAACT
CGAGTGGAGCTATTGGATCAAGGCACTAGGTGGACATATGAATCAAG
GACCTTTATTGGTGTGAAGTGGGTTCAATTGTGGACAGACGTGGAATC
ACATTGAAATGGGACTTCATGCTTCTTTGGTGCTATAATAATCTTTTC
CGTTTGTATGAAGAAGGTGGGTGCAGATAAAAAATCTACTTGTGAAGGAGCA
TACTACACATTTGTAAACAAGGGGGTGATATTGGTGAGCTTGATTTTC
GGTTTCCAATAGGAGCTCCCTGGCATGGGATCCAAGCATTTTTGACCACC
AATCAACTTAAGACGTATGATAAAGCAAGAAATGCTGTGGCTCTTGCCCT
AAGTCCAGTTGTGAAGGCCCTCATTGATCCAGATGGAGCAATGACGGACA
TAAAGAAATTTAGATAGTACAGCTTCTCTGATTGGTTTTTGTCTAAAGGT
GGCACACGCACAAGTATCCAAGAAATGTGGGATCCTGTTGCATATGCCCT
TGGGTTCAATTGACTGTGATAACATCAGGCTCGATGTATGCTCACTATAT
TCTCCTTTGTTGGCCATAAAACAGAGGTTCTGCTACTACGCTGCTTTAAG
GGTCTCCGGATGTTTATTTAAGCGGTCCCAATTAGAAGTATATTGAAGA
TAAAGGAGGCAGGTTTTCAATTTAAGGTGGGGATGCAGACAGATACTCTATG
ATACGTCTGCTGATGGAGAAACATATGTTACAGGGCTTGCCATGTCAAAA
TCAACTACCAAGAAAGTTGTGAAGGCTGATGCTTATGTTGACGATGTGA
TGTCCTTGGAAATCAAAAGATTACTTCCATCTCAGTGGAGGGAATCAAGT
TTTTTGATAATATTTATGAGCTAGTTGGAGTACCTGTTGTAAACGGTTCAA
CTTAGATATGATGGCTGGGTACGGAGTTACAAGATCTAGAGCCATCAAG
GCAATCAAGGCAAGCTGTGGGTTTAGATAACCTTCTAGACCTCCAGATG
CGATTATTTCTGTTTGGTGGGACTTAGCACTCACTTCTCCGGAAGATTAC

TACATTGAAGGACAAGGATCATTGCTCCAATGTGTTCTTACACCCGGTGA
TCCTTACATGCCCTTGGTAAATGAAGAAATCATAAAGAGAGTATCAAAGC
AGGTTTTGACTTTATTTCCGTATCCCAAGGCTTAGAGGTTATCTGGTCA
TCCGTAGTAAAAATGGGCAATCTTTATATCGCGAAGGACCTGGCAAAGA
TCCTTCAGACCAGATCAAAAGACACCTGTGAAAAATTTCTTCTTGTCTG
GTTCTTATACAAAACAGGATTACATAGACAGCATGGAGGGGGCAACTTTA
TCCGGAAGACAGGCTTCAGCATATATATGTGAAGCTGGTGAAGAACTAGT
TGCAATTGAAGAAGAAGCTTGCTGCCGTTGAATCTCCAGAATGCCACAGT
CTGCTGCAGTCGCTGATGAGCTGAGTCTTCTAGAAAGCATTATGGGTTCT
GTCATAAAGGATGGTTCCTTCTGTTTTCTGAGTTTCTGCTGTGTTTTCAT
CCTTGTTTCATCCTTTTTCATGTGATGCACAAGAACTGAAAGCATGTGGAT
TGGATGCAATATATCAATTTGGAGATTCAATTTAGATACAGGCAATGCT
GTTCTAGAAATTTTCTATCCTTCCCTTATGACCAATTTCCCTACGGCATAAC
AGTTAAACATGCAACTGGTAGACCTTCAGATGGACTACTTATGGTTGATT
ACATTGCACAAGCAGCTGGTCTTCCATTTGTTGAACCATATGAAACCCA
AAATCCAATTTTTCCCATGGAGTAGATTTTGAGTTGCTGGTGTAACGT
TGTGACAGCAAGAACTCTAGTTAAATGGCATATCCACCTTTTGTCACTA
ACCATTTCTTAACATTGCAACTTGGGTGGTTTGAGAAGCACTTGTCAACT
ATCTGCAGTGATCCTAAAGCCTGCCAAGAGAACTGAAGTCAGCCCTCTT
TATGGTTGGGACAATGGGATCTAATGATTATTTCTTGCATTGAGTCGAA
ATAAAACCTTGAGGAGATCAAGAACACCATGGTGCTGTAGTGGTGCAG
ACCATAACGGGAAGTTGCCAAAAAATTTATCGGCCATGGTGTGTTTCGAGT
GGTTGTCCTGGACTTCATCAATTGGGTGTTTACCAGGCATACTTACAG
CATTTGAGACCAATACATCTGTTTCATGATGCACAAGGCTGCCTGAAAGAT
TTCAACGACATGTTTGTGATCACAATGATCATCTGAAAACAGCTCTTGA
AGGATTGAGAAAAAGAAATTTCTAACGTGCATGTCGTATATGCTGATAACT
ACAGTGCAGTGCATATATCATAGACAATTTGTCAAAGCTTGGATTTAAA
GCATTACGAGAAGCCTGTTGTGGAACCGGAGGAAAAATACAATTACAGTGT
TGATCAATTAAGTTTGCCTGTGGACTTCTGGCATAACCATATTGTTCAA
ACCCTAGAGAACATGTCTTTTGGGATGGAGGACATTTTCTCATCAACA
AACAAGTTCTTATCAGATTGGCTACTTAGAGACATGTTACCAAAGCTTCA
ATGCGCAGCT

>0
GCCATGGCTGCCACGTCCCGCGCGCCGTCCACGCTGGCCCCGGCGTCTGT
CTCCGCCGCCGGC-----GGCA
GCAGGAGCGCGCGCCGGTGCCGAATCCCAAGGTCGCGGTGGGGGTGGGG
GTGCGGTGCTCGCTCGACAGC-----AACGTCTCCGACATGGCCGT
CAACGCTCCGAAAGGATTGTTCCCGCCGAGCCGAGCACTACAGGGGCG
CGAAGCTGAAGGTGGCCATCATAGGGGCCGGCTCGCCGGCATGTCCACC
GCGGTGGAGCTCTTGAGACCAGGGCCATGAGGTTGATCTGTACGAGTCCCG
GCCGTTTATCGGCGGCAAGTCGGTCTTTTGTGGATAGGAAAGGGAACC
ATATCGAGATGGGGCTGCACGTGTTCTTCGGGTGTTACAGCAATCTTTTC
CGCCTTATGAAGAAGGTTGGAGCTGATAATAATCTGCTAGTGAAGGAGCA
TACCCATACATTTGTAATAAAGGAGGACGATTGGTGAACCTTGATTTTC
GGTTTCTGTGGGAGCTCCGTTACATGGTATCCAAGCATTCTACGAAC
AACCAACTCAAGGTTTATGATAAAGCAAGAAATGCCGTTGCTCTTGCTCT
AAGCCCAGTTGTTTCGAGCTCTTGTTGATCCAGATGGTGCATTGCAGCAAG
TACGGGATTTGGATGATGTAAGTTTCAGCGATTGGTTCTTGTGCAAGGTT
GGTACTCGAGAGAGCATCAAGGATGTGGGATCCTGTTGCCTATGCTCT
TGGTTTCATTGACTGTGATAATATCAGTGCACGTTGCATGCTTACCATT
TCACTCTGTTTGGCACAAAAACAGAGGCATCTTATTACGCATGCTAAAG
GGTTCACCTGATGTTTATCTGAGTGGTCCAATAAAGAAGTACATAACAGA
CAGGGGTGGTAGGTTTACCTGAGGTGGGGATGTAGGGAGGTTCTCTATG
ATAAGTCACTGATGGGAAACCTATGTTAAAGGCCTTCTCCTATCCAAG
GCTACAAGTAGAGAGATAATCAAAGCAGATGCATATGTCGCAGCTTGTGA
TGTCGCGGGGATCAAAAGACTTTTACCTTCTGAATGGAGGCAATGGGATA
CATTTGACAACATCTACAAGTTAGATGGTGTTCTGTAGTCACAGTACAG
CTTCGTTATAATGGATGGGTTACAGAACTTCAAGATTTGGAGAAATCAAG
ACAACCTGAAAAAGGCAGTTGGCTTGGATAATCTTCTTACACTCCAGATG
CAGATTTTTCATGTTTTTCAGACCTTGCACTTTCATCTCCTGCTGACTAC
TACATTGAAGGACAAGGTTCTTGATCCAAGCTGTGCTAACCCCTGGCGA
TCCTTACATGCCATTGCCGAATGAGGAGATAATTAGCAAGGTTCAAAAGC
AGGTCTTAGAATTGTTCCCGTCATCACAAGGCTTGGAACTTACATGGTTCG
AGTGTGGTGAAAAATCGGTCAATCATTGTACCGCAGTCACCAGGAAATGA
TCCATTTAGACCTGATCAAAAGACACCAAGTTAAAAACTTCTTCTGTCTG
GCTCTTACACAAAACAGGACTACATTGACAGCATGGAAGGGGCAACTCTC
TCAGGCAGAGAAACCGCGCCTACATCTGTGGTGCAGGAGAGGAGCTGCT
CGCCCTCCGAAAGAAGCTCATTGTGACGACAGCGAG-----AAGG
CCAGGGGTAAAGTTCGAGGCCCTTCAGACAAGC-----


```

-----
-----

GA.1.14.11.12.fasta
>A
---GCAACGGAATGCATTGCAACGGTCCCTCAAATATTCAGT-----
-----GAAAACAAAACCAAAGAGGATTCTTCGATCT
TCGATGCAAAGCTCCTTAATCAGCACTCGCACCATACCTCAACAGTTC
GTATGGCCCGACCACGAGAAACCTTCT--ACGGATGTTCAACCTCTCCA
AGTCCCACCTCATAGACCTAGCCGGTTTCTCTCCGGCGACTCGTGCTTGG
CATCGGAGGCTACTAGACTCGTCTCAAAGGCTGCAACGAAACATGGCTTC
TTCTAATCACTAACCATGGTGTGATGAGAGCCTCTGTCTCGTGCTTA
TCTGCATATGGACTCTTTCTTTAAGGCCCGGCTTGTGAGAAGCAGAAGG
CTCAGAGGAAGTGGGGTGAGAGCTCCGGTTACGCTAGTAGTTTCGTCCGG
AGATTCTCCTCAAAGCTCCCGTGGAAGGAGACTCTGTCGTTTAAAGTTCTC
TCCCGAG--GAGAAGATCCATTCCCAAACCGTTAAAGACTTTGTTCTA
AGAAAATGGGCGATGGATACGAA--GATTTCCGGGAAGGTTTATCAAGAA
TACGCGGAGGCCATGAACACTCTCTCACTAAAGATCATGGAGCTTCTTGG
AATGAGTCTTGGGGT-----GAGAGGAGATAT--TTTAAAGAGTTT
TCGAAGACAGCGATTCAATATTCCGGTTGAATTACTACCCGCAAGTCAAG
CAACCGGAGCTTGCACTAGGGACAGGACCCCACTGCGACCCAACATCTCT
AACCATACCTTCATCAAGACCAAGTTGGCGGTCTGCAAGTTTTCTGGGACA
ACAAATGGCAATCCATTCTCTAACCCTCACGCTTTCTGGTGGAACATA
GGCGACACCTTCATGGCTCTAACGAGA--AGATACAAGAGTTGTTTGCA
TCGGGCGGTGGTGAACAGCGAGAGAGAAAGGAAGACGTTTGCATTCTTCC
TATGTCCGAAAGGGGAAAAAGTGGTGAAAGCCACCAGAAGAACTAGTAAAC
GGAGTGAAGTCTGGTGAAAGAAAGTATCCTGATTTTACGTGGTCTATGTT
TCTTGAGTTTACACAGAAAGCATTATAGGGCAGACATGAACACTCTTGACG
AGTTCTCAATTTGGCTTAAGAACAGAAGAAGTTTC-----
---
>P
ATGGCAATAGATTGCATCAAAACCATGCCATCCATAACCACCCCTCATCA
CCACCCAAAAGATCAGGATCAATGCAAAGATGATGGCAAGTCTTTTGTTC
TTGATGCACAAGTTCTT--CGACACCAGACAAACATACCCAGCAGTTC
ATTTGGCCTGAC--CAAAAGCCTAAT--ATTAATGCACCTGAACCTCA
AGTCCCACCTTGTAGACTTGGGTGATTTCTCTCTGTAACCTGTTGCTG
CAGTGGAAGCTTCAAGACTTGTGGTGAAGCATGTCAAAACATGGTTTC
TTTCTAGTTGTTAATCATGGAGTTGATAAAACGCTCATTGCCATGCTCA
TAATTACATGGACACCTTCTTCGAATTGCCACTTTCTGAGAAACAAAAGG
CCCAAAGAAAGATTGGCGAGTCTTGTGGATATGCTAGCAGCTTTACTGGC
AGGTTTTCTCTAAACTTCCATGGAAAGAAACGCTTTCTTTTCGCTACAC
AGCT-----GAGAAGAATTCAACAAACACATCGAGGAATACTTTTCATA
ACAGAATGGGGGAAGATTTGCT--GAATTCGGGAGTGTGTATCAGGAC
TACTGTGAGGCCATGAGCACTTTGTCACTAGGATCATGGAGCTATTAGG
CATGAGCCTTGGTGTG-----AGCAGAGAACAT--TTCAGGGAATTCT
TTGATGAGAATGATTCAATAATGAGGCTCAACTACTACCCTCCATGCCAA
AAACCTGACCTTACTTTAGGCACCGGTCTCTATTGTGATCCAACCTCTTT
GACCATCCTTTCATCAAGACCAAGTGGGTGGTCTTCAAGTGTGTTGGACA
ACGAATGGCGTTTCGATTAGCCCAATTTTGACGCTTTTGTGTTAACATT
GGTGACACCTTCATGGCTCTATCGAATGGTATATACAAGAGCTGTTTGCA
CCGAGCAGTGGTGAACAGCCAAACACCAAGAAATCCCTCGCTTTCTTTT
TGTGTCCAAAGAAATGACAAGATGGTAACCTCCACCACATGAATTAGTGAC
-----ACATGCAATCCGAGAATATATCCAGATTTACATGGCCTATGTT
GCTCGAATTACACAAAAGCATTACCGAGCTGACATGAAGACACTTGAGG
TGTTACAAACTGGCTTCATCAACAAAGTTTTAGC-----
---
>R
---GCAATAGAGTGCATCAAAGCGTGCCATCGTCAGTCCAT-----
-----CAAAAGGAAGAACGTAGCAAAGAAGAGCAGAAATCATTAGTTT
TTGATGCCAAATTTTTA--AGATACCAATCAAACATTCCTCAACATTT
GTATGGCCGGATGATGAAAAGCCAATC--GTCAATGCACCAGAAGTGGC
TGTTCCGCTAATCGACTTAGGTGGGTCTTGTCTGGTGACCCTGTTTCTG
CAACGGAAAGCTTCAAAGCTAGTTGGCGAGGCATGTCAAAAGCACGGTTTC
TTCTTGTGTTTAAACACGGTGTGATCAAAAGCTCATAGCTGATGCGCA
CCGTTACATGGACAACCTTCTTTGAATTGCCACTCTCCGATAAGCAGAAAG
CTCAGAGGAAAGTTGGTGAGCACTGTGGTTATGCTAGTAGCTTTACTGGC
AGGTTCTCCACTAAACTTCTTGGAAAGAAACACTTTCTTTTCGCTATTTC
TGCC-----CAGCAGAACTACCAAAACTTGTGCAAGACTATTTTCATA
ACACAATGGGTGAATCTTTTTTG--GAATTCGGGAAGGTCTACCAAGCAG
TACTGCGAGGCTATGAGCACTCTTTCATTAGGAATCATGGAACTTTGGG
AATGAGCTTGGGAGTA-----AGCAAAGCCAC--TTCAGGGAATTTT
TTGAAGAAAACGATTCAATAATGAGGCTGAACATTATCCACCGTGTCAA
AAACCTGACCTTACTTTAGGAAGTGGACCTCATTCGATCCACATCATT

```

AACCATCCTTCACCAAGATCAAGTTGGTGGTCTTCAAGTGTATGTAGACA
 ATGAATGGCGTTCAATCAGCCCCAATTTTCAAGCCTTCGTGCTCAACATT
 GGAGATACTTTTCATGGCACTTTTCAAATGGCAAATATAAGAGTTGCTTGCA
 TAGAGCAGTAGTGAACAGCGAGACAACAAGAAATCATTGGCTTTCTTTC
 TATGTCCAAAGAGTGACAAGATGATAACACCACCAACGAAGCTAGTGGAC
 -----AATTGGAACCAAGAATATATCCAGATTTTACATGGTCAATGCT
 TCTTGAATTTACACAGAAGCATTACAGAGCAGACATGAAGACTCTTGAGA
 TGTTTTCAAATGGGTTCAACAGAGAAAG-----

 >O
 ---AGCATGGTGGTG-----
 -----CAGCAGGAGCAGGAGGTGGTGT
 TCGACGCGGCGGTGCTG---AGCGGGCAGACGGAGATCCCGTCGCACTTC
 ATATGGCCGGCGGAGGAGAGCCCGGGTCGGTGGCGGTGGAGGAGCTGGA
 GGTGGCGCTGATCGACGTG-----GGGGCGGGGCGGAGA
 GGTGCTGCGGTGGTCCGGCAGGTGGGGGAGGCGTGCAGAGAGGCACGGCTTC
 TTCTTGGTGGTTAACCACGGCATCGAGGCGGCGCTGCTGGAGGAGGCGCA
 CCGGTGCAATGGACGCCTTCTTACGCTGCCGCTGGGGGAGAAGCAGCGGG
 CGCAGCGGCGCGGGGGAGAGCTGCGGCTACGCCAGCAGCTTCACGGGG
 CGCTTCGCGTCCAAGCTGCCGTGGAAGGAGACGCTGTCGTTCCGGTACTC
 ATCGGCTGGAGATGAAGAGGCGGAGGAGGCGTGGGTGAGTACCTGGTGC
 GGAAGCTCGGGGCGGAGCACGGGCGGCGCTGGGCGAGGTGTACTCGCGC
 TACTGCCACGAGATGAGCCGCTGTCGCTGGAGCTGATGGAGGTGCTCGG
 GGAGAGCCTGGGCATCGTCGGAGACGGCGCCACTACTTCCGGCGATTCT
 TCCAGCGCAACGACTCCATCATGCGCCTCAACTACTACCGGGCGTGCCAG
 AGGCCACTCGACACGCTGGGACCGGTCGCACTGCGACCCCACTCGCT
 CACCATCTCCACCAGGACCACGTCGGCGGCTGGAGGTGTGGGCGGAGG
 GGCGGTGGCGCGCCATCCGCCCTCGCCCCGGGGCGCTCGTCGTCAACGTC
 GGCAGACCTTCATGGCGCTCTCCAACGCCAGGTACCGCAGCTGCCTGCA
 CCGGGCGGTCTGCAACAGCACGGCGCCTCGCCGCTCGCTGGCTTCTTCC
 TCTGCCCGAGATGGACACGGTGGTGCGCCCGCCGGAGGAGCTGGTCGAC
 -----GACCACCACCCGAGGGGTACCCGGACTTCACGTGGCGGGCGCT
 GCTGGACTTCACGCAGCGCCACTACAGGGCGACATGCGCACGCTTCAGG
 CCTTCTCCGACTGGCTTAATCATCATCGTCACCTGCAACCAACAATATAC
 TCC
 >V
 ---TCTATAGTTTGTGTCGAAGGTAACCCACCATCCATGTTT-----
 -----AATCCGCTACAGACGACCACAAAAACAGGAACCACTAGTGT
 TTGATGCCTCGGTTCTT---CGGCACCAATCCAACATTCCCAAGCAGTTC
 ATATGGCCGGATGCAGAAAAGCCAGGG---GACAAGGCCACAGAGCTGTC
 TGTTCCACTCATAGACTTGGGAGGCTTCTCTCTGGAGACCTGCTGCAG
 CCATGGAAGCAACTAGGCTTGTGAGGGAGGCATGCCAGAAAGCATGGCTTC
 TTCTTTGTTGTTAACCATGGGGTTGATGATAAACTCATATAAAAGCTCA
 CCAGTACATGGACAGCTTCTTTGGGTTGCCTTTGGCCAAGAAGCAAAGGG
 CTCAGAGGAAGCTGGGTGAGCACTGTGGATATGCCAGCAGCTTCATTGGC
 AGGTTCTCTTTCCCAAGCTGCCATGGAAAGAGACCCTTTCTTTACGTATT
 AGCT-----GAGAAGAAGTCATCCAATGCTGTGCGAGGAGTATTTCTCA
 AAAAAATGGGTGAAGATTTCAGT---GAATTCGGGCAAGGTACCAGGAC
 TACTGTGAGGCCATGAGCACGCTTTCCCTTGTGATCATGGAGCTGCTGGG
 GATGAGCCTGGGTATA-----GGCGGAGCCCAT---TTCAGGGAATTTT
 TCGAGGAAAACGATTTCGATAATGAGACTCAACTACTACCCCCATGCCTA
 AAACCCGACCTCACTTTAGGAACAGGGCCTCATTGTGATCCAACATCTTT
 AACCATCCTTCATCAGGACCAAGTGGGTGGCCTTCAAGTGTTTGTAGATG
 ACAAATGGTGGTCCATTAGCCCAAATTTGATGCCTTTGTTGTTAATC
 GGCAGACCTTCATGGCTCTTTCAAATGGAAGGTACAAGAGCTGCTTACA
 TAGGGCAGTGGTAAACAGCCAACTCCAAGAAATCTCTTGCTTTCTTTT
 TGTGTCCAGAGAAAGATAAGGTGGTGAGGCCACCAACTGAACTGGTGGAC
 -----ACCAACAGCCCTAGAATATATCCGGATTTCACTTGGTCCAATCT
 CCTCGAATTCACCCAGAAGCATTACAGGGCTGACATGAAAACACTTGAGG
 TTTTCTCTAGCTGGCTTCAACAGAAAACAGCTGAAGCGGTG-----

GA.1.14.11.13.fasta

>A

GTAATTGTGTTACAGCCAGCCAGTTTTGATAGCAACCTCTATGTTAATCC
 AAAATGCAAACCGCGTCCGGTTTTA-----ATCCCTGTTATAGACT
 TAACCGACTCAGATGCCAAAACCAAATCGTCAAGGCATGTGAAGAGTTT
 GGGTTCTTCAAAGTCATCAACCATGGGGTCCGACCCGATCTTTGACTCA
 GTTGGAGCAAGAAGCCATCAACTTCTTTGCTTTGCATCACTCTCTCAAAG
 ACAAAGCGGGTCCACCTGACCCGTTTGGTTACGGTACTAAAAGGATTGGA
 CCCAATGGTGACCTTGGCTGGCTG---TACATTCTCCTTAATGCTAATCT
 TTGCCTTGAGTCTACAAAACACCGCCATTTTCCGGCACACCCCTGCAA
 TTTTCAGAGAGGCAGTGAAGAGTACATTAAAGAGATGAAGAGAATGTG

AGCAAATTTCTGGAAATGGTAGAGGAAGAGCTAAAGATAGAGCCAAAGGA
 GAAGCTGAGCCGTTTGGTGAAAGTGAAGAAAGTGATTCTGTCCTGAGAA
 TGAACCATTACCCGGAGAGGAAGAG-----ACTCCGGTCAAGGAA
 GAG-----ATTGGGTTCTGGTGAGCACACTGATCCACAGTTGATCACT
 GCTCAGATCAAACGACACAGAGGGTTTGCAAATCTGTGTCAAAGATGGAA
 CATGGGTTGATGTTACACCTGATCACTCCTCTTTCTTCGTTCTTGTGCGGA
 GATACTCTTCAGGTGATGACAAACGGAAGATTCAAGAGTGTGAAACATAG
 AGTGGTGACAAATACAAGAGGTCAAGGATATCGATGATCTACTTTCGCAG
 GTCTCCTTTGAGCGAGAAGATTGCACCATTATCATGCCTTGTGCCAAG
 CAAGATGATTGCCCTTTATAATGAGTTTACTTGGTCTCAATACAAGTTATC
 TGCTTACAAAATAAGCTTGGTGACTATAGGCTTGGTCTCTTTGAGAAAC
 GACCTCCATTTCTCTATCCAATGTT
 >P
 ---GTAGTACTTTCCCACTAGCTCTAGAA---CCTTCTCAGTTATCAA
 AACATGCAAGCCTATA---GGCTTGTTTTCCGAGATACCCGTTATAGACT
 TGACAGACCTCATGCCAAGACTCTCATAATCAAGGCCTGTGAAGAGTTT
 GGATTCTTCAAGCTGGTAAATCATGGAGTTCCAATGGAAGTCATGACCAA
 ATTGGAAGCTCTTGCTACCAACTTCTTTAACCTCCCCAGCCTGAGAAAG
 ACAAGGCTGGACCTCCTAACCTTTTGGCTACGGAAACAAGAAAATTGGC
 CCTAATGGTGATGTTGGTTGGGTTGAATATCTCCTCCTTAACACCAACCC
 TCAAATCTCCTCCCAAAA---ACGTCCATTTTTCAGGAAAACCCACAGA
 TTTTCCGCTCTGCTGTTGAAGATTACATACTGGCAGTGAAGAGAATGGCT
 TTTGAAGTATTGGAATTGATGGCTGATGGCTGGAGATTGAGTCAAGG---
 -ATGTTCAAGTAGGCTGTTGAGGGATGATAAGAGTGATTGATGTTTCAGGC
 TAAACCACTACCCACCATGCTCAGAG-----CTGCAAGCATTGAGTGGT
 GGAATTTGCTGATGGGTTTGGCGAGCACACAGACCCACAGATAATATCTGT
 TCTAAGATCTAACACACAAGTGGCCTGCAAATTTGTCTGAAAGAGGGAA
 CTTGGGTTTCAGTCCCCCTGATCAGACCTCTTTTTTCATCAATGTTGGT
 GATGCCTTGCAAGTGATGACAAATGGAAGGTTG---AGTGCAAGCACAG
 GGTTTTGGCTGATCCCCGAAACCAAGGATTTCCATGATTTTCTTCGGTG
 GTCCACCTTTAAGTGAAAAGATAGCACCTTTACCTCCTTGATGGCAGAA
 AGAGGAGGAAGCTTATACAAGGAGTTCACATGGTTTGAATACAAAAGGTC
 CGCATACAAGTCAAGGCTAGCTGATTACAGACTTGGGCTATTGAAAACCTG
 CAGGCCAA-----
 >R
 ---GTGGTTCTTTCACAAACAGCATTAGAC---CATTTCTCTTAATCAA
 AACATGCAAACCATCTTCAGGTTTGTTAATGGGATTCCAGTGATAGACT
 TGACAGACCTGAAGCCAAAAACCTCTTAGTTCAAGCTTGTCAAGAAATTT
 GGCTTCTTTAAATAGTCAATCATGGTGTCCCTATTGATTTTCATGACTCG
 ATTAGAGTCTCTTGCTATCAACTTTTTAACCTTCTCAATCTGAGAAAG
 ATAAAGCTGGTCTCTGACCTTTTGGTTATGGCAATAAAGAAATTGGC
 CCTAATGGTGATATCGGTTGGATTGAGTATCTTCTCCTCAATACCAATCC
 TCAAATTACTTCTCAAACCTACCTCTCCATTTTCAAAGAAAACCCACAAA
 TTTTCAGGCTCTGCTGTGGAAGATTACATATTGGCAGTGAAGAAAATGGCT
 TATGAAGTACTGGAAGTATGGCTGATGGATTGGGGATAAAGCCAAGGAA
 TGTATTGAGTAGGATGTTAAAGGATGATAAGAGTGATTGATGTTTCAGGC
 TAAATTACTATCCACCATGTCCAGAGCTTCAACTGCAACCATTAAGCGGT
 AGAAATCTGGTTGGGTTTGGGGAGCATACAGACCCACAGATAATTTCTGT
 CTTAAGATCTAACACACAACCTGGCCTGCAAATCTGTCTGAGAGATGGGA
 CTTGGGTTTTCAGTCCACCTGATCAGACCTCCTTTTTTCATCAATGTTGGT
 GATGCCTTACAGGTAATGACAAACGGAAGGTTTAAAGTGTAAGCATAG
 AGTTTTGGCTGACACAATGAGATCAAGGATTTCAATGATATATTTGGCG
 GTCCACCTTTGAGTGAAAAGATTGCACCTTTACCTTCTGTGTTGAGAGAA
 GGCGAAGAAAGCTTATATGAACAGTTCACATGGTGTGAATACAAGAGGTC
 AGCTTACAAGTCCAGG-----
 >O
 GTGGTTCTCGCTGGCCCGCCGCGCTCGAT---CACATCCGCTGCTGAG
 GTCGCGGACCCCGGC---GACGTCTTCTCCGCGTGCCTGCTGACG
 TCGGACGCCCCGGCGCGGCGAGGGCCGTGGTGACGCTGCGAGCGGTAC
 GGGTTCTTCAAGGTCGTCAACCACGGCGTGGCCACGGACACGATGGACAA
 GGCCGAGTCGGAGGCGGCTCAGGTTCTTCTCCAGACGCGCCGACAAGG
 ACCGCTCCGGCCCGGCTACCCGTTCCGGGTACGGCAGCAAGCGGATCGGG
 TTCAATGGCGACATGGGGTGGCTCGAGTACCTCCTCCTCGCCCTCGACGA
 CGCGTCGCTCGCCGACGCTGCACCGTC-----CCGTCTGCGCGG
 TCTTCCGGGCGCTCTGAACGAGTACATCTCGGGGTTGCGGAAGGTGGCG
 GTGCGGGTGATGGAGGCGATGTCGGAGGGGCTGGGCATTGCGCAGGCGGA
 CGCGCTGAGCGCGCTGGTGACGGCGGAAGGGAGCGACAGGTGTTCCGCG
 TGAACCACTACCCGCGCTGCCGCGC-----CTGCAGGGGCTC---GGC
 TGCACGCTACCCGGCTTCGGCGAGCACACCGACCCGAGCTCGTCTCCGT
 GCTCCGCTCAAACGGCACGTCGGGCTGCAAGTCCGCTCCGCGACGGCC
 AGTGGGTGTCGCTGCCCTCCGACCGGACTCCTTCTTCGTCAACGTGCGG
 GACTCGTTGAGGTTCTGACCAATGGGAGGTTCAAGAGCGTGAAGCACAG
 GGTGGTGGCCAAACAGCCTAAAGTCTAGGGTTTCCTTCATCTACTTTGGAG
 GGCCACCGTTAGCACAGAGGATTGCACCATTGCCACAGCTGCTGGGGAG

GGAGAGCAGAGCCTGTACAAGGAGTTCACATGGGATGAGTACAAGAAGGC
 TGCCACAAATCAAGGCTTGGAGACAACAGGCTGGCCAGTTTGAGAAGA
 AG-----
 >V
 ---GTTGTGTTATCGCAGAATGGTCGTGAG---CAGTTGCCTCTGATCAA
 GGCATGCAAGCCCAATTATGGC---TTCAGTGAATTCCTGTGATTGACA
 TGACAAGCCCTCAGCCAAGGCTCAC---GTTGAGGCCGTGTGAGGAGTTT
 GGATTCTTCAAGCTGGTCAATCATGGAGTTCCAATGGAGTTCATGTCCAG
 ATTTGGAGGATGAGGGTATCAAGTTCTTTAGCCTGCCTCAGTCTGAGAAGG
 ACCTGGCTGGTCTCTCAGCCTTTTGGGTATGGCAACAGGAGAATTGGC
 CCCAACGGTGATGTAGGGTGGATTGAGTATCTCCTCCTCCACCAATCC
 GGATTCATTTCCAGATATGCTCCTCCATTTCCCGGAAAAACCTGAAA
 TTTTCGGGTCTGAGTGGTTGATTATGTGTCATCCGTGAAGAATCATGACA
 TATGAAGTGCTGGAAATGATAGCTGAAGGGCTGAACATAGAACCAAGGAA
 TGTGCTGAGCAGGCTTCTCAAGGACGAGGAGAGCGACTCCTGCTTCAGGG
 TCAACCACTATCCACCATGCCAGAG-----CTTGAAGCATTGCGCGGC
 CGAGTTTTGGTTGGATTCCGAGAGCACACAGACCCACAGATAATATCTGT
 GCTCAGGTCCAACAACACCAATGGCCTGGAAATATGTCTCAAGGATGGGA
 CTTGGGTCTCAGTCCACCTGACCAGGACTCCTTTTTCATCAATGTTGGT
 GACTCTCTTCAGGTGATGACCAACGGGAAGTTTAAAGAGCGTGAAGCATAG
 AGTTGTGACAGAGAGCAGGAAGGCCAGAGTGTCAATGATCTACTTTGGAG
 GGCCACCTTTAACTGAAATCATAGCACCTTTACCCTCACTGGTGAAAGAA
 GGAGAAGAACTCTGTACAAGGAGTTCACATGGGCTGACTACAAAAAATC
 TGTTTATAAATCAAGGCTGGCTGATAACAGGCTTGGGCTCTTTGAGAAAA
 AAGAGCAGCCA-----

GA.1.14.11.15.fasta

>A
 ---CCTGCTATG---TTAACAGATGTGTTTAGAGGCCATCCCATTACCT
 CCCACACTCTCACATACCTGACTTCACATCTCTCCGGGAGCTCCCGGATT
 CTTACAAGTGGACCCCTAAAGACGATCTCCTC-----TTCTCCGCTGCT
 CCTTCTCCTCCGGCCACCGGTGAAAAACATCCCTCTCATCGACCTCGACCA
 CCCGGACGCGACTAACCAAATCGGTGTCATGCATGTAGAATTGGGGTGCCT
 TCCAAATCTCAACCACGGCGTGCCTTTGGGACTTCTCCAAGACATTGAG
 TTTCTCACCGGTAGTCTCTTCGGGCTACCTGTCCAACGCAAGCTTAAGTC
 TCGTCGGTCGGAGACAGGTGTGTCCGGCTACGGCGTCTGCTCGTATCGCAT
 CTTTCTTCAATAAGCAAATGTGGTCCGAAGGTTTACCATCACTGGCTCG
 CCTCTC---AACGATTTCCGTAACCTTTGGCCC-----CAACATCA
 CCTCAACTACTGCGATATCGTTGAAGAGTACGAGGAACATATGAAAAAGT
 TGGCATCGAAATTGATGTGGTTAGCACTAAATTCACCTTGGGGTCAGCGAA
 GAAGACATTGAATGGGCCAGTCTCAGTTCAGATTTAAAC---TGGGCCCA
 AGCTGCTCTCCAGCTAAATCACTACCCGGTTTGTCTGAACCGGACCGAG
 CCATGGGTCTAGCAGCTCATACCGACTCCACCCTCCTAACCATTTCTGTAC
 CAGAACAATAACCGCCGGTCTACAAGTATTTGCGGATGATCTTGGT---TG
 GGTCAACCGTGCCACCGTTTCTGGCTCGCTCGTGGTTAACGTTGGTGACC
 TCTTCCACATCCTATCCAATGGATTGTTTAAAGCGTGTGACCGCGCT
 CGGGTTAACCAAACCAGAGCCCGGTTATCTGTAGCATTCTTTGGGGTCC
 GCAATCTGATATCAAGATATCACCTGTACCGAAGCTGGTTAGTCCCGTTG
 AATCGCCTCTATACCAATCGGTGACATGGAAAGAGTATCTTCGAACAAA
 GCAACTCACTTCAACAAAGCTCTTTCAATGATTAGA-----
 -----AATCACAGAGAAGAA-----

>P
 ---CCTTCAAGATCATTGACAGACGCCTTTAGATCCCACCCGTCATCT
 ACACCAAAAGCACCTT---GACTTTTCTCTCTTCAAGAAATACCTGACT
 CACACAAATGGACTCAACTTGATGACATTGAA-----CAACAACATCCT
 TCAGTTGAATCCTTCATCACGGAATCTGTCCCGTTATCGATCTCTTAGA
 CCCTAACGTCCTTCAAAACATAGGCAATGCATGCAAACTTGGGGTGTGC
 TTCAAGTCACAAACCATGGCATCCCTATTAGCCTTCTTGAGAGCGTTGAG
 GGTGTTAGTAGGAGCCTTTTCTTTTACCTGTCCAGCAAAAACCTCAAAGC
 AGCTAGATCACCCGATGGTGTCTGGCTATGGTGTGGCTAGGATTTCTT
 CATTTTTCTCAAAGCTCATGTGGTCAGAGGGATTACCATAGTTGGTTCT
 CCACTT---GAGCACTTTCGCCAATTTGGCCC-----CAAGATTA
 CACCAAAATCTGCGATGTAATTGAAGAACATGAGAAAGAAATGCAAAAGC
 TAGCAAGGAGGTTGACATGGTTAATGCTGGGCTCATTGGGCATAACCAAG
 AAAGATTTAAACTGGGCTGGACCGAAAGGTGAATCCAAA---GAAGGTGG
 TGCAGCCTTGCAACTGAACTCCTACCCGGCTTGCCAGATCCAGATCTGG
 CAATGGGTCTTGCTGCACACACAGACTCCACTCTGCTGACCATTTCTTAC
 CAGAACAACACAAGTGGATTGCAAGTGCTGAAAGAGGGAATTGGG---TG
 GGTCAACCGTCCCAACCAATCCAGGTGGGCTTGTGTAAATGTAGGCGACC
 TACTCCATATCCTATCAAACGGGTTGTACCCGAGTGTGCTCCACAGAGCG
 GTGGTTAACCGAACCAACACAGATTATCCATAGCGTACCTCTACGGGCC
 ACCATCAAGTGTTCAAATATCACCCATACAAAACTTGTAGGCCCAAATC

ATCTCTCTCTATCGCCCAATCACTTGAATGAATACCTTGTGCTAAA
 GCTAAGCATTTCAACAAAGCATTATCATCGGTTAGGATTTGTGCTCTCT
 AAATGGA---CTAGTTGATGTAACGATCAT-----
 -----AATAGTGTAAGTTGGT
 >R
 GCCCCACAAAGA-----

 -----AAGATTTGGGGT-----
 -----GGCGATGGTATC-----

 -----TACAAAAGAAAT
 CCAATT-----
 -----GATATAACAGAAGAATACCAAATAGAGATGCAAAAAC
 TTGCAGTCAGATTGATGTGGTTAATGCTGGGCTCCTTAGGCATAACTAGC
 GAAGATGTCACCTGGGCTGGCCAGAGGTGATTTCAAGGGGCTTCTGC
 TGCCGCCTTACAACTGAATTACTACCCGGCTTGCCCTGACCCGGATAAGG
 CCATGGGTCTTGCTGCCCATAGTCACTCAACTCTGCTTACCATCCTACAC
 CAGAATAGTACCAGTGGATTGCAAGTACAAAAAGGAACCGGG---TG
 GGTACCGTTCCGCCTATTCCGGGTGGTCTTGAGTCCATGTAGGAGACC
 TCCTTCACATCCTATCCAACGGGTGTACCCAGTGTATTACCCGGGCG
 GTAGTTAACCGGACCAAGCATAGGCTATCCGTGGCTTATCTGTATGGGCC
 ACCATCAAGTATCCAAATATCACCTTATCTAAATTAGTAGGCCGAGTC
 AACCACTCTTTACAAAGCAGTCACTTTCTTTGAGTATCTTGCACTAAA
 GCTAAACACTTCAACGAAGCACTATCAAGGGTTAGGGTATGTGCTCTCT
 GAGTGGATCATTACTTGATGTAATGAGCACAGCAACAACAACAACA
 GATAATAATAATAGGGTAAAGTAGGA
 >0
 ---CCGACG-----CCGTCGCACTT
 GAAGAACCCGCTCTGCTTCGACTTCCGGGCGGCGAGGCGGGTGCCGGAGA
 CGCACGCGTGGCCGGGCTGGACGACCACCGGTGGTGGACGGCGGGG
 GCGGGCGGCGAG-----GACGCGGTGCCGGTGGTGGACGTCGGGGC
 GGGCGACGCGGCGGCGGGTGGCGGGGCGGCGGAGCAGTGGGGCGCGT
 TCCTTCTGGTCGGGCGGCGTCCGGCGGCGCTGCTGTCGCGCGTCGAG
 GAGCGCTGCGCCGCGTGTCTCCCTGCCGCGTGGGAGAAGATGCGCGC
 CGTCCGCGGCCCCGCGGAGCCCTGCGGCTACGGCTCGCCGCCCATCTCT
 CCTTCTTCTCAAGCTCATGTGGTCCGAGGGCTACACCTTCTCCCTTCC
 TCCTTCCGCTCCGAGCTCCGCCGCTCTGGCCCAAGTCCGGCGACGACTA
 CCTCTCTTCTGTGACGTG---GAGGAGTTTCAAGGAGATGCGGCGGC
 TAGCCGACGAGTTGCTGAGGTTGTTCTTGAGGGCGCTGGGGCTCACCGC
 GAGGAGGTGCGCGGAGTCGAGGCGGAGAGGAGGATCGGCGAGAGGATGAC
 GCGGACGGTGCACTCAACTGGTACCCGAGGTGCCCGAGCCGCGGCGAG
 CGTGGGGCTCATCGCGCACACGGACTCGGGCTTCTTCACTTCGTGCTC
 CAGAGCCTCGTCCCGGGGCTGCAAGTGTTCGTCGAGGGCCGACCGGTG
 GGTGGCGGTGCCGGCGGTGGCGGGGGCTTCTGTCGTCAACGTGCGGACC
 TCTTCCACATCCTCACCAACGGCGCTTCCACAGCGTCTACCAACGCGCC
 GTCGTGAACCGGACCGCGACCGGGTCTCGCTCGGCTACTTCTCGGGCC
 GCCGCGGAGCCGAGGTGGCGCGCTGCCGAGGCGGTGCCGGCGGCC
 GGAGCCCCGCTACCGCGCTGTACGTGGCGGAGTACATGGCGTCCGC
 AAGAAGGCTTCCGACCGGCGGCTCCGCCCTCAAGATGGTCTCCACCGA
 CGCCGCCCGCGCGGACGAACACGACGACGTGCGCGCGCGCGGACG
 TCCACGCA-----
 >V
 ---CCTTCGAGA---ATCTCTGATGCCTTCAAAGCCACCCTCTCCACCT
 CAATCACAGACACTTG---GACTTGAACCTAGTACAAGAATTACCCGACT
 TATATGAATGGCCGACTTATATGCATGGGCAAGGGTGGACGAGAACCCT
 TCTGGAGACTCGTTAATCACTGAGTCCGTGCCGGTTATTGATCTACCGA
 CCCTAATGCCTCAGAACTCGTGGGTGATGCATGCAAGAGCTGGGGCGTCT
 TCCAAGTCACCAACCATGGCATCCCCGGCAGCCTCCTCGACGACATTGAG
 TCTGTGGCAGGAGCCTTTTCTCTACCTGCCAGCAAAAGTGAAGGC
 CGCTCGTCCCCGTGATGGGGTTGCTGGCTATGGCTCGCCAGGATTTCTCT
 CTTTCTTCAACAAGCTCATGTGGTATGAAGGCTTACCATCTTCGGATCT
 CCACTT---GAACACGCTCGCCAACTTTGGCCT-----CAAGACTA
 CACCAAATCTGTGATGTCACGGAGGATTTGAGAAAGAGATGAACCAAC
 TTGCAGAGAGGCTCATGTGGCTCATGCTTGGATCACTGGGGATCACTAAA
 GAAGATCTTAACTGGGTGGCTCAAAGGCGACTTCAA-----
 -GCTGCCCTGCAAGTTGAACTCTTACCCAGCTTGTCAGAGCCGGACAGGG
 CGATGGGTCTCGCTGCCACACTGACTCTAGCCTCTTCAATCCTCTAC
 CAGAACACCGTCAGTGGTCTGCAAGTCCAACGAGAGGGAGCCGGA---TG
 GATCACAGTTCTCCACTCCCGGGTCACTTGTATCAACGTTGGTGACC
 TCCTCCACATACTATCCAATGGGGTGTCCCTAGTGTGCTACACAGGGCC
 TTGGTTAACCGGACAAAGCATCGACTCTCTGTGGCTACCTCTACGGACC
 ACCAGCCGGTGTCCAATCTCGCCCGTGCCGAAGCTGGTTGACTCGACTC

ACCCTCCTCTCTACCGCCAGTCACATGGAGCGAGTACCTCTGCACCAAG
 GCAAAGCACTTTGACAAGGCTCTTTTCATTAGTTCGGCTGTGTATGCCAG
 GAACGGA---TTCATCGATGTAACGATCAC-----
 -----AATGGGGTGAAAGTAGGT

GA.1.14.13.78.fasta

>A

-----GCCTTCTTCTCCATGATCTCCAT
 TCTCCTTGGCTTTGTTATCTCCTCCTTCATCTTCATCTTCTTCTTCAAGA
 AACTTCTCTCCTTCTCCAGAAAGAACATGTCTGAAGTCTCCACTCTCCCC
 ---TCTGTTCCAGTGGTACCAGGTTTCTGTTATTGGGAAGTTGCTGCA
 ACTAAAAGAGAAGAAACCTCACAGACTTTCACTAGATGGTCAGAGATTT
 ATGGTCCTATTTACTCTATAAAGATGGGTTCTTCTTCTTATTGTCTC
 AATTCTACTGAGACTGCCAAAGAGGCCATGGTGACGCGGTTTTCTGTCTAT
 CTCAACGAGGAAGTTGTCAAATGCGTTGACAGTCCTTACTTGTGACAAAT
 CTATGGTTGCTACTAGTGATTATGATGATTTCCACAAGTTGGTGAAACGG
 TGTCTCTTGAACGGTCTTTTGGGTGCTAATGCACAGAAACGAAAAAGACA
 TTACAGAGATGCACTCATTGAAATGTGTCTTCCAAGTTGCATGCCCATG
 CTAGGGACCATCCACAA---GAACCTGTAACTTCAGAGCTATATTTGAG
 CATGAGCTTTTTCGGTGTAGCATTGAAGCAAGCTTTTGGGAAAGATGTGGA
 ATCCATTTATGTTAAAGAACTCGGTGTGACTTTGTGAAAGACGAGATCT
 TCAGAGTTTTAGTACATGACATGATGGAAGGTGCAATTGATGTTGATTGG
 AGAGACTTCTTCCCATACTTGAATGGATTCCAAATAAAAGTTTTGAAGC
 AAGAATCCAGCAAAAAGCATAAACGTAGACTCGCAGTGATGAATGCTCTGA
 TTCAAGATCGACTGAAGCAGAATGGTTCAGAATCGGATGATGATTGCTAT
 CTCAACTTCTTGATGTCTGGAAGCG---AAAACACTAACCAAGGAGCAAAT
 TGCTATCTTGGTTTGGGAGACGATTATCGAGACAGCTGACACTACTTTGG
 TTACAACCTGAATGGGCCATCTATGAGCTCGCTAAGCATCCAAGTGTTCAA
 GATCGTCTGTGCAAGAAATCCAAATGTCTGCGGAGGAGAAAAGTTCAA
 AGAAGAGCAATTGTCTCAAGTTCCTTATCTCAATGGAGTCTTTCATGAAA
 CGCTTAGGAAATACAGTCTCTCTCTAGTTCCCATTCGCTACGCCAC
 GAGGATACGCAAAATCGGAGGCTATCATGTCCCTGCAGGAAGTGAGATTGC
 AATAACATATATGGATGCAACATGGATAAGAAGCGTTGGGAGAGACCAG
 AGGACTGGTGGCCGAGCGGTTTCTTGATGATGGCAAATATGAAACGTCA
 GATCTTCCACAAGACAATGGCGTTTGGAGCGGGAAGAGGGTTTTGTGCTGG
 TGTCTTTCAAGCATCTCTCATGGCAGGCATTGTCTATTGGAAGATTAGTGC
 AAGAATTCGAGTGGAAGCTTAGAGATGGCGAAGAAGAGAAATGTGGATACA
 TATGGCTTGACCTCTCAGAAGCTTTATCCTCTTATGGCTATTATCAATCC
 AAGG---CGTTC

>P

 -----GTGCCAGGATGGCCAGTCATAGGGAATTTGCTGCA
 ACTGAAGGAGAAGAAACACACAAGACGTTCTTAAGGTGGGCTGAGGCTT
 ATGGGCCAATCTATTCCATAAAGACCGGTGCTTCTACTGTGATTGTTCTC
 AATTCAACTGAAGTTGCAAGGAGGCCATGGTGACTAGATATTCTGTCAT
 ATCAACAAGAAAAACTATCAAGGCCTTGGAAGTTCTTACAGATAATAAAT
 CTATGGTTGCAACAAGTGATTATGGCGACTTTCACAAGATGGTGAAGCGG
 TATATACTTACAAATGTACTCGGAGCTGGTGCTCAGAGACGACACCGGGG
 TCACAGGGATACCTTGGTAGAAATGTTTCAAGTCAATTGCTTGATCATA
 TAAAGACCAATCCTCAATTACAAGCTGTGGATTTCAGGGAATATTTGAG
 TCTGAACTTTTTGGATTGTCAATGAAAGAAGCTTTAGGAAAAGATATGGA
 ATCTCTATATGTAGATGAACCTTCAGGCCACTTTATCAAGAGAGGAGATTT
 TCAATGTTCTGGTGCTTGATCCTATGGAGGGTGCAATTGATGTTGACTGG
 AGAGATTTCTTTCCATATTTGAGATGGATTCTTAATAAGGGTTTTGAAAT
 GAAAATCGAGCGAATGAATTTTCGAGGCAATCGGTGATGAATGCCCTTG
 TGCAGGAGCAGAAAGAAACGAATTGCATCTGGAGAGGAAATAAATTGTTAT
 ATTGACTACTTATTATCGGAAGGCAAACTACACTTACGGAGAAGCAAAAT
 TGGTATGCTGGTTTGGGAGACAATTATTGAGACATCAGATACAATATGG
 TCACTACAGAATGGGCCATGTATGAACTAGCGAAGAACCACAAATGCCAG
 GATCGGCTCTATCATGAAATCCAAATGTCTGTGGGTCTGAGAAGCTTAA
 AGAGGAACACCTGTCCCAACTGCCATACTTAAATGCTGTTTTCCATGAAA
 CTATTAGAAAGTACAGTCTGCTCTATTATACCTTTGAGATATGCTCAT
 GAAGATACACAGATAGGGGGCTACTATGTTCCAGCTGGAAGTGAGATCGC
 TATTAACATTTATGGATGTAACATGGACAAGAAGCGTTGGGAAAACCTG
 AAGAGTGGAAAGCCTGAGAGATTCTTTGAC---GGCAAATATGATCCCATG
 GACTTGCACAAGACAATGGCATTGTGAGCAGGAAAGAGGTGATGCGCAGG
 TGTCTTTCAAGCGTCTTGTATAGCATCTGCCACGATTGGCAAGGTAGCAC
 AAGAATTTGAATGGAGACTTAAAGATGGAGAGGAAGAACATGTTGATACT
 GTTGGACTACCACTCGCAAGCTTCAGCCGCTGCATGTCATGATAAAAAAC
 AAGA-----
 >R

GATGCTGTACCCCTAACATTCTTCTAACTTTCAGACAATGACTTATGC
CACCCCTGCAGCTGTTGGTGGTTTGTCTCTCCATGTTCTGTTATGAAA
AGCTGATCTCCAATCAAAGAAAAGACATGCCAGA-----CTTCCT
---CCTGTACCAGAAGTACCAGGGTGGCCAGTGATAGGGAATTTGCTGCA
ACTCAAAGAGAAAGAAACCCCATAAAGACTTTCATGAGATGGGCTGAGACAT
ATGGACCTATTTATTCGATTAGAACTGGTGCTTCTACTATTATTGTTCTC
AATTCAACTGACGTTGCTAAAGAGGCCATGGTGACTAGATATTCATCCAT
TTCAACAAGAAAGCTATCAAAAGCCTTGGTAATCCTTACCCAGGATAAAT
GTATGGTTGCAACAAGTGATTATGATGAGTTCCACAAGATGGTGAAACGA
TATTTACTTACAAATGTAAGTGGGAGCCAATGCTCAGAGGCGACATCGGTG
CCACAGGGACACCCTGATCAAAAATATTTGAGCCGGTTTCATGCTCATG
TAAAGACCAATCCTGAA---CAAAGTGTAAATTTTCAGGGATATTTTTGAG
TCTGAACTTTTTGGACTATCACTTAAAGAAGCATTGGGAAAGGATGTGCA
ATCCATTATGTTGATGAACTTGGGACCACTCTGTCAAGGAAGGAGATTT
TTGAGGTTTTGGTGCTTGACCAATGGAAGGGGCAATTGATGTGGACTGG
AGAGATTTCTTCCCATATTTAAGATGGGTTCTAATAAGAGCTGGGAAAT
GAAAATCCAGCAGATGCATTTTCGTAGGAAGGCAGTGATGGGTGCCCTCA
TCAAGGAGCAGAAAGAAACGAATTGCTCTAGGGGAGGAGGTAAATTGTTAT
CTCGATTTCTTATCATCCGAA-----AGCACACTCACGGATGAACAAAT
CAGAATGTTGATCTGGGAGACAATTGTTGAGACATCAGATACGACTATGG
TCGCTACAGAATGGGCCATCTACGAACCTTGCTAAGAATCCCAACTGCCAG
GATCTGCTCTATCGCCAAATTCAAAATGTATGTGGGTCTGAGAAGATCAC
AGAGGAACACCTGTCTAAACTGCCATACTTGAATGCTGTTTTCCATGAAA
CTATTAGAAAGTACAGTCTGCTCCAGTTATACCTCTGCGTTATGTACAT
GAAGATACAGAATTAGGGGGATACTACATTCTGCAGGAAGTGAGATTGC
AATTAATATTTATGGGTGCAACATGGACAAGAATCGATGGGTAAACCCCTG
AAGAGTGGAATCCTGAGAGATTTCTCAAC---GGAACTATGATTCCGGG
GACTTGCAATAAGACAATGGCTTTTGGAGCCGGAAGAGGGCTTGTGCAGG
TGCTCTTCAGGCATCATTAATAGCCAGTACATCAATTGGCAGACTGGTG
AAGAATTTGCAATGGGGACTGAAAGCTGGAGAGGAAGAAGATGTTGATATT
GTTGGACTGACAACTCGCAAGCTTCATCCATTGCATCTCTCCATAAAGCC
AAGA---AGACTG

>O

-----GAGTCGCT
GCTCGCAGCCGGTGCAGGAGGGATCGGTGTGGCGGCGGCGGCCGTCGGGG
GGTTTATCGCGGCGGCGACACTCGCCGTGGCACCGCCCAAGAACCGCCGC
AACCCGCTCCAGCTGTTCTGGTTTACCCATAATCGGAAATCTGCATCA
ATTGAAAGAAAAGAAGCCTCATCAGACCTTCACAAAATGGGCTGAAATTT
ATGGCCCAATCTACACTATAAGGACCGGGGCTTCTTCCGTAGTTGTGCTC
AATTCAACTGAAGTAGCCAAAGGAGGCGATGGTTGCAAAATCTCATCCAT
ATCTACCCGAAAGCTATCCAAAGCACTAACAGTGCTTAGTCATGATAAAT
CCATGTTTGTACTACAGCGACAGTGGTGATTTCCACAAAATGGGGAAGCGT
TATATCATGTTAAGCATGCTGGGTACTTCTGCACAGAAACAATTTCTGTA
CACAAAGAGATATGATCATCAATAACATGTTAAGCACTTTCCATCAACTGG
TGAAAGATGACCACAT---GCTCCTCTGATATTTCAGAGATGTTTTCAAG
GATGAGCTATTCCGGTTGTCCATGATCCAGAGCTTAGGAGAGGATGTGAG
TTCAGTCTATGTGGATGAATTTGGGAGGGACATTTGGAAGGAAGAAATCT
ACAATGCTACTGTGACTGACATGATGATGTGCGCAATTGAGGTCGACTGG
AGAGATTTCTTCCCTACCTCAGCTGGGTTCCAAACAAGAGCTTCGAAAC
AAGAGTGTTTACTACAGAATCTAGACGAACGCGGTGATGCGCGCCTTGA
TCAAGCAGCAGAAGGAAGGATTGTGCGTGGAGAGGCAAGGACATGCTAT
CTGGACTTCTTGTGTCAGAG-----AACACACTGACAGATGAGCAACT
AATGATGCTAGTGTGGGAGGCACTCATAGAGGCTGCAGATACTACCTTGG
TCACCACAGAATGGGCCATGTATGAGCTTGCCAAGAACCCTGACAAACAG
GAACGGCTTTACCAAGAGATCCGGGAGGTGTGCGGCGACGAGCGGTAC
CGAGGAGCACCTGCCGTGGCTGCCGTACCTCAATGCCGTCTTCCAGGAGA
CGCTGCGCGGCCACTCCCTGTCCCGCTCATACCTCCGAGGTTCTGTCAC
GAGGACACCATGCTCGCCGGCTACGATGTCCCGCCGGCACCGAGATGGT
GATCAACCTGTATGGGTGCAACATGAACAAGAAGGAGTGGGAGTCGCCGG
AGGAGTGGGCGCCGAGAGGTTGCGCGGT---GGGAGGTTCAAGGTGGCG
GACATGTACAAGACGATGGCGTTGCGCGCCGGGAGGAGGTTCTGCGCAGG
GAGCTGCGAGGCGACTCACATCGCGTGCAGCGCCATCGCGCGCTTCTGTC
AGGAGTTTGGGTGGAGGCTGAGGGAGGTTGACGAGGAGAAGGTGGACACC
GTGCAGCTCACCGCTACAAGCTCCACCGCTTCATGTCCACCTCACGCC
CAGAGGAAGGATG

>V

-----TTGATGTTGATGAATG
GGTATTGGTTGATG-----
---GGGTTGCAGATATCCAGGGCTGCCGCTGATTGGGAATTTGCTGCA
ACTGAAAGAGAGAAGAACACACAAGACTTTCGCAAGGTGGGCTGAGACTT
ATGGCCCAATCTTTCCATCAGGACCGGAGCTTCTACTATGATCGTTCTG
AATTCCTCCGAAGTTGCTAAAGAGGCAATGGTGACCAGGTTCTCATCCAT
CTCAACAAGGAAGTTATCAATGCCTTGAAGATACTTACCTTCGATAAAT
GCATGGTTGCCACAAGTGACTACAATGACTTCCACAAGATGGTCAAAGGG

TTTATACTCAGAAATGTTTTGGGAGCCCCGCTCAGAAGCGACACCGATG
CCACAGAGACACCTTGATTGAAAAATTTTCGAAATACTTGCATGCCCATG
TGAAGACCTCTCCTCTT - - GAACCAAGTGGTTCTCAAAAAAATTTTCGAG
TCTGAAATTTTCGGATTAGCGTTGAAACAAGCTTTGGGCAAGGACATTGA
ATCCATTTATGTGGAAGAGCTTGGCACCACCTTTGTCAAGGGAGGAGATAT
TTGCGGTTTTGGTTGTTGATCCG - - GCAGGGGCCATTGAGGTGGACTGG
AGAGATTTCTTCCCATATCTGAGTTGGATTCTAATAAGAGCATGGAAT
GAAAATCCAGAGAATGGATTTCCGCAGGGGAGCATTAAATGAAGGCCCTTA
TCGGGGAGCAGAAAGAAAAGGATTGGTTCAAGGAGGAGAGAAGAAATTGCTAC
ATCGATTTCTTATTATCCGAAGCA - - ACAACACTCACGGAGAAACAAT
AGCCATGTTGATTGGGAGACAATCATCGAGATATCAGATACTACTCTGG
TCACATCTGAATGGGCCATGTATGAACCTTGCTAAGGATCCAAACCGGCAG
GAAATTTTATCGGGAAATCCGCAAAATTTGTGGCTCCAATAAGGTTAC
AGAGGAAAACTTGCTAAGCTTCCGTACTTAAATTCGTTTTCCATGAGA
CTCTTAGGAAGTACAGTCCAGCTCCCATAGTCCCTGTTGCGTATGCGCAT
GAAGATACCCAATTGGGGGGATACCATATTCCTGCTGGAAGTCAGATCGC
TATAAACATATATGGGTGCAACATGAACAAGAAACAATGGGAAAAATCCTG
AGGAGTGGGAAGCCTGAAAGATTCTGGAT - - GAGAAATACGATCTAATG
GATTTGCACAAGACGATGGCCTTTGGAGGAGGAAAGAGGGTGTGTGCAGG
CGCTCTTCAGGCAATGCTAATAGCTTGCACATCAATGGTAGGTTCTGTGC
AGGAGTTTGAGTGGAACTGATGGGAGGGAGGAAGAAAATGTGGATACC
GTAGCACTCACCAGTCAGAACTCCATCCCATGCAAGCGATCATAAAGGC
AAGA - - GAG - -

GA.1.14.13.79.fasta

>A

-----ACGGAAACGGGTTTGATTCT
CATGTGGTTCCTTTGATAATATTGGGATTGTTTGTGTTTGAAATGGGTGC
TGAAGAGAGTGAATGTTTGGATTATGTGTCCAACTTGGT - - GAGAAA
AAACACTATCTGCCACCAGGTGATTTGGGATGGCCAGTCATAGGCAACAT
GTGGTCCTTTCTTAGAGCTTTCAAAACATCTGATCCTGAATCCTTCATCC
AATCCTACATCACCAGGTATGGGCGTACCGGGATTTATAAAGCACACATG
TTTGGGTACCCATGTGTACTAGTAACAACACCAGAGACTTGTAGGCGAGT
TCTAACAGATGATGATGCCTTCCACATAGGTTGGCCAAAATCTACCATGA
AACTCATTGGCAGGAAGTCCTTTGTTGGTATCTCCTTTGAAGAACACAAG
AGGCTCAGGCGTTTGACTTCTGCTCCTGTCAATGGCCCTGAAGCTCTCTC
TGTATACATACAGTTCATTGAAGAACTGTTAATACTGATCTAGAAAAAT
GGTCCAAA - - - - ATGGGAGAAATCGAGTTCTTATCTCATTGCGTAAG
CTTACGTTTAAGGTCATTATGTACATATTTCTCAGCAGTGAGAGTGAGCA
TGTCATGGATTCAATTGGAACGGGAGTATACCAACCTTAACATGGAAGTTC
GAGCAATGGGTATTAATCTTCTGGGTTTGCTTATCATAGAGCTTTGAAG
GCAAGGAAAAAAGCTTGTAGCTGCCTTTTCAGTCCATAGTACTAACCGAAG
AAATCAAAGGAAGCAGAATATTTATCCAAT - - - - AGAAAAGATATGC
TGGATAATCTAATAGATGTTAAAGATGAAAACGGAAGAGTCTTAGATGAT
GAGGAAATTATTGACCTTTTGTGATGTATCTTAATGCGGGTCATGAATC
TTCTGGACACCTCACTATGTGGGTACCATTTTGATGCAAGAACACCCCTA
TGATTCTGCAAAAGGCTAAGGAAGAACAAGAGAGGATCGTAAAAAGAGA
GCACCGGACAGAGAAG - - TTGACGCTTAAAGAGACACGTGAAATGGTGT
TCTTTCTCAGGTTATTGATGAGACCTTCGAGTAATTACATTCTCTCTGA
CGGCCTTCAGGGAAGCAAGAGTGTGTCAAATGGATGGCTATATAATT
CCAAAAGGCTGGAAGGTTCTGACTTGGTTTAGGAACGTCCATTTGGACCC
TGAATCTACCCGGATCCAAAGAAATTTGATCCTTCAAGATGGGAGGGAT
ACACACCAAAAGCAGGCACATTCCCTTCTTTGGTTTGGGAAGCCACCTA
TGCCCGGGAACGATCTTGCCAAGCTCGAGATTCCATTTTTCTTCATCA
TTTCTCTCTCAATACCGGGTGGAAAGGAGCAATCCCGGATGTCGGGTGA
TGTTCTTGCTCACAATCGACCCAAAGATAATTGCCTAGCAAGAATCACC
AGAACGATGCCA - - - - -

>P

-----ATGGAGTCGGGATCTATCTG
GGTGGTCCTAGCAGTCATATTTGGTGGATTAGGGGTTGGGAAATGGATT
TAAAGAAAGTGAATTGGTGGTTGTATGAAGCCCAACTGGGT - - GAAAAG
CAGTACTCACTACCTCCAGGGGATTTGGGCTGGCCTTTCATTGGCAATAT
GTGGTCCTTTCTCAGAGCTTTCAAATCCAGCGATCCTGATTCTTTCATGC
GCACCTTCATCAACAAATATGGAGATAATGGAATCTACAAGGCCTTCATG
TTTGGGAACCCGAGTGATTTCGTAACAACACCTGAAGCATGCAGAAGAGT
ATTAAGTAACGATGATGCATTTAAGCCTGGATGGCCAATTTCCACACTGA
AGCTTATCGGAAGGAAATCCTTCATTGATATTTCTTACGAAGAGCACAAAG
CGCTTCGCGCTCTGACGTCCGCTCCTGTCAATGGTCATGAAGCGTTGTC
CGTTTACATTCATATATAGAAGAAAATGTGATAGCTATGTTGGAAAAAT
GGACCACA - - - - -ATGGGGAAGATCGAGTTCTTGAAGTCAAGTGAGAAAA
CTTACCTTCAAAATAATCATGTATATATTTCTTAGCTCTGAGAGTGAGGT
GGTCATGGAGGCTTTGGAGAAGGATTACACCACTCTTAATTATGGAGTCA
GAGCCATGGCGATCAATCTTCTGGATTGCTTACTATAAAGCATTGAAG

GCCCGCAAAAGACTTGTGCTATATTCCAATCTATAGTGGATGAGCGTAG
AAATTTAAGGAAGAACAGTGCTCGTAACGCA - - - AAGAAGAAAGACATGA
TGGACTCTCTGTTGGGTGTTGAAGATGAAAATGGTAGGAAATTGACTGAT
GAAGAAATCATAGATGTAATTTTGATGTACTTGAATGCAGGCCATGAATC
CTCTGGCCATATCACGACATGGGCTACTATTTTTCTCCAGGAACATCCAG
AATTTCTCCAAAAGCTAAGGAAGAGCAAGAACAGATTGTA AAAAGGAGG
CCACCAGCACAAAATGGATTGTCACTTAAGGAAGTTCGAGAAATGGATTA
TCTTTCCAAGGTGATTGATGAACACTTCGTTTGATCACATTCTCGCTTA
CTGTTTTCCGAGAAGCAAAAACAGATTTTCAGTATAAACGGATACATCATT
CCAAAGGGTTGGAAAGTTCTGGTTTGTTTCAGGACAGTTCATTTGGATCC
AGAAATCTATCAAAATCCAAAGGAATTTAATCCTTCTAGATGGGATAATT
ACACCCCAAGCAGGGACTTTCTTCTCTTTGGAGCTGGAAGCAGGCTG
TGCCCTGGAATGATCTTGCCAAGCTTGAAATATCTATTTTTCTGCATTA
TTTTCTCTTGATTATAGGCTCGAACGTGAAAATCCTGAATGTCGTTGGA
TGTTCTCCTCATACCAAGACCAACAGACAACGTGTGGCTAGAATCAAG
AAAGTTTCATCCACATCTGTA

>R

-----GAGATGGGTTTCGTGTG
GGTGGTCTGATATGGATATCGGGGGGTTTTTGGTGCTTGAAATGGATTCT
TAAAGAGAGTAAATTGCTGGCTCTATGAAAACCAACTCGGG - - - GAAATG
CAGTATTCCTGCCTCCAGGCGATTTGGGGTGGCCTTTTCATTGGTAATAT
GTGGTCTTTTCTTAGAGCTTTCAAGTCCAACGACCTGATTCTTTTCATGC
GCAACTTCACTGCCAGATACGGGAGTGGTGAATATACAAGGCCTTCATG
TTTGGGAAGCCAGTGTGTTGGTAACAACATCAGAAGCATGCAAAAGGGT
ATTAACAGATGACGACGCTTTTAAGCCTGGTTGGCCAAGCTCCACCACGG
AGCTCATTGGCAAGAAATCATTGTTGTTTTCGTACGAAGAGCACAAAG
CGTCTTCGCAAGGCTCACAGCCTCTCCTGTCAATGGACATGAAGCTCTTTC
CGTCTACATGCACTACATTGAAGATAAAGTCAAATCTGCCTTGAAAAAT
GGTCCACT - - - - ATGGGAGAGATCCAGTTCCTGACTCAACTCAGAAAA
CTTACTTTTCAGGATCATTATGTATATCTTCTTAGCTCCGAGAGTCACTC
CGTCATGGAGGGCTTTGGAGAGGGAATATACCACTCTTAACATATGGGGTCA
GAGCCATGGCCATTAATCTTCTGGATTGCTTACTATAAAGCGCTCAAG
GCGCGCAAAAACCTTGTGCTGTACTACAATTTGTAGTGGATGCGCGAAG
AAATCAGAAAGAGGGCGCAGAGCCTAATTCATCAAGAAGAAAGACATGA
TGGACGCTCTTTTGGATGTTGAAGATGAAAAGGGTAGAAAATTGAGCGAT
GAAGAAATCGTTGATGTTCTGTGAATGTACTTGAATGCAGGCCATGAATC
TTCTGGCCATATCACTATGTGGGCTACTGTTTCTTTCAGGAGCACCTG
AATTTCTCCAAAAGGCCAGGCAAGAACAAGAGGAGATCAAAAAAGGAGG
CCACCGACGCAAAAAGGATTGACGCTTAAAGAAGTCCGAGACATGGAATA
TCTTTCTAAGGTCAATTGATGAAACTCTTCGTTTGATAACCTTCTCGCTTG
TAGTTTTCCGAGAGGCAAAAACAAATGTCAATATAAGCGGATATGTCATT
CCCAAGGTTTGGAAAGTCTAGTCTGGTTCAGAAGTGTTCATTGGATCC
TGAAATATACCCAAATCCAAGGGAATTCAATCCTTCCAGATGGGATAATC
ATACAGCAAAAGGCAGGAACTTTCTTCCGTTTGGAGCAGGAAGTAGGATG
TGCCCTGGAATGATCTAGCCAACTCGAAATGCTATTTTTCTGCATCA
TTTTCTGCAATTATGAGCTAGAACGTCTCAATCCTGGAAGCTCAATGA
TGTACTTGCCCTCATTCAAGACCAAAAGACAACGTGTTGGCAAGAATCAAG
AAAATTCCTTCTGCA - - - - -

>O

GTGATGGAGGGCATGGGCATGGCGGCGGCGTGGGCGGCGGGGACCTGTG
GGTGCTGGCGGCGGCGGTGGTGCCGGCGTGGTGCTGGTCGACGCGGTGG
TGCGGAGGGCGCACGACTGGGTTTCGCGTGGCGGCGCTGGGGGCGGAGAGG
AGGTCGAGGTTGCCCGGGGGAGATGGGGTGGCCGATGGTGGGCAGCAT
GTGGGCGTTCTCCGCGCTTCAAGTCCGGCAACCCGACGCCTTCATCG
CCTCCTTCATCCGACGGTTTGGGCGGACAGGGGTGTACAGGACGTTCC - - -
TTCAGCAGCCCGACGATCCTGGCGGTGACGCCGAGGCGTGCAAGCAGGT
GCTCATGGACGACGAGGGCTTCGTCACCGGCTGGCCCAAGGCCACCGTCA
CCCTCATCGGCCCCAAATCCTTCGTCAACATGCTTACGACGACCAACGCG
CGCATCCGAAGCTCACCGCGCCCCCATCAACGGCTTCGACGCCCTCAC
CACCTACCTCTCCTTCATCGACCAGACCGTCGTGCGCTCCCTCGCCGCT
GGTCCTCGCCGAGTCCGGCCAGGTCGAGTTCCTCACCGAGCTCAGGCGC
ATGACCTTCAAGATCATCGTCCAGATCTTCATGAGCGGCGCCGACGACG
CACCATGGAGGCCCTGGAGCGGAGCTACACCGACCTCAACTACGGCATGC
GCGCCATGGCCATCAACCTCCCGGCTTCGCTACTACCGCGCGCTCAGG
GCTCGCCGGAAGCTCGTGTCCGTGCTGCAAGGTTGTGCTGACGGCCGGAG
GGCGCCGCGCCCAAGGGCTTCAAACGCTCC - - - - - GGGCCGACATGA
TGGACCGCTCATCGAGGCCGAGGACGAACGCGGCCGCGCCTCGCCGAC
GACGAGATCGTCGACGTCCTCATCATGTACCTCAACGCCGGCCACGAGTC
CTCCGGCCACATCACCATGTGGGCCACCGTCTTCTCCAGGAGAACCCCG
ACATCTTCGCAAGAGCAAAGGCTGAGCAAGAGGAGATCATGAGAAGCATT
CCAGAACGCGAGAACGGATTAACCTCAGGGACTTCAAGAAGATGCACCT
CCTCTCACAGGTTGTGACGAGACACTTCGCTGCGTCAACATCTCCTTCG
TGTCCTTCGCTCAGGCCACAAGAGACATCTTTGTGAACGGTTATCTTATC
CCCAAGGGGTGGAAGGTTCAAGTGTGGTACAGAAGTGTGCACATGGATGA
CCAAGTTTATCCTGACCCCAAAATGTTCAACCTTCAAGATGGGAGGGAC

CCCCTCCGAAAGCCGGAACATTCCCTTCCATTTGGACTGGGAGCAAGACTG
 TGCCCTGGAAATGATCTTGCAAAGCTGGAGATCTCTGTCTTCCCTCCATCA
 TTTTCTCCTGGGTACAAAGCTGAAGAGGGCAAATCCAAAGTGCAGGGTGA
 GATATCTGCCTCATCCCCGGCCTGTGGACAACGTCTTGGCGACGATCACC
 AAAGTTTCCGATGAACAC ---
 >V
 -----GAGTTGGGTATGATCTG
 GGTGGCCTTTGGGGCCATTTTGGGAGGTGATTGGGTGTGAAATGGGTTT
 TGAGGAGAGCAAATTCATGGGTTTATGAAGTTAAGCTGGGA --- GAGAAG
 AGGTATTCTCTTCCACCAGGTGATCTTGGTTGGCCATTGATTGGTAATAT
 GTGGTCTTTTCTCAGAGCTTTCAAGTCCACTGATCCTGATTCTTTCATCT
 CCAGCTTTATCACCAGATTTGGACAACTGGTATGTACAAGGTGTTAATG
 TTTGGGAACCCTAGCATCATTGTTACTATACCTGAAGCATGCAAAAGAGT
 CTTAACAGACGATCAAATTTCAAGCCGGGCTGGCCTACATCCACCATGG
 AACTGATTGGAAGGAAATCATTTATTGGCATTACTAATGAGGAGCACAAA
 CGGCTCCGCCGGTTAACAGCAACTCCAGTCAATGGGCATGAAGCATTGTCT
 CATTTACATGCAATATATTGAAGACAACGTTATATCTGCCTTGAACAAAT
 GGGCTGCG --- ATGGGAGAATTTGAGTTCTTAAGTGCCTCCGAAAG
 CTTACTTTTAAAGATAATTATGTACATATTTCTTAGCTCAGAGAGCGAGCA
 TGTAAATGGAAGCTCTGGAGAGGGAATATACTTCACTCAACTATGGAGTTA
 GGTCCATGGCAATCAATCTCCCGGGTTTGTCTTACCATAAAGCACTCAAG
 GCTCGGAAAAATCTTGTGAACATTTTCAATCTATAAGTGAATGAGCGAAG
 GGATAGGAAGAAAGGCAATTCTCAAACAATG --- AAAAAAGACATGA
 TGGATGCTCTGCTGGACATCGAAGATGAAATGGAAGAAAATGAGTGAT
 GAGGAAATCATAGACATTCTTGTAAATGTACTTGAATGCCGGTCATGAATC
 TTCTGCTCACGTCACGATGTGGGCTACTGTCAAGCTACAAGAAAACCTG
 AATTTTTCAAAGAGCCAAGGCAGAGCAAGAGGAGATCATAAGGAAGAGA
 CCACCTAACAGAGAGATTGACGCTCAAGGAAATTCGAGAAATGGAGTA
 TCTTCCCAAGGTGATCGATGAAACGCTTCGTTGGATTACATTCTCATTG
 TGGTATTCCGAGAGGCAAAAGCAGACATCAATATATGTGGTTATACATT
 CCAAGGGATGGAAGGTTCTGGTTTGGTTAGATCCCTTCACTTTGACCC
 TGAACATATCCAGACCCAAAGGAGTTTAATCCTTGCAGATGGGATGATT
 ATACAGCAAAACCAGGAACCTTCTTCCCTTTGGATTAGGCAGCAGGCTG
 TGCCTTGCCAATGATCTTGCCAAGCTTGAAATTTCTGTTTTCTACACCA
 CTTTCTCCTTAATTATCAGCTGGAACGGCTGAATCCTGGGTGCCCGCGCA
 TGTACTTACCACACTCAAGGCCAGAGACAATTGCTTGGCAATAGTCCGA
 AAAGTAGCAGCTGAATCAGAA

GA.4.2.3.19.fasta

>A
 TCTATCAACCTTCGCTCCTCCGGTTGTTCTGCTCCGATCTCAGCTACTTT
 GGAACGGAGATTGGACTCAGAAGTACAGACAAGAGCTAACAAATGTGAGCT
 TTGAGCAAACAAGGAGAAGATTAGGAAGATGTTGGAGAAAGTGGAGCTT
 TCTGTTTTCGGCTACGATACTAGTTGGGTAGCAATGGTTCCATCACCG ---
 -AGCTCCCAAAATGCTCCACTTTTCCACAGTGTGTGAAATGGTTATTGG
 ATAATCAACATGAAGATGGATCTTGGGGACTTGATAACCATGACCATCAA
 TCTCTTAAGAAGGATGTGTTATCATCTACACTGGCTAGTATCCTCGCGTT
 AAAGAAGTGGGAATTGGTGAAAGACAATAAACAAGGGTCTCCAGTTTA
 TTGAGCTGAATTCTGCATTAGTCACTGATGAAACCATACAGAAACCAACA
 GGGTTTGATATTATTTTCTGGGATGATTAAATATGCTAGAGATTTGAA
 TCTGACGATTCCATTGGGCTCAGAAGTGGTGGATGACATGATACGAAAAA
 GAGATCTGGATCTTAAATGTGAT --- AGTGAAAAGTTTTCAAAGGGAAGA
 GAAGCATATCTGGCCTATGTTTTAGAGGGGACA --- AGAAACCAAGATTG
 GGATTTGATAGTCAAATATCAAAGGAAAAATGGGTCACTGTTTGATTCTC
 CAGCCACAACAGCAGCTGCTTTTACTCAGTTTGGGAATGATGGTTGTCTC
 CGTTATCTCTGTTCTCTCCTTCAGAAATTCGAGGCTGCAGTTCTTCAGT
 TTATCCATTTGATCAATATGCACGCCTTAGTATAATTGTCACTCTTGAAA
 GCTTAGGAATTGATAGAGATTTCAAACCGAAATCAAAGCATATTGGAT
 GAAACCTATAGATATTGGCTTCGTGGGGATGAAGAAATATGTTTGGACTT
 GGCCACTTGTGCTTTGGCTTTTCGATTATTGCTTGCTCATGGCTATGATG
 TGTCTTACGATCCGCTAAAACCATTTGCAGAAGAATCTGGTTTCTCTGAT
 ACTTTGGAAGGATATGTTAAGAATACGTTTTCTGTGTTAGAATTATTTAA
 GGCTGCTCAA --- AGTTATCCACATGAATCAGCTTTGAAGAAGCAGT
 GTTGTGGACTAAACAATATCTGGAGATGGAATTGTCCAGCTGGGTTAAG
 ACCTCTGTTTCTGA --- GATAAATACCTCAAGAAAGAGGTGAGGATGCTCT
 TGCTTTTCCCTCTATGCAAGCCTAGAAAGATCAGATCACAGGAGAAAAA
 TACTCAATGGTTCTGCTGTGAAAAACACCAGAGTTACAAAAACCTCATAT
 CGTTTGCACAATATTTGCACCTCTGATATCCTGAAGTTAGCTGTGGATGA
 CTTCAATTTCTGCCAGTCCATACACCGTGAAGAAATGGAACGCTTTGATA
 GGTGGATTGTGGAGAATAGATTGCAGGAAGTGAATTTGCCAGACAGAAG
 CTGGCTTACTGTTATTTCTCTGGGGCTGCAACTTTATTTTCTCTGAACT
 ATCTGATGCTCGTATATCGTGGGCCAAAGGTGGAGTACTTACAACGGTTG
 TAGACGACTTCTTTGATGTTGGAGGGTCCAAAGAAGAACTGGAAAAACCTC

ATACACTTGGTCGAAAAGTGGGATTTGAACGGTGTTCCTGAGTACAGCTC
AGAACATGTTGAGATCATATTCTCAGTTCTAAGGGACACCATTCTCGAAA
CAGGAGACAAAGCATTACCTATCAAGGACGCAATGTGACACACCACATT
GTGAAAAATTTGGTTGGATCTGCTCAAGTCTATGTTGAGAGAAGCCGAGTG
GTCCAGTGACAAGTCAACACCAAGCTTGGAGGATTACATGGAAAAATGCGT
ACATATCATTTGCATTAGGACCAATTGTCTCCAGCTACCTATCTGATC
GGACCTCCACTTCCAGAGAAGACAGTCGATAGCCACCAATATAATCAGCT
CTACAAGCTCGTGAGCACTATGGGTCGTCTTCTAAATGACATACAAGGTT
TTAAGAGAGAAAAGCGCGGAAGGGAAGCTGAATGCGGTTTCATTGCACATG
AAACACGAGAGAGACAATCGCAGCAAAGAAGTGATCATAGAATCGATGAA
AGGTTTAGCAGAGAGAAAAGAGGGAAGAAATTGCATAAGCTAGTTTGGAGG
AGAAAGGAAGTGTGGTTCCAAGGGAATGCAAAGAAGCGTTCTTGAAAAATG
AGCAAAGTGTGGTGAACCTATTTTACAGGAAGGACGATGGATTACATCAAA
TGATCTGATGAGTCTT---GTTAAATCAGTGATCTACGAGCCTGTTAGCT
TACAGGAAGAATCTTTAACT-----

>P

----- CAGAACT
TTCATGGAACCTAAGGAAAGGATCAAGAAGATGTTTGATAAGATTGAATTG
TCAGTGCTTTCATATGACACTGCTTGGGTGGCAATGGTCCCATCTCCA--
-GATTGTCCGAAACTCCTTGTTTTCCAGAGTGCACAAAATGGATTTTGG
AAAATCAACTTGGTGATGGCTCCTGGAGTCTTCCT---CATGGCAATCCA
TTATTAGTTAAGGATGCTCTTTCATCTACATTAGCGTGCATCCTTGCATT
GAAGCGATGGGGTATCGGTGAAGAACAATAAATAAAGGCCTTCGATTTA
TTGAGTTGAATTCGCTTCAGTTACGGATAACGAGCAACATAAACCAATT
GGATTTGATATAATATTTCTGGCATGATTGAATATGCCATAGATTTGGA
TTTGAACCTCCCTTTGAAGCCGACAGATATAAATTCATGCTCCACAGGA
GGGCTTTGGAGCTTACAAGTGGC---GGTGGCAAGAACTTGGAGGGAAGA
AGAGCCTACTTAGCATATGTTTCGGAAGGAATTGGAAAATTACAGGATTG
GGAAATGGCCATGAAAATATCAAAGAAAGAAATGGATCACTGTTCAATTAC
CATCCACCACAGCAGCTGCCTTTATTCATATTCAAGATGCTGAGTGTCTC
CATTATATTCTGTTCACTCTTACAGAAGTTTGGGAATGCAAGTTCCAACCAT
TTATCCTTTGGATATATATGCTCGTCTTCTATGGTTGATGCTCTTGAAA
GGTTGGGAATCGATCGGCATTTTAGGAAGGAAAGAAAATTTGTTTTGGAC
GAAACATACCGATTTTGGTTGCAGGGGGAGGAAGAGATATTTTCTGATAA
TGCCACTTGTGCTTTGGCATTAGGATATTACGTTTGAACGGATATGATG
TCTCTCAAGAT-----CATTTCTCTAAT
TCACTGGGAGGATATTTGAAGGATTGCGGAGCTGCCTTAGAGTTGTACAG
AGCTCTGCAGCTA---AGTTATCCAGATGAATCACTTCTGGAAAAACAAA
ATTCTCGGACAAGCTATTTCTGAAACAGGGATTATCCAACGTTTCACTT
TGTGGAGATAGGCTTCGTAAAAATATTATCGGAGAGGTGCATGATGCTCT
CAATTTTTCTGACCATGCAAAATTTGCAACGCTTAGCTATCAGAAGAAGAA
TTAAACAT---TATGCTACAGATGATACGAGGATTTTGAACCTTCGTAT
CGTTGTTGCACTATTGGTAACCAAGGATTTTCTCAAATGGCTGTAGAAGA
CTTCAATATCTGTCAATCAATACAGCGTGAAGAATTTAAACATATCGAGA
GGTGGGTTGTAGAGAGGAGACTGGACAAGCTAAAGTTTGTAGGCAGAAG
GAGGCCTACTGTTACTTCTCTGCTGCAGCAACTCTCTTCGCTCCAGAACT
ATCTGATGCACGCATGTCATGGGCAAAAAATGGTGTGCTTACTACTGTTG
TTGATGACTTCTTTGATGTTGGTGGTTCTGAAGAAGAACTGGTAAACCTT
ATTGAATTGATTGAGAGGTGGGATGTCAATGGCAGTGCTGATTTTTGTTC
TGAGGAAGTTGAGATCATATATTCGGCAATTCACAGCACTATAAGTGAGA
TAGGAGACAAATCTTTCGGATGGCAAGGACGCGATGTGAAAAGTCAGGTT
ATCAAGATTTGGTTGGATTTGCTCAAAATCCATGTTGACAGAAGCACAAATG
GTCAAGTAACAAATCAGTGCCGACCCTTGAA---TATATGACAACATGCAC
ATGTATCGTTGCTCTAGGGCCTATTGTTCTTCCAGCTCTGTATTTTGTG
GGGCCTAAGCTTTTCAAGAGGAGGTTGCTGGACATCCTGAATTGCTTAATCT
ATACAAGGTTACGAGCACTTGCAGGCGTCTGCTCAATGACTGGAGAAGCT
TTAAGAGAGAATCTGAAGAAGGGAATTTGAATGCCATCTCATTGTACATG
ATTACAGCGGTTGGTGTCTCAACTGAAGAAGAGACCATCGAACATTTTAA
AGGATTGATCGACAGCCAGAGAAGACAATTGCTTCAATTAGTTTTCAGG
AAAAGGATAGTATAATTCCTAGACCCTGCAAGGATTTGTTTTGGAACATG
ATAAAATTTATGCACAGTCTACATGAAGGATGATGGATTCACTTCAAA
CGAGATGAGAAATGTT---GTCAAGGCAATAATAAATGAACCCATCTCTC
TAGATGAATTA-----

>R

-----TTTGATAAGATTGAACTT
TCGGTTTCTCCGTATGACACTGCTTGGGTAGCAATGATCCCCTCTTTG--
-AATTCCTGACAAAGCTCCTTTCTTTCTGAGTGCACAAAATGGATAGTTG
ATAATCAACTTAGTGATGGTCTTTGGGGTCTTCT--CATCACCATCCA
TTGTTAATAAAGGATACCCTTTTCATCCACATTAGCATGCGTCTTGCAATT
GAAGAAATGGGGAGTTGGTGAACCCCTGGTTAACAAGGTCTGCAGTTTA
TTGAATTAATCAACTTCATTAATGATGAGAAACAACATACGCCAATT
GGATTTGATATAATTTCTGCTATGCTTGAGCATGCTAAAGAGTTGGC
TTTGAACCTGCCTTTGAAATCAGATGTTATAGATGCAATGCTCCACAGGA
GAGATGTGGATCTTAAGAGTGGCTCTGGTGGCAGCAATACAGAGGGAAGA
AAAGCCTACTTAGCTTACATTGCAGAAGGGATCGGAAAATTCAGGATTG
GGAAATGGTCATGAAGTATCAACGAAAGAATGGTTCCTGTTCAATTCAC
CATCTACCCTGCAGCTGCTTTTAGTCATCTTCGTAATGCTGATTGTCTT
CAGTATCTGCAGTCTGTTTTACAGAAATATGGAATGCAGTTCCAATAT
TTATCCACTAGACGTATATTCTCGTCTTCTTATGGTTGACATTCTTGAAA
GGTTGGGAATTGATCGGCATTTTCAGGAAGGAATAAAATTAGTACTGGAA
GAAACATACAGATATTGGTTGCAGGGGAATGAAGAAATTTCTAGATTG
TATTACCTGCGCCATGGCATTTCGAATATTACGTGTCAATGGATATGATG
TCTCTTCAGATGTATTTACCAATTTACAGAGGAC--CATTTCTTCGAT
TCGCTTGGAGGATATTTGAAGGACACGAGGACTGTGTTAGAGTTATACAG
AGCTTCACAAATA--CTTTATCCGGATGAACCACTTTTGGAGAAACAAA
ATTCATGGCAAAATCATTTCCTGGAGAAGTGCTATCCAGTGGTTCCAGT
TATGCTGACGGACCCAGGGAATGTATTACTGAAGTGGTGCATAATGCTCT
CAATTGTCCTACTATGCAGATTTGGAACGTTTAACTAACAGGAGGAGCA
TTGAGAAT--TATAATGTAGATGAAACAAGGATTTTAAAGCCTCATAT
CGTTGTTTGAACACTGGGAACCAACATTTTCTTAAACTGGCGGTGGAAGA
CTTCAACCTCTGCCAATTGATACACCAGGAAGAACTTCAACAACCTGGGA
GGTGGGTGTGGAGAAAAGATTGAACAAGCTAAAGTTTGGCAGGCAGAAAG
CTGGGTATTGTTACTTCTCTGCTGCAGCTACTCTTTGCTCCTGAATT
ATCTGATGCTCGCTTATCATGGGCGAAAAACGGTGTGCTTACTACTGTTG
TTGACGACTTCTTGATGTTGGCGGTTCTGTAGAGGAACTGATAACCTT
ATCCAGTTGATTGAGAAGTGGGATGTAGACGAGAGCACTCACTTCTGTTT
TGAGCAAGTTGAGATCATATTTTCAGCACTTCGCAGCACCATTTCTGAGA
TTGGAGACAAAGCATTTCACATGGCAAGGACGTAAGTCACAAGTCATGTC
ATCAAAATTTGGTTGGATTTGCTGAAGTCAATGTTGACGGAACTCTCTG
GACGAAAAGCAAGTCAATACCTACACTTGACGAGTATATGATTAATGGCT
ACGTGTCATTTGCTTTAGGACCAATTGCTCTCCAGCTCTTTTCTTGTT
GGACCTAAACTTACAGAGGAGGATGTTAGGGACCTGAATTGCATGATCT
GTTTAAAGCTATGGGTACCTGTGGCGTCTCCTCAATGATTGGAGAGGAT
TTCAGAGAGAATCCAAAGAAAGGAAATGAATGCTGTCTCACTGCACATG
ATTCAAGGAAACGGTGGTGAATGAAGAAGAGGCCATTAGAAAGATAAA
GGGTTTAAATTAAACAGTCAGAGGAGTGAAGTCTGAGACTAGTTTTGCGGG
AAAAGAACAGTAATATCCAAGAGCATGCAAGGATTTGTTTTGGAAATG
ATCAAAGTGTGACCTTTTTACTTGAAGGATGATGGATTCACTTCAAA
TGAAATCAGTACTGCA-----AATGCAGTGATTACTGAACCTGTCGCAT
TCCATGGACCTTTTCTGTGCTGGCTTTTATCTGCCACTATTTATGTGCAT
TCTATGTAGAAAATATCAATCCACTATTTGATGGTTTGTATATACGGT
GTATGATGCTTCTGGTTTCACTAATGCAGGGAAACCCATGGTAAGGG
AAACAATGAATGCTGCAACATTGCATGTCTTCATGTCACTAGCAGCGGTG
AAAGATACAACGAGAAAGAT
>0

----TTTCAGTTAGAATTAGTGAACGTCGTCATGCACCAGAGGAAAGCAA
TAGAGGATACAAGGAAGAAGAAGAAGCAGCAGCTTCACAAATTCGAGATG
TTGCCGTCTCCCTACGACACGGCATGGGTGGCCATGGTGCCTCTGCCCGG
ATCATCATCTCAGCTTCCATGCTTCCACAATGCGTTGAGTGGATACTAC
AGAACCAGCAGAGCAATGGTCTTTGGGACCTC-----AACCAGCTGGAC
TCAATCACCAAGGATGCTCTTGTGCCACTGGCTTGTGTTCTTGCACT
CAGGAGA-----GGACTGCTTTTTTA
TTGGAAGGAATTTCTCCATTGCTATGGATGAGCAGCTGGCTGCTCCTATT
GGATTCAATATAACTTTTCTGGTATGCTTAGCTCTGTAATTGAGATGGG
TTTGGAGGTTCTATTGGACAAACCGACGTCGAAAGAGTACTTCACTCC
AGGAAACGGAATTGAAAGAGAA--TATGAGGAGAACTATCGTGGGAGA
AACACATATATGGCATATGTGTCAGAAGGATTAGGGAATGCCAAGACTG
GAATGAAGTTATGAATTTCCAGAGGAAGAAGGATCATTGTTCAACTCCC
TTTCCATAACAGCTGCGGTGTTAGTCCATAATTATGATGCTAAAGCTCAC
CGATACTTAAATTTGCTTCTGAACAAATTTGGCACTGCAGTG-----
-TATACGAAAAATATACATCGCCAACCTTCCATGCTTGATGCGCTTGAAA
ATATGGGAATATCTCGGCATTTTATGATGGTGAAGATCAAGAGTACTGGAC
ATGACATACAGTTGTTGGTTACAGAGAGATGAGGAAGTTATGCTGGACAT
CACAACTGTGCAATGGCATTTCGTATTCTTCGGATGAATGGATATGATG
TTTCTTCTGATGACCTGTGTCATATCGCTGAAGTTTCTGATTTCCATAGT
TCGCATCAAGGATATTTGAGTGACACAAGAACATTACTAGAATAACAA
GGCTTCAGAAGTCAGTGTGAGATAATGAGTTTATTCTAGATAGAATAG
GCTCTTGGTCAGGTGCGTTACTGAAGGAACAGCTTAGCTCTGGTGGCTTG

TGACCTCCGTGCTGTG-----AAATCAGTCATTTATGAACCTGTTCTTC
TTGCTGAATTC-----

GA.5.5.1.13.fasta

>A
TCTCTTCAGTATCATGTTCTAAACTCCATTCCAAGTACAACCTTTCTCAG
TTCTACTAAAACA-----ACAATATCTT
CTTCTTTCTTACCATCTCAGGATCTCCTCTCAAT-----
-----GTCGCTAGAGACAAATCCAGAAGCGGT-----
-----TCCATACATTGTTCAAAGCTT-----
-----CGAACTCAAGAATACATTAATTCTCAAGAG-----GTTCAA
CATGATTTGCCTCTAATACATGAGTGG-----CA
ACAGCTTCAAGGAGAAGATGCTCCTCAGATTAGTGTGGAAGTAATAGTA
ATGCATTCAAAGAAGCAGTGAAGAGTGTGAAAACGATCTTGAGAAACCTA
ACGGACGGGGAAAATTACGATATCGGCTTACGATACAGCTTGGGTTGCATT
GATTCGAT-----GCCGGAGATAAAACTCCGGCGTTTCCCTCCGCCGTGA
AATGGATCGCCGAGAACCAACTTTCCGATGGTTCTTGGGGAGATGCGTAT
CTCTTCTCTTATCATGATCGTCTCATCAATACCTTGCATGCGTCGTTGC
TCTAAGATCATGGAATCTCTTCTCATCAATGCAACAAAGGAATCACGT
TTTTCCGGGAAAAATATTGGGAAGCTAGAAGACGAAAAATGATGAGCATATG
CCAATCGGATTCGAAGTAGCATTCCCATCGTTGCTTGAGATAGCTCGAGG
AATAAACATTGATGTACCGTACGATTCTCCGGCTTAAAGATATATACG
CCAAGAAAGAGCTAAAGCTTACAAGGATACAAAAGAGATAATGCACAAG
ATACCAACAACATTGTTGCATAGTTTGGAGGGGATGCGTGATTAGATTG
GGAAAAGCTCTTGAAACTTCAATCTCAAGACGGATCTTCTCTTCTCTC
CTTCCTCTACCGCTTTTGCATTTCATGCAGACCCGAGACGTAAGTGCCTC
GAGTATTTGCGAAATGCCGTCAAACGTTTCAATGGAGGAGTTCCCAATGT
CTTCCCGTGGATCTTTTCGAGCACATATGGATAGTGGATCGGTTACAAC
GTTTAGGGATATCGAGATACTTTGAAGAAGAGATTAAGAGTGTCTTGAC
TATGTCCACAGATATTGGACCGACAATGGCATATGTTGGGCTAGA---TG
TTCCCTCCAAGACATCGATGATACAGCCATGGCATTTAGGCTCTTAAGAC
AACATGGATACCAAGTGTCCGCAGATGTATTCAAGAACTTTGAGAAAGAG
GGAGAGTTTTTCTGCTTTGTGGGGCAATCAAACCAAGCAGTAACCGGTAT
GTTCAACCTATACCGGGCATCACAAATTGGCGTTTCCAAGGGAAGAGATAT
TGAAAAACGCCAAAGAGTTTTCTTATAATTATCTGCTAGAAAAACGGGAG
AGAGAGGAGTTGATTGATAAGTGGATTATAATGAAAGACTTACCTGGCGA
GATTGGGTTTGCCTTAGAGATTCCATGGTACGCAAGCTTGCCTCGAGTAG
AGACGAGATTCTATATTGATCAATATGGTGGAGAAAACGACGTTTGGATT
GGCAAGACTCTTTATAGGATGCCATACGTGAACAATAATGGATATCTGGA
ATTAGCAAAACAAGATTACAACAATTGCCAAGCTCAGCATCAGCTCGAAT
GGGACATATTTCCAAAAGTGGTATGAAGAAAATAGGTTAAGTGAAGTGGGGT
GTGCGCAGAAGTGAGCTTCTCGAGTGTTACTACTTAGCGGCTGCAACTAT
ATTTGAATCAGAAAGGTCACATGAGAGAATGGTTTGGGCG---TCAAGTG
TATTGGTTAAAGCCATTCTTCTCTTTT-----GGGGAATCC
TCTGACTCCAGAAGAAGCTTCTCCGATCAGTTTCAATACATTGCCAATGC
TCGACGA-----AGTGATCATCACTTTAATGACA
GGAACATGAGATTGGACCGACCAGGATCGGTTTCAGGCCAGTCGGCTTGCC
GGAGTGTTAATCGGGACTTTGAATCAAATGTCTTTTGACCTTTTCATGTC
TCATGGC-----CGC---GTTAACAATCTCCTCTATCTATCGTGGGGAG
ATTGGATGGAAAAATGGAACTATATGGAGAT-----GAAGGAGAAGGA
GAG-----
-CTCATGGTGAAGATGATAATTCTAATGAAGAACAATGACCTAACTAACT
TCTTACC-----CACACTCACTTCGTTCTGCTCGCGGAAATCATCAAT
CGAATCTGTCTTCTCGC---CAATACTTAAAGGCAAGGAGAAAC---GA
TGAGAAG-----GAGAAG-----ACAATAAAGA
GTATGGAGAAGGAGATGGGGAAAATGGTTGAGTTAGCATTGTCGGAGAGT
GACACATTT---CGTGACGTGAGCATC-----ACGTTTCTTGATGTAGC
AAAAGCATTTTACTACTTTGCTTTATGTGGCGATCAT---CTCCAAACTC
ACATCTCCAAAGTCTTGTTTCAAAAAGTC---
>P

-----GTTTCTTTTTCATTAGGTGTTTGGTTGTAT-----
-----GGGGCTAAAGACAAAAAAGATAACATT-----
-----CATACAAGGTGCAGTGCTATATCCA
AGCCTCGCACTCAAGAATATGCAGATTTGTTTCAA-----CAACAA
AATGGTCTGCCATTGATA---AACTGGCCTCACGATATT-----GTGGA
GGATGACGCAGAAGAAGATGCTGCTAAGGTCTCTGTA-----GCAA
AGGAAATAAGCAAGCGCGTCAGAACTATTAAAGCAATGCTGGATATGATG

GAGGATGGAGAGATAAGCATCTCAGCTTATGACACCGCATGGGTTGCTCT
 TGTGGAAGATATTAATGGGAGTGGCCTTCCCAATTCCCATCTAGCCTTC
 AGTGGATTGCCAACAATCAGCTTCCAGATGGTTCGTGGGGTGACGCTGGA
 ATATTTTGGCGCATGATAGGTTGATTAATACGCTAGCTTGCCTGGTAGC
 ATTGAAATCTTGGAACTTGCATCAAGATAAATGCGAAAAAGGGATGTTAT
 TTTTATAGAGACAACTTATGCAAGCTTGAAGATGAAAAATGCTGAGCACATG
 CCTATTGGATTTGAAGTTGCTTTCCCTTCACCTCTTGGCGATAGCCAAAGA
 ATTAGACATCGAAGTTCGGGATGATTCTACTTTCCCTTCAAGAGATTTATG
 CCAGGAGAAATTTAAAGCTGAAAAGGATTTCCAAGGAC - - -ATGCACAAT
 GTGCCCACGACCCTACTCCATAGCTTGGAAAGGAATGCGAGGTCTGGACTG
 GAAACAGCTTATAAAATTGCAGTGCCTAGATGGGTCACTTATTCTCTC
 CATCCTCTACTGCCTTCGCACTCAGTCAAATAAAGATAAAACTGCTTG
 GAATATTTTGAATAAGGCGGTCCAAAGATTCAAAGGGGGAGTTCCCAATGT
 CTACCCGGTTGACTTATTTGAGCACATCTGGGTTGTGGATCGCTGCAGC
 GTCTTGGAGTCTCAAGATATTTGCTCTCAGAGATCAATGAATGTGTCGAT
 TATATTCATAGATATTGGACTGAAAAATGGCATCTGTTGGGCGAGAAATTC
 TAATGTCATGATATTGATGACACGCCCATGGGATTACAGGATCTTAGAT
 TACATGGGCACCAAGGTTTCTGCCGATGTGTTCAAGCATTTTGAGAAAGGC
 GGTGAATTTTTCTGCTTCGAGGGCAGTCAACAGGTGCTGTTACGGGAAT
 TTCAACCTGTATAGGGCTTCTCAGGTTCTGTTTCCGGGAGAGAAAGATAC
 TTGAGGATGCCAAGGAATATTCATTGCAATTCCTAAGGAAAAACAAGCT
 GCTAATGAACCTCTAGATAAATGGATTATAACGAAGGACTTGCCTGGAGA
 GGTGGGTTTTGCATTGGAGATTCCATGGTATGCGAGCTTGCCTCGAGTGG
 AGACTAGATTCTTCATAGAACAGTATGGTGGCGAGGACGACGTGTGGATT
 GGCAAGACCTTTACAGAATGTCATATATAACAACAGTGAGTATCTCCA
 GCTGGCAAAATTAGACTACAACAACAGTCCAAGCATTACATCGATTGAAT
 GGGAAAAATTTCAAAGTGGTATGAAGAATGTAATCTAAGAGATTTCCGGG
 ATAAGCAGAAGAACCTTATCTTTTCTTACTTTTTGGCAGCAGCCAGTAT
 CTTTGAGCCGAAAGGTCCAAAGAGCGACTTGCATGGGCTACCACTACAG
 TCTTGCTCGACATTGTCGGCTTATTTCCCTGAGAATCATACAATTCT
 AGCGAGCAGAGGAGAGCCTTCATACATGAATTCAGTTACGGTATTAGCAT
 AAATGGAAGGGAGGAAATGATTATTTATGCTGATTCTCGTGTGTTGCA
 GGAGGTCGGGT - - - - -AGAAAGAAGACAAGGCAGGAG - - - - -CTCGTT
 AAGCTGCTGCTTGGAAACCCTAAATCAGCTCTCTTTGGGCGCATTAGTGGT
 TCATGGT - - - - -CGAGACATTAGCCACAGTTTACGCCATGCTTGGGAAA
 AGTGGCTCTTGATCTGGGAATTGGAAGGAGATAGACGCCAAGGGGAAGCA
 GAA - - - - -
 -CTACTAGTGCAACAATAAATCTCACC GCGGCTATTTGGTCTCCGAGG
 AGCTCTTGGCCCATCATCTCAGTATGAACAACCTGTCGACCTGACAAAC
 AGAATATGCTATCAACTTGATCATTACAAGAAAAACAAGGTACATTACAA
 TGGAAAGTACAGCACTATTACCAGTAATACTGACAGAATTACAACCCAC
 AAATAGAAATGACAGATGCAAGAACCTTGTGCAGTTAGTAGTCCAAATCCC
 TCAGATGGA - - -ATAGATTCTAATATCAAGCAACGTTCTTCAAGTAGC
 CAAGAGTTTTTATTATTCTGCAATCTGTGATCCTGGGACTATCAACTATC
 ACATAGCTAAAGTGCTCTTTGAGAGAGTGCCT
 >R
 TCTTCTCAATCCCTTCACCTTCTCTACCGCCTCTTCTTTCTCC - - - -
 - - - - -ACTCCCAATT
 CTCGTCTCTACTCTCCTCTGGTGGTGGGCTTTTC - - - - -
 - - - - -GGGGTTAAAGACACAAGAATTACCTTTGAT - -
 - - - - -GTTCTGCAAGATGCAGTGCCTATCCA
 ATCCACCCACTCAGGAATACCCAGATGTCTTTCAA - - - - -AGA
 AATGGTGTGGCAGTGATA - - -AAGCGC - - -CGTGAGGTT - - - - -GTGGA
 AGATGCCATAGAAAGAGCAAGTTACTAAGATTTCCGTA - - - - -TCGA
 ATGAGATCCTTAAACGTGTCCAAACGTTAAATCGATGTTGGATTGATG
 GAAGATGGAGAGATAAGCATATCAGCTTATGATACAGCATGGGTTGCTCT
 TGTGGAAGATATTAATGGAAGTGGTGTCTCAATCCCGTCCAGTTTGC
 AATGGATTGCTAATAATCAACTATCTGATGGTTCGTGGGCGATGGTGAT
 ATCTTTACTGCGCATGACAGGATTCTCAATACATTAGCTTGTGTTGTTGC
 CTTGAAATCATGGAACATACATCCTGATAAGTGTGAAAGAGGAATGAAGT
 ATTTTAAAGAAAATCTATGCAAGCTTGAAGATGAGAATGCTGAGCACATG
 CCCATTGGTTTTGAAGTTGCTTTCCCTTCACCTCTTGAGTTAGCAAAAAA
 TTTAGACATTGAAGTTCCCGAGGATTCTCCAGTCTTGAAGAGATCTATG
 CCAGCAGAAATATAAGCTAACAAAGATACCAAAAGACATAATGCACAAA
 GTGCCCACCACACTACTCCATAGCTTGGAAAGGAATGCCAGGTCTGGAGTG
 GGAAGGCTTCTGAAATTGCAGTGCCAGATGGGTCACTTCTGTTCTCTC
 CATCCTCCACTGCTTTTGCACATGCAAACTAAAAATGAAATTTGCTTA
 GCATATCTAAACAAGATCGTTCAAAGATTTAATGGAGGAGTACCAACGT
 CTACCCTGTTGACTTATTCGAACATATCTGGGCTGTTGATCGCGTTCAAC
 GCCTTGAATATCGAGATATTTTCAAGAGGAGCTTAAAGAATGTATTGAC
 TATGTTGCCAGATACTGGGAAGAAGATGGTATTTGCTGGGCAAGAAATTC
 TGCAGTTATGATATTGATGACACAGCCATGGGATTCAGGTTGCTTAGAC
 TATATGGCCACGAAGTTTCTTCTGATGTGTTCAAGCATTTTAAAGAAAGGC
 GATACTTTCTTCTGCTTCGCCGACAGTCAACGCAAGGCTGTTACCGGAAT
 GTACAACCTGTATAGAGCTTCTCAGGTATTGTTTCTGGAGAAAAAGTCC

TCGAGGAAGCCAAGGAATACTCGTCCAGTTTCCTAAAAGAAAAGCAAGAG
GCTAACGAAGTACTTGACAAATGGATTATAACGAAAGACTTGCCGGGAGA
GGTTAAGTATGCATTGGATATTCATGGTACGCGAGCTTGCCTCGGGTGG
AGTCGAGATTCTACTTGGAACAGTACGGTGGTGAGGATGATGTCTGGATC
GGCAAGACTCTTTATAGGATGCCGTATGTGAACAACAATGAGTATTTGGA
CCTTGCAAAGCTTGACTACAACAATTGCCAAGCGTTACATCGCAAGGAGT
GGGACAACCTTTCAAAGTGGTACGAAGAATGTGAATTGGGAACTTTGGG
GTAAGCAGAAGAGAAGTCTGTAGCATATTTGTAGCGGCGGCCAGCAT
TTTTGAGCCAGAAAGGTCAAAGAGAGACTTGCATGGGCTAAGACCACAA
CCTTGCTCCATACAATCGAGTCTTATTTT--GATGCCAGCAACAGTACC
TATGAGCAAAGGACGGCTTTTGTCCACGAATTCAAAAACGGTGTGCA--
-----TCCATACCCCACTTAAATGCAA
GGAAGTTG-----GAAGTAAAGACAAATGAGGAA-----CTTGTC
AGGATAGCGATTGGAATCCTAAACGATGTCTCTTTGGATACACTACTGGC
TCATGGG-----AAAGACATTAGCCACGATTTACGTATGCTTGGGAAA
AGTGGCTGTTGAAGTGGGCGAAGGAGGAGAAATACACCAAGGAACAGGG
GAA-----
-TTGCTAGTGAAAAACAATAACTCTCACCGCTGGAGGTTCCACACCTGAT-
-----CATCATAAATATGCACAGCTCTTTCAACTGACAGAT
AAATTATGCTATCAGCTGGCTCATTACAGAAAAAACAGGTACAAGGCAA
TAAGAAAAGC-----ACAACCCAG
AAATAGAGTCAGACATGCAACAACCTTGTGCAGTTAGCAATCCAAAATTCC
TCAGATGAA--ATAGATTCCGAGATCAAGCAAACATTTTTCATGGTAGC
CAAAAGCTTTTACTACCAAGCCATCTCTGATCTGGGACTTTGAACTACC
ATATTGCTAGAGTACTCTTTGAGAGAGTATAC
>0
---ATTCACCTCCACTCCCCGCCGACGGCGCCCGCCGCATTGCGCGGCGC
CGGCTCGGCGGACTGGCGGCGGCGGCGGCGGTGGTCTATGGTCGTCTGTCT
CCCGCGCTCCGGTGGCTAAAGGTGGCCATCTTCGTCCGTGCGTTTGGCGG
CGCGGCGGCGGACGACGGCGGCGGCGAGGATCATCACGCCGACGGCGGCGG
CGGCGGCGGAGGAGGAGCGGCGTGGAGGGCGCGGGCCACCACCGCGGCG
TGTGAGCTCCAGCAGTACAGCCAAAGGGCTGCAAGCCAACATCATCGAA
CATGAGACCCCCGGATCACGAAATGGCCCAATGAATCACGCGACCTCGA
CGATCACCAACAAAAACACGAGGCTGATGAGGAGGCA-----GATG
ATGAGCTGCAGCCACTGGTCGAGCAGGTGAGGTGATGCTGTCTGTCCATG
GAGGACGGCGCGATCACGCGCTCGGCGTACGACACGGCGTGGGTGGCGCT
GGTGGCGGCGGTGGACGGCGAGGGCGGCGACGCACTTCCCGGCCGCGTGC
GGTGGATCGTTCGGCAGCCAGCTCGCCGACGGGTGCTGGGGCGACGAGGCG
CTCTTCTCCGCTACGACCGCGTCTCAACACCCTCGCTGCGTGTCTGCGC
CCTCACCAAGATGGTCCCTCCACCATGACCAAGTCAAGCAAGGGCTTCAAGT
TTCTGAATCTGAATTTGGAGGTTAGCAGAGGAGGACCGGATACGATG
CCGATTGGGTTTGAGATTGCATTCCCTTCTCTTGTGGAGGCACTAGGGG
TTTGGGTATTGATTTCCCATATGATCACCTGCTCTCAAGGGCATTTATG
CAAACAGAGAATCAAGCTTAAGAGGATTCCAAAGGACATGATGCATATA
GTCCCAACTTCAATTCTGCATAGCCTTGAAGGGATGCCTGGGCTGGATTG
GCAGAGGCTTCTGAAGCTCCAATGCAGTGATGGATCCTTCTTGTCTCCC
CTTCAGCTACTGCTTATGCTCTCATGCAGACCGGTGACAAGAAATGCTTC
GCGTACATCGACAGGATCATTAAAGAAATTCGACGGAGGCGTTCGGAACGT
TTACCCGGTCGATCTTTTTGAGCACATATGGGTTGTCGATCGGTTGGAGC
GTCTTGGGATCTCGCGGTACTTCCAACGAGAGATTGAACAGAACATGGAC
TATGTCAACAGGCACTGGACTGAAGATGGGATTTGCTGGGCTAGGAACCTC
CAATGTAAAAGAAAGTGGATGACACCGCTATGGCTTTCGTCTACTACGCC
TCCATGGATACAATGTATACCAAGTGTGTTCAAGAAATTTGAGAAGGAT
GGGGAGTTCTTCTGTTTTGTGGGCAATCAACTCAAGCAGTCACTGGGAT
GTATAACCTGAACAGAGCATCTCAGATAAGTTTTCCAGGAGAAGACATTT
TGCAGCGTGCAAGGAATTTCTCATATGAGTTCCTTAGAGAAAGAGAAGCC
CAGGGGACACTTTCATGATAAATGGATCATCTCCAAGGACCTACCAGGAGA
GGTACAATACACACTAGATTTTCTTGGTATGCGAGCTTGCCACGCGTCG
AGGCAAGAACATACATAGGTCAATATGGTGGAATGATGACGTCTGGATT
GGAAAGACACTCTACAGGATGCCAATTGTGAATAACGCTACATATCTCGA
GTTGGCGAAACAGGATTTCAACCGTTGTCAAGCTCTACATCAGCATGAGT
TGCAGGGTCTACAAAAGTGGTTCATTGAGAATGGCTGGAAGCTTTTGGG
ATGACACCTGAAGATGTTTTGAGAGCTTATTTTTGGCTGCCGCGTGCAT
TTTGAACCAAACCGTGCCTCTGAGCGACTTGCATGGGCTAGAGTGTGAG
TGCTGGCCAAACACTATTCTAGGCATTTT-----TACAGCGATATG
TCAAGCATGAAAAGG-----ATGGAGCGTTTCATGTGGAGC-----
-----AGCCTCTATGAAGAAAATGGCA
ATGTTTTGGGG-----CTAGAAGGATATGCAAAGATGGAATCCTTGCG
AGGACACTTTGTCAACTTATAGATTTGTTGTCTCAAGAGACACCGCCAGT
TCGAGAAGGTCAAAGTGTATTCTAATCTCATAAGATGTGCTTGGATTG
AATGGATGATGCAACAAATCAATATGAAGGATGGCAGATATGACAAAGGC
AGAGTTATGCATCCAGGTCATGCACTGTTCTAATAAAGAAACATGTTT
ACTTATTGCTCAAATTGTTGAAATTTGTGCTGGACGAATTGAGGAGGCAG
CATCTATGATAAATAACACCGAAGGTTCTTGGTTTATTCAACTTGCTTCC
TCTATTTGCGATTCTCTT---CATGCCAAGATGTTACTTTACAG---GA

TACCAAGAAAAATGAGACAACAATAAATCAA-----ATTGACAAGG
AAATTGAGTTGGGTATGCAAGAACTTGCTCAATATCTTCTTCCAAGAGTT
GATGATAGAAGAATTAACAACAAAACCAAGCAGACCTTCTTGAGCATTGT
GAAAAGCTGTTACTATGCTGCCAATTGCTCACCACATATGCTTGACCAAC
ACATTTCTGAAGTGATTTTGGCAAGTTATT
>V
TCTTCTCAATCCACTATTCTCTTCCGCGCTCTCTCTCCAAAGGGTCCCCT
ACCTTCT-----CCTTCTCTT
CGACACCCAAATATGCTGCTGGTACTGGATGGTGG-----
-----GTGGCTAAAGACAAAAGGGCTCACTTCAAT--
-----TTAAGGAGAGAGTGCAAGGCCTTACCAA
AACCCCAAACCTCAAGAATACGTGGACGTATTTCAA-----
AGTGGTCTGCCAGTGATA---AAGTGG---CATGAGATT-----GTGGA
AGATGATATAGAAGAGGAAGCTTTCCAGAAGGTTTGT-----AGGTCAA
ATGAGATTAAGGAACGTGTGGATTCCATAAAATCGATGCTGAGTTCAATG
GAAGATGGAGAGATCAGCATATCGGCTTATGACACAGCCTGGGTGGCTCT
CGCGCAGGACCTCAATGGAGGAGGACTCCGCAATTCCCATCAAGCCTAG
AGTGGATCGCCAACAATCAGCTCATGGACGGGTCATGGGGAGACGACTGC
TTATTCTATGCTCACGACCGGATAATCAACACATTGGCGTGTGTGATCGC
ATTGAAATCGTGGAAAATTCATCCGAACAAGTGCGAGAAAGGGGTGTCAT
TCATCAAAGAAAACATAAGCAAGCTTGAAGGAGAAGGCTGAGCACATG
CCTATAGGATTCGAAAGTAGCTTTCCCTTCTCTCTCGAGATAGCTTGGAA
CTTGGGCATCGAAGTACCAATGATTCGCCCTTCTTGAAGGAGATCTACG
CCCAGAGAAATCTGAAGCTCACAAGATACCATATGACATAATGCACATA
GTGCCCACAACACTACTCCACAGCTTGAAGGAATGGCAGGTCTAGACTG
GGAAAAGCTTTCTAAACTACAGTGTCAAGATGGGTCATTTTGTTTTCTC
CATCGTCCACTGCCTATGCCCTCATGCAGACCAAGATGAGAATTGCTTT
AGATACCTAAACAAAGTGGTTGAGAGATTCAAGTGGAGGAGTCCCAATGT
TTACCCTGTGCACTTGTGAGCACATATGGCTGTGGATCGGCTGCAGC
GCCTTGGGATATCGCGGTATTTGAGCCTGAGATTAAGAATGCATCAAT
TATGTTGCCAGATATTGGAAGAAAATGGAATCTGCTGGGTAAAACTC
CGAAGTTCACGATATTGATGACACTGCTATGGGGTTTAGGCTCCTTAGGT
TACATGGCCATGATGTTTCTGCTGATGTGTTCAAGTATTTGAGAAAGGT
GGTGAGTTCTTCTGCTTGTGGGCAGTCATCCAGGCTGTGTCCGGGAT
GTTCAACCTCTACAGAGCTTCCAGGTGCTGTTCCCTGGAGAAACCATT
TTGAGAATGCCAAGAAGTTCTCTTCCAAGTTTCTAAGGGAACCAAGCT
TGTGACCAGCTCCTGGATAAGTGGATCATAATGAAGGATTTGCCTGGTGA
GATAGGGATATGCGCTGGATGTTCCATGGTATGCAAGCTTGCCTCGTGTGG
AGACCAGAATTTATATTGAACAGTACGGTGGAAAAGATGATGTATGGATT
GGCAAGACCCCTTACAGGATGCCTTATGTAAACAATAATGATTATCTTGA
GCTTGCAAACTGGACTTCAATAACTGCCAAGCCCTGCATCAGCTTGAAT
GGGACAGCATTCAACAGTGGTACACGCAATGCCATCTTGGGGAGTTTGGG
GTTAGCAGAAAAGCCTCTCTCTTGGCTTATTTCTAGCAGCAGCCTGCAT
ATTTGAGCTGAAAGGTCAATTGAGAGGCTTGCTTGGGCGAAAACCGCCA
TCTTGGTCTTGGCAGTCCGGTCGTATTTT-----AGCAAGGAAACATTT
ATTAAGCAGAGGAGAGCCTTCTTCTTAAATTTGGATACAGCCCAAGTGG
AGGAGAC-----TACATGAAGAAGATCAATGGAA
GGCCTGTGGGATTCCTTAAGAACAAGAGCGGAGAGGAA-----CTTGTC
GGACTCTTACTCGGAACCTTAAATCAACTCTCATTGGATACACTGGTGGC
TCATGGC-----AGAGACATACGACACATCTTACGTCAAACCTTGGGAG-
--TGGCTGATGAAACATTTGGATGAAGGAGATGGGTACCGAGGAGAGGCA
GAG-----
-CTCCTGGTGCGGACCATAAATCTCTGTGCTGGTCCGTCTCTCAGAGG
AGTTGTTGGCT---CATCCTCAGTACAGGCGTCTCTCAACTCACTAAC
AGAATCTGCAACGATCTTGGTCTCTTTACATTGCACAAGGGTCAC---AA
TGGCACCTACAATCCAGAAAATGGCAGCCCT-----ATGAGCCATA
GGGTAGAGTCAGACATGCAAGAACTTACGAAGATAGTCTGGAAAACCTAC
CCTGATGGC---ATCAACCCAGAAAATCAAGAGGACTTTTGAGATGGTGGG
CAAGAGTTTCTACTATGCTGCCTACTGTGGACCTGCAACTATCAGTGCTC
ATATTGCTAAAGTCCTCTTTGAGAGAGCAGCC

GLY.2.7.2.3.fasta

>A
---GCTTCCGCTGCCGCAAGTTCCGCCTTTTCACTCCTTAAGTCCACCGG
CGCTGTTGCTTCCCTCCGCCGGAACCTCGCGCACGTGCCTCCCTTCTGCCAA
TTCCCTCTACCTCTGTT-----TCCGCGCGTCTCTAGGCTTCTCC
GCCACTCTAGATTCCCGTCGTTTCTCTCTCCACGTTGCTTCCAAAGTTGA
A-----TCGGTGCCTGGGAAGGGAAGCAGAGGAGTGGTTTCTATGG
CGAAGAAGAGCGTCGGAGATCTGACCTCAGCTGATTTGAAGGGGAAGAAG
GTTTTCTGAGAGCTGATCTCAATGTACCTCTCGAT---AATCAGACTAT
CACTGACGATACCAGAATCCGTGCCGCCATTCCAACGATCAAGTATTTGA
TTGAAAATGGTGCTAAAGTTATCCTCTCCACTCATCTGGGAAGGCCAAAG
GGAGTCACCCCAAAGTTTAGTTTGGCTCCTCTTGTCCCTAGGCTCTCCGA

GCTTCTTGGTATTGAGGTCACGAAAGCTGATGATTGTATTGGCCAGAAG
TGGAAAGCTTGGTGGCTTCTCTACCA - - - GGTGGAGTTTTGCTTCTTGAG
AACGTCAGGTTTTACAAGGAGGAAGAGAAGAACGATCCTGAGTTTGCTAA
GAAGCTTGCTTCTCTAGCTGACCTTTATGTCAATGATGCTTTTCGGAAGCTG
CTCAGAGAGCTCATGCTTCTACCGAAGGAGTCACTAAGTTCTTGAAGCCT
TCAGTTGCTGGTTTCTTTTGCAAAAGGAAGTGGACTACCTAGTTGGTGC
TGTTTCAAACCCAAAGAGACCATTTCAGCCATAGTGGGTGGTTCCAAAG
TCTCATCCAAGATTGGAGTTATTGAATCGCTTCTGGAGAAAGTGTGATATT
CTTCTTCTTGGTGGTGAATGATCTTCACATTCTACAAGGCACAGGGTCT
TTCAGTTGGTTCGTCCTTGTGAAGAAGACAAGCTTGAATTGGCTACAG
AACTCCTTGCCAAAGCTAAGGCCAAAGGAGTCTCTCTTTTGTGCAACA
GATGTTGTAGTTGCTGACAAGTTTGTCTCTGATGCCAACAGCAAGATTGT
GCCTGCATCAGGCATTGAGGACGGATGGATGGGACTGGACATTGGTCCAG
ACTCTATCAAACTTTCAACGAAGCTCTGGACACAACACAAACAGTCATT
TGGAAATGGACCTATGGGAGTTTTCGAGATGGAAAAGTTTGGGCTGGAAC
AGAGGCGATAGCGAATAAACTAGCAGAGCTAAGTGAAAAAGGAGTGACAA
CGATAATAGGAGGAGGAGACTCAGTGGCTGCAGTGGAGAAAGTAGGAGTA
GCAGGAGTCATGAGTCACATCTCCACTGGTGGTGGAGCCAGCTTGGAGCT
GTTGGAAGGAAAAGTACTTCCCGGTGTGATCGCCCTTGATGAAGCAATCC
CAGTCACTGTT

>P

ATGGCCTCCGCCACAGCACCAACAACCTCTCTCTCCTTAAACA - - - -
- - - - GCCGCTCTCCACCTCCACATCTGTCCGGGCTCCCTACTCCCAG
TCTCCACCTCCGGCTCCGCACAACAGCCTCCGGGGTCTTGGTTTCTCT
GCTGCA - - - GACACA - - - TTGTTTAGCTCTCATGTGGTGCCAAAATACG
A - - - TCTTTCAAGTCAAATGGTAAAGCACAAGAGCAGTGGTGTCTATGG
CAAAGAAGAGTGTGGTGACTTGACAGCTGCTGATTTGAAAGGGAAGAAA
GTCTTCGTAGAGCTGATTTGAATGTGCCTCTTGATGATAATCAGAAATAT
TACTGATGATACTAGAATTAGAGCCGCTATCCCTACTATTAATACTTGA
TTTCAAATGGGGCTAAAGTTATTCTTTCTAGTCATCTGGGACGACCAAAA
GGTGTACTCCAAAATTCAGCTTGGCACCTCTTGTACCCCGGTTATCTGA
ACTCCTTGGAATTCAGGTTGTGAAAGCTGATGATTGCATTGGCCAGAGG
TAGAAAAATTGGTTGCTTCACTTCCTGATGGTGGTGTCTTCTTCTTGAG
AATGTGAGGTTTTTACAAGGAAGAAGAGAAGAATGAACCTGAATTTGCAAA
GAAGCTTGCTTCCCTTGGCAGATCTTTATGTAATGATGCATTTGGAAGT
CACATAGAGCACATGCCCTCACTGAGGGGGTCACAAAATCTTGAAACCA
TCTGTTTCTGGTTTCTCTTGCAGAAGGAGCTGGACTACCTTGTGGGGC
AGTTTCAACCCCAAGAGGCCATTTGCTGCCATTGTGGGTGGTTCCAAAG
TCTCATCCAAGATTGGAGTCATTGAGTCACTCCTGGAGAAGTGTGATATC
CTTCTTTTGGGTGGAGGTATGATTTTACATTTTACAAGGCTCAAGGCCT
CCCAGTAGGTTTCATCCTTGGTTGAGGAAGATAAGCTTGGTTTGGCAACAT
CGCTCCTTGAGAAGGCCAAGGCCAAAAGGAGTGTCTCTTTTGTACCCAGT
GATGTGATAATTGCAGACAAGTTTGTCTCTGATGCAATAGCAAGATTGT
ACCTGCATCAGCCATCCAGATGGTTGGATGGGACTGGACATTGGACCAG
AATCTGTGAAGACATTCAGTGAAGCTTTGGGAACCAACCAACTGTCATC
TGGAAACGACCCATGGGAGTGTGTTGAGTTTGACAAGTTTGCAGTTGGGAC
AGAGGCCATTGCAAGAAGCTTGTGAGCTTAGCGGCAAGGGAGTTACAA
CTATTATTGGAGGTGGAGATTCTGTTGCGGCTGTAGAAAAAGTAGGAGTT
GCTGATGTTATGAGCCACATATCAACTGGCGGTGGTGCCAGCTTGGAGTT
GTTGGAAGGGAAGAGCTTCCAGGTGTCTTGCACTAGATGAAGTCGAAC
GAGTTGCTGTG

>R

- - - GCCTCTGCTACAGCACCCACTACCCTTTCTTCTCAAGACT - - - -
-ACTGCTTCTTCTTCTTCCACCACCCGGGCGCGCTTCCCTTCTCCCTG
TTTCTGGCCTCCGCCCC - - - - -ACCACTTTCGCGCTCTCGGCTTCGCT
GCTGCT - - - GACCCT - - - GCTTTTCTCACCACGTTGCATCAAAAATCCG
T - - - - -TCTTTTGGTTCTGGCAAGGCGTCCAGGGCTGTTGTCTCCATGG
CTAAGAAGAGCGTTGGTGATCTGACTGCTGCTGATTTAAAAGGGAAGAAG
GTTTTTGAAGGGCTGATTTGAATGTTCTTTGGATGATAACCAGAAAAAT
TACTGATGATACTAGAATTAGAGCTGCTATCCCTACTGTTAAGCATTTGA
TTCAGAAATGGGGCTAAAGTCATTCTTTCCAGTCATTTGGGAAGGCCAAAA
GGTGTCTACTCCAAAATTCAGCTTGGCACCACTTGTCCCGATTATCAGA
ACTCCTTGGCATTGAGGTTGTGAAGGCTGATGATTGATTGGTCCAGAAG
TAGAAAAATTGGTGGCTTCACTTCTGAAGGAGGTGTTCTTCTCCTGGAG
AATGTGAGGTTTTACAAGGAGGAGGAAAAGAATGAGCCTGAATTTGCGAA
GAAGCTGGCTTCCCTAGCAGATCTATGTGAATGATGCATTTGGAAGT
CACATAGAGCACATGCCCTCAACCGAGGGAGTTACAAAATCTTAAAGCCT
TCTGTTGCTGGCTTCTCTTGCAGAAGGAGCTGGACTACCTGGTAGGTGC
AGTTTCAAGCCCAAGAGGCCATTTGCTGCCATTGTTGGTGGTTCAAAGG
TCTCATCAAAAATTTGAGTTATTGAGTCACTCCTTGAGAAGTGTGATATT
CTGCTTTTGGGTGGAGGAATGATTTTACATTTTACAAGGCACAAGGCCT
GTCAGTGGGTTTCATCCCTGGTTGAGGAAGATAAGCTTGATTTGGCAACAT
CTCTCCTTGCAGAGGCCAAATGAAAGGAGTGTCTCTTTTGTACCCACT
GATGTGGTAATTGCAGACAAGTTTGTCCCGATGCAACAGTAAGGTTGT
GGCTGCCCTCAGCCATCCCTGATGGCTGGATGGGATTGGATGTCGGCCAG

AATCTGTTAAGACATTCAATGAAGCTTTGGAACCACTAAAACGTGCATC
TGGAAATGGACCAATGGGAGTGTTGAGTTTGACAAGTTTGCAAGTTGGAAC
AGAGGCTATTGCAAAGAAGCTAGCAGAGCTCAGTGGCAAGGGAGTCACCA
CAATTATTGGAGGTGGAGATTCTGTTGCAGCTGTAGAGAAAGTAGGAGTT
GCTGAGGTTATGAGCCACATATCAACTGGTGGTGGTGCAGCTTGGAATT
GTTGGAAGGCAAGAGCTTCCAGGTGTTCTTGCTCTTGATGAAGCTACAC
CAGTTGCTGTG

>O

---GCGTCCGCGGCGCGCCGACCACGTCTCGCTCTCCCTCGCC-----
-GCCAGGCGCGCGCGGCGGTGGGGCCCGCGCGCTCGCCGCTACGGC
GCGGGGGGCTGGCGGCCGGGTGCCAGCCGGCGCGGTCTGCTGGCGTTCCG
GCGGCGCGGACCCG---CGGCTGGCGACCCACGTGCGCTGCGGGTGCCG
GCAGGCGGCGTCTGCTCTCGCGGGGACGCGCGCGCTCGCCACCATGG
CGAAGAAGAGCGTGGGGACCTCACCGCGCCGACCTCGAGGGGAAGCGC
GTGTTCTGTCGCGCCGACCTCAACGTCCCGCTCGACGACAACAGAACAT
CACCGACGACACCGGTGTCGCGCCGCCATCCCCACCATCCAGTACCTTA
TCAAGAACGGTGCCAAGGTATCTCTCCAGCCACCTGGGTGCCCAAAG
GGTGTTACTCCCAAGTTTAGCTTGGCTCCCTTGTACCGCGATTGTCTGA
ACTTCTCGGCATACAGGTGACTAAGGCAGAAGATGTTATTGGACCAGATG
TTGAGAAATTGGTATCTGAATTGCCAAATGGTAGTGTGTTGCTCCTTGAA
AATGTTAGATTTTACAAGGAAGAGGAGAAGAATGACCCAGAGTTTGCCAA
GAAGCTTGCCCTCCTTAGCAGATCTCTATGTCAATGATGCATTTCGGAACAG
CCCACAGAGCACATGCATCAACTGAGGGAGTTACCAAGTTTTTAAAGCCT
TCTGTGCGAGGATTCTTTTGCAGAAGGAGCTTGACTACCTTGTGGAGC
TGTTTCAAGCCCTAAACGCCCATTTGCTGCCATCGTGGGTGGTTCAAAGG
TGTCATCCCAAGATTGGGGTTATCGAATCCCTTTTGAGAAATGTGATATC
CTTCTTCTAGGTGGTGGTATGATCTTCACATTTTACAAGGCACAAGGACT
CTCTGTTGGTCTCTCTTGGTTGAGGAAGACAAACTTGAGCTTGCAACAT
CTCTCCTTGCAAAGGCAAAAGAAAAGGGTGTCTCCCTTCTGTTGCCATCT
GACGTATCATTTGCTGATAAGTTTGCTCTGATGCTAACAGCCAGGTTGT
TCCAGCATCTGCTATTCCTGATGGTGGATGGGTCTGGATATTGGTCCAG
ATTAGTCGCGCTATTAGCTCTACCTGGAACAACACAGACAGTCATT
TGGAAATGGACCTATGGGTGCTTCGAATTTGACAAGTTTGAGTAGGAAC
TGAGGCTTATGCAAAGAAGTTGGCAGACCTTAGCGGCAAGGGTGTACAA
CCATTATTGGAGGAGGTGATTCTGTTGCCGCTGTTGAGAAGGTGGGAGTT
GCAGATGTTATGAGCCACATTTCAACGGGAGGTGGTGTAGCTTGGAGCT
GTTGGAAGGAAAGGAACCTCTGGAGTCATTGCACCTTGATGAAGCTGTCA
CTGTGGCCGTA

>V

-----G
CCAAGAAGAGCGTGGGTGACCTCACGGGGGCGGATTTGAAGGGGAAGAAG
GTGTTTGTGAGGGTGGATTTGAATGTGCCTCTTGATGACAACCAGAAGAT
TACTGACGATACCAGGGTTCGAGCTGCCATTCCTACTATCAAGTATTTGA
TCGACAATGGTGCTAAAGTCATTCTTTCTAGCCATTTGGGAAGACCAAAG
GGAGTAACTCCAAAATTCAGCTTGGCGCCTCTGTGCCAAGGTTATCTGA
GCTTCTTGGCATTGAGATTGTGAAGGCTGATGACTGTATTGGCCAGAGG
TAGAAAAGTTAGTGACTGCACTTCTGATGGTGGTGTACTCCTTCTCGAG
AATGTGAGGTTTTACAAGGAGGAAGAAAAGAATGAACCCGAATTTGCAAA
GAAGCTTGCCCTCCATAGCTGATCTCTATGTGAATGATGCATTTGGTACTG
CACATAGAGCCCATGCCCTCAACTGAGGGGGTCACTAAGTTCTTGAAGCCA
TCTGTTGCTGGTTTCTTTTACAGAAGGAAGTGGACTATCTCGTTGGGGC
AGTTTCAAGCCCAAAGAGGCCATTTGCTGCCATTGTGGGTGGTTCAAAGG
TCTCATCCAAGATCGGAGTCATTGAATCTCTCCTCGAAAAGTGTGATATC
CTACTTCTAGGTGGAGGAATGATCTTCACATTTTACAAGGCACAGGGTCT
CTCAGTGGGTTTCATCCCTGGTAGAGGAAGATAAGCTAAGCCTTGCTACAT
CACTTCTCGAGAAGGCCAAGGCAAAAGGAGTATCTCTCTGCTACCCACC
GATGTTGTGATTGCTGACAAGTTTGCCCTGATGCAACAGCAAGATTGT
GCCAGCATCAGCCATTCTGATGGATGGATGGGATTGGATATTGGACCTG
ATTCTATTAAGATATTCAATGAAGCTTTGGAGACCACCAAAACAGTTATC
TGGAACGGACCGATGGGAGTGTGTTGAGTTTGATAAGTTTGCAAGGAAC
AGTGGCCGTTGCAAAGAAGCTAGCGGAGCTGAGTGGGAAGGATGTGACAA
CAATCATTGGAGGTGGAGATTCTGTTGCAGCTGTGGAGAAAGTAGGAGTC
GCTGATGTGATGAGCCACATATCAACGGGTGGAGGTGCCAGTTTGGAGTT
GTTGGAAGGAAAGAACTCCCTGGTGTGTTGCCCTTGATGAAGCTACTC
CAGTTCCAGTG

GLY.4.1.2.13.fasta

>A

-----GCGTCTGCTAGCTTCGTTAAGCCTAACACCTCTCTTCTCCATG

GATCGGCCAACGC---TCCTTTGCTCACACCTCTGCTTCTTCTCT----
 -----CCTCCTCCTCGAGTCTCCTTCGCGATCCGCGCCGGT-----
 GCTTACTCCGACGAGCTTGTTAAACCGCCAAAAGCATTGCATCCCCTGG
 GAGAGGTATCTTGGCGATCGATGAGTCCAATGCAACCTGTGGGAAGAGGC
 TTGCTTCTATCGGCTTGATAAACACCGAGGACAACCGTCAGGCCTACAGG
 CAACTTCTGCTTACCACTCCTGGCCTCGGCGATTACATCTCTGGTTCCAT
 TCTCTTCGAGGAGACTCTTACCAGTCCACCAAGGACGGTAAGACCTTTG
 TCGATTGCTTGGCGATGCCAACATCGTCCCTGGCATCAAAGTTGACAAG
 GGCTTGTCTCCCTAGCCGGTTCCAACGAAGAGTCTTGGTGCCAAGGCTT
 GGATGGATTGGCCTCACGCTCTGCTGAGTACTACAAGCAAGGCGCTCGTT
 TTGCCAAGTGGAGGACAGTGGTGAGTGTTCCTGCGGTCTTCAGCACTG
 GCTGTGAAGGAAGCTGCGTGGGGGCTGGCTCGTATGCAGCCATCTCTCA
 GGATAATGGTCTTGTCCCATTTGTGGAGCCAGAGATCCTTCTGGACGGGG
 ACCACCCAATAGAGAGGACTCTGGAGGTGGCAGAGAAAGTGTGGTCAGAG
 GTGTTCTTCTACTTGGCACAGAAACGTCATGTTTGAGGGCATTCTGTT
 GAAGCCGAGCATGGTCACCCAGGCGCTGAGCACAAGAACAGGCGCTCTC
 CCGAGACCGTTGCAGATTTACGCTCACCATGCTGAAAAGGAGGGTTCCT
 CCGGCTGTCCAGGGATCATGTTTCTGTGAGGAGGACAATCAGAGGCAGA
 GGCCCACTGAACCTGAACGCCATGAACAGAGCCCAAACCCATGGCATG
 TGTCTTCTCATACGCACGTGCCCTGCAGAACTCCGTGCTCAGAACATGG
 CAAGGCAAGCCGGAGAAGATTGAGGCTCGCAGAAGGCACTGTTGGTGAG
 GGCAAGGCCAACTCACTGGCCAGCTCGGCAATACTCAGCCGAGGGAG
 AGAACGAGGATGCCAAGAAAGGAATGTTGTCAAGGGTTACACCTAC
 >P
 ---ATGGCCTGTGCAAGCTTTGTAAAGCTTAACGCAGCGTCGTCTTCATG
 GACGGGCCAAAAGTCTTCTTTTCGGTAAGCGATCTCCGGGATCCTCA----
 -----ACTCGCCGAGTCTCCTTCGATCCGCGCCAGC-----
 TCTTACACTGACGAGCTCGTCCAACTGCTAACTCATTGCATCACCTGG
 TCGTGGTATCCTTGCCATAGATGAATCAAATGCAACTTGTGGGAAGAGGT
 TGGCATCTATTGGCTTGATAAACTGAAACCAACCGACAAGCATAACAGA
 CAACTTCTACTGACCACTCCTGGCCTTGGTGAATACATTTCCGGTGCCAT
 TCTCTTCGAGGAGACATAACAGTCTACAACCTGATGGAAGGAAGTTGCG
 TGGATTGCCTGCGTGATGAGAACATTGTACCTGGCATCAAAGTTGACAAG
 GGTTTAGTTCCCTGCCAGGATCAAACAATGAGTCTTGGTGCCAAGGTTT
 GGATGGATTGGCCTCAAGGTCTGCTGAATATTACAAGCAAGGCGCACGTT
 TTGCTAAGTGGAGGACTGTTGTGAGCATTCCATGTGGCCCATCTGCTCTG
 GCTGTCAAGGAAGCTGCATGGGGACTTGACAGATATGCTGCCATTTCTCA
 GGATAATGGTCTTGTGCCATT---GAGCCTGAGATTCTGCTTGATGGGG
 ACCATCCAATTGACAGGACCTTGAAGTTGCTGAGAAGGTCTGGGCAGAA
 GTCTTTTACTACTTGGCTGAAAACAATGTTGTGTTTGAGGGCATCCTACT
 TAAGCCTAGCATGGTAACGCCAGGGGCGGAACACAAGGAGAAGGCTTCCC
 CAGATACCATAGCCAAATATACACTCACAATGCTTAAAGGAGAGTACCT
 CCTGCAGTTCTTGGCATAATGTTTTGTGCGGAGGACAATCTGAAGTGCA
 AGCAACACTCAATCTCAATGCAATGAACCAAAGCCCCAACCCGTGGCATG
 TTTCTTCTCATATGCACGTGCACTGCAGAACCCGTGCTCAAGACATGG
 CAGGGACGCCCTGAGAACGTGGAAGCTGCTCAGAAATCACTTTTGGTGCG
 TGCCAAGGCTAACTCGCTGGCCAGCTTGGTAGATATTCCGCTGAGGGTG
 AAAGCGAGGAAGCTAAGAAGGGAATGTTGTAAGGGCTACACCTAC
 >R
 -----GCCTGTGCAAGCTTTGCAAAGCTAAACGCAGCTTCTTCTACCTG
 GATCGGCCAACAG---TCTTTTGGTCAGCGA---CCTGGATCTTCTCGG
 CTGTTTTTGAACCTCGCCGAGTCTCCCTACCGATCCGTGCCAGC-----
 TCTTACAAGGACGAGCTCGTTCAAACAGCTAAACTGTTGCATCACCTGG
 ACGTGGAATCCTCGCTATTGATGAATCGAATGCAACTTGTGGGAAGAGGT
 TGGCATCTATTGGGTTGGACAACAATGAGACTAACCGGCAGGCATATAGA
 CAACTCTTACTGACCACTCCTGGTCTTGGTGAATATATTTCTGGTGCAAT
 TCTTTTTGAGGAACTCTTTACCAGTCCACCACAGATGGAAAGAAAGTTTG
 TGGATTGCTTGCCTGATCAGAACATTGTTCTTGGTATCAAAGTTGACAAG
 GGTGTTGGTGCTCTTCCAGGATCAAACAATGAGTCTTGGTGCCAAGGTTT
 GGATGGATTGGCATCAAGATCTGCTGAGTACTACAAGCAAGGTGCTCGTT
 TTGCTAAGTGGAGAACAGTTGTGAGCATTCCCTGTGGTCTTCTGCTTTG
 GCTGTTAAAGGAAGCTGCATGGGGACTTGTGCTGTTATGCTGCCATTTCTCA
 GGACAATGGTCTTGTGCCATTGTTGAACCTGAGATTCTTCTTGTGAGGAG
 ACCATGGAATTGAGAGAACACTTGAAGTCGCAGAGAAGGTCTGGGCTGAA
 GTCTTTTCTACTTGGCGGAGAACAATGTTGTGTTTGAGGGTATCCTGCT
 TAAGCCTAGCATGGTAACACCAGGTGCTGAACACAAGGAGAAAGCTTCTC
 CTGATACCATTGCCAAATACACACTCACAATGCTCAGAAAGAGTACCA
 CCTGCAGTTCTTGGATCATGTTTTTGTGCGGAGGACAATCTGAAGTTGA
 AGCGACCTTAACCTCAATGCAATTAACCAAAGCCCCAACCCCTGGCATG
 TGTCTTCTCATATGCACGTGCACTGCAGAACTCTGTGCTTAAGACATGG
 CAGGGGCATCCAGAGAATGTGGAAGCTGCACAGAAGGCACTTTTGGTACG
 TGCCAAGGCAAACTCCTTGGCACAGCTTGGAAAGTATTCTGCTGAGGGTG
 AAAACGAGGAAGCTAAGAAGGGCATGTTGCTCAAGGGCTACACCTAT
 >O
 GCGATGCTGACGGCCAAGCTCACCTCCCCCCCCGCCGCCACCACC---TG

GCTC-----CCCGGCGGCGGCCGCC
GATCCGCGCCGCCCGCCGTGCTACGGTG--ATCCGCGCCGCCCGCCGT
TCCTACGCGCGACGAGCTCGTCTCCACCGCGAAATCTGTTGCTTCCCCAGG
GCGTGGTATCCTGGCAATTGATGAGTCGAATGCCACATGCGGAAAGAGAT
TAGCATCCATTGGTTTGGACAACACAGAAGTTAACCGCCAGGCTTACAGG
CAGCTTTTACTGACCACTGCTGGTCTTGGTGAATATATTTCTGGTGTCTAT
CCTTTTTGAGGAACTCTTTATCAGTCAACCACTGATGGTAAGAAAGTTTG
TTGACTGCTTGAAGGATCAGAATATCATGCCCGGTATCAAGGTCGACAAG
GGCTTGGTTCCATTGCTGGGTCCAACAATGAATCTTGGTGCCAAGGCCT
AGATGGTTTGGCTTCAAGGTGTGCTGAGTACTACAAGCAGGGGGCACGCT
TCGCTAAGTGGCGGACTGTTGTTAGCATCCCTTGTGGTCCCTCAGCATT
GCAGTCAAGGAAGCGGCATGGGGACTTGTCTGATATGCTGCCATTGTCTA
GGACAATGGCTTAGTGCCAATTGTTGAGCCAGAGATCCTTCTTGATGGTG
ACCATGCGATCGAGAGAACTCTTGAAGTGGCAGAGAAAGTGTGGTCTGAG
GTATTCTTCTACCTGGCCCAAAACAATGTTCTTTTGAGGGTATCCTGCT
GAAACCCAGCATGGTGACCCCTGGAGCTGAACACAAGCAGAAGGCCACTC
CAGAAGCCATTGCGAAGCACACCCTTACAATGCTGAGGAGGAGAGTCCG
CCTGTGCTGCCCTGGAATCATGTTCTTTCTGGTGGGCAATCTGAGGTGGA
GGCAACCTGAACCTGAACGCGATGAACCAAGAACCAACCCATGGCATG
TGTCCTTCTCATACGCCCGGGCTCTCCAGAACTCGGTGCTGAAGACATGG
CAGGGGCGCCCCGAGAAGCTGGAGGCAGCGCAGAAGGCACTGCTGGTCCG
TGCCAAGGCGAACTCGCTGGCTCAGCTCGGTGCTACACCGGCGAGGGCG
AGAGCGATGAGGCCAAGAAGGGA--TTCCAGAAGGGCTACACTTAC
>V
-----GCCTCTGTGACCTTTGCGAAGTTTAAAGCCTCGTCTTCGAGTG
GATCGGACAACAG---TCCTTCTCTCAACGC---CAGGGATCGTCC---G
CTCGCTTCCCGGCACGCGAGTCTCGGTCCCGATCCGTGCCGGG-----
TCGTAATCCGAGGAACTCGTTACAGACTGCTAAACTGTTGCATCACCTGG
CCGTGGTATACTAGCCATTGATGAGTCAATGCTACCTGCGGGAAGAGGT
TGGCATCTATTGGTCTGGACAATACAGAACCAACCGTCAGGCCATATAGA
CAACTTCTACTGACCACTCCTGGCCTTGGTGAATATATTTCTGGGGCCAT
CCTTTTTGAGGAAACACTGTACCACTCAACAATGATGGAAAGAAAGTTTG
TGGACTGCTTGCCTGAAAAGAAAATTTGCTGGCATCAAAGTTGATAAG
GGTTTGGTTCTCTACAGGATCAACAATGAATCTTGGTGCCAAGGCCT
GGATGGTTTGGCTTCAAGATCTGCTGAATACTATAAGCAAGGTGCTCGCT
TTGCAAAGTGGCGCACTGTTGTTAGCATTCCCTGTGGTCTTCTGCCTTG
GCTGTCAAAGAAGCTGCATGGGGTCTTGACGCTATGCTGCTATTTCTCA
GGACAATGGTCTTGTGCCCATAGTAGAACCTGAGATTCTTCTTGATGGGG
ATCACCAATTGATAGGACACTCGAAGTTGCAGAGAAAGTGTGGTGGAG
GTCTTCTTTACTTGGCACAAAACAATGTCTTATTTGAAGGAATCCTACT
TAAGCCTAGCATGGTAACCCAGGAGCTGAACACAAGGAGAAGGCTTCTC
CAGAGACCATCGCCAAATATACACTAACAATGCTTAGAAGGAGAGTACCT
CCTGCAGTTCTTGGAAATC---TTTTGTCTGGAGGGCAATCTGAAGTGG
AGCAACATTAAACCTGAATGCAATGAATCAAAGCCCCAACCCATGGCATG
TTTCTTCTCATATGCTCGTGCCTGCAAAACACTGTCCTAAAGACGTGG
CAAGGACATCTGAGAATGTTGAAGCGGCACAGAAGTCTCTTTTAGTTCG
GGCAAAGGCAAACTCCCTGGCTCAGCTTGGAAAGTACTCTGCTGAAGGTG
AAAGTGAGGAAGCTAAGAAAGGAATGTTGTCCAGGGCTATACCTAC

GLY.5.3.1.9c.fasta

>A
GCGTCATCAACCGCTTTGATTTGTGATACCGAAGCGTGAAGGATTTGAA
GGGACATGTAGAAGATATTAAGAAGACTCATTTGCGTGATTTGATGAGTG
ATGCTAATAGATGCCAGTCCATGATGATGGAGTTTGATGGGTTGCTGTTG
GATTATTCTCGACAGCGTGCAACTGTTGAGACAATGGACAAGCTTTTGAA
CTTGGCAAAGGCTTCTCAATTGACAGAGAAGATCAGCCGCATGTTCAATG
GGGAGCATATTAACAGTACAGAGAACAGATCAGTTCTTCATGTTGCGCTC
CGTGCTCCAAAGGATGCAGTTATCAAGGCTGATGGAATGAATGTGGTTCC
AGAAGTGTGGAACGTTCTAGATAAGATCAAGGAATTTCTGACAAAATTC
GCTCTGGTTCTATGGGTTGGAGCCACTGGCAAAACGCTGAAAGATGTCATT
GCCATTGGTATTGGTGGTAGCTTCTTAGGTCCACTGTTTGTCCACACGGC
TCTCCAAACAGATCCTGAAGCTCTAGAGTCTGCTAAAGGACGCCAGCTGC
GATTTCTTGCAAAATATTGATCCTGTTGATGTTGCTAGAAATATCAGTGGA
CTAAATCCAGAACTACTCTAGTTGTGGTGGTCTCGAAAACGTTTACAAC
AGCTGAAACAATGCTTAACGCCAGAACATTGAGGGAATGGATAACAGCTG
CTCTTGGGGCTTCACTGTTGCAAAACATATGGTTGCTGTGAGCACTAAT
CTTGGCGTTAGTAGAGAAGTTTGGTATTGACCCGAACAATGCATTTGCATT
TTGGGACTGGGTTGGTGAAGGTACAGTGTGTCAGTGCGGTTGGAGTCT
TACCTTTGCTCTGCAAGTATGGCTTCTCCATGGTTGAGAAGTTTTTGAAG
GGAGCTTCAAGCATTGATCAGCATTTCCAGTCCACACCGTTTCGAGAAGAA
TATACCTGTGCTTTAGGGTGTGAGTGTATGGAATGTATCATTTCTTG
GATATCCTGCTAGGGCCATCTTACCTTATTCGCAAGCCCTTGAGAAATTT
GCTCCACACATTCAACAG-----

```

----GTTAGTATGGAGAGTAATGGAAAGGGAGTCTCAATTGATGGTCTAC
CTCTCCCGTTTCGAGACTGGTGAGATTGATTTTGGTGAACCTGGAACAAAT
GGTCAACACAGCTTTTACCAACTCATTACCAGGGACGCGTAATCCCTTG
TGATTTTCATTGGCATTGTGAAGAGTCAGCAACCTGTGTACCTTAAGGGAG
AGGTGGTCAGC---CACGACGAGCTCATGTCAAACCTTTTTCACAGCCT
GATGCTCTTGCATATGGAAAACTCCTGAACAGCTGCAGAAAGAGAATGT
TTCAGAAAAATCTCATTCCCCATAAGACATTCTCTGGAATCGACCTTCTC
TTAGCCTTCTACTTCCAGAATTGACTGCTTACAATGTTGGCCAGTTGTTG
GCTATCTATGAACACAGAGTAGCAGTTCAAGGCTTTGTGTGGGGTATCAA
TTCGTTTGACCAAGTGGGGCGTTGAGCTAGGAAAAGTTCTGGCTACTCAGG
TCAGGAAACAGCTTCATTATCAGCACTCAAGGAACCGCTCCCGAGGGA
TTCAATTACAGTACCACCACACTTTTGAACGATATCTCGAGACAAGTTC
CGAG-----CCCCAGATG
>P
ATGGCTTCATCGACTTTAATTTGCGAAACCCAGCCATGGAAGGACTTGAA
GTCTCATGTTGAGGACATTAAGAAGACTCATTACGCGAGTTGTTGAGTG
ACACCGACCGATGCAAGTCAATGGTGGTTGATTTTGATGGAATAACGCTT
GACTACTCAGCCCAACGAGCCACCCCTGGTACCATGGATAACCTTTATAA
TCTGGCAGAGGCAGCTCATCTCAAGGAAAAGATTGATCGCATGTTCAATG
GGGAGCATATAAACAGCACTGAGAATAGGTGAGTGTTCATGTAGCTCTT
CGTGCTCCAAGGGATGCAGTTATGCAAAGTGATGGCAAGAATGTTGTACC
AGATGCTTGGAATGTTCTGGACAAGATCAAGGATTTCTCTGAGAGGGTCC
GCAATGGTTCTTGGGTTGGAGCCACAGGAAAAGCACTTACTGATGTTATT
TCAGTCGGCATTGGTGGCAGCTTCTGGGTCCTCTTTTTGTGCATACCGC
TCTTCAAACAGATCCAGAGGCCAGCAATGTGCTACAGGACGCCAATTGC
GATTTCTTGCAAATGTAGATCCTATTGATGTTGCTAGAAATATTGAGGC
CTCAAACCTGAGACTACTCTCGTTGTTGTGGTTTCAAAGACTTTTACCAC
AGCTGAAACTATGCTGAATGCTCGAACAATAAGGGCATGGATTTCAAAG
AACTCGGACCATCTGCAGTTGCCAAGCATATGGTTCAGTTAGCACTAAT
CTTATGCTTGTGGAAGTTTGGCATTGATCCTAACAAATGCTTTTGCATT
CTGGGACTGGGTTGGTGAAGATATAGTGTCTGTAGTGTGTTGGCGTGT
TACCTTTGTCTCTCAATATGGTTTCAGAGTTGTTGATAAGTCTTGAAA
GGAGCATCAAGCATTGATCAGCATTTCTATTCTGCACCATTTGAGAAAA
TCTACCTGTACTTCTAGGTATGTTGAGCGTGTGGAATGTTTCAATTTTTG
GATATCTGCAAGAGCCATCTTACCTTATTCTCAAGCCCTGGAGAAATTT
GCACCGCACATCCAGCAGGCAACCTACTTGAGCTCACTTTCCCTTCCCT
TTTAGTCAGCATGGAAAGTAATGGCAAGGGGGTATCTATTGATGGTAAAC
CACTTCCCTTTGAAACTGGTGAAATGATTTTGGTGAACCAAGGAAACAAAT
GGTCAGCATAGCTTTTATCAACTAATTCACCAGGGACGTGTGATTCCCTG
TGACTTTATTGGTGTGTGAAGAGTCAGCAACCGGTATACCTAGAAGGAG
AGGTTGTAATAACCATGATGAATCATGTCCAACCTTTTCCGCCAGCCA
GATGCCCTTGCATACGGGAAGACACCAGAAGAGTTGCAAAAAGAGAATGT
TCTGCAGCATCTCATACCTCACAAGACCTTCTCTGGCAATAGGCCTTCTC
TCAGCCTTCTGCTCTCTTCAATTGGATGCTTACAAAATTTGACAGTTGTTG
GCGATCTATGAACACAGAGTTGCTGTGCAAGGCTTCAATGGGGCATCAA
TTCTTTTGGACAGTGGGAGTTGAGTTAGGAAAGTCGTTGGCTACTCAAG
TTAGAAAGCAACTTCATGCATCTCGTGTGAAGGGAGAACCAGTTGAGGGT
TTTAATTTTCAGCACTACAACAATGCTAACAAAATATCTGAAGGAAAGTTT
GGAAGTTCCAGCCAATCCTCCAACAATCTTCTCTCGGATA
>R
---GCTTTATCGGCTCTAATCTGTGATACTCAGCCCTGGAAGGACTTGAA
GGCTCATGTTGAGGACATTAAAAAAATCATTACGCAATTTGTTGGGCG
ACACTGAACGATGCAAGTCCCTG---GTTGAATTTGATGGAATCATGCTT
GACTACTCAAGGCAAAACAGCCACTATTGACACACTGGATAAGCTCTATAA
CTTGGCAGAGGCAGCTCATCTCAAGGAAAAGATTATCGC---TTTAATG
GAGAGCGTATCAACAGCACAGAGAATAGGTCCGTGCTTCATGTTGCTCTT
CGTGCTCCACGGGATGCAATTGTAACAGTGATGGAAGAATGTTGTACC
AGATGTTTGGAAATGTTCTTGACAAGATTAAGGAGTTTTCTGAGAGGGTTC
GCAATGGTTCTTGGGTTGGAGCTACAGGGAAACCACTGACGGATGTTATT
GCCATTGGTATTGGTGGCAGTTTCTTAGGTCTCTTTTTGTGCACACTGC
TCTTCAAACAGATCCAGAGGCTAGCAATCTGCAAAGGGCCGCCAGCTGC
GTTTTCTTGCAAATGTTGATCCAATTGACGTTGCTAGAAATATTGACAGC
CTAAGCCCTGAAACCACCCTCGTTGTTGTGGTATCAAAAATTTTACTAC
TGCTGAACT---CTGAATGCACGAACCTAAGGGCATGGATTTCAAGAG
AACTTGGGCCCTTCTGCAGTTGCAAAGCACATGGTGGCAGTTAGCACAAAT
CTTACGCTTGTGAAAAGTTTGGTATTGATCCTAATAATGCTTTTGCATT
CTGGGACTGGGTTGGTGGCAGGTATAGTGTGTCAGTGTGTTGGAAGTGT
TGCCTTTGTCTCTCCAATATGGTTTCTCAATTGTTGAGAAATCTTGAAA
GGAGCATCAAGCATTGATCAGCACTTTTCTTACGCGCCATTTGAGAAAAA
TATTCCTGTACTTCTAGGTTTGTGAGCGTATGGAATGTTTCAGTTCTTG
GATATCCTGCAAGAGGCACTTACCTTATTTCCCAAGCTCTGGAGAAACTT
GCCCCACACATCCAACAG-----
----GTTAGCATGGAAAGTAATGGCAAGGGGTATCCATGGATGGTGTGC
CACTTCCCTTTGAGACTGGCGAAATGATTTTGGTGAACCAAGGACAAAT
GGACAGCACAGCTTTTATCAACTAATTCACCAGGGACGTGTAATCCCTG

```

TGATTTTATTGGTGTGTGAAGAGTCAGCAACCTGTATACCTCGAAGGAG
 AGGTTGTCAATAACCATGATGAACCTCATGTCCAACCTTTTTTGCACAACCA
 GATGCGCTTGCTTATGGGAAGACAGCAGAGCAGTTGCAAAAGGAGAATGT
 TCCACAGCATCTAATTCCTCACAAGACCTTCTCCGGAACAGGCCTTCTG
 TCAGTATTCTGCTTTTCGTCACCTGGATGCTTACAAAATTGGACAGCTGTTG
 GCAATCTATGAACACAGAGTAGCAGTGCAAGGCTTCGTTTGGGGAATCAA
 TTCTTTTGACCAGTGGGGAGTTGAATTAGGGAAGTCACTGGCAACTCAAG
 TCAGAAAGCAACTCAATGCATCTCGTACAAAGGGAGAACAGTTGAGGGC
 TTTAATTTTCAGCACTACGACAATGCTAACAAGATACCTTGAGGAACTTC
 GGAAGTGCCAGCCAAGCCTCAAACTCTTCTACCTCGGATA
 >0
 --- GCGTCGTCGGCGCTAATCTGCGACACCGAGCAGTGGAAGGGCCTCCA
 GCGCGACGTCGGGGCGATTTCAGAAGACGCACCTGCGCGATCTGATGGATG
 ATGCCGAGCGCTGCAAGGCAACA --- GCTGAGTATGAAGGCATATTTCTG
 GATTACTCGAGGACGCTGCAACTGGCGAGACCATGGAGAAGCTGTTTAA
 ATTGGCAGAGGCGGCAAGCTCAAGGAGAAGATTGAGAAG --- TTTAGTG
 GTGACAAGATAAATAGCACAGAGAACAGATCTGTGCTTCATGTAGCTCTA
 AGGGCTCCAAGAGACGAAGTAATAAAAAGTGATGGGGTCAATGTGGTTCC
 CGAAGTTTGGGGTGTAAAAGATAAAATCAAGCAGTTTTTCAGAACTTTTA
 GGAGTGGATCATGGGTTGGGGCAACTGGTAAAGCATTGACAAATGTTGTG
 TCAGTTGGAATAGGTGGTAGCTTTCTTGGTCCTCTGTTCTGTCATGCTGC
 CCTCCAGACAGATCCAGAAGCTGCAGAATCTGCCAAAGGGCGACAGTTAA
 GATTCTTGCAAAATGTTGACCCTGTTGATGTTGCACGAAGCATCAAAGAT
 TTAGATCCTGAAACAACACTTGTGTGGTTGTCTCAAAGACCTTCACAAC
 AGCTGAAACA --- TTAATGCTCGAACTCTCAAGGAGTGGATTGTCTCTT
 CTCTTGGACCTGATGCTGTTGCGAAACATATGATTGCTGTGAGTACCAAT
 CTTGAGCTTGTGGAGAAGTTTGAATTGACCCTAAAAATGCTTTTGCATT
 TTGGGACTGGGTTGGTGGTCGCTATAGTGTTCAGTGCTGTTGGTGTCC
 TGCCCTTATCTCTTCAATATGGTTTTCCGATTGTTTCAAGAAATTTTGGAG
 GGTGCGACGAGCATCGACAAACACTTCCGTTTCACTTTCATTGAGAAAAA
 CATACTGTACTCCTTGGTTTGTGAGTGTGTGGAATGTTTCTTTCTTG
 GATATCCAGCTAGAGCAATATTGCCCTATTCCAGGCACTTGAGAAATTT
 GCACCACATATTCAGCAG-----
 --- CTTAGCATGGAGAGTAATGGAAAGGGTGTCTCTATTGATGGCGTTC
 AATTGTCCTTTGAGACTGGTGAAATTGATTTTGGTGAACCTGGAACCAAT
 GGGCAACACAGCTTCTATCAATTAATCCATCAGGGGAAGAGTTATTCCTTG
 TGATTTTATCGGCGTCGTGAAAAGCCAGCAACCCGTTTACTTGAAAGGGG
 AAATTGTGAGCAATCATGACGAATTGATGTCCAATTTCTTTGCTCAGCCT
 GATGCACCTTGCTTATGGAAAGACCCCTGAACAACCTGCACAGCGAGAAAGT
 ACCTGAACATCTTATCTCTCATAAGACTTTTTCAGGGCAACCGACCATCAC
 TTAGTTTATTGCTGCCCTCATTATCTGCTTATGAGATTGGACAGCTTTTA
 TCCATCTACAGACACCGGATTGCAAGTTGAGGGTTTCTATGGGGAATAAA
 CTCATTTGACCAGTGGGGAGTGGAACTGGGAAAGTCTCTGGCTTCTCAAG
 TGAGAAAATCTCTGCATGCATCCCGCATGGAAGGAAAGCCTGTCCAGGGG
 TTCAACAGCAGCACTGCAAGTTTGTGACACGATATCTCGCTGTTGAGCC
 ATCCACTCCTTACAACACTACAACA --- ATGCCGAAAGTT
 >V
 --- GCTTCTCTACTCTAATCTGCGATACTGAGCCATGGAAGAACTTGAG
 GGCTCACGTTGAGGATATTAAGAAAGACTCACTTGCCTGATTTGATGAGTG
 ACACGAAGAGATGCCAGTCAATG --- GTCGAGTTTGACGGGTTGCTGTTG
 GACTACTCACGGCAGTGTGCCACTCTGGAAACCATGGATAAGCTCTTCAA
 ATTGGCAGAGGCAAGAAATCTCAAAACAAAAAATTGGCAGA --- TACAATG
 GAGAGCGTATAAACAGTACAGAGAATAGGCCGGTACTTCATGTAGCTCTC
 CGTGCTCCAAGGGATTCTGTTATACAGAGTGATGGCAAAAATGTGGTCCC
 AGATGTCTGGAAAGTTCTAGACAAGATCCAAGTATTTTCTGAGACAGTCC
 GCAGTGGTTCTGGGTTGGAGCCACTGGAAAAGCATTGACAGATGTAATT
 TCTGTGGGCATTGGTGGCAGCTTCTTAGGCCCTCTTTTTGTGCACACTGC
 TCTGCAACAGATCCAGAGGCTAGTAAATTGCAAAAGGGCGCCAGCTGC
 GTTTCCTTGCAAAATGTTGATCCTGTCGATGTGGCTAGAAATATCACAGGG
 CTGAACCTGAAACTACTTTAGTTGTGGTGGTTTCAAAGACTTTTACAAC
 AGCTGAAACT --- TTGAATGCCCGTACTCTAAGGGAATGGATTTTCAGCTG
 CTCTAGGGCCTTCAGCAGTAGCAAAGCATATGGTGGCTGTCACTACTAAT
 CTTACGCTTGTAGAACAGTTTGGCATTGATCCTAATAATGCTTTTGCATT
 CTGGGACTGGGTAGGAGGCCGATATAGTGTTCAGTGCTGTTGGTGTGC
 TGCCCTTATCTCTCAATATGGTTTCTCAGTTGTTGAGGGTTGG-----

 -----GGGAGGCTCGCG-----

 -----GCTCGAGGAGCAGCTGCAGAAGGT-----
 -----GGATGGGAC
 GCGGAAAGCGACTCATGGGGATTATGCAGA-----
 -----GTGGGCTATTGAGA-----

```

-----
-----
-----TGGGCTTTG-----
-----TGGTGGAGCCTAGTGGGT-----
-----
-----CTGAGGGTT
-----

```

GLY.5.3.1.9p.fasta

```

>A
GCCTCTCTCTCAGGCCTATACTCTTCTCTCCATCTCTCAAACCTGCCAA
AAACCATTCTTTAAAGCA-----TTGCCGGCGC
AATCTCGAGATTCTTCTCTTTCCACATACCTCCAAA-----CCCACC
AATCTACCGTTGACTCTCTCCTCGGCTCGTTCTGTGGCACGCGACATTTT
TCATGCGGAT-----TCGAAGAAAGAGCTTCTTAAGG
ATCCGGATGCGTTGTGGAAGCGATACCTCGATTGGTTTTATCAGCAGAAA
GAGCTTGGGTTGTATCTGGATATTAGTCGGGTCGGATTACCGATGAGTT
TGTGGCTGAAATGGAGCCAAAGTTTCAAGCTGCGTTTAAAGGCTATGGAGG
ATCTTGAAAAAGGGTCTATAGCGAATCCTGATGAAGGAAGAATGGTTGGA
CATTACTGGCTTAGGAACCTAAGCTCGCACCTAAGCCTACTTTAAAGAC
CTTGATCGAGAACACACTTGATTCTATTTGCGCTTTCTCCGACGACATAA
TCTCTGGAAGATAAAGCCACCATCTTCTCCTGAGGGTCGTTTTACTCAG
ATACTTTCTGTTGGCATTGGAGGCTCGGCTCTTGGACCTCAATTTGTCGC
TGAGGCCTTGGCTCTGATAATCCTCCA-----ATAAGATTCATTGACA
ACACCGACCCTGCTGGAATTGATCATCAGATTGCACAACCTGGGCCAGAG
CTGGCCTCGACTTTAGTAGTTGTCATCTCAAAGAGTGGAGGTAATCCTGA
AACTAGAAATGGACTACTGGAAGTGCAGAAAGCATTCCGA---GCTGGTC
TGAATTTTCGCAAAACAGGGTGTTGCAATAACCCAAGAGAATCATTACTG
GATAATACGGCAAGAATTGAAGGTTGGCTAGCTAGATTTCTATGTACGA
CTGGGTGGGTGGAAGAACATCAATAATGTCTGCAGTTGGTCTGCTTCCAG
CAGCACTACAGGGGATTAATGTTAGGGAGATGCTTACTGGTGCTGCTTTA
ATGGATGAGGCTACTAGGACAACCTTCGATCAAGAATAACCTGCAGCGCT
CTTAGCTATGTGTTGGTACTGGGCTTCCAATGGCGTTGGTTCCAAGGATA
TGTTTGTCTTCTTACAAGGATAGCTTATTGCTATTTAGTCGGTATCTG
CAGCAGCTGGTCATGGAATCACTAGGAAAGGAGTTTGATCTTGACGGTAA
CACTGTTAATCAAGGGTTAACTGTGTATGGAATAAGGGAAGCACAGATC
AGCACGCCTACATTCAACAGCTGAGAGATGGTGTGCACAATTTCTTTGCA
ACCTTCATAGAAGTGTTACGGGACAGACCCCTGGTCACGATTGGGAGCT
TGAGCCAGGTGTCACTTGTGGAGACTATCTCTTTGGGATGCTACAGGGAA
CTAGATCTGCTTTATATGCAAAACGGTAGAGAATCCATCAGTGTACAAATC
CAGGAAGTGACACCAACATCTGTTGGGGCTATTATAGCTCTTTATGAAAG
AGCTGTTGGTTTATATGCCTCAATTGTGAACATAAATGCTTACCACCAAC
CCGGTGTGGAAGCTGGGAAAAAGGCAGCAGAGAAGTTCTCGCCCTGCAA
AAGCGGGTATTGTGAGTTCTTAATGAAGCCACTTGTAAGATCCGGTAGA
GCCATTGACACTGGAAGAGATAGCTGATCGTTGTGCTGCTCCTGAGGAAA
TAGAAATGATATACAAGATCATAGCGCACATGTCTGCAAAACGACAGAGTT
CTGATAGCAGAAGGAACTGTGGATCGCCACGAAGTATCAAAGTGATTTT
GGCGGAGTGCAATGTGGATGAC-----CTGTACGCA
>P

```

```

-----
-----
-----
-----
-----CTAGAAAAAG
ATCCTAATTCTCTATGGAGGAGATACGTTGACTGGCTCTACCAGCACAA
GAATTAGGGTTGTATTTGGATGTGAGTCGGATCGGTTTACCGGACGAGTT
TGTTTCGGAATGGAGCCCCGATTCCATAAGGCGTTTAAAGACATGGAGG
AATTAGAGAAGGGCGCGATTGCCAATCCTGATGAAGGGAGAATGGTTGGG
CATTATTGGTTGAGGAATTCACCTTAGCGCCTAAGTCGTTTTTGAAGAAC
TCAGATTGATAAGGCTCTTGATGCTGTTTGTGATTTTGTGATCAAGTCG
TCAGTGGTAAGATTAAAGACTCCGGACGGG-----GGGCGTTTACTCAA
ATACTTTCTGTTGGGATTGGTGGTTGAGCTCTCGGTCCGAGTTTGTTC
GGAGGCATTGGCGCCCGATAATCCTCCTCAAGATAAGGTTTCAATTGATA
ATACAGACCCTGCTGGAATTGATCATCAGATTGCTCAGCTTGGACCGAG
CTGGCTTCTACTCTTGTCATTGTGATTCAAAGAGTGGAGGAACCCCTGA
AACTAGAAATGGTTTGTGGAAGTACAGCAGGCCTTTCTGTAAGCTGGAC
TGGATTTTGCAAAACAGGGTGTGCTATAACACAAGAAAAATTCATTACTA
GACAACACTGCAAGAATTGAGGGCTGGTTGGCTAGATTCCCTATGTTTGA
TTGGGTGGGTGGTAGGACATCTGAAATGTCTGCAGTTGGCCTGCTTCCAG
CAGCGCTTCAGGGGATTGACATCAGAGAAATGCTAGCTGGTGCGGCATTG
ATGGATGAGGCAATAGGACCACAGTGTCTAGGAATAACCTGCAGCTTT
GCTAGCTTTATGTTGTTGTTGGCTTCTGAGGGGGTGGGATCTAAGGACA
TGTTTGTGCTTCCATACAAGGACAGCCTATTGCTATTCAAGTAGGTATTG

```

CAACAGCTGGTCAATGGAATCACTTGGAAGGAATTTGATCTTGATGGTAA
 TCGGGTAAATCAAGGACTTACCGTCTATGGAATAAAGGGAGCACGGATC
 AGCATGCTTACATTCAACAACTGAGAGAAGGAGTGCACAATTTCTTTGTA
 ACGTTTCATTGAAGTGCTACGAGATAGACCCCTGGCCATGATTGGGAGCT
 TGAACCAAGGTGTACATGTGGTACTACCTATTTGGAATGTTACAGGGAA
 CTAGGTCAAGCTCTGTATGCCAAGGACAGGGAATCTATAACAGTTACAGTG
 CAAGAAGTGACCCCAAGATCTGTTGGGGCTCTTATTGGGCTTTATGAGCG
 GGCAGTTGGGATATATGCTCACTTGTCAACATCAATGCTTATCATCAAC
 CTGGTGTGGGAAGCAGGGAAGAAAGCAGCAGGAGAAGTTTATGCTCTTCAG
 AAGCGGGTTTTGGCAGTGCTCAATGAGGCAAGCTGCAAAACAGCCCGTGA
 ACCATTGACAATCGAGGAAGTAGCCGATCGTTGCCATGCTACCGAAGATA
 TTGAAATGATATACAAGATAATTGCACACATGGCTGCTAATGACAGAGCA
 CTTATAGCTGAAGGCAGCTGTGGTTACCGCGCAGTCTCAAAGTTTTCT
 TGGAGAGTGTAATGTGGATGAG-----
 >R
 GCCTCACTCTCCGGTCTCTGCTCCTCGTCTCCATCTCTCAAAACAAAAAC
 TTCTCTTCTTAAACAGCT-----CTCAACTCTGCTCCTCTTAAACAC-
 -----TCACTCGCATTTCACCTCGC-----ACTTCC
 CGATTTACCCCTGCACAGTCAATCGCGCGGAGATTTCCGCTGATTTATC
 CAAAACAACTATAAATATATCGGTCAAACGGAGCAAGGACTAGAGAAAG
 ATCCAAACAGTCTATGGAGAAGATACACGAGTGGCTGTACCGACACAAG
 GAGTTAGGGCTTTATTTGGATGTGAGTCGGATCGGGTTACGGACAACCT
 TGTAAGAGATGGATCCCGCTTTCAAAGGCGTTTAAGGATATGGAAG
 AACTTGAGAAAGGTGCAATTGCGAATCCTGATGAAGGGAGAATGGTTGGT
 CATTATTGGTTGAGGAATCCGGGCTTGTCTCTAAAGCTTTTTGAAACA
 ACAGATTGATAAAGCTCTCGACGCCGTTTGTAGTTTGTGCGGACGTCG
 TCAGTGGTAAGATTAAGCCCCGAATTACCAGAAGGCCGTTTACTCAA
 ATACTTTCTGTAGGTATTGGAGGTTTCACTCTTGGACCACAATTTGTTGC
 CGAGGCTCTTGCCTGATAATCCTCTCTCAAGATAAGATTTATTGATA
 ATACGGATCGACGCTGGAATTGATCATCAGATTGCTCAGCTTGGTCTGAG
 CTTTCTTCTACTCTTGTGATAGTGATTTCTAAGAGTGGAGGTACCCCTGA
 AACTAGAAATGGTTATTGGAAGTTGAGAAGGCCCTTTCGTGAAGCTGGCC
 TGGAGTTTGCAAAACAGGGAGTTGCTATAACGCAAGAAAATTCATTACTA
 GACCAACTGCAAGAATTGAAGGATGGTTAGCTAGATTCCCATGTTTGA
 CTGGGTTGGTGGGAGGACATCTGAAATGTCTGCAGTGGGCCCTTCTCCAG
 CAGCACTTCAGGGGATTGACATTAGGGAATGCTTGTGCGCGCTCACTG
 ATGGATGAGGCAATAGAACCACAGTGCTTAGAAATAATCCTGCAGCTAT
 GCTAGCTTATGCTGTGTTACTGGGCTCTGATGGGATAGGATCCAAAGGACA
 TGGTCGCTCTCTTATAAGGACAGTCTGTTACTATTTAGTAGGTATTTA
 CAGCAACTAGTCATGGAATCACTTGGGAAGGAGTTTGACCTGGATGGTAA
 TCGGGTGAATCAGGGACTTACAGTATATGGAATAAAGGGAGCACAGATC
 AGCACGCTACATTCAACAACTAAGAGAGGGGTACACAATTTCTTTGTG
 ACGTTTCATTGAAGTGCTGCGTGATAGACCTCCTGGTCATGATTGGGAGCT
 CGAACCAAGGTGTACATGTGGTGAATCTTATTGGAATGCTACAGGGAA
 CTAGGTCTGCATTGTATGCTAATGATAGGGAGTCCATAACTGTTACAGTG
 CAAGAAGTGACACCTAGAACTGTGGGGGCCCTTGTGCACTTTATGAGCG
 AGCGGTTGGAATTTATGCATCACTGGTCAATATCAATGCTTATCATCAAC
 CCGGTGTAGAAGCTGGGAAAAAGGCTGCAGGGGAAGTATTAGCTCTTCAG
 AAGCGTGTGTTAGCAGTGCTTAATGAGGCAAGTTGTAAGAGCCTGTTGA
 ACCTTTGACACTTGAAGAAGTAGCTGAACGTTGCCATGCAGAAGAAGATA
 TTGAAATGATATACAAGATTACTGCACATGGCTGCCAATGATAGAGCA
 CTAATTGCCGAAGGTAATTGTGGTTACCACGCAGCATCAAAGTGTCTCT
 TGGGAGTGTAATGTTGAAGCA-----TCATATGCT
 >O
 GCGTCCATCTCCGGCGCGGCGGCGCTCCGTCCTCGTCGGCGGCGTGACAG
 GCTGCGGCTGCGGCGGCAAGCTCCTCCTCCGCCCCCTCCACCTCCGCCTCC
 GCGCCCCCACTCCATCGCGGACCTCTCCCGCTCCTCCTCGTCTCGTCG
 GAGCAGCAGCCCTCCTCTACCCGCGGCGGCGGCTCGCTCCAGGCCGCGG
 GCAGAACGGC-----AGCACCCGCGGCGCGGTGGAGAAGG
 ACCCGATCAAGCTGTGGGAGAGGTACGTGGAGTGGCTGTACAGCACAAAG
 GAGCTCGGCCTGTTCTGTGGACGTGAGCCGGATGGGGTTACGGAGGAGTT
 CTTCCGCGGATGGAGCCGCGG--CAGCGCGCTTCGCGGCGATGCGGG
 AGCTCGAGAAGGGGGCCATCGCAACCCGACGAGGGCCGATGGTTGGC
 CATTACTGGCTCCGCGACCCAGGCCTCGCCCCAACTCCTTCTCCGCAC
 CAAGATCGAGACCACCTCGACCGCATCTCGCTTCTCCCAAGACGTCG
 TCTCCGGCAAGATTAAACCTCCGTCATCTCAGCTGGCCGATTACTCAA
 ATACTCTCAATAGGAATTGGAGGATCAGCTTTGGGGCTCAATTTGTTTC
 AGAGGCACTTGACCTGATAACCCTCCATTGAAGATACGATTTATTGACA
 ACACTGATCCTGCTGGAATCGACCATCAAATTGCTCAACTAGGACCAGAG
 CTTGCACTACTCTTGAATTGTTATTTCTAAGAGTGGAGGCACACCTGA
 AACCCGGAATGGTCTGCTTGAAGTGCAAAAAGCCTTCAGAGATGCTGGGC
 TGGATTTCTCGAAACAGGGAGTTGCAGTTACTCAAGAAAATTTCTGTGTA
 GACAACACTGCCAGAATAGAGGGATGGTTAGCTAGATTTCCCATGTTTGA
 CTGGGTTGGTGGAGAACATCAGAAATGTGAGCTGTGCGGATTGCTTCCAG
 CTGCATTACAGGGTATCGATATCAAAGAAATGTTAGTTGGCGCAGCACTA

ATGGATGAGGAAACAGGAATACTGTGGTTAAGGAAAATCCAGCAGCCTT
ACTTGCTTTGTGCTGGTATTGGGCATCGGAAGGAATAGGAAATAAGGATA
TGGTTGTACTTCCGTACAAGGATAGTTTGTACTTTTGTAGTAGATATTTG
CAACAGCTTGTG - - - GAATCTCTTTGGGAAAGAATTTGACCTTGATGGCAA
TCGGGTTAATCAAGGTCTAACTGTGTATGGCAACAAAGGAAGCACCAGTC
AGCATGCCTACATTAGCAGCTGAGAGAAGGTGTACACAACCTCTTCGTT
ACTTTTATTGAGGTTTTGCGTGATAGGCCTCCTGGTCATGATTGGGAGCT
TGAACCTGGAGTTACATGTGGTGACTATTTGTTCCGGGATGTTGCAGGGAA
CCCGTTCTGCTCTTTATTCTAATGACCGAGAATCCATCTCCGTTACTGTG
CAAGAAGTTACTCCTAGGGCTGTTGGAGCGCTGGTTGCTCTATATGAACG
TGCTGTGGGGATTATGCTTCTTTAGTAAATATCAATGCCTATCATCAGC
CTGGTGTTGAGGCTGGGAAAAAGGCAGCAGGTGAAGTGTGGCACTTCAG
AAGAGGGTGTGACTGTTCTGAATGAGGCCAGCTGCAAGGATCCTGCCGA
GCCGTTGACGCTGGATCAAATAGCAGAGCGCTGTCTATTGCCCTGAAGATA
TTGAGATGATATACAAAATAATCCAGCACATGGCAGCAAATGATAGAGCG
CTCATAGCAGAAGGCAGCTGCGGTTCTCCTCGAAGCATCAAGGTTTACCT
TGGAGAATGCAATGTAGATGAGGATATGCTTGCTGCT

>V

GCTTCCGCTCCTCCGCATCTGCTCTTCTCATATCCGTTCAAATCCAAGCA
CTTCACCGCCAGATCATCG - - - - - CCCTCCTCTACAATCATGCCTTCTT
TTAGAATAGACTCTCTCACATTCCTCACTCGCTAACTGGACGATCGG
ACTCTGGTTCTCACTCCGTCCTGGCTCGGGAGGTATCGCGGACTTGTG
CAAAAGCGAC - - - - - CCGTCTCCAAGAAGAAGGGACTAGAGAAGG
ACCCGGGAGCTCTGTGGCGCAGATACGTCGACTGGTTGTACAGCACAAAG
GAGTTAGGGCTTTTTCTCGATGTGAGTCGGATCGGATTTTCGGAGGAGTT
CGTGGAGGAGATGGAGCCAGGTTCCAGGCGGCTTTTCGCGCGATGCAGG
AGCTGGAGAAGGGAGCCATTGCGAATCCCGACGAAGGACGGATGGTTGGT
CATTACTGGCTGAGGAGCTCTAAGCTCGCACCCCAACCGTTTCTGAAGTT
GCAGATTGAGAATACTCTTGAGGCTGTTTGAAGTTCGCCGAGGACGTCG
TCAGCGGTACAAATTAACCTCCATCTTCTCCAGAGGGTCGCTTACACAT
GTCCTTTCTGTTGGAATTGGAGGTTCCGGCTCTTGGACCTCAGTTTGTG
AGAGGCATTGGCTCCTGATAATCCTCCTCTCAAGATAAGATTCAATTGATA
ATACAGATCCAGCTGGAATTGATCACCAGATTGCCAGCTTGGCCCCGAG
CTGGCCTCTACAATAGTAATTGTCAATTTCAAAGAGTGAGGCACTCCTGA
AACTAGGAATGGTTTGTGGAAGTACAGAAGGCCTTTCGTGAAGCTGGTC
TGGATTTTGCAAAACAGGGTGTTGCCATTACACAAGAAAATTCATTATTA
GACAACACTGCTAGAAATTGAGGGTTGGTTAGCTAGATTTCTATGTTTGA
CTGGGTGGGTGGAAGGACATCTGAAATGTCTGCAGTTGGCCTACTTCCTG
CTGCACCTCAGGGGATTGATATCAGAGAAATGCTTGCTGGTGCTTCATTG
ATGGATGAGGCTAATAGAACCCTGTGGTTAGGAATAACCCTGCAGCATT
GCTAGCTTTATGTTGGTATTGGGCTCTGAGGGAGTAGGATCCAAGGATA
TGGTTATTCTTCTCTTACAAGGACAGCTTATTATTATTCACTAGGTATTTG
CAGCAACTGGTC - - - GAATCAATTGGAAGGAGTTTGACCTGGATGGAAA
TCGGGTGAATCAAGGACTTACGGTCTACGGGAATAAAGGCAGCACAGATC
AACATGCATACATTCAACAATTGAGAGAGGGTGTGCACAACCTCTTTGTG
ACATTCAATTGAAGTTCTGCGTGATAGACCCCTGGTCATGATTGGGAGCT
TGAACCAAGGTGTACATGTGGTGACTACCTCTTCGGAATGCTTCAGGGAA
CAAGGTCAAGCTCTATATGCCAAGGACCGGAATCAGTTACAGTCACTGTG
CAGGAAGTGACGGCTAGATCTGTGGGGCTATGATAGCATTGTATGAGCG
AGCAGTTGGAAATATATGCCTCACTAGTCAACATCAATGCTTACCATCAAC
CCGGTGTGGAAGCTGGGAAAAAGGCAGCAGGAGAAGTATTAGCTCTTCAG
AAGCGGGTCTGCGAGTTCTGAATGAGGCCAGCTGTAAGGAGCCCGTTGA
ACCATTGACACTTGATGAAGTAGCTGAACGCTGCCATGCGCCTGAAGATA
TCGAAATGAAATTACAAGATTATTGCTCATATGGCCGCCAATGATAGAGCT
CTTATTGCTGAAGGCAGTTGCGGTTACCCCGGAGCATCAAAGTTTTCCT
TGGGGAGTGTTATGTCGATGAC - - - - - CTGTATGCT

GLY.5.4.2.2.fasta

>A

- - - ACGTCGACCTACACGAGATTGATACCGTATTTCTCTTTTCTAGGTT
CGCCGGCGCTAAGTATTCGCCGTTATTGCCGCTCCGTCGTTTACTTTAT
CTACTTCCGGC - - - - - ATTCATATCAGAACCAAACCGAATTCTCGT - - -
- - - - - TTTCACTCGATTATAGCATCCTCCTCATCTTCATCTGT
AGTCGCCGGAAGTATGATGCAATTGAGATTAATCGTTGCCGACGAAGCCCTA
TTGAAGGACAAAGACTGGAAGTACGCGACTCCGGAAGAAAGGTGAAGTT
TTCATGGAAGATAATTATTTAGCCAAGTGGATTGAGGCTGTTTAAATTC
ATTGCCACTTGAGGATTATAAAGATGCAACGTTGGTTTTAGGAGGTGATG
GTCGTTACTTTAACAAAGAAGCTTCACAGATCATTATCAAAATTGCTGCT
GGGAATGGAGTTGGACAAATTTAGTTGGAAGGAAGGCATTTTGTCAAC
CCCAGCGGTTTCCGCTGTAATCAGGAAAAGAAAG - - - - -
- - - - - GCAAATGGAGGATTCAATTGAGTGCAAGCCATAATCCTGGTGGA
CCTGAATATGACTGGGGTATCAAGTTCAATTACAGCAGTGGTCAGCCTGC
TCCTGAAACCTTACTGACAAAATATATGAAATACACTTTCAATCTCTG

AGATTAAGGTGGCAGAGATTCTGATATTGATCTTTCTCAAGTTGGAGTG
ACCAAAATATGGGAATTTTCAGTGTAGAAGTTATAGATCCTGTTTCCGACTA
TTTGGAGCTCATGGAGGATGTTTTCGACTTTGATCTCATCAGAGGTCTTC
TTTCAAGATCAGATTTTGGGTTTCATGTTTGATGCCATGCATGCTGTTACT
GGTGCTTTATGCGAAACCAATTTTGTGACAATCTTGGAGCCAAACCGGA
TTCAATTTCCAACGGAGTTCCCTGGAAGATTTTGGGCATGGCCATCCTG
ATCCTAATTTGACATATGCAAAGGATTTGGTTGATGTCATGTATCGTGAC
AATGGACCGGATTTTCGGAGCAGCAAGTGATGGTGATGGTGACAGAAATAT
GGTTTTAGGAAACAAATTCCTTTGTCACCTCCTCAGACTCTGTAGCTATCA
TCGCAGCCAATGCACAAGAAGCAATTCATATTTCCGTGCTGGTCCTAAG
GGGCTGGCAGCTTCTATGCCGACAAGTGGTGCTCTTGATCGTGTGGCGA
GAAGTTGAAGCTTCCTTTCTTCGAGGTTCCCACTGGATGGAATTTTTTG
GAAATCTTATGGATGCTGGAAAATTGTCAATCTGTGGAGAAGAGAGCTTT
GGCAGAGTTTCTGATCACATCCGTGAGAAGGATGGAATATGGGCTGTGTT
GGCGTGGCTTTCGATCCTTGCTCATCGCAACAAGGACAGAAACCAGGGG
ACAAGTTGGTGCTGTTGCAGATGTTGTGAAAGAGTATTGGGCGACATAC
GGAAGGAATTTTTTTCAAGATACGACTATGAAGAATGTGAATCTGAAGG
AGCCAATAAAATGATCGAGTATCTTAGAGAAATCTTGCTAAGAGCAAGG
CCGGTGACGTTTATGGAACTATGTCCTCCAGTTTGCAGATGACTTCTCA
TACACAGACCCTGTTGATGGAAGTGTTCATCCAAGCAAGGTGTTCCGTT
TGTGTTACGGATGGATCCAGAATCATATTTGTTTTATCCGGAAGTGGCT
CAGCTGGAGCTACAGTGCGAATATACATTGAACAATTTGAGCCAGACGTT
TCAAAACATGATGTGGATGCACAAATAGCTCTTAAACCTTTAATCGATCT
TGCAATTGCTGTATCAAAGCTGAAGGACTTCACAGGGAGGGAGAAGCCCA
CTGTCATTACA

>P

-----GCAATTAATTCAGTGCCGACAAAGCCAA
TTGAGGGTCAGAAAGACAGGGACAAAGTGGTCTGCGAAAGAAAGGTTAAGGTT
TTCCTGAGGAAAACTACCTTGCAAAATGGATTTCAGGCATTGTTTAATTC
ATTGCGACACAGAGGATTACAAGAATGGAGTGCTGGTCTTGGGAGGGGACG
GTCGTTACTTCAATCGAGAAGCTTCACAGACAATCATCAAAATTGCTGCA
GGGAATGGTGTTGGCAAAATCCTTGTCGGCAAGGAAGGTATTATGTCCAC
GCCAGCTGTTTCTGCC-----
-----GCCAACGGTGGATTTATAATGAGTGCAAGTCATAATCCTGGAGGG
CCTGAATATGATTGGGGTATCAAGTTTAATTACAGCAGTGGGCAACCTGC
ACCTGAATCCATCACTGACAAGATATATGAAAACACCCCTTTCTATATCTG
AGATTAAGATGGCGGACATTCTGATGTCGACCTTTCTAGCCTCGGAGTC
ACAGAATATGAAAACCTTCATTGTGCAAGTAGTAGACCCAGTTTCGGATTA
TTTGAAGTAATGGAGAATGTATTTGATCTTGAGCTCATCAAAAGTCTCC
TTTCACGGTCAGATTTTCAGGTTTGTATTGATGCCATGCATGCTGTTACT
GGTGCTTACGCAAAACCATCTTTGTGGACAAGCTTGGGGCCAGTCCGGA
TTCAATTTCAAATGGGGTACCTCTGGAGGATTTTGGTCACGGCCATCCAG
ATCCTAATCTTACATATGCCAAGGATTTGGTCAACATCATGTATGGTGAG
AATGGACAGATTTTGGAGCTGCAAGTGATGGAGATGGGGACAGAAAC--
-ATCCTAGGTAGAGGGTTTTTGTCACTCCTTCAGATTCCGTTGCTATCA
TTGCTGCCAATGACAGGAAGCCATTCCATACTTTATGAGTGCCCTAAG
GGTTTAGCTCGTCCATGCCGACAAGTGGTGCTCTAGATCGTGTGGCTGA
AAAATTGAATCTTCCATTTTTTGGAGTCCCTACTGGTTGGAATTTTTTG
GGAATCTT--GATGCTGGAAAAGTGTCAATTTGTTGGTGAAGAAAGTTTT
GGAAGTGGTTCTGATCACATTCTGTGAGAAGGATGGCATATGGGCTGTTTT
AGCTTGGCTTTCAATTATTGCTTACAGAAACAAAGACAAGAAACCAGGGG
AGAAGTTAGTATCTGTTGCTGATGTTGCGAAAGAGCATTGGGAAACCTTT
GGAAGAAATTTCTTTCTAGATATGACTATGAAGAATGTGAATCTGAAGG
TGCCAACAAAATGATCCAGAATCTTAGAGATATAGTCTCTAAAAGCAAAC
CAGGTGATAAGTATGGAATTTACACCCTTCAGTTTGTGATGACTTTACG
TACACT--CCTGTGGATGGAAGTGTGGTATCAAAGCAAGGTGTCCGGTT
TGTTTTACCCGATGGATCAAGGATTATATCCGTTTATCTGGAAGTGGAT
CAGCAGGTGCAACTGTGAGAATCTACATCGAACAGTATGAACCAGATGTC
TCTAAACATGAGATGGATGCCAAGTAGCCCTAAAACCACTAATAGATCT
GGCATTGCTGTATCAAAGTTGAAAGACTTCACAGGAAGGGACAAGCCAA
CAGTCATCACA

>R

GCGTCCTCTTCTGTTTAAAGATAGAAACCATCTTCTCTTCCACATTCTC
TAAGCTTAAGCAATCTAATCCAAATCGAAATGCCACTAATATCATTTCTT
CTTCTAATTCTTCTCGTCTTTTGTATCCAGAACTTCTTATTTAAGAACA
ACCTCACCTTCTTTTCTGCTCCCTCTCTATCAAAGCTTCTCTTTCGACTGC
CATCGCTGAACCAAGAGGATCAAGATTAATTCACCTTCCAACCTAAGCCTA
TTGAAGGTGAGAAGACAGGAAGTGTGGTCTTCGAAAGAAGGTTAAAGTT
TTCATGCAAGAAATACCTAGCAAAATGGATCCAGGCATTATTTAATTC
ATTACGCCAGAGGATTACAAGAATGGGTTATTGGTCTTAGGAGGAGATG
GTCGTTATTTCAACAAAGAAGCTGCACAGATAATCATCAAAATTCGGGCT

GGAAATGGTGTTGGTAAAATCTTAGTGGGCAAGGAAGGTATTATATCTAC
ACCAGCTGTTTCTGCTATAATCAGAAAGCACAGG-----
-----GCGAATGGTGGATTATAATGAGTGCAAGTCATAATCCTGGAGGC
CCTGAATATGATTGGGGTATCAAGTTTAAATTATAGCAGTGGGCAACCTGC
ACCTGAATCAATCACCGACAAGATCTATGGGAATACCCTTTCTATCTCTG
AGATTAAATGGCAGACATTCTGATGTTGATCTTTCTCTCTGGAGTT
GCAAAACATGGAACCTTCATTATTGAAGTAGTGGACCCAGTTTCTGATTA
TTTGGAGCTAATGGAGAATGTATTTGATTTTGAACATCAGAAAGTCTTC
TTTCACGTTAGATTTCAGGTTTATATTGATGCCATGCATGCTGTTACT
GGTGCTTATGCAAAGCCCATCTTTGTGGACAAGCTGGGGGCTAGTCCGAA
TTCAATTTTAAATGGAAGTCCTTTGGAGGATTTTGGGCATGGTCATCCTG
ATCCTAATCTTACATATGCCAAGGATTTGGTGAATATT--TATGGTGAG
AATGGGCTGATTGTTGGCGCTGCAAGTGATGGAGATGGTGACAGAAAT--
-ATCCTAGGTAGAGGGTTTTTGTACTCCTTCAGATTCTGTTGCTATCA
TTGCTGCCAATTGCGAGGGAGCAATCCATACTTCAAGAATGGCCCTAAG
GGTTTAGCTCGGTCCATGCCAACAAAGTGGTGCACTGGATCGTGTGGCTGA
AAAAATTAATCTACCTTTTTTTTGGAGTCCCACTGGCTGGAATTTTTT
GAAATCTTATGGATGCTGGAAAGTTGTCAATTTGTGGGAAGAAAGTTTT
GGAAGTGGTTCTGATCACATTCTGTGAGAAGGATGGCATATGGGCAGTGTT
AGCTTGGCTTTCAATTATTGCTCACAGAAATAGGGACAAGAAACAGGGG
AGAAGTTAATATCTGTTTCTGATATTGTGAAAGAATACTGGGCAACATAT
GGAAGAAATTTCTTTTCTAGATACGACTATGAAGAATGCAAGTCTGAAGG
AGCAATAAAATGATAGACTATCTTAGAGATTTGATTTCTAAGAGCAAAC
CGGGTGAAGATGGAAGTTATGTCCTTCAGTTTGCTGATGACTTCAAG
TACACTGATCCTGTGCGATGGAAGTGTGGTCTCAAAGCAAGGGGTCCGGTT
TGTTTTTCTGATGGGTCAAGGATCATTTTCCGGTTATCTGGAAGTGGAT
CTGCAGGTGCAACTGTTAGAATGTATATCGAACAGTTTGAACCTGATGTG
TCTAAACACGAGATGGATGCCCAAACAGCCTTAAACCGTTAATAGACCT
GGCATTGTCTGTATCAAAGTTGGAGGACTTACAGGAAGGGAGAAGCCTA
CAGTCATCACA
>O
---GCCTCGCACGCGCTCCGCCTCCACCCGCTGCTCTTCTCCGCCGCCGC
CGCGCGCCCGGCTCCGCTCGCGGCGCGGCCCGGTGGTGGT-----
-----GCCCGCCGGGTCCACCGC---
-----CGCCACTCTCTCGCCGTCGTCCGGTGCTCCTCCTCCGC
C-----GCCCAGGCGCTCAAGATCAAGTCGATTCGACCAAGCCCG
TTGAGGGGCGAGAAGACCGGGACCAAGTGGGTTGAGGAAGAAGGTGAAAGTG
TTCCAGCAGGAGAATTACCTCGCTAATTGGATTCAAGGCTCTGTTCAATT
ATTGCCCGCGGAGGATTATGTTGGTGGAAACCCTTGCTGTTGGTGGTGATG
GCCGATACCTTTAAACAAGGATGCTGCTCAGATTATCACTAAAATTGCAAGCT
GGGAATGGTGTTGGGAAGATCCTAGTTGGCAGGAACGGTCTGCTGTCAAC
GCCTGCTGTATCTGCAGTAATTCTGTAAGACAA-----
-----GCCAATGGTGGCTTCATCATGAGTGCAAGCCATAATCCAGGTGGG
CCAGATAATGATTGGGGTATCAAGTTCAACTATAGCAGTGGGCAGCCAGC
ACCAGAGACAATTACCGACCAATATATGGAACACACTTTTCGATTTCTG
AAATAAAACGGCAGATATTCTGATGTTGATTTGTCTCTCTAGAGTT
GTAAGCTATGGTGATTTACCGTTGAAGTGATAGACCCTGTCTTGGACTA
CCTTGAGCTAATGGAGAATGTGTTTGACTTCCAATTATCAAGGGCTTGT
TGCTCTCGCCAGATTTTCAGGTTTGATTTGATGCCATGCATGCTGTGACT
GGTGATATGCGGATCCTATTTTGTGAGAACTTGGAGCTGATCCGGA
CTATATATTAATGGTGTTCCACTTGAAGATTTTGGCAATGGTCACCCTG
ATCCTAATTTAACTTATGCCAAAGAGCTTGTTTACCATGTTTGGAAGC
GGAGCACCTGACTTTGGTGCAAGTATGGTGATGGTGATCGAAACAT
GATTCTTGGAAGAAGGTTCTTTGTTACCATCAGACTCTGTTGCAATAA
TTGCAGCGAATGCACAGGCAGCAATTCCTTATTTCCAATCTGGTCCAAAA
GGTCTTGCTAGATCAATGCCAACGAGTGGTGCTTGTATCGTGTAGCTGA
TAAATTGAATGTTCCGTTCTTTGAGGTACCAACAGGATGGAATTTTTTG
GAAACCTAATGGATGCAGGTAATTGTCTATATGTGGAGAGGAAGTTTT
GGGACAGGATCTGATCACATCAGGGAGAAGGATGGCATATGGGCTGTTCT
AGCTTGGCTGTCCATACTTGACACCGGAACAAGGATAAGAAGGCCGGGG
AGAGATTAGTGTGAGTGAAGATGTAGCTAGGGAACACTGGGCAACCTAT
GGAAGGAATTTCTTCCAGATATGATTATGAGGAGTGTGAATCTGAGAG
TGCAATAAGATGATGGAGCATCTTAGAGATGTGATCGCAAAAAGCAAGC
CTGGAGAGAAATATGGAACATATACCCTTCAGTTTGCCGATGATTTCAAGT
TACACTGATCCGGTGGATGGTAGCACTGTATCTAAACAAGGGCTTCGATT
TGATTACCCGATGGATCTAGGATTATCTCCGCTTTTGGGAACCGGAT
CTGCTGGAGCAACAATCCGTATATACATTGAGCAATTCGAGTCTGATGCC
TCAAAGCATGATCTGGATGCACAAATAGCTTTGAAGCCTTTAATAGACCT
AGCTCTATCTGTTTCAAAGTTGAAGGACTTCACTGGAAGAGATAAGCCTA
CTGTCATAACA
>V
-----GATAGTGTCTTCTTTTCAACCCTCAA
ACTCACTCCACTACCAATGCCTCTTTCTCCAATCGCCACTTCGCTTCTC
TTTCT-----CTTCTCTCTCGCAAGTTCCTATCAGGAGA---
-----CTCTCCGTTAAAGCCTCTTCTCTGCGCCTTCCACCTC

CACTGCCAACTCTCAGACCATCAAGATCATATCGATGCCACAAAGCCGA
 TTGAAGGCCAAAAGACTGGAACCGCGGTCTACGAAAGAAGGTTAAAGTT
 TTTATAGAAAGAAAATTACCTTGCGAATTGGATTCAAGCGCTGTTCAATTC
 ATTGCCACCTGAGGATTACAAGGATGGGGTGTGGTTTTAGGCGGTGATG
 GCCGATATTTCAACCGGGAAGCTGCACAGATAATTATCAAAATTGCAGCA
 GGAATGGCGTTGGAATAATAGTTGGCAAGGAAGGCATCATGTCAAC
 TCCAGCTGTTTCTGCTGTAATCCGGAAGCAAAAGGCAAGCATATTCCTC
 TCCTTGCCAATGGAGGTTTTATAATGAGTGCAAGCCATAATCCTGGTGGG
 CCTGAATATGATTGGGGCATCAAGTTTTAATTACAGCAGTGGACAACCTGC
 ACCTGAAACCATCACAGACAAGATTTATGGGAACACTCTTTCGATCTCTG
 AAATTAAGGTA - - - GACATCCCAGACATTGATCTTCTTGTCTTGGAGTC
 ACAACATATGGGAATTTTCAGTGTTGAAGTTGTAGATCCAGTTTCTGATTA
 TTTGGAGCTACTGGAGAATGTATTTGACTTCCAGCTCATTAAAGGTCTTC
 TTTGCGGTGTCAGATTTTCAGGTTTACGTTTGTATGCCATGCATGCAGTTACA
 GGCGCTATGCAAACCCATCTTGTGGACAAGCTAGGAGCCAGTCCAGA
 TTCAATTTCAAATGGAGTACCTCTGGAAGATTTTGGGCATGGTCATCCAG
 ATCCTAATCTTACGTATGCTAAGGATTTGGTGAATATTATGTATGGTGT
 AATGGACCTGATTTTGGAGCCGCAAGTGATGGAGATGGTGATAGAAACAT
 GATTCTAGGAAGGAGTTTTTTTTGTACACCTTCAGACTCCGTTGCTATTA
 TTGCTGCAAAATGCTCAAGAAAGCCATTCCATATTTTCAGAGTGGCCCCAAG
 GGTTAGCTCGATCTATGCCAACGAGTGGTCTTGTATCGTGTGCTGA
 AAAATTGACTCTTCCATTCTTTGAGGTCCCACTGGATGGAAATTTTTTG
 GGAACCTT - - - GATGCTGGAAAGTTGTCTATTTGTGGGAAGAAAGTTT
 GGAACAGGTTCTGATCACATTCGTGAGAAGGATGGCATATGGGCTGTTCT
 AGCTTGGCTTCAATCATTGCATATAGAAACAAGGACAAGAAACCTGGAG
 AGAAATTGGCATCTGTTGCTGATGTCGTAAGGAGCACTGGGCAACATAT
 GGAAGGAACCTCTTTCTAGATATGACTATGAAGAATGTGAATCTGAAGG
 TGCCAATAAATGATAGAATATCTTAGAGATTTGGTCTCTAAGAGCAAGG
 CAGGTGAAAAGTATGGAATATTGCACTTCAATTTGCGGATGACTTCTCA
 TACACTGATCCTGTGGATGGAAGTGTGGCATCAAAGCAAGGTGTTGATT
 TGTTTTACGGATGGATCAAGGATCATATATCGTTTATCAGGAACCGGTT
 CTGCAGGTGCAACTGTTAGAATTTATATTGAACAGTTTGAGCCAGATGTC
 TCTAAACATGAAATGGATGCCCAAGCAGCATTGAAGCCATTGATAGATT
 GGCGTTATCTTTGTCGAAGCTCAAGGACTTCACAGGAAGGGAGAAACCTA
 CAGTGATCAC

LUT.LUT1.fasta

>A
 ---GAGTCTTCACTCTTTTCTCCATCTTCTCTTCTTACTCTTCTCTCTT
 CACTGCAAAA - - - CCTACGCGTCTTTTATCACCAAAACCCAAATTCACAT
 TC - - - TCCATCAGATCCTCCATTGAGAAACCCAAACCC - - - - -
 - - - - - AAACCTCGAGACCAATTCATCGAAATCCCAATCATGGGTCAATCC
 CGATTGGCTCACAACTCACTCGTACCCTTTCC - - - - - TCAGGA - - - A
 AAAACGACGAGTCAGGTATACCAATCGCGAACGCGAAGCTCGACGATGTC
 GCTGATCTCCTCGGAGGTGCTCTCTTCTTACCTCTCTACAAATGGATGAA
 G - - - TACGGACCCATTTACCGTCTCGCTGCTGGTCTCGTAATTTCTGTA
 TTGTGAGCGACCCAGCGATAGCTAAACATGTTTTGAGGAATTTATCCA - - -
 - - - AAGTACGCTAAAGGCTTAGTCGCTGAAGTCTCTGAATTTCTATTTGG
 TTCGGGTTTCGCTATCGCTGAAGGACCTTTTGGACAGCGAGGCGTAGAG
 CGGTGGTTCCATCGCTTACAGGAGGTATTTGTCTGTGATTGTGGAGAGA
 GTATTCTGCAATGTGCAGAGAGGCTTGTGAGAAGTTGCAGCCTTATGC
 AGAAGACGGAAGTGTGTAATATGGAAGCGAAGTTCTCTCAGATGACAC
 TTGATGTCATTGGGTTGTCTCTTTTTAACTACAATTTGATTCTTTGACT
 ACTGATAGTCTGTGATTGAAGCTGTTTACACTGCTCTTAAAGAAGCTGA
 GCTTCGTTCTACTGATCTTCTGCCATATTGGAAG - - - - -
 - - - - - ATCGATGCATTGTGTAAGATAGTCCCGAGA
 CAGGTGAAAGCTGAAAAGGCTGTAACCTTTGATAAGGGAACTGTTGAAGA
 CCTTATTGCTAAGTGTAAGAAATTGTGAAAGAGAAGGCGAAAGAATCA
 ATGATGAGGAGTATGTAATGATGCTGACCCAAGTATCCTGCGTTTCTTG
 CTTGCAAGCAGAAGAGGTATCAAGTGTGCAATTACGGGATGATCTTCT
 CTAATGCTCGTAGCGGGTCATGAAACCACTGGATCTGTCTCACTTGGGA
 CACTTTATCTCCTAAGTAAGAACTCATCTGCATTAAGGAAAGCACAAAGAA
 GAAGTAGACAGAGTGTTAGAAGGAAGAAACCCGGCTTTCGAGGATATAAA
 GGAGTTGAAGTACATCACTCGTTGTATAAACGAGTCAATGCGTCTCTATC
 CTCATCCTCCTGTCTTGATAAGAAGAGCTCAAGTTCTGACATTCTTCTCT
 GGGAACTATAAGGTCAATACCGGACAAGACATTATGATTTCACTCTATAA
 CATCCATCGTTCTTCCGAGGTATGGGAAAAAGCTGAGGAATTTCTGCGCTG
 AACGATTGCAATAGATGGCGCAATCCCTAACGAAACAAACACTGATTTT
 AAATTCATCCCATTCAGTGGAGGGCTAGAAAAATGTGTAGGCGATCAGTT
 TGCATTGATGGAGGCAATTGTGGCACTCGCGGTGTTTCTTACGCGGTAA
 ACGTTGAGCTGGTTCTGATCAGACCATTAGCATGACCACAGGAGCAACC
 ATACACACCACCAATGGATTGTATGAAGGTGAGCCAAAGG - - - - -
 - - - - -

```

>P
-----TCACTCACCTTCTCCGTCTCCTCTCTTCATCCTCTCTACAAGTA
CAAACCCATC---TCCACCACCACTCTCCCTCCAAAACCTCGG-----T
TTCTGTCCATCAAACTCTCTTTAGACGACGAAACC-----
-----CAAAAACAAACCTCTTGGGTCAGCCC
AGACTGGCTAACGTCCCTCACTCGCTCCTTATCC-----TTAGGC---C
GAAACGATGATTGAGGGATTCCCATTTGCCAGTGCTAAGCTTGAAGATGTG
TCTGATCTTTTAGGTGGTGCTTTGTTTCTCCCTTTGTTTAAATGGATGAA
TGATTATGGGCCCATTTACCGTCTCGCTGCTGGCCCTAGAAATTTTCGTTA
TAGTTAGC---CCTGCTATTGCTAAACATGTTTTAAAAAATTATGGTACT
---AAGTATGCCAAAGGCCTTGTTGCTGAAGTCTCTGAGTTTCTCTCGG
CTCCGGTTTTGCTATCGCCGAGGGCCCACTTTGGACTGCAAGGCGCAGGG
CAGTGGTGCCCTTCACTGCACAGGAAGTATTTGTCAGTGATAGTTGAAAGG
GTGTTTTGCAAAATGTGCAGAAAGGTTGGTGGAGAAGTTACAAGCTGATGC
ATTGAATGGCAATGCTGTAATATGGAAGAGAAGTTTTCTCAATTGACAC
TCGATGTCATTGGTCTTTCAGTATTTCACTACAATTTTGATTCAATTAAC
ACTGATAGTCTGCTTATTGATGCTGTTTACACTGCATTGAAAGAGGCAGA
GGCTCGTGCCACTGACCTTTTGCCATATTGGAAGGCATGCAAAATCCATGA
GTTTCTTCATTATCTCCGTATTGATGCTTTGTGAAGATTATACCGAGG
CAAATAAAGCGGCAAGGCGGTTATGGTTATTAGGCAAACTGTTGAGGA
ACTTATTGAGAAATGCAAAAAGATTGTGGAATTTGAGGGTGAGAAAATAA
ATGAGGAGGAATATGTGAACGACAATGATCCGAGCATTCTCCGGTTTTTG
CTCGCCAGCAGGGAAGAGGTTTCAAGTGTACAGTTACGAGATGACCTACT
GTCAATGTTGGTTGCTGGCCATGAGACAACCTGGTTCAGTATTGACTTGGA
CTTTGTATCTCTGAGTAAGGATCTTCTACATTGATGAAAGCACAAGAA
GAGGTTGACCGAGTACTACAGGGAAGGCCTCCTACCTATGATGATATAAA
GGATCTAAAGTATTTGACACGCTGCATAAATGAGTCCCTACGTCTCTACC
CACATCCTCCGGTCTTGATAAGAAGAGCTCAAGTTGCTGATGTGCTTCCT
GGTAATTATAAGGTTAATGCTGGACAAGATATAATGATTTCAGTGTACAA
TATTCATCACTCCTCCAAGGCTGGGAGAGAGCAGAAGAGTTTGTGCCAG
AGAGATTTGGCTTGGATGGCCAGTGCCTAATGAATCAAACACAGATTTT
AGGTTCACTTCTTTCAGTGGAGGACCCGTAATGCGTTGGTGACCAGTT
TGCTTTACTTGAAGCCATTGTTGCACTTGCAATCTTTCTACAGAAAATTA
ACTTTGAGCTGGTTCTGATCAGAACATCCGCATGACTACTGGAGCAACA
ATACACACTACAGATGGCCTGTACATGAAACTCAGCCAAAGAGACACT--
-----
>R
-----CCTACTCTTATTCTTTCTGATCTTTCTCTCATCTTTCACTCCC
TCTCATCAAG---CCCGCCACAATTCTCACTTCAAAGCCTCAT-----T
TCATCTCAATCAAATCATCACTGGACAACAAAACACCA-----
-----AAACCCAATTCCAATAATACAAAACCCAGTTCTTGGGTAAGCCC
AGATTGGCTAACTTCACTCACTAAAACCTCACA-----TTAAACCAAAA
AACATGATGACTCTGGTATTCCTATCGCTAGTGCTAAGCTTGAGGATGTG
TCTGATCTTTTGGGTGGTGCTTTGTTTCTTCTTTGTTTAAATGGATGAA
TGAGTATGGTCTATCTATAGGCTTGCTGCTGGTCTAGAAATTTTGTG
TTGTTAGTGACCCCTGCTATTGCTAAACATGTGCTTAGAAATTTATGGTAT
GGAAAATATGCTAAAGGCTTGTGCTGAAGTTTCTGAGTTTTTGTGTTGG
GTCTGGTTTTGCTATTGCTGAGGGATCCCTGTGGACGGCAAGGCGCAGGG
CTGTTGTTCCATCTCTTCATAAGAAGTATCTGTCAAGTTATAGTTGATAGA
GTATTTTGGCAAGTGCTGCTCAGAGATTAGTGGAAGAGCTTCAACCTGATGT
GCTTAATGGCACTGCAGTAAATATGGAAGAAAATTTCTCAATTAACCTC
TTGATGTTATTGGTCTTTCAGTATTTAACTACAATTTTGATTCCCTCACC
ACTGATAGCCCTGTTATTGGTGCTGTTTATACTGCATTGAAAGAGGCAGA
GGCTCGTTCTACAGATCTTTTGCCATATTGGAAG-----
-----GTCAAAGCTTTACGTAAGATAATTCTAGA
CAAATAAAGAGTGAAAAAGCAGTTACAGTAATAAGGCAACAGTCGAAGA
ACTTATTGTAAATGCAAGAGATTGTGGACACTGAGGATGAAAGAATAG
ATGATGAGGAATATGTAATGACACTGATCCAAGCATCCTTCGCTTCTTG
CTTGCCAGCAGGGAAGAGGTTTCAAGCGTACAATTACGTGATGACCTGTT
ATCGATGTTAGTTGCTGGCCATGAGACAACCTGGTTCCGTTGTTGACTTGGA
CATTGTATCTTCTGAGTAAGGATCCTGTCTCACTACTGAAGGCACAAGAA
GAGGTTGACCGTGATTACAAGGAAAGCCTCCATCATATGAAGATATAAA
GGATCTGAAGTTTCTGACACGCTGCATAAATGAGTCACTGCGTCTATATC
CACATCCTCCTGTCTTGTGAGAAGAGCTCAGGTTGCTGATGTACTCCCA
GGAAATTAAGAAGTTAATGCTGGTCAAGACATAATGATTTCTGTATACAA
TATCCATCATTTCTCGAAGGCTGGGATAGAGCAGAAGAGTTTGTGCCAG
AAAGATTTGACTTGGAAAGGCCAGTGCCTAATGAAACAAATACAGATTTT
AGGTTTATCTCTTTTAGTGGAGGACCCGTAATGTGTTGGTGATCAATT
TGCTTTACTCGAAGCCATTGTTGCCCTCGCGATCTTTCTACAGAATATGA
ACTTTGAACTGGTTCCTGATCAAAAGATTAGC---ACTACTGGGGCAACA
ATACATACAACAAATGGCTTGTACATGAAACTCGGGCAAAAGGAAAAAAT
CCAGGCA-----TTTGCTTTGTCTTCTGCAAGG
>O
GCCGCCGCCGCCGCCGCCGCCGTCCCGTGCGTACCATTCTGTGCCGCC
T-----CCTCCGCCATTGGTCTCGCCGCGTCTCCGCGTGCGC

```

ACGTCCGCCTCCGCCTGCGGCCGCCAAGGAGCAGCGGCGGTGGAGGCGGA
 GCGGAGCGGGGGAGACGAGCCGCCCATCACACCTCGTGGGTGAGCCC
 CGACTGGCTCACGGCGCTCTCCGCTCGGTGGCAACCCGCTCGGC - - - G
 GGGGCGAGCTCGGGGATCCCGTCGCTCCGCCAAGCTCGACGACGTG
 CGGGACCTCCTCGGCGGCGCGCTCTTCTCCCTCTCTTCAAGTGGTTCCG
 CGAGGAAGGCCCGTCTACCGCTCGCGGCGGGGCGCGGGATCTCGTCG
 TCGTCAGCGATCCCGCGCTTGCAGGCACGTGCTGCGTGGGTACGGTTCG
 ---AGGTACGAGAAGGGGCTCGTCGCCGAGGTTTCCGAGTTCCTCTTCGG
 CTCGGGTTCCGCATCGCCGAGGGCGCTCTCTGGACGGTGAGACGTGAT
 CAGTTGTACCATCTCTACACAAACGATTTCTCTCGGTGATGGTTGACAGA
 GTTTTTGTAAATGTGCTGAGAGATTAGTGGAGAAGCTTGAGACATCTGC
 TTTAAGTGGCAAACCTGTAATATGGAAGCAAGGTTCTCTCAAATGACTT
 TAGATGTGATTGGTTTGTCTTGTCAATTACAATTTTATTCCCTCACA
 TCAGATAGCCCTGTTATTGATGCTGTTTACACTGCACTCAAGGAAGCAGA
 ACTTCGTCTACAGATCTTTTACCATACTGGAAG-----
 -----ATTGATTGCTGTGCAAGATTGTTCTAGA
 CAAATAAAGCAGAAAAGGCGAGTTAACATCATCAGGAATACCGTTGAGGA
 CCTAATTACCAAATGCAAGAAGATTGTAGATGCTGAGAATGAACAAATTG
 AGGGTGAGGAATATGTAAATGAGGCAGACCCTAGCATCTCGGATTCCTA
 CTTGCTAGCCGTGAAGAGGTAACCAAGTGTGACGTTACGTGATGATCTATT
 GTCAATGTTAGTTGCTGGTCATGAAACAACAGGCTCTGTACTGACGTGGA
 CTATTTATCTTCTCAGTAAGGATCCAGCAGCGCTGAGGAGAGCTCAAGCA
 GAGGTTGACCGTGTCTACAAGGTAGACTCCCAGATATGAAGATCTAAA
 AGAGCTGAAGTACTTGATGCGCTGTATAAATGAGTCTATGCGGCTTTATC
 CACACCCACCTGTGTTGATACGGCGAGCCATAGTTGATGATGTGCTTCCC
 GGAACATAAAGATCAAAGCTGGTCAAGATATTATGATTTTCAAGTGTACAA
 TATACACAGGTCACCTGAGGTTTGGGACAGAGCTGATGATTTTATTCTG
 AGAGATTTGATTTAGAGGGACCTGTTCCAAATGAGACAAACACTGAATAC
 AGATTTATCCCATTCAGTGGAGGTCTCGGAAATGTGTTGGAGATCAGTT
 TGCTCTCTTGGAAAGCAATTGTGCACTTGCTGTTGTGTTGCAAGAGATGG
 ACATTGAGCTTGTGCCAGATCAAAAAATTAACATGACTACTGGGGCCACA
 ATTCATACAACCAATGGCTGTATATGAATGTAAGTCTGCGTAAAGTTGA
 CAGGAACCTGATTTTGCACTCAGTGGGTCCAGA
 >V
 -----CCTAGTTTGATGTCTCTCTCTTTTTCGCACTCCTCACACCCCC
 TACCTTCAACTTCAAACCCACCACACTCCCTCCTAATCCCTCACTTCCCT
 ATCTCTCCATAACATGTTCTCAGACAACAACCT-----
 -----AAACCAACCAGAAGCAGCTCCTGGGTGAGCCC
 TGACTGGCTGACCTCTCTGAGCCGGTCTCTCACC-----ATGGGC-----
 --CAAGACGACTCCGGTATACCCATGGCCACTGCTAAACTGGATGACGTG
 TCTGAGCTCCTGGGTGGTGCCCTCTTCTTCTTTGTTCAAGTGGATGAA
 TATGTACGGACCCATTATACCGGCTCGCCGCGGGGCGCCGAATTTTGTGG
 TTGTGAGTGACCCCGCTATTGCTAAGCATGTTCTGAGGAATTATGGC---
 ---AAATACGCCAAGGGTCTCGTGGCCGAGGTTTCGGAGTTTTTGTTCGG
 ATCGGGTTTTGCCATTGCGGAAGGCGAGCTTTGGACGGTGAGGCGCAGGG
 CTGTGGTTCTTCTCTTCAACGCGATACTTATCCATAATTGTTGATCGG
 GTGTTTTGCAAATGTGCTGAGAGATTGGTAGAGAATCTTAGAACTGATGC
 ACTCAATGGCAGTGCTGTAAACATGGAGGAGAAGTTTTCTCAATTGACTC
 TGGATGTCATCGGTCTATCTGTGTTCAACTACAATTTTGATTCACTTACT
 GCTGATAGTCCTGTTATTGATGCAAGTTTACACTGCATTGAAAGAGGCTGA
 GGCTCGATCAACTGACCTTTTACCATACTGGAAGGCAATTCCTAACTCAC
 ACCCT-----TTCTTACTTTTTTCTTTATAATCCCTAGA
 CAAATAAAGCTGAAGAATCAGTAACAGTGATAAGGAAAAGTGTAGAAGA
 ACTTATTGCTAAGTGCAAAGAAATTGTGAAAGAGAGGGGTGAGAGGATTG
 ATGAGGATGAATATGTTAATGACTCAGATCCAAGCATCCTTCGTTTCTTG
 CTTGCTAGCAGAGAAGAGGTTTCAAGTGTGCAATTACGAGATGACCTTTT
 ATCAATGTTGGTTGCTGGGCATGAGACCACTGGTTCAAGTGTGACTTGGA
 CATTATATCTTTAAGTAAGAACTCTTCTTCTGATGAAAGCACAAGAA
 GAGGTCGACAGGGTTTTACAGGGAAGGTCTCCCTCGTATGAAGATATAAA
 GGATCTCAAGTTCTTAACACGTTGCATAAATGAGTCCATGCGTCTATACC
 CTCATCCTCCTGTTTTGATAAGAAGAGCTAAAGTCGCTGATGTGCTTCT
 GGAATTTACAAGGTCAATGCTGGTCAAGATATAATGATTTCAAGTATATA
 TATCCATCACTCCTCTCAGGTCTGGGAAAGAGCAGAAGATTTTCTGCCGG
 AAAGATTGACTTGAAGGTCCTGTGCCTAATGAAACAAATACAGATTTC
 AGATTCAATTCATTTAGTGGGGGCCACGTAAGTGTGTTGGTGATCAGTT
 TGCTTTGCTGGAAGCGATTGTTGCACCTTGAATCTTTGTGACGACATGA
 ATTTTGAAGTGGTTCTGATCAGAATATTAGCATGACTACCGGGGCAACT
 ATTCATACAACAAATGGCTTGATGAAATTGACTCAAAGGCAAAATTC
 ATCTGCA-----TTTGTTCACATCTTCTAGG

pre-BR.1.14.99.7.fasta

>A

-----GAGTCACAATTATGGAATTGGATCTTACCTCTT
TTGATCTCTTCTCTCCTCATCTCCTTCGTCGCTTTCTATGGATTCTTCGT
CAAACCGAAGCGGAACGGT-----CTCCGTCACGATCGGAAAA
CTGTTTCTACCGTCACCTCCGACGTCGGATCTGTTAATATTACCGGAGAT
ACTGTC-----GCTGATGTCATTGTTGTTGGAGCTGGTGTGCTGG
TTCTGCTCTTGCTTATACTCTTGGAAGGATAAACGCCGAGTTCATGTGA
TTGAAAGAGATTTATCGGAGCCTGATCGTATTGTTGGGAGTTGTTACAG
CCTGGGGGTTACCTCAAGTTACTGGAGTTGGGAATTGAAGATTGTGTGGA
AGAAATAGATGCTCAGCGTGTGTATGGTTATGCACTTTTTAAAAATGGGA
AACGCATTTCGCTTAGCTTATCCTTTGGAGAAGTTTACGAAGATGTATCT
GGAAGGAGCTTTCACAATGGACGTTTTATTCAAAGAATGCGGGAGAAAAGC
TGCTTCACTTCCCAATGTTTCAGCTAGAGCAAGGAACAGTTCTTTCTCTTC
TAGAAGAGAATGGGACTATCAAAGGTGTGAGATATAAGAATAAAGCAGGA
GAGGAACAAACCGCATTTCAGCTTTGACTATAGTTTGTGATGGTTGTTT
CTCAAACCTGCGTCGCTCTTTGTGCAATCCTCAGGTCGAGGTGCCTTCTT
GTTTTGTGCGGTTGGTCCTAGAGAACTGCAATCTCCCATATGCAAACCAT
GGACATGTCGTCTTAGCAGATCCATCACCATTGATGTATCCAATTAG
TAGCACAGAGGTGCGGTGCCTAGTTGATGTTCCCGGCCAGAAAGTGCCGT
CCATTGCAAATGGTGAAATGAAAACTATTTGAAGACTGTTGTGGCTCCT
CAGATGCTCATGAGGTCTATGACTCTTTCATTGCTGCGGTTGATAAAGG
AAATATTAAGTCCATGCCAAACAGAAGCATGCCAGCTTCTCCTTATCCTA
CTCCAGGGGCTCTGTTAATGGGAGATGCATTTAACATGCGTCACTCCTTG
ACGGGTGGAGGAATGACGGTTGCATTAGCTGACATTGTTGCTCTGCGTAA
TCTCCTTAGACCGCTGCGTGATCTTAGTGACGGCGCTAGTCTCTGCAAA
ATCTTGAATCATTTTACTCTGCGAAAGCCAGTTGCAGCAACAATCAAC
ACCCTTGCGAATGCTCTTTACCAAGTGTCTGTTTCATCAGAAAATGAAGC
AAGAAACGAGATGAGGGAAGCTTGCTTCGATTATCTGGGCTCGGGGTA
TGTGCACAGTGACCAAGTATCTTTGCTTTGCGGTTTGAACCTCGACCA
TTAACACTTGTCTGCCATTTCTTTGCGGTTGCGGTTTATGGAGTCATACG
GTTGTTAATCCCATTCCTTCCCAAAACGAATCTGGCTTGGAGCCAAAT
TAATCTCGGGAGCATCGGGGATAATATTTCCGATAATAAAGCGGAAGGA
GTG---CAGATGTTTTTCCAGCAACTGTACCTGCATACTACTACAAAGC
TCCT---ACAGTTGGAGAAACCAATGTTCA
>P

-----GCGGATCAGTGCAATGCCAGGATGGATCTTAGCTACT
CTGTTTGGTCTC-----TCTGCTATCTATTATTTGGTAACCAA
GAAAAAGGAACAGAGAGCTAATAAGGATTTGTTGGAGCGAGAAGCGAGT
CCGTTAATAGCATAGCCACCACC-----AGTGGAGAATGCAGATCCAGT
AATGTTTCCGAT---GTCGACGTTATCATTGTCGGCGCGGTTGCTGG
CGCTGCTCTCGCTCATACTCTTGGCAAGGATGGACGTCAGTGCAATGTCA
TTGAAAGAGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG
CCAGGGGGTTACCTTAAGTTAATTGAATTGGGACTTGAAGATTGTGTGGA
GAAAATTGATGCTCAGAGGGTGTGTTGGTTATGCACTTTTCAAGGATGGAA
AACATACTCGACTCTCTTATCCTTTGGAAAAGTTCCACTCAAATGTAGCT
GGAAGGAGCTTTTCACAATGGGCGTTTATACAGAGAATGCGGGACAAAGC
TGCGTCCCTCCCAATGTACAATTGGAGCAAGGAAGTGCATATCTCTAC
TTGAAGATAAAGGGACTATTAGAGGCGTGCAGTACAAACAAAGATGGG
CAAGAGCTGAAGGCATTTGCACCTCTGACAATTGTTTGTGATGGCTGTTT
CTCAAACCTGCGCCGCTCCCTTTGCAACCCTAAGGTGGATGTGCCCTCTT
GTTTTGTTGGCATGGTCTGGAGAATTGCCAGCTTCCATGTGCAAAATCAT
GCTCATGTTATCTTAGGGGATCCGTCTCCGATTTTGTATCTCTATTAG
CAGTACCGAGGTCGCTGTCTGGTTGATGTACCTGGTCAGAAAGGTTCCCTT
CCATTTCAAGTGGTGAAATGGCAAAATATTTGAAGACTGTGGTGGCACCT
CAGCTTCCCGAGAAAGTTATGATGCTTTCTAGCTGCTGTTGATAAAGG
AAATATAAGGACAATGCCAAACAGAAGCATGCCAGCTGCTCCTTATCCTA
CTCCTGGTGCAATTGTTGATGGGCGATGCATTCAACATGCGTCATCCTCTT
ACCGGGGAGGAATGACTGTTGCAATGTTGCTGACATTGTTGTTTTACGGAA
TCTTCTCAGGCCTTTGCGCAACCTTAATGATGCACCTACACTCTGCAAAAT
ATCTTGAATCTTTCTACACCTTGGCAAGCCTGTAGCATCCACAATCAAT
ACACTGGCAGGTGCCCTCTACAAGGTTTTTTGTGCTTCACTGATCAAGC
AATGAAGGAAATGCGTCAGGCGTGCTTTGATTATTTAAGTCTTGGAGGCG
TTTTTTCAGAGGACCGGTGCTTTGCTCTCAGGCTTAAACCCCGTCCG
TTGAGCTTGGTTGCCACTTCTTTGCGGTTGCTATATACGGGGTTGGGCG
TTTGTTACTGCCATTCCCTTCACTAAACGCATCTGGATTGGAGCTAGAT
TGATTTCAAGTGATCAGGAATCATCTTCCCATCATTAGGGCAGAAGGA
GTTAGACAGATGTTCTTCCCTGCAACTGTTCCAGCATATTAT---AGAGC

TCCT---CCAGTC-----AAG-----
>R

-----GCGGATAATTACTTGTTAGGATGGATCTTATGTTCT
ATAATAGGACTGTTTGGATTGTATTATATGGTGATTTTGGTGGTCAAAAG
AGAAGAAGAAGATAACAACAGAAAAGCTCTGTTACAAGCAAGAAGTGACT
CCGCCAAAACATATGTCCGCCGTGAGTCAAAATGGAGAATGCAGATCCGAT
AACCTCGCCGAT---GCTGACATCATCATCGTTGGTGCTGGCGTTGCTGG
TTCCGCTCTGGCTCATACTCTCGCAAGGATGGGCGTAGAGTGCATGTTA
TTGAAAGAGACTTGACTGAACCTGACAGAATTGTTGGTGAATTGCTACAA
CCGGGGGGCTATCTAAAGTTAATTGAGTTAGGACTTGAAGATTGTGTGGA
AGAAATTGATGCACAGCGGGTGTGTTGGTTATGCACTTTTATGGATGGGA
AACACACCCAACCTCTCTTATCCCTTGGAAAAGTTTCACTCTGATGTAGCA
GGGCGGAGCTTCCACAATGGGCGTTTTATACAGAGGATGCGGGAGAAAGC
TTCATCTATTCTAATGTACGATTGGAACAAGGAAGTGTCACTTCTCTAA
TTGAAGAAAAAGGGATAATCCGAGGAGTGGTTTACAAAACGAAAAGTGGT
GAAGAGCTCACAGCTTTTGCACCACTAACAATTGTTTGGGATGGTTGTTT
CTCAAACTTGCCTCGCTCCCTTTGCAACCCCAAGGTGGATGTGCCCTCTT
GTTTTGTTGGTTTGGTCTTGGAGGATTGCAAACTTCCATATCAATATCAT
GGGCACGTTGTATTGGCAGACCTTCCCAATTCTGTTCTATCAGATCAG
TAGTACAGAAGTTCGCTGTCTAGTTGATGTACCTGGCCAGAAGGTTCCCTT
CTATTTCAAATGGGGAGATGGCAAAATATTTGAAGAATGTGGTGGCACCT
CAGGTTCCGCCAGAAATATATGATTCTTTGTAGCTGCAGTTGATAAGGG
AAACATTCGGACAATGCCAAACAGAAGCATGCCTGCTTCTCTTATCCTA
CTCCTGGTGCATTATTGATGGGTGATGCATTCAACATGCGCCATCCTCTG
ACTGGGGGAGGAATGACTGTGGCACTATCTGACATTGTTGTGTTGGGGA
ACTTCTGAAGCCCTTGCCTGACCTGCATGATGCACCTACTCTCTGCAGAT
ATCTTGAACTCTTCTACACCTTGCAGAAAGCCTGTGGCATCTACAATCAAT
ACGCTTGCAGGTGCCCTCTACAAGGTGTTTTGTGCTTCATCTGATGAAGC
AAGGAATGAAATGCGGCAAGGCTTGCTTTGATTATTTGAGTCTAGGAGGTG
TATTCTCAACTGGACCAATCTCTTTGCTCTCGGGTTTAAATCCTCGTCCA
TTGAGCTTGGTTGTCCATTTCTTTGCTGTGGCAATTTATGGTGTGGCCG
TTTGTTACTGCCATTCCCTTACCTAAGCGCGTCTGGGTTGGAGCTAGGT
TGATATCGGGTGATCAGGAATCATCTTCCCAATTATCAAGGCAGAAAGGA
GTTAGACAGATGTTCTTCCCGCAACCGTCCCAGCATATTAT--AGAGC
TCCT---CCTGTT-----GAATGCAAT
>O

AAAACCGGAGGGAGGAGTCTCCTCTTCTTCTCCTTCCCTCCGCTCCC
GATAAATCGGAACAATCTCGAAGCATCCGCCCGTCTCCGCCTGCTCCTCC
CACCAACACACCGCTGCTTCCCTTCTCCCTCTCGCCCGCCGCCGCTC
GCCGCGACGCGCATCGGCGCGGCCACGACCGCAGCAGAGAAGCCACCC
ACCACGCGCGGAGCGGCCGAGGAGGCCCTACCACTATTGGTGCGGGATGG
CGGAGGTGCGCGCGGCCACCGGGCAGCTATTGGCGTGCCTGTTGCCACG
CTCCTCGCG-----GCGATCTTCTCGCCGCCGCGCTTCT
TGGCAGCGCGCGCGCGCGCGCGCGCGCGCTTGGCAGGGAAGCCTGCGG
CGGTGGGTGCTGCGGGGTGGCCGACGCGGAAGGG---TGCGCGGGGAT
GGCCGG-----ACCGACGTATCGTGTGCGAGCGGGGTGCGCGG
ATCTGCGCTGGCATAACGCTCGGAAAGGATGGCCGTGCTGTGCATGTTA
TAGAGAGAGACCTAACAGAGCCTGATAGAATTGTGCGTGAAGTGTACAA
CCTGGTGGCTACCTGAAATTGATCGAGTTGGGTCTCGAGGATTGTGTTCA
AGAAATAGATGCTCAGCGTGTCTTGGTTACGCGTTATTCAAAGATGGGA
AAGACACAAAACCTCCTATCCCTTGGAGAAGTTCCACTCAGATGTTGCT
GGTAGGAGCTTTCACAATGGACGGTTTATACAGAGGATGCGCCAGAAAGC
TGCATCTTTGCCCTAATGTTCAATTGGAGCAAGGAAGTGTACATCATTGG
TTGAAGAAGATGGTACAGTTAAGGGTGTAAATACAAGACCAAGTCAGGT
GAAGAATTAAGCATATGCACCTCTGACAATTGTATGCGATGGCTGTTT
CTCAAACCTTCCGCGCGCCCTTTGCTCTCAAAGGTTGATGTACCATCTT
GTTTTGTTGGGCTGGTCTGAGAAATTGTCAACTTCTCATGCAAACCAT
GGCCATGTTGTCTGCGCAATCTTACCTATCCTATTTTACCAATAAG
CAGCACTGAAGTTGCTGTTTGGTTGATGTCCCTGGTCAGAAGGTGCCCT
CCATAGCAAAACGGTGAAATGGCAAAATATCTCAAACAGTGGTTGCACCT
CAGATTCCTCCAGAAATCTATGATTATTATAGCAGCCATTGATAAGGG
AAGCATAAGAAACAATGCCAAACAGGAGCATGCCGGCTGCTCCACATCCAA
CCCCTGGTGCATTTTGTGGGTGATGCATTCAACATGCGGCATCCTTTG
ACTGGTGGCGGAATGACTGTTGCATTATCTGACATTGTTGTGCTACGTAA
TCTTCTCAAGCCTCTCCGCAATCTGCATGATGCATCTGCTCTTTGCAAT
ACCTTGAATCATTCTATACACTGCGGAAGCGGTTGCTTCTACCATAAAC
ACATTAGCTGGTGTCTATACAAGGTTTTAGTGCCTCACCTGATCAGGC
TAGGAATGAGATGCGCAAGCCTGCTTTGATTACTTGAGCCTTGGAGGTG
TCTTTTCAAATGGGCTATTGCTCTTCTGTCTGGTCTGAATCCTCGACCA
TTGAGTTTAGTGGCACATTTCTTTGCTGTGCTATCTATGGTGTGCGTCTG

pre-BR.2.5.1.21.fasta

$\geq P$

GCTGAAGCTGAAAATGGCAGCTAAGCATGCTGAAAAACAGATCCACGTG
 AACCTCACTGGGCTTTCTGTTATTCTATGCTCCCTAGAGTCTCTCGTAGC
 TTTGCTCTCGTTATCCAACTTGACACAGAACTCCGTAACGCTGTATG
 CGTATTTTATTTGGTCTTCGAGCCCTTGACACTGTTGAGGATGATACAA
 GCATACCTACAGATGTCAAAGTGCCCTATCCTGATAGCTTTTACC GCCAC
 ATATATGATCATGACTGGCATTCTCATGTGGTACCAAGGAGTACAAGGT
 TCTCATGGACCAGTTTCATAATGTTTCAAATGCTTTTCTGGAGCTTGGAA
 AAGGTTATCAGGAGGCAATCGAGGATATTACCAAAAGAAATGGGTGCAGGA
 ATGGCAAAGTTTATCTTGAAGGAGGTGGAAAGCATTGATGACTATGATGA
 ATATTGCCACTATGTAGCAGGACTTGTGGACTGGGCCTGTCCAACTCT
 TCCATGCATCTGGATTAGAAGATTGGCACCAGAT-----AGCATCTCC
 AATTCAATGGGTTTGTCTTTCAGAAAACAAACATTATTCGTGATTATTT
 GGAGGACATAAAGCAGATACCTAAGTCACGCATGTTTTGGCCTCGCGAGA
 TTTGGAGCAAATATGTCAACAACTTGAGGACTTGAAATATGAAGAGAAC
 TCGGTGAAGGCAGTGCACTGCTGAATGATATGGTTACCAATGCCTTGAT
 ACATATGGATGATTGCTTGAAGTACTTGTCTGAATTGCGGGATCCTGCTA
 TATTTGGTTTTGTGCTATTCTCAGATCATGGCGATTGGAACCTAGCA
 CTGTGCTACAACAATGTCAATGTCTTCAAGGTTAGTGAAGATGAGGCG
 AGGTCTTACCGCTCAAGTTATTCATCAAACGAAAACAATGGATGATGTCT
 ATGGTGCTTTCTTCGACTTCTCTTGTATGCTGAAGTTCAAGGTTGACAA
 AGTGATCCTAATGCAATAAAACCTTGAGCAGGCTGGAAGCAGCACAAAA
 ATCTTGCAAGGGAATCTGGGGCTCTAAACAAAAGGAAATCTTACATAATTA
 GGAATGAGCCAAAATATAATTCTGTTCTGATCGTCTACTTTTCATTATA
 TTGTCTATTATTTTCGCGTATCTCTGCCAACAGATCGAGTTAC
 >R
 ---GGTAGTGTTGGAGCGATTTTGAGACACCCAGATGATTTTTACCCGTT
 ATTGAAGTTAAAAATGGCAGCTAGGAATGCAGAGAAGCAGATCCCACCAG
 AACCTCGCTGGGGTTTCTGTTATTCTATGCTTCATAAGGTTTCTCGTAGC
 TTTGCTCTTGTATTCAACAGCTTGGCACTGAGCTCCGTGATGCTGTCTG
 CATATTCTATTTGGTCTTTCGAGCCCTTGATACTGTTGAGGATGATACAA
 GCATACCTACAGACGTCAAAGTGCCCTATTCTGATAGATTTTCATAAACAC
 ATATACAATCCTGAATGGCACTTTTCATGTGGCACAAGGATTATAAAGT
 TCTCATGGACCAGTTTCATCATGTTTCAACTGCTTTTCTTGAGCTTGGAA
 AAAGTTATCAGGAGGCGATTGAGGATATTACAAAAGAAATGGGTGCAGGA
 ATGGCTAAATTCATATGCAAAGAGGTGGAAACAATTGATGACTATGATGA
 GTACTGCCACTATGTAGCAGGACTTGTGGACTAGGACTGTCCAAGCTTT
 TCCATGCCTCTGGAAACGGAAGATTGGCACCAGAT-----ACCCTCTCC
 AACTCA---GGTTATTTCTTCAGAAAACAACTTATTCGAGATTATCT
 GGAGGATATAAATGAGATACCTAAGTCACGCATGTTTTGGCCTCATCAAA
 TCTGGAGCAAATATGTTACTAACTCGAGGATTGAAATATGAAGAGAAC
 TCGGTCAAGGCAGTGCAATGCTTGAACGATATGGTTACTAATGCTTTGAT
 CCATATGGATGATTGCTTAAATACATGTCTGCACTGCGCGATCCTTCTA
 TATTTCTGTTTTTGGCAATCCCTCAGATCATGGCAATTGGAACCTTAGCA
 TTGTGCTACAACAACGTTGAAGTATTCAAGGTTAGTGAAGATGAGGCG
 TGGTCTTACTGCGAAGGTCAATTGACCGAACAAAGACTATGGCTGATGTAT
 ATGGTGCTTTTTTATGACTTCTCCTGT---CTTAAGTCCAAGGTTGAAATG
 AATGATCCTAATGCAGAAAAGACATTGAGCCGGTTGGAAGCATTACAGAA
 ATCTTGCCGGGAATCTGGCCTCCTAAGCAGAAGGAGATCTTACATA---A
 GGAGTGAGCCAAAGAAATTAATTCTGTTATTGTTGCTCTACTTTTCATCATA
 TTGTCCATTATTTTCGATATCTCTCGGCGAATCGACCAAGTAAC
 >O
 -----GGGGTGCTGTGAGGCCGAGGAGGTGTTGCCGCT
 GGTGAAGCTGCGGGTGGCGGCGGGGCGGATCAAGCGGCAGATCCCGCCGG
 AGGAGCACTGGGCCCTTCGCCACCATGCTGCAGAGGCTCTCCCGGAGC
 TTCGCGCTCGTCATCCAGCAGCTCGGCCCGACCTCCGCAATGCCGTGTG
 CATCTTCTATCTCGTGCTCCGAGCCCTCGACACTGTTGAGGACGACACTA
 GCATTCTGCCGCGGTGAAGGTGCCGATCCTTAAGGAATTCATCGGCAT
 ATCTACAACCGGACTGGCATTATTCATGTGGAACAAAAGACTACAAATT
 ACTGATGGATAAGTTTCGCCCTGTCTCCACGGCTTTCTTGAGCTTGGTC
 AAGGTTATCAAGAGGCAATTGAAGAAATCACTAGGCTAATGGGAGCAGGA
 ATGGCAAATTTATCTGCAAGGAGGTTGAAACTGTTGATGACTACAATGA
 GTACTGTCACTATGTAGCAGGGCTAGTGGGTTATGGGCTTTCCAGGCTCT
 TTCATGTGTTGGGACGGAAGATCTGGCTTCAGAT-----TCACTTTCA
 AATTCAATGGGCTTGTCTTGCAGAAAATCAATATAATTAGGGATTATTT
 GGAGGACATAAAGCAGATACCAAGTCACGTATGTTCTGGCCTCGAGAAA
 TATGGAGTAAATATGTCAATAAACTCGAGGATTGAAATACGAGGAAAT
 TCAGAAAAGGCAGTTCAAGTGTGAAATGATATGGTGACTAACGCTCTGTC
 TCATGCTGAAGACTGCCCTCAATACATGTGAGCATTGAAGGATCATGCCA
 TTTTCCGTTTTTGTGCAATACCTCAGATAATGGCAATTGGGACATGTGCT
 ATTTGCTACAATAATGTGAATGTCTTTAGAGGAGTTGTTAAGATGAGGCG
 TGGGCTCACTGCACGAGTAATTGATGAGACAAACAAATGTGATGATGTCT
 ATACTGCTTTCTATGAGTTCTCTTCGCTGATAGAATCGAAGATTGATAAT
 AATGATCCAAATGCTTCCCTAACGCGGAAACGTTGATGCGATAAAGAG
 AACCTGCAAGTCATCTTGCTCACTA---AAGAGAAGGGGATACGAT---T
 TGGAGAAGTCAAAGTACAATCCATGCTGATAATGGTTGTAATCTGTTG

GTGGCTATTGTTTTGGGCATGATATATGCCAAG-----
 >V
 ---GGCAGTTTGGGAGCGATTTTGACGCATCCGGATGACATATACCCACT
 GCTGAAGCTGAAAATGGCGGTACGCCACGCCGAGAGCAGATCCCGCCGG
 AGCCGCACCTGGGCCCTTCTGCTACACTCTGCTCCACAAGGTCTCTCGCAGC
 TTCGGTCTCGTCATTCAACAGCTCGGCACCGAGCTCCGCAACGCCATATG
 CATTCTCTATTTGGTTCTTCGAGCACTCGACACTGTTGAGGATGATACAA
 GCATTCGTACAGATGTCAAAGTGCCTATTCTGATAGCTTTTCATCGTCAT
 ATATATGACCGTGACTGGCATTGTCATGTGGTACAAAGGACTACAAGGT
 TCTCATGGACCAATTCCATCATGTTTCAACTGCTTTTCTGGAGCTTGAAA
 GAGGTTATCAGGAGGCAATTGAAGACATCACCAAACGAATGGGGGCAGGA
 ATGGCAAATTTATTTGCAAAGAGGTGAAACCATTGATGACTATGATGA
 ATATTGCCACTATGTAGCAGGACTTGTGGATTAGGATTGTCCAAGCTTT
 TCCATGCCTCCAAGTTGGAAGATTGGCTTCTGAT-----GAACTCTCC
 AACTCA---GGTTATTTCTTCAGAAAACAAATATAATCCGAGATTATCT
 GGAAGGATATAAATGAGATCCCAAAGTCACGCATGTTTGGCCCCGTGAGA
 TTTGGAGTAAATATGTTAACAACTTGAGGACTTGAAAGAGGAGGAAAAC
 TCAATCAAAGCAGTGCAATGCTTGAATGACATGGTCACTAATGCTTTAAT
 ACATATGGAAGATTGCCTGACATACATGTCTGCTTTCAGAGTCCAGCAA
 TATTTGATTTTGTGCTATCCACAGATCATGGCAATCGGGACACTAGCT
 TTATGCTACAACAATATTGAGGTCTTCAGAGGAGTAGTAAAAATGAGGCG
 TGGTCTTACTGCCAAAGTTATTGACAGAACAAAGCGATGTCTGATGTCT
 ATGGTGTCTTCTCGATTTTCTTGT---CTGAAGTCCAAGGTTGACAAG
 AATGACCCGAATGCTACAAAAGCATTGAGCAGGCTAGAAGCAGTACAGAA
 AATTTGCAGGGAATCTGGAGCCCTTACCAAAGGAAATCCTATGTAATCA
 GGAGCGAACCAAGATATAATTACAGCTCTGATTGTGGCATTCTCATCATA
 CTGTCCATCATTTTCGCGTATCTATCTGCCAACCGTCAGAATAAT

pre-BR.5.4.99.8.fasta

>A
 TGGAAACTGAAGATCGCGGAAGGA---GGTAGTCCATGGCTTAGAACCCAC
 CAATAATCACGTCGGAAGACAGTTTTGGGAGTTCGATCCGAATCTCGGT-
 --ACTCCTGAGGATCTCGCCGCCGTGGAAGAAGCTAGGAAGCTTTTTTCG
 GATAATCGATTCTGTGCAAGAACATAGCGCCGATCTGCTTATGCGCCTTCA
 GTTTTCAAGAGAAAAATTTGATTAGCCAGTTTTACCTCAAGTCAAAATCG
 AAGACACTGATGATGTTACAGAGGAGATGGTGGAAACCACGTTAAAGAGG
 GGTCTAGATTTCTATTCAACTATACAGGCACACGACGGGCACTGGCCAGG
 TGATTATGGTGGTCTATGTTTCTTCTCCAGGACTGATAATTACACTCT
 CCATAACTGGAGCACTGAATACAGTATTGTGCGAACACATAAACAAGAA
 ATGCGCCGTTATCTCTATAATCACAGAAATGAGGAC---GGAGGTTGGGG
 TTTACATATTGAGGGCCCTAGCACCATGTTTGGGTCTGTGTTGAACATG
 TTAATCTAAGGTTGCTTGGAGAAGGACCTAACGATGGAGATGGAGATATG
 GAGAAAGGACGAGACTGGATACTAAATCATGGTGGTGTACCAATATTAC
 ATCTTGGGGGAAAATGTGGCTATCGGTACTTGGAGCTTTTGAATGCTCG
 GAAATAACCCACTGCCACCTGAGATATGGCTTCTCCCATATTTCTGCCA
 ATACATCCAGGAAGGATGTGGTGCCATTGTGCAATGGTGTACTTGCCGAT
 GTCGATTTGTATGGAAAAAGGTTTGTGGGTCCCATAACGTCACACTGTTT
 TATCACTGAGAAAAGGAGCTTTTACAGTACCATAT---CAAGTCAACTGG
 AATGAAGCACGCAACCTTTGCGCAAAGGAGGATTTATACTACCCACATCC
 ACTTGTGCAAGATATTCTTTGGGCATCACTTCATAAGATTGTTGAGCCTG
 TTCTGATGCGATGGCCTGGTGCAAAATTTGAGAGAAAAGGCTATAAGAACC
 GCAATAGAACATATTCTATTATGAAGATGAGAATACTAGGTACATCTGCAT
 AGGTCCCGTGAACAAGGTATTAATATGCTTTGCTGTTGGGTAGAAGACC
 CAACTCAGAGGCTTTCAAGTTGCACCTACCAAGAATCCATGACTTTCTC
 TGGTTAGCTGAAGATGGAATGAAGATGCAAGGTTATAACGGAAGCCAGCT
 ATGGGATACAGGTTTTGCTATTCAAGCGATTTTGGCAACTAACCTCGTCG
 AAGAATATGGGCCCGTTTTGGAAAAAGCACATTCAATTTGTCAAGAATTCC
 CAGGTGTTAGAAGACTGCCCTGGAGATCTGAATTACTGGTATCGCCACAT
 TTCTAAAGGGGCTTGGCCTTTCTCAACTGCAGATCACGGTTGGCCCATCT
 CTGACTGCACCGCAGAAGGACTGAAAGCTGCTCTTTTGTATCCAAAGTT
 CCCAAGGCGATTGTTGGTGAACCAATAGATGCAAAACGGTTATATGAAGC
 TGTTAATGTTATCATTTCTTTACAGAATGCAGATGGAGGCCTCGCAACAT
 ATGAGCTCACCAGGTATACCTTTGGTTAGAGCTAATCAACCCAGCAGAA
 ACCTTTGGCGATATTGTTATTGATTATCCTTACGTGGAATGTACATCAGC
 TGCATATCCAAAGCTTTGATATCATTTTCAAAGCTGTATCCTGGTCACTGAA
 AGAAGGAAGTAGAG---TGCATTGAGAAGGCGGTTAAGTTCAATGAATCC
 ATTCAAGCAGCAGATGGCTCATGGTATGGATCATGGGCTGTTTGTCTCAC
 GTATGGTACGTGGTTTGGAGTGAAAGGGCTGGTAGCTGTTGGAAAAACAT
 TGA AAAAATCTCCACATGTTGCTAAAGCTTGTGAATTTCTATTGTGAAAA
 CAACAACCTTGGGGCGCTGGGGAGAAAGCTATCTTTATGTCAAGACAA
 GGTCTATTCAAACCTTGATGGCAACAGATCTCACGTCGTGAATACAGCAT
 GGGCTATGCTCGCACTCATTGGTGTGGGCAAGCTGAGGTAGACCGGAAA
 CCACTACACCGGGCTGCAAGATACTTGATTAATGCTCAATGGAGAATGG

TGATTTTCCACAACAGGAAATAATGGGAGTCTTCAATAGGAAGTGCATGA
 TAACATATGCCGCGTATCGAAACATTTTCCGATATGGGCTTTGGGGGAG
 TACCGTTGTCAGGTATTATTGCAACAAGGAGAA - - -
 >P
 TGGAGGCTAAAGATAGCAGCAGGA - - - GGAAATCCATGGCTAAGAACAAA
 TAATGATCATATTGGAAGACAGATTTGGGAATTCGATCCGAACTAACTC
 TATCTCCCGAAGAGATCTCAGAGATCGAAAATGCGCGTCAGAACTTTACT
 AAAAATCGGTTTCAATCCAAGCATAGTGCAGATCTAATCATGCGCATGCA
 GTTTGAGAAGGAGAATCCGGTGTGCGAAGTGTGCGCAGGTGAAATTGA
 AGGAGAGTGAGAAAGTTACGGAGGAGGCGGTTACTACAACGTTGAAGAGA
 GGTGGATTATTATTCGAGTATTCAAGCTCATGATGGACACTGGCCTGG
 TGATTATGGTGGTCTATGTTTCTTATGCCAGGATTGGTGATTACTTTGT
 CAATTACTGGAGCACTACATGCAGTGTATCAGATGAACATAAAAAGGAG
 ATAATTCGGTACCTTTACAATCATCAGGCATGTGATTATGGGGGATGGGG
 TTTGCATATTGAGGGACCGAGTACGATGTTTGGTTCGGTGTGAATTATG
 TTTACTTTAAGGTTGCTTGGAGAAGGGCCGAATGATGGAGAAGGTGCGATG
 GAGAAAGCACGTGATTGGATTTTGAATCATGGTGGTGCAGTATGATAAC
 ATCATGGGGGAAAATGTGGCTTTCACTACTTGGAGTGTGGAGTGGTCTG
 GAAATAATCCAATGCCCCCTGAGATGTGGCTTCTTCTTATTTGCTCCCA
 GTTCATCCAGGAAGGATGTGGTGTCACTGCCGGATGGTCTATCTGCCTAT
 GTCCTACTTATATGGGAAGCGGTTTGTGGCCCAATCACACCAACAATTT
 TATCTTTGGGAAAAGGAGCTGTTCACTGTCCCATATCATGAAATAGAATGG
 AATCAAGCGCGCACCTTATGTGCAAGGAGGACTTATATTACCCTCACCC
 TCTGGTGCAGATGTACTTTGGGCATTGCTTGACAAGGCTGCCGAACCTG
 TTCTAATGCATTGGCTGGAAAGAAGTTGAGAGAAAAGGCACTTTGTACT
 GCAATTGAGCACATACACTATGAAGATGAAAATACTCGTTATCTTTGCAT
 AGGACCTGTAAATAAGGTGTTAAATATGCTTTGTTGTTGGGTGGAAGACC
 CTAATCAGAAGCTTTCAAGTTGCATATTCTAGAAATACAAGATTACCTC
 TGGCTTGCTGAAGATGGAATGAAAATGCAGGGTTACAATGGAAGTCAATT
 GTGGGACACTGCTTTTGGCGTTCAAGCAATTATTTCAACTAATCTTGTG
 AGGAATACAGTCCAACTTTAAAAAAGCA - - - CCATTCGTTAAAAATTCA
 CAGGTTTTGGAAGATTGCCCTGGAGATCTTCATTTCTGGTATCGCCATAT
 TTCCAAAGGTGCATGGCCTTTCTCACTGCAGATCATGGATGGCCCATCT
 CAGATTGCACAGCAGAGGGGACTAAAAGCTGCTCTTATTATCAAAAATT
 CCATCTGAGATTGTTGGGGAACCGTTAGTTGCAAACCGATTTTATGATGC
 AGTAAATGTCCTCTCTCATTACAGAAATGGGGATGGTGGTTTTGCAACAT
 ATGAGCTCACGAGATCTTACAGCTGGTTAGAGTTGATCAACCCAGCTGAA
 ACTTTTGGTGACATTGTTATTGATTACCCATATGTTGAATGTAATTCAGC
 TGCAATTCAGCTTTGACATCGTTTAAAGAAATTATATCTGGACACCGGT
 CAGAAGAAATAGAAAGTTGTATCAGAAAGGCGACTATGTTCAATTGAAAGT
 ATCCAGGAGAAGGATGGCTCG - - - TATGGCTCATGGGGTGTGTTGCTTAC
 CTATGGCATATGTTTGGCATAAAAGGGCTGGTGGCTGCTGGAAAAAACT
 TCAATAATAGCTCTAGCATTCTGTAAGCCTGTGATTTTCTGCTGTCCAAA
 CAGTGTCTTCTGGTGGTTGGGGAGAGAGTTATCTTTCATGTCAAAAACAA
 GACATATTCAAATATTGAAGGGAACAGAGCTCATGTGGTAAATACAGCAT
 GGGCTATGTTGTCTCTCATTGAAGCTGGGCAAGGCTGAGAGAGAGCCAGAA
 CCATTGCACCGTGCAGCAAGATATTTGATAAATCCCAGATGGAAAATGG
 AGACTTTCCGACAGGAAATCATGGGAGTTTCAATAGGAATTGTATGA
 TAACATATGCTGCCTACAGGGACATTTCCCCATTTGGGCATTGGGAGAA
 TACCGGTGTCGGGTGTTGCAGGCCAAGGCTTCA - - -
 >R
 TGGAAAGCTAAGGATTGCAGAAGGAAGTGGAAATCCATGGCTTAGAACTAC
 AAACGATCACATTGGGAGACAAGTCTGGGAATTTGATTATCTAAAAATCG
 GATCTCCAGAAGAGCTCTCACAAATCGAAAATGCTCGTCAAAATTTCAACC
 AAAAACCCTTTATTACAAGCATAGTTCAAGTCTACTCATGCGCATTCAT
 GTTTTCGAAGGAAAATCCAATATGTGAAGTGTACCGCAAGTAAAAGTGA
 AAGAATCAGAACAGTTACCGAGGAGAAAGTGAATAATACGTTAAGAAGG
 GCATTGAATTACTATTATCTATTCAAGCTGATGATGGACATTGGCCTGG
 TGATTATGGCGGTCTATGTTTCTAATGCCTGGCTTGATTATAGCTCTGT
 CTATCACTGGCGCATTTGAATGCAATTTTATCTGAGGAACATAAGCGCGAA
 ATGTGTCTGATATCTATACAACCACCAAGATAGAGAT - - - GGCAGTTGGGG
 CTTACGATTGAGGGACCAAGCACAATGTTGGGAGTGTATTGTGTTATG
 TTTCTGTTAAGGCTGCTTGGTGAAAGGTCCCAATGAGGGAGAAGGGGAGTC
 GAGAGAGGCCGTAATTGGATTCTCAAACATGGTGGTGTACTGCAATTAC
 GTCCTGGGGAAAATGTGGCTTTCACTACTTGGAGCATATGAGTGGTCTG
 GTAATAATCCCTTACCCCTGAGATGTGGCTTCTCCCTTACATCTCCCA
 GTCCATCCAGGTTCTTAT - - - - - GTATATCTTCCA - -
 - TGTCATATTTATATGGGAAAAGGTTGTTGGTCCAATCACACCAACAGTT
 TGCTTTTGAGAAGGAGCTGTACTGTCCCATATCATGAAATAGACTGG
 AATCAAGCACGCAACCAATGTGCAAGGAAGATTTGACTATCCACATCC
 TATGTTGCAAGATGTACTTTGGGCAACTCTTCATAAGTTTGTGTAACCTA
 TTCTTATGCATTGGCTGGAAAAAGGTTGAGAGAAAAGGCTATTCAAAC
 GCGATTGAACACATACACTATGAAGATGAAAATACTCGTTACATATGCAT
 AGGGCTGTAAACAAGGTGTTAAATATGCTTTGCTGTTGGGTGGAAGACC
 CAAACTCAGAAGCTTTTAAAGTTGCATCTTCAAGATTATATGACTACCTC

AAGATACAGAGGATGTTACAGAGGATGCAGTGACAAACACATTAAGAAGG
GCTATCAATTTTCATTCAACTCTCCAGGCTCATGATGGACACTGGCCAGG
GGATTATGGAGGTCCCATTGTTCTTTTGGCCGGCTTGATCATTACTCTTT
CGATTACTGGGGCGCTGAATGCTGTCTGTGCAAAAGAACATAGACAAGAG
ATGTGCCGTTATCTCTATAACCATCAGAACAAAGAT - - - GGTGGGTGGGG
TTTACATATCGAGGGCCCAAGCACCATTGTTTGGTACTGTCTTGAACATG
TAACTTTAAGGTTGCTTGGTGAAGGAGCTAATGATGCAGATGGGGCAATG
GAGAAAGGGCGTGACTGGATTCTGAATCATGGTGGTGAACCTGCAATAAC
CTCATGGGGAAAAATGTGGCTTTTCAGTACTTGGAGTATTTGAATGGTCTG
GCAATAATCCTCTGCCCTGAAATATGGCTTCTTCCCTATATCCTACCA
GTCCATCCAGGAAGGATGTGGTGTCACTGCAGGATGGTTTATTGCTTAT
GTCATACTTATATGGGAAGAGGTTTGTGGTCTCTCACACCTACAGTAT
TGTCTTTGAGGAAGGAGCTCTATACTGTTCCATATCATGAAATCGATTGG
AATCAAGCAGCAAAACCTATGTGCCAAGGAAGATCTGTACTACCCACATCC
ATTGGTGCAGGATATACTTTGGACATCCCTCGACAAAGTTGTTGAACCTA
TTCTAATGCATTGGCCTGGGAAAAAGTTGAGAGAGAAGGCTCTTCGCACT
GTCTTGGAGCATGTACACTACGAAGACGAGAACACTCGCTATATATGCAT
AGGTCTGTGATAACAAGGTTTTAAATATGCTTTGCTGTTGGGTGGAAGATC
CAAACCTCAGAGGCTTTCAAGTTACATCTTCTAGAAATTTTGATTTCCTG
TGGCTTGCTGAAGATGGCATGAAAATGCAGGGCTATAATGGAAGTCAATT
ATGGGACACAGCCTTTTGTGTTCAAGCAATTATTTCAACCGACCTAGTGG
AAGAATATGGTCCAACTCTAAGAAAAGCACATGCATACCTTGAAGAAGTCA
CAGGTTTTAGAAGATTGTCCAGGCAATCTTGATTATTGGTATCGCCACAT
TTCAAAAGGTGCTTGGCCTTTTTCAACCGGTGATCATGGATGGCCCATCT
CAGACTGCATGCGGAAGGATTGAAGGCTATTCTTTTATTATCAAAAATT
CCATCAGAAAATTGTTGGTGAACCATTAGATGCGAAGCAGTTATACGATGC
TGTAATGTAATTCTTTCTTACAGAATGGAGATGGCGGCTTTGCAACCT
ATGAACCTCACTAGATCTTATGCTTGGTGGAGTTAATCAATCTGCTGAA
ACTTTTGGCGACATTGTTATTGATTATCCTTACGTGGAATGTACATCAGC
TGCACTTCAAGCTTTGACATCATTTAATAAATTTTATCCGGGGCATCGGA
GAGAAGAAATAGAACATTGTATTAATAAAGCTACAATGTTTATTGAAAAG
ATACAAGCATCAGATGGCTCATGGTATGGCTCTTGGGGAGTATGCTTCAC
CTATGCCATATGGTTTGGGATAAAAGGATTAGTGGCTGCTGGAAAGAACT
ACAATAATTGCTCTAGCATTTGTAAAGCATGTGATTTTCTGTTGTCCAAG
CAGCTTGCTTCTGGTGGTTGGGGCGAGAGTTACCTTTCTGTGTCAGAACAA
GGTGTATTCAAATCTTGATGGCAACAGATCTCATGTAGTAAATACCGCAT
GGGCTATGCTGGCTCTCATTGATGCTGGGCGAGGCTGAGAGAGACCCAACA
CCATTGACACAGAGCAGCAGCAGTATTGATAAATCCCAAATGGAGAATGG
AGATTTTCCACAGGAGGAGATCATGGGGGTCTTCAATAGGAAGTGTATGA
TCACTTATGCTGCATACAGAAACATCTTCCCCATTTGGGCTCTGGGTGAA
TACCGATGTGCGGTTCTACGGGGTCCCTTGACAAAA

pre-BR.5.5.1.9.fasta

>A
-----TCAGGATC
TTCTTCACCGAGCTTGTGGCTGGCACCGAATCCGAGCAAGAGATGGGGAG
AGCTCTTCTTCTCTTCTACACTCCTTTTTGGCTCACTCTCTGTTTAGGC
ATCGTTGTTCTTCTTACAAGCTTTACGAGACATTACGGAGTTGGAGTATCT
GCTTCTAGCTTTGGTTTCAGCTGTTCTGCTTTCTGTTATACCCATGTTAC
TCGTTGGAAAGGCCGACAGAAGTTA - - - - - TGTTGGAAGGACCGCTAT
TGGGTTAAGGCAATCTCTGGATAATTGTTTTAGCTATGTCGGAAACTA
CTTTTGGACTCACTATTTCTTTAAAGTTCTTGGAGCGTCTATACTTTTC
CATCATGGAATAATGAATAATGTGCCTCACACAACATTC - - - TTCCACAT
GTTTGCTTCTCTTTTACCACGTTGCATCGAACATTACTCTTCGGAGGCT
ACGCCATTCCACTGCTGATTTACCAGATTCTCTGAAATGGTGTGTTGAGG
CTGCATGGATACTTGCGCTTTCTTATTTTCATTGCATACCTTGGAGACTATT
GCTATCGCAATTTTCTTACTATGAGTTTGTGGACCGAAGTGCCATGTA
CAGAGTTGGATGCTTTTCTATGCCATTACTTCATTGTGAGCTTCCCAA
TGTTCCTCAGGATGGATGAGAAATCAACTGATGAATGGGATTATCTCGA
GTGGCTGTTGATGCTTTGGGTGCTGCTATGTTGGTAACAATCATTCTTGA
TCTATGGCGTCTCTTCTTGGGACCTATAGTTCCTTACCGGAGGGACAAA
ACTGCCCTCAGTCTGGATTACCATGGTTCTCCAAT - - - - -
>P
-----AGTTCTAG
TTCTAGTCCAGTTTATGGTTGGCTCCAGACCCTAGTAAAAGATGGGGAG
AGATTTTCTTTCTTTGTTATACTCCTTTTTGGCTCACTCTTTGTTGGGA
ATTGTGATTCTTTGCAAGCTGTACGAGAGCTTTACAGAATTGGAATACTT
GCTCTGGGTTGATCTCAGCTGTTCTTCTTTCTTGATACCAATGATGT
TTGTGCGTAAGGCTGATAGCAATTG - - - - - TGTTGGAAGGACTGTTAT
TGGGTCAAGGCCAGTCTATGGATAATCATTTTTAGCTATGTTGGAAATTA
CTTCTGGACCCACTATTTTTTCACAGTTTTGGGAGCTTCTTATACATTTT
CATCTTGAAGATGAATAATGTACCACATACAACCTTCTTCTGACACAT
GTTTGCTTCTGTTTTATCACGTCACCTCAAACATGACAATTGCGAGACT

```

ACGGCATGCGGTTGCTGACTTGCCAGATTATCTTCGATGGGCTGCTGAGG
GTGCATGGATTTTGGCTCTTTCATACTTCATAGCATACCTGGAGACTTTA
GCTGTTTTCTAATTTCCCTTACTATGAGTTTGTGGACCGGGCTTCAATGTA
CAAAGTAGGTTCCTTGTGTTTATGCCATCTACTTCTTTGTAAGCTTCCCAA
TGTTTCTGAGGATTGATGAGAAACCTGGTGATCTATGGGAGTTACCAAGA
GTGGCTGTTGATTCCCTGGGTGCTGCAATGTTGGTTACAATAATACTTGA
CTTGTGGCGAATCTTTTGGGGCTATTGTTCCACTTCCAGTAACAAAAC
AATGCCTTCAACAGGGAATACCTGCTT-----
>R
-----GGCGATGATACTACAAGTAAGGAGTCGGGTTT
TTATACTTCGAGTTTATGGTTGGCTCCTAACGAGAGCAAAAGATGGGGCG
AAGTTTTCTTCTCTGCTATACTCCTTCTGGCTTACTCTGCTTAGGC
ATTGTTGTTCCTTACAAGCTTTATGAGAGCTTTACGGAATTGGAGTATCT
GCTTCTAGCTATGGTTTCTGCAGCTCCTTCTTCTGATTCCAATGATTT
TTGTTGGAAGGTAAATTATAGGTTATCAACCTCAATTGTGAACCTACTG
TTGTGGCAGGCTAGTCTCTGGATAATCATTTTCAGTTACGTTGGCAATTA
CTTTTGGAGCATTATTTCTTTACAGTTTGGGAGCTTCTTATACATTTT
CATCATGGAAGATGAATGATGTACCACACACAACTTCTTCTGACTCAT
GTATGCTTCTGTTTTATCATGTCACTTCAAATATGACACTTCGTAGACT
GCGGCATGCTATTGCTGATTTGCCAGATAAACTTCGGTGGGCTACTGAGG
CAGCATGGATCTTGGCTCTTGCATATTTATAGCATACCTGGAGACCTG
GCCATCTCTAATTTTCCATACTACGAGTTTGTGATCGGTCTTCCATGTA
CAAAGTTGGGTCCTTGTGTTTATGCCATCTACTTCATCGTAAGCTTCCCAA
TGTTTTTGAGGATTGATGAGAAACCTGGTGACTTATGGGACTTACCGAGA
GTGGCTGTAGATGCTCTGGGTGCAGCAATGCTGGTAACGATAATACTCGA
CTTGTGGCGCATCTTTCTTGGACCTATTGTTCCACTATCAGATACAAAAC
AGTGCTTCAACCAGGACTGCCCTGGATTCTGCACATGCCAATGAAACT
>O
GCAGCCCCGCGGGCGGGCGGGCGGCGGAAGGCGAGGAATGGAGGAGGGG
GGGAGGAGCAGCGGTGGATGGCGGCTGACGGGAGCAAGCGGTGGGGG
AGACGTTCTTCTGCTCTACACGCCCTTCTGGCTCACGCTCTGCCTCGGC
GTCGTCGTCCTTCAAGCTCTACGAGAGGTTACGGAAGCTGGAGTACCT
AGTTGTTGGGTTGGTGCCACAGTGCCCTGCCTTCGTCATCCGCTCTTCC
TCGTAGGCAAGGCAGATAGTGTAGA-----AGTTTGAAGATCGGTAT
TGGGTTAAGGCTAATATTTGGATTATAATTTTCAGCTATGTTGGTAATTA
CTTTTGGACGCTTACTTCTTACAGTTCTTGGTGATCATACACTTTCC
CATCATGGAGGATGAACAACGTACCCCATACAACATTTCTCTGACCCAC
GCTTGCTTTCTTTTCTATCACATGACATCAAAT---AGCCTTCGCAAAAT
ACATCATTTACAGCTCATTTGCCCCAGTTTCTTCGATGGTCATTTGAAG
CTGCATGGGTTTTAGCACTCTCATACTTCATAGCTTACCTGGAACCTTA
GCCATTGCAAAATTTCCGTATTATGAATTCATCGACAGGGATATAATGTA
CAAAGTAGGCTCACTGTTTTATGCAATATACTTCATCGTCAGCTTCCCAA
TGTTTTCAAGGATCGATGAAAATGAA---GAGAAGTGGAGCCTTTCCAGA
GTGGCCGTCGACGCCCTTGGGTGCAGCTATGCTTGTAACAATAATCCTCGA
CCTGTGGCGCATATTCTTGGGGCAATTGTACCCATCCCTGAATCAAGAC
GCTGCGGTCAACCGGGCTTGCATGGTTCCAGGTGCAAGAACGAGAGCGTC
>V
-----GGCGGTACAGGGG
GGCTTCTTCGAGTCTGTGGTTGGCTACGAATCCGAGCAAGAGATGGGGAG
AGTTGTTTTTCTTCTCTACACTCCTTCTGGCTCACTTTGTGCTCGGG
ATTGTCGTTCCGTACAAGCTATACGAGAACTTTACGGAATTGGAATACTT
GCTTCTCGGATTGGTTTTCAGCTGTTCTTCTTCTGATACCTATGCAAT
TTGCGGGGGAGGTTGATAGGAGCCTG-----TGCTTGAAGGACCGTTAT
TGGGTAAGGCCAGTTTGTGGATAGTAATCTTTAGTTATGTTGGGAATTA
CTTCTGGACCCATTATTTCTTTACAGTTTGGGAGCTTCATATACCTTTC
CATCATGGAAGATGAATAATGTACCACACACCACTTTCTTCTCACACAT
GTCTGCTTCTATTCTACCATATGTCTGTCAAACATAACGCTTCGTAGACT
ACGACATTCTATTGCTGATTTGCCAGAGAAAATTCAGTGTTTACTGAGG
CTGCATGGATTTTAGCTCTTGCCTATTTTCATAGCCTATTTAGAGACTATA
GCTATTTCTAATTTTCCATATTATCAGTTTGTGGACCGGGCATCCATGTA
CAAAGTTGGGTCCTTGTGTTTATGCTATTTATTTCAATTGTAAGCTTCCCA-
--TTTCTGAGGATTGATGAGAAACCTGGGAATACCTGGGACTTGCCAGG
GTGGCTGTAGATGCTCTGGGAGCTGCAATGCTTGTACAATCATTCTTGA
TCTGTGGCGCATCTTTCTGGGACCTATTGTTCTGATCCAGAAATCAAACC
AGTGCCATCAGCCAGGACTGCCATGGTTTCCAGGACACGGCCCAAAGCT

```

pre-BR.CY051G1.fasta

```

>A
-----GAATTGGATTTCGGAGAACAATTTGTTGAAGACGGGTTTGGTTAT
AGTGGCGACACTTGTTATAGCCAACTCATCTTCTTCTTCTTCACTTCTG
ATTCTAAGAAGAGCGCTTCTCCTACTCTTAAAGCTTGGCTCCATTG
GTTGGAAGTCTTATCAAATTTGAAAGGACCTATTATTATGCTTAGAGA
GGAATACCTAAGCTTGGAAAGTGTTTACTGTTAATCTTGTTCACAAAA

```

AGATTACTTTTCTTATTGGTCCTGAAGTCTCTGCTCATTTTTTCAAAGCT
TCTGAATCTGATCTTAGTCAGCAGGAAGTGTATCAGTTCAATGTCCCTAC
TTTTGGTCTGGAGTTGTTTTCGATGTTGATTATTCTGTTTCGTGAGGAGC
AGTTTCGGTCTTCACTGAGGCACTTAGAGTTAAAGTTGAAGGGTTAT
GTGGATATGATGGTTACTGAAGCTGAGGATTACTTCTCTAAATGGGGAGA
GAGTGGTGAAGTTGATATTAAGGTTGAGCTAGAGAGGCTCATCATCTTGA
CTGCAAGTAGATGTTTACTGGGTGAGAAAGTTCTGTGATCAGCTTTTTGAT
GATGTCTCTGCTTTGTTCCATGACCTTGACAATGGAATGCTTCCCATCAG
TGTTCTCTTCCCATATCTCCAATTCCAGCTCACCGCCGTCGTGACCGTG
CCCGAGAAAAGCTTTGAGGATTTTCGCAAAAATCATTGGGTGAGAAAA
CGCTCTGAAAAACAGAGAACGACATGCTGCAGTGTTTCATCGAATCAAA
GTACAAAGATGGTAGACAGACAACCGAATCTGAAGTCACTGGTTTGCTCA
TTGCTGCTCTGTTTGCAGGACAACACAGAGCTCTATCACTTCCACCTGG
ACCGGTGCTTATCTGATGCGATACAAAGAGTACTTCTCAGCTGCTCTTGA
TGAGCAGAAGAACCTGATTGCGAAACATGGAGACAAGATCGATCATGATA
TCTTATCCGAGATGGATGTTCTCTACCGCTGCATTAAGGAAGCGTTGAGG
CTTCACCTCCACTCATCATGTTAATGAGAGCCTCGCACAGTGATTTTCAG
CGTGACAGCTCGGGATGGAAAACTTACGATATCCCAAAGGGTCACATCG
TTGCAACCTCCCCTGCATTTGCCAACCGCTTACCGCACATCTTCAAAGAC
CCCGACACCTACGACCCAGAAAGATTCTCCCCTGGAAGAGAAGAGGACAA
AGCCGAGGGGCACTTCTGACATTGCATTGCGAGGGGGAAGGCACGGGT
GCCCTGGAGAGCCGTTTGCTTACCTGCAGATCAAAGCCATATGGAGTCAT
TTGTTGAGGAACCTTCGAGCTTGAGCTAGTTTACCCTGCTGAGATTGA
CTGGAACGCTATGGTGGTTGGAGTTAAAGGCAATGTGATGGTGCGTTACA
AGAGGCGCCAGCTTTCT - - - - -

>P

- - - ATGACTAAAGATACGGATAACAAGTTCTTGAACGTGGGTCTGCTCAT
TTTGCCCACTCTACTAGTGCCCAAGCTTATCTCTGCATTATAATGCCTA
GATCT - - - CAAAAGCGGTTGCCCTCCGGTTATGAAGGGTTGGCCT - - - TTG
ATAGGAGGGCTCATTTCGATTCTTAAAGGACCCATCGTTATGCTTCGTGA
AGAGTACCCAAACTTGGCAGTGTTTTTACAGTGAACCTAGCTAACTGGA
AGATTACTTTCTTGATTGGCCCTGAGGTGTCTGCTCACTTCTTCAAGGCC
TCCGAAGCAGATTTGAGCCAGCAGGAGGTGTATCAGTTCAATGTGCCTAC
TTTTGGTCTGGTGTGTTGTTGATGTGGATTACTCTATAAGGCAAGAGC
AATTTGATTCTTTACAGAGTCTTTAAGAGTGAGCAAACTCAAGGGATAC
GTGGATCAGATGGTGGTAGAAGCAGAGGACTACTTCTCAAAATGGGGAGA
CAGTGGGGTGGTAGACATAAAGTATGAGCTTGAGCATCTGATCATTTTGA
CAGCCAGTAGATGTCTACTTGGACGAGAAGTTCTGTATAAGCTTTTTGAT
GATGTGTCTGCTCTATTCCATGACCTTGACAATGGAATGCTCCCAATCAG
TGTTTTATTCCCATACTTGCCCATCCCAAGCTCATCGCCGCCGTGATCGGG
CACGCAAGAAGCTTGCAAAATCTTTGCAAAATATCATAAATCCCCTAAG
CTTGCTGGAAGAATCAGAGAATGACATGTTGCAAGTCTTCACTGACTCGAA
GTATAAAGATGGTCGCCCCAACCAACCGAGTCTGAGATCACAGGCTTGCTCA
TTGCTGCTCTCTTTGCTGGGCAGCACACCAAGTTCCATCACCTCCACTTGG
ACAGGAGCCTATCTTCTGCGACATAATGAATACCTATCTGCTGTGTTGGA
GGAGCAGAGAAGAACTTGATGAAAAAGCACGGGAACAAAGTTGATCAAGATA
TTTTATCCGAGATGGGTGCTTTCATCGGTGCATCAAGGAGGCCCTCAGA
CTCCATCTCCACTAATTATGCTTTTACGGAGCTCACATAGTGACTTTAG
CGTGACAACACGAGATGGGAAAGAATATGACATCCCAAAGGGCCACATAG
TCGCGACATCTCCAGCATTTGCTAACCGTCTTCTCATGTCTTCAAGGAC
CCAGAAAGGTATGATCCTGACAGATTTGCTGCTGGGAGAGAAGAAGACAA
AGCGGCAGGGGCACTTTTCATATATTTCTTTTGGAGGTGGCAGGCATGGGT
GCCTTGGCGAACCATTGCTTACTTGCAGATAAAGGCTATATGGAGTCAT
TTGCTGAGGAATTTTGAATTTGAGCTCATATCTCTTTCCCTGAGACAGA
CTGGAATGCAATGGTCGTTGGTGTGAAGGACAAGGTGATGGTGCGCTACA
AGCGGCGTGAGCTTTCAGTTAAT - - -

>R

- - - - - GATTCCGATAACAACCTAATGAACGTGCTTCTAGTAAT
CGCAGCCACACTGTAGTGCCGAAGCTTATCTCTGCATTTGTATTGCCCA
GATCT - - - AGAAAGCGTCTTCTCCGGTCGTCAAAGGCTTGCCC - - - CTG
GTCGGAGGGCTCTTCCGGTTCTTGAAGGAGCCATCATTATGCTTCGGGA
TGAGTATCCGAAGCTTGGGAGTGTTTTCACTGTCAACCTTGCTCAAAAGA
AGATCACCTTCTTTATTGGCCCTGAAGTCTCTGCTCACTTCTTCAAGGCC
CCAGAGTCTGACCTCAGCCAGCAGGAGGTTTATCAATTCAATGTTCTCTAC
TTTCGGACCTGGCGTTGTCTTTGATGTCGATTATTCATTGCGCAAGAGC
AATTCCGCTTCTTTACCGAGGCTCTTAGAGTCCATAAACTTAAGGGCTAT
GTCGATCAGATGGTCATGGAAGCCAGGATTACTTCTCAAAGTGGGGAGA
TAGTGGTGAAGTAGACTTAAAGTATGAACTCGAGCATCTGATTATTTGA
CAGCCAGTAGATGTCTTCTGGGTGAGAAAGTTCTGTATAAGCTTTTCGAT
GATGTGTCTGCTCTATTCCATGATCTGGACAATGGAATGCTCCCCATCAG
TGATTATTCCCATACCTGCCCATTCAGCTCACCGACGCCGTGACAGAG
CCCGCAAGAAGCTGGCAGAAATCTTTGCAAAATATCATAGCTTCCCGCAA
CTTGCTGAAAAACAGAGAATGACATGCTGCAGTGCTTTATTGATTGCAA
GTACAAAGATGGTCGCCCCAACCACTGAAGCCGAGGTTACTGGTCTGCTTA
TTGCTGCACTGTTTGGTGGGCAGCACACCAAGTTCCATCACCTCACTTGG

≥ 0 $\geq V$

pre-BR.DWF1.fasta

>A

-----TCGGATCTTCAGACAC
CGCTTGTGAGGCCCAAGAGGAAGAAGACTTGGGTTGATTACTTTGTCAAG
TTCAGATGGATCATTGTCTATCTTCATCGTCTTCCATTCTCAGCCACATT
CTACTTCCATCTACCTCGGGGACATGTGGTCAGAGTCCAAGTCCCTTG
AGAAACGTGAGAAGAACACGACGAGAATGTCAAGAAAGTCATCAAAGG
CTTAAGGGTAGGGATGCTTCCAAGGACGGGCTTGTCTGCACTGCTCGTAA
GCCCTGGATCGCTGTTGGAATGAGGAACGTTGACTACAAGAGAGCCCGGC
ATTTGAGGTTGACTTGGGGGAGTTCCGTAACATCCTTGAGATCAACAAG
GAGAAGATGACTGCTAGAGTGGAGCCTCTTGTTAACATGGGACAGATTTT
CCGTGCTACCGTCCCAATGAACCTGTCTCTCGCTGTTGTTGCTGAGCTTG
ATGACCTTACCGTTGGTGGACTTATCAATGGATATGGTATTGAAGGAAGC
TCTCACATCTACGGTTTGTGCTGATACCGTTGAGGCTTACGAGATTGT
TCTTGCGGGTGGAGAGCTTGTCCGCGCCACAAGGGATAATGAGTATTCTG
ATCTTTACTACGCAATCCCGTGGTCGCAAGGAACCTTTGGACTCCTTGTA
GCTGCTGAGATCAGGCTTATTAAAGTCAAGGAGTACATGAGACTCACTTA
CATACCAGTCAAGGGTGATCTTCAAGCCTTAGCTCAAGGTTACATTGATT
CTTTTGCTCCCAAGACGGTGAC-----AAGTCGAAAATCCCGGAT
TTCGTCGAAGGCATGGTTTACAATCCAACGGAAGGAGTGATGATGGTTGG
AACATATGCATCTAAAGAAGAGGCAAGAAAGAAAGGGAACAAATCAACA
ATGTGGGATGGTGGTTCAAGCGTGGTTCTACCAGCACGCGCAGACCGCC
CTGAAAAAGGGACAGTTTGTGAGTACATCCCAACTCGTGAATACCA
CAGGCACACAAGGTGCTTGTACTGGGAAGGGAAGCTTATTCTCCATTTG
GTGATCAGTTCTGGTTTAGGTACCTCTTAGGTTGGTTGATGCCTCCAAAG
GTCTCTCTTCTTAAGGCCACTCAAGGTGAAGCTATCAGGAATATTACCA
TGATATGCATGTTATTACGGATATGCTTGTTCCTCTTTACAAGGTTGGCG
ATGCACTCGAATGGGTCCACCGCGAAATGGAGGTGTATCCAATTTGGCTT
TGCCACACAAACTCTTCAAGCAGCCAATCAAGGCCAAATCTACCCAGA
GCCAGGCTTCGAGTACGAAAAACAGACAAGGAGACACAGAAGATGCACAGA
TGTACACTGATGTTGGAGTCTACTACGCACCTGGCTGTGTCTTAAGAGGT
GAAGAGTTTGATGGATCAGAAGCAGTGCCTAGGATGGAGAAATGGCTGAT
AGAGAACCATGGATTCCAGCCTCAGTACGCGGTGTCTGAGCTCGACGAGA
AGAGCTTCTGGAGAATGTTTAAATGGTGAATTGTATGAGGAGTGCCGCAAG
AAGTATAGAGCTATTGGAAACGTTTCATGAGTGTGTTACTACAAGTCCAAGAA
AGGAAGGAAGACTGAGAAAGAAGTTAGAGAAGCCGAACAAGCTCATCTCG
AAACTGCTTATGCCGAGGCAGAT-----

>P

-----ATGTCCGATCTTGAGGCCC
CCTTG---CGCCCGAAGAGGAAGAAGGTGTGGGTAGACTATTTTGTGCAG
TTCCGATGGATTTTAGTTATTTTCGTCGTTCTCCCAATCTCCTTCACCT
TTACTTTCTCACTTATCTCGGGGATGTCAAATCAGAGATGAAGTCCTTACA
AACAGCGTCAAAGGAACATGATGAAATGTTAAAAAAGTTGTGAAACGT
CTTAAGAGAGGAATCCGTCAAAGGATGGTCTTGTCTGCACTGCTCGTAA
ACCCTGGATTGCTGTTGGAATGCGAAATGTTGACTACAAACGGGCTCGGC
ACTTCGAAGTTGATTATCAGCTTTTCGTAATATCCTTGAAATTGATAAG
GAGAGAATGATTGCTAGAGTTGAGCCTCTTGTAATATGGGTCAGATCAG
CAGGGCAAGTGTCCCAATGAATCTTCTCTTGCACTGCTTGCAGAACTTG
ATGATCTCACTGTTGGTGGGCTCATTAAATGGTTATGGGATTGAAGGAAGC
TCTCATATCTATGGCTTGTCTCCGACACTGTTGTGGCATATGAGATTGT
TTTGGCAGATGGTCAGGTTGTTAGAGCTACCAAGGACAATGAATACTCTG
ATCTTTTCTATGCCATCCCTGGTCTCAGGGAACACTTGGGCTTCTTGTC
TCTGCTGAGATCAAGCTTATTCCTATTAAGGAATACATGAGGCTGACCTA
CAAACCTGTAGTGGGTAATCTGAAAGAAGTTGCGCAGGCCTATATAGACT
CTTTTGACCCAGAGATGGAGATCAGGATAACCCAGAGAAGGTTCCAGAC
TTTGTGGAGACTATGATTATAACTCTACCGAAGGTGTGATGATGACTGG
GAGATATGCCTCCAAAGAAGAGGCCAAGAAGAAGGGAATGTGATTAA
ATGTTGGTTGTTGGTTTAAACCGTGGTTCTATCAGCATGCGCAGACGCC
CTAAAGAAAGGAGAGTTTGTGGAGTACATTCCAACCAGAGAATATTATCA
CAGGCACACAAGGTGTTGTACTGGGAGGGGAAGCTTATACTTCCATTG
CTGATCAATGGTGGTTTAGATTCTCTTAGGCTGGATGATGCCTCCAAAG
GTTTCTCTTCTCAAGGCCACTCAAGGTGAAGCAATCCGAACTATTACCA
TGAGATGCATGTCATTACGGACATGCTTGTTCCTCTTTACAAGGTTGGGG
ATGCCCTTGAATGGGTCCACCAAGAGATGGAGGTATACCCCTTTGGCTC
TGTCGCGACAGGTTGTTCAAGCTCCCTGTGAAAATATGGTGTATCCAGA
GCCAGGATTCGAGCATCAGCACAAACAGGGAGATACATCCTATGCTCAGA
TGTACACTGATGTCGGGGTGTATTATTCGCCCCGACCTGTGCTGAGGGGT
GAGGTGTTTGATGGTGCAGATGCTGTTCTGTAATGGAGGACTGGTTGAT
AGAAAACCGTGGCTTCCAGCCACAGTATGCAGTGTCCGAGCTAAGTGAGA
AGAAATCTGGAGGATGTTTGTGGTGACCTCTATGAACACTGCAGGAAG
AAATATGGAGCTGTGGGAACCTTCATGAGCGTGTACTACAAATCCAAGAA

AGGAAGGAAGACAGAGAAGGAGGTGCAGGAAGCTGAACAAGCCACCTTG
 AGACTGCTTATGCAGAGGCTACT-----
 >R

 -----TCAGATCTTGAGGCC
 CCCTG---CGTCCAAAGAGGAAAAAGGTGTTGGTGGACTATTTTGTTCAG
 TTCCGATGGATTTTGGTTATTTTGTGCTCCTGCCAATCTCCTTTACAAT
 TTAATTTCTCACTTATCTTGGGGATGTTAAATCCTCAATGAAGTCTTACA
 AGCAACGCCAGAAAGGAACATGATGAAAATGTTAAGAAAGTTGTGAAACGT
 CTAAGCAGAGGAATCCATCAAAGGATGGTCTTGTATGCACAGCCCGTAA
 ACCCTGGATTGCCGTTGGAATGCGGAATGTAGACTATAAGCGGGCTCGGC
 ACTTTGAAGTTGATTTGTGAGCCTTCCGTCATATTCTTGACATTGATAAA
 GACAGAATGATTGCAAAAAGTGGAGCCCTTGTCAACATGGGGCAGATCAC
 TAGGGCCACAGTCCCAATGAATCTTGCCCTTGCAAGTGGTGTGAGCTTG
 ATGATCTTACTGTTGGTGGGCTCATCAACGGTTATGGGATTGAAGGAAGC
 TCTCATATATATGGCTTGTCTTCTGACACTGTCGTAGCTTATGAGATTGT
 TCTGGCAGATGGTGGGTTGTTAGGGCTACCAAGGACAATGAATATTCTG
 ATCTTTTCTATGGCATCCCATGGTCTCAAGGAACACTTGGGCTTCTTGTCT
 TCTGCTGAGATTAAGCTTATACCCATTAAAGAGTATATGAAGCTTACATA
 CAAACCTGCAAGTGGGCAATCTAAAAGATCTTGACAGGCATATGTGGACT
 CTTTTCGCCCCAAAGATCTTGACCAGGATAACCCAGACAAGGTTCTGAT
 TTTGTAGAGGGCATGATTTATTCTCAACTGAAGGTGTAATGATGACAGG
 GAGGTATGCATCTAAAGAAGAAGCCAAGAAGAAGGGCAATGTTATTAACA
 GTGTTGGTTGGTGGTTTAAACCTTGGTTTACCAGCATGCACAAACAGCC
 TTAAGAAGAGGAGGTTTGTAGAGTACATCCCAACAGGGAATATTACCA
 CAGGCACACAAGGTGTTTGTACTGGGAGGGAAGCTTATTCTTCCATTG
 GCGACCAAGTGGTGGTTAGGTTTCTCTTGGCTGGTTGATGCCTCCCAAG
 GTTTCACCTGCTTAAAGCTACCCAAAGGTGAAGCAATCAGAAACTATTACCA
 TGAGAATCATGTCTTACAGGATATGCTTGTCTCTTTACAAGGTTGGAG
 ATGCCCTATTGTGGGTCCACCGTGAGATGGAGGTATATCCCATTTGGCTC
 TGTCCACACAGATTGTTCAAGCTGCCCATGAAACTATGGTGTATCCTGA
 ACCAGGATTTGAGCATACGCGCAGACAGGGGGACAGTCCATATGCTCAAA
 TGTACACAGATATTGGGGTGTATTATGCACCGGGACCTATATTAAGAGGT
 GAGGTGTTTGTGTTGCTGAGGCAGTTCTGATGATGGAGAGTTGGTTGAT
 TGAGAACCATGGATTCCAGCCACAATATGCAGTTTCCGAGCTGAATGAGA
 AGAATTTCTGGAGAATGTTTGTGCTGGGCTGTACGAGCAGTGCAGAAAG
 AAATACGGAGCTGTGCGAACCTTCATGAGTGTCTACTACAAGTCCAAGAA
 AGGAAGGAAGACTGAGAAGGAGGTGCAAGAAGCTGAGCAAGCCACCTGG
 AACCTGCTTACGCAGAAGTCGATCAACCAGCAGAC
 >0
 GTAAATGGAAGTGTGATGGATGTAACGTGTGTCATACAGCGGAGAAGAT
 ATTAATCAATAATCATAGCCGGACTGCAGCCATGGCAGATCTGCAGGAGC
 CCCTCGTTCTGCCGAAGAGGAAGAAGGTTTGGTGGACTACTTGGTAAAG
 TTCCGATGGATTCTGGTGATCTTGTGGTGTCTCCCATTTCCGCTCTGAT
 CTACTTCAATATCTATTTGGGCGATGTCTGGTCT--GCCAATCTGAGA
 AACGTGCGCCAGAAAGGAACATGATGACAATGTGCAAAAAGTTGTGAAGCG
 CTAAGCAGAGGAACCCAAAGAAGGATGGCCTTGTGTCACAGCTAGGAA
 GCCTGGAATTGCTGTTGGCATGCGCAATGTAGACTACAAGCGTGTAGGC
 ATTTTGAGGTTGACCTTTCCGCTTCAGGAACATTCTTGAGATTGACAGA
 GAGAGAATGGTTGCCAAGGTTGAGCCTTGTGCAACATGGGCCAGATAAC
 CAGAGCTACATGCCCAATGAACCTTGCCCTTGCAAGTTGTTGCTGAGCTTG
 ATGACCTTACTGTTGGGGGACTGATCAATGGGTATGGTATTGAAGGGAGC
 TCTCACCTCTATGGTCTTTCTCTGACACTGTTGTGCGCGTGGAAAGTTGT
 TCTTGCAGACGGTGGAGTTGTTAGAGCCACTAAGGATAATGAGTACTCTG
 ACCTTTTCTATGGCATTCCCTGGTCCAGGGAACACTTGGGTTTCTTGT
 TCCGCTGAGATCAAACCTATTCCCATCAAGGAATACATGAGGCTCACATA
 TACTCCAGTTAAAGGGTCACTGAAGGAGATAGCACAAGGTTATTGTGATT
 CGTTTGCACACGAGATGGTGAT-----CCTGCAAGGTCACAGAC
 TTCGTTGAGGGAATGGGTACACAGAAAATGAGGGTGTGATGACTGG
 TGTTTATGCTTCCAAAGAAGAGGCAAGAAGAAGGGCAATAAGATCAACT
 GTGTGCGGTGGTGGTTCAAGCCTTGGTTTACCAACATGCTCAGACAGCA
 CTCAAGAAAGGGTGAAGTTTGTGGAGTACATTCCAACAAGAGAGTACTACCA
 CCGTCAACCCGGTGTCTGTACTGGGAGGGGAAGCTGATCTTGCCATTG
 GCGACCAATTCTGGTTCAGGTTCTCTTGGGCTGGCTGATGCCACCAAAG
 GTGTCTCTGCTCAAGGCCACACAGGGTGAATCTATCAGGAATTACTACCA
 TGACAACCATGTGATTCAAGACATGCTGGTTCCCTTGTACAAAGTTGGAG
 ATGCTCTTGAAGTTTGTTCACAAGGAA---GAGGTTTATCCACTGTGGCTG
 TGCCCGCACCGGCTCTACAAGCTCCCTGTGAAAACC---GTGTACCCAGA
 GCCTGGCTTTGAGCACCACACAGGCAAGGTGACACTAGCTATGCCCAGA
 TGTTACCCGATGTTGGTGTGACTATGCTCCTGGTGTGCTGAGGGGC
 GAGGAGTTCAATGGCGCTAGCTGTCCACAGGCTGGAGCAGTGGCTGAT
 TGAGAACCACAGCTACCAGCCACAGTACGCTGTATCTGAGCTCAACGAGA
 AGGACTTCTGGAGGATGTTTGTGCTTCTCACTACGAGCATTGCCGCCAA
 AAGTATGGTGCCGTCGGTACCTTTATGAGCGTCTACTACAAGTCCAAGAA
 GGAAGGAAGACTGAGAAGGAGGTGCAGGAAGCCGAGGCCGCATCCTCG

```

AGCCAGCCTACGCTGATGAGGCG-----
>V
-----TCGGATCTTCAGGCTC
CCTTG---CGTCCCAAGAGGAAAAAATTTGGGTGGACTATTTTGTTTCAC
TTCCGATGGATTATTGTCATTTTTGTTGTCCTTCTATCTCCTTCACTTT
GTACTTCCCTCACATATCTTGGAGATGTCAGATCTGAATCAAAATCTTTCA
AGCAGCGTCAGGAGGAACATAATGAAAATGTCAAAAAGTCATAAAACGT
CTCAAAGAGAGGAACCCATCAAGGGATGGCCTTGCTGTCACAGCCCGGAA
ACCATGGATTGCTGTTGGAATGAGAAATGTTGACTATAAGCGGGCTCGGC
ATTTTGAAGTTGATCTTTGAGCTTTGAGAAATATCCTGGACATTGACAAA
GAGAGAATGATTGCCAGATGTGAACCCCTAGTCAACATGGGGCAGATTAG
CAGGGTTAGTGTCCTCAATGAATCTTGCCCTTGCTGTGGTTGCTGAGCTTG
ATGATCTTACAGTTGGTGGCCTCATCAATGGCTATGGAATTGAAGGAAGC
TCTCACATTTATGGCCTATTCTGACACTGTTGTGGCTTATGAAATCAT
TTTGGCTGATGGCGGGCTAGTTAGAGCTACCAAAGACAATGAGTACTCTG
ATCTTTTCTATGCTATTCCATGGTCTCAGGGAACACTGGGGCTTCTTGT
GCCGCTGAGATCAAGCTTATACCCATTAAGGAATACATGAAGTTGACTTA
CAAACAGTAGTGGGAAATCTGAAAGACCTTGCGCAG-----
-----GACGGAGATCAGGATAATCCTGAGAAGGTTCCAGAC
TTGTAGAAACCATGATTTACAATCCTACTGAAGCTGTGTATGACAGG
GAGATATGCCTCAAAAGAAGAGGCTAAGAAAGAAAGGAAATGTGATTAAAC
GTGTTGGGTGGGTGACAAGCCCTGGTTCTATCAACATGCACAGACAGCC
CTAAAGAAAGGGGAGTTTGTGGAGTACATCCCAACAGGGAATATTACCA
TAGGCACACTAGGTGCTGTATTGGGAGGGAAAACTTATTCTTCCATTG
CAGATCAATGGTGGTTTAGGTTTTGTTTGGGTGGTTGATGCCACCAAAG
GTTTCTCTCTCAAGGCTACTCAAGGTGAAGCTATCAGAACTATTACCA
TGAGATGCATGTAATTCAGGACATGCTTGTTCGCTGTACAAAGTTGGGG
ATGCTCTAGAATGGGTACATCATGAGATGGAGGTATACCAATTTGGCTC
TGCCACACCGATTGTACAAGCTTCCT-----
-----GGTGACACACATTATGCCCAGA
TGATACACAGATGTGGGGGTGTACTATGCGCCAGGGCCTGTCTTGAGGGGT
GAGCAGTTTGATGGTGCAGAAGCAGTTCGCCGAATGGAGAAGTGGTTGAT
TGAAAACCATGGATTCCAGCCACAATATGCAGTGTCTGAGCTGACTGAAA
AGAACTTCTGGAGGATGTTTGATGCTGGGCTTTATGAGCACTGCAGAAAG
AAGTATGGAGCAGTGGGGACTTTTATGAGTGTTTACTACAAATGCAAGAA
GGGGAAGAAGACCGAGAAGGAAGTGCAGGAGGCGGAGCAAGCACAACCTG
AGACACCTTATGCTGAGGCTGAT-----

```

```

pre-BR.DWF5.fasta
>A

```

```

-----GCGGAGACTGTACATTCTCCGATCGTTA
CTTACGCATCGATGTTATCGCTTCTCGCCTTCTGTCCACCTTTCTGTCATT
CTCCTATGGTACACAATGGTTCATCAGGATGGTTCTGTTACTCAGACCTT
TGCTTCTTTTGGGAGAATGGA---GTTCAAGGACTTATCAACATATGGC
CAAGACCCACTTTGATTGCTTGGAAAAATTATATTTTGTATGGAGCATT
GAAGCTATTCTTTCAGCTGCTTCTGCCTGGTAAAAGAGTTGAGGGTCCAAT
ATCTCCAGCCGGAACCGACAGTTTACAAGGCCAATGGTCTGGCTGCTT
ACTTTGTGACACTAGCAACCTATCTTGGTCTTTGGTGGTTTGAATCTTC
AACCTGCAATTGTCTATGATCACTTGGGTGAAATATTTTCGGCACTAAT
ATTCGGAAGCTTCATATTTTGTGTTTTGTTGTACATAAAAGGCCATGTTG
CACCTTCATCAAGTGAATCTGGTTCATGTGGTAACCTAATAATTGACTTC
TATTGGGGCATGGAGTTGTACCCTCGAATTGGTAAGAGCTTTGACATCAA
GGTGTTTACTAATTGCAGATTCGGAATGATGTCTTGGGCAGTTCTTGACAG
TCACGTAATGCATAAAACAGTATGAAATAAATGGCAAAGTATCTGATTCA
ATGCTGGTGAACACCATCCTGATGCTGGTGTATGTCACAAAATCTTCTG
GTGGGAAGCTGGTTATTGGAACACCATGGACATTGCATGACCGAGCTG
GATTCTATATATGCTGGGGTTGTCTAGTGTGGGTGCCTTCTGTCTACACT
TCTCCAGGCAATGTACCTGTGAACACCCCGTCAAACTCGGAACTCAGTT
GGCAATATACATTCTCGTTGCAGGAATTCTGTGCATTTACATAAACTATG
ACTGTGATAGACAAAGGCAAGAGTTGAGGAGGACAAACGGGAAATGTTTG
GTTTGGGGAAGGGCCCCGTCAAAGATTGTGGCGTCGTATACTACAACATC
TGGTGAACA---ACTAGTCTTCTCTTAACGTCTGGATGGTGGGATTGG
CTCGTCATTTTCCATTATGTTTCTGAGATCTTAAGTGCTTTCTTCTGGACC
GTACCGGCTCTCTTCGATAACTTCTTGGCATACTTCTACGTCATATTTCT
CACCTTCTTCTCTTTGATCGAGCCAAGAGAGACGATGACCGATGCCGAT
CAAAGTATGGGAAATATTGGAAGCTGTATTGTGAGAAAGTC-----
-----
-----
-----
-----

```


-----AGGATTATCCCTGGAATCT
AC

pre-BR.HYD1.fasta

```
>A
---GAGGAG-----TTGGCGCATCCGTACGTTCCGAGAGATCTGAACCT
ACCCGGATACGTACCAATCTCAATGTCAATGTCCTCCATCGTCTCTATCT
ACCTCGGTTCTTCCCTCCTTGTGTCTCCCTCGTCTGGCTTCTCTCGGG
AGG-----AAGAAAGCTAAACTTGATAAGTTGCTTATGTGTTGGTG
GACATTCACCTGGTCTCACTCATGTTATTCTCGAGGGCTATTTGTTTTCT
CCCTGAGTTTTTCAAGGACAACACTTCTGCTTATCTTGCTGAAGTTTG
AAAGAATATAGCAAAGGTGATTGAGATACGTAGGTAGAGATTCTGCAGT
TGATATCTGTTGAAGGGATCACTGCTGTTATCGTTGGCCAGCTTCTCTCC
TAGCTATTTATGCCATTGCTAAGGAGAAGTCGTATAGCTACGTGCTTCAG
CTTGCATTTTCGGTTTGCCAGCTCTACGGATGTTTGGTTTATTTCAATTAC
TGCTATCTTGGGAAGGAGACAACCTTGCCACAACCTCTTCTACTATTACT
CCTACTACATCGGGGCTAACTGTTGGTGGGTCTTGATACCTTCACTCATT
TCTTTCGTTGCTGGAAGAAAGATTTGTGACGCTGCTGCCATTGCCAACAA
CAACGTCGAGACAAAGACGAAGAAGAAACCCGT
>P
ATGGAGGGG-----CAGCCGCATCCATACGCACCAAGAGATCTGAAACT
GCCTGGCTATGTACCTAATTTCCCTCACTCAATCCACCATTTGTCGGCGTCT
ATTTACTCACCTCCCTTCTTGTGCTCTCTCATCTGGATCCTCTCTGGA
AGG---TCTCGTAAGATAACAAAA---AATAGGTTGCTCATGTGCTGGTG
GGCTTTCACTGGTCTCACACACTTGATTCTCGAGGGTTATTTGCTTTCT
CTCCAGAATTTTACAAGGACAAGACTGCTCATTATCTGGCTGAAGTTTGG
AAAGAATATAGCAAAGGTGATTCAAGATATGCAGCAAGGGATGCTGCAAC
AGTTACTGTTGAAGGATTGACTGCTGTTCTCGAGGGACCAAGCTAGCCTCC
TGGCAGTGATGCTATTGCTTCAGGAAAATCGTACAGCTACATACTTCAG
TTTGCTGTTTGTGTTGGGACAGCTCTATGGAACAGCTGTGATTTCTTAAC
TGCCTACTTGGGAAGGTGATCATTTTGTACCAAGTCCGTACCACTATTACG
TATACTATATTGGTGCAAATGCCTCCTGGGTTGTAATTCCTCGCTCATC
GCTATGCGTTGTTGGAAGAAAGATTTGCTCAGCAGTC-----
-CAAGTTCATGGCCAGAAAAGGACCAAACTCGT
>R
---GAGGGG-----CAAGCGCATCCGTACGTGCCACAGATCTGAAACT
GGAGGACTACGTCCCTGTATTTCTCTCACAGTCGACAATCCTTGCCGTCT
TCGGAACCTCGTCCCTTGTCTCCTCGTCTCCCTCGTCTGGTTCAATTTCTGGA
AGT---TCGCGTAAAGCAGCTAAAATTGATAGATGGTTG---TGCTGGTG
GGCTTCACTGGTCTCACACACATTATCCTAGAGGGTTATTTTGTGTTTCT
CCCTGAGTTTTTACAAGGATAAGACTTCTTTTTATCTTGCTGAAGTTTGG
AAAGAATACAGTAAAGGTGATTCAAGATATGCTGCAAGAGATTCTGCAGT
AGTGGCTGTGGAAGGATTGACCGCAGTTTTAGAGGGTCCAGCAAGTCTTC
TGGCAGTCTATGCTATTGCTGCTGGGAAGTCGTATAGCTACATACTTCAG
TTTGCCATTTCACTCGGGCAGCTTTATGGAAGTGTGTATTTCTTAAC
TCCTACTTGGATGGCTATAACTATGCATCGAGTCCATATTACCATATG
CATACTACATCGGTGCTAATGCTTCCTGGGTTGTTATACCATCACTCATT
TGTATCCGTAGCTGGAAGAAGATATGTGACGATTT-----
-CAAGTGAAGGCCAGAGAAAGACCAAAAGTGGT
>O
-----GGGCACCCCCACCCCCACCTTACGCGCCGGCGGAGCTTACCT
CCCGGGCTTCGTGCTCTCCAATGTCCAGGCCCAAATCCTCGTGCCCT
ACCTCGCCACCTCCCTCTTCTCCTCCTCGCCGTCTGGCTCATCTCCGGG
AGATGCAGTCGTAGGCTTTCGACACCGACCGCTGGCTCATGTGCTGGTG
GGCTTCAACGGCTCACCCACATTATCATCGAGGGAACCTTTGTCTTTG
CTCCTAATTTCTTCTCCAACCAAACCTTCTTACTTCGATGAAGTTTG
AAAGAGTACAGCAAAGGTGACTCCAGATATGTCGCCAGAGACCTGTACT
TGTTACAGTTGAAGGAATTACAGCTGTCTTGAAGGCCCTGCTTCACTCC
TTGCTGTCTATGCCATCGCATCGGGCAAGTCTTACAGCCATATCCTCCAG
TTCACTGTCTGTCTTGGTCAGCTCTATGGATGCCTGGTGTACTTTATTAC
AGCCTACTTGGATGGCTTCAACTTCTGGACTAGCCCGTTCTACTTCTGGG
CTTATTTCAATTGGTGAAACAGCTCGTGGGTTGTTATACCAACTATGATC
GCCATAAGGAGCTGGAAGAAGATTTGTGACGATTT-----
-----CAAGGTGAA---AAGGTGAAGACTAAA
>V
---GAGGGAGCGCTCAAAGCCACCCATATGTGCCAAAAGATCTAAAAC
GCCAGGTTATGTGCAAGTTGTCTCTCACAGTCCACCATTTGTCGGAAGTTT
ATGGCTCTCTTCTGTGCTTGTGTATCTGTAATATGGATCATCTCTGGG
CGG---TTTCCGAAAACATCTAAGATTGATAGATTGCTCATGTGCTGGTG
GGCATTACGGGTCTAACACATGCAATCCTTGAGGGCTATTTTGTGTTCT
CTCCAGAATTTACAAGGATAAACTGCTTGTACCTGGCTGAAGTGTGG
AAAGAATACAGCAAAGGTGATTCAAGATATGCAGCCAGAGATGCTGGCGT
CATTGCTGTTGAAGGGATTACTGCAGTTTTAGAAGGTCCAGTAGCCTTC
```

TAGCACTGTATACTATTGCCACAAAAAGGTCATACAGCTACATACTTCAG
 TTGACCATTCTCTGGGCGAGCTCTATGGCACTGCTGTATATTTATAAC
 ATCTTACTTGGATGGTGACAATTTTCGCTGCAAGCGCATATTACTACTATG
 CATACTACGTTTTTGCAAATGCTTTTTGGGTTGTAATACCCTCACTCATT
 GTTATCCGTTGTTGGAAGAAAATATGTGCAGCATT-----
 -CAGGTCCAAGGTCAGCAAAAGACCAAGTTCGG

pre-BR.SMT1.fasta

>A
 -----GATCTCGCGTCGAATCTTGGTGGGAAGATCGA
 TAAATCCGATGTTCTCACGCGCGTCGAAAAGTATGAGCAATATCATGTCT
 TCCATGGAGGAAACGAGGAAGAGAGAAAAGCCAATTATACTGACATGGTT
 AATAAGTACTATGACCTTGCTACTAGCTTTTATGAGTACGGATGGGGAGA
 ATCCTTCCATTTTGACAAAAGATGGAAGGAGAATCGCTTCGAGAGAGTA
 TTAAGCGACACGAGCACTTTCTTGCTCTTACGCTTGGCATCCAACGAGGA
 CAAAAGGTTACTGGATGTAGGATGTGGAATTGGTGGACCGTTGAGGGAAAT
 TGCAAGATTTCAGCAATTCAGTTGTTACCGGGCTCAACAATAACGAATACC
 AGATCACCAGAGGCAAGGAACTAAACCGACTTGCAAGGTGTCGACAAGACA
 TGTAACTTTGTCAAGGCTGACTTCATGAAGATGCCATTCCCGGAAAACAG
 TTTGATGTCAGTTTATGCAATAGAAGCAACTTGCCATGCACCTGATGCGT
 ATGGATGCTACAAAGAGATCTACAGAGTGCTAAAGCCTGGGCAATGTTTT
 GCTGCCTATGAGTGGTGTATGACTGATGCATTTGACCTGATAACGCCGA
 ACATCAGAAAATAAAGGGAGAGATAGAGATTGGAGATGGTCTTCCTGACA
 TTAGGCTGACTACAAAATGCCTCGAAGCTCTGAAGCAGGCCGGTTTTGAA
 GTGATATGGGAAAAGGATCTGGCCAAGGACTCGCCGGTCCCATGGTACTT
 ACCTCTTGACAAAAATCATTTCTCACTCAGTAGCTTCCGCCCTACAGCTG
 TTGGACGATTTCATAACCAAAAACATGGTCAAGATCCTTGAATACATAAGA
 CTTGCACCTCAAGGAAGCCAAAGGGTCTCAAATTTCTGAGCAGGCTGC
 GGAAGGATTAGTCGACGGTGAAGGAGAGAGATTTTCACGCCAATGTATT
 TCTTCTTGGCCCGGAAGCCAGAG-----

>P
 -----TTGGATCTCGCTACTGGTCTTGGCGGAAAGATCGA
 TAAAAGTGATGTTCTCTCTGCTGTCGAGAAAGTATGAGAAGTATCATGTCT
 GTTATGGAGGTGATGAGGAAGAGAGAAAAGCCAACCTACTCTGACATGGTT
 ---AAATATTATGATCTTGTTACCAGCTTTTATGAGTTTGGCTGGGGGGA
 GTCTTTCCATTTTGTCTCTAGATTCAAAGGGGAGTCTCTTCGCGAGAGCA
 TCAAGCGACATGAACACTTTCTCGCTTTACAACCTTGGCCTGAAACCTGGA
 CAGAAGGTGTTGGATGTAGGATGTGGAATTGGCGGGCCTTTAAGAGAAAT
 TGTCTGATTCAGCTCAACATCAGTAACAGGGTTGAACAACAATGAGTATC
 AGATATCAAGGGGAAAGGAACTAAATCGCACTGCAGGAGTTGACAGGACC
 TGTGATTTTGTGAAGGGTGATTTTCATGAAAATGCCATTCCCTGATAACAT
 TTAT---GCTGTATATGCAATTGAAGCTACCTGCCATGCACCAAGATGCAT
 ATGGATGCTACAGTGAGATTTACAGGGTATTGAAGCCAGGACAATGTTTT
 GCTGCATATGAGTGGTGCATGACTGATTCTTTGATCCACATAACCAAGA
 ACATCAAAAAATTAAGGCAGAAATAGAGATTGGTGATGGCCTTCAGACA
 TCAGGTTGACTGGACAATGCATTGATGCTCTAACGAAAGCTGGTTTTGAG
 GTCATATGGTCAAAAGATCTTGCAAGTGGGCTACCTGTGCCATGGTACTT
 GCCTTTGGACACAAGTCACTTCTCACTGAGTAGCTTCCGTTTAAAGCTG
 TTGGGAGATTCTTACAAAAAATATGGTCAAAGCCCTGGAATTTGTAGGA
 CTTGCTCCCAAGGAAGTCAAAGGGTTCAAGATTTTCTTGAGAAGGCTGC
 AGAGGGATTAGTTGAAGGTGAAGGAAGGAGATTTTACCCCGCTGTTTT
 TCTTCGTGGCTCGGAAGCCA-----

>R
 TCTAAAGCCGGCGCATTGGATCTCGCCTCCGGTGTGGTGGTAAGATCGA
 TAAAAACGAAGTTCTCTCCGCCGTCGAAAAATATGAGAAGTATCATGTCT
 ATTATGGAGGTGCAGAAGAAGAGAGAAAAGCCAACCTACACTGACATGGTC
 AATAAACTACTATGATCTTGTTACTAGCTTTTATGAGTTCCGGCTGGGGAGA
 GTCTTTCCATTTTGCACACAGATTTAAAGGGGAGTCTCTTCGAGAAAAGCA
 TCAAACGACATGAACACTTTCTGGCTTTACAGCTTGGCCTGAAACCTGGA
 CAGAAGGTTTTGGATGTTGGATGTGGAATTGGTGGACCACTAAGAGAAAT
 TTCTCGTTTCAGTTTGACATCAGTTACAGGGTTGAACAACAATGAATATC
 AGATCACAAGAGGAAAGGTAATAATCGCATAGCAGGAGTGGACAAGACC
 TGTGATTTGTAAAGGCCGACTTCATGAAAATGCCATTTCTCGCCAATAG
 TTTTGATGCAGTGATGCAATTGAAGCTACTTGCCATGCACCAAGATGCAT
 ACGGATGCTACAGTGAGATTTACAGAGTATTGAAGCCTGGGCAATTTTTT
 GCCGCATATGAGTGGTGCATGACTGATTCTTTGATCCCAATAACCAAGA
 ACATCAGAAAAATTAAGGCAGAGATCGAAATTGGCGATGGCCTTCCAGATA
 TTAGGTTAACTGGGCAATGCATTGAAGCTCTCAAAAAAGCTGGTTTTGAT
 GTCATATGGTCAGGTGATCTTACAGCAAGCTCACCTGTCCCATGGTACTT
 GCCTCTGGACAAAAATCAATACTCTGACTGGTTTCCGGTTAACAGCTA
 TTGGAAGATTTTCACTAGAAACATGATAAAGGCTCTAGAATTTGCAGGA

CTTGCTCCGAAGGGAAGTCAAAGGGTTCAAGATTTTCTTGAGAAGGCTGC
GGAGGGATTAGTGGAAGGTGGAAGGAAGGAGATTTTCACACCAATGTTCT
TCTTCCTGGCTCAGAAACCACATTCAGAGAACCAG-----

>0

TCGAGGTCTGGGGCCATGGATCTTGCATCTGGACTCGGGGGGAAGTCAC
CAAGGATGAGGTCAAGTCGGCTGTCGATGAGTATGAGAAATACCATGGAT
ACTATGGAGGAAAAGAAGCAAGAAAGTCCAACTACACAGACATGGTT
AATAAACTATGATCTTGCCACCAGCTTCTATGAGTATGGTTGGGGTGA
ATCCTTCCACTTTGCTCACAGATGGAATGGAGAATCCCTACGTGAAAGCA
TTAAGCGACATGAGCACTTTCTTGCCCTTCAGCTTGGAGTGAACCAGGA
ATGAAGGTTTTGGACGTCGGTTGTGGAATAGGCGGACCACTAAGAGAAAT
TGCCAAATTTAGCTTGGCTTCAGTCACTGGATTGAACAACAATGAGTACC
AGATAACTAGGGGAAAAGGAGCTTAATCGGGTAGCAGGAGTTAGTGGAAT
TGCGACTTTGTGAAGGCAGACTTCATGAAGATGCCGTTTTCTGATAACAC
TTTTGATGCTGTCTATGCCATTGAGGCAACATGCCACGCACCTGATCCGG
TTGGCTGCTATAAGGAGATCTATCGTGTATTGAAGCCTGGCCAGTGTTTT
GCTGTGTATGAGTGGTGTATTACTGATCACTATGAACCGAACAATGCAAC
ACACAAGAGGATCAAGGATGAAATCGAGCTTGGCAATGGCCTGCCAGATA
TCGAAGCACTCAACAATGTCTTCAAGCTGCAAAAGATGCTGGGTTTGAG
GTTATTTGGGATAAAGACCTTGCTGAAGATTCCCCGGTTCCTTGGTACCT
GCCCTTGGATCCAAGCCGTTTTCTTGAGTAGCTTCCGTCTGACTACTG
TGGGGCGTGCAATTACTCGACAATGGTGAAGGCATTGGAATACGTCGGC
CTTGCCCCACAAGGAAGTGAAAGGGTATCTAATTTTTAGAGAAGGCAGC
TGAAGGTCTCGTGAAGGTGGCAAGAAGGAAATCTTCACGCCGATGTACT
TCTTTTTGGTTTCGAAGCCTATCTCAGAA-----

>V

TCGAAGGCTGGAGCTTTGGATCTCGCATCTGGTGTGCGCGGGAAGATTCA
GAAAGATGAAGTGCTTTCCGCTGTTGAAAAGTATGAGAAGTATCATGTCT
GTTATGGTGGTGAAGAAGAAGAGAGGAAAGCTAATTACAGTGACATGGTT
AATAAATATTATGATCTTGTACCAGCTTCTATGAATTTGGATGGGGAGA
GTCTTTCCATTTTGACCCAGATGGAAGGGGAGTCTCTCCGAGAGAGCA
TCAGGCGACATGAGCACTTCCTTGCTTTGCAACTAGGTGTGAAGCCTGGA
CAGAAGGTGTTGGATGTAGGATGTGGAATTGGTGGACCACTAAGAGAAAT
TGCTCGATTTAGCTCAACATCGGTTACGGGGTTGAATAACAATGAATATC
AAATAACAAGAGGAAGGGAAGTGAAGTGCATTGCTGGAGTGGACAAAACC
TGTGACTTTGTGAAGGCTGACTTCATGAAAATGCCATTTTCTGATAATAC
TTTTGATGCGGTATATGCGATTGAAGCAACCTGTCATGCTCCTGATGCGC
TTGGATGCTATAAAGAGATCTACAGAGTATTGAAGCCTGGCCAGTGTTTT
GCTGCATATGAATGGTGCATGACTGATGCCTTTGATCCCAATAACCAAGA
ACACCAAAAAATCAAGGCAGAAAATAGAAATTGGTGATGGCCTTCAGACA
TCAGATTGACTAGACAATGCCTTGAAGCCCTAAAACAAGCAGGTTTTGAG
GTTATTTGGGAGAAAGACCTAGCTGTGGGCTCACCTCTTCTTGGTACTT
GCCTTTAGATAAAAGTCACTTCTCATTAAAGTAGCTTCCGTCTAACAGCTG
TTGGACGTTTCTATTACTAAAAACATGGTCAAGGCTCTAGAATTTGTGGGA
CTTGCTCCAAAGGGAAGTCAAAGAGTTCAAGCTTTTCTAGAGAAAGCTGC
AGAAGGGCTAGTTGATGGTGGAAAGAAAGAGATCTTCACACCATGTATT
TCTTCTTGGCTCGAAAACCAAACCTCAGATGGTCAGTTATGTGGCAGTTGG
ACCCAA

pre-BR.STE1.fasta

>A

---GCGGCGGATAATGCTTATCTGATGCAGTTTGTGACGAAACCTCTTT
TTACAACCGAATCGTTCTGAGTCATCTTTGCGCGCGAATCTATGGGAAC
CCTTACCTCATTTTCTCCAGACATGGCTCCGAAATTACCTCGCGGAACC
CTACTATACTTCATCTCCGTTTTCTCTGGTGTCTACATCTATTACCT
TAAAATCAACGTTTACCTTCCCAAAGATGCAATTCCTACAATAAAGGCTA
TGCGTTTGCAAATGTTTGTGGCAATGAAGGCTATGCCATGGTACACTCTT
CTTCCAACGTGCTCCGAGAGTATGATTGAACGTGGTTGGACCAATGTTT
TGCTAGCATAGGCGAATTCCGGCTGGATTCTGTATTTTGTATACATCGCCA
TCTATCTTGTGTTTTCGTTGAGTTTGGTATTTATTGGATGCACAGAGAGCTT
CATGACATTAAGCCTCTCTATAAGTATCTCCATGCCACCCATCATATCTA
CAACAAGCAGAATACACTCTCTCCATTTGCCGGGCTTGCAATTTACCCAG
TAGACGGGATACTTCAGGCTGTACCGCATGTGATAGCGCTGTTTATAGTG
CCAATTCATTTCACTCATATAGGTCTTTTGTTCATGGAAGCGATATG
GACGGCGAATCATGACTGCATCCATGGCAACATCTGGCCAGTAATGG
GTGCAGGATACCATACGATACACCACACGACATACAAGCATAACTATGGT
CATTATACCATATGGATGGATTGGATGTTTGGCTCTCTTAGGGATCCTCT
CTTAGAAGAAGATGACAACAAAGACAGCTTCAAGAAAGCAGAG-----
--

>P

-----GAGGGAGACAAGTACTTGCAGGAATTCTTGGTTGAGACATCAAT

GTTCAACAGCATTGTTTTGGGTCATTTGTTACCGTCCAATTGGTGGGTCA
CTCTGCCTCATTTCTTGACAGACTTGGCTTCGTAACCTACATTGCTGGAAC
CTCCTTTACTTTCATCTCTGGTTTCCTCTGGTGCTTCTATATTTATTACTT
GAAACGCAACGTTTATGTCCCTAAAGATACCATTACCTTCAAACAGAGCCA
TGCTCTTACAAATATATGTTGCTATGAAGGCTATGCCATGGTACACTCTT
CTTCCAACAGTTTCTGAGTACATGATTGAAAATGGCTGGACAAAAGTGT
CTCTAGCATATCTGAAATTGGTTGGTTCGCCTACATCATATATTTTCCTA
TATATCTTGTTATTGTGGAGTTTGGGATTTACTGGATGCATAGAGAATTG
CATGATATAAAACCTCTATACAAGTATCTTCATGCAACCCATCACATCTA
CAACAAACAGAACACTCTATCTCCATTTGCTGGTTTGGCATTTCACCCAG
TTGACGGGATTCTACAAGCTGTACCACATGTTATAGCTCTGTTTCTTG
CCAACCCATTTAGATCGCACGTAGCCCTCTTATTCATCGAGGCTATATG
GACAGCAAACATCCATGATTGCATCCATGCCAACCTATGGCCCGTGATGG
GTGCTGGGTACCACACCATTCACCATACCACATACAAGCATAATTACGGC
CATTATACTATATGGATGGATTGGATGCTGGGAACCCTTCGTGACCCA - -
-----GAGGATGATTGATGCCAGAGAGCGCAGAAGGTGC

AA

>R

GAGGTGGTGAGGAACAAGCAGTTGCAGTATTTCTTGGAAGAGACAGCGAT
GTATAACACATTGTCTTAGGACATCTATTGCCCTTCTGGATGGTGGGCCA
CGCTTCCAGGTTCTTGACAGCTGGCTTCGTAACCTACATTGCTGGAAC
CTCCTTTTACTTCTCTGCTGGCTTCTGTGGTGTCTTACATTTACTACTT
GAAACGCAACGTTTATCTTCTAAAGATGCTATCCCTACAACAAAAGCCA
TGCTCTTGCAAATATATGTTGCTATGAAGGCTATGCCATGGTACTGTGCT
CTTCCATCTCTTCTGAGTACATGGTTGAAAATGGCTGGACTAAGTGT
CTCATACATATCTGATGTTGGCTGTTTGGCTACATCATATATTTAATTG
CGTATCTTGTTATGGTGGAGTTTGGGATTTACTGGATGCACAGAGAATTG
CATGATATAAAACCTTTATACAAGTATCTTCATGCAACCCATCACATCTA
TAACAAGCAGAATACTTTCTCTTTTGGCGGTTTGGCATTTCATCCGC
TTGACGGAATACTCCAAGCAGTCCACATGTTATTGCTCTCTTCATTGTA
CCTACACATTTCAAGACACACATAGGTCTCTTATTCCTAGAGGCCATATG
GACCGCAACATTTCATGATTGCATACATGGTAAGCTATGGCCGGTGATGG
GTGCTGGCTATCACACCATCCACCATACTACATACCGCCACAATTATGGT
CATTATACCATGATGGATGGATTGGATGCTAGGAACCCTTCGTGACCCT - -
-----GAGGAAGATGCATTCGAGAAGGTGAAG - - - - -
--

>O

GCGGGCGGCGGCGGCGAGTACCTGCGCCAGTTCTGTCGAGGAGACGGCCTG
GTACAACGAGATCTTCTCAGCCATGTGGTCCCGGGCGACTGGTGGCGCG
CCCTCCCCACCCGCTCCAGTCTGTGGCTCCGCAACGGCTCGGCGGCTAC
CTCATCTACTTCGCTGCGGCTTCTCTGGTGCTTCGTATCTACTACTG
GAAGCGCCACGCCCTACATCCCCAAAGATTCTATACCTACAATCGAAGCTA
TGAAGAAGCAAATAATTGTTGCATCAAAGGCTATGCCTCTCTATTGTGCC
CTTCCAACCTTATCTGAGTACATGGTTGAGAATGGATGGACACAGTGTTA
TGTTAATATCAGTGAAGTTGGTTGGCCAATGTACCTGGTTTATCTGGCTT
TATATCTTATCTTTGTTGAGTTTGGAAATTTACTGGATGCACAGAGAGTTG
CATGACATAAAGCCATTGTACAAGTACCTGCACACATACCACCATATTTA
CAACAAGGAGAATACCCTATCACCATTTCAGGACTAGCATTCCATCCAC
TGGATGGGATTTTGAAGCCATACCGCATGTGTTTGCCTCTACCTTATC
CCAACACACTTCAAGGACACACATTGCTCTCTTGTTCATAGAGGCCGTGTG
GACAACTAACATCCATGACTGCATTACGGCAAGGTTTGGCCGGTCATGG
GTGCTGGCTATCACACCATTCACCATACAACATACCGTCACAACATATGGC
CACTACACCGTGTGGATGGACTGGATGTTGCGCACCCCTTCGAGAGCCA - -
-----GAAGATATCTTGAAGAAGGAT - - - - -
--

>V

-----GACGATCAGCTTCACCTTTTCGTGCGCGAGACTTCCTT
CTACAACCGCATTGTTCTCGGCACATTCTTGCCGGCCAAACTCTGGGAGA
CCTGGCCCCACTTCGTCCAGACGTGGCTCCGTAACCTACATTGGTGGAAAC
CTAATCTACCTCGTCTCTGGATTCTCTGGTGCTTCTACATTTATTATCT
CAAGCGCAACGTTTATGTACCTAAAGATGCCATCCCTACAACAAGAGCCT
AT - - - TTGCAAATATATGTTGCCATGAAGGCTATGCCATGGTACACTGCT
CTTCCAACGCTCTCTGAGTACATGGTAGAAAATGGTTGGACAAGGTGTTA
TTCTAGCATAAGTGATGTCGGTTGGCCTGCCTACTTTGTTAATTTAGCAG
TATATCTTGTTATTGTGGAGTTTGGTATTTATTGGATGCACAGGGAAGTTG
CATGACATAAAGCCTCTCTACAAGTATCTTCATGCAACACATCATATCTA
CAACAAACAGAACACCCCTTTCTCCTTTTGTGGTTTGGCATTTCATCCAC
TAGATGGGATATTGCAAGCATTGCCCCATGTTCTAGCACTCTTCTTTTT
CGGACCCATTTTTTGACGCACGTACTGCTTATTTCATTGAGGCCATATG
GACAGCAAACATTTCATGACTGCATTGATGGCAAGATCTGGCCAGTAATGG
GTGCGGCTACCAACACCATTCACCACTACCTACCGGCATAAATATGGT
CACTATACAGTATGGATGGATTGGATGTTTGGAACTCTTTGTGACCCT - -
-----GAGGAGTCACCAAGAAGGTGACT - - - - -
--

TRP.1.1.1.267.fasta

>A

ATGACATTAACTCACTATCT --- CCAGCTGAATCCAAAGCTATTTCTTT
CTTGATACCTCCAGGTTCAATCCAATCCCTAACTCTCAGGTGGGTTTA
GTTTGAGGAGGAGGAATCAAGGAGAGGTTTTGGAAAAGGTGTTAAGTGT
TCAGTGAAGTGCAGCAGCAACAACCTCCTCCAGCATGGCTGGGAG
AGCTGTCCCTGAGGCGCCTCGTCAATCTTGGGATGGACCAAAACCATCT
CTATCGTTGGATCTACTGGTTCTATTGGCACTCAGACATTGGATATTGTG
GCTGAGAATCCTGACAAATTCAGAGTTGTGGCTCTAGCTGCTGGTTCGAA
TGTTACTCTACTTGCTGATCAGGTAAGGAGATTTAAGCCTGCATTGGTTG
CTGTTAGAAACGAGTCACTGATTAATGAGCTTAAAGAGGCTTTAGCTGAT
TTGGACTATAAACTCGAGATTATTCCAGGAGAGCAAGGAGTGATTGAGGT
TGCCCGACATCCTGAAGCTGTAACCGTTGTTACCGGAATAGTAGGTTGTG
CGGGACTAAAGCCTACGGTTGCTGCAATTGAAGCAGGAAAGGACATTGCT
CTTGCAAAACAAAGAGACATTAATCGCAGGTGGTCTTTTCGTGCTTCGCT
TGCCAACAAACATAATGTAAGATTCTTCCGGCAGATTGAGAACATTCTG
CCATATTTCAAGTGTATTCAAGGTTTGCCTGAAGGCGCTCTGCGCAAGATA
ATCTTGACTGCATCTGGTGGAGCTTTTAGGGATTGGCTGTGAAAAGCT
AAAGGAAGTTAAAGTAGCGGATGCGTTGAAGCATCAAACCTGGAACATGG
GAAAGAAAATCACTGTGGACTCTGCTACGCTTTTCAACAAGGGTCTTGAG
GTCATTGAAGCGCATTATTTGTTTGGAGCTGAGTATGACGATATAGAGAT
TGTCATTCATCCGCAAAAGTATCATACATTCCATGATTGAAACACAGGATT
CATCTGTGCTTGCTCAATTGGGTTGGCCTGATATGCGTTTACCGATTCTC
TACACCATGTCTATGGCCCGATAGAGTTCTTGTCTGAAGTAACTTGCC
AAGACTTGACCTTTGCAAACCTCGGTTTCATTGACTTTCAAGAAACAGACA
ATGTGAAATACCCATCCATGGATCTTGCTTATGCTGCTGGACGAGCTGGA
GGCACAATGACTGGAGTTCTCAGCGCCGCCAATGAGAAAGCTGTTGAAAT
GTTCAATTGATGAAAAGATAAGCTATTTGGATATCTTCAAGGTTGTGGAAT
TAACATGCGATAAACATCGAAACGAGTTGGTAACATCACCGTCTCTTGAA
GAGATTGTTCACTATGACTTGTGGGACGCTGAATATGCCGCGAATGTGCA
GCTTTCTCTGGTGCTAGGCCAGTTTCATGCA

>P

ATGGCACTTAATATTCTATCT --- CCAGCTGAAATCAAGGCCATCTCTTT
CTTGATTTCTACTAAATCCAATCGCCTACCTAAGCTTCAAGGTGGGCTTA
GTTTGAAGAGGAAGGATTGTGGG- - - - - GGGAGAAGAATTCAAGTGT
TCAGTT- - - - - CAGAATCAGAATCAGCCACCTCCAGCTTGGCCAGGAAG
AGCCTGTCCAGAACCTGGACGCAAGACATGGGATGGTCTTAAGCCTATAT
CAATTGTTGGATCTACTGGTTCCATCGGAACGCAGACTTTGGACATTGTA
GCGGAGAATCCAGATAAATCAAAGTTGTGGCACTCGCAGCTGGGTCAAA
TGTTACTCTTCTTGCAGATCAGGTGAGGACATTCAAACCTCAACTGATTG
CTGTTAGAAACGAGTCATTAGTTGATGAAATCAAAGAGGCCCTGGCTGAT
GTTGAGGAAAAGCCGAGATTATTCTGGGGAGCAAGGAGTTGTAGAGGT
TGCTCGTCATCCAGATGCTGTCAAGTGTAGTTACAGGAATAGTAGGTTGCG
CAGGTCTAAAGCCTACGGTTGCTGCAATAGAAGCTGGAAAAGACATATGC
TTGGCCAATAAAGAGACATTAATTGCTGGGGTCTTTTGTCTCCCTCT
TGCTCACAAATATAACGTGAAAATTCTTCCAGCTGATTCTGAACATTCTG
CCATTTTTCAAGTGTATTCAAGGCTGCGGAGGGTGCATTGCGGCGCATC
ATTTTAACCTGCTTCTGGTGGGGCTTTCAAGGATTGGCTGTTGAGAAAT
GAAAGAAGTTAAAGTAGCTGATGCTTTGAAGCATCCCACTGGAATATGG
GTAAAAAGATTACTGTAGACTCTGCTACCCTTTTCAACAAGGGTTTAGAA
GTCATTGAAGCCCACTATTTGTTCCGAGCTGATTATGATAATATTGATAT
CGTAATTCACTCCACAGTCTATAATACATTCAATGATTGAAACACAGGACT
CATCTGTTCTCGCACAGTTGGGGTGGCCTGATATGCGCTTGCCTATCCTT
TACACAATGTCTATGGCCAGACAGAGTTTACTGTTCTGAAATTAATTGGCC
TCGCCTAGATCTTTGCAAGCTTGGATCACTAACTTTTAAAGCTCCTGACA
ATGTAAAAATACCCATCTATGGATCTTGCTTATGCTGCTGGACGAGCTGGA
GGCACCATGACAGGAGTCTCAGTGCTGCTAATGAGAAAGCTGTAGAAAT
GTTTCATAGATGAAAAGATAAGCTATCTTGACATTTTCAAGGTTGTGGAGC
TAACATGTGATAAGCACCGAGCAGAATTAGTGGTCTCACCCCTCTCTTGAG
GAAATTATACATTATGACTTGTGGGCTCGAGAATATGCTGCTAACCTGCA
ACACTCTTCTGGTCCAAGTCTGTTTTTGT

>R

--- GCGCTTAATTTGCTTTCT --- CCAGCTGAACTCAAATCCGTTTCTTT
CTTAGATTCCACCAGGTCCAGTCAGCTCCCCAAGCTTCCAGGCAGTTTCA
GTTTGAAAAGGAAGGAA- - - - - TTTGGAAGGAAAGTGCAGTGT
TCAGCA- - - - - CAGGCACCACCACGCTGGCCAGGAAG
AGCAGTTGCAGAGCCAGGCCGTAACCTTGGGATGGTCTTAAGCCTATTT
CAATAGTTGGATCCACTGGCTCTATTGGTACTCAGACGTTGGATATTGTG
GCTGAGAAATCCAGATAAATTCAGAGTTGTAGCACTTGTGCTGGTTCCAA
TGTTACTCTTCTGCTGATCAGGTTAAGACTTTCAAACCTCAGCTTGTTT
CTGTTAGAGATGAGTCATTAGTTGATGAGCTTAAAGAGGCTTTAGCTGAT
GTTGATGAAAAGCCTGAGATTATCTGGAGAGCAAGGAATTGTTGAGGT
TGCTCGCCATCCAGATGCTGTCAAGTGTAGTTACTGGAATTGTAGGTTGTG

CAGGCTTGAAGCCTACTGTTGCTGCAATAGAAGCTGGTAAAGACATATGC
 TTGGCCAATAAAGAGACTTTAATTGCTGGAGGTCCTTTTGCCTCCCTCT
 TGCTCACAATATAATGTAAAAATCTTCCAGCTGATTCGGAACATTCTG
 CCATATTTCAAGTGCATTGAGGGCTGCCTGAGGGTGCAGTGCAGGCGCATT
 ATTTTAACCTGCCTCTGGTGGTGTCTTTCAGGGATTGGCCTGTTGAGAAAT
 GAAAGATGTTAAAGTAGCTGATGCTTTAAAGCATCCGAATTGGAATATGG
 GGAAGAAAGATTACTGTAGACTCTGCTACCCTTTTCAATAAGGGACTAGAA
 GTAATTGAAGCCCATTTATTTGTTGGAGCTGACTATGATAACATTGAGAT
 AGTAATTCATCCTCAATCCATTATACATTCAATGGTGGAAACAGAGGATT
 CATCTGTTCTTGACAGTTGGGGTGGCCTGATATGCGTATTCCAATTCTA
 TACACCATGTGATGGCCTGACAGAATTTACTGCTCTGAGATAACCTGGCC
 TCGCCTTGATCTTTGCAAGCTTGGGTCTTTAACATTTAAAGCTCCTGACA
 ATGTAAAATACCATCTATGGAGCTTGCTATGCAGCTGGACGGCTGGA
 GGCACCATGACAGGAGTCTCAGTGCTGCTAACGAGAAAGCTGTAGAGAT
 GTTCATCGATGAACAGATCAGCTACCTTGACATTTTCAAGATTGTGGAGC
 TAACATGCGACAAGCATCGGGCAGAATTGGTGACTTCAACCTCTCTTGAA
 GAAATTTACTATGACTTGTGGGCGGAGAATATGCCGCCAGTTTGCA
 GCCCTCTTCTGGTCTAAGTCTGTTCTCGCA
 >0
 --- GCGCTCAAGGTCGTCTCTTTCCCCGGGGACTTGGCCGCGGTCTCATT
 CCTCGACTCCAACAGAGGAGGAGCTTTCAACCAAGCTCAAAGTGGACCTCC
 CGTTTCAAAGCAGGGACAGAAGAGCAGTTTCCCTGAGAAGGACTTGTGT
 TCAATG-----CAACAGGCTCCACCACAGCATGGCCTGGTCG
 AGCCGTGTTGAACCTGGGAGGAGGTCATGGGATGGCCCCAAGCCTATCT
 CAATTGTTGGCTCAACCGGTTCTATTGGCACACAGACATTGGACATAGTT
 GCGGAGAAATCCAGATAAAATCCGGGTTGTTGCTCTTGCTGCTGGCTCCAA
 TGTGACTCTTCTAGCTGATCAGGTGAAAACATTCAAACCAAAGCTTGTG
 CTGTAAGAAATGAGTCATTAGTTGATGAGCTAAAGGAAGCCTTAGCTGAT
 TGTGATTGGAAGCCAGAAATTATTCCTGGTGAGCAAGGTGTCATAGAGGT
 TGCTCGCCACCCAGATGCAGTTACAGTTGTTACTGGGATAGTAGGGTGTG
 CAGGACTGAAGCCTACAGTTGCTGCAATTGAAGCTGGGAAAGATATAGCA
 TTGGCGAACAAAGAGACACTTATTGCAAGTGGTCCTTTTGTGCTTCCCT
 TGCACAAAAGCACAAGTGAAAATACTTCTGCTGATTCTGAGCACTCTG
 CTATATTTCAAGTGATATACAAGGCTTGCCCGAAGGAGCACTTCGCCGCA
 ATTTTGACTGCATCAGGTGGTGTCTTTCAGGGACTGGCCAGTTGACAAGTT
 GAAAGAAGTAAAGTTGCTGATGCTTTAAAGCACCCGAACCTGGAATATGG
 GGAAGAAGATTACTGTAGATTCTGCTACATTATTCACAAGGGTTTAGAA
 GTTATTGAAGCACATTATTTATTTGGTGTGAATACGATGACATTGAAAT
 TGTGATCCACCCACAATCTATCATACACTCTATGATTGAAACCCAGGATT
 CATCTGTGTTGGCTCAACTGGGATGGCCAGAT---CGGATACCAATCTTA
 TACACCATGTCTTGGCCAGACAGAATCTATTGCTCAGAGGTACCTGGCC
 CCGACTAGACTTTTGCAAGCTGGGTTCACTGACATTCAAAGCTCCTGACA
 ATGTGAAATACCCGTCGATGGATCTCGCCTATGCAGCTGGAAGAGCTGGG
 GGCACCATGACAGGAGTTCTGAGTGCTGCTAATGAGAAGGCTGTGGAGTT
 GTTCATCGATGAAAAGATCGGGTACCTGGACATCTTCAAGGTGGTGGAGC
 TGACATGCGACGCTCATCGGAATGAGCTAGTAACAAGGCCATCACTGGAG
 GAGATCATACATTATGATCTGTGGGCGAGGGAGTATGCTGCCAGCCTACA
 GCCATCCACTGGCCTCAGCCCTGTACCTGTC
 >V
 --- GCTCTGAATCTCCTCCCT---CCGGCAGAAATTAAGTTAATTCGTT
 CCTGGATTCCACTAAATCCACTCGCCTCCCCAAGCTTCCAGGTGGATTG
 TTTTAAAGAGAAAGGATTATGGAGCAACATTTGGTAGAAGAGTTAACTGT
 TCAGCA-----CAGCCGCTCTCCAGCTGGCCAGGACG
 AGCTGTTCTCTGAGCTGGGGCGCAAGACTTGGGATGGTCCAAGCCTCTCT
 CAGTTGTTGGATCTACTGGCTCCATTGGAACCTCAGACACTGGACATAGTT
 GCAGAGAATCCTGATAAATTCAGAATTGTGGCACTTGCAAGCTGGTTCCAA
 TGTGACTCTTCTTGCTGATCAGGTGAAGACGTTCAAGCCTCAATTAGTTG
 CTGTGAGAAATGAGGCATTAGTCAATGAACCTCAAAGAGGCTTTGGCTGAC
 TGTGAACACAAGCCTGAGATCATCCCTGGAGAGCAAGGTGTAATTGAGGT
 TGCCCGCCACCCAGAGGCACTGCTAGTGACAGGAATAGTAGGTTGCG
 CAGGTTTAAACCTACAGTGGCTGCAATAGAAGCAGGGAAAGACATAGCC
 TTAGCCAAACAAGAGACCCTAATTGCTGGGGGTCCTTTTGTCTTCTCTCT
 TGCACACAAGCACAATGTAAAGATCTTCTGCTGACTCAGAACACTCTG
 CAATATTTCAAGTGATTAGGGTTTGCCTGAGGGAGCACTTCGGCGCATT
 CTTTAACTGCATCTGGAGGGGCACTCAGGGATTGGCCCGTTGATAAAT
 GAAAGAAGCTAAAGTTTCAGATGCTTTGAAGCATCTAACTGGAATATGG
 GGAAGAAAGATCAGGTTGACTCAGTACCTTTTCAATAAGGGTTTAGAA
 GTTATTGAAGCACACTATCTGTTTGGGGCTGAATATGATGATATTGAGAT
 CATAATTCATCCACAATCTATCATACATTCTATGGTGGAAACACAGGATT
 CGTCTGCTCTGGCACAGTTGGGGTGGCCTGATATGCGTTTGCCAATCCTC
 TATACTTTATCGTGGCCAGACAGAATTTACTGCTCTGAGATAAAGTTGGCC
 ACGACTTGATCTTTGCAAGCTTGGGTCACTGACATTTAAACCTCTGACA
 ATGTGAAATATCCATCCATGGATCTTGCTATGCTGCTGGACGGGCTGGA
 GGCACCATGACAGGAGTTCTTAGTGACGCTAATGAGAAGGCTGTGGAGAT
 GTTATTGATGAAAAGATTGGCTATCTAGATATATTCAAGGTGCTGGAAC

TAACATGTGATAAGCACCAGGCAGAGTTAGTAACTTCACCATCCCTCGAG
GAAATCGTACATTATGACCTCTGGGCCCGGACTATGCAGCCAGTTTACA
TCATTCTCCAGTTTAAATGCCTGTTCCAGCA

TRP.1.17.1.2.fasta

>A

---GCTGTTGCGCTCCAATTGACCCGATTATGCGTTCGACCGGATACTTT
CGTGC GGGAAGATCATCTCTCTGGATCCGGATCTCTCCGC---CGCCGGA
AAGCTTTATCAGTCCGGTGCTCGTCT-----GGCGATGAGAACGCT
CCTTCGCCATCGGTGGTGATGGACTCCGATTTTCGACGCCAAGGTGTTCCG
TAAGAAGTTGACGAGAGAAGCGATAATTACAATCGTAAAGGGTTTCGGTCATA
AGGAGGAGACACTCAAGCTCATGAATCGAGAGTACACCAGTGATATATTG
GAGACACTGAAAAACAAATGGGTATACTTATTCTTGGGGAGATGTTACTGT
GAAACTCGCTAAAGCATATGGTTTTTGTCTGGGGTGTGAGCGTGCTGTTT
AGATTGCATATGAAGCACGAAAGCAGTTTCCAGAGGAGAGGCTTTGGATT
ACTAACGAAATCATTCATAACCCGACCGTCAATAAGAGGTTGGAAGATAT
GGATGTTAAATTTATCCGGTTGAGGATTCAAAGAAACAGTTTGATGTAG
TAGAGAAAGATGATGTGGTTATCCTTCCTGCGTTTGGAGCTGGTGTTGAC
GAGATGTATGTTCTTAATGATAAAAAGGTGCAAAATTGTTGACACGACTTG
TCCTTGGGTGACAAAGGTTCTGGAACACGTTGAGAAGCACAAGAAAGGGG
AATACACATCAGTAATCCATGGTAAATATAATCATGAAGAGACGATTGCA
ACTGCGCTTTTTGACAGGAAAGTACATCATTGTAAGAAACATGAAAGAGGC
AAATTACGTTTGTGATTACATTCTCGGTGGCCAATACGATGGATCTAGCT
CCACAAAAGAGGAGTTTATGGAGAAATTCAAATACGCAATTTGGAAGGTT
TTCGATCCCGACAATGACCTTGTCAAAGTTGGTATTGCAAACCAAACAAC
GATGCTAAAGGGAGAAACAGAGGAGATAGGAAGATTACTCGAGACAACAA
TGATGCGCAAGTATGGAGTGGAAAATGTAAGCGGACATTTTCATCAGCTTC
AACACAATATGCGACGCTACTCAAGAGCGACAAAGACGCAATCTATGAGCT
AGTGGAGAGAAGATTGACCTCATGCTAGTGGTTGGCGGATGGAATTCAA
GTAACACCTCTCACCTTCAGGAAATCTCAGAGGCACGGGGAATCCCATCT
TACTGGATCGATAGTGAGAAACGGATAGGACCTGGGAATAAAATAGCCTA
TAAGCTCCACTATGGAGAACTGGTCTGAGAAAGGAAACCTTCTCCAAAGG
GACCAATAACAATCGGTGTGACATCAGGTGCATCAACCCCGGATAAGGTC
GTGGAAGATGCTTTGGTGAAGGTGTTGACATTAACGTTGAAGAGTTATT
GCAGCTGGCT

>P

---GCTATCTCTCTCCAACCTCTGCCGCTATCACTCCGCTCCGACCTCTC
CTCCGAC-----AATCGCGTACCTATCCGCCGC---CGTAGAA
CCACTTTCTCCGTCCGCTGCGCTGGC-----GGTGATGACTCT---
---ACTTCTTCAGTCTCTTTGGAATCGGAATTTGACGCCAAAGTCTTCAG
ACATAACCTGACAAGAAGCAAGAATTATAATCGTAGAGGTTTTGGCCACA
AAGAAGAGACACTTGAGCTCATGAATCGCGAGTACACCAGTGATATAATA
AAG-----AAAGAGAAATGGTTATGAGTATACATGGGGAAATGTTACAGT
CAAATTAGCCGGAAGCTTATGGATTTTGTCTGGGAGTAGAGCGAGCTGTT
AGATTGCTTACGAAGCCAGAAAACAGTTTCTGATGACAAGATTTGGATT
ACTAACGAGATTATCCACAATCCGACTGTTAATAAGCGGTTAGAGGAGAT
GGAAGTTGAAAACGTTCTGTTGAGGAAGGGAAGAAACAGTTTGAAGTTG
TAAATGGTGGTGATGTTGTGATTTTGCCTGCATTTGGAGCGGCAAGTGAT
GAGATGTTGACTTTGAGTAGCAAAAATGTACAAATTGTTGATACAATTG
CCCTTGGGTATCCAAGGTTTGGACTACTGTTGAGAAGCACAAGAAGGGAG
ATTATACCTCAATTATTCATGGAAAATATGCTCATGAGGAACTGTAGCA
ACCGCTTCTTTTTCAGGAAAGTACATTATTGTGAAGGATATGAAGAGGC
AATGTATGTGTGTGATTACATTCTTGGGGGTGAACCTTAATGGATCTAGCT
CAACCAGAGAGGAGTTTCTAGAGAAATTTAAAAATGCAGTTTCTAAGGGG
TTTGATCCTGATAGTGACCTGGTGAAACTTGGTATTGCAAAATCAAACAC
AATGCTCAAGGGAGAAAACAGAAGACATTGGAATTTGGTGAGAGGATCA
TGATGCGCAAGTATGGAGTGGAAAATGTAATGATCATTTCATAAGCTTC
AACACCATTTGTGATGCTACTCAGGAGCGACAAGATGCAATGTATAAACT
GGTGGAGGAAAAGTTGGATCTTATGTTAGTTGTTGGCGGGTGGAACTCAA
GTAACACCTCCCACTCCAAGAAATTTGCTGAGCACCATGGAATTCCTTCA
TACTGGATTGACAGTGAACAGAGGATAGGCCAGGAAACAAAATAGCTTA
TAAGTTGAACCATGGGGAGTTGGTTGAGAAAGAGAACTGGCTTCCACAAG
GCCCTATTACAATTGGTGTGACATCAGGCGCCTTACACCAGACAAGGTT
GTGGAAGATGCCCTTATCAAGGTGTTGACATCAAACGTGACGAAGCTTT
ACAAGTAGCT

>R

---GCAATCTCCCTACAGCTCTGCCGCTCTCTCTCCGCACTGATCTCTT
TTCTCGAGACAAC-----AACTCGTCTCCTCTTCTT-----CGCCGTA
AACCTTTCTCAATTGATGCTCTGCTGCTGCTTCCGACGATGAATCA
TCTTCTGCCTCTGTAGCTATGGACTCCGATTTTGATGCTAAGGTTTTTAG
GCATAATTTGACGAGAAGCAAGAATTACAATCGCAGAGGCTTTGGTCATA
AAGAAGAACTCTTCAACTCATGAGCCAAGAATATACCAAGTGATATTATT
AAGACTTTGAAGGAGAATGGGAATGAGTATACTTGGGGAAATGTTACTGT

TAAATTGGCTGAAGCTTATGGGTTTTGTTGGGGTGTGAAAGGGCTGTTC
AGATTGCTTATGAAGCTAGGAAACAATTTCTGATGAGAAGATTTGGATT
ACTAATGAGATTATTCATAACCCCTACTGTTAATAAGCGTTTGGAAAGAGAT
GAATGTTGAAAAATTCCTCTTGAGGAAGGGAGGAAACAGTTTGAAGTTG
TGAAACAATGGCGATGTTGTGATTTTGCCTGCTTTTGGAGCTGCAGTGGAT
GAGATGTTGACTCTGAGCAACAAAAATGTACAAATTGTTGATACAACTTG
CCCTTGGGTATCTAAGGTCTGGAATACTGTTGAGAAGCACAAAGAGGGAG
ATTATACTTCAATTATTCATGGTAAATATTCTCATGAGGAAACCATAGCT
ACTGCTTCTTTTGGCTGGAACGTATATCATTTGTCAAGAATATGAAGGAGGC
AATGTATGTATGTGATTACATTCTTGGTGGCCAACTTAATGGATCTAGCT
CAACCAAAGAGGAGTTTCTTAAGAAATTTAAAAACGCAGTATCTAAGGGA
TTTGATCCAGATGTTGACCTGGTTAAGGTTGGTATCGCAAATCAAACAAC
AATGCTTAAGGGGAGAAACAGAGATATTGGTAAATTGGTGGAGAAGCA
TGATGCAGAGATATGGTGTGAAAAATGTCAATGACCACTTTATTAGCTTC
AACACCACTTTGTGATGCTACTCAAGAACGACAAGACGCAATGTTAATTT
GGTGGAGGAGAAGCTGGATCTTATATTAGTAATTGGTGGGTGGAACCTAA
GCAACACTTTCACACCTCCAAGAAATTTGCTGAGCTCCGTGGAATTCCTTCA
TACTGGATTGACAGTGAACAGAGGATTGGTCCAGGAAATAGAATAGCTTA
TAAGTTGAATCATGGGGAGTTGGTTGAGAAAGAAAATTCCTACCAGAAG
GTCCCATTAACAATTGGTGAACATCTGGTGCCTCTACTCCAGACAAGGTT
GTGGAAGATGCCCTAGTCAAGGTGTTGACATCAAACGTGAGGAAGCCTT
GCAACTGGCA
>0
GCGACCATCACACGCAGCTC-----CGGTGGCCCTCCT
CTCGCG-----GCGGCTCGCCGAGCCGCCGTGCCAGGCGCG
CCCGTCCGTCCGTGCGGTGCGACTCC-----TCGGCAGCGTCGTCC
CTGTCCGCTCCGCGTCTGCTCGACGCCGACTTCGACAAGAAGCAGTTCCG
CCACAACCTCACCGCAGCGACAACCTACAACCGCAAGGGCTTCGGCCACA
AGAAGGAGACGCTCGAGCTCATGAGCCAGGAGTACACCAGCGACGTGATC
AAGACGCTCAAGGAGAACGGGAACAGCACACGTGGGGCCCCGTGACGGT
GAAGCTGGCGGAGGCGTACGGCTTCTGCTGGGCGTCGAGCGCGCCGTGC
AGATCGCTACGAGGCGCGCAAGCAGTTCCCCGACGATCGCATCTGGCTC
ACCAACGAGATCATCCACAACCCACCGTCAACAAGCGATTGGAGGACAT
GGGCGTCCAAACATTCTGTTGATGCAAGTATTAAGGATTTTCGATGTTG
TCGAGCAAGGGGATGTTGTCGTATTGCCTGCATTTGGAGCGGCCGTGGAG
GAGATGTACACGCTGAATGAGAAGAAGGTGCAGATAGTCGATACAACGTG
CCCTTGGGTTTCAAAGGTGTGGAACATGGTTGAAAAGCACAAGAAGGGTG
ACTACACTTCTATTATTATTCATGGAATACTCCCATGAAGAACTGTTGCC
ACTGCTTCTTTGCAGGACATACATCATTGTTAAGAACATAGCAGAGGC
TAGCTATGTATGTGATTATATACTTGGTGGCCAACTTGATGGTTCTAGCT
CCACAAAAGAGGAGTTCTTGAAGAAATCAAAAATGCTGTTTCTCCAGGA
TTTGATCCTGATGTTGACTTAGTGAAAGTTGGCATTGCCAATCAAACAAC
AATGCTTAAAGGAGAACTGAAGAAATGGGAAGCTTGTGAGAAGACAA
TGATGCGTAGGTTTGGAGTTGAGAATGTCAATGACCACTTCATTGCCTTC
AACACTATCTGTGATGCCACACAGGAAGACAAGATGCAATGTACCAGCT
AGTGAAGGAGAAAGTTGATCTTATTCTGTTGTTGGAGGATGGAACCTCA
GTAACACCTCTCATTTGCAAGAAATGGAGAAGCTCAGTGAATTCCGTCT
TACTGGATTGACAGTGAACAAAGGATTGGACCTGGAAACAAGATCAGCTA
CAAGTTAAATCATGGCGAGCTTGTGAGAAAGAGAAGCTGGTTACCAGAAG
GCCCTATCAATACTCGAGTTACTTCTGGCGCTCAACTCCGGACAAGGTT
GTTGAGGATGCGTTACAGAAGGTGTTGAGATCAAGCGCCAGGAAGTTT
GCAGGCCGCG
>V
---GCGATGTCTGCAACTCTGCCGCTTCTCCACCTTCTCCGACCGCTC
CCTTCCGGAGGCC---TTCGCCGAATCGGAGTGTTCCGG---CGCCGGA
AGCCTCTCTCCGTTCCGTGCTCAGGC-----GAGTCCGAGTCTTCT
TCTTCTCGTCTGTGGCTGTTGATTCCGATTTTGATGCGAAGGTGTTTCG
GCACAACCTTGACGAGGAGTAAAAATTATAATCGGAAAGGATTTGGACACA
AGGATGAGACTCTGGAGCTCATGAATCGCGAGTATACTAGTGACATCATA
AAGACATTGAAGGAAAAATGGAATGAGTATAAATGGGGAATGTTACAGT
GAAACTGGCAGAGCCTATGGATTTTGTGGGGTGTGGAGCGTGCTGTCC
AAATTGCTTATGAAGCCAGAAAACAATTTCTGAGGAGAAAAATTTGGATT
ACGAATGAAATTATTCAACCCCTACTGTTAACCAGCGATTGGCAGAGAT
GGAAGTCAAAGATATCCCATTTGATGATGGCCAGAAACAATTTGAAGTTG
TTGATAAGGGTGATGTTGTGATTTTGCCTGCTTTTGGAGCTGCAGTGGAT
GAGATGTTGACTTTGAGCAACAAAAATGTACAAATTTGTTGATACAACTTG
CCCATGGGTTTCTAAGGTTTGGAAATTTGTTGAGAAGCACAAAGAGGGAG
AGTACACTTCAATCATGGTAAATATTCTCATGAGGAGACTATTGCA
ACTGCATCTTTTGCAGGGAAGTATATCATTGTGAAGAATATGGCAGAGGC
AATGTATGTATGTGATTACATTCTTGGAGGAGAACTTGATGGATCTAGCT
CAACCAAGAGGAGTTTTTTGAGAAATTTAAATTTGCCATCTCTGAGGGG
TTTGATCCAGACATTGACTTGTCCAAAGTTGGTATTGCAATCAAACCA
AATGCTCAAGGGAGAACTGAAGAAATTTGGGAATTAGTAGAGAGGACTA
TGATGCGAAAAATATGGGTTGAAAAATGTCAATAATCACTTCATAAGCTTC
AACACAATTTGTGATGCTACTCAAGAGCGACAAGATGCAATGTATAAGCT

GGTGGAGGAGAACTGGACGTTATGTTAGTAGTTGGTGGGTGGAACCTCA
 GCAACACCTCACATCTACAAGAGATTGCGGAGGACCGTGAATCCCCTCG
 TATTGGATTGACAGTGAGAAGAGAATAGGTCCAGGAAACAGAATAAGTCA
 TAAGTTGATGCATGGGGAGCTGGTTGAGAAAGAAAACCTGGCTACCAGAGG
 GTCCATTACGATAGGTGTTACATCTGGTGCCTCTACCCAGATAAGGTT
 GTTGAAGATGTTCTCATCAAGGTATTCGACATTAACGGGAAGAAGCCTT
 GCAACTGGCA

TRP.2.2.1.7.fasta

>A
 GCTTCTTCTGCATTTGCTTTTCCTTCTTACATAATAACCAAAGGAGGACT
 TTCAACTGATTCTTGTAATCAACTTCTTTGTCTTCTTAGATCTTTGG
 TTACAGATCTTCCATCACCATGTCTGAAACCCAACAACATTTCCATTCA
 AACAGA --- AGAGCAAAAGTGTGTGCTTCACTTGCA ----- GAGAA
 GGGTGAATATTATTCAAACAGACCACCAACTCCATTACTTGACACTATTA
 ACTACCCCAATCCACATGAAAAATCTTTCTGTCAAGGAACTGAAACAACCT
 TCTGATGAGCTGAGATCAGACGTGATCTTTAATGTGTGAAAACCGGTGG
 ACATTTGGGGTCAAGTCTTGGTGTGTGGAGCTTACTGTGGCTCTTCATT
 ACATTTTCAATACTCCACAAGACAAGATTCTTTGGGATGTTGGTCATCAG
 TCTTATCTCTATAAGATTCTTACTGGGAGAAGAGGAAAG --- CCTACAAT
 GAGGCAAAACCAATGGTCTCTCTGGTTTCACCAAACGAGGAGAGAGTGAAC
 ATGATTGCTTTGGTACTGGACACAGCTCAACCACAATATCTGCTGGTTTA
 GGAATGGCGGTAGGAAGGGATTGAAGGGGAAGAACAACATGTGGTTGC
 TGTGATTGGTGATGGTGCATGACGGCAGGACAGGCTTATGAAGCCATGA
 ACAACGCCGGATATCTAGACTCTGATATGATTGTGATTCTTAATGACAAC
 AAGCAAGTCTCATTACCTACAGCTACTTTGGATGGACCAAGTCCACCTGT
 TGGTGCATTGAGCAGTGCTCTTAGTCGGTTACAGTCTAACC CGGCTCTCA
 GAGAGTTGAGAGAAGTCGCAAAGGGTATGACAAAGCAAATAGGCGGACCA
 ATGCATCAGTTGGCGGCTAAGGTAGATGAGTATGCTCGAGGAATGATAAG
 CGGGACTGGATCGTCACTGTTTGAAGAACTCGGTCTTTACTATATTGGTC
 CAGTTGATGGGCACAACATAGATGATTGGTAGCCATTCTTAAAGAAGTT
 AAGAGTACCAGAACCCACAGGACCTGTACTTATTCATGTGGTGACGGAGAA
 AGGTCGTGGTTATCCTTACGCGGAGAGAGCTGATGACAAATACCATGGTG
 TTGTGAAATTTGATCCAGCAACCGGTAGACAGTTCAAACCTACTAATAAG
 ACTCAATCTTACACAACCTTACTTTGCGGAGGCATTAGTCGAGAAGCAGA
 GGTAGACAAAGATGTGGTTGCGATTGATGACGCTATGGGAGGTGGAACCG
 GGTTAAATCTCTTTCAACGTCGCTTCCCAACAAGATGTTTCGATGTAGGA
 ATAGCGGAACAACACGCAAGTTACTTTTGTGCGGGTTTAGCCTGTGAAGG
 CCTTAAACCTTCTGTGCAATCTATTCGTCTTTCATGCAAG --- CGTGCTT
 ATGACCAAGTTGTCCATGATGTT -----
 GATTTGCAA ----- AAATTACC
 GGTGAGATTTGCAATGGATAGAGCTGGACTCGTTGGAGCT --- GGTCCGA
 CACATTGTGGAGCTTTGATGTGACATTTATGGCTTGTCTTCTTAACATG
 ATAGTGATGGCTCCATCAGATGAAGCAGATCTTTTAACATGGTTGCAAC
 TGCTGTTGCGATTGATGATCGTCTTCTTGTTCCTTACCCTAGAGGTA
 ACGGTATTGGAGTTGCATTACCTCCCGAAACAAAGGTGTTCCAATTGAG
 ATTGGGAAAGGTAGAATTTAAAGGAAGGAGAGAGATTGCGTTGTTGGG
 TTAGTGCTCGAGCTTCAAGAGCTGTTTAGGAGCGGCTGTAATGCTCGAAG
 AACGCGGA --- TACGTAAGTGTAGCGGATGCACGGTTTTGCAAGCCATTG
 GACCGTGCTCTCATTGCGAGCTTAGCTAAGTCGCACGAGGTTCTGATCAC
 GGTGGAAGAAGGTTCCATTGGAGGTTTTGGCTCGCACGTTGTTCAAGTTTC
 TTGCTCTCGATGGTCTTCTT --- GGCAAACCTCAAGTGAGACCAATGGTA
 CTGCCTGATCGATACATTGATCACGGTGCACAGCTGATCAACTAGCTGA
 AGCTGGACTCATGCCATCTCACATCGCAGCAACCGCACTTAACCTAATCG
 GTGCACCAAGGGAAGCTCTGTTT -----
 >P
 GCTCTCTCTGCATTTTCTTTGCCTGCTCATGTCAATAGTGTCAAG ---

 ----- AGAGAAGTACATGTCAAG
 AAAAGG --- CCAATGGGGTTTGTGCATCACTATCA ----- GAGAG
 CGGAGAGTTTCTTACAGAGGCCACCAACACCTCTCTTGGACACCGTAA
 ACTATCCAATTCATATGAAAAATCTATCTATCAAGGAGCTAAACAACCTA
 GCAGAGGAGCTGCGGTCCGATGTTATCTTCAATGTTTCGAAAACCTGGGGG
 TCACTTAGGCTCTAGCCTTGGTGTGTGGAGCTCACCGTGGCTCTTCACT
 ATGTTTTCAATAGCCCTCAAGACAAGATACTGTGGGATGTTGGCCATCAG
 GCTTACCCTCACAAGATTCTAACTGGGAGAAGAGACAAGATGCATACAAT
 AAGACAGACCAATGGATTGGCTGGTTTTACAAGCGATCGGAGAGTGAAT
 ATGATTGTTTTGGCACTGGTACAGCTCTACCACCATTTCTGCTGGCTTG
 GGAATGGCTGTGGGGAGAGATCTAAAGGAAGAGCGAACAATGTGGTTGC
 TGTATAGGTGATGGTGCCATGACAGCAGGACAAGCTTATGAAGCTATGA
 ACAATGCAAGGTACCTAGACTCTGAT --- ATCGTTATTCTTAATGACAAC
 AAACAGGTTTCTTTACCAACTGCCAATTTAGATGGGCCAATACCACCTGT
 CGGAGCCTTGAGCAGTGCTCTCAGTAGGCTGCAATCAAACAGGCCTCTCA

GAGAACTAAGAGAGGTTGCTAAGGGAGTTACAAAACAAATTGGTGGACCG
 ATGCATGAATTGGCAGCAAAAGGTTGACGAATATGCTCGTGAATGATCAG
 TGGTTCTGGATCAACCCTCTTTGAAGAGCTTGGTCTCTATTATATTGGTC
 CGGTGGACGGGCACAATATTGATGATCTTATTGCCATTCTCAAAGAGGTG
 AAGAGTACCAAAACAACAGGTCCTGTCTAATCCATGTTGTCAACCGAGAA
 AGGTGGGGGTATCCATATGCTGAGAGAGCTGCGGACAAGTACCACGGAG
 TGAACAAGTTTGATCCCGCAACTGGAAAGCAGTCCAAGGCAAGTGCTAGC
 ACACAGTCTTACACAACATATTTTGCAGAGGCTTTGATTGCAGAAGCTGA
 GCGGGATAAAGATGTTGTTGCAATTCATGCTGCAATGGGAGGAGGGACAG
 GCTTAAATCTCTTCTTCGCCGTTTTCCGACAAGATGTTTTGATGTGGGG
 ATCGCAGAACAGCATGCTGTTACTTTTCTGCGGACTTGCCTGTGAAGG
 TCTTAAACCTTTTTGCGCAATCTACTCCTTCTTCTTATCCTGTTAT
 ATGACAATACACAGCAATACCTGGTTGGCATTATATTGTTACCAGACACT
 GACTCAAGCCCCCATGTAGATACTATTATTAATGTAAAAATGAACTTCC
 AGTAAGATTTCGCAATGGACAGAGCTGGGTTGGTGGGAGCAGATGGTCCCA
 CACATTGTGGAGCTTTTGATGTCACTTATATGGCATGCCTCCCTAACATG
 GTTGTATTAGCTCCTTCTGATGAGGCAGAGCTTTCCACATGGTTGCTAC
 TGCTGCTGCTATAGATGATCGTCTAGCTGTTTCCGCTACCCGAGAGGTA
 ATGGTGTGGTGTTCAGTTGCCACCAGAAAACAAGGGCATTCTCTTGAG
 GTTGGAAAAGGCAGGATATTGATTGAAGGGGAGAGGGTGGCACTCTTAGG
 CTATGGTACTGCAGTCCAGAGCTGTTTAGCTGCTGCCCTTTAGTGAAC
 GCCACGGTATTCTATCTGACAGTTGCAGATGCCAGGTTCTGCAAGCCATTG
 GACAATGCCCTTATTCTGAGCTTGCAAAATCACACGAGGTTTTGATTAC
 AGTGGAAAGAAGGATCAATCGGGGGCTTCGGGTCTCATGTTGCTATTTTC
 TGGCCCTGGATGGTCTTCTTGATGGCAAACTAAAGTGGCGACCGGTTGTT
 CTTCTGATAGGTACATTGACCATGGATCGCCAGCTGATCAGCTGGTAGA
 AGCTGGTCTCACACCATCTCACATTGCAGCAACAGTATTCAATATACTCG
 GACAAAGAAGAACTCTCTGGAGATTATGTATCA-----
 >R
 GCTCTTTGTGCATTTTCATTTCTGCTCATGCAAATAAGGTCACAACAGC
 AAGTTCAGATCTTCAGAAGTCTAATTATGTTTCTTCTAATTTCTTATTGA
 AAACAGATCTTTTGGGC---CAATCCATGCAAAAATTTAATCAGGGAAG
 AAAAGG---CCAGCTGGGGTTTGTGCATCACTATCA-----GAGAC
 AGGAGAGTATCATTCACAGAGACCACCAACACCTCTCCTTGACACCATAA
 ACTATCCAATTCACATGAAAAATCTATCAATCAAGGAGCTAAAACAATA
 TCGGATGAGCTACGGTCGGATGTTATTTTCAATGTTTCTATAACTGGAGG
 TCACTTAGGTTCAAGTCTCGGTGTTGTTGAGCTCACTGTTGCCCTTCACT
 ATGTTTTTCAATGCTCCTCAAGATAAGATACTTTGGGATGTTGGCCATCAG
 TCTTACCCCATAAAAATCCTTACCGGAAGAAGAGACAAGATGCAACAAT
 GAGACAGACAAATGGACTTTCTGGTTTTCACGAAGCGATCAGAGAGTGAAT
 ATGATTGCTTTGGGACTGGCCATAGCTCTACCACAATTTCTGCAGGCTTG
 GGAATGGCTGTGGGCAGAGATTTAAAGGAAGGAAGAACAATGTTGTTGC
 TGTATAGGCGATGGTGAATGACAGCAGGACAAGCTTATGAAGCTATGA
 ATAATGCGGGATACCTAGACTCTGATATGATTGTTATTCTTAATGACAAC
 AAACAAGTTTCTTTACCAACTGCCAATCTTGATGGACCAATACCTCCTGT
 TGGAGCTTTGAGCAGTGCTCTCAGTAGATTGCAATCAAACAAGCCTCTCA
 GGGAAATTAAGAGAGGTTGCTAAGGGTGTACAAAGCGGATTGGTGGACCT
 ATGCATGAATTGGCAGCAAAAGTTGATGAATATGCACGTGGGATGATAAG
 TGGTTCTGGATCGACCTCTTCGAAGAGCTTGGATTATATTATATTGGTC
 CTGTGGACGGTCACAATGTCGATGATCTTGTGCTATTCTCAAAGAGGTT
 AAGAGTACCAAAACAACAGGTCCAGTCTTGATCCATGTTGTTACTGAGAA
 AGGACGTGGTTATCCATATGCTGAGAAAGCTGCAGATAAGTATCATGGAG
 TGACCAAGTTTGATCCTGCTACTGGAAGCAATTCAAGGGTAGTGCTAGT
 ACACAGTCTTACACAACATACTTCGAGAGGCTTTGATTGAGAAGCAGA
 AGTAGACAAGGATGTCGTTGCAATCCATGCTGCAATGGGAGGTGGAACAG
 GACTAAATCTCTTCTTCGTCGGTTCCGACTAGATGCTTTGATGTTGGG
 ATAGCAGAACAGCATGCTGTTACATTTGCTGCAGGTCTAGCTTGTGAGGG
 TCTTAAACCTTTTTGTGCAATCTACTCATCTTTCATGCAG---AGGGCAT
 ATGACCAGGTGGTACATGATGTG-----AAACTGCC
 GATTTGCAG-----AAACTGCC
 TGTAAGATTTGCAATGGATAGAGCAGGGCTGGTTGGAGCAGATGGTCCAA
 CACACTGTGGAGCTTTTGTGCTACTTTTATGGCATGCCTCCCTAACATG
 GTGGTGTGGCTCCTTCTGATGAGGCAGAACTATTTACATGGTTGCTAC
 TGCTGCAGCAATAGATGATCGTCTAGCTGTTTCCGATACCCAAGAGGTA
 ATGGTATTGGTGTTCAGCTGCCACCAGGAAACAAGGCATTCTCTTGAG
 GTTGGTAAAGGCAGGATATTAATTGAAGGAGAAAGAGTGGCACTCTTGGG
 TTATGGATCAGCAGTTCAAGAGCTGTTTAGCTGCTGCCTCTCTAATGGAAA
 CCTATGACTTACGCATAACAGTTGCCGATGCTAGATTCTGCAAACTTTG
 GATCATGCCCTCATTGCAAGTCTAGCAAAATCACATGAAGTATTGATTAC
 AGTTGAAGAAGGTTCAATTGGGGGCTTTGGATCTCATGTTGCCATTTTC
 TGGCTCTTGATGGTCTTCTTGATGGAAAGCTAAAGTGGCGACCACTATT
 CTTCCAGATAGGTACATTGACCATGGATCGCCAGCTGACCAGCTGGTAGA
 GGCAGGTCTGACACCATCTACATAGCAGCAACAATATTCAACATACTTG
 GAAATAAAGAGAAGCTCTACAGATAATGTACGA-----
 >0

GCCTCACGACGTTCTCC-----ATTTCGAGAGGAGGCTTCGT
 CGGCGCGCTGCCGACGAGGGGCAATTCGCTCCGGCGCGCGGAGCTCA
 GTCTCCAC-----AAGCTCCAGAGCAGGCCACACAAGGCTAGG
 CGGAGGTCGTCGTCGAGCATCTCGGCGTCGCTGTCCACGGAGAGGGAGGC
 GGCAGGAGTACCACTCGCAGCGGCCACCGACGCCGCTGCTGGACACGGTCA
 ACTACCCCATCCACATGAAGAACCTGTCCCTCAAGGAGCTCCAGCAGCTC
 GCCGACGAGCTCCGCTCCGACGTCATCTTCCACGCTCTCCAAGACCGGGGG
 ACATCTCGGGTCCAGCCTCGGCGTCGTCGAGCTACCGTCTGCGCTCCACT
 ACGTGTTCAACACGCTCAGGACAAGATCCTCTGGGACGTCGGCCACCAG
 TCGTACCCTCACAAGATTCTGACCGGGCGGCGGACAAGATGCCGACGAT
 GCGTCAGACCAACGGCTTGTCTGGGATTACCAAGCGGTGGAGAGCGAGT
 ACGACTCCTTCGGCACCAGGCCACAGCTCCACCACCATCTCCGCCGCCCTC
 GGGATGGCGGTGGGAGGGATCTCAAGGAGGGAAGAACAACGTGGTGGC
 GGTGATCGGCGACGGCGCATGACGGCGGGCAGGCGTACGAGGCGATGA
 ATAACGCGGGGTATCTCGACTCCGATATGATCGTGATTCTCAACGACAAC
 AAGCAGGTGTGCTGCTGCCGACGGCGACGCTCGACGGGCGGGCGCGCGGT
 GGGCGCGCTCAGCAGCGCCCTCAGCAAGCTGCACTCCAGCGCCCACTCA
 GGGAGCTCAGGGAGGTGGCAAAGGGCGTGACGAAGCAAATCGAGGGTCTG
 GTGACGAGCTGGCGGCGAAGGTGGACGAGTACGCCCGCGGCGATGATCAG
 CGGCTCCGGCTCGACGCTCTTCGAGGAGCTCGGCCCTTACTACATCGGCC
 CCGTCGACGGCCACAACATCGACGACCTCATCACCATCCTCCGCGAGGTC
 AAGAGCACCAAGACCAAGGCCCGGTGCTCATCCACGTCGTACCGAGAA
 AGGCCGCGGCTACCCCTACGCCGAGCGCGCGCGGACAAGTACCACGGCG
 TGGCGAAGTTGATCCGGCGACGGGGAAGCAATTCAAGTCGCCGGCGAAG
 ACGCTGTGTACACGAACACTTTCGCGGAGGCGCTCATCGCGAGGCGGA
 GCAGGACAAACAGGGTCTGGCCATCCACGCGGCGTGGGGGAGGCGACGG
 GGCTCAACTACTTCTCCGCCGCTTCCCGAACCGGTGCTTCGACGTCGGG
 ATCGCGGAGCAGCAGCGCGTACGTTTCGCCGCGGCGCTCGCTGCGAGGG
 CCTCAAGCCGTTCTGCGCCATCTACTCCTCCTTCTGCGAG---AGAGGCT
 ACGACAGGTGGTGCACGACGT-----
 GACCTCCAG-----AAGCTGCC
 GGTGAGGTTCCGCTATGGACAGGGCCGGGCTCGTGGGCGCGACGGGCCGA
 CGCACTGCGGCGGCTTCGACGTCACCTACATGGCGTGCCTGCCGAACATG
 GTCGTCATGGCCCGCTCCGACGAGGCGGAGCTTGCACATGCTCGCCAC
 CGCCGCGGCGCATCGACGACCGCCCTCCTGCTTCCGCTACCCAAGAGGCA
 ACGGCATCGGCGTCCCGCTACCAACCAACTACAAAGGCGTTCCTCTCGAG
 GTAGGCAAGGGGAGGGTACTGCTGGAGGGCGAGAGGGTGGCGCTGCTTGG
 GTACGGTTTCGGCGGTGCACTGCTCGCCGACGCGTCTGCTGGTGGAGC
 GGCACGGCTCAAGGTGACCGTTCGCCGACGCGAGGTTCTGCAAGCGCTG
 GACCAAAACGCTCATCAGGAGGCTGGCCAGCTCCACAGAGGTGCTCCTCAC
 CGTCGAGGAAGGCTCCATCGGCGGGTTCGGCTCCACGTCGCGCAGTTCA
 TGGCCCTCGACGGCTCCTCGACGGCAAACTCAAGTGGCGGCGCTGGTG
 CTACCCGACAGGTACATCGACCACGGGTACCGGCGGATCAGCTGGCGGA
 GGCAGGGCTGACGCCGTGCGACATCGCGGCGACGGTGTCAACGTGCTGG
 GCCAGGCGAGGGAGGCGCTCGCCATCATGACGGTGCCCAACGCG
 >V
 GCTCTGTACGCTCTCATTTCTGCCCATTTTAGCCAGGCTGCT-----
 -GCTTCAAATCCTCAGAGACTTACTCCTCAGTGTTCCCATTTGTTCTTGG
 GGGTGGATTGCGAG-----TGCCAATCCAGCAAAGGAGTAAGGCCAGG
 AAAAGG---CCAAATGGG-----
 -----AGACCACCAACTCCTCTCCTGGACACTATCA
 ATTATCCAATTCACATGAAAAATCTGTCTGTCAAGGAGCTGAAACAACCTC
 GCAGATGAACCTAAGGTCTGATGTTGTCTTCAATGTTTCCAAAACCTGGGGG
 TCACTTGGGCTCCAGCCTCGGGTTGTGGAGTCACTGTGGCTCTTCAAT
 ATGTCTTCAATGCCCTCAAGACAGGATACTATGGGATGTTGGTCTACAG
 TCTTACCACACAAAATTCTAACTGGGAGAAGAGATCAAATGCATACCAT
 GAGGCAACAGATGGGTTAGCGGGATTACCAAGCGCTCGGAGAGTGAAT
 ATGACTGCTTTGGAACCGGCCACAGTTCTACTACCATCTCAGCAGGCTTG
 GGA---GCGGTGCGCCGGGATCTAAAAGGAAAAACAACAACGTTATTGC
 TGTATAGGTGATGGAGCCATGACTGCAGGGCAAGCTTATGAAGCAATGA
 ACAATGCTGGGTACCTGGATTCTGACATGATTGTTATCCTTAATGACAAC
 AAGCAGGTTTCTTACCCTGCTACTCTAGATGGGCCATACCACCTGT
 AGGAGCTTTGAGCAGTGCTCTTAGTAGGTTACAATCAAACAGACCTCTTA
 GAGAATTACGAGAGGTTGCCAAGGGCGTTACCAAACAGATTGGCGGACCG
 ATGATGAATTTGGTGCAAAAGTTGATGAATATGCTCGTGGGATGATCAG
 TGGTTCTGGATCAACACTTTTTGAAGAGCTTGGACTCTATTATATAGGTC
 CTGTTGATGGCCACAACATAGATGACCTTGTGTCATTCTCAAGGAGGTT
 AAGAGTACCAAGACAACAGGTCCAGTTCTGATCCATGTTGTACAGAGAA
 AGGCCGCGGATATCCATATGCTGAGAAAGCTGAGATAAGTACCATGGAG
 TGACCAAGTTCGATCCTGCTACTGGAAAAAATCAAATCCAGTGCTCCT
 ACTCAGTCTACACAACATATTTGCGAGAGGCTTGTGTTGAGAAAGCAGA
 GGTGGACAAGGATATTGTTGCAATTCATGCAGCAATGGGGGGTGAACGG
 GCTTGAATCTCTCCATCGCCGGTCCCCACACGATGCTTGTGTTGGG
 ATAGCAGAACAGCATGCTGTACCTTTGCTGCTGGTCTAGCCTGTGAAGG
 CATTAACCTTTTGTGCAATCTACTCATCTTTCATGCGAG---AGAGCTT

ATGACCAGGCATGCCAATCCTATTACCATATCTTCTTAATCAAATGCTAT
 GATTTGCAG-----AAGCTGCC
 AGTGAAATTTGCAATGGACAGAGCTGGGCTGGTTGGAGCAGATGGCCCAA
 CACATTGCGGAGCTTTTGATGTCACTTTCATGGCTTGCCCTTCCAAACATG
 GTGGTGATGGCTCCTGCTGATGAGGCCGAGCTTTTTCACATGGTGCCAC
 AGCTGCTGCCATAGATGACAGGCCAGTTGTTTCCGGTACCCAAGAGGAA
 ATGGGGTGGGTGTTGAACTGCCACCAGGGAACAAAGGCATTCTATTGAG
 GTTGAAGGGGCGGAATATTGATTGAGGGGGAGAGATTGCACTCTTGGG
 CTATGGAACAGCAGTACAGAGCTGTTTGGTTGCGTCTTCTTTGCTGGAAC
 AACATGGCTTACGAATAACAGTCGAGATGCCCGCTTCTGCAAACATTG
 GACCATGCTCTTATTCTAGCCTAGCAAAATCACATGAAGTTTTGATTAC
 AGTAGAAGAAGGGTCAATTGGTGGTTTTGGGTCTCATGTTGCTCAGTTTT
 TGGCCCTTAAATGGTCTTCTTGATGGCACAACAAAGTGAGTCCATGGTT
 CTTCTGACCGGTACATAGACCATGGAGCGCCAGCAGACCAAGTTGGCCAT
 GCGGGTCTGACACCATCTCATATTGCAGCAACAGTATTCAATATACTTG
 GACAAACAAGGGAGGCCCTGGAGATCATGTTA-----

TRP.2.5.1.10.fasta

>A

---ACTACACTCAATCTA-----TCAATTTTTCC
 ATCAGTTAAATATCCAGC-----TCCGCATCCATACCCGATTCA
 TC-----AAAATCCAACCTTTCCCTTCTC-----
 CGTCGGAAGCTCTCTACCGTTTTATCGGTCAACGCCAGAGACGAAGGTAT
 CATCCACAACCAT-----TTCGATT
 TCACGTCATACATGATCGGAAAAGCTAATGCCGTCAATGAAGCTCTCGAC
 TCCGCCGCTCTCGCTCCGAGAACCA---ATCAAGATCCATGAAGCGATACG
 TTAATCGCTTCTCGCCAGAGGCAACGAGTGAGACCCGTGCTCTGTATCG
 CAGCGTGG---TTAGTGGGCGGCGAAGAATCCGTGGCGTTACCGGCGGCG
 TGGCGCGTGGAATGATCCACACAATGTCGTTGATCCACGACGATCTTCC
 ATGTATGGATAACGACGATCTCCGCCGCGGAAACCAACGAACCACAAAG
 TTTTCGGTGAAGACGTCGCGGTTTTAGCCGGAGACGCGCTTATATCGTTC
 GCGTTTGAGCACTTAGCAACG-----TCAACG--
 -GCTGTTTTCTCCGGCTAGAGTTGTTGAGCAATCGGAGAGTTAGCGAAAG
 CCATTGGATCAAAGGGCTCGTGGCGGGTCAAGTAGTGGATCTAACCAGT
 GGAGGAATGGACCAAAACGACGTCGGCTTAGAGGTTCTTGAGTTTATCCA
 CGTTCTATAAAACCGCCGTGTTGTTAGAGGCGGCTACGGTGTAGGAGCCA
 TCGTTGGTGGTGGTTCTGATGAAGAAGTTGAGAAGCTGAGAAGATTCGCG
 AGATGCATTGGATTGTTGTTTCAAGGTGGTGGACGATATTTTGACGTCGAC
 GAAGTCGTCGGAGGAATTGGGAAAGACCGCGGAAAGATCTT---GCCG
 ATAAGCTCAGCTATCCGAAACTAATGGGTCTTGAGAAATCTAAAGACTTT
 GCAGATAAGTTGTTGAGCGATGCTCACGAACAGCTTCATGGGTTTGATTC
 GAGTAGAGTTAAACCTTTACTAGCTTTGGCTAATTACATTGCCAAAAGAC
 AGAAT

>P

ATGACTTCTGTGAATCTGGGTCCATGGGTTTCAAGATAACTTCAATCATCAA
 CAAAAAACAAGATCCAGA-----TCCAGACCCATGTCAAACCTTCT
 TGTCCCCGGGCTTAAACCCCTTCAAATCGCTTCAACACATCAAAGACAG
 AGAAAACCCATGTCTTCAATTTGCGCAGTTCTTACAGAAGAAGAGACTCT
 TCAAGAAGATCAACAC-----AAACCCACTTTTGATT
 TCAAGTCTTATATGATACAGAAAGCCAATTCTGTTAACAAAGCATTAGAT
 ACTGCTGTTACTCTTAAAGAACCG---GCTAAATCCATGAGTCTATGCG
 TTAATCTCTTTTGGCTGGTGGCAAGAGGGTTCGGCCAGTGCTTTGTCTAG
 CTGCATGTGAGCTTGTGGTGGGTCTGAATCCATGGCTATGCCTGCTGCT
 TGTGCTGTAGAAATGATCCATACTATGTCATTAAATACATGATGATCTTCC
 TTGCATGGATAATGATGATCTTCGCCGCGGAAATCCACCAATCATATTG
 TTTTGGTGAGGATGTTGCGGTTTTGGCAGGGGATGCTTTACTGGCATTT
 GCATTTGAACATATTGCAGTG-----TCTACACT
 CAATGTTTCGCCTCTTAGAATTGTTTCGTGCAAGTTGGTGAATTGGCGAAAG
 TGATTGGTGCA---GGACTTGTGCTGGACAAGTTGTGGATATTTGTTCT
 GAAGGGTTG-----TCCGAAGTGGGGTTAGAACAGCTTGAATTTATTCA
 TATTCATAAGACTGCTAAGTTGTTGGAAGGTGCGGTTGTTTGGGGGCTA
 TATTAGGTGGAGGGACCGATGAGGAAGTTGAGAAATTGAGGCGATACGCG
 AGGAGTATTGGATTGTTGTTTCAAGTAGTGGATGATATCTTGATGTTAC
 CAAATCTTCACAAGAATTGGGAAAACGAGGGAAGATTGGTTGCGG
 ATAAAGTTACTTATCCTAAGTTAATGGGAATTGAGAAGTCTAGGGAGTTT
 GCTGAGAAGTTGCTTAACGAAGCTAGGGAGTTGCTTGCTGGATTGATCA
 AGAGAAGGCGGCTCCGTTGATTGCTTTGGCTAATTACATTGCTTACAGGC
 AAAAC

>R

---AGTTCTGTAAATTTAGGTTCTTGGGTACATACCTCTTCAATCTTCAA
 CCAAGCATCCAGATCCAGA-----TCCAAA-----TCTTTTT
 CTTCTAATCCTTAAAGGCTTGACAGATTCTGTTTTATGCTCAAAATCA
 AGGAAGCCCATTTCTTTTATTGCAGCAATTATTACAAAAGAAGAGAAAAAC

TCTTCAAGAAGAACAAGATAATTTTAATTCAAGACCACCCACTTTTGATT
TCAAGTCTTACATGCTCCAAAAAGCTGCTTCCATCAATCAAGCTTTAGAT
GCTGCTGTTCCGATTCTGTGAACCA---ACTAAATTCATGAATCTATGAG
GTATTCTCTTTTGGCTGGTGGCAAAAGGGTCAGACCAGCACTTTGTTTAG
CTGCTTGTGAACCTGGTTGGTGGCAATGAATCTATGGCTATGCCTGCTGCA
TGTGCTGTAGAAATGATTCATACTATGTCACTTATTCATGATGATCTTCC
TTGTATGGATAATGATGATTTACGCCGTGGAAAGCCCAATCATATTG
TTTTTGGTGAGGATGTGGCTGTTCTTGCAAGTGATGCATTATTAGCACTC
GCTTTTGAACATATCGCTATA-----GCCACTTT
AAATGTTTCTCCTGCTAGAATTGTTAGAGCAGTTGGGGAATTAGCTAAAG
CTATTGGGGCTGAAGGGTTAGTTGCTGGTCAAGTTGTTGATATATGTTCT
GAAGGATTA-----TCGGAAGTGGGCTTAGAAAACTTGAATTTATTCA
TGTCTATAAGACCGCAAGTTGCTGGAAGGGGCTGTGGTGTAGGGGCTA
TAATGGGTGGAGGGACTGACGAGGAAGTTGAGAACTGAGGAAATATGCT
AGGGATATTGGGCTGTTGTTTCAAGTGGTGGATGATATTCTTGATGTCAC
CAAATCATCTCAAGAGTTGGGAAAACTGCAGGGAAGGATTTGGTGGCAG
ATAAGGTTACATATCCGAAGTTGATGGGAATTGACAAGTCGAGGGAATTT
GCTGAGAAAGTTGAATAAGGAAGCTCAAGAGCAGTTGGCTGGGTTTGATCC
TGAGAAGGCTGCTCCATTGATTGCTTTGGCTAATTACATAGCTTACAGGC
AAAAAC
>O
-----GCTGCCTTCCC
CCCGCTCGCCGCTCCCGCTCCGCATCTCCCTCTCATCCCCGCCGCCG
CGATGGCTGGGACAGCAGGGGCGGCGGCGGCGTCTGACGCGCAGCACC
CGGAGGTTCTGCGCCATCGTGGCGACCGCGGCGGCTCACCGTGCCGGC
CGCGGCGGCGGCGGCG-----GCGACGGGTTTGATT
TTAACGCGTATATGGGTGAGAAGGCTGCGGCGGTGAACGGGCGCTGGAC
GGCTCGATCCCGGCGGACGAGCGCCGCGGCGCTCCACGAGGCGATGCG
GTACGCGCTGCTGGCGGGCGGGAAGCGGGTGGCGCCGCGCTGTGCTGG
CGCGTGGCGGCTGGTGGTGGGGGGCGGAGGCCGCGGCGATGCCGCGGCC
GCCGCGCTCGAGATGGTGCACACC---TCGCTCGTCCACGACGACCTCCC
CTGCATGGACGACGACGACCTCCGCGCGGGAAGCCACCTGCCACGTCG
TCTACGGGGAGCCCATCGCCGTGCTACCGGCGACGCGCTGCTCTCCCTC
TCCTTCCACACATGGCCAGGTTGACTCTTACCCTCCGGACATCGACGC
TGACAAGCACCCCGCCGCGTCTGTCGCGCCATTGGCGAGCTCGCGCGCT
GCATAGGCTCCGAGGGCTAGTCGCGGCGCAGTTGTTGATCTTGAGATG
ACTGGATCG---ACCGAACTGTACCTCTTGAACGCTTGTAGTACATCCA
TCTCCACAAAACCTGCTGCTTGGTGGCATCAGTGGTTATTGGGGCAA
TCTTGGGAGGTGGCTCTGATGAGCAGATTGAAAGTTTGGCGATGTATGCG
AGATCGATAGGATTGTTGTTCCAGGTTGTTGATGATATACTTGATGTGAC
CAAGTCTTCTGAGGAGCTCGGCAAGACAGCAGGGAAGGACTTGGCGAGTG
ACAAGACGCATACCCGAAATTACTTGGGCTGGAGAAATCACGGGAGTTT
GCAGAAAAGTTGCTTTCTGATGCAAGGGAACAACCTTTCAGGATTTGATCA
AGAGACCGCAGCACCACTTCTGCACCTGGCCAATTATATTGCCATCGGC
AGAAC
>V
---AATACTGTGAATCTGGGCACATGGGTGCATACCTCTTATGTGTTCAA
TCACGGGGGTAGATTAAAA-----TCCCCTGCTTGAGCCGTATGC
TT---CATCCTCTGAAGAACCCATCCTTTCCG---TTGTCTCAGAGGCT
GGACGACCCATGCTCTGCTGGTGGCGTTCTACCAAGGAGGAGGTTAGAGG
AGGGGACACCCAGAGAAGAG-----AAGCCGCTTTCAATT
TCAACGCTTACATCCTTGATAAGGCTAATTCCGTCAACAAAGCTTTGGAC
GACGCGGTTCCCATTAGGGAACCT---GTGAAGGTCCACGAGTCGATGCG
CTACTCCCTTCTCGCGGCGGCAACGCTGTCGCTCCGTTCTCTGTATCG
CGGCTGCGAGCTGGTGGGCGGGGATGAAGCTACGGCGATGCCGCGTCT
TGTGCGGTGGAGATGATTACACCATGTCTTTAATGCATGATGACCTTCC
TTGC---GACAACGACGATCTACGCCGTGGAAAGCCCAAAATCACAAAGG
TCTTCGGTGAGAAATGTTGCCATTCTCGCCGGAGACGCTCTTCTGGCGTTT
GCGTTTCGAGCACATGGCCACA-----GCGACGGT
TGGAGTTCCGCGGGGTGGATAGTTTCGTGCAAGTCGGCAATTGGCGAAAT
CGATTGGGTGAGAAGGACTTGTGTCAGGGCAAGTGGTTGATATATGTTCT
GAGGGCTTG-----AAAGATGTTGGATTGGAACATCTCGAATACATCCA
TGTTCAAGACCGCAGCTCTACTTGAAGGGGCGGTAGTTCTTGGGGCGA
TTCTCGGAGGAGATCAAATGAGGAAATTGAAAAATTGAGGAAATTTGCC
AGATGTATAGGGTTGCTCTTTCAGGTGGTTGATGATATTCTGGATGTTAC
CAAATCCTCCCAAGAGTTGGGAAGACTGCAGGCAAGATTTGGTGGCTG
ATAAGCTCACGTACCCGAAGCTTCTGGGTATTGAGAAATCTCGGGAATTA
GCTGAGCAACTGAATAAGGATGCCAAGGACGCTCTCCGGCTTCGATCC
AGACAAGGCCGCCCTTTGATTGCTTTATCTAATTATATTGCTTACAGGC
AAAAA

TRP.2.7.1.148.fasta
>A

GCAACGGCTTCTCCTCCATTTATCTCAACTCTCAGCTTC-----
 -----ACTCACTCTTCTTTCAAACCTTCTTCTT
 CTCTTTCATTTTCTCCGAAGCTTCTTCGA-----CCCCTCTTAAGC
 TTTTCCGTCAAAGCT-----TCCAGAAAGCAAGTAGAGATAGTGTTTGA
 TCCT---GAGAGGCTTAATAAGATAGGTGATGATGTTGACAAAGAAGCTC
 CTTTGTCCAGGCTTAAGCTCTTCTCACCTTGCAAGATCAATGTTTTCTTG
 AGGATCACCGGAAAGCGAGAAGATGGGTTTCATGATTTAGCCTCTTTGTT
 TCATGTGATTAGCTTAGGAGACACTATTAAATTCTCATTGTCACCATCAA
 AGTCTAAAGATCGTTTGTCTACTAACGTTCAAGGAGTCCCTGTTGGG---
 AGAAATCTGATTATAAAAGCACTTAACCTTTACAGGAAGAAAACCTGGTAG
 TAACAGATTCTTCTGGATTCAATTAGATAAGAAGGTGCCTACCGGGGCTG
 GACTCGGTGGTGGAAAGTAGTAATGCTGCAACTGCACTCTGGGCGGCAAA
 GAGCTCAATGGAGGTCTTGTCACTGAGAACGAACTCCAGGATTGGTCAAG
 TGAATTTGGGTGAGATATTCTTTCTTCTTCTCGCGA---GCTGCCTATT
 GTACCGGGAGAGGTGAGATTGTCCAAGACCTTCTCCACCTTTTCTCTT
 GATCTTCCGATGGTGCTCATAAAGCCCCGAGAAGCATGTTCCACTGCTGA
 AGTTTACAAACGTCTTCGTTTAGATCAGACGAGCAATATTAATCCCTTGA
 CATTACTAGAGAATGTGACCAAGCAATGGTGTGTCTCAAAGCATATGCGTA
 AACGATTTGGAACCGCCAGCGTTTTCACTTCTTCCATCTCTAAAACGCTT
 GAAGCAACGGATAATAGCATCTGGACGTGGGGAATACGATGCTGTGTTTA
 TGTCTGGGAGTGAAGCACTATTATCGGTATTGGTTCACCAGATCCTCCT
 CAATTTATATGATGATGAAGAATAACAAGACGTGTTCTTGTCTGAAGC
 AAACCTTTATGACGCGTGAGGCTAATGAATGGTACAAAGAACCTGCTTCTG
 CAAATGCTACTACCTCA-----TCCGCCGAATCTCGCATGGATTTT
 CAA
 >P

 -----CTGGTGTATGA
 TCCTGATAGAAGGGTAACAAGTTGGCGGATGAGGTGGATAAACAGGCTC
 CTCTCTCAAGGCTCACTTTTTCTCACCTTGCAAGGTTAATGCTTCTTG
 AGAATAACCAAGTAAGAGAGAAGATGGGTATCATGATTTGGCATCTCTATT
 TCATGTGATTAGTCTCGGAGATGTAATTAAGTTCTCTTGTCAACCATCAA
 AAAGAATTGATCGATTATCAACCAATGTGTCTGGTGTACCTCTTACTGAA
 GATAATTTGATAATTAAGCCCTTAACCTTTACAGGAAAAAGACTGGCAC
 TGACAATTTCTTTGGATTCAATTAGACAAAAAGGTGCCCACTGGAGCAG
 GACTTGGTGGTGGGAGCAGCAATGCCGCTACTGCCCTGTGGGCGGCAAA
 CAATTCAGTGGTGTCTTGCTACTGAGAAGGAACTCCAAGAATGGTCAAG
 TGAGATTGGTTCAGATATCCCTTTCTTTTTTCTCATGGAGCAGCCTACT
 GTACTGGCCGTGGTGAGATTGTTCAAGATATTCCTTCACCAGTATCTTCA
 GGCAATTCGAATGGTTCTCATGAAGCCCCGCACTCATGTTCAACTGCAGA
 AGTTTACAAGTGCTTTCAATTGGATAAAACAGTCAAGCTGATCCTTTAA
 CATTGCTGGAGAAGATCTCAAGGAATGGAATATCTCAA--GTATGTATC
 AATGATTTGGAACCTCCTGCCTTTGAAGTTCTCCCTCTCTCAAAGATT
 AAAACAGCGTATAGCTGCAGCTGGTCTGGACAATATGATGCTGTTTTCA
 TGTCTGGAAGTGAAGCACAATAGTTGGGATTGGCTCACCAGATCCTCCG
 CAATTTATATATGACGAAGATGAGTACCAGGATGTTTTGTGTGAGAAGC
 CAACCTCCTTGCTCGTGAAGCAAAATCAATGGTATCAACAACCTGCTTCAA
 CAAC--TGAGCTCT-----CCACCCGAATTTCCCGGTCTGTT
 GAG
 >R
 -----GCTTCTGCTCATTTCTCTGCAACAGCCATCTCTTTCAACCTTC
 TTGCAATTCCTTTGCTAAACTAGTCTTCTTTCATTTAAGCCAGCTGGGT
 CTATTTCTTTT-----CCTCGAACACAAAGAATCCCATTTATCAAA
 -----GCC-----ACCAAGAAACAGCTAGAGATAGTGTATGA
 TCCCGATGAAAGGTTAAACAAGTTAGGAGATGAAGTAGACAAGATTGCTC
 CTCTTTCAAGGCTCACTTTGTTCTCTCCTTGCAAGATTAAATGTTTTCTG
 CGAATAACTGATAAAAGAGAAGATGGGTATCATGATTTGGCATCTCTCTT
 TCATGTAAATTAGCTAGGAGATACAATTAAGTTCTCTTATCTCCTTCAA
 AATCAACAGACCGTTTATCAACTAATGTGGCTGGCGTTCCCTTGATGAA
 AGGAATTTGATTATTAAAGCCCTTAACCTATATAGGAAAAAGACGGGCAC
 TGACAACCTCTTTTGGATTCACTAGATAAGAAGGTGCCTACTGGGGCAG
 GGCTTGGTGGTGAAGCAGTAATGCTGCGACTGCATTGTGGGCGGCAAA
 CAGTTCACTGGCGGTCTTGCCACTGAGAAGGACCTTCTAGAATGGTCAAG
 TGAGATTGGTTCAGATATCTCCTTCTTTTTTCCCGTGGAGCAGCCTATT
 GTACTGGTCTGTTGAAATTGTCCAAGATATTCTTCTCCAGTTCTCTT
 GACATTCGAATGGTTCTTATAAAGCCCCAGGAAGCTGCCCAACTGGTGA
 AGTTTACAGCGCTTTCGGTTGGATCAAACAGTCAAGTTGATCCTTTAA
 CACTTCTGGACAAGATCTCAAAGAGTGAATATCTCAAGATGTTGTATA
 AATGACTTGGAAACCTCCTGCATTTGAAGTTCTCCCATCTTTAAAAGATT
 AAAACAGCGTATAGTTGCAGCAAGCCGTGGACAATACAATGCAGTTTTTA
 TGTCTGGAAGTGGTAGTACCATAGTTGGGATTGGTTCGCCGGATCCTCCA
 CAGTTTATTTATGACGACGATGACTACAAGGATGTTTTCTGTGTCAGAGGC
 TAACCTTCTGACTCGTGAAGCAAAATGAGTGGTACAAAGAACCTGCTTCAA

CTGCTGCTTGTGGCACA-----GAGTCTGATTTTTCCCAATCTACA
GAG
>0
---GCTTGCTCCACCCACCTCCTGTCCCAGAGCCTCTACCCGCTCAACCG
CGCTAACCCGGCAGCCGCGCGCGGGCACCTCCGGTTCCAGGCTTCTCCCA
GTGTGAGGCTCGGCTCCGGCACCAGTCGCCGCGGGCGCTGGGCCTGAGA
GTCGCGCGCTCGGCGGAGCAAGGGAGGAGGCAAGTGGAGGTTGAGTATGA
TCTACAAGCAAAGTTCAACAAGCTAGCGGACCAAATTGACCAGAATGCTG
GGATTACACGGCTGAACCTATTCTCACCTTGCAAAATTAATGTGTTCTTG
AGGATAACGGGTAAGAGACCAGATGGGTTCCATGACCTGGCTTCTCTGTT
TCATGTGATAAGTTTGGGTGATACTATCAAATTCTCATTGTCAACCAAGTA
AGAGCAAAGATCGCTTGTCAACCAATGTAGCAGGTGTCCAGTTGATGAA
AGCAATTTGATCATCAAGGCACTCAATCTTTACGCAAGAAAACCTGGAAC
TGACAACTTTTTTGGATTCTCTTGATAAGAAGGTCCTACTGGTGCTG
GTCTTGGTGGTGGAAGCAGTAATGCTGCAACTGCGCTGTGGGCGCCAAC
CAGTTTAGTGGTGCATTGCTTCAGAAAAGGAGCTTCAGGAGTGGTCTGG
AGAGATAGGATCGGATATTCCCTTCTTCTTTTCAAGGAGCAGCGTATT
GTACCGGTAGAGGAGAGATTGTTGAAGATATTCGGAATCCATTGCCAGCA
AATTTGCCGATGGTACTAGTAAAGCCACCTGAAGCATGCTCAACAGCTGA
AGTTTACAGCGGCTCAGGTTAGAGCACACAAGTCAAACCTGATCCCTTGG
TATTGCTCAAGGAAATTACTGAAAATGGGATATCACAGGATGCCTGTGTT
AATGATCTAGAACCTCCAGCATTGAGGTGTTGCCATCACTAAAGAGGTT
GAAGAAACGTATAATTGCTGCTAACCAGGGGAGATTATGATGCTGTTTTTA
TGTCAAGGAAGTGAAGCACAATTGTTGGGATTGGTTCACCAGATCCGCCT
GCATTTGTGATGATGATGACTACAAGGATACTTTTGTGTCAGAGGC
CTGCTTCCTCACTCGTAATGAGAACGAGTGGTACAGAGAACCAATCTCAT
CGAAAATCACTAGCGAGGAAGATTACCTCCGGAGGTAGCATCAGTTTCT
GAC
>V
GCTATGGCTTCATCACAAATTTCTTCTGCGATCACCAACTT---CACCTTC
CTGCAATTCATCAGAAGAACCACCTCCCTTCATTTAGACCACATGGGT
CATCTTCTTTTCGTCCAAAACCCCAATTTAGAGGACTCCATTTCTCAGA
GTCAGGGCTTCTGCT-----GGTAGAAAACAAGTGGAGATAGTGATAA
TCCAGATGAAAGGATAAACAATCTAGCAGATGAAGTGGATAAAGACGCTG
GGCTTTCAAGGCTCACTCTATTTTACCTTGCAAGATTAATGTTTTCTCTG
AGAATAACCGGTAAGAGAGCAGATGGATTTTCATGATTTGGCATCTCTCTT
TCATGTAATCAGTTTAGGAGATATAATTAAGTTCTCTTTGTCGCCATCAA
AAACAAAGGATCGCTTATCAACCAATGTGCCAGGGGTCCCTTGTATGAA
AGTAATTTGATCATAAAGGCACCTTAACCTTTACAGGAAAAAACTGGCAC
TGATAACTACTTTTGGGTTTCATCTTGATAAAAAGGTGCCGACTGGGGCAG
GGCTTGGTGGTGGCAGCAGTAATGCTGCAACAGCACTGTGGGCAGCAAT
CAATTTAGTGTTGCTTGGCTTGGCACTGAGAAGGAACTCCAACAATGGTCAAG
TGAGATTGGTTCGGATATTCCATTCTTCTTCTCTAATGGAGCAGCTTTTT
GTACTGGCAGAGGTGAGGTTGTTGAAGATATTTCCCTGCCAATACCTTGT
GACATTTCAATGGTTCTTGTAAAGCCACCGCAGGCGATGCTCCACCGCTGA
AGTTTACAGCACCTTCGGTTGGATCAAACTAGTAAGGTTGATCCCATAA
CCTTGCTTGAGAAGATCTCAAAGAACAGAATATCTCAAGATGTTTGTATA
AATGATCTAGAACCTCCTGCATTGAAAGTCCCTCCCATCTCTGAAAAGATT
AAAACAGCGTATAATTGCAGCAAGCCGTGGACAATATGATGCTGTTTTCA
TGCTCGGGAGG-----
-----TATAAA-----
-----ATT
CAG

TRP.5.3.3.2.fasta

>A
-----TCTGCTTCTTCTTTAT
TTAATCTCCATTGATTCGCCTCAGATCTCTCGCTCTTTCGTCTTCTTTT
TCTTCTTTCCGATTTGCCATCGTCTCTGTCAATCGATTTACCGAGAAAA
GTTACCGAATTTTCGTGCTTCTCTGGTACCGCTATGACAGATACTAAAG
ATGCTGGTATGGATGCTGTTTACAGAGACGTCTCATGTTTGAGGATGAATGC
ATTCTTGTGATGAAACTGATCGTGTGTTGGGGCATGACAGCAAGTATAA
TTGTCACTGATGGAAAAATATTGAAGCCAAGAATTTGCTGCACAGGGCTT
TTAGTGATTTTTATTCAACTCGAAGTATGAGTTGCTTCTCCAGCAAAGG
TCAAACACAAAGGTTACGTTCCCTCTAGTGTGGACTAACACTTGTTCAG
CCATCCTCTTTACCGTGAATCAGAGCTTATCCAGGACAATGCACTAGGTG
TGAGGAATGCTGCACAAAGAAAGCTTCTCGATGAGCTTGGTATTGTAGCT
GAAGATGTACCAAGTCGATGAGTTCACCTCCCTTGGGACGTATGCTGTACAA
GGCTCCTTCTGATGGCAAATGGGGAGAGCATGAACCTTGATTACTTGTCT
TCATCGTGCGAGACGTGAAGGTTCAACCAAACCCAGATGAAGTAGCTGAG
ATCAAGTATGTGAGCCGGGAAGAGCTGAAGGAGCTGGTGAAGAAAGCAGA

TGCGGAATGCTGCACAAAGAAAGCTGTTGGATGAACTCGGCATTCTGCT
 GAAGATGCACCTGTTGATCAGTTCACCTCCACTGGGCCGATGCTTTACAA
 GGCACCATCAGATGGCAAGTGGGGAGAGCATGAACTTGATTATCTACTCT
 TCATTGTCCGAGATGTTAATGTGAATCCAAACCTGATGAAGTTGCTGAT
 GTTAAAGTATGTGTCCCGAATGATCTGAAAGAGCTGTTGAGGAAAGCGGA
 TGCTGGTGAGGAAGGTCTGAAGCTGTCCCTTGTTGACTGGTGGTCTG
 ACAACTTCCTCTTCAAGTGGTGGGACCATGTTGAGAAGGGGACCTCCTT
 GAAGCTGCAGATATGAAAACCATTCACAAGTTGACT

altered.GA.5.5.1.13.fasta

>P

GCAAAGGAAATAAGCAAGCGCTCAGAACTATTAAGCAATGCTGGATATGATGGAGGAT
 GGAGAGATAAGCATCTCAGCTTATGACACCGCATGGGTGCTTGTGGAAGATATTAAT
 GGGAGTGGCCTTCCCAATTCCCATCTAGCCTTCAGTGGATTGCCAACAAATCAGCTTCCA
 GATGGTTGCTGGGTGACGCTGGAATATTTTTGGCGCATGATAGGTTGATTAATACGCTA
 GCTTGCCTGGTAGCATTGAAATCTTGGAACCTTCATCAAGATAAATGCGAAAAGGGATG
 TTATTTTTTAGAGACAACCTTATGCAAGCTTGAAGATGAAAATGCTGAGCACATGCCTATT
 GGATTTGAAGTTGCTTTCCCTTCACCTTCTTGCGATAGCCAAAGAAATTAGACATCGAAGTT
 CCGGATGATTCTACTTTCCCTCAAGAGATTTATGCCAGGAGAAATTTAAAGCTGAAAAGG
 ATTTCCAAGGAC - - -ATGCACAAATGTGCCACGACCCTACTCCATAGCTTGGAAAGGAATG
 CGAGGTCTGGACTGGAACAGCTTATAAAATTGCAGTGCCTAGATGGGTCACTTCTTATTC
 TCTCCATCTCTACTGCCTTCGCACTCAGTCAAACTAAAGATAAAAACCTGCTTGGAAATAT
 TTGAATAAGGCGGTCCAAGATTCAAAGGGGGAGTTCCCAATGTCTACCCGGTTGACTTA
 TTTGAGCACATCTGGGTTGTGGATCGCCTGCAGCGTCTTGAGATCTCAAGATATTTTCGTC
 TCAGAGATCAATGAATGTGTCGATTATTTATAGATATTGGACTGAAAATGGCATCTGT
 TGGGCGAGAAATTTCTAATGTCCATGATATTGATGACACGGCCATGGGATTGAGGATACCT
 AGATTACATGGGCACAGGTTTCTGCCGATGTGTTCAAGCATTTTGAAGGCGGTGAA
 TTTTTCTGCTTCGCAAGGCAAGTCAACAGGTGCTGTTACGGGAATTTTCAACCTGTATAGG
 GCTTCTCAGGTTCTGTTTCCGGGAGAGAAGATACTTGAGGATGCCAAGGAATATTCATTC
 GAATTCCTAAGGAAAAACAAGCTGCTAATGAACCTTCTAGATAAATGGATTATAACGAAG
 GACTTGGCCGGAGAGGTGGGTTTTGCATTGGAGATCCATGGTATGCGAGCTTGCCCTCGA
 GTGGAGACTAGATTCTTTCATAGAACAGTATGGTGGCGAGGACGACGTGTGGATTGGCAAG
 ACCCTTTACAGAATGTCATATATAACAACAGTGAGTATCTCCAGCTGGCAAAATTAGAC
 TACAACAACCTGCCAAGCATTACATCGCATTGAATGGGAAAAATTTCAAAAGTGGTATGAA
 GAATGTAATCTAAGAGATTTCCGGGATAAGCAGAAGAACCCTTATCTTTCTTACTTTTTG
 GCAGCAGCCAGTATCTTTGAGCCGGAAGGTCCAAAGAGCGACTTGATGGGCTACCACT
 ACAGTCTTGCTCGACATTGTGGGCTCTTATTTCCCTGAGAATCATAACAATTCTAGCGAG
 CAGAGGAGAGCCTTCATACATGAATTCAGTTACGGTATTAGCATAAATGGAAGGGAGGAA
 AATGATTATTTATGTCTGATTCTCGTGTGTTGCAGGAGGTGCGGT - - - - -AGAAAGAG
 ACAAGGCAAGGAG - - - - -CTCGTTAAGCTGCTGCTTGAACCCATAATCAGCTCTCTTGG
 GCGCATTAGTGGTTTCATGGT - - - - -CGAGACATTAGCCACAGTTTACGCCATGCTTGG
 GAAAAGTGGCTCTTGATCTGGGAATTGGAAGGAGATAGACGCCAAGGGGAAGCAGAA - - -
 -----CTACTAGTGCAACA
 ATAAATCTCACCGCCGGCTATTTGGTCTCCGAGGAGCTCTTGCCCATCATCCTCAGTAT
 GAACAACCTGTCGACCTGACAAACAGAATATGCTATCAACTTGATCATTACAAGAAAAAC
 AAGGTACATTACATGGAAGCTACAGCACTATTACCAGTAATACTGACAGAATTACAACC
 CCACAAATAGAATCAGACATGCAAGAACTTGTGCAGTTAGTAGTCCAAAAATCCCTCAGAT
 GGA - - -ATAGATTCTCAATATCAAGCAACGTTCTTCAAGTAGCCAAGAGTTTTATTAT
 TCTGCAATCTGTGATCTGGGACTATCAACTATCACATAGCTAAAGTCTCTTTGAGAGA
 GTGCCT

>R

TCGAATGAGATCCTTAAACGTGTCCAAACGTTAAATCGATGTTGGATTGATGGAAGAT
 GGAGAGATAAGCATATCAGCTTATGATACAGCATGGGTGCTTGTGGAAGATATTAAT
 GGAAGTGGTGCTCCTCAATCCCGTCCAGTTTGCAATGGATTGCTAATAATCAACTATCT
 GATGGTTGCTGGGGCGATGGTGATCTTTACTGCGCATGACAGGATTCTCAATACATTA
 GCTTGTGTTGTTGCCCTTGAATCATGGAACATACATCCTGATAAGTGTGAAAGAGGAATG
 AAGTATTTTAAAGAAAACCTTATGCAAGCTTGAAGATGAGAATGCTGAGCACATGCCATT
 GGTTTTGAAGTTGCTTTCCCTTCACCTTCTGAGTTAGCAAAAAATTTAGACATTGAAGTT
 CCCGAGGATTCTCCAGTCTTGAAGAGATCTATGCCAGCAGAAATATAAAGCTAACAAAG
 ATACCAAAAGACATAATGCACAAAGTGCCACCACACTACTCCATAGCTTGGAAAGGAATG
 CCAAGGTCTGGAGTGGGAAAAGCTTCTGAAATTGCAGTGGCCAGATGGGTCACTTCTGTTT
 TCTCCATCTCCACTGCTTTTGCATCATGCAAACTAAAAATGAAAATTGCTTAGCATAT
 CTAACAAAGATCGTTCAAAGATTAAATGGAGGAGTACCAACGCTTACCCTGTTGACTTA
 TTCGAACATATCTGGGCTTGTGATCGGTTCAACGCCTTGAATATCGAGATATTTTCAGA
 AAGGAGCTTAAAGAATGTTATTGACTATGTTGCCAGATACTGGGAAGAAGATGGTATTGTC
 TGGGCAAGAAATCTGCAGTTTCATGATATTGATGACACAGCCATGGGATTGAGTTGCTT
 AGACTATATGGCCACGAAGTTTCTTCTGATGTGTTCAAGCATTTTAAAGAAAGGCGTACT
 TTCTTCTGCTTCGCGGACAGTCAACGAGGCTGTTACCGGAATGTACAACCTGTATAGA
 GCTTCTCAGGATATTGTTTCTTGAGAAAAAGTCTCGAGGAAGCCAAGGAATACTCGTCC
 AGTTTCTAAAAGAAAAGCAAGAGGCTAACGAAGTACTTGACAAATGGATTATAACGAAA
 GACTTGGCCGGGAGAGGTTAAGTATGATTGGATATTCCATGGTACGCGAGCTTGCTCGG
 GTGGAGTCGAGATTCTACTTGAACAGTACGGTGGTGGGATGATGTCTGGATCGGCAAG
 ACTCTTTATAGGATCCGCTATGTGAACAACAATGAGTATTTGGACCTTGCAAGCTTGAC

TACAACAATTGCCAAGCGTTACATCGCAAGGAGTGGGACAACCTTCAAAAGTGGTACGAA
GAATGTGAATTGGGAACTTTGGGGTAAGCAGAAGAGAACTGCTGTTAGCATATTTTGTA
GCGGCGGCCAGCATTTTTGAGCCAGAAAGGTCAAAAGAGAGACTTGCAATGGGCTAAGACC
ACAACCTTGCTCCATACAATCGAGTCTTATTTT---GATGCCAGCAACAGTACCTATGAG
CAAAGGACGGCTTTTGTCCACGAATTCAAAAACGGTGTGCGCA-----
-----TCCATACCCCACTTAAATGCAAGGAAGTTG-----GAAGTAAAG
ACAAATGAGGAA-----CTTGTGAGGATAGCGATTGGAATCCTAAACGATGTCTCTTTG
GATACACTACTGGCTCATGGG-----AAAGACATTAGCCACGATTTACGTCATGCTTGG
GAAAAGTGGCTGTTGAAGTGGGCGAAGGAGGAGAAATACACCAAGGAACAGGGGAA---
-----TTGCTAGTGAAAACA
ATAACTCTCACCCTGGAGGTTCCACACCTGAT-----CATCATAAATAT
GCACAGCTCTTTCAACTGACAGATAAATTATGCTATCAGCTGGCTCATTACAGAAAAAAC
AAGGTACAAGGCAATAAGAAAAGC-----ACAACC
CCAGAAATAGAGTCAGACATGCAACAACCTTGTGCAGTTAGCAATCCAAAATTCCTCAGAT
GAA---ATAGATTCCGAGATCAAGCAAACATTTTTCATGGTAGCCAAAAGCTTTTACTAC
CAAGCCATCTCTGATCTGGGACTTTGAACTACCATATTGCTAGAGTACTCTTTGAGAGA
GTATAC

>A

AGTAATGCATTCAAAGAAGCAGTGAAGAGTGTGAAAACGATCTTGAGAAACCTAACGGAC
GGGGAATTTACGATATCGGCTTACGATACAGCTTGGGTTGCATTGATCGAT-----GCC
GGAGATAAAAGCTCCGGCTTTCCCTCCGCCGTGAAATGGATCGCCGAGAACCAACTTTCC
GATGGTTCTTGGGGAGATGCGTATCTCTTCTTATCATGATCGTCTCATCAATACCTT
GCATGCGTCGTTGCTCTAAGATCATGGAATCTCTTTCCTCATCAATGCAACAAAGGAATC
ACGTTTTTCCGGGAAAAATATTGGGAAGCTAGAAGACGAAAAATGATGAGCATATGCCAATC
GGATTGGAAGTAGCATTTCCCATCGTTGCTTGAGATAGCTCGAGGAATAAACATTGATGTA
CCGTACGATTCTCCGGTCTTAAAGATATATACGCCAAGAAAGAGCTAAAGCTTACAAGG
ATACCAAAAGAGATAATGCACAAGATACCAACAACATTGTTGCATAGTTTGGAGGGGATG
CGTGATTTAGATTGGGAAAAGCTCTTGAAACTTCAATCTCAAGACGGATCTTTCCTCTTC
TCTCCTTCTCTACCGCTTTTGCAATCATGCAGACCCGAGACAGTAACTGCCTCGAGTAT
TTGCGAAATGCCGTCAAACGTTTCAATGGAGGAGTTCCCAATGTCTTTCCCGTGGATCTT
TTCGAGCACATATGGATAGTGGATCGGTTACAACGTTTAGGGATATCGAGATACTTTGAA
GAAGAGATTAAAGAGTGTCTTGACTATGTCCACAGATATTGGACCGACAATGGCATATGT
TGGGCTAGA---TGTTCCCTCCAAGACATCGATGATACAGCCATGGCATTTAGGCTCTTA
AGACAACATGGATTGAGTACCAAGTGTCCGAGATGATTCAAGAACCTTGAGAAAGAGGGAGAG
TTTTTCTGCTTTGTGGGGCAATCAAACCAAGCAGTAACCGGTATGTTCAACCTATACCGG
GCATCACAAATGGCGTTTTCCAAGGGAAGAGATATTGAAAAACGCCAAAGAGTTTTCTTAT
AATTATCTGCTAGAAAAACGGGAGAGAGAGGAGTTGATTGATAAGTGGATTATAATGAAA
GACTTACCTGGCGAGATTGGGTTTGCCTTAGAGATTCCATGGTACGCAAGCTTGCCTCGA
GTAGAGACGAGATTCTATATTGATCAATATGGTGGAGAAAACGACGTTTGGATTGGCAAG
ACTCTTTATAGGATGCCATACGTGAACAATAATGGATATCTGGAATTAGCAAAACAAGAT
TACAACAATTGCCAAGCTCAGCATCAGCTCGAATGGGACATATTCCAAAAGTGGTATGAA
GAAAAAGTTAAGTGAAGTGGGTGTGCGCAGAAGTGAGCTTCTCGAGTGTACTACTTA
GCGGCTGCAACTATATTGAAATCAGAAAGGTCACATGAGAGAATGGTTTGGGCG---TCA
AGTGATTGGTTAAAGCCATTTCTTCTCTTTT-----GGGGAATCCTCTGAC
TCCAGAAGAAGCTTCTCCGATCAGTTTCAATACATTGCCAATGCTCGACGA-----
-----AGTGATCATCACTTTAATGACAGGAACATGAGATTGGACCGACCAGGA
TCGGTTACAGGCCAGTCGGCTTGCCGGAGTGTTAATCGGGACTTTGAATCAAATGTCTTTT
GACCTTTTCATGTCTCATGGC-----CGC---GTTAACAATCTCCTATCTATCGTGG
GGAGATTGGATGGAAAAATGAAACTATATGGAGAT-----GAAGGAGAAGGAGAG---
-----CTCATGGTGAAGATG
ATAATTCTAATGAAGAACAATGACCTAACTAATTCTTCACC-----CACACTCACTTC
GTTGCTCTCGCGGAAATCATCAATCGAATCTGTCTTCTCGC---CAATACTTAAAGGCA
AGGAGAAAC---GATGAGAAG-----GAGAAG-----ACAATA
AAGAGTATGGAGAGGAGATGGGAAAAATGGTTGAGTTAGCATTGTCGGAGAGTGACACA
TTT---CGTGACGTCAGCATC-----ACGTTTCTTGATGTAGCAAAAGCATTTTACTAC
TTTGCTTTATGTGGCGATCAT---CTCCAACTCACATCTCCAAAGTCTTGTTTCAAAA
GTC---

>V

TCAAATGAGATTAAGGAACGTGTGGATTCCATAAAATCGATGCTGAGTTCAATGGAAGAT
GGAGAGATCAGCATATCGGCTTATGACACAGCCTGGGTGGCTCTCGCGCAGGACCTCAAT
GGAGGAGGGACTCCGCAATTCCTATCAAGCCTAGAGTGGATCGCCAACAATCAGCTCATG
GACGGGTATGGGAGAGCAGTCTTATTCTATGCTCACGACCGGATAATCAACACATTG
GCGTGTGTGATCGCATTGAAATCGTGGAAAATTCATCCGAACAAGTGCAGAGAAAGGGGTG
TCATTATCAAAAGAAAACATAAGCAAGCTTGAAAAGGAGAAGGCTGAGCACATGCCTATA
GGATTGGAAGTAGCTTTCCCTTCTCTCTCGAGATAGCTTGGAACTTGGGCATCGAAGTA
CCCAATGATTCGCCCTTCTTGAAGGAGATCTACGCCAGAGAAATCTGAAGCTCACAAAG
ATACCATATGACATAATGCACATAGTGCCCAACAACACTACTCCACAGCTTGGAAAGGAATG
GCAGGTCTAGACTGGGAAAAGCTTCTAAACTACAGTGTCAAGATGGGTCAATTTTGTGTT
TCTCCATCGTCCACTGCCATGCCCTCATGCAGACCCAAGATGAGAATTGCTTTAGATAC
CTAAACAAAGTGGTTGAGAGATTCAAGTGGAGGAGTCCCAATGTTTACCCTGTGACTTG
TTTGAGCACATATGGGCTGTGGATCGGCTGCAGCGCTTGGGATATCGCGGTATTTTGAG
CCTGAGATTAAGAATGCATCAATTATGTTGCCAGATATTGGAAGAAAAATGGAATCTGC
TGGGCTAAAAACTCCGAAGTTCACGATATTGATGACACTGCTATGGGTTTAGGCTCCTT
AGGTTACATGGCCATGATGTTTCTGCTGATGTGTTCAAGTATTTGAGAAAGGTGGTGAG
TTCCTTCTGCTTTGCTGGGCGATCATCCAGGCTGTGTCGGGATGTTCAACCTCTACAGA

GCTTCCCAGGTGCTGTTCCCTGGAGAAACCATTTCTTGAGAATGCCAAGAAGTTCTCTTCC
 AAGTTTCTAAGGGAAAAACAAGCTTGTGACCAGCTCCTGGATAAGTGGATCATAATGAAG
 GATTTGCCCTGGTGAGATAGGGTATGCGCTGGATGTTCCATGGTATGCAAGCTTGCCTCGT
 GTGGAGACCAGAATTTATATTGAACAGTACGGTGGAAAAGATGATGTATGGATTGGCAAG
 ACCCTTTACAGGATGCCCTTATGTAAACAATAATGATTATCTTGAGCTTGCAAACTGGAC
 TTCAATAACTGCCAAGCCCTGCATCAGCTTGAATGGGACAGCATTCAACAGTGGTACACG
 CAATGCCATCTTGGGGAGTTTGGGGTTAGCAGAAAAGCCTCTCTTGGCTTATTTTCTA
 GCAGCAGCTGCATATTTGAGCCTGAAAGGTCAATTGAGAGGCTTGGCTTGGGCGAAAACC
 GCCATCTTGGTCTTGGCAGTCCGGTCGTATTTT-----AGCAAGGAAACATTTATTAAG
 CAGAGGAGAGCCTTCTTCTTAAATTTGGATACAGCCCAAGTGGAGGAGAC-----
 -----TACATGAAGAAGATCAATGGAAGGCCGTGGGATTCTTAAAGAACAAG
 AGCGGAGAGGAA-----CTTGTCTGGACTCTTACTCGGAACCTTAAATCAACTCTCATTG
 GATACACTGGTGGCTTATGGC-----AGAGACATACGACACATCTTACGTCAAACCTGG
 GAG---TGGCTGATGAAACATTTGGATGAAGGAGATGGGTACCGAGGAGAGGCAGAG---
 -----CTCCTGGTGCAGGACC
 ATAAATCTCTGTGCTGGTCTCTCTCAGAGGAGTTGTTGGCT---CATCCTCAGTAC
 AGGCGTCTCTCTCAACTCAACAGAATCTGCAACGATCTTGGTCTCTTTACATTGCAC
 AAGGGTCAC---AATGGCACTTACAATCCAGAAAATGGCAGCCCT-----ATGAGC
 CATAGGGTAGAGTCAGACATGCAAGAAGTTACGAAGATAGTGTGGAAAACCTACCCTGAT
 GGC---ATCAACCAGAAAATCAAGAGGACTTTTGAGATGGTGGGCAAGAGTTTCTACTAT
 GCTGCCTACTGTGGACCTGCAACTATCAGTGCTCATATTGCTAAAGTCTCTTTGAGAGA
 GCAGCC
 >0
 GATGATGAGCTGCAGCCACTGGTCGAGCAGGTGAGGTGATGCTGTCGTCCATGGAGGAC
 GGCAGGATCACCAGCTCGGCGTACGACACGGCGTGGTGGCGCTGGTGCCGCGGCTGGAC
 GGCAGGGCGGGCAGCGATTTCCGCGCCCGCTGCGGTGGATCGTGGCAGCCAGCTCGCC
 GACGGGTCTGGGGCAGCAGGGCGCTCTTCTCCGCTACGACCGCTCATCAACACCCCTC
 GCCTGCGTCTGCGCCCTCACCAGATGGTCCCTCCACCATGACAGTGAAGCAAGGGCTT
 CAGTTTCTGAATCTGAAGTTTGGAGGTTAGCAGAGGAGGAGCCGGATACGATGCCGATT
 GGGTTTGAGATTGCATTCCTTCTCTTGTGGAGGCAAGTGGGTTTGGGTATTGATTTT
 CCATATGATCACCCTGCTCTCAAGGGCATTATGCAAAACAGAGAAGTCAAGCTTAAGAGG
 ATTCCAAAGGACATGATGCATATAGTCCCAACTTCAATTCTGCATAGCCTTGAAGGGATG
 CCTGGGCTGGATTGGCAGAGGCTTCTGAAGCTCCAATGCAAGTATGGATCCTTCTTGTTC
 TCCCTTCAGCTACTGCTTATGCTCTCATGACAGCCGGTGACAAGAAATGCTTCGCGTAC
 ATCGACAGGATCATTAAGAAATTCGACGGAGGCGTCCGAACGTTTACCCGGTCGATCTT
 TTTGAGCACATATGGGTTGTGATCGGTTGGAGCGTCTTGGGATATCGCGGTAATTCCAA
 CGAGAGATTGAACAGAACATGGACTATGTCAACAGGCACTGGACTGAAGATGGGATTTGC
 TGGCTAGGAAGTCCAATGTAAAAGAGTGGATGACACCGCTATGGCTTTCCGTCTACTA
 CGCCTCCATGGATACAATGTATCACCAGTGTGTTCAAGAATTTTGAAGGATGGGGAG
 TTCCTTCTGTTTTGTGGGCAATCAACTCAAGCAGTCACTGGGATGTATAACCTGAACAGA
 GCATCTCAGATAAGTTTTCCAGGAGAAGACATTTTGCAGCGTGAAGGAATTTCTCATAT
 GAGTTCTTAGAGAAAAGAGAAGCCAGGGGACACTTCATGATAAATGGATCATCTCCAAG
 GACCTACCAGGAGAGGTACAATACACACTAGATTTTCTTGGTATGCGAGCTTGGCACGC
 GTCGAGGCAAGAACATACATAGGTCAATATGGTGAAATGATGACGTCTGGATTGAAAG
 AACTCTACAGGATGCCAATTTGAATAACGCTACATATCTCGAGTTGGCGAAACAGGAT
 TTCAACCGTTGTCAAGCTTACATCAGCATGAGTTGCAAGGCTACAAAAAGTGGTTTATT
 GAGAATGGCCTGGAAGCTTTTGGGATGACACCTGAAGATGTTTTGAGAGCTTATTTTTTG
 GCTGCCGCTGCAATTTTGAACCAACCGTGCCTCTGAGCGACTTGCATGGGCTAGAGTG
 TCAGTGCTGGCCAACTATTTCTAGGCATTTT-----TACAGCGATATGTCAAGC
 ATGAAAAGG-----ATGGAGCGTTTTCATGTGGAGC-----
 -----AGCCTCTATGAAGAAAATGGCAATGTTTTGGGG-----CTAGAAGGA
 TATGCAAAAGATGGAATCCTTGCAGGACACTTTGTCAACTTATAGATTTGTTGTCTCAA
 GAGACACCGCCAGTTCGAGAAGGTCAAAGTGTATTCTATAATCTCATAAGATGTGCTTGG
 ATTGAATGGATGATGCAACAAATCAATATGAAGGATGGCAGATATGACAAAGGCAGAGTT
 ATGCATCCAGGGTCATGCACTGTTTATAATAAAGAAACATGTTTACTTATTGCTCAAATT
 GTTGAAATTTGTGCTGGACGAATTGAGGAGGCAGCATCTATGATAAATAACACCGAAGGT
 TCTTGGTTTATTCAACTTGCTTCTCTATTTGCGATTCTCTT---CATGCCAAGATGTTA
 CTTTCACAG---GATACCAAGAAAAATGAGACAACAATAAATCAA-----ATTGAC
 AAGGAAATTGAGTTGGGTATGCAAGAACTTGCTCAATATCTTCTTCCAAGAGTTGATGAT
 AGAAGAATTAACAACAAAACCAAGCAGACCTTCTTGAGCATTGTGAAAAGCTGTTACTAT
 GCTGCCAATTGCTCACCACATATGCTTGACCAACACATTTCTGAAGTGATTTTGTAGCAA
 GTTATT

altered.TRP.2.2.1.7.fasta

>p

GCTCTCTCTGCATTTTCTTTGCCCTGCTCATGTCAATAGTGTCACA-----

 -----AGAGAAGTACATGTCAAGAAAAGG---CCAAATGGGGTTTGTGCATCA
 CTATCA-----GAGAGCGGAGAGTTTCTTTCAGAGAGGCCACCAACACCTCTCTTG
 GACACCGTAAACTATCCAATTCATATGAAAAATCTATCTATCAAGGAGCTAAAACAATA
 GCAGAGGAGCTGCGGTCCGATGTTATCTTCAATGTTTCGAAAACCTGGGGTCACTTAGGC
 TCTAGCCTTGGTGTGTGGAGCTCACCCTGGCTCTTCACTATGTTTTCAATAGCCCTCAA
 GACAAGATACTGTGGGATGTTGGCCATCAGGCTTACCCTCACAAGATTCTAACTGGGAGA

TCTGATGAGCTGAGATCAGACGTGATCTTTAATGTGTGCGAAAACCGGTGGACATTTGGGG
TCAAGTCTTGGTGTTGTGGAGCTTACTGTGGCTCTTCATTACATTTTCAATACTCCACAA
GACAAGATTCTTTGGGATGTTGGTCATCAGTCTTATCCTCATAGATTCTTACTGGGAGA
AGAGGAAAG---CCTACAATGAGGCAACCAATGGTCTCTCTGGTTTCACCAAACGAGGA
GAGAGTGAACATGATTGCTTTGGTACTGGACACAGCTCAACCACAATATCTGCTGGTTTA
GGAATGGCGGTAGGAAGGGATTGAAGGGGAAGAACAACAATGTGGTTGCTGTGATTGGT
GATGGTGGCATGACGGCAGGACAGGCTTATGAAGCCATGAACAACGCCGGATATCTAGAC
TCTGATATGATTGTGATTCTTAATGACAACAAGCAAGTCTCATTACCTACAGCTACTTTG
GATGGACCAAGTCCACCTGTTGGTGCAATTGAGCAGTGCTCTTAGTCGGTTACAGTCTAAC
CCGGCTCTCAGAGAGTTGAGAGAAGTCGCAAGGGTATGACAAGCAAAATAGGCGGACCA
ATGCATCAGTTGGCGGCTAAGGTAGATGAGTATGCTCGAGGAATGATAAGCGGGACTGGA
TCGTCACTGTTTGAAGAATCTGGTCTTTACTATATTGGTCCAGTTGATGGGCACAACATA
GATGATTTGGTAGCCACTCTTAAAGAAGTTAAGAGTACCAGAACCACAGGACCTGTACTT
ATTGATGTGGTGACGGAGAAAGGTGCTGGTTATCCTTACGCGGAGAGAGCTGATGACAAA
TACCATGGTGTGTGAAATTTGATCCAGCAACGGGTAGACAGTTCAAACTACTAATAAG
ACTCAATCTTACACAACCTTACTTTGCGGAGGCATTAGTCGCAGAAGCAGAGGTAGACAAA
GATGTGGTTGCGATTGATGACCCATGGGAGGTGGAACGGGTTAAATCTCTTCAACGT
CGCTTCCCAACAGGATGTTTTCGATGTAGGAATAGCGGAACAACACGAGTTACTTTTGCT
GCGGGTTTAGCCTGTGAAGGCTTAAACCTTCTGTGCAATCTATTGCTCTTTCATGCAG
---CGTGCTTATGACCAGGTTGTCATGATGTT-----
GATTTGCAA-----AAATTACCGGTGAGATTT
GCAATGGATAGAGCTGGACTCGTTGGAGCT---GGTCCGACACATTGTGGAGCTTTCGAT
GTGACATTTATGGCTTGTCTTCTAACATGATAGTATGGCTCCATCAGATGAAGCAGAT
CTCTTTAACATGGTTGCAACTGCTGTTGCGATTGATGATCGTCTTCTTGTTCCTGTTAC
CCTAGAGGTAACGGTATTGGAGTTGCATTACCTCCGGAACAAGGTGTTCCAATTGAG
ATTGGGAAGGTAGAATTTTAAAGGAAGGAGAGAGATTGCGTTGTTGGGTTATGGCTCA
GCAGTTCAGAGCTGTTTAGGAGCGGCTGTAATGCTCGAAGAACGCGGA--TACGTAAC
GTAGCGGATGACAGGTTTTGCAAGCCATTGGACCGTGCTCTCATTGCGAGCTTAGCTAAG
TCGCACGAGGTTCTGATCAGGTTGAAGAAGGTTCCATTGGAGGTTTTGGCTCGACGTT
GTTCAGTTTCTTGCTCTCGATGGTCTTCTT---GGCAAACTCAAGTGAGACCAATGGTA
CTGCCTGATCGATACATTGATCAGGTCACCGAGCTGATCAACTAGCTGAAGCTGGACTC
ATGCCATCTCACATCGCAGCAACCGCACTTAACCTAATCGGTGCACCAAGGGAAGCTCG
TTT-----
>V
GCTCTGTGACGCTCTCATTTCTGCCCCATTTTAGCCAGGCTGCT-----GCTTCAAAT
CCTCAGAGACTTACTCCTCAGTGTTCCCATTTGTTCTTGGGGGTGGATTTGCAG-----
TGCCAATCCCAGCAAAGGAGTAAGGCCAGGAAAGG---CCAATGGG-----
-----AGACCACCAACTCCTCTCCTG
GACACTATCAATTATCCAATTCACATGAAAAATCTGTCTGTCAAGGAGCTGAAACAACCTC
GCAGATGAACTAAGGTCTGATGTTGTCTTCAATGTTTCCAAAACCTGGGGGTCACTTGGGC
TCCAGCCTCGGGGTTGTGGAGCTCACTGTGGCTCTTCATTATGTCTTCAATGCCCTCAA
GACAGGATACATGAGGATGTTGGTCTCATCAGTCTTACCACACAAAAATCTAACTGGGAGA
AGAGATCAAATGCATACCATGAGGCAACAGATGGGTTAGCGGGATTACCAAGCGCTCG
GAGAGTGAATATGACTGCTTTGGAACCGGCCACAGTTCTACTACCATCTCAGCAGGCTTG
GGA---GCGGTGCGGCCGGGATCTAAAAGGAAAAACAACAACGTTATTGCTGTATAGGT
GATGGAGCAATGCTGCAAGGCCAAGCTTATGAAGCAATGAACAATGCTGGGTACCTGGAT
TCTGACATGATTGTTATCTTAATGACAACAAGCAGGTTTCTTTACCCACTGCTACTCTA
GATGGGCCCATACACCTGTAGGAGCTTTGAGCAGTGCTCTTAGTAGGTTACAATCAAAC
AGACCTCTTAGAGAATTACGAGAGGTTGCCAAGGGCGTTACCAAACAGATTGGCGGACCG
ATGCATGAATTGGCTGCGGAAAGTTGATGAATATGCTCGTGGGATGATCAGTGTTCTGGA
TCAACACTTTTTGAAGAGCTTGGACTCTATTATATAGGTCTGTTGATGGCCACAACATA
GATGACCTGTGTTGCCATTCTCAAGGAGGTTAAGAGTACCAAGACAACAGGTCCAGTTCTG
ATCCATGTTGTACAGAGAAAGGCCGCGGATATCCATATGCTGAGAAAGCTGCAGATAAG
TACCATGGAGTGACCAAGTTGATCCTGCTACTGGAACAACATTCAAATCCAGTGCTCCT
ACTCAGTCTACACAACATATTTTGAGAGGCTTTGATTGCAGAAGCAGAGGTGGACAAG
GATATTGTTGCAATTGATGAGCAATGGGGGTTGGAACGGGCTTGAATCTCTCCATCGC
CGGTTCCCCACACGATGCTTTGATGTTGGGATAGCAGAACAGCATGCTGTTACCTTTGCT
GCTGGTCTAGCCTGTGAAGGCATTAAACCTTTTTGTGCAATCTACTCATCTTTCATGCAG
---AGAGCTTATGACCAGGCTGCCAATCCTATTACCATATCTTCTTAATCAAATGCTAT
GATTTGCAG-----AAGCTGCCAGTGAAATTT
GCAATGGACAGAGCTGGGCTGGTTGGAGCAGATGGCCCAACACATTGCGGAGCTTTTGAT
GTCACTTTCAATGGCTTCCCTTCCAAACATGGTGGTGATGGCTCCTGCTGATGAGGCCGAG
CTTTTTCACATGGTGGCCACAGCTGCTGCCATAGATGACAGGCCAGTTGTTTCCGGTAC
CCAAGAGGAAATGGGGTGGGTGTTGAACTGCCACCAGGGAACAAGGCATTCTATTGAG
GTTGGAAGGGGCGCAATATTGATTGAGGGGGAGAGAGTTGCACTCTTGGGCTATGGAACA
GCAGTACAGAGCTTGTGGTGGCTCTTCTTGTGGAACAACATGGCTTACGAATAACA
GTCGAGATGCCCGCTTCTGCAAAACATTGGACCATGCTCTTATTGCTAGCCTAGCAAAA
TCACATGAAGTTTTGATTACAGTAGAAGAAGGTCATTTGGTGGTTTTGGGTCTCATGTT
GCTCAGTTTTTGGCCCTTAATGGTCTTCTTATGGCACAACAAGTGGAGTCCCATGGTT
CTTCTGACCGGTACATAGACCATGGAGCGCCAGCAGACAGTTGGCCATGGCGGGTCTG
ACACCATCTCATATTGCAAGCAACAGTATTCAATATACTTGGACAAAACAAGGAGGCCCTG
GAGATCATGTAA-----
>0
GCGCTCACGACGTTCTCC-----ATTTGAGAGGAGGCTTCGTGCGCGCGCTG
CCGACGAGGAGGCTATTCGCTCCGGCGGCGGAGCTCAGTCTCCAC-----

AAGCTCCAGAGCAGGCCACACAAGGCTAGGCGGAGGTCGTCGTCGAGCATCTCGGCGTCG
 CTGTCCACGAGAGGGAGGCGGCGAGTACCACTCGCAGCGGCCACCGACGCCGCTGCTG
 GACACGGTCAACTACCCCATCCACATGAAGAAGCTGTCCCTCAAGGAGCTCCAGCAGCTC
 GCCGACGAGCTCCGCTCCGACGTCATCTTCCACGTCCTCAAGACCGGGGACATCTCGGG
 TCCAGCCTCGGCGTCGTCGAGCTCACCCTCGCGCTCCACTACGTGTTCAACACGCTCAG
 GACAAGATCCTCTGGGACGTCGGCCACCAGTCGTACCCTCACAAGATTCTGACCGGGCGG
 CGCGACAAGATGCCGACGATGCGTCAGACCAACGGCTTGTGGGATTACCAAGCGGTGCG
 GAGAGCGAGTACGACTCCTTCCGACCGGCCACAGCTCCACCACCATCTCCGCCGCCCTC
 GGGATGGCGGTGGGAGGGATCTCAAGGGAGGGAAGAACAACGTGGTGCGGTGATCGGC
 GACGCGGCCATGACGGCCGGGAGGCGTACGAGGCGATGAATAACGCGGGGTATCTCGAC
 TCCGATATGATCGTGATTCTCAACGACAACAAGCAGGTGTCGTCGCCGACGGCGACGCTC
 GACGGGCGGCGCGGCCGCTGCGCGCTCAGCAGCGCCCTCAGCAAGCTGCACTCCAGC
 CGCCCACTCAGGAGCTCAGGAGGTGGCAAGGGCGTGACGAAGCAAATCGGAGGGTCG
 GTGCACGAGTGGCGGCGAAGGTGGACGAGTACGCGCGGGCATGATCAGCGGCTCCGGC
 TCGACGCTCTTCGAGGAGCTCGGCTCTACTACATCGGCCCGCTCGACGGCCACAACATC
 GACGACCTCATCACCATCTCCGCGAGGTCAAGAGCACCAAGACCACAGGCCCGGTGCTC
 ATCCACGTGCTCACCAGAAAGGCCGCGCTACCCCTACGCCGAGCGCGCGCGACAAG
 TACCACGGCGTGGCAAGTTTCGATCCGCGACGGGGAAGCAAGTTCAAGTCGCCGGCGAAG
 ACGCTGTGTACACGAAGTCTTCGCGGAGGCGCTCATCGCGAGGCGGAGCAGGACAAC
 AGGGTGTGGCCATCCACGCGGCCATGGGGGAGGCACGGGGCTCAACTACTTCTCCGCG
 CGCTTCCGAAACCGTGCTTCGACGTCGGGATCGCGAGCAGCAGCCGTACGTTCCGCC
 GCCGCCCTCGCTCGAGGGCCCTCAAGCCGTTCTGCGCCATCTACTCCTCTCTGACAG
 ---AGAGGCTACGACAGGTGGTGACGACGTG-----
 GACCTCCAG-----AAGCTGCCGGTGAGGTTT
 GCCATGGACAGGGCGGGCTCGTGGGCGCCGACGGCCGACGCACTGCGGCGCGTTTCGAC
 GTCACCTACATGGCGTGCTGCCGAACATGGTGTGTCATGGCCCGTCCGACGAGGCGGAG
 CTCTGCCACATGGTCGCCACCGCGCGGCCATCGACGACCGCCCTCCTGCTTCCGCTAC
 CCAAGAGGCAACGGCATCGGCGTCCCGCTACCAACCACTACAAAGGCGTTCCTCTCGAG
 GTAGGCAAGGGGAGGTTACTGCTGGAGGGCGAGAGGGTGGCGCTGCTTGGGTACGGTTCG
 GCGGTGCAAGTCTTCCCTGCGTCCGCGACGTCGCTGGTGAGCGGACGGCCTCAAGGTGACC
 GTCGCGACGCGAGGTTCTGCAAGCCGCTGGACCAACGCTCATCAGGAGGCTGGCCAGC
 TCCACGAGGTGCTCCTCACCCTCGAGGAAGGCTCCATCGGCGGGTTCGGCTCCACGTC
 GCGCAGTTATGGCCCTCGACGGCCTCCTCGACGGCAAACTCAAGTGGCGGCGCTGGTG
 CTACCCGACAGGTACATCGACACGAGGTCACCGCGGATCAGCTGGCGGAGGCGAGGGCTG
 ACGCCGTCGCACATCGCGGCGACGGTGTTCACAGTGTGGGCCAGGCGAGGGAGGCGCTC
 GCCATCATGACGGTGCCCAACGCG

altered.TRP.2.5.1.10.fasta

>P

TTTGATTCAAGTCTTATATGATACAGAAAGCCAATTCTGTTAAACAAGCATTAGATACT
 GCTGTTACTCTTAAAGAACCG---GCTAAAATCCATGAGTCTATGCGTTACTCTCTTTTG
 GCTGGTGGCAAGAGGGTTCGGCCAGTGCTTTGTCTAGCTGCATGTGAGCTTGTGGTGGG
 TCTGAATCCATGGCTATGCCTGCTGCTGTGTAGAAATGATCCATACTATGTCATTA
 ATACATGATGATCTTCCCTGCGATGGATAATGATGATCTTCGCCGCGGAAAATCCACCAAT
 CATATTGTTTTTGGTGAGGATGTTGCGGTTTTGGCAGGGGATGCTTTACTGGCATTGCA
 TTTGAACATATTGCAGTG-----TCTACACTCAATGTTTCGCCCT
 CTTAGAATTGTTGCTGCAGTTGGTGAATTGGCGAAAGTATTGGTGCA---GGACTTGTT
 GCTGGACAAGTTGTGGATATTTGTTCTGAAGGGTTG-----TCCGAAGTGGGGTTAGAA
 CAGCTTGAATTTATTATATTATTAAGACTGCTAAGTTGTTGGAAGGTGCGGTTGTTTTA
 GGGGCTATATTAGGTGGAGGGACCGATGAGGAAGTTGAGAAATTGAGGCGATACGCGAGG
 AGTATTGGATTGTTGTTTCAAGTAGTGGATGATATTCTTGATGTTACCAAATCTTCACAA
 GAATTGGGAAAACTGCAGGGAAGATTGTTGCGGATAAAGTTACTTATCCTAAGTTA
 ATGGGAATTGAGAAGTCTAGGGAGTTTGTGAGAAGTTGCTTAACGAAGCTAGGGAGTTG
 CTTGCTGGATTTGATCAAGAGAAGGCGGCTCCGTTGATTGCTTTGGCTAATTACATTGCT
 TACAGGCAAAAC

>R

TTTGATTCAAGTCTTACATGCTCCAAAAAGCTGCTTCCATCAATCAAGCTTTAGATGCT
 GCTGTTCCGATTGCTGAACCA---ACTAAAATTCATGAATCTATGAGGTATTCTCTTTTG
 GCTGGTGGCAAAAGGGTCAGACCAGCACTTTGTTTAGCTGCTGTGAACTGGTTGGTGGC
 AATGAATCTATGGCTATGCCTGCTGCTGTAGAAATGATTCACTATGTCACCT
 ATTATGATGATCTTCTTGTATGGATAATGATGATTTACGCCGTGGAAAGCCCACTAAT
 CATATTGTTTTTGGTGAGGATGTGGCTGTTCTTGCAGGTGATGCATTATTAGCACTCGCT
 TTTGAACATATCGCTATA-----GCCACTTTAAATGTTTCTCCT
 GCTAGAATTGTTAGAGCAGTTGGGAATTAGCTAAAGCTATTGGGCTGAAGGGTTAGTT
 GCTGGTCAAGTTGTTGATATATGTTCTGAAGGATTA-----TCGGAAGTGGGCTTAGAA
 AAACCTGAATTTATTATGCTCCATAAGACCGCAAAGTTGCTGGAAGGGGCTGTGGTGTTA
 GGGGCTATAATGGGTGGAGGGACTGACGAGGAAGTTGAGAACTGAGGAAATATGCTAGG
 GATATTGGGCTGTTGTTTCAAGTGGTGGATGATATTCTTGATGTCACCAAATCATCTCAA
 GAGTTGGGGAAGGCTGACGGGAAGGATTGGTGGCAGATAAGGTTACATATCCGAAGTTG
 ATGGGAATTGACAAGTCGAGGGAATTTGCTGAGAAGTTGAATAAGGAAGCTCAAGAGCAG
 TTGGCTGGGTTTGATCTGAGAAGGCTGCTCCATTGATTGCTTTGGCTAATTACATAGCT
 TACAGGCAAAAC

>A

TTCGATTTACGTCATACATGATCGGAAAAGCTAATGCCGTCAATGAAGCTCTCGACTCC
 GCCGTCTCGCTCCGAGAACCA---ATCAAGATCCATGAAGCGATACGTTACTCGCTTCTC
 GCCAGAGGCAAAACGAGTGAGACCCGTGCTCTGTATCGCAGCGTGG---TTAGTGGGCGGC
 GAAGAATCCGTGGCGTTACCGGCGCGTGCGCCGTGGAATGATCCACACAATGTCGTTG
 ATCCACGACGATCTTCCATGTATGGATAACGACGATCTCCGCCGCGGAAAACCAACGAAC
 CACAAAGTTTTCCGTTGAAGACGTGCGGTTTTAGCCGGAGACGCGCTTATATCGTTCGCG
 TTTGAGCACTTAGCAACG-----TCAACG---GCTGTTTCTCCG
 GCTAGAGTTGTTGAGCAATCGGAGAGTTAGCGAAAGCCATTGGATCAAAAGGGCTCGTG
 GCGGGTCAAGTAGTGGATCTAACCAGTGGAGGAATGGACCAAAACGACGTGCGCTTAGAG
 GTTCTTGAGTTTATCCACGTTCAAAAACCGCGTGTGTTAGAGGCGGCTACGGTGTTA
 GGAGCCATCGTTGGTGGTGGTTCGTATGAAGAAGTTGAGAAGCTGAGAAGATTCGCGAGA
 TGCAATTGGATTGTTGTTTCAGGTGGTGGACGATATTTGGACGTGACGAAGTCGTGCGAG
 GAATTGGGAAAGACCGCGGAAAAGATCTT---GCCGATAAGCTGACGTATCCGAAACTA
 ATGGGTCTTGAGAAATCTAAAGACTTTGCAGATAAGTTGTTGAGCGATGCTCACGAACAG
 CTTTCATGGGTTTGATTGAGTAGAGTTAAACCTTTACTAGCTTTGGCTAATTACATTGCC
 AAAAGACAGAAT

>V

TTCAATTTCAACGCTTACATCCTTGATAAGGCTAATTCGGTCAACAAAGCTTTGGACGAC
 GCCGTTCCCATTAGGGAACCT---GTGAAGGTCCACGAGTCGATGCGCTACTCCCTTCTC
 GCCGGCGGCAACGCTGTCCTGCCGTTCTCTGTATCGCGGCTGCGAGCTGGTGGGCGGG
 GATGAAGCTACGGCGATGCCGCTGCTTGTGCGGTGGAGATGATTCACACCATGCTTTTA
 ATGCATGATGACCTTCTTGC---GACAACGACGATCTACGCCGTGGAAGGCCACAAT
 CACAAGGTCTTCGGTGAGAATGTTGCCATTCTCGCCGGAGACGCTCTTCTGGCGTTTGC
 TTCGAGCACATGGCCACA-----GCGACGGTTGGAAGTTCCGCCG
 GGGTGGATAGTTGTCGAGTCGCGCAATTGGCGAAATCGATTGGGTGAGAAGGACTTGTT
 GCAGGGCAAGTGGTTGATATGTTCTGAGGGCTTG-----AAAGATGTTGGATTGGAA
 CATCTCGAATACATCCATGTTCAAGACCGCAGCTCTACTTGAGGGGGCGGTAGTTCTT
 GGGGCGATTCTCGGAGGAGGATCAATGAGGAAATTGAAAAATTGAGGAAATTGCCAGA
 TGTATAGGGTTGCTCTTTCAGGTGGTTGATGATATTCTGGATGTTACCAAATCCTCCAA
 GAGTTGGGGAAGACTGCAGGCAAGATTGTTGGTGGCTGATAAGCTCACGTACCCGAAGCTT
 CTGGGTATTGAGAAATCTCGGGAATTAGCTGAGCAACTGAATAAGGATGCCAAGGACCAG
 CTCTCCGGCTTCGATCCAGACAAGGCCGCCCTTTGATTGCTTTATCTAATTATATTGCT
 TACAGGCAAAAT

>0

TTTGATTTTAACGCGTATATGGGTGAGAAGGCTGCGGCGGTGAACCGGGCGCTGGACGCG
 TCGATCCCGGCGGACGAGCCGCCCGCGGCGCTCCACGAGGCGATGCGGTACGCGCTGCTG
 GCGGGCGGGAAGCGGGTGCGGCCCGCGCTGTGCCTGGCGCGTGCGCGGTGGTCGGGGGG
 CGGGAGGCTGGCGGATGCCCGCCGCCGCCGCGCTCGAGATGGTGCAACCC---TCGCTC
 GTCCACGACGACCTCCCTGCATGGACGACGACGACCTCCGCCGCGGGAAGCCACCTGC
 CACGTGCTCTACGGGAGGCCATCGCCGTGCTCACCGGCGACGCGCTGCTCTCCCTCTCC
 TTCCACCACATGGCCAGGTTGACTCCTACCTCCGGACATCGACGCTGACAAGCACCCC
 GCCCGCGTCTGTCGCGGCTTGGCGAGCTGCGCGCTGATAGGCTCCGAGGGCTTAGTC
 GCCGGCCAGGTTGTTGATCTTGAGATGACTGGATCG---ACCGAACTGTACCTCTTGAA
 CGTCTTGAGTACATCCATCTCCACAAAAGTGTGCTGCTTGGAGGATCAAGTGGTTATT
 GGGGCAATCTTGGGAGGTGGCTCTGATGAGCAGATTGAAAGTTTGCGCATGTATGCGAGA
 TCGATAGGATTGTTGTTCCAGGTTGTTGATGATATACTTGATGTGACCAAGTCTTCTGAG
 GAGCTCGGCAAGACAGCAGGGAAGGACTTGGCGAGTGACAAGACGACATACCCGAAATTA
 CTTGGGCTGGAGAAATCACGGGAGTTTGAGAAAAGTTGCTTTCTGATGCAAGGGAACAA
 CTTTCAGGATTTGATCAAGAGACCGCAGCACCCTTCTGCACCTGGCCAATTATATTGCC
 TATCGGCAGAAC

altered.TRP.2.7.1.148.fasta

>P

 -----CTGGTGTATGATCCTGATAGAAGGGTAAACAAGTTGGCGGATGAGGTGGAT
 AAACAGGCTCCTCTCTCAAGGCTCACTCTTTTCTACCTTGCAAGGTTAATGTCTTCTTG
 AGAATAACCAAGTAAGAGAGAAGATGGGTATCATGATTTGGCATCTCTATTTTCATGTGATT
 AGTCTCGGAGATGTAATTAAGTTCTCTTTGTCAACATCAAAAAGAATTGATCGATTATCA
 ACCAATGTGTCTGGTGTACCTCTTACTGAAGATAATTTGATAATTAAGCCCTTAACCTT
 TACAGGAAAAAGACTGGCACTGACAATTTCTTTTGGATTCAATTTAGACAAAAAGGTGCC
 ACTGGAGCAGGACTTGGTGGTGGGAGCAGCAATGCCGCTACTGCCCTGTGGGCGGCAAT
 CAATTCAGTGTTGCTTGTCTACTGAGAAGGAACTCCAAGAATGGTCAAGTGAGATTGGT
 TCAGATATCCCTTTCTTTTCTCATGGAGCAGCCTACTGTACTGGCGGTGGTGAGATT
 GTTCAAGATATTCCTTACCAGTATCTTCAGGCATTCCAATGGTTCTCATGAAGCCCCCG
 CAGTCATGTTCAACTGCAAGAAGTTTACAAGTGCTTTCAATTGGATAAAACAGTCAAGCT
 GATCCTTTAACATTGCTGGAGAAGATCTCAAGGAATGGAATATCTCAA---GTATGTATC
 AATGATTTGGAACCTCTGCTTGAAGTTCTCCCTCTCTCAAAAGATTAACAGCGT
 ATAGCTGCAGCTGGTGTGACAATATGATGCTGTTTTCATGTCTGGAAGTGAAGCACA
 ATAGTTGGGATTGGCTACCAAGATCTCCGCAATTTATATATGACGAAGATGAGTACCAG
 GATGTTTTGTGTCAGAAGCCAACCTTCTGCTGTAAGCAATCAATGGTATCAACAA
 CCTGCTCAACAACT---TGATGCTCT-----CCACCCGAATTTCCGGTCTGAG

>R
 -----GCTTCTGCTCATTTCTCTGCAACAGCCATCTCTTTCAACCTTCTTGCAATTCC
 TTTGCTAAAACTAGTCTTCCCTTCATTTAAGCCAGCTGGGTCTATTTCTTTT-----
 CCTCGAAGCACAAAGAATCCCATTTATCAAA-----GCC-----ACCAAGAAA
 CAGCTAGAGATAGTGTATGATCCCGATGAAAGGTTAAACAAGTTAGGAGATGAAGTAGAC
 AAGATTGCTCCTCTTTCAAGGCTCACTTTGTTCTCTCCTTGCAAGATTAATGTTTTCTCG
 CGAATAACTGATAAAAGAGAAGATGGGTATCATGATTTGGCATCTCTCTTTCATGTAATT
 AGTCTAGGAGATACAATTAAGTTCTCTTTATCTCCTTCAAAATCAACAGACCGTTTATCA
 ACTAATGTGGCTGGCGTTCCCCTTGATGAAAGGAATTTGATTATTAAGCCCTTAACCTA
 TATAGGAAAAAGACGGGCACTGACAACCTCTTTTGGATTCTATAGATAAGAAGGTGCCT
 ACTGGGGCAGGGCTTGGTGGTGGGAAGCAGTAATGCTGCGACTGCATTGTGGGCGGCAAA
 CAGTTTCAGTGGCGGTCTTGCCACTGAGAAGGACCTTCTAGAATGGTCAAGTGAGATTGGT
 TCAGATATCTCCTTCTTTTTTCCCGTGGAGCAGCCTATTGACTGGTGGTGGTGAATTT
 GTCCAAGATATTCTTCTCCAGTTCTCTTGACATTCCAATGGTTCTTATAAAGCCCCAG
 GAAGCCTGCCAACTGGTGAAGTTTACAAGCGCTTTCGGTTGGATCAAACAGTCAAGTT
 GATCCTTTAAGACTTCTGGACAAGATCTCAAAGAGTGAATATCTCAAGATGTTTGTATA
 AATGACTTGGAACTCTCGCATTTGAAGTTCTCCCATCTTAAAAAGATTAAACAGCGT
 ATAGTTGCAGCAAGCGTGGACAATACAATGCAGTTTTATGTCTGGAAGTGGTAGTACC
 ATAGTTGGGATTGGTTCGCGGGATCTCCACAGTTTATTTATGACGACGATGACTACAAG
 GATGTTTTCGTGTCAGAGGCTAACTTCTTGACTCGTGAAGCAATGAGTGGTACAAAGAA
 CCTGCTTCAACTGCTGCTTGTGGCACA-----GAGTCTGATTTTCCCAATCTGAG

>A
 GCAACGGCTTCTCTCCATTTATCTCAACTCTCAGCTTC-----
 -----ACTACTCTTCTTTCAAAACTTCTTCTTCTTCTTCTTCTTCTTCTCCGAAG
 CTTCTTCGA-----CCCCTTTAAGCTTTTCCGTCAAAGCT-----TCCAGAAAG
 CAAGTAGAGATAGTGTTTGATCCT--GAGAGGCTTAATAAGATAGGTGATGATGTTGAC
 AAAGAAGCTCCTTTGTCAGGCTTAAGCTCTTCTCACCTTGAAGATCAATGTTTTCTTG
 AGGATCACCGGAAGCGGAGAGATGGGTTTCATGATTAGCCTCTTTGTTTTCATGTGATT
 AGCTTAGGAGACACTATTAATTCTCATTGTCAACATCAAAGTCTAAAGATCGTTTGTCT
 ACTAACGTTCAAGGAGTCCCTGTTGGG--AGAAATCTGATTATAAAGCACTTAACCTT
 TACAGGAAGAAAACCTGGTAGTAACAGATTCTTCTGGATTCAATTAGATAAGAAGGTGCCT
 ACCGGGGCTGGACTCGGTGGTGGAAAGTAGTAATGCTGCAACTGCACCTGCGGCGGCAAT
 GAGCTCAATGGAGGTCTTGTCACTGAGAACGAACCTCCAGGATTGGTCAAGTGAATTTGGG
 TCAGATATTCTTTCTTCTTCTCGCGA--GCTGCCTATTGTACCGGAGAGGTGAGATT
 GTCCAAGACCTTCTCCACCTTTTCTCTTGATCTTCCGATGGTGCTCATAAAGCCCCGA
 GAAGCATGTTCCCACTGCTGAAGTTTACAACGCTCTTCGTTTAGATCAGACGAGCAATATT
 AATCCCTTGACATTACTAGAGAATGTGACCAGCAATGGTGTGTCTCAAAGCATATGCGTA
 AACGATTTGGAACCGCCAGCGTTTTCAGTTCTTCCATCTCTAAAACGCTTGAAGCAACGG
 ATAATAGCATCTGGACGTGGGAATACGATGCTGTGTTATGTCTGGGAGTGGAAAGCACT
 ATTATCGTATTGTTCCACAGATCCTCCTCAATTTATATATGATGATGAAGAATACAAG
 AACGTGTTCTTGTCTGAAGCAAACTTTATGACGCGTGAGGCTAATGAATGGTACAAAGAA
 CCTGCTTCTGCAATGCTACTACCTCA-----TCCGCCGAATCTCGCATGGATCAA

>V
 GCTATGGCTTCATCACAATTCTTCTGCGATCACCAACTT--CACCTTCTCTGCAATTCC
 ATCAGAAGAACCACCTCCCTTCATTTAGACCACATGGGTCACTTCTTTTCGTCCAAAA
 CCCCAATTTTCAGAGGACTCCATTTCTCAGAGTCAGGGCTTCTGCT-----GGTAGAAAA
 CAAGTGGAGATAGTGATAATCCAGATGAAAGGATAAACAATCTAGCAGATGAAGTGGAT
 AAGAACGCTGGGCTTTCAAGGCTCACTCTATTTTACCTTGCAAGATTAATGTTTTCTCG
 AGAATAACCGGTAAAGAGAGCAGATGGATTTCATGATTGGCATCTCTTTTCATGTAATC
 AGTTTAGGAGATATAATTAAGTTCTCTTTGTGCGCATCAAAAACAAAGGATCGCTTATCA
 ACCAATGTGCCAGGGGTCCCCTTGATGAAAGTAATTTGATCATAAAGGCACTTAACCTT
 TACAGGAAAAAACTGGCACTGATAACTACTTTTGGGTTTCTGATAAAAAGGTGCCG
 ACTGGGGCAGGGCTTGGTGGTGGCAGCAGTAATGCTGCAACAGCACTGTGGGCGCAAA
 CAATTTAGTGGTTGCCCTTGCCACTGAGAAGGAACCTCAACAATGGTCAAGTGAGATTGGT
 TCGGATATTCCATTCTTCTTCTCTAATGGAGCAGCTTTTGTACTGGCAGAGGTGAGGTT
 GTTGAAGATATTTCCCTGCCAATACCTTTGGACATTCCAATGGTTCTTGTAAGCCACCG
 CAGGCATGCTCCACCGCTGAAGTTTACAAGCACCTTCGGTTGGATCAAACAGTAAGGTT
 GATCCCATAACCTTGCTTGAGAAGATCTCAAAGAACAGAATATCTCAAGATGTTTGTATA
 AATGATCTAGAACCTCCTGCATTTGAAGTCTCCCATCTCTGAAAAGATTAAACAGCGT
 ATAATTGCAGCAAGCCGTGGACAATATGATGCTGTTTTCATGTCTGGGAGG-----

 -----TATAAA---
 -----CAG

>O
 ---GCTTGCTCCACCCACCTCCTGTCCAGAGCCTCTACCCGCTCAACCGCGCTAACCCG
 GCAGCCGCGCGCGGACCTCCGGTTCCAGGCTTCTCCAGTGTGAGGCTCGGCTCCGGC
 ACCAGTCCGCGCGCGGCTGGGCTGAGAGTCGCGCGCTCGGCGGAGCAAGGGAGGAGG
 CAAGTGGAGGTTGAGTATGATCTACAAGCAAAGTTCAACAAGCTAGCGGACCAAAATTGAC
 CAGAATGCTGGGATTACACGGCTGAACCTATTCTCACCTTGCAAAATTAATGTGTTCTTG
 AGGATAACGGGTAAAGAGACCAGATGGGTTCCATGACCTGGCTTCTCTGTTTCATGTGATA
 AGTTTGGGTGATCATCAAAATCTCATTGTCAACCAAGTAAGAGCAAAGATCGCTTGTCA

ACCAATGTAGCAGGTGTCCAGTTGATGAAAGCAATTTGATCATCAAGGCACTCAATCTT
TACCGCAAGAAAAGTGGAACTGACAACTTTTTTTGGATTCATCTTGATAAGAAGGTCCCT
ACTGGTGCTGGTCTTGGTGGTGGAAAGCAGTAATGCTGCAACTGCGCTGTGGGCCGCCAAC
CAGTTTAGTGGCTGCATTGCTTCAGAAAAGGAGCTTCAGGAGTGGTCTGGAGAGATAGGA
TCGGATATTCCCTTCTTCTTTTCAAGGAGCAGCGTATTGTACCGGTAGAGGAGAGATT
GTTGAAGATATTCGGAATCCATTGCCAGCAAATTTGCCGATGGTACTAGTAAAGCCACCT
GAAGCATGCTCAACAGCTGAAGTTTACAAGCGGCTCAGGTTAGAGCACACAAGTCAAAC
GATCCCTTGGTATTGCTCAAGGAAATTAAGTAAATGGGATATCACAGGATGCCTGTGTT
AATGATCTAGAACCTCCAGCATTGAGGTGTTGCCATCACTAAAGAGGTTGAAGAAACGT
ATAATTGCTGCTAACCAGGAGATTATGATGCTGTTTTATGTCAGGAAGTGGAAAGCACA
ATTGTTGGGATTGGTTACCAGATCCGCCTGCATTTGTGTATGATGATGATGACTACAAG
GATACTTTTGTGTCAGAGGCCTGCTTCCTCACTCGTAATGAGAACGAGTGGTACAGAGAA
CCAATCTCATCGAAAATCACTAGCGAGGAAGATTTACCTCCGGAGGTAGCATCAGTTGAC